ON SEQUENTIAL DESIGNS FOR MAXIMIZING THE SUM OF n OBSERVATIONS[1]

By R. N. Bradt, S. M. Johnson, and S. Karlin

1. Introduction. An important simple type of sequential design problem is as follows: We have two binomial random variables, X and Y, having parameters under the two hypotheses, H_1 and H_2, given by

$$\begin{array}{rccc} & & X & Y \\ (\zeta) & H_1 & p & q \\ (1-\zeta) & H_2 & q & p, \end{array}$$

where ζ is the a priori probability that H_1 is true. We wish to maximize the sum of n observations. The procedure for selecting an X or Y observation at each stage, of course, takes account of all the previous history.

A more realistic version of the design problem deals with the situation such that X and Y have parameters p and q, respectively, where an a priori distribution $F(p, q)$ is known. The problem holds interest for several reasons. It would appear to be one of the simplest problems in the sequential design of an experiment that can be posed; hence its analysis is a step towards obtaining a body of information relative to *specific* sequential design problems. It has not only this general interest but also, as it stands, it has applications in particular problems such as learning theory, biology, and medicine; see [1], for instance, in which applications in the latter two fields may be found. A discussion of problems of this general variety and of certain strategies has been published by Robbins [2]. More immediately, in the final section of this paper it is shown that the solution to the problem in which p has a priori distribution F and q is assumed known, explicitly obtained in Section 4, yields directly the solution of a problem in industrial inspection.

The type of problem known as the "Two-armed Bandit" is a special case of the preceding. In its "classical" formulation (whence the name), we have a slot machine with two arms, an X-arm and a Y-arm. When either arm is pulled, the machine pays off either one unit or nothing; and the probability of winning with one arm is p, and, with the other, q. A priori it is unknown which is which, but the probability ζ that it is the X-arm which has probability p of success is assumed known. One is allowed n plays, and a sequential design, or strategy, is desired which will maximize the expected winnings.

We shall use here for intuitive concreteness the gambling interpretation and terminology.

It has been conjectured for this problem that the optimal strategy is S_1 : on

Received Sept. 28, 1954.

[1] This work was supported in part by an Office of Naval Research contract at Stanford University.

each play choose the arm having, at that time, the maximum expected probability of paying off, i.e., play each time as though there were but one play remaining. This conjecture has been verified to hold for $n \leqq 8$.

The "Two-armed Bandit" problem can be generalized in two directions. The random variables may have distributions other than binomial. Sufficient conditions that S_1 be optimal are given and it is shown that, for the binomial case following S_1, the expected winnings per play tend to $\max[p, q]$ as $n \to \infty$.

The other direction of generalization was the problem as originally introduced: X and Y are binomial with parameters p and q having a priori distribution $F(p, q)$. In Section 3 it is shown that several properties which one intuitively expects the optimum strategy to possess are not, in general, characteristics of the optimal strategy—e.g., for p and q independently distributed, if only n is sufficiently large, S_1 is not optimal; the optimal strategy may not stay on a winner; and the expected gain on the r-th play is not necessarily a nondecreasing function of r. Also, S_k, the strategy which maximizes the expected winnings over the next k plays, is not always an improvement over S_{k-1}.

In Section 4, the parameter q and the a priori distribution $F(p)$ are assumed known. In this case the optimal strategy is determined *explicitly* and is shown to have those intuitive properties which have been previously noted not to be general characteristics of optimal strategies. These results are applied, in the final section, to obtain the optimal procedure in a certain industrial inspection problem.

2. The "Two-armed Bandit." 2.1. The statistical problem which goes under this general title is that of finding a design which will maximize the sum of n independent observations in the following situation: Let X and Y be real-valued random variables having cdf's F_i and G_i, respectively, under hypothesis $H_i (i = 1, 2)$ and ζ be the a priori probability that H_1 is the true hypothesis. The problem is to devise a sequential design which will maximize the expected value of the sum of n observations, each of which is to be an observation either of X or of Y.

Let f_i and g_i be the densities corresponding to F_i and G_i with respect to the measure ψ. Let $W_n(\zeta, S^*)$ denote the expected value of the sum of the n observations if ζ is the a priori probability for H_1 and the design, S^*, is used. If one observed X first and then continued for $n - 1$ steps following the optimal rule S^*, then the expected sum would be

$$\begin{aligned} A_n = \zeta \int_{-\infty}^{\infty} t f_1(t)\, d\psi + (1 - \zeta) \int_{-\infty}^{\infty} t f_2(t)\, d\psi \\ + \int_{-\infty}^{\infty} W_{n-1}\left(\frac{\zeta f_1(t)}{\zeta f_1(t) + (1 - \zeta) f_2(t)}, S^*\right) [\zeta f_1(t) + (1 - \zeta) f_2(t)]\, d\psi. \end{aligned} \tag{2.1.1}$$

Similarly, if Y were observed first and the optimal rule followed for the re-

maining $n - 1$ steps, the expected sum would be

$$B_n = \zeta \int_{-\infty}^{\infty} tg_1(t)\, d\psi + (1 - \zeta) \int_{-\infty}^{\infty} tg_2(t)\, d\psi + \int_{-\infty}^{\infty} W_{n-1}\left(\frac{\zeta g_1(t)}{\zeta g_1(t) + (1 - \zeta)g_2(t)}, S^*\right) [\zeta g_1(t) + (1 - \zeta)g_2(t)]\, d\psi. \tag{2.1.2}$$

Hence, $W_n(\zeta, S^*) = \max(A_n, B_n)$.

A natural design to be considered is that which requires that one maximize step by step—i.e., after the j-th observation, the a posteriori probability, ζ_j, is computed; and at the next step, the random variable corresponding to the maximum of

$$\int_{-\infty}^{\infty} t[\zeta_j f_1(t) + (1 - \zeta_j)f_2(t)]\, d\psi$$

and

$$\int_{-\infty}^{\infty} t[\zeta_j g_1(t) + (1 - \zeta_j)g_2(t)]\, d\psi$$

is observed. Denote this stepwise maximization design by S_1.

THEOREM 2.1. *If the likelihood ratios f_2/f_1 and g_2/g_1 have the same distributions under H_1 and also under H_2, then S_1 is the optimal design.*

PROOF. Since

$$\zeta_1 = \begin{cases} \dfrac{1}{1 + \dfrac{1 - \zeta}{\zeta} \dfrac{f_2(x)}{f_1(x)}} & \text{if } X \text{ is observed first,} \\[2ex] \dfrac{1}{1 + \dfrac{1 - \zeta}{\zeta} \dfrac{g_2(x)}{g_1(x)}} & \text{if } Y \text{ is observed first,} \end{cases} \tag{1}$$

and the likelihood ratios have the same distributions, the distribution of ζ_1 is independent of which random variable is observed first. Hence, the expected value of the optimal yield from the last $(n - 1)$ steps is independent of the choice for the first step. One can, therefore, maximize the expected sum of n observations by choosing at the first step the random variable having the larger expected value and continuing with the optimal design for the remaining steps.

Since all the random variables are assumed to be independent, the same argument shows that, given ζ_j, it is optimal to follow S_1 for the $(j + 1)$-st step.

An example in which the likelihood ratios are distributed alike is:

	X	Y	
H_1	$N(0, 1)$	$N(\nu, 1)$	
H_2	$N(\mu, 1)$	$N(0, 1)$	$(\mu > 0, \nu > 0)$,

with $\mu = \nu$. However, it can be shown that for $n = 2$ and $(1 - \zeta)\mu = \zeta\nu$, S_1 is optimal *only* if $\mu = \nu$.

2.2. A special case of the "Two-armed Bandit" of widespread interest (the "classical" case) is that in which the random variables have binomial distributions with parameters given by:

	X	Y
H_1	p	q
H_2	q	p.

A second example in which the likelihood ratios are distributed alike is furnished here if $p + q = 1$. Hence, for that case, S_1 is the optimal design. Indeed, it is a conjecture that for any choice of p and q, S_1 is optimal; it has been verified to be for $n \leqq 8$. Optimal or not, S_1 has the desirable property of being consistent; i.e., Theorem 2.2 holds.

THEOREM 2.2. *Following the design S_1, the expected value of the average of the first n observations converges to* $\max(p, q)$ *as* $n \to \infty$.

PROOF. Assume $p > q$. Then

$$W_1(\zeta, S_1) = \begin{cases} q + (p - q)\zeta & \text{for } \zeta \geqq \tfrac{1}{2}, \\ p - (p - q)\zeta & \text{for } \zeta \leqq \tfrac{1}{2}; \end{cases} \tag{1}$$

and if $\zeta \geqq \frac{1}{2}$,

$$\begin{aligned} W_n(\zeta, S_1) = W_1(\zeta, S_1) &+ W_{n-1}\left(\frac{p\zeta}{P_\zeta(X = 1)}\right) P_\zeta(X = 1) \\ &+ W_{n-1}\left(\frac{(1 - p)\zeta}{P_\zeta(X = 0)}\right) P_\zeta(X = 0); \end{aligned} \tag{2}$$

while if $\zeta \leqq \frac{1}{2}$,

$$\begin{aligned} W_n(\zeta, S_1) = W_1(\zeta, S_1) &+ W_{n-1}\left(\frac{q\zeta}{P_\zeta(Y = 1)}\right) P_\zeta(Y = 1) \\ &+ W_{n-1}\left(\frac{(1 - q)\zeta}{P_\zeta(Y = 0)}\right) P_\zeta(Y = 0), \end{aligned} \tag{3}$$

where $P_\zeta(Z = c) = \zeta P_r(Z = c \mid H_1) + (1 - \zeta)P_r(Z = c \mid H_2)$.

W_1 is clearly convex, symmetric about $\zeta = \frac{1}{2}$, and continuous. So is W_n, since by an inductive argument, W_n is symmetric about $\frac{1}{2}$, (2) and (3) are continuous, and each (by formal differentiation twice) is convex. Also it is easily seen that $W_n(\zeta, S_1) = n[(p - q)\zeta + q]$ for ζ near 1.

Let $a_n(\zeta, S_1) = 1/nW_n(\zeta, S_1)$. Then a_n is convex, continuous, and bounded above by p on [0, 1]. Furthermore, $|a_n'(\zeta, S_1)| \leqq p - q$. As a consequence of a more general result below, Lemma 3.3, a_n is nondecreasing in n. Hence, $a(\zeta, S_1) = \lim_{n\to\infty} a_n(\zeta, S_1)$ exists and is convex and continuous on [0, 1]. Moreover, since

$na_n(\zeta, S_1)$ satisfies (1) and (2),

$$(4)\quad a(\zeta, S_1) = \begin{cases} a\left(\dfrac{p\zeta}{P_\zeta(X = 1)}\right) P_\zeta(X = 1) + a\left(\dfrac{(1-p)\zeta}{P_\zeta(X = 0)}\right) P_\zeta(X = 0) & \zeta \geqq \frac{1}{2}, \\ a\left(\dfrac{q\zeta}{P_\zeta(Y = 1)}\right) P_\zeta(Y = 1) + a\left(\dfrac{(1-q)\zeta}{P_\zeta(Y = 0)}\right) P_\zeta(Y = 0) & \zeta \leqq \frac{1}{2}. \end{cases}$$

Suppose that the minimum of $a(\zeta, S_1)$ is assumed at $\zeta_0 \geqq \frac{1}{2}$. Then it also assumes its minimum at $p\zeta_0/P_{\zeta_0}(X = 1) > \zeta_0$. By iteration, it assumes its minimum at $(p^n\zeta_0)/[p^n\zeta_0 + q^n(1 - \zeta_0)]$, which tends to 1 as $n \to \infty$. Hence, ζ_0 could be taken to be 1. If, on the other hand, $\zeta_0 < \frac{1}{2}$, the analogous procedure shows that ζ_0 could be taken to be 0. Thus, the minimum of $a(\zeta, S_1)$ is assumed either at 0 or 1. But $a(0, S_1) = a(1, S_1) = p$, which establishes the theorem.

3. A generalized "Two-armed Bandit."[2] This section is concerned with the Bayes problem of maximizing the expected number of successes in n trials when at each trial we are free to choose between two binomial random variables, X and Y, whose probabilities of success, p and q, respectively, are unknown, but a known a priori distribution, $F(p, q)$, is specified.

The special case where $F(p, q)$ concentrates at the two fixed points (p, q) and (q, p) with probabilities ζ and $1 - \zeta$, respectively, leads to the "classical" problem considered in Section 2.2.

Let S denote a strategy for choosing between X and Y and $W_n[(p, q), S]$ denote the expected number of successes in following S for n plays for given p and q. Then the expected number of successes is

$$(3.1)\qquad W_n(F, S) = \int_0^1 \int_0^1 W_n[(p, q), S]\, dF(p, q).$$

We will find it convenient sometimes to express this as $W_n(dF, S)$. The best strategy is the one maximizing $W_n(F, S)$. Since n is finite, the maximum exists.

Example 1. Suppose $p + q = 1$ with probability 1; i.e., $F(p, q)$ is of the form $F(p, 1 - p)$. In this case a success or failure of X is equivalent to (gives the same information as) a failure or success with Y, respectively.

S_1 is optimal in this case; for let $F_{k-1}(p, q)$ denote the a posteriori probability after $(k - 1)$ plays and S_{n-k} the optimal strategy for $(n - k)$ plays. Then using X followed by S_{n-k} yields

$$\int p\, dF_{k-1} + W_{n-k}(p\, dF_{k-1}, S_{n-k}) \int p\, dF_{k-1}$$
$$+ W_{n-k}((1 - p)\, dF_{k-1}, S_{n-k}) \int (1 - p)\, dF_{k-1},$$

[2] This section represents an extension of some preliminary work by S. Johnson and S. Karlin at The RAND Corporation.

and using Y followed by S_{n-k} yields

$$\int q\, dF_{k-1} + W_{n-k}(q\, dF_{k-1}, S_{n-k}) \int q\, dF_{k-1}$$
$$+ W_{n-k}((1-q)\, dF_{k-1}, S_{n-k}) \int (1-q)\, dF_{k-1}.$$

Since $q = 1 - p$, X is the optimal play if and only if $\int p\, dF_{k-1} \geqq \int q\, dF_{k-1}$; i.e., S_1 is optimal. This example is related to the result of Theorem 2.1, which embraces a special case of $F(p, q) = F(p, 1 - p)$.

3.1. Our first task is to obtain the complete strategy for $n = 2$ when the a priori distribution is $F(p)G(q)$. It is important to notice that, if the number of trials is n, then only designs which are functions of the first n moments, $\mu_1, \cdots, \mu_n$ of F and $\mu_1', \cdots, \mu_n'$ of G, need be considered. This is a consequence of the fact that the expected yield for any strategy is an expression involving at most these moments. Thus all strategies describing a first move can be expressed in terms of functions $T_j(\mu_1, \cdots, \mu_n, \mu_1', \cdots, \mu_n')$ such that if

$$T_j(\mu_1, \cdots, \mu_n, \mu_1', \cdots, \mu_n') \geqq 0,$$

then X is chosen at the first trial; otherwise Y is used first.

Suppose for definiteness that $\mu_1 \geqq \mu_1'$; we determine necessary and sufficient conditions that X be used first when $n = 2$. Using the fact that on the last trial one chooses the random variable having greatest expected value, if X is used first the expected yield is

$$\mu_1 + \mu_1\left(\frac{\mu_2}{\mu_1}\right) + (1 - \mu_1) \max\left\{\mu_1', \frac{\mu_1 - \mu_2}{1 - \mu_1}\right\}. \tag{1}$$

Since (1) $\geqq 2\mu_1 \geqq \mu_1 + \mu_1' \geqq 2\mu_1'$, X followed by optimal is better than Y followed always by X, which is better than Y followed always by Y. Of the other two strategies starting with Y, the one requiring X if $Y = 1$ has expected yield $2\mu_1' + \mu_1\mu_1' - \mu_2'$, which can be shown to be less than or equal to that for the strategy requiring Y if $Y = 1$, namely,

$$\mu_1' + \mu_2' + (1 - \mu_1')\mu_1. \tag{2}$$

Upon comparison, (2) $\leqq$ (1) if and only if

$$\text{either} \quad \mu_2 \geqq \mu_2' \quad \text{or} \quad \mu_1 + \mu_1\mu_1' \geqq \mu_1' + \mu_2'. \tag{3.1.1}$$

Combining and rewriting in a symmetric form, we have

LEMMA 3.1. *If $n = 2$ and p and q are independent with moments μ_i and μ_i', then X is used on the first trial if and only if*

$$\max\{\mu_2 - \mu_1'\mu_1, \mu_1 - \mu_1'\} \geqq \max\{\mu_2' - \mu_1\mu_1', \mu_1' - \mu_1\}.$$

Our next theorem shows that in almost all circumstances of independent a priori distributions for p and q, the optimal design and S_1 cannot agree.

THEOREM 3.1. *If p and q are independent with a priori distributions $F(p) = \int_0^p \phi(t)\, dt$ and $G(q) = \int_0^q \psi(t)\, dt$, where ϕ and ψ are continuous and positive for*

$0 < t < 1$, *then there exists an* n *such that for* n *trials, the optimal design does not agree with* S_1.

PROOF. (By contradiction.) Suppose for definiteness that $1 > \int_0^1 t\phi(t)\,dt = b \geqq a = \int_0^1 t\psi(t)\,dt$. According to S_1 it is clear that X is used first and, by the Schwarz inequality, that we do not change random variables if a success occurs.

It is easily shown, in view of the hypothesis on ϕ, that if r and s tend to infinity so that $r/(r+s) \to t_0$, then

$$\frac{\int_0^1 t^{r+1}(1-t)^s\phi(t)\,dt}{\int_0^1 t^r(1-t)^s\phi(t)\,dt} \to t_0. \tag{1}$$

(This can also be obtained as a consequence of the law of large numbers where the relative frequency of success tends to t_0.) Hence, taking $t_0 > a + \epsilon/2$, we can choose r and s so that

$$\left|\frac{r}{r+s} - a\right| < \epsilon, a + \epsilon > \frac{\int_0^1 t^{r+i+1}(1-t)^s\phi(t)\,dt}{\int_0^1 t^{r+i}(1-t)^s\phi(t)\,dt} > a \qquad \text{for } i = 0 \text{ and } 1. \tag{2}$$

Furthermore, ϵ may be chosen sufficiently small that also

$$\int t^2\phi(t)\,dt > \left[\int t\phi(t)\,dt\right]^2 + 3\epsilon.$$

Now let $n = r + s + 2$ and suppose that the first r plays resulted in successes with X and the next s plays were failures with X. This agrees with the procedure prescribed by S_1 and has positive probability of occurrence. There are now two plays left and, from (2), S_1 requires X on the next step. However, Lemma 3.1 gives necessary and sufficient conditions that the use of X is optimal and we show that these are violated.

There are two steps left and the a posteriori probability distribution is $F^1(p)G(q)$ where

$$F^1(p) = \frac{p^r(1-p)^s\phi(p)}{\int_0^1 t^r(1-t)^s\phi(t)\,dt} \quad \text{and} \quad G(q) = \int_0^q \psi(t)\,dt.$$

On account of (2),

$$\int_0^1 p\,dF^1(p) > \int_0^1 q\psi(q)\,dq,$$

and on direct calculation it is seen that both of the inequalities of (3.1.1) are violated. Hence, following S_1, we arrive at a nonoptimal yield, and the theorem is established.

3.2. S_1 may be described as that procedure requiring at each step the random

variable that would be optimal, were there but one trial remaining. In a similar spirit, let S_j be the strategy which requires at each trial the random variable that would be optimal were there j trials remaining, with the understanding that if fewer than j trials remain, then the optimal procedure is followed. (The strategy S_2 for p and q independent is determined by the relations given in Lemma 3.1.)

We have, thus, a sequence of strategies, $S_1, S_2, \cdots, S_n$. For a series of n trials, S_n is the optimal strategy and hence $W_n(F, S_n) \geqq W_n(F, S_j)$ for all $j < n$. Intuitively one might expect that the $W_n(F, S_j)$ are nondecreasing in j; i.e., the more steps ahead we take into account, the better the strategy. However, it can be shown that there exist a priori distributions such that for $n = 3$, $W_3(F, S_1) > W_3(F, S_2)$. The details are omitted.

3.3. The next principle examined is that of "staying on a winner": Does the optimal strategy have the property that whenever a success occurs, the same random variable is required on the next trial? S_1, for instance, has this property. However, it is not always a characteristic of an optimal strategy, as the following example shows.

Suppose $F(p, q)$ concentrates probability 0.8 on (0.1, 0) and 0.2 on (0.9, 1). It can be shown that for this example we must stay on a loser but switch from a winner for the case of $n = 2$.

A property related to the intuitive notion of staying on a winner is that of "monotonicity," which we discuss for p and q independent. Let $S^*(n, FG)$ denote the optimal strategy for n trials against a priori FG and let $dF^s = p\,dF/\int p\,dF$ and $dF^f = (1 - p)\,dF/\int(1 - p)\,dF$, with G^s and G^f similarly defined. $S^*(n, FG)$ will be termed monotone if:

(i) $S^*(n, F^fG)$ allows X first $\Rightarrow S^*(n, F^sG)$ requires X first,

(ii) $S^*(n, FG^f)$ allows Y first $\Rightarrow S^*(n, FG^s)$ requires Y first,

and

(iii) $S^*(n, F^sG)$ allows Y first $\Rightarrow S^*(n, FG^s)$ requires Y first.

For instance, (i) is to be thought of as: if a prior "free" observation of X were allowed, then, if we might use X on the first trial, even if X had failed on the prior trial, we should certainly use X on the first trial, if the prior trial had resulted in a success with X.

LEMMA 3.2. If for $1 \leqq k \leqq n - 1$ and for all F and G, $S^*(k, FG)$ is monotone, then $S^*(n, FG)$ stays on a winner.

The proof is omitted.

With the aid of the results of Lemma 3.1, it can be verified directly that, for $n = 1$ and $n = 2$, the optimal strategy is monotone and, therefore, for $n = 3$ and p and q independent, the optimal strategy stays on a winner.

The general monotonicity property for independent parameters remains an open question.

3.4. In using any strategy, S, let $Z_r = 1$ if the random variable used on the r-th trial wins, and $Z_r = 0$ otherwise; i.e., Z_r is the contribution of the r-th trial. The last property considered is whether, for S^* denoting the optimal S,

$E[Z_r \mid S^*]$ is monotone increasing in r. We show first that $E[Z_r \mid S_1]$ is nondecreasing in r, and second that $E[Z_r \mid S^*]$ may decrease.

LEMMA 3.3. *$E[Z_r \mid S_1]$ is nondecreasing in r for every initial distribution F.*

PROOF. It is enough to prove the result for the first two trials. Suppose $Z_1 = X$, then $E[Z_1 \mid S_1] = \int p \, dF$. But $E[Z_2 \mid S_1] = E[E[Z_2 \mid Z_1, S_1]] \geqq E[E[X \mid Z_1, S_1]] = E[Z_1 \mid S_1]$.

In contrast to this result, consider the case of $n = 3$,

$$F(p) = p^5(1 - p)^3 / \int p^5(1 - p)^3 \, dp,$$

and $G(q) = q$. For optimal return, X should be employed first with expected return from the first trial of 0.6. If success results, then X is used again, while if failure occurs, then the criteria of Lemma 3.1 require Y. The a priori expected yield from the second trial is $64/110 < 66/110 = 0.6$.

4. The case of one known and one random probability of success. In this section we examine in detail the situation in which X has a binomial distribution with $p = \Pr(X = 1)$ unknown but selected by a known a priori distribution, F, while Y has a binomial distribution with known parameter, q.

The results of the preceding section were informative largely in a negative sense; there are many nice properties which optimal strategies do not possess. Many properties which seemed obvious but which were not in general enjoyed by optimal strategies in the general case, are held by the optimal strategy when one of the random variables has a known distribution. Hence, the rather detailed proofs in this section.

We establish a series of lemmas describing some properties and the form of the optimal strategy and then obtain an explicit statement of it.

LEMMA 4.1. *If Y is required at any trial according to an optimal strategy, then Y is required thereafter.*

PROOF. First, it is easily seen that if at any trial Y is required, then the optimal choice for the next trial is independent of whether Y wins or not.

Now suppose that at some trial, let us say at the first one, Y is required and is used r times, but that X is allowed on the $(r + 1)$-st trial. Then the expected winnings are

$$rq + W_{n-r-1}(F^s, q) \int_0^1 p \, dF + W_{n-r-1}(F^f, q) \int_0^1 (1 - p) \, dF, \tag{1}$$

where F^s and F^f are the a posteriori probabilities defined in Section 3.3, and $W_k(F, q)$ is the expected gain against a priori F in k trials pursuing an optimal strategy. But using X first followed by r trials of Y yields the same amount, contradicting the fact that Y was required on the first trial. Hence Y must be required throughout.

As a consequence of Lemma 4.1, we can characterize an optimal strategy. We use the notation

$$\mu_i = \int_0^1 p^i \, dF.$$

LEMMA 4.2. *There exists a function, Q, of n and F such that for n trials remaining and F the a priori distribution of p at that time, Y is required if and only if $q > Q(n, F)$.*

PROOF. From Lemma 4.1, Y is required if and only if

$$(1) \qquad nq > \mu_1 + \mu_1 W_{n-1}(F^s, q) + (1 - \mu_1)W_{n-1}(F^f, q) = nK_n(F, q).$$

Now $W_1(F, q) = \max\{q, \mu_1\}$ and hence is nondecreasing convex in q for all F. By easy induction, $W_n(F, q)$ is nondecreasing convex in q for all F and n and, hence, so is $K_n(F, q)$. Since $K_n(F, 0) = \mu_1 > 0$ and $K_n(F, 1) = 1 - (1 - \mu_1)/n < 1$, it follows that for each n and F there is a point $Q(n, F)$ such that $q > Q(n, F)$ if and only if $q > K_n(F, q)$—i.e., if and only if Y is required.

We shall adopt the convention that if $q = Q(n, F)$ we shall always use X, giving us a definite optimal strategy:

If $q > Q(n, F)$, use Y for all n trials. If $q \leqq Q(n, F)$, use X on the first trial and compute the a posteriori distribution of p, F', and compare q and $Q(n - 1, F')$, following the above rules for choice at the second trial, etc.

Having characterized the optimal strategy, we turn to a series of lemmas describing more precisely its form and properties.

LEMMA 4.3. *For all F and $n \geqq 2$, $Q(n, F) \geqq Q(n - 1, F)$.*

PROOF. Suppose the contrary. Then for $Q(n, F) < q < Q(n - 1, F)$, Y would be required on the first trial and X on the second, contradicting Lemma 4.1.

LEMMA 4.4. *For all F, q, and n,*

$$W_n(F^s, q) \geqq W_n(F, q) \geqq W_n(F^f, q).$$

PROOF.

$$\text{I}: q \geqq \max\{Q(n, F^f), Q(n, F), Q(n, F^s)\}.$$

Then $nq = W_n(F^s, q) = W_n(F, q) = W_n(F^f, q)$.

$$\text{II}: q \leqq \min\{Q(n, F^f), Q(n, F), Q(n, F^s)\}.$$

We proceed by induction. The lemma holds for $n = 1$, since for all q and F,

$$(1) \qquad \max\left\{q, \frac{\mu_2}{\mu_1}\right\} \geqq \max\{q, \mu_1\} \geqq \max\left\{q, \frac{\mu_1 - \mu_2}{1 - \mu_1}\right\}.$$

In the case under consideration,

$$(2) \quad W_n(F^s, q) = \frac{\mu_2}{\mu_1} + \frac{\mu_2}{\mu_1} W_{n-1}(F^{ss}, q) + \left(1 - \frac{\mu_2}{\mu_1}\right) W_{n-1}(F^{sf}, q) = \frac{\mu_2}{\mu_1} + A_n,$$

$$(3) \quad W_n(F, q) = \mu_1 + \mu_1 W_{n-1}(F^s, q) + (1 - \mu_1)W_{n-1}(F^f, q) = \mu_1 + B_n,$$

$$(4) \quad \begin{aligned} W_n(F^f, q) &= \frac{\mu_1 - \mu_2}{1 - \mu_1} + \frac{\mu_1 - \mu_2}{1 - \mu_1} W_{n-1}(F^{fs}, q) \\ &\quad + \left(1 - \frac{\mu_1 - \mu_2}{1 - \mu_1}\right) W_{n-1}(F^{ff}, q) = \frac{\mu_1 - \mu_2}{1 - \mu_1} + C_n. \end{aligned}$$

By the induction hypothesis,

$$(5)\qquad \begin{aligned} W_{n-1}(F^{ss}, q) \geqq W_{n-1}(F^{s}, q) \geqq W_{n-1}(F^{sf}, q) \\ = W_{n-1}(F^{fs}, q) \geqq W_{n-1}(F^{f}, q) \geqq W_{n-1}(F^{ff}, q); \end{aligned}$$

and it is easily shown that since

$$\frac{\mu_2}{\mu_1} > \mu_1 > \frac{\mu_1 - \mu_2}{1 - \mu_1},$$

$A_n \geqq B_n \geqq C_n$. Thus the lemma is established for this case.

As a consequence of Case II, $Q(n, F^f) \leqq \min\{Q(n, F), Q(N, F^s)\}$. For if (say) $Q(n, F) = \min\{Q(n, F), Q(n, F^s)\} < Q(n, F^f)$, then for $q = Q(n, F)$, $nq = W_n(F, q) < W_n(F^f, q)$, a contradiction of the case just established.

$$\text{III}: Q(n, F^f) < q \leqq \min\{Q(n, F), Q(n, F^s)\}.$$

Then $W_n(F^f, q) = nq \leqq \min\{W_n(F, q)\ W_n(F^s, q)\}$. But by induction argument parallel to that for Case II, it is shown that $W_n(F, q) \leqq W_n(F^s, q)$.

From Case III it follows by the same reasoning as above that $Q(n, F) \leqq Q(n, F^s)$. Hence, there is only one remaining case.

$$\text{IV}: Q(n, F^f) \leqq Q(n, F) < q < Q(n, F^s).$$

Immediately, $W_n(F^f, q) = W_n(F, q) = nq \leqq W_n(F^s, q)$, and the lemma is established.

Interspersed in the proof just completed is the proof of

LEMMA 4.5.

$$Q(n, F^f) \leqq Q(n, F) \leqq Q(n, F^s).$$

LEMMA 4.6. *Following the optimal strategy, if a success occurs on any trial, then the same random variable is used on the next trial—i.e., stay with a winner.*

PROOF. In view of Lemma 4.1, we need only show that if X is required and wins, then X is required on the next trial. It is clearly sufficient to show this for the first trial. Suppose to the contrary that $Q(n, F) \geqq q > Q(n - 1, F^s)$. By Lemma 4.3, $Q(n - 1, F^s) \geqq Q(n - 2, F^s) \geqq \cdots \geqq Q(1, F^s)$ and, clearly, $Q(1, F^s) \geqq \mu_2/\mu_1 > \mu_1$. Hence, $q > \mu_1$.

By Lemma 4.5, $q > Q(n - 1, F^f)$ also. Consequently Y is required on the second trial, regardless of the outcome of first. Then

$$(1)\qquad nq \leqq W_n(F, q) = \mu_1 + (n - 1)q.$$

Hence, $q \leqq \mu_1$, and we have a contradiction.

LEMMA 4.7. *The a priori expected value of the yield on the r-th step is nondecreasing in r when using an optimal strategy.*

PROOF. The proof can be obtained with the aid of the foregoing lemmas; it is left as an exercise for the reader.

As we have noted in Section 3, Lemma 4.7 is not true, in general, while Lemma 4.6. is the "stay on a winner" rule which, appealing as it is, does not hold in general.

With the above lemmas we are in a position to determine explicitly the value of $Q(n, F)$. Assume that $q = Q(n, F)$; then the optimal strategy has the following form for appropriate k_i.

(4.1) (A) Observe X until a failure occurs.

(B) There exists an integer $k_1 \geqq 0$ such that if at least k_1 successes preceded the first failure, continue with X; otherwise switch to Y for the remaining trials.

(C) There is an integer $k_2 \geqq 0$ attached to the second failure such that if at least $k_1 + k_2$ successes with X precede the second failure of X, continue with X; otherwise switch to Y for the remaining trials.

(D) In general, let S_r be the number of successes that precede the r-th failure of X. If $S_r \geqq k_1 + k_2 \cdots + k_r$, continue with X; otherwise switch to Y for the remaining trial.

Thus, any sequence $k = (k_1, k_2, \cdots, k_n)$ of integers, $0 \leqq k_i \leqq n$, corresponds to a strategy of the same form as the optimal.

Let E_k denote expectation given k and F, and E_{kp} denote expected value given k and p.[3] In using any strategy for n trials, X will be used a certain number, N_x, of times, and there will be a certain number, S_x, of successes with X; similarly for Y.

THEOREM 4.1.

$$Q(n, F) = \max_k \left\{ \frac{E_k[S_x]}{E_k[N_x]} \right\}.$$

PROOF. $q = Q(n, F)$ implies $nq = W_n(F, q)$. But since the optimal strategy corresponds to a sequence k, this is equivalent to $nq = \max_k\{E_k[S_x] + E_k[S_y]\}$. However, $E_k[S_y] = qE_k[N_y]$, and neither $E_k[S_x]$ nor $E_k[N_y]$ depends on q. Hence $q = Q(n, F)$ implies each of the following equivalent statements: $nq = \max\{E_k[S_x] + qE_k[N_y]\} \Leftrightarrow q \geqq E_k[S_x]/E_k[N_x]$ for all k with equality for some $\Leftrightarrow q = \max_k E_k[S_x]/E_k[N_x]$.

COROLLARY.

$$Q(n, F) = \max_k \left\{ \frac{\int_0^1 E_{kp}[N_x]p \, dF}{\int_0^1 E_{kp}[N_x] \, dF} \right\} = \max_k \left\{ \frac{\int_0^1 (n - E_{kp}[N_y])p \, dF}{\int_0^1 (n - E_{kp}[N_y]) \, dF} \right\}.$$

[3] The authors are indebted to the referee for suggesting the following derivation of $Q(n, F)$, which is somewhat simpler and more illuminating than that originally used to obtain the result.

We give two methods of evaluating $Q(n, F)$. The first proceeds by obtaining directly a formula for $E_{kp}[N_y]$ and yields

$$E_{kp}[N_y] = \sum_{j=1}^{n} (n-j)p^{j-\phi(j)-1}(1-p)^{\phi(j)+1}$$

$$\cdot \sum{}^{*} \prod_{i=0}^{\phi(j)-1} \binom{k_{i+1}+f_i-1}{f_i} \binom{j - \sum_{\nu=1}^{\phi(j)} k_\nu - \sum_{\nu=1}^{\phi(j)-1} f_\nu - 1}{\phi(j) - \sum_{\nu=1}^{\phi(j)-1} f_\nu}, \tag{4.2}$$

where $\phi(j) = \max\{i: \sum_{\nu=1}^{i} k_\nu + i < j\}$ = number of failures of X in the first $(j-1)$ trials and $\sum^*$ denotes the sum over all choices of f_i's such that $f_i \geqq 0$: $f_i = 0$, if $k_{i+1} = 0$, $f_0 = 0$; $\sum_{i=0}^{\nu} f_i \leqq \nu$ and $\sum_i f_i = \phi(j)$ (f_i denotes the number of failures between the $(k_1 + k_2 + \cdots k_i)$-th and $(k_1 + k_2 + \cdots + k_{i+1})$-th success).

The second proceeds by obtaining directly a formula for $E_{kp}[S_x] = E_{kp}[N_x]p$. While more complicated in appearance and derivation, it is the result of a direct counting.

$$E_{kp}[S_x] = \sum_{r=0}^{r=n} I_r, \tag{4.3}$$

where

$$\begin{aligned} I_r = {} & \sum_{a_1=0}^{1} \binom{a_1-1}{a_1-1} + \cdots \\ & + \binom{k_2+a_1-1}{a_1-1} \sum_{a_2=0}^{2-a_1} \left\{ \binom{a_2-1}{a_2-1} + \cdots + \binom{k_3+a_2-1}{a_2-1} \right\} \\ & + \cdots + \sum_{a_r=0}^{r-a_1-a_2\cdots a_{r-1}} \left\{ \binom{a_r-1}{a_r-1} + \cdots + \binom{k_{r+1}+a_r-1}{a_r-1} \right\} \\ & \cdot \left[\binom{b_r}{b_r} p + \binom{b_r+1}{b_r} p^r + \cdots + \binom{b_r + n - \sum_{i=1}^{r} k_i - r - 1}{b_r} \right. \\ & \left. p^{n - \sum_{i=1}^{r} k_i - r} \right] p^{k_1+k_2+\cdots+k_r}(1-p)^r, \end{aligned}$$

with $b_r = r - a_1 - a_2 - \cdots - a_r$, and we interpret $\binom{-1}{-1} = 1$ and $\binom{c}{-1} = 0$, for $c \neq -1$.

Some special cases are worth noting:

$$Q(2, F) = \frac{\int_0^1 (p + p^2)\, dF}{\int_0^1 (1 + p)\, dF},$$

$$Q(3, F) = \max \left\{ \frac{\int_0^1 (p + p^2 + p^3)\, dF}{\int_0^1 (1 + p + p^2)\, dF}, \frac{\int_0^1 (p + 2p^2)\, dF}{\int_0^1 (1 + 2p)\, dF} \right\},$$

Each term of $Q(3, F)$ can occur; e.g., for $F(p) = p$, the first is the maximum (value 13/22), while for $F(p) = p^{1/5}$, the second is the maximum (value 23/88).

The expression for $Q(n, F)$ cannot be simplified in any essential way, which again testifies to the complex nature of the optimal strategies in sequential design problems.

If one chooses $k_1 = r$, $k_2 = k_3 = \cdots = 0$, then

$$\frac{E_k[S_x]}{E_k[N_x]} = \frac{\mu_1 + \mu_2 + \cdots + \mu_r + (n - r)\mu_{r+1}}{1 + \mu_1 + \cdots + \mu_{r-1} + (n - r)\mu_r}, \tag{4.4}$$

where $\mu_i = \int_0^1 p^i\, dF(p)$. For distributions such that $\sum_{\gamma=1}^{\infty} \mu_\gamma = +\infty$, at least, a reasonable approximation to $Q(n, F)$ may be had by taking $r = n - 1$ and using

$$L(n, F) = \frac{\mu_1 + \mu_2 + \cdots + \mu_n}{1 + \mu_1 + \cdots + \mu_{n-1}}$$

in place of $Q(n, F)$. For the uniform distribution, Q and L coincide for $n \leqq 4$, but not for larger n. It is worth remarking that L shares many of the properties of Q.

LEMMA 4.8. *L is nondecreasing in n, and $L(n, F^s) \geqq L(n, F) \geqq L(n, F^f)$.*

The proof is omitted.

We close by noting that if the number of trials is sufficiently large, one should almost always commence by using X. More precisely, we see from (4.4) that as $n \to \infty$, $Q(n, F)$ becomes at least $(\mu_{r+1})/\mu_r$, and this for every r. But $(\mu_{r+1})/\mu_r$ is the expected value of p given r successes which will tend to the supremum of the spectrum of F as r increases. Clearly, if q is greater than the supremum of the spectrum of F, one would never play X; while if q is less than the supremum, for all sufficiently large n, X should be used first. Finally, for a fixed n we have the following intuitive result:

LEMMA 4.9. *Given n and F, Y should never be used if*

$$q \leqq \frac{\int_0^1 p(1-p)^{n-1}\,dF}{\int_0^1 (1-p)^{n-1}\,dF}.$$

The rigorous proof may be supplied by the reader.

5. An applied problem. An interesting problem in industrial inspection is closely related to the problem of Section 4. Suppose that lots of n items are produced by a process having probability p of producing a defective where p varies from lot to lot according to an a priori distribution, $F(p)$. Let the loss per defective item accepted be unity and the cost of inspection be c per item inspected ($c < 1$). Items are drawn and inspected (defective items found being replaced by good items at no additional cost) until a sequential stopping rule terminates inspection, at which point the remainder of the lot is accepted. A stopping rule is desired which will minimize the expected loss.

One may proceed to attack this problem in the spirit of Section 4 and find a completely analogous series of lemmas, culminating in the theorem that for n items remaining and a priori distribution F, it is optimal to inspect another item or to accept the remaining n according as c is less than or greater than $Q(n, F)$. The same result is more immediately obtained by noting that the problem is equivalent to finding a rule to maximize the gain if one wins c for each item not inspected, nothing for each good item inspected, and one for each defective item inspected (and replaced). This latter problem is precisely that treated in Section 4.

REFERENCES

[1] THOMPSON, W. R., "On the likelihood that one unknown probability exceeds another in view of the evidence of two samples," *Biometrika*, Vol. 25 (1933), pp. 285–294.

[2] ROBBINS, H., "Some aspects of the sequential design of Experiments," *Bull. Amer. Math. Soc.*, Vol. 58, No. 5 (1952), pp. 527–535.