AN ASYMPTOTICALLY OPTIMAL FIXED SAMPLE SIZE PROCEDURE FOR COMPARING SEVERAL EXPERIMENTAL CATEGORIES WITH A CONTROL

By Charles DeWitt Rcberts

National Institute of Arthritis and Metabolic Diseases

Summary. The basic problem considered here involves k experimental categories. The experimenter must decide none of the k categories is better than the control or decide a certain category is better. For this problem a fixed sample size procedure δ_m^* is given. With a definite loss function and a cost $c > 0$ per observation, δ_m^* and other fixed sample size procedures are compared in a certain asymptotic sense as $c \to 0$. In particular, δ_m^* is shown to be an optimal fixed sample size procedure in this asymptotic sense. By appealing to asymptotic results the procedure δ_m^* is compared with sequentially designed procedures.

1. Introduction and statement of results. Let $X^{(j)}$ be the random variable resulting from an observation on the jth category, $j = 1, 2, \cdots, k$. We denote the probability density of $X^{(j)}$ by $g(x, \tau_j)$. For simplicity it is supposed here that the larger the value of τ, the more desirable the category is. We say $\theta = 0$ when $\tau_1 = \tau_2 = \cdots = \tau_k = \tau_0$ and say $\theta = j$ when $\tau_1 = \cdots = \tau_{j-1} = \tau_{j+1} = \cdots = \tau_k = \tau_0$ and $\tau_j = \tau_0 + \Delta$ where $\Delta > 0$, as described in the following table [where $g_0(x) = g(x, \tau_0)$ and $g_1(x) = g(x, \tau_0 + \Delta)$]:

(1.1)

θ	$X^{(1)}$	$X^{(2)}$	$X^{(3)}$	$\cdots$	$X^{(k)}$
0	g_0	g_0	g_0		g_0
1	g_1	g_0	g_0		g_0
2	g_0	g_1	g_0		g_0
$\vdots$	$\vdots$				
k	g_0	g_0	g_0	$\cdots$	g_1

The decision D_0 is preferred if $\theta = 0$ or if none of the experimental categories is better than the control [that is, $\tau_s \leqq \tau_0$ for $s = 1, 2, \cdots, k$] in the model (1.1). The decision D_j is preferred if $\theta = j$ or if the jth experimental category has the maximum value of τ and is better than the standard [that is, $\tau_j = \max_{s(1 \leqq s \leqq k)}(\tau_s) > \tau_0$] in the model (1.1). This formulation is that of Paulson [5] and Roberts [6].

The fixed sample size procedure δ_m^* is described as follows: Let $X_i^{(j)}$ be the ith observation on $X^{(j)} = \log [g_1(X_i^{(j)})/g_0(X_i^{(j)})]$ for $j = 1, 2, \cdots, k$. Define W after n_j observations on $X^{(j)}$ to be the integer for which

$$\sum_{i=1}^{n_W} Z_i^{(W)} = \max_j \left\{ \sum_{i=1}^{n_j} Z_i^{(j)} \right\}.$$

[If W is not unique because $\max_j\{\sum_{i=1}^{n_j} Z_i^{(j)}\}$ is assumed for more than one

Received 27 January 1964; revised 4 April 1964.

category, select W by a random choice of those j for which the maximum is attained.] The procedure δ_m^* takes $n = n_1 = n_2 = \cdots n_k$ observations and makes terminal decision $\theta = W$ if $\sum_{i=1}^n Z_i^{(W)} > m$ and terminal decision $\theta = 0$ otherwise.

Now assign a cost of $c > 0$ per observation, a loss which equals 0 when a correct terminal decision is made and 1 when an incorrect decision is made, and a prior distribution that assigns probability $\xi_j > 0$ to $\theta = j$ with $\xi_0 + \xi_1 + \xi_2 + \cdots + \xi_k = 1$. For θ the state of nature [θ is one of $0, 1, 2, \cdots, k$], δ a procedure, and N the total sample size required let $L(\theta, \delta)$ equal the expected loss with procedure δ, $E_\theta N$ equal the expected sample size required, and $r(\theta, \delta) = L(\theta, \delta) + cE_\theta N$ be the risk of procedure δ when θ is the state of nature. Define $r(\delta)$, the expected risk with procedure δ by $r(\delta) = \sum_{j=0}^k \xi_j r(j, \delta)$. Define $\rho(\delta)$, the price of procedure δ, by $\rho(\delta) = \lim \sup_{c \to 0}[-r(\delta)/c \log c]$. Finally define $A_0 = \inf_t E_0 \exp(tZ_1^{(1)})$ and $B_0 = \inf_t E_1 \exp(-tZ_1^{(1)})$. Then A_0 and B_0 are both finite and positive.

The price of a procedure is a type of measure of its desirability where the more desirable procedures have smaller prices. It is shown in Theorem 1 that there is a certain minimal price possible for fixed sample size procedures. We state now

THEOREM 1. *Any fixed sample size procedure δ (whose sample sizes may depend on the cost c) has* $\rho(\delta) \geqq -k/\log\{\max(A_0, B_0)\}$.

With a certain choice of the sample sizes it is shown in Theorem 2 that δ_m^* has the minimal (fixed sample size) price and hence we would say that δ_m^* is asymptotically the best fixed sample size procedure possible. More precisely we have

THEOREM 2. *If* $n_1 = n_2 = \cdots = n_k = \log c/\log\{\max(A_0, B_0)\}$ *then for each fixed* m, $\rho(\delta_m^*) = -k/\log\{\max(A_0, B_0)\}$.

It is of interest to compare the procedure δ_m^* with procedures which have a sequential design. Suppose $I_0 = -E_0 Z_1^{(1)}$ and $I_1 = E_1 Z_1^{(1)}$ exist (finite) and are positive. Roberts [6] gives three different sequential procedures which have prices bounded above by $k\xi_0/I_1 + (1 - \xi_0)[1/I_1 + (k - 1)/I_0]$. It will be shown in Section 3 (Theorem 3) that $-\log\{\max(A_0, B_0)\} < \min(I_0, I_1)$. This shows that $\delta_1, \delta_2, \delta_3$ are each strictly better than the optimal fixed sample size procedure δ_m^*.

This problem also has been discussed in the non-sequential case by Paulson [4] and Karlin and Truax [3]. Paulson [5] considers a sequential procedure of the problem.

2. Applications. In many practical situations the consequence of making a wrong decision cannot be evaluated in economic terms. In such situations it would seem that a reasonable approach is to ask for a solution which has some desirable properties with regard to the probabilities $L(\theta, \delta)$ without explicitly using c, the cost of an observation. A conventional formulation of an optimum procedure when $n_1, n_2, \cdots, n_k$ are fixed is to ask fo a procedure δ which will minimize $[\max_{1 \leqq j \leqq k} L(j, \delta)]$ subject to the restriction that $L(0, \delta) = \alpha$.

Using the techniques of [3] or [4] with $n = n_1 = n_2 = \cdots = n_k$, it can be shown that the optimum procedure δ is to make the decision $\theta = W$ if $\sum_{i=1}^{n} Z_i^{(W)} > m$ and to make the decision $\theta = 0$ if $\sum_{i=1}^{n} Z_i^{(W)} \leqq m$, where m is determined by the requirement $P_0[\sum_{i=1}^{n} Z_i^{(W)} > m] = \alpha$ and P_θ indicates probability with state of nature θ. When we are at liberty to choose n, a reasonable procedure would be to select n as the smallest integer so that $L(j, \delta) \leqq \beta$ for $j = 1, 2, \cdots, k$ in addition to the requirement $L(0, \delta) = \alpha$.

If m and n are chosen to satisfy

$$kP_0 \left\{ \sum_{i=1}^{n} Z_i^{(1)} > m \right\} \leqq \alpha \tag{2.1}$$

and

$$P_1 \left\{ \sum_{i=1}^{n} Z_i^{(1)} \leqq m \right\} + (k-1) P_0 \left\{ \sum_{i=1}^{n} Z_i^{(1)} > m \right\} \leqq \beta, \tag{2.2}$$

then $L(0, \delta) \leqq \alpha$ and $L(j, \delta) \leqq \beta$ for $j = 1, 2, \cdots, k$. That is, the probability of selecting D_0 when $\theta = 0$ is at least $1 - \alpha$ and the probability of selecting D_j when $\theta = j$ is at least $1 - \beta$ for each $j, j = 1, 2, \cdots, k$.

In the case that (2.1) and (2.2) are satisfied a truncated sequential design procedure, which preserves error levels α, β and may save observations, can be used. The procedure is as follows: Take one observation on each of $X^{(1)}$, $X^{(2)}$, $\cdots$, $X^{(k)}$. Then the rule is to take one observation on category W. This procedure of one observation at a time is continued until there are n observations taken on one category $X^{(j)}$ (say). Then if $\sum_{i=1}^{n} Z_i^{(j)} > m$ stop and make decision $\theta = j$. If $\sum_{i=1}^{n} Z_i^{(j)} \leqq m$, continue the sampling only with $X^{(j)}$ deleted from further observation. We obey the following two rules:

(1) Stop sampling and make decision $\theta = j$ as soon as $\sum_{i=1}^{n} Z_i^{(j)} > m$ for some j.

(2) Stop sampling and make decision $\theta = 0$ if $\sum_{i=1}^{n} Z_i^{(j)} \leqq m$ for $j = 1, 2, \cdots, k$.

This truncated sequential design will not be an optimum one in the class of all closed sequential designs with the same α and β. In fact, it may be much less efficient than some other closed sequential designs in this class.

3. Proofs.

LEMMA 1 (Bounds of the sample mean). *Let $Y_1, Y_2, \cdots$ be independent and identically distributed random variables. Define for b fixed*

(1) $p_n = P\{(Y_1 + Y_2 + \cdots + Y_n)/n \geqq b\}$;

(2) $\varphi(t) = E \exp(tY_1)$ for all real t;

(3) $\psi(t) = e^{-bt}\varphi(t)$ *for all real* t;

(4) $T = \{t: -\infty < t < \infty, \varphi(t) < \infty\}$.

If

(a) $P(Y_1 = b) \neq 1$,

(b) *T is a non-degenerate interval,*

(c) *there exists a positive τ in the interior of T such that* $\psi(\tau) = \inf_{t \in T} \psi(t) = \rho$

(say), then $p_n \leqq \rho^n$ *and for every real number* ϵ *such that* $0 < \epsilon < \psi(\tau)$ *for* n *sufficiently large* $p_n \geqq (\rho - \epsilon)^n$.

PROOF. This is essentially due to Chernoff [2] but is stated and proved in this form by Bahadur and Rao [1].

Now define $A(t) = E_0 \exp(tZ_1^{(1)})$ and $B(t) = E_1 \exp(-tZ_1^{(1)})$. Let $R(j, s) = \prod_{i=1}^{s}[g_1(X_i^{(j)})/g_0(X_i^{(j)})]$ for $j = 1, 2, \cdots, k$ and $R(0, n_0) = 1$. Denote $\mathbf{n} = (n_1, n_2, \cdots, n_k)$ and let $\delta^0 = \delta^0(\mathbf{n})$ denote a procedure with Bayes terminal decision rule based on n_j observations on $X^{(j)}$.

LEMMA 2. *Given* $X_1^{(j)}, X_2^{(j)}, \cdots, X_{n_j}^{(j)}$ *for* $j = 1, 2, \cdots, k$ *mutually independent then a Bayes terminal decision rule makes decision* $\theta = s$ *if* $\xi_s R(s, n_s) > \xi_t R(t, n_t)$ *for* $s \neq t$ *and* $t = 0, 1, 2, \cdots, k$.

PROOF. Write $\xi_j(\mathbf{n}) = \xi_j R(j, n_j)/\{\sum_{i=0}^{k} \xi_i R(i, n_i)\}$ and $L(j, \xi(\mathbf{n})) = \sum_{i=0, i\neq j}^{k} \xi_i(\mathbf{n})$ for $j = 0, 1, \cdots, k$. Since the Bayes terminal decision rule makes decision j when $L(j, \xi(\mathbf{n})) < L(i, \xi(\mathbf{n}))$ for $i \neq j$, $i = 0, 1, 2, \cdots, k$, the proof now follows.

LEMMA 3. *The functions* $A(t)$ *and* $B(t)$ *are convex.*

LEMMA 4. *For* $0 \leqq t \leqq 1$, $A(t) < \infty$ *and* $B(t) < \infty$ *so that the intervals of (finite) convergence of* $A(t)$ *and* $B(t)$ *are each non-degenerate.*

LEMMA 5. (1) $0 < A_0 < 1$, $0 < B_0 < 1$ *and* (2) *there exists* a, $0 < a < 1$, *and* b, $0 < b < 1$, *such that* $A_0 = A(a)$, $B_0 = B(b)$.

Let P_i denote probability associated with $\theta = i$.

LEMMA 6.

(1) $P_s(R(j, n_j) \geqq e^m) \leqq e^{-am} A_0^{n_j}$ if $j \neq s$ *and* $P_j(R(j, n_j) \leqq e^m) \leqq e^{bm} B_0^{n_j}$ *for* $j = 1, 2, \cdots, k$.

(2) *If* $0 < \epsilon < \min(A_0, B_0)$, *then for* n_j *sufficiently large*

$$\text{(a) } P_0(R(j, n_j) > \xi_0/\xi_j) \geqq (A_0 - \epsilon)^{n_j}$$

and

$$\text{(b) } P_j(R(j, n_j) < \xi_0/\xi_j) \geqq (B_0 - \epsilon)^{n_j}.$$

PROOF. By application of Lemma 1 the proof of (1) follows. We have $P_0(R(j, n_j) > \xi_0/\xi_j) = P_0(\sum_{i=1}^{n_j} Z_i^{(j)} > \log(\xi_0/\xi_j)) \geqq P_0(\sum_{i=1}^{n_j} Z_i^{(j)} \geqq vn_j)$ for any $v > 0$ and n_j sufficiently large. We can choose v so close to 0 that $A^* = \inf_t e^{-vt} A(t) \geqq A_0 - \epsilon/2$ and then $A^* - \epsilon/2 \geqq A_0 - \epsilon$. By applying Lemma 1 again (2) (a) follows. Similarly (2) (b) follows.

LEMMA 7.

(1) *For all* $n_1, n_2, \cdots, n_k$, $L(0, \delta_m^*(\mathbf{n})) \leqq \sum_{i=1}^{k} P_0(R(i, n_i) \geqq e^m)$ *and* $L(j, \delta_m^*(\mathbf{n})) \leqq P_j(R(j, n_j) \leqq e^m) + \sum_{i=1, i\neq j}^{k} P_0(R(i, n_i) \geqq e^m)$.

(2) *If* $n_1, n_2, \cdots, n_k$ *are sufficiently large* $L(0, \delta^0(\mathbf{n})) \geqq \frac{1}{2}\sum_{i=1}^{k} P_0(R(i, n_i) > \xi_0/\xi_i)$ *and* $L(j, \delta^0(\mathbf{n})) \geqq \frac{1}{2} P_j(R(j, n_j) < \xi_0/\xi_j)$.

PROOF. The proof of (1) follows by definition. We have $L(0, \delta^0(\mathbf{n})) \geqq \sum_{i=1}^{k}\{P_0(R(i, n_i) > \xi_0/\xi_i) \prod_{s=1, s\neq i}^{k} P_0(R(s, n_s) < \xi_0/\xi_s)\}$ and so for $n_1, n_2, \cdots, n_k$ sufficiently large $L(0, \delta^0(\mathbf{n})) \geqq \frac{1}{2}\sum_{i=1}^{k} P_0(R(i, n_i) > \xi_0/\xi_i)$. Also since $L(j, \delta^0(\mathbf{n})) \geqq \prod_{i=1}^{k} P_j(R(i, n_i) < \xi_0/\xi_i)$ then for $n_1, n_2, \cdots, n_k$ sufficiently large $L(j, \delta^0(\mathbf{n})) \geqq \frac{1}{2} P_j(R(j, n_j) < \xi_0/\xi_j)$ which completes the proof of Lemma 7.

PROOF OF THEOREM 1. If $\rho(\delta^0) < \infty$, it follows for some $M > 0$ and c sufficiently small that $P_0(R(j, n_j) > \xi_0/\xi_j) \leqq -Mc \log c$ and $P_j(R(j, n_j) < \xi_0/\xi_j) \leqq -Mc \log c$ for $j = 1, 2, \cdots, k$. By Lemma 6 (2) it follows that $n_j \log (A_0 - \epsilon) \leqq \log c + \log(-M \log c)$ and $n_j \log(B_0 - \epsilon) \leqq \log c + \log(-M \log c)$. Thus for each ϵ, $0 < \epsilon < \min(A_0, B_0)$, $n_j \geqq \log c/\log(A_0 - \epsilon) + o(\log c)$ and $n_j \geqq \log c/\log(B_0 - \epsilon) + o(\log c)$ so that $n_j \geqq \log c/\max\{\log A_0, \log B_0\} + o(\log c)$. Therefore, $r(\delta^0(\mathbf{n})) \geqq [1 + o(1)]kc \log c/\max\{\log A_0, \log B_0\}$ which completes the proof.

PROOF OF THEOREM 2. By Lemmas 6 (1) and 7 (1) $r(\delta_m^*(\mathbf{n})) \leqq \xi_0 k e^{-am} A_0^{n_1} + (1 - \xi_0)[e^{bm} B_0^{n_1} + (k - 1)e^{-am} A_0^{n_1}] + ckn_1$. Now $A_0^{n_1} \leqq A_0^{\log c/\log A_0} = c = o(c \log c)$ and $B_0^{n_1} \leqq B_0^{\log c/\log B_0} = c = o(c \log c)$ so that $r(\delta_m^*(\mathbf{n})) \leqq [1 + o(1)]kc \log c/\log\{\max(A_0, B_0)\}$ which completes the proof.

THEOREM 3. *If* $I_0 = -E_0 Z_1^{(1)}$ *and* $I_1 = E_1 Z_1^{(1)}$ *exist (finite) then*

(1) $A_0 > e^{-I_0}$, $B_0 > e^{-I_1}$ *and*

(2) $-\log\{\max(A_0, B_0)\} < \min(I_0, I_1)$.

PROOF. Now $E_0 \exp(tZ_1^{(1)}) \geqq \exp(tZ_1^{(1)}) = e^{-tI_0}$ for each t with strict inequality holding for $t \neq 0$. We have $A_0 = E_0 \exp(aZ_1^{(1)})$ for $0 < a < 1$ by Lemma 5. Thus $A_0 > e^{-I_0}$. Similarly $B_0 > e^{-I_1}$ which proves (1). By applying (1) the proof of (2) follows.

REFERENCES

[1] BAHADUR, R. R. and RAO, R. R. (1960). On deviations of the sample mean. *Ann. Math. Statist.* **31** 1015–1027.

[2] 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.

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

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

[5] PAULSON, E. (1962). A sequential procedure for comparing several experimental categories with a standard or control. *Ann. Math. Statist.* **33** 438–443.

[6] ROBERTS, C. D. (1963). An asymptotically optimal sequential design for comparing several experimental categories with a control. *Ann. Math. Statist.* **34** 1486–1493.