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

TECHNICAL REPORT NO. 259

May 1975

ON MINIMAX MULTISTAGE ELIMINATION TYPE RULES FOR SELECTING THE LARGEST NORMAL MEAN

by

Ajit C. Tamhane

Prepared under contracts

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

and

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

TABLE OF CONTENTS

PAGE

LIST OF TABLES

Table PAGE

LIST OF ILLUSTRATIONS

CHAPTER 1

INTRODUCTION AND SUMMARY

§1.0 A brief historical background:

In many practical situations an experimenter is faced with the problem of selecting one (or more) out of $k \geq 2$ possible categories. Typically the categories under study are characterized by a certain (numerical) parameter and the experimenter is interested in choosing the category associated with the largest (or smallest) parameter. Thus, e.g., in agricultural field trials, an experimenter may be interested in choosing that variety of grain which has the highest mean yield. Or in a scientific laboratory, an experimenter may wish to choose that measuring instrument which has the least variability.

In spite of the fact that the experimenter may ultimately be seeking to make a selection decision in the above types of experimental situations, conventionally the test of homogeneity approach had usually been employed. In this latter approach one tests the null hypothesis that all of the categories are the same (with respect to the characterizing parameter) using some appropriate test such as the F test for homogeneity of means or Bartlett's test for homogeneity of variances. Whether the test employed yields statistically significant results or not, it is clear that it does not supply the information that the experimenter truly seeks.

The inappropriateness of tests of homogeneity for the problems which are inherently selection problems, was realized by some early writers such as Mosteller [1948] who treated so called "slippage" problem and

Paulson [1949, 1952a, 1952b] who treated problems relating to classification schemes, comparison with a control category and slippage tests. But Bechhofer [1954] was the first writer to precisely formulate certain possible objectives of the experimenter viz. a complete ranking of the categories or selection of the category having the largest parameter value etc. He proposed the so-called "indifference-zone approach" to the ranking and selection problem. Since Bechhofer's paper this area of statistical research has received considerable attention. As we study a specific problem in the subsequent chapters of the thesis, we shall at that time review in some detail the literature pertinent to that problem. The interested reader may find a complete bibliography of the research work in this field until 1968 in the excellent monograph by Bechhofer, Kiefer and Sobel [1968].

Another approach to the selection problem,the so-called "subset approach," was proposed by Gupta [1956, 1965]. Here the objective of the experimenter is to select a (preferably small) subset of the populations which contains the category with (say) the largest parameter value. It is implicitely presumed that the selected subset of the populations would be subjected to further intensive study. Thus the subset selection rules may be regarded as "screening" rules. For the most recent survey of the literature dealing with the subset selection approach the reader may refer to Gupta and Panchapakesan [1972]. Recently Santner [1973] developed a "restricted subset approach" to the selection problem. His formulation provides a unifying theory to the two approaches proposed by Bechhofer and Gupta.

When employing either the indifference-zone approach or the subset approach, the experimenter requires a rule which makes a correct decision with at least a certain preassigned probability. This latter quantity is analogous to the "power" of the test in hypothesis testing. Instead of setting a preassigned lower bound on the probability of making a correct decision, Somerville [1954] treated the selection problem by introducing a loss function which took into account the loss incurred due to a wrong decision and the cost due to sampling. He proposed using a rule which minimizes the maximum of the total expected loss.

In the above we have mentioned only a few of the important papers in the development of the ranking and selection theory. Preliminary to motivating the main theme of the present thesis, we next describe a specific problem associated with selecting the population (category) having the largest population mean from k normal populations. We shall also state some single-stage (fixed-sample size) and fully sequential rules that have been proposed to solve this problem and mention their merits and demerits.

§1.1 The normal means problem:

§1.1.1 Assumptions and notation:

Let $\Pi_1, \Pi_2, \ldots, \Pi_k$ be $k \geq 2$ normal populations with unknown population means $\mu_1, \mu_2, \ldots, \mu_k$ $(-\infty < \mu_i < \infty,\ 1 \leq i \leq k)$ and a common known population variance σ^2 $(0 < \sigma^2 < \infty)$. Let Ω denote the space of all parameter vectors $\underline{\mu} = (\mu_1, \mu_2, \ldots, \mu_k)^t$. Further let $\mu_{[1]} \leq \mu_{[2]} \leq \cdots \mu_{[k]}$ denote the ordered values of the μ_i. We assume that the experimenter has no prior knowledge regarding the correct pairing of the Π_i with the $\mu_{[j]}$ $(1 \leq i,j \leq k)$. We denote by $\Pi_{(j)}$ the population associated with $\mu_{[j]}$. If the r largest means $\mu_{[k]}, \mu_{[k-1]}, \ldots, \mu_{[k-r+1]}$ are equal then any one of the populations $\Pi_{(k)}, \Pi_{(k-1)}, \ldots, \Pi_{(k-r+1)}$ is regarded as "best."

The experimenter's goal is to select the best population. Throughout the present thesis we denote by $\Phi(\cdot)$ the standard normal cdf and by $\phi(\cdot)$ the corresponding density function.

§1.1.2 The indifference-zone approach and some associated rules:

In the following indifference-zone approach due to Bechhofer [1954], it is assumed that the experimenter specifies two constants $\{\delta^*, P^*\}$, $\delta^* > 0$ and $1/k < P^* < 1$ prior to the start of experimentation. Let $\Omega(\delta^*) = \{\mu\varepsilon\Omega | \mu_{[k]} - \mu_{[k-1]} \geq \delta^*\}$. $\Omega(\delta^*)$ is called the preference zone; the indifference zone being $\Omega_0(\delta^*) = \Omega - \Omega(\delta^*)$.

In this formulation the goal of the experimenter is to select any one of the populations associated with $\mu_{[k]}, \mu_{[k-1]}, \ldots, \mu_{[k-r+1]}$ if the r largest means are equal. The event of selecting any one of the populations $\Pi_{(k)}, \Pi_{(k-1)}, \ldots, \Pi_{(k-r+1)}$ when $\mu_{[k]} = \mu_{[k-r+1]}$ is called a correct selection w.r.t. this goal and is denoted by CS. The experimenter restricts attention to only those rules R which satisfy the probability requirement

$$(1.1.1) \qquad P_{\underline{\mu}}(CS|R) \geq P^* \quad \forall \underline{\mu} \varepsilon \Omega(\delta^*).$$

The intuitive meaning of the indifference zone now becomes clear. Since the probability requirement need not be guaranteed over the indifference zone $\Omega_0(\delta^*)$, the experimenter is, in a sense, indifferent to which population is selected when $\underline{\mu}$ lies in $\Omega_0(\delta^*)$.

We shall now describe three rules proposed for the above problem which are relevant to our thesis. Each rule tells how to sample, when to stop sampling and what terminal decision to make. Each rule is designed to guarantee the probability requirement (1.1.1).

Bechhofer [1954] proposed the following single-stage rule.

Bechhofer rule (R_0): Observe the sample means $\bar{X}_i$ $(1 \leq i \leq k)$ based on n_0 mutually independent observations from each Π_i and assert that the population associated with $\bar{X}_{[k]} = \max_{1 \leq i \leq k} \bar{X}_i$ is best. Here n_0 is the smallest positive integer which guarantees (1.1.1).

Let $d_0 = d_0(k,P^*)$ be the solution to the equation

$$\int_{-\infty}^{\infty} \Phi^{k-1}(x + d_0)d\Phi(x) = P^*. \tag{1.1.2}$$

Then n_0 is given by

$$n_0 = \left[\left(\frac{d_0\sigma}{\delta^*}\right)^2\right] + 1, \tag{1.1.3}$$

where $[x]$ denotes the largest integer $\leq x$. Bechhofer [1954] has provided tables of $d_0(k,P^*)$ for selected values of k and P^*.

Bechhofer, Kiefer and Sobel [1968] proposed the following open-ended, non-eliminating type of fully sequential rule.

Bechhofer, Kiefer and Sobel rule (BKS): Observe at each stage of the experiment a single random variable (r.v.) X_{ij} from each $\Pi_i (1 \leq i \leq k)$ where j is the stage index ($j = 1,2,\ldots$ ad. inf.). Let

$$Y_{im} = \sum_{j=1}^{m} X_{ij} \quad (1 \leq i \leq k) \tag{1.1.4}$$

denote the cumulative sums at the mth stage. Let $Y_{[1]m} \leq Y_{[2]m} \leq \cdots \leq Y_{[k]m}$ denote the corresponding ordered values of the Y_{im}. Define

(1.1.5) $$Z_m = \sum_{i=1}^{k-1} \exp\{-\frac{\delta^*}{\sigma^2}\left(Y_{[k]m} - Y_{[i]m}\right)\}.$$

Terminate the sampling at the first value of m such that

(1.1.6) $$Z_m \leq \frac{1-P^*}{P^*},$$

and assert that the population associated with $Y_{[k]m}$ is best. Sampling is continued if (1.1.6) is not satisfied.

Since there is no finite preassigned upper bound on the number of stages, the BKS rule is referred to as an open-ended rule. Further it is a non-eliminating type of rule since all populations are carried through the entire experiment.

Paulson rule (P): Paulson [1964] proposed a class of closed, eliminating type fully sequential rules $\mathscr{P}_\lambda$. For any fixed λ, $0 < \lambda < \delta^*$, define

(1.1.7) $$a_\lambda = \frac{\sigma^2}{(\delta^*-\lambda)} \log_e\left(\frac{k-1}{1-P^*}\right),$$

and let

(1.1.8) $$W_\lambda = [a_\lambda/\lambda].$$

At the first stage observe $X_{i1} (1 \leq i \leq k)$ and let $X_{[k]1} = \max_{1 \leq i \leq k} X_{i1}$. Retain the population Π_i for further sampling if $X_{i1} \geq X_{[k]1} - a_\lambda + \lambda$ $(1 \leq i \leq k)$. If there is only one population retained then stop sampling and assert that, that population is best. Otherwise take another observation from each retained population and repeat the process. In general, if at the mth stage $(m \leq W_\lambda)$, S_m is the subset of the populations still retained with at least two populations in S_m then define

$$(1.1.9) \qquad S_{m+1} = \{i \varepsilon S_m | Y_{im} \geq \max_{j \varepsilon S_m} Y_{jm} - a_\lambda + m\lambda\}.$$

If S_{m+1} consists of a single population then stop sampling and assert that, that population is best; otherwise continue sampling. If more than one population remains after the W_λth stage then take one more observation from each of these remaining populations and then terminate the experiment by choosing that one of the remaining populations associated with the largest cumulative sum.

It is seen that Paulson rule can eliminate "noncontending" populations as the experiment proceeds and if it does, these populations are eliminated permanently. Therefore it is said to have a screening feature with permanent elimination. Further, there is a preassigned finite upper bound $(W_\lambda + 1)$ on the total number of stages to terminate experimentation. Therefore it is referred to as a closed or a truncated rule.

In addition to the rules described above, a two-stage rule proposed by Alam [1970] is central to our thesis. We shall describe it in detail in Chapter 2.

§1.1.3 The subset approach and Gupta rule:

In the following subset approach due to Gupta [1956, 1965], it is assumed that the experimenter specifies one constant {P*}, $1/k < P^* < 1$, prior to the start of experimentation. The goal of the experimenter is to choose a non-empty subset of the populations which contains the best population. He restricts attention to those rules which guarantee the probability requirement

$$P_{\underline{\mu}}(CS|R) \geq P^* \qquad \forall \underline{\mu} \in \Omega \tag{1.1.10}$$

where the event CS (correct selection) refers to the selection of any subset of the populations which contains the best population.

In may be noted that (1.1.10) can be guaranteed without taking any observations by choosing the set of all k populations as the selected subset. However one would not use such a rule since clearly the experimenter would like the selected subset size (which in general is a r.v.) to be small in a certain sense. Another important point to be noted is that (1.1.10) must be guaranteed over the entire parameter space Ω (in constrast to the situation in (1.1.1)),and there is no indifference zone.

Gupta rule: Observe the sample means $\overline{X}_i (1 \leq i \leq k)$ based on n mutually independent observations from each population. Include the population Π_i in the subset iff

$$\overline{X}_i \geq \max_{1 \leq j \leq k} \overline{X}_j - h \tag{1.1.11}$$

where $h > 0$ and n are decided on prior to the experiment in such a way that (1.1.10) is guaranteed. Gupta showed that this can be accomplished by letting

$$\frac{hn^{1/2}}{\sigma} = d_0, \tag{1.1.12}$$

where $d_0 = d_0(k,P^*)$ is given by (1.1.2).

Note that an infinite number of choices are possible for the values of h and n. Therefore further criteria might be introduced to choose a "best" pair (h,n).

Before we proceed to the next section we point out that a central problem concerned with the construction of rules using either the indifference-zone approach to guarantee (1.1.1) or the subset approach to guarantee (1.1.10) is that of finding the infimum of the probability of a correct selection (PCS) over the appropriate region of the parameter space Ω, the region being the preference zone $\Omega(\delta^*)$ for the indifference-zone approach and the entire parameter space Ω for the subset approach. If these infima are equated to P^*, the constant d_0 of (1.1.2) can be determined and the corresponding probability requirements (1.1.1) and (1.1.10) will then be guaranteed. A parameter configuration for which such an infimum occurs is known as a <u>least favorable configuration</u> (LFC) of the parameters for the rule under study.

§1.2 <u>Motivation for developing two-stage screening rules</u>:

In single-stage rules there is no opportunity to learn from data as they are observed, in order to modify the sampling procedure. For example, single-stage rules used for the indifference-zone approach require, in general, a large amount of sampling relative to rules which "react" to the data. This becomes an important consideration when k, the number of populations is large and/or the experimenter's requirements are very stringent (small δ^* and/or large P^*.)

The BKS sequential rule does react to favorable parameter configurations. Thus if one population mean is much larger than all of the others, then the BKS rule terminates in very few stages. However, the BKS rule possesses the disadvantages that no screening is performed and that it is open ended. Paulson's rule possesses the desirable features of reacting to favorable parameter configurations and of screening and truncation. Monte Carlo (MC) studies by Ramberg [1966] have demonstrated that Paulson's rule is highly efficient (in terms of reducing the expected total sample size) relative to the single-stage rule R_0 and the sequential rule BKS which guarantee the same probability requirement (1.1.1).

However, the administration of any sequential experiment is more difficult than that of a single-stage experiment. Sequential rules become impractical when the time period that must elapse between the successive stages of the experiment is very large. For example, in agricultural field trials one must wait until the next planting season to conduct the next stage of the experiment.

The Gupta rule for subset selection also has a practical drawback. It does not explicitely take into account a possible ultimate goal of the experimenter namely that of selecting a <u>single</u> population which is best. Santner's restricted subset size approach makes an important contribution in this direction by placing a preassigned upper bound on the size of the subset.

Therefore what appears to be needed is a two- or a three-stage rule which has the property that it screens out the non-contending populations in the preliminary stages. Cohen [1959] and Alam [1970] have proposed such two-stage rules. We shall review their work in greater detail in the next chapter. Our objective in the present thesis is to develop two-stage (in one instance three-stage) eliminating type rules for certain selection problems. (The normal means problem described in §1.1.1 is one of the problems considered in the present thesis.)

In the following section we give an overview of the main results of the thesis.

§1.3 An outline and summary of the thesis:

We adopt the indifference-zone approach throughout the present thesis. Chapters 2, 3 and 4 are devoted to the consideration of the normal means problem described in §1.1.1 and §1.1.2.

In Chapter 2 we consider the two-stage permanent elimination type of rule (R_1) proposed by Alam [1970]. Alam proposes implementing the rule R_1 by choosing constants to minimize the maximum of the expected total sample size over a restricted region of the parameter space $\Omega(\delta^*)$, i.e. he proposes using a restricted minimax design criterion (R-minimax criterion). We indicate certain disadvantages associated with the use of such a criterion and propose instead that the rule R_1 be implemented by choosing constants to minimize the maximum of the expected total sample size over the entire parameter space Ω, i.e. we propose using an unrestricted minimax design criterion (U-minimax criterion). We also extend Alam's work in many ways for the case of $k \geq 2$ populations, the major portion of Alam's work being concerned with the case of 2 populations. We derive general exact expression for the PCS and the expected total sample size for rule R_1 for $k \geq 2$. We obtain a partial result concerning the infimum of the PCS of rule R_1 namely that the PCS is non-decreasing in the largest mean. Though we could not prove Alam's general ($k \geq 2$) conjecture regarding the LFC for rule R_1, we were able to prove his conjecture regarding the supremum of the expected total sample size for rule R_1. We next state certain constrained optimization problems associated with finding the constants necessary to implement rule R_1 using the R-minimax criterion and the U-minimax criterion. We close the second chapter by providing extensive tables of the constants required to implement rule R_1 for the case $k = 2$ for both criteria.

In Chapter 3 we propose a useful lower bound for the PCS of rule R_1. The lower bound is helpful because it has an easily computable infimum over $\Omega(\delta^*)$. Using this lower bound we have constructed tables of constants for implementing our two-stage rule with the U-minimax criterion for $k = 3(1)5$, 10 and 25. These constants form the basis of a conservative rule (which overprotects the experimenter) relative to the probability requirement (1.1.1), the overprotection being purchased at the expense of an increase in the expected total sample size. Employing MC techniques we study the relative performance of this lower bound on the PCS of rule R_1 and the exact PCS of rule R_1. We next consider the asymptotic ($P^* \to 1$) performance of these rules and derive the values of the asymptotic relative efficiencies of the single-stage rule R_0 w.r.t the two-stage rule R_1 using U-minimax and R-minimax criteria. We then propose a modification of our basic permanent elimination type two-stage rule R_1; the modification consists in allowing the "eliminated" populations in the first stage to be eligible for selection after the second stage. We refer to this as the "come-back" modification of rule R_1 and study its properties using MC methods. We propose another come-back modification of rule R_1 and derive its properties for $k = 2$.

In Chapter 4 we consider the question of optimality (in the sense of minimizing the maximum of the total expected sample size over Ω) of our rule R_1; the optimality is with respect to a restricted class of two-stage permanent elimination type rules. We show that the desired U-minimax rule is a Bayes rule with positive prior probabilities on the LFC and the parameter point where the maximum of the total expected sample size occurs. For $k = 2$ we derive the exact structure of the Bayes rule. We show that our natural selection U-minimax rule R_1 is Bayes with the above prior and hence U-minimax in the class of two-stage rules under consideration. For $k > 2$ our tentative results indicate that the rule R_1 is not U-minimax in the same class of two-stage rules.

In Chapter 5 we consider the normal means problem when the common variance is unknown. Our objective here is to provide a screening rule without any attempt at optimization. We give a three-stage rule and a two-stage rule which guarantee the experimenter's probability requirement (1.1.1). Expressions for the expected total sample sizes for these rules are derived. Using MC sampling techniques we compare the performance characteristics of both of these rules with the two-stage rule proposed by Bechhofer, Dunnett and Sobel [1954] which is used when the common variance is unknown. In general the three-stage rule is indicated as being superior whereas our two-stage rule is inferior.

We conclude in Chapter 6 with some suggestions for future research.

CHAPTER 2

A TWO-STAGE PERMANENT ELIMINATION RULE FOR THE NORMAL MEANS PROBLEM (COMMON KNOWN VARIANCE).

§2.0 Introduction:

In the present and the next chapter we consider the problem of selecting the population having the largest mean from k normal populations with a common known variance. This problem has been described in §1.1. We propose and study in detail certain two-stage screening type rules. The motivation for developing such rules was provided in §1.2.

We consider two types of rules: (1) the first type of rule permanently eliminates the "non-contending" populations after the first stage and permits the final selection to be made from among only those populations which enter the second stage. (2) the second type of rule *temporarily* eliminates the non-contending populations after the first stage and permits the final selection to be made from *all* of the populations i.e. the ones that were "eliminated" after the first stage and the ones that entered the second stage. We now briefly sketch the contents of the present chapter.

In §2.1 we describe the problem and propose our two-stage rule R_1. The rule is characterized by certain constants used to implement it. We recommend the U-minimax criterion (explained in §1.3) for the choice of these constants and provide a rationale for doing so. In §2.2 we derive the expression for the PCS of rule R_1 and study its infimum. We show that it is non-decreasing in the largest population mean but we have not been able to obtain the LFC of rule R_1. In §2.3 we derive an expression for the expected total sample size and show that its supremum takes place when all the population means are equal. In §2.4 we state the optimization problems which one must solve to obtain the constants necessary to implement rule R_1 using either the U-minimax criterion or the R-minimax criterion (explained in §1.3). In §2.5 we define a measure of relative efficiency of two competing rules and give the associated expressions for the relative efficiency of the single-stage rule R_0 with respect to our rule R_1. In §2.6 we give tables necessary to implement rule R_1 for k = 2 using the U-minimax criterion. For the purposes of comparison we also give corresponding tables for the R-minimax criterion. Finally we give some numerical comparisons of the discrete optimal and the continuous optimal solutions to the optimization problem associated with the U-minimax criterion.

§2.1 Preliminaries:

§2.1.1 Statement of the problem and the proposed rule R_1:

Consider the set up described in §1.1. We seek a two-stage rule which guarantees the probability requirement (1.1.1). The rule is to be one which employs a common non-random number of observations from each population in the first stage, takes another common non-random number of observations from each non-eliminated population in the second stage and makes the final selection only from the populations which entered the second stage. Clearly if only a single population enters the second stage then any reasonable rule would terminate sampling and assert that population as best.

The class of rules described in the previous section has the feature of permanent elimination. Using a permanent elimination type rule it could happen in an experiment that the largest first stage sample mean of those populations which are eliminated turns out to be bigger than the largest cumulative sample mean of those populations which are not eliminated. To avoid such undesirable contingencies we shall, in the next chapter, propose and study certain "come-back" type rules.

Another somewhat restrictive feature of the above class of rules is the non-random nature of the second stage sample size. Leaving aside the serious analytical difficulties associated with making the second stage sample size random, there is a practical advantage to be gained from making it non-random. This restriction allows the experimenter to know in advance an upper limit on the total number of observations to be taken and thus budget for them adequately. Clearly the same advantage can be gained by having a random but bounded second stage sample size.

We now propose our rule R_1.

Rule R_1: Let non-negative integers n_1, n_2 and a non-negative real constant h be specified prior to the start of experimentation. The following are the steps in rule R_1 utilizing (n_1, n_2, h) which are chosen to guarantee (1.1.1).

1. In the first stage, from each Π_i take n_1 independent observations $X_{ij}^{(1)}$ $(1 \le j \le n_1)$ and compute $\bar{X}_i^{(1)} = \sum_{j=1}^{n_1} X_{ij}^{(1)}/n_1$ $(1 \le i \le k)$. Let $\bar{X}_{[k]}^{(1)} = \max_{1 \le i \le k} \bar{X}_i^{(1)}$.

2. Choose a subset I of $\{1,2,\ldots,k\}$ where

$$I = \{ i \mid \bar{X}_i^{(1)} \geq \bar{X}_{[k]}^{(1)} - h \}. \tag{2.1.1}$$

3a. If I consists of a single population, stop sampling after the first stage and assert that, that single population is best.

3b. If not, go onto the second stage and take n_2 additional independent observations $X_{ij}^{(2)}$ $(1 \leq j \leq n_2)$ from each Π_i for $i\varepsilon I$. Compute the overall means $\bar{X}_i = (\sum_{j=1}^{n_1} X_{ij}^{(1)} + \sum_{j=1}^{n_2} X_{ij}^{(2)})/(n_1 + n_2)$ for $i\varepsilon I$ and assert that the population associated with $\max_{i\varepsilon I} \bar{X}_i$ is best. (We ignore the possibility of ties since the probability of an event involving tied means is zero.)

This rule had been proposed previously by Cohen [1959] and Alam [1970]. Due to the analytical difficulties involved, most of their work was limited to the special case $k = 2$. It will be our objective to extend this work to the general case $k \geq 2$, and choose the constants necessary to implement the rule using the U-minimax criterion.

Note that the limiting cases of $h = 0$ and $h = \infty$ yield Bechhofer's single-stage rule R_0 with n_1 and $n_1 + n_2$ as the single-stage sample sizes, respectively. Also note that the retention rule in the second step is of the type proposed by Gupta [1956, 65] in his subset selection rule (1.1.11).

§2.1.2 The design criterion:

There are an infinite number of combinations of (n_1, n_2, h) which will guarantee (1.1.1). To help us choose among them, we propose to study the expected total sample size at a certain parameter vector $\underline{\mu}$ of interest. We shall use that combination (n_1, n_2, h) which guarantees (1.1.1) and yields the minimum expected total sample size at the $\underline{\mu}$ of interest. The choice of the $\underline{\mu}$ at which the experimenter would like to minimize the expected total sample size subject to (1.1.1) leads to a choice of the optimal (n_1, n_2, h).

The total sample size N required by rule R_1 with the associated constants (n_1, n_2, h) is given by

$$N = kn_1 + T'n_2 \tag{2.1.2}$$

where T' is defined by (2.3.2). Note that T' (and hence N) is a r.v. Let $E_{\underline{\mu}}(N|R_1)$ denote the expected total sample size for rule R_1 when the underlying parameter vector is $\underline{\mu}$. The design criterion proposed by Alam [1970] can be stated as follows:

1. R-minimax criterion: For specified $\{\delta^*, P^*\}$ choose the three constants (n_1, n_2, h) necessary to implement rule R_1 so as to

$$(2.1.3)\quad \begin{cases} \text{minimize} \sup\limits_{\Omega(\delta^*)} E_{\underline{\mu}}(N|R_1), \\ \text{subject to } P_{\underline{\mu}}(CS|R_1) \geq P^* \quad \forall\, \underline{\mu}\in\Omega(\delta^*), \\ \qquad n_1,\ n_2 \text{ non-negative integers}, \\ \text{and } h \geq 0. \end{cases}$$

We denote by $(\tilde{n}_1, \tilde{n}_2, \tilde{h})$ a solution to the above optimization problem and the corresponding rule R_1 by $\tilde{R}_1$.

Alam's approach ignores what can happen to the expected total sample size if, unknown to the experimenter, $\underline{\mu}$ lies in $\Omega_0(\delta^*)$. Thus suppose n_0 is the sample size per population required by rule R_0 which is known to be "most-economical" in the sense of Hall [1958] in the class of all single-stage rules. Then Alam's formulation insures that $kn_0 \geq E_{\underline{\mu}}(N|\tilde{R}_1)$ for all $\underline{\mu}\varepsilon\Omega(\delta^*)$. But for $\underline{\mu}\varepsilon\Omega_0(\delta^*)$, it is possible to have $kn_0 < E_{\underline{\mu}}(N|\tilde{R}_1)$. Indeed, intuitively one would expect that the ratio $E_{EMC}(N|\tilde{R}_1)/kn_0$ where EMC represents the equal means configuration $(\mu_{[1]} = \cdots = \mu_{[k]})$, may grow large (> 1) as $P^* \to 1$ for fixed δ^*. Bechhofer [1960] has shown a similar undesirable feature of the Wald's sequential probability ratio test (WSPRT). It is to guard against this possibility that we propose a design criterion which will insure that the rule based on that criterion will have an expected total sample size uniformly (in $\underline{\mu}$) smaller than the fixed sample size required by rule R_0 which guarantees the same probability requirement (1.1.1) for any $\{\delta^*, P^*\}$. We now propose such a design criterion.

2. U-minimax criterion: For specified $\{\delta^*, P^*\}$ choose the three constants (n_1, n_2, h) necessary to implement rule R_1 so as to

$$(2.1.4)\quad \begin{cases} \text{minimize } \sup_{\Omega} E_{\mu}(N|R_1), \\ \text{subject to } P_{\underline{\mu}}(CS|R_1) \geq P^* \quad \forall\, \underline{\mu}\varepsilon\Omega(\delta^*), \\ \qquad n_1, n_2 \text{ non-negative integers,} \\ \text{and } h \geq 0. \end{cases}$$

We denote by $(\hat{n}_1, \hat{n}_2, \hat{h})$ a solution to the above optimization problem and the corresponding rule R_1 by $\hat{R}_1$.

If n_0 is the sample size per population required to guarantee the same probability requirement using rule R_0 then rule $\hat{R}_1$ has the desirable property that $E_{\underline{\mu}}(N|\hat{R}_1) \leq kn_0$ for all $\underline{\mu}\varepsilon\Omega$ for any specified $\{\delta^*, P^*\}$. Now we proceed with the derivation of various properties of rule R_1.

§2.2 PCS of rule R_1, and its behavior:

§2.2.1 A general expression for $P_{\underline{\mu}}(CS|R_1)$:

We first derive a general expression for $P_{\underline{\mu}}(CS|R_1)$. Our result is summarized in the following theorem.

Theorem 2.2.1: For any $\underline{\mu}\varepsilon\Omega$ we have

(2.2.1)

$$P_{\underline{\mu}}(CS|R_1) = \sum_{s\varepsilon\mathscr{S}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \prod_{i\varepsilon s} \left\{ \int_{x+(\delta_{ki}-h)n_1^{1/2}/\sigma}^{x+\delta_{ki}n_1^{1/2}/\sigma} \Phi\left[y+(x-z)(p/q)^{1/2} + \frac{\delta_{ki}}{\sigma}(n/q)^{1/2}\right]\right.$$

$$\left. d\Phi(z) \right\} \times \prod_{i\varepsilon s} \Phi\left[x+(\delta_{ki}-h)n_1^{1/2}/\sigma\right] d\Phi(y)\,d\Phi(x) + \sum_{j=1}^{k-1}\sum_{s\varepsilon\mathscr{S}_j} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}$$

$$\left[\int_{x-(\delta_{kj}+h)n_1^{1/2}/\sigma}^{x-\delta_{kj}n_1^{1/2}/\sigma} \int_{y+(x-u)(p/q)^{1/2} - \frac{\delta_{kj}}{\sigma}(n/q)^{1/2}}^{\infty} \left(\prod_{i\varepsilon s} \left\{ \int_{x-(\delta_{ij}+h)n_1^{1/2}/\sigma}^{x-\delta_{ij}n_1^{1/2}/\sigma} \right.\right.\right.$$

$$\Phi\left[v+(u-z)(p/q)^{1/2}+\frac{\delta_{ki}}{\sigma}(n/q)^{1/2}\right]d\Phi(z)\Bigg\} \times \prod_{i\varepsilon s} \Phi\left[x+(\delta_{ji}-h)n_1^{1/2}/\sigma\right]\Bigg)$$

$$d\Phi(v)d\Phi(u)\Bigg]\, d\Phi(y)d\Phi(x),$$

where $\mathcal{S}$ = collection of all subsets from $\{1,2,\ldots,k-1\}$,

$\mathcal{S}_j$ = collection of all subsets from $\{1,\ldots,j-1,j+1,\ldots,k-1\}$,

$n = n_1 + n_2$, $p = n_1/n$, $q = n_2/n = 1 - p$,

and $\delta_{ij} = \mu_{[i]} - \mu_{[j]}$ $(1 \leq i,j \leq k)$.

Proof: We denote by $\bar{X}^{(1)}_{(i)}$ and $\bar{X}_{(i)}$, respectively, the first stage and the cumulative mean from two stages from population $\Pi_{(i)}$ $(1 \leq i \leq k)$. Then

$$(2.2.2)\quad P_{\underline{\mu}}(CS|R_1) = \sum_{s\varepsilon\mathcal{S}} P_{\underline{\mu}}\{\bar{X}^{(1)}_{(k)} = \bar{X}^{(1)}_{[k]}, \bar{X}^{(1)}_{(i)} \geq \bar{X}^{(1)}_{[k]}-h \ \forall i\varepsilon s,\ \bar{X}^{(1)}_{(i)}$$

$$< \bar{X}^{(1)}_{[k]} - h \ \ \forall i\notin s,\ \bar{X}_{(k)} > \bar{X}_{(i)} \ \ \forall i\varepsilon s\} + \sum_{j=1}^{k-1}\sum_{s\varepsilon\mathcal{S}_j} P_{\underline{\mu}}\{\bar{X}^{(1)}_{(j)} = \bar{X}^{(1)}_{[k]},$$

$$\bar{X}^{(1)}_{(k)} \geq \bar{X}^{(1)}_{[k]} - h,\ \bar{X}^{(1)}_{(i)} \geq \bar{X}^{(1)}_{[k]} - h \ \ \forall i\varepsilon s,\ \bar{X}^{(1)}_{(i)} < \bar{X}^{(1)}_{[k]} - h \ \ \forall i\notin s;\ \bar{X}_{(k)} >$$

$$\bar{X}_{(j)},\ \bar{X}_{(k)} > \bar{X}_{(i)} \ \ \forall i\varepsilon s\}$$

$$= \sum_{s\varepsilon\mathscr{S}} A_s + \sum_{j=1}^{k-1} \sum_{s\varepsilon\mathscr{S}_j} B_{s,j} \quad .$$

We shall now evaluate the probabilities A_s and $B_{s,j}$. Throughout Chapters 2 and 3 we denote the r.v. $[(\bar{X}^{(1)}_{(i)} - \mu_{[i]})\, n_1^{1/2}/\sigma,$ $(\bar{X}_{(i)} - \mu_{[i]})n^{1/2}/\sigma]$ by (X_i, Y_i). Note that for $1 \leq i \leq k$, the r.v. (X_i, Y_i) has a standard bivariate normal distribution with correlation coefficient $= p^{1/2} = (n_1/n)^{1/2}$. In what follows we shall have occasion to use the relation

$$\Phi_2(a,b\,|\,\rho) = \int_{-\infty}^{a} \Phi\left[\frac{b-\rho z}{(1-\rho^2)^{1/2}}\right] d\Phi(z), \tag{2.2.3}$$

where $\Phi_2(\cdot,\cdot\,|\,\rho)$ represents the standard bivariate normal cdf with correlation coefficient $= \rho$.

We first consider

$$A_s = P_{\underline{\mu}}\{X_k + \delta_{ki}n_1^{1/2}/\sigma > X_i \geq X_k + (\delta_{ki} - h)n_1^{1/2}/\sigma \quad \forall i\varepsilon s, \tag{2.2.4a}$$

$$X_i < X_k + (\delta_{ki} - h)n_1^{1/2}/\sigma \quad \forall i\not\varepsilon s,\ Y_k + \delta_{ki}n^{1/2}/\sigma > Y_i \quad \forall i\varepsilon s\}$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \prod_{i\varepsilon s} \left\{ \Phi_2\left[x+\delta_{ki}n_1^{1/2}/\sigma,\ u+\delta_{ki}n^{1/2}/\sigma\,|\,p^{1/2}\right] - \Phi_2\left[x+(\delta_{ki} - h)\, n_1^{1/2}/\sigma,\right.\right.$$

$$\left.\left. u + \delta_{ki}n^{1/2}/\sigma\,|\,p^{1/2}\right] \right\} \times \prod_{i\not\varepsilon s} \Phi\left[x+(\delta_{ki} - h)n_1^{1/2}/\sigma\right] d\Phi_2(x,u\,|\,p^{1/2}).$$

This expression was obtained by conditioning on $(X_k, Y_k) = (x,u)$ and then integrating w.r.t. the joint density function of (X_k,Y_k). Applying (2.2.3) and (2.2.4a) we obtain

(2.2.4b)

$$A_s = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \prod_{i\varepsilon s} \left\{ \int_{x+(\delta_{ki}-h)n_1^{1/2}/\sigma}^{x+\delta_{ki}n_1^{1/2}/\sigma} \Phi\left[u/q^{1/2}-z(p/q)^{1/2} + \frac{\delta_{ki}}{\sigma}(n/q)^{1/2}\right] d\Phi(z) \right\} \times$$

$$\prod_{i\notin s} \Phi\left[x + (\delta_{ki} - h)n_1^{1/2}/\sigma\right] d\Phi_2(x,u|p^{1/2}).$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \prod_{i\varepsilon s} \left\{ \int_{x+(\delta_{ki}-h)n_1^{1/2}/\sigma}^{x+\delta_{ki}n_1^{1/2}/\sigma} \Phi\left[y + (x-z)(p/q)^{1/2} + \frac{\delta_{ki}}{\sigma}(n/q)^{1/2}\right] d\Phi(z) \right\} \times$$

$$\prod_{i\notin s} \Phi\left[x + (\delta_{ki} - h)n_1^{1/2}/\sigma\right] d\Phi(y)d\Phi(x).$$

The last equality was obtained by making the orthogonal transformation $y = \dfrac{u-p^{1/2}x}{(1-p)^{1/2}}$, $x = x$.

We next consider

(2.2.5a) $$B_{sj} = P_{\underline{\mu}}\{X_j - \delta_{kj}n_1^{1/2}/\sigma > X_k \geq X_j - (\delta_{kj} + h)n_1^{1/2}/\sigma,$$

$$X_j - \delta_{ij}n_1^{1/2}/\sigma > X_i \geq X_j - (\delta_{ij} + h)n_1^{1/2}/\sigma \quad \forall i\varepsilon s;$$

$$X_i < X_j - (\delta_{ij} + h)n_1^{1/2}/\sigma \quad \forall i \notin s;\ Y_k + \delta_{kj}n_1^{1/2}/\sigma > Y_j,\ Y_k + \delta_{ki}n^{1/2}/\sigma > Y_i$$

$$\forall i\varepsilon s\}. = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[\int_{x-(\delta_{kj}+h)n_1^{1/2}/\sigma}^{x-\delta_{kj}n_1^{1/2}/\sigma} \int_{t-\delta_{kj}n^{1/2}/\sigma}^{\infty}\left(\prod_{i\varepsilon s}\left\{\Phi_2\left[x - \delta_{ij}n_1^{1/2}/\sigma,\right.\right.\right.\right.$$

$$\left.\left.w + \delta_{ki}n^{1/2}/\sigma|p^{1/2}\right] - \Phi_2\left[x - (\delta_{ij} + h)n_1^{1/2}/\sigma,\ w + \delta_{ki}n^{1/2}/\sigma|p^{1/2}\right]\right\} \times$$

$$\left.\left.\prod_{i\notin s}\Phi\left[x - (\delta_{ij} + h)n_1^{1/2}/\sigma\right]\right)d\Phi_2(u,w|p^{1/2})\right]\ d\Phi_2(x,t|p^{1/2}).$$

This latter expression was obtained by conditioning on $(X_j,Y_j) = (x,t)$ and $(X_k,Y_k) = (u,w)$ and then integrating w.r.t. the joint density functions of (X_j,Y_j) and (X_k,Y_k). In this expression the integration is first w.r.t. (u,w) and then w.r.t. (x,t). Applying (2.2.3) to (2.2.5a) we obtain

$$\text{(2.2.5b)}\quad B_{s,j} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[\int_{x-(\delta_{kj}+h)n_1^{1/2}/\sigma}^{x-\delta_{kj}n_1^{1/2}/\sigma} \int_{t-\delta_{kj}n^{1/2}/\sigma}^{\infty}\left(\prod_{i\varepsilon s}\right.\right.$$

$$\left\{\int_{x-(\delta_{ij}+h)n_1^{1/2}/\sigma}^{x-\delta_{ij}n_1^{1/2}/\sigma}\Phi\left[w/q^{1/2} - z\ (p/q)^{1/2} + \frac{\delta_{ki}}{\sigma}(n/q)^{1/2}\right]d\Phi(z)\right\} \times$$

$$\left.\left.\prod_{i\notin s}\Phi\left[x - (\delta_{ij}+h)n_1^{1/2}/\sigma\right]\right)\ d\Phi_2(u,w/p^{1/2})\right]\ d\Phi_2(x,t|p^{1/2})$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[\int_{x-(\delta_{kj}+h)n_1^{1/2}/\sigma}^{x-\delta_{kj}n_1^{1/2}/\sigma} \int_{y+(x-u)(p/q)^{1/2}-\delta_{kj}n_1^{1/2}/\sigma}^{\infty}\left(\prod_{i\varepsilon s}\right.\right.$$

$$\left\{\int_{x-(\delta_{ij}+h)n_1^{1/2}/\sigma}^{x-\delta_{ij}n_1^{1/2}/\sigma} \Phi\left[v + (u-z)(p/q)^{1/2} + \frac{\delta_{ki}}{\sigma}(n/q)^{1/2}\right] d\Phi(z)\right\} \times$$

$$\left.\left.\prod_{i\neq s} \Phi\left[x - (\delta_{ij} + h)n_1^{1/2}/\sigma\right]\right) d\Phi(v)d\Phi(u)\right] d\Phi(y)d\Phi(x).$$

The last equality was obtained by making the orthogonal transformation $y = \frac{t-p^{1/2}x}{(1-p)^{1/2}}$, $x = x$ and $v = \frac{w-p^{1/2}u}{(1-p)^{1/2}}$, $u = u$. Combining (2.2.2), (2.2.4b) and (2.2.5b) we obtain the desired result (2.2.1).

Corollary 2.2.1: Let $\underline{\mu}(\delta)$ denote any $\underline{\mu}$ such that $\mu_{[1]} = \mu_{[2]} = \cdots = \mu_{[k-1]} = \mu_{[k]} - \delta$ where $\delta \geq 0$ ($\underline{\mu}(\delta)$ is known as a slippage configuration). Then we have

$$(2.2.6)\quad P_{\underline{\mu}(\delta)}(CS|R_1) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left\{\Phi\left[x + (\delta-h)n_1^{1/2}/\sigma\right] + \int_{x+(\delta-h)n_1^{1/2}/\sigma}^{x+\delta n_1^{1/2}/\sigma}\right.$$

$$\left.\Phi\left[y + (x-z)(p/q)^{1/2} + \frac{\delta}{\sigma}(n/q)^{1/2}\right] d\Phi(z)\right\}^{k-1} d\Phi(y)d\Phi(x) + (k-1) \times$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[\int_{x-(\delta+h)n_1^{1/2}/\sigma}^{x-\delta n_1^{1/2}/\sigma}\ \int_{y+(x-u)(p/q)^{1/2} - \frac{\delta}{\sigma}(n/q)^{1/2}}^{\infty}\left\{\Phi(x-hn_1^{1/2}/\sigma) +\right.\right.$$

$$\int_{x-hn_1^{1/2}/\sigma}^{x} \Phi\left[v + (u-z)(p/q)^{1/2} + \frac{\delta}{\sigma}(n/q)^{1/2}\right] \times$$

$$\left.\left.d\Phi(z)\right\}^{k-2} d\Phi(v)d\Phi(u)\right] d\Phi(y)d\Phi(x).$$

Proof: The proof is straightforward and is omitted.

Remark 2.2.1: Alam [1970] did not give any expression corresponding to our (2.2.1). For k = 2, (2.2.1) simplifies a great deal and we obtain for any $\underline{\mu}\varepsilon\Omega$

(2.2.7)
$$P_{\underline{\mu}}(CS|R_1) = \Phi\left[\left(\frac{\delta-h}{\sigma}\right)\left(\frac{n_1}{2}\right)^{1/2}\right] + \int_{\left(\frac{\delta-h}{\sigma}\right)\left(\frac{n_1}{2}\right)^{1/2}}^{\left(\frac{\delta+h}{\sigma}\right)\left(\frac{n_1}{2}\right)^{1/2}} \Phi\left[-u\left(\frac{p}{q}\right)^{1/2} + \frac{\delta}{\sigma}\left(\frac{n}{2q}\right)^{1/2}\right] d\Phi(u).$$

This result was given by Alam as his equation (3.5).

Alam also gave an expression for $P_{\underline{\mu}(\delta)}(CS|R_1)$ for $k > 2$; see his equation (3.24). However, we were not able to verify his general result and suspect that it is in error.

Remark 2.2.2: It is easy to check that letting $h \to 0$ in (2.2.1) yields

(2.2.8)
$$\lim_{h\to 0} P_{\underline{\mu}}(CS|R_1) = \int_{-\infty}^{\infty} \prod_{i=1}^{k-1} \Phi(x + \delta_{ki}n_1^{1/2}/\sigma)d\Phi(x).$$

It is straightforward but tedious to check that letting $h \to \infty$ in (2.2.1) yields

(2.2.9)
$$\lim_{h\to \infty} P_{\underline{\mu}}(CS|R_1) = \int_{-\infty}^{\infty} \prod_{i=1}^{k-1} \Phi(x + \delta_{ki}n^{1/2}/\sigma)d\Phi(x).$$

Note that (2.2.8) and (2.2.9) are the expressions for $P_{\underline{\mu}}(CS|R_0)$ i.e. the PCS associated with rule R_0, with respectively n_1 and n as the common sample sizes per population.

§2.2.2 Determination of the infimum of $P_{\underline{\mu}}(CS|R_1)$:

It is now our objective to determine the set of parameter points in the preference zone $\Omega(\delta^*)$ where the infimum of $P_{\underline{\mu}}(CS|R_1)$ occurs. Our reasons for doing so will be given later.

Definition 2.2.1: Any $\underline{\mu}_0 \varepsilon \Omega(\delta^*)$ such that $P_{\underline{\mu}_0}(CS|R) = \inf_{\Omega(\delta^*)} P_{\underline{\mu}}(CS|R)$ is referred to as a least favorable configuration (LFC) of the parameters for rule R.

If we know the LFC of the μ_i, then we can formally set

$$(2.2.10) \qquad P_{LFC}(CS|R_1) = P^*$$

Any values of (n_1,n_2,h) which make the l.h.s. of (2.2.10) equal to P* will guarantee the probability requirement (1.1.1).

As a step toward finding the LFC of the μ_i for rule R_1 we shall now study the behavior of $P_{\underline{\mu}}(CS|R_1)$ and in particular its monotonicity w.r.t. $\mu_{[i]}(1 \leq i \leq k)$.

Lemma 2.2.1: For fixed $\mu_{[i]}(1 \leq i \leq k-1)$ and for fixed n_1,n_2,h, σ and k, $P_{\underline{\mu}}(CS|R_1)$ is non-decreasing in $\mu_{[k]}$.

Proof:

$$\text{(2.2.11a)} \quad P_{\underline{\mu}}(CS|R_1) = P_{\underline{\mu}}\left\{ \bigcup_{s\varepsilon\mathscr{S}} \left[B_s \cap (\bar{X}_{(k)} > \bar{X}_{(i)} \quad \forall\, i\varepsilon s)\right] \right\} .$$

Here the event $B_s = \left[\bar{X}^{(1)}_{(k)} \geq \bar{X}^{(i)}_{[k]} - h,\ \bar{X}^{(1)}_{(i)} \geq \bar{X}^{(1)}_{[k]} - h \quad \forall\, i\varepsilon s,\right.$ $\left.\bar{X}^{(1)}_{(i)} < \bar{X}^{(1)}_{[k]} - h \quad \forall\, i\notin s\right]$ and $\mathscr{S}$ = the collection of <u>all</u> subsets from $\{1,2,\ldots,k-1\}$. Thus (2.2.11a) becomes

$$\text{(2.2.11b)} \quad P_{\underline{\mu}}(CS|R_1) = P\left\{ \bigcup_{s\varepsilon\mathscr{S}} \bigcup_{\substack{j\varepsilon s\\ j=k}} \left[(X_j + (\mu_{[j]} - \mu_{[i]})n_1^{1/2}/\sigma > \right.\right.$$

$$X_i \geq X_j + (\mu_{[j]} - \mu_{[i]} - h)n_1^{1/2}/\sigma \quad \forall\, i\varepsilon s,\ i = k,\ i \neq j;$$

$$X_i < X_j + (\mu_{[j]} - \mu_{[i]} - h)n_1^{1/2}/\sigma \quad \forall\, i\notin s) \cap$$

$$\left.\left.(Y_k + (\mu_{[k]} - \mu_{[i]})n^{1/2}/\sigma > Y_i \quad \forall\, i\varepsilon s)\right]\right\} \quad = P\{A(\underline{\mu})\}.$$

The event $A(\underline{\mu})$ is in the sigma algebra generated by the r.v.(X_i,Y_i) for $1 \leq i \leq k$.

Now consider a vector $\underline{\mu}' = (\mu'_{[1]},\mu'_{[2]},\ldots,\mu'_{[k]})$ where $\mu'_{[i]} = \mu_{[i]}$ for $1 \leq i \leq k - 1$ and $\mu'_{[k]} > \mu_{[k]}$. Then

$$\text{(2.2.12)} \qquad P_{\underline{\mu}'}(CS|R_1) = P\{A(\underline{\mu}')\}.$$

Now we shall show that $A(\underline{\mu}) \subseteq A(\underline{\mu}')$. For the sake of clarity we shall indicate in this section, the dependence of a r.v. X on the sample point ω by $X(\omega)$.

Pick $\omega\varepsilon A(\underline{\mu})$. Then the event

(2.2.13) $$[X_k(\omega) + (\mu_{[k]} - \mu_{[i]} + h)n_1^{1/2}/\sigma > X_i(\omega) \quad \forall i \neq k]$$

$$\Longrightarrow [X_k(\omega) + (\mu'_{[k]} - \mu_{[i]} + h)n_1^{1/2}/\sigma \geq X_i(\omega) \quad \forall i \neq k]$$

$$\Longrightarrow [X_j(\omega) + (\mu'_{[j]} - \mu'_{[i]})n_1^{1/2}/\sigma > X_i(\omega) \geq X_j(\omega) + (\mu'_{[j]} - \mu'_{[i]} - h) \times$$

$$n_1^{1/2}/\sigma \quad \forall i\varepsilon s', \; i = k, \; i \neq j; \; X_i(\omega) < X_j(\omega) + (\mu'_{[j]} - \mu'_{[i]} - h)n_1^{1/2}/\sigma$$

$$\forall i\varepsilon s']$$

for some set $s'\varepsilon\mathscr{S}$, $j = k$ or $j\varepsilon s'$.

To study the set s' consider the following two cases.

<u>Case 1</u>: $\overline{X}^{(1)}_{(k)}(\omega) = \overline{X}^{(1)}_{[k]}(\omega)$ and the set $s\varepsilon\mathscr{S}$ is selected. Thus the following event occurs:

$$[X_k(\omega) + (\mu_{[k]} - \mu_{[i]})n_1^{1/2}/\sigma > X_i(\omega) \geq X_k(\omega) + (\mu_{[k]} - \mu_{[i]} - h) \times$$

$$n_1^{1/2}/\sigma \quad \forall i\varepsilon s, \; X_i(\omega) < X_k(\omega) + (\mu_{[k]} - \mu_{[i]} - h)n_1^{1/2}/\sigma \quad \forall i\notin s].$$

Hence an increase in $\mu_{[k]}$ can only reduce the set s, and thus $s' \subseteq s$.

<u>Case 2</u>: $\overline{X}^{(1)}_{(j)}(\omega) = \overline{X}^{(1)}_{[k]}(\omega)$ for $j \neq k$ and the set $s\varepsilon\mathscr{S}$ is selected.

Thus the following event occurs:

$$[X_j(\omega) + (\mu_{[j]} - \mu_{[k]})n_1^{1/2}/\sigma > X_k(\omega) \geq X_j(\omega) + (\mu_{[j]} - \mu_{[k]} - h)\times$$

$$n_1^{1/2}/\sigma, X_j(\omega) + (\mu_{[j]} - \mu_{[i]})n_1^{1/2}/\sigma > X_i(\omega) \geq X_j(\omega) + (\mu_{[j]} - \mu_{[i]} - h)\times$$

$$n_1^{1/2}/\sigma \quad \forall i \varepsilon s,\ X_i(\omega) < X_j(\omega) + (\mu_{[j]} - \mu_{[i]} - h)n_1^{1/2}/\sigma \quad \forall i \notin s].$$

An increase in $\mu_{[k]}$ can at the most violate the first inequality in the above event. In that case we have the event

$$[X_k(\omega) + (\mu'_{[k]} - \mu'_{[i]})n_1^{1/2}/\sigma > X_i(\omega)\ \forall\ i \neq k,\ X_i(\omega) < X_k(\omega)$$

$$+ (\mu'_{[k]} - \mu'_{[i]} - h)n_1^{1/2}/\sigma\ \forall\ i \notin s].$$

Thus the new set $s' \subseteq s$.

From Case 1 and Case 2 we obtain

(2.2.14) $$[Y_k(\omega) + (\mu_{[k]} - \mu_{[i]})n_1^{1/2}/\sigma > Y_i(\omega) \quad \forall i \varepsilon s]$$

$$\Longrightarrow [Y_k(\omega) + (\mu'_{[k]} - \mu'_{[i]})n_1^{1/2}/\sigma > Y_i(\omega) \quad \forall i \varepsilon s'].$$

From (2.2.13) and (2.2.14) we find that $\omega \varepsilon A(\underline{\mu}')$ and $A(\underline{\mu}) \subseteq A(\underline{\mu}')$. Hence $P\{A(\underline{\mu}')\} \geq P\{A(\underline{\mu})\}$, and $P_{\underline{\mu}'}(CS|R_1) \geq P_{\underline{\mu}}(CS|R_1)$ which completes the proof.

Corollary 2.2.2: For $\delta \geq 0$, $P_{\underline{\mu}(\delta)}(CS|R_1)$ is non-decreasing in δ, the other parameters being kept fixed. In particular for $k = 2$, $P_{\underline{\mu}}(CS|R_1)$ achieves its infimum over the preference zone $\Omega(\delta^*)$ at any $\underline{\mu}$ satisfying $\mu_{[2]} - \mu_{[1]} = \delta^*$.

Proof: The proof is immediate upon noting that because of the translation invariant nature of R_1, $P_{\underline{\mu}}(CS|R_1)$ depends on $\underline{\mu}$ only through the differences $\delta_{ki}(1 \leq i \leq k - 1)$ and then applying Lemma 2.2.1.

We have not been able to use the previous method of proof to prove our intuitive conjecture that $P_{\underline{\mu}}(CS|R_1)$ is decreasing in $\mu_{[i]}$ for $i \neq k$. Alternative methods of proof also failed to arrive at this result. Thus the monotonicity of $P_{\underline{\mu}}(CS|R_1)$ in $\delta_{ki}(1 \leq i \leq k - 1)$ remains an open question. We believe that the following conjecture made by Alam [1970] is correct.

Conjecture 2.2.1 (Alam): For $k > 2$, the infimum (w.r.t.$\underline{\mu}$) of $P_{\underline{\mu}}(CS|R_1)$ over $\Omega(\delta^*)$ is achieved at any $\underline{\mu}$ satisfying $\mu_{[1]} = \cdots = \mu_{[k-1]} = \mu_{[k]} - \delta^*$. Thus the slippage configuration $\mu_{[1]} = \cdots = \mu_{[k-1]} = \mu_{[k]} - \delta^*$ is a LFC for rule R_1.

In §3.1 we derive a conservative lower bound for $P_{\underline{\mu}}(CS|R_1)$, and we prove that the bound achieves its infimum over $\Omega(\delta^*)$ at the conjectured LFC. For $k > 2$ we use this lower bound as a basis for computing tables from which (n_1,n_2,h) can be determined to guarantee (1.1.1). The computational advantages associated with this lower bound are also discussed in §3.1.

Remark 2.2.3: Suppose that the μ_i are location (scale) parameters (no assumption of normality being made). If we have available at each stage of the experiment real sufficient statistics, say, $Z_i^{(1)}$ and Z_i for μ_i which is a location (scale) parameter in the induced marginal distributions of $Z_i^{(1)}$ and Z_i $(1 \leq i \leq k)$ then for selecting the population associated with $\mu_{[k]}$ we can state two-stage rules analogous to R_1 (see e.g. Gupta [1965]). The method of proof of Lemma 2.2.1 would also apply in this general case. Further if the induced distributions have the property of monotone likelihood ratio (MLR) then Theorem 2.3.1 (to be proved in the next section) would also apply in this general case. We have restricted consideration to the case of normality both for the purpose of specificity, and also because of the importance of this special case.

We shall note a few more properties of the PCS expression before closing this section.

1. If (n_1,n_2,h) guarantee the probability requirement (1.1.1) when used with rule R_1 and if n_0 is the corresponding single stage sample size, then we have

$$n = n_1 + n_2 \geq n_0. \qquad (2.2.15)$$

This follows by considering a modified rule R_1' which includes all k populations in the second stage. Let $P^{*'}$ be the infimum of the PCS over $\Omega(\delta^*)$ for the modified rule. Then clearly $P^{*'} \geq P^*$. But the modified rule is simply a single-stage rule with $n = n_1 + n_2$ as the common sample size per population. Hence $n_1 + n_2 \geq n_0$.

2. It is intuitively clear that for any values of n_1, n_2 and h we have

$$\lim_{\mu_{[1]} \to \mu_{[k]}} P_{\underline{\mu}}(CS|R_1) = \frac{1}{k} . \tag{2.2.16}$$

We were able to obtain special proofs of (2.2.16) for k = 2 and 3. But we were unable to obtain a general proof for $k \geq 2$.

§2.3 Expected total sample size of rule R_1 and its behavior:

§2.3.1 A general expression for $E_{\underline{\mu}}(N|R_1)$:

In order to employ the U-minimax criterion that was proposed in §2.1.2, it is necessary to know the set of parameter points in Ω at which the supremum of $E_{\underline{\mu}}(N|R_1)$ occurs, and the corresponding expression for $E_{\underline{\mu}}(N|R_1)$. The general expression for $E_{\underline{\mu}}(N|R_1)$ would also be useful. We derive it in the following lemma.

Lemma 2.3.1: For any $\underline{\mu}\varepsilon\Omega$

$$E_{\underline{\mu}}(N|R_1) = kn_1 + n_2 \sum_{i=1}^{k} \int_{-\infty}^{\infty} \left\{ \prod_{j=1, j\neq i}^{k} \Phi[x + (\delta_{ij} + h)n_1^{1/2}/\sigma] - \prod_{j=1, j\neq i}^{k} \Phi[x + (\delta_{ij} - h)n_1^{1/2}/\sigma] \right\} d\Phi(x). \tag{2.3.1}$$

Proof: Let T denote the (random) size of the subset I where the set I is defined by (2.1.1). Define a new r.v. T' such that

$$(2.3.2) \qquad T' = \begin{cases} 0 & \text{if} \quad T = 1. \\ T & \text{if} \quad T > 1. \end{cases}$$

Then

$$E_{\underline{\mu}}(N|R_1) = kn_1 + n_2 E_{\underline{\mu}}(T')$$

$$= kn_1 + n_2\{E_{\underline{\mu}}(T|R_1) - P_{\underline{\mu}}(T = 1|R_1)\}$$

$$= kn_1 + n_2 \left\{ \sum_{i=1}^{k} P_{\underline{\mu}}(\overline{X}^{(1)}_{(i)} \geq \overline{X}^{(1)}_{(j)} - h \quad \forall j \neq i) - \right.$$

$$\left. \sum_{i=1}^{k} P_{\underline{\mu}}(\overline{X}^{(1)}_{(i)} > \overline{X}^{(1)}_{(j)} + h \quad \forall j \neq i) \right\}$$

$$= kn_1 + n_2 \left\{ \sum_{i=1}^{k} P[X_i + (\delta_{ij} + h)n_1^{1/2}/\sigma \geq X_j \quad \forall j \neq i] - \right.$$

$$\left. \sum_{i=1}^{k} P[X_i + (\delta_{ij} - h)n_1^{1/2}/\sigma > X_j \quad \forall j \neq i] \right\}$$

$$= kn_1 + n_2 \sum_{i=1}^{k} \int_{-\infty}^{\infty} \left\{ \prod_{j=1, j\neq 1}^{k} \Phi[x + (\delta_{ij} + h)n_1^{1/2}/\sigma] - \right.$$

$$\left. \prod_{j=1, j\neq 1}^{k} \Phi[x + (\delta_{ij} - h)n_1^{1/2}/\sigma] \right\} d\Phi(x).$$

This completes the proof.

§2.3.2 Determination of the supremum of $E_{\underline{\mu}}(N|R_1)$:

We now derive the main result of this section in the following theorem.

Theorem 2.3.1: For fixed n_1, n_2, h, σ and k, the supremum of $E_{\underline{\mu}}(N|R_1)$ over Ω occurs at any $\underline{\mu}$ satisfying $\mu_{[1]} = \mu_{[2]} = \cdots = \mu_{[k]}$ (EMC). Further

(2.3.3)
$$\sup_{\Omega} E_{\underline{\mu}}(N|R_1) = kn_1 + kn_2 \int_{-\infty}^{\infty} \left\{\Phi^{k-1}(x + hn_1^{1/2}/\sigma) - \Phi^{k-1}(x - hn_1^{1/2}/\sigma)\right\} d\Phi(x).$$

Proof: We must show that $\{E_{\underline{\mu}}(T|R_1) - P_{\underline{\mu}}(T=1|R_1)\}$ achieves its supremum over Ω at the EMC. Gupta [1965] has shown this for $E_{\underline{\mu}}(T|R_1)$. Thus it only remains for us to show that $P_{\underline{\mu}}(T=1|R_1)$ achieves its infimum at the EMC. We use the method due to Gupta [1965].

Set $\mu_{[1]} = \mu_{[2]} = \cdots = \mu_{[\ell]} = \mu < \mu_{[\ell+1]}$ for some ℓ $(1 \leq \ell \leq k = 1)$ and define $\delta_i = \mu_{[i]} - \mu$ for $\ell + 1 \leq i \leq k$. Define

(2.3.4)
$$Q(\mu) = P_{\underline{\mu}}\{T=1|R_1;\ \mu_{[1]} = \cdots = \mu_{[\ell]} = \mu\}$$
$$= \ell \int_{-\infty}^{\infty} \Phi^{\ell-1}(x - hn_1^{1/2}/\sigma) \prod_{j=\ell+1}^{k} \Phi[x - (\delta_j + h)n_1^{1/2}/\sigma]d\Phi(x)$$
$$+ \sum_{i=\ell+1}^{k} \int_{-\infty}^{\infty} \Phi^{\ell}[x + (\delta_i - h)n_1^{1/2}/\sigma] \prod_{j=\ell+1}^{k} \Phi[x + (\delta_{ij} - h)n_1^{1/2}/\sigma]d\Phi(x).$$

After differentiating w.r.t.μ and then interchanging the order of integration and summation in the first term and making appropriate substitutions we obtain

$$(2.3.5)\quad \frac{dQ(\mu)}{d\mu} = \frac{\ell n_1^{1/2}}{\sigma} \sum_{i=\ell+1}^{k} \int_{-\infty}^{\infty} \Phi^{\ell-1}[x + (\delta_i - h)n_1^{1/2}/\sigma] \prod_{j=\ell+1, j\neq i}^{k}$$

$$\Phi[x + (\delta_{ij} - h)n_1^{1/2}/\sigma]\{\phi(x - hn_1^{1/2}/\sigma)\phi(x + \delta_i n_1^{1/2}/\sigma) -$$

$$\phi(x + (\delta_i - h)n_1^{1/2}/\sigma)\phi(x)\}dx.$$

$$\leq 0.$$

The last result is obtained by noting that the quantity inside the braces in (2.3.5) is non-negative for every x and i for $\ell + 1 \leq i \leq k$ due to the MLR property of the normal density function. Now it follows that Q is non-increasing in μ and is in fact strictly decreasing in μ if h and n_1 are positive. Thus subject to $\mu_{[1]} = \cdots = \mu_{[\ell]}$, we see that $P_{\underline{\mu}}(T=1|R_1)$ is minimized by increasing the common value μ until $\mu = \mu_{[\ell+1]}$. Since this is true for each $\ell \leq k - 1$, it follows that $P_{\underline{\mu}}(T=1|R_1)$ is minimized and hence $E_{\underline{\mu}}(N|R_1)$ is maximized over Ω by setting all the $\mu_{[i]}$'s equal in (2.3.1).

<u>Corollary 2.3.1</u>: The supremum of $E_{\underline{\mu}}(N|R_1)$ over $\Omega(\delta^*)$ occurs at the slippage configuration $\mu_{[1]} = \cdots = \mu_{[k-1]} = \mu_{[k]} - \delta^*$ and

$$(2.3.6)\quad \sup_{\Omega(\delta^*)} E_{\underline{\mu}}(N|R_1) = kn_1 + n_2\left[\int_{-\infty}^{\infty} \left\{\Phi^{k-1}[x + (\delta^* + h)n_1^{1/2}/\sigma] -\right.\right.$$

$$\Phi^{k-1}[x + (\delta^* - h)n_1^{1/2}/\sigma] \Big\} d\Phi(x) + (k-1) \int_{-\infty}^{\infty} \Big\{ \Phi^{k-2}(x + hn_1^{1/2}/\sigma)$$

$$\Phi[x - (\delta^* - h)n_1^{1/2}/\sigma] - \Phi^{k-2}(x - hn_1^{1/2}/\sigma)\Phi[x - (\delta^* + h)n_1^{1/2}/\sigma] \Big\} d\Phi(x) \Big] .$$

Thus we have achieved the objective of determining the supremum of the expected total sample size function over Ω and $\Omega(\delta^*)$, respectively. We now can state the optimization problems associated with the U-minimax criterion and the R-minimax criterion. This is done in the following section.

§2.4 Optimization problems:

§2.4.1 Discrete optimization problems:

For the sake of convenience, in this section we state the optimization problems associated with the two design criteria for the general $k \geq 2$ case assuming the Conjecture 2.2.1 regarding the LFC to be true for $k > 2$.

We denote by $\psi(n_1,n_2,h;\ \delta^*,\sigma,k)$ the value of $P_{\underline{\mu}}(CS|R_1)$ at the conjectured LFC; the value of $\psi(n_1,n_2,h;\ \delta^*,\sigma,k)$ is obtained by substituting $\delta = \delta^*$ in (2.2.6). For the U-minimax criterion we have

Discrete optimization problem (U-minimax criterion):

For specified $\{\delta^*,P^*\}$ and given σ and k, choose the three constants (n_1,n_2,h) necessary to implement rule R_1 so as to

$$(2.4.1)\quad \begin{cases} \text{minimize } kn_1 + kn_2 \displaystyle\int_{-\infty}^{\infty} \left\{\Phi^{k-1}(x + hn_1^{1/2}/\sigma) - \Phi^{k-1}(x-hn_1^{1/2}/\sigma)\right\} \times \\ \qquad d\Phi(x), \\ \text{subject to } \psi(n_1,n_2,h;\ \delta^*,\sigma,\ k) \geq P^*, \\ n_1,n_2 \text{ non-negative integers} \\ \text{and } h \geq 0. \end{cases}$$

As in (2.1.4) we denote by $(\hat{n}_1,\hat{n}_2,\hat{h})$ a solution to the above optimization problem and the corresponding rule R_1 by $\hat{R}_1$.

For the R-minimax criterion, one only need replace the objective function in (2.4.1) by the expression (2.3.6). As in (2.1.3) we denote by $(\tilde{n}_1,\tilde{n}_2,\tilde{h})$ a solution to the corresponding optimization problem and the corresponding rule R_1 by $\tilde{R}_1$.

The problem (2.4.1) is an extremely complicated integer programming problem with a non-linear objective function. Ignoring the computational difficulties involved in the evaluation of $\psi(n_1,n_2,h;\ \delta^*,\sigma,k)$ expression, we remark here that (2.4.1) can be solved in principle by enumeration although the search is likely to be a costly one. More importantly, since the solution depends on δ^*, one requires a separate solution not just for each k and P^*-value but also for each δ^*-value. In view of these difficulties we shall (temporarily) abandon the requirement that n_1, n_2 must be integers; we reparametrize the problem and regard the new design parameters (which are functions of n_1,n_2 and h) as continuous. We use this continuous version as a large sample approximation. This continuous version is given in the following section.

§2.4.2 Continuous optimization problems:

We define the new design variables

$$c = \frac{hn_1^{1/2}}{\sigma}, \quad d_1 = \frac{\delta^* n_1^{1/2}}{\sigma}, \quad d_2 = \frac{\delta^* n_2^{1/2}}{\sigma}. \tag{2.4.2}$$

We regard c, d_1 and d_2 as non-negative continuous variables and note that the value of $P_{\underline{\mu}}(CS|R_1)$ at the conjectured LFC depends on δ^*, σ, n_1, n_2 and h only through c, d_1 and d_2. If we denote the corresponding value of the PCS by $\psi(c, d_1 d_2; k)$ then from (2.2.6) we have

$$\begin{aligned} \psi(c,d_1d_2;k) = {} & \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\Big\{\Phi(x+d_1-c) + \int_{x+d_1-c}^{x+d_1} \Phi\Big[y + \frac{d_1}{d_2}(x-z) + \\ & (d_1^2+d_2^2)/d_2\Big]d\Phi(z)\Big\}^{k-1} d\Phi(y)\,d\Phi(x) + (k-1)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\Bigg[\int_{x-d_1-c}^{x-d_1} \\ & \int_{y+(x-u)(d_1/d_2)-(d_1^2+d_2^2)/d_2}^{\infty} \Big\{\Phi(x-c) + \int_{x-c}^{x} \Phi\Big[v + \frac{d_1}{d_2}(u-z) + \\ & (d_1^2+d_2^2)/d_2\Big]d\Phi(z)\Big\}^{k-2} d\Phi(v)\,d\Phi(u)\Bigg]\, d\Phi(y)\,d\Phi(x). \end{aligned} \tag{2.4.3}$$

Clearly the optimization problem remains unaltered if we multiply the objective function by a constant $(\delta^*/\sigma)^2$. By this device we obtain the objective function in terms of the new design variables. Then for the U-minimax criterion we have

Continuous optimization problem (U-minimax criterion):

For specified P* and given k find (c,d_1,d_2) so as to

$$(2.4.4)\quad \begin{cases} \text{minimize } kd_1^2 + kd_2^2 \displaystyle\int_{-\infty}^{\infty} \left\{ \Phi^{k-1}(x + c) - \Phi^{k-1}(x-c) \right\} d\Phi(x), \\ \text{subject to } \psi(c,d_1 d_2;\ k) \geq P^*, \\ \text{and } c,d_1,d_2 \geq 0. \end{cases}$$

We denote by $(\hat{c},\hat{d}_1,\hat{d}_2)$ a solution to (2.4.4).

Correspondingly for the R-minimax criterion we have

Continuous optimization problem (R-minimax criterion):

For specified P* and given k find (c,d_1,d_2) so as to

$$(2.4.5)\quad \begin{cases} \text{minimize } kd_1^2 + d_2^2 \left[\displaystyle\int_{-\infty}^{\infty} \left\{ \Phi^{k-1}(x + d_1 + c) - \Phi^{k-1}(x-d_1 + c) \right\} \right. \\ \quad d\Phi(x) + (k-1) \displaystyle\int_{-\infty}^{\infty} \left\{ \Phi^{k-2}(x + c)\Phi(x-d_1 + c) - \right. \\ \quad \left. \left. \Phi^{k-2}(x-c)\Phi(x-d_1-c) \right\} d\Phi(x) \right], \\ \text{subject to } \psi(c,d_1,d_2;\ k) \geq P^*, \\ \text{and } c,d_1,d_2 \geq 0. \end{cases}$$

We denote by $(\tilde{c},\tilde{d}_1,\tilde{d}_2)$ a solution to (2.4.5).

An experimenter would use these values as follows: Suppose he has adopted the U-minimax criterion. Then for given k and P* he would read the corresponding $(\hat{c},\hat{d}_1,\hat{d}_2)$ values from an appropriate table. He would compute an approximately optimal integer solution <u>(approximate solution)</u> denoted by $(\hat{n}_1',\hat{n}_2',\hat{h}')$ as follows:

$$(2.4.6)\quad \hat{n}_1' = \left[\left(\frac{\hat{d}_1\sigma}{\delta^*}\right)^2\right] + 1,\ \hat{n}_2' = \left[\left(\frac{\hat{d}_2\sigma}{\delta^*}\right)^2\right] + 1,\ \hat{h} = \hat{c}\sigma/\hat{n}_1'^{1/2}$$

He would use rule R_1 with $(\hat{n}_1',\hat{n}_2',\hat{h}')$.

In §2.6.2 we make a study of the expected total sample sizes associated with $(\hat{n}_1,\hat{n}_2,\hat{h})$ and $(\hat{n}_1',\hat{n}_2',\hat{h}')$. One would clearly expect that for small δ*-values and/or large P*-values there would be little change in the expected total sample size by using the approximate result obtained from the continuous solution.

§2.5 <u>Relative efficiency of</u> R_0 <u>w.r.t.</u> $\hat{R}_1$ <u>and</u> $\tilde{R}_1$:

As a measure of relative performance of two competing rules R_0 and R_1 guaranteeing the same probability requirement (1.1.1) we consider the ratio of the expected total sample size of rule R_1 to the single-stage total sample size kn_0. We shall consider only the continuous versions of rules $\hat{R}_1$, $\tilde{R}_1$ and R_0. Although such rules are really approximations to the exact discrete solutions we use the same notation to denote these rules. We define the relative efficiency (RE) of rule R_0 w.r.t. rule $\hat{R}_1$ as follows

$$(2.5.1) \qquad RE_{\underline{\mu}}(\delta^*,P^*,k;\ \hat{R}_1/R_0) = [k\hat{d}_1^2 + \hat{d}_2^2 \sum_{i=1}^{k} \int_{-\infty}^{\infty} \Big\{ \prod_{j=1,j\neq i}^{k} \Phi(x + \frac{\delta_{ij}\hat{d}_1}{\delta^*} + \hat{c}) - \prod_{j=1,j\neq i}^{k} \Phi(x + \frac{\delta_{ij}\hat{d}_1}{\delta^*} - \hat{c}) \Big\} d\Phi(x)]/kd_0^2,$$

where d_0 is given by (1.1.2) and $(\hat{c},\hat{d}_1,\hat{d}_2)$ is given by (2.4.4).

Remark 2.5.1: Note that the above definition of RE differs from the conventional definition where the quantity defined in (2.5.1) would be regarded as RE of rule $\hat{R}_1$ w.r.t. rule R_0. We feel that our definition agrees with the intuitive meaning of efficiency.

Remark 2.5.2: RE depends on δ^* only through the ratios δ_{ij}/δ^* $(1 \leq i \neq j \leq k)$. For the EMC all the ratios equal 0 and for the conjectured LFC all the ratios equal 0 or 1. Thus RE_{EMC} and RE_{LFC} are independent of δ^*. Further since rule R_0 will always be used for the purpose of comparison with rules $\hat{R}_1$, $\tilde{R}_1$ and some other rules, we shall omit the dependence of the corresponding RE's on R_0 from the notation. Thus we write

$$(2.5.2) \qquad RE_{EMC}(P^*,k;\ \hat{R}_1) = [\hat{d}_1^2 + \hat{d}_2^2 \int_{-\infty}^{\infty} \Big\{ \Phi^{k-1}(x + \hat{c}) - \Phi^{k-1}(x - \hat{c}) \Big\} d\Phi(x)]/d_0^2.$$

We note that $1 \geq RE_{EMC}(P^*,k;\ \hat{R}_1) \geq RE_{\underline{\mu}}(\delta^*,P^*,k;\ \hat{R}_1)$ for all $\underline{\mu}\varepsilon\Omega$.

For rule $\tilde{R}_1$ we can define quantities analogous to (2.5.1) and (2.5.2). In particular we give the expression for $RE_{\underline{\mu}(\delta^*)}$ for rule $\tilde{R}_1$

$$(2.5.3)\quad RE_{\underline{\mu}(\delta^*)}(P^*,k;\ \tilde{R}_1) = \left[k\tilde{d}_1^2 + \tilde{d}_2^2\left\{\int_{-\infty}^{\infty}\left\{\Phi^{k-1}(x + \tilde{d}_1 + \tilde{c}) - \Phi^{k-1}(x - \tilde{d}_1 + \tilde{c})\right\}d\Phi(x) + (k-1)\int_{-\infty}^{\infty}\left[\Phi^{k-2}(x + \tilde{c})\Phi(x - \tilde{d}_1 + \tilde{c}) - \Phi^{k-2}(x - \tilde{c})\Phi(x - \tilde{d}_1 - \tilde{c})\right]d\Phi(x)\right\}\right]/kd_0^2$$

where $(\hat{c},\hat{d}_1,\hat{d}_2)$ is given by (2.4.5). Note that
$1 \geq RE_{\underline{\mu}(\delta^*)}(P^*,k;\ \tilde{R}_1) \geq RE_{\underline{\mu}}(\delta^*,P^*,k;\ \tilde{R}_1)$ for all $\underline{\mu}\varepsilon\Omega(\delta^*)$.

In the next section we give tables of constants necessary to implement rules $\hat{R}_1$ and $\tilde{R}_1$ for k = 2 and illustrate their use with a numerical example.

§2.6 Tables to implement rules $\hat{R}_1$ and $\tilde{R}_1$ (k = 2):

§2.6.1 Use of the tables:

We have used the computer to solve the continuous versions of the optimization problems for the U-minimax criterion (2.4.4) and the R-minimax criterion (2.4.5) for k = 2 and for P* = 0.55(0.05)0.95, 0.99, 0.999, 0.9995, 0.9999. Our results for the two criteria are given in Tables 2.6.1 and 2.6.2, respectively. These tables are exact in the sense that there is no overprotection in the PCS if the values given in the tables are used. Computational details regarding

Table 2.6.1

Rule $\hat{R}_1$ (U-minimax criterion)

(k = 2)

P^*	$\hat{d}_1$	$\hat{d}_2$	$\hat{c}$	$\hat{p} = \frac{\hat{d}_1^2}{\hat{d}_1^2 + \hat{d}_2^2}$	$\frac{\delta^{*2}}{k\sigma^2} E_{EMC}(N \mid \hat{R}_1)$	$RE_{EMC}(P^*,k;\ \hat{R}_1)$
.9999	4.53969	2.90871	.97215	.70895	24.90828	.90047
.9995	3.97423	2.67083	.95623	.68888	19.36875	.89442
.999	3.70620	2.57123	.94824	.67508	17.02480	.89137
.99	2.71894	2.09056	.91913	.62846	9.50902	.87850
.95	1.86206	1.61556	.88072	.57037	4.68287	.86540
.90	1.42699	1.31318	.85278	.54146	2.81831	.85799
.85	1.13913	1.09297	.84174	.52067	1.83314	.85319
.80	.91577	.90970	.82702	.50333	1.20384	.84982
.75	.72801	.74161	.81999	.49074	.77088	.84719
.70	.56240	.58661	.80783	.47894	.46481	.84515
.65	.41227	.43299	.80202	.47550	.25047	.84356
.60	.26982	.28775	.79714	.46787	.10816	.84248
.55	.13374	.14322	.79140	.46582	.02659	.84194

Table 2.6.2

Rule $\tilde{R}_1$ (R-minimax criterion)

(k = 2)

P*	$\tilde{d}_1$	$\tilde{d}_2$	$\tilde{c}$	$\tilde{p} = \frac{\tilde{d}_1^2}{\tilde{d}_1^2 + \tilde{d}_2^2}$	$\frac{\delta^{*2}}{k\sigma^2} E_{LFC}(N \mid \tilde{R}_1)$	$RE_{LFC}(P^*,k;\ \tilde{R}_1)$	$RE_{EMC}(P^*,k;\ \tilde{R}_1)$
.9999	3.48012	4.41205	1.91618	.38354	14.72598	.53237	1.01811
.9995	3.12390	3.90343	1.69916	.39042	12.14384	.56079	.99273
.999	2.95663	3.66058	1.60260	.39481	11.00005	.57593	.97828
.99	2.29313	2.73708	1.28026	.41243	6.99033	.64581	.92509
.95	1.65831	1.93466	1.05739	.42354	3.90295	.72127	.88560
.90	1.32245	1.49958	.92974	.43748	2.50218	.76175	.86724
.85	1.07886	1.18803	.90951	.45195	1.68966	.78641	.85695
.80	.88255	.96174	.87132	.45715	1.13904	.80408	.85162
.75	.71036	.77072	.84413	.45931	.74330	.81689	.84795
.70	.55468	.59769	.82505	.46273	.45455	.82650	.84547
.65	.40845	.43957	.81002	.46335	.24747	.83335	.84379
.60	.26907	.28914	.79865	.46410	.10758	.83801	.84245
.55	.13378	.14318	.79053	.46613	.02655	.84084	.84195

the construction of the tables are given in Appendix A1.

We give the following numerical example to illustrate the use of the tables: Suppose that an experimenter wishes to decide which one of two normal populations has the larger mean. The two populations have a common $\sigma = 2$ units. The experimenter wishes to make a correct selection with probability atleast 0.99 whenever $\mu_{[2]} - \mu_{[1]} \geq 0.50$ units.

1. U-minimax criterion $(\hat{R}_1)$: For $k = 2$, $P^* = 0.99$ we find from Table 2.6.1 that $\hat{d}_1 = 2.71894$, $\hat{d}_2 = 2.09056$ and $\hat{c} = 0.91913$. Hence using (2.4.6) we find that

$$\hat{n}_1' = \left[\left(\frac{2.71894 \times 2}{0.50}\right)^2\right] + 1 = 112.$$

$$\hat{n}_2' = \left[\left(\frac{2.09056 \times 2}{0.50}\right)^2\right] + 1 = 70.$$

$$\hat{h}' = \frac{0.91913 \times 2}{(112)^{1/2}} = 0.1735.$$

$$E_{EMC}(N|\hat{R}_1) \cong 304.289$$

From Table 1 in Bechhofer [1954] and using (1.1.3) we find that

$$2n_0 = 2 \times \left\{\left[\left(\frac{3.29 \times 2}{0.50}\right)^2\right] + 1\right\} = 348$$

Thus the experimenter would take 112 observations from each population in the first stage. If the difference between the observed first-stage means is more than 0.1735 units, he would stop sampling and assert that the population associated with the larger mean is better. Otherwise he would take 70 additional observations from each of the two populations and assert that the population producing the overall larger mean is better. By operating in this manner, the expected total sample size in the EMC would be somewhat greater than 304. The corresponding single-stage total sample size is 348 and the efficiency of rule R_0 relative to rule $\hat{R}_1$ would be slightly greater than 87.85% when (unknown to the experimenter) $\mu_1 = \mu_2$; clearly the efficiency decreases strictly as $\mu_{[2]} - \mu_{[1]}$ increases.

2. <u>R-minimax criterion</u> ($\tilde{R}_1$): For k = 2, P* = 0.99 we find from Table 2.6.2 that $\hat{d}_1 = 2.29313$, $\hat{d}_2 = 2.73708$ and $\hat{c} = 1.28026$. Hence using (2.4.6) we find that

$$\tilde{n}_1' = \left[\left(\frac{2.29313 \times 2}{0.50}\right)^2\right] + 1 = 85$$

$$\tilde{n}_2' = \left[\left(\frac{2.73708 \times 2}{0.50}\right)^2\right] + 1 = 120$$

$$\tilde{h}' = \frac{1.28026 \times 2}{(85)^{1/2}} = 0.278$$

$$E_{LFC}(N|\tilde{R}_1) \cong 223.691$$

$$E_{EMC}(N|\tilde{R}_1) \cong 321.90$$

The explanation of the various quantities is similar to the one given for rule $\hat{R}_1$.

In the next section we give some numerical results which can be used to compare the total expected sample size obtained by using Table 2.6.1 to give $(\hat{n}_1', \hat{n}_2', \hat{h}')$ for the <u>continuous</u> version rather than using $(\hat{n}_1, \hat{n}_2 \hat{h})$, the true <u>discrete</u> optimal solution.

§2.6.2 <u>A comparison of discrete and approximate solutions</u>:

The proposed comparison depends on the sample sizes involved (which in turn depend on $\{\delta^*, P^*\}$). We shall consider five values of P^*; for each P^* we fix five values of n_0, the single-stage sample size and find the respective δ^*-values. We choose n_0-values in three sample ranges 10-20, 105-125 and 255-275. For each $\{\delta^*, P^*\}$ combination, the discrete optimal solution $(\hat{n}_1, \hat{n}_2, \hat{h}\hat{n}_1^{1/2}/\sigma)$ is found by direct search. The approximate solution $(\hat{n}_1, \hat{n}_2, \hat{c})$ is obtained by first finding the continuous solution $(\hat{c}, \hat{d}, \hat{d}_2)$ using the method given in Appendix A1 and then using (2.4.6). All of the three solutions and the associated expected total sample sizes are given in Table 2.6.3.

The last two columns of the table which show the percentage deviation between the expected total sample sizes associated with the approximate solution and the discrete solution and the percentage deviation between the expected total sample sizes associated with the discrete solution and the continuous solution are of particular interest. We see that the difference between the <u>discrete optimum</u> (the expected total sample size associated with the discrete solution) and the continuous optimum is quite small even for small n_0-values and it decreases further as the sample size is increased. This behavior is not strictly monotone for the obvious reasons. The deviation

Table 2.6.3

A comparison of discrete, approximate and continuous optima (k = 2)

I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV
P^*	n_0	$\frac{\delta^*}{\sigma}$	Discrete optimal solution				Approximate optimal solution				Continuous Optimum	$\frac{XI-VII}{XII}\times 10^2$	$\frac{VII-XII}{XII}\times 10^2$
			$\hat{n}_1$	$\hat{n}_2$	$\frac{\hat{h}\hat{n}_1^{1/2}}{\sigma}$	$\frac{1}{k}E_{EMC}(N)$	$\hat{n}_1'$	$\hat{n}_2'$	$\hat{c}$	$\frac{1}{k}E_{EMC}(N)$	$\frac{1}{k}E_{EMC}(N)$		
	10	1.3820	7	4	.9255	8.9470	8	4	.9482	9.9899	8.9137	11.6566	.3729
	15	1.1284	11	5	.8986	13.3741	11	6	.9482	13.9848	13.3705	4.5660	.0272
0.999	20	.9772	14	7	1.0685	17.8506	15	7	.9482	18.4823	17.8274	3.5385	.1304
	110	.4167	79	38	.9570	98.0537	80	39	.9482	99.4015	98.0520	1.3746	.0017
	260	.2710	187	90	.9482	231.7689	188	91	.9482	233.2695	231.7644	.6474	.0020
	10	.5731	6	6	.8118	8.6044	7	6	.8528	9.7210	8.5799	12.9770	.2856
0.90	15	.4680	9	8	.9204	12.8789	10	8	.8528	13.6280	12.8698	5.8167	.0707
	20	.4053	12	11	.8881	17.1697	13	11	.8528	17.9885	17.1601	4.7685	.0560
	125	.1621	77	66	.8636	107.2643	78	66	.8528	107.9308	107.2428	.6214	.0200
	275	.1093	170	145	.8563	235.9994	171	145	.8528	236.7571	235.9538	.3211	.0193
	10	.3764	6	6	.7770	8.5037	6	6	.8270	8.6479	8.4984	1.6944	.0627
	15	.3073	9	9	.7769	12.7552	9	9	.8270	12.9718	12.7472	1.6977	.0628
0.80	20	.2661	12	12	.7769	17.0068	12	12	.8270	17.2957	16.9961	1.6990	.0630
	105	.1162	62	61	.8385	89.2514	63	62	.8270	90.3612	89.2341	1.2435	.0195
	255	.0745	151	149	.8271	216.7618	152	150	.8270	218.1966	216.7239	.6619	.0175

Table 2.6.3 (cont'd)

A comparison of discrete, approximate and continuous optima

I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV
P^*	n_0	$\frac{\delta^*}{\sigma}$	Discrete optimal solution				Approximate solution				Continuous Optimum	$\frac{XI-VII}{XII}\times 10^2$	$\frac{VII-XII}{XII}\times 10^2$
			$\hat{n}_1$	$\hat{n}_2$	$\hat{h}\hat{n}_1^{1/2}/\sigma$	$\frac{1}{k}E_{EMC}(N)$	$\hat{n}_1'$	$\hat{n}_2'$	$\hat{c}$	$\frac{1}{k}E_{EMC}(N)$	$\frac{1}{k}E_{EMC}(N)$		
.70	10	.2345	6	6	.7608	8.4564	6	7	.8078	9.0251	8.4519	6.7249	.05336
	15	.1915	9	9	.7608	12.6842	9	10	.8078	13.3215	12.6774	5.0249	.05363
	20	.1658	12	12	.7606	16.9115	12	13	.8078	17.6180	16.9024	4.1775	.05360
	115	.0692	66	72	.8107	97.2123	67	73	.8078	98.5471	97.2056	1.3731	.00688
	265	.0456	153	164	.8087	223.9403	153	167	.8078	225.1693	223.9278	.5488	.00560
.60	10	.1133	6	6	.7527	8.4327	6	7	.7971	8.9891	8.4254	6.5988	.08581
	15	.1925	9	9	.7527	12.6487	9	10	.7971	13.2701	12.6379	4.9129	.08571
	20	.0801	11	13	.8465	16.8561	12	13	.7971	17.5512	16.8488	4.1236	.04327
	125	.0321	71	80	.8006	105.2928	72	81	.7971	106.5881	105.2920	1.2303	.00074
	275	.0216	156	177	.7974	231.6035	157	178	.7971	233.0087	231.6018	.6067	.00074

of the approximate optimum from the discrete optimum is substantial for the small sample size range but it decreases rapidly as the sample size is increased. For the range of P*-values studied, the <u>approximate</u> optimum is always less than n_0, the corresponding single stage sample size. However for P*-values close to 1 and for very small sample sizes, the approximate optimum may in fact exceed n_0.

In Table 2.6.4 for a particular $\{\delta^*, P^*\}$ combination we give an indication of how the expected total sample size function behaves as n_1 and n_2 are varied systematically.

Table 2.6.4

Some (n_1, n_2, c) values in the vicinity of the optimum

$(\delta^* = .27103,\ P^* = .999,\ k = 2,\ \sigma = 1,\ n_0 = 260)$

n	n_1	n_2	c	$\frac{\delta^{*2}}{k\sigma^2}E_{EMC}(N)$	
	185	91	.98570	17.02678	
	186	90	.97320	17.02577	
276	187	89	.96069	17.02536	*
	188	88	.94819	17.02555	
	189	89	.93569	17.02636	
	185	92	.97300	17.02645	
	186	91	.96058	17.02548	
277	187	90	.94817	17.02511	**
	188	89	.93576	17.02535	
	189	88	.92355	17.02622	
	185	93	.96118	17.02774	
	186	92	.94884	17.02679	
278	187	91	.93651	17.02645	*
	188	90	.92419	17.02672	
	189	89	.91186	17.02762	

(* indicates the minimum for a fixed n;

** indicates the global minimum)

Thus for each fixed n a minimum is achieved for a certain combination of (n_1,n_2,c) values; c being chosen to satisfy the P*-requirement for the given (n_1,n_2) values. This minimum first decreases and then increases as n is increased. No occurrence of multiple local minima was observed.

CHAPTER 3

FURTHER TOPICS IN TWO-STAGE RULES FOR THE NORMAL MEANS PROBLEM (COMMON KNOWN VARIANCE)

§3.0 Introduction:

In the present chapter we continue the studies of the previous chapter and extend them in several directions. In §3.1 we derive a lower bound on the PCS associated with rule R_1 and obtain the LFC associated with that bound. In §3.2 we indicate the computational advantages derived from using the lower bound, and construct tables of constants necessary to implement rule R_1 based on the conservative lower bound for k > 2. We denote the corresponding conservative two-stage rule by $\overline{R}_1$. We study the gains achieved by $\overline{R}_1$ over the single-stage rule R_0 and also the losses incurred due to not using the exact rule R_1. In §3.3 we study the asymptotic ($P^* \to 1$) behavior of our two-stage rule with the U-minimax criterion ($\hat{R}_1$) and the R-minimax criterion ($\tilde{R}_1$). We show that the asymptotic relative efficiency (ARE) in the EMC of single-stage rule R_0 w.r.t. rule $\hat{R}_1$ is 1 and the ARE in the LFC of rule R_0 w.r.t. rule $\tilde{R}_1$ is 1/4 for every $k \geq 2$. In §3.3 we propose a "come-back" type modification R_2 of our basic rule R_1 and give the expression for the asoociated PCS. We give extensive MC sampling results comparing the PCS achieved by the conservative rule $\overline{R}_1$, the exact rule R_1 and the modified rule R_2 when the μ_i's are in the conjectured LFC. In §3.5 we propose another come-back type rule R_3 and derive its properties for k = 2. Some avenues for further research on this problem are described in §3.6.

§3.1 Some useful bounds on PCS of rule R_1

§3.1.1 Derivation of the bounds:

In this section we derive an upper and a lower bound on $P_{\underline{\mu}}(CS|R_1)$. The lower bound is particularly useful since we can prove that the value of the lower bound achieves its infimum over $\Omega(\delta^*)$ at the conjectured LFC. Thus if we set the value of this lower bound equal to P^* at any $\underline{\mu}$ such that $\mu_{[1]} = \mu_{[k-1]} = \mu_{[k]} - \delta^*$, then any values of n_1, n_2, h which guarantee this equality will also guarantee the probability requirement (1.1.1) for rule R_1. We shall denote by $\overline{R}_1$, the conservative rule based on (n_1, n_2, h) arrived at through the use of the lower bound.

A secondary advantage of the above mentioned lower bound is that since it is in the form of a sum of two univariate iterated integrals, its value can be calculated at a very fast speed on a digital computer. This advantage is very meaningful since for the exact rule determination of (n_1, n_2, h) to meet a certain design criterion would require a large number of computations of the expression (2.4.3) which involves four-variate integrals.

The following theorem provides desired bounds on $P_{\underline{\mu}}(CS|R_1)$.

Theorem 3.1.1: For any $\underline{\mu}\varepsilon\Omega$ we have

(3.1.1)
$$\int_{-\infty}^{\infty} \prod_{i=1}^{k-1} \Phi\left[x + \frac{(\delta_{ki} + h)n_1^{1/2}}{\sigma}\right] d\Phi(x) + \int_{-\infty}^{\infty} \prod_{i=1}^{k-1}$$

$$\Phi\left[x + \frac{\delta_{ki}n^{1/2}}{\sigma}\right] d\Phi(x) - 1 \leq P_{\underline{\mu}}(CS|R_1) \leq \int_{-\infty}^{\infty} \prod_{i=1}^{k-1} \Phi\left[x + \frac{(\delta_{ki} + h)n_1^{1/2}}{\sigma}\right] d\Phi(x).$$

Proof: First consider the lower bound.

$$1 - P_{\underline{\mu}}(CS|R_1) = P_{\underline{\mu}}(\text{Incorrect Selection}|R_1).$$

$$\leq P_{\underline{\mu}}(\overline{X}_{(k)}^{(1)} < \overline{X}_{(i)}^{(1)} - h \text{ for some } i \neq k)$$

$$+ P_{\underline{\mu}}(\overline{X}_{(k)} < \overline{X}_{(i)} \text{ for some } i \neq k).$$

$$= 1 - P_{\underline{\mu}}(\overline{X}_{(k)}^{(1)} \geq \overline{X}_{(i)}^{(1)} - h \quad \forall i \neq k) + 1$$

$$- P_{\underline{\mu}}(\overline{X}_{(k)} > \overline{X}_{(i)} \quad \forall i \neq k).$$

$$= 2 - P(X_k + \frac{(\delta_{ki} + h)n_1^{1/2}}{\sigma} > X_i \quad \forall i \neq k)$$

$$- P(Y_k + \frac{\delta_{ki}n^{1/2}}{\sigma} > Y_i \quad \forall i \neq k).$$

$$= 2 - \int_{-\infty}^{\infty} \prod_{i=1}^{k-1} \Phi\left[x + \frac{\delta_{ki}n^{1/2}}{\sigma}\right] d\Phi(x)$$

$$- \int_{-\infty}^{\infty} \prod_{i=1}^{k-1} \Phi\left[x + \frac{\delta_{ki}n^{1/2}}{\sigma}\right] d\Phi(x).$$

A rearrangement of the terms gives the desired lower bound. For the upper bound simply note that

$$P_{\underline{\mu}}(CS|R_1) \leq P_{\underline{\mu}}(\Pi_{(k)} \text{ is included in the set } I|R_1).$$

$$= P_{\underline{\mu}}(\overline{X}^{(1)}_{(k)} \geq \overline{X}^{(1)}_{(i)} - h \;\; \forall i \neq k).$$

$$= \int_{-\infty}^{\infty} \prod_{i=1}^{k-1} \Phi\left[x + \frac{(\delta_{ki} + h)n_1^{1/2}}{\sigma}\right] d\Phi(x).$$

Corollary 3.1.1: For all $\underline{\mu}\varepsilon\Omega(\delta^*)$

(3.1.2)
$$P_{\underline{\mu}}(CS|R_1) \geq \int_{-\infty}^{\infty}\Phi^{k-1}\left[x + \frac{(\delta^* + h)n_1^{1/2}}{\sigma}\right]d\Phi(x) + \int_{-\infty}^{\infty}\Phi^{k-1}\left[x + \frac{\delta^* n^{1/2}}{\sigma}\right]d\Phi(x) - 1$$

Proof: The proof follows immediately upon noting that the lower bound given in (3.1.1) is non-decreasing in each δ_{ki} $(1 \leq i \leq k - 1)$.

§3.1.2 Some properties of the lower bound:

1. If we let $h \to \infty$ in the expression for the lower bound in (3.1.1) we obtain $\int_{-\infty}^{\infty} \prod_{i=1}^{k-1} \Phi\left(x + \frac{\delta_{ki} n^{1/2}}{\sigma}\right) d\Phi(x)$ which is the expression for $P_{\underline{\mu}}(CS|R_0)$ using sample size n per population. Thus Bechhofer's single-stage rule R_0 is a special case of the conservative rule $\overline{R}_1$.

2. The lower bound is increasing in h for $h \geq 0$, for $n_1 \geq 0$ and for any $\underline{\mu}\varepsilon\Omega$. Thus if (n_1, n_2, h) are chosen to guarantee the probability re-

quirement using the r.h.s. of (3.1.2), and if n_0 is the corresponding single-stage sample size, then by letting $h \to \infty$ in the r.h.s. of (3.1.2) we find that

$$n_1 + n_2 \geq n_0 \tag{3.1.3}$$

for rule $\overline{R}_1$.

3. At the EMC the value of the lower bound is $\int_{-\infty}^{\infty} \Phi^{k-1}(x + c)d\Phi(x) + 1/k - 1$ which tends to $1/k$ iff $c = hn_1^{1/2}/\sigma \to \infty$. Similarly for $\delta^* > 0$ fixed, if the r.h.s. of (3.1.2) is set equal to P^* and if $P^* \to 1/k$ then either $n_1 \to 0$, $c \to 0$ and $n_2 \to \infty$ or $n_1 \to 0$, $n_2 \to 0$ and $c \to \infty$. Therefore as $P^* \to 1/k$, the two-stage conservative rule $\overline{R}_1$ behaves as a single-stage rule R_0.

§3.1.3 <u>An Optimization problem</u>:

We shall consider only the continuous optimization problem in the context of the conservative rule $\overline{R}_1$. Let c, d_1, d_2 be as defined by (2.4.2). Then the continuous optimization problem for the U-minimax criterion becomes:

For specified P^* and given k find (c, d_1, d_2) so as to

$$\text{minimize } d_1^2 + d_2^2 \int_{-\infty}^{\infty} \left[\Phi^{k-1}(x + c) - \Phi^{k-1}(x - c)\right] d\Phi(x) \tag{3.1.4}$$

$$\text{subject to} \int_{-\infty}^{\infty} \Phi^{k-1}(x + d_1 + c)d\Phi(x) +$$

$$\int_{-\infty}^{\infty} \Phi^{k-1}[x + (d_1^2 + d_2^2)^{1/2}]d\Phi(x) - 1 \geq P^*$$

and $c, d_1, d_2 \geq 0$.

We shall denote by $(\hat{\bar{c}}, \hat{\bar{d}}_1, \hat{\bar{d}}_2)$ a solution to the above optimization problem. One may analogously state the optimization problem for the rule $\bar{R}_1$ using the R-minimax criterion. The definition of $RE_{\underline{\mu}}(\delta^*, P^*, k; \hat{\bar{R}}_1)$ can be obtained by replacing $(\hat{c}, \hat{d}_1, \hat{d}_2)$ by $(\hat{\bar{c}}, \hat{\bar{d}}_1, \hat{\bar{d}}_2)$ in (2.5.1). Since $\bar{R}_1$ is based on the conservative lower bound on the PCS of rule R_1 we have $RE_{EMC}(P^*,k; \hat{\bar{R}}_1) \geq RE_{EMC}(P^*,k; \hat{R}_1)$ for all $k \geq 2$ and $P^* \varepsilon (k^{-1}, 1)$. Moreover, since the single-stage rule R_0 is a special case of rule $\bar{R}_1$ we have $1 \geq RE_{EMC}(P^*,k; \hat{\bar{R}}_1) \geq RE_{\underline{\mu}}(\delta^*,P^*,k; \hat{\bar{R}}_1)$ for all $\underline{\mu} \varepsilon \Omega$, $\delta^* > 0$, $k \geq 2$ and $P^* \varepsilon (k^{-1}, 1)$. Thus rule $\bar{R}_1$ is uniformly (in $\underline{\mu}$) atleast as good as rule R_0. However as an implication of Property 3 in §3.1.2 we might expect that for the range of P*-values close to 1/k and perhaps even somewhat higher, rule $\hat{\bar{R}}_1$ will be equivalent to rule R_0. This is in fact borne out by the numerical results.

§3.2 Tables to implement rule $\hat{\bar{R}}_1$ ($k > 2$):

§3.2.1 Discussion of the tables:

We have computed the continuous optimum solution for the optimization problem (3.1.4) for k = 3, 4, 5, 10 and 25 and for selected

Table 3.2.1

Rule $\hat{\bar{R}}_1$ (U-minimax criterion)

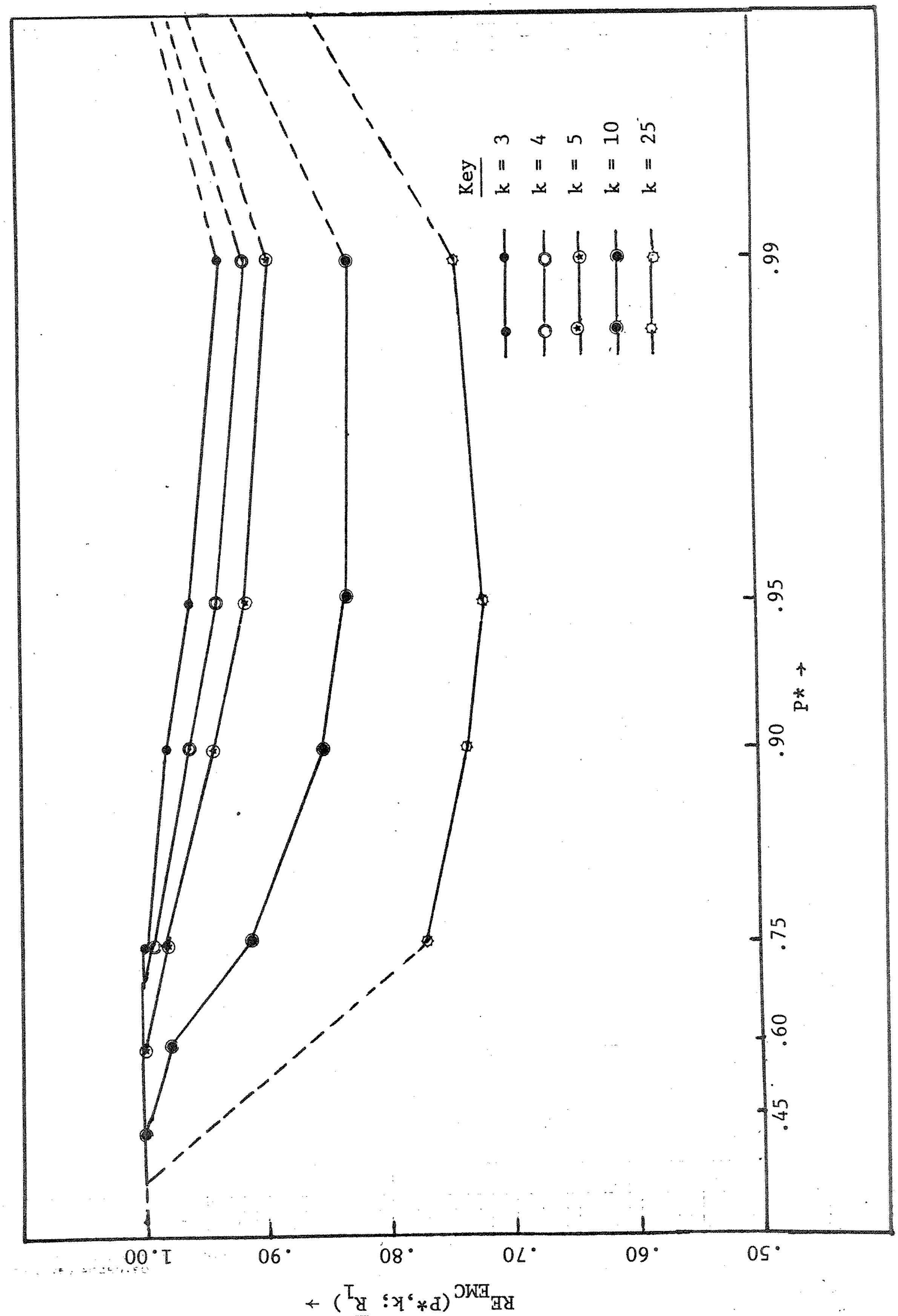
k	P*	$\hat{d}_1$	$\hat{d}_2$	$\hat{c}$	$\hat{p} = \frac{\hat{d}_1^2}{\hat{d}_1^2+\hat{d}_2^2}$	$\frac{\delta^{*2}}{k\sigma^2} E_{EMC}(N\|\hat{\bar{R}}_1)$	$RE_{EMC}(P^*,k;\hat{\bar{R}}_1)$
3	.99	2.93258	2.40832	1.24583	.59722	12.20554	.93280
	.95	2.08926	1.89737	1.63028	.54802	7.06860	.96242
	.90	1.66988	1.54913	2.18145	.53746	4.89325	.98381
	.75	1.049188	.97980	3.93352	.53416	2.05562	.99992
	.60 & below	<——— Use single-stage rule R_0 ———>					1.00000
4	.99	3.04319	2.58052	1.25962	.58172	13.17619	.91392
	.95	2.24000	2.11055	1.48060	.52973	7.98685	.93916
	.90	1.82625	1.78588	1.82453	.511173	5.78861	.96311
	.75	1.22030	1.17124	3.23650	.52050	2.82038	.99667
	.60 & below	<——— Use single-stage rule R_0 ———>					1.00000
5	.99	3.10350	2.73007	1.27117	.56375	13.75575	.89537
	.95	2.31844	2.26222	1.46036	.51227	8.55503	.91652
	.90	1.92094	1.97864	1.64030	.48521	6.35481	.94028
	.75	1.31909	1.33042	2.72804	.49432	3.37047	.98875
	.60	.96047	.91856	4.30382	.52229	1.76261	.99839
	.50 & below	<——— Use single-stage rule R_0 ———>					1.00000

Table 3.2.1 (contd.)

k	P*	$\hat{d}_1$	$\hat{d}_2$	$\hat{c}$	$\hat{p} = \frac{\hat{d}_1^2}{\hat{d}_1^2+\hat{d}_2^2}$	$\frac{\delta^{*2}}{k\sigma^2} E_{EMC}(N \mid \hat{R}_1)$	$RE_{EMC}(P^*,k;\hat{R}_1)$
10	.99	3.23639	3.16198	1.34527	.51163	14.96143	.83003
	.95	2.50946	2.77500	1.35292	.44988	9.77464	.83658
	.90	2.14663	2.53488	1.38302	.41763	7.57880	.85177
	.75	1.57123	1.97247	1.69804	.38821	4.69933	.91706
	.60	1.22475	1.34907	2.76484	.45181	3.07226	.98064
	.45	.95459	.92770	4.34334	.51429	1.76494	.99941
	.40 & below	<— Use single-stage rule R_0 —>					1.00000
25	.99	3.36341	3.65718	1.47829	.45823	15.84894	.74715
	.95	2.66460	3.32039	1.44012	.39173	10.70065	.73720
	.90	2.32701	3.13932	1.39717	.35461	8.49593	.73879
	.75	1.80000	2.85394	1.32342	.28458	5.59541	.76826
		<— No computations were made for lower P*-values. —>					

FIGURE 3.2.1

RE values of R_0 w.r.t. $\hat{\bar{R}}_1$ in the EMC plotted against P* for k = 3(1)5, 10 and 25.



ranges of P*-values. Our results are given in Table 3.2.1. The method of solution of the optimization problem, the details of computations, and the use of the table are the same as described in Appendix A1, and §2.6.1 respectively.

Figure 3.2.1 shows the variation of $RE_{EMC}(P^*,k;\hat{\hat{R}}_1)$ with P* for different values of k. For each fixed P*-value we see that $RE_{EMC}(P^*,k;\hat{\hat{R}}_1)$ decreases with k. This is to be expected since as k increases, the savings resulting from the screening aspect of our two-stage rule become more pronounced. In §3.2 we study the limiting behavior of $RE_{EMC}(P^*,k;\hat{R}_1)$ as $P^* \to 1$ for fixed k.

Figure 3.2.1 also indicates that for fixed k, we have that $RE_{EMC}(P^*,k;\hat{\hat{R}}_1)$ achieves a minimum at a certain intermediate value of P*. A more detailed picture of the behavior of $RE_{EMC}(P^*,k;\hat{\hat{R}}_1)$ in the range $P^* \geq .99$ will be presented in the next section for k = 2. It should be noted that for each k, $RE_{EMC}(P^*,k;\hat{\hat{R}}_1)$ becomes 1 as $\hat{\hat{R}}_1$ becomes equivalent to single-stage rule R_0 (i.e. $\hat{c} \to \infty$) for some value of P*, say $\overline{P}^*$, between 1/k and 1. From Figure 3.2.1 it would appear that $\overline{P}^*$ decreases as k increases.

In the next section we study for k = 2, the loss in relative efficiency due to using the conservative lower bound. To do so we will compare the <u>exact</u> continuous optimum (Table 2.6.1) and the <u>conservative</u> continuous optimum.

§3.2.2 A performance comparison of rules $\hat{R}_1$ and $\hat{\hat{R}}_1$ (k = 2):

The solution of the optimization problem (3.1.5) for k = 2 with selected values of P* is given in Table 3.2.2 below. These computations were made solely for the purpose of comparison with our results in

Table 3.2.2

Rule $\hat{\bar{R}}_1$ (U-minimax criterion) (k = 2)

P*	$\hat{\bar{d}}_1$	$\hat{\bar{d}}_2$	$\hat{\bar{c}}$	$\hat{\bar{p}} = \frac{\hat{\bar{d}}_1^2}{\hat{\bar{d}}_1^2 + \hat{\bar{d}}_2^2}$	$\frac{\delta^{*2}}{k\sigma^2} E_{EMC}(N \mid \hat{\bar{R}}_1)$	$RE_{EMC}(P^*,k;\ \hat{\bar{R}}_1)$	$RE_{EMC}(P^*,k;\ \hat{R}_1)$
.9999	4.47996	3.05941	1.12032	.68196	25.42155	.91903	.90047
.9995	3.90866	2.83627	1.13516	.65507	19.92597	.92015	.89442
.999	3.64005	2.72718	1.15306	.64048	17.60182	.92158	.89137
.99	2.62298	2.25388	1.30057	.57525	10.14257	.93704	.87850
.95	1.73928	1.66481	1.84202	.52187	5.26249	.97252	.86540
.90	1.30767	1.28841	2.59890	.50742	3.26026	.99253	.85799
.85 & below	←—— Use single-stage rule R_0 ——→					1.00000	.85319 & below

FIGURE 3.2.2

RE values of R_0 w.r.t. $\hat{R}_1$ and $\hat{\bar{R}}_1$ in the EMC plotted against P* for k = 2.

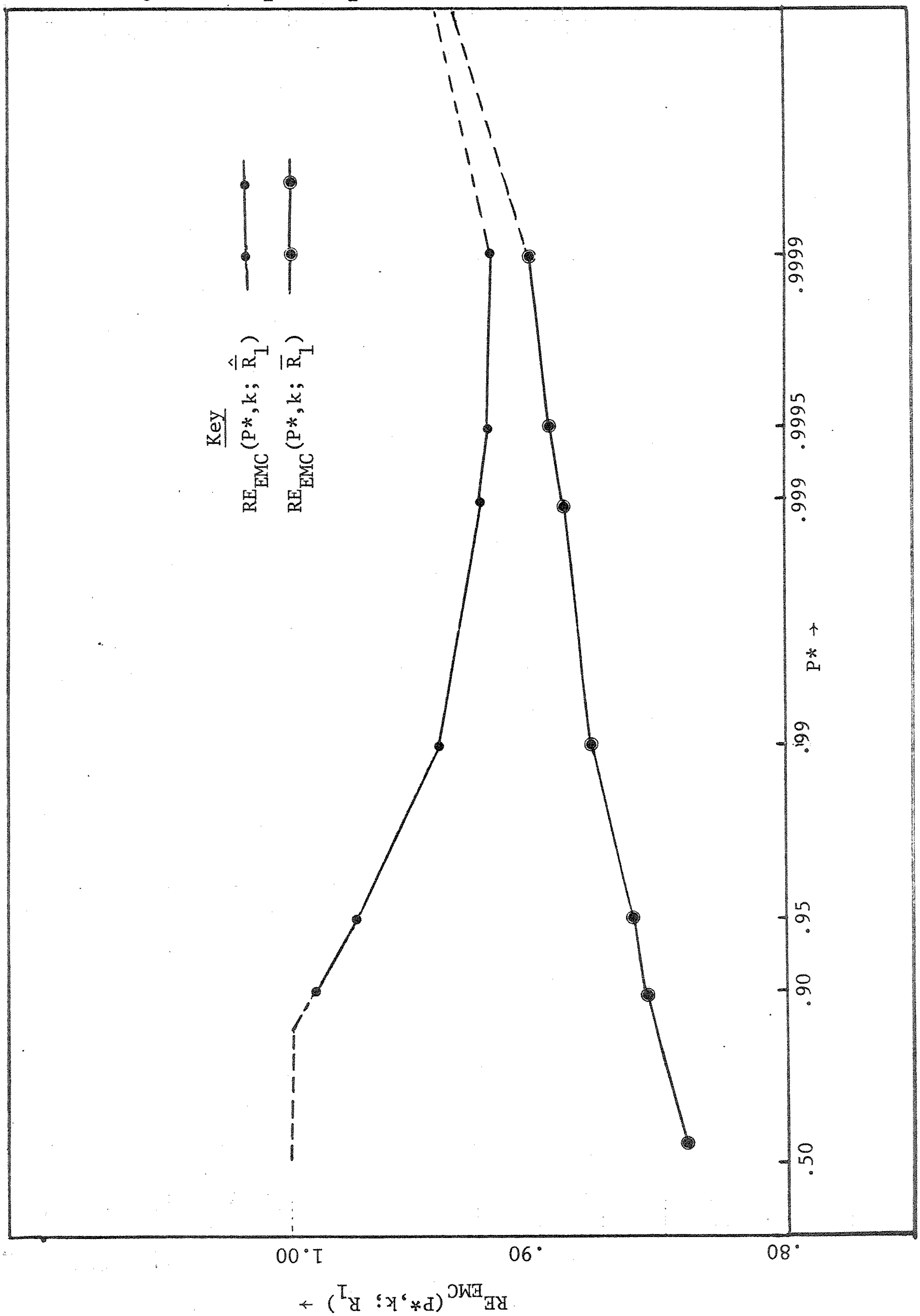


FIGURE 3.2.3

RE values of R_0 w.r.t. $\tilde{R}_1$ in the EMC and the LFC plotted against P* for k = 2.

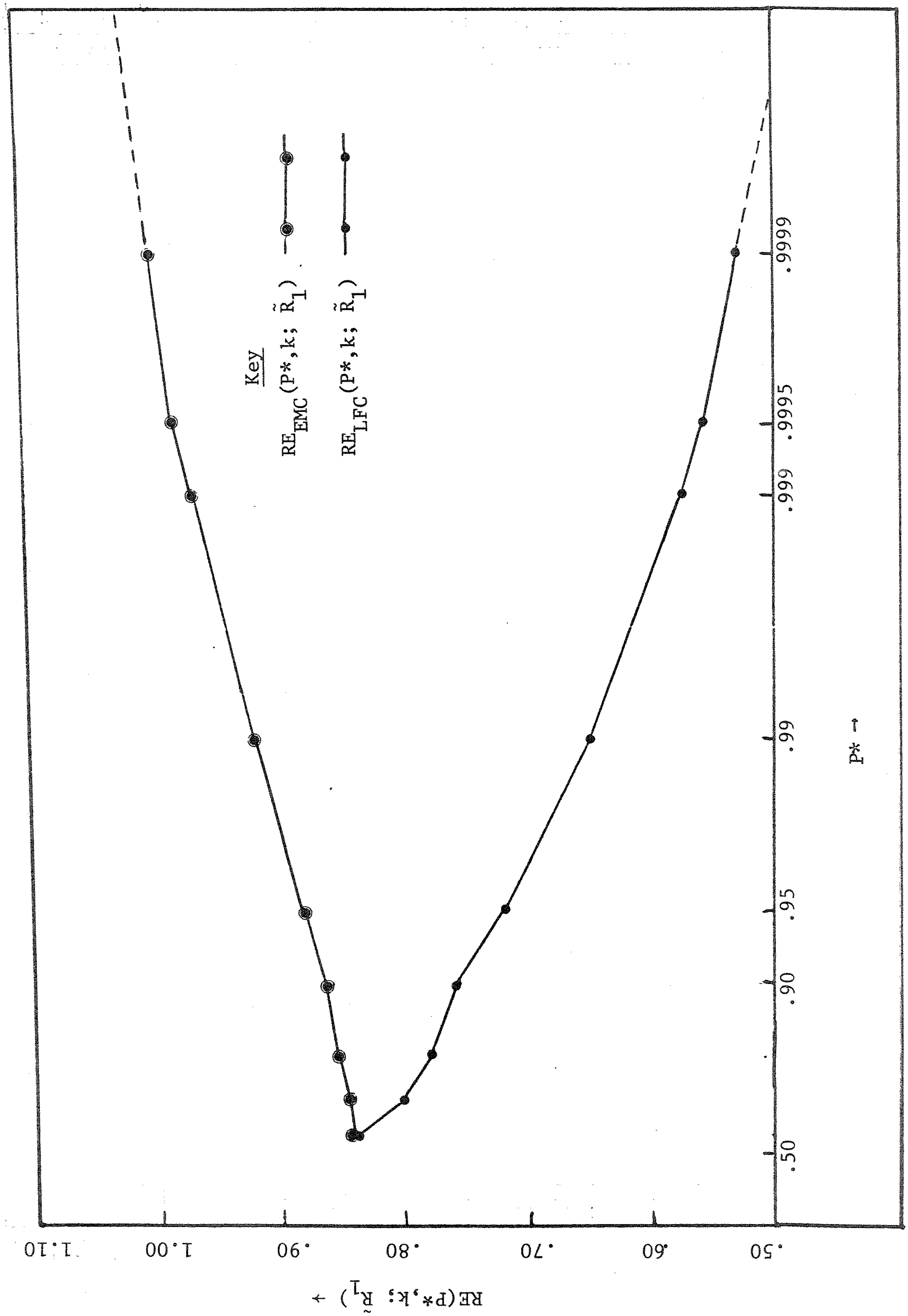


Table 2.6.1 in order to study the loss in relative efficiency. Since Table 2.6.1 is based on the exact expression for $P_{LFC}(CS|R_1)$, it is the one to be used in practice. The RE_{EMC} - values for rules $\hat{R}_1$ and $\hat{\hat{R}}_1$ for k = 2 are plotted against P* in Figure 3.2.2. It is seen that $RE_{EMC} \to 1$ for both $\hat{R}_1$ and $\hat{\hat{R}}_1$ as $P^* \to 1$. This result will be proved analytically in §3.3. The loss in efficiency due to using conservative rule $\hat{\hat{R}}_1$ instead of the exact rule $\hat{R}_1$ is significant for low P*-values; the difference in relative efficiencies between rules $\hat{R}_1$ and $\hat{\hat{R}}_1$ decreases rapidly as P* increases. Thus significant gains can be achieved in terms of the expected total sample size if the exact rule $\hat{R}_1$ is used for small to moderate values of P* and large values of k. Thus a great incentive exists for attempting the proof of the conjectured LFC for rule R_1.

In Figure 3.2.3 we have plotted RE_{EMC} - and RE_{LFC} - values for rule $\tilde{R}_1$ for k = 2. We note that $RE_{EMC}(P^*,k;\ \tilde{R}_1)$ increases with P* and is greater than 1 for P* = .9999. $RE_{LFC}(P^*,k;\ \tilde{R}_1)$ decreases with increasing P*. We shall prove in §3.3 that $RE_{LFC}(P^*,k;\ \tilde{R}_1) \to 1/4$ as $P^* \to 1$ for all $k \geq 2$.

§3.2.3 A brief comparison of rule $\hat{\hat{R}}_1$ with two sequential rules:

Ramberg [1966] made a MC sampling study of the sequential rules BKS and P both for the normal means problem. We shall compare the performance of our conservative rule $\hat{\hat{R}}_1$ with that of BKS_A and P_A rules by comparing the relative efficiencies in the EMC of rule R_0 relative to rules BKS_A and P_A. Here the subscript A means that the original sequential rules are adjusted to eliminate the overprotection in PCS.

We shall use Tables 16 and 19 from Ramberg [1966]. It might be mentioned that the comparisons with the Paulson rule given by Ramberg do not depict the Paulson rule when it is operating in its most efficient way. Fabian [1973] has recently shown that to minimize the expected total sample size in the EMC as $P^* \to 1$, the Paulson rule should be used with $\lambda = \delta^*/2$; Ramberg's results are for $\lambda = \delta^*/4$. Fabian, in the same paper, also gives an improved lower bound on the PCS of Paulson rule. Since Ramberg makes an <u>empirical</u> adjustment for the overprotection in PCS, we may assume that the results would not change much if one were to use Fabian's improved bound. The results are given in the following table.

<u>Table 3.2.3</u>

<u>A comparison of rule $\bar{R}_1$ with sequential rules</u>

(k = 10)

P^*	$RE_{EMC}(P^*,P_A)$ $\lambda=\delta^*/4$ MC estimate	$RE_{EMC}(P^*,BKS_A)$ MC estimate	$RE_{EMC}(P^*,\hat{\bar{R}}_1)$
.75	.56250	1.12500	.91706
.90	.55631	1.22072	.85177
.95	.56815	1.38699	.83658
.99	.61622	1.80444	.83003

We see that Paulson's adjusted rule (with $\lambda = \delta^*/4$) minimizes the expected total sample size in the EMC among the three rules for all the values of P* under consideration. It is possible that even the <u>exact</u> $\hat{R}_1$ is uniformly (in P*) dominated by the Paulson rule. Thus the gains achievable by a fully sequential rule allowing for elimination at each stage, appear to be significantly more than the gains achievable by our two-stage rule which allows elimination only at one stage. However, we have to trade off this advantage of the fully sequential rule against the administrative difficulties associated with its implementation.

Rule $\hat{\hat{R}}_1$, in turn, dominates rule BKS_A for the P*-values under consideration; for small P*-values since rule $\hat{\hat{R}}_1$ becomes equivalent to single-stage rule R_0, it will be in general dominated by BKS_A atleast for small values of k. We must, however, point out that to obtain a true picture of the relative performances of the sequential and the two-stage rules, one must use the <u>exact</u> rule $\hat{R}_1$ in any future comparison.

§3.3 <u>Asymptotic ($P^* \to 1$) behavior of rules</u> $\hat{R}_1$ <u>and</u> $\tilde{R}_1$

In what follows we shall keep δ^*, σ and k fixed throughout and only allow P* to vary. We shall assume that the conjectured LFC for rule R_1 is correct. We shall also assume that the solutions $(\hat{c},\hat{d}_1,\hat{d}_2)$ and $(\tilde{c},\tilde{d}_1,\tilde{d}_2)$ to the optimization problems (2.4.4) and (2.4.5) approach limiting values which may be finite or $+\infty$ as $P^* \to 1$.

<u>Definition 3.3.1</u>: For any fixed P*, let $\hat{c} = \hat{c}(P^*)$, $\hat{d}_1 = \hat{d}_1(P^*)$ and $\hat{d}_2 = \hat{d}_2(P^*)$ denote a solution to the optimization problem (2.4.4). Let $d_0 = d_0(P^*)$ denote the solution to equation (1.1.2). Then the asymptotic relative efficiency as $P^* \to 1$ of rule R_0 relative to rule $\hat{R}_1$

at any parameter configuration $\underline{\mu}\varepsilon\Omega$ is defined by

(3.3.1) $\quad ARE_{\underline{\mu}}(\delta^*,k;\ \hat{R}_1)$

$$= \lim_{P^*\to 1} RE_{\underline{\mu}}(\delta^*,P^*,k;\ \hat{R}_1)$$

$$= \lim_{P^*\to 1} \frac{1}{kd_0^2}\left[k\hat{d}_1^2 + \hat{d}_2^2 \sum_{i=1}^{k} \int_{-\infty}^{\infty} \left\{ \prod_{\substack{j=1\\ j\neq i}}^{k} \Phi\left[x + \frac{\delta_{ij}\hat{d}_1}{\delta^*} + \hat{c}\right] - \prod_{\substack{j=1\\ j\neq i}}^{k} \Phi\left[x + \frac{\delta_{ij}\hat{d}_1}{\delta^*} - \hat{c}\right]\right\} d\Phi(x)\right].$$

In particular recalling Remark 2.5.1 we have,

(3.3.2) $\quad ARE_{EMC}(k;\ R_1) =$

$$\lim_{P^*\to 1}\left[\frac{\hat{d}_1^2 + \hat{d}_2^2 \int_{-\infty}^{\infty} \left\{\Phi^{k-1}(x+\hat{c}) - \Phi^{k-1}(x-\hat{c})\right\} d\Phi(x)}{d_0^2}\right]$$

We can similarly define $ARE_{\underline{\mu}}(\delta^*,k;\ \tilde{R}_1)$. Here we give only the expression for $ARE_{LFC}(k;\ \tilde{R}_1)$.

(3.3.3) $$ARE_{LFC}(k;\ \tilde{R}_1) = \lim_{P^*\to 1} \frac{1}{kd_0^2}\left[k\tilde{d}_1^2 + \tilde{d}_2^2\left\{\int_{-\infty}^{\infty}\left[\Phi^{k-1}(x+\tilde{d}_1+\tilde{c}) - \right.\right.\right.$$

$$\left.\Phi^{k-1}(x+\tilde{d}_1-\tilde{c})\right] d\Phi(x) + (k-1)\int_{-\infty}^{\infty}\left[\Phi^{k-2}(x+\tilde{c})\Phi(x-\tilde{d}_1+\tilde{c}) - \right.$$

$$\Phi^{k-2}(x - \tilde{c})\Phi(x - \tilde{d}_1 - \tilde{c})\Big] d\Phi(x)\Big\}\Bigg] .$$

We now state two basic lemmas which are useful in proving the main theorems of this section.

Lemma 3.3.1: As $P^* \to 1$ we have $d_1 + c \to \infty$ and $d_1^2 + d_2^2 \to \infty$.

Proof: We have using (3.1.1)

$$P^* = P_{LFC}(CS|R_1) \leq \int_{\infty}^{\infty} \Phi^{k-1}(x + d_1 + c)d\Phi(x) \leq 1.$$

Therefore $P^* \to 1$ implies $d_1 + c \to \infty$. Next as $P^* \to 1$, $n = n_1 + n_2 \to \infty$. Hence $d_1^2 + d_2^2 \to \infty$.

We shall need the following special case of a lemma due to Bechhofer, Kiefer and Sobel [1968]. In the following we use the notation $a \sim b$ to mean that $a/b \to 1$ in the limit and by log we denote the natural logarithm.

Lemma 3.3.2 (Bechhofer, Kiefer and Sobel):

Let $H(u) = 1 - \int_{-\infty}^{\infty} \Phi^{k-1}(x + u)d\Phi(x)$. Then as $u \to \infty$

$$H(u) \sim \frac{(k-1)e^{-u^2/4}}{u\sqrt{\pi}} .$$

Corollary: As $P^* \to 1$, the solution d_0 to equation (1.1.2) is given by

(3.3.4) $$d_0^2 = 4\log(\frac{k-1}{1-P^*}) - 2\log\log(\frac{k-1}{1-P^*}) - 2\log 4\pi + o(1)$$

or $d_0^2 \sim 4\log(1-P^*)^{-1}$.

Now we state and prove the main result concerning the ARE of rule $\hat{R}_1$.

Theorem 3.3.1: For the U-minimax rule $\hat{R}_1$ we have,

$$\text{ARE}_{\text{EMC}}(k;\hat{R}_1) = 1 \quad \text{for all } k \geq 2.$$

Proof: From (3.1.1) we have

$$1 - \int_{-\infty}^{\infty} \Phi^{k-1}(x + \hat{d}_1 + \hat{c})d\Phi(x) \leq 1 - P_{LFC}(CS|\hat{R}_1)$$

$$\leq \left\{1 - \int_{-\infty}^{\infty} \Phi^{k-1}(x + \hat{d}_1 + \hat{c})d\Phi(x)\right\} + \left\{1 - \int_{-\infty}^{\infty} \Phi^{k-1}\left[x + (\hat{d}_1^2 + \hat{d}_2^2)^{1/2}\right]d\Phi(x)\right\}.$$

Therefore,

(3.3.5) $$1-P_{LFC}(CS|\hat{R}_1) = \left\{1 - \int_{-\infty}^{\infty} \Phi^{k-1}(x + \hat{d}_1 + \hat{c})d\Phi(x)\right\} +$$

$$\gamma\left\{1 - \int_{-\infty}^{\infty} \Phi^{k-1}\left[x + (\hat{d}_1^2 + \hat{d}_2^2)^{1/2}\right]d\Phi(x)\right\},$$

where $0 \leq \gamma \leq 1$ and $\gamma = \gamma(k,P^*,\hat{R}_1)$. We set $P_{LFC}(CS|\hat{R}_1) = P^*$ and use Lemmas 3.3.1 and 3.3.2 to obtain as $P^* \to 1$,

$$(3.3.6) \qquad 1 - P^* \sim \frac{(k-1)}{\sqrt{\pi}} \left\{ \frac{e^{-(\hat{d}_1 + \hat{c})^2/4}}{(\hat{d}_1 + \hat{c})} + \gamma \frac{e^{-(\hat{d}_1^2 + \hat{d}_2^2)/4}}{(\hat{d}_1^2 + \hat{d}_2^2)^{1/2}} \right\}.$$

Now we consider the following two possibilities for the limiting values of $\hat{c}$.

<u>Case (i)</u> $\lim\limits_{P^*\to 1} \hat{c} = \infty$:

In this case $\int_{-\infty}^{\infty} \{\Phi^{k-1}(x + \hat{c}) - \Phi^{k-1}(x - \hat{c})\}d\Phi(x) \to 1$ as $P^* \to 1$. Therefore we obtain,

$$(3.3.7) \quad ARE_{EMC}(k; \hat{R}_1) = \lim_{P^*\to 1} \frac{\hat{d}_1^2 + \hat{d}_2^2 \int_{-\infty}^{\infty} \{\Phi^{k-1}(x + \hat{c}) - \Phi^{k-1}(x-\hat{c})\} d\Phi(x)}{d_0^2}$$

$$= \lim_{P^*\to 1} \frac{\hat{d}_1^2 + \hat{d}_2^2}{d_0^2}$$

$$\geq 1.$$

The last step follows since $\hat{d}_1^2 + \hat{d}_2^2 \geq d_0^2$ for all P^* from (2.2.15).

<u>Case (ii)</u> $\lim\limits_{P^*\to 1} \hat{c} < \infty$:

Since $\hat{d}_1 + \hat{c} \to \infty$ as $P^* \to 1$ from Lemma 3.3.1, in this case we have

$\hat{d}_1 \to \infty$, $(\hat{c}/\hat{d}_1) \to 0$ and

$$1 - P^* \sim \frac{(k-1)}{\sqrt{\pi}} \left\{ \frac{e^{-\hat{d}_1^2/4}}{\hat{d}_1} + \gamma \frac{e^{-(\hat{d}_1^2 + \hat{d}_2^2)/4}}{(\hat{d}_1^2 + \hat{d}_2^2)^{1/2}} \right\} .$$

$$\sim \frac{(k-1)A}{\sqrt{\pi}} \frac{e^{-\hat{d}_1^2/4}}{\hat{d}_1} ,$$

where $1 \leq A < \infty$. Hence $\hat{d}_1^2 \sim 4 \log(1 - P^*)^{-1} \sim d_0^2$ using Lemma 3.3.2. Thus we obtain

$$\text{(3.3.8)} \quad \text{ARE}_{\text{EMC}}(k; \hat{R}_1) = \lim_{P^* \to 1} \frac{\hat{d}_1^2 + \hat{d}_2^2 \int_{-\infty}^{\infty} \{\Phi^{k-1}(x + \hat{c}) - \Phi^{k-1}(x-\hat{c})\} d\Phi(x)}{d_0^2}$$

$$\geq 1.$$

From (3.3.7) and (3.3.8) we have that $\text{ARE}_{\text{EMC}}(k; \hat{R}_1) \geq 1$ for all $k \geq 2$. But $\text{RE}_{\text{EMC}}(k,P^*; \hat{R}_1) \leq 1$ for all P^* implies that $\text{ARE}_{\text{EMC}}(k; \hat{R}_1) \leq 1$. Therefore $\text{ARE}_{\text{EMC}}(k; \hat{R}_1) = 1$ for all $k \geq 2$. Hence the theorem is proved.

This theorem tells us that the limiting relative efficiency (as $P^* \to 1$) of rule R_0 w.r.t. any two-stage rule R_1 in the EMC must be atleast as large as 1. In particular,

$$\text{(3.3.9)} \qquad \text{ARE}_{\text{EMC}}(k; \tilde{R}_1) \geq 1.$$

In the next theorem we study $\text{ARE}_{\text{LFC}}(k; \tilde{R}_1)$.

Theorem 3.3.2: For the R-minimax rule $\tilde{R}_1$ we have,

$$\text{(3.3.10)} \qquad ARE_{LFC}(k;\ \tilde{R}_1) = \frac{1}{4} \quad \text{for all} \quad k \geq 2.$$

Also as $P^* \to 1$, $\tilde{d}_1 - \tilde{c} = o(\tilde{d}_1) \to \infty$ and $\lim_{P^*\to 1} (\tilde{d}_2/\tilde{d}_1)^2 \geq 3$.

Proof: As $P^* \to 1$ we have,

$$\text{(3.3.11)} \qquad 1 - P^* \sim \frac{(k-1)}{\sqrt{\pi}} \left\{ \frac{e^{-(\tilde{d}_1 + \tilde{c})^2/4}}{(\tilde{d}_1 + \tilde{c})} + \gamma \frac{e^{-(\tilde{d}_1^2 + \tilde{d}_2^2)/4}}{(\tilde{d}_1^2 + \tilde{d}_2^2)^{1/2}} \right\},$$

where $0 \leq \gamma \leq 1$ and $\gamma = \gamma(k,P^*;\ \tilde{R}_1)$. The following proof proceeds by considering all the possible limiting values that $\tilde{d}_1$ and $\tilde{c}$ can take so that $\tilde{d}_1 + \tilde{c} \to \infty$ and evaluating $ARE_{LFC}(k;\ \tilde{R}_1)$ in each case. The actual limiting behavior of $\tilde{d}_1$ and $\tilde{c}$ would be dictated by the requirement that $ARE_{LFC}(k;\ \tilde{R}_1)$ be minimized.

Case (i) $\lim_{P^*\to 1} (\tilde{d}_1 - \tilde{c}) < \infty$:

In this case from (3.3.11) we obtain as $P^* \to 1$

$$\text{(3.3.12)} \qquad 1 - P^* \sim \frac{(k-1)}{\sqrt{\pi}} \left\{ \frac{e^{-\tilde{d}_1^2}}{2\tilde{d}_1} + \gamma \frac{e^{-(\tilde{d}_1^2 + \tilde{d}_2^2)/4}}{(\tilde{d}_1^2 + \tilde{d}_2^2)^{1/2}} \right\}$$

Subcase (ia) $\lim_{P^*\to 1} (\tilde{d}_2/\tilde{d}_1)^2 < 3$:

For this subcase, we have $\tilde{d}_1^2 > (\tilde{d}_1^2 + \tilde{d}_2^2)/4$ for P* arbitrarily close to

1. Therefore we can write

$$(3.3.13) \qquad 1 - P^* \sim \frac{(k-1)A}{\sqrt{\pi}} \left\{ \frac{e^{-(\tilde{d}_1^2 + \tilde{d}_2^2)/4}}{(\tilde{d}_1^2 + \tilde{d}_2^2)^{1/2}} \right\},$$

where $0 < A < \infty$, unless $\gamma \to 0$ at a rate rapid enough so that we have,

$$(3.3.14) \qquad 1 - P^* \sim \frac{(k-1)B}{\sqrt{\pi}} \cdot \frac{e^{-\tilde{d}_1^2}}{2\tilde{d}_1},$$

where $1 \leq B < \infty$. Suppose (3.3.14) holds. Then $\tilde{d}_1^2 \sim \log(1 - P^*)^{-1} \sim d_0^2/4$ using Lemma (3.3.2) and $\lim\limits_{P^* \to 1} (\tilde{d}_2/d_0)^2 < \frac{3}{4}$. Therefore

$$ARE_{EMC}(k; \tilde{R}_1) = \lim_{P^* \to 1} \left\{ \left(\frac{\tilde{d}_1}{d_0}\right)^2 + \frac{\tilde{d}_2^2}{kd_0^2} \left[\int_{-\infty}^{\infty} \{\Phi^{k-1}(x + \tilde{d}_1 + \tilde{c}) - \Phi^{k-1}(x + \tilde{d}_1 - \tilde{c})\} d\Phi(x) + (k-1) \int_{-\infty}^{\infty} \{\Phi^{k-2}(x + \tilde{c})\Phi(x - \tilde{d}_1 + \tilde{c}) - \Phi^{k-2}(x - \tilde{c})\Phi(x - \tilde{d}_1 - \tilde{c})\} d\Phi(x) \right] \right\} < \frac{1}{4} + \frac{3}{4} = 1.$$

This contradicts (3.3.9). Hence (3.3.13) holds and we have $(\tilde{d}_1^2 + \tilde{d}_2^2) \sim 4 \log(1 - P^*)^{-1} \sim d_0^2$ using Lemma 3.3.2. Now since $\lim\limits_{P^* \to 1} (\tilde{d}_2/\tilde{d}_1)^2 < 3$, it follows that $\lim\limits_{P^* \to 1} (\tilde{d}_1/\tilde{d}_0)^2 > \frac{1}{4}$ and consequently

$$ARE_{LFC}(k; \tilde{R}_1) > \frac{1}{4}.$$

<u>Subcase (ib)</u> $\lim_{P^*\to 1} (\tilde{d}_2/\tilde{d}_1)^2 \geq 3$:

In this case, as $P^* \to 1$, we have from (3.3.12)

$$1 - P^* \sim \frac{(k-1)A}{\sqrt{\pi}} \cdot \frac{e^{-\tilde{d}_1^2}}{2\tilde{d}_1}, \tag{3.3.15}$$

where $1 \leq A < \infty$. Therefore $\tilde{d}_1^2 \sim \log(1 - P^*)^{-1} \sim d_0^2/4$ and $\lim_{P^*\to 1} (\tilde{d}_1/d_0)^2 \geq \frac{3}{4}$. Now $\lim_{P^*\to 1} (\tilde{d}_1 - \tilde{c}) < \infty$ implies that $\tilde{d}_1 \to \infty$, $\tilde{c} \to \infty$ and,

$$0 < \lim_{P^*\to 1} \left[\int_{-\infty}^{\infty} \{\Phi^{k-1}(x + \tilde{d}_1 + \tilde{c}) - \Phi^{k-1}(x + \tilde{d}_1 - \tilde{c})\} d\Phi(x) + (k-1) \int_{-\infty}^{\infty} \{\Phi^{k-2}(x + \tilde{c})\Phi(x-\tilde{d}_1 + \tilde{c}) - \Phi^{k-2}(x-\tilde{c})\Phi(x - \tilde{d}_1 - \tilde{c})\} d\Phi(x) \right] < k. \tag{3.3.16}$$

Therefore we have,

$$\begin{aligned} ARE_{LFC}(k; \tilde{R}_1) = \lim_{P^*\to 1} \Bigg\{ \frac{\tilde{d}_1^2}{\tilde{d}_0^2} + \frac{\tilde{d}_1^2}{k\tilde{d}_0^2} \Bigg[\int_{-\infty}^{\infty} \{\Phi^{k-1}(x + \tilde{d}_1 + \tilde{c}) - \\ \Phi^{k-1}(x + \tilde{d}_1 - \tilde{c})\} d\Phi(x) + (k-1) \int_{-\infty}^{\infty} \{\Phi^{k-2}(x + \tilde{c})\Phi(x - \tilde{d}_1 + \tilde{c}) - \\ \Phi^{k-2}(x - \tilde{c})\Phi(x - \tilde{d}_1 - \tilde{c})\} d\Phi(x) \Bigg] \Bigg\} \\ > \frac{1}{4}. \end{aligned} \tag{3.3.17}$$

<u>Case (ii)</u>: $\lim_{P^*\to 1} (\tilde{d}_1 - \tilde{c}) = \infty$, $\tilde{d}_1 \to \infty$, $\tilde{c} \to \infty$, $(\tilde{d}_1 - \tilde{c})/\tilde{d}_1 \to 0$:

In this case also (3.3.12) holds and we have subcases (a) and (b) as in Case (i). The analysis in each subcase is the same as before except in Subcase (b) instead of (3.3.16) we have

$$(3.3.18)\quad \lim_{P^*\to 1} \left[\int_{-\infty}^{\infty} \{\Phi^{k-1}(x + \tilde{d}_1 + \tilde{c}) - \Phi^{k-1}(x + \tilde{d}_1 - \tilde{c})\} d\Phi(x) + (k-1) \int_{-\infty}^{\infty} \{\Phi^{k-2}(x + \tilde{c})\Phi(x - \tilde{d}_1 + \tilde{c}) - \Phi^{k-2}(x - \tilde{c})\Phi(x - \tilde{d}_1 - \tilde{c})\} d\Phi(x) \right] = 0.$$

Further $\lim_{P^*\to 1} (\tilde{d}_1/d_0)^2 = \frac{1}{4}$ and $\lim_{P^*\to 1} (\tilde{d}_2/d_0)^2 \geq \frac{3}{4}$. In view of (3.3.18) we may assume that

$$\lim_{P^*\to 1} \frac{\tilde{d}_2^2}{kd_0^2} \left[\int_{-\infty}^{\infty} \{\Phi^{k-1}(x + \tilde{d}_1 + \tilde{c}) - \Phi^{k-1}(x + \tilde{d}_1 - \tilde{c})\} d\Phi(x) + (k-1) \int_{-\infty}^{\infty} \{\Phi^{k-2}(x + \tilde{c})\Phi(x - \tilde{d}_1 + \tilde{c}) - \Phi^{k-2}(x - \tilde{c})\Phi(x - \tilde{d}_1 - \tilde{c})\} d\Phi(x) \right] = 0.$$

Therefore,

$$(3.3.19)\qquad ARE_{LFC}(k; \tilde{R}_1) = \lim_{P^*\to 1} \frac{\tilde{d}_1^2}{d_0^2} + 0 = \frac{1}{4}.$$

<u>Case (iii)</u> $\lim_{P^* \to 1} (\tilde{d}_1 - \tilde{c}) = \infty$ and $(\tilde{d}_1 - \tilde{c})/\tilde{d}_1$ does not approach 0 as $P^* \to 1$:

Therefore $\tilde{c}/\tilde{d}_1 \to B$ where $0 \leq B < 1$. Also denote the limiting value of $(\tilde{d}_2/\tilde{d}_1)^2$ by D. Then as $P^* \to 1$ we have

$$(3.3.20) \quad 1 - P^* \sim \frac{(k-1)}{\sqrt{\pi}} \left\{ \frac{e^{-\tilde{d}_1^2(1 + B)^2/4}}{\tilde{d}_1(1 + B)} + \gamma \frac{e^{-\tilde{d}_1^2(1 + D)/4}}{\tilde{d}_1(1 + D)^{1/2}} \right\} .$$

<u>Subcase (iiia)</u> $(1 + B)^2 < (1 + D)$:

Then we have

$$1 - P^* \sim \frac{(k-1)A}{\sqrt{\pi}} \frac{e^{-\tilde{d}_1^2(1 + B)^2}}{(1 + B)} ,$$

where $1 \leq A < \infty$. Using Lemma 3.3.2 we have,

$\tilde{d}_1^2 \sim \frac{4}{(1 + B)^2} \log(1 - P^*)^{-1} \sim d_0^2/(1 + B)^2$. Therefore we obtain,

$$(3.3.21) \qquad ARE_{LFC}(k; \tilde{R}_1) \geq \frac{1}{(1 + B)^2} > \frac{1}{4} .$$

<u>Subcase (iiib)</u> $(1 + B)^2 \geq 1 + D$:

From (3.3.20) we have,

$$(3.3.22) \qquad 1 - P^* \sim \frac{(k-1)A}{\sqrt{\pi}} \frac{e^{-\tilde{d}_1^2(1 + D)/4}}{\tilde{d}_1(1 + D)} ,$$

where $0 < A < \infty$. Using Lemma 3.3.2 we have,

$\tilde{d}_1^2 \sim \dfrac{4}{(1+D)} \log(1 - P^*)^{-1} \sim d_0^2/(1 + D)$. But $D \leq (1 + B)^2 - 1 < 3$.

Hence we obtain,

$$\text{(3.3.23)} \qquad \text{ARE}_{\text{LFC}}(k; \tilde{R}_1) \geq \lim_{P^*\to 1} \frac{\tilde{d}_1^2}{d_0^2} = \frac{1}{1 + D} > \frac{1}{4}.$$

<u>Case (iv)</u> $\lim_{P^*\to 1} (\tilde{d}_1 - \tilde{c}) = -\infty$:

In this case we have,

$$\lim_{P^*\to 1}\Bigg[\int_{-\infty}^{\infty} \{\Phi^{k-1}(x + \tilde{d}_1 + \tilde{c}) - \Phi^{k-1}(x + \tilde{d}_1 - \tilde{c})\}d\Phi(x) + (k-1)\int_{-\infty}^{\infty}$$

$$\{\Phi^{k-2}(x + \tilde{c})\Phi(x - \tilde{d}_1 + \tilde{c}) - \Phi^{k-2}(x - \tilde{c})\Phi(x - \tilde{d}_1 - \tilde{c})\}d\Phi(x)\Bigg] = k.$$

Hence we obtain,

$$\text{(3.3.24)} \quad \text{ARE}_{\text{LFC}}(k; \tilde{R}_1) = \lim_{P^*\to 1}\Bigg\{\frac{\tilde{d}_1^2}{d_0^2} + \frac{\tilde{d}_2^2}{kd_0^2}\Bigg[\int_{-\infty}^{\infty} \{\Phi^{k-1}(x + \tilde{d}_1 + \tilde{c}) -$$

$$\Phi^{k-1}(x + \tilde{d}_1 - \tilde{c})\}d\Phi(x) + (k-1)\int_{-\infty}^{\infty} \{\Phi^{k-2}(x + \tilde{c})\Phi(x - \tilde{d}_1 + \tilde{c}) -$$

$$\Phi^{k-2}(x - \tilde{c})\Phi(x - \tilde{d}_1 - \tilde{c})\}d\Phi(x)\Bigg]\Bigg\}$$

$$= \lim_{P^* \to 1} \frac{\tilde{d}_1^2 + \tilde{d}_2^2}{d_0^2}$$

$$\geq 1.$$

From Cases (i), (ii), (iii) and (iv) we find that subcase (iib) yields minimum $ARE_{LFC}(k;\ \tilde{R}_1) = \frac{1}{4}$. Hence the theorem is proved.

From Bechhofer, Kiefer and Sobel [1968] we know that for their sequential rule BKS,

(3.3.25) $$ARE_{LFC}(k;\ BKS/R_0) = \frac{1}{4}.$$

Thus the ratio of the expected total sample sizes in the LFC required by rules BKS and $\tilde{R}_1$ to guarantee the same probability requirement (1.1.1) goes to 1 as $P^* \to 1$. In particular, Wald's sequential probability ratio test (WSPRT) to test H_0: $\mu_1 - \mu_2 \geq \delta^*$ against H_1: $\mu_1 - \mu_2 \leq -\delta^*$ is a special case of BKS for $k = 2$. WSPRT is known to have the optimum property of requiring the least expected total number of observations at the parameter configuration $\mu_{[2]} - \mu_{[1]} = \delta^*$ among all tests with the same probabilities of Type I and Type II errors. If the error probabilities are set equal to $1 - P^*$ and if $P^* \to 1$ then the ARE in LFC of the best single-stage rule (which is R_0) w.r.t. WSPRT is known to be $\frac{1}{4}$. Thus for $k = 2$, we find that as $P^* \to 1$, the two-stage rule $\tilde{R}_1$ performs as well as WSPRT which is the optimum rule for the given testing problem. This is a somewhat surprising but a very important result.

Another quantity of interest might be $ARE_{EMC}(k;\ \tilde{R}_1)$. We know that $ARE_{EMC}(k;\ \tilde{R}_1) \geq 1$. However we have not been able to obtain an exact expression for $ARE_{EMC}(k;\ \tilde{R}_1)$. We conjecture that $1 < ARE_{EMC}(k;\ \tilde{R}_1) < \infty$

§3.4 A come-back modification of rule R_1 and some MC results:

§3.4.1 A come-back type rule R_2:

In §2.1.1 we noted that the permanent-elimination feature of rule R_1 might result in outcomes in which all of the populations entering the second stage would yield cumulative sample means that are smaller than the first stage sample mean of a population which was eliminated after the first stage. Although the first stage sample mean is based on a smaller number of observations, it would appear that for any parameter configuration $\underline{\mu}\varepsilon\Omega$ one can increase the PCS if rule R_1 is modified to allow the selection of the population associated with $\max_{i \notin I} \overline{X}_i^{(1)}$ as best in case of the above type of outcome. A general family of such rules will be called come-back type rules. We now propose a rule R_2 in this family.

Rule R_2: The first two steps in rule R_2 are the same as in the case of rule R_1. Thus if only a single population enters the second stage then we terminate sampling and assert that, that population is best. However, if more than one population enters the second stage then we assert that the population associated with $\max[\overline{X}_i,\ i\varepsilon I;\ \overline{X}_i^{(1)},\ i\notin I]$ is best.

For $k = 2$, rules R_1 and R_2 are identical. Further, if both the rules use the same values of n_1, n_2 and h then $E_{\underline{\mu}}(N|R_1) = E_{\underline{\mu}}(N|R_2)$ for any $\underline{\mu}\varepsilon\Omega$. The analysis of the PCS associated with R_2 is extremely involved for k > 2. In the following theorem we give without proof a general expression for $P_{\underline{\mu}}(CS|R_2)$ for any $\underline{\mu}\varepsilon\Omega$ for $k \geq 2$. The proof is omitted since it is very lengthy and tedious.

Theorem 3.4.1: For any $\underline{\mu}\varepsilon\Omega$ and for $k \geq 2$

$$P_{\underline{\mu}}(CS|R_2) = T_1 + T_2 + T_3 + T_4 \tag{3.4.1}$$

where $T_1 = \int_{-\infty}^{\infty} \prod_{i=1}^{k-1} \Phi\left[x + (\delta_{ki} - h)n_1^{1/2}/\sigma\right] d\Phi(x)$.

$$T_2 = \sum_{s\varepsilon\mathscr{S}^0} \Bigg[\int_{-\infty}^{\infty} \int_{x(q/p)^{1/2} - \frac{h}{\sigma}(n/q)^{1/2}}^{\infty} \prod_{i\varepsilon s} \left\{ \int_{x+(\delta_{ki}-h)n_1^{1/2}/\sigma}^{x+\delta_{ki}n_1^{1/2}/\sigma} \Phi\left[y + (x-z)(p/q)^{1/2} + \frac{\delta_{ki}}{\sigma}(n/q)^{1/2}\right] d\Phi(z) \right\} \prod_{i\not\varepsilon s} \Phi\left[x + (\delta_{ki} - h)n_1^{1/2}/\sigma\right] d\Phi(y)\, d\Phi(x) + \int_{-\infty}^{\infty} \int_{-\infty}^{x(q/p)^{1/2} - \frac{h}{\sigma}(n/q)^{1/2}} \prod_{i\varepsilon s} \left\{ \int_{x+(\delta_{ki}-h)n_1^{1/2}/\sigma}^{x+\delta_{ki}n_1^{1/2}/\sigma} \Phi\left[y + (x-z)(p/q)^{1/2} + \frac{\delta_{ki}}{\sigma}(n/q)^{1/2}\right] d\Phi(z) \right\} \times \prod_{i\not\varepsilon s} \Phi\left[p^{1/2}(xp^{1/2} + yq^{1/2}) + \delta_{ki}n_1^{1/2}/\sigma\right] d\Phi(y)\, d\Phi(x) \Bigg].$$

$$T_3 = \sum_{j=1}^{k-1} \sum_{s\varepsilon\mathscr{S}_j} \Bigg[\int_{-\infty}^{\infty} \int_{x(q/p)^{1/2} - \frac{h}{\sigma}(n/q)^{1/2}}^{\infty} \int_{x-(\delta_{kj}+h)n_1^{1/2}/\sigma}^{x-\delta_{kj}n_1^{1/2}/\sigma}$$

$$\int_{y+(x-u)(p/q)^{1/2}-\frac{\delta_{ki}}{\sigma}(n/q)^{1/2}}^{\infty} \prod_{i\varepsilon s} \left\{ \int_{x-(\delta_{ij}+h)n_1^{1/2}/\sigma}^{x-\delta_{ij}n_1^{1/2}/\sigma} \right.$$

$$\left. \Phi\left[v + (u-z)(p/q)^{1/2} + \frac{\delta_{ki}}{\sigma}(n/q)^{1/2}\right] d\Phi(z) \right\} \prod_{i \not\varepsilon s} \Phi\left[x - (\delta_{ij} + h)n_1^{1/2}/\sigma\right].$$

$$d\Phi(v)d\Phi(u)d\Phi(y)d\Phi(x) + \int_{-\infty}^{\infty}\int_{-\infty}^{x(q/p)^{1/2}-\frac{h}{\sigma}(n/q)^{1/2}} \int_{x-(\delta_{kj}+h)n_1^{1/2}/\sigma}^{x-\delta_{kj}n_1^{1/2}/\sigma}$$

$$\int_{x/(pq)^{1/2}-u(\frac{p}{q})^{1/2}-\frac{(\delta_{kj}+h)}{\sigma}(\frac{n}{q})^{1/2}}^{\infty} \prod_{i\varepsilon s} \left\{ \int_{x-(\delta_{ij}+h)n_1^{1/2}/\sigma}^{x-\delta_{ij}n_1^{1/2}/\sigma} \right.$$

$$\left. \Phi\left[v + (u-z)(p/q)^{1/2} + \frac{\delta_{ki}}{\sigma}(\frac{n}{q})^{1/2}\right] d\Phi(z) \right\} \prod_{i \not\varepsilon s}$$

$$\Phi\left[x - (\delta_{ij} + h)n_1^{1/2}/\sigma\right] d\Phi(v)d\Phi(u)d\Phi(y)d\Phi(x) + \int_{-\infty}^{\infty}\int_{-\infty}^{x(q/p)^{1/2}-\frac{h}{\sigma}(n/q)^{1/2}}$$

$$\int_{x-(\delta_{kj}+h)n_1^{1/2}/\sigma}^{x-\delta_{kj}n_1^{1/2}/\sigma} \int_{y+(x-u)(p/q)^{1/2}-\frac{\delta_{kj}}{\sigma}(n/q)^{1/2}}^{x/(pq)^{1/2}-u(p/q)^{1/2}-\frac{(\delta_{kj}+h)}{\sigma}(n/q)^{1/2}} \prod_{i\varepsilon s}$$

$$\left\{ \int_{x-(\delta_{ij}+h)n_1^{1/2}/\sigma}^{x-\delta_{ij}n_1^{1/2}/\sigma} \Phi\left[v + (u-z)(p/q)^{1/2} + \frac{\delta_{ki}}{\sigma}(n/q)^{1/2}\right] d\Phi(z) \right\} \times$$

$$\left. \prod_{i \not\varepsilon s} \Phi\left[p^{1/2}(up^{1/2} + vq^{1/2}) + \delta_{ki}n_1^{1/2}/\sigma\right] d\Phi(v)d\Phi(u)d\Phi(y)d\Phi(x) \right].$$

$$T_4 = \sum_{j=1}^{k-1} \sum_{s\varepsilon \mathscr{S}_j} \int_{-\infty}^{\infty} \int_{x(q/p)^{1/2} - \frac{h}{\sigma}(n/q)^{1/2}}^{\infty} \int_{x-(\delta_{kj}+h)n_1^{1/2}/\sigma}^{p^{1/2}(xp^{1/2}+yq^{1/2})-\delta_{kj}n_1^{1/2}/\sigma}$$

$$\prod_{i\varepsilon s} \left\{ \int_{x-(\delta_{ij}+h)n_1^{1/2}/\sigma}^{x-\delta_{ij}n_1^{1/2}/\sigma} \Phi\left[-z(p/q)^{1/2} + u/(pq)^{1/2} + \frac{\delta_{ki}}{\sigma}(n/q)^{1/2}\right] d\Phi(z) \right\} \times$$

$$\prod_{i \notin s} \Phi(u + \delta_{ki} n_1^{1/2}/\sigma)\, d\Phi(u)\, d\Phi(y)\, d\Phi(x).$$

In the above,

$\mathscr{S}^0$ = collection of all non-empty subsets from $\{1,2,\ldots,k-1\}$.

$\mathscr{S}_j$ = collection of all subsets from $\{1,\ldots,j-1,j+1,\ldots,k-1\}$.

$\mathscr{S}_j^0$ = collection of all non-empty subsets from $\{1,\ldots,j-1,j+1,\ldots,$ $k-1\}$.

The other notation is the same as in Theorem 2.2.1.

The complexity of the above result indicates the difficulties associated with an analysis of the simplest come-back type rule. In particular we were not able to prove that for the same values of n_1, n_2 and h we have $P_{\underline{\mu}}(CS|R_2) \geq P_{\underline{\mu}}(CS|R_1)$ for any $\underline{\mu}\varepsilon\Omega$. However in the MC results which we describe in the next section, we found that the estimates of $P_{\underline{\mu}(\delta^*)}(CS|R_2) \geq$ the estimates of $P_{\underline{\mu}(\delta^*)}(CS|R_1)$ for the same choice of (n_1,n_2,h) for all the cases that we studied.

We can define a U-minimax rule R_2 analogous to the definition of

R_1 in §2.4.1. If our conjecture that $P_{\underline{\mu}}(CS|R_2) \geq P_{\underline{\mu}}(CS|R_1)$ for the same choice of (n_1,n_2,h) for any $\underline{\mu}\epsilon\Omega$ is true, then for the same probability requirement (1.1.1) we would have $\min_{R_2} \max_{\Omega} E_{\underline{\mu}}(N|R_2) = E_{EMC}(N|\hat{R}_2) \leq E_{EMC}(N|\hat{R}_1) = \min_{R_1} \max_{\Omega} E_{\underline{\mu}}(N|R_1)$. This would be an improvement over R_1.

§3.4.2 MC sampling results for rules R_1 and R_2

Let the probability requirement (1.1.1) be preassigned and suppose (n_1,n_2,h) satisfy

$$(3.4.3) \quad \int_{-\infty}^{\infty} \Phi^{k-1}\left(x + \frac{(\delta^* + h)n_1^{1/2}}{\sigma}\right)d\Phi(x) + \int_{-\infty}^{\infty} \Phi^{k-1}\left(x + \frac{\delta^* n_1^{1/2}}{\sigma}\right)d\Phi(x) - 1 \geq P^*.$$

Then we know that rule R_1 using the same values of (n_1,n_2,h) satisfies the probability requirement (1.1.1). In particular $P_{\underline{\mu}(\delta^*)}(CS|R_1) \geq P^*$.

Let $\varepsilon_1(n_1,n_2,h,k) = P_{\underline{\mu}(\delta^*)}(CS|R_1) - P^*$, i.e. the overprotection afforded by rule R_1 using the same values of (n_1,n_2,h) and when the underlying parameter configuration is $\underline{\mu}(\delta^*)$. Also let $\varepsilon_2(n_1,n_2,h,k) = P_{\underline{\mu}(\delta^*)}(CS|R_2) - P_{\underline{\mu}(\delta^*)}(CS|R_1)$ where R_1 and R_2 both use the same values of (n_1,n_2,h).

We conducted MC sampling studies by simulating the operations of rules R_1 and R_2 to obtain estimates of the quantities $\varepsilon_1(n_1,n_2,h,k)$ and $\varepsilon_2(n_1,n_2,h,k)$ for various values of k and P^*. To obtain (n_1,n_2,h) we operated as follows: For given k and P^*, we fixed a $\delta^* > 0$ and $\sigma = 1$. We then chose $n_1 = (\hat{d}_1/\delta^*)^2$, $n_2 = (\hat{d}_2/\delta^*)^2$ and $h = \hat{c}/n_1^{1/2}$

where $(\hat{c},\hat{d}_1,\hat{d}_2)$ were obtained from Table 3.2.1. Note n_1 and n_2 are taken as continuous variables to insure exact equality in (3.4.3). For simulation purposes it does not matter if n_1 and n_2 are not integer valued since we only need to generate $\bar{X}_i^{(1)}$ and $\bar{X}_i$ $(1 \leq i \leq k)$ which depend on n_1 and n_2 through their variances. For this reason, our results do not depend on the value of δ^* chosen.

The simulation results are given in Tables 3.4.1 through 3.4.5 for k = 3,4,5,10 and 25. The tables also give many other related quantities such as estimates of the expected total sample sizes (in a reparametrized form) in the EMC and in the conjectured LFC, the corresponding expected subset sizes etc. The notation is mostly familiar except for the following three quantities.

(3.4.4)

ξ_1 = probability under $\underline{\mu}(\delta^*)$ configuration that R_1 terminates after the first stage.

ξ_2 = probability under $\underline{\mu}(\delta^*)$ configuration that R_1 terminates after the first stage with correct selection.

ξ_3 = probability under $\underline{\mu}(\delta^*)$ configuration that the best population enters the 2nd stage when using rule R_1.

Note that we have used the same notation to denote the quantities to be estimated and their estimates. In Figure 3.4.1 we have plotted

FIGURE 3.4.1.

Overprotection in the PCS afforded by $\hat{\bar{R}}_1$ in the LFC plotted against P^* for k = 3(1)5, 10 and 25.

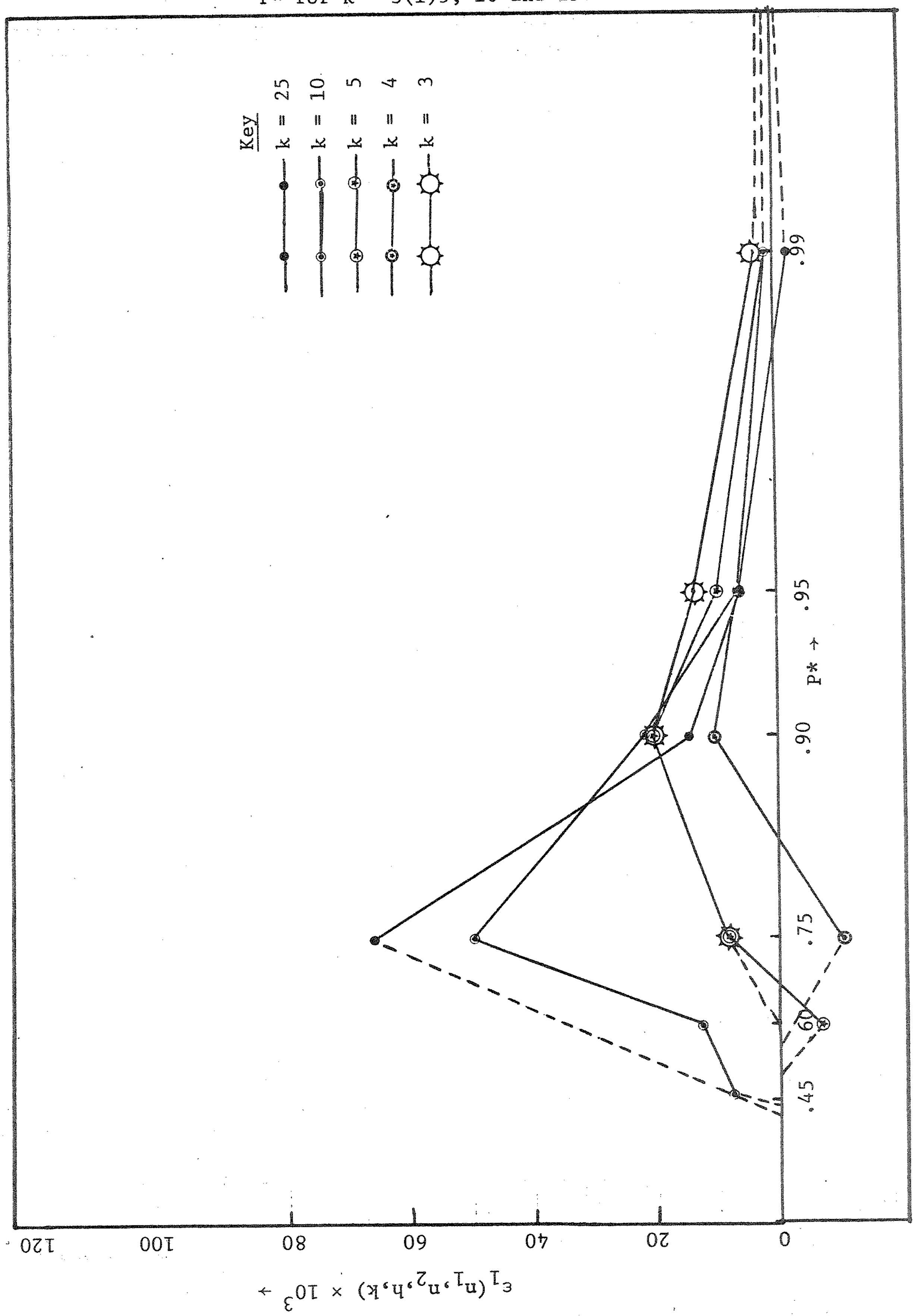


FIGURE 3.4.2

Excess PCS afforded by R_2 over that of R_1 in the LFC plotted against P^* for k = 3(1)5, 10 and 25.

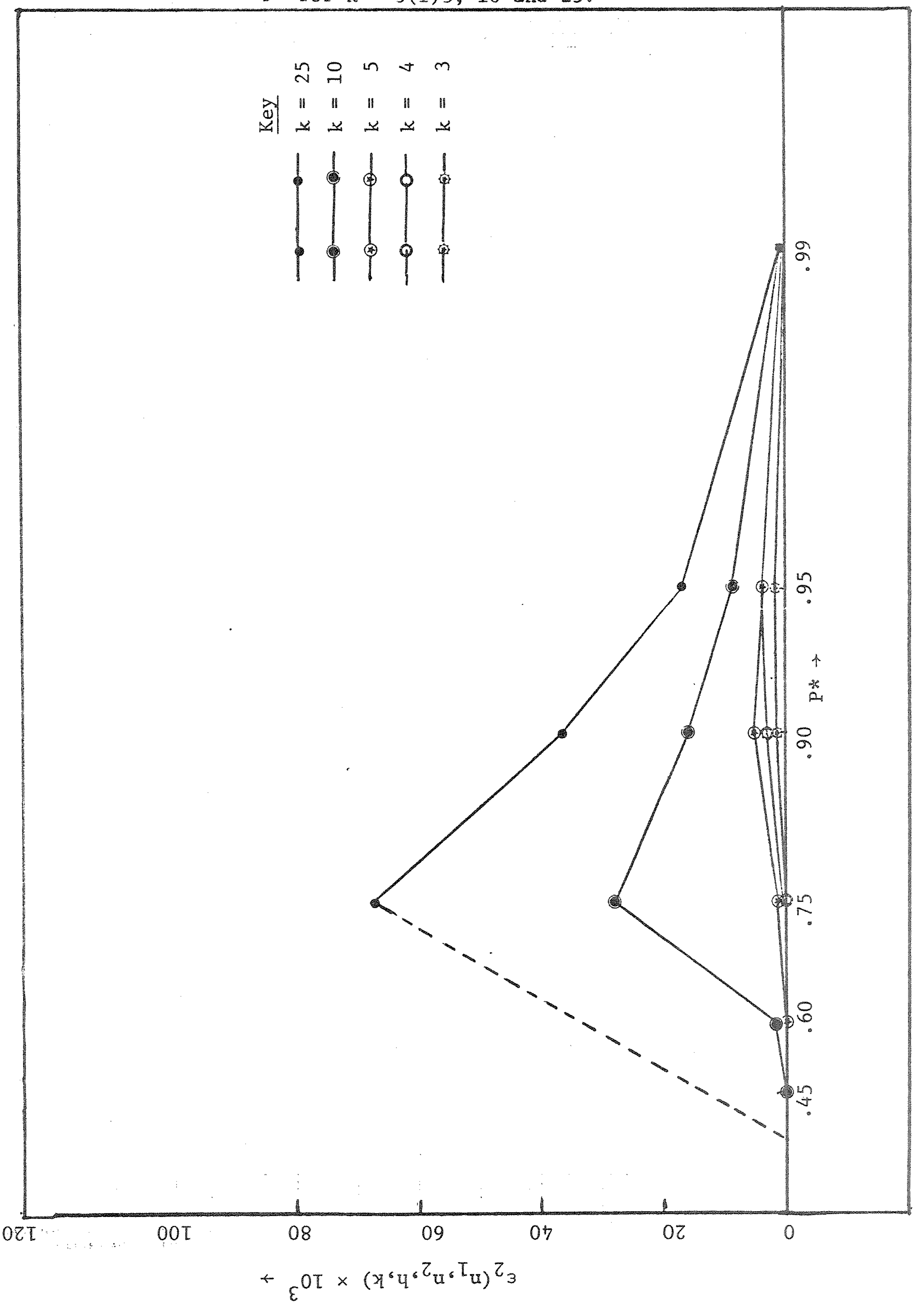


Table 3.4.1

MC sampling results for rules R_1 and R_2 (k = 3)

No. of experiments in each run = 3500.

P*	$P_{\underline{\mu}(\delta^*)}(CS\|R_1)$	$P_{\underline{\mu}(\delta^*)}(CS\|R_2)$	$\varepsilon_1(n_1,n_2,h,k)$	$\varepsilon_2(n_1,n_2,h,k)$	$\frac{\delta^{*2}}{k\sigma^2}E_{EMC}(N\|R_1)$	$E_{EMC}(T'\|R_1)$	$\frac{\delta^{*2}}{k\sigma^2}E_{\underline{\mu}(\delta^*)}(N\|R_1)$	$E_{\underline{\mu}(\delta^*)}(T'\|R_1)$	ξ_1	ξ_2	ξ_3
.99	.99314 (.00139)	.99343 (.00137)	.00314 (.00139)	.00029	12.26559 (.03723)	1.89600 (.01926)	9.34019 (.02741)	.38286 (.01418)	.82257 (.00646)	.82000 (.00649)	.99714 (.00090)
.95	.96343 (.00317)	.96429 (.00314)	.01343 (.00317)	.00086	7.10717 (.02000)	2.28514 (.01667)	5.83723 (.02494)	1.22686 (.02078)	.48029 (.00845)	.47543 (.00844)	.99371 (.00134)
.90	.91314 (.00476)	.91343 (.00475)	.01314 (.00476)	.00029	4.90787 (.00961)	2.64943 (.01220)	4.40597 (.01550)	2.02200 (.01938)	.21543 (.00695)	.21171 (.00691)	.99457 (.00124)
.75	.75829 (.00724)	.75829 (.00724)	.00829 (.00724)	.00000	2.05431 (.00083)	2.97971 (.00258)	2.04517 (.00140)	2.95114 (.00438)	.00343 (.00099)	.00314 (.00095)	.99971 (.00029)

(The standard errors of the estimates are given in round brackets under the corresponding estimates).

Table 3.4.2

MC sampling results for rules R_1 and R_2 $(k = 4)$

No. of experiments in each run = 3500

P^*	$P(CS\mid R_1)$ $\underline{\mu}(\delta^*)$	$P(CS\mid R_2)$ $\underline{\mu}(\delta^*)$	$\varepsilon_1(n_1,n_2,h,k)$	$\varepsilon_2(n_1,n_2,h,k)$	$\frac{\delta^{*2}}{k\sigma^2}E_{EMC}(N\mid R_1)$	$E_{EMC}(T'\mid R_1)$	$\frac{\delta^{*2}}{k\sigma^2}E_{\underline{\mu}(\delta^*)}(N\mid R_1)$	$E_{\underline{\mu}(\delta^*)}(T'\mid R_1)$	ξ_1	ξ_2	ξ_3
.99	.99314 (.00139)	.99429 (.00127)	.00314 (.00139)	.00115	13.25629 (.03563)	2.41086 (.02140)	10.07943 (.02769)	.50257 (.01664)	.78171 (.00698)	.78000 (.00700)	.99686 (.00095)
.95	.95657 (.00345)	.95943 (.00333)	.00657 (.00345)	.00286	8.04500 (.02266)	2.71857 (.02035)	6.52477 (.02603)	1.35343 (.02337)	.47400 (.00844)	.46771 (.00843)	.98857 (.00180)
.90	.91057 (.00482)	.91343 (.00475)	.01057 (.00482)	.00293	5.81674 (.014282)	3.11229 (.01791)	5.05493 (.01941)	2.15686 (.02435)	.25714 (.00739)	.25114 (.00733)	.98686 (.00193)
.75	.74000 (.00741)	.74000 (.00741)	−.01000 (.00741)	.00000	2.82114 (.00219)	3.88400 (.00638)	2.76676 (.00371)	3.72543 (.01082)	.01143 (.00180)	.01086 (.00175)	.99800 (.00076)

(The standard errors of the estimates are given in round brackets under the corresponding estimates).

Table 3.4.3

MC sampling results for rules R_1 and R_2 (k = 5)

No. of experiments in each run = 3500

P^*	$P_{\underline{\mu}(\delta^*)}(CS \mid R_1)$	$P_{\underline{\mu}(\delta^*)}(CS \mid R_2)$	$\varepsilon_1(n_1,n_2,h,k)$	$\varepsilon_2(n_1,n_2,h,k)$	$\frac{\delta^{*2}}{k\sigma^2}E_{EMC}(N \mid R_1)$	$E_{EMC}(T' \mid R_1)$	$\frac{\delta^{*2}}{k\sigma^2}E_{\underline{\mu}(\delta^*)}(N \mid R_1)$	$E_{\underline{\mu}(\delta^*)}(T' \mid R_1)$	ξ_1	ξ_2	ξ_3
.99	.99114 (.00158)	.99229 (.00148)	.00114 (.00158)	.00115	13.81110 (.03639)	2.80371 (.02442)	10.56318 (.02862)	.62486 (.01920)	.74686 (.00735)	.74486 (.00737)	.99657 (.00099)
.95	.96000 (.00331)	.96343 (.00317)	.01000 (.00331)	.00343	8.59343 (.02445)	3.14429 (.02389)	6.98678 (.02712)	1.57457 (.02650)	.44171 (.00839)	.43514 (.00838)	.98771 (.00186)
.90	.91343 (.00475)	.91829 (.00463)	.01343 (.00475)	.00486	6.38352 (.01785)	3.44000 (.02280)	5.46338 (.02195)	2.26486 (.02804)	.27857 (.00758)	.26943 (.00750)	.97771 (.00250)
.75	.75886 (.00723)	.75971 (.00722)	.00886 (.00723)	.00085	3.37251 (.00461)	4.58571 (.01294)	3.24719 (.00654)	4.23371 (.01837)	.02200 (.00248)	.02171 (.00246)	.99286 (.00142)
.60	.59371 (.00830)	.59371 (.00830)	-.00629 (.00830)	.00000	1.76436 (.00053)	4.97400 (.00311)	1.76021 (.00075)	4.94943 (.00446)	.00000 (.00000)	.00000 (.00000)	.99971 (.00029)

(The standard errors of the estimates are given in round brackets under the corresponding estimates).

Table 3.4.4

MC sampling results for rules R_1 and R_2 $(k = 10)$

No. of experiments in each run = 3500

P^*	$P_{\underline{\mu}(\delta^*)}(CS\|R_1)$	$P_{\underline{\mu}(\delta^*)}(CS\|R_2)$	$\varepsilon_1(n_1,n_2,h,k)$	$\varepsilon_2(n_1,n_2,h,k)$	$\frac{\delta^{*2}}{k\sigma^2}E_{EMC}(N\|R_1)$	$E_{EMC}(T'\|R_1)$	$\frac{\delta^{*2}}{k\sigma^2}E_{\underline{\mu}(\delta^*)}(N\|R_1)$	$E_{\underline{\mu}(\delta^*)}(T'\|R_1)$	ξ_1	ξ_2	ξ_3
.99	.99029 (.00166)	.99229 (.00148)	.00029 (.00166)	.00200	14.96363 (.03716)	4.49029 (.03717)	11.58199 (.02836)	1.10800 (.02836)	.63457 (.00814)	.63229 (.00815)	.99457 (.00124)
.95	.95743 (.00341)	.96629 (.00305)	.00743 (.00341)	.00886	9.77984 (.02897)	4.52229 (.03763)	7.94731 (.02793)	2.14257 (.03627)	.39829 (.00827)	.39057 (.00825)	.97943 (.00240)
.90	.92286 (.00451)	.93857 (.00406)	.02286 (.00451)	.01571	7.58342 (.02415)	4.63057 (.03758)	6.39395 (.02462)	2.77943 (.03831)	.27943 (.00758)	.26771 (.00748)	.95829 (.00338)
.75	.80077 (.00666)	.82886 (.00637)	.05077 (.00666)	.02809	4.71253 (.01506)	5.76714 (.03872)	4.26622 (.01647)	4.62000 (.04233)	.09457 (.00495)	.08571 (.00473)	.94086 (.00399)
.60	.61343 (.00823)	.61514 (.00822)	.01343 (.00823)	.00171	3.07570 (.00475)	8.65771 (.02611)	2.98308 (.00594)	8.14886 (.03261)	.00571 (.00127)	.00514 (.00121)	.98114 (.00230)
.45	.45829 (.00842)	.45829 (.00842)	.00829 (.00842)	.00000	1.76543 (.00049)	9.92514 (.00565)	1.75992 (.00077)	9.86114 (.00899)	.00029 (.00029)	.00029 (.00029)	.99943 (.00040)

(The standard errors of the estimates are given in round brackets under the corresponding estimates).

Table 3.4.5

MC sampling results for rules R_1 and R_2 (k = 25)

No. of experiments in each run = 2000

P*	$P_{\underline{\mu}(\delta^*)}(CS\|R_1)$	$P_{\underline{\mu}(\delta^*)}(CS\|R_2)$	$\varepsilon_1(n_1,n_2,h,k)$	$\varepsilon_2(n_1,n_2,h,k)$	$\frac{\delta^{*2}}{k\sigma^2}E_{EMC}(N\|R_1)$	$E_{EMC}(T'\|R_1)$	$\frac{\delta^{*2}}{k\sigma^2}E_{\underline{\mu}(\delta^*)}(N\|R_1)$	$E_{\underline{\mu}(\delta^*)}(T'\|R_1)$	ξ_1	ξ_2	ξ_3
.99	.98950 (.00228)	.99200 (.00199)	-.00050 (.00228)	.00250	15.84796 (.04867)	8.47750 (.09097)	12.47559 (.03521)	2.17400 (.06581)	.50000 (.01118)	.49700 (.01118)	.99450 (.00165)
.95	.95600 (.00459)	.97350 (.00359)	.00600 (.00459)	.01750	10.70219 (.03935)	8.16800 (.08923)	8.82993 (.03675)	3.92250 (.08332)	.27150 (.00994)	.26500 (.00987)	.97200 (.00369)
.90	.91450 (.00625)	.95150 (.00480)	.01450 (.00625)	.03700	8.50029 (.03438)	7.82650 (.08722)	7.24788 (.03358)	4.64950 (.08519)	.19850 (.00892)	.18850 (.00875)	.94050 (.00529)
.75	.81500 (.00868)	.88200 (.00721)	.06500 (.00868)	.06700	5.60515 (.02758)	7.25950 (.08467)	5.07148 (.02749)	5.62150 (.08437)	.11200 (.00705)	.09350 (.00651)	.86750 (.00758)

estimated overprotection $\varepsilon_1(n_1,n_2,h,k)$ as a function of P* for different values of k. For a fixed k, we notice that $\varepsilon_1(n_1,n_2,h,k)$ goes to 0 for very high and low values of P* and achieves a maximum at a certain intermediate value of P*. Since $P^* + \varepsilon_1(n_1,n_2,h,k) \leq 1$ and $\varepsilon_1(n_1,n_2,h,k) \geq 0$ it is clear that $\varepsilon_1(n_1,n_2,h,k) \to 0$ as $P^* \to 1$. The explanation for the observed behavior of $\varepsilon_1(n_1,n_2,h,k)$ over the lower range of P* is as follows: Since for given k and P*, we choose (n_1,n_2,h) from the optimal solution for the <u>conservative</u> rule $\bar{R}_1$, and since $\hat{c}$ (and hence h) increases as P* decreases, rule R_1 operates as if it is a single-stage rule R_0 with $(n_1 + n_2)$ as the common sample size per population. Hence $P_{\underline{\mu}(\delta^*)}(CS|R_1)$ approaches $P_{\underline{\mu}(\delta^*)}(CS|R_0)$ and also ℓ.h.s. of (3.4.3) approaches $P_{\underline{\mu}(\delta^*)}(CS|R_0)$. Thus eventually $\varepsilon_1(n_1,n_2,h)$ starts decreasing and approaches zero as P* is decreased. For fixed P* the behavior of $\varepsilon_1(n_1,n_2,h,k)$ w.r.t. k is not very clear for small values of k and extreme values of P*. But for k = 10 and 25, the overprotection appears to be substantially higher at intermediate P* values. It is expected that for fixed P* the overprotection would increase with k. For few extreme values of P* we observed "negative" overprotection. However these results are not significantly negative and may be ascribed to the statistical error of simulation. In Figure 3.4.2 we have also plotted estimated $\varepsilon_2(n_1,n_2,h,k)$ as a function of P* for different values of k. The behavior of $\varepsilon_2(n_1,n_2,h,k)$ with respect to P* and k is quite similar to the behavior of $\varepsilon_1(n_1,n_2,h,k)$ and for similar reasons. The estimated total expected sample sizes in the EMC compare well with the ones computed numerically and given in Table 3.2.1. For fixed k, the variation in ξ_1,ξ_2 and ξ_3, w.r.t. P* can be explained on the basis of variation in $\hat{\bar{c}}$ w.r.t. P*.

In closing, we note that the overprotection $\varepsilon_1(n_1,n_2,h,k)$ depends crucially on the (n_1,n_2,h)-values under consideration. Thus although the overprotection afforded by rule R_1 using the optimum values for rule $\hat{\bar{R}}_1$ goes to zero in the lower range of P*-values, this does not imply that rule $\hat{\bar{R}}_1$ performs as well as rule $\hat{R}_1$. In fact, potentially large gains in terms of relative efficiency are possible in the lower range of P*-values if we use the exact rule R_1. More research needs to be conducted in this direction, although for high P*-values, which a practitioner uses as a matter of habit, our present lower bound is found to be satisfactory.

The MC studies also indicate the superiority of rule R_2 over R_1. However, it should be pointed out that since the increase of additional PCS achieved by rule R_2 over rule R_1, namely $\varepsilon_2(n_1,n_2,h,k)$,depends on the (n_1,n_2,h)-values selected; the MC results do not give a true indication of the extent of potential gain in terms of relative efficiency, that is possible if we use rule R_2.

§3.5 <u>Some further extensions of come-back type rules</u>:

We consider the same setup as described in §1.1. The possibility of sampling from a single population in the second stage is ruled out in the most elementary come-back type rule R_2. Although, it appears reasonable that if the largest first stage sample mean is sufficiently bigger than all the others then we should stop sampling, a single "yardstick" of h may not be sufficient. For example, if the largest first stage sample mean is only "moderately" larger than the remaining first stage sample means, then we may want to perform more sampling on the population associated with that mean before we make our final selection. The following rule R_3 responds to such contingencies.

We now propose our rule R_3.

§3.5.1. Rule R_3 and its properties:

Rule R_3: Let non-negative integers n_1, n_2 and non-negative real constants $h_1 \geq h_2$ be specified prior to the start of experimentation. The following are the steps in rule R_3 specified by (n_1, n_2, h_1, h_2) which are chosen to guarantee (1.1.1).

1. In the first stage, from each Π_i take n_1 independent observations $X_{ij}^{(1)}$ $(1 \leq j \leq n_1)$ and compute $\overline{X}_i^{(1)} = \sum_{j=1}^{n_1} X_{ij}^{(1)}/n_1$ $(1 \leq i \leq k)$. Let $\overline{X}_{[1]}^{(1)} \leq \overline{X}_{[2]}^{(1)} \leq \cdots \leq \overline{X}_{[k]}^{(1)}$ be the corresponding ordered values.

2. If $\overline{X}_{[k]}^{(1)} - \overline{X}_{[k-1]}^{(1)} > h_1$, stop sampling at the first stage, and assert that the population associated with $\overline{X}_{[k]}^{(1)}$ is best.

3. If not, choose a subset I of $\{1,2,\ldots,k\}$ where

$$I = \{i \mid \overline{X}_i^{(1)} \geq \overline{X}_{[k]}^{(1)} - h_2\} \tag{3.5.1}$$

(Note I may consist of a single population.)

4. Proceed to the second stage and take n_2 additional independent observations $X_{ij}^{(2)}$ $(1 \leq j \leq n_2)$ from each Π_i for $i \varepsilon I$. Compute the cumulative sample means $\overline{X}_i = (\sum_{j=1}^{n_1} X_{ij}^{(1)} + \sum_{j=1}^{n_2} X_{ij}^{(2)})/(n_1 + n_2)$ for $i \varepsilon I$

and assert that the population associated with $\max\{\bar{X}_i, i\varepsilon I;\ \bar{X}_i^{(1)}, i\varepsilon I\}$ is best.

Note for $k = 2$, rule R_3 is the same as rule R_1 (and R_2) iff $h_1 = h_2$. The analysis of rule R_3 is quite involved and tedious. We present here the results concerning the PCS and the expected total sample size of R_3 for $k = 2$. For reasons of conciseness, we omit the proofs of the following results.

Theorem 3.5.1: For $k = 2$ and for any $\underline{\mu}\varepsilon\Omega$

$$(3.5.2)\quad P_{\underline{\mu}}(CS|R_3) = \Phi[A_1(\delta-h_1)] + \int_{A_1(\delta-h_1)}^{A_1(\delta-h_2)} \Phi(-A_2x + A_3\delta)d\Phi(x) +$$

$$\int_{A_1(\delta-h_2)}^{A_1(\delta+h_2)} \Phi(-A_4x + A_5\delta)d\Phi(x) + \int_{A_1(\delta+h_2)}^{A_1(\delta+h_1)} \Phi(-A_2x + A_3\delta)d\Phi(x),$$

where $A_1 = \frac{1}{\sigma}(\frac{n_1}{2})^{1/2}$, $A_2 = \left(\frac{1+p}{1-p}\right)^{1/2}$, $A_3 = \frac{1}{\sigma}\left[\frac{2n_1}{1-p^2}\right]^{1/2}$, $A_4 = \left(\frac{p}{1-p}\right)^{1/2}$, $A_5 = \frac{1}{\sigma}\left[\frac{n_1}{2p(1-p)}\right]^{1/2}$ and $\delta = \mu_{[2]} - \mu_{[1]}$. The other notation is as defined in Theorem 2.2.1.

Theorem 3.5.2: For $k = 2$, $P_{\underline{\mu}}(CS|R_3)$ is non-decreasing in δ for $\delta \geq 0$ and hence $\delta = \delta^*$ is a LFC for rule R_3.

Theorem 3.5.3: (i) For $k = 2$ and for any $\underline{\mu}\varepsilon\Omega$

$$(3.5.3)\quad E_{\underline{\mu}}(N|R_3) = 2n_1 + n_2\{\Phi[A_1(\delta + h_1)] + \Phi[A_1(\delta + h_2)] -$$

$$\Phi[A_1(\delta - h_1)] - \Phi[A_1(\delta - h_2)]\}.$$

(ii) $E_{\underline{\mu}}(N|R_3)$ is non-increasing in δ for $\delta \geq 0$. Hence $\sup_{\Omega} E_{\underline{\mu}}(N|R_3) = E_{\underline{\mu}}(N|R_3)\Big|_{\delta=0}$ and $\sup_{\Omega(\delta^*)} E_{\underline{\mu}}(N|R_3) = E_{\underline{\mu}}(N|R_3)\Big|_{\delta=\delta^*}$.

We may define a U-minimax rule $\hat{R}_3$ analogous to $\hat{R}_1$ and also the relative efficiency of R_0 w.r.t. $\hat{R}_3$. Since $\hat{R}_1$ is a special case of R_3 when $h_1 = h_2$ we have the following set of inequalities for k = 2.

$$(3.5.4) \quad 1 \geq RE_{EMC}(P^*,k;\ \hat{R}_1) \geq RE_{EMC}(P^*,k;\ \hat{R}_3) \geq RE_{\underline{\mu}}(\delta^*,P^*,k;\ \hat{R}_3)$$

for all $\underline{\mu}\varepsilon\Omega$ and $\delta^* > 0$.

We do not plan a detailed study of rule R_3 at this point. In the following section we shall make some comments about a two-stage rule of a very general nature which one may eventually want to analyze.

§3.5.2 Concluding remarks:

In the elimination type two-stage rules, whether of a permanent elimination nature (like rule R_1) or of a come-back nature (like rules R_2 and R_3), we assign a probability of zero or one to each population for it to be retained in the second stage. In general, one may assign probability p_i that the population Π_i $(1 \leq i \leq k)$ is retained in the second stage (n_2 is still non-random); p_i's will be some functions of the first stage sample means. Possibly p_i may be made to depend only on the difference between the largest first stage sample mean and $\overline{X}_i^{(1)}$, i.e., $p_i = p_i(\overline{X}_{[k]}^{(1)} - \overline{X}_i^{(1)})$. One such rule has been mentioned by

Bessler [1960]. In the special case of rule R_1 we have

$$(3.5.5) \qquad p_i = \begin{cases} 1 & \text{if } \bar{X}^{(1)}_{[k]} - \bar{X}^{(1)}_i \leq h. \\ 0 & \text{if } \bar{X}^{(1)}_{[k]} - \bar{X}^{(1)}_i > h. \end{cases}$$

A basic difficulty in analyzing any such rule will be to identify the LFC of the exact rule. For this and the computational reasons, the task of computing an optimum rule of the above general nature may turn out to be a formidable one. However even the development of a "reasonable" rule of the above type, which guarantees the probability requirement, will be a significant contribution to the field of ranking and selection procedures.

CHAPTER 4

OPTIMALITY OF RULE $\hat{R}_1$ IN A CLASS OF TWO-STAGE PERMANENT ELIMINATION TYPE RULES

§4.0 Introduction:

In the present chapter we investigate the nature of U-minimax two-stage permanent elimination type rules for selecting the largest mean from k normal populations when the common variance is known. Our main conclusion is that,in a restricted class of rules as described in §2.1.1, our natural selection rule $\hat{R}_1$ is U-minimax for $k = 2$. Although we were not able to draw any concrete conclusions for $k > 2$ due to the analytical complexities of the problem, our calculations indicate that the rule $\hat{R}_1$ is not U-minimax in the class of rules under study for $k > 2$. Our analysis also throws some light on the nature of U-minimax rules viz. how the populations to be retained in the second stage are selected by the U-minimax rule. A brief review of the contents of the present chapter follows.

In §4.1 we describe a very general set up including the underlying loss function for a k-decision identification problem. The design criterion is to minimize the expected total number of observations at some parameter point other than the parameter points specified by the k identification hypotheses, subject to guaranteeing a specified probability of making a correct decision. By imposing certain natural symmetry restrictions, we show in §4.2 that such rules are Bayes rules against a symmetric prior (defined in §4.2). We also give sufficient conditions under which the Bayes rule solves the underlying ranking problem and is U-minimax i.e. minimizes the maximum (over the parameter space) of the total expected sample size among all the symmetric rules which guarantee the specified probability requirement for the ranking problem.

In §4.3 we apply this theory to the normal means problem of §2.1.4 and characterize the Bayes two-stage rules for the identification problem. In §4.4 we consider the special case k = 2 and show that the Bayes rule solves the ranking problem and is U-minimax in the class of rules under consideration. Further, it has the same structure as rule $\hat{R}_1$ and therefore rule $\hat{R}_1$ is U-minimax for k = 2 being the Bayes rule.

§4.1 Preliminaries:

§4.1.1 Assumptions and notation:

Let $\Pi_1, \Pi_2, \ldots, \Pi_k$ be $k \geq 2$ populations with probability densities $g(x, \theta_i, \underline{\psi})$ w.r.t. some σ-finite measure ν. In the above $\theta_i \varepsilon \Theta$ $(1 \leq i \leq k)$ is the parameter of interest and $\underline{\psi}$ is a vector of nuisance parameters which we shall ignore in the sequel. Let the parameter space Ω be the collection of all parameter vectors $\underline{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)^t$. We assume that the following symmetry condition is satisfied by the joint density function.

Let $x_i = (x_{i1}, x_{i2}, \ldots, x_{in_i})^t$ for $1 \leq i \leq k$ and let $g(x_1, \ldots, x_k; \theta_1, \ldots, \theta_k)$ be the joint density function of the corresponding $\sum_{i=1}^{k} n_i$ dimensional random variable $(\underline{X}_1, \ldots, \underline{X}_k)$. Then

(4.1.1) $g(x_1,\ldots,x_k;\ \theta_1,\ldots,\theta_k) = g(x_{\tau_1},\ldots,x_{\tau_k};\ \theta_{\tau_1},\ldots,\theta_{\tau_k})$

for every permutation τ: $(1,2,\ldots,k) \to (\tau_1,\tau_2,\ldots,\tau_k)$ and for every possible vector $(x_1,\ldots,x_k)$.

§4.1.2 An identification problem:

Let $\underline{\omega}_i$ $(0 \leq i \leq k)$ be k completely specified states of nature. In particular, let $\underline{\omega}_1,\ldots,\underline{\omega}_k$ be k slippage configurations with $\underline{\omega}_i = (\theta,\ldots,\bar{\theta},\ldots,\theta)^t, \bar{\theta} \neq \theta$ $(1 \leq i \leq k)$ where $\bar{\theta}$ is in the ith place. Let ω_0 be some other parameter configuration. We assume that no two $g(x_1,\ldots,x_k;\ \underline{\omega}_i)$ and $g(x_1,\ldots,x_k;\ \underline{\omega}_j)$ are identical a.e. ν for $0 \leq i \neq j \leq k$.

We assume that there are exactly k possible terminal decisions $d_1,d_2,\ldots,d_k$ one of which must be made. Here the decision d_i corresponds to deciding that $\underline{\omega}_i$ is the true state of nature $(1 \leq i \leq k)$. We consider a general class of randomized decision rules with a symmetry restriction as follows.

Definition 4.1.1: A symmetric decision rule is a decision rule which is invariant w.r.t. the group operations of permutations τ: $(1,2,\ldots,k) \to (\tau_1,\tau_2,\ldots,\tau_k)$. That is, if at any stage of experimentation, after having observed $(x_1,\ldots,x_k)$ it assigns probability $P_{(n_1,\ldots,n_k)}(x_1,\ldots,x_k)$ to taking the vector of observations $\underline{n} = (n_1,n_2,\ldots,n_k)^t$ in the next stage where n_i is the number of observations from Π_i $(1 \leq i \leq k)$, then

(4.1.2a) $$P_{(n_1,\ldots,n_k)}(x_1,\ldots,x_k) = P_{(n_{\tau_1},\ldots,n_{\tau_k})}(x_{\tau_1},\ldots,x_{\tau_k}).$$

Further if it assigns probability $P_i(x_1,\ldots,x_k)$ to taking the terminal decision d_i $(1 \leq i \leq k)$ then

(4.1.2b) $$P_i(x_1,\ldots,x_k) = P_{\tau_i}(x_{\tau_1},\ldots,x_{\tau_k})$$

for all permutations $\tau: (1,2,\ldots,k) \to (\tau_1,\tau_2,\ldots,\tau_k)$.

Let P^*, $k^{-1} < P^* < 1$, be specified. Let $\Gamma'(P^*)$ denote the class of all symmetric rules γ which terminate with probability 1 and which satisfy

(4.1.3) $$P_{\underline{\omega}_i}(\gamma \text{ makes decision } d_i) = P_{\underline{\omega}_i}(d_i|\gamma) \geq P^* \quad \text{for} \quad i = 1,\ldots,k.$$

Our objective is to choose $\gamma^* \varepsilon \Gamma'(P^*)$ such that

(4.1.4) $$E_{\underline{\omega}_0}(N|\gamma^*) = \inf_{\gamma\varepsilon\Gamma'(P^*)} E_{\underline{\omega}_0}(N|\gamma)$$

where $E_{\underline{\omega}_0}(N|\gamma)$ is the expected value under $\underline{\omega}_0$ of the total random number of observations required by rule γ. Such a γ^* will be regarded as "optimum".

§4.1.3 <u>The underlying loss function</u>:

Loss due to a terminal decision is given by

$$(4.1.5) \qquad L(\underline{\omega}_i, d_j) = \begin{cases} 1 & \text{if } i \neq j. \\ 0 & \text{if } i = j \text{ or } i = 0. \end{cases}$$

Thus if the true state of nature is $\underline{\omega}_0$, it does not matter which terminal decision is made. In addition, C(N), the cost of sampling N observations is given by

$$(4.1.6) \qquad C(N) = \begin{cases} N & \text{if } \underline{\theta} = \underline{\omega}_0 \\ 0 & \text{otherwise} \end{cases}$$

This results in the following risk function $r(\underline{\theta},\gamma)$.

$$(4.1.7) \qquad r(\underline{\theta},\gamma) = \begin{cases} 1 - P_{\underline{\omega}_i}(d_i|\gamma) & \text{for } \underline{\theta} = \underline{\omega}_i \ (1 \leq i \leq k) \\ E_{\underline{\omega}_0}(N|\gamma) & \text{for } \underline{\theta} = \underline{\omega}_0. \end{cases}$$

For a similar loss function see, e.g., Weiss [1964].

We characterize the optimum rules in the next section.

§4.2 Characterization of optimum rules:

§4.2.1 U-minimax rules for the identification problem

We first state a few basic definitions and lemmas before considering the main result of this section which is given in Theorem 4.2.1.

Definition 4.2.1: A decision rule γ^0 is said to be a Bayes rule against a prior $B(\underline{\theta})$ if it minimizes the integrated risk

$R(B,\gamma) = \int_{\Omega} r(\underline{\theta},\gamma)dB(\underline{\theta})$ in the class of all decision rules.

We shall need only finite priors (i.e. $\int_{\Omega} dB(\underline{\theta}) < \infty$) which may be taken as probability measures on Ω. Here we are concerned only with a finite subset of the parameter space namely $(\underline{\omega}_0,\underline{\omega}_1,\ldots,\underline{\omega}_k)$. Let $B = (b_0,b_1,\ldots,b_k)$ be the prior on $(\underline{\omega}_0,\underline{\omega}_1,\ldots,\underline{\omega}_k)$ with $b_i \geq 0$ and $\sum_{i=1}^{k} b_i = 1$.

<u>Definition 4.2.2</u>: We call $B = (b_0,b_1,\ldots,b_k)$ a <u>symmetric prior</u> if $b_1 = b_2 = \cdots = b_k = b$ with $0 \leq b \leq k^{-1}$ and $b_0 = 1 - kb$. We shall denote this prior simply by b.

<u>Lemma 4.2.1</u>: For every symmetric decision rule γ for the loss and the sampling cost functions given by (4.1.5) and (4.1.6), the risk function $r(\underline{\theta},\gamma)$ is invariant w.r.t. the group operations of permutations τ: $(1,2,\ldots,k) \rightarrow (\tau_1,\tau_2,\ldots,\tau_k)$.

<u>Proof</u>: Follows from the invariance of the loss function, the joint density function, and the decision rule.

<u>Lemma 4.2.2</u>: Every symmetric Bayes rule for the loss and the sampling cost functions given by (4.1.5) and (4.1.6), is Bayes against a symmetric prior.

Proof: The proof is similar to the proof of Theorem 3.1 of Karlin and Traux [1960] and is omitted.

Now we state the main theorem of this section.

Theorem 4.2.1: (i) For a given $P^* \varepsilon (k^{-1},1)$ every symmetric optimum procedure is Bayes against a symmetric prior for some $b \varepsilon (0,k^{-1})$ i.e. it minimizes the expression,

$$R(b,\gamma) = (1-kb)E_{\underline{\omega}_0}(N|\gamma) + b\sum_{i=1}^{k}\{1 - P_{\underline{\omega}_i}(d_i|\gamma)\}. \tag{4.2.1}$$

(ii) Conversely if any γ^* minimizes $R(b,\gamma)$ for some $b\varepsilon(0,k^{-1})$, then γ^* is a symmetric optimum procedure for some $P^*\varepsilon(k^{-1},1)$.

Proof: (i) Let $\alpha_i(\gamma) = 1 - P_{\underline{\omega}_i}(d_i|\gamma)$ for $1 \leq i \leq k$. Consider a $(k + 1)$-dimensional risk set $A = \{(\alpha_1(\gamma),\ldots,\alpha_k(\gamma);\ E_{\underline{\omega}_0}(N|\gamma))|\gamma\varepsilon\Gamma'\}$ where Γ' = the collection of all symmetric decision rules (not just $\Gamma'(P^*)$). Clearly A is convex. For symmetric rules we have $\alpha_1(\gamma) = \cdots = \alpha_k(\gamma)$. Hence we need only consider $A_1 = \{(\alpha_1(\gamma);\ E_{\underline{\omega}_0}(N|\gamma))|\gamma\varepsilon\Gamma'\}$, the projection of A. Note that A_1 is also convex.

The rest of the proof of part (i) of the theorem is similar to the proof of Lemma 4.1 of Kiefer and Weiss [1957] and the reader is referred to that paper for the completion of the proof.

(ii) Let γ^* be a Bayes procedure against a symmetric prior $b\varepsilon(0,k^{-1})$

and let $\alpha_1(\gamma^*) = \alpha^* = 1 - P^*$. Then

$$R(b,\gamma^*) \leq R(b,\gamma) \quad \forall \gamma\varepsilon\Gamma'$$

$$\iff (1-kb)E_{\underline{\omega}_0}(N|\gamma^*) + kb\alpha_1(\gamma^*) \leq (1-kb)E_{\underline{\omega}_0}(N|\gamma) + kb\alpha_1(\gamma) \quad \forall\gamma\varepsilon\Gamma'$$

$$\iff \quad E_{\underline{\omega}_0}(N|\gamma^*) - E_{\underline{\omega}_0}(N|\gamma) \leq \frac{kb}{1-kb}(\alpha_1(\gamma) - \alpha^*) \quad \forall\gamma\varepsilon\Gamma'$$

$$\leq 0 \quad \forall\gamma\varepsilon\Gamma'(P^*).$$

Therefore γ^* is optimum for some $P^*\varepsilon(k^{-1},1)$. This completes the proof.

It should be noted that the above theorem does not give us a functional relationship between P^* and b. Thus we do not have a mechanism for constructing an optimal procedure for any given P^*. The above is simply a characterization theorem. Note that b increases with P^* and $b \to k^{-1}$ as $P^* \to 1$.

§4.2.2 <u>The underlying ranking problem and the corresponding U-minimax rules</u>:

Suppose $\Omega_1,\Omega_2,\ldots,\Omega_k$ are k symmetric disjoint non-empty subsets of Ω where $(\theta_1,\theta_2,\ldots,\theta_k)\varepsilon\Omega_i \implies (\theta_{\tau_1},\theta_{\tau_2},\ldots,\theta_{\tau_k})\varepsilon\Omega_{\tau_i}$ for $1 \leq i \leq k$ and for all permutations τ: $(1,2,\ldots,k) \to (\tau_1,\tau_2,\ldots,\tau_k)$ and for all $\underline{\theta}\varepsilon\Omega$. Let $\Omega_0 = \Omega - \bigcup_{i=1}^{k} \Omega_i$ be a non-empty subset of Ω. Then a typical ranking problem can be reformulated as follows:

There are exactly k terminal decisions $D_1,D_2,\ldots,D_k$ where the decision D_i corresponds to deciding that $\underline{\theta}\varepsilon\Omega_i$. The <u>probability re-</u>

quirement is

$$(4.2.2) \qquad P_{\underline{\theta}}(D_i|\gamma) \geq P^* \;\; \forall\, \underline{\theta}\varepsilon\Omega_i \;\; (1 \leq i \leq k)$$

where P^* $(k^{-1},1)$ is preassigned. Note that Ω_0 is the usual indifference zone and $\bigcup_{i=1}^{k} \Omega_i$ is the usual preference zone.

Let $\Gamma(P^*)$ be the class of all symmetric rules which terminate with probability 1 and which satisfy (4.2.2). Let $\underline{\omega}_i\varepsilon\Omega_i$ $(0 \leq i \leq k)$ be defined in §4.1.2 and let $\Gamma'(P^*)$ be the class of symmetric rules which satisfy (4.1.3). Then the following theorem gives the sufficient conditions for a Bayes rule (against a certain symmetric prior b) to be U - minimax, first in the class of rules $\Gamma'(P^*)$ and then in a smaller class of rules $\Gamma(P^*)$.

Theorem 4.2.2: (i) Let γ^* be a Bayes rule against a symmetric prior $b\varepsilon(0,k^{-1})$ on $(\underline{\omega}_0,\underline{\omega}_1,\ldots,\underline{\omega}_k)$. Let $\alpha_1(\gamma^*) = \alpha^* = 1 - P^*$ $(1 \leq i \leq k)$. Suppose

$$(4.2.3a) \qquad E_{\underline{\omega}_0}(N|\gamma^*) = \sup_{\underline{\theta}\varepsilon\Omega} E_{\underline{\theta}}(N|\gamma^*).$$

Then,

$$(4.2.3b) \qquad E_{\underline{\omega}_0}(N|\gamma^*) = \inf_{\gamma\varepsilon\Gamma'(P^*)} \; \sup_{\underline{\theta}\varepsilon\Omega} E_{\underline{\theta}}(N|\gamma).$$

(ii) Further suppose that $\underline{\omega}_i$ $(1 \leq i \leq k)$ are LF-configurations for γ^*

in the sense that

$$(4.2.4a) \qquad P_{\underline{\omega}_i}(D_i|\gamma^*) = \inf_{\underline{\theta}\varepsilon\Omega_i} P_{\underline{\theta}}(D_i|\gamma^*) \quad \text{for} \quad 1 \leq i \leq k.$$

Then,

$$(4.2.4b) \qquad E_{\underline{\omega}_0}(N|\gamma^*) = \inf_{\gamma\varepsilon\Gamma(P^*)} \sup_{\underline{\theta}\varepsilon\Omega} E_{\underline{\theta}}(N|\gamma).$$

Proof: (i) The idea of the following proof is taken from Weiss [1962].

Suppose $E_{\underline{\omega}_0}(N|\gamma^*) = \sup_{\underline{\theta}\varepsilon\Omega} E_{\underline{\theta}}(N|\gamma^*)$ but there exists some $\gamma\varepsilon\Gamma'(P^*)$ such that

$$\sup_{\underline{\theta}\varepsilon\Omega} E_{\underline{\theta}}(N|\gamma) < \sup_{\underline{\theta}\varepsilon\Omega} E_{\underline{\theta}}(N|\gamma^*).$$

$$\Longrightarrow \qquad E_{\underline{\omega}_0}(N|\gamma) < E_{\underline{\omega}_0}(N|\gamma^*) = \sup_{\underline{\theta}\varepsilon\Omega} E_{\underline{\theta}}(N|\gamma^*).$$

$$\Longrightarrow (1-kb)E_{\underline{\omega}_0}(N|\gamma) + kb\alpha_1(\gamma) < (1-kb)E_{\underline{\omega}_0}(N|\gamma^*) + kb\alpha_1(\gamma^*).$$

$\Longrightarrow$ γ^* is not a Bayes rule against the prior b which is a contradiction. In the above we have used the fact that $\alpha_1(\gamma) \leq \alpha_1(\gamma^*) = \alpha^*$. This completes the proof.

(ii) The proof of part (ii) is straightforward.

The import of Theorems 4.2.1 and 4.2.2 is the following. In order to find a U-minimax ranking procedure, we need to know the least favorable parameter configurations (both from the point of view of making a correct decision and the expected total sample size) for the Bayes rule. Then if one constructs a Bayes rule by putting a symmetric prior (See Definition 4.2.2) on these parameter points, it will correspond to a U-minimax ranking rule for a certain $P^* \varepsilon (k^{-1}, 1)$.

Clearly this does not provide us with a good recipe for the construction of the desired rules. However by making use of the previous theorems we may be able to characterize the U-minimax ranking rules in the following manner.

We first guess least favorable parameter configurations for the Bayes rule. By assuming a certain symmetric prior on these parameter points we construct a Bayes rule of the desired nature (viz. two-stage, fully sequential etc.). Then we check to see if the initial guesses regarding the least favorable configuration are correct. If not, then the whole procedure is to be repeated with new guesses. If our initial guesses are correct then we have a U-minimax ranking rule, although, often it may not be implementable because we do not know the P*-value guaranteed by it. In any case, we shall know the structure of the U-minimax ranking rule.

We shall follow this method for the normal means problem of §2.1.4 in the following two sections.

§4.3 The normal means problem:

§4.3.1 A preview:

Consider the setup described in §2.1.1, 2.1.2 and 2.1.4. We consider the class of two-stage permanent elimination type rules described in §2.1.4 with the added invariance restrictions of symmetry (w.r.t. permutation of the labels of the population) and translation invariance. Note that rule R_1 satisfies both of these invariance conditions. We denote this class by $\mathscr{C}$. We denote by $\mathscr{C}(\delta^*,P^*)$, a subset of $\mathscr{C}$ which meets the probability requirement (1.1.1).

In the present case $\Omega_i = \{\underline{\mu}\epsilon\Omega | \mu_i \geq \mu_j + \delta^* \ \forall j \neq i\}$. We shall construct a two-stage Bayes rule by putting a prior probability b on $\underline{\omega}_i = (\mu_0,\ldots,\mu_0 + \delta^*,\ldots,\mu_0)$ where $\mu_0 + \delta^*$ is in the ith place, for $1 \leq i \leq k$ and a prior probability (1-kb) on $\underline{\omega}_0 = (\mu_0,\mu_0,\ldots,\mu_0)$ where μ_0 is arbitrary but fixed and may be taken to be zero in view of the translation invariant nature of the rule. In other words, our initial guesses regarding the least favorable configurations are <u>slippage configurations</u> for the probability of making a correct decision and <u>equal means configuration</u> for the expected total sample size.

After the construction of the Bayes rule we shall check to determine whether or not the conditions of Theorem 4.2.1 are satisfied. If they are, then we shall have succeeded in characterizing the U-minimax rules in the class $\mathscr{C}$. In the following we have been able to do this for k = 2; for k > 2 we have been able to obtain only a partial characterization of the Bayes rule.

§4.3.2 <u>Reduction by translation invariance</u>:

Let n_1 and n_2 be the first and the second stage sample sizes per population respectively and let $n = n_1 + n_2$. By the usual backward induction method n_1 and n_2 for the Bayes rule will be determined at the end so as to minimize the integrated (w.r.t. the prior chosen) risk. If we restrict attention to translation invariant symmetric rules then the terminal decision after the second stage must be a function of the maximal invariant sufficient statistic $(\overline{X}_{i_1}-\overline{X}_{i_t}, \overline{X}_{i_2}-\overline{X}_{i_t}, \ldots, \overline{X}_{i_{t-1}}-\overline{X}_{i_t})$ where $I = (i_1, i_2, \ldots, i_t)$ = the set of populations that entered the second stage ($2 \leq t \leq k$). Similarly sampling/terminal decision after the first stage must be a function of the maximal invariant sufficient statistic $(\overline{X}_1^{(1)} - \overline{X}_k^{(1)}, \ldots, \overline{X}_{k-1}^{(1)} - \overline{X}_k^{(1)})$, a consequence of Theorems 1 and 3, p. 216 and 220 of Lehmann [1959].

Let $Y_i^{(1)} = \overline{X}_i^{(1)} - \overline{X}_k^{(1)}$ for $1 \leq i \leq k-1$. Then the joint density of $(Y_1^{(1)}, \ldots, Y_{k-1}^{(1)})$ under different $\underline{\omega}_i$'s is given by (for the derivation see p. 304 of Fergusson [1967])

$$(4.3.1) \qquad g(y_1^{(1)}, \ldots, y_{k-1}^{(1)}; \underline{\omega}_i) = k^{-1/2}(2\pi\sigma^2/n_1)^{-\frac{k-1}{2}} \times$$

$$\exp\left\{-\frac{n_1}{2\sigma^2}\left[\sum_{j=1}^{k} y_j^{(1)^2} - k\overline{y}^{(1)^2} - 2\delta^*(y_i^{(1)} - \overline{y}^{(1)}) + \frac{\delta^{*2}(k-1)}{k}\right]\right\} \quad (1 \leq i \leq k)$$

$$(4.3.2) \qquad g(y_1^{(1)}, \ldots, y_{k-1}^{(1)}; \underline{\omega}_0) = k^{-1/2}(2\pi\sigma^2/n_1)^{-\frac{k-1}{2}} \times$$

$$\exp\left\{-\frac{n_1}{2\sigma^2}\left[\sum_{j=1}^{k} y_j^{(1)^2} - k\overline{y}^{(1)^2}\right]\right\}$$

where

$$(4.3.3) \qquad \overline{y}^{(1)} = \frac{1}{k}\sum_{j=1}^{k} y_j^{(1)} \quad \text{and} \quad y_k^{(1)} \equiv 0.$$

Similarly let $Y_j = \overline{X}_{i_j} - \overline{X}_{i_t}$ for $1 \leq j \leq t-1$. Then

$$(4.3.4) \qquad g(y_1,\ldots,y_{t-1};\ \underline{\omega}_{1_\ell}) = t^{-1/2}(2\pi\sigma^2/n)^{-\frac{t-1}{2}} \times$$

$$\exp\left\{-\frac{n}{2\sigma^2}\left[\sum_{j=1}^{t} y_j^2 - t\bar{y}^2 - 2\delta^*(y - \bar{y}) + \frac{\delta^{*2}(t-1)}{t}\right]\right\} \text{ for } i_\ell \varepsilon I \quad (1 \le \ell \le t).$$

and

$$(4.3.5) \qquad g(y_1,\ldots,y_{t-1};\ \underline{\omega}_0) = t^{-1/2}\ (2\pi\sigma^2/n)^{-\frac{t-1}{2}} \times$$

$$\exp\left\{-\frac{n}{2\sigma^2}\left[\sum_{j=1}^{t} y_j^2 - t\bar{y}^2\right]\right\}$$

where

$$(4.3.6) \qquad \bar{y} = \frac{1}{t}\sum_{j=1}^{t} y_j \quad \text{and} \quad y_t \equiv 0.$$

Having reduced the problem by translation invariance, we proceed to construct the Bayes rule by the backward induction method.

§4.3.3 Construction of the Bayes rule:

Terminal decision rule after the second stage:

Suppose using the Bayes rule, the set $I = (i_1, i_2, \ldots, i_t)$ $(2 \le t \le k)$ entered the second stage. In order to compute the terminal decision, we renormalize the priors such that

$$(4.3.7) \qquad b_i' = \frac{b}{1-(k-t)b} \;\; \forall i \varepsilon I, \;\; b_i' = 0 \;\forall i \notin I, \;\; b_0' = \frac{1-kb}{1-(k-t)b} \; .$$

We take the decision D_i if $i \varepsilon I$ and

$$(4.3.8) \quad b_i' g(y_{i_1},\ldots,y_{i_t} ; \underline{\omega}_i) > b_j' g(y_{i_1},\ldots,y_{i_t} ; \underline{\omega}_j) \quad \forall j \neq i,\ j \varepsilon I.$$

$$\Longleftrightarrow \qquad y_i > y_j \quad \forall j \neq i,\ j \varepsilon I.$$

$$\Longleftrightarrow \qquad \bar{x}_i > \bar{x}_j \quad \forall j \neq i,\ j \varepsilon I.$$

Thus the Bayes rule chooses the population i as "best" (takes decision D_i) if it enters the second stage and produces the highest overall sample mean among all the populations that entered the second stage.

<u>Sampling/terminal decision rule after the first stage</u>:

Let the decision d_I' correspond to selecting the subset $I \subseteq \{1,2,\ldots,k\}$ to be retained for sampling in the second stage. The Bayes rule selects that decision d_I' which has the minimum conditional posterior expected loss $CPEL(d_I')$ associated with it. Clearly if $I = \{i\}$ for some i, then we stop sampling and make the terminal decision D_i. Thus

$$(4.3.9) \qquad CPEL(d_{\{i\}}') = b_0^{(1)} kn_1 + \sum_{j=1, j\neq i}^{k} b_j^{(1)} \qquad (1 \leq i \leq k)$$

where $b_i^{(1)}$ = the posterior on $\underline{\omega}_i$ after observing the first stage outcome $(0 \leq i \leq k)$. Similarly,

$$\text{(4.3.10)} \qquad \text{CPEL}(d_I') = b_0^{(1)}(kn_1 + |I|n_2) + \sum_{i\varepsilon I} b_i^{(1)} \left\{1 - P_{\underline{\omega}_i}(d_i|\mathscr{F}_1)\right\} + \sum_{i \notin I} b_i^{(1)} \quad \text{for} \quad |I| \geq 2$$

where $\mathscr{F}_1$ = sigma algebra generated by $(\overline{X}_1^{(1)}, \ldots, \overline{X}_k^{(1)})$ and $|I|$ = cardinality of the set I. We shall now derive the Bayes first stage elimination rule through the following lemmas.

Lemma 4.3.1: For $|I| = t$ fixed $(2 \leq t \leq k)$, to minimize $\text{CPEL}(d_I')$ it is necessary and sufficient to maximize the expression

$$\text{(4.3.11)} \qquad \sum_{i\varepsilon I} e^{C_1\overline{x}_i^{(1)}} \int_{-\infty}^{\infty} \prod_{\substack{j\varepsilon I \\ j\neq i}} \Phi\left[x + C_2(\overline{x}_i^{(1)} - \overline{x}_j^{(1)}) + C_3\right] d\Phi(x)$$

over all sets I such that $|I| = t$. Here $C_1 = \delta^* n_1/\sigma^2$, $C_2 = pn^{1/2}/\sigma q^{1/2}$, $C_3 = \delta^* n_2^{1/2}/\sigma$, $p = n_1/(n_1 + n_2)$ and $q = 1 - p$.

Proof: Note first that the conditional distribution of $\overline{X}_i$ given $\mathscr{F}_1$ under $\underline{\omega}_i$ is normal with mean = $p\overline{x}_i^{(1)} + q\delta^*$ and variance = $q\sigma^2/n$. The conditional distribution of $\overline{X}_j$ for $j \neq i$, given $\mathscr{F}_1$ under $\underline{\omega}_i$ is normal with mean $p\overline{x}_i^{(1)}$ and variance = $q\sigma^2/n$. Now fix a set I with $|I| = t$ and note that

$$(4.3.12)\quad P_{\underline{\omega}_i}(d_i|\mathscr{F}_1)$$

$$= P_{\underline{\omega}_i}(\bar{X}_{(i)} > \bar{X}_{(j)} \quad \forall j\varepsilon I,\ j \neq i|\mathscr{F}_1)$$

$$= P_{\underline{\omega}_i}\left\{\frac{\bar{X}_{(i)}-p\bar{x}^{(1)}_{(i)}-q\delta^*}{\sigma(q/n)^{1/2}} + \frac{p(\bar{x}^{(1)}_{(i)}-\bar{x}^{(1)}_{(j)})+q\delta^*}{\sigma(q/n)^{1/2}} > \frac{\bar{X}_{(j)}-p\bar{x}^{(1)}_{(j)}}{\sigma(q/n)^{1/2}} \quad \forall j\varepsilon I, j\neq i|\mathscr{F}_1\right\}$$

$$= \int_{-\infty}^{\infty} \prod_{\substack{j\varepsilon I\\ j\neq i}} \Phi\left[x + \frac{p(\bar{x}^{(1)}_{(i)} - \bar{x}^{(1)}_{(j)}) + q\delta^*}{\sigma(q/n)^{1/2}}\right] d\Phi(x).$$

Substituting (4.3.12) in (4.3.10) we note that since $|I|$ is fixed, $\mathrm{CPEL}(d_I')$ is minimized iff

$$- b\sum_{i\varepsilon I} g(y_1^{(1)},\ldots,y_{k-1}^{(1)};\ \underline{\omega}_i)\int_{-\infty}^{\infty}\prod_{\substack{j\varepsilon I\\ j\neq i}} \Phi\left[x + \frac{p(\bar{x}^{(1)}_{(i)} - \bar{x}^{(1)}_{(j)}) + q\delta^*}{\sigma\sqrt{q/n}}\right]d\Phi(x)$$

is minimized. Substituting for $g(y_1^{(1)},\ldots,y_{k-1}^{(1)};\ \underline{\omega}_i)$ from (4.3.1) and cancelling the common terms (for all sets I s.t. $|I| = t$) we find that $\mathrm{CPEL}(d_I')$ is minimized iff

$$\sum_{i\varepsilon I} e^{C_1\bar{x}_i^{(1)}}\int_{-\infty}^{\infty}\prod_{j\varepsilon I, j\neq i}\Phi\left[x + C_2(\bar{x}_i^{(1)} - \bar{x}_j^{(1)}) + C_3\right]d\Phi(x)$$

is maximized over all sets I s.t. $|I| = t$. This completes the proof of the lemma.

Lemma 4.3.2: Suppose that if $|I| = t$ is fixed with $1 \leq t \leq k$, then the Bayes rule chooses that set of t populations which are associated with $\bar{x}^{(1)}_{[k]}, \ldots, \bar{x}^{(1)}_{[k-t-1]}$ with ties (if any) broken by randomization.

Proof: Case (i) $t = 1$: Here $\mathrm{CPEL}(d'_{\{i\}}) = b_0^{(1)} kn_1 + \sum_{\substack{j=1 \\ j \neq i}}^{k} b_j^{(1)}$

is clearly minimized by choosing i associated with $\max_{1 \leq j \leq k} b_j^{(1)}$, i.e. the one associated with $\max_{1 \leq j \leq k} \bar{x}_j^{(1)} = \bar{x}^{(1)}_{[k]}$.

Case (ii) $t > 1$: Fix a set I with $|I| = t > 1$. Let $\alpha \varepsilon I$ and denote $I' = I - \{\alpha\}$. Regard $\bar{x}_i^{(1)}$ for $i \varepsilon I'$ as fixed and denote

$$(4.3.13) \qquad \Psi(\bar{x}_\alpha^{(1)}) = \sum_{i \varepsilon I} e^{C_1 \bar{x}_i^{(1)}} \int_{-\infty}^{\infty} \prod_{\substack{j \varepsilon I \\ j \neq i}} \Phi\left[x + C_2(\bar{x}_i^{(1)} - \bar{x}_j^{(1)}) + C_3\right] d\Phi(x).$$

Our objective is to study how $\Psi(\bar{x}_\alpha^{(1)})$ is maximized. In the following we shall denote by

$$(4.3.14) \qquad \Phi_t(a_i,\ i\varepsilon I;\ \{\rho\}) = P\{X_i < a_i \quad \forall i \varepsilon I\}$$

where $(X_1, \ldots, X_t)$ have a t-variate standard normal distribution with common correlation $= \rho$. Now consider,

$$(4.3.15) \qquad \frac{\partial \Psi(\bar{x}_\alpha^{(1)})}{\partial \bar{x}_\alpha^{(1)}} = C_1 e^{C_1 \bar{x}_\alpha^{(1)}} \int_{-\infty}^{\infty} \prod_{j \varepsilon I'} \Phi\left[x + C_2(\bar{x}^{(1)} - \bar{x}_j^{(1)}) + C_3\right] d\Phi(x) +$$

$$e^{C_1\bar{x}_\alpha^{(1)}} \int_{-\infty}^{\infty} C_2 \sum_{i\varepsilon I'} \prod_{\substack{j\varepsilon I' \\ j\neq i}} \phi\Big[x + C_2(\bar{x}_\alpha^{(1)} - \bar{x}_j^{(1)}) + C_3\Big]\phi\Big[x + C_2(\bar{x}_\alpha^{(1)} - \bar{x}_j^{(1)})$$

$$+ C_3\Big]d\Phi(x) - \sum_{i\varepsilon I'} C_2 e^{C_1\bar{x}_i^{(1)}} \int_{-\infty}^{\infty} \prod_{\substack{j\varepsilon I' \\ j\neq i}} \Phi\Big[x + C_2(\bar{x}_i^{(1)} - \bar{x}_j^{(1)}) + C_3\Big]\phi\Big[x +$$

$$C_2(\bar{x}_i^{(1)} - \bar{x}_\alpha^{(1)}) + C_3\Big]d\Phi(x).$$

We interchange the order of summation and integration in the second term and combine it with the last term. We also use the identity

$$(4.3.16) \quad \int_{-\infty}^{\infty} \prod_{i=1}^{n} \Phi(a_i x + b_i)\phi(cx + d)d\Phi(x) = \frac{\phi\left[\frac{d}{(c^2+1)^{1/2}}\right]}{(c^2+1)^{1/2}} \times$$

$$\Phi_n\left[\frac{b_i(c^2+1) - a_i cd}{\{(c^2+1)(a_i^2+c^2+1)^{1/2}\}} \quad (1 \leq i \leq n); \Sigma\!\!\!/\right],$$

where

$$\Sigma\!\!\!/_{ij} = \begin{cases} 1 & \text{if } i = j \\ \dfrac{a_i a_j}{(a_i^2+c^2+1)^{1/2}(a_j^2+c^2+1)^{1/2}} & \text{if } i \neq j \end{cases} \qquad (1 \leq i, j \leq n)$$

We thus obtain from (4.3.15)

$$(4.3.17) \quad \frac{\partial\Psi(\bar{x}_\alpha^{(1)})}{\partial\bar{x}_\alpha^{(1)}} = C_1 e^{C_1\bar{x}_\alpha^{(1)}} \int_{-\infty}^{\infty} \prod_{j\varepsilon I'} \Phi\Big[x + C_2(\bar{x}_\alpha^{(1)} - \bar{x}_j^{(1)}) + C_3\Big]d\Phi(x) +$$

$$C_2 \sum_{i\varepsilon I'} \left\{ e^{C_1\overline{x}_\alpha^{(1)}} \int_{-\infty}^{\infty} \prod_{\substack{j\varepsilon I' \\ j\neq i}} \Phi\left[x + C_2(\overline{x}_\alpha^{(1)} - \overline{x}_j^{(1)}) + C_3\right] \times \right.$$

$$\phi\left[x + C_2(\overline{x}_\alpha^{(1)} - \overline{x}_j^{(1)}) + C_3\right] d\Phi(x) - e^{C_1\overline{x}_i^{(1)}} \int_{-\infty}^{\infty} \prod_{\substack{j\varepsilon I' \\ j\neq i}} \Phi\left[x + C_2(\overline{x}_i^{(1)} - \overline{x}_j^{(1)}) + \right.$$

$$\left. \left. C_3\right] \phi\left[x + C_2(\overline{x}_i^{(1)} - \overline{x}_\alpha^{(1)}) + C_3\right] d\Phi(x) \right\}$$

$$= C_1 e^{C_1\overline{x}_\alpha^{(1)}} \int_{-\infty}^{\infty} \prod_{j\varepsilon I'} \Phi\left[x + C_2(\overline{x}_\alpha^{(1)} - \overline{x}_j^{(1)}) + C_3\right] d\Phi(x) + \frac{C_2}{\sqrt{2}} \sum_{i\varepsilon I'} e^{C_1\overline{x}_\alpha^{(1)}} \times$$

$$\phi\left[\frac{C_2(\overline{x}_\alpha^{(1)} - \overline{x}_i^{(1)}) + C_3}{\sqrt{2}}\right] \Phi_{t-2}\left[\left(\frac{C_2(\overline{x}_\alpha^{(1)} + \overline{x}_i^{(1)} - 2\overline{x}_j^{(1)}) + C_3}{\sqrt{6}}\right) \forall \underset{j\neq i}{j\varepsilon I'} ; \{1/3\}\right]$$

$$- e^{C_1\overline{x}_i^{(1)}} \phi\left[\frac{-C_2(\overline{x}_\alpha^{(1)} - \overline{x}_i^{(1)}) + C_3}{\sqrt{2}}\right] \Phi_{t-2}\left[\left(\frac{C_2(\overline{x}_\alpha^{(1)} + \overline{x}_i^{(1)} - 2\overline{x}_j^{(1)}) + C_3}{\sqrt{6}}\right)\right.$$

$$\left. \forall \underset{j\neq i}{j\varepsilon I'} ; \{1/3\}\right]$$

$$= C_1 e^{C_1\overline{x}_\alpha^{(1)}} \int_{-\infty}^{\infty} \prod_{j\varepsilon I} \Phi\left[x + C_2(\overline{x}_\alpha^{(1)} - \overline{x}_j^{(1)}) + C_3\right] d\Phi(x) + \frac{C_2}{\sqrt{2}} \sum_{i\varepsilon I'}$$

$$\Phi_{t-2}\left[\left(\frac{C_2(\overline{x}_\alpha^{(1)} + \overline{x}_i^{(1)} - 2\overline{x}_j^{(1)}) + C_3}{\sqrt{6}}\right) \forall \underset{j\neq i}{j\varepsilon I'} ; \{1/3\}\right] . (2\pi)^{-1} \times$$

$$\exp\left\{ -\frac{1}{4}\left[C_2^2(\overline{x}_\alpha^{(1)} - \overline{x}_i^{(1)})^2 + C_3^2 - 2C_2C_3(\overline{x}_\alpha^{(1)} + \overline{x}_i^{(1)})\right]\right\} \times$$

$$\left\{\exp\left[(C_1 - C_2C_3)\bar{x}_\alpha^{(1)}\right] - \exp\left[(C_1 - C_2C_3)\bar{x}_i^{(1)}\right]\right\}$$

$$= C_1 e^{C_1\bar{x}_\alpha^{(1)}} \int_{-\infty}^{\infty} \prod_{j\varepsilon I'} \Phi\left[x + C_2(\bar{x}_\alpha^{(1)} - \bar{x}_j^{(1)}) + C_3\right] d\Phi(x)$$

$$\geq 0.$$

In the above we have used the fact that $C_1 - C_2C_3 = 0$. Thus (4.3.11) is nondecreasing in $\bar{x}_\alpha^{(1)}$ for every $\alpha\varepsilon I$. Therefore given $\bar{x}_1^{(1)},\ldots,\bar{x}_k^{(1)}$, and subject to $|I| = t > 1$ fixed, (4.3.11) is maximized by choosing the subset I to be the set of populations associated with $\bar{x}_{[k]}^{(1)},\ldots,\bar{x}_{[k-t+1]}^{(1)}$. Using Lemma 4.3.1 the proof of the lemma is now complete.

Now let us suppose that the populations are labelled so that $\bar{x}_1^{(1)} \geq \bar{x}_2^{(1)} \geq \cdots \geq \bar{x}_k^{(1)}$. Let

$$a_i = \bar{x}_1^{(1)} - \bar{x}_i^{(1)} \quad (1 \leq i \leq k) \tag{4.3.18}$$

so that $0 \equiv a_1 \leq a_2 \leq \cdots \leq a_k$. Then the following lemma tells us how the Bayes rule chooses t, the size of the subset I.

Lemma 4.3.3: The Bayes rule chooses t, the size of the subset to be retained for sampling in the second stage ($1 \leq t \leq k$), so as to maximize a function $\Psi(t)$ where

$$(4.3.19) \qquad \Psi(t) = \begin{cases} e^{C_1\bar{a}} & \text{for } t = 1 \\ -C_4 t + \sum_{i=1}^{t} e^{-C_1(a_i-\bar{a})} \int_{-\infty}^{\infty} \prod_{\substack{j=1 \\ j\neq i}}^{t} \\ \Phi\left[x + C_2(a_j-a_i) + C_3\right] d\Phi(x), \end{cases}$$

where $\bar{a} = \frac{1}{k}\sum_{i=1}^{k} a_i$ and $C_4 = \left(\frac{1-kb}{kb}\right)n_2 \times \exp\left\{\frac{n_1\delta^{*2}}{2\sigma^2}\left(\frac{k-1}{k}\right)\right\}$.

Proof: Using the previous lemma we have

$$(4.3.20) \qquad CPEL(d_I') = \begin{cases} b_0^{(1)} kn_1 + \sum_{i=2}^{k} b_i^{(1)} & \text{for } |I| = 1 \\ b_0^{(1)}(kn_1 + tn_2) + \sum_{i=1}^{k} b_i^{(1)} - \sum_{i=1}^{t} b_i^{(1)} \times \\ \int_{-\infty}^{\infty} \prod_{\substack{j=1 \\ j\neq i}}^{t} \Phi\left[x + C_2(a_j-a_i) + C_3\right] d\Phi(x) \\ & \text{for } |I| = t \geq 2. \end{cases}$$

We want to choose I so that $CPEL(d_I')$ is minimized. The desired result is obtained by substituting

$$(4.3.21)\qquad b_i^{(1)} = \frac{b_i g(y_1^{(1)},\ldots,y_{k-1}^{(1)};\ \underline{\omega}_i)}{\sum_{j=0}^{k} b_j g(y_1^{(1)},\ldots,y_{k-1}^{(1)};\ \underline{\omega}_j)}$$

where $b_0 = 1-kb$ and $b_i = b$ for $1 \le i \le k$, substituting for g from (4.3.1) and (4.3.2), cancelling the common terms and rearranging the resulting ones. This completes the proof of the lemma.

The previous lemma indicates that the structure of the Bayes rule is quite complex. Therefore we could not prove that for the Bayes rule, for $k > 2$, the slippage configuration is a LFC and that the maximum of the total expected sample size occurs at the EMC. For $k = 2$, the structure of the Bayes rule is simple and the corresponding proofs can be derived easily. This is done in the next section.

§4.4 Special case $k = 2$:

For $k = 2$, we have only two possible decisions after the first stage.

(i) Stop sampling and take decision d_1 or randomize between d_1 and d_2 if $\bar{x}_1^{(1)} = \bar{x}_2^{(1)}$. (Recall our labelling of the populations is such that $\bar{x}_1^{(1)} \ge \bar{x}_2^{(1)}$.)

(ii) Enter the second stage with $I = \{1,2\}$.

Let $a = \bar{x}_1^{(1)} - \bar{x}_2^{(1)}$. Then using Lemma 4.3.3, we stop sampling

$$\iff \Psi(1) > \Psi(2),$$

$$(4.4.1) \quad \iff e^{\frac{C_1 a}{2}} > -2C_4 + e^{\frac{C_1 a}{2}} \times \Phi\left(\frac{C_2 a + C_3}{\sqrt{2}}\right) + e^{-\frac{C_1 a}{2}} \times \Phi\left(\frac{-C_2 a + C_3}{\sqrt{2}}\right),$$

$$\iff 2C_4 > e^{-\frac{C_1 a}{2}} \times \Phi(-C_2' a + C_3') - e^{\frac{C_1 a}{2}} \times \Phi(-C_2' a - C_3'),$$

where $C_i' = \frac{C_i}{\sqrt{2}}$ for $i = 2,3$.

<u>Theorem 4.4.1</u>: For $k = 2$ and for $1/2 < P^* < 1$, after the first stage the Bayes rule γ^* decides to stop sampling and chooses d_1 if $a > a^*$, decides to continue sampling i.e. makes decision $d'_{\{1,2\}}$ if $a < a^*$, and randomizes in any manner between the above two decisions if $a = a^*$. Here a^* is the unique positive solution in a of the equation

$$(4.4.2) \qquad e^{-\frac{C_1 a}{2}} \times \Phi(-C_2' a + C_3') - e^{\frac{C_1 a}{2}} \times \Phi(-C_2' a - C_3') = 2C_4.$$

<u>Proof</u>: First assume that a positive solution to the equation (4.4.2) exists. We shall show that then such a solution is unique. Denote the ℓ.h.s. of (4.4.2) by $h(a)$. Then

$$\frac{\partial h(a)}{\partial a} = -\frac{C_1}{2}\left\{e^{-\frac{C_1 a}{2}} \times \Phi(-C_2' a + C_3') + e^{\frac{C_1 a}{2}} \times \Phi(-C_2' a - C_3')\right\} +$$

$$C_2'\left\{e^{\frac{C_1 a}{2}} \times \phi(-C_2' a - C_3') - e^{-\frac{C_1 a}{2}} \times \phi(-C_2' a + C_3')\right\}$$

$$= -\frac{C_1}{2}\left\{e^{-\frac{C_1 a}{2}} \times \Phi(-C_2' a + C_3') + e^{\frac{C_1 a}{2}} \times \Phi(-C_2' a - C_3')\right\}$$

$$+ \frac{C_2'}{\sqrt{2\pi}} e^{-\frac{1}{2}(C_2'a^2 + C_3'^2)} \times \left\{ e^{\frac{a}{2}(C_1 - 2C_2'C_3')} - e^{-\frac{a}{2}(C_1 - 2C_2'C_3')} \right\}$$

$$= -\frac{C_1}{2} \left\{ e^{-\frac{C_1 a}{2}} \times \Phi(-C_2'a + C_3') + e^{\frac{C_1 a}{2}} \times \Phi(-C_2'a - C_3') \right\} + 0$$

≤ 0 and < 0 if $C_1 > 0$.

In the above we have used the fact that $C_1 - 2C_2'C_3' = 0$. Thus h(a) is strictly decreasing in a if $C_1 > 0$ (i.e. $n_1 > 0$) and hence if a solution exists to (4.4.2) that solution is then unique.

Now we shall show that a positive solution always exists to (4.4.2). Note $\lim_{a\to\infty} h(a) = 0 < 2C_4$. Therefore a positive solution will not exist to (4.4.2) if at a = 0 we have,

$$\Phi(C_3') - \Phi(-C_3') < 2C_4. \tag{4.4.3}$$

Since h(a) is decreasing in a, this implies that $\Psi(1) > \Psi(2)$ for all $a \geq 0$ and hence the Bayes rule γ^* always terminates after the first stage. Therefore γ^* becomes equivalent to the single-stage rule R_0. Hence for any $\delta^* > 0$ and $1/2 < P^* < 1$; if γ^*, R_0 and $\hat{R}_1$ are designed to guarantee the same probability requirement (1.1.1) then

$$E_{EMC}(N|\gamma^*) = n_1 = n_0 > E_{EMC}(N|\hat{R}_1) \tag{4.4.4}$$

(We here regard n_1, n_0 as non-negative continuous variables.) But (4.4.4) contradicts Theorem 4.2.1. Hence a positive solution always exists to (4.4.2) for $1/2 < P^* < 1$ and the theorem is proved.

<u>Corollary 4.4.1</u>: For $k = 2$, the Bayes rule γ^* is U-minimax in the class $\mathscr{C}(\delta^*,P^*)$.

<u>Proof</u>: Note γ^* has the same structure as rule R_1. But for $k = 2$, the LF-configuration for rule R_1 is $\mu_{[2]} - \mu_{[1]} = \delta^*$ and its expected total sample size is maximized at the EM-configuration. Hence the same is true for γ^*. Therefore γ^* satisfies the conditions of Theorem 4.2.2 and γ^* is U-minimax in the class $\mathscr{C}(\delta^*,P^*)$.

<u>Corollary</u>: Rule $\hat{R}_1$ is U-minimax in the class $\mathscr{C}(\delta^*,P^*)$ for $k = 2$.

<u>Proof</u>: Rules γ^* and $\hat{R}_1$ is a Bayes rule for $k = 2$ and is hence optimal (i.e. U-minimax) in a wider class of rules.

We cannot make a similar conclusion about rule R_1 for $k > 2$. But it appears from Lemma 4.3.3 that the structure of the U-minimax rule is quite complex. Lemmas 4.3.2 and 4.3.3 also indicate that "good" rules (from the point of view of the U-minimax criterion) in the class $\mathscr{C}$ choose the size of the subset based on the first stage outcome and do not preassign it as in the case of certain two-stage rules proposed by Somerville [1954] and Fairweather [1968]. These lemmas also indicate that "good" rules include in the retained subset the populations associated with the largest first stage sample means.

CHAPTER 5

A THREE-STAGE PERMANENT ELIMINATION RULE FOR THE NORMAL MEANS PROBLEM (COMMON UNKNOWN VARIANCE)

§5.0 Introduction:

In the present chapter we consider the problem of selecting the largest mean from k normal populations when the common variance is unknown. A two-stage non-screening type of rule (RS_0) had been proposed by Bechhofer, Dunnett and Sobel [1954] for this problem. We propose a three-stage rule which has the desirable feature of screening.

In §5.1 we describe the problem and review the previous work done. We propose our three-stage rule (RS_1) in §5.2 and show that it guarantees the probability requirement (5.1.1). We also derive an expression for the expected total sample size of RS_1. In §5.3 we show that a two-stage rule with the screening feature (RS_2) can also be constructed for this problem. We also derive an expression for the expected total sample size of RS_2. In §5.4 using MC sampling techniques, we compare the performances of rules RS_1 and RS_2 with RS_0. It is observed that, even in the absence of any formal optimization, and only with a limited heuristic choice of the design constants, RS_1 performs better than RS_0 in terms of the expected total sample size needed to guarantee the same probability requirement. However rule RS_2 is found to perform rather poorely in comparison to RS_0.

§5.1 Preliminaries:

§5.1.1 Assumptions and notation:

Let $\Pi_1, \Pi_2, \ldots, \Pi_k$ be $k \geq 2$ normal populations with unknown means $\mu_1, \mu_2, \ldots, \mu_k$ ($-\infty < \mu_i < \infty$; $1 \leq i \leq k$) and a common unknown variance $\sigma^2 (0 < \sigma^2 < \infty)$. Let Ω denote the space of all parameter vectors $\underline{\omega} = (\mu_1, \ldots, \mu_k; \sigma^2)^t$ and let $\mu_{[1]} \leq \mu_{[2]} \leq \cdots \leq \mu_{[k]}$ denote the ordered values of the μ_i. Let $\delta_{ij} = \mu_{[i]} - \mu_{[j]}$ for $1 \leq i, j \leq k$. It is not known which population is associated with $\mu_{[i]}$ $(1 \leq i \leq k)$. The experimenter's goal is to select the "best" population where any of the populations associated with $\mu_{[k]}, \mu_{[k-1]}, \ldots, \mu_{[k-r+1]}$ is regarded as best if the r largest population means are equal $(1 \leq r \leq k)$. Any such selection is regarded as a correct selection (CS). The experimenter restricts consideration to only those rules R which satisfy the probability requirement

$$P_{\underline{\omega}}(CS|R) \geq P^* \quad \forall \underline{\omega} \varepsilon \Omega(\delta^*) \tag{5.1.1}$$

where P^* $(1/k < P^* < 1)$ and $\delta^* > 0$ are preassigned constants and $\Omega(\delta^*) = \{\underline{\omega} \varepsilon \Omega | \mu_{[k]} - \mu_{[k-1]} \geq \delta^*\}$.

§5.1.2 Previous work:

Bechhofer, Dunnett and Sobel [1954] solved this problem using a generalization of a two-stage test of Student's hypothesis due to Stein [1945]. A brief history of the development of two-stage rules for the case of unknown variance follows. Dantzig [1940] showed the non-existence of a single-stage test of Student's hypothesis whose power is independent of the variance. Stein developed a two-stage test with his property. Dudewicz [1971] later extended Dantzig's idea to show that for the problem of ranking normal means, a single-stage rule, whose PCS is independent of the variance, does not exist. Paulson [1964] has given a truncated sequential rule for the present problem. His rule has the desirable feature of screening at each stage. But the performance of this rule has not been studied so far. Srivastava [1966] and Robbins, Sobel and Starr [1968] have proposed fully sequential rules; however, their rules are shown to satisfy the probability requirement (5.1.1) only asymptotically as $\delta^* \to 0$.

A more general problem of unknown and unequal variances remained unsolved for a long time. Recently Ofosu [1973] claimed to have solved this problem. But Bechhofer [1974] has pointed out some crucial errors in Ofosu's main proof and Rinott [1974] has shown that Ofosu's final result is incorrect. Dudewicz and Dalal [1971] have provided a solution to this problem, but the performance of their rule remains to be studied.

For the reasons discussed in §1.1 it would be desirable to have a screening type two-or three-stage rule. We propose in the following section a <u>three-stage</u> rule which allows for elimination of the "non-contenders" at the end of the second stage and makes the final selection only from the non-eliminated populations at the end of the third stage. The first stage is simply used to obtain a preliminary estimate of the common variance.

The question of optimality in terms of developing a U-minimax rule analogous to rule $\hat{R}_1$ is not addressed in the present work. Minimax considerations would be extremely involved since the expected total sample size depends on the underlying variance which is unknown. Thus an "optimal" choice of the constants to implement the rule, which guarantees the specified probability requirement, would be a function of the unknown common variance. In the present work, we restrict ourselves to providing a screening-type of rule which guarantees the specified probability requirement.

§5.2 Three-stage rule RS_1 and its properties:

§5.2.1 Proposed rule RS_1:

The steps in rule RS_1 are as follows:

1. Take $n_1 \geq 2$ independent observations $X_{ij}^{(1)}$ from each Π_i $(1 \leq i \leq k,\ 1 \leq j \leq n_1)$. Compute the sample means $\bar{X}_i^{(1)}$ $(1 \leq i \leq k)$ and a pooled estimate of the variance

$$S_1^2 = \frac{1}{k(n_1-1)} \sum_{i=1}^{k} \sum_{j=1}^{n_1} (X_{ij}^{(1)} - \bar{X}_i^{(1)})^2. \tag{5.2.1}$$

2. Take additional N_2 independent observations $X_{ij}^{(2)}$ from each Π_i $(1 \leq i \leq k,\ 1 \leq j \leq N_2)$ where $N_2 = \bar{N}_2 - n_1$,

$$\bar{N}_2 = \max\left\{n_1 + 2,\ \left[2\left(\frac{S_1 h_1}{\delta^*}\right)^2\right] + 1\right\}, \tag{5.2.2}$$

$[x]$ is the largest integer $\leq x$ and h_1 is a positive constant defined in (5.2.7) below.

3. Compute

$$(5.2.3) \qquad \tilde{\bar{X}}_i^{(2)} = \frac{1}{N_2}\sum_{j=1}^{N_2} X_{ij}^{(2)}, \quad \bar{X}_i^{(2)} = \frac{n_1\bar{X}_i^{(1)} + N_2\tilde{\bar{X}}_i^{(2)}}{n_1 + N_2} \qquad (1 \le i \le k)$$

and the second stage pooled estimate of the variance

$$(5.2.4) \qquad S_2^2 = \frac{1}{k(N_2-1)}\sum_{i=1}^{k}\sum_{j=1}^{N_2}(X_{ij}^{(2)} - \tilde{\bar{X}}^{(2)})^2.$$

4. Choose a subset I of populations where

$$(5.2.5) \qquad I = \{i \mid \bar{X}_i^{(2)} \ge \max_{1\le j\le k} \bar{X}_j^{(2)} - \lambda S_1(2/\bar{N}_2)^{1/2}\}$$

and λ is a positive constant defined in (5.2.7) below.

5a. If I consists of a single population then stop sampling and assert that, that population is best.

5b. If I consists of more than one population, then take N_3 additional independent observations $X_{ij}^{(3)}$ $(1 \le j \le N_3)$ for $i\varepsilon I$ where $N_3 = \bar{N}_3 - n_1 - N_2$,

$$(5.2.6) \qquad \bar{N}_3 = \max\{n_1 + N_2, \left[2\left(\frac{S_2h_2}{\delta^*}\right)^2\right] + 1\},$$

and conditioned on N_2, h_2 is a positive constant defined in (5.2.8) below. Compute the overall sample means $\bar{X}_i^{(3)}$ for $i\varepsilon I$ and assert that the population associated with $\max_{i\varepsilon I} \bar{X}_i^{(3)}$ is best.

We now show how to choose h_1, h_2 and λ to satisfy the probability requirement (5.1.1).

§5.2.2 PCS of rule RS_1:

In the following $F_{\nu,p}(\cdot,\cdot,\ldots,\cdot,\{\rho\})$ denotes the cdf of a p-variate equicorrelated central t-distribution with ν degrees of freedom (d.f.) and the common correlation = ρ; for tables see Krishniah and Armitage [1966].

Theorem 5.2.1: If $h_1 > 0$, $\lambda >$ are chosen to satisfy

(5.2.7) $$F_{\nu_1,k-1}(h_1 + \lambda,\ldots,h_1 + \lambda; \{1/2\}) = \beta_1,$$

and conditioned on $N_2 = n_2, h_2 > 0$ is chosen to satisfy

(5.2.8) $$F_{\nu_2,k-1}(h_2,\ldots,h_2; \{1/2\}) = \beta_2,$$

where $\nu_1 = k(n_1-1)$ and β_1, β_2 are preassigned constants such that $P^* < \beta_1,\beta_2 < 1$ and

(5.2.9) $$\beta_1 + \beta_2 - 1 = P^*,$$

then rule R_1 guarantees the probability requirement (5.1.1).

Proof: In the following we denote by $\overline{X}^{(\ell)}_{(i)}$, the overall sample mean up-to the ℓth stage from the population having the mean $\mu_{[i]}$ ($\ell = 2,3$; $1 \leq i \leq k$). We have

$$(5.2.10)\qquad 1 - P_{\underline{\omega}}(CS|RS_1) = P_{\underline{\omega}}(\text{Incorrect Selection}|RS_1)$$

$$\leq P_{\underline{\omega}}\left[\overline{X}^{(2)}_{(k)} < \overline{X}^{(2)}_{(i)} - \lambda S_1(2/\overline{N}_2)^{1/2} \text{ for some } i \neq k\right] +$$

$$P_{\underline{\omega}}\left[\overline{X}^{(3)}_{(k)} < \overline{X}^{(3)}_{(i)} \text{ for some } i \neq k\right]$$

$$= 2 - P_{\underline{\omega}}\left[\overline{X}^{(2)}_{(k)} \geq \overline{X}^{(2)}_{(i)} - \lambda S_1(2/\overline{N}_2)^{1/2} \quad \forall i \neq k\right]$$

$$- P_{\underline{\omega}}\left[\overline{X}^{(3)}_{(k)} > \overline{X}^{(3)}_{(i)} \quad \forall i \neq k\right].$$

$$= 2 - P_{\underline{\omega}}\left[\left(\frac{\overline{X}^{(2)}_{(i)} - \overline{X}^{(2)}_{(k)} + \delta_{ki}}{S_1}\right)\left(\frac{\overline{N}_2}{2}\right)^{1/2} \leq \frac{\delta_{ki}}{S_1}\left(\frac{\overline{N}_2}{2}\right)^{1/2} + \lambda \quad \forall i \neq k\right]$$

$$- P_{\underline{\omega}}\left[\left(\frac{\overline{X}^{(3)}_{(i)} - \overline{X}^{(3)}_{(k)} + \delta_{ki}}{S_2}\right)\left(\frac{\overline{N}_3}{2}\right)^{1/2} \leq \frac{\delta_{ki}}{S_2}\left(\frac{\overline{N}_3}{2}\right)^{1/2} \quad \forall i \neq k\right]$$

$$\leq 2 - P_{\underline{\omega}}\left[T^{(2)}_i \leq \frac{\delta^*}{S_1}\left(\frac{\overline{N}_2}{2}\right)^{1/2} + \lambda \ \forall i \neq k\right]$$

$$- P_{\underline{\omega}}\left[T^{(3)}_i \leq \frac{\delta^*}{S_2}\left(\frac{\overline{N}_2}{2}\right)^{1/2} \quad \forall i \neq k\right],$$

for all $\underline{\omega}\varepsilon\Omega(\delta^*)$. In the above

$$(5.2.11)\quad T^{(\ell)}_i = \left(\frac{\overline{X}^{(\ell)}_{(i)} - \overline{X}^{(\ell)}_{(k)} + \delta_{ki}}{S_{\ell-1}}\right)\left(\frac{\overline{N}_\ell}{2}\right)^{1/2} \quad (1 \leq i \leq k-1;\ \ell = 2,3).$$

It is straightforward to check that $(T^{(2)}_1,\ldots,T^{(2)}_{k-1})$ and $(T^{(3)}_1,\ldots,T^{(3)}_{k-1})$ each have a (k-1)-variate central t-distribution with equal correlation $= \frac{1}{2}$; the former has $k(n_1-1)$ d.f. and the latter has $k(N_2-1)$ d.f. (random) associated with it. By using $h_1 \leq \frac{\delta^*}{S_1}\left(\frac{\overline{N}_2}{2}\right)^{1/2}$,

$h_2 \leq \frac{\delta^*}{S_2}\left(\frac{\bar{N}_3}{2}\right)^{1/2}$ from (5.2.2) and (5.2.6), we obtain from (5.2.10) and (5.2.11) that

$$1-P_{\underline{\omega}}(CS|RS_1) \leq 2 - P\left[T_i^{(2)} \leq h_1 + \lambda \ (1 \leq i \leq k-1)\right]$$

$$- \sum_{n_2=2}^{\infty} P\left[T_i^{(3)} \leq h_2 \ (1 \leq i \leq k-1)|N_2 = n_2\right] \times P_{\underline{\omega}}(N_2=n_2).$$

$$= 2 - F_{\nu_1,k-1}(h_1 + \lambda,\dots,h_1 + \lambda; \{1/2\})$$

$$- \sum_{n_2=2}^{\infty} F_{\nu_2,k-1}(h_2,\dots,h_2; \{1/2\}) \times P_{\underline{\omega}}(N_2=n_2).$$

$$= 2 - \beta_1 - \beta_2 \sum_{n_2=2}^{\infty} P_{\underline{\omega}}(N_2 = n_2).$$

$$= 2 - \beta_1 - \beta_2.$$

$$= 1 - P^*.$$

Therefore $P_{\underline{\omega}}(CS|RS_1) \geq P^* \ \forall \underline{\omega}\varepsilon\Omega(\delta^*)$. This proves the theorem.

The design constants for this rule are n_1, h_1, λ (and thus β_1) and β_2. We shall usually choose n_1 to be a small number ≥ 2. If k is large (when our rule would be most useful because of its screening feature) then a small value of n_1 should be sufficient to yield a reasonably large number of d.f. for S_1^2. We shall make some comments regarding the choice of h_1, λ and β_2 in §5.4. We now derive an expression for the expected total sample size for rule R_1.

§5.2.3 Expected total sample size of rule RS_1:

Lemma 5.2.1: (i) The probability distribution of N_2 is given by

$$(5.2.12)\quad P_{\underline{\omega}}(N_2 = n_2) = \begin{cases} 0 & \text{for } n_2 < 2, \\ G_{\nu_1}\{a_1(n_1+2)^{1/2}\} & \text{for } n_2 = 2, \\ G_{\nu_1}\{a_1(n_1+n_2)^{1/2}\} - G_{\nu_1}\{a_1(n_1+n_2-1)^{1/2}\} & \\ \qquad \text{for } n_2 > 2, & \end{cases}$$

where $G_\nu(\cdot)$ is the cdf of the random variable $(\chi^2_\nu/\nu)^{1/2}$, $\nu_1 = k(n_1 - 1)$ and $a_1 = \delta^*/2^{1/2}h_1\sigma$.

(ii) The probability distribution of N_3 conditioned on N_2 is given by

$$(5.2.13)\quad P_{\underline{\omega}}(N_3 = n_3 | N_2 = n_2) = \begin{cases} G_{\nu_2}\{a_1(n_1+n_2)^{1/2}\} & \text{for } n_3 = 0 \\ G_{\nu_2}\{a_2(n_1+n_2+n_3)^{1/2}\} & \\ -G_{\nu_2}\{a_2(n_1+n_2+n_3-1)^{1/2}\} & \text{for } n_3 > 0 \end{cases}$$

where $\nu_2 = k(n_2 - 1)$, $a_2 = \delta^*/2^{1/2}h_2\sigma$ and h_2 is obtained from (5.2.8).

Proof: The proof is straightforward and is omitted.

If N denotes the total sample size using procedure RS_1 then we have

(5.2.14) $$N = kn_1 + kN_2 + TN_3$$

where $T = |I|$ if $|I| \geq 2$ and $T = 0$ if $|I| = 1$ where I is defined by (5.2.5). The following theorem gives an expression for the expected total sample size.

Theorem 5.2.2: For any $\underline{\omega}\varepsilon\Omega$ we have,

(5.2.15) $$E_{\underline{\omega}}(N|RS_1) = kn_1 + k\left[2 + \sum_{n_2=3}^{\infty} [1 - G_{\nu_1}\{a_1(n_1 + n_2)^{1/2}\}]\right] +$$

$$\sum_{n_2=2}^{\infty}\Bigg[\sum_{i=1}^{k} \int_{C(n_2-1)}^{C(n_2)} \int_{-\infty}^{\infty} \Bigg\{ \prod_{\substack{j=1 \\ j\neq i}}^{k} \Phi\left(x + \frac{\delta_{ij}(n_1 + n_2)^{1/2}}{\sigma} + \lambda u\right) -$$

$$\prod_{\substack{j=1 \\ j\neq i}}^{k} \Phi\left(x + \frac{\delta_{ij}(n_1 + n_2)^{1/2}}{\sigma} - \lambda u\right)\Bigg\} d\Phi(x)dG_{\nu_1}(u)\Bigg] \times \Bigg[\sum_{n_3=0}^{\infty} [1 -$$

$$G_{\nu_2}\{a_2(n_1 + n_2 + n_3)^{1/2}\}]\Bigg],$$

where

(5.2.16) $$C(n_2) = \begin{cases} 0 & \text{for } n_2 = 1 \\ a_1(n_1 + n_2)^{1/2} & \text{for } n_2 > 1. \end{cases}$$

Proof:

$$(5.2.17)\qquad E_{\underline{\omega}}(N|RS_1) = kn_1 + \sum_{n_2=2}^{\infty} n_2 P_{\underline{\omega}}(N_2 = n_2) +$$

$$\sum_{n_2=2}^{\infty} \sum_{t=2}^{k} \sum_{n_3=1}^{\infty} tn_3 P_{\underline{\omega}}(N_2 = n_2, T = t, N_3 = n_3)$$

$$= kn_1 + kA_1 + kA_2$$

We shall now evaluate A_1 and A_2. Using Lemma 5.2.1 and the formula for expectation in terms of tail probabilities we have,

$$(5.2.18)\qquad A_1 = 2G_{\nu_1}\{a_1(n_1 + 2)^{1/2}\} + \sum_{n_2=3}^{\infty} n_2[G_{\nu_1}\{a_1(n_1 + n_2)^{1/2}\}$$

$$- G_{\nu_1}\{a_1(n_1 + n_2 - 1)^{1/2}\}]$$

$$= 2 + \sum_{n_2=3}^{\infty} [1 - G_{\nu_1}\{a_1(n_1 + n_2)^{1/2}\}].$$

We next consider,

$$(5.2.19)\qquad A_2 = \sum_{n_2=2}^{\infty} \sum_{t=2}^{k} tP_{\underline{\omega}}(N_2 = n_2, T = t) \sum_{n_3=1}^{\infty} n_3 P_{\underline{\omega}}(N_3 = n_3|N_2 = n_2, T = t)$$

$$= \sum_{n_2=2}^{\infty} \sum_{t=2}^{k} tP_{\underline{\omega}}(N_2 = n_2, T = t) \sum_{n_3=1}^{\infty} n_3 P_{\underline{\omega}}(N_3 = n_3|N_2 = n_2).$$

$$= \sum_{n_2=2}^{\infty} \sum_{t=2}^{k} t P_{\underline{\omega}}(N_2 = n_2,\ T = t) \sum_{n_3=1}^{\infty} n_3 [G_{\nu_2}\{a_2(n_1 + n_2 + n_3)^{1/2}\}$$

$$- G_{\nu_2}\{a_2(n_1 + n_2 + n_3 - 1)^{1/2}\}].$$

$$= \sum_{n_2=2}^{\infty} \left\{ \sum_{i=1}^{k} P_{\underline{\omega}}(\Pi_i \text{ is included in subset I, } N_2 = n_2) \right.$$

$$\left. - \sum_{i=1}^{k} P_{\underline{\omega}}(\Pi_i \text{ is } \underline{\text{alone}} \text{ included in subset I, } N_2 = n_2) \right\}$$

$$\times \sum_{n_3=0}^{\infty} [1 - G_{\nu_2}\{a_2(n_1 + n_2 + n_3)^{1/2}\}] \quad .$$

$$= \sum_{n_2=2}^{\infty} \left\{ \sum_{i=1}^{k} P_{\underline{\omega}}(\overline{X}^{(2)}_{(i)} \geq \overline{X}^{(2)}_{(j)} - \lambda S_1/\overline{N}_2^{1/2} \quad \forall j \neq i,\ N_2 = n_2) \right.$$

$$\left. - \sum_{i=1}^{k} P_{\underline{\omega}}(\overline{X}^{(2)}_{(i)} > \overline{X}^{(2)}_{(j)} + \lambda S_1/\overline{N}_2^{1/2} \quad \forall j \neq i,\ N_2 = n_2) \right\} \times$$

$$\sum_{n_3=0}^{\infty} [1 - G_{\nu_2}\{a_2(n_1 + n_2 + n_3)^{1/2}\}] \quad .$$

$$= \sum_{n_2=2}^{\infty} \left[\sum_{i=1}^{k} P_{\underline{\omega}} \left\{ \frac{(\overline{X}^{(2)}_{(i)} - \mu_{[i]})(n_1 + n_2)^{1/2}}{\sigma} + \frac{\delta_{ij}(n_1 + n_2)^{1/2}}{\sigma} \right. \right.$$

$$+ \frac{\lambda S_1}{\sigma} \geq \frac{(\overline{X}^{(2)}_{(j)} - \mu_{[j]})(n_1 + n_2)^{1/2}}{\sigma} \quad \forall j \neq i,\ C(n_2 - 1) \leq \frac{S_1}{\sigma}$$

$$< C(n_2)\Bigg\} - \sum_{i=1}^{k} P_{\underline{\omega}}\Bigg\{\frac{(\overline{X}^{(2)}_{(i)} - \mu_{[i]})(n_1 + n_2)^{1/2}}{\sigma} + \frac{\delta_{ij}(n_1 + n_2)^{1/2}}{\sigma}$$

$$- \frac{\lambda S_1}{\sigma} \geq \frac{(\overline{X}^{(2)}_{(j)} - \mu_{[j]})(n_1 + n_2)^{1/2}}{\sigma} \quad \forall j \neq i,\ C(n_2-1) \leq \frac{S_1}{\sigma} < C(n_2)\Bigg\}\Bigg]$$

$$\times \sum_{n_3=0}^{\infty} [1 - G_{\nu_1}\{a_2(n_1 + n_2 + n_3)^{1/2}\}] \quad .$$

$$= \sum_{n_2=2}^{\infty}\Bigg[\sum_{i=1}^{k} \int_{C(n_2-1)}^{C(n_2)} \int_{-\infty}^{\infty} \Bigg\{\prod_{\substack{j=1\\ j\neq i}}^{k} \Phi\Big[x + \frac{\delta_{ij}(n_1 + n_2)^{1/2}}{\sigma} + \lambda u\Big] -$$

$$\prod_{\substack{j=1\\ j\neq i}}^{k} \Phi\Big[x + \frac{\delta_{ij}(n_1 + n_2)^{1/2}}{\sigma} - \lambda u\Big]\Bigg\} d\Phi(x)dG_{\nu_1}(u)\Bigg] \times$$

$$\sum_{n_3=0}^{\infty} [1 - G_{\nu_2}\{a_2(n_1 + n_2 + n_3)^{1/2}\}] \quad .$$

In the above we have used Lemma 5.2.1, the fact that $N_2 = n_2$ iff $C(n_2 - 1) \leq S_1/\sigma < C(n_2)$ and the formula for expectation in terms of tail probabilities. The last step is obtained by conditioning on $X = (\overline{X}^{(2)}_{(i)} - \mu_{[i]})(n_1 + n_2)^{1/2}/\sigma = x$ and $S_1/\sigma = u$ and integrating w.r.t the density functions of X and S_1/σ. Combining (5.2.17), (5.2.18) and (5.2.19) we obtain the desired result.

<u>Lemma 5.2.2</u>: If σ^2 and all the design constants of rule RS_1 are regarded as kept fixed then $\sup_{\Omega} E_{\underline{\omega}}(N|RS_1)$ occurs at the EMC($\mu_{[1]}$ =

$\mu_{[2]} = \cdots = \mu_{[k]}$) and

$$(5.2.20) \quad \sup_{\Omega} E_{\underline{\omega}}(N|RS_1) = kn_1 + k\left\{2 + \sum_{n_2=3}^{\infty} [1 - G_{\nu_1}\{a_1(n_1 + n_2)^{1/2}\}]\right\} +$$

$$k \sum_{n_2=2}^{\infty}\left[\int_{C(n_2-1)}^{C(n_2)} \int_{-\infty}^{\infty} \left\{\Phi^{k-1}(x + \lambda u) - \Phi^{k-1}(x - \lambda u)\right\} d\Phi(x) dG_{\nu_1}(u)\right] \times$$

$$\sum_{n_3=0}^{\infty} [1 - G_{\nu_2}\{a_2(n_1 + n_2 + n_3)^{1/2}\}] \quad .$$

Proof: The result of this lemma can be proved by using the method of proof of Theorem 2.3.1.

We now consider the two-stage rule RS_2.

§5.3 Two-stage rule RS_2 and its properties:

Our purpose in this section is simply to illustrate that the probability requirement (5.1.1) can be guaranteed using a two-stage screening type of rule. We do not recommend its use in practice since its performance is found to be poor when compared to rule RS_0.

§5.3.1 Proposed rule RS_2:

The steps in rule RS_2 are as follows:

1. First step is the same as in case of RS_1.

2. Choose a subset I of populations where

$$(5.3.1) \qquad I = \{\overline{X}^{(1)}_{(i)} \geq \max_{1\leq j\leq k} \overline{X}^{(1)}_{j} - \lambda S_1 (2/n_1)^{1/2} + \delta^*\}$$

and λ is a positive constant defined in (5.3.4) below.

3a. If I consists of less than 2 populations then stop sampling and assert that the population associated with $\max_{1\leq j\leq k} \overline{X}^{(1)}_{j}$ is best.

3b. If I consists of atleast 2 populations then take N_2 additional independent observations $\overline{X}^{(2)}_{ij}$ $(1 \leq j \leq N_2)$ for $i\varepsilon I$ where $N_2 = \overline{N}_2 - n_1$ and

$$(5.3.2) \qquad \overline{N}_2 = \max\left\{\left[2\left(\frac{S_1 h}{\delta^*}\right)^2\right] + 1,\ n_1\right\},$$

$h \leq \lambda$ is a positive constant defined in (5.3.3) below.

4. Compute the overall sample means $\overline{X}^{(2)}_{i}$ for $i\varepsilon I$ and assert that the population associated with $\max_{i\varepsilon I} \overline{X}^{(2)}_{i}$ is best.

We now show how to choose h and λ to satisfy the probability requirement (5.1.1) and also give an expression for the expected total sample size for this rule.

§5.3.2 PCS and expected total sample size of rule RS_2:

Theorem 5.3.1: If $\lambda \geq h > 0$ are chosen to satisfy

(5.3.3) $$F_{\nu_1,k-1}(h,\ldots,h;\ \{1/2\}) = \beta_1,$$

(5.3.4) $$F_{\nu_1,k-1}(\lambda,\ldots,\lambda;\ \{1/2\}) = \beta_2,$$

where $\nu_1 = k(n_1 - 1)$, β_1 and β_2 are preassigned constants such that $P^* < \beta_1 \leq \beta_2 < 1$ and

(5.3.5) $$\beta_1 + \beta_2 - 1 = P^*,$$

then rule RS_2 guarantees the probability requirement (5.1.1).

Proof: We first note that if T denotes the cardinality of set I defined by (5.3.1) then using $\lambda \geq h$ we have,

(5.3.6) $$T = 0 \iff \lambda S_1(2/n_1)^{1/2} < \delta^* \implies 2(S_1 h/\delta^*)^2 < n_1 \implies$$

$$\left[2\left(\frac{S_1 h}{\delta^*}\right)^2\right] + 1 \leq n_1 \implies \overline{N}_2 = n_1.$$

Hence in the following probability calculations we may assume that if sampling is terminated after the first stage because $T = 0$, the termination is really due to the fact that $\overline{N}_2 = n_1$. We now proceed to prove the main result of this section. We follow the same notation as in Theorem 5.2.1.

(5.3.7) $$1 - P_{\underline{\omega}}(CS|RS_2) = P_{\underline{\omega}}(\text{Incorrect Selection}|RS_2)$$

$$\leq P_{\underline{\omega}}\{\overline{X}^{(1)}_{(k)} < \overline{X}^{(1)}_{(i)} - \lambda S_1(2/n_1)^{1/2} + \delta^* \text{ for some } i \neq k,$$

$$\lambda S_1(2/n_1)^{1/2} > \delta^*\} + P_{\underline{\omega}}\{\overline{X}^{(2)}_{(k)} < \overline{X}^{(2)}_{(i)} \text{ for some } i \neq k\}$$

$$\leq P_{\underline{\omega}}\left\{\left(\frac{\overline{X}^{(1)}_{(i)} - \overline{X}^{(1)}_{(k)} + \delta_{ki}}{S_1}\right)\left(\frac{n_1}{2}\right)^{1/2} > \left(\frac{\delta_{ki} - \delta^*}{S_1}\right)\left(\frac{n_1}{2}\right)^{1/2} + \lambda \text{ for}\right.$$

$$\left.\text{some } i \neq k\right\} + P_{\underline{\omega}}\left\{\left(\frac{\overline{X}^{(2)}_{(i)} - \overline{X}^{(2)}_{(k)} + \delta_{ki}}{S_1}\right)\left(\frac{\overline{N}_2}{2}\right)^{1/2} >\right.$$

$$\left.\left(\frac{\delta_{ki}}{S_1}\right)\left(\frac{\overline{N}_2}{2}\right)^{1/2} \text{ for some } i \neq k\right\}$$

$$\leq 2 - P_{\underline{\omega}}\left\{T_i^{(1)} \leq \lambda \ \forall i \neq k\right\} - P_{\underline{\omega}}\left\{T_i^{(2)} \leq \frac{\delta^*}{S_1}\ \frac{\overline{N}_2}{2}^{\ 1/2} \ \forall i \neq k\right\}$$

for all $\underline{\omega} \varepsilon \Omega(\delta^*)$.

Now we note that $(T_1^{(\ell)}, \ldots, T_{k-1}^{(\ell)})$ have a (k - 1) - variate central t-distribution with common correlation = 1/2 and the d.f. = $k(n_1 - 1)$ for $\ell = 1,2$ and that $h \leq \frac{\delta^*}{S_1}\left(\frac{\overline{N}_2}{2}\right)^{1/2}$. Thus we obtain from (5.3.7)

(5.3.8) $$1 - P_{\underline{\omega}}(CS|RS_2) \leq 2 - F_{\nu_1,k-1}(\lambda,\ldots,\lambda; \{1/2\})$$

$$- F_{\nu_1,k-1}(h,\ldots,h; \{1/2\})$$

$$= 2 - \beta_1 - \beta_2$$

$$= 1 - P^*.$$

Therefore $P_{\underline{\omega}}(CS|RS_2) \geq P^* \; \forall \underline{\omega}\varepsilon\Omega(\delta^*)$ and the theorem is proved.

We may regard n_1, h and λ as the design constants for rule RS_2. As discussed earlier, n_1 would be chosen to be a small number but sufficiently large to yield a reasonably large number of d.f. for S_1^2.

We now state without proof the following result concerning the expected total sample size for rule RS_2.

Theorem 5.3.2: For any $\underline{\omega}\varepsilon\Omega$ we have

$$(5.3.9) \quad E_{\underline{\omega}}(N|RS_2) = kn_1 + \sum_{n=n_1+1}^{\infty} (n-n_1) \sum_{i=1}^{k} \int_{a(n-1)^{1/2}}^{an^{1/2}} \int_{-\infty}^{\infty} \Bigg\{ \prod_{\substack{j=1\\ j\neq i}}^{k} \Phi\left[x + \frac{(\delta_{ij} - \delta^*)n_1^{1/2}}{\sigma} + \lambda u\right] - \prod_{\substack{j=1\\ j\neq i}}^{k} \Phi\left[x + \frac{(\delta_{ij} - \delta^*)n_1^{1/2}}{\sigma} - \lambda u\right]\Bigg\} d\Phi(x)\, dG_{\nu_1}(u)$$

where $a = \delta^*/2^{1/2}h\sigma$. Further if σ^2 and all the other design constants of rule RS_2 are regarded as kept fixed then $\sup_{\Omega} E_{\underline{\omega}}(N|RS_2)$ occurs at the EMC.

Next we shall discuss the MC sampling results for the three rules RS_0, RS_1 and RS_2.

§5.4 Monte Carlo sampling studies:

We shall use the expected total sample sizes of rules RS_0, RS_1 and RS_2 in the EMC as measures of their performances. Whereas $E_{\underline{\omega}}(N|RS_0)$ remains unaffected by a change in the μ_i-configuration, we have noted, respectively in Lemma 5.3.2 and Theorem 5.2.2 that $E_{\underline{\omega}}(N|RS_1)$ and

$E_{\underline{\omega}}(N|RS_2)$ are maximized over Ω at the EMC. Thus for rules RS_1 and RS_2 the comparison with rule RS_0 is carried out in the "worst" parameter configuration. We also study $E_{\underline{\omega}}(N|RS_1)$ and $E_{\underline{\omega}}(N|RS_2)$ in the least favorable configuration (LFC) i.e. $\mu_{[1]} = \cdots = \mu_{[k-1]} = \mu_{[k]} - \delta^*$. The expression for the expected total sample size of rule RS_0 is given by

$$(5.4.1) \qquad E_{\underline{\omega}}(N|RS_0) = kn_1 + \sum_{n=n_1+1}^{\infty} (n-n_1)\left[G_{\nu_1}\{a_0 n^{1/2}\} - G_{\nu_1}\{a_0(n-1)^{1/2}\}\right]$$

where $a_0 = \delta^*/2^{1/2}h_0\sigma$, n_1 = the first stage sample size, $\nu_1 = k(n_1-1)$ and h_0 solves the equation

$$(5.4.2) \qquad F_{\nu_1,k-1}(h_0,\ldots,h_0;\ \{1/2\}) = P^*.$$

Due to the complicated nature of expressions (5.2.15), (5.3.9) and (5.4.1), a direct analytical comparison appears to be difficult. Therefore we conducted MC sampling studies for the three rules for $k = 4(2)\,10$ and $P^* = 0.90$ and 0.95. For each (k,P^*) combination we used two values of δ^*. The underlying value of σ^2 was kept fixed throughout to be 1. In each case, the first stage sample size n_1 was chosen to be the same for all the three rules.

The choice of design constants β_1 and β_2 for both the rules RS_1 and RS_2 was limited to $\beta_1 = \beta_2 = (1 + P^*)/2$ because tables for multivariate t-distribution with arbitrary percentage points are not yet available. Thus we had no flexibility in the choice of design constants for rule RS_2. We had some flexibility in the choice of design constants for rule RS_1 because we could choose h_1 and λ subject to the restriction that (5.2.7) is satisfied with $\beta_1 = (1 + P^*)/2$. If optimality is defined in terms of (say) minimizing $\sup_{\Omega} E_{\underline{\omega}}(N|R)$, it should be noted that the optimal choice of design constants is unknown to the experimenter since he does not know σ.

For rule RS_1, we chose $h_1 > \lambda$ ($h_1 \cong 2\lambda$ to 3λ) since a small λ-value results in a small subset size T which is a major factor in determining the total sample size. We would recommend higher values of h_1 for larger values of k, P^* and δ^* and vice versa although no precise recommendation can be given at this stage.

Our results for the MC sampling studies are given in Table 5.4.1. For each run 2000 experiments were conducted. The numbers inside the round brackets are the standard errors of the corresponding estimates. An inspection of the results reveals that in almost all cases $E_{LFC}(N)$ and $E_{EMC}(N)$ associated with rule RS_1 are smaller than $E_{\underline{\omega}}(N)$ (same for all $\underline{\omega}\varepsilon\Omega$) associated with rule RS_0. Thus RS_1 provides a distinct improvement over RS_0. It should be noted that in practice the μ_i-values would be spaced somewhat far apart and for such configurations substantial gains are possible by using rule RS_1. For the same reason, although rule RS_2 is found to perform poorly in the LFC and the EMC configurations, it is possible that RS_2 performs better than RS_0 when the μ_i-values are not close to each other.

Table 5.4.1 also gives estimates of probabilities of correct selection in the LFC for all the three rules. These numbers give some idea about the overprotection in P(CS), i.e. excess over the guaranteed P^*, afforded by each rule. In all the cases, rule RS_1 is

Table 5.4.1

Monte Carlo Sampling Results for RS_0, RS_1 and RS_2

($P^* = 0.95$)

k	δ^*	n_1	RS_0		RS_1					RS_2		
			$P_{LFC}(CS)$	$E_{\underline{\omega}}(N/k)$	h_1	λ	$P_{LFC}(CS)$	$E_{LFC}(N/k)$	$E_{EMC}(N/k)$	$P_{LFC}(CS)$	$E_{LFC}(N/k)$	$E_{EMC}(N/k)$
10	1.0	4	.954 (.005)	13.31 (.08)	2.20	.65	.971 (.004)	10.71 (.06)	12.16 (.06)	.955 (.005)	10.72 (.12)	12.85 (.12)
10	.25	4	.958 (.004)	205.47 (1.19)	2.04	.81	.966 (.004)	149.86 (.80)	179.79 (.85)	.952 (.005)	245.48 (1.72)	246.54 (1.71)
8	.8	5	.952 (.005)	18.97 (.10)	2.02	.74	.966 (.004)	14.50 (.08)	17.19 (.08)	.957 (.005)	16.61 (.18)	19.37 (.17)
8	.2	5	.951 (.005)	295.95 (.165)	1.96	.80	.959 (.004)	219.67 (1.21)	267.13 (1.29)	.950 (.005)	361.77 (2.40)	362.50 (2.38)
6	.6	6	.950 (.005)	30.44 (.17)	1.89	.77	.970 (.004)	23.44 (.14)	28.86 (.15)	.962 (.004)	30.27 (.31)	33.65 (.29)
6	.15	6	.946 (.005)	479.63 (2.78)	1.83	.825	.967 (.004)	361.34 (2.23)	452.57 (2.40)	.954 (.005)	601.63 (4.06)	601.65 (4.04)
4	.5	9	.954 (.005)	36.86 (.20)	1.70	.76	.967 (.004)	28.75 (.20)	36.33 (.24)	.954 (.005)	35.57 (.40)	41.03 (.36)
4	.125	9	.948 (.005)	582.26 (3.25)	1.65	.81	.960 (.004)	448.52 (3.32)	572.42 (3.53)	.950 (.005)	735.94 (4.94)	741.79 (4.83)

Table 5.4.2

Monte Carlo Sampling Results for RS_0, RS_1 and RS_2

($P^* = 0.90$)

k	δ^*	n_1	RS_0		RS_1					RS_2		
			$P_{LFC}(CS)$	$E_{\underline{\omega}}(N/k)$	h_1	λ	$P_{LFC}(CS)$	$E_{LFC}(N/k)$	$E_{EMC}(N/k)$	$P_{LFC}(CS)$	$E_{LFC}(N/k)$	$E_{EMC}(N/k)$
10	.8	4	.907 (.007)	15.37 (.09)	1.83	.71	.930 (.006)	12.28 (.07)	14.15 (.07)	.917 (.006)	13.63 (.15)	15.37 (.15)
10	.2	4	.909 (.006)	238.60 (1.38)	1.78	.76	.930 (.006)	185.58 (1.03)	216.07 (1.07)	.910 (.006)	295.53 (2.22)	296.94 (2.20)
8	.6	5	.903 (.007)	24.81 (.14)	1.71	.73	.930 (.006)	19.63 (.11)	22.94 (.12)	.912 (.006)	23.67 (.25)	26.05 (.24)
8	.15	5	.903 (.007)	389.55 (2.17)	1.66	.78	.927 (.006)	300.78 (1.75)	354.98 (1.87)	.903 (.007)	487.46 (3.48)	488.77 (3.45)
6	.5	6	.897 (.007)	31.64 (.18)	1.58	.75	.936 (.005)	25.66 (.16)	30. 3 (.17)	.911 (.006)	32.43 (.34)	35.17 (.33)
6	.125	6	.887 (.007)	498.73 (2.89)	1.54	.79	.931 (.006)	396.96 (2.65)	479.49 (2.71)	.904 (.007)	644.48 (4.73)	643.69 (4.72)
4	.4	9	.903 (.007)	39.81 (.22)	1.41	.73	.926 (.006)	34.24 (.26)	41.12 (.28)	.914 (.006)	41.99 (.47)	46.50 (.44)
4	.1	9	.898 (.007)	629.40 (3.51)	1.37	.77	.920 (.006)	526.95 (4.28)	643.33 (4.55)	.904 (.007)	841.53 (6.24)	846.89 (6.03)

found to provide the greatest overprotection and in most cases rule RS_2 provides greater overprotection than RS_0.

In conclusion, rule RS_1 is somewhat complicated to implement but provides a distinct improvement over rule RS_0 with a reasonable choice of the design constants h_1 and λ. Rule RS_2 is more simple to implement and in the MC sampling studies it is found in most cases to be inferior to RS_0 at the LFC and the EMC. Its use may be recommended in practice only when there is a reason to believe apriori that the μ_i-values are spaced far apart.

CHAPTER 6

SUGGESTIONS FOR FUTURE RESEARCH

In the present chapter we mention some important unsolved problems in the area of two-stage ranking rules and offer certain suggestions regarding the possible directions in which the present work can be extended.

1. The most important unsolved problem is to show that the slippage configuration $\mu_{[1]} = \cdots = \mu_{[k-1]} = \mu_{[k]} - \delta^*$ is a LF-configuration for the two-stage rule R_1 for $k > 2$. We have been able to obtain only a partial solution to this problem in Chapter 2.

2. More efficient algorithms are needed to evaluate the complicated multivariate integral expressions associated with the PCS of different types of screening rules and to obtain solutions to the constrained optimization problems associated with different design criteria.

3. As we pointed out in §3.5.2, it would be desirable to develop a generalized come-back type two-stage rule. This rule would assign a probability p_i $(0 \leq p_i \leq 1)$ for the population Π_i to be retained in the subset which is sampled in the second stage; p_i would be some function of the first stage means. It appears that an exact analysis of such rules would be extremely involved. Thus the research may have to be directed towards the development of somewhat conservative rules based

on certain efficient lower bounds on the PCS of the original rules.

4. We have carried out analytical studies of the asymptotic ($P^* \to 1$; δ^*,k fixed) behavior of our two-stage rule R_1 with U-minimax and R-minimax criteria. Since the screening type rules become more effective compared to the single-stage rule R_0 as k is increased, it would be interesting to study the asymptotic behavior of rules $\hat{R}_1$ and $\tilde{R}_1$ as $k \to \infty$ with δ^*,P^* kept fixed.

5. The extension of the work done for the normal means problem to the problem of normal variances is immediate and clear. We have completed this work. Instead of including it as a part of the present thesis, we plan to issue it as a separate report.

6. In the case of the normal means problem with a common unknown variance, it would be desirable to conduct more extensive Monte Carlo studies on rules RS_0, RS_1 and RS_2 to determine for which (δ^*,P^*,k)-values one rule is preferable to the others.

7. Finally, the philosophy of two-stage screening rules may be applied to the selection problems associated with other distributions, e.g., selecting the Bernoulli population associated with the highest probability of success.

APPENDIX A1

DETAILS OF CONSTRUCTION OF TABLES 2.6.1, 2.6.2 and 3.2.1

First we describe the method of solution of constrained continuous optimization problems (2.4.4), (2.4.5) and (3.1.4). The first two problems are associated with finding the constants necessary to implement our two-stage rule R_1 using the U-minimax and the R-minimax criterion for k = 2. The third problem is associated with finding the constants necessary to implement the conservative two-stage rule $\overline{R}_1$ using the U-minimax criterion for k > 2. The method of solution is the same in all the cases.

First we find a "reasonably good" discrete optimal solution. Then we use that solution as an initial guess in the program using a modified version of the steepest descent method to solve the continuous nonlinear programming (NLP) problem.

To find the discrete optimal solution we fix a $\delta^* > 0$ for any specified P* and take $\sigma = 1$. Using Table I in Bechhofer [1954] we then compute n_0, the single-stage sample size required. We use the inequality $n = n_1 + n_2 \geq n_0$ to restrict the region of search in the (n_1,n_2,h) space. We choose δ^* small enough and n_0 large enough so that the discrete optimal solution in the region $n_1 + n_2 \geq n_0$ would be fairly close to the continuous optimal solution. For each fixed (n_1,n_2) we solve for h setting the corresponding PCS expressions equal to P* and compute the associated objective function. We systematically vary n,n_1 and n_2 so as to move rapidly in the direction of the minimum of the objective function. Let $(\hat{n}_1, \hat{n}_2, \hat{h})$ be the minimizing solution in the case of rule R_1 (with obvious modifications in the notation in case of rule $\overline{R}_1$). We then take $d_1 = \delta^* \hat{n}_1^{1/2}$, $d_2 = \delta^* \hat{n}_2^{1/2}$ and $c = \hat{h}\hat{n}_1^{1/2}$ as the initial guesses for the NLP algorithm. If the initial guesses are fairly close to the continuous optimum solution $(\hat{c}, \hat{d}_1, \hat{d}_2)$, then the algorithm is found to converge in less than 15 iterations for the values of the step-size and the convergence criterion chosen. The convergence criterion is fixed throughout to be 1×10^{-5} but the step-size is changed to suit each situation depending on the rate of convergence of the solution. Although it may be possible to specify an even smaller convergence criterion and thereby achieve a better minimum, the possible improvement in the value of the objective function was deemed to be minimal and the additional computing cost involved excessively large. Thus we do not claim that our solutions represent the absolute optima but they are reasonably close to the optima.

Now we describe the details of computations. All the computations were carried out in double precision arithmetic on Cornell University's IBM 360/65 and IBM 370/168 machines. For computing the standard normal cdf, we used the Algorithm 304 of Hill and Joyce [1967]. The integrals are evaluated using the Romberg method with 2^{10} being the upper limit on the number of subdivisions of the interval of integration. For $k > 2$, we need to evaluate the integrals of the type $\int_{-\infty}^{\infty} \Phi^{k-1}(x + a)d\Phi(x)$. We replace this by $\int_{-6}^{6} \Phi^{k-1}(x + a)d\Phi(x)$. The error committed thus can easily seen to be bounded above by

$$\int_{|x|>6} \Phi^{k-1}(x + a)d\Phi(x) \leq 2.0253 \times 10^{-9} \quad \text{for all} \quad a.$$

since c, d_1 and d_2 are chosen to satisfy the equality for P*, the above error is in the conservative direction. However the objective function $d_1^2 + d_2^2 \times \int_{-\infty}^{\infty} \{\Phi^{k-1}(x + c) - \Phi^{k-1}(x - c)\}d\Phi(x)$ may be underestimated by the corresponding amount.

The tabulated values are rounded off in the fifth decimal place and are correct up to the fourth decimal place. All the relevant program listings are given in Appendix A2.

APPENDIX A2

COMPUTER PROGRAMS

The computer programs written in FORTRAN IV, are displayed in the present section. We describe below the tasks performed by each program, along with an explanation of the input and the output variables. The following symbols are common to most of the programs and hence are explained only at the outset: PSTAR = P*, K = k, DEL = δ^*/σ, R2PI = $1/\sqrt{2\pi}$, NUMSIG, NUM = the number of significant numerical digits, MAXIT = the number of maximum divisions of the range of integration required for the Romberg method of numerical integration, MAX = the number of maximum iterations for solving a system of nonlinear simultaneous equations, A = the lower limit of integration and B = the upper limit of integration.

Program 1: This program is used to identify, for $k > 2$, the discrete optimal solution $(\hat{n}_1, \hat{n}_2, \hat{c})$ to the optimization problem (3.1.4) where we restrict n_1 and n_2 to take nonnegative integer values in the design variables $d_1 = \delta^* n_1^{1/2}/\sigma$ and $d_2 = \delta^* n_2^{1/2}/\sigma$. For an explanation of the method of obtaining $(\hat{n}_1, \hat{n}_2, \hat{c})$ by a systematic enumeration, refer to Appendix A1. The meanings of the symbols used in the program for the input and the output variables are as follows: NI and NF = the lower and the upper limits, respectively, of the range of $(n_1 + n_2)$ values over which the search for the optimum is to be made, P1 and P2 = the lower and the upper limits, respectively, of the range of p-values (for each ($n_1 + n_2$)) over which the search for the optimum is to be made, X(1) = the initial guess for the solution in c of the probability requirement constraint in (3.1.4) regarded as an equality, AMBDA = the corresponding solution in c of the resulting equation, EN = the value of the corresponding objective function, N1 = n_1 and N2 = n_2. In addition PSTAR, K and DEL are also input variables.

The corresponding programs for k = 2 to find the exact discrete optimal solutions associated with $\hat{R}_1$ and $\tilde{R}_1$ have not been displayed here.

Program 2 (SUBROUTINE CSD): This subroutine solves a nonlinear programming problem by utilizing an algorithm based on a modified steepest descent method. This subroutine requires the user to supply a calling program where certain constants are specified and certain subroutines are provided. We thank Professor Bartel of the Department of Mechanical Engineering, Cornell University for providing us with this program.

Program 3: This program calls the subroutine CSD and provides it with the necessary input data and the subroutines necessary to solve the NLP problem (3.1.4). The corresponding programs to solve (2.4.4) and (2.4.5) for k = 2 have not been displayed here. The following input data are provided: N = the number of design variables, M = the number of inequality constraints; K = max(M,N), NK = k, PSTAR, ETA = the step size and EX = the convergence criterion. In addition, the initial guesses U(1), U(2) and U(3) for the final solution values $\hat{\bar{d}}_1$, $\hat{\bar{d}}_2$ and $\hat{\bar{c}}'$, respectively, are also input variables.

The program also provides the following five subroutines.

1) OBJ: Returns the value of the objective function in (3.1.4) for given values of the design variables.

2) PHI: Returns the values of the constraints in (3.1.4) where the constraints are rewritten in the form $P^* + 1 - \int_{-\infty}^{\infty} \Phi^{k-1}(x + d_1 + c)d\Phi(x) - \int_{-\infty}^{\infty} \Phi^{k-1}(x + (d_1^2 + d_2^2)^{1/2})d\Phi(x) \leq 0$, $-c \leq 0$, $-d_1 \leq 0$, $-d_2 \leq 0$.

3) DPHI: Returns the matrix of values of the first derivatives of the four constraints w.r.t. the three design variables.

4) DF: Returns the vector of values of the first derivatives of the objective function w.r.t. the three design variables.

5) WTI: Supplies the inverse of the weighting matrix which we have taken to be the identity matrix. In some cases, we also used the hessian matrix of the objective function as a weighting matrix.

Program 4 (REAL FUNCTION NORMAL): This program computes the value of the standard normal cdf for a given value of the argument. This program is based on the Algorithm 304 of Hill & Joyce [1967].

Program 5 (SUBROUTINE DEFINT): This program is a part of the Cornell Computing Library. It computes the value of the definite integral $\int_A^B F(x)dx$ by means of Romberg method of numerical integration. A separate user written subroutine YFUN is needed to evaluate the value of the integrand F(x).

Program 6 (SUBROUTINE SYSTEM): This program is also a part of the

Cornell Computing Library and it solves a system of N simultaneous nonlinear equations. A separate user written subroutine AUXFCN is needed to evaluate the values of the ℓ.h.s. of the N equations.

Program 7: This program simulates the operations of rule R_1 and R_2. The input data consists of PSTAR, K, DEL, AN1 = $\hat{d}_1$, AN_2 = $\hat{d}_2$, H = $\hat{c}$, N = the total number of experiments in one simulation run and NR(I,J) = the seed for generating the pseudo-random variable, namely the Jth stage sample mean from the Ith population ($1 \leq I \leq K$, J = 1,2). the important part of the output consists of the following quantities: PCS1 and PCS2 = the estimates of $P_{\underline{\mu}(\delta^*)}(CS|R_1)$ and $P_{\underline{\mu}(\delta^*)}(CS|R_2)$, respectively, EN and EN1 = the estimates of $E_{\underline{\mu}(\delta^*)}(N|R_1)$ and $E_{EMC}(N|R_1)$, respectively and ES and ES1 = the estimates of $E_{\underline{\mu}(\delta^*)}(T|R_1)$ and $E_{EMC}(T|R_1)$, respectively. The output also contains the estimates of various other quantities and the standard errors of all the estimates.

Program 8: This program simulates the operations of rules RS_0 and RS_2. For additional details regarding the choice of the design constants of the rules, refer to §5.4. The input data consists of PSTAR, K, N1 = n_1, N = the total number of experiments in one simulation run, H1 = $h_0/\sqrt{2}$ where h_0 is defined by (5.4.2), H2 = $h/\sqrt{2}$ where h is defined by (5.3.3), AMDA = $\lambda/\sqrt{2}$ where λ is defined by (5.3.4). The important part of the output consists of the following quantities: P1 and P2 = the estimates of $P_{LFC}(CS|RS_0)$ and $P_{LFC}(CS|RS_1)$, respectively, EN1 = the estimate of $E_{\underline{\omega}}(N|RS_0)$, EN21 and EN22 = the estimates of $E_{LFC}(N|RS_1)$ and $E_{EMC}(N|RS_2)$, respectively, and ES21 and ES22 = the estimates of the expected sizes of the subsets retained in the second stage by RS_2 in the LFC and in the EMC respectively. The output also contains the estimates of various other quantities and the standard errors of all the estimates.

Program 9: This program simulates the operation of rule RS_1. For additional details regarding the choice of the design constants of RS_1, refer to §5.4. The input data consists of PSTAR, DEL, K, H1 = $h_1/\sqrt{2}$, and AMBDA = $\lambda/\sqrt{2}$ where h_1 and λ are defined by (5.2.7), H3 = h_3 where h_3 solves the equation $\Phi_{k-1}(h_3\sqrt{2},\ldots,h_3\sqrt{2};\ \{1/2\}) = \beta_2$ and N = the total number of experiments in one simulation run. The percentage points of the (k - 1) - variate equally correlated t-distribution with equal correlation = 1/2 corresponding to the percentage point $\beta_2 = (1 + P^*)/2$ for various values of the d.f. are also read as input data. H3 is used as a normal approximation to this percentage point for large number of d.f. These percentage points are used in solving the equation (5.2.8) for $H_2 = h_2$. The important part of the output consists of the following quantities: P = the estimate of $P_{LFC}(CS|RS_1)$, EN and EN1 = the estimates of $E_{LFC}(N|RS_1)$ and $E_{EMC}(N|RS_1)$, respectively, ES and ES1 = the estimates of the expected sizes of the subsets retained in the third stage by RS_1 in the LFC and in the EMC, respectively. The output also contains the estimates of various other quantities and the standard errors of all the estimates.

Program 10 (SUBROUTINE ANORM, SUBROUTINE RAND): ANORM generates standard normal random variables. RAND generates random variables which are uniformly distributed over [0,1].

BIBLIOGRAPHY

1. Alam, K. [1970]: "A two-sample procedure for selecting the population with the largest mean from k normal populations", Ann. Inst. Statist. Math. (Tokyo), 22, 127-36.

2. Bahadur, R. R. [1950]: "On a problem in the theory of k populations", Ann. Math. Statist., 21, 362-75.

3. Bessler, S. A. [1960]: "Theory and applications of the sequential design of experiments, k-actions and infinitely many experiments, Part II - Applications", Tech. Rep. No. 56, Appl. Math. Statist. Labs, Stanford Univ., Stanford, California.

4. Bechhofer, R. E. [1954]: "A single-sample multiple decision procedure for ranking means of normal populations with known variances", Ann. Math. Statist., 25, 16-39.

5. Bechhofer, R. E. [1960]: "A note on the limiting relative efficiency of the Wald sequential probability ratio test", J. Amer. Statist. Assoc., 55, 660-63.

6. Bechhofer, R. E. [1974]: "A two-sample procedure for selecting the largest mean from several normal populations with unknown variances: some comments on Ofosu's paper", Tech Rep. No. 233, Dept. of Operations Research, Cornell Univ., Ithaca, N. Y.

7. Bechhofer, R. E., Dunnett, C. W. and Sobel, M. [1954]: "A two-sample multiple decision procedure for ranking means of normal populations with a common unknown variance", Biometrika, 41, 170-76.

8. Bechhofer, R. E., Kiefer, J. and Sobel, M. [1968]: Sequential Identification and Ranking Procedures, Statistical Research Monographs, Vol. III, The University of Chicago Press, Chicago, Illinois.

9. Cohen, D. S. [1959]: "A two sample decision procedure for ranking means of normal populations with a common known variance", Unpublished M.S. Thesis, Cornell Univ., Ithaca, N. Y.

10. Dantzig, G. B. [1940]: "On the non-existence of tests of Student's hypothesis having power functions independent of σ", Ann. Math. Statist., 11, 186-92.

11. Dudewicz, E. J. [1971]: "Non-existence of a single-sample selection procedure whose P(CS) is independent of the variances", South African Statist. Journal, 5, 37-39.

12. Dudewicz, E. J. and Dalal, S. R. [1971]: "Allocation of observations in ranking and selection with unequal variances", Unpublished manuscript, Univ. Of Rochester, Rochester, N. Y.

13. Fabian, V. [1974]: "Note on Anderson's sequential procedures with triangular boundary", Ann. of Statist., 2, 170-76.

14. Fairweather, W. R. [1968]: "Some extensions of Somerville's procedure for ranking means of normal populations", Biometrika, 55, 411-18.

15. Fergusson, T. S. [1967]: "Mathematical Statistics - A Decision Theoretic Approach", Academic Press, New York, N. Y.

16. Gupta, S. S. [1956]: "On a decision rule for a problem in ranking means", Inst. Stat. Mimeo. Ser. No. 150, Inst. Stat., University of North Carolina, Chapel Hill, N. C.

17. Gupta, S. S. [1965]: "On some multiple decision (selection and ranking) rules", Technometrics, 7, 225-45.

18. Gupta, S. S. and Panchapakesan, S. [1972]: "On multiple decision procedures", Journal of Math. Physical Sciences, 6, 1-72.

19. Hall, W. J. [1959]: "The most economical character of some Bechhofer and Sobel decision rules", Ann. Math. Statist., 30, 964-69.

20. Hill, I. D. and Joyce, S. A. [1967]: "Algorithm 304, Normal curve integral", Comm. of the A. C. M., 10, 374

21. Karlin, S. and Traux, D. [1960]: "Slippage problems", Ann. Math. Statist., 31, 296-304.

22. Kiefer, J. and Weiss, L. [1957]: "Some properties of generalized sequential probability ratio tests", Ann. Math. Statist., 28, 57-74.

23. Krishniah, P. R. and Armitage, J. V. [1966]: "Tables for multivariate t-distribution", Sankhya, Ser. B, 28, 31-56.

24. Lehmann, E. L. [1959]: "Testing Statistical Hypotheses", John Wiley & Sons Inc., New York, N. Y.

25. Milton, R. C. [1963]: "Tables of the equally correlated multivariate normal probability integral", Tech. Rep. No. 27, Dept. of Statistics, University of Minnesota, Minneapolis, Minnesota.

26. Mosteller, F. [1948]: "A k-sample slippage test for an extreme population", Ann. Math. Statist., 19, 58-65.

27. Ofosu, J. B. [1973]: "A two-sample procedure for selecting the population with the largest mean from several normal populations with unknown variances", Biometrika, 60, 117-24.

28. Paulson, E. [1949]: "A multiple decision procedure for certain problems in the analysis of variance", Ann. Math. Statist., 20, 95-98.

29. Paulson, E. [1952a]: "On the comparison of several experimental categories with a control", Ann. Math. Statist., 23, 239-46.

30. Paulson, E. [1952b]: "An optimum solution to the k-sample slippage problem for the normal distribution", Ann. Math. Statist, 23, 610-16.

31. Paulson, E. [1964]: "A sequential procedure for selecting the population with the largest mean from k normal populations", Ann. Math. Statist., 35, 174-80.

32. Ramberg, J. S. [1966]: "A comparison of the performance characteristics of two sequential procedures for ranking the means of normal populations", Tech. Rep. No. 4, Dept. of Ind. Eng. & Oper. Res., Cornell Univ., Ithaca, N. Y.

33. Rinott, Y. [1974]: "On two-stage procedures for selecting the largest mean from several normal populations with unknown variances", Unpublished report, Dept. of Mathematics, Cornell Univ., Ithaca, N.Y.

34. Robbins, H., Sobel, M. and Starr, N. [1968]: "A sequential procedure for selecting the best of k populations", Ann. Math. Statist., 39, 88-92.

35. Santner, T. J. [1973]: "A restricted subset selection approach to ranking and selection problems", Mimeo. Ser. No. 318, Dept. of Statistics, Purdue University, W. Lafayette, Indiana.

36. Somerville, P. N. [1954]: "Some problems of optimum sampling", Biometrika, 41, 420-29.

37. Srivastava, M. S. [1966]: "Some asymptotically efficient sequential procedures for ranking and slippage problems", J. Roy. Statist. Soc., Ser. B, 28, 370-80.

38. Stein, C. [1945]: "A two-sample test for a linear hypothesis whose power is independent of the variance", Ann. Math. Statist, 16, 243-58.

39. Weiss, L. [1962]: "On sequential tests which minimize the maximum expected sample size", J. Amer. Statist. Assoc., 57, 551-66.

40. Weiss, L. [1964]: "Sequential Bayes procedures which never observe more than a bounded number of observations", Ann. Inst. Statist. Math. (Tokyo), 15, 177-85.

41. Wilks, S. [1962]: "Mathematical Statistics", John Wiley & Sons, New York, N. Y.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER #259	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) ON MINIMAX MULTISTAGE PERMANENT ELIMINATION TYPE RULES FOR SELECTING THE LARGEST NORMAL MEAN		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Ajit C. Tamhane		8. CONTRACT OR GRANT NUMBER(s) DAHC04-73-C-0008 N00014-67-A-0077-0020
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Operations Research College of Engineering, Cornell University Ithaca, New York 14853		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Sponsoring Military Activity U.S. Army Research Office Durham, N. C. 27706		12. REPORT DATE May 1975
		13. NUMBER OF PAGES 162
14. CONTROLLING OFFICE NAME AND ADDRESS Sponsoring Military Activity Statistics and Probability Program Office of Naval Research Arlington, Virginia 22217		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Ranking and selection procedures, Indifference zone approach, Normal means problem, Multistage (2- and 3-stage) elimination type rules, Minimax design criterion.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The problem of selecting the population associated with the largest mean from $k \geqq 2$ normal populations is considered in the context of the indifference-zone approach. For the case of common known variance, the two-stage permanent elimination type rule (R_1) proposed by Alam [1970] is analyzed in great detail extending Alam's work in several directions for $k > 2$. A new design criterion (U-minimax criterion) is proposed which specifies that the constants necessary to implement R_1 be determined so as to minimize the

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. (con't)

maximum of the total expected sample size over the entire parameter space subject to a given probability requirement. For $k = 2$, the exact tables of these constants is constructed using the Alam's design criterion and the U-minimax criterion. For $k > 2$, because of the difficulty in computing the infimum of the probability of correct selection (PCS) a simple lower bound is proposed for the PCS and a table of constants necessary for implementing a conservative rule based on this lower bound is constructed using the U-minimax criterion. The relative performance of Bechhofer's single-stage rule R_0 with respect to the two-stage rule R_1 (with a given design criterion) is studied numerically for small sample sizes and analytically for large sample sizes. The question of optimality of U-minimax rule R_1 in terms of minimizing the maximum of the expected total sample size subject to a specified probability requirement in a given class of two-stage permanent elimination type rules is investigated in detail.

Two "comeback" type modifications of R_1 which allow a population eliminated at the end of the first stage to become eligible for final selection at the end of the second stage, are proposed and studied. Monte Carlo sampling techniques are used to compare the performance of one comeback type rule with the permanent elimination type rule R_1.

Finally, the case of common unknown variance is considered. Two permanent elimination type rules, a three-stage rule (RS_1) and a two-stage rule (RS_2), are proposed and shown to guarantee a specified probability requirement. Using Monte Carlo sampling techniques, their relative performances in comparison with the two-stage non-elimination type rule (RS_0) due to Bechhofer et.al. [1954] is studied. The performance of RS_1 is found to be superior, whereas the performance of RS_2 is found to be inferior.