The Annals of Mathematical Statistics
1968, Vol. 39, No. 3, 1040–1047

THE PERFORMANCE OF SOME SEQUENTIAL PROCEDURES FOR A RANKING PROBLEM[1]

By M. S. Srivastava and J. Ogilvie

University of Toronto

0. Summary. Srivastava (1966) has proposed two classes of asymptotically efficient sequential procedures for selecting the population with the largest mean when nothing is assumed about the form of the distribution functions except for the finiteness of unknown variance. In order to choose between these two classes, it seems desirable to study the performance of these two classes of sequential procedures for moderate sample sizes. In this paper, assuming that these populations are normal, the average sample sizes and the error probabilities actually obtained are computed to compare the two procedures. Results of calculation show that Procedure B (see Section 2) is better than Procedure A in that the average sample size is smaller while the error probabilities are almost the same (the superiority of procedure B was suggested by Srivastava (1966) for a different intuitive reason). At the same time the calculation helps to show the closeness of the error probability (both cases) to α. The calculation is on the lines of Ray (1957) and Robbins (1959) by generalizing a problem of the latter. Starr (1966) has recently carried out calculations to study the performance of a sequential procedure for finding fixed-width confidence bounds for the normal mean (see also Chow and Robbins, 1965).

1. Introduction. Consider k normal populations $\Pi_1, \Pi_2, \cdots, \Pi_k$; $\Pi_i : N(\mu_i, \sigma^2)$, where $N(\mu_i, \sigma^2)$ denotes a normal distribution with mean μ_i and variance σ^2. We assume that $\mu_1, \mu_2, \cdots, \mu_k, \sigma^2$ are unknown parameters, and the best category is the one with the largest μ. Denote the ranked μ's by

$$\mu_{[1]} \leqq \mu_{[2]} \leqq \cdots \leqq \mu_{[k]}. \tag{1.1}$$

The problem is to select the best category $\Pi_{[k]}$, i.e., the category with mean $\mu_{[k]}$ so that in each case the probability of making the correct decision exceeds a specified value (say, $1 - \alpha$) when the greatest mean exceeds all the other means by at least a specified amount (say, $d2^{\frac{1}{2}}$). Here d and α are constants which are specified by the experimenter in advance of the experiment on the basis of practical considerations. For this problem, 'known as a ranking problem', Srivastava (1966) has proposed two asymptotically 'efficient' sequential procedures; both satisfying

$$\lim_{d\to 0} P[\text{eliminating } \Pi_{[k]} \mid \mu_{[k]} - \mu_{[k-1]} \geqq d2^{\frac{1}{2}}] \leqq \alpha. \tag{1.2}$$

Received 5 July 1967.

[1] Research of the senior author was supported partially by the Office of Naval Research, Contract Nonr-1858(05) at Princeton University and partially by the Summer Research Institute of the Canadian Mathematical Congress. Research of the junior author was supported by grant APA-176 from the National Research Council of Canada.

In order to compare these two classes of sequential procedures we compute for both procedures

$$e(\lambda) = P[\text{eliminating } \Pi_{[k]} \mid \mu_{[k]} - \mu_{[k-1]} \geqq d2^{\frac{1}{2}}], \tag{1.3}$$

the error probabilities actually attained. Also, we compute the average sample size.

Corresponding to these two classes of asymptotically efficient sequential procedures (see Section 2) we obtain two Stein's (1945) two-stage procedures; one of them has been proposed by Bechhofer, Dunnett and Sobel (1954). Another sequential procedure has recently been proposed by Paulson (1964). Thus in all, we have five sequential procedures for the problem. The latter three procedures require a first stage sample to estimate the variance σ^2; the subsequent observations are not utilized in estimating σ^2. It seems intuitively inefficient not to utilize all of the sample. Thus, it would be desirable to compare these five procedures. Since no explicit formula of average sample size is available for Paulson's (1964) procedure, a comparison is not possible. Also, there is an arbitrariness in the choice of 'λ' in Paulson's procedure and it is not clear what considerations, if any, one should use in choosing 'λ'. Of the four remaining procedures only the two asymptotically efficient procedures proposed by Srivastava (1966) will be compared. The two based on Stein's two-stage procedure are asymptotically less efficient (Seelbinder (1953)). Also Starr (1966) has pointed out that a poor guess of initial sample size can be very costly in number of observations with a two-stage procedure while a sequential procedure is always reasonably efficient.

In the present investigation, the computation has been carried out for $\alpha = .05$ and $k = 2, 4, 6$. For $k = 2$ the two asymptotically efficient sequential procedures are identical.

2. The two procedures. Let $X_s^{(i)}$ denote the sth observation from category Π_i ($i = 1, 2, \cdots, k$, and $s = 1, 2, \cdots$). We assume throughout the paper that $\{X_s^{(i)}\}$ is a sequence of mutually independent random variables. The random variables $X_s^{(i)}$ are assumed to be normally distributed with mean μ_i and variance σ^2. Let

$$\begin{aligned} \bar{X}_n^{(i)} &= (1/n) \textstyle\sum_{s=1}^{n} X_s^{(i)}, \\ V_n &= (1/mk) \textstyle\sum_{i=1}^{k} \sum_{s=1}^{n} (X_s^{(i)} - \bar{X}_n^{(i)})^2, \qquad n > 1, \\ m &= n - 1. \end{aligned} \tag{2.1}$$

We now describe formally the two procedures.

Procedure A. For $x > 0$, let

$$\begin{aligned} \Phi(-x) &= (2\pi)^{-\frac{1}{2}} \textstyle\int_{-\infty}^{-x} e^{-t^2/2}\, dt, \\ \Phi_{km}(-x) &= [\Gamma(\tfrac{1}{2}(km+1))/(km\pi)^{\frac{1}{2}}\Gamma(\tfrac{1}{2}km)] \textstyle\int_{-\infty}^{-x} (1 + t^2/km)^{-(km+1)/2}\, dt. \end{aligned} \tag{2.2}$$

Suppose constants a and a_{km} are chosen so that

$$\Phi(-a) = \alpha/(k-1), \qquad \Phi_{km}(-a_{km}) = \alpha/(k-1). \tag{2.3}$$

TABLE 1

Error probabilities and average sample numbers for $k = 2$ where the two procedures are identical

λ	error probability $e(\lambda)$	average sample number $O(\lambda)$
0.75	.0150	3.01
1.00	.0347	3.85
1.25	.0509	5.00
1.50	.0610	6.51
1.75	.0661	8.42
2.00	.0677	10.74
2.25	.0671	13.48
2.50	.0655	16.63
2.75	.0635	20.16
3.00	.0616	24.07
3.25	.0599	28.33

TABLE 2

Comparison of error probabilities, $e(\lambda)$, and average sample numbers, $O(\lambda)$, for Procedures A and B

λ	$k = 4$				$k = 6$			
	$e_A(\lambda)$	$e_B(\lambda)$	$O_A(\lambda)$	$O_B(\lambda)$	$e_A(\lambda)$	$e_B(\lambda)$	$O_A(\lambda)$	$O_B(\lambda)$
.75	.0275	.0283	3.68	3.44	.0345	.0350	4.00	3.65
1.00	.0504	.0496	5.38	5.00	.0515	.0505	6.20	5.65
1.25	.0591	.0576	7.78	7.22	.0529	.0521	9.22	8.42
1.50	.0589	.0577	10.86	10.09	.0506	.0504	12.97	11.88
1.75	.0562	.0554	14.57	13.57	.0493	.0493	17.38	15.97
2.00	.0537	.0533	18.85	17.61	.0489	.0490	22.47	20.67
2.25	.0522	.0519	23.70	22.17	.0489	.0490	28.22	25.98
2.50	.0513	.0511	29.10	27.24	.0490	.0490	34.65	31.91
2.75	.0509	.0506	35.05	32.85	.0494	.0492	41.63	38.42
3.00	.0505	.0504	41.57	38.97	—	—	—	—
3.25	.0503	.0502	48.66	45.63	—	—	—	—

Then

$$\lim_{m\to\infty} a_{km} = a, \tag{2.4}$$

and the sequence $\{a_{km}\}$ determines a member of the class $\mathcal{C}$ of sequential procedures defined as follows:

(I) Sample one observation at a time from each population and stop according to the stopping variable N defined by

$$N = \text{smallest } n > 1 \text{ such that } V_n \leqq d^2 n/a_{km}^2, \tag{2.5}$$

where $m = n - 1$.

(II) When sampling is stopped at $N = n$, select the population with the largest sample mean as the best category.

PROCEDURE B. For $x > 0$, let

$$(2.6)\quad G_k(x) = (2\pi)^{-(k-1)/2}(\det \Sigma)^{-\frac{1}{2}} \int_{-\infty}^{x} \int_{-\infty}^{x} \cdots \int_{-\infty}^{x} \exp\left[\tfrac{1}{2}u'\Sigma^{-1}u\right] du_1 \cdots du_{k-1}$$
$$= (2\pi)^{-k/2} \int_{-\infty}^{\infty} \left(\int_{-\infty}^{z+x2^{\frac{1}{2}}} e^{-t^2/2}\, dt\right)^{k-1} e^{-z^2/2}\, dz,$$

where $U' = (U_1, \cdots, U_{k-1})$ follow a joint distribution with zero means and covariance matrix Σ,

$$\Sigma = \begin{pmatrix} 1 & \frac{1}{2} & \cdots & \frac{1}{2} \\ \frac{1}{2} & 1 & \cdots & \frac{1}{2} \\ \vdots & & & \vdots \\ \frac{1}{2} & \frac{1}{2} & \cdots & 1 \end{pmatrix}$$

and; let

$$G_{km}(x) = \frac{2^{(k-1)/2}\Gamma(\frac{1}{2}(km + k - 1))}{k^{1/2}(\frac{1}{2}\pi km)^{(k-1)/2}\Gamma(\frac{1}{2}km)} \int_{-\infty}^{+x}\int_{-\infty}^{+x} \cdots \int_{-\infty}^{+x}$$
$$\cdot[1 + 2(k-1)(k^2m)^{-1}\left(\textstyle\sum t_i^2 - 2/(k-1)\right.$$
$$(2.7)\qquad \left.\cdot \textstyle\sum_{i<j} t_i t_j\right)]^{-\frac{1}{2}(km+k-1)}\, dt_1 \cdots dt_{k-1}$$
$$= (2\pi)^{-k/2}(\Gamma(km/2))^{-1} \int_0^{\infty} \left[\int_{-\infty}^{\infty} \left(\int_{-\infty}^{z+ux(2/km)^{1/2}} e^{-t^2/2}\, dt\right)^{k-1}\right.$$
$$\left. \cdot\, e^{-z^2/2}\, dz \right] u(u^2/2)^{km/2-1} e^{-u^2/2}\, du$$

where $T = (T_1, \cdots, T_{k-1})$ follow a joint multivariate t-distribution. Suppose constants a^* and a_{km}^* are chosen so that

$$(2.8)\qquad G_k(a^*) = 1 - \alpha, \qquad G_{km}(a_{km}^*) = 1 - \alpha.$$

Then

$$(2.9)\qquad \lim_{m\to\infty} a_{km}^* = a^*.$$

The Steps (I) and (II) for Procedure B are the same as for A except that in (2.5) a_{km} is replaced by a_{km}^*.

3. Preliminaries for procedures A and B. Define

$$U_j^{(i)} = (X_j^{(i)} - \mu_i)/\sigma, \qquad i = 1, 2, \cdots, k; \quad j = 1, 2, \cdots, n,$$
$$(3.1)\quad Y_j^{(i)} = [U_1^{(i)} + \cdots + U_j^{(i)} - jU_{j+1}^{(i)}]/[j(j+1)]^{\frac{1}{2}}, \quad j = 1, 2, \cdots, n-1,$$
$$Y_n^{(i)} = n^{-\frac{1}{2}}(U_1^{(i)} + \cdots + U_n^{(i)}) = n^{\frac{1}{2}}(\bar{X}_n^{(i)} - \mu_i)/\sigma.$$

Then $U_j^{(i)}$ are independent $N(0, 1)$ and since the transformation to the $Y_j^{(i)}$, $j = 1, 2, \cdots, n$, is orthogonal, the same is true of $Y_j^{(i)}$. From (2.1) it is easily seen that

$$(3.2)\qquad V_n = (\sigma^2/km) \textstyle\sum_{i=1}^{k} \sum_{\beta=1}^{m} Y_\beta^{(i)2}, \qquad m = n - 1,$$

where $Y_\beta^{(i)2}$ are independent chi-square with one degree of freedom. V_n can also be written as

$$V_n = (2\sigma^2/km) \textstyle\sum_{\beta=1}^{m} Z_\beta, \qquad m = n - 1, \tag{3.3}$$

where Z_β's are independently and identically distributed with density

$$\{1/\Gamma(k/2)\} z^{k/2-1} e^{-z}, \qquad z > 0. \tag{3.4}$$

For computational convenience, we will consider in this paper the case when k is of the form

$$k = 2(r + 1), \qquad r = 0, 1, 2, \cdots. \tag{3.5}$$

Consequently from (2.5), N is the first positive integer $n = m + 1 > 1$, such that

$$\textstyle\sum_1^m Z_\beta \leqq (r + 1)(m + 1)m/\lambda^2 a_{km}^2 = b_{m+1,r,\lambda}, \tag{3.6}$$

where

$$\lambda = \sigma/d, \qquad b_{m+1,r,\lambda} = (r + 1)(m + 1)m/\lambda^2 a_{km}^2. \tag{3.7}$$

Let $p_m(\lambda, r, a_{km})$ be the probability that m is the first positive integer for which

$$Z_1 + \cdots + Z_m \leqq b_{m+1,r,\lambda}. \tag{3.8}$$

Then, from a generalization of Robbins' (1959) result (see Srivastava (1967)), we have

$$p_m = F_m(\infty) - F_{m+1}(\infty), \tag{3.9}$$

where

$$\begin{aligned} F_1(\infty) &= 1, \\ F_m(\infty) &= e^{-b_m}(\Gamma(r + 1))^{-1} \textstyle\sum_{j=1}^{m(r+1)-1} h_m^{(j)}(b_m), \qquad m \geqq 2; \end{aligned} \tag{3.10}$$

$$\begin{aligned} h_1(x) &= x^r, \\ h_m(x) &= \textstyle\sum_{j=1}^{m(r+1)-1} [(x - b_m)^j j!^{-1} h_m^{(j)}(b_m)], \quad x \geqq b_m, m = 2, 3, \cdots, \end{aligned} \tag{3.11}$$

where $h_m^{(j)}(b_m) = (d^j/dx^j) h_m(x)|_{x=b_m}$.

In the expressions (3.9), (3.10) and (3.11), for convenience of notation, we have omitted the arguments λ, r, a_{km}.

For computational purposes, we need to simplify (3.11) further for $m = 2, 3, \cdots, b_m \geqq x$ and $r = 0, 1, 2$.

For $r = 0$

$$h_m(x) = \textstyle\sum_{j=1}^{m-1} (x - b_m)^j (j!)^{-1} h_{m-j}(b_m); \tag{3.12}$$

for $r = 1$

$$\begin{aligned} h_m(x) = &\textstyle\sum_{j=1}^{m-1} (x - b_m)^{2j} [(2j)!]^{-1} h_{m-j}(b_m) \\ &+ \textstyle\sum_{j=1}^{m-1} (x - b_m)^{2j+1} [(2j + 1)!^{-1} h_{m-j}^{(1)}(b_m); \end{aligned} \tag{3.13}$$

for $r = 2$

$$\begin{aligned}(3.14)\qquad h_m(x) &= \sum_{j=1}^{m-1} (x - b_m)^{3j}[(3j)!]^{-1}h_{m-j}(b_m) \\ &+ \sum_{j=1}^{m-1} (x - b_m)^{3j+1}[(3j+1)!]^{-1}h_{m-j}^{(1)}(b_m) \\ &+ \sum_{j=1}^{m-1} (x - b_m)^{3j+2}[(3j+2)!]^{-1}h_{m-j}^{(2)}(b_m).\end{aligned}$$

To solve (3.13), a recurrence relation similar to (3.13) is needed for $h_m^{(1)}$. It can be obtained by differentiating (3.13). Similarly to solve (3.14), the recurrence relations needed for $h_m^{(1)}(x)$ and $h_m^{(2)}(x)$ can be obtained by differentiating (3.14) twice.

4. Evaluation of error probabilities $e(\lambda)$ and average sample number. Let δ_2 denote the parameter configuration $\mu_{[k]} \geqq \mu_{[k-1]} + d2^{\frac{1}{2}}$, and let δ_2^* denote the parameter configuration $\mu_k \geqq \mu_j + d2^{\frac{1}{2}}$ for $j = 1, 2, \cdots, k - 1$. It is obvious from the symmetry of the sequential procedure that

$$\begin{aligned}(4.1)\qquad e(\lambda) &= P\,[\text{incorrect decision} \mid \delta_2] \\ &= P\,[\Pi_k \text{ is eliminated} \mid \delta_2^*].\end{aligned}$$

Let

$$\begin{aligned}(4.2)\qquad \mu_k &= \mu_\nu + d_\nu 2^{\frac{1}{2}}, \qquad \nu = 1, 2, \cdots, k-1, \\ \lambda &= \sigma/d,\end{aligned}$$

$$(4.3)\qquad p_m(\lambda, r, a_{km}) = P_A(N = m+1); \qquad p_m^*(\lambda, r, a_{km}^*) = P_B(N = m+1),$$

where P_A and P_B denotes the probabilities under Procedures A and B respectively. Then (see Srivastava (1966)) for procedure A, $e_A(\lambda, r)$ is given by

$$\begin{aligned}(4.4)\qquad e_A(\lambda, r) &\leqq (k-1)P[N^{\frac{1}{2}}(\bar{X}_N^{(k)} - \bar{X}_N^{(\nu)} - d_\nu 2^{\frac{1}{2}})/\sigma 2^{\frac{1}{2}} \leqq -dN^{\frac{1}{2}}/\sigma \mid \delta_2^*] \\ &= (k-1)\sum_{m=1}^{\infty} P[(m+1)^{\frac{1}{2}}(\bar{X}_{m+1}^{(k)} - \bar{X}_{m+1}^{(\nu)} - d_\nu 2^{\frac{1}{2}})/\sigma 2^{\frac{1}{2}} \\ &\leqq -d(m+1)^{\frac{1}{2}}/\sigma \mid \delta_2^*, N = m+1]p_m(\lambda, r, a_{km}) \\ &= (k-1)\sum_{m=1}^{\infty} p_m(\lambda, r, a_{km})\Phi(-(m+1)^{\frac{1}{2}}/\lambda),\end{aligned}$$

since $(\bar{X}_n^{(k)} - \bar{X}_n^{(\nu)})$ is independent of V_{n-1}.

Similarly, for Procedure B, $e_B(\lambda, r)$ is given by

$$\begin{aligned}(4.5)\qquad e_B(\lambda, r) &\leqq P[N^{\frac{1}{2}}(\bar{X}_N^{(\nu)} - \bar{X}_N^{(k)} + d_\nu 2^{\frac{1}{2}})/\sigma 2^{\frac{1}{2}} > dN^{\frac{1}{2}}/\sigma \text{ for all} \\ &\qquad \nu = 1, 2, \cdots, k-1 \mid \delta_2^*] \\ &= 1 - \sum_{m=1}^{\infty} p_m^*(\lambda, r, a_{km}^*)G_k((m+1)^{\frac{1}{2}}/\lambda).\end{aligned}$$

Φ and G_k have been defined by (2.2) and (2.6) respectively.

The average sample numbers for the two Procedures A and B are respectively given by

$$\begin{aligned}(4.6)\qquad O_A(\lambda, r) &= 1 + \sum_{m=1}^{\infty} mp_m(\lambda, r, a_{km}) \\ &= 1 + \sum_{m=1}^{\infty} F_{m,\lambda,r}(\infty)\end{aligned}$$

and

$$O_B(\lambda, r) = 1 + \sum_{m=1}^{\infty} mp_m^*(\lambda, r, a_{km}^*) \tag{4.7}$$
$$= 1 + \sum_{m=1}^{\infty} F_{m,\lambda,r}^*(\infty).$$

5. Computational procedure. First the values of a_{km} and a_{km}^* were obtained for $r = 0, 1, 2$ and for $m = 1, 2, \cdots, 65$ with $\alpha = .95$. There are no special problems in calculating a_{km}. The a_{km}^* were obtained by inverse iteration using the secant method with an absolute error criterion of 10^{-4}. Starting values were taken from the table given by Dunnett (1955) to minimize the number of iterations. In calculating $G_{km}(x)$ Romberg integration (see, e.g., Ralston (1966)) was used for the two outer integrals. The innermost integral is a normal error function and a rational approximation was used. The maximum limits for the outer most integral were 0 and 17 with a convergence criterion of 10^{-5} relative error. The limits for the second integral were -7.5 and 7.0 with a relative error criterion of 10^{-7}. In both cases the limits were reduced automatically as long as the function value was less than 10^{-9}.

The polynomials $h_m(x)$ were calculated recursively. The values increase rapidly with m and require scaling. The original scaling is easily recovered when calculating the $F_{m,r,\lambda}(\infty)$. The individual terms of the series defining the $h_m(x)$ being of the form $(x - b_m)^j/j!$ have to be computed logarithmically to avoid overflow.

The number of terms required to compute $e_A(\lambda, r)$ and $O_A(\lambda, r)$ was found to depend primarily on λ. The series oscillate when r is 1 or 2 so that a certain amount of trial and error was needed to determine the appropriate number of terms. The final choice was 20 terms for λ less than 1.5, 40 terms for λ between 1.5 and 2.0 inclusive and 65 terms for λ greater than 2.0. The values in the table for $r = 2$, $\lambda = 3.00, 3.25$ were omitted because of the overflow problems that were encountered despite scaling.

The numerical accuracy of the computation was tested by perturbing the $b_{m,r,\lambda}$ by a rectangularly distributed random relative error in the interval $(-10^{-4}, 10^{-4})$ for $\lambda = 1.0, 2.5$ and $r = 0, 1, 2$ for both procedures. The maximum effect of the perturbation on the values of e_A and O_A was 1 in the fifth significant digit. The results of the computation are therefore probably good to 4 digits and certainly to 3.

Acknowledgment. The authors wish to thank Professor John Tukey for going through the first draft of this paper and to Professor Keith Hastings for his advice on numerical integration. The calculations were carried out by Mrs. Bryna Fellen to whom we express our sincere thanks.

REFERENCES

BECHHOFER, R. E., DUNNETT, C. W., and SOBEL, M. (1954). A two-sample multiple decision procedure for ranking means of normal populations with a common unknown variance. *Biometrika* **41** 170–176.

CHOW, Y. S. and ROBBINS, H. (1965). On the asymptotic theory of fixed-width sequential confidence intervals for the mean. *Ann. Math. Statist.* **36** 457–462.

DUNNETT, C. W. (1955). A multiple comparison procedure for comparing several treatments with a control. *J. Amer. Statist. Assoc.* **50** 1096–1121.

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

RALSTON, A. (1965). *A First Course in Numerical Analysis*. McGraw-Hill, New York.

RAY, W. D. (1957). Sequential confidence intervals for the mean of a normal distribution with unknown variance. *J. Roy. Statist. Soc. Ser. B* **19** 133–143.

ROBBINS, H. (1959). Sequential estimation of the mean of a normal population. *Probability Statistics—The Harald Crâmér Volume*, 235–245. Almquist and Wiksell, Uppsala.

SEELBINDER, B. M. (1953). On Stein's two-stage sampling scheme. *Ann. Math. Statist.* **24** 640–649.

SRIVASTAVA, M. S. (1966). Some asymptotically efficient sequential procedures for ranking and slippage problems. *J. Roy. Statist. Soc. B* **28** 370–380.

SRIVASTAVA, M. S. (1967). The performance of a sequential procedure for a slippage problem (in preparation).

STARR, N. (1966). The performance of a sequential procedure for the fixed-width interval estimation of the mean. *Ann. Math. Statist.* **37** 36–50.

STEIN, C. (1945). A two-sample test for a linear hypothesis whose power is independent of the variance. *Ann. Math. Statist.* **16** 243–258.