

NONPARAMETRIC RANKING PROCEDURES FOR

COMPARISON WITH A CONTROL

By

M. Haseeb Rizvi, Milton Sobel and George G. Woodworth

Ohio State University and Stanford University
University of Minnesota, Stanford University

TECHNICAL REPORT NO. 133

FEBRUARY 22, 1968

PREPARED UNDER OFFICE OF NAVAL RESEARCH CONTRACT
Nonr-225(52)

NASA Grant No. NGR36-008-040,

NASA always -

*Natl. Sci.
Fd.*

NSF Grant No. 3183, and Office of Naval Research

Navy

Contract Nonr-225(53)

NSF # G-3183

grant?

Reproduction in Whole or in Part is Permitted
for any Purpose of the United States Government

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

NONPARAMETRIC RANKING PROCEDURES FOR COMPARISON
WITH A CONTROL

By

M. Haseeb Rizvi, Milton Sobel and George G. Woodworth

1. Introduction and Summary.

A decision maker is confronted with k populations, π_1, \dots, π_k , (say, k lots of items available for purchase) and a control population π_0 and must, on the basis of random samples of common size n from π_0, \dots, π_k , select those which are at least as good as π_0 . We suppose that items are judged on the basis of a continuously distributed attribute X and that a known fraction α ($0 < \alpha < 1$) of the items in the control population are deficient (their X -values are too small). A population is considered to be better than the control if it has a smaller proportion of deficient items; that is, letting F_j , $j = 0, \dots, k$, denote the distribution function (df) of X for population π_j and $x_\alpha(F_j)$ its α^{th} quantile, π_i is better than π_0 if $x_\alpha(F_i) \geq x_\alpha(F_0)$. We also consider the possibility that F_0 is known in which case π_0 is called a standard and is not sampled. In section 2 we propose a nonparametric procedure R based on order statistics which guarantees a minimal preassigned probability P^* that, when each F_j is stochastically ordered with respect to F_0 , all populations better than the control will be selected; such a selection will be called a correct selection (CS). The corresponding problem of selecting a subset containing the best population (without any control) was treated in [11].

Since the trivial procedure R_0 of including all k populations in the selected subset also guarantees the probability requirement it is necessary to investigate the expected number of misclassifications; this is done exactly in section 3 and asymptotically in section 5. Exact results for known standard F_0 are given in section 4. Some other aspects of the problem are briefly discussed in section 8.

As a secondary problem we suppose that for some preassigned fraction δ^* the decision maker considers a population π_i to be δ^* -inferior to π_0 if more than $100(\alpha + \delta^*)$ percent of the items in π_i are as bad as at least one of the worst $100(\alpha - \delta^*)$ percent of the items in π_0 ; i.e., π_i is δ^* -inferior if $x_{\alpha - \delta^*}(F_0) \geq x_{\alpha + \delta^*}(F_i)$. In section 5 we give asymptotic expressions for the smallest sample size needed to guarantee that the expected proportion of δ^* -inferior populations selected by R will be less than a preassigned number β^* . An equally reasonable definition of π_i to be δ^* -inferior is that more than $100(\alpha + 2\delta^*)$ percent of the items in π_i are deficient. Our results with α replaced by $\alpha' = \alpha + \delta^*$ also apply to this problem.

We show in section 6 that for small values of δ^* a competing nonparametric procedure S based on rank sums and a competing asymptotically nonparametric procedure M based on sample means both require sample sizes proportional to the square of that required by R to achieve the same degree of rejection of δ^* -inferiors. For moderate δ^* -values it is shown that S requires a sample size which has the same order of magnitude as that required by R . In section 7 we study a related minimax procedure. We append tables for $\alpha = 1/2$ of (1) the

integer constant c needed to make procedure R explicit, (2) some required values to make the minimax procedure explicit and (3) efficiency comparisons of S with respect to R .

A Basic Inequality:

Let $\underline{X} = \{X_{ji}, 1 \leq j \leq n, 0 \leq i \leq k\}$ denote the combined sample, thus for each i X_{1i}, \dots, X_{ni} are independent random variables having the df $F_i(x)$. We regard $\omega = (F_0, F_1, \dots, F_k)$ as the unknown "parameter" and, for an arbitrary function ψ , use the symbol $E_{\omega} \psi(\underline{X})$ to denote the expected value of $\psi(\underline{X})$ computed under the assumption that ω is the true parameter value. The following lemma is used extensively in this paper; we state it without proof since it follows easily from Lemma 2.1 of [1].

Lemma 1.1. Let $\psi(\underline{x})$ be non-increasing in each x_{j0} , $j=1, \dots, n$, and non-decreasing in each x_{ji} , $1 \leq j \leq n$, $1 \leq i \leq k$, and let $\omega = (F_0, F_1, \dots, F_k)$ and $\omega' = (F'_0, F'_1, \dots, F'_k)$ satisfy $F_0(x) \leq F'_0(x)$ and $F_i(x) \geq F'_i(x)$ for $i = 1, \dots, k$ and all x , then

$$E_{\omega} \psi(\underline{X}) \leq E_{\omega'} \psi(\underline{X}) .$$

2. The Problem and the Proposed Procedure R (Unknown F_0).

Based on a common number n of observations from each of $k+1$ populations $(\pi_0, \pi_1, \dots, \pi_k)$, all $n(k+1)$ being independent, we want a procedure R that selects a subset of the k populations which (with high probability) will contain all populations better than π_0 , i.e., all π_i with $x_{\alpha}(F_i) \geq x_{\alpha}(F_0)$. To make this more precise,

we say F_i is as good as F_0 uniformly iff $F_i(x) \leq F_0(x)$ for all x and that F_i is worse than F_0 uniformly iff $x_\alpha(F_i) < x_\alpha(F_0)$ and $F_i(x) \geq F_0(x)$ for all x . Let Ω denote the space of all possible $(k+1)$ -tuples $\omega = (F_0, F_1, \dots, F_k)$ and let Ω_1 denote the subspace of Ω consisting of those ω such that for each $i (i = 1, 2, \dots, k)$ either F_i is as good as F_0 uniformly or F_i is worse than F_0 uniformly.

For any preassigned P^* with $2^{-k} < P^* < 1$ we want the procedure R to be such that

$$(2.1) \quad P\{CS|R\} \geq P^* \text{ whenever } \omega \in \Omega_1.$$

For any fixed α with $0 < \alpha < 1$ we assume that

$$(2.2) \quad 1 \leq (n+1)\alpha \leq n$$

and define the integer r by the inequalities

$$(2.3) \quad r \leq (n+1)\alpha < r+1.$$

It follows that $1 \leq r \leq n$.

We now define the procedure $R = R(c)$ in terms of an integer c and the order statistic Y_{ji} , where Y_{ji} is the j^{th} order statistic in a sample of size n from π_i ; since the F_i are unknown we take Y_{0i} to mean $-\infty$ for each i .

Procedure R :

The procedure $R(c)$ puts π_i in the selected subset for each $i (i = 1, 2, \dots, k)$ iff

$$(2.4) \quad Y_{ri} \geq Y_{r-c,0}.$$

The procedure R will be defined as that $R(c)$ for which c is the smallest integer ($0 \leq c \leq r$) such that $R(c)$ satisfies (2.1).

In order that the nonrandomized procedure be nondegenerate we limit the c -values to $0 \leq c \leq r-1$. We shall show that for any α and k a value of $c \leq r-1$ may not exist for all pairs (n, P^*) but if P^* is chosen not greater than some function $\bar{P}_0 = \bar{P}_0(n, \alpha, k)$, then a value of $c \leq r-1$ does exist that satisfies (2.1). \bar{P}_0 will be evaluated by setting $c = r-1$ in the $P(CS)$ and we show that \bar{P}_0 approaches unity as n increases. The values of P^* between \bar{P}_0 and 1 can be handled by the degenerate procedure $R_0(c = r)$ or by a randomized combination of the procedures for $c = r-1$ and $c = r$. The expressions for the $P(CS)$ etc. derived below all hold for $0 \leq c \leq r$ unless explicitly stated otherwise.

Letting $P\{CS|R\}$ denoted by $P_0(R)$ we now introduce other functions, some of which were suggested by Lehmann [7]. Some of these functions can be used as alternative criteria for developing new procedures. Let k_1 denote the number of π_i 's at least as good as π_0 , i.e., such that $x_\alpha(F_i) \geq x_\alpha(F_0)$; we denote the set of subscripts of these π_i by I_1 and refer to the corresponding set of populations as the superior set. Then $k_2 = k - k_1$ is the size of the set I_2 of subscripts of π_i 's in the interior set.

Let $P_1(R)$ denote the expected proportion of the k_1 superior populations that are correctly classified under procedure R . Let $P_2(R)$ denote the expected proportion of the k_2 inferior populations that are misclassified. If there are no superior (inferior) populations

then we define $P_1 = 0$ ($P_2 = 0$).

If we define a loss function $L = L(R; F_0, F_1, \dots, F_k)$ as the total number of misclassifications then we can write the expected loss or risk $E\{L|R\} = P_3(R)$ as

$$(2.5) \quad P_3(R) = k_1[1-P_1(R)] + k_2P_2(R) .$$

Obviously we would like R to be such that $P_0(R)$ and $P_1(R)$ are large while $P_2(R)$ and $P_3(R)$ are small. We shall therefore be interested in deriving the $\inf P_0(R)$, $\inf P_1(R)$, $\sup P_2(R)$, $\sup P_3(R)$, each taken over Ω_1 .

3. Exact Expressions for $P_i(R)$.

Let $dH_{ri}(y)$ and $H_{ri}(y)$ denote, respectively, the probability (density) element and the df of the r^{th} order statistic Y_{ri} in a sample of size n from the df $F_i(y)$. It is well known (and easy to show) that

$$(3.1) \quad dH_{ri}(y) = r \binom{n}{r} F_i^{r-1}(y) [1-F_i(y)]^{n-r} dF_i(y) ,$$

$$(3.2) \quad H_{ri}(y) = \sum_{j=r}^n \binom{n}{j} F_i^j(y) [1-F_i(y)]^{n-j} = G_r[F_i(y)] ,$$

where $G_r(p) = I_p(r, n-r+1)$ denotes the standard incomplete beta function

$$(3.3) \quad G_r(p) = r \binom{n}{r} \int_0^p x^{r-1} (1-x)^{n-r} dx .$$

Using the above notation, the probability of a correct selection under procedure R is given by

$$(3.4) \quad P_0(R) = P\{Y_{ri} \geq Y_{r-c,0}, i \in I_1\} = \int_{-\infty}^{\infty} \prod_{i \in I_1} [1 - H_{ri}(y)] dH_{r-c,0}(y) .$$

Similarly we obtain

$$(3.5) \quad P_1(R) = \frac{1}{k_1} \sum_{i \in I_1} \int_{-\infty}^{\infty} [1 - H_{ri}(y)] dH_{r-c,0}(y) ,$$

$$(3.6) \quad P_2(R) = \frac{1}{k_2} \sum_{i \in I_2} \int_{-\infty}^{\infty} [1 - H_{ri}(y)] dH_{r-c,0}(y) .$$

These in turn yield exact expression for $P_3(R)$.

We now obtain the infimum (or supremum) of these over Ω_1 . Consider $P_0(R)$. Since $G_r(p)$ is strictly increasing in p , it follows as in [11] that the infimum of $P_0(R)$ over Ω_1 is obtained by setting $F_i(y) = F_0(y)$ for $i \in I_1$ and minimizing over k_1 . Thus we obtain

$$(3.7) \quad \begin{aligned} \inf_{\Omega_1} P_0(R) &= \min_{0 \leq k_1 \leq k} \int_{-\infty}^{\infty} [1 - H_{r0}(y)]^{k_1} dH_{r-c,0}(y) \\ &= \int_0^1 [1 - G_r(u)]^k dG_{r-c}(u) = J_c(k) \quad (\text{say}). \end{aligned}$$

Since $G_r(u)$ is decreasing in r for any u (see e.g. [11]) it follows that

$$(3.8) \quad J_c(k) = k \int_0^1 G_{r-c}(u) [1 - G_r(u)]^{k-1} dG_r(u)$$

is an increasing function of c . Since $J_r(k) = P\{CS|R_0\} = 1$ it follows that our primary P^* -requirement in (2.1) has a solution for any n . Below we shall consider what values of P^* allow us to take $c \leq r-1$ and avoid the degenerate procedure R_0 of putting all k populations in the selected subset. Table 1 gives $(r-c)$ -values for

procedure R for some specified P^* when $\alpha = \frac{1}{2}$.

Similarly, we obtain the supremum of $P_2(R)$ (which is the same as the infimum of $P_1(R)$) by setting $F_1(y) = F_0(y)$ for $I \in I_2$ and maximizing (3.6) over k_2 , obtaining

$$(3.9) \quad \inf_{\Omega_1} P_1(R) = \sup_{\Omega_1} P_2(R) = \int_{-\infty}^{\infty} [1 - H_{r0}(y)] dH_{r-c,0}(y) \\ = \int_0^1 [1 - G_r(u)] dG_{r-c}(u) = J_c(1).$$

To find the supremum of $P_3(R)$ over Ω_1 we first show that $J_c(1) \geq 1/2$ for $0 \leq c \leq r$. Integration by parts in (3.9) gives

$$(3.10) \quad J_c(1) = \int_0^1 G_{r-c}(u) dG_r(u)$$

and we note that $J_0(1) = 1/2$. Since $G_r(x)$ is decreasing in r for any fixed x it follows that $J_c(1) \geq 1/2$ for $0 \leq c \leq r$. Hence, taking the supremum for fixed k_1 and then the maximum over k_1 ,

$$(3.11) \quad \sup_{\Omega_1} P_3(R) = \max_{0 \leq k_1 \leq k} \{k_1 \sup_{\Omega_1} [1 - P_1(R)] + k_2 \sup_{\Omega_1} P_2(R)\} \\ = \max_{0 \leq k_1 \leq k} \{k_1 [1 - J_c(1)] + (k - k_1) J_c(1)\} = k J_c(1).$$

In order to use the procedure R with $c \leq r-1$ and avoid the degenerate procedure R_0 for $c = r$, it is necessary to specify P^* not greater than \bar{P}_0 , where \bar{P}_0 is the value of $\inf_{\Omega_1} P_0(R)$ for $c = r-1$. From (3.7) we obtain

$$(3.12) \quad \bar{P}_0 = n \int_0^1 [G_{n-r+1}(v)]^k v^{n-1} dv = J_{r-1}(k).$$

An asymptotic expression for (3.7) is derived in section 5.

The value of $\bar{P}_1 = \inf_{\Omega_1} P_1(R)$ for $c = r-1$ (which also holds for \bar{P}_0 with $k = 1$) is

$$(3.13) \quad \bar{P}_1 = n \int_0^1 G_{n-r+1}(v) v^{n-1} dv = \binom{2n}{n}^{-1} \sum_{i=0}^{r-1} \binom{2n-i-1}{n-1} = 1 - \frac{\binom{2n-r}{n}}{\binom{2n}{n}}.$$

This is also the value of \bar{P}_2 , i.e., $\sup_{\Omega_1} P_2(R)$ for $c = r-1$. The smallest value that can be specified for P^* under Ω_1 using procedure R is easily seen to be $1/(k+1)$, obtained by setting $c = 0$ in (3.7).

It is also of some interest to investigate the infimum \underline{P}_0 of $P_0(R)$ under the set Ω of all possible configurations. The least favorable configuration here will occur for fixed F_0 when for each $i \in I_1$ F_i is as large as possible subject to $x_\alpha(F_i) \geq x_\alpha(F_0)$. We thus obtain k_1 binomial distributions with probability $1-\alpha$ at $x_\alpha(F_0)$ and the remaining mass at $-\infty$. Then

$$(3.14) \quad \inf_{\Omega} P_0(R) = \min_{1 \leq k_1 \leq k} G_{r-c}(\alpha) \left[\sum_{j=0}^{r-1} \binom{n}{j} \alpha^j (1-\alpha)^{n-j} \right]^{k_1} \\ = G_{r-c}(\alpha) [1 - G_r(\alpha)]^k.$$

To get an upper bound for (3.14) we first show that $G_r(\frac{r}{n+1})$ is decreasing in r . Writing

$$(3.15) \quad G_{r+1}(\frac{r+1}{n+1}) = G_{r+1}(\frac{r}{n+1}) + (n-r) \binom{n}{r} \int_{\frac{r}{n+1}}^{\frac{r+1}{n+1}} x^r (1-x)^{n-r-1} dx$$

and integrating $G_{r+1}(\frac{r}{n+1})$ by parts gives

$$(3.16) \quad G_r\left(\frac{r}{n+1}\right) - G_{r+1}\left(\frac{r+1}{n+1}\right) = \binom{n}{r} \left[\left(\frac{r}{n+1}\right)^r \left(\frac{n-r+1}{n+1}\right)^{n-r} - (n-r) \int_{\frac{r}{n+1}}^{\frac{r+1}{n+1}} x^r (1-x)^{n-r-1} dx \right].$$

Since the maximum of $x^r (1-x)^{n-r+1}$ is at $x = \frac{r}{n+1}$ we obtain from

(3.16) for any r

$$(3.17) \quad G_r\left(\frac{r}{n+1}\right) - G_{r+1}\left(\frac{r+1}{n+1}\right) \geq \binom{n}{r} \left(\frac{r}{n+1}\right)^r \left(\frac{n-r+1}{n+1}\right)^{n-r} \left[1 - (n-r) \left(\frac{n-r+1}{n+1}\right) \int_{\frac{r}{n+1}}^{\frac{r+1}{n+1}} \frac{dx}{(1-x)^2} \right] = 0.$$

Hence from (3.17) and the fact that $r/(n+1) \leq \alpha$, we obtain

$$(3.18) \quad G_r(\alpha) \geq G_r\left(\frac{r}{n+1}\right) \geq G_n\left(\frac{n}{n+1}\right) = \left(\frac{n}{n+1}\right)^n > \frac{1}{e}.$$

Thus from (3.14) we find that for any c -value

$$(3.19) \quad \underline{P}_0 \leq \left(1 - \frac{1}{e}\right)^k$$

which does not depend on α , r or n . Since this is less than $(.65)^k$ for any values of r , n , α we cannot use the least favorable configuration over Ω as a tool for formulating a ranking problem with the usual P^* -requirement.

4. Procedure R_1 for Known Standard.

In this case we do not sample the known standard and the form of the procedure changes. Let $x_{\alpha-\beta}(F_0)$ denote the $(\alpha-\beta)^{th}$ quantile of F_0 where β corresponds to $c/(n+1)$ in section 2.

Procedure R_1 :

For each $i (i = 1, 2, \dots, k)$ put F_i in the selected subset iff

$$(4.1) \quad Y_{ri} \geq x_{\alpha-\beta}$$

where β is the smallest number between 0 and α for which (2.1) holds.

Corresponding to the results in (3.4) through (3.9) we obtain for R_1

$$(4.2) \quad P_0(R_1) = P\{CS|R_1\} = P\{Y_{ri} \geq x_{\alpha-\beta}; i \in I_1\} = \prod_{i \in I_1} [1 - H_{ri}(x_{\alpha-\beta})]$$

$$(4.3) \quad P_1(R_1) = \frac{1}{k_1} \sum_{i \in I_1} [1 - H_{ri}(x_{\alpha-\beta})],$$

$$(4.4) \quad P_2(R_1) = \frac{1}{k_2} \sum_{i \in I_2} [1 - H_{ri}(x_{\alpha-\beta})],$$

$$(4.5) \quad \inf_{\Omega_1} P_0(R_1) = [1 - H_{r0}(x_{\alpha-\beta})]^k = [1 - G_r(\alpha - \beta)]^k = J'_\beta(k) \quad (\text{say}),$$

$$(4.6) \quad \inf_{\Omega_1} P_1(R_1) = 1 - H_{r0}(x_{\alpha-\beta}) = 1 - G_r(\alpha - \beta) = J'_\beta(1),$$

and the last result also holds for $\sup P_2(R_1)$ over Ω_1 .

If $r/(n+1) \geq 1/2$ then $\alpha \geq 1/2$ and $1-x \leq x$ for $x \geq \alpha$. It follows that for $r/(n+1) \geq 1/2$

$$(4.7) \quad 1 - G_r(\alpha) = r \binom{n}{r} \int_{\alpha}^1 x^{r-1} (1-x)^{n-r} dx \geq r \binom{n}{r} \int_{\alpha}^1 x^{n-r} (1-x)^{r-1} dx = G_r(\alpha),$$

so that $J'_0(1) = 1 - G_r(\alpha) \geq 1/2$. Since $J'_\beta(1)$ is strictly increasing in β for $0 \leq \beta \leq \alpha$, it follows that $J'_\beta(1) \geq 1/2$ for $r/(n+1) \geq 1/2$ and any β with $0 \leq \beta \leq \alpha$. Hence, corresponding to (3.11), we have for $r/(n+1) \geq 1/2$,

$$(4.8) \quad \sup_{\Omega_1} P_3(R_1) = \max_{0 \leq k_1 \leq k} \{k_1 [1 - J'_\beta(1)] + (k - k_1) J'_\beta(1)\} = k J'_\beta(1).$$

Since $J'_\beta(k)$ approaches 1 as $\beta \rightarrow \alpha$ we need not be concerned with the quantities \bar{P}_0, \bar{P}_1 , etc. when F_0 is known.

If we take the least favorable configuration over the set Ω of all possible configurations then we obtain, as in (3.14) through (3.19)

$$(4.9) \quad \inf_{\Omega} P_0(R_1) = [1 - G_r(\alpha)]^k \leq (1 - \frac{1}{e})^k \leq (.65)^k.$$

Hence the terminal remark of section 3 also holds for the case of known F_0 .

5. Asymptotic Properties of Procedure R .

Procedure R is constructed so that with high probability it retains those populations at least as good as the standard; it eliminates only those populations which, on the basis of a sample, appear to be definitely inferior. In this section we define a nonparametric measure, $\delta_\alpha(F, F_0)$, of the inferiority of a population with df F compared to the control population with df F_0 . It will be seen that

$$0 \leq \delta_\alpha(F, F_0) \leq \min(\alpha, \bar{\alpha}) \quad \text{provided } F(x) \geq F_0(x) \text{ for all } x \text{ and where } \bar{\alpha} = 1 - \alpha.$$

δ^* -Inferior Populations: For δ^* , a specified number between 0 and $\min(\alpha, 1 - \alpha)$, F is δ^* -inferior to F_0 if $F(x) \geq F_0(x)$ for all x and $\delta_\alpha(F, F_0) \geq \delta^*$. Let $P_2(\delta^* | R)$ denote the expected proportion of δ^* -inferiors in the subset selected by R; if there are no δ^* -inferiors then we define $P_2(\delta^* | R(c)) = 0$.

Recall that $R(c)$ is the selection procedure defined by (2.4).

In this section we obtain asymptotic expressions ($n \rightarrow \infty$) for

$$\inf_{\Omega_1} P_0(R(c)) \quad \text{and} \quad \sup_{\Omega_1} P_2(\delta^* | R(c)).$$

We use these to obtain asymptotic

expressions for the minimum sample size required by procedure R to guarantee for specified P^* and β^* , $\inf_{\Omega_1} P_0(R) \geq P^*$, and $\sup_{\Omega_1} P_2(\delta^*|R) \leq \beta^*$.

A measure of inferiority:

Let $F(x) \geq F_0(x)$ for all x and let $\delta(F, F_0)$ denote an arbitrary nonparametric measure of the degree of inferiority of F to F_0 . δ is nonparametric if and only if for continuous F and F_0

$$(5.1) \quad \delta(F, F_0) = \delta(F(F_0^{-1}), U), \text{ where } U \text{ is the uniform } (0,1) \text{ df.}$$

Being a measure of inferiority (degree of stochastic smallness) δ should also satisfy

$$(5.2) \quad F_0 = F \Rightarrow \delta(F, F_0) = 0$$

and

$$(5.3) \quad F'(x) \geq F(x) \geq F_0(x), \text{ for all } x \Rightarrow \delta(F', F_0) \geq \delta(F, F_0).$$

Let g be an arbitrary non-decreasing function of bounded variation on $(0,1)$; a general δ satisfying (5.1)-(5.3) is

$$\delta(F, F_0) = \int \left(\frac{F-F_0}{2} \right) dg \left(\frac{F+F_0}{2} \right).$$

One example of such a δ is already familiar, namely

$$\delta(F, F_0) = \int (F-F_0) d \left(\frac{F+F_0}{2} \right) = \int F dF_0 - \frac{1}{2}.$$

The measure $\delta_\alpha(F, F_0)$ which we propose is obtained by setting $g(u) = 0$ or 1 according as $u < \alpha$ or $u \geq \alpha$. It is easy to see that under the assumptions $F(x) \geq F_0(x)$, for all x , F and F_0 continuous,

this choice of g gives

$$(5.4) \quad \delta_{\alpha}(F, F_0) = \inf_x \left\{ \frac{F(x) - F_0(x)}{2} : \frac{F(x) + F_0(x)}{2} = \alpha \right\}.$$

Notice that if $F(x) + F_0(x) = 2\alpha$ then $F(x) = \alpha + \delta_{\alpha}(F, F_0)$ and $F_0(x) = \alpha - \delta_{\alpha}(F, F_0)$ so that $\delta_{\alpha}(F, F_0) \leq \min(\alpha, \bar{\alpha})$. We can also express (5.4) as

$$(5.5) \quad \delta_{\alpha}(F, F_0) = \inf_d \{d: F_0^{-1}(\alpha-d) \geq F^{-1}(\alpha+d)\},$$

provided we define $F_0^{-1}(u) = \inf_x \{x: F_0(x) > u\}$ and $F^{-1}(u) = \sup_x \{x: F(x) < u\}$. Thus $\delta_{\alpha}(F, F_0)$ is the smallest non-negative d such that $x_{\alpha-d}(F_0) \geq x_{\alpha+d}(F)$.

Asymptotic Expressions for $\inf_{\Omega_1} P_0(R(c))$:

It follows from (2.3) that $r/n \rightarrow \alpha$ as, $n \rightarrow \infty$. We shall consider two rates of growth as $n \rightarrow \infty$ for c in the procedure $R(c)$; Case i) $n^{-1/2}c \rightarrow (\alpha - \bar{\alpha})^{1/2}A$ where A is an arbitrary non-negative number and $\bar{\alpha} = 1 - \alpha$ and Case ii) for some ϵ ($0 < \epsilon < \alpha/2$), $\epsilon \leq c/n \leq \alpha - \epsilon$.

Case i) is involved in questions of Pitman efficiency and Case ii) in questions of Bahadur efficiency.

Case i): From (3.7) we conclude that

$$\inf_{\Omega_1} P_0(R(c)) = P\{Y_{ri} \geq Y_{r-c,0}, i = 1, \dots, k\},$$

where $F_1 = \dots = F_k = F_0$ are continuous. We can assume any convenient continuous form for this F_0 ; in particular, if F_0 is exponential then Y_{ri} and $Y_{r-c,0}$ are sums of independent random variables, from which it easily follows that (letting Φ denote the standard normal df)

$$(5.6) \quad \lim_{n \rightarrow \infty} \inf_{\Omega_1} P_0(R(c)) = \int_{-\infty}^{\infty} [\Phi(x+A)]^k d\Phi(x),$$

where $n^{-1/2}c \rightarrow A(\alpha \bar{\alpha})^{1/2}$. The integral in (5.6) occurs frequently in the literature of selection procedures and is extensively tabulated among others by Milton [10] and Gupta [5].

Case ii): In this case clearly $\inf_{\Omega_1} P_0(R(c)) \rightarrow 1$. Since $\inf_{\Omega_1} P_0(R(c)) = P\{Y_{ri} \geq Y_{r-c,0}, 1 \leq i \leq k\}$ when $F_1 = \dots = F_k = F_0$, it is clear that

$$P\{Y_{r1} < Y_{r-c,0}\} \leq 1 - \inf_{\Omega_1} P_0(R(c)) \leq kP\{Y_{r1} < Y_{r-c,0}\},$$

where $F_1 = F_0$.

The event $\{Y_{r1} < Y_{r-c,0}\}$ is the same as the event that at least r observations from population π_1 are among the $2r-c-1$ smallest observations from π_0 and π_1 together. Thus

$$(5.7) \quad P\{Y_{r1} < Y_{r-c,0}\} = \sum_{j=0}^{r-c-1} \frac{\binom{n}{j} \binom{n}{2r-c-j-1}}{\binom{2n}{2r-c-1}},$$

from which it is easy to obtain

$$(5.8) \quad P\{Y_{r1} < Y_{r-c,0}\} \leq (r-c) \frac{\binom{n}{r-c-1} \binom{n}{r}}{\binom{2n}{2r-c-1}},$$

and

$$(5.9) \quad P\{Y_{r1} < Y_{r-c,0}\} \geq \frac{\binom{n}{r-c-1} \binom{n}{r}}{\binom{2n}{2r-c-1}}.$$

Since $r/n \rightarrow \alpha$ and $0 < \epsilon \leq c/n \leq \alpha - \epsilon$, we can apply Stirling's

approximation to (5.8) and (5.9) to obtain:

$$P\{Y_{r-1} < Y_{r-c,0}\} \approx K_n \cdot \left(\frac{r - \frac{c}{2}}{r-c}\right)^{r-c} \left(\frac{r - \frac{c}{2}}{r}\right)^r \left(\frac{n-r+\frac{c}{2}}{n-r+c}\right)^{n-r+c} \left(\frac{n-r+\frac{c}{2}}{n-r}\right)^{n-r};$$

(5.8) implies that there exists an $\epsilon' > 0$ depending only on ϵ and α such that $K_n \leq n^{1/2}/\epsilon'$ and (5.9) implies that there exists an $\epsilon'' > 0$ depending only on ϵ and α such that $n^{-1/2}\epsilon'' \leq K_n$.

Thus if $c/n \rightarrow \gamma$, $0 < \gamma < \alpha$,

$$(5.10) \quad \lim_{n \rightarrow \infty} \left\{ -\frac{1}{n} \log[1 - \inf_{\Omega_1} P_0(R(c))] \right\} = I(\alpha-\gamma, \alpha-\frac{\gamma}{2}) + I(\alpha, \alpha-\frac{\gamma}{2}),$$

where $I(x, y) = x \log(\frac{x}{y}) + (1-x) \log(\frac{1-x}{1-y})$ is the Kullback-Leibler information number.

Asymptotic Expressions for $\sup_{\Omega_1} P_2(\delta^* | R(c))$:

Let $I_2(\delta^*)$ denote the set of subscripts of those F_i which are δ^* -inferior to F_0 and let $k_2(\delta^*)$ be the number of subscripts in $I_2(\delta^*)$. Then $P_2(\delta^* | R(c))$ is just (3.6) with k_2 and I_2 replaced by $k_2(\delta^*)$ and $I_2(\delta^*)$ respectively. The supremum of $1-H_{r-1}(y) = 1-G_r(F_i(y))$ over Ω_1 subject to $\delta(F_i, F_0) \geq \delta^*$ occurs when

$$(5.11) \quad F_i(x) = \begin{cases} F_0(x), & -\infty < x < x_{\alpha-\delta^*}(F_0) \\ \alpha+\delta^*, & x_{\alpha-\delta^*}(F_0) \leq x < x_{\alpha+\delta^*}(F_0) \\ F_0(x), & x_{\alpha+\delta^*}(F_0) \leq x < \infty \end{cases}$$

$$= F_1^*(F_0(x)), \text{ say.}$$

Thus

$$(5.12) \quad \sup_{\Omega_1} P_2(\delta^* | R(c)) = P\{Y_{r-1} \geq Y_{r-c,0}\},$$

where the latter probability is computed under the assumption that $F_0(x)$ is continuous and $F_1(x) = F_1^*(F_0(x))$.

Analogous to the two cases studied for $P_0(R(c))$ we consider as $n \rightarrow \infty$,

Case i) $n^{-1/2}c \rightarrow (\alpha \bar{\alpha})^{1/2}A$ and $n^{1/2}\delta^* \rightarrow (\alpha \bar{\alpha})^{1/2}f$, where A and f are arbitrary non-negative constants, and

Case ii) $c/n \rightarrow \gamma$, $0 \leq \gamma < \delta^*$, δ^* fixed, $0 < \delta^* < \min(\alpha, \bar{\alpha})$.

In case i) an argument similar to that used for $\inf_{\Omega_1} P_0(R(c))$ yields

$$(5.13) \quad \lim_{n \rightarrow \infty} \sup_{\Omega_1} P_2(\delta^* | R(c)) = \int_f^\infty [1 + \Phi(x+A) - \Phi(x-A)] d\Phi(x) + \Phi(A-f)[2\Phi(f)-1].$$

In case ii), by introducing U_{r-1} and $U_{r-c,0}$, where U_{r-1} and $U_{r-c,0}$ are the r^{th} and $(r-c)^{\text{th}}$ order statistics from two independent uniform $(0,1)$ samples each of size n , we can write (5.12) as

$$(5.14) \quad \begin{aligned} \sup_{\Omega_1} P_2(\delta^* | R(c)) &= P\{U_{r-1} \geq U_{r-c,0}, U_{r-1} < \alpha - \delta^*\} \\ &+ P\{\alpha - \delta^* \geq U_{r-c,0}, \alpha - \delta^* \leq U_{r-1} < \alpha + \delta^*\} \\ &+ P\{U_{r-1} \geq U_{r-c,0}, \alpha + \delta^* \leq U_{r-1}\}. \end{aligned}$$

Thus

$$(5.15) \quad \sup_{\Omega_1} P_2(\delta^* | R(c)) \leq P\{U_{r-c,0} \leq \alpha - \delta^*\} = P\{U_{r-1} \geq \alpha + \delta^*\}$$

and

$$(5.16) \quad \begin{aligned} \sup_{\Omega_1} P_2(\delta^* | R(c)) &\geq P\{U_{r-c,0} \leq \alpha - \delta^*\} \cdot P\{\alpha - \delta^* \leq U_{r-1} < \alpha + \delta^*\} \\ &\quad + P\{U_{r-1} \geq \alpha + \delta^*\} \cdot P\{U_{r-c,0} \leq \alpha + \delta^*\} \\ &\approx P\{U_{r-c,0} \leq \alpha - \delta^*\} + P\{U_{r-1} \geq \alpha + \delta^*\} \quad \text{as } n \rightarrow \infty. \end{aligned}$$

Letting $W(p)$ denote the sum of n Bernoulli random variables with parameter p , the right side of (5.15) which is the same as the second expression in (5.16) can be written as

$$(5.17) \quad P\{W(\alpha - \delta^*) \geq r - c\} + P\{W(\alpha + \delta^*) \leq r\}.$$

Then it follows from standard results on large deviations (eg. [4] Theorem 1) applied to (5.17) that

$$(5.18) \quad \begin{aligned} \lim_{n \rightarrow \infty} \left\{ -\frac{1}{n} \log \left[\sup_{\Omega_1} P_2(\delta^* | R(c)) \right] \right\} \\ = \min[I(\alpha - \gamma, \alpha - \delta^*), I(\alpha, \alpha + \delta^*)], \end{aligned}$$

where $I(x, y)$ is defined after (5.10).

Approximations to the Sample Size:

Let $n(P^*, \beta^*, \delta^* | R)$ be the smallest sample size required by procedure R to achieve $\inf_{\Omega_1} P_0(R) \geq P^*$ and $\sup_{\Omega_1} P_2(\delta^* | R) \leq \beta^*$. We now derive asymptotic expressions for $n(P^*, \beta^*, \delta^* | R)$ valid in three regions in the domain of the specified quantities (P^*, β^*, δ^*) ; the

first two regions correspond to cases i) and ii).

Region i): Let $0 < \beta^* < P^* < 1$ be fixed and δ^* small. Clearly, as $\delta^* \rightarrow 0$, $n(P^*, \beta^*, \delta^* | R) \rightarrow \infty$. It follows from (5.6) that $n^{-1/2} c \rightarrow A^*(\alpha \bar{\alpha})^{1/2}$, where A^* is the solution of the right side of (5.6) equated to P^* . Also, it follows from (5.13) that $n^{1/2} \delta^* \rightarrow f^*(\alpha \bar{\alpha})^{1/2}$ where f^* is the solution of the right side of (5.13) equated to β^* with A replaced by A^* .

Thus we have

$$(5.19) \quad n(P^*, \beta^*, \delta^* | R) \approx \alpha \bar{\alpha} (f^*)^2 / (\delta^*)^2$$

and

$$(5.20) \quad c \approx n^{1/2} A^*(\alpha \bar{\alpha})^{1/2}.$$

Region ii): Let $0 < \beta^* < 1$, $\delta^* > 0$ be fixed and P^* be close to 1.

It is easy to prove that $\frac{c}{n} \rightarrow \delta^*$; for if not then from (5.15) and

(5.16) one concludes that

$$\lim_{n \rightarrow \infty} \sup_{\Omega_1} P_2(\delta^* | R(c)) = \begin{cases} 0, & c/n \leq \delta^* - \epsilon \\ 1, & c/n \geq \delta^* + \epsilon \end{cases}$$

for any $\epsilon > 0$, but in fact $\sup_{\Omega_1} P_2(\delta^* | R(c)) = \beta^* \neq 0, 1$. Hence $\frac{c}{n} \rightarrow \delta^*$ and consequently γ of case ii) equals δ^* .

Therefore from (5.18) we have

$$(5.21) \quad \lim_{n \rightarrow \infty} \left\{ -\frac{1}{n} \log(1 - P^*) \right\} = I(\alpha - \delta^*, \alpha - \frac{\delta^*}{2}) + I(\alpha, \alpha - \frac{\delta^*}{2}).$$

Thus we have

$$(5.22) \quad n(P^*, \beta^*, \delta^* | R) \approx - \frac{\log(1-P^*)}{I(\alpha-\delta^*, \alpha-\frac{\delta^*}{2}) + I(\alpha, \alpha-\frac{\delta^*}{2})}$$

and, of course,

$$(5.23) \quad c \approx n\delta^* .$$

Region iii): Let $0 < P^* < 1$ and $0 < \delta^* < \min(\alpha, \bar{\alpha})$ be fixed and β^* small.

As in region i) we have $n^{-1/2}c \rightarrow A^*(\alpha, \bar{\alpha})^{1/2}$ so that $c/n \rightarrow 0$.

Since $\beta^* = \sup_{\Omega_1} P_2(\delta^* | R(c))$ we have from (5.18) (with $\gamma = 0$)

$$(5.24) \quad n(P^*, \beta^*, \delta^* | R) \approx -\log \beta^* / \min[I(\alpha, \alpha-\delta^*), I(\alpha, \alpha+\delta^*)] .$$

6. Efficiency Comparisons with Competing Procedures.

A nonparametric competitor S: Let R_{ji} ($1 \leq i \leq k$, $1 \leq j \leq n$) denote the rank of X_{ji} among $X_{10}, \dots, X_{n0}, X_{11}, \dots, X_{ni}$ (the smallest has rank 1) and let $R_{.i} = \sum_{j=1}^n R_{ji}$. Procedure $S(d)$ puts π_i in the selected subset iff

$$(6.1) \quad R_{.i} \geq d ,$$

where d is an integer not less than $n(n+1)/2$. Procedure S is determined by setting d equal to its largest value satisfying the condition $\inf_{\Omega_1} P_0(S(d)) \geq P^*$.

S is intimately related to a simultaneous inference procedure proposed by Steel (see [9] p. 143); in fact the d value needed to carry out S can be obtained from tables of the critical values of Steel's procedure with $1-P^*$ corresponding to the significance level.

To see this, notice that $R_{.i}$ is nondecreasing in observations from π_i and does not depend on observations from $\pi_{i'}, i' \neq 0, i$. Then by an obvious application of Lemma 1.1 we conclude that $P_0(S(d))$ is minimized over Ω_1 when $F_1 = F_2 = \dots = F_k = F_0$. Under this hypothesis the distribution of $(R_{.1}, \dots, R_{.k})$ is the same as that of $(n(2n+1)-R_{.1}, \dots, n(2n+1)-R_{.k})$. This is proved by taking $F_i, 0 \leq i \leq k$, to be uniform. Thus $Y_{ji} = (1-X_{ji})$, $0 \leq i \leq k, 1 \leq j \leq n$, are independent uniform random variables and if S_{ji} denotes the rank of Y_{ji} among $Y_{10}, \dots, Y_{n0}, Y_{1i}, \dots, Y_{ni}$, then clearly $S_{ji} = (2n+1)-R_{ji}$. The array $\{S_{ji}; 1 \leq i \leq k, 1 \leq j \leq n\}$ has the same distribution as $\{R_{ji}; 1 \leq i \leq k, 1 \leq j \leq n\}$ so $(R_{.1}, \dots, R_{.k})$ has the same distribution as $(\sum_j S_{j1}, \dots, \sum_j S_{jk}) = (n(2n+1)-R_{.1}, \dots, n(2n+1)-R_{.k})$ and consequently

$$(6.2) \quad \inf_{\Omega_1} P_0(S(d)) = P\left\{ \min_{1 \leq i \leq k} R_{.i} \geq d \right\} \\ = 1 - P\left\{ \max_{1 \leq i \leq k} R_{.i} > n(2n+1)-d \right\},$$

where these probabilities are computed under the assumption that $F_1 = F_2 = \dots = F_k = F_0$. If (6.2) is equated to P^* then $d = n(2n+1)+1-r^*$, where r^* can be obtained by entering table VIII, p. 250 of [9] at significance level $1-P^*$.

For fixed P^* from (57), p. 151 of [9] we obtain

$$(6.3) \quad d \approx n(2n+1)/2 - A^*n[(2n+1)/24]^{1/2},$$

where A^* is the solution of the right side of (5.6) equated to P^* .

Proportion of Inferiors Selected by S:

Let F^* denote an arbitrary (not necessarily continuous) df on the interval $0 < u < 1$ such that $F^*(u) \geq u$. We shall say that a df F_1 is F^* -inferior to F_0 if $F_1(x) \geq F^*(F_0(x))$, for all x ; $P_2(F^*|S)$ denotes the expected proportion of F^* -inferior populations in the subset selected by S and if no populations are F^* -inferior then we define $P_2(F^*|S) = 0$. Again applying Lemma 1.1 we conclude that

$$(6.4) \quad \sup_{\Omega_1} P_2(F^*|S) = P\{R_{.1} \geq d\},$$

where the latter probability is computed under the assumption that $F_1(x) = F^*(F_0(x))$.

From Lemma 3.2 of [2] we conclude that $W = n^{1/2}\{(R_{.1} - \frac{n(2n+1)}{2})/n^2 - (1/2 - \int F^* du)\}$ has the same limiting distribution ($n \rightarrow \infty$) as

$$Y = n^{-1/2} \sum_{j=1}^n [F_0(X_{j1}) + 1 - F^*(F_0(X_{j0}))] - 2n^{1/2}(1 - \int F^* du).$$

For purposes of analysis suppose that F^* depends on n and as $n \rightarrow \infty$ $F^*(u)$ approaches u at such a rate that $n^{1/2}(\int F^* du - 1/2) = O(1)$, then by application of the central limit theorem (as stated in [8], p. 295) we conclude that Y is asymptotically normal with mean zero and variance $1/6$. Hence from (6.3) and (6.4) we obtain

$$(6.5) \quad \sup_{\Omega_1} P_2(F^*|S) \approx \Phi\{2^{-1/2} A^* - (6n)^{1/2}(\int F^* du - 1/2)\},$$

when Φ is the standard normal df.

From (5.11) it is clear that F_1 is δ^* -inferior if and only if it is F^* -inferior with

$$(6.6) \quad F^*(x) = F_1^*(x) = \begin{cases} x & , \quad 0 < x < \alpha - \delta^* \quad \text{or} \quad \alpha + \delta^* \leq x < 1 \\ \alpha + \delta^* & , \quad \alpha - \delta^* \leq x < \alpha + \delta^* . \end{cases}$$

So if $P_2(\delta^*|S)$ denotes the expected proportion of δ^* -inferior populations in the subset selected by S we have from (6.5),

$$(6.7) \quad \sup_{\Omega_1} P_2(\delta^*|S) \approx \Phi(2^{-1/2} A^* - (24n)^{1/2} \delta^{*2}) ,$$

provided $n^{1/2} \delta^{*2} = o(1)$.

Defining $n(P^*, \beta^*, \delta^*|S)$ as in section 5 it follows from (6.7) that for fixed $0 < \beta^* < P^* < 1$ and small δ^*

$$(6.8) \quad n(P^*, \beta^*, \delta^*|S) \approx (z^* - 2^{-1/2} A^*)^2 / 24 (\delta^*)^4 ,$$

where $\Phi(z^*) = \beta^*$.

Asymptotic Relative Efficiencies for S Compared to R:

Comparison of (6.8) with (5.19) shows that for small δ^* the sample size required by S is proportional to the square of the sample size required by R . Thus the Pitman efficiency of S with respect to R is zero. It should be noted that if the extrema are taken over a smaller class than Ω_1 (such as a location parameter family) then the sample size comparison need not be so unfavorable to S , indeed, S may even require a smaller sample size than R .

Next we consider an efficiency comparison of the sort urged by Bahadur [3]. Here we hold δ^* and β^* fixed and study the behavior of the sample size as P^* approaches one.

In view of (6.4) and (6.5) and the asymptotic normality of $R_{.1}$ the assumption that $\sup_{\Omega_1} P_2(\delta^*|S) = \beta^*$ determines d and implies

$$(6.9) \quad d = ER_{.1} + O(\text{Var } R_{.1})^{1/2} = n^2 \left\{ \frac{3}{2} - \int F_1^* du \right\} + O(n^{3/2}) \approx n^2 [1 - 2(\delta^*)^2].$$

From (6.2) one obtains after some algebra using Bonferroni's inequality

$$(6.10) \quad -\frac{1}{n} \log[1 - \inf_{\Omega_1} P_0(S)] \approx -\frac{1}{n} \log P\{R_{.1} < d\},$$

where the latter probability is computed under the assumption that

$F_1 = F_0$. In [12] it is shown that

$$(6.11) \quad \lim_{n \rightarrow \infty} -\frac{1}{n} \log P\{R_{.1} < d\} = 2e_w(2(\delta^*)^2),$$

where $e_w(\rho)$ is given by the numerator of (3.4) of [6]. From (6.10) and (6.11) we obtain, for fixed β^* and δ^* and P^* approaching unity,

$$(6.12) \quad n(P^*, \beta^*, \delta^*|S) \approx -\log(1-P^*)/2e_w(2(\delta^*)^2)$$

and comparing (6.12) with (5.22) we have

$$(6.13) \quad \frac{n(P^*, \beta^*, \delta^*|R)}{n(P^*, \beta^*, \delta^*|S)} \approx \frac{2e_w(2(\delta^*)^2)}{I(\alpha - \delta^*, \alpha - \frac{\delta^*}{2}) + I(\alpha, \alpha - \frac{\delta^*}{2})}.$$

We shall call the right side of (6.13) the Bahadur efficiency of S with respect to R ; note that it is independent of β^* . Using line 13, p. 1762, of [6], $e_w(2(\delta^*)^2)$ can be evaluated by entering column 2 of Table I on that page for $\mu = 2^{-1/2} \Phi^{-1}[2(\delta^*)^2 + 1/2]$; we also do some additional calculations of $e_w(\cdot)$ and in Table 3 we tabulate this and the Bahadur efficiency of S with respect to R for $\alpha = \frac{1}{2}$.

An Asymptotically Nonparametric Competitor M:

Let \bar{X}_i be the sample mean from π_i , $0 \leq i \leq k$, and let $\Omega_1(B)$ be the subset of Ω_1 on which $v_4(F_0)/\sigma^4(F_0) \leq B < \infty$ where $\sigma^2(F_0)$ and $v_4(F_0)$ denote the variance and fourth central moment of F_0 . If $\sigma^2(F_0)$ is known it is possible to carry out the procedure $W(d)$ which retains those populations in the selected subset for which

$$(6.14) \quad n^{1/2} \bar{X}_i \geq n^{1/2} \bar{X}_0 - \sigma(F_0)d.$$

It follows from Lemma 1.1 that

$$(6.15) \quad \inf_{\Omega_1(B)} P_0(W(d)) = \inf_{F_0} \int [1 - F_0^{(n)}(x-d)]^k dF_0^{(n)}(x),$$

where $F_0^{(n)}$ is the df. of $n^{1/2}(\bar{X}_0 - \mu(F_0))/\sigma(F_0)$ and the second infimum is taken over those F_0 for which $v_4(F_0)/\sigma^4(F_0) \leq B$. If d remains bounded as $n \rightarrow \infty$ then using the Berry-Esseen bound

$$|F_0^{(n)}(x) - \Phi(x)| \leq \frac{C v_3(F_0) n^{-1/2}}{\sigma^3(F_0)} \leq C B^{3/4} n^{-1/2}, \text{ where } C \text{ is a constant}$$

and $v_3(F_0)$ is the third absolute central moment of F_0 , one can easily prove that the right side of (6.15) approaches the right side of (5.6) with $A = d$.

When $\sigma^2(F_0)$ is unknown, its estimate $S_0^2 = \sum_{j=0}^n (X_{j0} - \bar{X}_0)^2 / (n-1)$ has the property that $\sup_{F_0} P\{|1 - S_0/\sigma(F_0)| \geq \epsilon\} \leq B/n\epsilon^4, \epsilon > 0$. Define procedure $M(d)$ by replacing $\sigma(F_0)$ by S_0 in (6.14). Then it is easy to establish that

$$(6.16) \quad \lim_{n \rightarrow \infty} \{ \inf_{\Omega_1(B)} P_0(M(d)) \} = \int_{-\infty}^{\infty} \{\Phi(x+d)\}^k d\Phi(x).$$

With the further restriction that $\sigma(F_1) = \dots = \sigma(F_k) = \sigma(F_0)$ and $v_4(F_i)/\sigma^4(F_i) \leq B$, $i = 0, \dots, k$ one can use the pooled estimate $S^2 = \sum_{j=1}^n \sum_{i=0}^k (X_{ji} - \bar{X}_i)^2 / (k+1)(n-1)$ in place of S_0^2 and (6.16) will remain true.

We denote by M the procedure $M(d)$ with d determined so that $\lim_{n \rightarrow \infty} \inf_{\Omega_1(B)} P_0(M(d)) = P^*$; it follows from (6.16) that $d \rightarrow d^*$, the solution of the right side of (6.16) equated to P^* .

Proportion of Inferiors Selected by M for Fixed F_0 :

We define δ^* -inferior populations as usual, thus $F_1(x)$ is δ^* -inferior to F_0 iff $F_1(x) \geq F_1^*(F_0(x))$, where $F_1^*(F_0(x))$ is the right side of (5.11). Letting $P_2(\delta^*|M)$ denote the expected proportion of δ^* -inferiors selected by M and $\Omega_1(F_0)$ denote Ω_1 with F_0 held fixed, it follows from Lemma 1.1 that for any $\epsilon > 0$

$$(6.17) \quad \sup_{\Omega_1(F_0)} P_2(\delta^*|M) \leq P\{\bar{X}_1 \geq \bar{X}_0 - n^{-1/2} \sigma(F_0)(1+\epsilon)d\} \\ + P\{|1 - S_0 / \sigma(F_0)| > \epsilon\}$$

where the first probability on the right is computed with $F_1(x) = F_1^*(F_0(x))$ and the second probability depends only on F_0 . It follows easily from (6.17) that

$$(6.18) \quad \lim_{n \rightarrow \infty} \sup_{\Omega_1(F_0)} P_2(\delta^*|M) = \Phi(2^{-1/2} d^* - (\frac{n}{2})^{1/2} \frac{\mu(F_0) - \mu(F_1^*(F_0))}{\sigma(F_0)}),$$

provided $\delta^* \rightarrow 0$ such that

$$(6.19) \quad n^{1/2} [\mu(F_0) - \mu(F_1^*(F_0))] = n^{1/2} \int_{x_{\alpha-\delta^*}(F_0)}^{x_{\alpha+\delta^*}(F_0)} (x - x_{\alpha-\delta^*}(F_0)) dF_0 = o(1);$$

here $\mu(F)$ denotes the mean of the df F . Assuming F_0 has a positive derivative at its α^{th} -quantile and denoting it by $f_0(x_\alpha)$

(6.18) becomes

$$(6.20) \quad \lim_{n \rightarrow \infty} \sup_{\Omega_1(F_0)} P_2(\delta^* | M) = \Phi(2^{-1/2} z^* - \frac{(2n)^{1/2} (\delta^*)^2}{\sigma(F_0) f_0(x_\alpha)}) .$$

Thus for fixed F_0 , P^* and β^* we have (as $\delta^* \rightarrow 0$)

$$(6.21) \quad n(P^*, \beta^*, \delta^* | M) \approx (z^* - 2^{-1/2} z^*)^2 (\sigma(F_0) f_0(x_\alpha))^2 / 2 (\delta^*)^4 ,$$

where $\Phi(z^*) = \beta^*$.

Asymptotic Relative Efficiencies of M Compared to R:

Comparison of (6.21) and (5.19) shows that M , like S , requires sample sizes proportional to the square of that required by R for small δ^* .

In order to obtain Bahadur efficiency comparisons analogous to (6.13) for fixed F_0 one (essentially) needs a "large deviations" result for the t-statistic computed from a sample drawn from F_0 . To the authors' knowledge such a result is known only when F_0 is normal. Indeed if $F_1 = F_0$ is normal then $T = (2n)^{1/2} (\bar{X}_1 - \bar{X}_0) / S_0$ has the t-distribution with $(n-1)$ degrees of freedom. Klotz [6] shows that for a sequence r_n approaching a positive constant r_0 ,

$$\lim_{n \rightarrow \infty} -n^{-1} \log P\{T < -n^{1/2} r_n\} = \log(1+r_0^2)/2 .$$

Arguing as in the discussion leading up to (6.12) we conclude that with F_0 normal, δ^* and β^* fixed, as $P^* \rightarrow 1$,

$$(6.22) \quad n(P^*, \beta^*, \delta^* | M) \approx -2 \log(1 - P^*) / \log(1 + r_0^2(\alpha, \delta^*)) ,$$

where (with ϕ denoting the standard normal density function)

$$(6.23) \quad r_0(\alpha, \delta^*) = 2^{1/2} \{ \phi(\phi^{-1}(\alpha - \delta^*)) - \phi(\phi^{-1}(\alpha + \delta^*)) - 2\delta^* \phi^{-1}(\alpha - \delta^*) \} .$$

Thus, combining (6.22) and (5.22),

$$(6.24) \quad \frac{n(P^*, \beta^*, \delta^* | R)}{n(P^*, \beta^*, \delta^* | M)} \approx \frac{\log(1 + r_0^2(\alpha, \delta^*))}{2\{I(\alpha - \delta^*, \alpha - \frac{\delta^*}{2}) + I(\alpha, \alpha - \frac{\delta^*}{2})\}} .$$

We call the right side of (6.24) the Bahadur efficiency of M with respect to R when F_0 is normal; (6.24) is tabulated for $\alpha = \frac{1}{2}$ in Table 3. Since F_0 is normal it is not surprising that M becomes more efficient for larger values of δ^* .

7. A Minimax Procedure R' .

Another problem of interest to us is to define for a given $b \geq 0$ the risk function

$$(7.1) \quad P_4(R') = P_3(R') + b[1 - P_0(R')]]$$

and find the c -value such that for unknown F_0 the procedure $R' = R(c)$ minimizes the maximum of $P_4(R')$ over \mathcal{F} . This defines a new procedure R' that does not depend on any specified P^* ; we refer to it as the minimax procedure and with $J_c(\cdot)$ as defined by (3.7) obtain in a straightforward manner

$$(7.2) \quad \sup_{\Omega_1} P_4(R') = \max_{0 \leq k_1 \leq k} \{k_1[1-J_c(1)] + (k-k_1)J_c(1) + b[1-J_c(k_1)]\} .$$

If k is not large then we resort to a numerical computation for each value of k_1 in (7.2) to obtain the maximum, since an analytic maximization is difficult. Then the required c -value for the minimax procedure R' is the integer (with $0 \leq c \leq r$) that minimizes these maximal values. Table 2 gives c -values and the resulting minimax risks for $\alpha = \frac{1}{2}$, $b = k, 2k, 3k, k^2$ and selected values of n and k .

The trivial procedure R'_0 that selects one of the 2^k possible subsets at random without looking at any observations has the constant risk

$$(7.3) \quad P_4(R'_0) = \frac{k}{2} + b(1-2^{-k}) ,$$

which is an upper bound for the minimax risk for procedure R' .

For the case of known F_0 , the result analogous to (7.2) is obtained by recalling $J'_\beta(\cdot)$ of (4.5) and replacing $J_c(\cdot)$ by $J'_\beta(\cdot)$ in (7.2) and the minimax procedure R'_1 is then defined by taking β equal to the value that minimizes the maximum in the modified version of (7.2). The result (7.3) also holds for the trivial procedure in the case of known F_0 .

It should be noted that the minimax risk of R' in Table 2 is not necessarily monotonic in n for fixed k ; we believe that this is due to our forcing c to be an integer. If we use a suitable randomized procedure we can presumably take these "kinks" out of the minimax risk and make it monotonic.

8. Concluding Remarks.

Related Problems Solved by Procedure R:

The problem of selecting all populations with $x_{\alpha}(F_i) \leq x_{\alpha}(F_0)$ so that the probability of a correct selection is no less than P^* is solved by the procedure which selects π_i iff

$$(8.1) \quad Y_{n-r+1,i} \leq Y_{n-r+c+1,0} ,$$

where r is the integer satisfying (2.3) and c is the solution of (3.7) equated to P^* and may be obtained from a table giving c -values for procedure R corresponding to $1-\alpha$. This statement is proved simply by noting that if X has df $F(x)$ then $-X$ has df $1-F(-x)$ so that $-x_{1-\alpha}(F)$ is the α^{th} -quantile of $-X$.

The procedure defined by (8.1) also solves the classical problem of testing at level $1-P^*$ that at least one population is better than π_0 . Like Steel's procedure (section 6 and [9], p. 143), it has the property that, with probability at least P^* , one may correctly assert that all populations for which (8.1) is not true are better than π_0 .

We remark here that R has an unbiasedness property: if $F_i(x) \geq F_j(x)$ for all x then R is more likely to select π_j than π_i .

Scores Procedures:

Procedure S discussed in section 6 can be generalized by replacing the Wilcoxon statistic $R_{.i}$ in (6.1) by a two-sample scores statistic

$$(8.2) \quad T_i = \sum_{j=1}^n J_{n,R_{ji}}$$

with monotone scores $J_{n,1} \leq J_{n,2} \leq \dots \leq J_{n,2n}$; let us call this procedure S_J . It seems clear under the usual assumption (the step function $J_n(u) = J_{n,j}$, $(j-1)/2n \leq u < j/2n$, $1 \leq j \leq 2n$, converges in quadratic mean to a function $J(u)$) that S_J will still have zero Pitman efficiency compared to R . Under some additional assumptions on $J(u)$ (see [12]), there is a function $I_J(r_0)$ such that, when $F_1 = F_0$ and r_n is a sequence of constants approaching some constant r_0 ,

$$(8.3) \quad \lim_{n \rightarrow \infty} \left[-\frac{1}{n} \log P\{T_1 \geq nr_n\} \right] = I_J(r_0) .$$

In this case the Bahadur efficiency of S_J with respect to R will be the right side of (6.13) with the numerator replaced by $I_J(r^*)$, where r^* is the probability limit of $n^{-1}T_1$ when $F_1 = F_1^*(F_0)$ (see (5.11)), that is, $r^* = \int_0^1 J[(F_1^*(u) + u)/2] dF_1^*(u)$.

Acknowledgement.

The authors wish to thank Mr. Richard Freedman of Ohio State University for computing Tables 1 and 2.

Table 1: Largest values^s of r-c for which

$$\inf P_1(R) \geq P^* \text{ for } \alpha = \frac{1}{2} \text{ and } r = \frac{n+1}{2}.$$

$$P^* = .750$$

n	k=1	k=2	k=3	k=4	k=5	k=6	k=7	k=8	k=9
5	1	1	1	1	<u>1</u>	0#	0#	0#	0#
15	6	5	4	4	4	4	<u>4</u>	3	3
25	10	9	8	8	8	7	7	7	7
35	15	13	13	12	12	<u>12</u>	11	11	11
45	19	18	17	16	16	16	15	15	15
55	24	22	21	21	20	20	20	19	19
65	29	27	26	25	25	24	24	24	<u>24</u>

$$P^* = .900$$

5	1	0#	0#	0#	0#	0#	0#	0#	0#
15	3	3	3	3	3	<u>3</u>	2	2	2
25	8	7	7	6	6	6	6	<u>6</u>	5
35	12	11	10	10	10	<u>10</u>	9	9	9
45	16	15	14	14	14	13	13	13	13
55	21	19	19	18	18	17	17	17	17
65	25	24	23	22	22	21	21	21	21

$$P^* = .950$$

5	0#	0#	0#	0#	0#	0#	0#	0#	0#
15	3	3	2	2	2	2	2	2	2
25	7	6	6	5	5	5	5	5	<u>5</u>

$$P^* = .950 \text{ (cont.)}$$

n	k=1	k=2	k=3	k=4	k=5	k=6	k=7	k=8	k=9
35	11	10	9	9	<u>9</u>	8	8	8	8
45	15	14	13	13	12	12	12	12	11
55	19	18	17	16	16	16	16	15	15
65	23	22	21	20	20	20	<u>20</u>	19	19

$$P^* = .975$$

5	0#	0#	0#	0#	0#	0#	0#	0#	0#
15	2	2	2	<u>2</u>	1	1	1	1	1
25	6	5	5	4	4	4	4	4	4
35	10	9	8	8	<u>8</u>	7	7	7	7
45	13	12	12	11	11	11	11	<u>11</u>	10
55	17	16	16	15	15	<u>15</u>	14	14	14
65	22	20	19	19	19	18	18	18	18

$$P^* = .990$$

5	0#	0#	0#	0#	0#	0#	0#	0#	0#
15	2	1	1	1	1	1	1	1	1
25	5	4	4	<u>4</u>	3	3	3	3	3
35	8	7	7	7	6	6	6	6	6
45	12	11	10	10	10	<u>10</u>	9	9	9
55	16	15	14	14	13	13	13	13	<u>13</u>
65	20	18	18	17	17	17	16	16	16

§ Based on the equation $J_c(k) = P^*$; see (3.7). Other r-c values for $n > 65$ can be obtained from Table 3 of [11] by entering that table with the value of k increased by one. The underlined entries are the only values that differ from the corresponding entries (with k shifted by one) of Table 3 of [11]; in each case this value is exactly one larger than the value in [11].

Degenerate cases in which all the populations go into the selected subset with probability one.

Table 2: Minimax Risk and c-Values for $\alpha = \frac{1}{2}$

(In each cell the risk is followed by the c-value)

k	<u>b = k</u>				
	n = 5	n = 15	n = 25	n = 35	n = 45
1	.73810;1	.71524;1	.71454;2	.68381;2	.67324;2
2	1.52380;1	1.53786;2	1.57092;2	1.52802;3	1.52644;3
3	2.52380;1	2.46214;2	2.41041;3	2.47198;3	2.40396;4
4	3.57444;1	3.46214;2	3.39306;3	3.33592;4	3.39736;4
5	4.58335;2	4.34115;3	4.37100;4	4.33592;4	4.28882;5
6	5.50002;2	5.26354;3	5.25160;4	5.31906;5	5.28882;5
7	6.41669;2	6.26354;3	6.25160;4	6.22698;5	6.28882;5
8	7.33336;2	7.26354;3	7.25160;4	7.22698;5	7.20160;6
9	8.25003;2	8.27246;3	8.25160;4	8.22698;5	8.20160;6
<u>b = 2k</u>					
1	.78570;1	.76893;2	.80347;3	.76401;3	.78966;3
2	1.83334;2	1.73646;3	1.74840;4	1.67184;4	1.71118;5
3	2.75001;2	2.65885;3	2.62900;4	2.65953;5	2.69760;6
4	3.66668;2	3.74372;4	3.70368;5	3.68094;5	3.60480;6
5	4.58335;2	4.67965;4	4.62960;5	4.63780;6	4.66315;7
6	5.66664;2	5.61558;4	5.59264;5	5.57952;6	5.59578;7
7	6.74997;2	6.57663;4	6.66672;5	6.65196;6	6.60633;7
8	7.88844;2	7.64070;4	7.68320;6	7.65256;7	7.65624;8
9	9.08283;2	8.70477;4	8.64360;6	8.60913;7	8.61327;8

<u>$b = 3k$</u>					
k	n = 5	n = 15	n = 25	n = 35	n = 45
1	.91667;2	.86823;3	.80347;3	.83204;4	.80132;4
2	1.83334;2	1.79062;3	1.75480;4	1.77302;5	1.79840;6
3	2.75001;2	2.80779;4	2.77776;5	2.78268;6	2.79789;7
4	3.83330;2	3.74372;4	3.74080;5	3.72440;6	3.73052;7
5	4.99996;2	4.76884;4	4.80200;6	4.78285;7	4.78515;8
6	6.33282;2	5.84946;5	5.76240;6	5.73942;7	5.74218;8
7	7.70276;2	6.82437;5	6.72280;6	6.69599;7	6.69921;8
8	9.16030;2	7.79928;5	7.71280;6	7.78174;7	7.77346;8
9	10.67226;2	8.77419;5	8.79200;6	8.78148;8	8.76546;9

<u>$b = k^2$</u>					
1	.73810;1	.71524;1	.71454;2	.68381;2	.67324;2
2	1.83334;2	1.73646;3	1.74840;4	1.67184;4	1.71118;5
3	2.75001;2	2.80779;4	2.77776;5	2.78268;6	2.79789;7
4	4.22176;2	3.89698;4	3.84160;6	3.82628;7	3.82812;8
5	6.59058;2	4.87455;5	4.87120;6	4.87860;8	4.86970;9
6	9.98484;2	5.84946;5	5.88672;7	5.85432;8	5.84364;9
7	14.46795;2	6.95030;6	6.86784;7	6.91215;9	6.89528;10
8	19.92744;2	7.94320;6	7.93736;8	7.89960;9	7.88032;10
9	26.39700;2	8.93610;6	8.92953;8	8.92870;9	8.92719;11

Table 3: Bahadur Efficiency of S with respect to R and of M with respect to R (with F_0 normal df) when $\alpha = 1/2$.

δ^*	$I(\frac{1}{2} - \delta^*, \frac{1}{2} - \frac{\delta^*}{2})$ + $I(\frac{1}{2}, \frac{1}{2} - \frac{\delta^*}{2})$	$2e_w(2(\delta^*)^2)$	$\frac{1}{2} \log(1+r_0^2(\frac{1}{2}, \delta^*))$	Bahadur Efficiency [#] of	
				S with respect to R	M with respect to R
.050	.02707	.047665	.031579	.03058	.05833
.100	.01012	.034442	.022561	.1186	.2531
.150	.02312	.026085	.01319	.2632	.5705
.200	.04201	.01931	.04217	.4596	1.0040
.250	.06764	.04756	.1025	.7031	1.5150
.300	.1013	.1002	.2061	.9892	2.0350
.350	.1453	.1911	.3596	1.3150	2.4750
.400	.2035	.3428	.5661	1.6850	2.7820
.450	.2847	.6048	.8416	2.1220	2.9560
.500	$\frac{3}{2} \log(\frac{4}{3}) \doteq .4315$	$2\log(2) \doteq 1.3860$	∞	3.2130	∞

[#] See (6.13) and (6.24); here F_0 is assumed to be the normal df.

REFERENCES

- [1] Alam, K., and Rizvi, M.H. (1966). "Selection from Multivariate Normal Populations." Ann. Inst. Statist. Math. 18 307-318.
- [2] Andrews, F.C. (1954). "Asymptotic Behavior of Some Rank Tests for Analysis of Variance." Ann. Math. Statist. 29 724-735.
- [3] Bahadur, R.R. (1967). "Rates of Convergence of Estimates and Test Statistics." Ann. Math. Statist. 38 303-324.
- [4] Chernoff, H. (1952). "A Measure of Asymptotic Efficiency for Tests of a Hypothesis Based on the Sum of Observations." Ann. Math. Statist. 23 493-507.
- [5] Gupta, S.S. (1963). "Probability Integrals of Multivariate Normal and Multivariate t ." Ann. Math. Statist. 34 792-828.
- [6] Klotz, J. (1965). "Alternative Efficiencies for Signed Rank Tests." Ann. Math. Statist. 36 1759-1766.
- [7] Lehmann, E.L. (1961). "Some Model I Problems of Selection." Ann. Math. Statist. 32 990-1012.
- [8] Loève, M. (1963). Probability Theory. D. Van Nostrand Co., Inc., Princeton.
- [9] Miller, Jr., R.G. (1966). Simultaneous Statistical Inference, McGraw-Hill Book Co., New York.
- [10] Milton, R.C. (1963). "Tables of the Equally Correlated Multivariate Normal Probability Integral." Technical Report No. 27, Department of Statistics, University of Minnesota.
- [11] Rizvi, M.H., and Sobel, M. (1967). "Nonparametric Procedures for Selecting a Subset Containing the Population with the Largest α -Quantile." Ann. Math. Statist. 38 1788-1803.

REFERENCES (cont.)

- [12] Woodworth, G.G. (1967). "Large Deviations of Linear Rank Statistics." Technical Report No. 98, Department of Statistics, Stanford University.

UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R&D		
<i>(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)</i>		
1. ORIGINATING ACTIVITY (Corporate author) Department of Statistics Stanford University Stanford, California		2a. REPORT SECURITY CLASSIFICATION
		2b. GROUP
3. REPORT TITLE Nonparametric Ranking Procedures for Comparison with a Control		
4. DESCRIPTIVE NOTES (Type of report and inclusive dates) Technical Report		
5. AUTHOR(S) (Last name, first name, initial) M. Haseeb Rizvi, Milton Sobel and George G. Woodworth		
6. REPORT DATE February 22, 1968	7a. TOTAL NO. OF PAGES 38	7b. NO. OF REFS 12
8a. CONTRACT OR GRANT NO. Nonr 225(52)	9a. ORIGINATOR'S REPORT NUMBER(S) Technical Report No. 133	
b. PROJECT NO. NR-342-022		
c.	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) NASA Grant No. NGR36-008-040	
d.	NSF Grant No. 3183 and ONR 225(53)	
10. AVAILABILITY/LIMITATION NOTICES Distribution of this document is unlimited		
11. SUPPLEMENTARY NOTES		12. SPONSORING MILITARY ACTIVITY Logistics and Mathematical Science Branch Office of Naval Research, Washington, D.C.
13. ABSTRACT A decision maker is confronted with k populations π_1, \dots, π_k (say, k lots of items available for purchase) and a control population π_0 and must, on the basis of random samples of common size from π_0, \dots, π_k , select those which are at least as good as π_0 . It is assumed that items are judged on the basis of a continuously distributed attribute X and that a known fraction α ($0 < \alpha < 1$) of the items in the control population are deficient (their X -values are too small). Then π_1 is considered to be better than π_0 if it has a smaller proportion of deficient items, that is, if $x_\alpha(\pi_1) \geq x_\alpha(\pi_0)$ where $x_\alpha(\pi)$ denotes the α th quantile of π . A nonparametric procedure R based on order statistics is proposed as a solution for this problem; R guarantees with a preassigned probability P^* that, when each π_i is stochastically ordered with respect to π_0 , all populations better than π_0 will be selected. Further, for a preassigned fraction δ^* , π_1 is considered to be δ^* -inferior if $x_{\alpha-\delta^*}(\pi_0) \geq x_{\alpha+\delta^*}(\pi_1)$. Asymptotic expressions for the smallest sample size required to guarantee that the expected proportion of δ^* -inferiors selected by R will be less than a specified number β^* are obtained. Moreover, it is shown that for small values of δ^* a competing nonparametric procedure S based on rank sums and a competing asymptotically nonparametric procedure M based on sample means both require sample sizes proportional to the square of that required by R to achieve the same degree of rejection of δ^* -inferiors; for moderate values of δ^* , S requires a sample size which has the same order of magnitude as that required by R .		

DD FORM 1 JAN 64 1473

UNCLASSIFIED
Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Nonparametric Control population Quantile Stochastically ordered Order statistic 6*-inferior Rank sum Pitmann efficiency Bahadur efficiency						

INSTRUCTIONS

1. **ORIGINATING ACTIVITY:** Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization (*corporate author*) issuing the report.

2a. **REPORT SECURITY CLASSIFICATION:** Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2b. **GROUP:** Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. **REPORT TITLE:** Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. **DESCRIPTIVE NOTES:** If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. **AUTHOR(S):** Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. **REPORT DATE:** Enter the date of the report as day, month, year, or month, year. If more than one date appears on the report, use date of publication.

7a. **TOTAL NUMBER OF PAGES:** The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7b. **NUMBER OF REFERENCES:** Enter the total number of references cited in the report.

8a. **CONTRACT OR GRANT NUMBER:** If appropriate, enter the applicable number of the contract or grant under which the report was written.

8b, 8c, & 8d. **PROJECT NUMBER:** Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9a. **ORIGINATOR'S REPORT NUMBER(S):** Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9b. **OTHER REPORT NUMBER(S):** If the report has been assigned any other report numbers (*either by the originator or by the sponsor*), also enter this number(s).

10. **AVAILABILITY/LIMITATION NOTICES:** Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

- (1) "Qualified requesters may obtain copies of this report from DDC."
- (2) "Foreign announcement and dissemination of this report by DDC is not authorized."
- (3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through _____."
- (4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through _____."
- (5) "All distribution of this report is controlled. Qualified DDC users shall request through _____."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. **SUPPLEMENTARY NOTES:** Use for additional explanatory notes.

12. **SPONSORING MILITARY ACTIVITY:** Enter the name of the departmental project office or laboratory sponsoring (*paying for*) the research and development. Include address.

13. **ABSTRACT:** Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as (TS), (S), (C), or (U).

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. **KEY WORDS:** Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, roles, and weights is optional.