DEPARTMENT OF STATISTICS

The Ohio State University

OSU

12

ROBUST SELECTION PROCEDURES

BASED ON VECTOR RANKS*

by

Young Jack Lee

and

Edward J. Dudewicz

Technical Report No. 233
Department of Statistics
The Ohio State University
Columbus, Ohio 43210

April 1981

* Supported in part by the U. S. Army Research Office - Durham, and by Office of Naval Research Contract No. N00014-78-C-0543.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

14 | TR-233

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 9 Technical Rep[redacted]t, No. 233	2. GOVT ACCESSION NO. AD-A099398	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) 6 ROBUST SELECTION PROCEDURES BASED ON VECTOR RANKS.		5. TYPE OF REPORT & PERIOD COVERED
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) 10 Young Jack Lee, Edward J. Dudewicz	15	8. CONTRACT OR GRANT NUMBER(s) N00014-78-C-0543
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics, The Ohio State University, Columbus, Ohio 43210		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 042-403
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research, Department of The Navy, Arlington, Virginia 22217		12. REPORT DATE 11 Apr[redacted]81
		13. NUMBER OF PAGES iii + 35
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) 12 39		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)

Key Words and Phrases: Robust selection procedures, single-stage rule, block designs, asymptotic relative efficiency, means procedure, indifference-zone approach, counterexamples, least favorable configuration, large sample approximation.

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

Consider n blocks of k observations $(X_{1j},\ldots,X_{kj})$, $j=1,\ldots,n$. Suppose X_{ij} are independent and $P(X_{ij} \le x) = F(x - \eta_j - \theta_i)$ where η_j is the nuisance location parameter of the j^{th} block and θ_i is the location parameter corresponding to population π_i $(1 \le j \le n,\ \ 1 \le i \le k)$. We are interested in selecting

DD FORM 1473 1 JAN 73

Unclassified 406331

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. (continued)

populations associated with large location parameter θ. To this end compute $H_1 = \sum_{j=1}^{n} R_{ij}$ where $R_{ij} = [\#$ of $X_{i'j} \leq X_{ij}$ $(i \leq i' \leq k)]$, and $\overline{X}_i = n^{-1} \sum_{j=1}^{n} X_{ij}$, and base the terminal statistical decision on $\overline{X}_1,\ldots,\overline{X}_k$ (means procedure $\mathcal{R}_{MP}$) or $H_1,\ldots,H_k$ (vector rank procedure $\mathcal{R}_V$). Fix t $(1 \leq t < k)$ and consider the problem of selecting populations associated with the t largest θ's based on $\overline{X}_1,\ldots,\overline{X}_k$ or $H_1,\ldots,H_k$. In this paper we investigate large sample behavior (as well as some fixed sample behaviour) of $\mathcal{R}_V$. The asymptotic relative efficiency of $\mathcal{R}_V$ with respect to $\mathcal{R}_{MP}$ is also studied.

ROBUST SELECTION PROCEDURES BASED ON VECTOR RANKS*

Young Jack Lee and Edward J. Dudewicz

National Institutes of Health,
Bethesda, Maryland
and
The Ohio State University,
Columbus, Ohio

0. Summary

Consider n blocks of k observations $(X_{1j},\ldots,X_{kj})$, $j=1,\ldots,n$. Suppose X_{ij} are independent and $P(X_{ij}\le x) = F(x-\eta_j-\theta_i)$ where η_j is the nuisance location parameter of the j^{th} block and θ_i is the location parameter corresponding to population π_i $(1\le j\le n,\ 1\le i\le k)$.

We are interested in selecting populations associated with large location parameter θ. To this end compute $H_1 = \sum_{j=1}^{n} R_{ij}$ where $R_{ij} = [\#\text{ of } X_{i'j}\le X_{ij}\ (1\le i'\le k)]$, and $\bar{X}_i = n^{-1}\sum_{j=1}^{n} X_{ij}$, and base the terminal statistical decision on: $\bar{X}_1,\ldots,\bar{X}_k$ (<u>means</u> <u>procedure</u> $\underline{P_{MP}}$) or $H_1,\ldots,H_k$ (<u>vector</u> <u>rank</u> <u>procedure</u> $\underline{P_V}$). Fix t $(1\le t<k)$ and consider the problem of selecting populations associated with the t largest θ's based on: $\bar{X}_1,\ldots,\bar{X}_k$ or $H_1,\ldots,H_k$.

* This research was supported in part by the U. S. Army Research Office-Durham, and by Office of Naval Research Contract No. N00014-78-C-0543.

Key words and phrases. Robust selection procedures, single-stage rule, block designs, asymptotic relative efficiency, means procedure, indifference-zone approach, counterexamples, least favorable configuration, large sample approximation.

In this paper we investigate large sample behavior (as well as some fixed sample behavior) of P_V. The asymptotic relative efficiency of P_V with respect to P_{MP} is also studied.

1. Introduction

Let X_{ij} $(j = 1,2,\ldots,n_i;\ i = 1,2,\ldots,k)$ be independent random samples drawn from populations $\pi_1,\pi_2,\ldots,\pi_k$ with absolutely continuous distribution functions (df's) $F(x-\theta_i)$. Let $\theta_{[1]} \le \ldots \le \theta_{[k]}$ denote the ordered values of the unknown θ_i, and let $\pi_{(i)}$ denote the population associated with $\theta_{[i]}$; these associations are assumed completely unknown. Often for some fixed t $(1 \le t < k)$ an experimenter is interested in the problem of selecting the "so-called" t best populations, $\pi_{(k-t+1)},\ldots,\pi_{(k)}$. For the selection of the t best populations, Bechhofer (1954) proposed the means procedure (denoted by P_{MP}) which selects, as being the t best populations, the t populations yielding the t highest sample means $\bar{X}_i (= n_i^{-1} \sum_{j=1}^{n_i} X_{ij})$: Bechhofer requires that the probability that the so-selected t populations are the t best [when this occurs, a Correct Selection (CS) is said to occur] be at least P^* (a prespecified constant between $\binom{k}{t}^{-1}$ and 1) whenever $\theta_{[k-t+1]} - \theta_{[k-t]} \ge \delta^*$ (δ^* is a prespecified positive constant). A different procedure was proposed by Gupta (1956, 1965): rather than selecting the t populations associated with the t highest sample means, he selects a subset of the k populations (retaining in the selected subset all the populations yielding sample means close to the t highest sample means) and requires that the probability be at least P^* that the selected subset contains the t best (when this occurs, a CS is said to occur). Both Bechhofer and Gupta considered the case of normal distributions with common known variances; for the case of normal distributions but with (possibly different) unknown variances the reader is referred to Dudewicz and Dalal (1976). The robustness of the means procedure is broached in Lehmann (1963) and is under investigation, in a more general context, by one of the authors [YJL].

Lehmann (1963) and Bartlett and Govindarajulu (1968) based selection procedures on the joint ranks of the observations in the combined sample of $N = \sum n_i$ observations. Specifically, each observation X_{ij} is assigned a score $a_{ij} = E[Z(R_{ij})|G]$ where $Z(1) < \ldots < Z(N)$ denotes an ordered sample from any continuous df G and R_{ij} denotes the rank of X_{ij} in the combined sample. The selection procedures are then based on the quantities $n_i^{-1} \sum_j a_{ij}$ $(1 \le i \le k)$. Lehmann's approach uses a Bechhofer-type (indifference-zone) approach while Bartlett and Govindarajulu use a Gupta-type (subset-selection) approach. Bartlett and Govindarajulu also base some selection procedures on randomized scores (i.e., quantities $n_i^{-1} \sum_j Z(R_{ij})$ $(1 \le i \le k)$); but we have shown (details will not be given here) that in selection procedures based on randomized scores the probability of CS (denoted by P[CS]) is bounded away from 1 for any configuration of parameters and two different statisticians reach, with positive probability, two different conclusions from the same set of observations. This extends results of Jogdeo (1966) to ranking and selection problems. An extensive review of other selection procedures (including joint rank procedures) is provided in Lee and Dudewicz (1974).

The model usually assumed in the literature is that of the one-way analysis-of-variance model. The selection procedure investigated in this paper arises from the two-way analysis-of-variance type model where block effect enters: namely $P(X_{ij} \le x) = F(x - \eta_j - \theta_i)$ where η_j is a nuisance location parameter of the j^{th} block $(1 \le j \le n)$. In this case ranks within each block are preferable to joint ranks. McDonald (1972, 1973) makes subset-selection approaches to a selection problem by basing terminal decision rules on ranks within each block, and Dudewicz and Fan (1973) suggested an indifference-zone approach. In Section 2 we investigate, by an indifference-zone approach, selection procedures based on ranks within each block (we denote this procedure by P_V) under the <u>slipped</u> <u>parameter</u> <u>configuration</u> (SPC) $\theta_{[1]} = \cdots = \theta_{[k-t]} < \theta_{[k-t+1]} = \cdots = \theta_{[k]}$; all the results are asymptotic. In Section 3, we investigate the asymptotic relative efficiency (ARE) of P_V with respect to P_{MP} as $\theta_{[k-t+1]} - \theta_{[k-t]}$ tends to zero under the SPC assumption. The configuration of θ_i's minimizing $P[CS|P_V]$ is investigated in Section 4; in particular we show that the SPC is not necessarily least-favorable. In Section 5 we discuss practicality of the assumption of the SPC as an underlying configuration. In this article we denote by $O(a)$ a positive quantity such that $a^{-1}O(a)$ converges to a positive constant in the limit of a.

2. $P[CS|P_V]$ under the SPC: Asymptotic Results

We make the following probability requirement for given δ^* and P^* ($\delta^* > 0$, $\binom{k}{t}^{-1} < P^* < 1$):

(2.1) Probability requirement: We select the populations $\pi_{(k-t+1)},\ldots,\pi_{(k)}$ (i.e., we make a CS) with probability $P[CS] \geq P^*$ whenever $\theta_{[k-t+1]} - \theta_{[k-t]} \geq \delta^*$.

Consider the following single-stage procedure: Take n independent vectors $\underset{\sim}{X}_j = (X_{1j},\ldots,X_{kj})$ $(1 \leq j \leq n)$ (X_{ij} denotes the j^{th} observation from π_i); compute $H_i = \sum_{j=1}^{n} R_{ij}$ $(1 \leq i \leq k)$ where $R_{ij} = \{\#\text{ of } X_{i'j} \leq X_{ij}\ (1 \leq i' \leq k)\}$; and select (as being the t best populations) the populations associated with the t highest H_i's (breaking ties, if any, by randomization).

We first consider $t = 1$ and then generalize to $t \geq 1$. Let

$$\Omega_0 = \{\vec{\theta} \in R_k;\ \ \vec{\theta} = (\theta_{[1]},\ldots,\theta_{[k]})\},$$

$$\Omega_\theta(\delta^*,t) = \{\vec{\theta} \in \Omega_0:\ \theta_{[k-t+1]} - \theta_{[k-t]} \geq \delta^*\},$$

$$\omega_\theta(\delta^*,t) = \left\{ \vec{\theta} \in \Omega_0 : \theta_{[1]} = \cdots = \theta_{[k-t]} \le \theta_{[k-t+1]} - \delta^*, \atop \theta_{[k-t+1]} = \cdots = \theta_{[k]}, \quad \delta^* > 0 \right\},$$

and

$$\vec{\theta}_0(t): \theta_{[1]} = \cdots = \theta_{[k-t]} = \theta_{[k-t+1]} - \delta^*, \quad \theta_{[k-t+1]} = \cdots = \theta_{[k]}.$$

Lemma 2.1: For selection of $\pi_{(k)}$ under $\omega_\theta(\delta^*,1)$, $P[CS|P_V]$ is a nondecreasing function of $\theta_{[k]}$. Hence $\inf\limits_{\omega_\theta(\delta^*,1)} P[CS|P_V] = P[CS|P_V,\vec{\theta}_0(1)]$.

Proof

See Theorem 3.1 of McDonald (1972).■

Now we wish to determine a sample size n_θ which will guarantee $P[CS|P_V, \vec{\theta} \in \omega_\theta(\delta^*,1)]$ to be at least P^* for given $\delta^* > 0$, but we do not know how to determine the sample size for given P^* and δ^*. Rather, we find δ^* for given P^* and sample size n, namely we put δ^* as a function of n and P^* and then solve n for given δ^* and P^*. (This method was introduced by Lehmann (1963).) To this end we need to investigate the asymptotic determination of $P[CS|P_V]$ under the following configuration with $t = 1$:

$$(2.2) \quad \vec{\theta}_0(t,n): \theta_{[1]} = \cdots = \theta_{[k-t]}, \quad \theta_{[k-t+1]} - \theta_{[k-t]} = \delta(n),$$
$$\theta_{[k-t+1]} = \cdots = \theta_{[k]}.$$

Let $H_{(i)}$ be the sum of rank scores yielded by $\pi_{(i)}$. To show the dependence of $H_{(i)}$ on n, we write $H_{(i)}(n)$, and for the notational convenience, without loss of generality, we let

$\theta_{[i]} = \theta_i$, $\pi_{(i)} = \pi_i$, and thus $H_{(i)}(n) = H_i(n)$ $(1 \le i \le k)$. For large samples, since $\lim_{n\to\infty} P[H_k(n) - H_i(n) = 0,\ i \le k-1 | \vec{\theta}_0(n)] = 0$, we drop the randomization part of $P[CS|P_V]$. Thus

$$P[CS|P_V, \vec{\theta}_0(1,n)] \doteq P[H_k(n) - H_i(n) > 0,\ 1 \le i \le k-1 \mid \vec{\theta}_0(1,n)]$$

(2.3)

$$= P[\tfrac{1}{\sqrt{n}}(H_k(n) - H_i(n)) > 0,\ 1 \le i \le k-1 \mid \vec{\theta}_0(1,n)]$$

($A(x) \doteq B(x)$ means $|A(x) - B(x)| \to 0$ as x approaches a limit). We will approximate (2.3), and in the sequel we need the following.

Lemma 2.2: Let $G(x)$ be an absolutely continuous function possessing a quadratically integrable derivative $G'(x)$ and let $F(x)$ be an absolutely continuous df with a pdf $f(x)$. If $\int H^2(x) f^2(x)dx < \infty$, then

$$\lim_{h\to 0} \left| \int \frac{G(x+h) - G(x)}{h} H(x)dF(x) - \int G'(x)H(x)dF(x) \right| = 0 .$$

(For the special case $H(x) \equiv 1$ and $F(x) \equiv G(x)$ a.e., this is Lemma 3.4 of Mehra and Sarangi (1967).)

Proof

See Lemma 3.4 of Mehra and Sarangi (1967).■

Let

$$\delta = \lim_{n\to\infty} n^{1/2}\delta(n), \quad \delta > 0 \text{ fixed}$$

(the cases $n^{1/2}\delta(n) \to \infty$ and $n^{1/2}\delta(n) \to 0$, as $n \to \infty$, are covered after Theorem 2.5), and assume

(2.4) $\int f^2(x)dx < \infty$ ($f(x) \equiv$ pdf of the underlying df F).

(For some pdf's $\int f^2(x)dx$ does not exist. Df's satisfying (2.4) are characterized by Lemma 1.4.1 of Kagan, Linnik and Rao (1973).)

Lemma 2.3: Under the configuration of $\vec{\theta}_0(1,n)$, defining

$$V_i(n) = n^{-\frac{1}{2}}(H_k(n) - H_i(n)) \quad (1 \leq i \leq k-1),$$

we find that

$$\text{(2.5)} \quad \lim_{n\to\infty} E[V_i(n)] = \delta k \int f^2(x)dx \quad (1 \leq i \leq k-1),$$

$$\text{(2.6)} \quad \lim_{n\to\infty} \mathrm{Var}[V_i(n)] = k(k+1)/6 \quad (1 \leq i \leq k-1),$$

and

$$\text{(2.7)} \quad \lim_{n\to\infty} \mathrm{Cov}[V_i(n), V_{i'}(n)] = k(k+1)/12 \quad (1 \leq i \neq i' \leq k-1).$$

Proof

Defining

$$P_r^{(i)} = P[X_{i1} \text{ has the } r^{th} \text{ rank among } X_{11},\ldots,X_{k1} | \vec{\theta}_0(1,n)],$$

we have

$$\text{(2.8)} \quad E[V_i(n)] = n^{-\frac{1}{2}} \sum_{j=1}^{n} E(R_{kj} - R_{ij}) = n^{\frac{1}{2}} \sum_{r=1}^{k} r(P_r^{(k)} - P_r^{(i)}).$$

Let $\theta_1 = \ldots = \theta_{k-1} = \theta$ and $\theta_k = \theta + \delta(n)$.

$$P_r^{(k)} = \binom{k-1}{r-1} \int F^{r-1}(x-\theta)[1 - F(x-\theta)]^{k-r} dF(x-\theta-\delta(n))$$

$$= \binom{k-1}{r-1} \int F^{r-1}(x)[1 - F(x)]^{k-r} dF(x-\delta(n)).$$

Note that we do not lose any generality by letting $\theta = 0$.

Now

$$P_r^{(k)} = \binom{k-2}{r-2} \int F^{r-1}(x+\delta(n))[1 - F(x+\delta(n))]^{k-r} dF(x)$$

$$+ \binom{k-2}{r-1} \int F^{r-1}(x+\delta(n))[1 - F(x+\delta(n))]^{k-r} dF(x),$$

$$P_r^{(i)} = \binom{k-2}{r-2} \int F^{r-2}(x)F(x-\delta(n))[1 - F(x)]^{k-r} dF(x)$$

$$+ \binom{k-2}{r-1} \int F^{r-1}(x)[1 - F(x)]^{k-r-1}[1 - F(x-\delta(n))] dF(x),$$

and combining $P_r^{(k)}$ and $P_r^{(i)}$ and taking the limit yields

$$(2.9) \quad \lim_{n\to\infty} n^{\frac{1}{2}}(P_r^{(k)} - P_r^{(i)}) = \binom{k-1}{r-1}\delta \int \{\frac{d}{dx}F^{r-1}(x)[1-F(x)]^{k-r}\} dF(x)$$

$$+ \binom{k-2}{r-2}\delta \int F^{r-2}(x)[1-F(x)]^{k-r} f^2(x)dx$$

$$- \binom{k-2}{r-1}\delta \int F^{r-1}(x)[1-F(x)]^{k-r-1} f^2(x)dx,$$

hence (2.5) is obtained from (2.8). Define

$$P_{\ell,q}^{(i,j)} = P\left[\begin{matrix} X_{i1} \text{ has the } \ell^{th} \text{ rank, and } X_{j1} \text{ has the} \\ q^{th} \text{ rank among } X_{11},\ldots,X_{k1} \end{matrix} \;\middle|\; \vec{\theta}_0(1,n)\right].$$

Note that $P^{(i,j)}_{\ell,\ell} = 0$ $(i \neq j,\ 1 \leq \ell \leq k)$.

Then

$$\text{(2.10)} \qquad \lim_{n\to\infty} P^{(i)}_r = \frac{1}{k}, \quad \text{and} \quad \lim_{n\to\infty} P^{(i,j)}_{\ell,q} = \frac{1}{k(k-1)} .$$

(For details see Lee and Dudewicz (1974).) From (2.10), a computation shows (2.6) and (2.7). ∎

Thus we have obtained asymptotic moments of $V_i(n)$ $(1 \leq i \leq k-1)$. To evaluate $P[CS|P_V,\vec{\theta}_0(1,n)]$, we need to obtain an asymptotic distribution of $V(n) = (V_1(n),\ldots,V_{k-1}(n))'$. We can show that any linear combination of $(V_1(n),\ldots,V_{k-1}(n))'$ has an asymptotic normal distribution by a Lindeberg-Feller type central limit theorem (specifically see §26 of Gnedenko and Kolmogorov (1949)). Thus $V(n)$ has an asymptotic $(k-1)$-variate normal distribution with certain known mean and variance-covariance. The following lemma is proven in Lee and Dudewicz (1974).

Lemma 2.4: The $(k-1)$-variate random vector $(V_1(n),\ldots,V_{k-1}(n))'$ has an asymptotic $(k-1)$-variate normal distribution with mean $\delta k \int f^2(x)dx$ and variance-covariance

$$\sigma_{ij} = k(k+1)(1+\delta_{ij})/12$$

where $\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases}$

Now we are prepared to approximate $P[CS|P_V,\vec{\theta}_0(1,n)]$. Let $(U_1,\ldots,U_{k-1})'$ be a $(k-1)$-variate normal random vector satisfying $E(U_i) = 0$, $\sigma_{ij} = (1+\delta_{ij})/2$. Then

$$P[CS|P_V,\vec{\theta}_0(1,n)]$$

$$\doteq P\{[k(k+1)/6]^{-\frac{1}{2}}[V_i(n) - \delta k\int f^2(x)dx] > -[k(k+1)/6]^{-\frac{1}{2}}\delta k\int f^2(x)dx,\quad 1\le i\le k-1|\vec{\theta}_0(1,n)\}$$

$$\doteq P\{U_i > -[k(k+1)/6]^{-\frac{1}{2}}\delta k\int f^2(x)dx,\quad 1\le i\le k-1\}.$$

Letting $\Delta > 0$ satisfy

$$P(U_i > -\Delta,\quad 1\le i\le k-1) = P^*, \tag{2.11}$$

we find

Theorem 2.5: For $P[CS|P_V,\vec{\theta}_0(1,n)]$ to be asymptotically P^* $(1/k < P^* < 1)$, $\delta(n)$ should satisfy

$$\lim_{n\to\infty} n^{\frac{1}{2}}\delta(n) = [k(k+1)/6]^{\frac{1}{2}}(k\int f^2(x)dx)^{-1}\Delta .$$

We make several remarks on implications of Theorem 2.5:

(i) if $\lim_{n\to\infty} n^{\frac{1}{2}}\delta(n) = \infty$, then $\lim_{n\to\infty} P[CS|P_V,\vec{\theta}_0(1,n)] = 1$, and if $\lim_{n\to\infty} n^{\frac{1}{2}}\delta(n) = 0$, then $\lim_{n\to\infty} P[CS|P_V,\vec{\theta}_0(1,n)] = 1/k$, that is if $\delta(n) \ne O(n^{-\frac{1}{2}})$, then $P[CS|P_V,\vec{\theta}_0(1,n)]$ converges either to 1 or $1/k$, in which case we cannot relate n and δ^* (as $\delta^* \to 0$) for fixed P^* (for the cases of

P_{MP} and joint rank procedures (i) is implicit in Lehmann (1963));

(ii) when $\delta(n) = O(n^{-\frac{1}{2}})$, we can relate n and δ^* for given P^* via ($n_A(P_V) \equiv$ approximated n)

$$n_A(P_V) = (\Delta/\delta^*)^2[k(k+1)/6][k\int f^2(x)dx]^{-2}; \tag{2.12}$$

(iii) consider the question how good $n_A(P_V)$ is; namely letting n_{TRUE} denote the sample size which will guarantee $P[CS|P_V,\vec{\theta}_0(1)] \geq P^*$, does $n_A(P_V)/n_{TRUE}$ converge to 1 as $\delta^* \to 0$?; the answer is conjectured to be affirmative (see Lee and Dudewicz (1974)); and

(iv) the conjecture of (iii) justifies, in part, dropping the randomization part of $P[CS|P_V,\vec{\theta}_0(1,n)]$ as we did earlier. [Such dropping has been done without justification in the literature, e.g., p. 270 of Lehmann (1963), p. 295 of Puri and Puri (1968), p. 623 of Puri and Puri (1969), p. 377 of Bhapkar and Gore (1971), and p. 258 of Alam and Thompson (1971) among others.]

The P_{MP} version of (2.12) is due to Lehmann (1963). Namely, let m be the sample size for $P[CS|P_{MP},\vec{\theta}_0(1,m)] = P^*$ asymptotically. Then

$$m = 2(\Delta\sigma/\delta^*)^2 \tag{2.13}$$

where σ^2 is the variance of the underlying df, Δ satisfies (2.11), and $\delta^* = \theta_{[k]} - \theta_{[k-1]}$. [(2.13) is the equation (11) of Lehmann (1963).] Note that when the underlying df is normal, (2.13) is the sample size obtained by Bechhofer (1954), and is thus exact.

The results in this section apply so far only to the selection of the best population under $\vec{\theta}_0(1)$, but can be extended to the t best selection problem under configuration $\vec{\theta}_0(t)$. The proof is, of course, more complicated so we will state the results corresponding to Lemma 2.4 and Theorem 2.5 and refer to Lee and Dudewicz (1974) for proofs. We have

(2.14) $$P[CS|P_V,\vec{\theta}_0(t,n)]$$

$$\doteq P[H_\ell(n) - H_i(n) > 0, \quad k-t+1 \le \ell \le k, \quad 1 \le i \le k-t]$$

$$= P[V_{\ell i}(n) > 0, \quad k-t+1 \le \ell \le k, \quad 1 \le i \le k-t]$$

where

$$V_{\ell i}(n) = n^{-\frac{1}{2}}(H_\ell(n) - H_i(n)) \quad (k-t+1 \le \ell \le k, \quad 1 \le i \le k-t).$$

Lemma 2.6: $(V_{k-t+1,1}(n),\ldots,V_{k-t+1,k-t}(n),\ldots,V_{k,k-t}(n))'$ is a $t(k-t)$-variate random vector the limiting distribution of which, under the configuration $\vec{\theta}_0(t,n)$, is the distribution of a $t(k-t)$-variate normal vector $(U_{\ell i};\ k-t+1 \le \ell \le k,\ 1 \le i \le k-t)'$ with $E(U_{\ell i}) = \delta k \int f^2(x)dx$, $Var(U_{\ell i}) = k(k+1)/6$, $Corr(U_{\ell i},U_{\ell i'}) = Corr(U_{\ell i},U_{\ell' i}) = 1/2$, $Corr(U_{\ell i},U_{\ell' i'}) = 0$, $(k-t+1 \le \ell \ne \ell' \le k,$ $1 \le i \ne i' \le k-t)$, where $\delta = \lim_{n\to\infty} n^{\frac{1}{2}}\delta(n)$.

Let $(U_1,\ldots,U_{k-t},W_{k-t+1},\ldots,W_{k-1})'$ be a $(k-1)$-variate normal random vector with $E(U_i) = E(W_j) = 0$, $Var(U_i) = Var(W_j) = 1$, $Corr(U_i,U_{i'}) = Corr(W_j,W_{j'}) = 1/2$, $Corr(U_i,W_j) = -1/2$ $(1 \le i \ne i' \le k-t,\ \ k-t+1 \le j \ne j' \le k)$, and let $\Delta_t > 0$ satisfy

$$(2.15)\qquad P^* = tP[U_i > -\Delta_t,\ \ W_j > 0,\ \ 1 \le i \le k-t,\ \ k-t+1 \le j \le k].$$

Theorem 2.7: For (2.14) to be P^* $(1/\binom{k}{t} < P^* < 1)$ asymptotically under $\vec{\theta}_0(t,n)$, $\delta(n)$ should satisfy

$$(2.16)\qquad \delta = \lim_{n\to\infty} n^{\frac{1}{2}}\delta(n) = [k(k+1)/6]^{\frac{1}{2}}[k\int f^2(x)dx]^{-1}\Delta_t.$$

The implications of Theorem 2.7 are the same as those of Theorem 2.5. Therefore through Theorem 2.7, we can relate n to δ^* and P^* by

$$(2.17)\qquad n_A(P_V) = (\Delta_t/\delta^*)^2(k(k+1)/6)[k\int f^2(x)dx]^{-2}$$

(note the difference between Δ and Δ_t in (2.12) and (2.17)).

The P_{MP} equivalent of (2.16) is due to Puri and Puri (1969) and is

$$(2.18)\qquad m = 2(\Delta_t\sigma/\delta^*)^2$$

where m is the sample size for $P[CS|P_{MP},\vec{\theta}_0(t,m)]$ to be P^*, σ^2 is the variance of the underlying df, and Δ_t satisfies (2.15). (2.18) is the equation (4A.11) of Puri and Puri (1969).

In this section we have studied $P[CS|P_V]$ under the SPC; namely how to relate the necessary sample size to the minimum discrepancy worth detecting and the required P[CS] for large sample size under the assumption that the underlying df is known. Similar results for joint rank procedures were obtained by Lehmann (1963) and Puri and Puri (1969).

3. ARE of P_V under the SPC.

Suppose there are two different selection procedures P_1 and P_2 with the same probability requirement. We define an asymptotic relative efficiency (ARE) of P_1 with respect to P_2 as

$$\text{(3.1)} \qquad \mathrm{ARE}(P_1, P_2) = \lim_{\delta^* \to 0} \left(\frac{\text{Sample size for } P_2}{\text{Sample size for } P_1} \right).$$

To determine the ARE this way, we should be able to determine a sample size for given δ^* and P^*. We noted that when we let $\delta^* = \delta(n)$, and $\lim_{n\to\infty} n^{\frac{1}{2}}\delta(n) = c$ (an appropriate constant), P[CS] converges to P^* as $n \to \infty$. In other words, by letting $n = n(\delta^*)$ and requiring $\lim_{\delta^*\to 0} [n(\delta^*)]^{\frac{1}{2}}\delta^*$ to converge to some constant, P[CS] converges to P^* as $\delta^* \to 0$. Thus letting $n_{P_i}(\delta^*)$ $(i = 1,2)$ (the selection sample size for P_i for a given δ^*) satisfy $\lim_{\delta^*\to 0} [n_{P_i}(\delta^*)]^{\frac{1}{2}}\delta^* = c_i$ $(i = 1,2)$, we can determine the $\mathrm{ARE}(P_1,P_2)$ (as $\delta^* \to 0$). One may suspect that $\mathrm{ARE}(P_1,P_2)$ (as $\delta^* \to 0$) and $\mathrm{ARE}(P_1,P_2)$ (as $n_{P_2} \to \infty$) may be different. [Note that the latter quantity was used by Lehmann (1963) to compute the ARE of a joint rank procedure with respect to P_{MP}.] However we can show their equivalence as follows. If for given P^* and n_{P_i} $(i=1,2)$ $\delta(n_{P_i})$ is determined so that $\lim_{n_{P_i}} n_{P_i}^{\frac{1}{2}}\delta(n_{P_i}) = c_i$ $(i=1,2)$, then $P[CS|P_i] \doteq P^*$. But since P_1 and P_2 are required to satisfy the same probability requirement, we have $\delta(n_{P_1}) = \delta(n_{P_2})$, and also $\lim_{n_{P_1}\to\infty} n_{P_1}^{\frac{1}{2}}\delta(n_{P_1}) = c_1$ and $\lim_{n_{P_2}\to\infty} n_{P_2}^{\frac{1}{2}}\delta(n_{P_2}) = c_2$. Note that as $n_{P_2}\to\infty$, $\delta(n_{P_2}) = \delta(n_{P_1}) \to 0$ and thus $n_{P_1}\to\infty$. Therefore we have

$$\text{(3.2)}\qquad \underset{\delta^*\to 0}{\mathrm{ARE}}(P_1,P_2) = \underset{\delta(n_{P_1})=\delta(n_{P_2})\to 0}{\mathrm{ARE}}(P_1,P_2) = \underset{n_{P_2}\to\infty}{\mathrm{ARE}}(P_1,P_2).$$

By combining (2.17) and (2.18), we can compute the $\mathrm{ARE}(P_V,P_{MP})$ (as $\delta^*\to 0$) under $\vec{\theta}_0(t)$:

$$\text{(3.3)}\qquad \underset{\delta^*\to 0}{\mathrm{ARE}}(P_V,P_{MP}) = 12k\sigma^2\left[\int f^2(x)dx\right]^2/(k+1).$$

This $\mathrm{ARE}(P_V,P_{MP})$ is tabulated in Table 3.1 for several df's.

Table 3.1

df	ARE	k=2	k=3	k=5	k=10	k=30
Rectangular	$k/(k+1)$	.66667	.75000	.83333	.90909	.96774
Normal	$3k/[(k+1)\pi]$	.63662	.71620	.79578	.86812	.92413
Logistic	$k\pi^2/[9(k+1)]$	.73108	.82247	.91385	.99693	1.06125
Laplace	$3k/(2k+1)$	1.20000	1.28591	1.36364	1.42857	1.47549
Lower bound*	$.864k/(k+1)$	.57600	.64800	.72000	.78545	.83629

* The lower bound for $12\sigma^2\left[\int f^2(x)dx\right]^2$ was obtained by Hodges and Lehmann (1956) for the location parameter case.

Hodges and Lehmann (1962) aligned observations so that they are free of block effects, and applied joint-rank procedures to random block designs. Likewise we can align observations, apply joint-rank selection procedures, and thus obtain better efficiencies (in the order of $(k+1)/k$). But there are cases where alignments of block effects are not applicable, e.g. p. 485 of Hodges and Lehmann (1962).

In passing we can note that Lehmann's lemma (Lemma 1 of Lehmann (1963)) which leads to (2.13) (and hence to (2.18) as well), is only justified heuristically. We now give a proof. We need the following generalized Helly-Bray Lemma:

Lemma 3.1: (Generalized Helly-Bray Lemma). Let $Q_n \to Q$, a continuous df of a random variable, and let $\{g_n\}, g, h$ be continuous functions satisfying

(i) $|g_n(x)| \leq h(x)$ for all x

(ii) $g_n(x) \to g(x)$ uniformly on finite intervals, and

(iii) $\int h\, dQ_n \to \int h\, dQ$.

Then $\int g_n dQ_n \to \int g\, dQ$.

Proof

See Lemma 7.1.1 of Johnson and Roussas (1970).■

Theorem 3.2: (Lemma 1 of Lehmann (1963).) Let Δ satisfy (2.11) and let $\overline{X}_{(i)}$ be a sample mean (based on the sample size n) yielded by $\pi_{(i)}$ (the population associated with $\theta_{[i]}$), and let σ^2 be a variance of the underlying df F with a pdf f. Under the configuration $\vec{\theta}_0(1,n)$, if $\lim_{n\to\infty} n^{\frac{1}{2}}\delta(n) = 2^{\frac{1}{2}}\Delta\sigma$, then we have

$$\lim_{n\to\infty} P[\overline{X}_{(k)} \geq \overline{X}_{(i)},\ 1 \leq i \leq k-1 \mid \vec{\theta}_0(1,n)] = P^*.$$

Proof

Let $\lim_{n\to\infty} n^{\frac{1}{2}}\delta(n) = \delta$ (>0). Assuming, without loss of generality, that $E(X_{(i)}) = \theta_{[i]}$,

$$\lim_{n\to\infty} P[\overline{X}_{(k)} \geq \overline{X}_{(i)},\ 1 \leq i \leq k-1 \mid \vec{\theta}_0(1,n)]$$

$$= \lim_{n\to\infty} P\left[\frac{\overline{X}_{(k)} - \theta_{[k]}}{\sigma/\sqrt{n}} \geq \frac{\overline{X}_{(i)} - \theta_{[i]}}{\sigma/\sqrt{n}} - \frac{\theta_{[k]} - \theta_{[i]}}{\sigma/\sqrt{n}},\ 1 \leq i \leq k-1\right].$$

Let $Y_i(n) = \dfrac{n^{\frac{1}{2}}(\overline{X}_{(i)} - \theta_{(i)})}{\sigma}$ $(1 \leq i \leq k-1)$ and let $Y_i(n)$ be distributed as $F_n(\cdot)$. Then since the second moment of the underlying df exists, $F_n(y)$ converges to $\Phi(y)$ uniformly for all y as $n \to \infty$, where $\Phi(y) = \int_{-\infty}^{y} (2\pi)^{-\frac{1}{2}} \exp[-x^2/2]dx$. Thus for every given $\varepsilon > 0$ there exists an integer $n_1(\varepsilon)$ such that, whenever $n \geq n_1(\varepsilon)$,

$$|F_n(y + \delta/\sigma) - \Phi(y + \delta/\sigma)| < \varepsilon/2.$$

Now by the continuity assumption there exists an integer $n_2(\varepsilon)$

such that, whenever $n \geq n_2(\varepsilon)$,

$$\max\left(\int_{y\in(\delta/\sigma\,,\, n^{\frac{1}{2}}\delta(n)/\sigma)} dF_n(y) \quad , \int_{y\in(n^{\frac{1}{2}}\delta(n)/\sigma\,,\,\delta/\sigma)} dF_n(y) \right) < \varepsilon/2.$$

Hence, for all $n \geq \max(\, n_1(\varepsilon), n_2(\varepsilon)\,)$,

$$|F_n(y + n^{\frac{1}{2}}\delta(n)/\sigma) - \Phi(y + \delta/\sigma)| < \varepsilon.$$

Therefore we have

$$\prod_{i=1}^{k-1} F_n(x + n^{\frac{1}{2}}\delta(n)/\sigma) \rightarrow \Phi^{k-1}(x + \delta/\sigma),$$

and hence

$$\lim_{n\to\infty} P[\overline{X}_{(k)} \geq \overline{X}_{(i)},\ 1 \leq i \leq k-1 \mid \vec{\theta}_0(1,n)]$$

$$= \lim_{n\to\infty} P[Y_k(n) \geq Y_i(n) - n^{\frac{1}{2}}\delta(n)/\sigma,\ 1 \leq i \leq k-1]$$

$$= \lim \int \left[\prod_{i=1}^{k-1} F_n(y + n^{\frac{1}{2}}\delta(n)/\sigma)\right] dF_n(y)$$

$$= \int \Phi^{k-1}(x + \delta/\sigma)d\Phi(x).$$

The last equality is due to the generalized Helly-Bray Lemma. Thus letting $\delta = 2^{\frac{1}{2}}\Delta\sigma$, where Δ satisfies (2.11), the Theorem follows. ■

Note that if $\delta = 0$ or ∞, then $P[\overline{X}_{(k)} > \overline{X}_{(i)},\ 1 \leq i \leq k-1 \mid \vec{\theta}_0(1,n)]$ converges to $1/k$ or 1 respectively.

4. LFC and Counterexamples.

The configuration of θ_i's which minimizes P[CS] for any given selection procedure is called the least-favorable configuration (LFC). The SPC, $\vec{\theta}_0(t)$, is often least-favorable for selection procedures in the indifference-zone approach, and the equal-parameter configuration (EPC) $\theta_{[1]} = \cdots = \theta_{[k]}$ is often least-favorable in subset-selection approaches. Rizvi and Woodworth (1970) showed that

$$\inf_{\Omega_0(\delta^*,t)} P[CS] < P[CS \mid \vec{\theta}_0(t)] \quad (\inf_{\Omega_0} P[CS] < P[CS|EPC]) \quad \text{for}$$

selection procedures based on joint ranks in the indifference-zone approach (the subset-selection approach) for some df's. And McDonald (1972) also showed that $\inf_{\Omega_0} P[CS] < P[CS|EPC]$ for one of his subset-selection procedures based on vector ranks. In this section the counterexamples of Rizvi and Woodworth (1970) are modified to show that the SPC, $\vec{\theta}_0(t)$, is not the LFC for P_V. We consider two counterexamples: first, for the case of fixed δ^* and finite n; and second, for the case of $\delta^* \to 0$ (and thus $n \to \infty$).

Counterexample 4.1: Let $k = 3$, $t = 1$ and F be a continuous df which places probabilities of q and p (= 1-q) uniformly on the intervals $(0,\varepsilon)$ and $(1,1+\varepsilon)$ respectively, where ε (< 1/3) is a positive constant. Let $\delta^* = \varepsilon$, $0 \le \delta_2 \le \delta^*$, and $\vec{\theta}(\delta_2) = (\theta_1,\theta_2,\theta_3) = (0,\delta_2,\delta_2+\delta^*)$, where θ_i is the location parameter for π_i $(i = 1,2,3)$. Then for $n = 1$ $P[CS|P_V,\vec{\theta}(\delta_2)]$ is a constant for any δ_2 and for $n = 2$

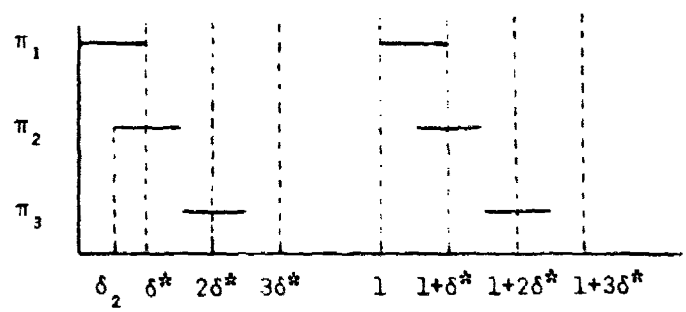
$\max_{\delta_2} P[CS|P_V,\vec{\theta}(\delta_2)] = P[CS|P_V,\vec{\theta}(0)]$ and $\min_{\delta_2} P[CS|P_V,\vec{\theta}(\delta_2)] = P[CS|P_V,\vec{\theta}(\delta^*)]$.

This constitutes a counterexample because $\vec{\theta}(0) = (0,0,\delta^*)$ is a SPC while $\vec{\theta}(\delta^*) = (0,\delta^*,2\delta^*)$ is not.

Proof

The supports of the distribution of the populations under the parameter configuration $\vec{\theta}(\delta_2)$ can be depicted as in Figure 4.1, where "heights" show the supports of df's under $\vec{\theta}(\delta_2)$.

Figure 4.1 Supports of df's under $\vec{\theta}(\delta_2)$



Note that π_3 (the best population) is separated from π_1 and π_2 in its support while π_1 and π_2 do not have disjoint supports. Fix $n=2$. Let B_i be 0, 1, or 2 according as 0, 1, or 2 observations from π_i are in the upper interval of the support of its distribution, let $B=(B_1,B_2,B_3)$, and let $b=(b_1,b_2,b_3)$ be a realization of B. Clearly $P[B=b|(\theta_1,\theta_2,\theta_3)] = \prod_{i=1}^{3} \binom{2}{b_i} p^{b_i} q^{2-b_i}$.

Let $R = \begin{pmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \end{pmatrix}$ be the matrix of ranks each row of which is a row vector of ranks $R_{ij} = 1,2,3$ and let

$r = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \end{pmatrix}$ be a typical realization of R. Given $R = r$ a CS (selection of π_3) occurs with probability 1 if $r_{13}+r_{23} > \max(r_{12}+r_{22}, r_{11}+r_{21})$, with probability 1/2 if either $r_{13}+r_{23} = r_{12}+r_{22} > r_{11}+r_{21}$ or $r_{13}+r_{23} = r_{11}+r_{21} > r_{12}+r_{22}$, with probability 1/3 if $r_{13}+r_{23} = r_{11}+r_{21} = r_{12}+r_{22}$, and with probability 0 otherwise. The conditional probability that $R = r$ given $B = b$ under $\vec{\theta}(\delta_2)$ involves 27 possible rank combinations. Many of the possible rank combinations are not equally likely (hence our situation differs from those of Rizvi and Woodworth (1970) and McDonald (1972), where the rank combinations are equally likely). One example of the computations is that of $P[CS|P_V,\vec{\theta}(\delta_2),B=b]$ for $b=(0,1,1)$. $b=(0,1,1)$ means that for π_1 both observations are from the lower support, and for π_2 and π_3 one of two observations is from the upper support; this can be expressed as:

$$\text{(i)} \quad \begin{pmatrix} [0,\delta^*], & [\delta_2,\delta_2+\delta^*] & , & [\delta_2+\delta^*,\delta_2+2\delta^*] \\ [0,\delta^*], & [1+\delta_2,1+\delta_2+\delta^*], & & [1+\delta_2+\delta^*,1+\delta_2+2\delta^*] \end{pmatrix},$$

and

$$\text{(ii)} \quad \begin{pmatrix} [0,\delta^*], & [1+\delta_2,1+\delta_2+\delta^*], & & [\delta_2+\delta^*,\delta_2+2\delta^*] \\ [0,\delta^*], & [\delta_2,\delta_2+\delta^*] & , & [1+\delta_2+\delta^*,1+\delta_2+2\delta^*] \end{pmatrix}$$

express supports from which observations for each population originate to have $b=(0,1,1)$. Given (i) there are two possible rank combinations, while given (ii) there are two other possibilities. We now compute the probability of each rank combination. Let $\delta_1 = \delta^* - \delta_2$. Then

$$P\left[\begin{pmatrix}1, & 2, & 3\\ 1, & 2, & 3\end{pmatrix}\right] = \{\, P[0<X_1<\delta_2, \delta_2<X_2<\delta_2+\delta^*, \delta_2+\delta^*<X_3<\delta_2+2\delta^*]$$

$$+P[\delta_2<X_1<\delta^*, \delta_2<X_1<X_2<\delta^*, \delta_2+\delta^*<X_3<\delta_2+2\delta^*]$$

$$+P[\delta_2<X_1<\delta^*, \delta^*<X_2<\delta_2+\delta^*, \delta_2+\delta^*<X_3<\delta_2+2\delta^*]\}$$

$$\times\ P[0<X_1<\delta^*, 1+\delta_2<X_2<1+\delta_2+\delta^*, 1+\delta_2+\delta^*<X_3<1+\delta_2+2\delta^*]$$

$$= q^4p^2(1-\delta_1^2/2\delta^{*2}).$$

Similarly

$$P\left[\begin{pmatrix}2, & 1, & 3\\ 1, & 2, & 3\end{pmatrix}\right] = q^4p^2\ \delta_1^2/2\delta^{*2}, \quad P\left[\begin{pmatrix}1, & 3, & 2\\ 1, & 2, & 3\end{pmatrix}\right] = q^4p^2(1-\delta_1^2/2\delta^{*2}), \quad \text{and}$$

$$P\left[\begin{pmatrix}1, & 3, & 2\\ 2, & 1, & 3\end{pmatrix}\right] = q^4p^2\ \delta_1^2/2\delta^{*2}.$$

Thus

$$P[CS\,|\,P_V, \vec{\theta}(\delta_2)\ , b=(0,1,1)] = 3/4 + \delta_1^2/(8\delta^{*2}).$$

For the other cases the method of computation is similar. In all but 9 cases, $P[CS\,|\,P_V,\vec{\theta}(\delta_2),b]$ equals $P[CS\,|\,P_V,\vec{\theta}(\delta^*),b]$; those 9 cases are listed in Table 4.1. Now

$$P[CS\,|\,P_V,\vec{\theta}(\delta_2)] - P[CS\,|\,P_V,\vec{\theta}(\delta^*)]$$

$$= [1-\delta_1^2/(2\delta^{*2})]\delta_1^2/(2\delta^{*2})[(4/3)q^4p^2 + (4/3)q^3p^3 + (4/3)qp^5] > 0,$$

and the difference is monotone increasing in δ_2 for $0 \le \delta_1 \le \delta^*$ (namely monotone decreasing in δ_2 for $0 \le \delta_2 \le \delta^*$). Thus we conclude that $P[CS\,|\,P_V,\vec{\theta}(\delta_2)]$ is maximized at $\vec{\theta}(0)$ (which is a SPC) and is minimized at $\vec{\theta}(\delta^*)$. The case of $n = 1$ is trivial and the Counterexample follows.

Table 4.1 *

b	P[B=b]	$P[CS \mid P_V, B=b, \vec{\theta}(\delta_2)]$
(0,1,0)	$2q^5p$	$1/2 + \delta_1^2/(4\delta^2)$
(0,1,1)	$4q^4p^2$	$3/4 + \delta_1^2/(8\delta^2)$
(1,0,0)	$2q^5p$	$1 - \delta_1^2/(4\delta^2)$
(1,0,1)	$4q^4p^2$	$1 - \delta_1^2/(8\delta^2)$
(1,1,0)	$4q^4p^2$	$1/6 + 1/3[1 - \delta_1^2/(2\delta^2)][\delta_1^2/(2\delta^2)]$
(1,1,1)	$8q^3p^3$	$3/4 + 1/6[1 - \delta_1^2/(2\delta^2)][\delta_1^2/(2\delta^2$
(1,2,1)	$4q^2p^4$	$1/4 + 5/24(\delta_1^2/\delta^2)$
(2,1,1)	$4q^2p^4$	$2/3 - 5/24(\delta_1^2/\delta^2)$
(2,2,1)	$2qp^5$	$2/3(1 - \delta_1^2/\delta^2)[\delta_1^2/(2\delta^2)]$

* Note that $\delta = \delta^*$ and $\delta_1 = \delta^* - \delta_2$. ■

We now show that the SPC is not the LFC for P_V even for large samples. One method of showing this is to show that the ratio of $P[CS \mid P_V]$ under a configuration different from the SPC to that under the SPC converges to some number smaller than 1 for fixed δ^* as $n \to \infty$. Another method of constructing a counterexample is to show that the ratio of sample size for a configuration different from the SPC to that for the SPC converges to some number smaller than 1 for fixed δ^* as $P^* \to 1$. However we have obtained a counterexample by another method originated by Rizvi and Woodworth (1970): we show that when the relation between n and $\delta(n)$ ($= \theta_{[k-t+1]} - \theta_{[k-t]}$) satisfies (2.16), $P[CS|P_V]$ converges (as $n \to \infty$) to some number smaller than P^* under a certain configuration of θ_i's different from the SPC, $\vec{\theta}_0(t)$, but still in $\Omega_\theta(\delta^*,t)$. This serves our purpose, because when the relation (2.16) holds between $\delta(n)$ and n, $P[CS|P_V]$ converges, as $n \to \infty$, to P^* under the SPC. [One may ask how much larger $P[CS|P_V]$ is under the SPC than under the configuration we will consider; this question is discussed in the next section.]

Consider the selection of the t best populations, when the underlying df is a logistic distribution with a location parameter. For simplicity take $k \geq 4$ (k even) and $t = k/2$. Without loss of generality drop [ ] around the ordered parameter values for convenience of notation; namely take $\theta_{[i]} = \theta_i$, $\pi_{(i)} = \pi_i$, and thus $H_{(i)}(n) = H_i(n)$ $(1 \leq i \leq k)$.

<u>Lemma 4.2</u>: Let $F(x) = (1+e^{-x})^{-1}$ and let

$$\vec{\theta}_1(t,n);\quad \theta_1 = \ldots = \theta_{k-t-1} = -\theta_0,\quad \theta_{k-t} = 0,\quad \theta_{k-t+1} = \delta(n),$$

$$\theta_{k-t+2} = \ldots = \theta_k = \theta_0,$$

where $\delta(n) > 0$ and is in the order of $O(n^{-\frac{1}{2}})$, $\theta_0 > 0$ fixed satisfying $\theta_0 > \delta(n)$, and $k = 2t$. Then

$$(4.1)\qquad \lim_{n\to\infty} P[CS|P_V, \vec{\theta}_1(t,n)] \leq \Phi[\Delta_t \rho((k+1)/k)^{\frac{1}{2}}]$$

where Δ_t satisfies (2.15),

$$(4.2) \qquad \rho = \{3^{\frac{1}{2}}\int H_0(2F-1)dF\}/\{\int H_0^2 dF - (\int H_0 dF)^2\}^{\frac{1}{2}},$$

$$(4.3) \qquad H_0(x) = (k-t-1)F(x-\theta_0) + 2F(x) + (t-1)F(x+\theta_0),$$

and

$$\lim_{n\to\infty} n^{\frac{1}{2}}\delta(n) = [k(k+1)/6]^{\frac{1}{2}}\,\Delta_t[\int kf^2(x)dx]^{-1}.$$

Proof

For large samples, dropping the randomization part we have

$$(4.4) \qquad P[CS|P_V, \vec{\theta}_1(t,n)] \doteq P[\max_{1\le i\le k-t} H_i(n) < \min_{k-t<j\le k} H_j(n) \mid \vec{\theta}_1(t,n)]$$

$$\le P[V(n) > 0 \mid \vec{\theta}_1(t,n)],$$

where $\qquad V(n) = n^{-\frac{1}{2}}(H_{k-t+1}(n) - H_{k-t}(n)).$

We will find an upper bound for (4.4) as $n \to \infty$ by finding $\lim_{n\to\infty} E[V(n)]$ and $\lim_{n\to\infty} Var[V(n)]$, and applying a Lindeberg-Feller type central limit theorem. The computations for $E[V(n)]$ and $Var[V(n)]$ are lengthy, and thus are omitted. In the limiting process using Olshen's lemma (Lemma (12) of Olshen (1967)), we have

$$\lim_{n\to\infty} E[V(n)|\vec{\theta}_1(t,n)] = [6(k+1)/k]^{\frac{1}{2}}\,\Delta_t \int H_0(x)[2F(x)-1]dF(x),$$

where $H_0(x)$ is given by (4.3), and $(0 < \theta_0 \le C(k,t,F))$

$$(4.5) \qquad \lim_{n\to\infty} Var[V(n)|\vec{\theta}_1(t,n)] \ge 2[\int H_1^2 dF - (\int H_1 dF)^2]$$

and $t \geq 2$, where $H_1 = (k-t-1)F(x+\theta_0) + 2F(x) + (t-1)F(x-\theta_0)$. Since we assume $k = 2t$, we have $H_1 \equiv H_0$. Thus from (4.4)

$$\lim_{n\to\infty} P[CS|P_V, \vec{\theta}_1(t,n)] \leq \lim_{n\to\infty} P[V(n) > 0 \mid \vec{\theta}_1(t,n)]$$

$$= \lim_{n\to\infty} P\left\{\frac{V(n)-\lim E[V(n)]}{\lim\sqrt{Var[V(n)]}} \geq -\frac{\lim E[V(n)]}{\lim\sqrt{Var[V(n)]}}\right\}$$

$$\leq \Phi[\Delta\rho((k+1)/k)^{\frac{1}{2}}],$$

where the last inequality is due to the asymptotic normality of $\{V(n) - \lim E[V(n)]\}/\{\lim Var[V(n)]^{\frac{1}{2}}\}$ due to a Lindeberg-Feller type central limit theorem and (4.5). ∎

Lemma 4.3: For any k and t, $1 \leq t < k$,

$$\lim_{P^*\to 1} \Phi^{-1}(P^*)/\Delta_t = 1$$

where Δ_t satisfies (2.15).

Proof

This follows from Lemma 2 of Rizvi and Woodworth (1970) upon noting that Δ_t, which satisfies (2.15), also satisfies

$$P[\max_{1\leq i\leq k-t} Z_i < \min_{k-t<j\leq k} Z_j + \sqrt{2}\,\Delta_t] = P^*$$

where Z_i $(1 \leq i \leq k)$ are independent standard normal random variables. [For the case $t = 1$, Dudewicz (1969) also obtained the result of Lemma 4.3 in a different form.]

Counterexample 4.4: Under the same setup as in Lemma 4.2,

$$\lim_{n\to\infty} P[CS|P_V,\vec{\theta}_1(t,n)] < P^* = \lim_{n\to\infty} P[CS|P_V,\vec{\theta}_0(t,n)].$$

Proof

Note that $0 \leq \rho < 1$, since $\rho = \mathrm{corr}(H_0, 2F-1)$ and H_0 and $2F-1$ are monotone increasing in x for fixed θ_0. Choose P^* and k large enough such that $[\Delta_t/\Phi^{-1}(P^*)](k+1)/k < 1/\rho$. Substituting this into (4.1), the inequality follows. The equality is due to Theorem 2.7. ■

Through Counterexamples 4.1 and 4.4, we have seen that the SPC minimizes $P[CS|P_V]$ neither when one has a fixed sample size nor when one lets δ^* $(= \theta_{[k-t+1]} - \theta_{[k-t]})$ tend to zero as $n \to \infty$. Note that the logistic df possesses a monotone likelihood ratio with respect to its location parameter and has a support independent of its location parameter; thus imposing additional conditions such as the above two will not obviate the difficulty in the LFC.

It is an open question whether (for selection of the t best by P_V)

$$\inf_{\omega_\theta(\delta^*,t)} P[CS|P_V] = P[CS|P_V,\vec{\theta}_0(t)].$$

5. Remarks on Selection Procedures based on Ranks.

In the literature of selection procedures based on ranks (either joint ranks or vector ranks) each contribution either imposes artificial restrictions on the parameter space (Puri and Puri (1968), (1969), Gupta and McDonald (1970), and McDonald (1972), (1973)), or is not able to find the LFC (Blumental and Patterson (1969)), or was partially invalidated by Rizvi and Woodworth (1970) (Lehmann (1963), and Bartlett and Govindarajulu (1968)). A conjecture as to why these procedures were invalidated is that the LFC's for them were sought in a parameter space where the P[CS] for certain parametric procedures is monotone while the P[CS] for rank procedures is not monotone (as is indicated by Gupta and McDonald (1970) and Blumental and Patterson (1969)).

For any procedures based on joint ranks or vector ranks, $\mathcal{P}_{RANK}$, define

$$R_{ID} = \frac{\inf_{\Omega_\theta(\delta^*,t)} P[CS|\mathcal{P}_{RANK}] - P[CS|\mathcal{P}_{RANK},\vec{\theta}_0(t)]}{P[CS|\mathcal{P}_{RANK},\vec{\theta}_0(t)]} \times 100$$

for the indifference zone approach, and

$$R_{SS} = \frac{\inf_{\Omega_\theta} P[CS|\mathcal{P}_{RANK}] - P[CS|\mathcal{P}_{RANK},EPC]}{P[CS|\mathcal{P}_{RANK},EPC]} \times 100$$

for the subset-selection approach. Then the quantities R_{ID} and R_{SS} merit study because small R_{ID} and R_{SS} may well justify the SPC assumption (which will simplify theoretical development) while large R_{ID} and R_{SS} imply that the SPC assumption may be of only theoretical interest. [This aspect was called to our attention by Dr. Gary C. McDonald.]

We have noted in Section 4 that P_V also suffers in the LFC unless the SPC is assumed. Hence we wish to compute R_{ID} in the case of Counterexample 4.1 (where $n = 2$, $k = 3$, and the LFC is relatively simple). Our results on R_{ID} for $p = .01(.01).99$ (and some typical values of $P[CS|P_V, SPC]$) are summarized in Table 5.1. The minimum of $P[CS|P_V, SPC]$ is .66146 (occurring at $p = .50$) and the maximum R_{ID} is 3.11234% (occurring at $p = .77$) out of the cases studied. These computations indicate that the assumption of the SPC as an underlying configuration may not be unreasonable. We propose that further study of R_{ID} and R_{SS} be carried out to see in how far this result generalizes to other cases.

Table 5.1 R_{ID}

p	$P[CS\|P_V, SPC]$	R_{ID} (%)
.01	.98991	.00327
.10	.89555	.27168
.20	.79887	.86522
.30	.72492	1.49768
.40	.67910	1.99793
.50	.66146 (minimum)	2.36220
.60	.67014	2.69315
.70	.70420	3.01292
.77	.74403	3.11234 (maximum)
.80	.76559	3.07632
.90	.86051	2.31874
.99	.98363	.32231

6. Discussions

We have studied mathematical properties of the vector rank procedure when applied to selecting the largest location parameter in randomized block models. Even if the data is not quantitative but ordinal, or is not from a location model but from a stochastically ordered family of distributions, the vector rank procedure is applicable.

An important competing selection procedure is based on the robust estimate of location parameters (Sen and Puri (1972)). The selection procedure based on robust location estimates does not have the LFC difficulty that the vector rank procedure suffers from, and its relative efficiency compared to the means procedure is that of the Mann-Whitney- Wilcoxon test versus the t-test. A serious disadvantage, however, is that the robust location estimate method is not applicable if the data is ordinal or from a non-location family: for example, see Lee and Dudewicz (1980) where the data is incomplete rank order scores or Lee (1980) where the distributional origin of the data is not known.

We now discuss how to choose a proper selection procedure to be applied. If the data is from a location family, then the robust procedure of choice should be based on robust location estimates. If the data originates from a location family but in the ordinal form, or from a stochastically increasing scale parameter family, then the vector rank procedure may be applied to selecting the population with the largest parameter of interest. In this latter case, it is possible that the $P(CS) \geq P^*$ requirement is not met. In doubtful cases, the multinomial category selection procedure (Lee, 1980) is a possible alternative.

As a final remark, note that the procedure considered here is, like most robust selection procedures, not nonparametric since the required sample size (say (2.12) or (2.17)) depends on the underlying distribution, but is less senstitive to departure from the assumed underlying distribution than the means procedure.

REFERENCES

1. Alam, K., and Thompson, J. R. (1971): "A Selection Procedure based on Ranks," Ann. Inst. Statist. Math., Vol. 23, pp. 253-262.

2. Bartlett, N. S., and Govindarajulu, Z. (1968): "Some Distribution-free Statistics and Their Application to the Selection Problem," Ann. Inst. Statist. Math., Vol. 20, pp. 79-97.

3. Bechhofer, R. E. (1954): "A Single-Sample Multiple Decision Procedure for Ranking Means of Normal Populations with Known Variances," Ann. Math Statist., Vol. 25, pp. 16-39.

4. Bhapkar, V. P. and Gore, A. P. (1971): "Some Selection Procedures based on U-statistics for the Location and Scale Problems," Ann. Inst. Statist. Math., Vol. 23, pp. 375-386.

5. Blumenthal, S. and Patterson, D. W. (1969): "Rank Order Procedures for Selecting a Subset Containing the Population with the Smallest Scale Parameter," Sankhyā, Series A, Vol. 31, pp. 37-42.

6. Dudewicz, E. J. (1969): "An Approximation to the Sample Size in Selection Problems," Ann. Math. Statist., Vol. 40, pp. 492-497.

7. Dudewicz, E. J. and Dalal, S. R. (1975): "Allocation of Observations in Ranking and Selection with Unknown Variances," Sankhyā, Vol. 37B, pp. 28-78.

8. Dudewicz, E. J., and Fan, C-L (1973): "Further Light on Nonparametric Selection Efficiency," Naval Research Logistics Quarterly, Vol. 20, pp. 737-743.

9. Gnedenko, B. V. and Kolmogorov, A. N. (1949): Limit Distributions for Sums of Independent Random Variables, translated by K. L. Chung in 1954, Addison-Wesley, Reading, Massachusetts.

10. Gupta, S. S. (1956): "On a Decision Rule for a Problem in Ranking Means," Mimeograph Series No. 150, Institute of Statistics, University of North Carolina, Chapel Hill, N.C.

11. Gupta, S. S. (1965): "On Some Multiple Decision (Selection and Ranking) Rules," Technometrics, Vol. 7, pp. 225-245.

12. Gupta, S. S., and McDonald, G. C. (1970): "On Some Classes of Selection Procedures Based on Ranks," Nonparametric Techniques in Statistical Inference, edited by M. L. Puri, University of Cambridge Press, Cambridge, England, pp. 491-514.

13. Hodges, J. L., Jr., and Lehmann, E. L. (1956): "The Efficiency of Some Nonparametric Competitors of the t-test," Ann. Math. Statist., Vol. 27, pp. 324-335.

14. Hodges, J. L., Jr., and Lehmann, E. L. (1962): "Rank Methods for Combination of Independent Experiments in Analysis of Variance," Ann. Math. Statist., Vol. 33, pp. 482-497.

15. Jogdeo, K. (1966): "On Randomized Rank Score Procedures of Bell and Doksum," Ann. Math. Statist., Vol. 37, pp. 1697-1703.

16. Johnson, R. S., and Roussas, G. G. (1970): "Asymptotically Optimal Tests in Markov Processes," Ann. Math. Statist., Vol. 41, pp. 918-938.

17. Kagan, A. M., Linnik, Y. U., and Rao, C. R. (1973): Characterization Problems in Mathematical Statistics, John Wiley & Sons, New York.

18. Lee, Y. J. (1980): "Nonparametric Selections in Blocked Data: Application to Motor Vehicle Fatality Rate Data," Technometrics, Vol. 22, pp. 535-542.

19. Lee, Y. J., and Dudewicz, E. J. (1974): "Nonparametric Ranking and Selection Procedures" Technical Report No. 105, Department of Statistics, The Ohio State University, Columbus, Ohio.

20. Lee, Y. J., and Dudewicz, E. J. (1980): "Awardee Selection Procedures, with Special Reference to the Frank Wilcoxon and Jack Youden Prizes," Technometrics, Vol. 22, pp. 121-124.

21. Lehmann, E. L. (1963): "A Class of Selection Procedures Based on Ranks," Mathe. Annal. Vol 150, pp. 268-275.

22. McDonald, G. C. (1972): "Some Multiple Comparison Selection Procedures Based on Ranks," Sankhyā, Series A, Vol. 34, pp. 53-64.

23. McDonald, G. C. (1973): "The Distribution of Some Rank Statistics in Block Design Selection Problems," Sankhyā, Series A, Vol. 35, pp. 187-204.

24. Mehra, K. L., and Sarangi, J. (1967): "Asymptotic Efficiency of Certain Rank Tests for Comparative Experiments," Ann. Math. Statist., Vol. 38, pp. 90-107.

25. Olshen, R. A. (1967): "Sign and Wilcoxon Tests for Linearity," Ann. Math. Statist., Vol. 68, pp. 1759-1769.

26. Puri, M. L., and Puri, P. S. (1968): "Selection Procedures Based on Ranks, Scale Parameter Case," Sankhya, Series A, Vol. 30, pp. 291-302.

27. Puri, M. L., and Puri, P. S. (1969): "Multiple Decision Procedures Based on Ranks for Certain Problems in Analysis of Variance," Ann. Math. Statist., Vol. 40, pp. 619-632.

28. Rizvi, M. H., and Woodworth, G. G. (1970): "On Selection Procedures Based on Ranks: Counterexamples Concerning Least Favorable Configurations," Ann. Math. Statist., Vol. 41, pp. 1942-1951.

29. Sen, P. K., and Puri, M. L. (1972): "On Some Selection Procedures in Two-Way Layouts," Z. Wahrscheinlichkeitstheorie verw. Geb., Vol. 22, pp. 242-250.