Some Results in the Theory of Subset Selection Procedures

by

Hwa-Ming Yang
Purdue University

Department of Statistics
Division of Mathematical Sciences
Mimeograph Series #80-22

August, 1980

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

INTRODUCTION

Selection and ranking (ordering) problems in statistical inference arise mainly because the classical tests of homogeneity are often inadequate in certain situations where the experimenter is interested in comparing k (≥ 2) populations, treatments or processes with the goal of selecting one or more worthwhile (good) populations. Mosteller (1948), Paulson (1949), Bahadur (1950) and Bahadur and Robbins (1950) were among the earliest research workers to recognize this inadequacy and to formulate the problem as a multiple decision problem aimed at the selection and ranking of the k populations.

In the thirty years since these early papers, selection and ranking problems have become an active area of statistical research. There have been two approaches to these problems, the 'indifference zone' approach and the 'subset selection' approach. In the first approach, due to Bechhofer (1954), the experimenter wishes to select one population (or a fixed number $t \geq 1$ of population) which is guaranteed to be the one of interest to him with a fixed probability P* whenever the unknown parameters lie outside some subspace of the parameter space, the so-called indifference zone. Important contributions using this approach have been made by Bechhofer and Sobel (1954), Bechhofer, Dunnett and Sobel (1954), Sobel and Huyett (1957), Sobel (1967), Bechhofer, Kiefer and Sobel (1968), Mahamunulu (1967), Desu and Sobel (1968, 1971) and

Tamhane and Bechhofer (1977, 1979) among others. A quite complete bibliography may be found in Gupta and Panchapakesan (1979) (see also Gibbons, Olkin and Sobel (1977)).

The second approach pioneered by Gupta (1956, 1963, 1965) assumes no a priori information about the parameter space. A single population is not necessarily chosen; rather a subset of the given k populations is selected depending on the outcome of the experiment. It is guaranteed to contain the population(s) of interest with probability which is at least equal to P* (the basic probability requirement) regardless of the true unknown configurations of the parameters. Some recent contributors in the category of subset selection include: Deely (1965), Gnanadesikan (1966), Gnanadesikan and Gupta (1970), Gupta (1967), Gupta and Studden (1970), Nagel (1970), Gupta and Nagel (1971), Gupta and Panchapakesan (1972), Rizvi and Sobel (1967), McDonald (1969), Gupta and McDonald (1970), Santner (1975), W. T. Huang (1972), D. Y. Huang (1975), Gupta and Huang (1975a, 1975b) and Gupta and Huang (1976).

Subset selection procedures can also be thought of as screening procedures which enable the experimenter to select a subset of populations (under study) which contains the populations of interest so that the populations in the selected subset can be further studies.

Sequential and multistage aspects of the ranking and selection problems, have been explored, based on the indifference zone approach by Bechhofer, Dunnett and Sobel (1954), Bechhofer (1958), Paulson (1962, 1963, 1964, 1967) and Bechhofer, Kiefer and Sobel (1968). Barson and Gupta (1972), Huang (1972), Gupta and Huang (1975). Gupta and Miescke (1979) and Carroll (1974) have investigated subset selection procedures, based on sequential sampling.

Contributions to optimum properties of subset selection procedures have been made by Lehmann (1961), Studden (1967), Deely and Gupta (1968), Berger (1977, 1979), Gupta and Hsu (1978), Gupta and Miescke (1978), Berger and Gupta (1980).

In the decision-theoretic approach to the subset selection problems, Goel and Rubin (1977), Chernoff and Yahav (1977), Bickel and Yahav (1977), Gupta and Hsu (1978), Miescke (1979), Gupta and Kim (1980), Gupta and Hsiao (1980) have given different formulations under different loss functions and carried out investigations which indicate that the Gupta-type maximum (minimum) means procedures are quite 'optimal' and 'robust'.

The main purpose of this thesis is to study some problems using the subset selection approach and provide procedures and results for some unsolved problems.

Chapter I considers the problem of selecting a subset containing all populations better than a control under an ordering prior. Here, by an ordering prior we mean that there exists a known simple or partial order relationship among the unknown parameters of the treatmetns (excluding the control). Three new selection procedures are proposed and studied. These procedures do meet the usual requirement that the probability of a correct selection is greater than or equal to a pre-determined number P*. Two of the three procedures use the isotonic regression over the sample means of the k treatments with respect to (wrt) the given ordering prior. Tables which are necessary to carry out the selection procedures with isotonic approach for the selection of unknown means of normal populations and gamma populations are given. Monte Carlo comparisons on the performance of several procedures for the normal or gamma means problem were carried out in

several selected cases; these are given in Table V and Table VI at the end of Chapter I. In each case ten thousand simulations were performed. The results of this study seem to indicate that the procedures based on isotonic estimators always have superior performance, expecially, when there are more than one bad populations (in comparison with the control).

Chapter II deals with a new 'Bayes-P^*' approach about the problem of selecting a subset which contains the 'best' of k populations. Here, by best we mean the (unknown) population with the largest unknown mean. The (non-randomized) Bayes-P^* rule refers to a rule with minimum risk in the class of (non-randomized) rules which satisfy the condition that the posterior probability of selecting the best is at least equal to P^*. Given the priors of the unknown parameters, two 'Bayes-P^*' subset selection procedures ψ^B and ψ^B_{NR} (randomized and non-randomized, respectively) under certain loss functions are obtained and compared with the classical maximum-type means procedure ψ^M. The comparisons of the performance of ψ^B with ψ^B_{NR} and ψ^M, based on Monte Carlo studies, indicate that the procedure ψ^B always has higher 'efficiency' and smaller expected selected size of the selected subset. Also ψ^B appears to be robust when the true distributions are not normal but are some other symmetric distributions such as, the logistic, the double exponential, Laplace, and the gross error model (the contaminated distribution).

CHAPTER I

SELECTION PROCEDURES FOR POPULATIONS BETTER THAN A CONTROL UNDER ORDERING PRIOR

1.1. Introduction

In this chapter, three new selection procedures are given for the problem of selecting a subset which contains all populations better than a standard or control under simple or partial ordering prior. Here by simple or partial ordering prior we mean that there exist known simple or partial order relationships (defined more specifically later in Section 1.2) among unknown parameters. The procedures described do meet the usual requirement that the probabilities of a correct selection are greater than or equal to a predetermined number P*, the so-called P* condition.

Many authors have considered the problem of comparing populations with a control under different types of formulations (see Gupta and Panchapakesan (1979)). Dunnett (1955) considered the problem of separating those treatments which are better than the control from those that are worse. Gupta and Sobel (1958), Gupta (1965), Naik (1975), Broström (1977) studied the problem of selecting a subset containing all populations better than the control. Lehmann (1961) discussed similar problems with emphasis on the derivation of a restricted minimax procedure. Kim (1979), Hsiao (1979) studied the problem of

selecting populations close to a control. In all these papers it is assumed that all populations are independent and that there is no information about the order of unknown parameters. However, in many situations, we may know something about the unknown parameters. What we know is always not the prior distributions but some partial or incomplete prior information, such as the simple or partial order relationship among the unknown parameters. This type of information about the ordering prior may come from the past experiences; or it may arise in the experiments where, for example, higher dose level of some drugs always has larger effect (side-effect) on the patients.

In Section 1.2 definitions and notations used in this chapter are introduced. In Section 1.3 we consider the problem for location parameters. We propose three types of selection procedures for the cases when the control parameter is known or not known (the scale parameter may or may not be assumed known). Some equivalent forms of the procedures are given, and their properties are discussed. In Section 1.4 the problem for scale parameters of the gamma distributions is considered and three analogous selection procedures are proposed. In both Section 1.3 and 1.4 simple ordering priors are assumed and some theorems in the theory of random walks are used. In Section 1.5 a selection procedure is given for the problem of selecting all populations better than the control under partial ordering prior. Section 1.6 deals with the use of Monte Carlo techniques to make comparisons among the selection procedures proposed in Section 1.3 and those in Section 1.4, respectively.

1.2. Notations and Definitions

Suppose we have k + 1 populations $\pi_0, \pi_1, \ldots \pi_k$. The population treatment π_0 is called the control or standard population. Assume that the random variables X_{ij} associated with $F(\cdot;\theta_i)$ and $X_{i1},\ldots,X_{in_i}$, $i = 1,\ldots,k$, is an independent sample from π_i. Assume that we have an ordering prior of $\theta_1,\ldots,\theta_k$. First we assume that the ordering prior is the simple order, so that without loss of generality, we may assume that, $\theta_1 \leq \ldots \leq \theta_k$. In Section 1.5 we will consider the partial ordering prior case. Note that the values of θ_i's are unknown.

Suppose our goal is to find a non-trivial (small) subset which contains all populations with parameter larger (smaller) than the control θ_0 (known or unknown) with probability not less than a given value P*.

The action space $\mathcal{A}$ is the class of all subsets of set {1, 2,...,k}. An action A is the selection of some subset of the k populations. $i \in A$ means that π_i is included in the selected subset.

Let $\underline{\theta} = (\theta_0, \theta_1,\ldots,\theta_k)$. Then the parameter space is denoted by Ω, where $\Omega = \{\underline{\theta} \in \mathbb{R}^{k+1} \mid \theta_1 \leq \theta_2 \leq \ldots \leq \theta_k; -\infty < \theta_0 < \infty\}$ is a subset of k + 1 dimensional Euclidean space $\mathbb{R}^{k+1}$.

The sample space is denoted by $\mathcal{X}$ where

$$\mathcal{X} = \{\underline{x} \in \mathbb{R}^{n_1+\ldots+n_k} \mid \underline{x} = (x_{11},\ldots x_{1n_1}, x_{21},\ldots,x_{k1},\ldots x_{kn_k})\}.$$

Definition 1.2.1. A (non-randomized) selection procedure (rule) $\delta(x)$ is a mapping from $\mathcal{X}$ to $\mathcal{A}$.

A population π_i $(i = 1,\ldots,k)$ is called a good population if $\theta_i \geq \theta_0$, and we say a selection procedure δ make a correct selection (CS) if the selected subset contains all good populations. A selection procedure δ satisfies the P*-condition if

$$P_{\underline{\theta}}(CS|\delta) \geq P^* \quad \text{for all} \quad \underline{\theta} \in \Omega$$

that is

$$\inf_{\underline{\theta}\in\Omega} P_{\underline{\theta}}(CS|\delta) \geq P^*. \tag{1.2.1}$$

Let $\mathcal{D} = \{\delta \,|\, \inf_{\underline{\theta}\in\Omega} P_{\underline{\theta}}(CS|\delta) \geq P^*\}$ be a collection of all selection procedures satisfying the P*-condition.

In the sequel we will use the isotonic estimators (see Barlow, Bartholomew, Bremner and Brunk (1972)). Hence we give the following definitions and theorems.

Definition 1.2.2. Let the set $\mathcal{T}$ be a finite set. A binary relation "$\lesssim$" on $\mathcal{T}$ is called a simple order if it is

(1) reflexive: $x \lesssim x$ for $x \in \mathcal{T}$

(2) transitive: $x, y, z \in \mathcal{T}$ and $x \lesssim y$, $y \lesssim z$ imply $x \lesssim z$

(3) antisymmetric: $x, y \in \mathcal{T}$ and $x \lesssim y$, $y \lesssim x$ imply $x = y$

(4) every two elements are comparable: $x, y \in \mathcal{T}$ imply either $x \lesssim y$ or $y \lesssim x$.

A partial order on $\mathcal{T}$ is a binary relation "$\lesssim$" on $\mathcal{T}$, such that it is (1) reflexive, (2) transitive, and (3) antisymmetric. Thus every simple order is a partial order. We use poset $(\mathcal{T},\lesssim)$ to denote the set $\mathcal{T}$ that has a partial order binary relation "$\lesssim$" on it.

Definition 1.2.3. A real-valued function f is called isotonic on poset $(\mathcal{I},\leq)$ if and only if (1) f is defined on $\mathcal{I}$, (2) if $x, y \in \mathcal{I}$, $x \leq y$ imply $f(x) \leq f(y)$.

Definition 1.2.4. Let g be a real-valued function on $\mathcal{I}$ and let W be a given positive function on $\mathcal{I}$. A function g* on $\mathcal{I}$ is called an isotonic regression of g with weights W if and only if:

(1) g* is an isotonic function on poset $(\mathcal{I},\leq)$

(2) $$\sum_{x\in\mathcal{I}} [g(x) - g^*(x)]^2 W(x) = \min_{f\in\mathfrak{F}} \sum_{x\in\mathcal{I}} [g(x) - f(x)]^2 W(x),$$

where $\mathfrak{F}$ is the class of all isotonic functions on poset $(\mathcal{I},\leq)$.

From Barlow, et. al. (1972), (see their Theorems 1.3, 1.6 and the corollary there), we have the following theorems.

Theorem 1.2.1. There exists one and only one isotonic regression g* of g with weight W on poset $(\mathcal{I},\leq)$.

Definition 1.2.5. A set S is convex if s_1 and $s_2 \in S$ and $0 \leq \alpha \leq 1$ then $\alpha s_1 + (1-\alpha)s_2 \in S$.

Definition 1.2.6. A set S is a cone if $s \in S$ then for any non-negative real number c, $cs \in S$.

Definition 1.2.7. A poset $(\mathcal{I},\leq)$ is a lattice if sup H and inf H exist for any finite non-empty subset H of $\mathcal{I}$.

If f and g are two isotonic functions on poset $(\mathcal{I},\leq)$, we define $f \wedge g$ and $f \vee g$ as

$$(f \wedge g)(t) = f(t) \wedge g(t) \equiv \min(f(t), g(t))$$

and

$$(f \vee g)(t) = f(t) \vee g(t) \equiv \max(f(t), g(t)).$$

Then we state the following:

Theorem 1.2.2. The class $\mathfrak{F}$ of all isotonic functions on poset $(\mathcal{T},\leq)$ is a convex cone and a lattice.

There are some algorithms, such as the "pool-adjacent-violation" algorithm (see page 13 of Barlow, et. al. (1972)) or Ayer, Brunk, Ewing, Reid and Silverman (1955) or the "up-and-down blocks" algorithm, Kruskal (1964), which show how to calculate the isotonic regression under simple order.

The following max-min formulas were given by Ayer et. al. (1955).

Theorem 1.2.3. (max-min formulas)

Assume that we have poset $(\mathcal{T};\leq)$ where $\mathcal{T} = \{\theta_1,\ldots,\theta_k\}$, $\theta_1 \leq \ldots \leq \theta_k$, and that function $g: \mathcal{T} \to \mathbb{R}$, then the isotonic regression g^* of g with weight W has the following formulas:

$$\begin{aligned} g^*(\theta_i) &= \max_{s\leq i} \min_{t\geq i} Av(s,t) \\ &= \max_{s\leq i} \min_{t\geq s} Av(s,t) \\ &= \min_{t\geq i} \max_{s\leq i} Av(s,t) \\ &= \min_{t\geq i} \max_{s\leq t} Av(s,t) \end{aligned}$$

where

$$Av(s,t) = \frac{\sum_{r=s}^{t} g(\theta_r)W(\theta_r)}{\sum_{r=s}^{t} W(\theta_r)}.$$

Corollary 1.2.1. $(g + c)^* = g^* + c$,

$(ag)^* = ag^*$, if $a > 0$.

Corollary 1.2.2. $[\rho(g^*)g + \varphi(g^*)]^* = \rho(g^*)g^* + \varphi(g^*)$, where ρ is a non-negative function and φ is an arbitrary function.

1.3. Proposed Selection Procedures for the Location Parameter Problem

To discuss some more general results, we assume that population π_i has an absolutely continuous location-scale distribution function $F(x; \mu_i, \sigma^2) = F(\frac{x-\mu_i}{\sigma})$, where $0 < F(x) < 1$ for all $x, -\infty < x < \infty$ and the values of $\mu_1, \ldots,$ and μ_k are unknown, but their ordering, say, $\mu_1 \leq \ldots \leq \mu_k$ is known. Note that in this case we replace $\underline{0}$ in the parameter space Ω by $\underline{\mu}$, all other quantities remaining the same.

Let us define the subspace $\Omega_i = \{\underline{\mu} \in \Omega | \ \mu_{k-i} < \mu_0 \leq \mu_{k-i+1}\}$ for $i = 1, \ldots, k-1$ and let subspace $\Omega_k = \{\underline{\mu} \in \Omega | \ \mu_0 \leq \mu_1\}$ and subspace $\Omega_0 = \{\underline{\mu} \in \Omega | \ \mu_k < \mu_0\}$, then we have $\Omega = \bigcup_{i=0}^{k} \Omega_i$. Note that the control μ_0 could be known or unknown. If μ_0 is unknown, we assume that the distribution of population π_0 is $F(x; \mu_0, \sigma^2)$ and we take independent observations $X_{01}, \ldots, X_{0n}$ from π_0 and the sample space $\mathcal{X}$ turns to $\{\underline{x} \in \mathbb{R}^{n_0+\ldots+n_k} | \ \underline{x} = (X_{01}, \ldots, X_{0n_0}, X_{21}, \ldots, X_{kn_k})\}$. Using the partition $\{\Omega_0, \ldots, \Omega_k\}$ of parameter space Ω, we have

$$\inf_{\underline{\mu} \in \Omega} P_{\underline{\mu}}(CS|\delta) = \inf_{1 \leq i \leq k} \{\inf_{\underline{\mu} \in \Omega_i} P_{\underline{\mu}}(CS|\delta)\},$$

for any selection procedure $\delta \in \mathcal{S}$. Hence the P*-condition is equivalent to

$$\inf_{\mu\in\Omega_i} P_{\underline{\mu}}(CS|\delta) \geq P^*, \quad \text{for} \quad i = 1,\ldots,k.$$

Note that $\inf_{\mu\in\Omega_0} P_{\underline{\mu}}(CS|\delta) = 1$ for any selection procedure δ since there exist no good population in this case.

Suppose $X_i = x_i$ is the outcome of the sample mean of population π_i, $i = 1,\ldots,k$. Let $\mathcal{T}$ denote the set $\{\mu_1, \mu_2,\ldots,\mu_k\}$ where $\mu_1 \leq \ldots \leq \mu_k$, and let $W(\mu_i) = n_i\sigma^{-2} \equiv w_i$, $g(\mu_i) = x_i$, $i = 1,\ldots,k$. Then by the max-min formulas, the isotonic regression of g is g*, where

$$g^*(\mu_i) = \max_{1\leq s\leq i} \min_{s\leq t\leq k} \frac{\sum_{j=s}^{t} x_j w_j}{\sum_{j=s}^{t} w_j}, \quad i = 1,\ldots,k.$$

The isotonic estimator of μ_i is denoted by $\hat{X}_{i:k}$, $i = 1,\ldots,k$ where

$$\hat{X}_{i:k} = \max_{1\leq s\leq i} \min_{s\leq t\leq k} \frac{\sum_{j=s}^{t} X_j w_j}{\sum_{j=s}^{t} w_j}$$

$$= \max_{1\leq j\leq i} \{\hat{\hat{X}}_{j:k}\} \tag{1.3.1}$$

where

$$\hat{\hat{X}}_{j:k} = \min \{X_j, \frac{X_j w_j + X_{j+1} w_{j+1}}{w_j + w_{j+1}}, \ldots, \frac{X_j w_j + \ldots + X_k w_k}{w_j + \ldots + w_k}\}. \tag{1.3.2}$$

1.3.1. Proposed Selection Procedure δ_1

Case I. μ_0 known, common variance σ^2 known, and common sample size n.

Definition 1.3.1. We define the procedure δ_1 as follows:

Step 1. Select π_i, $i = 1,\ldots,k$ and stop, if

$$\hat{X}_{1:k} \geq \mu_0 - d^{(1)}_{1:k} \frac{\sigma}{\sqrt{n}},$$

otherwise reject π_1 and go to step 2.

Step 2. Select π_i, $i = 2,\ldots,k$ and stop, if

$$\hat{X}_{2:k} \geq \mu_0 - d^{(1)}_{2:k} \frac{\sigma}{\sqrt{n}},$$

otherwise reject π_2 and go to step 3.

.
.
.

Step k-1. Select π_i, $i = k-1, k$ and stop, if

$$\hat{X}_{k-1:k} \geq \mu_0 - d^{(1)}_{k-1:k} \frac{\sigma}{\sqrt{n}},$$

otherwise reject π_{k-1} and go to step k.

Step K. Select π_k and stop, if

$$\hat{X}_{k:k} \geq \mu_0 - d^{(1)}_{k:k} \frac{\sigma}{\sqrt{n}},$$

otherwise reject π_k.

Here $d^{(1)}_{i:k}$'s are the smallest values such that $\delta_1 \in \mathcal{D}$, that is δ_1 satisfies the p*-condition.

1.3.2. On the Evaluation of $\inf_{\underline{\mu}\in\Omega_i} P_{\underline{\mu}}(CS|\delta_1)$ and the Value of the Constants $d^{(1)}_{1:k},\ldots,d^{(1)}_{k:k}$

For any $\underline{\mu}\in\Omega_i$, $1 \le i \le k$, let Z_i's i.i.d. $\sim F(\cdot; 0, 1)$ then

$$P_{\underline{\mu}}(CS|\delta_1)$$

$$= P_{\underline{\mu}}\Big(\bigcup_{j=1}^{k-i+1}\{\hat{X}_{j:k} \ge \mu_0 - d^{(1)}_{j:k}\frac{\sigma}{\sqrt{n}}\}\Big)$$

$$= P_{\underline{\mu}}\Big(\bigcup_{j=1}^{k-i+1}\bigcup_{r=1}^{j}\{\hat{\hat{X}}_{r:k} \ge \mu_0 - d^{(1)}_{j:k}\frac{\sigma}{\sqrt{n}}\}\Big)$$

$$\ge P_{\underline{\mu}}\Big(\bigcup_{j=1}^{k-i+1}\bigcup_{r=1}^{j}\{\hat{\hat{Z}}_{r:k} + \frac{\mu_r-\mu_0}{\sigma/\sqrt{n}} \ge - d^{(1)}_{j:k}\}\Big)$$

which is decreasing in μ_r, $r = 1,\ldots,k-i+1$.

Hence

$$\inf_{\underline{\mu}\in\Omega_i} P_{\underline{\mu}}(CS|\delta_1) \ge P(\hat{\hat{Z}}_{k-i+1:k} \ge - d^{(1)}_{k-i+1:k})$$

On the other hand,

$$\inf_{\underline{\mu}\in\Omega_i} P_{\underline{\mu}}(CS|\delta_1)$$

$$\le P_{\underline{\mu}^*}\Big(\bigcup_{j=1}^{k-i+1}\{\hat{X}_{j:k} \ge \mu_0 - d^{(1)}_{j:k}\frac{\sigma}{\sqrt{n}}\}\Big)$$

$$= P(\hat{\hat{Z}}_{k-i+1:k} \ge - d^{(1)}_{k-i+1:k})$$

whenever $\underline{\mu}^* = (\mu_0, -\infty,\ldots, -\infty, \underbrace{\mu_0,\ldots,\mu_0}_{i}) \in \bar{\Omega}_i$

Thus, we have

$$\inf_{\underline{\mu}\in\Omega_i} P_{\underline{\mu}}(CS|\delta_1) = P(\hat{\hat{Z}}_{k-i+1:k} \ge - d^{(1)}_{k-i+1:k}).$$

Since $\hat{\hat{Z}}_{k-i+1:k} = \min \{Z_{k-i+1}, \ldots, \frac{Z_{k-i+1} + \ldots + Z_k}{i}\}$ has the same distributions as

$$\hat{\hat{Z}}_{1:i} = \min \{Z_1, \ldots, \frac{Z_1 + \ldots + Z_i}{i}\},$$

let
$$V_i = \hat{\hat{Z}}_{1:i} \tag{1.3.3}$$
$$= \min_{1 \le r \le i} \frac{1}{r} \sum_{j=1}^{r} Z_j,$$

we have

$$\inf_{\underline{\mu} \in \Omega_i} P_{\underline{\mu}}(CS|\delta_1) = P(V_i \ge -d^{(1)}_{k-i+1:k}), \quad i = 1, \ldots, k. \tag{1.3.4}$$

Theorem 1.3.1. In case I, (μ_0 known, common known σ^2 and common sample size n), if $d^{(1)}_{k-i+1:k}$ is the solution of equation

$$P(V_i \ge -x) = P^* \tag{1.3.5}$$

where

$$V_i = \min_{1 \le r \le i} \frac{1}{r} \sum_{j=1}^{r} Z_j \quad \text{and} \quad Z_i \text{ are i.i.d.} \sim F(\cdot),$$

$i = 1, \ldots, k$ then δ_1 satisfies the P*-condition.

Proof. For any i, $1 \le i \le k$,

$$\inf_{\underline{\mu} \in \Omega_i} P_{\underline{\mu}}(CS|\delta_1) = P(V_i \ge -d^{(1)}_{k-i+1:k}) = P^*,$$

so δ_1 satisfies the P*-condition.

Therefore, the problem of finding the $d^{(1)}_{i:k}$'s reduces to finding the distributions of $V_1, \ldots,$ and V_k. This is achieved by using some theorems in the theory of random walk.

1.3.3. Some Theorems in the Theory of Random Walk

Suppose $Y_1, Y_2, \ldots$ are independent random variables with a common distribution H not concentrated on a half-axis, i.e. $0 < P(Y_1 < 0)$, $P(Y_1 > 0) < 1$. The induced random walk is the sequence of random variables

$$S_0 = 0,\ S_n = Y_1 + \ldots + Y_n,\quad n = 1, 2, \ldots .$$

Let

$$\tau_n = P(S_1 \leq 0, \ldots, S_{n-1} \leq 0,\ S_n > 0) \tag{1.3.6}$$

and

$$\tau(s) = \sum_{n=1}^{\infty} \tau_n s^n,\quad 0 \leq s \leq 1. \tag{1.3.7}$$

Then we have the following theorem which was discovered by Andersen (1953). Feller (1971) gave an elegant short proof.

Theorem 1.3.2.

$$\log \frac{1}{1-\tau(s)} = \sum_{n=1}^{\infty} \frac{s^n}{n} P(S_n > 0). \tag{1.3.8}$$

Theorem 1.3.3. (Feller (1971))

Let

$$p_n = P(S_1 > 0, \ldots, S_n > 0),$$

then

$$p(s) \equiv \sum_{n=1}^{\infty} p_n s^n = \frac{1}{1-\tau(s)}, \tag{1.3.9}$$

hence

$$\log p(s) = \sum_{n=1}^{\infty} \frac{s^n}{n} P(S_n > 0). \tag{1.3.10}$$

By symmetry, the probabilities

$$q_n = P(S_1 \leq 0,\ldots,S_n \leq 0) \tag{1.3.11}$$

have the generating function q given by

$$\log q(s) = \sum_{n=1}^{\infty} \frac{s^n}{n} P(S_n \leq 0). \tag{1.3.12}$$

Note: The above two theorems remain valid if the signs $>$ and $\leq$ are replaced by $\geq$ and $<$, respectively.

Now, let

$$U_j' = \max_{1\leq r\leq j} \frac{1}{r} \sum_{i=1}^{r} Z_i', \quad j = 1,2,\ldots, \tag{1.3.13}$$

and

$$V_j' = \min_{1\leq r\leq j} \frac{1}{r} \sum_{i=1}^{r} Z_i', \quad j = 1,2,\ldots, \tag{1.3.14}$$

where Z_i''s are i.i.d. with absolutely continuous c.d.f. $G(\cdot)$. We would like to apply Theorem 1.3.3 to get the distribution of U_j' and V_j', $j = 1,2,\ldots$.

Remark 1.3.1. The distribution of U_j', $j = 1,\ldots,k$ for some $k \geq 1$, will be used whenever our goal is changed to select a subset containing no population with parameter smaller than the control.

Theorem 1.3.4. The generating function q(s) of $P(U_j' \leq x)$, $j = 1,2,\ldots$ is

$$\sum_{j=1}^{\infty} s^j P(U_j' \leq x) = \exp\left[\sum_{n=1}^{\infty} \frac{1}{n} s^n P(S_n \leq 0)\right] \tag{1.3.15}$$

where

$$S_n = \sum_{i=1}^{n} (Z_i' - x), \quad n = 1, 2, \ldots,$$

if the distribution of $Y_1 = Z_1' - x$ is not concentrated on a half-axis.

Proof. Since the distribution of random variable $Y_i = Z_i' - x$ is not concentrated on a half-axis, and Y_i's are i.i.d. let $S_r = \sum_{i=1}^{r} (Z_i' - x)$, $r = 1, \ldots, k$. Then

$$\{U_j' \leq x\} = \{\max_{1 \leq r \leq j} \frac{1}{r} S_r \leq 0\} = \{S_1 \leq 0, \ldots, S_j \leq 0\}.$$

By Feller's Theorem 1.3.3, we complete the proof.

Similarly, $\{V_j' \geq x\} = \{S_i \geq 0, \quad i = 1, 2, \ldots, j\}$,

where

$$S_i = \sum_{r=1}^{i} (Z_r' - x).$$

Theorem 1.3.5. The generating function $p(s)$ of $P(V_j' \geq x)$ is

$$\sum_{j=1}^{\infty} s^j P(V_j' \geq x) = \exp\{\sum_{n=1}^{\infty} \frac{1}{n} s^n P(\tilde{S}_n \geq 0)\}, \qquad (1.3.16)$$

if the distribution of $Y_1 = Z_1' - x$ is not concentrated on a half-axis.

Corollary 1.3.1. Both Theorem 1.3.4 and Theorem 1.3.5 hold for all x such that $0 < G(x) < 1$.

Proof. Let $Y_1 = Z_1' - x$, then

$$P(Y_1 \leq 0) = G(x)$$

and

$$0 < G(x) < 1,$$

hence Y_1 is not concentrated on a half-axis.

Corollary 1.3.2. Both Theorem 1.3.4 and Theorem 1.3.5 hold for all x whenever $G = \Phi$, c.d.f. of $N(0,1)$, or $G = F$ which is defined at the beginning of Section 1.3.

Proof. Followed immediately by Corollary 1.3.1.

Note that in the case of location parameter of normal population, $P(U_n' \leq -x) = P(V_n' \geq x)$.

Let

$$\Delta_j(x) \equiv \Delta_j = P(S_j \geq 0), \quad j = 1, 2, \ldots,$$

$$a(s) = \sum_{n=1}^{\infty} \frac{s^n}{n} \Delta_n,$$

we have

$$p(s) \equiv \sum_{j=1}^{\infty} s^j P(V_j' \geq x) = \exp(a(s)).$$

Lemma 1.3.1. $p^{(n+1)}(s) = \sum_{j=0}^{n} \binom{n}{j} p^{(j)}(s)\, a^{(n+1-j)}(s)$, $\forall n$.

Proof. Since $p'(s) = p(s) \cdot a'(s)$, the result can be proved by induction on n.

Theorem 1.3.6. Under the assumption of Theorem 1.3.5

$$P(V'_{n+1} \geq x) = \frac{1}{(n+1)!} \lim_{s\to 0^+} \frac{d^{n+1}p(s)}{ds^{n+1}}$$

$$= \frac{1}{n+1} \sum_{j=0}^{n} P(V'_j \geq x)\Delta_{n-j+1}, \quad n = 0, 1, 2,\ldots \qquad (1.3.17)$$

where

$$P(V'_0 \geq x) \equiv 1, \ \forall x.$$

Proof. By Lemma 1.3.1, we have

$$P(V'_{n+1} \geq x) = \frac{1}{(n+1)!} \lim_{s\to 0^+} p^{(n+1)}(s)$$

$$= \sum_{j=0}^{n} \frac{1}{(n+1)!} \frac{n!}{j!(n-j)!} p^{(j)}(0)\ [(n-j)!\ \Delta_{n+1-j}]$$

$$= \frac{1}{n+1} \sum_{j=0}^{n} \frac{p^{(j)}(0)}{j!} \Delta_{n+1-j}$$

$$= \frac{1}{n+1} \sum_{j=0}^{n} P(V'_j \geq x)\Delta_{n+1-j}.$$

Similarly, we have

$$P(U'_{n+1} \leq x) = \frac{1}{n+1} \sum_{i=0}^{n} P(U'_{n-i+1} \leq x)\ P(S_i \leq 0). \qquad (1.3.18)$$

1.3.4. Limiting Distributions of U'_n and V'_n

Let $F_n(x) = P(U'_n \leq x)$ and $F_\infty(x)$ denote the limiting distribution function as $n \to \infty$ of U'_n. Suppose the distribution of random variable $Y_1 = Z'_1 - x$ is not concentrated on a half axis, then we have

$$1 - F_n(x) = P(S_1 > 0) + \sum_{r=2}^{n} P(S_1 \le 0, \ldots, S_{r-1} \le 0, S_r > 0),$$

$$1 - F_\infty(x) = \lim_{s \to 1^-} \tau(s),$$

and apply Andersen-Feller Theorem 1.3.2, we have

$$F_\infty(x) = \exp\{- \sum_{r=1}^{\infty} \frac{1}{r} P(S_r > 0)\}. \tag{1.3.19}$$

Similarly,

$$G_\infty(x) = \exp\{- \sum_{r=1}^{\infty} \frac{1}{r} P(S_r \le 0)\} \tag{1.3.20}$$

where

$$G_\infty(x) = P(V'_\infty \ge x).$$

Let

$$G_\infty(- d^{(1)}_{1:\infty}) = P^*. \tag{1.3.21}$$

If Z_i, $i = 1,\ldots,k$, are independent identically distributed $\sim N(0,1)$, then we can use the recurrence formula of Theorem 1.3.6 to solve the equations $P(V_i \ge - d_{k-i+1:k}) = P^*$, $i = 1,\ldots,k$. Hence in Case I, $\Delta_j(x) = \Phi(- x\sqrt{j})$.

Remark 1.3.2. From formula (1.3.4) we know that $d^{(1)}_{k-i+1:k}$ $(i = 1,\ldots,k)$ does not depend on k. And we have $d^{(1)}_{k-i+1:k} = d^{(1)}_{1:i}$. These values for $k = 1\ (1)\ 6, 10, \infty$ and $P^* = .99, .975, .95, .925, .9, .85, .8, .75, .7, .65$ are tabulated in Table I.

1.3.5. Some Other Forms of Selection Procedure δ_1

Lemma 1.3.2. $d_{1:i}^{(1)}$ is increasing in i.

Proof. By Remark 1.3.2 and the fact

$$V_{i+1} = \min (V_i, \frac{iV_i + Z_{i+1}}{i+1}).$$

Lemma 1.3.3. If c_j, $1 \le j \le i \le k$ is decreasing in j, then

$$\bigcup_{j=1}^{i} \{\hat{X}_{j:k} \ge -c_j\} = \bigcup_{j=1}^{i} \{\hat{\hat{X}}_{j:k} \ge -c_j\}$$

Proof.
$$\bigcup_{j=1}^{i} \{\hat{X}_{j:k} \ge -c_j\} \supseteq \bigcup_{j=1}^{i} \{\hat{\hat{X}}_{j:k} \ge -c_j\},$$

since

$$\hat{X}_{j:k} \ge \hat{\hat{X}}_{j:k}, \quad 1 \le j \le k.$$

On the other hand, if

$$\hat{X}_{r:k} \ge -c_r \quad \text{for some } r, \quad 1 \le r \le i$$

then

$$\hat{\hat{X}}_{s:k} \ge -c_r \quad \text{for some } s, \quad 1 \le s \le r,$$

since

$$\hat{X}_{r:k} = \max_{1 \le s \le r} \{\hat{\hat{X}}_{s:k}\}.$$

Because c_j is decreasing in j, this implies $\hat{\hat{X}}_{s:k} \ge -c_s$ for some s, $1 \le s \le r$.

Hence we have

$$\bigcup_{j=1}^{i} \{\hat{X}_{j:k} \geq - c_j\} \subseteq \bigcup_{j=1}^{i} \{\hat{\hat{X}}_{j:k} \geq - c_j\},$$

therefore the lemma is proved.

Definition 1.3.2. We define a selection procedure δ_1' by replacing the inequality in the ith step of procedure δ_1 by the inequality

$$\hat{\hat{X}}_{i:k} \geq \mu_0 - d'_{i:k} \frac{\sigma}{\sqrt{n}}, \quad i = 1,\ldots,k$$

where $d'_{i:k},\ldots,d'_{k:k}$ are the smallest values such that δ_1' satisfies the P*-condition.

Theorem 1.3.7. The selection procedure δ_1 and δ_1' are identical and $d^{(1)}_{i:k} = d'_{i:k}$, $i = 1,\ldots,k$.

Proof. For any i, $1 \leq i \leq k$, by Theorem 1.3.1

$$P^* = \inf_{\underline{\mu} \in \Omega_i} P_{\underline{\mu}}(CS|\delta_1) = P(\hat{\hat{Z}}_{k-i+1:k} \geq - d^{(1)}_{k-i+1:k}).$$

On the other hand, using the same arguments as Section 1.3.2, we have

$$P^* = \inf_{\underline{\mu} \in \Omega_i} P_{\underline{\mu}}(CS|\delta_i') = P(\hat{\hat{Z}}_{k-i+1:k} \geq - d'_{k-i+1:k}).$$

Hence we have $d^{(1)}_{i:k} = d'_{i:k}$, $i = 1,\ldots,k$.

Since $\hat{X}_{1:k} \equiv \hat{\hat{X}}_{1:k}$, the first step of δ_1 and δ_1' are identical. For $i = 2,\ldots,k$, the event

$$\{\text{select } \pi_i,\ldots,\pi_k|\delta_1\} = \{\bigcup_{j=1}^{i}(\hat{X}_{j:k} \geq \mu_0 - d^{(1)}_{j:k}\frac{\sigma}{\sqrt{n}})\}$$

$$= \{\bigcup_{j=1}^{i}(\hat{\hat{X}}_{j:k} \geq \mu_0 - d^{(1)}_{j:k}\frac{\sigma}{\sqrt{n}})\}$$

$$= \{\text{select } \pi_i,\ldots,\pi_k|\delta_1'\}$$

by Lemma 1.3.2 and Lemma 1.3.3. Hence selection procedures δ_1 and δ_1' are identical.

1.3.6. Some Other Proposed Selection Procedures δ_2, δ_3, δ_4

In Case I, we proposed some other selection procedures:

Definition 1.3.3. We define a selection procedure δ_2 by

δ_2: Select π_i if and only if $\hat{X}_{i:k} \geq \mu_0 - d\frac{\sigma}{\sqrt{n}}$ $\quad i = 1,\ldots,k$

where d is the smallest value such that δ_2 satisfies the P*-condition.

Theorem 1.3.8. Under assumptions of Case I, and selection procedure δ_2, if we select population π_i, then we will select populations π_j, for all $j \geq i$.

Proof. Since $\hat{X}_{i:k} \leq \hat{X}_{j:k}$ for all $j \geq i$.

Evaluation the Value d of δ_2

For any i, $1 \leq i \leq k$, we have

$$\inf_{\underline{\mu}\in\Omega_i} P_{\underline{\mu}}(CS|\delta_2) = \inf_{\underline{\mu}\in\Omega_i} P_{\underline{\mu}}(\hat{X}_{k-i+1:k} \geq \mu_0 - d\frac{\sigma}{\sqrt{n}})$$

$$= P(V_i \geq -d)$$

by the same argument for selection procedure δ_1 and here

$$V_i = \hat{\hat{Z}}_{1:i} = \min_{1 \le r \le i} \{\frac{1}{r} \sum_{j=1}^{r} Z_j\}.$$

We need the constant d such that $P(V_i \ge -d) \ge P^*$ holds for all i, $1 \le i \le k$. By Lemma 1.3.2 we have $d = d_{1:k}^{(1)}$. Hence we have the following theorem.

Theorem 1.3.9. Selection procedure δ_2 satisfies the P*-condition with $d = d_{1:k}^{(1)}$.

Corollary 1.3.3. If S_1 and S_2 are the selected subsets associated with selection procedures δ_1 and δ_2, respectively, then $S_1 \subset S_2$.

Proof. Proof follows from Lemma 1.3.2.

Definition 1.3.4. The procedure δ_3 is defined as follows:

Step 1. Select π_i, $i \ge 1$ and stop, if

$$\tilde{X}_1 \ge \mu_0 - d_1 \frac{\sigma}{\sqrt{n}},$$

otherwise reject π_1 and go to step 2.

Step 2. Select π_i, $i \ge 2$ and stop, if

$$\tilde{X}_2 \ge \mu_0 - d_2 \frac{\sigma}{\sqrt{n}},$$

otherwise reject π_2 and go to step 3.

. . .

Step k-1. Select π_i, $i \ge k - 1$ and stop, if

$$\tilde{X}_{k-1} \ge \mu_0 - d_{k-1} \frac{\sigma}{\sqrt{n}},$$

otherwise reject π_{k-1} and go to step k.

Step k. Select π_k and stop, if

$$\tilde{X}_k \geq \mu_0 - d_k \frac{\sigma}{\sqrt{n}},$$

otherwise reject π_k.

Here $\tilde{X}_j = \max\{X_1, \ldots X_j\}$ and d_i's are the smallest values such that δ_3 satisfies the P*-condition.

Evaluation of d_i's

For any i, $1 \leq i \leq k$,

$$\inf_{\underline{\mu} \in \Omega_i} P_{\underline{\mu}}(CS|\delta_3) = \inf_{\underline{\mu} \in \Omega_i} P_{\underline{\mu}}\left(\bigcap_{j=1}^{k-i+1} \{\tilde{X}_j \geq \mu_0 - d_j \frac{\sigma}{\sqrt{n}}\}\right)$$

$$= P_{\underline{\mu}}(Z_{k-i+1} \geq - d_{k-i+1} \frac{\sigma}{\sqrt{n}})$$

$$= P(Z_{k-i+1} \geq - d_{k-i+1})$$

$$= P^*, \quad Z_i \sim F(\cdot; 0, 1).$$

This implies $d_{k-i+1} = d$ for all i, and

$$d = - F^{-1}(1-P^*),$$

$$= F^{-1}(P^*), \quad \text{if F is symmetric}$$

$$= \Phi^{-1}(P^*), \quad \text{if } X_i \sim N(\mu_i, \sigma^2/n).$$

Similar to the selection procedure δ_1, we have the following theorem:

Theorem 1.3.10. Selection procedure δ_3 satisfies the P*-condition with $d_i = - F^{-1}(1-P^*)$.

Definition 1.3.5. Selection procedures δ_3' is defined as follows:

Step 1. Select π_i, $i \geq 1$ and stop, if

$$X_1 \geq \mu_0 - d\frac{\sigma}{\sqrt{n}},$$

otherwise reject π_1 and go to step 2.

Step 2. Select π_i, $i \geq 2$ and stop, if

$$X_2 \geq \mu_0 - d\frac{\sigma}{\sqrt{n}},$$

otherwise reject π_2 and go to step 3.

.
.
.

Step k-1. Select π_i, $i \geq k \geq 1$ and stop, if

$$X_{k-1} \geq \mu_0 - d\frac{\sigma}{\sqrt{n}},$$

otherwise reject π_k.

Here

$$d = -F^{-1}(1-P^*)$$
$$= F^{-1}(P^*) \text{ if } F \text{ is symmetric.}$$

Theorem 1.3.11. The selection procedures δ_3' satisfies the P*-condition.

Proof. For any i, $1 \leq i \leq k$,

$$\inf_{\underline{\mu}\in\Omega_i} P_{\underline{\mu}}(CS|\delta_3') \geq P(Z_{k-i+1} \geq -d) = P^*.$$

Theorem 1.3.12. The selection procedure δ_3 and δ_3' are identical.

Proof. The proof is simple hence it is omitted.

The following procedure δ_4 was given by Gupta and Sobel (1958), without assuming any ordering prior:

Definition 1.3.6. The selection procedure δ_4 is defined as follows:

δ_4: Select π_i if and only if $X_i \geq \mu_0 - d\frac{\sigma}{\sqrt{n}}$ $i = 1,\ldots,k$

where d is the smallest constant such that δ_4 satisfies the P*-condition.

It was shown that the value d is determined by the equation

$$F(-d) = 1 - P^{*\frac{1}{k}}$$

or

$$F(d) = P^{*\frac{1}{k}} \text{ if } F \text{ is symmetric.}$$

1.3.7. A Dual Problem

We start with the same assumptions as in Section 1.3.1 Case 1, but change our goal to select a subset which contains no bad populations; the definition of a correct selection (CS) will now be changed to select a subset that contains no bad populations and the P*-condition will be defined based on this new definition of correct selection (CS).

In location parameter case, this problem is a dual problem of the original problem, namely, "select a subset which contains all good populations under ordering prior assumption".

One method to solve this problem is that, first, change the signs of all statistics and the control to opposite sign; then use a procedure δ for selecting a subset which contains all "new good" populations,

where the "new good" populations are the "old bad" populations before changing signs; finally, reject the selected subset and keep the remainders as the desired selected subset. Let ψ_i, i = 1, 2, 3, 4 denote the above procedure which corresponds to δ_i, i = 1, 2, 3, 4, respectively.

Theorem 1.3.13. The selection procedure ψ_i, i = 1, 2, 3, 4 satisfies the P*-condition in which the correct selection (CS) means that it selects a subset which contains no bad population.

Proof. Given P* and observations, for any selection procedure ψ_i, i = 1, 2, 3, 4, after changing the signs of all associated statistics, the probability that the selected subset S contains all "new good" populations is not less than P*. If we reject the selected subset S, then the complement subset S^c of S contains any "new good" populations with probability less than 1-P*, but the "new good" populations are the originally bad populations so what we have is that the subset S^c contains any originally bad population with probability less than 1-P*, in other words, subset S^c contains no bad populations with probability greater than or equal to P*. Since this is true for all arbitrary true configurations, we have completed the proof.

Remark 1.3.3. It is easy to see that the value $d_{i:k}(\psi_1)$ of ψ_1 which was used by ψ_1 in the ith step is determined by the equation

$$P(U_{k-i+1} < d_{i:k}(\psi_1)) = P^*,$$

where

$$U_{k-i+1} = \max_{1\le r\le k-i+1} \frac{1}{r} \sum_{j=1}^{r} Z_j.$$

If the distribution F is symmetric, then

$$d_{i:k}(\psi_1) = d^{(1)}_{i:k}.$$

1.3.8. Some Proposed Selection Procedures $\delta^{(2)}_i$, i = 1, 2, 3, 4 When μ_0 is Unknown

Case II. μ_0 unknown, common σ^2 known, common sample size n.

Definition 1.3.7. We define a selection procedure $\delta^{(2)}_1$ by replacing the inequalities

$$\hat{X}_{i:k} \geq \mu_0 - d^{(1)}_{i:k} \frac{\sigma}{\sqrt{n}}, \quad i = 1,\ldots,k$$

in procedure δ_1 (Definition 1.3.1) with

$$\hat{X}_{i:k} \geq X_0 - d^{(2)}_{i:k} \frac{\sigma}{\sqrt{n}}, \quad i = 1,\ldots,k, \text{ respectively.}$$

Here $X_0 = \sum_{i=1}^{n} X_{0i}/n$, $d^{(2)}_{i:k}$, i = 1,...,k are the smallest constants such that the selection procedure $\delta^{(2)}$ satisfies the P*-condition.

Similar to the Case I, we have the following theorem:

Theorem 1.3.14. For any i, $1 \leq i \leq k$, $d^{(2)}_{k-i+1:k}$ is determined by the equation

$$\int_{-\infty}^{\infty} P(V_i \geq t - d^{(2)}_{k-i+1:k}) dF(t) = P^*. \qquad (1.3.22)$$

It is easy to see that $d^{(2)}_{k-i+1:k} = d^{(2)}_{1:i}$. The following theorem gives us an identical form of the selection procedure $\delta^{(2)}_1$.

Theorem 1.3.15. The selection procedure $\delta^{(2)}_1$ will not be changed if the statistics $\hat{X}_{i:k}$, $i = 1,\ldots,k$, are replaced by $\hat{\hat{X}}_{i:k}$, $i = 1,\ldots,k$, respectively.

Proof. The proof is the same as that in Case I and hence it is omitted.

The values $d^{(2)}_{1:i}$, $i = 1,\ldots,k$ are tabulated in Table II for $k = 1\ (1)\ 6, 8, 10,$ and $P^* = .99, .975, .95, .925, .90, .85, .80, .75, .70, .65$.

Similar to the Case I, we propose a selection procedure $\delta^{(2)}_2$ as follows:

Definition 1.3.8. We define a selection procedure $\delta^{(2)}_2$ by

$\delta^{(2)}_2$: Select π_i if and only if $\hat{X}_{i:k} \geq X_0 - d\frac{\sigma}{\sqrt{n}}$ $\quad i = 1,\ldots,k$

where d is the smallest value such that $\delta^{(2)}_2$ satisfies the P*-condition. Then, similar to Theorem 1.3.9 we have:

Theorem 1.3.16. Under assumptions of Case II, the selection procedure $\delta^{(2)}_2$ satisfies the P*-condition with $d = d^{(2)}_{1:k}$.

Next, we define a selection procedure $\delta^{(2)}_3$ which is similar to δ_3 but replace μ_0 by X_0, the sample mean of population π_0.

Definition 1.3.9. The selection procedure $\delta_3^{(2)}$ is defined by replacing $\tilde{X}_i \geq \mu_0 - d_i \frac{\sigma}{\sqrt{n}}$ in δ_3 (Definition 1.3.4) by $\tilde{X}_i \geq X_0 - d_i' \frac{\sigma}{\sqrt{n}}$, $i = 1,\ldots,k$ where $d_1',\ldots,d_k'$ are the smallest values such that $\delta_3^{(2)}$ satisfies the P*-condition.

Similar to Theorem 1.3.10 we have:

Theorem 1.3.17. The selection procedure $\delta_3^{(2)}$ satisfies the P*-condition with $d_i = d$, $i = 1,\ldots,k$ where d is determined by the equation

$$\int_{-\infty}^{\infty} [1 - F(t-d)]dF(t) = P^*, \qquad (1.3.23)$$

$$\int_{-\infty}^{\infty} F(d-t)dF(t) = P^*, \text{ if } F \text{ is symmetric.}$$

And $\delta_3^{(2)}$ will not be changed if the statistics $\tilde{X}_i$ is replaced by X_i, the sample mean of population π_i for $i = 1,\ldots,k$.

The following selection procedure $\delta_4^{(2)}$ was proposed by Gupta and Sobel (1958):

Definition 1.3.10. The selection procedure $\delta_4^{(2)}$ is defined by

$\delta_4^{(2)}$: Select π_i if and only if $X_i \geq X_0 - d \frac{\sigma}{\sqrt{n_i}}$ $\quad i = 1,\ldots,k$

where d is determined by the following equation if F is normal distribution:

$$\int_{-\infty}^{\infty} \prod_{i=1}^{k} [F(t\sqrt{\frac{n_i}{n_0}} + d)]f(u)du = P^*. \qquad (1.3.24)$$

For the special case $n_i = n$ $(i = 0, 1, \ldots, k)$

$$\int_{-\infty}^{\infty} F^k(t+d)f(t)dt = P^*. \qquad (1.3.25)$$

If F is normal distribution N(0,1), the tables of d-values satisfying the Equation (1.3.25) for several values of P* are given in Bechhofer (1954) for k = 1 (1) 10 and in Gupta (1956) for k = 1 (1) 50.

1.3.9. Some Proposed Selection Procedures $\delta_i^{(3)}$, i = 1, 2, 3, 4 When Common Variance σ^2 is Unknown

Case III. μ_0 known, common variance σ^2 unknown, $n_i = n > 1$.

In this case, we assume that $F(x) = \Phi(x)$ which is the c.d.f. of N(0,1).

Definition 1.3.11. We define the selection procedure $\delta_1^{(3)}$ by replacing the inequalities

$$\hat{X}_{i:k} \geq \mu_0 - d_{i:k}^{(1)} \frac{\sigma}{\sqrt{n}} \quad i = 1, \ldots, k$$

in procedure δ_1 (Definition 1.3.1) by

$$\hat{X}_{i:k} \geq \mu_0 - d_{i:k}^{(3)} \frac{S}{\sqrt{n}} \quad i = 1, \ldots, k, \text{ respectively,}$$

where $d^{(3)}$'s are the smallest values such that $\delta_1^{(3)}$ satisfies the P*-condition, S^2 denotes the pooled estimator of σ^2 based on $\nu = k(n-1)$, that is

$$S^2 = \sum_{i=1}^{k} \sum_{j=1}^{n} (X_{ij} - X_i)^2/\nu. \qquad (1.3.26)$$

Note that $\frac{\nu S^2}{\sigma^2}$ has the chi-square distribution χ^2_ν with ν degrees of freedom.

By using similar arguments as in Case I, we have:

Theorem 1.3.18. The equation which determines the constant $d^{(3)}_{k-i+1:k}$ is

$$P(V_i \geq - d_{k-i+1:k} \frac{S}{\sigma}) = P^* \tag{1.3.27}$$

or

$$\int_0^\infty P(V_i \geq - d_{k-i+1:k}\, y) q_\nu(y) dy = P^* \tag{1.3.28}$$

where

$$V_i \equiv \min_{1 \leq r \leq i} \frac{1}{r} \sum_{j=1}^{r} Z_j$$

and $q_\nu(y)$ is the density of $\frac{S}{\sigma} = \frac{\chi_\nu}{\sqrt{\nu}}$.

We can rewrite Formula (1.3.28) as

$$\int_0^\infty P(V_i \geq - d^{(3)}_{k-i+1:k} \sqrt{\frac{t}{\nu}}) d\chi^2_\nu(t) = P^*$$

or

$$\int_0^\infty P(V_i \geq - d^{(3)}_{k-i+1:k} \sqrt{\frac{2t}{\nu}}) \frac{t^{\frac{\nu}{2}-1} e^{-t}}{\Gamma(\frac{\nu}{2})} dt = P^*. \tag{1.3.30}$$

Remark 1.3.4. The values of $d^{(3)}_{k-i+1:k}$, $i = 1,\ldots,k$ depend on $\nu = k(n-1)$, hence $d^{(3)}_{k-i+1:k} \neq d^{(3)}_{1:i}$.

By using Rabinowitz and Weiss table (1959) (with $n = 0$, $N = 24$), we have evaluated and tabulated the values of $d^{(3)}_{k-i+1:k}$, $i = 1,\ldots,k$, in Table III, for $k = 2\ (1)\ 6$, $P^* = .99, .975, .95, .925, .90, .85, .80$, and $.75$ with common sample size $n = 3, 5, 9$, and 21.

For $k \geq 6$ and $n > 21$, we can use $d^{(1)}_{1:i}$ as an approximation of $d^{(3)}_{k-i+1:k}$.

Definition 1.3.12. We define the selection procedure $\delta^{(3)}_2$ by

$\delta^{(3)}_2$: Select π_i if and only if $\hat{X}_{i:k} \geq \mu_0 - d^{(3)} \frac{S}{\sqrt{n}}$ $i = 1,\ldots,k$

where S is defined as in procedure $\delta^{(3)}_1$, and $d^{(3)}$ is the smallest constant such that $\delta^{(3)}_2$ satisfies the P*-condition.

As before, it can be shown that $d^{(3)} = d^{(3)}_{1:k}$.

Remark 1.3.5. Theorem 1.3.7 still holds for Case III, i.e. the selection procedure $\delta^{(3)}_1$ will not be changed if we replace the isotonic statistics $\hat{X}_{i:k}$ by $\hat{\hat{X}}_{i:k}$, respectively. But this is not necessarily true for selection procedure $\delta^{(3)}_2$.

Definition 1.3.13. The selection procedure $\delta^{(3)}_3$ is defined to have the same form as procedure $\delta^{(2)}_3$ except that the inequality defined in the ith step of procedure $\delta^{(2)}_3$ is replaced by

$$X_i \geq \mu_0 - d \frac{S}{\sqrt{n}} \quad \text{for } i = 1,\ldots,k.$$

The proof of the following theorem uses the same arguments as that in Case I, hence it is omitted.

Theorem 1.3.19. The equation which determines the constant d of selection procedure $\delta_3^{(3)}$ is

$$\int_0^\infty \Phi(yd)q_\nu(y)dy = P^*. \tag{1.3.31}$$

Gupta and Sobel (1958) gave a selection procedure $\delta_4^{(3)}$ in this case. It is as follows:

$\delta_4^{(3)}$: Select π_i if and only if $X_i \geq \mu_0 - d\frac{S}{\sqrt{n_i}}$ $i = 1,\ldots,k$

and the equation which determines d is

$$\int_0^\infty \Phi^k(yd)q_\nu(y)dy = P^*. \tag{1.3.32}$$

1.3.10. Some Proposed Selection Procedures $\delta_i^{(4)}$, i = 1, 2, 3, 4 When Both Control μ_0 and Common Variance σ^2 are Unknown

Case IV. μ_0 unknown, common variance σ^2 unknown and common sample size n.

We assume that in this case distribution F is the c.d.f. N(0,1), and denoted by Φ. We replace μ_0 in each selection procedure $\delta_j^{(3)}$ by X_0, $1 \leq j \leq 4$, and get four procedures $\delta_j^{(4)}$, $1 \leq j \leq 4$, respectively. Let $\chi_\nu^2(t)$ denote the c.d.f. of the chi-square distribution with $\nu = k(n-1)$ degrees of freedom. The constant $d_{k-i+1:k}^{(4)}$, $i = 1,\ldots,k$, of procedure $\delta_1^{(4)}$ is determined by

$$\int_0^\infty \int_{-\infty}^\infty P(V_i \le u - d_{k-i+1:k}^{(4)} \sqrt{\frac{t}{\nu}}) d\Phi(u) d\chi_\nu^2(t) = P^*. \quad (1.3.33)$$

The constant d of procedure $\delta_2^{(4)}$ is

$$d = d_{1:k}^{(4)}.$$

The constants d of procedures $\delta_3^{(4)}$ and $\delta_4^{(4)}$ are determined by

$$\int_0^\infty \int_{-\infty}^\infty \Phi^r(u + td) d\Phi(u) d\chi_\nu^2(t) = P^* \quad (1.3.34)$$

with r = 1 and k, respectively, and their values for selected values of P*, k and ν are given in Gupta and Sobel (1957) and Dunnett (1955).

1.3.11. Properties of the Selection Procedures

Under simple ordering prior, it is natural to require that an ideal selection procedure is order-preserving as defined below:

Definition 1.3.14. A selection procedure δ is order-preserving if it selects π_i with parameter μ_i, and if $\mu_i < \mu_j$, then it also selects π_j. Procedure δ is weak order-preserving or monotone if

$$P(\pi_i \text{ is selected}|\delta) \le P(\pi_j \text{ is selected}|\delta) \text{ whenever } \mu_i < \mu_j.$$

It is easy to see that any order-preserving selection procedure is weak order-preserving, but the converse is not true.

Now, let $\delta_i^{(1)} = \delta_i$, i = 1, 2, 3, 4.

Theorem 1.3.20. The selection procedures $\delta_1^{(i)}$, $\delta_2^{(i)}$ and $\delta_3^{(i)}$ are order-preserving and procedure $\delta_4^{(i)}$ is monotone, for i = 1, 2, 3, 4.

Proof. The proof follows immediately from the definitions of the procedures.

Given observations $\underline{X} = \underline{x} = (x_0,\ldots,x_k)$ where x_i is the sample mean of population π_i, $i = 1,\ldots,k$, and $x_0 = \mu_0$ if μ_0 is known; otherwise x_0 is the sample mean of population π_0. Let

$$\psi_i(\underline{x}, \delta) = P(\pi_i \text{ included in the selected subset} | \underline{X} = \underline{x}, \delta) \quad \text{for } i = 1,\ldots,k.$$

Definition 1.3.15. A selection procedure δ is called translation-invariant if for any $\underline{x} \in \mathbb{R}^{k+1}$, $c \in \mathbb{R}$

$$\psi_i(x_0 + c, x_1 + c,\ldots,x_k + c; \delta) = \psi_i(x_0,\ldots,x_k; \delta) \quad i = 1,\ldots,k.$$

Theorem 1.3.21. The selection procedures $\delta_1^{(i)}$, $\delta_2^{(i)}$, $\delta_3^{(i)}$ and $\delta_4^{(i)}$ are translation-invariant for $i = 1, 2, 3, 4$.

Proof. By Corollary 1.3.1 the isotonic regression is a linear operator. On the other hand,

$$\frac{\sum_{j=1}^{n} (X_{ij} + c)}{n} = \frac{\sum_{j=1}^{n} X_{ij}}{n} + c$$

hence we have the result.

Expected Number (Size) of Bad Populations in the Selected Subset

Suppose the control μ_0 is known and we have common sample size n and common known variance σ^2; without loss of generality, we assume that $\mu_0 = 0$ and $\sigma/\sqrt{n} = 1$. Let $E(S'|\delta)$ denote the expected number of bad populations in the selected subset in using the selection procedure

δ, then for any j, $0 \le j \le k$,

$$\sup_{\underline{\mu}\in\Omega_{k-j}} E_{\underline{\mu}}(S'|\delta_1)$$

$$= \sup_{\underline{\mu}\in\Omega_{k-j}} \sum_{r=1}^{j} P_{\underline{\mu}}\Big(\bigcup_{\ell=1}^{r}\{\hat{\hat{X}}_{\ell:k} \ge - d^{(1)}_{\ell:k}\}\Big) -$$

$$= \sum_{r=1}^{j} P\Big(\bigcup_{\ell=1}^{r}\{\hat{\hat{Z}}_{\ell:j} \ge - d^{(1)}_{\ell:k}\}\Big) \qquad (1.3.35)$$

On the other hand, for procedure δ_2

$$\sup_{\underline{\mu}\in\Omega_{k-j}} E(S'|\delta_2) = \sum_{r=1}^{j} P\Big(\bigcup_{\ell=1}^{r}\{\hat{\hat{Z}}_{\ell:j} \ge - d^{(1)}_{1:k}\}\Big) \qquad (1.3.36)$$

Formula (1.3.36) is increasing in j and is greater than or equal to Formula (1.3.35), since

$$d^{(1)}_{\ell:k} = d^{(1)}_{1:k-\ell+1} < d^{(1)}_{1:k}.$$

Therefore, we have the following theorem.

Theorem 1.3.22. For any i, $0 \le i \le k$

$$\sup_{\underline{\mu}\in\Omega_i} E(S'|\delta_2) \ge \sup_{\underline{\mu}\in\Omega_i} E(S'|\delta_1),$$

$$\sup_{\underline{\mu}\in\Omega} E(S'|\delta_2) = \sup_{\underline{\mu}\in\Omega_0} E(S'|\delta_2).$$

Theorem 1.3.23. In Section 1.3.1, Case I, for any j, $0 \le j \le k$

$$\sup_{\underline{\mu}\in\Omega_{k-j}} E(S'|\delta_3) = j - q(1-q^j)/P^* \qquad (1.3.37)$$

where $q = 1 - P^*$.

Proof.

$$\sup_{\underline{\mu}\in\Omega_{k-j}} E(S'|\delta_3)$$

$$= \sup_{\underline{\mu}\in\Omega_{k-j}} \sum_{i=1}^{j} P_{\underline{\mu}}(\text{select } \pi_i|\delta_3)$$

$$= \sup_{\underline{\mu}\in\Omega_{k-j}} \sum_{i=1}^{j} P_{\underline{\mu}}(\max_{1\le r\le i} X_r \ge -d)$$

$$= \sum_{i=1}^{j} (1 - \prod_{r=1}^{i} F(-d))$$

$$= j - \sum_{i=1}^{j} q^i$$

$$= j - q(1-q^j)/P^*$$

where $q = (1-P^*)$.

Theorem 1.3.24. $\sup_{\underline{\mu}\in\Omega_{k-j}} E(S'|\delta_3)$ is increasing in j, hence

$$\sup_{\underline{\mu}\in\Omega_{k-j}} E(S'|\delta_3) = \sup_{\Omega_0} E(S'|\delta_3) = k - q(1-q^k)/P^*. \quad (1.3.38)$$

Proof. Since the function

$$f(x) = x - ab^{x+1}$$

is increasing in x, for $0 < a < 1$, $0 < b < 1$, and $0 < x < \infty$.

In Case I of Gupta (1965) showed that

$$\sup_{\underline{\mu}\in\Omega} E(S'|\delta_4) = kP^{*\frac{1}{k}}. \quad (1.3.39)$$

When the ordering prior among the unknown parameters is unknown, we can use the selection procedure of Gupta and Sobel (1958) or use the ordering of the sample means as the ordering of unknown parameters and apply the selection procedure which is originally used under ordering prior. With the latter approach, the substitution implies that the isotonic regression of the sample means turns to the usual ordered sample means, and that the selection procedures $\delta_j^{(i)}$ (i = 1, 2, 3, 4; j = 2, 3) are the same as $\delta_4^{(i)}$ (i = 1, 2, 3, 4), respectively, and the selection procedures $\delta_1^{(i)}$, i = 1, 2, 3, 4 become $\delta_5^{(i)}$, i = 1, 2, 3, 4 respectively, which are equivalent to the procedures proposed by Naik (1975) and Broström (1977), independently.

It has been proved in some quite general situations and studied by using Monte Carlo technique in some selected cases by Naik (1975) and Broström (1977), separately, that δ_5 (= $\delta_5^{(1)}$) is slightly better than δ_4. The values $d_{i:k}^{(j)}$ in the ith step of the procedure $\delta_5^{(j)}$, j = 1, 2, 3, 4, are given by Broström (1977) as follows:

$$d_{i:k}^{(1)} = F^{-1}[1 - (P^*)^{\frac{1}{i}}], \qquad (1.3.40)$$

$$\int_{-\infty}^{\infty} \Phi^i(x + d_{i:k}^{(2)})\varphi(x)dx = P^*, \qquad (1.3.41)$$

$$\int_0^{\infty} \Phi^i(xd_{i:k}^{(3)})q_\nu(x)dx = P^*, \qquad (1.3.42)$$

and

$$\int_0^{\infty}\int_{-\infty}^{\infty} \Phi^i(x + yd_{i:k}^{(4)})\varphi(x)q_\nu(y)dx\,dy = P^* \qquad (1.3.43)$$

where q_ν is the density of S/σ.

1.4. Selection Procedures for Scale Parameters of Gamma Populations

Suppose we have k + 1 independent populations $\pi_0,\ldots,\pi_k$. The population π_i has a gamma density function

$$g(x;\ \alpha_i,\ \theta_i) = \frac{1}{\Gamma(\frac{\alpha_i}{2})\theta_i^{\frac{\alpha_i}{2}}}\, x^{\frac{\alpha_i}{2}-1} e^{-\frac{x}{\theta_i}}, \qquad (1.4.1)$$

$i = 1,\ldots,k$. The ordering prior of $\theta_1,\ldots,$ and θ_k is assumed known, say, $0 < \theta_1 \le \theta_2 \le \ldots \le \theta_k < \infty$. Note that the values of $\theta_1,\ldots,$ and θ_k are unknown, α_i's are known.

In this section we define population π_i, $i = 1,\ldots,k$, as good population if the scale parameter $\theta_i \le \theta_0$. Let $\underline{\theta} = (\theta_0, \theta_1,\ldots,\theta_k)$, then the parameter space is denoted by Ω, where

$$\Omega = \{\underline{\theta} \in \mathbb{R}^{k+1} \mid 0 < \theta_1 \le \ldots \le \theta_k;\ \ 0 < \theta_0 < \infty\}$$

is a subspace of (k+1)-dimension Euclidean space $\mathbb{R}^{k+1}$.

Suppose we have independent observations X_{ij} $(i = 1,\ldots,n_i)$ from population π_i, $(i = 1,\ldots,k)$. Let $\nu_i = n_i\alpha_i$, then

$$X_i = \sum_{j=1}^{n_i} X_{ij}/n_i \text{ has density } g(\cdot;\ \nu_i,\ \theta_i/n_i),$$

and

$$X_i/\theta_i \text{ has density } g(\cdot;\ \nu_i,\ 1/n_i).$$

Suppose our goal is to select a subset which contains all good populations under the ordering prior with probability greater than or equal to P*, a predetermined value between zero and one.

Let Ω_i, $i =$,...,k, be the subspaces of parameter space Ω, such that $\Omega = \bigcup_{i=0}^{k} \Omega_i$ where

$$\Omega_i = \{\underline{\theta} \in \Omega | \theta_i \leq \theta_0 < \theta_{i+1}\}, \quad \text{if } i = 1, \ldots, \text{or } k-1,$$

$$= \{\underline{\theta} \in \Omega | \theta_k \leq \theta_0\}, \quad \text{if } i = k,$$

$$= \{\underline{\theta} \in \Omega | \theta_0 < \theta_1\}, \quad \text{if } i = 0.$$

1.4.1. Proposed Selection Procedures δ_i, i = 6, 7, 8, 9

Case I. Control θ_0 known and common sample size n.

Definition 1.4.1. The selection procedure δ_6 is defined as follows:

Step 1. Select π_i, $i \leq k$ and stop, if

$$\hat{X}_{k:k} \leq c_{k:k}\ \theta_0,$$

otherwise reject π_k and go to step 2.

Step 2. Select π_i, $i \leq k - 1$ and stop, if

$$\hat{X}_{k-1:k} \leq c_{k-1:k}\ \theta_0,$$

otherwise reject π_{k-1} and go to step 3.

.
.
.

Step k-1. Select π_i, $i \leq 2$ and stop, if

$$\hat{X}_{2:k} \leq c_{2:k}\ \theta_0,$$

otherwise reject π_2 and go to step k.

Step k. Select π_1 and stop, if

$$\hat{X}_{1:k} \leq c_{1:k}\ \theta_0,$$

otherwise reject π_1.

Here $c_{i:k}$ (≥ 1), $i = 1,\ldots,k$ are the smallest values such that the procedure δ_6 satisfies the P*-condition.

Theorem 1.4.1. Assume we have common sample size n and $\alpha_i = \alpha > 0$, $i = 1,\ldots,k$, and the constant $c_{i:k}$ (≥ 1) is determined by the equation

$$P(U_i \leq c_{i:k}) = P^*, \quad i = 1,\ldots,k \tag{1.4.2}$$

where

$$U_i \equiv \max_{1\leq s\leq i} \left(\frac{Y_s + \ldots + Y_i}{i-s+1}\right) \tag{1.4.3}$$

and Y_i are i.i.d. with density

$$g(\cdot;\ \nu,\ \tfrac{1}{n}),\ \text{and c.d.f. } G(\cdot;\ \nu,\ \tfrac{1}{n})$$

then the procedure δ_6 satisfies the P*-condition.

Proof. For any i, $1 \leq i < k$, if the unknown true $\underline{\theta} \in \Omega_i$, that is if there are i good populations, then, under the procedure δ_6

$$\inf_{\underline{\theta}\in\Omega_i} P_{\underline{\theta}}(CS|\delta_6)$$

$$= \inf_{\underline{\theta}\in\Omega_i} P_{\underline{\theta}}\left(\bigcup_{j=i}^{k} (\hat{X}_{j:k} \leq c_{j:k}\ \theta_0)\right)$$

$$= \inf_{\underline{\theta}\in\Omega_i} P_{\underline{\theta}}\left(\bigcup_{j=i}^{k} \left(\max_{1\leq s\leq j} \min_{j\leq t\leq k} \frac{X_s + \ldots + X_t}{t-s+1} \leq c_{j:k}\ \theta_0\right)\right)$$

$$= \inf_{\underline{\theta}\in\Omega_i} P_{\underline{\theta}}\left(\bigcup_{j=i}^{k} \left(\max_{1\leq s\leq j} \min_{j\leq t\leq k} \frac{\frac{X_s}{\theta_s}\theta_s + \ldots + \frac{X_t}{\theta_t}\theta_t}{t-s+1} \leq c_{j:k}\ \theta_0\right)\right)$$

$$= \inf_{\underline{\theta}\in\Omega_i} P_{\underline{\theta}}\Big(\bigcup_{j=i}^{k} \Big(\max_{1\le s\le j} \min_{j\le t\le k} \frac{Y_s\,\theta_s + \ldots + Y_t\,\theta_t}{t-s+1} \le c_{j:k}\,\theta_0 \Big) \Big)$$

$$= P_{\underline{\theta}=\underline{\theta}^*}\Big(\bigcup_{j=i}^{k} \Big(\max_{1\le s\le j} \min_{j\le t=i} \frac{Y_s\,\theta_0 + \ldots + Y_t\,\theta_0}{t-s+1} \le c_{j:k}\,\theta_0 \Big) \Big)$$

$$= P\Big(\max_{1\le s\le i} \frac{Y_s + \ldots + Y_i}{i-s+1} \le c_{i:k} \Big)$$

$$= P(U_i \le c_{i:k}),$$

where Y_i's are i.i.d with density $g(\cdot;\ \nu,\ \frac{1}{n})$,

$$\underline{\theta}^* = (\underbrace{\theta_0,\ \theta_0,\ldots,\theta_0}_{i+1},\ \infty,\ldots,\infty)$$

and

$$U_i = \max_{1\le s\le i} \frac{Y_s + \ldots + Y_i}{i-s+1}\ .$$

Corollary 1.4.1. $c_{i:k} = c_{i:i}$, $i = 1,\ldots,k$.

For any $x > 0$, let $S_n = \sum_{i=1}^{n} (Y_i - x)$, $n = 1, 2, \ldots$, $S_0 = 0$. Since $0 < P(Y_i - x \le 0) = G(x,\ \nu,\ \frac{1}{n}) < 1$, the distribution of $Y_1 - x$ is not concentrated on a half-axis. By Theorem 1.3.4 the probability generating function of cumulative distribution functions $P(U_k \le x)$, $x > 0$, $k = 1, 2, \ldots$, is given by

$$\exp\ \{ \sum_{k=1}^{\infty} \frac{s^k}{k} \cdot G(xkn;\ k\nu,\ 1)\}.$$

Hence by Theorem 1.3.6, we have the following recurrence formula for all $x > 0$

$$P(U_{k+1} \leq x)$$

$$= \frac{1}{k+1} \sum_{j=0}^{k} P(U_{k-j} \leq x) \cdot G(x(j+1)n; (j+1)\nu, 1) \qquad (1.4.4)$$

where

$$P(U_0 \leq x) \equiv 1.$$

When $x = 0$, both sides of Equation (1.4.4) equal to zero, hence it also holds for $x = 0$.

Note that

$$P(\frac{1}{r} \sum_{i=1}^{r} Y_i \leq x) = G(x; r\nu, \frac{1}{rn})$$

$$= G(xrn; r\nu, 1). \qquad (1.4.5)$$

The values $c_{k:k}(P^*, \alpha, n)$ which satisfy Equation (1.4.2) are tabulated in Table IV for $k = 1\ (1)\ 10$, $P^* = .99, .95, .90, .75$, $\alpha = 2, 4, 6$, and $n = 4, 6, 8, 10, 15, 20$.

Lemma 1.4.1. $c_{i:k} \leq c_{i+1:k}$ for all $1 \leq i \leq k - 1$.

Proof. The constants $c_{i:k}$ $(i = 1,\ldots,k)$ are determined by (1.4.2), respectively.

$U_{i+1} > U_i$ a.s. implies $c_{i:k} \leq c_{i+1:k}$ for all $1 \leq i \leq k - 1$.

Theorem 1.4.2. The selection procedure δ_6 will not be changed if the isotonic estimators $\hat{X}_{i:k}$, $i = 1,\ldots,k$ are replaced by $\hat{X}_{i:i}$, $i = 1,\ldots,k$

where

$$\hat{X}_{i:i} = \max_{1\le j\le i} \hat{\hat{X}}_{j:i}$$

$$= \max_{1\le s\le i} \frac{X_s + \ldots + X_i}{i-s+1}$$

Proof. The proof is similar to that of Theorem 1.3.7.

Next, we define a selection procedure by using an isotonic estimator and a fixed constant which depends on P*, k, common sample size n and common $\alpha_i = \alpha$.

Definition 1.4.2. The selection procedure δ_7 is defined by

δ_7: Select π_i if and only if $\hat{X}_{i:k} \le c\theta_0 \quad i = 1,\ldots,k$

where c (≥ 1) is the smallest value such that procedure δ_7 satisfies the P*-condition.

Corollary 1.4.2. The constant $c = c(P^*, k, n)$ of the selection procedure δ_7 equals to $c_{k:k}$ which is determined by Equation (1.4.2).

Proof. Follows immediately from Theorem 1.4.1 and Lemma 1.4.1.

Definition 1.4.3. The selection procedure δ_8 is defined as follows:

Step 1. Select π_j, $j \le k$ and stop, if

$$\frac{X_k}{\alpha_k} \le c_k\, \theta_0,$$

otherwise reject π_k and go to step 2.

Step 2. Select π_j, $j \leq k-1$ and stop, if

$$\frac{X_{k-1}}{\alpha_{k-1}} \leq c_{k-1}\,\theta_0,$$

otherwise reject π_{k-1} and go to step 3.

.
.
.

Step k-1. Select π_j, $j \leq 2$ and stop, if

$$\frac{X_2}{\alpha_2} \leq c_2\,\theta_0,$$

otherwise reject π_2 and go to step k.

Step k. Select π_1 and stop, if

$$\frac{X_1}{\alpha_1} \leq c_1\,\theta_0,$$

otherwise reject π_1.

Here the c_i's are the smallest real values (≥ 1) such that the procedure δ_8 satisfies the P*-condition.

Theorem 1.4.3. The constants c_i are determined by

$$\frac{1}{\Gamma(\frac{\nu_i}{2})} \int_0^{c_i \nu_i} u^{\frac{\nu_i}{2}-1} e^{-u}\,du = P^*, \quad i = 1,\ldots,k \qquad (1.4.6)$$

Proof. For arbitrary i, $1 \leq i \leq k$, if $\underline{\theta} \in \Omega_i$

$$\inf_{\underline{\theta} \in \Omega_i} P_{\underline{\theta}}(CS|\delta_8)$$

$$= \inf_{\underline{\theta}\in\Omega_i} P_{\underline{\theta}}\Big(\bigcup_{j=i}^{k} \Big(\frac{X_j}{\alpha_j} \le c_j\ \theta_0\Big)\Big)$$

$$= \inf_{\underline{\theta}\in\Omega_i} P_{\underline{\theta}}\Big(\bigcup_{j=i}^{k} \Big(\frac{X_j}{\theta_j\alpha_j} \le c_j\ \frac{\theta_0}{\theta_j}\Big)\Big)$$

$$= P_{\underline{\theta}^*}\Big(\bigcup_{j=i}^{k} \Big(\frac{X_j}{\theta_j\alpha_j} \le c_j\ \frac{\theta_0}{\theta_j^*}\Big)\Big)$$

$$= P(Z_i \le c_i\ \nu_i)$$

where $Z_i = \dfrac{n_i X_i}{\theta_i}$, $i = 1,\ldots,k$ are i.i.d. with the gamma density $g(\cdot;\ \nu_i,\ 1)$, $\underline{\theta}^* = (\underbrace{\theta_0,\ldots,\theta_0}_{i+1},\ \infty,\ldots,\infty)$. Hence c_i, $i = 1,\ldots,k$ are determined by (1.4.6). If $\nu_1 = \ldots = \nu_k$, then $c_1 = \ldots = c_k$.

The following selection procedure δ_9 was given by Gupta and Sobel (1958).

Definition 1.4.4. The selection procedure δ_9 is defined by

δ_9: Select π_i if and only if $\dfrac{X_i}{\alpha_i} \le c'\ \theta_0$ $i = 1,\ldots,k$

where c' is determined by

$$\prod_{i=1}^{k} \Big\{\frac{1}{\Gamma(\frac{\nu_i}{2})} \int_0^{\nu_i c'} u^{\frac{\nu_i}{2}-1} e^{-u} du\Big\} = P^*. \tag{1.4.7}$$

For $\nu_i = \nu$ $(i = 1,\ldots,k)$, it turns to

$$\frac{1}{\Gamma(\frac{\nu}{2})} \int_0^{\nu c'} u^{\frac{\nu}{2}-1} e^{-u} du = (P^*)^{\frac{1}{k}} \tag{1.4.8}$$

The left hand side is the c.d.f. of $\frac{1}{2}\chi^2_\nu$ with ν degrees of freedom, hence the value c' can be easily solved with the help of a table of chi-square distribution.

Application to the Selection of Variance of Normal Populations

If $\theta_i = 2\sigma_i^2$, $i = 0, 1,\ldots,k$ are the scale parameters for the $k + 1$ normal populations and x_{ij} $(j = 1,\ldots,n;\ i = 1,\ldots,k)$ are the n observations on the population π_i with the mean μ_i (known). We assume that the order $\sigma_1^2 \leq \ldots \leq \sigma_k^2$ is known.

In the application of selection procedure δ_6 or δ_7, what we need to do is to evaluate the isotonic regression of S_i^2 which is the sample variance of population π_i, $i = 1,\ldots,k$ and denote it by $S^2_{i:k}$, $i = 1,\ldots,k$ then directly apply δ_6 or δ_7. The constant we need is determined by Equations (1.4.2) and (1.4.4) where we replace ν by n, the reason being that $2nS_i^2/\theta_i$ has χ^2_n distribution with n degrees of freedom and $Y_i \quad S^2/\theta_i$ has the c.d.f.

$$G(2nt;\ n,\ 2) = G(t;\ n,\ \tfrac{1}{n}),$$

hence

$$P(\frac{1}{r}\sum_{i=1}^{r} Y_i \leq t) = G(trn;\ rn,\ 1).$$

The application of δ_8 is similar to that of δ_9 (see Gupta and Sobel (1958)). What we need to do is to replace X_i in δ_8 and δ_9 by S_i and replace ν in Equation (1.4.6) and (1.4.8) by n, $i = 1,\ldots,k$.

Remark 1.4.1. δ_9 (Gupta and Sobel (1958)) does not depend on the ordering prior and the sample sizes for each population need not be equal.

If the means μ_i, $i = 1,\ldots,k$ are unknown and common sample size $n > 1$, let $S_i^2 = \sum_{j=1}^{n} (X_{ij} - X_i)^2/n-1$ and use $n - 1$ in place of ν Equation (1.4.4), (1.4.6) and (1.4.8) which determined the constants $c_{i:k}$'s, c and c' for δ_6, δ_8, and δ_9, respectively.

1.4.2. Selection Procedure $\delta_i^{(2)}$, i = 6, 7, 8, 9

Case II. θ_0 unknown.

The assumptions are the same as in Case I except that n_0 observations, viz., $x_{01},\ldots,x_{0n_0}$ are taken on π_0.

For selection procedure $\delta_6^{(2)}$, the inequalities defining the procedure and corresponding to $\hat{X}_{i:k} \leq c_{i:k}\, \theta_0$ $(i = 1,\ldots,k)$ is replaced by $\hat{X}_{i:k} \leq c_{i:k}^{(2)}\, X_0$ $(i = 1,\ldots,k)$, respectively. The equation determining $c_{i:k}^{(2)}$ is obtained as before and is given by

$$\int_0^\infty P(U_i \leq c_{i:k}^{(2)}\, t) f(t)dt = P^* \qquad (1.4.9)$$

where $P(U_i \leq x)$ is the same as that in Thoerem 1.4.1, and f(t) is the p.d.f. of X_0 of population π_0.

If population π_0 has gamma distribution with density $g(x; \alpha_0, \theta_0)$ (α_0 known and θ_0 unknown), then $f(x) = g(x; n_0\alpha_0, \frac{1}{n_0})$.

For selection procedure $\delta_7^{(2)}$, the inequality defining the procedure is

$$\hat{X}_{k:k} < c' X_0$$

and it can be shown $c' = c^{(2)}_{k:k}$.

For selection procedure $\delta^{(2)}_8$, the inequality defining the procedure and corresponding the (k-i)th step is

$$\frac{X_i}{\nu_i} \leq c \frac{X_0}{\nu_0} \quad \text{where } \nu_0 = n_0\alpha_0. \tag{1.4.11}$$

The equation determining c of $\delta^{(2)}_8$ is given by

$$\int_0^\infty \int_0^{\frac{\nu_i}{\nu_0}ct} \frac{u^{\frac{\nu_i}{2}-1}e^{-u}}{\Gamma(\frac{\nu_i}{2})}\,du\,\frac{t^{\frac{\nu_0}{2}-1}e^{-t}}{\Gamma(\frac{\nu_0}{2})}\,dt = P^*. \tag{1.4.12}$$

For selection procedure $\delta^{(2)}_9$, the inequality defining the procedure is

$$\frac{X_i}{\nu_i} \leq c \frac{X_0}{\nu_0}, \tag{1.4.13}$$

and the equation determining d is given by Gupta and Sobel (1958) as follows:

$$\int_0^\infty \left[\prod_{i=1}^{k} \int_0^{\frac{\nu_i}{\nu_0}tc} \frac{u^{\frac{\nu_i}{2}-1}e^{-u}}{\Gamma(\frac{\nu_i}{2})}\,du\right] \frac{t^{\frac{\nu_0}{2}-1}e^{-t}}{\Gamma(\frac{\nu_0}{2})}\,dt = P^*. \tag{1.4.14}$$

1.5. Selection Rules for the Location Parameter Under Partial Ordering Prior Assumption

Assume that we have only a partial ordering prior of k unknown location parameters, that is the parameter space

$\Omega = \{\underline{\theta} | \underline{\theta} \in \mathbb{R}^k$ and there is a partial order relation "$\lesssim$" among θ_i's$\}$

Our approach is to partition the set $\{\theta_1,\ldots,\theta_k\}$ into several subsets, say $B_0,\ldots,B_\ell$, so that $B_i \cap B_j = \emptyset$, if $i \neq j$, $\bigcup_{j=1}^{\ell} B_j = \{\theta_1,\ldots,\theta_k\}$ and for each B_j $(j = 1,\ldots,\ell)$ there is a simple order on it and there is no order relation among the elements of subset B_0.

Let $b_i = |B_i|$, the number of elements contained in B_i, $i = 0,\ldots,\ell$, so we have

$$\sum_{i=0}^{\ell} b_i = k.$$

If we denote the new induced partial order by "$\lesssim'$", then we have a parameter space $\Omega' \supset \Omega$. We use an example to illustrate how to find an induced partial order.

Example. Suppose $k = 8$, and we have a partial ordering prior $\theta_1 \lesssim \theta_5$, $\theta_1 \lesssim \theta_8$, $\theta_1 \lesssim \theta_2 \lesssim \theta_3 \lesssim \theta_4$, and $\theta_2 \lesssim \theta_6 \lesssim \theta_7$. We use a "tree" to represent this partial ordering as in Figure 1.

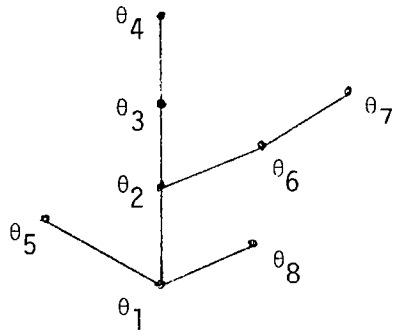

Figure 1. Original partial ordering

Then we have an induced partial ordering $\theta_1 \lesssim \theta_2 \lesssim \theta_3 \lesssim \theta_4$, $\theta_6 \lesssim \theta_7$ as in Figure 2.

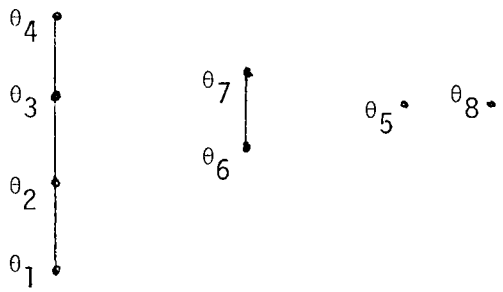

Figure 2. Induced partial ordering.

And

$$B_0 = \{\theta_5, \theta_8\}$$

$$B_1 = \{\theta_1, \theta_2, \theta_3, \theta_4\}$$

$$B_2 = \{\theta_6, \theta_7\}.$$

It is clear that the induced partial order is not unique, for example, we can partition $\{\theta_1,\ldots,\theta_8\}$ into three other subsets B_0', B_1', B_2' where

$$B_0' = \{\theta_5, \theta_8\}$$

$$B_1' = \{\theta_1, \theta_2, \theta_6, \theta_7\}$$

$$B_2' = \{\theta_3, \theta_4\}.$$

For the location parameter case, a selection procedure δ^P can be defined as follows:

Definition 1.5.1. We define a selection procedure δ^P as follows:

Suppose $B_0,\ldots,B_\ell$ are the induced subsets and that for each subset B_j, $j = 1,\ldots,\ell$ there is a simple order on it. We choose a proper

selection procedure δ for each subset B_j, such that the corresponding probability of a correct selection is not less than $P_j^* = P^{*\frac{b_j}{k}}$. For subset B_0 we may use selection procedure δ_4 or δ_5 with $P_0^* = P^{*\frac{b_0}{k}}$.

Theorem 1.5.1. The selection procedure δ^P satisfies the P*-condition.

Proof.

$$\inf_{\theta\in\Omega} P_\theta(CS|\delta^P)$$

$$\geq \inf_{\theta\in\Omega'} P_\theta(CS|\delta^P)$$

$$\geq \prod_{i=1}^{\ell} \inf_{\Omega_{B_i}} P(CS|\delta^P)$$

$$\geq P^{*\sum_{i=0}^{\ell}\frac{b_i}{k}} = P^*$$

where Ω_{B_i} is the parameter space associated with the subset B_i.

Remark 1.5.1. For the selection problem of the scale parameter of the gamma-distributions, under partial ordering assumptions, an analogous selection procedure can be proposed easily, hence it is omitted.

1.6. Comparisons of the Performance of Basic Rules

1.6.1. The Location Parameters of Normal Distributions

In this section we use Monte Carlo techniques to compare the performance of selection procedures δ_1, δ_2, δ_3, and δ_4. Suppose we have

k independent populations, each population with distribution $N(\mu_i, \sigma^2)$, with common known variance σ^2 and common sample size n. Assume that the mean μ_0 of the control is known; without loss of generality we assume that $\mu_0 = 0$ and $\sigma/\sqrt{n} = 1$.

In the simulation, we use Rubin and Hinkle's RVP-Random Variable Package, Purdue University Computing Center, to generate random numbers. For each k, we generated one random number (variable) for each population, then applied each selection procedure separately and repeated it ten thousand times; we used the relative frequencies as an approximation of the exact values of the associated performance characteristics for each procedure. In Table V we use the following notations:

$\underline{\theta} = (\theta_1,\ldots,\theta_k)$, θ_i is the parameter of population π_i.

PS = P(CS)

PI = P(correctly rejecting all bad populations)

PC = P(correct classification of all population)

where the correct classification means that we select all good populations and reject all bad populations.

EI = Expected number (size) of bad populations contained in the selected subset.

$$EJ = \sum_{\theta_i<\theta_0} (\theta_i - \theta_0)^2 \, P(\pi_i \text{ is selected})$$

ES = Expected size of the selected subset.

Table V.1 consists of four parts, namely, the four values of k = 2, 3, 4, 5. For each value of k we assume that the first population is the one and only one bad population with parameter -1 which is less than the control $\theta_0 = 0$. A glance at the table indicates that the performance can roughly be ordered as follows:

$$\delta_3 \succsim [\delta_2, \delta_1] \succ \delta_4$$

i.e. procedure δ_3 is the best one and slightly better than δ_2 and δ_1, δ_2 and δ_1 are very close and both are better than δ_4. The performance is based on the characteristics PI, PC, EI, and EJ which are explained in Section 1.6.1. As the number of populations k increases from two to five and the three additional populations are good populations with parameter 1, 2, and 3, respectively, we find that $EI(\delta_i, k=5)-EI(\delta_i, k=2)$ $(i = 1, 2, 3, 4)$ is 0.0124, 0.0124, 0.0031, and 0.1211 respectively. This means that when k increases and the additional populations are good, then δ_4 is the most sensitive procedure with k.

Table V.2 has the same structure as Table V.1 but in Table V.2 we assume that we have two bad populations for each k. In this case based on the performance characteristics PI, PC, EI or $\tilde{EJ}$, we have the performance ordering as follows:

$$\delta_1 \succ \delta_2 \succ \delta_3 \succ \delta_4.$$

In Table V.3 we assume that we have three bad populations for $k \geq 3$, and that both populations are bad for k = 2, this table indicates the same trend as Table V.2, i.e. $\delta_1 \succ \delta_2 \succ \delta_3 \succ \delta_4$. If k is increased by adding strictly good (parameter strictly larger than control)

populations, then $EI(\delta_i)$, $i = 1,2$ does not increase but decreases. This is because $\hat{\hat{X}}_{k:k} > \hat{\hat{X}}_{k:k+1}$ a.s. $1 \leq i \leq k$.

In Table V.4 we assume for each k, $k = 2, 3, 4, 5$ that every population is bad. Based on the quantities PI, PC, EI, and EJ, we find that performance is as follows: $\delta_1 \succ \delta_2 \succ \delta_3 \succ \delta_4$. This is the same result as before.

In Table V.5 we assume that the ordering prior of unknown parameter is incorrect; i.e. the true configuration (-2, -1, 0, 1, 2) is replaced by (-1, -2, 1, 0, 1, 2). The simulation results indicate that, based on PI, PC, EI and EJ we have performance $\delta_1 \succ \delta_2 \succ [\delta_3, \delta_4]$. Thus here again δ_1 is the best. If we compare Table V.5 with Table V.2, we see that δ_4 does not change (the small differences are because of random fluctuations), $EI(\delta_3)$ and $EJ(\delta_3)$ increase quite appreciable. For example, in case $k = 5$, $EI(\delta_3)$ the increase is 0.3737 and $EJ(\delta_3)$ increases by 1.7534; $EI(\delta_2)$ decreases by 0.0375, and $EI(\delta_1)$ decreases by 0.0374; $EJ(\delta_2)$ increases by 0.7173 and $EJ(\delta_1)$ increases by 0.7171.

From these five tables, it appears that the over all performance of these procedures is: $\delta_1 \succ \delta_2 \succ \delta_3 \succ \delta_4$, if the ordering prior is correct; otherwise for an incorrect ordering procedure δ_4 seems to be more stable than the others.

1.6.2. Performance of Selection Rules for the Scale Parameters of Gamma Distributions

In this section we again use Monte Carlo techniques to compare the selection procedures δ_6, δ_7, δ_8, and δ_9. We assume that each of

populations has exponential distribution (gamma distribution with $\alpha = 2$) with unknown scale parameters $\theta_1,\ldots,\theta_k$, where $\theta_1 \leq \ldots \leq \theta_k$.

In the simulation study, the common sample size used four, and we used Rubin and Hinkle's RVP sub-program to generate random numbers (variables) with specified exponential distribution. The simulation results are based on two thousand sets of random numbers. The notations in Table VI.1 and Table VI.2 are the same as those in Section 1.6.1 but the meaning of a bad population is different in that now a bad population has a parameter which is greater than the control (assumed one).

A good population means that its parameter is less than or equal to one. The results of Table VI.1 and Table VI.2 uniformly indicate that we have the performance $\delta_6 \succ \delta_7 \dot{\succ} \delta_8 \succ \delta_9$, one exception being that δ_8 is slightly better than δ_7 if there is only one bad population.

TABLE I

Table of $d^{(1)}_{1:k}$ values (satisfying (1.3.5) and (1.3.21)) necessary to carry out the procedure δ_1 for the normal means problem under the simple ordering prior.

$d^{(1)}_{1:k}$ \ k	P^* = .99	.975	.95	.925	.90
1	2.3264	1.9600	1.6449	1.4395	1.2816
2	2.3337	1.9775	1.6780	1.4872	1.3430
3	2.3339	1.9787	1.6817	1.4942	1.3538
4	--	--	1.6823	1.4956	1.3563
5	--	--	1.6824	1.4960	1.3571
6	--	--	--	--	1.3573
∞	2.3340	1.9787	1.6824	1.4960	1.3574
$d^{(1)}_{1:k}$ \ k	P^* = .85	.80	.75	.70	.65
1	1.0364	.8416	.6745	.5244	.3853
2	1.1239	.9537	.8104	.6836	.5677
3	1.1428	.9815	.8473	.7300	.6238
4	1.1485	.9911	.8615	.7491	.6482
5	1.1506	.9951	.8679	.7584	.6607
6	1.1514	.9969	.8710	.7633	.6678
7	1.1517	.9977	.8727	.7661	.6720
10	1.1518	.9984	.8744	.7688	.6775
∞	1.1519	.9985	.8749	.7706	.6800

The '--' in Table I means that the value is the same as the preceding one in the same column.

TABLE II

Table of $d^{(2)}_{1:k}$ values (satisfying (1.3.22)) necessary to carry out the procedure $\delta^{(2)}_1$ for the normal means problem under simple ordering prior.

$d^{(2)}_{1:k}$ / k	P^*				
	.99	.975	.95	.925	.90
1	3.2886	2.7711	2.3258	2.0355	1.8122
2	3.3449	2.8494	2.4267	2.1530	1.9434
3	3.3605	2.8730	2.4589	2.1917	1.9874
4	3.3673	2.8840	2.4723	2.2105	2.0091
5	3.3711	2.8901	2.4832	2.2215	2.0219
6	3.3734	2.8941	2.4890	2.2286	2.0303
8	3.3761	2.8988	2.4960	2.2375	2.0406
10	3.3776	2.9014	2.5000	2.2426	2.0440
∞	3.3787	2.9032	2.5021	2.2448	2.0487
$d^{(2)}_{1:k}$ / k	P^*				
	.85	.80	.75	.70	.65
1	1.4656	1.1901	.9538	.7416	.5449
2	1.6198	1.3639	1.1453	.9496	.7689
3	1.6728	1.4247	1.2131	1.0240	.8495
4	1.6993	1.4554	1.2475	1.0620	.8909
5	1.7151	1.4738	1.2683	1.0850	.9161
6	1.7259	1.4860	1.2822	1.1004	.9330
8	1.7386	1.5013	1.2996	1.1198	.9543
10	1.7463	1.5105	1.3100	1.1285	.9618
∞	1.7476	1.5108	1.3100	1.1287	.9622

TABLE III. 1

Table of $d^{(3)}_{i:k} = D(i:k)$ values (satisfying (1.3.30)) necessary to carry out the procedure $\delta^{(3)}_1$ for the normal means problem with common sample size n (common variance unknown) under simple ordering prior.

n = 3

P^* =	.990	.975	.950	.925	.900	.850	.800	.750
D(1:2) =	6.6993	4.9977	3.8822	3.2780	2.8643	2.2934	1.8896	1.5716
D(2:2) =	6.4966	4.8106	3.6930	3.0802	2.6558	2.0605	1.6298	1.2830
D(1:3) =	4.5435	3.5731	2.8815	2.4867	2.2081	1.8134	1.5278	1.3001
D(2:3) =	4.5272	3.5534	2.8567	2.4569	2.1735	1.7688	1.4730	1.2345
D(3:3) =	4.4443	3.4604	2.7482	2.3337	2.0361	1.6039	1.2809	1.0148
D(1:4) =	3.6058	2.8997	2.3763	2.0701	1.8512	1.5374	1.3085	1.1252
D(2:4) =	3.6037	2.8965	2.3713	2.0633	1.8424	1.5244	1.2910	1.1027
D(3:4) =	3.5958	2.8853	2.3553	2.0427	1.8173	1.4901	1.2471	1.0490
D(4:4) =	3.5480	2.8241	2.2776	1.9501	1.7108	1.3572	1.0887	.8651
D(1:5) =	3.0674	2.4966	2.0637	1.8068	1.6216	1.3548	1.1594	1.0028
D(2:5) =	3.0670	2.4959	2.0625	1.8049	1.6189	1.3502	1.1525	.9934
D(3:5) =	3.0659	2.4940	2.0590	1.7999	1.6123	1.3398	1.1380	.9742
D(4:5) =	3.0613	2.4865	2.0473	1.7840	1.5922	1.3113	1.1007	.9278
D(5:5) =	3.0281	2.4406	1.9854	1.7081	1.5032	1.1974	.9630	.7666
D(1:6) =	2.7100	2.2230	1.8473	1.6223	1.4593	1.2234	1.0503	.9117
D(2:6) =	2.7099	2.2228	1.8470	1.6216	1.4583	1.2216	1.0474	.9072
D(3:6) =	2.7098	2.2224	1.8461	1.6202	1.4563	1.2178	1.0415	.8990
D(4:6) =	2.7091	2.2211	1.8436	1.6163	1.4509	1.2090	1.0289	.8821
D(5:6) =	2.7060	2.2156	1.8343	1.6033	1.4340	1.1842	.9959	.8406
D(6:6) =	2.6810	2.1788	1.7823	1.5379	1.3562	1.0832	.8726	.6955

TABLE III.2

Table of $d^{(3)}_{i:k} = D(i:k)$ values (satisfying (1.3.30)) necessary to carry out the procedure $\delta^{(3)}_1$ for the normal means problem with common sample size n (Common variance unknown) under simple ordering prior.

n = 5

P^* =	.990	.975	.950	.925	.900	.850	.800	.750
D(1:2) =	5.0853	4.0805	3.3310	2.8889	2.5700	2.1073	1.7637	1.4836
D(2:2) =	5.0177	3.9939	3.2210	2.7578	2.4194	1.9194	1.5396	1.2235
D(1:3) =	3.8313	3.1412	2.6072	2.2858	2.0518	1.7098	1.4550	1.2474
D(2:3) =	3.8269	3.1334	2.5941	2.2674	2.0279	1.6747	1.4084	1.1888
D(3:3) =	3.7914	3.0813	2.5205	2.1750	1.9180	1.5319	1.2341	.9836
D(1:4) =	3.1905	2.6405	2.2073	1.9440	1.7514	1.4695	1.2597	1.0892
D(2:4) =	3.1901	2.6397	2.2050	1.9401	1.7457	1.4597	1.2451	1.0694
D(3:4) =	3.1877	2.6347	2.1956	1.9261	1.7271	1.4312	1.2064	1.0201
D(4:4) =	3.1640	2.5965	2.1383	1.8520	1.6372	1.3119	1.0590	.8452
D(1:5) =	2.7891	2.3203	1.9470	1.7186	1.5513	1.3061	1.1238	.9762
D(2:5) =	2.7891	2.3201	1.9465	1.7178	1.5497	1.3028	1.1183	.9680
D(3:5) =	2.7889	2.3194	1.9448	1.7147	1.5451	1.2945	1.1058	.9507
D(4:5) =	2.7873	2.3158	1.9374	1.7033	1.5295	1.2701	1.0720	.9073
D(5:5) =	2.7693	2.2851	1.8893	1.6399	1.4518	1.1656	.9421	.7525
D(1:6) =	2.5082	2.0935	1.7608	1.5566	1.4064	1.1864	1.0230	.8910
D(2:6) =	2.5082	2.0935	1.7607	1.5563	1.4059	1.1851	1.0207	.8872
D(3:6) =	2.5081	2.0934	1.7604	1.5555	1.4046	1.1823	1.0158	.8799
D(4:6) =	2.5080	2.0929	1.7590	1.5530	1.4007	1.1751	1.0047	.8644
D(5:6) =	2.5068	2.0900	1.7528	1.5432	1.3871	1.1533	.9744	.8253
D(6:6) =	2.4922	2.0640	1.7109	1.4871	1.3179	1.0593	.8569	.6848

TABLE III. 3

Table of $d^{(3)}_{i:k} = D(i:k)$ values (satisfying (1.3.30)) necessary to carry out the procedure $\delta^{(3)}_1$ for the normal means problem with common sample size n (common variance unknown) under simple ordering prior.

n = 9

P^* =	.900	.975	.950	.925	.900	.850	.800	.750
D(1:2) =	4.5081	3.7260	3.1050	2.7240	2.4424	2.0241	1.7061	1.4426
D(2:2) =	4.4746	3.6720	3.0240	2.6191	2.3154	1.8553	1.4976	1.1953
D(1:3) =	3.5468	2.9598	2.4876	2.1963	1.9809	1.6618	1.4208	1.2224
D(2:3) =	3.5452	2.9557	2.4789	2.1824	1.9616	1.6311	1.3780	1.1671
D(3:3) =	3.5245	2.9189	2.4196	2.1031	1.8637	1.4981	1.2118	.9685
D(1:4) =	3.0155	2.5272	2.1312	1.8862	1.7051	1.4375	1.2364	1.0719
D(2:4) =	3.0153	2.5267	2.1298	1.8835	1.7008	1.4291	1.2232	1.0533
D(3:4) =	3.0143	2.5238	2.1231	1.8724	1.6850	1.4034	1.1869	1.0061
D(4:4) =	2.9990	2.4947	2.0746	1.8064	1.6027	1.2903	1.0447	.8356
D(1:5) =	2.6678	2.2412	1.8935	1.6777	1.5183	1.2829	1.1067	.9633
D(2:5) =	2.6678	2.2412	1.8932	1.6771	1.5171	1.2801	1.1018	.9557
D(3:5) =	2.6677	2.2409	1.8921	1.6749	1.5134	1.2729	1.0903	.9393
D(4:5) =	2.6670	2.2385	1.8865	1.6654	1.4998	1.2504	1.0582	.8974
D(5:5) =	2.6546	2.2140	1.8446	1.6078	1.4274	1.1503	.9319	.7456
D(1:6) =	2.4174	2.0344	1.7206	1.5256	1.3814	1.1686	1.0098	.8809
D(2:6) =	2.4174	2.0344	1.7206	1.5255	1.3811	1.1676	1.0077	.8774
D(3:6) =	2.4175	2.0344	1.7204	1.5250	1.3801	1.1652	1.0033	.8706
D(4:6) =	2.4174	2.0341	1.7195	1.5231	1.3769	1.1588	.9929	.8557
D(5:6) =	2.4169	2.0322	1.7146	1.5147	1.3647	1.1384	.9639	.8178
D(6:6) =	2.4064	2.0106	1.6772	1.4630	1.2995	1.0478	.8492	.6796

TABLE III. 4

Table of $d^{(3)}_{i:k}$ = D(i:k) values (satisfying (1.3.30)) necessary to carry out the procedure $\delta^{(3)}_1$ for the normal means problem with common sample size n (common variance unknown) under simple ordering prior.

n = 21

P^* =	.990	.975	.950	.925	.900	.850	.800	.750
D(1:2) =	4.2169	3.5394	2.9828	2.6332	2.3714	1.9770	1.6732	1.4190
D(2:2) =	4.1973	3.5006	2.9165	2.5422	2.2570	1.8187	1.4735	1.1790
D(1:3) =	3.3948	2.8606	2.4208	2.1456	1.9405	1.6342	1.4010	1.2078
D(2:3) =	3.3940	2.8583	2.4142	2.1342	1.9238	1.6058	1.3603	1.1544
D(3:3) =	3.3801	2.8290	2.3627	2.0622	1.8326	1.4785	1.1988	.9597
D(1:4) =	2.9192	2.4639	2.0881	1.8532	1.6785	1.4189	1.2228	1.0618
D(2:4) =	2.9192	2.4637	2.0872	1.8511	1.6749	1.4113	1.2104	1.0439
D(3:4) =	2.9187	2.4616	2.0818	1.8415	1.6607	1.3871	1.1755	.9979
D(4:4) =	2.9074	2.4374	2.0381	1.7801	1.5827	1.2777	1.0363	.8299
D(1:5) =	2.5999	2.1964	1.8628	1.6541	1.4991	1.2693	1.0967	.9557
D(2:5) =	2.5999	2.1964	1.8627	1.6537	1.4982	1.2669	1.0921	.9485
D(3:5) =	2.5999	2.1962	1.8619	1.6519	1.4950	1.2602	1.0811	.9326
D(4:5) =	2.5995	2.1945	1.8572	1.6435	1.4825	1.2388	1.0501	.8916
D(5:5) =	2.5899	2.1733	1.8187	1.5891	1.4132	1.1413	.9259	.7416
D(1:6) =	2.3667	2.0006	1.6975	1.5077	1.3668	1.1582	1.0020	.8750
D(2:6) =	2.3667	2.0006	1.6974	1.5076	1.3665	1.1574	1.0001	.8717
D(3:6) =	2.3667	2.0006	1.6973	1.5072	1.3657	1.1552	.9960	.8651
D(4:6) =	2.3667	2.0004	1.6966	1.5056	1.3629	1.1492	.9860	.8506
D(5:6) =	2.3664	1.9990	1.6924	1.4981	1.3516	1.1297	.9578	.8133
D(6:6) =	2.3578	1.9801	1.6577	1.4488	1.2887	1.0409	.8446	.6765

TABLE IV

The value $c_{k:k}$, $1 \leq k \leq 10$ (satisfying (1.4.2)) necessary to carry out the procedure δ_6 for the gamma means problem (with common sample size n and common α) under simple ordering prior.

$P^* = 0.99$

α	k	n					
		4	6	8	10	15	20
2	1	2.5113	2.1849	2.0000	1.8782	1.6964	1.5924
	2	2.5217	2.1922	2.0056	1.8829	1.6998	1.5951
	3	2.5222	2.1924	2.0059	1.8831	1.7000	1.5952
	10	--	--	--	--	--	--
4	1	4.0003	3.5817	3.3429	3.1845	2.9460	2.8081
	2	4.0113	3.5899	-.3494	3.1900	2.9501	2.8116
	3	4.0117	3.5902	3.3497	3.1902	--	--
	4	4.0118	--	--	--	--	--
	10	--	--	--	--	--	--
6	1	5.3725	4.8851	4.6050	4.4190	4.1372	3.9737
	2	5.3846	4.8939	4.6122	4.4252	4.1419	3.9776
	3	5.3851	4.8942	4.6125	4.4255	4.1421	3.9777
	4	--	--	--	--	--	--
	10	--	--	--	--	--	--

The '--' in Table IV means that the value is the same as the preceding one in the same column.

TABLE IV (continued)

$P^* = 0.95$

α	k	n					
		4	6	8	10	15	20
2	1	1.9385	1.7522	1.6435	1.5705	1.4591	1.3940
	2	1.9746	1.7783	1.6644	1.5882	1.4724	1.4048
	3	1.9797	1.7818	1.6672	1.5906	1.4741	1.4062
	4	1.9807	1.7825	1.6677	1.5910	1.4744	1.4065
	5	1.9809	1.7826	1.6678	1.5911	--	--
	6	1.9810	1.7827	1.6679	--	--	--
	10	--	--	--	--	--	--
4	1	3.2870	3.0346	2.8872	2.7879	2.6361	2.5470
	2	3.3288	3.0656	2.9125	2.8097	2.6527	2.5608
	3	3.3344	3.0697	2.9157	2.8124	2.6548	2.5625
	4	3.3354	3.0704	2.9163	2.8129	2.6551	2.5628
	5	3.3357	3.0705	2.9164	2.8130	2.6552	2.5629
	10	--	--	--	--	--	--
6	1	4.5519	4.2499	4.0732	3.9541	3.7715	3.6642
	2	4.5984	4.2850	4.1020	3.9790	3.7908	3.6803
	3	4.6045	4.2894	4.1057	3.9821	3.7931	3.6823
	4	4.6055	4.2901	4.1063	3.9826	3.7935	3.6826
	5	4.6058	4.2903	4.1064	3.9828	3.7936	3.6827
	10	--	--	--	--	--	--

TABLE IV (continued)

$p^* = 0.90$

α	k	n					
		4	6	8	10	15	20
2	1	1.6702	1.5458	1.4714	1.4206	1.3419	1.2951
	2	1.7285	1.5888	1.5063	1.4505	1.3646	1.3139
	3	1.7406	1.5974	1.5132	1.4563	1.3689	1.3175
	4	1.7440	1.5998	1.5151	1.4579	1.3700	1.3184
	5	1.7451	1.6005	1.5156	1.4584	1.3704	1.3187
	6	1.7455	1.6008	1.5158	1.4585	1.3705	1.3188
	7	1.7456	1.6009	1.5159	1.4586	--	--
	8	1.7457	--	--	--	--	--
	10	--	--	--	--	--	--
4	1	2.9427	2.7667	2.6615	2.5903	2.4799	2.4145
	2	3.0126	2.8191	2.7051	2.6279	2.5090	2.4389
	3	3.0264	2.8293	2.7133	2.6350	2.5144	2.4434
	4	3.0301	2.8320	2.7155	2.6368	2.5158	2.4445
	5	3.0313	2.8328	2.7162	2.6374	2.5162	2.4448
	6	3.0317	2.8331	2.7164	2.6376	2.5164	2.4449
	7	3.0318	2.8332	2.7165	2.6377	--	2.4450
	8	3.0319	--	--	--	--	--
	10	--	--	--	--	--	--
6	1	4.1495	3.9344	3.8064	3.7199	3.5855	3.5058
	2	4.2286	3.9948	3.8569	3.7635	3.6196	3.5385
	3	4.2439	4.0062	3.8663	3.7717	3.6259	3.5398
	4	4.2480	4.0092	3.8688	3.7737	3.6274	3.5411
	5	4.2492	4.0101	3.8695	3.7744	3.6279	3.5415
	6	4.2496	4.0104	3.8697	3.7746	3.6281	3.5416
	7	4.2498	4.0105	3.8698	3.7747	--	3.5417
	10	--	--	--	--	--	--

TABLE IV (continued)

$P^* = 0.75$

α	k	n					
		4	6	8	10	15	20
2	1	1.2774	1.2371	1.2106	1.1914	1.1600	1.1404
	2	1.3798	1.3155	1.2757	1.2480	1.2041	1.1776
	3	1.4108	1.3387	1.2948	1.2645	1.2168	1.1882
	4	1.4237	1.3483	1.3026	1.2711	1.2219	1.1924
	5	1.4300	1.3529	1.3063	1.2743	1.2243	1.1944
	6	1.4333	1.3553	1.3083	1.2760	1.2255	1.1954
	7	1.4353	1.3567	1.3093	1.2769	1.2262	1.1959
	8	1.4364	1.3575	1.3100	1.2774	1.2266	1.1963
	9	1.4371	1.3579	1.3103	1.2777	1.2268	1.1965
	10	1.4375	1.3582	1.3106	1.2779	1.2270	1.1966
4	1	2.4211	2.3534	2.3108	2.2808	2.2327	2.2033
	2	2.5514	2.4546	2.3957	2.3551	2.2913	2.2530
	3	2.5896	2.4838	2.4201	2.3763	2.3079	2.2669
	4	2.6052	2.4956	2.4299	2.3848	2.3145	2.2725
	5	2.6126	2.5012	2.4344	2.3887	2.3175	2.2750
	6	2.6165	2.5041	2.4368	2.3908	2.3191	2.2763
	7	2.6187	2.5057	2.4381	2.3919	2.3200	2.2770
	8	2.6199	2.5066	2.4389	2.3925	2.3204	2.2774
	9	2.6207	2.5072	2.4393	2.3929	2.3207	2.2777
	10	2.6212	2.5075	2.4395	2.3931	2.3209	2.2778
6	1	3.5301	3.4420	3.3873	3.3491	3.2883	3.2514
	2	3.6818	3.5606	3.4874	3.4370	3.3581	3.3107
	3	3.7257	3.5946	3.5158	3.4618	3.3776	3.3273
	4	3.7435	3.6082	3.5272	3.4717	3.3854	3.3338
	5	3.7518	3.6146	3.5324	3.4763	3.3889	3.3368
	6	3.7562	3.6178	3.5351	3.4786	3.3908	3.3383
	7	3.7586	3.6196	3.5366	3.4799	3.3917	3.3392
	8	3.7600	3.6207	3.5375	3.4806	3.3923	3.3396
	9	3.7608	3.6213	3.5379	3.4811	3.3926	3.3399
	10	3.7613	3.6216	3.5382	3.4813	3.3928	3.3400

TABLE V. 1

Simulation results for the comparative performance of various selection procedures for the normal means problem (notation explained in Section 1.6.1) under simple ordering prior.

$P^* = .900$

$k = 2$, $\underline{\theta} = (-1,0)$				
	δ_1	δ_2	δ_3	δ_4
PS	.9453	.9490	.9579	.9470
PI	.3854	.3854	.3937	.2676
PC	.3307	.3344	.3516	.2530
EI	.6146	.6146	.6063	.7324
EJ	.6146	.6146	.6063	.7324
ES	1.5599	1.5636	1.5642	1.6794
$k = 3$, $\underline{\theta} = (-1,0,1)$				
	δ_1	δ_2	δ_3	δ_4
PS	.9531	.9535	.9638	.9616
PI	.3741	.3741	.3826	.2044
PC	.3272	.3276	.3464	.1970
EI	.6259	.6259	.6174	.7956
EJ	.6259	.6259	.6174	.7956
ES	2.5771	2.5777	2.5803	2.7574
$k = 4$, $\underline{\theta} = (-1,0,1,2)$				
	δ_1	δ_2	δ_3	δ_4
PS	.9580	.9582	.9640	.9765
PI	.3664	.3664	.3834	.1683
PC	.3244	.3246	.3474	.1640
EI	.6336	.6336	.6166	.8317
EJ	.6336	.6336	.6166	.8317
ES	3.5902	3.5904	3.5801	3.8081
$k = 5$, $\underline{\theta} = (-1,0,1,2,3)$				
	δ_1	δ_2	δ_3	δ_4
PS	.9554	.9554	.9623	.9794
PI	.3730	.3730	.3906	.1465
PC	.3284	.3284	.3529	.1431
EI	.6270	.6270	.6094	.8535
EJ	.6270	.6270	.6094	.8535
ES	4.5812	4.5812	4.5714	4.8329

TABLE V. 2

Simulation results for the comparative performance of various selection procedures for the normal means problem (notation explained in Section 1.6.1) under simple ordering prior.

$P^* = .900$

$k = 2$, $\underline{\theta} = (-2,-1)$				
	δ_1	δ_2	δ_3	δ_4
PS	1.0000	1.0000	1.0000	1.0000
PI	.3420	.3252	.3001	.1719
PC	.3420	.3252	.3001	.1719
EI	.8673	.8841	.9389	1.0950
EJ	1.4952	1.5120	1.6559	2.1831
ES	.8673	.8841	.9389	1.0950
$k = 3$, $\underline{\theta} = (-2,-1,0)$				
	δ_1	δ_2	δ_3	δ_4
PS	.9535	.9573	.9696	.9632
PI	.3437	.3407	.3007	.1233
PC	.2972	.2980	.2703	.1175
EI	.8585	.8615	.9350	1.2126
EJ	1.4651	1.4681	1.6421	2.4996
ES	1.8120	1.8188	1.9046	2.1758
$k = 4$, $\underline{\theta} = (-2,-1,0,1)$				
	δ_1	δ_2	δ_3	δ_4
PS	.9596	.9606	.9715	.9747
PI	.3269	.3254	.2936	.0874
PC	.2865	.2860	.2651	.0851
EI	.8802	.8817	.9431	1.3062
EJ	1.5015	1.5030	1.6532	2.7378
ES	2.8387	2.8412	2.9142	3.2808
$k = 5$, $\underline{\theta} = (-2,-1,0,1,2)$				
	δ_1	δ_2	δ_3	δ_4
PS	.9562	.9564	.9690	.9765
PI	.3333	.3331	.2984	.0746
PC	.2895	.2895	.2674	.0725
EI	.8835	.8837	.9480	1.3712
EJ	1.5339	1.5341	1.6872	2.9450
ES	3.8386	3.8390	3.9167	4.3477

TABLE V. 3

Simulation results for the comparative performance of various selection procedures for the normal means problem (notation explained in Section 1.6.1) under simple ordering prior.

$P^* = .900$

$k = 2$, $\underline{\theta} = (-3,-2)$				
	δ_1	δ_2	δ_3	δ_4
PS	1.0000	1.0000	1.0000	1.0000
PI	.7551	.7380	.7342	.5912
PC	.7551	.7380	.7342	.5912
EI	.2632	.2803	.3035	.4395
EJ	1.1443	1.2127	1.4025	2.1590
ES	.2632	.2803	.3035	.4395
$k = 3$, $\underline{\theta} = (-3,-2,-1)$				
	δ_1	δ_2	δ_3	δ_4
PS	1.0000	1.0000	1.0000	1.0000
PI	.3362	.3156	.2837	.1090
PC	.3362	.3156	.2837	.1090
EI	.8937	.9166	1.0275	1.3290
EJ	1.6654	1.6952	2.1746	3.5318
ES	.8937	.9166	1.0275	1.3290
$k = 4$, $\underline{\theta} = (-3,-2,-1,0)$				
	δ_1	δ_2	δ_3	δ_4
PS	.9579	.9616	.9737	.9731
PI	.3257	.3225	.2801	.0759
PC	.2836	.2841	.2538	.0736
EI	.9118	.9160	1.0419	1.4675
EJ	1.7093	1.7165	2.2324	4.1380
ES	1.8697	1.8776	2.0156	2.4406
$k = 5$, $\underline{\theta} = (-3,-2,-1,0,1)$				
	δ_1	δ_2	δ_3	δ_4
PS	.9582	.9590	.9714	.9796
PI	.3292	.3281	.2877	.0602
PC	.2874	.2871	.2591	.0582
EI	.8962	.8976	1.0172	1.5283
EJ	1.6554	1.6577	2.1429	4.3912
ES	2.8536	2.8559	2.9884	3.5078

TABLE V. 4

Simulation results for the comparative performance of various selection procedures for the normal means problem (notation explained in Section 1.6.1) under simple ordering prior.

$P^* = .900$

$k = 2$, $\underline{\theta} = (-4,-3)$				
	δ_1	δ_2	δ_3	δ_4
PS	1.0000	1.0000	1.0000	1.0000
PI	.9613	.9560	.9585	.9130
PC	.9613	.9560	.9585	.9130
EI	.0392	.0445	.0448	.0876
EJ	.3563	.4040	.4263	.8493
ES	.0392	.0445	.0448	.0876
$k = 3$, $\underline{\theta} = (-4,-3,-2)$				
	δ_1	δ_2	δ_3	δ_4
PS	1.0000	1.0000	1.0000	1.0000
PI	.7587	.7359	.7300	.4997
PC	.7587	.7359	.7300	.4997
EI	.2599	.2835	.3201	.5574
EJ	1.1340	1.2324	1.5547	2.9908
ES	.2599	.2835	.3201	.5574
$k = 4$, $\underline{\theta} = (-4,-3,-2,-1)$				
	δ_1	δ_2	δ_3	δ_4
PS	1.0000	1.0000	1.0000	1.0000
PI	.3348	.3114	.2814	.0747
PC	.3348	.3114	.2814	.0747
EI	.9003	.9282	1.0440	1.4745
EJ	1.7013	1.7437	2.2947	4.3666
ES	.9003	.9282	1.0440	1.4745
$k = 5$, $\underline{\theta} = (-4,-3,-2,-1,-0.5)$				
	δ_1	δ_2	δ_3	δ_4
PS	1.0000	1.0000	1.0000	1.0000
PI	.1117	.1045	.0615	.0036
PC	.1117	.1045	.0615	.0036
EI	1.7460	1.7600	1.9734	2.4985
EJ	1.8147	1.8275	2.4965	5.0978
ES	1.7460	1.7600	1.9734	2.4985

TABLE V. 5

Simulation results for the comparative performance of various selection procedures for the normal means problem (notation explained in Section 1.6.1) under simple ordering prior.

$P^* = .900$

$k = 2$, $\underline{\theta} = (-1,-2)$				
	δ_1	δ_2	δ_3	δ_4
PS	1.0000	1.0000	1.0000	1.0000
PI	.5405	.5349	.2937	.1722
PC	.5405	.5349	.2937	.1722
EI	.8331	.8387	1.3151	1.0904
EJ	2.2116	2.2340	3.4232	2.1578
ES	.8331	.8387	1.3151	1.0904
$k = 3$, $\underline{\theta} = (-1,-2,1)$				
	δ_1	δ_2	δ_3	δ_4
PS	.9932	.9943	.9957	.9976
PI	.5365	.5349	.2987	.1190
PC	.5297	.5292	.2944	.1189
EI	.8347	.8363	1.3116	1.2154
EJ	2.2252	2.2316	3.4155	2.4919
ES	1.8279	1.8306	2.3073	2.2130
$k = 4$, $\underline{\theta} = (-1,-2,1,0)$				
	δ_1	δ_2	δ_3	δ_4
PS	.9921	.9923	.9973	.9746
PI	.5271	.5269	.2894	.0873
PC	.5192	.5192	.2867	.0849
EI	.8498	.8500	1.3235	1.3077
EJ	2.2685	2.2693	3.4553	2.7474
ES	2.8390	2.8395	3.3207	3.2822
$k = 5$, $\underline{\theta} = (-1,-2,1,0,2)$				
	δ_1	δ_2	δ_3	δ_4
PS	.9906	.9906	.9958	.9795
PI	.5317	.5316	.2937	.0711
PC	.5223	.5222	.2895	.0693
EI	.8461	.8462	1.3217	1.3593
EJ	2.2510	2.2514	3.4406	2.8830
ES	3.8341	3.8342	4.3173	4.3388

TABLE VI. 1

Simulation results for the comparative performance of various selection procedures for the gamma means problem (notation explained in Section 1.6.1) under simple ordering prior.

$P^* = .900$

$k = 4$, $\underline{\theta} = (.3,.6,.9,1.2)$				
	δ_6	δ_7	δ_8	δ_9
PS	.9840	.9845	.9865	.9890
PI	.2020	.1745	.1955	.0685
PC	.1860	.1590	.1820	.0670
EI	.7980	.8255	.8045	.9315
EJ	.0319	.0330	.0322	.0373
ES	3.7820	3.8100	3.7910	3.9205
$k = 5$, $\underline{\theta} = (.3,.6,.9,1.2,1.5)$				
	δ_6	δ_7	δ_8	δ_9
PS	.9905	.9910	.9965	.9885
PI	.0990	.0950	.0645	.0075
PC	.0895	.0860	.0610	.0075
EI	1.5040	1.5560	1.5890	1.7995
EJ	.1868	.1989	.2008	.2531
ES	4.4945	4.5470	4.5855	4.7880
$k = 4$, $\underline{\theta} = (.5,1,1.5,2)$				
	δ_6	δ_7	δ_8	δ_9
PS	.9685	.9700	.9795	.9765
PI	.2685	.2560	.2130	.0545
PC	.2370	.2260	.1925	.0535
EI	1.0790	1.1310	1.2110	1.4690
EJ	.5304	.5730	.6207	.8341
ES	3.0475	3.1010	3.1905	3.4455
$k = 5$, $\underline{\theta} = (.5,1,1.5,2,2.5)$				
	δ_6	δ_7	δ_8	δ_9
PS	.9690	.9695	.9875	.9785
PI	.2320	.2295	.1355	.0230
PC	.2010	.1990	.1230	.0230
EI	1.3605	1.4055	1.7555	2.0075
EJ	.9707	1.0520	1.4602	1.9862
ES	3.3295	3.3750	3.7430	3.9860

TABLE VI. 2

Simulation results for the comparative performance of various selection procedures for the gamma means problem (notation explained in Section 1.6.1) under simple ordering prior.

$P^* = .900$

$k = 4$, $\underline{\theta} = (.1,.6,1.1,1.6)$				
	δ_6	δ_7	δ_8	δ_9
PS	.9995	.9995	1.0000	.9995
PI	.0690	.0645	.0545	.0110
PC	.0685	.0640	.0545	.0110
EI	1.4940	1.5345	1.5315	1.7500
EJ	.2120	.2250	.2204	.2940
ES	3.4935	3.5340	3.5315	3.7495
$k = 5$, $\underline{\theta} = (.1,.6,1.1,1.6,2.1)$				
	δ_6	δ_7	δ_8	δ_9
PS	.9990	.9990	1.0000	.9995
PI	.0540	.0530	.0295	.0030
PC	.0530	.0520	.0295	.0030
EI	1.9205	1.9700	2.1385	2.4085
EJ	.6140	.6646	.7727	1.0577
ES	3.9195	3.9690	4.1385	4.4080
$k = 4$, $\underline{\theta} = (1.1,1.4,1.7,1.9)$				
	δ_6	δ_7	δ_8	δ_9
PS	1.0000	1.0000	1.0000	1.0000
PI	.0375	.0375	.0085	0
PC	.0375	.0375	.0085	0
EI	2.6890	2.7475	3.1155	3.2465
EJ	.6796	.7212	.8918	1.0608
ES	2.6890	2.7475	3.1155	3.2465
$k = 5$, $\underline{\theta} = (1.1,1.4,1.7,1.9,2.2)$				
	δ_6	δ_7	δ_8	δ_9
PS	1.0000	1.0000	1.0000	1.0000
PI	.0315	.0315	.0050	0
PC	.0315	.0315	.0050	0
EI	3.0780	3.1355	3.8600	3.9075
EJ	1.0652	1.1355	1.6638	1.9368
ES	3.0780	3.1355	3.8600	3.9075

CHAPTER II

BAYES - P^* SELECTION RULES FOR SELECTING A SUBSET CONTAINING THE BEST POPULATION

2.1. Introduction

Suppose we have $k(k \geq 2)$ independent populations $\pi_1,\ldots,\pi_k$ and that the random variable X_i associated with π_i has a distribution with unknown parameters θ_i, $i = 1,\ldots,k$. First, we give some definitions.

Definition 2.1.1. The populaiton π_i is the best population if $\theta_i \geq \theta_j$ for all $j \neq i$. If there are more than one populations satisfying this condition we arbitrarily tag one of them and call it the best one. A population which is not the "best" is called a "non-best" population.

Assume that we have n_i independent observations X_{ij}, $j = 1,\ldots,n_i$ for population π_i, $i = 1,\ldots,k$. Let $X_i = T_i(X_{i1},\ldots,X_{in_i})$ be a suitable estimator of θ_i, $i = 1,\ldots,k$, so that X_i's are indepedent. Usually X_i is a sufficient statistic for θ_i. Let $\underline{X} = (X_1,\ldots,X_k)$ and assume that for each θ_i, we have the (conditional) density $f(x_i;\theta_i)$ for population π_i as the probability density function of X_i, $i = 1,\ldots,k$.

The sample space is denoted by $\mathcal{X}$ where

$$\mathcal{X} = \{\underline{x} \mid \underline{x} = (x_1,\ldots,x_k) \in \mathbb{R}^k\} \subset \mathbb{R}^k.$$

The action space $\mathcal{A}$ consists of all $2^k - 1$ non-empty subsets of the set

$\{1,2,\ldots,k\}$. An action $A \in \mathcal{A}$ is the selection of some subset of the k populations. $i \in A$ means that π_i is included in the selected subset. An action $A \in \mathcal{A}$ is called a correct selection (CS) if the best population is included in the selected subset A.

Definition 2.1.2. A measurable function δ defined on $\mathcal{X} \times \mathcal{A}$ is called a selection procedure provided that for each $\underline{x} \in \mathcal{X}$, we have

$$0 \leq \delta(\underline{x},A)$$

and

$$\sum_{A \in \mathcal{A}} \delta(\underline{x},A) = 1$$

where $\delta(\underline{x},A)$ dentoes the probability that the subset A is selected when $\underline{x}$ is observed.

The individual selection probability $\psi_i(\underline{x})$ for the population π_i is then given by

$$\psi_i(\underline{x}) = \sum_{A \ni i} \delta(\underline{x},A)$$

where the summation is over all subsets A which contain i. If the selection probability $\psi_i(\underline{x})$ takes on only value 0 or 1, $i = 1,\ldots,k$ then the selecetion procedure $\delta(\underline{x},A)$ is completely specified; in this case it is called a non-randomized procedure.

Definition 2.1.3. Two selection procedures δ and δ' are equivalent if they have the same individual selection probabilities $\psi_i(\underline{x})$ and $\psi_i'(\underline{x})$ for all $\underline{x}$, $i = 1,\ldots,k$.

Hence we can use the following definition, replacing δ by ψ.

Definition 2.1.4. A subset selection rule ψ is a measurable mapping from $\mathcal{X}$ to $\mathbb{R}^k$,

$$\psi(\underline{x}) = (\psi_1(\underline{x}),\ldots,\psi_k(\underline{x}))$$

with

$$0 \leq \psi_i(\underline{x}) \leq 1, \quad i = 1,\ldots,k.$$

If ψ_i's are 0 or 1, the rule is non-randomized. Note that by Definition 2.1.2, we have $\sum_{i=1}^{k} \psi_i(\underline{x}) \geq 1$.

Suppose our goal is to find a nontrivial subset which contains the best population. A large body of literature exists in the area of subset selection procedures (see Gupta and Panchapakesan (1979)). Gupta (1956, 1965) gave maximum-type subset selection procedure. Gupta and Hsu (1978) studied the performance of Gupta-type maximum procedure, Seal-type average procdure (Seal (1955, 1957)) and the Bayes rules. Berger (1979) and Berger and Gupta (1980) proved that Gupta-type maximum procedure is minimax under certain loss functions. In the decision-theoretic approach to the subset selection problem, Goel and Rubin (1977), Chernoff and Yahav (1977), Bickel and Yahav (1977), Gupta and Hsu (1978) and Kim (1979) gave different formulation under different loss function. The loss functions proposed by these authors consist of two component losses and the results are quite sensitive to the ratio of the coefficients of the two components. Hence, whenever the ratio is unknown, we may wish to try some other method of attack.

On the other hand, as pointed out by many authors (see, for example, Bahadur (1950)) the testing of homogeneity of population means or variances is not a satisfactory solution to a comparison among several

populations. Gupta-type maximum procedure gives us a direct and efficient way to meet our goal. However, the LFC (least favorable and configuration), usually occurs when the distributions are identical, i.e., under the hypothesis of homogeneity. As usual, in many cases, the hypothesis of homogeneity is rejected at some small significant level α. Then we may wish to relax (modify) the so-called P^*- condition.

In this chapter we define the posterior-P^* condition in Section 2.2. Two Bayes-P^* selection procedures ψ^B_{NR} and ψ^B are proposed in Section 2.2 and Section 2.3 separately, and their properties are discussed in Section 2.4. In Section 2.5 we discuss their applications to normal distributions. In Section 2.6 procedure ψ^B is compared with Gupta-type maximum procedure. An application for the problem of selecting the $\max\{\frac{\mu_i - a}{\sigma_i}, i = 1,\ldots,k\}$ of the normal distributions $N(\mu_i,\sigma_i^2)$, $i = 1,\ldots,k$ is given in Section 2.7. In Section 2.8 we discuss their applications to the selection problems for Poisson distributions and Poisson processes and their relation to the selection of gamma distributions. Section 2.9 deals with comparisons of the performance of selection procedures ψ^B, ψ^B_{NR}, ψ^M and ψ^{MED}. Here ψ^M and ψ^{MED} are the maximum type selection procedures based on sample means and sample medians, respectively (see Gupta (1956, 1965) and Gupta and Singh (1980)). The comparisons are based on Monte Carlo studies. Robustness of these four procedures is studied in terms of the expected selected size and the efficiency (which will be defined in Section 2.9) where the robustness is in the sense of the effect on the efficiency of the rule when the k true distributions are not normal but, say, logistic, the double exponential distribution or the contaminated distribution (grass error model (Tukey (1960)).

2.2. Definitions of the Posterior-P^* Condition and the Non-randomized Bayes-P^* Procedure ψ^B_{NR}

Let $\theta_{[1]} \leq \cdots \leq \theta_{[k]}$ be the ordered unknown θ_i's. Suppose we have prior distribution τ for $\underline{\theta} = (\theta_1,\ldots,\theta_k)$ then the posterior probability of a correct selection under selection procedure ψ, given $\underline{X} = \underline{x}$, is

$$P(CS|\psi,\underline{X} = \underline{x}) = \sum_{i=1}^{k} \psi_i(\underline{x})p_i(\underline{x})$$

where

$$p_i(\underline{x}) = P(\pi_i \text{ is the best } |\underline{X} = \underline{x}).$$

It is clear

$$\sum_{i=1}^{k} p_i(\underline{x}) = 1.$$

Definition 2.2.1. Given a number $P^*(\frac{1}{k} < P^* < 1)$ and the prior τ, we say a selection procedure ψ satisfies the posterior-P^* condition if

$$P(CS|\psi,\underline{X} = \underline{x}) \geq P^* \quad \text{for all } \underline{x}.$$

Remark 2.2.1. The posterior-P^* condition is based on the prior distribution τ and is different from the usual so-called $\tilde{P}^*$-condition.

Definition 2.2.2. The loss function L_1 is defined by $L_1(\underline{\theta},A) = |A|$ where $|A|$ is the size (number) of populations associated with the selected set A. The loss function L_2 is defined by $L_2(\underline{\theta},A) = \sum_{i \in A} I_{\{\theta_i<\theta_{[k]}\}}(\underline{\theta})$ which is the number of the non-best populations selected by action A. Note that the indicator function

$$I_{\{\theta_i<\theta_{[k]}\}}(\theta) = 1, \quad \text{if } \theta_i < \theta_{[k]}$$
$$= 0, \text{ otherwise.}$$

Definition 2.2.3. Given a number $P^*(\frac{1}{k} < P^* < 1)$ and the prior τ, we define the class $\mathcal{S}_{NR}(\tau,P^*)$ as follows.

$$\mathcal{S}_{NR}(\tau,P^*) = \{\psi | \psi \text{ is any non-randomized rule which satisfies the posterior-}P^* \text{ condition}\} .$$

For the sake of convenience sometimes we will use $\mathcal{S}_{NR}$ instead of $\mathcal{S}_{NR}(\tau,P^*)$.

Definition 2.2.4. Given a number $P^*(\frac{1}{k} < P^* < 1)$, a prior τ, and the loss function L, a selection procedure $\psi \in \mathcal{S}_{NR}(\tau,P^*)$ is called a non-randomized Bayes-P^* procedure (rule) if ψ is a Bayes rule in the class $\mathcal{S}_{NR}(\tau,P^*)$.

Let $p_{[1]}(\underline{x}) \leq \ldots \leq p_{[k]}(\underline{x})$ be the ordered $p_i(\underline{x})$'s and $\pi_{(i)}$ be the population associated with $p_{[i]}(\underline{x})$, $i = 1,\ldots,k$, then a subset selection rule ψ is completely specified by $\{\psi_{(1)},\ldots,\psi_{(k)}\}$ where $\psi_{(i)}$ is defined by

$$\psi_{(i)}(\underline{x}) = P(\pi_{(i)} \text{ is selected } | \psi, \underline{X} = \underline{x}), \quad i = 1,\ldots,k .$$

Next, we propose a non-randomized selection rule which belongs to $\mathcal{S}_{NR}(\tau,P^*)$.

Definition 2.2.5. Given a number $P^*(\frac{1}{k} < P^* < 1)$, $\underline{X} = \underline{x}$, and a prior distribution τ, the selection rule ψ^B_{NR} is defined by $\{\psi_{(1)},\psi_{(2)},\ldots,\psi_{(k)}\}$ where

$$\psi_{(i)}(\underline{x}) = \begin{cases} 1, & \text{if } i \geq j(\underline{x}) \\ 0, & \text{otherwise} \end{cases}$$

and $j(\underline{x})$ is the maximum integer such that

$$\sum_{i=j}^{k} P_{[i]}(\underline{x}) \geq P^* .$$

Lemma 2.2.1. $\psi_{NR}^B \in \mathcal{S}_{NR}$.

Proof. Follows from the definition of ψ_{NR}^B.

Theorem 2.2.1. Given a number $P^*(\frac{1}{k} < P^* < 1)$, the prior τ, and the loss function L_1, the selection procedure ψ_{NR}^B is a non-randomized Bayes-P^* rule.

Proof. It is sufficient to show that the selection procedure ψ_{NR}^B has the the smallest posterior risk in the class $\mathcal{S}_{NR}(\tau,P^*)$. Given the observation $\underline{X} = \underline{x}$. Let the posterior risk of $\psi \in \mathcal{S}_{NR}(\tau,P^*)$ be $\gamma(\underline{x},\psi)$ then

$$\gamma(\underline{x},\psi_{NR}^B) = k - j + 1$$

and

$$\sum_{i=j+1}^{k} p_{[i]}(\underline{x}) < P^*$$

for some j, $1 \leq j \leq k$.

Hence the inequality

$$\gamma(\underline{x},\psi) < \gamma(\underline{x},\psi_{NR}^B)$$

is not true for any $\psi \in \mathcal{S}_{NR}(\tau,P^*)$, $\psi \neq \psi_{NR}^B$. Therefore, the result follows.

Theorem 2.2.2. Theorem 2.2.1 also holds when we replace the loss L_1 by L_2.

Proof. Under the loss function L_2, the posterior risk of the selection procedure $\psi \in \mathcal{D}_{NR}(\tau, P^*)$ is

$$\gamma(\underline{x},\psi) = \sum_{i=1}^{k} \psi_{(i)}(\underline{x})[1 - p_{[i]}(\underline{x})], \quad \text{given } \underline{X} = \underline{x}.$$

By Theorem 2.2.1, we have

$$\sum_{i=1}^{k} \psi^{B}_{NR(i)}(\underline{x}) \leq \sum_{i=1}^{k} \psi_{(i)}(\underline{x}) .$$

If

$$\sum_{i=1}^{k} \psi^{B}_{NR(i)}(\underline{x}) = \sum_{i=1}^{k} \psi_{(i)}(\underline{x})$$

then by definition of ψ^{B}_{NR}, we have

$$\sum_{i=1}^{k} \psi^{B}_{NR(i)}(\underline{x}) p_{[i]}(\underline{x}) \geq \sum_{i=1}^{k} \psi_{(i)}(\underline{x}) p_{[i]}(\underline{x}) .$$

On the other hand, if

$$\sum_{i=1}^{k} \psi^{B}_{NR(i)}(\underline{x}) < \sum_{i=1}^{k} \psi_{(i)}(\underline{x})$$

then

$$\sum_{i=1}^{k} \psi^{B}_{NR(i)}(\underline{x}) \leq \sum_{i=1}^{k} \psi_{(i)}(\underline{x}) - 1$$

$$\leq \sum_{i=1}^{k} \psi_{(i)}(\underline{x})(1 - p_{[i]}(\underline{x})) .$$

Therefore, we have

$$\gamma(\underline{x},\psi^{B}_{NR}) \leq \gamma(\underline{x},\psi) \quad \text{for all } \psi \in \mathcal{D}_{NR}(\tau, P^*) .$$

Corollary 2.2.1. For a given prior τ and the loss function $L = C_1L_1 + C_2L_2$ where C_1, $C_2 > 0$ then ψ^{B}_{NR} is a non-randomized Bayes-P^*

rule wrt the loss function L for all C_1, $C_2 > 0$.

Proof. For the given prior τ and the loss function L, the posterior risk of any procedure $\psi \in \mathcal{S}_{NR}$ is, given $\underline{X} = \underline{x}$, C_1, $C_2 > 0$,

$$\gamma(\underline{x},\psi) = C_1 \sum_{i=1}^{k} \psi_i(\underline{x}) + C_2 \sum_{i=1}^{k} \psi_i(\underline{x})[1 - p_{i}(\underline{x})]$$

$$\geq C_1 \sum_{i=1}^{k} \psi^{B}_{NR\ i}(\underline{x}) + C_2 \sum_{i=1}^{k} \psi^{B}_{NR(i)}(\underline{x})[1 - p_{[i]}(\underline{x})]$$

$$= \gamma(\underline{x},\psi^{B}_{NR})$$

wrt the loss function L.

Hence ψ^{B}_{NR} is a Bayes-P^* rule wrt the loss function L for all C_1, $C_2 > 0$.

2.3. Proposed Bayes-P^* Procedure ψ^B in General

Suppose we are interested in the randomized subset selection rule, and we would like to find such a rule which also satisfies the posterior-P^* condition and has the minimum risk wrt the loss function L_1 and L_2 and the prior distribution τ.

Definition 2.3.1. Given a prior τ, we define a class $\mathcal{S}(\tau,P^*)$ of selection rules, in which all rules satisfy the posterior-P^* condition, for any given observation $\underline{X} = \underline{x}$, that is,

$$\mathcal{S}(\tau,P^*) = \{\psi | P(CS|\psi,\underline{X} = \underline{x}) \geq P^* \quad \text{for all } \underline{x}\} .$$

Definition 2.3.2. Given a number $P^*(\frac{1}{k} < P^* < 1)$, a prior τ and a loss function L, a selection procedure $\psi \in \mathcal{S}(\tau,P^*)$ is called a Bayes-P^* rule if this procedure ψ is a Bayes rule in the class $\mathcal{S}(\tau,P^*)$.

For the sake of convenience, sometimes we will use $\mathcal{S}$ instead of $\mathcal{S}(\tau, P^*)$.

Definition 2.3.2. We define a subset selection procedure ψ^B as follows:

Given a prior τ and observation $\underline{X} = \underline{x}$, ψ^B is defined by

$$\{\psi^B_{(1)}, \psi^B_{(2)}, \ldots, \psi^B_{(k)}\},$$

where

$$\psi^B_{(k)}(\underline{x}) = 1$$

and

$$\psi^B_{(j)}(\underline{x}) = 1, \quad \text{if} \quad \sum_{i=j}^{k} p_{[i]}(\underline{x}) < P^*, \; j \neq k$$

$$= \nu, \quad \text{if} \begin{cases} \sum_{k=j+1}^{k} p_{[i]}(\underline{x}) < P^* \\ \sum_{i=j}^{k} p_{[i]}(\underline{x}) > P^* \\ \nu p_{[j]}(\underline{x}) + \sum_{i=j+1}^{k} p_{[i]}(\underline{x}) = P^* \end{cases}$$

$$= 0, \quad \text{otherwise.}$$

Example. If $k=3$, $P^*=.90$ and the posterior probabilities are: $p_1(\underline{x})=.05$, $p_2(\underline{x})=.80$, $p_3(\underline{x})=.15$, then we select the population π_2 (corresponding to $p_{[3]}(\underline{x})$) with probability 1. And we select π_3 with probability ν where ν is given by

$$\nu(.15) + .80 = .90$$

$$\nu = \frac{.10}{.15} = \frac{2}{3} .$$

By Definition 2.3.2 we have

$$P(CS|\psi^B, \underline{X} = \underline{x}) = \sum_{i=1}^{k} \psi^B_{(i)}(\underline{x}) p_{[i]}(\underline{x}) = P^* \quad \text{if } p_{[k]}(\underline{x}) \leq P^* .$$

Hence we have the following lemma.

Lemma 2.3.1. $\psi^B \in \mathcal{S}(\tau, P^*)$.

Definition 2.3.3. We define a subclass $\mathcal{S}'(\tau, P^*)$ of class $\mathcal{S}(\tau, P^*)$ by

$$\mathcal{S}'(\tau, P^*) = \{\psi \in \mathcal{S}(\tau, P^*) | \psi_{(i)}(\underline{x}) = \psi_{[i]}(\underline{x}) \quad \text{for all } \underline{x}\}$$

where $\psi_{[1]}(\underline{x}) \leq \ldots \leq \psi_{[k]}(\underline{x})$ are the ordered $\psi_i(\underline{x})$'s.

By the definition of $\mathcal{S}'(\tau, P^*)$ we have the following lemmas.

Lemma. 2.3.2. $\psi^B \in \mathcal{S}'(\tau, P^*)$.

Lemma 2.3.3. For all $\psi \in \mathcal{S}(\tau, P^*)$ there exists $\psi' \in \mathcal{S}'(\tau, P^*)$ such that $\gamma(\underline{x}, \psi') = \gamma(\underline{x}, \psi)$ wrt the loss function L_1, for all $\underline{x}$.

Theorem 2.3.1. Selection procedure ψ^B is a Bayes-P^* procedure in $\mathcal{S}(\tau, P^*)$ wrt the loss function L_1.

Proof. Given the observation $\underline{X} = \underline{x}$, and any selection procedure $\psi \in \mathcal{S}(\tau, P^*)$, if $p_{[k]}(\underline{x}) \geq P^*$, then $\psi^B_{(k)}(\underline{x}) = 1$ and $\psi^B_{(j)} = 0$ for all $j \neq k$. Hence we have

$$\gamma(\underline{x}, \psi^B) = 1 \leq \gamma(\underline{x}, \psi) \quad \text{for all } \psi \in \mathcal{S} .$$

If $p_{[k]}(\underline{x}) < P^*$, then we have $\psi^B_{(k)} = 1$ and $\sum_{i=1}^{k} \psi^B_{(i)}(\underline{x}) \tilde{p}_{[i]}(\underline{x}) = P^*$.

We will now show that for any ψ,

$$\gamma(\underline{x}, \psi) < \gamma(\underline{x}, \psi^B)$$

implies

$$P(CS|\psi,\underline{x}) < P(CS|\psi^B,\underline{x}) = P^* .$$

That is

$$\sum_{i=1}^{k} \psi_i(\underline{x}) < \sum_{i=1}^{k} \psi_i^B(\underline{x})$$

implies

$$\sum_{i=1}^{k} \psi_{(i)}(\underline{x})p_{[i]}(\underline{x}) < \sum_{i=1}^{k} \psi_{(i)}^B(\underline{x})p_{[i]}(\underline{x}) .$$

For any C, $1 \leq C < \sum_{i=1}^{k} \psi_i^B(\underline{x})$, we have $C = a + \eta$ where a is a positive integer and $0 \leq \eta < 1$.

It is easy to see that the maximum posterior proability of correct selection of procedure ψ with $\sum_{i=1}^{k} \psi_i(\underline{x}) = C$ is

$$\sum_{i=k-a+1}^{k} p_{[i]}(\underline{x}) + \eta p_{[k-a]}(\underline{x}) .$$

And it is less than $\sum_{i=1}^{k} \psi_{(i)}^B(\underline{x})p_{[i]}(\underline{x})$, since $C < \sum_{i=1}^{k} \psi_{(i)}^B(\underline{x})$. Therefore ψ^B is Bayes-P^* procedure in $\mathcal{S}(\tau,P^*)$.

Lemma 2.3.4. Given the loss function L_2, for all $\psi \in \mathcal{S}(\tau,P^*)$, there exists $\psi' \in \mathcal{S}'(\tau,P^*)$ such that, given $\underline{X} = \underline{x}$, $\gamma(\underline{x},\psi') \leq \gamma(\underline{x},\psi)$ where $\gamma(\underline{x},\psi)$ is the posterior risk wrt the prior τ and the loss function L_2.

Proof. Given $\underline{X} = \underline{x}$, let $\psi'_{(i)}(\underline{x}) = \psi_{[i]}(\underline{x})$ then

$$\sum_{i=1}^{k} \psi'_{(i)}(\underline{x})p_{[i]}(\underline{x}) = \sum_{i=1}^{k} \psi_{[i]}(\underline{x})p_{[i]}(\underline{x})$$

$$\geq \sum_{i=1}^{k} \psi_{(i)}(\underline{x}) p_{[i]}(\underline{x})$$

$$\geq P^*$$

hence $\psi' \in \mathcal{D}'(\tau, P^*)$.

Now,

$$\gamma(\underline{x}, \psi') = \sum_{i=1}^{k} \psi'_{(i)}(\underline{x})(1 - p_{[i]}(\underline{x}))$$

$$= \sum_{i=1}^{k} \psi_{[i]}(\underline{x})(1 - p_{[i]}(\underline{x}))$$

$$\leq \sum_{i=1}^{k} \psi_{(i)}(\underline{x})(1 - p_{[i]}(\underline{x}))$$

$$= \gamma(\underline{x}, \psi) .$$

Hence the proof is complete.

Theorem 2.3.2. Given the prior τ and the observation $\underline{X} = \underline{x}$, the procedure ψ^B is a Bayes-P^* procedure in the class $\mathcal{D}(\tau, P^*)$ when the loss function is L_2.

Proof. By Lemma 2.2.4, it is sufficient to show that

$$\gamma(\underline{x}, \psi^B) = \min_{\psi' \in \mathcal{D}} \gamma(\underline{x}, \psi')$$

where

$$\gamma(\underline{x}, \psi) = \sum_{i=1}^{k} \psi_i(\underline{x})(1 - p_i(\underline{x}))$$

$$= \sum_{i=1}^{k} \psi_i(\underline{x}) - \sum_{i=1}^{k} \psi_i(\underline{x}) p_i(\underline{x}) .$$

Let

$$i_0 = \min_{1 \le i \le k} \{i | \psi^B_{(i)}(\underline{x}) > 0\}$$

$$A_1 = \{i | \psi'_{(i)}(\underline{x}) - \psi^B_{(i)}(\underline{x}) > 0\}$$

$$A_2 = \{i | \psi'_{(i)}(\underline{x}) - \psi^B_{(i)}(\underline{x}) < 0\} ,$$

then $A_1 \cap A_2 = \phi$.

And we have

$$a_1 = \max A_1 = i_0 - 1, \quad \text{if} \quad \psi^B_{(i_0)}(\underline{x}) \ge \psi'_{(i_0)}(\underline{x})$$

$$= i_0 \quad , \quad \text{if} \quad \psi^B_{(i_0)}(\underline{x}) < \psi'_{(i_0)}(\underline{x})$$

$$a_2 = \min A_2 = i_0 \quad , \quad \text{if} \quad \psi^B_{(i_0)}(\underline{x}) \ge \psi'_{(i_0)}(\underline{x})$$

$$= i_0 + 1, \quad \text{if} \quad \psi^B_{(i_0)}(\underline{x}) < \psi'_{(i_0)}(\underline{x})$$

hence $a_1 < a_2$.

Therefore, we have

$$\gamma(\underline{x},\psi') - \gamma(\underline{x},\psi^B) = \sum_{i=1}^{k} (\psi'_{(i)}(\underline{x}) - \psi^B_{(i)}(\underline{x}))(1 - p_{[i]}(\underline{x}))$$

$$= \sum_{i \in A_1} (\psi'_{(i)}(\underline{x}) - \psi^B_{(i)}(\underline{x}))(1 - p_{[i]}(\underline{x})) + \sum_{i \in A_2} (\psi'_{(i)}(\underline{x}) - \psi^B_{(i)}(\underline{x}))(1 - p_{[i]}(\underline{x}))$$

$$\ge \sum_{i \in A_1} (\psi'_{(i)}(\underline{x}) - \psi^B_{(i)}(\underline{x}))(1 - p_{[a_1]}(\underline{x})) + \sum_{i \in A_2} (\psi'_{(i)}(\underline{x}) - \psi^B_{(i)}(\underline{x}))(1 - p_{[a_2]}(\underline{x}))$$

$$\geq \Big[\sum_{i\in A_1} (\psi'_{(i)}(\underline{x})+\psi^B_{(i)}(\underline{x})+\sum_{i\in A_2} (\psi'_{(i)}(\underline{x})+\psi^B_{(i)}(\underline{x}))\Big]$$

$$\cdot(1 - p_{[a_2]}(\underline{x}))$$

$$= \sum_{i=1}^{k} [\psi'_{(i)}(\underline{x}) - \psi^B_{(i)}(\underline{x})](1 - p_{[a_2]}(\underline{x}))$$

≥ 0 by Theorem 2.3.1.

Corollary 2.3.1. Procedure ψ^B is a Bayes-P* rule in $\mathfrak{D}(\tau,P^*)$ wrt the loss function $L = c_1L_1 + c_2L_2$, $c_1, c_2 > 0$.

Proof. Similar to Corollary 2.2.1. hence it is omitted.

2.4. Properties of ψ^B and ψ^B_{NR}

In this section we discuss some properties of selection procedure ψ^B and ψ^B_{NR}. The following definition of the ordering of distributions was introduced by Lehmann (1952) and further discussed by Lehmann (1955) and Alam (1973).

Definition 2.4.1. A subset $A \subset \mathbb{R}^k$ is monotone if $\underline{x} \in A$ and $\underline{y}$ satisfies $y_i \leq x_i$ for all $i = 1,\ldots,k$, implies $\underline{y} \in A$.

Definition 2.4.2. A family of probability distributions on $\mathbb{R}^k$, $\{F_{\underline{\theta}}: \underline{\theta} \in \Omega \subset \mathbb{R}^k\}$ has the stochastic increasing property (SIP) if $\underline{\theta} \in \Omega$, $\underline{\theta}' \in \Omega$, and $\theta_i \leq \theta'_i$ for all $i = 1,\ldots,k$, implies

$$\int_A dF_{\underline{\theta}} \geq \int_A dF_{\underline{\theta}'}$$

for all monotone sets A.

Let $f(\cdot,\theta_i)$ be the p.d.f. of population π_i. Let $\tau(\underline{\theta})$ be the given prior where θ_i's are mutually independent. Suppose for $\underline{X} = \underline{x}$, we have absolutely continuous posterior c.d.f. $G(\underline{\theta}|\underline{x})$. Hence we can write the p.d.f. as

$$g(\underline{\theta}|\underline{x}) = \prod_{i=1}^{k} g_i(\theta_i|\underline{x}) = \prod_{i=1}^{k} g_i(\theta_i|x_i) .$$

Let $G_i(\cdot|x_i)$ be the posterior c.d.f. associated with $g_i(\cdot|x_i)$, $i = 1,\ldots,k$.

Definition 2.4.3. The absolutely continuous posterior c.d.f. $G_i(\cdot|x_i)$, $i = 1,\ldots,k$, have the generalized (strictly) stochastic increasing property (G(S)SIP) if for any i, j, $1 \leq i, j \leq k$, $x_i(<) \leq x_j$ then

$$G_i(\cdot|x_i)(>) \geq G_i(\cdot|x_j) .$$

Note that if $G_i(\cdot|\cdot) \equiv G_j(\cdot|\cdot) \equiv G(\cdot|\cdot)$ for all i, j, $1 \leq i, j \leq k$, then the GSIP is the usual SIP.

Definition 2.4.4. A selection procedure ψ is monotone (ordered) if and only if for every $\underline{x} \in \mathbb{R}^k$, $x_i \leq x_j$ implies $\psi_i(\underline{x}) \leq \psi_j(\underline{x})$. ψ is monotone a.e. if ψ is monotone with the exception of a subset of observations having probability zero.

Theorem 2.4.1. If the prior $\tau(\underline{\theta})$ is such that we have absolutely continuous independent posterior distributions $G_i(\cdot|\cdot)$, $i = 1,2,\ldots,k$, with GSSIP, then for every $\underline{x} \in \mathbb{R}^k$, $x_i < x_j$ implies

$$p_i(\underline{x}) < p_j(\underline{x}) .$$

Hence both selection procedures ψ^B and ψ^B_{NR} are monotone (ordered) a.e.

Proof.

$$p_i(\underline{x}) \equiv P(\theta_i = \theta_{[k]}|\underline{x}) = \int \prod_{\ell \neq i} G_\ell(t|x_\ell)dG_i(t|x_i)$$

$$= \int \prod_{\ell \neq i,j} G_\ell(t|x_\ell)G_j(t|x_j)dG_i(t|x_i)$$

$$< \int \prod_{\ell \neq i,j} G_\ell(t|x_\ell)G_i(t|x_i)dG_i(t|x_i)$$

$$= 1 - \int \frac{d}{dt}\left[\prod_{\ell \neq j} G_\ell(t|x_\ell)\right]G_i(t|x_i)dt$$

$$< 1 - \int \frac{d}{dt}\left[\prod_{\ell \neq j} G_\ell(t|x_\ell)\right]G_j(t|x_j)dt$$

$$= \int \prod_{\ell \neq j} G_\ell(t|x_\ell)dG_j(t|x_j) = p_j(\underline{x}) \text{ if } x_i < x_j.$$

Since

$$\psi_i^B(\underline{x}) \leq \psi_j^B(\underline{x}) \quad \text{iff} \quad p_i(\underline{x}) < p_j(\underline{x})$$

and

$$\psi_{NR\ i}^B(\underline{x}) \leq \psi_{NR\ j}^B(\underline{x}) \quad \text{iff} \quad p_i(\underline{x}) < p_j(\underline{x}) .$$

Therefore, the procedures ψ^B and ψ_{NR}^B are ordered a.e.

Under GSIP assumptions, we can relabel the populations such that $x_1 \leq \ldots \leq x_k$, hence we have $p_i(\underline{x}) = p_{[i]}(\underline{x})$ and $\psi_i^B(\underline{x}) = \psi_{[i]}^B(\underline{x}) = \psi_{(i)}^B(\underline{x})$. Nagel (1970) defined the "just" property of a selection rule as follows:

Definition 2.4.5. A selection rule ψ is called "just", if and only if, for all $i = 1,\ldots,k$, $\psi_i(\underline{x}) \leq \psi_i(\underline{x}')$ whenever $x_i \leq x_i'$ and $x_j \geq x_j'$ for $j \neq i$.

We call ψ "just" a.e. (almost everywhere) if ψ is "just" with the exception of a subset of observations having probability zero.

Definition 2.4.6. A selection rule ψ is called translation-invariant if for all $\underline{x} \in \mathbb{R}^k$, and for all $c \in R$

$$\psi_i(x_1 + c,\ldots,x_k + c) = \psi_i(x_1,\ldots,x_k) \quad i = 1,\ldots,k.$$

Definition 2.4.7. A selection rule ψ is called scale-invariant if for all $\underline{x} \in \mathbb{R}^k$, and for all $c > 0$

$$\psi_i(x_1 \cdot c,\ldots,x_k \cdot c) = \psi_i(x_1,\ldots,x_k) \quad i = 1,\ldots,k.$$

Theorem 2.4.2. If the posterior distributions $G_i(\cdot|\cdot)$, $i = 1,2,\ldots,k$, have the GSSIP property then both selection procedures ψ^B and ψ^B_{NR} are "just" a.e. .

Proof. It is sufficient to show that

$$p_i(\underline{x}) < p_i(\underline{x}') \quad \text{whenever} \quad x_i' > x_i$$

$$\text{and } x_j' < x_j \quad \forall\, j \neq i .$$

For any fixed i,

$$\begin{aligned} p_i(\underline{x}) &= \int \prod_{j\neq i} G_j(t|x_j)\,dG_i(t|x_i) \\ &< \int \prod_{j\neq i} G_j(t|x_j')\,dG_i(t|x_i) \\ &= 1 - \int G_i(t|x_i)\,d \prod_{j\neq i} G_j(t|x_j') \\ &< 1 - \int G_i(t|x_i')\,d \prod_{j\neq i} G_j(t|x_j') \\ &= \int \prod_{j\neq i} G_j(t|x_j')\,dG_i(t|x_i') \\ &= p_i(\underline{x}') , \end{aligned}$$

which holds for all $i = 1,\ldots,k$. Hence the proof follows.

Definition 2.4.8. Given a number $P^*(\frac{1}{k} < P^* < 1)$, $\underline{X} = \underline{x}$ and a prior τ; for any selection procedure $\psi \in \mathcal{S}(\tau,P^*)$ the ratio of the posterior probability $P(CS|\psi,\underline{x})$ and the posterior expected selected size $E(S|\psi,\underline{x})$ is called the posterior-efficiency of ψ and is denoted by $EFF(\psi|\underline{x})$,

$$EFF(\psi|\underline{x}) = \frac{P(CS|\psi,\underline{x})}{E(S|\psi,\underline{x})} .$$

If $EFF(\psi|\underline{x}) \geq EFF(\psi'|\underline{x})$ for all $\psi' \in \mathcal{S}$ and all $\underline{x}$, then the selection procedure ψ is called "posterior most efficient" (PME) selection procedure in $\mathcal{S}(\tau,P^*)$.

Theorem 2.4.3. The non-randomized posterior-P^* selection procedure ψ^B_{NR} is the PME selection procedure in $\mathcal{S}_{NR}(\tau,P^*) = \mathcal{S}_{NR}$, given τ, P^*.

Proof. By Lemma 2.2.1, for all $\psi \in \mathcal{S}_{NR}$

$$\exists\, \psi' \in \mathcal{S}'_{NR} \ni EFF(\psi'|\underline{x}) \geq EFF(\psi|\underline{x}) \quad \forall\, \underline{x} ,$$

hence it is sufficient to show that:

Given $\tau(\underline{\theta})$, P^*, $\underline{x}$, $EFF(\psi^B_{NR}|\underline{x}) \geq EFF(\psi'|\underline{x})$ for all $\psi' \in \mathcal{S}'_{NR}(\tau,P^*)$.

We know that, in $\mathcal{S}_{NR}(\tau,P^*)$ hence in $\mathcal{S}'_{NR}(\tau,P^*)$, ψ^B_{NR} always has minimum selected size, i.e. $\forall\, \underline{x}$, $\sum_{i=1}^{k} \psi^B_{NRi}(\underline{x}) + c = \sum_{i=1}^{k} \psi'_i(\underline{x})$ for some integer c, $0 \leq c \leq k-1$.

$$EFF(\psi'|\underline{x}) = \frac{\sum_{i=1}^{k} \psi'_{(i)}(\underline{x}) p_{[i]}(\underline{x})}{\sum_{i=1}^{k} \psi'_{(i)}(\underline{x})}$$

$$\leq \frac{\sum_{i=1}^{k} \psi^B_{NR(i)}(\underline{x}) p_{[i]}(\underline{x}) + p_{[k-s-c+1]}(\underline{x}) + \ldots + p_{[k-s]}(\underline{x})}{\sum_{i=1}^{k} \psi^B_{(i)} + c}$$

if $\psi^B_{NR}(\underline{x}) = (0,\ldots,0,\underbrace{1,\ldots,1}_{s \text{ terms}})$.

$$EFF(\psi'|\underline{x}) \leq \frac{\sum_{i=k-s+1}^{k} \psi^B_{NR(i)}(\underline{x})p_{[i]}(\underline{x}) + cp_{[k-s]}(\underline{x})}{\sum_{i=1}^{k} \psi^B_{NR(i)}(\underline{x}) + c}$$

$$\leq \frac{\sum_{i=1}^{k} \psi^B_{NR(i)}(\underline{x})p_{[i]}(\underline{x})}{\sum_{i=1}^{k} \psi^B_{NR(i)}(\underline{x})}$$

$$= EFF(\psi^B_{NR}|\underline{x}) \quad .$$

The last inequality is obtained by

$$\sum_{i=1}^{k} \psi^B_{NR(i)}(\underline{x})p_{[i]}(\underline{x}) = \sum_{i=k-s+1}^{k} \psi^B_{NR(i)}(\underline{x})p_{[i]}(\underline{x})$$

$$\geq \left(\sum_{i=k-s+1}^{k} \psi^B_{NR(i)}(\underline{x})\right) p_{[k-s]}(\underline{x}).$$

Theorem 2.4.4. The randomized selection procedure ψ^B is the PME procedure in $\mathcal{S}(\tau,P^*) = \mathcal{S}$ for given τ, P^*.

Proof. It suffices to show that, given τ, P^*, $\underline{X} = \underline{x}$,

$$EFF(\psi^B|\underline{x}) \geq EFF(\psi'|\underline{x}), \quad \forall\, \psi' \in \mathcal{S}' .$$

Suppose $\psi^B(\underline{x}) = (0,\ldots,\underbrace{\nu,1,\ldots,1}_{s \text{ terms}})$ $\quad 0 \leq \nu < 1,\ 1 \leq s \leq k-1.$

By theorem 2.3.1 there exists $c \geq 0$ such that

$$\sum_{i=1}^{k} \psi^B_{(i)}(\underline{x}) + c = \sum_{i=1}^{k} \psi'_{(i)}(\underline{x}) \ .$$

If $0 \leq c < 1$, then

$$\mathrm{EFF}(\psi'|\underline{x}) = \frac{\sum_{i=1}^{k} \psi'_{(i)}(\underline{x}) p_{[i]}(\underline{x})}{\sum_{i=1}^{k} \psi'_{(i)}(\underline{x})}$$

$$\leq \frac{\sum_{i=1}^{k} \psi^{B}_{(i)}(\underline{x}) p_{[i]}(\underline{x}) + c p_{[k-s]}(\underline{x})}{\sum_{i=1}^{k} \psi^{B}_{(i)}(\underline{x}) + c}$$

$$\leq \frac{\sum_{i=1}^{k} \psi^{B}_{(i)}(\underline{x}) p_{[i]}(\underline{x})}{\sum_{i=1}^{k} \psi^{B}_{(i)}(\underline{x})}$$

$$= \mathrm{EFF}(\psi^{B}|\underline{x}) .$$

If $1 \leq c = \nu' + t + (1-\nu)$, $t \geq 0$ integer, $0 \leq \nu' \leq 1$ then

$$\sum_{i=1}^{k} \psi'_{(i)}(\underline{x}) p_{[i]}(\underline{x}) = \sum_{i=k-s+1}^{k} \psi^{B}_{(i)}(\underline{x}) p_{[i]}(\underline{x}) + \nu' p_{[k-s-t+1]}(\underline{x}) + p_{[k-s-t]}(\underline{x}) + \ldots + (1-\nu) p_{[k-s]}(\underline{x})$$

$$\geq \sum_{i=1}^{k} \psi^{B}_{(i)}(\underline{x}) p_{[i]}(\underline{x}) + c p_{[k-s]}(\underline{x})$$

hence by the same argument as above we have

$$\mathrm{EFF}(\psi'|\underline{x}) \leq \mathrm{EFF}(\psi^{B}|\underline{x}) .$$

Since $\underline{x}$ is arbitrary, the result holds for all $\underline{x}$.

2.5. Applications to Normal Model

Suppose we have k populations $\pi_1,\ldots,\pi_k$; population π_i has distribution $N(\mu_i,\sigma_i^2)$, where σ_i's are known and μ_i's are unknown. Assume that we

have independent observations $X_{i1},\ldots,X_{in_i}$, $i = 1,\ldots,k$. Let $X_i = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij}$ and let $\underline{X} = (X_1,\ldots,X_k)$.

Suppose we are interested in selecting a subset containing the best (the population having the largest mean) under the posterior-P^* condition, wrt some prior $\tau = \tau(\underline{\mu})$. Then to find a Bayes-$P^*$ selection procedure is equivalent, in some sense, to finding $p_i(\underline{x})$, which is the posterior probability of the event $\{\pi_i$ is the best$\}$, given observations $\underline{X} = \underline{x}$, wrt a given prior τ, for all $i = 1,\ldots,k$.

Case I. Assume that we have a common sample size n and a common known variance σ^2.

Ia. Suppose we have no prior information about the unknown parameters, and use the "non-informative" (Box and Tiao (1973)) or "locally uniform" prior $p(\mu_i)\propto c$ for each population.

The posterior density function g_i of μ_i, given $\underline{x}$ is the normal density with mean x_i and variance σ^2/n, i.e.,

$$g_i(\mu_i|\underline{x}) = \frac{\sqrt{n}}{\sqrt{2\pi}\,\sigma}\exp\left(-\frac{n(\mu_i-x_i)^2}{2\sigma^2}\right).$$

Hence

$$p_{[i]}(\underline{x}) = P(\mu_{(i)} = \mu_{[k]}|\underline{X} = \underline{x})$$

$$= \int_{-\infty}^{\infty} \prod_{j\neq i} \Phi\left(t + \frac{\sqrt{n}}{\sigma}(x_{[i]} - x_{[j]})\right)d\Phi(t)$$

$i = 1,\ldots,k.$

Here $\mu_{(i)}$ is the quantity corresponding to the $i^{\underline{th}}$ largest observation $x_{[i]}$.

Ib. If μ_i's are independent and have the identical prior distribution $N(\theta_0,\sigma_0^2)$ and $X_i|\mu_i \sim N(\mu_i,\sigma_1^2/n)$, then it is well known that the posterior density funciton g_i of μ_i, given $\underline{X} = \underline{x}$ is

$$g_i(\mu_i|\underline{x}) \sim N(\bar{\theta}_{x_i},\xi^2) \text{ with SIP property}$$

where

$$\bar{\theta}_{x_i} = \xi^2(\sigma_0^{-2}\theta_0 + n\sigma_1^{-2}x_i)$$

$$\xi^2 = (\sigma_0^{-2} + n_1\sigma_1^{-2})^{-1}.$$

Hence

$$p_{[i]}(\underline{x}) = \int_{-\infty}^{\infty} \prod_{j\neq i} \Phi(t + \xi n\sigma_1^{-2}(x_{[i]} - x_{[j]}))d\Phi(t).$$

The last expression for $p_{[i]}(\underline{x})$ is the same as that for the non-informative prior whenever $\sigma_0 \to \infty$.

Since $p_{[i]}(\underline{x}) = p_{[i]}(\underline{x} + \underline{b})$ and since the normal distribution has the strictly SIP, it follows that ψ^B and ψ^B_{NR} are "just" a.e. and translation-invariant in both case Ia and Ib.

Case II. Variance σ_i's are known but σ_i's and n_i's are not all equal.

IIa. Using the non-informative prior $p(\mu_i) \propto c$, i=1,...,k, we have

$$p_{(i)}(\underline{x}) = \int_{-\infty}^{\infty} \prod_{j\neq i} \Phi(t\,\frac{v_{(i)}}{v_{(j)}} + \frac{x_{[i]}-x_{[j]}}{v_{(j)}})d\Phi(t)$$

where $v_{(i)} = \frac{\sigma_{(i)}}{n_{(i)}}$ $i = 1,...,k$. $p_{(i)}$, $\sigma_{(i)}$ and $n_{(i)}$ are corresponding to $x_{[i]}$ and we have the following theorem.

Theorem 2.5.1. $p_{(i)}(\underline{x})$ is non-decreasing in i, i.e., $p_{(i)}(\underline{x}) \geq p_{[i]}(\underline{x})$.

Remark 2.5.1. From the above formula of $p_{(i)}(\underline{x})$, it is easy to see, increasing the sample size of the non-best populations will increase the probability that the best population to be selected, however, before doing this, we don't know which one is the best one.

In this case ψ^B and ψ^B_{NR} are "just" a.e. and translation-invariant.

Case III. Assume that priors are independent but not identical normal distributions, namely, $\mu_i \sim N(\theta_i, \sigma_{0i}^2)$, where θ_i's are not all equal; if the conditional distribution of X_i, given μ_i, is $N(\mu_i, \frac{\sigma_{1i}^2}{n_i})$, then the posterior density of μ_i, given $X_i = x_i$ is $g_i(\mu_i|x_i)$, which is the probability density function of normal distritubion $N(\bar{\theta}_{x_i}, \xi_i^2)$ where

$$\bar{\theta}_{x_i} = \xi_i^2(\sigma_{0i}^{-2}\theta_i + n_i\sigma_{1i}^{-2}x_i)$$

$$\xi_i^2 = (\sigma_{0i}^{-2} + \sigma_{1i}^{-2}n_i)^{-1} .$$

Hence we have

$$p_i(\underline{x}) = \int_{-\infty}^{\infty} \prod_{j\neq i} \Phi[t\frac{\xi_i}{\xi_j} + \frac{1}{\xi_j}(\bar{\theta}_{x_i} - \bar{\theta}_{x_j})]d\Phi(t).$$

If $\sigma_{0i} = \sigma_0$, $\sigma_{1i} = \sigma_1$ and $n_i = n$, $i = 1,\ldots,k$, then

$$\xi_i = \xi = (\sigma_0^{-2} + \sigma_1^{-2}n)^{-1} \qquad i = 1,\ldots,k$$

and

$$p_i(\underline{x}) = \int_{-\infty}^{\infty} \prod_{j\neq i} \Phi[t + \xi(\frac{\theta_i-\theta_j}{\sigma_0^2} + \frac{n(x_i-x_j)}{\sigma_1^2})]d\Phi(t) .$$

Case IV. The General Normal Model

Here we consider a more general prior. Suppose we have k populations, common sample size n for each population, and common known variance $\sigma^2 > 0$. The observation can reduce to $\underline{X} = (X_1, \ldots, X_k)$ where $X_i = \sum_{i=1}^{n} X_{ij}/n$, by sufficiency.

The "Normal Model" is defined as follows:

(a) $$\underline{X}|\underline{\mu} \sim N(\underline{\mu}, qI),\ q \equiv \frac{\sigma^2}{n}$$

where I is the $k \times k$ identity matrix.

So the X's are (conditionally) independent when $\underline{\mu}$ is given.

(b) $$\underline{\mu} \sim N(\theta_0 \underline{1},\ \gamma I + tU)$$

where $\theta_0 \in \mathbb{R}$, $\gamma > 0$, $t > -\frac{\gamma}{k}$,

$$\underline{1} = (1,\ldots,1) \quad \text{and} \quad U = \underline{1}'\underline{1}\ .$$

Here $\gamma > 0$ and $t > -\frac{\gamma}{k}$ are necessary and sufficient for $\gamma I + tU$ to be positive definite. This model was chosen by Chernoff and Yahav (1977) ($t > 0$), Gupta and Hsu (1978) and Miescke (1979).

By (a) and (b) we get the posterior distribution of $\underline{\mu}$, given $\underline{X} = \underline{x}$, and the distribution of $\underline{X}$ as follows:

$$\underline{\mu}|\underline{x} \sim N(\underline{\theta},\ aI + bU)$$

where

$$\underline{\theta} = \gamma(q+\gamma)^{-1}\underline{x} + qt((q+\gamma)(q+\gamma+kt))^{-1}\underline{x}\,U + q(q+\gamma+kt)^{-1}m\,\underline{1}$$

$$a = \gamma q(q + r)^{-1}$$

$$b = q^2 t(q+\gamma)^{-1}(q+\gamma+kt)^{-1}$$

$$\underline{X} \sim N(m\underline{1},(q+\gamma)I + tU)$$

Lemma 2.5.1. Let $\underline{Y} \sim N(\underline{\mu} + \rho\underline{1},aI + bU)$ with $\underline{\mu} \in \mathbb{R}^k$, $\rho \in \mathbb{R}$, $a > 0$ and $b > -a/k$. Then there exists a random vector $\underline{Z} \sim N(\underline{\mu},aI)$ such that $h(\underline{Y}) = h(\underline{Z})$ everywhere for every translation-invariant $h: \mathbb{R}^k \to \mathbb{R}^k$.

Proof. (See Miescke (1979)).

With this lemma, it is easy to get

$$p_i(\underline{x}) \equiv P(\mu_i = \mu_{[k]}|\underline{x})$$

$$= \int I_{\{\mu_i=\mu_{[k]}\}} d\Phi\left((\tfrac{\gamma}{q+\gamma})\underline{x}, \tfrac{\gamma q}{q+\gamma} I\right)(\underline{\mu})$$

where $\Phi_{(\underline{\mu},V)}$ is the normal distribution with mean $\underline{\mu}$ and variance-covariance matrix V.

We can rewrite $p_i(\underline{x})$ as

$$p_i(\underline{x}) = \int_{-\infty}^{\infty} \prod_{j\neq i} \Phi(t + (\tfrac{\gamma}{q(q+\gamma)})^{\frac{1}{2}}(x_i - x_j))d\Phi(t) .$$

Let $\gamma = \sigma_0^2$, $q = \sigma^2/n$, we have

$$p_i(\underline{x}) = \int_{-\infty}^{\infty} \prod_{j\neq i} \Phi\left(t + \left(\frac{\sigma_0^2}{\frac{\sigma^2}{n}\left(\frac{\sigma^2}{n} + \sigma_0^2\right)}\right)^{\frac{1}{2}}(x_i - x_j)\right)d\Phi(t).$$

The above expression is exactly the same as that of the independent prior Case I, Ib.

Case V. Under normal assumption as before, but suppose σ_i's are unknown and that neither σ_i's nor n_i's are all equal.

Suppose we have no prior information about (μ,σ), for each individual population π_i assign prior $p(\mu_i,\sigma_i)\propto\sigma_i^{-1}$ then we have (See Box and Tiao (1973)) that the posterior density of μ_i, given $\underline{X}_i=\underline{x}_i=(x_{i1},\ldots,x_{in_i})$ is

$$p(\mu_i|\underline{x}_i) = \frac{(s_i/\sqrt{n_i})^{-1}}{B(\frac{1}{2}\nu_i,\frac{1}{2})\sqrt{\nu_i}}\left[1+\frac{n_i(\mu_i-x_i)^2}{\nu_i s_i^2}\right]^{-\frac{1}{2}(\nu_i+1)}$$

where s_i^2 is the sample variance, $B(\cdot,\cdot)$ is a Beta function and $\nu_i = n_i-1$. Hence

$$p(t_i = \frac{\mu_i-x_i}{s/\sqrt{n_i}}\,|\,\underline{x}_i) = \frac{1}{B(\frac{1}{2}\nu_i,\frac{1}{2})\sqrt{\nu_i}}\left(1+\frac{t_i^2}{\nu_i}\right)^{-\frac{1}{2}(\nu_i+1)},$$

which is the density of the student's t distribution with $\nu_i(=n_i-1)$ degrees of freedom.

Using this result we can write the formula of $p_i(\underline{x})$ by

$$p_i(\underline{x}) = P(\mu_i > \mu_j,\ \forall\, j\neq i\,|\underline{x})$$

$$= \int \prod_{j\neq i} T\nu_j\left(t\,\frac{s_i/\sqrt{n_i}}{s_j/\sqrt{n_j}}+\frac{x_i-x_j}{s_j/\sqrt{n_j}}\right)dT\nu_i(t)$$

where $\nu_i = n_i - 1, \quad i = 1,\ldots,k$

$$\nu_i s_i = \sum_{r=1}^{n_i}(x_{ir}-x_i)^2$$

$T\nu_i$ is the c.d.f. of t distribution with ν_i degrees of freedom. When ν_i's are large, t distribution approaches normal distribution, hence, for large n_i, $i = 1,\ldots,k$, we can replace T by Φ.

Case VI. Suppose we are interested in finding a subset which contains the population with the smallest variance; i.e., we define the best population as the one with the smallest variance, and suppose that we have no prior information about σ. In this case, it is reasonable to assume that

$$p(\mu,\sigma)\propto\sigma^{-1}, \text{ if } \mu \text{ is unknown}$$

$$p(\sigma)\propto\sigma^{-1}, \text{ if } \mu \text{ is known.}$$

Let

$$\nu_j = n_j,\ \nu_j S_j^2 = \sum_{r=1}^{k} (X_{jr}-\mu)^2 \text{ if } \mu \text{ is known}$$

$$\nu_j = n_j - 1,\ \nu_j S_j^2 = \sum_{r=1}^{k} (X_{jr}-\bar{X}_j)^2 \text{ if } \mu \text{ is unknown, } n_j > 1 \quad j = 1,2,\ldots,k$$

$$\underline{S}^2 = (S_1^2,\ldots,S_k^2),\ \underline{X} = (X_{11},\ldots,X_{1n_1},\ldots,X_{kn_k})$$

and Y_ν be the random variable with c.d.f. χ^2_ν which is the χ^2 distribution with ν degrees of freedom.

Then for either case (μ known or unknown), we have

$$\begin{aligned}
p_i(\underline{x}) &= P(\sigma_i^2 = \sigma^2_{[1]} | \underline{X} = \underline{x}) \\
&= P(\sigma_i^2 \le \sigma_j^2,\ \forall\, j\neq i | \underline{X} = \underline{x}) \\
&= P\left(\frac{\nu_j s_j^2}{\sigma_j^2} \le \frac{\nu_i s_i^2}{\sigma_i^2}\left(\frac{\nu_j s_j^2}{\nu_i s_i^2}\right),\ \forall\, j\neq i | \underline{S}^2 = \underline{s}^2\right) \\
&= P\left(Y_{\nu_j} \le Y_{\nu_i}\left(\frac{\nu_j s_j^2}{\nu_i s_i^2}\right),\ \forall\, j\neq i | \underline{S}^2 = \underline{s}^2\right) \\
&= \int_0^\infty \prod_{j\neq i} \chi^2_{\nu_j}\left(u\,\frac{\nu_j s_j^2}{\nu_i s_i^2}\right) d\chi^2_{\nu_i}(u)
\end{aligned}$$

$$= \int_0^\infty \prod_{j\neq i} \chi^2_\nu\left(u \frac{s_j^2}{s_i^2}\right) d\chi^2_\nu(u) \quad \text{if } n_1=\ldots=n_k=\nu+1.$$

With these $p_1(\underline{x}),\ldots,p_k(\underline{x})$ we can apply Bayes-P* rules ψ^B and ψ^B_{NR} easily.

Lemma 2.5.2. In Case VI, ψ^B and ψ^B_{NR} are just * a.e. and (scale) translation invariant.

* Here the definition of the "just" property for a selection rule is

$$\psi_i(\underline{s}^2) \leq \psi_i(\underline{s}^{2'}) \text{ if } s_i^2 \geq s_i^{2'},\ s_j^2 \leq s_j^{2'},\ \forall\, j \neq i\ .$$

2.6. Comparison of Selection Rules ψ^B and ψ^M in the Normal Location Parameter Case

We have k normal populations with a common known variance σ^2 and common sample size n. For this case Gupta (1956) proposed and studied the procedure ψ^M.

ψ^M: Select π_i iff $X_i > X_{[k]} - d\frac{\sigma}{\sqrt{n}}$ $i = 1,\ldots,k$ where $d = d(k,P^*) > 0$ is to be determined by

$$\inf_{\underline{\theta} \in \Omega} P(CS|\psi^M) = P^*$$

and Ω is the parameter space.

We will show that $\psi^M \in \mathcal{D}_{NR}(\tau,P^*)$ where τ is the locally uniform prior distribution. For fixed P^* and k, let d be determined by

$$\int_{-\infty}^{\infty} \Phi^{k-1}(t + d)d\Phi(t) = P^*\ . \qquad (2.6.1)$$

Let

$$\mathcal{X} = \{\text{all possible observed values}\} = \mathbb{R}^k$$

$$\mathcal{X}_1 = \{\underline{x} \in \mathcal{X} \mid x_{[k]} - d\frac{\sigma}{\sqrt{n}} \leq x_{[1]}\}$$

$$\mathcal{X}_i = \{\underline{x} \in \mathcal{X} \mid x_{[i-1]} < x_{[k]} - d\frac{\sigma}{\sqrt{n}} \leq x_{[i]}\}, \quad 2 \leq i \leq k$$

$$\mathcal{X}_i^{(1)} = \{\underline{x} \in \mathcal{X} \mid x_{[1]} = x_{[i-1]} < x_{[k]} - d\frac{\sigma}{\sqrt{n}} \leq x_{[i]}\} \subset \mathcal{X}_i$$

$$\mathcal{X}_i^{(2)} = \{\underline{x} \in \mathcal{X} \mid x_{[1]} = x_{[i-1]} < x_{[k]} - d\frac{\sigma}{\sqrt{n}} = x_{[i]} = x_{[k-1]}\} \subset \mathcal{X}_i^{(1)}$$

then we have the following theorem.

Theorem 2.6.1. Given a number $P^*(\frac{1}{k} < P^* < 1)$ and locally uniform prior τ for each population π_i, $\underline{X} = \underline{x} \in \mathcal{X}_i$, then

$$P(CS|\psi^M, \underline{X} = \underline{x}) \geq q^*(i)$$

where

$$q^*(i) = \frac{k-i}{k-1}(1 - P^*) + P^* .$$

Hence

$$\psi^M \in \mathcal{S}_{NR}(\tau, P^*) .$$

Proof. It is sufficient to show that

$$\inf_{\underline{x} \in \mathcal{X}_i} \sum_{\ell=i}^{k} p_{[i]}(\underline{x}) = q^*(i) = \frac{k-i}{k-1}(i - P^*) + P^* .$$

Since $\underline{x} \in \mathcal{X}_i$,

$$P(CS|\psi^M, \underline{x}) \geq \inf_{x \in \mathcal{X}_i} P(CS|\psi^M, \underline{x})$$

$$= \inf_{\underline{x} \in \mathcal{X}_i} \sum_{\ell=i}^{k} p_{[i]}(\underline{x}) .$$

Without loss of generality we can assume $\frac{\sigma}{\sqrt{n}} = 1$.

Since

$$\sum_{i=1}^{k} p_{[i]}(\underline{x}) = 1 \quad \forall\, \underline{x} \in \mathcal{X}\,, \text{ and } \forall\, \ell \geq i$$

$p_{[\ell]}(\underline{x})$ is nonincreasing for all $x_{[j]}$, $j \leq i - 1$, we have

$$\inf_{\underline{x} \in \mathcal{X}_i} \sum_{\ell=i}^{k} p_{[\ell]}(\underline{x}) = \inf_{\underline{x} \in \mathcal{X}_i^{(1)}} \sum_{\ell=i}^{k} p_{[\ell]}(\underline{x})$$

$$= 1 - \sup_{x \in \mathcal{X}_i^{(1)}} \sum_{\ell=1}^{i-1} P_{[\ell]}(\underline{x})$$

$$= 1 - \sup_{x \in \mathcal{X}_i^{(1)}} \sum_{\ell=1}^{i-1} \int_{-\infty}^{\infty} \prod_{j \neq \ell} \Phi(t + x_{[\ell]} - x_{[j]}) d\Phi(t)$$

$$= 1 - \sup_{\underline{x} \in \mathcal{X}_i^{(1)}} \sum_{\ell=1}^{i-1} \int_{-\infty}^{\infty} \{ \prod_{j \geq i} \Phi(t + x_{[\ell]} - x_{[j]})$$

$$\cdot \prod_{\substack{j \neq \ell \\ j<i}} \Phi(t + x_{[\ell]} - x_{[j]}) \} d\Phi(t)$$

$$= 1 - \sup_{\underline{x} \in \mathcal{X}_i^{(1)}} \sum_{\ell=1}^{i-1} \int_{-\infty}^{\infty} \{ \prod_{j \geq i} \Phi(t + x_{[\ell]} - x_{[j]})$$

$$\cdot \Phi^{i-2}(t) \} d\Phi(t)$$

$$= 1 - \sum_{\ell=1}^{i=1} \int_{-\infty}^{\infty} \Phi(t - d) \Phi^{k-2}(t) d\Phi(t) \qquad (2.6.2)$$

$$= 1 - (i - 1) \int_{-\infty}^{\infty} \Phi(t-d)\Phi^{k-2}(t)d\Phi(t)$$

$$= (k - i) \int_{-\infty}^{\infty} \Phi^{k-2}(t)\Phi(t - d)d\Phi(t)$$

$$+ \int_{-\infty}^{\infty} \Phi^{k-1}(t + d)d\Phi(t) \qquad (2.6.3)$$

The superimum of (2.6.3) occurs when $\underline{x} \in \mathcal{X}_i^{(2)}$. The last equality follows from the identity

$$(k - 1) \int \Phi^{k-2}(t)\Phi(t - d)d\Phi(t)$$

$$= 1 - \int \Phi^{k-1}(t + d)d\Phi(t),$$

which can be shown by the integration by parts. By (2.6.1), the second term of (2.6.3) equals P*; then use the integration by parts to the first term of (2.6.2), we get

$$\inf_{x \in \mathcal{X}_i} \sum_{\ell=i}^{k} P_{[i]}(\underline{x}) = \frac{k-i}{k-1} [1 - P^*] + P^* \qquad (2.6.4)$$

$$= q^*(i) .$$

Remark 2.6.1. If the procedure ψ^M selects $\pi_{(k)}$ only, i.e. $\underline{X} = \underline{x} \in \mathcal{X}_k$, then by Theorem 2.6.1 we have $p_{[k]}(x) \geq P^*$ so that ψ^B or ψ^B_{NR} selects $\pi_{(k)}$ only. But the converse is not necessarily true.

Remark 2.6.2. For the case k = 2, $\psi^B_{NR} = \psi^M$ a.e. For any given $\underline{X} = \underline{x}$; if $\underline{x} \in \mathcal{X}_2$, then $p_{[2]}(\underline{x}) > P^*$, hence ψ^M and ψ^B_{NR} select the population $\pi_{(2)}$ associated $x_{[2]}$. If $\underline{x} \in \mathcal{X}_1$, and $x_{[2]} - d\frac{\sigma}{\sqrt{n}} < x_{[1]}$ then ψ^M and

ψ^B_{NR} select both populations π_1 and π_2. Since

$$P(X_{[2]} - d\frac{\sigma}{\sqrt{n}} = X_{[1]}) = 0,$$

we have $\psi^B_{NR} = \psi^M$ a.e. .

Remark 2.6.3. The above Theorem and Remark 2.6.1 gives us a lower bound on the value of $\sum_{\ell=i}^{k} P_{[\ell]}(\underline{x})$, over all $\underline{x} \in \mathcal{X}_i$. The exact value of the difference of the selected sizes between ψ^M and ψ^B depends on the observations.

2.7. Applications to Select $\max_{1\le i\le k} \theta_i$, $\theta_i = \frac{\mu_i - a}{\sigma_i}$ for Normal Distribution $N(\mu_i, \sigma_i^2)$, $i = 1,\ldots,k$

Let $\pi_1,\ldots,\pi_k$ be k independent normal populations with mean μ_i and variance σ_i^2, both μ_i and σ_i are unknown. For the goal of finding a random subset which contains the population with maximum $\theta_i = \frac{\mu_i - a}{\sigma_i}$ for some given constant a, we assume that apriori (μ_i, σ_i), $i = 1,\ldots,k$ are independent. Suppose we have n_i independent observations $X_{i1},\ldots,X_{in_i}$ from π_i, and let X_i be their sample mean, $i = 1,\ldots,k$.

Let $Y_1,\ldots,Y_n$ be i.i.d. $\sim N(\mu, \sigma^2)$. If no prior information is available to (μ, σ), we could assign a locally uniform prior $p(\mu, \sigma) \propto \sigma^{-1}$ to (μ, σ), (see Box and Tiao (1973)). And the posterior joint distribution of $\mu' = \mu - a$ and σ, given observations $\underline{Y} = \underline{y} = (y_1,\ldots,y_n)$ is given by

$$P(\mu', \sigma|\underline{y}) = k\sigma^{-(n+1)} \exp\{-\frac{1}{2\sigma^2}[\nu s^2 + n(y' - \mu')^2]\}$$

where

$$\left.\begin{aligned} y' &= y - a, \quad y = \sum_{i=1}^{n} y_i/n \\ \nu s^2 &= \sum (y_i - y)^2, \quad \nu \equiv n - 1 \\ k &= \sqrt{\frac{n}{2\pi}}\,[\tfrac{1}{2}\Gamma(\tfrac{\nu}{2})]^{-1}(\tfrac{\nu s}{2})^{\frac{\nu}{2}}. \end{aligned}\right\} \qquad (2.7.1)$$

Let $\xi = \sqrt{n}\,(\mu-a)/\sigma$, with (2.7.1) the posterior distribution of ξ, given $\underline{Y} = \underline{y}$ is

$$p(\xi|\underline{Y} = \underline{y}) = p(\xi|t)$$

$$= \{2^{\frac{\nu}{2}-1}\Gamma(\tfrac{\nu}{2})\}^{-1}(\frac{\nu}{\nu+t^2})^{\frac{\nu}{2}} \exp\{-\frac{1}{2}(\frac{\xi\nu}{\nu+t^2})\}\,\Gamma(\nu)I_{\nu-1}(\frac{t\xi}{(\nu+t^2)^{1/2}})$$

where

$$t \equiv \sqrt{n}(y - a)/s, \quad \nu \equiv n - 1$$

$$I_\nu(x) = \int_0^\infty (\sqrt{2\pi}\,\Gamma(\nu))^{-1}u^\nu \exp\{-\frac{1}{2}(u + x)^2\}du.$$

Now, let $p(\mu_i, \sigma_i) \propto \sigma_i^{-1}$ be the assigned locally uniform prior to (μ_i, σ_i). Then let $\underline{x} = (x_{11},\ldots,x_{1n_1},\ldots,x_{kn_k})$, we have

$$p_i(\underline{x}) = P(\theta_i = \theta_{[k]}\,|\underline{x})$$

$$= P(\frac{\mu_i - a}{\sigma_i} = \max_{1\le j\le k}(\frac{\mu_j - a}{\sigma_j})\,|\underline{x})$$

$$= P\left(\sqrt{\frac{n_j}{n_i}}\,\xi_i > \xi_j \;\; \forall\, j \neq i \;|\underline{t}\right)$$

$$= \int \prod_{j\neq i} G_{\xi_j}\left(\sqrt{\frac{n_j}{n_i}}\, z|\underline{t}\right) d\, G_{\xi_i}(z|\underline{t}) \qquad (2.7.2)$$

$$= \int \prod_{j\neq i} G_{\xi_j}(z|\underline{t})\, d\, G_{\xi_i}(z|\underline{t}) \quad \text{if } n_1 = \ldots = n_k = n,$$

where G_ξ is the posterior c.d.f. of ξ, given $\underline{x}$ or $\underline{t}$.

By (2.7.2), the Bayes-P* procedure is completely specified.

If the prior distribution for (μ, σ) is the conjugate distribution (see Raiffa and Schlaifer (1960)), then

$$p(\mu, \sigma) \propto \exp\{-\frac{1}{2\sigma^2} n'(\mu-m')^2\} \frac{1}{\sigma} \cdot \exp\{-\frac{1}{2\sigma^2}(\nu'v')\}\sigma^{-\nu'+2}$$
$$= p(\mu|\sigma)\; p(\sigma)$$

that is

$$p(\mu|\sigma) \sim N(m', \sigma^2/n'), \; n' > 0$$

$$p(\sigma) \ni \frac{\nu'v'}{\sigma^2} \sim \chi^2_{\nu'}, \; v', \nu' > 0.$$

Let

$$x' = \frac{nx + n'm'}{n + n'}, \; x \text{ is the sample mean.}$$

$$u^2 = \{(n-1)s^2 + \nu'v' + [(nn')/(n+n')](x - m')^2\}/\nu^*$$

$$\nu^* = (n-1) + \nu' + 1$$

$$\xi^* = (n+n')^{1/2}(\mu - a)/\sigma$$

$$t^* = (n+n')^{1/2}(x' - a)/u,$$

the posterior distribution of ξ^*, given $\underline{x}$ is $p(\xi^*|\underline{x}) = p(\xi^*|t^*)$ which has the same form as $p(\xi|t)$, but replace ξ, t, ν by ξ^*, t^*, ν^*.

Thus for the conjugate prior case, we get

$$p_i(\underline{\underline{x}}) = p(\theta_i = \theta_{[k]} \mid \underline{\underline{x}})$$

$$= \int \prod_{j \neq i} G_{\xi_j^*}\left(\sqrt{\frac{n_j}{n_i}}\, z \mid \underline{t}^*\right) d\, G_{\xi_i^*}(z \mid \underline{t}^*) \qquad (2.7.3)$$

$$= \int \prod_{j \neq i} G_{\xi_j^*}(z \mid t^*) d\, G_{\xi_i^*}(z \mid t^*) \quad \text{if } n_i = n$$

where G_{ξ^*} is the posterior c.d.f. of ξ^* given $\underline{\underline{x}}$ or $\underline{t}$.

Note that (2.7.3) has the same form as (2.7.2), but repalce ξ, t by ξ^*, t^*.

2.8. Applications to Poisson Distributions and Poisson Processes

2.8.1. Poisson Distributions Case

Suppose that $\pi_1,\ldots,\pi_k$ are k independent Poisson populations, where the independent observations $X_{i1},\ldots,X_{in_i}$ from π_i have the Poisson density with parameter λ_i; denoted by $P(\cdot|\lambda_i)$, $i = 1,\ldots,k$.

Let $Y_1,\ldots,Y_n$ be i.i.d. with $p(\cdot|\lambda)$. If we use non-informative prior $p(\lambda) \propto \lambda^{-1/2}$ (Box and Tiao (1973)), then given $\underline{Y} = \underline{y} = (y_1,\ldots,y_n)$ we have the posterior density as follows:

$$p(\lambda|\underline{y}) = c\lambda^{ny - \frac{1}{2}} \exp(-n\lambda)$$

where

$$y = \frac{1}{n}\sum_{i=1}^{n} y_i \text{ and } c = n^{(ny + \frac{1}{2})}[\Gamma(ny + \tfrac{1}{2})]^{-1}.$$

We see that $2n\lambda|\underline{y} \sim \chi^2_{2ny+1}$, the chi-square distribution with 2ny+1 degrees of freedom. Hence by using non-informative prior $p(\lambda_i) \propto \lambda_i^{-1/2}$ for each population π_i, we have

$$p_i(\underline{x}) \equiv p(\lambda_i = \lambda_{[k]} \,|\underline{x})$$

$$= \int_{-\infty}^{\infty} \prod_{j\neq i} \chi^2_{\xi_j}(z \frac{n_i}{n_j}) d\, \chi^2_{\xi_i}(z)$$

where

$$\xi_i = 2n_i\, x_i + 1, \quad x_i = \sum_{j=1}^{n_i} x_{ij}/n_i.$$

If $n_1 = \ldots = n_k$, then

$$p_i(\underline{x}) = \int_0^{\infty} \prod_{j\neq i} \chi^2_{\xi_j}(z) d\, \chi^2_{\xi_i}(z).$$

With $p_i(\underline{x})$, i = 1,...,k, we can apply Bayes-P* selection rules ψ^B and ψ^B_{NR} easily to select a subset which contains the population with the largest parameter λ. On the other hand, if we are interested in selecting the population with the smallest parameter λ, then

$$p_i(\underline{x}) = \int_0^{\infty} \prod_{j\neq i} [1 - \chi^2_{\xi_j}(z \frac{n_j}{n_i})] d\, \chi^2_{\xi_j}(z)$$

$$= \int_0^{\infty} \prod_{j\neq i} [1 - \chi^2_{\xi_j}(z)] d\, \chi^2_{\xi_i}(z) \quad \text{if } n_1 = \ldots = n_k .$$

In this case, the simulation results for selection procedures ψ^B and ψ^B_{NR} are tabulated on Table VII.

2.8.2. Poisson Processes Case

Suppose we have k independent Poisson processes $\{X^{(1)}(t)\},\ldots,\{X^{(k)}(t)\}$ with expected arrival times equal to $\frac{1}{\lambda_1},\ldots,\frac{1}{\lambda_k}$, respectively. Hence for the processes $\{X^{(i)}(t)\}$, the probability that there are m_i arrivals until time t_i is

$$p(X^{(i)}(t_i) = m_i|\lambda_i, t_i) = \frac{(t_i\lambda_i)^{m_i}}{m_i!} e^{-t_i\lambda_i}$$

$$i = 1, 2,\ldots,k, \quad m_i = 0, 1, 2,\ldots$$

If there exists no prior information, then we use the non-informative prior $p(\lambda_i) \propto \lambda_i^{-1/2}$ for all processes. Therefore, we get the posterior density function of λ_i, given (m_i, t_i) as follows:

$$p(\lambda_i|X^{(i)}(t_i) \equiv m_i, t_i) \equiv p(\lambda_i|m_i, t_i)$$

$$= \frac{(t_i\lambda_i)^{m_i + \frac{1}{2} - 1}}{\Gamma(m_i + \frac{1}{2})} t_i e^{-t_i\lambda_i}.$$

Thus $2t_i\lambda_i$ has χ^2 distribution with $2m_i+1$ degrees of freedom, given the number m_i of arrivals before time t_i.

Let $\underline{m} = (m_1,\ldots,m_k)$ and $\underline{t} = (t_1,\ldots,t_k)$, then it can be shown that the Poisson process $\{X^{(i)}(\underline{t})\}$ has the maximum parameter (or minimum mean waiting time), given $(\underline{m}, \underline{t})$ is

$$p_i(\underline{m}, \underline{t}) = \int_0^\infty \prod_{j\neq i} \chi^2_{2m_j+1}(y\,\frac{t_j}{t_i})d\,\chi^2_{2m_i+1}(y) \quad i = 1,\ldots,k. \qquad (2.8.1)$$

Here we list two special cases which are of interest.

(a) Observations of all processes are obtained in a common time interval $[s_i, t + s_i]$. Since Poisson process is stationary, we can assume that $s_i = 0$, and $t_1 = \ldots = t_k = t$. In this case

$$p_i(\underline{m}, \underline{t}) = \int_0^\infty \prod_{j\neq i} \chi^2_{2m_j+1}(y)d\,\chi^2_{2m_i+1}(y)$$

which is independent of t.

(b) All m_i's are equal, i.e. we fix m first, then get observations $\underline{t}$. Hence

$$p_i(m, \underline{t}) = \int_0^\infty \prod_{j\neq i} \chi^2_{2m+1}(y\,\frac{t_j}{t_i})d\,\chi^2_{2m+1}(y).$$

There is an alternative way to approach the cases (a) and (b). Let T_i be the waiting time of the nth arrival in the ith process, then T_i has a gamma distribution with density given by

$$p(t) = \frac{\lambda_i}{\Gamma(m_i)}(\lambda_i t)^{m_i-1} e^{-\lambda_i t} \quad t > 0.$$

If we have only non-informative prior $p(\lambda) \propto \lambda^{-1/2}$; then, given m_i and t_i, $2t_i\lambda_i$ has posterior distribution $\chi^2_{2m_i+1}$, therefore the formula of $p_i(\underline{m}, \underline{t})$ we get here is exactly the same as before.

Remark 2.8.1. Under non-informative prior, in comparing the subset selection problem in k Poisson distributions with the problem in k Poisson processes, it is easily seen that Poisson distributions model is a special case of Poisson processes model, namely, $t_i = n_i$ where n_i denotes the sample size of the ith Poisson population.

2.8.3. Relation Between Selection from Poisson Processes and Selection from Populations with Gamma or Exponential Distribution

Suppose we have k independent populations, the ith population having the gamma distribution with parameters $\alpha = m_i$ (known), $\beta = 1/\lambda_i$ (unknown). Since the random variable T_i, the waiting time until m_i arrivals in a Poisson process with parameter λ_i, has a gamma distribution with parameters $\alpha = m_i$, $\beta = 1/\lambda_i$. If the m_i's are given and if the goals for both selection problems are the same, namely, to select a subset containing the population (process) with the largest parameter λ, then it is easily seen that these are identical problems. Note that in the selection problem of Poisson processes, m_i's might not be the preassigned values but are given random observations whenever t_i's are preassigned values. In this case, the selection problem of Poisson processes is different from that of the gamma distributions.

If the process associated with the minimum parameter λ (or the maximum waiting time) is the best, then the posterior probability of process $\{X^{(i)}(t)\}$ to be the best is analogous to the one obtained

before with the modifications that the integrand function

$$\prod_{j\neq i} \chi^2_{2m_j+1}(y \frac{t_j}{t_i})$$

of (2.8.1) is replaced by

$$\prod_{j\neq i} [1 - \chi^2_{2m_j+1}(y \frac{t_j}{t_i})].$$

2.9. Comparison of the Performance of ψ^B, ψ^B_{NR}, ψ^M and ψ^{MED}

Let π_i, $i = 1,\ldots,k$ be k independent populations, where π_i has the associated c.d.f. $F(x, \theta_i) = F(x - \theta_i)$ with unknown location parameter θ_i. Let $f(x, \theta_i) = f(x - \theta_i)$ be the p.d.f. The goal is to find a small (nontrival) subset which contains the best.

The following subset selection procedure ψ^{MED} based on sample medians is due to Gupta and Singh (1980).

ψ^{MED}: Select π_i if and only if $\tilde{X}_i \geq \tilde{X}_{[k]} - d$

where $\tilde{X}_i$ is the median of the 2m+1 random observations from population π_i and $\tilde{X}_{[k]} = \max_i \tilde{X}_i$. The value d is determined by the following equation so that the P*-condition is met.

$$\int_{-\infty}^{\infty} G(u + d)^{k-1} g(u)du = P^*$$

where

$$g(u) = \frac{(2m+1)!}{(m!)^2} [F(u)]^m [1 - F(u)]^m f(u)$$

$$G(u) = I_{F(u)}(m + 1, m + 1)$$

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

In this section we use Monte Carlo simulation techniques to compare the performance of selection procedures ψ^B, ψ^B_{NR}, ψ^M and ψ^{MED} in the normal means problem. Because both rules ψ^M and ψ^{MED} are not based on any prior information about the unknown parameters, we assume that the prior distribution τ for both ψ^B and ψ^B_{NR} is locally uniformly distributed. Since the selection procedure ψ^M satisfies both the P*-condition and the posterior-P* condition wrt the locally uniform priors, it makes sense to compare the Bayes-P* procedures ψ^B and ψ^B_{NR} with ψ^M and compare ψ^M with ψ^{MED} in terms of efficiency which is the ratio of the probability of a correct selection to the expected selected size. For studying the robustness of these four rules, ψ^B, ψ^B_{NR}, ψ^M and ψ^{MED}, we change the true distribution to non-normal distributions, namely, the logistic, Laplace (the double exponential) and the gross error model (the contaminated distribution), but keep the selection procedure unchanged (i.e. still based on the normal assumption). The Monte Carlo simulation results for both equal distances of the parameters and slippage cases are tabulated. In the simulation study all generated random variables are adjusted to have variance 1. Each time we generate five random variables with the given distribution of each population, then apply the selection procedures. The simulation process is repeated 100 times for each random variable. The relative frequency of selecting the population π_i is

used as an approximation to the probability of selecting the population π_i. The sum of relative frequency of selecting each population π_i, i = 1,...,k is treated as an approximation of the expected selected size. The efficiency EFF of each selection procedure is approximated by the ratio of relative frequency of selecting the best one to the expected size. The simulation results indicate that in all cases we have the performance

$$\psi^B \succ \psi^B_{NR} \succ \psi^M.$$

It should be noted that in the above comparison of the performance, we restrict attention to these rules which satisfy the posterior-P* condition. For small sample size, the efficiency of rule ψ^M tends to be larger than ψ^{MED} under P*-condition.

Remark 2.9.1. The Laplace distribution has the density function

$$f(x - \theta) = \frac{1}{2} e^{-|x-\theta|} \quad -\infty < x < \infty$$

for which the variance is 2.

The logistic distribution has the density function

$$f(x - \theta) = \frac{e^{-(x-\theta)}}{(1+e^{-(x-\theta)})^2}$$

for which the variance Var (X) = $\frac{\pi^2}{3}$.

The gross error model we used has the density function

$$f(x - \theta) = (1 - \varepsilon)\varphi(x - \theta) + \frac{\varepsilon}{4} \varphi(\frac{x-\theta}{4}) \qquad \varepsilon = .15$$

for which φ is the p.d.f. of N(0,1) and the variance

Var (X) = $(1 - \varepsilon) + \varepsilon \cdot 4^2 = 3.25$.

The efficiency of a selection procedure ψ is defined by

$$EFF_{\theta}(\psi) = \frac{P_{\theta}(CS|\psi)}{E_{\theta}(S|\psi)}$$

where $E_{\theta}(S|\psi)$ is the expected selected size.

Discussion and Conclusion

For Table VIII.1 and Table VIII.2 (equal distances case) the P* is .99 and .90 respectively, the common sample size n = 5, k = 5. If the k populations have normal distributions with the unknown parameter configuration $(\theta,\ldots,\theta+(k-1)\Delta)$. common variance 1. From both tables the performance based on either the efficiency or the expected selected size is

$$\psi^{B} \succ \psi^{B}_{NR} \succ \psi^{M}$$

if the posterior-P* condition is considered, and

$$\psi^{M} \succ \psi^{MED}$$

under the P*-condition.

When the true distributions are not normal, but the logistic, the Laplace or the gross error model, the results are very close to the normal case, hence the four rules are robust. From Table VIII.2 all efficiencies are larger than the corresponding ones in Table VIII.1. This is to be expected because the value of P* is smaller in the second table.

For Table X.1 and X.2 (slippage case) the P* is .99 and .90 respectively, the common sample size n = 5, k = 5. If the k populations

have normal distributions with unknown parameter configurations $(\theta,\ldots,\theta+\Delta)$, common variance 1. From both tables the performance based on either the efficiency or the expected selected size is

$$\psi^B \succ \psi^B_{NR} \succ \psi^M$$

if the posterior-P* condition is considered, and if $\Delta\sqrt{n} > 1$

$$\psi^M \succ \psi^{MED}$$

under the P*-condition.

Note that in both equal distances and slippage cases when $\Delta\sqrt{n} > 1$, that is the population means are not very close, the procedures ψ^B and ψ^B_{NR}, wrt the locally uniform priors, always satisfy not only the posterior-P* condition but also $P_{\underline{\theta}}(CS|\psi^B \text{ or } \psi^B_{NR}) \geq P^*$, and the expected selected size of the selection procedure ψ^B or ψ^B_{NR} is much less than the selection procedures ψ^M and ψ^{MED}. For example, in the normal equal distances case, $P^* = .99$, $k = 5$, $\Delta\sqrt{n} = 4$,

$$E(S|\psi^{MED}) - E(S|\psi^B_{NR}) = 0.382;$$

in the normal slippage case, $P^* = .99$, $k = 5$, $\Delta\sqrt{n} = 4$.

$$E(S|\psi^{MED}) - E(S|\psi^B_{NR}) = 1.560.$$

TABLE VII

For procedures ψ^B and ψ^B_{NR} and the parameter configurations $(.5,\ldots,.5\cdot k)$ of k Poisson populations, this table gives the values (based on simulation) of the probability of selecting the population with parameter $.5i$, $i=1,\ldots,k$ and the expected selected size ES. The prior distribution for each population is $p(\lambda) \propto \lambda^{-\frac{1}{2}}$.

n = 10

k		P^* = 0.99		0.95		0.90		0.75	
		ψ^B	ψ^B_{NR}	ψ^B	ψ^B_{NR}	ψ^B	ψ^B_{NR}	ψ^B	ψ^B_{NR}
2	1	.993	1.000	.966	.990	.950	.990	.946	.990
	2	.670	.820	.450	.670	.324	.490	.110	.190
	ES	1.663	1.820	1.415	1.660	1.274	1.480	1.055	1.180
3	1	.998	1.000	.990	1.000	.967	1.000	.929	1.000
	2	.741	.850	.388	.630	.321	.510	.171	.280
	3	.205	.330	.120	.130	.065	.110	.021	.050
	ES	1.944	2.180	1.498	1.760	1.354	1.620	1.121	1.330
4	1	.998	1.000	.981	1.000	.966	.990	.911	.950
	2	.745	.820	.560	.730	.259	.400	.175	.290
	3	.369	.490	.070	.140	.067	.100	.025	.050
	4	.075	.110	.041	.060	.024	.040	0	0
	ES	2.187	2.420	1.653	1.930	1.316	1.530	1.112	1.290
5	1	.996	1.000	.981	.990	.985	1.000	.950	.980
	2	.737	.850	.503	.650	.367	.550	.128	.260
	3	.355	.470	.102	.160	.133	.210	.013	.030
	4	.067	.090	.015	.030	.013	.020	.006	.010
	5	0	0	.006	.010	0	0	0	0
	ES	2.154	2.410	1.607	1.840	1.498	1.780	1.097	1.280

TABLE VIII. 1

Efficiency (EFF) and Expected Selected Size (ES) (based on simulation) of ψ^{B}, ψ^{B}_{NR}, ψ^{M} and ψ^{MED} when the unknown means of the k populations are $\theta,\ldots,\theta + (k-1)\Delta$; the common variance is 1, common sample size n = 5 and the prior for ψ^{B} and ψ^{B}_{NR} is locally uniformly distributed.

TABLE VIII. 1
k = 5, P* = .99

		normal		logistic		Laplace		gross error	
		EFF	ES	EFF	ES	EFF	ES	EFF	ES
$\Delta\sqrt{n} = .5$	ψ^B	.254	3.809	.250	3.963	.259	3.774	.248	3.988
	ψ^B_{NR}	.238	4.110	.233	4.290	.238	4.120	.230	4.300
	ψ^M	.208	4.810	.207	4.840	.210	4.720	.207	4.840
	ψ^{MED}	.208	4.810	.202	4.940	.202	4.940	.201	4.980
$\Delta\sqrt{n} = 1$	ψ^B	.333	2.977	.332	3.005	.336	2.941	.329	3.032
	ψ^B_{NR}	.305	3.280	.304	3.290	.302	3.280	.304	3.290
	ψ^M	.250	4.000	.234	4.280	.246	4.030	.233	4.290
	ψ^{MED}	.232	4.310	.224	4.460	.217	4.600	.214	4.670
$\Delta\sqrt{n} = 2$	ψ^B	.541	1.847	.541	1.839	.541	1.884	.559	1.778
	ψ^B_{NR}	.488	2.050	.481	2.080	.498	2.010	.510	1.960
	ψ^M	.417	2.400	.417	2.400	.515	2.410	.437	2.290
	ψ^{MED}	.368	2.720	.351	2.850	.362	2.760	.368	2.720
$\Delta\sqrt{n} = 4$	ψ^B	.825	1.212	.855	1.169	.821	1.217	.865	1.156
	ψ^B_{NR}	.730	1.370	.806	1.240	.746	1.340	.800	1.250
	ψ^M	.676	1.480	.694	1.440	.680	1.470	.671	1.490
	ψ^{MED}	.575	1.740	.578	1.730	.588	1.700	.571	1.750

TABLE VIII. 2

Efficiency (EFF) and Expected Selected Size (ES) (based on simulation) of ψ^{B}, ψ^{B}_{NR}, ψ^{M} and ψ^{MED} when the unknown means of the k populations are $\theta,\ldots,\theta + (k-1)\Delta$; the common variance is 1, common sample size n = 5 and the prior for ψ^{B} and ψ^{B}_{NR} is locally uniformly distributed.

TABLE VIII. 2
k = 5, P* = .90

		normal		logistic		Laplace		gross error	
		EFF	ES	EFF	ES	EFF	ES	EFF	ES
$\Delta\sqrt{n} = .5$	ψ^{B}	.334	2.449	.353	2.359	.332	2.447	.333	2.611
	ψ^{B}_{NR}	.303	2.870	.313	2.780	.316	2.820	.310	3.030
	ψ^{M}	.239	4.020	.252	3.930	.237	4.090	.231	4.280
	ψ^{MED}	.226	4.200	.231	4.240	.220	4.490	.215	4.650
$\Delta\sqrt{n} = 1$	ψ^{B}	.493	1.807	.523	1.817	.516	1.814	.502	1.839
	ψ^{B}_{NR}	.444	2.160	.454	2.180	.450	2.180	.454	2.160
	ψ^{M}	.327	3.030	.331	3.020	.330	3.000	.331	2.990
	ψ^{MED}	.297	3.270	.287	3.480	.292	3.420	.283	3.530
$\Delta\sqrt{n} = 2$	ψ^{B}	.773	1.275	.770	1.257	.756	1.291	.786	1.227
	ψ^{B}_{NR}	.671	1.490	.678	1.460	.664	1.490	.712	1.390
	ψ^{M}	.535	1.870	.541	1.830	.535	1.870	.579	1.710
	ψ^{MED}	.490	2.040	.490	2.040	.490	2.040	.543	1.840
$\Delta\sqrt{n} = 4$	ψ^{B}	.992	1.008	.981	1.020	.956	1.042	.990	1.010
	ψ^{B}_{NR}	.980	1.020	.935	1.070	.909	1.100	.962	1.040
	ψ^{M}	.877	1.140	.862	1.160	.840	1.190	.901	1.110
	ψ^{MED}	.719	1.390	.800	1.250	.820	1.220	.833	1.200

TABLE IX. 1

Efficiency (EFF) and Expected Selected Size (ES) (based on simulation) of ψ^B, ψ^B_{NR}, ψ^M and ψ^{MED} when the unknown means of the k populations are $\theta,\ldots,\theta,\theta + \Delta$; the common variance is 1, common sample size n = 5 and the prior for ψ^B and ψ^B_{NR} is locally uniformly distributed.

TABLE IX. 1
$k = 5, \quad P^* = .99$

		normal		logistic		Laplace		gross error	
		EFF	ES	EFF	ES	EFF	ES	EFF	ES
$\Delta\sqrt{n} = .5$	ψ^B	.207	4.192	.218	4.253	.226	4.203	.220	4.361
	ψ^B_{NR}	.204	4.520	.212	4.580	.214	4.530	.212	4.670
	ψ^M	.201	4.970	.202	4.950	.200	4.900	.202	4.890
	ψ^{MED}	.203	4.930	.202	4.940	.202	4.960	.200	5.000
$\Delta\sqrt{n} = 1$	ψ^B	.228	4.101	.240	4.133	.240	4.062	.231	4.264
	ψ^B_{NR}	.220	4.400	.224	4.460	.223	4.430	.216	4.580
	ψ^M	.202	4.940	.203	4.930	.203	4.870	.204	4.900
	ψ^{MED}	.204	4.890	.201	4.970	.200	4.980	.200	5.000
$\Delta\sqrt{n} = 2$	ψ^B	.278	3.598	.274	3.645	.263	3.701	.295	3.381
	ψ^B_{NR}	.252	3.970	.253	3.960	.242	4.090	.269	3.720
	ψ^M	.214	4.680	.211	4.750	.214	4.680	.222	4.510
	ψ^{MED}	.208	4.800	.207	4.840	.203	4.920	.203	4.920
$\Delta\sqrt{n} = 4$	ψ^B	.513	1.950	.528	1.893	.520	1.921	.511	1.955
	ψ^B_{NR}	.457	2.190	.474	2.110	.457	2.190	.444	2.250
	ψ^M	.352	2.840	.356	2.810	.348	2.870	.340	2.940
	ψ^{MED}	.267	3.750	.267	3.750	.256	3.910	.258	3.880

TABLE IX. 2

Efficiency (EFF) and Expected Selected Size (ES) (based on simulation) of ψ^B, ψ^B_{NR}, ψ^M and ψ^{MED} when the unknown means of the k populations are $\theta,\ldots,\theta,\theta + \Delta$; the common variance is 1, common sample size n = 5 and the prior for ψ^B and ψ^B_{NR} is locally uniformly distributed.

TABLE IX. 2
k = 5, P* = .90

		normal		logistic		Laplace		gross error	
		EFF	ES	EFF	ES	EFF	ES	EFF	ES
$\Delta\sqrt{n} = .5$	ψ^B	.245	2.813	.267	2.632	.238	2.734	.256	2.858
	ψ^B_{NR}	.243	3.250	.262	3.020	.228	3.160	.240	3.290
	ψ^M	.208	4.480	.219	4.430	.211	4.400	.212	4.490
	ψ^{MED}	.201	4.470	.204	4.600	.202	4.840	.204	4.900
$\Delta\sqrt{n} = 1$	ψ^B	.306	2.645	.349	2.508	.334	2.653	.302	2.871
	ψ^B_{NR}	.288	3.060	.315	2.890	.305	3.050	.275	3.310
	ψ^M	.219	4.420	.231	4.160	.228	4.380	.219	4.520
	ψ^{MED}	.228	4.250	.238	4.420	.207	4.730	.209	4.790
$\Delta\sqrt{n} = 2$	ψ^B	.450	2.149	.463	2.167	.433	2.127	.500	1.890
	ψ^B_{NR}	.396	2.500	.385	2.520	.385	2.470	.441	2.220
	ψ^M	.269	3.720	.272	3.670	.271	3.650	.303	3.270
	ψ^{MED}	.253	3.950	.247	4.050	.238	4.160	.249	4.010
$\Delta\sqrt{n} = 4$	ψ^B	.875	1.143	.862	1.155	.815	1.224	.862	1.153
	ψ^B_{NR}	.794	1.260	.775	1.290	.741	1.350	.800	1.250
	ψ^M	.645	1.550	.610	1.640	.585	1.710	.633	1.580
	ψ^{MED}	.461	2.170	.452	2.210	.495	2.020	.541	1.850

BIBLIOGRAPHY

BIBLIOGRAPHY

Alam, K. (1973). On a multiple decision rule. Ann. Statist., 1, 750-755.

Andersen, E. S. (1953). On the fluctuation of sums of random variables. Math. Scand., 1, 263-285.

Ayer, M., Brunk, H. D., Ewing, G. M., Reid, W. T. and Silverman, E. (1955). An empirical distribution function for sampling with incomplete information. Ann. Math. Statist., 26, 641-647.

Bahadur, R. R. (1950). On the problem in the theory of k populations. Ann. Math. Statist. 21, 362-375.

Bahadur, R. R. and Robbins, H. (1950). The problem of the greater mean. Ann. Math. Statist., 21, 469-487. Correction, 22 (1951), 310.

Barlow, R. E., Bartholomew, D. J., Bremner, J. M. and Brunk, H. D. (1972). Statistical Inference under Order Restrictions. John Wiley, New York.

Barron, A. M. and Gupta, S. S. (1972). A class of non-eliminating sequential multiple decision procedures. Operations Research - Verfahren (Ed. Henn, Künzi and Schubert). Verlag Anton Hain, Meisenheim am Glan, Germany, pp. 11-37.

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.

Bechhofer, R. E. (1958). A sequential multiple decision procedure for selecting the best one of several normal populations with a common unknown variance and its use with various experimental designs. Biometrics, 14, 408-429.

Bechhofer, R. E., Dunnet, C. W. and Sobel, M. (1954). A two-sample multiple-decision procedure for ranking means of normal populations with a common unknown variance. Biometrika, 41, 170-176.

Bechhofer, R. E., Kiefer, J. and Sobel, M. (1968). Sequential Identification and Ranking Procedures. Univ. of Chicago Press, Chicago.

Bechhofer, R. E. and Sobel, M. (1954). A single-sample multiple-decision procedure for ranking variances of normal populations (prelim. report), Abstract. Ann. Math. Statist., 24, 136-137.

Berger, R. L. (1977). Minimax, admissible and gamma-minimax multiple decision rules. Ph.D. Thesis (Mimeo. Ser. No. 489). Dept. of Statist., Purdue Univ., West Lafayette, Indiana.

Berger, R. L. (1979). Minimax subset relation for loss measured by subset size. Ann. Statist., 7, 1333-1338.

Berger, R. L. and Gupta, S. S. (1980). Minimax subset selection rules with applications to unequal variance problems. Scand. J. Statist., 7, 21-26.

Bickel, P. J. and Yahav, J. A. (1977). On selecting a subset of good populations. Statistical Decision Theory and Related Topics - II (Eds. S. S. Gupta and D. S. Moore). Academic, New York, pp. 37-55.

Box, G. E. P. and Tiao, G. C. (1973). Bayesian Inference in Statistical Analysis. Addison Wesley, Massachusetts.

Broström, G. (1977). An improved procedure for selecting all populations better than a standard. Tech. Report 1977-3, Inst. of Math. and Statist., Univ. of Umeå, Sweden.

Carroll, R. J. (1974). Asymptotically nonparametric sequential selection procedures. Mimeo. Ser. No. 944, Inst. of Statist., Univ. of North Carolina, Chapel Hill.

Chernoff, H. and Yahav, J. (1977). A subset selection problem employing a new criterion. Statistical Decision Theory and Related Topics - II (Eds. S. S. Gupta and D. S. Moore). Academic, New York, pp. 93-119.

Deely, J. J. (1965). Multiple decision procedures from an emprical Bayes approach. Ph.D. Thesis (Mimeo. Ser. No. 45). Dept. of Statist., Purdue Univ., West Lafayette, Indiana.

Deely, J. J. and Gupta, S. S. (1968). On the properties of subset selection procedures. Sankhyā Ser. A, 30, 37-50.

Desu, M. M. and Sobel, M. (1968). A fixed subset-size approach to a selection problem. Biometrika, 55, 401-410. Corrections and amendments: 63 (1976), 685.

Desu, M. M. and Sobel, M. (1971). Nonparametric procedures for selecting fixed-size subsets. Statistical Decision Theory and Related Topics (Eds. S. S. Gupta and J. Yackel). Academic, New York, pp. 255-273.

Dunnett, C. W. (1955). A multiple comparison procedure for comparing several treatments with a control. J. Amer. Statist. Assoc., 50, 1096-1121.

Feller, W. (1971). An Introduction to Probability Theory and Its Applications, Vol. II, 2nd Edition. J. Wiley & Sons, New York.

Gibbons, J. D., Olkin, I. and Sobel, M. (1977). Selecting and Ordering Populations. Wiley, New York.

Gnanadesikan, M. (1966). Some selection and ranking procedures for multivariate normal populations. Ph.D. Thesis. Dept. of Statist., Purdue Univ., West Lafayette, Indiana.

Gnanadesikan, M. and Gupta, S. S. (1970). Selection Procedures for multivariate normal distributions in terms of measures of dispersion. Technometrics, 12, 103-117.

Goel, P. K. and Rubin, H. (1977). On selecting a subset containing the best population - A Bayesian approach. Ann. Statist., 5, 969-983.

Gupta, S. S. (1956). On a decision rule for a problem in ranking means. Ph.D. Thesis (Mimeo. Ser. No. 150). Inst. of Statist., Univ. of North Carolina, Chapel Hill.

Gupta, S. S. (1963). On a selection and ranking procedure for gamma populations. Ann. Inst. Statist. Math., 14, 199-216.

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

Gupta, S. S. (1967). On selection and ranking procedures. Trans. Amer. Soc. Qual. Control., pp. 151-155.

Gupta, S. S. and Hsiao, P. (1980). On Γ-minimax, minimax and Bayes procedures for selecting populations close to a control. (submitted).

Gupta, S. S. and Hsu, J.C. (1978). On the performance of some subset selection procedures. Commun. Statist. - Simula. Computa., B7(6), 561-591.

Gupta, S. S. and Huang, D. Y. (1975a). On subset selection procedures for Poisson populations and some applications to the multinomial selection problems. Applied Statistics (Ed. R. P. Gupta). North Holland, Amsterdam, pp. 97-109.

Gupta, S. S. and Huang, D. Y. (1975b). On some parametric and non-parametric sequential subset selection procedures. Statistical Inference and Related Topics, Vol. 2 (Ed. M. L. Puri). Academic, New York, pp. 101-128.

Gupta, S. S. and Huang, D. Y. (1976). Selection procedures for the means and variances of normal populations: unequal sample sizes case. Sankhyā Ser. B, 38, 112-128.

Gupta, S. S. and Kim, W. C. (1980). Γ-minimax and minimax decision rules for comparison of treatments with a control. To appear in Recent Developments in Statistical Inference and Data Analysis. North Holland Publishing Company, 1980, 56-72.

Gupta, S. S. and McDonald, C. C. (1970). On some classes of selection procedures based on ranks. Nonparametric Techniques in Statistical Inference (Ed. M. L. Puri). Cambridge University Press, London, pp. 491-514.

Gupta, S. S., and Miescke, K. J. (1978). On subset selection procedures for ranking means of three normal distributions. Mimeo. Ser. No. 78-19. Dept. of Statist. Purdue Univ., W. Lafayette, Indiana. To appear in Sankhyā Ser. A.

Gupta, S. S. and Miescke, K. J. (1979). On the least favorable configurations in certain two-stage selection procedures. (Mimeo. Ser. No. 79-6). Dept. of Statist., Purdue Univ., West Lafayette, Indiana.

Gupta, S. S. and Nagel, K. (1971). On some contributions to multiple decision theory. Statistical Decision Theory and Related Topics (Eds. S. S. Gupta and J. Yackel). Academic, New York, pp. 79-102.

Gupta, S. S., Nagel, K. and Panchapakesan, S. (1973). On the order statistics from equally correlated normal random variables. Biometrika, 60, 403-413.

Gupta, S. S. and Panchapkesan, S. (1972). On a class of subset selection procedures. Ann. Math. Statist., 43, 814-822.

Gupta, S. S. and Panchapkesan, S. (1979). Multiple Decision Procedures: Theory and Methodology of Selection and Ranking Populations. John Wiley, New York.

Gupta, S. S. and Singh, A. K. (1980). On rules based on sample medians for selection of the largest location parameter. Commun. Statist.-Theor. Meth., A9(12), 1277-1298.

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.

Gupta, S. S. and Sobel, M. (1962a). On selecting a subset containing the population with the smallest variance. Biometrika, 49, 495-507.

Gupta, S. S. and Sobel, M. (1962b). On the smallest of several correlated F-statistics. Biometrika, 49, 509-523.

Gupta, S. S. and Studden, W. J. (1970). On some selection and ranking procedures with applications to multivariate populations. Essays in Probability and Statistics (Eds. R. C. Bose et. al.). University of North Carolina Press, Chapel Hill, pp. 327-338.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. Scand. J. Statist., 6, 65-70.

Hsiao, P. (1979). Some contributions to Gamma-minimax and empirical Bayes selection procedures. Ph.D. thesis (Mimeo. Ser. No. 79-16). Dept. of Statist., Purdue Univ., West Lafayette, Indiana.

Huang, D. Y. (1975). Some contributions to fixed sample and sequential multiple decision (selection and ranking) theory. Ph.D. thesis (Mimeo. Ser. No. 363). Dept. of Statist., Purdue Univ., West Lafayette, Indiana.

Huang, W. T. (1972). Some contributions to sequential selection and ranking procedures. Ph.D. thesis (Mimeo. Ser. No. 295). Dept. of Statist., Purdue Univ., West Lafayette, Indiana.

Kruskal, J. B. (1964). Nonmetric multidimensional scaling: a numerical method. Psychometrika, 29, 115-129.

Lehmann, E. L. (1952). Testing multiparameter hypotheses. Ann. Math. Statist., 23, 541-552.

Lehmann, E. L. (1955). Ordered families of distributions. Ann. Statist., 3, 413-428.

Lehmann, E. L. (1961). Some Model I problems of selection. Ann. Math. Statist., 32, 990-1012.

Mahamunulu, D. M. (1967). Some fixed-sample ranking and selection problems. Ann. Math. Statist., 38, 1079-1091.

McDonald, G. C. (1969). On some distribution-free ranking and selection procedures. Ph.D. thesis (Mimeo. Ser. No. 174). Dept. of Statist., Purdue Univ., West Lafayette, Indiana.

Miescke, K. J. (1979). Bayesian subset selection for additive and linear loss functions. Commun. Statist. - Theor. Meth., A8(12), 1205-1226.

Mosteller, F. (1948). A k-sample slippage test for an extreme population. Ann. Math. Statist., 19, 58-65.

Nagel, K. (1970). On subset selection rules with certain optimality properties. Ph.D. thesis (Mimeo. Ser. No. 222). Dept. of Statist., Purdue Univ., West Lafayette, Indiana.

Naik, U. D. (1975). Some selection rules for comparing p process with a standard. Comm. Statist., 4, 519-535.

Naik, U. D. (1977). Some subset selection problems. Commun. Statist.-Theor. Meth., Ab(10), 955-966.

Naik, U. D. (1978). On selection procedures based on the exceedance probability. Calcutta Statist. Assoc. Bull., 29-40.

Paulson, E. (1949). A multiple decision procedure for certain problems in the analysis of variance. Ann. Math. Statist., 20, 95-98.

Paulson, E. (1962). A sequential procedure for comparing several experimental categories with a standard or control. Ann. Math. Statist., 33, 438-443.

Paulson, E. (1963). A sequential decision procedure for choosing one of k hypothesis concerning the unknown mean of a normal distribution. Ann. Math. Statist., 34, 549-554.

Paulson, E. (1964). A sequential procedure for selecting the population with the largest mean from k normal populations. Ann. Math. Statist., 35, 174-180.

Paulson, E. (1967). Sequential procedures for selection of the best one of several binomial populations. Ann. Math. Statist., 38, 117-123.

Rabinowitz, P. and Weiss, G. (1959). Tables of abscissas and weights for numerical evaluation of integrals of the form $\int_0^\infty e^{-x} x^n f(x)dx$. Math. Tables and other Aids to Comput., Vol. XIII, 68, 285-294.

Rizvi, M. H. and Sobel, M. (1967). Nonparametric procedures for selecting a subset containing the population with the largest α-quantile. Ann. Math. Statist., 38, 1788-1803.

Santner, T. J. (1975). A restrict subset selection approach to ranking and selection problems. Ann. Statist., 3, 334-349.

Seal, K. C. (1955). On a class of decision procedures for ranking means of normal populations. Ann. Math. Statist., 26, 387-398.

Seal, K. C. (1957). An optimum decision rule for ranking means of normal populations. Calcutta Statist. Assoc. Bull., 7, 131-150.

Sobel, M. (1967). Nonparametric procedures for selecting the t populations with the largest α-quantiles. Ann. Math. Statist., 38, 1804-1816.

Sobel, M. and Huyett, M. J. (1957). Selecting the best one of several binomial populations. Bell System Tech. J., 36, 537-576.

Studden, W. J. (1967). On selecting a subset of k populations containing the best. Ann. Math. Statist., 38, 1072-1078.

Tamhane, A. C. and Bechhofer, R. E. (1977). A two-stage minimax procedure with screening for selecting the largest normal mean. Commun. Statist. - Theor. Meth., A6(11), 1003-1033.

Tamhane, A. C. and Bechhofer, R. E. (1979). A two-stage minimax procedure with screening for selecting the largest mean (II): An improved PCS lower bound and associated tables. Commun. Statist. - Theor. Meth., A8(4), 337-358.

Tukey, J. W. (1960). A survey of sampling from contaminated distributions. Contributions to Probability and Statistics (Eds. I. Olkin and others). Stanford Univ. Press, 39, 448-458.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Mimeograph Series #80-22	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE *(and Subtitle)* Some Results in the Theory of Subset Selection Procedures		5. TYPE OF REPORT & PERIOD COVERED 6. PERFORMING ORG. REPORT NUMBER Mimeo. Series #80-22
7. AUTHOR(s) Hwa-Ming Yang		8. CONTRACT OR GRANT NUMBER(s) ONR N00014-75-C-0455
9. PERFORMING ORGANIZATION NAME AND ADDRESS Purdue University Department of Statistics West Lafayette, IN 47907		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Washington, DC		12. REPORT DATE August 1980 13. NUMBER OF PAGES 137
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)*

Selection Procedures, Subset Selection, Stagewise Procedures, Better than Control, Ordering Prior, Isotonic Regression, Isotonic Estimators, Random Walk, Prior Distribution, Non-informative Prior, Posterior-P* Condition, Bayes-P* Procedure.

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

Selection and ranking (ordering) problems in statistical inference arise mainly because the classical tests of homogeneity are often inadequate in certain situations where the experimenter is interested in comparing k (≥ 2) populations, treatments or processes with the goal of selecting one or more worthwhile (good) populations.

Chapter I of this thesis considers the problem of selecting a subset containing all populations that are better than a control under the assumptions of

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)*

an ordering prior. Here, by an ordering prior we mean that there exists a known simple or partial order relationship among the unknown parameters of the treatments (excluding the control). Three new selection procedures are proposed and studied. These procedures do meet the usual requirement that the probability of a correct selection is greater than or equal to a pre-determined number P*. Two of the three procedures use the isotonic regression over the sample means of the k treatments with respect to the given ordering prior. Tables which are necessary to carry out the selection procedure with isotonic approach for the selection of unknown means of normal populations and gamma populations are given. Monte Carlo comparisons on the performance of several procedures for the normal or gamma means problem were carried out in several selected cases. The results of this study seem to indicate that the procedures based on isotonic estimators always have superior performance, especially, when there are more than one bad populations (in comparison with the control).

Chapter II deals with a new 'Bayes-P*' approach about the problem of selecting a subset which contains the 'best' of k populations. Here, by best we mean the (unknown) population with the largest unknown mean. The (non-randomized) Bayes-P* rule refers to a rule with minimum risk in the class of (non-randomized) rules which satisfy the condition that the posterior probability of selecting the best is at least equal to P*. Given the priors of the unknown parameters, two 'Bayes-P*' subset selection procedure ψ^B and ψ^B_{NR} (randomized and non-randomized, respectively) under certain loss functions are obtained and compared with the classical maximum-type means procedure ψ^M. The comparisons of the performance of ψ^B with ψ^B_{NR} and ψ^M, based on Monte Carlo studies, indicate that the procedure ψ^B always has higher 'efficiency' and smaller expected size of the selected subset. The studies also indicate that ψ^B is robust when the true distributions are not normal but are some other symmetric distributions such as, the logistic, the double exponential (Laplace) and the gross error model (the contaminated distribution).