# ON SLIPPAGE TESTS AND MULTIPLE DECISION (SELECTION AND RANKING) PROCEDURES*

by

Ashok K. Singh

Purdue University

Department of Statistics

Division of Mathematical Sciences

Mimeograph Series # 494

June 1977

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

# TABLE OF CONTENTS

# LIST OF TABLES

Table

# LIST OF FIGURES

# INTRODUCTION

In many practical situations the experimenter is faced with the following problem: given a set of k populations $\pi_1,\ldots,\pi_k$ with parameters $\theta_1,\ldots,\theta_k$, respectively, decide if (1) all $\pi_i$ are same, and if not (2) find the subset $\pi_{i_1},\ldots,\pi_{i_t}$ $(1 \leq t < k)$ which, in some sense, is better than the rest of the given popualtions. It is clear that the classical tests of homogeneity are not designed for this problem, since they fail to provide any information regarding the second question, which may be more meaningful than the first one. A partial answer to this question was provided by Mosteller (1948) who proposed a test of $H_0$: $\theta_1 = \ldots = \theta_k$ against the slippage alternatives $H_i$: $\theta_1 = \ldots = \theta_i - \Delta = \ldots = \theta_k$, $i = 1,\ldots,k$. Paulson (1949) proposed a multiple decision rule for testing slippage in normal means. Since then, many authors have contributed to the theory of slippage tests. Karlin and Truax (1960) have given a decision thoeretic formulation of the single slippage problem and have derived the class of symmetric Bayes proceedures. Hall and Kudo (1968), Hall, Kudo and Yeh (1968), and Kudo and Yeh (1970) have considered the slippage problem as a problem in testing of composite statistical hypotheses. Van Ryzin (1970) has discussed the slippage problem from the Empirical Bayes approach and has derived asymptotically optimal procedures. A discussion of some of the availabe results is given in Doornbos (1966).

The theory of slippage tests, however, cannot be applied to more general situations, and provides only partial answer. Bahadur (1950) is among the earliest authors to consider a different approach and investigate the so-called selection and ranking procedures. References could be made to Bechhofer (1954), Gupta (1956, 1965) Lehmann (1961).

Generally speaking, two formulations for selection and ranking problems have been considered. Suppose population $\pi_i$ has distribution function $F(x,\theta_i)$, $i = 1,\ldots,k$, and let $\theta_{[i]}$ denote the i-th ordered $\theta_i$; the correct pairing of $\theta_i$ and $\theta_{[j]}$ is unknown. To fix ideas, suppose $\pi_i$ is better than $\pi_j$ $(j \neq i)$ if $\theta_i \geq \theta_j$, and consider the problem of selection of the best population, i.e, the population associated with $\theta_{[k]}$.

The first formualtion is the indifference zone approach of Bechhofer (1954), in which a procedure is defined to select one population subject to the condition that the probability of selecting the best one is at least $P^*$ whenever $\theta_{[k]} - \theta_{[k-1]} \geq \delta^*$, where $P^*$ $(\frac{1}{k} < P^* < 1)$ and $\delta^* > 0$ are preassigned constants.

The second formulation is due to Gupta (1956) in which a random number of populations is chosen so that the probability that $\theta_{[k]}$ is included in the selected subset is at least $P^*$, where $P^* \in (\frac{1}{k},1)$ is a preassigned constant.

Selection and ranking problems have also been considered from a Bayes approach. The problem of selecting a subset containing the 'best' has been considered by Deely and Gupta (1968), where it is shown that under certain conditions the Bayes rules for a linear loss function selects exactly one population. Guttman and Tiao (1964) have studied some best population problems when the a priori distribution of parameters is known. Deely (1965) has investigated emperical Bayes procedures for the problem.

Interest has recently developed in Bayesian selection rules for nonlinear loss functions. Eaton (1967) has proved that, under certain conditions, the natural selection rule is (i) Bayes for any symmetric a priori distribution on $(\theta_1,\ldots,\theta_k)$ (ii) has uniformly minimum risk among symmetric rules and (iii) is minimax and admissible. Goel and Rubin (1975), Chernoff and Yahav (1977), Bickel and Yahav (1977) and Hsu (1977) have considered Bayes rules for selecting a subset containing the best population. Typically, these Bayes rules are difficult to compute. It has been observed by Chernoff and Yahav (1977) and Hsu (1977) that Gupta-type procedures are reasonable approximations of the Bayes procedures.

A slighly different situation arises when, in addition to the k populations $\pi_i$, a control population $\pi_0$ is given and the goal is to select a subset of populations which are 'better' than control. Gupta and Sobel (1958), Huang (1975) and others have considered this problem.

The present thesis consists of investigations of multiple decision problems mentioned earlier, and some related topics. In Chapter I a problem in multiple slippage testing is considered, and the class of symmetric invariant Bayes rules to test if all parameters $\theta_i$ are the same against the alternative hypotheses that some (unknown) subset of $\{\theta_1,\ldots,\theta_k\}$ of a given size t $(1 \leq t < k)$ has slipped to the right by the same amount $\Delta > 0$ have been derived for location and scale type densities admitting sufficient statistics. The t-slippage problem has also been considered in nonparametric situations, and locally best rules based on ranks have been derived.

The rest of the thesis pertains to Bayes and Gupta-type selection procedures. In Chapter II Bayes rules for the problem of selecting a subset containing the t best (t known, $1 \leq t \leq k$) have been investigated. For a loss function which is linear in $\theta_i$, it is shown that, under certain conditions, the Bayes rule selects exactly t population which are associated with the t largest observations. This generalizes a similar result of Deely and Gupta (1968) for the case of $t = 1$. When the loss functions are non-linear, Bayes rules are analytically and computationally intractable. Some special cases, namely, normal exponential, Poisson and gamma distributions have been considered and Bayes rules derived for natural conjugate a priori distributions of the parameters $\theta_i$.

In Chapter III some selection problems for normal populations have been considered and Gupta-type selection rules based on sample medians of odd number of observations have been investigated. Exact and asymptotic results are obtained. In the special case of equally spaced normal means, the proposed rule is numerically compared to the selection rule of Gupta (1965), which is based on sample means. As expected, the sample means procedure appears to be better than the procedure based on medians. It is shown that the asymptotic relative efficiency (ARE) of the rule based on sample median relative to the rule based on sample means, when the normal means are in a slippage configuration, is equal to $2/\pi$. However, in case the normal populations are contaminated, the rule based on sample medians shows a significant improvement over the rule based on means in terms of the ARE.

Chapter IV consists of investigations of Bayes and Gupta-type rules for selecting a subset which contains all populations 'close' to a given control. A slightly different problem, in which the t $(1 \leq t < k)$ populations 'closest' to control are to be selected, is also studied using Bayes and empirical Bayes approaches.

# CHAPTER I

# TESTING MULTIPLE SLIPPAGES

## 1.1 Introduction

The problem of multiple slippages can roughly be described as follows: We are given k populations $\pi_1,\ldots,\pi_k$. The goal is to decide, on the basis of a sample from each population, if all $\pi_i$ are same, and if not, which of the t $(1 \leq t \leq k-1)$ $\pi_i$ are different from the remaining k-t, where t is a known integer. A similar problem in multiple slippage testing involves an additional control population, and the goal then is to compare the k populations $\pi_1,\ldots,\pi_k$ to the control population and decide if all the populations have parameters equal to that of the control, and if not, which t of the k populations are different from control. The latter problem will be referred to as the controlled case. If $\pi_{i_1},\ldots,\pi_{i_t}$ are different from $\{\pi_j: j \neq i_\ell, \ell = 1,\ldots,t\}$ we say that the subset $\{\pi_{i_1},\ldots,\pi_{i_t}\}$ has 'slipped'. In this chapter we consider a special case, in which the amount of slippage is assumed to be the same for all the populations which have slipped.

Karlin and Truax (1960) have derived a class of symmetric Bayes procedures for the single slippage problem, which corresponds to $t = 1$. Hall and Kudo (1968), Hall, Kudo and Yeh (1968), and Kudo and Yeh (1970) have considered the problem of single slippage in the framework of testing composite statistical hypotheses.

In this chapter the results of Karlin and Truax (1960) have been extended to the case of multiple slippage (t-slippage) when t is a known integer between 1 and k-1. A discussion of some available rules for the t-slippage problem is given in Doornbos (1966).

In Section 1.2 definitions and notations used in this chapter are introduced, and a decision-theoretic formulation of the problem is given. In Section 1.3 the general form of the symmetric Bayes procedures for the t-slippage problem is derived. Some specific examples are considered in Section 1.4.

In Section 1.5 (1.6) the class of symmetric invariant Bayes procedures for the t-slippage problem has been derived for location (scale) parameter densities having a sufficient statistic.

Section 1.7 consists of investigation of several related problems in ranking and selection.

In Section 1.8 the problem of t-slippage, using the notions of slippage introduced by Karlin and Truax (1960), has been discussed in nonparametric situations. Some locally best tests based on ranks have been derived. In Section 1.9 selection of a rule from the class of procedures obtained in this chapter is discussed.

## 1.2 Decision-Theoretic Formulation of the Multiple Slippage Problem

Let $\underline{x} = (x_1,\ldots,x_k) \in \mathbb{R}^k$, the k-dimensional Euclidean space, be an observed value of a random vector $\underline{X} = (X_1,\ldots,X_k)$ which has probability density function (pdf) $f(x_1,\ldots,x_k;\ \theta_1,\ldots,\theta_k)$ with respect to a $\sigma$-finite measure $\mu$. The form of the function f is known and $\theta_i$'s are unknown, $\theta_i \in \Theta \subset \mathbb{R}$ for $i = 1,\ldots,k$ where the set $\Theta$ has at least two points.

Let $S_k$ be the symmetric group on $\{1,\ldots,k\}$ i.e., the group of all permutations $\psi$: $\{1,\ldots,k\} \to \{1,\ldots,k\}$. For $\underline{z} \in \mathbb{R}^k$ and $\psi \in S_k$ define $\underline{z}_\psi$ by $(\underline{z}_\psi)_i = z_{\psi i}$, where $(\underline{z}_\psi)_i$ represents the i-th coordinate of the vector $\underline{z}_\psi$.

The pdf $f(\underline{x};\theta)$ and the measure $\mu$ are assumed to be symmetric in the following sense:

$$f(\underline{x}_\psi;\ \underline{\theta}_\psi) = f(\underline{x};\underline{\theta}) \quad \forall\, \psi \in S_k,\ \underline{x} \in \mathbb{R}^k,\ \underline{\theta} \in \Theta^k \tag{1.2.1}$$

$$d\mu(\underline{x}_\psi) = d\mu(\underline{x}) \quad \forall\, \psi \in S_k. \tag{1.2.2}$$

We assume that $f(\underline{x};\underline{\theta})$ also has the property M of Eaton (1967).

<u>Definition 1.2.1</u>: A family of density functions $\{f_\alpha(\underline{x};\underline{\theta}): \alpha \in A\}$ is said to have property M if for each $\alpha \in A$, and each i,j $(i \neq j)$, $1 \leq i,j \leq 1$ the following holds:

$$x_i \geq x_j,\ \theta_i \geq \theta_j \Rightarrow f_\alpha(\underline{x};\underline{\theta}) \geq f_\alpha(\underline{x};\underline{\theta}_{\psi_{ij}}) \tag{1.2.3}$$

where $\psi_{ij} \in S_k$ is such that for $1 \leq \ell \leq k$,

$$\psi_{ij}(\ell) = \begin{cases} j & \text{if } \ell = i \\ i & \text{if } \ell = j \\ \ell & \text{otherwise.} \end{cases}$$

Let $\mathbb{S} = \{J \subset \{1,2,\ldots,k\} : |J| = t\}$

where $|J|$ denotes the size of the subset J. For any $J \in \mathbb{S}$, let

$$\Theta_J = \{\underline{\theta} = (\theta_1,\ldots,\theta_k) \in \Theta^k : \theta_j = \begin{matrix} \omega+\Delta & \text{if } j \in J \\ \omega & \text{if } j \notin J \end{matrix};\ j = 1,\ldots,k,\ \omega \in \mathbb{R}, \Delta > 0\}.$$

Also let $\Theta_0 = \{\underline{\theta} = (\theta_1,\ldots,\theta_k) \in \Theta^k : \theta_j = \omega;\ j = 1,\ldots,k,\ \omega \in \mathbb{R}\}$.

The problem is to test

$$H_0 : \underline{\theta} \in \Theta_0$$

against $\binom{k}{t}$ alternative hypotheses

$$H_J : \underline{\theta} \in \Theta_J, \quad J \in \mathbb{S}.$$

The action space $\mathcal{a} = \{a_0, a_J : J \in \mathbb{S}\}$ consists of $\binom{k}{t} + 1$ elements. The action $a_0$ corresponds to acceptance of $H_0$, and $a_J$ to the acceptance of $H_J$.

For $J \in \{0\} \cup \mathbb{S}$, let $L_J(\underline{\theta})$ denote the loss incurred in taking action $a_J$ when $\underline{\theta}$ is the true parameter value. Here $L: \Theta^k \times \mathcal{a} \to \mathbb{R}$ is a $\mu$-measurable function, and is assumed to satisfy the following:

(1) $L_J(\underline{\theta}) = L_{\psi^{-1}J}(\underline{\theta}_\psi)$ $\forall\ J \in \{0\} \cup \mathbb{S}$ and $\psi \in S_k$ where, for $J = \{j_1,\ldots,j_t\}$, $\psi^{-1}J = \{\psi^{-1}j_1,\ldots,\psi^{-1}j_t\}$

(2) $L_J(\underline{\theta}) < L_I(\underline{\theta})$ $\forall\ \underline{\theta} \in \Theta_J$, $I,J \in \{0\} \cup \mathbb{S}$, $I \neq J$.

(3) $L_I(\underline{\theta}) = L_J(\underline{\theta})$ $\forall \underline{\theta} \notin \Theta_L$, $I,J \neq L$. Otherwise $L_I(\underline{\theta}) = L_0(\underline{\theta})$ $\forall$ I.

For the problem of comparing the k populations $\pi_1,\ldots,\pi_k$ to a given control population $\pi_0$, the above needs some modifications. Here we are given an observation $(x_1,\ldots,x_k,y) \in \mathbb{R}^{k+1}$ from a population with density function $f(x_1,\ldots,x_k,y;\ \theta_1,\ldots,\theta_k,\theta_0) \equiv f(\underline{x},y;\ \underline{\theta},\theta_0)$ with respect to a σ-finite measure μ; $\theta_i \in \Theta \subset \mathbb{R}$, $i = 0,1,\ldots,k$.

The density function f is assumed to have the following symmetry:

$$f(\underline{x},y;\underline{\theta},\theta_0) = f(\underline{x}_\psi,y;\underline{\theta}_\psi,\theta_0)\ \ \forall \psi \in S_k, (\underline{x},y) \in \mathbb{R}^{k+1}, (\underline{\theta},\theta_0) \in \Theta^{k+1} \tag{1.2.4}$$

It is also assumed that the family $\{f_{(y,\theta_0)}(\underline{x};\underline{\theta}) \equiv f(\underline{x};y;\underline{\theta},\theta_0):$ $(y,\theta_0) \in \mathbb{R} \times \Theta\}$ has property M.

The problem is to test

$$H_0:\ (\underline{\theta},\theta_0) \in \Theta_0 = \{(\theta_1,\ldots,\theta_k,\theta_0) \in \Theta^{k+1}:\ \theta_i = \theta_0,\ i = 1,\ldots,k\}$$

against $\binom{k}{t}$ alternatives

$$H_J:\ (\underline{\theta},\theta_0) \in \Theta_J = \{(\theta_1,\ldots,\theta_k,\theta_0) \in \Theta^{k+1}:\ \theta_j = \begin{matrix}\theta_0+\Delta & \text{if } j \in J \\ \theta_0 & \text{if } j \notin J\end{matrix};\ j=1,\ldots,k,$$

$$\theta_0 \in \mathbb{R},\ \Delta>0\} \text{ where } J \in \mathcal{S}.$$

The action space $\mathcal{A} = \{a_0,a_J:\ J \in \mathcal{S}\}$ again has $\binom{k}{t}+1$ elements. The action $a_0$ decides in favor of $H_0$, and $a_J$ in favor of $H_J$.

The loss function L: $\Theta^{k+1} \times \mathcal{A} \to \mathbb{R}$ is a μ-measurable function, with $L_J(\underline{\theta},\theta_0)$ representing the loss of taking action $a_J$, when $(\underline{\theta},\theta_0)$ is the true value of the parameter. The assumptions on the loss functions in this case are

(1') $L_J(\underline{\theta},\theta_0) = L_{\psi^{-1}J}(\underline{\theta}_\psi,\theta_0)\ \forall\ J \in \mathcal{S},\ \psi \in S_k$ and $(\underline{\theta},\theta_0) \in \Theta^{k+1}$

(2') $L_J(\underline{\theta},\theta_0) < L_I(\underline{\theta},\theta_0)\ \forall\ (\underline{\theta},\theta_0) \in \Theta_J,\ I,J \in \{0\} \cup \mathcal{S},\ I \neq J.$

(3') $L_I(\underline{\theta},\theta_0) = L_J(\underline{\theta},\theta_0)\ \forall\ (\underline{\theta},\theta_0) \in \Theta_L,\ I,J \neq L.$ Otherwise

$L_I(\underline{\theta},\theta_0) = L_0(\underline{\theta},\theta_0)\ \forall\ I.$

## 1.3 Bayes Solutions

We first derive the Bayes procedures for the uncontrolled t-slippage problem. Since the problem is invariant with respect to the group $S_k$, we can restrict attention to symmetric procedures.

Definition 1.3.1: A symmetric decision function is a $\mu$-measurable vector function $\varphi = (\varphi_0,\ \varphi_I\colon I \in \mathcal{S})\colon \mathbb{R}^k \to [0,1]^m$ where $m = \binom{k}{t} + 1$, such that

(i) $\sum_{I\in\{0\}\cup\mathcal{S}} \varphi_I(\underline{x}) = 1 \qquad \forall\ \underline{x} \in \mathbb{R}^k$

(ii) $\varphi_I(\underline{x}) = \varphi_{\psi^{-1}I}(\underline{x}_\psi) \quad \forall\ \psi \in S_k,\ \underline{x} \in \mathbb{R}^k,\ I \in \{0\} \cup \mathcal{S}.$

In our problem, for $I \in \{0\} \cup \mathcal{S}$, $\varphi_I(\underline{x})$ represents the probability of accepting $H_I$, given $\underline{x}$ is observed. In the terminology of Ferguson (1967), $\varphi$ is a behavioral decision function.

Definition 1.3.2: If $\mathcal{D}$ denotes the set of all possible behavioral decision functions for the t-slippage problem, the risk function $\rho\colon \mathcal{D} \times \Theta^k \to \mathbb{R}$ is defined by

$$\rho(\varphi,\underline{\theta}) = \int_{\mathbb{R}^k} \sum_{I\in\{0\}\cup\mathcal{S}} L_I(\underline{\theta})\varphi_I(\underline{x})f(\underline{x},\underline{\theta})\ d\mu(\underline{x})$$

The following two results will be used in deriving the form of the Bayes procedures; these are slight modifications of Lemma 3.1 and Theorem 3.1 of Karlin and Truax (1960).

Lemma 1.3.1: The risk function of a symmetric decision function is a symmetric function of $\underline{\theta}$.

Theorem 1.3.1: Any symmetric Bayes procedure for the t-slippage problem is Bayes against a symmetric a priori distribution.

Let $\Omega = \Theta_0 \cup (\bigcup_{I \varepsilon \mathcal{S}} \Theta_I)$.

The set $\Omega$ can be represented as follows:

$$\Omega = \{(I,\omega,\Delta)\colon I \varepsilon \{0\} \cup \mathcal{S},\ \omega \varepsilon \mathrm{IR},\ \Delta > 0\}$$

where $I = 0$, $\Delta = 0$ iff$(I,\omega,\Delta) \in \Theta_0$, and $I \varepsilon \mathcal{S}$, $\Delta > 0$ iff$(I,\omega,\Delta) \in \Theta_I$. Let G be an a priori distribution on $\Omega$, and

$$\xi_0 = \int_{\Theta_0} dG$$

$$\xi_I = \int_{\Theta_I} dG.$$

Also let $G_0(\omega)$ denote the conditional distribution function of $\omega$ given $\Delta = 0$ and $G_I(\omega,\Delta)$ denote the joint conditional distribution given the set I has slipped.

It is easily seen that G is symmetric, i.e., $G(\underline{\theta}) = G(\underline{\theta}_\pi)$ $\forall\ \underline{\theta}\ \varepsilon\ \Omega$, $\pi\ \varepsilon\ S_k$, iff all $\xi_I$, $I\ \varepsilon\ \mathcal{S}$ are equal and all $G_I(\omega,\Delta)$, $I\ \varepsilon\ \mathcal{S}$ are identical. Then, since we are interested in symmetric procedures, using Theorem 1.3.1 we can take

$$\xi_I = \xi \quad \forall I \in \mathcal{S}, \text{ where } \xi \in [0,1],$$

and

$$G_I(\omega,\Delta) = \bar{G}(\omega,\Delta) \quad \forall I \in \mathcal{S},\ \omega \in \mathbb{R},\ \Delta > 0.$$

Let $f_0(\underline{x};\omega)$ and $f_J(\underline{x};\omega,\Delta)$ denote the density under $H_0$ and $H_J$, respectively. Then comparing Bayes posterior risks we obtain $\varphi_0(\underline{x}) = 1$ if $\underline{x} \in R_0$, where $R_0$, a symmetric set in $\mathbb{R}^k$, is given by

$$\begin{aligned} R_0 = \{\underline{x} \in \mathbb{R}^k\colon\ & \xi_0 \int [L_0(\omega) - L_J(\omega)] f_0(\underline{x};\omega)\, dG_0(\omega) \\ & + \xi \int [L_0(J,\omega,\Delta) - L_J(J,\omega,\Delta)] f_J(\underline{x};\omega,\Delta)\, d\bar{G}(\omega,\Delta) \\ & < 0 \quad \forall J \in \mathcal{S}\} \end{aligned} \tag{1.3.1}$$

and for $I \in \mathcal{S}$

$$\begin{aligned} \varphi_I(\underline{x}) = 1 \text{ if } & \xi_0 \int [L_I(\omega) - L_J(\omega)] f_0(\underline{x};\omega)\, dG_0(\omega) \\ & + \xi \int [L_I(J,\omega,\Delta) - L_J(J,\omega,\Delta)] f_J(\underline{x};\omega,\Delta)\, d\bar{G}(\omega,\Delta) \\ & + \xi \int [L_I(I,\omega,\Delta) - L_J(I,\omega,\Delta)] f_I(\underline{x};\omega,\Delta)\, d\bar{G}(\omega,\Delta) \\ & < 0 \quad \forall J \in \{0\} \cup \mathcal{S},\ J \neq I. \end{aligned} \tag{1.3.2}$$

For $J = 0$, the second term in the expression (1.3.2) vanishes, and (1.3.1) implies that $\underline{x} \notin R_0$. Using assumption (1) on $L_J(\underline{\theta})$, it can be seen that

$$L_I(I,\omega,\Delta) - L_I(J,\omega,\Delta) = -[L_I(J,\omega,\Delta) - L_J(J,\omega,\Delta)]$$

and hence taking $J \in \mathcal{S}$ in (1.3.2) gives

$$\varphi_I(\underset{\sim}{x}) = 1 \text{ if } \int [L_I(I,\omega,\Delta)-L_J(I,\omega,\Delta)][f_I(\underset{\sim}{x};\omega,\Delta)-f_J(\underset{\sim}{x};\omega,\Delta)]d\bar{G}(\omega,\Delta) < 0 \quad \forall J \neq I. \quad (1.3.3)$$

It follows from assumption (2) on the loss functions and the property M of the density f that the inequality in (1.3.3) holds iff the set I is such that

$$\{x_i: i \in I\} = \{x_{[k-t+1]},\ldots,x_{[k]}\} \quad (1.3.4)$$

where $x_{[1]} \leq \ldots \leq x_{[k]}$ are the ordered observations $x_i$.
Hence we have, for $I \in \mathcal{S}$

$\varphi_I(\underset{\sim}{x}) = 1$ if $\underset{\sim}{x} \notin R_0$ and the set I is as defined by (1.3.4).
Thus we have the following theorem:

Theorem 1.3.2: If $f(\underset{\sim}{x},\underset{\sim}{\theta})$ and $\mu$ satisfy the symmetry conditions (1.2.1) and (1.2.2), respectively, f has property M, and loss functions satisfy assumptions (1), (2), (3) of Section 1.2, then any symmetric Bayes procedure $\varphi$ is of the following form:

$\varphi_0(\underset{\sim}{x}) = 1$ if $\underset{\sim}{x} \in R_0$, a symmetric set in $\mathbb{R}^k$

$\varphi_I(\underset{\sim}{x}) = 1$ if $\underset{\sim}{x} \in R_0^c \cap \{\underset{\sim}{x} \in \mathbb{R}^k: \{x_i: i \in I\} = \{x_{[k-t+1]},\ldots,x_{[k]}\}\}$

Remarks:

1. The form of the set $R_0$ depends heavily on the density function f, as illustrated by the examples discussed in the next section.
2. For t = 1, Karlin and Truax (1960) assume that

$$L_j(\underset{\sim}{\theta}) = L_{\pi j}(\underset{\sim}{\theta}_\pi). \quad (1.3.5)$$

It follows from Eaton (1967) that assumption (1.3.5) of Karlin and Truax (1960) is stronger than our assumption (1) of Section 1.2. Moreover, it may not even be reasonable to assume (1.3.5). For example, suppose $k = 3$, $\underline{\theta} = (\omega+\Delta, \omega, \omega)$ and take $\pi$ as $\binom{1\ 2\ 3}{2\ 3\ 1}$. Then assumption (1.3.5) gives $L_1(\omega+\Delta, \omega, \omega) = L_2(\omega, \omega, \omega+\Delta)$.

A result similar to Theorem 1.3.2 holds for the controlled t-sliipage problem. The analysis involves some minor modifications, and hence is not included.

## 1.4 Examples.

In this section the set $R_0$ is explicitly obtained for normal, gamma and multinomial populations. Unless mentioned otherwise, it is assumed throughout the rest of this chapter that the loss functions satisfy assumptions (1), (2), (3) [(1'), (2'), (3')] of Section 1.2 for the uncontrolled [controlled] problem. We will discuss the two cases separately.

1. (a) The normal distribution with known variance. (Controlled case) We are given independent observations $x_1,\ldots,x_k$, $y$ from $k+1$ normal populations with means $\theta_1,\ldots,\theta_k,\theta_0$, respectively and variance 1. We make an additional assumption on the loss functions:

$$L_I(\theta_1+c,\ldots,\theta_k+c;\theta_0+c) = L_I(\theta_1,\ldots,\theta_k;\theta_0),\ \forall I \in \{0\} \cup \mathcal{S},\ c \in \mathbb{R},\ (\underline{\theta},\theta_0) \in \Omega.$$

Using invariance with respect to the group of translations attention can be restricted to procedures which depend only on the observed values of the random vector $\underset{\sim}{U} = (U_1,\ldots,U_k)$ where $U_1 = X_1-Y,\ldots,U_k = X_k-Y$. The joint distribution of $U_1,\ldots,U_k$, as given in Karlin and Truax (1960), is

$$f(u_1,\ldots,u_k;\omega_1,\ldots,\omega_k) = C.\exp[-\frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{k}\lambda^{ij}(u_i-\omega_i)(u_j-\omega_j)] \quad (1.4.1)$$

where C is the normalizing constant, independent of $\omega_i = \theta_i-\theta_0$, and

$$\lambda^{ij} = \begin{cases} \frac{k}{k+1} & \text{if} \quad i = j \\ -\frac{1}{k+1} & \text{if} \quad i \neq j \end{cases} .$$

It is easily seen from (1.4.1) that $f(\underset{\sim}{u};\underset{\sim}{\omega})$ satisfies the assumptions of Theorem 1.3.2 which therefore gives the class of symmetric invariant Bayes procedures; only the set $R_0$ remains to be found. Now, from (1.3.1), we have

$$R_0 = \{\underset{\sim}{u}\colon \xi_0(L_0-L_J)f_0(\underset{\sim}{u}) + \xi\int_0^\infty[L_0(J,\Delta)-L_J(J,\Delta)]f_J(\underset{\sim}{u};\Delta)d\bar{G}(\Delta) < 0 \ \forall\, J \in \mathcal{S}\} \quad (1.4.2)$$

where $f_J(\underset{\sim}{u};\Delta)$ is the joint density of $U_1,\ldots,U_k$ given that the subset $J \in \mathcal{S}$ has slipped to the right, and $f_0(\underset{\sim}{u})$ is the density when there is no slippage. Substituting $\omega_j = \Delta, \forall\, j \in J$ and $\omega_j = 0, \forall\, j \notin J$ in (1.4.1) we obtain

$$f_J(\underset{\sim}{u};\Delta) = f_0(\underset{\sim}{u}).\exp[-\frac{\Delta^2}{2}\sum_{i\in J}\sum_{j\in J}\lambda^{ij} + t\Delta\, h_J(\underset{\sim}{u})] \quad (1.4.3)$$

where $h_J(\underset{\sim}{u}) = \bar{u}_J - \frac{k}{k+1}\bar{u}$, $\bar{u}_J = \frac{\sum_{j\in J} u_j}{t}$, $\bar{u} = \frac{\sum_{i=1}^{k} u_i}{k}$.

It is clear from (1.4.3) that $f_J(\underline{u};\Delta)$ is an increasing function of $h_J(\underline{u})$ and therefore, from (1.4.2) we have

$$R_0 = \{\underline{u}: h_J(\underline{u}) < c \ , \forall J \in \mathcal{S}\}$$

$$= \{(\underline{x},y): \max_{j\varepsilon J} (\bar{x}_J - \frac{k\bar{x}+y}{k+1}) < c\} \tag{1.4.4}$$

where $\bar{x}_J = \dfrac{\sum_{j\varepsilon J} x_j}{t}$ .

Hence every symmetric invariant Bayes procedure has the form

$$\varphi_0(\underline{x},y) = 1 \text{ if } (\underline{x},y) \in R_0$$

$$\varphi_I(\underline{x},y) = 1 \text{ if } (\underline{x},y) \notin R_0 \text{ and } \{x_i: i \in I\} = \{x_{[k-t+1]},\ldots,x_{[k]}\}$$

where the set $R_0$ is given by (1.4.4).

(b) The normal distribution with known variance (uncontrolled case)

The additional assumption on the loss function is

$$L_I(\theta_1+c,\ldots,\theta_k+c) = L_I(\theta_1,\ldots,\theta_k),\ \forall\ c \in \mathbb{R},\ I \in \{0\} \cup \mathcal{S} \text{ and } (\theta_1,\ldots,\theta_k) \in \Theta^k.$$

The problem is again invariant with respect to the group of translations and a maximal invariant is $U_1 = X_1-X_k,\ldots,U_{k-1} = X_{k-1}-X_k$. The joint pdf of $U_1,\ldots,U_{k-1}$ is

$$f(u_1,\ldots,u_{k-1};\ \omega_1,\ldots,\omega_{k-1}) = C\exp[-\tfrac{1}{2}Q(\underline{u};\underline{\omega})]$$

where

$$Q(\underline{u};\underline{\omega}) = \sum_{i=1}^{k-1}\sum_{j=1}^{k-1} \lambda^{ij}(u_i-\omega_i)(u_j-\omega_j) \tag{1.4.5}$$

$$\omega_i = \theta_i - \theta_k, \ \lambda^{ij} = \begin{cases} \frac{k-1}{k} & \text{if } i = j \\ -\frac{1}{k} & \text{if } i \neq j \end{cases}.$$

It is easily verified that $f(\underline{u};\omega)$ has property M. We now calculate $f_J(\underline{u};\Delta)$, the joint pdf of $U_1,\ldots,U_{k-1}$ for $\underline{\theta} \in \Theta_J$, $J \varepsilon \mathcal{S}$. There are two possibilities:

(i) $k \in J$

Here

$$\omega_j = \begin{cases} 0 & \text{if } j \in J,\ j \neq k \\ -\Delta & \text{if } j \notin J \end{cases}$$

and thus

$$Q_J(\underline{u};\Delta) = Q_0(\underline{u}) + \frac{\Delta^2(k-t)t}{2k} - t\Delta[\bar{u}_J - \frac{k-1}{k}\bar{u}] \tag{1.4.6}$$

where $\bar{u}_J = \frac{\sum_{j\in J} u_j}{t}$, $u_k \equiv 0$, $\bar{u} = \frac{\sum_{i=1}^{k-1} u_i}{k-1}$

(ii) $k \notin J$

In this case $\omega_j = \begin{cases} \Delta \text{ if } j \in J \\ 0 \text{ if } j \notin J \end{cases}$.

Substituting the values of $\omega_j$ in the expression (1.4.5) we find that the form of $Q_J(\underline{u};\Delta)$ remains the same as (1.4.6).

Then, as in case (a), the set $R_0$ is given by

$$R_0 = \{\underline{u}\colon \bar{u}_J - \frac{k-1}{k}\bar{u} < c,\ \forall J \varepsilon \mathcal{S}\}$$

$$= \{\underline{x}\colon \bar{x}_J - \bar{x} < c,\ \forall J \varepsilon \mathcal{S}\}. \tag{1.4.7}$$

Hence the symmetric Bayes rule in terms of the observations $x_1,\ldots,x_k$ are given by

$$\varphi_0(\underline{x}) = 1 \quad \text{if} \quad \underline{x} \in R_0$$

$$\varphi_I(\underline{x}) = 1 \quad \text{if} \quad \underline{x} \notin R_0 \text{ and } \{x_i: i \in I\} = \{x_{[k-t+1]},\ldots,x_{[k]}\}$$

where $R_0$ is given by (1.4.7).

2. (a) The normal distribution with unknown variance (controlled case)

$y_j$ and $x_{ij}$ $(i = 1,\ldots,k;\ j = 1,\ldots,n)$ are n independent observations from k+1 independent normal populations with means $\theta_0$ and $\theta_i$ $(i = 1,\ldots,k)$ and a common unknown variance $\sigma^2$. The loss function is assumed to satisfy

$$L_I\left(\frac{\theta_1+\alpha}{\beta},\ldots,\frac{\theta_k+\alpha}{\beta},\frac{\theta_0+\alpha}{\beta},\frac{\sigma}{\beta}\right) = L_I(\theta_1,\ldots,\theta_k,\theta_0,\sigma)$$

$\forall\ \alpha \in \mathbb{R},\ \beta > 0$ and $I \in \{0\} \cup \{\mathbb{S}\}$.

The problem is invariant under groups of location and scale transformations, and a maximal invariant for both groups is

$$U_1 = \frac{\bar{X}_1-\bar{Y}}{S},\ldots,\ U_k = \frac{\bar{X}_k-\bar{Y}}{S}, \quad \text{where}$$

$$\bar{X}_i = \frac{\sum_{j=1}^{n} X_{ij}}{n}, \quad \bar{Y} = \frac{\sum_{j=1}^{n} Y_j}{n} \quad \text{and}$$

$$S^2 = \sum_{i=1}^{k}\sum_{j=1}^{n}(X_{ij}-\bar{X}_i)^2 + \sum_{j=1}^{n}(Y_j-\bar{Y})^2.$$

The joint pdf of $U_1,\ldots,U_k$, as given in Karlin and Truax (1960), is

$$f(\underset{\sim}{u};\underset{\sim}{\eta}) = c\int_0^\infty \exp\left[-\frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{k}\lambda^{ij}(u_i s-\eta_i)(u_j s-\eta_j)-\frac{s^2}{2}\right]s^{\frac{k+1}{n-1}}\,ds$$

where $$\lambda^{ij} = \begin{cases} \frac{kn}{k+1} & \text{if } i = j \\ -\frac{n}{k+1} & \text{if } i \neq j \end{cases}.$$

It is easily seen that the joint pdf of $U_1,\ldots,U_k$, given the set $J \in \mathcal{S}$ has slipped, is given by

$$f_J(\underset{\sim}{u};\delta) = C\,A(\delta)g(\underset{\sim}{u})\int_0^\infty \exp\left[t\delta h_J(\underset{\sim}{u})-\frac{s^2}{2}\right]s^{\frac{k+1}{n-s}}\,ds$$

where $\delta = \frac{\Delta}{\sigma}$, $A(\delta) = \exp\left[-\frac{n(k-t+1)}{2(k+1)}\delta^2\right]$

$$g(\underset{\sim}{u}) = \frac{1}{\left(\sum_{i=1}^{k}\sum_{j=1}^{k}\lambda^{ij}u_i u_j+1\right)^{\frac{n+k}{2(n-1)}}}$$

and $$h_J(\underset{\sim}{u}) = \frac{tn\left(\bar{u}_J-\frac{k}{k+1}\bar{u}\right)}{\left(\sum_{i=1}^{k}\sum_{j=1}^{k}\lambda^{ij}u_i u_j+1\right)^{\frac{1}{2}}}.$$

We now verify that the pdf $f(\underset{\sim}{u};\underset{\sim}{\eta})$ has property M.

For each fixed $s > 0$, the exponent of the integrand in the expression for $f(\underset{\sim}{u};\underset{\sim}{\eta})$ can be written as

$$Q_\Lambda(\underset{\sim}{u};\underset{\sim}{\eta}) = -\frac{1}{2}\left[(s\underset{\sim}{u}-\underset{\sim}{\eta})'\Lambda(s\underset{\sim}{u}-\underset{\sim}{\eta})\right]-\frac{s^2}{2}$$

where $$\Lambda = (\lambda^{ij}) = n\,I_k - \frac{n}{k+1}e'e,$$

$$e_{1\mathrm{x}k} = (1,\ldots,1) \in \mathbb{R}^k.$$

It follows from proposition (2.2) of Eaton (1967) that $f(\underline{u};\underline{\eta})$ has property M. Hence all the assumptions are met, and Theorem 1.3.2 gives the form of the symmetric invariant Bayes rules, except for the set $R_0$. The set $R_0$ is given by the inequalities

$$\xi_0(L_0-L_J)f_0(\underline{u}) + \xi\int_0^\infty [L_0(J,\delta)-L_J(J,\Delta)]f_J(\underline{u};\delta)d\bar{G}(\delta) < 0 \;, \forall J \in \mathcal{S}.$$

Since $f_J(\underline{u};\delta)$ is monotonically increasing in $h_J(\underline{u})$ we have

$$R_0 = \{\underline{u}: \frac{tn(\bar{u}_J - \frac{k}{k+1}\bar{u})}{\{1+n[\sum_{i=1}^k u_i^2 - \frac{(\sum_1^k u_i)^2}{k+1}]\}^{\frac{1}{2}}} < c \;, \forall J \in \mathcal{S}\}.$$

Changing to original observations $x_{ij}, y$, we have

$$\varphi_0(x_{ij},y) = 1 \text{ if } (x_{ij},y) \in R_0 = \{(x_{ij},y): \frac{\bar{x}_J - \frac{\bar{x}_. + \bar{y}}{k+1}}{[s^2 + n\sum_{i=1}^k (\bar{x}_i - \frac{\bar{x}_. + y}{k+1})^2 + n(\bar{y} - \frac{\bar{x}_. + y}{k+1})^2]} < c \;, \forall J \in \mathcal{S}\}$$

$$\varphi_I(x_{ij},\underline{y}) = 1 \text{ if } (x_{ij},y) \notin R_0 \text{ and}$$

$$\{x_i: i \in I\} = \{\bar{x}_{[k-t+1]},\dots,\bar{x}_{[k]}\}$$

where $\bar{x}_J = \frac{\sum_{j\in J} \bar{x}_j}{t}$, $\bar{x}_. = \sum_{j=1}^k \bar{x}_j$.

(b) The normal distribution with unknown variance (uncontrolled case)

Here $x_{ij}$ $(i = 1,\dots,k;\ j = 1,\dots,n)$ are n independent observations from k normal populations with means $\theta_i$ and a common unknown variance $\sigma^2$. Additional assumption on the loss function is

$$L_I(\frac{\theta_1+\alpha}{\beta},\ldots,\frac{\theta_k+\alpha}{\beta},\frac{\sigma}{\beta}) = L_I(\theta_1,\ldots,\theta_k,\sigma)$$

$\forall\ \alpha \in \mathbb{R},\ \beta > 0,\ I \in \{0\} \cup \mathcal{S}$.

Then, as in case (a), we have

$$R_0 = \{x_{ij}: \max_{J\varepsilon\mathcal{S}} [\frac{\bar{x}_J - \bar{x}}{\{\sum_{i=1}^{k}\sum_{j=1}^{n}(x_{ij}-\bar{x})^2\}^{\frac{1}{2}}}] < c\}$$

where $\bar{x} = \frac{\sum_{i=1}^{k}\bar{x}_i}{k}$, $\bar{x}_J = \frac{\sum_{i\in J}\bar{x}_i}{t}$

and the Bayes rules have the form

$$\varphi_0(x_{ij}) = 1 \quad \text{if} \quad x_{ij} \in R_0$$

$$\varphi_I(x_{ij}) = 1 \quad \text{if} \quad x_{ij} \notin R_0 \text{ and } \{x_i: i \in I\} = \{x_{[k-t+1]},\ldots,x_{[k]}\}.$$

3. (a) The gamma distribution (controlled case)

Here $x_1,\ldots,x_k,y$ are observations from k+1 independent gamma populations with unknown scale parameters $\theta_1,\ldots,\theta_k,\theta_0$, respectively and a common known shape parameter $p > 0$.

Additional assumption on the loss function is

$$L_I(\frac{\theta_1}{c},\ldots,\frac{\theta_k}{c},\frac{\theta_0}{c}) = L_I(\theta_1,\ldots,\theta_k,\theta_0)\ \forall\ c > 0,\ I \in \{0\} \cup \mathcal{S}.$$

The joint pdf of maximal invariants $U_1 = \frac{X_1}{Y},\ldots,U_k = \frac{X_k}{Y}$ is

$$f(u_1,\ldots,u_k;\ \omega_1,\ldots,\omega_k) = \frac{c \prod_{i=1}^{k} \omega_i^{-p}\, u_i^{p-1}}{\left(1 + \sum_{i=1}^{k} \frac{u_i}{\omega_i}\right)^{(k+1)p}},$$

where c is a constant, and $\omega_i = \frac{\theta_i}{\theta_0}$ .

It is easily verified that $f(\underline{u};\underline{\omega})$ satisfies the condition (1.2.3) and hence f has property M.

Then the set $R_0$ is given by the inequalities

$$\xi_0(L_0-L_J)f_0(\underline{u}) + \int_0^\infty [L_0(J,\Delta) - L_J(J,\Delta)]f_J(\underline{u};\delta)d\bar{G}(\delta) < 0,\ \forall\ J \in \mathcal{S}$$

where $\delta = \frac{\Delta}{\theta_0}$, and

$$f_J(\underline{u};\delta) = \frac{f_0(\underline{u})}{(1+\delta)^{pt}} \left[\frac{1}{\left(1 - \frac{\delta}{1+\delta}\frac{t\bar{u}_J}{1+k\bar{u}}\right)^{(k+1)p}}\right].$$

Since $f_J(\underline{u};\delta)$ is an increasing function of $\frac{t\bar{u}_J}{1+k\bar{u}}$, we have

$$R_0 = \{\underline{u}:\ \max_{J\in\mathcal{S}} \frac{t\bar{u}_J}{1+k\bar{u}} < c\}$$

$$= \{(\underline{x},y):\ \max_{J\in\mathcal{S}} \frac{\bar{x}_J}{y+k\bar{x}} < c\}$$

where $\bar{x}_J = \frac{\sum_{j\in J} x_j}{t}$, $\bar{x} = \frac{\sum_{j=1}^{k} x_j}{k}$

(b) The gamma distribution (uncontrolled case)

Similar analysis shows that the symmetric Bayes procedures for the uncontrolled case are given by

$$\varphi_0(\underset{-}{x}) = 1 \text{ if } \underset{-}{x} \in R_0 = \{\underset{-}{x}: \frac{\bar{x}_J}{\bar{x}} < c, \forall J \in \mathcal{S}\}$$

and

$$\varphi_I(\underset{-}{x}) = 1 \text{ if } \underset{-}{x} \notin R_0 \text{ and } \{x_i: i \in I\} = \{x_{[k-t+1]}, \ldots, x_{[k]}\}.$$

4. Multinomial Population

Let $\underset{-}{x} = (x_1, \ldots, x_k)$ be an observation from a multinomial distribution with parameters N and $\underset{-}{\theta} = (\theta_1, \ldots, \theta_k)$:

$$p(x_1, \ldots, x_k; \theta_1, \ldots, \theta_k) = N! \prod_{j=1}^{k} [\frac{\theta_j^{x_j}}{x_j!}], \quad \sum_{j=1}^{k} x_j = N, \quad \sum_{j=1}^{k} \theta_j = 1$$

$$0 \leq \theta_j \leq 1, \; 1 \leq j \leq k.$$

In the notation of Section 1.2 we have

$$\Theta_0 = \{\underset{-}{\theta}: \theta_1 = \ldots = \theta_k\} = \{(\tfrac{1}{k}, \ldots, \tfrac{1}{k})\} = \{\theta_0\}, \text{ say}$$

and for $J \in \mathcal{S}$

$$\Theta_J = \{\underset{-}{\theta}: \theta_j = \begin{cases} \omega+\Delta & \text{if } j \in J \\ \omega & \text{if } j \notin J \end{cases}\}$$

where $\Delta = \frac{1-k\omega}{t}$, $0 \leq \omega < \frac{1}{k}$.

Hence $$p_0(\underset{\sim}{x}) = \frac{N!}{k^N \prod_{j=1}^{k} (x_j!)}$$

and

$$p_J(\underset{\sim}{x};\omega) = k^N\, p_0(\underset{\sim}{x})\omega^N \left(1 + \frac{1-k\omega}{t\omega}\right)^{\sum_{j\in J} x_j} \tag{1.4.8}$$

It is easily verified that the multinomial distribution has property M, and hence the set $R_0$ is given by the inequalities

$$\xi_0(L_0-L_J)p_0(\underset{\sim}{x}) + \xi\int_0^{\frac{1}{k}} [L_0(J,\omega)-L_J(J,\omega)]p_J(\underset{\sim}{x};\omega)d\bar{G}(\omega) < 0 \quad \forall J \in \mathcal{S}. \tag{1.4.9}$$

It follows from (1.4.8) and (1.4.9) that

$$R_0 = \{\underset{\sim}{x}: \max_{J\in\mathcal{S}} (\sum_{j\in J} x_j) < c\}$$

and the class of symmetric Bayes procedures is given by Theorem 1.3.2.

## 1.5 Multiple Slippage Problem for Location in the Presence of a Sufficient Statistics (Controlled Case)

Here we have observations $x_1,\ldots,x_k,y$ from k+1 independent populations which are distributed as $f(x-\theta_1),\ldots,f(x-\theta_k)$, $f(y-\theta_0)$ respectively, where $x_i,y,\theta_i,\theta_0$ are all real. Assume that

(i) $f(x)$ is strictly positive

(ii) $f(x-\theta)$ has monotone likelihood ratio (MLR) in $x$ and $\theta$

(iii) $f(x)$ is bounded, and $f(0) = \max_{x \in \mathbb{R}} f(x)$

(There is no loss of generality in the latter part of assumption (iii), since if $f(x_0) = \max_x f(x)$, we can reparametrize the family of densities by setting $\theta_i' = \theta_i + x_0$)

(iv) $\prod_{i=1}^{k+1} f(x_i-\theta) = r(\underset{\sim}{x},y)q(\hat{\theta}-\theta)$, $x_{k+1} \equiv y$

where $\hat{\theta} = \hat{\theta}(\underset{\sim}{x},y)$ is the maximum likelihood estimate (MLE) of $\theta$ under $H_0$: $\theta_i = \theta_0 \forall\, i$.

Loss function is assumed to satisfy (1'), (2'), (3') of Section 1.2, and also the following:

$$L_I(\theta_1+c,\ldots,\theta_k+c,\ \theta_0+c) = L_I(\theta_1,\ldots,\theta_k,\theta_0)\ \forall\ c \in \mathbb{R},\ I \in \{0\} \cup \mathcal{S}.$$

The joint pdf of maximal invariants $U_i = X_i - Y$, $i = 1,\ldots,k$ is given by

$$p(\underset{\sim}{u};\underset{\sim}{\omega}) = \int_{\mathbb{R}} \prod_{i=1}^{k} f(u_i-\omega_i+s)f(s)ds$$

where $\omega_i = \theta_i - \theta_0$, $i = 1,\ldots,k$.

It follows from assumption (ii) on f that $p(\underset{\sim}{u};\underset{\sim}{\omega})$ has property M, and therefore the symmetric invariant Bayes procedures are given by Theorem 1.3.2, with $R_0$ defined as follows:

$$R_0 = \{\underset{\sim}{u} \in \mathbb{R}^k:\ \xi_0(L_0-L_J)p_0(\underset{\sim}{u})+\xi\int_0^\infty[L_0(J,\Delta)-L_J(J,\Delta)]p_J(\underset{\sim}{u};\Delta)d\bar{G}(\Delta) < 0 \quad \forall\ J \in \mathcal{S}\} \tag{1.5.1}$$

where

$$p_J(\underset{\sim}{u};\Delta) = \int_{\mathbb{R}} \prod_{j\in J} f(u_j-\Delta+s) \prod_{j\notin J} f(u_j+s)f(s)ds$$

$$= \int_{\mathbb{R}} \prod_{j \in J} \left[\frac{f(u_j-\Delta+s)}{f(u_j+s)}\right] \prod_{i=1}^{k} f(u_i+s)f(s)ds.$$

From assumption (iii) and the fact that $\hat{\theta}(\underline{x}+c) = \hat{\theta}(\underline{x})+c$, $\forall c \in \mathbb{R}$, $\underline{x} \in \mathbb{R}^k$ (see Lemma 5.2.1 of Karlin and Truax (1960)), we have

$$\prod_{i=1}^{k+1} f(u_i+s) = r(\underline{u})q(\hat{\theta}(\underline{u})+s)$$

where $\underline{u} = (u_1,\ldots,u_k,u_{k+1}) \in \mathbb{R}^{k+1}$, $u_{k+1} \equiv 0$

and

$$\prod_{j \in J} \left[\frac{f(u_j-\Delta+s)}{f(u_j+s)}\right] = \frac{q(\hat{\theta}_J(\underline{u})-\Delta+s)}{q(\hat{\theta}_J(\underline{u})+s)}$$

where $\hat{\theta}_J(\underline{u})$ is the MLE of $\theta$ based on the subset $\{u_j: j \in J\}$. Substituting these values in expression (1.5.1) for $R_0$, and interchanging the order of integration , we obtain

$$R_0 = \{\underline{u}: \int_{\mathbb{R}} \Psi(\hat{\theta}_J(\underline{u})+s)r(\underline{u})q(\hat{\theta}(\underline{u})+s)ds < 0, \forall J \varepsilon \mathcal{S}\}$$

where

$$\Psi(z) = \int_0^\infty \{\xi[L_0(J,\Delta)-L_J(J,\Delta)] \frac{q(z-\Delta)}{q(z)} - \xi_0(L_J-L_0)\}d\bar{G}(\Delta).$$

From Lemma 5.2.2 of Karlin and Truax (1960) we see that $q(\hat{\theta}-\theta)$ has MLR, and therefore $\Psi(z)$ is an increasing function (see Lemma 2, p. 74 of Lehmann (1959)).

Now

$$\int_{\mathbb{R}} \Psi(\hat{\theta}_J(\underline{u})+s)q(\hat{\theta}(\underline{u})+s)ds = \int_{\mathbb{R}} \Psi(z)q(z-[\hat{\theta}_J(\underline{u})-\hat{\theta}(\underline{u})])dz.$$

Since $q(z-[\hat{\theta}_J(\underline{u})-\hat{\theta}(\underline{u})])$ has MLR, and $\Psi(z)$ is increasing, the above integral is negative if $\hat{\theta}_J(\underline{u})-\hat{\theta}(\underline{u}) < c$ for all $J \in \mathcal{S}$, for some $c \in \mathbb{R}$,

Hence

$$R_0 = \{\underline{u}: \max_{J\in\mathcal{S}} [\hat{\theta}_J(\underline{u})-\hat{\theta}(\underline{u})] < c\}$$

$$= \{(\underline{x},y): \max_{J\in\mathcal{S}} [\hat{\theta}_J(\underline{x})-\hat{\theta}(\underline{x},y)] < c\} \tag{1.5.2}$$

by translation invariance of $\hat{\theta}(\underline{u})$.

(b) Uncontrolled Case

In this case assumption (iii) of case (a) is to be replaced by

$$\prod_{i=1}^{k} f(x_i-\theta) = r(\underline{x})q(\hat{\theta}-\theta).$$

The loss function is assumed to satisfy (1), (2), (3) of Section 1.2 and also the following:

$$L_I(\theta_1+c,\ldots,\theta_k+c) = L_I(\theta_1,\ldots,\theta_k)\ \forall\ c \in \mathbb{R} \quad I \in \{0\} \cup \mathcal{S}.$$

A maximal invariant is $\underline{U} = (U_1,\ldots,U_{k-1}) = (X_1-X_k,\ldots,X_{k-1}-X_k)$ which is distributed as

$$p(\underline{u};\omega) = \int_{\mathbb{R}} \prod_{i=1}^{k-1} f(u_i-\omega_i+s)f(s)ds \quad \text{where}$$

$\omega_i = \theta_i-\theta_k$, $i = 1,\ldots,k-1$.

Using the method of Example 2 (b) of Section 1.4, we can show that

$$p_J(\underline{u};\Delta) = \int_{\mathbb{R}} \prod_{j\in J} \left[\frac{f(u_j+z-\Delta)}{f(u_j+z)}\right] \prod_{i=1}^{k} f(u_i+z)dz, \text{ with } u_k \equiv 0.$$

$$= \int_{\mathbb{R}} \frac{q(\hat{\theta}_J+z-\Delta)}{q(\hat{\theta}_J+z)} r(\underline{u})q(\hat{\theta}+z)dz$$

and

$$p_0(\underline{u}) = \int_{\mathbb{R}} r(\underline{u})q(\hat{\theta}+z)dz.$$

Then, as in the controlled case, we have

$$R_0 = \{\underline{u}: \int_{\mathbb{R}} \Psi(\hat{\theta}_J)r(\underline{u})q(\hat{\theta}+s)ds < 0 \text{ , } \forall\ J \in \mathcal{S}\}$$

with $\Psi(\cdot)$ as defined in case (a)

$$= \{\underline{u}: \max_{J\in\mathcal{S}} [\hat{\theta}_J(\underline{u})-\hat{\theta}(\underline{u})] < c\}$$

$$= \{\underline{x}: \max_{J\in\mathcal{S}} [\hat{\theta}_J(\underline{x})-\hat{\theta}(\underline{x})] < c\} \tag{1.5.3}$$

We now prove that the class of Bayes rules of Theorem 1.3.2, with $R_0$ given by (1.5.2), is minimal essentially complete if the loss function satisfies two additional assumptions:

(i) $L: \mathcal{S} \times \Theta^k \to \mathbb{R}$ is bounded

(ii) $\gamma(\Delta) = L_0(J,\Delta) - L_J(J,\Delta)$ is monotonically increasing in $\Delta$

(By assumption (1) of Section 1.2, $\gamma(\Delta)$ is independent of J.)

<u>Lemma 1.5.1</u>: Let $D^{(m)} = \{\Delta_0,\Delta_1,\ldots,\Delta_m\}$. Given a symmetric invariant rule $\varphi$, we can find a symmetric invariant Bayes rule $\varphi^{(m)}$ such that

$$\rho(\varphi,\Delta) - \rho(\varphi^{(m)},\Delta) \geq 0 \ , \forall \ \Delta \in D^{(m)}.$$

Proof: In the terminology of Blackwell and Girshick (1954) consider the game $(D^{(m)}, \mathfrak{a}, \rho)$. Since the set of states of nature $D^{(m)}$ and the action space $\mathfrak{a}$ are finite, the set $W \subset \mathbb{R}^p$ $(p = \binom{k}{t}+1)$ defined by

$$W = \{\underline{w}(a) = (w_0(a),\ldots,w_m(a)): w_i(a) = L(a,\Delta_i)$$
$$i = 0,1,\ldots,m, a \in \mathfrak{a}\}$$

is closed. (Here $L(a,\Delta_i)$ represents the loss incurred in taking action $a \in \mathfrak{a}$ when $\Delta_i \in D^{(m)}$ is the state of nature.)

Moreover, since L is bounded, the set W is also bounded. It follows from Theorem 2, Section 2.10 of Ferguson (1967) that Bayes rules for the game $(D^{(m)}, \mathfrak{a}, \rho)$ form a complete class, and then there exists a Bayes rule $\varphi^{(m)}$ such that

$$\rho(\varphi,\Delta) - \rho(\varphi^{(m)},\Delta) \geq 0 \ , \forall \ \Delta \in D^{(m)}.$$

Theorem 1.5.1: Under assumptions of this section, the class of procedures of form

$$\varphi_0(\underline{x}) = 1 \text{ if } \max_{J \in \mathcal{S}} (\hat{\theta}_J - \hat{\theta}) < c$$

$$\varphi_I(\underline{x}) = 1 \text{ if } \max_{J \in \mathcal{S}} (\hat{\theta}_J - \hat{\theta}) \geq c \text{ and}$$
$$\{x_i : i \in I\} = \{x_{[k-t+1]},\ldots,x_{[k]}\}$$

is a minimal essentially complete class of symmetric invariant procedures.

Proof: We follow the proof of Theorem 5.2.2 of Karlin and Truax (1960). Since $\gamma(D)$ is a monotone function of $\Delta$, there exists a countable dense set $D = \{\Delta_0, \Delta_1, \ldots, \Delta_m, \ldots\}$ with $\Delta_0 = 0$ such that all points of discontinuity of $\gamma(\Delta)$ are contained in D. By Lemma 1.5.1, given a symmetric invariant rule $\varphi$ we can find Bayes rule $\varphi^{(m)}$, which is also symmetric, such that $\varphi^{(m)}$ dominates $\varphi$ at $\Delta_0$, $\Delta_1, \ldots, \Delta_m$. That is

$$\rho(\varphi^{(m)}, \Delta_i) - \rho(\varphi, \Delta_i) \leq 0 \text{ for } i = 0,1,\ldots,m \tag{1.5.5}$$

where, for any rule $\varphi'$

$$\begin{aligned}\rho(\varphi', \Delta) &= L_0(0)\int \varphi'_0(\underline{x})p_0(\underline{x})d\underline{x} + L_J(0) \sum_{I \in \mathcal{S}} \varphi'_I(\underline{x})p_0(\underline{x})d\underline{x} \\ &= -\gamma_0 \int \varphi'_0(\underline{x})p_0(\underline{x})d\underline{x} + L_J(0)\end{aligned} \tag{1.5.6}$$

where $\gamma_0 = L_J(0) - L_0(0) > 0$, and for $\Delta > 0$

$$\begin{aligned}\rho(\varphi', \Delta) &= L_0(\Delta)\int \varphi'_0(\underline{x})p_0(\underline{x})d\underline{x} + \sum_{\substack{I \in \mathcal{S} \\ I \neq J}} L_I(\Delta)\int \varphi'_I(\underline{x}) \cdot p_I(\underline{x};\Delta)d\underline{x} \\ &\quad + L_J(\Delta)\int \varphi'_J(\underline{x})p_J(\underline{x};\Delta)d\underline{x} \\ &= L_0(\Delta)\int [1-\varphi'_J(\underline{x})]p_J(\underline{x};\Delta)d\underline{x} + L_J(\Delta)\cdot\int \varphi'_J(\underline{x})p_J(\underline{x};\Delta)d\underline{x}\end{aligned}$$

by assumption (3) of Section 1.2

$$= -\gamma(\Delta)\int \varphi'_J(\underline{x})p_J(\underline{x};\Delta)d\underline{x} + L_0(\Delta). \tag{1.5.7}$$

By symmetry considerations the right hand sides of (1.5.6) and (1.5.7) are independent of J.

Since $\varphi^{(m)}$ is a symmetric invariant Bayes rule, there exists a constant $c_m$ such that

$$\varphi_0^{(m)}(\underset{\sim}{x}) = 1 \text{ if } \underset{\sim}{x} \in \{\max_{J \in \mathcal{S}} [\hat{\theta}_J - \hat{\theta}] < c_m\} = R_0^{(m)}, \text{ say}$$

$$\varphi_I^{(m)}(\underset{\sim}{x}) = 1 \text{ if } \underset{\sim}{x} \notin R_0^{(m)} \text{ and } \{x_i: i \in I\} = \{x_{[k-t+1]},\ldots,x_{[k]}\}.$$

Let $R_0^* = \{\max_{J \in \mathcal{S}} [\hat{\theta}_J - \hat{\theta}] < c^*\} \subset \mathbb{R}^k$, where $c^*$ is a constant satisfying

$$\int \varphi_0(\underset{\sim}{x}) p_0(\underset{\sim}{x}) d\underset{\sim}{x} = \int_{R_0^*} p_0(\underset{\sim}{x}) d\underset{\sim}{x} \tag{1.5.8}$$

Define a symmetric invariant Bayes rule $\varphi^*$ by

$$\begin{aligned} &\varphi_0^*(\underset{\sim}{x}) = 1 \text{ if } \underset{\sim}{x} \in R_0^* \\ &\varphi_I^*(\underset{\sim}{x}) = 1 \text{ if } \underset{\sim}{x} \notin R_0^* \text{ and } \{x_i: i \in I\} = \{x_{[k-t+1]},\ldots,x_{[k]}\}. \end{aligned} \tag{1.5.9}$$

We show that the rule $\varphi^*$ dominates $\varphi$ for all $\Delta > 0$. It suffices to show that

$$\rho(\varphi^*,\Delta) \leq \rho(\varphi,\Delta) \quad , \forall \Delta \in D$$

since any $\Delta' \notin D$ is a point of continuity of $\gamma(\Delta)$ from which it follows that $\rho(\varphi^*,\Delta') \leq \rho(\varphi,\Delta')$. Let $\Delta = \Delta_m$. Then it follows from (1.5.5), (1.5.6) and (1.5.7) that

$$\int \varphi_0^{(m)}(\underset{\sim}{x}) p_0(\underset{\sim}{x}) d\underset{\sim}{x} = \int_{R_0^{(m)}} p_0(\underset{\sim}{x}) d\underset{\sim}{x} \geq \int \varphi_0(\underset{\sim}{x}) p_0(\underset{\sim}{x}) d\underset{\sim}{x} \tag{1.5.10}$$

and $\int \varphi_J^{(m)}(\underset{\sim}{x}) p_J(\underset{\sim}{x};\Delta_i) d\underset{\sim}{x} \geq \int \varphi_J(\underset{\sim}{x}) p_J(x;\Delta_i) d\underset{\sim}{x}, \ i = 1,\ldots,m.$ (1.5.11)

It is clear from (1.5.8) and (1.5.10) that

$$R_0^{(m)} \supset R_0^*$$

and therefore

$$\int \varphi_J^*(\underline{x})p_J(\underline{x};\Delta_i)d\underline{x} \geq \int \varphi_J^{(m)}(\underline{x})p_J(\underline{x};\Delta_i)d\underline{x}$$

$$\geq \int \varphi_J(\underline{x})p_J(\underline{x};\Delta_i)d\underline{x} \ , \forall\, i = 1,\ldots,m.$$

Since the above holds for every m, the result follows.

The proof of minimality of the essentially complete class is **the same** as in Theorem 5.2.2 of Karlin and Truax (1966).

Remarks:

1. If the direction of slippage is unknown, we have a symmetric two-sided t-slippage problem. The symmetric invariant procedures for this problem are of the following form:

$$\varphi_0(\underline{x}) = 1 \quad \text{if} \quad \underline{x} \in R_0 = \{\underline{x}: \max_{\mathcal{S}} (|\hat{\theta}_J - \hat{\theta}|) < c\}$$

$$\varphi_I(\underline{x}) = 1 \quad \text{if} \quad \underline{x} \notin R_0 \text{ and } |\hat{\theta}_I - \hat{\theta}| = \max_{J \in \mathcal{S}} |\hat{\theta}_J - \hat{\theta}|.$$

2. It is clear from the proof of Theorem 1.5.1 that each symmetric invariant Bayes procedure for testing t-slippage in the location parameter case is uniformly most powerful in the class of all symmetric invariant procedures, having a preassigned error associated with $H_0$.

3. Similar results hold for the uncontrolled case.

Illustrations

1. Normal populations with a common known variance.

It is easy to see that $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, $-\infty < x < \infty$ satisfies all the assumptions of this section, with

$$\hat{\theta} = \begin{cases} \frac{k\bar{x}+y}{k+1} & \text{for the controlled case} \\ \frac{k\bar{x}}{k} & \text{for the uncontrolled case} \end{cases}$$

and $$\hat{\theta}_J = \frac{\sum_{j \in J} x_j}{t} = \bar{x}_J.$$

It is easily seen that the Bayes procedures obtained by substituting the expressions for $\hat{\theta}$ and $\hat{\theta}_J$ in (1.5.2) are same as the ones given in Example 1 of Section 1.4.

2. Location parameter of the exponential density

Here $$f(x-\theta) = \begin{cases} e^{-(x-\theta)} & \text{if } x \geq \theta \\ 0 & \text{otherwise} \end{cases}.$$

For the problem involving a control we have

$$\hat{\theta} = \min\{x_1,\ldots,x_k,y\}$$

and

$$\hat{\theta}_J = \min_{j \in J}\{x_j\}$$

and therefore the set $R_0$ is given by

$$R_0 = \{(\underset{\sim}{x},y): \max_{J \in \mathcal{S}}[\min_{j \in J} x_j - \min\{x_1,\ldots,x_k,y\}] < c\}. \tag{1.5.4}$$

The set $R_0$ for the uncontrolled case is obtained by replacing $(\underset{\sim}{x},y) = (x_1,\ldots,x_k,y)$ by $\underset{\sim}{x} = (x_1,\ldots,x_k)$ in (1.5.4).

## 1.6 Slippage Problem for Scale Parameter Distributions

We discuss the controlled case first. Here $x_i$, $i = 1,\ldots,k$ are independent observations from populations with density functions $\frac{1}{\theta_i} f(\frac{x}{\theta_i})$, $i = 1,\ldots,k$, respectively, and y is an observation from control population which has density $\frac{1}{\theta_0} f(\frac{x}{\theta_0})$; $x > 0$, $\theta_i > 0$, $i = 0,1,\ldots,k$.

As in location case, we assume that

(i) $f(x)$ is strictly positive for all $x > 0$.

(ii) $f(\frac{x}{\theta})$ has MLR in $x$ and $\theta$

(iii) MLE of $\theta_0$ exists under $H_0$: $\theta_1 = \ldots = \theta_k = \theta_0$, and is sufficient for $\theta$.

Loss functions are assumed to satisfy (1'), (2'), (3') of Section 1.2 and the following:

$$L_I(c\theta_1,\ldots,c\theta_k,c\theta_0) = L_I(\theta_1,\ldots,\theta_k,\theta_0)\ \forall\ c > 0,\ I \in \{0\} \cup \mathcal{S}.$$

Analysis similar to that for the location parameter t-slippage problem shows that every symmetric invariant Bayes rule has the form

$$\varphi_0(x_1,\ldots,x_k,y) = 1 \text{ if } (\underset{\sim}{x},y) \in R_0 = \{(\underset{\sim}{x},y) \in \mathbb{R}^{k+1}: \max_{J \in \mathcal{S}}[\frac{\hat{\theta}_J(\underset{\sim}{x},y)}{\hat{\theta}(\underset{\sim}{x},y)}] < c\} \tag{1.6.1}$$

$\varphi_I(\underline{x},y) = 1$ if $\underline{x} \notin R_0$ and $\{x_i: i \in I\} = \{x_{[k-t+1]},\ldots,x_{[k]}\}$.

An analogous result holds for the uncontrolled case.

Remark:

The class of symmetric invariant Bayes rules given by (1.6.1) is minimal essentially complete.

## 1.7 Applications

Selection of t-best populations under slippage configuration.

Let $x_i$ be an observation from population $\pi_i$ with pdf $f(x,\theta_i)$ $(i = 1,\ldots,k)$, $\theta_i \in \Theta \subset \mathbb{R}$; we assume that $f(x,\theta)$ has MLR in x and $\theta$. It is also assumed that the parameters $\theta_i$ $(i = 1,\ldots,k)$ are in the following configuration:

$$\theta = \theta_{[1]} = \ldots = \theta_{[k-t+2]} < \theta_{[k-t+1]} = \ldots = \theta_{[k]} = \theta+\Delta, \quad \theta \in \mathbb{R}, \ \Delta > 0 \tag{1.7.1}$$

where $\theta_{[1]} \leq \ldots \leq \theta_{[k]}$ denote ordered $\theta_i$, and $t(1 \leq t < k)$ is a known integer. The problem is to identify the subset of populations which are associated with $\{\theta_{[k-t+1]},\ldots,\theta_{[k]}\}$.

The action space $\mathcal{a}$ consists of $\binom{k}{t}$ elements:

$$\mathcal{a} = \{a_J: J \in \mathcal{S}\}$$

where $\mathcal{S}$ is as defined in Section 1.2. Assume that the loss function satisfies

(i) $L_{\psi^{-1}I}(\underline{\theta}_\psi) = L_I(\underline{\theta})\ \forall\ I \in \mathcal{S},\ \psi \in S_k$

(ii) $L_J(\underline{\theta}) < L_I(\underline{\theta})$ if $\theta_j = \theta+\Delta, \forall j \in J, I \in \mathcal{S}$ $(I \neq J)$

(iii) $L_I(\underline{\theta}) = L_J(\underline{\theta})$ if $\theta_i = \theta$ for at least one $i \in I$

and $\theta_j = \theta$ for at least one $j \in J$.

Then we have the following theorem :

Theorem 1.7.1: If the above assumptions are satisfied, the symmetric Bayes rules for the problem of selecting the t best populations, when the parameters are in the slippage configuration given by (1.7.4), is to select the set associated with $\{x_{[k-t+1]},\ldots,x_{[k]}\}$.

2. Selection of the normal populations associated with largest probability $P(a < x \leq b)$.

Here we have k normal populations $N(\mu_i,1)$ $i = 1,\ldots,k$, and the problem is to select the subset of populations which are associated with the largest t values of the coverage of the set $A = (a,b] \subset \mathbb{R}$, $-\infty \leq a < b \leq \infty$, where t $(1 \leq t < k)$ is a known integer and the coverage of a set A for a population with distribution function $F(x,\mu)$ is defined as [c.f., Guttman (1961)]

$$\Psi(A,\mu) = \int_A dF(x,\mu).$$

The problem of selecting a subset containing the population associated with the largest coverage for $A = (-\infty,a]$ has been considered by Guttman (1961). Guttman has investigated Gupta type procedures for normal and exponential populations. Guttman and Tiao (1964) have considered the case $A = (-a,b]$ and have derived Bayes rules for two-parameter normal and exponential populations for the problem of selecting the best; the a priori distributions for the two parameters are taken to be independent and locally uniform. We consider the problem of selecting the t best populations.

We have

$$\Psi(\mu) \equiv \Psi((a,b],\mu) = \int_a^b d\Phi(x-\mu) = \Phi(b-\mu)-\Phi(a-\mu)$$

where $\Phi$ denotes the distribution function of a standard normal random variable. It is easily seen that

(i) $\Psi(\frac{a+b}{2} + \mu) = \Psi(\frac{a+b}{2} - \mu)$

i.e., $\Psi$ is symmetric about $\mu_0 = \frac{a+b}{2}$

(ii) $\Psi(\mu) \to 0$ as $|\mu| \to \infty$

(iii) $\frac{d\Psi}{d\mu} \geq 0$ if $\mu \leq \mu_0$

$\leq 0$ if $\mu \geq \mu_0$.

Hence $\Psi(\mu)$ is monotonically increasing in $(-\infty,\mu_0)$ and monotonically decreasing in $(\mu_0,\infty)$. It follows from (i) - (iii) that

$$|\mu_1-\mu_0| < |\mu_2-\mu_0| \Rightarrow \Psi(\mu_1) > \Psi(\mu_2). \tag{1.7.2}$$

Let $\theta_i = |\mu_i-\mu_0|$. It follows from (1.7.2) that the problem of selecting the populations associated with t largest $\Psi(\mu_i)$ is equivalent to the problem of selecting the t populations which correspond to $\{\theta_{[1]},\ldots,\theta_{[t]}\}$.

Let $Y_i = |X_i-\mu_0|$; it is easily shown that if $X_i$ is normal with mean $\mu_i$ and variance 1, the density of $Y_i$ is given by

$$f(y,\theta_i) = \phi(\theta_i+y) + \phi(\theta_i-y),\ \theta_i > 0,\ y > 0$$

where $$\phi(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}.$$

We now show that the function $f(y,\theta) = \phi(\theta+y) + \phi(\theta-y)$, $y > 0$, $\theta > 0$ has MLR in $y$ and $\theta$.

We have

$$\frac{\partial f}{\partial y} = (\theta-y)\phi(\theta-y)-(\theta+y)\varphi(\theta+y)$$

and

$$\frac{\partial^2 f}{\partial\theta\partial y} = \varphi(\theta+y)(\theta+y)^2 + \varphi(\theta-y)(\theta-y)^2$$

$$+ \varphi(\theta-y) - \varphi(\theta+y)$$

$$\geq 0 \text{ since } \varphi(\theta-y) \geq \varphi(\theta+y),\ \forall\ \theta > 0,\ y > 0.$$

It follows from Ex. 6 on p. 111 of Lehmann (1959) that $f(y,\theta)$ has MLR in $y$ and $\theta$.

Then assuming that the loss function $L_J(\underline{\theta})$ satisfies the conditions given in Ex. 1 discussed above, it can be easily seen that a symmetric Bayes rule for the problem is to select the subset of populations associated with $\{y_{[1]},\ldots,y_{[t]}\}$.

3. Selection in terms of entropy for binomial distributions

Let $X_i$, $i = 1,\ldots,k$ be $k$ independent binomial random variables with distributions given by

$$P(X=x,p_i) = \binom{n}{x}p_i^x(1-p_i)^{n-x},\ x = 0,1,\ldots,n,\ i = 1,\ldots,k.$$

Let $\theta_i = H(p_i) = -p_i \log p_i-(1-p_i) \log (1-p_i)$. The function

$H(p_i)$ is the entropy function associated with the binomial population $\pi_i$ [c.f., Gupta and Huang (1976)].

Let $\theta_{[1]} \leq \cdots \leq \theta_{[k-t+1]} \leq \cdots \leq \theta_{[k]}$ denote the ordered $\theta_i$. The populations associated with $\theta_{[k-t+1]},\ldots,\theta_{[k]}$ are of interest. The problem of selecting a subset containing the population corresponding to $\theta_{[k]}$ has been considered by Gupta and Huang (1976). Here we find a Bayes rule for selecting the t populations associated with $\theta_{[k-t+1]},\ldots,\theta_{[k]}$.

Let $\mu_i = |p_i - \frac{1}{2}|$, $i = 1,\ldots,k$. It is shown in Gupta and Huang (1976) that the function $H(p)$ strictly decreases with $\mu = |p - \frac{1}{2}|$, and hence the problem of selecting $\{\theta_{[k-t+1]},\ldots,\theta_{[k]}\}$ is equivalent to the selection of $\{\mu_{[1]},\ldots,\mu_{[t]}\}$.

Assume that $\underset{\sim}{p} = (p_1,\ldots,p_k) \in \bigcup_{J \in \mathcal{S}} \Omega_J = \Omega$ where

where $$\Omega_J = \left\{ \underset{\sim}{p}:\ p_j = \begin{cases} \frac{1}{2} + \omega & \text{if } j \in J,\ \omega \in \mathbb{R},\ \Delta \neq 0 \\ \frac{1}{2} + \omega + \Delta & \text{if } j \notin J \quad |\omega| < |\omega+\Delta| \end{cases} \right\}.$$

The action space $\mathcal{A} = \{a_J\colon J \in \mathcal{S}\}$ consists of $\binom{k}{t}$ elements.

The loss function $L_J(\underset{\sim}{p})$ is assumed to satisfy (i)and (ii) of Ex. 1 of this section and also the following:

(iii) $L_J(p) < L_I(\underset{\sim}{p}),\ \forall\, \underset{\sim}{p} \in \Omega_J,\ I,J \in \mathcal{S},\ I \neq J$

(iv) $L_J(p) = L_J(\underset{\sim}{1}-\underset{\sim}{p}),\ \forall\, \underset{\sim}{p} \in \Omega,\ J \in \mathcal{S}$

where $\underset{\sim}{1} = (1,\ldots,1)$ is a row vector in $\mathbb{R}$.

Since the problem is invariant with respect to the group $G = \{e,g\}$, where e is the identity element and $gx_i = n-x_i$, we can restrict attention to procedures based on maximal invariants $Y_i = |X_i - \frac{n}{2}|$, $i = 1,\ldots,k$.

Let $Q(y,\theta_i)$ denote the probability distribution of the random variable $Y_i = |X_i - \frac{n}{2}|$, given $\theta_i = |p_i - \frac{1}{2}|$. Then $Q(y,\theta_i)$ has MLR in $y$ and $\theta_i$ [c.f., Sobel and Starr (1975)] and hence a Bayes rule against an exchangeable prior is to select the populations associated with $y_{[1]},\ldots,y_{[t]}$, the t smallest values of the observed values of $Y_i$.

## 4. Selecting the biased coin

We are given $k \geq 2$ coins, one of which may be biased. Our goal is to decide, on the basis of n tosses of each coin, whether all the coins are fair, and, if not, which one is biased.

Let $p_i$ = probability of getting Head for i-th coin ($i = 1,\ldots,k$). Then $\theta_i = |p_i - \frac{1}{2}|$ is called the bias of the i-th coin [c.f., Sobel and Starr (1975)]. The problem is to test

$$H_0: \ \theta_1 = \ldots = \theta_k = 0$$

against k alternatives

$$H_j: \ \theta_i = \begin{cases} \Delta \text{ if } & i = j \\ 0 \text{ if } & i \neq j \end{cases}, \quad j = 1,\ldots,k, \ 0 < \Delta \leq \frac{1}{2}.$$

Let $x_i$ be the number of heads in n tosses of coin i. Invariance reduces the problem to consideration of tests based on maximal invariants $Y_1 = |X_1 - \frac{1}{2}|, \ldots, Y_k = |X_k - \frac{1}{2}|$.

The random variables $Y_i$, $i = 1,\ldots,k$ are independent, and $Y_i$ is distributed as

$$Q(y,\theta_i) = \binom{n}{\frac{n}{2}+y}[\theta_i^{\frac{n}{2}+y}(1-\theta_i)^{\frac{n}{2}-y} + \theta_i^{\frac{n}{2}-y}(1-\theta_i)^{\frac{n}{2}+y}].$$

Let $Q_0(\underline{y},\underline{\theta})$ and $Q_j(\underline{y},\underline{\theta})$ denote the distribution of $\underline{Y} = (Y_1,\ldots,Y_k)$ under $H_0$ and $H_j$ respectively. Then simple calculation yields

$$Q_0(\underline{y}) = \prod_{i=1}^{k} \left[\frac{\binom{n}{\frac{n}{2}+y_i}}{2^{n-1}}\right]$$

and $Q_j(\underline{y};\Delta) = Q_0(\underline{y})h(y_j;\Delta)$

where

$$h(y;\Delta) = 2^{n-1}[\Delta(1-\Delta)]^{\frac{n}{2}}[(\frac{\Delta}{1-\Delta})^y + (\frac{1-\Delta}{\Delta})^y]. \tag{1.7.3}$$

For each fixed $\Delta$, we have

$$\frac{dh}{dy} = 2^{n-1}[\Delta(1-\Delta)]^{n-1}[\{(\frac{1-\Delta}{\Delta})^y - (\frac{\Delta}{1-\Delta})^y\}\log\frac{1-\Delta}{\Delta}]$$

$$\geq 0 \;,\; \forall\, y \quad \text{since } 0 < \Delta < \frac{1}{2},$$

Hence the function $h(y;\Delta)$ increases with $y$. Assume that the loss function satisfies assumptions (1'), (2'), (3') of Section 1.2, with $\theta_0 = 0$, and $t = 1$. Then all the assumptions of Theorem 1.3.2 are met, and a Bayes rule against an exchangeable a priori distribution has the form

$\varphi_0(\underset{\sim}{y}) = 1$ if $\underset{\sim}{y} \in R_0$, a symmetric set

$\varphi_j(\underset{\sim}{y}) = 1$ if $\underset{\sim}{y} \notin R_0$ and $y_j = \max\limits_{1\le i\le k} y_i$.

Here the set $R_0$ is given by the following inequalities:

$$\xi_0(L_0-L_j)Q_0(\underset{\sim}{y}) + \xi\int_0^{\frac{1}{2}}[L_0(j,\Delta)-L_j(j,\Delta)]Q_j(\underset{\sim}{y};\Delta)d\bar{G}\Delta < 0 \quad \forall\ j = 1,\ldots,k.$$

Substituting for $Q_j(\underset{\sim}{y};\Delta)$ we get

$$\int_0^{\frac{1}{2}}\{\xi[L_0(j,\Delta)-L_j(j,\Delta)]h(y_j;\Delta)-\xi_0(L_j-L_0)\}Q_0(\underset{\sim}{y})\cdot d\bar{G}(\Delta) \quad \forall\ j = 1,\ldots,k$$

where $h(y;\Delta)$ is given by (1.7.3).

Since the integrand is an increasing function of $y_j$, we have

$$R_0 = \{\underset{\sim}{y}: \max_{1\le i\le k} y_i < c\}$$

## 1.8 Nonparametric Slippage Tests

Let $X_{ij}$ $(j = 1,\ldots,n)$ be n independent observations from population $\pi_i$ with continuous cdf $F_i$ $(i = 1,\ldots,k)$. Let $g: [0,1]\to[0,1]$ be a continuous distribution function, $g(x) \not\equiv x$. The subset $\{\pi_j: j \in J\}$, where $J \in \mathcal{S}$, is said to have slipped if

$$F_j = \begin{cases} g(F) & \text{if } j \in J \\ F & \text{if } j \notin J \end{cases}.$$

The problem is to test

$$H_0: F_i = F, \quad i = 1,\ldots,k$$

against $\binom{k}{t}$ alternatives

$$H_J: \{\pi_j: j \in J\} \text{ has slipped.}$$

Karlin and Truax (1960) have considered the case $t = 1$ for the following two types of slippage:

(i) $g(F) = (1-\lambda)F + \lambda F^2$, $0 < \lambda < 1$

(ii) $g(F) = F^{1+\lambda}$, $\lambda > 0$.

We will investigate the general case $(1 \leq t < k)$

(i) $g(F) = (1-\lambda)F + \lambda F^2$.

Let $r_{ij}$ be the rank of $X_{ij}$ in the combined sample, and $R = (r_{ij})$, $i = 1,\ldots,k$, $j = 1,\ldots,n$. Since the problem is invariant under monotone transformations, we can restrict attention to procedures based on $R$.

Let $P_J^{(\lambda)}(R)$ denote the distribution of $R = (r_{ij})$ under $H_J$. It follows from a result of Lehmann (1953) that

$$P_J^{(\lambda)}(R) = \frac{1}{\binom{nk}{nk-nt}} E\left[ \prod_{\substack{1 \leq \ell \leq n \\ j \in J}} \{(1-\lambda)+2\lambda\, U^{(r_{j\ell})}\} \right] \tag{1.8.1}$$

where $U^{(m)}$, $1 \leq m \leq nk$ is the m-th order statistic from a sample of size nk from a uniform distribution on [0,1]. Differentiating (1.8.1) with respect to $\lambda$ we get

$$\frac{d}{d\lambda} P_J^{(\lambda)}(R) = \frac{1}{\binom{nk}{nk-nt}} E\,[\prod_{\substack{1<\ell<n \\ j\in J}} \{(1-\lambda)+2\lambda\, U^{(r_{j\ell})}\} \cdot$$

$$\cdot\{\sum_{j\in J} \sum_{\ell=1}^{n} \frac{2\, U^{(r_{j\ell})}-1}{(1-\lambda)+2\lambda U^{(r_{j\ell})}}\}]$$

$$\frac{d}{d\lambda} P_J^{(\lambda)}(R)\Big|_{\lambda=0} = \frac{1}{\binom{nk}{nk-nt}} \sum_{j\in J} (\frac{2r_{j.}}{nk+1}) - nt$$

where $r_{j.} = \sum_{\ell=1}^{n} r_{j\ell}$.

Following Karlin and Truax (1960), we find regions $C_0$, $\{C_J: J \in \mathcal{S}\}$ in the set of all possible ranks such that

$$P_0(C_0) = 1-\alpha, \quad 0 < \alpha < 1$$

$P_J^{(\lambda)}(C_J)$ is maximum over $\mathcal{S}$ for small $\lambda$

and the procedure is symmetric. It follows that $C_I$ has the form

$$\max_{J\in\mathcal{S}} \{\sum_{j\in J} r_{j.}\} = \sum_{j\in I} r_{j.} > \gamma_1 \tag{1.8.2}$$

where $\gamma_1$ is given by

$$P_0(\max_{J\in\mathcal{S}} \sum_{j\in J} r_{j.} \leq \gamma_1) = 1-\alpha \tag{1.8.3}$$

(ii) $g(F) = F^{1+\lambda}$, $\lambda > 0$.

As before, we will consider procedures based on ranks only.

Let $w_1 < \ldots < w_{nk}$ be the sequence obtained by ordering the combined sample $\{x_{ij}: i = 1,\ldots,k,\ j = 1,\ldots,n\}$. For each $J \in \mathcal{S}$, we define a new sequence $Z_1^{(J)},\ldots,Z_{nk}^{(J)}$ as follows:

$$Z_i^{(J)} = \begin{cases} 1 & \text{if } w_i \text{ comes from } \{\pi_j: j \in J\} \\ 0 & \text{otherwise} \end{cases}.$$

Let $P_J^{(\lambda)}(\underline{z})$ denote the probability of a rank order $\underline{z}$ under $H_J$. Then from a result of Savage (1956) we have

$$P_J^{(\lambda)}(\underline{z}) = \frac{(nk-nt)!(nt)!(1+\lambda)^{nt}}{\prod_{i=1}^{nk} [\sum_{\ell=1}^{i} \{z_\ell^{(J)}(1+\lambda)+1-z_\ell^{(J)}\}]} \tag{1.8.4}$$

Differentiating (1.8.4) at $\lambda = 0$ we get

$$\frac{d}{d\lambda} P_J^{(\lambda)}(\underline{z})\Big|_{\lambda=0} = \frac{(nk-nt)!(nt)!}{(nk)!} [nt - \sum_{i=1}^{nk} \sum_{\ell=1}^{i} (\frac{z_\ell^{(J)}}{i})].$$

Then, as in case (i), the symmetric rule $\varphi$ which maximizes $P_J^{(A)}(C_J)$ for $\lambda$ in the neighborhood of zero among the class of rules satisfying

$$P_0(C_0) = 1-\alpha$$

is given by

$$\varphi_0(\underline{z}) = 1 \text{ if } \min_{J\in\mathcal{S}} \sum_{i=1}^{nk} \sum_{\ell=1}^{i} (\frac{z_\ell^{(J)}}{i}) \geq \gamma_2 \tag{1.8.5}$$

$$\varphi_I(\underline{z}) = 1 \text{ if } \min_{J\in\mathcal{S}} \sum_{i=1}^{nk} \sum_{\ell=1}^{i} (\frac{z_\ell^{(J)}}{i}) = \sum_{i=1}^{nk} \sum_{\ell=1}^{i} (\frac{z_\ell^{(I)}}{i}) < \gamma_2. \tag{1.8.6}$$

## 1.9 Computation of the Constant c

The Bayes rule derived in this chapter are of the form

$$\varphi_0(\underset{\sim}{x}) = 1 \text{ if } \underset{\sim}{x} \in \{\underset{\sim}{x}: \max_{J \in \mathcal{S}} U_J(\underset{\sim}{x}) < c\} = R_0, \text{ say,}$$

$$\varphi_I(\underset{\sim}{x}) = 1 \text{ if } \underset{\sim}{x} \notin R_0,\ U_I(\underset{\sim}{x}) > U_J(\underset{\sim}{x}),\ \forall J \neq I$$

where $U_J(\underset{\sim}{x})$ is a symmetric function of $\underset{\sim}{x}$; the constant c depends on the a priori distribution G. Following Karlin and Truax (1960), c may be determined so that $E(\varphi_0|H_0) = 1-\alpha$; this distribution of $\max_{J \in \mathcal{S}} U_J(\underset{\sim}{x})$ is needed for this purpose. For normal and gamma populations, the distribution of $U_J$ are given in Doornbos (1966). For nonparametric slippage tests, the constant c can be computed from direct calculations. In general case the distribution has not been worked out and approximations are needed.

# CHAPTER II

# A BAYES APPROACH TO SELECTION OF t BEST POPULATIONS

## 2.1 Introduction

Let $\pi_1,\ldots,\pi_k$ be k independent populations, which have probability densities $f(x,\theta_k),\ldots,f(x,\theta_k)$, respectively, with respect to a $\sigma$-finite measure $\mu$. Let $\theta_{[1]} \leq \cdots \leq \theta_{[k]}$ be the ordered $\theta_i$. For a given integer t, $1 \leq t < k$, the subset of populations associated with $\theta_{[k-t+1]},\ldots,\theta_{[k]}$ will be referred to as the set of t-best populations. In many practical situations the set of t-best populations is of interest. Multiple decision (selection and ranking) rules for the problem have been considered in the literature by Bechhofer (1954), Carroll, Gupta and Huang (1974), and others. Gupta's approach, which is commonly known as the subset selection formulation, is to select a subset S of random size $|S|$ $(t \leq |S| \leq k)$ such that S contains the set of t-best populations with probability at least P*, where P* is a preassigned constant. In Bechhofer's indifference zone formulation, a subset of size t is chosen so that the probability of selecting the t-best populations is at least P* whenever $\theta_{[k-t+1]} - \theta_{[k-t]} \geq \delta^*$, where $\delta^* > 0$ is a constant specified in advance.

In this chapter, the problem of selecting a subset containing the t-best populations has been considered from Bayes approach. Deely and Gupta (1968) have considered the problem of selecting a subset containing the best, which corresponds to $t = 1$. It is proved there that under certain assumptions on $f(x,\theta)$ the Bayes rule selects exactly one population associated with the largest observation, provided the loss is linear in $\theta_i$, and the a priori distributions of $\theta_i$ are independent. The problem for general t is formulated in 2.2 and it is shown in Section 2.3 that if the joint distribution has property M of Section 1.2, and the a priori distribution $g(\underline{\theta})$ on on $\Theta^k$ is exchangeable, then for a loss function which is linear in $\theta_i$, the Bayes rule will select exactly t populations corresponding to the t largest observations. For $t = 1$ Bayes rules for non-linear loss functions have recently been considered by Goel and Rubin (1975), Chernoff and Yahav (1977), and Bickel and Yahav (1977). In Section 2.4 we derive some Bayes rules for selection of a subset containing the t best populations, when the loss functions are non-linear in $\theta_i$; normal, exponential, Poisson and binomial populations have been discussed.

## 2.2 Formulation of the Problem

Let $\pi_1,\ldots,\pi_k$ be k independent populations with densities $f(x,\theta_1),\ldots,f(x,\theta_k)$, respectively, with respect to a $\sigma$-finite measure $\mu$, $\theta_i \in \Theta \subset \mathbb{R}$ $(i = 1,\ldots,k)$. The goal is to select a subset containing the t best populations, i.e., the populations associated with $\theta_{[k-t+1]},\ldots,\theta_{[k]}$, where $\theta_{[1]} \leq \cdots \leq \theta_{[k]}$ are ordered $\theta_i$.

The action space $\mathcal{G}$ consists of $\sum_{m=t}^{k} \binom{k}{m}$ elements:

$$\mathcal{G} \doteq \{S \subset \{1,\ldots,k\}: \ |S| \geq t\}$$

where $|S|$ denotes the size of the selected subset S.

The loss function $L(\underline{\theta},S): \ \Theta^k \times \mathcal{G} \to \mathbb{R}$ gives the amount of loss incurred in selecting a subset $S \in \mathcal{G}$, when $\underline{\theta} \in \Theta^k$ is the true value of the parameter.

The following loss functions will be considered:

$$L_1(\underline{\theta},S) = c|S| + \left[ \frac{\sum_{i=k-t+1}^{k} \theta_{[i]}}{t} - \frac{\sum_{j\in S} \theta_j}{|S|} \right] \qquad (2.2.1)$$

where $c \geq 0$ is a known constant.

$$L_2(\underline{\theta},S) = c|S| + \beta \left[ \frac{\sum_{i=k-t+1}^{k} \theta_{[i]}}{t} - \frac{\sum_{j\in S} \theta_j}{S} \right] + \sum_{i=k-t+1}^{k} I_{\{\theta_{[i]} \notin S'\}} \qquad (2.2.2)$$

where $\beta$ and c are given non-negative constants, $I_A$ represents the indicator function of a set A, and

$$S' = \{\theta_i: \ i \in S\}$$

(iii) $L_3(\underline{\theta},S) = c|S| + \sum_{i=1}^{t} [\theta_{[k-i+1]} - \theta_{\{S_{t-i+1}\}}] + \sum_{j\in B} (\theta_{[k]} - \theta_j)$ . (2.2.3)

where $\theta_{\{S_1\}} \leq \cdots \leq \theta_{\{S_t\}}$ are the t largest $\theta_i$ in the subset S, c is a known nonnegative constant and

$$B = \{i \in S: \ \theta_i \neq \theta_{\{S_\ell\}}, \ \ell = 1,\ldots,t\}$$

We assume that

(i) $f(x,\theta)$ has monotone likelihood ration in x and $\theta$

(ii) $g(\underline{\theta})$, the a priori distribution on $\Theta^k$, is exchangeable or permutation symmetric. That is,

$$g(\theta_1,\ldots,\theta_k) = g(\theta_{\psi 1},\ldots,\theta_{\psi k})$$

for all permutations $\psi$: $\{1,\ldots,k\} \to \{1,\ldots,k\}$.

(iii) the measure $\mu$ is also symmetric, i.e.,

$$d\mu(\underline{x}) = d\mu(\underline{x}_\psi).$$

Since only Bayes rules will be considered in this chapter, attention can be restricted to the class of non-randomized rules (see Ferguson (1967), Section 1.8).

It is easily verified that the loss functions $L_1$, $L_2$ and $L_3$ satisfy monotonicity and invariance properties of Eaton (1967), and hence, in each case, the Bayes rule d* for selecting a subset containing $\{\theta_{[k-t+1]},\ldots,\theta_{[k]}\}$ is given by

$$r(d^*,\underline{x}) = \min_{t\le j\le k} r(d_j,\underline{x}) \tag{2.2.4}$$

where $r(d,\underline{x})$ is the Bayes posterior risk, given $\underline{x}$, of a rule d, and $d_j$ in the rule which selects the subset of populations associated with $\{x_{[k-j+1]},\ldots,x_{[k]}\}$; $x_{[1]} \le \cdots \le x_{[k]}$ denote the ordered observations.

In the next two sections we derive Bayes rules for the loss functions given above.

## 2.3 Bayes Rule for the Loss Function $L_1$.

Assume that the loss in selecting a subset S when $\underline{\theta}$ is the parameter is $L_1(\underline{\theta},S)$, where $L_1$ is given by (2.2.1).

Deely and Gupta (1968) have considered the case t = 1 with loss function given by

$$L(S,\underline{\theta}) = \sum_{q \in S} \alpha_q(\theta_{[k]} - \theta_q), \ \alpha_q \geq 0$$

It is shown that under certain mild conditions on the density f and the constants $\alpha_q$, the Bayes rule selects only one population. In this section we prove a similar result for the general case.

Let $x_i \in \mathbb{R}$ be an observation from $\pi_i$, i = 1,...,k. The posterior risk of a decision $\delta$: $\mathcal{X}^k \to \mathcal{A}$ is given by

$$r(\delta,\underline{x}) = c|S| + E\left[\frac{\sum_{i=k-t+1}^{k} \theta_{[i]}}{t} - \frac{\sum_{j \in S} \theta_j}{|S|} \;\Big|\; \underline{x}\right]$$

where rule $\delta$ selects the subset S with probability one.

It follows from (2.2.4) that the search for a Bayes rule for the problem can be restricted to the rules $d_j$ (j = t,...,k).

Let $\theta_{(i)}$ denote the parameter values associated with $x_{[i]}$, i = 1,...,k. Then

$$r(d_{m+1},\underline{x}) - r(d_m,\underline{x}) = c + E\left[\frac{\sum_{\alpha=k-m+1}^{k} (\theta_{(\alpha)} - \theta_{(k-m)})}{m(m+1)} \;\Big|\underline{x}\right]$$

We need the following Lemma:

Lemma 2.3.1: If a density $h(\underset{\sim}{x};\underset{\sim}{\theta})$ has property M, as defined in Section 1.2, and $g(\underset{\sim}{\theta})$ is exchangeable, then

$$E(\theta_{(i)}|\underset{\sim}{x}) \geq E(\theta_{(j)}|\underset{\sim}{x}) \quad \forall\ i \geq j.$$

Proof: Let $B = \{\underset{\sim}{\theta} \in \Theta^k: \ \theta_{(i)} \geq \theta_{(j)}\}$. Then

$$\int_{\Theta^k} [\theta_{(i)} - \theta_{(j)}] h(\underset{\sim}{x};\underset{\sim}{\theta}) g(\underset{\sim}{\theta}) d\underset{\sim}{\theta} = \int_B [\theta_{(i)} - \theta_{(j)}] h(\underset{\sim}{x};\underset{\sim}{\theta}) g(\underset{\sim}{\theta}) d\underset{\sim}{\theta}$$

$$+ \int_{B^c} [\theta_{(i)} - \theta_{(j)}] h(\underset{\sim}{x};\underset{\sim}{\theta}) g(\underset{\sim}{\theta}) d\underset{\sim}{\theta}$$

$$= \int_B [\theta_{(i)} - \theta_{(j)}] h(\underset{\sim}{x};\underset{\sim}{\theta}) g(\underset{\sim}{\theta}) d\underset{\sim}{\theta}$$

$$+ \int_B [\theta_{(j)} - \theta_{(i)}] h(\underset{\sim}{x};\underset{\sim}{\theta}_{\psi_{ij}}) g(\underset{\sim}{\theta}) d\underset{\sim}{\theta}$$

where $\psi_{ij}$ is the permutation which interchanges coordinates i and j of a vector $\underset{\sim}{\theta} = (\theta_1,\ldots,\theta_k)$, keeping other components fixed, and $\underset{\sim}{\theta}_{\psi_{ij}}$ is as in Section 1.2. Thus

$$E(\theta_{(i)} - \theta_{(j)}|\underset{\sim}{x}) = \int_B [\theta_{(i)} - \theta_{(j)}][h(\underset{\sim}{x};\underset{\sim}{\theta}) - h(\underset{\sim}{x};\underset{\sim}{\theta}_{\psi_{ij}})] g(\underset{\sim}{\theta}) d\underset{\sim}{\theta} \geq 0.$$

It follows from Lemma 2.3.1 that

$$E[\theta_{(\alpha)} - \theta_{(k-m)}|\underset{\sim}{x}] \geq 0 \text{ for } \alpha = k-m+1,\ldots,k.$$

Hence

$$r(d_{m+1},\underset{\sim}{x}) \geq r(d_m,\underset{\sim}{x}) \text{ for } m=t,t+1,\ldots,k-1.$$

Theorem 2.3.1: If $f(x,\theta)$ has MLR in x and $\theta$, and the a priori distribution on $\Theta^k$ is exchangeable, the Bayes rule selects exactly t populations associated with the largest t observations.

Proof: Let $h(\underset{\sim}{x};\underset{\sim}{\theta}) = \sum_{i=1}^{k} f(x_i,\theta_i)$.

It is clear from definition that $h(\underset{\sim}{x};\underset{\sim}{\theta})$ has property M, and the result follows from Lemma 2.3.1.

## 2.4 Bayes Rules for the Loss Functions $L_2$ and $L_3$.

The Bayes posteriori risk for rule $d_j$, when the loss function is $L_2(\underset{\sim}{\theta},S)$ given by (2.2.2), is

$$r(d_j,\underset{\sim}{x}) = H(j) + \frac{\beta}{t} \sum_{i=k-t+1}^{k} E(\theta_{[i]}|\underset{\sim}{x}) \quad t \leq j \leq k \tag{2.4.1}$$

where

$$H(j) = cj - \frac{\beta}{j} \sum_{i=k-j+1}^{k} E(\theta_{(i)}|\underset{\sim}{x}) + \sum_{p=k-t+1}^{k} \sum_{q=1}^{k-j} P[\theta_{(q)} = Q_{[p]}|\underset{\sim}{x}) \tag{2.4.2}$$

It is clear that the Bayes rule $d^{(2)}$ in this case is obtained by minimizing $H(j)$, $t \leq j \leq k$. Hence

$$d^{(2)} = d_\ell \quad \text{if} \quad H(\ell) = \min_{t \leq j \leq k} H(j) \tag{2.4.3}$$

When the loss function is $L_3(\underset{\sim}{\theta},S)$ given by (2.2.3), the Bayes posterior risk in using $d_j$ is

$$r(d_j,\underline{x}) = cj + E\left[\left\{\sum_{i=k-t+1}^{k} \theta_{[i]} - \sum_{i\in S} \theta_i\right\} + (j-t)\theta_{[k]} \,\middle|\, \underline{x}\right] \tag{2.4.4}$$

and therefore

$$\Delta_m \equiv r(d_{m+1},\underline{x}) - r(d_m,\underline{x}) = c + E[\theta_{[k]} - \theta_{(k-m)}\ \underline{x}].$$

It follows from Lemma 2.3.1 that $\Delta_{m+1} \geq \Delta_m$, $t \leq m \leq k$, and hence the Bayes rule $d^{(3)}$ is given by

$$d^{(3)} = \begin{cases} d_k & \text{if } A = \{j:\ \Delta_j \geq 0\} \text{ is null set} \\ d_\ell, & \text{where } \ell = \min\{i : i \in A\} \text{ if} \\ & \quad A \text{ is non-null} \end{cases} \tag{2.4.5}$$

2.5 Specific Examples.

In this section, we compute Bayes procedures for the loss functions $L_2$ and $L_3$ in several specific cases.

1. Selection of t best normal means.

Let population $\pi_i$ be normal with mean $\theta_i \in (-\infty,\infty)$ and a known variance $\sigma^2$ $(i = 1,\ldots,k)$, and let $\bar{x}_i$ be the mean of an independent sample $x_{i_1},\ldots,x_{i_n}$ of size n from $\pi_i$; $\pi_i$ and $\pi_j$ $(i, j = 1,\ldots,k,\ i \neq j)$ are assumed to be independent. We compute the Bayes rule when the loss function is $L_2(\underline{\theta},S)$ given by (2.2.2) and the a priori distribution on $\Theta^k$ is k-variate normal with mean vector $\underline{\mu} = (\mu,\ldots,\mu)'$ and covariance matrix $\Sigma = \tau^2 I_k$ where $I_k$ is the identity matrix of order k, i.e.,

$$g(\theta_1,\ldots,\theta_k) = \sum_{i=1}^{k} [\frac{1}{\tau}\,\phi(\frac{\theta_i - \mu}{\tau})]$$

where $\phi(x)$ is the density of a standard normal random variable.

The posterior distribution of $\underline{\theta} = (\theta_1,\ldots,\theta_k)$ given the observation vector $\underline{x} = (x_1,\ldots,x_k)$ is

$$g^*(\underline{\theta}|\underline{x}) = \prod_{i=1}^{k} [\frac{1}{v}\;\phi(\frac{\theta_i - m_i}{v})] \tag{2.5.1}$$

where

$$m_i = \frac{\frac{\mu\sigma^2}{n} + \bar{x}_i\tau^2}{\frac{\sigma^2}{n} + \tau^2}, \quad i = 1,\ldots,k$$
$$v = \left[\frac{\frac{\tau^2\sigma^2}{n}}{\tau^2 + \frac{\sigma^2}{n}}\right]^{\frac{1}{2}} \tag{2.5.2}$$

Let $\bar{x}_{[1]},\ldots,\bar{x}_{[k]}$ be **ordered** sample means, and let $\theta_{(i)}$ be the (unknown) population mean associated with $\bar{x}_{[i]}$, $i = 1,\ldots,k$. Then

$$E(\theta_{(i)}|\bar{x}_1,\ldots,\bar{x}_k) = m_{[i]} \tag{2.5.3}$$

where $m_{[1]} \leq \cdots \leq m_{[k]}$ are the ordered posterior means.

We also have

$$P(\theta_{(q)} = \theta_{[p]}|\bar{x}_1,\ldots,\bar{x}_k)$$

$$= \int_{\Theta} P_{\underset{\sim}{x}}[\max_{1\le a\le p-1} \theta_{(i_a)} < y, \min_{j\in R(i_a)} \theta_{(j)} > y \ \forall\{i_1,\ldots,i_{p-1}\}\subset K(q)] \, dG_{\theta_{(q)}}(y) \tag{2.5.4}$$

where

$$K(q) = \{1,\ldots,q-1,q+1,\ldots,k\}$$

$$R(i_a) \equiv R(i_1,\ldots,i_{p-1}) = K(q)\cap\{i_1,\ldots,i_{p-1}\}^c$$

$P_{\underset{\sim}{x}}$ is the joint conditional distribution of $\underset{\sim}{\theta}$ given $\underset{\sim}{x}$, and $G(\cdot)$ is the posterior cumulative distribution function (cdf) of $\theta_{(\ell)}$.

From (2.5.1) we have

$$G_{\theta_{(\ell)}}(y) = \Phi\left(\frac{y - m_{[\ell]}}{v}\right), \ \ell = 1,\ldots,k \tag{2.5.5}$$

where $\Phi(\cdot)$ is the cdf of a standard normal random variable.

From (2.5.4) and (2.5.5) we obtain

$$P(\theta_{(q)} = \theta_{[p]} | \bar{x}_1,\ldots,\bar{x}_k)$$

$$= \int_{-\infty}^{\infty} \sum_{\{i_1,\ldots,i_{p-1}\}\subset K(q)} \prod_{a=1}^{p-1} \Phi\left(u + \frac{\bar{x}_{[q]} - \bar{x}_{[i_a]}}{\frac{\tau^2\sigma^2}{n}(\frac{\sigma^2}{n}+\tau^2)^{\frac{1}{2}}} \cdot \tau^2\right) \cdot$$

$$\cdot \prod_{j\in R(i_a)} \left[1 - \Phi\left(u + \frac{\bar{x}_{[q]} - \bar{x}_{[j]}}{\frac{\tau^2\sigma^2}{n}(\frac{\sigma^2}{n}+\tau^2)^{\frac{1}{2}}} \cdot \tau^2\right)\right]$$

$$\cdot \ \phi(u)du \tag{2.5.6}$$

Using (2.5.3) and (2.5.6), the function H(j) given by (2.4.2) can be computed for each j, $t \leq j \leq k$ and then the Bayes rule $d^{(2)}$ can be obtained from (2.4.3).

2. Selection of normal populations associated with t smallest variances

Here $\pi_1,\ldots,\pi_k$ are independent normal populations each with mean zero and variances $\frac{1}{\theta_1},\ldots,\frac{1}{\theta_k}$, respectively. We wish to select a subset containing $\theta_{[k-t+1]},\ldots,\theta_{[k]}$ when the loss function is given by (2.2.2) or (2.2.3).

Let $x_{ij}$ $(i = 1,\ldots,k;\ j = 1,\ldots,n)$ be an independent sample from $\pi_i$. Then

$$s_i^2 = \frac{1}{n} \sum_{j=1}^{n} X_{ij}^2$$

is sufficient for $\theta_i$.

Assume that $\theta_1,\ldots,\theta_n$ are independent, and $\theta_i$ has natural conjugate gamma-2 priori density[see p. 54, Raiffa and Schlaifer (1961)]. Then the joint a priori distribution of $\underset{\sim}{\theta} = (\theta_1,\ldots,\theta_k)$ is

$$g(\underset{\sim}{\theta}) = \prod_{i=1}^{k} \{A\, e^{-\frac{1}{2}\theta_i a'b'}\, \theta_i^{\frac{1}{2}a'-1}\} \equiv \prod_{i=1}^{k} f_{\gamma_2}(\theta_i|b',a') \quad (2.5.7)$$

where A is the normalizing constant for $f_{\gamma_2}(\cdot|b',a')$, the pdf of a gamma-2 distribution.

From Raiffa and Schlaifer (1961) we have the posterior distribution of $\underset{\sim}{\theta}$ as

$$g^*(\theta_1,\ldots,\theta_k|s_1^2,\ldots,s_k^2) = \prod_{i=1}^{k} f_{\gamma_2}(\theta_i|b_i'',a'') \qquad (2.5.8)$$

where

$$a'' = a'+n$$
$$b_i'' = \frac{a'b' + ns_i^2}{a''} \qquad (2.5.9)$$

Let $\theta_{(k-j+1)}$ be the parameter associated with $s_{[j]}^2$ where $s_{[1]}^2 \leq \cdots \leq s_{[k]}^2$. Then

$$E(\theta_{(k-j+1)}|s_1^2,\ldots,s_k^2) = \frac{a' + n}{a'b' + ns_{[i]}^2} \qquad (2.5.10)$$

We now have

$$H(j) = cj - \frac{\beta}{j} \sum_{i=1}^{j} \frac{a' + n}{a'b' + ns_{[i]}^2} + \sum_{p=k-t+1}^{k} \sum_{q=1}^{k-j} P(\theta_{(q)} = \theta_{[p]}|s_1^2,\ldots,s_k^2)$$

We can compute $P(\theta_{(q)} = \theta_{[p]}|s_1^2,\ldots,s_k^2)$ by the method indicated in example 1.

The Bayes rule $d^{(2)}$ for the loss function $L_2(\underline{\theta},S)$ is given by (2.4.3).

When the loss function is $L_3(\underline{\theta},S)$, we need to compute $E(\theta_{[k]}|s_1^2,\ldots,s_k^2)$. Let $G_{[k]}(\theta)$ denote the cdf of $\theta_{[k]}$, given $s_1^2,\ldots,s_k^2$. Then

$$G_{[k]}(\theta) = \prod_{i=1}^{k} F_{\gamma_2}(\theta|b_i'',a'')$$

where $F_{\gamma_2}$ stands for the cdf of a gamma-2 population. The function $G_{[k]}(\theta)$ can be evaluated by using either of the following relations given in Raiffa and Schlafer (1961):

$$F_{\gamma_2}(z|b,a) = I(bz\sqrt{\tfrac{1}{2}a}\ ,\ \tfrac{1}{2}a - 1) = G_p(\tfrac{1}{2}a|\tfrac{1}{2}abz)$$

where $I(\cdot,\cdot)$ is the gamma function tabulated by Pearson (1934) and $G_p$ is the cumulative Poisson function.

Now, since $\theta_{[k]}$ is a positive random variable, we have

$$E(\theta_{[k]}|s_1^2,\ldots,s_k^2) = \int_0^\infty [1 - G_{[k]}(\theta)]d$$

Substituting the values of $E(\theta_{[k]}|s_1^2,\ldots,s_k^2)$ and $E(\theta_{(k-m)}|s_1^2,\ldots,s_k^2)$ in the expression for $\Delta_m$, we can compute the Bayes rule $d^{(3)}$ given by (2.4.5).

3. Binomial populations

Let $\pi_1,\ldots,\pi_k$ be k independent binomial populations with probabilities of success $\theta_i,\ldots,\theta_k$, respectively, $0 < \theta_i < 1$. Assume that $\theta_i$ are independent, and have a priori distribution $f_\beta(\cdot|a,b)$ where $f_\beta(\cdot|a,b)$ is the beta pdf given by

$$f_\beta(\theta|a,b) = \frac{1}{B(a,b)}\ \theta^{a-1}(1-\theta)^{b-1},\ a,\ b > 0 \qquad (2.5.11)$$

Let $x_i$ be the number of successes in independent trials from $\pi_i$ $(i = 1,\ldots,k)$. Then the posterior distribution of $\underline{\theta} = (\theta_1,\ldots,\theta_k)$ given $\underline{x} = (x_1,\ldots,x_k)$ is

$$g^*(\underset{\sim}{\theta}|\underset{\sim}{x}) = \prod_{i=1}^{k} f_\beta(\theta_i|x_i + a,\ n + b - x_i)$$

Thus

$$E(\theta_{(i)}|\underset{\sim}{x}) = \frac{x_{[i]} + a}{n + a + b}$$

The Bayes rule $d^{(2)}$ ($d^{(3)}$) for the loss $L_2(\underset{\sim}{\theta},S)$ ($L_3(\underset{\sim}{\theta},S)$) can be obtained as in Example 2 discussed above.

4. Poisson populations

Here $\pi_1,\ldots,\pi_k$ are independent Poisson distributions with parameters $\theta_1,\ldots,\theta_k$, respectively, $\theta_i > 0$, $i = 1,\ldots,k$. Assume all $\theta_j$ are independent, and $\theta_i$ has the natural conjugate a priori distribution $f_{\gamma_1}(\cdot|a',b')$, where

$$f_{\gamma_1}(\theta|a',b') = Be^{-\theta b'}\theta^{a'-1},\ \theta \geq 0,\ a',b' > 0 \qquad (2.5.12)$$

Let $x_i$ be an observation from $\pi_i$. Then the posterior distribution of $\underset{\sim}{\theta} = (\theta_1,\ldots,\theta_k)$ is given by

$$g^*(\underset{\sim}{\theta}|\underset{\sim}{x}) = \prod_{i=1}^{k} (f_{\gamma_1}(\theta_i|x_i + a',\ 1 + b')$$

and

$$E(\theta_{(i)}|\underset{\sim}{x}) = \frac{x_{[i]} + a'}{1 + b'}$$

The Bayes rule $d^{(2)}$ and $d^{(3)}$ can be computed as before.

5. A selection problem in life testing.

Let $\pi_1,\ldots,\pi_k$ be independent exponential populations with scale parameters $\theta_1,\ldots,\theta_k$, respectively. Suppose n items from each of the k populations are put on a life test, and the experiment is continued until the first r failures from $\theta_i$ $(i = 1,\ldots,k)$ are observed. This scheme of sampling is called Type II censored sampling.

Let $x_{i1},\ldots,x_{ir}$ denote the first r naturally ordered observations from $\pi_i$, and set

$$T_i \equiv T_{ir} = \sum_{j=1}^{r} x_{ij} + (n - r)x_{ir}, \quad i = 1,\ldots,k \tag{2.5.13}$$

$T_i$ represents the total accumulated life at the termination of the test on population $\pi_i$. The sample likelihood for $\pi_i$, conditional on $\theta_i$, is

$$\ell(x_{i1},\ldots,x_{ir}|\theta_i) = \frac{n!}{(n - r)!\theta_i^r} e^{-\frac{T_{ir}}{\theta_i}}, \quad i = 1,\ldots,k \tag{2.5.14}$$

Assume that $\underset{\sim}{\theta} = (\theta_1,\ldots,\theta_k)$ has the following a priori distribution:

$$g(\underset{\sim}{\theta}) = \prod_{i=1}^{k} \left[ \frac{e^{-\frac{\mu}{\theta_i}} \left(\frac{\mu}{\theta_i}\right)^{\nu+1}}{\mu\Gamma(\nu)} \right], \quad 0 < \theta_i < \infty, \ \mu, \nu > 0$$

Then the posterior distribution of $\underset{\sim}{\theta} = (\theta_1,\ldots,\theta_k)$ given $\underset{\sim}{T} = (T_1,\ldots,T_k)$ is

$$g^*(\underline{\theta}|\underline{T}) = \prod_{i=1}^{k} \left[ \frac{e^{-\frac{\mu + T_i}{\theta_i}} \left(\frac{\mu + T_i}{\theta_i}\right)^{r+\nu+1}}{(\mu + T_{ir})\,\Gamma(r + \nu)} \right]$$

and

$$E(\theta_{(i)}|\underline{T}) = \frac{\mu + T_{[i]}}{r + \nu - 1}, \quad r + \nu > 1$$

where $T_{[1]} \leq \cdots \leq T_{[k]}$ denote the ordered accumulated life times, and $\theta_{(i)}$ is the parameter associated with $T_{[i]}$.

The Bayes rules $d^{(2)}$ and $d^{(3)}$ can now be computed as in earlier examples.

# CHAPTER III

# ON SOME RULES BASED ON SAMPLE MEDIANS FOR SELECTION OF THE LARGEST NORMAL MEAN

## 3.1 Introduction

Many of the classical statistical procedures which are superior to their competitors under the normal model have one drawback: their behavior is seriously affected if a few gross errors are present in the sample. For example, consider the problem of point estimation of the mean $\theta$ of a normal population. It is well known that the sample mean is a uniformly most powerful unbiased estimate of $\theta$, but is not a very good estimate if some wild observations are present in the sample. Hodges and Lehmann (1963) have proposed a class of estimates which are based on rank test statistics; these estimates are approximately normally distributed for large samples. Gupta and Huang (1974) have investigated selection procedures based on one-sample Hodges-Lehmann estimates of location for the problem of selecting a subset containing the largest t ($1 \leq t < k$) location parameters, when the sample size is large. Gupta and Leong (1977) have discussed a selection rule based on sample medians for double exponentials. In this chapter we consider selection procedures based on sample medians for normal populations. For the problem of selecting a subset containing the largest normal mean, when the means are equally spaced, a rule based on medians is compared to the rule proposed by Gupta (1965), which is based on sample means. It appears from the numerical work done that, as expected, Gupta's rule is superior. This may be due to the fact that the tails of a normal distribution are not very thick, and hence the probability of getting extreme observations is small. In case the underlying distributions have thicker tails, e.g., logistic, double exponential [see Gupta and Leong (1977)], extreme observations are more frequent and have a serious effect on the sample means, but not on the sample medians. In these situations, sample median should be a better estimate of location than the sample mean, as indicated by the fact that for a double exponential population, the sample median has smaller variance than the sample mean, if sample size $n \geq 7$ [see Chu and Hotelling (1955)].

Apart from having a simpler form for its density function, the sample median as an estimate of location has some other advantages over the sample mean. Intuitively any reasonable estimate of location should have a distribution which, in some sense, is centered on the true location parameter value. It is shown by Hodges and Lehmann (1963) that the sample median has a distibution which is symmetric about the true value if the underlying distribution is symmetric,and in case the underlying distribution is not symmetric, the sample median is a median unbiased estimate of location,i.e., the median of its distribution coincides with the true location parameter.

In Section 3.2 we introduce the notations to be used in this chapter. In Section 3.3 we investigate procedures based on sample medians for selection of a subset containing the largest normal mean. In Section 3.4 the problem of selecting a subset containing all normal populations better than a control is discussed, and a rule based on sample medians is proposed. In Section 3.5 the selection rule proposed in Section 3.3 is compared to Gupta's procedure based on sample means [c.f., Gupta (1965)], when the means are equally spaced.

In Section 3.6 selection rules based on medians of large samples are discussed. In Section 3.7 asymptotic relative efficiency (ARE) of the proposed rule with respect to Gupta's rule for selecting a subset containing the largest normal mean is computed, when the means are in slippage configuration. It is also shown that, in case the normal populations are contaminated, the proposed rule based on sample medians improves on the rule based on sample means in terms of ARE.

In Section 3.8 a test of homogeneity based on sample medians is proposed, and a relation between the test and selection rule of Section 3.3 is established. In Section 3.9 the distribution of a statistic useful in some selection and ranking problems is derived.

## 3.2 Preliminaries and Notations

Let $X_1, \ldots, X_{2m+1}$ be independent observations from a population with cumulative distribution function (cdf) $F(x,\theta)$ and probability density function (pdf) $f(x,\theta)$, $x,\theta \in \mathbb{R}$, the real line and $m \geq 1$. Then the sample median $\tilde{X}$ is given by

$$\tilde{X} = X_{[m+1]}$$

where $X_{[1]} \leq \cdots \leq X_{[2m+1]}$ are ordered $X_i$.

The pdf of $\tilde{X}$ is

$$g(x,\theta) = c_m [F(x,\theta)]^m [1-F(x,\theta)]^m f(x,\theta) \tag{3.2.1}$$

where $c_m = \frac{(2m+1)!}{(m!)^2}$

and its cdf is given by

$$G(x,\theta) = c_m \int_{-\infty}^{x} [F(u,\theta)]^m [1-F(u,\theta)]^m f(u,\theta)\, du$$

$$= c_m \int_0^{F(x,\theta)} u^m (1-u)^m du = I_{F(x,\theta)} (m+1,\ m+1) \quad \begin{matrix} 0\leq y\leq 1 \\ p,q>0 \end{matrix} \tag{3.2.2}$$

where $I_y$ $(p,q)$ is the incomplete beta function:

$$I_y (p,q) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \int_0^y u^{p-1} (1-u)^{q-1} du \tag{3.2.3}$$

The following result from Karlin (1968) will be needed:

Lemma 3.2.1: If $f(x,\theta)$ has monotone likelihood ratio (MLR) in x and $\theta$, then $g(x,\theta)$ given by (3.2.1) has MLR in x and $\theta$.

## 3.3 A Procedure Based on Sample Medians for Selecting a Subset Containing the Largest Normal Mean

Let $\pi_1,\ldots, \pi_k$ be k independent normal populations with means $\theta_1,\ldots, \theta_k$ and a common known variance $\sigma^2$. Let $\tilde{X}_i$ be the sample median of $n = 2m+1$ $(m\geq 1)$ observations from $\pi_i$. The pdf g and the cdf G of $\tilde{X}_i$ are obtained from (3.2.1) and (3.2.2) by substituting $F(x,\theta_i) = \Phi(\frac{x-\theta_i}{\sigma})$, $f(x,\theta) = \frac{1}{\sigma}\phi(\frac{x-\theta_i}{\sigma})$, and are given by

$$g(x,\theta_i) = \frac{c_m}{\sigma} [\Phi(\frac{x-\theta}{\sigma})]^m [1-\Phi(\frac{x-\theta}{\sigma})]^m \phi(\frac{x-\theta}{\sigma}) \tag{3.3.1}$$

$$G(x, \theta_i) = I_{\Phi(\frac{x-\theta_i}{\sigma})} \quad (m+1, m+1) \tag{3.3.2}$$

where $\Phi(\cdot)$ is cdf and $\phi(\cdot)$ is pdf of a standard normal random variable.

Let $\theta_{[1]} \leq \cdots \leq \theta_{[k]}$ be the ordered $\theta_i$. For the problem of selecting a subset containing the population associated with $\theta_{[k]}$, consider the following rule $R_1$:

$$R_1 \text{: select } \pi_i \quad \text{iff} \quad \tilde{X}_i \geq \tilde{X}_{[k]} - d_1\sigma \tag{3.3.3}$$

where $\tilde{X}_{[1]} \leq \cdots \leq \tilde{X}_{[k]}$ are ordered sample medians, and $d_1 \geq 0$ is a constant depending on a preassigned $P^*$ $(\frac{1}{k} < P^* < 1)$ such that

$$P(CS \mid R_1) \geq P^* \tag{3.3.4}$$

Let $\tilde{X}_{(i)}$ be the sample median associated with $\theta_{[i]}$ $(i=1,\ldots,k)$. Then

$$P(CS \mid R_1) = P(\tilde{X}_{(k)} \geq \tilde{X}_{[k]} - d_1 \sigma)$$

$$= P(\tilde{X}_{(j)} \leq \tilde{X}_{(k)} + d_1, \ j=1,\ldots,k-1)$$

$$= P(\tilde{X}_{(j)} - \theta_{[j]} \leq \tilde{X}_{(k)} - \theta_{[k]} + \theta_{[k]} - \theta_{[j]} + d_1\sigma, \ j=1,\ldots,k-1)$$

$$= \frac{c_m}{\sigma} \int_{-\infty}^{\infty} \left[ \prod_{j=1}^{k-1} I_{\Phi} \left(u + \frac{\theta_{[k]} - \theta_{[j]}}{\sigma} + d_1\right) \quad (m+1, m+1) \right]$$

$$\cdot \ \Phi^m(u) \ [1-\Phi(u)]^m \ \phi(u) \ du$$

It is clear from the above expression that the infimum of $P(CS \mid R_1)$ occurs when all $\theta_i$ are equal, and hence constant $d_1 = (d_1 \ k,n,P^*)$ is obtained by solving

$$\frac{c_m}{\sigma} \int_{-\infty}^{\infty} [I_{\Phi\,(u+d_1)}\,(m+1,m+1)]^{k-1} \;\Phi^m(u)\; [1-\Phi(u)]^m$$

$$\cdot\phi(u)\; du = P^* \tag{3.3.5}$$

Expected Size of the Selected Subset

The rule $R_1$ defined by (3.3.3) and (3.3.4) selects a random number of populations and it is desirable that the size of the selected subset be small, and also the ranks of the selected subset be large, where the population associated with $\theta_{[i]}$ is given rank i, (i=1,...,k). Gupta (1965) has proposed the expected size of the selected subset and expected sum of ranks of selected populations as criteria of efficiency of selection rules.

Let S and SR be random variables denoting the size of the selected subset and the sum of ranks of the selected populations, respectively. Then following Gupta (1965) we have

$$E_{\underline{\theta}}\,(S|R_1) = \sum_{i=1}^{k} P_{\underline{\theta}}(i) \tag{3.3.6}$$

and

$$E_{\underline{\theta}}\,(SR|R_1) = \sum_{i=1}^{k} i\; P_{\underline{\theta}}(i) \tag{3.3.7}$$

where $P_{\underline{\theta}}$ (i) is the probability that the rule $R_1$ selects the population associated with $\theta_{[i]}$, i=1,...k, and is given by

$$P_{\underline{\theta}}(i) = \frac{c_m}{\sigma} \int_{-\infty}^{\infty} \left[ \prod_{\substack{j=1\\ j\neq i}}^{k} I_{\Phi(u+\frac{\theta_{[i]}-\theta_{[j]}}{\sigma} + d_1)}\,(m+1,m+1) \right]$$

$$\cdot\Phi^m(u)\;\; [1-\Phi(u)]^m\; \phi(u)\; du \tag{3.3.8}$$

Some Properties of $R_1$

(i) Upper Bound on $E_{\underline{\theta}}(S|R_1)$

It follows from Lemma 3.2.1 that $g(x,\theta_i)$ given by (3.3.1) has MLR in x and θ. Hence, from Theorem 1 of Gupta (1965), we have

$$\max_{\underline{\theta}\in\Theta} E_{\underline{\theta}}(S|R_1) = \max_{\theta\in\Theta_0} E_{\underline{\theta}}(S|R_1) = kP^*$$

where $\Theta_0 = \{(\theta_1,\ldots,\theta_k) \in \Theta\colon \theta_1 = \ldots = \theta_k\}$.

(ii) Property of Monotonicity

It is clear from expression (3.3.8) that the rule $R_1$ is strongly monotone [see Santner (1975)], i.e., $P_{\underline{\theta}}(i)$ is non-decreasing (nonincreasing) in $\theta_{[i]}$ $(\theta_{[j]}, j\neq i)$, when other components of $\underline{\theta}$ remain fixed. The following two properties of $R_1$ are immediate consequences of strong monotonicity:

(a) Monotonicity: $P_{\underline{\theta}}(i) \geq P_{\underline{\theta}}(j) \quad \forall\ 1\leq j\leq i\leq k$

(b) Unbiasedness: $P_{\underline{\theta}}(k) \geq P_{\underline{\theta}}(i) \quad \forall\ 1\leq i\leq k$

(iii) Minimaxity with Respect to the Expected Subset Size

A selection rule $R^*$ is said to be minimax with respect to S if

$$\sup_{\Theta} E_{\underline{\theta}}(S|R^*) = \inf_{R} \sup_{\Theta} E_{\underline{\theta}}(S|R)$$

where inf is over all selection rules which satisfy the $P^*$-condition [see Berger (1977)].

The density function $g(x,\theta_i)$ given by (3.3.1) is clearly of location type, and it has MLR in x and $\theta_i$ by Lemma 3.2.1. It follows from Theorem 1.4.2 of Berger (1977) that the selection rule

$R_1$ is minimax with respect to S among all rules based on sample medians.

### 3.4 Selection of a Subset Containing All Normal Populations Better Than a Control

Here we have k+1 independent normal populations $\pi_0, \pi_1, \ldots, \pi_k$, where $\pi_i$ has mean $\theta_i$ and a known variance $\sigma^2$ $(i=1,\ldots,k)$. The population $\pi_0$ is a standard or control population; $\theta_0$ may or may not be known. The population $\pi_i$ is said to be better than control if $\theta_i \geqq \theta_0$. Our interest is in the subset of all populations which are better than control. Gupta (1965) has investigated a rule based on sample means which selects a subset that contains all normal populations better than control with probability at least $P^*$. We propose a rule based on sample medians. Two cases are possible:

A. <u>$\theta_0$ known</u>

We are given medians $\tilde{X}_i$ of $n = 2m + 1$ $(m \geqq 1)$ independent observations from $\pi_i$ $(i=1,\ldots,k)$. Consider the rule $R_A$ defined as follows:

$$R_A\text{: select } \pi_i \text{ iff } \tilde{X}_i \geqq \theta_0 - a\sigma \tag{3.4.1}$$

where a is chosen so as to satisfy the basic $P^*$ condition. Let $k_1$, $k_2$ denote the true (unknown) number of populations with $\theta_i \geqq \theta_0$ and $\theta_i < \theta_0$, respectively. Then $k_1 + k_2 = k$. Also let primes refer to the $k_1$ populations which are better than control. Then

$$P(CS|R_A) = \prod_{i=1}^{k_1} P(\tilde{X}_i' \geqq \theta_0 - a\sigma)$$

$$= \prod_{i=1}^{k_1} \left[ 1 - I_{\Phi\left(\frac{\theta_0-\theta_i'}{\sigma} - a\right)} (m+1,m+1) \right] \tag{3.4.2}$$

It is clear from (3.4.2) that the minimum of $P(CS|R_A)$ under the restriction $\theta_i' \geqq \theta_0$ occurs at $\theta_i' = \theta_0$ , $i=1,\ldots k_1$, and therefore a lower bound on $P(CS|R_A)$ is obtained by setting $\theta_i' = \theta_0$ and $k_1=k$:

$$P(CS|R_A) \geqq [1 - I_{\Phi(-a)} (m+1,m+1)]^k$$

$$= [I_{\Phi(a)} (m+1,m+1)]^k$$

Hence a is obtained from the equation

$$I_{\Phi(a)} (m+1,m+1) = (P^*)^{1/k} \tag{3.4.3}$$

The expected subset size for the rule $R_A$ is given by

$$E(S|R_A) = \sum_{i=1}^{k} P(\pi_i \text{ is selected})$$

$$= \sum_{i=1}^{k} [1-I_{\Phi\left(\frac{\theta_0 - \theta_i}{\sigma} -a\right)} (m+1,m+1)] \tag{3.4.4}$$

B. $\theta_0$ unknown

In this case 2m+1 independent observations are taken from $\pi_0$. Let $X_0$ be the median of this sample. We propose the following rule

$$R_B\text{: select } \pi_i \text{ iff } \tilde{X}_i \geq \tilde{X}_0 - b\sigma \tag{3.4.5}$$

where constant b is chosen to satisfy $P^*$ condition. We have, as in

Case A,

$$P(CS|R_B) = \frac{c_m}{\sigma} \int_{-\infty}^{\infty} \prod_{i=1}^{k_1} [1-I_{\Phi(u+\frac{\theta_0 - \theta_i'}{\sigma} - b)}(m+1,m+1)]$$

$$\cdot \Phi^m(u)[1-\Phi(u)]^m \phi(u)\, du \qquad (3.4.6)$$

$$\geqq \frac{c_m}{\sigma} \int_{-\infty}^{\infty} [1-I_{\Phi(u-b)}(m+1,m+1)]^k$$

$$\cdot \Phi^m(u)[1-\Phi(u)]^m \phi(u)\, du \qquad (3.4.7)$$

The constant b is obtained by equating right hand side of (3.4.7) to $P^*$. The expected subset size for the rule $R_B$ is obtained as in Case A.

Remarks:

(i) It is clear from expressions for P(select $\theta_i$) for rules $R_A$ and $R_B$ that, in either case

$$P(\text{select } \theta_i) \geqq P(\text{select } \theta_j) \qquad \text{if} \qquad \theta_i \geqq \theta_j$$

(ii) If $\theta_i \to \infty \quad \forall_i = 1,\ldots,k$ and $\theta_0$ is finite, then $E(S) \to K$ in each case.

### 3.5 Comparison between $R_1$ and Gupta's Procedure Based on Sample Means when the Normal Means Are Equally Spaced

Let $\pi_1,\ldots,\pi_k$ be k independent normal populations with means $\theta, \theta + \delta\sigma, \ldots, \theta + (k-1)\delta\sigma$ and a common known variance $\sigma^2$; $\delta > 0$ is a known constant. Let $X_{ij}$ $(j=1,\ldots,n)$ be a sample of size $n=2m+1 (m \geqq 1)$ from $\pi_i$ $(i=1,\ldots,k)$, and let $\tilde{X}_i$, $\overline{X}_i$ be the median and the mean of observations from $\pi_i$:

$$\tilde{X}_i = X_{[m+1]} \qquad \text{where } x_{[1]} \leq \cdots \leq x_{[2m+1]}$$

$$\bar{X}_i = \frac{\sum_{j=1}^{2m+1} X_{ij}}{2m+1}$$

For the problem of selecting a subset containing the largest mean $\theta+(k-1)\delta\sigma$, Gupta (1965) has proposed the following rule R:

$$R: \text{select } \pi_i \quad \text{iff} \quad \bar{X}_i \geq \bar{X}_{[k]} - \frac{d\sigma}{\sqrt{2m+1}} \tag{3.5.1}$$

where $d \geq 0$ is given by

$$\int_{-\infty}^{\infty} \Phi^{k-1}(u+d) \ \phi(u) \ du = P^* \tag{3.5.2}$$

We will compare the rule R to the rule $R_1$ defined by (3.3.3) and (3.3.5).

Let $P(i,k,P^*,\delta\ ,n|R')$ denote the probability with which a rule R' selects the population associated with the i-th largest mean $(i=1,\ldots,k)$. Then from Gupta (1965) we have

$$P(i,k,P^*,\delta\ ,n|R) = P(i,k,P^*,\delta\ \sqrt{n}\ |R)$$

$$=\int_{-\infty}^{\infty} [\prod_{\substack{j=1\\ j\neq i}}^{k} \Phi(x+d-(j-i)\delta\sqrt{n})]\phi(x)\ dx] \tag{3.5.3}$$

The expression for $P,(i,k,P^*,\delta\ ,n|R_1)$ is obtained from (3.3.8) by substituting $\theta_{[i]} = \theta+(i-1)\delta\sigma$ , $i=1,\ldots,k$:

$$P(i,k,P^*,\delta,n|R_1)= \frac{c_m}{\sigma} \int_{-\infty}^{\infty} [\prod_{\substack{j=1\\ j\neq 1}}^{k} I_{\Phi(u+(i-j)\delta+d_1)}(m+1,m+1)]$$

$$\cdot\Phi^m(u)[1-\Phi(u)]^m \phi(u)\ du \tag{3.5.4}$$

For the rule $R_1$, the probabilities $P(i,k,P^*,\delta,n|R_1)$ given by (3.5.4) have been computed for k=2(1)5, n=3,5, $\delta$=0.5 (0.5) 5.0 and $P^*$= .90,.95. The numerical integration was done by the Gauss-Hermite integration formula. Tables are given at the end of this chapter. Tables of $P(i,k,P^*,\delta,n|R)$ are available in Gupta (1965).

Next, let $\Psi(k,P^*,\delta,n|R')$ and $\Psi_1(k,P^*,\delta,n|R')$ denote the expected sum of ranks and expected average rank of the selected populations for a rule R', respectively. Then

$$\Psi(k,P^*,\delta\ ,n|R') = \sum_{i=1}^{k} i\ P(i,k,P^*,\delta,n|R') =k\Psi_1(k,P^*,\delta,n|R') \tag{3.5.5}$$

For the rule R, tables of values of $\Psi_1$ and the expected proportion of the population retained in the subset $(= \frac{1}{k} E(S|R))$ are available in Gupta (1965). We have computed the values of these functions for the rule $R_1$, for values of k,n,$\delta$ and $P^*$ mentioned above. For instance if $P^*$=. 90, k=5, n=3, $\delta=1.5/\sqrt{3}$, then the rule R based on sample means selects the second best and third best populations with probabilities .781 and .357, respectively. The corresponding probabilities for the rule $R_1$ are .822 and .467, in that order. The probability of correct selection (selecting the best) has to be greater than .90 for both the rules and is actually equal to .998 for the rule R, and .997 for the rule $R_1$.

It appears from these tables that the rule R based on sample means is superior to the rule $R_1$ based on sample medians, and that, as expected , the performance of R relative to $R_1$ improves as sample size is increased.

Remarks:

(i) It is shown in Gupta (1965) that the rule R has the following desirable properties:

(1) For fixed $P^*$ and k

$P(1,k,P^*,\delta,n) \downarrow \ \delta\sqrt{n}$

and

$P(k,k,P^*,\delta,n) \uparrow \ \delta\sqrt{n}$

(2) For fixed $P^*,i,\delta$ and n

$P(i,k,P^*,\delta,n) \downarrow \ k, \quad 1 \leqq i \leqq k$

(3) For fixed $k,P^*$ and $(i-j)\delta$

$\lim_{n\to\infty} \Psi\,(P^*,k,\delta,n) = k$

It is clear from expression (3.5.4) that the rule $R_1$ has similar properties.

(ii) It follows from (3.5.4) and (3.5.5) that

$$\text{(A)}\quad \Psi(P^*,k,\delta,n|R_1) > \frac{c_m}{\sigma}\sum_{i=1}^{k} i\int_{-\infty}^{\infty}\Big[\prod_{j=2}^{k} I_{\Phi(x+d_1-(j-1)\delta)}(m+1,m+1)\Big]$$

$$\cdot\,\Phi^m(x)\,[1-\Phi(x)]^m\,\phi(x)\,dx$$

$$(B)\ \Psi(P^*, k, \delta, n | R_1) > \frac{k(k+1)c_m}{2\sigma}$$

$$\cdot \int_{-\infty}^{\infty} [I_{\Phi(x+d_1-(k-1)\delta)}\ (m+1, m+1)]^{k-1}$$

$$\cdot \Phi^m(x)[1-\Phi(x)]^m\ \phi(x)\ dx$$

## 3.6 A Selection Rule Based on Medians of Large Samples

Let $f(x)$ be the pdf of a continuous random variable $X$ and let $\xi$ be its unique median. Let $\tilde{X}$ be the sample median of $n=2m+1$ observations on $X$. The distribution of $\tilde{X}$ under certain conditions on $f(x)$ is known [see Cramer (1946)] to be asymptotically normal with mean $\xi$ and variance

$$\frac{1}{4[f(\xi)]^2\ (2m+1)} .$$

The above result will be used to investigate a rule for selecting a subset containing the largest normal mean.

Let $\tilde{X}_i$ be the sample median of $(2m+1)$ independent observations from $\pi_i$ $(i=1,\ldots,k)$, where $\pi_i$ is normal with mean $\theta_i$ and a known variance $\sigma^2$. Then, for large $n$, $\tilde{X}_i$ is asymptotically normal with mean $\theta_i$ and variance

$$\frac{\pi\sigma^2}{2(2m+1)} .$$

For the problem of selecting a subset containing $\theta_{[k]}$, we propose the following rule:

$$R_2\text{: select } \pi_i \quad \text{iff} \quad \tilde{X}_i \geq \tilde{X}_{[k]} - d_2\sigma \frac{\sqrt{\pi}}{\sqrt{2(2m+1)}} \qquad (3.6.1)$$

where $d_2 \geq 0$ is to be determined from the basic $P^*$ condition.

We have

$$P(CS|R_2) = P(\tilde{X}_{(k)} \geq \tilde{X}_{[k]} - d_2 \frac{\sigma\sqrt{\pi}}{\sqrt{2(2m+1)}})$$

$$= \int_{-\infty}^{\infty} [\prod_{i=1}^{k-1} \Phi(u+ \frac{\theta_{[k]}-\theta_{[i]}}{\frac{\sigma\sqrt{\pi}}{\sqrt{2(2m+1)}}} + d_2)]\ \phi(u)\ du \tag{3.6.2}$$

It follows from (3.6.2) that the equation for $d_2$ is

$$\int_{-\infty}^{\infty} \Phi^{k-1}(u+d_2)\ \phi(u)\ du = P^* \tag{3.6.3}$$

Tables of values of $d_2$ satisfying (3.6.3) for selected values of $P^*$ are available in Bechhofer (1954) for k = 1(1)10 and in Gupta (1956) for k = 1(1)50.

Expected Size of the Selected Subset

$$E(S|R_2) = \sum_{i=1}^{k} P(\text{population associated with } \theta_{[i]} \text{ is selected})$$

$$= \sum_{i=1}^{k} \int_{-\infty}^{\infty} \left[\prod_{\substack{j=1 \\ j\neq i}}^{k} \Phi(u+ \frac{\theta_{[i]}-\theta_{[j]}}{\frac{\sigma\sqrt{\pi}}{\sqrt{2(2m+1)}}} + d_2)\right] \phi(u)\ du \tag{3.6.4}$$

3.7 Asymptotic Relative Efficiency (ARE) of the Rule $R_2$ Relative to Gupta's Rule R

(i) Normal Populations

Let $\pi_1, \ldots, \pi_k$ be normal populations with means $\theta_1, \ldots, \theta_k$, respectively, and a common known variance $\sigma^2$. Assume that $\theta_i (i=1,\ldots,k)$ are in following slippage configuration:

$$\theta_i = \begin{cases} \theta+\sigma\Delta & \text{if } i=i_o;\ \Delta>0 \text{ is unknown} \\ \theta & \text{if } i\neq i_o \end{cases}$$

The population $\pi_{i_o}$ has the largest mean and is called the 'best' population; the index $i_o$ is not known.

Our interest is in the relative performance of the following two selection rules:

$$R_2\text{: select } \pi_i \text{ iff } \tilde{X}_i \geq \max_{1\leq j\leq k} \tilde{X}_j - \frac{d_2\sigma\sqrt{\pi}}{\sqrt{2n}}$$

$$R\text{ : select } \pi_i \text{ iff } \overline{X}_i \geq \max_{1\leq j\leq k} \overline{X}_j - \frac{d\sigma}{\sqrt{n}}$$

where $\tilde{X}_i$ is the median and $\overline{X}_i$ the mean of $n=2m+1$ observations from $\pi_i$, $i=i,\ldots,k$, $m\geq1$. The constants d and $d_2$ are given by equations (3.5.2) and (3.6.3), respectively. It is obvious that

$$d_2=d \tag{3.7.1}$$

Let $S^*$ be the number of non-best populations in the selected subset. Then small values of $S^*$ are desirable, and therefore, consistent with the basic $P^*$-condition, we would like to keep the expected value of $S^*$ as small as possible.

It is intuitively clear that the performance of any reasonable selection rule R' should improve as the sample size is increased. For a given $0<\varepsilon<1$, let $N_{R'}(\varepsilon)$ be the number of observations needed so that

$$E(S^*|R') = \varepsilon$$

We will use the following definition of ARE [see McDonald (1969)]:

Definition 3.7.1: The ARE of the rule $R_2$ relative to the rule R is defined to be

$$\text{ARE}(R_2,R;\theta) = \lim_{\varepsilon \downarrow 0} \frac{N_R(\varepsilon)}{N_{R_2}(\varepsilon)} \tag{3.7.2}$$

We have

$$E(S^*|R_2) = \sum_{i=1}^{k-1} P(\tilde{X}_{(i)} \geqq \tilde{X}_{[k]} - d_2 \frac{\sigma\sqrt{\pi}}{\sqrt{2n}})$$

$$= \sum_{i=1}^{k-1} P(\tilde{X}_{(j)} \leqq \tilde{X}_{(i)} + d_2 \frac{\sigma\sqrt{\pi}}{\sqrt{2n}};\ j=1,\ldots,k,\ j\neq i)$$

$$=(k-1)\int_{-\infty}^{\infty} \Phi(u - \frac{\Delta}{\frac{\sqrt{\pi}}{\sqrt{2n}}} + d_2)\ \Phi^{k-2}(u+d_2)\ \phi(u)du$$

By definition of $N_{R_2}(\varepsilon)$ we have

$$(k-1)\int_{-\infty}^{\infty} \Phi(u - \frac{\Delta}{\frac{\sqrt{\pi}}{\sqrt{2N_{R_2}(\varepsilon)}}} + d_2)\ \Phi^{k-2}(u+d_2)\ \phi(u)\ du = \varepsilon \tag{3.7.3}$$

Similarly

$$E(S^*|R) = (k-1)\int_{-\infty}^{\infty} \Phi(u - \frac{\Delta}{\sqrt{1/n}} + d)\ \Phi^{k-2}(u+d)\ \phi(u)\ du$$

and therefore

$$(k-1) \int_{-\infty}^{\infty} \Phi(u- \frac{\Delta}{\sqrt{1/N_R(\varepsilon)}} +d) \; \Phi^{k-2} (u+d) \; \phi(u) \, du = \varepsilon \tag{3.7.4}$$

Since $d_2=d$, (3.7.3) and (3.7.4) lead to

$$\int_{-\infty}^{\infty} [ \Phi(u - \frac{\Delta}{\sqrt{\pi/2N_{R_2}(\varepsilon)}} + d) - \Phi(u - \frac{\Delta}{\sqrt{1/N_R(\varepsilon)}} + d)] \cdot \Phi^{k-2} (u+d) \; \phi(u) \, du = 0 \tag{3.7.5}$$

Using the fact that $\Phi$ is strictly increasing, it can be seen from (3.7.5) that

$$\frac{\pi}{2N_{R_2}(\varepsilon)} = \frac{1}{N_R(\varepsilon)}$$

and hence

$$\text{ARE } (R_2,R) = \lim_{\varepsilon \downarrow 0} N_R(\varepsilon)/N_{R_2}(\varepsilon)$$

$$= \frac{2}{\pi} = .64 \tag{3.7.6}$$

(ii) <u>Contaminated Normal Populations</u>

Suppose in the course of sampling from population $\pi_i$ ($i=1,\ldots,k$) something happens to the system and gives rise to wild observations. In these situations the pdf of $\pi_i$ can be written as

$$f(x,\theta_i) = \alpha f_1(x,\theta_i) + (1-\alpha) \; f_2(x,\theta_i), \quad 0<\alpha<1 \tag{3.7.7}$$

This means that the experimenter is sampling from a population with pdf $f_1(x,\theta_i)$ 100 $\alpha$ percent of the time, and from $f_2(x,\theta_i)$ with 100 $(1-\alpha)$ percent of the time $(i=1,\ldots,k)$. The presence of observations from $f_2(x,\theta_i)$ is termed as <u>contamination</u>. For our discussion we will assume that

$$f_1(x,\theta_i) = \frac{1}{\sigma}\ \phi\left(\frac{x-\theta_i}{\sigma}\right)$$

$$f_2(x,\theta_i) = \frac{1}{\sqrt{b}\ \sigma}\ \phi\left(\frac{x-\theta_i}{\sqrt{b}\ \sigma}\right) \qquad (i=1,\ldots,k)$$

where $\theta_i = \begin{cases} \theta + \Delta\sigma & \text{if } i=i_o \ (i_o \text{ is unknown}), \Delta > 0 \ \text{unknown} \\ \theta & \text{if } i \neq i_o \end{cases}$

Given $n$ independent observations from $\pi_i$, let $\tilde{X}_i$ be the median and $\overline{X}_i$ the mean from $\pi_i$, $i=1,\ldots,k$, where $n \geqq 3$ is an odd integer. It is known [see Rohtagi (1976)] that $\tilde{X}_i$ and $\overline{X}_i$ both are asymptotically normal, each with mean $\theta_i$ and variances $\tilde{\sigma}^2$ and $\bar{\sigma}^2$, respectively, where

$$\tilde{\sigma}^2 = \frac{\pi\sigma^2}{2n}\ \frac{1}{\{\alpha + \frac{1-\alpha}{\sqrt{b}}\}^2} \tag{3.7.8}$$

$$\bar{\sigma}^2 = \frac{\sigma^2}{n}\ [\alpha + (1-\alpha)b] \tag{3.7.9}$$

For the problem of selecting a subset containing the (unknown) best population $\pi_{i_o}$, consider the following two rules:

$$R_2^* : \text{select } \pi_i \ \text{ iff } \ \tilde{X}_i \geqq \tilde{X}_{[k]} - d_2^*\ \tilde{\sigma}$$

$$R^* : \text{select } \pi_i \ \text{ iff } \ \overline{X}_i \geqq \overline{X}_{[k]} - d^*\ \bar{\sigma}$$

It is easy to see that $d_2^*$ and $d^*$ both satisfy the equation (3.5.2) and hence

$$d_2^* = d^* = d, \text{ say.}$$

Then

$$E(S^*|R_2^*) = (k-1)\int_{-\infty}^{\infty} \Phi(u - \frac{\Delta\sigma}{\tilde{\sigma}} + d)\, \Phi^{k-2}(u+d)\, \phi(u)\, du \tag{3.7.10}$$

and

$$E(S^*|R^*) = (k-1)\int_{-\infty}^{\infty} \Phi(u - \frac{\Delta\sigma}{\bar{\sigma}} + d)\, \Phi^{k-2}(u+d)\, \phi(u)\, du \tag{3.7.11}$$

Equating the right hand sides of (3.7.10) and (3.7.11) to $\varepsilon$ and solving for $N_{R_2^*}(\varepsilon)$ and $N_{R^*}(\varepsilon)$ we get

$$\frac{\pi\sigma^2}{2N_{R_2^*}(\varepsilon)} \quad \frac{1}{[\alpha + \frac{1-\alpha}{\sqrt{b}}]^2} = \frac{\sigma^2}{N_{R^*}(\varepsilon)} \quad [\alpha + (1-\alpha)b]$$

or

$$\frac{N_{R^*}(\varepsilon)}{N_{R_2^*}(\varepsilon)} = \frac{2}{\pi} \quad [\alpha + (1-\alpha)\, b]\ [\alpha + \frac{1-\alpha}{\sqrt{b}}]^2$$

$$\to \infty \quad \text{as} \quad b \to \infty .$$

This shows that for large values of b the rule $R_2^*$ based on medians is much better than the rule $R^*$ based on sample means. In fact, it can be seen from a result in Rohtagi (1976) that the ARE $(R_2^*, R)$ is close to 1 when b=9 and $\alpha$=.915, and as the differences b-9>0 or $.915 - \alpha > 0$ increase the rule $R_2^*$ shows a significant improvement over $R^*$ in terms of the ARE.

## 3.8 A Test of Homogeneity Based on $\tilde{X}_{[k]} - \tilde{X}_{[1]}$

Let $\pi_1, \ldots, \pi_k$ be k independent normal populations with means $\theta_1, \ldots, \theta_k$, respectively, and a common known variance $\sigma^2$. As before, let $\tilde{X}_i$ denote the sample medians of $n=2m+1$ $(m \geqq 1)$ observations from $\pi_i$ $(i=1,\ldots,k)$, and $\tilde{X}_{[1]},\ldots,\tilde{X}_{[k]}$ be their ordered values. For the hypothesis

$$H_o : \theta_1 = \ldots = \theta_k$$

consider the following test:

$$\text{Reject } H_o \text{ if } R = \tilde{X}_{[k]} - \tilde{X}_{[1]} > \gamma \tag{3.8.1}$$

We find the constant $\gamma$ so that

$$P_{H_o} ( R > \gamma ) \leqq \alpha$$

where $\alpha$ is the size of the test.

The following result gives the constant $\gamma$, and also establishes a relationship between the test given by (3.8.1) and the selection rule $R_1$ of Section 3.3.

Theorem 3.8.1

For $0<\alpha<1$, let $\gamma$ satisfy

$$P_{H_o} (\tilde{X}_k \geqq \tilde{X}_{[k]} - \gamma) \geqq 1 - \frac{\alpha}{k}$$

Then

$$P_{H_o} (R > \gamma) \leqq \alpha .$$

Proof: The proof is similar to that of Theorem 6.1 of Gupta and Leong (1977), and hence omitted.

3.9 On the Distribution of the Statistic Associated with $R_1$ when the Underlying Distributions Are Normal

Let $\tilde{X}_i$ $(i=0,1,\ldots,k)$ be k+1 sample medians of $n=2m+1$ $(m\geq 1)$ independent observations from a standard normal population. Define

$$Z_i = \tilde{X}_i - \tilde{X}_o \qquad (i=1,\ldots,k)$$

The r.v.'s $Z_i$ are correlated and the distribution of $Z = \max_{1\leq i\leq k} Z_i$ is needed in some ranking and selection problems. For standard double exponential population the distribution of Z has been computed by Gupta and Leong (1977) for selected values of k,n, and $\alpha$. In this section we give an expression for the distribution of Z, and also provide a short table for its upper percentage points for $P^*=\alpha=.75,.85,.90,.95,.99$, $k=2(1)5$, $n=3(2)11$.

Let $F(\cdot)$ be the cdf of Z. Then $F(z)=P(Z\leq z)=P(\tilde{X}_i\leq\tilde{X}_o+z, i=1,\ldots,k)$

It is easy to see that

$$F(z)= c_m \int_{-\infty}^{\infty} [I_{\Phi(z+x)}(m+1,m+1)]^k\ \Phi^m(x)\ [1-\Phi(x)]^m\ \phi(x)\ dx \qquad (3.9.1)$$

Computations for upper percentage points of F were done using Gauss-Hermite quadrature based on 20 nodes for the numerical integration.

## TABLE IA

For the rule $R_1$ and the configuration $(\theta, \theta + \delta\sigma, \ldots, \theta + (k-1)\delta\sigma)$ this table gives the probability of selecting the normal population with rank $i$ when the population with mean $\theta + (i-1)\delta\sigma$ has rank $i$, $i=1,2,\ldots,k$; the common variance $\sigma^2$ is assumed to be known.

$P^* = .90,\ n = 3$

| k | i | $\delta\sqrt{n}$ .5 | 1.0 | 1.5 | 2.0 | 2.5 | 3.0 | 3.5 | 4.0 | 4.5 | 5.0 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 1 | .836 | .749 | .643 | .524 | .404 | .291 | .196 | .123 | .072 | .039 |
| | 2 | .944 | .971 | .986 | .994 | .997 | .999 | 1.000 | 1.000 | 1.000 | 1.000 |
| 3 | 1 | .778 | .585 | .359 | .171 | .061 | .016 | .003 | .000 | .000 | .000 |
| | 2 | .882 | .828 | .745 | .640 | .521 | .400 | .288 | .194 | .121 | .070 |
| | 3 | .959 | .983 | .993 | .997 | .999 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 4 | 1 | .704 | .381 | .118 | .019 | .001 | .000 | .000 | .000 | .000 | .000 |
| | 2 | .818 | .648 | .423 | .217 | .084 | .024 | .005 | .001 | .000 | .000 |
| | 3 | .907 | .865 | .793 | .697 | .583 | .462 | .344 | .240 | .156 | .094 |
| | 4 | .969 | .989 | .996 | .998 | .999 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 5 | 1 | .612 | .193 | .019 | .001 | .000 | .000 | .000 | .000 | .000 | .000 |
| | 2 | .740 | .425 | .142 | .024 | .002 | .000 | .000 | .000 | .000 | .000 |
| | 3 | .845 | .688 | .467 | .251 | .102 | .031 | .007 | .001 | .000 | .000 |
| | 4 | .923 | .887 | .822 | .733 | .624 | .504 | .384 | .274 | .183 | .113 |
| | 5 | .975 | .992 | .997 | .999 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |

## TABLE IIA

For the rule $R_1$ and the configuration $(\theta, \theta + \quad, \ldots, \theta + (k-1) \quad)$ this table gives the expected average rank of the selected subset (top) and the expected proportion of the populations selected in the subset (bottom) when the normal population with mean $\theta + (i-1)$ has rank $i$, $i=1,2,\ldots,k$; the common variance is assumed to be known.

$P^* = .90,\ n = 3$

| k | .5 | 1.0 | 1.5 | 2.0 | 2.5 | 3.0 | 3.5 | 4.0 | 4.5 | 5.0 |
|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 1.361 | 1.345 | 1.307 | 1.256 | 1.199 | 1.145 | 1.098 | 1.061 | 1.036 | 1.019 |
| | .890 | .860 | .814 | .759 | .701 | .645 | .598 | .562 | .536 | .519 |
| 3 | 1.806 | 1.730 | 1.610 | 1.481 | 1.367 | 1.272 | 1.193 | 1.129 | 1.081 | 1.047 |
| | .873 | .799 | .699 | .603 | .527 | .472 | .430 | .398 | .374 | .357 |
| 4 | 2.234 | 2.057 | 1.831 | 1.634 | 1.479 | 1.358 | 1.261 | 1.180 | 1.117 | 1.071 |
| | .849 | .721 | .582 | .483 | .417 | .372 | .337 | .310 | .289 | .274 |
| 5 | 2.639 | 2.323 | 1.995 | 1.745 | 1.561 | 1.422 | 1.311 | 1.220 | 1.146 | 1.091 |
| | .819 | .637 | .489 | .401 | .346 | .307 | .278 | .255 | .237 | .223 |

## TABLE IB

For the rule $R_1$ and the configuration $(\theta, \theta + \delta\sigma, \ldots, \theta + (k-1)\delta\sigma)$ this table gives the probability of selecting the normal population with rank i when the population with mean $\theta + (i-1)\delta\sigma$ has rank i, $i=1,2,\ldots,k$; the common variance $\sigma^2$ is assumed to be known.

$P^* = .95$, $n = 3$

| k | $\delta\sqrt{n}$ / i | .5 | 1.0 | 1.5 | 2.0 | 2.5 | 3.0 | 3.5 | 4.0 | 4.5 | 5.0 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 1 | .910 | .850 | .768 | .665 | .548 | .427 | .312 | .213 | .136 | .080 |
| | 2 | .974 | .988 | .995 | .998 | .999 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 3 | 1 | .871 | .718 | .500 | .277 | .118 | .037 | .009 | .001 | .000 | .000 |
| | 2 | .938 | .902 | .842 | .758 | .653 | .535 | .414 | .301 | .204 | .129 |
| | 3 | .982 | .993 | .998 | .999 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 4 | 1 | .784 | .477 | .173 | .033 | .003 | .000 | .000 | .000 | .000 | .000 |
| | 2 | .875 | .732 | .517 | .292 | .126 | .041 | .010 | .002 | .000 | .000 |
| | 3 | .940 | .909 | .851 | .770 | .668 | .551 | .430 | .315 | .216 | .138 |
| | 4 | .982 | .994 | .998 | .999 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 5 | 1 | .742 | .306 | .043 | .002 | .000 | .000 | .000 | .000 | .000 | .000 |
| | 2 | .842 | .565 | .235 | .052 | .006 | .000 | .000 | .000 | .000 | .000 |
| | 3 | .914 | .798 | .602 | .369 | .176 | .063 | .016 | .003 | .000 | .000 |
| | 4 | .961 | .938 | .894 | .828 | .739 | .631 | .512 | .391 | .281 | .188 |
| | 5 | .990 | .997 | .999 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |

## TABLE IIB

For the rule $R_1$ and the configuration $(\theta, \theta + \delta\sigma, \ldots, \theta + (k-1)\delta\sigma)$ this table gives the expected average rank of the selected subset (top) and the expected proportion of the populations selected in the subset (bottom) when the normal population with mean $\theta + (i-1)\delta\sigma$ has rank i, $i=1,2,\ldots,k$; the common variance $\sigma^2$ is assumed to be known.

$P^* = .95$, $n = 3$

| $\delta\sqrt{n}$ / k | .5 | 1.0 | 1.5 | 2.0 | 2.5 | 3.0 | 3.5 | 4.0 | 4.5 | 5.0 |
|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 1.429 | 1.413 | 1.379 | 1.330 | 1.273 | 1.213 | 1.156 | 1.107 | 1.068 | 1.040 |
| | .942 | .919 | .881 | .831 | .774 | .713 | .656 | .607 | .568 | .540 |
| 3 | 1.898 | 1.834 | 1.725 | 1.597 | 1.474 | 1.369 | 1.279 | 1.201 | 1.136 | 1.086 |
| | .930 | .871 | .780 | .678 | .590 | .524 | .474 | .434 | .401 | .376 |
| 4 | 2.321 | 2.161 | 1.938 | 1.731 | 1.565 | 1.434 | 1.327 | 1.237 | 1.162 | 1.103 |
| | .895 | .778 | .635 | .523 | .449 | .398 | .360 | .329 | .304 | .284 |
| 5 | 2.792 | 2.514 | 2.178 | 1.904 | 1.699 | 1.543 | 1.419 | 1.315 | 1.225 | 1.150 |
| | .890 | .721 | .555 | .450 | .384 | .339 | .306 | .279 | .256 | .238 |

## TABLE IC

For the rule $R_1$ and the configuration $(\theta, \theta + \delta\sigma, \ldots, \theta + (k-1)\delta\sigma)$ this table gives the probability of selecting the normal population with rank $i$ when the population with mean $\theta + (i-1)\delta\sigma$ has rank $i$, $i=1,2,\ldots,k$; the common variance $\sigma^2$ is assumed to be known.

$P^* = .90$, $n = 5$

| $\delta\sqrt{n}$ | | .5 | 1.0 | 1.5 | 2.0 | 2.5 | 3.0 | 3.5 | 4.0 | 4.5 | 5.0 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| k | i | | | | | | | | | | |
| 2 | 1 | .838 | .754 | .653 | .539 | .422 | .312 | .216 | .139 | .084 | .047 |
| | 2 | .943 | .969 | .985 | .993 | .997 | .999 | 1.000 | 1.000 | 1.000 | 1.000 |
| 3 | 1 | .782 | .596 | .379 | .191 | .073 | .021 | .005 | .001 | .000 | .000 |
| | 2 | .882 | .832 | .753 | .652 | .538 | .422 | .311 | .215 | .139 | .084 |
| | 3 | .958 | .982 | .993 | .997 | .999 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 4 | 1 | .710 | .400 | .134 | .024 | .002 | .000 | .000 | .000 | .000 | .000 |
| | 2 | .822 | .657 | .442 | .239 | .098 | .031 | .007 | .001 | .000 | .000 |
| | 3 | .907 | .868 | .799 | .708 | .599 | .483 | .368 | .264 | .177 | .110 |
| | 4 | .967 | .988 | .995 | .998 | .999 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 5 | 1 | .620 | .214 | .025 | .001 | .000 | 000 | .000 | .000 | .000 | .000 |
| | 2 | .746 | .444 | .161 | .031 | .003 | .000 | .000 | .000 | .000 | .000 |
| | 3 | .848 | .698 | .486 | .274 | .119 | .039 | .010 | .002 | .000 | .000 |
| | 4 | .922 | .890 | .828 | .743 | .639 | .525 | .408 | .299 | .205 | .131 |
| | 5 | .974 | .991 | .996 | .999 | .999 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |

## TABLE IIC

For the rule $R_1$ and the configuration $(\theta, \theta + \delta\sigma, \ldots, \theta + (k-1)\delta\sigma)$ this table gives the expected average rank of the selected subset (top) and the expected proportion of the populations selected in the subset (bottom) when the normal population with mean $\theta + (i-1)\delta\sigma$ has rank $i$, $i=1,2,\ldots,k$; the common variance $\sigma^2$ is assumed to be known.

$P^* = .90$, $n = 5$

| $\delta\sqrt{n}$ | .5 | 1.0 | 1.5 | 2.0 | 2.5 | 3.0 | 3.5 | 4.0 | 4.5 | 5.0 |
|---|---|---|---|---|---|---|---|---|---|---|
| k | | | | | | | | | | |
| 2 | 1.361 | 1.346 | 1.311 | 1.262 | 1.208 | 1.155 | 1.107 | 1.069 | 1.042 | 1.023 |
| | .890 | .862 | .819 | .766 | .710 | .655 | .608 | .569 | .542 | .523 |
| 3 | 1.807 | 1.735 | 1.621 | 1.495 | 1.382 | 1.288 | 1.209 | 1.144 | 1.093 | 1.056 |
| | .874 | .803 | .708 | .613 | .537 | .481 | .439 | .405 | .380 | .361 |
| 4 | 2.236 | 2.068 | 1.850 | 1.654 | 1.499 | 1.378 | 1.280 | 1.198 | 1.132 | 1.083 |
| | .851 | .728 | .593 | .492 | .425 | .378 | .344 | .316 | .294 | .278 |
| 5 | 2.643 | 2.342 | 2.019 | 1.770 | 1.583 | 1.443 | 1.332 | 1.240 | 1.164 | 1.105 |
| | .822 | .647 | .499 | .409 | .352 | .313 | .284 | .260 | .241 | .226 |

TABLE ID

For the rule $R_1$ and the configuration $(\theta, \theta + \delta\sigma, \ldots, \theta + (k-1)\delta\sigma)$ this table gives the probability of selecting the normal population with rank $i$ when the population with mean $\theta + (i-1)\delta\sigma$ has rank $i$, $i=1,2,\ldots,k$; the common variance $\sigma^2$ is assumed to be known.

$P^* = .95$, $n = 5$

| $\delta\sqrt{n}$ | | .5 | 1.0 | 1.5 | 2.0 | 2.5 | 3.0 | 3.5 | 4.0 | 4.5 | 5.0 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| k | i | | | | | | | | | | |
| 2 | 1 | .912 | .854 | .776 | .678 | .566 | .449 | .337 | .236 | .155 | .095 |
| | 2 | .974 | .987 | .994 | .998 | .999 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 3 | 1 | .875 | .729 | .521 | .304 | .137 | .047 | .012 | .002 | .000 | .000 |
| | 2 | .939 | .905 | .848 | .769 | .670 | .557 | .441 | .328 | .230 | .150 |
| | 3 | .981 | .993 | .997 | .999 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 4 | 1 | .824 | .543 | .230 | .053 | .006 | .000 | .000 | .000 | .000 | .000 |
| | 2 | .901 | .778 | .582 | .359 | .174 | .064 | .018 | .004 | .001 | .000 |
| | 3 | .953 | .928 | .881 | .811 | .721 | .615 | .499 | .383 | .277 | .187 |
| | 4 | .986 | .995 | .998 | .999 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 5 | 1 | .753 | .335 | .054 | .003 | .000 | .000 | .000 | .000 | .000 | .000 |
| | 2 | .849 | .585 | .264 | .065 | .009 | .001 | .000 | .000 | .000 | .000 |
| | 3 | .917 | .809 | .622 | .399 | .202 | .078 | .023 | .005 | .001 | .000 |
| | 4 | .962 | .941 | .899 | .837 | .754 | .652 | .539 | .422 | .311 | .215 |
| | 5 | .989 | .997 | .999 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |

TABLE IID

For the rule $R_1$ and the configuration $(\theta, \theta + \delta\sigma, \ldots, \theta + (k-1)\delta\sigma)$ this table gives the expected average rank of the selected subset (top) and the expected proportion of the populations selected in the subset (bottom) when the normal population with mean $\theta + (i-1)\delta\sigma$ has rank $i$, $i=1,2,\ldots,k$; the common variance $\sigma^2$ is assumed to be known.

$P^* = .95$, $n = 5$

| $\delta\sqrt{n}$ | .5 | 1.0 | 1.5 | 2.0 | 2.5 | 3.0 | 3.5 | 4.0 | 4.5 | 5.0 |
|---|---|---|---|---|---|---|---|---|---|---|
| k | | | | | | | | | | |
| 2 | 1.429 | 1.414 | 1.382 | 1.336 | 1.282 | 1.224 | 1.168 | 1.118 | 1.078 | 1.047 |
| | .943 | .921 | .885 | .838 | .783 | .724 | .668 | .618 | .578 | .547 |
| 3 | 1.899 | 1.839 | 1.736 | 1.613 | 1.492 | 1.387 | 1.298 | 1.220 | 1.153 | 1.100 |
| | .932 | .876 | .789 | .690 | .602 | .535 | .484 | .444 | .410 | .383 |
| 4 | 2.357 | 2.216 | 2.007 | 1.801 | 1.629 | 1.493 | 1.383 | 1.289 | 1.208 | 1.140 |
| | .916 | .811 | .673 | .556 | .475 | .420 | .379 | .347 | .319 | .297 |
| 5 | 2.799 | 2.536 | 2.208 | 1.935 | 1.728 | 1.569 | 1.445 | 1.340 | 1.250 | 1.172 |
| | .894 | .733 | .568 | .461 | .393 | .346 | .312 | .285 | .262 | .243 |

TABLE III

Upper $100(1 - P^*)$ percentage points of $Z = \max_{1 \leq i \leq k} ( X_i - X_0 )$ where $X_0, X_1, \ldots, X_k$ are iid sample median random variables in samples of size $n = 2m+1$ ( $m \geq 1$ ) from the standard normal distribution.

| k \ n | 3 | 5 | 7 | 9 | 11 |
|---|---|---|---|---|---|
| | .638 | .511 | .445 | .409 | .393 |
| | .980 | .784 | .676 | .614 | .582 |
| 1 | 1.213 | .969 | .832 | .751 | .710 |
| | 1.558 | 1.245 | 1.065 | .956 | .900 |
| | 2.208 | 1.766 | 1.522 | 1.398 | 1.491 |
| | .959 | .768 | .667 | .610 | .579 |
| | 1.276 | 1.019 | .876 | .792 | .744 |
| 2 | 1.493 | 1.192 | 1.019 | .915 | .855 |
| | 1.816 | 1.452 | 1.239 | 1.105 | 1.030 |
| | 2.429 | 1.943 | 1.676 | 1.533 | 1.606 |
| | 1.125 | .901 | .783 | .715 | .675 |
| | 1.432 | 1.142 | .980 | .884 | .828 |
| 3 | 1.642 | 1.310 | 1.117 | 1.000 | .931 |
| | 1.854 | 1.563 | 1.333 | 1.184 | 1.099 |
| | 2.551 | 2.040 | 1.761 | 1.609 | 1.671 |
| | 1.235 | .989 | .859 | .784 | .738 |
| | 1.536 | 1.223 | 1.048 | .945 | .883 |
| 4 | 1.742 | 1.389 | 1.182 | 1.057 | .982 |
| | 2.049 | 1.639 | 1.396 | 1.238 | 1.145 |
| | 2.634 | 2.106 | 1.819 | 1.661 | 1.715 |

For given k, n and $P^* = .75$ (top), .85 (second), .90 (third), .95 (fourth), .99 (bottom), the entries in this table are the values of d which satisfy

$$\int G^k(x+d)\, g(x)\, dx = P^*$$

where $G(.)$ is the cdf and $g(.)$ the pdf of the median of a sample of size n from a standard normal population ; $n \geq 3$ is an odd integer.

# CHAPTER IV

# ON SELECTION OF POPULATIONS CLOSE TO A CONTROL

## 4.1 Introduction

Let $\pi_0, \pi_1, \dots, \pi_k$ be (k+1) independent populations with densities $f(x,\theta_0)$, $f(x,\theta_1),\dots,f(x,\theta_k)$, respectively, $\theta_i \in \Theta \subset \mathbb{R}$ $(i=0,1,\dots,k)$. The population $\pi_0$ is a control population. Let $E= [a(\theta_0), b(\theta_0)]$ be an interval in IR, and let $D=\{\theta_1,\dots,\theta_k\}$ Then the set D can be partitioned as follows:

$$D = D_1 \cup D_2 \ , \quad \text{where} \quad D_1 = D \cap E \ , \ D_2 = D \cap E^c$$

The subset of all populations with parameters in $D_1$ is of interest in many practical situations. The problem of selection of a subset containing all populations better than a standard corresponds to $E=(\theta_0,\infty)$. Gupta and Sobel (1958) have considered this problem for normal, gamma and binomial populations and have investigated a procedure for selecting a subset which contains all populations better than a standard with probability at least $P^*$, where $0<P^*<1$ is a preassigned constant. W.T. Huang (1975) has derived a Bayes rule for the problem of partitioning a set of k normal populations with respect to a control.

In Section 4.2 we consider the case $E=[A+\theta_o, B+\theta_o]$ where A and $B(-\infty<A<B<\infty)$ are known constants. The problem of selecting a subset containing populations with parameters in $[A+\theta_o, B+\theta_o]$ arises in many situations. For example, in dealing with matching parts, one is interested in populations which are close to a control. A Bayes procedure has been derived for the above problem when the underlying distributions are normal. The problem has also been considered from the subset selection approach and modified forms of rules proposed by Gupta and Sobel (1958) for selection of normal and gamma populations have been investigated in Section 4.3.

In the next two sections we discuss a slightly different problem. The goal in this case is to select the subset of a given size t $(1\leq t\leq k)$ of populations associated with $d_{[1]},\ldots,d_{[t]}$, where $d_i \equiv d(\theta_i,\theta_o)$ is a distance function and $d_{[1]} \leq\ldots\leq d_{[k]}$ represent the ordered $d_i$. In Section 4.4 , we have derived Bayes procedures for selection of t populations associated with t smallest values of $d(\theta_i-\theta_o)=(\theta_i-\theta_o)^2$; normal, binomial, Poisson and exponential populations are discussed.

In Section 4.5 the problem of selecting the t populations closest to control has been considered from the Empirical Bayes approach; Poisson, geometric and binomial populations have been discussed. Empirical Bayes subset selection rules for the problem of selection of largest or smallest parameter have been investigated by Deely (1965).

## 4.2 A Bayes Rule for Selecting All Normal Populations Close to Control

Let $\pi_0, \pi_1, \ldots, \pi_k$ be k+1 independent normal populations with means $\theta_0, \theta_0, \ldots, \theta_k$ ($\theta_i \in \Theta$ IR) and a common known variance $\sigma^2$. We say that $\pi_i$ is "close" to $\pi_0$ if $A+\theta_0 \le \theta_i \le B+\theta_0$, where A and B ($-\infty<A<B<\infty$) are given constants. The goal is to select a subset containing all populations close to $\Pi_0$.

The action space $\mathscr{A}$ consists of all subsets of $\{1,\ldots,k\}$, including the null set. Assume that the loss function $L:\Theta^{k+1} \times \mathscr{A} \to IR$ is given by

$$L(\underline{\theta},S)= \sum_{i\notin S} I_{[A+\theta_0,B+\theta_0]}(\theta_i) + \sum_{i\in S} I_{[A+\theta_0,B+\theta_0]^c}(\theta_i) \tag{4.2.1}$$

where $I_D(\cdot)$ is the indicator function of a set D.

The mean $\theta_0$ of population $\pi_0$ may or may not be known. We consider the two cases separately.

(1) $\underline{\theta_0 \text{ known}}$

Let $x_{i1},\ldots,x_{in}$ be an independent sample of size n from $\pi_i (i=1,\ldots,k)$. Suppose the a priori distribution $g(\underline{\theta})$ of $\underline{\theta}=(\theta_1,\ldots, \theta_k)$ is given by

$$g(\underline{\theta})= \prod_{i=1}^{k} \left[ \frac{1}{\tau} \ (\frac{\theta_i-\mu}{\tau}) \right] ,$$

where $\phi$ is the density of a standard normal random variable.

Then the posterior distribution of $\underline{\theta}$ is

$$g^*(\underline{\theta}) = \prod_{i=1}^{k} \left[ \frac{1}{v} \ \phi(\frac{\theta_i - m_i}{v}) \right],$$

where $m_i$ and $v$ are given by (2.5.2).

For $a = A + \theta_0$, $b = B + \theta_0$, let

$$S_1 = \{i : m_i \in [a,b],\ \Psi(m_i) \geq \tfrac{1}{2}\} \tag{4.2.2}$$

$$\text{where } \Psi(y) = \Phi(\tfrac{b-y}{v}) - \Phi(\tfrac{a-y}{v}),\ -\infty < y < \infty \tag{4.2.3}$$

Let $d_1$ be the rule which selects $S_1$ with probability one. We show that $d_1$ is a Bayes rule for the problem. As we are considering only Bayes rules, it suffices to show that

$$r(d_1,\underline{x}) \leq r(d,\underline{x}) \quad \text{for any nonrandomized rule } d \tag{4.2.4}$$

where $r(d,\underline{x})$, the Bayes posterior risk in using the rule $d$, is given by

$$r(d,\underline{x}) = \sum_{i \notin S} \Psi(m_i) + \sum_{i \in S} [1 - \Psi(m_i)] \tag{4.2.5}$$

Let a rule $d$ choose a set $S$ ($S \neq S_1$) with probability one. We show that $r(d,\underline{x}) \geq r(d_1,\underline{x})$. The following cases need to be considered:

(i) There exist $i$ and $j$ ($1 \leq i,\ j \leq k$) such that

$i \in S^c \cap S_1$, $j \in S \cap S_1^c$

Let

$$S' = (i,j)\,S$$

where $(i,j)S$ denotes the subset obtained from $S$ by replacing $j$ by $i$. Letting $d'$ denote the rule which selects the set $S'$ with probability one, we have

$$r(d,\underline{x}) - r(d',\underline{x}) = 2\,(\Psi(m_i) - \Psi(m_j))$$

where $\Psi(\cdot)$ is given by (4.2.3).

We can see, as in Section 1.7, that the function $\Psi(y)$ is symmetric about $\frac{a+b}{2}$, and strictly decreases with $|y - \frac{a+b}{2}|$. Since $m_i \in [a,b]$, $m_j \notin [a,b]$, we have

$$|m_i - \frac{a+b}{2}| < |m_j - \frac{a+b}{2}|$$

and hence $\Psi(m_i) > \Psi(m_j)$. Thus

$$r(d,\underline{x}) - r(d',\underline{x}) = 2\,(\Psi(m_i) - \Psi(m_j)) > 0$$

Since $S_1$ can be obtained from S by the operation used in (i), the inequality (4.2.4) follows.

(ii) $S \subset S_1$

It is easily seen from (4.2.5) that

$$r(d,\underline{x}) - r(d_1,\underline{x}) = \sum_{i \in S^c \cap S_1} [2\,\Psi(m_i) - 1] \geq 0$$

by definition of $S_1$.

(iii) $S \supset S_1$

In this case we have

$$r(d,\underline{x}) - r(d_1,\underline{x}) = \sum_{i \in S \cap S_1^c} [1 - 2\Psi(m_i)] \geq 0$$

by definition of $S_1$. It follows that $d_1$ is a Bayes rule for the problem.

(2) $\theta_0$ unknown

Here we are given samples $x_{i1},\ldots,x_{in}$ from all k+1 normal populations. $\pi_i$ (i= ,1,...,k); $\sigma^2$ is assumed to be known. Suppose that the a priori distribution of $\underline{\theta}=(\theta_0,\theta_1,\ldots,\theta_k)$ is

$$g(\theta_0,\theta_1,\ldots,\theta_k) = \prod_{i=0}^{k} \left[\frac{1}{\tau}\ \phi\left(\frac{\theta_i-\mu}{\tau}\right)\right] \tag{4.2.6}$$

Then $g^*(\theta_0,\theta_1,\ldots,\theta_k)$, the posterior distribution of $\underline{\theta}$ given the observations, is

$$g^*(\theta_0,\theta_1,\ldots,\theta_k) = \prod_{i=0}^{k} \left[\frac{1}{v}\ \phi\left(\frac{\theta_i-m_i}{v}\right)\right] \tag{4.2.7}$$

where v and $m_i$ are given by (4.2.2).

We wish to find a Bayes rule for the problem of selecting a subset containing all population for which $\theta_i\varepsilon[A+\theta_0,B+\theta_0]$, when the loss function is

$$L(\underline{\theta},S) = \sum_{i\notin S} I_{[A+\theta_0,B+\theta_0]}(\theta_i) + \sum_{i\varepsilon S}[1-I_{[A+\theta_0,B+\theta_0]}(\theta_i)]$$

Let $S_2=\{i:m_i-m_0\varepsilon\ [A,B]\ ,\ \psi^*(m_i-m_0)\geq 1/2\}$ (4.2.8)

where $\psi^*$ is same as $\psi$ given by (4.2.3),with v replaced by $v\sqrt{2}$. Also,let $d_2$ be the rule which selects the subset $S_2$ defined by

(4.2.8) with probability one. Then using the fact that the posterior distribution of $\theta_i-\theta_0$ is $N(m_i-m_0, 2v^2)$ we can show, as in Case (1), that

$$r(d_2, \underline{x}) \leq r(d, \underline{x})$$

i.e., $d_2$ is the Bayes rule in this case.

## 4.3 Selection of All the Populations Close to Control from Subset Selection Approach

Let $\pi_1, \ldots, \pi_k$ be k independent populations with densities $f(x,\theta_1), \ldots, f(x,\theta_k)$ $(x \in IR;\ \theta_i \in \Theta \subset IR,\ i=1,\ldots,k)$, respectively. Also let $\pi_0$ be a control population with density $f(x,\theta_0)$. Let $d: \Theta \to IR$ be a function which measures the distance of a point $\theta \in \Theta$ from $\theta_0$. We will say that the population $\pi_i$ is close to $\pi_0$ if $0 \leq d(\theta_i) \leq a$. In this section we will propose and study procedures to select a subset which contains the set $\{\pi_i : d(\theta_i) \leq a\}$ with probability at least $P^*$.

### Location Parameter- Normal Populations with Common Known Variance

Let $\pi_i$ be normal with mean $\theta_i$ and variance $\sigma^2 (i=0,1,\ldots,k)$, and define

$$d(\theta_i) = |\theta_i - \theta_0|$$

### Case A. $\theta_0$ known

A sample of size $n_i$ is taken from $\pi_i$ $(i=1,\ldots,k)$. Let $\bar{x}_i$ be the sample mean from $\pi_i$. For selecting a subset containing all populations $\pi_i$ with $|\theta_i-\theta_0| \leq a$, consider the following rule $R_A$:

$$\text{select } \pi_i \text{ iff } \theta_0 - \frac{d\sigma a}{\sqrt{n_i}} \leq \bar{x}_i \leq \theta_0 + \frac{d\sigma a}{\sqrt{n_i}} \tag{4.3.1}$$

The constant $d$ is chosen to satisfy

$$P(CS|R_A) \geq P^* \quad , \quad 0 < P^* < 1$$

where CS stands for correct solution, i.e. the selection of all $\pi_i$ with $|\theta_i - \theta_0| \leq a$. Let $k_1$ and $k_2$ denote the true number of populations with $|\theta_i - \theta_0| \leq a$ and $|\theta_i - \theta_0| > a$, respectively, so that $k_1 + k_2 = k$. If we let primes refer to values associated with $k_1$ populations with $|\theta_i - \theta_0| \leq a$, then

$$\begin{aligned} P(CS|R_A) &= \prod_{i=1}^{k_1} P(\theta_0 - \frac{\sigma da}{\sqrt{n_i}} \leq \bar{x}_i' \leq \theta_0 + \frac{\sigma da}{\sqrt{n_i'}}) \qquad (4.3.2) \\ &= \prod_{i=1}^{k_1} [\Phi(\frac{(\theta_0 - \theta_i')\sqrt{n_i'}}{\sigma} + ad) - \Phi(\frac{(\theta_0 - \theta_i')\sqrt{n_i'}}{\sigma} - ad)] \end{aligned}$$

where $\Phi(\cdot)$ represents the cdf of a standard normal random variable. Consider the function

$$\Psi(u) = \Phi(\frac{(\theta_0 - u)\sqrt{n_i'}}{\sigma} + ad) - \Phi(\frac{(\theta_0 - u)\sqrt{n_i'}}{\sigma} - ad) \qquad (4.3.3)$$

We can verify that

(i) $\lim_{u \to \infty} \Psi(u) = 0 = \lim_{u \to -\infty} \Psi(u)$

(ii) $\Psi(\theta_0 + u) = \Psi(\theta_0 - u)$, i.e., the function $\Psi(u)$ is symmetric about $u = \theta_0$.

(iii) $$\begin{aligned} \frac{d}{du} \Psi(u) &= \frac{\sqrt{n_i'}}{\sigma} [\phi(\frac{(\theta_0 - u)\sqrt{n_i'}}{\sigma} - ad) - \phi(\frac{(\theta_0 - u)\sqrt{n_i'}}{\sigma} + ad)] \\ &> 0 \text{ if } u < \theta_0 \\ &= 0 \text{ if } u = \theta_0 \\ &< 0 \text{ if } u > \theta_0 \end{aligned}$$

Hence the function $\Psi(u)$ is increasing if $u < \theta_0$ , decreasing if $u > \theta_0$ and attains its maximum at $u=\theta_0$. It follows that

$$\inf_{|u-\theta_0|\leq a} \Psi(u) = \Psi(\theta_0-a) = \Psi(\theta_0+a)$$

Hence

$$\inf P(CS|R_A) = \prod_{i=1}^{k_1} [\Phi(a(\frac{\sqrt{n_i'}}{\sigma} + d)) - \Phi(a(\frac{\sqrt{n_i'}}{\sigma} -d))]$$

$$\geq \prod_{i=1}^{k} [\Phi(a(\frac{\sqrt{n_i}}{\sigma} +d)) - \Phi(a(\frac{\sqrt{n_i}}{\sigma} - d))]$$

The constant d is obtained by equating the right hand side of the above inequality to $P^*$. For unequal sample sizes, computation of d is difficult. If $n_i=n \; \forall \; i=1,\ldots,k$, we have

$$P(CS|R_A) \geq [\Phi(a(\frac{\sqrt{n}}{\sigma}+ d)) - \Phi(a(\frac{\sqrt{n}}{\sigma} - d))]^k$$

and hence d can be found by solving

$$\Phi(a(\frac{\sqrt{n}}{\sigma} + d)) - \Phi(a( \frac{\sqrt{n}}{\sigma} -d) =(P^*)^{1/k} \qquad (4.3.4)$$

Case B. $\theta_0$ unknown

In this case observations from all the k+1 populations are taken. Let $\bar{x}_i$ be the sample mean of $n_i$ observations from $\pi_i$ $(i=0,1,\ldots,k)$. Consider the following selection rule:

$$R_B\text{: select } \pi_i \text{ iff } \bar{x}_0 - \frac{D\sigma a}{\sqrt{n_i}} \leq \bar{x}_i \leq \bar{x}_0 + \frac{D\sigma a}{\sqrt{n_i}} \qquad (4.3.5)$$

Simple calculation gives

$$P(CS|R_B) = \int_{-\infty}^{\infty} \prod_{i=1}^{k_1} [\Phi(\frac{\sqrt{n_i'}}{\sqrt{n_0}} y + \frac{\sqrt{n_i'}}{\sigma} (\theta_0 - \theta_i') + Da) - \Phi(\frac{\sqrt{n_i'}}{\sqrt{n_0}} y + \frac{\sqrt{n_i'}}{\sigma} (\theta_0 - \theta_i') - Da)] \phi(y)\, dy \tag{4.3.6}$$

where $\theta_i'$, $n_i'$ $(i=1,\ldots,k)$ and $k_1$ are as in case (a).

Define

$$\Psi(\theta,y) = \Phi(\frac{\sqrt{n_i'}}{\sigma} [\frac{\sigma}{\sqrt{n_0}} y + \theta_0 - \theta] + Da) - \Phi(\frac{\sqrt{n_i'}}{\sigma} [\frac{\sigma}{\sqrt{n_0}} y + \theta_0 - \theta] - Da) \tag{4.3.7}$$

We can easily verify that, for a fixed $y \in \mathrm{IR}$

(i) $\Psi(y,\theta)$ is continuous $\forall\ \theta \in \mathrm{IR}$ and hence attains its minimum in the compact set $\{|\theta_i' - \theta_0| \leq a\}$.

(ii) $\lim\limits_{\theta \to \pm\infty} \Psi(y,\theta) = 0$

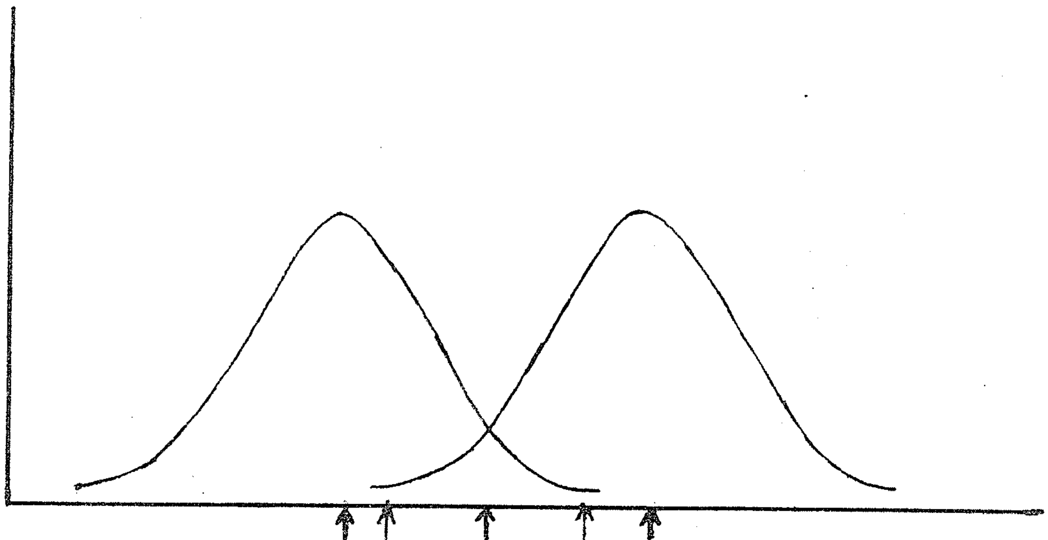
(iii) $\Psi(\frac{\sigma}{\sqrt{n_0}} y + \theta_0 + \theta) = \Psi(\frac{\sigma}{\sqrt{n_0}} y + \theta_0 - \theta)$

that is, for each fixed y, $\Psi(y,\theta)$ is symmetric about $\theta = \frac{\sigma}{\sqrt{n_0}} y + \theta_0$.

(iv) $$\frac{d}{d\theta} \Psi(y,\theta) = \frac{1}{\sigma} \frac{\sqrt{n_i'}}{\sqrt{2\pi}} [ e^{-1/2 (\frac{\sqrt{n_i'}}{\sigma} [\frac{\sigma}{\sqrt{n_0}} y + \theta_0 - \theta] - Da)^2} - e^{-1/2 (\frac{\sqrt{n_i'}}{\sigma} [\frac{\sigma}{\sqrt{n_0}} y + \theta_0 - \theta] + Da)^2} ]$$

$$\begin{matrix} > \\ = \\ < \end{matrix} 0 \text{ if } \theta \begin{matrix} < \\ = \\ > \end{matrix} \frac{\sigma}{\sqrt{n_0}} y + \theta_0$$

It follows from (i) to (v) above that, for fixed $y \in \mathbb{R}$, the form of $\Psi$ is as shown in Figure 1.



$\frac{\sigma}{\sqrt{n_0}}\,y+\theta_0$, $\theta_0-a$, $\theta_0$, $\theta_0+a$, $\frac{\sigma}{\sqrt{n_0}}\,y+\theta_0$

( $y < 0$ ) ( $y > 0$ )

Figure 1. The Function $\Psi(\theta)$

It is clear from Figure 1 that

$$\inf_{\theta_i' \in \{|\theta_i'-\theta_0| \leq a\}} \Psi(\theta_i') = \begin{cases} \Psi(\theta_0+a) & \text{if} \quad y < 0 \\ \Psi(\theta_0-a) & \text{if} \quad y > 0 \end{cases} \tag{4.3.8}$$

From (4.3.6) and (4.3.8) we have

$$\inf_{\{|\theta_i'-\theta_0| \leq a\}} P(CS|R_B) = \int_{-\infty}^{0} \prod_{i=1}^{k_1} \Psi(\theta_0+a,y)\phi(y)\,dy + \int_{0}^{\infty} \prod_{i=1}^{k_1} \Psi(\theta_0-a,y)\phi(y)\,dy$$

$$\geq \int_{-\infty}^{0} \prod_{i=1}^{k} \Phi[\frac{\sqrt{n_i}}{\sqrt{n_0}} y + a(D - \frac{\sqrt{n_i}}{\sigma}))$$

$$- \Phi(\frac{\sqrt{n_i}}{\sqrt{n_0}} y - a(D + \frac{\sqrt{n_i}}{\sigma}))] \quad \phi(y)\, dy$$

$$+ \int_{0}^{\infty} \prod_{i=1}^{k} [\Phi(\frac{\sqrt{n_i}}{\sqrt{n_0}} y + a(D + \frac{\sqrt{n_i}}{\sigma})$$

$$- \Phi(\frac{\sqrt{n_i}}{\sqrt{n_0}} y - a(D - \frac{\sqrt{n_i}}{\sigma}))] \quad \phi(y)\, dy \tag{4.3.9}$$

Since the function on the right hand side of (4.3.9) increases from 0 to 1 as d ranges from 0 to $\infty$, the equation for $D=D(P^*)$ can be solved for any $0 \leq P^* \leq 1$.

Scale Parameter- Gamma Populations with Known Shape Parameters

Here $\pi_i$ $(i = 0,1,\ldots,k)$ has density

$$g(x ; \theta_i, \alpha_i) = \frac{\theta_i^{-\alpha_i/2}}{\Gamma(\frac{\alpha_i}{2}))} x^{(\alpha_i/2)-1} e^{-x/\theta_i} \tag{4.3.10}$$

where $x, \theta_i, \alpha_i > 0$, and $\alpha_i$ are known constants. Let $G(x; \theta_i, \alpha_i)$ denote the corresponding cdf. In this case we say that $\pi_i$ is close to $\pi_0$ if

$$\frac{1}{\beta} \leq \frac{\theta_i}{\theta_0} \leq \beta$$

where $\beta > 1$ is a given constant.

Case A. $\theta_0$ known.

Consider the rule $R_A'$: select $\pi_i$ iff $\frac{T_i}{\nu_i} \in [\frac{\beta\theta_0}{c}, \beta\theta_0 c]$,

where

$$T_i = \sum_{j=1}^{n_i} X_{ij}$$

$$\nu_i = n_i\alpha_i \quad , \; i = 1,\ldots,k$$

and $c > 1$ is chosen so as to satisfy the P*-condition.

Using the fact that the cdf of $\frac{T_i}{\theta_i}$ is $G(t) \equiv G(t;1,\nu_i)$ we obtain

$$P(CS|R_A') = \prod_{i=1}^{k_1} [G(\frac{\beta\theta_0 c\nu_i'}{\theta_i'};1,\nu_i) - G(\frac{\beta\theta_0\nu_i'}{c\theta_i'};1,\nu_i)] \tag{4.3.11}$$

Here, as before, $k_1$ is number of populations $\pi_i$ with $\frac{1}{\beta} \leq \frac{\theta_i}{\theta_0} \leq \beta$, and the primes refer to values corresponding to the populaitons close to $\pi_0$. Define

$$H_i(\theta_i') = G(\frac{\beta\theta_0 c\nu_i'}{\theta_i'};1,\nu_i) - G(\frac{\beta\theta_0\nu_i'}{c\theta_i'};1,\nu_i) \tag{4.3.12}$$

It is easy to verify that the function $H_i(\theta_i')$ is increasing in $\theta_i'$ if $\theta_i' \leq \beta\theta_0$, and hence

$$\inf_{\frac{\theta_0}{\beta} \leq \theta_i' < \beta\theta_0} P(CS|R_A') = \prod_{i=1}^{k_1} H_i(\frac{\theta_0}{\beta})$$

$$\geq \prod_{i=1}^{k} H_i(\frac{\theta_0}{\beta})$$

Hence the constant c is given by the following equation:

$$\prod_{i=1}^{k} [G(\beta^2 c \nu_i;1,\nu_i) - G(\frac{\beta^2\nu_i}{c};1,\nu_i)] = P^* \tag{4.3.13}$$

<u>Case B. $\theta_0$ unknown.</u>

In this case, consider the rule $R_B'$

$$\text{select } \pi_i \text{ iff } \frac{\beta T_0}{\nu_0 C} \le \frac{T_i}{\nu_i} \le \frac{\beta T_0 C}{\nu_0}$$

where $T_i$ $(i = 0,1,\ldots,k)$ are as defined in case A, and $C > 1$ is a constant.

Then, as in case A, we have

$$P(CS|R_B') \ge \int_0^\infty \prod_{i=1}^{k} [G(\frac{\beta^2 C\nu_i}{\nu_0} u;1,\nu_i) - G(\frac{\beta^2\nu_i}{C\nu_0} u;1,\nu_i)]$$

$$\cdot g(u;1,\nu_0)du \tag{4.3.14}$$

The constant C can be obtained by equating the right hand side of (4.3.14) to P*.

<u>Application to Selecting Variances of Normal Populations</u>

Let $\pi_i$ be a normal population with mean $\mu_i$ and variance $\sigma_i^2$ $(i = 0,1,\ldots,k)$. We will say that $\pi_i$ is close to $\pi_0$ if $\frac{1}{\beta} < \frac{\theta_i}{\theta_0} < \beta$ where $\theta_i = 2\sigma_i^2$ $(i = 0,1,\ldots,k)$ and $\beta > 1$ is a known constant.

For the case where the $\mu_i$ $(i = 0,1,\ldots,k)$ are known, we consider a rule based on $S_i^2 = \sum_{j=1}^{n_i} (X_{ij} - \mu_i)^2/n_i$ which is sufficient for $\sigma_i^2$.

If $\sigma_0$ is known, the rule is : select $\pi_i$ iff

$$\frac{2\beta\sigma_0^2}{d} \leq S_i^2 \leq 2\beta d\sigma_0^2 \qquad 1 < d$$

Using the fact that $\frac{n_i S_i^2}{\theta_i}$ is distributed as $\frac{1}{2}\chi^2_{n_i}$ random-variable, we can show that the equation for d is the same as (4.3.14) with $\nu_i = n_i$. If the means $\mu_i$ $(i = 0, 1,\ldots,k)$ are unknown and $n_i > 1$ $(i = 1,\ldots,k)$, then we use $\bar{x}_i$ in place of $\mu_i$ and $n_i' - 1$ for $n_i$. The constant d is given by (4.3.14) with $\nu_i = n_i - 1$.

## 4.4 Bayes Rules for Selection of t Populations Closest to Control.

Let $\pi_0, \pi_1,\ldots,\pi_k$ be k+1 independent populations with density function $f(x,\theta_0), f(x,\theta_1),\ldots,f(x,\theta_k)$, $i = 0, 1,\ldots,k$. The goal here is to select a subset S of a given size t $(1 \leq t < k)$ from $\{1,\ldots,k\}$ such that

$$\max_{j\in S} (\theta_i - \theta_0)^2 \leq (\theta_i - \theta_0)^2 \qquad , \forall\, i \notin S$$

The action space $\mathcal{a}$ consists of $\binom{k}{t}$ elements:

$$\mathcal{a} = \{S \subset \{1,\ldots,k\} : \ |S| = t\}$$

where $|S|$ denotes the size of a subset S.

A reasonable loss function L: $\Theta^{k+1} \times \mathcal{A} \to \mathbb{R}$ in this situation is

$$L(\underline{\theta},S) = \sum_{j \in S} [(\theta_j - \theta_0)^2 - \min_{1 \le i \le k} (\theta_i - \theta_0)^2] \quad (4.4.1)$$

We will derive Bayes rules in several examples.

1. Normal populations with common known variance.

Let $\pi_i$ be normal with mean $\theta_i$ and a known variance $\sigma^2$, $i = 0, 1, \ldots, k$. The mean $\theta_0$ of the control population $\pi_0$ may or may not be known. We discuss the two cases separately.

A. $\theta_0$ known

Let $x_{ij}$ $(j = 1, \ldots, n)$ be an independent sample of size n from $\pi_i$ $(i = 1, \ldots, k)$, and $\bar{x}_i$ the mean of the sample, i.e.,

$$\bar{x}_i = \frac{\sum_{j=1}^{n} x_{ij}}{n}, \quad i = 1, \ldots, k$$

Assume that the a priori distribution of $\underline{\theta} = (\theta_1, \ldots, \theta_k)$ is

$$g(\underline{\theta}) = \prod_{i=1}^{k} [\frac{1}{\tau} \phi(\frac{\theta_i - \mu}{\tau})]$$

where $\phi(x)$ is the pdf of a standard normal random variable. Then the posterior distribution of $\underline{\theta}$ given $(\bar{x}_1, \ldots, \bar{x}_k)$ is

$$g^*(\underline{\theta}|\underline{x}) = \prod_{i=1}^{k} [\frac{1}{v} \phi(\frac{\theta_i - m_i}{v})]$$

where $m_i$ and $v$ are given by (2.5.2)

The Bayes posterior risk in choosing a subset $S \in \mathcal{Q}$ is given by

$$r(S,\underline{x}) = \sum_{j \in S} E[\{(\theta_j - \theta_0)^2 - \min_{1 \leq i \leq k} (\theta_i - \theta_0)^2\}|\bar{\underline{x}}] \tag{4.4.2}$$

where $\bar{\underline{x}} = (\bar{x}_1,\ldots,\bar{x}_k)$.

It follows from (4.4.2) that a Bayes rule for the problem is to select the subset $S^*$ defined by

$$S^* = \{j: \sum_{j \in S^*} E\{(\theta_j - \theta_0)^2|\bar{\underline{x}}\} = \min_{S \in \mathcal{Q}} \sum_{j \in S} E\{(\theta_j - \theta_0)^2|\bar{\underline{x}}\} \tag{4.4.3}$$

Simple calculation gives

$$E[(\theta_j - \theta_0)^2|\underline{x}] = D_j + v^2 \tag{4.4.4}$$

where $D_j = (M_j - \theta_0)^2$

Let $D_{[1]} \leq \cdots \leq D_{[k]}$ denote the ordered $D_j$. It is clear from (4.4.3) that Bayes rule will select the subset of populations associated with $\{D_{[1]},\ldots,D_{[t]}\}$.

B. $\theta_0$ unknown

Let $\bar{x}_i$ be the mean of n independent observations from $\pi_i$ $(i = 0, 1,\ldots,k)$ and assume that the a priori distribution of $(\theta_0, \theta_1,\ldots,\theta_k)$ is

$$g(\theta_0, \theta_1,\ldots,\theta_k) = \prod_{i=0}^{k} \{\frac{1}{\tau} \phi(\frac{\theta_i - \mu}{\tau})$$

As in Case A we have

$$E[(\theta_j - \theta_0)^2 | \bar{x}_0, \bar{x}_1, \ldots, \bar{x}_k] = D_j' + 2v^2 \text{ where } D_j' = (m_j - m_0)^2$$

and therefore the Bayes rule selects the populations associated with $\{D'_{[1]}, \ldots, D'_{[t]}\}$.

2. Selection in terms of variances of normal populations

Here $\pi_i$ is normal with mean zero and variance $\frac{1}{\theta_i}$ $(i = 0, 1, \ldots, k)$.

A. $\theta_0$ known

Here $x_{i1}, \ldots, x_{in}$ is an independent sample from $\pi_i$ $(i = 1, \ldots, k)$. Let

$$s_i^2 = \frac{1}{n} \sum_{j=1}^{n} x_{ij}^2 \qquad i = 1, \ldots, k$$

Assuming that $\underline{\theta} = (\theta_1, \ldots, \theta_k)$ has a priori distribution $g(\underline{\theta})$ given by (2.5.7) we have, as in example 2 of Section 2.5,

$$E\{(\theta_j - \theta_0)^2 | s_1^2, \ldots, s_k^2\} = \frac{[(\frac{2}{a''}+1) - \theta_0 b_i'']^2}{(b_i'')^2(\frac{2}{a''}+1)} + \frac{2\theta_0{}^2}{2+a''}$$

where $a''$, $b_i''$ are given by (2.5.9)

Hence the Bayes rule will select the populations associated with t smallest $D_j$, where

$$D_j = \left[\frac{\frac{2}{a''}+1}{b_j''} - \theta_0\right]^2$$

B. $\theta_0$ unknown

In this case a sample $x_{01},\ldots x_{0n}$ is taken from $\pi_0$, and let $s_0^2 = \sum_{j=1}^{n} x_{0j}^2$. Let the a priori distribution of $(\theta_0, \theta_1,\ldots,\theta_k)$ be

$$g(\theta_0, \theta_1,\ldots,\theta_k) = \prod_{i=0}^{k} f_{\gamma_2}(\theta_i|b', a')$$

Then, as in case A, we have

$$E[(\theta_j - \theta_0)^2 | s_0^2, s_1^2,\ldots,s_k^2] = \frac{2 + a''}{a''(b_j'')^2} - \frac{2}{b_j''b_0''} + \frac{2 + a''}{a''(b_0'')^2}$$

$$= \frac{\left[\frac{2 + a''}{a''b_j''} - \frac{1}{b_0''}\right]^2}{\frac{2 + a''}{a''}} + \frac{\frac{2 + a''}{a''} - \frac{a''}{2 + a''}}{(b_0'')^2}$$

Thus the Bayes rule will select the populations associated with t smallest $D_j'$, where

$$D_j' = \left[\frac{2 + a''}{a''b_j''} - \frac{1}{b_0''}\right]^2$$

3 Binomial Populations

Here we are given $\pi_0, \pi_1,\ldots,\pi_k$ with probabilities of success $\theta_0, \theta_1,\ldots,\theta_k$, respectively. As before, we consider two cases:

A. $\theta_0$ known.

Let $x_i$ be the number of successes in n independent trials from

$\pi_i$ $(i = 1,\ldots,k)$. Assume $\theta_i$ are independent, and each has the natural conjugate beta prior with parameters a and b, given by (2.5.11). Then, conditional on $\underset{\sim}{x} = (x_1,\ldots,x_k)$, $\theta$'s are independent, with $\theta_i$ having a beta distribution with parameter $(x_i + a)$ and $(n + b - x_i)$, and it is easy to show that

$$E[(\theta_j - \theta_0)^2 | \underset{\sim}{x}] = \frac{(x_j + a)(x_j + a + 1)}{(n + a + b)(n + a + b + 1)} - \frac{2\theta_0(x_j + a)}{n + a + b} + \theta_0^2$$

$$= \frac{[x_j + a + \frac{1}{2} - (n + a + b + 1)\theta_0]^2 + (n + a + b + 1)\theta_0(1 - \theta_0) - \frac{1}{4}}{(n + a + b)(n + a + b + 1)}$$

Setting $D_j = [x_j + a + \frac{1}{2} - (n + a + b + 1)\theta_0]^2$, we see that the Bayes rule selects the populations associated with t smallest $D_j$.

B. $\theta_0$ unknown

In this case, n independent trials are performed on $\pi_0$ also; let $x_0$ be the number of success from $\pi_0$. Then, if $\theta_0$, $\theta_1,\ldots,\theta_k$ are assumed to be independent, each with natural conjugate beta prior $f(\cdot|a,b)$, we have, as in case A,

$$E[(\theta_j - \theta_0)^2 | x_0, x_1,\ldots,x_k] = \frac{(x_j + a)(x_j + a + 1)}{(n + a + b)(n + a + b + 1)} - \frac{2(x_j+a)(x_0+a)}{(n + a + b)^2}$$

$$+ \frac{(x_0 + a)(x_0 + a + 1)}{(n + a + b)(n + a + b + 1)}$$

$$= \frac{D_j' + (2 + \frac{1}{n + a + b})(x_0 + a)(1 - \frac{x_0 + a}{n + a + b})}{(n + a + b)(n + a + b + 1)}$$

where $$D_j' = [x_j + a + \tfrac{1}{2} - \frac{n + a + b + 1}{n + a + b}(x_0 + a)]^2$$

Hence the Bayes rule selects populations associated with the t smallest $D_j'$.

4. Poisson populations

Here $\pi_i$ is Poisson with parameter $\theta_i$ ($i = 0, 1, \ldots, k$).

A. $\theta_0$ known

Assume $\theta_1, \ldots, \theta_k$ are independent, and each has the natural conjugate a priori distribution $f_{\gamma_1}(\cdot|a', b')$ defined by (2.5.12). Then, if $x_i$ is observed from $\pi_i$, we have, as in examples 4 of Section 2.5,

$$E[(\theta_j - \theta_0)^2 | \underline{x}] = D_j - \frac{1}{4(1 + b')^2} - \frac{\theta_0}{1 + b'}$$

where $$D_j = \left[\frac{x_j + a'}{1 + b'} - (\theta_0 - \frac{1}{2(1 + b')})\right]^2$$

and the Bayes rule is to select the populations corresponding to $\{D_{[1]}, \ldots, D_{[t]}\}$.

B. $\theta_0$ unknown

Here we observe $x_i$ from $\pi_i$ ($i = 0, 1, \ldots, k$). Then assuming $(\theta_0, \theta_1, \ldots, \theta_k)$ has a priori distribution

$$g(\theta_0, \theta_1, \ldots, \theta_k) = \prod_{i=0}^{k} f_{\gamma_1}(\theta_i | a', b')$$

we have,

$$E[(\theta_j - \theta_o)^2 | x_0, x_1, \ldots, x_k] = D'_j + \frac{2(a' + x_0)}{(1 + b')^2} - \tfrac{1}{4}$$

where $$D'_j = [\frac{x_j - (x_0 - \frac{1}{2})}{1 + b'}]^2$$

The Bayes rule in this case selects the populations corresponding to the subset $\{D'_{[1]}, \ldots, D'_{[t]}\}$.

5. A problem in life testing

Let $\pi_0, \pi_1, \ldots, \pi_k$ be k independent exponential populations with parameters $\theta_0, \theta_1, \ldots, \theta_k$ respectively, i.e., the density of the i-th population is

$$f(x|\theta_i) = \frac{1}{\theta_i} e^{-\frac{x}{\theta_i}}, \quad x > 0, \theta_i > 0, i = 0, 1, \ldots, k$$

Here, $\theta_0$ may or may not be known. As before, the two cases will be discussed separately. We will assume, in each case, that $\theta_i$ are independent, and each has the same a priori density $h(\cdot)$. The following a priori distributions, used by Bhattacharya (1967), will be considered:

$$\text{(i)} \quad h(\theta) = \frac{(a-1)(\alpha\beta)^{a-1}}{(\beta^{a-1} - \alpha^{a-1})\theta^a} I_{[\alpha,\beta]}(\theta), \; 0 < \alpha < \beta, \; a > 1 \qquad (4.4.5)$$

where $I_{[\alpha,\beta]}(\cdot)$ is the indicator function of $[\alpha,\beta]$.

$$\text{(ii)} \quad h(\theta) = \frac{1}{\lambda} e^{-\frac{\theta}{\lambda}}, \; \theta > 0, \; \lambda > 0 \qquad (4.4.6)$$

(iii) $$h(\theta) = \frac{e^{-\frac{\mu}{\theta}} (\frac{\mu}{\theta})^{\nu+1}}{\mu\Gamma(\nu)}, \quad \theta > 0, \quad \mu, \nu > 0 \tag{4.4.7}$$

The prior density (4.4.7) is the natural conjugate prior density for $\theta$.

A. $\theta_0$ known.

Here n items from $\pi_i$ $(i = 1,\ldots,k)$ are put on a life test, and the experiment continued until the first r failures are observed. Let $T_i$ be the accumulated life at the termination of the test on $\pi_i$, defined by (2.5.13). Then the probability density of $T_i$ is given by (2.5.14). Suppose the a priori distribution of $\underline{\theta} = (\theta_1,\ldots,\theta_k)$ is

$$g(\underline{\theta}) = \prod_{i=1}^{k} h(\theta_i)$$

where $h(\cdot)$ is one of the three a priori denisties given above.

The posterior density $g^*(\underline{\theta}|\underline{T})$ of $\underline{\theta} = (\theta_1,\ldots,\theta_k)$, given $\underline{T} = (T_1,\ldots,T_k)$, in each case is obtained from results of Bhattacharya (1967).

When $h(\theta)$ is given by (4.4.5) we have

$$g^*(\underline{\theta}|\underline{T}) \propto \prod_{i=1}^{k} [\frac{1}{\theta_i^{r+a}} e^{-\frac{T_i}{\theta_i}} I_{[\alpha,\beta]}(\underline{\theta}_i)]$$

That is, $\theta_i$ are conditionally independent and have truncated inverted gamma posterior distributions (see Raiffa and Schlaifer (1961)). It follows from Bhattacharya (1967) that

$$E(\theta_i|\underline{T}) = \frac{\gamma^*(r+a-2,T_i)}{\gamma^*(r+a-1,T_i)} \; T_i$$

$$E(\theta_i^2|\underline{T}) = \frac{\gamma^*(r+a-3,T_i)}{\gamma^*(r+a-1,T_i)} \cdot \; T_i^2$$

where $\gamma^*(n,y) = \gamma(n, \frac{y}{\alpha}) - \gamma(n, \frac{y}{\beta})$

and $\gamma(n,x) = \int_0^x e^{-s} s^{n-1} ds$ is the incomplete gamma function defined for $x > 0$.

Hence $E[(\theta_j - \theta_0)^2|\underline{T}] = D_j^{(1)} + \theta_0^2$

where

$$D_j^{(1)} = \frac{\gamma^*(r+a-3,T_j)}{\gamma^*(r+a-1,T_j)} \; T_j^2 - 2\theta_0 \frac{\gamma^*(r+a-2,T_j)}{\gamma^*(r+a-1,T_j)} \; T_j$$

For $h(\theta)$ given by (4.4.6) we have

$$g^*(\underline{\theta}|\underline{T}) \propto \prod_{i=1}^{k} \left[\frac{1}{\theta_i^r} e^{-\left(\frac{T_i}{\theta_i} + \frac{\theta_i}{\lambda}\right)}\right], \; \theta_i > 0 \quad \forall \, i$$

The normalizing constant $\prod_{i=1}^{k} \left[\int_0^\infty \frac{1}{\theta^r} e^{-\left(\frac{T_i}{\theta} + \frac{\theta}{\lambda}\right)} d\theta\right]^{-1}$

is evaluated by using an integral representation of $K_\nu(z)$, the modified Bessel function of the third kind of order $\nu$, given in Erdelyi et al (1953) [see Bhattacharya (1967)].

Then

$$E(\theta_i|\underline{T}) = \sqrt{\lambda T_i} \; \frac{K_{r-2}(2\sqrt{T_i/\lambda})}{K_{r-1}(2\sqrt{T_i/\lambda})} \; , \quad E(\theta_i^2|\underline{T}) = \lambda T_i \frac{K_{r-3}(2\sqrt{T_i/\lambda})}{K_{r-1}(2\sqrt{T_i/\lambda})}$$

and hence

$$E[(\theta_j - \theta_0)^2 | \underset{\sim}{T}] = D_j^{(2)} + \theta_0^2$$

where

$$D_j^{(2)} = \frac{\lambda T_i \, K_{r-3}(2\sqrt{T_i/\lambda}) - 2\theta_0\sqrt{\lambda T_i}\, K_{r-2}(2\sqrt{T_i/\lambda})}{K_{r-1}(2\sqrt{T_i/\lambda})}$$

In case $h(\theta)$ is given by (4.4.7), we have

$$g^*(\underset{\sim}{\theta}|\underset{\sim}{T}) = \prod_{i=1}^{k} \left[\frac{e^{-\frac{\mu + T_i}{\theta_i}} \left(\frac{\mu + T_i}{\theta_i}\right)^{r+\nu+1}}{(\mu + T_i)\Gamma(r + \nu)}\right], \; \theta_i > 0$$

It follows from Bhattacharya (1967) that

$$E[(\theta_j - \theta_0)^2|\underset{\sim}{T}] = \left[\frac{\mu + T_i}{r + \nu - 2} - \theta_0\right]^2 \frac{r + \nu - 2}{r + \nu - 1} + \frac{\theta_0^2}{r + \nu - 1}$$

Hence the Bayes rule in the first two cases is to select populations associated with the corresponding t smallest D-statistics, and in the last case, is to select the t populations for which $\frac{\mu + T_i}{r + \nu - 2}$ is closest to $\theta_0$.

B. $\theta_0$ unknown

Here items from $\pi_0$ also are put on the same life test. Let $T_0$ be the total accumulated life until r failures. In this case, we assume that the a priori distribution of $\underset{\sim}{\theta} = (\theta_0, \theta_1, \ldots, \theta_k)$ is

$$g(\theta_0, \theta_1, \ldots, \theta_k) = \prod_{i=0}^{k} h(\theta_i)$$

where $h(\cdot)$ is one of the density functions considered above. Then, proceeding as in case A, we can show that the Bayes rule is to select t populations associated with t smallest values of

$$\text{(i)} \quad \frac{\gamma^*(r + a - 3, T_i)}{\gamma^*(r + a - 1, T_i)} T_i^2 - \frac{2\gamma^*(r + a - 2, T_i)\gamma^*(r + a - 2, T_0)}{\gamma^*(r + a - 1, T_i)\gamma^*(r + a - 1, T_0)}$$

$$+ \frac{\gamma^*(r + a - 3, T_0)}{\gamma^*(r + a - 1, T_0)} T_0^2 \qquad \text{when } h(\cdot) \text{ is given by (4.4.5)}$$

$$\text{(ii)} \quad \lambda T_i \frac{K_{r-3}(2\sqrt{T_i/\lambda})}{K_{r-1}(2\sqrt{T_i/\lambda})} - 2\lambda\sqrt{T_i T_0} \frac{K_{r-2}(2\sqrt{T_i/\lambda})K_{r-2}(2\sqrt{T_0/\lambda})}{K_{r-1}(2\sqrt{T_i/\lambda})K_{r-1}(2\sqrt{T_0/\lambda})}$$

$$+ \lambda T_0 \frac{K_{r-3}(2\sqrt{T_0/\lambda})}{K_{r-1}(2\sqrt{T_0/\lambda})} \qquad \text{when } h(\cdot) \text{ is given by (4.4.6)}$$

It can also be seen that, for $h(\cdot)$ given by (4.4.7), the Bayes rule selects the t populations for which $(\mu + T_i)$ is closest to $\frac{r + \nu - 2}{r + \nu - 1} (\mu + T_0)$.

## 4.5 Selection of the Population Closest to Control-an Emprical Bayes Approach

In this Section we derive empirical Bayes rules for selecting the population closest to control.

We are given k+1 ($k \geq 2$) independent populations $\pi_0, \pi_1, \ldots, \pi_k$ with densities $f(x,\theta_0), f(x,\theta_1), \ldots, f(x,\theta_k)$, respectively, with respect to a $\sigma$-finite measure $\mu$. Our interest is in the population with the smallest value of $d(\theta_j;\theta_0) = (\theta_j - \theta_0)^2$.

We now formulate the problem in the empirical Bayes framework of Robbins (1964).

We are given:

(i) a parameter space $\Theta^{k+1} \subset \mathbb{R}^{k+1}$, where $\Theta$ is a subset of $\mathbb{R}$, the real line.

(ii) an action space $\mathcal{A} = \{a_1, \ldots, a_k\}$, where $a_i$ is the action which selects $\pi_i$.

(iii) a continuous loss function L: $\Theta^{k+1} \times \mathcal{A} \to \mathbb{R}$, with $L(\underline{\theta}, a_j)$ denoting the loss incurred when $\underline{\theta} \in \Theta^{k+1}$ is the state of nature and action $a_j$ is taken. For our discussion

$$L(\underline{\theta}, a_j) = (\theta_j - \theta_0)^2 - \min_{1 \leq i \leq k} (\theta_i - \theta_0)^2, \quad 1 \leq j \leq k \tag{4.5.1}$$

(iv) an a priori distribution $G(\theta_0, \theta_1, \ldots, \theta_k) = \prod_{j=0}^{k} F(\theta_j)$ of $(\theta_0, \theta_1, \ldots, \theta_k)$ on $\Theta^{k+1}$, where F is an absolutely continuous (a.c.) distribution function on $\Theta$.

(v) an observable random variable $X_j \in \mathcal{X}_j$, $j = 0, 1, \ldots, k$ on which $\sigma$-finite measure $\mu$ is defined; for $\theta_j \in \Theta$, $x_j$ has probability density $f(\cdot|\theta_j)$ with respect to $\mu$.

We wish to find a decision function

$$t_G : \mathscr{X} = \mathop{X}_{j=0}^{k} \mathscr{X}_j \to \mathcal{G}$$

such that

$$R(t_G, G) = \int_{\Theta^{k+1}} \int_{\mathscr{X}} L(t_G(x), \underline{\theta}) h(\underline{x}|\underline{\theta}) d\mu(\underline{x}) dG(\underline{\theta}) \tag{4.5.2}$$

$$= \min_t R(t, G)$$

where $h(\underline{x}|\underline{\theta}) = \prod_{j=0}^{k} f(x_j|\theta_j)$, $\underline{x} = (x_0, x_1, \ldots, x_k)$, $\underline{\theta} = (\theta_0, \theta_1, \ldots, \theta_k)$

and $d\mu(\underline{x}) = d\mu(x_0) \times d\mu(x_1) \times \ldots \times d\mu(x_k)$

Since we are interested in rules which are Bayes with respect to a priori distribution $G \in \mathcal{G}$, attention can be restricted to nonrandomized rules .

We assume that a Bayes rule $t_G$ for each $G \in \mathcal{G}$ exists. Then,

$$R(t_G, G) = \int_{\mathscr{X}} \min_{1 \leq j \leq k} H_G(a_j, \underline{x}) \, d\mu(\underline{x}) \leq R(t, G), \ \forall \ t \tag{4.5.3}$$

where

$$H_G(a, \underline{x}) = \int_{\Theta^{k+1}} L(a, \underline{\theta}) f(\underline{x}|\underline{\theta}) dG(\underline{\theta}) \tag{4.5.4}$$

Knowing G, we can select according to the rule $t_G$, and incur the minimum Bayes risk $R(t_G, G)$ defined by (4.6.3). In many situations G is unknown and $t_G$ cannot be computed.

Suppose now that the above probelm occurs repeatedly and independently, with the same unknown G throughout. Thus we have

$$(\theta_{j1}, X_{j1}), \quad (\theta_{j2}, X_{j2}), \ldots$$

a sequence of pairs of random variables from $\pi_j$ $(j = 0, 1, \ldots, k)$ each pair being independent of all the other paris, where $x_{jn}$ given $\theta_{jn} = \theta_j$ follows the distribution $f(\cdot|\theta_j)$. A decision about $\underline{\theta}_n = (\theta_{0n}, \theta_{1n}, \ldots, \theta_{kn})$ is to be made after observing $\{\underline{x}_i = (x_{0i}, x_{1i}, \ldots, x_{ki}), \ i = 1, \ldots, n\}$. Following Robbins (1964) we will use some function $t_n(\underline{x}_1, \ldots, \underline{x}_n; \cdot)$ of $\underline{x}_1, \ldots, \underline{x}_n$ so that $t_n(\cdot)$ will be close to $t_G(\cdot)$ in the sense of Bayes risk. To make this more precise, we state a definition from Robbins (1964):

Definition 4.6.1: An empirical decision function $T = \{t_n\}$ is said to be asymptotically optimal (a.o) relative to G if

$$R_n(T,G) = \int_{\mathscr{X}} EH_G(t_n(\underline{x}), \underline{x}) d\mu(\underline{x}) \to R(t_G, G)$$

where $H_G$ is given by (4.6.4).

We wish to find a sequence $T = \{t_n\}$ of decision rules such that T is a.o. relative to every $G \in \mathcal{G}$. To this end, set

$$\begin{aligned} \Delta_G(a_j, \underline{x}) &= \int_{\Theta^{k+1}} [L(a_j, \underline{\theta}) - L(a_1, \underline{\theta}) h(\underline{x}|\underline{\theta}) dG(\underline{\theta}) \\ &= \int_{\Theta^{k+1}} [(\theta_j^2 - 2\theta_0\theta_j) - (\theta_1^2 - 2\theta_0\theta_1)] h(\underline{x}|\underline{\theta}) \, dG(\underline{\theta}) \qquad \text{4.5.5)} \end{aligned}$$

from (4.6.1)

Let $\Delta_{in}(\underline{x}) \equiv \Delta_{in}(\underline{x}_1,\ldots,\underline{x}_n;\underline{x})$ $(i = 1,\ldots,k;\ n \geq 1)$ be a sequence such that

$$\Delta_{in}(\underline{x}) \overset{p}{\to} \Delta_G(a_i,\underline{x})$$

Define

$$t_n(\underline{x}) = a_\ell, \text{ where } \ell\ (1 \leq \ell \leq k) \text{ is such that}$$

$$\Delta_{\ell n} = \min\{0,\Delta_{2n}(\underline{x}),\ldots,\Delta_{kn}(\underline{x})\} \qquad (4.5.6)$$

$$,\Delta_{1n} = 0$$

It follows from Corollary 1 of Robbins (1964) that the sequence $T = \{t_n\}$ is a.o. relative to G.

We will find the sequence $\Delta_{in}$ for some discrete distributions. Following Robbins (1955), let

$$f_G(x_j) = \int_\Theta f(x_j|\theta)\, dF(\theta),\ j = 0, 1,\ldots,k \qquad (4.5.7)$$

$$E[\theta_j^p|x_0, x_1,\ldots,x_k] = \frac{\int_\Theta \cdots \int_\Theta \theta_j^p \prod_{i=0}^{k} [f(x_i|\theta_i)dF(\theta_i)]}{\int_\Theta \cdots \int_\Theta \prod_{i=0}^{k} f(x_i|\theta_i)dF(\theta_i)}$$

$$= \frac{\int_\Theta \theta_j^p\, f(x_j|\theta_j)\, dF(\theta_j)}{f_G(x_j)} \qquad (4.5.8)$$

We consider several examples.

(i) Poisson populations

$$\text{Here } f(x|\theta) = \frac{e^{-\theta}\theta^{x}}{x!} \quad , \; x = 0, 1, \ldots, \; \theta > 0$$

Let $x_{j1}, \ldots, x_{jn}$ be a sample of size n from $\pi_j$, $j = 0, 1, \ldots, k$. Also let $\mathfrak{F}$ be the family of all a.c. distributions on the positive real line, and define a family of distributions $\mathcal{G}$ on $\mathbb{R}^{k+1}$ by

$$\mathcal{G} = \{G: \; G(\theta_0, \theta_1, \ldots, \theta_k) = \prod_{j=0}^{k} F(\theta_j), \; F \in \mathfrak{F}\} \quad (4.5.9)$$

From (4.6.8) we have, for p = 1, 2,

$$E(\theta_j^p | x_0, x_1, \ldots, x_k) = \frac{(x_j+p)!}{x_j!} \quad \frac{f_G(x_j+p)}{f_G(x_j)}$$

Let

$$p_{jn}(x) = \frac{\text{number of } X_{j1}, \ldots, X_{jn} = x}{n} \quad (4.5.10)$$

Set

$$h_{jn}^{(p)}(\underset{\sim}{x}) = \frac{(x_j+p)!}{x_j!} \quad \frac{p_{jn}(x_j+p)}{p_{jn}(x_j)} \quad , \; j = 0, 1, \ldots, k \quad (4.5.11)$$

Since the empirical frequence function $p_{jn}(x)$ converges in probability to the (unknown) marginal density $f_G(\cdot)$, we have

$$h_{jn}^{(p)}(\underset{\sim}{x}) \xrightarrow[n\to\infty]{P} E(\theta_j^p | \underset{\sim}{x}) \qquad \forall \; G \in \mathcal{G} \quad (4.5.12)$$

It follows that the sequence $\Delta_{jn}$ defined by

$$\Delta_{jn}(\underset{-}{x}_1,\ldots,\underset{-}{x}_n;\underset{-}{x}) = [h_{jn}^{(2)}(\underset{-}{x}) - 2h_{0n}^{(1)}(\underset{-}{x})h_{jn}^{(1)}(\underset{-}{x})] - [h_{1n}^{(2)}(\underset{-}{x}) - 2h_{0n}^{(1)}(\underset{-}{x})h_{1n}^{(1)}(\underset{-}{x})] \tag{4.5.13}$$

coverges in probability to $\Delta_G(a_j,\underset{-}{x})$ given by (4.6.5).

(ii) Geometric Distribution

Let $x_{j1},\ldots,x_{jn}$ be a sample size n from $\pi_j$ which has a geometric distribution given by

$$f(x|\theta_j) = (1 - \theta_j)\theta_j^x,\ x = 0,\ 1,\ 2,\ldots,\ 0 < \theta_j < 1,\ j = 0,1,\ldots,k.$$

Here $\mathfrak{F}$ is the family of a.c. distributions on (0,1) and $\mathcal{G}$ is defined by (4.6.9).

From (4.6.7) and (4.6.8) we have

$$E(\theta_j^p|\underset{-}{x}) = \frac{f_G(x_j+p)}{f_G(x_j)},\quad p = 1,2$$

In this case we define

$$h_{jn}^{(p)}(\underset{-}{x}) = \frac{\text{number of } X_{j1},\ldots,X_{jn} \text{ which equal } x_j+p}{\text{number of } X_{j1},\ldots,X_{jn} \text{ which equal } x_j}$$

and $\Delta_{jn}(\underset{-}{x}_1,\ldots,\underset{-}{x}_n;\underset{-}{x})$ by (4.6.13).

(iii) Binomial populations

Here $f(x|\theta) = \binom{r}{x}\ \theta^x(1-\theta)^{r-x}$, $x = 0,\ 1,\ldots,r,\ 0 < \theta < 1$

The family $\mathcal{G}$ is same as in example (ii) discussed above.

In this case, no asymptotically optimal solution seems to exist. Following Robbins (1964) we propose a rule which will do about as well as the (unknown) Bayes rule $t_G$ for large samples.

It is easily seen that

$$E(\theta_j^p|\underline{x}) = \frac{\binom{r}{x_j} f_{G,r+p}(x_j+p)}{\binom{r+p}{x_j+p} f_{G,r}(x_j)} \qquad x_j = 0, 1, \ldots, r \ , \ ( r > 2 ) \quad p = 1, 2$$

$$= \varphi_{G,r}^{(p)}(x_j) \text{ ,say}$$

where

$$f_{G,r}(y) = \int_0^1 \binom{r}{y} \theta^y (1-\theta)^{r-y} \, dG(\theta)$$

Let $x_{j1}^{(p)}, \ldots, x_{jn}^{(p)}, \ldots$ denote the sequence of number of successes in 1st r-p out of the r trials (which provided $x_{j1}, \ldots, x_{jn}, \ldots$), with $X_{ji}^{(0)} \equiv X_{ji}$.

Define

$$H_{jn}^{(p)}(\underline{x}) = \frac{\binom{r-p}{x_j} h_{jn}^{(r)}(x_j+p)}{\binom{r}{x_j+p} h_{jn}^{(r-p)}(x_j)}$$

where $$h_{jn}^{(r-p)}(\underline{x}) = \frac{\text{number of } x_{j1}^{(p)}, \ldots, x_{jn}^{(p)} \text{ which equal } x_j}{n}$$

Since $h_{jn}^{(r-p)}(\underline{x}) \overset{p}{\to} f_{G,r-p}(x_j)$ as $n \to \infty$, we have

$$H_{jn}^{(p)}(\underline{x}) \overset{p}{\to} \frac{\binom{r-p}{x_j} f_{G,r}(x_j+p)}{\binom{r}{x_j+p} f_{G,r-p}(x_j)} = \varphi_{G,r-p}^{(p)}(x_j)$$

It follows that the sequence $\Delta_{jn}$ defined below converges in probability to $\Delta_G(a_j,\underline{x})$:

$$\Delta_{jn}(\underline{x}_1,\ldots,\underline{x}_n;\underline{x}) = [H_{jn}^{(2)}(\underline{x}) - 2H_{jn}(\underline{x})\, H_{0n}(\underline{x})] - [H_{1n}^{(2)}(\underline{x}) - H_{1n}(\underline{x})H_{0n}(\underline{x})] \tag{4.5.14}$$

For large r, throwing away two observations does not sacrifice much information, but it is not at all clear that the sequence $\Delta_{jn}$ defined by (4.6.14) is the best sequence.

# BIBLIOGRAPHY

Bahadur, R. R. (1950). On the problem in the theory of k populations. Ann. Math. Statist. 21: 362-375.

Bechhofer, R. E. (1954). A single-sample multiple decision procedure for ranking means of normal populations with known variances. Ann. Math. Statist. 25: 16-39.

Berger, R. (1977). Minimax, Admissible and Gamma-Minimax Multiple Decision Rules. Ph.D. thesis. Department of Statistics, Purdue University, West Lafayette, IN 47907.

Bhattacharya, S. K. (1967). Bayesian approach to life testing and reliability estimation. J. Amer. Statist. Assoc. 62: 48-62.

Bickel, P. J. and Yahav, J. A. (1977). On selecting a set of good populations. Statistical Decision Theory and Related Topics II. (ed. Shanti S. Gupta and David S. Moore), 37-55. Academic Press, NY.

Blackwell, D. and Girshick, M. A. (1954). Theory of Games and Statistical Decisions. John Wiley, NY.

Carroll, R. J., Gupta, S. S., and Huang, D. Y. (1975). On selection procedures for the t best populations and some related problems. Comm. Statist. 4: 987-1008.

Chernoff, H. and Yahav, J. A. (1977). A subset selection problem employing a new criterion. Statistical Decision Theory and Related Topics II. (ed. Shanti S. Gupta and David S. Moore), 93-119. Academic Press, NY.

Chu, J. T. and Hotelling, H. (1955).The moments of the sample median. Ann. Math. Statistist. 26: 593-606.

Cramer, H. (1946). Mathematical Methods of Statistics. Princeton University Press, p. 369.

Deely, J. J. (1965). Multiple Decision Procedures from an Empirical Bayes Approach. Ph.D. thesis ( Mimeo Series No. 45 ).Department of Statistics, Purdue University, W. Lafayette, IN 47907.

Deely, J. J. and Gupta, S. S. (1968). On the properties of subset selection procedures. Sankhyā Ser. A. 30: 37-50.

Doornbos, R. (1966). Slippage Tests. Mathematical Research Center, Amsterdam.

# BIBLIOGRAPHY

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

Erdelyi, A. et al (1953). Higher Transcendental Functions II. McGraw-Hill, NY.

Ferguson, T. S. (1967). Mathematical Statistics - A Decision Theoretic Approach. Academic Press, NY.

Goel, P, K., and Rubin, H. (1975). On selecting a subset containing the best population - A Bayesian approach. Mimeo. Ser. No. 432, Department of Statistics, Purdue University, West Lafayette, IN 47907.

Gupta, S. S. (1956). On a decision rule for a problem in ranking means. Mimeo. Ser. No. 150. Inst. of Statist. University of North Carolina, Chapel Hill, NC .

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

Gupta, S. S. and Huang, D. Y. (1976). Subset selection procedures for the entropy function associated with the binomial populations. Sankhyā, Ser. A. 38.

Gupta, S. S. and Huang, D. Y. (1974). Nonparametric subset selection procedures for the t best populations. Bull. Inst. Math. Acad. Sinica. 2: 377-386.

Gupta, S. S. and Leong, Y. K. (1976). Some results on subset selection procedures for double exponential populations. Proceedings on the Workshop on Decision Information For Tactical Command and Control (ed. R. M. Thrall, C. P. Tsokos, J. C. Turner). Robert M. Thrall and Associates, Houston, TX.

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.

Guttman, I. (1961). Best populations and tolerance regions. Ann. Inst. Stat. Math. 13: 9-25.

Guttman, I. and Tiao, G. (1964). A Bayesian approach to some best population problems. Ann. Math. Statist. 35: 825-835.

Hall, I. J., and Kudo, A. (1968). On slippage tests - I - a generalization of Neyman-Pearson's lemma. Ann. Math. Statist. 39: 1693-1699.

Hall, I. J., Kudo, A. and Yeh, N. C. (1968). On slippage tests - II - similar slippage tests. Ann. Math. Statist. 39: 2029-2037.

Hodges, J. L. and Lehmann, E. L. (1963). Estimates of location based on rank tests. Ann. Math. Statist. 34: 598-611.

Hsu. J. (1977). On some decision theoretic contributions to the problem of subset selection. Ph.D. Thesis. Department of Statistics, Purdue University, West Lafayette, IN 47907.

Huang, W. T. (1975). Bayes approach to a problem of partioning k normal populations. Bull. Inst. Math. Acad. Sinica. 3: 87-97.

Karlin, S. (1968). Total Positivity. Stanford University Press. p. 155.

Karlin, S. and Truax, D. (1960). Slippage problems. Ann. Math. Statist. 31: 296-324.

Kudo, A. and Yeh, N. C. (1970). On slippage tests - III - least favorable distribution. Tamkang J. of Maths. 1: 57-63.

Lehmann, E. L. (1953). The power of rank tests. Ann. Math. Statist. 24: 23-43.

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

Lehmann, E. L. (1959). Testing Statistical Hypotheses. John Wiley, NY.

Mc Donald, G. C. (1969). On Some Distribution-free Ranking and Selection Procedures. Ph.D. thesis ( Mimeo Series No. 174 ). Department of Statistics, Purdue University, W. Lafayette, IN47907.

Mosteller, F. (1948). A k-sample slippage test for an extreme population. Ann. Math. Statistist. 19: 58-65.

Pearson, K. (1934). Tables of the Incomplete Γ - function. Cambridge University Press, England.

Paulson, E. (1949). A multiple decision procedure for certain problems in the analysis of variance. Ann. Math. Statist. 20: 95-98.

Raiffa, H. and Schlaifer, R. (1961). Applied Statistical Decision Theory. Harvard Business School.

Robbins, H. (1955). An empirical Bayes approach to statistics. Proc. 3rd Berkeley Symp. Math. Stat. Prob. University of California Press. 155-163.

Robbins, H. (1964). The empirical Bayes approach to statistical decision problems. Ann. Math. Statist. 35: 1-19.

Rohtagi, V. K. (1976). An Introduction to Probability Theory and Mathematical Statistics. John Wiley, p. 581-583.

Santner, T. J. (1975). A restricted subset selection approach to ranking and selection problems. Ann. Statist. 3: 334-349.

Savage, I. R. (1956). Contributions to the thoery of rank order statistics - the two-sample case. Ann. Math. Statist. 27:590-615.

Sobel, M. and Starr, N. (1975). Selecting the coin with the greatest bias. Sankhyā Ser. A. 37: 197-210.

Van Ryzin, J. (1970). Empirical Bayes slippages tests. Tech. Report No. 258, University of Wisconsin.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

# REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER
Mimeograph Series #494

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)
ON SLIPPAGE TESTS AND MULTIPLE DECISION (SELECTION AND RANKING) PROCEDURES

5. TYPE OF REPORT & PERIOD COVERED
Technical

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)
Ashok K. Singh

8. CONTRACT OR GRANT NUMBER(s)
N00014-75-C-0455

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Department of Statistics
Purdue University
West Lafayette, IN 47907

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
NR 042-243

11. CONTROLLING OFFICE NAME AND ADDRESS
Office of Naval Research
Washington, DC

12. REPORT DATE
June 1977

13. NUMBER OF PAGES
134

14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)

15. SECURITY CLASS. (of this report)
Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Slippage test, subset selection, decision theoretic, invariant procedure, minimal essentially complete class, Bayes procedure, property M, nonparametric slippage test, monotone likelihood ratio, location parameter, scale parameter, sample median, asymptotic relative efficiency, contamination, Gauss-Hermite quadrature.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Suppose k independent populations are given and we are interested in t $(1 \leq t < k)$ of these populations which are 'better' than the rest. Two different formulations, namely testing t slippages, and selection of a subset (of a random size or a fixed size t) containing the t 'better' populations, are considered herein. Both of these formulations provide answers to certain relevant questions which are not answered by the classical tests of homogeneity.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601 |

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

A t-slippage problem can roughly be described as follows: Test the homogeneity of the given $k(\geq 2)$ populations against $\binom{k}{t}$ alternative hypotheses that t of these populations are different from the rest, that is to say, these have slipped. In Chapter I the form of symmetric invariant Bayes procedures has been derived for the t-slippage problem, when the underlying densities are of location or scale type, and have a sufficient maximum likelihood estimate of the parameter in case the populations are homogeneous. Locally best tests of some t-slippage problems in nonparametric situations have also been derived.

In Chapter II the problem of selecting a subset of a random size which contains the populations associated with t largest parameters is investigated. It is shown that for a loss function which is linear in the parameters the Bayes rule for any symmetric a priori distribution selects exactly t populations, provided assumptions of symmetry and monotonicity on the density function are satisfied. For some non-linear loss functions, Bayes rules have been investigated for normal, exponential, Poisson and binomial populations, when the a priori distributions are known.

Chapter III deals with the problem of selecting a subset containing the normal population associated with the largest mean. Gupta-type rules based on sample medians have been investigated. Numerical comparisons are carried out in the case of equally spaced normal means which show that the selection rule based on the sample mean is 'better' than the proposed rule based on sample medians. However, if the normal populations are contaminated, the proposed rule based on sample medians is superior to the rule based on sample means in terms of the asymptotic relative efficiency.

Chapter IV consists of investigations of procedures for the following problems: (i) To select a subset of random size which contains all populations 'close' to a given control population, and (ii) to select the t populations closest to control. Bayes and Gupta type-rules have been investigated.