Reprinted from *Nonparametric Techniques in Statistical Inference*, edited by Madan Lal Puri, C.U.P. 1970.

# ON SOME CLASSES OF SELECTION PROCEDURES BASED ON RANKS†

SHANTI S. GUPTA AND GARY C. MCDONALD

## 1 INTRODUCTION AND SUMMARY

The shortcomings of the classical tests of homogeneity, i.e. testing the hypothesis of equality of parameters, have long been known. Given $k$ populations and from each population a fixed number of observations whose distribution depends on a parameter $\theta_i$, concluding that all $\theta_i$ are not equal may not be sufficient. Often the experimenter is interested in ascertaining which population is associated with the largest (or smallest) $\theta$, which populations possess the $t$ largest (or smallest) $\theta$, etc. Suppose the experimenter is interested in identifying which one of the $k$ populations possesses the largest $\theta$, the so-called 'best' population. The parameter $\theta$ may be, for example, the mean, the variance, some quantile, or some function of these quantities. Basically, there have been two approaches to ranking and selection problems, the 'indifference zone' approach and the 'subset selection' approach. In the first a single population is chosen and is guaranteed to be the best with probability $P^*$ whenever a certain indifference zone condition holds. For example, in case the populations have normal distributions with a common known variance and unknown means the experimenter may be interested in guaranteeing this probability to be at least $P^*$ whenever the two largest means are separated by a distance greater than $d^*$. This formulation is due to Bechhofer [5]. The second approach requires no specifications of the parameter space. However, a single population is not necessarily chosen; rather a subset of the given $k$ populations is selected which is guaranteed to contain the best population with probability $P^*$, the basic probability requirement in these procedures. In this sense the number of populations in the selected subset is a random variable. This formulation is due to Gupta [7, 10].

In the past ten years many papers have appeared on both formulations of the selection problem. As can be expected, most of this research has been devoted to rules which assume a specific distributional form of the

† This research was supported in part by the Office of Naval Research Contract NONR-1100(26) and the Aerospace Research Laboratories Contract AF 33(615)67C1244 at Purdue University. Reproduction in whole or part is permitted for any purposes of the United States Government.

underlying observations; e.g. normal, binomial, multinomial, etc. Barlow and Gupta [2] and Barlow, Gupta and Panchapakesan [3] have considered the problem of selecting a subset containing the largest (smallest) quantile of a given order and a subset containing the largest (smallest) mean. They assume the observations from each population have a distribution which belongs to certain restricted families, e.g. IFR (Increasing Failure Rate) distributions, IFRA distributions, etc. Distribution-free selection procedures, most of which are based on joint ranks of the observations, have been studied by Lehmann [13]; Patterson [18]; Dudewicz [6]; Rizvi and Sobel [22]; Bartlett and Govindarajulu [4]; Puri and Puri [20, 21]; and McDonald [15].

The present paper deals with three classes of nonrandomized distribution-free ranking and selection procedures under the subset selection formulation. The main problem is to select a subset of $k$ given populations which contains the 'best' population with probability of at least $P^*$. The random variables associated with a fixed population are assumed to be independent identically distributed with a continuous distribution function depending on a scalar parameter. This parameter is assumed to stochastically order the $k$ distribution functions, and the 'best' population is the stochastically largest (smallest) population. The procedures presented depend on the individual observations of a given population only through their ranks in the combined sample. In other words, one is not required to have at hand the actual observations from each population; it suffices to have these ranks, which in some preference-type tests or lost data problems may be the only information available to an experimenter.

In § 2 the problem is formally stated and the three classes of rules are defined. In § 3 the probability of making a correct selection, i.e. selecting the population with the largest parameter, using these rules is investigated. For all these classes this probability is shown to be a nondecreasing function in the largest parameter. For one of the classes this probability is further shown to be nonincreasing in all parameters but the largest. For the other two classes of rules, § 4 provides bounds on the probability of correct selection, which in turn provides conservative bounds on the constants needed for the actual implementation of these rules. Section 5 presents exact expressions for the means, variances and covariances of the statistics upon which our selection rules are based. In § 6 some distribution theory is presented which arises from consideration of one selection procedure based on the rank sums of each population. Section 7 discusses some properties of these selection rules, e.g. local optimality and monotonicity, and makes some comparison between rules of the three classes in terms of the expected number of populations included in the selected subset.

## 2 FORMULATION OF THE PROBLEM AND THREE CLASSES OF RULES

Let $\pi_1, \ldots, \pi_k$ be $k\ (\geqslant 2)$ independent populations. The associated random variables $X_{ij}, j = 1, \ldots, n_i$; $i = 1, \ldots, k$, are assumed independent and to have a continuous distribution $F_{\theta_i}(x)$, where $\theta_i$ belong to some interval $\Theta$ on the real line. Suppose $F_\theta(x)$ is a stochastically increasing (SI) family of distributions, i.e. if $\theta_1$ is less than $\theta_2$, then $F_{\theta_1}(x)$ and $F_{\theta_2}(x)$ are distinct and $F_{\theta_2}(x) \leqslant F_{\theta_1}(x)$ for all $x$. Examples of such families of distributions are: (1) any location parameter family, i.e. $F_\theta(x) = F(x-\theta)$; (2) any scale parameter family, i.e. $F_\theta(x) = F(x/\theta), x > 0, \theta > 0$; (3) any family of distribution functions whose densities possess the monotone likelihood ratio (or $TP_2$) property. Let $R_{ij}$ denote the rank of the observation $x_{ij}$ in the combined sample; i.e. if there are exactly $r$ observations less than $x_{ij}$ then $R_{ij} = r+1$. These ranks are well-defined with probability one, since the random variables are assumed to have a continuous distribution. Let $Z(1) \leqslant Z(2) \leqslant \ldots \leqslant Z(N)$ denote an ordered sample of size $N = \sum_{i=1}^{k} n_i$ from any continuous distribution $G$, such that

$$-\infty < a(r) \equiv E[Z(r)|G] < \infty \quad (r = 1, \ldots, N).$$

With each of the random variables $X_{ij}$ associate the number $a(R_{ij})$ and define

$$H_i = n_i^{-1} \sum_{j=1}^{n_i} a(R_{ij}) \quad (i = 1, \ldots, k). \tag{2.1}$$

Using the quantities $H_i$, we wish to define procedures for selecting a subset of the $k$ populations. Letting $\theta_{[i]}$ denote the $i$th smallest unknown parameter, we have

$$F_{\theta_{[1]}}(x) \geqslant F_{\theta_{[2]}}(x) \geqslant \ldots \geqslant F_{\theta_{[k]}}(x) \quad (\forall x). \tag{2.2}$$

The population whose associated random variables have the distribution $F_{\theta_{[k]}}(x)$ will be called the best population. In case several populations possess the largest parameter value $\theta_{[k]}$, one of them is tagged at random and called the best. A 'Correct Selection' (CS) is said to occur if and only if the best population is included in the selected subset. In the usual subset selection problem one wishes to select a subset such that the probability is at least equal to a preassigned constant $P^*$ $(1/k < P^* < 1)$

that the selected subset includes the best population. Mathematically, for a given selection rule $R$,

$$\inf_{\Omega} P(\mathrm{CS}|R) \geqslant P^*, \tag{2.3}$$

where

$$\Omega = \{\theta = (\theta_1, \dots, \theta_k) : \theta_i \in \Theta, i = 1, 2, \dots, k\}. \tag{2.4}$$

The following three classes of selection procedures, which choose a subset of the $k$ given populations, and which depend on the given distribution $G$, will be considered:

$$R_1(G): \text{select } \pi_i \text{ iff } H_i \geqslant \max_{1 \leqslant j \leqslant k} H_j - d \quad (i = 1, \dots, k, d \geqslant 0), \tag{2.5}$$

$$R_2(G): \text{select } \pi_i \text{ iff } H_i \geqslant c^{-1} \max_{1 \leqslant j \leqslant k} H_j \quad (i = 1, \dots, k, c \geqslant 1), \tag{2.6}$$

$$R_3(G): \text{select } \pi_i \text{ iff } H_i \geqslant D \quad (i = 1, \dots, k, -\infty < D < \infty). \tag{2.7}$$

It should be noted that rules $R_1(G)$, $R_2(G)$, and $R_3(G)$ are equivalent if $k = 2$. The procedures $R_1(G)$ (and their randomized analogs) have been suggested by Bartlett and Govindarajulu [4] for continuous distributions differing by a location parameter. The procedure $R_2(G)$ will be studied in this paper *only* for the case where $H_i \geqslant 0$ for all $i$. The constants $d$ and $c$ are usually chosen to be as small as possible, $D$ as large as possible, while satisfying the probability requirement (2.3). The number of populations included in the selected subset is a random variable which takes values 1 to $k$ inclusive for rules $R_1(G)$ and $R_2(G)$. The subset chosen by rule $R_3(G)$, however, could possibly be empty.

Another class of selection rules which includes $R_1(G)$ and $R_3(G)$ as special cases, and depends on an index $t$ ($1 \leqslant t \leqslant \infty$), can be defined as follows when $H_i$ are nonnegative:

$$R(G): \text{select } \pi_i \text{ iff } H_i \geqslant \left( \frac{1}{k-1} \sum_{\substack{j=1 \\ j \neq i}}^{k} H_j^t \right)^{1/t} - d_t \quad (i = 1, \dots, k;\ d_t \geqslant 0). \tag{2.8}$$

For $t = 1$, this rule reduces to a rule of the form $R_3(G)$ since the sum of all the $H_j$ is constant, and for $t = \infty$, $R(G)$ reduces to a rule of the type $R_1(G)$.

Let $\pi_{(i)}$ be the population associated with $\theta_{[i]}$, the $i$th smallest $\theta_i$. Then the probability of making a correct selection using the procedures $R_i(G)$, $i = 1, 2, 3$, is given, respectively, by

$$P(\mathrm{CS}|R_i(G)) = \begin{cases} P(H_{(k)} \geqslant \max_{1 \leqslant j \leqslant k} H_{(j)} - d) & (i = 1), \\ P(H_{(k)} \geqslant c^{-1} \max_{1 \leqslant j \leqslant k} H_{(j)}) & (i = 2), \\ P(H_{(k)} \geqslant D) & (i = 3). \end{cases} \tag{2.9}$$

The corresponding rules for choosing a subset of the $k$ populations which contains the population with the smallest parameter, $\pi_{(1)}$, are:

$$R_1'(G)\text{: select } \pi_i \text{ iff } H_i \leqslant \min_{1\leqslant j\leqslant k} H_j + d' \quad (i = 1, \ldots, k;\ d' \geqslant 0), \quad (2.10)$$

$$R_2'(G)\text{: select } \pi_i \text{ iff } H_i \leqslant c' \min_{1\leqslant j\leqslant k} H_j \quad (i = 1, \ldots, k;\ c' \geqslant 1), \quad (2.11)$$

$$R_3'(G)\text{: select } \pi_i \text{ iff } H_i \leqslant D' \quad (i = 1, \ldots, k;\ -\infty < D' < \infty). \quad (2.12)$$

The constants $d'$, $c'$ and $D'$ are obtained as before. No more consideration will be given to these three rules; results and methods developed for $R_1(G)$, $R_2(G)$ and $R_3(G)$ will have an obvious analog for $R_1'(G)$, $R_2'(G)$ and $R_3'(G)$, respectively.

## 3 THE INFIMUM OF THE PROBABILITY OF A CORRECT SELECTION

We start with a lemma, which is essentially the same as Lemma 4.2 in Mahamunulu [14] and Lemma 2.1 in Alam and Rizvi [1] both being a generalization of a result of Lehmann [12, p. 112, no. 11] for more than one dimension. We state our version without proof.

*Lemma* 3.1. *Let* $\mathrm{X} = (X_{11}, \ldots, X_{1n_1}, \ldots, X_{k1}, \ldots, X_{kn_k})$ *be a vector valued random variable of* $\sum_{i=1}^{k} n_i (\geqslant 1)$ *independent components with* $X_{ij}$ *having the distribution* $F_{\theta_i}(x), j = 1, \ldots, n_i; i = 1, \ldots, k$. *Suppose* $F_\theta(x)$ *is a* SI *family of distributions. Let* $\Psi$ *be a function of* $x_{11}, \ldots, x_{1n_1}, \ldots, x_{k1}, \ldots, x_{kn_k}$ *which, for any fixed* $i$, *is a nondecreasing (nonincreasing) function of* $x_{i1}, \ldots, x_{in_i}$ *when the other components of* x *are held fixed. Then* $E_\theta[\Psi(\mathrm{X})]$ *is a nondecreasing (nonincreasing) function of* $\theta_i$.

*Theorem* 3.1. *For rules* $R_i(G)$, $i = 1, 2, 3, p_s(R_i(G))$, *the probability of including the population* $\pi_{(s)}$ *in the selected subset is nondecreasing in* $\theta_{[s]}$ *and, hence,*

$$\inf_{\Omega} p_s(R_i(G)) = \inf_{\Omega_s} p_s(R_i(G)) \quad (s = 1, \ldots, k), \quad (3.1)$$

*where*

$$\Omega_s = \{\theta \in \Omega\colon \theta_{[s]} = \theta_{[s-1]}\} \quad (3.2)$$

*and* $\theta_{[0]}$ *is the least admissible value of* $\theta$.

*Proof.* We will prove it for the rule $R_1(G)$. Let

$$\Psi_s(\mathrm{X}) = \begin{cases} 1 & \text{if } H_{(s)} \geqslant \max_{j \neq s} H_{(j)} - d, \\ 0 & \text{otherwise.} \end{cases} \quad (3.3)$$

Let $R_{(i)j}$ be the rank of $X_{(i)j}, j = 1, \ldots, n_{(i)}$, and consider an observation $x_{(s)l}$ for some fixed $l$, $1 \leqslant l \leqslant n_{(s)}$. As $x_{(s)l}$ increases and the other observations remain fixed, either:

(1) $x_{(s)l}$ surpasses first an $x_{(m)j}$, $m \neq s$, so $R_{(s)l}$ increases by 1 and $R_{(m)j}$ decreases by 1; or
(2) $x_{(s)l}$ surpasses first an $x_{(s)j}$, $j \neq l$, so $R_{(s)l}$ increases by 1 and $R_{(s)j}$ decreases by 1; or
(3) $x_{(s)l}$ does not surpass any other observation, so all ranks remain the same.

In all three cases, $H_{(s)}$ is nondecreasing and $H_{(j)}$, $j \neq s$, is nonincreasing and hence so is $\max_{j \neq s} H_{(j)}$. Therefore, $\Psi_s(\mathrm{x})$ is a nondecreasing function of $x_{(s)j}$, $j = 1, \ldots, n_{(s)}$. By Lemma 3.1, $E_\theta[\Psi_s(\mathrm{X})] = p_s(R_1(G))$ is a nondecreasing function of $\theta_{[s]}$. A similar argument proves the result for $R_2(G)$ and $R_3(G)$.

In particular, for $s = k$, Equation (3.1) can be written as

$$\inf_{\Omega} P(\mathrm{CS}|R_i(G)) = \inf_{\Omega_k} P(\mathrm{CS}|R_i(G)). \tag{3.4}$$

*Remark.* If $H_i$ in (2.1) is redefined to be

$$H_i^* = n_i^{-1} \sum_{j=1}^{n_i} Z(R_{ij})$$

and rules $R_1^*(G)$, $R_2^*(G)$ and $R_3^*(G)$ are defined by (2.5), (2.6) and (2.7) with $H_i$ replaced by $H_i^*$, $i = 1, \ldots, k$, then Theorem 3.1 holds with $R_i(G)$ replaced by $R_i^*(G)$. Thus, Theorem 3.1 is valid for randomized, as well as nonrandomized, selection procedures.

In the case of $R_3(G)$ we can say more on the infimum of the probability of a correct selection.

*Theorem* 3.2. *For the procedure* $R_3(G)$,

$$\inf_{\Omega} P(\mathrm{CS}|R_3(G)) = \inf_{\Omega_0} P(\mathrm{CS}|R_3(G)), \tag{3.5}$$

*where*

$$\Omega_0 = \{\theta \in \Omega: \theta_{[1]} = \ldots = \theta_{[k]}\}. \tag{3.6}$$

*Proof.* Let

$$\Psi(\mathrm{X}) = \begin{cases} 1 & \text{if } H_{(k)} \geqslant D \\ 0 & \text{otherwise.} \end{cases}$$

By an argument similar to the one employed in the proof of Theorem 3.1, we have that $E_\theta[\Psi(\mathrm{X})]$ is a nondecreasing function of $\theta_{[k]}$ and a nonincreasing function of $\theta_{[j]}$, $j = 1, \ldots, k-1$. This completes the proof.

For $\theta \in \Omega_0$, the quantity $P_\theta(\mathrm{CS}|R_i(G))$, $i = 1, 2, 3$ is independent of the common underlying distribution, $F_0(x)$. In other words, the distribution of the statistics $\max_{1 \leqslant j \leqslant k} H_j - H_i$ or $H_i$, or any other statistics involving $H_i$ does not depend on $F_0(x)$. It is in this sense that the procedures of this paper are distribution-free.

From Theorem 3.1, if $k = 2$, the probability of a correct selection using either rule $R_1(G)$ or $R_2(G)$ is minimized when the two populations are identically distributed. The same result is true in a slippage configuration, i.e. if $\theta_{[1]} = \ldots = \theta_{[k-1]}$ then the probability of a correct selection is minimized when $\theta_{[1]} = \theta_{[2]} = \ldots = \theta_{[k]}$.

It should be pointed out that a theorem similar to Theorem 3.2 involving $R_1(G)$ does not hold in general. This fact is established by means of a counterexample constructed by Rizvi and Woodworth [23] using distributions having two finite disjoint intervals for their supports and lacking the MLR property. McDonald [15] uses the same type of distributions to show that for $k = 3$, $P(\mathrm{CS}|R_1(G))$ need not be monotonic in $\theta_{[2]}$. The main difficulty arises out of the fact that the statistics $H_{(i)}$ are not independent.

We will next obtain bounds on $P(\mathrm{CS}|R_s(G))$, $s = 1, 2$, before investigating further the rule $R_3(G)$.

## 4 BOUNDS ON $P(\mathrm{CS}|R_i(G))$, $i = 1, 2$

We will assume that $n_i = n$, $i = 1, \ldots, k$. First consider rule $R_1(G)$. We have

$$H_i = n^{-1} \sum_{j=1}^{n} a(R_{ij}) \quad (i = 1, \ldots, k). \tag{4.1}$$

It is easy to see that

$$(k-1)^{-1} \sum_{j=1}^{k-1} H_{(j)} \leqslant \max_{1 \leqslant j \leqslant k-1} H_{(j)} \leqslant n^{-1} \sum_{r=N-n+1}^{N} a(r). \tag{4.2}$$

Using the inequalities (4.2) in the relation

$$P(\mathrm{CS}|R_1(G)) = P(H_{(k)} \geqslant \max_{1 \leqslant j \leqslant k-1} H_{(j)} - d),$$

we obtain

$$P\left(H_{(k)} \geqslant n^{-1} \sum_{r=N-n+1}^{N} a(r) - d\right) \leqslant P(\mathrm{CS}|R_1(G))$$
$$\leqslant P\left(H_{(k)} \geqslant (k-1)^{-1} \sum_{j=1}^{k-1} H_{(j)} - d\right). \tag{4.3}$$

Letting $\sum_{r=1}^{N} a(r) = A$, it follows that $\sum_{j=1}^{k} H_{(j)} = A/n$, a constant. Using this relation in (4.3), and defining

$$u(d,k,n) = [A - nd(k-1)]/nk, \tag{4.4}$$

$$v(d,k,n) = n^{-1} \sum_{r=N-n+1}^{N} a(r) - d, \tag{4.5}$$

we have
$$P(H_{(k)} \geqslant v) \leqslant P(\mathrm{CS}|R_1(G)) \leqslant P(H_{(k)} \geqslant u), \tag{4.6}$$
and hence

$$\inf_{\Omega} P(H_{(k)} \geqslant v) \leqslant \inf_{\Omega} P(\mathrm{CS}|R_1(G)) \leqslant \inf_{\Omega} P(H_{(k)} \geqslant u). \tag{4.7}$$

For the rule $R_2(G)$, we get a corresponding expression

$$\inf_{\Omega} P(H_{(k)} \geqslant v') \leqslant \inf_{\Omega} P(\mathrm{CS}|R_2(G)) \leqslant \inf_{\Omega} P(H_{(k)} \geqslant u'), \tag{4.8}$$

where
$$u'(d,k,n) = n^{-1}A[1+c(k-1)]^{-1} \tag{4.9}$$

and
$$v'(d,k,n) = (nc)^{-1} \sum_{r=N-n+1}^{N} a(r). \tag{4.10}$$

From Theorem 3.2, we know that the infima over $\Omega$ of expressions of the form $P(H_{(k)} \geqslant K)$ are attained when $\theta_{[1]} = \ldots = \theta_{[k]}$.

For the particular case where $a(r) = r$, we have

$$nH_i = \sum_{j=1}^{n} R_{ij} = T_i \quad \text{say}. \tag{4.11}$$

The $T_i$ are the rank-sum statistics, and in this case we denote the selection rules $R_j(G)$ by simply $R_j$. The expressions given above take the form

$$A = N(N+1)/2, \tag{4.12}$$

and
$$\sum_{r=N-n+1}^{N} a(r) = n(2N-n+1)/2. \tag{4.13}$$

Thus, equations (4.4), (4.5), (4.9) and (4.10) reduce to

$$u(d,k,n) = (N+1)/2 - d(k-1)/k, \tag{4.14}$$

$$v(d,k,n) = (2N-n+1)/2 - d, \tag{4.15}$$

$$u'(d,k,n) = k(N+1)/[2+2c(k-1)], \tag{4.16}$$

$$v'(d,k,n) = (2N-n+1)/2c. \tag{4.17}$$

In the special case $a(r) = r$ a more useful form of the lower bound appearing in (4.7) is given in the next theorem.

*Theorem* 4.1. *If $U$ is the Mann–Whitney statistic associated with samples of size $n$ and $(k-1)n$ taken from identically distributed populations, then in the case where $a(r) = r$,*

$$\inf_{\Omega} P(\text{CS}|R_1) \geqslant P(U \leqslant nd). \tag{4.18}$$

*Proof.* We first recall that the Mann–Whitney $U$ statistic, calculated from the samples $x_1, \ldots, x_p$ and $y_1, \ldots, y_q$ of sizes $p$ and $q$ from two independent populations, is the number of times an $x_i$ precedes a $y_j$. If $T_x$ denotes the rank-sum of the $x$'s in the combined sample, then $U$ and $T_x$ are related by

$$U + T_x = pq + p(p+1)/2. \tag{4.19}$$

In our present case with samples from $k$ populations, we need to evaluate $P(H_{(k)} \geqslant v)$ when all the populations are identical. Considering whether the observations came from the $\pi_{(k)}$ or any one of the rest, we have from (4.7), (4.14) and (4.19) with $p = n$ and $q = (k-1)n$,

$$\begin{aligned}\inf_{\Omega} P(\text{CS}|R_1) &\geqslant P(T_{(k)} \geqslant nv) \\ &= P(U \leqslant n^2(k-1) + n(n+1)/2 - nv) \\ &= P(U \leqslant nd).\end{aligned} \tag{4.20}$$

A similar theorem holds for rule $R_2(G)$.

Since $\sum_{j=1}^{k} H_{(j)} = A/n$, we see that

$$\max_{1 \leqslant j \leqslant k} H_j \geqslant A/nk. \tag{4.21}$$

Hence, a sufficient, but not necessary, condition for the selection rule $R_3(G)$ to select a nonempty subset is that $P^*$ be sufficiently large so that

$$D \leqslant A/N. \tag{4.22}$$

For large $n$, this sufficiency condition for rule $R_3(G)$ is satisfied if $P^* > \frac{1}{2}$. For rule $R_3$, i.e. when $a(r) = r$, the condition (4.22) is $D \leqslant (N+1)/2$. As an example, with $k = 3$, $n = 5$ the sufficient condition $D \leqslant 8$ is satisfied for $P^* \geqslant 0{\cdot}523$ and for such values a nonempty subset will be selected.

The evaluation of the constants $D = D(k, n, P^*)$ for the rule $R_3$ can be effected as follows:

$$P^* \leqslant P(T_i \geqslant Dn) = P(U \leqslant n^2(k-\tfrac{1}{2}) - n(D-\tfrac{1}{2})), \tag{4.23}$$

using (4.19) and considering all populations identically distributed. Hence, $D$ is the largest integer satisfying the inequality (4.23). The Mann–Whitney $U$-statistic has been well-tabulated by Milton [16] and others.

## 5 MOMENTS OF THE $H_i$

In this section we will derive the means, variances and covariances of the $H_i$ assuming the independent random variables $X_{ij}$ have the continuous distribution $F_i(x)$, $j = 1, \ldots, n_i$; $i = 1, \ldots, k$. Let $p^{(i)}_{j_1, \ldots, j_{n_i}}$ be the probability that the $n_i$ observations from the population $\pi_i$ have ranks $j_1, \ldots, j_{n_i}$ in the combined sample. Then,

$$E(H_i) = n_i^{-1} \sum_{(j_1, \ldots, j_{n_i})} [a(j_1) + \ldots + a(j_{n_i})] \, p^{(i)}_{j_1, \ldots, j_{n_i}} \quad (i = 1, \ldots, k), \tag{5.1}$$

where the summation is over all possible subsets of size $n_i$ in the set of integers 1 through $N$. Alternatively,

$$E(H_i) = n_i^{-1} \sum_{l=1}^{N} a(l) \, p_l^{(i)} \quad (i = 1, \ldots, k), \tag{5.2}$$

where $p_l^{(i)}$ is the probability that any one observation from $\pi_i$ has rank $l$ in the combined sample.

Let $p^{(i)}_{l,m}$ be the probability that any two of the observations from $\pi_i$ have ranks $l$ and $m$ in the combined sample. Then,

$$E(H_i^2) = n_i^{-2} \sum_{(j_1, \ldots, j_{n_i})} [a(j_1) + \ldots + a(j_{n_i})]^2 \, p^{(i)}_{j_1, \ldots, j_{n_i}}. \tag{5.3}$$

Alternatively,

$$E(H_i^2) = n_i^{-2} \left[ \sum_{l=1}^{N} (a(l))^2 p_l^{(i)} + 2 \sum_{l=1}^{N} \sum_{m=l+1}^{N} a(l) \, a(m) \, p^{(i)}_{l,m} \right]. \tag{5.4}$$

Hence,

$$\begin{aligned} n_i^2 \operatorname{var}(H_i) &= \sum_{l=1}^{N} (a(l))^2 p_l^{(i)} + 2 \sum_{l=1}^{N} \sum_{m=l+1}^{N} a(l) \, a(m) \, p^{(i)}_{l,m} - \left( \sum_{l=1}^{N} a(l) \, p_l^{(i)} \right)^2 \\ &= \sum_{l=1}^{N} (a(l))^2 p_l^{(i)} (1 - p_l^{(i)}) + 2 \sum_{l=1}^{N} \sum_{m=l+1}^{N} a(l) \, a(m) \, (p^{(i)}_{l,m} - p_l^{(i)} p_m^{(i)}). \end{aligned} \tag{5.5}$$

In a similar manner one can show that for $i \neq j$,

$$n_i n_j \operatorname{cov}(H_i, H_j) = \sum_{l=1}^{N} \sum_{m=1}^{N} a(l) \, a(m) \, (p^{(i,j)}_{l,m} - p_l^{(i)} p_m^{(j)}), \tag{5.6}$$

where $p^{(i,j)}_{l,m}$ is the probability that one observation from population $\pi_i$ has rank $l$ and one observation from $\pi_j$ has rank $m$. Note that $p^{(i,j)}_{l,l} = 0$, $i \neq j, l = 1, \ldots, N$.

As we see above, the computation of these moments depends upon the evaluation of $p_l^{(i)}$, $p^{(i)}_{l,m}$ and $p^{(i,j)}_{l,m}$. To evaluate $p_l^{(i)}$, choose one of the observations from $\pi_i$ to have rank $l$. Ranks $1, 2, \ldots, l-1$ are then assumed by $l-1$ of the remaining $N-1$ observations. These $l-1$ observations consist of $r_j$ observations from $\pi_j, j = 1, \ldots, k$, subject to the conditions

$$B: \begin{cases} 0 \leqslant r_i \leqslant n_i - 1, \\ 0 \leqslant r_j \leqslant n_j, \\ r_1 + \ldots + r_k = l-1, \end{cases} \quad (j = 1, \ldots, k; j \neq i). \tag{5.7}$$

Thus,
$$p_l^{(i)} = \sum_B n_i \binom{n_i - 1}{r_i} \int \left\{ \prod_{\substack{j=1 \\ j \neq i}}^{k} \binom{n_j}{r_j} F_j^{r_j} \bar{F}_j^{n_j - r_j} \right\} F_i^{r_i} \bar{F}_i^{n_i - r_i - 1} dF_i, \tag{5.8}$$

where $F_j \equiv F_j(x)$, $\bar{F}_j \equiv 1 - F_j(x), j = 1, \ldots, k$. Taking the summation inside the integral yields

$$p_l^{(i)} = n_i \int A_l^{(i)}(x)\, dF_i(x) \quad (i = 1, \ldots, k;\ l = 1, \ldots, N), \tag{5.9}$$

where
$$A_l^{(i)}(x) = \sum_B \binom{n_i - 1}{r_i} F_i^{r_i} \bar{F}_i^{n_i - r_i - 1} \left\{ \prod_{\substack{j=1 \\ j \neq i}}^{k} \binom{n_j}{r_j} F_j^{r_j} \bar{F}_j^{n_j - r_j} \right\}. \tag{5.10}$$

In a similar fashion we can obtain expressions for the probabilities $p_{l,m}^{(i)}$ and $p_{l,m}^{(i,j)}$ in terms of the given distributions and sample sizes.

In the special cases where $F_i(x) = F(x)$ and $n_i = n, i = 1, \ldots, k$, we have:

$$E(H_i) = N^{-1} \sum_{l=1}^{N} a(l) = A/N, \tag{5.11}$$

$$n^2 \operatorname{var}(H_i) = k^{-2}(k-1) \sum_{l=1}^{N} (a(l))^2 - 2(k-1)\, k^{-2}(N-1)^{-1} \sum_{l=1}^{N} \sum_{m=l+1}^{N} a(l)\, a(m), \tag{5.12}$$

$$n^2 \operatorname{cov}(H_i, H_j) = k^{-1}(N-1)^{-1} \sum_{\substack{l=1 \\ l \neq m}}^{N} \sum_{m=1}^{N} a(l)\, a(m) - k^{-2} A^2 \quad (i \neq j). \tag{5.13}$$

If, in addition, $a(l) = l, l = 1, \ldots, N$, then $nH_i = T_i$, and, hence,

$$E(T_i) = n(N+1)/2, \tag{5.14}$$

$$\operatorname{var}(T_i) = n^2(k-1)(N+1)/12, \tag{5.15}$$

$$\operatorname{cov}(T_i, T_j) = -n^2(N+1)/12 \quad (i \neq j), \tag{5.16}$$

which agree with the known expressions for this special case.

Asymptotic forms for the moments of $H_i$ have been given by Puri [19].

## 6 THE EXACT AND ASYMPTOTIC DISTRIBUTION OF $\max_{1 \leqslant j \leqslant k} T_j - T_i$ FOR IDENTICALLY DISTRIBUTED POPULATIONS

In this section the random variables $X_{ij}, j = 1, \ldots, n_i$; $i = 1, \ldots, k$, are assumed independent identically distributed with a continuous distribution $F(x)$. In this case the $H_i$ are exchangeable random variables if $n_i = n, i = 1, \ldots, k$. It should be noted that in a slippage-type configuration (see § 3), the constants required to implement rules $R_i(G)$, $i = 1, 2, 3$, are determined from the basic probability requirement $P(\mathrm{CS}|R_i(G)) \geqslant P^*$ calculated with identically distributed populations. But the exact distributions of the relevant statistics, e.g. $\max_{1 \leqslant j \leqslant k} H_j - H_1$, are not known for the general scores $a(R_{ij})$. However, in the case $a(R_{ij}) = R_{ij}$ the procedures $R_i(G)$ reduce to the rank sum procedures $R_i$, $i = 1, 2, 3$. The distribution of the statistic $\max_{1 \leqslant j \leqslant k} T_j - T_1$, both exact and asymptotic, is somewhat easier to obtain than the corresponding distribution of the statistic $\max_{1 \leqslant j \leqslant k} T_j / T_1$. For some results concerning the latter statistic, see McDonald [15]. Our concern here will be the former which is tantamount to considering rule $R_1$. Corresponding to rule $R_3$ is the statistic $T_1$, the distribution of which has been well-treated elsewhere.

For $k = 2$, the rules $R_1$, $R_2$ and $R_3$ are all equivalent. The constants required to implement these rules are obtained in a manner as described at the end of § 4. Some of these values are given in Table 2 where they are compared with asymptotic solutions.

Now suppose $k = 3$ and that we have $n_i$ observations from the $i$th population. The quantities $T_i$ can be obtained if each observation in the ordered sample is replaced by an $i$ if it came from the $i$th population. Now one has only to consider a sequence of length $\Sigma n_i$ consisting of $n_1$ 1's, $n_2$ 2's, and $n_3$ 3's. Since the random variables are identically distributed, each of the $\binom{\Sigma n_i}{n_1, n_2, n_3}$ different sequences are equally likely. Hence, to find $P[T_1 \geqslant \max_{1 \leqslant j \leqslant 3} T_j - m]$, it suffices to count the number of sequences which possess the attribute $[T_2 - T_1 \leqslant m, T_3 - T_1 \leqslant m]$. The recursion formula presented here is of the same type as that given by Odeh [17] in tabulating the distribution of the maximum rank sum. Let

$$S = n_1 + n_2 + n_3, \tag{6.1}$$

and define

$$N(n_1, n_2, n_3 | m_2, m_3)$$
$$= \text{number of sequences in which } T_2 - T_1 \leqslant m_2 \text{ and } T_3 - T_1 \leqslant m_3. \quad (6.2)$$

The following symmetry holds:

$$N(n_1, n_2, n_3 | m_2, m_3) = N(n_1, n_3, n_2 | m_3, m_2). \quad (6.3)$$

Then, by conditioning on the parent population of the last element in a sequence, the following recursion formula is obtained:

$$N(n_1, n_2, n_3 | m_2, m_3) = N(n_1 - 1, n_2, n_3 | m_2 + S, m_3 + S)$$
$$+ N(n_1, n_2 - 1, n_3 | m_2 - S, m_3) + N(n_1, n_2, n_3 - 1 | m_2, m_3 - S), \quad (6.4)$$

with the boundary conditions:

(1) If for any $i \geqslant 2$, $m_i < [n_i(n_i+1) - n_1(1+2S-n_1)]/2$, then

$$N(n_1, n_2, n_3 | m_2, m_3) = 0.$$

(2) If for every $i \geqslant 2$, $m_i \geqslant [n_i(1+2S-n_i) - n_1(n_1+1)]/2$, then

$$N(n_1, n_2, n_3 | m_2, m_3) = \binom{S}{n_1, n_2, n_3}.$$

(3) $N(0, n_2, n_3 | m_2, m_3)$ = number of sequences of $n_2$ 2's and $n_3$ 3's such that $S(S+1)/2 - m_3 \leqslant T_2 \leqslant m_2$, so

(*a*) if $S(S+1)/2 - m_3 > m_2$, $N(0, n_2, n_3 | m_2, m_3) = 0$,

(*b*) if $m_2 < n_2(n_2+1)/2$, $N(0, n_2, n_3 | m_2, m_3) = 0$,

(*c*) if $m_3 < n_3(n_3+1)/2$, $N(0, n_2, n_3 | m_2, m_3) = 0$,

(*d*) if (*a*) through (*c*) do not hold, the term can be evaluated from a Mann–Whitney table.

(4) $N(n_1, 0, n_3 | m_2, m_3)$ = number of sequences of $n_1$ 1's and $n_3$ 3's such that $T_1 \geqslant \max\{-m_2, L(S(S+1)/4 - m_3/2)\} \equiv M$, where $L(x)$ is the smallest integer not less than $x$, so

(*a*) if $M > n_1(n_1 + 2n_3 + 1)/2$, $N(n_1, 0, n_3 | m_2, m_3) = 0$,

(*b*) if $M \leqslant n_1(n_1+1)/2$, $N(n_1, 0, n_3 | m_2, m_3) = \binom{S}{n_1}$,

(*c*) if (*a*) and (*b*) fail to hold, the term can be evaluated from a Mann–Whitney table.

(5) $N(n_1, n_2, 0 | m_2, m_3) = N(n_1, 0, n_2 | m_3, m_2)$, so condition (4) applies.

It follows from (6.3) that at an 'equal $n_i$, equal $m_i$ stage', equation (6.4) can be written as

$$N(n, n, n | m, m) = N(n-1, n, n | m + 3n, m + 3n)$$
$$+ 2N(n, n-1, n | m - 3n, m). \quad (6.5)$$

In order to get $P[T_1 \geqslant \max_{1\leqslant j\leqslant 3} T_j - m]$ for values of $m \geqslant 0$, one uses the following relation:

$$P[T_1 \geqslant \max_{1\leqslant j\leqslant 3} T_j - m] = N(n_1, n_2, n_3|m, m)\binom{S}{n_1, n_2, n_3}^{-1}. \qquad (6.6)$$

A recursion formula similar to (6.4) can be written for an arbitrary number of populations. The quantity $N(n, n, n|m, m)$ was computed for $n = 2, 3, 4, 5$; $m = 0, 1, \ldots, 2n^2$. Using (6.6), $P[T_1 \geqslant \max_{1\leqslant j\leqslant 3} T_j - m]$ was then obtained to five decimal places, the fifth being rounded. These computations are given in Table 1.

Asymptotically, we have the following theorem as a special case of a more general result applying to the statistics $H_i$ with populations not necessarily identically distributed. The proof follows directly from Puri [19] and is omitted.

*Theorem* 6.1. *Let $X_{ij}, j = 1, \ldots, n$; $i = 1, \ldots, k$, be independent identically distributed random variables with a continuous distribution function. Then*

$$P[T_k \geqslant \max_{1\leqslant j\leqslant k} T_j - m] \approx \int_{-\infty}^{\infty} [\Phi(x+m/z)]^{k-1}\phi(x)\,dx \quad (m \geqslant 0), \qquad (6.7)$$

*where $\Phi(\cdot)$ and $\phi(\cdot)$ are the cumulative distribution function and density of a standard normal random variable, respectively, and*

$$z = z(n, k) = n[k(nk+1)/12]^{\frac{1}{2}}. \qquad (6.8)$$

Integrals of the type appearing on the right-hand side of equation (6.7) have been considered by Gupta [8]. Table 1 in [8] gives $h$ values satisfying the equation

$$\int_{-\infty}^{\infty} [\Phi(x+h\sqrt{2})]^{k-1}\phi(x)\,dx = P^* \qquad (6.9)$$

for $P^* = 0{\cdot}99$, $0{\cdot}975$, $0{\cdot}95$, $0{\cdot}90$, $0{\cdot}75$ and $k = 2(1)51$. If $\tilde{m}$ denotes the value of $m$ based on the normal approximation, then from (6.9) one obtains

$$\tilde{m} = hn[k(nk+1)/6]^{\frac{1}{2}}, \qquad (6.10)$$

$h$ being the entry of Table 1 of [8] corresponding to the given $P^*$ and $k$.

*Remarks.* (1) By using (6.10) one can obtain an asymptotic value of $m$ (and, hence, $d$) in rule $R_1$ when a slippage configuration in $\Omega$ exists (as shown in § 3) for $k = 2(1)51$ and for any common sample size $n$, $n$ large. (2) In general $\tilde{m}$ will not be an integer. So for the solution the smallest integer not less than $\tilde{m}$, $L(\tilde{m})$ should be taken. This method was used to calculate an asymptotic value of $m$ for $k = 2, 3$; $n = 2(1)25$ and $P^* = 0{\cdot}99$, $0{\cdot}975$,

Table 1. *For a given $n$, the left column is the value $N(n,n,n|m,m)$, i.e. the number of sequences in which $\max_{1\leqslant j\leqslant 3} T_j - T_1 \leqslant m$; the right column is the quantity $P(T_1 \geqslant \max_{1\leqslant j\leqslant 3} T_j - m)$. The rank sums $T_i, i = 1,2,3$ are based on random variables $X_{ij}, j = 1,\ldots,n; i = 1,2,3$, which are independent identically distributed*

| | $n = 2$ | | $n = 3$ | | $n = 4$ | | $n = 5$ | |
|---|---|---|---|---|---|---|---|---|
| $m = 0$ | 38 | 0·42222 | 600 | 0·35714 | 12,268 | 0·35405 | 262,686 | 0·34712 |
| 1 | 44 | 0·48889 | 702 | 0·41786 | 13,500 | 0·38961 | 283,426 | 0·37453 |
| 2 | 54 | 0·60000 | 808 | 0·48095 | 14,958 | 0·43169 | 305,560 | 0·40378 |
| 3 | 62 | 0·68889 | 912 | 0·54286 | 16,322 | 0·47105 | 327,738 | 0·43308 |
| 4 | 70 | 0·77778 | 1,004 | 0·59762 | 17,734 | 0·51180 | 349,236 | 0·46149 |
| 5 | 76 | 0·84444 | 1,112 | 0·66190 | 19,162 | 0·55302 | 372,410 | 0·49211 |
| 6 | 84 | 0·93333 | 1,206 | 0·71786 | 20,588 | 0·59417 | 394,770 | 0·52166 |
| 7 | 88 | 0·97778 | 1,294 | 0·77024 | 21,884 | 0·63157 | 416,774 | 0·55074 |
| 8 | 90 | 1·00000 | 1,374 | 0·81786 | 23,274 | 0·67169 | 439,432 | 0·58068 |
| 9 | | | 1,438 | 0·85595 | 24,500 | 0·70707 | 461,534 | 0·60988 |
| 10 | | | 1,490 | 0·88690 | 25,708 | 0·74193 | 482,468 | 0·63755 |
| 11 | | | 1,544 | 0·91905 | 26,846 | 0·77478 | 504,104 | 0·66614 |
| 12 | | | 1,584 | 0·94286 | 27,956 | 0·80681 | 524,454 | 0·69303 |
| 13 | | | 1,618 | 0·96309 | 28,906 | 0·83423 | 543,924 | 0·71876 |
| 14 | | | 1,644 | 0·97857 | 29,842 | 0·86124 | 563,152 | 0·74417 |
| 15 | | | 1,664 | 0·99048 | 30,616 | 0·88358 | 581,280 | 0·76812 |
| 16 | | | 1,674 | 0·99643 | 31,336 | 0·90436 | 598,016 | 0·79024 |
| 17 | | | 1,678 | 0·99881 | 31,952 | 0·92214 | 614,640 | 0·81220 |
| 18 | | | 1,680 | 1·00000 | 32,496 | 0·93784 | 629,818 | 0·83226 |
| 19 | | | | | 32,936 | 0·95053 | 643,940 | 0·85092 |
| 20 | | | | | 33,338 | 0·96214 | 657,292 | 0·86857 |
| 21 | | | | | 33,644 | 0·97097 | 669,588 | 0·88481 |
| 22 | | | | | 33,912 | 0·97870 | 680,548 | 0·89929 |
| 23 | | | | | 34,126 | 0·98488 | 690,952 | 0·91304 |
| 24 | | | | | 34,296 | 0·98978 | 700,152 | 0·92520 |
| 25 | | | | | 34,424 | 0·99348 | 708,408 | 0·93611 |
| 26 | | | | | 34,518 | 0·99619 | 715,862 | 0·94596 |
| 27 | | | | | 34,574 | 0·99781 | 722,454 | 0·95467 |
| 28 | | | | | 34,614 | 0·99896 | 728,098 | 0·96213 |
| 29 | | | | | 34,634 | 0·99954 | 733,120 | 0·96877 |
| 30 | | | | | 34,644 | 0·99983 | 737,408 | 0·97443 |
| 31 | | | | | 34,648 | 0·99994 | 741,054 | 0·97925 |
| 32 | | | | | 34,650 | 1·00000 | 744,192 | 0·98339 |
| 33 | | | | | | | 746,838 | 0·98689 |
| 34 | | | | | | | 749,008 | 0·98976 |
| 35 | | | | | | | 750,826 | 0·99216 |
| 36 | | | | | | | 752,296 | 0·99411 |
| 37 | | | | | | | 753,464 | 0·99565 |
| 38 | | | | | | | 754,402 | 0·99689 |
| 39 | | | | | | | 755,130 | 0·99785 |
| 40 | | | | | | | 755,660 | 0·99855 |
| 41 | | | | | | | 756,048 | 0·99906 |
| 42 | | | | | | | 756,316 | 0·99942 |
| 43 | | | | | | | 756,496 | 0·99966 |
| 44 | | | | | | | 756,608 | 0·99980 |
| 45 | | | | | | | 756,680 | 0·99989 |
| 46 | | | | | | | 756,720 | 0·99995 |
| 47 | | | | | | | 756,740 | 0·99998 |
| 48 | | | | | | | 756,750 | 0·99999 |
| 49 | | | | | | | 756,754 | 0·99999 |
| 50 | | | | | | | 756,756 | 1·00000 |

0·95, 0·90, 0·75. These results are presented in Table 2. Exact $m$ values are given in parentheses where they are known. In most cases where the asymptotic value and exact value do not agree, the asymptotic value is larger and, hence, a conservative constant for the rule $R_1$. From the values given in this table, it is seen that $-1 \leqslant L(\tilde{m}) - m \leqslant 3$ for $k = 2$, and $0 \leqslant L(\tilde{m}) - m \leqslant 3$ for $k = 3$.

## 7 REMARKS ON THE PROPERTIES OF THE SELECTION RULES

### Expected size of the selected subset

All the selection rules discussed in this paper select a subset of size $S$, where $S$ is an integer-valued random variable. Since $R_1(G)$ and $R_2(G)$

Table 2. *For given values of $k$, $n$, $P^*$, this table gives the smallest integer $m$ based on asymptotic theory which satisfies $P[T_k \geqslant \max_{1 \leqslant j \leqslant k} T_j - m] \geqslant P^*$. The rank sums $T_i, i = 1, \ldots, k$, are based on random variables $X_{ij}, j = 1, \ldots, n$; $i = 1, \ldots, k$, which are independent identically distributed. Exact $m$ values, where known, are given in parentheses*

| | $P^*$ | | | | |
|---|---|---|---|---|---|
| $n$ | 0·99 | 0·975 | 0·95 | 0·90 | 0·75 |
| | | | $k = 2$ | | |
| 2 | 7 | 6 | 5 | 4 | 2 (2) |
| 3 | 11 | 9 | 8 (7) | 6 (5) | 4 (3) |
| 4 | 17 | 14 (14) | 12 (12) | 9 (8) | 5 (4) |
| 5 | 23 (21) | 19 (19) | 16 (15) | 13 (13) | 7 (7) |
| 6 | 30 (28) | 25 (24) | 21 (20) | 17 (16) | 9 (8) |
| 7 | 37 (35) | 31 (31) | 26 (25) | 21 (21) | 11 (11) |
| 8 | 45 (44) | 38 (36) | 32 (32) | 25 (24) | 13 (14) |
| 9 | 53 (51) | 45 (45) | 38 (37) | 30 (29) | 16 (15) |
| 10 | 62 (60) | 52 (52) | 44 (44) | 34 (34) | 18 (18) |
| 11 | 71 (69) | 60 (59) | 51 (51) | 40 (39) | 21 |
| 12 | 81 (80) | 68 (68) | 57 (58) | 45 (44) | 24 |
| 13 | 91 (89) | 77 (77) | 65 (65) | 50 (51) | 27 |
| 14 | 102 (100) | 86 (84) | 72 (72) | 56 (56) | 30 |
| 15 | 113 (111) | 95 (95) | 80 (79) | 62 (63) | 33 |
| 16 | 124 (122) | 105 (104) | 88 (88) | 69 (68) | 36 |
| 17 | 136 (133) | 114 (113) | 96 (95) | 75 (75) | 40 |
| 18 | 148 (146) | 124 (124) | 104 (104) | 82 (82) | 43 |
| 19 | 160 (157) | 135 (133) | 113 (113) | 88 (89) | 47 |
| 20 | 172 (170) | 145 (144) | 122 (122) | 95 (96) | 50 |
| 21 | 185 | 156 | 131 | 102 | 54 |
| 22 | 199 | 168 | 141 | 110 | 58 |
| 23 | 212 | 179 | 150 | 117 | 62 |
| 24 | 226 | 191 | 160 | 125 | 66 |
| 25 | 240 | 203 | 170 | 133 | 70 |

Table 2 (*cont.*)

| | $k = 3$ | | | | |
|---|---|---|---|---|---|
| 2 | 10 (8) | 9 (7) | 8 (7) | 6 (6) | 4 (4) |
| 3 | 18 (15) | 15 (14) | 13 (13) | 11 (11) | 7 (7) |
| 4 | 27 (25) | 23 (22) | 20 (19) | 17 (16) | 11 (11) |
| 5 | 37 (35) | 32 (31) | 28 (27) | 23 (23) | 15 (15) |
| 6 | 48 | 41 | 36 | 30 | 19 |
| 7 | 60 | 52 | 45 | 37 | 24 |
| 8 | 73 | 63 | 55 | 45 | 29 |
| 9 | 87 | 75 | 65 | 54 | 35 |
| 10 | 101 | 88 | 76 | 63 | 40 |
| 11 | 117 | 101 | 87 | 72 | 46 |
| 12 | 133 | 115 | 99 | 82 | 53 |
| 13 | 149 | 129 | 112 | 92 | 59 |
| 14 | 167 | 144 | 125 | 103 | 66 |
| 15 | 185 | 160 | 138 | 114 | 73 |
| 16 | 203 | 176 | 152 | 125 | 81 |
| 17 | 222 | 192 | 167 | 137 | 88 |
| 18 | 242 | 209 | 181 | 149 | 96 |
| 19 | 262 | 227 | 197 | 162 | 104 |
| 20 | 283 | 245 | 212 | 175 | 112 |
| 21 | 304 | 263 | 228 | 188 | 121 |
| 22 | 326 | 282 | 244 | 201 | 130 |
| 23 | 349 | 301 | 261 | 215 | 138 |
| 24 | 371 | 321 | 278 | 229 | 148 |
| 25 | 395 | 341 | 296 | 244 | 157 |

select non-empty subsets, $S$ in these cases takes values 1 through $k$. As pointed out in § 4, $R_3(G)$ under certain conditions will select a non-empty subset; but generally for $R_3(G)$, $S$ takes values 0 through $k$. For all these rules:

$$E(S) \equiv E(S|R_i(G)) = \sum_{j=1}^{k} p_j(R_i(G))$$

$$= P(\mathrm{CS}|R_i(G)) + \sum_{j=1}^{k-1} p_j(R_i(G)), \tag{7.1}$$

where $p_j(R_i(G))$ is as defined in Theorem 3.1. In general, it is difficult to obtain the exact expressions for $E(S)$. But asymptotic expressions can be obtained. We consider $R_1$ and $R_3$. Assuming $n_i = n$, for large $n$, the distribution of $\mathbf{T}' = (T_1, \ldots, T_k)$ is approximately a multivariate normal distribution with mean vector $\boldsymbol{\mu}'_T = (\mu_1, \ldots, \mu_k)$ and variance-covariance matrix $\Sigma_T = (\sigma_{ij})$, where

$$\mu_i = E(T_i), \quad \sigma_i^2 = \mathrm{var}\,(T_i) \quad \text{and} \quad \sigma_{ij} = \mathrm{cov}\,(T_i, T_j);\ i,j = 1, \ldots, k;\ i \neq j.$$

Let

$$\mathbf{W} = A\mathbf{T}, \tag{7.2}$$

where $A$ is a $(k-1)\times k$ matrix given by

$$A = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 & -1 \\ 0 & 1 & 0 & \dots & 0 & -1 \\ \vdots & \vdots & & & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & -1 \end{pmatrix}. \tag{7.3}$$

Thus $$W_i = T_i - T_k \quad (i = 1, \dots, k-1). \tag{7.4}$$

Then, for large $n$, $\mathbf{W}' = (W_1, \dots, W_{k-1})$ is approximately distributed as a multivariate normal random vector with mean vector given by $\mu_W = A\,\mu_T$ and the variance-covariance matrix $\Sigma_W = A\,\Sigma_T A'$.

Now, for $\nu = 1, \dots, k$, we define

$$\mathbf{W}^\nu = A_\nu \mathbf{T}, \tag{7.5}$$

where $A_\nu$ is the $(k-1)\times k$ matrix obtained from matrix $A$ defined in (7.3) by moving column $j$ to column $j+1$, $j = \nu, \nu+1, \dots, k-1$ and replacing column $\nu$ by column $k$. The matrix $A_k$ is $A$. Thus,

$$W_i^\nu = T_i - T_\nu \quad (i = 1, \dots, k;\ i \neq \nu). \tag{7.6}$$

The random vector $\mathbf{W}^\nu$ is asymptotically distributed as a multivariate normal random vector with mean vector $\mu_\nu = A_\nu\,\mu_T$ and variance-covariance matrix $\Sigma_\nu = A_\nu \Sigma_T A_\nu'$. Hence, we can state

*Theorem* 7.1. *If $\Sigma_\nu$ is non-singular for $\nu = 1, \dots, k$; then*

$$E(S|R_1) \approx \sum_{\nu=1}^{k} K_\nu \int_{-\infty}^{d} \dots \int_{-\infty}^{d} \exp\left[-(\mathbf{w}^\nu - \mu_\nu)' \Sigma_\nu^{-1} (\mathbf{w}^\nu - \mu_\nu)/2\right] \prod_{\substack{i=1 \\ i\neq\nu}}^{k} dw_i^\nu, \tag{7.7}$$

*where* $K_\nu = [(2\pi)^{k-1} |\Sigma_\nu|]^{-\frac{1}{2}}$. *For $R_3$, we have*

$$E(S|R_3) \approx \sum_{\nu=1}^{k} \Phi[(\mu_\nu - D)/\sigma_\nu]. \tag{7.8}$$

A similar result can be derived for rule $R_2$.

## Some Monte Carlo results

In order to compare the performance of selection rules $R_1$ and $R_3$, some Monte Carlo studies were made. Normal and logistic distributions with variance unity were studied for different configurations of their means. For $k = 3$ and $n = 2, 3, 4$, these configurations were taken to be $(0{\cdot}1, 0, 0)$, (0·2, 0, 0), (0·5, 0, 0), (1·0, 0, 0), (2·0, 0, 0), (0·1, 0·1, 0), (0·2, 0·2, 0), (0·5, 0·5, 0), (1·0, 1·0, 0), (2·0, 2·0, 0). The number of simulations were 500 or 1000. The logistic distribution was chosen because equally spaced scores such as ranks yield locally most powerful tests for the location parameter of this distribution. Since there are only a finite number of different possible $d$ and $D$ values for rules $R_1$ and $R_3$, there are also only a finite number of values for $\inf_{\Omega} P(\mathrm{CS}|R_1)$ and $\inf_{\Omega} P(\mathrm{CS}|R_3)$; therefore, in general, it is not possible to determine $d$ and $D$ such that both infima yield the same $P^*$. In such cases there is no definite way of comparing these two rules. For our purposes $d$ and $D$ were determined such that they yielded approximately the same $P^*$ in the case of identical distributions. Then the ratio of $kP(\mathrm{CS}|R)$ and $E(S|R)$ was computed for both rules $R_1$ and $R_3$. The bigger ratio for a rule indicates it to be better than the other. For example, for $k = 3$, $n = 2$, then $D = 2$ and $d = 3$ give the probability 14/15 for the identical case. Using the configuration (0·1, 0, 0), we find that for the normal means the two ratios are 1·012 for $R_1$ and 1·005 for $R_3$ so that $R_1$ seems slightly better than $R_3$. Using the configuration (0·5, 0, 0), it was found that $R_3$ was slightly better than $R_1$; the ratios being 1·045 for $R_1$ and 1·049 for $R_3$.

Our Monte Carlo studies showed no significant uniform superiority of either of these procedures. However, $R_3$ seemed to perform slightly better than $R_1$ in the cases where the two highest parameters are equal. No difference in the performance of $R_1$ and $R_3$ was noticeable when we changed from logistic to normal populations. In all cases the frequency of correct selections for $R_1$ was higher than the theoretical value exactly calculated for the identical distributions. Thus, there was no indication that the infimum of the probability of a correct selection does not take place when all populations are identically distributed as normal or logistic distributions under shift in location.

## Local optimality and monotonicity of $R_3$

If we use the scores which lead to locally most powerful rank tests of the hypothesis $\boldsymbol{\theta} = (0, 0, \ldots, 0)$ against $\boldsymbol{\theta} = \alpha(\theta_1, \theta_2, \ldots, \theta_k)$ (see, Hajék and Šidák [11]), then rules of the type $R_3$ have the property that the probability of a correct selection increases fastest among all rules based on ranks in the neighborhood of $\boldsymbol{\theta} = (0, 0, \ldots, 0)$. For example, the locally most powerful rank test for shift in location of the logistic distribution is based on rank sums $T_i$ which for the two-sample case is the Wilcoxon or Mann–Whitney statistic. Hence, the selection rule $R_3$ based on rank sums is locally optimum in the above sense, provided the underlying distributions are logistic differing only in their location parameters. This result has been shown by K. Nagel.†

A selection rule is called monotone if $\theta_i \geqslant \theta_j$ implies that the population with parameter $\theta_i$ is selected with larger probability than that with parameter $\theta_j$. It can be shown that $R_3$ is monotone if one uses nondecreasing scores and if $F_\theta(x)$ is a stochastically increasing family of distributions.

## Asymptotic relative efficiency (ARE) of the rules $R_1$, $R_2$ and $R_3$ relative to a normal means procedure $R$

We consider here the case of two populations, and so the rules $R_1$, $R_2$ and $R_3$ are equivalent. Hence, we will be concerned with $R_1$ and $R$ here. Suppose $\pi_1$ and $\pi_2$ are two independent normal populations with common variance unity. Let the means of $\pi_{(1)}$ and $\pi_{(2)}$ be 0 and $\theta(\geqslant 0)$, respectively. A sample of size $n$ is drawn from each population. Based on

$$X_{ij}, j = 1, \ldots, n \quad (i = 1, 2),$$

let $T_i$ and $\bar{X}_i$ be the rank sum and sample mean, respectively, from $\pi_i$, $i = 1, 2$. The procedures to be compared are:

$$R_1\text{: select } \pi_i \text{ iff } T_i \geqslant \max_{j=1,2} T_j - nd \quad (d \geqslant 0), \tag{7.9}$$

$$R\text{: select } \pi_i \text{ iff } \bar{X}_i \geqslant \max_{j=1,2} \bar{X}_j - b \quad (b \geqslant 0). \tag{7.10}$$

The constants $d$ and $b$ are chosen so that the probability of a correct selection is bounded below by a given number $P^*$, $\frac{1}{2} < P^* < 1$, for all $\theta$; i.e.

$$\inf_{\theta \geqslant 0} P(\pi_{(2)} \text{ is selected}) \geqslant P^*. \tag{7.11}$$

Procedure $R$ has been investigated by Gupta [10]. Let $S^*$ denote the number of nonbest populations in the selected subset. Since

$$T_{(1)} + T_{(2)} = n(2n+1),$$

$$\begin{aligned} E(S^*|R_1) &= P(T_{(1)} \geqslant T_{(2)} - nd) \\ &= P\{\sigma^{-1}(T_{(1)} - \mu) \geqslant -\sigma^{-1}[\mu - n(2n+1-d)/2]\} \\ &\approx \Phi\{\sigma^{-1}[\mu - n(2n+1-d)/2]\}, \end{aligned} \tag{7.12}$$

† On subset selection rules with certain optimality properties. Department of Statistics, Purdue University. Mimeograph Series No. 222 (1970).

where $$\mu = E(T_{(1)}) = n(3n+1)/2 - n^2\Phi(\theta 2^{-\frac{1}{2}}), \tag{7.13}$$

$$\sigma^2 = \text{var}\,(T_{(1)})$$
$$= n^2\left[\Phi(\theta 2^{-\frac{1}{2}}) + 2(n-1)\int_{-\infty}^{\infty}\Phi^2(x+\theta)\,\phi(x)\,dx - (2n-1)\,\Phi^2(\theta 2^{-\frac{1}{2}})\right]. \tag{7.14}$$

These moments can be obtained from § 5. For $k = 2$ moment expressions are also given in Wilks [24, p. 460]. Now set the right-hand side of (7.12) equal to $\epsilon > 0$ and obtain

$$\mu - n(2n+1-d)/2 = \sigma\Phi^{-1}(\epsilon). \tag{7.15}$$

From Theorem 3.1, the appropriate value of $d$ is obtained from (7.11) when $\theta = 0$. Equation (6.10) provides a large sample solution for $d$; namely,

$$d \approx h_1 n^{\frac{1}{2}}, \tag{7.16}$$

where $h_1$ is independent of $n$ and $\theta$. Actually $h_1 = h(2/3)^{\frac{1}{2}}$, where $h$ is the appropriate value obtained from Gupta [8]. Using (7.13), (7.14) and (7.16) in (7.15) and simplifying yields

$$n + h_1 n^{\frac{1}{2}} - 2n\Phi(\theta 2^{-\frac{1}{2}})$$
$$= 2\Phi^{-1}(\epsilon)\left[\Phi(\theta 2^{-\frac{1}{2}}) + 2(n-1)\int_{-\infty}^{\infty}\Phi^2(x+\theta)\,\phi(x)\,dx - (2n-1)\,\Phi^2(\theta 2^{-\frac{1}{2}})\right]^{\frac{1}{2}}, \tag{7.17}$$

or upon rearrangement,

$$n(1 - 2\Phi(\theta 2^{-\frac{1}{2}})) + h_1 n^{\frac{1}{2}} = 2\Phi^{-1}(\epsilon)\,(2nB^2(\theta) + R(\theta))^{\frac{1}{2}}, \tag{7.18}$$

where $$B^2(\theta) = \int_{-\infty}^{\infty}\Phi^2(x+\theta)\,\phi(x)\,dx - \Phi^2(\theta 2^{-\frac{1}{2}}), \tag{7.19}$$

$$R(\theta) = \Phi(\theta 2^{-\frac{1}{2}}) - 2\int_{-\infty}^{\infty}\Phi^2(x+\theta)\,\phi(x)\,dx + \Phi^2(\theta 2^{-\frac{1}{2}}). \tag{7.20}$$

For large $n$, the $R(\theta)$ term in (7.18) can be ignored and then that equation simplifies to

$$n^{\frac{1}{2}} \approx [2^{\frac{3}{2}}\Phi^{-1}(\epsilon)\,B(\theta) - h_1]\,[1 - 2\Phi(\theta 2^{-\frac{1}{2}})]^{-1}. \tag{7.21}$$

Thus, $$n \equiv n_{R_1}(\epsilon) \approx [2^{\frac{3}{2}}\Phi^{-1}(\epsilon)\,B(\theta) - h_1]^2\,[1 - 2\Phi(\theta 2^{-\frac{1}{2}})]^{-2}. \tag{7.22}$$

Now consider rule $R$:

$$E(S^*|R) = P[\bar{X}_{(1)} \geqslant \bar{X}_{(2)} - b] = \Phi[(b-\theta)\,(n/2)^{\frac{1}{2}}]. \tag{7.23}$$

Again, $b$ is obtained from (7.11) when $\theta = 0$ and is given by

$$b = h_1(3/n)^{\frac{1}{2}}. \tag{7.24}$$

Setting the right-hand side of (7.23) equal to $\epsilon$ and using (7.24) yields

$$n \equiv n_R(\epsilon) = [(3^{\frac{1}{2}}h_1 - 2^{\frac{1}{2}}\Phi^{-1}(\epsilon))/\theta]^2. \tag{7.25}$$

The asymptotic relative efficiency of $R_1$ relative to $R$ is defined to be

$$\mathrm{ARE}(R_1, R;\, \theta) = \lim_{\epsilon \downarrow 0} [n_R(\epsilon)/n_{R_1}(\epsilon)]. \tag{7.26}$$

From equations (7.22) and (7.25),

$$\mathrm{ARE}(R_1, R;\, \theta) = \{[2\Phi(\theta 2^{-\frac{1}{2}}) - 1]/2\theta B(\theta)\}^2. \tag{7.27}$$

If $\theta$ is allowed to decrease to 0, then

$$\lim_{\theta \downarrow 0} \mathrm{ARE}(R_1, R;\, \theta) = 3/\pi = 0{\cdot}9549. \tag{7.28}$$

## Asymptotic relative efficiency of $R_2$ relative to Gupta's gamma procedure $R'$

Let $\pi_1, \pi_2$ be two independent exponential populations with independent associated random variables $X_{ij}, j = 1, \ldots, n;\, i = 1, 2$. The density function of $X_{ij}$ is

$$f_i(x) = \begin{cases} \theta_i^{-1} e^{-x/\theta_i} & (x > 0,\, i = 1, 2), \\ 0 & (x \leqslant 0), \end{cases} \tag{7.29}$$

where $1 = \theta_{[1]} \leqslant \theta_{[2]} = \theta$.

Procedure $R_2$ is given by (2.6) with $T_i$ in the place of $H_i$ and Procedure $R'$ is given by

$$R'\colon \text{select } \pi_i \text{ iff } \bar{X}_i \geqslant b^{-1} \max_{j=1,2} \bar{X}_j \quad (b \geqslant 1). \tag{7.30}$$

The constants $c$ and $b$ are chosen so that

$$\inf_{\theta \geqslant 1} P(\pi_{(2)} \text{ is selected}) \geqslant P^*. \tag{7.31}$$

Procedure $R'$ has been studied by Gupta [9]. Computing $n_{R_2}(\epsilon)$ and $n_{R'}(\epsilon)$ as before in the case of $R_1$ and $R$, we obtain

$$n_{R_2}(\epsilon) \approx 4(\theta+1)^2\,(\theta-1)^{-2}\,[6^{-\frac{1}{2}}\Phi^{-1}(P^*) - B(\theta)\,\Phi^{-1}(\epsilon)]^2 \tag{7.32}$$

and

$$n_{R'}(\epsilon) = 2(\log\theta)^{-2}\,[\Phi^{-1}(\epsilon) - \Phi^{-1}(P^*)]^2, \tag{7.33}$$

where

$$B^2(\theta) = 1 - 2(1+\theta)^{-1} + (2\theta+1)^{-1} + \theta(2+\theta)^{-1} - 2\theta^2(1+\theta)^{-2}. \tag{7.34}$$

Hence,

$$\mathrm{ARE}(R_2, R';\, \theta) = \lim_{\epsilon \downarrow 0} [n_{R'}(\epsilon)/n_{R_2}(\epsilon)] = [(\theta-1)/4(\theta+1)\,B(\theta)\log\theta]^2. \tag{7.35}$$

Letting $\theta$ decrease to 1 yields

$$\lim_{\theta \downarrow 1} \mathrm{ARE}(R_2, R';\, \theta) = \tfrac{3}{4}. \tag{7.36}$$

ACKNOWLEDGEMENTS

The authors are very thankful to Dr S. Panchapakesan and Mr K. Nagel for valuable assistance during the writing of this paper. Thanks are also due to Mr M. Ward for his assistance in some of the programming work and to Mrs P. McLaughlin for typing earlier drafts as well as the final paper.

## REFERENCES

[1] Alam, K. and Rizvi, M. H. (1966). Selection from multivariate normal populations. *Ann. Inst. Statist. Math., Tokyo* 18, 307–18.

[2] Barlow, R. E. and Gupta, S. S. (1969). Selection procedures for restricted families of probability distributions. *Ann. Math. Statist.* 40, 905–17.

[3] Barlow, R. E., Gupta, S. S. and Panchapakesan, S. (1969). On the distribution of the maximum and minimum of ratios of order statistics. *Ann. Math. Statist.* 40, 918–34.

[4] Bartlett, N. S. and Govindarajulu, Z. (1968). Some distribution-free statistics and their application to the selection problem. *Ann. Inst. Statist. Math., Tokyo* 20, 79–97.

[5] 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.

[6] Dudewicz, E. J. (1966). The efficiency of a nonparametric selection procedure: largest location parameter case. Department of Industrial Engineering and Operations Research, Cornell University, Technical Report No. 14.

[7] Gupta, S. S. (1956). On a decision rule for a problem in ranking means, Institute of Statistics, University of North Carolina, Chapel Hill, N.C. Mimeograph, Series No. 150.

[8] Gupta, S. S. (1963). Probability integrals of multivariate normal and multivariate *t*. *Ann. Math. Statist.* 34, 792–828.

[9] Gupta, S. S. (1963). On a selection and ranking procedure for gamma populations. *Ann. Inst. Statist. Math., Tokyo* 14, 199–216.

[10] Gupta, S. S. (1965). On some multiple decision rules. *Technometrics* 7, 225–45.

[11] Hájek, J. and Šidák, Z. (1967). *Theory of Rank Tests.* New York: Academic Press.

[12] Lehmann, E. L. (1959). *Testing Statistical Hypotheses.* New York: John Wiley

[13] Lehmann, E. L. (1963). A class of selection procedures based on ranks. *Math. Annalen* 150, 268–75.

[14] Mahamunulu, D. M. (1967). Some fixed-sample ranking and selection problems. *Ann. Math. Statist.* 38, 1079–91.

[15] McDonald, G. C. (1969). On some distribution-free ranking and selection procedures. Department of Statistics, Purdue University. Mimeograph, Series No. 174.

[16] Milton, R. C. (1964). An extended table of critical values for the Mann-Whitney (Wilcoxon) two-sample statistic. *J. Am. Statist. Ass.* 59, 925–34.

[17] Odeh, R. E. (1967). The distribution of the maximum sum of ranks. *Technometrics* 9, 271–8.

[18] Patterson, D. W. (1966). Contributions to the theory of ranking and selection problems, Ph.D. Thesis, State University of Rutgers. New Jersey: New Brunswick.

[19] Puri, M. L. (1964). Asymptotic efficiency of a class of $c$-sample tests. *Ann. Math. Statist.* 35, 102–21.

[20] Puri, P. S. and Puri, M. L. (1968). Selection procedures based on ranks: scale parameter case. *Sankhyā*, Series A **30**, 291–302.

[21] Puri, P. S. and Puri, M. L. (1969). Multiple decision procedures based on ranks for certain problems in analysis of variances. *Ann. Math. Statist.* **40**, 619–32.

[22] Rizvi, M. H. and Sobel, M. (1967). Nonparametric procedures for selecting a subset containing the population with the largest $\alpha$-quantile. *Ann. Math. Statist.* 38, 1788–803.

[23] Rizvi, M. H. and Woodworth, G. G. (1968). On selection procedures based on ranks: counterexamples concerning the least favorable configurations. Department of Operations Research and Department of Statistics, Standford University, Technical Report No. 114.

[24] Wilks, S. S. (1962). *Mathematical Statistics*. New York: John Wiley.