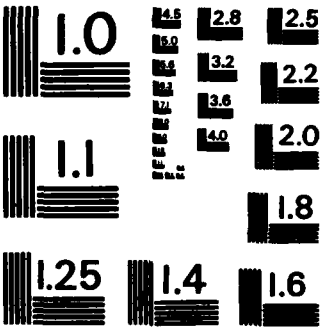
AD-A159 154 ON THE LEAST FAVORABLE CONFIGURATION OF A SELECTION PROCEDURE BASED ON RANKS(U) PURDUE UNIV LAFAYETTE IN DEPT OF STATISTICS S S GUPTA ET AL. AUG 85 TR-85-19 1/1

UNCLASSIFIED N00014-84-C-0167 F/G 12/1 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AD-A159 154

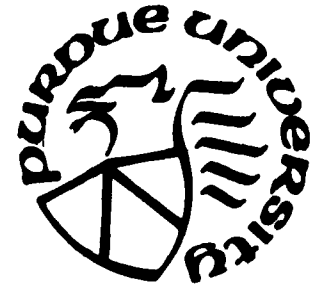
ON THE LEAST FAVORABLE CONFIGURATION OF A SELECTION PROCEDURE BASED ON RANKS

by

Shanti S. Gupta* and Takashi Matsui
Purdue University Purdue University and
Dokkyo University

Technical Report #85-19

PURDUE UNIVERSITY



DEPARTMENT OF STATISTICS

DTIC
ELECTE
SEP 1 3 1985
S D
A

DTIC FILE COPY

This document has been approved for public release and sale; its distribution is unlimited.

85 9 11 009

ON THE LEAST FAVORABLE CONFIGURATION OF A SELECTION PROCEDURE BASED ON RANKS

by

Shanti S. Gupta* and Takashi Matsui
Purdue University Purdue University and
Dokkyo University

Technical Report #85-19

DTIC
ELECTE
SEP 1 3 1985
S D
A

Department of Statistics
Purdue University

...ment has been approved
for public release and sale; its
distribution is unlimited.

August 1985

*The research of this author was supported in part by the ONR Contract N00014-84-C-0167 at Purdue University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

(A)

ON THE LEAST FAVORABLE CONFIGURATION OF A SELECTION PROCEDURE BASED ON RANKS

by

Shanti S. Gupta
Purdue University

and

Takashi Matsui
Purdue University and
Dokkyo University

ABSTRACT

This document ~~Let us~~ consider two types of statistics based on the sums of combined (Wilcoxon type) ranks and vector (Friedman type) ranks. Underlying populations are supposed to belong to the location or scale parameter family of distributions.

Two approaches - subset selection and indifference zone - of ranking and selection procedures based on these statistics are considered in an asymptotic framework for selecting the population with the largest parameter value. The least favorable configurations of parameters are discussed by computing the exact moments of these statistics and introducing an assumption of order relation between the gaps of parameters.

Acces
NTIS
DTIC TAB
Unannounced
Justification
By
Distribution/
Availability Codes
Dist | Avail and/or Special
A-1

DTIC COPY INSPECTED

ON THE LEAST FAVORABLE CONFIGURATION OF A SELECTION PROCEDURE BASED ON RANKS

Shanti S. Gupta* and Takashi Matsui
Purdue University Purdue University and
Dokkyo University

I. INTRODUCTION

Let k populations $\pi_1, \pi_2, \pi_k, \ldots, \pi_k$ be given. A cumulative distribution function (c.d.f) of population π_i is denoted by $F_{\theta_i}(x)$, which is assumed to belong to the location or scale family of distributions. A parameter θ_i is taken from some interval Θ on the real line. $F_{\theta_i}(x)$ is expressed as $F_{\theta_i}(x) = F(x-\theta_i)$ or $F_{\theta_i}(x) = F(X/\theta_i)$ depending on whether it belongs to location or scale family. $F_{\theta_i}(x)$ will be denoted as $F_i(x)$ or F_i for simplicity. Let the ordered parameters of $\theta_1, \theta_2, \ldots, \theta_k$ be denoted as $\theta_{[1]} \leq \theta_{[2]} \leq \cdots \leq \theta_{[k]}$. Then we have

$$F_{\theta_{[1]}}(x) \geq F_{\theta_{[2]}}(x) \geq \cdots \geq F_{\theta_{[k]}}(x) \tag{1.1}$$

for all x. We call the population associate with $F_{\theta_{[k]}}(x)$ the best population. Hereafter, we assume that the population π_k is the best population, without loss of generality.

Take n observations $X_{i1}, X_{i2}, \ldots, X_{in}$ from populations $\pi_i (i=1,2,\ldots,k)$ and consider the following two types of ranks and rank sum statistics. As we mention later, note that when we are dealing with scale parameters, absolute values of the observations are used for obtaining the ranks.

(I) Combined (Wilcoxon type) ranks

*The research of this author was supported in part by the Office of Naval Research Contract N00014-84-C-0167 at Purdue University.

Consider the combined rank of observation X_{ij} among all kxn observations. We denote the rank of X_{ij} by $R_{ij}^{(1)}$ where $R_{ij}^{(1)} = s$ if X_{ij} is the s-th smallest among $X_{11},X_{12},\ldots X_{1n},\ldots,X_{k1},X_{k2},\ldots,X_{kn}$. Also we define

$$H_i^{(1)} = \frac{1}{n} \sum_{j=1}^{n} R_{ij}^{(1)}, \quad i = 1,2,\ldots,k. \tag{1.2}$$

and

$$\underline{H}^{(1)} = (H_1^{(1)}, H_2^{(1)}, \ldots,H_k^{(1)})'. \tag{1.3}$$

(II) Vector (Freedman type) ranks

Consider the rank of observations X_{ij} among $X_{1j},X_{2j}, \ldots,X_{kj}$. We denote the rank of X_{ij} by $R_{ij}^{(2)}$ where $R_{ij}^{(2)} = s$ if X_{ij} is the s-th smallest among $X_{1j},X_{2j},\ldots,X_{kj}$ $(j = 1,2,\ldots,n)$. Rank sum statistic is defined as

$$H_i^{(2)} = \sum_{j=1}^{n} R_{ij}^{(2)}, \quad i = 1,2,\ldots,k \tag{1.4}$$

and

$$\underline{H}^{(2)} = (H_1^{(2)}, H_2^{(2)},\ldots,H_k^{(2)})'. \tag{1.5}$$

Now let us consider the following two approaches of ranking and selection procedures of selecting the best population based on rank statistics $\underline{H}^{(1)}$ and $\underline{H}^{(2)}$. The first approach is a subset selection approach due to Gupta (see Gupta and Panchapakesan (1979)) and we select a subset of populations using following selection procedures.

$$R(\alpha,\beta,1)\text{: Select } \pi_i \text{ if and only if } H_i^{(\alpha)} \geq \max H_j^{(\alpha)} - d_\beta$$
$$i = 1,2,\ldots,k;\ d_\beta \geq 0;\ \alpha,\beta = 1,2. \tag{1.6}$$

These same types of rules are used for selecting the best population with either largest location (β=1) or scale (β=2) parameters, using statistics $\underline{H}^{(\alpha)}$, $\alpha = 1,2$. The use of these rules is warranted by the Theorem 4.2 which we mention later. In fact, $H_k^{(\alpha)}$ corresponds to π_k in the sense stated in the theorem for both location and scale parameter cases. A correct selection (CS) is said to occur if and only if the best population (in our case π_k) is included in the selected subset. Our aim is to select a subset satisfying

$$\inf_{\Omega} P(CS|R(\alpha,\beta,1)) \geq P^* \tag{1.7}$$

where $\alpha,\ \beta = 1,2$; $1/k < P^* < 1$ and $\Omega = \{\underline{\theta} = (\theta_1,\theta_2,\ldots,\theta_k);\ \theta_i \in \Theta,\ i = 1,2,\ldots,k\}$.

Another approach we study is the indifference zone approach due to Bechhofer (1954). The procedure is stated as follows.

$$R(\alpha,\beta,2)\text{: Select the population associate with } H_k^{(\alpha)} \text{ as the best.} \tag{1.8}$$

In this case, the rules $R(\alpha,\beta,2)$, α,β=1,2 are requested to satisfy the following probability requirement;

$$P(CS|R(\alpha,\beta,2)) \geq P^* \text{ whenever } \psi_\beta(\theta_k,\theta_i) \geq c_\beta + \delta_\beta^* \tag{1.9}$$

where $\alpha,\ \beta = 1,2$, $1/k < P^* < 1$, $\delta_\beta^* > 0$ is a given constant.

$$\psi_\beta(\theta_i,\theta_j) = \begin{cases} \theta_i - \theta_j & \text{when } \beta = 1 \\ \theta_i/\theta_j & \text{when } \beta = 2 \end{cases} \tag{1.10}$$

and

$$c_\beta = \begin{cases} 0 & \text{when } \beta = 1 \\ 1 & \text{when } \beta = 2 \end{cases} \tag{1.11}$$

Selection procedures - both subset selection and indifference zone approaches - based on the statistics $\underline{H}^{(1)}$ are studied by many authors including Lehmann (1963), Bartlett and Govindarajule (1968), Gupta and McDonald (1970), Puri and Puri (1968), (1969), Alam and Thompson (1971). Also procedures based on $\underline{H}^{(2)}$ are studied by McDonald (1972),(1973), Matsui (1974), Lee and Dudewicz (1974). A summary of procedures based on ranks is seen in Gupta and McDonald (1980), Gupta and Panchapakesan (1984).

A parameter configuration which gives the infimum of the probability of a correct selection is called the least favorable configuration (LFC). It is fairly troublesome to obtain the LFC for both rules $R(\alpha,\beta,1)$ and $R(\alpha,\beta,2)$ using statistics $\underline{H}^{(1)}$, $\underline{H}^{(2)}$ and still an open question in general ($\alpha,\beta = 1,2$). Including the counter example due to Rizvi and Woodworth (1970), Lee and Dudewicz (1974) and several approaches done by above cited authors, perspective discussion on the LFC is given in Gupta and McDonald (1980).

A purpose of this paper is to discuss the LFC in an asymptotic framework. An order relation is assumed to hold between the gaps of parameters (1.10). This assumption is similar to those considered by Puri and Puri (1968),(1969), Alam and Thompson (1971). The LFC's of the procedures are studied by using the exact moments of the combined and the vector rank statistics $\underline{H}^{(\alpha)}$, $\alpha = 1,2$, for

location and scale parameter cases ($\beta = 1,2$) and for both subset selection and indifference zone approaches.

In Section 2, asymptotic distributions of $\underline{H}^{(\alpha)}$, $\alpha = 1,2$ are considered under the assumption of order relation between gaps of parameters. PCS and LFC are investigated in Section 3. Moments results are given in Section 4 as an Appendix.

2. Asymptotic Property

2.1 Moments of Ranks

Let us define the mean vector and variance-covariance matrix of $H^{(1)}$ by $\underline{\mu}_\beta^{(1)}$ and $\underline{\Lambda}_\beta^{(1)}$ according as we are dealing with location ($\beta = 1$) or scale ($\beta = 2$) parameters. Under the population model we considered in Section 1, the elements of $\underline{\mu}_\beta^{(1)}$ and $\underline{\Lambda}_\beta^{(1)}$ are given as follows. These relations are obtained from more general results given in Theorem 4.1 of Appendix.

$$\mu_{\beta i}^{(1)} = kn\int G^* dF_i + \frac{1}{2}\,, \quad i = 1,2....,k \tag{2.1}$$

$$\lambda_{\beta ij}^{(1)} = \begin{cases} k(3n-1)\int G^* dF_i - 2k(2n-1)\int F_i G^* dF_i + k^2 n\int G^{*2} dF_i \\ - k\int H^* dF_i - k^2 n(\int G^* dF_i)^2 - (n-1)\sum(\int F_m dF_i)^2 - \frac{1}{12}\,, & i = j \\ kn(2 - \int F_j dF_i)\int G^* dF_j + kn(2 - \int F_i dF_j)\int G^* dF_i - n\sum\int F_m dF_i \int F_m dF_j \\ - 2kn\int F_j G^* dF_i - 2kn\int F_i G^* df_j + \int F_i df_j \int F_j dF_i \\ + \int F_i^2 dF_j + \int F_j^2 dF_i - 1 & i \neq j \end{cases} \tag{2.2}$$

where

$$G^*(x) = \frac{1}{k}\sum_{j=1}^{k} F_j(x) \tag{2.3}$$

$$H^*(x) = \frac{1}{k} \sum_{j=1}^{k} F_j^2(x) \tag{2.4}$$

In case of vector rank $R_{ij}^{(2)}$, the moments results are given in Matsui (1985) from which we obtain mean vector $\underline{\mu}_\beta^{(2)}$, variance-covariance matrix $\underline{\Lambda}_\beta^{(2)}$ of statistic $\underline{H}^{(2)}$ as follows;

$$\mu_{\beta i}^{(2)} = kn\int G^* dF_i + \frac{n}{2}, \quad i = 1,2,\ldots,k \tag{2.5}$$

$$\lambda_{\beta ij}^{(2)} = \begin{cases} n[2k\int G^* dF_i - 2k\int F_i G^* dF_i + k^2\int G^{*2} dF_i - k\int H^* dF_i \\ \qquad - k^2(\int G^* dF_i)^2 - 1/12]. & i = j \\ n[k(2-\int F_j dF_i)\int G^* df_j + k(2 - \int F_i dF_j)\int G^* dF_i \\ - \sum_{m=1}^{k} \int F_m dF_i \int F_m dF_j - 2k\int F_j G^* dF_i - 2k\int F_i G^* dF_j \\ + \int F_i dF_j \int F_j dF_i + \int F_i^2 dF_j + \int F_j^2 dF_i - 1], & i \neq j \end{cases} \tag{2.6}$$

2.2 Assumption

Let the gap of parameters θ_i and θ_j be $\psi_\beta(\theta_i,\theta_j)$ as given in (1.10), according as the c.d.f. $F_{\theta_i}(x)$ be location ($\beta = 1$) or scale ($\beta = 2$) family of distribution. When we treat the scale parameter, both of combined rank $R_{ij}^{(1)}$ or vector rank $R_{ij}^{(2)}$ are given to the absolute value of observation X_{ij} from c.d.f. $F_{\theta_i}(x)$ ($i = 1,2,\ldots,k$; $j = 1,2,\ldots,n$). Thus the c.d.f. $G_{\theta_i}(x)$ of $|X_{ij}|$ is given as $G_{\theta_i}(x) = F_{\theta_i}(x) - F_{\theta_i}(-x)$ or

$$G_{\theta_i}(x) = G(x/\theta_i) = F(x/\theta_i) - F(-x/\theta_i), \quad x \geq 0 \tag{2.7}$$

We assume the following relation to hold between the gaps of parameters

$\psi_\beta(\theta_i,\theta_j)$. Note here that although we use the same notation $F_i(x)$ for both location or scale cases, we should read c.d.f. $F_i(x)$ to be $G_i(x)$ given in (2.7), in case we are dealing with scale parameter.

We assume that

$$\psi_\beta(\theta_i,\theta_j) = c_\beta + \kappa_{\beta ij}\, n^{-\frac{1}{2}} + o(n^{-\frac{1}{2}}),\ \beta = 1,2. \tag{2.8}$$

where c_β is given by (1.11).

Then putting

$$I_{\beta ij} = \sqrt{n}\{\int F_j(x)dF_i(x) - \int F_i(x)df_i(x)\} \tag{2.9}$$

we have the following lemma.

Lemma 2.1

For $\psi_\beta(\theta_i,\theta_j)$ ($\beta = 1,2$) given by (2.8), we have the following

$$I_{\beta ij} = K_{\beta ij} + o(1) \tag{2.10}$$

where

$$K_{\beta ij} = \begin{cases} \kappa_{1ij}\int f^2(x)dx & \text{when } \beta = 1 \\ \kappa_{2ij}\int xf^2(x)dx & \text{when } \beta = 2 \end{cases} \tag{2.11}$$

$$i,j = 1,2,\ldots,k;\ i \neq j.$$

Example:

When $F(x)$ is normal $N(0,1)$, we have for $\psi_\beta(\theta_i,\theta_j)$ given by (2.8).

$$I_{1ij} = \frac{1}{2\sqrt{\pi}} \kappa_{1ij} + o(1) \tag{2.12}$$

$$I_{2ij} = \frac{1}{\pi} \kappa_{2ij} + o(1) \tag{2.13}$$

2.3 Asymptotic Distribution

Let us define

$$W_i^{(\alpha)} = \frac{1}{\sqrt{n}} (H_k^{(\alpha)} - H_i^{(\alpha)}), \ \alpha = 1,2 \tag{2.14}$$

that is

$$\underline{W}^{(\alpha)} = \frac{1}{\sqrt{n}} \underline{A}\, \underline{H}^{(\alpha)}, \ \alpha = 1,2. \tag{2.15}$$

where $\underline{W}^{(\alpha)} = (W_1^{(\alpha)}, W_2^{(\alpha)}, \ldots, W_k^{(\alpha)})'$, $\underline{A} = (-\underline{E}_{(k-1)}, \underline{J}_{(k)})_{(k-1)xk}$ and $\underline{E}_{(k-1)}$ is a unit matrix of order k-1, $\underline{J}_{(k)} = (1,1,\ldots,1)'$ kx1. $\underline{W}^{(\alpha)}$ has a mean vector $\underline{\eta}_\beta^{(\alpha)}$, variances-covariance matrix $\underline{\Sigma}_\beta^{(\alpha)}$ such that

$$\underline{\eta}_\beta^{(\alpha)} = \frac{1}{\sqrt{n}} \underline{A}\underline{\mu}_\beta^{(\alpha)} \tag{2.16}$$

$$\underline{\Sigma}_\beta^{(\alpha)} = \frac{1}{n} \underline{A}\underline{\Lambda}_\beta^{(\alpha)}\underline{A}' \tag{2.17}$$

Elements of $\underline{\eta}_\beta^{(\alpha)}$ and $\underline{\Sigma}_\beta^{(\alpha)}$ are given as

$$\underline{\eta}_{\beta i}^{(\alpha)} = \frac{1}{\sqrt{n}} (\mu_{\beta k}^{(\alpha)} - \mu_{\beta i}^{(\alpha)}) \ , \ i = 1,2,\ldots,k-1 \tag{2.18}$$

$$\sigma_{\beta ij}^{(\alpha)} = \frac{1}{n}\left(\lambda_{\beta ij}^{(\alpha)} - \lambda_{\beta ik}^{(\alpha)} - \lambda_{\beta kj}^{(\alpha)} + \lambda_{\beta kk}^{(\alpha)}\right),\ i,j = 1,2,\ldots,k-1 \tag{2.19}$$

where $\mu_{\beta i}^{(\alpha)}$ and $\lambda_{\beta ij}^{(\alpha)}$ are given by (2.1) through (2.6).

Now under the assumption (2.8) using lemma 2.1, we have for $\beta = 1,2$ and $\alpha = 1,2$

$$n_{\beta i}^{(\alpha)} = \frac{1}{\sqrt{n}}\left\{n\int \sum_{j=1}^{k} F_j dF_k - n\int \sum_{j=1}^{k} F_j dF_i\right\}$$

$$\longrightarrow \sum_{j=1}^{k-1} K_{\beta kj} - \sum_{\substack{j=1 \\ j\neq i}}^{k} K_{\beta ij} \quad (\equiv \tilde{n}_{\beta i}^{(\alpha)}) \tag{2.20}$$

as $n \to \infty$, where $K_{\beta ij}$ is given by (2.11). Also since (2.8) is assumed, we have

$$\lambda_{1ij} \to \begin{cases} -k/12 & \text{for } i \neq j \\ (k^2 - k)/12 & \text{for } i = j \end{cases}$$

and

$$\lambda_{2ij} \to \begin{cases} -(k+1)/12 & \text{for } i \neq j \\ (k^2 - 1)/12 & \text{for } i = j \end{cases}$$

Thus $\sigma_{\beta ij}^{(\alpha)}$ goes to the following limit.

$$\sigma_{\beta ij}^{(\alpha)} \longrightarrow \begin{cases} 2v_\alpha & \text{for } i = j \\ v_\alpha & \text{for } i \neq j \end{cases} \tag{2.21}$$

where

$$v_\alpha = \begin{cases} k^2/12 & \text{when } \alpha = 1 \\ k(k+1)/12 & \text{when } \alpha = 2 \end{cases} \tag{2.22}$$

Thus by applying the central limit theorem, we have the following asymptotic distribution of $\underline{W}^{(\alpha)}$.

$$\underline{W}^{(\alpha)} \sim N(\underline{\tilde{\eta}}_{\beta}^{(\alpha)}, \underline{\Sigma}_{\beta}^{(\alpha)}), \ \beta = 1,2 \tag{2.23}$$

where $\underline{\tilde{\eta}}^{(\alpha)} = (\tilde{\eta}_{\beta 1}^{(\alpha)}, \tilde{\eta}_{\beta 2}^{(\alpha)}, \ldots, \tilde{\eta}_{\beta(k-1)}^{(\alpha)})'$ with elements given by (2.20) and

$$\underline{\Sigma}_{\beta}^{(\alpha)} = v_{\alpha}(\underline{E}_{(k-1)} + \underline{G}_{(k-1)}). \tag{2.24}$$

where $\underline{G}_{(k-1)} = \underline{J}_{(k-1)}\underline{J}'_{(k-1)}$.

3. PCS and LFC

Since the asymptotic distribution of $\underline{W}_{\beta}^{(\alpha)}$ $(\alpha,\beta=1,2)$ is given by (2.23), probability of a correct selection for rule $R(\alpha,\beta,m)$ $(\alpha,\beta,m=1,2)$ is given as

$$\begin{aligned} P(CS|R(\alpha,\beta,m)) &= Pr(\underline{W}_{\beta}^{(\alpha)} \geq -\delta(\beta,m)\underline{J}_{(k-1)}) \\ &= Pr(\underline{U}_{\beta}^{(\alpha)} \leq (\tilde{\mu}_{\beta}^{(\alpha)} + \delta(\beta,m)\underline{J}_{(k-1)})/\sqrt{v_{\alpha}}) \end{aligned} \tag{3.1}$$

where

$$\delta(\beta,m) = \begin{cases} d\varrho/\sqrt{n} & \text{when } m = 1 \\ 0 & \text{when } m = 2 \end{cases} \tag{3.2}$$

$$\underline{U}_{\beta}^{(\alpha)} = (\underline{W}_{\beta}^{(\alpha)} - \underline{\eta}_{\beta}^{(\alpha)})/\sqrt{v_{\alpha}} \tag{3.3}$$

and

$$\underline{U}_{\beta}^{(\alpha)} \sim N(\underline{0}_{(k-1)}, \underline{E}_{(k-1)} + \underline{G}_{(k-1)}) \tag{3.4}$$

For the subset selection approach ($\alpha = 1$), since

$$\kappa_{\beta kj} - \kappa_{\beta ij} \geq 0$$

and

$$\kappa_{\beta kj} \geq 0$$

for large n, we have

$$\underline{\tilde{\eta}}_{\beta}^{(1)} \geq \underline{0}_{(k-1)}.$$

Also for indifference zone approach ($\alpha = 2$), taking the requirement

$$\psi_{\beta}(\theta_k, \theta_i) \geq c_{\beta} + \delta_{\beta}^*$$

in mind, we have

$$\underline{\tilde{\eta}}_{\beta}^{(2)} \geq \begin{cases} (k\int f^2(x)dx)\sqrt{n}\ \delta_1^* \underline{J} & \text{when } \beta = 1 \\ (k\int xf^2(x)dx)\ \sqrt{n}\ \dfrac{\delta_2^*}{1+\delta_2^*}\ \underline{J} & \text{when } \beta = 2 \end{cases}$$

Thus we have the following.

Theorem 3.1

Under the assumption of order restriction (2.8) and for large n, the LFC of the PCS for rules $R(\alpha,\beta,1)$ $(\alpha,\beta=1,2)$ are given when

$$\kappa_{\beta ki} = 0,\ i = 1,2,\ldots,k-1;\ \alpha,\beta = 1,2 \tag{3.5}$$

and for rules $R(\alpha,\beta,2)$ $(\alpha,\beta=1,2)$ are given when

$$\kappa_{\beta ki} = c_{\beta} + \delta_{\beta}^*,\ i = 1,2,\ldots,k-1;\ \alpha,\beta=1,2. \tag{3.6}$$

Under the LFC, $P(CS|R(\alpha,\beta,m))$ is evaluated as follows.

$$P(CS|R(\alpha,\beta,m)) \geq Pr(\underline{U}_{\beta}^{(\alpha)} \leq ((\gamma(\beta,m) + \delta(\beta,m))/\sqrt{v_{\alpha}})\underline{J}_{(k-1)}) \tag{3.7}$$

where v_{α} is given by(2.22), $\delta(\beta,m)$ is given by(3.2) and $\gamma(\beta,m)$ is defined as

$$\begin{cases} \gamma(\beta,1) = 0 \quad \text{for } \beta = 1,2 \\ \gamma(\beta,2) = \begin{cases} (k\int f^2(x)dx)\sqrt{n}\ \delta_1^* & \text{for } \beta = 1 \\ (k\int xf^2(x)dx)\sqrt{n}\ \dfrac{\delta_2^*}{1+\delta_2^*} & \text{for } \beta = 2. \end{cases} \end{cases} \tag{3.8}$$

By using the evaluation formula of the integral over the domain of the normal of the type (3.4), (see Gupta (1963)), we have the following reduced form of the expression (3.7).

$$P(CS|R(\alpha,\beta,m)) \geq \int \Phi^{k-1}\{x + ((\gamma(\beta,m) + \delta(\beta,m))/\sqrt{v_{\alpha}})\}\, d\Phi(x) \tag{3.9}$$

for $\alpha,\beta,m = 1,2$, where $\Phi(x)$ is the c.d.f. of Normal $N(0,1)$.

The (relative asymptotic) efficiency of two selection procedures R_1 and R_2 is considered in the following way. Let us define the efficiency of procedure R_2 relative to procedure R_1 be the ratio of sample sizes

$$Eff(R_1,R_2) = n_1/n_2 \tag{3.10}$$

where n_i satisfies

$$P(CS|R_i)_{LFC} = P^* \ , \ i = 1,2.$$

Then using the Theorem 3.1, we have the following.

$$\mathrm{Eff}(R(1,\beta,1),R(2,\beta,1)) = (1+1/k)(d_1/d_2)^2 \ , \ \beta = 1,2,$$

$$\mathrm{Eff}(R(\alpha,1,1),R(\alpha,2,1)) = (d_1/d_2)^2 \ , \qquad \alpha = 1,2,$$

$$\mathrm{Eff}(R(1,\beta,2),R(2,\beta,2)) = k/(k+1) \ , \qquad \beta = 1,2,$$

$$\mathrm{Eff}(R(\alpha,1,2),R(\alpha,2,2)) = (\delta_1{}^*\delta_2{}^*/(1+\delta_2{}^*))^2 \ (\int xf^2(x)dx/\int f^2(x)dx)^2 \ , \ \alpha = 1,2.$$

4. Appendix

Let us give the moments of combined ranks under the following population model.

Let k populations $\pi_1, \pi_2, \ldots, \pi_k$ be given. The c.d.f. of population π_s is denoted by $F_s(x)$ and is assumed to be continuous in $x(s=1,2, \ldots,k)$. Take n_s observations $X_{s1}, X_{s2}, \ldots, X_{sn_s}$ from population $\pi_s(s = 1,2. \ldots, k)$ and consider the combined (Wilcoxon type) rank R_{sj} of X_{sj} in such a way as we stated in Section 1. Then we have the following mean, variance and covariances of the rank R_{sj}.

Theorem 4.1

$$E(R_{sj}) = N\int GdF_s + \frac{1}{2} \tag{4.1}$$

$$V(R_{sj}) = 2N\int GdF_s - 2N\int F_sGdF_s + N^2\int G^2dF_s - N\int HdF_s - N^2(\int GdF_s)^2 - 1/12 \tag{4.2}$$

$$\mathrm{Cov}(R_{si},R_{sj}) = 3N\int GdF_s - 4N\int F_sGdF_s - \sum_{m=1}^{k} n_m(\int F_mdF_s)^2 - 1/12 \tag{4.3}$$

$$\begin{aligned}\mathrm{Cov}(R_{si},R_{tj'}) &= N(2 - \int F_tdF_s)\int GdF_t + N(2 - \int F_sdF_t)\int GdF_s \\ &\quad - \sum_{m=1}^{k} n_m\int F_mdF_s\int F_mdF_t - 2N\int F_tGdF_s - 2N\int F_sGdF_t \\ &\quad + \int F_sdF_t\int F_tdF_s + \int F_s{}^2dF_t + \int F_t{}^2dF_s - 1\end{aligned} \tag{4.4}$$

where $s,t = 1,2,\ldots,k$, $s\neq t$; $i,j = 1,2,\ldots,n_s$, $i\neq j$; $j' = 1,2,\ldots,n_t$ and

$$N = \sum_{m=1}^{k} n_m \tag{4.5}$$

$$G(x) = \frac{1}{N} \sum_{m=1}^{k} n_m F_m(x) \tag{4.6}$$

$$H(x) = \frac{1}{N} \sum_{m=1}^{k} n_m F_m^{\,2}(x) \tag{4.7}$$

Proof:

Let us give the sketch of proofs for (4.1) and (4.3) above. The remaining results are also obtained similarly.

Mean:

$$\begin{aligned} \Pr(R_{11} = s) = \sum_{A} \Pr(&a_1 \text{ of } X_1\text{'s}, a_2 \text{ of } X_2\text{'s},\ldots,a_k \text{ of } X_k\text{'s} \\ &\le X_1 \le (n_1-a_1-1) \text{ of } X_i\text{'s}, (n_2-a_2) \text{ of } X_2\text{'s},\ldots, (n_k-a_k) \\ &\text{of } X_k\text{'s}) \end{aligned} \tag{4.8}$$

where $a_i (i=1,2,\ldots,k)$ is an integer such that

$$\begin{cases} 0 \le a_1 \le n_1 - 1,\ 0 \le a_i \le n_i (i=2,3,\ldots,k) & (4.9) \\ \sum_{j=1}^{k} a_j = s - 1 & (4.10) \end{cases}$$

and "a_i of X_i's", "(n_i-a_i) of X_i's" should be read that a_i variables out of $(X_{i1}, X_{i2}, \ldots, X_{in_i})$ and remaining (n_i-a_i) variables, and so forth. Further, summation $\sum_{A}$ is taken for all tuples $(a_1,a_2,\ldots,a_k)$ of integers which satisfy the relations (4.9) and (4.10). From (4.8), we have

$$E(R_{11}) = \int \sum_{s=1}^{N} \sum_{A} s\binom{n_1-1}{a_1}\binom{n_2}{a_2}\cdots\binom{n_k}{a_k} F_1^{a_1} F_2^{a_2} \cdots F_k^{a_k}$$

$$\times (1-F_1)^{n_1-a_1-1}(1-F_2)^{n_2-a_2}\cdots(1-F_k)^{n_k-a_k} dF_1 \qquad (4.11)$$

By changing the order of summation, we first add for s, and we have

$$E(R_{11}) = \int \sum_{A_1} \binom{n_1-1}{a_1}\binom{n_2}{a_2}\cdots\binom{n_{k-1}}{a_{k-1}} F_1^{a_1} F_2^{a_2} \cdots F_{k-1}^{a_{k-1}}$$

$$\times (1-F_1)^{n_1-a_1-1}(1-F_2)^{n_2-a_2}\cdots(1-F_{k-1})^{n_{k-1}-a_{k-1}}(n_k F_k + \sum_{j=1}^{k-1} a_j + 1)dF_1,$$

where the summation $\sum_{A_1}$ is taken for all tuples $(a_1,a_2,\ldots,a_{k-1})$ of integers which satisfy the relation (4.9). Adding in turn for a_{k-1}, a_{k-2}, $\ldots$,a_1 we have the result for $E(R_{11})$.

Covariance:

For $s < t$, we have

$$\Pr(R_{11} = s,\ R_{21} = t) = \sum_{B} \Pr(a_1 \text{ of } X_1\text{'s},\ a_2 \text{ of } X_2\text{'s},\ \ldots,\ a_k \text{ of } X_k\text{'s}$$

$$\leq X_{11} \leq b_1 \text{ of } X_1\text{'s},\ b_2 \text{ of } X_2\text{'s},\ \ldots,\ b_k \text{ of } X_k\text{'s} \leq X_{21} \leq c_1 \text{ of } X_1\text{'s},\ c_2 \text{ of}$$

$$X_2\text{'s},\ \ldots\ c_k \text{ of } X_k\text{'s}) \qquad (4.12)$$

where a_i, b_i, $c_i (i = 1,2,\ldots,k)$ are integers such that

$$\begin{cases} a_i + b_i + c_i = \nu_i,\ i = 1,2,\ldots,k & (4.13) \\ \sum_{j=1}^{k} a_j = s - 1,\quad \sum_{j=1}^{k} b_j = t - s - 1,\quad \sum_{j=1}^{k} c_j = n - t & (4.14) \end{cases}$$

and $\nu_i = n_i - 1$ for $i = 1,2$, $\nu_i = n_i$ for $i = 3,4\ldots,k$.

Summation $\sum_{B}$ is taken for all tuples $(a_1,\ldots,a_k,b_1,\ldots,b_k,c_1,\ldots,c_k)$ which satisfy the relations (4.13) and (4.14). Then

$$I_1 \equiv \sum_{s<t} s\ t\ \Pr(R_{11} = s,\ R_{21} = t)$$

$$= \int\int_{x<y} \sum_{s<t} \sum_{B} \sum_{i=1}^{k} P_i(x,y)\,dF_1(x)\,dF_2(y) \qquad (4.15)$$

where

$$P_i(x,y) = \binom{\nu_i}{a_i,b_i,c_i} F_i^{a_i}(x)(F_i(y)-F_i(x))^{b_i}(1-F_i(y))^{c_i},\ i = 1,2,\ldots,k.$$

By changing the order of summation, we first add for s then for t and we have

$$I_1 = \int\int_{x<y} \sum_{s<t} \sum_{B_1} C_1 \prod_{i=1}^{k-1} P_i(x,y)\,dF_1(x)\,dF_2(y)$$

where

$$C_1 = \alpha_1 + \beta_1 \sum_{j=1}^{k-1} a_j + \gamma_1 \sum_{j=1}^{k-1} b_j + \left(\sum_{j=1}^{k-1} a_j\right)^2 + \left(\sum_{j=1}^{k-1} a_j\right)\left(\sum_{j=1}^{k-1} b_j\right)$$

and

$$\alpha_1 = n_k(n_k-1)F_k(x)F_k(y) + 3n_kF_k(x) + n_kF_k(y) + 2$$

$$\beta_1 = n_kF_k(x) + n_kF_k(y) + 3$$

$$\gamma_1 = n_kF_k(x) + 1.$$

Summation $\sum_{B_1}$ is taken for all tuples $(a_1,\ldots,a_{k-1},b_1,\ldots,b_{k-1},c_1,\ldots,c_{k-1})$ which satisfies the condition (4.13). By adding in turn for a set (a_i,b_i,c_i) $i = k-1, k-2,\ldots,1$, we have reduced form of I_1. By proceeding the similar steps for $\sum_{s>t} s\ t\ \Pr(R_{11} = s,\ R_{21} = t)$, we have the covariance relation for Cov (R_{11},R_{21}).

For rank sums

$$T_s = \sum_{j=1}^{n_s} R_{sj},\ s = 1,2,\ldots,k \qquad (4.16)$$

we have

$$E(T_s) = n_s E(R_{sj}) \ , \ s = 1,2,\ldots,k \tag{4.17}$$

$$\text{Cov}\ (T_s,T_t) = n_s n_t\ \text{Cov}\ (R_{sj},R_{tj'}),\ s,t = 1,2,\ldots,k,\ s \neq k \tag{4.18}$$

and for variance

$$\begin{aligned} V(T_s) &= \sum_{j=1}^{n_s} V(R_{sj}) + \sum_{i\neq j}^{n_s} \text{Cov}\ (R_{si},R_{sj}) \\ &= Nn_s(3n_s-1)\int G dF_s - 2Nn_s(2n_s-1)\int F_s G dF_s \\ &\quad + N^2 n_s \int G^2 dF_s - Nn_s \int H dF_s - N_2 n_s (\int G dF_s)^2 \\ &\quad - n_s(n_s-1)\sum_{m=1}^{k} n_m (\int F_m dF_s)^2 - n_s{}^2/12 \end{aligned} \tag{4.19}$$

Especially if $F_i(x) = F(x)$ for all i, then we have

$$E(T_s) = n_s(N+1)/2 \tag{4.20}$$

$$V(T_s) = n_s(N-n_s)(N+1)/12 \tag{4.21}$$

$$\text{Cov}\ (T_s,T_t) = -n_s n_t(N+1)/12 \tag{4.22}$$

Also for k=2, we have the following

$$E(T_i) = n_i(n_i+1)/2 + n_i n_j \int F_j dF_i,\ i,\ j = 1,2;\ j \neq i \tag{4.23}$$

$$V(T_i) = n_i n_j (2n_i - 1)\int F_j dF_i + n_i n_j (n_j - 1)\int F_j^2 dF_i$$

$$+ n_i n_j (n_i - 1)\int F_i^2 dF_j - n_i n_j (n_i + n_j - 1)\left(\int F_j^2 dF_i\right)^2 - n_i n_j (n_i - 1)$$

$$i,j = 1,2;\ i \neq j \qquad (4.24)$$

$$\text{Cov}\,(T_1,T_2) = n_1 n_2 [n_1 \int F_1 dF_2 + n_2 \int F_2 dF_1 - (n_1 + n_2 - 1)\int F_1 dF_2 \int F_2 dF_1$$

$$- (n_1 - 1)\int F_1^2 dF_2 - (n_2 - 1)\int F_2^2 dF_1 - 1] \qquad (4.25)$$

Finally we give a property which lies between ranks and distributions (parameters). Let $F_i(x)$'s be stochastically increasing family of distribution specified by parameter θ_i. Then we have the following.

Theorem 4.2

$E(R_s) \geq E(R_t)$ if and only if $F_s \leq F_t$ where $s,t = 1,2,\ldots,k$.

References

Alam, K. and Thompson, J. R. (1971). A selection procedure based on ranks Ann. Inst. Statist. Math., 23, 253-262.

Bartlett, N. S. and Govindarajule, Z. (1968). Some distribution-free statistics and their application to the selection problem, Ann. Inst. Statist. Math., 20, 79-97.

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

Gupta, S. S. (1963). Probability integrals of multivariate normal and multivariate t, Ann Math. Statist., 34, 792-828.

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

Gupta, S. S. and McDonald, G. C. (1980). Nonparametric procedures in multiple decisions (Ranking and selection procedures), Mimeo Ser. No. 80-13, Dept. of Statistics, Purdue University, West Lafayette, Indiana.

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

Gupta, S. S. and Panchapekasan, S. (1984). Subset selection procedures: Review and an assessment, Tech. Rep. No. 84-4, Dept. of Statistics, Purdue University, West Lafayette, Indiana.

Lee, Y. J. and Dudewicz, E. J. (1974). Nonparametric ranking and selection procedures. Tech. Rep. No. 105, Dept. of Statistics, The Ohio State University, Columbus, Ohio.

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

Matsui, T. (1974). Asymptotic behavior of a selection procedure based on rank sums, J. Japan Statist. Soc., 4, 57-64.

Matsui, T. (1985). Moments of rank vector with applications to selection and ranking. J. Japan Statist. Soc. 15, 17-25.

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

McDonald, G. C. (1973). The distribution of some rank statistics with applications in block design selection problems, Sankhya, Ser. A, 35, 187-204.

Puri, M L. and Puri, P. S. (1969). Multiple decision procedures based on ranks for certain problems in analysis of variance, Ann. Math. Statist., 40, 619-632.

Puri, P. S. and Puri, M. L. (1968). Selection procedures based on ranks: scale parameter case. Sankhya, Ser. A, 30, 291-302.

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

AD-A159154

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: Technical Report #85-18	2. GOVT ACCESSION NO
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): On the Least Favorable Configuration of a Selection Procedure Based on Ranks	5. TYPE OF REPORT & PERIOD COVERED: Technical
6. PERFORMING ORG. REPORT NUMBER: Technical Report #85-19	
7. AUTHOR(s): Shanti S. Gupta and Takashi Matsui	8. CONTRACT OR GRANT NUMBER(s): N00014-84-C-0167
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Purdue University, Department of Statistics, West Lafayette, IN 47907	10. PROGRAM ELEMENT, PROJECT, TASK, AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: Office of Naval Research, Washington, DC	12. REPORT DATE: August 1985
	13. NUMBER OF PAGES: 20
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)

Selection Procedures, Combined Ranks, Least Favorable Configuration, Moments of Ranks, Wildoxon Type, Friedman Type

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

Let us consider two types of statistics based on the sums of combined (Wilcoxon type) ranks and vector (Friedman type) ranks. Underlying populations are supposed to belong to the location or scale parameter family of distributions.

Two approaches - subset selection and indifference zone - of ranking and selection procedures based on these statistics are considered in an asymptotic framework for selecting the population with the largest parameter value. The least favorable configurations of parameters are discussed by computing the

DD FORM 1 JAN 73 1473

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

exact moments of these statistics and introducing an assumption of order relation between the gaps of parameters.

END

FILMED

11-85

DTIC