NONPARAMETRIC SELECTION, RANKING AND TESTING

by

Sayaji N. Hande
Purdue University

Technical Report # 90–06

Department of Statistics
Purdue University

January 1990

△

Nonparametric Selection, Ranking and Testing*

Sayaji N. Hande

Department of Statistics, Purdue University

Abstract

In this paper we consider the problem of ranking (partitioning) k populations according to the parameter which is defined as functionals of the distribution functions on the underlying populations. We obtain minimax rules for general loss functions, Bayes rules for some specific loss functions and propose approximate non-randomized minimax rules. We also derive restricted minimax rules for selecting a subset of populations which are better than a control. Some nonparametric "optimal" tests are derived for different hypotheses written in terms of the parameter as a functional of the underlying distribution function.

Key Words: Selection and ranking, nonparametric, comparison with a control, testing, minimax decision rules.

AMS 1985 subject classification: 62G99, 62C20.

1 Introduction

In practice, the experimenter is often faced with the problem of comparing k populations, for example, comparing k different treatments in clinical trials, or comparing k different varieties

*This research was supported in part by the Office of Naval Research Contract N00014-88-K-017 and NSF Grants DMS-86066964, DMS-8702620 at Purdue University.

of grain in an agricultural experiment. The classical tests of homogeneity never answer the question "what next?" if the hypothesis is rejected. Mosteller (1948) and Paulson (1949), and Bahadur (1950) were among the first research workers to recognize the inadequacy of such tests for homogeneity and to reformulate the problem as a multiple decision problem concerned with the ranking and selection of k populations.

One approach pioneered by Bechhofer (1954) has been to allow the experimenter to select one population which is guaranteed to be of interest to him with a fixed probability P^*, whenever the unknown parameters lie outside some subset of the parameter space. This has been termed as the indifference zone approach. In contrast to the indifference zone approach, Gupta (1956) proposed a formulation in which the experimenter obtains a subset of k populations for which there is a fixed minimum probability P^*, over the entire parameter space, that the population of interest is selected. For an extensive review of the subset selection methodology see Gupta and Panchapakesen (1979) and Gupta and Panchapakesen (1986).

In this paper we consider a decision theoretic formulation of the ranking problem in the nonparametric setup. Let the distribution function F on R^p be characterized by the functional $\theta(F) = \int g dF$, where g is a known real-valued bounded function on R^p and $\theta = \theta(F)$ is the parameter of interest.

Consider the following examples.

(1) SELECTING THE BEST:

Company A produces a product whose observable quality is represented by a random variable Y. Company B has discovered k new products of the same "type" and wants to select one of those k products which will beat the product of company A in the market. Let us suppose $X(i)$ represents the quality of the i th product of company B for $i = 1, \ldots, k$. A customer will select the product of the company A instead of a specified ith product of company B if Y is grater than $X(i)$. Hence in this problem the parameter of interest is $\theta(i) = Pr(X(i) > Y)$ and company B wants to select the product for which $\theta(i)$ is largest. Here g is the distribution function of Y.

The function $Pr(X < Y)$ is of considerable importance in many practical situations, such as clinical trials, genetics, and reliability. For the estimation of the parameter $Pr(X < Y)$ and for related references see Brownie (1988), Simnoff, Hochber and Reiser (1986). In Section 3 "optimal" non-parametric tests for the various hypothesis for the parameter $\theta = Pr(X < Y)$ are derived.

(2) REGRESSION:

Let $X = (X_1, X_2, \ldots, X_p)$ be a p dimensional random vector which has the distribution function F. We want to test whether x_1 is well approximated by $h(x_2, x_3, \ldots x_p)$, where h is a known real-valued function on $R^{(p-1)}$. Define $\theta(F) = \int d(X_1 - h(X_2, X_3, \ldots, X_p)dF$, where d is an appropriate non-negative function on R. In this situation $g(x) = d(x_1 - h(x_2, x_3, \ldots, x_p))$. We may want to test

$$H_o : \theta(F) < \theta_0 \text{ vs } H_1 : \theta(F) > \theta_0,$$

where θ_0 is a known constant.

(3) SELECTING A SUBSET OF THE POPULATIONS CONTAINING A POPULATION BETTER THAN THE CONTROL:

Let $x(\alpha)$ be the α th quantile of the control. There are k populations, $\Pi_1, \Pi_2, \ldots, \Pi_k$. The population Π_i is associated with the distribution function F_i on R , for $i = 1, 2, \ldots, k$.

We say the population Π_i is

$$\text{"good" if } \int_{-\infty}^{x_\alpha} dF_i > \alpha$$

and

$$\text{"bad" if } \int_{-\infty}^{x_\alpha} dF_i > \alpha - \delta.$$

In this problem $g(x) = I_{(-\infty,x_\alpha)}(x)$.

It is important to consider a non-parametric model, since often in practice, especially for the new treatments, there is not much information which could lead us to assume some parametric model.

In the next section, we will derive a minimax procedure for the selection and ranking problem, we also obtain a restricted minimax procedure for the problem when the populations are compared with a control.

Our procedures, however, are randomized. We feel that the randomization is unavoidable in the present situation, since as is known, certain properties of the risk function can be improved only by using randomization. In some examples we will also prove that these procedures are unique and admissible. In Section 4 we will derive some "optimal" non-parametric tests.

Most of the existing results on non-parametric models, in general are asymptotic. The finite sample results, which are presented here may be of use to check the optimality of the existing procedures (tests) or for proving optimality of new tests.

It should be pointed out that results presented here do not apply to the problem of selecting the population with the largest α th quantile (or largest location parameter). Also these results do not apply to the problem of selecting a subset of the population which contains the population with largest α th quantile (or location parameter). Considerable amount of work has been done on those kinds of problems. See Barlow and Gupta (1969), Gupta and McDonald (1970), Gupta and Huang (1974), Rizvi and Sobel (1967), Sobel (1967). An extensive review of non-parametric selection and ranking procedures is in Desu and Bristol (1986).

2 Selection And Ranking

There are k populations $\Pi_1, \Pi_2 \ldots, \Pi_k$. The population Π_i is associated with the cumulative distribution function $F_i(.)$ on R^p, for $i = 1, 2, \ldots, k$. The population Π_i is characterized by the real-valued function,

$$\theta(F_i) = \int_{R^p} g(x) dF_i(x) \ ;$$

where g is a known, real-valued bounded function on R^p.

Define $\theta_i \equiv \theta(F_i)$ for $i = 1, 2, \ldots, k$ and

$$F = (F_1, F_2, \ldots, F_k), \ \ \theta = \theta(F) = (\theta_1, \theta_2, \ldots, \theta_k).$$

Let

$$\mathcal{F} = \{(F_1, F_2, \ldots, F_k) : F_i \text{ is distribution on } R^p \ \}$$

and

$$\Theta = \{(\theta(F_1), \theta(F_2), \ldots, \theta(F_k) : F_i \text{ is distribution on } R^p \ \}.$$

Let $X_{i1}, X_{i2}, \ldots, X_{in}$ be the n independent random vectors from population Π_i.

Problem (I) General Ranking Problem:

On the basis of a set of observations we wish to partition the set of the co-ordinate values of the k dimensional parameter vector $\theta = (\theta_1, \theta_2, \ldots, \theta_k)'$ in to r disjoint subsets, say $S_1, S_2, \ldots, S_r$, such that S_1 contains the t_1 largest components of θ, S_2 contains next t_2 largest components of θ and $\ldots$, S_r contains the t_r smallest components of θ. The size of each subset is fixed in advance and $\sum_{i=1}^{r} t_i = k$.

Let the the action space $\mathcal{A}$, be the set of all possible partition of the set $\{1, 2, \ldots, k\}$ in to r subsets $S_1, S_2, \ldots, S_r$ of size $t_1, t_2, \ldots, t_r$, respectively. For $a \in \mathcal{A}$ let $a = (S_{a,1}, S_{a,2}, \ldots, S_{a,r})$. A decision rule $\delta = \delta(.)$,

$$\delta(.) \ \ = \ \ \{\delta(., a) : a \in \mathcal{A}\} \ ; \qquad (1)$$

is a measurable function on R^{npk}, such that

$$0 \leq \delta(.,a) \leq 1$$

and

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

If $X = x$ is observed then the decision rule δ takes the action a with the probability $\delta(x,a)$.

Let $\mathcal{D}$ be the class of all decision rules. We will consider the loss functions which are "invariant", "non-negative" and "monotone." This type of loss structure is considered by several authors, for example, see Eaton (1967), Gupta and Mieske (1984). Let $L(.,.)$, a real valued measurable function, be a loss function on $\Theta \times \mathcal{A}$. Hence if one takes action a and if the true parameter is θ then the loss is $L(\theta, a)$. Formally we write the conditions on the loss function as:

[1] $L(\theta, a) \;\; \geq \;\; 0$

[2] For every parmutation π on

$$\{1, 2, \ldots, k\} \;\; L(\pi(\theta), \pi(a)) = L(\theta, a) \;\; \forall a \in \mathcal{A}$$

[3] Let $\theta_i \geq \theta_j$ and $a = (S_1, S_2, \ldots, S_r)$, $a' = (S'_1, S'_2, \ldots, S'_r)$ such that, for r_1 and r_2 such that $1 \leq r_1 \leq r_2 \leq r \; \forall \, t \neq r_1$ and $i \neq r_2$ $S_t = S'_t$ and $S'_{r_1} = (S_{r_1} - \{j\}) \cup \{i\})$, $S'_{r_2} = (S_{r_2} - \{i\}) \cup \{j\})$, then,

$$L(\theta, a) \leq L(\theta, a');$$

[4] For every $a \in \mathcal{A}$ $L(\theta, a)$ is a continuous function of θ .

The risk function of the decision rule δ is given by

$$\begin{aligned} R_1(F,\delta) &= E_F L(\theta(F),\delta) \\ &= \sum_{a\in\mathcal{A}} L(\theta(F),\delta) E_F \delta(X,a). \end{aligned}$$

A minimax rule will be derived for the problem described above.

Problem (II) Selecting "good" populations:

We describe this problem as in Lehmann (1961). Let there be a fixed value θ_0 and let Δ be a fixed positive real number. The population Π_i is said to be good (positive) if $\theta_i(F) \geq \theta_0+\Delta$ and bad if $\theta_i(F) \leq \theta_0$. We wish to select a subset of the populations containing good populations, provided there exists at least one good population.

For this problem we will consider two loss functions, one will guard against selecting too many bad populations and the other one will make sure that good populations are being selected. As in Lehmann (1961) the following criteria will be used for measuring how well the procedure carries out the task,

(S1) The expected number of good populations.

(S2) The expected proportion of good populations.

(S3) The probability of selecting at least one good population, provided there exists one.

(S4) The probability of including the "best" population provided it is "good".

The following criteria are considered for measuring the performance of the procedure.

(R1) The number of bad populations in the selected subset.

(R2) The proportion of bad population in the selected subset.

For a subset selection procedure δ, $S(\theta(F),\delta)$ is given by (S1), (S2), (S3) or (S4) and $R(\theta(F),\delta)$ is given by (R1) or (R2). Let

$$\mathcal{F}' = \left\{ F: \ \theta(F_i) \geq \theta_0 + \Delta \quad \text{for some} \quad i, \ 1 \leq i \leq k \right\}$$

and $\mathcal{D}_s$ be the class of all procedures for selecting a subset of good populations. We will construct a restricted minimax procedure for this problem, that is, we will construct a procedure $\delta_{**} \in \mathcal{D}_s$ which minimizes $\sup R(\theta(F), \delta)$ among all $\delta \in \mathcal{D}_s$ for which

$$\inf_{F \in \mathcal{F}'} S(\theta(F), \delta) \geq p \,,$$

where p is a given fixed number.

To prove the main results, we need results from Eaton (1967) and from Lehmann (1961). For sake of completeness, we state them.

Theorem 2.1 *: Let the random variable X_i have density $p_{\theta_i}(x); i = 1, 2, \ldots k$, $p_\theta(x)$ has monotone likelihood ratio in x, and let $R_1(\theta, \delta)$ be as defined in Problem I.*

Let

$$\begin{aligned} B_a &= \{x : \, x_{i_1} \geq x_{i_2} \geq \ldots \geq x_{i_s} \, \forall \, i_j \in S_{a,j} \, \forall j\} \\ H(x) &= \{a : a \in \mathcal{A} \, ; x \in B_a\} \,. \end{aligned}$$

And let

$$n(x) = \text{ number of elements in} \, \{a : a \in H(x)\} \,.$$

Define

$$\delta'(x, a) = \frac{1}{n(x)} \quad \text{if } a \in H(x) \tag{2}$$

$$= 0 \;\; \text{otherwise.} \tag{3}$$

Then the rule δ' minimizes $\sup_\theta R_1(\theta, \delta)$ among all $\delta \in \mathcal{D}$.

Theorem 2.2 *: Let $X_1, X_2, X_3, \ldots, X_k$ be the independent random variables with probability densities $p_{\theta_1}(x), p_{\theta_2}(x), \ldots, p_{\theta_k}(x_k)$ respectively. Let $p_\theta(x)$ has monotone likelihood ratio in x. Define $R(\theta, \delta)$ and $S(\theta, \delta)$ as in Problem II, where δ is a subset selection procedure. Let*

$\delta_s = (\delta_{1s}, \ldots, \delta_{ks})$, *where*

$$\begin{aligned}\delta_{is} &= 1 \quad \textit{if } X_i > c \\ &= \lambda_0 \quad \textit{if } X = c \\ &= 0 \quad \textit{otherwise.}\end{aligned}$$

Where λ_0 *and* c *are determined by the equation,* $E_{\theta_0+\Delta}\delta_{1s} = p$, *then the rule* δ_s *minimizes* $\sup_\theta R(\theta, \delta)$ *among all rules* $\delta \in \mathcal{D}_s$ *such that*

$$\inf_{\theta \in \Omega'} S(\theta, \delta) \geq p,$$

where,

$$\Omega' = \left\{ (\theta_1, \theta_2, \ldots, \theta_k) : \quad \theta_i \geq \theta_0 + \Delta \quad \textit{for some } \; i \right\}.$$

The supremum of $R(\theta, \delta_s)$ *is attained at* $\theta = (\theta_0, \theta_0, \ldots, \theta_0)$ *and the infrimum of* $S(\theta, \delta_s)$ *is attained at* $\theta = (\theta_0 + \delta, \theta_0, \ldots, \theta_0)$.

Let $X_1, X_2, \ldots X_k$, be independent binomial random variables with parameters $(n, \theta_1), (n, \theta_2), \ldots, (n, \theta_k)$ respectively. For $\delta_s(.) = \delta_s(x)$,

$$\sup_{\theta \in \Omega} R(\theta, \delta) = n\left[P_{\theta_0}(X > c) + \lambda \; P_{\theta_0}(X = c)\right] \tag{4}$$

$$= h(\theta_0, \Delta, p) \text{ say.} \tag{5}$$

Here c is a non-negative real number and λ in $(0, 1)$ are chosen such that,

$$P_{\theta_0+\Delta}(X_1 > c) + \lambda \; P_{\theta_0+\Delta}(X_1 = c) = p. \tag{6}$$

If Y_i is a sequence of binomial random variables with parameters (n, p_i), and as $i \longrightarrow \infty$, p_i converges to p_0, then the sequence of Y_i converges weakly to Y_0, we note that, for a fixed p, $h(\theta_0, \Delta, p)$ is a continuous function of θ_0 and Δ.

Making transformation on g, if necessary

$$g(x) \longrightarrow \frac{g(x) - \inf g(x)}{\sup g(x) - \inf g(x)}$$

and observing that

$$\int [ag(x) + b] \, dF(x) = a \ \ \theta(F) + b \, ,$$

without any loss of generality we assume that

$0 \le g(x) \le 1 \ \forall \ \ x \in R^p$ and $\sup g(x) = 1, \ \ \inf g(x) = 0.$

Lemma 2.1 *: If $\delta \in \mathcal{D}_s$, a subset selection procedure for Problem II and*

$$\inf_{F \in \mathcal{F}'} S(\theta(F), \delta) \ge p,$$

then

$$\sup_{F \in \mathcal{F}} R(\theta, \delta) \ge h(\theta_0, \Delta_0, p),$$

where $h(\theta_0, \Delta_0, p)$ is as defined in (4).

Proof:

We know that $\inf g(x) = 0$ and $\sup g(x) = 1$. Let $\delta \in \mathcal{D}_s$. Fix $\epsilon > 0$, and get a and b such that $g(a) = \epsilon_1$, $g(b) = 1 - \epsilon_2$ and $0 < \epsilon_1 + \epsilon_2 < \epsilon$.

Let P_i be the probability measure induced by a distribution function F_i. // Define

$$\mathcal{F}_0 = \mathcal{F}_{0(\epsilon_1, \epsilon_2)} = \left\{ F : \begin{array}{ll} P_i(\{b\}) = p_i; & P_i(\{a\}) = 1 - p_i \\ 0 \le p_i \le 1; & \text{for } \ i = 1, 2 \ldots, k \end{array} \right\}. \tag{7}$$

If $F = (F_1, F_2, \ldots, F_k) \in \mathcal{F}_0$, then

$$\begin{aligned} \theta_i &= \theta(F_i) \\ &= \epsilon_1(1 - p_i) + (1 - \epsilon_2)p_i \\ &= \epsilon_1 + (1 - \epsilon_1 - \epsilon_2)p_i. \end{aligned}$$

Therefore, $\theta(F_i) \leq \theta_0$ if and only if

$$p_i \leq \frac{\theta_0 - \epsilon_1}{1 - \epsilon_1 - \epsilon_2},$$

and $\theta(F_i) \geq \theta_0 + \Delta$ if and only if

$$p_i \geq \frac{\theta_0 - \epsilon_1}{1 - \epsilon_1 - \epsilon_2} + \frac{\Delta}{1 - \epsilon_1 - \epsilon_2}.$$

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

$$T_i = \# \{X_{ij} : \; X_{ij} = b \;\; j = 1, 2, \ldots, n.\}.$$

Note that for a class of distribution functions $\mathcal{F}_0$, the statistics $T = (T_1, T_2, \ldots, T_k)$ is a complete sufficient statistics. We also note that $T_1, T_2, \ldots, T_k$ are independent and they have binomial distributions with parameters $(n, p_1), (n, p_2), \ldots, (n, p_k)$ respectively. Since $\mathcal{F}_0 \cap \mathcal{F}' \subset \mathcal{F}'$, and

$$\inf_{F \in \mathcal{F}'} S(\theta(F), \delta) \geq p \;\; ,$$

we have,

$$\inf_{\mathcal{F}_0 \cap \mathcal{F}'} S(\theta(F), \delta) \geq p.$$

Also the binomial family possess the monotone likelihood ratio property. By Theorem 1.2.2 we have,

$$\sup_{F \in \mathcal{F}_0} R(\theta(F), \delta) \geq h(\frac{\theta_0 - \epsilon_1}{1 - \epsilon_1 - \epsilon_2}, \frac{\Delta}{1 - \epsilon_1 - \epsilon_2}, p). \tag{8}$$

And since $\mathcal{F} \supset \mathcal{F}_0$, we have,

$$\begin{aligned} \sup_{F \in \mathcal{F}} R(\theta(F), \delta) \;\; &\geq \;\; \sup_{F \in \mathcal{F}_0} R(\theta(F), \delta) \qquad (9) \\ &\geq \;\; h(\frac{\theta_0 - \epsilon_1}{1 - \epsilon_1 - \epsilon_2}, \frac{\Delta}{1 - \epsilon_1 - \epsilon_2}, p). \qquad (10) \end{aligned}$$

Since $\epsilon > 0$ is arbitrary, letting $\epsilon \longrightarrow 0$, we have

$$\frac{\theta_0 - \epsilon_1}{1 - \epsilon_1 - \epsilon_2} \longrightarrow \theta_0$$

and

$$\frac{\Delta}{1 - \epsilon_1 - \epsilon_2} \longrightarrow \Delta.$$

As we noticed before h is a continuous function, letting $\epsilon \longrightarrow 0$, we get

$$\sup_{F \in \mathcal{F}} R(\theta(F), \delta) \geq h(\theta_0, \Delta, p). \tag{11}$$

This completes the proof of the lemma.

Suppose that X_i has a binomial distribution with parameter (n, θ_i), for $i = 1, 2, \ldots, k$. Let

$$\sup_{\theta} R_1(\theta, \delta') = R, \tag{12}$$

where δ' is as defined in (2).

Lemma 2.2 *If $\delta \in \mathcal{D}$ is a decision rule for the Problem I then*

$$\sup_{F \in \mathcal{F}} R_1(\theta(F), \delta) \geq R$$

where R is as defined in (12).

Proof:-

Fix $\epsilon \geq 0$ and get ϵ_1, ϵ_2 and a, b, $\mathcal{F}_0 = \mathcal{F}_{0,(\epsilon_1,\epsilon_2)}$ and $T = (T_1, T_2, \ldots, T_k)$ as in Lemma 1. Observe that $T_1, T_2, \ldots, T_k$ are independent binomial with parameters $(n, p_1), (n, p_2), \ldots, (n, p_k)$ respectively. Here $\theta_i = \theta_{(F_i)} = \epsilon_1 + (1 - \epsilon_1 - \epsilon_2)p_i$ for each i, $0 \leq p_i \leq 1$ if and only if $\epsilon_1 \leq \theta_i \leq 1 - \epsilon_2$. By Theorem 1 if δ is any decision rule then,

$$\sup_{F \in \mathcal{F}_0} R_1(\theta(F), \delta) \geq \sup_{F \in \mathcal{F}_0} R_1(\theta(F), \delta_0)$$

where $\delta_0(x,a) = \delta'(T,a)$ and δ' as defined by (2).

Since

$$\sup_{F\in\mathcal{F}} R_1(\theta(F),\delta) \geq \sup_{F\in\mathcal{F}_0} R_1(\theta(F),\delta),$$

we have

$$\sup_{F\in\mathcal{F}} R_1(\theta(F),\delta) \geq \sup_{F\in\mathcal{F}_{0,(\epsilon_1,\epsilon_2)}} R_1(F,\delta_0) \;=\; R_{(\epsilon_1,\epsilon_2)} \quad \text{say.}$$

For $F \in \mathcal{F}_0$,

$$R_1(\theta(F),\delta_0) = \sum_{a\in\mathcal{A}} E_F \delta'(T,a) \;\; L(\theta(F),a).$$

The expectation of $\delta'(T,a)$ depends on F only through $\theta = \theta(F)$ and $\theta = (\epsilon_1 + (1 - \epsilon_1 - \epsilon_2)p_1, \epsilon_1 + (1-\epsilon_1-\epsilon_2)p_2, \ldots, \epsilon_1 + (1-\epsilon_1-\epsilon_2)p_k) \longrightarrow (p_1,p_2,\ldots,p_k)$ as $\epsilon_1+\epsilon_2 \longrightarrow 0$. We know that for every $a \in \mathcal{A}$, $L(.,a)$ is continuous over $[0,1]^k$, hence it is uniformly continuous over $[0,1]^k$. We have $R_{(\epsilon_1,\epsilon_2)} \longrightarrow R$ as $\epsilon_1 + \epsilon_2 \longrightarrow 0$. Hence

$$\sup_{F\in\mathcal{F}} R_1(F,\delta) \geq R, \tag{13}$$

and this completes the proof of the lemma.

Let a random variavle X have a binomial distribution with the parameters $(n, \theta_0 + \Delta)$. Let $c = c(p) \geq 0$ and $\lambda = \lambda(p) \in [0,1)$ such that,

$$P(X > c) + \lambda \;\; P(X = c) = p.$$

Let $Z_1, Z_2, \ldots Z_k$ be the independent Bernoulli random variables with parameters $p_1, p_2, \ldots p_k$. Define,

$$\psi(p_1,p_2,\ldots,p_k) = \qquad P(Z_1 + Z_2 \ldots Z_k > c(p)) \tag{14}$$

$$+ \;\; \lambda(p) \;\; P(Z_1 + Z_2 \ldots + Z_k = c(p)). \tag{15}$$

Let $Z = (Z_{11}, Z_{12}, \ldots, Z_{1n}, \ldots, Z_{k1}, Z_{k2}, \ldots Z_{kn})$ be a random vector. Components of Z for $X = (X_{11}, X_{12}, \ldots X_{1n}, \ldots X_{k1}, X_{k2}, \ldots, X_{kn})$ are independent Bernoulli with

$$P(Z_{ij} = 1|x) = g(X_{ij}) \ \forall \ i, \ \ j.$$

Let $T_i = \sum_{j=1}^{n} Z_{ij}$; $T = (T_1, T_2, \ldots T_k)$. Let δ' be as in Theorem2.1 . Define

$$\delta''(x, a) = E(\delta'(T, a)|X = x). \tag{16}$$

Theorem 2.3 *:*

(i) The decision rule $\delta''(.,.) = \delta''(x, a)$ is a minimax rule for a class of loss functions defined for the Problem I .

(ii) Let $\delta_i(x) = \psi(g(x_{i1}), g(x_{i2}), \ldots, g(x_{in}))$; for $i = 1, 2 \ldots, k$. Then the decision rule $\delta_{**} = (\delta_1, \delta_2, \ldots, \delta_k)$ is a restricted minimax procedure for the Problem II .

Proof:

(i) .

$$\begin{aligned} &R_1(\theta(F), \delta'') \\ &= E_F \sum_{a \in \mathcal{A}} \delta''(X, a) L(\theta(F), a) \\ &= \sum_{a \in \mathcal{A}} [E_F \delta''(X, a)] L(\theta(F), a) \\ &= \sum_{a \in \mathcal{A}} E[E_F \delta'(T, a|X = x)] L(\theta(F), a) \\ &= \sum_{a \in \mathcal{A}} E_F \delta'(T, a) L(\theta(F), a). \end{aligned}$$

We notice that $T_1, T_2, \ldots, T_k$ are independent random variables. The marginal distribution of T_i is binomial with the parameters $(n, \theta(F_i))$ for $i = 1, 2, \ldots, k$.

Hence

$$\begin{aligned} \sup_{f \in \mathcal{F}} R_1(F, \delta'') &= \sup_{0 \leq \theta_i \leq 1} [\sum_{a \in \mathcal{A}} E_F \delta'(T, a) L(\theta, a)] \\ &= R. \end{aligned}$$

By *Lemma* 1.2.2 the result follows.

(*ii*) .

Observe that $\delta_1(X), \delta_2(X), \ldots \delta_k(X)$ are independent.

$$\begin{aligned} E_F\delta_i(X) &= E_F\psi(g(x_{i1}, g(x_{i2}), \ldots, g(x_{in})) \\ &= P(T_i > c(p)) + \lambda(p)\ P(T_i = c(p)). \end{aligned}$$

Here $T_1, T_2, \ldots, T_k$ are independent binomial with parameters $(n, \theta(F_1)), (n, \theta(F_2)), \ldots,$ $(n, \theta(F_k))$ respectively.

For $F \in \mathcal{F}$,

$$S(\theta(F), \delta_{**}(.)) = S(\theta(F), \delta_s(.)),$$

where

$$\delta_s(.) = \delta_s(T) \quad \text{and} \quad T = (T_1, T_2, \ldots, T_K).$$

The decision rule δ_s is as before and

$$R(\theta(F), \delta_{**}) = R(\theta(F), \delta_s)$$

By the choice of $c = c(p)$ and $\lambda = \lambda(p)$ and by (11) we have

$$\begin{aligned} \sup_{F \in \mathcal{F}} R(\theta(F), \delta_{**}) &= \sup_{\theta} R(\theta, \delta_s) \\ &= h(\theta_0, \Delta, p) \end{aligned}$$

and

$$\begin{aligned} \inf_{F \in \mathcal{F}'} S(\theta, \delta_s) &= \inf_{\theta \in \Omega'} S(\theta, \delta_s) \\ &= p, \end{aligned}$$

and this completes the proof.

Remark 2.1

Notice that the minimax procedures we have established are randomized. To avoid randomization, for example in Problem I, we may take action $a \in \mathcal{A}$ if $\delta'(x, a) \geq \delta'(x, a')$ for all $a' \in \mathcal{A}$. We feel that this procedure may be approximate minimax and will be more useful in practical situations. However it is difficult to establish analytic properties of this approximated procedure.

3 Discussion and Examples

In this section we consider a few examples (problems) and consider the Bayes rules with respect to particular priors and see how different they are from the minimax rules.

Example 3.1

Nonparametric Bayes Procedures

Let θ, as defined before, be the parameter of interest. Let us suppose that we are interested in selecting population Π_i for which θ_i is largest. Here

$$\theta_i = \theta(P_i) = \int_{\mathcal{X}} g(t)\, dP_i(t).$$

and P_i is a probability measure corresponding to population Π_i. Let for each $i = 1, 2, \ldots, k$, the probability measures $P_1, P_2, \ldots, P_k$ are independently, identically distributed with common Dirichlet distribution $D(\alpha)$. The probability measure P on $\mathcal{X}$ is said to have Dirichlet distribution with parameter α if for any k, $A_1, A_2, \ldots, A_k$ is some measurable partition of $\mathcal{X}$ then $(P(A_1), P(A_2), \ldots, P(A_k))$ has Dirichlet distribution with parameters $(\alpha(A_1), \alpha(A_2), \ldots, \alpha(A_k))$. See Ferguson (1973) and Ferguson (1974) for more discussion and for relative information.

Let the loss function be of the form

$$L(\theta, i) = \max_j \theta_j - \theta_i.$$

That is the loss for selecting $i-th$ population is $max_j\theta_j - \theta_i$. Then according to the Bayes rule we select the population Π_i for which $E(\theta(P_i)|X = x)$ is largest. Following Ferguson (1973) we know that the posterior distribution of P_i follows a Dirichlet distribution with parameter $\alpha + nP_{n,i}$. Hence the conditional expectation of θ_i is

$$\int_{\mathcal{X}} g \; d(\alpha + nP_{i,n}).$$

Hence $E(\theta(P_i)|X = x)$ is largest if and only if $\hat{\theta}_i = n^{-1}\sum_{j=i}^{n} g(x_{ij})$ is largest. The procedure we have established, will also select the ith population with high probability if $\hat{\theta}_i$ is large.

Example 3.2

Let $X_{i1}, X_{i2}, \ldots X_{in}$ be the observable random vectors from the population Π_i for $i = 1, 2, \ldots, k$. Let P_i be the probability measure generated by the random variable X_{i1} . Let $\theta_i = P_i(A)$ be the parameter of interest. Let $X_i = (X_{i1}, X_{i2}, \ldots X_{in})$ and

$$T_i(X_i) = \sum_{j=1}^{n} I_A(X_{ij}).$$

Then according to the theorem above the procedure which ranks population Π_i according to the rank of T_i in $T_1, T_2, \ldots, T_k$ is the minimax procedure for any permutation invariant loss function. Here $g(x) = I_A(x)$ is an indicator function of set A.

We will show that this procedure is also Bayes procedure when P_i , $i = 1, 2, \ldots, k$ are independently identically distributed with the Dirichlet prior. To see this, notice that the posterior distribution of $P_i(A)$ has beta distribution with parameter $p = \alpha(A) + nP_{i,n}(A)$ and $q = c - p$, for $i = 1, 2, \ldots, k$. Where $c = \alpha(\mathcal{X}) + n$, is a fixed constant. We also notice that if a random variable Y has beta distribution with parameters p and $c - p$ (c is a fixed constant) then y has a monotone likelihood ratio in p. From these facts it is straightforward to prove that if L is any invariant loss function then the rule of ranking k population according to ranks of T_i $i = 1, 2, \ldots, k$, is a Bayes rule, and hence it is admissible.

The same argument holds for the (restricted) subset selection procedures.

4 Testing

Let $X_1, X_2, \ldots, X_n$ be observable independent random vectors with a common distribution function F on R^P.

Let $\theta(F) = \int g dF$ be the parameter of interest, where g is a real-valued bounded function on R^p such that $\sup g(x) = 1$ and $\inf g(x) = 0$. We will construct "optimal" tests for testing the hypothesis,

$$H_0 : \theta(F) \in \Theta_0 \quad \text{Vs} \quad H_1 : \theta(F) \in \Theta_1 .$$

Here $\Theta_1 = \Theta_0^c$ and Θ_0 is of the form $\{\theta : \theta \leq \Theta_0\}$, $\{\theta : \theta \geq \theta_0\}$ or $\{\theta : \theta = \theta_0\}$.

Comparison between the tests of level α is made on the basis of the "power" of the test. The power of a test ϕ_0 at F for $\theta(F) \in \Theta_1$, $Pr(\phi_0$ Rejects $H_0|F)$ is a function of F and not $\theta(F)$ alone. We will take a conservative view to choose the test. We will select the test of level α , which maximizes the minimum power. We need the following definitions.

Definition 4.1 *The function* $\beta(\theta) = \beta_\phi(\theta)$ *is called a minimum power function of the test* ϕ *if*

$$\beta_\phi(\theta) = \inf_{F:\, \theta(F)=\theta} Pr(\phi \quad Rejects \quad H_0|F)$$

Definition 4.2 *The test* ϕ *of the level* α *is said to be the least uniformly most powerful test (LUMP) if for any test* ϕ' *of level level* α,

$$\beta_\phi(\theta) \geq \beta_{\phi'}(\theta) \ \forall \ \theta \in \Theta_1.$$

Definition 4.3 *The test* ϕ *of level* α *is said to be the least uniformly most powerful unbiased test (LUMPU) of level* α *if*

$$\beta_\phi(\theta) \geq \alpha \ \forall \ \theta \in \ \Theta$$

and if ϕ' is any other test of level α with

$$\beta_{\phi'} \geq \alpha \ \forall \, \theta \ \in \Theta_1,$$

then

$$\beta_\phi(\theta) \ \geq \beta_{\phi'}(\theta) \quad \forall \ \theta \in \ \Theta_1.$$

Let

$$h(p_1, p_2, \ldots, p_n) = Pr(Z_1 + Z_2 + \ldots + Z_n > c) + \lambda \ P(Z_1 + Z_2 + \ldots + Z_n = c),$$

where $Z_1, Z_2, \ldots, Z_n$ are independent Bernolli with parameters $p_1, p_2, \ldots, p_n$ respectively. $\lambda = \lambda(\alpha)$ and $c = c(\alpha)$ are chosen such that, if Z is a binomial random variable with parameters (n, θ_0), then

$$Pr(Z > c(\alpha) + \lambda(\alpha)P(X = c(\alpha)) = \alpha.$$

Theorem 4.1 *For testing*

$$H_0 : \theta(F) \leq \theta_0 \quad Vs \ \ H_1 : \theta(F) > \theta_0$$

the test $\phi(x) = h(g(x_1), g(x_2), \ldots, g(x_n))$ is a least uniformly most powerful test.

Proof:

Let us fix an arbitrarily small $\epsilon > 0$, get a , $b \in R^p$ such that $g(a) = \epsilon_1$, $g(b) = 1 - \epsilon_2$ and $\epsilon_1 + \epsilon_2 < \epsilon$ Let

$$\mathcal{F}_0 = \left\{ F : \begin{array}{c} F(x) = 0 \ \forall \ x < a \\ = 1 - p \ \forall \ a \in \ [a, b) \\ = 1 \ \forall \ x \geq b \end{array} \right\}.$$

For the class $\mathcal{F}_0$ $,T = \#\{X_i : \quad X_i = b\}$ is a sufficient statistics and have a binomial distribution with the parameters (n, p), where

$$\begin{aligned}\theta(F) &= \epsilon_1(1-p) + (1-\epsilon_2)p \\ &= \epsilon_1 + (1-\epsilon_1-\epsilon_2)p\end{aligned}$$

and $\theta > \theta_0$ if and only if

$$p > \frac{\theta_0 - \epsilon_1}{1-\epsilon_1-\epsilon_2} = p_{(\epsilon_1,\epsilon_2)} \text{ say.}$$

For the class $\mathcal{F}_0$ the UMP test is $\phi_1(T)$, where

$$\begin{aligned}\phi_1(T) &= 1 \quad \text{if } \theta > c(\epsilon_1,\epsilon_2) \\ &= \lambda(\epsilon_1,\epsilon_2) \quad \text{if } T = c(\epsilon_1,\epsilon_2) \\ &= 0 \quad \text{if } T < T = c(\epsilon_1,\epsilon_2)\ .\end{aligned}$$

The constant $c = c(\epsilon_1,\epsilon_2)$ and $\lambda = \lambda(\epsilon_1,\epsilon_2)$ are chosen such that, $Pr(X > c) + \lambda\ P(X = c) = \alpha$, where X is a binomial random variable with parameters $(n, p_{(\epsilon_1,\epsilon_2)})$. Power of the test $\phi_1(T)$ at $\theta = \theta_1 > \theta_0$ is $Pr(T > c(\epsilon_1,\epsilon_2)) + \lambda(\epsilon_1,\epsilon_2)\ \ Pr(T = c(\epsilon_1,\epsilon_2))$. Let

$$h_1(\frac{\theta_1+\epsilon_1}{1-\epsilon_1-\epsilon_2}, \frac{\theta_0-\epsilon_1}{1-\epsilon_1-\epsilon_2}, \alpha) = Pr(T > c(\epsilon_1,\epsilon_2)) + Pr(T = c(\epsilon_1,\epsilon_2)).$$

So, if ϕ' is any test of level α , then

$$\begin{aligned}\beta_{\phi'}(\theta_1) &= \inf_{F:\theta(F)=\theta_1} Pr(\phi' \text{ Rejects } H_0|F) \\ &\leq \inf_{\theta(F)=\theta_1,\, F\in\mathcal{F}_0} Pr(\phi'\text{Rejects } H_0|F) \\ &\leq h_1(\frac{\theta_1+\epsilon_1}{1-\epsilon_1-\epsilon_2}, \frac{\theta_0-\epsilon_1}{1-\epsilon_1-\epsilon_2}, \alpha).\end{aligned}$$

We know that h_1 is a continuous function, letting $\epsilon \longrightarrow 0$, we have

$$\begin{aligned}\beta_{\phi'}(\theta_1) &\leq h(\theta_1,\theta_0,\alpha) \\ Pr(\phi \text{ rejects } H_0\ |F) &= E_F\phi(g(X_1), g(X_2), \ldots, g(X_n))\end{aligned}$$

$$
\begin{aligned}
&= Pr(Z_1 + Z_2 + \ldots + Z_n > c) \\
&\quad + \lambda \, Pr(Z_1 + Z_2 + \ldots + Z_n = c).
\end{aligned}
$$

Here $Z_1, Z_2, \ldots, Z_n$ are independent Bernolli with common parameter $\theta(F)$.

Hence

$$\beta_{\phi}(\theta_1) = h(\theta_1, \theta_0, \alpha)$$

that is

$$\beta_{\phi'}(\theta_1) \leq \beta_{\phi}(\theta_1) \quad \forall \, \theta_1 > \theta_0.$$

This proves the theorem.

ACKNOWLEDGMENTS:

My sincere thanks go to my dissertation advisor, Professor Shanti S. Gupta. His suggestions, criticism and enthusiasm have been a source of constant encouragement in the preparation of this report.

References

[1] Bahadur, R. R. (1950) *" On a problem in the theory of k populations "* Annals of Mathematical Statistics 2, 362-375.

[2] Barlow, R. E. and Gupta, S. S. (1969) *" Selection procedures for restricted families of probability distribution "* Annals of Mathematical Statistics 40, 905-917.

[3] Barlow, R. E. Gupta, S. S. and Panchapakesan, S. (1969) *" On the distribution of the maximum and minimum of ratios of order statistics "* Annals of Mathematical Statistics 40, 918-934.

[4] Bechhofer, R. E. (1954) *" A singel sample multiple decision procedure for ranking means of normal populations with known variances "* Annals of Mathematical Statistics 25, 16-39.

[5] Brownie, C. (1988) *" Reader reaction : Estimating $Pr(X < Y)$ in categorized data using 'ROC' analysis "* Biometrice 44, 615-620.

[6] Desu, M. M. and Bristol, D. R. (1986) *" Rizvi and Sobel type non parametric selection procedures : Review and Extensions "* AJMMS 6, 87-107.

[7] Eaton, M. L. (1967) *" Some optimum properties of ranking procedures "* Annals of Mathematical Statistics 38, 124-137.

[8] Gupta, S. S. (1956) *" On a decision rule for a problem of ranking means "* Memeograph series No. 150 , Institute of statistics, University of North Carolina, Chapel Hill.

[9] Gupta, S. S. and Huang, D. Y. (1974) *" Non parametric subset selection procedures for the 't' best populations "* Bulletin of Mathematics , Academia Sinica 2, 377-386.

[10] Gupta, S. S. and McDonald, G. C. (1970) *" On classes of selection procedures based on ranks "* Nonparametric Techniques in statistical Inference , Ed M. L. puri , Cambridge University Press, London, 419-514.

[11] Gupta, S. S. and Miescke, K. J. (1984) *"Sequential selection procedures -decision theoretic approach"* Annals of Statistics 12, 336-350.

[12] Gupta, S. S. and Panchapakesan, S. (1986) *" Subset selection procedures : Review nad Assessments "* JAMMS 5, 235-311.

[13] Gupta, S. S. and Panchapakesan (1979) *" Multiple Decision Procedures "* Wiley, New York.

[14] Lehmann, E. L. (1961) *" Some model I problems of selection "* Annals of Mathematical Statistics 32, 990-1012.

[15] Mosteller, F. (1948) *" A k-sample slippage test for an extrem population "* Annals of Mathematical Statistics 19, 58-65.

[16] Paulson, E. (1949) *" A multiple decision procedure for certain problems in analysis of variance "* Annals of Mathematical Statistics 20, 95-98.

[17] Rizvi, M. H. and Sobel, M. (1967) *" Nonparametric procedures for selecting a subset containing the population with largest α -th quantile "* Annals of Mathematical Statistics 38, 1788-1803.

[18] Simonoff, J. S. Hochber, Y. and Reiser, B (1986) *" Alternative estimation procedure for $Pr(X < Y)$ in categorized data "* Biometrics 42, 895-907.

[19] Sobel, M. (1967) *" Nonparametric procedures for selecting the 't' populations with the largest α -th quantile "* Annals of Mathematical Statistics 38, 1804-1816.

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified	**1b. RESTRICTIVE MARKINGS**
2a. SECURITY CLASSIFICATION AUTHORITY	**3. DISTRIBUTION/AVAILABILITY OF REPORT** Approved for public release, distribution unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) Technical Report #90-06	**5. MONITORING ORGANIZATION REPORT NUMBER(S)**
6a. NAME OF PERFORMING ORGANIZATION Purdue University — **6b. OFFICE SYMBOL** *(If applicable)*	**7a. NAME OF MONITORING ORGANIZATION**
6c. ADDRESS *(City, State, and ZIP Code)* Department of Statistics, West Lafayette, IN 47907	**7b. ADDRESS** *(City, State, and ZIP Code)*
8a. NAME OF FUNDING/SPONSORING ORGANIZATION Office of Naval Research — **8b. OFFICE SYMBOL** *(If applicable)*	**9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER** NSF Grants DMS-8606964, DMS-87-02620 N00014-88-K-0170
8c. ADDRESS *(City, State, and ZIP Code)* Arlington, VA 22217-5000	**10. SOURCE OF FUNDING NUMBERS** PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. / WORK UNIT ACCESSION NO.

11. TITLE *(Include Security Classification)*
Nonparametric Selection, Ranking and Testing (Unclassified)

12. PERSONAL AUTHOR(S)
Sayaji N. Hande

13a. TYPE OF REPORT	**13b. TIME COVERED**	**14. DATE OF REPORT** *(Year, Month, Day)*	**15. PAGE COUNT**
Technical	FROM ____ TO ____	January 1990	23

16. SUPPLEMENTARY NOTATION

17. COSATI CODES		
FIELD	GROUP	SUB-GROUP

18. SUBJECT TERMS *(Continue on reverse if necessary and identify by block number)*
Selection and Ranking; Nonparametric; Comparison with a Control; Testing; Minimax Decision Rules

19. ABSTRACT *(Continue on reverse if necessary and identify by block number)*

In this paper we consider the problem of ranking (partitioning) k populations according to the parameter which is defined as a functionals of the distribution functions on the underlying populations. We obtain minimax rules for general loss functions, Bayes rules for some specific loss functions and propose approximate non-randomized minimax rules. We also derive restricted minimax rules for selecting a subset of populations which are better than a control. Some nonparametric "optimal" tests are derived for different hypotheses written in terms of the parameter as a functional of the underlying distribution function.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS	**21. ABSTRACT SECURITY CLASSIFICATION** Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL Shanti S. Gupta	**22b. TELEPHONE** *(Include Area Code)* (317) 494-6030 — **22c. OFFICE SYMBOL**