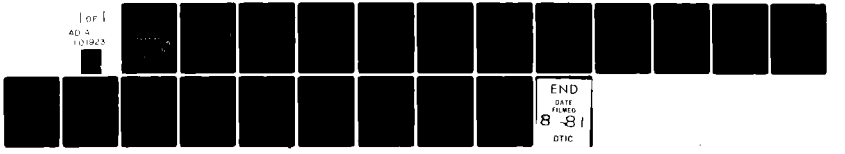
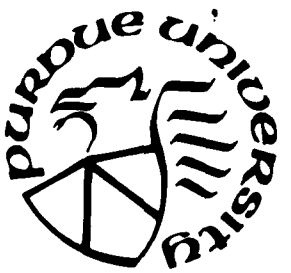
AD-A101 923 PURDUE UNIV LAFAYETTE IN DEPT OF STATISTICS F/G 12/1
SOME LOCALLY OPTIMAL SUBSET SELECTION RULES,(U)
JUN 81 D HUANG, S PANCHAPAKESAN N00014-75-C-0455
UNCLASSIFIED MMS-81-23 NL



(12)

PURDUE UNIVERSITY



DEPARTMENT OF STATISTICS

DIVISION OF MATHEMATICAL SCIENCES

(12)

(6) SOME LOCALLY OPTIMAL SUBSET SELECTION RULES*

(10) By Deng-Yuan Huang and S. Panchapakesan
Academia Sinica, Taipei and Southern Illinois University

(9) Mimeograph Series, #81-23

(11) NNS-84-

Department of Statistics
Division of Mathematical Sciences

11 June 1981 (13)

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

SOME LOCALLY OPTIMAL SUBSET SELECTION RULES*

By Deng-Yuan Huang and S. Panchapakesan
Academia Sinica, Taipei and Southern Illinois University

1. Introduction. Let $\pi_0, \pi_1, \ldots, \pi_k$ be k + 1 independent populations where π_i has the associated density function $f(x,\theta_i)$ with the unknown parameter θ_i belonging to an interval Θ of the real line. In this paper, we are concerned with two types of goals. The first is to select from $\pi_1, \ldots, \pi_k$ those populations, if any, that are better (to be defined) than π_0 which is the control or standard population. The other is to select a subset (preferably small in size) of the k populations $\pi_1, \ldots, \pi_k$ that will contain the best (suitably defined) among them. In the recent years, the attention has increasingly been focussed on the construction of optimal selection rules. Some of the important papers in this direction are Bickel and Yahav [1], Chernoff and Yahav [2], Goel and Rubin [4], Gupta and Hsu [5], Gupta and Huang [6,7], and Gupta and Kim [10]. These investigations deal with the symmetric case which implies equal sample sizes. There have been some investigations in the unequal sample sizes case but these relate to ad hoc and heuristic procedures and are not generally successful in establishing the least favorable configuration (LFC) for the probability of a correct decision. For many classical procedures in the literature for selecting a subset containing the best, the LFC turns out to be $\theta_1 = \ldots = \theta_k$. This provides the motivation for seeking selection rules which are optimal in a suitable sense in a neighborhood of every equi-parameter point. When comparing these populations with π_0, the local optimality concerns with the

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

ability of the rule to reject any population that is inferior to π_0, and to choose a population which is superior while all others are inferior.

Section 2 deals with selecting populations that are superior to a control and as the applications indicate we can handle unequal sample sizes. The next section discusses the usual goal of selecting a subset containing the best and locally optimal rules based on ranks. Though the density function $f(x,\theta)$ is assumed to be known but for the value of θ, a rule based on ranks is justified on the basis of usual robustness considerations in that the ranks are insensitive to outliers and there could be possible deviations from the model.

For detailed discussions on multiple decision problems and general decision theoretic approach, the reader is referred to Gupta and Huang [8] and Gupta and Panchapakesan [12].

2. Comparison with a control. We want to construct a rule to select all populations that are better than the control. Let $\{X_{ij}\}$, $j = 1,\ldots,n_i$, denote the sample observations from π_i, $i = 0,1,\ldots,k$. The selection rule will depend upon the observations through the statistics T_{io}, $i = 1,\ldots,k$, where T_{ij} is suitably defined to indicate the difference between π_i and π_j. For fixed $n_0,\ldots,n_k$, we assume that T_{ij} has a density function $g_{\tau_{ij}}(t_{ij})$ depending on the parameter τ_{ij}. Usually the T_{ij} are invariantly sufficient for τ_{ij}. We use the parameter τ_{ij} as our measure of 'distance' between π_i and π_j. Now, τ_{ii} is the same for all i and this common value is denoted by τ^*. In the case of a location parameter θ, we may have $\tau_{ij} = |\theta_i - \theta_j|$ and in this case, $\tau^* = 0$. On the other hand, if θ is a scale parameter, we can define $\tau_{ij} = \max\left(\frac{\theta_i}{\theta_j}, \frac{\theta_j}{\theta_i}\right)$, which gives $\tau^* = 1$. We define population π_i to be superior if $\tau_{io} > \tau^*$ and inferior otherwise.

Let $\Omega = \{\underline{\tau} | \underline{\tau} = (\tau_{1o},\ldots,\tau_{ko})\}$, $\Omega_{io} = \{\underline{\tau} | \tau_{io} \leq \tau^*\}$ and $\Omega^*_{io} = \{\underline{\tau} | \tau_{jo} = \tau^* < \tau_{io}, j=1,\ldots,k; j\neq i\}$. In other words, Ω_{io} is the set of configurations for which π_i is inferior and Ω^*_{io} is the set of configurations for which π_i is superior and all other populations are 'just inferior'.

2.1. Optimality requirement. Let $0 < \gamma < 1$ be given. We wish to derive a rule which satisfies

$$\sup_{\Omega_{io}} P_{\underline{\tau}}[\pi_i \text{ is selected}] \leq \gamma \text{ for } i = 1,\ldots,k, \text{ and} \tag{2.1}$$

$$\text{maximizes } \sum_{i=1}^{k} \frac{\partial}{\partial \tau_{io}} P_{\underline{\tau}}[\pi_i \text{ is selected} | \Omega^*_{io}]\Big|_{\tau_{io}=\tau^*} \tag{2.2}$$

among all rules satisfying (2.1). We note that the condition (2.1) controls the error probabilities and the condition (2.2) maximizes the efficiency in a certain sense of the rule in picking out the superior populations in a neighborhood of the configuration $\underline{\tau} = (\tau^*,\ldots,\tau^*)$.

Let $h_{\underline{\tau}}(t_1,\ldots,t_k)$ denote the joint density of $T_{1o},\ldots,T_{ko}$ with respect to a σ-finite measure μ. A selection rule is given by $\delta(\underline{t}) = \{\delta_1(\underline{t}),\ldots,\delta_k(\underline{t})\}$, where $\delta_i(\underline{t})$ is the probability of selecting π_i when $\underline{t} = (t_1,\ldots,t_k)$ is observed. Further, let $h_{\tau^*}(\underline{t})$ denote the density $h_{\underline{\tau}}(\underline{t})$ when $\underline{\tau} = (\tau^*,\ldots,\tau^*)$, and let $h^{(i)}_{\tau^*}(\underline{t})$ denote the partial derivative $\frac{\partial}{\partial \tau_{io}} h_{\underline{\tau}}(\underline{t})$ evaluated at $\underline{\tau} = (\tau^*,\ldots,\tau^*)$. Finally, we need to assume certain regularity conditions, namely, that $h_{\underline{\tau}}(\underline{t})$ is continuously differentiable with respect to each component of $\underline{\tau}$ and $\frac{\partial}{\partial \tau_{io}} h_{\underline{\tau}}(\underline{t})$ is integrable. Under these regularity conditions, it is easy to see that

$$\frac{\partial}{\partial \tau_{io}} P_{\underline{\tau}}[\pi_i \text{ is selected} | \Omega^*_{io}]\Big|_{\tau_{io}=\tau^*} = \int_{\mathcal{T}} \delta_i(\underline{t}) h^{(i)}_{\tau^*}(\underline{t}) d\mu(\underline{t})$$

where $\mathcal{T}$ is the sample space of $(T_{1o},\ldots,T_{ko})$.

Summarizing the above discussion, we are seeking a rule $\delta(\underline{t})$ which satisfies

$$\int_{\mathcal{T}} \delta_i(\underline{t}) h_{\underline{\tau}}(\underline{t}) d\mu(\underline{t}) \leq \gamma \text{ for } \underline{\tau} \in \Omega_{io}, \; i = 1,\ldots,k, \tag{2.3}$$

and maximizes, among all rules satisfying (2.3), the expression

$$\sum_{i=1}^{k} \int_{\mathcal{T}} \delta_i(\underline{t}) h_{\tau*}^{(i)}(\underline{t}) d\mu(\underline{t}). \tag{2.4}$$

2.2. A locally optimal rule. We now state and prove the main theorem of this section.

Theorem 2.1. Under all the assumptions stated previously, a rule $\delta^o(\underline{t})$ which satisfies (2.3) and maximizes (2.4) among all rules satsifying (2.3) is given by

$$\delta_i^o(\underline{t}) = \begin{cases} 1 & \\ \lambda & \text{if } h_{\tau*}^{(i)}(\underline{t}) = c_i h_{\tau*}(\underline{t}) \\ 0 & \end{cases} \begin{matrix} > \\ \\ < \end{matrix} \tag{2.5}$$

where $0 < \lambda < 1$ and c_i are determined by

$$\int_{\mathcal{T}} \delta_i^o(\underline{t}) h_{\tau*}(\underline{t}) d\mu(\underline{t}) = \gamma, \quad i = 1,\ldots,k. \tag{2.6}$$

Proof. The proof is straightforward by noting that for any rule δ satisfying (2.3) we have

$$\sum_{i=1}^{k} \int_{\mathcal{T}} [\delta_1(\underline{t}) - \delta_i^o(\underline{t})][h_{\tau*}^{(i)}(\underline{t}) - c_i h_{\tau*}(\underline{t})] d\mu(\underline{t}) \leq 0. \tag{2.7}$$

2.3 Applications to special cases. We consider the following cases: (A) normal means comparison; common known variance and unequal sample sizes, (B) normal means comparison; common unknown variance and unequal sample

sizes, and (C) gamma scale parameters comparison; unequal degrees of freedom.

Case A: Normal means comparison; common known variance and unequal sample sizes. Let π_i be $N(\theta_i,\sigma^2)$ and $\bar{X}_i$ be the sample mean based on n_i independent observations from π_i, $i = 0,1,\ldots,k$. We take $T_{io} = \bar{X}_i - \bar{X}_o$ and $\tau_{io} = \theta_i - \theta_o$. Here $\tau^* = 0$. The joint density $h_{\underline{\tau}}(t_1,\ldots,t_k)$ of $T_{1o},\ldots,T_{ko}$ is that of a multivariate normal distribution with mean vector $\underline{\tau}$ and covariance matrix Σ given by

$$\Sigma = \frac{\sigma^2}{n_o}\begin{bmatrix} 1+\alpha_1^2 & & & \\ & \ddots & & 1 \\ & & \ddots & \\ & 1 & & \ddots \\ & & & 1+\alpha_k^2 \end{bmatrix},$$

where $\alpha_i^2 = n_o/n_i$, $i = 1,\ldots,k$. It is easy to verify that

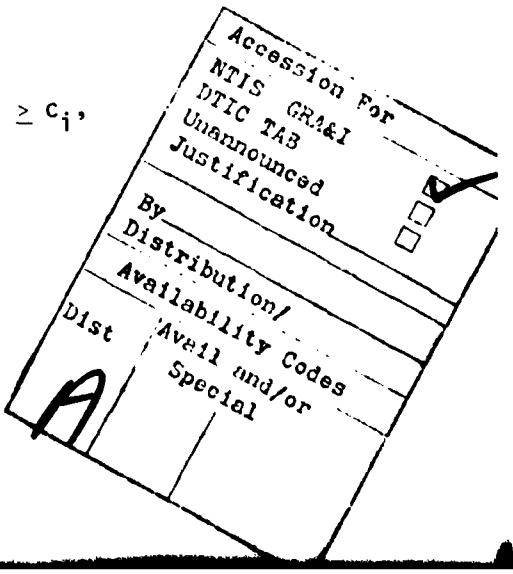
$$h^{(i)}_{\tau^*}(\underline{t})/h_{\tau^*}(\underline{t}) = \sum_{j=1}^{k} t_j \sigma^{ij},$$

where $\Sigma^{-1} = (\sigma^{ij})$.

In the particular case in which $n_o \neq n_1 = \ldots = n_k$, we have $\alpha_1 = \ldots = \alpha_k = \alpha$ (say). In this case, it is easy to evaluate Σ^{-1}. The rule δ^o is given by

$$\delta_i^o(\underline{t}) = \begin{cases} 1 & t_i - \frac{1}{k+\alpha^2}\sum_{j=1}^{k} t_j \geq c_i, \\ 0 & \text{otherwise,} \end{cases} \tag{2.8}$$

where c_i is determined by



$$(2.9) \qquad P_{\tau^*}\{Z_i - \frac{1}{k+\alpha^2} \sum_{j=1}^{k} Z_j \geq c_i\} = \gamma,$$

$Z_1,\ldots,Z_k$ being normal variables with $E(Z_i) = 0$, $V(Z_i) = \frac{\sigma^2}{n_o}(1+\alpha^2)$, and $E(Z_iZ_j) = \frac{\sigma^2}{n_o}$. The left hand side of (2.9) is equal to

$$P_{\tau^*}\{(k+\alpha^2-1)Z_i - \sum_{j\neq i} Z_j > c_i(k+\alpha^2)\} = 1 - \Phi(c_i(k+\alpha^2)/\eta),$$

where $\Phi(\cdot)$ is the standard normal cdf and

$$\eta^2 = \frac{\sigma^2}{n_o}[(1+\alpha^2)(k+\alpha^2-1)^2 + (1+\alpha^2)(k-1) - 2(k-1)(k+1-\alpha^2) + (k-1)(k-2)].$$

Thus,

$$(2.10) \qquad c_i = \frac{\eta}{k+\tau^2}\Phi^{-1}(1-\gamma).$$

Remark 2.1. The rule δ^0 given by (2.8) is of the form: select π_i if and only if $t_i \geq a\frac{1}{k-1}\sum_{j\neq i} t_j - c$, where $c > 0$ and $a < 1$. If, however, we take $a = 1$, then we have the so-called average-type rule.

Case B: Normal means comparison; common unknown variance and unequal sample sizes. As in Case A, $\bar{X}_i$ is the sample mean based on n_i observations from π_i. Let s^2 be the usual pooled unbiased estimator of σ^2 based on $\nu = (k+1)(n-1)$ degrees of freedom. Define $Z_i = (\bar{X}_i - \bar{X}_o)/\sqrt{\frac{1}{n_i}+\frac{1}{n_o}}$ and $T_{io} = Z_i/s$, $i = 1,\ldots,k$. Let $\Omega = (\rho_{ij})$ where $\rho_{ij} = \mathrm{Corr}(Z_i,Z_j)$. Then $\rho_{ij} = [(1+\alpha_i^2)(1+\alpha_j^2)]^{-\frac{1}{2}}$ where $\alpha_i^2 = n_o/n_i$. Let $A = (a_{ij})$ be the inverse of Ω. Then the joint density $h_{\underline{T}}(t_1,\ldots,t_k)$ of the T_{io} is the multivariate t-density (see Dunnett and Sobel [3]) given by

$$(2.11) \quad h_{\underline{T}}(t_1,\ldots,t_k) = \frac{A^{\frac{1}{2}}\Gamma(\frac{\nu+k}{2})}{(\nu\pi)^{k/2}\Gamma(\frac{\nu}{2})}[1+\frac{1}{\nu}\sum_{i,j} a_{ij}(t_i-\tau_{io})(t_j-\tau_{jo})]^{-\frac{1}{2}(\nu+k)}$$

where $\tau_{io} = \theta_i - \theta_o$, $i = 1,\ldots,k$.

Now, it is easy to verify that

$$\frac{h^{(i)}_{\tau^*}(\underline{t})}{h_{\tau^*}(\underline{t})} = \frac{\nu+k}{\nu} \frac{a_{ii}t_i + \sum_{j\neq i} a_{ij}t_j}{1+\frac{1}{\nu}\sum_{i,j} a_{ij}t_it_j} = \ell(\underline{t};\nu,k,\{a_{ij}\}), \text{ say.}$$

Thus the rule δ^o is given by

$$\delta^o_i(\underline{t}) = \begin{cases} 1 & \text{if } \ell(\underline{t};\nu,k,\{a_{ij}\}) \geq c \\ 0 & \text{otherwise.} \end{cases} \tag{2.12}$$

In the particular case in which $n_o \neq n_1 = n_2 = \ldots = n_k$, we get $\rho_{ij} = 1/(1+\alpha^2)$; where $\alpha_1 = \ldots = \alpha_k = \alpha$. This gives

$$a_{ij} = \begin{cases} \frac{(1+\alpha^2)(k-1+\alpha^2)}{\alpha^2(k+\alpha^2)}, & i = j \\ -\frac{1+\alpha^2}{\alpha(k+\alpha^2)}, & i \neq j. \end{cases}$$

The constant c_i is given by

$$P_{\tau^*}[\ell(T_{1o},\ldots,T_{ko};\nu,k,\{a_{ij}\}) \geq c_i] = \gamma. \tag{2.13}$$

It can be easily seen that (2.13) can be rewritten as

$$P[d_i\underline{Y}'A\underline{Y} - A_i'\underline{Y} + \nu d_i \leq 0] = \gamma, \tag{2.14}$$

where $d_i = c_i/(\nu+k)$, $A' = [A_1 \ldots A_k]$, and $\underline{Y}' = (Y_1 \ldots Y_k)$ has the multivariate t density given by (2.11) with $\tau_{io} = 0$, $i = 1,\ldots,k$. The left hand side of (2.14) is equal to

$$P[d_i(\underline{Y} - \frac{1}{2d_i}A^{-1}A_i)'A(\underline{Y} - \frac{1}{2d_i}A^{-1}A_i) + \nu d_i - \frac{1}{4d_i}A_i'A^{-1}A_i \leq 0]$$

$$= P[d_i(\underline{Z} - \frac{s}{2d_i}A^{-1}A_i)'A(\underline{Z} - \frac{s}{2d_i}A^{-1}A_i) + s^2\nu d_i - \frac{s^2}{4d_i}A_i'A^{-1}A_i \leq 0]$$

where $\underline{Z} \sim N(\underline{0},\sigma^2\Omega)$,

$$= P[d_i(\frac{1}{\sigma}\underline{Z} - \frac{\sqrt{\nu}\,s}{\sigma}\ \frac{1}{2d_i\sqrt{\nu}}A^{-1}A_i)'A(\frac{1}{\sigma}\underline{Z} - \frac{\sqrt{\nu}\,s}{\sigma}\ \frac{1}{2d_i\sqrt{\nu}}A^{-1}A_i)$$

$$+ \frac{\nu s^2}{\sigma^2}d_i - \frac{\nu s^2}{\sigma^2}\ \frac{1}{4d_i\nu}A_i'A^{-1}A_i \le 0].$$

In order to evaluate the above probability, we note that $U = \frac{\nu s^2}{\sigma^2} \sim \chi^2_\nu$ and that, conditioned on $U = u$, the quadratic form has a non-central χ^2 distribution on k degrees of freedom with the noncentrality parameter $\lambda = \lambda(u,d_i) = \frac{u}{4\nu d_i^2}A_i'A^{-1}A_i$. Thus the constant c_i is given by

$$(2.15) \qquad \int_0^\infty P[d_i\{V_i + u - \lambda(u,d_i)\} \le 0]g_\nu(u)du = \gamma,$$

where V_i has a noncentral χ^2 distribution on k degrees of freedom with noncentrality parameter $\lambda(u,d_i)$, and $g_\nu(u)$ is the central χ^2 density on ν degrees of freedom. Tables for the constant d_i are expected to be part of a future paper.

Case C: Gamma scale parameters comparison; unequal shape parameters.

Let π_i $(i = 0,1,\ldots,k)$ be a gamma population with the density

$$(2.16) \qquad f(x;\theta_i,\nu_i) = \frac{x^{\nu_i-1}}{\Gamma(\nu_i)\theta_i^{\nu_i}}\exp\{-x/\theta_i\}, \quad x \ge 0, \quad \theta > 0.$$

We take $\tau_{io} = \theta_i/\theta_o$ so that $\tau^* = 1$. Let $X_1,\ldots,X_k$ be independent observations from these populations. Define $T_{io} = X_i/X_o$, $i = 1,\ldots,k$. The joint density of $T_{1o},\ldots,T_{ko}$ is easily derived to be

$$(2.17) \qquad h_{\underline{\tau}}(\underline{t}) = \frac{\Gamma(m)}{\left\{\prod_{i=0}^{k}\Gamma(\nu_i)\tau_{io}^{\nu_i}\right\}}\ \frac{\prod_{i=1}^{k}t_i^{\nu_i-1}}{\left[1+\sum_{i=1}^{k}\frac{t_i}{\tau_{io}}\right]^m}$$

where $m = \nu_0 + \ldots + \nu_k$. From this, we get

$$\frac{h^{(i)}_{\tau^*}(\underline{t})}{h_{\tau^*}(\underline{t})} = \frac{mt_i}{1+t_1+\ldots+t_k} - \nu_i. \tag{2.18}$$

Thus the rule δ^0 is given by

$$\delta^0_i(\underline{t}) = \begin{cases} 1 & mt_i \geq (\nu_i + c_i)(1 + t_1 + \ldots + t_k) \\ 0 & \text{otherwise.} \end{cases} \tag{2.19}$$

The constant c_i is given by

$$P_{\tau^*}\left(\frac{mX_i}{X_0+\ldots+X_k} \geq \nu_i + c_i\right) = \gamma \tag{2.20}$$

where $X_0,\ldots,X_k$ are independent standardized ($\theta = 1$) gamma variables with degrees of freedom $\nu_0,\ldots,\nu_k$, respectively. It is known that $X_i/(X_0+\ldots+X_k)$ has a beta distribution with parameters ν_i and $m - \nu_i$, denoted by $B(\nu_i, m-\nu_i)$. Thus $(\nu_i + c_i)m^{-1}$ is the upper 100 γ% point of $B(\nu_i, m-\nu_i)$ distribution. So c_i can be obtained from the tables of the incomplete beta function.

It should be pointed out that the above problem includes that of comparing normal variances based on unequal sample sizes.

3. A locally strongly monotone invariant subset selection rule based on ranks. In this section, we consider the goal of selecting a subset containing the best population. Subset selection procedures based on ranks have been earlier studied by several authors notably among whom are Gupta, Huang and Nagel [9], Gutpa and McDonald [11], McDonald [13,14], and Nagel [15]. A detailed account of these procedures is available in Gupta and Panchapakesan [12]. Of these authors, Nagel [15] and Gupta, Huang and Nagel [9] have discussed locally optimal selection rules but using

different criteria from ours. The setup of our investigation is analogous to that of Gupta, Huang and Nagel [9]. We assume a parametric model but seek procedures based on ranks for robustness reasons as explained in Section 1.

3.1. Preliminaries. Let $\pi_1,\ldots,\pi_k$ be k independent populations where π_i has the associated distribution function $F(x,\theta_i)$ assumed to be absolutely continuous with density $f(x,\theta_i)$ which is known except for the value of the parameter θ_i. The parameter θ_i belongs to some interval on the real line containing the origin. It is further assumed that the density $f(x,\theta)$ satisfies the set of conditions in (3.3). The population associated with the largest θ_i is called the best population. In case of a tie, one of the contenders is assumed to be tagged as the best. Let $\Omega = \{\underline{\theta} | \underline{\theta} = (\theta_1,\ldots,\theta_k)\}$ and $\Omega_0 = \{\underline{\theta} | \theta_1 = \ldots = \theta_k\}$. Let $\{X_{ij}\}$, $j = 1,\ldots,n$, be independent observations from π_i and let R_{ij} denote the rank of X_{ij} in the pooled sample of $N = kn$ observations. The smallest observation has rank 1 and the largest N.

Definition 3.1. A rank configuration is an N-tuple $\Delta = (\Delta_1,\ldots,\Delta_N)$, $\Delta_i \in \{1,2,\ldots,k\}$, where $\Delta_i = j$ means that the ith smallest observation in the pooled sample comes from π_j.

Let $\mathcal{C} = \{\Delta\}$ dentoe the set of all rank configurations for fixed k and n. For fixed Δ, let $\mathcal{X}_\Delta = \{\underline{x} \in \mathcal{X} \mid \Delta_{\underline{x}} = \Delta\}$, where $\mathcal{X} = \{\underline{x}: \underline{x} = (x_1,\ldots,x_N)\}$ and $\Delta_{\underline{x}}$ denotes the rank configuration of $\underline{x} = (x_1,\ldots,x_N)$. The decision space $\mathcal{A}$ consists of all 2^k subsets (including the empty set) of the set $\{1,2,\ldots,k\}$. Any subset is denoted by d so that $\mathcal{A} = \{d | d \subseteq \{1,2,\ldots,k\}\}$. Any decision $d = \{i_1,i_2,\ldots,i_r\}$ corresponds to selecting the subset of the k populations comprising $\pi_{i_1},\ldots,\pi_{i_r}$. Let $\delta(\Delta,d)$ denote the probability

that the decision d is made if the rank configuration Δ is observed.

Definition 3.2. A rank selection rule is a measurable function defined on $\mathcal{C} \times \mathcal{D}$ such that for each $\Delta \in \mathcal{C}$, (i) $\delta(\Delta,d) \geq 0$ and

(ii) $\sum_{d \in \mathcal{D}} \delta(\Delta,d) = 1$.

Let $p_i(\Delta)$, $i = 1,\ldots,k$, denote the individual selection probabilities of the k populations, where

$$p_i(\Delta) = \sum_{d \ni i} \delta(\Delta,d) \tag{3.1}$$

the summation being over all subsets containing i. Let P_α ($\alpha = 1,\ldots,k$) denote the probability of including the population π_α in the selected subset; in other words, $P_\alpha = E_{\underline{\theta}}[p_\alpha(\Delta)]$, $\alpha = 1,\ldots,k$. Any decision d that corresponds to the selection of the best population is called a correct decision. We assume that π_k is the best population and denote the probability of a correct decision by $P(CD|\delta,\Delta)$. Clearly, $P(CD|\delta,\Delta) = E_{\underline{\theta}}[p_k(\Delta)]$.

Definition 3.3. A selection rule is strongly monotone if, for each $i = 1,\ldots,k$, P_i increases in θ_i when all other components of $\underline{\theta}$ are fixed, and P_i decreases in θ_j, $j \neq i$, when all other components of $\underline{\theta}$ are fixed.

Let G denote the group of permutations g of the integers $(1,2,\ldots,k)$. We write $g(1,\ldots,k) = (g1,\ldots,gk)$. Let h denote the inverse of g and define $g(x_1,\ldots,x_k) = (x_{h1},\ldots,x_{hk})$ and $gd = \{i|hi \in d\}$, $d \in \mathcal{D}$. Also, for any $\Delta \in \mathcal{C}$, let $\bar{g}$ be defined by $\bar{g}\Delta = (g\Delta_1,\ldots,g\Delta_N)$. Thus $\bar{g}$ is induced by g. Let $\bar{G}$ be the group $\{\bar{g}\}$. Finally, let $G(i,j) = \{g \in G|gi = j\}$.

Definition 3.4. A selection rule δ is invariant under permutation if and only if $\delta(\bar{g}\Delta,gd) = \delta(\Delta,d)$ for all $\Delta \in \mathcal{C}$, $d \in \mathcal{D}$, $g \in G$ and $\bar{g} \in \bar{G}$.

An invariant selection rule is completely specified by any one of its individual selection probabilities. We wish to derive an invariant rank

selection rule δ such that

$$\inf_{\underline{\theta} \in \Omega_0} P_{\underline{\theta}}(CD|\delta,\Delta) = P^* \tag{3.2}$$

where $P^*(\frac{1}{2^k} < P^* < 1)$ is specified in advance and it is locally optimal in the sense that it is strongly monotone in a neighborhood of any point $\underline{\theta}_0$ in Ω_0. Obviously, the infimum of the probability of a correct decision over that neighborhood is attained at $\underline{\theta}_0$.

3.2. Derivation of a locally optimal rule. We assume that the density $f(x,\theta)$ satisfies the following set of conditions:

(i) $f(x,\theta)$ is absolutely continuous in θ for almost every x;

(ii) the limit

$$\dot{f}(x,0) = \lim_{\theta \to 0} \frac{1}{\theta} [f(x,\theta) - f(x,0)] \tag{3.3}$$

exists for almost every x;

(iii) $\lim_{\theta \to 0} \int_{-\infty}^{\infty} |\dot{f}(x,\theta)|dx = \int_{-\infty}^{\infty} |\dot{f}(x,0)|dx < \infty$ where $\dot{f}(x,\theta) = \frac{\partial}{\partial \theta} f(x,\theta)$.

Now, we note that the distribution of the ranks does not depend upon the underlying distributions of the X_{ij} when $\underline{\theta} \in \Omega_0$ and thus $P_{\underline{\theta}}(CD|\delta,\Delta)$ is constant in Ω_0. Hence we can satisfy (3.2) by choosing any point $\underline{\theta}_0$ in Ω_0 and satisfying $P_{\underline{\theta}_0}(CD|\delta,\Delta) = P^*$. With no loss of generality, we assume that $\underline{\theta}_0 = (0,\ldots,0)$.

The probability that the rank configuration Δ is observed under $\underline{\theta}$ with $\theta_i \neq 0$, $i = 1,\ldots,k$, is given by (see Gupta, Huang and Nagel [9])

$$P_{\underline{\theta}}(\Delta) = A_0 + \sum_{i=1}^{k} \theta_i A_i(\Delta,\underline{\theta}) \tag{3.4}$$

where

$$A_o = \int_{-\infty}^{\infty} \int_{-\infty}^{x_N} \cdots \int_{-\infty}^{x_2} \prod_{i=1}^{N} f(x_i,0)dx_1 \cdots dx_N,$$

$$A_i(\Delta,\underline{\theta}) = \sum_{\substack{j \\ \Delta_j = i}} \int_{-\infty}^{\infty} \int_{-\infty}^{x_N} \cdots \int_{-\infty}^{x_2} \frac{f(x_j,\theta_i)-f(x_j,0)}{\theta_i}$$

$$\times \prod_{e=1}^{j-1} f(x_e,0) \prod_{e=j+1}^{N} f(x_e,\theta_{\Delta_e})dx_1 \cdots dx_N, \quad i = 1,\ldots,k.$$

Here we define $\prod_{i=1}^{0} f(x_i,0) = 1$.

The following lemma summarizes some results of Gupta, Huang and Nagel [9] for an invariant rule.

Lemma 3.1. For an invariant rank selection rule,

$$(3.5) \quad P_{\underline{\theta}}(CD|\delta,\Delta) = E_{\underline{\theta}} p_k(\Delta)$$

$$= \sum_{\Delta \in C} [A_o + \frac{1}{(k-1)!} \sum_{g \in G(k,k)} \sum_{i=1}^{k} \theta_{hi} A_i(\Delta, g\underline{\theta})] p_k(\Delta).$$

There exists an $\epsilon > 0$ such that for all $\underline{\theta}$ with $0 < |\theta_i| < \epsilon$, $i = 1,\ldots,k$, $A_i(\Delta,\underline{\theta})$ is approximately equal to $A_i(\Delta) = \sum_{j:\Delta_j=i} B_j$, where

$$B_j = \int_{-\infty}^{\infty} \int_{-\infty}^{x_N} \cdots \int_{-\infty}^{x_2} \cdot f(x_j,0) \prod_{\substack{e=1 \\ e \neq j}}^{N} f(x_e,0)dx_1 \cdots dx_N.$$

In this case,

$$(3.6) \quad P(CD|\delta,\Delta) \approx \sum_{\Delta \in C} p_k(\Delta)[A_o + \frac{1}{(k-1)} \{(U-\theta_k)V + (k\theta_k - U)A_k(\Delta)\}],$$

where $U = \sum_{i=1}^{k} \theta_i$ and $V = \sum_{i=1}^{k} A_i(\Delta) = \sum_{i=1}^{nk} B_i$, independent of Δ.

Now, for deriving a locally strongly monotone selection procedure, consider

$$(3.7) \qquad P_\alpha \approx \sum_{\Delta \in C} [A_0 + \frac{1}{(k-1)} \{(U - \theta_\alpha)V + (k\theta_\alpha - U)A_\alpha(\Delta)\}]p_\alpha(\Delta).$$

For any $\alpha = 1,\ldots,k$, we get

$$(3.8) \qquad \begin{cases} \frac{\partial P_\alpha}{\partial \theta_\alpha} \approx \sum_{\Delta \in C} A_\alpha(\Delta)p_\alpha(\Delta) \\ \frac{\partial P_\alpha}{\partial \theta_\beta} \approx \sum_{\Delta \in C} \frac{1}{(k-1)} \{V - A_\alpha(\Delta)\}p_\alpha(\Delta), \ \beta \neq \alpha. \end{cases}$$

For the strong monotonicity property, we need to define $p_\alpha(\Delta)$ such that $\sum_{\Delta \in C} A_\alpha(\Delta)p_\alpha(\Delta) > 0$ and $\sum_{\Delta \in C} \{V - A_\alpha(\Delta)\}p_\alpha(\Delta) < 0$. This is obviously accomplished by defining

$$(3.9) \qquad p_\alpha(\Delta) = \begin{cases} 1, & A_\alpha(\Delta) > V + D \\ \rho, & A_\alpha(\Delta) = V + D \\ 0, & A_\alpha(\Delta) < V + D \end{cases}$$

where D and ρ are to be chosen such that

$$(3.10) \qquad V + D > 0 \text{ and } P_{\underline{\theta}_0}[A_\alpha(\Delta) > V + D] + \rho P_{\underline{\theta}_0}[A_\alpha(\Delta) = V + D] = P^*.$$

Summarizing the above discussion, we obtain the following theorem.

Theorem 3.1. Let $f(x,\theta)$ satisfy the conditions in (3.3). Then the invariant subset selection rule based on ranks defined by (3.9) is strongly monotone in a neighborhood of any point $\underline{\theta}_0 \in \Omega_0$ and $\inf_{\underline{\theta} \in \Omega_0} P_{\underline{\theta}}(CD|\delta,\theta) = P^*$ provided constants ρ and D satisfying (3.10) exist.

Remarks. (1) It is possible that D and ρ satisfying (3.10) may not exist. In such a case, the rule selects the empty subset. (2) When D and

ρ satisfying (3.10) exist, the rule (3.9) is also locally optimal in the sense it has the maximum $p_{\underline{\theta}}(CD|\delta, \Delta)$ in a neighborhood of $\underline{\theta}_o$ among all invariant subset selection rules based on ranks satisfying $\inf_{\underline{\theta} \in \Omega_o} p_{\underline{\theta}}(CD|\delta,\Delta)$ This follows from the theorem of Gupta, Huang and Nagel [9]. (3) Rules of the type (3.9) are shown to be "just" by Nagel [15] provided that B_i is non-decreasing in i, which is true for location parameters if and only if $f(x)$ is strongly unimodal (i.e., $-\log f(x)$ is concave). Nagel [15] has also shwon that for a just rule $\inf_{\underline{\theta} \in \Omega} p_{\underline{\theta}}(CD|\delta,\Delta) = \inf_{\underline{\theta} \in \Omega_o} p_{\underline{\theta}}(CD|\delta,\Delta)$.

3.3. A special case. If $f(x,\theta)$ is the logistic density, i.e., $f(x,\theta) = e^{-(x-\theta)}/[1+e^{-(x-\theta)}]^2$, then we get $B_i = r + is$, where the actual values of r and $s > 0$ are immaterial. The rule becomes

$$P_\alpha(\Delta) = \begin{cases} 1 & > \\ \rho & \sum_{j=1}^{n} R_{ij} = C \\ 0 & < \end{cases} \tag{3.11}$$

where ρ and C satisfy

$$P_{\underline{\theta}_o}\left(\sum_{j=1}^{n} R_{ij} > C\right) + \rho P_{\underline{\theta}_o}\left(\sum_{j=1}^{n} R_{ij} = C\right) = P^*.$$

The above rule is a randomized version of the so-called R_3 type rule of Gupta and McDonald [11] who proposed it on an ad hoc basis. It should also be pointed out that the rule is 'just' in the sense of Nagel [15] which, as pointed out earlier, implies that the infimum of the probability of a correct decision over Ω is P*.

REFERENCES

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

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

[3] Dunnett, C. W., and Sobel, M. (1955). Approximations to the probability integral and certain percentage points of a multivariate analogue of Students' t-distribution. Biometrika 42, 258-260.

[4] Goel, P. K. and Rubin, H. (1977). On selecting a subset containing the best populations - a Bayesian approach. Ann. Statist. 5, 969-983.

[5] Gupta, S. S. and Hsu, J. C. (1978). On the performance of some subset selection procedures. Commun. Statist.-Simula. Computa. B7, 561-591.

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

[7] Gupta, S. S. and Huang, D. Y. (1980). A note on optimal subset selection procedures. Ann. Statist. 8, 1164-1167.

[8] Gupta, S. S. and Huang, D. Y. (1981). Multiple Statistical Decision Theory: Recent Developments. Lecture Notes in Statistics, Vol. 8, Springer-Verlag, New York.

[9] Gupta, S. S., Huang, D. Y., and Nagel, K. (1979). Locally optimal subset selection procedures based on ranks. Optimizing Methods in Statistics (Ed. J. Rustagi), Academic Press, New York, pp 251-262.

[10] Gupta, S. S. and Kim, W. C. (1980). Γ-minimax and minimax decision rules for comparison of treatments with a control. Recent Developments in Statistical Inference and Data Analysis (Ed. K. Matusita), North-Holland Publishing Co., Amsterdam, pp. 55-71.

[11] Gupta, S. S. and McDonald, G. C. (1970). On some classes of selection procedures based on ranks. Nonparametric Techniques in Statistical Inference (Ed. M. L. Puri), Cambridge University Press, London, pp. 419-514.

[12] Gupta, S. S. and Panchapakesan, S. (1979). Multiple Decision Procedures: Theory and Methodology of Selecting and Ranking Populations, John Wiley & Sons, New York.

[13] McDonald, G. C. (1972). Some multiple comparison selection procedures based on ranks. Sankhyā Ser. A 34, 53-64.

[14] McDonald, G. C. (1973). The distribution of some rank statistics with applications in block design selection problems. Sankhyā Ser. A 35, 187-203.

[15] Nagel, K. (1970). On subset selection rules with certain optimality properties. Ph.D. Thesis. Also Mimeo. Ser. No. 222, Dept. of Statist., Purdue Univ., West Lafayette, Indiana.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Mimeograph Series #81-23	2. GOVT ACCESSION NO. AD-A102 923	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Some Locally Optiaml Subset Selection Rules		5. TYPE OF REPORT & PERIOD COVERED
		6. PERFORMING ORG. REPORT NUMBER Mimeo. Series #81-23
7. AUTHOR(s) Deng-Yuan Huang and S. Panchapakesan		8. CONTRACT OR GRANT NUMBER(s) N00014-75-C-0455
9. PERFORMING ORGANIZATION NAME AND ADDRESS Purdue University, Dept. of Statistics, W.Lafayette, Indiana		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS		12. REPORT DATE June 1981
		13. NUMBER OF PAGES 17
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release, distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Subset selection, best population, comparison with control, local optimality, rules based on ranks, normal, gamma, logistic.

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

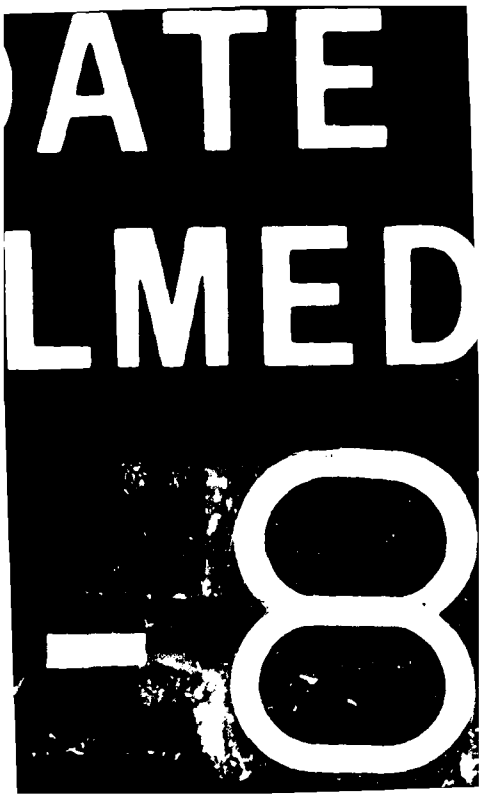
(See reverse side)

DD FORM 1 JAN 73 1473

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Let $\pi_0, \pi_1, \ldots, \pi_k$ be k = 1 independent populations where π_i has the associated density function $f(x,\theta_i)$ with the unknown parameter belonging to an interval Θ of the real line. Two types of problems are studied: (1) to select from $\pi_1, \ldots, \pi_k$ those populations, if any, that are better (to be suitably defined) than π_0 which is the control population; and (2) to select from $\pi_1, \ldots, \pi_k$ a subset preferably of small size so as to contain the best population. For both problems, some locally optimal selection rules are derived. The optimality criteria employed in the two problems are differnt. Further, the procedure for the second problem is based on ranks though the densities are assumed to be kno-n but for the values of the parameters. The rule in the first case is applied to the special cases of (i) normal means comparison with common known variance and unequal sample sizes; (ii) normal means comparison with common unknown variance and unequal sample sizes, and (iii) gamma scale parameters comparison with unequal shape parameters. The rank procedure is specialized to the case of logistic distributions.


ATE
LMED
-8