A SEQUENTIAL PROCEDURE FOR COMPARING SEVERAL EXPERIMENTAL CATEGORIES WITH A STANDARD OR CONTROL[1]

By Edward Paulson

Queens College

1. Introduction and summary. A problem that seems to be of some practical interest is that of comparing k experimental categories with a standard (or control) in order to determine if any of the experimental categories are better than the standard (or control) and if so, to find the best one. We might for example be concerned with evaluating several new drugs in relation to a placebo, or with evaluating several new alloys in relation to some standard alloy. This problem has been discussed in the non-sequential case in [1] and [2].

In the present paper we will be concerned with a sequential treatment of the problem. In our notation, the subscript 0 will always refer to the standard (or control), the subscript i will range from 0 to k, and the subscript j will range from 1 to k. Let X_i be the random variable resulting from a measurement with the ith category π_i. We will assume the probability distribution of X_i is the same for each i except for a single parameter θ, which might vary from category to category. The probability density of X_i (or the probability of $X_i = x$ in the discrete case) will be denoted by $f(x, \theta_i)$ and this is assumed known, except for the value of θ_i. For simplicity, we will suppose that the larger the value of θ, the more desirable a category is. Let R_0 denote the relation $\theta_1 = \theta_2 = \cdots = \theta_k = \theta_0$, and let R_j denote the relation $\theta_1 = \cdots = \theta_{j-1} = \theta_{j+1} = \cdots = \theta_k = \theta_0$ and $\theta_j = \theta_0 + \Delta$, where $\Delta > 0$. Let D_0 denote the decision that none of the experimental categories is better than the standard (or control) [that is, $\theta_s \leqq \theta_0$ for $s = 1, 2, \cdots, k$] and let D_j denote the decision that π_j is the best experimental category and is better than the standard [that is, $\theta_j = \max(\theta_1, \theta_2, \cdots, \theta_k) > \theta_0$]. We will develop a sequential procedure for choosing one of the $k + 1$ decisions $(D_0, D_1, \cdots, D_k)$ so that the probability of selecting D_0 when R_0 is true is $\geqq 1 - \alpha$, and the probability of selecting D_j when R_j is true is $\geqq 1 - \beta$ for each $j, j = 1, 2, \cdots, k$. First we will treat the case where θ_0 is known and $f(x, \theta)$ is arbitrary, and then consider the case where θ_0 is unknown and

$$f(x, \theta) = (2\pi\sigma^2)^{-\frac{1}{2}} \exp[-\tfrac{1}{2}(x - \theta)^2/\sigma^2].$$

A rather interesting feature of the sequential procedures that are developed is that inferior categories can be eliminated before the final stage of the experiment, which tends to decrease the number of measurements required to reach a decision. An investigation is made (when θ_0 is known) of the efficiency of the

Received May 31, 1961; revised August 7, 1961.

[1] This work was done at Cornell University and was supported by the Office of Naval Research.

sequential procedure when compared with the corresponding "best" non-sequential procedure, which indicates that a substantial saving in the average number of measurements required to reach a decision is possible with the sequential procedure.

2. The sequential procedure when θ_0 is known. Let $X_{j\alpha}$ denote the αth measurement with category π_j. We will assume that for each j the measurements $\{X_{j\alpha}\}$, $\alpha = 1, 2, \cdots$, are independent and have the same distribution as X_j. Let

$$y_{j\alpha} = \log [f(X_{j\alpha}, \theta_0 + \Delta)/f(X_{j\alpha}, \theta_0)].$$

Let a and b be two constants with $b < 0 < a$. We suppose the sampling is done in the following manner. At the first stage one measurement is taken with each experimental category. If no decision is taken at the $(r-1)$st stage, then at the rth stage one measurement is taken with each experimental category which has not been eliminated on any of the first $r-1$ stages, where a category π_j is eliminated whenever $\sum_\alpha y_{j\alpha} \leqq b$.

For each r, let $M = M(r)$ denote the integer for which

$$\sum_{\alpha=1}^{r} y_{M\alpha} = \max_j \left\{ \sum_{\alpha=1}^{r} y_{j\alpha} \right\}$$

for all j for which $\sum_{\alpha=1}^{s} y_{j\alpha} > b$ for $s = 1, 2, \cdots, r-1$. [In the discrete case, if M is not uniquely defined at any stage because $\max_j (\sum_\alpha y_{j\alpha})$ is assumed on a set of s values of j, we can select M by any chance mechanism which assigns probability $1/s$ to each of these s values]. We now propose the following procedure for choosing between $D_0, D_1, \cdots, D_k$. At the rth stage $r = 1, 2, \cdots$,

(1)
(a) If $\sum_{\alpha=1}^{r} y_{M\alpha} \geqq a$, select D_M.
(b) If $\sum_{\alpha=1}^{r} y_{M\alpha} \leqq b$, select D_0.
(c) If $b < \sum_{\alpha=1}^{r} y_{M\alpha} < a$, postpone making a decision, go on to the $(r+1)$st stage of the experiment, and repeat the above procedure.

Let n = the number of stages required to reach a decision using the sequential procedure (1), and let n_j = the smallest integer for which $\sum_{\alpha=1}^{n_j} y_{j\alpha} \geqq a$ or $\sum_{\alpha=1}^{n_j} y_{j\alpha} \leqq b$. Then clearly $n \leqq \max(n_1, n_2, \cdots, n_k)$. It easily follows that the sequential procedure (1) will terminate with probability one, and also that $E(n)$ is always finite, provided $P[y_{j\alpha} = 0] < 1$.

We will now determine the constants a and b as functions of α and β. Let $P(D_i \mid R_s)$ denote the probability of choosing D_i when R_s is true. By symmetry, $P(D_j \mid R_j)$ is independent of $j, j = 1, 2, \cdots, k$. Since the probability of coming to a decision is one, we have

$$\begin{aligned} 1 - P(D_0 \mid R_0) &= \sum_{j=1}^{k} P(D_j \mid R_0) \\ &\leqq \sum_{j=1}^{k} P\left[\sum_{\alpha=1}^{r} y_{j\alpha} \geqq a \quad \text{for some} \quad r < \infty \mid R_0 \right]. \end{aligned}$$

Also

$$1 - P(D_j \mid R_j) = 1 - P(D_1 \mid R_1)$$
$$= P(D_0 \mid R_1) + \sum_{s=2}^{k} P(D_s \mid R_1)$$
$$\leqq P\left[\sum_{\alpha=1}^{r} y_{1\alpha} \leqq b \quad \text{for some} \quad r < \infty \mid R_1\right]$$
$$+ \sum_{s=2}^{k} P\left[\sum_{\alpha=1}^{r} y_{s\alpha} \geqq a \quad \text{for some} \quad r < \infty \mid R_1\right].$$

Now it is well known (see [3]) that $P[\sum_{\alpha=1}^{r} y_{j\alpha} \geqq a \text{ for some } r < \infty \mid R_0] \leqq e^{-a}$, $P[\sum_{\alpha=1}^{r} y_{1\alpha} \leqq b \text{ for some } r < \infty \mid R_1] \leqq e^{b}$, and $P[\sum_{\alpha=1}^{r} y_{s\alpha} \geqq a \text{ for some } r < \infty \mid R_1] \leqq e^{-a}$ for $s > 1$. In order to satisfy the requirement that $P(D_0 \mid R_0) \geqq 1 - \alpha$ and $P(D_j \mid R_j) \geqq 1 - \beta$, we therefore determine a and b so that $ke^{-a} \leqq \alpha$ and $e^{b} + (k-1)e^{-a} \leqq \beta$. Among all pairs (a, b) which satisfy these two inequalities, we will select the pair which minimizes the upper bound for the expected number of measurements required to reach a decision when R_j is true. First let

$$c_0 = E(y_{1\alpha} \mid R_0), \qquad c_1 = E(y_{1\alpha} \mid R_1), \qquad \lambda^* = kc_0\beta\{\alpha(k-1)[c_0 - (k-1)c_1]\}^{-1},$$

and $\lambda = \min(1, \lambda^*)$. We assume that $E(y_{j\alpha})$ exists and $P[y_{j\alpha} = 0] < 1$, which together with the standard inequality $E \log X < \log E(X)$ when $P[X = 1] < 1$, implies that $c_1 > 0$ and $c_0 < 0$. Now using the upper bound given in (3) in the next section, it is a straightforward matter to show that the required pair (a, b) is given by $a = \log[k/\lambda\alpha]$ and $b = \log[\beta - (k-1)\lambda\alpha/k]$.

3. The efficiency of the sequential procedure (θ_0 known). Let T = the total number of measurements required by the sequential procedure (1) to reach a decision. In this section we will attempt to give some indication of the efficiency of the sequential procedure by finding upper bounds for $E(T \mid R_0)$ and $E(T \mid R_j)$, and then comparing these bounds with the sample size required for the comparable optimum non-sequential test. The comparison will be carried out explicitly only for the case $f(x, \theta)$ is the normal probability density function with a known standard deviation. However because of the central limit theorem one would strongly suspect that for small values of Δ the comparison is valid for a general $f(x, \theta)$ subject to some mild restrictions.

It is easy to see that $T \leqq \sum_{j=1}^{k} n_j$. Therefore

$$E(T) \leqq \sum_{j=1}^{k} E(n_j), \qquad E(T \mid R_0) \leqq kE(n_1 \mid R_0),$$

$$E(T \mid R_j) = E(T \mid R_1) \leqq E(n_1 \mid R_1) + (k-1)E(n_2 \mid R_1).$$

Neglecting the excess over the boundary, it is obvious that

$$E(n_2 \mid R_1) = E(n_1 \mid R_0) \leqq b/c_0$$

and $E(n_1 \mid R_1) \leqq a/c_1$. We now find that

$$E(T \mid R_0) \leqq kb/c_0, \tag{2}$$

$$E(T \mid R_j) \leqq a/c_1 + (k-1)b/c_0. \tag{3}$$

We now consider the special case $f(x, \theta) = (2\pi\sigma^2)^{-\frac{1}{2}} \exp\{-\frac{1}{2}[(x-\theta)/\sigma)]^2\}$ For simplicity, take $\theta_0 = 0$ and $\sigma = 1$. Then for the sequential procedure, $y_{1\alpha} = \Delta(X_{1\alpha} - \Delta/2)$

$$E(y_{1\alpha} \mid R_0) = -\Delta^2/2, \qquad E(y_{1\alpha} \mid R_1) = \Delta^2/2, \qquad \text{and} \qquad \lambda^* = \beta[\alpha(k-1)]^{-1}.$$

Now we discuss the non-sequential procedure for choosing between $D_0, D_1, \cdots, D_k$ based on N measurements with each experimental category. Let $\bar{x}_j = \sum_{\alpha=1}^{N} X_{j\alpha}/N$, let $\bar{x}_M = \max(\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_k)$ let $w = 1 - (1-\alpha)^{1/k}$, and let λ_p be defined by the equation $(2\pi)^{-\frac{1}{2}} \int_{\lambda_p}^{\infty} \exp(-t^2/2)\, dt = p$. Consider the procedure

$$\begin{aligned} &\text{(a) If } \bar{x}_M < \lambda_w N^{-\frac{1}{2}}, \text{ select } D_0. \\ &\text{(b) If } \bar{x}_M \geqq \lambda_w N^{-\frac{1}{2}}, \text{ select } D_M. \end{aligned} \tag{4}$$

It is a straightforward matter using the methods of [2] or [4] to show that (4) is optimum in the following sense: among all non-sequential procedures based on N measurements with each experimental category for which $P(D_0 \mid R_0) = 1 - \alpha$ and for which $P(D_j \mid R_j)$ has the same value for each $j, j = 1, 2, \cdots, k$, the procedure given in (4) is best in that it maximizes $P(D_j \mid R_j)$.

For a fixed value of α, the value of $P(D_j \mid R_j)$ for the procedure in (4) is a monotonically increasing function of N. We now have to find $N = N(\alpha, \beta)$ so that the additional restriction $P(D_j \mid R_j) = 1 - \beta$ is satisfied. Since

$$P(D_j \mid R_j) = P(D_1 \mid R_1) = P[\bar{x}_1 = \max(\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_k) \text{ and } \bar{x}_1 > \lambda_w N^{-\frac{1}{2}} \mid R_1],$$

we can apply Bonferroni's Inequality, and we get

$$\begin{aligned} P[\bar{x}_1 > \lambda_w N^{-\frac{1}{2}} \mid R_1] - (k-1)P[\bar{x}_1 < \bar{x}_2 \mid R_1] &\leqq P[D_j \mid R_j] \\ &\leqq P[\bar{x}_1 > \lambda_w N^{-\frac{1}{2}} \mid R_1]. \end{aligned} \tag{5}$$

If we now determine N so that $P[\bar{x}_1 > \lambda_w N^{-\frac{1}{2}} \mid R_1] = 1 - \beta$, we find

$$N = (\lambda_w + \lambda_\beta)^2/\Delta^2 \tag{6}$$

Using (5) it can be shown tbat the value of N given by (6) is a good approximation to the value of $N(\alpha, \beta)$ required in order to have $P(D_0 \mid R_0) = 1 - \alpha$ and $P(D_j \mid R_j) = 1 - \beta$ when α and β are fairly small. Equally important, however, is the fact that the value of N given by (6) is conservative from the standpoint of comparing the sequential procedure with the non-sequential procedure, since the true value of $N(\alpha, \beta)$ is actually somewhat greater than the value given by (6).

We are now in a position to make a numerical comparison between the sequential and the non-sequential procedure, and this is done in Table I.

TABLE I

Values of kN and the corresponding upper bounds for $E(T \mid R_0)$ and $E(T \mid R_i)$ for the normal distribution with unit variance*

α	β	k	kN	Upper bound for $E(T \mid R_0)$	Upper bound for $E(T \mid R_i)$
.05	.05	2	26	15	15
		5	79	32	38
		10	178	62	71
.01	.05	2	36	12	17
		5	102	32	38
		10	224	62	71
.05	.01	2	37	21	21
		5	108	48	54
		10	239	94	103
.01	.01	2	48	21	21
		5	136	48	54
		10	293	94	103

* All values are multiplied by Δ^2.

4. A sequential procedure for the normal distribution when θ_0 is unknown. We now consider the situation when θ_0 is unknown and the random variables $\{X_i\}$ are normally distributed with means $\{\theta_i\}$ and a common variance σ^2. Although our primary concern is with the case when σ^2 is known, some remarks will be made at the end of this section for the case σ^2 is unknown.

A sequential procedure will be given which is similar to that of Section 2. Let $\lambda' = \min\{1, [\beta/\alpha(k-1)]\}$, let $a' = \log[k/\alpha\lambda']$ and $b' = \log[\beta - (k-1)\lambda'\alpha/k]$. Let

$$Z_{j\alpha} = \log[\exp\{-(4\sigma^2)^{-1}(X_{j\alpha} - X_{0\alpha} - \Delta)^2\}/\exp\{-(4\sigma^2)^{-1}(X_{j\alpha} - X_{0\alpha})^2\}]$$
$$= \Delta(X_{j\alpha} - X_{0\alpha} - \Delta/2)/2\sigma^2.$$

A category π_j is rejected if at any stage $\sum_\alpha Z_{j\alpha} \leqq b'$. The sampling procedure used is the following: at the first stage, one measurement is taken with the standard (or control) category and one with each of the k experimental categories while at the rth stage one measurement is taken with the standard (or control) and one measurement with each experimental category that has not been eliminated during the preceding $r-1$ stages. For each r let $M' = M'(r)$ be the integer for which $\sum_{\alpha=1}^{r} Z_{M'\alpha} = \max_j (\sum_{\alpha=1}^{r} Z_{j\alpha})$ for all j for which π_j has not been eliminated in the preceding $r-1$ stages. The sequential procedure now follows: at each stage r, $r = 1, 2, \cdots$, if $\sum_{\alpha=1}^{r} Z_{M'\alpha} \geqq a'$, select $D_{M'}$, if $\sum_{\alpha=1}^{r} Z_{M'\alpha} \leqq b'$, select D_0, while if $b' < \sum_{\alpha=1}^{r} Z_{M'\alpha} < a'$, go on to

the $(r + 1)$st stage of experimentation and continue this procedure until a decision is reached. Exactly as in Section 2, it follows that $P(D_0 \mid R_0) \geqq 1 - \alpha$ and $P(D_j \mid R_j) \geqq 1 - \beta$. Although we are now not able to find any upper bound for $E(T \mid R_j)$ which is not too crude to be useful, it seems likely that the efficiency of the above sequential procedure when compared with the corresponding nonsequential procedure is close to the efficiency of (1) in relation to (4).

We conclude with a few remarks about the situation where σ^2 is unknown. We will suppose that an estimate s^2 of σ^2 is available, where fs^2/σ^2 has the χ^2 distribution with f degrees of freedom. This estimate might result from past experience, or might be based on a preliminary sample. We now note that the sequential procedure just given can still be used if σ^2 is replaced by s^2, a' is replaced by

$$\bar{a} = \{[\lambda'\alpha/k]^{-2/f} - 1\}f/2$$

and b' is replaced by

$$\bar{b} = -\{[\beta - (k - 1)\lambda'\alpha/k]^{-2/f} - 1\}f/2,$$

where $\lambda' = \min\{1, [\beta/\alpha(k - 1)]\}$. If this is done, then for the resulting sequential procedure

$$1 - P(D_0 \mid R_0) \leqq E[k \exp(-\bar{a}s^2/\sigma^2)] = k[1 + 2\bar{a}/f]^{-f/2} = \lambda'\alpha \leqq \alpha,$$

and in the same way $1 - P(D_j \mid R_j) \leqq \beta$.

REFERENCES

[1] PAULSON, E. (1952). On the comparison of several experimental categories with a control. *Ann. Math. Statist.* **23** 239–246.

[2] KARLIN, S. and TRUAX, D. (1960). Slippage problems. *Ann. Math. Statist.* **31** 296–324.

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

[4] PAULSON, E. (1952) An optimum solution to the k-sample slippage problem for the normal distribution. *Ann. Math. Statist.* **23** 610–616.