Lecture Notes in Statistics

Edited by D. Brillinger, S. Fienberg, J. Gani,
J. Hartigan, and K. Krickeberg

20

Mathematical Learning Models– Theory and Algorithms

Proceedings of a Conference

Edited by
Ulrich Herkenrath, Dieter Kalin,
and Walter Vogel

Springer-Verlag
New York Berlin Heidelberg Tokyo

ON A CLASS OF LEARNING ALGORITHMS WITH SYMMETRIC BEHAVIOR UNDER SUCCESS AND FAILURE

by

M. R. Meybodi and S. Lakshmivarahan
School of Electrical Engineering and Computer Science
University of Oklahoma, Norman, Oklahoma

1. Introduction:

Learning algorithms have been extensively studied in Mathematical Psychology [13] [14] [18] for decades and more recently in Learning Automata Theory and Mathematical Statistics [1] [2] [8] [11] [16]. In Mathematical Psychology the interest in learning stems from the desire to understand the observed animal learning and associated changes in their behavior. However, in Learning Automata Theory and Mathematical Statistics the aim is to build algorithms that exhibit prespecified behaviour. Our interests, in this paper, are in the latter approach.

We study a new class of absorbing barrier algorithms of the reward-penalty type with identical behaviour under the occurrence of success and failure. Necessary and sufficient conditions for strong absolute expediency of this class of algorithm is obtained.

The concept of absolute expediency was originally introduced by Lakshmivarahan and Thathachar [3] in 1973. Since then this concept has played a major role in the analysis and design of ε-optimal absorbing barrier algorithms. [1] [2] [4] [5] [7]. The results of this paper sheds further light on this concept. It should be interesting to note that ε-optimal non-absorbing learning algorithms have also been recently developed in [15] and discussed extensively in the book [21].

A number of learning algorithms of the reward-penalty type whose behaviour is asymmetrical with respect to the occurrence of success and failure have been extensively studied in the literature - Varshavskii and Vorontsova [10], Fu [8], McMurty and Fu [9], Shapiro and Narendra [20], Viswanathan and Narendra [19], Chandrasekaran and Shen [17], Lakshmivarahan and Thathachar [3] and Sawaragi and Baba [7] to mention a few. For a similar class of general learning algorithms discussed in this paper Aso and Kimura [5] derived necessary and sufficient conditions for absolute expediency [3]. Interestingly Aso and Kimura [5] call this class of algorithms as "Stochastic Vector Automaton" algorithms. Recently for the same class of algorithms considered in this paper Herkenrath, Kalin and Lakshmivarahan [6] derived necessary and sufficient conditions for absorption at the vertices of the unit simplex of proper dimension.

In section 2, the algorithm and the statement of problem are given. Necessary and sufficient conditions for strong absolute expediency and convergence of the

algorithm with probability one are established in section 3. ϵ-optimality is derived in section 4.

2. LEARNING ALGORITHM and STATEMENT OF PROBLEM

There are M ($2 \leq M < \infty$) coins. At time instant k ($= 0,1,2....$) the coin i ($1 \leq i \leq M$) is chosen for tossing with probability $P_i(k)$ where $P(k) = (P_1(k), P_2(k), ...P_M(k))^T$, $\sum_{i=1}^{M} P_i(k) = 1$, $0 \leq P_i(k)$ and T denotes transpose. In tossing, the i^{th} coin falls head (tail) with probability d_i ($c_i = 1 - d_i$). It is assumed that (1) $0 < d_i < 1$, $i = 1, 2, ...M$, (2) the d_i's are all distinct, that is, $d_i \neq d_j$ for all i and j. (3) d_i's do not depend on k and (4) d_i's are all unknown. The outcome of falling head is called success and falling tail is called failure. Let $D = (d_1, d_2 ...d_M)^T$ and without loss of generality assume

$$d_1 > d_2 > d_3 ... > d_M \tag{1}$$

Any vector D that satisfies the above conditions is called an admissible D. The average probability of success at stage k denoted by $\eta(k)$ is

$$\eta(k) = \sum_{i=1}^{M} P_i(k)\, d_i \tag{2}$$

Our basic problem is to make the average probability of success as close to its maximum (that is, d_1) as possible. The primary interest in and the challenge of this problem arise basically from the fact that the success probabilities (d_i's) of the coin are unknown.

As a first step towards the solution to this problem, in this paper, we propose to change p(k) in a learning algorithm. The key idea of the learning algorithm may be stated in words as follows: Increase (decrease) the probability of choosing the i^{th} coin at time (k + 1), if it was chosen for tossing at time k and the toss resulted in success (failure). In particular, let

$$S_M = \{ P \mid P = (P_1, P_2, \ldots P_M)^T,\ 0 \leq P_i,\ \sum_{i=1}^{M} P_i = 1 \}$$

be the M-dimensional unit simplex and let

$$V_M = \{ e_i \mid i = 1, 2, ...M \} \text{ where } e_i = (0, 0, ...1, ...0)^T$$

be the i^{th} unit i vector of dimension M. Further, let $S_M^0 = \{ P \mid P = (P_1\ P_2, ...P_m)^T$ $0 < P_i$, $\sum_{i=1}^{M} P_i = 1 \}$ Clearly, V_M corresponds to the corners of vertices of S_M.
Let $f_s^i [.]$, $g_s^i [.]$ be continuous functions such that

$$f_s^i : S_M \to [0,1],\ g_s^i : S_M \to [0,1],\ i, s = 1, 2, ...M$$

Formally the above algorithm may be stated as follows

$$\left.\begin{aligned} P_s(k+1) &= P_s(k) - \theta\, f_s^i [P(k)], s \neq i \\ P_i(k+1) &= P_i(k) + \theta\, f_i^i [P(k)] \end{aligned}\right\} \text{ if the toss of coin i resulted in success} \tag{3}$$

and

$$\left.\begin{aligned} P_i(k+1) &= P_i(k) - \theta\, g_i^i [P(k)] \\ P_s(k+1) &= P_s(k) + \theta\, g_s^i [P(k)], s \neq i \end{aligned}\right\} \text{ if the toss of coin i resulted in failure}$$

where $0 < \theta \leq 1$ is called the step length parameter and the following cons condition (4):

$$\begin{aligned} &\text{either } f_s^i [.] \equiv 0 \text{ for all } p \in S_M \\ &\text{or } f_s^i [p] \leq P_s \quad s \neq i \text{ and } f_i^i [p] = \sum_{s \neq i} f_s^i [p] \end{aligned} \tag{4}$$

and

$$\begin{aligned} &\text{either } g_s^i [p] \equiv 0 \text{ for all } p \in S_M \\ &\text{or } g_s^i [p] \leq P_s \text{ and } g_i^i [p] = \sum_{s \neq i} g_s^i [p] \end{aligned}$$

for all i, s = 1, 2, ...M,
imply that $p(k + 1) \in S_M$ if p(k) does. However in order to make the algorithm non trivial and interesting either $f_s^i \equiv 0$ or $g_s^i \equiv 0$ but not both.

Remark 1. If the coin i is chosen for tossing, it is clear from (3) that success (failure) increases (decreases) the probability of its choice. The increase and decrease are called reward and penalty and hence (3) is called reward-penalty algorithm. If $f_s^i \neq 0$ but $g_s^i \equiv 0$ then (3) is called reward-inaction algorithm, if $f_s^i \equiv 0$ but $g_s^i \neq 0$ then it is called inaction- penalty algorithm

Since the vector D, of success probabilities and the functions $f_s^i [.]$ and $g_s^i [.]$ (for all i and s) are independent of k, $\{p(k)\}, k \geq 0$ is clearly a discrete time Markov process with stationary transition function over the state space S_M. To quantify the behaviour of the process $\{p(k)\}$, and hence of $\{\eta(k)\}$ we introduce the following definitions. Let $I = \{1,2,....M\}$ $E = \{success, Failure\}$. The algorithm (3) defines a mapping T: $S_M \times I \times E \to S_M$ where

$$P(k+1) = T [p(k), i(k), e(k)] \tag{5}$$

$i(k) \in I$ is the coin chosen for tossing and $e(k) \in E$ is the outcome of the toss of that coin at time k.

Definition 1: A state $p \in S_M$ is said to be absorbing state if and only if $p = T [p, i, e,]$ with probability one (6)

Definition 2: A learning algorithm T is said to be absorbing if and only if there is at least one absorbing state.

For reasons that will become apparent in this paper our interests are in the class of learning algorithms for which V_M is the only set of absorbing states. We call such an algorithm absorbing barrier algorithms

Definition 3: A learning algorithm T is said to be

a) optimal if

$$\lim_{k \to \infty} E [\eta(k)] = d_1 \tag{7}$$

b) ϵ-optimal if for all $p(0) = p \in S_M^0$

$$\lim_{k \to \infty} | E [\eta(k)] - d_1 | < \epsilon,\ \epsilon > 0 \tag{8}$$

c) absolutely expedient if

$$E [\eta(k+1) \mid p(k) = p] \geq \eta(k) \text{ with probability one} \tag{9}$$

for all admissible D (satisfying the assumptions given at the beginning of this

section) and for all $p \in S_M$, with strict inequality in (9) holding good for all $p \in S_M^o$.

d) strongly absolutely expedient if

$$E[\eta(k+1) \mid p(k) = p] \geq \eta(k) \text{ with probability one} \qquad (10)$$

for all admissible D and for all $p \in S_M$ with strict inquality in (10) holding good for all $p \in (S_M - V_M)$.

STATEMENT OF PROBLEM: Our aim in this paper is to find conditions for ϵ-optimality of the general class of learning algorithm given in (3).

Remark 2: Algorithm (3) is a generalization of many known algorithms in the sense that the functions in (3) used for updating the probabilities depend on the coin chosen as well. Such generalized algorithms are considered in [5] and [6]. In most of the earlier papers the functions that are used in updating are independent of the coin that is chosen for tossing, [1] [2] [3] [4] [7] [8] [9] [10] [12] [13].

3. Convergence with probability one:

As a first step towards ϵ-ptimality, in this section we derive conditions on the algorithm such that the Markov process $\{p(k)\}$ $k \geq 0$ converges with probability one. To this end we begin be rewriting the consistency conditions (4) in a form more suitable for our analysis. Let

(C.1) $$f_i^i[p] = \alpha[i,p](1 - p_i), \quad f_s^i[p] = \beta[i,s,p]\, p_s$$

and

$$\sum_{s \neq i} \beta[i,s,p]\, p_s = \alpha[i,p](1 - p_i)$$

and

(C.2) $$g_i^i[p] = \gamma[i,p]\, p_i, \quad g_s^i[p] = \Delta[i,s,p](1 - p_s)$$

and

$$\gamma[i,p]\, p_i = \sum_{s \neq i} \Delta[i,s,p](1 - p_s)$$

where $\alpha, \gamma : I \times S_M \to [0,1]$ and $\beta, \Delta : I \times I \times S_M \to [0,1]$.

The basic rule that governs the choice of functions in the above form is that if a term is substracted from p_j then it is made proportional to p_j and if a term is added to p_j then it is made proportional to $(1 - p_j)$ irrespective of which coin is chosen for the toss and whether the toss results in success or failure.

Remark 3: Our choice of the functions $g_s^i[.]$ is quite untraditional in the sense that in all most all the papers in Mathematical Psychology [18] [17] [13] [14] $g_s^i[p]$ is made proportional to p_s for all i and s, $s \neq i$. Also in almost all the papers on Learning Automata [3] [4] [5] [7] $g_i^i[p]$ is made proportional to $(1 - p_i)$ and $g_s^i[p]$ is made proportional to p_s for all $s \neq i$. Because of this there is a disparity in the behaviour of the algorithm (3) under success and failure. However, our present choice of functions $g_s^i[.]$ for all i and s given in (C.2) induce identical behaviour of the algorithm (3) under success and failure.

The following theorem is immediate

Theorem 1: The learning algorithm (3) is an absorbing barrier learning a and only if (A.1) and (A.2) hold true.

(A.1) For all $P \in S_M - V_M$ there exists $1 \leq s \leq M$ such that $[\alpha[s,p] + \gamma[s,p]]\, p_s > 0$

(A.2) For all $1 \leq s \leq M$, $\gamma[s, e_s] = 0$

Proof: Refer [6]

A typical choice of functions in the algorithm (3) that satisfies the conditions (C.1) - (C.2) and (A.1) - (A.2) are given in the following example.

Let $0 < C_1, C_2 < 1$.

$$\beta[i,s,p] = C_1 (1 - p_i)(1 - p_s), \quad \alpha[i,p] = C_1 \sum_{s \neq i} p_s (1 - p_s)$$

$$\Delta[i,s,p] = C_2\, p_i\, p_s^2 (1 - p_i), \quad \gamma[i,p] = C_2 (1 - p_i) \sum_{s \neq i} p_s^2 (1 - p_s)$$

for all i, s = 1, 2, ...M.

Remark 4: The demonstration of the symmetry in the properties of the algofithm (3) under success and failure alluded to inremark 3 is evidenced by theorem 1. Notice that (3) is an absorbing barrier learning algorithm if $\alpha[i,p] \neq 0$ and $\gamma[i,p] \neq 0$ or $\alpha[i,p] \neq 0$ and $\gamma[i,p] \neq 0$ or both $\alpha[i,p]$ and $\gamma[i,p] \neq 0$. This is in sharp contrast with the properties of the currently available absolutely expedient learning algorithm [3] [4] [5] [7] wherein the reward-penalty and reward-inaction algorithms are absorbing barrier type but the inaction-penalty is not. In fact in all the inaction-penalty algorithms of the absolutely expedient type known so far [3], every state in S_M is an absorbing state. Our modified definition of strong absolute expediency is in fact motivated by the existence of the absorbing barrier algorithms of the reward-penalty, reward-inaction and inaction-penalty types.

Remark 5: For some $1 \leq j \leq M$ if $p_j(k) = 0$, then it follows from (3) and (C.1) - (C.2) that $p_j(k^*) = 0$ for all $k^* \geq k$. In other words, during learning process if p(k) reaches the boundary $p_j = 0$ of the simplex S_M, then p(k) will continue to reamin in that boundary.

Henceforth in this paper we will only be concerned with the absorbing barrier learning algorithms, that is, algorithm (3) under the conditions (A.1) - (A.2) of theorem 1.

Theorem 2: Necessary and sufficient conditions for the absorbing barrier learning algorithm (3) to be strongly absolutely expedient are:

(S.1) $$\sum_{j \neq i} p_j\, \beta[i,j,p] = \sum_{j \neq i} p_j\, \beta[j,i,p]$$

and

(S.2) $$\sum_{j \neq i} p_i (1 - p_j)\, \Delta[i,j,p] = \sum_{j \neq i} p_j (1 - p_i)\, \Delta[j,i,p]$$

for all i = 1, 2, ...M

Sufficiency: Define $\Delta x(k) = x(k+1) - x(k)$ and let

$$\Delta\eta(k) = E[\Delta\eta(k) \mid p(k)] = \sum_{i=1}^{M} E[\Delta p_i(k) \mid p(k)]\, d_i \qquad (11)$$

It can be seen by direct computation that

$$E[\Delta p_i(k) \mid p(k)=p] = p_i(1-p_i)\,d_i\,\alpha[i,p] - p_i^2 c_i\,\gamma[i,p] - \sum_{j\neq i} p_j p_i d_j\,\beta[j,i,p] + \sum_{j\neq i} p_j(1-p_i)c_j\,\delta[j,i,p] \quad (12)$$

Substituting (12) in (11) and inview of (C.1) and (C.2) we obtain

$$\Delta\eta(k) = \Delta\eta_1(k) + \Delta\eta_2(k) \quad (13)$$

where

$$\Delta\eta_1(k) = \sum_{i=1}^{M} p_i d_i^2 \sum_{j\neq i} p_j\,\beta[i,j,p] - \sum_{i=1}^{M} p_i d_i \sum_{j\neq i} p_j d_j\,\beta[j,i,p] \quad (14)$$

and

$$\Delta\eta_2(k) = -\sum_{i=1}^{M} p_i d_i c_i \sum_{j\neq i}\delta[i,j,p](1-p_j) + \sum_{i=1}^{M}(1-p_i)d_i \sum_{j\neq i} p_j c_j\,\delta[j,i,p] \quad (15)$$

After simplification it can be shown that [23]

$$\Delta\eta_1(k) = 1/2 \sum_{i=1}^{M}\sum_{j>i} p_i p_j (d_i-d_j)^2\,[\delta[i,j,p] + \delta[j,i,p]]$$

$$\geq 0 \text{ with equality holding only if } p \in V_m \quad (16)$$

Similarly it can be shown that [23]

$$\Delta\eta_2(k) = 1/2 \sum_{i=1}^{M}\sum_{j>i} p_i(1-p_j)(d_i-d_j)^2\,[\delta[i,j,p] + \delta[j,i,p]] \quad (17)$$

$$\geq 0 \text{ with equality holding good only if } p \in V_m$$

From (16) and (17) sufficiency follows.

Necessity: $\Delta\eta(k)$ can be represented as a quadratic and linear term in the vector D as follows:

$$\Delta\eta(k) = D^T A D + D^T B \quad (18)$$

where $A = [A_{ij}]$ and $B = [B_1, B_2, \ldots, B_m]^T$ with

$$A_{ii} = p_i(1-p_i)\,\delta[i,p] + p_i^2\,\gamma[i,p]$$

$$A_{ij} = -[p_i p_j\,\beta[j,i,p] + (1-p_i)\,p_j\,\delta[j,i,p]]$$

and

$$B_i = p_i^2\,\gamma[i,p] + (1-p_i)\sum_{j\neq i} p_j\,\delta[j,i,p]$$

for all $i,j = 1,2,\ldots M$

From the definition of strong absolute expediency it follows that $\Delta\eta(k)$ attains its minimum value zero either when all d_i are equal or when $p \in V_m$. Since every member of V_m is absorbing, on V_m, it is easily seen that $\Delta\eta(k)$ attains its minimum value for all admissible D vectors. In the following we shall derive conditions for the minimum of $\Delta\eta(k)$ when $d_i = d$ for all $i = 1,2,\ldots,M$, $0 < d < 1$. Necessary conditions for minimum are obtained by setting the derivative of $\Delta\eta(k)$ (with respect to d_i) at the point $d_i = d$ for all $i = 1,2,\ldots,M$ equal to zero, that is,

$$\left.\frac{\partial\Delta\eta}{\partial d_i}\right|_{d_i=d} = 0 \text{ for all } i = 1,2,\ldots,M \quad (19)$$

From (18), the equation (19) take the form

$$(A + A^T)\,d\,1 + B = 0 \quad (20)$$

where $1 = (1,1,\ldots,1)^T$ is an M dimensional column vector of all ones.

Rewriting (20) we get

$$d\,K_i + L_i = 0$$

for all $i = 1,2,\ldots,M$ and all $0 < d < 1$

where

$$K_i = p_i(1-p_i)\,\delta[i,p] + p_i^2\,\gamma[i,p] - p_i\sum_{j\neq i} p_j\,\beta[j,i,p] - (1-p)\sum_{j\neq i} p_i\,\delta[j,i,p]$$

and

$$L_i = p_i^2\,\gamma[i,p] + (1-p_i)\sum_{j\neq i} p_j\,\delta[j,i,p].$$

(21) is true for all $0 < d < 1$ only if $L_i = 0$ and $K_i = 0$ for all $i = 1,2,\ldots,M$.

$L_i = 0$ leads to the condition

$$p_i^2\,\gamma[i,p] = (1-p_i)\sum_{j\neq i} p_j\,\delta[j,i,p]. \quad (22)$$

Using (C.2), from (22) we obtain

$$p_i\sum_{j\neq i}(1-p_j)\,\delta[i,j,p] = (1-p_i)\sum_{j\neq i} p_j\,\delta[j,i,p]. \quad (23)$$

which in fact is (S.2). Substituting (23) in $K_i = 0$, we obtain

$$p_i(1-p_i)\,\delta[i,p] = p_i\sum_{j\neq i} p_j\,\delta[j,i,p].$$

Once again, using (C.1) we get $p_i\sum_{j\neq i} p_j\,\delta[i,j,p] = p_i\sum_{j\neq i} p_j\,\delta[j,i,p]$

which is the same as (S.1). Hence the theorem.

Corollary 1: The Markov process $\{p(k)\}$, $k \geq 0$ as generated by the algorithm (3) under the conditions (C.1) - (C.2), (A.1) - (A.2) and (S.1) - (S.2) converges to V_M with probability one.

Proof: Since $\{p(k)\}$ is a Markov process, from theorem (2) we get

$$E[\Delta\eta(k) \mid P(r): 0 < r < k] = E[\Delta\eta(k) \mid P(k)] \geq 0$$

This inturn implies that $\{\eta(k)\}$, $k \geq 0$ is a submartingale [22]. By martingale theorem $\lim_{k\to\infty}\eta(k)$ exists and hence $\lim_{k\to\infty} p(k) = p^*$ exists with probability one. As $\Delta\eta(k) = 0$ only on V_M, it follows that $P^* \in V_M$ with probability one.

4. ε - Optimality: In the previous section it was established that $P^* \in V_M$ with probability one. In this section we set out to quantify the distribution of P^*.

To this end, define

$$\Gamma_i(p) = \text{Prob}[P^* = e_i \mid P(0) = p] \quad (24)$$

for $i = 1,2,\ldots,M$. Notice $\sum_{i=1}^{M}\Gamma_i(p) = 1$ for all $p \in S_M$

In view of the corollary 1 and (24) we obtain, for all $p(0) = p \in S_M^0$

$$\lim_{k\to\infty} E[\eta(k)] = \sum_{i=1}^{M}\Gamma_i(p)\,d_i \quad (25)$$

To compute $\Gamma_i(p)$ we need the following: Let $C[S_M]$ be the class of all continuous functions from S_M to the real line. If $f(.) \in C[S_M]$ define the operator U as follows:

$$U f(p) = E[f(p(k+1)) \mid p(k) = p]$$

Clearly the operator U is linear and positive [12]

Definition 4: A function $f: S_M \to$ real line is called super regular (regular, sub

regular) if

$$f(p) \geq (=, \leq) \; U f(p)$$

for all $p \in S_M$.

With these preliminaries, we now state two important propositions that lead an algorithm for quantifying $\Gamma_i(p)$.

Proposition 1: $\Gamma_i(p)$ is the only continuous solution of the functional equation.

$$U \Gamma_i(p) = \Gamma_i(p) \tag{26}$$

satisfying the boundary condition

$$\Gamma_i(e_i) = 1 \text{ and } \Gamma_i(e_j) = 0 \text{ for all } i,j, i \neq j$$

Notice $\Gamma_i(p)$ satisfying (26) by definition is a regular function. This functional equation is extremely difficult to solve. Hence in the following we establish upper and lower bounds on $\Gamma_i(p)$.

Proposition 2: Let $f_i(p) \in C[S_M]$ be super (sub) regular function with $f_i(e_i) = 1$ and $f_i(e_j) = 0$ for all i,j, $i \neq j$ then

$$f_i(p) \geq (\leq) \; \Gamma_i(p)$$

The proof of these propositions are rather involved and we refer the reader to Norman [12] for an elegant proof.

Thus, if we can find two functions $h_i^{(1)}(p)$ and $h_i^{(2)}(p)$ which are super and sub-regular functions respectively and satisfying the boundary conditions

$$h_i^{(1)}(e_i) = h_i^{(2)}(e_i) = 1, \; h_i^{(\ell)}(e_j) = 0 \text{ for all } j \neq i, \; \ell = 1,2.$$

then from proposition 2 it follows that $h^{(2)}(p) \leq \Gamma_i(p) \geq h_i^{(1)}(p)$

Consider a function

$$\Psi_i[x_i, p] = \text{ExP}[-x_i p_i / \theta]$$

where $x_i > 0$ is a parameter and $\text{ExP}[\chi] = e^{\chi}$

Recognizing the fact that

$$\phi[x_i, p] = \frac{1 - \text{ExP}[-x_i p_i/\theta]}{1 - \text{ExP}[-x_i/\theta]}$$

is sub (super) regular whenever $\Psi_i[x_i, p]$ is super (sub) regular, it follows from propositions 2 that

$$\phi[Y_i, p] \leq \Gamma_i(p) \leq \phi[Z_i, p]$$

where Y_i, Z_i are two constants such that $\phi[Y_i, p]$ and $\phi_i[Z_i, p]$ are sub and super regular function respectively. Note that function $\phi_i[x_i, p]$ depends on just one compenent of P and thus leads to conservative results. However, it gives rise to expressions which are easily manageable.

The problem of getting bounds on $\Gamma_i(p)$ now reduces to one of finding two positive constants Y_i and Z_i such that $\phi_i[Y_i, p]$ is sub and regular and $\phi_i(Z_i, p)$ is super regular. Further, from (25) and the inequality (1) it is clear that for ϵ - optimality we need to concentrate only (on the lower bound)on $\Gamma_i(p)$.

It can be seen that (by dropping the subscript 1 from x, for convenience)

$$U \Psi_1[x, p_1] - \Psi_1[x, p_1] = -x F_1[x,p] \Psi_1[x,p]$$

where

$$F_1[x,p] = p_1 d_1 \alpha[1,p](1-p_1) V[-x \alpha[1,p](1-p_1)]$$

$$-p_1 C_1 V[1,p] p_1 V[x \gamma[1,p] p_1]$$

$$-\sum_{j \neq 1} p_j d_j \beta[j,1,p] p_1 V[x \beta[j,1,p] p_1]$$

$$+\sum_{j \neq 1} p_j C_j \lambda[j,1,p](1-p_1) V[-x \lambda[j,1,p](1-p_1)]$$

and $V[\pi] = (\text{ExP}(\pi) - 1)/\pi$ if $\pi \neq 0$ and 1 if $\pi = 0$.

Clearly, $F_1[x,p] \geq 0$ implies $\phi_1[x,p]$ regular.

Since $\alpha[i,p]$, $\gamma[i,p]$, $\lambda[i,j,p]$ and $\beta[i,j,p]$ are all bounded above by unity and as the $V[.]$ is strictly monotonically increasing, it follows

$F_1[x,p] \geq 0$ if

$$G[x,p] \triangleq \frac{V[-x(1-p_1)]}{V[x p_1]} \geq \frac{p_1^2 C_1 \gamma[1,p] + \sum_{j \neq 1} p_1 p_j d_j \beta[j,1,p]}{\sum_{j \neq 1} p_j (1-p_1) C_j \lambda[j,1,p] + p_1 (1-p_1) d_1 \alpha[1,p]}$$

It can be shown using the properties of the $V[.]$ function [4] [12] that

$$G[x,p] \geq \frac{1}{V[x]}$$

Further, using (C.1) - (C.2) and (S.1) - (S.2) it can be seen that

$$e^* \triangleq \frac{C_1 A + d_2 B}{C_M A + d_1 B} \geq \frac{p_1^2 C_1 \gamma[1,p] + \sum_{j \neq 1} p_1 p_j d_j \beta[j,1,p]}{\sum_{j \neq 1} p_j (1-p_1) C_j \lambda[j,1,p] + p_1 (1-p_1) d_1 \alpha[1,p]}$$

where $A = \sum_{j \neq 1} p_j (1-p_1) C_j \lambda[j,1,p]$ and $B = \sum_{j \neq 1} p_1 p_j d_j \beta[j,1,p]$

In view of the inequality (1) it follows that $e^* \leq 1$. From (27) and (28)

$$F_1[x,p] \geq 0 \text{ if } \frac{1}{V[x]} \geq e^* \tag{29}$$

Since $e^* < 1$; $V[x] = \frac{1}{e^*}$ has an unique solution $x = y > 0$ so that $\phi[y,p]$ is subregular

Remark 7: For the reward-incation algorithm (refer remark1) e^* reduces to $\frac{d_2}{d_1}$ and for the inaction - penalty algorithm e^* reduces to $\frac{C_1}{C_M}$. Thus in either of these special cases there exists an unique solution $x=y > 0$ for the equation $V[x] = \frac{1}{e^*}$ and hence in both these special cases there exists a lower bound on $\Gamma_1(p)$. This is in sharp contrast with the existing results in the literature where in for the inaction-penalty (absolutely expedient) algorithms [4] no lower bound on $\Gamma_1(p)$ has been established. One of the reasons for this anamoly is that none of the currently available inaction-penalty (absolutely expedient) algorithms [4] are of the absorbing barrier type.

Having established the lower bound on $\Gamma_1(p)$ we now state our main result.

Theorem 3: For every $\epsilon > 0$ and $p(0) = p \in S_M^o$ there esists $0 < \theta^* < 1$ such that for $0 < \theta < \theta^*$, the learning algorithm under the conditions (C.1) - (C.2), (A.1) - (A.2) and (S.1) - (S.2) is such that

$$\lim_{k \to \infty} | E[\eta(k)] - \lambda_1 | < \epsilon \tag{30}$$

Proof: From corollary 1 it follows that the limit in the left hand side of (30) exists

From (25)

$$\lim_{k\to\infty} E[\eta(k)] = \Gamma_1(p) d_1 + \sum_{j=1} \Gamma_j(p) d_j$$

That is,

$$\lim_{k\to\infty} E[\eta(k) - d_1] \leq (1-\Gamma_1(p))\quad (d_2 - d_1)$$

Since $\emptyset[7,p] \leq \Gamma_1(p)$, Combining this with the above,

$$\lim_{k\to\infty} |E[\eta(k)] - d_1| \leq [1-\emptyset_1[y,p]]\ |(d_2 - d_1)|$$

For any $p \in S_M^o$, we know that

$$\lim_{\theta\to 0} \emptyset_1[x,p] = 1$$

Combining these we see that for any $\epsilon > 0$ there exists $0 < \theta^* < 1$ such that for all $0 < \theta < \theta^*$

$$\lim_{k\to\infty} |E[\eta(k)] - d_1| < \delta\quad |(d_2 - d_1)|$$

The theorem follows by choosing $\delta = \epsilon / |(d_2 - d_1)|$.

5. Conclusions: A new class of absorbing barrier absolutely expedient learning algorithms, whose behaviour under the action of success and failure are identical, is introduced. Conditions for the ϵ - optimality of this class of learning algorithms are derived.

6. References:

[1] K.S. Narendra and M.A.L. Thatachar. "Learning Automata - A survey. " IEEE Transactions on Systems, Man, and Cybernetics. Vol. 4, pp. 323-334, 1974.

[2] K.S. Narendra and S. Lakshmivarahan. "Learning Automata - A critique." Journal of Cybernetics and Information Sciences - Special Issue on Learning Automata, Vol. 1, pp. 53-66, 1978.

[3] S. Laksmivarahan and M.A.L. Thatachar. "Absolutely Expedient Learning Algorithms for Stochastic Automata." IEEE Transactions on Systems, Man and Cybernetics, Vol. 3, pp. 281-286, 1973.

[4] S. Laksmivarahan and M.A.L. Thathachar. "Bounds on the Probability of Convergence of Learning Automata.: IEEE Transactions on Systems, Man and Cybernetics. Vol. 6, pp. 756-763, 1976.

[5] H. Aso and M. Kimura. "Absolute Expediency of Learning Automata." Information Sciences, Vol. 17, pp. 91-112, 1979.

[6] U. Herkenrath, D. Kalin and S. Lakshmivarahan. "On a General Class of Absorbing Barrier Learning Algorithms." School of EECS Technical report - 8001 University of Oklahoma, March, 1980. (Also in Information Sciences Vol. 24 pp. 255-263, 1981.

[7] Y. Sawaragi and N. Baba. "Two ϵ- optimal Non-Linear Reinforcement Schemes for Stochastic Automata." IEEE Transactions on Systems, Man and Cybernetics. Vol. 4, pp. 126-131, 1974.

[8] K.S. Fu. "Stochastic Automata Models for Learning Systems." in Computers and Information Sciences II (Ed) by J.T. Tou, Academic Press, 1967.

[9] G.J. McMurtry and K.S. Fu. "A Variable Structure Automaton Used in Multimodal Search Technique. " IEEE Transactions on Automatic Control, Vol 11, pp. 379-387, 1966.

[10] V.I. Varshavskii and I.P. Vorontsova. "On the Behaviour of Stochastic Automata with Variable Structure, " Automation and Remote Control, Vol. 24, pp. 327-333, 1963.

[11] M.L. Tsetlin. "Automaton Theory and Modelling of Biological Systems" Academic Press, 1973.

[12] M.F. Norman. "On a Linear Model with Two absorbing Barriers." Journal of Mathematical Psychology. Vol. 5, pp. 225-241, 1968.

[13] M.F. Norman. Markov Processes and Learning Models. Academic press, New York, 1972.

[14] M. Iosifescu and R. Theodorescu. "Random Process and Learning." Springer Verlag, 1969.

[15] S. Lakshmivarahan. " ϵ - optimal Learning Algorithms - Non Absorbing Barrier Type." Technical Report, EECS 7901, School of EECS, University of Oklahoma, February, 1979.

[16] I.H. Witten. " The Apparent Conflict Between Estimation and Control - A Survey of Two Aremd-Bandit Problem." Journal of Franklin Institute Vol. 301, pp. 161-189, 1976.

[17] B. Chandrasekaran and D.W.C. Shen. "On Expediency and Convergence in Variable Sturcture Automata." IEEE Transactions on Systems Science and Cybernetics. Vol. 4, pp. 52-60, 1968.

[18] R.R. Bush and F. Mosteller. Stochastic Models for Learning. John Wiley, New York, 1958.

[19] R. Viswanathan and K.S. Narendra. "Expedient and Optimal Variable Structure Stochastic Automata." Dunharm Lab. TR-37, 1970. Yale University.

[20] I.J. Shapiro and K.S. Narendra. "Use of Stochastic Automata for Parameter Self Optimisation with Multimodal Performance Criteria." IEEE Transactions on Systems, Man and Cybernetics. Vol. 5, pp. 352-360, 1969.

[21] S. Lakshmivarahan. Learning Algorithms: Theory and Applications. Springer Verlag, New York, 1981.

[22] J.L. Dobb. Stochastic Processes. John Wiley, 1955.

[23] M.R. Meybodi and S. Lakshmivarahan. " ϵ- Optimality of a general class of absorbing barrier learning algorithms." School of EECS Technical Report, August 1981. University of Oklahoma, Norman, Oklahoma, U.S.A.