The Annals of Mathematical Statistics
1971, Vol. 42, No. 2, 630–642

A METHOD FOR THE CONSTRUCTION OF SEQUENTIAL SELECTION PROCEDURES[1]

BY DAVID G. HOEL[2]
Oak Ridge National Laboratory

1. Introduction. Paulson [13] has given a sequential procedure for selecting the normal population with the largest mean when the k available populations have a common known variance. His formulation employed the indifference zone approach, which, briefly described, is the situation where the probability of making a correct decision is required to be at least a given amount whenever the greatest mean exceeds all the others by a certain specified amount. This approach has been studied extensively by Bechhofer, Kiefer and Sobel [2] in the context of ranking and identification problems.

Paulson's procedure combines two properties which other indifference zone procedures do not typically possess, namely: the procedure is truncated and fully sequential. Truncation is self-explanatory, but by fully sequential we mean that the procedure is able to discontinue taking observations from a particular population before the procedure terminates sampling altogether. A fully sequential procedure will often make a substantial savings in the total number of observations taken by quickly eliminating "bad" populations if they are present among the populations under consideration. However, in studying the performance of ranking procedures the parameters are usually assumed to be in the "least favorable configuration" (all populations the same except one, which is at the minimum required distance for which detection is expected) or in the "equal parameter configuration" (all populations the same). In both instances "bad" populations are not present. Therefore, this should be kept in mind when comparing a fully sequential procedure with one which is not fully sequential.

Paulson's procedure for normal means has been extended [10] to the Koopman–Darmois family with the same measure of distance as that considered by Bechhofer, Kiefer and Sobel [2]. The procedures of this extension are not truncated. However, it has been shown by Bechhofer, Kiefer and Sobel [2] that for many members of the Koopman–Darmois family and for this particular distance measure, no truncated procedure exists.

We present in this paper a technique for the construction of Paulson-type procedures which are both truncated and fully sequential. Briefly, these procedures are formed by choosing a statistic of the observations from any two of the populations and performing a modified sequential probability ratio test (SPRT) based on this statistic. This is done simultaneously for all pairs of populations, and if a

Received January 30, 1970.

[1] Research sponsored by the U.S. Atomic Energy Commission under contract with the Union Carbide Corporation.

[2] Now at the National Institute of Environmental Health Sciences.

particular SPRT terminates then an appropriate population is removed from a set of contending populations. This is continued until only one population belongs to the set at which time sampling terminates and this population is selected as "best." It is shown how Paulson's procedure may be obtained in this manner. In Section 3 a procedure for selecting the normal population with the smallest variance is obtained by this method. The procedure is both location and scale invariant and uses the ratio of variances as the distance measure between populations. Finally in Section 4 a procedure based on rank-order statistics is given for the selection of the stochastically largest population. Also, Monte Carlo results are presented for both procedures.

2. The sequential procedure. Let $\Pi_1, \Pi_2, \cdots, \Pi_k$ represent k populations ($k \geqq 2$) and X_{ij} denote the jth observation from population Π_i. It is assumed that the observations $\{X_{ij}\}$ are independent random variables for all i and j. The distributions associated with the populations $\Pi_1, \Pi_2, \cdots, \Pi_k$ are not completely known and we wish to construct a selection procedure which attempts to pick the "best" (in some sense) population from the above set of k populations. For example, if the observations drawn from Π_i are normally distributed with unknown mean $\theta_i (i = 1, \cdots, k)$ and known variance σ^2 then we may be interested in selecting the population with the largest mean.

The selection procedure will depend upon the observations through a sequence of statistics $T_{ij}(n) (n = 1, 2, \cdots)$ which are defined to be functions

$$T_{ij}(n) = f_n(X_{i1}, \cdots, X_{in}; X_{j1}, \cdots, X_{jn}) \tag{1}$$

of the first n observations from populations Π_i and $\Pi_j (i, j = 1, \cdots, k)$. In a given problem the function f_n is chosen so as to indicate the differences between the populations in a reasonable way. In the above normal mean problem, for example, a choice of f_n might be $\sum_{k=1}^{n} (X_{ik} - X_{jk})$. Now we assume that $\mathbf{T}_{ij}(n) = (T_{ij}(1), \cdots, T_{ij}(n))$ has a joint probability density function $g_{n,\tau_{ij}}(\mathbf{T}_{ij}(n))$ depending on the parameter τ_{ij}. (Usually $T_{ij}(1), T_{ij}(2), \cdots$ has been chosen so that it is both a sufficient and transitive sequence and also invariantly sufficient for τ_{ij} (see [8]).) We use the parameter τ_{ij} as our measure of distance between the populations Π_i and Π_j. Next define $\tau = \max_i \min_{j, j \neq i} \{\tau_{ij}\}$ and let Π^* denote the population Π_i for which $\min_{j, j \neq i} \{\tau_{ij}\} = \tau$. If more than one population has this property we pick one to be Π^* and refer to it as the "best" population. Again returning to the above normal mean problem, we find that $\tau_{ij} = \theta_i - \theta_j$ and $\tau = \theta_{[k]} - \theta_{[k-1]} (\theta_{[1]} \leqq \cdots \leqq \theta_{[k]})$ and thus Π^* is the population with the largest mean $\theta_{[k]}$. If instead $\tau_{ij} = \theta_j - \theta_i$ then Π^* would be the population with the smallest mean.

As an error requirement for the selection procedure we use the indifference zone approach (see [2] for a detailed discussion). Basically, two constants τ^* and P^* with $\tau^* > \tau_{ii}$, $1 > P^* > 1/k$ are specified and we wish to select the best population Π^* with probability at least P^* whenever $\tau \geqq \tau^*$. When Π^* is chosen we

say the procedure has made the correct selection (CS). Thus we require a procedure for which

$$P(\text{CS}) \geqq P^* \quad \text{whenever} \quad \tau \geqq \tau^*. \tag{2}$$

Next we make the assumption that the functions $\{f_n\}$ have been chosen so that there exists a monotonically decreasing function h such that

$$\tau_{ji} = h(\tau_{ij}). \tag{3}$$

In the above example of normal means, we have $h(t) = -t$ which is decreasing and satisfies (3).

Using the function h to specify $\tau_0 = h(\tau^*)$ we define the log-likelihood ratios

$$l_n(\mathbf{T}_{ij}(n)) = \log g_{n,\tau_1}(\mathbf{T}_{ij}(n)) - \log g_{n,\tau_0}(\mathbf{T}_{ij}(n)) \tag{4}$$

upon which the procedure is based. In (4) we assume τ_1 is a fixed constant greater than τ_0.

Sequential selection procedure. Begin at stage one by taking a single observation from each of the k populations. Next calculate the values of the $k(k-1)$ log-likelihood ratios $l_1(\mathbf{T}_{ij}(1))$ $(i = 1, \cdots, k; j = 1, \cdots, k; i \neq j)$. If for some i

$$l_1(\mathbf{T}_{ij}(1)) \geqq a$$

where $a = \log\{(k-1)/(1-P^*)\}$ we eliminate population Π_j from further consideration. If only one population remains we terminate the sampling and declare this population to be the best. If more than one population remains we proceed to the next stage (second) and take a single observation from each of the *remaining* populations. The log-likelihood ratios for the contending populations are again computed and the same elimination rule is used except that $l_2(\mathbf{T}_{ij}(2))$ everywhere replaces $l_1(\mathbf{T}_{ij}(1))$. We continue in this manner until only one population is left at which time the procedure is terminated with the declaration that this remaining population is best. If at any stage (n say) in the procedure all the remaining populations are eliminated, we then select as best from among those populations sampled at stage n that Π_i for which $\min_{j \neq i} l_n(\mathbf{T}_{ij}(n))$ is a maximum.

Varying the choice of the constant τ_1 clearly changes the selection procedure. The only restriction is that $\tau_1 > \tau_0$ and thus we have a family of selection procedures for a given problem. If we pick $\tau_1 = \tau^*(= h(\tau_0))$, then by considering the log-likelihoods as functions of the original observations we see that $l_n(\mathbf{T}_{ij}(n)) = -l_n(\mathbf{T}_{ji}(n))$. This implies that the comparison made by the procedure between any two populations is a sequential probability ratio test with stopping boundaries $(-a, a)$. So in a sense the procedure consists of $k(k-1)/2$ simultaneous SPRT's when $\tau_1 = \tau^*$. In all applications so far observed, the procedure is truncated whenever τ_1 is chosen in the open interval (τ_0, τ^*). Also it has been observed that there is less over protection $(P(\text{CS}) - P^*)$ if the restriction $\tau_1 \geqq \tau_*$ is made where $\tau_* = h(\tau_{ii})$. Therefore we generally will choose $\tau_* \leqq \tau_1 \leqq \tau^*$ and recommend $\tau_1 = (\tau_* + \tau^*)/2$. Some further discussion on the choice of τ_1 is given in Section 5.

We now give in the following theorem sufficient conditions for the procedure to satisfy the error requirement (2).

THEOREM. *Assume that for each i and j*

$$P_{\tau_{ij}}\{l_n(\mathbf{T}_{ij}(n)) \geqq a \quad \text{for some} \quad n\} \tag{5}$$

is a non-decreasing function of τ_{ij}. *Then the sequential procedure satisfies the error requirement* (2) *provided the procedure terminates with probability one.*

PROOF. Without loss of generality we assume that Π_k is the best population Π^*. Since the procedure terminates with probability one, we have

$$\begin{aligned} 1-P(\text{CS}) &\leqq \textstyle\sum_{i=1}^{k-1} P[\Pi_i \quad \text{eliminates} \quad \Pi_k] \\ &\leqq \textstyle\sum_{i=1}^{k-1} P_{\tau_{ik}}[l_n(\mathbf{T}_{ik}(n)) \geqq a \quad \text{for some} \quad n]. \end{aligned}$$

Now since Π_k is the best population we have by assuming $\tau \geqq \tau^*$ that $\tau_{ki} \geqq \tau \geqq \tau^*(i = 1, \cdots, k-1)$. It then follows from (3) and the definition of h that $\tau_{ik} = h(\tau_{ki}) \leqq h(\tau^*) = \tau_0$. By the monotonicity condition (5) of the theorem we then have $1-P(\text{CS}) \leqq (k-1)P_{\tau_0}[l_n(\mathbf{T}_{ik}(n)) \geqq a$ for some $n]$. Finally, using a basic inequality of Wald's (page 146 of [16]) we have $P_{\tau_0}[l_n(\mathbf{T}_{ik}(n)) \geqq a$ for some $n] \leqq e^{-a}$ and thus $1-P(\text{CS}) \leqq (k-1)e^{-a} = 1-P^*$ which completes the proof.

One can interpret $P_{\tau_{ij}}\{l_n(\mathbf{T}_{ij}(n)) \geqq a$ for some $n\}$ as the probability of accepting H_1 in an SPRT of $H_0: \tau = \tau_0$ vs $H_1: \tau = \tau_1$ with stopping boundaries $(-\infty, a)$. Therefore, condition (5) of the theorem is satisfied if this SPRT has a monotone operating characteristic (OC) function. A sufficient condition for this monotonicity is the one given by Ghosh [7]. His condition is that the family of densities $g_{n,\tau}$ possesses a monotone likelihood ratio. For other sufficient conditions and further discussion of the OC of an SPRT the reader is referred to [8] and [9].

As a simple illustration of the procedure, we continue with the example of selecting the normal population with the greatest mean. Since we chose $T_{ij}(n) = \sum_{k=1}^{n}(X_{ik}-X_{jk})$ it follows that

$$l_n(\mathbf{T}_{ij}(n)) = [(\tau_1-\tau_0)/2\sigma^2]\textstyle\sum_{k=1}^{n}(X_{ik}-X_{jk})+n(\tau_0^2-\tau_1^2)/4\sigma^2. \tag{6}$$

Now if Π_i and Π_j both remain up to stage n then Π_j is eliminated if

$$\sum_{k=1}^{n} X_{jk} \leqq \sum_{k=1}^{n} X_{ik} - \frac{2\sigma^2 a}{\tau_1-\tau_0} - n(\tau_0+\tau_1)/2. \tag{7}$$

Thus the procedure is equivalent to the one which eliminates Π_j if

$$\sum_{k=1}^{n} X_{jk} \leqq \max_i \sum_{k=1}^{n} X_{ik} - \frac{2\sigma^2 a}{\tau_1-\tau_0} - n(\tau_0+\tau_1)/2 \tag{8}$$

where the max is taken over those Π_i which are in contention at stage n. If for some n

$$-n(\tau_0+\tau_1)/2 \geqq \frac{2\sigma^2 a}{\tau_1-\tau_0}$$

then sampling is terminated and the population with the largest value of $\sum_{k=1}^{n} X_{ik}$ is chosen as best. Thus if $\tau_1 < -\tau_0 = \tau^*$ the procedure is truncated and since $g_{n,\tau}$ possess a monotone likelihood ratio we have by the theorem that the procedure satisfies the error requirement (2) for $\tau_1 \leqq -\tau_0 (\tau_1 = -\tau_0$ is permitted, since although the procedure is not truncated, it terminates with probability one.)

This procedure is recognized as precisely Paulson's [13] procedure for selecting the normal population with the largest mean. He requires that $\tau_0 < \tau_1 < -\tau_0$ and recommends the value $\tau_1 = -\tau_0/2$ which is equal to $(\tau_* + \tau^*)/2$.

3. Normal variances. In this section we apply the selection procedure to the problem of choosing the normal population with the smallest variance. To begin, assume that X_{ij} is normally distributed with unknown mean μ_i and unknown variance ${\sigma_i}^2$. Let the ranked values of ${\sigma_i}^2$ be denoted by $\sigma^2_{[1]} \leqq \sigma^2_{[2]} \leqq \cdots \leqq \sigma^2_{[k]}$ and define the best population Π^* to be the one with the smallest variance $\sigma^2_{[1]}$.

Bechhofer and Sobel ([3], [4]) have given both a fixed sample and a sequential procedure for choosing Π^*. However, their sequential procedure is neither truncated nor fully sequential. The measure of distance between $\Pi_{[i]}$ and $\Pi_{[j]}$ they consider is $\sigma^2_{[i]}/\sigma^2_{[j]}$ for $i \geqq j$ and for this measure the error requirement (2) becomes

$$P(\mathrm{CS}) \geqq P^* \quad \text{if} \quad \sigma^2_{[2]}/\sigma^2_{[1]} \geqq \tau^*. \tag{9}$$

In applying the procedure to this problem we define

$$\bar{X}_{im} = \sum_{j=1}^{m} X_{ij}/m$$
$$s^2_{im} = \sum_{j=1}^{m} (X_{ij} - \bar{X}_{im})^2/(m-1)$$

and assign $T_{ij}(n) = s^2_{in}/s^2_{jn}$, $n = 2, 3, \cdots$. This choice of T_{ij} seems to be a natural one if we wish to have the ratio of variances as the distance measure. Now from the results of Hall, Wijsman and Ghosh [8], the sequence $T_{ij}(2), T_{ij}(3), \cdots$ can be seen to be invariantly sufficient for ${\sigma_i}^2/{\sigma_j}^2$. Thus $l_n(\mathbf{T}_{ij}(n))$ is a function only of $T_{ij}(n)$ and in particular

$$l_n(\mathbf{T}_{ij}(n)) = (n-1)\log\left\{\left(\frac{1/\tau_0 + T_{ij}(n)}{1/\tau_1 + T_{ij}(n)}\right)\left(\frac{\tau_0}{\tau_1}\right)^{\frac{1}{2}}\right\} \tag{10}$$

with $\tau_{ij} = {\sigma_j}^2/{\sigma_i}^2$ and $\tau = \sigma^2_{[2]}/\sigma^2_{[1]}$. If we instead defined $\tau_{ij} = {\sigma_i}^2/{\sigma_j}^2$ then $\tau = \sigma^2_{[k]}/\sigma^2_{[k-1]}$ and the procedure's objective would be to select the population with the largest variance.

To show when the procedure is truncated we consider only the comparisons made between Π_i and Π_j, namely, $l_n(\mathbf{T}_{ij}(n))$ and $l_n(\mathbf{T}_{ji}(n))$. Now since $T_{ji} = 1/T_{ij}$ we have from (10) that if

$$(n-1)\log\left\{\frac{1/\tau_0 + T_{ij}(n)}{1/\tau_1 + T_{ij}(n)}\left(\frac{\tau_0}{\tau_1}\right)^{\frac{1}{2}}\right\} \geqq a \tag{11}$$

population Π_j is eliminated and if

$$(n-1)\log\left\{\frac{1/\tau_0 + 1/T_{ij}(n)}{1/\tau_1 + 1/T_{ij}(n)}\left(\frac{\tau_0}{\tau_1}\right)^{\frac{1}{2}}\right\} \geqq a \tag{12}$$

population Π_i is eliminated. The left-hand side of (11) is a decreasing function of $T_{ij}(n)$ and thus either Π_i or Π_j must be eliminated by the time that

$$n-1 \geqq a \log^{-1}\left\{\frac{1/\tau_0+1}{1/\tau_1+1}\left(\frac{\tau_0}{\tau_1}\right)^{\frac{1}{2}}\right\}$$

provided that the expression within the braces is greater than one. Since this is true for any pair of populations, the procedure can continue for at most

$$1+a \log^{-1}\left\{\frac{1/\tau_0+1}{1/\tau_1+1}\left(\frac{\tau_0}{\tau_1}\right)^{\frac{1}{2}}\right\} \tag{13}$$

stages provided $\tau_1 < 1/\tau_0$. Finally, by setting $T_{ij}(n) = 1$ we see from (11) and (12) that the procedure is not truncated for $\tau_1 \geqq 1/\tau_0$.

Now, the procedure terminates w.p.1 for $\tau_1 = 1/\tau_0$ (see [8]) and the densities $\{g_{n\tau}\}$ possess a monotone likelihood ratio. Thus the procedure satisfies the error requirement (2) for $\tau_1 \leqq 1/\tau_0$. Finally, the procedure clearly depends only on the T_{ij}'s and is therefore both location and scale invariant.

To illustrate the procedure a series of Monte Carlo trials was performed at the points $k = 3$, $\tau_0 = .5$, $\tau_1 = 1, 1.5, 2$ and $P^* = .75, .90, .95, .99$. The chosen values of τ_1 correspond to τ_*, $(\tau_* + \tau^*)/2$, τ^* which were discussed in the previous section. At each point 500 trials were run with the parameters in the least favorable configuration (i.e. $\sigma^2_{[2]} = \sigma^2_{[3]} = \tau^*\sigma^2_{[1]}$.) The results are given in Table 1 with $E(N)$ and $E(S)$, respectively, denoting the average number of observations and stages observed. Column 5 gives the number of stages required for the fixed sample procedure with probability of correct selection equal to the P^* of Column 1. These values were obtained by interpolation in the tables of Bechhofer and Sobel [3]. Column 6 contains the procedure's truncation point as given in (13). Finally, in Table 2 the observed number of stages are given with the parameters in the equal parameter configuration (i.e. $\sigma^2_{[1]} = \sigma^2_{[2]} = \sigma^2_{[3]}$.)

On inspection of Table 1 it is first noticed that for $\tau_1 = 1$ there is considerable amount of over-protection ($P(\mathrm{CS}) - P^*$) especially for smaller values of P^*. Next $E(S)$ and $E(N)$ for the recommended value $\tau_1 = 1.5$ come fairly close to the corresponding values for $\tau_1 = 2$. However, when one goes to Table 2 we see that $\tau_1 = 1.5$ comes out best. This is particularly true when $\tau_1 = 1.5$ is compared with the non-truncated procedure $\tau_1 = 2$. In fact, for the case $P^* = .99$ and $\tau_1 = 2$ it was found that 18% of the trials took more than 139 stages, the upper bound for the procedure with $\tau_1 = 1.5$. Also one of the 500 trials took more than 400 stages to terminate. Therefore it appears that truncated procedures can often make considerable savings in undesirable cases without giving up too much when the parameters are in the least favorable configuration.

One further point should be made concerning Table 1. It is that $E(N)$ is not much smaller than $3E(S)$. A fully sequential procedure will make its best savings when several "bad" populations are present which are then quickly eliminated. We could for example set $\sigma^2_{[3]}$ quite high and would then expect to observe $E(N)$ much closer to $2E(S)$ than $3E(S)$.

Before leaving the problem of normal variances it should be mentioned that another measure of distance has been considered by Bechhofer, Kiefer and Sobel [2]. They give a sequential procedure for selecting Π^* when the distance between $\Pi_{[j]}$ and $\Pi_{[i]}$ is $1/\sigma^2_{[j]} - 1/\sigma^2_{[i]}$ for $i \geqq j$. Also they show that there does not exist a truncated procedure for the problem and thus we would not expect our method to be of use. However, a copy of Paulson's procedure for normal means is as follows.

TABLE 1

Empirical properties of the variance selection procedure with the parameters in the least favorable configuration

$\tau_1 = 1$				Number of stages for fixed sample case	Truncation point
P^*	$P(CS)$	$E(N)$	$E(S)$		
.75	.910	45.6 (.6)	16.5 (.2)	10.7	36.3
.90	.960	65.3 (.8)	23.7 (.3)	22.7	51.9
.95	.986	79.1 (.9)	28.6 (.4)	32.6	63.6
.99	.996	107.2 (1.1)	38.2 (.4)	56.5	91.0
$\tau_1 = 1.5$				Number of stages for fixed sample case	Truncation point
P^*	$P(CS)$	$E(N)$	$E(S)$		
.75	.860	35.9 (.6)	13.3 (.3)	10.7	55.0
.90	.930	52.3 (.9)	19.6 (.4)	22.7	78.9
.95	.962	63.0 (1.1)	23.3 (.4)	32.6	96.9
.99	.996	91.3 (1.4)	33.6 (.6)	56.5	138.7
$\tau_1 = 2$				Number of stages for fixed sample case	
P^*	$P(CS)$	$E(N)$	$E(S)$		
.75	.828	33.0 (.7)	12.5 (.3)	10.7	
.90	.932	51.7 (1.2)	19.6 (.5)	22.7	
.95	.972	62.6 (1.2)	23.8 (.5)	32.6	
.99	.996	87.1 (1.6)	32.8 (.7)	56.5	

($\tau^* = 2$, 500 simulations per point)

TABLE 2

Observed number of stages for the variance selection procedure with the parameters in the equal population configuration

	$E(S)$		
P^*	$\tau_1 = 1$	$\tau_1 = 1.5$	$\tau_1 = 2$
.75	20.1 (.3)	16.6 (.4)	17.2 (.5)
.90	31.6 (.4)	28.2 (.5)	32.4 (1.0)
.95	40.5 (.5)	36.7 (.7)	46.0 (1.5)
.99	61.4 (.6)	59.2 (1.0)	92.7 (3.1)

Note: The values in the parentheses are the observed standard errors.

At stage m population Π_i is eliminated if $s_{im}^2 > \min_j s_{jm}^2 + 2a/\Delta$ where the minimum is taken over the remaining populations. To see that this procedure satisfies

$$P(\text{CS}) \geqq P^* \quad \text{whenever} \quad 1/\sigma_{[1]}^2 - 1/\sigma_{[2]}^2 \geqq \Delta$$

simply write s_{im}^2/σ_i^2 as a sum of independent chi-square variables and apply Paulson's argument.

4. Selecting the stochastically largest population. In this section we consider the more interesting application of the selection procedure; that of choosing the stochastically largest population. We accomplish this by using Lehmann alternatives and applying the sequential rank test of Bradley, Merchant and Wilcox in [6].

First, we assume that an observation X_{ij} from population Π_i has a distribution G^{θ_i} where G is an unknown continuous distribution function and θ_i is a positive unknown parameter. With this set up, the stochastically largest population Π^* is that population associated with $G^{\theta_{[k]}}$ where $\theta_{[1]} \leqq \theta_{[2]} \leqq \cdots \leqq \theta_{[k]}$.

Following the method of Bradley ([5], [6]) we set $T_{ij}(n)$ equal to the vector of ranks $(s_{1n}(ij), \cdots, s_{nn}(ij))$ of the n observations $X_{i1}, \cdots, X_{in}$ among the combined sample $X_{i1}, \cdots, X_{in}, X_{j1}, \cdots, X_{jn}$. Now, since we may write $G^{\theta_i} = (G^{\theta_j})^{\theta_i/\theta_j}$ we have from [11] that the density of $T_{ij}(n)$ is

$$\frac{\tau_{ij}^n}{\binom{2n}{n}} \prod_{\nu=1}^{n} \frac{\Gamma(s_{\nu n}(ij) + \nu(\tau_{ij}-1))\Gamma(s_{\nu+1,n}(ij))}{\Gamma(s_{\nu+1,n}(ij) + \nu(\tau_{ij}-1))\Gamma(s_{\nu n}(ij))}$$

where $\tau_{ij} = \theta_i/\theta_j$ and $s_{n+1,n} = 2n+1$. Also $T_{ij}(n)$ is sufficient for $\mathbf{T}_{ij}(n)$ (see [8]) and thus

$$l_n(\mathbf{T}_{ij}(n)) = n \log(\tau_1/\tau_0) + \sum_{\nu=1}^{n} \left\{ \log \frac{\Gamma(s_{\nu n}(ij) + \nu(\tau_1 - 1))}{\Gamma(s_{\nu+1,n}(ij) + \nu(\tau - 1))} - \log \frac{\Gamma(s_{\nu n}(ij) + \nu(\tau_0 - 1))}{\Gamma(s_{\nu+1,n}(ij) + \nu(\tau_0 - 1))} \right\}. \tag{14}$$

From $\tau_{ij} = \theta_i/\theta_j$ we see that $\tau = \theta_{[k]}/\theta_{[k-1]}$ and $h(t) = 1/t$. This in turn yields the error requirement

$$P(\text{CS}) \geqq P^* \quad \text{whenever} \quad \theta_{[k]}/\theta_{[k-1]} \geqq \tau^*. \tag{15}$$

Here $\theta_{[k]}/\theta_{[k-1]}$ is the power which you must raise the distribution of the second stochastically largest population to obtain the distribution of the stochastically largest. Also the right-hand side of (15) implies that

$$P[X_{[k]} \geqq X_{[k-1]}] \geqq \tau^*/(1+\tau^*) \tag{16}$$

where $X_{[k-1]}$ and $X_{[k]}$ are random variables with distribution $G^{\theta_{[k-1]}}$ and $G^{\theta_{[k]}}$, respectively. This gives some insight into what different choices of τ^* imply. Also useful is Table 5 of [17] which gives the mean and variance of G^θ for various θ when G is the normal distribution function.

In applying the theorem to establish that the procedure satisfies (15) several difficulties are encountered. The main one concerns the monotonicity condition (5). It has not been established that the sequential rank test which draws observations in pairs has a monotone OC function which would be sufficient for (5). However, empirical evidence [5], [6] certainly indicates that the OC is in fact monotone. Also, it has been established [9] that the sequential, rank test which draws its observations singly (alternating between populations) [12] does have a monotone OC function.

Turning now to the problem of truncation, consider, as before, only the two populations Π_i and Π_j. Suppose, after n observations are drawn from each population and the combined sample ordered, that the observations appear in alternating order from Π_i and Π_j. Specifically, we set $s_{kn}(ij) = 2k$ where $s_{kn}(ij)$ is the rank of the kth observation from Π_i in the combined sample. Similarly, if $u_{kn}(ij)$ is the rank of the kth observation from Π_j, we have $u_{kn}(ij) = 2k-1$. From (14) we then find

$$l_n(\mathbf{t}_{ij}(n)) = \log\left\{\left(\frac{\tau_1}{\tau_0}\frac{\tau_0+1}{\tau_1+1}\right)\prod_{k=1}^{n-1}\left(\frac{\tau_0+1+1/k}{\tau_1+1+1/k}\right)\left(\frac{\tau_1}{\tau_0}\frac{\tau_0+1}{\tau_1+1}\right)\right\} \tag{17}$$

and

$$l_n(\mathbf{t}_{ji}(n)) = \log\left\{\prod_{k=1}^{n}\left(\frac{\tau_0+1-1/k}{\tau_1+1-1/k}\right)\left(\frac{\tau_1}{\tau_0}\frac{\tau_0+1}{\tau_1+1}\right)\right\}$$

where $\mathbf{t}$ represents a realization of the random variable $\mathbf{T}$. Now a term in the product defining $l_n(\mathbf{t}_{ij}(n))$ can be shown to be $\leqq 1$ if $k \geqq 1/(\tau_0\tau_1 - 1)$ provided $\tau_1 \geqq 1/\tau_0 = \tau^*$. Since

$$l_n(\mathbf{t}_{ij}(n)) \geqq l_n(\mathbf{t}_{ji}(n)) \tag{18}$$

we then have that this alternating sample will terminate the procedure only if $l_k(\mathbf{t}_{ij}(k)) \geqq a$ where k is the greatest integer in $1/(\tau_0\tau_1 - 1)$. Thus for a sufficiently large a we have exhibited a realization for which the procedure does not terminate with $\tau_1 \geqq \tau^*$. On the other hand if $\tau_1 < \tau^*$ then it can be shown that $l_n(\mathbf{t}_{ji}(n))$ becomes arbitrarily large as n increases.

In order to find a truncation point for the procedure we would like to use the following. If for some realization both $l_n(\mathbf{t}_{ij}(n)) \geqq a$ and $l_n(\mathbf{t}_{ji}(n)) \geqq a$ then the procedure is truncated and n is an upper bound to the truncation point. This would be true if following result is true.

CONJECTURE. Let $\mathbf{t}_{ij}(n)$ and $\mathbf{t}^*_{ij}(n)$ be two realizations of $\mathbf{T}_{ij}(n)$. Now if

$$l_n(\mathbf{t}_{ij}(n)) \geqq l_n(\mathbf{t}^*_{ij}(n)) \tag{19}$$

then

$$l_n(\mathbf{t}_{ji}(n)) \leqq l_n(\mathbf{t}^*_{ji}(n)).$$

The idea here is that if $l_n(\mathbf{t}_{ij}(n)) \geqq l_n(\mathbf{t}^*_{ij}(n))$ then $\mathbf{t}_{ij}(n)$ favors the elimination of Π_j

more than $\mathbf{t}_{ij}^*(n)$ does. Thus we would not expect $l_n(\mathbf{t}_{ji}(n)) \geqq l_n(\mathbf{t}_{ji}^*(n))$ which indicates $\mathbf{t}_{ji}(n)$ favoring the elimination of Π_i more than $\mathbf{t}_{ji}^*(n)$ does. Also, if the family of densities $\{g_{n\tau_{ij}}\}$ which form these likelihood ratios has a monotone likelihood ratio then the conjecture is true. Now if $l_n(\mathbf{t}_{ij}(n)) \geqq a$ and $l_n(\mathbf{t}_{ji}(n)) \geqq a$ then for any realization $\mathbf{t}_{ij}^*(n)$ we have from (19) that either $l_n(\mathbf{t}_{ij}^*(n)) \geqq a$ or $l_n(\mathbf{t}_{ji}^*(n)) \geqq a$. Hence n is an upper bound to the truncation point. Thus from our example of the alternating realization in the previous paragraph, we find that the procedure is truncated for $\tau_1 < \tau^*$ (provided of course that (19) is true.)

For the case $\tau_1 = \tau^*$, which corresponds to the SPRT, Savage and Sethuramen [15] have established that the sequential rank test terminates with probability one except at one point. They, however, strongly conjecture that the test also terminates at this point.

In order to specifically construct an upper bound to the procedure we use the following computing technique. Begin by letting $s_{11}(ij) = 1$. Now at stage n we set $s_{k,n}(ij) = s_{k,n-1}(ij)$ for $k = 1, \cdots, n-1$ and let

$$s_{nn}(ij) = 2n-1 \quad \text{if} \quad l_{n-1}(\mathbf{t}_{ij}(n-1)) \geqq l_{n-1}(\mathbf{t}_{ji}(n-1))$$

$$= 2n \quad \text{otherwise.}$$

This method produces a realization for which $l_n(\mathbf{t}_{ij}(n))$ and $l_n(\mathbf{t}_{ji}(n))$ are close to being equal for each n. Hopefully $l_n(\mathbf{t}_{ij}(n))$ and $l_n(\mathbf{t}_{ji}(n))$ will both remain below a until they simultaneously exceed it for some n. Then by using the conjectured results, the point, n, at which they exceed a is declared the truncation point for the procedure. In Table 3 some Monte Carlo results are given for the sequential procedure and the conjectured truncation points are included. It should be mentioned that for Tables 3 and 4, a total of 8000 trials were performed on truncated procedures and no violation of the truncation points was observed. Also in each instance the two likelihoods exceeded a for the first time with the same values of n. Thus we know without the conjecture that the listed truncation point minus one is a lower bound to the procedure's truncation point.

As mentioned, Tables 3 and 4 contain Monte Carlo results for the selection procedure. Both the format of the simulation and conclusions about the results are much the same as with the variance selection procedure experiment given in Tables 1 and 2.

5. Choice of τ_1. In previous sections it was suggested that τ_1 be chosen equal to $(\tau_* + \tau^*)/2$. Some insight into the choice of τ_1 can be obtained by considering the problem of selecting which of two normal populations has the larger mean.

Let X_{ij} be normally distributed with mean δ_i and variance $\frac{1}{2}$ $(i = 1, 2; j = 1, 2, \cdots)$. Then from Section 2 the selection procedure essentially is:

If $\sum_{k=1}^{n} (X_{ik} - X_{jk}) > a/(\tau_1 - \tau_0) + n(\tau_0 + \tau_1)/2$ stop and declare $\delta_j < \delta_i$.

If $\sum_{k=1}^{n} (X_{ik} - X_{jk}) < -a/(\tau_1 - \tau_0) - n(\tau_0 + \tau_1)/2$ stop and declare $\delta_j > \delta_i$.

TABLE 3

Empirical properties of the selection procedure based upon ranks with the parameters in the least favorable configuration

$\tau_1 = 1.0$				
P^*	$P(CS)$	$E(N)$	$E(S)$	Truncation point
.75	.900	24.8 (.3)	9.1 (.1)	19
.90	.960	34.7 (.4)	12.6 (.2)	27
.95	.986	40.9 (.4)	14.7 (.2)	32
.99	.996	57.8 (.5)	20.6 (.2)	46
$\tau_1 = 1.5$				
P^*	$P(CS)$	$E(N)$	$E(S)$	Truncation point
.75	.874	19.1 (.3)	7.3 (.1)	28
.90	.922	27.6 (.4)	10.4 (.2)	40
.95	.984	33.4 (.5)	12.4 (.2)	49
.99	.994	47.4 (.7)	17.4 (.3)	70
$\tau_1 = 2.0$				
P^*	$P(CS)$	$E(N)$	$E(S)$	
.75	.872	18.7 (.4)	7.3 (.2)	
.90	.962	27.0 (.5)	10.3 (.2)	
.95	.960	32.9 (.7)	12.5 (.3)	
.99	.996	46.5 (.8)	17.4 (.3)	

($\tau^* = 2$, 500 simulations per point)

TABLE 4

Observed number of stages with the parameters in the equal population configuration

	$E(S)$		
P^*	$\tau_1 = 1.0$	$\tau_1 = 1.5$	$\tau_1 = 2.0$
.75	11.1 (.1)	9.0 (.2)	10.0 (.3)
.90	16.5 (.2)	15.0 (.3)	17.0 (.5)
.95	21.0 (.2)	19.4 (.3)	23.8 (.7)
.99	32.0 (.3)	31.6 (.5)	48.3 (1.5)

Now if we define $Y_k = X_{ik} - X_{jk}$ and $\mu = \delta_i - \delta_j$ then the above selection procedure can be considered to be a sequential test of

$$H_0: \mu = -\tau^* \text{ vs. } H_1: \mu = \tau^*$$

where $Y_1, Y_2, \cdots$ are normal random variables with mean μ and variance one. This sequential test is:

(20) If $\sum_{k=1}^{n} Y_k > -\log(1-P^*)/(\tau_1+\tau^*)+n(\tau_1-\tau^*)/2$ stop and accept H_1.

If $\sum_{k=1}^{n} Y_k < \log(1-P^*)/(\tau_1+\tau^*)-n(\tau_1-\tau^*)/2$ stop and accept H_0.

The test's error probabilities are $\alpha = \beta = 1-P^*$ and we still require $\tau_1 > -\tau^*$ $(= \tau_0)$.

The stopping region in the $(n, \sum_{k=1}^{n} Y_k)$ plane consists of two intersecting straight lines. This is the same region as the one given by Anderson [1] (also considered by Paulson [14].) From (20) we see that the x-intercept is $2\log(1-P^*)/(\tau_1^2-\tau^{*2})$ and the y-intercepts are $\pm \log(1-P^*)/(\tau_1+\tau^*)$. Now if we compare the two regions $\tau_1 = \delta(< 0)$ and $\tau_1 = -\delta$ we see that their x-intercepts are the same and the y-intercepts are larger in magnitude for $\tau_1 = \delta$. Thus the $\tau_1 = -\delta$ region will always terminate first and hence will offer less overprotection and smaller sample sizes. This illustrates why it was suggested that $\tau_1 \geqq \tau_*(\tau_* = 0$ in this example.)

Anderson has given an approximate expression ((4.73) of [1]) for the probability of accepting H_0 in the test (20). It was observed using his approximation that the overprotection decreased with increasing τ_1. For the recommended choice $\tau_1 = (\tau^* + \tau_*)/2$ the overprotection based on the approximation was not too severe. Also the simulation results in Table 1–Table 4 indicate a preference for $(\tau^* + \tau_*)/2$ over either τ^* or τ_*.

Finally we should remark that from Section 4 a truncated sequential rank test can be developed on similar lines to (20). The test, however, is one of

$$H_0: \theta = 1/\tau^* \quad \text{vs.} \quad H_1: \theta = \tau^*$$

instead of the usual

$$H_0: \theta = 1 \quad \text{vs.} \quad H_1: \theta = \tau^*.$$

6. Final comments. The empirical results of Table 1–Table 4 showed that the truncated procedure with $\tau_1 = (\tau_* + \tau^*)/2$ compared favorably with the non-truncated procedure $\tau_1 = \tau^*$. However, more generally, a major difficulty with the application of sequential procedures is the positive probability of requiring a large number of observations. Often the experimenter is unwilling to take this risk even though he realizes that he stands to save, on the average, many observations. This aspect of the truncated selection procedure should be remembered when making comparisons with nontruncated selection procedures.

The second feature of the selection procedure, that of being fully sequential, may in many instances be unimportant. Such a case is when the major cost involved is with the sampling stage and not the observation itself. On the other hand, there are many situations where the observations themselves are costly. A particular example is the comparison of several medical treatments. In this situation it is essential to quickly eliminate inferior treatments instead of continuing them until sampling is completely terminated.

Finally, we may add that of the other two-sample sequential tests which may be used to construct corresponding selection procedures, the two most notable are Wald's sequential double dichotomy test and the two-sample sequential t-test. In the latter test, one may base the estimate of variance in the t-statistic upon all the observations instead of only those from the pair of populations being compared.

REFERENCES

[1] ANDERSON, T. W. (1960). A modification of the sequential probability ratio test to reduce the sample size. *Ann. Math. Statist.* **31** 165–197.

[2] BECHHOFER, R. E., KIEFER, J., and SOBEL, M. (1968). Sequential ranking and identification procedures. Univ. of Chicago Statistical Research Monograph Series.

[3] BECHHOFER, R. E. and SOBEL, M. (1954). A single-sample multiple decision procedure for ranking variances of normal populations. *Ann. Math. Statist.* **25** 273–289.

[4] BECHHOFER, R. E. and SOBEL, M. (1956). A scale invariant sequential multiple decision procedure for selecting the population with the smallest variance from k normal populations (preliminary report). Abstract, *Ann. Math. Statist.* **27** 219.

[5] BRADLEY, R. A. (1965). Topics in rank-order statistics. *Proc. Fifth Berkeley Symposium Math. Statist. Prob.* **1** 593–607.

[6] BRADLEY, R. A., MERCHANT, SARLA, D., and WILCOXON, F. (1966). Sequential rank test II. Modified two-sample procedures. *Technometrics* **8** 615–623.

[7] GHOSH, J. K. (1960). On the monotonicity of the OC of a class of sequential probability ratio tests. *Calcutta Statist. Assoc. Bull.* **2** 139–144.

[8] HALL, W. J., WIJSMAN, R. A., and GHOSH, J. K. (1965). The relationship between sufficiency and invariance with applications in sequential analysis. *Ann. Math. Statist.* **36** 575–614.

[9] HOEL, D. G. (1970). On the monotonicity of the OC of an SPRT. *Ann. Math. Statist.* **41** 310–314.

[10] HOEL, D. G. and MAZUMDAR, M. (1968). An extension of Paulson's selection procedure. *Ann. Math. Statist.* **39** 2067–2074.

[11] LEHMANN, E. L. (1953). The power of rank tests. *Ann. Math. Statist.* **24** 23–43.

[12] PARENT, E. A. (1965). Sequential ranking procedures. Stanford Univ., Department of Statistics, Tech. Report No. 80.

[13] PAULSON, E. (1964a). A sequential procedure for selecting the population with the largest mean from k normal populations. *Ann. Math. Statist.* **35** 174–180.

[14] PAULSON, E. (1964b). Sequential estimation and closed sequential decision procedures. *Ann. Math. Statist.* **35** 1048–1058.

[15] SAVAGE, I. R. and SETHURAMAN, J. (1966). Stopping time of a rank-order sequential probability ratio test based on Lehmann alternatives. *Ann. Math. Statist.* **37** 1154–1160.

[16] WALD, A. (1947). *Sequential Analysis.* Wiley, New York.

[17] WILCOXON, F., RHODES, L. J. and BRADLEY, R. A. (1963). Two sequential two-sample grouped rank tests with applications to screening experiments. *Biometrics* **19** 58–84.