AN ASYMPTOTICALLY OPTIMAL SEQUENTIAL DESIGN FOR COMPARING SEVERAL EXPERIMENTAL CATEGORIES WITH A CONTROL

By Charles DeWitt Roberts[1]

University of North Carolina

Summary. The basic problem is to decide none of k experimental categories is better than the control or decide a certain category is better. For this problem three sequential procedures are examined with specification of how procedures are carried out in practice. With a definite loss function and a cost $c > 0$ per observation the three sequential procedures and fixed sample size procedures are compared in a certain asymptotic sense as $c \to 0$. In particular, one of the procedures is shown to be optimal in this asymptotic sense. By appealing to asymptotic results a discussion of the relative merits of the three sequential procedures as considered in practice is given.

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	$\cdots$	g_0
1	g_1	g_0	g_0	$\cdots$	g_0
2	g_0	g_1	g_0	$\cdots$	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 is better than the control [that is, $\tau_j > \tau_0$] in the model (1.1). This formulation is that of Paulson [4].

The three sequential procedures to be considered are denoted by δ_1, δ_2, δ_3

Received May 27, 1963.

[1] Now at the National Institute of Arthritis and Metabolic Diseases.

and are described as follows: Let $b < 0 < a$, $X_i^{(j)}$ be the ith observation on $X^{(j)}$ and

$$Z_i^{(j)} = \log [g_1(X_i^{(j)})/g_0(X_i^{(j)})] \quad \text{for} \quad j = 1, 2, \cdots, k.$$

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 category, select W by a random choice of those j for which the maximum is attained.]

Procedure δ_1. Take one observation on each of $X^{(1)}, X^{(2)}, \cdots, X^{(k)}$. Then select after each single observation a category on which to sample next. The selection rule is to take one observation next on $X^{(W)}$.

Procedure δ_2. Select at random an order to examining the categories, and then one-by-one decide if a category is better than the control. If an order of $(i_1, i_2, \cdots, i_k)$ is chosen, sample first on $X^{(i_1)}$, then on $X^{(i_2)}, \cdots$, then on $X^{(i_k)}$ so that once sampling is begun on $X^{(i_j+1)}$ no more observations are taken on $X^{(i_j)}$.

Procedure δ_3. Sample in k (or less)—tuples of one observation on each category beginning with a k-tuple. After each observation decide a category is better than the control and stop further sampling or continue sampling after (possibly) eliminating categories that appear no better than the control. This is the procedure suggested by Paulson [4].

Finally for δ_1, δ_2, and δ_3 the following three rules apply:

(i) Stop sampling on $X^{(j)}$ as soon as $b < \sum_{i=1}^{n_j} Z_i^{(j)} < a$ is violated for some n_j.

(ii) If for some n_j, $\sum_{i=1}^{n_j} Z_i^{(j)} \geqq a$ stop further sampling and make decision D_W.

(iii) As soon as $\sum_{i=1}^{n_W} Z_i^{(W)} \leqq b$ and observations have been taken on all k categories, stop further sampling and make decision D_0.

In Section 2 it will be shown for the three procedures how to choose one of the $k + 1$ decisions $(D_0, D_1, \cdots, D_k)$ so that 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, $j = 1, 2, \cdots, k$.

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 $p(\delta)$, the price of procedure δ, by $p(\delta) = \lim\sup_{c\to 0} [-r(\delta)/c \log c]$. Finally it is supposed that $I_0 = -E_0Z_1^{(1)}$ and $I_1 = E_1Z_1^{(1)}$ exist (finite) and are 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 procedures which operate in a measurable fashion. Also Theorem 1 gives lower bounds for the prices of δ_2 and δ_3 which show that both δ_2 and δ_3 could not achieve the minimal value and hence could not be asymptotically optimal. We state now

THEOREM 1.

(i) *Any procedure* δ *has* $p(\delta) \geqq k\xi_0/I_0 + (1 - \xi_0)/I_1$. *With any choice of* a, b [*possibly depending on the cost* c],

(ii) $p(\delta_2) \geqq k\xi_0/I_0 + (1 - \xi_0)[1/I_1 + (k - 1)/2I_0]$ *and*

(iii) $p(\delta_3) \geqq k\xi_0/I_0 + (1 - \xi_0)[1/I_1 + (k - 1)/(I_0 + I_1)]$.

With the choice of $a = -b = -\log c$ it is shown in Theorem 2 that procedure δ_1 has the minimal price and hence we would say that δ_1 is asymptotically optimal. With the same choice of a and b it is shown that δ_2 attains the lower bound for procedures of the form of δ_2. Also with the choice of $a = -b = -\log c$ an upper bound is given for the price of δ_3. More precisely we have

THEOREM 2. *With* $a = -b = -\log c$

(i) $p(\delta_1) = k\xi_0/I_0 + (1 - \xi_0)/I_1$,

(ii) $p(\delta_2) = k\xi_0/I_0 + (1 - \xi_0)[1/I_1 + (k - 1)/2I_0]$,

(iii) $p(\delta_3) \leqq k\xi_0/I_0 + (1 - \xi_0)[1/I_1 + (k - 1)/\max(I_0, I_1)]$.

An experimenter may feel that the asymptotically optimal procedure δ_1 is troublesome or undesirable in practice. Thus one may prefer to perform procedure δ_2 or δ_3. By making use of Theorems 1 and 2 we see that procedure δ_2 is better than δ_3 if $I_1 < I_0$ and procedure δ_3 is better than δ_2 if $2I_0 < I_1$. From Theorem 2 we see that if I_1/I_0 or I_0/I_1 is small then δ_3 is approximately optimal, and if I_1/I_0 is small then δ_2 is approximately optimal.

It is of interest to compare the three sequential procedures with fixed sample size procedures. It is shown by Theorem 3 in view of Theorem 2 that δ_1, δ_2, and δ_3 are each strictly better than any fixed sample size procedure. Although the proof will not be given here, the following theorem is shown by Roberts [5].

THEOREM 3. *Any fixed sample size procedure* δ (*whose sample sizes may depend on the cost* c) *has* $p(\delta) > k/\min(I_0, I_1)$.

For discussion about the sequential design of experiments and general asymptotic results see Chernoff [1]. In particular, asymptotic optimality is in the same sense. More recent results are obtained by Kiefer and Sacks [3].

For the problem considered here the method of Chernoff [1] suggests a procedure which is similar to δ_1 [when $a = b = -\log c$] and will be asymptotically optimal as $c \to 0$. However this procedure is dependent upon the cost c and in practice δ_1 probably would be more desirable to an experimenter. Kiefer and Sacks [3] consider a type of two-stage design procedure which is asymptotically optimal as $c \to 0$. This design allows that during the second stage the choice of the next category on which to take the next observation does not depend upon the second stage observations. In both cases modification is involved because the model (1.1) leads to games which have payoff matrices to which the problems of Chernoff, Kiefer, and Sacks do not directly apply.

2. Applications. If α, β [error probability levels] are specified and g_0, g_1 are assumed known, denote

$$I_0 = E_0 \{\log [g_0(X_1^{(1)})/g_1(X_1^{(1)})]\}, \qquad I_1 = E_1 \{\log [g_1(X_1^{(1)})/g_0(X_1^{(1)})]\}$$

and let

$$\lambda = \min \{1, k\beta I_0/\alpha(k-1)[I_0 + (k-1)I_1]\}. \tag{2.1}$$

For any of the three procedures, we suggest choosing

$$a = \log (k/\lambda\alpha) \quad \text{and} \quad b = \log [\beta - (k-1)\alpha\lambda/k] \tag{2.2}$$

which is what is suggested by Paulson [4] for procedure δ_3. Let P_θ indicate probability when θ is the state of nature. For the three procedures we will have

$$1 - P_0(D_0) = \sum_{j=1}^{k} P_0(D_j) \leqq \sum_{j=1}^{k} P_0\left[\sum_{i=1}^{r} Z_i^{(j)} \geqq a \quad \text{for some } r < \infty\right]$$

and

$$1 - P_j(D_j) \leqq P_j\left[\sum_{i=1}^{r} Z_i^{(j)} \leqq b \quad \text{for some } r < \infty\right]$$
$$+ \sum_{s=1, s\neq j}^{k} P_j\left[\sum_{i=1}^{r} Z_i^{(s)} \geqq a \quad \text{for some } r < \infty\right].$$

It is well-known [see Wald [6]] that $P_0[\sum_{i=1}^{r} Z_i^{(s)} \geqq a$ for some $r < \infty] \leqq e^{-a}$ and $P_s[\sum_{i=1}^{r} Z_i^{(s)} \leqq b$ for some $r < \infty] \leqq e^{b}$ for $s = 1, 2, \cdots, k$. Thus in order to satisfy the requirement that $P_0(D_0) \geqq 1 - \alpha$ and $P_j(D_j) \geqq 1 - \beta$, we therefore determine a and b so that $ke^{-a} \leqq \alpha$ and $e^{b} + (k-1)e^{-a} \leqq \beta$. For the three procedures $a/I_1 - (k-1)b/I_0$ is an approximate upper bound for the expected sample size if θ is one of $1, 2, \cdots, k$. If we minimize this approximate upper bound [that is, $a/I_1 - (k-1)b/I_0$] with respect to a and b subject to $ke^{-a} \leqq \alpha$ and $e^{b} + (k-1)e^{-a} \leqq \beta$ we have the values for a and b given by (2.1) and (2.2).

Example. Suppose g_0 and g_1 are probability density functions of normal distributions with means μ and $\mu + \Delta$, variances 1 and σ^2, respectively. Then

$$Z_1^{(1)} = [\sigma^2(X_1^{(1)} - \mu)^2 - (X_1^{(1)} - \mu - \Delta)^2 - 2\sigma^2 \log \sigma]/2\sigma^2,$$
$$I_0 = (\Delta^2 + 1 - \sigma^2 + 2\sigma^2 \log \sigma)/2\sigma^2,$$
$$I_1 = (\Delta^2 - 1 + \sigma^2 - 2 \log \sigma)/2$$

so that δ_2 is suggested over δ_3 if $\sigma^2 < 1$ and δ_3 is suggested over δ_2 if $\sigma^2 > y(\Delta)$ where $y(\Delta)$ is the greatest value of y for which $y^2 + y(1 + \Delta^2) - 3y \log y - 2(1 + \Delta^2) = 0$. In particular, if $\mu = 0$ and $\Delta = 1$ then δ_2 is suggested over δ_3 if $\sigma^2 < 1$ and δ_3 is suggested over δ_2 if $\sigma^2 > 2.17$.

3. Proofs.

LEMMA 1. *For any random variable* Y

(i) $EY \leqq \log Ee^Y$ *when* EY *exists*,
(ii) $P(Y \geqq 0) \leqq Ee^{hY}$ if $h \geqq 0$.

LEMMA 2. (Wald's Equation). *Suppose that*

(i) $Y_1, Y_2, \cdots$ *are identically distributed random variables*,
(ii) N *is a random variable whose values are the positive integers*,
(iii) *the event* $\{N = j\}$ *and the random variable* Y_k *are independent for* $j < k$,
(iv) $E|Y| < \infty$ *and* $EN < \infty$ *then* $E(\sum_{j=1}^{N} Y_j) = (EN)(EY_1)$.

PROOF. See Johnson [2].

Let N_j denote the number of observations taken on $X^{(j)}$. Let

$$R(j, s) = \prod_{i=1}^{s} [g_1(X_i^{(j)})/g_0(X_i^{(j)})] \quad \text{for} \quad j = 1, 2, \cdots, k$$

and denote $R(0, N_0) = 1$. The notation P_θ indicates probability when θ is the state of nature.

LEMMA 3. *If* S *is an event such that* $P_0(S) > 0$ *and the procedure terminates with probability* 1 *for* $\theta = 0, 1, 2, \cdots, k$ *then* $P_t(S) = P_s(S)E_s[R(t, N_t)/R(s, N_s) \mid S]$ *for* $s, t = 0, 1, 2, \cdots, k$.

PROOF. Our sample space consists of points $\mathbf{X} = (X_1^{(1)}, X_2^{(1)}, \cdots; X_1^{(2)}, X_2^{(2)}, \cdots; \cdots; X_1^{(k)}, X_2^{(k)}, \cdots)$. We have $P_t(S) > 0$ for $t = 1, 2, \cdots, k$. Suppose $s = 1, t = 2$ and define

$$\begin{aligned} S_{ij} &= \{\mathbf{X} : N_1 = i \quad \text{and} \quad N_2 = j\} \\ \psi_{ij} &= i \text{ if } \mathbf{X} \in S_{ij} \\ &= 0 \text{ otherwise} \\ \phi(j, k) &= 1 \text{ if } \mathbf{X} \in S \cap S_{jk} \\ &= 0 \text{ otherwise.} \end{aligned}$$

It follows that $\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \psi_{ij} = 1$ with probability 1 for each θ. Now

$$\begin{aligned} P_2(S) &= \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} P_2(S \cap S_{ij}) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} E_2\psi_{ij}\phi(i, j) \\ &= \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} E_1\{\psi_{ij}\phi(i, j)R(2, j)/R(1, i)\} \\ &= \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} E_1\{\psi_{ij}\phi(N_1, N_2)R(2, N_2)/R(1, N_1)\} \\ &= E_1\{\phi(N_1, N_2)R(2, N_2)/R(1, N_1)\}. \end{aligned}$$

Also $P_1(S) = E_1\phi(N_1, N_2)$ so we have $P_2(S) = P_1(S)E_1\{R(2, N_2)/R(1, N_1) \mid S\}$ which completes the proof when $s = 1$, $t = 2$. The proofs in the other cases are similar.

Let M_j be the least n_j such that $b < \sum_{i=1}^{n_j} Z_i^{(j)} < a$ is violated.

PROOF OF THEOREM 1. If $p(\delta) < \infty$ then by the definition of $p(\delta)$ it follows that $P_i(D_j) \leqq -Mc \log c$ if $i \neq j$ for some $M > 0$ and c sufficiently small. By Lemmas 1, 2, 3 we have

$$\begin{aligned}-I_0E_0N_j &= E_0 \sum_{i=1}^{N_j} Z_i^{(j)} \\ &= P_0(\text{not } D_0)E_0\left(\sum_{i=1}^{N_j} Z_i^{(j)} \mid \text{not } D_0\right) + P_0(D_0)E_0\left(\sum_{i=1}^{N_j} Z_i^{(j)} \mid D_0\right) \\ &\leqq P_0(\text{not } D_0) \log E_0(R(j, N_j) \mid \text{not } D_0) + P_0(D_0) \log E_0(R(j, N_j) \mid D_0) \\ &= P_0(\text{not } D_0) \log \{P_j(\text{not } D_0)/P_0(\text{not } D_0)\} + P_0(D_0) \log \{P_j(D_0)/P_0(D_0)\}.\end{aligned}$$

But

$$\lim\ \inf_{c\to 0} \frac{P_0(D_0) \log[P_j(D_0)/P_0(D_0)]}{\log c} \geqq \lim_{c\to 0} \frac{P_0(D_0) \log(-Mc \log c)}{\log c} = 1$$

and

$$\lim_{c\to 0} P_0(\text{not } D_0) \log [P_j(\text{not } D_0)/P_0(\text{not } D_0)] = 0.$$

Thus $E_0N_j \geqq -[1 + o(1)] \log c/I_0$ for $j = 1, 2, \cdots, k$. Similarly $E_jN_j \geqq -[1 + o(1)] \log c/I_1$ for $j = 1, 2, \cdots, k$. Therefore it follows that

$$r(\delta) \geqq c\sum_{j=0}^{k} \xi_jE_jN \geqq -[1 + o(1)]\{k\xi_0/I_0 + (1 - \xi_0)/I_1\}c \log c$$

which completes the proof of (i). By similar methods (ii) will follow. Let us now prove (iii). Let us show $E_1N_2 \geqq -[1 + o(1)]\{1/(I_0 + I_1)\} \log c$. Define

$$A = \left\{\sum_{i=1}^{M_3} Z_i^{(3)} \leqq b, \sum_{i=1}^{M_4} Z_i^{(4)} \leqq b, \cdots, \sum_{i=1}^{M_k} Z_i^{(k)} \leqq b\right\},$$

$$S_0 = \left\{M_2 < M_1, \sum_{i=1}^{M_2} Z_i^{(2)} \leqq b\right\} \cap A, \qquad S_1 = \left\{M_2 \geqq M_1, \sum_{i=1}^{M_1} Z_i^{(1)} \geqq a\right\} \cap A,$$

$S_2 =$ complement of $S_0 \cup S_1$.

Proceeding as in the proof of (i),

$$\begin{aligned}E_1\sum_{i=1}^{N_2} (-Z_i^{(1)} + Z_i^{(2)}) = -(I_1 + I_0)E_1N_2 &\leqq P_1(S_0) \log [P_2(S_0)/P_1(S_0)] \\ &+ P_1(S_1) \log [P_2(S_1)/P_1(S_1)] + P_1(S_2) \log [P_2(S_1)/P_1(S_2)].\end{aligned}$$

However $P_2(S_0) \leqq -2Mc \log c$, $P_2(S_1) \leqq -Mc \log c$ and $P_1(S_0) + P_1(S_1) \to 1$ $c \to 0$. Thus $-(I_1 + I_0)(E_1N_2)/\log c \geqq 1$ as $c \to 0$. Similarly for $j \neq s$, $E_jN_s \geqq$

$-[1 + o(1)]\{1/(I_0 + I_1)\} \log c$ for $j, s = 1, 2, \cdots, k$. Thus we have

$$r(\delta_3) \geqq -[1 + o(1)]\{k\xi_0/I_1 + (1 - \xi_0)[1/I_1 + (k - 1)/(I_0 + I_1)]\}c \log c$$

which completes the proof of Theorem 1.

For the remaining remarks let $a = -b = -\log c$.

Lemma 4. *For* $j = 1, 2, \cdots, k$

(i) $P_0\{R(j, M_j) \geqq 1/c\} \leqq c$

(ii) $P_j\{R(j, M_j) \leqq c\} \leqq c$.

Proof. Suppose $j = 1$. Let A_n be the set in the sample space on which we have $M_1 = n$ and $\sum_{i=1}^{M_1} Z_i^{(1)} \geqq -\log c$. Let B_n be the set on which $M_1 = n$ and $\sum_{i=1}^{M_1} Z_i^{(1)} \leqq \log c$. Then on A_n, $R(1, n) \geqq 1/c$ so that $\prod_{i=1}^{n} g_0(X_i^{(1)}) \leqq c\prod_{i=1}^{n} g_1(X_i^{(1)})$. Therefore $P_0(A_n) \leqq cP_1(A_n)$ so that

$$P_0\Big(\sum_{i=1}^{M_1} Z_i^{(1)} \geqq -\log c\Big) = \sum_{n=1}^{\infty} P_0(A_n) \leqq c\sum_{n=1}^{\infty} P_1(A_n)$$

$$= cP_1\Big(\sum_{i=1}^{M_1} Z_i^{(1)} \geqq -\log c\Big) \leqq c$$

which proves (i) when $j = 1$. The other cases are very similar.

Lemma 5. *For* $j = 1, 2, \cdots, k$

(i) $E_0M_j = -[1 + o(1)] \log c/I_0$

(ii) $E_jM_j = -[1 + o(1)] \log c/I_1$.

Proof. By the Theorem 1 argument we have $E_0M_j \geqq -[1 + o(1)] \log c/I_0$ and $E_jM_j \geqq -[1 + o(1)] \log c/I_1$. To show that also $E_jM_j \leqq -[1 + o(1)] \log c/I_0$ a type of argument which can be found in the proof of Lemma 2 of [1] will be used. If $\epsilon > 0$ and $n_j \geqq -(1 + \epsilon) \log c/I_1$ we have for $t \leqq 0$

$$P_j\left(\sum_{i=1}^{n_j} Z_i^{(j)} \leqq -\log c\right) \leqq P_j\left(\sum_{i=1}^{n_j} Z_i^{(j)} \leqq n_jI_1/(1 + \epsilon)\right)$$

$$\leqq [E_j \exp \{t[Z_1^{(j)} - I_1/(1 + \epsilon)]\}]^{n_j}.$$

But $Z_i^{(j)} - I_1/(1 + \epsilon)$ has positive mean and finite moment generating function for $-1 \leqq t \leqq 0$ and $\theta = j$. Hence the left-hand derivative of the moment generating function is positive at $t = 0$. Thus there is a $t_j^* = t_j^*(\epsilon)$ so that $E_j \exp \{t_j^*[Z_1^{(j)} - I_1/(1 + \epsilon)]\} \leqq d_j$ for some $d_j = d_j(\epsilon)$, $0 < d_j < 1$. Thus it follows that $P_j(\sum_{i=1}^{n_j} Z_i^{(j)} \leqq -\log c) \leqq d_j^{n_j}$ for $n_j \geqq -(1 + \epsilon) \log c/I_1$.

Therefore $E_jM_j = -[1 + o(1)] \log c/I_1$ which proves (ii). The proof of (i) is similar.

Proof of Theorem 2. For δ one of $\delta_1, \delta_2, \delta_3$ by Lemma 4, $L(i, \delta) \leqq kc = o(c \log c)$ for $i = 0, 1, 2, \cdots k$. Note in the proof of Theorem 1 that $E_jN_j \geqq -[1 + o(1)] \log c/I_1$ for $j \geqq 1$. For δ_1 let us show that $E_jN \leqq -[1 + o(1)] \log c/I_1$ for $j = 1, 2, \cdots, k$. We have $E_jN_j \leqq E_jM_j = -[1 + o(1)] \log c/I_1$. Let us show $E_1N_2 = o$ (log c). It is sufficient to show there is a τ, $0 < \tau < 1$, so that after n_1 observations on $X^{(1)}$ and n_2 on $X^{(2)}$ we have $P_1\{\sum_{i=1}^{n_1} Z_i^{(1)} \leqq \sum_{i=1}^{n_2} Z_i^{(2)}\} \leqq \tau^{n_2}$. Now we have for $t \leqq 0$

$$P_1\{\sum_{i=1}^{n_1} Z_i^{(1)} \leqq \sum_{i=1}^{n_2} Z_i^{(2)}\} \leqq \{E_1 \exp [tZ_1^{(1)}]\}^{n_1}\{E_1 \exp [-tZ_1^{(2)}]\}^{n_2}.$$

But $-Z_1^{(2)}$ has a positive mean and finite moment generating function for $-1 \leqq t \leqq 0$ when $\theta = 1$. Hence the left-hand derivative of the moment generating function is positive at $t = 0$. Thus there is a t_1^*, $-1 < t_1^* < 0$, so that $E_1 e^{-t_1^* Z_1^{(2)}} \leqq \tau$ for some τ, $0 < \tau < 1$. Now since $\psi_1(t) = E_1 e^{tZ_1^{(1)}}$ is convex and $\psi_1(0) = \psi_1(-1) = 1$ then $\psi_1(t_1^*) \leqq 1$. Thus $E_1 N_2 = o$ (log c). Similarly $E_j N_s = o$ (log c) for $s, j = 1, 2, \cdots, k$ and $j \neq s$. Therefore $r(\delta_1) \leqq -[1 + o(1)]\{k\xi_0/I_0 + (1 + \xi_0)/I_1\}c \log c$ which proves (i). By use of Lemmas 4 and 5, $r(\delta_2) = -[1 + o(1)]\{k\xi_0/I_0 + (1 - \xi_0)[1/I_1 + (k - 1)/2I_0]\}c \log c$ which proves (ii). Also by Lemmas 4 and 5, $r(\delta_3) \leqq -[1 + o(1)]\{k\xi_0/I_0 + (1 - \xi_0)[1/I_1 + (k - 1)/\max (I_0, I_1)]\}c \log c$ which completes the proof of Theorem 2.

4. Acknowledgment. The author is deeply indebted to Professor Wassily Hoeffding for suggesting the problem, and for his comments and encouragement.

REFERENCES

[1] CHERNOFF, H. (1959). Sequential design of experiments. *Ann. Math. Statist.* **30** 755–770.

[2] JOHNSON, N. L. (1959). A proof of Wald's theorem on cumulative sums. *Ann. Math. Statist.* **30** 1245–1247.

[3] KIEFER, J. and SACKS, J. (1963). Asymptotically optimum sequential inference and design. *Ann. Math. Statist.* **34** 705–750.

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

[5] ROBERTS, C. D. (1962). An asymptotically optimal sequential design for comparing several experimental categories with a standard or control. Institute of Statistics, Mimeo Series No. 344, Univ. of North Carolina. (Also Ph.D. dissertation, Univ. of N. C.)

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