ERGODIC LEARNING AUTOMATA CAPABLE OF INCORPORATING APRIORI INFORMATION

B. J. Oommen

SCS-TR-89
May 1986

School of Computer Science
Carleton University
Ottawa, Ontario
K1S 5B6 Canada

A preliminary version of some of these results will be presented as an invited paper at the 1986 IEEE International Conference on Systems, Man and Cybernetics, Atlanta, Georgia.

This research was partially supported by the Natural Science and Engineering Research Council of Canada.

Ergodic Learning Automata Capable Of Incorporating Apriori Information*

B. J. Oommen†

Abstract

We consider learning automata which update their action probabilities on the basis of the responses they get from a random environment. The automata update the probabilities whether the environment responds with a reward or a penalty. Learning automata are said to be ergodic if the distribution of the limiting action probability vector is independent of the initial distribution.

In this paper, we present an ergodic scheme which can take into consideration apriori information about the action probabilities. This is the only reported scheme in the literature capable of achieving this. The mean and the variance of the limiting distribution of the automaton is derived, and it is shown that the mean is **not** independent of the apriori information. Further, it is shown that the expressions for the above quantities are general cases of the corresponding quantities derived for the familiar L_{RP} scheme. Finally, it is shown that by constantly updating the parameter quantifying the apriori information, a resultant linear scheme can be obtained. This scheme is indeed counter-intuitive, for it shown to be of a **Reward-Reward** flavour and is yet absolutely expedient. This demonstrates that Absolutely Expedient schemes have far more general properties that those known in the past [8].

---

* A preliminary version of some of these results will be presented as an invited paper at the 1986 IEEE International Conference on Systems, Man and Cybernetics, Atlanta, Georgia. Partially supported by the Natural Sciences and Engineering Research Council of Canada.

† School of Computer Science, Carleton University, Ottawa, Canada K1S 5B6.

I. Introduction

Learning automata have been extensively studied by researchers in the area of adaptive learning and biological modelling [20, 21]. The intention is to design a learning machine which interacts with an environment and which dynamically learns the optimal action which the environment offers. The literature on learning automata is extensive. We refer the reader to a review paper by Narendra and Thathachar [11] and a recent excellent book by Lakishmivarahan [5] for a review of the various families of learning automata. The latter reference also discusses in fair detail some of the applications of learning automata which include game playing [7], pattern recognition and hypothesis testing [11], priority assignment in a queueing system [9] and telephone routing [12, 13]. Other applications not found in [5] include the computation of optimal discriminant functions [19] and object partitioning [16].

The learning process of the automaton can be described as follows. Consider Fig. I. The environment with which the automaton interacts offers the latter a finite set of actions. The automaton is constrained to choose one of these actions. Once the action is chosen, the automaton is penalized by the environment, the penalty probability being dependent on the action chosen. A learning automaton is one which learns the action with the minimum penalty probability and which ultimately chooses this "more frequently" compared to the other actions.

Broadly speaking, learning automata can by classified into two categories: fixed structure and variable structure automata. A fixed structure automaton is one whose transition and output functions are time invariant. Examples of such automata are the Tsetlin, Krylov and Krinsky automata [20, 21]. By far, most of the research in this area has involved the second category, namely, Variable Structure Stochastic Automata (VSSA). Automata in this category possess transition and output functions which evolve as the learning process proceeds. It can be shown that a VSSA is completely defined by a set of **action probability** updating functions [10, 11,23].

Learning automata can also be broadly classified in terms of their Markovian representations. Generally speaking, learning automata are either ergodic [1-3, 5, 6, 11-15,17,18, 20] or possess absorbing barriers [5, 8, 11, 14]. Automata in the former class converge with a distribution which is **independent** of the initial distribution of the action probabilities. This feature is desirable when interacting with a nonstationary environment - for the automaton does not "lock itself" into choosing any one action. However, if the environment is stationary an automaton with an absorbing barrier is preferred. Various absolutely expedient schemes which ideally interact with such environments have been proposed in the literature [5, 8, 10, 11]. Since we shall be concentrating on ergodic schemes, we would refer the reader to some of the ingenious schemes of this class especially due to Lakishmivarahan [5], Flerov [3], Tsypkin and Poznyak [17, 22], El-Fatteh [1, 2] and Thathachar et al [15, 18]. More recently, we [14] have proposed some ergodic schemes that update probabilities in a discretized probability space and yet have excellent learning capabilities.

The most simple ergodic scheme known is probably the Linear Reward-Penalty (L_{RP}) scheme. In this case the action probability decrements are made linearly proportional to the probabilities themselves and are made irrespective of the response of the environment. The limiting probability vector converges in distribution, and the form of this distribution has been known only for the symmetric version of the L_{RP} scheme which is a one-parameter probability updating algorithm. The latter scheme is a special case of a variety of learning automata which were shown to be ergodic in the mean [15, 18]. In this case, the mean action probability vector is the state probability of an ergodic Markov chain [15]. In [15], the general problem of a two action scheme being EM was studied; necessary and sufficient conditions were derived for the property to hold. These properties were later generalized for the multi-action case in [18].

Although there is an extensive work done in the area of ergodic schemes, there is one problem common to all : none of these schemes can take into consideration the apriori information about the action probabilities. This is because, in the case of all the ergodic automata known, the limiting **distribution** of the action probability vector is indeed independent of the initial action probabilities themselves.

In this paper we present an ergodic scheme which overcomes this drawback. We first present an analytical method for specifying the apriori information about the actions, and prove that the resulting scheme is ergodic. Subsequently, it is shown that whereas the mean of the limiting action probability distribution is independent of the initial action probability vector, the mean of this distribution is **not** independent of the **apriori** information. The scheme is also shown to be ergodic in the mean. Expressions for the mean and the variance of the limiting distribution are derived and it is shown that these are general cases of the well known **Linear Reward-Penalty** Scheme. Experimental results involving this learning automaton are also presented.

Finally, it is shown that by constantly updating the parameter describing the apriori information, a resultant linear scheme can be obtained. The latter scheme completely baffles the theory of adaptive learning for it is of a **Reward-Reward** flavour and is **simultaneously Absolutely Expedient**. More than a decade ago, Lakshmivarahan and Thathachar [8] derived necessary and sufficient conditions for schemes to be absolutely expedient. The new learning automaton introduced in this paper does not satisfy their conditions and is yet absolutely expedient. This demonstrates that Absolutely Expedient schemes have far more general properties than those which have been known in the past. As opposed to what was believed earlier, we show that absolutely expedient schemes need not merely be of a Reward-Penalty nature. Indeed, by making the automaton systematically **increase** the probability of an action chosen even when it is **penalized**, the designer can obtain absolute expedient characteristics. This is completely counter-intuitive.

The organization of the paper is as follows. We first introduce the terminology used in the literature and explain the Linear Reward-Penalty (L_{RP}) automaton. We then present the scheme which incorporates the apriori information and derive its asymptotic properties. We also prove some fundamental theorems regarding the rate of convergence of the scheme and of the limiting action probabilities. We shall then present a linear Reward-Reward automaton which is absolutely expedient. We shall also show why this does not satisfy the conditions [8] "required" for absolute expediency and shall thus illustrate that the class of absolute expedient automata is far more general than has been believed. We finally present simulation results of some of the automata we have introduced.

I. 1 Fundamentals

The automaton considered in this paper (Fig. 1) selects an action a(n) at each instant 'n' from a finite action set $\{a_i \mid i = 1 \text{ to } R\}$. The selection is done on the basis of a probability distribution **p**(n), an R x 1 vector, where, $\mathbf{p} = [p_1, p_2, ..., p_R]^T$, and,

$$p_i(n) = Pr[a(n) = a_i], \qquad \text{with}$$

$$\Sigma_{1 \le i \le R} \; p_i(n) = 1 \qquad \text{for all n.} \qquad (1)$$

The selected action serves as the input to the environment which gives out a response b(n) at time 'n'. b(n) is an element of B = {0, 1}. The response '1' is said to be a 'penalty'. The environment penalizes the automaton with the penalty c_i, where,

$$c_i = Pr[b(n) = 1 \mid a(n) = a_i] \qquad (i = 1 \text{ to } R). \qquad (2)$$

Thus the environment characteristics are specified by the set of penalty probabilities $\{c_i\}$ ($i = 1$ to R). On the basis of the response $b(n)$, the action probability vector $\mathbf{p}(n)$ is updated and a new action is chosen at $(n+1)$.

The $\{c_i\}$ are unknown initially and it is desired that as a result of interaction with the environment the automaton arrives at the action which evokes the minimum penalty response in an expected sense. It may be noted that if

$$c_L = \min (c_i) \qquad (3)$$

this is attained when $p_L = 1$ and $p_i = 0$ for all $i \neq L$.

The average penalty received at the n^{th} time instant is

$$M(n) = \Sigma_{1 \leq i \leq R} \; p_i(n)\, c_i$$

With no apriori information, the automaton chooses the actions with equal probability. The expected penalty is thus initially M_0.

$$M_0 = \Sigma_{1 \leq i \leq R} \; p_i(0)\, c_i$$

$$= 1/R \left[\Sigma_{1 \leq i \leq R} \; c_i \right] \qquad (\text{since } p_i(0) = 1/R).$$

An automaton is said to learn **expediently** if, as time tends towards infinity, the expected penalty is less than M_0. The automaton is **absolutely expedient** if $E[M(n+1) \mid \mathbf{p}(n)] < M(n)$. Note that in this case $M(n)$ is a supermartingale [8].

It is ε-optimal if for any arbitrary $\varepsilon > 0$, in the limit, $E[M(n)] < c_{min} + \varepsilon$, where $c_{min} = \min\{c_i\}$. This can be achieved by a suitable choice of some parameter of the automaton. Thus the limiting value of $E[M(n)]$ can by as close to c_{min} as desired.

Throughout this paper, we shall be considering 2-action automata, i.e., with R=2. The properties of the analogous R-action schemes are currently being investigated.

I. 2 The L_{EM} Scheme

The two-action Linear Reward-Penalty (L_{RP}) scheme, which is a probability updating algorithm having two parameters a, b < 1, is given below for $i, j = 1, 2$; $i \neq j$.

$$
\begin{aligned}
p_i(n+1) &= bp_i(n) && \text{if } a(n) = a_j \text{ and } b(n) = 0 \\
&= (1-b) + bp_i(n) && \text{if } a(n) = a_i \text{ and } b(n) = 0 \\
&= ap_i(n) && \text{if } a(n) = a_i \text{ and } b(n) = 1 \\
&= (1-a) + ap_i(n) && \text{if } a(n) = a_j \text{ and } b(n) = 1
\end{aligned}
$$

To simplify the notation, unless explicitly stated we use p_i to refer to the probability $p_i(n)$. In this form of the L_{RP} scheme, $E[p_i(n+1) \mid p_i]$ has the expression :

$$E[p_i(n+1) \mid p_i] = p_i^2 (a-b)(c_i-c_j) + p_i\{1-c_i(1-b)- c_j(1+b-2a)\} + c_j(1-a)$$

where $i, j = 1, 2$ and $i \neq j$.

Observe that $E[p_i^k(n+1)]$ is dependent on $E[p_i^r(n+1)]$ for some $r \geq k$ for all $k > 1$. Because of this the form of the limiting distribution of the general L_{RP} scheme is unknown. However, if $b = a$, the term containing p_i^2 disappears from the above expression and renders it ergodic in the mean [18]. Using distance diminishing operators [5] it can be shown that when $b = a$ the limiting distribution of p_i is normal with the following mean and variance :

$$E[\, p_i(\infty)\,] = c_j / (c_i + c_j) \qquad i, j = 1, 2 \,;\, i \neq j$$

$$\mathrm{Var}[\, p_i(\infty)] = \frac{c_1\, c_2\, (1-a)}{(c_1 + c_2)^2\, [(1-a) + 2a(c_1 + c_2)]} \tag{4}$$

We refer to the symmetric L_{RP} scheme with $b = a$ as the L_{EM} scheme (for Linear scheme possessing Ergodicity of the Mean). A two parameter EM scheme whose properties are superior to those of the L_{EM} scheme has also been discovered [18].

II. Learning With Apriori Information

The general property common to all of the ergodic learning schemes is that the final distribution of the action probability vector is independent of the initial action probabilities. Though this is advantageous in that the automaton does not get locked into an action, the main drawback of all these schemes is that there is no way of using apriori information in the learning process. This is not true of the class of absorbing schemes [5], for in those schemes, the apriori probabilities can be represented in terms of the initial action probabilities. Indeed, for these schemes, $\Gamma_i(\mathbf{p})$, the limiting action probability of choosing a_i is dependent on the initial probability vector, **p**.

Our intention here is to devise a scheme which is ergodic and which simultaneously can incorporate the apriori information the designer has about the environment. Before we proceed to present the scheme, we shall consider a basic question : How does using apriori information differ from continuously using the responses of the environment to the automaton?

Consider the scenario in which an **absorbing** learning automaton interacts with the environment for numerous experiments. In each experiment, let us suppose the initial probability vector is $\mathbf{p}(0) = [0.5, 0.5]^T$. Thus, [5, 10], the expected probability of converging to action a_1 is well defined as Γ_1 [$\mathbf{p}(0)$]. This means that in any particular experiment, the automaton will converge to a_2 with probability Γ_2 [$\mathbf{p}(0)$], which is $1 - \Gamma_1$ [$\mathbf{p}(0)$]. Let us consider the scenario that after a large number of experiments the automaton has converged in a_1 60% of the time and in a_2 40% of the time. Thus, this itself yields **apriori** information about the environment **before** the next experiment is conducted. In other words, it **could** be advantageous to run the next experiment with another initial probability vector other than $[0.5, 0.5]^T$. For example, a possible starting action probability vector would be $[0.6, 0.4]^T$. Observe that this "pre-biasing" of the automaton's choice of a_1 will lead to a higher chance of converging to a_1, and also to a faster convergence.

It is easy to see that such a scheme **can** be implemented with an absorbing automaton. However, for any of the known ergodic schemes, such a strategy would be impossible because $E[p_i(\infty)]$ will indeed be independent of the initial action probability vector. This is exactly where we shall make our first contribution.

The scheme which we introduce shall be called the **Apriori Linear Reward-Penalty** (AL_{RP}) scheme. The automaton operates as follows for i, j = 1, 2,

$i \neq j$. Whenever an action a_i is chosen and rewarded, the probability p_j is decremented linearly to ap_j, and hence, the probability p_i is correspondingly incremented. This, of course, is identical to the way the L_{RP} automaton operates. However, on choosing a_i and being **penalized**, the automaton updates p_j to $ap_j + q_j$, where q_j is a **constant** and it reflects the apriori information about action a_j. Correspondingly, p_i gets updated to maintain the vector **p** as a probability vector.

Formally, the scheme is given below for $i, j = 1, 2$; $i \neq j$. Let $0 < a < 1$. Then,

$$
\begin{aligned}
p_i(n+1) \quad &= ap_i && \text{if } a(n) = a_j \text{ and } b(n) = 0 \\
&= 1 - ap_j && \text{if } a(n) = a_i \text{ and } b(n) = 0 \\
&= ap_i + q_i && \text{if } a(n) = a_j \text{ and } b(n) = 1 \\
&= 1 - ap_j - q_j && \text{if } a(n) = a_i \text{ and } b(n) = 1 \qquad (5)
\end{aligned}
$$

Observe that the scheme always increments the action probability if the action is rewarded. However, it does not necessarily decrement the action probability on being penalized. Observe too that the scheme reduces to the L_{EM} scheme if $q_1 = q_2 = 1 - a$. We shall now consider how to specify q_i and q_j so as to incorporate the apriori knowledge of the actions.

Consider the case when a_j is chosen and penalized. In this case, p_i is updated to have the value $ap_i + q_i$. Thus, q_i must be less than $1 - a$ to ensure that $p_i(n+1)$ is always a probability. Similarly, q_j must be less than 1-a. In our strategy, q_i and q_j are appropriately chosen so that **their ratio** is equal to the ratio of the apriori action probabilities x_i and x_j. Explicitly, we require that, if $\{x_i\}$ is the set of apriori probabilities,

$$q_i / q_j = x_i / x_j, \qquad \text{and ,}$$

$$0 < q_i, q_j < 1\text{-}a. \qquad (6)$$

We shall now prove some limiting properties of the AL_{RP} scheme. To simplify the notation, we shall consider the probability $p_1(n)$. Similar expressions can easily be obtained for $p_2(n)$.

Theorem I

The AL_{RP} automaton is Ergodic in the mean.

Proof : It is required to prove that the following vector equation holds:

$$E[\mathbf{p}(n)] = M^T E[\mathbf{p}(n)]$$

where, M is a stochastic matrix with no absorbing barriers. Observe that by virtue of (5), $p_1(n+1)$ has the following distribution :

$p_1(n+1)$	$= ap_1$	with probability $p_2(1-c_2)$
	$= 1 - ap_2$	with probability $p_1(1-c_1)$
	$= ap_1 + q_1$	with probability $p_2 c_2$
	$= 1 - ap_2 - q_2$	with probability $p_1 c_1$

Thus, $E[p_1(n+1) \mid \mathbf{p}] = ap_1p_2(1-c_2) + (1-ap_2)\,p_1\,(1-c_1)$
$+ (ap_1+q_1)p_2c_2 + (1-ap_2-q_2)p_1c_1.$

Quite amazingly, all the quadratic terms in p_1^2, p_1p_2 and p_2^2 cancel.

Thus, $E[p_1(n+1) \mid \mathbf{p}] = (1-q_2c_1)\,p_1 + (q_1c_2)\,p_2.$

Taking expectations again we get,

$$E[p_1(n+1)] = (1 - q_2c_1)\,E[p_1(n)] + (q_1c_2)\,E[p_2(n)]. \quad (7)$$

Similarly, if we computed $E[p_2(n+1)]$ we would obtain,

$$E[p_2(n+1)] = (q_2c_1)\,E[p_1(n)] + (1 - q_1c_2)\,E[p_2(n)] \quad (8)$$

Combining (7) and (8) we get the matrix equation :

$E[\mathbf{p}(n+1)] = M^T E[\mathbf{p}(n)]$, where,

$$M = \begin{bmatrix} 1-q_2c_1 & q_2c_1 \\ q_1c_2 & 1-q_1c_2 \end{bmatrix} \quad (9)$$

Clearly M is a stochastic matrix which has no absorbing barriers if $0 < q_1, q_2 < 1$. Hence the theorem!

We now prove that the limiting value of $E[p_1(n)]$, though independent of the initial value of $\mathbf{p}(0)$, is **not** independent of the value of apriori probabilities specified in terms of q_1 and q_2.

Theorem II

For the AL_{RP} scheme, the limiting value of $E[\mathbf{p}(\infty)]$ is independent of the initial value of $\mathbf{p}(0)$ but **not** independent of the quantities $\{q_i\}$ specifying the apriori information.

Proof : From Theorem I, we know that the expected probability vector obeys the following stochastic equation :

$$E[\mathbf{p}(n+1)] = M^T E[\mathbf{p}(n)]$$

where M is given by (9).
From the elementary theory of Markov chains, [4], we know that the limiting value of $E[\mathbf{p}(n)]$ is obtained by solving :

$$E[\mathbf{p}(\infty)] = M^T E[\mathbf{p}(\infty)]. \qquad (10)$$

Expanding (10) for $E[p_1(\infty)]$ we obtain,

$$E[p_1(\infty)] = (1 - q_2c_1) E[p_1(\infty)] + (q_1c_2) E[p_2(\infty)].$$

Since $E[p_2(\infty)] = 1 - E[p_1(\infty)]$, we have,

$$E[p_1(\infty)] = (1 - q_2c_1) E[p_1(\infty)] + (q_1c_2)(1 - E[p_1(\infty)]).$$

Simplifying, we obtain, $E[p_1(\infty)] = q_1c_2 / (q_1c_2 + q_2c_1)$

Clearly $E[p_1(\infty)]$ is independent of $\mathbf{p}(0)$. But, it is not independent of $\{q_i\}$. Hence the theorem!

Remarks

1. Observe that $E[p_1(\infty)]$ can be equivalently written as :

$$E[p_1(\infty)] = \frac{q_1 / c_1}{q_1 / c_1 + q_2 / c_2}.$$

Thus $E[p_1(\infty)]$ is merely a weighted function of the reciprocals of the penalty probabilities, and the weighting parameters are exactly the quantities $\{q_i\}$. Since the latter are proportional to the apriori probabilities $\{x_i\}$, the weighting parameters can equivalently be considered to be the apriori probabilities themselves.

2. By way of example, let us suppose $c_1 = 0.2$ and $c_2 = 0.8$. Then, with no apriori information $E[p_1(\infty)]$ would be 0.8. However, if the apriori probability vector was $[0.75 \quad 0.25]^T$, the value of $E[p_1(\infty)]$ would be as high as 0.923. Indeed, in general, as the apriori knowledge is more informative, the value of $E[p_1(\infty)]$ gets arbitrarily close to unity.

Corollary I. With no apriori information, the AL_{RP} scheme is always expedient.

Proof : If no apriori information is available, clearly $x_1 = x_2 = 0.5$. Hence, $E[p_1(\infty)] = c_2 / (c_1 + c_2)$. Thus, the limiting penalty $M(\infty)$ is :

$$M(\infty) = 2c_1c_2 / (c_1 + c_2).$$

The result follows since, $M(\infty) < M_0 = (c_1 + c_2) / 2$ for all environments.

Theorem III

The limiting variance of the probability $p_1(n)$ for the AL_{RP} scheme is given by:

$$\mathrm{Var}(p_1(\infty)) = \frac{c_1c_2q_1q_2\,[(1-a)^2 + (c_2q_1 + c_1q_2)(q_1 + q_2 + 2(a-1))]}{(c_2q_1 + c_1q_2)^2\,[(1-a)^2 + (c_2q_1 + c_1q_2)(2a)]} \qquad (11)$$

Proof : Consider the AL_{RP} scheme defined by (5). The updating of $p_1(n)$ hence yields the conditional distribution of $p_1(n+1)$ as :

$$\begin{aligned} p_1(n+1) &= ap_1 && \text{with probability } (1-p_1)(1-c_2) \\ &= 1 - a(1-p_1) && \text{with probability } p_1(1-c_1) \\ &= ap_1 + q_1 && \text{with probability } (1-p_1)\,c_2 \\ &= 1 - a(1-p_1) - q_2 && \text{with probability } p_1\,c_1 \end{aligned} \qquad (12)$$

Considering the second conditional moment of $p_1(n+1)$, we get from (12),

$$E[\,p_1^2(n+1) \mid \mathbf{p}\,] = (ap_1 + q_1)^2 (1-p_1)c_2 + (1 - a(1-p_1) - q_2)^2 (p_1 c_1) + (ap_1)^2 (1-p_1)(1-c_2) + (1 - a(1-p_1))^2 (p_1(1-c_1))$$

Interestingly, all the cubic terms in p_1 cancel. After much simplification we get :

$$E[\,p_1^2(n+1) \mid \mathbf{p}\,] = p_1^2 [\,-2ac_2q_1 - 2ac_1q_2 + 2a - a^2] + p_1 [2ac_2q_1 - c_2q_1^2 - 2c_1q_2 + 2ac_1q_2 + c_1q_2^2 + 1 - 2a + a^2] + c_2q_1^2$$

Taking expectations again and taking the limiting as $n \to \infty$ we get,

$$E[\,p_1^2(\infty)\,] = E[\,p_1^2(\infty)\,] [\,-2a\,(c_2q_1 + c_1q_2) - (1-a)^2 + 1] + E[\,p_1(\infty)\,] [2a(c_2q_1 + c_1q_2) + (1-a)^2 - c_2q_1^2 + c_1q_2^2 - 2c_1q_2] + c_2q_1^2$$

Substituting the expression for $E[\,p_1(\infty)\,]$ from Theorem II and solving for $E[\,p_1^2(\infty)\,]$ yields after considerable manipulation :

$$E[\,p_1^2(\infty)\,] = \frac{c_2q_1\,[2a(c_2q_1 + c_1q_2) + (1-a)^2 + c_1q_2\,(q_1 + q_2 - 2)]}{(c_2q_1 + c_1q_2)\,[2a\,(c_2q_1 + c_1q_2) + (1-a)^2]}$$

But, $\mathrm{Var}[\,p_1(\infty)\,] = E[\,p_1^2(\infty)\,] - E[\,p_1(\infty)\,]^2$. Thus,

$$\mathrm{Var}[\,p_1(\infty)\,] = E[\,p_1^2(\infty)\,] - [\,c_2q_1 / (c_1q_2 + c_2q_1)\,]^2$$

which yields the expression in (11) after further simplification. Hence the theorem!

Remark : When $q_1 = q_2 = (1-a)$, the expression for the variance given by (11) reduces to :

$$\mathrm{Var}[\,p_1(\infty) \mid q_1 = q_2 = 1-a\,] = \frac{c_1c_2\,(1-a)}{(c_1 + c_2)^2\,[(1-a) + 2a(c_1 + c_2)]}$$

Indeed, as expected, this is the variance of $p_1(\infty)$ for the L_{EM} scheme.

Theorem IV

The rate of convergence of the AL_{RP} scheme is controlled completely by the apriori probabilities and penalty probabilities.

Proof : The rate of convergence of the Markov process is determined by the eigenvalue of the stochastic matrix M (defined in (9)) less than unity [4]. Since one eigenvalue is always unity, the trace of M indicates that the second eigenvalue is $1 - q_1c_2 - q_2c_1$. Note that this is **not** dependent on the parameter of the scheme, 'a', but **only** on $\{q_i\}$ and the unknown penalty probabilities.

Remark : If estimates of the penalty probabilities are available (though the actions to which these penalty probabilities correspond to are unknown), it is interesting to note that the designer that specify q_1 and q_2 so as to maximize the rate of convergence. Indeed, this rate of convergence can, in many cases, be maximized so that $E[p_1(n)]$ converges to its final value in **one** step. This can be done, for example, if $(c_1 + c_2) > 1$ and if $q_1 = q_2 = 1\text{-a}$, for in this case, the second eigenvalue can be exactly set to zero. However, note that in general, q_1 and q_2 need not be functions of 'a', except that they must be such that $0 < q_1, q_2 < (1\text{-a})$.

We shall now proceed to study the scheme defined by (5) and shall show that by continuous updating, it naturally leads to a counter-intuitive Reward-Reward scheme that is **absolutely expedient.**

III. A Reward-Reward Absolutely Expedient Scheme

When the field of learning automata was in its infancy, Lakshmivarahan and Thathachar [8] proved a powerful result which classified all absolutely expedient schemes. This was done by proving necessary and sufficient conditions for a learning automaton to be absolutely expedient. Indeed, this result was so all encompassing that it included all the well known linear and nonlinear ε-optimal VSSA of the day.

However, we shall show in this section that their result does not completely define the set of **all** absolutely expedient schemes. The basic underlying premise of the work in [8] was that if an action was chosen and rewarded, the corresponding action probability was incremented by a non-negative amount. Analogously, if the action chosen was penalized, the corresponding action probability was decremented by a non-negative amount. By removing the constraint, we shall show that

Reward-Reward schemes can be absolutely expedient.

Consider the AL_{RP} learning scheme defined by (5). As stated earlier, q_i and q_j are merely quantities that represent the user's apriori information about the actions. Observe that in **any one particular run,** a good estimate for the apriori information of the actions can be obtained by studying p_1 and p_2 themselves. Thus, the scheme that we propose merely continuously updates q_i and q_j as linear functions of p_i and p_j respectively. Explicitly, the scheme is obtained from (5) by substituting q_i and q_j as bp_i and bp_j respectively, where a, b ≥ 0, with 0 < a+b < 1. Thus, the L_{RR} scheme is defined as below :

$$
\begin{aligned}
p_i(n+1) &= ap_i && \text{if } a(n) = a_j \text{ and } b(n) = 0 \\
&= 1 - ap_j && \text{if } a(n) = a_i \text{ and } b(n) = 0 \\
&= (a + b)p_i && \text{if } a(n) = a_j \text{ and } b(n) = 1 \\
&= 1 - (a + b)p_j && \text{if } a(n) = a_i \text{ and } b(n) = 1 \qquad (13)
\end{aligned}
$$

Theorem V

The linear scheme defined by (13) is of a Reward-Reward flavour.

Proof : This result is obvious, since, by virtue of (13), the probability $p_i(n)$ is increased every time a_i is chosen. Correspondingly, the probability $p_i(n)$ is decremented every time a_j ($j \neq i$; i, j = 1, 2) is chosen.

The L_{RR} automaton systematically increases its probability of choosing a_i whenever a_i is chosen. One would be quite amazed if such a "stubborn" behaviour is beneficial. Indeed it is, because the L_{RR} automaton is absolutely expedient.

Theorem VI

The L_{RR} scheme defined by (13) is absolutely expedient.

Proof : By virtue of (13) observe that $p_1(n+1) - p_1(n)$ has the following distribution :

$$
\begin{aligned}
p_1(n+1) - p_1(n) &= -(1-a)\, p_1 && \text{with probability } p_2(1-c_2) \\
&= 1 - ap_2 - p_1 && \text{with probability } p_1(1-c_1) \\
&= -(1 - a - b)\, p_1 && \text{with probability } p_2 c_2 \\
&= 1 - (a + b)\, p_2 - p_1 && \text{with probability } p_1 c_1
\end{aligned}
$$

But $1 - p_1 = p_2$. Hence $E[\, p_1(n+1) - p_1(n) \mid \mathbf{p}]$ has the form :

$$E[p_1(n+1) - p_1(n) \mid \mathbf{p}] = -(1-a)p_1p_2(1-c_2) + (1-a)p_1p_2(1-c_1) - (1-a-b)p_1p_2c_2 + (1-a-b)p_1p_2c_1$$
$$= bp_1p_2(c_2 - c_1). \quad (14)$$

Similarly, $E[p_2(n+1) - p_2(n) \mid \mathbf{p}] = bp_1p_2(c_1 - c_2)$. (15)

Let $\Delta M(n) = E[M(n+1) - M(n) \mid \mathbf{p}]$.
Since $M(n+1) = c_1 p_1(n+1) + c_2 p_2(n+1)$, we have,

$$\Delta M(n) = c_1 E[p_1(n+1) - p_1(n) \mid \mathbf{p}] + c_2 E[p_2(n+1) - p_2(n) \mid \mathbf{p}]$$

Substituting (14) and (15) we obtain,

$$\Delta M(n) = c_1 bp_1p_2(c_2 - c_1) + c_2 bp_1p_2(c_1 - c_2)$$
$$= -bp_1p_2(c_2 - c_1)^2 < 0.$$

Hence, M(n) is a supermartingale and the theorem is proved.

Remarks

(1) Observe that the L_{RR} automaton has exactly two absorbing barriers, namely the unit vectors $[0, 1]^T$ and $[1, 0]^T$.

(2) Apart from being absolutely expedient we also believe that the scheme is ε-optimal. In other words, we believe that if $c_1 < c_2$, the probability that the automaton converges into $[1, 0]^T$ can be made as close to unity as desired. This result is yet unproven but various simulation results validify the conjecture.

(3) The reason why the L_{RR} scheme does not fall into the set of absolutely expedient schemes studied in [8] is because in the latter work, the only linear scheme that satisfied the conditions, required that the automaton was to ignore the penalty responses. This led to the familiar L_{RI} scheme. However, if we remove the constraint and actually permitted probability "increments" on being penalized, it would lead to a Reward-Reward Scheme. The fact that such a scheme is absolutely expedient is quite astonishing. Indeed, it opens the avenue for the whole family of absolutely expedient schemes to be grown to include nonlinear Reward-Reward Strategies too. Certainly, absolutely expedient schemes are far more general that what was known so far. This could be one of the most interesting areas of research in learning theory in the near future.

(4) Psychologically, we believe that the L_{RR} scheme, which we have

proposed, is a very good model for the stubborn learner. This is to our knowledge, the first model for absolutely expedient stubborn learning.

IV Experimental Results

The AL_{RP} scheme has been simulated for various environments in which $c_2 = 0.8$ and c_1 was varied from 0.1 to 0.7. Although the initial starting probability was always $[0.5, 0.5]^T$, the experiments were conducted for various apriori probabilities. To obtain accurate simulation results, each simulation was performed **100** times, and the average results of these experiments is reported below.

In Fig. II, the variation of $\hat{E}\,[\,p_1(\infty)\,]$ is plotted when $c_2 = 0.8$ and c_1 varied from 0.1 to 0.7. In all the cases, 'a' was set equal to 0.9 and q_1 and q_2 were made to be (1-a) times the apriori probabilities x_1 and x_2 respectively. Observe that if $x_1 = x_2 = 0.5$, the resulting scheme is the L_{EM} scheme. In the case when $x_1 = 0.75$ and $x_2 = 0.25$ the final value of $E\,[\,p_1(\infty)\,]$ is always higher than the corresponding figure for the L_{EM} scheme. For example, when $c_1 = 0.1$ and $c_2 = 0.8$, the value for $\hat{E}\,[p_1(\infty)]$ was 0.8948, for apriori probabilities $x_1 = x_2 = 0.5$. The corresponding value when $x_1 = 0.75$ and $x_2 = 0.25$ was 0.9621.

To demonstrate the power of the AL_{RP} scheme, we also considered various values of apriori probabilities x_1, in which the apriori information was **completely misleading**. Clearly, whenever $c_1 < c_2$, any apriori information suggesting that a_2 is the better action is misleading. To study how the AL_{RP} automaton performed in such a scenario various experiments were performed in which $x_1 = 0.25$ and $x_2 = 0.75$ even though $c_1 < c_2$. Even with such erroneous apriori information the automaton had an $\hat{E}[p_1(\infty)\,]$ as high as 0.7418 when $c_1 = 0.1$ and $c_2 = 0.8$. A plot of $\hat{E}\,[p_1(\infty)]$ with c_1 is given in Fig. II. Fig. III shows the variation of the exact (as given by Theorem II) and the estimated values of $E\,[p_1(\infty)]$ for various values of x_1. The two curves are almost identical. Observe too the monotonic increase of $E\,[p_1(\infty)]$ with x_1. As anticipated, the value of the latter quantity equals unity in the limit when x_1 is unity.

To demonstrate the learning behaviour of the automaton with time we have also plotted the variation of $\hat{E}\,[p_1(n)]$ as a function of the time parameter 'n'. This is shown in Fig. IV for the case when $c_1 = 0.4$ and $c_2 = 0.8$ and for various values of x_1. Observe the fast convergence of the scheme.

Analogous to Fig. III, where we plotted the variation of $\hat{E}[p_1(\infty)]$ with x_1, in Fig. V we have plotted the variation of $\hat{Var}[p_1(\infty)]$ with x_1. To do this, $\hat{Var}$ was computed as a function x_1, where,

$$\hat{Var} = [\Sigma_{1 \le j \le N} (p_{1j}^*(\infty) - E[p_1(\infty)])^2] / N$$

In the above expression, N is the number of experiments and $p_{1j}^*(\infty)$ is the final value of p_1 in the j^{th} experiment. Note that we have used the exact value $E[p_1(\infty)]$ in the computation instead of the sample mean of the final value. This is to avoid the errors that would be encountered by ignoring the effect of the variance of the sample mean. Form the graph, we observe that the variance has a peak near the value of $x_1 = 0.2$. Observe too that the variance is zero when $x_1 = 0$ or unity. This is intuitively appealing.

Finally to consider how the variance of the scheme varies with 'a', the **exact** expression for Var $[p_1(\infty)]$, as given by (11), has been plotted as a function of 'a'. Observe that this is quite different from plotting the quantity as a function of x_1. When a = 0, the scheme is purely an "additive" scheme. Similarly, when a = 1, the scheme is the L_{RP} scheme in which the probabilities remain the same and hence the scheme is devoid of learning. In the open interval (0, 1), the value of the variance tends to decrease with 'a'. This is identical to the behaviour of the L_{EM} scheme. Remarkably, however, the decrease seems to be almost linear, as can be seen in Fig. VI.

V. Conclusions

In this paper, we have considered Learning Automata whose structure evolve with time. Such automata, termed as Variable Structure Stochastic Automata (VSSA) are fully defined by action probability updating rules. Variable Structure Stochastic Automata (VSSA) fall into two major classes : those which have absorbing barriers and those which are ergodic.

One of the problems that exists with all the ergodic schemes available in the literature is that there schemes **cannot** take into consideration any **apriori** information about the action probabilities. This is because, in the case of all the ergodic automata known, the distribution of the action probability vector is indeed independent of the initial action probabilities.

In this paper we have presented an ergodic scheme which overcomes this drawback. We first presented an analytical method for specifying the apriori information about the actions, and proved that the resulting scheme is ergodic in the mean. Subsequently, it is shown that whereas the mean of the limiting action probability distribution is independent of the initial action probability vector, the mean of this distribution is **not** independent of the **apriori** information. Various asymptotic properties of this scheme has also been derived.

Finally, it has been shown that by constantly updating the parameter describing the apriori information, a resultant linear scheme can be obtained. The latter scheme is of a **Reward-Reward** flavour and is **simultaneously Absolutely Expedient.** This result demonstrates that Absolutely Expedient schemes are far more extensive than what researchers have thought.

We are currently investigating the multi-action case of ergodic automata incorporating apriori information. Also under investigation is the existence of two-action (and multi-action) **nonlinear** schemes which are of a Reward-Reward nature and simultaneously absolutely expedient.

Acknowledgements

I would like to thank David Ma for implementing the automata for me and David Ng for so painstakingly preparing the manuscript. I am also grateful to James Dean for helping me in the algebraic manipulations required to derive Theorem III.

References

[1] El-Fatteh, Y. M., "Gradient Approach for Recursive Estimation and Control in Finite Markov Chains", Advances in Applied Prob., Vol. 13, 1981, pp.778-803.

[2] El-Fatteh, Y. M., "Multi Automaton Games : A Rationale for Expedient Collective Behavior", Systems and Control Letters, Vol. 1, 1982, pp.332-339.

[3] Flerov, Y. A., "Some Classes of Multi-Input Automata", Journal of Cybernetics, Vol. 2, 1972, pp.112-122.

[4] Isaacson, D. L., and Madson, R. W., "Markov Chains: Theory and Applications",

Wiley, 1976.

[5] Lakishmivarahan, S., "Learning Algorithms Theory and Applications", Springer-Verlag, New York, 1981.

[6] Lakishmivarahan, S., "ε-Optimal Learning Algorithms Non-absorbing Barrier Type", Technical Report EECS 7901, February 1979, School of Electrical Engineering and Computer Sciences, University of Oklahoma, Norman, Oklahoma.

[7] Lakishmivarahan, S., "Two Person Decentralized Team With Incomplete Information", Applied Mathematics and Computation, Vol. 8, pp.51-78, 1981.

[8] Lakishmivarahan, S., and Thathachar, M. A. L., "Absolutely Expedient Algorithms for Stochastic Automata", IEEE Trans. on Syst. Man. and Cybern., Vol. SMC-3, 1973, pp. 281-286.

[9] Meybodi, M. R., "Learning Automata and Its Application to Priority Assignment in a Queueing System With Unknown Characteristics", Ph. D. Thesis, School of Electrical Engineering and Computer Science, University of Oklahoma, Norman, Oklahoma.

[10] Narendra, K. S., and Thathachar, M. A. L., Forthcoming book on learning automata.

[11] Narendra, K. S., and Thathachar, M. A. L., "Learning Automata -- A Survey", IEEE Trans. Syst. Man. and Cybern., Vol. SMC-4, 1974, pp.323-334.

[12] Narendra, K. S., and Thathachar, M. A. L., "On the Behaviour of a Learning Automaton in a Changing Environment With Routing Applications", IEEE Trans. on Syst. Man. and Cybern., Vol. SMC-10, 1980, pp.262-269.

[13] Narendra, K. S., Wright, E. and Mason, L. G., "Application of Learning Automata to Telephone Traffic Routing", IEEE Trans on Syst. Man. and Cybern., Vol. SMC-7, 1977, pp.785-792.

[14] Oommen, B. J., "Absorbing and Ergodic Discretized Two Action Learning Automata", IEEE Trans. on Syst. Man. and Cybern., March/ April 1986,

pp.282-293.

[15] Oommen, B. J., and Thathachar, M. A. L., " Multi-Action Learning Automata Processing Ergodicity of the Mean", Information Sciences, June 1985, pp.183-198.

[16] Oommen, B. J., and Ma, D. Y. C., "Deterministic Learning Automata Solutions to the Object Partitioning Problem", Submitted for Publication (Also available as the SCS-TR-84 from Carleton University, Ottawa, Canada).

[17] Poznyak, A. S., "Use of Learning Automata for the Control of Random Search", Automation and Remote Control, Vol.33, No.12, 1972, pp.1992-2000.

[18] Thathachar, M. A. L., and Oommen, B. J., "Learning Automata Possessing Ergodicity of the Mean : The Two Action Case", IEEE Trans. on Syst. Man. and Cybern., Vol. SMC-13, 1984, pp. 1143-1148.

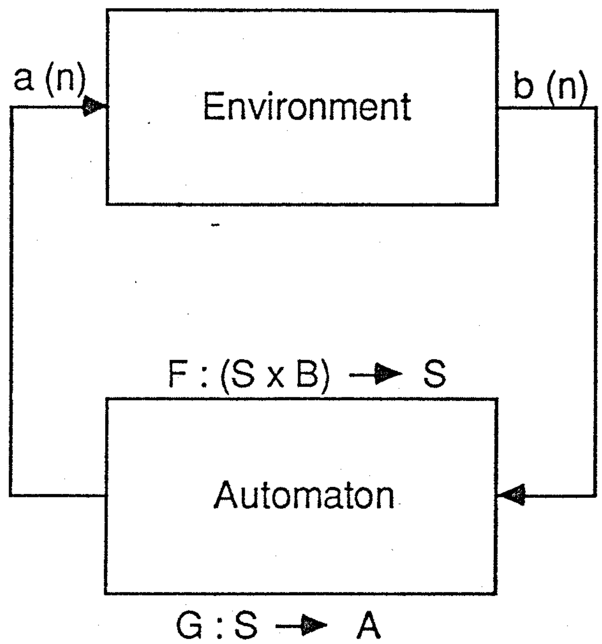
[19] Thathachar, M. A. L., and Sastry, P. S., "Learning Optimal Discriminant Functions Through a Cooperative Game of Automata", Submitted for Publication. (Also available as Report No EE/64/1985 from the Indian Institute of Science, Bangalore, India.)

[20] Tsetlin, M. L., "On the Behaviour of Finite Automata in Random Media", Automat. Telemekh., Vol. 22, 1961, pp. 1345-1354.

[21] Tsetlin, M. L., "Automaton Theory and the Modelling of Biological Systems", New York and London, Academic, 1973.

[22] Tsypkin, Y. Z. and Poznyak, A. S., "Finite Learning Automata", Engineering Cybernatics, Vol. 10, 1972, pp.478-490.

[23] Varshavskii, V. I., and Vorontsova, I. P., "On the Behaviour of Stochastic Automata With Variable structure", Automat. Telemek. (USSR), Vol. 24, 1963, pp. 327-333.



$$b\,(n) \in \{0, 1\} = B$$
$$s\,(n) \in \{s_1, s_2, \ldots, s_N\} = S$$
$$a\,(n) \in \{a_1, a_2, \ldots, a_R\} = A$$

Fig I : The Automaton-Environment interaction

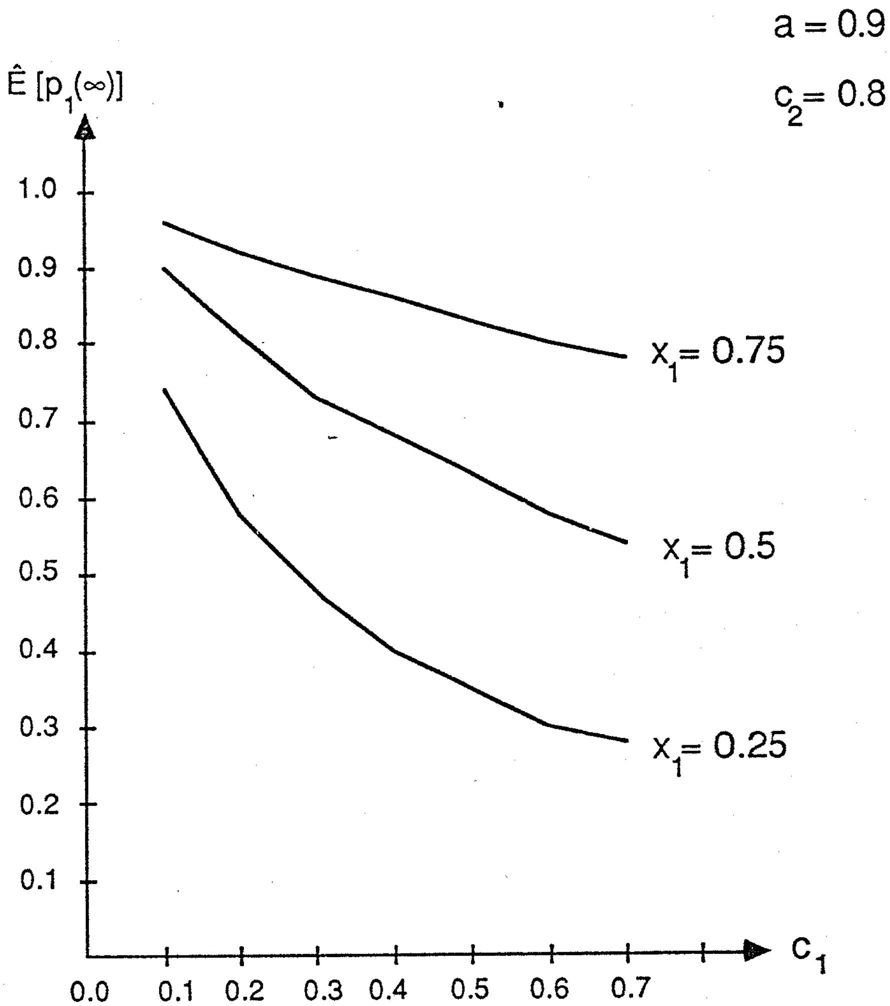


Fig. II : Variation of $\hat{E}[p_1(\infty)]$ with c_1 when $c_2 = 0.8$ for various apriori probabilities x_1.

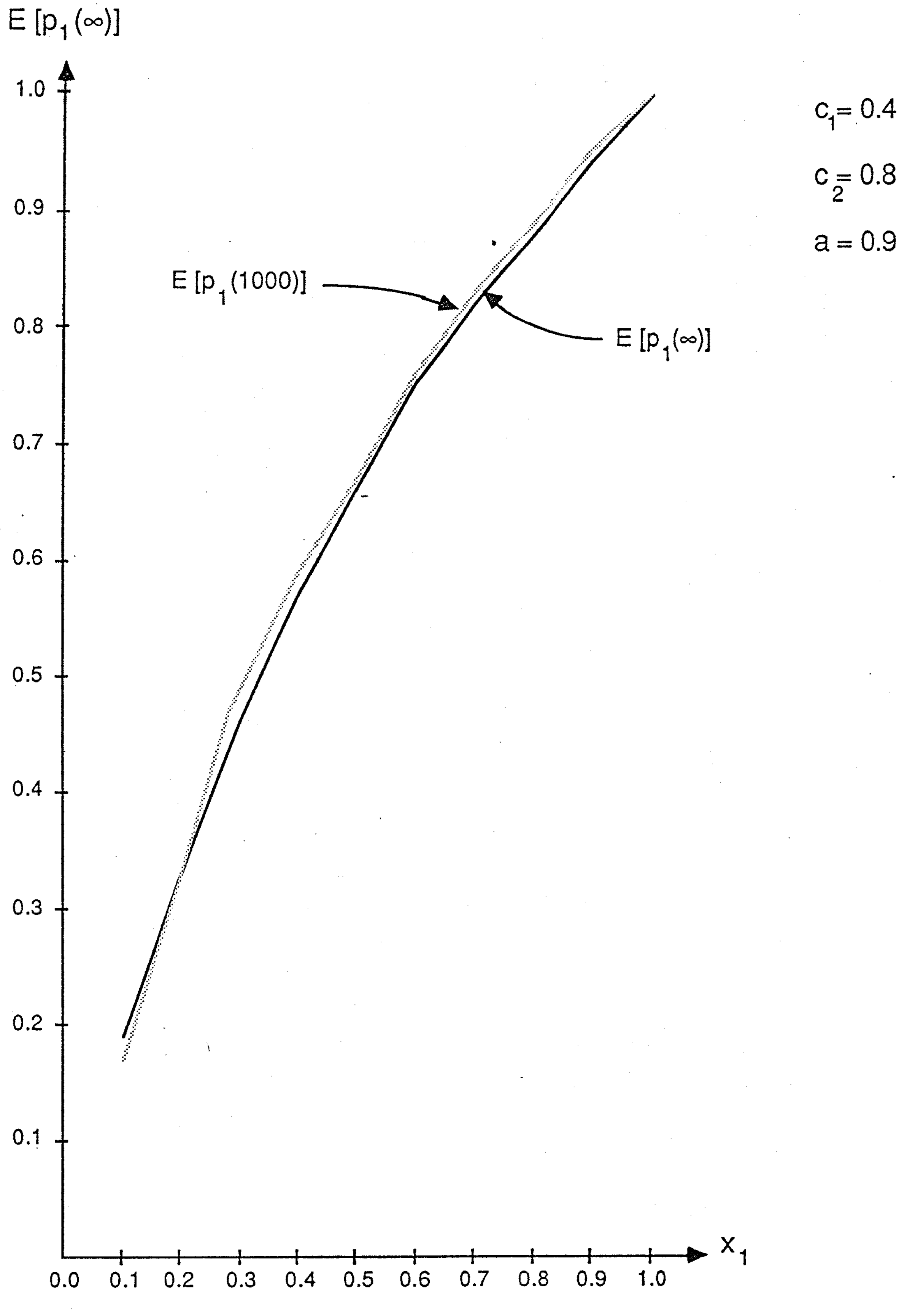


Fig. III : Variation of the exact and estimated values of $E[p_1(\infty)]$ when $c_1 = 0.4$ and $c_2 = 0.8$ with varying values of apriori probabilities x_1.

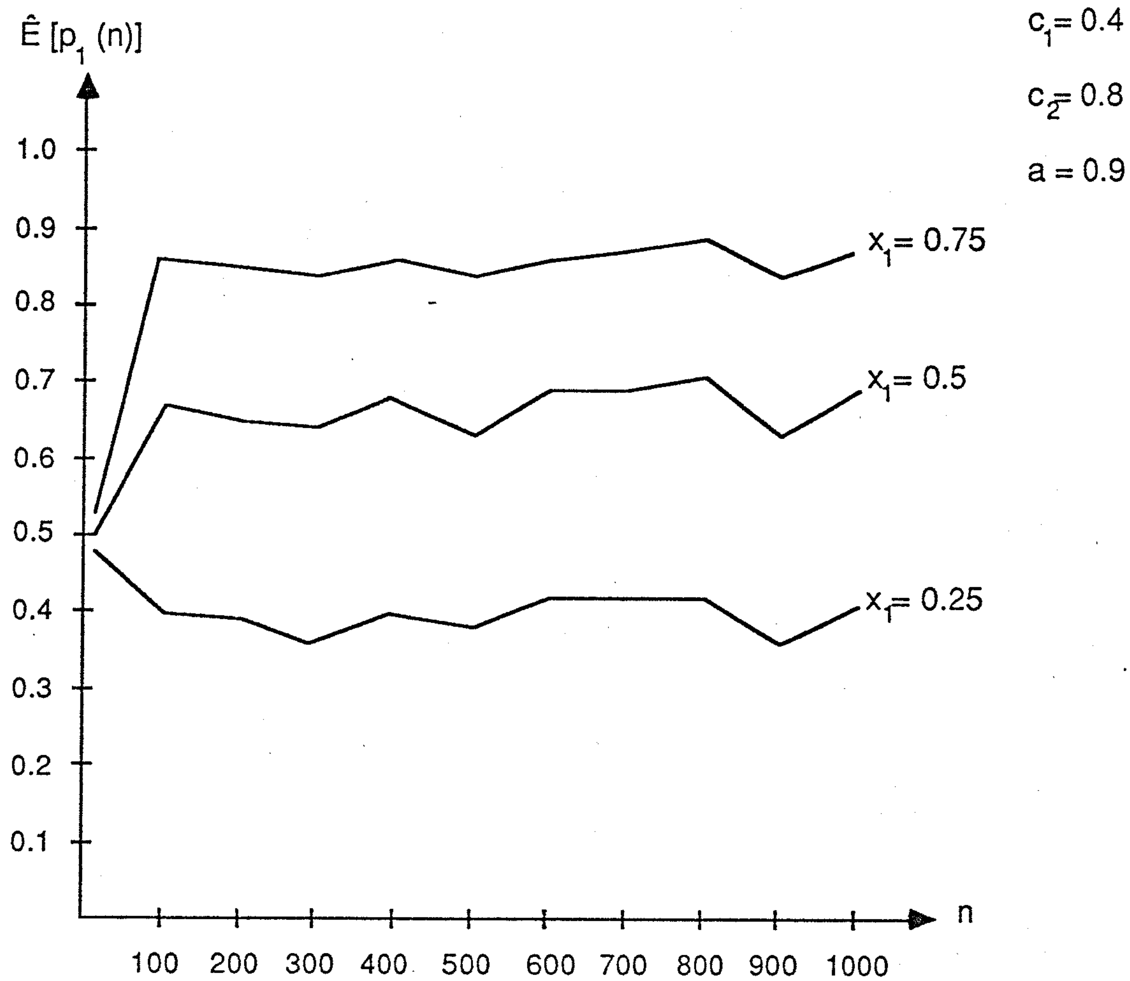


Fig. IV : Rate of increase of $\hat{E}[p_1(n)]$ with time for various apriori probabilities. In this case $c_1 = 0.4$, $c_2 = 0.8$ and $a = 0.9$.

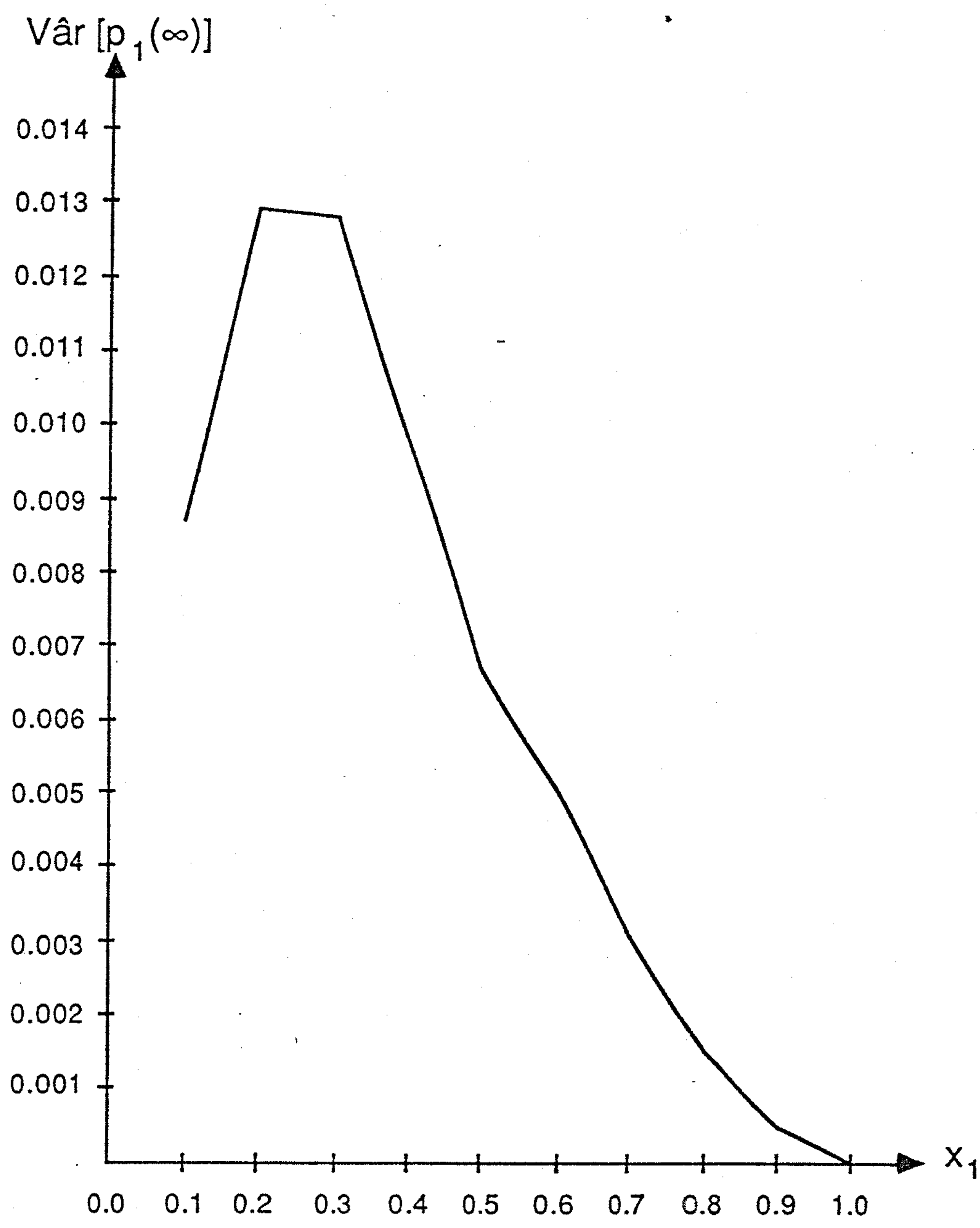


Fig. V : Variation of the Variance of $p_1(\infty)$ with the apriori probability x_1.
In this case $c_1 = 0.4$, $c_2 = 0.8$ and $a = 0.9$.

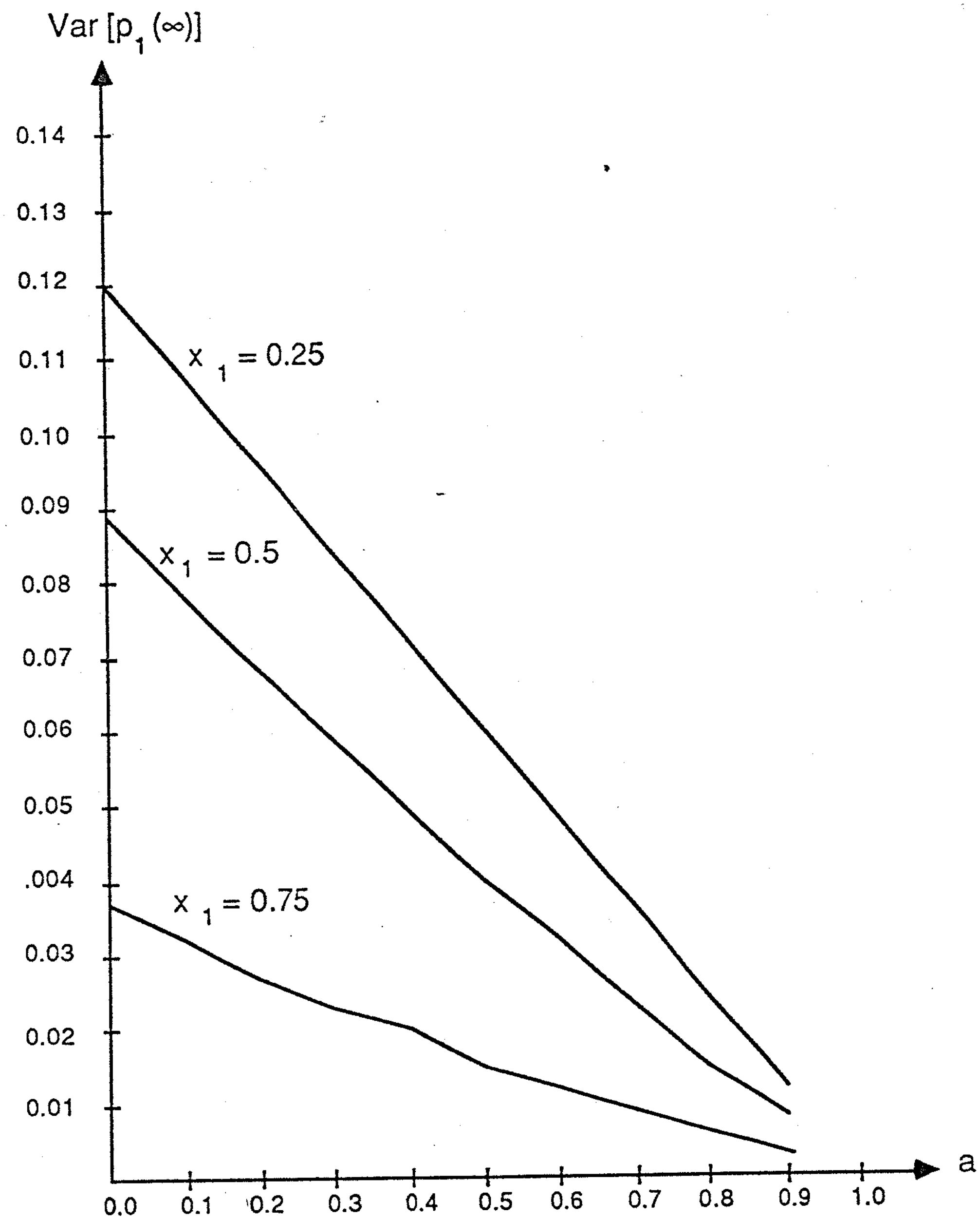


Fig. VI : Variation of the exact variance of p_1 (∞) with a for various values of x_1 . In all cases $c_1 = 0.4$ and $c_2 = 0.8$.

Carleton University, School of Computer Science
Bibliography of Technical Reports
Publications List

School of Computer Science
Carleton University
Ottawa, Ontario, Canada
KIS 5B6

SCS-TR-1	**The Design of CP-6 Pascal** Jim des Rivieres and Wilf R. LaLonde, June 1982
SCS-TR-2	**Single Production Elimination in LR(1) PARSERS: A Synthesis** Wilf R. LaLonde, June 1982
SCS-TR-3	**A Flexible Compiler Structure That Allows Dynamic Phase Ordering** Wilf R. LaLonde and Jim des Rivieres, June 1982
SCS-TR-4	**A Practical Longest Common Subsequence Algorithm for Text Collation** Jim des Rivieres, June 1982
SCS-TR-5	**A School Bus Routing and Scheduling Problem** Wolfgang Lindenberg, Frantisek Fiala, July 1982
SCS-TR-6	**Routing Without Routing Tables: Labelling and Implicit Routing in Networks** Nicola Santoro and Ramez Khatib, July 1982
SCS-TR-7 Out-of-print	**Concurrency Control in Large Computer Networks** Nicola Santoro and Jeffrey B. Sidney, July 1982
SCS-TR-8 Out-of-print	**Order Statistics on Distributed Sets** Nicola Santoro and Jeffrey B. Sidney, July 1982
SCS-TR-9	**Oligarchical Control of Distributed Processing Systems** Moshe Krieger and Nicola Santoro, August 1982
SCS-TR-10	**Communication Bounds for Selection in Distributed Sets** Nicola Santoro and Jeffrey B. Sidney, September 1982
SCS-TR-11	**A Simple Technique for Converting from a Pascal Shop to a C Shop** Wilf R. LaLonde and John R. Pugh, November 1982
SCS-TR-12	**Efficient Abstract Implementations for Relational Data Structures** Nicola Santoro, December 1982
SCS-TR-13	**On The Message Complexity of Distributed Problems** Nicola Santoro, December 1982
SCS-TR-14	**A Common Basis for Similarity Measures Involving Two Strings** R.L. Kashyap and B.J. Oommen, January 1983

Carleton University, School of Computer Science
Bibliography of Technical Reports

SCS-TR-15 **Similarity Measures for Sets of String**
R.L. Kashyap and B.J. Oommen, January 1983

SCS-TR-16 **The Noisy Substring Matching Problem**
R.L. Kashyap and B.J. Oommen, January 1983

SCS-TR-17 **Distributed Election in a Circle Without a Global Sense of Orientation**
E. Korach, D. Rotem, and N. Santoro, January 1983

SCS-TR-18 **A Geometrical Approach to Polygonal Dissimilarity and the Classification of Closed Boundaries**
R.L. Kashyap and B.J. Oommen, January 1983

SCS-TR-19 **Scale Preserving Smoothing of Polygons**
R.L. Kashyap and B.J. Oommen, January 1983

SCS-TR-20 **Not-Quite-Linear Random Access Memories**
Jim des Rivieres, Wilf R. LaLonde, and Mike Dixon, March 1983

SCS-TR-21 **Shout Echo Selection in Distributed Files**
D. Rotem, N. Santoro, J.B. Sydney, March I983.

SCS-TR-22 **Distributed Ranking**
E. Korach, D. Rotem, N. Santoro, March I983.

SCS-TR-23 **A Reduction Technique for Selection in Distributed Files : I**
N. Santoro, J.B. Sidney, April I983.
Replaced by SCS-TR-69

SCS-TR-24 **Learning Automata Possessing Ergodicity of the Mean : The Two Action Case**
M.A.L. Thathachar and B.J. Oommen, May I983.

SCS-TR-25 **Actors - The Stage is Set**
John R. Pugh, June I983.

SCS-TR-26 **On the Essential Equivalence of Two Families of Learning Automata**
M.A.L. Thathachar and B.J. Oommen, May I983.

SCS-TR-27 **Generalized Krylov Automata and their Applicability to Learning in Nonstationary Environments**
B.J. Oommen, June I983.

SCS-TR-28 out-of-print **Actor Systems: Selected Features**
Wilf R. LaLonde, July I983.

SCS-TR-29 **Another Addendum to Kronecker's Theory of Pencils**
M.D. Atkinson, August I983.

SCS-TR-30 **Some Techniques for Group Character Reduction**
M.D. Atkinson and R.A. Hassan, August I983.

SCS-TR-3I **An Optimal Algorithm for Geometrical Congruence**
M.D. Atkinson, August I983.

Carleton University, School of Computer Science
Bibliography of Technical Reports

SCS-TR-32 — **Multi-Action Learning Automata Possessing Ergodicity of the Mean**
B.J. Oommen and M.A.L. Thathachar, August 1983.

SCS-TR-33 Out-of-print — **Fibonacci Graphs, Cyclic Permutations and Extremal Points**
N. Santoro and J. Urrutia, December 1983

SCS-TR-34 — **Distributed Sorting**
D. Rotem, N. Santoro, and J.B. Sydney, December 1983.

SCS-TR-35 — **A Reduction Technique for Selection in Distributed Files: II**
N. Sanatoro, M. Scheutzow, and J.B. Sidney, December 1983.

SCS-TR-36 — **The Asymptotic Optimality of Discretized Linear Reward-Inaction Learning Automata**
B.J. Oommen and Eldon Hansen, January 1984.

SCS-TR-37 — **Geometric Containment is Not Reducible to Pareto Dominance**
N. Santoro, J.B. Sidney, S.J. Sidney, and J. Urrutia, January 1984.

SCS-TR-38 — **An Improved Algorithm for Boolean Matrix Multiplication**
N. Santoro and J. Urrutia, January 1984.

SCS-TR-39 — **Containment of Elementary Geometric Objects**
J. Sack, N. Santoro and J. Urrutia, February 1984

SCS-TR-40 — **SADE: A Programming Environment for Designing and Testing Systolic Algorithms**
J.P. Corriveau and N. Santoro, February 1984.

SCS-TR-41 — **Intersection Graphs, {B_1}-Orientable Graphs and Proper Circular Arc Graphs**
Jorge Urrutia, February 1984.

SCS-TR-42 — **Minimum Decompositions of Polygonal Objects**
J. Mark Keil and Jörg-R. Sack, March 1984.

SCS-TR-43 — **An Algorithm for Merging Heaps**
Jörg-R. Sack and Thomas Strothotte, March 1984.

SCS-TR-44 — **A Digital Hashing Scheme for Dynamic Multiattribute Files**
E.J. Otoo, March 1984

SCS-TR-45 — **Symmetric Index Maintenance Using Multidimensional Linear Hashing**
E.J. Otoo, March 1984

SCS-TR-46 — **A Mapping Function for the Directory of a Multidimensional Extendible Hashing**
E.J. Otoo, March 1984.

SCS-TR-47 — **Translating Polygons in the Plane**
Jörg-R. Sack, March 1984.

SCS-TR-48 — **Constrained String Editing**
J. Oommen, May 1984.

Carleton University, School of Computer Science
Bibliography of Technical Reports

SCS-TR-84 **Deterministic Learning Automata Solutions to the Object Partitioning Problem**
B. John Oommen, D.C.Y. Ma, November 1985

SCS-TR-85 **Selecting Subsets of the Correct Density**
M.D. Atkinson, December 1985

SCS-TR-86 **Robot Navigation In Unknown Terrains Using Learned Visiblity Graphs. Part I: The Disjoint Convex Obstacles Case**
B. J. Oommen, S.S. Iyengar, S.V.N. Rao, R.L. Kashyap, February 1986

SCS-TR-87 **Breaking Symmetry In Synchronous Networks**
Greg N. Frederickson, Nicola Santoro, April 1986

SCS-TR-88 **Data Structures and Data Types: An Object-Oriented Approach**
John R. Pugh, Wilf R. LaLonde and David A. Thomas, April 1986

SCS-TR-89 **Ergodic Learning Automata Capable of Incorporating Apriori Information**
B. J. Oommen, May 1986

SCS-TR-90 **Iterative Decomposition of Digital Systems and Its Applications**
Vaclav Dvorak, May 1986.

SCS-TR-91 **Actors In a Smalltalk Multiprocessor: A Case for Limited Parallelism**
Wilf R. LaLonde, Dave A. Thomas and John R. Pugh, May 1986

SCS-TR-92 **ACTRA - A Multitasking/Multiprocessing Smalltalk**
David A. Thomas, Wilf R. LaLonde, and John R. Pugh, May 1986

SCS-TR-93 **Why Exemplars are Better Than Classes**
Wilf R. LaLonde, May 1986

SCS-TR-94 **An Exemplar Based Smalltalk**
Wilf R. LaLonde, Dave A. Thomas and John R. Pugh, May 1986

SCS-TR-95 **Recognition of Noisy Subsequences Using Constrained Edit Distances**
B. John Oommen, June 1986

SCS-TR-96 **Guessing Games and Distributed Computations In Synchronous Networks**
J. van Leeuwen, N. Santoro, J. Urrutia and S. Zaks, June 1986.

SCS-TR-97 **Bit vs. Time Tradeoffs for Distributed Elections In Synchronous Rings**
M. Overmars and N. Santoro, June 1986.

SCS-TR-98 **Reduction Techniques for Distributed Selection**
N. Santoro and E. Suen, June 1986.

SCS-TR-99 **A Note on Lower Bounds for Min-Max Heaps**
A. Hasham and J.-R. Sack, June 1986.

Carleton University, School of Computer Science
Bibliography of Technical Reports

SCS-TR-67 **Heaps in Heaps**
T. Strothotte, J.-R. Sack, November 1984. Revised April 1985.

SCS-TR-68 out-of-print **Partial Orders and Comparison Problems**
M.D. Atkinson, November 1984. See Congressus Numerantium 47 ('86), 77-88

SCS-TR-69 **On the Expected Communication Complexity of Distributed Selection**
N. Santoro, J.B. Sidney, S.J. Sidney, February 1985.

SCS-TR-70 **Features of Fifth Generation Languages: A Panoramic View**
Wilf R. LaLonde, John R. Pugh, March 1985.

SCS-TR-71 **Actra: The Design of an Industrial Fifth Generation Smalltalk System**
David A. Thomas, Wilf R. LaLonde, April 1985.

SCS-TR-72 **Minmaxheaps, Orderstatisticstrees and their Application to the Coursemarks Problem**
M.D. Atkinson, J.-R. Sack, N. Santoro, T. Strothotte, March 1985.

SCS-TR-73 **Designing Communities of Data Types**
Wilf R. LaLonde, May 1985.

SCS-TR-74 **Absorbing and Ergodic Discretized Two Action Learning Automata**
B. John Oommen, May 1985.

SCS-TR-75 **Optimal Parallel Merging Without Memory Conflicts**
Selim Akl and Nicola Santoro, May 1985

SCS-TR-76 **List Organizing Strategies Using Stochastic Move-to-Front and Stochastic Move-to-Rear Operations**
B. John Oommen, May 1985.

SCS-TR-77 **Linearizing the Directory Growth in Order Preserving Extendible Hashing**
E.J. Otoo, July 1985.

SCS-TR-78 **Improving Semijoin Evaluation in Distributed Query Processing**
E.J. Otoo, N. Santoro, D. Rotem, July 1985.

SCS-TR-79 **On the Problem of Translating an Elliptic Object Through a Workspace of Elliptic Obstacles**
B.J. Oommen, I. Reichstein, July 1985.

SCS-TR-80 **Smalltalk - Discovering the System**
W. LaLonde, J. Pugh, D. Thomas, October 1985.

SCS-TR-81 **A Learning Automation Solution to the Stochastic Minimum Spanning Circle Problem**
B.J. Oommen, October 1985.

SCS-TR-82 **Separability of Sets of Polygons**
Frank Dehne, Jörg-R. Sack, October 1985.

SCS-TR-83 **Extensions of Partial Orders of Bounded Width**
M.D. Atkinson and H.W. Chang, November 1985.

SCS-TR-49 Out-of-print	**O(N) Election Algorithms in Complete Networks with Global Sense of Orientation** Jörg-R. Sack, Nicola Santoro, Jorge Urrutia, May 1984
SCS-TR-50 ________	**The Design of a Program Editor Based on Constraints** Christopher A. Carter and Wilf R. LaLonde, May 1984.
SCS-TR-51 ________	**Discretized Linear Inaction-Penalty Learning Automata** B.J. Oommen and Eldon Hansen, May 1984.
SCS-TR-52 ________	**Sense of Direction, Topological Awareness and Communication Complexity** Nicola Santoro, May 1984.
SCS-TR-53 ________	**Optimal List Organizing Strategy Which Uses Stochastic Move-to-Front Operations** B.J. Oommen, June 1984.
SCS-TR-54 ________	**Rectilinear Computational Geometry** J. Sack, June 1984.
SCS-TR-55 ________	**An Efficient, Implicit Double-Ended Priority Queue** M.D. Atkinson, Jörg-R. Sack, Nicola Santoro, T. Strothotte, July 1984.
SCS-TR-56 ________	**Dynamic Multipaging: A Multidimensional Structure for Fast Associative Searching** E. Otoo, T.H. Merrett, August 1984.
SCS-TR-57 ________	**Specialization, Generalization and Inheritance** Wilf R. LaLonde, John R. Pugh, August 1984.
SCS-TR-58 ________	**Computer Access Methods for Extendible Arrays of Varying Dimensions** E. Otoo, August 1984.
SCS-TR-59 ________	**Area-Efficient Embeddings of Trees** J.P. Corriveau, Nicola Santoro, August 1984.
SCS-TR-60 ________	**Uniquely Colourable m-Dichromatic Oriented Graphs** V. Neumann-Lara, N. Santoro, J. Urrutia, August 1984.
SCS-TR-61 ________	**Analysis of Distributed Algorithms for Extrema Finding in a Ring** D. Rotem, E, Korach and N. Santoro, August 1984.
SCS-TR-62 out-of-print	**On Zigzag Permutations and Comparisons of Adjacent Elements** M.D. Atkinson, October 1984. See Information Processing Letters 21 ('85) 187-189
SCS-TR-63 ________	**Sets of Integers with Distinct Differences** M.D. Atkinson, A. Hassenklover, October 1984.
SCS-TR-64 ________	**Teaching Fifth Generation Computing: The Importance of Small Talk** Wilf R. LaLonde, Dave A. Thomas, John R. Pugh, October 1984.
SCS-TR-65 ________	**An Extremely Fast Minimum Spanning Circle Algorithm** B.J. Oommen, October 1984.
SCS-TR-66 ________	**On the Futility of Arbitrarily Increasing Memory Capabilities of Stochastic Learning Automata** B.J. Oommen, October 1984. Revised May 1985.