SOME CONTRIBUTIONS TO SEQUENTIAL SELECTION AND RANKING PROCEDURES

by

Wen-Tao Huang

Purdue University

Department of Statistics
Division of Mathematical Sciences
Mimeograph Series #295

July 1972

*This research was supported by the Office of Naval Research Contract N00014-67-A-0226-00014 at Purdue University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

INTRODUCTION

When an experimenter encounters a problem of comparing k categories such as k new drugs, k kinds of wheats or k new machines, etc., the classical tests of homogeneity are usually not sufficient to answer the various questions. His main interest is not only knowing if those categories are identical but, which category (or categories) is the best or the worst for his special purpose. Similarly, to a gambler, the question of most concern is which number on the roulette wheel comes up most frequently and not so much the question of whether the roulette wheel is fair or not. All these and related important practical problems constitute the investigations in the theory of selection and ranking.

In the theory of selection and ranking procedures, there are two basic approaches to the problem. One is called the 'indifference zone' approach and the other is called the 'subset selection' approach. In the former approach, usually the experimenter specifies a value d before the experiment, which is some sort of difference between the parameters of the best and second best population. This is called the indifference zone. The so-called best and the second best populations are to be defined in terms of some desirable characteristic by the experimenter. Usually, these are the populations associated with the largest (smallest) and the second largest (second smallest) parameters which are under consideration. In the simplest case this approach leads to the selection of a unique population. When the population selected is truly the best, it is called a correct selection (CS). The probability of a CS is at least $P^*(\frac{1}{k} < P^* < 1)$, a preassigned value, whenever the difference between the best and second best is at least equal to d. When the indifference zone is not assumed by the experimenter, he can use another approach, the subset selection approach. In this situation, the populations may have the most general configuration of the parameters. The usual procedures select a non-empty subset of populations. The number of the populations in the subset selected is usually a random variable taking values between 1 and k inclusive. In this case one requires that the probability that the subset selected contains the best (or a best) population is at least P^*. Details of this approach may be found in Gupta [21] where references to the other earlier papers are given. A rather complete reference to literatures in the field, especially for sequential procedures, is provided by Bechhofer and Kiefer and Sobel [6].

This thesis consists of three chapters. In Chapter 1, some classes of sequential selection procedures are constructed. The formulation is in terms of the indifference zone approach.

Procedures for both binomial and normal populations are derived. Also a class of procedures is constructed for the binomial populations in the case when the value d is positive but not specified. For normal populations, a sufficient condition is given for the construction of sequential procedures so that in a finite number of stages a unique population is selected under the subset selection formulation. The P*-condition is satisfied for all cases. The confidence interval estimators for the largest parameters for both the binomial and normal populations are also obtained. A class of sequential procedures for the selection of the stochastically smallest (largest) population is derived under some mild conditions.

In Chapter 2, k given normal populations are partitioned into two subgroups with respect to a control so that one group is worse and the other group is better than the control. The partition is in terms of population means (univariate case) and the generalized variances (multivariate case), respectively. For the former case Bayes and empirical Bayes procedures are derived.

In Chapter 3, the k-armed-bandit problem is formulated as a selection problem. The solution is not necessarily restricted to the binomial populations. The problem is also studied in terms of a game. A maximin strategy is derived and an asymptotic optimality property is shown to hold. Some numerical computations for the maximin strategy are given for normal and binomial populations.

CHAPTER 1

SOME CLASSES OF SEQUENTIAL PROCEDURES

1.0 Introduction

In the problem of selection and ranking, some sequential procedures have been proposed and studied. The purpose of using a sequential procedure is not only to reduce the costs of sampling, but, in some cases, it achieves a goal which can not be achieved by any kind of fixed-sample size procedure. For instance, in the problem of selecting the population corresponding to the largest mean among a finite set of normal populations with a common unknown variance, no fixed sample size procedure would satisfy the P*-condition if we use the indifference zone approach. Accordingly, studies of sequential procedures are important in ranking and selection problems.

As fixed sample size procedures in ranking and selection problem, there are two formulations of sequential procedures, that is, indifference zone formulation and subset selection formulation. The sequential procedures of Paulson [41], [43], Bechhofer and Kiefer and Sobel [6] Robbins, Sobel and Starr [49] and others belong to the former formulation while the procedures of Barron and Gupta [4] belong to the latter formulation.

In this chapter we review the types of known procedures. We treat the cases when the given populations are respectively binomial, normal and nonparametric for some formulations of problems. It is emphasized that the procedures proposed in this chapter treat the case where the conditions imposed are somewhat between indifference zone formulation and subset selection formulation. More exactly, it treats the case under a sole assumption that the difference between the largest and the second largest parameters is positive but unknown. Furthermore, a sufficient condition is given for these procedures so that in finite stages a unique population will be selected and the P*-condition is satisfied under the subset selection formulation. Some comparisons among rules in terms of the expected sample size required are investigated. Confidence interval for the largest parameter is also studied.

1.1 Various Types of Sequential Rules

In the selection and ranking problems, sequential procedures which satisfy the P*-condition can be classified as one of the following four types. Before we state the various types of rules, we introduce some definitions and notation.

Let $\pi_1, \pi_2, \ldots, \pi_k$ be k ($k \geq 2$) given populations such that π_i has cdf $F(x;\theta_i)$, where θ_i is a real-valued parameter, $i=1,2,\ldots,k$. We say π_i is the best if θ_i is the largest (or the smallest) among k parameters $\theta_1, \theta_2, \ldots, \theta_k$. Let X_{ir} denote the rth random observation from π_i for $i=1,2,\ldots,k$, $r=1,2,\ldots$. Let $\{d(n);\ n=1,2,\ldots\}$

denote a sequence of positive real-valued numbers. Let $T_i(X_1,X_2,\ldots,X_n)$ denote a statistic of the random variables $X_1,X_2,\ldots,X_n$ for $i=1,2,\ldots,k$. Then we have the following types of rules.

(a) Elimination Type Rule

At the first stage, a certain number of observations are taken from each population. According to these observations, some comparison is made among them and as a result, some populations are considered "bad" and are rejected. For these populations that are rejected, we do not take any further observations from them and no comparisons will be made with them any more. At each stage, a certain number of observations are drawn from those populations which are not rejected in the preceding stages. According to all these observations which are sampled from the remaining populations a comparison is made among them and as a result, some populations will be rejected. The sampling and comparison procedures continue until only one population is left which is considered to be the best. More generally, let $\{n_i;\ i=1,2,\ldots\}$ denote a sequence of positive integers and $\{h_i;\ i=1,2,\ldots\}$ denote a sequence of functions such that $h_i(\cdot)$ with s variables is a Lebesgue-measurable function on R^s. Then, at the rth stage, n_r observations are drawn from those t populations, say, $\pi_{i1},\pi_{i2},\ldots,\pi_{it}$ which have not been rejected in the preceding stages. Let T_{ij} denote the statistic from π_{ij}, then, reject π_{ij} if

$$T_{ij} \leq h_r(T_{i1}, T_{i2}, \ldots, T_{it}) - d(r),$$

for $r=1,2,\ldots$. The procedure of sampling continues until a unique population is left. A procedure which follows the above sampling scheme and stops in finite stages is said to belong to the elimination type. In most cases, we take n_i to be a fixed positive integer and $T_i(X_1,\ldots,X_n)$ as a sufficient statistic for θ_i; $h_i(\cdot)$ is some fixed function, say, $h(\cdot)$ for each $i=1,2,\ldots$. For example, in Paulson [41], $n_i \equiv 1$, $T_i(X_1,X_2,\ldots,X_n) = \sum_{i=1}^{n} X_i$, $h_i(Y_1,Y_2,\ldots,Y_s) = \max\{Y_1,Y_2,\ldots,Y_s\}$ and $d(n) = a_\lambda - n\lambda$ where a_λ and λ are given such that $0 < \lambda < \Delta$, $a_\lambda = [\sigma^2/(\Delta-\lambda)] \log[(k-1)/\alpha]$ for given Δ,σ and α. In Paulson [43], n_i is chosen according to a random observation from a Poisson population with a preassigned positive integer-valued parameter J. $T_i(X_1,X_2,\ldots,X_n)$ is the number of successes minus the number of failures in n tosses of observations $X_1,X_2,\ldots,X_n$ using the coin π_i. $h_i(Y_1,Y_2,\ldots,Y_s)=\max\{Y_1,Y_2,\ldots,Y_s\}$, $d(n) = -[(\log\alpha/\log\lambda)+nA(\lambda)]$ with $A(\lambda)=J[d(\lambda^2-1)-(\lambda-1)^2]/(\lambda \log \lambda)$, λ is a parameter to be chosen before the experiment such that $1 < \lambda < (1+d)/(1-d)$ for given d. In Nomachi [38] $n_i = 1$, $h_i(Y_1,\ldots,Y_s) = \min\{Y_1,Y_2,\ldots,Y_s\}$, $-d(n) = K(\lambda) + n\lambda$ where λ is some parameter to be chosen before the experiment and $K(\lambda)$ is some given function of λ.

We say a sequential rule is closed or truncated if there exists some positive integer M, say, such that the process of sampling does not exceed M stages, i.e. a decision would be made no later than M stages. Procedures described in [38], [41], [43] are all closed.

(b) All Sampling Type Rule

At each stage, a certain number of observations are drawn from each population. The sampling procedure continues until a certain condition is satisfied and as soon as the sampling procedure terminates, a decision (selection) is made. More exactly, let $\{h_n(Y_1, Y_2, \ldots, Y_k);\ n=1,2,\ldots\}$ be a sequence of measurable functions on R^k and let $\{D_n(k,d,P^*),\ n=1,2,\ldots\}$ be a sequence of functions of given parameters k,d and P*. At stage r, n_r observations are drawn from each of the populations. Let T_{ir} denote a statistic based on all preceeding observations drawn from π_i. Then, the rule calls for termination of sampling if

$$S_r : h_r(T_{1r}, T_{2r}, \ldots, T_{kr}) \leq D_r(k,d,P^*) \quad \text{holds.}$$

When S_r holds at stage r, in most cases we select π_i which is associated with $T_{[k]}$ as the best where $T_{[1]} \leq T_{[2]} \leq \cdots \leq T_{[k]}$ is the ordered values of $T_{1r}, T_{2r}, \ldots, T_{kr}$. In most cases, T_{ir} is a sufficient statistic for the parameter of interest.

The Bechhofer-Kiefer-Sobel procedure [6] (BKS-procedure) assumes $T_{in}(X_1, X_2, \ldots, X_n) = \sum_{j=1}^{n} X_{ij}$, $h_n(Y_1, Y_2, \ldots, Y_k) = \sum_{i=1}^{k-1} \exp(-dZ_i/\sigma^2)$ where $Z_i = Y_{[k]} - Y_{[i]}$ and $Y_{[1]} \leq Y_{[2]} \leq \cdots \leq Y_{[k]}$ are the ordered values of $Y_1, Y_2, \ldots, Y_k$. $D_n(k,d,P^*) = (1-P^*)/P^*$

and $n_i \equiv 1$. This procedure treats a selection problem for given k normal populations with a common known variance using the indifference zone approach.

The Robbins-Sobel-Starr procedure [49] (RSS-procedure) which selects a unique population with the largest mean of k normal populations with a common unknown variance using indifference zone approach, assumes $T_{in}(X_1,X_2,\ldots, X_n) = \sum_{i=1}^{n} (X_i - \overline{X})^2$ where $\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$, $h_n(Y_1,Y_2,\ldots, Y_k) = [k(n-1)]^{-1} \cdot \sum_{i=1}^{k} Y_i$, $D_n(k,d,P^*) = Cn$ where $C = d^2/h^2$ and h is the solution of the equation $\int_{-\infty}^{\infty} \Phi^{k-1}(y+h)\, d\Phi(y) = P^*$. $\Phi(x)$ denotes the standard normal cdf. For special values of P^*, h has been tabulated in Gupta [20]. Also $n_i = 1$. The BKS and RSS procedures are not closed. The inverse sampling procedures of Panchapakesan [39], Sobel and Weiss [52] also belong to this type.

(c) Non-Elimination Type Rule

At each stage, a certain number of observations are taken from each population and a population π_i, say, is either considered rejected, accepted, or no decision is made. As soon as π_i is considered either rejected or accepted, π_i is called tagged and its rejection or acceptance is determined although at each stage observations are still drawn from π_i. The sampling procedure terminates as soon as each of the k populations is tagged.

Let $\{I_n;\ n=1,2,\ldots\}$ be a sequence of intervals of the real line R. Let $L_n = \{x\varepsilon R \mid x<y\ \forall\ y\varepsilon I_n\}$ and $U_n = \{z\varepsilon R \mid z>y\ \forall\ y\varepsilon I_n\}$ for $n=1,2,\ldots$. Let $\{n_i;\ i=1,2,\ldots\}$ be a sequence of positive integers. Let $T_{ir}\ (X_1,X_2,\ldots,\ X_n)$ denote a suitably chosen statistic of observations from π_i. At stage r, n_r observations are drawn from each population and let $T_{ir} = T_{ir}(X_1,X_2,\ldots,\ X_{n_r})$ denote the statistic based on n_r observations from π_i. If π_i has not been tagged in the preceeding r-1 stages, then,

accept π_i if $T_{ir}\ \varepsilon\ U_r$,

reject π_i if $T_{ir}\ \varepsilon\ L_r$,

make no decision if $T_{ir}\ \varepsilon\ I_r$.

Let $S(t) = \{\pi_i \mid T_{ir}\ \varepsilon\ I_r$ for $r=1,2,\ldots,t\}$. Then, we continue sampling from each population as long as $S(t) \neq \phi$, for $t=1,2,\ldots$. Of course, the choice of $\{I_n; n=1,2,\ldots\}$ is not arbitrary. We need to choose $\{I_n\}$ so that $P\{S(t) = \phi,\ t<\infty\} = 1$.

It should be pointed out that when $I_n = [a_n, +\infty)$, $L_n = (-\infty, a_n)$, $U_n = \phi$ where $a_n = h_n(T_{i_1 n},\ldots T_{i_t n})$ for some $t, 1\leq t\leq k$, the procedure is still different from the all sampling type because of the different sampling scheme. However, if, we take $I_n = (-\infty, b_n)$ and $U_n = (b_n, +\infty)$, $L_n = \phi$, where $b_n = h_n(T_{1n}, T_{2n},\ldots, T_{kn})$ and the sampling procedure terminates as soon as one or more populations are tagged, then the rule belongs to (b), the all sampling type.

Barron [3] and Barron and Gupta [4] gave a class of procedures of this type using subset selection approach. Their procedure selects a subset of k normal populations with known variances such that the probability that the selected subset contains the population associated with the largest mean is at least P*, ($\frac{1}{k} < P^* < 1$), a preassigned value. The procedure assumes $I_n = (\delta n-\gamma_1, \delta n+\gamma_2)$ where $\delta\varepsilon(o,1)$ and γ_1 and γ_2 are positive integers. Based on Gupta's procedure R of (1.10) in [21] for n = 1, a new random variable Y_{ij} is defined such that $Y_{ij} = 1$ if the j^{th} observation from π_i is selected according to R and $Y_{ij} = 0$ otherwise. The selection problem is thus transformed into a random walk problem by this construction of new random variables. It is important to note that when $\delta\varepsilon(P_{k-1}, P_k)$, there exists a $\gamma=\gamma(\delta,\varepsilon)$ such that a procedure with $I_n = (\delta n-\gamma, \delta n+\gamma)$ selects the best population with probability at least $1-\varepsilon$ and selects the next best with probability at most ε assuming that $P_1, P_2, \ldots, P_k$ are given where P_i is the probability that the i^{th} best is selected according to R for n = 1.

This type of rule seems more economic than that of (a) in the sense that at each stage, it permits not only rejection but also acceptance so that it takes shorter time to stop. However, it seems possible to reduce the costs of sampling from those populations which have been tagged. [3] and [4] are the only known procedures belonging to this type in selection problems that have been investigated.

(d) Acceptance Type Rule

Contrary to the elimination type, it is desirable to construct a class of procedures which draw samples from each population at each stage until one or more populations are accepted. This type of rules uses, of course, the subset selection approach. More specifically, let $\{J_n;\ n=1,2,\ldots\}$ be a sequence of right half line intervals and $\{h_n(Y_1,Y_2,\ldots,Y_k);\ n=1,2,\ldots\}$ be a sequence of measurable functions. At stage r, n_r observations are drawn from each population. Let T_{ir} denote a sufficient statistic for θ_i based on all observations drawn from π_i. Then π_i is accepted, for the first r if $T_{ir} \varepsilon J_r$, for $r = 1,2,\ldots$. The sampling procedure terminates as soon as at least one population is accepted. For instance, let $\pi_1,\pi_2,\ldots,\pi_k$ be k normal populations with common known variance. We want to select a subset which contains the best population with probability at least P^*, $(\frac{1}{k} < P^* < 1)$. We may define $T_{ir}(X_1,X_2,\ldots,X_n) = \sum_{j=1}^{n} X_j$ where X_j are iid from π_i. $h_r(Y_1,Y_2,\ldots,Y_k) = \sum_{i=1}^{k} \alpha_{ir} Y_{[i]}$ where $\alpha_{ir} \geq o$ and $\sum_{i=1}^{k} \alpha_{ir} = 1$ for $r = 1,2,\ldots$ and $Y_{[1]} \leq Y_{[2]} \leq \ldots \leq Y_{[k]}$ are the ordered values of $Y_1,Y_2,\ldots,Y_k$. Take $J_r = (h_{ir}(T_{1r},T_{2r},\ldots,T_{kr}),+\infty)$. We need to define $\{h_r(\cdot),\ r=1,2\ldots\}$ and $\{\alpha_{ir},\ i=1,2,\ldots,k,\ y=1,2,\ldots\}$ so that the P^*-condition is satisfied and the expected selected size is minimized. When π_i is a coin with probability of tossing a head p_i, $i=1,2,\ldots,k$ we may define $T_i(X_1, X_2, \ldots, X_n)$ as the total number of heads of n tosses of π_i. Let $\{C(n,t); n=1,2,\ldots, t=0,1,2,\ldots\}$ be a positive monotone increasing sequence with respect to each index. Define $T_n = \sum_{j=1}^{k} T_j(X_1,\ldots,X_n)$. At stage r, accept π_i if for the first time that

$$T_i(X_1, X_2, \ldots, X_{n_r}) \geq C(r, T_{n_r})$$

where T_{n_r} is the total number of heads of all tosses of k coins, $r=1,2,\ldots$. When this procedure is single-stage, this becomes the Gupta-Nagel procedure (GN Procedure) of (2.4) in [24]. The values $\{C(1,n); n=1,2\ldots\}$ are tabulated in [24]. It should be pointed out that when $T \leq k$ (T is the total number of heads of all tosses of k coins), the GN procedure selects all coins as best. Therefore, when k is rather large and p_i is samll for each i, the single-stage procedure is quite undesirable. Furthermore, when k is large and the confir-uration is of slippage or equally spaced type, this type of rule seems more favorable than that of the elimination type.

All these four types have their own features. For the same formulations of problem, they may lead to different types of rules. Also, procedures of the same type may be used as solutions to different formulations of problems. For instance, the Paulson procedure [41], an elimination type, and BKS procedure [6], an all sampling type, are both designed for the indifference zone formulation. The Barron-Gupta procedure [4], of non-elimination type, and the procedures presented in section 1.3, of elimination type are both designed for the subset formulation. Comparison can be made between two procedures of different types, as the comparison made by Perng [44] in terms of the expected number of observations needed to terminate the sampling procedure when the P*-condition is satisfied.

In the following, we investigate some classes of procedures which belong to the elimination type.

1.2 Sequential Procedures for Selecting the Best Coin

A. Notation and Assumptions

Let $\pi_1, \pi_2, \ldots, \pi_k$ be k coins and $X_{i1}, X_{i2}, \ldots$ be independent Bernoulli trials of tossing π_i with $P(X_{i1} = 1) = 1 - P(X_{i1} = 0) = p_i$ for $i = 1, 2, \ldots, k$. Let $S_{in} = \sum_{j=1}^{n} X_{ij}$. It is known that S_{in} are independent with binomial distribution $B(n; p_i)$ for each $i = 1, 2, \ldots, k$. We assume $0 < p_i < 1$ and assume without loss of generality that $p_{[k]} = p_k$ where $0 < p_{[1]} \leq p_{[2]} \leq p_{[3]} \leq \cdots \leq p_{[k]}$ are the ordered values of $p_1, p_2, \ldots, p_k$. For a given value d, $0 < d < 1$, we assume $p_{[k]} - p_{[k-1]} \geq d$. For given P*, $\frac{1}{k} < P^* < 1$, we use $\alpha = (k-1)/(1-P^*)$ throughout this chapter. By a correct selection (CS) we mean π_k is selected and by an incorrect selection (IS) we mean some π_i other than π_k is selected. Using a rule R (defined below), let N denote the stopping variable of R and let S_N denote the total number of successes in N tosses of π_i if R selects π_i. Let $\{c(n);n=1,2...\}$ be a sequence of positive numbers.

By sequential sampling of k coins, we need to select a unique coin corresponding to $p_{[k]}$. We do not know the correct pairing between $p_{[i]}$ and π_j.

B. A Class of Procedures $R(n_o;c(n))$

Let n_o be a positive integer. A selection rule $R(n_o;c(n))$ is defined as follows.

(i) At the first stage, n_o observations are drawn from each coin and we reject π_i if

$$\frac{S_{i\,no}}{n_o} \leq \max_{j=1,2,\ldots,k} \frac{S_{j\,no}}{n_o} - c(n_o).$$

Let T_1 denote the set of coins which are not rejected at the first stage and let t_1 denote the number of coins in T_1. If $t_1 = 1$ stop sampling and select the unique coin in T_1 as best. Otherwise, take one more toss of each coin in T_1. The sampling procedure stops as soon as all but one coin are rejected.

(ii) In general, at stage r, let T_{r-1} denote the set of coins which are left in the preceeding r-1 stages. Take one more toss of each coin and we reject π_i if

$$\frac{S_{i\,n_o+r-1}}{n_o+r-1} \leq \max_{j\varepsilon T_{r-1}} \frac{S_{j,n_o+r-1}}{n_o+r-1} - c(n_o+r-1)$$

Now let T_r denote the set of coins which are left and let t_r denote the number of coins in T_r. If $t_r=1$, we stop sampling and select the unique coin in T_r as best. Otherwise, continue sampling.

Let $R(n_o,n_1;c(n))$ denote a closed (truncated) sequential rule which operates as $R(n_o;c(n))$ except that the stages of sampling do not exceed n_1-n_o+1 $(n_1 \geq n_o)$. And at stage n_1, one more toss is taken from each coin in T_{n_1} and if $t_{n_1} > 1$, we select π_i if

$$\frac{S_i(n_1-n_o+1)}{n_1-n_o+1} = \max_{j \varepsilon T_{n_1-1}} \frac{S_{j,}(n_1-n_o+1)}{n_1-n_o+1}$$

If a tie occurs, we select a coin by a random mechanism. We note that $\{n_o,n_1,\{c(n)\}\}$ uniquely defines R.

<u>Lemma 1.2.1</u> For given d, $0 < d < 1$, if n_1 is such that $c(n_1) \leq d$, then,

$$P\{IS|R(1,n_1;c(n))\} \leq (k-1) \sum_{n=1}^{n_1} [P\{|\frac{S_{kn}}{n} - p_k| \geq {}^{c(n)}/2\} + P\{\frac{S_{jn}}{n} - p_j \geq {}^{c(n)}/2\}].$$

Proof: According to $R(1,n_1;c(n))$ it is clear that an incorrect selection occurs only if there is some $i \neq k$ and some n $1 \leq n \leq n_1-1$ such that $\frac{S_{kn}}{n} \leq \frac{S_{in}}{n} - c(n)$ or there is some $j \neq k$ such that $\frac{S_{kn_1}}{n_1} \leq \frac{S_{jn_1}}{n_1}$. Therefore, we have

$$(1.2.1)\quad P\{IS|R(1,n_1;c(n))\} \leq P\{\frac{S_{kn}}{n} \leq \frac{S_{in}}{n} - c(n) \text{ for some } n, 1 \leq n \leq n_1-1 \text{ for some } i\varepsilon T_{n-1}\} + P\{\frac{S_{kn_1}}{n_1} \leq \frac{S_{jn_1}}{n_1} \text{ for some } j\varepsilon T_{n_1-1}\}$$

Let A_{in} and B_j denote respectively the following events

$$A_{in} = \{\frac{S_{kn}}{n} \leq \frac{S_{in}}{n} - c(n)\} \qquad \text{for } 1 \leq n \leq n_1-1$$

$$B_j = \{\frac{S_{kn_1}}{n_1} \leq \frac{S_{jn_1}}{n_1}\} \qquad , j \neq k .$$

By assumption that $p_k - p_j \geq d > o$, we have

$$B_j \subset \{|\frac{S_{kn_1}}{n_1} - p_k| \geq \frac{d}{2}\} \cup \{\frac{S_{jn_1}}{n_1} - p_j \geq \frac{d}{2}\}$$

Since $c(n_1) \leq d$, we have thus

$$(1.2.2)\quad B_j \subset \{|\frac{S_{kn_1}}{n_1} - p_k| \geq \frac{c(n_1)}{2}\} \cup \{\frac{S_{jn_1}}{n_1} - p_j \geq \frac{c(n_1)}{2}\}.$$

There are three possible cases for the event A_{in}.

(i) If $p_k \varepsilon (\frac{S_{kn}}{n}, \frac{S_{in}}{n})$, then, we have either

$p_k - \frac{S_{kn}}{n} \geq c(n)/2$, or,

$p_k - \frac{S_{kn}}{n} < c(n)/2$. The latter implies that

$\frac{S_{in}}{n} - p_i > c(n)/2$ because $\frac{S_{in}}{n} - \frac{S_{kn}}{n} \geq c(n)$ and $p_k > p_i$.

This concludes that

$$(1.2.3) \qquad A_{in} \subset \{|\frac{S_{kn}}{n} - p_k| \geq c(n)/2\} \cup \{\frac{S_{in}}{n} - p_i \geq c(n)/2\}.$$

(ii) If $p_k \leq \frac{S_{kn}}{n}$, then, since $p_i < p_k \leq \frac{S_{kn}}{n}$, we have

$$\frac{S_{in}}{n} - p_i \geq \frac{S_{in}}{n} - \frac{S_{kn}}{n} \geq c(n) > c(n)/2 \ .$$

Therefore, (1.2.3) still holds.

(iii) If $p_k \geq \frac{S_{in}}{n}$, then, $p_k - \frac{S_{kn}}{n} \geq \frac{S_{in}}{n} - \frac{S_{kn}}{n} > c(n)/2$.

This conculdes that (1.2.3) holds true for all cases. It follows from (1.2.1), (1.2.2) and (1.2.3) that

$$P\{IS|R(1,n_1;c(n))\} \leq P\{\bigcup_{i=1}^{k-1} \bigcup_{n=1}^{n_1-1} A_{in} \cup \bigcup_{j=1}^{k-1} B_j\}$$

$$\leq (k-1)\ P\{\bigcup_{n=1}^{n_1-1} A_{in} \cup B_j\} \qquad \text{for some } j \neq k$$

$$\leq (k-1) \sum_{n=1}^{n_1} [P(|\frac{S_{kn}}{n} - p_k| \geq c(n)/2) + P(\frac{S_{jn}}{n} - p_j \geq c(n)/2)]$$

The proof is thus complete.

Kambo and Kotz [29] (Kraft [30] gives a correction) give the following exponential bounds for binomial probabilities.

<u>Lemma 1.2.2</u> (Kambo-Kotz-Kraft). For $0 < p_i < 1$ and $c > 0$ we have

$$P\left(\frac{S_{in}}{n} - p_i \geq c\right) < \exp\left(-2nc^2 - \frac{4}{9}nc^4\right).$$

Let $\{c_1(n); n=1,2,\ldots\}$ be a positive monotone decreasing sequence. Define

$$(1.2.4) \qquad d_1(n) = \left\{\left[\frac{9\ln\left(\frac{2k-2}{1-P^*}\right)}{4n} + \frac{81}{16}\right]^{\frac{1}{2}} - \frac{9}{4}\right\}^{\frac{1}{2}}.$$

Let N be the stopping rule of $R(m_o, m_1; c_1(n))$, for some positive integers m_o and m_1. Define

$$(1.2.5) \qquad A_N = \frac{S_N}{N} - d_1(N)$$

$$(1.2.6) \qquad B_N = \frac{S_N}{N} + d_1(N)$$

$$(1.2.7) \qquad I_N = (A_N', B_N'), \text{ where } A_N' = \max(o, A_N) \text{ and}$$

$$B_N' = \min(1, B_N).$$

Then, we have the following

<u>Theorem 1.2.1</u> If m_o, m_1 and $c_1(n)$ are so chosen that

(1) $c(m_1) \leq d$

(2) $\displaystyle\sum_{n=m_o}^{m_1} \exp[-(nc_1^2(n)/2) - (nc_1^4(n)/36)] \leq \frac{1-P^*}{3(k-1)}$

then,

(a) $P\{CS \mid R(m_o, m_1; c(n))\} \geq P^*$

(b) $P\{I_n \ni p_{[k]}\} \geq 2P^*-1$

Proof: (a) We note that Lemma 1.2.1 holds for $R(m_o,m_1;c(n))$ when the summation over n is from m_o to m_1. Using this fact, we set $c = c_1(n)/2$ in Lemma 1.2.2. It follows from assumption (2) we have

$$P\{IS|R(m_o,m_1;c(n))\} \leq 1-P^*.$$

(b) Let $A = \{p_{[k]} \varepsilon I_N\}$, $B = \{\pi_k$ is selected$\}$.

Then, by (a) we have $P(B) \geq P^* > 0$.

Hence, $P(A) = P(A\cap B) + P(A\cap B^c) \geq P(A\cap B)$.

By definition of A and B we note that

$$A\cap B = \{|\frac{S_{kN}}{N} - p_k| < d_1(N)\} \cap \{\pi_k \text{ is selected}\}.$$

It follows that Lemma 1.2.2 and definition (1.2.4) of $d_1(n)$ that

$$P\{I_N \ni p_{[k]}\} \geq P(A\cap B) = P(\{|\frac{S_{kN}}{N} - p_k| < d_1(N)\} \cap\{\pi_k \text{ is selected}\}.$$

$$\geq P^* - (1-P^*)$$

$$= 2P^* - 1.$$

The proof is complete.

<u>Remark 1.2.1</u>: (1) The most economical choices of m_o and m_1 are given by $m_1 = \min\{n|c_1(n) \leq d\}$, $m_o = \min\{n|\sum_{m=n}^{m_1} \exp[-nc^2(n)/2] - (n\, c^4(n)/36)] \leq (1-P^*)/3(k-1)$ if $m_1 \geq m_o$.

(2) When $m_1 = m_o$, $R(m_o,m_o,c(n))$ becomes the Sobel-Huyett Procedure [51].

We denote $R_1 = R(m_o,m_1;c_1(n))$.

When there is some information that all p_i are large or all are small, we can improve the given procedure $R(m_o,m_1;c(n))$ by using sharper inequalities.

The following bounds are given by Kambo and Kotz [29] and Kraft [30].

<u>Lemma 1.2.3</u> (Kambo-Kotz-Kraft) If $c \geq 0$ and $0 < p_i < 1$, $q_i = 1-p_i$, then,

(1) $P\{\frac{S_{in}}{n} - p_i \geq c\} < \exp[-(n\ c^2)/(2\ p_i q_i) - (4\ n\ c^4/3)]$ if $p_i \geq q_i$

(2) $P\{\frac{S_{in}}{n} - p_i \geq c\} < (q_i/c\sqrt{n})\ \exp[-2\ n\ c^2 - (4\ n\ c^4/9)]$

if $n \geq 3$ and $p_i + c \leq \frac{1}{2}$

<u>Remark 1.2.2</u>: We note that the following holds

(1)' $P\{\frac{S_{in}}{n} - p_i \leq -c\} < \exp[-(n\ c^2/2\ p_i q_i) - (4\ n\ c^4/3)]$ if $p_i \leq q_i$

(2)' $P\{\frac{S_{in}}{n} - p_i \leq -c\} < (p_i/c\sqrt{n})\ \exp[-2\ n\ c^2 - (4\ n\ c^4/9)]$ if $n \geq 3$

and $q_i + c \leq \frac{1}{2}$.

For given $\varepsilon(0 < \varepsilon < \frac{1}{2})$, we define $R_2 = R_2(r_o,r_1;c_2(n))$ where r_o and r_1 are defined as follows.

(1.2.8) $$r_\varepsilon = \min\{n \mid c_2(n) \leq \varepsilon\}$$

$$r_o = \max\{3,\ r_\varepsilon\}$$

$$r_1 = \min\{n \mid c_2(n) \leq d\}$$

Define $d_2(n)$ to be the positive solution of equation

(1.2.9) $$4\ n\ d_2^4(n) + 6\ n\ d_2^2(n) - 3\ \ell n\ \left(1 + \frac{\exp(8\ n\ d^4(n)/9)}{\sqrt{n}\ d_2(n)}\right)$$

$$+ 3\ \ell n\ [(1-P^*)/(k-1)] \geq 0$$

Let N_2 be the stopping rule of R_2 and define

(1.2.10) $$A_{N_2} = \frac{S_{N_2}}{N_2} - d_2(N_2)$$

$$B_{N_2} = A_{N_2} + 2\ d_2(N_2)$$

(1.2.11) $$I_{N_2} = (A'_{N_2}, B'_{N_2}) \text{ where } A'_{N_2} = \max(o, A_{N_2}),$$

$$B'_{N_2} = \min(1, B_{N_2}).$$

Theorem 1.2.2 Suppose for some $\varepsilon (0 < \varepsilon < \frac{1}{2})$ we have either $p_i \geq \frac{1}{2} + \varepsilon$ for all $i = 1,2,\ldots,k$ or, $p_j \leq \frac{1}{2} - \varepsilon$ for all $j = 1,2,\ldots,k$. If $\{c_2(n)\}$ is a positive monotone decreasing sequence such that

$$\sum_{n=r_0}^{r_1} [1+\frac{4}{\sqrt{n}c_2(n)} \exp(nc_2^4(n)/18)] \exp[-(nc_2^2(n)/2)-(nc_2^4(n)/12)]$$

$\leq (1-P^*)/(k-1)$, then

(a) $P\{CS|R_2(r_0,r_1;c_2(n))\} \geq P^*$

(b) $P\{I_{N_2} \ni p_{[k]}\} \geq 2P^*-1$,

where r_0, r_1 and I_{N_2} are defined by (1.2.8), (1.2.9) and (1.2.11).

Proof: (a) Suppose $p_i \leq \frac{1}{2} - \varepsilon$ for all i. Then $q_i > p_i$ for all i. Also for every $n \geq r_0$ $c_2(n) \leq \varepsilon$ since $c_2(n)$ is monotone decreasing and $c_2(r_0) \leq \varepsilon$. Hence, by (2) of lemma 1.2.3,

$$P(\frac{S_{in}}{n} - p_i \geq c) < (1/c\sqrt{n}) \exp[-2nc^2-(4nc^4/9)] .$$

Now for any n ($n \geq r_0$), take $c=c_2(n)/2$. It follows thus

$$P(\frac{S_{in}}{n} - p_i \geq c_2(n)/2) < (2/\sqrt{n}c_2(n)\exp[-(nc_2^2(n)/2)- -(nc_2^4(n)/36)]. \tag{1.2.12}$$

Again, since $q_i \geq p_i$ for all i, it follows from (1) of lemma 1.2.3. that

$$P(\frac{S_{in}}{n} - p_i \leq \frac{-c_2(n)}{2}) < \exp[-(nc_2^2(n)/2)-(nc_2^4(n)/12)] \tag{1.2.13}$$

by taking $c=c_2(n)/2$ for $n \geq r_0$.

It follows from Lemma 1.4.1 (1.2.11) and (1.2.13) that

$$P(CS|R_2) \geq 1-(k-1) \sum_{n=r_0}^{r_1} [1+(4/\sqrt{n}c_2(n))\exp(nc^4(n)/18)] .$$

$$\cdot \exp[-(nc_2^2(n)/2)-(nc_2^4(n)/12)] .$$

$$\geq P^* .$$

Applying the same argument to the case $p_i \geq \frac{1}{2} + \varepsilon$ we come to the same conclusion.

(b) The proofs for (b) are analogous to the corresponding proofs of (b) in Theorem 1.2.1.

This completes the proof.

Remark 1.2.3 (1) In general, $c_2(n)$ defined in Theorem 1.2.2 will be smaller than $c_1(n)$ defined in Theorem 1.2.1 for the same P^*, and thus the set of suitable $c_2(n)$ will contain that of $c_1(n)$ since sharper inequalities are used.

(2) $d_2(n)$ defined by (1.2.9) will be in general smaller than $d_1(n)$ defined in Theorem 1.2.1. The best $d_2(n)$ in (1.2.9) is the solution of the equation (1.2.9) when equality holds.

(3) Under the assumptions of Theorem 1.2.1 and Theorem 1.2.2, we take $c_1(n)$ and $c_2(n)$ as small as possible.

When d is unknown but positive, Paulson [43] and BKS [6] procedures are not applicable. In this case, m_1 in Theorem 1.2.1 and r_1 of (1.2.8) may tend to infinity. Let N_1 and N_2 be respectively the stopping variables of $R_1(m_0) = R_1(m_0;c_1(n))$ and $R_2(r_1) = R_2(r_1;c_2(n))$. We have the following

<u>Corollary 1.2.1</u> If $p_{[k]}-p_{[k-1]} > 0$, and $\{c_1(n)\}$ and $\{c_2(n)\}$ are two positive decreasing sequences such that

(1) $c_i(n) \to 0$, $i = 1,2$

(2) $\sum_{n=m_0}^{\infty} \exp[-(nc_1^2(n)/2)-(nc_1^4(n)/36)] \leq \frac{1}{3\alpha}$ for some $m_0 \geq 1$

(3) $$\sum_{n=r_0}^{\infty} [1+(4/\sqrt{n}c_2(n))\exp(nc_2^4(n)/18)]\exp[-(nc_2^2(n)/2) - (nc_2^4(n)/12)] \leq \frac{1}{\alpha}$$

where r_0 is defined by (1.2.8).

Then,

(a) $P(N_i < \infty) \geq P^*$ $\quad i = 1,2$

(b) $P\{CS|R_1(m_0)\} \geq 2P^*-1$

(c) $P\{CS|R_2(r_1)\} \geq 2P^*-1$ if $p_i \geq \frac{1}{2} + \varepsilon$ for all i or, $p_j \leq \frac{1}{2} - \varepsilon$ for all j

(d) $P\{p_{[k]} \in I_{N_i}\} \geq 3P^*-2$, $i = 1,2$

(e) when $k = 2$, $P(N_i < \infty)=1$, $P(CS|R_i) \geq P^*$ and $P\{I_{N_i} \ni p_{[2]}\} \geq$ $\geq 2P^*-1$, $i = 1,2$.

Proof: (a) It suffices to show the case $i = 1$. Let $p_{[k]}-p_{[k-1]}=\delta>0$. Then, $p_k-p_i \geq \delta$ for $i \neq k$. By the strong law of large numbers, there exists n_0 such that for $n \geq n_0$ $(S_{kn}/n)-p_k > -\delta/4$ and $p_i-(S_{in}/n) \geq -\delta/4$ with probability one. Therefore, $(S_{kn}/n)-(S_{in}/n) =$

$$[(S_{kn}/n)-p_k] + (p_k-p_i) + [p_i-(S_{in}/n)] > -\delta/4+\delta-\delta/4$$
$$= \delta/2 \qquad \text{WP1} \quad \text{for } n \geq n_0.$$

Since $c_1(n)$ decreases to 0, there exists some $n_1 \geq n_0$ such that $c_1(n) < \delta/2$ if $n \geq n_1$.

This shows that $\frac{S_{in}}{n} \leq \max\limits_{j\epsilon T_{n-1}} (\frac{S_{jn}}{n}, \frac{S_{kn}}{n}) - c_1(n)$ wp1 for $n \geq n_1$.

Let $A=\{\pi_k$ is selected$\}$ and $B_{in}=\{\frac{S_{in}}{n} \leq \max\limits_{j\epsilon T_{n-1}} \frac{S_{jn}}{n} - c_1(n)\}$.

We have

$$B_{in} = (B_{in} \cap A) \cup (B_{in} \cap A^c) \text{ and thus}$$

$$\begin{aligned} P(B_{in}) &\geq P(B_{in} \cap A) \\ &= P(\{\frac{S_{in}}{n} \leq \max_{j\epsilon T_{n-1}} (\frac{S_{jn}}{n}, \frac{S_{kn}}{n}) - c_1(n)\} \cap A) \text{ if } n \geq n_1 \\ &= P(A) \\ &\geq P^* \qquad \text{for } i \neq k. \end{aligned}$$

(b) $P(CS|R_1(m_0)) \geq p(\{N_1 < \infty\} \cap A) \geq P^*-(1-P^*)=2P^*-1$.

(c) The proof is analogous to (b).

(d) By Theorem 1.2.1 and Theorem 1.2.2 that

$$P\{p_{[k]} \in I_{N_i}\} \geq 2P^*-1 \text{ when } N_i < \infty \text{ wp1}.$$

Hence, $P\{p_{[k]} \in I_{N_i}\} \geq P(\{p_{[k]} \in I_{N_i}\} \cap \{N_i < \infty\}) \geq (2P^*-1)-(1-P^*)$

$$= 3P^*-2.$$

The proof for (e) follows directly from the proof of (a) (b) and (c). This completes the proof.

<u>Corollary 1.2.2</u> If $p_i \neq p_j$ for $i \neq j$, then under the same asymptions of Corollary 1.2.1, we have

(a) $p(N_i < \infty) = 1$

(b) $p(CS|R_i) \geq P^*$

(c) $p(p_{[k]} \varepsilon I_{N_i}) \geq 2P^*-1$ for $i = 1,2$.

Proof: (a) For each fixed j, $(j = 1,2,\ldots,k-1)$, following the same argument of the proof of (a) in Corollary 1.2.1, there is some r and n(j,r) such that $p_r > p_j$ and for $n \geq n(j,r)$

$$\frac{S_{jn}}{n} \leq \frac{S_{rn}}{n} - c_1(n) \leq \max_{i \varepsilon T_{n-1}} \frac{S_{in}}{n} - c_1(n) \quad \text{wp1} ,$$

since $c_i(n) \to 0$. For (b), (c) the proofs are obvious.

<u>Corollary 1.2.3</u> If we change the sampling scheme of R_1 and R_2 to the non-elimination type of sampling (Barron-Gupta type), then, under the same assumptions of Corollary 1.2.1, we have the same results (a), (b) and (c) of Corollary 1.2.2.

Proof: Because of the non-elimination type of sampling, S_{kn} is always compared to other S_{in}. Using the same arguments of Corollary 1.2.2 the results follow immediately.

We see that when P* is sufficiently near 1, the procedures R_1 and R_2 may come to a decision in finite stages. However, if we use a sequence of closed procedures we have asymptotic results.

Let $\{n_i; i = 1,2,\ldots\}$ and $\{m_i; i = 1,2,\ldots\}$ be two positive increasing sequences of integers such that $n_1 \geq m_0$ and $m_1 \geq r_0$.

Let $R_{1i}(m_0) = R_1(m_0,n_i;\ c_1(n))$ and $R_{2j}(r_0) = R_2(r_0,m_j;\ c_2(n))$ we have the following asymptotic result.

<u>Corollary 1.2.4</u> Under the same assumptions as those of Corollary 1.2.1, let $p_{[k]}-p_{[k-1]} = d > 0$, we have

(1) $\lim\limits_{d\to 0} \lim\limits_{j\to\infty} P(CS|R_{ij}(m_0)) \geq P^*$ $\quad i = 1,2.$

(2) $\lim\limits_{j\to\infty} P\{p_{[k]} \in I_{N_{ij}}\} \geq 2P^*-1$ $\quad i = 1,2$

where N_{ij} is the stopping rule of R_{ij} for $i = 1,2$. The proofs follow immediately from Theorem 1.2.1 and Theorem 1.2.2.

For given ℓ $(0 < \ell < \frac{1}{2})$, let $n_1(\ell)$ and $n_2(\ell)$ denote respectively the smallest integers such that $d_1(n_1(\ell)) \leq \ell$ and $d_2(n_2(\ell)) \leq \ell$ which are defined by (1.2.4) and (1.2.9) respectively. Let $\bar{N}_i = \max\{N_i,n_i(\ell)\}$ $i = 1,2$, where N_i are stopping rules of some closed procedure R_i. Then, we have

<u>Corollary 1.2.5</u> Under the same assumptions of Theorem 1.2.1 and Theorem 1.2.2, $I_{\bar{N}_i}$ is a $(2P^*-1)$-confidence interval for $p_{[k]}$ with length $(I_{\bar{N}_i}) \leq 2\ell$ using $R_1(m_0,m_1;c_1(n))$ and $R_2(r_0,r_1;c_2(n))$ respectively, where

$$I_{\bar{N}_i} = \left(\frac{S_{\bar{N}_i}}{\bar{N}_i} - d_i(\bar{N}_i),\ \frac{S_{\bar{N}_i}}{\bar{N}_i} + d_i(\bar{N}_i)\right) \qquad i = 1,2.$$

Proof: It follows immediately from definitions of $\bar{N}_i$ and Theorem 1.2.1 and Theorem 1.2.2.

C. Asymptotic Bounds for Expected Sample Sizes for Selection Rules.

For given k, P* and $d(p_{[k]}-p_{[k-1]} \geq d > 0)$, let R_0 and R_{00} denote respectively the Paulson procedure (P-procedure) [43] and Bechhofer-Kiefer-Sobel procedure (BKS-procedure) [6]. We outline R_0 and R_{00} as follows.

First we describe the P-procedure. Let $\{N_{ir};\ i = 1,2,\ldots,k,\ r = 1,2,\ldots\}$ be a double sequence of independent random variables such that each has a Poisson distribution with mean J, where J is a positive integer to be chosen in advance of the experiment. Let S_{ir} and F_{ir} denote respectively the number of successes and the number of failures when N_{ir} observations are taken from π_i at the rth stage of the experiment. $N_{ir} = 0$ implies $S_{ir} = F_{ir} = 0$. Let λ be a parameter such that $1 < \lambda < (1+d)/(1-d)$, where d defines the indifference zone and λ is also to be chosen in advance of the experiment. Define $A(\lambda) = J[d(\lambda^2-1)-(\lambda-1)^2]/(\lambda \ln \lambda)$ and let $N(\lambda)$ denote the largest integer which is less than $\ln\alpha/(A(\lambda)\ln\lambda)$. For the first stage, we take N_{i1} observations from π_i, for i = 1,2,...,k. We reject π_i at the first stage if

$$S_{i1}-F_{i1} \leq \max_{1\leq j\leq k} (S_{j1}-F_{j1})-(\ln\alpha/\ln\lambda)+A(\lambda).$$

As the P-procedure is of elimination type, at the start of rth stage, let T_{r-1} denote the set populations which are not rejected and t_{r-1} denote the number of elements in T_{r-1}. We take N_{ir} observations from π_i for $i \in T_{r-1}$. Then, reject π_j if

$$\sum_{n=1}^{r} (S_{jn}-F_{jn}) \leq \max_{i\in T_{r-1}} \left[\sum_{n=1}^{r}(S_{in}-F_{in})\right]-(\ln\alpha/\ln\lambda)+rA(\lambda).$$

If there is only one population left, stop sampling and accept this population as the best, otherwise continue to the (r+1)st stage. If $t_{N(\lambda)} > 1$, take one more observation from each population in $T_{N(\lambda)}$ and select π_i if

$$\sum_{n=1}^{N(\lambda)+1} (S_{in}-F_{in}) = \max_{j\varepsilon T_{N(\lambda)}} \sum_{n=1}^{N(\lambda)+1} (S_{jn}-F_{jn}).$$

If a tie occurs, select any one by a random mechanism.

Next, we describe the BKS-procedure. At the nth stage of experiment ($n = 1,2,\ldots$) take an observation from each π_i and compute the sample sums $S_{in} = \sum_{j=1}^{n} X_{ij}$ ($i = 1,2,\ldots,k$) which employ X_{ij} from all (n-1) preceding observations. Let $S_{[1]n} \le S_{[2]n} \le \cdots \le S_{[k]n}$ denote the ordered values of $S_{1n}, S_{2n},\ldots,S_{kn}$. Define $D_{in} = S_{[k]n}-S_{[i]n}$ ($i = 1,2,\ldots,k-1$) and let $Z_n = \sum_{i=1}^{k-1} \exp[-2\ell n \frac{1+d}{1-d} D_{in}]$. We continue sampling until for the first time we have $Z_n \le (1-P^*)/P^*$, say at the nth stage. Then, we select π_i which corresponds to $S_{[k]n}$. If a tie occurs, break it by a random mechanism.

We estimate various asymptotic expected sample sizes needed to terminate sampling using R_0, R_{00} and R_1 in this section. Some asymptotic form is obtained in the sense of Perng[44].

<u>Theorem 1.2.3</u> Let N_{00i} and N_{0i} denote respectively the number of stages needed to eliminate π_i using R_{00} and R_0 for any $i=1,2,\ldots,k-1$. Let π_k be the best population and let $p_k-p_i=h_i d$, $i=1,2,\ldots,k-1$.

Then, as $P^* \to 1$, we have

(a)(B-K-S) $\quad E(N_{00i}) = \{\ell n(1-P^*)^{-1}/2d\ \ell n[(1+d)/(1-d)]\}$

$$+ o(\ell n(1-P^*))$$

if $h_i \equiv 1$

(b) $\quad E(N_{0i}) = (\ell n\alpha)/\{J[d(\lambda^2-1)-(\lambda-1)^2] + 2J\lambda h_i d\ \ell n\lambda\}$

$$+ o(\ell n\alpha)$$

Proof: (a) is given in [5, p. 270] with $\nu = \delta^*_{k,k-1}\cdot\Delta_{k,k-1} =$ $2\{\ell n[(1+d)/(1-d)]\}\cdot d$. We now prove (b).

(b) Let $(E,\mathcal{B},P)$ be the underlying probability space. Let $E_0 = \{\pi_k \text{ is selected}\}$, $E_1 = E-E_0$.

According to R_0, N_{0i} is the smallest n such that

$$\sum_{r=1}^{n}(S_{ir}-F_{ir}) \le \max_{j\varepsilon T_{n-1}}\left[\sum_{r=1}^{n}(S_{jr}-F_{jr})\right]-(\ell n\alpha/\ell n\lambda)+nA(\lambda).$$

Let J_{jn} and J_{in} be respectively the numbers of observations taken from π_j and π_i at stage n, $n = 1,2,\ldots$. Throughout the proof, i will be fixed and $j \neq i$. Then, it follows from the strong law of large numbers that

$$(1.2.14)\quad \sum_{r=1}^{n}[(S_{jr}-F_{jr})/(J_{j1}+J_{j2}+\ldots+J_{jn})]-\sum_{r=1}^{n}[(S_{ir}-F_{ir})/(J_{i1}+J_{i2}+\ldots+J_{in})]$$

$$\to (p_j-q_j)-(p_i-q_i) = 2(p_j-p_i) \quad \text{wp1} \quad \text{as } n \to \infty,$$

where $q_r = 1-p_r$, $r = i,j$. Define

(1.2.15) $\quad Q_{rn} = nJ/(J_{r1}+J_{r2}+\ldots+J_{rn}) \quad r = 1,2,\ldots,k$

$$T_{jr} = S_{jr}-F_{jr} \quad \text{for } j = 1,2,\ldots,k,\ j\neq i,\ r = 1,2,\ldots .$$

Then,

(1.2.16) $\quad \sum_{r=1}^{n} [T_{jr}/(J_{j1}+J_{j2}+\ldots+J_{jn})] = (\sum_{r=1}^{n} T_{jr}/nJ)\, Q_{jn}.$

Since $J_{r1}, J_{r2},\ldots,J_{rn}$ are independent random variables with a common Poisson distribution with mean J, it follows again from the law of large numbers that

(1.2.17) $\quad Q_{rn} \to 1$ WP1 as $n \to \infty$ for $r = 1,2,\ldots,k.$

For $\delta > 0$, it follows from (1.2.14), (1.2.16), (1.2.17) and the Egoroff theorem that on E_0 there exists A_j, B_j and $n_j(\delta)$ such that

$$A_j \cup B_j = E_0$$

$$P(B_j) < \delta \text{ and}$$

for $n \geq n_j(\delta)$, we have, on A_j, that

(1.2.18) $\quad 2(p_j-p_i)-\delta \leq [(\sum_{r=1}^{n} T_{jr})Q_{jn}/nJ]-[(\sum_{r=1}^{n} T_{ir})Q_{in}/nJ]$

$$\leq 2(p_j-p_i)+\delta$$

(1.2.19) $\quad 1-\delta \leq Q_{jn} \leq 1+\delta$ for $j = 1,2,\ldots,k,\ j \neq i$.

For $n \geq n_j(\delta)$, it follows from (1.2.8) and (1.2.9) that

$$2(p_j-p_i)-\delta-\{\delta[\sum_{r=1}^{n}(|T_{jr}|+|T_{ir}|)]/nJ\} \leq \sum_{r=1}^{n}(T_{jr}-T_{ir})/nJ$$

$$\leq 2(p_j-p_i)+\delta+\{\delta[\sum_{r=1}^{n}(|T_{jr}|+|T_{ir}|)]/nJ\} .$$

It is obvious that $n_j(\delta)$ can be chosen sufficiently large so that

$$(\sum_{r=1}^{n} |T_{jr}|/nJ) + (\sum_{r=1}^{n} |T_{ir}|/nJ) \leq 2(|p_j-q_j|+|p_i-q_i|) \leq 4.$$

Hence, for every $n \geq n_j(\delta)$, we have on A_j that

$$\text{(1.2.20)}\quad nJ[2(p_j-p_i)-5\delta] \leq \sum_{r=1}^{n}(T_{jr}-T_{ir}) \leq nJ[2(p_j-p_i)+5\delta].$$

Define $n(\delta) = \max\limits_{\substack{j=1,2,\dots,k\\ j\neq i}} n_j(\delta)$, $B = \bigcup\limits_{\substack{j=1\\ j\neq i}}^{k} B_j$, $E_2 = E_0-B$.

Then,

$P(B) < k\delta$ and for $n \geq n(\delta)$, we have on E_2 that

$$\text{(1.2.21)}\quad nJ[2(p_k-p_i)-5\delta] \leq \max \sum_{r=1}^{n}(T_{jr}-T_{ir}) \leq nJ[2(p_k-p_i)+5\delta]$$

where the maximum is taken over those populations which are left at the (n-1)st stage. We note that π_k is never rejected on E_2. Now, by the definition of N_{0i}, and (1.2.21) we have either

$$N_{0i} \leq n(\delta)\ , \text{ or,}$$

$$\ln\alpha/\{\ln\lambda[2J(p_k-p_i) + A(\lambda) + 5\delta J]\} \leq N_{0i} \leq$$

$$\leq \ln\alpha/\{\ln\lambda[2J(p_k-p_i) + A(\lambda) - 5\delta J]\} + 1.$$

Since $\ln\lambda[2J(p_k-p_i) + A(\lambda) + 5\delta J]$ is bounded and since as $P^* \to 1$, we have $\ln\alpha \to \infty$, then, when P^* is sufficiently near 1, we have

$$\text{(1.2.22)}\quad n(\delta) \leq \ln\alpha/\{\ln\lambda[2J(p_k-p_i) + A(\lambda) + 5\delta J]\} \leq N_{0i} \leq$$

$$\leq \ln\alpha/\{\ln\lambda[2J(p_k-p_i) + A(\lambda) - 5\delta J]\} + 1\ .$$

Define $E_3 = \{\omega | N_{0i}(\omega) \leq n(\delta)\} \cap E_2$.

Then, we have

(1.2.23) $E(N_{0i}) = \int_{E_1} N_{0i}dP + \int_{E_3} N_{0i}dP + \int_{E_2-E_3} N_{0i}dP + \int_B N_{0i}dP$.

According to R_0 and the definitions of E_1 and B, we have

(1.2.24) $P(E_1) < 1-P^*$, $P(B) < k\delta$ and $0 \leq N_{0i} \leq N(\lambda)$

where $N(\lambda) = [\ell n\alpha/(A(\lambda)\ell n\lambda)]$ and $[\cdot]$ denotes the largest integer. Hence,

(1.2.25) $[1/(A(\lambda)\ell n\alpha)]-1 \leq N(\lambda)/\ell n\alpha \leq [1/(A(\lambda)\ell n\lambda)]+1$

where the upper and lower bounds are independent of P*. Now, by (1.2.23), (1.2.24) and (1.2.21), we have

(1.2.26) $E(N_{0i}) \leq N(\lambda)(1-P^*)+\{\ell n\alpha/\ell n\lambda[2J(p_k-p_i)+A(\lambda)-5\delta J]\}+1+N(\lambda)k\delta$

On the other hand, we have

(1.2.27) $E(N_{0i}) \geq \{\ell n\alpha/\{\ell n\lambda[2J(p_k-p_i)+A(\lambda)+5\delta J]\}\}P(E_2-E_3)$.

We note that as $P^* \to 1$, $p(E_3) \to 0$ because of definition of E_3 and (1.2.22). Therefore, as $P^* \to 1$

$$P(E_2-E_3) \to 1-P(B) \geq 1-k\delta.$$

Now, it follows from (1.2.26), (1.2.27) and (1.2.25) that as $P^* \to 1$, we have

(1.2.28) $(1-k\delta)/\{\ell n\lambda[2J(p_k-p_i)+A(\lambda)+5\delta J]\} \leq E(N_{0i})/\ell n\alpha \leq$

$$\leq 1/\{\ell n\lambda[2J(p_k-p_i)+A(\lambda)-5\delta J]\}+k\delta/(A(\lambda)\ell n\lambda) .$$

Since (1.2.28) holds for any arbitrarily small positive δ, we thus conclude that as $P^* \to 1$,

$$E(N_{0i}) = \{\ell n\alpha/\{\ell n\lambda[2J(p_k-p_i)+A(\lambda)]\}\}+o(\ell n\alpha).$$

This completes the proof.

<u>Corollary 1.2.6</u> If $p_i < p_{i+1}$ for $i = 1,2,\ldots,k-1$, then

$$\lim_{P^*\to 1} P\{N_{01} < N_{02} < \ldots < N_{0k}\} = 1.$$

Proof: Let $\varepsilon_{ij} = p_i - p_j$, $i,j = 1,2,\ldots,k$.

It follows from (1.2.22) that on E_2, there exists r_1 $(0<r_1<1)$ such that whenever $P^* \geq r_1$, we have

(1.2.29) $N_{01} \leq \log \alpha/\{\log\lambda[2J\varepsilon_{k1}+A(\lambda)-5\delta J]\}+1$ for arbitrary fixed $\delta > 0$.

Again, by (1.2.22), there exists r_2 $(0 < r_2 < 1)$ such that whenever $P^* \geq r_2$ we have

(1.2.30) $N_{02} \geq \log \alpha/\{\log\lambda[2J\varepsilon_{k2}+A(\lambda)+5\delta J]\}$ for arbitrary fixed $\delta > 0$. We note that $\varepsilon_{k1}-\varepsilon_{k2} = p_2-p_1 > 0$ and δ can be made arbitrarily small and λ and J are fixed values. When P^* is sufficiently near 1, $\log \alpha$ is sufficiently large and therefore there exists r_3, $1 > r_3 \geq \max(r_1,r_2)$ and $\delta_1 > 0$ so that whenever $P^* \geq r_3$ and $\delta < \delta_1$, we have

$$\begin{aligned} N_{01} &\leq \log \alpha/\{\log\lambda[2J\varepsilon_{k1}+A(\lambda)-5\delta J]\}+1 \\ &< \log \alpha/\{\log\lambda[2J\varepsilon_{k2}+A(\lambda)+5\delta J]\} \\ &\leq N_{02} \qquad \text{by (1.2.29), (1.2.30).} \end{aligned}$$

Since $P(E_2) \geq P^*-k\delta$, using the analogous argument for $N_{03}, N_{04},\ldots, N_{0k}$, we conclude the results immediately.

This completes the proof.

<u>Corollary 1.2.7</u> Let T_{00} and T_0 denote respectively the total number of observations needed to conclude a selection using R_{00} and R_0. Let $h_i d = p_k - p_i$ $i = 1,2,\ldots,k-1$.

Then, as $P^* \to 1$

(1) $E(T_{00}) = \{k\ \ell n(1-P^*)^{-1}\}/\{2d\ \ell n[(1+d)/(1-d)]\}+o(\ell n(1-P^*))$

when $h_i \equiv 1$ for $i \neq k$

(2) $$E(T_0) = \ell n\alpha[\sum_{i=1}^{k-2}(B(\lambda)+2J\lambda h_i d\ell n\lambda)^{-1}+2(B(\lambda) + 2J\lambda h' d\ell n\lambda)^{-1}]+o(\ell n\alpha)$$

where

$$B(\lambda) = J[d(\lambda^2-1)-(\lambda-1)^2],\ h' = \min_{j=1,2,\ldots,k-1} h_j \text{ and } \alpha=(k-1)/(1-P^*).$$

(3) For slippage case, $h_i = 1$, and $k = 2$, we have

$$\lim_{P^*\to 1} \frac{E(T_{00})}{E(T_0)} \geq 1 \text{ or } < 1 \text{ according to } 2d\ \ell n(\frac{1+d}{1-d}) \leq B(\lambda)+2J\lambda d\ell n\lambda$$

or

$$2d\ \ell n(\frac{1+d}{1-d}) > B(\lambda)+2J\lambda d\ell n\lambda\ .$$

Proof: According to the definition of R_{00} and R_0, (1) and (2) follow from Theorem 1.2.3. (3) follows immediately from (1) and (2).

Remark 1.2.4 In order that the P-procedure be more efficient in the sense of (3), it is necessary to take a large J.

Theorem 1.2.4 Using the same notation as in Theorem 1.2.3, and letting $h' = \min_{1\leq i\leq k-1} h_i$, $h'' = \max_{1\leq j\leq k-1} h_j$, we have the following asymptotic results:

(1) (a) As $d \to 0$,

$$\frac{\ell n(P^*\alpha)}{2h''d\ell n(\frac{1+d}{1-d})} + o\ (\frac{1}{d\ell n(\frac{1+d}{1-d})}) \leq E(N_{00i}) \leq \frac{\ell n(P^*\alpha)}{2h'd\ell n(\frac{1+d}{1-d})} + o\ (\frac{1}{d\ell n(\frac{1+d}{1-d})})$$

(b) when $P^* \to 1$ and follow $d \to 0$ or when $d \to 0$ and follow $P^* \to 1$,

$$\frac{\ln\alpha}{h'' d \ln(\frac{1+d}{1-d})} + o(\frac{\ln\alpha}{d\ln(\frac{1+d}{1-d})}) \le E(N_{00i}) \le \frac{\ln\alpha}{h' d\ln(\frac{1+d}{1-d})} + o(\frac{\ln\alpha}{d\ln(\frac{1+d}{1-d})})$$

(2) (a) As $d \to 0$

$$\frac{\ln\alpha}{\ln\lambda[A(\lambda)+2Jh_i d]} + o(\frac{1}{\ln\lambda[A(\lambda)+2Jh_i d]}) \le E(N_{0i}) \le N(\lambda)(2-P^*)+o(N(\lambda))$$

(b) when $d \to 0$ and then $P^* \to 1$

$$\frac{\ln\alpha}{\ln\lambda[A(\lambda)+2Jh_i d]} + o(\frac{\ln\alpha}{\ln\lambda[A(\lambda)+2Jh_i d]}) \le E(N_{0i}) \le N(\lambda)+o(N(\lambda))$$

where $A(\lambda) = J[d(\lambda^2-1)-(\lambda-1)^2]/(\lambda\ln\lambda)$ and $N(\lambda)$ is the largest integer no larger than $\ln\alpha/(A(\lambda)\ln\lambda)$.

(3) when we take $c_1(n) = n^{\gamma-1/2}$, $0 < \gamma < 1/2$, and use $R_1(\gamma) = R_1(m_0, m_1; n^{\gamma-1/2})$, we have, as $d \to 0$,

$E(N_i) \le (1-P^*)(\frac{1}{d})^{\frac{2}{1-2\gamma}} + o((\frac{1}{d})^{\frac{2}{1-2\gamma}})$ where N_i denote the total number of samples needed to reject π_i, $i \ne k$.

Proof: (1) (a) N_{00i} is the smallest n such that

(1.2.31) $$W_n \le (1-P^*)/P^* .$$

By the strong law of large numbers we can show that

$$\frac{D_{in}}{n} \to p_{[k]}-p_{[i]} \quad \text{WP1} .$$

By Egoroff's theorem, for $\delta > 0$ there exists A and B such that $E = A \cup B$, $p(B) < \delta$ and

$\frac{D_{in}}{n}$ converges uniformly on A where E is the whole space.

Hence, there exists $n(\delta)$ such that for $n \geq n(\delta)$ and on A

$n(p_{[k]}-p_{[i]}-\delta) \leq D_{in} \leq n(p_{[k]}-p_{[i]}+\delta)$ i.e.

$$(1.2.32) \quad \sum_{i=1}^{k-1} \exp\{-2n\ln(\tfrac{1+d}{1-d})(h_i d-\delta)\} \geq W_n \equiv \sum_{i=1}^{k-1} \exp\{-2\ln(\tfrac{1+d}{1-d})D_{in}\}$$
$$\geq \sum_{i=1}^{k-1} \exp\{-2\, n\ln(\tfrac{1+d}{1-d})h_i d+\delta\}.$$

It follows from (1.2.31) and (1.2.32) that

$$\ln\{\sum_{i=1}^{k-1} \exp\{-2\, N_{00i}\ell_n(\tfrac{1+d}{1-d})(h_i d-\delta)\}\} \geq \ln(\tfrac{1-P^*}{P^*}) \geq$$
$$\ln\{\sum_{i=1}^{k-1} \exp\{-2\, N_{00i}\ln(\tfrac{1+d}{1-d})(h_i d+\delta)\}\}$$

Noting that

$$\ln\{\sum_{i=1}^{k-1} \exp[-2\, N_{00i}\ln(\tfrac{1+d}{1-d})h_i d+\delta]\}$$

$$\geq \ln\{(k-1)\exp\{-2\, N_{00i}\ln(\tfrac{1+d}{1-d})h''d\}\}$$

$$= \ln(k-1) - 2\, N_{00i}h''d\, \ln(\tfrac{1+d}{1-d}) ,$$

we have, for $n \geq n(\delta)$ and on A,

$$(1.2.33) \qquad \frac{\ln P^*\alpha}{2(h''d+\delta)\ln(\frac{1+d}{1-d})} \leq N_{00i} \leq \frac{\ln(P^*\alpha)}{2(h'd-\delta)\ln(\frac{1+d}{1-d})} + 1 .$$

Let $C = \{\omega | N_{00i}(\omega) < n(\delta)\} \cap A$

(1.2.34) then, by (1.2.33), $\lim_{d\to 0} P(C) = 0.$

Hence, we have

(1.2.35) $E(N_{00i}) = \int_{A-C} N_{00i}\,dP + \int_C N_{00i}\,dP + \int_B N_{00i}\,dP$.

It follows from (1.2.33), (1.2.34) and (1.2.35) that

$$\frac{\ell n P^*\alpha}{2h''} \leq \lim_{d\to 0} \frac{E(N_{00i})}{\left(\frac{1}{d\,\ell n\left(\frac{1+d}{1-d}\right)}\right)} \leq \frac{\ell n P^*\alpha}{2h'}$$

since δ can be made arbitrarily small.

This shows (a).

By some modifications, we see that

$$\frac{1}{2h''} \leq \lim_{P^*\to 1} \lim_{d\to 0} \frac{E(N_{00i})}{\left(\frac{\ell n\alpha}{d\,\ell n\left(\frac{1+d}{1-d}\right)}\right)} \leq \frac{1}{2h'}$$

this proves (b).

(2) We note that $1 < \lambda < \frac{1+d}{1-d}$. Hence, as $d \to 0$, $\lambda \to 1$ and

$\ell n\lambda \to 0$.

It follows from (1.2.22) that when d is sufficiently near 0,

$N_{0i} \geq n(\delta)$. Hence $\lim_{d\to 0} P(E_3) = 0$. By (1.2.23) and (1.2.22) we have

(1.2.36) $E(N_{0i}) \geq \left(\frac{\ell n\alpha}{\ell n\lambda}\right) \frac{P(E_2 - E_3)}{2Jh_i d + A(\lambda) + 5\delta J}$.

Hence, $\lim_{d\to 0} \frac{E(N_{0i})}{\left(\frac{1}{\ell n\lambda(2Jh_i d + A(\lambda) + 5\delta J)}\right)} \geq (\ell n\alpha)(1-k\delta).$

Since δ can be made arbitrarily small, we have

$$E(N_{0i}) \geq \frac{\ell n\alpha}{\ell n\lambda(2Jh_i d+A(\lambda))} + o\left(\frac{1}{\ell n\lambda(2Jh_i d+A(\lambda))}\right)$$

On the other hand, by (1.6.9) and (1.6.8), we have

$$E(N_{oi}) < N(\lambda)(1-P^*) + \frac{\ell n\alpha}{\ell n\lambda[2Jh_i d+A(\lambda)-5\delta J]} + 1 + N(\lambda)k\delta$$

for arbitrary $\delta > 0$.

Let $\delta \to 0$, we have thus

$$\begin{aligned} (1.2.37) \quad E(N_{0i}) &< N(\lambda)(1-P^*) + \frac{\ell n\alpha}{\ell n\lambda(2Jh_i d+A(\lambda))} + 1 \\ &\leq N(\lambda)(1-P^*) + \frac{\ell n\alpha}{(\ell n\lambda)A(\lambda)} + 1 \\ &= N(\lambda)(2-P^*) + 1 . \end{aligned}$$

Hence, we have

$$E(N_{0i}) < N(\lambda)(2-P^*)+o(N(\lambda)) \text{ where } N(\lambda)=\ell n\alpha/(A(\lambda)\ell n\lambda).$$

By some modifications of (1.2.36) and (1.2.37), let $P^* \to 1$, we also have

$$\ell n\alpha/[\ell n\lambda(2Jh_i d+A(\lambda))]+o(\ell n\alpha/\ell n\lambda(2Jh_i d+A(\lambda))$$

$$\leq E(N_{0i}) \leq N(\lambda)+o(1/N(\lambda)).$$

(3) The proofs are similar to the proofs of (3) of Theorem 1.3.3.

D. Application of Binomial Sequential Selection Procedures to the Selection of the Best Cell Problem in Multinomial Distributions

The problem of selecting the particular of k multinomial cells with the highest probability (the so-called best cell) has been considered in several papers. For example, Bechhofer, Elmaghraby and Morse [5] give a fixed sample size procedure. Cacoullos and Sobel [8] investigate an inverse-sampling procedure under an indifference zone formulation. Panchapakesan [39] also studies an inverse sampling procedure under subset selection formulation. Under the same formulation, Gupta and Nagel [23] consider a fixed sample size procedure. However, when the problem involves more than one multinomial population, the procedures just mentioned are no more available. For instance, a gambler faces t different kinds of roulette wheels or s different kinds of loaded dice. What kind of gambling strategies should he use with his small amount of money? We note that he can observe as long as he likes before he actually starts playing. Under this situation a modified procedure based on binomial selection rules are given below.

Let π be a multinomial population with k cells. Let them be $c_1, c_2, \ldots, c_k$ with positive probabilities $p_1, p_2, \ldots, p_k$ respectively such that $\sum_{i=1}^{k} p_i = 1$. Let $X_1, X_2, \ldots$ be observable independent random variables from π. Without loss of generality, we assume

X_1 takes k values so that $P(X_1 = i) = p_i$, $i = 1,2,\ldots,k$. We define a sequence of new random variables $\{Y_{ij}\}$ as follows:

$$Y_{11} = 1 \text{ if } X_1 = 1 \qquad Y_{21} = 1 \text{ if } X_2 = 2 \text{ and so on.}$$
$$\quad = 0 \text{ otherwise} \qquad\quad = 0 \text{ otherwise.}$$

In general, we define

(1.2.41) $Y_{ij} = 1$ if the $(i+(j-1)k)$th observation from π takes value i.

$= 0$ otherwise

where $i = 1,2,\ldots,k$, $j = 1,2,\ldots$.

By our definition of (1.2.41), it is obvious that $\{Y_{ij};\ i = 1,2,\ldots,k,\ j = 1,2,\ldots\}$ is a sequence of independent random variables. Furthermore, $\{Y_{ij};\ j = 1,2,\ldots\}$ are independent Bernoulli trials with $P(Y_{ij} = 1) = p_i = 1-P(Y_{ij} = 0)$ for $i = 1,2,\ldots,k$. Let $(\Omega,\mathcal{B},P)$ be the underlying probability space on which X_1 is a random variable. Let P_i be the probability measure on $(\Omega,\mathcal{B})$ induced by Y_{i1}, $i = 1,2,\ldots,k$. Then, we can consider Y_{ij} as the jth toss of coin π_i which obeys distribution law according to $(\Omega,\mathcal{B},P_i)$ for $i = 1,2,\ldots,k$. Using $X_1,X_2,\ldots,$ to select the cell corresponding to the largest value of p_i is equivalent to selecting the best coin based on Y_{ij}.

<u>Formulation of the problem</u>

Let $\pi_1,\pi_2,\ldots,\pi_t$ denote t multinomial populations such that π_i has k_i cells with respective cell probabilities $p_{i1}, p_{i2},\ldots,p_{ik_i}$ for

$i = 1,2,\ldots,t$. Let $k_1+k_2+\ldots+k_t = k$. Let $p_{[1]} \leq p_{[2]} \leq \cdots \leq p_{[k]}$ be the ordered values of all k cell probabilities. For a given value d $(0 < d < 1)$, we assume the indifference zone $p_{[k]} - p_{[k]} \geq d$. By a correct selection we mean a unique cell corresponding to $p_{[k]}$ is selected. Taking observations sequentially from each π_i, we seek to find the cell so that the probability of a correct selection is at least P^* $(\frac{1}{k} < P^* < 1)$, a preassigned value under the indifference zone formulation.

Let X_{ij} be the jth observation of π_i for $i = 1,2,\ldots,t$ and $j = 1,2,\ldots$. For fixed i $(i = 1,2,\ldots,t)$, according to a sequence of observable random variables $X_{i1}, X_{i2},\ldots$, we define a sequence of new random variables $Y(i,j,r)$ according to (1.2.41) for $j = 1,2,\ldots,k_i$, $r = 1,2,\ldots$. More exactly, define

(1.2.42) $Y(i,j,r) = 1$ if the $(j+(r-1)k_i)$th observation is from cell j of π_i

$= 0$ otherwise

for $j = 1,2,\ldots,k_i$, $r = 1,2,\ldots$.

Define

(1.2.43) $$S(i,j,n) = \sum_{r=1}^{n} Y(i,j,r),\ n = 1,2,\ldots .$$

Then, it is obvious that the random variable $S(i,j,n)$ is distributed as binomial with density $b(n,p_{ij})$.

We describe procedures $\tilde{R}_{00}$, $\tilde{R}_0$ and $\tilde{R}_1$ as follows:

Procedure $\tilde{R}_{00}$:

On each stage, we take k_i observations from π_i for all $i = 1,2,\ldots,t$. According to definitions (1.2.42) and (1.2.43), we have corresponding observations $S(i,j,n)$. In general, at stage r, we have $\{S(i,j,r);\ j = 1,2,\ldots,k_i,\ i = 1,2,\ldots,t\}$, a set of k values. We use R_{00}, the BKS-procedure introduced at the beginning of last section, to select the best coin. As soon as $S(i,j,n_0)$ is selected, say, at stage n_0, we select the jth cell of π_i as best.

Procedure $\tilde{R}_0$:

At the first stage, k_i observations are taken from π_i for all $i = 1,2,\ldots,t$. We have k corresponding observations $S(i,j,1)$, $j = 1,2,\ldots,k_i$, $i = 1,2,\ldots,t$. According to rule R_0, the P-procedure, some cells may be rejected. Suppose s_{i1} cells of π_i are rejected, we sample $r_{i1} = (k_i - s_{i1})$ observations from π_i for the second stage. Let the numbers of cells of π_i that are not rejected be $i(1)$, $i(2),\ldots,i(r_{i1})$. According to its ascending order. We define

$$Y(i\ ,\ i(j),2) = 1 \text{ if jth observation of } \pi_i \text{ is from cell } i(j)$$
$$= 0 \text{ otherwise}$$

and $S(i,i(j),2) = \sum_{r=1}^{2} Y(i,i(j),r)$.

The sampling scheme proceeds and on each stage some cells may be rejected according to R_0. Sampling procedure stops as soon as a unique cell is left. We select this cell as best.

Procedure $\tilde{R}_1$:

Choose a certain procedure R_1 in the class of procedures proposed in Theorem 1.2.1. Following the same sampling scheme as $\tilde{R}_0$, we use R_1 to select the best cell.
Let $\tilde{N}_1$ be the stopping rule of $\tilde{R}_1$. Then, we have

Theorem 1.2.5 For given d and P*, we have

(1) $P\{CS|\tilde{R}_i\} \geq P^*$ $i = 00,0,1$

(2) $P\{p_{[k]} \in I_{\tilde{N}_1}\} \geq 2P^*-1$

where $I_{\tilde{N}_1}$ is defined by (1.2.4) through (1.2.7) replacing N by $\tilde{N}_1$.

(3) $P\{p_{[k]} \in I_{\tilde{N}_1}\} \geq 2P^*-1$ with length $(I_{\tilde{N}_1}) \leq 2\ell$

where $I_{\tilde{N}_1}$ are defined by Corollary 1.2.3 replacing N_1 by $\tilde{N}_1$.
The proofs follow immediately from our definitions of $\tilde{R}_{00}$, $\tilde{R}_0$ and $\tilde{R}_1$ and Theorem 1.2.1 and Corollary 1.2.3.

For $k_i = 2$, our problem becomes a special case of t coins. Let $\pi_1,\pi_2,\ldots,\pi_k$ be k coins such that π_i has p_i of coming up a head and probability q_i of coming up a tail for $i = 1,2,\ldots,k$. Let $r_{[1]} \leq r_{[2]} \leq \cdots \leq r_{[2k]}$ denote the order values of all p_i and q_i. For given d $(0 < d < 1)$, we assume $r_{[2k]}-r_{[2k-1]} \geq d$. Let $p_{(1)}q_{(1)} = \min_{j=1,2,\ldots,k} p_jq_j$ (which is unique because of our indifference zone assumption). Let $\pi_{(1)}$ be the coin corresponding to $p_{(1)}q_{(1)}$. We want to select a coin which has the minimum variance. By a correct selection, we mean $\pi_{(1)}$ is selected.

Corollary 1.2.8

(1) $P\{CS|\tilde{R}_i\} \geq P^*$, $i = 00, 0, 1$.

(2) $P\{p_{(1)}q_{(1)} \varepsilon (A'_{\tilde{N}_1}(1-B'_{\tilde{N}_1}), B'_{\tilde{N}_1}(1-A'_{\tilde{N}_1}))\} \geq 2P^*-1$

where $A'_{\tilde{N}_1}$ and $B'_{\tilde{N}_1}$ are defined by (1.2.4) through (1.2.7) replacing N_1 by $\tilde{N}_1$.

Proof: We note that for each i, $p_i q_i = p_i - p_i^2$ is a monotone decreasing function of p_i if $p_i > 1/2$. Similarly, it is a monotone decreasing function of q_i if $q_i > 1/2$. Therefore, it is clear that $p_i q_i = p_{(1)}q_{(1)}$, the minimum value, if either $p_i = r_{[2k]}$ or, $q_i = r_{[2k]}$. When $p_i = q_i = 1/2$ for all i, this contradicts our assumption of indifference zone. Therefore, to select $\pi_{(1)}$ is equivalent to select π_i such that the best cell is in π_i. (1) is thus an immediate consequence of Theorem 1.2.5.

For (2), we note that $p_{(1)}q_{(1)} = r_{[2k]}\, r_{[1]}$ and $\{r_{[2k]} \varepsilon (A'_{\tilde{N}_1}, B'_{\tilde{N}_1})\}$ is the same event as $\{r_{[2k]}r_{[1]} \varepsilon (A'_{\tilde{N}_1}(1-B'_{\tilde{N}_1}), B'_{\tilde{N}_1}(1-A'_{\tilde{N}_1})\}$ in probabilistic sense.

Finally, let $\tilde{T}_{00}$ and $\tilde{T}_0$ denote, respectively, the total number of observations needed to select the best cell using $\tilde{R}_{00}$ and $\tilde{R}_0$. Then, by our definitions of $\tilde{R}_{00}$ and $\tilde{R}_0$ and Theorem 1.2.3, we have •

<u>Corollary 1.2.9</u> As $P^* \to 1$, we have

(1) $E(\tilde{T}_{00}) = k\{\ln(1-P^*)^{-1}/d\ln[d/(1-d)]\} + o(\ln\alpha^{-1})$ if $h_{ij} \equiv 1$

(2) $$E(\tilde{T}_0) = \ln\alpha\{\sum_{(i,j)\in T} \frac{1}{J[d(\lambda^2-1)-(\lambda-1)^2]+2J\lambda h_{ij} d\ln\lambda} + \frac{2}{J[d(\lambda^2-1)-(\lambda-1)^2]+2J\lambda d\ln\lambda}\} + o(\ln\alpha)$$

if $p_{[k]}-p_{ij} = h_{ij}d \geq d$, where $T=\{(i,j), j=1,2,\ldots,k_i, i=1,\ldots,t\}$ excluding the cells corresponding to $p_{[k]}$ and $p_{[2k-1]}$.

1.3. Sequential Procedures for Selecting the Best Population Among k Normal Populations

A. Introduction and Formulations of Problems.

Paulson [41], Bechhofer, Kiefer and Sobel [6] and Robbins, Sobel and Starr [49] present various kinds of sequential procedures for selecting a population corresponding to the largest mean of k normal populations employing the indifference zone formulation. Barron and Gupta [4] present a class of non-elimination type of procedures under subset selection formulation. Fabian [15] improves Paulson's procedure [41] in the sense that the domain of parameter λ, which is to be chosen before the experiment, can be extended to $[0,\Delta)$ from original $(0,\Delta)$ and the probability of a CS is increased from $(1-\alpha)$ to $(1-\beta(\lambda)\alpha)$ where $\beta(\lambda) \leq 1$.

In this section, we propose a class of selection rules of elimination type for both the indifference zone approach and the subset selection approach. It should be emphasized that a unique population is selected satisfying the P*-condition under the subset selection formulation. Some asymptotic bounds on the expected sample sizes are also discussed.

Let $\pi_1,\pi_2,\ldots,\pi_k$ be k normal populations such that π_i has the distribution function $\Phi(x;\ \theta_i,\sigma^2)$, where θ_i is unknown and σ^2 is known, $i = 1,2,\ldots,k$. Let $\theta_{[1]} \leq \theta_{[2]} \leq \cdots \leq \theta_{[k]}$ be the ordered values of k means and let $\pi_{(i)}$ denote the population which is associated with $\theta_{[i]}$. We assume no prior information about this correct association. We define, for given $d > 0$,

$\Omega(d) = \{(\theta_1,\theta_2,\ldots,\theta_k)|\theta_{[k]}-\theta_{[k-1]} \geq d\}$. Let R^k denote the k-dimensional Euclidean space. Let Ω denote the parameter space in the problem at hand. Then our problem is to define a rule which samples sequentially and selects one population so as to guarantee with a prescribed probability P^* $(\frac{1}{k} < P^* < 1)$ that $\pi_{(k)}$ is selected. We assume $\pi_k = \pi_{(k)}$.

<u>Case 1</u>

$\Omega = \Omega(d)$, $d > 0$, d may or may not be specified.

<u>Case 2</u>

$\Omega = R^k$ $(k \geq 2)$.

B. A Class of Rules $R_3 = R_3(n_0,n_1;d(n))$

Let x_{ij} denote the jth observation from π_i and define

$$S_{in} = \sum_{j=1}^{n} X_{ij} \quad \text{for } i = 1,2,\ldots,k,\ n = 1,2,\ldots\ .$$

The rule $R_3(n_0,n_1;d(n))$ has the same form as the rule $R(n_0,n_1;c(n))$ of B of Section 1.2. with $d(n) = nc(n)$. When n_1 is not specified, the procedure is not closed. Let N denote the stopping variable of $R_3(n_0;d(n))$. We must show that $P(N < \infty) = 1$ in our problem so that a selection is possible and thus a correct selection makes sense. In this case, we denote the procedure by $R_3(n_0,N;d(n))$.

Case 1. $\Omega = \Omega(d)$, d specified.

We shall first discuss the truncated case and define

$$(1.3.1) \quad \beta = \sqrt{\pi}(1-P^*)/3(k-1)$$

$$(1.3.2) \quad f(v(n)) = \exp(nv(n))[(nv(n))^{1/2}+(nv(n)+(4/\pi))^{1/2}]$$

where $v(n)$ is a positive sequence.

$$(1.3.3) \quad I_N = (\frac{S_N}{N} - \ell(N),\ \frac{S_N}{N} + \ell(N)) \text{ for some positive function } \ell(n).$$

<u>Theorem 1.3.1</u> For given $d > 0$, P^* and k $(k \geq 2)$, if $\{d(n); n=1,2,\ldots\}$ is a positive sequence such that $d(n)/n$ decreases to zero and n_0 and n_1 are positive integers such that

(1) $d(n_1)/n_1 \leq d$

(2) $\sum_{n=n_0}^{n_1} [f(\frac{d^2(n)}{8\sigma^2 n^2})]^{-1} \leq \beta$

then,

(a) $P\{CS|R_3(n_0,n_1;d(n))\} \geq P^*$

(b) $P\{I_N \ni \theta_{[k]}\} \geq 2P^*-1$

where N is the stopping rule of R_3 and $\ell(n) = \frac{\sigma}{\sqrt{n}} \Phi^{-1}(\frac{1+P^*}{2})$.

Proof: For any i, $\frac{S_{in}}{n}$ is distributed according to $\Phi(\theta_i, \frac{\sigma^2}{n})$.

Therefore,

$$(1.3.4) \quad P\{\frac{S_{in}}{n} - \theta_i > a\} = \frac{\sqrt{n}}{\sqrt{2\pi}\sigma} \int_a^\infty \exp(-nt^2/2\sigma^2)dt.$$

We note that Lemma 1.2.1 holds for our case. Hence, we have

$$(1.3.5) \quad P\{CS|R_3\} \geq 1-(k-1) \sum_{n=n_0}^{n_1} [P(|\frac{S_{kn}}{n} - \theta_k| > \frac{d(n)}{2n}]+P(\frac{S_{jn}}{n} - \theta_j > \frac{d(n)}{2n})]$$

we note that the normal density $n(x;0, \frac{\sigma^2}{n})$ is symmetric about 0. Using a well-known inequality (see for example 7.1.13 of [55])

$$(1.3.6) \quad \int_a^\infty \exp(-t^2) \leq \exp(-a^2)/[a+(a^2+\frac{4}{\pi})^{1/2}] \quad \text{for } a > 0,$$

it follows from (1.3.2), (1.3.4) and (1.3.5) that

$$(1.3.7) \quad P\{CS|R_3\} \geq 1 - \frac{3(k-1)}{\sqrt{\pi}} \sum_{n=n_0}^{n_1} [f(\frac{d^2(n)}{8\sigma^2n^2})]^{-1} .$$

By our assumption (2) and definition of β, (a) follows immediately.

Follow the analogous argument in Theorem 1.2.1 and use our $\ell(n)$, (b) is concluded. The existence of n_0 and n_1 follows from our assumption that $d(n)/n$ decreases to 0 and the convergence of the series $\sum_{n=1}^{\infty} [f(d^2(n)/8\sigma^2n^2)]^{-1}$.

Remark 1.3.1: If we take $d(n) = c_1 n^{1/2+\gamma}$, $\frac{1}{2} > \gamma > 0$, c_1 positive constant, then, $n_1 = [[(d/c_1)^{\frac{2}{2\gamma-1}}]]$ if $f(c_1^2 n_1^{2\gamma}/8\sigma^2) \geq \frac{1}{\beta}$ and $n_0 = \min\{n \mid \sum_{m=n}^{n_1} [f(\frac{c_1^2 m^{2\gamma}}{8\sigma^2})]^{-1} < \beta\}$. Here c_1 is used to adjust n_0 and n_1.

Corollary 1.3.1 Under the assumption that the parameter space is $\Omega = \Omega^+ \equiv \{(\theta_1, \theta_2, \ldots, \theta_k) \mid \theta_{[k]} > \theta_{[k-1]}\}$ and if N is the stopping variable of $R_3(n_0, N; d(n))$ such that

(1) $d(n)/n$ decreases to 0

(2) $\sum_{n=n_0}^{\infty} [f(d^2(n)/8\sigma^2 n^2)]^{-1} \leq \beta$

we have

(a) $P(N < \infty) \geq P^*$

(b) $P(CS \mid R_3(n_0 N, d(n)) \geq 2P^* - 1$

(c) $P\{I_N \ni p_{[k]}\} \geq 3P^*-2$, $P^* > \frac{2}{3}$

(d) when $k = 2$, $P(N < \infty) = 1$, $P(CS \mid R_3) \geq P^*$, $P\{p_{[2]} \in I_N\} \geq 2P^*-1$.

The proofs are analogous to that of Corollary 1.2.1.

Case 2 $\Omega = R^k$

In this we have

Lemma 1.3.1 Let X_{ij} be iid with cdf $\Phi(\theta, \sigma^2)$ for $i = 1,2$ and $j = 1,2,\ldots$. Define $S_{in} = X_{i1} + X_{i2} + \ldots + X_{in}$, $i = 1,2$.

Then, using rule $R_3(1, N; d(n))$,

$$P(N < \infty) = 1 \quad \text{if} \quad \sum_{n=1}^{\infty} \frac{d(n)}{n\sqrt{2n}\,\sigma} \exp\left(\frac{-d^2(n)}{4n\sigma^2}\right) \text{ diverges.}$$

Proof. Define $Y_n = X_{1n} - X_{2n}$ for $n = 1,2,\ldots$. Then $Y_1, Y_2 \ldots$ are iid with cdf $\Phi(0, 2\sigma^2)$. Let $Z_n = Y_n/\sqrt{2}\,\sigma$. Then $Z_1, Z_2, \ldots$ are iid with cdf $\Phi(0, 1)$. Let $V_n = Y_1 + Y_2 + \ldots + Y_n$ and $U_n = Z_1 + Z_2 + \ldots + Z_n$. Then $V_n = \sqrt{2}\,\sigma U_n$. Let $a(x) = \int_{|t|>x} t^2 d\Phi(t;0,1)$ for $x \geq 0$. Then, it is obvious that

$$0 \leq a(x) = \frac{2}{\sqrt{2\pi}} \int_x^\infty t^2 e^{-t^2/2} dt \leq \frac{2}{\sqrt{2\pi}} e^{-x^2/4} \int_x^\infty t^2 e^{-t^2/4}\, dt$$

$$\leq \frac{2e^{-x^2/4}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t^2 e^{-t^2/4}\, dt = 4\sqrt{2}\, e^{-x^2/4} .$$

Then, it follows that $a(x)\, \ln\ln x \leq 4\sqrt{2}\, \ln\ln x/\exp(x^2/4) \to 0$ as $x \to \infty$, i.e. $a(x) = o((\ln\ln x)^{-1})$. It follows from Feller [16] that

$$P(U_n > \sqrt{n}\, d(n) \text{ i.o.}) = 1 \quad \text{if, and only if} \sum^{\infty} \frac{d(n)}{n} \exp(-d^2(n)/2)$$

diverges.

Or, equivalently, $P(V_n \geq d(n) \text{ i.o.}) = 1$ if, and only if

$\sum^{\infty} (d(n)/n\sqrt{2n}\,\sigma) \exp(-d^2(n)/4n\sigma^2)$ diverges. According to R_3,

$$P(N < \infty) = 2 \sum_{n=1}^{\infty} P(S_{1n} - S_{2n} \geq d(n)) \geq P(S_{1n} - S_{2n} \geq d(n) \text{ i.o.})$$

$= P(V_n \geq d(n) \text{ i.o.})$. This proves the lemma.

<u>Theorem 1.3.2</u> Let $f(V(n))$, β and I_N be defined by (1.3.1) through (1.3.3). For given $P^*(\frac{1}{k} < P^* < 1)$, if $d(n)$ satisfies

(1) $\sum_{n=1}^{\infty} (d(n)/n\sqrt{2n}\,\sigma)\exp(-d^2(n)/4n\sigma^2)$ diverges

(2) $\sum_{n=n_0}^{\infty} [f(d^2(n)/8\sigma^2n^2)]^{-1} \leq \beta$ for some $n_0 \geq 1$.

Then, using $R_3(n_0, N, d(n))$, we have

(a) $P(N < \infty) = 1$

(b) $P\{CS|R_3\} \geq P^*$

(c) $P\{I_N \ni \theta_{[k]}\} \geq 2P^*-1$

where $\ell(n) = (\sigma/\sqrt{n})\ \Phi^{-1}((1+P^*)/2)$ for I_N.

Proof: It suffices to show (a). (b) and (c) follow from (a) and Theorem 1.3.1. To show (a), we consider the following cases.

(i) $\Omega = \Omega_0 \equiv \{(\theta, \theta, \ldots, \theta); \theta \in R\}$.

For any fixed i $(i = 1,2,\ldots,k)$ let N_i denote the stage at which π_i is rejected. Then $P(N_i < \infty) = \sum_{n=1}^{\infty} P(\max_{j\in T_{n-1}} (S_{jn} - S_{in}) \geq d(n))$

$\geq P(\max_{j\in T_{n-1}} (S_{jn} - S_{in}) \geq d(n)$ i.o.)

$\geq P(S_{jn} - S_{in} \geq d(n)$ i.o.)

$=1$ (by assumption (a) and lemma 1.3.1).

Hence, $N_i < \infty$ WP1 for $i = 1,2,\ldots,k$.

Hence, $N \leq N_1 + N_2+\ldots+N_k < \infty$ WP1

(ii) $\Omega = \Omega_1 \equiv \{(\theta_1, \theta_2,\ldots,\theta_k) \mid \theta_{[1]}=\theta_{[2]}=\ldots=\theta_{[\ell]} < \theta_{[\ell+1]} = \theta_{[\ell+2]} =\ldots= \theta_{[k]}$ for some ℓ, $1 \leq \ell \leq k-1\}$

We consider $\pi_{(1)}, \pi_{(2)}, \ldots, \pi_{(\ell)}$ first where $\pi_{(i)}$ is associated with $\theta_{[i]}$. Let $d'(n)$ and T'_n be respectively associated with $d(n)$ and when our present $k = \ell$. We note that $d'(n)$ is the same as $d(n)$ and thus assumption (1) is satisfied. Let N' denote the stopping variable for selecting $\pi_{(1)}, \pi_{(2)}, \ldots, \pi_{(\ell)}$. Then, by (i), $N' < \infty$ WP1. Furthermore, for any $i \leq \ell$, we note that $\pi_{(i)}$ is rejected at stage n if

(1.3.8) $S_{(i)n} \leq \max_{j \in T'_{n-1}} S_{jn} - d'(n)$, where $S_{(i)n}$ is associated with $\pi_{(i)}$.

By noting the fact that

(1.3.9) $$\{S_{(i)n} \leq \max_{j \in T'_{n-1}} S_{jn} - d'(n)\} \subset \{S_{(i)n} \leq \max_{j \in T_{n-1}} S_{jn} - d(n)\}$$

We conclude that in a finite stages at most one population of $\pi_{(1)}, \ldots, \pi_{(\ell)}$ is left. Same treatment can be applied to these populations $\pi_{(\ell+1)}, \pi_{(\ell+2)}, \ldots, \pi_{(k)}$ which are associated with $\theta_{[\ell+1]}, \theta_{[\ell+2]}, \ldots, \theta_{[k]}$ respectively.

The only thing left to be considered is the case that $\pi_{(i)}$ and $\pi_{(j)}$ $(i \leq \ell < j)$ are left. This case is considered in the following.

(iii) when $k = 2$, $\theta_{[1]} < \theta_{[2]}$.

Let $\theta_{[2]} - \theta_{[1]} = \varepsilon > 0$. It follows from the strong law of large numbers that

$S_{in}/n \to \theta_i$ WP1 as $n \to \infty$ $i = 1,2$. Assume $\theta_2 = \theta_{[2]}$.

Then, there exists $n_1(\varepsilon)$ and $n_2(\varepsilon)$ such that

$$P(|\frac{S_{in}}{n} - \theta_i| < \varepsilon/4 \quad n \geq n_i(\varepsilon)) = 1,\ i = 1,2.$$

Let $n(\varepsilon) = n_1(\varepsilon) + n_2(\varepsilon)$. Then, for $n \geq n(\varepsilon)$

$$\begin{aligned}\frac{S_{2n}}{n} - \frac{S_{1n}}{n} &= (\frac{S_{2n}}{n} - \theta_2) - (\frac{S_{1n}}{n} - \theta_1) + (\theta_2 - \theta_1)\\ &> \varepsilon - \varepsilon/4 - \varepsilon/4\\ &= \varepsilon/2 \qquad \text{WP1.}\end{aligned}$$

Hence, $S_{2n} - S_{1n} > \varepsilon/2n$ WP1 for $n \geq n(\varepsilon)$.

The convergence condition of (2) implies that the order of $d(n)$ is less than n. This implies that

$$S_{2n} - S_{1n} \geq d(n) \quad \text{WP1} \quad \text{for } n \geq n(\varepsilon)$$

i.e. $P(N \leq n(\varepsilon) < \infty) = 1$.

(iv) $k = 3$, $\theta_{[1]} < \theta_{[2]} < \theta_{[3]}$.

Without losing generality, we may assume $\theta_{[1]}=\theta_1$, $\theta_{[2]}=\theta_2$.
Consider π_1 and π_2 and we note that

$$\{S_{1n} \leq S_{2n} - d(n)\} \subset \{S_{1n} \leq \max_{i\varepsilon T_{n-1}} (S_{in}) - d(n)\}.$$

Similarly, for the case of π_1 and π_3 or π_2 and π_3. By this fact and the conclusion of (iii), we conclude that $P(N < \infty) = 1$.

(v) For any configuration we may have is included in one of the above four cases or some combinations of the cases given. The proof is thus complete.

Remark 1.3.2. Similar results can be obtained for the case of binomial populations.

C. Asymptotic Bounds on the Expected Sample Sizes of Rules.

In this section we compute some asymptotic bounds of expected sample sizes of the Paulson procedure (P-procedure) [41], Bechhofer-Kiefer-Sobel procedure (BKS procedure) [6] and $R_3(\gamma) = R_3(n_0,n_1;\ d(n))$ with $d(n) = n^{1/2+\gamma}$, $0 < \gamma < \frac{1}{2}$.

We describe briefly the BKS-procedure and P-procedure. Let X_{ij} be the jth observation from π_i. Define $S_{in} = \sum_{j=1}^{n} X_{ij}$. Let $S_{[1]n} \leq S_{[2]n} \leq \cdots \leq S_{[k]n}$ be the ordered values of $S_{1n},\ldots,S_{kn}$. Define $D_{in} = S_{[k]n} - S_{[i]n}$ for $i = 1,2,\ldots,k-1$. Under the formulation $\Omega = \Omega(d)$, the BKS procedure, which will be denoted by R_{00}, is defined as follows.

Define $W_n = \sum_{i=1}^{k-1} \exp\{-d\ D_{in}/\sigma^2\}$, $n = 1,2,\ldots.$ At each stage, one observation is drawn from each π_i and continue sampling until for the first time, say at stage n, that $W_n \leq (1-P^*)/P^*$. We select the population associated with $S_{[k]n}$.

The P-procedure which will be denoted by R_0 is as follows. For a preassigned λ, $(0 < \lambda < d)$, define $a = (\sigma^2/(d-\lambda))\log\alpha$ $(\alpha=(k-1)/(P^*-1))$. Then $R_0=R_3(1,m_1;\ d(n))$ with $d(n)=a-\lambda n$ and

$m_1 = [a/\lambda]+1$ where $[\cdot]$ denotes the largest integer.

Let N_{ri} denote the number of observations needed to eliminate π_i using R_r where $r = 00, 0, 3$ where $R_3 = R_3(\gamma)$. It should be pointed out that the choice of $R_3(\gamma)$ in the class of R_3 proposed in Theorem 1.3.1 is by no means good in the sense of minimizing its asymptotic expected sample size.

<u>Theorem 1.3.3</u> For $\Omega = \Omega(d)$, let $\theta_k - \theta_i = h_i d$ ($h_i \geq 1$) for $i = 1,2,\ldots,k-1$. Also let $h' = \min_{j=1,2,\ldots,k-1} h_j$, $h'' = \max_{j=1,\ldots,k-1} h_j$. Then, we have the following asymptotic bounds.

(1a) As $d \to 0$

$$(\sigma^2 \ln\alpha P^*)/(h''d^2) + o(1/d^2) \leq E(N_{00i}) \leq (\sigma^2 \ln\alpha P^*)/(h'd^2) + o(1/d^2)$$

(1b) As $d \to 0$ and then $P^* \to 1$ or $P^* \to 1$ and then $d \to 0$, we have

$$(\sigma^2 \ln\alpha/h''d^2) + o(\ln\alpha/d^2) \leq E(N_{00i}) \leq (\sigma^2 \ln\alpha/h'd^2) + o(\ln\alpha/d^2)$$

when $h_i = 1$, i.e. the slippage case, we have equality.

(2a) As $d \to 0$

$$aP^*/\lambda(1+\zeta h_i) + o(a/\lambda) \leq E(N_{0i}) \leq (1-P^*)\frac{a}{\lambda} + aP^*/\lambda(1+\zeta h_i) + o(\frac{a}{\lambda})$$

(2b) As $d \to 0$ and then $P^* \to 1$ or as $P^* \to 1$ and then $d \to 0$, we have

$$E(N_{0i}) = a/\lambda(1+\zeta h_i) + o(a/\lambda)$$

where $a = (\sigma^2/(d-\lambda))\ln\alpha$ and $\zeta = \lim_{d\to 0} (d/\lambda) > 1$.

We note that $a/\lambda \geq \sigma^2 \ln\alpha/d^2$.

(3) As $d \to 0$

$$E(N_{3i}) \leq (1-P^*)(1/d)^{2/1-2\gamma} + o((1/d)^{2/1-2\gamma}) \qquad (0 < \gamma < \tfrac{1}{2}).$$

Proof: (1a) N_{00i} is the smallest n such that

$$(1.3.10) \qquad W_n \leq (1-P^*)/P^*.$$

By the strong law of large numbers we can show that

$$(1.3.11) \qquad D_{in}/n \to \theta_{[k]} - \theta_{[i]} \quad \text{WP1}.$$

By Egoroff's Theorem, for $\delta > 0$, there exists A and B such that $E = A \cup B$ and $P(B) < \delta$ and D_{in}/n converges uniformly on A, where $(E,\mathcal{B},P)$ is the underlying probability space. Hence, there exists $n(\delta)$ such that for $n \geq n(\delta)$ we have

$$n(\theta_{[k]}-\theta_{[i]}-\delta) \leq D_{in} \leq n(\theta_{[k]}-\theta_{[i]}+\delta).$$

It follows then that for $n \geq n(\delta)$

$$(1.3.12) \quad \sum_{i=1}^{k-1} \exp[-dn(\theta_{[k]}-\theta_{[i]}-\delta)/\sigma^2] \geq W_n \equiv \sum_{i=1}^{k-1} \exp[-d\, D_{in}/\sigma^2]$$
$$\geq \sum_{i=1}^{k-1} \exp[-dn(\theta_{[k]}-\theta_{[i]}+\delta)/\sigma^2].$$

It follows from (1.3.10) and (1.3.12) that

$$(1.3.13) \quad \ln\{\sum_{i=1}^{k-1} \exp[-d\, N_{00i}(\theta_{[k]}-\theta_{[i]}-\delta)/\sigma^2]\} \geq \ln[(1-P^*)/P^*)] \geq$$
$$\ln\{\sum_{i=1}^{k-1} \exp[-d\, N_{00i}(\theta_{[k]}-\theta_{[i]}+\delta)/\sigma^2]\}.$$

We note that

$$(1.3.14)\quad \ln\{\sum_{i=1}^{k-1} \exp[-d\ N_{00i}(\theta_{[k]}-\theta_{[i]}+\delta)/\sigma^2]\}$$

$$\geq \ln\{(k-1)\exp[-d\ N_{00i}(h''d+\delta)/\sigma^2]\}$$

$$= \ln(k-1)-d\ N_{00i}(h''d+\delta)/\sigma^2\ .$$

Similarly, we have

$$(1.3.15)\quad \ln\{\sum_{i=1}^{k-1} \exp[-d\ N_{00i}(\theta_{[k]}-\theta_{[i]}-\delta)/\sigma^2]\} \leq \ln(k-1)$$

$$-\ d\ N_{00i}(h'd-\delta)/\sigma^2\ .$$

It follows from (1.3.13), (1.3.14) and (1.3.15) that

$$(1.3.16)\quad n(\delta) \leq [\sigma^2/d(h''d+\delta)][\ln(k-1)-\ln((1-P^*)/P^*)] \leq N_{00i} \leq$$

$$[\sigma^2/d(h'd-\delta)][\ln(k-1)-\ln((1-P^*)/P^*)]$$

when d sufficiently nears 0.

Let $C = \{N_{00i} < n(\delta)\} \cap A$.

Then, we see from (1.3.16) that

$$(1.3.17)\quad \lim_{d\to 0}\ P(C) = 0.$$

We have thus

$$(1.3.18)\quad E(N_{00i}) = \int_C N_{00i}dP + \int_{A-C} N_{00i}dP + \int_B N_{00i}dP\ .$$

It follows from (1.3.16) and (1.3.18) we have

$$(1.3.19)\quad [\sigma^2/d(h''d+\delta)][\ell n(k-1)-\ell n((1-P^*)/P^*]]P(A-C) \le E(N_{00i}) \le$$

$$[\sigma^2/d(h'd-\delta)][\ell n(k-1)-\ell n((1-P^*)/P^*)]P(A-C)+P(B)\cdot K$$

$$+n(\delta)P(C)\text{ , for some constant K.}$$

By (1.3.17) and definition of A we note that

$$(1.3.20)\quad \lim_{d\to 0} P(C) = 0 \quad\text{and}\quad \lim_{\delta\to 0} P(A) = 1.$$

It follows from (1.3.19) and (1.3.20) that

$$\lim_{\delta\to 0}\ \lim_{d\to 0} \frac{E(N_{00i})}{[\sigma^2/d(h'd-\delta)]\ell n(P^*\alpha)} \le 1$$

similarly, we have

$$\lim_{\delta\to 0}\ \lim_{d\to 0} \frac{E(N_{00i})}{[\sigma^2/d(h''d+\delta)]\ell n(P^*\alpha)} \ge 1 .$$

This shows (1a).

Divide each side of (1.3.19) by $\ell n\ \alpha$ and take $d \to 0$ and then $P^* \to 1$, we have

$$E(N_{00i}) \ge (\sigma^2\ell n\alpha/h''d^2)+o((\ell n\alpha/d^2)) \text{ and}$$

$$E(N_{00i}) \le (\sigma^2\ell n\alpha/h'd^2)+o((\ell n\alpha/d^2)).$$

This shows (1b).

(2a) According to R_0, N_{0i} is the smallest n such that

$$(1.3.21)\qquad \max(S_{jn}-S_{in}) \ge a-n\lambda\ , \quad 1 \le n \le m_1$$

where $a = [\sigma^2/(d-\lambda)]\ell n\ \alpha,\ \ 0 < \lambda < d,\ m_1 = [\frac{a}{\lambda}]+1.$

Let $E_0 = \{\pi_k \text{ is selected}\}$, $E_1 = E - E_0$.

By using the strong law of large numbers and the Egoroff's Theorem and similar arguments in (1a), we have for $\delta > 0$ there exist A_j and B_j such that

(1.3.22) $\quad E = A_j \cup B_j$ and $P(B_j) < \delta$

and there exists $n(\delta)$ such that on A_j and $n \geq n(\delta)$

(1.3.23) $\quad n(\theta_j-\theta_i-\delta) \leq S_{jn}-S_{in} \leq n(\theta_j-\theta_i+\delta)$.

(1.3.24) Let $n_0(\delta) = \max_j n_j(\delta)$, $B = \bigcup_{j=1}^{k} B_j$, $E_2 = E_0 - B$.

Then, on E_2 for $n \geq n(\delta)$

$$\max_{j\varepsilon T_{n-1}} n(\theta_j-\theta_i-\delta) \leq \max_{j\varepsilon T_{n-1}} (S_{jn}-S_{in}) \leq \max_{j\varepsilon T_{n-1}} n(\theta_j-\theta_i+\delta) .$$

π_k is not eliminated on E_2 , we have thus

(1.3.25) $\quad n(\theta_k-\theta_i-\delta) \leq \max_{j\varepsilon T_{n-1}} (S_{jn}-S_{in}) \leq n(\theta_k-\theta_i+\delta)$.

It follows from (1.3.21) and (1.3.25) that

(1.3.26) $\quad a(\theta_k-\theta_i-\delta-\lambda)^{-1} \leq N_{0i} \leq a(\theta_k-\theta_i+\lambda-\delta)^{-1}+1$.

Let $E_3 = \{N_{0i} \leq n_0(\delta)\}$.

Then,

(1.3.27) $\quad E(N_{0i}) = \int_{E_1} N_{0i}dP+\int_{E_3} N_{0i}dP+\int_{E_2-E_3} N_{0i}dP+\int_B N_{0i}dP.$

It follows from (1.3.21), (1.3.26) and (1.3.27) that

$$a(h_i d-\delta-\lambda)^{-1}P(E_2-E_3) \le E(N_{0i}) \le (a/\lambda)(1-P^*)+n_0(\delta)+(ak\delta/\lambda)$$

$$+\ a(h_i d-\delta+\lambda)^{-1}P^* \ .$$

Hence, we have

$$(1.3.28)\qquad \frac{P(E_2-E_3)}{[(\frac{d}{\lambda})h_i+\frac{\delta}{\lambda}+1]} \le \frac{E(N_{0i})}{(\frac{a}{\lambda})} \le (1-P^*) + \frac{n_0(\delta)}{(\frac{a}{\lambda})} + k\delta$$

$$+\ \frac{P^*}{[(\frac{d}{\lambda})h_i - \frac{\delta}{\lambda}+1]}$$

we note that $\lambda(d-\lambda) < \lambda\ d < d^2$ by definition of λ. Hence,

$\frac{a}{\lambda} = \frac{\sigma^2 \ell n\alpha}{\lambda(d-\lambda)} > \frac{\sigma^2 \ell n\alpha}{d^2}$. Hence as $d \to 0$, $(\frac{a}{\lambda}) \uparrow \infty$.

Let $\lim\limits_{d\to 0} \frac{d}{\lambda} = \zeta > 1$. Then, it follows from (1.3.28) that

$$\frac{P^*}{(\zeta h_i+1)} \le \lim_{d\to 0} \frac{E(N_{0i})}{(\frac{a}{\lambda})} \le (1-P^*) + k\delta + \frac{P^*}{(\zeta h_i+1)} \ .$$

Since $\lim\limits_{d\to 0} P(E_3) = 0$ and let δ decrease to 0, we have

$$\frac{aP^*}{\lambda(1+\zeta h_i)} + o(\frac{a}{\lambda}) \le E(N_{0i}) \le (1-P^*)(\frac{a}{\lambda}) + \frac{aP^*}{\lambda(1+\zeta h_i)} + o(\frac{a}{\lambda})$$

as $d \to 0$. This proves (1a).

It follows from (1.3.28) that if we take $P^* \to 1$ and then $d \to 0$, we have

$$E(N_{0i}) = \frac{a}{\lambda(1+\zeta h_i)} + (\frac{a}{\lambda}) \ .$$

By (1.3.28) if we take $d \to 0$ first and then $P^* \to 1$, we obtain the same result. This shows (2b).

(3) According to R_3, N_3 is the smallest n such that

$$(1.3.29) \quad \max_{j \in T_{n-1}} (S_{jn}-S_{in}) \geq n\, d(n) = n^{3/2-\gamma} \qquad (m_0 \leq n \leq m_1).$$

Using same notations of (2a) and same arguments of (2a) we have for $\delta > 0$ there exist A_j and B_j such that $E_0 = A_j \cup B_j$ and $P(B_j) < \delta$ and there exist $n_j(\delta)$ such that for $n \geq n_j(\delta)$ on A_j we have

$$n(\theta_j-\theta_i-\delta) \leq S_{jn}-S_{in} \leq n(\theta_j-\theta_i+\delta) .$$

Let $n_0(\delta) = \max_j n_j(\delta)$ and $B = \cup B_j$, $E_2 = E_0-B$.

We can obtain that

$$(1.3.30) \quad n(\theta_k-\theta_i-\delta) \leq \max_{j \in T_{n-1}} (S_{jn}-S_{in}) \leq n(\theta_k-\theta_i+\delta) \text{ on } E_2.$$

It follows from (1.3.29) and (1.3.30) we have, on E_2,

$$N_3(\theta_k-\theta_i-\delta) \leq N_3^{3/2-\gamma} \leq N_3(\theta_k-\theta_i+\delta)+1.$$

On E_2, we have either

$$N_3 \leq n_0(\delta) \quad \text{or}$$

$$(1.3.31) \quad (\theta_k-\theta_i+\delta)^{-1} \leq N_3^{-0.5+\gamma} \leq (\theta_k-\theta_i-\delta)^{-1}+1$$

(1.3.32) Define $E_3 = \{N_3 \leq n_0(\delta)\}$.

We note that $m_1 \geq [[d^{\frac{-2}{1-2\gamma}}]]$ $(0 < \gamma < \frac{1}{2})$ with $\sum_{m_0}^{m_1} [f(d^2(n)/8\sigma^2 n^2)]^{-1} \leq \beta$ where $f(x)$ and β are both independent of d. When d sufficiently nears 0, we can consider $m_1 = d^{\frac{-2}{1-2\gamma}}$.

Since

$$
\begin{aligned}
E(N_3) &= \int_{E_1} N_3 dP + \int_{E_3} N_3 dP + \int_{E_2 - E_3} N_3 dP + \int_B N_3 dP \\
&\leq m_1 P(E_1) + n_0(\delta) + m_1 k\delta + (h_i d - \delta)^{\frac{-2}{1-2\gamma}} P(E_2 - E_3) \\
&\leq d^{\frac{-2}{1-2\gamma}}(1 - P^* + k\delta) + (h_i d - \delta)^{\frac{-2}{1-2\gamma}} + n_0(\delta)
\end{aligned}
$$

noting $n_0(\delta)/(\frac{1}{d})^{\frac{1-2\gamma}{2}} \to 0$ as $d \to 0$ and

$$
\lim_{\delta \to 0} \frac{d^{\frac{1-2\gamma}{2}}}{(h_i d - \delta)^{\frac{2}{1-2\gamma}}} = \left(\frac{d}{h_i d}\right)^{\frac{2}{1-2\gamma}} = \left(\frac{1}{h_i}\right)^{\frac{2}{1-2\gamma}} \quad \left(\frac{2}{1-2\gamma} > 1\right),
$$

we see that

$$
\lim_{\delta \to 0} \lim_{d \to 0} \frac{E(N_3)}{(\frac{1}{d})^{\frac{2}{1-2\gamma}}} \leq 1 - P^* .
$$

This shows (3).

<u>Remark 1.3.2</u>. Since $a/\lambda \geq \sigma^2 \ell n\alpha/d^2$, when σ^2 is not big, R_{00} is better than R_0 in the sense that $E(N_{00i}) \leq E(N_{0i})$ when $P^* \to 1$ and $d \to 0$ in the slippage case.

1.4 A Class of Nonparametric Selection Procedures

In the field of nonparametric selection and ranking problems, some results have been obtained. Barlow and Gupta [2] considered the problem of selecting a subset containing the population which is associated with the largest quantile of a given order when the distributions of the populations are not specified but belong to some restricted family of distributions. Gupta and McDonald [22] considered the problem of selecting a subset containing the population associated with the largest parameter when the population belong to a stochastically increasing family of distributions and the selection procedures are based on ranks. Rizvi and Sobel [46] considered the problem of selecting a subset containing the population associated with the largest α-quantile when the populations considered satisfy some indifference zone condition. Other papers are: Lehman [36], Puri and Puri [45].

In this section we consider the problem of selecting a single population with the distribution which is stochastically smaller (larger) than the others when the distributions of populations considered satisfy some conditions which may or may not depend on the parameters. The procedure is sequential and is based on two kinds of ranks.

A. Notation and Assumptions

Let $\pi_1, \pi_2, \ldots, \pi_k$ be k populations. Assume that π_i is distributed according to continuous cumulative distribution $F_i(x)$. The parametric forms of $F_1(x), F_2(x), \ldots, F_k(x)$ are not specified. We assume there exists one π_i, unknown to the experimenter, such that $F_i(x) > F_j(x)$ for every x in R and for all $j \neq i$. We assume that all $F_j(x)$ have same support. Let $\pi_{(k)}$ denote the population which is associated with the above $F_i(x)$. By a correct selection (CS), we mean $\pi_{(k)}$ is selected.

Under the above assumptions, we want to sample sequentially and base on some functions of ranks of observations, we want to select the unique population in a finite number of stages of sampling such that the P*-condition is satisfied.

Definition 1.4.1

Let X_{ij} denote the jth observation of π_i, $i = 1,2,\ldots,k$, $j = 1,2,\ldots$. Let $R_{ij}(n)$ denote the rank of X_{ij} among the pooled observations when n independent observations are drawn from each of k populations.
Let $r_{ij}(n)$ denote the rank of X_{ij} in the n observations drawn from π_i. For any i and j, we define

(1.4.1) $d_n(i,j) = \max(r_{i\alpha}(n) - r_{j\beta}(n) + 1, 0)$ where the maximum

is taken over those α and β such that $R_{i\alpha}(n) < R_{j\beta}(n)$. For any j, we define

$$(1.4.2) \qquad D_n(j) = \max_{\substack{i=1,2,\ldots,k \\ i\neq j}} d_n(i,j).$$

B. A Class of Rules $R_4(m_0;e(n))$

Let m_0 be a positive integer and let $\{e(n), n=1,2,\ldots\}$ be a sequence of positive numbers which are to be specified before the experiment. A non-elimination sequential procedure $R_4(m_0;e(n))$ is defined as follows:

(i) m_0 observations are drawn from each population. Compute $D_{m_0}(i)$, $i = 1,2,\ldots,k$. We consider π_i to be rejected if, and only if, $D_{m_0}(i) \geq e(m_0)$ and we say π_i is tagged.

If at this stage only one population is left, stop sampling and select this population as the best. Otherwise, draw one more observation from each of k populations and consider only those which are not tagged. If π_j is not tagged in the first stage and $D_{m_0+1}(j) \geq e(m_0+1)$,

we consider π_j to be rejected and π_j is tagged at this second stage.

(ii) In general, at the nth stage, one more observation is drawn from each of k populations. If π_i is not tagged in the preceeding stages, we consider π_i to be rejected if, and only if,

$D_{m_0+n-1}(i) \geq e(m_0+n-1)$ and π_i is tagged at this stage. We stop sampling as soon as only one population is not tagged or as soon as all populations are tagged.

In the former case we select the one which is not tagged as the best. In the latter case we proceed as follows.

(iii) Let $\pi_{i_1}, \pi_{i_2}, \ldots, \pi_{i_r}$ be the populations which are tagged at the last stage and only at the last stage of sampling. We select any one of them as best by a random mechanism.

Before giving the main result, we mention a result due mainly to Gnedenko and Korolyuk [19]. Darling and Robbins [11] also show it and give an upper bound for a certain probability as the following.

<u>Lemma 1.4.1</u> (Darling-Robbins) If $F(x) \leq G(x)$ for every $x \in R$, then

$$P\{\sup_x (F_n(x)-G_n(x)) \geq \frac{s}{n}\} \leq \frac{(n!)^2}{(n^2-s^2)!} \leq \exp(-s^2/n+1)$$

for $s = 0,1,2,\ldots,n$, where $F_n(x)$ and $G_n(x)$ are, respectively, the empirical distributions based on n observation from $F(x)$ and $G(x)$, respectively.

<u>Theorem 1.4.1</u> If

(1) $e(n)/n$ tends to 0 as $n \to \infty$ and

(2) $\sum_{n=m_0}^{\infty} \exp(-e^2(n)/n+1) \leq (1-P^*)/(k-1)$ for some $m_0 \geq 1$

then,

(a) $P(N < \infty | R_4(m_0;e(n)))=1$

(b) $P(CS|R_4(m_0;e(n))) \geq P^*$

where N is the stopping variable associated with R_4.

Proof: We may assume $\pi_k = \pi_{(k)}$ for convenience.

(a) According to the definition of (1.4.1) we can show that

(1.4.3) $d_n(i,j) \geq e(n)$ if, and only if $\sup_x (F_{in}(x)-F_{jn}(x)) \geq e(n)/n$

where F_{in} and F_{jn} are respectively the empirical distributions of F_i and F_j. By the well-known Glivenko-Cantelli lemma, we have

$$P\{\lim_{n\to\infty} \sup_x (F_i(x)-F_{in}(x)) = 0\} = 1, \quad i = 1,2,\ldots,k.$$

Since, for each fixed i $(i = 1,2,\ldots,k-1)$,

$$\sup_x (F_{kn}(x)-F_{in}(x)) \to \sup(F_k(x)-F_i(x)) = d_{ki} > 0, \text{ say, WP1}$$

and $e(n)/n$ tends to 0, it follows that there is some finite n_i such that

$$\sup_x (F_{kn}(x)-F_{in}(x)) \geq e(n)/n \quad \text{WP1 for } n \geq n_i .$$

According to (1.4.3) we have

$$d_{n_i}(k,i) \geq e(n_i) \quad \text{WP1.}$$

Hence, we have shown that

$$P(D_{n_i}(i) \geq e(n_i)|R_4(m_0;e(n)))=1 \quad \text{for } i = 1,2,\ldots,k-1.$$

It follows obviously that

$$N \leq n_1+n_2+\ldots+n_{k-1} < \infty \quad \text{WP1.}$$

(b) We note that

$P\{\pi_k \text{ is rejected} | R_4\} = P\{\text{For at least one } i \ (i=1,2,\ldots,k-1)$

there is some m_i. $m_0 \le m_i < \infty$ such that $D_{m_i}(k) \ge e(m_i)\}$

$\le (k-1)P\{d_{m_i}(i,k) > e(m_i)\}$ for some $m_0 \le m_i < \infty$, $i \ne k$

$$\le (k-1) \sum_{n=m_0}^{\infty} P(d_n(i,k) > e(n))$$

$< 1-P^*$ (this follows from (1.4.3) and Lemma 1.4.1 and Assumption a(2)). This completes the proof.

Remark 1.4.1: If we change our sampling scheme of R_4 to that of R_1 and use the same argument as Corollary 1.3.1 we obtain (a) $P(N < \infty | R_4) \ge P^*$ and (b) $P(CS|R_4) \ge 2P^*-1$ under the same assumptions (1) and (2) of Theorem 1.4.1. We note that the sampling scheme of elimination type is better than that of non-elimination type because the former saves the costs of sampling.

CHAPTER 2

PARTITIONING OF k NORMAL POPULATIONS WITH RESPECT TO A CONTROL

2.0 Introduction

In many of the experimental situations the experimenter is confronted with the problem of partitioning k populations into two classes. Usually one class is better than a control and the other is worse. The terms better and worse are up to the experimenter and depend on his particular goal. Paulson [40] made the initial efforts in this direction. He considered the problem of selecting the best one of k categories when comparing k-1 categories with a standard control. He treated the case of k normal populations with a common unknown variance and the case of k binomial populations. He controlled the probability of selecting the standard as best when other categories are equal to the standard. Later, Dunnett [14] considered the same problem of normal populations. But his goal is to select all those treatments which are better than the control. He controlled the probability of selecting the control when the treatments are all equal to the control. Gupta and Sobel [25] considered a problem of selecting all populations (normal or binomial) which are better than or equal to a control under arbitrary configuration. Their procedures controlled the probability that all good (better than equal to the control) populations are selected. Some of these results could be applied to several other distributions in the Koopman-Darmois family. Lehmann [35] considered some model 1 problems using a general decision theoretic approach and gave some criterions of selection procedures. He obtained a minimax solution and illustrated some specific examples. Krishnaiah [31] considered the problem of selecting multivariate normal populations better than a control on the basis of the linear combinations of the elements of covariance matrices. Krishnaiah and Rizvi [32] considered the problem of selecting multivariate normal populations better than a control based on various scalar quantities, such as linear combination of the elements of mean vector , Mahalanobis distances of populations from a control etc. Using various definitions of positive (good) populations and negative (bad) populations (which are earlier defined by Lehmann [35]), they studied the probability of selecting all positive populations, the expected proportion of positive populations in the subset selected and the expected proportion of negative populations in the subset selected. Recently, Tong [54] considered the problem of partitioning k univariate normal populations according to their location parameters with respect to a control. The problem is set up in the indifference zone formulation. He controlled the probability of a misclassification. The problem is investigated in terms of single-stage and multiple-stage procedures. Some optimum and asymptotic properties are shown.

In this chapter, we consider the problem of partitioning k normal populations into disjoint exhaustive classes with respect to a given control. When the comparison with respect to the control is in terms of the location parameter, we consider the univariate case using Bayes and empirical Bayes approaches. When it is based on scale parameter, we consider the multivariate case (p-dimensional) and control the probability of a misclassification in the sense of Tong [54] which will be defined later. When p=1 and p=2 we also investigate the sequential procedures. An optimum property is also discussed.

2.1 Definitions and Notation

Let π_0, $\pi_1, \ldots, \pi_k$ be k+1 populations with parameters of distribution functions θ_0, $\theta_1, \ldots, \theta_k$, respectively. For given values ρ_1 and ρ_2 with $\rho_2 < \rho_1$ and a given function g defined on a subset of R^2 we define

$$
\begin{aligned}
(2.1.1) \qquad \mathcal{P}_G &= \{\pi_i : | g(\theta_i,\theta_0) \geq \rho_1\} \\
\mathcal{P}_B &= \{\pi_i : | g(\theta_i,\theta_0) \leq \rho_2\} \\
\mathcal{P}_I &= \{\pi_i : | \rho_2 < g(\theta_i,\theta_0) < \rho_1\} \\
\mathcal{P} &= \{\pi_1, \pi_2, \ldots, \pi_k\} .
\end{aligned}
$$

Based on n observations from each population ($n \geq 1$, preassigned), we want to partition $\mathcal{P}$ into two disjoint exhaustive subsets, say, S_G and S_B. S_G and S_B, for example, might be subsets of good and bad populations, respectively. Let X_i denote the sample space of

π_i for $i = 0,1,2,\ldots,k$ and let $X = X_0 \times X_1 \times \ldots \times X_k$ be the cartesian product. Define

$$(2.1.2) \qquad K = \{1,2,\ldots,k\}$$
$$SK = \{S | S \subset K\}$$

We note that there are 2^k elements including empty set in SK. A decision function d for our problem is a measurable function from X into SK so that for an observation $x \in X$, if $d(x) = S$, we partition θ into $S_G = \{\pi_i | i \in S\}$ and $S_B = \theta - S_G$. Define

$$(2.1.3) \qquad S_M(d,x) = \{\pi_i | i \in (d(x) \cap \bar{\theta}_B) \cup ((K - d(x)) \cap \bar{\theta}_G)\}$$

where $\bar{\theta}_B = \{i \mid \pi_i \in \theta_B\}$ and $\bar{\theta}_G = \{j \mid \pi_j \in \theta_G\}$.

<u>Definition 2.1.1</u> If $\pi_i \in S_M(d,x)$, we say under observation of x, π_i is misclassified by d. If $S_M(d,x) = \phi$, the empty set, we say d(x) is a correct decision (CD).

<u>Definition 2.1.2</u> A loss function L(.,.) is a non-negative function on SK $\times\ \Omega$ (Ω denotes the parameter space of $\pi_0, \pi_1, \ldots, \pi_k$) such that

$$L(S,\omega) = 0 \quad \text{if} \quad \bar{\theta}_G \subset S \subset \bar{\theta}_G \cup \bar{\theta}_I \quad \forall\, \omega \in \Omega.$$

We define

$$(2.1.4) \qquad \begin{aligned} L_{1i}(S,\omega) &= \alpha \quad \text{if} \quad i \in \bar{\theta}_G \quad \text{and } i \notin S \\ &= \beta \quad \text{if} \quad i \in \bar{\theta}_B \quad \text{and } i \in S \\ &= 0 \qquad\qquad\qquad\quad \text{otherwise} \end{aligned}$$

$$L_{2i}(S,\omega) = \alpha(\theta_i-\theta_0) \text{ if } i \in \bar{\mathcal{P}}_G \text{ and } i \notin S$$

$$= \beta(\theta_0-\theta_i) \text{ if } i \in \bar{\mathcal{P}}_B \text{ and } i \in S$$

$$= 0 \qquad \text{otherwise}$$

(2.1.5)
$$L_1(S,\omega) = \sum_{i=1}^{k} L_{1i}(S,\omega)$$

$$L_2(S,\omega) = \sum_{i=1}^{k} L_{2i}(S,\omega).$$

We use the above notation and definitions throughout this chapter unless otherwise stated.

2.2 Location Parameter for Univariate Case

Let $\pi_0, \pi_1,\ldots,\pi_k$ be k+1 univariate normal populations with mean $\theta_0, \theta_1,\ldots,\theta_k$ and variances $\sigma_0^2, \sigma_1^2,\ldots,\sigma_k^2$, respectively. We define $g(x,y) = x-y$ for $x,y \in R$.

Tong [54] treat the case $\rho_2 < \rho_1$ and $\sigma_i^2 = \sigma^2$, $i = 0,1,2,\ldots,k$ for both known and unknown σ^2. In this section we treat the problem for known σ_i^2 and $\rho_2 \leq \rho_1$ using Bayes approach.

Deely and Gupta [13] considered the problem of subset selection of normal populations using Bayes approach. Deely [12] also considered the same problem using empirical Bayes approach. We also consider our problem using empirical Bayes approach.

A. Bayes Procedure

We assume θ_i have a prior distribution each distributed according to normal distribution with known mean λ_i and known variance φ_i^2 and we also assume σ_i^2 are known for $i = 0,1,2,\ldots,k$. Let the configuration of $\pi_0, \pi_1,\ldots,\pi_k$ be $\Omega = \{(\theta_0, \theta_1,\ldots,\theta_k) | \theta_i \in R\}$ and G_i

denote the normal prior distribution of θ_i and G be the joint prior distribution on Ω such that G is the product of G_i which assumes known mean λ_i and known variance φ_i^2. Let x_{ij} denote the jth observation from π_i and let $\bar{x}_i = \sum_{j=1}^{n} x_{ij}/n$, the sample mean of π_i for $i = 0,1,2,\ldots,k$, $j = 1,2,\ldots$.

<u>Theorem 2.2.1</u> Using loss functions L_1 and L_2, defined by (2.1.5), respectively, we have Bayes procedures R_{1G} and R_{2G}, respectively, defined by the following:

(i) R_{1G}: If S is a subset of $\{1,2,\ldots,k\}$ and S^c is the complement of S such that

$$\alpha \sum_{i\varepsilon S^c} [1-\Phi(J(\underset{\sim}{x}_i,\rho_1))] + \beta \sum_{j\varepsilon S} \Phi(J(\underset{\sim}{x}_j,\rho_2))$$

$$= \min_{T\subset\{1,2,\ldots,k\}} \{\alpha \sum_{i\varepsilon T^c} [1-\Phi(J(\underset{\sim}{x}_i,\rho_1))] + \beta \sum_{j\varepsilon T} \Phi(J(\underset{\sim}{x}_j,\rho_2))\}$$

then, we classify π_i in S_G if, and only if, $i \varepsilon S$, where $\underset{\sim}{x}_i$ denotes n observations from π_i, $\Phi(x)$ denotes cdf of the standard normal distribution and

$$J(\underset{\sim}{x}_i,\rho) = [(n\varphi_0^2\,\bar{x}_0+\lambda_0\sigma_0^2)/(\sigma_0^2+n\varphi_0^2)-(n\varphi_i^2\,\bar{x}_i+\lambda_i\sigma_i^2)/$$

$$(\sigma_i^2+n\varphi_i^2)+\rho]/[(\sigma_i^2\varphi_i^2)/(\sigma_i^2+n\varphi_i^2)+(\sigma_0^2\varphi_0^2)/(\sigma_0^2+n\varphi_0^2)]^{1/2}$$

for $i = 1,2,\ldots,k$.

(2.2.1) (ii) R_{2G}: Let $f_i=f(\alpha_i,\beta_i,\alpha_0,\beta_0,\rho_1) \equiv [\beta_i C(\alpha_i-\rho_1,\beta_i;$

$\alpha_0,\beta_0)/(2\pi(\beta_0+\beta_i))^{1/2}]+\alpha_i-\alpha_0+H(\alpha_i,\beta_i;\alpha_0,\beta_0;1,\rho_1)$

$-\alpha_i\Phi((\alpha_0-\alpha_i+\rho_1)/(\beta_0+\beta_i)^{1/2})$ for $i = 1,2,\ldots,k$, and $g_j = g(\alpha_j,\beta_j,\alpha_0,\beta_0,\rho_2) \equiv H(\alpha_j,\beta_j;\alpha_0,\beta_0;\ 1,\rho_2)$

$+\alpha_j\Phi((\alpha_0-\alpha_j+\rho_2)/(\beta_0+\beta_j)^{1/2})-\beta_j C(\alpha_j-\rho_2,\beta_j;\alpha_0,\beta_0)/$

$[2\pi(\beta_0+\beta_j)]^{1/2}$ for $j = 1,2,\ldots,k$.

If S is a subset of $\{1,2,\ldots,k\}$ and S^c is the complement of S such that

$$\alpha \sum_{i\epsilon S^c} f_i + \beta \sum_{j\epsilon S} g_j = \min_{T\subset\{1,2,\ldots,k\}} (\alpha \sum_{i\epsilon T^c} f_i+\beta \sum_{j\epsilon T} g_j),$$

then, $\pi_i \in S_G$ if, and only if, $i \in S$, where

$$C(\alpha,\beta;\ \gamma,\delta) = \exp[-(\alpha-\gamma)^2/2(\beta+\delta)] \tag{2.2.2}$$

$$\alpha_i = (n\varphi_i^2\ \bar{x}_i+\sigma_i^2\lambda_i)/(\sigma_i^2+n\varphi_i^2)$$

$$\beta_i = \sigma_i^2\varphi_i^2/(\sigma_i^2+n\varphi_i^2) \quad \text{for } i = 0,1,2,\ldots,k.$$

$$H(\alpha,\beta;\gamma,\delta;\ a,b) = \int_{-\infty}^{\infty} x\Phi(\alpha,\beta;ax+b)d\Phi(\gamma,\delta;x)$$

$\Phi(\alpha,\beta;x)$ is the cdf of normal distribution with mean α, variance β.

Proof: Since the proofs of (i) and (ii) are analogous and computations for (i) are easier than (ii), we thus outline the proof of (ii).

For a decision function d, the Bayes risk is given by

$$R(d,G) = E_G\{E_X L_2(d(X),\omega)\} \tag{2.2.3}$$

$$= \int_\Omega dG(\omega)\int_X L_2(d(x),\omega)\ f(x|\omega)dx\ .$$

By Fubini theorem, we have

$$R(d,G) = \int_X \varphi_G(d,x)dx$$

where

(2.2.4) $\varphi_G(d,x)= \int_\Omega L_2(d(x),\omega)f(x|\omega)dG(\omega)$.

To find a Bayes solution d_G, it suffices to require

$\varphi_G(d_G,x) \leq \varphi_G(d,x)$ for almost all $x \in X$ and any d.

Hence, we need

(2.2.5) $d_G(x) = S$ such that $\varphi_G(S,x) = \min\limits_{E\subset K} \varphi_G(E,x)$ for almost all $x \in X$. Now

$$\varphi_G(d,x) = \sum_{i=1}^{k} \int_\Omega L_{2i}(d,\omega)f(x|\omega)dG(\omega)$$

$$=\alpha \sum_{i\in S^c} \int_{\Omega_{1i}} (\theta_i-\theta_0)f(x_i|\omega)dG(\omega)+\beta \sum_{j\in S}\int_{\Omega_{2j}} (\theta_0-\theta_j)f(x_j|\omega)dG(\omega)$$

where $d(x) = S$ and

(2.2.6) $$\Omega_{1i} = \{(\theta_0,\theta_1,\ldots,\theta_k)|\theta_i-\theta_0 \geq \rho_1\}$$

$$\Omega_{2i} = \{(\theta_0,\theta_1,\ldots,\theta_k)|\theta_i-\theta_0 \leq \rho_2\}.$$

Let G_i and g_i denote respectively the cdf and pdf of θ_i. Then

(2.2.7) $$\varphi_G(d,x) = \alpha \sum_{i\in S^c} \prod_{j\neq i} \int_{-\infty}^{\infty} f(x_j|\theta_j)dG_j(\theta)\int_{-\infty}^{\infty}\int_{\theta_0+\rho_1}^{\infty} (\theta_i-\theta_0) f(x_i|\theta_i)dG_i(\theta)f(x_0|\theta_0)dG_0(\theta_0) + \beta \sum_{j\in S} \prod_{i\neq j} \int_{-\infty}^{\infty} f(x_i|\theta_i) dG_i(\theta)\int_{-\infty}^{\infty}\int_{-\infty}^{\theta_0+\rho_2} (\theta_0-\theta_j)f(x_j|\theta_j)dG_j(\theta)f(x_0|\theta_0)dG_0(\theta_0) .$$

We note that

$$(2.2.8)\quad \int_{-\infty}^{\infty} f(x_j|\theta_j)dG_j(\theta_j)=\int_{-\infty}^{\infty} f(x_j|\theta)g_j(\theta)d\theta=\int_{-\infty}^{\infty} g_j(\theta|x_j)f_{G_j}(x_j)d\theta,$$

where

$$(2.2.9)\quad f_{G_j}(x_j)=\int_{-\infty}^{\infty} f(x_j|\theta)dG_j(\theta)$$

$$g_j(\theta|x_j)=f(x_j|\theta)g_j(\theta)/f_{G_j}(x_j).$$

Through some calculation it is seen that $g_j(\theta|x_j)$ is a normal density with mean α_j and variance β_j which are defined by (2.2.2). We also note that

$$(2.2.10)\quad \int_{\theta_0+\rho_1}^{\infty} \theta f(x_i|\theta)dG_i(\theta)=\int_{\theta_0+\rho_1}^{\infty} \theta g_i(\theta|x_i)f_{G_i}(x_i)d\theta$$

$$=f_{G_i}(x_i)\{(\frac{\beta_i}{2\pi})^{1/2}\exp[-(\theta_0+\rho_1-\alpha_i)^2/2\beta_i]+\alpha_i[1-\Phi((\theta_0+\rho_1-\alpha_i)/\sqrt{\beta_i})]\}$$

where α_i and β_i are defined by (2.2.2).

We have then

$$(2.2.11)\quad \int_{-\infty}^{\infty} f(x_0|\theta_0)dG_0(\theta_0)\int_{\theta_0+\rho_1}^{\infty} \theta f(x_i|\theta)dG_i(\theta)$$

$$=f_{G_i}(x_i)f_{G_0}(x_0)\{\beta_i C(\alpha_i-\rho_1,\beta_i;\alpha_0,\beta_0)/[2\pi(\beta_0+\beta_i)]^{1/2}+\alpha_i$$

$$-\alpha_i\Phi((\alpha_0-\alpha_i+\rho_1)/(\beta_0+\beta_i)^{1/2})\}$$

where $C(\alpha,\beta;\gamma,\delta)$ is defined by (2.2.2).

Similarly, we have

$$(2.2.12)\quad \int_{-\infty}^{\infty} \theta_0 f(x_0|\theta_0)dG_0(\theta_0)\int_{\theta_0+\rho_1}^{\infty} f(x_i|\theta)dG_i(\theta)$$

$$=f_{G_i}(x_i)f_{G_0}(x_0)[\alpha_0-H(\alpha_i,\beta_i;\alpha_0,\beta_0;1,\rho_1)]$$

where

$$H(\alpha,\beta;\gamma,\delta;a,b) = \int_{-\infty}^{\infty} x\Phi(\alpha,\beta;ax+b)d\Phi(\gamma,\delta;x)$$

as defined by (2.2.2).

It follows from (2.2.11) and (2.2.12) that

$$\begin{aligned}
(2.2.13)\quad & \int_{-\infty\ \theta_0+\rho_1}^{\infty} (\theta-\theta_0)f(x_i|\theta)dG_i(\theta)f(x_0|\theta_0)dG_0(\theta_0) \\
&= f_{G_i}(x_i)f_{G_0}(x_0)\{\beta_i C(\alpha_i-\rho_1,\beta_i;\alpha_0,\beta_0)/[2\pi(\beta_0+\beta_i)]^{1/2}+\alpha_i \\
&\quad -\alpha_i\Phi((\alpha_0+\rho_1-\alpha_i)/(\beta_0+\beta_i)^{1/2}) - \alpha_0+H(\alpha_i,\beta_i;\alpha_0,\beta_0;1,\rho_1)\}.
\end{aligned}$$

Similarly, we also have

$$\begin{aligned}
(2.2.14)\quad & \int_{-\infty}^{\infty}\int_{-\infty}^{\theta_0+\rho_2} (\theta_0-\theta)f(x_i|\theta)f(x_0|\theta_0)dG_0(\theta_0)dG_i(\theta) \\
&= f_{G_i}(x_i)f_{G_0}(x_0)[H(\alpha_i,\beta_i;\alpha_0,\beta_0;1,\rho_2)-\beta_i C(\alpha_i-\rho_2,\beta_i;\alpha_0,\beta_0)/ \\
&\quad [2\pi(\beta_0+\beta_i)]^{1/2} + \alpha_i\Phi((\alpha_0-\alpha_i+\rho_2)/(\beta_0+\beta_i)^{1/2})\} .
\end{aligned}$$

It follows from (2.2.5), (2.2.7), (2.2.8), (2.2.13) and (2.2.14) that (ii) holds. This completes the proof.

B. Empirical Bayes Procedures

When the mean λ_i of the random variable associated with the prior distribution θ_i is unknown, the Bayes procedures R_{1G} and R_{2G} are not available. This is most the case in practical situations. However, the data itself reveals something about λ_i. Robbins [48] first proposed the idea of the asymptotic optimal rule which achieves the Bayes risk as sample size increases to infinity.

Deely [12] used this idea to treat subset selection problems in Bayes formulation. He assumed λ_i is finite but unknown while the variance φ_i^2 of θ_i is assumed known.

Here we treat the same problem in part A from an empirical Bayes approach using loss functions L_1 and L_2 respectively. We assume λ_i and φ_i^2 are both finite but unknown.

Let X be a random variable distributed according to cumulative distribution function $F(x;\theta)$ where θ belongs to certain parameter space, say, Θ on which a distribution G is assigned, G may or may not be known to us. Let $d(\cdot)$ be a decision function and $L(.,.)$ be a loss function defined on the sample space and $\mathcal{a} \times \Theta$ respectively ($\mathcal{a}$ is action space). Then the risk of d is defined by

$$R(d,G) = \int_{\Theta} \int_X L(d(x),\theta)dF(x;\theta)dG(\theta).$$

d_G is called a Bayes decision rule if

$$R(d_G,G) = \inf_d R(d,G)=R(G), \text{ say.}$$

We recall a definition of Robbins [48].

<u>Definition 2.2.1</u> Let $T=\{d_n\}$ be a sequence of decision functions such that $d_n(x) = d_n(x_1,x_2,\ldots,x_n;x)$ which is a function of x and whose form depends on the preceding n observations $x_1,x_2,\ldots,x_n$. Usually x is the present or new observation. Let

$$R_n(T,G) = \int_{\Theta} \int_X E(L(d_n(x_1,x_2,\ldots,x_n;x))dF(x;\theta)dG(\theta)$$

where the expectation is taken with respect to $x_1, x_2, \ldots, x_n$. If $\lim_{n\to\infty} R_n(T,G) = R(G)$, the Bayes risk, we say T is asymptotically optimal (a.o.) relative to G. We call such a.o. procedure an empirical Bayes procedure. Under the formulation and notation of our problem, we restate a result of Deely [12] as follows.

<u>Lemma 2.2.1</u> (Deely) Let $\underset{\sim}{X}_1, \underset{\sim}{X}_2 \ldots$ be independently identically distributed random vectors with $\underset{\sim}{X}_i$ having density $f(X|\omega)$ and consisting of (k+1)-components which are independent random variables. Let $\bar{G}_n$ be a cumulative distribution function on the parameter space Ω. Suppose $\bar{G}_n$ is a function of $\underset{\sim}{X}_1, \underset{\sim}{X}_2, \ldots, \underset{\sim}{X}_n$ and d_G denote a Bayes solution with respect to G. If

(i) $\lim_{n\to\infty} \bar{G}_n(\omega; \underset{\sim}{x}_1, \underset{\sim}{x}_2, \ldots, \underset{\sim}{x}_n) = G(\omega)$ wp1 for every continuity point ω of G where probability is taken with respect to $\underset{\sim}{x}_1, \underset{\sim}{x}_2, \ldots, \underset{\sim}{x}_n$.

(ii) $L(S,\omega)f(\underset{\sim}{x}|\omega)$ is continuous in ω and finite with respect to G for every $S \subset K$ and $x \in X$.

(iii) $\int_\Omega L(S,\omega)dG(\omega) < \infty$ for every $S \subset K$. Then,

$$T = \{d_{\bar{G}_n}\} \text{ is a.o. relative to } G.$$

We recall our notation that $\bar{x}_{in} = \bar{x}_i = \frac{1}{n}\sum_{j=1}^{n} x_{ij}$ and $S_{in}^2 = S_i^2 = \frac{1}{n-1}\sum_{j=1}^{n}(x_{ij}-\bar{x}_i)^2$ and $\Phi(x;\theta,\sigma^2)$ denotes the cdf of a normal distribution with mean θ and variance σ^2. Under the formulation of our problem we have

Lemma 2.2.2 Let $G = \prod_{i=0}^{k} \Phi(x_i;\lambda_i,\ \sigma_i^2+\varphi_i^2)$ and $\bar{G}_n = \prod_{i=0}^{k} \Phi(x_i;\bar{X}_i,S_i^2)$. Then, the condition (i) of lemma 2.2.1 is satisfied where $\omega = (\omega_1,\omega_2)$ with

$\omega_1 = \{(\lambda_0,\lambda_1,\ldots,\lambda_k);\lambda_i \in R\}$, $\omega_2 = \{(\sigma_0^2+\varphi_0^2,\sigma_1^2+\varphi_1^2,\ldots,\sigma_k^2+\varphi_k^2)\}$.

Proof: Firstly, we show for the case k=0. By simple calculation we see that X_{0i} are iid with unconditional normal density with mean λ_0 and variance $\sigma_0^2+\varphi_0^2$. It is well-known that $\bar{X}_{0n}$ and S_{0n}^2 are independent for every n and $\bar{X}_{0n} \to \lambda_0$ wpl and $S_{0n}^2 \to \sigma_0^2+\varphi_0^2$ wpl by the strong law of large numbers. We note that $\{S_{0n}^2;\ n = 1,2,\ldots\}$ is not an independent sequence of random variables and the strong law of large numbers can not be applied directly. Using the Helmert transformations we can achieve our goal.

Let

$$Y_{0i} = (X_{0i}-\lambda_0)/(\sigma_0^2+\varphi_0^2)^{1/2} \quad i = 1,2,\ldots .$$

Then, $\{Y_{0i},\ i = 1,2,\ldots\}$ are iid with standard normal density.
Also, $S_{0n}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_{0i}-\bar{X}_{0n})^2 = [(\sigma_0^2+\varphi_0^2)/(n-1)]\sum_{i=1}^{n}(Y_{0i}-\bar{Y}_{0n})^2$
$= (\sigma_0^2+\varphi_0^2)S_{0n}'^2$ where $S_{0n}'^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_{0i}-\bar{Y}_{0n})^2$.

Define

$$U_1 = (Y_{01}-Y_{02})/\sqrt{2}$$

$$U_2 = (Y_{01}+Y_{02}-2Y_{03})/\sqrt{6}$$

$$\vdots$$

$$U_i = (Y_{01}+Y_{02}+\dots+Y_{0i} - (i-1)Y_{0\,i+1})/\sqrt{i(i-1)}$$
$$\vdots$$
$$U_n = (Y_{01}+Y_{02}+\dots+Y_{0n})/\sqrt{n}\ .$$

We see that

$$(n-1)\ S_{0n}^{\prime 2} = \sum_{i=1}^{n-1} U_i^2 \quad \text{where } U_i \text{ are iid with } E(U_1^2) = 1.$$

Therefore,

$$S_{0n}^{\prime 2} = \frac{1}{n-1}\sum_{i=1}^{n-1} U_i^2 \to 1 \quad \text{wp1 by the strong law of large numbers.}$$

This shows $S_{0n}^2 \to \sigma_0^2+\varphi_0^2$ wp1. Since $\Phi(x;\alpha,\beta)$ is a continuous function of α and $\beta > 0$, we have then

(2.2.15) $\Phi(x;\ \bar{X}_{0n},\ S_{0n}^2) \to \Phi(x;\lambda_0,\ \sigma_0^2+\varphi_0^2)$ wp1 $\forall x \in R$.

Similarly, we also have

(2.2.16) $\Phi(x;\ \bar{X}_{in},\ S_{in}^2) \to \Phi(x;\lambda_i,\ \sigma_i^2+\varphi_i^2)$ wp1 $\forall x \in R$, for $i=1,2,\dots,k$.

Define

$$H(x_0,x_1,\dots,x_k) = \prod_{j=0}^{k} x_j \quad \text{for } x_i \geq 0,\ i = 1,2,\dots,k.$$

It is obvious that H is a continuous function of each variable. By usual (ε,δ)-argument and by (2.2.15) and (2.2.16) we can show that

$$\bar{G}_n = H(\bar{G}_{n0},\ \bar{G}_{n1},\dots,\bar{G}_{nk}) \to H(G_0,G_1,\dots,G_k) = G \quad \text{wp1}$$

where

$\bar{G}_{ni} = \Phi(x;\ \bar{X}_{in},\ S_{in}^2)$ and $G_i = \Phi(x;\lambda_i,\sigma_i^2+\varphi_i^2)$, $i = 0,1,2,\dots,k$.

This completes the proof.

Theorem 2.2.2 Under the formulation of our problem, the following rules $\tilde{R}_{1n}$ and $\tilde{R}_{2n}$, using loss functions L_1 and L_2, respectively, are a.o. with respect to G where G is any normal distribution with a finite mean and a finite variance.

(i) $\tilde{R}_{1n}$ partitions $\mathcal{P}$ into S_G and S_B such that

$S_G = \{\pi_i | i\varepsilon S\}$ and $S_B = \{\pi_j | j\notin S\}$ for some $S \subset K$, where S satisfies

$$\alpha \sum_{i\varepsilon S^c} [1-\Phi(J'(\underset{\sim}{x}_i;\rho_1)] + \beta \sum_{j\varepsilon S} \Phi(J'(\underset{\sim}{x}_j;\rho_2))$$
$$= \min_{T\subset\{1,2,\ldots,k\}} \{\alpha \sum_{i\varepsilon T^c} [1-\Phi(J'(x_i;\rho_1))]+\beta \sum_{j\varepsilon J} \Phi(J'(\underset{\sim}{x}_j;\rho_2))\},$$

where

$$J'(\underset{\sim}{x}_i;\rho) = \{[n(S_{0n}^2-\sigma_0^2)\bar{x}_{0n}+\lambda_0\sigma_0^2]/[\sigma_0^2+n(S_{0n}^2-\sigma_0^2)]$$
$$-[n(S_{in}^2-\sigma_i^2)\bar{x}_{in}+\lambda_i\sigma_i^2]/[\sigma_i^2+n(S_{in}^2-\sigma_i^2)]+\rho\}/$$
$$\{[\sigma_i^2(S_{in}^2-\sigma_i^2)]/[\sigma_i^2+n(S_{in}^2-\sigma_i^2)]+[\sigma_0^2(S_{0n}^2-\sigma_0^2)/$$
$$[\sigma_0^2+n(S_{0n}^2-\sigma_0^2)]\}^{1/2}.$$

(ii) $\tilde{R}_{2n}$ partitions $\mathcal{P}$ into S_G and S_B such that

$S_G = \{\pi_i | i\varepsilon S\}$ and $S_B = \{\pi_j | j\notin S\}$ for some $S \subset K$ where S satisfies

$$\alpha \sum_{i\varepsilon S^c} f_i + \beta \sum_{j\varepsilon S} g_j = \min_{T\subset\{1,2,\ldots,k\}} (\alpha \sum_{i\varepsilon T^c} f_i + \beta \sum_{j\varepsilon T} g_j)$$

where

$f_i = f(\tilde{\alpha}_i, \tilde{\beta}_i, \tilde{\alpha}_0, \tilde{\beta}_0, \rho_1)$ and $g_j = g(\tilde{\alpha}_j, \tilde{\beta}_j, \tilde{\alpha}_0, \tilde{\beta}_0, \rho_2)$

which are defined by (2.2.1) and (2.2.2) for i, j = 1,2,...,k, and where

$$\tilde{\alpha}_i = [n(S_{in}^2 - \sigma_i^2)\bar{X}_{in} + \sigma_i^2 \bar{X}_{in}]/[\sigma_i^2 + n(S_{in}^2 - \sigma_i^2)]$$
$$\tilde{\beta}_i = \sigma_i^2(S_{in}^2 - \sigma_i^2)/[\sigma_i^2 + n(S_{in}^2 - \sigma_i^2)] \quad \text{for } i = 1,2,\ldots,k.$$

Proof: It follows from Theorem 2.2.1, Lemma 2.2.1 and Lemma 2.2.2 that it suffices to check the conditions (ii) and (iii) of Lemma 2.2.1. It is obvious that

$$L_1(S,\omega)f(\underset{\sim}{x}|\omega) \leq (\alpha+\beta)\left(\frac{1}{2\pi}\right)^{\frac{k+1}{2}}\left(\prod_{i=0}^{k}\sigma_i^{-n}\right) < \infty \quad \forall\, x \in R^{n(k+1)}, S \subset K.$$

$$\int_\Omega L_1(S,\omega)dG(\omega) \leq \alpha+\beta < \infty$$

$\int_\Omega L_2(S,\omega)dG(\omega) \leq 2(\alpha+\beta)E_G(|\theta|) < \infty$ since $G_i(\theta)$ has finite absolute moment. Finally, we note that

$$L_2(S,\omega)f(\underset{\sim}{x}|\omega) \leq (\alpha+\beta) \max_{1\leq i\leq k} |\theta_i - \theta_0| < \infty \text{ a.s. with respect to } G.$$

2.3 Scale Parameter for Multivariate Case

Gnanadesikan [17] and Gnanadesikan and Gupta [18] consider the subset selection problem of k multivariate normal populations in terms of the generalized variance. Their papers study some approximations to the distributions of the generalized sample variance which are useful for our problem in this section. Here we consider partitioning of k multivariate normal populations with respect to a control in terms of the generalized variance. We treat the problem with single stage and multiple stage procedures.

A. Definitions and Notation

Let π_0, $\pi_1,\ldots,\pi_k$ be k+1 multivariate normal populations such that $N(\underset{\sim}{x};\underset{\sim}{\mu}_i,\Sigma_i)$ is the pxp multivariate distribution of π_i where $\underset{\sim}{\mu}_i$ and Σ_i are both unknown, $i = 0,1,2,\ldots,k$. Let $g(x,y) = y/x$ for $x > 0$ and $0 < \rho_1 < 1$, $\rho_2 > \rho_1$ and $\theta_i=|\Sigma_i|$. Then, according to (2.1.1) we have

$$\text{(2.3.1)} \qquad \begin{aligned} \mathcal{O}_G &= \{\pi_i \mid |\Sigma_i| \leq \rho_1|\Sigma_0|\} \\ \mathcal{O}_B &= \{\pi_j \mid |\Sigma_j| \geq \rho_2|\Sigma_0|\} \\ \mathcal{O}_I &= \{\pi_i \mid \rho_1|\Sigma_0| < |\Sigma_i| < \rho_2|\Sigma_0|\}. \end{aligned}$$

Let $\underset{\sim}{x}_{ij}$ denote the jth random observation of π_i for $i = 0,1,2,\ldots,k$, $j = 1,2,\ldots,n$ for some preassigned n. Based on these (k+1)n observations we need to partition the k populations into two disjoint exhaustive subsets S_G and S_B with respect to a control π_0 in terms of the generalized variance. For a preassigned $P^*((\frac{1}{2})^k < P^* < 1)$, we require that the probability of correct decision is at least P^*, where the correct decision (CD) is defined by Definition 2.1.1.

Let $\underset{\sim}{\Sigma} = \{\Sigma_0, \Sigma_1,\ldots,\Sigma_k\}$, $\underset{\sim}{\mu} = \{\underset{\sim}{\mu}_0, \underset{\sim}{\mu}_1,\ldots,\underset{\sim}{\mu}_k\}$ and $\Omega = \{\underset{\sim}{\Sigma},\underset{\sim}{\mu}\}$. Let S_i denote the sample covariance matrix of n observations from π_i and $|S_i|$ denote the determinant of S_i for $i = 0,1,2,\ldots,k$.

B. Single Stage Procedure

For a constant $C = C(n,k,P^*,p) > 0$, we define procedure $R_4 = R_4(C)$ as follows:

$$\pi_i \in S_G \text{ if } |S_i| < C|S_0| \text{ and}$$

$$\pi_j \in S_B \text{ if } |S_j| \geq C|S_0|.$$

It is clear then that the worst configuration with respect to R_4 is $\Sigma_0(q)$ ($0 \leq q \leq k$, integer q) i.e.

$$\inf_{\Omega} P\{CD|R_4\} = \inf_{\Sigma_0(q)} P\{CD|R_4\} \tag{2.3.2}$$

where

$$\Sigma_0(q) = \{|\Sigma_{ir}| = \rho_1|\Sigma_0|, |\Sigma_{is}| = \rho_2|\Sigma_0|;\ r=1,2,\ldots,q,\ s=q+1,\ldots,k\}.$$

Without loss of generality we may assume

$$\Sigma_0(q) = \{|\Sigma_i| = \rho_1|\Sigma_0|, |\Sigma_j| = \rho_2|\Sigma_0|; i=1,2,\ldots,q, j=q+1,\ldots,k\}. \tag{2.3.3}$$

we have then

$$\begin{aligned} P\{CD|R_4, \Sigma_0(q)\} &= P\{|S_i| < C|S_0|, |S_j| \geq C|S_0|, i=1,2,\ldots,q, \\ &\qquad j = q+1,\ldots,k|\Sigma_0(q)\} \\ &= P\{\frac{|S_i|}{|\Sigma_i|}(n-1)^p < C\frac{1}{\rho_1}\frac{|S_0|}{|\Sigma_0|}(n-1)^p, \frac{|S_j|}{|\Sigma_j|}(n-1)^p \\ &\quad \geq C\frac{1}{\rho_2}\frac{|S_0|}{|\Sigma_0|}(n-1)^p, i = 1,2,\ldots,q, j = q+1,\ldots,k\} \\ &= P\{A_i < C\frac{1}{\rho_1}A_0, A_j > C\frac{1}{\rho_2}A_0, i = 1,2,\ldots,q, j = q+1,\ldots,k\} \\ &\qquad \text{for } 0 \leq q \leq k, \end{aligned} \tag{2.3.4}$$

where

$A_i = \frac{|S_i|}{|\Sigma_i|}(n-1)^p$ has the same distribution as $\prod_{j=1}^{p} \chi^2(n-j)$ where $\chi^2(n-j)$ is chi-square distributed with d.f. n-j and $\chi^2(n-1), \chi^2(n-2), \ldots, \chi^2(n-p)$ are independent. We note that $A_0, A_1, \ldots, A_k$ are iid. This leads to the following

<u>Lemma 2.3.1</u> $\inf_{\Omega} P\{CD|R_4\} = \min_{0\leq q\leq k} P\{f_i < \lambda, g_j < \lambda,$ $i = 1,2,\ldots,q,\ j = q+1,\ldots,k\}$ where

$$\lambda = (\rho_2/\rho_1)^{1/2} > 1,\ f_i = A_i/A_0 \tag{2.3.5}$$

and $g_j = A_0/A_j$, where A_r are iid, each being the product of p independent factors, the jth factor being distributed as a chi-square variable with (n-j) degrees of freedom, $r = 0,1,\ldots,k$, $j = 1,2,\ldots,p$.

Proof: It follows from (2.3.2) and (2.3.4) that

$$\inf_{\Omega} P\{CD|R_4\} = \min_{0\leq q\leq k} P\{CD|R_4, \Omega = \Sigma_0(q)\}$$

$$= \min_{0\leq q\leq k} P\{f_i < C\frac{1}{\rho_1}, \frac{1}{g_j} > C\frac{1}{\rho_2},\ i = 1,2,\ldots,q;\ j = q+1,\ldots,k\}$$

where f_i and g_j are defined by (2.3.5). Take $C = (\rho_1\rho_2)^{1/2}$ and define $\lambda = (\rho_2/\rho_1)^{1/2}$ then, the result follows.

We note that n is involved in the degrees of freedom of the product factors of A_i.

In order to satisfy the P*-condition of the probability of correct decision, there are two approaches. The first one assumes n fixed and $\lambda^2 = \rho_2/\rho_1$ changing. For given n, P*-condition can always be satisfied by increasing λ. However, this does not satisfy our formulation. For fixed λ^2, we need to find the smallest n so that the P*-condition is satisfied.

We need a lemma which dues to Anderson [1] and Cramér [10].

<u>Lemma 2.3.2</u> (Anderson-Cramér) $\sqrt{n}\ [(|S_i|/|\Sigma_i|)-1]$ is asymptotically normally distributed with mean 0 and variance 2p.

It follows from lemma 2.3.2 that as n is large

$$P\{CD|R_4,\Sigma_0(q)\}=P\{\frac{\sqrt{n}}{\sqrt{2p}}\left(\frac{|S_i|}{|\Sigma_i|}-1\right) < \frac{C\sqrt{n}}{\rho_1\sqrt{2p}}\left(\frac{|S_0|}{|\Sigma_0|}-1\right)+\frac{C\sqrt{n}}{\rho_1\sqrt{2p}}$$

$$-\frac{\sqrt{n}}{\sqrt{2p}},\ \frac{\sqrt{n}}{\sqrt{2p}}\left(\frac{|S_j|}{|\Sigma_j|}-1\right) > \frac{C\sqrt{n}}{\rho_2\sqrt{2p}}\left(\frac{|S_0|}{|\Sigma_0|}-1\right)+\frac{C\sqrt{n}}{\rho_2\sqrt{2p}}-\frac{\sqrt{n}}{\sqrt{2p}} \quad \text{for}$$

$$i = 1,2,\ldots,q,\ j = q+1,\ldots,k\}$$

$$\simeq P\{X_i-\lambda X_0 < \frac{\sqrt{n}(C-\rho_1)}{\sqrt{2p}\ \rho_1},\ \lambda X_j-X_0 > \frac{-\lambda\sqrt{n}(\rho_2-C)}{\sqrt{2p}\ \rho_2},\ i = 1,2,\ldots,q,$$

$$j = q+1,\ldots,k\},$$

where $X_0,X_1,\ldots,X_k$ are iid with standard normal cdf and $\lambda = (\rho_2/\rho_1)^{1/2} > 1$. Let $Z_i = X_i-\lambda X_0$, $Z_j = X_0-\lambda X_j$, $1 \leq i \leq q$, $q+1 \leq j \leq k$ and take $C = \sqrt{\rho_1\rho_2}$.

Then, when n is large, we have

<u>**Theorem 2.3.1**</u> $\inf\limits_{\Omega} P\{CD|R_4\} \simeq \min\limits_{0\leq q\leq k} P\{Z_i < \sqrt{\frac{n}{2p}}\,(\lambda-1),\ i=1,\ldots,k\}$

where Z_i are identically distributed with common cdf $\Phi(0,1+\lambda^2)$ such that $\mathrm{Cov}(Z_r,Z_s) = \lambda^2$ for $1 \leq r,\ s \leq q$, $\mathrm{Cov}(Z_r,Z_s) = 1$ for $q+1 \leq r,s \leq k$, $\mathrm{Cov}(Z_r,Z_s) = \lambda \quad 1 \leq r \leq q,\ q+1 \leq s \leq k,$

or equivalently, we have

$$\inf_{\Omega} P\{CD|R_4\} \simeq \min_{0\leq q\leq k} \int_{-\infty}^{\infty} \Phi^q(\lambda x+\sqrt{\frac{n}{2p}}\,(\lambda-1))[1-\Phi(\frac{x}{\lambda}+\frac{\sqrt{n}(\lambda-1)}{\lambda\sqrt{2p}})]^{k-q}d\Phi(x).$$

Proof: The proof is obvious and thus omitted.

Remark 2.3.1 (i) Since $\lambda > 1$, it is conjectured that minimum value attains at $q=k$.

(ii) When p is large, the approximation is good. When p is small, we have the following exact and conservative results.

(1) Case $p=1$

$$A_i = \chi^2(n-1) \quad \text{for } i = 0,1,2,\ldots,k\ .$$

It follows directly from lemma 2.3.1 that

$$\text{(2.3.6)} \quad \inf_{\Omega} P\{CD|R_4\} = \min_{0\leq q\leq k} \int_0^{\infty} G^q_{n-1}(\lambda x)[1-G_{n-1}(\tfrac{x}{\lambda})]^{k-q}g_{n-1}(x)dx$$

where $G_{n-1}(x)$ and $g_{n-1}(x)$ are respectively the cdf and pdf of $\chi^2(n-1)$.

(2.3.7) Let $I(n;p) = \int_0^p x^{n-1}e^{-x}dx$ be the incomplete Gamma function and

$$J(n;p) = \Gamma(n)-I(n;p).$$

Then

$$G_{n-1}(\lambda x) = I(\tfrac{n-1}{2}\ ;\ \tfrac{\lambda x}{2})/\Gamma(\tfrac{n-1}{2}).$$

Hence, we have

$$(2.3.8)\quad \inf_{\Omega} p\{CD|R_4\} = \min_{0\leq q\leq k} \int_0^{\infty} I^q(\tfrac{n-1}{2};\ \tfrac{\lambda x}{2})J^{k-q}(\tfrac{n-1}{2};\ \tfrac{x}{2\lambda})\frac{g_{n-1}(x)dx}{[\Gamma(\frac{n-1}{2})]^k}$$

Since I(n;p) and J(n;p) are increasing and decreasing in p respectively, it follows thus

$$(2.3.9)\quad \inf_{\Omega} p\{CD|R_4\} = \min_{0\leq q\leq k} \left(\int_0^{x_0} + \int_{x_0}^{\infty}\right) I^q(\tfrac{n-1}{2};\ \tfrac{\lambda x}{2})J^{k-q}(\tfrac{n-1}{2};\ \tfrac{x}{2\lambda}) \times \frac{g_{n-1}(x)dx}{[\Gamma(\frac{n-1}{2})]^k}$$

$$\geq \int_0^{x_0} I^k(\tfrac{n-1}{2};\ \tfrac{\lambda x}{2})\ \frac{g_{n-1}(x)}{\Gamma^k(\frac{n-1}{2})}\,dx + \int_{x_0}^{\infty} J^k(\tfrac{n-1}{2};\ \tfrac{x}{2\lambda})\ \frac{g_{n-1}(x)}{\Gamma^k(\frac{n-1}{2})}\,dx$$

where x_0 is uniquely defined by

$$(2.3.10)\qquad I(\tfrac{n-1}{2};\ \tfrac{\lambda x_0}{2}) = J(\tfrac{n-1}{2};\ \tfrac{x_0}{2\lambda}).$$

Define

$$(2.3.11)\quad L(k,n,\lambda,x_0) = \frac{1}{\Gamma^k(\frac{n-1}{2})}\left\{\int_0^{x_0} I^k(\tfrac{n-1}{2};\ \tfrac{\lambda x}{2})g_{n-1}(x)dx + \int_{x_0}^{\infty} J^k(\tfrac{n-1}{2};\ \tfrac{x}{2\lambda})\ g_{n-1}(x)dx\right\}$$

where x_0 is defined by (2.3.10).

Let $s = s(n)$ denote the unique value such that

(2.3.12) $I(\frac{n-1}{2};\ s) = \frac{1}{2}\Gamma(\frac{n-1}{2})$. Then, since $\lambda > 1$, we see that

$J(\frac{n-1}{2};\ \frac{1}{2\lambda}(\frac{2}{\lambda}s)) = J(\frac{n-1}{2};\ \frac{s}{\lambda^2}) \geq J(\frac{n-1}{2};\ s) = I(\frac{n-1}{2};\ s)$. This shows that $x_0 > \frac{2}{\lambda} s$. On the other hand, we see that

$J(\frac{n-1}{2}; \frac{1}{2\lambda}(2\lambda s)) = I(\frac{n-1}{2}; s) < I(\frac{n-1}{2}; \lambda^2 s) = I(\frac{n-1}{2}; \frac{\lambda}{2}(2\lambda s))$. This shows that $x_0 < 2\lambda s$. Hence, we can conclude that

(2.3.13) $x_0 \,\varepsilon (\frac{2}{\lambda} s, 2\lambda s)$ where x_0 and s are defined respectively by (2.3.10) and (2.3.12).

(2) Case p=2

$$A_i = \chi^2(n-1)\chi^2(n-2) \quad \text{for } i = 0,1,2,\ldots,k.$$

It is well-known that $2A_i^{1/2}$ is distributed according to the law of the cdf of $\chi^2(2n-4)$. Therefore, it follows from the Lemma 2.3.1 that

$$(2.3.14) \quad \inf_{\Omega} P\{CD|R_4\} = \min_{0\leq q\leq k} \int_0^\infty G_{2n-4}^q(\sqrt{\lambda}x)[1-G_{2n-4}(\frac{x}{\sqrt{\lambda}})]^{k-q} g_{2n-4}(x)dx$$

where $G_{2n-4}(x)$ and $g_{2n-4}(x)$ are cdf and pdf of $\chi^2(2n-4)$. Using same definitions of I and J, we can conclude that

$$(2.3.15) \quad \inf_{\Omega} P\{CD|R_4\} = \min_{0\leq q\leq k} \int_0^\infty I^q(n-2; \frac{\sqrt{\lambda}}{2}x)J^{k-q}(n-2; \frac{x}{2\sqrt{\lambda}})\frac{g_{2n-4}(x)dx}{\Gamma^k(n-2)}$$

$$\geq L(k, 2n-3, \lambda, x_1)$$

where L is defined by (2.3.11) and x_1 is defined by

(2.3.16) $I(n-2; \frac{\sqrt{\lambda}x_1}{2}) = J(n-2; \frac{x_1}{2\sqrt{\lambda}})$. It is shown by (2.3.13) that

(2.3.17) $x_1 \varepsilon (\frac{2}{\sqrt{\lambda}} S_1, 2\sqrt{\lambda} S_1)$ where S_1 satisfies $I(n-2; S_1) = \frac{1}{2}\Gamma(n-2)$.

It follows from (2.3.6)-(2.3.17) that we conclude in the following

<u>Corollary 2.3.1</u> Let $C = (\rho_1 \rho_2)^{1/2}$ and $\lambda = \frac{\rho_2}{\rho_1}$.

(a) When p=1, the infimum of the probability of correct selection is given by (2.3.6).

If n is the smallest integer such that $L(k,n,\lambda,x_0) \geq P^*$ then,

$\inf_{\Omega} P\{CD|R_4\} \geq P^*$ where $L(k,n,\lambda,x_0)$, x_0 are defined by (2.3.11), (2.3.10), (2.3.12) and (2.3.13).

(b) When p=2, the inf of PCS is given by (2.3.14).

If n is the smallest positive integer such that $L(k,2n-3,\lambda,x_1) \geq P^*$, then, $\inf_{\Omega} P\{CD|R_4\} \geq P^*$ where x_1 is defined by (2.3.16), (2.3.17).

(3) Case $p \geq 3$, Hoel [28] suggested approximating the distribution of $A_i^{1/p}$ by the distribution of Y having density

$$g(y) = \eta^{1/2p(n-p)} y^{[1/2p(n-p)-1]} e^{-\eta y} / \Gamma(\tfrac{p}{2}(n-p)) \tag{2.3.18}$$

where $\eta = (p/2)[1-(p-1)(p-2)/2n]^{1/p}$.

We see that when p=1 and p=2, the approximations are exact. Gnanadesikan and Gupta [18] made a study of this approximation by generating random samples from the Gamma distribution and comparing the distribution of the variate $A_i^{1/p}$ generated from these samples which obey the distribution law of (2.3.18). They found that the Hoel's approximation decreases in accuracy as p increases. When p=3, they suggested using F-distribution. The approximation is found improving with n. Thus, when p=3, we have

$$\inf_{\Omega} P\{CD|R_4\} \simeq \min_{0 \leq q \leq k} \int_0^{\infty} G_{3n-9}(3\sqrt{\lambda}x)\left[1-G_{3n-9}\left(\frac{x}{3\sqrt{\lambda}}\right)\right]^{k-q} dG_{3n-9}(x) \tag{2.3.19}$$

where $G_{3n-9}(x)$ is the cdf of $\chi^2(3n-9)$.

When p is bigger than 3, it is favorable to use the approximation given by Theorem 2.3.1.

C. A Minimax Property of R_4

Let R_p denote the p-dimensional Euclidean space and R_p^{k+1} denote the (k+1) product space of R_p. For $c > 0$, we define a transformation T_c from R_p^{k+1} into R_p^{k+1} such that a point $\underset{\sim}{x}=(\underset{\sim}{x}_0,\underset{\sim}{x}_1,\ldots,\underset{\sim}{x}_k)$ in R_p^{k+1} is transformed into $T_c\ \underset{\sim}{x} = (c\underset{\sim}{x}_0,c\underset{\sim}{x}_1,\ldots,c\underset{\sim}{x}_k)$ in R_p^{k+1} where $\underset{\sim}{x}_i$ is in R_p $(i=0,1,\ldots,k)$. By defining the usual operations on the set of all T_c, $G = \{T_c;\ c > 0\}$ becomes a non-trivial group. Let $|\Sigma| = (|\Sigma_0|,|\Sigma_1|,\ldots,|\Sigma_k|)$ denote the vector value of the generalized variance of k+1 multivariate normal populations. Then, G induces $\bar{G}$, a group of transformations on the R_{k+1}^+, (the positive quadrant of (k+1)-dimensional Eucleadian space) such that, for $\bar{T}_c \in \bar{G}$, $\bar{T}_c(|\Sigma_0|,|\Sigma_1|,\ldots,|\Sigma_k|) = (c^2|\Sigma_0|,c^2|\Sigma_1|,\ldots,c^2|\Sigma_k|)$. Define $R_i = |S_i|/|S_0|$, for a fixed sample size n, where S_i are sample variance for $i=0,1,2,\ldots,k$. Then, the vector $\underset{\sim}{R} = (R_1,R_2,\ldots,R_k)$ is maximal invariant with respect to G and also $\underset{\sim}{\theta} = (|\Sigma_1|/|\Sigma_0|,|\Sigma_2|/|\Sigma_0|,\ldots,|\Sigma_k|/|\Sigma_0|)$ is maximal invariant with respect to $\bar{G}$. Therefore, it follows that (see, for example, [33]) the distribution of $\underset{\sim}{R}$ depends only on $\underset{\sim}{\theta}$. Furthermore (see, for example, [33]), each rule in the class of all invariant decision rules under G must be a function of $\underset{\sim}{R}$ and thus its distribution depends on $|\Sigma_0|,|\Sigma_1|,\ldots,|\Sigma_k|$ only through $\underset{\sim}{\theta}$. Let

$$(2.3.20)\qquad \theta_i = |\Sigma_i|/|\Sigma_0| \text{ and } \psi = \{\underset{\sim}{\theta}|\underset{\sim}{\theta}=(\theta_1,\theta_2,\ldots,\theta_k)\}.$$

Define loss functions L_i by

$$L_i(.,.) = 1 \quad \text{if } \pi_i \text{ is misclassified}$$
$$= 0 \quad \text{otherwise}$$

for $i = 1,2,\ldots,k$.

Then, the risk function r is given by

$$(2.3.21) \qquad r(S,\omega) = E \sum_{i=1}^{k} L_i(S,\omega) = \sum_{i=1}^{k} P\{\pi_i \text{ is misclassified } |R_4\}$$

where $\omega = \{\underset{\sim}{\Sigma},\underset{\sim}{\mu}\}$ and $S \subset K$ such that $i \in S$ implies $\pi_i \in S_G$. By Lemma 2.3.1, we see that

$$(2.3.22) \qquad r(S,\omega) \leq k[1-H(\lambda,n)]$$

where

$$H(\lambda,n) = \min_{0\leq q\leq k} P\{f_i < \lambda, g_j < \lambda,\ i=1,2,\ldots,q,\ j=q+1,\ldots,k\}$$

which is defined by (2.3.5).

We note that the equality of (2.3.22) holds when θ_i is either ρ_1 or ρ_2(because of (2.3.2)). We define a prior distribution Q on ψ (which is defined by (2.3.20)) such that

(2.3.23) $Q = Q_1 x Q_2 x \ldots x Q_k$, a product probability measure of each Q_i, with $Q_i(\{\rho_1\})=Q_i(\{\rho_2\}) = 1/2$. Then, it is obvious that $Q(\{\theta\}) = (1/2)^k \quad \forall\theta\epsilon\psi_0$, where

$$(2.3.24) \qquad \psi_0 = \{\underset{\sim}{\theta}_0 | \underset{\sim}{\theta}_0 = (\theta_1,\theta_2,\ldots,\theta_k),\ \theta_i \text{ is either } \rho_1 \text{ or } \rho_2\}.$$

Then, $\psi_0 \subset \psi$ and $Q(\psi_0) = 1$ and

$$r(R_4,\omega_0) = k[1-H(\lambda,n)] \geq r(R_4,\omega) \quad \text{for any } \omega, \text{ where}$$

$\omega_0 = \{\psi_0,\underset{\sim}{\mu}\}$. Therefore, we have

(2.3.25) $r(R_4,\omega_0) = \sup_{\omega\varepsilon\psi} r(R_4,\omega) \ \forall\ \omega_0\varepsilon\psi_0$ with $Q(\psi_0) = 1$.

Then, it can be concluded (see, for example, [34]) that

Theorem 2.3.2 R_4 is minimax in the class of invariant rules.

D. Sequential Procedures (for p = 1,2)

Finally, we treat the problem formulated in 2.3 for the case of univariate and bivariate normal populations by using truncated (closed) sequential procedures. These procedures control the probabilities of misclassification and possess the monotonicity property which will be defined later.

Procedure R_5 $(m_0,m_1;\ m_2,m_3;\ A_n,B_n;\ \lambda)$

For given $\rho_0(\rho_0 < 1)$, $\rho_1(\rho_1 > \rho_0)$ and P^* $((\frac{1}{2})^k < P^* < 1)$, let λ be a value such that $1 < \lambda < \min(\rho_1^2,\ 1/\rho_0^2)$ and let $\alpha = (1-P^*)/2k$ and let $\beta_n = \alpha^{1/n-1}$ $(n \geq 2)$.

Define

$$A_n = \rho_1^2(\lambda\beta_n-1)/\lambda(\lambda-\beta_n) \tag{2.3.26}$$

$$B_n = \rho_0^2\lambda(\lambda-\beta_n)/(\lambda\beta_n-1)$$

$$m_0 = [\ln(1/\alpha)/\ln\lambda] + 2$$

$$m(A) = \left[\frac{\ln(1/\alpha)}{\ln(\lambda(\rho_1^2+1)/(\lambda^2+\rho_1^2))}\right] + 1$$

$$m(B) = \left[\frac{\ln(1/\alpha)}{\ln((\lambda+\rho_0^2\lambda)/(\rho_0^2\lambda^2+1))}\right] + 1$$

where [x] denotes the largest integer not exceeding x.

(2.3.27) $m_1 = \min\{m(A), m(B)\}$

$m_2 = \min\{n | A_n \geq B_n \text{ for } m_1 \leq n \leq m_3\}$

$m_3 = \max\{m(A), m(B)\}$.

For convenience, we use $R_5(\lambda)$ instead of $R_5(m_0,m_1;m_2,m_3;A_n,B_n;\lambda)$.

$R_5(\lambda)$ is defined as follows in two cases:

(i) when $m_1=m(A)$

(1) We note that on each stage of sampling, one observation is always drawn from π_0. Draw m_0 observations from each of k populations. If

$$S^2_{im_0} < A_{m_0} S^2_{0m_0}, \text{ put } \pi_i \text{ in } S_G.$$

If $$S^2_{jm_0} > B_{m_0} S^2_{0m_0}, \text{ put } \pi_j \text{ in } S_B.$$

If all k populations are classified, stop sampling and the disjoint exhaustive classes S_G and S_B are obtained. Otherwise, draw one more observation from those populations which are not yet classified. If π_i is not classified in the first stage and $S^2_{im_0+1} < A_{m_0+1} S^2_{0m_0+1}$, classify π_i in S_G, or if $S^2_{im_0+1} > B_{m_0+1} S^2_{0m_0+1}$, classify $_i$ in S_B. Otherwise, continue sampling on the third stage. This procedure continues until all populations are classified.

(2) On stage (m_1-m_0+1), take one more observation from these populations which are not yet classified. If π_i is sampled on this stage and $S^2_{im_1} < S^2_{0m_1}$, classify π_i in S_G and classify π_i in S_B if $S^2_{im_1} > B_{m_1} S^2_{0m_1}$.

(3) If all k populations are not yet classified after the stage (m_1-m_0+1), take (m_2-m_1) observations from these populations which are not classified. Then, classify π_i in S_B if $S^2_{im_2} > B_{m_2} S^2_{0m_2}$. Same sampling procedure stated in (1) continues until at stage (m_3-m_0+1).

(4) On stage (m_3-m_0+1), one more observation is drawn from these populations which are not yet classified. Then, classify π_i in S_B if $S^2_{im_3} > S^2_{0m_3}$, otherwise, classify π_i in S_G.

(ii) When $m_1 = m(B)$

(1)' This part follows (1) of (i).

(2)' At stage (m_1-m_0+1), take one more observation from these which are not yet classified. Classify π_i in S_B if $S^2_{im_1} > S^2_{0m_1}$ and classify π_i in S_G if $S^2_{im_1} < A_{m_1} S^2_{0m_1}$.

(3)' If sampling procedure does not stop after (m_1-m_0+1)th stage, take (m_2-m_0) observations from those which are not classified. Classify π_i in S_G if $S^2_{im_2} < A_{m_2} S^2_{0m_2}$. Same sampling procedure stated in (1)' continues until the stage (m_3-m_0+1).

(4)' At stage (m_3-m_0+1), one more observation is drawn from those which are not classified. Classify π_i in S_G if $S^2_{im_3} < S^2_{0m_3}$. otherwise, classify π_i in S_B.

According to the preceding sampling procedure, we have the following

<u>Theorem 2.3.3</u> Let p=1 and λ satisfies $1 < \lambda < \min(\rho_1^2, 1/\rho_0^2)$.

(a) $P\{CD|R_5(\lambda)\} \geq P^*$.

(b) (Monotonicity property) If $\sigma_i^2 \leq \sigma_j^2 < \rho_0\sigma_0^2$, then

$P\{\pi_i \varepsilon S_G|R_5(\lambda)\} \geq P\{\pi_j \varepsilon S_G|R_5(\lambda)\}$. Also, if $\sigma_i^2 \geq \sigma_j^2 > \rho_1\sigma_0^2$, then,

$P\{\pi_i \varepsilon S_B|R_5(\lambda)\} \geq P\{\pi_j \varepsilon S_B|R_5(\lambda)\}$.

Before we prove this theorem, we state a result due to Cox [9] and Paulson [40].

<u>Lemma 2.3.3</u> (Cox-Paulson) Let $g_n(f|\phi^2) = \dfrac{\Gamma(n-1)}{\Gamma^2(\frac{n-1}{2})} \times \dfrac{f^{(n-3)/2}}{\phi^{n-1}(1+f/\phi^2)^{n-1}}$, the density of f. Then, for $\gamma > 1$,

$$P\{\frac{g_n(f|\phi^2)}{g_n(f|\phi^2/\gamma^2)} < \alpha \text{ for at least one } n,\ n = 2,3,\ldots\} \leq \alpha.$$

Proof of Theorem 2.3.3. (a) We give a proof for the case $m_1 = m(A)$. The proof for the case $m_1 = m(B)$ is analogous. According to $R_5(\lambda)$, a population π_i is misclassified if (i) $\pi_i \varepsilon \mathscr{G}_G$ and $S_{in}^2 > B_n S_{0n}^2$ for some n, where either $m_0 \leq n \leq m_1$ or $m_2 \leq n \leq m_3$, or, (ii) $\pi_i \varepsilon \mathscr{G}_B$ and $S_{in}^2 < A_n S_{0n}^2$ for some n, where $m_0 \leq n \leq m_1$, or $S_{in}^2 \leq B_n S_{0n}^2$ for all n, $m_2 \leq n \leq m_3$. We see that

(2.3.28) $P\{S_{in}^2 > B_n S_{0n}^2$ for some n, $m_0 \le n \le m_1$ or $m_2 \le n \le m_3 | \sigma_i^2 < \rho_0 \sigma_0^2\}$

$$= P\{\frac{(S_{in}^2/\sigma_i^2)}{(S_{0n}^2/\sigma_0^2)} \cdot (\frac{\sigma_i^2}{\sigma_0^2}) > B_n \text{ for some } n,\ m_0 \le n \le m_1 \text{ or } m_2 \le n \le m_3 | \frac{\sigma_i^2}{\sigma_0^2} < \rho_0\}$$

$\le P\{f(n-1,n-1) > B_n/\rho_0$ for at least one n, $n=2,3,\ldots\}$

$\le \alpha$ (By taking $\phi^2 = \rho_0$ and $\gamma = \lambda$ and using (2.3.26) and Lemma 2.3.3) where $f(n-1,n-1)$ is F-distributed with d.f. $n-1$, $n-1$.

(2.3.29) $P\{S_{in}^2 < A_n S_{0n}^2$ for some n, $m_0 \le n \le m_1$, or $S_{in}^2 \le B_n S_{0n}^2$ for all n, $m_2 \le n \le m_3 | \sigma_i^2 > \rho_1 \sigma_0^2\}$

$$= P\{\frac{(S_{in}^2/\sigma_i^2)}{(S_{0n}^2/\sigma_0^2)} \cdot (\frac{\sigma_i^2}{\sigma_0^2}) < A_n \text{ for some } n,\ m_0 \le n \le m_1 \text{ or}$$

$$\frac{(S_{in}^2/\sigma_i^2)}{(S_{0n}^2/\sigma_0^2)} (\frac{\sigma_i^2}{\sigma_0^2}) \le B_n \text{ for some } n,\ m_2 \le n \le m_3 | \sigma_i^2/\sigma_0^2 > \rho_1\}$$

$\le P\{f(n-1,n-1) < A_n/\rho_1$ for some n, $m_0 \le n \le m_1$ or $m_2 \le n \le m_3\}$

$\le P\{f(n-1,n-1) < A_n/\rho_1$ for some n, $n=2,3,\ldots\}$

$\le \alpha$ (by noting that $B_n \le A_n$ for $m_2 \le n \le m_3$ and $1 < \lambda < \min(\rho_1^2, 1/\rho_0^2)$ and using Lemma 2.3.3).

It follows from (2.3.28) and (2.3.29) that

$$P\{CD|R_5(\lambda)\} = 1-P\{\pi_i \text{ is misclassified for } i=1,2,\ldots,k|R_5(\lambda)\}$$

$$\geq 1-\sum_{i=1}^{k} P\{\pi_i \text{ is misclassified } |R_5(\lambda)\}$$

$$\geq 1-2k\ \alpha$$

$$= 1-(1-P^*)$$

$$= P^*.$$

Finally, we note that A_n and B_n are respectively monotone increasing and decreasing functions of n and also $A_{m_0} \leq 1$ and $B_{m_0} \geq 1$. Hence the definitions of (2.3.26) and (2.3.27) are well-defined. This completes the proof of (a).

(b) It suffices to show the case $\sigma_i^2 \leq \sigma_j^2 < \rho_0\sigma_0^2$ since the proof of other case is quite similar. For the case $m_1=m(A)$, we note that

$$P\{\pi_j\varepsilon S_G|R_5(\lambda),\ \sigma_i^2 \leq \sigma_j^2 < \rho_0\sigma_0^2\}$$

$$= P\{S_{jn}^2 < A_n\ S_{0n}^2 \text{ for some n, } m_0 \leq n \leq m_1, \text{ or } S_{jn}^2 \leq B_n\ S_{0n}^2 \text{ for every n, } m_2 \leq n \leq m_3|\rho_0 > \sigma_j^2/\sigma_0^2 \geq \sigma_i^2/\sigma_0^2\}$$

$$= P\{f(n-1,n-1) < A_n(\sigma_0^2/\sigma_j^2) \text{ for some n, } m_1 \leq n \leq m_2 \text{ or,}$$

$$f(n-1,n-1) < B_n(\sigma_0^2/\sigma_j^2) \text{ for every n, } m_2 \leq n \leq m_3|\frac{\sigma_0^2}{\sigma_i^2} \geq \frac{\sigma_0^2}{\sigma_j^2} \geq \frac{1}{\rho_0}\}$$

$$\leq P\{f(n-1,n-1) < A_n(\sigma_0^2/\sigma_i^2) \text{ for some n, } m_1 \leq n \leq m_2 \text{ or,}$$

$$f(n-1,n-1) < B_n(\sigma_0^2/\sigma_i^2) \text{ for every n, } m_2 \leq n \leq m_3\}$$

$$= P\{\pi_i\varepsilon S_G|R_5(\lambda),\ \sigma_i^2 \leq \sigma_j^2 < \rho_0\sigma_0^2\}$$

This completes the proof of (b).

When p=2, we note that $A_{in} = (|S_{in}^2|/|\Sigma_i|)(n-1)^2$ $= \chi^2(n-1)\chi^2(n-2)$ and $2A_{in}^{1/2} = \chi^2(2n-4)$. Hence, $(|S_{in}^2|/|S_{on}^2|)^{1/2} \times (|\Sigma_0|/|\Sigma_i|)^{1/2} = 2A_{in}^{1/2}/2A_{i0}^{1/2} = f(2n-4, 2n-4)$, the random variable distributed according to F-distribution with d.f. 2n-4, 2n-4.

Define

$$(2.3.30)\quad \begin{aligned} A'_n &= A_{2n-2} \\ B'_n &= B_{2n-2} \\ m'_0 &= [\ln(1/\alpha)/\ln\lambda']+2 \text{ where } 1 < \lambda' < \min(\rho_1, \frac{1}{\rho_0}) \\ m'(A) &= \left[\frac{\ln(1/\alpha)}{\ln\left(\frac{\lambda'(\rho_1+1)}{\lambda'^2+\rho_1}\right)}\right]+1 \\ m'(B) &= \left[\frac{\ln(1/\alpha)}{\ln\left(\frac{\lambda'+\rho_0\lambda'}{\lambda'^2\rho_0+1}\right)}\right]+1 \\ m'_1 &= \min\{m'(A), m'(B)\} \\ m'_3 &= \max\{m'(A), m'(B)\} \\ m'_2 &= \min\{n \mid A'_n \geq B'_n,\ m'_1 \leq n \leq m'_3\} . \end{aligned}$$

Instead of S_{in}^2 used in case p=1, we use $|S_{in}^2|$, the generalized sample variance, for our case p=2. For convenience, we use $R'_5(\lambda')$ to denote $R_5(m'_0, m'_1;\ m'_2, m'_3;\ A'_n, B'_n; \lambda')$. Then, it follows from Theorem 2.3.3 and (2.3.30) that we have an immediate result as follows.

<u>Corollary 2.3.2</u> (a) For bivariate normal populations, we have $P\{CD|R_5'(\lambda')\} \geq P^*$ where $1 < \lambda' < \min(\rho_1, 1/\rho_0)$.

(b) (Monotonicity property) If $|\Sigma_i| \leq |\Sigma_j| < \rho_0 |\Sigma_0|$, then

$P\{\pi_i \varepsilon S_G|R_5'(\lambda')\} \geq P\{\pi_j \varepsilon S_G|R_5'(\lambda')\}$. Also, if

$|\Sigma_i| \geq |\Sigma_j| > \rho_1 |\Sigma_0|$, then,

$P\{\pi_i \varepsilon S_B|R_5'(\lambda')\} \geq P\{\pi_j \varepsilon S_B|R_5'(\lambda')\}$.

CHAPTER 3

SELECTION APPROACH TO A K-ARMED-BANDIT PROBLEM

3.0 Introduction and Summary

Robbins [47] proposed a finite memory sampling scheme for tossing one of two coins sequentially. For his scheme the average number of heads tends to the largest probability of a head of the two coins when the length of memory tends to infinity. Later, many authors studied this two-armed-bandit problem extensively and thoroughly. Smith and Pyke [50] constructed a class of rules for this problem. They considered the problem from a more general point of view and gave an insight into the possible constructions of optimal rules. All these rules treat the case of infinite tossings. Bradt, Johnson and Karlin [7] considered the problem of finite tosses of two coins assuming a non-trivial prior probability on the parameter space. They showed some optimal properties and discussed some applications.

In this chapter we study a problem of the k-armed-bandit in a more general formulation by using selection approach. We investigate a maximin procedure for finite numbers of sampling of k populations without assuming prior distribution. We also show an asymptotic optimality of the maximum procedure.

3.1 Notation and Formulation of the Problem

Let $\pi_1, \pi_2, \ldots, \pi_k$ be k populations such that π_i has cdf $F(x;\theta_i)$ for $i = 1,2,\ldots,k$. Let Ω denote the parameter space of $\theta = (\theta_1,\theta_2,\ldots,\theta_k)$. For $d > 0$, we define $\Omega(d) = \{\theta = (\theta_1,\theta_2,\ldots,\theta_k) | \theta_{[k]} - \theta_{[k-1]} \geq d\}$. Let X_{ij} denote the jth independent observation from π_i and define $S_{in} = \sum_{j=1}^{n} X_{ij}$, for $i = 1,2,\ldots,k$, $n = 1,2,\ldots$. Let $W(k,n; X_1,X_2,\ldots,X_n)$ denote a statistic depending on $X_1,X_2,\ldots,X_n$ which are random observations from the k populations.

For a given positive integer n, we need to draw n samples from the k populations. On each stage one sample is permitted to be drawn from any one of the k populations. When n samples are drawn, say, $x_1,x_2,\ldots,x_n$, a reward $W(k,n; x_1,x_2,\ldots,x_n)$ is obtained where the function W is defined before the experiment. The problem is how to design a good sampling scheme so that the expected reward, $E_\theta W(\cdot)$ is maximum when θ is the true configuration.

When $n \leq k$, it is reasonable to take one sample from each of any n populations since we have no information about its distributions at all. We therefore confine ourselves to the case

$n \geq k + 1$. We also restrict ourselves to the case where the mean of each population is finite. We confine ourselves to the case where $W(k,n; X_1,X_2,\ldots,X_n) = \sum_{i=1}^{n} X_i$.

Let R be a selection procedure for selecting the unique population associated with $\theta_{[k]}$, the largest parameter, when $\Omega = \Omega(d)$, $d > 0$. Then, we define a random variable $\pi(R;m)$, which takes value in $\{1,2,\ldots,k\}$, such that $\pi(R;m) = i$ means π_i is selected based on m observations of each population using rule R. Let $\gamma_i(m) = P\{\pi(R;m) = i\}$, $i = 1,2,\ldots,k$. Define $\gamma(m) = \inf_{\Omega(d)} P\{CS|R\}$. Without loss of generality, we assume $\theta_{[k]}$ to be θ_k.

3.2 A Maximin Strategy

By a test block U(m) we mean a sequence of random outcomes $\{X_{11}, X_{21},\ldots,X_{k1}, X_{12}, X_{22},\ldots,X_{k2}, X_{1m}, X_{2m},\ldots,X_{km}\}$ where X_{ij} is the jth independent observation of π_i, $i = 1,2,\ldots,k$. By a trial block V(i,m) we mean a sequence of m random outcomes of π_i, i.e. $\{X_{i1}, X_{i2},\ldots,X_{im}\}$. For a given integer m, $0 \leq m \leq [\frac{n}{k}]$, a UV(R;m) scheme is a strategy of sampling which follows a test block U(m) first and then follows a trial block $V(\pi(R;m),n-km)$. Let W(UV(R;m)) denote the reward function using the UV(R;m) scheme. Then, it is obvious that $W(UV(R;m)) = \sum_{i=1}^{k} S_{im}+S_{\pi(R;m),n-kn}$. Let $X = \{0, 1,2,\ldots,[\frac{n}{k}]\}$ and $Y = \Omega(d)$ for given $d > 0$. And, for given R, let $G(x,y) = E_y W(UV(R;m))$ for $x \in X$ and $y \in Y$. Now if we consider a zero-sum two person game such that a statistician (player I) plays a game against nature (player II) with reward G, then, following the UV scheme, we have a game (X, Y, G). In this game, player II tries to minimize the reward while player I wants to maximize the reward. Hence, a good strategy for player I is a maximin strategy, i.e., player I needs to choose $m^* = m(R,k,n) \in X$ so that

$$(3.2.1) \qquad G(m^*,\underline{\theta}) = \max_{0 \le m \le [\frac{n}{k}]} \inf_{\underline{\theta} \in \Omega(d)} E_{\underline{\theta}} W(UV(R;m)).$$

It is obvious that a maximin strategy always exists for our problem.

According to our notation we see that

$$(3.2.2) \qquad G(m,\underline{\theta}) = E_{\underline{\theta}} W(UV(R;m) = E_{\underline{\theta}}[\sum_{i=1}^{k} S_{im} + S_{\pi(R;m),n-km}]$$
$$= m \sum_{i=1}^{k} t(\theta_i) + (n-km) \sum_{i=1}^{k} \gamma_i(R;m) t(\theta_i),$$

where $E_{\theta_i} X_{i1} = t(\theta_i)$ for $i = 1,2,\ldots,k$. As a special case, when X_{i1} is unbiased for θ_i, we have

$$(3.2.3) \qquad G(m,\underline{\theta}) = m \sum_{i=1}^{k} \theta_i + (n-km) \sum_{i=1}^{k} \gamma_i(R;m)\theta_i.$$

Let

$$(3.2.4) \qquad L(R;m) = md + (n-km)\gamma(R;m)d \quad \text{and}$$

$$U(R;m) = md + (n-km)\gamma_k(R;m|\underline{\theta}=\underline{\theta}_0), \text{ where } \underline{\theta}_0 = (0,0,\ldots,0,d).$$

Then, if $\theta_i \ge 0$ and $t(\theta_i) = \theta_i$ for $i = 1,2,\ldots,k$, we have

$$(3.2.5) \quad \max_{m\varepsilon X} L(R;m) \le \max_{m\varepsilon X} \inf_{\underline{\theta}\varepsilon Y} G(m,\underline{\theta}) \le \max_{m\varepsilon X} U(m,\underline{\theta}_0).$$

Let X* denote the convex hull of X. Then, if $[\frac{n}{k}] = \ell$, X* is a closed convex subset of R^{ℓ}. Let $\gamma(R;m) = (\gamma_1(R;m), \gamma_2(R;m),\ldots,\gamma_k(R;m))$ and $\underline{t}(\underline{\theta}) = (t(\theta_1),t(\theta_2),\ldots,t(\theta_k))$. Then we have the following theorem:

Theorem 3.2.1 If $\Omega(d)$ is a closed bounded convex subset of R^k and $\gamma_i(m|\underline{\theta})$ is continuous in $\underline{\theta}$ for each m, t(z) is convex and continuous in z and $\underline{t}(\underline{\theta})\cdot\gamma(R;x)$ is convex in $\underline{\theta}$ for each x, then, (X*, $\Omega(d)$, G) has a value and player II has a good pure strategy and player I has a good strategy which is a mixture of at most $\min(k+1,\ell)$ pure strategies.

Proof: For $x \in X^*$ and $\underline{\theta} \in \Omega(d)$, there exists $\underline{\alpha} = (\alpha_0,\alpha_1,\ldots,\alpha_k)$ such that $\alpha_i \ge 0$, $\sum_{i=0}^{\ell} \alpha_i=1$ and $x = \sum_{i=1}^{k} i\alpha_i$. We note that $G(x,\underline{\theta}) = (\sum_{i=1}^{\ell} i\alpha_i)(\sum_{i=1}^{k} t(\theta_i))+\sum_{i=0}^{\ell}(n-ki\alpha_i)\sum_{j=1}^{k}\gamma_j(i\alpha_i|\underline{\theta})t(\theta_j)$. Hence, we see

that $G(x,\underline{\theta})$ is continuous in x and convex in $\underline{\theta}$. It

follows from our assumption and a result of Game Theory (see, for example, Blackwell and Girshick p. 53) that the theorem holds. Here we extend the discrete variable m of $\gamma_i(R;m)$ to a continuous variable by the polygonal interpolation. This completes the proof.

Definition 3.2.1 A selection procedure R is monotonic if $\gamma(R;m)$ is monotone increasing with respect to m such that $\lim_{m\to\infty} \gamma(R;m) = 1$.

Definition 3.2.2 (Hall [27]) A selection rule R is most economical if, for any rule R',

$$\gamma(R;m) \leq \gamma(R';n) \text{ implies } n \geq m.$$

Let C_1 denote a non-empty set of all monotonic rules for some selection problems and let C_2 denote the set of all most economical rules. Let $C = C_1 \cap C_2$. Then, it is clear that in the indifference zone formulation, the Bechhofer type procedure is in C_1. It has been shown in Hall [27] that R_{SH}, the Sobel-Huyett procedure [51], and R_B, the Bechhofer procedure, are both in C_2. Hence, R_{SH} and R_B are both in C.

Let $t(\theta)$ and $L(R;m)$ be defined by (3.2.2) and (3.2.4). A most economical rule has the following property:

Corollary 3.2.1 If $t(\theta) = \theta$ and $R \varepsilon C_2$, then, for any R',

$$(3.2.6) \qquad \max_{0\leq s\leq \ell} L(R';s) \leq \max_{0\leq s\leq \ell} L(R;s)+\min\{n\delta(R;\ell),k\gamma(R;\ell+1)\}d$$

where $\ell = [\frac{n}{k}]$ and $\delta(R;\ell) = \max_{0\leq m\leq \ell} \{\gamma(R;m+1)-\gamma(R;m)\}$.

Proof: Since $R \varepsilon C_2$, for any R', we have

$$\gamma(R';m) < \gamma(R;m+1) \quad \text{for} \quad m = 0,1,2,\ldots,\ell .$$

Hence,

(3.2.7) $\max_s L(R';s) \leq md + (n-km)\gamma(R;m+1)d$ for some m, $0 \leq m \leq \ell$.

On the other hand, we see that

$$\max_m L(R;m) \geq (m+1)d+[n-k(m+1)]\gamma(R;m+1)d \text{ for some}$$

$0 \leq m \leq \ell-1$. Or,

(3.2.8) $$\max_m L(R;m) \geq [md+(n-km)\gamma(R;m+1)d]+[d-k\gamma(R;m+1)d]$$

for some $0 \leq m \leq \ell-1$.

It follows from (3.2.7) and (3.2.8) that

(3.2.9) $\max_s L(R';s) \leq \max_s L(R;s)+ k\gamma(R;\ell+1)d$

since $R \in C_2$.

Furthermore, we note that

$$\max_s L(R';s) \leq md+(n-km)[\gamma(R;m)+\delta(R;\ell)]d \text{ for some}$$

m $(0 \leq m \leq \ell)$ where $\delta(R;n)$ is defined by (3.2.6).

Hence, we have

(3.2.10) $\max_s L(R';s) \leq \max_{m'} \{m'd+(n-km')d\gamma(R;m')\}+(n-km)\delta(R;\ell)d$

for some m .

It follows from (3.2.9) and (3.2.10) that

$$\max_s L(R';s) \leq \max_s L(R;s)+\min\{n\delta(R;\ell)\ k\gamma(R;\ell+1)\}d$$

this completes the proof.

Henceforth we confine ourselves to C whenever it is non-empty.

To find an infimum in $\Omega(d)$ there is no general rule to follow since it depends on $\underline{t}$ and $\underline{\gamma}$. However, we have the following sufficient conditions for the solution of the least favorable configuration when k = 2.

<u>Corollary 3.2.2</u> Suppose $t(x)$ is twice differentiable and $\gamma_1(R;m)$ is also twice partially differentiable with respect to θ_1 and θ_2.

Define

$$(3.2.11)\quad A(\theta_1,\theta_2) = mt'(\theta_1)+(n-2m)\{\frac{\partial\gamma_1}{\partial\theta_1}[t(\theta_1)-t(\theta_2)]+\gamma_1 t'(\theta_1)\}$$

$$B(\theta_1,\theta_2) = mt'(\theta_2)+(n-2m)\{\frac{\partial\gamma_1}{\partial\theta_2}[t(\theta_1)-t(\theta_2)]+(1-\gamma_1)t'(\theta_2)\}$$

$$C(\theta_1,\theta_2) = mt''(\theta_1)+(n-2m)\{\frac{\partial^2\gamma_1}{\partial\theta_1^2}[t(\theta_1)-t(\theta_2)]+2\frac{\partial\gamma_1}{\partial\theta_1}t'(\theta_1)+ \\ +\gamma_1 t''(\theta_1)\}$$

$$D(\theta_1,\theta_2) = (n-2m)\{\frac{\partial^2\gamma_1}{\partial\theta_1\partial\theta_2}[t(\theta_1)-t(\theta_2)]+\frac{\partial\gamma_1}{\partial\theta_1}[t'(\theta_1)-t'(\theta_2)]\}$$

$$E(\theta_1,\theta_2) = mt''(\theta_2)+(n-2m)\{\frac{\partial^2\gamma_1}{\partial\theta_2^2}[t(\theta_1)-t(\theta_2)]+(1-\gamma_1)t''(\theta_2)- \\ -2\frac{\partial\gamma_1}{\partial\theta_2}t'(\theta_2)\}\ .$$

If $\underline{\theta}^* = (\theta_1,\theta_2)$ satisfies $|\theta_1-\theta_2| \geq d$ such that

$$A(\theta_1,\theta_2)=0,\ B(\theta_1,\theta_2)+0 \text{ and } C(\theta_1,\theta_2)E(\theta_1,\theta_2) > D^2(\theta_1,\theta_2),$$

then, $\underline{\theta}^*$ is the least favorable configuration.

Proof: The proof is straightforward.

Corollary 3.2.3 If $k = 2$ and $F(x;\theta_i) = \Phi(x;\theta_i,1)$, the cdf of a normal distribution with mean $\theta_i \geq 0$ and variance 1, and $R = R_B$, then, $\theta^* = (0,d)$ is a least favorable configuration and m^* $(0 \leq m^* \leq [\frac{n}{k}])$ is a maximin strategy for player I if $m = m^*$ maximizes $(n-2m)d\Phi(\sqrt{\frac{m}{2}}\,d)+md$.

Proof: Let $\underline{\theta} = (\theta,\theta+d')$ with $\theta \geq 0$ and $d' \geq d$. Then, we have

$$\gamma_2(R_B;m) = \int_{-\infty}^{\infty} \Phi(x+\sqrt{m}\,d')d\Phi(x) = \Phi(\sqrt{\frac{m}{2}}\,d') \text{ and}$$

$$\begin{aligned} EW(UV(R_B;m)) &= (2\theta+d')m+(n-2m)[\gamma_1\theta+(1-\gamma_1)(\theta+d')] \\ &= n\theta+(n-2m)d'\gamma_2+md'. \end{aligned}$$

This is a monotone increasing function of θ and d'. Letting $\theta \downarrow 0$ and $d' \downarrow d$, we have the result. This completes the proof.

m^* of Corollary 3.2.3 is tabulated in table 1 at the end of this chapter.

For $k = 2(1)7$, the maximum value m^* for L defined by (3.2.4) for binomial populations is also tabulated in table 2.

It is natural to ask how good is the maximin strategy when n increases to infinity. We have the following result to see its asymptotic behavior.

Lemma 3.2.1 If $\{n_i;n=1,2,\ldots\}$ and $\{m_i;i=1,2,\ldots\}$ are monotone increasing sequences of positive integers such that $n_i \to \infty$, $m_i \to \infty$ and $m_i = o(n_i)$ as $i \to \infty$, then, $EW(UV(R;m_i))/n_i \to \max_{1\leq j\leq k} t(\theta_j)$ $\forall R \in C$ as $i \to \infty$.

Proof: We note that $EW(UV(R;m_i)) = m_i \sum_{j=1}^{k} t(\theta_j)+(n_i-km_i)\cdot \sum_{j=1}^{k} \gamma_j(m_i)t(\theta_j)$. Hence,

$$EW(UV(R;m_i))/n_i = r_i \sum_{j=1}^{k} t(\theta_j)+(1-kr_i)\sum_{j=1}^{k}\gamma_j(R;m_i)t(\theta_j)$$

where $r_i = m_i/n_i$. Since $r_i \to 0$ and $\gamma_k(R;m_i) \to 1$, $\gamma_j(R;m_i) \to 0$ as $i \to \infty$ we have thus

$$EW(UV(R;m_i))/n_i \to \max_{1\leq j\leq k} t(\theta_j) \quad \text{as } i \to \infty.$$

This completes the proof.

Corollary 3.2.4 Under the same assumptions of Lemma 3.2.1

$$W(UV(R;m_i))/n_i \to \max_{1\leq j\leq k} t(\theta_j) \quad \text{WP1} \quad \forall R \in C.$$

According to the UV scheme, we note that

$$W(UV(R;m_i))/n_i = (S_{1m_i}+S_{2m_i}+\ldots+S_{km_i})/n_i + $$

$$+ S_{\pi(R;m_i),n_i-km_i}/n_i$$

$$= \sum_{j=1}^{k} \frac{S_{jm_i}}{m_i}\cdot\frac{m_i}{n_i} + \frac{S_{\pi(R;m_i),n_i-km_i}}{n_i-km_i}\cdot\frac{n_i-km_i}{n_i} .$$

By the strong law of large numbers we see that

$\frac{S_{jm_i}}{m_i} \to t(\theta_j)$ a.s. as $i \to \infty$. By our assumption we have

$\frac{m_i}{n_i} \to 0$ and $\frac{n_i - km_i}{n_i} \to 1$ as $i \to \infty$. This follows that

$$W(UV(R;m_i))/n_i \to t(\theta_j) \text{ a.s. for some } j \text{ as } i \to \infty,$$

If $t(\theta_j) \neq t(\theta_k) = \max\limits_{1 \le j \le k} t(\theta_j)$, this comes to a contradiction with Lemma 2.3.1. The proof is thus complete.

Let $\{n_i;\ i = 1,2,\ldots\}$ be a strictly increasing sequence of positive integers. For a given $R \in C$, let $\{m_i^*;\ i = 1,2,\ldots\}$ and $\{\underline{\theta}_i^*;\ i = 1,2,\ldots\}$ be respectively the associated maximin strategies and least favorable configurations. Then, we have the following theorem.

<u>Theorem 3.2.2</u> If $\{\underline{\theta}_i^*;\ i = 1,2,\ldots\}$ is bounded in R^k and $t(\theta_{ij}^*)$ is bounded for each $j = 1,2,\ldots,k$ and $i = 1,2,\ldots$ where $\underline{\theta}_i^* = \{\theta_{i1}^*, \theta_{i2}^*, \ldots, \theta_{ik}^*\}$, then, there exists a subsequence $\{n_{i_j};j=1,2,\ldots\}$ of $\{n_i\}$ such that

$$G(m_{i_j}^*, \underline{\theta})/n_{i_j} \to \max_{1 \le r \le k} t(\theta_r) \text{ as } j \to \infty \quad \forall \underline{\theta} \in \Omega(d).$$

Proof: (i) Let $\ell_i = [\frac{n_i}{k}]$ for $i = 1,2,\ldots,$. We are going to show that $\{m_i^*;\ i = 1,2,\ldots\}$ is unbounded. Suppose there is some integer M such that $m_i^* < M$ for $i = 1,2,\ldots,$ then, we note that for

$m_i^* < M \leq \ell_i$ for sufficiently large i,

$$G(m_i^*,\underline{\theta}_i^*) < G(M,\underline{\theta}_i^*) \text{ if } n_i > \frac{k \sum_{j=1}^{k} [M\gamma_j(M)-m_i^*\gamma_j(m_i^*)]t(\theta_{ij}^*)}{\sum_{j=1}^{k} [\gamma_j(M)-\gamma_j(m_i^*)]t(\theta_{ij}^*)} .$$

Since $t(\theta_{ij}^*)$ is bounded for all i and j, it is easy to see that there exists some i_0 such that $G(m_{i_0}^*,\underline{\theta}_{i_0}^*) < G(M,\underline{\theta}_{i_0}^*)$. This contradicts the assumption that $m_{i_0}^*$ is a maximin strategy.

(ii) Suppose $m_i^*/n_i \to \alpha$ $(0 \leq \alpha \leq 1)$. If $\underline{\theta}_i^*$ does not tend to a limit, we can choose a subsequence $\{\underline{\theta}_{i_j}^*, j = 1,2,\ldots\}$ of $\{\underline{\theta}_i^*\}$ such that $\underline{\theta}_{i_j}^* \to \underline{\theta}^*$, say, since $\{\theta_i^*\}$ is bounded. Then, $t(\theta_j^*)$ is finite where $\underline{\theta}^* = (\theta_1^*,\theta_2^*,\ldots,\theta_k^*)$. Then, the associated subsequences $\{m_{i_j}^*\}$ and $\{n_{i_j}^*\}$ satisfy $m_{i_j}^* \to \infty$ and $m_{i_j}^*/n_{i_j} \to \alpha$ $(0 \leq \alpha \leq 1)$. We note that

$$G(m_{i_j}^*,\underline{\theta}^*)/n_{i_j} \to \alpha \sum_{j=1}^{k} t(\theta_j^*)+(1-k\alpha)t(\theta_k^*), \text{ assuming}$$

$t(\theta_k^*) \equiv \max_{1\leq j\leq k} t(\theta_j^*)$, as $j \to \infty$, since $m_{i_j}^* \to \infty$ and $R \in C$. Since $\alpha \sum_{j=1}^{k} t(\theta_j^*)+(1-k\alpha)t(\theta_k^*)$ is a decreasing function of α $(0 \leq \alpha \leq 1)$, hence, we can conclude that $\alpha = 0$.

(iii) If m_i^*/n_i does not tend to a limit, we can choose subsequences $\{m_{i_j}^{!*}; j = 1,2,\ldots\}$ and $\{n_{i_j}^!; j = 1,2,\ldots\}$ such that $m_{i_j}^*/n_{i_j} \to \beta$ $(0 \leq \beta \leq 1)$, since $0 \leq m_i/n_i \leq 1$ for each i. Again, we can choose sequences $\{m^*_{i_j}; j = 1,2,\ldots\}$ and $\{n_{i_j}; j = 1,2,\ldots\}$ which are subsequences of $\{m'^*_{i_j}\}$ and $\{n'_{i_j}\}$, respectively such that $m^*_{i_j} \to \infty$ and $m^*_{i_j}/n_{i_j} \to \beta$ and the associated $\{\underline{\theta}^*_{i_j}\}$ satisfies $\underline{\theta}^*_{i_j} \to \theta^*$, say. By the same argument of (ii), we can show that $\beta = 0$.

It follows from (i), (ii) and (iii) that there always subsequences $\{n_{i_j}\}$ and $\{m^*_{i_j}\}$ such that $m^*_{i_j} \to \infty$ and $m^*_{i_j}/n_{i_j} \to 0$. It follows from Lemma 3.2.1 that

$$G(m^*_{i_j},\underline{\theta})/n_{ij} \to \max_{1\le r\le k} t(\theta_r) \text{ as } j \to \infty \quad \forall\, \underline{\theta} \in \Omega(d).$$

The proof is complete.

<u>Corollary 3.2.5</u> Under the same assumptions of Theorem 3.2.2, we have

$$W(UV(R;m^*_i))/n_i \to \max_{1\le j\le k} t(\theta_j) \quad \text{WP1} \quad \text{if } R \in C \text{ for } \Omega = \Omega^+$$

where $\Omega^+ = \{\underline{\theta} = (\theta_1,\theta_2,\ldots,\theta_k) | \theta_{[k]} > \theta_{[k-1]}\}$

Proof: When $\underline{\theta} \in \Omega(d)$, the argument is the same as that of Corollary 3.2.4. Since our arguments in Theorem 3.2.2 do not depend on d, hence, the whole arguments hold as long as $\theta_{[k]}-\theta_{[k-1]} > 0$.

Let $\{n_i; i=1,2,\ldots\}$ and $\{m_i; i=1,2,\ldots\}$ be two sequences of positive integers such that $km_i \leq n_i$ for every i and as $i \to \infty$, $m_i \to \infty$ and $m_i/n_i \to 0$. Let R be a monotonic selection procedure. By a $U_iV_i(R;n_i,m_i)$ scheme we mean a procedure which follows a sequence of test blocks $U_i(R;m_i)$ and trial blocks $V_i(\pi(R;m_i), n_i-km_i)$ and they are ordered by $U_1,V_1,\ldots,U_n,V_n,\ldots$.

<u>Theorem 3.2.3</u> Following the U_iV_i scheme we have

$$E_{\underline{\theta}}W(U_i V_i)/(n_1+n_2+\ldots+n_i) \to t(\theta_k) \text{ as } i \to \infty \quad \underline{\theta} \in \Omega(d)$$

where $W(U_iV_i)$ denotes the total sum of observations up to V_i.

Proof: Let T_s denote the total sum of observations in the first 2s blocks. Then, we have

$$T_s = \sum_{i=1}^{k} \sum_{j=1}^{m_1+\ldots+m_s} X_{ij} + \sum_{i=1}^{s} S_{\pi(R;m_i)n_i-km_i}.$$ It follows that

$$t(\theta_k) \geq ET_s/(n_1+n_2+\ldots+n_s) \geq [(\sum_{i=1}^{k} t(\theta_i))(\sum_{i=1}^{s} m_i)/(n_1+n_2+\ldots+n_s)]+$$

$$+ t(\theta_k)\sum_{i=1}^{s} \gamma(m_i)(n_i-km_i)/(n_1+n_2+\ldots+n_s).$$

By our assumption that $m_i/n_i \to 0$, it is easy to see that

$(\sum_{i=1}^{k} t(\theta_i))(\sum_{i=1}^{s} m_i)/(n_1+n_2+\ldots+n_s) \to 0$ as $s \to \infty$ for any arbitrary fixed $\underline{\theta}$. Since $\gamma(m_i) \to 1$, for $0 < \varepsilon < \frac{1}{k}$, there exists an $i_0=i_0(\varepsilon)$ such that for $i \geq i_0$, $\gamma(m_i) \geq 1-\varepsilon$ and $m_i \leq \varepsilon\, n_i$.

Then we have,

$$\gamma(m_i)(n_i-km_i) \geq (1-\varepsilon)(n_i-k\varepsilon n_i) = n_i(1-\varepsilon)(1-k\varepsilon) .$$

Hence, for $s \geq i_0$, we have

$$\sum_{i=1}^{s} \gamma(m_i)(n_i-km_i)/(n_1+n_2+\ldots+n_s) = \Big(\sum_{i=1}^{i_0-1} + \sum_{i=i_0}^{s}\Big)\gamma(m_i)(n_i-km_i)/\Big(\sum_1^s n_i\Big)$$

$$\geq \Big[\sum_{i=1}^{i_0} \gamma(m_i)(n_i-km_i)/\Big(\sum_1^s n_i\Big)\Big]+(1-\varepsilon)(1-k\varepsilon).$$

Letting $s \to \infty$ and then $\varepsilon \to 0$, completes the proof.

Table 1

m*-value for Normal Populations When k = 2

d \ m*	0	1	2	3	4	5	6	7	8	9
0.01	1(2)	3(6)	9(6)	15(6)	21(6)	27(6)	33(6)	39(6)	45(6)	51(6)
0.05	-	-	-	-	-	-	-	-	-	-
0.10	-	-	-	-	-	27(7)	34(6)	40(6)	46(6)	52(6)
0.20	-	-	9(7)	16(6)	-	28(6)	37(7)	41(6)	47(6)	53(7)
0.30	-	-	-	-	22(7)	29(6)	35(7)	42(7)	49(7)	56(8)
0.50	-	3(7)	10(7)	17(7)	24(7)	31(8)	39(9)	48(9)	57(9)	66(11)
0.70	-	-	10(8)	18(8)	26(10)	36(10)	46(13)	59(13)	72(16)	88(18)
0.90	-	3(8)	11(8)	19(11)	30(13)	43(17)	60(20)	80(24)	105(33)	138(40)
1.00	-	-	11(10)	21(12)	33(16)	49(20)	70(29)	98(36)	135(49)	185(16)
2.00	-	3(17)	20(48)	69(132)						
3.00	-	3(65)	68(133)							
5.00	-	3(198)								
7.00	-	-								

d \ m*	10	11	12	13	14	15	16	17
0.01	57(6)	63(6)	69(6)	75(6)	81(6)	87(6)	93(6)	99(6)
0.05	-	-	-	-	-	-	93(7)	100(6)
0.10	58(6)	64(6)	70(6)	76(6)	82(7)	89(6)	95(6)	101(6)
0.20	60(7)	67(6)	73(7)	80(7)	87(7)	94(6)	100(7)	107(8)
0.30	64(7)	71(8)	79(7)	86(8)	94(8)	102(8)	110(9)	119(8)
0.50	77(11)	88(11)	99(12)	111(14)	125(14)	139(15)	154(16)	170(17)
0.70	106(20	127(22)	150(26)	177(23)				
0.90	178(23)							

Table 1 (cont'd)

d \ m*	18	19	20	21	22	23	24	25
0.01	105(6)	111(6)	117(6)	123(6)	129(6)	135(6)	141(6)	147(6)
0.05	106(6)	112(6)	118(6)	124(6)	130(6)	136(6)	142(6)	148(6)
0.10	106(7)	114(6)	120(6)	126(6)	132(7)	139(6)	145(6)	151(7)
0.20	115(7)	122(7)	129(7)	136(7)	143(8)	151(7)	157(8)	166(7)
0.30	127(9)	136(9)	145(9)	154(10)	164(9)	173(10)	183(10)	193(8)
0.50	187(14)							

d \ m*	26	27	28	29	30	31	32	33
0.01	153(6)	159(6)	165(6)	171(6)	177(6)	183(6)	189(6)	195(6)
0.05	154(6)	160(6)	166(6)	172(6)	178(6)	184(6)	190(6)	196(5)
0.10	158(6)	164(6)	170(7)	177(6)	183(7)	190(6)	196(5)	
0.20	173(8)	181(8)	189(7)	196(5)				

(i) A dot indicates the same value as above

(ii) For d = 0.05, m* = 20 the entry 118(6) in the table, shows that for n from 118 to 123(=118+6-1), the m*-value is 20.

Table 2

Maximum value of m for L of (3.2.4) for binomial populations using $R=R_{SH}$

For k = 2, d = 0.20, m* = 6 the entry 36-41 in the table, shows that for n from 36 to 41, the m*-value of L is 6.

K = 2

d \ m*	1	2	3	4	5	6	8	9	10
0.05	2-8	9-14	15-20	21-26	27-32	33-38	39-51	52-57	58-60
0.10	2-8	9-14	15-20	21-26	27-32	33-39	40-52	53-59	60
0.20	2-8	9-14	15-21	22-28	29-35	36-41	42-60		
0.30	2-8	9-15	16-22	23-31	32-40	41-47	48-60		
0.50	2-9	10-19	20-32	33-49	50-60				
0.80	2-15	16-54	55-60						

K = 3

d \ m*	1	2	3	4	5	6	7	8	9	10
0.05	3-8	9-17	18-29	30-38	39-47	48-55	56	57-74	75-83	84-90
0.10	3-8	9-19	20-29	30-38	39-46	47-55		56-75	76-84	85-90
0.20	3-9	10-20	21-29	30-39	40-48	49-57		58-81	82-90	
0.30	3-9	10-21	22-31	32-43	44-53	54-62		63-90		
0.50	3-10	11-24	25-40	41-61	62-89	90				
0.80	3-16	17-63	64-90							

Table 2 (cont'd.)

K = 4

d \ m*	1	2	3	4	5	6
0.05	4-11	12-19	20-31	32-49	50-61	62-72
0.10	4-11	12-19	20-37	38-49	50-60	61-71
0.20	4-11	12-20	21-37	38-49	50-61	62-71
0.30	4-11	12-23	24-38	39-52	53-66	67-76
0.50	4-12	13-29	30-48	49-72	73-103	104-120
0.80	4-18	19-70	71-120			

d \ m*	8	9	10	11
0.05	73-98	99-109	110-120	
0.10	72-96	97-108	109-119	120
0.20	72-101	102-114	115-120	
0.30	77-120			

K = 5

d \ m*	1	2	3	4	5	6
0.05	5-14	15-24	25-33	34-52	53-75	76-88
0.10	5-14	15-23	24-34	35-59	60-73	74-86
0.20	5-14	15-22	23-41	42-59	60-73	74-85
0.30	5-13	14-24	25-45	46-61	62-77	78-90
0.50	5-14	15-31	32-55	56-82	83-116	117-141
0.80	5-19	20-76	77-150			

d \ m*	8	9	10	11
0.05	89-120	121-134	135-148	149-150
0.10	87-117	118-131	132-145	146-150
0.20	86-120	121-136	137-150	
0.30	91-140	141-150		
0.50	142-150			

K = 6

d \ m*	1	2	3	4	5	6
0.05	6-17	18-29	30-40	41-51	52-84	85-103
0.10	6-17	18-28	29-38	39-57	58-86	87-100
0.20	6-16	17-27	28-41	42-68	69-85	86-98
0.30	6-16	17-26	27-51	52-70	71-89	90-103
0.50	6-16	17-33	34-62	63-92	93-129	130-156
0.80	6-21	22-82	83-180			

d \ m*	8	9	10	11
0.05	104-143	144-160	161-176	177-180
0.10	101-138	139-154	155-170	171-180
0.20	99-140	141-157	158-176	177-180
0.30	104-159	160-180		
0.50	157-180			

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] Anderson, T. W. (1966). An introduction to Multivariate Statistical Analysis. Wiley, New York.

[2] Barlow, R. E. and Gupta, S. S. (1969). Selection procedures for restricted families of probability distributions. Ann. Math. Statist. (40) 905-917.

[3] Barron, A. (1968). A class of sequential multiple decision procedures. Ph.D. Thesis, Department of Statistics, Purdue University, West Lafayette, Indiana.

[4] Barron, A. and Gupta, S. S. (1972). In Methods of Operations Research (R. Henn, H. P. Kurzi and H. Schubert, ed.) Proceedings of the Oberwolfach Conference on Operations Research. pp. 124-144, Verlag Anton Hain-Meisenheim.

[5] Bechhofer, R. E., Elmaghraby, S. and Morse, N. (1959). A single-sample multiple-decision procedure for selecting the multinomial event which has the highest probability. Ann. Math. Statist. (30) 102-119.

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

[7] Bradt, R. N., Johnson, S. M. and Karlin, S. (1955). On sequential designs for maximizing the sum of n observations. Ann. Math. Statist. (26) 1060-1074.

[8] Cacoullas, T. and Sobel, M. (1966). An inverse sampling procedure for selecting the most probable event in a multinomial distribution. Multivariate Analysis: Proceedings of an International Symposium. (P. R. Krishnaiah ed.) Academic Press, N. Y., pp. 423-455.

[9] Cox, D. R. (1952). Sequential tests for composite hypothesis. Proc. Cambridge Philos. Soc. (48) 290-299.

[10] Cramér, H. (1946). Mathematical Methods of Statistics. Princeton University Press, Princeton, New Jersey.

[11] Darling, D. A. and Robbins, H. (1969). Some non-parametric sequential tests with power one. Proc. Nat. Acad. Sci. (61) 804-809.

[12] Deely, J. J. (1965). Multiple decision procedures from an empirical Bayes approach. Ph.D. Thesis, Department of Statistics, Purdue University, West Lafayette, Indiana.

[13] Deely, J. J. and Gupta, S. S. (1968). On the properties of subset selection procedures. Sankhyā Ser. A. (30) 37-50.

[14] Dunnett, C. W. (1955). A multiple comparison procedure for comparing several treatments with a control. J. Amer. Statist. Assoc. (50) 1096-1121.

[15] Fabian, V. (1970). A note on the Paulson's sequential ranking procedure. Tech. Report RM-260, VF-9, August 1970, Department of Statistics and Probability, Michigan State University.

[16] Feller, W. (1946). The law of the iterated logarithm for identically distributed random variables. Ann. of Math. (47) 631-638.

[17] Gnanadesikan, M. (1966). Some selection and ranking procedures for multivariate normal populations. Ph.D. Thesis, Department of Statistics, Purdue University, West Lafayette, Indiana.

[18] Gnanadesikan, M. and Gupta, S. S. (1970). A selection procedure for multiple normal distributions in terms of the generalized variances. Technometrics (12) 103-117.

[19] Gnedenko, B. V. and Korolynk, V. S. (1961). On the maximum discrepancy between two empirical distribution functions. IMS and AMS. Selected Translations in Mathematical Statistics and Probability 1 (1961) 13-16.

[20] Gupta, S. S. (1956). On a decision rule for a problem in ranking means. Inst. Stat. Mimeo. Series No. 150, Inst. Stat., University of North Carolina, Chapel Hill, North Carolina.

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

[22] Gupta, S. S. and McDonald, G. C. (1970). On some classes of selection procedures based on ranks. Nonparametric Techniques in Statistical Inference. Ed. by M. L. Puri, C.U.P. 1970. pp. 491-514.

[23] Gupta, S. S. and Nagel, K. (1967). On selection and ranking procedures and order statistics from the multinomial distribution. Sankhyā Ser. B. (29) 1-34.

[24] Gupta, S. S. and Nagel, K. (1971). On some contributions to multiple decision theory. In Statistical Decision Theory and Related Topics (S. S. Gupta and J. Yackel ed.) Academic Press, New York.

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

[26] Hall, W. J. (1958). Most economical multiple-decision rules. Ann. Math. Statist. (29) 1079-1094.

[27] Hall, W. J. (1959). The most-economic character of some Bechhofer and Sobel decision rules. Ann. Math. Statist. (30) 964-968.

[28] Hoel, P. G. (1937). A significance test for component analysis. Ann. Math. Statist. (8) 149-158.

[29] Kambo, N. S. and Kotz, S. (1966). On exponential bounds for binomial probabilities. Ann. Inst. Statist. Math. Tokyo (18) 277-287.

[30] Kraft, O. (1968). A note on exponential bounds for binomial probabilities. Ann. Inst. Statist. Math. Tokyo (20) 219-220.

[31] Krishnaiah, P. R. (1967). Selection procedures based on covariance matrices of multivariate normal populations. In Blanch Anniversary Volume (B. Mond, ed.), pp. 147-160, Aerospace Research, U. S. Air Force.

[32] Krishnaiah, P. R. and Rizvi, M. H. (1966). Some procedures for selection of multivariate normal populations better than a control. Multivariate Analysis: Proceedings of an International Symposium. (R. Krishnaiah ed.) Academic Press, New York, pp. 477-490.

[33] Lehmann, E. L. (1959). Testing Statistical Hypotheses. John Wiley and Sons, New York.

[34] Lehmann, E. L. Notes of the Theory of Estimation. Chapter 4, p. 19.

[35] Lehmann, E. L. (1961). Some Model 1 problems of selection. Ann. Math. Statist. (32) pp. 990-1012.

[36] Lehmann, E. L. (1963). A class of selection procedures based on ranks. Math. Ann. (150) 268-275.

[37] Nagel, K. (1970). On subset selection rules with certain optimality properties. Mimeo Series No. 222, Department of Statistics, Purdue University.

[38] Nomachi, Y. (1969). A closed sequential procedure selecting the population with minimum variance from several normal populations. Kumamoto Journal of Science, Ser. A, (8) 129-134.

[39] Panchapakesan, S. (1971). On a subset selection procedure for the most probable event in a multinomial distribution. In Statistical Decision Theory and Related Topics (S. S. Gupta and J. Yackel ed.), Academic Press, New York.

[40] Paulson, E. (1952). On the comparison of several experimental categories with a control. Ann. Math. Statist. (23) 239-246.

[41] Paulson, E. (1964). A sequential procedure for selecting the population with the largest mean from k normal populations. Ann. Math. Statist. (35) 174-180.

[42] Paulson, E. (1964). Sequential estimation and closed sequential decision procedures. Ann. Math. Statist. (35) 1048-1058.

[43] Paulson, E. (1967). Sequential procedures for selection of the best one of several binomial populations. Ann. Math. Statist. (38) 117-123.

[44] Perng, S. K. (1969). A comparison of the asymptotic expected sample sizes of two sequential procedures for ranking problem. Ann. Math. Statist. (40) 2198-2202.

[45] Puri, P. S. and Puri, M. L. (1968). Selection procedures based on ranks: scale parameter case. Sankhyā Ser. A. 291-302.

[46] Rizvi, M. H. and Sobel, M. (1967). Nonparametric procedures for selecting a subset containing the population with the largest α-quantile. Ann. Math. Statist. (38) 1788-1803.

[47] Robbins, H. (1956). A sequential decision problem with finite memory. Proc. Nat. Acad. Sci. (42) 920-923.

[48] Robbins, H. (1955). An empirical Bayes approach to statistics. Proceedings of the Third Berkeley Symposium on Statistics and Probability. I, 157-164.

[49] Robbins, H., Sobel, M. and Starr, N. (1968). A sequential procedure for selecting the best of k populations. Ann. Math. Statist. (39) 88-92.

[50] Smith, C. V. and Pyke, R. (1965). The Robbins-Isbell two-armed-bandit problem with finite memory. Ann. Math. Statist. (36) 1375-1386.

[51] Sobel, M. and Huyett, M. (1957). Selecting the best one of several binomial populations. Bell System Tech. J. (36) 537-576.

[52] Sobel, M. and Weiss, G. (1970). Inverse sampling and other selection procedures for tournaments with two or three players. In Non-parametric Techniques in Statistical Inference (M. L. Puri, ed), pp. 515-543. Cambridge University Press.

[53] Somerville, P. N. (1957). Optimum sampling in binomial populations. J. Amer. Statist. Assoc. (52) 494-502.

[54] Tong, Y. L. (1969). On partitioning a set of normal populations by their locations with respect to a control. Ann. Math. Statist. (40) 1300-1324.

[55] U. S. Department of Commerce (1964). Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. National Bureau of Standards, Applied Mathematics Ser. 55.