SOME CONTRIBUTIONS TO EMPIRICAL BAYES, SEQUENTIAL AND LOCALLY OPTIMAL SUBSET SELECTION RULES*

by

TaChen Liang
Purdue University

Technical Report #84-31

Department of Statistics
Purdue University

August 1984

*This research was supported by the Office of Naval Research Contract N00014-84-0167 at Purdue University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

INTRODUCTION

In many practical situations, the experimenter is faced with the problem of comparing k (≥ 2) populations, and desires to select the best among them. In such situations, the classical tests of homogeneity are inadequate in the sense that they have not been designed to meet this goal and do not provide an answer to this question. Mosteller (1948), Paulson (1949), Bahadur (1950) and Bahadur and Robbins (1950) were among the earliest research workers to recognize this inadequacy. For overcoming this shortcoming, they formulated the problem as a multiple decision problem aimed at the selection and ranking of the k populations. The attempts to formulate the decision problem to answer the realistic goal set the stage for the development of the selection and ranking theory.

In the three and one-half decades since these early studies, ranking and selection problems have become an active area of statistical research. The two main approaches that have been used in formulating a selection and ranking problem are indifference zone approach and subset selection approach. To fix ideas, suppose that there are k (≥ 2) populations $\pi_1, \ldots, \pi_k$, and population π_i is characterized by a parameter θ_i for each $i = 1, \ldots, k$. π_i is said to be better than π_j if $\theta_i > \theta_j$. Consider the problem of selecting the best population, i.e., the population associated with $\theta_{[k]} = \max_{1 \leq i \leq k} \theta_i$. In the

indifference zone approach, due to Bechhofer (1954), the goal is to select one population as the best population. The experimenter is required to specify an "indifference zone" in the parameter space and the procedure determines the smallest sample size so that a certain probability condition is satisfied whenever the unknown parameter lies in the "preference zone." Contributions using this approach in the decision-theoretic framework have been made by Lehmann (1966), Eaton (1967) and Alam (1973), among others. In the subset selection approach pioneered by Gupta (1956, 1965), the goal is to select a nonempty subset of the k populations so that the best population is included in this selected subset with a minimum guaranteed probability $P^*(\frac{1}{k} < P^* < 1)$. The size of the selected subset is not determined in advance but depends on the observed data themselves. Selection of any subset consistent with the goal is called a correct selection (CS) and the probability of a correct selection using a rule R is denoted by $P(CS|R)$. The requirement that $P(CS|R) \geq P^*$ is referred as the P^*-condition. It should be emphasized that in the subset selection framework, there is no indifference zone specification. The P^*-condition must be satisfied whatever be the configuration of the unknown parameters. Decision-theoretic contributions in this framework have been made by Studden (1967), Deely and Gupta (1968), Bickel and Yahav (1977), Chernoff and Yahav (1977), Goel and Rubin (1977), Gupta and Hsu (1978), Miescke (1979) and Bjørnstad (1981) in the fixed sample case and by Gupta and Miescke (1984a, b) in two-stage or sequential sampling cases. When the prior distributions of the parameters of interest are known, some Bayes rules have been studied

by Deely and Gupta (1968), Bickel and Yahav (1977), Chernoff and Yahav (1977), Goel and Rubin (1977), Gupta and Hsu (1978) and Gupta and Miescke (1984a). When the prior distributions are unknown, Deely (1965), Singh (1977), Gupta and Hsiao (1983) and Gupta and Leu (1983) formulated some selection problems in terms of empirical Bayes framework.

There have also been attempts in the literature to study the performance of some existing selection rules and to construct optimal selection rules. For the class of subset selection rules satisfying the P*-condition, Berger (1979) investigated minimaxity taking the size of selected subset as the loss function. Berger and Gupta (1980) obtained minimax rules in the class of nonrandomized, just and invariant rules when the risk is measured by the maximum probability of including a nonbest population. Bjørnstad (1980) compared three minimax rules for the normal means problem. Gupta and Huang (1977) derived a Γ-minimax procedure for selecting the best population. Gupta and Kim (1980) considered some minimax and Γ-minimax rules for partitioning k populations in comparison with a standard or a control. Gupta, Huang and Nagel (1979) investigated some locally optimal subset selection rules based on ranks. Recently, Huang and Panchapakesan (1982) and Huang, Panchapakesan and Tseng (1984) also derived other locally optimal subset selection rules with different optimality criteria.

The present thesis consists of investigations of some multiple decision (selection and ranking) problems. Some contributions to the empirical Bayes rules, sequential subset selection procedures and

locally optimal subset selection rules have been made under the subset selection framework.

Chapter I deals with the problem of selecting good populations through the empirical Bayes approach. The empirical Bayes approach in statistical theory is appropriate when one is confronted repeatedly and independently with the same decision problem. In such instances, it is reasonable to formulate the component problem in the sequence as Bayes decision problems with respect to an unknown prior distribution on the parameter space, and then use the accumulated observations to improve the decision rule at each stage. This approach is due to Robbins (1955, 1964). Many such empirical Bayes rules have been shown to be asymptotically optimal in the sense that the risk for the nth decision problem converges to the optimal Bayes risk which would have been obtained if the prior distribution was known and the Bayes rule with respect to this prior distribution was used. Two selection problems have been studied in Chapter I: selecting populations better than a control or a standard and selecting all good populations among k populations. For the problem of selecting populations better than a control or a standard, it is assumed that the populations are binomially distributed. For the problem of selecting all good populations among k populations, we assume that the populations have Pareto distributions. For each problem, a nonrandomized Bayes rule is derived for a linear loss function. Later, this Bayes rule is represented in terms of the marginal probability density function of the random observations. Based on this Bayes rule, a sequence of empirical

Bayes rules for selecting good populations is derived. For each problem, the rate of convergence of the sequence of the empirical Bayes rules is also studied. For the problem of selecting populations better than a control or a standard, two sequences of smoothed decision rules have also been studied when the prior distribution is symmetrical about $P = \frac{1}{2}$, but its functional form is still unknown. Some Monte Carlo studies have been carried out. The results indicate that the smoothed competitors actually perform better than the original ones.

Chapter II deals with the problem of selecting the best population through sequential subset selection approach. In most fixed sample size cases, the subset selection rules always select a nonempty subset containing the best population with guaranteed probability P^* ($\frac{1}{k} < P^* < 1$). However, those rules do not offer inference about the measure of separation between each selected population and the unknown best population. It may happen that a poor population may be contained in the selected subset. Recently, Hsu (1981, 1982) and Hsu and Edwards (1983) studied some methods to derive simultaneous upper confidence intervals for all measures of separation between each population and the unknown best population under the location model. This motivates us to study selection rules such that, with guaranteed probability P^*, not only the best population is selected, but also, one can state that separation between each selected population and the unknown best population is bounded by some prespecified value. In Chapter II, some sequential subset selection procedures achieving the goal described above are derived. Basically, these procedures are formed

by choosing an invariant statistic of the parameter of interest, based on the observations from each pair of two of the k populations and by performing a modified sequential probability ratio test (MPRST) based on this statistic. This is done simultaneously for all pairs of populations and if a particular MSPRT terminates, then an appropriate population is removed from the set of contending populations. This is continued until only one population belongs to this set or some statistical evidence indicates that all the populations remaining in this set are not far from the unknown best population. At each stage, these procedures also provide some statistical inference about the bounds on the measure of separation between each remaining population and the unknown best population. In Section 2.3, we study the problem of selecting the best among k populations belonging to the exponential family of distributions. We use the same measure of separation as that considered by Bechhofer, Kiefer and Sobel (1968). For this particular measure of separation, we consider an appropriate transformation of the random observations taken from any two populations. With this transformation, the likelihood function of the new statistics can be factored into two parts, one of which, obtained by a conditional argument, and termed the conditional likelihood function, is a function only of the parameter of interest. Based on this conditional likelihood function, a sequential subset selection procedure is derived. This sequential subset selection procedure achieves the goal described above. At each stage, it also provides some statistical inference about the bounds on the measure of separation between each remaining population and the unknown best population.

In practical situations, it sometimes happens that the order of the observations are easily obtained whereas the actual measurements themselves are not available due to excessive cost or physical constraints. In a problem of this type, one may desire to investigate decision rules based on ranks. Gupta and McDonald (1970) studied three classes of subset selection rules based on ranks for selecting a subset containing the best among k populations under the situation that the underlying distributions are unknown. When the form of the underlying distributions is known, but its value depends on some unknown parameter, Gupta, Huang and Nagel (1979) and Huang and Panchapakesan (1982) studied the problem of deriving some subset selection rules, based on ranks, which are locally optimal in some sense. All the studies mentioned above only considered the situation where the ranks are completely observed. Due to the design reasoning or cost consideration, it sometimes happens that ranks are only partially observed according to some censoring scheme. In Chapter III, we assume that the form of the underlying distribution is known, but there exists some unknown parameter. Based on the partial ranks, our goal is to derive locally optimal subset selection rules for selecting a subset containing the best population. Problems are formulated according to whether the sample sizes from the k populations are equal or not. Locally optimal subset selection rules R_1 (for the equal sample sizes case) and R_2 (for the unequal sample sizes case) are derived and some local monotonicity of R_1 and R_2 are discussed. Finally, a class of compatible censoring schemes is considered. We

observe that the properties of local optimality of R_1 or R_2 can be extended to those locally optimal rules which are based on the partial rank configurations censored by any compatible censoring scheme.

CHAPTER I

EMPIRICAL BAYES RULES FOR SELECTING GOOD POPULATIONS

1.1. Introduction

The empirical Bayes approach in statistical decision theory is appropriate when one is confronted repeatedly and independently with the same decision problem. In such instances, it is reasonable to formulate the component problem in the sequence as Bayes decision problems with respect to an unknown prior distribution on the parameter space and then use the accumulated observations to improve the decision rule at each stage. This approach is due to Robbins (1955, 1964). Many such empirical Bayes rules have been shown to be asymptotically optimal in the sense that the risk for the nth decision problem converges to the optimal Bayes risk which would have been obtained if the prior distribution was known and the Bayes rule with respect to this prior distribution was used.

Empirical Bayes rules have been derived for multiple decision problems by Deely (1965). He considered selecting a subset containing the best population. Van Ryzin (1970), Huang (1975), Van Ryzin and Susarla (1977) and Singh (1977) also studied some multiple decision problem by using empirical Bayes approach. Recently, Gupta and Hsiao (1983) studied some empirical Bayes rules for selecting good populations with respect to a standard or a control. In their paper, the

underlying population π_i is uniformly distributed with parameter θ_i, $i = 0, 1, \ldots, k$, and π_0 is a control population. π_i is said to be good if $\theta_i \geq \theta_0$ and to be bad if $\theta_i < \theta_0$. Let $a \subseteq \{1, 2, \ldots, k\}$ be an action. With the loss function $L(\underset{\sim}{\theta}, a) = \sum_{j \in a} (\theta_0 - \theta_j) I_{(0, \theta_0)}(\theta_j) + \sum_{j \notin a} (\theta_j - \theta_0) I_{(\theta_0, \infty)}(\theta_j)$, where $\underset{\sim}{\theta} = (\theta_0, \theta_1, \ldots, \theta_k)$, they proposed some empirical Bayes rules for the problem of selecting good populations with respect to a standard or a control.

For a similar problem, if the underlying populations have binomial distribution, then, in general, it is hard or impossible to find a sequence of empirical Bayes rules (see Robbins (1964), Samuel (1963) and Singh (1977)). In Section 1.2, we are concerned with this problem. Two cases have been studied: one is that the prior distribution is completely unknown and the other is that the prior distribution is symmetrical about $p = \frac{1}{2}$, but its form is still unknown. In each case, empirical Bayes rules are derived and the rate of convergence of corresponding empirical Bayes rules is also studied. In each case, the order of the rate of convergence is $O(\exp(-c_i n))$ for some $c_i > 0$, $i = 1, 2$. For the case when the prior distribution is symmetrical about $p = \frac{1}{2}$, in order to improve the performance of the sequence of empirical Bayes rules, two smoothing methods are studied. Some Monte Carlo studies have also been carried out. The results indicate that the smoothed competitors actually perform better than the original one.

Suppose now, an experimenter has k (≥ 2) different treatments. He is interested in finding those populations among the k treatments

which are good. In Section 1.3, we consider comparison among k Pareto populations. Let $\theta_i > 0$ be the location parameters of Pareto population π_i. Let $\theta_{[1]} \leq \theta_{[2]} \leq \cdots \leq \theta_{[k]}$ denote the ordered parameters of $\theta_1, \ldots, \theta_k$. For a given $\Delta \geq 0$, population π_i is said to be good if $\theta_i < \theta_{[1]} + \Delta$, and to be bad if $\theta_i \geq \theta_{[1]} + \Delta$. We are interested in selecting all good populations. When $\Delta = 0$, the problem is similar to that considered by Deely (1965), Van Ryzin (1970) and Van Ryzin and Susarla (1977). However, when $\Delta > 0$, the problem becomes more difficult since the determination of the Bayes rule depends on the value $\int_{\Omega} \theta_{[1]} \, f(\underset{\sim}{x}|\underset{\sim}{\theta})dG(\underset{\sim}{\theta})$ which is hard to evaluate when the prior distribution G is completely unknown. In Subsection 1.3.2, a Bayes rule for this decision problem is derived and this Bayes rule is represented as a function of the marginal density functions of the random observations. Corresponding to this Bayes rule, a sequence of empirical Bayes rules is derived in Subsection 1.3.3. Finally, the rate of convergence of the empirical Bayes rules is studied in Subsection 1.3.4.

1.2. Empirical Bayes Rules for Selecting Good Populations--Binomial Populations Case

Let $\pi_0, \pi_1, \ldots, \pi_k$ denote k + 1 populations and let X_i be a random observation from π_i. Assume that $X_i \sim B(N_i, p_i)$, where $p_i \in (0, 1)$ and N_i is fixed and known. Let π_0 be the control population. For each i = 1, ..., k, population π_i is said to be good if $p_i \geq p_0$ and to be bad if $p_i < p_0$, where the control parameter p_0 is either known or unknown. Our goal is to derive some empirical Bayes

rules to select all the good populations and exclude all the bad populations.

In the following, Subsection 1.2.1 deals with the formulation of empirical Bayes approach for our selection problem. In Subsection 1.2.2, we consider the case when the prior distribution is completely unknown. In Subsection 1.2.3, we consider the case when the prior distribution is partially known. In each case, empirical Bayes rules are derived. The rates of convergence of these empirical Bayes rules are also studied in the above subsections. For the case when the prior distribution is symmetrical about $p = \frac{1}{2}$, in order to improve the performance of the sequence of empirical Bayes rules, two smoothing methods are studied in Subsection 1.2.4. In Subsection 1.2.5, Monte Carlo studies are carried out for some specified prior distributions. The results indicate the smoothed competitors always perform better than the original empirical Bayes rules.

1.2.1. Formulation of the Empirical Bayes Approach

When the control parameter p_0 is known, the empirical Bayes framework can be formulated as follows:

(1) Let $\Omega = \{\underset{\sim}{p} | \underset{\sim}{p} = (p_1, \ldots, p_k), p_i \in (0, 1) \text{ for } i = 1, 2, \ldots, k\}$. For each $\underset{\sim}{p} \in \Omega$, define $A(\underset{\sim}{p}) = \{i | p_i \geq p_0\}$, $B(\underset{\sim}{p}) = \{i | p_i < p_0\}$. That is, $A(\underset{\sim}{p})(B(\underset{\sim}{p}))$ is the set of indices of good (bad) populations.

(2) Let $\mathcal{A} = \{a | a \sqsubset \{1, 2, \ldots, k\}\}$ be the action space. When action a is taken, it means that population π_i is selected as a good population if $i \in a$, and excluded as a bad population if $i \notin a$.

(3) The loss function $L(\underset{\sim}{p}, a)$ is defined as follows:

$$(1.2.1) \qquad L(\underset{\sim}{p}, a) = \sum_{i \in A(\underset{\sim}{p})-a} (p_i - p_0) + \sum_{i \in a-A(\underset{\sim}{p})} (p_0 - p_i)$$

where the first summation is the loss due to not selecting some good populations and the second summation is the loss due to selecting some bad populations.

(4) Let $dG(\underset{\sim}{p}) = \prod_{i=1}^{k} dG_i(p_i)$ be the prior distribution over the parameter space Ω, where $G_i(\cdot)$ are unknown for all $i = 1, 2, \ldots, k$.

(5) For each i, let (X_{ij}, P_{ij}), $j = 1, 2, \ldots$, be pairs of random variables associated with population π_i, where X_{ij} is observable but P_{ij} is not observable. P_{ij} has distribution G_i. Conditional on $P_{ij} = p_{ij}$, X_{ij} is binomially distributed with parameters N_i and p_{ij}. Some additional observations $\underset{\sim}{Y}_{ij} = (Y_{ij1}, \ldots, Y_{ij\gamma_i})$ are also available. Conditional on $P_{ij} = p_{ij}$, X_{ij} and Y_{ijm}, $m = 1, 2, \ldots, \gamma_i$, are independent and identically distributed. The jth stage observations are denoted by $\underset{\sim}{Z}_j$. That is, $\underset{\sim}{Z}_j = ((X_{1j}, \underset{\sim}{Y}_{1j}), \ldots, (X_{kj}, \underset{\sim}{Y}_{kj}))$.

(6) Let $\underset{\sim}{X} = (X_1, \ldots, X_k)$ be the present observation. Conditional on $\underset{\sim}{p} = (p_1, \ldots, p_k)$, $\underset{\sim}{X}$ has probability function

$$(1.2.2) \qquad f(\underset{\sim}{x}|\underset{\sim}{p}) = \prod_{i=1}^{k} f_i(x_i|p_i) = \prod_{i=1}^{k} \binom{N_i}{x_i} p_i^{x_i} (1 - p_i)^{N_i - x_i}.$$

Finally, since we are interested in Bayes rule, we can restrict our attention to the nonrandomized rules.

(7) Let $D = \{d | d = \chi \to \mathcal{A}$, being measurable$\}$ be the set of nonrandomized rules, where $\chi = \prod_{i=1}^{k} \{0, 1, \ldots, N_i\}$. For each $d \in D$, let $r(G, d)$ denote the associated Bayes risk. Then, $r(G) = \inf_{d \in D} r(G, d)$ is the minimum Bayes risk.

When the control parameter p_0 is unknown, for the related framework, the indices in the associated notations should begin at 0 instead of at 1. In the sequel, (0) will be used to show this additional fact.

We now consider decision rules $d_n(\underset{\sim}{x}, \underset{\sim}{Z}_1, \ldots, \underset{\sim}{Z}_n)$ whose form depends on $\underset{\sim}{x}$ and $\underset{\sim}{Z}_j$, $j = 1, \ldots, n$.

Definition 1.2.1. A sequence of decision rules $\{d_n(\underset{\sim}{x}, \underset{\sim}{Z}_1, \ldots, \underset{\sim}{Z}_n)\}_{n=1}^{\infty}$ is said to be asymptotically optimal or empirical Bayes relative to the prior distribution G if

$$(1.2.3) \qquad r_n(G, d_n) \equiv \sum_{\underset{\sim}{x} \in \chi} E \int_{\Omega} L(\underset{\sim}{p}, d_n(\underset{\sim}{x}, \underset{\sim}{Z}_1, \ldots, \underset{\sim}{Z}_n)) \cdot f(\underset{\sim}{x} | \underset{\sim}{p})\, dG(\underset{\sim}{p}) \to r(G) \text{ as } n \to \infty.$$

The expectation E in (1.2.3) is taken with respect to $(\underset{\sim}{Z}_1, \ldots, \underset{\sim}{Z}_n)$. For simplicity, $d_n(\underset{\sim}{x}, \underset{\sim}{Z}_1, \ldots, \underset{\sim}{Z}_n)$ will be denoted by $d_n(\underset{\sim}{x})$.

For constructing a sequence of empirical Bayes rules, we first need to find the minimum Bayes risk and the associated Bayes rule, say d_G. From (1.2.1), the Bayes risk with associated decision rule d is

$$
\begin{aligned}
r(G, d) &= \int_{\Omega} \sum_{\underset{\sim}{x}\in\chi} L(\underset{\sim}{p}, d(\underset{\sim}{x}))\, f(\underset{\sim}{x}|\underset{\sim}{p})\, dG(\underset{\sim}{p}) \\
&= \sum_{\underset{\sim}{x}\in\chi} \int_{\Omega} L(\underset{\sim}{p}, d(\underset{\sim}{x}))\, f(\underset{\sim}{x}|\underset{\sim}{p})\, dG(\underset{\sim}{p}) \\
&= \sum_{\underset{\sim}{x}\in\chi} \int_{\Omega} \Bigg[\sum_{i\in A(\underset{\sim}{p})-d(\underset{\sim}{x})} (p_i - p_0) \\
&\quad + \sum_{i\in d(\underset{\sim}{x})-A(\underset{\sim}{p})} (p_0 - p_i) \Bigg] f(\underset{\sim}{x}|\underset{\sim}{p})\, dG(\underset{\sim}{p}) \qquad (1.2.4) \\
&= \sum_{\underset{\sim}{x}\in\chi} \int_{\Omega} \Bigg[\sum_{i\in d(\underset{\sim}{x})} (p_0 - p_i) \\
&\quad + \sum_{i=1}^{k} (p_i - p_0) I_{(p_0,1)}(p_i) \Bigg] f(\underset{\sim}{x}|\underset{\sim}{p})\, dG(\underset{\sim}{p}) \\
&= \sum_{\underset{\sim}{x}\in\chi} \sum_{i\in d(\underset{\sim}{x})} \int_{\Omega} (p_0 - p_i)\, f(\underset{\sim}{x}|\underset{\sim}{p})\, dG(\underset{\sim}{p}) \\
&\quad + \sum_{\underset{\sim}{x}\in\chi} \sum_{i=1}^{k} \int_{\Omega} (p_i - p_0) I_{(p_0,1)}(p_i) f(\underset{\sim}{x}|\underset{\sim}{p})\, dG(\underset{\sim}{p}),
\end{aligned}
$$

where $I_A(\cdot)$ is the indicator function of set A.

The second term in the right-hand side of (1.2.4) is a constant and does not affect the determination of the Bayes rule.

Let $\phi_{iG}(\underset{\sim}{x}) = \int_{\Omega} (p_0 - p_i) f(\underset{\sim}{x}|\underset{\sim}{p}) dG(\underset{\sim}{p})$. After integration, one obtains

$$(1.2.5) \qquad \phi_{iG}(\underset{\sim}{x}) = \begin{cases} \left[W_0(x_0) f_i(x_i) - W_i(x_i) f_0(x_0)\right] \prod_{\substack{j=1 \\ j\neq i}}^{k} f_j(x_j) & \\ & \text{if } p_0 \text{ is unknown,} \\ \left[p_0 f_i(x_i) - W_i(x_i)\right] \prod_{\substack{j=1 \\ j\neq i}}^{k} f_j(x_j) & \text{if } p_0 \text{ is known.} \end{cases}$$

where

$$(1.2.6) \qquad f_i(x) = \int_0^1 f_i(x|p)\, dG_i(p) \qquad \text{and}$$

$$(1.2.7) \qquad W_i(x) = \int_0^1 p f_i(x|p)\, dG_i(p) = \int_0^1 \binom{N_i}{x} p^{x+1} (1-p)^{N_i - x}\, dG_i(p).$$

Since $f_i(x)$, the marginal probability function of X_i, is always positive for all $x = 0, 1, \ldots, N_i$, $i = (0), 1, \ldots, k$, the Bayes rule d_G can be obtained as follows:

$$(1.2.8) \qquad d_G(\underset{\sim}{x}) = \{i | \Delta_{iG}(\underset{\sim}{x}) \leq 0\},$$

where

$$(1.2.9) \qquad \Delta_{iG}(\underset{\sim}{x}) = \begin{cases} W_0(x_0) f_i(x_i) - W_i(x_i) f_0(x_0) & \text{if } p_0 \text{ is unknown,} \\ & \\ p_0 f_i(x_i) - W_i(x_i) & \text{if } p_0 \text{ is known.} \end{cases}$$

Now, for each $i = (0), 1, \ldots, k$, and for each $n = 1, 2, \ldots$, let $W_{in}(x_i) \equiv W_{in}(x_i; (X_{i1}, \underset{\sim}{Y}_{i1}), \ldots, (X_{in}, \underset{\sim}{Y}_{in}))$ be an estimator of $W_i(x_i)$ and $f_{in}(x_i) \equiv f_{in}(x_i; (X_{i1}, \underset{\sim}{Y}_{i1}), \ldots, (X_{in}, \underset{\sim}{Y}_{in}))$ be an estimator of $f_i(x_i)$. Define

$$\Delta_{in}(\underset{\sim}{x}) = \begin{cases} W_{0n}(x_0)f_{in}(x_i) - W_{in}(x_i)f_{0n}(x_0) & \text{if } p_0 \text{ is unknown} \\ p_0 f_{in}(x_i) - W_{in}(x_i) & \text{if } p_0 \text{ is known.} \end{cases} \tag{1.2.10}$$

Then, define

$$d_n(\underset{\sim}{x}) = \{i \mid \Delta_{in}(\underset{\sim}{x}) \leq 0\}. \tag{1.2.11}$$

If $W_{in}(x) \xrightarrow{P} W_i(x)$ and $f_{in}(x) \xrightarrow{P} f_i(x)$ for all $x = 0, 1, \ldots, N_i$, where "$\xrightarrow{P}$" means convergence in probability, then $\Delta_{in}(\underset{\sim}{x}) \xrightarrow{P} \Delta_{iG}(\underset{\sim}{x})$ for all $\underset{\sim}{x} \in \chi$. Therefore, from a corollary of Robbins (1964), it follows that $r_n(G, d_n) \to r(G)$ as $n \to \infty$. So, the sequence of decision rules $\{d_n(\underset{\sim}{x})\}$ defined in (1.2.11) is asymptotically optimal for our selection problem. Hence, in the following, all we have to do is to find sequences of estimators, say $\{W_{in}(x)\}$ and $\{f_{in}(x)\}$, $i = (0), 1, \ldots, k$, satisfying $W_{in}(x) \xrightarrow{P} W_i(x)$ and $f_{in}(x) \xrightarrow{P} f_i(x)$ for all $x = 0, 1, \ldots, N_i$.

1.2.2. A Case when the Prior Distribution is Completely Unknown

In this subsection, we study the case when the prior distribution G is completely unknown. Robbins (1964) and Samuel (1963), respectively, pointed out that there was no way of approximating $W_i(x)$ just by using the observations $(X_{i1}, \ldots, X_{in})$. In order to remedy this deficiency, we take, at each stage, some more observations $(Y_{ij1}, \ldots, Y_{ij\gamma_i})$ in our model where γ_i can be any positive integer. For simplicity, let $\gamma_i = 1$ for all i.

Estimation of $W_i(x)$ and $f_i(x)$

A usual estimator of $f_i(x)$ can be given as follows:

$$(1.2.12) \qquad f_{in}(x) = \frac{1}{n} \sum_{j=1}^{n} I_{\{x\}}(X_{ij}) \quad \text{for } x = 0, 1, \ldots, N_i.$$

Then $f_{in}(x)$ is an unbiased estimator of $f_i(x)$, and by the strong law of large numbers, $f_{in}(x) \to f_i(x)$ with probability 1 for each $x = 0, 1, \ldots, N_i$. Hence, $f_{in}(x) \xrightarrow{P} f_i(x)$ for all $x = 0, 1, \ldots, N_i$.

For the estimation of $W_i(x)$, we consider the following. Define

$$(1.2.13) \qquad V_{ij}(x) = Y_{ij}\, I_{\{x\}}(X_{ij}).$$

Then,

$$
\begin{aligned}
E[V_{ij}(x)] &= E[Y_{ij}\, I_{\{x\}}(X_{ij})] \\
&= E_{G_i}\left[E[Y_{ij}\, I_{\{x\}}(X_{ij})|B(N_i, P)]\right] \\
&= E_{G_i}\left[E[Y_{ij}|B(N_i, P)]E[I_{\{x\}}(X_{ij})|B(N_i, P)]\right] \\
&= E_{G_i}\left[N_i P\binom{N_i}{x}P^x(1-P)^{N_i-x}\right] \\
&= N_i \int_0^1 \binom{N_i}{x}P^{x+1}(1-P)^{N_i-x}\, dG_i(P) \\
&= N_i\, W_i(x),
\end{aligned}
\tag{1.2.14}
$$

where $E[h(X)|B(N, P)]$ denotes the expectation of $h(X)$ when $X \sim B(N, P)$. In (1.2.14), the second and third equalities are obtained from condition (5) of Subsection 1.2.1.

We then define

$$
W_{in}(x) = \frac{1}{n}\sum_{j=1}^{n} V_{ij}(x)/N_i\,. \tag{1.2.15}
$$

Since $V_{ij}(x)$, $j=1,2,\ldots$, are i.i.d. and bounded, it is easy to show that $W_{in}(x) \to W_i(x)$ with probability 1 for all $x = 1, 0, \ldots, N_i$. Now, let $\Delta_{in}(\underset{\sim}{x})$ and $d_n(\underset{\sim}{x})$ be defined as in (1.2.10) and (1.2.11), respectively. From the discussion of Subsection 1.2.1 and the construction of the sequence of decision rules $\{d_n\}_{n=1}^{\infty}$ through (1.2.10) ~ (1.2.12) and (1.2.15), we get the following result.

Theorem 1.2.2. For our decision problem, the sequence of decision rules $\{d_n\}_{n=1}^{\infty}$ is asymptotically optimal relative to the prior distribution G.

Rate of Convergence of Empirical Bayes Rules $\{d_n\}$

Let $\{\delta_n\}_{n=1}^{\infty}$ be a sequence of empirical Bayes rules relative to the prior distribution G. Since the Bayes rule d_G achieves the minimum Bayes risk r(G) relative to G, $r_n(G, \delta_n) - r(G) \geq 0$ for all $n = 1, 2, \ldots$. Thus, the nonnegative difference $r_n(G, \delta_n) - r(G)$ is used as a measure of the optimality of the sequence of empirical Bayes rules $\{\delta_n\}_{n=1}^{\infty}$.

Definition 1.2.3. The sequence of empirical Bayes rules $\{\delta_n\}_{n=1}^{\infty}$ is said to be asymptotically optimal at least of order α_n relative to G if $r_n(G, \delta_n) - r(G) \leq O(\alpha_n)$ as $n \to \infty$ where $\lim_{n\to\infty} \alpha_n = 0$.

In the following, we want to evaluate the behavior of the sequence of empirical Bayes rules $\{d_n\}$ described in Theorem 1.2.2.

For each $i = 1, \ldots, k$; $n = 1, 2, \ldots$, define

$$S_i = \{\underset{\sim}{x} \epsilon \chi | \Delta_{iG}(\underset{\sim}{x}) < 0\},$$

$$T_i = \{\underset{\sim}{x} \epsilon \chi | \Delta_{iG}(\underset{\sim}{x}) > 0\},$$

$$S_{in} = \{\underset{\sim}{x} \epsilon \chi | \Delta_{in}(\underset{\sim}{x}) \leq 0\},$$

where $\Delta_{iG}(\underset{\sim}{x})$ and $\Delta_{in}(\underset{\sim}{x})$ are defined in (1.2.9) and (1.2.10), respectively. Let

$$\varepsilon_1 = \min_{\substack{\underset{\sim}{x}\in S_i \\ 1\le i\le k}} (-\Delta_{iG}(\underset{\sim}{x})), \quad \varepsilon_2 = \min_{\substack{\underset{\sim}{x}\in T_i \\ 1\le i\le k}} (\Delta_{iG}(\underset{\sim}{x})) \text{ and } \varepsilon = \min(\varepsilon_1, \varepsilon_2).$$

Since χ is a finite space, therefore $\varepsilon > 0$. Now,

$$0 \le r_n(G, d_n) - r(G)$$

$$= \sum_{\underset{\sim}{x}\in\chi} \left\{ E\left[\sum_{i\in d_n(\underset{\sim}{x})} \Delta_{iG}(\underset{\sim}{x}) \prod_{\substack{j=1 \\ j\neq i}}^{k} f_j(x_j) \right] - \sum_{i\in d_G(\underset{\sim}{x})} \Delta_{iG}(\underset{\sim}{x}) \prod_{\substack{j=1 \\ j\neq i}}^{k} f_j(x_j) \right\}$$

$$= \sum_{\underset{\sim}{x}\in\chi} \left\{ E\left[\sum_{i=1}^{k} \Delta_{iG}(\underset{\sim}{x}) I_{S_{in}}(\underset{\sim}{x}) \prod_{\substack{j=1 \\ j\neq i}}^{k} f_j(x_j) \right] \right.$$

$$\left. - \sum_{i=1}^{k} \Delta_{iG}(\underset{\sim}{x}) I_{S_i}(\underset{\sim}{x}) \prod_{\substack{j=1 \\ j\neq i}}^{k} f_j(x_j) \right\} \qquad (1.2.16)$$

$$= \sum_{\underset{\sim}{x}\in\chi} \sum_{i=1}^{k} \Delta_{iG}(\underset{\sim}{x}) \left[E\, I_{S_{in}}(\underset{\sim}{x}) - I_{S_i}(\underset{\sim}{x}) \right] \prod_{\substack{j=1 \\ j\neq i}}^{k} f_j(x_j)$$

$$= \sum_{i=1}^{k} \sum_{\underset{\sim}{x}\in S_i} (-1)\Delta_{iG}(\underset{\sim}{x}) P\{\Delta_{in}(\underset{\sim}{x}) > 0\} \prod_{\substack{j=1 \\ j\neq i}}^{k} f_j(x_j)$$

$$+ \sum_{i=1}^{k} \sum_{\underset{\sim}{x}\in T_i} \Delta_{iG}(\underset{\sim}{x}) P\{\Delta_{in}(\underset{\sim}{x}) \le 0\} \prod_{\substack{j=1 \\ j\neq i}}^{k} f_j(x_j)$$

$$\leq \sum_{i=1}^{k} \left\{ \sum_{\underset{\sim}{x}\in S_i} P\{\Delta_{in}(\underset{\sim}{x}) > 0\} + \sum_{\underset{\sim}{x}\in T_i} P\{\Delta_{in}(\underset{\sim}{x}) \leq 0\} \right\}$$

where the last inequality is due to the fact that $0 < f_j(x_j) \leq 1$ and $|\Delta_{iG}(\underset{\sim}{x})| \leq 1$. From (1.2.16), it suffices to consider the behavior of $P\{\Delta_{in}(\underset{\sim}{x}) > 0\}$ when $\underset{\sim}{x} \in S_i$ and that of $P\{\Delta_{in}(\underset{\sim}{x}) < 0\}$ when $\underset{\sim}{x} \in T_i$ as $n \to \infty$ for each $i = 1, 2, \ldots, k$.

For each $\underset{\sim}{x} \in S_i$,

$$P\{\Delta_{in}(\underset{\sim}{x}) > 0\} = P\{\Delta_{in}(\underset{\sim}{x}) - \Delta_{iG}(\underset{\sim}{x}) > -\Delta_{iG}(\underset{\sim}{x})\}$$

$$\leq P\{\Delta_{in}(\underset{\sim}{x}) - \Delta_{iG}(\underset{\sim}{x}) > \varepsilon\}$$

(1.2.17)

$$= \begin{cases} P\{W_{0n}(x_0)f_{in}(x_i) - W_{in}(x_i)f_{0n}(x_0) - W_0(x_0)f_i(x_i) \\ \quad + W_i(x_i)f_0(x_0) > \varepsilon\} \quad \text{when } p_0 \text{ is unknown,} \\ P\{p_0 f_{in}(x_i) - W_{in}(x_i) - p_0 f_i(x_i) + W_i(x_i) > \varepsilon\} \\ \qquad\qquad \text{when } p_0 \text{ is known.} \end{cases}$$

When p_0 is known,

$$P\{\Delta_{in}(\underset{\sim}{x}) > 0\} \leq P\{p_0 f_{in}(x_i) - p_0 f_i(x_i) - W_{in}(x_i) + W_i(x_i) > \varepsilon\}$$

$$\leq P\{p_0 f_{in}(x_i) - p_0 f_i(x_i) > \frac{\varepsilon}{2}$$

(1.2.18)

$$\text{or } W_{in}(x_i) - W_i(x_i) < -\frac{\varepsilon}{2}\}$$

$$\leq P\{p_0 f_{in}(x_i) - p_0 f_i(x_i) > \frac{\varepsilon}{2}\}$$

$$+ P\{W_{in}(x_i) - W_i(x_i) < -\frac{\varepsilon}{2}\}$$

$$\leq P\{f_{in}(x_i) - f_i(x_i) > \frac{\varepsilon}{2}\}$$

$$+ P\{W_{in}(x_i) - W_i(x_i) < -\frac{\varepsilon}{2}\}$$

since $p_0 \in (0, 1)$.

When p_0 is unknown,

$$P\{\Delta_{in}(\underset{\sim}{x}) > 0\} \leq P\{W_{0n}(x_0)f_{in}(x_i) - W_{in}(x_i)f_{0n}(x_0) - W_0(x_0)f_i(x_i)$$

$$+ W_i(x_i)f_0(x_0) > \varepsilon\}$$

$$\leq P\{W_{0n}(x_0)f_{in}(x_i) - W_0(x_0)f_i(x_i) > \frac{\varepsilon}{2}\}$$

$$+ P\{W_{in}(x_i)f_{0n}(x_0) - W_i(x_i)f_0(x_0) < -\frac{\varepsilon}{2}\}$$

(1.2.19)
$$\leq P\{f_{in}(x_i)[W_{0n}(x_0) - W_0(x_0)] > \frac{\varepsilon}{4}\}$$

$$+ P\{W_0(x_0)[f_{in}(x_i) - f_i(x_i)] > \frac{\varepsilon}{4}\}$$

$$+ P\{f_{0n}(x_0)[W_{in}(x_i) - W_i(x_i)] < -\frac{\varepsilon}{4}\}$$

$$+ P\{W_i(x_i)[f_{0n}(x_0) - f_0(x_0)] < -\frac{\varepsilon}{4}\}$$

$$\leq P\{W_{0n}(x_0) - W_0(x_0) > \tfrac{\varepsilon}{4}\} + P\{f_{in}(x_i) - f_i(x_i) > \tfrac{\varepsilon}{4}\}$$

$$+ P\{W_{in}(x_i) - W_i(x_i) < -\tfrac{\varepsilon}{4}\} + P\{f_{0n}(x_0) - f_0(x_0) < -\tfrac{\varepsilon}{4}\}.$$

In (1.2.19), the last inequality is due to the fact that $0 \leq W_i(x_i)$, $f_i(x_i) \leq 1$ and that $0 \leq W_{in}(x_i)$, $f_{in}(x_i) \leq 1$ where the latter can be easily checked from (1.2.12), (1.2.13) and (1.2.15).

(1.2.18) and (1.2.19) show that it suffices to consider the behavior of $P\{|f_{in}(x_i) - f_i(x_i)| > \delta\}$ and $P\{|W_{in}(x_i) - W_i(x_i)| > \delta\}$ for some $\delta > 0$. For this purpose, we note that Bernstein's inequality is useful.

Bernstein's Inequality (see Ibragimov and Linnik (1971), page 169). Let $Z_1, Z_2, \ldots,$ be a sequence of independent random variables with mean 0 and variances $\mathrm{Var}(Z_i) = \beta_i$ for $i = 1, 2, \ldots$. Write $B_n = \beta_1 + \cdots + \beta_n$ and $S_n = Z_1 + \cdots + Z_n$.

Lemma (Bernstein's inequality). Suppose that, for some $H > 0$, and integer $m \geq 2$, $E[Z_i^m] \leq \frac{1}{2}\beta_i H^{m-2} m!$. Then, for $0 < t \leq \frac{1}{2} B_n^{\frac{1}{2}} H^{-1}$,

(1) $P\{S_n \geq 2tB_n^{\frac{1}{2}}\} < e^{-t^2}$

(2) $P\{S_n \leq -2tB_n^{\frac{1}{2}}\} < e^{-t^2}$.

From (1.2.13) and (1.2.15),

$$(1.2.20)\qquad W_{in}(x) - W_i(x) = \frac{1}{n}\sum_{j=1}^{n}\left[\frac{Y_{ij}I_{\{x\}}(X_{ij})}{N_i} - W_i(x)\right] = \frac{1}{n}\sum_{j=1}^{n} A_{ij}$$

where $A_{ij} = Y_{ij}I_{\{x\}}(X_{ij})/N_i - W_i(x)$. It is easy to see that A_{ij}, $j = 1,\ldots,n$, are i.i.d. with mean 0 and finite variance, say $\beta_i = \beta_i(x)$ since $|A_{ij}| \le 1$. Therefore, taking $H = 1$, we then have, for $m \ge 2$,

$$\begin{aligned} E[A_{ij}^m] &\le E[|A_{ij}|^m] \\ &\le E[|A_{ij}|^2] \\ &= \beta_i \\ &\le \frac{1}{2}\beta_i m! \ . \end{aligned}$$

Thus, by Bernstein's inequality, for any $\delta > 0$,

$$\begin{aligned} &P\{|W_{in}(x) - W_i(x)| > \delta\} \\ &= P\left\{\left|\sum_{j=1}^{n} A_{ij}\right| > n\delta\right\} \\ &= P\left\{\left|\sum_{j=1}^{n} A_{ij}\right| > n^{\frac{1}{2}}\delta\beta_i^{-\frac{1}{2}}B_n^{\frac{1}{2}}\right\} \quad \text{where } B_n = n\beta_i \\ &\le P\left\{\left|\sum_{j=1}^{n} A_{ij}\right| > 2B_n^{\frac{1}{2}}\min\left(\frac{1}{2}n^{\frac{1}{2}}\delta\beta_i^{-\frac{1}{2}}, \frac{1}{2}n^{\frac{1}{2}}\beta_i^{\frac{1}{2}}\right)\right\} \end{aligned} \qquad (1.2.21)$$

$$\leq 2 \exp\{-[\min(\tfrac{1}{2} n^{\frac{1}{2}} \delta \beta_i^{-\frac{1}{2}}, \tfrac{1}{2} n^{\frac{1}{2}} \beta_i^{\frac{1}{2}})]^2\}$$

$$= 2 \exp\{-\frac{n}{4} \min(\delta^2 \beta_i^{-1}, \beta_i)\}.$$

Similarly, from (1.2.12), $f_{in}(x) - f_i(x) = \frac{1}{n} \sum_{j=1}^{n} [I_{\{x\}}(X_{ij}) - f_i(x)]$

$= \frac{1}{n} \sum_{j=1}^{n} B_{ij}$ where $B_{ij} = I_{\{x\}}(X_{ij}) - f_i(x)$. Also, B_{ij}, $j = 1, 2, \ldots, n$, are i.i.d. with mean 0 and $|B_{ij}| \leq 1$, and hence with finite variance, say $\alpha_i \equiv \alpha_i(x)$. Taking $H = 1$ and applying Bernstein's inequality again, we obtain

$$(1.2.22) \quad P\{|f_{in}(x) - f_i(x)| > \delta\} \leq 2 \exp\{-\frac{n}{4} \min(\delta^2 \alpha_i^{-1}, \alpha_i)\}.$$

Thus, if we take $\delta = \frac{\varepsilon}{4}$, from (1.2.18) and (1.2.19), for $\underset{\sim}{x} \varepsilon S_i$,

$$P\{\Delta_{in}(\underset{\sim}{x}) > 0\} \leq O(\exp\{-\frac{n}{4} \min(\delta^2 \alpha_i^{-1}(x_i), \alpha_i(x_i))\})$$

$$(1.2.23) \qquad + O(\exp\{-\frac{n}{4} \min(\delta^2 \beta_i^{-1}(x_i), \beta_i(x_i))\}).$$

Similarly, for each $\underset{\sim}{x} \varepsilon T_i$, $i = 1, 2, \ldots, k$, following similar arguments, we also get the conclusion given below:

$$P\{\Delta_{in}(\underset{\sim}{x}) \leq 0\} \leq O(\exp\{-\frac{n}{4} \min(\delta^2 \alpha_i^{-1}(x_i), \alpha_i(x_i))\})$$

$$(1.2.24) \qquad + O(\exp\{-\frac{n}{4} \min(\delta^2 \beta_i^{-1}(x_i), \beta_i(x_i))\}).$$

where $\alpha_i(x)$, $\beta_i(x)$ are defined as above.

Now, let $c_1 = \frac{1}{4}\min(b_1, b_2)$ where $b_1 = \min_{m \le i \le k}\left[\min_{0 \le x \le N_i}(\delta^2\alpha_i^{-1}(x), \alpha_i(x))\right]$, $b_2 = \min_{m \le i \le k}\left[\min_{0 \le x \le N_i}(\delta^2\beta_i^{-1}(x), \beta_i(x))\right]$, here $m = 1$ if p_0 is known and $m = 0$ if p_0 is unknown. It is clear that $c_1 > 0$ since $\beta_i(x) > 0$, $\alpha_i(x) > 0$ and χ is finite. Thus, we have the following theorem:

Theorem 1.2.4. Let $\{d_n\}_{n=1}^{\infty}$ be the sequence of empirical Bayes rules described in Theorem 1.2.2. Then, $r_n(G, d_n) - r(G) \le O(\exp\{-c_1 n\})$ for some $c_1 > 0$. That is, the sequence of empirical Bayes rules $\{d_n\}_{n=1}^{\infty}$ is asymptotically optimal at least of order $\exp\{-c_1 n\}$ relative to the prior distribution G.

1.2.3. A Case When $G_i(\cdot)$ are Symmetric about $p = \frac{1}{2}$

In this subsection, we suppose that there is sufficient information to tell us that $G_i(\cdot)$ are symmetric about $p = \frac{1}{2}$ for all $i = (0), 1, \ldots, k$. Further, we also assume that N_i are even integers for all $i = (0), 1, \ldots, k$.

Estimation of $W_i(x)$ and $f_i(x)$

Under the above assumptions, $f_i(x) = f_i(N_i - x)$ for all $x = 0, 1, \ldots, N_i$. Therefore, it is reasonable to use

$$(1.2.25)\qquad f_{in}^1(x) \equiv f_{in}^1(N_i - x) = \begin{cases} \frac{1}{2n}\sum_{j=1}^{n} I_{\{x, N_i - x\}}(X_{ij}) & \text{for } x \neq \frac{N_i}{2}, \\ \frac{1}{n}\sum_{j=1}^{n} I_{\{x\}}(X_{ij}) & \text{for } x = \frac{N_i}{2} \end{cases}$$

to estimate $f_i(x)$.

For $W_i(x)$, $x = 0, 1, \ldots, N_i$, we will construct a sequence of consistent estimators $\{W_{in}^1(x)\}$, in terms of $f_{in}^1(y)$, $y \in \{0, 1, \ldots, N_i\}$, by using the observations $(X_{ij}, j = 1, \ldots, n)$ only. The following lemma is very helpful for the above purpose.

Lemma 1.2.5. Suppose that the prior distribution $G_i(\cdot)$ is symmetric about $p = \frac{1}{2}$. Then

(a) $W_i(x) = \frac{x+1}{N_i - x} W_i(N_i - x - 1)$ for each $x = 0, 1, \ldots, N_i - 1$.

(b) $W_i(x) + W_i(N_i - x) = f_i(x) = f_i(N_i - x)$ for each $x = 0,1,\ldots, N_i$.

(c) Furthermore, if N_i is an even integer, then, $W_i\left(\frac{N_i}{2}\right) = \frac{1}{2} f_i\left(\frac{N_i}{2}\right)$.

Proof: (a) Under the assumption that $G_i(\cdot)$ is symmetric about $p = \frac{1}{2}$, then, for each $x = 0, 1, \ldots, N_i - 1$, we have

$$
\begin{aligned}
W_i(x) &\equiv \int_0^1 p\binom{N_i}{x} p^x (1-p)^{N_i - x}\, dG_i(p) \\
&= \int_0^1 \binom{N_i}{x} p^{x+1} (1-p)^{N_i - x}\, dG_i(p) \\
&= \int_0^1 \binom{N_i}{x} p^{N_i - x} (1-p)^{x+1}\, dG_i(p) \qquad (1.2.26) \\
&= \frac{\binom{N_i}{x}}{\binom{N_i}{N_i - x - 1}} \int_0^1 p\binom{N_i}{N_i - x - 1} p^{N_i - x - 1} (1-p)^{x+1} dG_i(p)
\end{aligned}
$$

$$= \frac{\binom{N_i}{x}}{\binom{N_i}{N_i-x-1}} W_i(N_i - x - 1)$$

$$= \frac{x+1}{N_i - x} W_i(N_i - x - 1).$$

(b) Similarly, for each $x = 0, 1, \ldots, N_i$,

$$W_i(x) + W_i(N_i - x)$$

$$= \int_0^1 \binom{N_i}{x} p^{x+1}(1-p)^{N_i-x} dG_i(p) + \int_0^1 \binom{N_i}{x} p^{N_i-x+1}(1-p)^x dG_i(p)$$

(1.2.27)

$$= \int_0^1 \binom{N_i}{x} p^{x+1}(1-p)^{N_i-x} dG_i(p) + \int_0^1 \binom{N_i}{x} p^x(1-p)^{N_i-x+1} dG_i(p)$$

$$= \int_0^1 \binom{N_i}{x} p^x(1-p)^{N_i-x} dG_i(p)$$

$$= f_i(x).$$

(c) In (1.2.27), by taking $x = \frac{N_i}{2}$, then the result $W_i\left(\frac{N_i}{2}\right) = \frac{1}{2} f_i\left(\frac{N_i}{2}\right)$ is obtained.

Theorem 1.2.6. Suppose that $G_i(\cdot)$ is symmetric about $p = \frac{1}{2}$ and N_i is an even integer. Then, for each $x = 0, 1, \ldots, N_i$, $W_i(x)$ can be represented as a linear function of $f_i(y)$, $y = 0, 1, \ldots, N_i$.

Proof: First, from Lemma 1.2.5(a) and(b), for each $x = 0, 1, \ldots, N_i - 1$,

$$W_i(N_i - x) = f_i(N_i - x) - W_i(x)$$

(1.2.28)

$$= f_i(N_i - x) - \frac{x + 1}{N_i - x} W_i(N_i - x - 1).$$

By taking $x = \frac{N_i}{2} - 1 + z$ and after some simple computation, we have

$$W_i\left(\frac{N_i}{2} - z\right) = \frac{N_i + 2 - 2z}{N_i + 2z} f_i\left(\frac{N_i}{2} - z + 1\right)$$

(1.2.29)

$$- \frac{N_i + 2 - 2z}{N_i + 2z} W_i\left(\frac{N_i}{2} - z + 1\right).$$

Next, from Lemma 1.2.5 (c), we have

$$W_i\left(\frac{N_i}{2}\right) = \frac{1}{2} f_i\left(\frac{N_i}{2}\right). \quad (1.2.30)$$

Then, by (1.2.29), (1.2.30) and induction, we conclude that for each $z = 1, 2, \ldots, \frac{N_i}{2}$, $W_i\left(\frac{N_i}{2} - z\right)$ can be represented as a linear function of $f_i(y)$, $y \in \{0, 1, \ldots, N_i\}$.

Finally, by Lemma 1.2.5 (b), we also see that for each $x = \frac{N_i}{2} + 1, \ldots, N_i$, $W_i(x)$ can be represented as a linear function of $f_i(y)$, $y \in \{0, 1, \ldots, N_i\}$. Hence, the proof of this theorem is completed.

By Theorem 1.2.6, for each $x = 0, 1, \ldots, N_i$,

$$W_i(x) = \sum_{y=0}^{N_i} \beta(N_i, x, y) f_i(y), \quad (1.2.31)$$

where the coefficients $\beta(N_i, x, y)$ depend on N_i, x and y. Also, the values of $\beta(N_i, x, y)$ can be obtained from (1.2.30) and the iterative relation (1.2.29).

For example, for $N_i = 4$, we have

$$\begin{cases} W_i(0) = \frac{1}{4} f_i(3) - \frac{1}{12} f_i(2) \\ W_i(1) = \frac{1}{3} f_i(2) \\ W_i(2) = \frac{1}{2} f_i(2) \\ W_i(3) = f_i(3) - \frac{1}{3} f_i(2) \\ W_i(4) = f_i(4) - \frac{1}{4} f_i(3) + \frac{1}{12} f_i(2) \end{cases}$$

We then define

$$(1.2.32) \qquad W_{in}^1(x) = \sum_{y=0}^{N_i} \beta(N_i, x, y)\, f_{in}^1(y)$$

where $f_{in}^1(y)$ have been defined in (1.2.25).

Now, define

$$(1.2.33) \qquad \Delta_{in}^1(\underset{\sim}{x}) = \begin{cases} W_{0n}^1(x_0) f_{in}^1(x_i) - W_{in}^1(x_i) f_{0n}^1(x_0) & \text{if } p_0 \text{ is unknown,} \\ p_0 f_{in}^1(x_i) - W_{in}^1(x_i) & \text{if } p_0 \text{ is known,} \end{cases}$$

and

$$(1.2.34) \quad d_n^1(\underset{\sim}{x}) = \{i | \Delta_{in}^1(\underset{\sim}{x}) \leq 0\}.$$

From (1.2.25), it is clear that $f_{in}^1(x) \to f_i(x)$ with probability 1 as $n \to \infty$ for each $x \in \{0, 1, \ldots, N_i\}$. Therefore, from (1.2.31) and (1.2.32), $W_{in}^1(x) \to W_i(x)$ with probability 1 as $n \to \infty$ for each $x \in \{0, 1, \ldots, N_i\}$. Thus, we have the following theorem:

Theorem 1.2.7. Suppose that the prior distributions $G_i(\cdot)$ are symmetric about $p = \frac{1}{2}$ and N_i are even integers for all $i = (0), 1, \ldots, k$. Then, the sequence of decision rules $\{d_n^1\}_{n=1}^{\infty}$ is asymptotically optimal relative to the prior distribution G.

Rate of Convergence of Empirical Bayes Rules $\{d_n^1\}$

We now consider the rate of convergence of the empirical Bayes rules $\{d_n^1\}$. Following the same discussion as given in (1.2.16) through (1.2.19), and the fact that the estimators $\{f_{in}^1(x)\}$ defined in (1.2.25) share the same property as that defined in (1.2.12), it suffices to consider the behavior of $P\{W_{in}^1(x) - W_i(x) > \delta\}$ and $P\{W_{in}^1(x) - W_i(x) < -\delta\}$ as $n \to \infty$ for some $\delta > 0$, for each $x \in \{0, 1, \ldots, N_i\}$, $i = (0), 1, \ldots, k$. From (1.2.31) and (1.2.32), for each $x \in \{0, 1, \ldots, N_i\}$,

$$P\{W_{in}^1(x) - W_i(x) > \delta\}$$

$$(1.2.35) \quad = P\left\{ \sum_{y=0}^{N_i} \beta(N_i, x, y)\left[f_{in}^1(y) - f_i(y)\right] > \delta \right\}$$

$$\leq \sum_{y=0}^{N_i} P\left\{\beta(N_i, x, y)\left[f_{in}^1(y) - f_i(y)\right] > \delta_1\right\}$$

where $\delta_1 = \frac{\delta}{N_i + 1}$. If $\beta(N_i, x, y) = 0$ for some $0 \leq y \leq N_i$, then $P\{\beta(N_i, x, y)[f_{in}^1(y) - f_i(y)] > \delta_1\} = 0$. So, we assume $\beta(N_i, x, y) \neq 0$. When $\beta(N_i, x, y) > 0$, then

$$P\{\beta(N_i, x, y)[f_{in}^1(y) - f_i(y)] > \delta_1\}$$

(1.2.36)

$$= P\left\{f_{in}^1(y) - f_i(y) > \frac{\delta_1}{\beta(N_i, x, y)}\right\}$$

where $\delta_1/\beta(N_i, x, y) > 0$. When $\beta(N_i, x, y) < 0$,

$$P\{\beta(N_i, x, y)[f_{in}^1(y) - f_i(y)] > \delta_1\}$$

(1.2.37)

$$= P\{f_{in}^1(y) - f_i(y) < \delta_1/\beta(N_i, x, y)\}$$

where $\delta_1/\beta(N_i, x, y) < 0$. In either case, the problem can be reduced to considering the convergence rate of $P\{|f_{in}^1(y) - f_i(y)| > \delta_2\}$ as $n \to \infty$ for some $\delta_2 > 0$. Similarly, for the convergence rate of $P\{W_{in}^1(x) - W_i(x) < -\delta\}$ where $x \in \{0, 1, \ldots, N_i\}$ and $\delta > 0$, we also get a similar result. Therefore, by applying Bernstein's inequality and following an argument similar to that of (1.2.20) and (1.2.21), we conclude the following theorem:

Theorem 1.2.8. Let $\{d_n^1\}_{n=1}^{\infty}$ be the sequence of empirical Bayes rules defined in (1.2.34). Then, $\{d_n^1\}_{n=1}^{\infty}$ is asymptotically optimal at least of order $\exp\{-c_2 n\}$ relative to the prior distribution G for some $c_2 > 0$.

1.2.4. Smooth Empirical Estimation of $f_i(x)$ and $W_i(x)$

In this subsection, we also assume that $G_i(\cdot)$ are symmetric about $p = \frac{1}{2}$ and N_i are even integers for all $i = (0), 1, \ldots, k$. In Subsection 1.2.3, the marginal frequency functions $f_i(x)$, $x \in \{0, 1, \ldots, N_i\}$, $i = (0), 1, \ldots, k$, are estimated in terms of the empirical frequency functions $f_{in}^1(x)$, regardless of properties associated with the marginal function $f_i(x)$. In this subsection, by considering some properties related to $f_i(x)$ and $W_i(x)$, two methods for smoothing the estimators $f_{in}^1(x)$ and $W_{in}^1(x)$ are studied.

We first need the following lemma.

Lemma 1.2.9. Suppose that $G_i(\cdot)$ is symmetric about $p = \frac{1}{2}$ and N_i is an even integer. Then,

$$(1.2.38) \quad (y + 1)\, f_i(y + 1) \leq (N_i - y)\, f_i(y) \quad \text{and}$$

$$(1.2.39) \quad W_i(y) \leq W_i(N_i - y)$$

for all $y \in \left\{0, 1, \ldots, \frac{N_i}{2} - 1\right\}$.

Proof: We prove the inequality of (1.2.38) first. It suffices to show that

$$(1.2.40) \quad \int_0^1 p^x (1 - p)^{N_i - x}\, dG_i(p) \leq \int_0^1 p^y\, (1 - p)^{N_i - y}\, dG_i(p)$$

for all integers x, y such that $0 \leq y < x \leq \frac{N_i}{2}$.

By the symmetry of $G_i(p)$,

$$\int_0^1 p^x (1-p)^{N_i-x}\, dG_i(p)$$

$$= \int_0^{\frac{1}{2}} p^x (1-p)^{N_i-x}\, dG_i(p) + \int_{\frac{1}{2}}^1 p^x (1-p)^{N_i-x}\, dG_i(p)$$

(1.2.41)

$$= \int_0^{\frac{1}{2}} p^x (1-p)^{N_i-x}\, dG_i(p) + \int_0^{\frac{1}{2}} (1-p)^x p^{N_i-x}\, dG_i(p)$$

$$= \int_0^{\frac{1}{2}} \left[p^x (1-p)^{N_i-x} + (1-p)^x p^{N_i-x} \right] dG_i(p).$$

Note that for $0 \le y < x \le \frac{N_i}{2}$ and $p \in [0, \frac{1}{2}]$, the inequality

$$p^{N_i-y} (1-p)^y \le p^{N_i-x} (1-p)^x \le p^x (1-p)^{N_i-x}$$

(1.2.42)

$$\le p^y (1-p)^{N_i-y}$$

is always true. Also,

$$p^x (1-p)^{N_i-x}\, p^{N_i-x} (1-p)^x = p^y (1-p)^{N_i-y}\, p^{N_i-y} (1-p)^y. \tag{1.2.43}$$

Then, by (1.2.42) and (1.2.43), we obtain

$$p^{N_i-x} (1-p)^x + p^x (1-p)^{N_i-x} \le p^{N_i-y} (1-p)^y + p^y (1-p)^{N_i-y} \tag{1.2.44}$$

for all $p \in [0, \frac{1}{2}]$ as $0 \le y < x \le \frac{N_i}{2}$.

Therefore, from (1.2.41) and (1.2.44), the inequality (1.2.40) is obtained.

For the inequality (1.2.39), we consider

$$W_i(N_i - y) - W_i(y)$$

$$= \int_0^1 \binom{N_i}{y} p^{N_i-y+1}(1-p)^y dG_i(p) - \int_0^1 \binom{N_i}{y} p^{y+1}(1-p)^{N_i-y} dG_i(p)$$

$$= \binom{N_i}{y} \int_0^1 p^{y+1}(1-p)^y \left[p^{N_i-2y} - (1-p)^{N_i-2y}\right] dG_i(p)$$

$$= \binom{N_i}{y} \int_0^{\frac{1}{2}} p^{y+1}(1-p)^y \left[p^{N_i-2y} - (1-p)^{N_i-2y}\right] dG_i(p)$$

$$(1.2.45) \quad + \binom{N_i}{y} \int_{\frac{1}{2}}^1 p^{y+1}(1-p)^y \left[p^{N_i-2y} - (1-p)^{N_i-2y}\right] dG_i(p)$$

$$= \binom{N_i}{y} \int_{\frac{1}{2}}^1 (1-p)^{y+1} p^y \left[(1-p)^{N_i-2y} - p^{N_i-2y}\right] dG_i(p)$$

$$+ \binom{N_i}{y} \int_{\frac{1}{2}}^1 p^{y+1}(1-p)^y \left[p^{N_i-2y} - (1-p)^{N_i-2y}\right] dG_i(p)$$

$$= \binom{N_i}{y} \int_{\frac{1}{2}}^1 p^y (1-p)^y [2p-1] \left[p^{N_i-2y} - (1-p)^{N_i-2y}\right] dG_i(p)$$

$$\geq 0$$

Since for $0 \leq y \leq \frac{N_i}{2} - 1$, $p \in [\frac{1}{2}, 1]$, $2p - 1 \geq 0$ and $p^{N_i-2y} \geq (1-p)^{N_i-2y}$.

Therefore, $W_i(y) \leq W_i(N_i - y)$ for $y \in \{0, 1, \ldots, \frac{N_i}{2} - 1\}$.

Lemma 1.2.10. Let $U(x)$, $h(x)$ be nonnegative functions defined on $\{0, 1, \ldots, N\}$ and satisfy

(i) $U(x) = \frac{x+1}{N-x} U(N-x-1)$ for all $x = 0, 1, \ldots, N-1$.

(ii) $U(x) + U(N-x) = h(x) = h(N-x)$ for all $x = 0, 1, \ldots, N$

and

(iii) $U(x) \leq U(N-x)$ for all $x = 0, 1, \ldots, \frac{N}{2} - 1$,

where N is an even positive integer. Then,

(iv) $(x+1)\, h(x+1) \leq (N-x)\, h(x)$ for all $x = 0, 1, \ldots, \frac{N}{2} - 1$.

Proof: Note that from (i), $(N-x)\, U(x) = (x+1)\, U(N-x-1)$. Then, by (ii), we obtain

$$(1.2.46) \quad (N-x)[h(x) - U(N-x)] = (x+1)[h(x+1) - U(x+1)].$$

Hence, from (1.2.46), we have

$$\begin{aligned} &(N-x)\, h(x) - (x+1)\, h(x+1) \\ &\quad = (N-x)\, U(N-x) - (x+1)\, U(x+1) \\ &\quad \geq (N-x)\, U(x) - (x+1)\, U(x+1) && \text{(by (iii))} \\ &\quad \geq (N-x)\, U(x) - (x+1)\, U(N-x-1) && \text{(by (iii))} \end{aligned} \tag{1.2.47}$$

$$= (N - x)\left[U(x) - \frac{x + 1}{N - x} U(N - x - 1)\right]$$

$$= 0 \quad \text{(by (i))}.$$

Hence, the proof of this lemma is completed.

We note that conditions (i), (ii) and (iv) of Lemma 1.2.10 do not imply that $U(x) \leq U(N - x)$ for all $x = 0, 1, \ldots, \frac{N}{2} - 1$. The following example illustrates this fact.

Example 1.2.11. Take $N = 4$. Let

$$\begin{cases} h(0) = \frac{1}{10} \\ h(1) = \frac{3}{10} \\ h(2) = \frac{2}{10} \\ h(3) = \frac{3}{10} \\ h(4) = \frac{1}{10} \end{cases}, \qquad \begin{cases} U(0) = \frac{7}{120} \\ U(1) = \frac{8}{120} \\ U(2) = \frac{12}{120} \\ U(3) = \frac{28}{120} \\ U(4) = \frac{5}{120} \end{cases}$$

Then, conditions (i), (ii) and (iv) are satisfied but $U(4) < U(0)$.

From Lemma 1.2.9, the inequalities (1.2.38) and (1.2.39) are always true for all $y = 0, 1, \ldots, \frac{N_i}{2} - 1$. However, the empirical frequency functions $f^1_{in}(x)$ and the functions $W^1_{in}(x)$ do not always satisfy the above inequalities. Hence, it is reasonable to consider some smoothing of $f^1_{in}(x)$ and $W^1_{in}(x)$, which will satisfy the above inequalities. Two smoothing methods, based on $f^1_{in}(x)$ and $W^1_{in}(x)$ respectively, are given as follows.

Method 1. Smoothing Based on $f^1_{in}(x)$

Given $\Delta_1 \geq 0$ (for $\Delta_1 > 0$, Δ_1 is chosen small), for each $y = 0, 1, \ldots, \frac{N_i}{2} - 1$, let $\varepsilon_{in}(\Delta_1, y) = \min(\Delta_1, f^1_{in}(y) + f^1_{in}(y+1))$. Let m_1 stand for the number of times the smoothing process is carried out. Algorithmically, first $m_1 = 0$.

Step 1. $m_1 = m_1 + 1$.

Check whether $(N_i - y)f^1_{in}(y) - (y+1)f^1_{in}(y+1) - \varepsilon_{in}(\Delta_1, y) \leq 0$ or not. If not, for $y \leq \frac{N_i}{2} - 2$, let

$$a_{in}(\Delta_1, y) = [(y+1)f^1_{in}(y+1) - (N_i - y)f^1_{in}(y) + \varepsilon_{in}(\Delta_1, y)]/(N_i + 1),$$

$$f^0_{in}(y) = f^0_{in}(N_i - y) = f^1_{in}(y) + a_{in}(\Delta_1, y),$$

$$f^0_{in}(y+1) = f^0_{in}(N_i - y - 1) = f^1_{in}(y) - a_{in}(\Delta_1, y) \quad \text{and}$$

$$f^0_{in}(x) = f^1_{in}(x) \quad \text{for all } x \neq y, y+1, N_i - y - 1, N_i - y.$$

For $y = \frac{N_i}{2} - 1$, let

$$a_{in}(\Delta_1, y) = [(y+1)f^1_{in}(y+1) - (N_i - y)f^1_{in}(y) + \varepsilon_{in}(\Delta_1, y)] \times \frac{4}{3N_i + 2},$$

$$f^0_{in}(y) = f^0_{in}(N_i - y) = f^1_{in}(y) + \frac{1}{2} a_{in}(\Delta_1, y),$$

$$f^0_{in}(y+1) = f^1_{in}(y+1) - a_{in}(\Delta_1, y) \quad \text{and}$$

$$f^0_{in}(x) = f^1_{in}(x) \quad \text{for all } x \neq \frac{N_i}{2} - 1, \frac{N_i}{2} \text{ and } \frac{N_i}{2} + 1.$$

Step 2. Check whether $(N_i - y)f_{in}^0(y) - (y + 1)f_{in}^0(y + 1) - \varepsilon_{in}(\Delta_1, y) \geq 0$ for all $y = 0, 1, \ldots, \frac{N_i}{2} - 1$ or not.

If yes, go to step 3.

If no, let $f_{in}^1(x) = f_{in}^0(x)$ for all $x = 0, 1, \ldots, N_i$, and go to step 1.

Step 3. Define $W_{in}^0(x) = \sum_{y=0}^{N_i} \beta(N_i, x, y)f_{in}^0(y)$, $x = 0, 1, \ldots, N_i$.

Remark 1.2.12 (1). We note that when the above smoothing procedure stops, then the smooth estimators $f_{in}^0(y)$ have the property that $(y + 1)f_{in}^0(y + 1) \leq (N_i - y)f_{in}^0(y)$ for all $y = 0, 1, \ldots, \frac{N_i}{2} - 1$.

(2). However, it is possible that the above smoothing procedure never stops. In this situation, we can set up a maximal smoothing time to stop this procedure. When this happens, the inequality that $(y + 1)f_{in}^0(y + 1) \leq (N_i - y)f_{in}^0(y)$ for all $y = 0, 1, \ldots, \frac{N_i}{2} - 1$, is not guaranteed.

Based on the smooth estimators $f_{in}^0(x)$ and $W_{in}^0(x)$, we define decision rules $d_n^0(\cdot)$, $n = 1, 2, \ldots$, as follows:

$$(1.2.48) \quad d_n^0(\underset{\sim}{x}) = \{i | W_{0n}^0(x_0)f_{in}^0(x_i) - W_{in}^0(x_i)f_{0n}^0(x_0) \leq 0\}.$$

Method 2. Smoothing Based on $W_{in}^1(x)$

Given $\Delta_2 \geq 0$ (for $\Delta_2 > 0$, Δ_2 is chosen small), for each $y = 0, 1, \ldots, \frac{N_i}{2} - 1$, let $\delta_{in}(\Delta_2, y) = \min(\Delta_2, [W_{in}^1(y)+W_{in}^1(N_i-y)]/2)$

and $b_{in}(\Delta_2, y) = [W^1_{in}(y) - W^1_{in}(N_i - y)]/2 + \delta_{in}(\Delta_2, y)$. We start with a variable m_2 which stands for the number of times the smoothing carried. At first $m_2 = 0$.

Step 1. $m_2 = m_2 + 1$.

Check whether $W^1_{in}(N_i - y) \geq W^1_{in}(y) + \delta_{in}(\Delta_2, y)$ or not. If not, for $y = 0$, let

$$W^*_{in}(N_i) = W^1_{in}(N_i) + c(0)b_{in}(\Delta_2, 0),$$

$$W^*_{in}(0) = W^1_{in}(0) - d(0)b_{in}(\Delta_2, 0),$$

$$W^*_{in}(N_i - 1) = N_i W^*_{in}(0) \quad \text{and}$$

$$W^*_{in}(x) = W^1_{in}(x) \quad \text{for all } x \neq 0, N_i - 1, N_i.$$

For $1 \leq y \leq \frac{N_i}{2} - 1$, let

$$W^*_{in}(N_i - y) = W^1_{in}(N_i - y) + c(y)b_{in}(\Delta_2, y),$$

$$W^*_{in}(y) = W^1_{in}(y) - d(y)b_{in}(\Delta_2, y),$$

$$W^*_{in}(N_i - y - 1) = \frac{N_i - y}{y + 1} W^*_{in}(y)$$

$$W^*_{in}(y - 1) = \frac{y}{N_i - y + 1} W^*_{in}(N_i - y) \quad \text{and}$$

$$W^*_{in}(x) = W^1_{in}(x) \quad \text{for all } x \neq y - 1, y, N_i - y - 1, N_i - y.$$

Here, $c(y) = 2(N_i - y + 1)/(N_i + 2)$, $d(y) = 2(y + 1)/(N_i + 2)$ for $y = 0, 1, \ldots, \frac{N_i}{2} - 1$.

Step 2. Check whether $W^*_{in}(y) + \delta_{in}(\Delta_2, y) \leq W^*_{in}(N_i - y)$ for all $y = 0, 1, \ldots, \frac{N_i}{2} - 1$ or not.

If yes, go to step 3.

If no, let $W^1_{in}(x) = W^*_{in}(x)$ for all $x = 0, 1, \ldots, N_i$ and go to step 1.

Step 3. Let $f^*_{in}(y) = W^*_{in}(y) + W^*_{in}(N_i - y)$ for all $y = 0, 1, \ldots, N_i$.

Remark 1.2.13. (1) We note that when the above smoothing procedure stops, then the smooth estimators $W^*_{in}(y)$ satisfy that $W^*_{in}(y) \leq W^*_{in}(N_i - y)$ for all $y = 0, 1, \ldots, \frac{N_i}{2} - 1$, $W^*_{in}(y) = \frac{y + 1}{N_i - y} W^*_{in}(N_i - y - 1)$ for all $y = 0, 1, \ldots, N_i - 1$ and $W^*_{in}(y) + W^*_{in}(N_i - y) = f^*_{in}(y) = f^*_{in}(N_i - y)$ for all $y = 0, 1, \ldots, N_i$. Then by Lemma 1.2.10, $(y + 1)f^*_{in}(y + 1) \leq (N_i - y)f^*_{in}(y)$ for all $y = 0, 1, \ldots, \frac{N_i}{2} - 1$. Therefore, method 2 is better than method 1 in this sense.

(2) It is also possible that the above smoothing procedure never stops. Hence, we can set up a maximal smoothing time to stop this procedure. When this happens, for the smooth estimators, the inequality properties of (1.2.38) and (1.2.39) are not guaranteed.

Based on the smooth estimators $f^*_{in}(x)$ and $W^*_{in}(x)$, we define decision rules $d^*_n(\cdot)$, $n = 1, 2, \ldots$, as follows:

$$(1.2.49)\quad d_n^*(\underset{\sim}{x}) = \{i \mid W_{0n}^*(x_0) f_{in}^*(x_i) - W_{in}^*(x_i) f_{0n}^*(x_0) \leq 0\}.$$

1.2.5. Monte Carlo Studies

For the sequence of decision rules $\{\delta_n(\underset{\sim}{x})\}_{n=1}^{\infty}$, the conditional Bayes risk at stage n + 1 given $(\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_n)$ is

$$(1.2.50)\quad R_n(G, \delta_n) = \int_{\Omega} \sum_{\underset{\sim}{x}\in\mathcal{X}} L(\underset{\sim}{p}, \delta_n(\underset{\sim}{x}))\, f(\underset{\sim}{x}|\underset{\sim}{p})\, dG(\underset{\sim}{p}).$$

To measure the performance of the sequence of decision rules $\{\delta_n(\underset{\sim}{x})\}_{n=1}^{\infty}$, computing the overall risk $r_n(G, \delta_n) = ER_n(G, \delta_n)$ is needed, where the expectation E is taken with respect to $(\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_n)$. For the small sample situation, it is impossible to analytically determine such values. Therefore, Monte Carlo simulation is employed.

In this subsection, we have carried out some Monte Carlo studies to see the performance of the sequences of decision rules $\{d_n^1\}$, $\{d_n^0\}$ and $\{d_n^*\}$. We let, conditional on p_i, $X_i \sim B(N_i, p_i)$ where N_i are even integers for i = 0, 1 and p_0 is treated as unknown. We also assume that

$$(1.2.51)\quad G_i(p) = \int_0^p \frac{\Gamma(2\alpha_i + 2)}{[\Gamma(\alpha_i + 1)]^2} y^{\alpha_i} (1 - y)^{\alpha_i}\, dy, \quad i = 0, 1.$$

Then,

$$(1.2.52)\quad f_i(x) = \binom{N_i}{x} \frac{\Gamma(2\alpha_i + 2)}{[\Gamma(\alpha_i + 1)]^2} \times \frac{\Gamma(x + \alpha_i + 1)\, \Gamma(N_i + \alpha_i - x + 1)}{\Gamma(N_i + 2\alpha_i + 2)},$$

$$(1.2.53)\quad W_i(x) = \binom{N_i}{x} \frac{\Gamma(2\alpha_i + 2)}{[\Gamma(\alpha_i + 1)]^2} \times \frac{\Gamma(x + \alpha_i + 2)\,\Gamma(N_i + \alpha_i - x + 1)}{\Gamma(N_i + 2\alpha_i + 3)}$$

for $x = 0, 1, \ldots, N_i$; $i = 0, 1$.

Hence, the Bayes rule d_G is:

$$\text{Select } \pi_1 \text{ as good} \quad \text{iff} \quad W_0(x_0) f_1(x_1) \leq W_1(x_1) f_0(x_0) ,$$

$$\text{iff} \quad \frac{x_0 + \alpha_0 + 1}{N_0 + 2\alpha_0 + 2} \leq \frac{x_1 + \alpha_1 + 1}{N_1 + 2\alpha_1 + 2} .$$

A random sample of size 50 was generated by computing from a population having $f_i(x)$ $(i = 0, 1)$ as probability function. For each $n = 1, 2, \ldots, 50$, the conditional Bayes risks $R_n(G, d_n^1)$, $R_n(G, d_n^0)$ and $R_n(G, d_n^*)$ were calculated. One hundred repetitions were performed. Estimates of the overall risks $r_n(G, d_n^1)$, $r_n(G, d_n^0)$ and $r_n(G, d_n^*)$ were obtained by averaging the associated conditional Bayes risks and the standard deviations of the estimated overall risks were also obtained based on these repeated samples.

In Table I, we consider the combinations of different N_i's and α_i's values for our decision problem. We let $\hat{r}_n(G, d_n)$ denote the average of 100 $R_n(G, d_n)$ values obtained from simulation. The standard deviation associated with $\hat{r}_n(G, d_n)$ is given in the corresponding parentheses. It is easy to see that the performances of the sequences of decision rules $\{d_n^0\}$ and $\{d_n^*\}$ are always better than that of $\{d_n^1\}$, as we expect. For the comparison of $\{d_n^0\}$ and $\{d_n^*\}$, for the cases that $(N_0, N_1, \alpha_0, \alpha_1) = (2, 2, 4, 4)$ and $(N_0, N_1, \alpha_0, \alpha_1) = (2, 2, 6, 6)$,

both of them have the same performance. For the other $(N_0, N_1, \alpha_0, \alpha_1)$'s, the performance of $\{d_n^*\}$ is always better than that of $\{d_n^0\}$, except the case that $(N_0, N_1, \alpha_0, \alpha_1) = (2, 2, 4, 6)$. Even in this case, the difference between $\hat{r}_n(G, d_n^0)$ and $\hat{r}_n(G, d_n^*)$ is very small for each $n = 1, 2, \ldots, 50$. It is also interesting to note that in most cases, $\hat{r}_n(G, d_n^*)$ has the smallest standard deviation while $\hat{r}_n(G, d_n^1)$ has the largest standard deviation. This fact indicates that the behavior of the sequence of decision rules $\{d_n^*\}$ is more stable than the others.

1.3. Empirical Bayes Rules for Selecting Good Populations--Pareto Populations Case

In the previous section, we considered the construction of empirical Bayes rules for selecting good populations in comparison with a control or standard population. However, an experimenter may be concerned about comparisons among k (≥ 2) different populations. He may be interested in finding the superior (or the inferior) populations among the k populations. For example, a sociologist may be concerned about the development of the environment of different local areas. An economist might be interested in the incomes of different groups. One of the most commonly used models for the above situations is a Pareto model. For income example, economists believe that there should be some minimum income for a family to subsist on. Similarly, for the environmental problem, there should also be some minimum acceptable level for a person to live in some area. However, it may be hard to define what kind of level it should be, and these levels may change with time. Social workers, generally, are interested in

TABLE I.1

Simulation results for the comparative performance of sequences of empirical Bayes rules $\{d_n^1\}$, $\{d_n^0\}$ and $\{d_n^*\}$.

$(N_0, N_1, \alpha_0, \alpha_1) = (2, 2, 4, 4)$, $r(G) = 0.05287$

n	$\hat{r}_n(G, d_n^1)$	$\hat{r}_n(G, d_n^0)$	$\hat{r}_n(G, d_n^*)$
1	0.08936	0.06299	0.06299
	(0.00118)	(0.00073)	(0.00073)
2	0.08426	0.05907	0.05907
	(0.00171)	(0.00062)	(0.00062)
3	0.09188	0.05638	0.05638
	(0.00213)	(0.00053)	(0.00053)
5	0.08745	0.05452	0.05452
	(0.00227)	(0.00038)	(0.00038)
10	0.08299	0.05298	0.05298
	(0.00222)	(0.00010)	(0.00010)
15	0.07899	0.05287	0.05287
	(0.00251)	(0.00000)	(0.00000)
20	0.07767	0.05287	0.05287
	(0.00243)	(0.00000)	(0.00000)
25	0.07849	0.05287	0.05287
	(0.00234)	(0.00000)	(0.00000)
30	0.07481	0.05287	0.05287
	(0.00212)	(0.00000)	(0.00000)
35	0.07328	0.05287	0.05287
	(0.00215)	(0.00000)	(0.00000)
40	0.07215	0.05287	0.05287
	(0.00203)	(0.00000)	(0.00000)
45	0.07200	0.05287	0.05287
	(0.00213)	(0.00000)	(0.00000)
50	0.07157	0.05287	0.05287
	(0.00212)	(0.00000)	(0.00000)

TABLE I.2

Simulation results for the comparative performance of sequences of empirical Bayes rules $\{d_n^1\}$, $\{d_n^0\}$ and $\{d_n^*\}$.

$(N_0, N_1, \alpha_0, \alpha_1) = (2, 2, 4, 6)$, $r(G) = 0.04965$

n	$\hat{r}_n(G, d_n^1)$	$\hat{r}_n(G, d_n^0)$	$\hat{r}_n(G, d_n^*)$
1	0.08093	0.06135	0.06135
	(0.00103)	(0.00068)	(0.00068)
2	0.07724	0.05747	0.05747
	(0.00147)	(0.00066)	(0.00066)
3	0.07771	0.05459	0.05447
	(0.00215)	(0.00058)	(0.00059)
5	0.07706	0.05209	0.05212
	(0.00220)	(0.00032)	(0.00032)
10	0.07659	0.05112	0.05097
	(0.00210)	(0.00012)	(0.00014)
15	0.07314	0.05090	0.05129
	(0.00211)	(0.00013)	(0.00014)
20	0.07034	0.05107	0.05138
	(0.00203)	(0.00013)	(0.00014)
25	0.07006	0.05093	0.05082
	(0.00215)	(0.00014)	(0.00014)
30	0.06743	0.05085	0.05115
	(0.00195)	(0.00013)	(0.00014)
35	0.06682	0.05078	0.05121
	(0.00197)	(0.00014)	(0.00015)
40	0.06724	0.05059	0.05118
	(0.00192)	(0.00012)	(0.00014)
45	0.06632	0.05061	0.05111
	(0.00194)	(0.00013)	(0.00014)
50	0.06676	0.05074	0.05127
	(0.00184)	(0.00013)	(0.00014)

TABLE I.3

Simulation results for the comparative performance of sequences of empirical Bayes rules $\{d_n^1\}$, $\{d_n^0\}$ and $\{d_n^*\}$.

$(N_0, N_1, \alpha_0, \alpha_1) = (2, 2, 6, 6)$, $r(G) = 0.04896$

n	$\hat{r}_n(G, d_n^1)$	$\hat{r}_n(G, d_n^0)$	$\hat{r}_n(G, d_n^*)$
1	0.07640	0.05626	0.05626
	(0.00089)	(0.00056)	(0.00056)
2	0.07188	0.05284	0.05284
	(0.00126)	(0.00043)	(0.00043)
3	0.07577	0.05121	0.05121
	(0.00160)	(0.00038)	(0.00038)
5	0.07272	0.04965	0.04965
	(0.00175)	(0.00022)	(0.00022)
10	0.07098	0.04896	0.04896
	(0.00164)	(0.00000)	(0.00000)
15	0.07111	0.04896	0.04896
	(0.00189)	(0.00000)	(0.00000)
20	0.07090	0.04896	(0.04896)
	(0.00173)	(0.00000)	(0.00000)
25	0.06931	0.04896	(0.04896)
	(0.00185)	(0.00000)	(0.00000)
30	0.06904	0.04896	0.04896
	(0.00176)	(0.00000)	(0.00000)
35	0.06938	0.04896	0.04896
	(0.00181)	(0.00000)	(0.00000)
40	0.06855	0.04896	0.04896
	(0.00171)	(0.00000)	(0.00000)
45	0.06860	0.04896	0.04896
	(0.00169)	(0.00000)	(0.00000)
50	0.06681	0.04896	0.04896
	(0.00169)	(0.00000)	(0.00000)

TABLE I.4

Simulation results for the comparative performance of sequences of empirical Bayes rules $\{d_n^1\}$, $\{d_n^0\}$ and $\{d_n^*\}$.

$(N_0, N_1, \alpha_0, \alpha_1) = (4, 4, 4, 4)$, $r(G) = 0.04114$

n	$\hat{r}_n(G, d_n^1)$	$\hat{r}_n(G, d_n^0)$	$\hat{r}_n(G, d_n^*)$
1	0.08641	0.07389	0.06350
	(0.00030)	(0.00123)	(0.00123)
2	0.09043	0.06071	0.05480
	(0.00168)	(0.00164)	(0.00116)
3	0.08626	0.06396	0.05026
	(0.00069)	(0.00131)	(0.00092)
5	0.08595	0.05851	0.04613
	(0.00174)	(0.00156)	(0.00069)
10	0.08479	0.05546	0.04274
	(0.00145)	(0.00147)	(0.00035)
15	0.07992	0.05596	0.04189
	(0.00161)	(0.00147)	(0.00016)
20	0.07994	0.05536	0.04148
	(0.00156)	(0.00145)	(0.00010)
25	0.07514	0.05358	0.04155
	(0.00185)	(0.00133)	(0.00011)
30	0.07674	0.05446	0.04144
	(0.00157)	(0.00141)	(0.00011)
35	0.07458	0.05449	0.04144
	(0.00156)	(0.00139)	(0.00010)
40	0.07024	0.05312	0.04168
	(0.00158)	(0.00130)	(0.00013)
45	0.06855	0.05267	0.04158
	(0.00145)	(0.00126)	(0.00013)
50	0.06749	0.05187	0.04158
	(0.00152)	(0.00119)	(0.00012)

TABLE I.5

Simulation results for the comparative performance of sequences of empirical Bayes rules $\{d_n^1\}$, $\{d_n^0\}$ and $\{d_n^*\}$.

$(N_0, N_1, \alpha_0, \alpha_1) = (4, 4, 4, 6)$, $r(G) = 0.03937$

n	$\hat{r}_n(G, d_n^1)$	$\hat{r}_n(G, d_n^0)$	$\hat{r}_n(G, d_n^*)$
1	0.08071	0.06913	0.06074
	(0.00027)	(0.00102)	(0.00103)
2	0.08244	0.05919	0.05334
	(0.00145)	(0.00139)	(0.00094)
3	0.07920	0.05979	0.04862
	(0.00072)	(0.00126)	(0.00088)
5	0.07858	0.05437	0.04439
	(0.00154)	(0.00140)	(0.00071)
10	0.07701	0.05012	0.04082
	(0.00152)	(0.00123)	(0.00015)
15	0.07278	0.05212	0.04098
	(0.00158)	(0.00118)	(0.00017)
20	0.07153	0.05074	0.04057
	(0.00152)	(0.00120)	(0.00014)
25	0.06943	0.05023	0.04099
	(0.00152)	(0.00116)	(0.00018)
30	0.06777	0.05017	0.04058
	(0.00153)	(0.00121)	(0.00012)
35	0.06606	0.04839	0.04084
	(0.00146)	(0.00107)	(0.00015)
40	0.06493	0.04839	0.04082
	(0.00150)	(0.00111)	(0.00014)
45	0.06345	0.04853	0.04078
	(0.00138)	(0.00115)	(0.00014)
50	0.06300	0.04867	0.04080
	(0.00136)	(0.00111)	(0.00012)

TABLE I.6

Simulation results for the comparative performance of sequences of empirical Bayes rules $\{d_n^1\}$, $\{d_n^0\}$ and $\{d_n^*\}$.

$(N_0, N_1, \alpha_0, \alpha_1) = (4, 4, 6, 6)$, $r(G) = 0.03970$

n	$\hat{r}_n(G, d_n^1)$	$\hat{r}_n(G, d_n^0)$	$\hat{r}_n(G, d_n^*)$
1	0.07415	0.06354	0.05607
	(0.00020)	(0.00088)	(0.00087)
2	0.07609	0.05526	0.04950
	(0.00117)	(0.00119)	(0.00073)
3	0.07333	0.05737	0.04669
	(0.00057)	(0.00104)	(0.00066)
5	0.07158	0.05513	0.04275
	(0.00110)	(0.00108)	(0.00051)
10	0.07102	0.05097	0.04079
	(0.00132)	(0.00103)	(0.00024)
15	0.06987	0.05202	0.04006
	(0.00102)	(0.00106)	(0.00009)
20	0.06907	0.05035	0.04001
	(0.00122)	(0.00105)	(0.00009)
25	0.06632	0.04973	0.04004
	(0.00121)	(0.00100)	(0.00010)
30	0.06748	0.05067	0.04004
	(0.00118)	(0.00099)	(0.00010)
35	0.06669	0.05024	0.03997
	(0.00136)	(0.00101)	(0.00009)
40	0.06712	0.05034	0.04004
	(0.00118)	(0.00098)	(0.00009)
45	0.06626	0.05110	0.03987
	(0.00129)	(0.00101)	(0.00006)
50	0.06373	0.05042	0.03992
	(0.00126)	(0.00102)	(0.00008)

those who have a low level or standard of living. From the past and current information, they may select some subgroups and render help to improve their living conditions. In this section, we consider k Pareto populations and suggest a method to decide what populations are around the low level situation.

1.3.1. Formulation of the Problem

Let $\pi_1, \ldots, \pi_k$ be k (≥ 2) populations, and let X_i be a random observation for a certain characteristic of π_i. Assume that X_i has a Pareto distribution and write $X_i \sim Pa(\theta_i, \alpha)$ where $\alpha > 0$ and $\theta_i > 0$ for all $i = 1, 2, \ldots, k$. That is, conditional on θ_i, X_i has density $f_i(x|\theta_i) = \frac{\alpha\theta_i^{\alpha}}{x^{\alpha+1}} I_{(\theta_i,\infty)}(x)$. Let $\Omega = \{\underset{\sim}{\theta} = (\theta_1, \ldots, \theta_k) | \theta_i > 0$ for $i = 1, \ldots, k\}$ be the parameter space. Let $G = \prod_{i=1}^{k} G_i$ be a prior distribution over Ω. Suppose that the following information is available for all $i = 1, 2, \ldots, k$.

$$(1.3.1)\begin{cases} \text{(i)} & G_i(c) = 0 \text{ for some known positive constant } c. \\ \text{(ii)} & G_i(\cdot) \text{ has a continuous probability density function } g_i. \\ \text{(iii)} & \sup_{\theta \geq c} g_i(\theta) \leq M < \infty. \\ \text{(iv)} & \int_c^{\infty} \theta dG_i(\theta) < \infty. \end{cases}$$

From (i), the parameter space Ω can be reduced to $\Omega = \{\underset{\sim}{\theta} = (\theta_1, \ldots, \theta_k) | \theta_i \geq c\}$. For each $\underset{\sim}{\theta} \in \Omega$, let $A_{\Delta}(\underset{\sim}{\theta}) = \{i | \theta_i < \theta_{[1]} + \Delta\}$,

$B_\Delta(\underset{\sim}{\theta}) = \{j|\theta_j \geq \theta_{[1]} + \Delta\}$ where $\theta_{[1]} = \min_{1\leq j\leq k} \theta_j$ and $\Delta > 0$ is a prespecified constant. Population π_i is said to be good if $i \in A_\Delta(\underset{\sim}{\theta})$ and to be bad if $i \in B_\Delta(\underset{\sim}{\theta})$. We are interested in selecting all good populations.

Let $\mathcal{A} = \{a|a \subseteq \{1, 2, \ldots, k\}, a \neq \phi\}$ be the action space. When action a is taken, it means that π_i is selected as a good population if $i \in a$ and rejected as a bad population if $i \notin a$. For each $a \in \mathcal{A}$ and $\underset{\sim}{\theta} \in \Omega$, define the loss function

$$
\begin{aligned}
L(\underset{\sim}{\theta}, a) = & \sum_{j\in a} (\theta_j - \theta_{[1]} - \Delta) I_{[\theta_{[1]}+\Delta,\infty)}(\theta_j) \\
& + \sum_{j\notin a} (\theta_{[1]} + \Delta - \theta_j) I_{[c,\theta_{[1]}+\Delta)}(\theta_j).
\end{aligned} \tag{1.3.2}
$$

The first summation is the loss due to selecting bad populations and the second summation is the loss due to rejecting good populations.

Let $(X_{i1},\Theta_{i1}), \ldots, (X_{in}, \Theta_{in})$ be pairs of random variables associated with population π_i. $\Theta_{i1}, \ldots, \Theta_{in}$ are independently distributed with common distribution G_i and are unobservable. Conditional on $\Theta_{ij} = \theta_{ij}$, $X_{ij} \sim Pa(\theta_{ij}, \alpha)$. Let $\underset{\sim}{X}_j = (X_{1j}, \ldots, X_{kj})$ denote the previous jth observations taken from $\pi_1, \ldots, \pi_k$, respectively. Let $\underset{\sim}{X} = (X_1, \ldots, X_k)$ be the present observations. Conditional on $\underset{\sim}{\theta} = (\theta_1, \ldots, \theta_k)$, $\underset{\sim}{X}$ has a joint density

$$
f(\underset{\sim}{x}|\underset{\sim}{\theta}) = \prod_{i=1}^{k} f_i(x_i|\theta_i) \quad \text{where } f_i(x_i|\theta_i) = \frac{\alpha\theta_i^\alpha}{x_i^{\alpha+1}} I_{[\theta_i,\infty)}(x_i). \tag{1.3.3}
$$

Let $\chi = \prod_{i=1}^{k} \chi_i$ where $\chi_i = [\theta_i, \infty)$. Let $D = \{d | d = \chi \to A\}$ be the decision space. For each $d \in D$, let $r(G, d)$ denote the associated Bayes risk. Then, $r(G) = \inf_{d \in D} r(G, d)$ is the minimum Bayes risk.

Let $d_n(\underset{\sim}{x}) \equiv d_n(\underset{\sim}{x}, \underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_n)$ be a decision rule, based on the present observation $\underset{\sim}{x}$, and the previous n observations $\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_n$. The sequence of decision rules $\{d_n(\underset{\sim}{x})\}_{n=1}^{\infty}$ is said to be asymptotically optimal relative to the prior distribution G if $r_n(G, d_n) \to r(G)$ as $n \to \infty$, where $r_n(G, d_n) = \int_{\chi} E \int_{\Omega} L(\underset{\sim}{\theta}, d_n(\underset{\sim}{x}, \underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_n)) f(\underset{\sim}{x} | \underset{\sim}{\theta}) dG(\underset{\sim}{\theta}) d\underset{\sim}{x}$ and the expectation is taken with respect to the σ-finite measure generated by $(\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_n)$.

Note that for $\Delta = 0$, our problem is reduced to similar cases studied by Deely (1965), Van Ryzin (1970) and Van Ryzin and Susarla (1977) for other distributions. We see that when $\Delta = 0$, the determination of the Bayes rule is independent of the value $\int_{\Omega} \theta_{[1]} f(\underset{\sim}{x} | \underset{\sim}{\theta}) dG(\underset{\sim}{\theta})$, and hence the computation of the Bayes rule is quite easy in this case. However, when $\Delta > 0$, the determination of the Bayes rule depends on the value $\int_{\Omega} \theta_{[1]} f(\underset{\sim}{x} | \underset{\sim}{\theta}) dG(\underset{\sim}{\theta})$ which, in general, is very hard to evaluate.

In Subsection 1.3.2, a Bayes rule for our selection problem is derived. Later, this Bayes rule is represented as a function of the marginal density functions of X_i, $i = 1, 2, \ldots, k$. Corresponding to this Bayes rule, a sequence of empirical Bayes rules is derived in

Subsection 1.3.3. Finally, the rate of convergence of the empirical Bayes rules is studied in Subsection 1.3.4.

1.3.2. A Bayesian Framework

Determination of a Bayes Rule

Let a_i, $i = 1, \ldots, 2^k - 1$, denote the $2^k - 1$ nonempty subsets of $\{1, \ldots, k\}$. When $\underset{\sim}{x} = (x_1, \ldots, x_k)$ are observed, let $P_i(\underset{\sim}{x})$ be the probability of taking action a_i. Let d be the randomized decision rule determined by $(P_i(\underset{\sim}{x}), i = 1, \ldots, 2^k - 1)$. Then, under the prior distribution G, the Bayes risk associated with d is

$$r(G, d) = \int_\Omega \int_\chi \sum_{i=1}^{2^k-1} P_i(\underset{\sim}{x})\, L(\underset{\sim}{\theta}, a_i)\, f(\underset{\sim}{x}|\underset{\sim}{\theta}) d\underset{\sim}{x}\, dG(\underset{\sim}{\theta})$$

$$= \int_\chi \sum_{i=1}^{2^k-1} P_i(\underset{\sim}{x}) \int_\Omega L(\underset{\sim}{\theta}, a_i)\, f(\underset{\sim}{x}|\underset{\sim}{\theta}) dG(\underset{\sim}{\theta}) d\underset{\sim}{x}$$

$$= \int_\chi \sum_{i=1}^{2^k-1} P_i(\underset{\sim}{x}) \int_\Omega \Big[\sum_{j \in a_i} (\theta_j - \theta_{[1]} - \Delta) I_{[\theta_{[1]}+\Delta, \infty)}(\theta_j)$$

$$+ \sum_{j \notin a_i} (\theta_{[1]} + \Delta - \theta_j) I_{[c, \theta_{[1]}+\Delta)}(\theta_j) \Big]$$

$$\cdot f(\underset{\sim}{x}|\underset{\sim}{\theta}) dG(\underset{\sim}{\theta}) d\underset{\sim}{x} \qquad (1.3.4)$$

$$= \int_\chi \sum_{i=1}^{2^k-1} P_i(\underset{\sim}{x}) \int_\Omega \Big[\sum_{j \in a_i} (\theta_j - \theta_{[1]} - \Delta)$$

$$+ \sum_{j=1}^{k} (\theta_{[1]} + \Delta - \theta_j) I_{[c, \theta_{[1]}+\Delta)}(\theta_j) \Big]$$

$$\cdot f(\underset{\sim}{x}|\underset{\sim}{\theta}) dG(\underset{\sim}{\theta}) d\underset{\sim}{x}$$

$$= \int_{\chi} \sum_{i=1}^{2^k-1} P_i(\underset{\sim}{x}) \int_{\Omega} \sum_{j\in a_i} (\theta_j - \theta_{[1]} - \Delta) f(\underset{\sim}{x}|\underset{\sim}{\theta}) dG(\underset{\sim}{\theta}) d\underset{\sim}{x}$$

$$+ \int_{\chi} \int_{\Omega} \sum_{j=1}^{k} (\theta_{[1]} + \Delta - \theta_j) I_{[c, \theta_{[1]}+\Delta)}(\theta_j) f(\underset{\sim}{x}|\underset{\sim}{\theta}) dG(\underset{\sim}{\theta}) d\underset{\sim}{x}.$$

In (1.3.4), the second equality is obtained from the condition (iv) of (1.3.1) and by applying Fubini's theorem. Since the second term at the right-hand side of (1.3.4) is independent of the choice of action a_i, thus, the Bayes rule can be determined by considering only the first term.

Hence take action a_i with probability 1 if

$$\int_{\Omega} \sum_{j\in a_i} (\theta_j - \theta_{[1]} - \Delta) f(\underset{\sim}{x}|\underset{\sim}{\theta}) dG(\underset{\sim}{\theta})$$

$$= \min_{1\le m\le 2^k-1} \int_{\Omega} \sum_{j\in a_m} (\theta_j - \theta_{[1]} - \Delta) f(\underset{\sim}{x}|\underset{\sim}{\theta}) dG(\underset{\sim}{\theta}) . \tag{1.3.5}$$

In general, let

$$\mathcal{A}(\underset{\sim}{x}) = \left\{ a_i \in \mathcal{A} \,\middle|\, \int_{\Omega} \sum_{j\in a_i} (\theta_j - \theta_{[1]} - \Delta) f(\underset{\sim}{x}|\underset{\sim}{\theta}) dG(\underset{\sim}{\theta}) \right.$$

$$\left. = \min_{1\le m\le 2^k-1} \int_{\Omega} \sum_{j\in a_m} (\theta_j - \theta_{[1]} - \Delta) f(\underset{\sim}{x}|\underset{\sim}{\theta}) dG(\underset{\sim}{\theta}) \right\}.$$

A randomized decision rule d satisfying $\sum_{a\in\mathcal{A}(\underset{\sim}{x})} P\{d(\underset{\sim}{x}) = a\} = 1$ is also a Bayes rule. Since for our problem, we can restrict our attention to

nonrandomized rules. A nonrandomized Bayes rule d_G can be obtained as follows:

Let $R_{Gj}(\underset{\sim}{x}) = \int_{\Omega} (\theta_j - \theta_{[1]} - \Delta) f(\underset{\sim}{x}|\underset{\sim}{\theta}) dG(\underset{\sim}{\theta})$. Then,

(1.3.6)
$$\begin{cases} \text{(i) If } R_{Gj}(\underset{\sim}{x}) < 0 \text{ for some } j, \text{ then} \\ \quad d_G(\underset{\sim}{x}) = \{j | R_{Gj}(\underset{\sim}{x}) < 0\} \\ \quad\quad = \{j | x_j < \Delta + c\} \sqcup \{j | x_j \geq \Delta + c \text{ and } R_{Gj}(\underset{\sim}{x}) < 0\} \\ \quad\quad = N_G(\underset{\sim}{x}, \Delta + c) \sqcup S_G^1(\underset{\sim}{x}, \Delta + c) \\ \quad \text{where } N_G(\underset{\sim}{x}, \Delta + c) = \{j | x_j < \Delta + c\} \text{ and} \\ \quad\quad S_G^1(\underset{\sim}{x}, \Delta + c) = \{j | x_j \geq \Delta + c \text{ and } R_{Gj}(\underset{\sim}{x}) < 0\}. \\ \text{(ii) If } R_{Gj}(\underset{\sim}{x}) \geq 0 \text{ for all } j, \text{ then} \\ \quad d_G(\underset{\sim}{x}) = \{\min\{m | R_{Gm}(\underset{\sim}{x}) = \min_{1 \leq j \leq k} R_{Gj}(\underset{\sim}{x})\}\} \\ \quad\quad = S_G^2(\underset{\sim}{x}, \Delta + c). \end{cases}$$

We note that $R_{Gj}(\underset{\sim}{x}) \geq 0$ implies that $x_j \geq c + \Delta$.

A Representation of the Bayes Rule d_G

Note that $f_i(x|\theta_i) = \frac{\alpha \theta_i^{\alpha}}{x^{\alpha+1}} I_{[\theta_i, \infty)}(x)$. Thus, the marginal density function $f_i(x)$ of the random variable X_i is:

$$f_i(x) = \int_c^x f_i(x|\theta)dG_i(\theta) = \int_c^x \frac{\alpha\theta^\alpha}{x^{\alpha+1}} dG_i(\theta) \quad \text{for } x \geq c,$$

(1.3.7)

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

So, the marginal distribution function $F_i(x)$ of X_i is:

$$F_i(y) = \int_{\theta=c}^{\infty} F_i(y|\theta)dG_i(\theta)$$

$$= \int_{\theta=c}^{\infty} \int_{x=\theta}^{y} f_i(x|\theta)dx\, dG_i(\theta)$$

(1.3.8)

$$= \int_{\theta=c}^{\infty} \int_{x=\theta}^{y} \frac{\alpha\theta^\alpha}{x^{\alpha+1}} dx\, dG_i(\theta)$$

$$= \int_{\theta=c}^{y} (1 - \frac{\theta^\alpha}{y^\alpha})dG_i(\theta)$$

$$= G_i(y) - \frac{y}{\alpha} f_i(y)$$

Therefore, the posterior distribution $G_i(\theta|x_i)$, when $X_i = x_i$ is observed, is:

$$G_i(\theta|x_i) = \int_c^\theta dG_i(z|x_i)$$

$$= \int_c^\theta \frac{f_i(x_i|z)}{f_i(x_i)} dG_i(z)$$

(1.3.9)

$$= \frac{1}{f_i(x_i)} \int_c^\theta \frac{\alpha z^\alpha}{x_i^{\alpha+1}} dG_i(z)$$

$$= \frac{\theta^{\alpha+1} f_i(\theta)}{x_i^{\alpha+1} f_i(x_i)} \quad \text{for } c \leq \theta \leq x_i \ ,$$

and

$$G_i(\theta|x_i) = \begin{cases} 0 & \text{if } \theta < c \ , \\ \\ 1 & \text{if } \theta \geq x_i \ . \end{cases}$$

We now are going to find the posterior distribution of $\Theta_{[1]} = \min_{1\leq j\leq k} \Theta_j$ when $\underset{\sim}{X} = \underset{\sim}{x}$ are observed. Let $G_{[1]}(\cdot|\underset{\sim}{x})$ denote the posterior distribution of $\Theta_{[1]}$ given that $\underset{\sim}{X} = \underset{\sim}{x}$ are observed. Then

$$\begin{aligned} G_{[1]}(\theta|\underset{\sim}{x}) &= P\{\Theta_{[1]} \leq \theta|\underset{\sim}{x}\} \\ &= 1 - P\{\Theta_{[1]} > \theta|\underset{\sim}{x}\} \\ &= 1 - P\{\Theta_i > \theta \text{ for all } i = 1, \ldots, k|\underset{\sim}{x}\} \\ &= 1 - \prod_{i=1}^{k} P\{\Theta_i > \theta|x_i\} \\ &= 1 - \prod_{i=1}^{k} [1 - G_i(\theta|x_i)] \\ &= \begin{cases} 0 & \text{if } \theta < c \ , \\ \\ 1 - \prod_{i=1}^{k} \left[1 - \dfrac{\theta^{\alpha+1} f_i(\theta)}{x_i^{\alpha+1} f_i(x_i)}\right] & \text{if } c \leq \theta \leq x_{[1]} \ , \\ \\ 1 & \text{if } \theta \geq x_{[1]} \ , \end{cases} \end{aligned} \tag{1.3.10}$$

where $x_{[1]} = \min_{1\leq i\leq k} x_i$.

Then,

$$R_{Gj}(\underset{\sim}{x}) = \int_{\Omega} (\theta_j - \theta_{[1]} - \Delta)dG(\underset{\sim}{\theta}|\underset{\sim}{x})f(\underset{\sim}{x})$$

(1.3.11)

$$= (E_{G_j}[\Theta_j|x_j] - E_G[\Theta_{[1]}|\underset{\sim}{x}] - \Delta)f(\underset{\sim}{x})$$

where

$$E_{G_j}[\Theta_j|x_j] = \int_c^{x_j} \theta dG_j(\theta|x_j)$$

(1.3.12)

$$= x_j - \int_c^{x_j} G_j(\theta|x_j)d\theta$$

$$= x_j - \int_c^{x_j} \frac{\theta^{\alpha+1}f_j(\theta)}{x_j^{\alpha+1}f_j(x_j)}\, d\theta,$$

$$E_G[\Theta_{[1]}|\underset{\sim}{x}] = \int_c^{x_{[1]}} \theta dG_{[1]}(\theta|\underset{\sim}{x})$$

$$= x_{[1]} - \int_c^{x_{[1]}} G_{[1]}(\theta|\underset{\sim}{x})d\theta$$

(1.3.13)

$$= x_{[1]} - \int_c^{x_{[1]}} \left(1 - \prod_{i=1}^{k} \left[1 - \frac{\theta^{\alpha+1}f_i(\theta)}{x_i^{\alpha+1}f_i(x_i)}\right]\right)d\theta$$

$$= c + \int_c^{x_{[1]}} \prod_{i=1}^{k} \left[1 - \frac{\theta^{\alpha+1}f_i(\theta)}{x_i^{\alpha+1}f_i(x_i)}\right] d\theta$$

and $f(\underset{\sim}{x}) = \prod_{i=1}^{k} f_i(x_i)$ is the marginal joint density function of $\underset{\sim}{x}$.

Hence,

$$R_{Gj}(\underset{\sim}{x}) = \Bigg[x_j - \int_c^{x_j} \frac{\theta^{\alpha+1} f_j(\theta)}{x_j^{\alpha+1} f_j(x_j)}\, d\theta - c$$

$$- \int_c^{x_{[1]}} \prod_{i=1}^{k} \left[1 - \frac{\theta^{\alpha+1} f_i(\theta)}{x_i^{\alpha+1} f_i(x_i)}\right] d\theta - \Delta\Bigg] f(\underset{\sim}{x})$$

(1.3.14)

$$= (x_j - c - \Delta) f(\underset{\sim}{x}) - \frac{f(\underset{\sim}{x})}{x_j^{\alpha+1} f_j(x_j)} \int_c^{x_j} \theta^{\alpha+1} f_j(\theta)\, d\theta$$

$$- \frac{1}{\prod_{i=1}^{k} x_i^{\alpha+1}} \int_c^{x_{[1]}} \prod_{i=1}^{k} \left[x_i^{\alpha+1} f_i(x_i) - \theta^{\alpha+1} f_i(\theta)\right] d\theta.$$

Therefore, from (1.3.6) and (1.3.14), the Bayes rule d_G can be represented in terms of the marginal density functions f_i, $i = 1, 2, \ldots, k$.

1.3.3. Empirical Bayes Analysis

For each $j = 1, 2, \ldots, k$; $n = 1, 2, \ldots$, let $f_{jn}(x; X_{j1}, \ldots, X_{jn}) \equiv f_{jn}(x)$ be a nonnegative density estimator of $f_j(x)$. Let $f_n(\underset{\sim}{x}) = \prod_{j=1}^{k} f_{jn}(x_j)$. Define

$$R_{nj}(\underset{\sim}{x}) \equiv R_{nj}(\underset{\sim}{x};\ \underset{\sim}{X}_1,\ \ldots,\ \underset{\sim}{X}_n)$$

$$= (x_j - \Delta - c)f_n(\underset{\sim}{x}) - \frac{f_n(\underset{\sim}{x})}{x_j^{\alpha+1} f_{jn}(x_j)} \int_c^{x_j} \theta^{\alpha+1} f_{jn}(\theta)\,d\theta$$

(1.3.15)

$$- \frac{1}{\prod_{i=1}^{k} x_i^{\alpha+1}} \int_c^{x_{[1]}} \prod_{i=1}^{k} \left[x_i^{\alpha+1} f_{in}(x_i) - \theta^{\alpha+1} f_{in}(\theta) \right] d\theta.$$

Next, define the decision rule $d_n(\underset{\sim}{x}) \equiv d_n(\underset{\sim}{x};\ \underset{\sim}{X}_1,\ \ldots,\ \underset{\sim}{X}_n)$ as follows:

(1.3.16)

(i) If $x_j < c + \Delta$ for some j or $R_{nj}(\underset{\sim}{x}) < 0$ for some j, let

$$d_n(\underset{\sim}{x}) = \{j | x_j < \Delta + c\} \bigsqcup \{j | x_j \geq \Delta + c \text{ and } R_{nj}(\underset{\sim}{x}) < 0\}$$

$$= N_G(\underset{\sim}{x},\ \Delta + c) \bigsqcup S_n^1(\underset{\sim}{x},\ \Delta + c).$$

(ii) If $x_j \geq \Delta + c$ and $R_{nj}(\underset{\sim}{x}) \geq 0$ for all j, let

$$d_n(\underset{\sim}{x}) = \{\min\{m | R_{nm}(\underset{\sim}{x}) = \min_{1 \leq j \leq k} R_{nj}(\underset{\sim}{x})\}\}$$

$$= S_n^2(\underset{\sim}{x},\ \Delta + c).$$

We then have the following result.

Theorem 1.3.1. Let G be a prior distribution over the parameter space Ω. Suppose that $\int_c^\infty \theta dG_i(\theta) < \infty$ for all $i = 1, \ldots, k$. If $R_{nj}(\underset{\sim}{x}) \xrightarrow{P} R_{Gj}(\underset{\sim}{x})$ a.e. $(\underset{\sim}{x})$ as $n \to \infty$, then $r_n(G, d_n) \to r(G)$ as $n \to \infty$. Therefore, the sequence of decision rule $\{d_n\}$ is asymptotically optimal relative to the prior distribution G.

Proof: The finiteness of $\int_c^\infty \theta dG_i(\theta)$ for all $i = 1, \ldots, k$, implies that the Corollary 1 of Robbins (1964) can be applied here. Thus, it suffices to show that for each $\underset{\sim}{x} \in \chi$,

$$\int_\Omega L(\underset{\sim}{\theta}, d_n(\underset{\sim}{x}))f(\underset{\sim}{x}|\underset{\sim}{\theta})dG(\underset{\sim}{\theta}) \xrightarrow{P} \int_\Omega L(\underset{\sim}{\theta}, d_G(\underset{\sim}{x}))f(\underset{\sim}{x}|\underset{\sim}{\theta})dG(\underset{\sim}{\theta}) \quad \text{as } n \to \infty. \tag{1.3.17}$$

Now, d_G is a Bayes rule, so,

$$\begin{aligned} 0 &\leq \int_\Omega L(\underset{\sim}{\theta}, d_n(\underset{\sim}{x}))f(\underset{\sim}{x}|\underset{\sim}{\theta})dG(\underset{\sim}{\theta}) - \int_\Omega L(\underset{\sim}{\theta}, d_G(\underset{\sim}{x}))f(\underset{\sim}{x}|\underset{\sim}{\theta})dG(\underset{\sim}{\theta}) \\ &= \sum_{j \in d_n(\underset{\sim}{x})} R_{Gj}(\underset{\sim}{x}) - \sum_{j \in d_G(\underset{\sim}{x})} R_{Gj}(\underset{\sim}{x}) \\ &= \sum_{j \in d_n(\underset{\sim}{x})} \left[R_{Gj}(\underset{\sim}{x}) - R_{nj}(\underset{\sim}{x})\right] + \sum_{j \in d_G(\underset{\sim}{x})} \left[R_{nj}(\underset{\sim}{x}) - R_{Gj}(\underset{\sim}{x})\right] \\ &\quad + \sum_{j \in d_n(\underset{\sim}{x})} R_{nj}(\underset{\sim}{x}) - \sum_{j \in d_G(\underset{\sim}{x})} R_{nj}(\underset{\sim}{x}) \end{aligned} \tag{1.3.18}$$

$$\leq \sum_{j \in d_n(\underset{\sim}{x})} \left[R_{Gj}(\underset{\sim}{x}) - R_{nj}(\underset{\sim}{x})\right] + \sum_{j \in d_G(\underset{\sim}{x})} \left[R_{nj}(\underset{\sim}{x}) - R_{Gj}(\underset{\sim}{x})\right]$$

since by the definition of $d_n(\underset{\sim}{x})$, $\sum_{j \in d_n(\underset{\sim}{x})} R_{nj}(\underset{\sim}{x}) - \sum_{j \in d_G(\underset{\sim}{x})} R_{nj}(\underset{\sim}{x}) \leq 0$.

Therefore,

$$0 \leq \int_{\Omega} L(\underset{\sim}{\theta}, d_n(\underset{\sim}{x}))f(\underset{\sim}{x}|\underset{\sim}{\theta})dG(\underset{\sim}{\theta}) - \int_{\Omega} L(\underset{\sim}{\theta}, d_G(\underset{\sim}{x}))f(\underset{\sim}{x}|\underset{\sim}{\theta})dG(\underset{\sim}{\theta})$$

$$(1.3.19) \quad \leq \sum_{j \in d_n(\underset{\sim}{x})} |R_{Gj}(\underset{\sim}{x}) - R_{nj}(\underset{\sim}{x})| + \sum_{j \in d_G(\underset{\sim}{x})} |R_{nj}(\underset{\sim}{x}) - R_{Gj}(\underset{\sim}{x})|$$

$$\leq 2 \sum_{j=1}^{k} |R_{Gj}(\underset{\sim}{x}) - R_{nj}(\underset{\sim}{x})|$$

which converges to zero in probability by assumption.

Hence, (1.3.17) holds. Thus, $\{d_n\}$ is asymptotically optimal relative to the prior distribution G.

In the following, we give a sufficient condition under which $R_{nj}(\underset{\sim}{x}) \to R_{Gj}(\underset{\sim}{x})$ with probability 1. Then, the assumption in Theorem 1.3.1 that $R_{nj}(\underset{\sim}{x}) \xrightarrow{P} R_{Gj}(\underset{\sim}{x})$ a.e. $(\underset{\sim}{X})$ as $n \to \infty$, will be satisfied.

Theorem 1.3.2. Let G be a prior distribution over the parameter space Ω, satisfying $G_i(c) = 0$ for some $c > 0$ and $\int_c^{\infty} \theta dG_i(\theta) < \infty$ for all $i = 1, \ldots, k$. Let $f_{jn}(x) = f_{jn}(x; X_{j1}, \ldots, X_{jn})$ be a density estimator of $f_j(x)$, based on x and $(X_{j1}, \ldots, X_{jn})$. Suppose that

(1.3.20) $\sup_{x \geq c} |f_{jn}(x) - f_j(x)| \to 0$ with probability 1 as $n \to \infty$,

for all $j = 1, \ldots, k$.

Then, $R_{nj}(\underset{\sim}{x}) \to R_{Gj}(\underset{\sim}{x})$ with probability 1 for each $\underset{\sim}{x} \in \chi$ and $j = 1, \ldots, k$.

Proof: Note that $0 \leq f_j(x) = \int_c^x \frac{\alpha\theta^\alpha}{x^{\alpha+1}} dG_j(\theta) \leq \int_c^x \frac{c}{x} dG_j(\theta) \leq \frac{\alpha}{c} < \infty$.

So, $f_j(x)$, $j = 1, \ldots, k$, are bounded functions. Since $\sup_{x \geq c} |f_{jn}(x) - f_j(x)| \to 0$ with probability 1 as $n \to \infty$, then, there exists some positive integer, say n_0, such that for all $n \geq n_0$, $f_{jn}(x)$, $j = 1, \ldots, k$, are uniformly bounded functions. That is, there exists some constant $M_1 > 0$ such that $|f_{jn}(x)| \leq M_1$ for all $x \geq c$, $n \geq n_0$ and $j = 1, \ldots, k$.

Now, for each $\underset{\sim}{x} \in \chi$,

$$|R_{nj}(\underset{\sim}{x}) - R_{Gj}(\underset{\sim}{x})|$$

$$\leq |x_j - c - \Delta| \left| f_n(\underset{\sim}{x}) - f(\underset{\sim}{x}) \right| + \frac{1}{x_j^{\alpha+1}} \left| \frac{f_n(\underset{\sim}{x})}{f_{jn}(x_j)} \int_c^{x_j} \theta^{\alpha+1} f_{jn}(\theta) d\theta \right.$$

(1.3.21)

$$\left. - \frac{f(\underset{\sim}{x})}{f_j(x_j)} \int_c^{x_j} \theta^{\alpha+1} f_j(\theta) d\theta \right|$$

$$+ \frac{1}{\prod_{i=1}^k x_i^{\alpha+1}} \int_c^{x_{[1]}} \left| \prod_{i=1}^k \left[x_i^{\alpha+1} f_{in}(x_i) - \theta^{\alpha+1} f_{in}(\theta) \right] \right|$$

$$- \prod_{i=1}^{k} \left[x_i^{\alpha+1} f_i(x_i) - \theta^{\alpha+1} f_i(\theta) \right] \Big| d\theta$$

$$\leq |x_j - c - \Delta| |f_n(\underset{\sim}{x}) - f(\underset{\sim}{x})|$$

$$+ \frac{1}{x_j^{\alpha+1}} \left| \prod_{\substack{i=1 \\ i \neq j}}^{k} f_{in}(x_i) \right| \int_c^{x_j} \theta^{\alpha+1} |f_{jn}(\theta) - f_j(\theta)| d\theta$$

$$+ \left| \prod_{\substack{i=1 \\ i \neq j}}^{k} f_i(x_i) - \prod_{\substack{i=1 \\ i \neq j}}^{k} f_{in}(x_i) \right| \frac{1}{x_j^{\alpha+1}} \int_c^{x_j} \theta^{\alpha+1} |f_{jn}(\theta) - f_j(\theta)| d\theta$$

$$+ \frac{1}{\prod_{i=1}^{k} x_i^{\alpha+1}} \int_c^{x_{[1]}} \left| \prod_{i=1}^{k} \left[x_i^{\alpha+1} f_{in}(x_i) - \theta^{\alpha+1} f_{in}(\theta) \right] \right.$$

$$\left. - \prod_{i=1}^{k} \left[x_i^{\alpha+1} f_i(x_i) - \theta^{\alpha+1} f_i(\theta) \right] \right| d\theta.$$

By the uniform boundedness of $f_j(x)$, $f_{jn}(x)$, $n \geq n_0$ and the assumption that $\sup_{x \geq c} |f_{jn}(x) - f_j(x)| \to 0$ with probability 1 as $n \to \infty$, we see that for each $\underset{\sim}{x} \in \chi$, each term on the right-hand side of (1.3.21) tends to 0 with probability 1 as $n \to \infty$. Hence, $R_{nj}(\underset{\sim}{x}) - R_{Gj}(\underset{\sim}{x}) \to 0$ with probability 1 as $n \to \infty$ for each $\underset{\sim}{x} \in \chi$.

Estimation of Density Function $f_j(x)$

Theorems 1.3.1 and 1.3.2 imply that if sequences of uniformly consistent estimators $\{f_{jn}(x)\}$ of $f_j(x)$, $j = 1, \ldots, k$, can be

constructed, then a sequence of empirical Bayes rules for our selection problem can also be obtained. Before searching for the consistent estimators $f_{jn}(x)$, a lemma about the property of uniform continuity of $f_j(x)$ is given below.

Lemma 1.3.3. Suppose that the prior distribution $G_j(\cdot)$ satisfies the following conditions:

(1) $G_j(c) = 0$ for some $c > 0$.

(2) $G_j(\cdot)$ has a continuous probability density function g_j.

(3) $\sup_{\theta \geq c} g_j(\theta) \leq M < \infty$.

(4) $\int_c^\infty \theta dG_j(\theta) < \infty$.

Then, the marginal density function $f_j(x)$ is uniformly continuous. This lemma can be verified directly. We omit the proof here.

We now define

$$(1.3.22) \quad f_{jn}(x) = \frac{1}{nh(n)} \sum_{m=1}^{n} \phi\left(\frac{x - X_{jm}}{h(n)}\right)$$

where $\phi(x)$ is a certain density function and the sequence of positive constants $\{h(n)\}$ satisfying that $\lim_{n\to\infty} h(n) = 0$.

Let $V_{jn} = \sup_{x\in R} |f_{jn}(x) - f_j(x)|$. Nadaraya (1965) studied some situation in which $V_{jn} \to 0$ with probability 1 as $n \to \infty$. We cite his main result here.

Theorem (Nadaraya). Suppose that $\phi(x)$ is a function of bounded variation, $f_j(x)$ is a uniformly continuous density function and the series $\sum_{n=1}^{\infty} \exp(-rnh^2(n))$ converges for every positive value r. Then

$V_{jn} = \sup_{x\in R} |f_{jn}(x) - f_j(x)| \to 0$ with probability 1 as $n \to \infty$.

In our model, under the assumptions of (1.3.1), $f_j(x)$ is uniformly continuous for each $j = 1, \ldots, k$. Thus, by Nadaraya's theorem, we can construct a sequence of estimators $\{f_{jn}\}$ such that $V_{jn} \to 0$ with probability 1 for each $j = 1, \ldots, k$. Then, by Theorems 1.3.1 and 1.3.2, a sequence of empirical Bayes rules for our selection problem can also be obtained.

A Special Case when $\Delta = 0$

In case that $\Delta = 0$, we only need the assumption (iv) of (1.3.1). When $\Delta = 0$,

$$0 \le R_{Gj}(\underset{\sim}{x}) = \int_{\Omega} (\theta_j - \theta_{[1]})\, f(\underset{\sim}{x}|\underset{\sim}{\theta})dG(\underset{\sim}{\theta})$$

$$= \int_{\Omega} \theta_j f(\underset{\sim}{x}|\underset{\sim}{\theta})dG(\underset{\sim}{\theta}) - \int_{\Omega} \theta_{[1]} f(\underset{\sim}{x}|\underset{\sim}{\theta})dG(\underset{\sim}{\theta}). \tag{1.3.23}$$

Let d_G^1 be the Bayes rule for the case when $\Delta = 0$. Then,

$$d_G^1(\underset{\sim}{x}) = \{\min\{m | R_{Gm}(\underset{\sim}{x}) = \min_{1\le j\le k} R_{Gj}(\underset{\sim}{x})\}\}$$

$$= \{\min\{m \Big| \int_{\Omega} \theta_m f(\underset{\sim}{x}|\underset{\sim}{\theta})dG(\underset{\sim}{\theta}) = \min_{1\le j\le k} \int_{\Omega} \theta_j f(\underset{\sim}{x}|\underset{\sim}{\theta})dG(\underset{\sim}{\theta})\}\}. \tag{1.3.24}$$

Note that here the determination of the Bayes rule d_G^1 is independent of the value $\int_\Omega \theta_{[1]} f(\underset{\sim}{x}|\underset{\sim}{\theta})dG(\underset{\sim}{\theta})$. Let

$$T_{Gj}(\underset{\sim}{x}) = \int_\Omega \theta_j f(\underset{\sim}{x}|\underset{\sim}{\theta})dG(\underset{\sim}{\theta})$$

$$= x_j f(\underset{\sim}{x}) - \frac{1}{x_j^{\alpha+1}} \prod_{\substack{i=1 \\ i\neq j}}^{k} f_i(x_i) \int_0^{x_j} \theta^{\alpha+1} f_j(\theta)d\theta \tag{1.3.25}$$

$$= x_j f(\underset{\sim}{x}) - \frac{m_j(\alpha,x_j)}{x_j^{\alpha+1}} \prod_{\substack{i=1 \\ i\neq j}}^{k} f_i(x_i)$$

where

$$m_j(\alpha, x_j) = \int_0^{x_j} \theta^{\alpha+1} f_j(\theta)d\theta = E[X_j^{\alpha+1} I_{(0,x_j)}(X_j)]. \tag{1.3.26}$$

Therefore, let $f_{jn}(x) \equiv f_{jn}(x, X_{j1}, \ldots, X_{jn})$ be a density estimator of $f_j(x)$, based on x and $(X_{j1}, \ldots, X_{jn})$. Let

$$m_{jn}(\alpha, x) = \frac{1}{n} \sum_{m=1}^{n} X_{jm}^{\alpha+1} I_{(0,x]}(X_{jm}). \tag{1.3.27}$$

Define

$$T_{nj}(\underset{\sim}{x}) = x_j f_n(\underset{\sim}{x}) - \frac{m_{jn}(\alpha,x_j)}{x_j^{\alpha+1}} \prod_{\substack{i=1 \\ i\neq j}}^{k} f_{in}(x_i), \tag{1.3.28}$$

and

$$(1.3.29)\quad d_n^1(\underset{\sim}{x}) \equiv d_n^1(\underset{\sim}{x}, \underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_n) = \{\min\{m \mid T_{nm}(\underset{\sim}{x}) = \min_{1\le j\le k} T_{nj}(\underset{\sim}{x})\}\}.$$

Then, similar to Theorem 1.3.1, we get the following theorem.

Theorem 1.3.4. Let $G = \prod_{i=1}^{k} G_i$ be a prior distribution over the parameter space Ω, satisfying that $\int_0^\infty \theta dG_i(\theta) < \infty$ for all $i = 1, \ldots, k$. Suppose that $T_{nj}(\underset{\sim}{x}) \xrightarrow{P} T_{Gj}(\underset{\sim}{x})$ a.e. $(\underset{\sim}{X})$ as $n \to \infty$. Then, the sequence of decision rules $\{d_n^1\}$ is asymptotically optimal relative to the prior distribution G.

Proof: The proof of this theorem is similar to that of Theorem 1.3.1 and hence omitted.

Note that for each $x \ge 0$, $m_{jn}(\alpha, x)$ can be viewed as the sample mean of bounded i.i.d. random variables $X_{jm}^{\alpha+1} I_{(0,x]}(X_{jm})$, $m = 1, \ldots, n$. It is also an unbiased estimator of $m_j(\alpha, x)$. Thus, $m_{jn}(\alpha, x) \to m_j(\alpha, x)$ with probability 1 as $n \to \infty$. What remains is that we need to find a sequence of estimators $\{f_{jn}(x)\}$ such that $f_{jn}(x) \xrightarrow{P} f_j(x)$ a.e. (X_j) as $n \to \infty$. If all the conditions of (1.3.1) are satisfied, we can still apply Nadaraya's theorem to derive such a sequence of estimators. However, here, only condition (iv) of (1.3.1) is available. Thus, we need to consider some other approach. First we give a lemma similar to Lemma 1.3.3 as follows:

Lemma 1.3.5. The continuity of the prior distribution $G_i(\theta)$ implies the continuity of $f_i(x)$.

Now, define

$$(1.3.30)\qquad f_{in}(x) = \frac{1}{nh(n)} \sum_{j=1}^{n} \psi\left(\frac{x - X_{ij}}{h(n)}\right),$$

where the kernel density function $\psi(\cdot)$ satisfying

$$(1.3.31)\qquad (i)\ \sup_{x\in R} \psi(x) < \infty,\quad (ii)\ \lim_{|x|\to\infty} |x\psi(x)| = 0,$$

and the sequence of positive constants $\{h(n)\}$ satisfying

$$(1.3.32)\qquad (i)\ \lim_{n\to\infty} h(n) = 0,\quad (ii)\ \lim_{n\to\infty} nh(n) = \infty.$$

Then,

Theorem 1.3.6. Let G be a prior distribution over the parameter space Ω, satisfying (1) $\int_0^\infty \theta dG_i(\theta) < \infty$ and (2) $G_i(\theta)$ is a continuous function of θ for all $i = 1, \ldots, k$. Let $\{f_{in}(x)\}$ be a sequence of estimators defined in (1.3.30), where the kernel density function ψ and the sequence of positive constants $\{h(n)\}$ satisfy the conditions of (1.3.31) and (1.3.32), respectively. Then, the sequence of decision rules $\{d_n^1\}$ is asymptotically optimal relative to the prior distribution G.

Proof: By our assumption and by Lemma 1.3.5, $f_j(x)$, $j = 1, \ldots, k$, are continuous function. Then by Parzen (1962), $f_{jn}(x) \xrightarrow{P} f_j(x)$ a.e. (X_j) as $n\to\infty$ for each $j = 1, \ldots, k$. Also, we have shown that

$m_{jn}(\alpha, x) \to m_j(\alpha, x)$ with probability 1 as $n \to \infty$. Therefore, $T_{nj}(\underset{\sim}{x}) \xrightarrow{P} T_{Gj}(\underset{\sim}{x})$ a.e. $(\underset{\sim}{X})$ as $n \to \infty$. Hence, this theorem is a direct result of Theorem 1.3.4.

1.3.4. Rate of Convergence of the Empirical Bayes Rules

In this subsection, we assume that $\alpha > 1$. This assumption and condition (iv) of (1.3.1) together imply that $E[X_i] < \infty$ since

$$E[X_i] = E_{G_i}[E[X_i|\theta_i]] = \frac{\alpha}{\alpha - 1} E_{G_i}[\theta_i] < \infty.$$

When $\Delta > 0$, let $S_j = \{\underset{\sim}{x} | j \in d_G(\underset{\sim}{x})\}, S_{nj} = \{\underset{\sim}{x} | j \in d_n(\underset{\sim}{x})\}$. Then, we have the following theorem:

Theorem 1.3.7. Assume that $P\{\underset{\sim}{x} \in S_{nj} - S_j\} = O(\alpha_n)$, $P\{\underset{\sim}{x} \in S_j - S_{nj}\} = O(\beta_n)$ for all $\underset{\sim}{x} \in \chi$, where $\lim_{n\to\infty} \alpha_n = \lim_{n\to\infty} \beta_n = 0$. Let $\gamma_n = \max(\alpha_n, \beta_n)$. Then, the sequence of empirical Bayes rules $\{d_n\}$ is asymptotically optimal at least of order $\{\gamma_n\}$ relative to the prior G.

Proof: We see that

$$\begin{aligned} 0 &\leq r_n(G, d_n) - r(G) \\ &= \int_\chi \left\{ E\left[\sum_{j \in d_n(\underset{\sim}{x})} R_{Gj}(\underset{\sim}{x}) \right] - \sum_{j \in d_G(\underset{\sim}{x})} R_{Gj}(\underset{\sim}{x}) \right\} d\underset{\sim}{x} \\ &= \int_\chi \left\{ E\left[\sum_{j=1}^{k} R_{Gj}(\underset{\sim}{x})\, I_{S_{nj}}(\underset{\sim}{x}) \right] - \sum_{j=1}^{k} R_{Gj}(\underset{\sim}{x})\, I_{S_j}(\underset{\sim}{x}) \right\} d\underset{\sim}{x} \end{aligned} \tag{1.3.33}$$

$$= \sum_{j=1}^{k} \int_{\chi} R_{Gj}(\underset{\sim}{x})\ E\left[I_{S_{nj}}(\underset{\sim}{x}) - I_{S_j}(\underset{\sim}{x})\right] d\underset{\sim}{x}$$

$$= \sum_{j=1}^{k} \int_{\chi} R_{Gj}(\underset{\sim}{x})\left[p\{\underset{\sim}{x}\ \varepsilon S_{nj} - S_j\} - p\{\underset{\sim}{x}\ \varepsilon S_j - S_{nj}\}\right] d\underset{\sim}{x}$$

$$= \sum_{j=1}^{k} \int_{H_j^1 \cup H_j^2} R_{Gj}(\underset{\sim}{x})\left[p\{\underset{\sim}{x}\ \varepsilon S_{nj} - S_j\} - p\{\underset{\sim}{x}\ \varepsilon S_j - S_{nj}\}\right] d\underset{\sim}{x}$$

where $H_j^1 = \{\underset{\sim}{x} | R_{Gj}(\underset{\sim}{x}) < 0\}$ and $H_j^2 = \{\underset{\sim}{x} | R_{Gj}(\underset{\sim}{x}) \geq 0\}$.

When $\underset{\sim}{x} \varepsilon H_j^1$, $-\Delta f(\underset{\sim}{x}) \leq R_{Gj}(\underset{\sim}{x}) = \int_{\Omega} (\theta_j - \theta_{[1]} - \Delta) dG(\underset{\sim}{\theta} | \underset{\sim}{x}) f(\underset{\sim}{x}) < 0$.
Hence,

$$\int_{H_j^1} R_{Gj}(\underset{\sim}{x})\left[p\{\underset{\sim}{x}\ \varepsilon S_{nj} - S_j\} - p\{\underset{\sim}{x}\ \varepsilon S_j - S_{nj}\}\right] d\underset{\sim}{x}$$

(1.3.34)

$$\leq \int_{H_j^1} \Delta\left[p\{\underset{\sim}{x}\ \varepsilon S_{nj} - S_j\} + p\{\underset{\sim}{x}\ \varepsilon S_j - S_{nj}\}\right] f(\underset{\sim}{x}) d\underset{\sim}{x}$$

When $\underset{\sim}{x} \varepsilon H_j^2$, $0 \leq R_{Gj}(\underset{\sim}{x}) = \int_{\Omega} (\theta_j - \theta_{[1]} - \Delta) dG(\underset{\sim}{\theta} | \underset{\sim}{x}) f(\underset{\sim}{x})$

$$\leq \int_{\Omega} \theta_j dG(\underset{\sim}{\theta} | \underset{\sim}{x}) f(\underset{\sim}{x})$$

$$\leq x_j f(\underset{\sim}{x}).$$

Hence,

$$\int_{H_j^2} R_{Gj}(\underset{\sim}{x})\left[P\{\underset{\sim}{x}\ \varepsilon S_{nj} - S_j\} - p\{\underset{\sim}{x}\ \varepsilon S_j - S_{nj}\}\right] d\underset{\sim}{x}$$

(1.3.35)

$$\leq \int_{H_j^2} x_j\left[p\{\underset{\sim}{x}\ \varepsilon S_{nj} - S_j\} + p\{\underset{\sim}{x}\ \varepsilon S_j - S_{nj}\}\right] f(\underset{\sim}{x}) d\underset{\sim}{x}\ .$$

From (1.3.33) ~ (1.3.35), we obtain

$$0 \le r_n(G, d_n) - r(G)$$

$$\le \sum_{j=1}^{k} \int_{H_j^1 \cup H_j^2} (\Delta + x_j)\left[p\{\underset{\sim}{x} \in S_{nj} - S_j\} + p\{\underset{\sim}{x} \in S_j - S_{nj}\}\right] f(\underset{\sim}{x})d\underset{\sim}{x}$$

$$= \sum_{j=1}^{k} \int_{\chi} (\Delta + x_j)\left[p\{\underset{\sim}{x} \in S_{nj} - S_j\} + p\{\underset{\sim}{x} \in S_j - S_{nj}\}\right] f(\underset{\sim}{x})d\underset{\sim}{x} \qquad (1.3.36)$$

$$= \sum_{j=1}^{k} \int_{\chi} (\Delta + x_j)\, O(\gamma_n) f(\underset{\sim}{x})d\underset{\sim}{x}$$

$$= \left(\sum_{j=1}^{k} (\Delta + E[X_j])\right) O(\gamma_n)$$

$$= O(\gamma_n).$$

This implies that the sequence of decision rules $\{d_n\}$ is asymptotically optimal at least of order $\{\gamma_n\}$ relative to the prior G.

Similarly, for $\Delta = 0$, let $S_j^1 = \{\underset{\sim}{x} | j \in d_G^1(\underset{\sim}{x})\}$, $S_{nj}^1 = \{\underset{\sim}{x} | j \in d_n^1(\underset{\sim}{x})\}$. Then, we have the following theorem:

Theorem 1.3.8. Assume that $p\{\underset{\sim}{x} \in S_{nj}^1 - S_j^1\} = O(\alpha_n^1)$, $p\{\underset{\sim}{x} \in S_j^1 - S_{nj}^1\} = O(\beta_n^1)$ for all $\underset{\sim}{x} \in \chi$, where $\lim_{n\to\infty} \alpha_n^1 = \lim_{n\to\infty} \beta_n^1 = 0$. Let $\gamma_n^1 = \max(\alpha_n^1, \beta_n^1)$. Then, the sequence of empirical Bayes rules $\{d_n^1\}$ is asymptotically optimal at least of order $\{\gamma_n^1\}$ relative to the prior distribution G.

Proof: Note that from (1.3.25), $0 \leq T_{Gj}(\underset{\sim}{x}) \leq x_j f(\underset{\sim}{x})$. Thus,

$$0 \leq r_n(G, d_n^1) - r(G)$$

$$= \int_{\mathcal{X}} \left\{ E\left[\sum_{j \in d_n^1(\underset{\sim}{x})} T_{Gj}(\underset{\sim}{x}) \right] - \sum_{j \in d_G^1(\underset{\sim}{x})} T_{Gj}(\underset{\sim}{x}) \right\} d\underset{\sim}{x}$$

$$= \int_{\mathcal{X}} \left\{ E\left[\sum_{j=1}^{k} T_{Gj}(\underset{\sim}{x})\, I_{S_{nj}^1}(\underset{\sim}{x}) \right] - \sum_{j=1}^{k} T_{Gj}(\underset{\sim}{x})\, I_{S_j^1}(\underset{\sim}{x}) \right\} d\underset{\sim}{x}$$

$$= \sum_{j=1}^{k} \int_{\mathcal{X}} T_{Gj}(\underset{\sim}{x}) E\left[I_{S_{nj}^1}(\underset{\sim}{x}) - I_{S_j^1}(\underset{\sim}{x}) \right] d\underset{\sim}{x} \tag{1.3.37}$$

$$\leq \sum_{j=1}^{k} \int_{\mathcal{X}} T_{Gj}(\underset{\sim}{x}) \left[p\left\{ \underset{\sim}{x} \in S_{nj}^1 - S_j^1 \right\} + p\left\{ \underset{\sim}{x} \in S_j^1 - S_{nj}^1 \right\} \right] d\underset{\sim}{x}$$

$$\leq \sum_{j=1}^{k} \int_{\mathcal{X}} x_j f(\underset{\sim}{x}) \left[p\left\{ \underset{\sim}{x} \in S_{nj}^1 - S_j^1 \right\} + p\left\{ \underset{\sim}{x} \in S_j^1 - S_{nj}^1 \right\} \right] d\underset{\sim}{x}$$

$$= \sum_{j=1}^{k} \int_{\mathcal{X}} x_j f(\underset{\sim}{x})\, O(\gamma_n^1) d\underset{\sim}{x}$$

$$= \left(\sum_{j=1}^{k} E[X_j] \right) O(\gamma_n^1)$$

$$= O(\gamma_n^1).$$

Hence, the sequence of decision rules $\{d_n^1\}$ is asymptotically optimal at least of order $\{\gamma_n^1\}$ relative to the prior G.

CHAPTER II

SEQUENTIAL SUBSET SELECTION PROCEDURES

2.1. Introduction

An important class of statistical problems is concerned with the selection and ranking of k populations. The selection and ranking may be defined in terms of a parameter of the population which may physically represent the mean, the variance or some quantile of the distribution. Consider designing and analyzing an experiment for comparing k populations $\pi_1, \ldots, \pi_k$. The quality or goodness of population π_i is characterized by the value of an unknown parameter θ_i for $i = 1,2,\ldots,k$. Let $\delta_{ij} = \delta(\theta_i, \theta_j)$ be a measure of separation between π_i and π_j, and let $\overline{\delta} = \max_{1\le i\le k} \min_{j\neq i} \{\delta_{ij}\}$. Population π_i is called the best population if π_i is the unique population having the property that $\min_{j\neq i}\{\delta_{ij}\} = \overline{\delta}$. If more than one population has this property, one of them is tagged and considered as the best population.

Suppose that observations can be obtained from the k populations sequentially. It is often desirable to terminate sampling from a population as soon as there is statistical evidence that it is not the best population, and this population is eliminated from further consideration. Selection through sequential comparison with elimination provides a significant advantage. To achieve a certain accuracy, it requires, on the average, substantially fewer samples than the fixed

sample size procedure. This average saving can be very large, especially due to the rejection of the populations that are clearly not the best.

In sequential selection and ranking procedures, contributions have been made to select the best population by using the indifference zone approach. The simplest formulation of the indifference zone approach is the situation where one may wish to select only a single population and guarantee with probability P^* that the selected population is the best population provided some other condition on the parameters is satisfied, usually an indifference zone. However, in the real situation, it is hard or impossible to specify that condition. Therefore, as in the fixed sample size case (see Gupta (1965)), a reasonable and useful approach is to derive some sequential selection procedure to select a small subset containing the best population. This sequential selection procedure should satisfy the P^*-condition regardless of all the possible configurations of the parameters. Thus, in this type of formulation, there is no indifference zone configuration. However, the size of the selected subset is a random variable. We desire that each of the selected population is not far from the best population. Therefore, some inference is needed here to assert that each selected population is within some prespecified distance from the best population.

In Section 2.2, some parametric sequential subset selection procedures achieving the goal described above are derived. These procedures are formed by choosing an invariant statistic of the parameters of interest, based on the observations from any pair of

two of the k populations and by performing a modified sequential probability ratio test (MSPRT) based on this statistic. This is done simultaneously for all pairs of populations and if a particular MSPRT terminates, then an appropriate population is removed from the set of contending populations. This is continued until only one population belongs to this set or some statistical evidence shows that all the populations remaining in this set are within some prespecified distance from the unknown best population. At each stage, these procedures also provide some statistical inference about the bounds on the measure of separation between each remaining population and the unknown best population.

In Section 2.3, we study the problem of selecting the best population among k populations belonging to the class of exponential family of distributions. We use the same measure of separation as that considered by Bechhofer, Kiefer and Sobel (1968) for selecting one or more fixed number of populations using the indifference zone approach. For this particular measure of separation, we consider an appropriate transformation of the random observations taken from any two populations. With this transformation, the likelihood function of the new statistics can be factored into two parts, one of which, obtained by a conditional argument, and termed the conditional likelihood function, is a function only of the parameter of interest. Based on this conditional likelihood function, a sequential subset selection procedure is derived. This sequential subset selection procedure achieves the goal described above. At each stage, it also provides some statistical inference about the bounds of separation

between each remaining population and the unknown best population.

Finally, some applications of these sequential selection procedures are studied in Section 2.4. Simulation studies are carried out to see how these sequential selection procedures perform.

2.2. Sequential Subset Selection Procedures Based on Invariant Statistics

2.2.1. Formulation of the Problem

Let $\pi_1, \ldots, \pi_k$ represent $k(\geq 2)$ populations and let X_{in} denote the nth observation from population π_i, $i = 1, \ldots, k$. It is assumed that the observations $\{X_{in}\}$ are independent random variables for all i and n. Suppose that X_{in} has distribution function $F(x|\theta_i)$ depending on some unknown parameter θ_i for all $i = 1, \ldots, k$. Let $\underset{\sim}{\theta} = (\theta_1, \ldots, \theta_k)$ and let $\Omega = \{\underset{\sim}{\theta} | \underset{\sim}{\theta} = (\theta_1, \ldots, \theta_k)\}$ be the parameter space. For each pair i and j, let $\delta_{ij} = \delta(\theta_i, \theta_j)$ be a measure of separation between populations π_i and π_j where $\delta(\theta_i, \theta_j)$ is a function of θ_i and θ_j, and increasing (decreasing) in θ_i (θ_j) when θ_j (θ_i) is fixed, and satisfies that $\delta(\theta, \theta) = \delta_0$ for all θ. Define $\overline{\delta} = \max_{1\leq i\leq k} \min_{j\neq i}\{\delta_{ij}\} = \max_{1\leq i\leq k} \overline{\delta}_i$ where $\overline{\delta}_i = \min_{j\neq i} \{\delta_{ij}\}$. Population π_i is called the best population if π_i is the unique population having the property that $\overline{\delta}_i = \overline{\delta}$. If more than one population has this property, one of them is tagged, and considered as the best population. We use (k) to denote the index of the best population. For example, if the observations taken from π_i have normal distribution with unknown mean θ_i and common known variance σ^2, we may be

interested in selecting the population with the largest mean. Thus, we define $\delta_{ij} = \theta_i - \theta_j$. Then $\overline{\delta} = \theta_{[k]} - \theta_{[k-1]}$ where $\theta_{[1]} \leq \theta_{[2]} \leq \ldots \leq \theta_{[k]}$ are the ordered parameters of $\theta_1, \ldots, \theta_k$. Therefore, $\pi_{(k)}$ is the population having the largest mean $\theta_{[k]}$.

Suppose that observations can be obtained from the k populations sequentially. The selection procedure will depend upon the observations through a sequence of statistics $\{T_{ij}(n), n \geq 1\}$ which are defined to be functions

$$(2.2.1) \qquad T_{ij}(n) = T_n(X_{i1}, \ldots, X_{in}; X_{j1}, \ldots, X_{jn})$$

of the first n observations from populations π_i and π_j. In a given problem, the function T_n is chosen as to indicate the difference between the populations in a reasonable way. Let $\tilde{T}_{ij}(n) = (T_{ij}(1), \ldots, T_{ij}(n))$. We assume that $\tilde{T}_{ij}(n)$ has a joint probability density $g_n(\tilde{T}_{ij}(n)|\delta_{ij})$ depending on the parameter δ_{ij}. Usually, $T_{ij}(1), T_{ij}(2), \ldots,$ are chosen so that it is both a sufficient and transitive sequence and also invariantly sufficient for δ_{ij} (see Hall, Wijsman and Ghosh (1965) and Hoel (1971)).

We assume that there is no information about the separation δ_{ij}, $i, j = 1, \ldots, k$, $i \neq j$, and we use the subset selection approach. However, the size of the selected subset is a random variable. Thus, we desire that each selected population should not be far from the best population. Let S denote the selected subset and CS denote the event that the best population $\pi_{(k)}$ is selected in the subset S. Let $\delta_{i(k)}$ denote the measure of separation from the population π_i to

the best population $\pi_{(k)}$. Then, by the property of function $\delta(\cdot, \cdot)$, $\delta_{i(k)} \leq \delta_0$. For a prespecified value $\delta_* < \delta_0$, population π_i is said to be good if $\delta_{i(k)} \geq \delta_*$. Now let $CD(\delta_*)$ denote the event that $\pi_{(k)} \epsilon S$ and for each $\pi_i \epsilon S$, $\delta_* \leq \delta_{i(k)} \leq \delta_0$. Clearly, $CD(\delta_*)$ implies CS. Thus, for a prespecified value P^*, $1/k < P^* < 1$, we require a sequential subset selection procedure satisfying

$$(2.2.2) \qquad P_{\underset{\sim}{\theta}}\{CD(\delta_*)\} \geq P^* \quad \text{for all } \underset{\sim}{\theta} \epsilon \Omega.$$

2.2.2. Sequential Subset Selection Procedure $P_1(\delta_*, \delta_1)$

Let $h(\cdot)$ be a monotonically decreasing function such that $h(\delta_{ij}) = \delta_{ji}$. Let $\delta_* < \delta_0$ be a prespecified value used to specify the event $CD(\delta_*)$. Then $\delta_0 = h(\delta_0) < h(\delta_*)$. Let $a \leq \delta_0 < \delta_1 < h(\delta_*)$. Consider the likelihood ratio statistics

$$(2.2.3) \qquad L_{ij}(n, \delta_1, a) = \frac{g_n(\tilde{T}_{ij}(n)|\delta_1)}{g_n(\tilde{T}_{ij}(n)|a)}, \quad (n \geq n_0)$$

which have been used by Hoel (1971) and Gupta and Huang (1975), respectively, to construct sequential selection procedures where n_0 is the initial sampling size of the procedure. For simplicity, we assume that $n_0 = 1$. We now define a sequential subset selection procedure $P_1(\delta_*, \delta_1)$ as follows:

Let $B_0 = \{\pi_1, \ldots, \pi_k\}$. First, we take one observation from each population. Compute the values of the $2k(k - 1)$ statistics $L_{ij}(1, \delta_1, \delta_0)$ and $L_{ij}(1, \delta_1, \delta_*)$. Eliminate π_j from further consideration if there is some π_i such that $L_{ij}(1, \delta_1, \delta_0) \geq \frac{k - 1}{1 - P^*}$.

Let B_1 be the set of remaining populations. For each $\pi_i \varepsilon B_1$, π_i is labelled as good if $L_{ij}(1, \delta_1, \delta_*) \geq \frac{k - 1}{1 - P^*}$ for all $\pi_j \varepsilon B_1 - \{\pi_i\}$. Let $|A|$ denote the number of elements in set A. Then, the procedure terminates at stage 1 if either $|B_1| = 1$ or all populations in B_1 have been labelled as good. In this situation, B_1 is taken as the selected subset S. Otherwise, we go to the next stage. In general, for each $n \geq 1$, let

$$(2.2.4) \qquad B_n = \{\pi_i \varepsilon B_{n-1} | L_{ji}(n, \delta_1, \delta_0) < \frac{k - 1}{1 - P^*}$$

$$\text{for all } \pi_j \varepsilon B_{n-1} - \{\pi_i\}\}.$$

That is, B_n is the set of populations which have not been eliminated up to stage n. For each $\pi_i \varepsilon B_n$, label π_i as a good population if

$$(2.2.5) \qquad L_{ij}(n, \delta_1, \delta_*) \geq \frac{k - 1}{1 - P^*} \quad \text{for all} \quad \pi_j \varepsilon B_n - \{\pi_i\}.$$

If either $|B_n| = 1$ or all the populations in B_n have been labelled as good, then the procedure terminates and we take $S = B_n$. Otherwise, we go to the next stage. The procedure is continued in this way.

2.2.3. Probability of a Correct Selection

Let $g_m(t|\tilde{t}(m - 1), \delta)$ denote the conditional probability density function of $T_{ij}(m)$ given that $\tilde{T}_{ij}(m - 1) = \tilde{t}(m - 1)$, and $L_{ij}(n, \delta_1, a)$ be the statistic defined in (2.2.3). Then, the statistic $L_{ij}(n, \delta_1, a)$ can be rewritten as:

$$(2.2.6)\qquad L_{ij}(n, \delta_1, a) = \frac{g_1(T_{ij}(1)|\delta_1)}{g_1(T_{ij}(1)|a)} \prod_{m=2}^{n} \frac{g_m(T_{ij}(m)|\tilde{T}_{ij}(m-1), \delta_1)}{g_m(T_{ij}(m)|\tilde{T}_{ij}(m-1), a)},$$

$$n \geq 1,$$

where $\prod_{m=2}^{n}$ [ ] = 1 if $n = 1$. For each $n \geq 1$, let $\mathcal{F}_{ij}(n)$ denote the σ-field generated by $(T_{ij}(1), \ldots, T_{ij}(n))$. Then,

Lemma 2.2.1. $\{L_{i(k)}(n, \delta_1, \delta_{i(k)}), P_{\underset{\sim}{\theta}}, \mathcal{F}_{i(k)}(n), n \geq 1\}$ forms a nonnegative martingale for each $i \neq (k)$.

Proof: This lemma can be easily verified by direct computation.

Theorem 2.2.2. Let $\{B_n\}$ be the sequence of sets of remaining populations defined in (2.2.4). Then,

$$(2.2.7)\qquad P_{\underset{\sim}{\theta}}\Big\{L_{i(k)}(n, \delta_1, \delta_{i(k)}) < \frac{k-1}{1-P^*} \text{ for all } \pi_i \in B_{n-1} - \{\pi_{(k)}\}$$

$$\text{for all } n \geq 1\Big\} \geq P^* \text{ for all } \underset{\sim}{\theta} \in \Omega.$$

Proof: Note that for any $\underset{\sim}{\theta} \in \Omega$,

$$P_{\underset{\sim}{\theta}}\Big\{L_{i(k)}(n, \delta_1, \delta_{i(k)}) < \frac{k-1}{1-P^*} \text{ for all } \pi_i \in B_{n-1} - \{\pi_{(k)}\}$$

$$\text{for all } n \geq 1\Big\}$$

$$= 1 - P_{\underset{\sim}{\theta}}\Big\{L_{i(k)}(n, \delta_1, \delta_{i(k)}) \geq \frac{k-1}{1-P^*}$$

$$\text{for some } \pi_i \in B_{n-1} - \{\pi_{(k)}\} \text{ for some } n \geq 1\Big\}$$

(2.2.8)

$$\geq 1 - \sum_{\substack{i=1 \\ i \neq (k)}}^{k} P_{\underset{\sim}{\theta}}\Big\{L_{i(k)}(n, \delta_1, \delta_{i(k)}) \geq \frac{k-1}{1-P^*} \text{ for some } n \geq 1\Big\}.$$

Next, by Lemma 2.2.1, $\Big\{L_{i(k)}(n, \delta_1, \delta_{i(k)}), P_{\underset{\sim}{\theta}}, \mathcal{F}_{i(k)}(n), n \geq 1\Big\}$ forms a nonnegative martingale. Hence,

$$P_{\underset{\sim}{\theta}}\Big\{L_{i(k)}(n, \delta_1, \delta_{i(k)}) \geq \frac{k-1}{1-P^*} \text{ for some } n \geq 1\Big\} \leq \frac{1-P^*}{k-1}. \tag{2.2.9}$$

Then by (2.2.8) and (2.2.9), we obtain

$$P_{\underset{\sim}{\theta}}\Big\{L_{i(k)}(n, \delta_1, \delta_{i(k)}) < \frac{k-1}{1-P^*} \text{ for all } \pi_i \in B_{n-1} - \{\pi_{(k)}\}$$

$$\text{for all } n \geq 1\Big\}$$

(2.2.10)

$$\geq 1 - \sum_{\substack{i=1 \\ i \neq (k)}}^{k} \frac{1-P^*}{k-1} = P^*.$$

Therefore, the proof of this theorem is completed.

For each $a \leq \delta_0$ (the value a is chosen so that the joint probability density function $g_n(\tilde{T}_{ij}(n)|a)$ is well defined), let $A_{ij}(n, a)$ denote the event that $L_{ij}(n, \delta_1, a) < \frac{k-1}{1-P^*}$. That is,

$A_{ij}(n, a) = \left\{L_{ij}(n, \delta_1, a) < \frac{k - 1}{1 - P^*}\right\}$. In the following, we assume that the condition of (2.2.11) is satisfied.

$$(2.2.11) \qquad \bigcap_{n=1}^{m} A_{ij}(n, b) \sqsubset \bigcap_{n=1}^{m} A_{ij}(n, a) \text{ for all } m \geq 1 \text{ for } b < a \leq \delta_0.$$

The implication of (2.2.11) is that the values of the statistics $L_{ij}(n, \delta_1, a)$ for $n \geq 1$, never exceed the boundary level $\frac{k - 1}{1 - P^*}$ before that of the statistics $L_{ij}(n, \delta_1, b)$, $n \geq 1$ when $b < a \leq \delta_0$. A sufficient condition for (2.2.11) is that $A_{ij}(n,b) \sqsubset A_{ij}(n,a)$ for all $n \geq 1$.

For each $n \geq 1$, π_i, $\pi_j \in B_{n-1}$, as $i \neq j$, define

$$(2.2.12) \qquad D_{ij}^1(n) = \inf\left\{a \leq \delta_0 \middle| L_{ij}(n, \delta_1, a) < \frac{k - 1}{1 - P^*}\right\} \text{ if } \{\ \} \neq \phi$$

$$= \delta_0 \text{ if } \{\ \} = \phi$$

where ϕ denotes the empty set. Also, let $D_{ii}^1(n) = \delta_0$.

Under the assumption (2.2.11), if $D_{ij}^1(n) < \delta_0$, then $L_{ij}(n, \delta_1, a) < \frac{k - 1}{1 - P^*}$ for all $D_{ij}^1(n) < a \leq \delta_0$ and $L_{ij}(n, \delta_1, b) \geq \frac{k - 1}{1 - P^*}$ for all $b < D_{ij}^1(n)$.

For each $n \geq 1$, as $\pi_i \in B_{n-1}$, define

$$(2.2.13) \qquad D_i^1(n) = \max_{1 \leq m \leq n} \left[\min_{\pi_j \in B_{m-1}} D_{ij}^1(m)\right].$$

As $\pi_i \notin B_{n-1}$, let $n_i = \max\{m | \pi_i \in B_{m-1}\}$ and define $D_i^1(n) = D_i^1(n_i)$.

Note that by the definition of $D_i^1(n)$, for each $i = 1, 2, \ldots, k$, $\{D_i^1(n), n \geq 1\}$ is an increasing sequence and bounded above by δ_0. The value $D_i^1(n)$ will be used at stage n as an estimator of a lower bound of $\delta_{i(k)}$.

Theorem 2.2.3. Suppose that for each pair i and j, $i, j = 1, \ldots, k$, $i \neq j$, $\bigcap_{n=1}^{m} A_{ij}(n, b) \sqsubset \bigcap_{n=1}^{m} A_{ij}(n, a)$ for all $m \geq 1$ as $b < a \leq \delta_0$.

Then

$$P_{\underset{\sim}{\theta}}\left\{L_{i(k)}(n, \delta_1, \delta_{i(k)}) < \frac{k-1}{1-P^*} \text{ for all } \pi_i \in B_{n-1} - \{\pi_{(k)}\} \right.$$

$$\left. \text{for all } n \geq 1\right\}$$

(2.2.14)

$$\leq P_{\underset{\sim}{\theta}}\left\{\pi_{(k)} \in S \text{ and } \delta_{i(k)} \geq D_i^1(n) \text{ for all } \pi_i \in B_{n-1} \right.$$

$$\left. \text{for all } n \geq 1\right\}$$

for all $\underset{\sim}{\theta} \in \Omega$.

Proof: Since $\delta_{i(k)} \leq \delta_0$ for all i, then, under the assumption, we have

$$P_{\underset{\sim}{\theta}}\Big\{L_{i(k)}(n, \delta_1, \delta_{i(k)}) < \frac{k-1}{1-P^*} \text{ for all } \pi_i \in B_{n-1} - \{\pi_{(k)}\}$$

$$\text{for all } n \geq 1\Big\}$$

$$\leq P_{\underset{\sim}{\theta}}\Big\{L_{i(k)}(n, \delta_1, \delta_0) < \frac{k-1}{1-P^*} \text{ and } \delta_{i(k)} \geq D^1_{i(k)}(n)$$

$$\text{for all } \pi_i \in B_{n-1} - \{\pi_{(k)}\} \text{ for all } n \geq 1\Big\}$$

$$\leq P_{\underset{\sim}{\theta}}\Big\{\pi_{(k)} \in S \text{ and } \delta_{i(k)} \geq D^1_{i(k)}(n) \text{ for all } \pi_i \in B_{n-1} - \{\pi_{(k)}\}$$

$$\text{for all } n \geq 1\Big\}$$

$$= P_{\underset{\sim}{\theta}}\Big\{\pi_{(k)} \in S \text{ and } \delta_{i(k)} \geq D^1_{i(k)}(n) \text{ for all } \pi_i \in B_{n-1} \tag{2.2.15}$$

$$\text{for all } n \geq 1\Big\}$$

$$\leq P_{\underset{\sim}{\theta}}\Big\{\pi_{(k)} \in S \text{ and } \delta_{i(k)} \geq \min_{\pi_j \in B_{n-1}} D^1_{ij}(n) \text{ for all } \pi_i \in B_{n-1}$$

$$\text{for all } n \geq 1\Big\}$$

$$= P_{\underset{\sim}{\theta}}\Big\{\pi_{(k)} \in S \text{ and } \delta_{i(k)} \geq D^1_i(n) \text{ for all } \pi_i \in B_{n-1}$$

$$\text{for all } n \geq 1\Big\}.$$

Therefore, the proof of this theorem is completed.

The following corollary follows immediately from Theorem 2.2.2 and Theorem 2.2.3.

Corollary 2.2.4. Suppose that for each pair i and j, i, j = 1, ..., k, $i \neq j$, $\bigcap_{n=1}^{m} A_{ij}(n, b) \sqsubset \bigcap_{n=1}^{m} A_{ij}(n, a)$ for all $m \geq 1$ as $b < a \leq \delta_0$. Then

$$(2.2.16) \quad P_{\underset{\sim}{\theta}}\left\{\pi_{(k)} \in S \text{ and } \delta_{i(k)} \geq D_i^1(n) \text{ for all } \pi_i \in B_{n-1} \text{ for all } n \geq 1\right\} \geq P^*$$

for all $\underset{\sim}{\theta} \in \Omega$.

Theorem 2.2.5. Suppose that for each pair i and j, i, j = 1, ..., k, $i \neq j$, $\bigcap_{n=1}^{m} A_{ij}(n, b) \sqsubset \bigcap_{n=1}^{m} A_{ij}(n, a)$ for all $m \geq 1$ as $b < a \leq \delta_0$. Then,

$$(2.2.17) \quad P_{\underset{\sim}{\theta}}\{CD(\delta_*) | P_1(\delta_*, \delta_1)\} \geq P^* \text{ for all } \underset{\sim}{\theta} \in \Omega,$$

provided that the procedure $P_1(\delta_*, \delta_1)$ terminates with probability one.

Proof: Note that when the procedure $P_1(\delta_*, \delta_1)$ terminates, each selected population must have been labelled as good at some stage. Let $\pi_i \in S$ and π_i has been labelled as good at stage n. Then, $L_{ij}(n, \delta_1, \delta_*) \geq \frac{k - 1}{1 - P^*}$ for all $\pi_j \in B_n - \{\pi_i\}$. By definition of

$D^1_{ij}(n)$ and the assumption (2.2.11), $D^1_{ij}(n) \geq \delta_*$ for all $\pi_j \in B_n - \{\pi_i\}$, and so, $D^1_{ij}(n) \geq \delta_*$ for all $\pi_j \in B_n$. Hence from (2.2.15), we have

$$P_{\underset{\sim}{\theta}}\Big\{L_{i(k)}(n, \delta_1, \delta_{i(k)}) < \frac{k-1}{1-P^*} \text{ for all } \pi_i \in B_{n-1} - \{\pi_{(k)}\} \text{ for all } n \geq 1\Big\}$$

$$\leq P_{\underset{\sim}{\theta}}\Big\{\pi_{(k)} \in S \text{ and } \delta_{i(k)} \geq D^1_{i(k)}(n) \text{ for all } \pi_i \in B_{n-1} \text{ for all } n \geq 1\Big\}$$

$$\leq P_{\underset{\sim}{\theta}}\Big\{\pi_{(k)} \in S \text{ and } \delta_{i(k)} \geq D^1_{i(k)}(n) \text{ for all } \pi_i \in B_n \text{ for all } n \geq 1\Big\}$$

(2.2.18)

$$= P_{\underset{\sim}{\theta}}\Big\{\pi_{(k)} \in S \text{ and } \delta_{i(k)} \geq \min_{\pi_j \in B_n} D^1_{ij}(n) \text{ for all } \pi_i \in B_n \text{ for all } n \geq 1\Big\}$$

$$\leq P_{\underset{\sim}{\theta}}\Big\{\pi_{(k)} \in S \text{ and } \delta_{i(k)} \geq \min_{\pi_j \in B_n} D^1_{ij}(n) \text{ for all } \pi_i \in S \text{ for all } n \geq 1\Big\}$$

$$= P_{\underset{\sim}{\theta}}\Big\{\pi_{(k)} \in S \text{ and } \delta_{i(k)} \geq \max_{n \geq 1} (\min D^1_{ij}(n)) \text{ for all } \pi_i \in S\Big\}$$

$$\leq P_{\underset{\sim}{\theta}}\Big\{\pi_{(k)} \in S \text{ and } \delta_{i(k)} \geq \delta_* \text{ for all } \pi_i \in S\Big\}$$

$$= P_{\underset{\sim}{\theta}}\Big\{CD(\delta_*) | P_1(\delta_*, \delta_1)\Big\}.$$

Then, by Theorem 2.2.2 and the inequality of (2.2.18), we conclude this theorem.

Corollary 2.2.6. Suppose that for each pair i and j, i, j = 1, ...k,

$i \neq j$, $\bigcap_{n=1}^{m} A_{ij}(n, b) \sqsubset \bigcap_{n=1}^{m} A_{ij}(n, a)$ for all $m \geq 1$ as $b < a \leq \delta_0$.

Then $P_{\underset{\sim}{\theta}}\{CS | P_1(\delta_*, \delta_1)\} \geq P^*$ for all $\underset{\sim}{\theta} \in \Omega$, provided that the procedure $P_1(\delta_*, \delta_1)$ terminates with probability one.

Proof: This corollary can be obtained directly from Theorem 2.2.5 and the fact that $CD(\delta_*) \sqsubset CS$.

Remarks 2.2.7.(1). Corollary 2.2.4 implies that one can assert at the prespecified confidence level P* simultaneously that at each stage, none of the eliminated populations is the best population and that the measure of separation from each remaining population, say π_i, to the best population is bounded by D_i^1.

(2) Theorem 2.2.5 implies that at the conclusion of the experiment, one can assert at the prespecified confidence level P* simultaneously that the best population is selected and also only good populations are selected in this selected subset.

(3) The sequential comparison inference is based on the inclusion relation of (2.2.11). On the event $\left\{L_{i(k)}(n, \delta_1, \delta_{i(k)}) < \frac{k - 1}{1 - P^*}\right.$ for all $\pi_i \in B_{n-1} - \{\pi_{(k)}\}$ for all $n \geq 1\Big\}$, the best population will

never be eliminated. Therefore, at each stage, the comparison between each remaining population and the unknown best population is made.

(4) Hsu and Edwards (1983) also obtained some result similar to that of corollary 2.2.4 for the location parameters. However, our approach is quite different from theirs. Their procedure is based on some location invariant statistics while ours is based on the modified likelihood ratio of the probability density functions of some invariant sufficient statistics. Our approach can be applied to more general cases but theirs is valid only for location parameter case.

2.2.4. A Property of the Selection Procedure $P_1(\delta_*, \delta_1)$

Consider two prespecified values $\delta_{*1} < \delta_{*2} < \delta_0$. Let $\{B_n(\delta_{*i})\}$ be the sequence of sets of remaining populations defined in (2.2.4) by the procedure $P_1(\delta_{*i}, \delta_1)$ for $i = 1, 2$. Also, let N_i denote the stopping time and S_i denote the selected subset of the selection procedure $P_1(\delta_{*i}, \delta_1)$, $i = 1, 2$. Then, we have the following theorem.

Theorem 2.2.8. Suppose that for each pair i and j, $i, j = 1, \ldots, k$,

$$i \neq j, \ \bigcap_{n=1}^{m} A_{ij}(n, b) \sqsubset \bigcap_{n=1}^{m} A_{ij}(n, a) \text{ for all } m \geq 1 \text{ as } b < a \leq \delta_0.$$

Then, $N_1 \leq N_2$ and $S_1 \sqsupset S_2$.

Proof: Let $N = \min(N_1, N_2)$. Since both the procedures $P_1(\delta_{*1}, \delta_1)$ and $P_1(\delta_{*2}, \delta_1)$ use the same eliminating rule, thus, for each

$n \leq N$, $B_n(\delta_{*1}) = B_n(\delta_{*2})$. Now, $\delta_{*1} < \delta_{*2} \leq \delta_0$. By the assumption, if $\pi_i \in B_n(\delta_{*2})$ is labelled as a good population at stage n when the procedure $P_1(\delta_{*2}, \delta_1)$ is applied, then π_i must also be labelled as a good population not later than n when the procedure $P_1(\delta_{*1}, \delta_1)$ is applied. This fact implies that $N_1 \leq N_2$. Now, by the definition of B_n, $B_n \supseteq B_m$ as $m \geq n$. Therefore, $B_{N_1}(\delta_{*1}) \supseteq B_{N_2}(\delta_{*2})$. That is, $S_1 \supseteq S_2$.

2.2.5. Sequential Selection Procedure for Selecting a Good Population

Consider a selection problem among k populations. In most applications, we are usually content with selecting a good population. With this goal, what we want is in fact a selection procedure P having the property that

$$(2.2.19) \qquad P_{\underset{\sim}{\theta}}\left\{\pi_i \text{ is selected and } \delta_{i(k)} \geq \delta_*\right\} \geq P^* \text{ for all } \underset{\sim}{\theta} \in \Omega.$$

Also, the procedure should stop as soon as we are confident that a good population has been found. A sequential selection procedure achieving this goal can be obtained from $P_1(\delta_*, \delta_1)$ with a little modification on its stopping strategy. We now describe it as follows.

Let $\{B_n\}$ be the sequence of sets of remaining populations defined in (2.2.4). For each $\pi_i \in B_n$, π_i is labelled as good if $L_{ij}(n, \delta_1, \delta_*) \geq \frac{k-1}{1-P^*}$ for all $\pi_j \in B_n - \{\pi_i\}$. The procedure

terminates at stage n as soon as one of the populations in B_n has been labelled as good. If there is only one population in B_n having been labelled as good, we then select this population as a good population. If more than one population have been labelled as good, we usually select the one with the largest $D_i^1(n)$ value. We denote this modified sequential selection procedure by $MP_1(\delta_*, \delta_1)$.

The following theorem guarantees the desired confidence level of selecting a good population by applying the selection procedure $MP_1(\delta_*, \delta_1)$.

Theorem 2.2.9. Suppose that the condition $\bigcap_{n=1}^{m} A_{ij}(n, b) \sqsubset \bigcap_{n=1}^{m} A_{ij}(n,a)$ is true for all $m \geq 1$ and $b < a \leq \delta_0$. Then,

$$P_{\underset{\sim}{\theta}}\left\{\pi_i \text{ is selected and } \delta_{i(k)} \geq \delta_* | MP_1(\delta_*, \delta_1)\right\} \geq P^*$$

for all $\underset{\sim}{\theta} \in \Omega$.

Proof: Let $A(\underset{\sim}{\theta}) = \left\{L_{i(k)}(n, \delta_1, \delta_{i(k)}) < \frac{k-1}{1-P^*} \text{ for all } \pi_i \in B_{n-1} - \{\pi_{(k)}\} \text{ for all } n \geq 1\right\}$. Then by Theorem 2.2.2, $P_{\underset{\sim}{\theta}}(A(\underset{\sim}{\theta})) \geq P^*$ for all $\underset{\sim}{\theta} \in \Omega$. We note that on the event $A(\underset{\sim}{\theta})$, under the assumption of this theorem, $L_{i(k)}(n, \delta_1, \delta_0) < \frac{k-1}{1-P^*}$ for all $\pi_i \in B_{n-1} - \{\pi_{(k)}\}$ for all $n \geq 1$. Therefore, $\pi_{(k)}$ can never be eliminated in comparison with another population at any stage. Let $B(\underset{\sim}{\theta}) = \left\{\pi_i | \delta_{i(k)} < \delta_*\right\}$. That is, $B(\underset{\sim}{\theta})$ is the set of bad

populations. It suffices to show that on the event $A(\underset{\sim}{\theta})$, any population in $B(\underset{\sim}{\theta})$ cannot be labelled as good when the procedure $MP_1(\delta_*, \delta_1)$ terminates.

Let M be the stopping time of the sequential selection procedure $MP_1(\delta_*, \delta_1)$. On $A(\underset{\sim}{\theta})$, since $\pi_{(k)}$ can never be eliminated prior to the stopping time M, then $\pi_{(k)}' \in B_n$ for all $1 \leq n \leq M$. Moreover, for each $\pi_i \in B(\underset{\sim}{\theta})$, $\delta_{i(k)} < \delta_* \leq \delta_0$ and on $A(\theta)$, $L_{i(k)}(n, \delta_1, \delta_{i(k)}) < \frac{k-1}{1-P^*}$ for all $1 \leq n \leq M$. Then, by the assumption, $L_{i(k)}(n, \delta_1, \delta_*) < \frac{k-1}{1-P^*}$ which means that π_i cannot be labelled as good. Hence, π_i cannot be selected as a good population. Therefore,

$$P_{\underset{\sim}{\theta}}\left\{\pi_i \text{ is selected and } \delta_{i(k)} \geq \delta_* | MP_1(\delta_*, \delta_1)\right\} \geq P^*$$

for all $\underset{\sim}{\theta} \in \Omega$.

2.3. Sequential Subset Selection Procedures for Exponential Family Populations

2.3.1. Introduction

Let $\pi_1, \ldots, \pi_k$ be $k(\geq 2)$ populations in which we may observe the independent random variables $X_1, \ldots, X_k$, respectively. We assume that the distribution of X_i is a member of the exponential family so that X_i has a frequency function $f(x|\theta_i)$ of the form

(2.3.1) $$f(x|\theta_i) = \exp\{Q(\theta_i)P(x) + R(x) - \phi(\theta_i)\}$$

where $f(x|\theta)$ represents the probability density at x for continuous variable or the probability of obtaining the observed value x for discrete variable and θ is the unknown real parameter on which the value of the frequency function $f(x|\theta)$ depends. Here, $P(x)$ and $R(x)$ do not involve the parameter θ, while $Q(\theta)$ and $\phi(\theta)$ do not depend on x. The function $Q(\theta)$ is assumed to be a continuous, strictly increasing function of θ.

Define $\delta_{ij} = \delta(\theta_i, \theta_j) = Q(\theta_i) - Q(\theta_j)$ as a measure of separation between populations π_i and π_j. This particular separation measure had been considered by Bechhofer, Kiefer and Sobel (1968). Let $\overline{\delta} = \max_{1\le i\le k} \min_{j\neq i} \{\delta_{ij}\}$. Population π_i is called the best population if π_i is the unique population such that $\min_{j\neq i} \{\delta_{ij}\} = \overline{\delta}$. If more than one population has this property, one of them is tagged as the best population. Let (k) denote the index of the best population and $\pi_{(k)}$ the best population. For $\overline{\delta} \ge \Delta^*$ for some given Δ^* (> 0) which is known, a sequential procedure for selecting the best population through the indifference zone approach was derived by Hoel and Mazumdar (1968).

In this section, we assume that there is no information about δ_{ij} for all i, j = 1, ..., k. Thus, subset selection approach is more appropriate here. For a prespecified value $\delta^* > 0$, we say that population π_i is good if $Q(\theta_{(k)}) - Q(\theta_i) \le \delta^*$. Let S denote the

selected subset, CS be the event that $\pi_{(k)} \in S$ and $CD(\delta^*)$ be the event that $\pi_{(k)} \in S$ and $\delta_{(k)i} \leq \delta^*$ for all $\pi_i \in S$. We require a sequential subset selection procedure to select a subset S such that $P_{\underset{\sim}{\theta}}\{CD(\delta^*)\} \geq P^*$ for a preassigned probability level $P^* \in (\frac{1}{k}, 1)$ for all $\underset{\sim}{\theta} = (\theta_1, \ldots, \theta_k) \in \Omega$ where $\Omega = \{\underset{\sim}{\theta} = (\theta_1, \ldots, \theta_k) | f(x|\theta_i)$ is well defined for all $i = 1, \ldots, k\}$ is the parameter space.

For this particular separation measure $\delta_{ij} = Q(\theta_i) - Q(\theta_j)$, some appropriate transformation of the random observations taken from a pair of any two populations is considered. With this transformation, the likelihood function of the new statistics can be factored into two parts, one of which, obtained by a conditional argument, and termed the conditional likelihood function, is a function only of the parameter of interest. Based on this conditional likelihood function, a sequential subset selection procedure is derived.

2.3.2. Some Properties Associated with Exponential Family Distributions

Let X_{in} denote the nth observation taken from population π_i. It is assumed that the random observations $X_{i1}, X_{i2}, \ldots,$ are i.i.d., with the frequency function $f(x|\theta_i) = \exp\{Q(\theta_i)P(x) + R(x) - \phi(\theta_i)\}$ for $i = 1, \ldots, k$. For each pair (i, j), $1 \leq i, j \leq k$, $i \neq j$, let

$$(2.3.2.) \quad \begin{cases} W_{ij}(n) = (P(X_{in}) - P(X_{jn}))/2 \quad , \\ \\ Z_{ij}(n) = (P(X_{in}) + P(X_{jn}))/2 \quad . \end{cases}$$

Then, $(W_{ij}(n), Z_{ij}(n))$ have a joint frequency function

$$(2.3.3)\qquad g(w, z|\theta_i, \theta_j) = \exp\Big\{\delta_{ij}w + (Q(\theta_i) + Q(\theta_j))z + a(z, w) - \phi(\theta_i) - \phi(\theta_j)\Big\}C\ I_A(w, z),$$

and given that $Z_{ij}(n) = z$, the conditional frequency function of $W_{ij}(n)$ is

$$(2.3.4)\qquad g(w|z, \theta_i, \theta_j) = \exp\Big\{\delta_{ij}w + a(z, w) - \psi(z, \delta_{ij})\Big\}I_A(w, z),$$

where

$$(2.3.5)\quad \begin{cases} \text{(i)} & a(z, w) = R(z + w) + R(z - w)\ , \\ \text{(ii)} & \psi(z, \delta) = \log \int_{w\varepsilon A_z} \exp\{\delta w + a(z, w)\}dw\ , \\ \text{(iii)} & \text{Set A is the support of the random vector } (W_{ij}(n), Z_{ij}(n)) \text{ and } A_z = \{w|(w, z)\ \varepsilon A\}, \\ \text{(iv)} & C \text{ is a constant such that } g(w, z|\theta_i, \theta_j) \text{ is a frequency function.} \end{cases}$$

Note that for each z, $(w, z)\ \varepsilon A$ iff $(-w, z)\ \varepsilon A$. Thus, $w\ \varepsilon A_z$ iff $-w\ \varepsilon A_z$. Also, $a(z, w) = a(z, -w)$. Hence, by the definition of $\psi(z, \delta)$, $\psi(z, \delta) = \psi(z, -\delta)$.

From (2.3.4), given that $Z_{ij}(n) = z$, the conditional frequency function $g(w|z, \theta_i, \theta_j)$ of $W_{ij}(n)$ depends on (θ_i, θ_j) only through the difference $\delta_{ij} = Q(\theta_i) - Q(\theta_j)$. Hence, this conditional frequency function is also denoted by $g(w|z, \delta_{ij})$. That is,

$$g(w|z, \delta_{ij}) = \exp\left\{\delta_{ij}w + a(z, w) - \psi(z, \delta_{ij})\right\}I_A(w, z). \tag{2.3.6}$$

It is clear that $g(w|z, \delta_{ij})$ has exponential family distribution and hence has monotone likelihood ratio property. Therefore, $E_\delta[a(W_{ij}(n))|Z_{ij}(n)]$ is a nondecreasing function of δ whenever $a(\cdot)$ is any nondecreasing function of w. Also from the usual theory for the exponential family distribution, we know that given $Z_{ij}(n) = z$, the conditional mean and variance of $W_{ij}(n)$, for any δ_0, are

$$\begin{cases} \text{(i)} \quad E_{\delta_0}\left[W_{ij}(n)|Z_{ij}(n) = z\right] = \dfrac{\partial}{\partial\delta}\psi(z, \delta)\Big|_{\delta=\delta_0} = \psi_\delta(z, \delta_0), \\ \\ \text{(ii)} \quad \mathrm{Var}_{\delta_0}(W_{ij}(n)|Z_{ij}(n) = z) = \dfrac{\partial^2}{\partial\delta^2}\psi(z, \delta)\Big|_{\delta=\delta_0} \\ \\ \qquad\qquad = \psi_{\delta\delta}(z, \delta_0) > 0. \end{cases} \tag{2.3.7}$$

The following lemma is easily obtained from the above discussion.

Lemma 2.3.1. For each z belonging to the domain of $Z_{ij}(n)$,

(i) $\psi(z, \delta) = \psi(z, -\delta)$ for all δ.

(ii) $\psi(z, \delta)$ is a convex function of δ.

(iii) $\psi(z, \delta)$ is decreasing in $\delta \in (-\infty, 0]$ and increasing in $\delta \in [0, \infty)$ and $\psi(z, 0) = \min_{\delta} \psi(z, \delta)$.

(iv) Let a_1, a_2, b_1, b_2 be points such that $a_1 \leq a_2 < b_2$, $a_1 < b_1 \leq b_2$, then,

$$\frac{\psi(z, b_1) - \psi(z, a_1)}{b_1 - a_1} \leq \frac{\psi(z, b_2) - \psi(z, a_2)}{b_2 - a_2} .$$

For each z, define

(2.3.8) $$I(b|z) = b\psi_{\delta}(z, b) - [\psi(z, b) - \psi(z, 0)] \quad \text{where} \quad b \in R.$$

By the convexity property of $\psi(z, b)$, $I(b|z) \geq 0$ for all z if $b > 0$, $I(b|z) \leq 0$ for all z if $b \leq 0$. In (2.3.7) and (2.3.8), we replace z with the random variable $Z_{ij}(n)$. Then $\psi_{\delta}(Z_{ij}(n), b)$ and $I(b|Z_{ij}(n))$ are nonnegative (nonpositive) random variables if $b \geq 0$ (≤ 0).

Lemma 2.3.2.

(i) $E_{\underset{\sim}{\theta}}\left[I(\delta_{ij}|Z_{ij}(n))\right] > (=, <)\ 0$ iff $\delta_{ij} > (=, <)\ 0$.

(ii) $E_{\underset{\sim}{\theta}}\left[\psi_\delta(Z_{ij}(n), \delta_2) - \psi_\delta(Z_{ij}(n), \delta_1)\right] > (=, <)\ 0$

iff $\delta_2 > (=, <)\delta_1$.

Proof: We prove part (i) only. The argument for proving part (ii) is similar to that for part (i).

Without loss of generality, we assume that $\theta_i > \theta_j$. Then $\delta_{ij} > 0$ and so $I(\delta_{ij}|Z_{ij}(n)) \geq 0$. Suppose that $E_{\underset{\sim}{\theta}}[I(\delta_{ij}|Z_{ij}(n))] = 0$. This implies that $\delta_{ij}\psi_\delta(z, \delta_{ij}) = \psi(z, \delta_{ij}) - \psi(z, 0)$ a.e. $(Z_{ij}(n))$. By mean-value theorem, for each z, there exists $\zeta \equiv \zeta(z, \delta_{ij}) \in (0, \delta_{ij})$ such that $\delta_{ij}\psi_\delta(z, \zeta) = \psi(z, \delta_{ij}) - \psi(z, 0)$. Hence, $\psi_\delta(z, \zeta) = \psi_\delta(z, \delta_{ij})$ a.e. $(Z_{ij}(n))$. Now, since $\psi_\delta(z, a)$ is an increasing function of a, hence, $\psi_\delta(z, a) = \psi_\delta(z, \delta_{ij})$ for all $a \in [\zeta, \delta_{ij}]$. Thus, $\psi_{\delta\delta}(z, a) = 0$ for all $a \in (\zeta, \delta_{ij})$. By (2.3.5), $\psi_{\delta\delta}(z, a)$ is a continuous function of a. So, $\psi_{\delta\delta}(z, \delta_{ij}) = 0$. This holds a.e. $(Z_{ij}(n))$. That is, $\psi_{\delta\delta}(Z_{ij}(n), \delta_{ij}) = 0$ a.e. Hence $\mathrm{Var}_{\delta_{ij}}(W_{ij}(n)|Z_{ij}(n)) = 0$, which implies that $W_{ij}(n)$ is exactly a function of $Z_{ij}(n)$. By the transformation (2.3.2), it implies that X_{in} is exactly determined by X_{jn}, which is a contradiction. So, $E_{\underset{\sim}{\theta}}[I(\delta_{ij}|Z_{ij}(n))] > 0$ when $\delta_{ij} > 0$.

2.3.3. Sequential Subset Selection Procedure $P_2(H, \delta^*)$

Let $\delta^* > 0$ be the prespecified value used to define the event $CD(\delta^*)$. Let $H(y)$ be a distribution function defined on $[0, \delta^*]$ such that for some $C_0 \in (0, \delta^*)$, the interval $[0, C_0]$ is contained in the support of H. Let $a \geq 0$. For each pair (i, j) and $n \geq 1$, define

$$(2.3.9) \qquad L_{ij}(n, a) = \frac{\int_0^{\delta^*} \prod_{m=1}^{n} g(W_{ij}(m) | Z_{ij}(m), y)\, dH(y)}{\prod_{m=1}^{n} g(W_{ij}(m) | Z_{ij}(m), -a)},$$

where $g(w|z, \delta)$ is the conditional frequency function defined in (2.3.6).

We now define a sequential subset selection procedure $P_2(H, \delta^*)$ as follows.

Let $B_0 = \{\pi_1, \ldots, \pi_k\}$. For each $n \geq 1$, define

$$(2.3.10) \qquad B_n = \left\{\pi_i \in B_{n-1} \middle| L_{ji}(n, 0) < \frac{k - 1}{1 - P^*} \text{ for all } \pi_j \in B_{n-1} - \{\pi_i\}\right\}.$$

That is, B_n is the set of remaining populations up to stage n. At stage n, $\pi_i \in B_n$ is labelled as good if $L_{ij}(n, \delta^*) \geq \frac{k - 1}{1 - P^*}$ for all $\pi_j \in B_n - \{\pi_i\}$. If either $|B_n| = 1$ or all the populations in B_n have been labelled as good, then the procedure terminates and we take $S = B_n$. Otherwise, we go to the next stage. The procedure is continued in this way.

In the next subsection, we will give a sufficient condition under which we prove that the sequential subset selection procedure $P_2(H, \delta^*)$ terminates with probability one.

2.3.4. Termination with Probability One

We first assume that $H(\cdot)$ is a distribution function defined on $[0, \delta^*]$ such that for some $C_0 \in (0, \delta^*)$, the interval $[0, C_0]$ is contained in the support of H. It is also assumed that the conditional variance $\psi_{\delta\delta}(z, b)$ of $W_{ij}(n)$, given that $Z_{ij}(n) = z$, satisfies condition (2.3.11).

$$\psi_{\delta\delta}(z, b) \leq M(b) \text{ a.e. } (Z_{ij}(n)) \text{ for each } b > 0 \tag{2.3.11}$$

where $M(\cdot)$ is bounded on $[a, \infty)$ for all $a > 0$. That is, there exists a finite function $q(\cdot)$ such that $M(b) \leq q(a)$ for all $b \geq a > 0$. It is not hard to verify that many exponential family distributions, including normal, exponential, binomial, satisfy the condition (2.3.11).

Let π_1 and π_2 be two populations belonging to the same class of exponential family distributions with parameters θ_1 and θ_2, respectively. Without loss of generality, we assume that $\theta_1 \geq \theta_2$.

Define stopping times T_1 and T_2 as follows:

$$T_1 = \min\left\{n \mid L_{12}(n, \delta^*) \geq \frac{k - 1}{1 - P^*}\right\}, \tag{2.3.12}$$

$$T_2 = \min\left\{n \mid L_{12}(n, 0) \geq \frac{k - 1}{1 - P^*}\right\} . \tag{2.3.13}$$

Lemma 2.3.3. Let $L_{ij}(n, a)$, $n \geq 1$, be the statistics defined in (2.3.9) where $H(\cdot)$ is a distribution function on $[0, \delta^*]$ such that for some $C_0 \in (0, \delta^*)$, the interval $[0, C_0]$ is contained in the support of H. Suppose that $\psi_{\delta\delta}(z, b) \leq M(b)$ a.e. $(Z_{ij}(n))$ for each $b > 0$ where $M(\cdot)$ is bounded on $[a, \infty)$ for all $a > 0$. Then, $P_{\underset{\sim}{\theta}}\{T_2 < \infty\} = 1$ if $\theta_1 > \theta_2$.

Proof: The proof is analogous to that of Lemma 1 of Pollak and Siegmund (1975). Note that $\delta_{12} > 0$ and

$$L_{12}(n, 0) = \int_0^{\delta^*} \exp\Big\{y \sum_{m=1}^{n} W_{12}(m) - \sum_{m=1}^{n} \Big[\psi(Z_{12}(m), y) - \psi(Z_{12}(m), 0)\Big]\Big\} dH(y).$$

For convenience, in the following, we let $W(m) = W_{12}(m)$, $Z(m) = Z_{12}(m)$, $\lambda(m, a, b) = \psi(Z(m), a) - \psi(Z(m), b)$, $\delta_0 = \delta_{12}$ and $\alpha = \frac{k - 1}{1 - P^*}$.

For each n, define $T_{2n} = \min\{T_2, n\} - 1$, so that $T_{2n} + 1$ is a bounded stopping time.

Case 1. When $\delta_0 \leq C_0$. From the definition of T_2, we have

$$\begin{aligned} &\log \alpha \\ &> \log L_{12}(T_{2n}, 0) \\ &= \delta_0 \sum_{m=1}^{T_{2n}} W(m) - \sum_{m=1}^{T_{2n}} \lambda(m, \delta_0, 0) \end{aligned} \tag{2.3.14}$$

$$+ \log \int_0^{\delta^*} \exp\Big\{(y - \delta_0) \sum_{m=1}^{T_{2n}} W(m) - \sum_{m=1}^{T_{2n}} \lambda(m, y, \delta_0)\Big\} dH(y)$$

$$\geq \delta_0 \sum_{m=1}^{T_{2n}} \Big[W(m) - \psi_\delta(Z(m), \delta_0)\Big]$$

$$- \sum_{m=1}^{T_{2n}} \Big[\delta_0 \psi(Z(m), \delta_0) - \lambda(m, \delta_0, 0)\Big]$$

$$+ \log \int_{|y-\delta_0|<\varepsilon} \exp\Big\{(y - \delta_0) \sum_{m=1}^{T_{2n}} W(m)$$

$$- \sum_{m=1}^{T_{2n}} \lambda(m, y, \delta_0)\Big\} dH(y).$$

By using Taylor's expansion,

$$\lambda(m, y, \delta_0) \equiv \psi(z, y) - \psi(z, \delta_0)$$

(2.3.15)

$$= (y - \delta_0)\psi_\delta(z, \delta_0) + \tfrac{1}{2}(y - \delta_0)^2 \psi_{\delta\delta}(z, \zeta)$$

where $\zeta = \zeta(z, y, \delta_0)$ is between y and δ_0. By the assumption (2.3.11), we can find an ε so small that $\frac{1}{2}|y - \delta_0|\psi_{\delta\delta}(z(m), \zeta) \leq 1$ a.e. for all $y \in (\delta_0 - \varepsilon, \delta_0 + \varepsilon)$. We then obtain

$$\log \int_{|y-\delta_0|<\varepsilon} \exp\Big\{(y - \delta_0) \sum_{m=1}^{T_{2n}} W(m) - \sum_{m=1}^{T_{2n}} \lambda(m, y, \delta_0)\Big\} dH(y)$$

$$\geq \log \int_{|y-\delta_0|<\varepsilon} \exp\Big\{(y - \delta_0) \sum_{m=1}^{T_{2n}} \Big[W(m) - \psi_\delta(Z(m), \delta_0)\Big] - \varepsilon T_{2n}\Big\} dH(y) \tag{2.3.16}$$

$$\geq \log H(\delta_0 - \varepsilon, \delta_0 + \varepsilon) + (d - \delta_0) \sum_{m=1}^{T_{2n}} \Big[W(m) - \psi_\delta(Z(m), \delta_0)\Big] - \varepsilon T_{2n}$$

where $H(\delta_0 - \varepsilon, \delta_0 + \varepsilon) = \int_{|y-\delta_0|<\varepsilon} dH(y)$, $d = \int_{|y-\delta_0|<\varepsilon} y dH(y)/$ $H(\delta_0 - \varepsilon, \delta_0 + \varepsilon)$ and the second inequality is obtained by Jensen's inequality. From (2.3.8), (2.3.14) and (2.3.16), we have

$$\sum_{m=1}^{T_{2n}} \Big[I(\delta_0|Z(m)) - \varepsilon\Big] + d \sum_{m=1}^{T_{2n}} \Big[W(m) - \psi_\delta(Z(m), \delta_0)\Big] \leq \log \alpha - \log H(\delta_0 - \varepsilon, \delta_0 + \varepsilon). \tag{2.3.17}$$

Therefore,

$$\sum_{m=1}^{T_{2n}+1}\left[I(\delta_0|Z(m)) - \varepsilon\right] + d\sum_{m=1}^{T_{2n}+1}\left[W(m) - \psi_\delta(Z(m), \delta_0)\right]$$

$$\leq \log\alpha - \log H(\delta_0 - \varepsilon, \delta_0 + \varepsilon) + I(\delta_0|Z(T_{2n}+1)) - \varepsilon \tag{2.3.18}$$

$$+ d\left[W(T_{2n}+1) - \psi_\delta(Z(T_{2n}+1), \delta_0)\right].$$

Now, consider the expectation $E_{\underset{\sim}{\theta}}$ of both sides of (2.3.18). Since

$$E_{\underset{\sim}{\theta}}\left[|W(m) - \psi_\delta(Z(m), \delta_0)|\right]$$

$$= E_{\underset{\sim}{\theta}}\left[E_{\delta_0}\left[|W(m) - \psi_\delta(Z(m), \delta_0)|\,\middle|\,Z(m)\right]\right]$$

$$\leq \left\{E_{\underset{\sim}{\theta}}\left[E_{\delta_0}\left[|W(m) - \psi_\delta(Z(m), \delta_0)|^2\,\middle|\,Z(m)\right]\right]\right\}^{\frac{1}{2}} \tag{2.3.19}$$

$$= \left\{E_{\underset{\sim}{\theta}}\left[\psi_{\delta\delta}(Z(m), \delta_0)\right]\right\}^{\frac{1}{2}}$$

$$\leq M^{\frac{1}{2}}(\delta_0)$$

and

$$E_{\underset{\sim}{\theta}}\left[W(m) - \psi_\delta(Z(m), \delta_0)\right]$$

$$= E_{\underset{\sim}{\theta}}\left[E_{\delta_0}\left[W(m) - \psi_\delta(Z(m), \delta_0)|Z(m)\right]\right] \tag{2.3.20}$$

$$= 0.$$

Also,

$$0 < E_{\underset{\sim}{\theta}}\left[I(\delta_0|Z(m))\right]$$

$$= E_{\underset{\sim}{\theta}}\left[\delta_0\psi_\delta(Z(m), \delta_0) - \lambda(m, \delta_0, 0)\right]$$

$$\leq E_{\underset{\sim}{\theta}}\left[\delta_0\psi_\delta(Z(m), \delta_0)\right] \tag{2.3.21}$$

$$= \delta_0 E_{\underset{\sim}{\theta}}\left[E_{\delta_0}\left[W(m)|Z(m)\right]\right]$$

$$= \delta_0 E_{\underset{\sim}{\theta}}[W(m)] < \infty$$

Thus, by Wald's lemma, $E_{\underset{\sim}{\theta}}\left[\sum_{m=1}^{T_{2n}+1} [W(m) - \psi_\delta(Z(m), \delta_0)]\right] = 0$ and

$$E_{\underset{\sim}{\theta}}\left[\sum_{m=1}^{T_{2n}+1} [I(\delta_0|Z(m)) - \varepsilon]\right] = \left[E_{\underset{\sim}{\theta}}(T_{2n} + 1)\right] \times \left[E_{\underset{\sim}{\theta}}[I(\delta_0|Z(1))] - \varepsilon\right].$$

Also, by Schwarz inequality,

$$E_{\underset{\sim}{\theta}}\left[W(T_{2n} + 1) - \psi_\delta(Z(T_{2n} + 1), \delta_0)\right]$$

$$\leq E_{\underset{\sim}{\theta}}\left[\left|W(T_{2n} + 1) - \psi_\delta(Z(T_{2n} + 1), \delta_0)\right|\right] \tag{2.3.22}$$

$$\leq \left\{E_{\underset{\sim}{\theta}}\left[(W(1) - \psi_\delta(Z(1), \delta_0))^2\right] \times \left[E_{\underset{\sim}{\theta}}(T_{2n} + 1)\right]\right\}^{\frac{1}{2}}.$$

Here,

$$E_{\underset{\sim}{\theta}}\left[(W(1) - \psi_\delta(Z(1), \delta_0))^2\right]$$

$$= E_{\underset{\sim}{\theta}}\left[E_{\delta_0}[(W(1) - \psi_\delta(Z(1), \delta_0))^2|Z(1)]\right]$$

$$= E_{\underset{\sim}{\theta}}\left[\psi_{\delta\delta}(Z(1), \delta_0)\right] \tag{2.3.23}$$

$$\leq M(\delta_0).$$

Hence, $\left|E_{\underset{\sim}{\theta}}\left[W(T_{2n}+1) - \psi_\delta(Z(T_{2n}+1), \delta_0)\right]\right| \leq M^{\frac{1}{2}}(\delta_0) \times \left[E_{\underset{\sim}{\theta}}(T_{2n}+1)\right]^{\frac{1}{2}}$.

Similarly,

$$E_{\underset{\sim}{\theta}}\left[I(\delta_0|Z(T_{2n}+1))\right]$$

$$\leq \delta_0 E_{\underset{\sim}{\theta}}\left[\psi_\delta(Z(T_{2n}+1), \delta_0)\right]$$

(2.3.24)

$$= \delta_0 E_{\underset{\sim}{\theta}}\left[W(T_{2n}+1)\right]$$

$$\leq \delta_0\left\{E_{\underset{\sim}{\theta}}\left[W^2(1)\right] \times E_{\underset{\sim}{\theta}}\left[T_{2n}+1\right]\right\}^{\frac{1}{2}}.$$

Therefore,

$$\left\{E_{\underset{\sim}{\theta}}\left[I(\delta_0|Z(1))\right] - \varepsilon\right\} \times \left\{E_{\underset{\sim}{\theta}}\left[T_{2n}+1\right]\right\}$$

$$\leq \left\{\delta_0\left[E_{\underset{\sim}{\theta}} W^2(1)\right]^{\frac{1}{2}} + dM^{\frac{1}{2}}(\delta_0)\right\} \times \left\{E_{\underset{\sim}{\theta}} T_{2n}+1\right\}^{\frac{1}{2}}$$

$$- \log H(\delta_0 - \varepsilon, \delta_0 + \varepsilon) + \log \alpha,$$

and so,

$$\left\{E_{\underset{\sim}{\theta}}\left[I(\delta_0|Z(1))\right] - \varepsilon\right\}\left\{E_{\underset{\sim}{\theta}} T_{2n}+1\right\}^{\frac{1}{2}}$$

(2.3.25)

$$\leq \left\{\delta_0\left[E_{\underset{\sim}{\theta}} W^2(1)\right]^{\frac{1}{2}} + dM^{\frac{1}{2}}(\delta_0)\right\} - \log H(\delta_0 - \varepsilon, \delta_0 + \varepsilon) + \log \alpha.$$

Now, by Lemma 2.3.2, ε can be chosen so small that $E_{\underset{\sim}{\theta}}[I(\delta_0)|Z(1)] - \varepsilon > 0$. While the right-hand side of (2.3.25) is independent of n,

so, as $n \to \infty$, the left-hand side of (2.3.25) is still bounded. Hence, $E_{\underset{\sim}{\theta}}(T_2) < \infty$ and so, $P_{\underset{\sim}{\theta}}\{T_2 < \infty\} = 1$.

Case 2. When $\delta_0 > C_0$. From (2.3.14), we also have

$$\sum_{m=1}^{T_{2n}+1} I(C_0|Z(m)) + d \sum_{m=1}^{T_{2n}+1} \left[W(m) - \psi_\delta(Z(m),\delta_0)\right]$$

$$+ \sum_{m=1}^{T_{2n}+1} \left\{d\left[\psi_\delta(Z(m), \delta_0) - \psi_\delta(Z(m), C_0)\right] - \varepsilon\right\}$$

$$(2.3.26) \leq \log \alpha - \log H(C_0 - \varepsilon, C_0 + \varepsilon)$$

$$+ d\left[W(T_{2n}+1) - \psi_\delta(Z(T_{2n}+1), \delta_0)\right]$$

$$+ d\left[\psi_\delta(Z(T_{2n}+1), \delta_0) - \psi(Z(T_{2n}+1), C_0)\right].$$

Following an argument analogous to the above, we have $E_{\underset{\sim}{\theta}}[T_2] < \infty$, and so $P_{\underset{\sim}{\theta}}\{T_2 < \infty\} = 1$.

Lemma 2.3.4. Let $L_{ij}(n, a)$, $n \geq 1$, be the statistics defined in (2.3.9) where $H(\cdot)$ is a distribution function on $[0, \delta^*]$ such that for some $C_0 \in (0, \delta^*)$, the interval $[0, C_0]$ is contained in the support of H. Then, $P_{\underset{\sim}{\theta}}\{T_1 < \infty\} = 1$ if $\theta_1 = \theta_2$.

Proof: Note that here $\delta_{12} = 0$ since $\theta_1 = \theta_2$. From (2.3.6),

$$L_{12}(n, \delta^*) = \int_0^{\delta^*} \exp\left\{(y + \delta^*) \sum_{m=1}^{n} W(m) + \sum_{m=1}^{n} \lambda(m, \delta^*, y)\right\} dH(y).$$

By Jensen's inequality, we have

$$\frac{1}{n} \log L_{12}(n, \delta^*)$$

$$(2.3.27) \quad \geq \frac{1}{n} \int_0^{\delta^*} \left\{(y + \delta^*) \sum_{m=1}^{n} W(m) + \sum_{m=1}^{n} \lambda(m, \delta^*, y)\right\} dH(y)$$

$$= \frac{1}{n} \sum_{m=1}^{n} V_m$$

where $V_m = \int_0^{\delta^*} \{(y + \delta^*)W(m) + \lambda(m, \delta^*, y)\} dH(y)$, $m = 1, \ldots, n$, are i.i.d. By Lemma 2.3.1 and the fact that $\delta_{12} = 0$, then,

$$E_{\underset{\sim}{\theta}}[V_m]$$

$$= \int_0^{\delta^*} E_{\underset{\sim}{\theta}}\{(y + \delta^*)W(m) + \lambda(m, \delta^*, y)\} dH(y)$$

$$= \int_0^{\delta^*} E_{\underset{\sim}{\theta}}\{\lambda(m, \delta^*, y)\} dH(y)$$

$$(2.3.28) \quad = \int_0^{\delta^*} E_{\underset{\sim}{\theta}}[\psi(Z(m), \delta^*) - \psi(Z(m), y)] dH(y)$$

$$\geq \int_0^{C_0} E_{\underset{\sim}{\theta}}[\psi(Z(m), \delta^*) - \psi(Z(m), y)] dH(y)$$

$$\geq \int_0^{C_0} E_{\underset{\sim}{\theta}}[\psi(Z(m), \delta^*) - \psi(Z(m), C_0)]dH(y)$$

$$= E_{\underset{\sim}{\theta}}[\psi(Z(m), \delta^*) - \psi(Z(m), C_0)]H(C_0)$$

$$> 0 .$$

By strong law of large numbers and by (2.3.27) and (2.3.28), we obtain

$$\liminf_{n\to\infty} \frac{1}{n} \log L_{12}(n, \delta^*) \geq E_{\underset{\sim}{\theta}}[\psi(Z(1), \delta^*) - \psi(Z(1), C_0)]H(C_0) \text{ a.e.,}$$

while $\frac{1}{n} \log \frac{k - 1}{1 - P^*} \to 0$ as $n \to \infty$. Therefore, $P_{\underset{\sim}{\theta}}\{T_1 < \infty\} = 1$ when $\theta_1 = \theta_2$.

Theorem 2.3.5. Let $P_2(H, \delta^*)$ be the sequential subset selection procedure defined in Subsection 2.3.3 where $H(\cdot)$ is a distribution function defined on $[0, \delta^*]$ such that for some $C_0 \in (0, \delta^*)$, the interval $[0, C_0]$ is contained in the support of H. Suppose that $\psi_{\delta\delta}(z, b) \leq M(b)$ a.e. $(Z_{ij}(n))$ for each $b > 0$ where $M(\cdot)$ is bounded on $[a, \infty)$ for all $a > 0$. Then, the procedure $P_2(H, \delta^*)$ terminates with probability one.

Proof: It suffices to show that, by applying the procedure $P_2(H, \delta^*)$ for any two populations, say π_1 and π_2, with probability one, the event that either one of them will be eliminated (in comparison with the other) or both of them are labelled as good, occurs. This follows from Lemma 2.3.3 and Lemma 2.3.4 immediately. Therefore, the proof of Theorem 2.3.5 is completed.

2.3.5. Probability of a Correct Selection

For each $n \geq 1$, let $\mathcal{F}_{ij}(n)$ denote the σ-field generated by $(W_{ij}(m), Z_{ij}(m), m = 1, 2, \ldots, n)$. Then, we have

Lemma 2.3.6.

(i) $E_{\underset{\sim}{\theta}}\left[L_{ij}(n, 0)\middle|\mathcal{F}_{ij}(n-1)\right] \leq L_{ij}(n-1, 0)$ for all $n \geq 1$ if $\theta_i < \theta_j$ where $L_{ij}(n, 0) \equiv 1$ if $n = 0$.

(ii) $E_{\underset{\sim}{\theta}}\left[L_{i(k)}(n, \delta_{(k)i})\middle|\mathcal{F}_{i(k)}(n-1)\right] = L_{i(k)}(n-1, \delta_{(k)i})$ for all $n \geq 1$ for all $\underset{\sim}{\theta} \in \Omega$.

Proof: Note that $(W_{ij}(n), Z_{ij}(n))$ is independent of $\mathcal{F}_{ij}(n-1)$ for all $n \geq 1$. Thus,

$$E_{\underset{\sim}{\theta}}\left[L_{ij}(n, 0)\middle|\mathcal{F}_{ij}(n-1)\right]$$

$$(2.3.29) = E_{\underset{\sim}{\theta}}\left[\int_0^{\delta^*} \exp\left\{y \sum_{m=1}^{n} W_{ij}(m) - \sum_{m=1}^{n}\left[\psi(Z_{ij}(m), y) - \psi(Z_{ij}(m), 0)\right]\right\} dH(y)\middle|\mathcal{F}_{ij}(n-1)\right]$$

$$= \int_0^{\delta^*} \exp\left\{y \sum_{m=1}^{n-1} W_{ij}(m) - \sum_{m=1}^{n-1}\left[\psi(Z_{ij}(m), y) - \psi(Z_{ij}(m), 0)\right]\right\} A(\underset{\sim}{\theta}, n)\, dH(y)$$

where $A(\underset{\sim}{\theta}, n) = E_{\underset{\sim}{\theta}}\left[\exp\left\{y\ W_{ij}(n) - \psi(Z_{ij}(n), y) + \psi(Z_{ij}(n), 0)\right\}\right]$ and the second equality is obtained from Fubini's theorem and by the property that $(W_{ij}(n), Z_{ij}(n))$ is independent of $\mathcal{F}_{ij}(n-1)$. Also,

$$
\begin{aligned}
& A(\underset{\sim}{\theta}, n) \\
& = E_{\underset{\sim}{\theta}}\left[E_{\delta_{ij}}\left[\exp\left\{y\ W_{ij}(n) - \psi(Z_{ij}(n), y) + \psi(Z_{ij}(n), 0)\right\} \middle| Z_{ij}(n)\right]\right] \\
& = E_{\underset{\sim}{\theta}}\left[\exp\left\{\psi(Z_{ij}(n), y + \delta_{ij}) - \psi(Z_{ij}(n), \delta_{ij}) - \psi(Z_{ij}(n), y)\right.\right. \\
& \qquad \left.\left. + \psi(Z_{ij}(n), 0)\right\}\right] \\
& \leq 1.
\end{aligned}
\tag{2.3.30}
$$

The last inequality follows from Lemma 2.3.1 and the fact that $\delta_{ij} \leq 0$. Therefore,

$$
\begin{aligned}
& E_{\underset{\sim}{\theta}}\left[L_{ij}(n, 0) | \mathcal{F}_{ij}(n-1)\right] \\
& \leq \int_0^{\delta^*} \exp\left\{y \sum_{m=1}^{n-1} W_{ij}(m) - \sum_{m=1}^{n-1}\left[\psi(Z_{ij}(m), y)\right.\right. \\
& \qquad \left.\left. - \psi(Z_{ij}(m), 0)\right]\right\} dH(y) \\
& = L_{ij}(n-1, 0).
\end{aligned}
\tag{2.3.31}
$$

Hence, the proof of (i) of this theorem is completed.

Similarly, we can prove that

$$E_{\underset{\sim}{\theta}}\left[L_{i(k)}(n, \delta_{(k)i}) \middle| \mathcal{F}_{i(k)}(n-1)\right] = L_{i(k)}(n, \delta_{(k)i}) \text{ for all } n \geq 1$$

for all $\underset{\sim}{\theta} \in \Omega$.

Theorem 2.3.7. Let $P_2(H, \delta^*)$ be the sequential subset selection procedure defined above where $H(\cdot)$ is a distribution defined on $[0, \delta^*]$ such that for some $C_0 \in (0, \delta^*)$, the interval $[0, C_0]$ is contained in the support of H. Suppose the condition that $\psi_{\delta\delta}(z, \delta) \leq M(b)$ a.e. $(Z_{ij}(n))$ for each $b > 0$ where $M(\cdot)$ is bounded on $[a, \infty)$ for all $a > 0$. Then,

(i) $P_{\underset{\sim}{\theta}}\left\{CS|P_2(H, \delta^*)\right\} \geq P^*$ for all $\underset{\sim}{\theta} \in \Omega$.

(ii) $P_{\underset{\sim}{\theta}}\left\{L_{i(k)}(n, \delta_{(k)i}) < \frac{k-1}{1-P^*} \text{ for all } \pi_i \in B_{n-1} - \{\pi_{(k)}\}\right.$

$\left.\text{for all } n \geq 1\right\} \geq P^*$ for all $\underset{\sim}{\theta} \in \Omega$.

Proof: (i) Let ICS denote the event that the best population is not selected. Then, from Theorem 2.3.5,

$$P_{\underset{\sim}{\theta}}\{CS|P_2(H, \delta^*)\} = 1 - P_{\underset{\sim}{\theta}}\{ICS|P_2(H, \delta^*)\}.$$

Here,

$$P_{\underset{\sim}{\theta}}\{ICS|P_2(H, \delta^*)\}$$

$$= P_{\underset{\sim}{\theta}}\left\{L_{i(k)}(n, 0) > \frac{k-1}{1-P^*} \text{ for some } n \geq 1 \text{ for some } i \neq (k)\right\}$$

(2.3.32)

$$\leq \sum_{\substack{i=1 \\ i\neq(k)}}^{k} P_{\underset{\sim}{\theta}}\left\{L_{i(k)}(n, 0) > \frac{k-1}{1-P^*} \text{ for some } n \geq 1\right\}$$

$$\leq \sum_{\substack{i=1 \\ i\neq(k)}}^{k} \frac{1-P^*}{k-1}$$

$$= 1 - P^*$$

where the last inequality is obtained from the fact that $\left\{L_{i(k)}(n, 0), P_{\underset{\sim}{\theta}}, \mathcal{F}_{i(k)}(n), n \geq 1\right\}$ forms a nonnegative supermartingale which can be obtained directly from Lemma 2.3.6 (i).

Therefore, $P_{\underset{\sim}{\theta}}\{CS|P_2(H, \delta^*)\} \geq P^*$.

For the proof of part (ii) of this theorem, we see that, from Lemma 2.3.6 (ii), $\left\{L_{i(k)}(n, \delta_{(k)i}), P_{\underset{\sim}{\theta}}, \mathcal{F}_{i(k)}(n), n \geq 1\right\}$ forms a nonnegative martingale. Then following an argument similar to that of Theorem 2.2.2, we can prove that

$$P_{\underset{\sim}{\theta}}\left\{L_{i(k)}(n, \delta_{(k)i}) < \frac{k-1}{1-P^*} \text{ for all } \pi_i \in B_{n-1} - \{\pi_{(k)}\}\right.$$

$$\left.\text{for all } n \geq 1\right\} \geq P^* \text{ for all } \underset{\sim}{\theta} \in \Omega.$$

For each pair (i, j), $1 \leq i, j \leq k$, $i \neq j$, and each $a \geq 0$, let $A_{ij}(n, a)$ denote the event that $L_{ij}(n, a) < \frac{k - 1}{1 - P^*}$. That is, $A_{ij}(n, a) = \left\{L_{ij}(n, a) < \frac{k - 1}{1 - P^*}\right\}$. The following result is very helpful for obtaining some sequential estimate of an upper bound on $\delta_{(k)i}$, $i = 1, 2, \ldots, k$.

Theorem 2.3.8. Let $L_{ij}(n, a)$, $n \geq 1$, be the statistics defined in (2.3.9). Then,

$$A_{ij}(n, b) \sqsubset A_{ij}(n, a) \text{ for all } n \geq 1 \text{ if } b > a \geq 0. \tag{2.3.33}$$

Proof: Suppose that the statement of (2.3.33) is not true. Then, there exists some n such that

$$L_{ij}(n, b) < \frac{k - 1}{1 - P^*} \leq L_{ij}(n, a). \tag{2.3.34}$$

By (2.3.9), we have

$$L_{ij}(n, b) = L_{ij}(n, a) \exp\left\{(b - a) \sum_{m=1}^{n} W_{ij}(m) + \sum_{m=1}^{n} \left[\psi(Z_{ij}(m), b) - \psi(Z_{ij}(m), a)\right]\right\}. \tag{2.3.35}$$

(2.3.34) and (2.3.35) together imply that

$$(2.3.36)\qquad \sum_{m=1}^{n} W_{ij}(m) < \frac{-1}{b - a} \sum_{m=1}^{n} \left[\psi(Z_{ij}(m), b) - \psi(Z_{ij}(m), a)\right].$$

Then, by (2.3.34) and (2.3.36)

$$1 < \frac{k - 1}{1 - P^*}$$

$$\leq L_{ij}(n, a)$$

$$(2.3.37)\qquad = \int_0^{\delta^*} \exp\Big\{(y + a) \sum_{m=1}^{n} W_{ij}(m) + \sum_{m=1}^{n} \left[\psi(Z_{ij}(m), a) - \psi(Z_{ij}(m), y)\right]\Big\}\, dH(y)$$

$$\leq \int_0^{\delta^*} \exp\Big\{-(y + a) \sum_{m=1}^{n} \left[\frac{\psi(Z_{ij}(m), y) - \psi(Z_{ij}(m), a)}{y - (-a)} - \frac{\psi(Z_{ij}(m), -a) - \psi(Z_{ij}(m), -b)}{(-a) - (-b)}\right]\Big\} dH(y)$$

By Lemma 2.3.1 and the fact that $-b < -a \leq y$ for each $y \in [0, \delta^*]$, we have, for each $m = 1, 2, \ldots, n$,

$$\frac{\psi(Z_{ij}(m), y) - \psi(Z_{ij}(m), -a)}{y + a} - \frac{\psi(Z_{ij}(m), -a) - \psi(Z_{ij}(m), -b)}{b - a} \geq 0.$$

With this fact and from (2.3.37), we obtain $1 < \frac{k - 1}{1 - P^*} \leq L_{ij}(n, a) < 1$, which is a contradiction. Therefore, $A_{ij}(n, b) \sqsubset A_{ij}(n, a)$ for all $n \geq 1$ when $b > a \geq 0$. This completes the proof of Theorem 2.3.8.

From Theorem 2.3.8, it immediately follows that $\bigcap_{n=1}^{m} A_{ij}(n, b) \sqsubset \bigcap_{n=1}^{m} A_{ij}(n, a)$ for all $m \geq 1$ whenever $b > a \geq 0$.

For each $n \geq 1$, π_i, $\pi_j \in B_{n-1}$, as $i \neq j$, define

$$(2.3.38) \quad D_{ij}^2(n) = \begin{cases} \sup\left\{a \geq 0 \middle| L_{ij}(n, a) < \frac{k - 1}{1 - P^*}\right\} & \text{if } \{\ \} \neq \phi, \\ 0 & \text{if } \{\ \} = \phi. \end{cases}$$

Also, let $D_{ii}^2(n) = 0$. From Theorem 2.3.8, if $D_{ij}^2(n) > 0$, then $L_{ij}(n, a) < \frac{k - 1}{1 - P^*}$ for all $a \in [0, D_{ij}^2(n))$ and $L_{ij}(n, b) \geq \frac{k - 1}{1 - P^*}$ for all $b > D_{ij}^2(n)$.

For each $n \geq 2$, if $\pi_i \in B_{n-1}$, define

$$(2.3.39) \quad D_i^2(n) = \min_{1 \leq m \leq n} \left(\max_{\pi_j \in B_{m-1}} D_{ij}^2(m) \right).$$

If $\pi_i \notin B_{n-1}$, let $n_i = \max\{m | \pi_i \in B_{m-1}\}$ and define $D_i^2(n) = D_i^2(n_i)$.

By the definition of $D_i^2(n)$, for each $i = 1, 2, \ldots, k$, $\{D_i^2(n)\}$ is a decreasing sequence and bounded below by zero. The value $D_i^2(n)$ will be used at stage n as an estimator of an upper bound of $\delta_{(k)i}$.

Theorem 2.3.9. With $L_{ij}(n, a)$, $n \geq 1$, defined as before, we have

$$P_{\underset{\sim}{\theta}}\Big\{L_{i(k)}(n, \delta_{(k)i}) < \frac{k-1}{1-P^*} \text{ for all } \pi_i \in B_{n-1} - \{\pi_{(k)}\} \text{ for all } n \geq 1\Big\}$$

$$\leq P_{\underset{\sim}{\theta}}\Big\{\pi_{(k)} \in S \text{ and } \delta_{(k)i} \leq D_i^2(n) \text{ for all } \pi_i \in B_{n-1} \text{ for all } n \geq 1\Big\}. \tag{2.3.40}$$

Proof: By Theorem 2.3.8, we have

$$P_{\underset{\sim}{\theta}}\Big\{L_{i(k)}(n, \delta_{(k)i}) < \frac{k-1}{1-P^*} \text{ for all } \pi_i \in B_{n-1} - \{\pi_{(k)}\} \text{ for all } n \geq 1\Big\}$$

$$\leq P_{\underset{\sim}{\theta}}\Big\{L_{i(k)}(n, 0) < \frac{k-1}{1-P^*} \text{ and } \delta_{(k)i} \leq D_{i(k)}^2(n) \text{ for all } \pi_i \in B_{n-1} - \{\pi_{(k)}\} \text{ for all } n \geq 1\Big\}$$

$$\leq P_{\underset{\sim}{\theta}}\Big\{\pi_{(k)} \in S \text{ and } \delta_{(k)i} \leq \max_{\pi_j \in B_{n-1}} D_{ij}^2(n) \text{ for all } \pi_i \in B_{n-1} \text{ for all } n \geq 1\Big\}$$

$$= P_{\underset{\sim}{\theta}}\Big\{\pi_{(k)} \in S \text{ and } \delta_{(k)i} \leq D_i^2(n) \text{ for all } \pi_i \in B_{n-1} \text{ for all } n \geq 1\Big\}. \tag{2.3.41}$$

This completes the proof of Theorem 2.3.9.

An immediate consequence of Theorem 2.3.7 and Theorem 2.3.9 is that

$$P_{\underset{\sim}{\theta}}\Big\{\pi_{(k)} \in S \text{ and } \delta_{(k)i} \le D_i^2(n) \text{ for all } \pi_i \in B_{n-1} \text{ for all } n \ge 1\Big\} \ge P^* \tag{2.3.42}$$

for all $\underset{\sim}{\theta} \in \Omega$ whenever the procedure $P_2(H, \delta^*)$ terminates with probability one. This result offers a sequential comparison inference, with confidence level P^*, as follows: simultaneously, at each stage n, the best population is not eliminated and the separation between each remaining population, say π_i, and the unknown best population is not larger than the value $D_i^2(n)$ for all $n \ge 1$. Another consequence of Theorem 2.3.7 and Theorem 2.3.9 is that when the procedure $P_2(H, \delta^*)$ terminates, the event $CD(\delta^*)$ is guaranteed with probability at least P^*. We state this result without proof as a corollary as follows.

Corollary 2.3.10. Suppose that $\psi_{\delta\delta}(z, b) \le M(b)$ a.e. $(Z_{ij}(n))$ for each $b > 0$ where $M(\cdot)$ is bounded on $[a, \infty)$ for all $a > 0$. Then, the sequential subset selection procedure $P_2(H, \delta^*)$ satisfies $P_{\underset{\sim}{\theta}}\{CD(\delta^*) | P_2(H, \delta^*)\} \ge P^*$ for all $\underset{\sim}{\theta} \in \Omega$.

2.4. Applications

In this section, we give some examples of the application of the sequential selection procedures derived in Section 2.2.

2.4.1. Selecting the Population with the Largest Normal Mean

Let $\pi_1, \ldots, \pi_k$ be k populations and let X_{in} denote the nth observation taken from population π_i. Assume that X_{in} has normal distribution with unknown mean θ_i and a common known variance $\sigma^2 = 1$, for $i = 1, 2, \ldots, k$. Define the measure of separation between π_i and π_j as $\delta_{ij} = \theta_i - \theta_j$. Then, $\delta_0 = 0$ and $\overline{\delta} = \theta_{[k]} - \theta_{[k-1]}$, and the population with the largest mean is considered as the best population. For a given $\delta^* > 0$, π_i is said to be good if $\theta_{[k]} - \theta_i \leq \delta^*$ and bad if $\theta_{[k]} - \theta_i > \delta^*$. For a preassigned value $P^* \in (\frac{1}{k}, 1)$, we wish to derive a sequential subset selection procedure satisfying $P_{\underset{\sim}{\theta}}\left\{\pi_{(k)} \in S \text{ and } \theta_{[k]} - \theta_i \leq \delta^* \text{ for each } \pi_i \in S\right\} \geq P^*$ for all $\underset{\sim}{\theta} \in \Omega$.

For each $n \geq 1$, define $T_{ij}(n) = S_{in} - S_{jn}$ where $S_{in} = \sum_{m=1}^{n} X_{im}$. Let $\delta_* = -\delta^*$ and let $0 < \delta_1 < \delta^*$. Then,

$$(2.4.1) \qquad \log L_{ij}(n, \delta_1, 0) = \frac{\delta_1}{2}(S_{in} - S_{jn}) - \frac{n\delta_1^2}{4},$$

and

$$(2.4.2) \qquad \log L_{ij}(n, \delta_1, \delta_*) = \frac{\delta_1 + \delta^*}{2}(S_{in} - S_{jn}) + \frac{n(\delta^{*2} - \delta_1^2)}{4}.$$

In order to apply the procedure $P_1(\delta_*, \delta_1)$ to this selection problem, we need to assert that this procedure terminates with probability one.

Lemma 2.4.1. For the problem of selecting the population with the largest mean among k normal populations, the sequential subset selection procedure $P_1(\delta_*, \delta_1)$ terminates with probability one if $0 < \delta_1 < \frac{\delta^*}{2}$.

Proof: It suffices to show that for any two populations, say π_1 and π_2, with probability one, the event E, that either one of them will be eliminated (in comparison with the other) or both of them are labelled as good, occurs. Without loss of generality, we assume that $\theta_1 \geq \theta_2$.

First consider the case that $\theta_1 - \theta_2 > \frac{\delta_1}{2}$. Define

$$(2.4.3) \qquad T_1 = \min\left\{n \middle| L_{12}(n, \delta_1, 0) \geq \frac{k-1}{1-P^*}\right\}.$$

By the strong law of large numbers,

$$\frac{1}{n} \log L_{12}(n, \delta_1, 0) \to \frac{\delta_1}{2}\left(\theta_1 - \theta_2 - \frac{\delta_1}{2}\right) > 0 \text{ a.e. as } n \to \infty.$$

while

$$\frac{1}{n} \log \frac{k-1}{1-P^*} \to 0 \text{ as } n \to \infty. \quad \text{Hence, } P_{\underset{\sim}{\theta}}\{T_1 < \infty\} = 1.$$

Next, consider the case that $0 \leq \theta_1 - \theta_2 \leq \frac{\delta_1}{2}$. Define

$$(2.4.4) \qquad T_{ij} = \min\left\{n \middle| L_{ij}(n, \delta_1, \delta_*) \geq \frac{k-1}{1-P^*}\right\}$$

for $i, j = 1, 2, i \neq j$, and

$$(2.4.5) \qquad T_2 = \max(T_{12}, T_{21}).$$

By the strong law of large numbers again,

$$\frac{1}{n} \log L_{12}(n, \delta_1, \delta_*) \to \left(\theta_1 - \theta_2 + \frac{\delta^* - \delta_1}{2}\right)(\delta_1 + \delta^*)/2 > 0$$

a.e. as $n \to \infty$

and

$$\frac{1}{n} \log L_{21}(n, \delta_1, \delta_*) \to \left(\theta_2 - \theta_1 + \frac{\delta^* - \delta_1}{2}\right)(\delta_1 + \delta^*)/2 > 0$$

a.e. as $n \to \infty$.

Hence, $P_{\underset{\sim}{\theta}}\{T_{ij} < \infty\}$ for $i, j = 1, 2, i \neq j$ and so, $P_{\underset{\sim}{\theta}}\{T_2 < \infty\} = 1$.

From the above discussion, we conclude that $P_{\underset{\sim}{\theta}}\{E\} = 1$, and so, the sequential subset selection procedure $P_1(\delta_*, \delta_1)$ terminates with probability one.

In order to guarantee the P*-condition for the event $CD(\delta_*)$, from Theorem 2.2.5, it suffices to verify the condition (2.2.11). This can be easily verified.

In the following, two simulation studies have been carried out to describe the performance of the procedure $P_1(\delta_{*1}, \delta_1)$. We take $k = 4$, $P^* = 0.95$, $\delta^* = 1.5$, $\delta_1 = 0.5$, and $(\theta_1, \theta_2, \theta_3, \theta_4) = (0., -1.0, -1.5, -2.5)$ and $(\theta_1, \theta_2, \theta_3, \theta_4) = (0., -0.1, -1.0, -2.5)$,

respectively. Using random normal deviates generated according to $(\theta_1, \theta_2, \theta_3, \theta_4)$, two examples of sequential comparison inference are given in Table II.1 and Table II.2, respectively. In Table II.1, the final selected subset $S = \{\pi_1\}$ and in Table II.2, $S = \{\pi_1, \pi_2\}$.

If we are content with selecting a good population, then the sequential selection procedure $MP_1(\delta_*, \delta_1)$ can be applied for this goal. In Table II.1, the procedure $MP_1(\delta_*, \delta_1)$ terminates at stage 4 and selects population π_1 as a good population. In Table II.2, the procedure $MP_1(\delta_*, \delta_1)$ terminates at stage 6 and selects π_1 as a good population.

2.4.2. Selecting the Population with the Smallest Normal Variance

We assume that the random observation X_{in} from population π_i has normal distribution with unknown mean θ_i and unknown variance σ_i^2, for $i = 1, 2, \ldots, k$. Let $\sigma^2_{[1]} \leq \cdots \leq \sigma^2_{[k]}$ be the ordered values of $\sigma_1^2, \ldots, \sigma_k^2$. Let $\underset{\sim}{\sigma}^2 = (\sigma_1^2, \ldots, \sigma_k^2)$. Define the measure of separation between populations π_i and π_j as $\delta_{ij} = \frac{\sigma_j^2}{\sigma_i^2}$. Then, $\overline{\delta} = \frac{\sigma^2_{[2]}}{\sigma^2_{[1]}}$ and $\delta_0 = 1$. The population with the smallest variance is considered as the best population.

In applying the selection procedure $P_1(\delta_*, \delta_1)$ to this selection problem, we let $\overline{X}_{in} = \frac{1}{n} \sum_{m=1}^{n} X_{im}$, $S^2_{in} = \frac{1}{n} \sum_{m=1}^{n} (X_{im} - \overline{X}_{in})^2$ and define

$T_{ij}(n) = \frac{S_{jn}^2}{S_{in}^2}$ for $n \geq 2$. This choice of $T_{ij}(n)$ seems to be a natural one if we have the ratio of variances as the measure of separation. From the result of Hall, Wijsman and Ghosh (1965), the sequence $T_{ij}(2), T_{ij}(3), \ldots,$ can be seen to be invariantly sufficient for $\delta_{ij} = \frac{\sigma_j^2}{\sigma_i^2}$.

For a preassigned $\delta^* > 1$, let $\delta_* = \frac{1}{\delta^*}$ and let $\delta_1 \in (1, \delta^*)$. Then,

$$(2.4.6) \qquad L_{ij}(n, \delta_1, 1) = \left[\frac{\delta_1^{\frac{1}{2}} (1 + T_{ij}(n))}{\delta_1 + T_{ij}(n)}\right]^{n-1},$$

$$(2.4.7) \qquad L_{ij}(n, \delta_1, \delta_*) = \left[\frac{(\delta_1\delta^*)^{\frac{1}{2}} (\delta^{*-1} + T_{ij}(n))}{\delta_1 + T_{ij}(n)}\right]^{n-1}.$$

Lemma 2.4.2. For the problem of selecting the population with the smallest variance among k normal populations, the sequential subset selection procedure $P_1(\delta_*, \delta_1)$ terminates with probability one if $\delta^* > \delta_1^2$.

Proof: It suffices to show that for any two populations, say π_1 and π_2, with probability one, at least one of them will be eliminated or both of them will be labelled as good. Without loss of generality, we assume that $\sigma_1^2 \leq \sigma_2^2$.

For the case when $\delta_{12} = \frac{\sigma_2^2}{\sigma_1^2} > \delta_1^{\frac{1}{2}}$, define

$$(2.4.8) \qquad N_1 = \min\left\{n \mid L_{12}(n, \delta_1, 1) \geq \frac{k-1}{1-P^*}\right\} .$$

Note that $T_{12}(n)$ is a consistent estimator of δ_{12}. Hence, as $n \to \infty$,

$$\frac{1}{n} \log L_{12}(n, \delta_1, 1) \to \log\left[\frac{\delta_1^{\frac{1}{2}}(1+\delta_{12})}{\delta_1 + \delta_{12}}\right] > 0 \text{ while } \frac{1}{n} \log \frac{k-1}{1-P^*} \to 0$$

as $n \to \infty$. Therefore, $P_{\underset{\sim}{\sigma}^2}\{N_1 < \infty\} = 1$.

For the case when $\delta_{12} \leq \delta_1^{\frac{1}{2}}$, define

$$(2.4.9) \qquad N_{ij} = \min\left\{n \mid L_{ij}(n, \delta_1, \delta_*) \geq \frac{k-1}{1-P^*}\right\}$$

for $i, j = 1, 2$, $i \neq j$, and

$$(2.4.10) \qquad N_2 = \max(N_{12}, N_{21}).$$

With the same argument, we can show as $\delta^* > \delta_1^2$, $P_{\underset{\sim}{\sigma}^2}\{N_{ij} < \infty\} = 1$ for $i, j = 1, 2$, $i \neq j$ and so $P_{\underset{\sim}{\sigma}^2}\{N_2 < \infty\} = 1$.

From the above discussion, we conclude that with probability one, at least one of π_1 and π_2 will be eliminated or both of them will be labelled as good. This completes the proof of Lemma 2.4.2.

In order to guarantee the P*-condition for the event $CD(\delta_*)$, we need to verify the condition (2.2.11). This can be done just by showing that $L_{ij}(n, \delta_1, a)$ is a nonincreasing function of a for $a \in (0, 1]$.

TABLE II.1

A simulation result for the performance of the procedure $P_1(\delta_*, \delta_1)$.

$P^* = 0.95$, $\delta^* = 1.5$, $\delta_1 = 0.5$, $(\theta_1, \theta_2, \theta_3, \theta_4) = (0, -1.0, -1.5, -2.5)$, $S = \{\pi_1\}$.

n	$\bar{X}_{1n}$	$\bar{X}_{2n}$	$\bar{X}_{3n}$	$\bar{X}_{4n}$	B_n	$D_1(n)$	$D_2(n)$	$D_3(n)$	$D_4(n)$
1	0.038	-.1517	-1.991	-2.665	{1, 2, 3, 4}	2.627	6.093	6.802	7.864
2	0.090	-0.742	-1.968	-3.190	{1, 2, 3, 4}	2.048	3.989	5.898	7.864
3	0.376	-0.478	-1.308	-3.179	{1, 2, 3, 4}	1.509	3.555	4.882	7.864
4	-0.037	-1.117	-1.326	-2.786	{1, 2, 3, 4}	1.025	3.555	3.990	6.577
5	0.273	-0.936	-1.580	-3.006	{1, 2, 3, 4}	0.734	3.555	3.990	6.577
6	0.523	-0.782	-1.663	-2.987	{1, 2, 3}	0.533	3.555	3.990	6.577
7	0.581	-0.735	-1.765	-	{1, 2, 3}	0.417	3.555	3.990	6.577
8	0.618	-0.915	-1.800	-	{1, 2}	0.232	3.555	3.990	6.577
9	0.258	-1.096	-	-	{1, 2}	0.232	3.555	3.990	6.577
10	0.353	-0.936	-	-	{1, 2}	0.214	3.489	3.990	6.577
11	0.419	-0.971	-	-	{1, 2}	0.120	3.489	3.990	6.577
12	0.409	-0.875	-	-	{1, 2}	0.120	3.417	3.990	6.577
13	0.378	-0.954	-	-	{1, 2}	0.065	3.417	3.990	6.577
14	0.366	-0.991	-	-	{1, 2}	0.023	3.417	3.990	6.577
15	0.293	-0.945	-	-	{1, 2}	0.023	3.265	3.990	6.577
16	0.410	-0.988	-	-	{1, 2}	0.	3.265	3.990	6.577

TABLE II.2

A simulation result for the performance of the procedure $P_1(\delta_*, \delta_1)$.

$P^* = 0.95$, $\delta^* = 1.5$, $\delta_1 = 0.5$, $(\theta_1, \theta_2, \theta_3, \theta_4) = (0, -0.1, -1.0, -2.5)$, $S = \{\pi_1, \pi_2\}$.

n	$\bar{X}_{1n}$	$\bar{X}_{2n}$	$\bar{X}_{3n}$	$\bar{X}_{4n}$	B_n	$D_1(n)$	$D_2(n)$	$D_3(n)$	$D_4(n)$
1	0.038	-0.617	-1.491	-2.665	{1, 2, 3, 4}	3.395	4.863	6.057	7.864
2	0.090	0.158	-1.468	-3.190	{1, 2, 3, 4}	2.985	2.827	5.191	7.864
3	0.376	0.422	-0.808	-3.179	{1, 2, 3, 4}	2.445	2.335	4.137	7.864
4	-0.037	-0.217	-0.826	-2.786	{1, 2, 3, 4}	1.868	2.315	3.189	6.577
5	0.273	-0.036	-1.080	-3.006	{1, 2, 3, 4}	1.510	2.292	3.189	6.577
6	0.523	0.118	-1.163	-2.987	{1, 2, 3}	1.250	2.289	3.189	6.577
7	0.581	0.165	-1.265	-	{1, 2, 3}	1.115	2.200	3.189	6.577
8	0.618	-0.015	-1.300	-	{1, 2, 3}	0.804	2.200	3.189	6.577
9	0.258	-0.196	-1.472	-	{1, 2, 3}	0.804	2.107	3.189	6.577
10	0.353	-0.036	-1.563	-	{1, 2}	0.804	1.948	3.189	6.577
11	0.419	-0.071	-	-	{1, 2}	0.730	1.948	3.189	6.577
12	0.409	0.025	-	-	{1, 2}	0.730	1.850	3.189	6.577
13	0.378	-0.054	-	-	{1, 2}	0.692	1.850	3.189	6.577
14	0.366	-0.091	-	-	{1, 2}	0.624	1.850	3.189	6.577
15	0.293	-0.045	-	-	{1, 2}	0.624	1.677	3.189	6.577
16	0.410	-0.088	-	-	{1, 2}	0.514	1.677	3.189	6.577
17	0.408	-0.001	-	-	{1, 2}	0.514	1.677	3.189	6.577
18	0.397	0.084	-	-	{1, 2}	0.514	1.566	3.189	6.577
19	0.368	0.079	-	-	{1, 2}	0.514	1.506	3.189	6.577
20	0.337	0.136	-	-	{1, 2}	0.514	1.344	3.189	6.577

CHAPTER III

LOCALLY OPTIMAL SUBSET SELECTION RULES BASED ON RANKS UNDER JOINT TYPE II CENSORING

3.1. Introduction

Let $\pi_1, \ldots, \pi_k$ be k (≥ 2) independent populations where π_i has the associated distribution function $F(x, \theta_i)$ and density $f(x, \theta_i)$ with the unknown parameter θ_i belonging to an interval (a, b) of the real line. Our goal is to select a subset (preferably small in size) of the k populations $\pi_1, \ldots, \pi_k$ that will contain the best (suitably defined) among them.

In practice, it sometimes happens that the actual values of the random variables can only be observed under great cost, or not at all, while their ordering is readily observable. This occurs for instance in life-testing when one only observes the order in which the parts under investigation fail without being able to record the actual time of failure. In problems of this type, one may desire to investigate decision rules based on ranks.

In dealing with the goal specified above, Gupta and McDonald (1970) studied three classes of subset selection rules based on ranks for selecting a subset containing the best among k populations when the underlying distributions are unknown. When the form of the underlying distribution is known but the values of the parameters θ_i, $i = 1, \ldots, k$, are unknown, Gupta, Huang and Nagel (1979) studied some

locally optimal subset selection rules based on ranks. The latter study leads to the conclusion that the class of subset selection rules R_3 of Gupta and McDonald (1970) is locally optimal in some sense. Huang and Panchapakesan (1982) also studied the problem of deriving some subset selection rules, based on ranks, which are locally optimal in the sense that the rules have the property of strong monotonicity. All the studies mentioned above only considered the situation where the ranks are completely observed.

We now consider a problem as follows: Suppose that there are k different devices and we want to select the best among them. From each kind of device, say π_i, n_i prototypes are taken for experiment and the $N = \sum_{i=1}^{k} n_i$ prototypes are simultaneously put on a life test. Due to design reasoning or cost consideration, the experiment terminates as soon as the first r failures among the N devices are observed for some predetermined value r, where $1 \leq r \leq N$. Based on these r observations, we want to ascertain which device is associated with the largest (expected) lifetime. Since we are only concerned with the first r failures, we call this censoring scheme as a joint type II censoring. In this chapter, our goal is to derive some locally optimal subset selection rules to select a subset containing the best based on the ranks under the joint type II censoring.

In the following sections, we assume that the functional form of the density function $f(x, \theta)$ is known but the value of the parameter θ is unknown. In Section 3.2, problems are formulated according to whether the sample sizes from the k different populations are equal or

not. Some properties related to the ranks under the joint type II censoring are also given there. In Section 3.3, we study the problem in the case when $n_1 = \cdots = n_k = n$. Following the earlier setup of Gupta, Huang and Nagel (1979), a locally optimal subset selection rule R_1 is derived. The property of local monotonicity related to the rule R_1 is also discussed. When the sample sizes are unequal, a locally optimal subset selection rule R_2 is derived in Section 3.4. In Section 3.5, a class of compatible censoring schemes is considered. We see that the properties of local optimality of R_1 or R_2 can be extended to those locally optimal rules which are based on the partial rank configurations censored by any compatible censoring scheme.

3.2. Formulation of the Problem

Let $\pi_1, \ldots, \pi_k$ be k (≥ 2) populations and let $f(x, \theta_i)$ be the density function associated with the population π_i for $i = 1, \ldots, k$. Let $\theta_{[1]} \leq \cdots \leq \theta_{[k]}$ be the ordered parameters of $\theta_1, \ldots, \theta_k$. Of course, the correct pairing of the ordered and unordered θ_i is unknown to us. The population associated with $\theta_{[k]}$ is called the best population. In case of a tie, one of the contenders is tagged and is called the best. Let $\Omega = \{\underset{\sim}{\theta} | \underset{\sim}{\theta} = (\theta_1, \ldots, \theta_k)\}$ and $\Omega_0 = \{\underset{\sim}{\theta} \in \Omega | \theta_1 = \cdots = \theta_k\}$. Let $\{X_{ij}, j = 1, \ldots, n_i\}$ be independent observations from π_i and let R_{ij} denote the rank of X_{ij} in the pooled sample of the $N = \sum_{i=1}^{k} n_i$ observations. The smallest observation has rank 1 and the largest has rank N. Let $x_1 \leq \cdots \leq x_N$ denote the ordered observations.

Definition 3.2.1. A rank configuration is an N-tuple $\underset{\sim}{\Delta} = (\Delta_1, \ldots, \Delta_N)$, $\Delta_i \in \{1, \ldots, k\}$ where $\Delta_i = j$ means that the ith smallest observation in the pooled sample comes from π_j.

Let $\mathcal{L} = \{\underset{\sim}{\Delta}\}$ denote the set of all rank configurations for the fixed constants n_i, $i = 1, \ldots, k$. For each $\underset{\sim}{\Delta} \in \mathcal{L}$, let $\chi_{\underset{\sim}{\Delta}} = \{\underset{\sim}{x} = (x_1, \ldots, x_N) \in \chi | \underset{\sim}{\Delta}_{\underset{\sim}{x}} = \underset{\sim}{\Delta}\}$, where $\chi = \{\underset{\sim}{x} | \underset{\sim}{x} = (x_1, \ldots, x_N)\}$ and $\underset{\sim}{\Delta}_{\underset{\sim}{x}}$ denotes the rank configuration of $\underset{\sim}{x} = (x_1, \ldots, x_N)$.

Let r be a predetermined integer such that $1 \leq r \leq N$. Under the joint type II censoring scheme, only the first r smallest observations in the pooled sample of the N $(X_{ij}, j = 1, \ldots, n_i; i = 1, \ldots, k)$ are observed. That is, for the rank configuration $\underset{\sim}{\Delta}_{\underset{\sim}{x}} = (\Delta_1, \ldots, \Delta_N)$, only the subvector $(\Delta_1, \ldots, \Delta_r)$ is observable. For this preassigned value r, let C_r be a function defined on $\mathcal{L}$ such that for each $\underset{\sim}{\Delta} = (\Delta_1, \ldots, \Delta_N) \in \mathcal{L}$, $C_r(\underset{\sim}{\Delta}) = (\Delta_1, \ldots, \Delta_r) = \underset{\sim}{\Delta}(r)$. Let $\mathcal{L}_r = C_r(\mathcal{L})$. Then, $\mathcal{L}_r = \{\underset{\sim}{\Delta}(r) | \underset{\sim}{\Delta} \in \mathcal{L}\}$. Hence, for each $\underset{\sim}{\Delta}(r) \in \mathcal{L}_r$, $\max(0, r - \sum_{\substack{j=1 \\ j \neq i}}^{k} n_j) \leq r_i \equiv \sum_{j=1}^{r} I_{\{i\}}(\Delta_j) \leq \min(r, n_i)$ for each $i = 1, \ldots, k$, and $\sum_{i=1}^{k} r_i = r$. We call $\underset{\sim}{\Delta}(r)$ as a joint type II censored rank configuration.

For each $\underset{\sim}{\Delta}(r) \in \mathcal{L}_r$, define the set $\mathcal{L}(\underset{\sim}{\Delta}(r)) = \{\underset{\sim}{\Delta} \in \mathcal{L} | C_r(\underset{\sim}{\Delta}) = \underset{\sim}{\Delta}(r)\}$. Let $|A|$ denote the number of elements in the set A. Then,

$$(3.2.1) \qquad |\mathcal{L}(\underset{\sim}{\Delta}(r))| = \prod_{m=1}^{k} \binom{N - r - \sum_{i=1}^{m-1} (n_i - r_i)}{n_m - r_m},$$

where $\sum_{i=1}^{0} \equiv 0$. Also,

$$(3.2.2) \qquad \sum_{\underset{\sim}{\Delta}(r)\in\mathcal{L}_r} |\mathcal{L}(\underset{\sim}{\Delta}(r))| = N! \Big/ \left(\prod_{j=1}^{k} n_j! \right).$$

Let $\mathcal{D}$ be the decision space consisting of all the 2^k subsets of the set $\{1, \ldots, k\}$. Any subset is denoted by d so that $\mathcal{D} = \{d | d \sqsubset \{1, \ldots, k\}\}$. A decision d is the selection of a subset of the k populations. The fact that $i \in d$ means that population π_i is included in the selected subset if decision d is made. Let $\delta(\underset{\sim}{\Delta}(r), d)$ denote the probability that the decision d is made if the censored rank configuration $\underset{\sim}{\Delta}(r)$ is observed. Let $\alpha_i(\underset{\sim}{\Delta}(r))$, $i = 1, \ldots, k$, denote the individual selection probability of the k populations, where

$$(3.2.3) \qquad \alpha_i(\underset{\sim}{\Delta}(r)) = \sum_{d \ni i} \delta(\underset{\sim}{\Delta}(r), d),$$

the summation being over all the subsets containing i.

Definition 3.2.2. A subset selection rule R based on the censored ranks is a measurable mapping from $\mathcal{L}_r$ into $[0, 1]^k$ such that

$$R(\underset{\sim}{\Delta}(r)) = (\alpha_1(\underset{\sim}{\Delta}(r)), \ldots, \alpha_k(\underset{\sim}{\Delta}(r))).$$

Let $P_i(\underset{\sim}{\theta})$ denote the probability of including the population π_i in the selected subset when $\underset{\sim}{\theta} = (\theta_1, \ldots, \theta_k)$ are the true parameters. That is, $P_i(\underset{\sim}{\theta}) = E_{\underset{\sim}{\theta}}[\alpha_i(\underset{\sim}{\Delta}(r))]$ where the expectation is over the set $\mathcal{L}_r$. Any decision d that corresponds to the selection of the best

population is called a correct selection (CS). The probability of a correct selection is denoted by $P_{\underset{\sim}{\theta}}(CS|R)$ when the subset selection rule R is applied.

Let G denote the group of permutations g of the integers 1,..., k. We write $g(1, \ldots, k) = (g1, \ldots, gk)$. Let h denote the inverse of g and define $g(\theta_1, \ldots, \theta_k) = (\theta_{h1}, \ldots, \theta_{hk})$. For $n_1 = n_2 = \cdots = n_k = n$, for each $\underset{\sim}{\Delta} \in \mathcal{L}$, $\underset{\sim}{\Delta}(r) \in \mathcal{L}_r$, let $\overline{g}$ and $\overline{\overline{g}}$ be defined by $\overline{g}\underset{\sim}{\Delta} = (g\Delta_1, \ldots, g\Delta_N)$ and $\overline{\overline{g}}\underset{\sim}{\Delta}(r) = (g\Delta_1, \ldots, g\Delta_r)$, respectively. Thus, both $\overline{g}$ and $\overline{\overline{g}}$ are induced from g. Let $\overline{G} = \{\overline{g}\}$ and $\overline{\overline{G}} = \{\overline{\overline{g}}\}$. It is easy to see that $C_r(\overline{g}\underset{\sim}{\Delta}) = \overline{\overline{g}}(C_r(\underset{\sim}{\Delta}))$. Also, $\underset{\sim}{\Delta} \in \mathcal{L}(\underset{\sim}{\Delta}(r))$ iff $\overline{g}\underset{\sim}{\Delta} \in \mathcal{L}(\overline{\overline{g}}\underset{\sim}{\Delta}(r))$. Hence,

$$(3.2.4) \qquad |\mathcal{L}(\underset{\sim}{\Delta}(r))| = |\mathcal{L}(\overline{\overline{g}}\underset{\sim}{\Delta}(r))|$$

for each $\underset{\sim}{\Delta}(r) \in \mathcal{L}_r$ and for each $\overline{\overline{g}} \in \overline{\overline{G}}$.

Definition 3.2.3. A subset selection rule R on $\mathcal{L}_r$ is invariant under permutation if and only if $(\alpha_1(\overline{\overline{g}}\underset{\sim}{\Delta}(r)), \ldots, \alpha_k(\overline{\overline{g}}\underset{\sim}{\Delta}(r))) = g(\alpha_1(\underset{\sim}{\Delta}(r)), \ldots, \alpha_k(\underset{\sim}{\Delta}(r)))$ for all $\underset{\sim}{\Delta}(r) \in \mathcal{L}_r$, $g \in G$ and $\overline{\overline{g}}$ induced from g.

Let $f(x, \theta_i)$ be the density function associated with population π_i, with the parameter θ_i belonging to some interval (a, b) of the real line, where $-\infty \leq a < b \leq \infty$. Let $\Omega = \{\underset{\sim}{\theta} | \underset{\sim}{\theta} = (\theta_1, \ldots, \theta_k)\}$, $\Omega_0 = \{\underset{\sim}{\theta} \in \Omega | \theta_1 = \cdots = \theta_k\}$ and $\Omega_i = \{\underset{\sim}{\theta} \in \Omega | \theta_i \geq \theta_j \text{ for all } j \neq i\}$. Furthermore, let the density $f(x, \theta)$ have the following properties:

Condition A:

$$(3.2.5)\quad \begin{cases} \text{(i)} & f(x, \theta) \text{ is absolutely continuous in } \theta \text{ for every } x; \\ \text{(ii)} & \text{the limit } \dot{f}(x, \theta_0) = \lim_{\theta \to \theta_0} (f(x, \theta) - f(x, \theta_0))/(\theta - \theta_0) \\ & \text{exists for every } x, \text{ for each } \theta_0 \in (a, b); \\ \text{(iii)} & \lim_{\theta \to \theta_0} \int_{-\infty}^{\infty} |\dot{f}(x, \theta)| dx = \int_{-\infty}^{\infty} |\dot{f}(x, \theta_0)| dx < \infty \\ & \text{holds for every } \theta_0 \in (a, b). \end{cases}$$

Goal 1: In the case when $n_1 = \cdots = n_k = n$, our goal is to derive an invariant subset selection rule R, based on the joint type II censored ranks, such that

$$(3.2.6)\quad \begin{cases} \text{(i)} & \inf_{\underset{\sim}{\theta}_0 \in \Omega_0} P_{\underset{\sim}{\theta}_0}(CS|R) = P^* \text{ where } P^* \in (\frac{1}{k}, 1) \text{ is prespecified;} \\ \text{(ii)} & P_{\underset{\sim}{\theta}}(CS|R) \text{ is as large as possible for all } \underset{\sim}{\theta} \text{ in a} \\ & \text{neighborhood of } \underset{\sim}{\theta}_0 \in \Omega_0. \end{cases}$$

Note that for each $\underset{\sim}{\theta}_0 \in \Omega_0$, $P_{\underset{\sim}{\theta}_0}(CS|R)$ will be interpreted as the probability of selecting a specified population.

Goal 2: In the unequal sample sizes case, we consider the class of all subset selection rules satisfying $P_1(\underset{\sim}{\theta}_0) = \cdots = P_k(\underset{\sim}{\theta}_0) = P^*$ for

all $\underset{\sim}{\theta}_0 \in \Omega_0$. Among those rules, we are interested in the one which maximizes

$$(3.2.7) \qquad \sum_{i=1}^{k} \frac{\partial}{\partial \theta_i} \log P_{\underset{\sim}{\theta}}(CS|\underset{\sim}{\theta} \in \Omega_i)\Big|_{\underset{\sim}{\theta}=\underset{\sim}{\theta}_0 \in \Omega_0}.$$

3.3. A Locally Optimal Subset Selection Rule for Goal 1

In this section, we will derive an invariant subset selection rule, based on the joint type II censored ranks for Goal 1, i.e., the conditions of (3.2.6) are satisfied.

For each $\underset{\sim}{\theta} \in \Omega$, $\underset{\sim}{\Delta}(r) \in \mathcal{L}_r$, let $P_{\underset{\sim}{\theta}}(\underset{\sim}{\Delta}(r))$ denote the probability that the joint type II censored rank configuration $\underset{\sim}{\Delta}(r)$ is observed under $\underset{\sim}{\theta}$. Also, let $P_{\underset{\sim}{\theta}}(\underset{\sim}{\Delta})$, $\underset{\sim}{\Delta} \in \mathcal{L}$, denote the probability that the complete rank configuration $\underset{\sim}{\Delta}$ is observed under $\underset{\sim}{\theta}$. Then,

$$(3.3.1) \qquad P_{\underset{\sim}{\theta}}(\underset{\sim}{\Delta}) = (n!)^k \int_{-\infty}^{\infty} \int_{-\infty}^{x_N} \cdots \int_{-\infty}^{x_2} \prod_{j=1}^{N} f(x_j, \theta_{\Delta_j}) dx_1 \cdots dx_N.$$

It is also clear that for each $\underset{\sim}{\Delta}(r) \in \mathcal{L}_r$,

$$(3.3.2) \qquad P_{\underset{\sim}{\theta}}(\underset{\sim}{\Delta}(r)) = \sum_{\underset{\sim}{\Delta} \in \mathcal{L}(\underset{\sim}{\Delta}(r))} P_{\underset{\sim}{\theta}}(\underset{\sim}{\Delta}).$$

Let $\underset{\sim}{\theta}_0 = (\theta_0, \ldots, \theta_0) \in \Omega_0$, where $\theta_0 \in (a, b)$. By applying a simple algebraic computation, $P_{\underset{\sim}{\theta}}(\underset{\sim}{\Delta})$ can be written as follows:

$$(3.3.3)\qquad P_{\underset{\sim}{\theta}}(\underset{\sim}{\Delta}) = (n!)^k\left[A_0(\theta_0) + \sum_{i=1}^{k}(\theta_i - \theta_0)A_i(\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0, \underset{\sim}{\theta})\right]$$

where $A_0(\theta_0) = \int_{-\infty}^{\infty}\int_{-\infty}^{x_N}\cdots\int_{-\infty}^{x_2}\prod_{j=1}^{N} f(x_j, \theta_0)dx_1\cdots dx_N = \frac{1}{N!}$ which is independent of θ_0, and

$$(3.3.4)\qquad A_i(\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0, \underset{\sim}{\theta}) = \sum_{\substack{j=1\\ \Delta_j=i}}^{N}\int_{-\infty}^{\infty}\int_{-\infty}^{x_N}\cdots\int_{-\infty}^{x_2} q(i, j, \underset{\sim}{\theta}_0, \underset{\sim}{\theta}, \underset{\sim}{x})dx_1\cdots dx_N$$

for each $i = 1, \ldots, k$, where $\underset{\sim}{x} = (x_1, \ldots, x_N)$ and

$$(3.3.5)\qquad q(i,j,\underset{\sim}{\theta}_0,\underset{\sim}{\theta},\underset{\sim}{x}) = \frac{f(x_j,\theta_i)-f(x_j,\theta_0)}{\theta_i - \theta_0}\prod_{m=1}^{j-1} f(x_m,\theta_0)\prod_{m=j+1}^{N} f(x_m,\theta_{\Delta_m}).$$

Here, we define $\prod_{j=1}^{0} \equiv 1$, $\prod_{j=N+1}^{N} \equiv 1$ and $[f(x_j,\theta_i) - f(x_j,\theta_0)]/(\theta_i - \theta_0) = 0$ if $\theta_i = \theta_0$.

Also, for $\underset{\sim}{\theta}_0 \in \Omega_0$, we have

$$(3.3.6)\qquad P_{\underset{\sim}{\theta}_0}(\underset{\sim}{\Delta}(r)) = \frac{(n!)^k}{N!}\,|\mathcal{L}(\underset{\sim}{\Delta}(r))|.$$

Now, let $T(\underset{\sim}{\Delta}(r), \underset{\sim}{\theta}_0, \underset{\sim}{\theta}) = P_{\underset{\sim}{\theta}}(\underset{\sim}{\Delta}(r))/P_{\underset{\sim}{\theta}_0}(\underset{\sim}{\Delta}(r))$ and define

$$(3.3.7) \qquad T^*(\underset{\sim}{\Delta}(r), \underset{\sim}{\theta}_0) = \lim_{\|\underset{\sim}{\theta}-\underset{\sim}{\theta}_0\|\to 0} T(\underset{\sim}{\Delta}(r), \underset{\sim}{\theta}_0, \underset{\sim}{\theta})$$

where $||\underset{\sim}{\theta} - \underset{\sim}{\theta}_0|| = \max_{1\le i\le k} |\theta_i - \theta_0|$.

By (3.3.3) and (3.3.6),

$$(3.3.8) \qquad T(\underset{\sim}{\Delta}(r),\underset{\sim}{\theta}_0,\underset{\sim}{\theta}) = 1 + \frac{N!}{|\mathcal{L}(\underset{\sim}{\Delta}(r))|} \sum_{\underset{\sim}{\Delta}\in\mathcal{L}(\underset{\sim}{\Delta}(r))} \sum_{i=1}^{k} (\theta_i-\theta_0)A_i(\underset{\sim}{\Delta},\underset{\sim}{\theta}_0,\underset{\sim}{\theta}).$$

Thus, if $\underset{\sim}{\theta} = (\theta_1, \ldots, \theta_k)$ is in the neighborhood of $\underset{\sim}{\theta}_0$ with $\theta_i \neq \theta_0$ for all $i = 1, \ldots, k$, then,

$$(3.3.9) \qquad T^*(\underset{\sim}{\Delta}(r),\underset{\sim}{\theta}_0) \approx 1 + \frac{N!}{|\mathcal{L}(\underset{\sim}{\Delta}(r))|} \sum_{\underset{\sim}{\Delta}\in\mathcal{L}(\underset{\sim}{\Delta}(r))} \sum_{i=1}^{k} (\theta_i-\theta_0)$$

$$\times \lim_{||\underset{\sim}{\theta}-\underset{\sim}{\theta}_0||\to 0} A_i(\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0, \underset{\sim}{\theta}).$$

Under the Condition A, following an argument analogous to a theorem (page 71) of Hajek and Sidak (1967), we have

$$(3.3.10) \qquad \lim_{||\underset{\sim}{\theta}-\underset{\sim}{\theta}_0||\to 0} A_i(\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0, \underset{\sim}{\theta}) = A_i^*(\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0) = \sum_{\substack{j=1 \\ \Delta_j=i}}^{N} B_j(\underset{\sim}{\theta}_0),$$

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

$$B_j(\underset{\sim}{\theta}_0) = \int_{-\infty}^{\infty}\int_{-\infty}^{x_N}\cdots\int_{-\infty}^{x_2} \dot{f}(x_j,\theta_0) \prod_{\substack{m=1\\ m\neq j}}^{N} f(x_m, \theta_0)dX_1 \cdots dX_N$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{x_N}\cdots\int_{-\infty}^{x_2} \frac{\dot{f}(x_j,\theta_0)}{f(x_j,\theta_0)} \prod_{m=1}^{N} f(x_m,\theta_0)dX_1 \cdots dX_N \tag{3.3.11}$$

$$= \frac{1}{(j-1)!(N-j)!}\int_0^1 u^{j-1}(1-u)^{N-j}\,\phi(u, f, \theta_0)du$$

and

$$\phi(u, f, \theta_0) = \dot{f}(F^{-1}(u,\theta_0),\theta_0)/f(F^{-1}(u,\theta_0),\theta_0),\ u\in(0,1). \tag{3.3.12}$$

That is, there exists an $\varepsilon > 0$ such that as $0 < ||\underset{\sim}{\theta} - \underset{\sim}{\theta}_0|| < \varepsilon$, $A_i(\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0, \underset{\sim}{\theta})$ is approximately equal to $A_i^*(\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0)$ for each $i = 1, \ldots, k$. Hence

$$T^*(\underset{\sim}{\Delta}(r), \underset{\sim}{\theta}_0) \approx 1 + \frac{N!}{|\mathcal{L}(\underset{\sim}{\Delta}(r))|}\sum_{\underset{\sim}{\Delta}\in\mathcal{L}(\underset{\sim}{\Delta}(r))}\sum_{i=1}^{k}(\theta_i-\theta_0)A_i^*(\underset{\sim}{\Delta},\underset{\sim}{\theta}_0)$$

$$= 1 + N!\sum_{i=1}^{k}(\theta_i - \theta_0)T_i^*(\underset{\sim}{\Delta}(r), \underset{\sim}{\theta}_0) \tag{3.3.13}$$

where

$$T_i^*(\underset{\sim}{\Delta}(r), \underset{\sim}{\theta}_0) = \frac{1}{|\mathcal{L}(\underset{\sim}{\Delta}(r))|}\sum_{\underset{\sim}{\Delta}\in\mathcal{L}(\underset{\sim}{\Delta}(r))} A_i^*(\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0). \tag{3.3.14}$$

Lemma 3.3.1. Suppose that the density function $f(x, \theta)$ satisfies the Condition A. For each $\theta_0 \in (a, b)$, let $V(\theta_0) = \sum_{j=1}^{N} B_j(\underset{\sim}{\theta}_0)$ where $\underset{\sim}{\theta}_0 = (\theta_0, \ldots, \theta_0) \in \Omega_0$. Then $V(\theta_0) = 0$ for all $\theta_0 \in (a, b)$.

Proof: Note that for each $\theta_0 \in (a, b)$,

$$\sum_{j=1}^{N} B_j(\underset{\sim}{\theta}_0) = \sum_{j=1}^{N} \int_{-\infty}^{\infty} \int_{-\infty}^{x_N} \cdots \int_{-\infty}^{x_2} \dot{f}(x_j, \theta_0) \prod_{\substack{m=1 \\ m \neq j}}^{N} f(x_m, \theta_0) dx_1 \cdots dx_N$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{x_N} \cdots \int_{-\infty}^{x_2} \sum_{j=1}^{N} \dot{f}(x_j, \theta_0) \prod_{\substack{m=1 \\ m \neq j}}^{N} f(x_m, \theta_0) dx_1 \cdots dx_N$$

(3.3.15)

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{x_N} \cdots \int_{-\infty}^{x_2} \left[\frac{d}{d\theta} \prod_{m=1}^{N} f(x_m, \theta) \right] \Bigg|_{\theta=\theta_0} dx_1 \cdots dx_N$$

$$= \frac{d}{d\theta} \int_{-\infty}^{\infty} \int_{-\infty}^{x_N} \cdots \int_{-\infty}^{x_2} \prod_{m=1}^{N} f(x_m, \theta) dx_1 \cdots dx_N \Bigg|_{\theta=\theta_0},$$

where the last equality is obtained under Condition A.

Therefore $V(\theta_0) = 0$ for all $\theta_0 \in (a, b)$ since $\int_{-\infty}^{\infty} \int_{-\infty}^{x_N} \cdots \int_{-\infty}^{x_2} \prod_{m=1}^{N} f(x_m, \theta)\, dx_1 \cdots dx_N = \frac{1}{N!}$ which is independent of θ. This completes the proof of this lemma.

Lemma 3.3.2. Let $\underset{\sim}{\theta} \in \Omega$ and let $P_i(\underset{\sim}{\theta}) = E_{\underset{\sim}{\theta}}[\alpha_i(\underset{\sim}{\Delta}(r))]$ be the probability of including population π_i in the selected subset under $\underset{\sim}{\theta}$ by applying an invariant subset selection rule R. Let $G(i) = \{g \in G | gi = i\}$. Then

$$P_i(\underset{\sim}{\theta}) = \sum_{\underset{\sim}{\Delta}(r) \in \mathcal{L}_r} \left[\frac{(n!)^k}{N!} |\mathcal{L}(\underset{\sim}{\Delta}(r))| + \frac{(n!)^k}{(k-1)!} W(\underset{\sim}{\Delta}(r), \underset{\sim}{\theta}, \underset{\sim}{\theta}_0, G(i)) \right] \alpha_i(\underset{\sim}{\Delta}(r)) \tag{3.3.16}$$

where

$$W(\underset{\sim}{\Delta}(r), \underset{\sim}{\theta}, \underset{\sim}{\theta}_0, G(i)) = \sum_{\underset{\sim}{\Delta} \in \mathcal{L}(\underset{\sim}{\Delta}(r))} \sum_{g \in G(i)} \sum_{j=1}^{k} (\theta_{hj} - \theta_0) \cdot A_j(\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0, g\underset{\sim}{\theta}), \tag{3.3.17}$$

h is the inverse of $g \in G(i)$ and $\underset{\sim}{\theta}_0 = (\theta_0, \ldots, \theta_0) \in \Omega_0$.

Proof: By the invariance property of the invariant subset selection rule R, we have

$$\begin{aligned} P_i(\underset{\sim}{\theta}) &= E_{\underset{\sim}{\theta}}[\alpha_i(\underset{\sim}{\Delta}(r))] \\ &= \frac{1}{(k-1)!} \sum_{g \in G(i)} E_{g\underset{\sim}{\theta}}[\alpha_i(\bar{\bar{g}}\underset{\sim}{\Delta}(r))] \\ &= \frac{1}{(k-1)!} \sum_{g \in G(i)} E_{g\underset{\sim}{\theta}}[\alpha_{hi}(\underset{\sim}{\Delta}(r))] \text{ where } h = g^{-1} \in G(i) \end{aligned} \tag{3.3.18}$$

$$= \frac{1}{(k-1)!} \sum_{g\in G(i)} E_{g\underset{\sim}{\theta}}[\alpha_i(\underset{\sim}{\Delta}(r))]$$

$$= \frac{1}{(k-1)!} \sum_{g\in G(i)} \left[\sum_{\underset{\sim}{\Delta}(r)\in \mathcal{L}_r} \alpha_i(\underset{\sim}{\Delta}(r)) P_{g\underset{\sim}{\theta}}(\underset{\sim}{\Delta}(r)) \right]$$

$$= \sum_{\underset{\sim}{\Delta}(r)\in \mathcal{L}_r} \left[\frac{1}{(k-1)!} \sum_{g\in G(i)} P_{g\underset{\sim}{\theta}}(\underset{\sim}{\Delta}(r)) \right] \alpha_i(\underset{\sim}{\Delta}(r)).$$

Then, from (3.3.2), (3.3.3) and (3.3.18), we obtain

$$P_i(\underset{\sim}{\theta}) = \sum_{\underset{\sim}{\Delta}(r)\in \mathcal{L}_r} \left[\frac{1}{(k-1)!} \sum_{g\in G(i)} \sum_{\underset{\sim}{\Delta}\in \mathcal{L}(\underset{\sim}{\Delta}(r))} P_{g\underset{\sim}{\theta}}(\underset{\sim}{\Delta}) \right] \alpha_i(\underset{\sim}{\Delta}(r))$$

$$= \sum_{\underset{\sim}{\Delta}(r)\in \mathcal{L}_r} \left[\frac{1}{(k-1)!} \sum_{g\in G(i)} \sum_{\underset{\sim}{\Delta}\in \mathcal{L}(\underset{\sim}{\Delta}(r))} (n!)^k \right.$$

$$\left. \times \left(\frac{1}{N!} + \sum_{j=1}^{k} (\theta_{hj} - \theta_0) A_j(\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0, g\underset{\sim}{\theta}) \right) \right] \alpha_i(\underset{\sim}{\Delta}(r))$$

(3.3.19)

$$= \sum_{\underset{\sim}{\Delta}(r)\in \mathcal{L}_r} \left[\frac{(n!)^k}{N!} |\mathcal{L}(\underset{\sim}{\Delta}(r))| + \frac{(n!)^k}{(k-1)!} \sum_{\underset{\sim}{\Delta}\in \mathcal{L}(\underset{\sim}{\Delta}(r))} \sum_{g\in G(i)} \sum_{j=1}^{k} \right.$$

$$\left. (\theta_{hj} - \theta_0) A_j(\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0, g\underset{\sim}{\theta}) \right] \alpha_i(\underset{\sim}{\Delta}(r))$$

$$= \sum_{\underset{\sim}{\Delta}(r)\in \mathcal{L}_r} \left[\frac{(n!)^k}{N!} |\mathcal{L}(\underset{\sim}{\Delta}(r))| + \frac{(n!)^k}{(k-1)!} W(\underset{\sim}{\Delta}(r), \underset{\sim}{\theta}, \underset{\sim}{\theta}_0, G(i)) \right]$$

$$\times \alpha_i(\underset{\sim}{\Delta}(r)).$$

This completes the proof of Lemma 3.3.2.

Lemma 3.3.3. Suppose that the density function $f(x, \theta)$ satisfies the requirements of Condition A. Let $G(i) = \{g \in G | gi = i\}$ and $A_j^*(\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0) = \sum_{\substack{m=1 \\ \Delta_m = j}}^{N} B_m(\underset{\sim}{\theta}_0)$. Then

$$\sum_{g \in G(i)} \sum_{j=1}^{k} (\theta_{hj} - \theta_0) A_j^*(\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0) = (k-2)!(k\theta_i - U) A_i^*(\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0) \tag{3.3.20}$$

for each $i = 1, \ldots, k$, for each $\underset{\sim}{\theta} \in \Omega$, $\underset{\sim}{\theta}_0 \in \Omega_0$ where $U = \sum_{j=1}^{k} \theta_j$ and h is the inverse of $g \in G(i)$ and $\underset{\sim}{\Delta} \in \mathcal{L}$.

Proof: First note that $\sum_{j=1}^{k} A_j^*(\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0) = \sum_{j=1}^{k} \sum_{\substack{m=1 \\ \Delta_m = j}}^{N} B_m(\underset{\sim}{\theta}_0) = \sum_{m=1}^{N} B_m(\underset{\sim}{\theta}_0) = 0$ which is obtained from Lemma 3.3.1. Now,

$$\begin{aligned}
&\sum_{g \in G(i)} \sum_{j=1}^{k} (\theta_{hj} - \theta_0) A_j^*(\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0) \\
&= \sum_{g \in G(i)} \sum_{j=1}^{k} \theta_{hj} A_j^*(\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0) \\
&= \sum_{j=1}^{k} \left[A_j^*(\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0) \sum_{g \in G(i)} \theta_{hj} \right] \\
&= \sum_{\substack{j=1 \\ j \neq i}}^{k} A_j^*(\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0) \sum_{g \in G(i)} \theta_{hj} + A_i^*(\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0) \sum_{g \in G(i)} \theta_{hi}
\end{aligned} \tag{3.3.21}$$

$$= \sum_{\substack{j=1 \\ j\neq i}}^{k} A_j^*(\underset{\sim}{\Delta},\underset{\sim}{\theta}_0)\left[(k-2)! \sum_{\substack{m=1 \\ m\neq i}}^{k} \theta_m\right] + (k-1)!\theta_i A_i^*(\underset{\sim}{\Delta},\underset{\sim}{\theta}_0)$$

$$= (k-2)!(U-\theta_i) \sum_{\substack{j=1 \\ j\neq i}}^{k} A_j^*(\underset{\sim}{\Delta},\underset{\sim}{\theta}_0) + (k-1)!\theta_i A_i^*(\underset{\sim}{\Delta},\underset{\sim}{\theta}_0)$$

$$= (k-2)!(k\theta_i - U)A_i^*(\underset{\sim}{\Delta},\underset{\sim}{\theta}_0)$$

where the first and the last equalities are obtained due to the fact that $\sum_{j=1}^{k} A_j^*(\underset{\sim}{\Delta},\underset{\sim}{\theta}_0) = 0$. This completes the proof of Lemma 3.3.3.

Theorem 3.3.4. Let $\underset{\sim}{\theta} \in \Omega$ be any point in the neighborhood of $\underset{\sim}{\theta}_0 \in \Omega_0$. Let $P_i(\underset{\sim}{\theta}) = E_{\underset{\sim}{\theta}}[\alpha_i(\underset{\sim}{\Delta}(r))]$ be the probability of including population π_i in the selected subset under $\underset{\sim}{\theta}$ by applying an invariant subset selection rule R. Then, under the condition A, for each i = 1,2,...,k,

$$(3.3.22)\quad P_i(\underset{\sim}{\theta}) \approx E_{\underset{\sim}{\theta}_0}\left\{\left[1 + \frac{(k\theta_i - U)N!}{k-1} T_i^*(\underset{\sim}{\Delta}(r),\underset{\sim}{\theta}_0)\right]\alpha_i(\underset{\sim}{\Delta}(r))\right\}.$$

Proof: It is trivial that under the condition A, $|A_i(\underset{\sim}{\Delta},\underset{\sim}{\theta}_0,\underset{\sim}{\theta})| < \infty$ and $|A_i^*(\underset{\sim}{\Delta},\underset{\sim}{\theta}_0)| < \infty$ for all i = 1, ..., k. It is also clear that $(\theta_j - \theta_0)A_i(\underset{\sim}{\Delta},\underset{\sim}{\theta}_0,g\underset{\sim}{\theta}) = (\theta_j - \theta_0)\, A_i^*(\underset{\sim}{\Delta},\underset{\sim}{\theta}_0)$ if $\theta_j = \theta_0$. Thus, we assume that $\theta_j \neq \theta_0$ for each j = 1, ..., k. Then, by the assumption and (3.3.10), we can choose $\varepsilon > 0$ so small that as $||\underset{\sim}{\theta} - \underset{\sim}{\theta}_0|| < \varepsilon$ where $||\underset{\sim}{\theta} - \underset{\sim}{\theta}_0|| = \max_{1\leq i\leq k} |\theta_i - \theta_0|$, we have $A_i(\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0, g\underset{\sim}{\theta}) \approx A_i^*(\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0)$ for

all $g\varepsilon G$ and so $(\theta_{hj} - \theta_0)\, A_i(\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0, g\underset{\sim}{\theta}) \approx (\theta_{hj} - \theta_0)\, A_i^*(\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0)$ for all $g\varepsilon G$ where h is the inverse of g. Thus, either $\min_{1\le i\le k} |\theta_i - \theta_0| = 0$ or $\min_{1\le i\le k} |\theta_i - \theta_0| > 0$, if $||\underset{\sim}{\theta} - \underset{\sim}{\theta}_0|| < \varepsilon$, we have

$$\sum_{g\varepsilon G(i)} \sum_{j=1}^{k} (\theta_{hj} - \theta_0) A_j(\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0, g\underset{\sim}{\theta})$$

$$\approx \sum_{g\varepsilon G(i)} \sum_{j=1}^{k} (\theta_{hj} - \theta_0) A_j^*(\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0) \tag{3.3.23}$$

$$= (k-2)!(k\theta_i - U) A_i^*(\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0)$$

where the last equality is due to Lemma 3.3.3. Then, by (3.3.19) and (3.3.23), we obtain

$$\begin{aligned} & P_i(\underset{\sim}{\theta}) \\ & \approx \sum_{\underset{\sim}{\Delta}(r)\varepsilon \mathcal{L}_r} \left[\frac{(n!)^k}{N!} |\mathcal{L}(\underset{\sim}{\Delta}(r))| + \frac{(n!)^k}{(k-1)!} \sum_{\underset{\sim}{\Delta}\varepsilon \mathcal{L}(\underset{\sim}{\Delta}(r))} (k-2)! \right. \\ & \qquad \left. \times (k\theta_i - U) A_i^*(\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0) \right] \alpha_i(\underset{\sim}{\Delta}(r)) \\ & = E_{\underset{\sim}{\theta}_0} \left\{ \left[1 + \frac{(k\theta_i - U)N!}{k-1} \frac{1}{|\mathcal{L}(\underset{\sim}{\Delta}(r))|} \sum_{\underset{\sim}{\Delta}\varepsilon \mathcal{L}(\underset{\sim}{\Delta}(r))} A_i^*(\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0) \right] \right. \\ & \qquad \left. \times \alpha_i(\underset{\sim}{\Delta}(r)) \right\} \\ & = E_{\underset{\sim}{\theta}_0} \left\{ \left[1 + \frac{(k\theta_i - U)N!}{k-1} T_i^*(\underset{\sim}{\Delta}(r), \underset{\sim}{\theta}_0) \right] \alpha_i(\underset{\sim}{\Delta}(r)) \right\}. \end{aligned} \tag{3.3.24}$$

This completes the proof of Theorem 3.3.4.

Now, define subset selection rule R_1 as follows:

$$(3.3.25)\quad \alpha_i(\underset{\sim}{\Delta}(r)) = \begin{cases} 1 & \text{if } T_i^*(\underset{\sim}{\Delta}(r), \underset{\sim}{\theta}_0) > c(\underset{\sim}{\theta}_0); \\ \rho(\underset{\sim}{\theta}_0) & \text{if } T_i^*(\underset{\sim}{\Delta}(r), \underset{\sim}{\theta}_0) = c(\underset{\sim}{\theta}_0); \\ 0 & \text{if } T_i^*(\underset{\sim}{\Delta}(r), \underset{\sim}{\theta}_0) < c(\underset{\sim}{\theta}_0); \end{cases}$$

where the constants $c(\underset{\sim}{\theta}_0)$ and $\rho(\underset{\sim}{\theta}_0)$, $(0 \leq \rho(\underset{\sim}{\theta}_0) < 1)$, depend on the parameter θ_0, and can be determined by

$$(3.3.26)\quad P_{\underset{\sim}{\theta}_0}\{T_1^*(\underset{\sim}{\Delta}(r),\underset{\sim}{\theta}_0) > c(\underset{\sim}{\theta}_0)\} + \rho(\underset{\sim}{\theta}_0)P_{\underset{\sim}{\theta}_0}\{T_1^*(\underset{\sim}{\Delta}(r),\underset{\sim}{\theta}_0) = c(\underset{\sim}{\theta}_0)\} = P^*.$$

We then have the following theorem.

Theorem 3.3.5. Suppose that the density function $f(x,\theta)$ satisfies the Condition A. Then, the subset selection rule R_1 maximizes $P_{\underset{\sim}{\theta}}(CS|R)$ in a neighborhood of $\underset{\sim}{\theta}_0 \in \Omega_0$, among all invariant subset selection rules, based on the joint type II censored ranks, satisfying $\inf_{\underset{\sim}{\theta}\in\Omega_0} P_{\underset{\sim}{\theta}}(CS|R) = P^*$.

Proof: Without loss of generality, we assume that π_k is the best population. Then by Theorem 3.3.4, for any $\underset{\sim}{\theta} \in \Omega_k$ in a neighborhood of $\underset{\sim}{\theta}_0 \in \Omega_0$,

$$P_{\underset{\sim}{\theta}}(CS|R) = P_k(\underset{\sim}{\theta})$$

(3.3.27)

$$\approx E_{\underset{\sim}{\theta}_0}\left\{\left[1 + \frac{(k\theta_k - U)N!}{k-1} T_k^*(\underset{\sim}{\Delta}(r), \underset{\sim}{\theta}_0)\right]\alpha_k(\underset{\sim}{\Delta}(r))\right\}.$$

Since $k\theta_k - U = \sum_{j=1}^{k-1} (\theta_k - \theta_j) \geq 0$, then by Neyman-Pearson lemma, we conclude this theorem.

Local Monotonicity of the Subset Selection Rule R_1

Let R be a subset selection rule and $P_i(\underset{\sim}{\theta})$ be the associated probability of including population π_i in the selected subset for each i = 1, ..., k, when $\underset{\sim}{\theta}$ is the true parameter.

Definition 3.3.6. A subset selection rule R is locally strongly monotone at point $\underset{\sim}{\theta}_0 \in \Omega_0$ if for each i = 1, ..., k, $\left.\frac{\partial P_i(\underset{\sim}{\theta})}{\partial \underset{\sim}{\theta}_i}\right|_{\underset{\sim}{\theta}=\underset{\sim}{\theta}_0} \geq 0$ and $\left.\frac{\partial P_i(\underset{\sim}{\theta})}{\partial \theta_j}\right|_{\underset{\sim}{\theta}=\underset{\sim}{\theta}_0} \leq 0$ for all $j \neq i$.

Note that by definition of $P_i(\underset{\sim}{\theta})$,

$$P_i(\underset{\sim}{\theta}) = E_{\underset{\sim}{\theta}}[\alpha_i(\underset{\sim}{\Delta}(r))]$$

(3.3.28)

$$= \sum_{\underset{\sim}{\Delta}(r) \in \mathcal{L}_r} P_{\underset{\sim}{\theta}}(\underset{\sim}{\Delta}(r))\alpha_i(\underset{\sim}{\Delta}(r))$$

$$= \sum_{\underset{\sim}{\Delta}(r) \in \mathcal{L}_r} \left[\sum_{\underset{\sim}{\Delta} \in \mathcal{L}(\underset{\sim}{\Delta}(r))} P_{\underset{\sim}{\theta}}(\underset{\sim}{\Delta})\right]\alpha_i(\underset{\sim}{\Delta}(r))$$

where $P_{\underset{\sim}{\theta}}(\underset{\sim}{\Delta})$ is defined in (3.3.1).

Under the condition A, for each $j = 1, \ldots, k$, we have

$$\left.\frac{\partial P_{\underset{\sim}{\theta}}(\underset{\sim}{\Delta})}{\partial \theta_j}\right|_{\underset{\sim}{\theta}=\underset{\sim}{\theta}_0} = \sum_{\substack{m=1 \\ \Delta_m=j}}^{N} B_m(\underset{\sim}{\theta}_0) \times (n!)^k$$

(3.3.29)

$$= A_j^*(\underset{\sim}{\Delta},\underset{\sim}{\theta}_0)(n!)^k$$

where $B_m(\underset{\sim}{\theta}_0)$, $A_j^*(\underset{\sim}{\Delta},\underset{\sim}{\theta}_0)$ are defined in (3.3.11) and (3.3.10), respectively.

Therefore, we have

$$(3.3.30)\quad \left.\frac{\partial P_i(\underset{\sim}{\theta})}{\partial \theta_i}\right|_{\underset{\sim}{\theta}=\underset{\sim}{\theta}_0} = (n!)^k \sum_{\underset{\sim}{\Delta}(r)\in \mathcal{L}_r} \left[\sum_{\underset{\sim}{\Delta}\in \mathcal{L}(\underset{\sim}{\Delta}(r))} A_i^*(\underset{\sim}{\Delta},\underset{\sim}{\theta}_0)\right] \alpha_i(\underset{\sim}{\Delta}(r)),$$

and

$$(3.3.31)\quad \left.\frac{\partial P_i(\underset{\sim}{\theta})}{\partial \theta_j}\right|_{\underset{\sim}{\theta}=\underset{\sim}{\theta}_0} = (n!)^k \sum_{\underset{\sim}{\Delta}(r)\in \mathcal{L}_r} \left[\sum_{\Delta\in \mathcal{L}(\underset{\sim}{\Delta}(r))} A_j^*(\underset{\sim}{\Delta},\underset{\sim}{\theta}_0)\right] \alpha_i(\underset{\sim}{\Delta}(r))$$

$$\forall\ j \neq i.$$

The following lemmas are needed for deriving the locally strong monotonicity of the subset selection rule R_1.

Lemma 3.3.7. Let $g\in G$ and $\bar{g}\in\bar{G}$ and $\bar{\bar{g}}\in\bar{\bar{G}}$ are that induced from g. Then, for any $\underset{\sim}{\Delta}\in \mathcal{L}$, $\underset{\sim}{\Delta}(r)\in \mathcal{L}_r$, we have

(1) $A^*_{gi}(\overline{g}\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0) = A^*_i(\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0)$ and

(2) $T^*_{gi}(\overline{\overline{g}}\underset{\sim}{\Delta}(r), \underset{\sim}{\theta}_0) = T^*_i(\underset{\sim}{\Delta}(r), \underset{\sim}{\theta}_0)$.

Proof: (1) By definition of $A^*_i(\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0)$, we have

$$A^*_i(\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0) = \sum_{\substack{j=1 \\ \Delta_j=i}}^{N} B_j(\underset{\sim}{\theta}_0)$$

$$= \sum_{\substack{j=1 \\ g\Delta_j=gi}}^{N} B_j(\underset{\sim}{\theta}_0) \tag{3.3.32}$$

$$= \sum_{\substack{j=1 \\ (\overline{g}\underset{\sim}{\Delta})_j=gi}}^{N} B_j(\underset{\sim}{\theta}_0)$$

$$= A^*_{gi}(\overline{g}\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0)$$

This completes the proof of part (1).

(2) By the definition of $T^*_i(\underset{\sim}{\Delta}(r), \underset{\sim}{\theta}_0)$ and the result of part (1), we have

$$T^*_{gi}(\overline{\overline{g}}\underset{\sim}{\Delta}(r), \underset{\sim}{\theta}_0) = \frac{1}{|L(\overline{\overline{g}}\underset{\sim}{\Delta}(r))|} \sum_{\underset{\sim}{\Delta}\in L(\overline{\overline{g}}\underset{\sim}{\Delta}(r))} A^*_{gi}(\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0)$$

$$= \frac{1}{|L(\underset{\sim}{\Delta}(r))|} \sum_{\overline{g}\underset{\sim}{\Delta}\in L(\overline{\overline{g}}\underset{\sim}{\Delta}(r))} A^*_{gi}(\overline{g}\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0)$$

$$= \frac{1}{|L(\underset{\sim}{\Delta}(r))|} \sum_{\underset{\sim}{\Delta}\in L(\underset{\sim}{\Delta}(r))} A^*_{gi}(\overline{g}\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0) \tag{3.3.33}$$

$$= \frac{1}{|\mathcal{L}(\underset{\sim}{\Delta}(r))|} \sum_{\underset{\sim}{\Delta}\in\mathcal{L}(\underset{\sim}{\Delta}(r))} A_i^*(\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0)$$

$$= T_i^*(\underset{\sim}{\Delta}(r), \underset{\sim}{\theta}_0).$$

This completes the proof of part (2).

Now, we see that for each $i = 1, \ldots, k$, $A_i^*(\underset{\sim}{\Delta},\underset{\sim}{\theta}_0)$ depends on $\underset{\sim}{\Delta}$ only through whether $\Delta_j = i$ or not for each $j = 1, \ldots, N$, and when $\Delta_j \neq i$, then $A_i^*(\underset{\sim}{\Delta},\underset{\sim}{\theta}_0)$ is independent of the value of Δ_j. Similarly, $T_i^*(\underset{\sim}{\Delta}(r), \underset{\sim}{\theta}_0)$ depends on $\underset{\sim}{\Delta}(r)$ only through whether $\Delta_j = i$ or not for each $j = 1, \ldots, r$, and when $\Delta_j \neq i$, then $T_i^*(\underset{\sim}{\Delta}(r), \underset{\sim}{\theta}_0)$ is independent of the value of Δ_j. Thus, for the subset selection rule R_1, $\alpha_i(\underset{\sim}{\Delta}(r))$ depends on $\underset{\sim}{\Delta}(r)$ only through whether $\Delta_j = i$ or not for each $j = 1, \ldots, r$.

Let $g\in G(i)$. Since g does not change the position of index i, therefore, for each $\underset{\sim}{\Delta}(r) \in \mathcal{L}_r$, $\alpha_i(\bar{\bar{g}}\underset{\sim}{\Delta}(r)) = \alpha_i(\underset{\sim}{\Delta}(r))$ where $\bar{\bar{g}}\in\bar{\bar{G}}$ is induced from g. Now, according to the value of $\alpha_i(\underset{\sim}{\Delta}(r))$, the set $\mathcal{L}_r$ can be partitioned into three classes, say,

$\mathcal{L}_r = \mathcal{L}_r^i(0) \cup \mathcal{L}_r^i(1) \cup \mathcal{L}_r^i(\rho(\underset{\sim}{\theta}_0))$ where

$$(3.3.34)\quad \mathcal{L}_r^i(\beta) = \{\underset{\sim}{\Delta}(r) \in \mathcal{L}_r | \alpha_i(\underset{\sim}{\Delta}(r)) = \beta\}$$

for $\beta = 0$, 1 or $\rho(\underset{\sim}{\theta}_0)$.

Lemma 3.3.8. Let $g \in G(i)$ and $\bar{\bar{g}} \in \bar{\bar{G}}$ be the one induced from g. Then

$$\bar{\bar{g}}(\mathcal{L}_r^i(\beta)) = \mathcal{L}_r^i(\beta) \quad \text{for each } \beta = 0, 1 \text{ or } \rho(\underset{\sim}{\theta}_0).$$

Proof: For each β, let $\underset{\sim}{\Delta}(r) \in \mathcal{L}_r^i(\beta)$. Then $\alpha_i(\underset{\sim}{\Delta}(r)) = \beta$ and so $\alpha_i(\bar{\bar{g}}\underset{\sim}{\Delta}(r)) = \beta$ since $g \in G(i)$. Therefore $\bar{\bar{g}}\underset{\sim}{\Delta}(r) \in \mathcal{L}_r^i(\beta)$. That is, $\bar{\bar{g}}(\mathcal{L}_r^i(\beta)) \sqsubset \mathcal{L}_r^i(\beta)$. Also, $\bar{\bar{g}}\mathcal{L}_r = \mathcal{L}_r$. Thus, if $\bar{\bar{g}}\mathcal{L}_r^i(\beta) \underset{\neq}{\sqsubset} \mathcal{L}_r^i(\beta)$ for some β, we then have $\bar{\bar{g}}\mathcal{L}_r \underset{\neq}{\sqsubset} \mathcal{L}_r$ which is a contradiction. Therefore, $\bar{\bar{g}}(\mathcal{L}_r^i(\beta)) = \mathcal{L}_r^i(\beta)$ for each $\beta = 0, 1$ or $\rho(\underset{\sim}{\theta}_0)$.

Lemma 3.3.9. For each fixed i and $m \neq i$, $j \neq i$ and $m \neq j$, we have

$$\sum_{\underset{\sim}{\Delta}(r) \in \mathcal{L}_r^i(\beta)} \left[\sum_{\underset{\sim}{\Delta} \in \mathcal{L}(\underset{\sim}{\Delta}(r))} A_j^*(\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0) \right] \alpha_i(\underset{\sim}{\Delta}(r))$$

(3.3.35)

$$= \sum_{\underset{\sim}{\Delta}(r) \in \mathcal{L}_r^i(\beta)} \left[\sum_{\underset{\sim}{\Delta} \in \mathcal{L}(\underset{\sim}{\Delta}(r))} A_m^*(\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0) \right] \alpha_i(\underset{\sim}{\Delta}(r))$$

for each $\beta = 0, 1$ or $\rho(\underset{\sim}{\theta}_0)$.

Proof: Let $g \in G(i)$ and satisfies that $gm = j$, $gj = m$. Then,

$$\sum_{\underset{\sim}{\Delta}(r) \in \mathcal{L}_r^i(\beta)} \left[\sum_{\underset{\sim}{\Delta} \in \mathcal{L}(\underset{\sim}{\Delta}(r))} A_j^*(\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0) \right] \alpha_i(\underset{\sim}{\Delta}(r))$$

$$= \beta \sum_{\underset{\sim}{\Delta}(r) \in \mathcal{L}_r^i(\beta)} \left[\sum_{\underset{\sim}{\Delta} \in \mathcal{L}(\underset{\sim}{\Delta}(r))} A_{gj}^*(\bar{g}\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0) \right]$$

(3.3.36)

$$= \beta \sum_{\underset{\sim}{\Delta}(r)\in \mathcal{L}_r^i(\beta)} \left[\sum_{\underset{\sim}{\Delta}\in \mathcal{L}(\underset{\sim}{\Delta}(r))} A_m^*(\bar{g}\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0) \right]$$

$$= \beta \sum_{\underset{\sim}{\Delta}(r)\in \mathcal{L}_r^i(\beta)} \left[\sum_{\underset{\sim}{\Delta}\in \mathcal{L}(\bar{\bar{g}}\underset{\sim}{\Delta}(r))} A_m^*(\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0) \right]$$

$$= \beta \sum_{\underset{\sim}{\Delta}(r)\in \bar{\bar{g}}\mathcal{L}_r^i(\beta)} \left[\sum_{\underset{\sim}{\Delta}\in \mathcal{L}(\underset{\sim}{\Delta}(r))} A_m^*(\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0) \right]$$

$$= \beta \sum_{\underset{\sim}{\Delta}(r)\in \mathcal{L}_r^i(\beta)} \left[\sum_{\underset{\sim}{\Delta}\in \mathcal{L}(\underset{\sim}{\Delta}(r))} A_m^*(\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0) \right]$$

This completes the proof of Lemma 3.3.9.

The following corollary is a direct application of Lemma 3.3.9.

Corollary 3.3.10. For each fixed i and $m \neq i$, $j \neq i$, we have

$$(3.3.37) \quad \left.\frac{\partial P_i(\underset{\sim}{\theta})}{\partial \theta_j}\right|_{\underset{\sim}{\theta}=\underset{\sim}{\theta}_0} = \left.\frac{\partial P_i(\underset{\sim}{\theta})}{\partial \theta_m}\right|_{\underset{\sim}{\theta}=\underset{\sim}{\theta}_0}.$$

Theorem 3.3.11. Suppose that the density function $f(x,\theta)$ satisfies the Condition A. Then, the subset selection rule R_1 is locally strongly monotone at each $\underset{\sim}{\theta}_0 \in \Omega_0$.

Proof: By Corollary 3.3.10, for each $m \neq i$, we have

$$\left.\frac{\partial P_i(\underset{\sim}{\theta})}{\partial \theta_m}\right|_{\underset{\sim}{\theta}=\underset{\sim}{\theta}_0} = \frac{1}{k-1} \sum_{\substack{j=1 \\ j\neq i}}^{k} \left.\frac{\partial P_i(\underset{\sim}{\theta})}{\partial \theta_j}\right|_{\underset{\sim}{\theta}=\underset{\sim}{\theta}_0}$$

$$(3.3.38) \qquad = \frac{(n!)^k}{k-1} \sum_{\substack{j=1 \\ j\neq i}}^{k} \sum_{\underset{\sim}{\Delta}(r)\in \mathcal{L}_r} \left[\sum_{\underset{\sim}{\Delta}\in \mathcal{L}(\underset{\sim}{\Delta}(r))} A_j^*(\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0) \right] \alpha_i(\underset{\sim}{\Delta}(r))$$

$$= \frac{(n!)^k}{k-1} \sum_{\underset{\sim}{\Delta}(r)\in \mathcal{L}_r} \left[\sum_{\underset{\sim}{\Delta}\in \mathcal{L}(\underset{\sim}{\Delta}(r))} \sum_{\substack{j=1 \\ j\neq i}}^{k} A_j^*(\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0) \right] \alpha_i(\underset{\sim}{\Delta}(r))$$

$$= -\frac{(n!)^k}{k-1} \sum_{\underset{\sim}{\Delta}(r)\in \mathcal{L}_r} \left[\sum_{\underset{\sim}{\Delta}\in \mathcal{L}(\underset{\sim}{\Delta}(r))} A_i^*(\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0) \right] \alpha_i(\underset{\sim}{\Delta}(r))$$

$$= -\frac{1}{k-1} \frac{\partial}{\partial \theta_i} P_i(\underset{\sim}{\theta}) \Big|_{\underset{\sim}{\theta}=\underset{\sim}{\theta}_0},$$

where the last second equality is due to Lemma 3.3.1.

Therefore, it suffices to prove that $\frac{\partial}{\partial \theta_i} P_i(\underset{\sim}{\theta}) \Big|_{\underset{\sim}{\theta}=\underset{\sim}{\theta}_0} \geq 0$ for each $\underset{\sim}{\theta}_0 \in \Omega_0$. Now,

$$\sum_{\underset{\sim}{\Delta}(r)\in \mathcal{L}_r} |\mathcal{L}(\underset{\sim}{\Delta}(r))| T_i^*(\underset{\sim}{\Delta}(r), \underset{\sim}{\theta}_0)$$

$$= \sum_{\underset{\sim}{\Delta}(r)\in \mathcal{L}_r} \sum_{\Delta\in \mathcal{L}(\underset{\sim}{\Delta}(r))} A_i^*(\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0)$$

$$= \sum_{\underset{\sim}{\Delta}\in \mathcal{L}} A_i^*(\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0) \tag{3.3.39}$$

$$= \sum_{\underset{\sim}{\Delta}\in \mathcal{L}} \sum_{j=1}^{N} B_j(\underset{\sim}{\theta}_0) I_{\{i\}}(\Delta_j)$$

$$= \sum_{j=1}^{N} \left[B_j(\underset{\sim}{\theta}_0) \sum_{\underset{\sim}{\Delta}\in \mathcal{L}} I_{\{i\}}(\Delta_j) \right]$$

$$= \frac{(N-1)!}{(n!)^{k-1}(n-1)!} \sum_{j=1}^{N} B_j(\underset{\sim}{\theta}_0)$$

$$= 0,$$

since $\sum_{j=1}^{N} B_j(\underset{\sim}{\theta}_0) = 0$ which is due to Lemma 3.3.1 under Condition A.

Then, by (3.3.25), (3.3.38) and (3.3.39), we see that

$$(3.3.40) \qquad \sum_{\underset{\sim}{\Delta}(r)\in \boldsymbol{L}_r} \left[\sum_{\underset{\sim}{\Delta}\in \boldsymbol{L}(\underset{\sim}{\Delta}(r))} A_i^*(\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0) \right] \alpha_i(\underset{\sim}{\Delta}(r)) \geq 0$$

Therefore $\left.\frac{\partial P_i(\underset{\sim}{\theta})}{\partial \theta_i}\right|_{\underset{\sim}{\theta}=\underset{\sim}{\theta}_0} \geq 0$. Hence, the subset selection rule R_1 is locally strongly monotone at each $\underset{\sim}{\theta}_0 \in \Omega_0$.

Remarks 3.3.12. (1) Note that when r = N, that is, in the complete rank configuration case, this locally optimal subset selection rule R_1 turns out to be the one studied by Gupta, Huang and Nagel (1979).

(2) This locally optimal subset selection rule R_1 is based on the weighted rank sum $B_j(\underset{\sim}{\theta}_0) = \frac{1}{(j-1)!(N-j)!} \int_0^1 u^{j-1} (1-u)^{N-j} \phi(u, f, \theta_0) du$ where $\phi(u, f, \theta) = \dot{f}(F^{-1}(u, \theta), \theta)/f(F^{-1}(u, \theta), \theta)$. In general, $\phi(u, f, \theta)$ depends on θ. However, it is independent of θ if θ is a location parameter (see Gupta, Huang and Nagel (1979)). In this situation, the value $B_j(\underset{\sim}{\theta}_0)$ is independent of $\underset{\sim}{\theta}_0$. Therefore, the two constants $c(\underset{\sim}{\theta}_0)$ and $\rho(\underset{\sim}{\theta}_0)$, which are used to determine the rule R_1, are also independent of $\underset{\sim}{\theta}_0$ for each fixed p* value.

(3) Suppose that $\theta > 0$ is a scale parameter, that is, $f(x, \theta) = \theta h(\theta(x - \mu))$ for some function $h(\cdot)$. Let $\theta_1, \theta_2 > 0$ such that $\theta_2 = \beta\theta_1$. Then, $\phi(u, f, \theta_2) = \frac{1}{\beta}\phi(u, f, \theta_1)$. Therefore, $B_j(\underset{\sim}{\theta}_2) = \frac{1}{\beta}B_j(\underset{\sim}{\theta}_1)$ for each $j = 1, \ldots, N$, where $\underset{\sim}{\theta}_i = (\theta_i, \ldots, \theta_i) \in \Omega_0$, $i = 1, 2$. In this situation, for each fixed p* value, we have $c(\underset{\sim}{\theta}_2) = \frac{1}{\beta}c(\underset{\sim}{\theta}_1)$ and $\rho(\underset{\sim}{\theta}_2) = \rho(\underset{\sim}{\theta}_1)$.

Huang and Panchapakesan (1982) also derived a subset selection rule, say R_{HP}, based on the complete rank configurations, which can be represented as follows:

$$\alpha_i(\underset{\sim}{\Delta}) = \begin{cases} 1 & \text{if } A_i^*(\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0) > V(\theta_0) + D \\ \rho & \text{if } A_i^*(\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0) = V(\theta_0) + D \\ 0 & \text{if } A_i^*(\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0) < V(\theta_0) + D \end{cases} \tag{3.3.41}$$

where D and $\rho (0 \leq \rho < 1)$ are chosen so that

$$P_{\underset{\sim}{\theta}_0}\{A_i^*(\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0) > V(\theta_0) + D\} + \rho P_{\underset{\sim}{\theta}_0}\{A_i^*(\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0) = V(\theta_0) + D\} = P^* \quad \text{and } V(\theta_0) + D > 0. \tag{3.3.42}$$

The rule R_{HP} is always locally strongly monotone provided the constants D and ρ satisfying (3.3.42) exist. However, as pointed out by themselves, it is possible that the D and ρ satisfying (3.3.42) may not exist. In such a case, the rule R_{HP} selects the empty subset.

The following example indicates that the rule R_{HP} always selects the empty subset when $p^* > \frac{1}{2}$.

Example 3.3.13. Let k (≥ 2) and n be positive integers and let $N = kn$. Let $f(x,\theta)$ be the logistic density $f(x,\theta) = e^{-(x-\theta)}/[1 + e^{-(x-\theta)}]^2$, $-\infty < x < \infty$, $-\infty < \theta < \infty$. It is clear that $f(x,\theta)$ satisfies the Condition A. Then by Lemma 3.3.1, $V(\theta_0) = 0$ for all $\underset{\sim}{\theta}_0 \in \Omega_0$. Also, $\phi(u, f, \theta) = 2u - 1$, which leads to equally spaced scores and

$$(3.3.43)\quad B_j(\underset{\sim}{\theta}_0) = \frac{2j}{(N+1)!} - \frac{1}{N!} .$$

Note that $B_j(\underset{\sim}{\theta}_0) + B_{N+1-j}(\underset{\sim}{\theta}_0) = 0$ for each $j = 1, \ldots, N$. Therefore, for each $\underset{\sim}{\Delta} = (\Delta_1, \ldots, \Delta_N) \in \mathcal{L}$, let $\underset{\sim}{\Delta}^1 = (\Delta_1^1, \ldots, \Delta_N^1)$ where $\Delta_j^1 = \Delta_{N+1-j}$ for each $j = 1, \ldots, N$. Then $\underset{\sim}{\Delta}^1 \in \mathcal{L}$. By (3.3.43) and the definition of $A_i^*(\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0)$, we have $A_i^*(\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0) + A_i^*(\underset{\sim}{\Delta}^1, \underset{\sim}{\theta}_0) = 0$ for all $\underset{\sim}{\Delta} \in \mathcal{L}$, $\underset{\sim}{\theta}_0 \in \Omega_0$, which implies that $P_{\underset{\sim}{\theta}_0}\{A_i^*(\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0) > 0\} \leq \frac{1}{2}$ for all $\underset{\sim}{\theta}_0 \in \Omega_0$. Hence, for $p^* > \frac{1}{2}$, there exist no D and ρ ($0 \leq \rho < 1$) such that (3.3.42) is satisfied.

However, for the subset selection rule R_1, the corresponding two constants $c(\underset{\sim}{\theta}_0)$ and $\rho(\underset{\sim}{\theta}_0)$ always exist when $p^* \in (\frac{1}{k}, 1)$, and the rule R_1 is always locally strongly monotone which is guaranteed by Theorem 3.3.11.

3.4. Derivation of a Locally Optimal Rule for Goal 2

In this section, we also assume that the density function $f(x,\theta)$ satisfies the Condition A of (3.2.5). The selection problem is studied for the unequal sample sizes case. We consider the class of rules R such that

$$(3.4.1) \quad P_1(\underset{\sim}{\theta}_0) = \cdots = P_k(\underset{\sim}{\theta}_0) = P^* \quad \text{for all } \underset{\sim}{\theta}_0 \in \Omega_0.$$

Among these rules, we will find the one which is locally optimal in the sense that it maximizes

$$(3.4.2) \quad \sum_{i=1}^{k} \frac{\partial}{\partial \theta_i} \log P_{\underset{\sim}{\theta}}\{CS | \underset{\sim}{\theta} \in \Omega_i\}\Big|_{\underset{\sim}{\theta}=\underset{\sim}{\theta}_0 \in \Omega_0}.$$

The requirement (3.4.1) is reasonable, since for $\underset{\sim}{\theta}_0 = (\theta_{01}, \ldots, \theta_{0k}) \in \Omega_0$, $\theta_{01} = \cdots = \theta_{0k} = \theta_0$, and in this case, it can be verified that the distribution of the (censored) ranks is independent of the underlying distribution, and so, is also independent of the common value of the parameter θ_0.

Note that $P_{\underset{\sim}{\theta}}\{CS | \underset{\sim}{\theta} \in \Omega_i\} = P_i(\underset{\sim}{\theta} | \underset{\sim}{\theta} \in \Omega_i)$. Then,

$$\sum_{i=1}^{k} \frac{\partial}{\partial \theta_i} \log P_{\underset{\sim}{\theta}}\{CS | \underset{\sim}{\theta} \in \Omega_i\}\Big|_{\underset{\sim}{\theta}=\underset{\sim}{\theta}_0 \in \Omega_0}$$

$$(3.4.3) \quad = \sum_{i=1}^{k} \frac{1}{P_i(\underset{\sim}{\theta}_0)} \frac{\partial}{\partial \theta_i} P_i(\underset{\sim}{\theta} | \underset{\sim}{\theta} \in \Omega_i)\Big|_{\underset{\sim}{\theta}=\underset{\sim}{\theta}_0 \in \Omega_0}$$

$$= \frac{1}{P^*} \sum_{i=1}^{k} \frac{\partial}{\partial \theta_i} P_i(\underset{\sim}{\theta} | \underset{\sim}{\theta} \in \Omega_i)\Big|_{\underset{\sim}{\theta}=\underset{\sim}{\theta}_0 \in \Omega_0}.$$

Thus, if a rule R satisfies that $P_i(\underset{\sim}{\theta}_0) = P^*$ and maximizes $\frac{\partial}{\partial\theta_i} P_i(\underset{\sim}{\theta}|\underset{\sim}{\theta} \in \Omega_i)\Big|_{\underset{\sim}{\theta}=\underset{\sim}{\theta}_0}$ for all i = 1, ..., k, it also maximizes (3.4.2).

Now, for any $\underset{\sim}{\theta} \in \Omega$,

$$P_i(\underset{\sim}{\theta}) = E_{\underset{\sim}{\theta}}[\alpha_i(\underset{\sim}{\Delta}(r))]$$

$$(3.4.4) \qquad = \sum_{\underset{\sim}{\Delta}(r)\in\mathcal{L}_r} P_{\underset{\sim}{\theta}}(\underset{\sim}{\Delta}(r))\alpha_i(\underset{\sim}{\Delta}(r))$$

$$= \sum_{\underset{\sim}{\Delta}(r)\in\mathcal{L}_r} \left[\sum_{\underset{\sim}{\Delta}\in\mathcal{L}(\underset{\sim}{\Delta}(r))} P_{\underset{\sim}{\theta}}(\underset{\sim}{\Delta})\right]\alpha_i(\underset{\sim}{\Delta}(r)).$$

With an argument similar to that of (3.3.3), we have

$$(3.4.5) \quad P_{\underset{\sim}{\theta}}(\underset{\sim}{\Delta}) = \left(\prod_{j=1}^{k} n_j!\right) \times \left[A_0(\theta_0) + \sum_{j=1}^{k} (\theta_j - \theta_0)A_j(\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0, \underset{\sim}{\theta})\right],$$

where $A_0(\theta_0)$ and $A_j(\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0, \underset{\sim}{\theta})$ are defined as before and $\underset{\sim}{\theta}_0 = (\theta_0, \ldots, \theta_0) \in \Omega_0$. Hence,

$$P_i(\underset{\sim}{\theta}) = \left(\prod_{j=1}^{k} n_j!\right) \sum_{\underset{\sim}{\Delta}(r)\in\mathcal{L}_r} \left[\sum_{\underset{\sim}{\Delta}\in\mathcal{L}(\underset{\sim}{\Delta}(r))} \left[A_0(\theta_0) + \sum_{j=1}^{k} (\theta_j - \theta_0)A_j(\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0, \underset{\sim}{\theta})\right]\right]\alpha_i(\underset{\sim}{\Delta}(r)).$$

Under the assumptions of Condition A, we have

$$\frac{\partial}{\partial\theta_i} P_i(\underset{\sim}{\theta}|\underset{\sim}{\theta} \in \Omega_i)\Big|_{\underset{\sim}{\theta}=\underset{\sim}{\theta}_0}$$

$$= \left(\prod_{j=1}^{k} n_j!\right) \sum_{\underset{\sim}{\Delta}(r)\in\mathcal{L}_r} \left[\sum_{\underset{\sim}{\Delta}\in\mathcal{L}(\underset{\sim}{\Delta}(r))} A_i^*(\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0)\right]\alpha_i(\underset{\sim}{\Delta}(r)) \tag{3.4.7}$$

$$= \left(\prod_{j=1}^{k} n_j!\right) \sum_{\underset{\sim}{\Delta}(r)\in\mathcal{L}_r} [|\mathcal{L}(\underset{\sim}{\Delta}(r))|T_i^*(\underset{\sim}{\Delta}(r), \underset{\sim}{\theta}_0)]\alpha_i(\underset{\sim}{\Delta}(r))$$

$$= N!\, E_{\underset{\sim}{\theta}_0}[T_i^*(\underset{\sim}{\Delta}(r), \underset{\sim}{\theta}_0)\, \alpha_i(\underset{\sim}{\Delta}(r))].$$

Thus, we define a subset selection rule $R_2(\underset{\sim}{\Delta}(r)) = (\alpha_1(\underset{\sim}{\Delta}(r)), \ldots, \alpha_k(\underset{\sim}{\Delta}(r)))$ as follows:

$$\alpha_i(\underset{\sim}{\Delta}(r)) = \begin{cases} 1 & \text{if } T_i^*(\underset{\sim}{\Delta}(r), \underset{\sim}{\theta}_0) > c_i(\underset{\sim}{\theta}_0); \\ \rho_i(\underset{\sim}{\theta}_0) & \text{if } T_i^*(\underset{\sim}{\Delta}(r), \underset{\sim}{\theta}_0) = c_i(\underset{\sim}{\theta}_0); \\ 0 & \text{if } T_i^*(\underset{\sim}{\Delta}(r), \underset{\sim}{\theta}_0) < c_i(\underset{\sim}{\theta}_0); \end{cases} \tag{3.4.8}$$

where constants $c_i(\underset{\sim}{\theta}_0)$ and $\rho_i(\underset{\sim}{\theta}_0)$, $(0 \le \rho_i(\underset{\sim}{\theta}_0) < 1)$, $i = 1, \ldots, k$, are determined such that

$$P_{\underset{\sim}{\theta}_0}\left\{T_i^*(\underset{\sim}{\Delta}(r),\underset{\sim}{\theta}_0) > c_i(\underset{\sim}{\theta}_0)\right\} + \rho_i(\underset{\sim}{\theta}_0)P_{\underset{\sim}{\theta}_0}\left\{T_i^*(\underset{\sim}{\Delta}(r),\underset{\sim}{\theta}_0) = c_i(\underset{\sim}{\theta}_0)\right\} = p^*. \tag{3.4.9}$$

Theorem 3.4.1. Suppose that the density function $f(x, \theta)$ satisfies the condition A. Then, among all subset selection rules satisfying $P_1(\underset{\sim}{\theta}_0) = \cdots = P_k(\underset{\sim}{\theta}_0) = P^*$, the subset selection rule R_2 is locally optimal in the sense that it maximizes $\sum_{i=1}^{k} \frac{\partial}{\partial \theta_i} \log P_{\underset{\sim}{\theta}}\{CS | \underset{\sim}{\theta} \in \Omega_i\}\Big|_{\underset{\sim}{\theta}=\underset{\sim}{\theta}_0 \in \Omega_0}$.

Proof: This theorem follows immediately from (3.4.7), (3.4.8) and by applying Neyman-Pearson lemma.

Local Monotonicity of the Subset Selection Rule R_2

Let R be a subset selection rule and $P_i(\underset{\sim}{\theta})$ be the associated probability of including population π_i in the selected subset for each $i = 1, \ldots, k$, when $\underset{\sim}{\theta}$ is the true parameter.

Definition 3.4.2. A subset selection rule R is locally weakly monotone at point $\underset{\sim}{\theta}_0 \in \Omega_0$ if $\frac{\partial P_i(\underset{\sim}{\theta})}{\partial \theta_i}\Big|_{\underset{\sim}{\theta}=\underset{\sim}{\theta}_0} \geq 0$ and $\sum_{\substack{j=1 \\ j \neq i}}^{k} \frac{\partial P_i(\underset{\sim}{\theta})}{\partial \theta_j}\Big|_{\underset{\sim}{\theta}=\underset{\sim}{\theta}_0} \leq 0$

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

We first note that under the Condition A, Lemma 3.3.1 still holds, that is, $\sum_{j=1}^{N} B_j(\underset{\sim}{\theta}_0) = 0$, $\underset{\sim}{\theta}_0 \in \Omega_0$, in the unequal sample sizes case.

The following lemma is similar to that of Lemma 3.3.7.

Lemma 3.4.3. (Unequal sample sizes case) Suppose that the density function $f(x, \theta)$ satisfies the condition A. Then, for each $i = 1, \ldots, k$, $\underset{\sim}{\theta}_0 \in \Omega_0$, we have

$$(3.4.10)\qquad \sum_{\underset{\sim}{\Delta}(r)\in\mathcal{L}_r} |\mathcal{L}(\underset{\sim}{\Delta}(r))| T_i^*(\underset{\sim}{\Delta}(r), \underset{\sim}{\theta}_0) = \sum_{\underset{\sim}{\Delta}\in\mathcal{L}_r} \sum_{\underset{\sim}{\Delta}\in\mathcal{L}(\underset{\sim}{\Delta}(r))} A_i^*(\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0) = 0.$$

Proof: For completeness, the proof is given as follows.

Along the line of the argument of (3.3.39), we have

$$\sum_{\underset{\sim}{\Delta}(r)\in\mathcal{L}_r} \sum_{\underset{\sim}{\Delta}\in\mathcal{L}(\underset{\sim}{\Delta}(r))} A_i^*(\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0)$$

$$(3.4.11)\qquad = \sum_{j=1}^{N} B_j(\underset{\sim}{\theta}_0) \sum_{\underset{\sim}{\Delta}\in\mathcal{L}} I_{\{i\}}(\Delta_j)$$

$$= \sum_{j=1}^{N} B_j(\underset{\sim}{\theta}_0) \frac{(N-1)!}{(n_i-1)! \prod_{\substack{m=1 \\ m\neq i}}^{k} n_m!}$$

$$= 0$$

since $\sum_{j=1}^{N} B_j(\underset{\sim}{\theta}_0) = 0$ under the Condition A.

Theorem 3.4.4. Suppose that the density function $f(x, \theta)$ satisfies the Condition A. Then the subset selection rule R_2 is locally weakly monotone at each $\underset{\sim}{\theta}_0 \in \Omega_0$.

Proof: From (3.4.7)

$$(3.4.12)\quad \frac{\partial P_i(\underset{\sim}{\theta})}{\partial \theta_i}\Big|_{\underset{\sim}{\theta}=\underset{\sim}{\theta}_0} = \left(\prod_{j=1}^{k} n_j!\right) \sum_{\underset{\sim}{\Delta}(r)\in \mathcal{L}_r} |\mathcal{L}(\underset{\sim}{\Delta}(r))| T_i^*(\underset{\sim}{\Delta}(r), \underset{\sim}{\theta}_0)\alpha_i(\underset{\sim}{\Delta}(r)).$$

Then by (3.4.8) and Lemma 3.4.3, we conclude that $\frac{\partial P_i(\underset{\sim}{\theta})}{\partial \theta_i}\Big|_{\underset{\sim}{\theta}=\underset{\sim}{\theta}_0} \geq 0$

where the strict inequality holds if $T_i^*(\underset{\sim}{\Delta}(r), \underset{\sim}{\theta}_0) \neq 0$ for some $\underset{\sim}{\Delta}(r) \in \mathcal{L}_r$.

Next, from (3.4.6), for each i = 1, ..., k, we have

$$\begin{aligned}
&\sum_{\substack{j=1\\ j\neq i}}^{k} \frac{\partial P_i(\underset{\sim}{\theta})}{\partial \theta_j}\Big|_{\underset{\sim}{\theta}=\underset{\sim}{\theta}_0} \\
&= \sum_{\substack{j=1\\ j\neq i}}^{k} \sum_{\underset{\sim}{\Delta}(r)\in \mathcal{L}_r} \left[\sum_{\underset{\sim}{\Delta}\in \mathcal{L}(\underset{\sim}{\Delta}(r))} A_j^*(\underset{\sim}{\Delta}, \theta_0)\right]\alpha_i(\underset{\sim}{\Delta}(r)) \prod_{j=1}^{k} n_j! \\
&= \left(\prod_{j=1}^{k} n_j!\right) \sum_{\underset{\sim}{\Delta}(r)\in \mathcal{L}_r}\left[\sum_{\underset{\sim}{\Delta}\in \mathcal{L}(\underset{\sim}{\Delta}(r))} \sum_{\substack{j=1\\ j\neq i}}^{k} A_j^*(\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0)\right]\alpha_i(\underset{\sim}{\Delta}(r)) \qquad (3.4.13)\\
&= -\left(\prod_{j=1}^{k} n_j!\right) \sum_{\underset{\sim}{\Delta}(r)\in \mathcal{L}_r} \left[\sum_{\underset{\sim}{\Delta}\in \mathcal{L}(\underset{\sim}{\Delta}(r))} A_i^*(\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0)\right]\alpha_i(\underset{\sim}{\Delta}(r)) \\
&= -\left(\prod_{j=1}^{k} n_j!\right) \sum_{\underset{\sim}{\Delta}(r)\in \mathcal{L}_r} |\mathcal{L}(\underset{\sim}{\Delta}(r))| T_i^*(\underset{\sim}{\Delta}(r), \underset{\sim}{\theta}_0)\alpha_i(\underset{\sim}{\Delta}(r)) \\
&= -\frac{\partial P_i(\underset{\sim}{\theta})}{\partial \theta_i}\Big|_{\underset{\sim}{\theta}=\underset{\sim}{\theta}_0} \\
&\leq 0,
\end{aligned}$$

where the third equality is due to the fact that $\sum_{j=1}^{k} A_j^*(\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0) = 0$ for each $\underset{\sim}{\Delta} \in \mathcal{L}$, $\underset{\sim}{\theta}_0 \in \Omega_0$.

Therefore, the subset selection rule R_2 is locally weakly monotone at each $\underset{\sim}{\theta}_0 \in \Omega_0$.

3.5. A Generalization

In the preceding sections, we studied some locally optimal subset selection rules, based on the joint type II censored rank configurations. We now turn our consideration to the case when a specified censoring scheme C makes $\underset{\sim}{x} = (x_1, \ldots, x_N)$ (or $\underset{\sim}{\Delta}_{\underset{\sim}{x}}$) partially observable. First, we introduce a definition as follows:

Let $\mathcal{L} = \{\underset{\sim}{\Delta}\}$ denote the set of all rank configurations for the fixed constants $n_1, \ldots, n_k$.

Definition 3.5.1. A censoring scheme C is said to be compatible with the rank configurations $\underset{\sim}{\Delta} \in \mathcal{L}$ if C induces a partition of $\mathcal{L}$.

It is not hard to see that the joint type II censoring introduced previously is compatible with the rank configurations $\underset{\sim}{\Delta} \in \mathcal{L}$.

For a compatible censoring scheme C, we denote the mapping as $C(\underset{\sim}{\Delta}) = \underset{\sim}{\Delta}_C$, $\underset{\sim}{\Delta} \in \mathcal{L}$, and define the following terms

$$\mathcal{L}_C = \{\underset{\sim}{\Delta}_C \mid \exists\, \underset{\sim}{\Delta} \in \mathcal{L} \text{ such that } C(\underset{\sim}{\Delta}) = \underset{\sim}{\Delta}_C\}.$$

$$\mathcal{L}(\underset{\sim}{\Delta}_C) = \{\underset{\sim}{\Delta} \in \mathcal{L} \mid C(\underset{\sim}{\Delta}) = \underset{\sim}{\Delta}_C\}.$$

Now, for each $\underset{\sim}{\Delta}_c \in \mathcal{L}_C$, let $P_{\underset{\sim}{\theta}}(\underset{\sim}{\Delta}_c)$ be the probability that the censored rank configuration $\underset{\sim}{\Delta}_c$ is observed when $\underset{\sim}{\theta}$ is the true parameter. Then,

$$(3.5.1) \qquad P_{\underset{\sim}{\theta}}(\underset{\sim}{\Delta}_c) = \sum_{\underset{\sim}{\Delta} \in \mathcal{L}(\underset{\sim}{\Delta}_c)} P_{\underset{\sim}{\theta}}(\underset{\sim}{\Delta}).$$

Also, for $\underset{\sim}{\theta}_0 \in \Omega_0$, we have

$$(3.5.2) \qquad P_{\underset{\sim}{\theta}_0}(\underset{\sim}{\Delta}_c) = \frac{\prod_{j=1}^{k} n_j!}{N!} |\mathcal{L}(\underset{\sim}{\Delta}_c)|.$$

Let $T(\underset{\sim}{\Delta}_c, \underset{\sim}{\theta}_0, \underset{\sim}{\theta}) = P_{\underset{\sim}{\theta}}(\underset{\sim}{\Delta}_c)/P_{\underset{\sim}{\theta}_0}(\underset{\sim}{\Delta}_c)$. By an argument analogous to that of Section 3.3, we have

$$(3.5.3) \qquad T(\underset{\sim}{\Delta}_c, \underset{\sim}{\theta}_0, \underset{\sim}{\theta}) = 1 + \frac{N!}{|\mathcal{L}(\underset{\sim}{\Delta}_c)|} \sum_{\underset{\sim}{\Delta} \in \mathcal{L}(\underset{\sim}{\Delta}_c)} \sum_{i=1}^{k} (\theta_i - \theta_0) A_i(\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0, \underset{\sim}{\theta}).$$

Then, under the assumption of the Condition A, we have

$$(3.5.4) \qquad \left.\frac{\partial T(\underset{\sim}{\Delta}_c, \underset{\sim}{\theta}_0, \underset{\sim}{\theta})}{\partial \theta_i}\right|_{\underset{\sim}{\theta}=\underset{\sim}{\theta}_0} = \frac{N!}{|\mathcal{L}(\underset{\sim}{\Delta}_c)|} \sum_{\underset{\sim}{\Delta} \in \mathcal{L}(\underset{\sim}{\Delta}_c)} A_i^*(\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0)$$

$$= N!\, T_i^*(\underset{\sim}{\Delta}_c, \underset{\sim}{\theta}_0)$$

where

$$(3.5.5)\quad T_i^*(\underset{\sim}{\Delta}_c, \underset{\sim}{\theta}_0) = \frac{1}{|\mathcal{L}(\underset{\sim}{\Delta}_c)|} \sum_{\underset{\sim}{\Delta}\in\mathcal{L}(\underset{\sim}{\Delta}_c)} A_i^*(\underset{\sim}{\Delta}, \underset{\sim}{\theta}_0).$$

Thus, the likelihood ratio function $T(\underset{\sim}{\Delta}_c, \underset{\sim}{\theta}_0, \underset{\sim}{\theta})$ of the censored rank configuration $\underset{\sim}{\Delta}_c$ is increasing in θ_i in the neighborhood of $\underset{\sim}{\theta}_0$ if $T_i^*(\underset{\sim}{\Delta}_c, \underset{\sim}{\theta}_0) \geq 0$.

Note that $T_i^*(\underset{\sim}{\Delta}_c, \underset{\sim}{\theta}_0) = T_i^*(\underset{\sim}{\Delta}(r), \underset{\sim}{\theta}_0)$ if the censoring scheme C is the joint type II censoring scheme C_r.

Following an argument analogous to that of preceding sections, we can also derive some locally optimal subset selection rules based on the statistics $T_i^*(\underset{\sim}{\Delta}_c, \underset{\sim}{\theta}_0)$. We can see that these derived locally optimal subset selection rules have some properties similar to that of R_1 (in the equal sample sizes case) or R_2 (in the unequal sample sizes case), without any further assumptions. With these properties, it is possible for the experimenter to use a suitable compatible censoring scheme.

BIBLIOGRAPHY

Alam, K. (1973). On a multiple decision rule. Ann. 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 Robbins, H. (1950). The problem of the greater mean. Ann. Math. Statist., 21, 469-487. Correction, 22 (1951), 310.

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

Bechhofer, R. E., Kiefer, J. and Sobel, M. (1968). Sequential Identification and Ranking Procedures. Univ. of Chicago Press, Chicago.

Berger, R. L. (1979). Minimax subset selection for loss measured by subset size. Ann. Statist., 7, 1333-1338.

Berger, R. L. and Gupta, S. S. (1980). Minimax subset selection rules with applications to unequal variance (unequal sample size) problems. Scand. J. Statist., 7, 21-26.

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

Bjørnstad, J. F. (1980). Comparison of three minimax subset selection procedures. Technometrics, 22, 617-620.

Bjørnstad, J. F. (1981). A decision-theoretic approach to subset selection. Commun. Statist.-Theor. Meth., A10, 2411-2433.

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

Deely, J. J. (1965). Multiple decision procedures from an empirical Bayes approach. Ph.D. Thesis (Mimeo. Ser. No. 45), Dept. of Statist., Purdue Univ., West Lafayette, IN.

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

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

Goel, P. K. and Rubin, H. (1977). On selecting a subset containing the best population. Ann. Statist., 5, 969-983.

Gupta, S. S. (1956). On a decision rule for a problem in ranking means. Ph.D. Thesis (Mimeo. Ser. No. 150). Inst. of Statist., Univ. of North Carolina, Chapel Hill, North Carolina.

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

Gupta, S. S. and Hsiao, P. (1983). Empirical Bayes rules for selecting good populations. J. Statist. Plan. Infer., 8, 87-101.

Gupta, S. S. and Hsu, J. C. (1978). On the performance of some subset selection procedures. Commun. Statist.-Simula. Computa., B7, 561-591.

Gupta, S. S. and Huang, D. Y. (1975). On some parametric and non-parametric sequential subset selection procedures. Statistical Inference and Related Topic, Vol. 2 (Ed. M. L. Puri). Academic Press, New York, 101-128.

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

Gupta, S. S., Huang, D. Y. and Nagel, K. (1979). Locally optimal subset selection procedures based on ranks. Optimizing Methods in Statistics, (Ed. J. S. Rustagi). Academic Press, New York, 251-260.

Gupta, S. S. and Kim, W. C. (1980). Γ-minimax and minimax decision rules for comparison of treatments with a control. Recent Developments in Statistical Inference and Data Analysis (Ed. K. Matusita). North Holland, 55-71.

Gupta, S. S. and Leu, L. Y. (1983). On Bayes and empirical Bayes rules for selecting good populations. Technical Report No. 83-37, Dept. of Statist., Purdue Univ., West Lafayette, IN.

Gupta, S. S. and Miescke, K. J. (1984a). On two-stage Bayes selection procedures. Sankhyā Ser. B, 46.

Gupta, S. S. and Miescke, K. J. (1984b). Sequential selection procedures--a decision-theoretic approach. Ann. Statist., 12, 336-350.

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), Oxford University Press, Oxford, 491-514.

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

Gupta, S. S. and Panchapakesan, S. (1984). Subset selection procedures: a review and an assessment. Technical Report No. 84-4, Dept. of Statist., Purdue University, West Lafayette, Indiana.

Hajek, J. and Sidak, Z. (1967). Theory of Rank Tests, Academic Press, New York.

Hall, W. J., Wijsman, R. A. and Ghosh, J. K. (1965). The relationship between sufficiency and invariance with applications in sequential analysis. Ann. Math. Statist., 36, 575-614.

Hoel, D. G. (1971). A method for the construction of sequential selection procedures. Ann. Math. Statist., 42, 630-642.

Hoel, D. G. and Mazumdar, M. (1968). An extension of Paulson's selection procedure. Ann. Math. Statist., 39, 2067-2074.

Hsu, J. C. (1981). Simultaneous confidence intervals for all distances from the best. Ann. Statist., 9, 1034-1039.

Hsu, J. C. (1982). Simultaneous inference with respect to the best treatment in block design. J. Amer. Statist. Assoc., 77, 461-467.

Hsu, J. C. and Edwards, D. G. (1983). Sequential multiple comparisons with the best. J. Amer. Statist. Assoc., 78, 965-971.

Huang, D. Y. and Panchapakesan, S. (1982). Some locally optimal subset selection rules based on ranks. Statistical Decision Theory and Related Topics - III, Vol. 2, (Eds. S. S. Gupta and J. O. Berger), Academic, New York, 1-14.

Huang, D. Y., Panchapakesan, S. and Tseng, S. T. (1984). Some locally optimal subset selection rules for comparison with a control. J. Statist., Plan. Infer., 9, 63-72.

Huang, W. T. (1975). Bayes approach to a problem of partitioning k normal populations. Bull. Inst. Math. Acad. Sinica, 3, 87-97.

Ibragimov, I. A. and Linnik, Yu. V. (1971). Independent and Stationary Sequences of Random Variables. Wolters-Noordhoff Publishing Groningen.

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

Miescke, K. J. (1979). Bayesian subset selection for additive and linear loss functions. Commun. Statist.-Theor. Meth., A8, 1205-1226.

Mosteller, F. (1948). A k-sample slippage test for an extreme population. Ann. Math. Statist., 19, 58-65.

Nadaraya, E. A. (1965). On nonparametric estimate of density functions and regression curves. Theor. Probability Appl., 10, 186-190.

Parzen, E. (1962). On estimation of a probability density function and mode. Ann. Math. Statist., 33, 1065-1076.

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

Pollak, M. and Siegmund, D. O. (1975). Approximations to the expected sample size of certain sequential tests. Ann. Statist., 3, 1267-1282.

Robbins, H. (1955). An empirical Bayes approach to statistics. Proc. 3rd Berkeley Symp. Math. Statist. Prob., 155-163.

Robbins, H. (1964). The empirical Bayes approach to statistical decision problems. Ann. Math. Statist., 35, 1-20.

Van Ryzin, J. (1970). On some nonparametric empirical Bayes multiple decision problems. Proc. First Internat. Symp. Non-parametric Techniques in Statistic Inference, (Ed. M. L. Puri), 585-603.

Van Ryzin, J. and Susarla, V. (1977). On the empirical Bayes approach to multiple decision problems. Ann. Statist., 5, 172-181.

Samuel, E. (1963). An empirical Bayes approach to the testing of certain parametric hypothesis. Ann. Math. Statist., 34, 1370-1385.

Singh, A. K. (1977). On slippage tests and multiple decision (selection and ranking) procedures. Ph.D. Thesis. (Mimeo. Ser. No. 494), Dept. of Statist., Purdue University, West Lafayette, IN.

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

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: Technical Report #84-31	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): Some Contributions to Empirical Bayes, Sequential and Locally Optimal Subset Selection Rules	5. TYPE OF REPORT & PERIOD COVERED: Technical
6. PERFORMING ORG. REPORT NUMBER: Technical Report #84-31	
7. AUTHOR(s): TaChen Liang	8. CONTRACT OR GRANT NUMBER(s): N00014-84-0167
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Purdue University, Department of Statistics, West Lafayette, IN 47907	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: Office of Naval Research, Washington, DC	12. REPORT DATE: August 1984
	13. NUMBER OF PAGES: 171
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release, distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Ranking and selection, Empirical Bayes Rules, Subset Selection, Sequential Selection, Locally Optimal, Joint Type II Censoring, Asymptotically Optimal.

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

Selection and ranking problems in statistical inference arise mainly because the classical tests of homogeneity are often inadequate in certain situations where the experimenter is interested in comparing k ($\geq$ 2) populations with the goal of selecting the best or selecting good populations. One of the main approaches for selection and ranking problems is through the subset selection formulation which was pioneered by Gupta (1956, 1965). Since then, especially within the last ten years, different frameworks have been developed under the subset selection formulation (see Gupta and Panchapakesan (1979, 1984)).

(over)

DD FORM 1 JAN 73 1473

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

In this thesis, some results on empirical Bayes rules, sequential subset selection rules have been obtained. Chapter I of this thesis deals with the problem of selecting good populations through the empirical Bayes approach. Two selection problems have been studied: selecting populations better than a control or a standard and selecting all good populations among k populations. For the problem of selecting populations better than a control or a standard, we assume that the underlying populations are binomially distributed. For the problem of selecting all good populations among k populations, it is assumed that the underlying populations have Pareto distributions. For each problem, a nonrandomized Bayes rule is derived for a linear loss function. Later, this Bayes rule is represented in terms of the marginal probability density function of the random observations. Based on this Bayes rule, a sequence of empirical Bayes rules for selecting the good populations is derived. In each problem the rate of convergence of the sequence of the empirical Bayes rules is also studied.

Chapter II deals with the problem of selecting the best population through the sequential subset selection approach. We use a modified sequential probabilistic ratio test to construct sequential selection procedures to select a subset such that (a) a population is eliminated as soon as there is statistical evidence indicating that it is not the best population, and (b) when the procedure terminates and a subset is selected, one can assert, at some prespecified confidence level, the following: simultaneously, the best population is selected and the measure of separation between each selected population and the unknown best population is bounded by some prespecified value.

In Chapter III, we consider some selection problems based on ranks under joint type II censoring. Our goal is to derive locally optimal subset selection rules for selecting a subset containing the best population. Problems are formulated according to whether the sample sizes from the k different populations are equal or not. Some properties associated with the partial ranks under the joint type II censoring are given. Locally optimal subset selection rules R_1 (for the equal sample sizes case) and R_2 (for the unequal sample sizes case) are derived and some locally monotone properties of R_1 and R_2 are also discussed.