CONTRIBUTIONS TO THE "TWO-ARMED BANDIT" PROBLEM[1]

By Dorian Feldman

Michigan State University

0. Summary. The Bayes sequential design is obtained for an optimization problem involving the choice of experiments. Given are experiments A, B, densities p_1, p_2, a positive integer N and a number $\xi \varepsilon [0, 1]$. A sequence of N observations is to be made such that at each stage either A or B is observed, the loss being 1 if the experiment with density p_2 is chosen, 0 otherwise. ξ is the prior probability that A has density p_1. If the mean of p_1 is bigger than the mean of p_2 one obtains a more common version of the "two-armed bandit" (see e.g. [1]). The principal result of this paper is a proof of optimality for the procedure which at each stage chooses the experiment with higher posterior probability of being correct. Some attention is also given to the problem of computing risk functions.

1. Introduction. A prototype for the class of sequential design problems to be considered is the following: Let (A, B) be a pair of binomial experiments. An observation on A or B yields 1 for a success, 0 for a failure and the probabilities of success $p_1 > p_2$ are given, but it is not known which probability attaches to which experiment. The problem is to find a sequential procedure for taking observations so as to maximize the expected number of successes in a given (finite) number of trials. Of particular interest is the class of Bayes procedures when prior probabilities are assigned to the possible alternatives. This is the problem originally termed the "two-armed bandit" (TAB) because of its interpretation as an optimization problem for playing a game on a two-armed slot machine. The importance and applicability of this problem, and, in fact, of some of its generalizations, have been noted in papers by Bradt, Johnson, Karlin [1]; Bradt, Karlin [2] and Robbins [3].

Let the ordered pair (p, p') denote the assignment of p to A and p' to B and let H_A, H_B, be the hypotheses $(p, p') = (p_1, p_2)$ and $(p, p') = (p_2, p_1)$ respectively. We distinguish the following conditions on the problem:

(i) (A, B) represents a pair of binomial experiments $(1, 0)$

(ii) H_A, H_B are the symmetric singletons (p_1, p_2) and (p_2, p_1) respectively;

(iii) the total number of trials is finite;

(iv) the expected sum of the observations is to be maximized.

Conditions (i), (ii) can be relaxed in obvious ways. An optimality criterion which leads to the same procedure as (iv) for the TAB but which is more flexible will now be described. Let the experiment with higher probability of success, p_1,

Received December 26, 1961; revised April 10, 1962.

[1] This paper was prepared with the partial support of the Office of Naval Research (Nonr-222-43) and National Science Foundation grant #G18976. This paper in whole or in part may be reproduced for any purpose of the United States Government.

Part of a dissertation for the PhD degree, University of California, Berkeley.

be called "preferred" or "correct" while the one with p_2 is "wrong." Thus, under H_A, A is correct and B is wrong and under H_B, B is correct and A is wrong. Consider the procedure which minimizes the expected number of "mistakes", i.e., observations on the wrong experiments. Clearly, this procedure is precisely the one that satisfies (iv). Stating the criterion this way has the advantage that under a more general set-up the problem can be formulated without reference to the specific nature of the preference. Also, it allows a natural extension to the infinite case.

Some problems involving generalized versions of (i)-(iv) have been studied in [1], [2], [3], the results of [1] being of particular interest here. Realizing its limitations, we shall, nevertheless, refer to the following version of conditions (i)-(iv) as the generalized two-armed bandit (GTAB):

(i)′ (A, B) represents a pair of arbitrary, fixed experiments on a sample space $(T, \mathfrak{B})$;

(ii)′ H_A, H_B are the symmetric singletons (p_1, p_2), (p_2, p_1), the p_i's being densities with respect to some measure λ over $(T, \mathfrak{B})$;

(iii)′ the number of trials may be finite or infinite;

(iv)′ the expected number of observations on the experiment with density p_2 is to be minimized.

The equivalence of (iv)′ and criteria such as (iv) is guaranteed by the symmetry conditions (ii), (ii)′ whereby mistakes on A and B have equal weight. This will not be the case under condition (ii)″: H_A, H_B are arbitrary singletons $(p, p') = (p_1 p_2)$ and $(p, p') = (q_1, q_2)$ respectively. We shall briefly consider this case in Section 5.

Let $\xi_0 = \xi$ be the prior probability of H_A and let ξ_i, $i = 1, 2, \cdots$ be the successive posterior probabilities of H_A. The following procedure has been conjectured optimum for the TAB: at stage i observe A or B according as $\xi_{i-1} > \frac{1}{2}$ or $\xi_{i-1} < \frac{1}{2}$ with indifference at $\xi_{i-1} = \frac{1}{2}$. We denote this procedure by π^*. We shall show in Section 2 and 3 that π^* is optimum for the GTAB whether the number of trials is finite or not. The optimality of π^* has already been verified in the following cases:

(a) TAB, $N \geqq 10$ (N is the number of trials)

(b) under the following restriction of conditions (i)′, (ii)″, (iv) with N finite: p_1/q_1, p_2/q_2 have identical distributions under both H_A and H_B. (Bradt, Johnson, Karlin [1]).

In Section 4 we compute the risk for the infinite case for binomials with specific values of p_1, p_2.

2. GTAB finite case. Under conditions (i)′, (ii)′, (iii)′ of Section 1 we want to find a sequential procedure for taking observations so as to achieve (iv)′. Let ξ denote the prior probability of H_A. A procedure π_N will be evaluated on the basis of its risk function $R_N(\xi)$ which represents the expected number of mistakes given ξ and using π_N when N observations are contemplated.

An approach which has been found useful for handling multistage decision

problems of this sort is that of "dynamic programming" whereby present actions are compared on the assumption of future optimal behavior. Thus, in the present case suppose $N+1$ trials are contemplated, π_N is the optimum procedure for N trials, and π^A_{N+1}, π^B_{N+1} are defined as follows: π^A_{N+1} observes A first then follows π_N; π^B_{N+1} observes B first then follows π_N. $R^A_{N+1}(\xi)$, $R^B_{N+1}(\xi)$ are the risk functions of π^A_{N+1} and π^B_{N+1} respectively. If ξ is the prior probability of H_A then after one observation the posterior probability of H_A is given by

$$(1)\qquad \begin{aligned} \xi_A(t) &= \frac{p_1(t)\xi}{p_1(t)\xi + p_2(t)(1-\xi)} && \text{if } t \,\varepsilon\, T \text{ is observed on } A, \\ \xi_B(t) &= \frac{p_2(t)\xi}{p_2(t)\xi + p_1(t)(1-\xi)} && \text{if } t \,\varepsilon\, T \text{ is observed on } B. \end{aligned}$$

The risk of π^A_{N+1} (π^B_{N+1}) is the probability of a mistake on A (B) for the first trial plus the expectation of the future risk R_N over the possible values of ξ_A (ξ_B). Since π^A_{N+1}, π^B_{N+1} exhaust the possible optimum courses of action, the risk function $R_{N+1}(\xi)$ of π_{N+1}, the optimum procedure for $N+1$ trials, must satisfy

$$(2)\qquad \begin{aligned} R_{N+1}(\xi) &= \min\,[R^A_{N+1}(\xi), R^B_{N+1}(\xi)] \\ R^A_{N+1}(\xi) &= (1-\xi) \\ &\quad + \int R_N\left(\frac{p_1(t)\xi}{p_1(t)\xi + p_2(t)(1-\xi)}\right)[p_1(t)\xi + p_2(t)(1-\xi)]\,d\lambda(t) \\ R^B_{N+1}(\xi) &= \xi \\ &\quad + \int R_N\left(\frac{p_2(t)\xi}{p_2(t)\xi + p_1(t)(1-\xi)}\right)[p_2(t)\xi + p_1(t)(1-\xi)]\,d\lambda(t). \end{aligned}$$

Then π_{N+1} as determined by (2) is to choose A first if $R^A_{N+1}(\xi) < R^B_{N+1}(\xi)$, B first if $R^B_{N+1}(\xi) < R^A_{N+1}(\xi)$ with indifference if the two risks are equal. In either case π_N is followed for the last N trials. Because the observations are independent, conditioned on the choices of A and B, the relevant information after i stages is given by ξ_i, the posterior probability of H_A, and $N-i$, the number of trials remaining. Hence the optimum procedure for N trials is determined if we know $\pi_{k,1}(\xi)$, the optimum first choice as a function of ξ when the total number of trials is $k = 1, \cdots, N$. What we shall show is that actually the optimal first choice is independent of the total number of trials. Let $\pi_{k,1}(\xi)$ be 1 or 0 according as A or B is to be used. We shall show that for every N the procedure π^*_N determined by

$$(3)\qquad \begin{aligned} \pi^*_{k,1}(\xi) = \pi_1(\xi) &= 1 \quad \text{if} \quad \xi \geqq \tfrac{1}{2} \\ &= 0 \quad \text{if} \quad \xi < \tfrac{1}{2} \end{aligned}$$

is optimal. Before doing so, a few preliminary results will be established.

PROPERTIES OF R_N. For every N, $R_N(\xi)$ satisfies the following

(4) (a) $R_N(\xi)$ is a continuous function of ξ
(b) $R_N(\xi)$ is symmetric about $\xi = \frac{1}{2}$, i.e., $R_N(\xi) = R_N(1 - \xi)$
(c) $R_N(0) = R_N(1) = 0$
(d) $R^A_{N+1}(\xi) = R^B_{N+1}(1 - \xi)$.

These properties are immediate by inspection and induction.

LEMMA 2.1. *Let π^{AB}_N be the procedure which chooses A then B for the first two trials and then follows π_{N-2}. Let π^{BA}_N be the procedure which chooses B then A for the first two trials and then follows π_{N-2}. Then for every N and ξ*

$$R^{AB}_N(\xi) = R^{BA}_N(\xi). \tag{5}$$

PROOF. For the first two trials both π^{AB}_N and π^{BA}_N incur a risk of 1. Because of independence the order of receiving information is irrelevant and hence the same ξ_2 will be obtained by both. Since from the third trial on the two procedures follow π_{N-2}, the risk functions must be identical.

LEMMA 2.2. *Let ξ_A (ξ_B) be the posterior probability of H_A given ξ and an observation on $A(B)$. Then ξ_A, ξ_B are stochastically increasing functions of ξ,* i.e., *$P[\xi_A \geqq z|\xi]$, $P[\xi_B \geqq z|\xi]$ are increasing functions of ξ.*

PROOF. We prove this only for ξ_A, the proof for ξ_B being analogous. First note that the likelihood ratio p_1/p_2, for an observation on A, is stochastically larger under H_A than under H_B. This follows from

$$\begin{aligned} P[(p_1/p_2) \geqq r \,|H_A, A \text{ observed}] &= \int_{p_1/p_2 \geqq r} p_1 \, d\lambda \\ &\geqq r \int_{p_1/p_2 \geqq r} p_2 \, d\lambda = rP[(p_1/p_2) \geqq r \,|H_B, A \text{ observed}] \end{aligned} \tag{6}$$

and

$$\begin{aligned} P[(p_1/p_2) < r \,|H_A, A \text{ observed}] &= \int_{p_1/p_2 < r} p_1 \, d\lambda \\ &< r \int_{p_1/p_2 < r} p_2 \, d\lambda = rP[(p_1/p_2) < r \,|H_B, A \text{ observed}]. \end{aligned} \tag{7}$$

The desired result follows by taking $r > 1$ in (6) and $r < 1$ in (7). From this and the definition, (1), of ξ_A, follows

$$\begin{aligned} P[\xi_A \geqq z|H_A] &= P\{(p_1/p_2) \geqq [(1 - \xi)/\xi] \cdot [z/(1 - z)]|H_A, \text{A observed}\} \\ &> P\{(p_1/p_2) \geqq [(1 - \xi)/\xi] \cdot [z/(1 - z)] \mid H_B, A \text{ observed}\} \\ &= P[\xi_A \geqq z \mid H_B]. \end{aligned} \tag{8}$$

Since $(1 - \xi)/\xi$ is a decreasing function of ξ, $P[\xi_A \geqq z \mid H_A]$ and $P[\xi_A \geqq z \mid H_B]$ are non-decreasing functions of ξ. Now

$$P[\xi_A \geqq z \mid \xi] = \xi P[\xi_A \geqq z \mid H_A] + (1 - \xi)P[\xi_A \geqq z \mid H_B] \tag{9}$$

is a convex combination of non-decreasing functions of ξ, the first of which, by (8), is uniformly larger than the other. Hence as ξ increases so does $P[\xi_A \geqq z \mid \xi]$.

We are now ready to prove

THEOREM 2.1. *For every N the procedure π_N^*, which chooses A or B at stage i according as $\xi_{i-1} \geqq \frac{1}{2}$ or $\xi_{i-1} < \frac{1}{2}$, is optimal.*

PROOF. We shall show by induction that the first-choice functions (3) are optimal. Define the functions

$$\Delta_N(\xi) = R_N^B(\xi) - R_N^A(\xi). \qquad N = 1, 2, \cdots. \tag{10}$$

By virtue of the properties (4) a sufficient condition that π_N^* be optimal is that $\Delta_N(\xi)$ be strictly increasing in ξ. Specifically, (4) guarantees that for every N there exists $z_N \varepsilon (0, \frac{1}{2})$ such that

$$\xi \leqq z_N \Rightarrow \Delta_N(\xi) < 0, \xi \geqq 1 - z_N \Rightarrow \Delta_N(\xi) > 0,$$

while $\Delta_N(\frac{1}{2}) = 0$ for all N. Hence if $\Delta_N(\xi)$ is strictly increasing, π_N^* is optimal. The inductive hypothesis will then be

H_N : $\Delta_k(\xi)$ is strictly increasing in ξ for all $k \leqq N$. This is clearly true for $\Delta_1(\xi) = 2\xi - 1$. Let π_{N+1}^{AB}, π_{N+1}^{BA} be the procedures of Lemma 2.1 where the continuation after the first two trials is π_{N-1}^*. Then, define

$$\begin{aligned} \Delta_{N+1}^A(\xi) &= R_{N+1}^{AB}(\xi) - R_{N+1}^A(\xi) \\ \Delta_{N+1}^B(\xi) &= R_{N+1}^{BA}(\xi) - R_{N+1}^B(\xi) \end{aligned} \tag{11}$$

so that by Lemma 2.1,

$$\Delta_{N+1}(\xi) = \Delta_{N+1}^A(\xi) - \Delta_{N+1}^B(\xi). \tag{12}$$

Let $\delta_A(\xi)$ be the random variable whose expectation is $\Delta_{N+1}^A(\xi)$, i.e., $\delta_A(\xi)$ is the difference in the number of mistakes between π_{N+1}^{AB} and π_{N+1}^A. $\delta_B(\xi)$ is the analogous random variable for $\Delta_{N+1}^B(\xi)$. Since the first observation is the same for π_{N+1}^{AB}, π_{N+1}^A the difference in risk will depend only on the last N trials and on whether the posterior probability of H_A after the first observation is $\geqq \frac{1}{2}$ or $< \frac{1}{2}$. Thus

$$E[\delta_A(\xi) \mid \xi_A] = \pi_1^*(\xi_A)[R_N^B(\xi_A) - R_N^A(\xi_A)] \tag{13}$$

and, similarly

$$E[\delta_B(\xi) \mid \xi_B] = [1 - \pi_1^*(\xi_B)][R_N^A(\xi_B) - R_N^B(\xi_B)], \tag{14}$$

$\pi_1^*(\xi)$ being the first-choice function (3). Under H_N, (13) and (14) are increasing and decreasing respectively and the monotonicity is strict except, in the case of (13), where $\pi_1^*(\xi_A) = 0$, and in the case of (14), where $\pi_1^*(\xi_B) = 1$. Taking expectations in (13) and (14) we get by Lemma 2.2 and the fact that monotonicity is preserved for functions of stochastically ordered random variables, $\Delta_{N+1}^A(\xi)$ is strictly increasing except for the set of ξ's such that $P[\xi_A < \frac{1}{2}] = 1$ and $\Delta_{N+1}^B(\xi)$ is strictly decreasing except for the set of ξ's such that

$P[\xi_B \geqq \frac{1}{2}] = 1$. Since these sets are clearly disjoint, the difference (12) is strictly increasing. This proves H_{N+1} and hence the optimality of π_1^*.

It should be noted that the choice of A at $\xi = \frac{1}{2}$ is merely for the convenience of dealing with a fixed procedure. From the properties (4), of R_N, it can readily be seen that either choice at this point is as good as the other.

3. GTAB infinite case. Let π^* denote the extension of procedure π_N^* to the case where the number of observations to be made is infinite. Its risk $R(\xi)$ is easily seen to be

$$R(\xi) = \lim_{N\to\infty} R_N(\xi) \tag{15}$$

where $R_N(\xi)$ is the risk of π_N^*. Since π_N^* is optimal for every N, the only question concerning the optimality of π^* is finiteness of its risk function, and, because of symmetry, we need only show that under, say, H_A, the expected number of observations on B is finite.

Let $\xi = \xi_0, \xi_1, \cdots$ be the consecutive probabilities of H_A. Then under H_A and π^*, a mistake is made every time the event $[\xi_i < \frac{1}{2}]$ occurs and we want to show

$$\sum_{i=0}^{\infty} P[\xi_i < \tfrac{1}{2} \mid H_A, \pi^*, \xi] < \infty. \tag{16}$$

To facilitate the proof we transform ξ_i to the (equivalent) statistic

$$x_i = \log[\xi_i/(1 - \xi_i)]. \tag{17}$$

The procedure π^* in terms of the x-process becomes: choose A or B according as $x_i \geqq 0$ or $x_i < 0$. We wish now to show

$$\sum_{i=0}^{\infty} P[x_i < 0 \mid H_A, \pi^*, x] < \infty. \tag{18}$$

The property of the x-process which simplifies the proof is its transitions:

$$\begin{aligned} x_{i+1} &= x_i + \log[p_1(t)/p_2(t)] \text{ if } t \text{ is observed on } A \\ &= x_i + \log[p_2(t)/p_1(t)] \text{ if } t \text{ is observed on } B. \end{aligned} \tag{19}$$

Let $u_i = \log p_1(t_i)/p_2(t_i)$ if the ith observation is t_i on A and let $u_i = \log p_2(t_i)/p_1(t_i)$ if the ith observation is t_i on B. Let $S_n(x) = x + \sum_{i=1}^{n} u_i$, $n \geqq 1$, $S_0(x) = x$. Then (18) is equivalent to

$$\sum_{i=0}^{\infty} P[S_i(x) < 0 \mid H_A, \pi^*, x] < \infty. \tag{20}$$

Let $\phi_A(v)$ be the moment generating function of an observation u on A, $\phi_B(v)$ the same for an observation on B. Then, under H_A,

$$\begin{aligned} \phi_A(v) &= Ee^{vu} = \int [(p_1(t)/p_2(t))]^v p_1(t)\, d\lambda(t) \\ \phi_B(v) &= Ee^{vu} = \int [(p_2(t)/p_1(t))]^v p_2(t)\, d\lambda(t). \end{aligned} \tag{21}$$

Clearly,

$$\phi_A(-\tfrac{1}{2}) = \phi_B(-\tfrac{1}{2}) = \int [p_1(t)p_2(t)]^{\frac{1}{2}}\, d\lambda(t) = \rho < 1 \tag{22}$$

as long as p_1, p_2 are distinct. Now, consider the moment generating function of $S_n(x)$ at $v = -\frac{1}{2}$

$$\begin{aligned}\phi_n^x(-\tfrac{1}{2}) &= E \exp[-\tfrac{1}{2}S_n(x)] = E\{\exp[-\tfrac{1}{2}S_{n-1}(x)]E[\exp -\tfrac{1}{2}u_n \mid S_{n-1}(x)]\} \\ &= \rho\phi_{n-1}^x(-\tfrac{1}{2}) = e^{-x/2}\rho^n.\end{aligned} \tag{23}$$

Since $\exp[-\frac{1}{2}S_n(x)] > 1$ for $S_n(x) < 0$ we have

$$P[S_n(x) < 0 \mid H_A, \pi^*, x] \leqq \phi_n^x(-\tfrac{1}{2}) = e^{-x/2}\rho^n \tag{24}$$

which proves that the series (20) converges and thereby that π^* is optimal.

4. Computation of $R(\xi)$. It has already been noted in Section 3 that the transformation (17) has simplifying properties. These properties can be further exploited if the recursion formula (2) is expressed in terms of this transformation. From (19) we get

$$\begin{aligned}R_{N+1}(x) &= \min [R_{N+1}^A(x), R_{N+1}^B(x)] \\ R_{N+1}^A(x) &= 1 - \xi(x) + \int R_N(x + \log [p_1(t)/p_2(t)]) \\ &\qquad \cdot [p_1(t)\xi(x) + p_2(t)(1 - \xi(x))]\, d\lambda(t) \\ R_{N+1}^B(x) &= \xi(x) + \int R_N(x + \log [p_2(t)/p_1(t)]) \\ &\qquad \cdot [p_2(t)\xi(x) + p_1(t)(1 - \xi(x))]\, d\lambda(t)\end{aligned} \tag{25}$$

where we have used the same notation for risk functions, though here the domain is $[-\infty, +\infty]$, and $\xi(x) = e^x/e^x + 1$. A further simplification can be obtained by writing

$$\begin{aligned}&p_1(t)\xi(x) + p_2(t)(1 - \xi(x)) \\ &= [p_1(t)p_2(t)]^{\frac{1}{2}} \frac{\exp[x + \frac{1}{2}\log p_1(t)/p_2(t)] + \exp[-\frac{1}{2}\log p_1(t)/p_2(t)]}{e^x + 1} \\ &= \frac{[p_1(t)p_2(t)]^{\frac{1}{2}}}{e^{x/2} + e^{-x/2}} \{\exp \tfrac{1}{2}[x + \log p_1(t)/p_2(t)] + \exp - \tfrac{1}{2}[x + \log p_1(t)/p_2(t)]\}\end{aligned} \tag{26}$$

and

$$\begin{aligned}p_2(t)\xi(x) + p_1(t)[1 - \xi(x)] &= \frac{[p_1(t)p_2(t)]^{\frac{1}{2}}}{e^{x/2} + e^{-x/2}} \\ &\cdot \{\exp \tfrac{1}{2}[\log x + \log p_1(t)/p_2(t)] + \exp - \tfrac{1}{2}[x + \log p_1(t)/p_2(t)]\}.\end{aligned} \tag{27}$$

Multiplying $R_{N+1}(x)$ by $e^{x/2} + e^{-x/2}$, the expressions in brackets can be incorporated into the recursion property to get

$$
\begin{aligned}
f_{N+1}(x) &= \min [f^A_{N+1}(x), f^B_{N+1}(x)] \\
f^A_{N+1}(x) &= e^{-x/2} + \rho \int f_N(x + u(t))p(t)\, d\lambda(t) \qquad (28) \\
f^B_{N+1}(x) &= e^{x/2} + \rho \int f_N(x - u(t))p(t)\, d\lambda(t)
\end{aligned}
$$

where $f_{N+1}(x)$ is related to risk $R_{N+1}(x)$ by

$$f_{N+1}(x) = (e^{x/2} + e^{-x/2})R_{N+1}(x) \qquad (29)$$

and where

$$
\begin{aligned}
\rho &= \int [p_1(t)p_2(t)]^{\frac{1}{2}}\, d\lambda(t) < 1, \qquad (30) \\
\rho p(t) &= [p_1(t)p_2(t)]^{\frac{1}{2}}, \qquad u(t) = \log [p_1(t)/p_2(t)].
\end{aligned}
$$

The equation (28) can be used directly to prove optimality of π^*_N. Also, for the infinite case (28) becomes a functional equation which can be shown to have a unique, finite solution. We shall use (28) to get the risk function of π^* for the infinite case when the experiments are binomials chosen so that the likelihood ratios are simple integers. First, we introduce the following notation:

$p_1 = 1 - q_1 > p_2 = 1 - q_2$ are the probabilities of success,
$c_1 = \log p_1/p_2$, $c_2 = \log q_2/q_1$,
$p = (p_1p_2)^{\frac{1}{2}}$, $q = (q_1q_2)^{\frac{1}{2}}$,
$f(x)$ is the normalized risk function for an infinite number of trials.

In terms of the above, the functional equation to be solved becomes

$$
\begin{aligned}
f(x) = \min [&e^{-x/2} + pf(x + c_1) + qf(x - c_2), \\
&e^{x/2} + pf(x - c_1) + qf(x + c_2)]
\end{aligned} \qquad (31)
$$

since $f(x)$ is the normalized risk of π^* we must have

$$
\begin{aligned}
f(x) &= e^{x/2} + pf(x - c_1) + qf(x + c_2), && x \leqq 0 \\
&= e^{-x/2} + pf(x + c_1) + qf(x - c_2), && x \geqq 0
\end{aligned} \qquad (32)
$$

or, by symmetry,

$$f(x) = e^{-x/2} + pf(x + c_1) + qf(|x - c_2|), \qquad 0 \leqq x \leqq \infty. \qquad (33)$$

$f(x)$ has an interpretation in terms of the following random walk problem: for any x transitions are to

$$
\begin{aligned}
&x + c_1 \text{ with probability } p \\
&|x - c_2| \text{ with probability } q \qquad (34) \\
&\infty \text{ with probability } 1 - p - q.
\end{aligned}
$$

$e^{-x/2}$ is the cost incurred everytime x is visited and it is desired to find $E_x \sum_{i=0}^{\infty} e^{-x_i/2}$ where x is the initial state and the x_i's all succeeding states. Choose c_1, c_2 integers and consider only integral states x. Let Q_{xj} be the expected number of visits from x to j. Then

$$f(x) = \sum_{j=0}^{\infty} Q_{xj} e^{-j/2} = \sum_{j=0}^{\infty} Q_{xj} \alpha^j, \qquad \alpha = e^{-\frac{1}{2}}. \tag{35}$$

Let $\phi_x(u) = \sum_{j=0}^{\infty} Q_{xj} u^j$, so that $\phi_x(\alpha) = f(x)$. Let $\mathbf{P} = \{p_{ij}\}$ be the transition matrix of the random walk. Then from the relation $\mathbf{Q} = \{Q_{ij}\} = [\mathbf{I} - \mathbf{P}]^{-1}$ one gets (looking at $\mathbf{Q}(\mathbf{I} - \mathbf{P}) = \mathbf{I}$) that

$$(1 - pu^{c_1} - qu^{-c_2})\phi_x(u) = u^x - q\sum_{j=1}^{c_2} Q_{x,c_2-j}(u^{-j} - u^j), \quad |u| < 1. \tag{36}$$

α is a root on the left and we use l'Hospital's rule to get

$$\phi_x(\alpha) = \left[x\alpha^x + q\sum_{j=1}^{c_2} jQ_{x,c_2-j}(\alpha^{-j} + \alpha^j)\right](q_2c_2 - p_2c_1)^{-1}. \tag{37}$$

The problem now remains to evaluate Q_{xj}, $j = 0, 1, \cdots, c_2 - 1$, which can be done explicitly for small values of c_1, c_2 by evaluating (36) at the roots of $1 - pu^{c_1} - qu^{-c_2}$.

EXAMPLES.

(a) $c_2 = 1$ (α is the only root needed). From (36) we get $qQ_{x0} = \alpha^{x+1}/1 - \alpha^2$. Hence

$$\begin{aligned} \phi_x(\alpha) &= \left[x + \frac{1+\alpha^2}{1-\alpha^2}\right]\frac{\alpha^x}{q_2 - p_2 c_1} \\ &= \left[x + \frac{e+1}{e-1}\right]\frac{e^{-x/2}}{q_2 - p_2 c_1} = f(x), \qquad x \geqq 0. \end{aligned} \tag{38}$$

The risk of π^* in this case is

$$\begin{aligned} R(x) &= f(x)/(e^{x/2} + e^{-x/2}) \\ &= \{x + [(e+1)/(e-1)]\}/(q_2 - p_2c_1)(e^x + 1), \qquad x \geqq 0. \end{aligned} \tag{39}$$

(b) $c_2 = r$ (any positive real no.), $c_1 = kr$, k an integer and x is some integral multiple of r. Then, applying (38) one easily gets

$$f(x) = \left[x + \frac{c_2(e^{c_2}+1)}{e^{c_2}-1}\right]\frac{e^{-x/2}}{q_2 c_2 - p_2 c_1}, \qquad x \geqq 0, \tag{40}$$

and the risk for π^* is

$$\begin{aligned} R(x) &= \left[x + \frac{c_2(e^{c_2}+1)}{e^{c_2}-1}\right]/(q_2 c_2 - p_2 c_1)(e^x + 1), x \geqq 0, \\ &= \frac{q_2(c_2 + x) + q_1(c_2 - x)}{(q_2 - q_1)(q_2 c_2 - p_2 c_1)(e^x + 1)}, \qquad x \geqq 0. \end{aligned} \tag{41}$$

5. The condition (ii)″. When condition (ii)′ of Section 1 is violated, conditions (iv) and (iv)′ no longer lead to the same procedure and π_N^* will not necessarily be optimal for either. However, the approach used in Section 2 yields some results worth noting when condition (ii)″ holds. In this case we are still given a pair of singletons but they are no longer symmetric. Lemmas 2.1 and 2.2 still hold but not properties (4) (b) and (4) (d) (under, say, (iv)′). But, because (4) (a) and (4) (c) still hold, the monotonicity properties will also hold and therefore we can make the following statement: for every N, the optimal first-choice function $\pi_{N,1}(\xi)$ is given by

$$\begin{aligned} \pi_{N,1}(\xi) &= 1 \quad \text{if} \quad \xi \geqq \xi_N \\ &= 0 \quad \text{if} \quad \xi < \xi_N . \end{aligned} \tag{42}$$

6. Acknowledgment. I wish to express sincere and grateful thanks to Professor David Blackwell for suggesting the problem, for his enthusiastic interest, and for his invaluable guidance.

REFERENCES

[1] BRADT, R. N. JOHNSON, S. M. AND KARLIN, S. (1956). On sequential designs for maximizing the sum of n observations. *Ann. Math. Statist.* **27** 1060–1074.

[2] BRADT, R. N. AND KARLIN, S. (1956). On the design and comparison of certain dichotomous experiments. *Ann. Math. Statist.* **27** 390–409.

[3] ROBBINS, H. (1952). Some aspects of the sequential design of experiments. *Bull. Amer. Math. Soc.* **58** 527–535.