ON BAYES AND GAMMA-MINIMAX SUBSET SELECTION RULES*

by

Woo-Chul Kim
Purdue University

Department of Statistics
Division of Mathematical Sciences
Mimeograph Series #79-7

May 1979

*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

Historically, most problems of statistical inference have been formulated as those of estimation or testing of hypotheses. In many practical situations the experimenter is faced with the problem of comparing several populations. Suppose we have $k(k \geq 2)$ populations $\pi_1,\ldots,\pi_k$ whose qualities are characterized by real-valued parameters $\theta_1,\ldots,\theta_k$, respectively. The classical approach to this problem is that of testing the hypothesis of homogeneity, i.e., H_0: $\theta_1 =\ldots= \theta_k$. But in many situations, the goal of the experimenter is not just to decide whether all the parameters are equal or not. One of the more frequently occurring situations for which this is so arises when the experimenter wishes to find a subset of $\{\pi_1,\ldots,\pi_k\}$ which, in some sense, is better than the rest of the given populations. Mosteller (1948), Paulson (1949) and Bahadur (1950) were among the first research workers to recognize the inadequacy of the classical tests of homogeneity and to formulate the problem as multiple decision problems known as ranking and selection problems.

Two formulations for selection and ranking problems have been considered in the classical framework. To fix ideas suppose π_i is better than π_j if $\theta_i > \theta_j$. Consider the problem of selecting the 'best' population, i.e., the population associated with

$\max_{1 \le i \le k} \theta_i$. The first formulation is called the 'indifference zone' approach of Bechhofer (1954); the experimenter is allowed to select only one population which is the best one with a preassigned minimum probability P*, whenever the unknown parameters lie outside a zone of indifference. Contributions using this approach in the decision-theoretic framework have been made by Bahadur and Goodman (1952), Lehmann (1966), Eaton (1967) and Alam (1973) among others. The second formulation due to Gupta (1956, 1965) is known as the 'subset selection' approach in which the experimenter selects a subset of random size depending on the observed data such that it contains the best population with at least probability P*. Decision-theoretic contributions in this framework have been made by Studden (1967), Deely and Gupta (1968), Berger (1977), Chernoff and Yahav (1977), Bickel and Yahav (1977), Goel and Rubin (1977), Hsu (1977) and Gupta and Hsu (1978). Especially the last five preceding papers deal with Bayes selection rules. An up-to-date and comprehensive bibliography for both these approaches can be found in Gupta and Panchapakesan (1979). There also have been attempts in the literature to modify these basic formulations. In one such modification, Fabian (1962) called π_i good if $\theta_i \ge \max_{1 \le j \le k} \theta_j - \Delta$ and bad, otherwise, where the positive constant Δ is usually specified by the experimenter. The goal in this framework is to select the good populations and screen out the bad populations. Contributions along these lines have also been made by Fabian (1962), Desu (1970), Santner (1976), Panchapakesan and Santner (1977) and Broström (1978). Optimality of some of these rules has been studied by Bjørnstad (1978).

A slightly different situation arises when, in addition to the k treatment populations, a control population π_0 is considered and the goal is to partition the k treatment populations with regard to π_0. Following the initial investigation of Paulson (1952), Dunnett (1955), Gupta and Sobel (1958), Bhattacharyya (1956,1958), Lehmann (1961), Tong (1969), Randles and Hollander (1971) and Miescke (1979), among others, have studied this problem in several different formulations.

The present thesis consists of investigations of multiple decision problems mentioned in the preceding two paragraphs, and also some related topics. In Chapter 1, the problem of selecting good populations is studied from a decision-theoretic Bayesian point of view. We consider a loss function which seems natural for this problem. A theorem is proved to help find the Bayes rule with respect to a permutationally symmetric prior. Some properties of the Bayes rule are derived when an iid prior is assumed. Then to get insight on the performance of the Bayes rule the loss is assumed to be c_1 if we select a bad population and c_2 if we exclude a good population. The rest of Chapter 1 pertains to further simplification and approximation of the Bayes rules since these are often analytically and computationally intractable. Some special cases, namely, normal and gamma distributions are considered with specific prior distributions. Especially, it is shown that, for the normal populations, classical rules of the type proposed by Gupta (1956) and studied by Desu (1970) in this framework turn out to be close approximations to the Bayes rules. In this connection, Monte Carlo studies are also performed to see how well the classical rules proposed by Seal (1955) and by Gupta (1956) approximate the Bayes rules in terms of overall risks wrt exchangeable priors. The results of this study indicate that the rules of the type proposed by Gupta perform almost as well as the Bayes rules throughout the cases studied. Similar results have been found by Chernoff and Yahav (1977) and Gupta and Hsu (1978) under different frameworks. Since Bayes rules typically require numerical integrations to implement, this makes them usually unsuitable for practical use. Therefore, it was deemed useful to provide tables which give the 'best' classical rules with performances that are sufficiently close to those of the Bayes rules. The tables also provide the average number of bad populations selected that of good ones excluded and the proportion of times that the Bayes rules coincide with these classical rules.

Chapter 2 deals with the problem of partitioning k treatment populations with regard to a control population. The goal is to partition the k treatment populations into 'better' populations, 'worse' populations or 'close' populations in an optimal way. Loss function similar to the one in Bhattacharyya (1958) is assumed, and Γ-minimax rules and minimax rules are derived for the known control case when the parameters of interest are location parameters. Normal populations with unknown means and known variances are studied as a special example. When the parameter of the control population is unknown, both Γ-minimax rules and minimax rules (in a certain class of decision rules) are derived for the location parameter populations. Normal populations with unknown means and known variances, and normal populations with known means and unknown variances are provided as examples. The results regarding the minimax rules generalize a result of Bhattacharyya (1958) for the normal populations with a known control population. In the last section, comparisons are made between the Γ-minimax rules and the corresponding Bayes rules wrt independent normal priors for the case of normal populations with a common known variance. Tables are provided which shed light on the relative performance of these rules.

Chapter 3 deals with a selection problem in reliability theory and another for the selection of scale parameters. The first part deals with the problem of selecting components (units) for parallel and series systems from k populations with exponentially distributed lifelength times. Loss functions are assumed which are inversely proportional to the expected lifelength of the system corresponding to a possible choice of the units of the system. A similar problem has been studied by Broström (1977) for the 1-out-of-2 system when there are only two populations. An optimal rule is given for the series system, and the Bayes rule wrt natural conjugate prior is derived for the 1-out-of-2 system when we have k populations. Tables to implement the Bayes rule are provided at the end of the chapter. The second part of Chapter 3 deals with the investigation of the selection procedures based on robust estimators of measures of dispersion for selecting the populations in terms of scale parameters. Several selection procedures for this problem have been proposed by many researchers, most of them being extensions of k-sample tests. The approach here differs from these in that estimators of population dispersions are directly employed in constructing the selection procedures. We study the problem of selecting t populations associated with the t smallest scale parameters under the indifference-zone approach and the problem of selecting a subset containing the population associated with the smallest scale parameter. Large sample solutions for both problems are derived and asymptotic relative efficiencies (following the definitions of Lehmann (1963) and McDonald (1969)) of the proposed procedures are studied. These turn out to be same as those of the corresponding estimators in Bickel and Lehmann (1976).

CHAPTER 1

SELECTION OF GOOD POPULATIONS

1.1 Introduction

Suppose we have k populations $\pi_1,\ldots,\pi_k$ $(k \geq 2)$ from which we wish to select a subset which contains 'good' populations, where the quality of the i-th population is characterized by the unknown parameter θ_i. A population π_i is said to be

$$\begin{aligned} &\text{good if } \theta_i \geq \theta_{[k]}-\Delta, \\ &\text{bad if } \theta_i < \theta_{[k]}-\Delta \end{aligned} \tag{1.1.1}$$

where $\theta_{[k]} = \max_{1 \leq j \leq k} \theta_j$ and Δ is a specified positive constant. Clearly we wish to select a subset containing as many good populations as possible while excluding the bad ones. Therefore, any reasonable loss function should have two loss components, i.e., one incurred by excluding the good populations and another due to including the bad ones. Selection problems for the related goals have been considered in the literature by Fabian (1962), Desu (1970), Carroll, Gupta and Huang (1975), Santner (1976), Panchapakesan and Santner (1977), Broström (1978) and Bjørnstad (1978). In all these papers the selection rules are chosen to control one component of the loss, and then the other operating characteristics of these rules are studied. Recently Miescke (1978) has studied this problem from Bayesian point of view. We will treat this problem in the Bayesian framework considering two loss components simultaneously.

In Section 1.2 some definitions and notations are introduced, and a decision-theoretic formulation of the problem is given. We describe loss functions which seem natural in this framework, and, for these loss functions, a theorem is proved to find the Bayes rule wrt a permutationally symmetric prior. Some properties of Bayes rules are also given when an iid prior is assumed.

In Section 1.3 it is assumed that the loss is c_1 if a bad population is selected and c_2 if a good population is excluded; the Bayes rule wrt an exchangeable prior is derived. For the same loss function some results about the minimax rules are obtained for $k = 2$. In Section 1.4 a simplification of the Bayes rule is given by assuming a particular form of the posterior distribution, and it is shown that, in some sense, some rules studied in the past are natural approximations to the Bayes rule.

Normal and gamma populations are studied as special examples in Section 1.5. For the normal populations the rules of the type proposed by Gupta (1956) are shown to be asymptotically equivalent to the Bayes rule (as the sample size approaches infinity), and also it is shown that these are also extended Bayes rules. Further simplification of the Bayes rule for gamma populations is given.

Section 1.6 consists of Monte Carlo comparisons of the performances of the Bayes rule, the rules of the type proposed by Gupta (1956), and those proposed by Seal (1955) when the prior is assumed to be a permutationally symmetric multivariate normal distribution. The results of the Monte Carlo study indicate that the rules of the type proposed by Gupta perform almost as well as the Bayes rule throughout the cases studied, while those proposed by Seal perform poorly in most cases.

1.2 Formulation of the problem

Let $X_1,\ldots,X_k$ denote the random variables representing the k populations $\pi_1,\ldots,\pi_k$, respectively. Let S_k be the symmetric group of all permutations Ψ: $\{1,2,\ldots,k\} \to \{1,2,\ldots,k\}$, and Ψ_{ij} denote an element in S_k which interchanges i and j leaving all other elements of $\{1,2,\ldots,k\}$ fixed. For $\underline{x} \in R^k$ and $\Psi \in S_k$ define $\underline{x}_\Psi$ by $(\underline{x}_\Psi)_i = x_{\Psi^{-1}i}$.

From now on it is assumed that

(i) given $\underline{\theta} = (\theta_1,\ldots,\theta_k) \in \Theta^k$, the random variables $X_1,\ldots,X_k$ are independently distributed with X_i having p.d.f. $f(\cdot,\theta_i)$ and

(ii) $f(x,\theta)$ has the monotone likelihood ratio (MLR) property in x and θ.

The action space $\mathcal{A}$ consists of all possible non-empty subsets of $\{1,\ldots,k\}$. The action $a \in \mathcal{A}$ is interpreted as the action of selecting the populations $\{\pi_i, i \in a\}$ which are asserted to be the good populations. We will restrict our attention to loss functions which are invariant under S_k and monotone in the sense of Eaton (1967), that is, for any $\psi \in S_k$, and $a \in \mathcal{A}$,

$$L(\underline{\theta}_\psi, \Psi a) = L(\underline{\theta}, a) \quad \text{and}$$
$$L(\underline{\theta}, a) \leq L(\underline{\theta}, \Psi_{ij} a) \text{ for } \underline{\theta} \in \Theta^k,\ \theta_i \geq \theta_j,\ i \in a \text{ and } j \notin a \tag{1.2.1}$$

where Ψa denotes the image of a under $\Psi \in S_k$. Note that the problem is invariant under S_k in the above framework, and therefore it seems reasonable to consider a prior τ invariant under S_k. In the remainder of this chapter we will consider only such a permutationally symmetric prior.

Since Bayes rules are of main interest, attention can be restricted to the non-randomized decision rules $\delta: R^k \to a$. From this point for sake of simplicity we will use δ also for the action $\delta(\underline{x})$ taken by the rule δ whenever no confusion arises. Let $\mathcal{D}$ denote the class of all non-randomized decision rules. Let $r_\tau(\delta, \underline{x})$ denote the posterior risk of a decision rule $\delta \in \mathcal{D}$, given $\underline{x}$, when the prior is given by τ. Let $x_{[1]} \leq x_{[2]} \leq \cdots \leq x_{[k]}$ denote the ordered observations where the ties for a label are broken at random, and $\pi_{(i)}$ and $\theta_{(i)}$ denote the π and θ associated with $x_{[i]}$, $i = 1,\ldots,k$. For $j = 1,\ldots,k$, let δ_j denote the decision rule which chooses the subset associated with the j largest observations with probability one.

By partitioning the action space $\mathcal{A}$ into k components $\mathcal{A}_j$, $j = 1,\ldots,k$, where $\mathcal{A}_j$ consists of all the subsets of size j, Goel and Rubin (1977) proved the following result which seems useful in the selection problem.

Lemma 1.2.1. If the prior distribution τ of $\underline{\theta}$ is permutationally symmetric on Θ^k, then the Bayes rule δ^* for the loss function satsifying (1.2.1) satisfies

$$r_\tau(\delta^*,\underline{x}) = \underset{1\le j\le k}{\mathrm{Min}}\ r_\tau(\delta_j,\underline{x}).$$

When our goal is to select a subset containing the 'best' population, i.e., the population associated with $\theta_{[k]}$, most of the loss functions considered in the literature contain two loss components; the optimal solution is provided by the rule which selects all the populations if we were to study the problem wrt one of the two loss components only, and the optimal solution is to select the population $\pi_{(k)}$ associated with $x_{[k]}$ if the other component only is considered. In our formulation the experimenter is willing to accept all the populations which are reasonably close to the 'best' while screening out all the bad ones. Therefore, it seems reasonable that any loss function should contain two components: the first one depending only on the bad populations selected in the subset and the second depending only on the good populations which are not selected in the subset. Such a loss function reflects the loss due to misclassification of good or bad populations. One such general loss function can be written as follows: For $\underline{\theta} \in \Theta^k$ and $a \in \mathcal{G}$,

$$L(\underline{\theta},a) = \sum_{i\in a} L_B(\theta_i-\theta_{[k]}+\Delta) + \sum_{i\notin a} L_G(\theta_i-\theta_{[k]}+\Delta), \tag{1.2.2}$$

where $L_B(L_G)$ is a non-increasing (non-decreasing) function such that $L_B(y) = 0$ for $y \ge 0$ and $L_G(y) = 0$ for $y < 0$. By this loss function we mean that there is no loss for the correct judgment, the loss $L_B(\theta_i-\theta_{[k]}+\Delta)$ incurred by selecting a bad population π_i is non-decreasing in θ_i, the parameter associated with π_i. Similar arguments hold for the second component $L_G(\cdot)$ of the loss. It is easy to see that the loss function of the type (1.2.2) satisfies (1.2.1). With this loss function, we have the following result.

Lemma 1.2.2. Assume the prior distribution, $\tau(\underline{\theta})$, of $\underline{\theta}$ is permutationally symmetric on Θ^k and the loss function is given by (1.2.2). Then the following relation holds.

$$D_i - D_{i-1} \geq 0, \ i = 2,\ldots,k-1, \tag{1.2.3}$$

where $D_i = r_\tau(\delta_{i+1},\underline{x}) - r_\tau(\delta_i,\underline{x})$ for $i = 1,\ldots,k-1$.

Proof. It follows from (1.2.2) that D_i is given by

$$\begin{aligned} D_i &= E[L_B(\theta_{(k-i)}-\theta_{[k]}+\Delta)|\underline{x}] - E[L_G(\theta_{(k-i)}-\theta_{[k]}+\Delta)|\underline{x}] \\ &= E[\ell(\theta_{(k-i)}-\theta_{[k]}+\Delta)|\underline{x}] \quad \text{where } \ell(\cdot) = L_B(\cdot) - L_G(\cdot). \end{aligned}$$

Therefore, D_i-D_{i-1} can be written as

$$\begin{aligned} D_i-D_{i-1} &= n(\underline{x}) \sum_{j=0}^{2} \int_{\Theta_j} [\ell(\theta_{(k-i)}-\theta_{[k]}+\Delta)-\ell(\theta_{(k-i+1)}-\theta_{[k]}+\Delta)] \\ &\qquad f(\underline{x},\theta)d\tau(\underline{\theta}) \\ &= n(\underline{x})\int_{\Theta_1} [\ell(\theta_{(k-i)}-\theta_{[k]}+\Delta)-\ell(\theta_{(k-i+1)}-\theta_{(k)}+\Delta)][f(\underline{x},\theta)- \\ &\qquad f(\underline{x},\underline{\theta}_\psi)]d\tau(\theta) \\ &\geq 0, \end{aligned}$$

where $\Theta_0 = \{\underline{\theta} \in \Theta^k | \theta_{(k-i)} = \theta_{(k-i+1)}\}$, $\Theta_1 = \{\underline{\theta} \in \Theta^k | \theta_{(k-i)} < \theta_{(k-i+1)}\}$, $\Theta_2 = \{\underline{\theta} \in \Theta^k | \theta_{(k-i)} > \theta_{(k-i+1)}\}$, $f(\underline{x},\underline{\theta}) = \prod_{i=1}^{k} f(x_i,\theta_i)$, $n(\underline{x})$ is a normalizing factor and $\underline{\theta}_\psi$ is obtained from $\underline{\theta}$ by interchanging the components $\theta_{(k-i)}$ and $\theta_{(k-i+1)}$. The second equality follows from the symmetry of τ and the last inequality follows from the MLR property of $f(x,\theta)$ and the monotonicity of $\ell(\cdot)$.

From Lemma 1.2.1 and Lemma 1.2.2, we derive a Bayes rule for a loss function given by (1.2.2).

Theorem 1.2.1. Assume the loss function is given by (1.2.2). Then the Bayes rule δ^* wrt a permutationally symmetric prior τ is given by

$$\delta^* = \delta_{i^*} \quad \text{for } i^* = \min\{i: D_i \geq 0, \ i = 1,\ldots,k-1\}$$

where min ϕ is defined to be k.

Before we state some properties of the Bayes rule, we recall the following definitions (see Nagel (1970) and Santner (1975).)

Definition 1.2.1. Let $\delta(\underline{x},i)$ denote the individual probability of including the population π_i in the selected subset $\delta(\underline{x})$. Then a selection rule δ is called just if and only if $\delta(\underline{x},i) \leq \delta(\underline{x}',i)$ whenever $x_i \leq x_i'$ and $x_j \geq x_j'$ for $j \neq i$.

Definition 1.2.2. A selection rule δ is said to be strongly monotone if and only if, for any $i = 1,\ldots,k$,

$$E_{\underline{\theta}}[\delta(\underline{x},i)] \text{ is } \uparrow \text{ in } \theta_i \text{ when all other components of } \underline{\theta} \text{ are fixed,}$$
$$\text{is } \downarrow \text{ in } \theta_j \ (j \neq i) \text{ when all other components of } \underline{\theta} \text{ are fixed.}$$

Corollary 1.2.1. Assume that $\theta_1,\ldots,\theta_k$ are, a priori, independently identically distributed and the loss function is given by (1.2.2). Then the Bayes rule δ^* in Theorem 1.2.1 is just and strongly monotone.

Proof. It follows from Theorem 1.2.1 that the Bayes rule δ^* selects π_i if and only if $x_i = x_{[k]}$ and/or $\int \ell(y)dQ_i(y|\underline{x}) < 0$, where

$$Q_i(y|\underline{x}) = \begin{cases} 1-\int \prod_{j\neq i} G(z+\Delta-y|x_j)dG(z|x_i) & \text{if } y < \Delta \\ 1 & \text{otherwise} \end{cases}$$ and $G(\cdot|x_i)$ is

the posterior cdf of θ_i, given $\underline{x}$. Therefore, to show that δ^* is just, it suffices to show that $Q_i(y|\underline{x})$ is stochastically smaller than $Q_i(y|\underline{x}')$ whenever $x_i \leq x_i'$ and $x_j \geq x_j'$ for all $j \neq i$, since $\ell(\cdot)$ is non-increasing. This follows from the fact that $G(\cdot|x) \geq G(\cdot|x')$ if $x \leq x'$. Similarly, we can show that δ^* is strongly monotone using a theorem in Barr and Rizvi (1966).

Remark 1.2.1. If a selection rule δ is strongly monotone, then δ is monotone (see Santner (1975)), i.e. $E_{\underline{\theta}}[\delta(\underline{x},i)] \geq E_{\underline{\theta}}[\delta(\underline{x},j)]$ if $\theta_i \geq \theta_j$.

1.3 Loss function $c_1 \sum_{i\in a} I_{(-\infty,0)}(\theta_i-\theta_{[k]}+\Delta)+c_2 \sum_{i\notin a} I_{[0,\infty)}(\theta_i-\theta_{[k]}+\Delta)$

Even though Theorem 1.2.1 describes a Bayes rule for the loss function of the type (1.2.2), we need a more specific loss function as well as more assumptions about the prior distribution of $\underline{\theta}$ to specify the Bayes rule more explicitly. Examples of the loss functions satisfying (1.2.2) might be given as follows.

$$L_B(\theta_i-\theta_{[k]}+\Delta) = c_1 I_{(-\infty,0)}(\theta_i-\theta_{[k]}+\Delta),\ L_G(\theta_i-\theta_{[k]}+\Delta) = c_2 I_{[0,\infty)}(\theta_i-\theta_{[k]}+\Delta), \tag{1.3.1}$$

$$L_B(\theta_i-\theta_{[k]}+\Delta) = (\theta_i-\theta_{[k]}+\Delta)^-,\ L_G(\theta_i-\theta_{[k]}+\Delta) = (\theta_i-\theta_{[k]}+\Delta)^+ \tag{1.3.2}$$

where $I_A(y)$ is the usual indicator function, y^- is the negative part of y, y^+ is the positive part of y, $c_1 > 0$ and $c_2 > 0$.

From now on we will, unless otherwise mentioned, consider only the loss function given by (1.3.1) and assume that given $W = w$, $\theta_1,\ldots,\theta_k$ are iid random variables with a density and the distribution of W is known. Such a prior distribution of $\underline{\theta}$ will be called <u>exchangeable</u>. Note that we may assume that $c_1+c_2 = 1$ without loss of generality, and c_1 and c_2 can be interpreted as the relative weights of the losses due to two different sources. It is easy to see that, given $\underline{X} = \underline{x}$ and $W = w$, θ_i's are, a posteriori, independently distributed and the distribution of θ_i depends on $\underline{x}$ only through x_i. Let $G_i(\cdot) = G(\cdot|x_{[i]},w)$ and $H(w|\underline{x})$ denote the posterior cdf of $\theta_{(i)}$, given $\underline{x}$ and w, and the conditional cdf of W given $\underline{x}$, respectively. The bounds on D_i in the following lemma seem to be useful to simplify the Bayes rule.

<u>Lemma 1.3.1</u>. For $i = 1,\ldots,k-1$, $D_i = c_1-P(\theta_{(k-i)} \geq \theta_{[k]}-\Delta|\underline{x})$ satisfies the following relations.

$$D_i = c_1-\iint \prod_{j\neq k-i} G_j(z+\Delta)dG_{k-i}(z)dH(w|\underline{x})$$

$$D_i \geq c_1 - \iint G_k(z+\Delta)dG_{k-i}(z)dH(w|\underline{x}) \equiv c_1-u_1(i),$$

$$D_i \geq c_1 - \iint G^i_{k-i+1}(z+\Delta)dG_{k-i}(z)dH(w|\underline{x}) \equiv c_1-u_2(i) \text{ and} \tag{1.3.3}$$

$$D_i \leq c_1 - \iint G^{k-i-1}_{k-i}(z+\Delta)G^i_k(z+\Delta)dG_{k-i}(z)dH(w|\underline{x}) \equiv c_1-v_1(i).$$

Proof. The first inequality follows from the fact that $\theta_{[k]} \geq \theta_{(k)}$ and the next inequalities follow from the fact that $G_i(\cdot) \geq G_j(\cdot)$ for $i < j$.

Next, we state a theorem giving a simpler version of the Bayes rule from (1.3.3) and Theorem 1.2.1.

Theorem 1.3.1. Assume the loss function is given by (1.3.1). Then the Bayes rule δ^* wrt an exchangeable prior is given as follows. Let $u(i) = \min\{u_1(i), u_2(i)\}$ for $i = 1,\ldots,k-1$, then

(i) $u(1) \leq c_1 \Rightarrow \delta^* = \delta_1$

(ii) let $i_0 = \min\{i: u(i) \leq c_1, i = 2,\ldots,k-1\}$ and $j_0 = \max\{j|c_1 < v_1(j), j=1,\ldots,i_0-1\}$,

then $\delta^* = \delta_{i^*}$ where $i^* = \min\{m: D_m \geq 0, j_0+1 \leq m \leq i_0\}$.

Corollary 1.3.1. For $k = 2$, the Bayes rule δ^* is given by

$$\delta^* = \begin{cases} \delta_1 & \text{if } \iint G_2(z+\Delta)dG_1(z)dH(w|\underline{x}) \leq c_1 \\ \delta_2 & \text{otherwise.} \end{cases}$$

Remark 1.3.1. When the loss function is given by (1.3.2), it is easy to see that D_i can be written as $D_i = \iint[1-\prod_{j\neq k-i} G_j(z)]G_{k-i}(z)dzdH(w|\underline{x})-\Delta$, and results analogous to the above can be obtained in a similar way.

In fact, recently Miescke (1978) has studied this problem with such a loss function using a different approach.

At this point some comments about the case for $k = 2$ are in order because of the special structure of the problem in this case. As stated in Corollary 1.3.1, we can completely specify the Bayes rule in this case. Furthermore, we can specify an essentially complete class in this case provided we make the further assumption that $f(x_i,\theta_i) = f(x_i-\theta_i)$ with $f(\cdot)$ being the density wrt the Lebesgue measure. The loss function given by (1.3.1) can be written as follows.

$L(\underline{\theta},a)$

	$a = \{2\}$	$a = \{1,2\}$	$a = \{1\}$	
$\theta_1 - \theta_2 < -\Delta$	0	c_1	1	
$-\Delta \le \theta_1 - \theta_2 \le \Delta$	$1-c_1$	0	$1-c_1$	$(0 < c_1 < 1)$
$\theta_1 - \theta_2 > \Delta$	1	c_1	0	

With this loss function the problem is invariant under the group of translations as well as under S_k. Therefore, it seems reasonable to consider rules invariant under both. Then the decision rule should depend on $\underline{x}$ only through x_1-x_2, and our problem becomes a montone multiple decision problem (see Ferguson (1967)) with X_1-X_2 having pdf, say, $q(\cdot-(\theta_1-\theta_2))$, given $\underline{\theta} = (\theta_1,\theta_2)$. Note that $q((x_1-x_2)-(\theta_1-\theta_2))$ has the MLR property in x_1-x_2 and $\theta_1-\theta_2$ (see Ibragimov (1956)). Hence Theorem 6.1 of Ferguson (1967) in this case leads to the following result.

Theorem 1.3.2. For $k = 2$, assume that, X_i has pdf $f(x_i-\theta_i)$ $(i = 1,2)$ given $\underline{\theta}$, wrt the Lebesgue measure. Suppose the loss function is given by (1.3.1). Then the following rules R_d form an essentially complete class among the translation and permutation invariant rules:

Rule R_d: Select π_i if and only if $x_i \geq x_{[k]}-d$, $d \geq 0$.

Corollary 1.3.2. Under the assumptions of Theorem 1.3.2, a minimax rule R_{d_m} is given as follows.

c_1-values	d_m-values	minimax risks
$c_1 \geq Q(\Delta)$	0	$1-Q(\Delta)$
$Q(\Delta-d_0) < c_1 < \frac{1}{2}$	d_0	$c_1Q(d_0-\Delta)+(1-c_1)Q(-d_0-\Delta)$
otherwise	$\Delta-Q^{-1}(c_1)$	$(1-c_1)[c_1+Q(-2\Delta+Q^{-1}(c_1))]$

Here, $Q(y) = \int_{-\infty}^{y} q(x)dx$ and d_0 is the value determined by $c_1q(d-\Delta) = c_2q(d+\Delta)$.

Proof. From Theorem 1.3.2 and the generalized Hunt-Stein theorem it suffices to find a d-value which minimizes the supremum of the risk associated with the rules R_d for $d \geq 0$. By invariance under S_k, it is sufficient to consider the case when $\eta = \theta_1-\theta_2 \geq 0$. It is easy to see that

$$E_{\underline{\theta}}[L(\underline{\theta},R_d)] = \begin{cases} c_2[Q(-d-\eta)+Q(-d+\eta)] & \text{if } 0 \leq \eta \leq \Delta \\ c_1Q(d-\eta)+c_2Q(-d-\eta) & \text{if } \Delta < \eta. \end{cases}$$

Furthermore, for $0 \leq \eta \leq \Delta$,

$$\frac{\partial}{\partial \eta} E_{\underline{\theta}}[L(\underline{\theta},R_d)] = c_2[-q(-d-\eta)+q(-d+\eta)]$$

$$= c_2 q(d+\eta)[\frac{q(\eta-d)}{q(\eta+d)} - 1]$$

$$\geq c_2 q(d+\eta)[\frac{q(-d)}{q(\ d)} - 1] = 0 \text{ by the MLR property of } q.$$

Therefore, $\sup_{\underline{\theta}} E_{\underline{\theta}}[L(\underline{\theta},R_d)] = \text{Max}[c_2Q(-d-\Delta)+c_2Q(-d+\Delta),\ c_1Q(d-\Delta)+c_2Q(-d-\Delta)]$.

The first term in the right side is non-increasing in d, and the derivative of the second term wrt d is $c_1q(d-\Delta)-c_2q(-d-\Delta)$ which is non-negative for $d \geq d_0$. Hence the result follows.

Example 1.3.1.

(A) Two normal populations with unknown means and a common known variance:

Suppose X_1 and X_2 are independent normal random variables with means θ_1 and θ_2, respectively, and a common known variance $\gamma^2 > 0$. Then the minimax rule R_{d_m} is given as follows:

c_1-values	d_m-values	minimax risks
$\Phi(\frac{\Delta}{\sqrt{2}\gamma}) \leq c_1 < 1$	0	$1-\Phi(\frac{\Delta}{\sqrt{2}\gamma})$
$\Phi(\frac{\Delta-d_0}{\sqrt{2}\gamma}) < c_1 < \frac{1}{2}$	$d_0 = \gamma^2\Delta^{-1}\log(c_1^{-1}-1)$	$c_1\Phi(\frac{d_0-\Delta}{\sqrt{2}\gamma}) + (1-c_1)\Phi(\frac{-d_0-\Delta}{\sqrt{2}\gamma})$.
otherwise	$\Delta - \sqrt{2}\gamma\ \Phi^{-1}(c_1)$	$(1-c_1)[c_1+\Phi(\Phi^{-1}(c_1) - \frac{\sqrt{2}\Delta}{\gamma})]$

Here, $\Phi(\cdot)$ denotes the cdf of a standard normal distribution.

(B) Two gamma populations with a common shape parameter:

Suppose X_1 and X_2 are independent gamma random variables with a common shape parameter $\alpha(\alpha > 0)$ and unknown scale parameter β_1 and β_2, respectively. A population π_i is then defined to be good if $\beta_i \geq \Delta^{-1} \max_{1 \leq j \leq k} \beta_j$ and bad, otherwise, where $\Delta(\Delta > 1)$ is a positive constant. By considering the associated location parameter problem, we can get the following minimax rule R_{d_m}, which selects π_i if and only if $x_i \geq d_m^{-1} x_{[k]}$, $d_m \geq 1$.

c_1-values	d_m-values	minimax risks
$F(\Delta) < c_1 < 1$	1	$1-F(\Delta)$
$F(\Delta d_0^{-1}) < c_1 < \frac{1}{2}$	d_0	$c_1F(d_0\Delta^{-1})+(1-c_1)F(d_0^{-1}\Delta^{-})$
otherwise	$\Delta/F^{-1}(c_1)$	$(1-c_1)[c_1+F(F^{-1}(c_1)/\Delta^2)]$

Here, $d_0 = ((c_2/c_1)^{\frac{1}{2\alpha}}\Delta-1)/(\Delta-(c_2/c_1)^{\frac{1}{2\alpha}})$ and F is the cdf of the F-distribution with degrees of freedom 2α and 2α.

1.4 Further simplification of the Bayes rule

In this section, we make more assumptions that the posterior cdf of $\theta_{(i)}$ is of the form $G_i(\cdot|x_{[i]},w) = G(\frac{-b_1t_i-b_0w}{b_2})$ where $t_i = t(x_{[i]})$ is an increasing function of $x_{[i]}$, $b_1 > 0$ and $b_2 > 0$.

To simplify the forthcoming formulae we introduce the following notations; for fixed Δ^* and for $i = 1,\dots,k-1$, let

$$\ell_i(y) = \int G^i(y+z)dG(z)$$

$$m_i(y|\Delta^*) = \int G^{k-i-1}(z+\Delta^*)G^i(y+z+\Delta^*)dG(z).$$

It follows from the above specification of the posterior that, for $i = 1,\ldots,k-1$,

$$D_i = c_1 - \int \prod_{j\neq k-i} G(z+\frac{b_1}{b_2}(t_{k-i}-t_j)+\frac{\Delta}{b_2})dG(z),$$

$$u_1(i) = \ell_1[\frac{b_1}{b_2}(t_{k-i}-t_k)+\frac{\Delta}{b_2}]$$

$$u_2(i) = \ell_i[\frac{b_1}{b_2}(t_{k-i}-t_{k-i+1})+\frac{\Delta}{b_2}] \text{ and}$$

$$v_1(i) = m_i[\frac{b_1}{b_2}(t_{k-i}-t_k)|\frac{\Delta}{b_2}].$$

The following well-known results are stated here for completeness for providing bounds on D_i (see Hardy, Littlewood and Pólya (1934), Theorem 43 and Theorem 108).

Lemma 1.4.1. (1) (Tchebycheff) If Z is a random variable, then, for non-decreasing real-valued functions f_1 and f_2, $E[f_1(Z)f_2(Z)] \geq E[f_1(Z)]E[f_2(Z)]$ provided the expectations exist.

(2) (Karamata, Schur) If ϕ is a convex function on the real line, then $\Psi(\underline{y}) = \sum_{i=1}^{p} \phi(y_i)$ is a Schur-convex function of $\underline{y} = (y_1,\ldots,y_p)$.

Lemma 1.4.2. Under the assumptions made in this section, the following relations hold.

$$(1.4.2) \qquad c_1-u_3(i) \leq D_i \leq c_1-v_2(i),\ i = 1,\ldots,k-1$$

where $u_3(i) = \ell_{k-1}[\frac{b_1}{b_2}(t_{k-i}-\frac{1}{k-1}\sum_{j\neq k-i} t_j)+\frac{\Delta}{b_2}]$ and

$$v_2(i) = \prod_{j\neq k-i} \ell_1[\frac{b_1}{b_2}(t_{k-i}-t_j)+\frac{\Delta}{b_2}].$$

Proof. The upper bound on D_i is an immediate consequence of the repeated applications of (1) in Lemma 1.4.1. To obtain the lower bound $c_1-u_3(i)$, it suffices to show that $G(\cdot)$ is log-concave, since the log-concavity of G implies the Schur-concavity of $\prod_{i=1}^{p} G(y_i)$ in $\underset{\sim}{y} = (y_1,\ldots,y_p)$ from (2) in Lemma 1.4.1. Let $g(\theta_i|x_i,w)$ denote the posterior density of θ_i, given $\underline{x}$ and w, then, for $y < y'$ and $x_i < x_i'$,

$$G(y'|x_i',w)G(y|x_i,w)-G(y|x_i',w)G(y'|x_i,w)$$

$$= P(\theta_i \leq y|x_i,w)P(y < \theta_i \leq y'|x_i',w)-P(\theta_i \leq y|x_i',w)P(y < \theta_i \leq y'|x_i,w)$$

$$= \iint_{\substack{s\leq y\\ y<t<y'}} [g(s|x_i,w)g(t|x_i',w)-g(s|x_i',w)g(t|x_i,w)]dsdt$$

≥ 0 from the MLR property of $g(\theta_i|x_i,w)$ in θ_i and x_i for fixed w. The last inequality and the specified form of $G(\cdot|x_i,w)$ imply the log-concavity of G (see, for example, Lehmann (1959) p. 330). Hence the proof is completed.

The next result is an immediate consequence of Lemma 1.4.1 and Theorem 1.3.1.

Theorem 1.4.1. Assume the prior of $\underset{\sim}{\theta}$ is exchangeable and satisfies the specification of the posterior cdf in this section. Then the Bayes rule δ^* for the loss function given in (1.3.1) is given as follows.

(i) $$t_{k-1} \leq \max\{t_k + \frac{b_2}{b_1}(\ell_1^{-1}(c_1)-\frac{\Delta}{b_2}), \frac{1}{k-1}\sum_{j\neq k-1} t_j + \frac{b_2}{b_1}(\ell_{k-1}^{-1}(c_1)-\frac{\Delta}{b_2})\} \Rightarrow \delta^* = \delta_1$$

(ii) Let $i_0 = \min\{i: t_{k-i} \leq r_i,\ i = 2,\ldots,k-1\}$ and

$$j_0 = \max\{j: t_{k-j} > t_k + \frac{b_2}{b_1} m_j^{-1}(c_1|\frac{\Delta}{b_2}) \text{ or}$$

$$\prod_{i \neq k-j} \ell_1(\frac{b_1}{b_2}(t_{k-j} - t_i) + \frac{\Delta}{b_2}) > c_1,\ j = 1,\ldots,i_0-1\} \text{ where}$$

$$r_i = \max\{t_k + \frac{b_2}{b_1}(\ell_1^{-1}(c_1) - \frac{\Delta}{b_2}),\ t_{k-i+1} + \frac{b_2}{b_1}(\ell_i^{-1}(c_1) - \frac{\Delta}{b_2}),$$

$$\frac{1}{k-1} \sum_{j \neq k-i} t_j + \frac{b_2}{b_1}(\ell_{k-1}^{-1}(c_1) - \frac{\Delta}{b_2})\}, \text{ then } \delta^* = \delta_{i^*} \text{ for}$$

$$i^* = \min\{i: D_i \geq 0,\ j_0 + 1 \leq i \leq i_0\}.$$

Note that the above result can be written in terms of rules of the following types, which have been studied in the past.

Rule $\delta^m(d)$: Select $\pi_{(i)}$ iff $t_i = t_{max}$ and/or $t_i > t_{max} - d$,

Rule $\delta^a(d)$: Select $\pi_{(i)}$ iff $t_i = t_{max}$ and/or $t_i > \frac{1}{k-1} \sum_{j \neq i} t_j - d$,

Rule $\delta^0(\underline{d})$: Select $\pi_{(k)}$ and select $\pi_{(k-1)},\ldots,\pi_{(k-i^*+1)}$ where $i^* = \min\{i | t_{k-i} \leq t_{max} - d(i),\ i = 1,\ldots,k-1\}$ and $\underline{d} = (d(1),\ldots,d(k-1))$.

Here, $t_{max} = \max_{1 \leq j \leq k} t_j = t(x_{[k]})$. Rules of the first type were first proposed by Gupta (1956) for the problem of selecting a subset which includes the 'best' population, and later studied by Desu (1970) for selecting a subset consisting of only good populations. Rules δ^a are modified versions of the rules proposed by Seal (1955) and studied by many others for the selection of the 'best' population. The last rule $\delta^0(\underline{d})$ is of the type studied by Broström (1978). Let $d_1 = \frac{\Delta}{b_1} - \frac{b_2}{b_1} \ell_1^{-1}(c_1)$, $d_2 = \frac{\Delta}{b_1} - \frac{b_2}{b_1} \ell_{k-1}^{-1}(c_1)$ and $d(i) = \frac{-b_2}{b_1} m_i^{-1}(c_1|\frac{\Delta}{b_2})$ for

$i = 1,\ldots,k-1$, then $\delta^m(d_1)$, $\delta^a(d_2)$ and $\delta^0(\underline{d})$ are the 'approximate' Bayes rules suggested by Theorem 1.4.1. Since $v_1(i) \geq m_{k-1}(\frac{b_1}{b_2}(t_{k-i}-t_k)|\frac{\Delta}{b_2}) = \ell_{k-1}(\frac{b_1}{b_2}(t_{k-i}-t_k)+\frac{\Delta}{b_2})$, $\delta^m(d_2)$ is a special case of $\delta^0(\underline{d})$. Note that $\delta^m(d_1)$ and $\delta^a(d_2)$ ($\delta^m(d_2)$ and $\delta^0(\underline{d})$) select larger (smaller) subsets than the Bayes rule, and that they all coincide for $k = 2$.

Corollary 1.4.1. For $k = 2$, the Bayes rule δ^* is given by

$$\delta^* = \begin{cases} \delta_1 & \text{if } t_1-t_2+\frac{\Delta}{b_1} \leq \frac{b_2}{b_1}\ell_1^{-1}(c_1) \\ \delta_2 & \text{otherwise.} \end{cases}$$

The next result is helpful in eliminating some unnecessary computations in finding the Bayes rule.

Corollary 1.4.2. If $c_1 \geq \int G^i(z+\frac{\Delta}{b_2})dG(z)$ for some $i = 1,\ldots,k-1$, then the Bayes rule δ^* selects at most i populations.

1.5 Some specific examples

(A) Normal populations

Here we assume that $\pi_1,\ldots,\pi_k$ are normal populations with unknown means $\theta_1,\ldots,\theta_k$ and a common known variance σ^2, and that we have independent samples with a common sample size n for each population. By sufficiency we can reduce the problem to that based on sample means $X_1,\ldots,X_k$. The loss function is assumed to be given by (1.3.1). We will consider a permutationally symmetric prior defined as follows.

Given $W = w$, $(\theta_1,\ldots,\theta_k)$ have a multivariate normal distribution with $E(\theta_i) = w$, $Var(\theta_i) = \beta^2$ and $Cor(\theta_i,\theta_j) = \rho$ where $\rho > -(k-1)^{-1}$ and W has a known distribution $H(\cdot)$. (1.5.1)

A prior of the above type with $\rho \geq 0$ has been used in Goel and Rubin (1977) and in Chernoff and Yahav (1977). The following well known representation is useful for reducing the above prior to a simpler one.

<u>Lemma 1.5.1</u>. If $Y_1,\ldots,Y_k$ are equally correlated normal random variables with $E(Y_i) = 0$, $Var(Y_i) = 1$ and $Cor(Y_i,Y_j) = \rho$, $i,j=1,2,\ldots,k$, $j \neq i$, where $-(k-1)^{-1} < \rho < 1$, then Y_i's can be written as $Y_i = \sqrt{1-\rho}\, Z_i - (\sqrt{\rho^-} + \sqrt{\rho^+})Z_0$ where $(Z_0,\ldots,Z_k)$ have a multivariate normal distribution with $E(Z_i) = 0$, $Var(Z_i) = 1$, $i = 0,1,\ldots,k$, and

$$Cor(Z_i,Z_j) = \begin{cases} \sqrt{\rho^-}/\sqrt{1-\rho} & \text{if } i = 0 < j \leq k \\ 0 & \text{otherwise.} \end{cases}$$

The next result follows from Lemma 1.5.1, the invariance of the loss function and the detailed calculation of the posterior distribution.

<u>Lemma 1.5.2</u>. Let δ be a translation invariant rule, i.e., $\delta(\underline{x}) = \delta(\underline{x}+b\underline{1})$ for any real b where $\underline{1} = (1,\ldots,1)'$, then the overall risk $r(\tau,\delta)$ of the rule δ wrt the prior τ given in (1.5.1) can be written as

$$r(\tau,\delta) = \iint L(\underline{\theta},\delta(\underline{x}))\,dN(\underline{\theta}\,|\,(\sigma_0^{-2}+n\,\sigma^{-2})^{-1} n\,\sigma^{-2}\underline{x},(\sigma_0^{-2}+n\,\sigma^{-2})^{-1}I)\,dN(\underline{x}\,|\,\underline{0},(n^{-1}\sigma^2+\sigma_0^2)I) \quad (1.5.2)$$

where $N(\cdot|\underline{\mu},\Sigma)$ denotes the cdf of a multivariate normal distribution with mean $\underline{\mu}$ and covariance matrix Σ, I is the kxk identity matrix and $\sigma_0^2 = (1-\rho)\beta^2$.

It should be pointed out that similar reduction has been done in Chernoff and Yahav (1977) for the case $\rho \geq 0$, and in Gupta and Hsu (1978). Note that the right side in (1.5.2) is the overall risk of rule δ when $\theta_1,\ldots,\theta_k$ are, a priori, iid normal random variables with mean 0 and variance σ_0^2, and that the Bayes rule wrt the prior given in (1.5.1) can be taken as translation invariant. Hence we can reduce the prior given in (1.5.1) to this iid normal prior. Obviously the posterior cdf of $\theta_{(i)}$ satisfies the specification of the posterior cdf in Section 1.4 with $b_1 = (\sigma_0^{-2}+n\sigma^{-2})^{-1}n\sigma^{-2}$, $b_2 = (\sigma_0^{-2}+n\sigma^{-2})^{-\frac{1}{2}}$, $b_0 = 0$, $t_i = x_{[i]}$ and $G(\cdot) = \Phi(\cdot)$. It follows from the preceeding that, for $i = 1,\ldots,k-1$,

$$\ell_i(y) = \int \Phi^i(y+z)d\Phi(z)$$

$$m_i(y|\Delta^*) = \int \Phi^{k-i-1}(z+\Delta^*)\Phi^i(z+y+\Delta^*)d\Phi(z) \quad \text{and} \tag{1.5.3}$$

$$D_i = c_1 - \int \prod_{j\neq k-i} \Phi(z+n\sigma^{-2}(\sigma_0^{-2}+n\sigma^{-2})^{-\frac{1}{2}}(x_{[k-i]}-x_{[j]})+(\sigma_0^{-2}+n\sigma^{-2})^{\frac{1}{2}}\Delta)\, d\Phi(z).$$

The Bayes rule δ^* can be obtained by numerically integrating D_i using Gauss-Hermite or other methods of quadrature (we eliminate unnecessary computations using Theorem 1.4.1.). Note that $\ell_1(y) = \Phi(y/\sqrt{2})$ and therefore δ^* selects only one population $\pi_{(k)}$ if $c_1 \geq \Phi((\sigma_0^{-2}+n\sigma^{-2})^{\frac{1}{2}}\Delta/\sqrt{2})$.

The 'approximate' Bayes rules suggested in Section 1.4 are determined by $t_i = x_{[i]}$, $d_1 = d_1(n,\sigma_0) = \Delta(1+\frac{\sigma^2}{n}\frac{1}{\sigma_0^2})$ -

$\frac{\sigma}{\sqrt{n}}(1+\frac{\sigma^2}{n}\frac{1}{\sigma_0^2})^{\frac{1}{2}}\sqrt{2}\,\Phi^{-1}(c_1)$, $d_2 = d_2(n,\sigma_0) = \Delta(1+\frac{\sigma^2}{n}\frac{1}{\sigma_0^2}) -$ $\frac{\sigma}{\sqrt{n}}(1+\frac{\sigma^2}{n}\frac{1}{\sigma_0^2})^{\frac{1}{2}}\ell_{k-1}^{-1}(c_1)$ and etc. The Bayes rule wrt the diffuse prior, i.e., $\sigma_0^2 \to \infty$, is very often of interest, and it as well as the corresponding 'approximate' rules can be obtained by formally taking $\sigma_0^{-2} = 0$ or equivalently $\sigma_0^2 = \infty$ in the above. From these expressions we see that $d_1(n,\sigma_0) - d_2(n,\sigma_0) = \frac{\sigma}{\sqrt{n}}(1+\frac{\sigma^2}{n}\frac{1}{\sigma_0^2})^{\frac{1}{2}}$ $(\ell_{k-1}^{-1}(c_1)-\sqrt{2}\,\Phi^{-1}(c_1))$ would be small, especially when $\sigma_0^{-2} = 0$, for sufficiently large n. Therefore, one might expect that $\delta^m(d_1(n,\sigma_0))$ and $\delta^m(d_2(n,\sigma_0))$ would be close to the Bayes rule for sufficiently large n.

Let us consider the rules $\delta^m(h(n,\sigma_0,d))$ for various values of d where $h(n,\sigma_0,d) = \Delta(1+\frac{\sigma^2}{n}\frac{1}{\sigma_0^2}) - \frac{\sigma}{\sqrt{n}}d(1+\frac{\sigma^2}{n}\frac{1}{\sigma_0^2})^{\frac{1}{2}}$. Let $(\pi_{(i)}\in\delta)$ denote an event $(\pi_{(i)}\in\{\pi_j : j \in \delta(\underline{x})\})$ for any rule δ, and $r(\sigma_0,\delta)$ denote the overall risk of δ when the prior is given by (1.5.1). Since $(\pi_{(k-i)}\notin \delta^m(d_1(n,\sigma_0))) \subset (D_i \geq 0)$ and $\delta^m(d_1(n,\sigma_0)) \subset \delta^m(h(n,\sigma_0,d))$ for $d < \sqrt{2}\,\Phi^{-1}(c_1)$,

$$r(\sigma_0,\delta^m(h(n,\sigma_0,d))) - r(\sigma_0,\delta^m(d_1(n,\sigma_0)))$$

$$= E[\sum_{i=1}^{k-1} D_i\{I_{(\pi_{(k-i)}\in\delta^m(h(n,\sigma_0,d)))} - I_{(\pi_{(k-i)}\in\delta^m(d_1(n,\delta_0)))}\}]$$

$$= E[\sum_{i=1}^{k-1} D_i\{I_{(\pi_{(k-i)}\in\delta^m(h(n,\sigma_0,d)),\pi_{(k-i)}\notin\delta^m(d_1(n,\sigma_0)))}\}]$$

$$\geq 0,$$

where D_i is defined in (1.5.3) and the expectations are taken wrt the marginal distribution of X's. Similarly, it can be shown that $r(\sigma_0,\delta^m(h(n,\sigma_0,d))) \geq r(\sigma_0,\delta^m(d_2(n,\sigma_0)))$ for $d > \ell_{k-1}^{-1}(c_1)$. Therefore, we may consider the rules $\delta^m(h(n,\sigma_0,d))$ only for $d \in [\sqrt{2}\,\Phi^{-1}(c_1), \ell_{k-1}^{-1}(c_1)]$ as long as the overall risk is concerned. Furthermore, denoting the Bayes rule for sample size n by δ_n^* and $\delta^m(h(n,\sigma_0,d))$ by δ^m for fixed $d \in [\sqrt{2}\,\Phi^{-1}(c_1),\ell_{k-1}^{-1}(c_1)]$, it is easy to see that $r(\sigma_0,\delta^m)-r(\sigma_0,\delta_n^*)$

$$= E[\sum_{i=1}^{k-1} D_i\{I_{(\pi_{(k-i)}\notin\delta_n^*,\pi_{(k-i)}\in\delta^m)} - I_{(\pi_{(k-i)}\in\delta_n^*,\pi_{(k-i)}\notin\delta^m)}\}]$$

$$\leq E[\sum_{i=1}^{k-1}(c_1-\ell_{k-1}(d))I_{(\pi_{(k-i)}\notin\delta_n^*,\pi_{(k-i)}\in\delta^m)} + (\Phi(d/\sqrt{2})-c_1)I_{(\pi_{(k-i)}\in\delta_n^*,\pi_{(k-i)}\notin\delta^m)}]$$

$$\leq E[\sum_{i=1}^{k-1}(\Phi(d/\sqrt{2})-\ell_{k-1}(d))I_{(\pi_{(k-i)}\notin\delta^m(d_2(n,\sigma_0)),\pi_{(k-i)}\in\delta^m(d_1(n,\sigma_0)))}]$$

$$= [\Phi(d/\sqrt{2})-\ell_{k-1}(d)]\sum_{i=1}^{k-1}\{F_i[\frac{\sigma}{\sqrt{n}}\frac{1}{\sigma_0}\ell_{k-1}^{-1}(c_1)-(1+\frac{\sigma^2}{n}\frac{1}{\sigma_0^2})^{\frac{1}{2}}\frac{\Delta}{\sigma_0}] - F_i[\frac{\sigma}{\sqrt{n}}\frac{\sqrt{2}}{\sigma_0}\Phi^{-1}(c_1)-(1+\frac{\sigma^2}{n}\frac{1}{\sigma_0^2})^{\frac{1}{2}}\frac{\Delta}{\sigma_0}]\}$$

$$\text{where } F_i(y) = \begin{cases} \frac{k!}{(i-1)!(k-i-1)!}\int_{-\infty}^{y}[\int_{-\infty}^{\infty}\Phi^{i-1}(u+v)\varphi(u+v)[\Phi(v)-\Phi(u+v)]^{k-i-1}\varphi(v)dv]du & \text{if } y \leq 0 \\ 1 & \text{if } y > 0, \end{cases}$$

and $\varphi(\cdot)$ denotes the pdf of the standard normal distribution. The first inequality follows from the fact that

$$(\pi_{(k-i)} \notin \delta_n^*,\ \pi_{(k-i)} \in \delta^m) \subset (D_i \geq 0,\ x_{[k-i]} > x_{[k]} - h(n,\sigma_0,d))$$
$$\subset (0 \leq D_i < c_1 - \ell_{k-1}(d)) \text{ and}$$
$$(\pi_{(k-i)} \in \delta_n^*,\ \pi_{(k-i)} \notin \delta^m) \subset (c_1 - \ell_1(d) \leq D_i < 0).$$

The second inequality is obvious from the fact that $\delta^m(d_2(n,\sigma_0)) \subset \delta_n^* \subset \delta^m(d_1(n,\sigma_0))$ and $\delta^m(d_2(n,\sigma_0)) \subset \delta^m \subset \delta^m(d_1(n,\sigma_0))$, and the last equality follows from the marginal distributions of X_i's. It follows from the preceeding that $\lim_{n\to\infty} n^\alpha[r(\sigma_0,\delta^m) - r(\sigma_0,\delta_n^*)] = 0$ for $\alpha \in [0, \frac{1}{2})$. Hence we have proved the next result.

<u>Theorem 1.5.1</u>. Let $r(\sigma_0,\delta)$ denote the overall risk of a rule δ wrt the prior distribution given in (1.5.1) where $\sigma_0^2 = \beta^2(1-\rho)$, then

$$\text{(i)}\quad r(\sigma_0,\delta^m(h(n,\sigma_0,d))) \begin{cases} \geq r(\sigma_0,\delta^m(d_1(n,\sigma_0))) & \text{if } d < \sqrt{2}\,\Phi^{-1}(c_1) \\ \geq r(\sigma_0,\delta^m(d_2(n,\sigma_0))) & \text{if } d > \ell_{k-1}^{-1}(c_1) \end{cases}$$

and

(ii) for $d \in [\sqrt{2}\,\Phi^{-1}(c_1), \ell_{k-1}^{-1}(c_1)]$, $\lim_{n\to\infty} n^\alpha[r(\sigma_0,\delta^m(h(n,\sigma_0,d))) - r(\sigma_0, \delta_n^*)] = 0$

for any $\alpha \in [0, \frac{1}{2})$.

It is interesting to note that, for k = 2, the Bayes rule wrt the diffuse prior, which selects π_i iff $x_i = x_{max}$ and/or $x_i > x_{max} - \Delta + \frac{\sigma}{\sqrt{n}}\sqrt{2}\,\Phi^{-1}(c_1)$, coincides with the minimax rule in some cases (see Example 1.3.1), and with the rule studied by Desu (1970) if $c_1 = P^*$. Also for $k > 2$, consider the rules $\delta^m(h(n,\infty,d))$ for various values of d. Before we state some properties of these rules we recall the following definition (see Ferguson (1967)).

Definition 1.5.1. A decision rule δ_0 is called an extended Bayes rule if, for every $\epsilon > 0$, there exists a prior τ such that $r(\tau,\delta_0) < \inf_{\delta} r(\tau,\delta) + \epsilon$.

It can be easily shown that any extended Bayes rule δ_0 is ϵ-admissible for any $\epsilon > 0$, i.e., there does not exist a rule δ for which $R(\underline{\theta},\delta) < R(\underline{\theta},\delta_0)-\epsilon$ for all $\underline{\theta} \in \Theta^k$, where $R(\underline{\theta},\delta) = E_{\underline{\theta}}L(\underline{\theta},\delta(\underline{x}))$. In a manner similar to the one in the proof of Theorem 1.5.1 we can show that $\lim_{\sigma_0\to\infty} [r(\sigma_0,\delta^m(h(n,\infty,d))-r(\sigma_0,\delta_n^*)]=0$ for $d \in [\sqrt{2}\,\Phi^{-1}(c_1),\ \ell_{k-1}^{-1}(c_1)]$; therefore, we have the following result.

Theorem 1.5.2. The decision rules $\delta^m(h(n,\infty,d))$ for $d \in [\sqrt{2}\,\Phi^{-1}(c_1),$ $\ell_{k-1}^{-1}(c_1)]$, which selects π_i iff $x_i = x_{max}$ and/or $x_i > x_{max}-\Delta+ \frac{\sigma}{\sqrt{n}}\, d$, are extended Bayes and therefore ϵ-admissible for any $\epsilon > 0$.

The above arguments indicate that the performance of rules of the type δ^m would be close to that of the Bayes rule when σ_0 is large, but we could not make similar arguments for the other 'approximate' rules. Hence we carried out Monte Carlo study to see how well these rules (studied in the past under a different framework) perform compared with the Bayes rule. The results of the Monte Carlo study are given in the next section.

(B) Gamma populations

Here, we consider a problem of selecting good populations out of k gamma populations in terms of unknown scale parameters based on nk independent observations, assuming a common known shape parameter. By sufficiency we reduce the data to the k independent gamma random variables $X_1,\ldots,X_k$ with a common known shape parameter α ($\alpha > 0$) and unknown scale parameters $\beta_1,\ldots,\beta_k$, respectively. Population π_i is said to be good if $\beta_i \geq \Delta^{-1} \max_{1\leq j\leq k} \beta_j$, and bad, otherwise. Here Δ is a preassigned constant greater than 1.

We also assume that the loss structure is the same as that in (1.3.1), i.e., the loss is c_1 for selecting a bad population and c_2 if we exclude a good one. Further it is assumed that $\beta_1,\ldots,\beta_k$ are, a priori, independently distributed inverse gamma random variables, i.e., the prior pdf of $\underline{\beta} = (\beta_1,\ldots,\beta_k)$ is

$$\tau(\underline{\beta}) = \prod_{i=1}^{k} \left[\frac{b^a}{\Gamma(a)\beta_i^{a+1}} e^{-b\beta_i^{-1}}\right], \ \beta_i > 0, \text{ for } i = 1,\ldots,k \tag{1.5.4}$$

where $a > 0$ and $b > 0$ are known. Then it is easily observable that, given $\underline{x} = (x_1,\ldots,x_k)$, $\beta_1,\ldots,\beta_k$ are, a posteriori, independent and have the same distribution as that in (1.5.4) except that a and b are replaced by $a + \alpha$ and $b + x_i$, respectively.

It can be easily shown that, in the associated location parameter problems, all the assumptions in Section 1.4 are satisfied. Hence we have the results analogous to those in Section 1.4 with the following modifications. For $i = 1,\ldots k-1$, let $\ell_i(y) = \int_0^\infty [1-G(yz)]^i dG(z)$, and $m_i(y|\Delta^{-1}) = \int_0^\infty [1-G(yz)]^i [1-G(\Delta^{-1}z)]^{k-i-1} dG(z)$

where $G(\cdot)$ is the cdf of gamma distribution with shape parameter $a + \alpha$ and scale parameter 1. Then, for $i = 1,\ldots,k-1$,

$$
\begin{aligned}
D_i &= c_1 - \int_0^\infty \prod_{j \neq k-i} [1-G(t_j t_{k-i}^{-1} \Delta^{-1} y)] dG(y) \\
u_1(i) &= \ell_1(t_k \, t_{k-i}^{-1} \Delta^{-1}) \\
u_2(i) &= \ell_i(t_{k-i+1} \, t_{k-i}^{-1} \Delta^{-1}) \\
u_3(i) &= \ell_{k-1}(t_{k-i}^{-1} \Delta^{-1} (\prod_{j \neq k-i} t_j)^{\frac{1}{k-1}}) \qquad (1.5.5) \\
v_1(i) &= m_i(t_k \, t_{k-i}^{-1} \Delta^{-1} | \Delta^{-1}) \text{ and} \\
v_2(i) &= \prod_{j \neq k-i} \ell_1(t_j \, t_{k-i}^{-1} \Delta^{-1}), \text{ where } t_i = x_{[i]} + b.
\end{aligned}
$$

Note that $\ell_1(y) = 1-F(y)$ where $F(\cdot)$ denotes the cdf of an F random variable with degrees of freedom $2(a+\alpha)$ and $2(a+\alpha)$. In addition to the above bounds on D_i, we provide another bound in the following, using the fact that $1-G(\cdot)$ is log-concave (log-convex) if $a + \alpha \geq 1$ ($a + \alpha < 1$, respectively).

$$
D_i \begin{cases} \geq c_1 - \ell_{k-1}(\frac{1}{k-1} \sum_{j \neq k-i} t_j t_{k-i}^{-1} \Delta^{-1}) & \text{if } a + \alpha \geq 1 \\ \leq c_1 - \ell_{k-1}(\frac{1}{k-1} \sum_{j \neq k-i} t_j t_{k-i}^{-1} \Delta^{-1}) & \text{if } a + \alpha < 1. \end{cases} \qquad (1.5.6)
$$

Therefore a result analogous to Theorem 1.4.1 can be obtained with obvious modifications from (1.5.5) and (1.5.6). In this case also, the Bayes rule can be found by numerically integrating D_i using Gauss-Laguerre quadrature while we eliminate unnecessary computations using the result analogous to Theorem 1.4.1. Note that, for $k = 2$, the Bayes rule wrt the diffuse prior, i.e., $a \to 0$ and $b \to 0$, coincides with the minimax rule in some cases (see Example 1.3.1.).

1.6 Results of the Monte Carlo study for the normal populations.

In this section we are assuming the normal model in Section 1.5 (A). In the preceeding sections we have seen that rules studied in the past help to find the Bayes rule and that they are, in some sense, natural approximations of the Bayes rule. Among them δ^m and δ^a are perhaps the best well-known selection rules. Hence it would be worthwhile to investigate the performance of these rules in this Bayesian framework, since they have their own merits and are also easy to use. Some optimalities of these rules can be found in the literature by Gupta and Studden (1966), Berger (1977), Gupta and Miescke (1978), Bjørnstad (1978), Chernoff and Yahav (1977), Gupta and Hsu (1978) among others. Especially the last two papers are much related to our work in that they studied the performance of these rules in Bayesian framework for the problem of selecting a subset containing the 'best' population.

For our Monte Carlo study we may assume that $\sigma/\sqrt{n} = 1$ without any loss of generality. We recall that δ^m and δ^a can be written in the following more familiar forms:

$$\begin{aligned} &\delta^m(d): \text{ Select } \pi_i \text{ if and only if } x_i \geq x_{max}-d,\ d \geq 0 \\ &\delta^a(d): \text{ Select } \pi_i \text{ if and only if } x_i = x_{max} \text{ and/or } x_i \geq \frac{1}{k-1}\sum_{j\neq i} x_j - d. \end{aligned} \tag{1.6.1}$$

We carried out the Monte Carlo study for the cases $k = 3$ and $k = 9$. The remaining relevant parameters in this study are c_1 (or c_2), σ_0^2 and Δ. We use $c = c_2/c_1$ for the tabulation purpose since c, being the ratio of two different types of losses, seems more appealing than c_1. The ranges of the parameter values which were studied are as follows.

$$\begin{cases} \Delta = .25,\ .5,\ 1.0,\ \sigma_0 = (1.5)^i (i = -2(1)6), \\ c = 2^{i-1}\ (i = 1,\dots,5) \text{ for both } \Delta = .25 \text{ and } \Delta = .5 \\ c = 2^{i-2}\ (i = 1,\dots,6) \text{ for } \Delta = 1.0. \end{cases}$$

For each of parameter sets (c, σ_0, Δ), 400 simulations were carried out for $k = 3$ and 100 simulations were performed for $k = 9$. In each simulation the generation of the random vector $\underline{x} = (x_1,\dots,x_k)$ according to its marginal distribution was involved, and then the Bayes actions and the corresponding risks were obtained by numerically integrating D_i's in (1.5.3). The optimal values of d in δ^m and δ^a are estimated by minimizing the average regrets corresponding to sufficiently fine grids of the estimated constants d, where the range of these trial values are determined from the preliminary computations and Theorem 1.5.1.

The estimated Bayes risks, the estimated regrets incurred by the optimal δ^m and δ^a are given in Table I at the end of this chapter along with sample standard deviations of these estimates. For those obvious cases when the Bayes rule selects only one population or all the populations, we did not tabulate and, as a result, these cells are left blank in the table. Table II gives the average number of bad populations selected and that of the good ones excluded for the rules considered, along with proportions of times that the optimal δ^a and the optimal δ^m coincide with the Bayes rule. From Table I, we can observe that the performance of the rule δ^m is almost as good as that of the Bayes rule throughout the cases studied, and that it becomes remarkably better as the prior variance ($\sigma_0^2 = \beta^2(1-\rho)$) becomes large. This agrees with the argument in Section 1.5. Also we observe that, for k = 3, rule δ^a performs reasonably well when the prior is concentrated, and in fact it performs better than δ^m in a few extreme cases when c is very large. However, the performance of δ^a is poor for moderately large σ_0. Especially, for k = 9, the performance of δ^a is disastrous and it was observed that, for most values of σ_0, optimal δ^a selects only one population for small values of c and it tends to select much larger subsets than the Bayes rule as soon as c becomes large (roughly $c > 4$). Overall, the rule δ^a performs rather poorly when k = 9.

Similar behavior of the rule δ^m has been observed in Chernoff and Yahav (1977), and in Gupta and Hsu (1978) for the problem of selecting a subset containing the 'best' population. On the otherhand performance of δ^a is worse than that observed in Gupta and Hsu (1978), and it seems that rule δ^a has little to recommend for the goal of selecting 'good' populations while rule δ^m performs almost as well as the Bayes rule provided the value of d is chosen properly. This indicates that the proper use of rule δ^m can lead to efficient statistical method since it behaves fairly well in various formulations and also is easy to use and interpret. However, we point out that the choice of d should depend on the loss structure of the particular problem at hand, and we suggest the tabulation of the operating characteristics such as the number of bad populations selected or that of excluded good ones before setting P* based on an intuitive feeling, if one wants to use this rule. For this reason the estimated optimal d-values for rules δ^m have been provided in Table III.

TABLE I BAYES RISKS AND REGRETS

THE ENTRY ON TOP OF EACH BOX IS THE ESTIMATED BAYES RISK AND THE NUMBERS IN THE SECOND AND THIRD ROW ARE THE REGRETS INCURRED BY THE OPTIMAL δ^m AND THE OPTIMAL δ^a IN THAT ORDER. THE NUMBERS IN THE PARENTHESES ARE THE SAMPLE STANDARD DEVIATIONS OF THE ESTIMATES.

K=3, Δ =.25

σ_0 \ C	1	2	4	8	16
.44	.5516 (.0050) .0046 (.0006) .0185 (.0013)	.4612 (.0024) .0028 (.0005) .0131 (.0013)	.2970 (.0006) .0002 (.0001) .0002 (.0001)	.1659 (.0003) .0000 (.0000) .0000 (.0000)	
.67	.4801 (.0061) .0039 (.0006) .0023 (.0004)	.4337 (.0045) .0049 (.0006) .0087 (.0010)	.3049 (.0020) .0016 (.0004) .0015 (.0004)	.1807 (.0006) .0006 (.0002) .0006 (.0002)	.0971 (.0002) .0001 (.0001) .0000 (.0000)
1.00	.3657 (.0081) .0025 (.0001) .0028 (.0006)	.3442 (.0066) .0041 (.0006) .0087 (.0013)	.2712 (.0038) .0020 (.0004) .0027 (.0005)	.1728 (.0017) .0014 (.0003) .0014 (.0003)	.0964 (.0008) .0005 (.0001) .0008 (.0002)
1.50	.2675 (.0083) .0014 (.0004) .0026 (.0006)	.2651 (.0079) .0031 (.0005) .0197 (.0023)	.2028 (.0046) .0020 (.0004) .0080 (.0013)	.1340 (.0028) .0014 (.0003) .0037 (.0007)	.0827 (.0014) .0009 (.0002) .0019 (.0004)
2.25	.2087 (.0090) .0009 (.0002) .0024 (.0005)	.1883 (.0073) .0007 (.0002) .0250 (.0028)	.1414 (.0053) .0013 (.0003) .0178 (.0022)	.0999 (.0032) .0010 (.0002) .0093 (.0012)	.0592 (.0017) .0004 (.0001) .0050 (.0007)
3.38	.1254 (.0078) .0004 (.0002) .0011 (.0003)	.1280 (.0070) .0005 (.0002) .0222 (.0028)	.0969 (.0048) .0003 (.0002) .0337 (.0035)	.0699 (.0031) .0008 (.0002) .0199 (.0018)	.0380 (.0018) .0004 (.0001) .0104 (.0010)
5.06	.0805 (.0065) .0000 (.0000) .0011 (.0004)	.0923 (.0062) .0005 (.0003) .0142 (.0022)	.0680 (.0046) .0009 (.0003) .0370 (.0043)	.0498 (.0028) .0002 (.0001) .0245 (.0021)	.0245 (.0016) .0000 (.0000) .0125 (.0011)
7.59	.0645 (.0064) .0000 (.0000) .0016 (.0004)	.0534 (.0055) .0001 (.0001) .0104 (.0020)	.0447 (.0039) .0003 (.0002) .0229 (.0035)	.0317 (.0024) .0000 (.0000) .0315 (.0032)	.0210 (.0014) .0000 (.0000) .0162 (.0013)
11.39	.0417 (.0050) .0000 (.0000) .0516 (.0002)	.0437 (.0047) .0000 (.0000) .0110 (.0021)	.0239 (.0027) .0000 (.0000) .0126 (.0027)	.0227 (.0021) .0000 (.0000) .0264 (.0039)	.0090 (.0010) .0000 (.0000) .0185 (.0014)

K=9, Δ =.25

σ_0 \ C	1	2	4	8	16
.44		1.2362 (.0131) .0109 (.0023) .0287 (.0043)	1.1135 (.0090) .0182 (.0034) .2674 (.0099)	.7255 (.0036) .0083 (.0013) .0426 (.0041)	.4019 (.0010) .0016 (.0005) .0047 (.0011)
.67		.9353 (.0157) .0117 (.0033) .0201 (.0038)	.8451 (.0169) .0203 (.0031) .1850 (.0085)	.6359 (.0081) .0153 (.0023) .1785 (.0088)	.3854 (.0038) .0104 (.0023) .0459 (.0042)
1.00		.7238 (.0213) .0124 (.0029) .0201 (.0038)	.6007 (.0168) .0147 (.0028) .1235 (.0104)	.4881 (.0125) .0136 (.0024) .2617 (.0193)	.2947 (.0058) .0085 (.0014) .1171 (.0063)
1.50		.4943 (.0233) .0069 (.0025) .0227 (.0041)	.4536 (.0181) .0117 (.0025) .0627 (.0075)	.3489 (.0125) .0091 (.0019) .0804 (.0105)	.2221 (.0074) .0075 (.0018) .0510 (.0066)
2.25		.3370 (.0199) .0070 (.0019) .0222 (.0043)	.2939 (.0190) .0095 (.0029) .0662 (.0101)	.2095 (.0113) .0030 (.0011) .0572 (.0105)	.1484 (.0075) .0059 (.0013) .0419 (.0054)
3.38		.2480 (.0208) .0030 (.0011) .0189 (.0043)	.1814 (.0161) .0028 (.0010) .0648 (.0101)	.1352 (.0103) .0029 (.0009) .0708 (.0090)	.0853 (.0062) .0020 (.0007) .0478 (.0081)
5.06		.1416 (.0171) .0024 (.0012) .0186 (.0047)	.1148 (.0120) .0012 (.0008) .0527 (.0097)	.0964 (.0092) .0019 (.0008) .0609 (.0102)	.0700 (.0057) .0024 (.0007) .0523 (.0087)
7.59		.0942 (.0131) .0005 (.0004) .0098 (.0038)	.0825 (.0106) .0011 (.0006) .0384 (.0084)	.0529 (.0067) .0009 (.0006) .0444 (.0091)	.0296 (.0034) .0001 (.0001) .0340 (.0049)
11.39		.0599 (.0111) .0000 (.0000) .0151 (.0046)	.0575 (.0093) .0000 (.0000) .0372 (.0086)	.0393 (.0056) .0001 (.0001) .0536 (.0081)	.0170 (.0027) .0000 (.0000) .0234 (.0042)

TABLE I (CONTINUED)

K=3, Δ =.50

σ_0 \ c	1	2	4	8	16
	.4609 (.0016)	.3297 (.0012)	.1995 (.0008)		
.44	.0024 (.0005)	.0006 (.0002)	.0000 (.0000)		
	.0334 (.0029)	.0033 (.0008)	.0001 (.0001)		
	.4686 (.0053)	.3897 (.0026)	.2558 (.0011)	.1456 (.0001)	.0773 (.0003)
.67	.0038 (.0006)	.0029 (.0005)	.0010 (.0003)	.0001 (.0001)	.0000 (.0000)
	.0487 (.0028)	.0185 (.0017)	.0024 (.0006)	.0000 (.0000)	.0000 (.0000)
	.3872 (.0076)	.3424 (.0056)	.2511 (.0028)	.1532 (.0014)	.0869 (.0006)
1.00	.0036 (.0006)	.0027 (.0005)	.0022 (.0004)	.0011 (.0003)	.0006 (.0001)
	.0175 (.0018)	.0160 (.0016)	.0053 (.0008)	.0016 (.0004)	.0006 (.0002)
	.2813 (.0088)	.2710 (.0068)	.2005 (.0045)	.1361 (.0024)	.0803 (.0011)
1.50	.0027 (.0006)	.0023 (.0005)	.0020 (.0004)	.0009 (.0002)	.0008 (.0002)
	.0115 (.0015)	.0180 (.0024)	.0078 (.0013)	.0033 (.0006)	.0014 (.0003)
	.2033 (.0083)	.1833 (.0072)	.1561 (.0049)	.0987 (.0030)	.0590 (.0016)
2.25	.0010 (.0005)	.0010 (.0004)	.0011 (.0003)	.0009 (.0003)	.0004 (.0001)
	.0103 (.0014)	.0318 (.0034)	.0148 (.0020)	.0082 (.0011)	.0048 (.0006)
	.1470 (.0083)	.1274 (.0065)	.1081 (.0050)	.0662 (.0028)	.0439 (.0016)
3.38	.0006 (.0002)	.0004 (.0002)	.0007 (.0003)	.0005 (.0002)	.0004 (.0001)
	.0084 (.0013)	.0343 (.0038)	.0318 (.0030)	.0159 (.0016)	.0079 (.0009)
	.0984 (.0072)	.1207 (.0060)	.0726 (.0044)	.0457 (.0025)	.0310 (.0016)
5.06	.0000 (.0000)	.0008 (.0004)	.0000 (.0000)	.0002 (.0001)	.0001 (.0000)
	.0045 (.0009)	.0241 (.0032)	.0379 (.0039)	.0203 (.0020)	.0122 (.0011)
	.0663 (.0061)	.0564 (.0050)	.0462 (.0038)	.0298 (.0021)	.0179 (.0013)
7.59	.0003 (.0001)	.0002 (.0001)	.0002 (.0001)	.0000 (.0000)	.0002 (.0001)
	.0045 (.0010)	.0184 (.0030)	.0299 (.0041)	.0277 (.0032)	.0150 (.0012)
	.0335 (.0043)	.0392 (.0043)	.0287 (.0030)	.0200 (.0019)	.0118 (.0011)
11.39	.0000 (.0000)	.0000 (.0000)	.0000 (.0000)	.0001 (.0001)	.0000 (.0000)
	.0026 (.0013)	.0109 (.0022)	.0243 (.0041)	.0300 (.0038)	.0215 (.0014)

K=9, Δ =.50

σ_0 \ c	1	2	4	8	16
	1.5947 (.0148)	1.5322 (.0119)	1.0429 (.0037)	.5953 (.0016)	
.44	.0198 (.0033)	.0296 (.0046)	.0070 (.0018)	.0011 (.0001)	
	.0603 (.0052)	.2614 (.0170)	.0306 (.0049)	.0017 (.0006)	
	1.1473 (.0193)	1.2235 (.0173)	.9849 (.0118)	.6407 (.0051)	.3672 (.0018)
.67	.0083 (.0021)	.0223 (.0041)	.0222 (.0031)	.0112 (.0023)	.0042 (.0012)
	.0083 (.0021)	.2021 (.0099)	.3029 (.0146)	.0735 (.0065)	.0110 (.0020)
	.7783 (.0234)	.8272 (.0240)	.7342 (.0179)	.5097 (.0106)	.3219 (.0055)
1.00	.0072 (.0028)	.0238 (.0039)	.0191 (.0035)	.0138 (.0026)	.0088 (.0019)
	.0108 (.0029)	.1071 (.0092)	.3424 (.0164)	.2757 (.0126)	.0931 (.0062)
	.5637 (.0237)	.6126 (.0261)	.5207 (.0194)	.3625 (.0118)	.2497 (.0066)
1.50	.0064 (.0021)	.0186 (.0041)	.0107 (.0024)	.0101 (.0023)	.0095 (.0016)
	.0096 (.0028)	.0654 (.0078)	.2469 (.0148)	.4252 (.0230)	.1834 (.0073)
	.4021 (.0249)	.3444 (.0202)	.3389 (.0181)	.2324 (.0124)	.1500 (.0076)
2.25	.0043 (.0018)	.0051 (.0024)	.0062 (.0019)	.0067 (.0016)	.0041 (.0009)
	.0054 (.0016)	.0443 (.0068)	.1223 (.0141)	.1368 (.0172)	.1280 (.0156)
	.2420 (.0209)	.2662 (.0211)	.1925 (.0154)	.1559 (.0120)	.0954 (.0067)
3.38	.0015 (.0010)	.0032 (.0010)	.0032 (.0013)	.0019 (.0007)	.0023 (.0007)
	.0053 (.0018)	.0424 (.0081)	.0815 (.0130)	.1092 (.0136)	.0771 (.0163)
	.1668 (.0191)	.1284 (.0150)	.1590 (.0152)	.0924 (.0093)	.0723 (.0056)
5.06	.0004 (.0003)	.0005 (.0004)	.0032 (.0012)	.0017 (.0008)	.0008 (.0004)
	.0128 (.0029)	.0304 (.0071)	.0683 (.0109)	.0880 (.0134)	.0727 (.0114)
	.1114 (.0153)	.0972 (.0134)	.1101 (.0123)	.0583 (.0071)	.0333 (.0036)
7.59	.0001 (.0001)	.0012 (.0009)	.0009 (.0005)	.0003 (.0003)	.0000 (.0000)
	.0041 (.0021)	.0357 (.0078)	.0852 (.0143)	.0783 (.0121)	.0382 (.0074)
	.0622 (.0121)	.0754 (.0115)	.0698 (.0113)	.0421 (.0064)	.0253 (.0035)
11.39	.0001 (.0001)	.0000 (.0000)	.0014 (.0007)	.0000 (.0000)	.0002 (.0002)
	.0059 (.0026)	.0236 (.0069)	.0441 (.0099)	.0629 (.0107)	.0491 (.0079)

TABLE I (CONTINUED)

K=3, Δ=1.0

σ_0 \ c	.5	1	2	4	8	16
.44	.1933(.0022) .0000(.0000) .0012(.0004)	.1462(.0018) .0000(.0000) .0001(.0001)				
.67	.3610(.0026) .0028(.0005) .0908(.0060)	.3213(.0027) .0016(.0003) .0315(.0034)	.2276(.0023) .0011(.0003) .0035(.0009)	.1435(.0017) .0000(.0000) .0001(.0001)		
1.00	.3460(.0054) .0028(.0005) .0385(.0024)	.3515(.0046) .0018(.0004) .0774(.0044)	.2969(.0032) .0023(.0005) .0293(.0028)	.1989(.0019) .0011(.0003) .0087(.0012)	.1158(.0012) .0002(.0001) .0013(.0005)	.0625(.0007) .0002(.0001) .0002(.0001)
1.50	.2750(.0072) .0012(.0003) .0154(.0015)	.2948(.0066) .0022(.0005) .0385(.0034)	.2613(.0052) .0030(.0006) .0228(.0024)	.1922(.0032) .0022(.0005) .0135(.0016)	.1227(.0018) .0008(.0002) .0056(.0008)	.0706(.0009) .0002(.0001) .0018(.0004)
2.25	.1832(.0074) .0005(.0003) .0070(.0011)	.2117(.0075) .0009(.0003) .0389(.0041)	.1922(.0061) .0009(.0003) .0272(.0032)	.1488(.0042) .0011(.0003) .0150(.0019)	.1015(.0025) .0009(.0002) .0067(.0010)	.0587(.0013) .0005(.0002) .0031(.0006)
3.38	.1202(.0063) .0003(.0001) .0045(.0009)	.1438(.0069) .0005(.0002) .0296(.0033)	.1344(.0060) .0008(.0003) .0435(.0046)	.1049(.0040) .0009(.0003) .0198(.0024)	.0661(.0026) .0004(.0002) .0141(.0015)	.0424(.0015) .0004(.0001) .0074(.0009)
5.06	.0841(.0059) .0000(.0000) .0023(.0006)	.0997(.0067) .0005(.0003) .0231(.0030)	.0888(.0055) .0005(.0003) .0478(.0052)	.0682(.0039) .0004(.0002) .0385(.0036)	.0482(.0025) .0003(.0001) .0222(.0020)	.0275(.0013) .0001(.0000) .0097(.0010)
7.59	.0536(.0051) .0000(.0000) .0018(.0005)	.0625(.0052) .0001(.0001) .0182(.0028)	.0608(.0046) .0002(.0001) .0384(.0047)	.0476(.0033) .0000(.0000) .0366(.0043)	.0335(.0022) .0001(.0001) .0251(.0022)	.0171(.0012) .0000(.0000) .0125(.0012)
11.39	.0361(.0042) .0000(.0000) .0011(.0004)	.0376(.0041) .0000(.0000) .0086(.0019)	.0416(.0042) .0001(.0001) .0329(.0048)	.0323(.0029) .0001(.0001) .0438(.0059)	.0207(.0019) .0001(.0001) .0358(.0025)	.0138(.0011) .0001(.0001) .0186(.0013)

K=9, Δ=1.0

σ_0 \ c	.5	1	2	4	8	16
.44	1.2743(.0069) .0203(.0044) .1686(.0144)	1.0626(.0097) .0073(.0022) .0175(.0048)	.7235(.0076) .0000(.0000) .0005(.0005)			
.67	1.3251(.0184) .0168(.0030) .1037(.0073)	1.4650(.0147) .0256(.0052) .6476(.0328)	1.2372(.0097) .0284(.0062) .1879(.0180)	.8253(.0065) .0053(.0013) .0242(.0050)	.4712(.0045) .0020(.0009) .0041(.0015)	
1.00	.8967(.0251) .0131(.0026) .0193(.0030)	1.0771(.0300) .0212(.0038) .1833(.0116)	1.0578(.0200) .0283(.0050) .5566(.0264)	.8267(.0119) .0245(.0045) .3253(.0190)	.5314(.0063) .0122(.0025) .0991(.0085)	.3081(.0029) .0053(.0009) .0294(.0035)
1.50	.6065(.0210) .0043(.0014) .0053(.0017)	.7150(.0280) .0178(.0042) .1088(.0095)	.7205(.0264) .0154(.0035) .3235(.0219)	.6264(.0203) .0119(.0025) .6319(.0311)	.3998(.0117) .0103(.0020) .3399(.0165)	.2565(.0060) .0092(.0014) .1375(.0081)
2.25	.3695(.0215) .0030(.0010) .0072(.0021)	.4724(.0264) .0071(.0019) .0685(.0088)	.4437(.0269) .0061(.0019) .1925(.0177)	.4031(.0199) .0055(.0014) .4051(.0293)	.2852(.0132) .0082(.0022) .5105(.0170)	.1771(.0071) .0049(.0012) .2417(.0087)
3.38	.2431(.0197) .0021(.0010) .0029(.0012)	.2580(.0227) .0021(.0013) .0299(.0062)	.2926(.0237) .0050(.0018) .1266(.0172)	.2417(.0154) .0043(.0013) .2275(.0239)	.1732(.0112) .0022(.0009) .3120(.0349)	.1138(.0065) .0032(.0009) .2348(.0230)
5.06	.1355(.0168) .0002(.0002) .0023(.0011)	.1680(.0167) .0006(.0005) .0257(.0061)	.1920(.0165) .0006(.0003) .1062(.0159)	.1211(.0124) .0024(.0014) .1277(.0193)	.1027(.0078) .0010(.0006) .1334(.0158)	.0610(.0051) .0006(.0003) .0921(.0144)
7.59	.1245(.0146) .0002(.0001) .0065(.0019)	.1149(.0158) .0011(.0006) .0279(.0066)	.1331(.0158) .0015(.0007) .0642(.0120)	.1126(.0115) .0019(.0009) .1009(.0170)	.0566(.0070) .0005(.0003) .1124(.0198)	.0403(.0045) .0001(.0001) .0678(.0134)
11.39	.0846(.0124) .0000(.0000) .0030(.0014)	.0617(.0113) .0000(.0000) .0111(.0041)	.0668(.0122) .0001(.0001) .0262(.0073)	.0694(.0088) .0000(.0000) .0931(.0185)	.0364(.0049) .0000(.0000) .0501(.0118)	.0249(.0030) .0000(.0000) .0653(.0183)

TABLE II OPERATING CHARACTERISTICS

THE ROWS IN EACH BOX CORRESPOND TO THE BAYES RULE, THE OPTIMAL δ^m AND THE OPTIMAL δ^a FROM TOP TO BOTTOM. THE ENTRIES IN THE FIRST COLUMN AND THE SECOND COLUMN ARE THE AVERAGE NUMBERS OF BAD POPULATIONS SELECTED AND THAT OF GOOD POPULATIONS EXCLUDED, RESPECTIVELY. THE THIRD COLUMN GIVES THE PROPORTIONS OF TIMES THAT THE OPTIMAL δ^m AND THE OPTIMAL δ^a COINCIDE WITH THE BAYES RULE.

$K=3, \Delta = .25$

σ_0 \ c	1			2			4			8			16		
	.5563	.5468		1.1233	.1272		1.4443	.0100		1.4855	.0010				
.44	.5384	.5739	.79	1.1654	.1134	.87	1.4571	.0073	.93	1.4877	.0007	1.00			
	.3772	.7628	.56	1.2407	.0917	.74	1.4610	.0062	.98	1.4877	.0007	1.00			
	.4146	.5455		.8736	.2137		1.3201	.0511		1.5245	.0127		1.6209	.0018	
.67	.3973	.5707	.85	.8852	.2103	.83	1.3307	.0422	.91	1.5251	.0133	.94	1.6234	.0017	.99
	.3920	.5729	.88	.9306	.1932	.80	1.3365	.0462	.94	1.5295	.0128	.95	1.6139	.0022	.99
	.3221	.4093		.6206	.2060		1.0017	.0685		1.3409	.0268		1.5074	.0082	
1.00	.3420	.3942	.93	.6346	.2050	.87	1.0077	.0835	.90	1.3400	.0235	.90	1.5526	.0059	.93
	.2874	.4496	.91	.6393	.2097	.83	1.0144	.0868	.91	1.3446	.0280	.92	1.5082	.0090	.93
	.2358	.2991		.4592	.1631		.6964	.0794		.9282	.0348		1.2094	.0123	
1.50	.2159	.3217	.94	.4222	.1912	.88	.6684	.0889	.90	.8586	.0401	.90	1.2667	.0096	.93
	.2009	.3392	.93	.4428	.2028	.75	.7783	.0689	.85	.9597	.0338	.88	1.2490	.0119	.88
	.1957	.2218		.3185	.1233		.4821	.0562		.6723	.0283		.8231	.0115	
2.25	.1778	.2414	.96	.3109	.1222	.95	.4574	.0640	.94	.6378	.0338	.93	.8376	.0110	.95
	.1681	.2542	.93	.2502	.1949	.76	.5639	.0530	.81	.7706	.0266	.81	.9434	.0093	.83
	.1168	.1340		.2222	.0609		.3427	.0355		.4529	.0220		.5004	.0091	
3.38	.1147	.1369	.98	.2193	.0630	.99	.3430	.0358	.99	.4553	.0225	.96	.5056	.0092	.96
	.0979	.1550	.96	.1494	.1506	.82	.4224	.0576	.75	.6239	.0230	.70	.6803	.0063	.74
	.0767	.0843		.1658	.0556		.2382	.0254		.3219	.0158		.3268	.0056	
5.06	.0767	.0843	1.00	.1646	.0559	.99	.2191	.0314	.95	.3154	.0169	.98	.3268	.0057	1.00
	.0628	.1004	.97	.0950	.1133	.85	.0992	.1064	.76	.5264	.0103	.68	.5487	.0051	.72
	.0679	.0612		.1014	.0295		.1466	.0193		.2287	.0071		.2936	.0036	
7.59	.0679	.0613	1.00	.1032	.0288	.99	.1529	.0181	.99	.2287	.0071	1.00	.2916	.0041	.99
	.0519	.0803	.97	.0468	.0724	.91	.0555	.0707	.86	.3958	.0217	.72	.5830	.0031	.69
	.0336	.0449		.0896	.0207		.0661	.0133		.1582	.0057		.1130	.0025	
11.39	.0359	.0436	1.00	.0696	.0207	1.00	.0661	.0133	1.00	.1604	.0055	1.00	.1154	.0024	1.00
	.0328	.0516	.99	.0423	.0609	.91	.0267	.0389	.93	.1557	.0358	.81	.4351	.0020	.66

TABLE II. (CONTINUED)

K=9, Δ =.25

σ_0 \ c	2			4			8			16		
.44	.9407 .9116 .5428	1.3839 1.4143 1.6259	 .58 .44	3.4612 3.4553 6.9046	.5266 .5507 .0000	 .51 .27	5.7415 6.0164 6.9130	.0985 .0734 .0000	 .55 .21	6.5694 6.6933 6.9129	.0164 .0104 .0000	 .83 .68
.67	.7468 .7585 .5135	1.0296 1.0413 1.1763	 .74 .63	2.1765 2.1248 .5055	.5122 .5506 1.1613	 .52 .04	4.1201 4.3468 7.3297	.2004 .1891 .0000	 .47 .00	5.6209 5.4525 7.3326	.0583 .0799 .0000	 .53 .08
1.00	.7046 .6837 .4780	.7335 .7625 .8769	 .71 .65	1.3833 1.3500 .4348	.4051 .4918 .7965	 .64 .23	2.8797 3.0711 3.5859	.1892 .1805 .3353	 .60 .38	3.7508 3.9663 6.4562	.0787 .0742 .0340	 .59 .06
1.50	.5596 .5598 .3589	.4617 .4719 .5910	 .87 .67	1.2072 1.2989 1.0619	.2652 .2569 .3798	 .70 .45	2.1261 2.0641 2.1176	.1268 .1447 .2182	 .69 .36	2.7257 2.6391 3.5485	.0656 .0790 .0684	 .69 .42
2.25	.4568 .4438 .2656	.2771 .2941 .4060	 .85 .67	.8300 .9275 .6402	.1599 .1475 .2902	 .81 .49	1.1611 1.1197 1.1827	.0906 .0932 .1522	 .87 .55	1.8939 1.9523 2.9112	.0389 .0413 .0202	 .75 .35
3.38	.3624 .3855 .2213	.1902 .1838 .2896	 .90 .72	.5194 .5062 .3922	.0969 .1037 .2098	 .90 .60	.8295 .8057 1.0425	.0484 .0546 .1014	 .87 .44	1.0252 1.0459 1.8259	.0266 .0273 .0273	 .91 .44
5.06	.2154 .2111 .1140	.1047 .1104 .1833	 .95 .82	.3548 .3300 .1075	.0542 .0500 .1675	 .97 .71	.6197 .5349 .6540	.0310 .0382 .0953	 .91 .60	.8333 .8293 1.5762	.0185 .0250 .0314	 .85 .36
7.59	.1170 .1175 .0735	.0828 .0833 .1193	 .98 .92	.2510 .2121 .0814	.0403 .0514 .1307	 .95 .75	.3252 .3438 .3429	.0189 .0175 .0666	 .98 .67	.3090 .4049 .8059	.0074 .0063 .0173	 .98 .52
11.39	.1206 .1206 .0690	.0296 .0296 .0780	 1.00 .88	.2203 .2203 .1455	.0168 .0168 .0820	 1.00 .81	.2595 .2596 .4435	.0121 .0111 .0490	 .99 .57	.2239 .2239 .0139	.0037 .0037 .0087	 1.00 .60

TABLE II. (CONTINUED)

K=3, Δ =.50

σ_0 \ C	1			2			4			8			16		
.44	.7662	.1556		.9525	.0184		.9940	.0009							
	.7661	.1504	.91	.9631	.0140	.97	.9940	.0009	1.00						
	.9684	.0202	.65	.9974	.0008	.94	.9950	.0000	1.00						
.67	.5349	.4023		.9472	.1109		1.2105	.0172		1.2916	.0024		1.3024	.0007	
	.5113	.4337	.86	.9617	.1080	.89	1.2335	.0126	.96	1.3029	.0011	.99	1.3001	.0009	1.00
	.5249	.5093	.64	1.1306	.0469	[illegible]	1.2409	.0125	.95	1.2961	.0018	1.00	1.3071	.0005	1.00
1.00	.3634	.4051		.6767	.1752		.9962	.0649		1.2349	.0179		1.3737	.0065	
	.3417	.4359	.87	.6810	.1770	.90	1.0264	.0601	.91	1.2505	.0160	.94	1.3978	.0056	.95
	.3007	.5089	.76	.7794	.1479	.74	1.0775	.0512	.87	1.2254	.0209	.94	1.3814	.0066	.94
1.50	.2529	.3098		.4931	.1599		.7029	.0749		.9663	.0323		1.1990	.0104	
	.2518	.3162	.92	.4971	.1614	.92	.7208	.0728	.90	1.0072	.0282	.93	1.2116	.0104	.92
	.1932	.3926	.81	.5378	.1646	.81	[illegible]	.0622	.85	1.0207	.0292	.87	1.1416	.0154	.90
2.25	.1848	.2219		.3204	.1147		.5417	.0597		.7023	.0233		.8208	.0114	
	.1908	.2178	.98	.3331	.1099	.96	.5727	.0533	.96	.7143	.0227	.95	.8071	.0127	.95
	.1238	.3034	.85	.3289	.1532	.75	.6485	.0516	.82	.7927	.0211	.84	.8491	.0147	.84
3.38	.1498	.1443		.2131	.0845		.3733	.0418		.4674	.0168		.5754	.0107	
	.1542	.1412	.97	.2186	.0824	.99	.3740	.0425	.93	.4708	.0162	.99	.5711	.0113	.95
	.1019	.2089	.87	.1675	.1588	.77	.5551	.0361	.73	.6128	.0158	.77	.7127	.0105	.77
5.06	.0983	.0984		.1693	.0527		.2370	.0315		.2346	.0159		.4252	.0063	
	.0971	.0985	1.00	.1684	.0543	.97	.2351	.0321	.99	.2571	.0183	.97	.4347	.0053	.99
	.0603	.1454	.92	.0834	.1318	.82	.3569	.0489	.75	.5049	.0111	.74	.6374	.0060	.72
7.59	.0633	.0696		.0948	.0372		.1453	.0212		[illegible]	.0102		.2352	.0044	
	.0617	.0714	.99	.0950	.0374	.99	.1460	.0214	.99	[illegible]	.0102	1.00	.2471	.0039	.99
	.0409	.1006	.95	.0399	.0923	.89	.0997	.0701	.82	.3873	.0163	.73	.5066	.0034	.70
11.39	.0306	.0364		.0724	.0226		.0919	.0129		.1302	.0062		.1567	.0028	
	.0306	.0364	1.00	.0724	.0226	1.00	.0919	.0129	1.00	.1259	.0069	1.00	.1591	.0027	1.00
	.0196	.0504	.98	.0266	.0618	.92	.0387	.0527	.89	.2758	.0218	.77	.5383	.0018	.60

TABLE II. (CONTINUED)

K=9, Δ =.50

σ_0 \ C	1			2			4			8			16		
.44	.9132 1.0130 .3485	2.2761 2.2159 2.9614	 .52 .18	3.0660 3.0756 5.3807	.7653 .8049 .0000	 .43 .01	4.8163 4.6959 5.3676	.0934 .0883 .0000	 .75 .52	5.2616 5.3533 5.3721	.0095 .0016 .0000	 .92 .90			
.67	.4778 .3761 .3761	1.8169 1.9351 1.9351	 .78 .78	1.6250 1.7540 .3805	1.0223 .9918 1.9482	 .62 .02	3.3380 3.5822 5.4389	.3971 .3533 .0000	 .33 .00	4.9058 4.7214 6.4282	.1076 .1432 .0000	 .64 .14	5.9593 6.0081 6.4291	.0178 .0131 .0000	 .87 .61
1.00	.4314 .4686 .3322	1.1253 1.1025 1.2461	 .65 .78	.9308 .9346 .3246	.1755 .8092 1.2393	 .59 .21	2.0583 2.0634 .3307	.4006 .4257 1.2630	 .59 .04	3.2729 3.1622 7.0365	.1643 .1635 .0000	 .59 .00	4.5250 4.6468 7.0559	.0592 .0603 .0000	 .56 .03
1.50	.3508 .3322 .2604	.7766 .8060 .8662	 .87 .84	.7108 .6561 .2761	.5636 .6188 .8788	 .69 .36	1.3536 1.4223 .2985	.3100 .3087 .8649	 .72 .11	2.2131 2.2493 .5227	.1312 .1379 .8208	 .72 .03	3.2553 3.4348 7.3640	.0619 .0608 .0000	 .54 .01
2.25	.2795 .2937 .2149	.5248 .5191 .6003	 .91 .86	.4351 .4602 .1661	.2990 .2941 .4992	 .92 .53	.9762 .9984 .7470	.1795 .1217 .3903	 .78 .37	1.3742 1.3010 1.4132	.0896 .1063 .2386	 .77 .39	1.8831 1.9061 2.7452	.0417 .0446 .1239	 .73 .38
3.38	.1924 .1930 .1477	.2916 .2831 .3469	 .95 .88	.3815 .3913 .2572	.2086 .2085 .3343	 .88 .63	.6387 .6179 .5281	.0810 .0902 .2105	 .90 .58	.8702 .8632 1.1838	.0866 .0896 .1503	 .90 .38	1.3012 1.2559 1.2748	.0200 .0247 .1036	 .85 .47
5.06	.2040 .2043 .1218	.1697 .1701 .2775	 .98 .81	.1862 .1867 .0833	.0935 .1000 .1966	 .98 .81	.4834 .5266 .3837	.0779 .0711 .1881	 .89 .58	.5403 .5690 .5487	.0354 .0348 .1345	 .90 .50	.8848 .9327 1.6729	.0215 .0193 .0495	 .91 .29
7.59	.0900 .0951 .0641	.1328 .1279 .1668	 .99 .94	.1805 .1950 .0695	.0556 .0501 .1646	 .98 .79	.3384 .3313 .2935	.0530 .0559 .1702	 .95 .66	.3745 .3659 .7204	.0188 .0202 .0827	 .99 .55	.3790 .3791 .7343	.0117 .0117 .0270	 .98 .50
11.39	.0639 .0589 .0398	.0605 .0655 .0934	 .99 .94	.1262 .1262 .0497	.0501 .0501 .1236	 1.00 .85	.2143 .1842 .0590	.0335 .0429 .1277	 .94 .77	.2707 .2818 .4663	.0135 .0147 .0557	 .99 .60	.3238 .3270 .7934	.0065 .0067 .0291	 .93 .44

TABLE II. (CONTINUED)

K=3, Δ =1.0

σ_0 \ C	.5			1			2			4			8			16		
.44	.2797	.0205		.2912	.0012													
	.2806	.0188	1.00	.2912	.0002	1.00												
	.2918	.0000	.97	.2925	.0000	1.00												
.67	.3582	.3668		.5272	.1155		.6525	.0152		.7129	.0011							
	.3668	.3579	.87	.5475	.0984	.92	.6636	.0113	.97	.7129	.0011	1.00						
	.6415	.0726	.43	.6861	.0195	.63	.6861	.0037	.94	.7169	.0002	1.00						
1.00	.2353	.5674		.4200	.2830		.7025	.0941		.8704	.0311		1.0064	.0045		1.0337	.0018	
	.2356	.5751	.88	.4355	.2710	.92	.7314	.0831	.93	.8795	.0302	.95	1.0111	.0041	.99	1.0385	.0019	.99
	.1180	.9715	.52	.6386	.2191	.58	.8651	.0568	.75	.9811	.0143	.85	1.0188	.0044	.97	1.0433	.0015	.99
1.50	.1696	.4858		.3104	.2793		.5297	.1270		.7252	.0590		.9434	.0202		1.0876	.0070	
	.1667	.4953	.95	.3064	.2877	.92	.5494	.12[illegible]8	.90	.7453	.0566	.92	.9642	.0184	.96	1.0713	.0083	.97
	.1000	.6712	.73	.3776	.2890	.72	.6358	.1082	.77	.8686	.0399	.80	1.0624	.0099	.85	1.1270	.0064	.92
2.25	.1009	.3477		.2184	.2049		.3531	.1118		.5182	.0564		.7213	.0240		.8310	.0104	
	.1055	.3399	.96	.2093	.2158	.95	.3389	.1201	.95	.5113	.0595	.94	.7200	.0252	.94	.8551	.0095	.96
	.0629	.4448	.87	.1772	.3238	.73	.4469	.1056	.80	.6153	.0510	.82	.8235	.0187	.86	.8600	.0119	.88
3.38	.0671	.2263		.1356	.1519		.2606	.0713		.3476	.0442		.4153	.0225		.5818	.0087	
	.0648	.2316	.98	.1324	.1562	.98	.2480	.0787	.96	.3564	.0431	.97	.4268	.0215	.98	.5751	.0095	.97
	.0433	.2875	.92	.0640	.2828	.79	.3974	.0681	.78	.4713	.0380	.80	.5916	.0163	.78	.7469	.0062	.81
5.06	.0449	.1626		.0964	.1030		.1598	.0533		.2341	.0268		.3199	.0142		.3716	.0060	
	.0449	.1626	1.00	.0970	.1034	.98	.1669	.0504	.99	.2305	.0282	.98	.3180	.0148	.98	.3741	.0059	.99
	.0297	.1999	.95	.0358	.2058	.84	.0543	.1778	.76	.3926	.0353	.72	.5687	.0080	.73	.5578	.0047	.78
7.59	.0332	.0545		.0674	.0577		.1039	.0368		.1583	.0200		.2426	.0073		.2287	.0038	
	.0332	.0545	1.00	.0638	.0615	.93	.1141	.0345	.99	.1603	.0195	1.00	.2449	.0071	1.00	.2311	.0037	1.00
	.0208	.1246	.96	.0281	.1333	.68	.0306	.1335	.82	.2729	.0371	.78	.4766	.0064	.73	.4201	.0052	.75
11.39	.0218	.0648		.0403	.0349		.0837	.0206		.1065	.0138		.1402	.0058		.1911	.0027	
	.0210	.0665	1.00	.0403	.0349	1.00	.0855	.0198	1.00	.1107	.0129	1.00	.1425	.0056	1.00	.1959	.0026	1.00
	.0138	.0843	.97	.0140	.0785	.93	.0199	.1018	.87	.0363	.0861	.84	.4804	.0035	.64	.5273	.0015	.65

TABLE II. (CONTINUED)

K=9, Δ =1.0

σ_0 \ C	.5	1	2	4	8	16
.44	1.3924 1.0381 1.3927 1.0984 .58 2.1643 .0000 .11	1.9926 .1325 2.0550 .0849 .81 2.1601 .0000 .75	2.1648 .0029 2.1648 .0029 1.00 2.1719 .0000 .99			
.67	.5538 2.8675 .5373 2.9510 .62 .1742 3.9379 .12	1.6325 1.2974 1.6381 1.3430 .56 4.2251 .0000 .01	2.8473 .4321 2.8424 .4772 .56 4.2752 .0000 .12	3.8555 .0678 3.9488 .0511 .76 4.2477 .0000 .62	4.1572 .0104 4.1682 .0113 .93 4.2768 .0000 .89	
1.00	.2988 2.0925 .2352 2.2588 .70 .1715 2.4051 .58	.8150 1.3392 .8561 1.3404 .62 .1683 2.3525 .16	1.7229 .7252 1.8779 .6902 .58 .1729 2.3952 .02	2.9470 .2967 3.0915 .2912 .54 5.7603 .0000 .02	4.0555 .0909 3.9255 .1209 .66 5.6146 .0000 .17	4.7400 .0312 5.0089 .0200 .65 5.7389 .0000 .35
1.50	.1947 1.4302 .1623 1.5078 .87 .1534 1.5289 .86	.5275 .9024 .4802 .9852 .70 .1563 1.4912 .26	1.0839 .5388 1.1259 .5409 .70 .1540 1.4890 .10	2.0030 .2823 1.9110 .3202 .66 .1469 1.5362 .02	2.6489 .1187 2.8192 .1090 .67 6.6510 .0008 .00	3.4654 .0560 3.6722 .0528 .54 6.6994 .0000 .01
2.25	.1571 .7943 .1401 .8374 .90 .1143 .9016 .85	.3851 .5597 .3572 .6019 .80 .1286 .9532 .48	.6039 .3637 .5833 .3831 .83 .1097 .8995 .31	1.1882 .2068 1.2177 .2063 .80 .1212 .9799 .16	1.8056 .0951 1.8583 .0978 .69 7.1606 .0000 .00	2.2956 .0447 2.1875 .0567 .74 7.1113 .0005 .00
3.38	.1031 .5231 .0965 .5385 .94 .0793 .5793 .92	.1921 .3239 .1992 .3210 .96 .0720 .5038 .76	.4279 .2250 .4329 .2300 .91 .1345 .5615 .53	.7161 .1232 .7603 .1174 .85 .3436 .5007 .40	1.1018 .0571 1.0418 .0671 .92 1.1293 .4047 .38	1.4396 .0310 1.4617 .0330 .80 3.9144 .1258 .06
5.06	.0661 .2742 .0696 .2678 .99 .0484 .3165 .94	.1376 .1984 .1483 .1891 .98 .0583 .3291 .79	.3050 .1355 .2722 .1527 .95 .0912 .4017 .59	.3050 .0752 .3234 .0735 .96 .2887 .2388 .59	.6648 .0324 .6924 .0301 .93 1.5093 .0769 .35	.8152 .0139 .8158 .0145 .94 1.5755 .0643 .37
7.59	.0929 .1876 .0831 .2078 .97 .0527 .2874 .86	.1225 .1073 .1086 .1234 .97 .0455 .2403 .81	.2137 .0928 .2286 .0876 .92 .0713 .2604 .69	.3130 .0625 .3149 .0644 .94 .3099 .1894 .57	.3697 .0174 .3791 .0169 .97 .7931 .0909 .47	.5527 .0082 .5340 .0095 .98 1.2700 .0355 .41
11.39	.0454 .1631 .0454 .1631 1.00 .0317 .1994 .95	.0569 .0665 .0569 .0665 1.00 .0230 .1225 .91	.1308 .0349 .1376 .0316 .99 .0236 .1227 .83	.2497 .0243 .2497 .0243 1.00 .0788 .1834 .70	.2103 .0147 .2192 .0136 .99 .2337 .0682 .71	.3205 .0064 .3205 .0064 1.00 .4235 .0694 .67

TABLE III.

ESTIMATED d-VALUES FOR THE OPTIMAL δ^m

σ_0 \ c	.5	1	2	4	8	16	.5	1	2	4	8	16
	K=3, Δ=.25						K=9, Δ=.25					
.44		.6046	2.1681	3.7070	5.2520				.4076	1.6880	3.1407	4.1256
.67		.2266	1.3308	2.4711	3.1836	3.9854			.2717	1.2091	2.2187	2.8316
1.00		.2737	1.0657	1.7797	2.5294	3.3921			.2454	1.1081	1.8152	2.4658
1.50		.1057	.7517	1.4668	2.1038	2.8730			.3251	1.0582	1.5509	2.0677
2.25		.1079	.8520	1.3554	1.9026	2.6029			.5292	1.1970	1.5469	2.1488
3.38		.1833	.7986	1.4192	1.8781	2.4361			.6631	1.0959	1.6487	2.1910
5.06		.2648	.8280	1.2892	1.7989	2.4360			.5197	1.2944	1.8346	2.0436
7.59		.1686	.8242	1.4017	1.9438	2.4229			.6326	1.0[illegible]14	1.9337	2.4431
11.39		.2569	.8643	1.4390	1.9710	2.2446			.8844	1.4214	1.9083	2.4002
	K=3, Δ=.50						K=9, Δ=.50					
.44		2.1572	3.7576	4.6502				.7906	2.0710	3.4991	5.0995	
.67		1.0391	2.1839	3.3917	4.2796	4.6762		.0000	1.2464	2.2379	2.8148	3.9145
1.00		.6040	1.4808	2.3435	3.1461	3.8496		.2222	.8727	1.6505	2.2304	2.9658
1.50		.4939	1.2570	1.9361	2.6181	3.1760		.1333	.6862	1.4193	2.0923	2.6452
2.25		.5304	1.1760	1.8682	2.2840	2.8038		.2595	.9927	1.4305	1.9776	2.5138
3.38		.4552	1.1332	1.6755	2.3222	2.7550		.4240	.9194	1.5243	1.8893	2.4630
5.06		.4940	1.0801	1.6863	2.1657	2.7569		.4329	.8559	1.3859	1.9873	2.5174
7.59		.3775	1.0029	1.6789	2.2435	2.7378		.4683	1.0987	1.3360	1.9157	2.5915
11.39		.4938	1.1011	1.6307	2.0799	2.7224		.2981	1.1815	1.3571	2.0548	2.5919
	K=3, Δ=1.0						K=9, Δ=1.0					
.44	3.9204	5.3177					2.6400	4.2282	5.2869			
.67	1.7627	2.9075	3.9441	5.0438			.7982	1.8799	2.9796	4.1874	5.1159	
1.00	.8969	1.7525	2.6293	3.3965	4.1991	4.7542	.1615	1.0949	2.0000	2.7778	3.1879	4.2627
1.50	.5250	1.2341	2.0514	2.6403	3.3854	3.7850	.0263	.7834	1.5887	2.2136	2.8867	3.4756
2.25	.5081	1.0389	1.6517	2.3356	2.9320	3.5011	.1360	.7598	1.3617	2.0292	2.6420	2.9813
3.38	.3056	.9678	1.5832	2.2664	2.8035	3.2363	.1543	.7958	1.3381	2.0995	2.4958	2.9756
5.06	.4045	.9447	1.6608	2.1144	2.6623	3.2459	.3714	1.0135	1.4671	2.0125	2.7515	3.0522
7.59	.4046	.9013	1.5872	2.1949	2.7446	3.2540	.3567	.7652	1.4258	1.9957	2.6211	3.0094
11.39	.3753	.9977	1.5598	2.2023	2.7268	3.2337	.2749	1.0730	1.6401	2.2073	2.7092	3.1258

CHAPTER 2

Γ-MINIMAX RULES FOR PARTITIONING TREATMENTS WITH RESPECT TO A CONTROL

2.1 Introduction

In many fields of research one is faced with the problem of comparing k experimental categories with reference to a 'standard' or a 'control'. Following the initial investigation by Paulson (1952), this problem has been studied in several different formulations by Dunnett (1955), Gupta and Sobel (1958) and Lehmann (1961) among others. Tong (1969) has studied a problem where the treatment populations are to be partitioned into two sets, one consisting of 'better' populations and another consisting of 'worse' populations. Later Randles and Hollander (1971) applied Γ-minimax principle to the same problem.

Let $\pi_1,\ldots,\pi_k$ denote the k experimental categories or 'treatment' populations and let π_0 denote the 'control' population. We assume that each population is characterized by a real-valued location parameter θ_i $(i = 0,1,\ldots,k)$. We consider a problem in which the treatment populations $\pi_1,\ldots,\pi_k$ are to be classified as 'better' than, 'worse' than or 'close' to the control π_0 if the corresponding parameter values are much larger than, much smaller than or sufficiently close to the value of θ_0. Similar problems have been considered in Bhattacharyya (1956,1958) and in Seeger (1972) when $\theta_0,\ldots,\theta_k$ are means of independent normal distributions. We apply the Γ-minimax principle to this problem. Γ-minimax principle is known as one of the techniques for the use of incomplete prior information. It is assumed that although a prior distribution on the states of nature is not available, it is known to belong to some family, Γ, of distributions. It then requires the decision maker to select a decision rule which minimizes the supremum of the overall Bayes risk over distributions in Γ. Such a rule is called a Γ-minimax rule. Such a principle was first used by Robbins (1951) and independently by Hodges and Lehmann (1952) and Menges (1966). The name Γ-minimax was first used by Blum and Rosenblatt (1967). This principle has been applied to various problems by Jackson, O'Donovan, Zimmer and Deely (1970), Solomon (1972a, 1972b), DeRouen and Mitchell (1974), Gupta and Huang (1975,1977) ,Berger (1977) and Miescke (1979).

In Section 2.2 definitions and notations are introduced, and a formulation of the problem is given. The loss function and the incomplete prior are introduced. Results analogous to standard ones on minimaxity are given to help find Γ-minimax decision rules.

In Section 2.3 a Γ-minimax decision rule is derived for the case in which θ_0 is known and a minimax decision rule is found for the same case when a specific loss function is assumed. A solution is provided for the example in which $\theta_0,\ldots,\theta_k$ are unknown means of normal distributions.

In Section 2.4, the case in which the control population parameter θ_0 is unknown is treated. Rules are derived which are Γ-minimax among procedures for which the classification of the i-th population depends only on x_i and x_0 where $x_0,\ldots,x_k$ are independent random variables representing populations $\pi_0,\ldots,\pi_k$, respectively. Specific examples are again given. One example is to classify normal populations by their locations (means) wrt the mean of a normal control population. Another is to classify normal populations by their variances.

Section 2.5 consists of comparisons of Γ-minimax rules with Bayes rules wrt independent normal priors for the case of normal populations with common known variance.

2.2 Formulation of the problem

Let $X_0,X_1,\ldots,X_k$ be k+1 independent random variables representing the control π_0 and the k treatment populations $\pi_1,\ldots,\pi_k$, respectively, with X_i having probability density function $f_i(x-\theta_i)$ with respect to the Lebesgue measure on the real line R where $\theta_i \in \Theta = R$, $i = 0, 1,\ldots,k$. The random variables $X_0,\ldots,X_k$ may represent sufficient statistics or other statistics based on which we wish to make statistical decisions. We assume that each $f_i(\cdot)$ $(i = 0,1,\ldots,k)$ is symmetric about the origin and strongly unimodal, i.e., $f_i(\cdot)$ is log-concave on the real line. Hence $f_i(x-\theta_i)$ has the monotone likelihood ratio (MLR) property. Obviously, we do not need any observations from π_0 when θ_0 is assumed known; therefore, it will be understood that, in such a case, the random variable X_0 is deleted from our consideration.

The action space $\mathcal{A}$ can be written as $\mathcal{A} = \mathcal{A}_1 \times \ldots \times \mathcal{A}_k$ where $\mathcal{A}_i = \{1,2,3\}$ for $i = 1,\ldots,k$. The action $a = (a_1,\ldots,a_k) \in \mathcal{A}$ is to be interpreted in such a way that, for $i = 1,\ldots,k$, the treatment population π_i is classified as 'worse' if $a_i = 1$, 'close' if $a_i = 2$ and 'better' if $a_i = 3$. The loss $L(\underline{\theta},a)$ incurred by the action $a \in \mathcal{A}$ for $\underline{\theta} = (\theta_0,\theta_1,\ldots,\theta_k)$ is assumed to be of the following form.

$$L(\underline{\theta},a) = \sum_{i=1}^{k} L_i(\underline{\theta},a_i) \tag{2.2.1}$$

where $L_i(\underline{\theta},a_i)$ is defined below and denotes the loss in each component problem incurred by the component action a_i. For arbitrary but fixed positive constants Δ_1 and Δ_2 such that $\Delta_1 < \Delta_2$, we define five disjoint and exhaustive subregions R_W, R_{I_1}, R_c, R_{I_2}, R_B of the real line R by $R_W = (-\infty, -\Delta_2]$, $R_{I_1} = (-\Delta_2,-\Delta_1)$, $R_c = [-\Delta_1,\Delta_1]$, $R_{I_2} = (\Delta_1,\Delta_2)$ and $R_B = [\Delta_2,\infty)$, and define $L_i(\underline{\theta},a_i)$ as in the next table.

Table of loss $L_i(\underline{\theta},a_i)$

State of nature \ Action a_i	1	2	3
$\theta_i-\theta_0 \in R_W$	0	ℓ_1	$\ell_1+\ell_3$
$\theta_i-\theta_0 \in R_{I_1}$	0	0	ℓ_4
$\theta_i-\theta_0 \in R_c$	ℓ_2	0	ℓ_2
$\theta_i-\theta_0 \in R_{I_2}$	ℓ_4	0	0
$\theta_i-\theta_0 \in R_B$	$\ell_1+\ell_3$	ℓ_1	0

$(\ell_i \geq 0, i=1,\ldots,4)$ (2.2.2)

This loss function reduces to the one considered in Bhattacharyya (1958) and in Seeger (1972) if $\ell_1 = \ell_2 = \ell_4 = \ell_1+\ell_3 = 1$. Note that the above loss function assumes the indifference zones in the sense that we do not distinguish between the actions 1 and 2 (2 and 3) when $\theta_i-\theta_0 \in R_{I_1}$ ($\theta_i-\theta_0 \in R_{I_2}$, respectively). It should be pointed out that Bhattacharyya (1958) considered this loss function to avoid an irregularity of a similar loss function without indifference zones. In fact, Bhattacharyya (1956) derived an admissible minimax decision rule assuming a simple 0-1 type loss function when θ_0 is assumed known and $\theta_1,\ldots,\theta_k$ are unknown means of normal distributions. However, the irregularity of such a loss function has been pointed out in the sense that the minimax risk does not tend to zero even if the sample sizes increase indefinitely, and the same problem has been studied afresh by Bhattacharyya (1958) with a loss function of the type given in (2.2.2).

For given $\underline{x} = (x_0,x_1,\ldots,x_k)$ consider decision rules of the form

$$\delta(\underline{x}) = (\delta_1(\underline{x}),\ldots,\delta_k(\underline{x})) \tag{2.2.3}$$

where $\delta_i(\underline{x}) = (\delta_i(1|\underline{x}),\ \delta_i(2|\underline{x}),\ \delta_i(3|\underline{x}))$ and, for j=1,2,3, $\delta_i(j|\underline{x})$ denotes the conditional probability of taking action j in the i-th component problem. Note that there is no loss of generality in considering decision rules of the form given in (2.2.3). The risk function of a rule δ for fixed $\underline{\theta}$ is then $R(\underline{\theta},\delta) = \sum_{i=1}^{k} R_i(\underline{\theta},\delta_i)$ where $R_i(\underline{\theta},\delta_i) = E_{\underline{\theta}}[L_i(\underline{\theta},\delta_i(\underline{x}))]$. For a prior distribution $\tau(\underline{\theta})$ of $\underline{\theta}$, the overall risk of a rule δ wrt τ is denoted by $r(\tau,\delta) = \sum_{i=1}^{k} r_i(\tau,\delta_i)$ where $r_i(\tau,\delta_i) = \int R_i(\underline{\theta},\delta_i)d\tau(\underline{\theta})$.

Suppose that partial prior information is available to a decision maker such that, for each i, he can specify $\gamma_i = P[\theta_i - \theta_0 \in R_W \cup R_B]$ and $\gamma_i' = P[\theta_i - \theta_0 \in R_C]$ where $\gamma_i + \gamma_i' \le 1$ for i = 1,...,k. Let Γ denote the class of all such prior distributions, i.e.,

$$\Gamma = \{\tau(\underline{\theta}): \int_{\theta_i - \theta_0 \in R_W \cup R_B} d\tau(\underline{\theta}) = \gamma_i, \quad \int_{\theta_i - \theta_0 \in R_C} d\tau(\underline{\theta}) = \gamma_i' \quad \text{for } i = 1,\ldots,k\}. \tag{2.2.4}$$

Now our goal is to find a Γ-minimax rule δ^* as defined below.

Definition 2.2.1. A rule δ^* is a Γ-minimax decision rule if

$$\sup_{\tau \in \Gamma} r(\tau, \delta^*) = \inf_{\delta} \sup_{\tau \in \Gamma} r(\tau, \delta),$$

and $\sup_{\tau \in \Gamma} r(\tau, \delta^*)$ is called the Γ-minimax value.

Many authors have found Γ-minimax rules by finding Bayes rules with respect to 'least favorable' priors in the class Γ. However, such a method was found not to lead to the solution of our problem; the following results analogous to standard results on minimaxity are found useful.

Lemma 2.2.1. Suppose $\{\tau_n, n = 1,2,\ldots\}$ is a sequence of priors in Γ. If $\overline{\lim_n} \inf_{\delta} r(\tau_n, \delta) \ge c$ and if $\sup_{\tau \in \Gamma} r(\tau, \delta^*) \le c$, then δ^* is a Γ-minimax decision rule and c is the Γ-minimax value.

Proof. The result is an immediate consequence of the following inequalities.

$$\sup_{\tau\in\Gamma} \inf_{\delta} r(\tau,\delta) \geq \overline{\lim_{n}} \inf_{\delta} r(\tau_n,\delta)$$

$$\geq c$$

$$\geq \sup_{\tau\in\Gamma} r(\tau,\delta^*)$$

$$\geq \inf_{\delta} \sup_{\tau\in\Gamma} r(\tau,\delta)$$

$$\geq \sup_{\tau\in\Gamma} \inf_{\delta} r(\tau,\delta).$$

The next result is useful when we have certain invariance under a finite group. Following Ferguson (1967), let $G = \{g_1,\ldots,g_N\}$, $\bar{G}$ and $\tilde{G}$ denote the group of transformations on the sample space $\mathscr{X}$, the induced group on the parameter space and that on the action space, respectively.

Lemma 2.2.2. Suppose that a given decision problem is invariant under a finite group. If $\bar{g}\tau \in \Gamma$ for any $\tau \in \Gamma$ and $g \in G$ where $\bar{g}\tau(B) = \tau(\bar{g}^{-1}(B))$, then a Γ-minimax rule within the class of invariant behaviorial decision rules is Γ-minimax.

Proof. Let δ^I denote the invariant rule generated from a given rule δ, i.e., $\delta^I(x,A) = \frac{1}{N}\sum_{j=1}^{N} \delta(g_j(x), \tilde{g}_j(A))$ for each $x \in \mathscr{X}$ and $A \subset \mathcal{Q}$. Then clearly δ^I is invariant, and

$$\sup_{\tau\in\Gamma} r(\tau,\delta^I) = \sup_{\tau\in\Gamma} \frac{1}{N}\sum_{j=1}^{N} r(\tau,\delta^{g_j}) \text{ where } \delta^{g_j}(x,A) = \delta(g_j(x),\tilde{g}_j(A))$$

$$= \sup_{\tau\in\Gamma} \frac{1}{N}\sum_{j=1}^{N} r(\bar{g}_j\tau,\delta)$$

$$\leq \frac{1}{N}\sum_{j=1}^{N} \sup_{\tau\in\Gamma} r(\bar{g}_j\tau,\delta)$$

$$\leq \frac{1}{N} \sum_{j=1}^{N} \sup_{\tau \in \Gamma} r(\tau,\delta) \text{ since } \bar{g}\tau \in \Gamma \text{ for any } \tau \in \Gamma \text{ and } g \in G$$

$$= \sup_{\tau \in \Gamma} r(\tau,\delta).$$

This completes the proof.

It should be pointed out that our decision problem is invariant under the symmetry about the origin, and that randomized decision rules and behavioral rules are equivalent since the distributions of X's are assumed to be continuous.

2.3 Known control

In this section θ_0 is assumed known and thus we may assume $\theta_0 = 0$ without loss of generality. Hence $\underline{x}$ and $\underline{\theta}$ in this section denote $(x_1,\ldots,x_k)$ and $(\theta_1,\ldots,\theta_k)$, respectively.

Lemma 2.3.1. Suppose that a decision rule $\delta(\underline{x})$ of the form in (2.2.3) is determined by $\delta_i(1|\underline{x}) = I_{(-\infty,-d_i)}(x_i)$, $\delta_i(2|\underline{x}) = I_{[-d_i,d_i]}(x_i)$ and $\delta_i(3|\underline{x}) = I_{(d_i,\infty)}(x_i)$ for $d_i \geq 0$ and $i = 1,\ldots,k$. Then, for $i=1,\ldots,k$,

$$\sup_{\tau \in \Gamma} r_i(\tau,\delta_i) \leq r_i \tag{2.3.1}$$

where, for $i = 1,\ldots,k$,

$$r_i = \int_{d_i}^{\infty} [\ell_3\gamma_i f_i(x+\Delta_2)+\ell_2\gamma_i'(f_i(x-\Delta_1)+f_i(x+\Delta_1))+\ell_4(1-\gamma_i-\gamma_i') f_i(x+\Delta_1)]dx + \int_{-\infty}^{d_i} \ell_1\gamma_i f_i(x-\Delta_2)dx.$$

Proof. It follows from (2.2.2) that, for $\theta_i \leq -\Delta_2$,

$$R_i(\underline{\theta},\delta_i) = \ell_1 \int_{-d_i}^{\infty} f_i(x-\theta_i)dx + \ell_3 \int_{d_i}^{\infty} f_i(x-\theta_i)dx$$

$$\leq \ell_1 \int_{-d_i}^{\infty} f_i(x+\Delta_2)dx+\ell_3 \int_{d_i}^{\infty} f_i(x+\Delta_2)dx$$

$$= \ell_1 \int_{-\infty}^{d_i} f_i(x-\Delta_2)dx+\ell_3 \int_{d_i}^{\infty} f_i(x+\Delta_2)dx \text{ by the symmetry of } f_i(\cdot).$$

In a similar manner it can be shown that

$$R_i(\underline{\theta},\delta_i) \begin{cases} \leq \ell_1 \int_{-\infty}^{d_i} f_i(x-\Delta_2)dx+\ell_3 \int_{d_i}^{\infty} f_i(x+\Delta_2)dx & \text{for } \theta_i > \Delta_2, \\ \leq \ell_4 \int_{d_i}^{\infty} f_i(x+\Delta_1)dx & \text{for } \Delta_1 < |\theta_i|<\Delta_2. \end{cases}$$

If $-\Delta_1 \leq \theta_i \leq \Delta_1$, then $R_i(\underline{\theta},\delta_i) = \ell_2[\int_{-\infty}^{-d_i} f_i(x-\theta_i)dx+\int_{d_i}^{\infty} f_i(x-\theta_i)dx]$,

therefore, $R_i'(\underline{\theta},\delta_i) = \ell_2[f_i(d_i-\theta_i)-f_i(-d_i-\theta_i)]$

$$= \ell_2 \, f_i(\theta_i+d_i)[\frac{f_i(\theta_i-d_i)}{f_i(\theta_i+d_i)} -1],$$

where R_i' denotes the derivative of R_i w.r.t. θ_i.

It follows from the MLR property of $f_i(x-\theta_i)$ that $R_i'(\underline{\theta},\delta_i)$ has at most one change of sign, from negative to positive if there is any sign change at all; therefore, $R_i(\underline{\theta},\delta_i)$ attains it supremum over $\theta_i \in [-\Delta_1,\Delta_1]$ at either $\theta_i = -\Delta_1$ or $\theta_i = \Delta_1$. Hence it follows from the symmetry of $f_i(\cdot)$ that, for $\theta_i \in [-\Delta_1,\Delta_1]$,

$$R_i(\underline{\theta},\delta_i) \leq \ell_2 \int_{d_i}^{\infty} [f_i(x-\Delta_1)+f_i(x+\Delta_1)]dx.$$

It follows from (2.2.4) that $\sup_{\tau\in\Gamma} r_i(\tau,\delta_i) \leq r_i$.

Now we derive the Γ-minimax rule in the following result.

Theorem 2.3.1. Assume that independent random variables $X_1,\ldots,X_k$ have strong unimodal symmetric p.d.f's $f_1(x_1-\theta_1),\ldots,f_k(x_k-\theta_k)$, respectively, and that the loss function is given by (2.2.1) and (2.2.2). Then the Γ-minimax rule δ^* is given by $\delta_i^*(1|\underline{x}) = I_{(-\infty,-d_i)}(x_i)$, $\delta_i^*(2|\underline{x}) = I_{[-d_i,d_i]}(x_i)$ and $\delta_i^*(3|\underline{x}) = I_{(d_i,\infty)}(x_i)$ for $i = 1,\ldots,k$ where each d_i is determined by $d_i = c_i^+ = \max(c_i,0)$ with c_i being defined by

$$\begin{aligned}&\ell_3\gamma_i f_i(x+\Delta_2)+\ell_2\gamma_i'(f_i(x-\Delta_1)+f_i(x+\Delta_1))+\ell_4(1-\gamma_i-\gamma_i')f_i(x+\Delta_1)\\ &\quad \leq, > \ell_1\gamma_i f_i(x-\Delta_2) \quad \text{as} \quad x \geq, < c_i.\end{aligned} \tag{2.3.2}$$

Proof. We first show that the decision rule δ^* is well defined by verifying the existence of a c_i satisfying (2.3.2). Note that the difference between both sides in (2.3.2) can be written as

$$\begin{aligned}&\ell_3\gamma_i f_i(x+\Delta_2)+\ell_2\gamma_i'(f_i(x-\Delta_1)+f_i(x+\Delta_1))+\ell_4(1-\gamma_i-\gamma_i')f_i(x+\Delta_1)-\ell_1\gamma_i f_i(x-\Delta_2)\\ &= f_i(x-\Delta_2)[\ell_3\gamma_i \frac{f_i(x+\Delta_2)}{f_i(x-\Delta_2)} + \ell_2\gamma_i'(\frac{f_i(x-\Delta_1)}{f_i(x-\Delta_2)} + \frac{f_i(x+\Delta_1)}{f_i(x-\Delta_2)})+\ell_4(1-\gamma_i-\gamma_i')\\ &\qquad \frac{f_i(x+\Delta_1)}{f_i(x-\Delta_2)} -\ell_1\gamma_i].\end{aligned}$$

Then the existence of such a c_i follows from the MLR property of $f_i(x-\theta_i)$. Consider a sequence of prior distributions $\{\tau_n\}$ of $\underline{\theta}$ under which $\theta_1,\ldots,\theta_k$ are independent, $P(\theta_i=\Delta_2) = P(\theta_i = -\Delta_2) = \gamma_i/2$, $P(\theta_i=\Delta_1) = P(\theta_i = -\Delta_1) = \gamma_i'/2$ and $P(\theta_i = -\Delta_1-n^{-1}) = P(\theta_i=\Delta_1+n^{-1}) = (1-\gamma_i-\gamma_i')/2$. Then clearly $\tau_n \in \Gamma$ for $n > (\Delta_2-\Delta_1)^{-1}$, and the overall risk of the Bayes rule wrt τ_n for the i-th component problem can be written as, for $i = 1,\ldots,k$,

$$\inf_{\delta_i} r_i(\tau_n,\delta_i) = \int_{-\infty}^{\infty} p_n(x)dx,$$

where $p_n(x) = \frac{1}{2}\min\{p_n(1,x),\ p(2,x),\ p_n(3,x)\}$,

$$p_n(1,x) = \ell_2\gamma_i'(f_i(x-\Delta_1)+f_i(x+\Delta_1))+\ell_4(1-\gamma_i-\gamma_i')f_i(x-\Delta_1-\tfrac{1}{n}) + (\ell_1+\ell_3)\gamma_i f_i(x-\Delta_2),$$

$$p(2,x) = \ell_1\gamma_i(f_i(x+\Delta_2)+f_i(x-\Delta_2)) \text{ and}$$

$$p_n(3,x) = (\ell_1+\ell_3)\gamma_i f_i(x+\Delta_2)+\ell_4(1-\gamma_i-\gamma_i')f_i(x+\Delta_1+\tfrac{1}{n})+\ell_2\gamma_i'(f_i(x+\Delta_1)+f_i(x-\Delta_1)).$$

Since strongly unimodality of $f_i(\cdot)$ implies the continuity of $f_i(\cdot)$, $p_n(x)$ converges, as $n \to \infty$, to $p(x) = \frac{1}{2}\min_{1\le j\le 3} p(j,x)$ where $p(2,x)$ is defined as above,

$$p(1,x) = \ell_2\gamma_i'(f_i(x-\Delta_1)+f_i(x+\Delta_1))+\ell_4(1-\gamma_i-\gamma_i')f_i(x-\Delta_1)+(\ell_1+\ell_3)\gamma_i f_i(x-\Delta_2)$$

and

$$p(3,x) = (\ell_1+\ell_3)\gamma_i f_i(x+\Delta_2)+\ell_4(1-\gamma_i-\gamma_i')f_i(x+\Delta_1)+\ell_2\gamma_i'(f_i(x+\Delta_1) + f_i(x-\Delta_1)).$$

Since $p_n(x)$ is bounded above by $p(2,x)$ which is integrable, it follows from the Lebesgue convergence theorem that

$$\lim_n \inf_{\delta_i} r_i(\tau_n,\delta_i) = \int_{-\infty}^{\infty} p(x)dx = \frac{1}{2}\int_{-\infty}^{\infty} \min_{1\le j\le 3} p(j,x)dx. \qquad (2.3.3)$$

Furthermore, it can be easily shown that, for any $\Delta > 0$, $f_i(x-\Delta)-f_i(x+\Delta) \geq 0, \leq 0$ as $x \geq 0, \leq 0$ and therefore $p(1,x) \geq, \leq p(3,x)$ as $x \geq 0, \leq 0$. It follows from (2.3.3) that

$$\lim_n \inf_{\delta_i} r_i(\tau_n,\delta_i) = \frac{1}{2}\int_{-\infty}^{0} \min_{1\leq j\leq 2} p(j,x)dx + \frac{1}{2}\int_{0}^{\infty} \min_{2\leq j\leq 3} p(j,x)dx$$

$$= \int_0^\infty \min_{2\leq j\leq 3} p(j,x)dx$$

$$= \int_0^\infty \min[\ell_3\gamma_i f_i(x+\Delta_2)+\ell_4(1-\gamma_i-\gamma_i')f_i(x+\Delta_1)+ \quad (2.3.4)$$

$$\ell_2\gamma_i'(f_i(x+\Delta_1)+f_i(x-\Delta_1)),\ \ell_1\gamma_i f_i(x-\Delta_2)]dx$$

$$+ \int_{-\infty}^{0} \ell_1\gamma_i f_i(x-\Delta_2)dx.$$

The second equality in the above follows from the fact that $p(1,x) = p(3,-x)$ and $p(2,x) = p(2,-x)$. It follows from (2.3.2) and (2.3.4) that

$$\lim_n \inf_{\delta_i} r_i(\tau_n,\delta_i) = \int_{d_i}^{\infty} [\ell_3\gamma_i f_i(x+\Delta_2)+\ell_4(1-\gamma_i-\gamma_i')f_i(x+\Delta_1)+\ell_2\gamma_i'$$

$$(f_i(x+\Delta_1)+f_i(x-\Delta_1))]dx$$

$$+ \int_{-\infty}^{d_i} \ell_1\gamma_i f_i(x-\Delta_2)dx$$

$$\geq \sup_{\tau\in\Gamma} r_i(\tau,\delta_i^*) \text{ by Lemma 2.3.1.}$$

Therefore,

$$\lim_n \inf_{\delta} r(\tau_n,\delta) = \lim_n \sum_{i=1}^{k} \inf_{\delta_i} r_i(\tau_n,\delta_i)$$

$$\geq \sum_{i=1}^{k} \sup_{\tau\in\Gamma} r_i(\tau,\delta_i^*) \quad (2.3.5)$$

$$\geq \sup_{\tau\in\Gamma} r(\tau,\delta^*).$$

Here, the first equality follows from the fact that the Bayes rule consists of the Bayes rules for the component problems. Lemma 2.2.1 then yields that the decision rule δ^* is Γ-minimax among all decision rules.

We can derive a minimax rule using the proof of the above result. For this purpose, assume that $\ell_1 = \ell_2 = \ell_4 = 1$ and $\ell_3 = \ell < 1$. Let us consider a rule δ^M of the type given in Lemma 2.3.1 where each d_i is determined so that, for $F_i(x) = \int_{-\infty}^{x} f_i(t)dt$,

$$F_i(d_i-\Delta_2) + \ell F_i(-d_i-\Delta_2) = F_i(-d_i-\Delta_1)+F_i(-d_i+\Delta_1). \qquad (2.3.6)$$

Note that the existence of such a positive d_i follows from strong unimodality and the symmetry of $f_i(\cdot)$. Let us define γ_i and $\gamma_i'=1-\gamma_i$ for each $i = 1,\ldots,k$ by

$$\gamma_i = [f_i(d_i-\Delta_1)+f_i(d_i+\Delta_1)]/[f_i(d_i-\Delta_2)-\ell f_i(d_i+\Delta_2)+f_i(d_i-\Delta_1)+f_i(d_i+\Delta_1)].$$

Since $\gamma_i \in [0,1]$, we can consider a family of priors, Γ given in (2.2.4). Then it follows from Theorem 2.3.1 that the corresponding Γ-minimax rule is of the same type as δ^M except that now $d_i^* = \max(c_i,0)$ where c_i is determined so that

$$H(c_i) = \gamma_i[\ell f_i(c_i+\Delta_2)-f_i(c_i-\Delta_2)]+\gamma_i'(f_i(c_i-\Delta_1)+f_i(c_i+\Delta_1)) = 0.$$

Since $H(d_i) = 0$ and $d_i > 0$, $d_i^* = d_i$, i.e., the rule δ^M is the Γ-minimax procedure ; therefore it follows from (2.3.5) and (2.3.1) that, for some sequence of priors,

$$\liminf_{n} \inf_{\delta} r(\tau_n,\delta) \geq \sum_{i=1}^{k} \gamma_i[F_i(d_i-\Delta_2) + \ell F_i(-d_i-\Delta_2)]+\gamma_i'[F_i(-d_i-\Delta_1)+ F_i(-d_i+\Delta_1)]$$

$$= \sum_{i=1}^{k}[F_i(-d_i-\Delta_1)+F_i(-d_i+\Delta_1)] \quad \text{by (2.3.6)}$$

$$= \sum_{i=1}^{k} \sup_{\underline{\theta}} R_i(\underline{\theta},\delta^M)$$

$$\geq \sup_{\underline{\theta}} R(\underline{\theta},\delta^M).$$

The last equality in the above follows from similar arguments as in the proof of Lemma 2.3.1. Therefore, we have the next general result which includes Bhattacharyya's (1958) result as a special case.

Corollary 2.3.1. Under the assumptions in Theorem 2.3.1, if $\ell_1=\ell_2=\ell_4=1$ and $\ell_3=\ell < 1$, then a minimax decision rule δ^M is given by $\delta_i^M(1|\underline{x}) = I_{(-\infty,-d_i)}(x_i)$, $\delta_i^M(2|\underline{x}) = I_{[-d_i,d_i]}(x_i)$ and $\delta_i^M(3|\underline{x}) = I_{(d_i,\infty)}(x_i)$ where each d_i is determined by (2.3.6).

The following example illustrates the application of these results.

Example 2.3.1. Suppose π_i represents normal population $N(\theta_i,\sigma^2)$ for $i = 0,1,\ldots,k$ with θ_0 and σ^2 known. We assume that a random sample of size n_i is taken from each of the k populations $\pi_1,\pi_2,\ldots,\pi_k$. By sufficiency we can restrict our attention to the decision rules depending only on the sample means $x_1,\ldots,x_k$ where x_i has normal distribution with mean θ_i and variance $\sigma_i^2 = \sigma^2/n_i$ for $i = 1,\ldots,k$.

(A) Γ-minimax rule: Application of Theorem 2.3.1 yields the Γ-minimax decision rule δ^* of the type in Theorem 2.3.1 with $x_i-\theta_0$ replacing x_i and d_i being $\sigma_i\max(t_i,0)$ where t_i is determined by

$$\ell_3\gamma_i e^{-2(\lambda_i+\epsilon_i)t_i} + \ell_2\gamma_i'[e^{-2\epsilon_i(t_i-\lambda_i)} + e^{-2\lambda_i(t_i-\epsilon_i)}] + \ell_4(1-\gamma_i-\gamma_i') e^{-2\lambda_i(t_i-\epsilon_i)} \tag{2.3.7}$$

$$- \ell_1\gamma_i = 0$$

where $\lambda_i + \epsilon_i = \Delta_2/\sigma_i$ and $\lambda_i - \epsilon_i = \Delta_1/\sigma_i$ for $i = 1,\ldots,k$.

(B) Minimax rule: We assume $\ell_1 = \ell_2 = \ell_4 = 1$ and $\ell_3 = \ell < 1$. Then a minimax decision rule δ^M is of the type in Corollary 2.3.1 with $x_i - \theta_0$ replacing x_i and d_i being $\sigma_i t_i$ where t_i is determined by

$$\Phi(t_i-\lambda_i-\epsilon_i)+\ell\Phi(-t_i-\lambda_i-\epsilon_i) = \Phi(-t_i-\lambda_i+\epsilon_i) + \Phi(-t_i+\lambda_i-\epsilon_i) \tag{2.3.8}$$

with λ_i and ϵ_i defined as in (A).

2.4 Unknown control

In this section we will consider the case when θ_0 is unknown and will derive a Γ-minimax decision rule δ^* in the class $\mathcal{D}_0$ of decision rules $\delta = (\delta_1,\ldots,\delta_k)$ for which δ_i depends only on X_0 and X_i. We state the following well-known lemma of Ibragimov (1956).

Lemma 2.4.1. The convolution of two strongly unimodal probability density functions is also strongly unimodal.

It follows that the pdf of $Y_i = X_i - X_0$ given by

$$g_i(y-(\theta_i-\theta_0)) = \int_{-\infty}^{\infty} f_i(x+y-\theta_i)f_0(x-\theta_0)dx, \tag{2.4.1}$$

is strongly unimodal and symmetric about the origin. The next result follows from this fact and Lemma 2.3.1.

Lemma 2.4.2. Suppose that a decision rule $\delta(\underline{x})$ is defined by $\delta_i(1|\underline{x}) = I_{(-\infty,-d_i)}(x_i-x_0)$, $\delta_i(2|\underline{x}) = I_{[-d_i,d_i]}(x_i-x_0)$ and $\delta_i(3|\underline{x}) = I_{(d_i,\infty)}(x_i-x_0)$ for $d_i \geq 0$, $i = 1,\ldots,k$.
Then, for $i = 1,\ldots,k$,

$$\sup_{\tau\in\Gamma} r_i(\tau,\delta_i) \leq r_i \tag{2.4.2}$$

where, for $i = 1,\ldots,k$,

$$r_i = \int_{d_i}^{\infty} [\ell_3\gamma_i g_i(y+\Delta_2)+\ell_2\gamma_i'(g_i(y-\Delta_1)+g_i(y+\Delta_1))+\ell_4(1-\gamma_i-\gamma_i')g_i(y+\Delta_1)]dy$$
$$+ \int_{-\infty}^{d_i} \ell_1\gamma_i g_i(y-\Delta_2)dy.$$

We now proceed as in Theorem 2.3.1 by considering the following sequence $\{\tau_n, n = 1,2,\ldots\}$ of prior distributions.
Under τ_n, (i) $\theta_1-\theta_0,\ldots,\theta_k-\theta_0$ are independent,

(ii) $P[\theta_i-\theta_0=\Delta_2] = \gamma_i/2 = P[\theta_i-\theta_0 = -\Delta_2]$,

$P[\theta_i-\theta_0=\Delta_1] = \gamma_i'/2 = P[\theta_i-\theta_0 = -\Delta_1]$,

$P[\theta_i-\theta_0=\Delta_1+ \frac{1}{n}] = (1-\gamma_i-\gamma_i')/2 = P[\theta_i-\theta_0 = -\Delta_1- \frac{1}{n}]$ and

(iii) θ_0 has uniform distribution over $[-n,n]$ and is independent of $\theta_1-\theta_0,\ldots,\theta_k-\theta_0$.

It can be easily shown that the Bayes rule in $\mathcal{D}_0$ w.r.t. τ_n is determined by Bayes rules for component problems and that the Bayes rule for the i-th component problem depends on $\underline{x}$ only through x_i and x_0. It follows from simple computation of the posterior risk of each possible action that the overall risk of the Bayes rule for the i-th component problem is given by

$$\inf_{\delta\in\mathcal{D}_0} r_i(\tau_n,\delta_i) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{1}{4n}\, p_n(s,t)dsdt$$

where $p_n(s,t) = \min\{p_n(1,s,t),\ p_n(2,s,t),\ p_n(1,-s,-t)\}$,

$$p_n(1,s,t) = \ell_2\gamma_i' \int_{-n}^{n}[f_i(s-u-\Delta_1)+f_i(s-u+\Delta_1)]f_0(t-u)du +$$

$$+\ \ell_4(1-\gamma_i-\gamma_i')\int_{-n}^{n} f_i(s-u-\Delta_1-\tfrac{1}{n})f_0(t-u)du +$$

$$+\ (\ell_1+\ell_3)\gamma_i \int_{-n}^{n} f_i(s-u-\Delta_2)f_0(t-u)du \quad \text{and}$$

$$p_n(2,s,t) = \ell_1\gamma_i \int_{-n}^{n}[f_i(s-u+\Delta_2)+f_i(s-u-\Delta_2)]f_0(t-u)du.$$

From change of variables $s = nv-w$ and $t = nv+w$, it follows that

$$\begin{aligned}\inf_{\delta\in\mathcal{D}_0} r_i(\tau_n,\delta_i) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{1}{2}\, p_n(nv-w,\ nv+w)dvdw\\ &\geq \int_{-\infty}^{\infty}\Big[\int_{-1}^{1}\frac{1}{2}\, p_n(nv-w,\ nv+w)dv\Big]dw.\end{aligned}\tag{2.4.3}$$

Note that $h_n(v,w) = p_n(nv-w,\ nv+w)$ can be written as

$$h_n(v,w) = \min\{h_n(1,v,w),\ h_n(2,v,w),\ h_n(1,-v,-w)\}$$

where $$h_n(1,v,w) = \ell_2\gamma_i' \int_{n(v-1)}^{n(v+1)}[f_i(z-w-\Delta_1)+f_i(z-w+\Delta_1)]f_0(z+w)dz +$$

$$+\ \ell_4(1-\gamma_i-\gamma_i')\int_{n(v-1)}^{n(v+1)} f_i(z-w-\Delta_1-\tfrac{1}{n})f_0(z+w)dz +$$

$$+ (\ell_1+\ell_3)\gamma_i \int_{n(v-1)}^{n(v+1)} f_i(z-w-\Delta_2)f_0(z+w)dz \quad \text{and}$$

$$h_n(2,v,w) = \ell_1\gamma_i \int_{n(v-1)}^{n(v+1)} [f_i(z-w+\Delta_2)+f_i(z-w-\Delta_2)]f_0(z+w)dz.$$

Furthermore it is easy to see that, for any $(v,w) \in (-1,1)xR$, $h_n(v,w)$ converges to $h(w) = \min\{h(1,w), h(2,w), h(1,-w)\}$ where

$$h(1,w) = \ell_2\gamma_i' \int_{-\infty}^{\infty} [f_i(z-w-\Delta_1)+f_i(z-w+\Delta_1)]f_0(z+w)dz +$$

$$+ \ell_4(1-\gamma_i-\gamma_i')\int_{-\infty}^{\infty} f_i(z-w-\Delta_1)f_0(z+w)dz +$$

$$+ (\ell_1+\ell_3)\gamma_i \int_{-\infty}^{\infty} f_i(z-w-\Delta_2)f_0(z+w)dz \quad \text{and}$$

$$h(2,w) = \ell_1\gamma_i \int_{-\infty}^{\infty} [f_i(z-w+\Delta_2)+f_i(z-w-\Delta_2)]f_0(z+w)dz.$$

Since $h_n(v,w)$ is bounded above by $h(2,w)$ which is integrable, it follows from (2.4.3) and the Lebesgue convergence theorem that

$$\underline{\lim}_n \inf_{\delta \in \mathcal{D}_0} r_i(\tau_n,\delta_i) \geq \int_{-\infty}^{\infty} h(w)dw$$
$$= \frac{1}{2}\int_{-\infty}^{\infty} \min\{p(1,y),p(2,y),p(1,-y)\}dy \qquad (2.4.4)$$

where $p(1,y) = \ell_2\gamma_i'[g_i(y-\Delta_1)+g_i(y+\Delta_1)]+\ell_4(1-\gamma_i-\gamma_i')g_i(y-\Delta_1) +$

$(\ell_1+\ell_3)\gamma_i g_i(y-\Delta_2)$ and

$$p(2,y) = \ell_1\gamma_i[g_i(y+\Delta_2)+g_i(y-\Delta_2)].$$

The identity in (2.4.4) follows from (2.4.1) and a change of variable. Since $p(1,y) \geq, \leq p(1,-y)$ as $y \geq 0, \leq 0$, (2.4.4) yields that

$$\underline{\lim_n} \inf_{\delta\in\mathcal{D}_0} r_i(\tau_n,\delta_i) \geq \int_0^\infty \min[\ell_3\gamma_i g_i(y+\Delta_2)+\ell_4(1-\gamma_i-\gamma_i')g_i(y+\Delta_1) + \ell_2\gamma_i'(g_i(y+\Delta_1)+g_i(y-\Delta_1)), \ell_1\gamma_i g_i(y-\Delta_2)]dy + \int_{-\infty}^0 \ell_1\gamma_i g_i(y-\Delta_2)dy. \tag{2.4.5}$$

Now consider a decision rule δ^* defined by $\delta_i^*(1|\underline{x}) = I_{(-\infty,-d_i)}(x_i-x_0)$, $\delta_i^*(2|\underline{x}) = I_{[-d_i,d_i]}(x_i-x_0)$ and $\delta_i^*(3|\underline{x}) = I_{(d_i,\infty)}(x_i-x_0)$ for $i=1,\ldots,k$ where each d_i is determined by $d_i = \max(c_i,0)$ so that

$$\ell_3\gamma_i g_i(y+\Delta_2)+\ell_2\gamma_i'(g_i(y-\Delta_1)+g_i(y+\Delta_1))+\ell_4(1-\gamma_i-\gamma_i')g_i(y+\Delta_1) \leq, \geq \ell_1\gamma_i g_i(y-\Delta_2) \text{ as } y \geq, \leq c_i. \tag{2.4.6}$$

From Lemma 2.4.1, Lemma 2.4.2, (2.4.5) and (2.4.6) it follows that

$$\underline{\lim_n} \inf_{\delta\in\mathcal{D}_0} r_i(\tau_n,\delta_i) \geq \sup_{\tau\in\Gamma} r_i(\tau,\delta_i^*),$$

and this in turn implies that

$$\begin{aligned}\overline{\lim_n} \inf_{\delta\in\mathcal{D}_0} r(\tau_n,\delta) &= \overline{\lim_n} \sum_{i=1}^k \inf_{\delta\in\mathcal{D}_0} r_i(\tau_n,\delta_i)\\ &\geq \sum_{i=1}^k \underline{\lim_n} \inf_{\delta\in\mathcal{D}_0} r_i(\tau_n,\delta_i)\\ &\geq \sum_{i=1}^k \sup_{\tau\in\Gamma} r_i(\tau,\delta_i^*)\\ &\geq \sup_{\tau\in\Gamma} r(\tau,\delta^*).\end{aligned}$$

Lemma 2.2.1 then yields the next result.

Theorem 2.4.1. Assume that independent random variables $X_0,\ldots,X_k$ have strongly unimodal symmetric pdf's $f_0(x_0-\theta_0),\ldots,f_k(x_k-\theta_k)$, respectively, and that the loss function is given by (2.2.1) and (2.2.2). Then the Γ-minimax decision rule δ^* in $\mathcal{D}_0$ is given by $\delta_i^*(1|\underline{x}) = I_{(-\infty,-d_i)}(x_i-x_0)$, $\delta_i^*(2|\underline{x}) = I_{[-d_i,d_i]}(x_i-x_0)$ and $\delta_i^*(3|\underline{x}) = I_{(d_i,\infty)}(x_i-x_0)$ for $i = 1,\ldots,k$ where $d_i = \max(c_i,0)$ with c_i being determined by (2.4.6).

Remark 2.4.1. Note that the symmetry of $f_i(\cdot)$ can be replaced by the symmetry of $g_i(\cdot)$ for Theorem 2.4.1 to hold. It can be easily shown that the relation (2.4.4) holds without the assumption of the symmetry of $f_i(\cdot)$. Then all the steps after (2.4.4) still remain true provided $g_i(\cdot)$ is symmetric. It should also be pointed out that the symmetry of $g_i(\cdot)$ follows when $f_0(\cdot), f_1(\cdot),\ldots,f_k(\cdot)$ are identical.

The next result follows in exactly the same manner as Corollary 2.3.1 was proved..

Corollary 2.4.1. Under the assumptions in Theorem 2.4.1, if $\ell_1 = \ell_2 = \ell_4 = 1$ and $\ell_3 = \ell < 1$, then a minimax decision rule δ^M in $\mathcal{D}_0$ is given by $\delta_i^M(1|\underline{x}) = I_{(-\infty,-d_i)}(x_i-x_0)$, $\delta_i^M(2|\underline{x}) = I_{[-d_i,d_i]}(x_i-x_0)$ and $\delta_i^M(3|\underline{x}) = I_{(d_i,\infty)}(x_i-x_0)$ for $i = 1,\ldots,k$ where $d_i > 0$ is determined so that, for $G_i(x) = \int_{-\infty}^{x} g_i(t)dt$,

$$G_i(d_i-\Delta_2)+\ell G_i(-d_i-\Delta_2) = G_i(-d_i-\Delta_1)+G_i(-d_i+\Delta_1).$$

We illustrate the application of the above results by the following examples.

Example 2.4.1. Consider the same problem in Example 2.3.1 except that θ_0 is unknown. In this case a random sample of size n_0 is also taken from the control π_0. Let $X_0, X_1, \ldots, X_k$ denote the sample means corresponding to $\pi_0, \ldots, \pi_k$, respectively. Here, $\sigma_0^2 = \sigma^2/n_0$.

(A) Γ-minimax rule: The Γ-minimax rule δ^* in $\mathcal{D}_0$ given in Theorem 2.4.1 is determined by $d_i = (\sigma_i^2+\sigma_0^2)^{\frac{1}{2}} \max(t_i, 0)$ where t_i satisfies (2.3.7) with $\lambda_i + \epsilon_i = \Delta_2(\sigma_i^2+\sigma_0^2)^{-\frac{1}{2}}$ and $\lambda_i - \epsilon_i = \Delta_1(\sigma_i^2+\sigma_0^2)^{-\frac{1}{2}}$.

(B) Minimax rule: Assume $\ell_1 = \ell_2 = \ell_4 = 1$ and $\ell_3 = \ell < 1$. Then the minimax rule δ^M in $\mathcal{D}_0$ given in Corollary 2.4.1 is determined by $d_i = (\sigma_i^2+\sigma_0^2)^{\frac{1}{2}} t_i$ with t_i satisfying (2.3.8) where λ_i and ϵ_i are defined as in the above. This offers a partial solution of the problem in Section 5.1 of Bhattacharyya (1956).

Example 2.4.2. Assume that π_i represents normal population $N(0, \sigma_i^2)$ for $i = 0, 1, \ldots, k$ with σ_i^2 unknown, and that we have a random sample of size n taken from each population π_i. Consider a problem of partitioning the treatment populations in terms of variances $\sigma_1^2, \ldots, \sigma_k^2$ with a loss structure analogous to that given by (2.2.1) and (2.2.2), i.e. a loss function obtained from the latter by substituting $\log \sigma_i^2$, $\log \Delta_1$ and $\log \Delta_2$ for θ_i, Δ_1 and Δ_2, respectively. Thus Δ_1 and Δ_2 here are assumed such that $1 < \Delta_1 < \Delta_2$. By sufficiency we need to consider only the decision rules depending on $S_1^2, \ldots, S_k^2$ where S_i^2 denotes the sample variance corresponding to

π_i. Since $nS_i^2/\sigma_i^2 (i = 0,1,\ldots,k)$ are independently distributed chi-square random variables with degrees of freedom n, it can be easily seen that the associated location parameter problem satisfies the assumptions in Theorem 2.4.1 except the symmetry which are not necessary in this problem because of Remark 2.4.1. Therefore, with obvious modifications we have the following results.

Let $\mathcal{D}_0$ denote the class of decision rules $\delta = (\delta_1,\ldots,\delta_k)$ for which δ_i depends only on s_0^2 and s_i^2, and let T_i denote s_i^2/s_0^2 for $i = 1,\ldots,k$.

(A) <u>Γ-minimax rule</u>: A Γ-minimax rule δ^* in $\mathcal{D}_0$ is given b $\delta_i^*(1|T_i) = I_{(0,d_i^{-1})}(T_i)$, $\delta_i^*(2|T_i) = I_{[d_i^{-1},d_i]}(T_i)$ and $\delta_i^*(3|T_i) = I_{(d_i,\infty)}(T_i)$ for $i = 1,\ldots,k$. With each d_i being determined by $d_i = \max(c_i,1)$ so that

$$\ell_3\gamma_i\left(\frac{\Delta_2+y}{1+\Delta_2 y}\right)^n+\ell_2\gamma_i'\left[\left(\frac{\Delta_2+y}{\Delta_1+y}\right)^n+\left(\frac{\Delta_2+y}{1+\Delta_1 y}\right)^n\right]\left(\frac{\Delta_1}{\Delta_2}\right)^{\frac{n}{2}}+\ell_4(1-\gamma_i-\gamma_i')\left(\frac{\Delta_2+y}{1+\Delta_1 y}\right)^n\left(\frac{\Delta_1}{\Delta_2}\right)^{\frac{n}{2}} \tag{2.4.7}$$

$$\le, \ge \ell_1\gamma_i \quad \text{as} \quad y \ge, \le c_i.$$

(B) <u>Minimax rule</u>: Assume $\ell_1 = \ell_2 = \ell_4 = 1$ and $\ell_3 = \ell < 1$. Then the minimax rule δ^M in $\mathcal{D}_0$ is the same as δ^* in (A) except that $d_i = d$ for $i = 1,\ldots,k$ where d is determined so that

$$\text{(2.4.8)} \quad G_n(d/\Delta_2)+\ell[1-G_n(d\Delta_2)] = G_n(\Delta_1/d)+1-G_n(d\Delta_1)$$

where G_n denotes the distribution function of F-distribution with degrees of freedom n and n.

We note that if π_i represents $N(\mu_i,\sigma_i^2)$ with both μ_i and σ_i^2 unknown, then the above results still hold with n-1 replacing n where s_i^2 in such a case is the best unbiased estimator of σ_i^2.

2.5 Comparison of Γ-minimax rules with Bayes rules

When we represent our a priori information about the parameters by prior distributions over the parameter space, one method for the use of such information is to find a rule which is Γ-minimax with respect to the class of such priors. Another way is to select one of such priors and use the corresponding Bayes rule. Thus Bayes rules wrt prior distributions in Γ are natural competitors for a Γ-minimax rule.

In this section we consider k+1 normal populations $N(\theta_i,\sigma^2)$ with σ^2 known, and will derive Bayes rules wrt a normal prior and then we compare them with the corresponding Γ-minimax rules from both points of view. Assume that $(\theta_0,\ldots,\theta_k)$ have prior distribution τ_0 under which

(i) $\theta_0,\ldots,\theta_k$ are independent and (2.5.1)

(ii) each θ_i has a (marginal) normal distribution with mean μ_i and variance v_i^2.

Let $x_0,\ldots,x_k$ denote the sample means based on samples of size n_i $(i = 0,1,\ldots,k)$. To simplify forthcoming formulae, let us introduce the following notations.

$$\sigma_i^2 = \sigma^2/n_i,\ b_i = [(\sigma_i^{-2}+v_i^{-2})^{-1}+(\sigma_0^{-2}+v_0^{-2})^{-1}]^{\frac{1}{2}},$$

$$\lambda_i+\epsilon_i = \Delta_2/b_i, \quad \lambda_i-\epsilon_i = \Delta_1/b_i \tag{2.5.2}$$

$$m_i = (\sigma_i^{-2}x_i+v_i^{-2}\mu_i)(\sigma_i^{-2}+v_i^{-2})^{-1}, \; y_i = (m_i-m_0)/b_i.$$

The following theorem describes the Bayes rule.

Theorem 2.5.1. Assume the prior τ_0 as specified in (2.5.1). Then the Bayes rule δ^B wrt τ is given by $\delta_i^B(1|\underline{y}) = I_{(-\infty,-d_i)}(y_i)$, $\delta_i^B(2|y) = I_{[-d_i,d_i]}(y_i)$ and $\delta_i^B(3|y) = I_{(d_i,\infty)}(y_i)$ for $i = 1,\ldots,k$ where each $d_i = \max(c_i,0)$ is determined so that

$$\begin{aligned}&\ell_3\Phi(-\lambda_i-\epsilon_i-y) + \ell_4[\Phi(-\lambda_i+\epsilon_i-y) - \Phi(-\lambda_i-\epsilon_i-y)]\\ &+ \ell_2[\Phi(\lambda_i-\epsilon_i-y)-\Phi(-\lambda_i+\epsilon_i-y)] - \ell_1\Phi(-\lambda_i-\epsilon_i+y)\\ &\geq 0, \leq 0 \quad \text{as} \quad y \leq c_i, \; y \geq c_i\end{aligned} \tag{2.5.3}$$

with Φ denoting the distribution function of N(0,1) distribution.

Proof. It suffices to find the Bayes rule for each of the k component problems. This reduces to comparison of posterior risks of three possible actions. Let $p_1(y_i)$, $p_2(y_i)$ and $p_3(y_i)$ denote the posterior risks of action 1,2 and 3, respectively, in the i-th component problem, then it can be shown that

$$\begin{aligned}p_1(y) &= (\ell_1+\ell_3)\Phi(-\lambda_i-\epsilon_i+y)+\ell_4[\Phi(\lambda_i+\epsilon_i-y) - \Phi(\lambda_i-\epsilon_i-y)]+\\ &\quad + \ell_2[\Phi(\lambda_i-\epsilon_i-y) - \Phi(-\lambda_i+\epsilon_i-y)],\\ p_2(y) &= \ell_1[\Phi(-\lambda_i-\epsilon_i-y) + \Phi(-\lambda_i-\epsilon_i+y)] \text{ and}\\ p_3(y) &= p_1(-y).\end{aligned}$$

Note that $p_1(y)-p_3(y)$ can be written as $E_y H(z)$ where Z has normal distribution with mean y and variance 1 and $H(\cdot)$ is given by

$$H(z) = \begin{cases} \ell_1+\ell_3 & \text{if } z \geq \lambda_i+\epsilon_i \\ \ell_4 & \text{if } \lambda_i-\epsilon_i < z < \lambda_i+\epsilon_i \\ 0 & \text{if } -\lambda_i+\epsilon_i \leq z \leq \lambda_i-\epsilon_i \\ -\ell_4 & \text{if } -\lambda_i-\epsilon_i < z < -\lambda_i+\epsilon_i \\ -(\ell_1+\ell_3) & \text{if } z \leq -\lambda_i-\epsilon_i. \end{cases}$$

Since the density of normal distribution $N(y,1)$ has MLR property, it follows that $p_1(y)-p_3(y)$ has at most one change of sign. Furthermore, it can be shown that $p_1(y)-p_3(y)$ is increasing on $(-\epsilon_i,\epsilon_i)$ and $p_1(0)-p_3(0) = 0$. Thus $p_1(y)-p_3(y) \geq 0, \leq 0$ as $y \geq 0$, $y \leq 0$. Similarly, we can show that $p_3(y)-p_2(y) \geq 0, \leq 0$ as $y \leq c_i$, $y \geq c_i$ for some real number c_i unless $p_3(y)-p_2(y) \leq 0$ for all y. Therefore the result follows.

Thus comparison can be made between the Γ-minimax rule δ^* given in Example 2.4.1 and the Bayes rule δ^B given in Theorem 2.5.1 under the relations $\gamma_i = \Phi[(-\Delta_2+\mu_i-\mu_0)(v_i^2+v_0^2)^{-\frac{1}{2}}]+\Phi[(-\Delta_2-\mu_i+\mu_0)(v_i^2+v_0^2)^{-\frac{1}{2}}]$ and $\gamma_i' = \Phi[(\Delta_1-\mu_i+\mu_0)(v_i^2+v_0^2)^{-\frac{1}{2}}]-\Phi[(-\Delta_1-\mu_i+\mu_0)(v_i^2+v_0^2)^{-\frac{1}{2}}]$. We compare these rules under the assumption that $\ell_1 = \ell_2 = \ell_4 = 1$, $\ell_3 = \ell$, $n_i = n$ and $v_i^2 = v_0^2$ for $i = 1,\ldots,k$. There are two ways of any meaningful comparison of these rules. One way is to examine the increase in overall risk wrt τ_0 resulting from the use of the Γ-minimax rule. Another way is to compare these rules in terms of $\sup_{\tau\in\Gamma} r(\tau,\delta)$. When

$n_i = n$ and $v_i^2 = v_0^2$, the Bayes rule depends on $\underline{x}$ only through $x_1-x_0,\ldots,x_k-x_0$ and it can be shown that $\sup_{\tau\in\Gamma} r(\tau,\delta^B) = \sum_{i=1}^{k} \sup_{\tau\in\Gamma} r_i(\tau,\delta_i^B)$. Thus it suffices to compare these rules with respect to classification of one population. We choose π_1 for this purpose without loss of generality.

Now we introduce the parameters used in the comparison as follows.

$$\beta_1 = \frac{v_0^2}{\sigma_0^2} = \frac{nv_0^2}{\sigma^2},\ \beta_2 = \frac{\Delta_2+\Delta_1}{2\sqrt{2v_0^2}},\ \beta_3 = \frac{\Delta_2-\Delta_1}{2\sqrt{2v_0^2}} \text{ and } \beta_4 = \frac{\mu_1-\mu_0}{\sqrt{2v_0^2}}.$$

The overall risk wrt τ_0 of these rules can be written as

$$\begin{aligned}
&\ell\Phi(-A-B-C) + \Phi(A-B-C) + \Phi(D-E) + \ell\Phi(-D-E)\\
&- \Phi_0(-A-B-C,-D-E;\rho) + (1-\ell)\Phi_0(-A-B-C, D-E;\rho)\\
&- \Phi_0(-A+B-C,D-E;\rho) + \Phi_0(A-B-C,-D-E;\rho)\\
&- \Phi_0(-A+B-C,-D-E;\rho)-\Phi_0(A-B-C,D-E;\rho)\\
&+ \Phi_0(-A+B-C,D-E;\rho)-\Phi_0(A-B-C,-D-E;\rho)\\
&- \Phi_0(A+B-C,D-E;\rho)+(1-\ell)\Phi_0(A+B-C,-D-E;\rho)
\end{aligned}$$

where $\Phi_0(\cdot,\cdot;\rho)$ is the cdf of a bivariate normal distribution with zero means, unit variances and correlation coefficient ρ, $A=\beta_2$, $B=\beta_3$, $C = \beta_4$, $\rho = \beta_1^{\frac{1}{2}}(1+\beta_1)^{-\frac{1}{2}}$ for both δ^B and δ^*, $D = d_1\beta_1^{-\frac{1}{2}}$ for δ^B, $D = \max(t_1,0)(1+\beta_1)^{-\frac{1}{2}}$ for δ^*, $E = \rho^{-1}\beta_4$ for δ^B and $E = \rho\beta_4$ for δ^* with d_1 and t_1 being those in Theorem 2.5.1 and in Example 2.4.1. Also $\sup_{\tau\in\Gamma} r_1(\tau,\delta_1)$ for both rules can be written as

$$\gamma_1[\Phi(R+|S|-T-U)-\Phi(-R+|S|-T+U)+\ell\Phi(-R+|S|-T-U)]$$

$$+\ \gamma_1'[\Phi(-R-S+T-U)+\Phi(-R+S-T+U)]^{\vee}[\Phi(-R+S+T-U)+\Phi(-R-S-T+U)]$$

$$+\ (1-\gamma_1')\Phi(-R+|S|-T+U)$$

where $x^{\vee}y = \max(x,y)$, $T = \beta_2\beta_1^{\frac{1}{2}}$, $U = \beta_3\beta_1^{\frac{1}{2}}$ for both δ^B and δ^*, $S = \beta_4\beta_1^{-\frac{1}{2}}$ for δ^B, $S = 0$ for δ^*, $R = \beta_1^{-\frac{1}{2}}(1+\beta_1)^{\frac{1}{2}}d_1$ for δ^B, $R = \max(t_1,0)$ with d_1 and t_1 being those in Theorem 2.5.1 and in Example 2.4.1. Note that y_1, λ_1 and ϵ_1 in Theorem 2.5.1 can be written as $y_1 = \dfrac{\sqrt{\beta_1}}{\sqrt{1+\beta_1}}\dfrac{x_1-x_0}{\sqrt{2}\ \sigma/\sqrt{n}} + \dfrac{\beta_4}{\sqrt{1+\beta_1}}$, $\lambda_1 = \beta_2\sqrt{1+\beta_1}$ and $\epsilon_1 = \beta_3\sqrt{1+\beta_1}$. Thus the constant d_1 does not change for the different values of β_4 when β_1,β_2 and β_3 are fixed. Particularly, if $\ell = 1$, then d_1 is easily found to be $\beta_2\sqrt{1+\beta_1}$ which does not vary for different values of β_3. Table IV and Table V give $r_1(\tau_0,\delta_1)$ and $\sup_{\tau\in\Gamma} r_1(\tau,\delta_1)$ for $\delta_1 = \delta_1^B$, δ_1^* for $\ell = 0$ and $\ell = 1$, respectively. As by products they also provide the constants to implement these rules. It can be observed from these tables that the increase in overall risk wrt τ_0 from the use of δ^* is only slight compared to that in $\sup_{\tau\in\Gamma} r_1(\tau,\delta_1)$ from the use of δ^B. In this sense, δ^* is more robust against other formulation. As it can be expected, such a robustness of δ^* becomes more prominent as the difference between the prior means (β_4) increases and the prior variance (β_1) gets smaller. When the prior variance is large and we have the same prior means, both rules compare favorably with each other. In many cases, we can observe that the Γ-minimax decision rule compares favorably with the given Bayes rule in terms of overall risk.

TABLE IV

Lists the overall risks and the values of $\sup_{\tau \in \Gamma} r(\tau,\delta)$ of Bayes rules and Γ-minimax rules when $\ell=0$.

Entries	γ_1 / γ_1'	t_1^+	$r_1(\tau_0,\delta_1^*)$ / $\sup_{\tau\in\Gamma} r_1(\tau,\delta_1^*)$	d_1	$r_1(\tau_0,\delta_1^B)$ / $\sup_{\tau\in\Gamma} r_1(\tau,\delta_1^B)$	t_1^+	$r(\tau_0,\delta_1^*)$ / $\sup_{\tau\in\Gamma} r_1(\tau,\delta_1^*)$	d_1	$r_1(\tau_0,\delta_1^B)$ / $\sup_{\tau\in\Gamma} r_1(\tau,\delta_1^B)$	t_1^+	$r_1(\tau_0,\delta_1^*)$ / $\sup_{\tau\in\Gamma} r_1(\tau,\delta_1^*)$	d_1	$r_1(\tau_0,\delta_1^B)$ / $\sup_{\tau\in\Gamma} r_1(\tau,\delta_1^B)$
β_4 \ β_1	.25					1.0				4.0			
	$\beta_2 = 0.5$					$\beta_3 = 0.25$							
1.0	.6388	.0000	.3590	.4239	.2173	.0000	.2488	.6488	.1652	.4211	.1199	1.1142	.0783
	.1210		.4552		.9648		.3622		.7339		.2184		.4203
0.5	.5069	.7865	.3935		.3361	.6156	.2614		.2351	.7703	.1276		.1067
	.1747		.4680		.8022		.3841		.5554		.2370		.3140
0.0	.4533	1.3128	.3924		.3852	.8832	.2537		.2536	.9151	.1285		.1181
	.1974		.4386		.4928		.3729		.3879		.2365		.2372
	$\beta_2 = 0.5$					$\beta_3 = 0.125$							
1.0	.6983	.0000	.3930	.3807	.2587	.0000	.2864	.6231	.2097	.0933	.2001	1.1101	.1189
	.1814		.4961		.9695		.4097		.7734		.2816		.5404
0.5	.5806	.5402	.4520		.3957	.4107	.3321		.2963	.5105	.2121		.1615
	.2595		.5344		.8274		.4577		.6201		.3312		.4359
0.0	.5320	1.1559	.4599		.4544	.7288	.3341		.3315	.6963	.2103		.1785
	.2923		.5164		.5527		.4568		.4626		.3413		.3492
	$\beta_2 = 0.5$					$\beta_3 = 0.05$							
1.0	.7342	.0000	.4140	.3462	.2853	.0000	.3103	.6015	.2395	.0000	.2413	1.1057	.1493
	.2176		.5250		.9731		.4471		.3117		.3261		.6276
0.5	.6269	.3366	.4898		.4330	.2363	.3840		.3368	.2637	.2899		.2021
	.3090		.5768		.8460		.5038		.5703		.3999		.5306
0.0	.5823	1.0509	.5027		.4978	.6091	.3828		.3764	.4765	.2856		.2232
	.3473		.5680		.5937		.5170		.5190		.4210		.4391

* $t_1^+ = \max(t_1,0)$ in Example 2.4.1 and d_1 is the constant for implementing the Bayes rule in Theorem 2.5.1.

Table IV: ℓ = 0 (continued)

β_4 \ β_1	.25					1.0				4.0			
				β_2 = 0.8			β_3 = 0.4						
1.0	.4346	1.0534	.3188	.8760	.2111	.8964	.1739	1.1279	.1279	1.2911	.0614	1.7888	.0439
	.1935		.3908		.9486		.2805		.6152		.1288		.2510
0.5	.2865	2.1529	.2670		.2326	1.4814	.1633		.1475	1.6783	.0578		.0515
	.2761		.2829		.7468		.2350		.4222		.1245		.1718
0.0	.2301	2.6845	.2247		.2191	1.7657	.1525		.1506	1.8620	.0548		.0537
	.3108		.2294		.4411		.2028		.2754		.1157		.1209
				β_2 = 0.8			β_3 = 0.2						
1.0	.5228	.8653	.3909	.8639	.2824	.6857	.2539	1.1241	.1931	1.0265	.1369	1.7888	.0923
	.2898		.4770		.9598		.3793		.7074		.2574		.4408
0.5	.3753	2.3613	.3594		.3163	1.5336	.2424		.2237	1.7230	.1152		.1083
	.4042		.3739		.7920		.3450		.5344		.2692		.3449
0.0	.3173	3.1771	.3151		.3050	2.0022	.2404		.2298	2.0544	.1133		.1131
	.4515		.3172		.5226		.3066		.3794		.2546		.2667
				β_2 = 0.8			β_3 = 0.08						
1.0	.5778	.7448	.4380	.8527	.3309	.5047	.3190	1.1200	.2411	.6070	.2407	1.7888	.1353
	.3470		.5365		.9699		.4541		.7840		.3573		.5860
0.5	.4358	3.0492	.4311		.3742	1.8475	.3137		.2801	1.8908	.1632		.1588
	.4758		.4357		.8308		.4278		.6286		.3960		.4890
0.0	.3789	4.9653	.3789		.3658	2.9800	.3509		.2887	2.6419	.1939		.1659
	.5285		.3789		.5827		.3785		.4658		.3679		.3967

TABLE V

Lists the overall risks and the values of $\sup_{\tau \in \Gamma} r(\tau,\delta)$ of Bayes rules and Γ-minimax rules when $\ell=1$.

Entries	γ_1 / γ_1'	t_1^+	$r_1(\tau_0,\delta_1^*)$ / $\sup_{\tau\in\Gamma} r_1(\tau,\delta_1^*)$	d_1	$r_1(\tau_0,\delta_1^B)$ / $\sup_{\tau\in\Gamma} r_1(\tau,\delta_1^B)$	t_1^+	$r_1(\tau_0,\delta_1^*)$ / $\sup_{\tau\in\Gamma} r_1(\tau,\delta_1^*)$	d_1	$r_1(\tau_0,\delta_1^B)$ / $\sup_{\tau\in\Gamma} r_1(\tau,\delta_1^B)$	t_1^+	$r_1(\tau_0,\delta_1^*)$ / $\sup_{\tau\in\Gamma} r_1(\tau,\delta_1^*)$	d_1	$r_1(\tau_0,\delta_1^B)$ / $\sup_{\tau\in\Gamma} r_1(\tau,\delta_1^B)$
β_4 \ β_1			.25				1.0				4.0		
						$\beta_2 = 0.5$		$\beta_3 = 0.25$					
1.0	.6388	.9905	.4570	.5590	.2422	.5867	.2438	.7071	.1715	.5504	.1110	1.1180	.0784
	.1210		.5930		1.5502		.4532		.9926		.2333		.4353
0.5	.5069	1.6237	.4460		.3626	.9486	.2689		.2427	.8264	.1244		.1069
	.1747		.4963		1.1276		.4152		.6692		.2425		.3204
0.0	.4533	1.9515	.4230		.4043	1.1319	.2733		.2714	.9538	.1268		.1183
	.1974		.4488		.6070		.3906		.4329		.2398		.2401
						$\beta_2 = 0.$		$\beta_3 = 0.125$					
1.0	.6983	1.0867	.5218	.5590	.2929	.5941	.2947	.7071	.2195	.4229	.1674	1.1180	.1194
	.1814		.6655		1.6167		.5435		1.0979		.3258		.5696
0.5	.5806	1.7899	.5260		.4325	.9767	.3361		.3082	.6898	.1941		.1620
	.2595		.5746		1.2157		.5098		.7758		.3496		.4498
0.0	.5320	2.1609	.5085		.4817	1.1782	.3474		.3437	.8297	.1985		.1791
	.2923		.5299		.6991		.4865		.5298		.3529		.3560
						$\beta_2 = 0.5$		$\beta_3 = 0.05$					
1.0	.7342	1.2017	.5688	.5590	.3264	.6241	.3280	.7071	.2523	.3634	.2059	1.1180	.1500
	.2176		.7117		1.6582		.6084		1.1721		.4027		.6698
0.5	.6269	2.0195	.5872		.4778	1.0508	.3828		.3522	.6180	.2445		.2030
	.3090		.6242		1.2762		.5775		.[illegible]566		.4372		.5520
0.0	.5823	2.4774	.5683		.5316	1.2912	.4017		.[illegible]323	.7657	.2510		.2242
	.3473		.5817		.7627		.5545		.6032		.4452		.4502

Table V: ℓ = 1 (continued)

β_4 \ β_1	.25					1.0				4.0			
					β_2 = 0.8		β_3 = 0.4						
1.0	.4346	1.3771	.3387	.8944	.2143	.9763	.1737	1.1314	.1282	1.2921	.0614	1.7889	.0439
	.1935		.4057		1.3140		.2876		.6840		.1288		.2511
0.5	.2865	2.2654	.2698		.2344	1.5052	.1640		.1477	1.6785	.0578		.0515
	.2761		.2836		.8885		.2360		.4410		.1246		.1719
0.0	.2301	2.7491	.2253		.2198	1.7785	.1529		.1508	1.8621	.0548		.0537
	.3108		.2295		.4753		.2031		.2802		.1157		.1209
					β_2 = 0.8		β_3 = 0.2						
1.0	.5228	1.4395	.4248	.8944	.2883	.8959	.2465	1.1314	.1938	1.0409	.1355	1.7889	.0923
	.2898		.5021		1.4114		.3982		.8181		.2580		.4413
0.5	.3753	2.5823	.3642		.3195	1.5998	.2448		.2243	1.7242	.1151		.1083
	.4042		.3745		.9929		.3470		.5702		.2693		.3451
0.0	.3173	3.2996	.3157		.3062	2.0341	.2420		.2303	2.0548	.1133		.1131
	.4515		.3172		.5774		.3070		.3900		.2546		.2667
					β_2 = 0.8		β_3 = 0.08						
1.0	.5778	1.6000	.4891	.8944	.3393	.8909	.2966	1.1314	.2423	.7167	.2239	1.7889	.1353
	.3470		.5665		1.4765		.4861		.9276		.3617		.5872
0.5	.4358	3.3848	.4337		.3789	1.9689	.3222		.2811	1.8946	.1632		.1588
	.4758		.4357		1.0748		.4289		.6799		.3960		.4894
0.0	.3789	5.0783	.3789		.3675	3.0076	.3520		.2895	2.6422	.1939		.1659
	.5285		.3789		.6540		.3785		.4823		.3679		.3969

CHAPTER 3

SELECTION PROCEDURES FOR A PROBLEM IN RELIABILITY AND FOR A PROBLEM OF SCALE PARAMETERS

3.1 Introduction

In this chapter we discuss two selection problems, one arising in reliability theory and another for symmetric scale parameter populations. The first problem deals with an ℓ-out-of-m system where m components are to be placed and at least ℓ of them should function to make the system work. In many situations several brands (populations) of components are available from which we need to choose m components for a system. Note that it is allowed to draw more than one component from a population. It is assumed that the lifelength of a component from population π_i is exponentially distributed with mean λ_i^{-1} for $i = 1,\ldots,k$ and that the components in the system are statistically independent. Broström (1977) considered the 1-out-of-2 system when only two populations are available. He assumed a loss function depending on (λ_1,λ_2) only through λ_1/λ_2 so that the problem is invariant under the scale transformation, and then studied the admissibility and the minimaxity of some rules. In Section 3.2, we consider two cases; (A) m-out-of-m, i.e., series system and (B) 1-out-of-2 system when $k(k \geq 2)$ populations are available. We assume a loss function inversely proportional to the expected lifelength of the system, and derive a uniformly best decision rule among the permutation invariant rules for the series system and a Bayes rule wrt a natural conjugate prior for the 1-out-of-2 system. Tables to implement the Bayes rule are provided at the end of this chapter.

The second part of this chapter consists of the investigation of the selection procedures for scale parameters of symmetric distributions. Contributions for this problem have been made by Puri and Puri (1969), Blumenthal and Patterson (1969), Gupta and McDonald (1970), Bhapkar and Gore (1971) and McDonald (1977) among others. Since ranking the populations with regard to scale parameters is equivalent to ranking them in terms of measures of dispersion, we can consider selection procedures based on estimators of measures of dispersion. In Section 3.3, we consider two problems; (A) selection of the t 'best' populations under the indifference-zone approach, (B) selecting a subset containing the 'best' population. We consider selection procedures based on the p-th power sample deviations and on the trimmed standard deviations for problem (A), and derive a large sample solution of the sample size required by the basic probability condition. Asymptotic relative efficiencies of the procedures in (B) are shown to be the same as those in (A) as well as those of the corresponding estimators of Bickel and Lehmann (1976).

3.2 Some selection problems arising in reliability theory

Let $\pi_1,\ldots,\pi_k$ $(k \geq 2)$ denote the available populations (brands) and assume that each component from π_i has an exponentially distributed lifelength with mean lifelength λ_i^{-1} for $i = 1,\ldots,k$. Based on n independent observations $X_{i1},\ldots,X_{in}$ from each π_i $(i = 1,\ldots,k)$, we want to find an 'optimal' selection rule assuming a suitable loss function. By sufficiency the problem can be reduced to the one based on $X_1,\ldots,X_k$ where $X_i = \sum_{j=1}^{n} X_{ij}$ has gamma distribution with mean $n\lambda_i^{-1}$ and variance $n\lambda_i^{-2}$. Throughout this section let $x_{[1]} \leq \ldots \leq x_{[k]}$ denote the ordered observations $x_1,\ldots,x_k$, and $\pi_{(i)}$ and $\lambda_{(i)}$ denote the π and the λ associated with $x_{[i]}$ for $i = 1,\ldots,k$. Given $\underline{x} = (x_1,\ldots,x_k)$, the posterior risk of a decision rule δ will be denoted by $r(\delta,\underline{x})$.

(A) The series system

Here we denote the action space by $\mathcal{A} = \{(i_1,\ldots,i_m): 1 \leq i_1 \leq \ldots \leq i_m \leq k\}$ where $(i_1,\ldots,i_m)$ is to be interpreted as the action of drawing the j-th component from π_{i_j} for $j = 1,\ldots,m$. For the series system the expected lifelength of the system for the action $(i_1,\ldots,i_m)$ is easily seen to be $(\sum_{j=1}^{m} \lambda_{i_j})^{-1}$. We will consider a loss function which is inversely proportional to the expected lifelength corresponding to an action, i.e., loss function is assumed to be

$$L(\underline{\lambda}, (i_1,\ldots,i_m)) = \sum_{j=1}^{m} \lambda_{i_j}. \tag{3.2.1}$$

Thus the posterior risk of a decision rule δ, which leads to an action $(i_1,\ldots,i_m) \in \mathcal{A}$ with probability 1 is given by

$$r(\delta,\underline{x}) = E[\sum_{j=1}^{m} \lambda_{i_j} | \underline{x}]. \tag{3.2.2}$$

Let $N_s = \{\underline{n}_s = (n_1,\ldots,n_s)\colon n_1 \geq \ldots \geq n_s \geq 1, \sum_{j=1}^{s} n_j = m, n_i$'s are integers$\}$ and let $a \wedge b = \min(a,b)$. The next result leads to considerable reduction in the number of decision rules to be compared for the Bayes rule when the prior of $\underline{\lambda}$ is assumed to be permutationally symmetric.

Lemma 3.2.1. If the prior distribution of $\underline{\lambda}$ is permutationally symmetric on $(0,\infty)^k$, then the Bayes rule δ^* is determined by

$$r(\delta^*,\underline{x}) = \min_{1\leq s\leq k\wedge m} \ \min_{\underline{n}_s\in N_s} r(\delta_{\underline{n}_s},\underline{x}) \tag{3.2.3}$$

where $\delta_{\underline{n}_s}$ selects an action of drawing n_j components from $\pi_{(k-j+1)}$ for $j = 1,\ldots,s$.

Proof. For $s = 1,\ldots,k \wedge m$, let us define $\mathcal{A}_s^*$ to be $\mathcal{A}_s^* = \{(i_1,\ldots,i_s; n_1,\ldots,n_s)\colon \underline{n}_s \in N_s, 1 \leq i_j \leq k, i_j \neq i_{j'}$ for $j \neq j'\}$ where $(i_1,\ldots,i_s; n_1,\ldots,n_s)$ is interpreted as drawing n_j components from π_{i_j} $(j = 1,\ldots,s)$. Note that the action space $\mathcal{A}$ can be partitioned into $k\wedge m$ components $\mathcal{A}_s$ $(s = 1,\ldots,k\wedge m)$ where we should choose s different populations for m components. Again, $\mathcal{A}_s$ can be written as

$$\mathcal{A}_s = \bigcup_{\underline{n}_s\in N_s} \mathcal{A}_{\underline{n}_s} \text{ where } \mathcal{A}_{\underline{n}_s} = \{(i_1,\ldots,i_s)\colon (i_1,\ldots,i_s; n_1,\ldots,n_s)\in \mathcal{A}_s^*\}.$$

The loss function given by (3.2.1) can be written as $L(\underline{\lambda},a) = \sum_{j=1}^{s} n_j\lambda_{i_j}$ for $a \in \mathcal{A}_{\underline{n}_s}$. Now consider a decision problem with the action space $\mathcal{A}_{\underline{n}_s}$, the above loss function and the observation vector $\underline{x}$. Clearly this problem is equivalent to partitioning k populations $\pi_1,\ldots,\pi_k$ into s+1 subsets $(\gamma_1,\ldots,\gamma_s,\gamma_{s+1})$ where γ_j is of size 1 for j = 1,...,s, γ_{s+1} is of size k-s and the action $(i_1,\ldots,i_s)$ corresponds to the action $(\{\pi_{i_1}\},\ldots,\{\pi_{i_s}\},\{\pi_i: 1\le i\le k, i \neq i_1,\ldots,i_k\})$. Note that this decision problem is invariant under the permutation group, and that the loss function satisfies the monotonicity and the invariance properties of Eaton (1967) with λ_i^{-1} being the parameter θ_i in Eaton's paper. Since the density $f(x,\lambda_i)$ of x_i has the monotone likelihood ratio property in x and $\theta_i = \lambda_i^{-1}$, it follows from Eaton's result that the rule which assigns $\pi_{(k-j+1)}$ to γ_j for j = 1,...,s and the remaining brands to γ_{s+1} is Bayes wrt any permutationally symmetric prior distribution of $\underline{\lambda}$. Hence, the result follows.

The following lemma is needed for the main result.

Lemma 3.2.2. Assume that $X_1,\ldots,X_k$, given $\underline{\theta} = (\theta_1,\ldots,\theta_k) \in \Theta^k$, are independently distributed random variables with X_i having pdf $f(x,\theta_i)$. If $f(x,\theta)$ has the monotone likelihood ratio (MLR) property in x and θ, and if the prior distribution, $\tau(\underline{\theta})$, of $\underline{\theta} = (\theta_1,\ldots,\theta_k)$ is permutationally symmetric on Θ^k, then, for $i \geq j$,

$$E[g(\theta_{(i)})|x] \geq E[g(\theta_{(j)})|\underline{x}]$$

provided $g(\cdot)$ is non-decreasing on Θ where $\theta_{(i)}$ is the θ associated with $x_{[i]}$.

Proof. Let $\Omega_0 = \{\underline{\theta} \in \Theta^k: \theta_{(i)} \geq \theta_{(j)}\}$, then

$$\int_{\Theta^k} [g(\theta_{(i)})-g(\theta_{(j)})]f(\underline{x},\underline{\theta})d\tau(\underline{\theta})=[\int_{\Omega_0} +\int_{\Omega_0^c}][g(\theta_{(i)})-g(\theta_{(j)})]f(x,\underline{\theta})d\tau(\underline{\theta})$$

$$= \int_{\Omega_0} [g(\theta_{(i)})-g(\theta_{(j)})][f(\underline{x},\underline{\theta})-f(\underline{x},\underline{\theta}')d\tau(\underline{\theta})$$

where $f(\underline{x},\underline{\theta}) = \prod_{i=1}^{k} f(x_i,\theta_i)$ and $\underline{\theta}'$ is obtained from $\underline{\theta}$ by interchanging $\theta_{(i)}$ and $\theta_{(j)}$, keeping other components fixed. Therefore,

$$E[g(\theta_{(i)})-g(\theta_{(j)})|\underline{x}] = n(\underline{x})\int_{\Omega_0} [g(\theta_{(i)})-g(\theta_{(j)})][f(\underline{x},\underline{\theta})-f(\underline{x},\underline{\theta}')] d\tau(\underline{\theta})$$

where $n(\underline{x})$ is a normalizing factor. The result follows from the MLR property of $f(x,\theta)$ and the fact that $g(\theta_{(i)})-g(\theta_{(j)}) \geq 0$ for $\underline{\theta} \in \Omega_0$ if g is non-decreasing.

Now we state the following result.

<u>Theorem 3.2.1</u>. For any permutationally symmetric prior distribution of $\underline{\lambda}$ on $(0,\infty)^k$, the Bayes rule $\delta^* = \delta_{\underline{n}_1}$, i.e., δ^* draws all m components from $\pi_{(k)}$.

<u>Proof</u>. It follows from (3.2.2) that $r(\delta_{\underline{n}_s},\underline{x})$ can be written as

$$r(\delta_{\underline{n}_s},\underline{x}) = E[\sum_{j=1}^{s} n_j \lambda_{(k-j+1)}|\underline{x}] \text{ for } \underline{n}_s \in N_s.$$

Therefore

$$r(\delta_{\underline{n}_s},\underline{x})-E[(m-s+1)\lambda_{(k)}+\sum_{j=2}^{s}\lambda_{(k-j+1)}|\underline{x}]$$

$$= E[(m-\sum_{j=2}^{s}n_j)\lambda_{(k)}+\sum_{j=2}^{s}n_j\lambda_{(k-j+1)}|\underline{x}]-E[(m-s+1)\lambda_{(k)}+\sum_{j=2}^{s}\lambda_{(k-j+1)}|\underline{x}]$$

$$= E[\sum_{j=2}^{s} (n_j - 1)(\lambda_{(k-j+1)} - \lambda_{(k)}) | \underline{x}]$$

$$\geq 0 \text{ by Lemma 3.2.2.}$$

Thus $\underset{\underline{n}_s \in N_s}{\text{Min}} r(\delta_{\underline{n}_s}, \underline{x}) = r(\delta_s, \underline{x})$ where $\delta_s = \delta_{\underline{n}_s^*}$ with $\underline{n}_s^* = (m-s+1, 1, \ldots, 1) \in N_s$, i.e., δ_s draws $(m-s+1)$ components from $\pi_{(k)}$ and one component from each $\pi_{(k-j+1)} (j = 2, \ldots, s)$. Since, for any s; $2 \leq s \leq k \wedge m$,

$$r(\delta_s, \underline{x}) - r(\delta_1, \underline{x}) = E[\sum_{j=2}^{s} (\lambda_{(k-j+1)} - \lambda_{(k)})]$$

$$\geq 0 \text{ by Lemma 3.2.2,}$$

the result follows from Lemma 3.2.1.

The next result follows from considering a permutation symmetric prior which gives mass 1/k! at each parameter point obtained from a fixed parameter $\underline{\lambda} \in (0,\infty)^k$ by permuting its components.

Corollary 3.2.1. The 'natural' rule δ^* is uniformly best among the permutation invariant rules.

It follows from the above result that the natural rule δ^* is admissible and minimax among all decision rules.

Remark 3.2.1. If we consider a loss function $L_1(\underline{\lambda}, (i_1, \ldots, i_m)) = (m \underset{1 \leq i \leq k}{\text{Min}} \lambda_i)^{-1} - (\sum_{j=1}^{m} \lambda_{i_j})^{-1}$, it can be easily shown that Lemma 3.2.1 holds for this loss function. Assuming an exchangeable prior for $\underline{\lambda}$ on $(0,\infty)^k$, it can be verified that the Bayes rule δ^* for the loss function L_1 satisfies $r(\delta^*, \underline{x}) = \underset{1 \leq s \leq k \wedge m}{\text{Min}} r(\delta_s, \underline{x})$ where the rule δ_s is as in the proof of Theorem 3.2.1. Even though this is a considerable reduction in a number of candidates for the Bayes rule, specification of it seems difficult except when $m = 2$. Further simplification of the Bayes rule with respect to a specific prior would be interesting along with some numerical results.

(B) The 1-out-of-2 system

Here, the action space is $\mathcal{a} = \{(i,j): 1 \leq i \leq j \leq k\}$ where (i,j) is the action of drawing one component each from π_i and π_j, respectively. For the 1-out-of-2 system the expected lifelength of the system for the action (i,j) is given by $\lambda_i^{-1}+\lambda_j^{-1}-(\lambda_i+\lambda_j)^{-1}$. Again as in (A), the loss function is assumed to be given by

$$L(\underline{\lambda},(i,j)) = (\lambda_i^{-1}+\lambda_j^{-1}-(\lambda_i+\lambda_j)^{-1})^{-1}. \tag{3.2.4}$$

As mentioned earlier, Broström (1977) considered a scale invariant loss function obtained by dividing (3.2.4) by $L(\underline{\lambda},(1,2))$, i.e., the loss incurred by an 'intermediate' action which we don't have in the case when $k > 2$.

We will derive a Bayes rule with respect to a prior distribution of $\underline{\lambda}$. The prior distribution of $\underline{\lambda}$ is assumed to be independent natural conjugate gamma distribution with two parameters, i.e., the joint a priori pdf of $\underline{\lambda}$ is given by

$$\tau(\underline{\lambda}) = \prod_{i=1}^{k} [\frac{\beta^\alpha}{\Gamma(\alpha)} \lambda_i^{\alpha-1} e^{-\beta\lambda_i}], \ \alpha > 0 \text{ and } \beta > 0. \tag{3.2.5}$$

The improper prior corresponding to the vague prior knowledge can be given by the above with $\alpha = \beta = 0$. It can be easily observed that, given $\underset{\sim}{X} = \underset{\sim}{x}$, $\lambda_{(1)},\ldots,\lambda_{(k)}$ are, a posteriori, independently distributed gamma random variables with mean $(n+\alpha)\ (x_{[i]}+\beta)^{-1}$ and variance $(n+\alpha)\ (x_{[i]}+\beta)^{-2}$.

As in (A), the action space $\mathcal{a}$ can be partitioned into $\mathcal{a}_1 = \{(i,i): i = 1,\ldots,k\}$ and $\mathcal{a}_2 = \{(i,j): 1 \le i < j \le k\}$. Then the decision problem with $\mathcal{a}_s$ (s = 1,2) is equivalent to partitioning $\pi_1,\ldots,\pi_k$ into two subsets γ_1, and γ_2 with γ_1 being of size s and γ_2 being of size k-s. Then by the same arguments as in Lemma 3.2.1, we have the next result.

Lemma 3.2.3. Assume that the prior of $\underset{\sim}{\lambda}$ on $(0,\infty)^k$ is permutationally symmetric. Then the Bayes rule δ^* is given by

$$r(\delta^*,\underset{\sim}{x}) = \mathrm{Min}\{r(\delta_1,\underset{\sim}{x}),\ r(\delta_2,\underset{\sim}{x})\} \tag{3.2.6}$$

where δ_1 chooses 2 components from $\pi(k)$, and δ_2 chooses one component from $\pi_{(k)}$ and another from $\pi_{(k-1)}$.

Now we state the Bayes rule.

Theorem 3.2.2. Assume that the prior is given by (3.2.5). Then the Bayes rule δ^* is given by

$$\delta^* = \begin{cases} \delta_1 & \text{if } x_{[k-1]}+\beta \le c(x_{[k]}+\beta) \\ \delta_2 & \text{if } x_{[k-1]}+\beta > c(x_{[k]}+\beta) \end{cases} \tag{3.2.7}$$

where $c = H^{-1}_{\alpha,n}(0) \in (0,1)$, $H_{\alpha,n}(c) = E[\frac{UV\,(U+cV\,)}{U^2+c^2V^2+cUV}] - \frac{2}{3}\,(n+\alpha)$ for $c > 0$ and U, V are iid gamma random variables with mean $(n+\alpha)$ and variance $(n+\alpha)$.

Proof. It follows from (3.2.4) and (3.2.5) that

$$r(\delta_1,\underline{x}) = \frac{2}{3} E[\lambda_{(k)}|\underline{x}] = \frac{2}{3} \frac{n+\alpha}{x_{[k]}+\beta} \text{ and}$$

$$r(\delta_2,\underline{x}) = E[\lambda_{(k)}\lambda_{(k-1)}(\lambda_{(k)}+\lambda_{(k-1)})(\lambda^2_{(k)}+\lambda_{(k)}\lambda_{(k-1)}+\lambda^2_{(k-1)})^{-1}|\underline{x}]$$

$$= \frac{1}{x_{[k]}+\beta} E[\frac{UV(U+r_kV)}{U^2+r_kUV+r_k^2V^2}] \text{ for } r_k = \frac{x_{[k-1]}+\beta}{x_{[k]}+\beta}$$

where U and V are gamma variables with mean $(n+\alpha)$ and variance $(n+\alpha)$. Since $H_{\alpha,n}(t)$ is non-increasing in $t > 0$, $r(\delta_1,\underline{x}) \geq r(\delta_2,\underline{x})$ if and only if $H_{\alpha,n}(r_k) \leq 0$, i.e., $r_k \geq H^{-1}_{\alpha,n}(0)$. Furthermore, it can be easily observable that

$$H_{\alpha,n}(1) = E[\frac{UV(U+V)}{U^2+UV+V^2}] - \frac{2}{3}(n+\alpha) < 0$$

which implies $0 < H^{-1}_{\alpha,n}(0) < 1$. Hence the result follows from Lemma 3.2.3.

As it was pointed out, the Bayes rule wrt the vague prior knowledge can be obtained from (3.2.7) by taking $\alpha = \beta = 0$. It can be easily shown that $x_1(x_1+\beta)^{-1},\ldots,x_k(x_k+\beta)^{-1}$ are marginally independent beta random variables with mean $n(n+\alpha)^{-1}$ when the prior is given by (3.2.5). It follows that the overall risk of the rule δ_1 is given by

$$r(\delta_1) = \frac{2}{3}(n+\alpha)\beta^{-1} E[Z_{[1]}] \leq \frac{2}{3}\frac{\alpha}{\beta}$$

where $Z_{[1]}$ is the smallest order statistic based on a sample of size k from beta distribution with mean $\alpha(n+\alpha)^{-1}$. Therefore the overall risk of the Bayes rule δ^* is finite. Furthermore it can be shown that the risk function $R(\underline{\lambda},\delta)$ is a continuous function of $\underline{\lambda}$ for non-randomized rule δ. It follows that the Bayes rule δ^* is admissible in the class of non-randomized decision rules.

Remark 3.2.2. If a loss function is given by $L_2(\underline{\lambda},(i,j)) = \frac{3}{2}(\min_{1\le i\le k} \lambda_i)^{-1} - [\lambda_i^{-1}+\lambda_j^{-1}-(\lambda_i+\lambda_j)^{-1}]$, it follows from similar methods that the Bayes rule wrt the prior given by (3.2.5) is given by (3.2.7) with $c = G_{\alpha,n}^{-1}(0) \in (0,1)$ and $G_{\alpha,n}(t) = \frac{1}{2}\frac{1}{n+\alpha-1} - \frac{t}{n+\alpha-1} + E[\frac{t}{tU+V}]$ for $t > 0$ where U and V are iid gamma random variables with mean and variance equal to $(n+\alpha)$.

At the end of this chapter tables of the constants c are provided to implement the Bayes rules given in Theorem 3.2.2 and Remark 3.2.2. Values of c are found by numerically integrating $H_{\alpha,n}(c)$ and $G_{\alpha,n}(c)$ using Laguerre polynomials. In doing this the first fifteen Laguerre polynomials were used (see Abramowitz and Stegun (1964)).

3.3 Some selection procedures for symmetric scale parameter populations

Let $\pi_1,\ldots,\pi_k$ denote k independent populations with continuous cumulative distribution functions $F_1(x) = F(x/\sigma_1),\ldots,F_k(x) = F(x/\sigma_k)$, respectively, where $\sigma_i > 0$, $i = 1,2,\ldots,k$, and F is symmetric about the origin. Suppose we are interested in ranking or screening these populations with regard to some measures of dispersion. One measure of dispersion for a symmetric distribution G is a non-negative functional $\tau(G)$ or equivalently $\tau(Z)$ with Z being a random variable with distribution function G (see Bickel and Lehmann (1976)) such that

$\tau(aZ) = a\tau(Z)$ for $a > 0$

$\tau(Z+b) = \tau(Z)$ for any real b and (3.3.1)

$\tau(Z) \leq \tau(Z')$ whenever $|Z'|$ is stochastically larger than $|Z|$.

It follows that the measure of dispersion of F_i is $\tau(F_i) = \sigma_i \tau(F)$ for $i = 1,\ldots,k$. Hence ranking the populations with regard to the measure of dispersion becomes equivalent to the ranking in terms of σ_i's. This leads to selection and ranking problem in terms of σ_i's.

Several procedures for this problem have been proposed. When we know the functional form of F, some suitable estimators of σ_i's are usually used for selection or ranking purpose. For example we might use sample standard deviations for normal populations and sample mean deviations for double exponential populations. When we do not assume the functional form of F, we may use the estimators of some measures of dispersion of F_i's and study the robustness of the selection procedures. This is the approach taken in this section.

(A) <u>Selection of the t 'best' populations - Indifference-zone approach</u>

Let $X_{i1},\ldots,X_{in}(i=1,\ldots,k)$ be the independent observations from population π_i with cdf $F_i(x)=F(\frac{x-\theta_i}{\sigma_i})$ where F is continuous and symmetric about the origin. Here $\sigma_1,\ldots,\sigma_k$ are unknown and $\theta_1,\ldots,\theta_k$ may be either known or unknown. Let $\sigma_{[1]} \leq \ldots \leq \sigma_{[k]}$ be the ordered σ_i's where no a priori information about the correct pairing of π_i and $\sigma_{[i]}$ is assumed. Our goal is to select the $t(1 \leq t \leq k-1)$ populations associated with t smallest scale parameters $\sigma_{[1]},\ldots,\sigma_{[t]}$ based on the independent observations $X_{i1},\ldots,X_{in}$, $i = 1,\ldots,k$.

The indifference-zone formulation, due to Bechhofer (1954), of this problem may be briefly described as follows;

(i) Choose an appropriate statistic T and observe $T_i=T_i^{(n)}=T(X_{i1},\ldots,X_{in})$ for $i = 1,\ldots,k$.

(ii) Use the natural procedure R_T which selects populations associated with t smallest T_i values.

(iii) For preassigned $P^* \in (1/\binom{k}{t},1)$ and $\Delta > 1$, determine the smallest sample size $n=n(k,P^*,\Delta)$ such that $\inf_{\underline{\sigma}\in\Omega(\Delta)} P(CS|\underline{\sigma}) \geq P^*$ where $\Omega(\Delta)=\{\underline{\sigma}=(\sigma_1,\ldots,\sigma_k): \sigma_{[t+1]} \geq \Delta\sigma_{[t]}\}$ and CS stands for a correct selection of the t populations associated with $\sigma_{[1]},\ldots,\sigma_{[t]}$. (3.3.2)

Several optimum properties of the above procedure with suitable statistic T have been proved under certain assumptions on F (see, for example, Bahadur and Goodman (1952), Lehmann (1966), Eaton (1967)).

First we will consider the case when the location parameters are known.

<u>Case (I): $\theta_1,\ldots,\theta_k$ known</u>

In this case we may assume $\theta_1 =\ldots= \theta_k = 0$ without loss of generality. Let us consider the following selection procedures based on robust estimators studied by Bickel and Lehmann (1976); for $1 \leq p \leq 2$ and $0 < \alpha \leq \frac{1}{2} \leq \beta < 1$,

Procedure R(p): Use $T(X_{i1},\ldots,X_{in})=(\frac{1}{n}\sum_{j=1}^{n}|X_{ij}|^p)^{\frac{1}{p}}$ in (3.3.2)

Procedure R(α,β): Use $T(X_{i1},\ldots,X_{in})=[\sum_{j=[n\alpha]+1}^{[n\beta]} X^2_{i[j]}/([n\beta]-[n\alpha])]^{\frac{1}{2}}$ in (3.3.2), where $X^2_{i[1]} \leq \cdots \leq X^2_{i[n]}$ denote the ordered $X^2_{i1},\ldots,X^2_{in}$ and $[\cdot]$ is the greatest integer function. Note that R(2) is the procedure proposed by Bechhofer and Sobel (1954) for the normal populations. This procedure will be taken as the standard one for the comparison of procedures.

The next result gives the least favorable configurations for the above procedures.

Lemma 3.3.1. Let R_T denote the procedure defined by (3.3.2) where the statistic T is scale invariant, i.e., $T(aX_{i1},\ldots,aX_{in}) = aT(X_{i1},\ldots,X_{in})$ for any $a > 0$. Then the following holds:

$$\inf_{\underline{\sigma}\in\Omega(\Delta)} P(CS|\underline{\sigma}) = P(CS|\underline{\sigma} = (\underbrace{1,\ldots,1}_{t\text{-times}},\Delta,\ldots,\Delta)) \tag{3.3.3}$$

Proof. Using a theorem of Barr and Rizvi (1966) it follows that $P(CS|\underline{\sigma}) = P(\max_{1\leq j\leq t} T_{(j)} < \min_{t+1\leq j\leq k} T_{(j)}|\underline{\sigma})$ with $T_{(1)},\ldots,T_{(k)}$ being associated with $\sigma_{[1]} \leq \cdots \leq \sigma_{[k]}$ is a non-decreasing function of $\sigma_{[t+1]},\ldots,\sigma_{[k]}$ and non-increasing in $\sigma_{[1]},\ldots,\sigma_{[t]}$. Therefore the result follows from the scale invariance of T.

It follows from the above result that we have the same least favorable configuration as long as the statistic T is scale invariant. However, this slippage configuration of parameters is not the least favorable for the procedures based on ranks (see Rizvi and Woodworth (1970)). Now the sample size n required by the basic probability condition can be found by

$$P(\max_{1\leq j\leq t} T_j < \min_{t+1\leq j\leq k} T_j|\underline{\sigma}) = P^* \tag{3.3.4}$$

subject to the condition

$$\underline{\sigma} = (1,\ldots,1,\Delta,\ldots,\Delta). \tag{3.3.5}$$

To solve this we need the complete knowledge about F. But if we do not wish to rely on the distributional assumption on F, we can find a large sample solution.

Large Sample Solution and Asymptotic Efficiency

To find the large sample solution for n, following Lehmann (1963), we use the device of replacing Δ by Δ_n ar ' determine the large sample solution of n required to satisfy (3.3.4) subject to the condition

$$\sigma_{[k]} = \ldots = \sigma_{[t+1]} = \Delta_n \sigma_{[t]} = \Delta_n \, \sigma_{[1]}. \tag{3.3.6}$$

As Lehmann (1963) and Puri and Puri (1969) have pointed out, considering Δ_n as a sequence depending on n is only a mathematical device to approximate the actual situation and in practice Δ_n will be identified with the given value of Δ.

Assume that $\sqrt{n}\,(T(Z_1,\ldots,Z_n)-\tau(F))$ is asymptotically normally distributed with mean 0 and variance $v^2(F)$ when $Z_1,\ldots,Z_n$ are iid random variables with cdf F(z), where $\tau(F)$ is a measure of dispersion of F satisfiying (3.3.1). Let $Y_i = T(Z_{i1},\ldots,Z_{in})$ for $i = 1,\ldots,k.$, where $Z_{11},\ldots,Z_{kn}$ are kn iid random variables with cdf $F(\cdot)$. Then from (3.3.6), we see that equation (3.3.4) is equivalent to $P(Y_i \le \Delta_n Y_j,\ i = 1,\ldots,t,\ j = t+1,\ldots,k) = P^*$. This implies, after taking logarithmetic transformation, that

$$\lim_{n\to\infty} P(U_i - U_j \leq \sqrt{n}\ (\log \Delta_n)\tau(F)/v(F),\ i = 1,\ldots,t,\ j = t+1,\ldots,k) = P^*$$

where $U_1,\ldots,U_k$ are iid standard normal variables. This equation will be satisfied if and only if

$$\Delta_n = 1 + \frac{\Delta^*}{\sqrt{n}}\frac{v(F)}{\tau(F)} + o(n^{-\frac{1}{2}})$$

where Δ^* is determined by

$$P^* = t \int_{-\infty}^{\infty} \Phi^{t-1}(x)\Phi^{k-t}(\Delta^* - x)d\Phi(x). \tag{3.3.7}$$

Thus we have proved the following result.

<u>Lemma 3.3.2</u>. Consider the procedure R_T in (3.3.2) with T being scale invariant and satisfying the properties in the above paragraph. Let n be the solution of (3.3.4) subject to (3.3.6), then as $n \to \infty$

$$\Delta_n = 1 + \frac{\Delta^*}{\sqrt{n}}\frac{v(F)}{\tau(F)} + o(n^{-\frac{1}{2}}) \tag{3.3.8}$$

where Δ^* is determined by (3.3.7).

It follows from the above result that, for given Δ and P^*, a large sample solution of n is given by

$$n = \left(\frac{\Delta^*}{\Delta - 1}\right)^2 \frac{v^2(F)}{\tau^2(F)}. \tag{3.3.9}$$

The values of Δ^* satisfying (3.3.7) can be found in Bechhofer (1954). From the central limit theorem it follows that if $\int x^4 dF(x) < \infty$, then (3.3.9) holds for R(p) ($1 \leq p \leq 2$) with $\tau(F) = (\int |x|^p dF(x))^{\frac{1}{p}}$ and $v^2(F) = p^{-2}\ (\int |x|^p dF(x))^{\frac{2}{p} - 2} [\int |x|^{2p} dF(x) - (\int |x|^p dF(x))^2]$.

Therefore, the large sample solution for the procedure R(p) is given by

$$n(p) = \left(\frac{\Delta^*}{\Delta-1}\right)^2\left[\frac{E|Z|^{2p}}{(E|Z|^p)^2} - 1\right] p^{-2} \tag{3.3.10}$$

where the expectations are taken wrt the distribution $F(\cdot)$. Let us consider $R(\alpha,\beta)$ and let $T_n(\alpha,\beta) = \left[\frac{1}{[n\beta]-[n\alpha]} \sum_{j=[n\alpha]+1}^{[n\beta]} Z^2_{[j]}\right]^{\frac{1}{2}}$ where $Z^2_{[1]} \leq \cdots \leq Z^2_{[n]}$ are the ordered $Z_1^2,\ldots,Z_n^2$ with $Z_1,\ldots,Z_n$ being iid random variables with cdf $F(\cdot)$. Then the following theorem (see, for example, Stigler (1973)) gives the sufficient condition for the procedure $R(\alpha,\beta)$ to satisfy the assumptions in the above lemma.

Theorem 3.3.1. Let $G(\cdot)$ denote the cdf of Z_1^2. Assume that $a = G^{-1}(\alpha)$ and $b = G^{-1}(\beta)$ are uniquely determined. Then as $n \to \infty$, the limiting distribution of $\sqrt{n}\,(T_n^2(\alpha,\beta)-\tau^2(\alpha,\beta))$ is normal with mean 0 and variance $v^2(\alpha,\beta)$ where $\tau(\alpha,\beta) = \left[\frac{1}{\beta-\alpha}\int_a^b y\,dG(y)\right]^{\frac{1}{2}}$, $v^2(\alpha,\beta) = (\beta-\alpha)^{-2}[(\beta-\alpha)c+(b-\tau^2(\alpha,\beta))^2\beta(1-\beta)+(a-\tau^2(\alpha,\beta))^2\alpha(1-\alpha)$ $-2(b-\tau^2(\alpha,\beta))(a-\tau^2(\alpha,\beta))\alpha(1-\beta)]$ and $c=\frac{1}{\beta-\alpha}\int_a^b y^2dG(y)-\tau(\alpha,\beta)^4$.

It follows that under the assumptions in the above theorem, the large sample solution for the procedure $R(\alpha,\beta)$ is given by

$$n(\alpha,\beta)= \frac{1}{4}\left(\frac{\Delta^*}{\Delta-1}\right)^2 \frac{v^2(\alpha,\beta)}{\tau^4(\alpha,\beta)}. \tag{3.3.11}$$

Now comparison between the procedures is in order. To this end, the procedure $R(2)$ is taken as the standard one and, following Lehmann (1963), the asymptotic relative efficiency (ARE) of a procedure wrt $R(2)$ is then defined to be the limiting ratio of the reciprocals of the corresponding sample size required to achieve the same minimum probability of a correct selection over $\Omega(\Delta)$. These are given as follows:

$$\text{ARE}(R(p),R(2);F) = \frac{p^2}{4}\left[\frac{EZ^4}{(EZ^2)^2}-1\right]\Big/\left[\frac{E|Z|^{2p}}{(E|Z|^p)^2}-1\right]$$
$$\text{ARE}(R(\alpha,\beta),R(2);F) = \left[\frac{EZ^4}{(EZ^2)^2}-1\right]\frac{\tau^4(\alpha,\beta)}{v^2(\alpha,\beta)} \tag{3.3.12}$$

where the expectations are taken wrt the distribution $F(\cdot)$. Note that the above ARE's are the same as those of the estimators in Bickel and Lehmann (1976) where one can find several examples of the ARE's. For example, $\text{ARE}(R(p),R(2);F) \geq \frac{3}{4\sqrt{\pi}} \frac{\Gamma^2(\frac{p+1}{2})}{\Gamma(p+\frac{1}{2})} \cdot p^2$ for all F which is the cdf of a scale mixture of normal distributions with a common mean.

Case (II): $\theta_1,\ldots,\theta_k$ unknown

In this case we use $T_i = T(X_{i1},\ldots,X_{in}) = (\frac{1}{n}\sum_{j=1}^{n}|X_{ij}-\bar{X}_i|^p)^{\frac{1}{p}}$ for the procedure R(p) and $T_i = [\sum_{j=[n\alpha]+1}^{[n\beta]}(X_{i[j]}-M_i)^2/([n\beta]-[n\alpha])]^{\frac{1}{2}}$ for the procedure $R(\alpha,\beta)$ where M_i is the median of $X_{i1},\ldots,X_{in}$ and $\bar{X}_i = \frac{1}{n}\sum_{j=1}^{n} X_{ij}$. Then it follows from Bickel and Lehmann (1976) that the results analogous to the case (I) hold for R(p) without any further assumption, and hold for $R(\alpha,\beta)$ with the further assumption that F is differentiable with positive and continuous derivative f.

(B) Selection of a subset containing the 'best' population

Here we consider the problem of selecting a subset of k populations which includes the population associated with the smallest scale parameter with at least probability P*. This problem has been studied by Gupta and Sobel (1962) and Gupta (1965) when a specific form of the cdf is assumed. Also nonparametric or robust selection procedures have been proposed and studied by Sobel (1967), Blumenthal and Patterson (1969), Wong (1976) and McDonald (1977) among others.

Let $X_{i1},\ldots,X_{in}$ be the independent observations from π_i with cdf $F_i(x) = F(\frac{x-\theta_i}{\sigma_i})$ for $i = 1,\ldots,k$. Here, F is assumed to be continuous and symmetric about the origin, σ_i is unknown and θ_i is known. We may assume $\theta_1 = \ldots = \theta_k = 0$. Let $\sigma_{[1]} \leq \ldots \leq \sigma_{[K]}$ denote the ordered σ's and $\pi_{[i]}$ denote the population π associated with $\sigma_{[i]}$. Our goal is to select a subset of random size depending on the observed data so that, for given $P^*(\frac{1}{k} < p^* < 1)$,

$$\inf_{\underline{\sigma}\in\Omega} P(CS|\underline{\sigma}) \geq P^* \tag{3.3.13}$$

where $\Omega = \{\underline{\sigma}=(\sigma_1,\ldots,\sigma_k): \sigma_i > 0,\ i = 1,\ldots,k\}$ and CS denotes the correct selection of a subset which includes $\pi_{[1]}$.

We will consider the procedure $R^*(p)(1 \leq p \leq 2)$ which includes π_i in the selected subset if $T_i \leq c^{-1} \min_{1\leq j\leq k} T_j$ where $T_i = T(X_{i1},\ldots,X_{in}) = (\frac{1}{n}\sum_{j=1}^{n}|X_{ij}|^p)^{\frac{1}{p}}$ and $c(0 < c \leq 1)$ is determined to satisfy (3.3.13). Then it can be easily shown that, for the procedure $R^*(p)$, the infimum of $P(CS|\underline{\sigma})$ occurs when all the σ_i's are equal; therefore, the constant $c = c(n,k,p,P^*)$ is obtained by $\int_0^\infty [1-G(cx)]^{k-1}dG(x) = P^*$, where G is the cdf of T_1 when $\sigma_1=1$. By the same reason as in (A) one can find a large sample solution of c if he wishes to avoid the distributional assumption on F. Similarly as in (A) we can get the large sample solution of c as in the next result.

Lemma 3.3.3. Let c be the constant satisfying the basic probability condition. Then, as $n \to \infty$,

$$c = 1 - d\, n^{-\frac{1}{2}} p^{-1} \left[\frac{E|Z|^{2p}}{(E|Z|^{p})^{2}} - 1\right]^{\frac{1}{2}} + o(n^{-\frac{1}{2}}) \tag{3.3.14}$$

where the expectations are taken wrt the distribution $F(\cdot)$ and d is determined by

$$\int \Phi^{k-1}(x+d)\,d\Phi(x) = P^*. \tag{3.3.15}$$

Values of d satisfying (3.3.15) are available in Gupta (1956) for various values of k and P^*. Let S_p and S_p^* denote the random size of the selected subset and the random number of the non-best populations in the selected subset, respectively for the procedure $R^*(p)$. It follows from the result in Gupta and Sobel (1962) that $\sup_{\underline{\sigma}\in\Omega} E(S_p|\underline{\sigma}) = kP^*$ provided $f(x) = F'(x)$ is log-concave. Since small values of S_p^* are desirable, we would like to keep $E(S_p^*|\underline{\sigma})$ as small as possible. Let us consider the following slippage configuration;

$$\Delta\sigma_{[1]} = \sigma_{[2]} = \ldots = \sigma_{[k]} \quad \text{for some } \Delta > 1. \tag{3.3.16}$$

Then for a given $\epsilon > 0$, we would like to have small sample size n where n is determined, subject to (3.3.16), by

$$E(S_p^*|\underline{\sigma}) = \epsilon\,. \tag{3.3.17}$$

Following McDonald (1969) we define the asymptotic relative efficiency $ARE(p,2;\Delta,F)$ of $R^*(p)$ to be the limiting ratio of

$n(2,\epsilon)$ to $n(p,\epsilon)$ as $\epsilon \to 0$. To find the large sample solution of (3.3.17), we use the device of replacing Δ by Δ_n.

Lemma 3.3.4. Assume (3.3.14) and (3.3.16). Then as $n \to \infty$,

$$\Delta_n = 1+\Delta^* n^{-\frac{1}{2}} p^{-1}\left[\frac{E|Z|^{2p}}{(E|Z|^p)^2} - 1\right]^{\frac{1}{2}} + o(n^{-\frac{1}{2}}) \tag{3.3.18}$$

where the expectations are taken wrt $F(\cdot)$ and Δ^* is determined by $(k-1) \int \Phi^{k-2}(d-x)\Phi(d-\Delta^*-x)d\Phi(x) = \epsilon$ for d given by (3.3.15).

Hence the large sample solution of (3.3.17), keeping in terms of order $n^{\frac{1}{2}}$, is given by

$$n(p,\epsilon) = \left(\frac{\Delta^*}{\Delta-1}\right)^2 p^{-2}\left[\frac{E|Z|^{2p}}{(E|Z|^p)^2} - 1\right].$$

It follows that the ARE of $R^*(p)$ relative to $R(2)$ is given by

$$ARE(p,2;\Delta,F) = \frac{p^2}{4}\left[\frac{EZ^4}{(EZ^2)^2} - 1\right] \Big/ \left[\frac{E|Z|^{2p}}{(E|Z|^p)^2} - 1\right]$$

which is the same as that in (A).

Table VI

Lists c-values to implement the Bayes rule in Theorem 3.2.2. which depends on n and α through the quantity $m = n+\alpha$.

m	c	m	c	m	c
1.0	.3033	11.0	.8870	21.0	.9388
1.5	.4392	11.5	.8916	21.5	.9402
2.0	.5342	12.0	.8958	22.0	.9415
2.5	.6021	12.5	.8997	22.5	.9428
3.0	.6530	13.0	.9034	23.0	.9440
3.5	.6925	13.5	.9067	23.5	.9451
4.0	.7240	14.0	.9099	24.0	.9462
4.5	.7496	14.5	.9128	24.5	.9473
5.0	.7710	15.0	.9156	25.0	.9483
5.5	.7890	15.5	.9182	25.5	.9493
6.0	.8044	16.0	.9206	26.0	.9502
6.5	.8177	16.5	.9229	26.5	.9511
7.0	.8293	17.0	.9251	27.0	.9520
7.5	.8396	17.5	.9271	27.5	.9529
8.0	.8486	18.0	.9291	28.0	.9537
8.5	.8567	18.5	.9309	28.5	.9545
9.0	.8640	19.0	.9326	29.0	.9552
9.5	.8706	19.5	.9343	29.5	.9560
10.0	.8766	20.0	.9359	30.0	.9567
10.5	.8820	20.5	.9374	30.5	.9574

Table VII

Lists c-values to implement the Bayes rule in Remark 3.2.2. which depends on n and α through the quantity $m = n+\alpha$.

m	c	m	c	m	c
1.5	.7854	11.5	.9817	21.5	.9904
2.0	.8592	12.0	.9825	22.0	.9907
2.5	.8957	12.5	.9832	22.5	.9909
3.0	.9173	13.0	.9839	23.0	.9911
3.5	.9315	13.5	.9845	23.5	.9913
4.0	.9415	14.0	.9851	24.0	.9915
4.5	.9490	14.5	.9856	24.5	.9916
5.0	.9548	15.0	.9861	25.0	.9918
5.5	.9594	15.5	.9866	25.5	.9920
6.0	.9631	16.0	.9870	26.0	.9921
6.5	.9662	16.5	.9874	26.5	.9923
7.0	.9688	17.0	.9878	27.0	.9924
7.5	.9711	17.5	.9882	27.5	.9926
8.0	.9730	18.0	.9885	28.0	.9927
8.5	.9747	18.5	.9888	28.5	.9928
9.0	.9762	19.0	.9891	29.0	.9930
9.5	.9776	19.5	.9894	29.5	.9931
10.0	.9788	20.0	.9897	30.0	.9932
10.5	.9798	20.5	.9900	30.5	.9933
11.0	.9808	21.0	.9902	31.0	.9934

BIBLIOGRAPHY

BIBLIOGRAPHY

Abramowitz, M. and Stegun, I. A. (1964). Handbook of Mathematical Functions. N.B.S. Applied Mathematics Series, 55.

Alam, K. (1973). On a multiple decision rule. Ann. Math. Statist. 1: 750-755.

Bahadur, R. R. (1950). On the problem in the theory of k populations. Ann. Math. Statist. 21: 362-375.

Bahadur, R. R. and Goodman, L. (1952). Impartial decision rules and sufficient statistics. Ann. Math. Statist. 23: 553-562.

Barr, D. R. and Rizvi, M. H. (1966). An introduction to selection and ranking procedures. J. Amer. Statist. Assoc. 61: 640-646.

Bechhofer, R. E. (1954). A single-sample multiple decision procedure for ranking means of normal populations with known variance. Ann. Math. Statist. 25: 16-39.

Bechhofer, R. E. and Sobel, M. (1954). A single-sample multiple-decision procedure for ranking variances of normal populations. Ann. Math. Statist. 25: 273-289.

Berger, R. (1977). Minimax, Admissible and Gamma-Minimax Multiple Decision Rules. Ph.D. Thesis. Dept. of Statist., Purdue Univ., West Lafayette, Indiana.

Bhapkar, V. P. and Gore, A. P. (1971). Some selection procedures based on U-statistics for the location and scale problems. Ann. Inst. Statist. Math. 23: 375-386.

Bhattacharyya, P. K. (1956). Comparison of the means of k normal populations. Calcutta Statist. Assoc. Bull. 7: 1-20.

Bhattacharyya, P. K. (1958). On the large sample behavior of a multiple decision procedure. Calcutta Statist. Assoc. Bull. 8: 43-47.

Bickel, P. J. and Lehmann, E. L. (1976). Descriptive statistics for nonparametric models. III. Dispersion. Ann. Math. Statist. 4: 1139-1158.

Bickel, P. J. and Yahav, J. A. (1977). On selecting a subset of good populations. Statistical Decision Theory and Related Topics II (Ed. S. S. Gupta and D. S. Moore). Academic Press, New York: 37-55.

Bjørnstad, J. F. (1978). On optimal subset selection procedures. Ph.D. Thesis. Univ. of California, Berkeley, California.

Blum, J. R. and Rosenblatt, J. (1967). On partial information in statistical inference. Ann. Math. Statist. 38: 1671-1678.

Blumenthal, S. and Patterson, D. W. (1969). Rank order procedures for selecting a subset containing the population with the smallest scale parameter. Sankyā Ser. A. 31: 37-42.

Broström, G. (1977). Subset selection with a reliability application. Statistical Research Report No. 1977-2, Institute of Mathematics and Statistics, Univ. of Umeå.

Broström, G. (1978). Sequentially rejective subset selection procedures. Statistical Research Report No. 1978-1, Inst. of Mathematical Statistics, Univ. of Umeå.

Carroll, R. J., Gupta, S. S. and Huang, D. Y. (1975). On selection procedures for the t best populations and some related problems. Comm. Statist. 4: 987-1008.

Chernoff, H. and Yahav, J. (1977). A subset selection problem employing a new criterion. Statistical Decision Theory and Related Topics II (Ed. S. S. Gupta and D. S. Moore). Academic Press, New York: 93-119.

DeRouen, T. A. and Mitchell, T. J. (1974). A G_1-minimax estimator for a linear combination of binomial probabilities. J. Amer. Statist. Assoc. 69: 231-233.

Deely, J. J. and Gupta, S. S. (1968). On the properties of subset selection procedures. Sankhyā Ser. A. 30: 37-50.

Desu, M. M. (1970). A selection problem. Ann. Math. Statist. 41: 1596-1603.

Dunnett, C. W. (1955). A multiple comparison procedure for comparing several treatments with a control. J. Amer. Statist. Assoc. 50: 1096-1121.

Eaton, M. L. (1967). Some optimum properties of ranking procedures. Ann. Math. Statist. 38: 124-137.

Fabian, V. (1962). On multiple decision methods for ranking population means. Ann. Math. Statist. 33: 248-254.

Ferguson, T. S. (1967). Mathematical Statistics: A Decision Theoretic Approach. Academic Press, New York.

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. Mimeo. Ser. No. 150. Inst. of Statist., Univ. of North Carolina, Chapel Hill, North Carolina.

Gupta, S. S. (1965). On some multiple decision (selection and ranking) rules. Technometrics 7: 225-245.

Gupta, S. S. and Huang, D. Y. (1975). On Γ-minimax classification procedures. Proceedings of the 40th session of the international statistical institute, Vol. 46, Book 3: 330-335.

Gupta, S. S. and Huang, D. Y. (1977). On some Γ-minimax subset selection and multiple comparison procedures. Statistical Decision Theory and Related Topics II (Ed. S. S. Gupta and D. S. Moore). Academic Press, New York: 139-155.

Gupta, S. S. and Hsu, J. C. (1978). On the performance of some selection procedures. Comm. Statist. B7(6): 561-591.

Gupta. S. S. and McDonald, G. C. (1970). On some classes of selection procedures based on ranks. Nonparametric Techniques in Statistical Inference (Ed. M. L. Puri), Cambridge Univ. Press, London, 491-514.

Gupta, S. S. and Panchapakesan, S. (1979). Multiple Decision Procedures: Theory and Methodology of Selecting and Ranking Populations. John Wiley, New York. (To be published.)

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. (1962). On selecting a subset containing the population with the smallest variance. Biometrika 49: 495-507.

Gupta, S. S. and Studden, W. J. (1966). On some aspects of selection and ranking procedures with applications. Mimeo. Ser. No. 81, Dept. of Statist., Purdue Univ., West Lafayette, Indiana.

Hardy, G. H., Littlewood, J. E. and Pólya, G. (1952). Inequalities. 2nd Ed., Cambridge University Press, Cambridge.

Hodges, J. L., Jr. and Lehmann, E. L. (1952). The use of previous experience in reaching statistical decisions. Ann. Math. Statist. 23: 396-407.

Hsu, J. (1977). On some decision theoretic contributions to the problem of subset selection. Ph.D. Thesis. Dept. of Statist., Purdue Univ., West Lafayette, Indiana.

Ibragimov, J. A. (1956). On the composition of unimodel distributions (Russian). Teoriya veroyatnostey 1: 283-288.

Jackson, D. A., O'Donovan, T. M., Zimmer, W. J., and Deely, J. J. (1970). G_2-minimax estimators in the exponential family. Biometrika, 57: 439-443.

Lehmann, E. L. (1959). Testing Statistical Hypotheses. John Wiley and Sons, New York.

Lehmann, E. L. (1961). Some model I problems of selection. Ann. Math. Statist. 32: 990-1012.

Lehmann, E. L. (1963). A class of selection procedures based on ranks. Math. Ann. 150: 268-275.

Lehmann, E. L. (1966). On a theorem of Bahadur and Goodman. Ann. Math. Statist. 37: 1-6.

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.

McDonald, G. C. (1977). Subset selection procedures based on sample ranges. Comm. Statist. A6: 1055-1079.

Menges, G. (1966). On the "Bayesification" of the minimax principle. Unternehmesforschung. 10: 81-91.

Miescke, K. J. (1978). Bayesian subset selection for additive and linear loss functions. Mimeo. Ser. No. 78-23, Dept. of Statist., Purdue Univ., West Lafayette, Indiana.

Miescke, K. J. (1979). Γ-minimax selection procedures in simultaneous testing problems. Mimeo. Ser. No. 79-1, Dept. of Statist., Purdue Univ., West Lafayette, Indiana.

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.

Panchapakesan, S. and Santner, T. J. (1977). Subset selection procedures for Δ_p-superior populations. Comm. Statist. A6: 1081-1090.

Paulson, E. (1949). A multiple decision procedure for certain problems in analysis of variance. Ann. Math. Statist. 20: 95-98.

Paulson, E. (1952). On the comparison of several experimental categories with a control. Ann. Math. Statist. 23: 239-246.

Puri, P. S. and Puri, M. L. (1969). Selection procedures based on ranks: Scale parameter case. Sankhyā Ser. A. 30: 291-302.

Randles, R. H. and Hollander, M. (1971). Γ-minimax selection procedures in treatment versus control problems. Ann. Math. Statist. 42: 330-341.

Rizvi, M. H. and Woodworth, G. G. (1970). On selection procedures based on ranks: counter examples concerning the least favorable configurations. Ann. Math. Statist. 41: 1942-1951.

Robbins, H. (1951). Asymptotically subminimax solutions of compound statistical decision problems. Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability. University of California Press, Berkeley and Los Angeles: 131-148.

Santner, T. J. (1975). A restricted subset selection approach to ranking and selection problems. Ann. Statist. 3: 334-339.

Santner, T. J. (1976). A two stage procedure for selecting δ*-optimal means in the normal model. Comm. Statist. A5: 283-292.

Seal, K. C. (1955). On a class of decision procedures for ranking means of normal populations. Ann. Math. Statist. 26: 387-398.

Seeger, P. (1972). A note on partitioning a set of normal populations by their locations with respect to two controls. Ann. Math. Statist. 43: 2019-2023.

Sobel, M. (1967). Nonparametric procedures for selecting the t populations with the largest α-quantiles. Ann. Math. Statist. 38: 1804-1816.

Solomon, D. L. (1972a). Λ-minimax estimation of a multivariate location parameter. J. Amer. Statist. Assoc. 67: 641-646.

Solomon, D. L. (1972b). Λ-minimax estimation of a scale parameter. J. Amer. Statist. Assoc. 67: 647-649.

Stigler, S. M. (1973). The asymptotic distribution of the trimmed mean. Ann. Statist. 1: 474-477.

Studden, W. J. (1967). On selecting a subset of k populations containing the best. Ann. Math. Statist. 38: 1072-1078.

Tong, Y. L. (1969). On partitioning a set of normal populations by their locations with respect to a control. Ann. Math. Statist. 40: 1300-1324.

Wong, W. Y. (1976). Some Contributions to Subset Selection Procedures. Ph.D. Thesis (Mimeo. Ser. No. 454). Dept. of Statist., Purdue Univ., West Lafayette, Indiana.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Mimeograph Series #79-7	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) On Bayes and Gamma-minimax Subset Selection Rules		5. TYPE OF REPORT & PERIOD COVERED Technical
		6. PERFORMING ORG. REPORT NUMBER Mimeo. Series #79-7
7. AUTHOR(s) Woo-Chul Kim		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 May 1979
		13. NUMBER OF PAGES 106
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CL... (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 of Good Populations, Bayes Rule, Minimax Rule, Extended Bayes Rule, Monte Carlo Study, Γ-minimax Rule, Strong Unimodality, Series Systems, Parallel Systems, Reliability, Measure of Dispersion, Asymptotic Relative Efficiency.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Selection and ranking problems arise because the classical tests of homogeneity are often inadequate in practice when the experimenter wishes to make decisions regarding $k(\geq 2)$ populations or treatments. Chapter 1 deals with the problem of selecting 'good' populations from a set of k given populations $\pi_1,\ldots,\pi_k$ where the quality of each population π_i is characterized by an unknown parameter θ_i. A population π_i is called 'good' if $\theta_i \geq \max_{1\leq j\leq k} \theta_j - \Delta$ and

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE UNCLASSIFIED

S/N 0102-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered) 'bad' otherwise. In the preceding Δ is assumed known and is positive. Assuming a permutationally symmetric prior of $(\theta_1,\ldots,\theta_k)$ and the monotone likelihood ratio of the probability density function $f(x,\theta_i)$ associated with π_i, a Bayes rule is derived when a loss function is of a certain type. Also, some properties of the Bayes rule are given when an independently and identically distributed prior is assumed. The rest of Chapter 1 pertains to further simplification and approximation of the Bayes rules when the loss is c_1 for selecting a 'bad' population and c_2 for excluding a 'good' one. It turns out that, in some sense, some rules which are explicitly computable and have been studied in the past provide approximations to the Bayes rule. This is an interesting result since very often Bayes rules are analytically and computationally intractable. Further, it is shown that rules of the type proposed by Gupta (1956) are extended Bayes rules for the case of normal populations problem. Monte Carlo studies are carried out to see how well the rules proposed by Seal (1955) and by Gupta (1956) approximate the Bayes rules in terms of overall risks with respect to normal exchangeable priors.

Chapter 2 deals with the problem of partitioning k treatment populations with regard to a control population. The goal is to partition the k treatment populations into 'better' populations, 'worse' populations or 'close' populations in an optimal way. It is assumed that population π_i is characterized by a parameter θ_i and observable independent random variables X_i has probability density function $f_i(x_i-\theta_i)$ for all $i = 0,1,\ldots,k$, with $f_i(\cdot)$ being strongly unimodal and symmetric. Assuming a specific loss function, Γ-minimax rules are derived when Γ consists of prior distributions under which we can specify the probability of being 'close' to the control and being 'better' or 'worse' than it. Minimax rules are also derived. For the normal means problem, Bayes rules are derived with respect to independent normal priors, and comparisons between Γ-minimax rules and the corresponding Bayes rules are made.

In Chapter 3, a selection problem arising in reliability theory and another for the selection in terms of scale parameters are considered. The first problem deals with the selection of components (units) for parallel and series systems from k populations (brands) with exponentially distributed life-length times. An optimal rule is derived for the series system and the Bayes rule with respect to a natural conjugate prior is derived for the 1-out-of-2 system. The second problem deals with the investigation of the selection procedures based on robust estimators of measures of dispersion for selecting the populations in terms of scale parameters. Large sample solutions and the asymptotic relative efficiencies of the proposed procedures are studied.