DEPARTMENT OF OPERATIONS RESEARCH

COLLEGE OF ENGINEERING
CORNELL UNIVERSITY
ITHACA, NEW YORK 14850

DEPARTMENT OF OPERATIONS RESEARCH
COLLEGE OF ENGINEERING
CORNELL UNIVERSITY
ITHACA, NEW YORK

TECHNICAL REPORT NO. 102

March 1970

ASYMPTOTICALLY OPTIMAL RANKING AND SELECTION PROCEDURES

by

Vijay Singh Bawa

Prepared under contracts

DA-31-124-ARO-D-474, U.S. Army Research Office-Durham

Nonr-401(53), Office of Naval Research

and

National Science Foundation Grant GP 7798

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1 Abstract

In this thesis*, we develop single stage (fixed sample size) asymptotically optimal (minimax) procedures for ranking populations in the presence of nuisance parameters, when the populations are ranked according to a parameter of the distribution and the so-called indifference-zone approach to ranking and selection problems is employed. We adapt methods proposed by Weiss and Wolfowitz in developing asymptotically optimal (minimax) procedures for a certain class of 2-decision tests of composite hypotheses problems in the presence of nuisance parameters to multiple decision ranking and selection problems in the presence of nuisance parameters.

For the problem of selecting the "best" population, asymptotically optimal procedures are developed for situations in which the joint density function of the observations satisfies certain mild regularity conditions (similar to those imposed by Weiss and Wolfowitz). The method of analysis and basic theory is developed in detail for this case. The basic results are extended to develop asymptotically optimal procedures for certain other ranking goals considered in the literature. Some examples are included to illustrate the applicability of the results to specific distributions.

For ranking and selection problems with joint density function of observations not satisfying the regularity conditions, i.e., non-regular cases, we illustrate the applicability of the basic method by developing asymptotically optimal procedures for ranking non-regular exponential and

*A thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Field of Operations Research, Cornell University, June 1970.

uniform distributions.

The results in this thesis can be thought of as generalizing the basic method of Weiss and Wolfowitz for 2-decision hypothesis testing problem to multiple decision ranking and selection problems. As a consequence of our results, we show the asymptotically optimal character of certain so-called natural selection procedures which already have been proposed in the literature. We also develop single-stage asymptotically optimal procedures for certain problems for which heretofore no single-stage procedures had been proposed.

1.2 Outline of the Thesis

In Section 1.1, we have given an overview of the problem considered below and of the results obtained. In the present section we outline the contents of the various chapters.

In Section 1.3 we give a brief introduction to ranking and selection problems. In Section 1.4, we introduce the basic method proposed by Weiss and Wolfowitz [55] in developing asymptotically minimax tests of composite hypotheses. In Section 1.5, we point out that by treating Ranking and Selection problems in the framework of statistical decision theory, the basic method of Weiss and Wolfowitz can be extended to develop asymptotically optimal ranking procedures.

In Chapter 2, we consider the problem of selecting the "best" population. The notation used throughout the thesis is defined. Mild regularity conditions imposed on the density functions are specified. Some preliminary results of statistical decision theory are included. The ranking problem

is structured as a decision theory problem and an asymptotically optimal procedure for a particular zero-one loss function is obtained. The rate of convergence of the decision variables to the asymptotically normally distributed variables is studied. The general results are applied to certain specific distributions, and asymptotically optimal procedures noted in each case. Large sample applications of our results are illustrated by indicating how the procedure would be used in ranking means of normal population with common unknown variance.

In Chapter 3, we extend our basic method to develop asymptotically optimal procedures for certain other ranking goals. Asymptotically optimal procedures are developed for the problem of selecting a fixed-size subset to contain the best population and for the problem of selecting one of the t best populations. We discuss certain other general ranking goals which have been considered in the literature and develop asymptotically optimal procedures for two additional ones.

In Chapter 4, we note some density functions not satisfying the regularity conditions, and develop for non-regular exponential and uniform distributions, asymptotically optimal procedures for selecting the best population and for certain other ranking goals.

1.3 Ranking and Selection Problems

Bechhofer [4] in his pioneering paper pointed out the inappropriateness of the traditional practice of testing null hypotheses and proposed for a certain class of problems, termed ranking problems, the basic concepts of his multiple decision ranking approach. These methods for ranking,

or partially ranking, a group of populations on the basis of an experiment are of great practical importance, especially in connection with the problem of selecting the best populations from a set of competing ones.

Since the paper by Bechhofer, a considerable number of papers have been written on ranking and selection problems. Although there are many ways in which such problems can be formulated, the two most common formulations in the literature are the "Indifference Zone" approach, as proposed by Bechhofer [4], and the "Subset" approach, credited to Gupta [20]. In this thesis, we concentrate on the indifference-zone approach to ranking and selection problems. Unless otherwise stated, in all references to ranking and selection problems in this thesis, it is understood that we refer to the indifference-zone approach to the problem.

The general nature of investigations in ranking and selection problems may be summarized as follows:

Let $\Pi_1,\Pi_2,\ldots,\Pi_k$ denote k populations $(k \geq 2)$ being ranked, with $F(\cdot;\theta_t)$ denoting the distribution function of Π_t $(t = 1,2,\ldots,k)$. Here θ_t is a vector of population parameters. The populations are ranked on the basis of a well defined scalar $\psi_t = \psi(\theta_t)$, $t = 1,2,\ldots,k$. Each θ_t may be completely or partly unknown, but the functional form ψ is known. (In most problems, ψ is one of the parameters of the distribution.) The ordered ψ_t are denoted by $\psi_{[1]} \leq \psi_{[2]} \leq \cdots \leq \psi_{[k]}$. It is assumed that the pairing of the Π_t $(t = 1,2,\ldots,k)$ with the $\psi_{[j]}$ $(j = 1,2,\ldots,k)$ is completely unknown. Let $\Pi_{(t)}$ denote the population associated with $\psi_{[t]}$.

It is assumed that the larger the value of ψ, the better the associated population; $\Pi_{(k)}$, the population associated with $\psi_{[k]}$, is denoted the "best" population. In general, $\Pi_{(k-t+1)}$ $(t = 1,2,\ldots,k)$ is denoted the "t^{th} best" population. (Analogously, in appropriate situations, we may consider smaller values of ψ as more desirable and denote $\Pi_{(1)}$ as the "best" population.)

Selection Procedures (SP) are set up in such a way as to guarantee certain requirements on the probability of "Correct Selection," where the definition of correct selection depends on the ranking "Goal" being considered. Some of the goals considered in the literature are

i) Select the t best populations $(1 \leq t < k)$
   a) with regard to order
   b) without regard to order

ii) Select s of the t best populations $(1 \leq s \leq t < k)$.

iii) Select a subset of s populations to contain the t best populations $(1 \leq t \leq s < k)$.

iv) Select a subset of s populations to contain at least d of t best populations.

For a given goal, with associated definition of correct selection, we are interested in any Selection Procedure (SP) which guarantees the probability requirement

$$\inf_{\Omega(\delta^*)} P(CS|SP) \geq P^* \tag{1.1}$$

where $\omega = (\theta_1,\theta_2,\ldots,\theta_k)$, Ω is the set of all possible ω, and $\Omega(\delta^*)$, called the preference zone, is a certain subspace of the parameter space

Ω, defined according to the goal being considered and a certain distance function $\delta(\cdot,\cdot)$. The function $\delta(b,a)$ measures the "distance" between two populations with values of parameter of interest ψ being b and a respectively, with $a \leq b$. It is assumed to be i) non-negative ii) equal to zero if and only if $a = b$, and iii) increasing in b for fixed a and decreasing in a for fixed b.

For example, for Goal i), $\delta_t = \delta(\psi_{[k-t+1]}, \psi_{[k-t]})$ measures the distance between the set of t best populations $(\Pi_{(k)},\ldots,\Pi_{(k-t+1)})$ and the remaining $(k-t)$ worst populations; and $\Omega(\delta^*)$ may be defined as

$$\Omega(\delta^*) = \{\omega \in \Omega \mid \delta_t \geq \delta^*\} \tag{1.2}$$

If ψ_t is the location parameter of Π_t $(t = 1,2,\ldots,k)$, then one may choose $\delta_t = \psi_{[k-t+1]} - \psi_{[k-t]}$ as a natural distance function for the problem; but for other cases $\delta(\cdot,\cdot)$ may be defined using practical and/or theoretical considerations.

Here $\{(\delta^*, P^*)\}$ with $\delta^* > 0$, $A(k) < P^* < 1$ are constants, specified prior to start of experimentation. $A(k)$ is a lower bound on the specified probability P^*, which depends on the number of populations (k) and the goal being considered. It is the probability which could be achieved by selecting at random and not carrying out any experimentation.

Then for a given goal and distance function, and specified constants (δ^*,P^*), a selection procedure $SP(\delta^*,P^*)$ satisfying (1.1) is defined.

A selection procedure comprises a terminal decision rule and a sampling rule (and associated stopping rule if sequential sampling is used). The sampling rule may be single stage (fixed sample size), two or more stages or fully sequential.

In the field of ranking and selection problems, selection procedures have been proposed for the particular problem at hand. In view of the above, a particular ranking problem is characterized by: i) the form of distribution function $F(\cdot,\cdot)$, ii) the scalar $\psi(\cdot)$ by which populations are ranked iii) the ranking goal and iv) the distance function $\delta(\cdot,\cdot)$.

Thus, for example, Bechhofer [4] proposed a single stage procedure for selecting the best normal population, when populations are ranked according to their means (with known variances), and the difference between the largest and second largest mean is the distance function. In [4], Bechhofer also proposed single stage selection procedures for certain other goals for ranking means of normal populations with known variances. Tables of sample sizes required to meet the basic probability requirement (1.1) were provided. These tables would also be useful in obtaining a large sample approximation to sample size in using single stage selection procedures for ranking parameters of certain other distributions. It is also useful as an approximation for ranking variances of normal populations, a problem for which tables of exact sample size for a single stage selection procedure were provided by Bechhofer and Sobel [8].

Bechhofer, Dunnett and Sobel [5] proposed a two stage procedure for ranking means of normal populations with a common unknown variance. In the proposed procedure, the unknown variance is estimated from observations obtained in the first stage, and an additional random number of observations are taken at the second stage, the number depending on the outcome of the first stage are determined so as to guarantee the basic probability requirement.

Paulson [36] proposed a class of sequential procedures for selecting the normal population with the largest mean, the populations having a common variance; when the common variance is known, the sequential procedure is closed. Paulson [37] also proposed a sequential procedure for selecting the best binomial population. Hoel and Mazumdar [26] have extended Paulson's open sequential procedure to solve the problem of selecting the best from the class of Koopman-Darmois family of distributions.

Bechhofer, Kiefer and Sobel [7] have proposed sequential procedures (including generalizations of Paulson's [36] procedure) for ranking problems associated with the Koopman-Darmois family of distributions. Perng [38] has recently compared the asymptotic expected sample sizes of the two sequential procedures of Paulson [36] and Bechhofer, Kiefer and Sobel [7] for problem of ranking normal means with common known variance. In the literature, there are no other sequential procedures proposed for the ranking problems which satisfy the basic probability requirement. Robbins, Sobel and Starr [41], Srivastava [46] and Srivastava and Oglivie [47] have proposed sequential procedures, for the problem of ranking means of populations with common unknown variance, which satisfy the probability requirement only asymptotically (as $\delta^* \to 0$).

Since this discussion is not intended to be an extensive review of the literature, we refer the reader to the monograph by Bechhofer, Kiefer and Sobel [7] for a comprehensive bibliography on Ranking and Selection problems (for the indifference zone approach and other approaches considered in the literature). Also see Ramberg [39] for some recent work on certain ranking problems associated with multivariate normal populations.

In this thesis, we are only concerned with single stage procedures and the following discussion refers only to such procedures for ranking and selection problems. In most of the work done in this field, the selection procedures which have been proposed initially were developed more or less on an "intuitive" basis (so-called natural selection procedures) to satisfy the basic probability requirement imposed on the procedures. Bahadur and Goodman [1] considered a class of multiple-decision rules which they called impartial (invariant under permutations of the populations). Their results are applicable to the problem of selecting the best population and imply that Bechhofer's ([4]) and Bechhofer and Sobel's ([8]) natural selection procedures are minimax rules (in fact, uniformly minimum risk rules) among the class of impartial decision rules.

Hall [25] removed the restriction of impartiality and proved the optimality of the natural selection procedures by proving their minimax character (by introducing a suitable zero-one type loss function). Hall's results are applicable not only for the problem of selecting the best population, but also for the problem of ranking a specified number of populations, with or without regard to order. Hall's results apply to

problems wherein the ranking parameter is a location or scale parameter and for which there is a sufficient statistic (for each sample size) with a monotone likelihood ratio. In situations where the ranking parameter is not a location or scale parameter, Hall showed the "most economical" character of natural selection procedures for a specified location of the ranking parameter. Since for many problems, a least favorable location (of the ranking parameter) can be determined, Hall concluded that the optimal character of the natural selection procedure can be shown to hold irrespective of the location of the ranking parameter. Thus, Hall's result applies to the problem of selecting the best population whenever there is a sufficient statistic with a monotone likelihood ratio, and therefore, in particular, if its distribution is in the exponential family.

Lehmann [31] extended the results of Bahadur and Goodman [1] to show the optimality of natural selection procedures, among the class of impartial decision rules, for problems of ranking (with or without regard to order) populations, when the ranking parameter has a sufficient statistic with a monotone likelihood ratio. Lehmann [31] also showed certain other optimum properties of the natural selection procedures and provided an alternate proof to results of Hall [25]. Eaton [17] has shown the optimality of natural selection procedures, among the class of impartial decision rules, when the ranking parameter has a sufficient statistic which has a certain monotonicity property (defined in [17] and similar to the rankability condition in [7]). Eaton's results thus extend the optimality of selection procedures to a larger class of density functions. Fabian [19] has shown certain other optimum properties of natural selection

procedures for ranking problems.

As a consequence of the results in [1], [25], [31] and [17], we note that the natural selection procedures proposed for the following problems are indeed optimal among the class of all single-stage procedures. (The references noted in brackets indicate the papers in which the procedures were proposed).

(i) Ranking means of univariate normal population with common known variance (Bechhofer [4]).

(ii) Ranking variances of univariate normal populations (Bechhofer and Sobel [8]).

(iii) Selecting the best of several binomial populations (Sobel and Huyett [44]).

(iv) Selecting the multinomial event which has the highest probability (Bechhofer, Elmaghraby and Morse [6]).

(v) Selecting the bivariate normal population with largest correlation coefficient (Ramberg [39]; also given as an example in Eaton [17]) .

(vi) Selecting the component with the largest mean in ranking from a single multivariate normal population with common known variance and covariance of the components. (Given as an example in Eaton [17] and Milton [34]).

In most of the work done in ranking problems, populations are ranked according to values of a certain parameter in the distribution of the populations. Other unknown parameters in the distribution, if any, would constitute "nuisance" parameters for the ranking problem. In the work cited above on optimality of single stage ranking procedures, not much explicit consideration is given to the nuisance parameters. For example, in [17], it is assumed that a sufficient statistic exists for any unknown nuisance parameters, and the nuisance parameters are such that the basic

probability requirement is guaranteed for all possible values of unknown nuisance parameters by a finite sample size single stage procedure. For problems, where due to unknown nuisance parameters, no finite sample size single-stage procedure satisfying the basic probability requirement exists, one may like to develop single-stage procedures which satisfy the probability requirement asymptotically (as $\delta^* \to 0$). No previous work seems to have been done for such problems. The procedures developed in this thesis are applicable for such problems. These procedures which satisfy the probability requirement asymptotically, are also asymptotically optimal among the class of all decision rules.

In most of the work on ranking problems, large sample approximations to the single-stage sample size (needed to guarantee the probability requirement) are suggested. It would be interesting to study the asymptotic properties of the large sample approximations to the natural selection procedures. Asymptotically optimal procedures developed in this thesis answer this question and we show the asymptotically optimal character of certain natural selection procedures which already have been proposed in the literature.

Finally, in the field of ranking and selection problems, there are certain problems for which no single stage procedures have been proposed (for example ranking scale parameters of Weibull distributions with known location parameter and common (known or unknown) shape parameter). Asymptotically optimal procedures developed in this thesis are applicable to such problems too, thus indicating the wide applicability of our results.

1.4 Asymptotically minimax tests of composite hypotheses

For testing a simple hypothesis versus a simple alternative, the Neyman-Pearson Lemma provides an optimal test. That is among the class of all fixed sample size tests with level of significance $\leq \alpha$ $(0 < \alpha < 1)$, the optimal test maximizes the power. For a certain class of composite hypotheses testing problems, Neyman [35] obtained asymptotically optimal tests in a certain class of asymptotically similar tests. Lecam [29] extended Neyman's results to obtain asymptotically optimal tests among a larger class of asymptotically similar tests. Bartoo and Puri [3] and Buhler and Puri [11] have extended Neyman's result to slightly more general setups.

Very recently, Weiss and Wolfowitz [55] have obtained asymptotically minimax (optimal) procedures for a certain class of composite hypotheses testing problems. Using the basic method developed by the authors ([52] and [53]) in the general theory of asymptotically efficient estimators, Weiss and Wolfowitz obtained asymptotically optimal tests of hypotheses in the presence of nuisance parameters. There are no arbitrary restrictions on the class of tests among which optimal tests are being developed; hence the tests are asymptotically optimal tests among the class of all tests. The general theory is developed in [55] for the class of density functions satisfying some mild regularity conditions; but the basic idea can be used for the non-regular cases as well, each such non-regular case requiring special analysis.

The basic method of analysis and the problem considered by Weiss and Wolfowitz may be summarized as follows:

Let $X_1, X_2, \ldots, X_n$ be independent, identically distributed random variables with the common density function $f(\cdot;\theta)$. The density function is characterized by the vector parameter $\theta = (\theta_1, \theta_2, \ldots, \theta_p, \theta_{p+1})$, $p \geq 1$. $(\theta_1, \theta_2, \ldots, \theta_p)$ are the unknown nuisance parameters, and θ_{p+1} is the parameter being tested. Let H_0 be the (null) hypothesis that $\theta_{p+1} = \bar{\theta}_{p+1}$ and H_1 be the (alternative) hypothesis that $\theta_{p+1} = \bar{\theta}_{p+1} + \frac{c}{\sqrt{n}}$, where $\bar{\theta}_{p+1}$ is a given constant and c is a given positive constant.

Weiss and Wolfowitz [55] consider the problem of testing the null hypothesis H_0 versus the alternative hypothesis H_1. They obtain an asymptotically optimal test. That is, among the class of all tests (of H_0 versus H_1) which (is the limit as $n \to \infty$) have level of significance less than or equal to α $(0 < \alpha < 1)$, the optimal test maximizes (in the limit as $n \to \infty$), the power function.

In order to develop an asymptotically optimal test, Weiss and Wolfowitz first solve the following sequence of artificial problems (one for each n):

H_0 and H_1 are the two hypotheses as given above. The statistician does not know $(\theta_1, \ldots, \theta_p)$ but does know that

$$(1.3) \qquad |\theta_i - \bar{\theta}_i| \leq \frac{M_n(\bar{\theta})}{\sqrt{n}} \quad i = 1,2,\ldots,p \quad \text{and}$$

$$\theta_{p+1} = \bar{\theta}_{p+1} \quad \text{or} \quad \bar{\theta}_{p+1} + \frac{c}{\sqrt{n}}$$

and wishes to test H_0 versus H_1. Here $M_n(\bar{\theta})$ is a positive function of n such that $M_n(\bar{\theta}) \to \infty$, $\frac{M_n(\bar{\theta})}{\sqrt{n}} \to 0$. The above problem is an

artificial problem since it is assumed that the statistician knows $(\bar{\theta}_1, \bar{\theta}_2, \ldots, \bar{\theta}_p)$.

An asymptotically optimal (minimax) procedure is developed by constructing a Bayes decision rule for the problem of testing hypothesis H_0 versus H_1 (when the loss function is zero-one: the loss if zero or one according as the correct or incorrect decision is made) relative to the following apriori distribution: a total mass of b is spread uniformly over the set $\bar{\theta}_i - \frac{M_n(\bar{\theta})}{\sqrt{n}} \leq \theta_i \leq \bar{\theta}_i + \frac{M_n(\bar{\theta})}{\sqrt{n}}$ $i = 1,2,\ldots,p$ and $\theta_{p+1} = \bar{\theta}_{p+1}$ and a total lass of (1-b) is spread uniformly over the set $\bar{\theta}_i - \frac{M_n(\bar{\theta})}{\sqrt{n}} \leq \theta_i \leq \bar{\theta}_i + \frac{M_n(\bar{\theta})}{\sqrt{n}}$ $i = 1,2,\ldots,p$ and $\theta_{p+1} = \bar{\theta}_{p+1} + \frac{c}{\sqrt{n}}$.

Weiss and Wolfowitz [55] obtained a Bayes decision rule for the above problem and studied the asymptotic properties of the decision rule for a class of density function satisfying certain mild regularity conditions (stated in [55] and very similar to the regularity conditions imposed for our problem in Section 2.1). Since they are interested in asymptotic behavior of the decision rule, the apriori mass b $(0 < b < 1)$ is adjusted in such a way that the level of significance for the artificial hypothesis testing problem is equal to a specified level α $(0 < \alpha < 1)$. Then, the Bayes decision rule for this specially selected prior is, by the very nature of being a Bayes decision rule, an asymptotically optimal procedure for the sequence of artificial problems. It is shown in [55] that the asymptotically minimax (optimal) procedure is a function

of $(\bar{\theta}_1,\theta_2,\ldots,\bar{\theta}_p)$, which was assumed to be known for the sequence of artificial problems.

To obtain an asymptotically minimax procedure for the real problem, Weiss and Wolfowitz [55] propose using the asymptotically minimax procedure for the artificial problem, with $(\bar{\theta}_1,\bar{\theta}_2,\ldots,\bar{\theta}_p)$ in the decision variable replaced by estimators $(\hat{\theta}_1(n),\hat{\theta}_2(n),\ldots,\hat{\theta}_p(n))$ of the unknown nuisance parameter. These estimators of $(\theta_1,\theta_2,\ldots,\theta_p)$, based on $X_1,X_2,\ldots,X_n$ satisfy the following consistency condition; for any $\epsilon > 0$, there exists $D(\epsilon) < \infty$ such that

$$(1.4) \qquad \underset{\substack{\theta_1,\theta_2,\ldots,\theta_p \\ H_0 \text{ or } H_1 \text{ is true}}}{P}\{\sqrt{n}|\hat{\theta}_i(n) - \theta_i) < D(\epsilon) \quad i = 1,2,\ldots,p\} > 1 - \epsilon$$

Because of the above consistency condition, the proposed decision rule has the same asymptotic properties as the asymptotically optimal rule for the artificial problem, and is hence asymptotically optimal for the real problem.

The basic method (outlined above without going into any mathematical analysis) proposed by Weiss and Wolfowitz is very powerful, yet very simple, and is adopted in this thesis to develop asymptotically optimal ranking procedures.

1.5 Ranking problems in the framework of statistical decision theory and some additional remarks:

In the field of ranking and selection problems, Sommerville [45] was the first author explicitly to consider the ranking problem in the framework of statistical decision theory. By introducing a loss structure into the ranking problem, Sommerville obtained a minimax decision rule which balanced the cost of experimentation against the expected loss associated with a wrong decision. However, Sommerville considered only prespecified selection procedures, whereas we are concerned with developing selection procedures which are in some sense optimal in the class of all decision rules. Bechhofer, Kiefer and Sobel (p. 46 of [7]) also point out how the ranking problem could be reduced to a decision theory problem by introducing a suitable loss function.

In the work cited in Section 1.3 on showing the optimality of natural selection procedures, the ranking problem is treated in the framework of statistical decision theory. The optimality of natural selection procedures is proved by treating the problem as a multiple-decision problem and showing the minimax character of the procedure, by introducing a suitable loss function into the problem. In this thesis too, we treat the ranking problem in the framework of Wald's [48] statistical decision theory as a multiple decision problem, wherein the number of decisions depend on the number of populations being ranked and on the ranking goal being considered. A zero-one type loss function, suitable for the ranking goal being considered, is introduced and adopting the basic method proposed by Weiss and Wolfowitz [55] for 2-decision problems in the presence of unknown nuisance parameters, we proceed to

develop an asymptotically minimax procedure for the multiple-decision ranking problem in the presence of unknown nuisance parameters. As in [55], we first develop an asymptotically minimax procedure for a sequence of artificial problems, which in turn gives an asymptotically minimax procedure for the ranking problem.

In our analysis, we make no assumption about the form of the distributions (except that they satisfy certain mild regularity conditions, given later on in Section 2.1) or the existence of any sufficient statistic (for each sample size) for the ranking parameter. The problem is solved for an arbitrary, but fixed, location of the ranking parameter. If the ranking parameter is a location or scale parameter admitting a sufficient statistic, then the results hold irrespective of the specified location of the parameter and hence the procedure developed is an asymptotically optimal ranking procedure. In other situations one may be able to find "least favorable location" of the ranking parameter and thus the solution to the ranking problem at such a location gives an asymptotically optimal procedure. If in the worst case, one cannot find such a least favorable location, the procedure developed by our method gives an asymptotically optimal identification procedure, for any arbitrary, but fixed, location of ranking parameter.

Since the problem of selecting the best population seems to be of most practical interest, and also because the main ideas in the theoretical development of asymptotically optimal (minimax) ranking procedures are illustrated in this case, we treat in detail, the development of an asymptotically optimal procedure for the problem of selecting the best population. For certain additional ranking goals, we develop asymptotically optimal procedures by reducing the analysis to one very similar to the problem of selecting the best population.

CHAPTER 2

ASYMPTOTICALLY OPTIMAL PROCEDURE FOR SELECTING THE BEST POPULATION

2.1 Notation, Assumptions and Regularity Conditions

Let $X_{ti}(i = 1,2,\ldots,n)$ denote observations from population Π_t $(t = 1,2,\ldots,k)$, and write $X_i = (X_{1i},X_{2i},\ldots,X_{ki})$. In this thesis, we assume that $X_1,X_2,\ldots,X_n$ are independent and identically distributed random vectors.

Let $\theta_1,\theta_2,\ldots,\theta_p$ denote unknown nuisance parameters, common to each of the k populations. ψ_t, $(t = 1,2,\ldots,k)$ denotes a scalar valued parameter of population Π_t. The populations are ranked according to the values of the parameter ψ (the larger the value of ψ, the better the population is considered to be). For convenience in notation, let

$$\psi_t = \theta_0 - \theta_{p+t} \qquad t = 1,2,\ldots,k \tag{2.1}$$

Thus, θ_0 may be regarded as the common value of the ranking parameters ψ and θ_{p+1} $(t = 1,2,\ldots,k)$ may be viewed as shifts of the parameter from the common value θ_0.

The density function of the random vector X is represented by $f(x; \theta_0,\theta_1,\theta_2,\ldots,\theta_p,\theta_{p+1},\ldots,\theta_{p+k})$. Denoting $(\theta_0,\theta_1,\ldots,\theta_{p+k})$ by ϑ, the density function will be commonly represented as $f(x;\theta)$.

The basic theory is developed for a class of density functions satisfying certain mild regularity conditions, which we state in the form of the following assumptions:

Assumption 1: $\frac{\partial^3}{\partial\theta_\alpha \partial\theta_\beta \partial\theta_\gamma} f(x;\theta)$ exists and is continuous for all x all θ and all α, β, γ $(\alpha,\beta,\gamma = 1,2,\ldots,p+k)$.

Assumption 2: The $(p+k)\times(p+k)$ matrix whose $(i,j)^{th}$ element is

$$E_\theta\{\frac{\partial}{\partial\theta_i} \log f(X;\theta) \frac{\partial}{\partial\theta_j} \log f(X;\theta)\} \qquad (i,j = 1,2,\ldots,p+k)$$

is assumed to exist and is positive definite for all θ.

Fix a positive quantity L which will remain fixed throughout the analysis. For any positive quantity M, and any $\bar{\theta} = (\bar{\theta}_0,\bar{\theta}_1,\ldots,\bar{\theta}_{p+k})$, let $R_M^n(\bar{\theta})$ denote the following region in θ-space.

$$(2.2) \qquad \begin{aligned} |\theta_i - \bar{\theta}_i| &\leq \frac{M}{\sqrt{n}} \qquad i = 1,2,\ldots,p \\ |\theta_i - \bar{\theta}_i| &\leq \frac{L}{\sqrt{n}} \qquad i = p+1,p+2,\ldots,p+k \end{aligned}$$

Assumption 3: For any $\bar{\theta}$, there exists a sequence of positive values $\{M_n(\bar{\theta})\}$ with

$$(2.3) \qquad \begin{aligned} \lim_{n\to\infty} M_n(\bar{\theta}) &= \infty \\ \lim_{n\to\infty} \frac{M_n(\bar{\theta})}{n^{1/2}} &= 0 \end{aligned}$$

and such that, if $\{\omega(n)\}$ and $\{\mu(n)\}$ are any two sequences in $\bar{\theta}$-space with $\omega(n) \in R^n_{M_n(\bar{\theta})}(\bar{\theta})$ and $\mu(n) \in R^n_{M_n(\bar{\theta})}(\bar{\theta})$ for each n, then

$$\frac{M_n^3(\bar{\theta})}{n^{3/2}} \sum_{i=1}^{n} \left| \left[\frac{\partial^3}{\partial\theta_\alpha \partial\theta_\beta \partial\theta_\gamma} \log f(X_i;\theta) \right]_{\mu(n)} \right|$$

converges stochastically to zero as n increases, when $\omega(n)$ is the true parameter point for X. Also this convergence is uniform in $\mu(n)$ and $\omega(n)$ over $R^n_{M_n(\bar{\theta})}(\bar{\theta})$.

Assumption 4:

$$\left| \int_{-\infty}^{\infty} \frac{\partial^2}{\partial\theta_\beta \partial\theta_\gamma} f(x;\theta) \Big|_{\mu(n)} \frac{\partial \log f(x;\theta)}{\partial\theta_\alpha} \Big|_{\bar{\theta}} dx \right|$$

is uniformly bounded for all $\mu(n) \in R^n_{M_n(\bar{\theta})}(\bar{\theta})$ for all α, β, γ $(\alpha,\beta,\gamma = 1,2,\ldots,p+k)$.

For $\alpha,\beta = 1,2,\ldots,p+k$, let

$$B_n(\alpha,\beta;\bar{\theta}) = -\frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2}{\partial\theta_\alpha \partial\theta_\beta} \log f(X_i;\theta) \Big|_{\bar{\theta}} \tag{2.4}$$

and

$$B(\alpha,\beta;\bar{\theta}) = -E_{\bar{\theta}} \left\{ \frac{\partial^2}{\partial\theta_\alpha \partial\theta_\beta} \log f(X;\theta) \Big|_{\bar{\theta}} \right\} \tag{2.5}$$

Assumption 5: If $\omega(n)$ is the true parameter point for X, $\omega(n) \in R^n_{M_n(\bar{\theta})}(\bar{\theta})$, then $B_n(\alpha,\beta;\bar{\theta})$ converges stochastically to $B(\beta,\alpha;\bar{\theta})$ and $B(\alpha,\beta;\bar{\theta})$ is a continuous function of $\bar{\theta}$, for all α,β $(\alpha,\beta = 1,2,..,p+k)$

In our analysis, we consider only those procedures which take equal size samples from each population. We do not assume the existence of a sufficient statistic for the ranking parameter or the nuisance parameters; and the decision rule is given in terms of the density function $f(x;\theta)$.

Apart from the regularity conditions stated above, we make no other assumptions about the form of the density functions. Thus, the results obtained in this chapter (and also in Chapter 3) are applicable to the problem of ranking several univariate populations according to the values of a certain parameter, each population having common unknown nuisance parameters. The results also apply to problems of ranking several multivariate populations, the populations being ranked according to values of a scalar valued parameter and each population having common unknown nuisance parameters. The results are also applicable to certain ranking problems associated with a single multivariate population. The results are applicable in situations where, apart from the regularity conditions stated above, certain symmetry conditions (on the density function $f(x;\theta)$) hold. These symmetry conditions are needed in showing that a symmetric prior distribution gives a minimax decision rule. Thus, our results apply to the problem of ranking from a multinomial distribution as well as ranking means or variances of a single multivariate normal population with common (known or unknown) correlation coefficient.

2.2 Some Preliminary Results of Statistical Decision Theory

In order to put the ranking problems considered in this thesis into the decision theoretic framework, and present a general theoretical

development of an optimal procedure for different types of ranking problems, we define some notation and present some preliminary decision theoretic results which will be used later on. The notation is similar to that used by Weiss [49].

Let $X_1,\ldots,X_n$ be the observable random variables, on the values of which the decision is to be based. Let x be an index for the possible sets of values of $(X_1,X_2,\ldots,X_n)$. Let $f(x;\theta)$ denote the joint pdf of $(X_1,\ldots,X_n)$, where θ is an index for the possible joint distributions.

Let D be an index for the possible decisions; that is, a particular value of D indicates a particular decision. In the case where there is only a finite number of decisions, say h, we can list the decisions in a particular order and let D_i $(i = 1,2,\ldots,h)$ indicate the i^{th} decision. Since for ranking problems, the number of decisions is finite (this total number depending upon the number of populations and the specific goal being considered), this notation will be used.

In the decision theory formulation, different ranking problems would be analyzed in the same way, differing only in the total number of possible decisions h, and interpretation of each decision (depending on the ranking goal). For example, for selecting the best population, the total number of possible decisions h is equal to the number of populations k; and D_i may refer to selecting Π_i as the best population. For the problem of selecting a fixed size subset of size s $(s < k)$ to contain the best population, the total number of possible decisions h is $\binom{k}{s}$ and each D_i may refer to selecting a particular subset of size s as the best subset.

Let $W(\theta,D,x)$ denote the loss incurred when x is the observed value of $(X_1,\ldots,X_n)$, D is the decision made and θ is the true parameter value. For a large class of problems, $W(\ )$ may be independent of X.

Definition: A decision rule s is defined by nonnegative numbers $s(D;x)$, where $s(D;x)$ is the probability assigned by decision rule s to choosing a decision D when x is observed.

When the total number of possible decisions is finite, say h, then we have for each x,

$$\sum_{i=1}^{h} s(D_i;x) = 1 \tag{2.6}$$

Definition: The expected loss, incurred when using decision rule s, and the true joint probability distribution is given by θ, is denoted by $r(\theta;s)$, and often called the risk function.

For a problem with a finite number h of possible decisions and joint pdf $f(x;\theta)$ (for random variables $(X_1,X_2,\ldots,X_n)$)

$$r(\theta;s) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \sum_{i=1}^{h} W(\theta;D_i;x)\, f(x;\theta)\, s(D_i;x)\, dx_1,\ldots,dx_n \tag{2.7}$$

Let

$$M(s) = \max_{\theta} r(\theta;s) \tag{2.8}$$

Definition: The expected risk for a decision rule s, with respect to a cdf $B(\theta)$ for chance variable θ, is denoted $R(s; B(\theta))$ and given by

$$R(s;B(\theta)) = \underset{B(\theta)}{E} \{r(\theta;s)\} \tag{2.9}$$

Definition: A decision rule s is a "Bayes decision rule relative to $B(\theta)$" if for every decision rule t,

$$R(s;B(\theta)) \leq R(t;B(\theta)) \tag{2.10}$$

Definition: A decision rule s is called a minimax decision rule, if for every decision rule t,

$$M(s) \leq M(t)$$

$B(\theta)$, used for constructing a Bayes decision rule is often called an "apriori distribution." We would like to point out that θ is an unknown vector and not a chance variable. The introduction of the cdf $B(\theta)$ is just a technical device to enable one to define a Bayes decision rule for the case of an infinite number of possible distributions (indexed by θ).

From a Bayesian viewpoint, one may specify some particular cdf $B^*(\theta)$ and construct a Bayes decision rule relative to the specific apriori distribution. In this thesis, however, we are only interested in minimax decision rules, and Bayes decision rules are only used as a

technical device to construct such minimax decision rules.

If $B(\theta)$ has a pdf $b(\theta)$, then

$$R(s;B(\theta)) = \int_{\theta} r(\theta;s)b(\theta)d\theta \tag{2.11}$$

For a problem with a finite number h of possible decisions, and $B(\theta)$ having a pdf $b(\theta)$, we obtain using (2.7) and (2.11)

$$\begin{aligned} R(s;B(\theta)) &= \int_{\theta} \{ \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \sum_{i=1}^{h} s(D_i;x)\, W(\theta;D_i;x)\, f(x;\theta)dx_1 \cdots dx_n \} b(\theta)d\theta \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} [\sum_{i=1}^{h} s(D_i;x)\, k(D_i;x)]dx_1 \cdots dx_n \end{aligned} \tag{2.12}$$

where

$$k(D_i;x) = \int_{\theta} W(\theta;D_i;x)\, f(x;\theta)b(\theta)d\theta \tag{2.13}$$

Using the above representation, we easily see that "s is a Bayes decision rule relative to $B(\theta)$, if for each x, $s(D;x)$ is set equal to zero for every D_i for which $k(D_i;x)$ is greater than $\min_{1 \le i \le h} \{k(D_i,x)\}$."

We end this section by stating, without proof, two well-known theorems which enable one to recognize a decision rule as a minimax decision rule.

Theorem 2.1. If s is a Bayes decision rule relative to $B(\theta)$, and if $r(\bar{\theta};s) = M(s)$ for every $\bar{\theta}$ which is a point of increase of $B(\theta)$, then s is a minimax decision rule.

If the number of possible distribution functions is finite, and if we let $b(\theta_1),\ldots,b(\theta_m)$ denote the apriori distribution for the possible values $(\theta_1,\ldots,\theta_m)$ of θ, then we obtain the following special form of Theorem 2.1

Theorem 2.2. If s is a Bayes decision rule relative to $b(\theta_1),b(\theta_2),\ldots,b(\theta_m)$ and if $r(\theta;s) = m(s)$ for every θ for which $b(\theta)$ is positive, then s is a minimax decision rule.

2.3 Selecting the best population

In this chapter, we derive, in detail, the basic results needed to develop an asymptotically optimal procedure for the problem of selecting the best population, when the density function of observations satisfies the regularity conditions given in Section 2.1. Since we are dealing with the indifference zone approach to the ranking problem, we need to define the preference zone for the problem at hand. As we are interested in developing a procedure (really a sequence of procedures) which is asymptotically optimal, we define the following sequence of preference zones (one for each n).

$$\Omega(\delta^*(n)) = \{\omega \in \Omega \mid \psi_{[k]} - \psi_{[k-1]} \geq \frac{c}{\sqrt{n}}\} \tag{2.14}$$

where $c > 0$, ψ is the parameter being ranked, and $\psi_{[1]} \leq \psi_{[2]} \leq \cdots \leq \psi_{[k]}$ denote the ordered parameters.

Also, for any sequence of procedures $\{T(n)\}$, let

$$P(T) = \lim_{n\to\infty} \inf_{\Omega(\delta^*(n))} P\{CS \mid T(n)\} \tag{2.15}$$

Then, throughout this thesis, an asymptotically optimal selection procedure is defined as a procedure T (really a sequence of procedures (one for each n)) such that among the class of all procedures T',

$$P(T) \geq P(T') \tag{2.16}$$

In order to develop an asymptotically optimal procedure (that is a procedure for which (2.16) holds), we formulate the problem as a decision theory problem with a particular zero-one type loss function, and obtain an asymptotically minimax decision rule for the associated multiple decision problem. This is done in detail in the next section.

We would like to point out here that in developing selection procedures (for ranking problems) which are asymptotically optimal, the problem of defining a suitable preference zone is solved in a nice way. We are interested in defining a sequence of preference zones in such a way that the distance between the best and the second best population approaches zero as $n \to \infty$. That we require such a sequence for the asymptotic theory is clear by the fact that if, for example, the distance between best and second best population were some constant, then for any procedure using a consistent estimator of the ranking parameter, $P(T)$ (as given by (2.15)) would be equal to one (that is, one is able to select the best population with probability approaching one as $n \to \infty$).

Thus, the preference zone as defined in (2.15) is appropriate for the ranking problem. The distance function used in defining the preference zone is a "natural" distance function for the problem of ranking populations according to values of the location parameters. For asymptotic theory, this is an appropriate distance function for ranking any parameters of a distribution, a result not too surprising in view of the fact (as we shall see in the next section) that the problem (at least asymptotically) reduces to one of ranking means of normal populations for which the distance function used is a natural distance function.

For asymptotic theory, in the sequence of preference zones (as defined by (2.14)) the rate of convergence of the distance function to zero (consequently the rate of convergence of preference zones to the whole parameter space) is $1/\sqrt{n}$. This is directly related to the normalizing constant ($\sqrt{n}$ for the class of problems being considered), for which the decision variables have a limiting distribution. If the rate of convergence is too slow, the problem reduces to a degenerate case while if the rate of convergence is too fast, the decision rules will not be able to distinguish the best population among the set of competing ones.

For $i = 1,2,\ldots,k$ let

$$(2.17) \qquad H_i: \begin{cases} \theta_{p+j} = \overline{\theta}_{p+j} + c/\sqrt{n} & j = 1,2,\ldots,k \\ & \neq i \\ \theta_{p+i} = \overline{\theta}_{p+i} - c/\sqrt{n} \end{cases}$$

where $c > 0$ and $\overline{\theta}_{p+j} = \overline{\theta}_{p+i}$ all i,j $(i,j = 1,2,\ldots,k)$ are known values. $\overline{\theta}_{p+j}$ may be taken to be zero, with no loss in generality.

In view of (2.1), (2.17) represents a restricted parameter configuration, in which if H_i is the true state of nature, then Π_i is the best population.

We now discribe the basic method used in developing asymptotically optimal ranking procedures. In order to develop an asymptotically optimal procedure for the ranking problem, we develop an asymptotically optimal procedure for an associated identification problem, with $\underline{\theta}_0$, the common location of the ranking parameter, as the least favorable location of the ranking parameter. We first solve the problem for a restricted parameter configuration, given by (2.17), and then show that the procedure developed is minimax overall parameter configurations. As in [55], we first solve a sequence of artificial problems, the solution to which suggests an optimal procedure for the real ranking problem.

2.4 Asymptotically Optimal Procedure for Zero-one Type Loss Functions

For the problem of selecting the best population, let D_i denote selecting H_i (as given by (2.17)) as the true hypothesis (equivalently selecting Π_i as the best population). The loss function is given, for $i = 1,2,\ldots,k$, by

$$(2.18)\quad W(\theta,D_i;x) = W(\theta;D_i) = \begin{cases} 0 & \text{if } H_i \text{ is the true hypothesis} \\ 1 & \text{otherwise} \end{cases}$$

2.4.1 Preliminary Sequence of Artificial Problems

For the loss function (2.18), and with the joint density functions of the populations satisfying the regularity conditions (of Section 2.1),

we would like to develop an asymptotically optimal procedure for the problem of selecting the best population. Before proceeding to the real problem, we first solve a sequence of artificial k-decision problems (one for each n).

Suppose it is known that for $i = 1,2,\ldots,p$, we have $\bar{\theta}_i - \frac{M_n(\bar{\theta})}{\sqrt{n}} \leq \theta_i \leq \bar{\theta}_i + \frac{M_n(\bar{\theta})}{\sqrt{n}}$ and $(\theta_{p+1},\theta_{p+2},\ldots,\theta_{p+k})$ satisfy one of the k hypotheses given by (2.17). $(\bar{\theta}_1,\ldots,\bar{\theta}_p)$ are known constants and $0 < c \leq L$. We wish to test which one of the k hypotheses $H_1,H_2,\ldots,H_k$ is the true one.

For the above problem, which is an artificial one because we assume $\bar{\theta}_1,\ldots,\bar{\theta}_p$ are known, we construct a Bayes Decision Rule relative to the following apriori distribution: For $j = 1,2,\ldots,k$ a total mass of b_j is spread uniformly over the set

$$\bar{\theta}_i - \frac{M_n(\bar{\theta})}{\sqrt{n}} \leq \theta_i \leq \bar{\theta}_i + \frac{M_n(\bar{\theta})}{\sqrt{n}} \quad i = 1,2,\ldots,p \quad \text{and} \quad H_j \text{ is true}$$

where $b_j \geq 0$ for $j = 1,2,\ldots,k$, and

$$\sum_{j=1}^{k} b_j = 1$$

The prior distribution $(b_j,\ j = 1,2,\ldots,k)$ is arbitrary, but fixed. Later we select the prior in such a way as to obtain a minimax decision rule for the problem at hand.

If we let D_i denote the decision to select H_i as the true hypothesis and compute $k(D_i;x)$ for each i, then it can be seen that a Bayes decision rule relative to the above apriori distribution is given as follows:

Select H_ℓ as the true hypothesis if

$$J_n(\ell|j) \geq \frac{b_j}{b_\ell} \qquad \ell,j = 1,2,\ldots,k \tag{2.19}$$

where for $j,\ell = 1,2,\ldots,k$

$$J_n(\ell|j) = \frac{\int_{\bar{\theta}_1-\frac{M_n(\bar{\theta})}{\sqrt{n}}}^{\bar{\theta}_1+\frac{M_n(\bar{\theta})}{\sqrt{n}}} \cdots \int_{\bar{\theta}_p-\frac{M_n(\bar{\theta})}{\sqrt{n}}}^{\bar{\theta}_p+\frac{M_n(\bar{\theta})}{\sqrt{n}}} \prod_{i=1}^{n} f(X_i;\theta,H_\ell)\,d\theta_1,\ldots,d\theta_p}{\int_{\bar{\theta}_1-\frac{M_n(\bar{\theta})}{\sqrt{n}}}^{\bar{\theta}_1+\frac{M_n(\bar{\theta})}{\sqrt{n}}} \cdots \int_{\bar{\theta}_p-\frac{M_n(\bar{\theta})}{\sqrt{n}}}^{\bar{\theta}_p+\frac{M_n(\bar{\theta})}{\sqrt{n}}} \prod_{i=1}^{n} f(X_i;\theta,H_j)\,d\theta_1,\ldots,d\theta_p} \tag{2.20}$$

Here $f(X_i;\theta,H_j)$ denotes the joint pdf of the observations when θ is the parameter value and $(\theta_{p+1},\ldots,\theta_{p+k})$ are as given by hypothesis H_j. For notational convenience, we let

$$J_n(\ell|j) = \frac{\int\cdots\int \prod_{i=1}^{n} f(X_i;\theta,H_\ell)\,d\theta_1,\ldots,d\theta_p}{\int\cdots\int \prod_{i=1}^{n} f(X_i;\theta,H_j)\,d\theta_1,\ldots,d\theta_p}$$

where the integration is over the region as given by (2.20). Throughout the thesis, unless otherwise specified, $\int\cdots\int$ indicates integration over a region as given by the last expression for the integral where the limits are specified.

We investigate the asymptotic properties of $J_{\cdot}(\ell|j)$. We now define certain notation used frequently later on. Let $\bar{\theta}$ denote $(\bar{\theta}_0,\bar{\theta}_1,\ldots,\bar{\theta}_p,\bar{\theta}_{p+1},\ldots,\bar{\theta}_{p+k})$. Also for $\alpha = 1,2,\ldots,p+k$, let

$$A_n(\alpha;\bar{\theta}) \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial}{\partial\theta_\alpha} \log f(X_i;\theta)\Big|_{\bar{\theta}} \tag{2.21}$$

Denote by $F_n(\alpha,\beta;\bar{\theta})$, $\alpha,\beta = 1,2,\ldots,p$, the $(\alpha,\beta)^{th}$ element of the inverse of the $(p\times p)$ matrix whose $(\alpha,\beta)^{th}$ element is $B_n(\alpha,\beta;\bar{\theta})$, where $B_n(\alpha,\beta;\bar{\theta})$ is as given by (2.4). This inverse exists with probability approaching one as n increases.

For $\alpha,\beta = 1,2,\ldots,p$, let $F(\alpha,\beta;\bar{\theta})$ denote the $(\alpha,\beta)^{th}$ element of the inverse of the $(p\times p)$ matrix whose $(\alpha,\beta)^{th}$ element is $B(\alpha,\beta;\bar{\theta})$. By our assumptions, this inverse exists. Also by our assumptions, if the true parameter point for X is $\omega(n) \in R^n_{M_n(\bar{\theta})}(\bar{\theta})$, then $F_n(\alpha,\beta;\bar{\theta})$ converges stochastically to $F(\alpha,\beta;\bar{\theta})$ as n increases.

Using the above notation, we get by expanding around $\bar{\theta}$, for $\ell = 1,2,\ldots,k$,

$$\begin{aligned}\sum_{i=1}^{n} \log f(X_i;\theta,H_\ell) &= \sum_{i=1}^{n} \log f(X_i;\bar{\theta}) + \sum_{\alpha=1}^{p+k} \sqrt{n}(\theta_\alpha-\bar{\theta}_\alpha)A_n(\alpha;\bar{\theta}) \\ &\quad - \frac{1}{2}\sum_{\alpha=1}^{p+k}\sum_{\beta=1}^{p+k} \sqrt{n}(\theta_\alpha-\bar{\theta}_\alpha)\sqrt{n}(\theta_\beta-\bar{\theta}_\beta)B_n(\alpha,\beta;\bar{\theta}_\beta) + Q_n(\theta_1,\ldots,\theta_p|\ell)\end{aligned} \tag{2.22}$$

where,

$$Q_n(\theta_1,\ldots,\theta_p|\ell) = \frac{1}{6}\sum_{\alpha=1}^{p+k}\sum_{\beta=1}^{p+k}\sum_{\gamma=1}^{p+k}(\theta_\alpha-\bar{\theta}_\alpha)(\theta_\beta-\bar{\theta}_\beta)(\theta_\gamma-\bar{\theta}_\gamma)\cdot$$

(2.23)

$$\sum_{i=1}^{n}\frac{\partial^3}{\partial\theta_\alpha\partial\theta_\beta\partial\theta_\gamma}\log f(X_i;\theta)\Big|_{\mu(n)}$$

and in (2.22) and (2.23), we have set

$$\theta_{p+j} - \bar{\theta}_{p+j} = \frac{c}{\sqrt{n}} \qquad \begin{matrix} j \neq \ell \\ j = 1,2,\ldots,k \end{matrix}$$

$$\theta_{p+\ell} - \bar{\theta}_{p+\ell} = -\frac{c}{\sqrt{n}}$$

and $\mu(n) \in R^n_{M_n(\bar{\theta})}(\bar{\theta})$.

We now prove a lemma which will be useful in studying asymptotic properties of $J_n(\ell|j)$.

Lemma 2.1 For $\ell = 1,2,\ldots,k$, $Q_n(\theta_1,\ldots,\theta_p|\ell)$ converges stochastically to zero as $n \to \infty$ for all $\mu(n) \in R^n_{M_n(\bar{\theta})}(\bar{\theta})$.

Proof: From (2.23), we see that

$$\left|Q_n(\theta_1,\ldots,\theta_p|\ell)\right| \leq \frac{1}{6}\frac{M_n^3(\bar{\theta})}{n^{3/2}}\sum_{\alpha=1}^{p+k}\sum_{\beta=1}^{p+k}\sum_{\gamma=1}^{p+k}\left|\sum_{i=1}^{n}\frac{\partial^3}{\partial\theta_\alpha\partial\theta_\beta\partial\theta_\gamma}\log f(X_i;\theta)\Big|_{\mu(n)}\right|$$

$\to 0$ as $n \to \infty$ by our assumption.

Thus $Q_n(\theta_1,\ldots,\theta_p|\ell)$ converges stochastically to zero as $n \to \infty$ for all $\mu(n) \in R^n_{M_n(\bar{\theta})}(\bar{\theta})$.

Q.E.D.

Using (2.22), we obtain, for $\ell = 1,2,\ldots,k$,

$$\prod_{i=1}^{n} f(X_i;\theta,H_\ell) = \exp\Big\{ \sum_{i=1}^{n} \log f(X_i;\theta,H_\ell) \Big\}$$

$$= \exp\Big\{ \sum_{i=1}^{n} \log f(X_i;\bar{\theta}) + \sum_{\alpha=1}^{p+k} \sqrt{n}(\theta_\alpha-\bar{\theta}_\alpha) A_n(\alpha;\bar{\theta})\Big|_{H_\ell} \tag{2.24}$$

$$- \frac{1}{2} \sum_{\alpha,\beta=1}^{p+k} \sqrt{n}(\theta_\alpha-\bar{\theta}_\alpha)\sqrt{n}(\theta_\beta-\bar{\theta}_\beta)\, B_n(\alpha,\beta;\bar{\theta})\Big|_{H_\ell} + Q_n(\theta_1,\theta_2,\ldots,\theta_p|\ell)\Big\}$$

Substituting from (2.24) in (2.20), we obtain, for $j,\ell = 1,2,\ldots,k$,

$$J_n(\ell|j) =$$

$$\frac{\displaystyle\int\cdots\int \exp\Big\{ \sum_{i=1}^{n} \log f(X_i;\bar{\theta}) + \sum_{\alpha=1}^{p+k} \sqrt{n}(\theta_\alpha-\bar{\theta}_\alpha) A_n(\alpha;\bar{\theta})\Big|_{H_\ell} - \frac{1}{2} \sum_{\alpha,\beta=1}^{p+k} \sqrt{n}(\theta_\alpha-\bar{\theta}_\alpha)\sqrt{n}(\theta_\beta-\bar{\theta}_\beta) B_n(\alpha,\beta;\bar{\theta})\Big|_{H_\ell} + Q_n(\theta_1,\ldots,\theta_p|\ell)\Big\} d\theta_1,\ldots,d\theta_p}{\displaystyle\int\cdots\int \exp\Big\{ \sum_{i=1}^{n} \log f(X_i;\bar{\theta}) + \sum_{\alpha=1}^{p+k} \sqrt{n}(\theta_\alpha-\bar{\theta}_\alpha) A_n(\alpha;\bar{\theta})\Big|_{H_j} - \frac{1}{2} \sum_{\alpha,\beta=1}^{p+k} \sqrt{n}(\theta_\alpha-\bar{\theta}_\alpha)\sqrt{n}(\theta_\beta-\bar{\theta}_\beta) B_n(\alpha,\beta;\bar{\theta})\Big|_{H_j} + Q_n(\theta_1,\ldots,\theta_p|j)\Big\} d\theta_1,\ldots,d\theta_p}$$

where $H_\ell(H_j)$ inside the bracket is used to indicat that hypothesis $H_\ell(H_j)$ is true.

Simplifying the above expression, we obtain, for $\ell, j = 1,2,\ldots,k$,

(2.25)

$$J_n(\ell|j) =$$

$$\frac{\exp\left[-2cA_n(p+\ell;\bar{\theta}) - \frac{1}{2}\sum_{\alpha=p+1}^{p+k}\sum_{\beta=p+1}^{p+k}\sqrt{n}(\theta_\alpha-\bar{\theta}_\alpha)\sqrt{n}(\theta_\beta-\bar{\theta}_\beta)B_n(\alpha,\beta;\bar{\theta})\Big|_{H_\ell}\right]}{\exp\left[-2cA_n(p+j;\bar{\theta}) - \frac{1}{2}\sum_{\alpha=p+1}^{p+k}\sum_{\beta=p+1}^{p+k}\sqrt{n}(\theta_\alpha-\bar{\theta}_\alpha)\sqrt{n}(\theta_\beta-\bar{\theta}_\beta)B_n(\alpha,\beta;\bar{\theta})\Big|_{H_j}\right]} \cdot$$

$$\frac{\int\cdots\int \exp\left[\sum_{\alpha=1}^{p}\sqrt{n}(\theta_\alpha-\bar{\theta}_\alpha)A_n(\alpha;\bar{\theta}) - \frac{1}{2}\sum_{\alpha=1}^{p}\sum_{\beta=1}^{p}\sqrt{n}(\theta_\alpha-\bar{\theta}_\alpha)\sqrt{n}(\theta_\beta-\bar{\theta}_\beta)B_n(\alpha,\beta;\bar{\theta}) - c\sum_{\alpha=1}^{p}\sqrt{n}(\theta_\alpha-\bar{\theta}_\alpha)\Big[\sum_{\substack{\gamma=p+1\\ \neq p+\ell}}^{p+k}B_n(\alpha,\gamma;\bar{\theta}) - B_n(\alpha,p+\ell;\bar{\theta})\Big] + Q_n(\theta_1,\ldots,\theta_p|\ell)\right] d\theta_1\ldots d\theta_p}{\int\cdots\int \exp\left[\sum_{\alpha=1}^{p}\sqrt{n}(\theta_\alpha-\bar{\theta}_\alpha)A_n(\alpha;\bar{\theta}) - \frac{1}{2}\sum_{\alpha=1}^{p}\sum_{\beta=1}^{p}\sqrt{n}(\theta_\alpha-\bar{\theta}_\alpha)\sqrt{n}(\theta_\beta-\bar{\theta}_\beta)B_n(\alpha,\beta;\bar{\theta}) - c\sum_{\alpha=1}^{p}\sqrt{n}(\theta_\alpha-\bar{\theta}_\alpha)\Big[\sum_{\substack{\gamma=p+1\\ \neq p+j}}^{p+k}B_n(\alpha,\gamma;\bar{\theta}) - B_n(\alpha,p+j;\bar{\theta})\Big] + Q_n(\theta_1,\ldots,\theta_p|j)\right] d\theta_1\ldots d\theta_p}$$

If we denote by $J_n'(\ell|j)$, the above expression for $J_n(\ell|j)$ with $Q_n(\theta_1,\ldots,\theta_p|\ell)$ and $Q_n(\theta_1,\ldots,\theta_p|j)$ being removed from numerator and denominator respectively, then we obtain the following useful lemma.

Lemma 2.2. For $\ell, j = 1,2,\ldots,k$

$$\frac{J_n(\ell|j)}{J_n'(\ell|j)} \xrightarrow{\text{stochastically}} 1 \quad \text{as} \quad n \to \infty,$$

for all true parameter points $\mu(n) \in R^n_{M_n(\bar{\theta})}(\bar{\theta})$.

Proof: Using the law of the mean for integrals, we can write

$$J'_n(\ell|j) = \frac{e^{\bar{Q}_n(\ell)}}{e^{\bar{Q}_n(j)}} J_n(\ell|j)$$

where $\bar{Q}_n(i)$ $(i = j,\ell)$ is between the minimum and maximum values taken by $Q_n(\theta_1,\ldots,\theta_p;i)$ in the region of integration. Using Lemma 2.1, it follows immediately that if $\mu(n)$ is the true parameter point for X and $\mu(n) \in R^n_{M_n(\bar{\theta})}(\bar{\theta})$, then $\frac{J'_n(\ell|j)}{J_n(\ell|j)}$ converges stochastically to one as $n \to \infty$.

Q.E.D.

As a direct consequence of lemma 2.2, to study the asymptotic properties of $J_n(\ell|j)$, we need only study the asymptotic properties of $J'_n(\ell|j)$. If we let $w_\alpha = \sqrt{n}(\theta_\alpha - \bar{\theta}_\alpha)$, $\alpha = 1,2,\ldots,p$, we get for $\ell,j = 1,2,\ldots,k$,

(2.26)

$$J_n'(\ell|j) =$$

$$\frac{\exp\left[-2cA_n(p+\ell;\bar{\theta}) - \frac{1}{2}\sum_{\alpha,\beta=p+1}^{p+\ell}\sqrt{n}(\theta_\alpha-\bar{\theta}_\alpha)\sqrt{n}(\theta_\beta-\bar{\theta}_\beta)B_n(\alpha,\beta;\bar{\theta})\Big|_{H_\ell}\right]}{\exp\left[-2cA_n(p+j;\bar{\theta}) - \frac{1}{2}\sum_{\alpha,\beta=p+1}^{p+j}\sqrt{n}(\theta_\alpha-\bar{\theta}_\alpha)\sqrt{n}(\theta_\beta-\bar{\theta}_\beta)B_n(\alpha,\beta;\bar{\theta})\Big|_{H_j}\right]} \cdot$$

$$\frac{\int_{-M_n(\bar{\theta})}^{M_n(\bar{\theta})}\cdots\int_{-M_n(\bar{\theta})}^{M_n(\bar{\theta})} \exp\left[\sum_{\alpha=1}^{p} w_\alpha\left[A_n(\alpha;\bar{\theta}) - c\left[\sum_{\substack{\gamma=p+1\\ \neq p+\ell}}^{p+k} B_n(\alpha,\gamma;\bar{\theta}) - B_n(\alpha,p+\ell;\bar{\theta})\right]\right] - \frac{1}{2}\sum_{\alpha,\beta=1}^{p} w_\alpha w_\beta B_n(\alpha,\beta;\bar{\theta})\right] dw_1\cdots dw_p}{\int_{-M_n(\bar{\theta})}^{M_n(\bar{\theta})}\cdots\int_{-M_n(\bar{\theta})}^{M_n(\bar{\theta})} \exp\left[\sum_{\alpha=1}^{p} w_\alpha\left[A_n(\alpha;\bar{\theta}) - c\left[\sum_{\substack{\gamma=p+1\\ \neq p+j}}^{p+k} B_n(\alpha,\gamma;\bar{\theta}) - B_n(\alpha,p+j;\bar{\theta})\right]\right] - \frac{1}{2}\sum_{\alpha,\beta=1}^{p} w_\alpha w_\beta B_n(\alpha,\beta;\bar{\theta})\right] dw_1\cdots dw_p}$$

Since $M_n(\bar{\theta}) \to \infty$ as $n \to \infty$, it would be tempting to set in the limits of integration $M_n(\bar{\theta}) = \infty$ and conclude that the resulting value of the expression (2.26) would have the same asymptotic behavior as $J_n'(\ell|j)$. However, since the integrand is also a function of n, through $B_n(\alpha,\beta;\bar{\theta})$, a careful analysis is required.

Before proceeding to that, denote by $J_n''(\ell|j)$ the value of $J_n'(\ell|j)$ if in the limits of integration in tho expression for $J_n'(\ell|j)$ ((2.26)),

$M_n(\bar{\theta})$ are replaced by ∞. By treating w_α ($\alpha = 1,2,\ldots,p$) formally as having a multivariate normal distribution with mean vector 0 and covariance matrix given by $\left(\left(F_n(\alpha,\beta;\theta)\right)\right)$, we get for $\ell,j = 1,2,\ldots,k$,

(2.27)

$$J_n''(\ell|j) =$$

$$\frac{\exp\left[-2cA_n(p+\ell;\bar{\theta}) - \frac{1}{2}\sum_{\alpha,\beta=p+1}^{p+k}\sqrt{n}(\theta_\alpha-\bar{\theta}_\alpha)\sqrt{n}(\theta_\beta-\bar{\theta}_\beta)B_n(\alpha,\beta;\bar{\theta})\Big|_{H_\ell}\right]}{\exp\left[-2cA_n(p+j;\bar{\theta}) - \frac{1}{2}\sum_{\alpha,\beta=p+1}^{p+k}\sqrt{n}(\theta_\alpha-\bar{\theta}_\alpha)\sqrt{n}(\theta_\beta-\bar{\theta}_\beta)B_n(\alpha,\beta;\bar{\theta})\Big|_{H_j}\right]} \cdot$$

$$\cdot \frac{\exp\left[\frac{1}{2}\sum_{\alpha,\beta=1}^{p}\{A_n(\alpha;\bar{\theta}) - c(\sum_{\substack{\gamma=p+1\\ \neq p+\ell}}^{p+k} B_n(\alpha,\gamma;\bar{\theta}) - B_n(\alpha,p+\ell;\bar{\theta}))\}F_n(\alpha,\beta;\bar{\theta})\cdot \{A_n(\beta;\bar{\theta}) - c(\sum_{\substack{\gamma=p+1\\ \neq p+\ell}}^{p+k} B_n(\beta,\gamma;\bar{\theta}) - B_n(\beta,p+\ell;\bar{\theta}))\}\right]}{\exp\left[\frac{1}{2}\sum_{\alpha,\beta=1}^{p}\{A_n(\alpha;\bar{\theta}) - c(\sum_{\substack{\gamma=p+1\\ \neq p+j}}^{p+k} B_n(\alpha,\gamma;\bar{\theta}) - B_n(\alpha,p+j;\bar{\theta}))\}F_n(\alpha,\beta;\bar{\theta})\cdot \{A_n(\beta;\bar{\theta}) - c(\sum_{\substack{\gamma=p+1\\ \neq p+j}}^{p+k} B_n(\beta,\gamma;\bar{\theta}) - B_n(\beta,p+j;\bar{\theta}))\}\right]}$$

After some simplification, we get

(2.28)

$$J_n''(\ell|j) =$$

$$\exp\left[2cA_n(p+j;\theta) - 2cA_n(p+\ell;\bar{\theta}) + 2c \sum_{\alpha,\beta=1}^{p} A_n(\alpha;\bar{\theta})F_n(\alpha,\beta;\bar{\theta}) \cdot \left\{B_n(\beta,p+\ell;\bar{\theta}) - B_n(\beta,p+j;\bar{\theta})\right\} \right]$$

$$\exp[c^2 V_n(\bar{\theta};\ell|j)]$$

where,

$$c^2 V_n(\bar{\theta};\ell|j) = \frac{1}{2} \sum_{\alpha,\beta=p+1}^{p+k} \sqrt{n}(\theta_\alpha - \bar{\theta}_\alpha)\sqrt{n}(\theta_\beta - \bar{\theta}_\beta) B_n(\alpha,\beta;\bar{\theta})\Big|_{H_j}$$

$$- \frac{1}{2} \sum_{\alpha,\beta=p+1}^{p+k} \sqrt{n}(\theta_\alpha - \bar{\theta}_\alpha)\sqrt{n}(\theta_\beta - \bar{\theta}_\beta) B_n(\alpha,\beta;\bar{\theta})\Big|_{H_\ell}$$

(2.29)

$$+ \frac{c^2}{2} \sum_{\alpha,\beta=1}^{p} F_n(\alpha,\beta;\bar{\theta}) \left[\begin{array}{l} \{ \sum_{\gamma=p+1}^{p+k} B_n(\alpha,\gamma;\bar{\theta}) - 2B_n(\alpha,p+j;\bar{\theta}) \} \{ \sum_{\gamma=p+1}^{p+k} B_n(\beta,\gamma;\bar{\theta}) - 2B_n(\beta,p+j;\bar{\theta}) \} \\ - \{ \sum_{\gamma=p+1}^{p+k} B_n(\alpha,\gamma;\bar{\theta}) - 2B_n(\alpha,p+\ell;\bar{\theta}) \} \{ \sum_{\gamma=p+1}^{p+k} B_n(\beta,\gamma;\bar{\theta}) - 2B_n(\beta,p+\ell;\bar{\theta}) \} \end{array} \right]$$

After some simplification, we get, for $\ell, j = 1,2,\ldots,k$,

(2.30)

$$V_n(\bar{\theta};\ell|j) =$$

$$\frac{1}{2}\left[\begin{array}{l} \sum\limits_{\substack{\alpha=p+1 \\ \neq p+j}}^{p+k}\ \sum\limits_{\substack{\beta=p+1 \\ \neq p+j}}^{p+k} B_n(\alpha,\beta;\bar{\theta}) - 2\sum\limits_{\substack{\beta=p+1 \\ \neq p+j}}^{p+k} B_n(\beta,p+j;\bar{\theta}) + B_n(p+j;p+j;\bar{\theta}) \\ -\sum\limits_{\substack{\alpha=p+1 \\ \neq p+\ell}}^{p+k}\ \sum\limits_{\substack{\beta=p+1 \\ \neq p+\ell}}^{p+k} B_n(\alpha,\beta;\bar{\theta}) + 2\sum\limits_{\substack{\beta=p+1 \\ \neq p+\ell}}^{p+k} B_n(\beta,p+\ell;\bar{\theta}) - B_n(p+\ell;p+\ell;\bar{\theta}) \end{array}\right]$$

$$+2\sum_{\alpha,\beta=1}^{p} F_n(\alpha,\beta;\bar{\theta}) \left[\begin{array}{l} \{\sum\limits_{\gamma=p+1}^{p+k} B_n(\alpha,\gamma;\bar{\theta})\}\cdot\{B_n(\beta,p+\ell;\bar{\theta}) - B_n(\beta,p+j;\bar{\theta})\} \\ -B_n(\alpha,p+\ell;\bar{\theta})B_n(\beta,p+\ell;\bar{\theta}) + B_n(\alpha,p+j;\bar{\theta})B_n(\beta,p+j;\bar{\theta}) \end{array}\right]$$

We now proceed to examine the asymptotic properties of $J_n'(\ell|j)$. A lemma, which will be useful in determining the asymptotic distribution of $J_n'(\ell|j)$ is first proved.

For $\alpha = 1,2,\ldots,p+k$, let

$$(2.31) \qquad \bar{A}_n(\alpha;\bar{\theta}) = A_n(\alpha;\bar{\theta}) - \sum_{\beta=1}^{p+k} \sqrt{n}(\omega_\beta(n)-\bar{\theta}_\beta)B(\alpha,\beta;\bar{\theta})$$

Lemma 2.3 If the true parameter point for X is $\omega(n)$, $\omega(n) \in R^n_{M_n(\bar{A})}(\bar{\theta})$, then $\bar{A}_n(1;\bar{\theta}),\bar{A}_n(2,\bar{\theta}),\ldots,\bar{A}_n(p+k;\bar{\theta})$ have asymptotically a joint normal distribution, with zero means and covariance between $\bar{A}_n(\alpha;\bar{\theta})$ and $\bar{A}_n(\beta;\bar{\theta})$ given by $B(\alpha,\beta;\bar{\theta})$.

Proof: If $\omega(n)$, $\omega(n) \in R^n_{M_n(\bar{\theta})}(\bar{\theta})$, is the true parameter point for X, then for $\alpha = 1,2,\ldots,p+k$,

$$E_{\omega(n)}[A_n(\alpha;\bar{\theta})] = E_{\omega(n)}\left[\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial}{\partial\theta_\alpha} \log f(X_i;\theta)\Big|_{\bar{\theta}}\right]$$

$$= \sqrt{n} \int_{-\infty}^{\infty} f(x,\omega(n)) \frac{\partial}{\partial\theta_\alpha} \log f(x;\theta)\Big|_{\bar{\theta}} dx \tag{2.32}$$

$$= \sqrt{n} \int_{-\infty}^{\infty} f(x;\bar{\theta} + (\omega(n) - \bar{\theta})) \frac{\partial}{\partial\theta_\alpha} \log f(x;\theta)\Big|_{\bar{\theta}} dx$$

Expanding $f(x;\omega(n))$ around $\bar{\theta}$ and denoting, for notational convenience, $\frac{\partial}{\partial\theta_\alpha} \log f(x;\theta)\Big|_{\bar{\theta}}$ by $\frac{\partial}{\partial\theta_\alpha} \log f(x;\bar{\theta})$, we obtain for $\alpha = 1,2,\ldots,p+k$,

$$E_{\omega(n)}[A_n(\alpha;\bar{\theta})] = \sqrt{n} \int_{-\infty}^{\infty} f(x;\bar{\theta}) \frac{\partial}{\partial\theta_\alpha} \log f(x;\bar{\theta})dx$$

$$+ \sum_{\beta=1}^{p+k} \sqrt{n}(\omega_\beta(n)-\bar{\theta}_\beta) \int_{-\infty}^{\infty} \frac{\partial}{\partial\theta_\beta} f(x;\bar{\theta}) \frac{\partial}{\partial\theta_\alpha} \log f(x;\bar{\theta})dx$$

$$+ \frac{\sqrt{n}}{2} \sum_{\beta,\gamma=1}^{p+k} (\omega_\beta(n)-\bar{\theta}_\beta)(\omega_\gamma(n)-\bar{\theta}_\gamma) \int_{-\infty}^{\infty} \frac{\partial^2 f(x;\mu(n))}{\partial\theta_\beta \partial\theta_\gamma} \frac{\partial}{\partial\theta_\alpha} \log f(x;\bar{\theta})dx \tag{2.33}$$

where $\mu(n) \in R^n_{M_n(\bar{\theta})}(\bar{\theta})$.

Thus,

$$E_{\omega(n)}[A_n(\alpha;\bar{\theta})] = \sum_{\beta=1}^{p+k} \sqrt{n}(\omega_\beta(n)-\bar{\theta}_\beta) \int_{-\infty}^{\infty} \frac{\partial}{\partial\theta_\beta} \log f(x;\bar{\theta}) \frac{\partial}{\partial\theta_\alpha} \log f(x;\bar{\theta}) f(x;\bar{\theta})dx$$

$$+ \frac{\sqrt{n}}{2} \sum_{\beta,\gamma=1}^{p+k} (\omega_\beta(n)-\bar{\theta}_\beta)(\omega_\gamma(n)-\bar{\theta}_\gamma) \int_{-\infty}^{\infty} \frac{\partial^2 f(x;\mu(n))}{\partial\theta_\beta \partial\theta_\gamma} \frac{\partial}{\partial\theta_\alpha} \log f(x;\bar{\theta})dx$$

or,

$$E_{\omega(n)}[A_n(\alpha;\bar{\theta})] = \sum_{\beta=1}^{p+k} \sqrt{n}(\omega_\beta(n)-\bar{\theta}_\beta)B(\alpha,\beta;\bar{\theta})$$

(2.34)

$$+ \frac{\sqrt{n}}{2} \sum_{\beta,\gamma=1}^{p+k} (\omega_\beta(n)-\bar{\theta}_\beta)(\omega_\gamma(n)-\bar{\theta}_\gamma) \int_{-\infty}^{\infty} \frac{\partial^2 f(x;\mu(n))}{\partial\theta_\beta\,\partial\theta_\gamma} \frac{\partial}{\partial\theta_\alpha} \log f(x;\bar{\theta})dx$$

where $\mu(n) \in R^n_{M_n(\bar{\theta})}(\bar{\theta})$.

From (2.34) and the assumptions made in Section 2.1, it follows that

$$\left| E_{\omega(n)}\left[\bar{A}_n(\alpha;\bar{\theta}) - \sum_{\beta=1}^{p+k} \sqrt{n}(\omega_\beta(n)-\bar{\theta}_\beta)B(\alpha,\beta;\bar{\theta})\right]\right|$$

(2.35)

$$< K(\bar{\theta}) \frac{M_n(\bar{\theta})}{n^{1/2}}$$

where $K(\bar{\theta})$ is a fixed positive constant. Thus from (2.35), it follows that for $\alpha = 1,2,\ldots,p+k$, $E_{\omega(n)}\bar{A}_n(\alpha;\bar{\theta})$ converges to zero.

Also note that, for $\alpha,\beta = 1,2,\ldots,p+k$,

$$\mathrm{Cov}_{\omega(n)}[\bar{A}_n(\alpha;\bar{\theta}),\bar{A}_n(\beta;\bar{\theta})] = \mathrm{Covar}_{\omega(n)}[A_n(\alpha;\bar{\theta}),A_n(\beta;\bar{\theta})]$$

$$= E_{\omega(n)}[A_n(\alpha;\theta)A_n(\beta;\bar{\theta})] - E_{\omega(n)}[A_n(\alpha;\bar{\theta})]E_{\omega(n)}[A_n(\beta;\bar{\theta})]$$

$$= \int_{-\infty}^{\infty} \frac{\partial}{\partial\theta_\alpha} \log f(x;\bar{\theta}) \frac{\partial}{\partial\theta_\beta} \log f(x;\bar{\theta})\, f(x;\omega(n))dx$$

$$- E_{\omega(n)}(A_n(\alpha;\bar{\theta}))E_{\omega(n)}(A_n(\beta;\bar{\theta}))$$

Using algebraic simplification as for $L_{\omega(n)}[A_n(a;\bar{\theta})]$, and following the assumptions made in Section 2.1, it can be shown that the covariance between $\bar{A}_n(a;\theta)$ and $\bar{A}_n(\beta;\theta)$ converges to $B(a,\beta;\bar{\theta})$ for all $\omega(n) \in K^n_{M_n(\bar{\theta})}(\bar{\theta})$.

The asymptotic joint normality of $A_n(1;\bar{\theta}), A_n(2;\bar{\theta}), \ldots, A_n(p+k;\bar{\theta})$ is a standard result used frequently in the literature (See, for example, p. 500 of Cramer [13]). Using that, the proof of the lemma is hence complete.

Q.E.D.

Substituting from (2.31) in (2.26), we obtain, for $i,j = 1,2,\ldots,k$,

(2.36)

$$J_n^*(i|j) =$$

$$\frac{\exp\left\{-2cA_n(p+i;\bar{\theta}) - \frac{1}{2}\sum_{a,\beta=p+1}^{p+k}\sqrt{n}(\theta_a-\bar{\theta}_a)\sqrt{n}(\theta_\beta-\bar{\theta}_\beta)B_n(a,\beta;\bar{\theta})\Big|_{H_i}\right\}}{\exp\left\{-2cA_n(p+j;\bar{\theta}) - \frac{1}{2}\sum_{a,\beta=p+1}^{p+k}\sqrt{n}(\theta_a-\bar{\theta}_a)\sqrt{n}(\theta_\beta-\bar{\theta}_\beta)B_n(a,\beta;\bar{\theta})\Big|_{H_j}\right\}} \cdot$$

$$\frac{\int_{-M_n(\bar{\theta})}^{M_n(\bar{\theta})}\cdots\int_{-M_n(\bar{\theta})}^{M_n(\bar{\theta})} \exp\left(\sum_{a=1}^{p} w_a\bar{A}_n(a;\bar{\theta}) + \sum_{\gamma=1}^{p+k}\sqrt{n}(\omega_\gamma(n)-\bar{\theta}_\gamma)B(a,\gamma;\bar{\theta})\right.}{\left. -c\left[\sum_{\gamma=p+1}^{p+k} B_n(a,\gamma;\bar{\theta}) - 2B_n(a,p+i;\bar{\theta})\right] - \frac{1}{2}\sum_{a,\beta=1}^{p} w_a w_\beta B_n(a,\beta;\bar{\theta})\right) dw_1,\ldots,dw_p}$$

$$\int_{-M_n(\bar{\theta})}^{M_n(\bar{\theta})}\cdots\int_{-M_n(\bar{\theta})}^{M_n(\bar{\theta})} \exp\left(\sum_{a=1}^{p} w_a\bar{A}_n(a;\bar{\theta}) + \sum_{\gamma=1}^{p+k}\sqrt{n}(\omega_\gamma(n)-\bar{\theta}_\gamma)B(a,\gamma;\bar{\theta})\right.$$

$$\left. -c\left[\sum_{\gamma=p+1}^{p+k} B_n(a,\gamma;\bar{\theta}) - 2B_n(a,p+j;\bar{\theta})\right] - \frac{1}{2}\sum_{a,\beta=1}^{p} w_a w_\beta B_n(a,\beta;\bar{\theta})\right) dw_1,\ldots,dw_p$$

Rearranging terms, we get

$$J_n'(\ell|j) =$$

$$\frac{\exp\left[-2cA_n(p+\ell,\bar{\theta}) - \frac{1}{2}\sum_{\alpha,\beta=p+1}^{p+k}\sqrt{n}(\theta_\alpha-\bar{\theta}_\alpha)\sqrt{n}(\theta_\beta-\bar{\theta}_\beta)B_n(\alpha,\beta;\bar{\theta})\Big|_{H_\ell}\right]}{\exp\left[-2cA_n(p+j;\bar{\theta}) - \frac{1}{2}\sum_{\alpha,\beta=p+1}^{p+k}\sqrt{n}(\theta_\alpha-\bar{\theta}_\alpha)\sqrt{n}(\theta_\beta-\bar{\theta}_\beta)B_n(\alpha,\beta;\bar{\theta})\Big|_{H_j}\right]} \cdot$$

$$\frac{\int\cdots\int \exp\left[-\frac{1}{2}\sum_{\alpha,\beta=1}^{p}(w_\alpha-t_{\alpha,\ell}(n))(w_\beta-t_{\beta,\ell}(n))B_n(\alpha,\beta;\bar{\theta})\right] dw_1,\ldots,dw_p}{\int\cdots\int \exp\left[-\frac{1}{2}\sum_{\alpha,\beta=1}^{p}(w_\alpha-t_{\alpha,j}(n))(w_\beta-t_{\beta,j}(n))B_n(\alpha,\beta;\bar{\theta})\right] dw_1,\ldots,dw_p}$$

where, for $\alpha = 1,2,\ldots,k$, $\ell = 1,2,\ldots,k$,

$$t_{\alpha,\ell}(n) = \sum_{\beta=1}^{p} F_n(\alpha,\beta;\bar{\theta})\left[\bar{A}_n(r;\bar{\theta}) + \sum_{\gamma=1}^{p+k}\sqrt{n}(w_\gamma(n)-\bar{\theta}_\gamma)B(\beta,\gamma;\bar{\theta}) - c\left[\sum_{\substack{\gamma=p+1\\ \gamma\neq p+\ell}}^{p+k} B_n(\beta,\gamma;\bar{\theta}) - B_n(\alpha,p+\ell;\bar{\theta})\right]\right] \tag{2.38}$$

and

$$A_i(n) = \frac{1}{2}\sum_{\alpha,\beta=1}^{p} t_{\alpha,i}(n)t_{\beta,i}(n)B_n(\alpha,\beta;\bar{\theta}) \tag{2.39}$$

If we let $u_\alpha = w_\alpha - t_{\alpha,i}(n)$ and $v_\alpha = w_\alpha - t_{\alpha,j}(n)$, then after some simplification, we obtain, for $i,j = 1,2,\ldots,k$,

(2.40)

$$J_n'(\ell|j) =$$

$$J_n''(\ell|j) \quad \frac{\int_{-M_n(\bar{\theta})-t_{1,\ell}(n)}^{M_n(\bar{\theta})-t_{1,\ell}(n)} \cdots \int_{-M_n(\bar{\theta})-t_{p,\ell}(n)}^{M_n(\bar{\theta})-t_{p,\ell}(n)} \exp\left[-\frac{1}{2}\sum_{\alpha,\beta=1}^{p} u_\alpha u_\beta B_n(\alpha,\beta;\bar{\theta})\right] du_1,\ldots,du_p}{\int_{-M_n(\bar{\theta})-t_{1,j}(n)}^{M_n(\bar{\theta})-t_{1,j}(n)} \cdots \int_{-M_n(\bar{\theta})-t_{p,j}(n)}^{M_n(\bar{\theta})-t_{p,j}(n)} \exp\left[-\frac{1}{2}\sum_{\alpha,\beta=1}^{p} v_\alpha v_\beta B_n(\alpha,\beta;\bar{\theta})\right] dv_1,\ldots,dv_p}$$

In order to show that $J_n'(\ell|j)$ is arbitrarily close to $J_n''(\ell|j)$, it is sufficient to show that the integrals in the numerator and denominator of (2.4) are, for sufficiently large n, with any probability less than one, within an arbitrary positive constant ϵ of their common limit $(2\pi)^{p/2}\,|B|^{-1/2}$, where $|B|$ is the determinant of the matrix B, whose $(\alpha,\beta)^{th}$ element is $B(\alpha,\beta;\bar{\theta})$. For that it suffices to show that the limits of integration converge stochastically to $(-\infty,\infty)$ as $n \to \infty$.

This will be sufficient, because then by replacing the limits of integration by $(-\infty,\infty)$, the effect on the numerator of $J_n'(\ell|j)$ is to multiply it by $q_n(\ell)$ where $q_n(\ell)$ converges stochastically to one. Similarly, the effect in the denominator is to multiply it by $q_n(j)$, where $q_n(j)$ converges stochastically to one.

From (2.38), we obtain for $\alpha = 1,2,\ldots,p$, $\ell = 1,2,\ldots,k$,

$$t_{\alpha,\ell}(n) = \sum_{\beta=1}^{p} F_n(\alpha,\beta;\bar{\theta}) \left[\bar{A}_n(\beta;\bar{\theta}) + \sum_{\gamma=1}^{p+k} \sqrt{n}(\omega_\gamma(n)-\bar{\theta}_\gamma) B(\beta,\gamma;\bar{\theta}) \right.$$
$$\left. -c\left[\sum_{\substack{\gamma=p+1 \\ \neq p+\ell}}^{p+k} B_n(\beta,\gamma;\bar{\theta}) - B_n(\beta,p+\ell;\bar{\theta}) \right] \right]$$
$$= t'_{\alpha,\ell} + \sum_{\gamma=1}^{p+k} \sqrt{n}(\omega_\gamma(n)-\bar{\theta}_\gamma) \sum_{\beta=1}^{p} [F_n(\alpha,\beta;\bar{\theta}) B(\beta,\gamma;\bar{\theta})]$$

or,

$$t_{\alpha,\ell}(n) = t'_{\alpha,\ell} + \sum_{\gamma=1}^{p+k} \sqrt{n}(\omega_\gamma(n)-\bar{\theta}_\gamma)\delta_n(\alpha,\gamma) \tag{2.41}$$

where,

$$\delta_n(\alpha,\gamma) = \sum_{\beta=1}^{k} F_n(\alpha,\beta;\bar{\theta}) B(\beta,\gamma;\bar{\theta}) \tag{2.42}$$

and $t'_{\alpha,\ell}$ remains bounded with probability approaching one as $n \to \infty$.

Since $F_n(\alpha,\beta;\bar{\theta})$ converges stochastically to $F(\alpha,\beta;\bar{\theta})$ $(\alpha,\beta = 1,2,..,p)$ as n increases, it follows that

$$\delta_n(\alpha,\gamma) \xrightarrow[n\to\infty]{\text{stoch.}} \begin{cases} 0 & \alpha \neq \gamma \\ 1 & \alpha = \gamma \end{cases} \qquad \gamma = 1,2,\ldots,p. \tag{2.43}$$

It also follows that $\delta_n(\alpha, p+\ell)$, $\ell = 1,2,\ldots,k$ remains bounded with probability approaching one as $n \to \infty$.

Thus, if $\omega(n) \in R^n_{L_n(\bar{\theta})}(\bar{\theta})$, we obtain from (2.41), for $\alpha = 1,2,..,p$, $\ell = 1,2,...,k$,

$$(2.44) \quad |t_{\alpha,\ell}(n)| < |t'_{\alpha,\ell}| + L \cdot \sum_{\ell=1}^{k} |\delta(\alpha,p+\ell)| + L_n(\bar{\theta}) \sum_{\gamma=1}^{p} |\delta_n(\alpha,\gamma)|$$

where $\sum_{\gamma=1}^{p} |\delta_n(\alpha,\gamma)|$ converges stochastically to one as $n \to \infty$, for $\alpha = 1,2,...,p$. This implies that there exists a sequence $\{\epsilon_n\}$, $\epsilon_n > 0$ $\lim_{n\to\infty} \epsilon_n = 0$, such that,

$$(2.45) \quad \lim_{n\to\infty} P\{ \sum_{\gamma=1}^{p} |\delta_n(\alpha,\gamma)| < 1 + \epsilon_n \quad \alpha = 1,2,...,p\} = 1 .$$

From (2.44) and (2.45), we thus get that for any given $\phi > 0$, there exists $K(\phi) < \infty$, such that

$$(2.46) \quad P_{\omega(n)}\{|t_{\alpha,\ell}(n)| < K(\phi) + L_n(\bar{\theta})(1+\epsilon_n)\} > 1 - \phi$$

for any $\omega(n) \in R^n_{L_n(\bar{\theta})}(\bar{\theta})$.

Choose the sequence $\{L_n(\bar{\theta})\}$ to satisfy the following properties.

$$(2.47) \quad \lim_{n\to\infty} L_n(\bar{\theta}) = \infty$$

$$(2.48) \quad \frac{L_n(\bar{\theta})}{M_n(\bar{\theta})} < 1 \quad \text{for all } n \quad \text{and} \quad \lim_{n\to\infty} \frac{L_n(\bar{\theta})}{M_n(\bar{\theta})} = 1$$

and

$$(2.49) \quad \lim_{n\to\infty} [M_n(\bar{\theta}) - (1+\epsilon_n)L_n(\bar{\theta})] = \infty .$$

That it is not difficult to find such a sequence is illustrated by the simple example $L_n(\bar{\theta}) = (1+\epsilon_n)^{-1}[M_n(\bar{\theta}) - M_n^{1/2}(\bar{\theta})]$. Using (2.46) and the sequence $\{L_n(\bar{\theta})\}$ as selected above, we get that with probability greater than $1 - \phi$,

$$-M_n(\bar{\theta}) - t_{\alpha,\ell}(n) < -M_n(\bar{\theta}) + K(\phi) + L_n(\bar{\theta})(1+\epsilon_n)$$

$$= -[M_n(\bar{\theta}) - (1+\epsilon_n)L_n(\bar{\theta})] + K(\phi) \tag{2.50}$$

$$\to -\infty \quad \text{as} \quad n \to \infty,$$

and,

$$M_n(\bar{\theta}) - t_{\alpha,\ell}(n) > M_n(\bar{\theta}) - K(\phi) - L_n(\bar{\theta})(1+\epsilon_n) \tag{2.51}$$

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

Using the above result and lemma 2.2, we have thus proved the following.

Lemma 2.4 For $\ell, j = 1,2,\ldots,k$.

$$\log J_n(\ell|j) = \log J_n''(\ell|j) + Z_n(\ell|j) \tag{2.52}$$

where $Z_n(\ell|j)$ converges stochastically to zero as $n \to \infty$ for all parameter points $\omega(n)$, $\omega(n) \in R^n_{L_n(\bar{\theta})}(\bar{\theta})$.

As a consequence of the above lemma, in order to study asymptotic behavior of $J_n(\ell|j)$, we need only study the asymptotic behavior of $J''_n(\ell|j)$.

Using (2.28) and lemma 2.4, the Bayes decision rule for the problem reduces to the following:

For $\ell = 1,2,\ldots,k$, select H_ℓ if

$$2c(A_n(p+j;\theta) - A_n(p+\ell;\bar{\theta})) + 2c \sum_{\alpha,\beta=1}^{p} A_n(\alpha;\bar{\theta})F_n(\alpha,\beta;\bar{\theta})\left(B_n(\beta,p+\ell;\bar{\theta}) - B_n(\beta,p+j;\theta)\right)$$

$$+ c^2 V_n(\bar{\theta};\ell|j) + Z_n(\ell|j) \geq \log \frac{b_j}{b_\ell} \quad j = 1,2,\ldots,k \tag{2.53}$$

where $v_n(\bar{\theta};\ell|j)$ is given by (2.30).

The Bayes decision rule divides the sample space into k mutually exclusive and exhaustive regions, each of which a particular hypothesis is selected. This is so irrespective of whether $f(\cdot;\cdot)$ represents a joint density function or a joint probability mass function. Using the regularity conditions imposed on $f(\cdot;\cdot)$, we do not have to worry about the ties introduced into the problem due to equalities in relation (2.53). Thus the results obtained in this section are not restricted to assuming the existence of a joint probability density function for the joint distribution of the random variables.

For $\ell,j = 1,2,\ldots,k$, let

$$W_n(\ell|j) = A_n(p+j;\bar{\theta}) - A_n(p+\ell;\bar{\theta}) + \sum_{\alpha,\beta=1}^{p} A_n(\alpha;\bar{\theta})F_n(\alpha,\beta;\bar{\theta}) \cdot$$

(2.54)

$$\cdot\left(B_n(\beta,p+\ell;\bar{\theta}) - B_n(\beta,p+j;\bar{\theta})\right)$$

Rearranging terms in (2.53) and using (2.54), the Bayes decision rule reduces to the following:

For $\ell = 1,2,\ldots,k$, select H_ℓ if

$$W_n(\ell|j) \geq \frac{1}{2c}\left\{\log \frac{b_j}{b_\ell} - c^2 V(\bar{\theta};\ell|j) + Z_n'(\ell|j)\right\} \quad j = 1,2,\ldots,k, \tag{2.55}$$

where, $Z_n'(\ell|j)$ converges stochastically to zero as $n \to \infty$, if the true parameter point $\omega(n) \in R^n_{L_n(\bar{\theta})}(\bar{\theta})$. Also $V(\bar{\theta};\ell|j)$ is a non random continuous function of $\bar{\theta}$ for all ℓ,j $(\ell,j = 1,2,\ldots,k)$. $V(\bar{\theta};\ell|j)$ is obtained from (2.30) with the random variables in (2.30) replaced by the constants to which they converge. Because of the symmetry introduced in the problem by redefining the parameters ψ by (2.1), we may conclude for the problem of ranking from several univariate or multivariate populations, when the observations between populations are independent, that

$$V(\bar{\theta};\ell|j) = 0 \quad \text{for all } \ell,j \quad (\ell,j = 1,2,\ldots,k) \tag{2.56}$$

In obtaining the above result, we have used the following realtions

$$\begin{aligned} B(\alpha,\beta;\bar{\theta}) &= B(\alpha,\gamma;\bar{\theta}) \qquad \alpha = 1,2,\ldots,p \\ B(\beta,\beta;\bar{\theta}) &= B(\gamma,\gamma;\bar{\theta}) \qquad \beta,\gamma = p+1,p+2,\ldots,p+k \end{aligned} \tag{2.57}$$

and

$$(2.58) \qquad B(\beta,\gamma;\bar{\theta}) = 0 \qquad \beta \neq \gamma, \quad \beta,\gamma = p+1,p+2,\ldots,p+k$$

If for the problem of ranking from a single multivariate problem (for which (2.57) holds but not (2.58)), we assume that

$$(2.59) \qquad B(\beta,\gamma;\bar{\theta}) = B(\delta,\eta;\bar{\theta}) \quad \beta \neq \gamma, \quad \delta \neq \eta, \quad \beta,\gamma,\delta,\eta = p+1,p+2,..,p+k$$

then (2.56) still holds. From now on we assume that this is so and hence the results are applicable in ranking from a single multivariate population for cases for which (2.59) holds.

Thus the Bayes decision rule may be rewritten as follows:

For $\ell = 1,2,\ldots,k$, select H_ℓ if

$$(2.60) \qquad W_n(\ell|j) \geq \frac{1}{2c}\left(\log \frac{b_j}{b_\ell} + Z'_n(\ell|j)\right) \qquad j = 1,2,\ldots,k.$$

To study the asymptotic behavior of the Bayes decision rule, we proceed to study the asymptotic joint distribution of the random variables $W_n(\ell|j)$. The asymptotic distribution is given by the following theorem.

Theorem 2.3 If $\omega(n)$ is the true parameter point for X, $\omega(n) \in R^n_{L_n(\bar{\theta})}(\bar{\theta})$, then for fixed ℓ ($\ell = 1,2,\ldots,k$), $\{W_n(\ell|j);\ j = 1,2,\ldots,k,\ j \neq \ell\}$ have asymptotically a joint normal distribution with mean and covariance matrix given by

$$EW_n(\ell|j) = \sum_{\gamma=p+1}^{p+k} \sqrt{n}(\omega_\gamma(n)-\bar{\theta}_\gamma)[g(p+j;\gamma;\bar{\theta}) - g(p+\ell;\gamma;\bar{\theta})] \quad j \neq \ell$$

and

$$\mathrm{Cov}(W_n(\ell|j),W_n(\ell|i)) = s(\bar{\theta};\ell,i,j) \quad i,j \neq \ell$$

where $g(\ )$ and $s(\ )$ are continuous known functions of $\bar{\theta}$ ($g(\cdot)$ and $s(\cdot)$ defined later on in the proof).

Proof: For each fixed ℓ, $W_n(\ell|j)$ are linear combinations of $A_n(\alpha;\theta)$, $\alpha = 1,2,\ldots,p+k$; it follows from Lemma 2.3 and a standard result in literature on linear combination of random variables with asymptotic joint multivariate normal distribution (see for example Rao [40]) that for fixed ℓ, $\{W_n(\ell|j);\ j \neq \ell\}$ have an asymptotic joint normal distribution. It only remains to determine the mean and covariance matrix.

Rewriting $W_n(\ell|j)$ in terms of $\bar{A}_n(\alpha;\bar{\theta})$, we get from (2.71) and (2.54),

$$W_n(\ell|j) =$$

$$\left\{ \bar{A}_n(p+j;\bar{\theta}) + \sum_{\gamma=1}^{p+k} \sqrt{n}(\omega_\gamma(n)-\bar{\theta}_\gamma)B(p+j;\gamma;\bar{\theta}) - \sum_{\alpha,\beta=1}^{p} F_n(\alpha,\beta;\bar{\theta})B_n(\beta,p+j;\bar{\theta}) \right.$$
$$\left. \cdot (\bar{A}_n(\alpha;\bar{\theta}) + \sum_{\gamma=1}^{p+k} \sqrt{n}(\omega_\gamma(n)-\bar{\theta}_\gamma)B(\alpha,\gamma;\bar{\theta})) \right\}$$

$$- \left\{ \bar{A}_n(p+\ell;\bar{\theta}) + \sum_{\gamma=1}^{p+k} \sqrt{n}(\omega_\gamma(n)-\bar{\theta}_\gamma)B(p+\ell,\gamma;\bar{\theta}) - \sum_{\alpha,\beta=1}^{p} F_n(\alpha,\beta;\bar{\theta})B_n(\beta,p+\ell;\bar{\theta}) \right.$$
$$\left. \cdot (\bar{A}_n(\alpha;\bar{\theta}) + \sum_{\gamma=1}^{p+k} \sqrt{n}(\omega_\gamma(n)-\bar{\theta}_\gamma)B(\alpha,\gamma;\bar{\theta})) \right\}$$

or,

$$W_n(\ell|j) = \bar{W}_n(\ell|j) + W_n^{(1)}(\ell|j) + W_n^{(2)}(\ell|j) \tag{2.61}$$

where,

$$\begin{aligned}\bar{W}_n(\ell|j) = &\{\bar{A}_n(p+j;\bar{\theta}) - \sum_{\alpha,\beta=1}^{p} F_n(\alpha,\beta;\bar{\theta})\bar{A}_n(\alpha;\bar{\theta})B_n(\beta,p+j;\bar{\theta})\} \\ &-\{\bar{A}_n(p+\ell;\bar{\theta}) - \sum_{\alpha,\beta=1}^{p} \bar{A}_n(\alpha;\bar{\theta})F_n(\alpha,\beta;\bar{\theta})B_n(\beta,p+\ell;\bar{\theta})\}\end{aligned} \tag{2.62}$$

(2.63)

$$\begin{aligned}W_n^{(1)}(\ell|j) = &\sum_{\gamma=p+1}^{p+k} \sqrt{n}(\omega_\gamma(n)-\bar{\theta}_\gamma)\{B(p+j;\gamma;\bar{\theta}) - \sum_{\alpha,\beta=1}^{p} B(\alpha,\gamma;\bar{\theta})F_n(\alpha,\beta;\bar{\theta})B_n(\beta,p+j;\bar{\theta})\} \\ &- \sum_{\gamma=p+1}^{p+k} \sqrt{n}(\omega_\gamma(n)-\bar{\theta}_\gamma)\{B(p+\ell,\gamma;\bar{\theta}) - \sum_{\alpha,\beta=1}^{p} B(\alpha,\gamma;\bar{\theta})F_n(\alpha,\beta;\bar{\theta})B_n(\beta,p+\ell;\bar{\theta})\}\end{aligned}$$

and

(2.64)

$$W_n^{(2)}(\ell|j) = \sum_{\gamma=1}^{p} \sqrt{n}(\omega_\gamma(n)-\bar{\theta}_\gamma)\left\{\begin{array}{l}[B(p+j,\gamma;\bar{\theta}) - \sum_{\alpha,\beta=1}^{p} B(\alpha,\gamma;\bar{\theta})F_n(\alpha,\beta;\bar{\theta})B_n(\beta,p+j;\bar{\theta})] \\ -[B(p+\ell,\gamma;\bar{\theta}) - \sum_{\alpha,\beta=1}^{p} B(\alpha,\gamma;\bar{\theta})F_n(\alpha,\beta;\bar{\theta})B_n(\beta,p+\ell;\bar{\theta})]\end{array}\right\}$$

From (2.42) and (2.43), we see that if $\omega(n)$ is the true parameter point for X, $\omega(n) \in R^n_{M_n(\bar{\theta})}(\bar{\theta})$, then for all ℓ,γ ($\ell,\gamma = 1,2,\ldots,k$), $(B(p+\ell,\gamma;\bar{\theta}) - \sum_{\alpha,\beta=1}^{k} B(\alpha,\gamma;\bar{\theta})F_n(\alpha,\beta;\bar{\theta})B_n(\beta,p+\ell;\bar{\theta}))$ converges stochastically to zero as n increases.

Clearly, if we choose a sequence $\{M_n(\bar{\theta})\}$ converging to infinity such that $\sum_{\gamma=1}^{p} M_n(\bar{\theta})|\{B(p+\ell,\gamma;\bar{\theta}) - \sum_{\alpha,\beta=1}^{p} B(\alpha,\gamma;\bar{\theta})F_n(\alpha,\beta;\bar{\theta})B_n(\beta,p+\ell;\bar{\theta})\}|$ converges stochastically to zero as n increases, when $\omega(n)$ is the true parameter point for X and $\omega(n) \in R^n_{M_n(\bar{\theta})}(\bar{\theta})$; then

$$W_n^{(2)}(\ell|j) \xrightarrow[n \to \infty]{\text{stoch}} 0 \tag{2.65}$$

From now on we assume that the sequence $\{M_n(\bar{\theta})\}$ (on which we had placed no restriction except as given by (2.3)) is so chosen.

For $\ell = 1,2,\ldots,k$, let

$$g(p+\ell,\gamma;\bar{\theta}) = B(p+\ell,\gamma;\bar{\theta}) - \sum_{\alpha,\beta=1}^{p} B(\alpha,\gamma;\bar{\theta})F(\alpha,\beta;\bar{\theta})B(\beta,p+\ell;\bar{\theta}) \tag{2.66}$$

$$\gamma = p+1,p+2,\ldots,p+k \ .$$

Then from (2.61), (2.65)-(2.66), it follows that when the true parameter point for X, $\omega(n) \in R^n_{L_n(\bar{\theta})}(\bar{\theta})$, the asymptotic joint distribution of $W_n(\ell|j)$ is normal with mean of $W_n(\ell|j)$ as required.

For $\ell,j = 1,2,\ldots,k$, $j \neq \ell$, let

$$\bar{W}_n(\ell|j) = R_n(j) - R_n(\ell) \tag{2.67}$$

where, for $\ell = 1,2,\ldots,k$,

$$R_n(\ell) = \bar{A}_n(p+\ell;\bar{\theta}) - \sum_{\alpha,\beta=1}^{p} \bar{A}_n(\alpha;\bar{\theta})F_n(\alpha,\beta;\bar{\theta})B_n(\beta,k+\ell;\bar{\theta}) \tag{2.68}$$

Then, in view of results obtained above, asymptotically, we get

$$\begin{aligned} \mathrm{Cov}(W_n(\ell|j),W_n(\ell|i)) &= \mathrm{Cov}(\bar{W}_n(\ell|j),\bar{W}_n(\ell|i)) \\ &= E[R_n^2(\ell)] + E[R_n(i)R_n(j)] \\ &\quad -E[R_n(\ell)R_n(i)] - E[R_n(\ell)R_n(j)] \end{aligned}$$

Because of the symmetry in the problem, we get

$$\mathrm{Cov}(W_n(\ell|j),W_n(\ell|i)) = E[R_n^2(\ell)] - E[R_n(\ell)R_n(i)] \; . \tag{2.70}$$

For $\ell = 1,2,\ldots,k$, let

$$P_n(\alpha,p+\ell) = \sum_{\beta=1}^{p} F_n(\alpha,\beta;\bar{\theta})B_n(\beta,p+\ell;\bar{\theta}) \tag{2.71}$$

and,

$$P(\alpha,p+\ell) = \sum_{\beta=1}^{p} F(\alpha,\beta;\bar{\theta})B(\beta,p+\ell;\bar{\theta}) \; . \tag{2.72}$$

Then, using (2.67), (2.68), (2.71) and lemma 2.3, we get for $i \neq j$,

$$\begin{aligned} \mathrm{Cov}(W_n(\ell|j), W_n(\ell|i)) &= B(p+\ell, p+\ell; \bar{\theta}) - B(p+\ell, p+i; \bar{\theta}) \\ &= B(p+\ell, p+\ell; \bar{\theta}) - B(p+\ell, p+j); \bar{\theta} \end{aligned} \tag{2.73}$$

Denoting the right hand side of (2.73) by $s(\theta;\ell,i,j)$ completes the proof of the theorem.

Q.E.D.

Because of the inherent symmetry in the structure of the ranking problem and because of the way in which we have redefined our parameter (Eq. (2.1)), we note that the decision variables have asymptotically a normal distribution with a common correlation coefficient. This makes it easier to compute certain probability integrals needed in computing the Probability of Correct Selection and thus the Bayes risk for the problem. Tables of integrals for equicorrelated multivariate normal variables by Milton [34] and Gupta [21] can be used to obtain the Bayes risk for any particular prior distribution.

In order to determine an asymptotically minimax decision rule for the problem it follows immediately from Theorem 2.1, Theorem 2.3 and (2.60) that because of the symmetry in the problem, the Bayes decision rule given by (2.60) is an asymptotically minimax decision rule if we set, for $j = 1,2,\ldots,k$, $b_j = \frac{1}{k}$. Thus, an asymptotically minimax decision rule for the sequence of artificial k-decision problems is as given below:

For $\ell = 1,2,\ldots,k$, select H_ℓ if

$$W_n(\ell|j) \geq 0 \tag{2.73a}$$

where $W_n(\ell|j)$ is given by (2.54). The decision variables have a limiting equicorrelated multivariate normal distribution with mean and covariance matrix as given by Theorem 2.3. Recalling that $\frac{L_n(\bar{\theta})}{M_n(\bar{\theta})} \to 1$ as $n \to \infty$, the decision rule given by (2.73a) is an asymptotically minimax decision rule for the sequence of artificial k-decision problems.

2.4.2 Asymptotically Optimal Procedure for the "Real" Problem

The above sequence of problems was artificial because we assumed $\bar{\theta}_1,\bar{\theta}_2,\ldots,\bar{\theta}_p$ to be known. We now proceed to the real problem, where nothing is known about the values of the nuisance parameters $(\theta_1,\ldots,\theta_p)$; and with the loss function as given by (2.18), we want to develop an optimal procedure to select the best population.

We first define some additional notation. Let $\hat{\theta}_1(n),\hat{\theta}_2(n),\ldots,\hat{\theta}_p(n)$ be estimators based on X of $\theta_1,\ldots,\theta_p$, such that for any $\epsilon > 0$, there exists $D(\epsilon) < \infty$, such that

$$P_{\theta_1,\ldots,\theta_p,\ \{H \text{ is true}\}}\{\sqrt{n}|\hat{\theta}_i(n)-\theta_i| < D(\epsilon) \quad i = 1,2,\ldots,p\} > 1 - \epsilon \tag{2.74}$$

where {H is true} implies that one of the p hypotheses given by (2.17) is true.

It is worth noting here that the above condition on the estimators, is only a consistency condition, certainly desirable for any estimator. Maximum Probability Estimators of $(\theta_1,\ldots,\theta_p)$ could be used, although any set of consistent estimators satisfying (2.74) would suffice.

For notational convenience, let

$$(2.75)\qquad \hat{\theta}(n) = (\tilde{\theta}_0,\hat{\theta}_1(n),\hat{\theta}_2(n),\ldots,\hat{\theta}_p(n),\bar{\theta}_{p+1},\ldots,\bar{\theta}_{p+k})$$

Also define for $\ell = 1,2,\ldots,k$,

$$(2.76)\qquad T_n'(\ell) = -\{A_n(p+\ell;\hat{\theta}(n)) - \sum_{\alpha,\beta=1}^{p} F(\alpha,\beta;\hat{\theta}(n))B(\beta,p+\ell;\hat{\theta}(n))\cdot A_n(\alpha;\hat{\theta}(n))\}$$

and,

$$(2.77)\qquad T_n(\ell) = -A_n(p+\ell;\hat{\theta}(n))$$

Also define for $\ell \neq j$ $(\ell,j = 1,2,\ldots,k)$

$$(2.78)\qquad \hat{W}_n(\ell|j) = T_n(\ell) - T_n(j) = T_n'(\ell) - T_n'(j) \quad .$$

Then, an asymptotically optimal decision rule for the problem of selecting the best population is given by the following theorem:

Theorem 2.4 An asymptotically optimal decision rule for selecting the best population is given as

Select H_ℓ (Π_ℓ as the best population) if

$$T_n(\ell) \geq T_n(j) \qquad j \neq \ell \qquad \ell, j = 1,2,\ldots,k \quad . \tag{2.79}$$

Proof: It suffices to show that the decision rule defined above has the same asymptotic probability of correct decision as the Bayes decision for the artificial problem, which was also shown to be asymptotically minimax. It will be sufficient because then the decision rule has, asymptotically, the same "risk" as a Bayes and minimax decision rule and hence is an asymptotically minimax (optimal) decision rule.

In order to do that, we obtain the asymptotic distribution of $\hat{W}_n(\ell|j)$. From (2.78), we get

(2.80)

$$\hat{W}_n(\ell|j) =$$

$$A_n(p+j;\theta+\overline{\hat{\theta}(n)-\theta}) - \sum_{\alpha,\beta=1}^{p} A_n(\alpha;\theta+\overline{\hat{\theta}(n)-\theta})F(\alpha,\beta;\hat{\theta}(n))B(\beta,p+j;\hat{\theta}(n))$$

$$- \{A_n(p+\ell;\theta+\overline{\hat{\theta}(n)-\theta}) - \sum_{\alpha,\beta=1}^{p} A_n(\alpha;\theta+\overline{\hat{\theta}(n)-\theta})F(\alpha,\beta;\hat{\theta}(n))B(\beta,p+\ell;\hat{\theta}(n)\}$$

Expanding around θ, we obtain for $\alpha = 1,2,\ldots,p+k$,

$$(2.81)\quad A_n(\alpha;\hat{\theta}(n)) = A_n(\alpha;\theta) + \sum_{\gamma=1}^{p+k} \sqrt{n}(\hat{\theta}_\gamma(n)-\theta_\gamma)(-B_n(\alpha,\gamma;\bar{\theta})) + R_n(\alpha)$$

where $R_n(\alpha)$ converges stochastically to zero as $n \to \infty$. Substituting from (2.81) in (2.80), we get

(2.82)

$$\tilde{W}_n(\ell|j) =$$

$$\begin{bmatrix} A_n(p+j;\theta) + \sum_{\gamma=1}^{p+k} \sqrt{n}(\theta_\gamma-\hat{\theta}_\gamma(n))B(p+j,\gamma;\theta) - \sum_{\alpha,\beta=1}^{p} F(\alpha,\beta;\bar{\theta})B(\beta,p+j;\bar{\theta}) \cdot \\ \{A_n(\alpha;\theta) + \sum_{\gamma=1}^{p+k} \sqrt{n}(\theta_\gamma-\hat{\theta}_\gamma(n))B(\alpha,\gamma;\theta)\} \end{bmatrix}$$

$$- \begin{bmatrix} A_n(p+\ell;\theta) + \sum_{\gamma=1}^{p+k} \sqrt{n}(\theta_\gamma-\hat{\theta}_\gamma(n))B(p+\ell,\gamma;\theta) - \sum_{\alpha,\beta=1}^{p} F(\alpha,\beta;\theta)B(\beta,p+\ell;\theta) \cdot \\ \{A_n(\alpha;\theta) + \sum_{\gamma=1}^{p+k} \sqrt{n}(\theta_\gamma-\hat{\theta}_\gamma(n))B(\alpha,\gamma;\theta)\} \end{bmatrix}$$

$$+ Q_n(\ell|j)$$

where $Q_n(\ell|j)$ converges stochastically to zero as n increases. Rearranging terms, we get

(2.83)

$$\hat{W}_n(\ell|j) =$$

$$\left\{ \left(A_n(p+j;\theta) - \sum_{\alpha,\beta=1}^{p} F(\alpha,\beta;\theta)B(\beta,p+j;\theta)A_n(\alpha;\theta)\right) \right.$$
$$\left. -\left(A_n(p+\ell;\theta) - \sum_{\alpha,\beta=1}^{p} F(\alpha,\beta;\theta)B(\beta,p+\ell;\theta)A_n(\alpha;\theta)\right) \right\}$$

$$+ \sum_{\gamma=p+1}^{p+k} \sqrt{n}(\theta_\gamma - \hat{\theta}_\gamma(n)) \left[\begin{array}{l} \left(B_n(p+j,\gamma;\theta) - \sum_{\alpha,\beta=1}^{p} F(\alpha,\beta;\theta)B(\beta,p+j;\theta)B_n(\alpha,\gamma;\theta)\right) \\ -\left(B_n(p+\ell,\gamma;\theta) - \sum_{\alpha,\beta=1}^{p} F(\alpha,\beta;\theta)B(\beta,p+\ell;\theta)B_n(\alpha,\gamma;\theta)\right) \end{array} \right]$$

$$+ \sum_{\gamma=1}^{p} \sqrt{n}(\theta_\gamma - \hat{\theta}_\gamma(n)) \left[\begin{array}{l} \left(B_n(p+j,\gamma;\theta) - \sum_{\alpha,\beta=1}^{p} F(\alpha,\beta;\theta)B(\beta,p+j;\theta)B_n(\alpha,\gamma;\theta)\right) \\ -\left(B_n(p+\ell,\gamma;\theta) - \sum_{\alpha,\beta=1}^{p} F(\alpha,\beta;\theta)B(\beta,p+\ell;\theta)B_n(\alpha,\gamma;\theta)\right) \end{array} \right]$$

$$+ Q_n(\ell|j) .$$

Since in (2.83), $\hat{W}_n(\ell|j)$ is of the same form as $W_n(\ell|j)$ in (2.60) and because of the properties of estimators $(\hat{\theta}_1(n),\ldots,\hat{\theta}_p(n))$ (i.e. (2.74)) it follows that $\hat{W}_n(\ell|j)$ has the same asymptotic normal distribution with mean and covariance matrix as given by Theorem 2.3.

Since the decision rule given by (2.79) is equivalent to selecting H_ℓ if $\hat{W}_n(\ell|j) \geq 0$, as in (2.73), it follows immediately that the decision rule given by (2.79) has the same asymptotic Bayes risk and is hence an asymptotically minimax procedure for the ranking problem.

Q.E.D.

In the development of an asymptotically optimal procedure for the problem of selecting the best population, we considered a restricted parameter configuration (as given by (2.17)). The general parameter configuration may be represented as follows:

For $\ell = 1,2,\ldots,k$,

$$(2.84) \qquad H_\ell: \begin{cases} \theta_{p+\ell} = \overline{\theta}_{p+\ell} - c/\sqrt{n} \\ \theta_{p+j} = \overline{\theta}_{p+j} + c_j/\sqrt{n} \qquad j = 1,2,\ldots,k \neq \ell \end{cases}$$

where, $c_j \geq c > 0$.

A word of explanation is in order here. In (2.84), hypothesis H_ℓ implies that Π_ℓ is the best population; and the distance between the best population and the other populations are $\frac{c+c_j}{\sqrt{n}}$, $(j = 1,2,\ldots,k, j \neq \ell)$. For the case where c_j are all equal, we get the restricted configuration given by (2.17) and the case where one or all of c_j are different, we get a general parameter configuration.

Now, we must show that the procedure given by Theorem 2.4 is indeed asymptotically minimax over the parameter space. This follows very easily since for a general parameter configuration given by (2.80), the probability of correct selection increase, whenever for any j, $c_j > c$. This implies hence that the configuration given by (2.17) is the least favorable configuration and hence the procedure given by Theorem 2.4 is indeed asymptotically minimax.

2.5 The Rate of Convergence of Decision Variables to an Asymptotic Normal Distribution

In our analysis, we have shown that our decision rule (procedure) is asymptotically optimal. In order to completely specify the test, asymptotic normality of the decision variables is used. It would be, both from a theoretical and practical viewpoint, very useful to determine the rate of convergence to the asymptotic normal distribution. From a practical viewpoint, to an experimenter, who for large sample size n, would act as if the asymptotic distribution is the actual distribution (which it is, to a close approximation) it would be helpful to know how fast the asymptotic results take over. The results on the rate of convergence would supply that information.

In this section, we present the main results on the rate of convergence in Central Limit Theorems for a set of independent random variables (one dimensional or multidimensional). The results on rate of convergence for one dimensional random variables are useful for the problem of ranking two populations, which can be easily shown to be a one parameter problem. The results for the one dimensional case would equally well apply to the paper by Weiss and Wolfowitz [55].

Rate of Convergence in One Dimensional Central Limit Theorem

Let $X_1, X_2, \ldots X_n$ be independent and identically distributed random variables with $EX = 0$, $\text{Var } X = 1$ and $E|X|^3 < \infty$. Denote $\sum_{i=1}^{n} X_i$ by S_n and $\int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt$ by $\Phi(x)$. Let

$$A_n = \sup_x \left| P\left(\frac{S_n}{\sqrt{n}} \leq x\right) - \Phi(x) \right| \tag{2.85}$$

We state, without proof, the following theorems (on the rate of convergence in a central limit theorem) of Berry [10], Esseen [18] and Katz [28] in

Theorem 2.5 (Berry-Esseen)

$$A_n \leq \frac{C\, E|X|^3}{\sqrt{n}} \tag{2.86}$$

where C is an absolute constant.

Theorem 2.6 (Katz)

$$A_n \leq C_1 \frac{E\{X^2 g(x)\}}{g(\sqrt{n})} \tag{2.87}$$

where C_1 is an absolute constant and $g(x)$ is defined on the real line and satisfies

i) $g(x) \geq 0$, even, non-decreasing in $[0,\infty)$ and

ii) $x/g(x)$ is defined for all x and non-decreasing on $[0,\infty)$.

Theorem 2.6 is useful if one cannot make the assumption about $E|X|^3 < \infty$, but would be much harder to apply, since the function $g(x)$ has to be specified. In case one can assume that $E|X|^3 < \infty$, Theorem 2.5 would be very convenient to use.

We note that by the Central Limit Theorem, $A_n \to 0$ and from (2.86), A_n converges at a rate of $n^{-1/2}$. The constant C in Theorem 2.5 is given by the following result due to Zolotarev [56].

Theorem 2.7 (Zolotarev) In Theorem 2.5,

$$C \leq 1.301 \tag{2.88}$$

In our analysis, the random variables $\{X_i,\ i = 1,2,\ldots,n\}$ would be replaced by a suitable function of the original observations, as defined by the decision variables $T_n(\ell)$ (see (2.78)). If besides the regularity conditions, we also assume the existence of the 3rd absolute moments, we can use Theorem 2.5 and Theorem 2.7 directly. The rate of convergence is of the order of $n^{-1/2}$ and if the 3rd absolute moment of the normalized decision variables is small (it depends on the exact parametric form of the probability density function of the populations being ranked) the large sample results will be effective quite rapidly.

Rate of Convergence in Multidimensional Central Limit Theorem

For the general ranking problem for k populations $(k \geq 3)$, we need multidimensional analogues of the Berry-Esseen Theorem.

Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables $(X_i \equiv (X_{1i}, X_{2i}, \ldots, X_{ki}))$ each with distribution function F. Let

$$\begin{aligned} EX_{ti} &= \mu_t \\ E(X_{ti} - \mu_t)^2 &= \sigma_t^2 \\ E|X_{ti} - \mu_t|^3 &= \beta_t \end{aligned} \qquad \text{for } t = 1,2,\ldots,k \tag{2.89}$$

$$\text{and} \quad \alpha_t = \frac{\beta_t}{\sigma_t^{3/2}}$$

Also let Λ be the determinant of the covariance matrix and Λ_{tt} the corresponding principal minors of Λ $(t = 1,2,\ldots,k)$. Let

$S_n = \sum_{i=1}^{n} Y_i$ and let

$$F_n(x) = P\{\frac{S_n}{\sqrt{n}} \leq x\} \tag{2.90}$$

where S_n and x are k dimensional vectors. Let $G(x)$ denote the cdf of normal random variables with same first and second moments as that of $F_n(\cdot)$.

Using the above notation, we have the following theorem by Sazanov [42] (stated without proof) on the rate of convergence.

Theorem 2.9 (Sazanov)

$$\sup_{x \in R^k} |F_n(x) - G(x)| \leq C_2(k) \left(\sum_{t=1}^{k} \frac{\Lambda_{tt}}{\Lambda} \alpha_t\right) n^{-1/2} \tag{2.91}$$

As before, we note that the rate of convergence is of the order $n^{-1/2}$, and if the constant terms in (2.91) are small enough, asymptotic results would be effective quite rapidly. It may be noted, that for our analysis, since the decision variables $\hat{W}_n(i|j)$ (whose rate of convergence is being determined) have the same correlation coefficient and the same variances, in (2.91) we would get

$$\begin{aligned} \Lambda_{tt} &= \lambda \\ \alpha_t &= \alpha \end{aligned} \qquad t = 1,2,\ldots,k \tag{2.92}$$

and hence a certain simplification would be achieved in determining the right hand side of (2.91).

2.6 Examples

In this chapter, we have developed asymptotically optimal procedures for selecting the best population, when the density function of the observations satisfies certain mild regularity conditions (as given in Section 2.1). It is interesting to note that for a large class of density functions (satisfying the above mentioned regularity conditions), an asymptotically optimal procedure for selecting the best population (as given by Theorem 2.4) takes such a simple form. The populations are ranked according to the value of the statistic T_ℓ (as given by (2.77), the population associated with the largest value of T_ℓ being selected as the best population.

We now give some examples to illustrate the applicability of the results to specific distributions. The first set of examples indicates the asymptotic optimality of selection procedures already proposed in the literature. In the second example we give an asymptotically optimal procedure for selecting the normal population with the largest mean, the populations having common unknown variance. We also indicate how the asymptotically optimal procedure would be used in practice for large sample sizes. In the third example, we develop an asymptotically optimal procedure for the problem of ranking Weibull distributions according to the value of the scale parameter, when the populations have known location parameters and common (known or unknown) shape parameter.

1. Asymptotic Optimality of Certain Selection Procedures already Proposed in the Literature

As a consequence of the results in this chapter, the selection procedures already proposed in the literature for the following problems are asymptotically optimal (the references in brackets indicate the papers in which the procedures were proposed):

(i) Ranking variances of normal populations with known or unknown means (Bechhofer and Sobel [8]).

(ii) Selecting the best of several binomial populations (Sobel and Huyett [44]).

(iii) Selecting the multinomial event with the highest probability (Bechhofer, Elmaghraby and Morse [6]).

(iv) Selecting the bivariate normal population with the largest correlation coefficient (Ramberg [39]; also given as an example in Eaton [17]).

(v) Selecting the component with the largest mean in ranking from a single multivariate population with common known variance and covariance of the components. (Given as an example in Eaton [17] and Milton [34]).

2. Ranking Means of Normal Populations with Common Unknown Variance

An asymptotically optimal procedure for selecting the normal population with the largest mean, populations having a common unknown variance, is to select the population associated with the largest sample mean.

The procedure is the same, as one may expect, as one proposed by Bechhofer [4] for ranking means of normal populations with common known

variance. However, there is a difference in the probability statements one can make, depending on whether the common variance is known or unknown. In case the common variance is known, the experimenter specifies, prior to experimentation, two constants (δ^*, P^*), with $\delta^* > 0$, $\frac{1}{k} < P^* < 1$. The sample size (n) needed to guarantee the probability requirement is given as a solution to

$$\frac{\delta^*\sqrt{n}}{\sigma} = \lambda(P^*) \tag{2.93}$$

where, σ^2 is the common known variance and $\lambda(P^*)$ is, as given in [4], the solution to

$$\int_{-\infty}^{\infty} F^{k-1}(x + \lambda(P^*))f(x)dx = P^* \tag{2.94}$$

with

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

and

$$F(x) = \int_{-\infty}^{x} f(y)dy$$

$\lambda(P^*)$ is tabled in Bechhofer [4] (also in Milton [34] and Gupta [21]) and thus one determines the required sample size $n(\delta^*,P^*)$ for any prespecified constants (δ^*,P^*).

In case the common variance (σ^2) is unknown, for a sample of size n, the experimenter computes a consistent estimate $\hat{\sigma}^2(n)$, say the sample variance, of the unknown variance σ^2 and acts as if $\hat{\sigma}^2(n)$ were the actual variance σ^2. For any sample size n and the computed $\hat{\sigma}^2(n)$, one determines the pair $\lambda(\delta^*,P^*)$ satisfying (2.93). Then, for a specified δ^* one finds the probability (P^*) that is guaranteed. (or for a given P^* one can find the δ^* for which P^* is guaranteed.)

3. Ranking Scale Parameters of Weibull Distribution with Known Location Parameter and Common (known or unknown) Shape Parameter

For a random variable X having a Weibull distribution, the cdf $G(x;\theta)$ is given as

$$G(x;\theta) = \begin{cases} 0 & x < \theta_1 \\ 1 - \exp\left(-\left(\frac{x-\theta_1}{\theta_2}\right)^{1/\theta_3}\right) & x \geq \theta_1 \end{cases} \tag{2.95}$$

where θ_1 is the location parameter, θ_2 is the scale parameter and θ_3 is the shape parameter, $\theta = (\theta_1,\theta_2,\theta_3)$ and the parameter space $\Omega = \{\theta | -\infty < \theta_1 < \infty,\ \theta_2 > 0,\ \theta_3 > 0\}$.

In case the location parameter θ_1 is known, we may take θ_1 to be zero with no loss in generality and get

$$G(x;\theta) = \begin{cases} 0 & x < 0 \\ 1 - \exp\left(-\left(\frac{x}{\theta_2}\right)^{1/\theta_3}\right) & x \geq 0 \end{cases} \tag{2.96}$$

where $\theta = (\theta_2, \theta_3)$ and the parameter space $\Omega = \{\theta | \theta_2 > 0, \quad \theta_3 > 0\}$.

Defining the random variable Y as $\log X$, denoting θ_2 by e^B and denoting (B, θ_3) by θ we find that the cdf $F(y;\theta)$ of the random variable Y is given as

$$F(y;\theta) = 1 - \exp\left\{ - \exp\left(\frac{Y-B}{\theta_3}\right)\right\} \tag{2.97}$$

For the random variable Y having cdf as given by (2.97), B is the location parameter and θ_3 is the scale parameter. Thus the problem of ranking scale parameters of Weibull distributions reduces to the problem of ranking location parameters of a distribution as given by (2.97).

The density function $f(y;\theta)$ of observations, with cdf as given by (2.97), is given as

$$f(y;\theta) = \frac{1}{\theta_3} \exp\left(\frac{Y-B}{\theta_3}\right) \exp\left(-\exp\left(\frac{Y-B}{\theta_3}\right)\right) \tag{2.98}$$

For the ranking problem at hand, with density function of observations given by (2.98), the regularity conditions of Section 2.1 are satisfied. Thus the results obtained earlier in this chapter hold and an asymptotically optimal procedure is as given by Theorem 2.4.

Let X_{ti} $(i = 1,2,\ldots,n)$ denote independent observations from population Π_t $(t = 1,2,\ldots,k)$ each with cdf given by (2.96). Let

$$Y_{ti} = \log X_{ti} \qquad \begin{matrix} i = 1,2,\ldots,n \\ t = 1,2,\ldots,k \end{matrix} \tag{2.99}$$

and, for $\ell = 1,2,\ldots,k$

$$T_{\ell} = \sum_{i=1}^{n} (X_{ti})^{1/\theta_3} \tag{2.100}$$

Let the ranked value of T_{ℓ} be denoted by

$$T_{[1]} \leq T_{[2]} \leq \cdots\cdots \leq T_{[k]} \tag{2.101}$$

For selecting the Weibull population with largest scale parameter, when the location parameters are known and the populations have a common known shape parameter (θ_3), as asymptotically optimal procedure (as given by Theorem 2.4) is to select the population associated with $T_{[k]}$.

Let $Z_1 \leq Z_2 \leq \cdots\cdots \leq Z_{nk}$ denote the ordered values of $\{Y_{ti},\ i = 1,2,\ldots,n,\ t = 1,2,\ldots,k\}$. Then, as given by Weiss [50],

$$\hat{\theta}_3 = -\frac{nk+1}{nk-1} \sum_{j=1}^{nk-1} \{(1-j/nk)\log(1-j/nk)\}(Z_{j+1} - Z_j) \tag{2.102}$$

is a consistent estimate of θ_3.

For $\ell = 1,2,\ldots,k$, let

$$\hat{T}_{\ell} = \sum_{i=1}^{n} (X_{ti})^{1/\hat{\theta}_3} \tag{2.103}$$

where $\hat{\theta}_3$ is as given by (2.102). Let the ordered values of $\hat{T}_\ell$ be denoted by

$$\hat{T}_{[1]} \leq \hat{T}_{[2]} \leq \cdots \leq \hat{T}_{[k]} \tag{2.104}$$

For selecting the Weibull distribution with largest scale parameter, when the location parameters are known and the populations have a common unknown shape parameter (θ_3), an asymptotically optimal procedure (as given by Theorem 2.4) is to select the population associated with $\hat{T}_{[k]}$.

CHAPTER 3

ASYMPTOTICALLY OPTIMAL PROCEDURES FOR CERTAIN ADDITIONAL RANKING GOALS

In the last chapter, we have developed asymptotically optimal procedures for selecting the best population, for situations in which the joint density function of the observations satisfies certain mild regularity conditions. In this chapter, we extend the basic results to develop asymptotically optimal procedures for certain other ranking goals.

The following two general ranking goals have been considered in the literature:

(i) Selecting a subset of size s to contain at least d of the t best populations, with $\max(1,s+t+1-k) \leq d \leq \min(s,t)$ (which implies $\max(s,t) \leq k-1$).

d,s,t are integers specified prior to experimentation.

(ii) Select the k_s "best" populations, the k_{s-1} "second best" populations, the k_{s-2} "third best" populations, etc., and finally the k_1 "worst" populations.

$k_1,k_2,\ldots,k_s$ $(s \leq k)$ are integers such that $\sum_{i=1}^{s} k_i = k$ and $(k_1,k_2,\ldots,k_s,s)$ are specified prior to experimentation.

Bechhofer [4] in his paper alluded to the ranking goal (i), but it was formulated and formally considered by Mahamunulu [33] and Desu and Sobel [14]. It has been called Goal I by Mahamunulu [33] and we too shall use that designation. Two particular cases of Goal I are of special interest: (a) Selection of a subset of size s $(\geq t)$ which contains the t best populations, and (b) Selection of a subset of size s $(\leq t)$ which contains any s of the t best populations. In Section 3.1, we consider a special case of (a), with $t = 1$. We develop

an asymptotically optimal procedure for this special case of Goal I. This case is treated in detail, since it brings out the main steps needed in extending the basic results in Chapter 2 to develop asymptotically optimal procedures for more general ranking goals. In Section 3.2, we treat a special case of (b) with $s = 1$. We develop an asymptotically optimal procedure for this special case of Goal I. In Section 3.3, we give, without proof, an asymptotically optimal procedure for Goal I. Bechhofer [4] formulated goal (ii) and considered two cases in detail, namely, $s = 2$, $k_1 = k-t$, $k_2 = t$ and $s = t+1$, $k_1 = k-t$, $k_2 = k_3 = \ldots = k_{t+1} = 1$. For the second case, we are interested in selecting the t best populations with regard to order. Bechhofer [4] considered this special case and provided tables (to be used to determine the sample size needed for the proposed single-stage procedure) for the case $k = t = 3$. This special case of goal (ii) has been formulated as Goal II in Bechhofer, Kiefer and Sobel [7] and we refer to it as Goal II too. In Section 3.4, we obtain an asymptotically optimal procedure for Goal II. In Section 3.5, we point out certain other possible goals in ranking and selection problems. The method developed in this thesis could be adapted to these cases.

3.1 Selecting a Fixed-Size Subset to Contain the Best Population

In order to develop an asymptotically optimal procedure for selecting a fixed-size subset to contain the best population, we formulate the problem in decision theoretic structure. For $i = 1,2,\ldots,k$, let

$$
\begin{aligned}
&\theta_{p+j} = \bar{\theta}_{p+j} + c/\sqrt{n} \qquad j = 1,2,\ldots,k \neq i \\
(3.1) \qquad &H_i: \\
&\theta_{p+i} = \bar{\theta}_{p+i} - c/\sqrt{n}
\end{aligned}
$$

where $c > 0$ and $\bar{\theta}_{p+j} = \bar{\theta}_{p+i}$ all i,j $(i,j = 1,2,\ldots,k)$ are known values. $\bar{\theta}_{p+j}$ may be taken to be zero with no loss in generality.

In view of (2.1), (3.1) represents a restricted parameter configuration, in which if H_i is the true state of nature, then Π_i is the best population.

We now describe the basic method used in developing an asymptotically optimal ranking procedure for this problem. In order to develop such a procedure for the ranking problem, we develop an asymptotically optimal procedure for an associated identification problem, with $\bar{\theta}_0$, the common location of the ranking parameter, as the least favorable location of the ranking parameter. We first solve the problem for a restricted parameter configuration, given by (3.1), and then show that the procedure developed is minimax over all parameter configurations.

Let $D(I)$ denote the decision to select $\Pi_{i_1},\Pi_{i_2},\ldots,\Pi_{i_s}$ as the subset (of size s) containing the best populations, where $I = \{(i_1,i_2,..,i_s); i_t \neq i_{t'},\ t,t' = 1,2,\ldots,s,\ i_t,i_{t'} = 1,2,\ldots,k\}$. I, the index set, is a s-tuple whose components inidcate the s populations included in the selected subset.

The loss function for the multiple decision problem is given as

$$(3.2) \qquad W(\theta;D(I)) = \begin{cases} 0 & \text{if } H_{i_1} \text{ or } H_{i_2} \text{ or } \ldots H_{i_s} \text{ is the true hypothesis} \\ 1 & \text{otherwise} \end{cases}$$

where $I = (i_1,i_2,\ldots,i_s)$.

There are $\binom{k}{s}$ possible decisions, and we thus have a $\binom{k}{s}$-decision problem with simple loss function given by (3.2).

Preliminary Sequence of Artificial Problems

We now develop an asymptotically optimal procedure for a sequence of artificial $\binom{k}{s}$-decision problems (one for each n).

Suppose it is known that $\bar{\theta}_i - \frac{M_n(\bar{\theta})}{\sqrt{n}} \leq \theta_i \leq \bar{\theta}_i + \frac{M_n(\bar{\theta})}{\sqrt{n}}$, $i = 1,2,\ldots,p$ and $(\theta_{p+1},\theta_{p+2},\ldots,\theta_{p+k})$ satisfy one of the hypotheses given by (3.1). $\bar{\theta}_1,\bar{\theta}_2,\ldots,\bar{\theta}_p$ are known constants and $0 < c \leq L$. We wish to select a subset of size s, and the loss function is given by (3.2).

For the above problem, we construct a Bayes decision rule relative to the following prior distribution:

For $j = 1,2,\ldots,k$, a total mass of b_j is spread uniformly over the set

$$(3.3) \qquad \bar{\theta}_i - \frac{M_n(\bar{\theta})}{\sqrt{n}} \leq \theta_i \leq \bar{\theta}_i + \frac{M_n(\bar{\theta})}{\sqrt{n}} \qquad (i = 1,2,\ldots,p)$$

and H_j is true where; for $j = 1,2,\ldots,k$, we have $b_j \geq 0$, and

$$\sum_{j=1}^{k} b_j = 1 \quad .$$

The prior distribution $(b_j,\ j = 1,2,\ldots,k)$ is arbitrary, but fixed. Later we select the prior in such a way as to obtain a minimax decision rule for the problem at hand.

Before we develop a Bayes decision rule, we define some notation used later on. For $j = 1,2,\ldots,k$, let

$$(3.4) \qquad b_j(n) = b_j \int_{\bar{\theta}_1 - \frac{M_n(\bar{\theta})}{\sqrt{n}}}^{\bar{\theta}_1 + \frac{M_n(\bar{\theta})}{\sqrt{n}}} \cdots\cdots \int_{\bar{\theta}_p - \frac{M_n(\bar{\theta})}{\sqrt{n}}}^{\bar{\theta}_p + \frac{M_n(\bar{\theta})}{\sqrt{n}}} \prod_{i=1}^{n} f(X_i;\theta,H_j)\,d\theta_1,\ldots,d\theta_p \quad .$$

Also let,

$$(3.5) \qquad \bar{b}_j(n) = b_j(n) \Big/ \sum_{t=1}^{k} b_t(n)$$

Then, we may view $\bar{b}_j(n)$ as the posterior probability for the hypothesis H_j, if we consider b_j as the prior probability.

In terms of notation of Chapter 2, for $\ell, j = 1,2,\ldots,k$,

$$(3.6) \qquad J_n(\ell|j) = \frac{b_j}{b_\ell} \frac{b_\ell(n)}{b_j(n)}$$

Let the ranked values of $b_j(n)$ be denoted by

$$b_{[1]}(n) \leq b_{[2]}(n) \leq \cdots\cdots \leq b_{[k]}(n) \tag{3.7}$$

In order to develop a Bayes decision rule for this problem, we compute $k(D(I);x)$ for each I and (in view of the results in Section 2.2) use that to construct a Bayes decision rule.

For $I = (i_1, i_2, \ldots, i_s)$,

$$\begin{aligned} k(D(I);x) &= \sum_{j=1}^{k} b_j(n) - b_{i_1}(n) - b_{i_2}(n) - b_{i_3}(n) - \ldots - b_{i_3}(n) \\ &= \sum_{j=1}^{k} b_j(n) - \sum_{i_t \in I} b_{i_t}(n) \end{aligned} \tag{3.8}$$

From (3.8), it follows directly that a Bayes decision rule is to select the s populations associated with the s largest values of $b_j(n)$. In view of (3.7), a Bayes decision rule is to select populations associated with $(b_{[k-s+1]}(n), \ldots, b_{[k]}(n))$.

Using (3.6), the Bayes decision rule can be seen to reduce to the following:

Select $\Pi_{i_1}, \Pi_{i_2}, \ldots, \Pi_{i_s}$ as the subset containing the <u>best</u> population if

$$\begin{aligned} J_n(i_1|j) &\geq \frac{b_j}{b_{i_1}} \\ J_n(i_2|j) &\geq \frac{b_j}{b_{i_2}} \qquad \text{for } j \neq (i_1, i_2, \ldots, i_s) \\ &\vdots \qquad\qquad\qquad j = 1, 2, \ldots, k \\ J_n(i_s|j) &\geq \frac{b_j}{b_{i_s}} \end{aligned} \tag{3.9}$$

The decision rule, as defined by (3.9), divides the sample space into $\binom{k}{s}$ mutually exclusive and collectively exhaustive regions, on each of which a distinct set of s populations will be selected as the best subset. Note, as a check, that when $s = 1$, (3.9) reduces to the decision rule given in Chapter 2 for the problem of selecting the best population.

We now obtain asymptotic properties of the Bayes decision rule given by (3.9). Since the assumptions and regularity conditions of Section 2.1 hold, the asymptotic properties of $J_n(\ell|j)$, $\ell,j = 1,2,\ldots,k$, $\ell \neq j$, obtained in Section 2.4 hold. Thus Lemmas 2.1-2.4 and Theorem 2.3 could be used to reduce the Bayes decision rule to the following:

Select $\Pi_{i_1},\Pi_{i_2},\ldots,\Pi_{i_s}$ as the subset containing the best population if

$$
\begin{aligned}
W_n(i_1|j) &\geq \tfrac{1}{2c}\left(\log \frac{b_j}{b_{i_1}} + Z_n'(i_1|j)\right) \\
W_n(i_2|j) &\geq \tfrac{1}{2c}\left(\log \frac{b_j}{b_{i_2}} + Z_n'(i_2|j)\right) \\
&\vdots \\
W_n(i_s|j) &\geq \tfrac{1}{2c}\left(\log \frac{b_j}{b_{i_s}} + Z_n'(i_s|j)\right)
\end{aligned}
\qquad (3.10)
$$

$$
j \neq (i_1,i_2,\ldots,i_s)
$$

$$
j = 1,2,\ldots,k
$$

where, as before, $Z_n'(i_t|j)$, $(t = 1,2,\ldots,s)$ converges stochastically to zero as n increases, if the true parameter point $\omega(n) \in R^n_{L_n(\bar{\theta})}(\bar{\theta})$.

Also $W_n(i_1|j)$ are as defined by (2.54) and their asymptotic distribution is as given by Theorem 2.3. Thus, the decision variables have, asymptotically, a multi-variate normal distribution with mean and variance given by Theorem 2.3.

Using Theorem 2.1 and the natural symmetry in the multi-decision problem, it follows directly that the Bayes decision rule given by (3.10) is an asymptotically minimax decision rule if we set $b_j = \frac{1}{k}$ for $j = 1,2,\ldots,k$. Thus, an asymptotically minimax decision rule is given as:

Select $\Pi_{i_1},\Pi_{i_2},\ldots,\Pi_{i_s}$ as the subset containing the best population if

$$
\begin{aligned}
&W_n(i_1|j) \geq 0 \qquad j \neq (i_1,i_2,\ldots,i_s)\\
&W_n(i_2|j) \geq 0 \qquad j = 1,2,\ldots,k\\
&\quad\vdots\\
&W_n(i_s|j) \geq 0
\end{aligned}
\tag{3.11}
$$

Recalling that as $\frac{L_n(\bar{\theta})}{M_n(\bar{\theta})} \to 1$, the decision rule given by (3.11) is an asymptotically minimax decision rule for the sequence of artificial $\binom{k}{s}$-decision problems.

Optimal Procedure for the "Real" Problem

The above sequence of problems was artificial because we assumed $\bar{\theta}_1,\bar{\theta}_2,\ldots,\bar{\theta}_p$ to be known. We now proceed to the real problem, where nothing is known about the values of the nuisance parameters $(\theta_1,\theta_2,\ldots,\theta_p)$.

As before, let $\hat{\theta}_1(n),\hat{\theta}_2(n),\ldots,\hat{\theta}_p(n)$ denote estimators of $\theta_1,\theta_2,\ldots,\theta_p$ respectively, based on the observations $(X_1,\ldots,X_n)$. We assume that these estimators satisfy the consistency condition, as given by (2.74); one may use Maximum Probability Estimators (MPE) of $(\theta_1,\ldots,\theta_p)$ as the estimators $(\hat{\theta}_1(n),\ldots,\hat{\theta}_p(n))$.

For notational convenience, let

$$\hat{\theta}(n) = (\bar{\theta}_0,\hat{\theta}_1(n),\hat{\theta}_2(n),\ldots,\hat{\theta}_p(n),\bar{\theta}_{p+1},\ldots,\bar{\theta}_{p+k}) \tag{3.12}$$

For $\ell = 1,2,\ldots,k$, let

$$T_\ell(n) = -A_n(p+\ell;\hat{\theta}(n)) \tag{3.13}$$

Denote the ranked values of $T_\ell(n)$ by

$$T_{[1]}(n) \leq T_{[2]}(n),\ldots, \leq T_{[k]}(n) \tag{3.14}$$

Then, an asymptotically optimal decision rule for the problem of selecting a subset of size s to contain the best population is given by the following theorem:

Theorem 3.1. An asymptotically optimal decision rule for selecting a subset of size s to contain the best population is to select the s populations associated with $T_{[k]}(n),T_{[k-1]}(n),\ldots,T_{[k-s+1]}(n)$.

Proof: It suffices to show that the decision rule given by Theorem 3.1 has the same asymptotic properties as the minimax decision rule (given by (3.11)) for the sequence of artificial problems. That follows directly from the proof of Theorem 2.4, completing the proof of the present theorem.

Q.E.D.

In the development of an asymptotically optimal procedure for the ranking problem, we considered a restricted parameter configuration given by (3.1). The general parameter configuration may be represented as (2.84). As in Chapter 2, it follows very easily that the parameter configuration given by (3.1) is the least favorable configuration and hence the procedure given by Theorem 3.1 is indeed asymptotically minimax over all parameter configurations.

In order to compute the probability of correct selection for the selection procedure given by Theorem 3.1, one may use tables given in Desu and Sobel [14] and Mahamunulu [32] for some special cases.

3.2 Selecting One of the t Best Populations

In order to develop an asymptotically optimal procedure for this new goal, we formulate the problem in decision theoretic structure.

Let $I = \{(i_1,i_2,\ldots,i_k);\ i_j = 0 \text{ or } 1, \text{ exactly } t\ i_j \text{ equal to } 1,\ j = 1,2,\ldots,k\}$. There are $\binom{k}{t}$ possible I and for any particular I, let

$$\text{(3.14)} \qquad \Pi(I):\ \theta_{p+j} = \begin{cases} \bar{\theta}_{p+j} - \dfrac{c}{\sqrt{n}} & \text{if } i_j = 1 \\ \bar{\theta}_{p+j} + \dfrac{c}{\sqrt{n}} & \text{if } i_j = 0 \end{cases}$$

where $c > 0$ and $\bar{\theta}_{p+j} = \bar{\theta}_{p+i}$ all i,j $(i,j = 1,2,\ldots,k)$ are known values. Here $\bar{\theta}_{p+j}$ may be taken to be zero with no loss in generality.

In view of (2.1), (3.14) represents a restricted parameter configuration, in which if $H(I)$ is the true state of nature, then the t populations for which $i_j = 1$ in I are the t best populations. For the problem at hand, this is the least favorable configuration.

In order to develop an asymptotically optimal procedure for the ranking problem, we develop an asymptotically optimal procedure for an associated identification problem, with $\bar{\theta}_0$, the common location of the ranking parameter, as the least favorable location of the ranking parameter. We solve the problem for a restricted parameter configuration given by (3.14). Since (3.14) represents the least favorable location, the procedure developed is asymptotically minimax over all parameter configuration.

For $j = 1,2,\ldots,k$, let D_j denote the decision to select Π_j as one of the t best populations. There are k possible decisions and $\binom{k}{t}$ states of nature; the simple loss function for the problem is given as

$$W(\theta,D_j) = \begin{cases} 0 & \text{if } i_j = 1 \text{ in } I \\ 1 & \text{otherwise} \end{cases} \tag{3.15}$$

Preliminary Sequence of Artificial Problems

We now develop an asymptotically optimal procedure for a sequence of artificial k-decision problems (one for each n).

Suppose it is known that $\bar{\theta}_i - \frac{M_n(\bar{\theta})}{\sqrt{n}} \leq \theta_i \leq \bar{\theta}_i + \frac{M_n(\bar{\theta})}{\sqrt{n}}$, $i = 1,2,\ldots,p$, and $(\theta_{p+1},\ldots,\theta_{p+k})$ satisfy one of the $\binom{k}{t}$ hypothesis given by (3.14). Here $(\bar{\theta}_1,\bar{\theta}_2,\ldots,\bar{\theta}_p)$ are known constants and $0 < c \leq L$. We wish to select one of the t best populations, with loss function given by (3.15).

For the above problem, we construct a Bayes decision rule relative to the following prior distribution:

For $j = 1,2,\ldots,\binom{k}{t}$, a total mass of b_j is spread uniformly over the set

$$\bar{\theta}_i - \frac{M_n(\bar{\theta})}{\sqrt{n}} \leq \theta_i \leq \bar{\theta}_i + \frac{M_n(\bar{\theta})}{\sqrt{n}} \quad i = 1,2,\ldots,p \tag{3.16}$$

and H_j is true; where $b_j \geq 0$ for $j = 1,2,\ldots,\binom{k}{t}$,

$$\sum_{j=1}^{\binom{k}{t}} b_j = 1 ,$$

and H_j is one of the $\binom{k}{t}$ hypothesis given by (3.14).

The prior distribution is arbitrary, but fixed. Later we select the prior in such a way as to obtain a minimax decision rule for the problem at hand.

Before developing a Bayes decision rule, we define some notation used later on. For $I = 1,2,\ldots,\binom{k}{t}$, let

$$(3.17)\quad b_I(n) = b_I \int_{\bar{\theta}_1 - \frac{M_n(\bar{\theta})}{\sqrt{n}}}^{\bar{\theta}_1 + \frac{M_n(\bar{\theta})}{\sqrt{n}}} \cdots\cdots \int_{\bar{\theta}_p - \frac{M_n(\bar{\theta})}{\sqrt{n}}}^{\bar{\theta}_p + \frac{M_n(\bar{\theta})}{\sqrt{n}}} \prod_{i=1}^{n} f(X_i;\theta,H(I))\,d\theta_1,\ldots,d\theta_p$$

Let the ranked values of $b_I(n)$ be denoted by

$$(3.18)\qquad b_{[1]}(n) \le b_{[2]}(n) \le \cdots\cdots \le b_{[(\binom{k}{t})]}(n)$$

Also for $j = 1,2,\ldots,\binom{k}{t}$, let

$$(3.19)\qquad \bar{b}_j(n) = b_j(n) \Big/ \sum_{\ell=1}^{\binom{k}{t}} b_\ell(n)$$

Then we may view $\bar{b}_j(n)$ as the aposteriori probability for hypothesis H_j, if we consider b_i as the apriori probability.

Using the above notation, for $j = 1,2,\ldots,k$, we obtain

$$(3.20)\qquad k(D_j;x) = \sum_{\text{all } I} b_I(n) - \sum_{\{I = (i_1,i_2,\ldots,i_k) \mid i_j = 1\}} b_I(n)$$

Thus a Bayes decision rule is to select population Π_j (decision D_j) as a best population if $\sum_{\{I \mid i_j = 1\}} b_I(n)$ is maximized.

Note that for each population Π_j, $j = 1,2,\ldots,k$ there are $\binom{k-1}{t-1}$ terms in the summation $\sum_{\{I \mid i_j = 1\}} b_I(n)$. Also there are many common terms occurring in each summation being compared. We rewrite the Bayes decision rule (for convenience in analysis) as follows:

Select Π_ℓ if

$$(3.21) \qquad \sum_{\{I|i_\ell=1\}} b_I(n) \geq \sum_{\{I|i_j=1\}} b_I(n) \qquad \begin{matrix} j \neq \ell \\ \ell,j = 1,2,\ldots,k \end{matrix}$$

Eliminating the common terms from the above, the Bayes decision rule reduces to

Select Π_ℓ if

$$(3.22) \qquad \sum_{\substack{\{I|i_\ell=1\} \\ i_j=0}} b_I(n) \geq \sum_{\substack{\{I|i_j=1\} \\ i_\ell=0}} b_I(n) \qquad \begin{matrix} j \neq \ell \\ \ell,j = 1,2,\ldots,k \end{matrix}$$

Now there are exactly $\binom{k-2}{t-1}$ terms in each summation above. Because of the symmetry in the two summations, there is for every term in the summation at the left, exactly one term in the summation at the right which has the same I except for the obvious change from $(i_\ell = 1,\ i_j = 0)$ to $(i_\ell = 0,\ i_j = 1)$. Thus we may rewrite the Bayes decision rule as:

Select Π_ℓ as one of the t best populations if

$$(3.23) \qquad \frac{b_{I(i_\ell=1,\ i_j=0;FP)}^{(n)}}{b_{I(i_j=1,\ i_\ell=0;FP)}^{(n)}} \geq 1 \qquad \begin{matrix} j \neq \ell \\ \ell,j = 1,2,\ldots,k \end{matrix}$$

where FP is a fixed permutation of the remaining $(k-2)$ terms of I.

We now study the asymptotic properties of the Bayes decision rule given by (3.23). The basic results obtained in Section 2.4 apply in this case too. The method of analysis parallels that used in Section 2.4 and hence will not be repeated here. An asymptotically minimax decision rule reduces to the following:

For $\ell = 1,2,\ldots,k$ select Π_ℓ as one of the t best population if

$$W_n(\ell|j) \geq 0 \tag{3.24}$$

where $W_n(\ell|j)$ is as defined by (2.54).

The decision rule given by (3.24) is Bayes decision rule for the artificial k-decision problem for the symmetric prior distribution, $\{b_j = \frac{1}{\binom{k}{t}},\ j = 1,2,\ldots,\binom{k}{t}\}$. As in Chapter 2, the symmetric prior gives a minimax decision rule because of the natural symmetry in the ranking problem being considered.

Optimal Procedure for the "Real" Problem

The above sequence of problems was artificial because we assumed $\overline{\theta}_1,\overline{\theta}_2,\ldots,\overline{\theta}_p$ were known. We now proceed to the real problem where nothing is known about the nuisance parameters $(\theta_1,\theta_2,\ldots,\theta_p)$.

Let $\hat{\theta}_1(n),\hat{\theta}_2(n),\ldots,\hat{\theta}_p(n)$ be estimators of $\theta_1,\theta_2,\ldots,\theta_p$ respectively, based on the observations $(X_1,\ldots,X_n)$. We assume, as in Chapter 2, that these estimators satisfy the consistency condition (2.74). MPE could be

used, although any set of consistent estimators would suffice.

For notational convenience, let

$$(3.25) \qquad \hat{\theta}(n) = (\bar{\theta}_0, \hat{\theta}_1(n), \hat{\theta}_2(n), \ldots, \hat{\theta}_p(n), \bar{\theta}_{p+1}, \bar{\theta}_{p+2}, \ldots, \bar{\theta}_{p+k})$$

Using (3.13) and (3.14), an asymptotically optimal procedure for the real problem is given by the following theorem.

Theorem 3.2. An asymptotically optimal decision rule for selecting one of the t best populations is to select the population associated with $T_{[k]}(n)$.

The proof of the theorem parallels the proof of Theorem 2.4.

3.3 Asymptotically Optimal Procedure for Goal I

In Sections 3.1 and 3.2, we have considered special cases of Goal I and developed asymptotically optimal procedures for the special cases. In this section, we obtain an asymptotically optimal procedure for the more general Goal I.

The analysis for this problem is very similar to the one outlined in Section 3.2 and will not be repeated. The least favorable configuration for this problem is given by (3.14).

Let I be as defined earlier in Section 3.2 and let J denote $(j_1, j_2, \ldots, j_k)$, where for $t = 1,2,\ldots,k$, j_t is equal to one or zero and exactly s of the j_t are equal to one. Also let $D(J)$ denote the decision to select the s populations for which $j_t = 1$ in J.

For the ranking problem being considered, there are $\binom{k}{t}$ states of nature and $\binom{k}{s}$ possible decisions. The loss function for the $\binom{k}{s}$-decision problem is given as

$$W(\theta;D(J)) = \begin{cases} 0 & \text{if } \sum_{\ell=1}^{k} i_\ell \cdot j_\ell \geq d \\ 1 & \text{otherwise} \end{cases} \tag{3.26}$$

when $I = (i_1, i_2, \ldots, i_k)$ is the state of nature and $D(J)$ is the decision made.

The loss function, given by (3.26), implies that the loss is zero if the selected subset of size s contains at least d of the t best populations; otherwise the loss is equal to one.

For the ranking problem being considered, with the loss function defined by (3.26), using (3.11) and (3.12), an asymptotically optimal procedure is given by the following theorem.

Theorem 3.3. An asymptotically optimal procedure for selecting a subset of size s to contain at least d of the t best populations is to select the s populations associated with $T_{[k]}(n), T_{[k-1]}(n), \ldots, T_{[k-s+1]}(n)$.

The proof of the theorem parallels that of Theorem 2.4 (with added notational complexities) and will not be repeated.

3.4 Asymptotically Optimal Procedure for Goal II

In order to develop an asymptotically optimal procedure for the ranking problem, we formulate the problem in a decision theoretic structure.

Let $I = (i_1, i_2, \ldots, i_k)$, where for $j = 1,2,\ldots,k$, i_j is equal to an integer between 0 and t. Exactly (k-t) i_j are zero and of the remaining exactly one i_j is equal to integer m ($m = 1,2,\ldots,t$). Thus each I identifies a particular partition of the k population into t ordered "best set" of populations and a set of unordered (k-t) "worst" populations. ("best set" here implies the set of populations in which the best, second best, third best, etc., and finally t^{th} best population are identified). Thus there are $\frac{k!}{(k-t)!}$ distinct possible I, each corresponding to a particular state of nature. (We are excluding from our consideration any ties which may occur for the set of t best populations). For any one of the $\frac{k!}{(k-t)!}$ I, let

$$(3.27) \quad H(I): \quad \theta_{p+j} = \begin{cases} \bar{\theta}_{p+j} + c/\sqrt{n} & \text{if } i_j = 0 \\ \bar{\theta}_{p+j} - (2i_j - 1)\frac{c}{\sqrt{n}} & \text{if } i_j > 0 \end{cases} \qquad j = 1,2,\ldots,k .$$

where $c > 0$ and $\bar{\theta}_{p+j} = \bar{\theta}_{p+i}$ ($i,j = 1,2,\ldots,k$) are known values. $\bar{\theta}_{p+j}$ may be taken to be zero with no loss in generality. Here, (3.27) represents a parameter configuration in which the best population has a value of the parameter of interest ψ, $\frac{2c}{\sqrt{n}}$ units greater than the second best population, which in turn is $\frac{2c}{\sqrt{n}}$ units greater than the third best population, etc., and finally the t^{th} best population has a value of the parameter ψ $\frac{2c}{\sqrt{n}}$ units greater than the remaining (k-t) worst populations. It follows directly that this type of parameter configuration is the LFC for the problem at hand under the indifference zone approach that we are considering.

(See Bechhofer [4] for definition of the indifference zone for such a problem). However the indifference zone approach to this problem does not restrict itself to considering least favorable configurations (LFC's) in which the i^{th} best and (i+1)st best populations $(i = 1,2,\ldots,t)$ have the same distance between them $(= \frac{2c}{\sqrt{n}}$ in our notation). The more general setup with $\psi_{[k-i]} - \psi_{[k-i-1]} = \frac{c_i}{\sqrt{n}}$ $(i = 0,1,2,\ldots,t-1)$ could be considered too; but since that only introduces more notational complexities in the problem without adding anything new, we will only consider the LFC as given by (3.27) (with the understanding that suitable changes would be made in computing the probability of correct selection when using the procedure for a more general setup. The procedure given below is optimal even for the more general setup.)

Let $J = (j_1,j_2,\ldots,j_k)$, where for $a = 1,2,\ldots,k$, j_a is equal to an integer between 0 and t. Exactly $(k-t)$ j_a are zero and of the remaining, exactly one j_a is equal to the integer m $(m = 1,2,\ldots,t)$.

Let $D(J)$ denote the decision to select Π_{j_a} $(a = 1,2,\ldots,k)$ as the $(t-j_a+1)^{st}$ best population (for $j_a > 0$). The remaining $(k-t)$ populations are identified as the set of worst populations. Since there are $\frac{k!}{(k-t)!}$ distinct J's, we get $\frac{k!}{(k-t)!}$ distinct possible decisions.

With the above notation, the loss function for this problem is given as follows:

For any particular decision $D(J)$ and state of nature I,

$$W(\theta;D(J)) = \begin{cases} 0 & \text{if } J = I \\ 1 & \text{otherwise} \end{cases} \tag{3.28}$$

The simple loss function (3.28) implies that the loss is zero when the right set of t populations are identified as the best set; and the loss is one otherwise.

Thus we have a multiple decision problem with $\frac{k!}{(k-t)!}$ possible decisions and $\frac{k!}{(k-t)!}$ possible states of nature (exactly one decision is correct for each state of nature). This problem is very similar to the problem of selecting the best population (k possible decisions and k states of nature) considered in Chapter 2. The analysis involved in developing an asymptotically optimal procedure for this problem is a complete repetition of the analysis presented in Section 2.4, with obvious modifications introduced by using (3.27) instead of (2.17) for the least favorable configuration for the ranking problem. We thus do not repeat the analysis and simply given an asymptotically optimal procedure for the problem.

Using (3.11) and (3.12), an asymptotically optimal procedure is given by the following theorem.

Theorem 3.4 An asymptotically optimal procedure for selecting the t best populations with regard to order is to select the populations associated with $T_{[k]}(n), T_{[k-1]}(n), \ldots, T_{[k-t+1]}(n)$ as the best, the second best, etc., and the t^{th} best population respectively.

3.5 Certain Other Possible Goals in Ranking and Selection Problems

The ranking goals considered thus far include almost all the goals considered in the literature on ranking and selection problems (using the indifference zone approach). As pointed out earlier, it is assumed that

the common functional form of the density function of observations from each of the populations is known. The populations are ranked according to values of a parameter of interest ψ. There may or may not be other unknown parameters in the distribution, which, when present, would be classified as nuisance parameters for the ranking problem.

In actual practical applications, there may be situations where the criterion for "goodness" of populations is more complicated and must be properly interpreted before a ranking procedure is developed. For example, in many engineering applications, the quality of manufactured goods is characterized by the product meeting some fixed specifications. Thus, for example, one may be interested in selecting a manufacturing process which has the highest probability of coverage of the specification interval. For this problem "Converge Probability" is the criterion for "goodness" of populations. Guttman [22] and Guttman and Milton [23] consider this problem and have developed procedures for selecting a (random) subset to contain the best population for normal and exponential density functions, when a one sided fixed tolerance region is the criterion for "goodness".

In the example cited above and other related ranking problems, the criterion for "goodness" of populations may be specified, but ranking procedures can be developed only after the criterion is translated into a goal involving the parameters of the population. Two special cases of interest may arise. First, it may be possible to redefine the parameters of the populations in such a way that the "criterion" is translated into a parameter of interest ψ and there may or may not be any

nuisance parameters. If the redefined parameters satisfy the assumptions and regularity conditions of Section 2.1, then the entire analysis carried out earlier in this thesis would be applicable. If the regularity conditions are not satisfied, one may proceed to develop asymptotically optimal procedure as outlined in Chapter 4 for non-regular cases. Secondly, if after redefining the parameters, the nuisance parameters are functions of the parameter being ranked, then special analysis would be needed for each particular case. To our knowledge, the only paper in the literature dealing with such a case is by Chambers and Jarratt [12], which deals with the problem of ranking means of populations when the variances are a known function of the means being ranked.

We do not propose to consider problems falling into this second category in this thesis. However, it should be pointed out that the general method developed would be applicable to this class of problems, separate analysis being required for each specific case (each specific case being characterized by a known functional relation between the nuisance parameters and the parameter being ranked).

In our analysis we have assumed that the correct pairing of the populations and the ranked parameters are unknown, but the values of the parameters of interest are assumed to be unknown. Dunnett [16] and Guttman and Tiao [24] deal with the normal means problem when prior information is available about the possible values of parameters ψ_i $(i = 1,2,..,k)$ and/or the correct pairings of the populations and ranked parameters. The ranking goals considered are the same as the ones considered by us.

Although we do not concern ourselves with such Bayesian analyses of the ranking problem, we should mention that the procedures that we have developed are also asymptotically optimal from a Bayesian viewpoint; that is, when the apriori information is available about the possible pairings of populations and the ranked parameters. This is so since asymptotically minimax (optimal) ranking procedures were constructed from Bayes decision rules for suitably defined multiple decision problems by suitable choice of prior distribution.

Lastly we would like to point out that, in our analysis, we have assumed that we know the functional form of the joint density function of observations from the set of populations. This was used explicitly in constructing asymptotically optimal decision rule. Hence, we are not concerned with a class of nonparametric ranking problems, in which the form of the joint density function is not known. Different type of analysis would be needed for such ranking problems.

CHAPTER 4

RANKING PROBLEMS IN NON-REGULAR FAMILIES OF DISTRIBUTIONS

To this point, we have developed asymptotically optimal procedures for certain ranking goals, for situations in which the joint density function of the observations satisfies certain mild regularity conditions imposed in Section 2.1. There are many functions which do not satisfy these regularity conditions, but which occur frequently in practice.

The general method developed in previous chapters to obtain an asymptotically optimal procedure for certain ranking goals, using the regularity conditions, can often be used for the non-regular cases too. However each non-regular case must be treated separately. The basic idea of using local Bayes rules to develop asymptotically optimal procedures in the presence of several nuisance parameters would be useful in all such cases, the actual analysis being different in each case (due to say, different normalizing constants and the actual functional forms of the distributions).

In order to illustrate the applicability of our method to non-regular cases, we consider some particular non-regular density functions and develop asymptotically optimal procedures for the ranking goals considered earlier in this thesis for regular cases.

Before we consider such problems, we list certain non-regular density functions which may occur frequently in practical situations. The list contains most of the interesting known cases. Most of these are taken from Weiss and Wolfowitz ([51] and [54]).

4.1 Some Non-Regular Density Functions

Let $f(x|\theta)$ denote the density function of a random variable X, characterized by parameter θ $(= (\theta_1,\theta_2,\ldots,\theta_m))$, which in general may be multidimensional. Let Θ denote the parameter space, a suitably defined subspace of m dimensional Euclidean space (m being the dimension of θ). Whenever $m = 1$, the parameter is denoted by θ and for $m > 1$, θ denotes the vector of parameters, each component being denoted as θ_i, $i = 1,2,\ldots,m$.

The following constitute some interesting "non-regular" univariate density functions.

I $\quad f(x|\theta) = 0 \quad$ for $x < \theta_1$

$\quad f(\theta_1+|\theta) = h(\theta) > 0$

where, $\Theta = \{(\theta_1,\theta_2)|-\infty < \theta_1 < \infty,\ 0 < \theta_2 < \infty\}$

Example

(4.1) $$f(x|\theta) = \begin{cases} \frac{1}{\theta_2} e^{-\frac{(x-\theta_1)}{\theta_2}} & \text{for } x \geq \theta_1 \\ 0 & \text{otherwise} \end{cases}$$

and $\Theta = \{(\theta_1,\theta_2)|-\infty < \theta_1 < \infty,\ 0 < \theta_2 < \infty\}$.

II $\quad f(x|\theta) = 0$ if $x < \theta$ or $x > B(\theta)$

$\quad f(\theta+|\theta) = g(\theta) > 0$

$\quad f(B(\theta)-|\theta) = h(\theta) > 0$.

g, h and $B' = \frac{dB}{d\theta}$ are continuous functions of θ. Also $B(\theta) > \theta$. (Note that I is a special case of II with $B(\theta) = \infty$). Θ is suitably defined for each particular case.

Examples (i) $B' < 0$

$$f(x|\theta) = \begin{cases} \frac{\theta}{1-\theta^2} & \text{for} \quad \theta < x < 1/\theta \\ 0 & \text{otherwise} \end{cases} \tag{4.2}$$

where,

$$\Theta = \{\theta | 0 < \theta < 1\}$$

(ii) $B' > 0$

$$f(x|\theta) = \begin{cases} 1 + \frac{3}{2}(x-\theta) & \text{for} \quad \theta < x < \theta + 2/3 \\ 0 & \text{otherwise} \end{cases} \tag{4.3}$$

where,

$$\Theta = \{\theta | -\infty < \theta < \infty\}$$

(iii) One of the end points is a constant; i.e., $B' = 0$.

This case can be represented by the following:

$$f(x|\theta) = \begin{cases} 1/\theta & \text{for} \quad 0 < x < \theta \\ 0 & \text{otherwise} \end{cases} \tag{4.4}$$

where, $\Theta = \{\theta | 0 < \theta < \infty\}$

III

$$(4.5) \qquad f(x|\theta) = \begin{cases} B & \text{for } 0 \leq x \leq \theta \\ \frac{1-Bx}{1-\theta} & \text{for } \theta < x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

where B is a known constant and

$$\Theta = \{\theta | 0 < \theta < 1\}$$

IV

$$f(x|\theta) = 0 \qquad x < \theta_1 \quad \text{or} \quad x > \theta_2$$

$$f(\theta_1 + |\theta) = g((\theta_1, \theta_2)) > 0$$

$$f(\theta_2 - |\theta) = h((\theta_1, \theta_2)) > 0$$

g,h are continuous functions of (θ_1, θ_2).

Example

$$(4.6) \qquad f(x|(\theta_1, \theta_2)) = \begin{cases} \frac{1}{\theta_2 - \theta_1} & \text{for } \theta_1 \leq x \leq \theta_2 \\ 0 & \text{otherwise} \end{cases}$$

where,

$$\Theta = \{(\theta_1, \theta_2) | \theta_2 > \theta_1\}$$

V Double Exponential Density (Laplace distribution)

$$(4.7) \qquad f(x|\theta) = \frac{1}{2\theta_2} e^{-\frac{1}{\theta_2}|x-\theta_1|} \quad \text{for } -\infty < x < \infty$$

where,

$$\Omega = \{(\theta_1,\theta_2) | -\infty < \theta_1 < \infty, \quad 0 < \theta_2 < \infty\}$$

VI Weibull Distribution

$$(4.8) \qquad F(x|\theta) = \begin{cases} 0 & \text{for } x < \theta_1 \\ 1 - e^{-\left(\frac{x-\theta_1}{\theta_2}\right)^{1/\theta_3}} & \text{for } x \geq \theta_1 \end{cases}$$

where,

$F(x|\theta) = \int_{-\infty}^{x} f(y|\theta)dy$, is the cdf of the Weibull distribution.

Also,

$$\Omega = \{(\theta_1,\theta_2,\theta_3) | -\infty < \theta_1 < \infty, \quad 0 < \theta_2 < \infty, \quad 0 < \theta_3 < \infty\}$$

For this class of distributions, θ_1 is the "location" parameter, θ_2 is the "scale" parameter and θ_3 is the "shape" parameter. If $\theta_3 = 1$, we obtain the exponential density function (as given by (4.1)) a special case of the class of Weibull distribution. If θ_1 is known, (4.8) reduces to the regular case.

Before proceeding to develop optimal procedures for some of the density functions listed above, we review the literature in ranking and selection procedures for non-regular cases.

4.2 Brief Review of Literature on Ranking Problems in Non-Regular Cases

Sobel [43] developed sequential procedures for ranking scale parameters of exponential populations, (as given by (4.1) in Section 4.1 and θ_2 is the scale parameter) with known location parameters (same or different for each population) or a common unknown location parameter. In the case of a common unknown location parameter, Sobel assumed that if the unknown location parameter was greater than or equal to zero, zero is taken to be the value of the unknown location (nuisance) parameter and a procedure satisfying the basic probability requirement was developed.

Bechhofer, Kiefer and Sobel [7] developed sequential procedures for ranking problems associated with the Koopman-Darmois family of distributions; these would be applicable to ranking scale parameters of exponential populations when the location parameters are known.

Barr and Rizvi [2] developed a simple-stage procedure for ranking uniform distributions (density function as given by (4.4) in Section 4.1). Using a zero-one type loss function (as we have done throughout), Barr and Rizvi showed that the procedure they developed is minimax and is a most-economical decision rule (in the sense of Hall [25]). They also showed that the selection procedure may easily be extended to a larger class of non-regular distributions, given in Hogg and Craig [27]. This class of distributions corresponds to Case II in Section 4.1.

Dudewicz [15] in determining the efficiency of a non-parametric selection procedure given by Bechhofer and Sobel [9] (for the location parameter case) against parametric alternatives, proposed single stage procedures for ranking from uniform distributions (uniform between

$(\theta-1/2,\ \theta+1/2))$. (The nonparametric selection procedure of Bechhofer and Sobel [9] is also given as a selection procedure in Bechhofer, Elmaghraby and Morse [6]). Dudewicz [15] proposed a midrange procedure (i.e., populations are ranked according to the values of $\frac{(\text{max. of obs})+(\text{min. of obs})}{2}$) and a means procedure (i.e., populations are ranked according to the values of the sample means) for ranking from uniform distributions.

Mahamunulu ([32], [33]) proposed a single-stage procedure for ranking k populations $\{\Pi_i,\ i = 1,2,\ldots,k\}$, the selection procedure based on suitable statistics $T_1,T_2,\ldots,T_k$. T_i is computed from a random sample of size n from Π_i $(i = 1,2,\ldots,k)$. The proposed procedure is applicable to ranking populations according to Goal I (defined earlier in Chapter 3) and for situations in which T_i is an absolutely continuous random variable and its distribution function is stochastically increasing in the parameter being ranked (see Lehmann [30] for definition of stochastic increasing family of distributions).

The procedure developed by Mahamunulu is applicable to ranking problems associated sith the following non-regular families of distributions: (i) uniform distribution (Case II in Section 4.1) (ii) exponential distribution (Case I in Section 4.1) and (iii) double exponential distribution (Case V in Section 4.1). The proposed procedures are applicable when no nuisance parameters are present (or they are arbitrarily removed from explicit consideration). Mahamunulu also indicated how the tables developed for the case of ranking means of normal distribution with

common known variances could be used as an approximation for ranking from other distributions, whenever the statistics used for ranking the populations (T_i) converged asymptotically to a normal distribution.

Mahamunulu ([33], p. 1082) pointed out that if a sufficient statistic of fixed dimension for all n, exists (for the parameter being ranked), then T_i is some appropriate function of the sufficient statistic. The choice of T_i becomes a problem only when such a sufficient statistic does not exist. In the latter case, the author advocates using a statistic such that the induced family of distributions is a stochastically increasing family of distributions.

4.3 Asymptotically Optimal Procedures for Ranking Non-Regular Exponential Distributions

We now proceed to develop asymptotically optimal procedures for certain ranking problems associated with populations having non-regular exponential distributions. We develop, in detail, asymptotically optimal procedures for selecting the best population (defined appropriately). We then develop an asymptotically optimal procedure for other ranking goals considered earlier in this thesis. Later, we consider non-regular uniform density functions and develop asymptotically optimal procedures for the above mentioned ranking goals. The method of analysis used in this section carries over to most of the non-regular cases.

Formulation of the Problem

We formulate the problem in decision theoretic structure using the notation defined in Chapter 2.

Let X_{ti} $i = 1,2,\ldots,n$ denote n independent observations from Π_t, $t = 1,2,\ldots,k$, each with pdf $f_t(\cdot|\cdot)$ given as

$$(4.9) \qquad f_t(x|\theta) = f(x|L_t,S_t) = \begin{cases} \frac{1}{S_t} e^{-\left(\frac{x-L_t}{S_t}\right)} & x \geq L_t \\ 0 & \text{otherwise} \end{cases}$$

where L_t, S_t denote the two unknown parameters of the distribution characterizing population Π_t. Here L_t denotes the location parameter and S_t, the scale parameter.

Two cases of interest arise here.

(i) L_t is the parameter being ranked; S_t is the nuisance parameter.

(ii) $1/S_t$ is the parameter being ranked; L_t is the nuisance parameter.

Each case is treated separately below. We first define some additional notation.

Case (i) S_t is the common unknown nuisance parameter

For $t = 1,2,\ldots,k$, let

$$(4.10) \qquad \psi_t = L_t$$

$$(4.11) \qquad S_t = \theta_1$$

and

$$(4.12) \qquad \psi_t = \theta_0 - \theta_{1+t} \quad .$$

Also for $i = 1,2,\ldots,k$, let

$$(4.13) \qquad H_i = \begin{cases} \theta_{1+i} = \overline{\theta}_{1+i} - c/n \\ \theta_{1+j} = \overline{\theta}_{1+j} + c/n \qquad j = 1,2,\ldots,k \neq i \end{cases}$$

where $c > 0$ and $\overline{\theta}_{1+i} = \overline{\theta}_{1+j}$, all i,j $(i,j = 1,2,\ldots,k)$ are known values. $\overline{\theta}_{1+i}$ is taken to be zero, with no loss in generality.

The above relations imply that we are ranking location parameters, with the common unknown scale parameter being the nuisance parameter.

In the parameter configuration given by (4.13), selecting H_i as the true hypothesis is equivalent to selecting Π_i as the best population $(i = 1,2,\ldots,k)$. The above hypotheses also imply that the best population has a ψ-value which is $\frac{2c}{n}$ units greater than that of the remaining (k-1) populations. This corresponds to the least favorable configuration (LFC) for the ranking problem under consideration under the indifference zone approach.

For $i = 1,2,\ldots,k$, let D_i denote the decision to select Π_i as the best population. The loss function for the k-decision problem is given as

$$(4.14) \qquad W(\theta;D_i) = \begin{cases} 0 & \text{if } H_i \text{ is the true hypothesis} \\ 1 & \text{otherwise} \end{cases}$$

Asymptotically optimal decision rules (procedures) will be developed for the above problem.

Case (ii) L_t is the common unknown nuisance parameter

For $t = 1,2,\ldots,k$, let

$$\psi_t = 1/S_t \tag{4.15}$$

$$L_t = \theta_1 \tag{4.16}$$

and,

$$\psi_t = \theta_0 - \theta_{1+t} \tag{4.17}$$

Also for $i = 1,2,\ldots,k$, let

$$H_i: \begin{cases} \theta_{i+i} = \bar{\theta}_{1+i} - c/\sqrt{n} \\ \theta_{1+j} = \bar{\theta}_{1+j} + c/\sqrt{n} \qquad j = 1,2,\ldots,k,\ j \neq i \end{cases} \tag{4.18}$$

where $c > 0$, $\bar{\theta}_{i+j} = \bar{\theta}_{1+i}$ all i,j $(i,j = 1,2,\ldots,k)$ are known constants $\bar{\theta}_{1+i}$ is taken to be zero, with no loss of generality.

In this case θ_1, the common unknown location parameter, is the nuisance parameter for the problem of ranking scale parameters. Here, (4.18) represents the least favorable configuration for the ranking problem under consideration.

For $i = 1,2,\ldots,k$, let D_i denote the decision to select Π_i as the best population. The loss function for the k-decision problem is given by (4.14).

Preliminary Sequence of Artificial Problems: S_t is the nuisance parameter

Before proceeding to the real problem, we solve a sequence of artificial k-decision problems (one for each n) for case (i).

Suppose it is known that $\theta_1 = \bar{\theta}_1$ and $(\theta_2,\theta_3,\ldots,\theta_{k+1})$ satisfy one of the k hypotheses given by (4.13). $\bar{\theta}_1$ is a known constant and $0 < c \leq L$. We wish to test which one of the k hypotheses $H_1,H_2,\ldots,H_k$ (given by (4.13)) is the true hypothesis. The loss function is of the zero-one type, given by (4.14).

For the above problem, we construct a Bayes decision rule relative to the following apriori distribution:

For $j = 1,2,\ldots,k$, b_j is the apriori probability that H_j is true hypothesis, where for $j = 1,2,\ldots,k$, $b_j \geq 0$ and $\sum_{j=1}^{k} b_j = 1$.

In order to obtain a Bayes decision rule, we must compute for $i = 1,2,\ldots,k$, $k(D_i;x)$, and it can be easily seen that relative to the above apriori distribution, and the loss function given by (4.14), a Bayes decision rule reduces to the following:

For $\ell = 1,2,\ldots,k$, select H_ℓ as the true hypothesis (equivalently Π_ℓ as the best population) if

$$J_n(\ell|j) \geq \frac{b_j}{b_\ell} \qquad j = 1,2,\ldots,k,\ j \neq \ell \tag{4.19}$$

where, for $\ell, j = 1,2,\ldots,k$,

$$J_n(\ell|j) = \frac{\prod_{i=1}^{n} f(X_i;\bar{\theta}_1,H_\ell)}{\prod_{i=1}^{n} f(X_i;\bar{\theta}_1,H_j)} \tag{4.20}$$

Here $f(X_i;\bar{\theta}_1,H_j)$ denotes the joint pdf of $(X_{1i},X_{2i},\ldots,X_{ki})$, when $(\bar{\theta}_1,\theta_2,\ldots,\theta_{k+1})$ is the true parameter value for X and $(\theta_2,\theta_3,\ldots,\theta_{k+1})$ satisfy H_j, as given by (4.13).

Since we assume that the observations are independent,

$$f(X_i;\theta,H_j) = \prod_{t=1}^{k} f_t(X_{ti};\theta,H_j) \tag{4.21}$$

where $f_t(X_{ti};\theta,H_j)$ is given by (4.9).

Using (4.9)-(4.12) and (4.21), for $j = 1,2,\ldots,k$

$$\prod_{i=1}^{n} f(X_i;\theta,H_j) = \prod_{t=1}^{k} \left[\prod_{i=1}^{n} f_t(X_{ti};\theta,H_j) \right] \tag{4.22}$$

where, for $t \neq j$

$$\prod_{i=1}^{n} f_t(X_{ti};\theta,H_j) = \begin{cases} 0 & \text{if } \min_{1\le i\le n}(X_{ti}) < \theta_0 - c/n \\ \theta_1 \exp\left(-\frac{1}{\theta_1}\sum_{i=1}^{n}(X_{ti}-\theta_0+c/n)\right) & \text{otherwise} \end{cases} \tag{4.23}$$

and for $t = j$

$$(4.24)\quad \prod_{i=1}^{n} f_t(X_{ti};\theta,H_t) = \begin{cases} 0 & \text{if } \min\limits_{1\le i\le n} (X_{ti}) < \theta_0 + c/n \\ \theta_1^{-n} \exp\left(-\frac{1}{\theta_1} \sum\limits_{i=1}^{n} (X_{ti}-\theta_0-c/n)\right) & \text{otherwise} \end{cases}$$

For $t = 1,2,\ldots,k$, let

$$(4.25)\quad Y_t = \min_{1\le i\le n} (X_{ti})$$

and denote the ranked values of Y_t by

$$(4.26)\quad Y_{[1]} \le Y_{[2]} \le \cdots \le Y_{[k]}$$

Using (4.21)-(4.25), we obtain, for $j = 1,2,\ldots,k$,

$$(4.27)\quad \prod_{i=1}^{n} f(X_i;\bar{\theta}_1,H_j) = \begin{cases} \bar{\theta}_1^{-nk} \exp\left[-\frac{1}{\bar{\theta}_1} \sum\limits_{t=1}^{k} \sum\limits_{i=1}^{n} (X_{ti}-\theta_0)\right] \exp\left[-\frac{c}{n\bar{\theta}_1}(k-2)\right] \\ \quad \text{if } (Y_j > \theta_0 + c/n) \bigcap\limits_{\substack{i=1 \\ \neq j}}^{k} (Y_i > \theta_0 - c/n) \\ 0 \quad \text{otherwise} \end{cases}$$

If we denote, for $j = 1,2,\ldots,k$,

$$(4.28)\quad b_j(n) = b_j \prod_{i=1}^{n} f(X_i,\bar{\theta}_1,H_j)$$

then the Bayes decision rule may be rewritten as follows:

For $\ell = 1,2,\ldots,k$, select Π_ℓ as best population if

$$b_\ell(n) \geq b_j(n) \qquad \begin{matrix} j \neq \ell \\ j = 1,2,\ldots,k \end{matrix} \tag{4.29}$$

From (4.27), we note that for any given set of observations, the joint density function under each H_j is the same, whenever the joint density function is positive under each H_j. For certain sets of observations, the joint density function is equal to zero under each H_j (such cases represent regions of probability zero and may be ignored from consideration in constructing Bayes decision rules). For the remaining sets of possible observations, the joint density function may be zero under some hypotheses and equal to the same positive value under the remaining hypotheses (as is clear from (4.27)).

Since we are interested in an optimal (minimax) decision rule for the artificial sequence of problems, we need not rewrite (4.29) to develop simplified expressions for a Bayes decision rule for any general apriori distribution. We note, by the inherent symmetry in the problem that a minimax decision rule will be given by the Bayes decision rule with the prior given by

$$b_j = \frac{1}{k} \qquad j = 1,2,\ldots,k \tag{4.30}$$

For this prior, the Bayes decision rule, which is a minimax decision rule too, is not unique because of ties in the $b_j(n)$ (as defined by

(4.28)). A minimax decision rule is given by the following:

Select π_ℓ as the best population if

$$Y_\ell \geq Y_j \qquad j = 1,2,\ldots,k \quad (j \neq \ell) \tag{4.31}$$

For the minimax decision rule given by (4.31), one is interested in computing the probability of correct selection. This will be needed in showing that the decision rule for the real problem is asymptotically minimax.

Note that if X_{ti} $(i = 1,2,\ldots,n)$ have a distribution given by (4.9), then

$$\begin{aligned} P\{nY_t \leq y\} &= P\{\min_{1\leq i\leq n} X_{ti} \leq y/n\} \\ &= 1 - P\{\min_{1\leq i\leq n} X_{ti} \geq y/n\} \\ &= 1 - \prod_{i=1}^{n} (e^{-(y/n - \psi_t)/S_t}) \end{aligned}$$

or,

$$P\{nY_t \leq y\} = 1 - e^{-y/S_t} e^{n\psi_t/S_t} \tag{4.32}$$

From the above, it follows that

$$P\{n(Y_t - \psi_t) \leq y\} = 1 - e^{-y/S_t} \tag{4.33}$$

Thus from (4.12), (4.13) and (4.31) and denoting the common known S_t by $\bar{\theta}_1$, we obtain

$$P(CS) = P(Y_i \leq Y_{(k)} \quad i = 1,2,\ldots,k)$$

where $Y_{(k)}$ is the observation from the population with largest ψ. Rearranging, we obtain

$$P(CS) = P(n(Y_i - (\theta_0 - c/n) \leq n(Y_{(k)} - (\theta_0 + \frac{c}{n}) + 2c \quad i = 1,2,\ldots,k$$

$$= P(Z_i \leq Z_{(k)} + 2c \quad \text{all } i)$$

where the distribution of Z_i is given by (4.33).

Thus,

$$P(CS) = \int_0^\infty \left(1 - e^{-(\frac{z+2c}{\bar{\theta}_1})}\right)^{k-1} \frac{1}{\bar{\theta}_1} e^{-z/\bar{\theta}_1} \, dz$$

$$= \int_0^\infty \left(1 - e^{-\frac{2c}{\bar{\theta}_1}} e^{-t}\right)^{k-1} e^{-t} \, dt$$

Denoting e^{-t} by u,

$$P(CS) = \int_0^1 (1 - e^{-2c}u)^{k-1} \, du \tag{4.34}$$

where,

$$c^* = e^{-\frac{2c}{\theta_1}} \tag{4.35}$$

Evaluating the right hand side of (4.34), one can determine the P(CS) achieved for the optimal procedure for any given sample size n.

Asymptotically Optimal Procedure for the Real Problem

The above sequence of problems was artificial because we assumed $\overline{\theta}_1$ was known. We now develop an asymptotically optimal procedure for the real problem, where nothing is known about the value of θ_1.

Since the minimax procedure for the artificial problem did not use the information about θ_1, it follows immediately that an asymptotically optimal procedure for the problem is given by the following theorem.

Theorem 4.1 An asymptotically minimax procedure for selecting the best population is given by the following:

Select Π_ℓ as best population if

$$Y_\ell \geq Y_j \qquad \ell, j = 1,2,\ldots,k \quad j \neq \ell$$

where Y_ℓ is as defined by (4.25).

Proof: The decision variables have the same asymptotic distribution as the decision variables of the minimax (and Bayes) decision rule for the

the artificial problem. Thus the above procedure has the same asymptotic Bayes risk and is hence an asymptotically minimax decision rule.

Q.E.D.

It is interesting to note here that unlike the regular case, where local Bayes rules were used to construct an asymptotically optimal procedure, in the non-regular case being considered use of a simple Bayes rule for the artificial problem (where the nuisance parameter is known with certainty) enables one to construct an asymptotically optimal ranking procedure.

Preliminary Sequence of Artificial Problems: L_t is the Nuisance Parameter

We first solve a sequence of artificial problems (one for each n), before proceeding to the real problem for case (ii) (i.e., (4.15) - (4.18) hold).

Suppose it is known that $\theta_1 = \bar{\theta}_1$ and $(\theta_2,\theta_3,\ldots,\theta_{k+1})$ satisfy one of the k hypotheses given by (4.18). $\bar{\theta}_1$ is a known constant and $0 < c \leq L$. We wish to test which one of the k hypotheses, given by (4.18), is the true hypothesis; and the loss function is of the zero-one type, given by (4.14).

For the above problem, we construct a Bayes decision rule relative to the following apriori distribution:

For $j = 1,2,\ldots,k$, b_j is the apriori probability that H_j is the true hypothesis, where for $j = 1,2,\ldots,k$, $b_j \geq 0$ and $\sum_{j=1}^{k} b_j = 1$.

It can be seen that for the above problem, a Bayes decision rule relative to the above prior, reduces to the following:

For $\ell = 1,2,\ldots,k$ select H_ℓ as the true hypothesis (equivalently Π_ℓ as best population) if

$$J_n(\ell|j) \geq \frac{b_j}{b_\ell} \qquad \begin{matrix} j = 1,2,\ldots,k \\ \neq \ell \end{matrix} \tag{4.36}$$

where, for $\ell, j = 1,2,\ldots,k$

$$J_n(\ell|j) = \frac{\prod_{i=1}^{n} f(X_i;\bar{\theta}_1,H_\ell)}{\prod_{i=1}^{n} f(X_i;\bar{\theta}_1,H_j)} \tag{4.37}$$

Here $f(X_i;\bar{\theta}_1,H_j)$ denotes the joint pdf of $(X_{1i},X_{2i},\ldots,X_{ki})$ when $(\bar{\theta}_1,\theta_2,\ldots,\theta_{k+1})$ is the true parameter point and $(\theta_2,\theta_3,\ldots,\theta_{k+1})$ satisfy H_j, as given by (4.18).

From (4.9), it follows that

$$\prod_{i=1}^{n} f(X_i;\bar{\theta}_1,H_j) = \begin{cases} (\theta_0+c/\sqrt{n})(\theta_0-c/\sqrt{n})^{k-1} \exp\left[-(\theta_0+c/\sqrt{n}) \sum_{i=1}^{n} (X_{ji}-\bar{\theta}_1) - (\theta_0-c/\sqrt{n}) \sum_{i=1}^{n}\sum_{t=1}^{k} (X_{ti}-\bar{\theta}_1)\right] \\ \qquad \text{if } Y > \bar{\theta}_1 \\ 0 \qquad \text{otherwise} \end{cases} \tag{4.38}$$

where,

$$Y = \min_{\substack{1 \le t \le k \\ 1 \le i \le n}} \{X_{ti}\} \tag{4.39}$$

From (4.38), we get

$$J_n(\ell|j) = \frac{0}{0} \quad \text{if} \quad Y \le \bar{\theta}_1 \tag{4.40}$$

which occurs with probability zero and can be ignored; and

$$J_n(\ell|j) = \exp\{2c(Z_j - Z_\ell)\} \quad \text{otherwise} \tag{4.41}$$

where, for $\ell = 1,2,\ldots,k$,

$$Z_\ell = \sum_{i=1}^{n} (X_{\ell i} - \bar{\theta}_1)/\sqrt{n} \tag{4.42}$$

Thus the Bayes decision rule reduces to the following:

For $\ell = 1,2,\ldots,k$ select Π_ℓ as the best population if

$$Z_\ell \le Z_j - \frac{1}{2c} \log \frac{b_j}{b_\ell} \qquad j = 1,2,\ldots,k \ (\neq \ell) \tag{4.43}$$

In order to develop a minimax procedure for the problem, due to the inherent symmetry in the problem, the prior $\{b_j = \frac{1}{k} \quad j = 1,2,\ldots,k\}$ gives a minimax decision rule. Thus, a minimax decision rule for the artificial

problem, is given by the following:

Select Π_ℓ as the best population if

$$Z_\ell \leq Z_j \qquad \ell, j = 1,2,\ldots,k \quad \ell \neq j \tag{4.44}$$

The minimax decision rule is completely specified by (4.44); but we need to study the asymptotic behavior of the decision variables in order to develop an asymptotically optimal procedure for the real problem.

For $\ell = 1,2,\ldots,n$ and when $\bar{\theta}_1$ is the common location parameter, we have

$$E(X_{\ell i} - \bar{\theta}_1) = S_\ell \tag{4.45}$$

$$\text{Var } X_{\ell i} = S_\ell^2 \tag{4.46}$$

and by a central limit theorem,

$$\lim_{n\to\infty} P\left\{\sqrt{n}\,\frac{\dfrac{\sum_{i=1}^{n}(X_{\ell i} - \bar{\theta}_1)}{n}}{S_\ell} - S_\ell \leq y\right\} = \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}}\, e^{-z^2/2}\, dz \tag{4.47}$$

Thus (4.47) could be used for the asymptotic distribution of the decision variables.

Optimal Procedure for the Real Problem

The above sequence of problems was artificial because we assumed that $\bar{\theta}_1$ was known. We now develop asymptotically optimal procedure for the problem, where nothing is known about θ_1.

Let $\hat{\theta}_1(n)$ a consistent estimator of θ_1 be defined as

$$\hat{\theta}_1(n) = \min_{\substack{1\le t\le k \\ 1\le i\le n}} \{X_{ti}\} \tag{4.48}$$

where $\hat{\theta}_1(n)$ satisfies

$$P_{\theta_1}\{n|\hat{\theta}_1(n) - \theta_1| < D(\varepsilon)\} > 1 - \varepsilon \tag{4.49}$$

{H is true}

where {H is true} implies one of the p hypotheses given by (4.18) are true.

For $\ell = 1,2,\ldots,k$, let

$$\hat{Z}_\ell = \frac{1}{\sqrt{n}}\{\sum_{i=1}^{n} (X_{ti} - \hat{\theta}_1(n))\} \tag{4.50}$$

Then, an asymptotically optimal decision rule is given by the following theorem.

Theorem 4.2 An asymptotically optimal decision rule for the problem of selecting the best population (largest scale parameter) is given by:

For $\ell,j = 1,2,\ldots,k$, select Π_ℓ as the best population if

(4.51) $$\hat{Z}_\ell \leq \hat{Z}_j \qquad j \neq \ell$$

Proof: It is easy to see that

(4.52) $$\hat{Z}_\ell - \hat{Z}_j = Z_\ell - Z_j = \frac{1}{\sqrt{n}} \left(\sum_{i=1}^{n} X_{\ell i} - \sum_{i=1}^{n} X_{ji} \right)$$

Thus the decision variables given by (4.51) have the same asymptotic distribution as the minimax (and Bayes) decision rule for the artificial problem. Thus the decision rule given by (4.51) is an asymptotically minimax decision rule for the problem.

Q.E.D.

It may be noted that we could easily have defined $\hat{Z}_\ell$ by $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_{\ell i}$. However the above definition (i.e., $\hat{Z}_\ell$ given by (4.50)) was intentional since it allows us to observe what the optimal decision rule would look like when the populations have different unknown location parameters as the nuisance parameters for the problem. In such a situation, if we let, for $t = 1,2,\ldots,k$

(4.53) $$\hat{L}_t = \min_{1 \leq i \leq n} X_{ti}$$

and

(4.54) $$\hat{Z}_t = \frac{1}{\sqrt{n}} \left\{ \sum_{i=1}^{n} (X_{ti} - \hat{L}_t) \right\}$$

then an asymptotically optimal procedure is given by Theorem 4.2 with $\hat{Z}_t$ as defined by (4.54) (instead of (4.50)). We again note in this case that by using a simple Bayes rule one is able to construct an asymptotically optimal ranking procedure.

Optimal Procedures for Other Ranking Goals

In Chapter 3, we considered two general ranking goals which have been considered in the literature, and developed asymptotically optimal procedures for situations in which certain regularity conditions hold. We now develop an optimal procedure for the two general ranking goals considered in Chapter 3, for the case of the non-regular exponential density functions.

Instead of repeating the detailed analysis of Chapter 3, we will state the optimal ranking procedures for each goal in terms of the following theorems (using the notation defined earlier in the chapter).

Case (i) L_t is the ranking parameter; S_t is the common unknown nuisance parameter

Asymptotically optimal procedures for Goal I and Goal II are given by the following theorem:

Theorem 4.3 (i) An asymptotically optimal procedure for Goal I, (to select s populations to contain at least d of the t best populations) is to select the s populations associated with $Y_{[k]}, Y_{[k-1]}, \ldots, Y_{[k-s+1]}$.

(ii) An asymptotically optimal procedure for Goal II (selecting t best ordered populations, $1 \leq t \leq k$) is to select the t populations associated with $Y_{[k]}, Y_{[k-1]}, \ldots, Y_{[k-t+1]}$ as the best, the second best, etc., the t^{th} best population respectively.

Case (ii) $1/S_t$ is the ranking parameter; L_t is the common unknown nuisance parameter

Asymptotically optimal procedures for Goal I and Goal II are given by the following theorem:

Theorem 4.4 (i) An asymptotically optimal procedure for Goal I is to select populations associated with $\hat{Z}_{[k]}, \hat{Z}_{[k-1]}, \ldots, \hat{Z}_{[k-s+1]}$ as the best set of populations.

(ii) An asymptotically optimal procedure for Goal II is to select populations associated with $\hat{Z}_{[k]}, \hat{Z}_{[k-1]}, \ldots, \hat{Z}_{[k-t+1]}$ as the best, the second best, etc., the t^{th} best populations, respectively.

In this case $\hat{Z}_\ell$ is as defined by (4.50) and the ordered values are denoted by

$$\hat{Z}_{[1]} < \hat{Z}_{[2]} < \ldots < \hat{Z}_{[k]} \tag{4.55}$$

One may note here that Theorem 4.4 could be generalized to the case of different (unknown) location parameters L_t (by replacing $\hat{Z}_\ell$ as defined by (4.50), by $\hat{Z}_\ell$ as defined by (4.54)).

4.4 Asymptotically Optimal Procedure for Ranking Non-Regular Uniform Distributions

We now develop asymptotically optimal procedure for a non-regular uniform distribution.

Let X_{ti}, $i = 1,2,\ldots,n$ denote n independent observations from Π_t, $t = 1,2,\ldots,k$, each with pdf $f_t(\cdot|\cdot)$ given as

$$(4.56) \quad f_t(x|\theta) = f(x|\theta_{1t},\theta_{2t}) = \begin{cases} \frac{1}{\theta_{2t}-\theta_{1t}} & \text{if } \theta_{1t} \leq x \leq \theta_{2t} \\ 0 & \text{otherwise} \end{cases}$$

where θ_{1t},θ_{2t} denote the two unknown parameters characterizing population Π_t.

Two cases of interest arise here:

(i) θ_{2t} is the parameter being ranked; θ_{1t} is the nuisance parameter.

(ii) θ_{1t} is the parameter being ranked; θ_{2t} is the nuisance parameter.

The two cases are in a sense very similar and a solution to one would suggest a solution to the other. It may also be noted that the analysis for each of the two cases of interest is very similar to the analysis in Section 4.3 for ranking location parameters (L_t) of non-regular exponential populations, when S_t is the nuisance parameter. To avoid repetition, we omit the detailed analysis and the asymptotically optimal procedures are given in terms of the following theorems, which we state without proof.

We first define notation to be used in the statement of the theorems.

For $t = 1,2,\ldots,k$, let

$$U_t = \min_{1\le i\le n} (X_{ti}) \tag{4.57}$$

$$V_t = \max_{1\le i\le n} (X_{ti}) \tag{4.58}$$

and let the ranked values of U_t and V_t be denoted by

$$U_{[1]} < U_{[2]} < \cdots < U_{[k]} \tag{4.59}$$

and

$$V_{[1]} < V_{[2]} < \cdots < V_{[k]} \tag{4.60}$$

An asymptotically optimal procedure for selecting the best population is given by the following theorem.

Theorem 4.5 (i) An asymptotically optimal procedure for selecting the population associated with the largest θ_{2t} is to select the population associated with $V_{[k]}$.

(ii) An asymptotically optimal procedure to select the population associated with the largest θ_{1t} is to select the population associated with $U_{[k]}$.

For the two general goals considered in Chapter 3 (Goal I and Goal II), asymptotically optimal procedures for the problem at hand are as given by the following theorems:

Case (i) θ_{2t} is the ranking parameter

Theorem 4.6 (i) An asymptotically optimal procedure for Goal I (to select s populations to contain d of the t best populations) is to select the s populations associated with $V_{[k-s+1]},V_{[k-s]},\ldots,V_{[k]}$.

(ii) An asymptotically optimal procedure for Goal II (selecting t best ordered populations, $1 \leq t \leq k$) is to select the populations associated with $V_{[k]},V_{[k-1]},\ldots,V_{[k-t+1]}$ as the best, second best,..., t^{th} best population respectively.

Case (ii) θ_{1t} is the ranking parameter

Theorem 4.7 (i) An asymptotically optimal procedure for Goal I is to select s populations associated with $U_{[k]},U_{[k-1]},\ldots,U_{[k-s+1]}$ as the best set of populations.

(ii) An asymptotically optimal procedure for Goal II is to select the populations associated with $U_{[k]},U_{[k-1]},\ldots,U_{[k-t+1]}$ as the best, second best,...,t^{th} best population respectively.

Since the purpose of this chapter was only to indicate how the method used for developing asymptotically optimal procedures for regular cases could be used for the non-regular cases as well, we do not indicate the optimal procesures for the Laplace and Weibull distributions (the two

remaining non-regular density functions listed in Section 4.1). The method used in Sections 4.3 and 4.4 would be used for these cases too and the analysis is very similar to the one outlined in Section 4.3.

BIBLIOGRAPHY

[1] Bahadur, R. R. and Goodman, L. "Impartial decision rules and sufficient statistics," Annals of Mathematical Statistics, Vol. 23 (1952), pp. 553-562.

[2] Barr, D. R. and Rizvi, M. H. "Ranking and selection problems of uniform distributions," Trabajos Estadist, Vol. 17 (1966), pp. 15-31.

[3] Bartoo, J. B. and Puri, P.S. "On optimal asymptotic tests of composite hypotheses," Annals of Mathematical Statistics, Vol. 38 (1967), pp. 1845-1852.

[4] Bechhofer, R. E. "A single-sample multiple decision procedure for ranking means of normal populations with known variances," Annals of Mathematical Statistics, Vol. 25 (1954), pp. 16-39.

[5] Bechhofer, R. E., Dunnett, C. W. and Sobel, M. "A two-sample multiple decision procedure for ranking means of normal populations with a common unknown variance," Biometrika, Vol. 41 (1954), pp. 170-176.

[6] Bechhofer, R. E., Elmaghraby, S. and Morse, N. "A single sample multiple-decision procedure for selecting the multinomial event which has the highest probability," Annals of Mathematical Statistics, Vol. 30 (1959), pp. 102-119.

[7] Bechhofer, R. E., Kiefer, J. and Sobel, M. Sequential Identification and Ranking Procedures, Statistical Research Monographs, Vol. 3, The University of Chicago Press, (1968).

[8] Bechhofer, R. E. and Sobel, M. "A single-sample multiple-decision procedure for ranking variances of normal populations," Annals of Mathematical Statistics, Vol. 25 (1954), pp. 273-289.

[9] Bechhofer, R. E. and Sobel, M. "Non-Parametric multiple decision procedures for selecting that one of k populations which has the highest probability of yielding the largest observations (Preliminary Report), Annals of Mathematical Statistics, Vol. 29 (1958), p. 325.

[10] Berry, A. C. "The accuracy of the Gaussian approximation to the sum of independent variates," Trans. American Mathematical Society, Vol. 49 (1941), pp. 122-136.

[11] Buhler, W. J. and Puri, P.S. "On optimal asymptotic tests of composite hypotheses with several constraints," Zeitschriftfuer Wahrscheinlichkeitstheorie, Vol. 5 (1966), pp. 71-88.

[12] Chambers, M. L. and Jarratt, P. "Use of double sampling for selecting best populations," Biometrika, Vol. 51 (1964), pp. 49-64.

[13] Cramer, H. Mathematical Methods of Statistics, Princeton University Press, Princeton (1966).

[14] Desu, M. M. and Sobel, M. "A fixed subset-size approach to the selection problem," Biometrika, Vol. 55 (1968), pp. 401-410.

[15] Dudewicz, E. J. "The efficiency of a nonparametric selection procedure: largest location parameter case," Technical Report No. 14, Department of Operations Research, Cornell University, Ithaca, N.Y. (1966).

[16] Dunnett, C. W. "On selecting the largest of k normal population means," Journal of Royal Statistic Soc., B, Vol. 22 (1960), pp. 1-40.

[17] Eaton, M. L. "Some optimum properties of ranking procedures," Annals of Mathematical Statistics, Vol. 38 (1967), pp. 124-137.

[18] Esseen, C. G. "A moment inequality with an application to the Central Limit Theorem," Skand. Aktuarietidskr., Vol. 39 (1956), pp. 160-170.

[19] Fabian, V. "On multiple decision methods for ranking population means," Annals of Mathematical Statistics, Vol. 33 (1962), pp. 248-254.

[20] Gupta, S. S. "On a decision rule for a problem in ranking means," Mimeo. Series No. 150, Inst. of Stat., Univ. of North Carolina, Chapel Hill, North Carolina (1956).

[21] Gupta, S. S. "Probability integrals of multivariate normal and multivariate t," Annals of Mathematical Statistics, Vol. 34 (1963), pp. 792-828.

[22] Guttman, I. "Best populations and tolerance regions," Annals of Inst. of Statistical Mathematics (Tokyo), Vol. 13 (1961), pp. 9-26.

[23] Guttman, I. and Milton, R. C. "Procedures for a best population problem when the criterion of bestness involves a fixed tolerance region," Annals of Inst. of Statistical Mathematics (Tokyo), Vol. 21 (1969), pp. 149-161.

[24] Guttman, I. and Tiao, G. C. "A Bayesian approach to some best population problems," Annals of Mathematical Statistics, Vol. 35 (1964), pp. 825-835.

[25] Hall, W. J. "The most economical character of some Bechhofer and Sobel decision rules," Annals of Mathematical Statistics, Vol. 30 (1959), pp. 964-969.

[26] Hoel, D. G. and Mazumdar, M. "An extension of Paulson's selection procedure," Annals of Mathematical Statistics, Vol. 39 (1968), pp. 2067-2074.

[27] Hogg, R. V. and Craig, A. T. "Sufficient statistics in elementary distribution theory," Sankhya, The Indian Journal of Statistics, Vol. 17 (1956), pp. 209-216.

[28] Katz, M. L. "Note on the Berry-Esseen Theorem," Annals of Mathematical Statistics, Vol. 34 (1963), pp. 1107-1108.

[29] LeCam, L. "On the asymptotic theory of estimation and testing hypothesis," Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1, pp. 129-156. Univ. of California Press, Berkeley and Los Angeles (1956).

[30] Lehmann, E. L. Testing Statistical Hypotheses, Wiley, New York (1959).

[31] Lehmann, E. L. "On a theorem of Bahadur and Goodman," Annals of Mathematical Statistics, Vol. 37 (1966), pp. 1-6.

[32] Mahamunulu, D. M. "On a generalized goal in fixed-sample ranking and selection problems," Technical Report No. 72, Dept. of Statistics, Univ. of Minnesota, Minneapolis, Minn. (1966).

[33] Mahamunulu, D. M. "Some fixed-sample ranking and selection problems," Annals of Mathematical Statistics, Vol. 38 (1967), pp. 1079-1091.

[34] Milton, R. C. "Tables of the equicorrelated multivariate normal probability integrals," Technical Report No. 27, Dept. of Statistics, Univ. of Minnesota, Minneapolis, Minn. (1963).

[35] Neyman, J. "Optimal tests of composite statistical hypotheses," pp. 213-234; in the Harald Cramer Volume, ed. by Grenander, U., Almquist and Wiksell, Stockholm (1959).

[36] Paulson, E. "A sequential procedure for selecting the population with the largest mean from k normal populations," Annals of Mathematical Statistics, Vol. 35 (1964), pp. 174-180.

[37] Paulson, E. "Sequential procedures for selecting the best one of several binomial populations," Annals of Mathematical Statistics, Vol. 38 (1967), pp. 117-123.

[38] Perng, S. K. "A comparison of the asymptotic sample size of two sequential procedures," Annals of Mathematical Statistics, Vol. 40 (1969), pp. 2198-2202.

[39] Ramberg, J. S. "A multiple decision approach to the selection of the best set of predictor variables," Technical Report No. 79, Dept. of Operations Research, Cornell Univ., Ithaca, N.Y. (1969).

[40] Rao, C. R. Linear Statistical Inference and its Applications, Wiley, New York (1967).

[41] Robbins, H., Sobel, M. and Starr, N. "A sequential procedure for selecting the largest of k means," Annals of Mathematical Statistics Vol. 39 (1968), pp. 88-92.

[42] Sazanov, V. V. "On the multidimensional central limit theorem," Sankhya, The Indian Journal of Statistics, Series A, Vol. 30, (1968), pp. 181-204.

[43] Sobel, M. "Sequential procedures for selecting the best exponential population," pp. 99-110 in Proceedings of Third Berkeley Symposium on Math. Stat. and Probability, Vol. 5, Univ. of Calif. Press, Berkeley, Calif. (1956).

[44] Sobel, M. and Huyett, M. "Selecting the best one of several binomial populations," Bell System Tech. Journal, Vol. 36 (1957), pp. 537-576.

[45] Sommerville, P. N. "Some problems of optimum sampling," Biometrika, Vol. 41 (1954), pp. 420-429.

[46] Srivastava, M. S. "Some asymptotically efficient sequential procedures for ranking and slippage problems," Journal of Royal Statistical Soc., Series B, Vol. 28 (1966), pp. 370-380.

[47] Srivastava, M.S. and Oglivie, J. "The performance of some sequential procedures for a ranking problem," Annals of Mathematical Statistics, Vol. 39 (1968), pp. 1040-1047.

[48] Wald, A. Statistical Decision Functions, Wiley, New York (1950).

[49] Weiss, L. Statistical Decision Theory, McGraw Hill, N.Y. (1961).

[50] Weiss, L. "On the estimation of scale parameters," Naval Research Logistics Quarterly, Vol. 8 (1961), pp. 245-256.

[51] Weiss, L. and Wolfowitz, J. "Generalized maximum likelihood estimators," Theory of Probability and its Applications, Vol. 11 (1966), pp. 181-204.

[52] Weiss, L. and Wolfowitz, J. "Maximum Probability Estimators," Annals of Inst. of Stat. Math (Tokyo), Vol. 19 (1967), pp. 193-206.

[53] Weiss, L. and Wolfowitz, J. "Maximum Probability Estimators with a General Loss Function," to appear in Proceedings of the International Symposium on Probability and Information Theory, held at McMaster University, Hamilton, Ontario, April 4 and 5, 1968.

[54] Weiss, L. and Wolfowitz, J. "Maximum Probability Estimators and Asymptotic Sufficiency," submitted to Annals of Inst. of Stat. Math. (Tokyo).

[55] Weiss, L. and Wolfowitz, J. "Asymptotically minimax tests of composite hypotheses," Zeitschriftfuer Wahrscheinlichkeitstheorie Vol. 14 (1969), pp. 161-168.

[56] Zolotarev, V. M. "An absolute estimate of the remainder term in the central limit theorem," Theory of Probability and its Applications, Vol. 11 (1966), pp. 95-105.

UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)	2a. REPORT SECURITY CLASSIFICATION
Department of Operations Research College of Engineering, Cornell University Ithaca, New York 14850	Unclassified
	2b. GROUP

3. REPORT TITLE

Asymptotically Optimal Ranking and Selection Procedures

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Technical Report, March 1970

5. AUTHOR(S) (First name, middle initial, last name)

Bawa, Vijay S.

6. REPORT DATE	7a. TOTAL NO OF PAGES	7b. NO OF REFS
March 1970	132	56

8a. CONTRACT OR GRANT NO	9a. ORIGINATOR'S REPORT NUMBER(S)
DA-31-124-ARO-D-474 b. XXXXXXXXX Nonr-401(53) c. NSF GP-7798 d.	Technical Report No. 102 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10. DISTRIBUTION STATEMENT

Distribution of this document is unlimited

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
Sponsoring military activity U. S. Army Research Office Durham, N.C. 27706	Logistics and Mathematical Statistics Branch, Office of Naval Research Washington, D.C 20360

13. ABSTRACT

Single-stage asymptotically optimal (minimax) procedures are developed for ranking populations in the presence of nuisance parameters, when the populations are ranked according to a parameter of the distribution and the so-called indifference-zone approach to ranking and selection problems is employed. This is accomplished by adapting methods used by Weiss and Wolfowitz (for 2-decision tests of composite hypotheses problems in the presence of nuisance parameters) to multiple-decision ranking and selection problems in the presence of nuisance parameters.

For the problem of selecting the "best" population (and for certain other ranking and selection goals), asymptotically optimal procedures are developed for situations in which the joint density function of the observations satisfies certain mild regularity conditions. In addition, the applicability of the basic method is demonstrated by developing asymptotically optimal procedures for ranking non-regular exponential and uniform distributions. The asymptotically optimal character of certain so-called natural selection procedures which already have been proposed in the literature is proved. Single-stage asymptotically optimal procedures are derived for certain problems for which heretofore no single-stage procedures had been proposed.

DD FORM 1473

Unclassified
Security Classification

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
asymptotically optimal procedures						
decision theory						
mathematical statistics						
minimax procedures						
nuisance parameters						
ranking procedures						
selection procedures						