COMMENTS ON AUTOMATA IN RANDOM MEDIA*

Martin E. Hellman† and Thomas M. Cover‡

UDC 62-507

In this paper several approaches are presented to the problem of optimizing the design of a finite automaton for the hypothesis testing problem and the related two-armed bandit problem. It is noted that the two-armed bandit formulation is equivalent to a fundamental question raised by Tsetlin and his colleagues concerning the unknown optimal design of automata in random media. A solution of this problem is given by appropriate application of other work which is presented in condensed and unified form here. Closely related problems involving Markov switching hypotheses and multiple hypotheses remain unsolved.

1. Introduction

We will consider two related problems with finite memory constraints. The first is the hypothesis testing problem (HTP) and is formulated as follows: An infinite sequence of independent identically distributed random variable $\{X_n\}_{n=1}^{\infty}$ is generated sequentially, where X_n is distributed according to the probability measure P defined on the probability space $(\mathscr{X}, \mathscr{B}, P)$, and the sequence is distributed according to the corresponding infinite product measure. Consider the two-hypothesis testing problem:

$$\begin{aligned} H_0 &: P = P_0, \\ H_1 &: P = P_1. \end{aligned} \tag{1}$$

The a priori probabilities of hypotheses H_0 and H_1 are π_0 and $\pi_1 = 1 - \pi_0$, and are assumed known. We wish to make a sequence of decisions $\{d_n\}_{n=1}^{\infty}$, where $d_n \in \{H_0, h_1\}$ depends on $X_1, X_2, \ldots, X_n$ only through a finite valued, updatable statistic T. To be explicit,

$$\begin{aligned} T_n &= f(T_{n-1}, X_n) \in \{1, 2, \ldots, m\}, \\ d_n &= d(T_n) \in \{H_0, H_1\}, \end{aligned} \tag{2}$$

where T_n is the value of T at time n. We shall call T_n the state of the memory at time n and call m the size of memory. The objective is to find the pair of functions (f, d) which minimizes the long run probability of error

$$P(e) = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{\infty} P_n(e), \tag{3}$$

where

$$P_n(e) = \Pr\{d_n \neq H\} \tag{4}$$

and H is the underlying true hypothesis (which we may consider to have been drawn initially according to the prior probabilities π_0, π_1). Let p* denote the minimum achievable P(e) given by

*The results involving automata in random media (otherwise known as the two-armed bandit problem with time-invariant finite memory) were presented at the Symposium on Information Theory, Dubna, June 19-25, 1969. This work was supported under contract AF 49 (638) 1517, and under the National Science Foundation Graduate Fellowship Program. The Publisher thanks the authors for providing the original English manuscript.

†Massachusetts Institute of Technology, formerly with IBM Research and Stanford University.

‡Stanford University.

Published in Problemy Peredachi Informatsii, Vol. 6, No. 2, pp. 21-30, April-June, 1970.

$$P^* = \inf_{(f,d)} P(e). \tag{5}$$

The second problem to be considered is the two-armed bandit problem (TABP) [1-14]. This problem is closely related to the HTP and is stated as follows: Given are two experiments A and B. Whenever experiment A is performed, the experimental outcome Y is distributed according to the probability measure P_A. Similarly, whenever experiment B is performed, the experimental outcome Y is distributed according to P_B. Two hypotheses exist concerning P_A and P_B:

$$\begin{aligned} H_0 &: P_A = P_0 \text{ and } P_B = P_1, \\ H_1 &: P_A = P_1 \text{ and } P_B = P_0, \end{aligned} \tag{6}$$

where P_0 and P_1 are known measures defined on the probability space $(\mathscr{Y}, \mathscr{B}, \cdot)$ and $\mathscr{Y}$ is the space of experimental outcomes with Borel field $\mathscr{B}$. The a priori probabilities π_0 and π_1 are known. We wish to make a sequence of choices of experiment $\{e_n\}_{n=1}^{\infty}$ where $e_n \in \{A, B\}$ depends on the first n−1 observations $X_i = (e_i, Y_i)$, i = 1, 2, . . . , n−1, only through a finite-valued updatable statistic T. That is, for memory size m,

$$\begin{aligned} T_n &= f(T_{n-1}, X_n) \in \{1, 2, \ldots, m\}, \\ e_n &= e(T_{n-1}) \in \{A, B\}, \\ X_n &= (e_n, Y_n). \end{aligned} \tag{7}$$

Note that an observation consists of the experiment e performed and the resultant experimental outcome. Thus $X \in \{A, B\} \times Y \equiv \mathscr{X}$. It is assumed, for all i, j, that Y_i and Y_j are conditionally independent, conditioned on e_i and e_j, i.e.,

$$\Pr\{Y_i, Y_j | e_i, e_j\} = \Pr\{Y_i | e_i, e_j\} \Pr\{Y_j | e_i, e_j\}. \tag{8}$$

The objective is to find the pair of functions (f, e) which maximizes r*, the expected long-run proportion of uses of the experiment associated with P_0. Therefore, under H_0 one would like to perform A exclusively, while under H_1 one would like to perform B exclusively. Thus the TABP requires both a test between H_0 and H_1 and utilization of the results of the test in order to obtain a large proportion of successes (uses of P_0). Let

$$r^* = \sup_{(f,e)} r \tag{9}$$

denote the maximum achievable proportion of successes, where the maximum is taken over all algorithms (f, e) with fixed memory size m. (The dependence on m will not be explicitly noted.)

It is obvious that both the HTP and the TABP with finite memory are closely related to the behavior of finite automata in random media considered by Tsetlin and his colleagues [15-25]. The states in memory correspond to the internal states of the automaton; the observations are inputs and the decisions are outputs. The two hypotheses correspond to two different environments. In the HTP the automaton merely generates a sequence of guesses of the state of its environment; while in the TABP the automaton is able to affect the behavior of its environment by its actions (outputs).

Obviously, the maximal obtainable performance depends upon the nature of the restrictions we place on the mappings f, d, and e. The most restrictive condition is that the mappings be constant. This is trivial and uninteresting.

The first case of interest (and perhaps the most interesting case) is that in which the mappings are time-invariant, deterministic, Borel functions. The quantities P* and r*, and the associated optimal (f, d) and (f, e), are not known for this case, unless P_0 and P_1 represent continuous probability distributions [26]. Next, consider the mapping f to be stochastic, but still time-invariant [18, 19]. In this case Hellman and Cover [26] have found simple expressions for P* and r* and have characterized the optimal (f, d) and (f, e). In general P* > 0 and r* < 1, so that finite memory of this type precludes the gathering of perfect information about the environment.

Another option is to allow the mappings to be time-varying. That is, one specifies a sequence of functions $\{f_n\}_{n=1}^{\infty}$, where f_n is used to update T at time n. In this case Cover [9, 27] has shown that P* = 0 and r* = 1, for any finite memory m.

Adaptive schemes have also been considered by Chandrasekaran, Shen, Varshavskii, and Vorontsova [12, 13, 20]. Here the mappings are changed at each step, but the new mapping depends on the data. That is,

$$f_{n+1} = F(f_n, X_n). \tag{10}$$

where F is a fixed mapping and f_n is a function defined on $\{1, 2, \ldots, m\} \times \mathscr{X}$. The mappings d and e are similarly determined. Again, in this case P* and r* = 1 [11, 12, 13, 20].

One can raise objections to each of the last three models. Randomized rules require generation of random numbers; time-varying rules require an infinite clock; and adaptive rules require additional (perhaps infinite) memory capacity to store the current function f_n. The model chosen depends on the individual application.

The time-invariant, deterministic algorithms specified in Eqs. (2) and (7) have finite memory in all respects. If one believes that randomization requires no memory, or if one allows randomization by an independent outside agency, then allowing stochastic updating algorithms will not increase the memory. Thus we shall be interested here in the study of time-invariant stochastic rules as models of finite memory learning systems.

The next section gives a brief outline of the steps necessary to put bounds on P(e) and r. In §3 it is shown that these bounds are tight in the sense they can be approached arbitrarily closely. In §§2, 3, 4, the results of [14] and [26] are redeveloped from a unified point of view. This, together with the application of these results to the Automata in Random Media problem, is the contribution of this paper.

2. Bounds on Performance

Since we are concerned with randomized rules, the state transition function f may be specified by a family of mxm state-transition matrices indexed by $x \in X$. That is, $P(x) = [P_{ij}(x)]$, where

$$P_{ij}(x) = \Pr\{T_n = j \mid T_{n-1} = i, X_n = x\}. \tag{11}$$

From this reformulation it is easy to see that under H_t, t = 0 or 1, the sequence T_n, together with some initial state T_0, forms a Markov chain over the state space $S = \{1, 2, \ldots, m\}$.

Similarly, in the TABP, the mapping e can be specified by the m-vector α, where

$$\alpha_i = \Pr\{e_n = A \mid T_{n-1} = i\}, \qquad i = 1, 2, \ldots, m. \tag{12}$$

In the HTP, although d is allowed to be randomized, elementary decision theoretic considerations show that there is no loss in considering deterministic mappings only. Thus S may be partitioned into

$$S_0 = \{i \in S : d(i) = H_0\} \text{ and } S_1 = \{i \in S : d(i) = H_1\}. \tag{13}$$

As has been mentioned, the sequence T_n forms a Markov chain under H_0 and under H_1. Since P(x) is the state transition matrix conditioned on X = x, the state transition matrices conditioned on H_0 and H_1, $P^0 = [P^0_{ij}]$, and $P^1 = [P^1_{ij}]$ are given by

$$P_{ij}{}^t = \int_{\mathscr{X}} P_{ij}(x)\, dP_t(x), \; t = 0, 1, \tag{14}$$

for the HTP, while for the TABP

$$P_{ij}{}^t = \alpha_i \int_{\mathscr{Y}} P_{ij}(A, y)\, dP_t(y) + (1 - \alpha_i) \int_{\mathscr{Y}} P_{ij}(B, y)\, dP_{\bar{t}}(y), \tag{15}$$

where t = 0, 1 and $\bar{t} = 1 - t$.

The matrics P^0 and P^1 determine the stationary probability of occupation vectors μ^0 and μ^1, where

$$\mu_i{}^t = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \Pr\{T_n = i \mid H_t\}. \tag{16}$$

(If P^0 or P^1 determines a reducible Markov chain, the initial state is also needed. However, it can be shown [26] that these rules are suboptimal and may thus be neglected.) Finally, we may express

$$P(e) = \pi_0 \sum_{i \in S_1} \mu_i{}^0 + \pi_1 \sum_{i \in S_0} \mu_i{}^1 \equiv \pi_0 r_0 + \pi_1 r_1 \tag{17}$$

and

$$r = \pi_0 \sum_{i \in S} \alpha_i \mu_i{}^0 + \pi_1 \sum_{i \in S} (1 - \alpha_i) \mu_i{}^1 \equiv \pi_0 r_0 + \pi_1 r_1. \tag{18}$$

In order to minimize P(e), we wish μ_i^0 to be large for $i \in S_0$ and small for $i \in S_1$, while wishing the opposite behavior for μ_i^1. This may be achieved only if P^0 is 'very different' from P^1. (Similar remarks hold for maximizing r). The following lemma bounds this difference:

LEMMA 1. Let

$$\bar{l} = \sup \frac{P_0(A)}{P_1(A)} \text{ and } \underline{l} = \inf \frac{P_0(A)}{P_1(A)}, \tag{19}$$

where the supremum and infimum are taken over all sets A such that $P_0(A) + P_1(A) > 0$. Also let

$$L = \max\{\bar{l},\ 1/\underline{l}\}. \tag{20}$$

Then for the HTP

$$\underline{l} \leqslant P_{ij}{}^0 / P_{ij}{}^1 \leqslant \bar{l}, \quad \forall i,\ j \in S, \tag{21}$$

while for the TABP

$$1/L \leqslant P_{ij}{}^0 / P_{ij}{}^1 \leqslant L, \quad \forall i, j \in S. \tag{22}$$

Proof. Let $\nu = P_0 + P_1$ and $f_t(x) = dP_t(x)/d\nu(x)$, $t = 0, 1$, and $l(x) = f_0(x)/f_1(x)$. Note that $\underline{l} \le l(x) \le \bar{l}$, a.e. and that $dP_0 = f_0(x)d\nu = l(x)f_1(x)d\nu$ and $dP_1 = f_1(x)d\nu$. Then use (14) and (15) to obtain (21) and (22).

Definition. Let $\gamma = \bar{l}/\underline{l}$ for the HTP and $\gamma = L^2$ for the TABP.

Definition. Let the state likelihood ratio of state i be denoted by

$$\lambda_i = \mu_i{}^0 / \mu_i{}^1. \tag{23}$$

LEMMA 2. For an irreducible automaton, if the state likelihood ratios are arranged in nondecreasing order, then

$$\lambda_{i+1} / \lambda_i \leqslant \gamma \text{ for all i.} \tag{24}$$

Proof. Suppose the lemma were false. Then for some $k \in S$, $\lambda_{k+1}/\lambda_k > \gamma$ or

$$\mu_i{}^0 / \mu_i{}^1 \leqslant \lambda_k \text{ for } i \in C \equiv \{1,\ 2, \ldots, k\} \tag{25}$$

and

$$\mu_i{}^0 / \mu_i{}^1 > \lambda_k \gamma \text{ for } i \in C' \equiv \{k+1,\ k+2, \ldots, m\}. \tag{26}$$

When the Markov chain is in the steady state, the probability of a transition from C to C' must equal the probability of a transition from C' to C [30]. That is,

$$\sum_{i \in C} \sum_{j \in C'} \mu_i{}^t P_{ij}^{\,t} = \sum_{i \in C'} \sum_{j \in C} \mu_i{}^t P_{ij}{}^t, \quad t = 0, 1. \tag{27}$$

Considering the HTP, use (21) and (25) to obtain

$$\sum_{i \in C'} \sum_{j \in C} \mu_i{}^0 P_{ij}{}^0 \leqslant \lambda_k \bar{l} \sum_{i \in C} \sum_{j \in C'} \mu_i{}^1 P_{ij}{}^1, \tag{28}$$

and (22) and (26) to obtain

$$\sum_{i \in C'} \sum_{j \in C} \mu_i{}^0 P_{ij}{}^0 > \lambda_k (\bar{l}/\underline{l})\, \underline{l} \sum_{i \in C'} \sum_{j \in C} \mu_i{}^1 P_{ij}{}^1. \tag{29}$$

But, using (27), the right-hand sides of (28) and (29) are equal, as are the left-hand sides, a contradiction.

Note that the Markov chain must be irreducible to obtain (29).

Definition. Let the spread of an automaton be the ratio of its maximum state likelihood ratio to its minimum state likelihood ratio.

LEMMA 3. The spread of an m-state automaton is less than or equal to γ^{m-1}.

Proof. If the automaton is irreducible, m−1 applications of Lemma 2 yield the desired result. If the automaton is reducible, it can be shown [26] that the spread is less than γ^{m-2}, an even stronger bound.

THEOREM 1: For the HTP

$$P^{*} = \min\left\{\pi_0, \pi_1, \frac{2(\pi_0\pi_1\gamma^{m-1})^{1/2} - 1}{\gamma^{m-1} - 1}\right\} \tag{30}$$

is a lower bound on P(e), and for the TABP

$$r^{*} = \max\left\{\pi_0, \pi_1, \frac{\gamma^{m-1} - 2(\pi_0\pi_1\gamma^{m-1})^{1/2}}{\gamma^{m-1} - 1}\right\} \tag{31}$$

is an upper bound on r. Note that if $\pi_0 = \pi_1 = 1/2$ then (30) and (31) reduce to

$$P^{*} = 1/(\gamma^{(m-1)/2} + 1), \tag{32a}$$

$$r^{*} = \gamma^{(m-1)/2}/(\gamma^{(m-1)/2} + 1). \tag{32b}$$

Proof. From Lemma 3, there exists a real number k such that

$$k \leqslant \mu_i^0/\mu_i^{-1} \leqslant k\gamma^{m-1}, \quad \forall i \in S. \tag{33}$$

Then for the HTP

$$\alpha \sum_{i \in S_1} \mu_i^0 \geqslant k \sum_{i \in S_1} \mu_i^1 = k\left(1 - \sum_{i \in S_0} \mu_i^1\right) = k(1 - \beta) \tag{34}$$

and

$$\beta = \sum_{i \in S_0} \mu_i^1 \geqslant (1/k\gamma^{m-1})(1 - \alpha) \tag{35}$$

or

$$\alpha\beta \geqslant (1/\gamma^{m-1})(1 - \alpha)(1 - \beta). \tag{36}$$

But Lagrange minimization of $P(e) = \pi_0\alpha + \pi_1\beta$ subject to the constraint (36) yields the expression (30) as desired.

For the TABP the constraint is

$$r_0 r_1 \leqslant \gamma^{m-1}(1 - r_0)(1 - r_i); \tag{37}$$

and maximizing $r = \pi_0 r_0 + \pi_1 r_1$ subject to the constraint (37) yields the desired expression (31).

3. A Class of ε-Optimal Automata

In the previous section bounds were established on performance. In this section it is shown that these bounds are the tightest possible.

First, consider the HTP. Let $\mathscr{H} = \{x : l(x) = \bar{l}\}$ and $\mathscr{T} = \{x : l(x) = \underline{l}\}$. For the moment assume that $p_t = P_t(\mathscr{H})$ and $q_t = P_t(\mathscr{T})$, t = 0, 1, are all nonzero. Thus by definition of $l(x)$

$$p_0/p_1 = \bar{l} \text{ и } q_0/q_1 = \underline{l}. \tag{38}$$

We claim that the automata with state transition algorithm

$$P_{ij}(x) = \begin{cases} 1, & \text{if } x \in \mathscr{H} \text{ and } 3 \leqslant j = i + 1 \leqslant m, \\ & \text{or } x \in \mathscr{T} \text{ and } 1 \leqslant j = i - 1 \leqslant m - 2, \\ \delta, & \text{if } x \in \mathscr{H} \text{ and } j = i + 1 = 2, \\ k\delta, & \text{if } x \in \mathscr{T} \text{ and } j = i - 1 = m - 1, \\ 0, & \text{otherwise} \end{cases} \tag{39}$$

and decision function

$$d(i) = \begin{cases} H_0, & i > m/2, \\ H_1, & i \leqslant m/2 \end{cases} \tag{40}$$

can achieve $P(e) \leq P^{*} + \varepsilon$ for any $\varepsilon > 0$, merely by proper choice of δ and k; i.e., this class of automata is ε-optimal.

That this is so can be seen by solving (27) for $\underline{\mu}^0$ and $\underline{\mu}^1$, yielding

$$\underline{\mu}^0 = (a/\delta, a(p_0/q_0), a(p_0/q_0)^2, \ldots, a(p_0/q_0)^{m-1}/k\delta), \tag{41}$$

$$\underline{\mu}^1 = (b/\delta, b(p_1/q_1), b(p_1/q_1)^3, \ldots, b(p_1, q_1)^{m-1}/k\delta), \tag{42}$$

where a and b satisfy $\Sigma\mu_i^0 = \Sigma\mu_i^1 = 1$. Now $\delta \to 0$ implies that $\mu_i^t \to 0$, for t = 0, 1 and i = 2, 3, . . . , m−1. Thus

$$\mu_1{}^t + \mu_m{}^t \to 1 \text{ for } \quad t = 0, 1 \tag{43}$$

Also

$$\alpha = P(e/H_0) \to \mu_1{}^0, \tag{44}$$

$$\beta = P(e/H_1) \to \mu_m{}^1. \tag{45}$$

Thus

$$\alpha\beta \to \frac{ab}{k\delta^2}(p_1/q_1)^{m-1} \tag{46}$$

and

$$(1-\alpha)(1-\beta) \to \frac{ab}{k\delta^2}(p_0/q_0)^{m-1}. \tag{47}$$

But $p_0 = \bar{l}p_1$ and $q_0 = \underline{l}q_1$, so

$$\alpha\beta \to \left(\frac{p\underline{l}q_0}{p_0 q_1}\right)^{m-1} \quad (1-\alpha)(1-\beta) \tag{48}$$

$$= (\underline{l}/\bar{l})^{m-1}(1-\alpha)(1-\beta) \tag{49}$$

$$= (1/\gamma)^{m-1}(1-\alpha)(1-\beta). \tag{50}$$

This is just the equation (36) of the lower boundary of the operating characteristic. Since α can be forced to any value between zero and one by varying k, any point of the lower boundary can be approached as closely as desired. For the optimal value k = k*, P(e) → P* as $\delta > 0$ tends to zero. Note that $\delta = 0$ yields P(e) ≫ P* in general. Thus the limiting automaton is bad.

Now dropping the assumption that $\mathcal{H}$ and $\mathcal{T}$ have nonzero probability, we can approximate $\mathcal{H}$ and $\mathcal{T}$ for $\bar{l} < \infty$, by

and

$$\mathcal{H}_\varepsilon = \{x : l(x) \leqslant \bar{l} - \varepsilon\} \tag{51}$$

$$\mathcal{T}_\varepsilon = \{x : l(x) \geqslant \underline{l} + \varepsilon\}. \tag{52}$$

The case $\bar{l} = \infty$ is easily disposed of by a separate argument. From the definitions of $\bar{l}$ and $\underline{l}$ it is seen that $P_t(\mathcal{H}_\varepsilon)$ and $P_t(\mathcal{T}_\varepsilon)$ are both nonzero for any $\varepsilon > 0$. By letting $\varepsilon \to 0$ and $\delta \to 0$ with $\varepsilon > \bar{0}$, $\delta > 0$ and k = k*, we again approach P* as closely as desired. Thus P* is a tight bound, even though it cannot be achieved [26].

The ε-optimal solution to the TABP is very similar to that of the HTP. If $\bar{l} < 1/l$ then $L = \bar{l}$ and if $l(y_0) = f_0(y_0)/f_1(y_0) = \bar{l} = L$ then $x_0 = (A, y_0)$ is the observation which most supports H_0, and $x_1 = (B, y_0)$ is the observation which most supports H_1. If $\nu(y_0) > 0$, then the automaton with

$$P_{ij}(x) = \begin{cases} 1, & \text{if} \quad x = x_0 \text{ and } 2 \leqslant j = i + 1 \leqslant m \\ & \text{or} \quad x = x_1 \text{ and } 1 \leqslant j = i - 1 \leqslant m - 1, \\ 0, \text{otherwise} & \end{cases} \tag{53}$$

and

$$\alpha_i = \begin{cases} 1/2, & 2 \leqslant i \leqslant m-1, \\ \delta, & i = 1, \\ 1 - k\delta, & i = m, \end{cases} \tag{54}$$

approaches the upper boundary (37) to the operating characteristic, and for the optimal value k = k*, r → r* as $\delta > 0$ approaches zero.

If $\bar{l} < 1/\underline{l}$, then $L = (1/\underline{l})$. Now, if $l(y_1) = l$ and $\nu(y_1) > 0$, let $x_2 = (B, y_1)$ and $x_3 = (A, y_1)$. Then the automaton with

$$P_{ij}(x) = \begin{cases} 1, & \text{if } \quad x = x_3 \text{ and } 3 \leqslant j = i+1 \leqslant m \\ & \text{or } \quad x = x_3 \text{ and } 1 \leqslant j = i-1 \leqslant m-2, \\ \delta, & \text{if } \quad x = x_3 \text{ and } j = i+1 = 2 \\ k\delta, & \text{if } \quad x = x_3, \\ 0, & \text{otherwise} \end{cases} \tag{55}$$

and

$$\alpha_i = \begin{cases} 0, & i = 1, \\ \frac{1}{2}, & 2 \leqslant i \leqslant m-1, \\ 1, & i = m \end{cases} \tag{56}$$

approaches the upper boundary (37) to the operating characteristic. Again r → r* for k = k*.

If $\nu(y_0) = 0$ or $\nu(y_1) = 0$ then, as in the HTP, suitable approximations can be found so that r → r*.

4. Examples

We are given two coins, A and B. One coin has probability P_0 of showing heads and the other has probability P_1 of showing heads. It is not known which coin has which bias, and there is equal probability of either labeling ($\pi_0 = \pi_1 = 1/2$).

a) If $P_0 = 3/4$ and $P_1 = 1/4$ then $\overline{l} = 3$, $\underline{l} = 1/3$ and L = 3. Thus $\gamma = 9$ for both the HTP and TABP. A five state memory (m = 5) can ε-achieve $P^* = 1/(\gamma^{(m-1)/2} + 1) = 1/82$ and $r^* = \gamma^{(m-1)/2}/(\gamma^{(m-1)/2} + 1) = 81/82$. Thus the limiting probability of error is 1/82 for the HTP and the limiting proportion of uses of the "best" coin is 81/82 for the TABP.

b) If $P_0 = 0.99...99$ and $P_1 = 0.99...90$ then $\overline{l} \approx 1$, $\underline{l} \approx 1/101$ and $L \approx 10$. Thus $\gamma \approx 10$ for the HTP and $\gamma \approx 100$ for the TABP. Thus, for a five state memory $P^* \approx 1/101$ and $r^* \approx 10{,}000/10{,}001$. Note the much improved "probability of error" in the TABP with respect to the HTP. In this case, two coins are better than one.

c) If $P_0 = 0.501$ and $P_1 = 0.499$, then $L = \overline{l} = 1/\underline{l} \approx 1.004$. Thus for a five-state memory $P^* \approx 0.496$ and $r^* \approx 0.504$, little better than with no memory at all. In fact, approximately 500 states are required to obtain $P^* = 0.01$ or $r^* = 0.99$.

5. Conclusions

The outline presented here of recent results on optimal algorithms for the finite memory constrained hypothesis testing problem (HTP) and the two-armed bandit problem (TABP) points out the similarity of the two problems. Both problems require artificially randomized transition rules, unless (in the HTP) P_0 and P_1 represent continuous probability distributions.

A notable feature of the ε-optimal algorithm is its simple nature: move up one state on extreme observations which support H_0, and move down one state on extreme observations that support H_1. If the number of trials is finite, the optimal algorithm is not yet known. However, for a large but finite number of trials, the optimal algorithm should be similar to the one presented here. High (or low) likelihood ratio observations will still cause upward (or downward) transitions, although the events need not be as extreme as before and transitions need not be between adjacent states. Furthermore, randomization will still be required, although the optimal values of δ will not be arbitrarily close to zero.

It is not clear to us what the optimal algorithm would be in the important "switching environment" formulation put forth by Tsetlin [15, 16], in which the underlying hypothesis (state of nature) H_0 or H_1 is not constant from trial to trial, but obeys a Markov process.

We wish to thank D. Sagalowicz for his contributions to the ideas in this work and T. Kailath for delivering a talk for us on the subject of this paper at the Dubna, USSR, Conference on Information Theory, June, 1969.

LITERATURE CITED

1. Herbert Robbins, "Some aspects of the sequential design of experiments," Bull. Amer. Math. Soc., 58, 529-532 (1952).
2. R. N. Bradt, S. M. Johnson, and S. Karlin, "On sequential designs for maximizing the sum of n observations," Ann. Math. Statist., 27, 1060-1074 (1956).
3. R. N. Bradt and S. Karlin, "On the design and comparison of certain dichotomous experiments," Ann. Math. Statist., 27, 390-409 (1956).
4. Dorian Feldman, "Contributions to the 'two-armed bandit' problem," Ann. Math. Statist., 33, 817-856 (1962).
5. Herbert Robbins, "A sequential decision problem with a finite memory," Proc. Nat. Acad. Sci., 42, 920-933 (1956).
6. J. R. Isbell, "On a problem of Robbins," Ann. Math. Statist., 30, 606-610 (1959).
7. C. V. Smith and R. Pyke, "The Robbins–Isbell two-armed bandit problem with finite memory," Ann. Math. Statist., 36, 1375-1386 (1965).
8. S. M. Samuels, "Randomized rules for the two-armed bandit with finite memory," Ann. Math. Statist., 39, No. 6, 2103-2107 (1968).
9. Thomas M. Cover, "A note on the two-armed bandit problem with finite memory," Info. and Control, 12, No. 3, 371-377 (1968).
10. K. S. Fu and T. J. Li, "On the behavior of learning automata and its applications," Purdue University Technical Report No. TR-EE 68-20 (1968).
11. K. S. Fu and T. J. Li, "Formulation of learning automata and automata games," Information Sciences, 1, 237-256 (1969).
12. B. Chandrasekaran, "Contributions to the theory of learning automata," Ph. D. Dissertation, University of Pennsylvania, May (1967).
13. B. Chandrasekaran and D. W. C. Shen, "Adaption of stochastic automata in nonstationary environments," Proceedings of the National Electronics Conference; Sixth Symposium on Discrete Adaptive Processes, Chicago, Illinois (October, 1967).
14. Thomas M. Cover and Martin E. Hellman, "The two-armed bandit problems with time-invariant finite memory," to appear in IEEE Transactions on Information Theory (March, 1970).
15. M. L. Tsetlin, "Certain problems in the behavior of finite automata," Dokl. Akad. Nauk SSSR, 139, No. 4, 830-833 (1961).
16. M. L. Tsetlin, "On the behavior of finite automata in random media," Avtomatika i Telemekhanika, 22, No. 10, 1345-1354 (1961).
17. M. L. Tsetlin, "A game between a finite automaton and an opponent using a mixed strategy," Dokl. Akad. Nauk SSSR, 149, No. 1, 52-53 (1963).
18. V. Yu. Krylov, "One one automaton that is asymptotically optimal in a random medium," Avtomatika i Telemekhanika, 24, No. 9, 1226-1228 (1963).
19. V. Yu. Krylov and M. L. Tsetlin, "Examples of games with automata," Dokl. Akad. Nauk SSSR, 149, No. 2, 284-287 (1963).
20. Y. I. Varshavskii and I. P. Vorontsova, "On the behavior of stochastic automata with a variable structure," Avtomatika i Telemekhanika, 24, No. 3, 353-360 (1963).
21. V. L. Stefanyuk, "Example of a problem in the joint behavior of two automata," Avtomatika i Telemekhanika, 24, No. 6, 781-784 (1963).
22. I. M. Gel'fand, I. I. Pyatetskii-Shapiro, and M. L. Tsetlin, "Certain classes of games and automata games," Dokl. Akad. Nauk SSSR, 152, No. 4, 845-848 (1963).
23. S. L. Ginzburg, V. Yu. Krylov, and M. L. Tsetlin, "One example of a game for identical automata," Avtomatika i Telemekhanika, 25, No, 5, 668-672 (1964).
24. D. I. Kalinin and I. M. Rotvain, "Some asymptotic estimates for games of automata in distribution," Avtomatika i Telemekhanika, 27, No. 4, 119-121 (1966).
25. N. T. Kandelaki and G. N. Tsertsvadze, "Behavior of certain classes of stochastic automata in random media," Avtomatika i Telemekhanika, 27, No. 6, 115-119 (1966).
26. Martin E. Hellman and Thomas M. Cover, "Learning with finite memory," to appear in Ann Math. Stat. (1970).
27. Thomas M. Cover, "Hypotheses testing with finite statistics," Ann. Math. Stat., 40, No. 3, 828-835 (1969).