A SEQUENTIAL DESIGN FOR THE TWO ARMED BANDIT

By Walter Vogel[1]

Universität Tübingen and University of Chicago[2]

1. Introduction. Let the two random variables (r.v.) X and Y, with $E(X) = p$ and $E(Y) = q$, describe the outcomes of two experiments, Ex I and Ex II. An experimenter, who does not know the values of p and q, has to perform a sequence of experiments, and at each step he may choose between Ex I and Ex II. He has to stop after n steps, and he wishes to maximise the sum of all outcomes. His decision between Ex I and Ex II at the kth step will depend on the corresponding decisions at prior steps and on the outcomes of these prior experiments. We call a plan, which fixes his sequence of decisions according to his previous knowledge, a strategy.

Robbins [6] shows that it is easy to find a strategy so that the arithmetic mean of n outcomes tends $(n \rightarrow \infty)$ towards max (p, q) with probability 1. Bradt, Johnson and Karlin [3] try to find a best strategy for fixed n rather than asymptotically. They assume known *a priori* distributions for the values of p and q. For other approaches see Robbins [7] Isbell [5], Bellman [2] and Vogel [8].

The purpose of this paper is to describe a class of strategies, which results from the following kind of restriction. In the first $2k$ steps we perform each of Ex I and Ex II k times. Then the rest of the $n - 2k$ steps are made either with Ex I alone or with Ex II alone. The decision whether to continue with Ex I or with Ex II will be made with the help of a sequential probability ratio test for double dichotomies. Therefore k is a r.v. that will be denoted by K when appropriate.

Strategies of this kind are not exceptionally good ones (in the sense of the loss-function defined in Section 3). But when a strategy is applied in practice it may be found economic to do only one sort of experiment for most of the steps. Perhaps the equipment of the other sort of experiment can be used for other purposes; perhaps the shift from one experiment to the other is costly. For such reasons it may be quite natural to use only those strategies described above. Another justification for treating this class of strategies are the results in [8], for which the Theorems 2 and 3 of this paper are needed.

Section 2 contains some auxiliary material. Except for Theorem 1, which we give in a slightly more general form than needed for the rest of this paper, nothing here is new, but we found it convenient to summarize some definitions and easy-to-prove formulas in one section.

The loss-function and an approximation to the loss-function will be derived

Received January 19, 1959; revised January 18, 1960.

[1] Research carried out in part at the Statistical Research Center, University of Chicago, under the sponsorship of the Statistics Branch, Office of Naval Research. Reproduction in whole or in part is permitted for any purpose of the United States Government.

[2] Present address: Mathematisches Institut der Universität Tübingen, Germany.

in Section 3. Section 4 is devoted to a minimax theorem for the approximate loss-function. In Section 5 we give some results for $n \to \infty$. It is assumed in Sections 2–5 that the r.v's X and Y are binomially distributed. In Section 6 we consider a more general case.

2. Some remarks about sequential plans. Let Z_i ($i = 1, 2, 3 \cdots$) be a sequence of r.v.'s with $E(Z_i) = m_i$, and let $|Z_i| \leqq A < \infty$. We make no assumptions about the dependence of the Z_i. Let $N > 0$ be an integer-valued r.v. with $E(N) < \infty$. We assume that $E(Z_i \mid N = n) = m_i$ for $i > n$.

THEOREM 1: *From the assumptions made above it follows that*

$$\left(\sum_{i=1}^{N} Z_i\right) = E\left(\sum_{i=1}^{N} m_i\right).$$

PROOF: It is obviously sufficient to assume $m_i = 0$ and to prove $E(\sum_{i=1}^{N} Z_i) = 0$. From $E(N) < \infty$ it follows that $0 \leqq nP(N \geqq n) \leqq \sum_{k=n}^{\infty} kP(N = k) \to 0$, $(n \to \infty)$. Therefore

$$(2.1) \quad \left| E\left(\sum_{i=1}^{n} Z_i \mid N \geqq n\right) P(N \geqq n) \right| \leqq AnP(N \geqq n) \to 0 \quad \text{and}$$

$$\left| E\left(\sum_{i=1}^{N} Z_i \mid N \geqq n\right) P(N \geqq n) \right| \leqq AE(N \mid N \geqq n) P(N \geqq n)$$

$$= A \sum_{k=n}^{\infty} kP(N = k) \to 0.$$

From

$$E\left(\sum_{i=1}^{N} Z_i\right) = E\left(\sum_{i=1}^{N} Z_i \mid N < n\right) P(N < n) + E\left(\sum_{i=1}^{N} Z_i \mid N \geqq n\right) P(N \geqq n)$$

and the last relation it follows that

$$(2.2) \qquad E\left(\sum_{i=1}^{N} Z_i \mid N < n\right) P(N < n) \to E\left(\sum_{i=1}^{N} Z_i\right).$$

As $E(Z_i \mid N = k) = 0$ for $i > k$ we have

$$E\left(\sum_{i=1}^{n} Z_i \mid N < n\right)$$

$$(2.3) \qquad = E\left(\sum_{i=1}^{N} Z_i \mid N < n\right) + E\left(\sum_{i=N+1}^{n} Z_i \mid N < n\right)$$

$$= E\left(\sum_{i=1}^{N} Z_i \mid N < n\right).$$

Now

$$0 = E\left(\sum_{i=1}^{n} Z_i\right)$$

$$= E\left(\sum_{i=1}^{n} Z_i \mid N < n\right) P(N < n) + E\left(\sum_{i=1}^{n} Z_i \mid N \geqq n\right) P(N \geqq n).$$

(2.1) shows that the last term converges to zero. The other term on the right side converges to $E(\sum_{i=1}^{N} Z_i)$ because of (2.2) and (2.3). This proves the theorem.

If the Z_i are independent and identically distributed and if $\{N = n\}$ is defined on the first n Z_i the theorem reduces to the well known formula

$$E\left(\sum_{i=1}^{N} Z_i\right) = E(N)E(Z_i)$$

We describe now a random walk with absorbing barriers. Let X_i and Y_i $(i = 1, 2, 3, \cdots \mu)$ be independent r.v.'s with $P(X_i = 1) = p = 1 - P(X_i = 0)$, $P(Y_i = 1) = q = 1 - P(Y_i = 0)$ and $0 < p, q < 1$. Let further $Z_i = X_i - Y_i$ and $U_k = \sum_{i=1}^{k} Z_i$. We define some events (α is a positive integer).

$$A_{1,k} = \{-\alpha < U_i < +\alpha \text{ for } i < k \text{ and } U_k = +\alpha\}, \quad (k = 1, 2, \cdots \mu)$$

$$A_{2,k} = \{-\alpha < U_i < +\alpha \text{ for } i < k \text{ and } U_k = -\alpha\}, \quad (k = 1, 2, \cdots \mu)$$

$$B_\nu = \{-\alpha < U_i < +\alpha \text{ for } i \leqq \mu \text{ and } U_\mu = \nu\}$$

$$(\nu = -\alpha + 1, -\alpha + 2, \cdots \alpha - 2, \alpha - 1)$$

$$A_1 = \sum_{k=1}^{\mu} A_{1,k}, \qquad A_2 = \sum_{k=1}^{\mu} A_{2,k}, \qquad B = \sum_{-\alpha<\nu<\alpha} B_\nu .$$

We have $P(A_1) + P(A_2) + P(B) = 1$.

Let $p(1 - q) = r$, $q(1 - p) = s$, $1 - r - s = t$ and $u = r/s$. Here r, s and t are the probabilities that $Z_i = 1, -1$ or 0.

We will prove

LEMMA 1:

$$P(A_{1,k} \mid A_{1,k} + A_{2,k}) = u^\alpha/(u^\alpha + 1) = \gamma$$

$$P(A_{2,k} \mid A_{1,k} + A_{2,k}) = 1/(u^\alpha + 1) = 1 - \gamma.$$

PROOF: $P(A_{1,k}) = \sum_\tau c_{k,\tau} r^\rho s^\sigma t^\tau$, where $c_{k,\tau}$ is the number of admissible paths which the point (i, U_i) describes in the plane before reaching (k, α). The paths have ρ steps up, σ steps down and τ horizontal steps with the conditions $\rho + \sigma + \tau = k$ and $\rho - \sigma = \alpha$. The summation-range for τ is $0 \leqq \tau \leqq k - \alpha$. By introducing k and α instead of ρ and σ we get

$$P(A_{1,k}) = (r/s)^{\alpha/2}(rs)^{k/2} \sum_\tau c_{k,\tau} t^\tau (rs)^{-\tau/2}$$

and likewise

$$P(A_{2,k}) = (r/s)^{-\alpha/2}(rs)^{k/2} \sum_\tau c_{k,\tau} t^\tau (rs)^{-\tau/2}.$$

Taking account of $P(A_{1,k} + A_{2,k}) = P(A_{1,k}) + P(A_{2,k})$ the lemma follows immediately.

From Lemma 1 follows

$$
\begin{aligned}
P(A_1) &= \gamma \sum_{k=1}^{\mu} P(A_{1,k} + A_{2,k}) = (1 - P(B))u^{\alpha}/(u^{\alpha} + 1) \\
P(A_2) &= (1 - \gamma) \sum_{k=1}^{\mu} P(A_{1,k} + A_{2,k}) = (1 - P(B))/(u^{\alpha} + 1).
\end{aligned} \tag{2.4}
$$

We now define an integral valued r.v., K, by $\{K = k\} = A_{1,k} + A_{2,k}$ for $k < \mu$ and $\{K = \mu\} = A_{1,\mu} + A_{2,\mu} + B$. An application of Theorem 1 gives

$$
\begin{aligned}
&E\left(\sum_{i=1}^{K} (X_i + Y_i)\right) = (p + q)E(K) \quad \text{and} \\
&\qquad E(U_K) = E\left(\sum_{i=1}^{K} (X_i - Y_i)\right) = (p - q)E(K).
\end{aligned} \tag{2.5}
$$

From (2.4) follows

$$
\begin{aligned}
E(U_K) &= \alpha P(A_1) - \alpha P(A_2) + \sum_{-\alpha < v < \alpha} v P(B_v) \\
&= \alpha((u^{\alpha} - 1)/(u^{\alpha} + 1))(1 - P(B)) + \sum_{-\alpha < v < \alpha} v P(B_v).
\end{aligned}
$$

As $P(B) \to 0$ for $\mu \to \infty$, we have $E(K) = (p - q)^{-1} E(U_K) \to \alpha/(p - q)((u^{\alpha} - 1)/(u^{\alpha} + 1))$. Obviously $E(K)$ is a monotone increasing function of μ so that

$$
E(K) \leqq \frac{\alpha}{(p - q)} \left(\frac{u^{\alpha} - 1}{u^{\alpha} + 1}\right). \tag{2.6}
$$

In Section 5 we are interested in sequences of such random walks as described above by α, μ, U_k and K. In order to define a sequence of random walks let $n = 2\mu$ be an even number and let $X_i^{(n)}$ and $Y_i^{(n)}$ $(i = 1, 2, 3 \cdots \mu)$ be independent r.v.'s with $P(X_i^{(n)} = 1) = p_n = 1 - P(X_i^{(n)} = 0)$ and $P(Y_i^{(n)} - 1) = q_n = 1 - P(Y_i^{(n)} = 0)$. We assume $p_n \geqq q_n$, $p_n - q_n = mn^{-\frac{1}{2}} + o(n^{-\frac{1}{2}})$, p_n, $q_n \to p$ and $0 < p < 1$. Let α_n be an integer such that

$$
\alpha_n = an^{\frac{1}{2}} + o(n^{\frac{1}{2}}), \qquad a > 0.
$$

Putting $p_n(1 - q_n)/q_n(1 - p_n) = u_n$ we get

$$
u_n^{\alpha_n} \to v^a \qquad (\text{with } v = \exp m/p(1 - p)). \tag{2.7}
$$

PROOF OF (2.7):

$$
\begin{aligned}
\ln u_n^{\alpha_n} &= \alpha_n \ln\left(1 + \frac{p_n - q_n}{q_n(1 - p_n)}\right) \\
&= (an^{\frac{1}{2}} + o(n^{\frac{1}{2}})) \ln\left(1 + \frac{mn^{-\frac{1}{2}}}{p(1 - p)} + o(n^{-\frac{1}{2}})\right) \to am/p(1 - p).
\end{aligned}
$$

Formula (2.7) will be useful in Section 5.

Let the r.v.'s $K^{(n)}$ and $U_k^{(n)}$ be defined correspondingly to K and U_k. We want to evaluate the limit (for $n \to \infty$) of $n^{-1}E(K^{(n)})$. In the following argument we do not give a proof but proceed heuristically. "Let the reader who has never used this sort of reasoning exhibit the first counter example" (see [4] p. 395).

The point $(k, n^{-\frac{1}{2}}U_k^{(n)})$ describes a random walk in the plane. Absorbing barriers are $n^{-\frac{1}{2}}U_k^{(n)} = \pm a + o(1)$ and $k = n/2 = \mu$. $K^{(n)}$ is the "time" taken to reach the boundaries. We will now construct a suitable Wiener-process and compute $E(T)$ where T is the time taken to reach the boundaries in this Wiener-Process. Then we conclude heuristically that

$$(2.8)\qquad n^{-1}E(K^{(n)}) \to E(T).$$

We have $E(n^{-\frac{1}{2}}U_k^{(n)}) = n^{-\frac{1}{2}}kE(X_i^{(n)} - Y_i^{(n)}) = k\cdot m\cdot n^{-1} + o(n^{-1})$, and Var $(n^{-\frac{1}{2}}U_k^{(n)}) = n^{-1}k$ Var $(X_i^{(n)} - Y_i^{(n)}) \approx 2kp(1-p)n^{-1}$. Now let $n \to \infty$ and $k \to \infty$ so that $k/n = t$ is fixed. Then $n^{-\frac{1}{2}}U_k^{(n)}$ is in the limit normally distributed with mean tm and variance $2tp(1-p)$. We will therefore approximate $n^{-\frac{1}{2}}U_k^{(n)}$ by a Wiener-process V_t with absorbing barriers given by $V_t = \pm a$ and by $t = \frac{1}{2}$ and with mean and variance as stated above. The formulas for this simple kind of process are well known (see [1] p. 47). Let $f(x, t) =$

$$(4\pi p(1-p)t)^{-\frac{1}{2}} \exp - \frac{2mx - m^2t}{4p(1-p)} \sum_{s=-\infty}^{\infty} (-1)^s \exp - \frac{(x-2as)^2}{4p(1-p)t};$$

then $f(x, t)$ satisfies the diffusion equation

$$\frac{\partial f}{\partial t} + m\frac{\partial f}{\partial x} = p(1-p)\frac{\partial^2 f}{\partial x^2}$$

and the boundary-conditions $f(\pm a, t) = 0$. For the time T taken to reach the boundaries $x = \pm a$ the probability density is

$$g(t) = -\frac{\partial}{\partial t}\int_{-a}^{+a} f(x, t)\, dx$$

(see [1] p. 48) so that $E(T) = \int_0^{\frac{1}{2}} tg(t)\, dt + \int_{-a}^{+a} \frac{1}{2}f(x, \frac{1}{2})\, dx$. The first term is related to absorption on the boundaries $x = \pm a$, and the second term takes into account absorption on the boundary $t = \frac{1}{2}$. Integration by parts gives

$$(2.9)\qquad E(T) = \int_{-a}^{+a}\int_0^{\frac{1}{2}} f(x, t)\, dt\, dx.$$

(2.9) gives an expression for the limit in (2.8).

By the same heuristic argument we can get an expression for the limit of the probability of the event

$$(2.10)\qquad B^{(n)} = \{-\alpha_n < U_k^{(n)} < +\alpha_n,\ k \leqq n/2\}.$$

Let $\alpha_n = an^{\frac{1}{2}} + o(n^{\frac{1}{2}})$, $p_n - q_n = mn^{-\frac{1}{2}} + o(n^{-\frac{1}{2}})$, $a > 0$, $m > 0$, $p_n, q_n \to p$

and $0 < p < 1$. It then follows that

$$P(B^{(n)}) \to \int_{-a}^{+a} f(x, \tfrac{1}{2})\, dx. \tag{2.11}$$

3. The loss-function. We come now back to the problem stated in Section 1. Let $P(X = 1) = 1 - P(X = 0) = p$ and $P(Y = 1) = p = 1 - P(Y = 0)$ and let $n = 2\mu$ be an even integer. X and Y are the two r.v.'s between which the experimenter has to choose at each step and n is the total number of steps, fixed in advance. The strategy runs as follows: Begin sequentially with pairs of Ex I and Ex II (i.e. of X and Y) until a decision is reached to end this part of the strategy and to continue with Ex I or Ex II alone. Thus, observe X_1, Y_1 in the first two steps, and then decide either to observe another pair or to continue entirely with Ex I, or to continue entirely with Ex II. While pairs are still being observed we may describe the first $2k$ steps by X_1, Y_1; X_2, Y_2; $\cdots$ X_k, Y_k (all assumed to be independent). The decision at this point is based upon $U_k = \sum_{i=1}^{k} (X_i - Y_i)$ and an integer $\alpha > 0$. If $-\alpha < U_k < +\alpha$, another pair is observed; if $U_k \geqq +\alpha (\leqq -\alpha)$ we stop observing pairs and use only Ex I (Ex II) for the rest of the n steps. Let K be the random number of observed pairs $(0 < K \leqq n/2)$. α, μ $(n = 2\mu)$ and the r.v.'s U_k and K form a lay-out as described in Section 2 and we may use formulas (2.4)–(2.6).

The expected sum for all n steps, say W, is

$$W = E\left(\sum_{i=1}^{K} (X_i + Y_i)\right) + \gamma E\left(\sum_{i=2K+1}^{n} X_i\right) + (1 - \gamma) E\left(\sum_{i=2K+1}^{n} Y_i\right).$$

Here γ is defined as in Lemma 1, Section 2. The first term is related to the part where pairwise observations are made, the second and third terms stem from the possibilities to continue with Ex I alone or with Ex II alone.

Using Theorem 1 and Lemma 1 we get

$$\begin{aligned} W = (p + q)E(K) + p \frac{r^\alpha}{r^\alpha + s^\alpha} (n - 2E(K)) \\ + q \frac{1}{r^\alpha + s^\alpha} (n - 2E(K)). \end{aligned} \tag{3.1}$$

where $r = p(1 - q)$ and $s = q(1 - p)$.

The best possible expected outcome of the whole sum is $n \max (p, q)$. We define the loss-function $L_n = L(\alpha, p, q)$ as $L_n = n \max (p, q) - W$. Let $\sigma = \max (p, q)$ and $\tau = \min (p, q)$, then

$$L_n = \frac{n(\sigma - \tau)}{u^\alpha + 1} + (\sigma - \tau)\left(\frac{u^\alpha - 1}{u^\alpha + 1}\right) E(K),$$

with $u = \sigma(1 - \tau)/\tau(1 - \sigma)$. We remark that the loss-function is symmetric in p and q, i.e. that $L(\alpha, \sigma, \tau) = L(\alpha, \tau, \sigma)$. A strategy simply consists in the choice of an integer $\alpha(0 < \alpha \leqq n/2)$.

Suppose, now, that a fixed n has been given and that we intend to use a strategy, i.e. we want to choose an α. Our intention is to minimize L_n. The question of which α to choose cannot be answered unambiguously. When some knowledge about p and q in the form of an *a priori* distribution is given, we can try to compute the expectation of L_n with respect to this *a priori* distribution. This will be a function of α only. As α is an integer between 0 and $n/2$ there exists at least one α, which gives a minimum. The actual computation will be difficult, because no exact formula for $E(K)$ is available. We therefore use an approximation. Let

$$M_n = M(\alpha, \sigma, \tau) = \frac{n(\sigma - \tau)}{u^\alpha + 1} + \alpha \left(\frac{u^\alpha - 1}{u^\alpha + 1}\right)^2. \tag{3.2}$$

From (2.6) follows $M_n \geqq L_n$ and for $n \to \infty$ (but α fixed) we have $L_n - M_n \to 0$. As long as α is small compared to n, we will use M_n as an approximation of L_n.

We illustrate the use of M_n by an example. Let (p, q) be either $(0.6, 0.4)$ or $(0.4, 0.6)$ and $n = 100$. Then $\sigma - \tau = 0.2$ and $\mu = 2.25$. An easy computation shows, that $\alpha = 3$ is the only integer that makes M_{100} a minimum. Since α is small compared to n, the approximation is justified.

4. A minimax theorem for the approximate loss-function. In this section we are concerned only with the approximate loss-function M_n, and we regard α not as an integer but as a continuous variable $(0 < \alpha \leqq n/2)$.

THEOREM 2: *For $n \geqq 4$ we have*

$$\min_{\alpha} \max_{\sigma,\tau} M(\alpha, \sigma, \tau) = \max_{\sigma,\tau} \min_{\alpha} M(\alpha, \sigma, \tau)$$

We will prove the theorem by showing

$$M(\alpha_n, \sigma, \tau) \leqq M(\alpha_n, \sigma_n, \tau_n) \leqq M(\alpha, \sigma_n, \tau_n). \tag{4.1}$$

THEOREM 3: *The asymptotic behavior of α_n, σ_n and τ_n is given by $\alpha_n = an^{\frac{1}{2}} + o(n^{\frac{1}{2}})$ with $a = 0.292 \cdots$, $\sigma_n - \tau_n = mn^{-\frac{1}{2}} + o(n^{-\frac{1}{2}})$ with $m = 1.89 \cdots$, $\sigma_n, \tau_n \to \frac{1}{2}$ and $u_n^{\alpha_n} \to 9.06 \cdots$.*

We prove Theorems 2 and 3 together and proceed in several steps.

(i) M_n is monotone increasing in $(\sigma - \tau)$ and we compute $\max(\sigma - \tau)$ under the condition $u =$ constant. We find $\max(\sigma - \tau) = (u^{\frac{1}{2}} - 1)/(u^{\frac{1}{2}} + 1)$ and $\sigma = 1 - \tau = u^{\frac{1}{2}}/(u^{\frac{1}{2}} + 1)$. Then (4.1) is equivalent to

$$M(\alpha_n, u) \leqq M(\alpha_n, u_n) \leqq M(\alpha, u_n),$$

where

$$M(\alpha, u) = \frac{n}{u^\alpha + 1}\left(\frac{u^{\frac{1}{2}} - 1}{u^{\frac{1}{2}} + 1}\right) + \alpha\left(\frac{u^\alpha - 1}{u^\alpha + 1}\right)^2.$$

(ii) The saddle-point of $M(\alpha, u)$ can be obtained by setting

$$\frac{\partial M}{\partial \alpha} = \frac{\partial M}{\partial u} = 0.$$

We will show that these equations have a common solution, α_n, u_n, that (under (iii) and (iv)) α_n gives a minimum for $M(\alpha, u_n)$, and that u_n gives a maximum for $M(\alpha_n, u)$.

Setting $\frac{\partial M}{\partial \alpha} = 0$ gives (after division by $(u^\alpha + 1)^{-2} u^\alpha \ln u$)

$$\frac{(u^\alpha - 1)^2}{u^\alpha \ln u} - n\,\frac{u^{\frac{1}{2}} - 1}{u^{\frac{1}{2}} + 1} + 4\alpha\,\frac{u^\alpha - 1}{u^\alpha + 1} = 0. \tag{4.2}$$

Setting $\frac{\partial M}{\partial u} = 0$ gives (after division by $\alpha u^{\alpha-1}(u^\alpha + 1)^{-2}$)

$$n\,\frac{u^{\frac{1}{2}}(u^\alpha + 1)}{\alpha u^\alpha (u^{\frac{1}{2}} + 1)^2} - n\,\frac{u^{\frac{1}{2}} - 1}{u^{\frac{1}{2}} + 1} + 4\alpha\,\frac{u^\alpha - 1}{u^\alpha + 1} = 0. \tag{4.3}$$

From (4.2) and (4.3) it follows that

$$n((u^{\frac{1}{2}} - 1)/(u^{\frac{1}{2}} + 1)) = (\alpha(u - 1)(u^\alpha - 1)^2)/(u^{\frac{1}{2}}(u^\alpha + 1) \ln u).$$

This, introduced in (4.2), gives

$$\frac{u - 1}{u^{\frac{1}{2}} \ln u} = \frac{u^\alpha + 1}{u^\alpha \ln u^\alpha} + \frac{4}{u^\alpha - 1}. \tag{4.4}$$

We now rewrite (4.4) and (4.2), but set $u^\alpha = x$, thus

$$\frac{u - 1}{u^{\frac{1}{2}} \ln u} = \frac{x + 1}{x \ln x} + \frac{4}{x - 1}, \tag{4.5}$$

$$n\left(\frac{u^{\frac{1}{2}} - 1}{u^{\frac{1}{2}} + 1}\right) \ln u = \frac{(x - 1)^2}{x} + 4\,\frac{x - 1}{x + 1} \ln x. \tag{4.6}$$

A simultaneous solution of (4.5) and (4.6) would yield a simultaneous solution of (4.2) and (4.3), which is desired. Now, the left side of (4.5) is a monotone increasing function of u. As u ranges from 1 to ∞ the left side ranges from 1 to ∞ also. The right side of (4.5) is a monotone decreasing function of x. Let c be the unique solution of

$$1 = ((x + 1)/(x \ln x)) + (4/(x - 1)), \qquad c = 9.06 \cdots \tag{4.7}$$

so that (4.5) defines a function, $x = A(u)$ say, which is monotone decreasing from c to 1.

The same kind of argument shows that (4.6) defines a function, $x = B(u)$ say, which is monotone increasing from 1 to ∞. So the two functions have exactly one point in common. Its abscissa is u_n and its ordinate, $x_n = u_n^{\alpha_n}$, gives α_n. This shows that there is exactly one pair, u_n, α_n, which satisfies $\frac{\partial M}{\partial \alpha} = \frac{\partial M}{\partial u} = 0$. It is easily seen that $0 < u_n < \infty$ and that $0 < \alpha_n < n/2$.

(iii) We show that $M(\alpha, u_n)$ has a minimum for $\alpha = \alpha_n$. The left side of (4.2), times a positive factor, is just $\partial M/\partial \alpha$. This left side is monotone increasing, therefore it is negative for $\alpha < \alpha_n$ and positive for $\alpha > \alpha_n$. Then this is true for $\frac{\partial M}{\partial \alpha}$ also and M has the desired minimum.

(iv) To show that $M(\alpha_n, u)$ has a maximum for $u = u_n$ is a bit more difficult, and, before doing so, we investigate the behavior of α_n as a function of n. The function $x = B(u)$ depends on n and we will write it as $B_n(u)$. As is seen from (4.6), the following holds for all $u > 1$: If $n_1 > n_2$ then $B_{n_1}(u) > B_{n_2}(u)$ and if $n \to \infty$ then $B_n(u) \to \infty$. It is therefore clear that, if $n \to \infty$, then u_n is monotone decreasing towards 1 and $A(u_n)$ is monotone increasing towards c. Now $A(u_n) = x_n = u_n^{\alpha_n}$ and

$$\alpha_n = \ln x_n / \ln u_n \tag{4.8}$$

so that α_n is monotone increasing in n.

For $n \to \infty$ and $u \to 1$ the left side of (4.6) may be approximated by

$$(n \ln^2 u)/4$$

and the right side (as $x \to c$) by a constant. From this we see that $\ln u_n \approx bn^{-\frac{1}{2}}$ and from (4.8) that $\alpha_n \approx an^{\frac{1}{2}}$, where b and a are constants. A more careful examination gives $\ln u_n = bn^{-\frac{1}{2}} + o(n^{-\frac{1}{2}})$, $\sigma_n - \tau_n = mn^{-\frac{1}{2}} + o(n^{-\frac{1}{2}})$ and $\alpha_n = an^{\frac{1}{2}} + o(n^{\frac{1}{2}})$.

A numerical computation shows that $\alpha_4 > \frac{1}{2}$. As $n \geqq 4$ was assumed, we may use $\alpha_n > \frac{1}{2}$ in the following proof that $M(\alpha_n, u)$ has a maximum for $u = u_n$.

$\frac{\partial M}{\partial u}$ is, aside from the positive factor $\alpha u^{\alpha-1}(u^\alpha + 1)^{-2}(u^{\frac{1}{2}} - 1)(u^{\frac{1}{2}} + 1)^{-1}$, given by

$$n \frac{(u^\alpha + 1)n^{\frac{1}{2}}}{\alpha u^\alpha (u - 1)} - n + 4\alpha \left(\frac{u^\alpha - 1}{u^\alpha + 1}\right)\left(\frac{u^{\frac{1}{2}} + 1}{u^{\frac{1}{2}} - 1}\right).$$

The first term is monotone decreasing in u, the second term is constant, and, for $\alpha > \frac{1}{2}$, the third term is monotone decreasing also. But then $\frac{\partial M}{\partial u}$ goes from positive to negative values at $u = u_n$, which is what we wished to show.

(v) The next step is the determination of the constant in $n^{-\frac{1}{2}}\alpha_n \to a$. From (4.7) we find $c = x_\infty = 9.06 \cdots$. Then (4.6) gives, for $n \to \infty$, $\frac{1}{4} n \ln^2 u_n \approx (c - 1)^2 c^{-1} + 4(c - 1)(c + 1)^{-1} \ln c = 14.21 \cdots$ or $\ln u_n \approx 7.54\ n^{-\frac{1}{2}}$ and by (4.8) we have

$$\alpha_n n^{-\frac{1}{2}} \approx (\ln c)/7.54 = 0.292 \cdots = a.$$

In a similar way it can be shown that $(\sigma_n - \tau_n)n^{\frac{1}{2}} \to m = 1.89 \cdots$. Furthermore, we have $x_n = u_n^{\alpha_n} \to c = 9.06$, and from $\sigma_n = 1 - \tau_n$ and $\sigma_n - \tau_n \to 0$ it follows that $\sigma_n, \tau_n \to \frac{1}{2}$. Thus Theorems 2 and 3 are proved.

In order to compute $\alpha_n = \alpha(n)$, we first get $x = A(u)$ from (4.5). In formula (4.6), or

$$n = \frac{(x-1)^2}{x} + 4\left(\frac{x-1}{x+1}\right)\ln x \Big/ \frac{u^{1/2}-1}{u^{1/2}+1}\ln u,$$

we put $x = A(u)$ and get $n = n(u)$. Formula (4.8), or $\alpha_n = (\ln A(u))/\ln u$, gives $\alpha = \alpha(u)$. So we have gotten a representation of $\alpha(n)$ in terms of the parameter u. This allows us to compute $\alpha(n)n^{-1/2}$. We find $n^{-1/2}\alpha(n) = 0.292$ for $n \geqq 100$, and for $n = 70$ the third decimal is influenced for the first time by two units. The whole computation was made with a slide rule. $n^{-1/2}\alpha_n$ changes only slowly with n. We therefore propose to compute the sequential plan (for $n \geqq 100$) by the simple formula

$$\alpha_n = [0.292\, n^{1/2}].$$

M_n is a good approximation only as long as $P(B^{(n)})$ is small (For the definition of $B^{(n)}$ see (2.10)). The asymptotic behavior of α_n, σ_n and τ_n is such that we can apply (2.11), which shows that $P(B^{(n)})$ does not vanish in the limit. Therefore M_n should not be used asymptotically if both the experimenter and nature use the strategies derived in this section.

5. The loss-function as $n \to \infty$. In the previous section we used the approximation M because no exact formula for the truncated sequential procedure is available. If the number of steps tends towards infinity the random walk will become a Wiener-process and we can use the results from the end of Section 2. In this way, we will get another approximation for L_n, valid when n is large. Let

$$\alpha_n = an^{1/2} + o(n^{1/2}), \qquad a > 0,$$
$$\sigma_n - \tau_n = mn^{-1/2} + o(n^{-1/2}), \qquad \sigma_n, \tau_n \to p, \qquad 0 < p < 1.$$

(α_n, σ_n and τ_n have now different meanings than those in Section 4).

THEOREM 4: *$\lim_{n\to\infty} n^{-1/2}L_n(\alpha_n, \sigma_n, \tau_n) = L_\infty(a, m, p)$, where L_∞ is given by Formula* (5.1) *below.*

PROOF:

$$n^{-1/2}L_n = \frac{(\sigma_n - \tau_n)n^{1/2}}{u_n^{\alpha_n}+1} + (\sigma_n - \tau_n)n^{1/2}\left(\frac{u_n^{\alpha_n}-1}{u_n^{\alpha_n}+1}\right)\cdot\frac{E(K^{(n)})}{n}.$$

We use (2.7)–(2.9) and get

$$L_\infty = m(v^a+1)^{-1} + m(v^a-1)(v^a+1)^{-1}E(T), \tag{5.1}$$

where $v = \exp\,(m/p(1-p))$ and $E(T)$ is given by formula (2.9).

L_∞ has a cumbersome formula because $E(T)$ has one, and it would be worthwhile to try to find a value a_0 which gives a saddle-point, i.e. for which

$$L_\infty(a_0, m, p) \leqq L_\infty(a_0, m_0, p_0) \leqq L_\infty(a, m_0, p_0).$$

Then $\alpha = a_0 n^{\frac{1}{2}}$ could be used as an approximation to a minimax strategy. We can make only one step in this direction, namely we can prove that L_∞ is monotone increasing in m for fixed v and a.

The first term of L_∞, $m(v^a + 1)^{-1}$, surely is monotone increasing and we will show that $mE(T)$ is also. Putting $2mt = \lambda$, we get

$$mE(T) = m \int_{-a}^{+a} \int_0^{\frac{1}{2}} f(x, t) dt\, dx = \int_{-a}^{+a} \int_0^m \left(\pi \frac{p(1-p)\lambda}{2m}\right)^{-\frac{1}{2}} \exp$$

$$- \frac{4x - \lambda}{8p(1-p)/m} \sum_{s=-\infty}^{+\infty} (-1)^s \exp - \frac{(x - 2as)^2}{2\lambda p(1-p)/m}\, d\lambda\, dx.$$

This shows that, if $v = \exp(m/p(1-p))$ is fixed, then $mE(T)$ is monotone increasing in m (the integrand is positive and depends only on v).

The maximum of m, if v is fixed, is $m_{\max} = (\ln v)/4$. Putting $m = m_{\max}$ (and consequently $p(1-p) = \frac{1}{4}$) we get

$$L_\infty = (\ln v/(4(v^a + 1)) + \tfrac{1}{4} \ln v((v^a - 1)/(v^a + 1))E(T),$$

with

$$E(T) = \int_{-a}^{+a} \int_0^{\frac{1}{2}} (\pi t)^{-\frac{1}{2}} \exp - \frac{8x \ln v - \ln^2 vt}{16} \sum_{s=-\infty}^{+\infty} (-1)^s \exp$$

$$- \frac{(x - 2as)^2}{t}\, dt\, dx.$$

Here L_∞ is a function of v and a only. We strongly suspect, that this function has a saddle-point, but we did not succeed in proving it.

6. Generalization to other random variables. In order to generalize the procedure to other than binomial r.v.'s we have to make strong assumptions about the mathematical form of their distribution functions. Let

$$P(V(t) \leqq v) = \int_{-\infty}^{v} f(x, t)\, d\mu(x), \qquad E(V(t)) = p(t)$$

and

$$\ln \frac{f(x, t_1) f(y, t_2)}{f(x, t_2) f(y, t_1)} = z(x, y; t_1, t_2).$$

We make the following two assumptions:

$$z(x, y; t_1, t_2) = g(t_1, t_2) h(x, y), \tag{6.1}$$

where g does not depend on x and y and h does not depend on t_1 and t_2;

$$g(t_1, t_2) > 0 \quad \text{whenever} \quad p(t_1) > p(t_2). \tag{6.2}$$

Now let the outcomes of Ex I and Ex II be described by the r.v.'s $V(t_1)$ and $V(t_2)$ so that $X = V(t_1)$ and $Y = V(t_2)$, and that the density for the pair Ex I, Ex II is $f(x, t_1) f(y, t_2)$. Besides this hypothesis H_1 we consider the hypothesis

H_2 that the pair Ex I, Ex II has the density $f(x, t_2)f(y, t_1)$, i.e. that $X = V(t_2)$ and $Y = V(t_1)$. All computations will be done under H_1.

Let $Z = g(t_1, t_2)h(X, Y)$ and $U_k = \sum_{i=1}^k Z_i$. The Z_i are independent realizations of Z. We consider the following random walk in the (k, u) plane. The walk starts at $(0, 0)$. Absorbing barriers are (i) $u = a$ and (ii) $u = -a$. There should be a third absorbing barrier at $k = n/2$ (n is assumed to be even), but we do the following computations without regard to it, and therefore get only approximations. As long as the walking point is not yet absorbed, its general position is (k, U_k). The usual approximation for the probability of being absorbed at (i) is $\gamma = e^a/(e^a + 1)$ (see [1], p. 95). The conditional probability, γ_k say, for absorption at (i), provided the walk ends after k steps, is approximately $\gamma_k = \gamma$. As this may not be well known, we prove it.

The hypotheses H_1 and H_2 are such that $\gamma_k(H_1) = 1 - \gamma_k(H_2)$ and $P(K = k \mid H_1) = P(K = k \mid H_2)$. Here K is the random number of the step at which the point is absorbed. We define the event $\mathfrak{A}$ as

$$\mathfrak{A} = \{U_k \geqq a; U_i < a \quad \text{for} \quad i < k\} = \left\{ \prod_{i=1}^k f(X_i, t_1)f(Y_i, t_2) \geqq e^a \prod_{i=1}^k f(X_i, t_2)f(Y_i, t_1);\ U_i < a \text{ for } i < k \right\}.$$

Then, for all k with $P(K = k) \neq 0$,

$$\begin{aligned}\gamma_k P(K = k) = P(\mathfrak{A} \mid H_1) &= \int_{\mathfrak{A}} \prod_{i=1}^k f(x_i, t_1)f(y_i, t_2) \prod_{i=1}^k d\mu(x_i)\, d\mu(y_i) \\ &\geqq e^a \int_{\mathfrak{A}} \prod_{i=1}^k f(x_i, t_2)f(y_i, t_1) \prod_{i=1}^k d\mu(x_i)\, d\mu(y_i) = e^a P(\mathfrak{A} \mid H_2) \\ &= e^a(1 - \gamma_k)P(K = k).\end{aligned}$$

It follows that $\gamma_k \geqq e^a(1 - \gamma_k)$ or $\gamma_k \geqq e^a/(e^a + 1)$, and therefore

$$e^a/(e^a + 1) = \gamma = \sum_{k=1}^{\infty} \gamma_k P(K = k) \geqq e^a/(e^a + 1).$$

But then $\gamma_k = e^a/(e^a + 1) = \gamma$.

A strategy is as follows: we first take pairs X, Y of observations, say k times, and the rest of the $n - 2k$ observations are made either with X alone or with Y alone. The number k and whether to continue with X or with Y is given by the sequential plan. (We choose X when the random walk stops at (i)).

The expected outcome W of all n experiments is approximately

$$(6.3) \quad W = (p(t_1) + p(t_2))E(K) + (n - 2E(K))(\gamma p(t_1) + (1 - \gamma)p(t_2)).$$

This follows in the same way as equation (3.1).

Our strategy is completely symmetric in the treatments of Ex I and Ex II. The loss $L = n \max(p(t_1), p(t_2)) - W$ does not change when the names of the experiments are changed. We therefore may assume that $p(t_1) \geqq p(t_2)$ and we compute the loss under this assumption.

$$L = \frac{n(p(t_1) - p(t_2))}{e^a + 1} + \frac{a(p(t_1) - p(t_2))}{E(Z \mid H_1)} \left(\frac{e^a - 1}{e^a + 1}\right)^2,$$

where the relations $E(K) \cdot E(Z) = E(U_K) = a(e^a - 1)/(e^a + 1)$ and $\gamma = e^a/(e^a + 1)$ have been used.

To use the strategy practically we must be able to compute Z from X and Y; i.e. $g(t_1, t_2)$ must be known. As it is essential that no complete information about t_1 and t_2 is available, we proceed as follows:

Let $a = \alpha g(t_1, t_2)$, draw a plan with absorbing barriers at $u = \pm \alpha$ and use $\bar{U}_k = \sum_{i=1}^{k} \bar{Z}_i$, (with $\bar{Z} = h(X, Y)$) in this plan. This means only a change of scale (for $g > 0$). Then, with $u = \exp g(t_1, t_2)$, we have

$$L = \frac{n(p(t_1) - p(t_2))}{u^\alpha + 1} + \frac{\alpha(p(t_1) - p(t_2))}{E(h(X, Y) \mid H_1)} \left(\frac{u^\alpha - 1}{u^\alpha + 1}\right)^2.$$

Now our strategy is feasible as soon as we have decided which α to use, for the functional form of h is known. The loss is a function of our strategy α and of the pair t_1, t_2 (but not of the ordered pair).

To give an example, let $f(x, t) = (2\pi)^{-\frac{1}{2}} \exp - (x - t)^2/2$. Then $z = (t_1 - t_2)(x - y)$. We choose $g = t_1 - t_2$ and $h = x - y$ for then $g > 0$ whenever $p(t_1) = t_1$ is greater than $t_2 = p(t_2)$. Further

$$L = n(t_1 - t_2)/(u^\alpha + 1) + \alpha(u^\alpha - 1)^2(u^\alpha + 1)^{-2}$$

with $u = \exp(t_1 - t_2)$ is the loss computed under the assumption $t_1 \geqq t_2$. Without assumptions about t_1 and t_2 the loss will be

$$L = n\eta/(u^\alpha + 1) + \alpha(u^\alpha - 1)^2(u^\alpha + 1)^{-2}$$

with $\eta = |t_1 - t_2|$ and $u = \exp \eta$.

We can easily find a minimax solution. The maximum of η if u is constant, is $\eta_{\max} = \ln u$. Inserting this in L we have

$$(6.4) \qquad L = n \ln u/(u^\alpha + 1) + \alpha(u^\alpha - 1)^2(u^\alpha + 1)^{-2}.$$

Now the formula

$$M = n \frac{u^{\frac{1}{2}} - 1}{u^{\frac{1}{2}} + 1} \Big/ (u^\alpha + 1) + \alpha(u^\alpha - 1)^2(u^\alpha + 1)^{-2}$$

from Section 4 (at the end of (i)) gives, as $n \to \infty$ and $u \to 1$,

$$M \approx n \ln u/4(u^\alpha + 1) + \alpha(u^\alpha - 1)^2(u^\alpha + 1)^{-2};$$

and it was shown that there is a minimax solution $\alpha_n n^{-\frac{1}{2}} \to 0.292$. A comparison with (6.4) gives the following minimax solution for the above example:

$$\alpha = 0.292\ (4n)^{\frac{1}{2}} = 0.584\ n^{\frac{1}{2}}.$$

For another example, let $P(X = 1) = p = 1 - P(X = 0)$ and $P(Y = 1) = q = 1 - P(Y = 0)$. Let the likelihood-quotient be Q, then $\ln Q = (x - y) \ln (p(1 - q)/q(1 - p))$. With $u = \exp g(p, q) = p(1 - q)/q(1 - p)$ for $p > q$ and $u = q(1 - p)/p(1 - q)$ for $p < q$ and with $\sigma - \tau = | p - q |$ we get (3.2) as loss-function.

REFERENCES

[1] M. S. BARTLETT, *"An Introduction to Stochastic Processes,"* Cambridge University Press, 1955.

[2] R. BELLMAN, "A problem in the sequential design of experiments," *Sankhyā*, Vol. 16 (1956) pp. 221–229.

[3] R. N. BRADT, S. M. JOHNSON, AND S. KARLIN, "On sequential designs for maximising the sum of n observations," *Ann. Math. Stat.*, Vol. 27 (1956) pp. 1060–1074.

[4] J. L. DOOB, "Heuristic approach to the Kolmogorov-Smirnov theorems," *Ann. Math. Stat.*, Vol. 20 (1949) pp. 393–403.

[5] J. R. ISBELL, "On a problem of Robbins," *Ann. Math. Stat.*, Vol. 30 (1959) pp. 606–610.

[6] H. E. ROBBINS, "Some aspects of the sequential design of experiments," *Bull. Amer. Math. Soc.*, Vol. 55 (1952), pp. 527–535.

[7] H. ROBBINS, "A sequential decision problem with a finite memory," *Proc. Nat. Acad. Sci.*, Vol. 42 (1956), pp. 920–923.

[8] WALTER VOGEL, "An asymptotic minimax theorem for the two armed bandit problem," *Ann. Math. Stat.*, Vol. 31 (1960), pp. 444–451.