AUTOMATA IN ENVIRONMENTS

by

Jerren Gould

Department of Statistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series #947
June 1974

JERREN GOULD. Automata in Environments (Under the direction of EDWARD J. WEGMAN).

Automata theory is the study of mathematical models of finite-state systems which admit certain inputs and emit certain outputs. The state transition behavior of previously considered models, such as Turing machines, finite automata, and probabilistic automata, has been assumed to be related only to the present input. In this dissertation we take the viewpoint that every system is situated within an environment which affects the initial state distribution and the state transition function. The Rabin-Paz model of a probabilistic automaton is generalized so that the initial state distribution is related to the initial configuration of the environment and that the state transition function is not only related to the present input but also to the present configuration of the environment.

Chapter 2 considers the case in which there is a deterministic rule to specify the configuration of the environment at any time. We show that such a model is properly more general and more powerful than the probabilistic automaton model. Sufficient conditions are given for an automaton with a deterministic environment rule to be "simulated" by a probabilistic automaton. We also consider the behavior of automata with environment sets that have certain metric or semi-group structure.

In Chapter 3 we consider the effects of small perturbations of the environments on the system. Conditions are given for automata in environments to have stable state distribution functions. We also give conditions for the set of tapes defined by an automaton to remain unchanged for sufficiently small perturbations of the environments.

Finally, we consider the case of automata operating in random environments; we assume that the realization of any environment is governed by some probabilistic structure. The behavior of systems with various environment processes is investigated. We encounter the dual nature of the randomness involved in automata in random environments. We have that the environments are random and that for each value the environment assumes, the probabilistic state transitions are defined. However, we find that there is a mean equivalent canonical representation for automata in random environments which eliminates the randomness due to probabilistic state transitions; for any automaton in random environments we can find another automaton in random environments with finite environment set, deterministic assignment of the initial state, and deterministic transitions which has a state distribution equivalent in the mean.

ACKNOWLEDGMENTS

I wish to express my sincere appreciation to Professor Edward J. Wegman for proposing this problem and for the illuminating suggestions he made during the course of the investigation. I am greatly indebted to him for his guidance and encouragement throughout my graduate career. It has been a very valuable experience to work under his direction.

I would also like to thank Professor W.L. Smith, the chairman of my dissertation committee, for his helpful suggestions and for his encouragement during my graduate study. Thanks are also due to Professors N.L. Johnson, G.D. Simons, S. Cambanis, G.A. Magó, and W.R. Wogen for participating as members of my dissertation committee.

The financial support provided by the Department of Statistics during my period of study in Chapel Hill is greatly appreciated.

The contributions of June Maxwell, who ably typed the manuscript, are earnestly appreciated.

Finally, I wish to express my heartfelt gratitude to my wife, Marsha, for being My Small Joy.

TABLE OF CONTENTS

CHAPTER I

Development of Automata Theory

1.1 The Automaton Concept.

Automata theory is the study of the characteristics of automatic devices. For some time there has been research into the nature of algorithms and the functions computable by following an algorithm. Researchers in automata theory have made an effort to apply and extend these results to more realistic models of physical systems. The automata models have been motivated particularly by the behavior of digital computers, but also by organic nerve networks, information transmission systems, synthesis of man's languages, learning behavior, games between automata, etc. These models are constructed to investigate their mathematical properties as well as to discover the characteristics and capabilities of similar phenomena in the real world.

In considering the theoretical limitation of the type of computation that could be carried out by an automatic calculating device, A.M. Turing [18] constructed an abstract machine capable of following a finite, consistent list of instructions. The Turing machine is a finite-state machine with an associated external memory. The memory has the form of a sequence of squares marked off on a linear tape. The tape is assumed to be infinite in both directions. A Turing machine has a finite alphabet of symbols and a finite internal state set. The machine is capable of five operations. It can read a square of the tape, erase a square of the tape and write a new symbol on that square, change internal state, move to an adjacent square, and halt. The instructions to the machine are contained in a finite list of quintuples of the form: (present state, symbol read, new state, symbol written, direction of motion). The third, fourth, and fifth components of the quintuple are determined uniquely as functions of the first and second components. Thus, for any symbol read and internal machine state, the next act of the machine is unambiguously specified. Although the tape is regarded as infinite in both directions, when the machine is started, the tape must be blank except for a finite number of squares. The Turing machine is started by placing the machine on a square of the tape and designating an initial state. The output is considered to be the expression on the tape when the machine halts.

Indeed, any computation that can be performed by a deterministic automatic calculating device can also be performed by a Turing machine. This fact led to "Church's thesis," the assumption that every algorithm has a Turing machine which can execute it. Thus, consideration of Turing machines gave rise to results concerning the theory of computable functions and other equivalent logic systems based on algorithms. The Turing machine model in many ways patterns the operation of a digital computer. The finite state set and finite set of quintuples form a "program." Both systems are capable of accepting any finite amount of input. The major drawback to the Turing machine model for the digital computer is its unbounded memory capacity via its infinite tape. This enables the Turing machine model to consider the entire class of computable functions.

In order to simulate more closely the characteristics of digital computers, the Turing machine model was amended by restricting the external memory tape to be of bounded length. This model, called a linear bounded automaton, operates in the same manner as a Turing machine except that the tape contains only a finite number of squares equal to the length of the input. The machine is not allowed to move beyond the bounds of the tape. The new model is not as general as the Turing machine. Linear bounded automata, however, are capable of computing recursive functions that can be handled by a finite, bounded memory, as in the case of digital computers.

1.2 Finite Automata

So far the discussion has centered on the transducer model of machines. That is, a machine acts as a transducer if its main purpose is to generate an output sequence that is a function of the input and the initial state of the machine. However, for the purposes of this paper we shall henceforth consider machines that act as acceptors. Instead of producing an output sequence, acceptors simply answer "yes" or "no" to questions asked of them. The transducer model and the acceptor model are equivalent in their computing capabilities.

We now formulate an acceptor model with a bounded memory capacity. This class of machines is called finite automata. For finite automata the questions take the form of finite sequences from a finite alphabet; such a sequence of symbols is called a tape. The machine itself shall be considered as a black box, capable of reading one symbol at a time from the tape and then advancing to the next symbol on the tape. In order to maintain the generality of the structure of the model, the concept of a finite internal state set is employed. As with Turing machines and linear bounded automata, this machine is capable of changing states based on the present state and latest input symbol. With finite automata the memory is characterized by the finite internal state set and, hence, is finite and bounded. To obtain the output, or answer, from the machine its state set is partitioned into two classes, the states for which the output is "yes" and the states for which the output is "no." The output is observed when the machine has read all of the symbols on the tape.

Let us now define these notions of finite automata in a more precise mathematical form. Let Σ be the finite input set, the alphabet, and let Σ^*, the set of tapes, be the class of all finite sequences of symbols from Σ. Let us also include Λ, the empty tape, in Σ^*. Note that Σ^* together with the operation of concatenation forms a free semigroup generated by Σ. We are now prepared to formalize the definition of finite automata.

<u>Definition 1.2.1</u>. A finite automaton (FA) over the alphabet Σ is a system $A = (S,M,s_0,F)$, where S is a finite nonempty set (the internal states of A), M is a mapping $M:S \times \Sigma \rightarrow S$ (the transition function of A), $s_0 \in S$ (the initial state of A), and $F \subset S$ (the set of final, or "yes," states of A).

For a finite automaton A the domain of definition of the transition function can be extended from $S \times \Sigma$ to $S \times \Sigma^*$ by the recursion as follows:

$$M(s,\Lambda) = s \qquad \forall s \in S$$

$$M(s,x\sigma) = M(M(s,x),\sigma) \quad \forall s \in S,\ \forall x \in \Sigma^*,\ \forall \sigma \in \Sigma .$$

The interpretation of $M(s,x)$ is elementary: it is that state of the machine obtained by beginning in state s and reading through tape x symbol by symbol, changing states according to the transition function.

Definition 1.2.2. The collection of all tapes $x \in \Sigma^*$ such that $M(s_0,x) \in F$ is called $T(A)$, the set of tapes accepted, or defined, by the automaton $A = (S,M,s_0,F)$.

Definition 1.2.3. The class of all definable sets of tapes, $\mathcal{T}$, is the collection of all sets of the form $T(A)$, for some finite automaton A. If $U \in \mathcal{T}$, then U is called a regular event.

In order to characterize the definable sets of tapes, we must make the following definitions.

Definition 1.2.4. An equivalence relation E over the set Σ^* of tapes is right invariant if whenever xEy, then $xzEyz$, $\forall z \in \Sigma^*$.

Definition 1.2.5. An equivalence relation E over the set of tapes is of finite index if there are only finitely many equivalence classes under the relation.

Hence, we now state a fundamental theorem due to A. Nerode.

Theorem 1.2.1. Let $U \subset \Sigma^*$. The following are equivalent:

1. $U \in \mathcal{T}$.
2. U is the union of some equivalence classes of a right invariant equivalence relation of finite index.
3. The explicit right invariant equivalence relation E defined by

the condition that $\forall x,y \in \Sigma^*$, xEy iff $\forall z \in \Sigma^*$ whenever $xz \in U$, then $yz \in U$, and conversely, is an equivalence relation of finite index.

If $U \in T$, then the number of equivalence classes under the relation E is the least number of internal states of any finite automaton defining U. Theorem 1.2.1 also leads to some simple facts about the class T.

<u>Theorem 1.2.2</u>. The class T is a Boolean algebra.

<u>Theorem 1.2.3</u>. The class T contains all finite sets of tapes.

We are also interested in characterizing the set $T(A)$ when we are not told how the black box A operates. All that is allowed is to supply input, or "question," the black box and observe the "answers." Rabin and Scott [17] exhibit an effective procedure[1] based only on a knowledge of a bound on the number of states of A whereby in a finite number of steps it can be decided whether $T(A)$ is empty, finite, or infinite. It is also possible to determine an upper bound on the number of tapes in $T(A)$ when $T(A)$ is finite. Rabin and Scott further discuss this type of result in comparing finite automata.

<u>Definition 1.2.6</u>. Two finite automata A and B are equivalent iff $T(A) = T(B)$.

<u>Definition 1.2.7</u>. The length of any tape $x \in \Sigma^*$ is the number of symbols in the sequence x. We denote the length of $x \in \Sigma^*$ by $L(x)$.

---

[1] An effective procedure is a set of rules which dictate, from moment to moment, precisely how to behave. Effective procedure is a synonym for algorithm.

<u>Theorem 1.2.4</u>. Let A_1 and A_2 be FA with n_1 and n_2 internal states, respectively. A_1 and A_2 are not equivalent iff there exists $x \in \Sigma^*$ such that $L(x) < n_1 n_2$ and x is accepted by one machine and not the other.

Rabin and Scott also introduced the model of a nondeterministic finite automaton. In contrast to a finite automaton which has only one way to change state in any particular situation, the nondeterministic machine allows a choice from a set of states.

<u>Definition 1.2.8</u>. A nondeterministic finite automaton over the alphabet Σ is a system $A = (S,M,S_0,F)$, where S is a finite non-empty set, M is a function from $S \times \Sigma$ taking values in the set of all subsets of S, $S_0 \subset S$, and $F \subset S$.

<u>Definition 1.2.9</u>. Let A be a nondeterministic finite automaton. The set of tapes $T(A)$ accepted, or defined, by A is the collection of all tapes $x = \sigma_0 \sigma_1 \ldots \sigma_{k-1}$ for which there exists a sequence $s_0, s_1, \ldots, s_k$ of internal states of A such that

1. $s_0 \in S_0$, $s_i \in M(s_{i-1}, \sigma_{i-1})$ for $i = 1,2,\ldots,k$,
2. $s_k \in F$.

This appears to be a generalization of finite automata; however, Rabin and Scott demonstrated an effective procedure to construct a finite automaton defining exactly the same set of tapes as any given nondeterministic finite automaton.

<u>Definition 1.2.10</u>. Let $A = (S,M,S_0,F)$ be a nondeterministic finite automaton. $\mathcal{D}(A)$ is the system $\mathcal{D}(A) = (P,Q,p_0,R)$, where P

is the set of all subsets of S, Q is a function $Q: P\times\Sigma \to P$ such that $Q(p,\sigma) = \bigcup_{s\in p} M(s,\sigma)$, $p_0 = S_0$, and R is the set of all subsets of S containing at least one element of F. Clearly, $\mathcal{D}(A)$ is a finite automaton and is equivalent to A.

Theorem 1.2.5. If A is a nondeterministic finite automaton, then $T(A) = T(\mathcal{D}(A))\in\mathcal{T}$.

1.3 Probabilistic Automata

So far we have assumed that our black box machine operates in an error-free deterministic manner or in a manner equivalent to such a machine. In physical machines, however, malfunctions may occur as a result of mechanical or electrical breakdown. Rabin [16] proposed the model of probabilistic automata to simulate the behavior of machines and systems with certain types of random errors and to investigate the more general class of machines. Here we combine the mathematical definitions of Rabin and Paz [13] to take on the most general form.

Definition 1.3.1. Let M_n denote the set of all n×n stochastic matrices and V_n denote the set of all n-dimensional stochastic vectors.

Definition 1.3.2. A probabilistic automaton (PA) over the alphabet Σ is a system $A = (S,M,\pi_0,F)$, where $S = \{s_1,\ldots,s_n\}$ is a finite set (the set of internal states of A), M is a function $M: \Sigma \to M_n$ (the matrix transition function) such that $m_{ij}(\sigma)$ is the probability of changing to state s_j under input σ given that the machine is in state s_i, $\pi_0\in V_n$ (the initial state distribution) and $F\subset S$ (the set of final states).

Probabilistic automata are models for systems capable of a finite number of internal states and able to receive inputs $\sigma\in\Sigma$. When in state s_i and if the input is σ, then the system may go into any one of the states $s_j \in S$, and the probability of going into state s_j is $m_{ij}(\sigma)$. These transition probabilities are assumed to remain fixed and be independent of time and other inputs. Thus, the chance of an "error" on any transition depends only on the present state of the machine and the input symbol read. We now extend the definition of M to the set of all tapes of finite length, Σ^*, by the rule: if $x = \sigma_1 \sigma_2 \ldots \sigma_k$, then $M(x) = M(\sigma_1)M(\sigma_2)\ldots M(\sigma_k)$.

Definition 1.3.3. Let $D(\mathcal{M}_n)$ denote the subset of $\mathcal{M}_n$ such that if $M \in D(\mathcal{M}_n)$, then $m_{ij} = 0$ or 1. Also let $D(\mathcal{V}_n)$ be the subset of $\mathcal{V}_n$ such that if $v \in D(\mathcal{V}_n)$, then $v_i = 0$ or 1.

By restricting the transition function M to take values in $D(\mathcal{M}_n)$ and the initial distribution function π_0 to take values in $D(\mathcal{V}_n)$ we can now see that the class of probabilistic automata includes the class of finite automata.

Definition 1.3.4. Let $A = (S, M, \pi_0, F)$ be a PA. We define the state distribution of A after input tape $x \in \Sigma^*$ as $\pi(x) = \pi_0 M(x)$. Let η^F be an n-dimensional column vector such that $\eta_i^F = \begin{cases} 1 & s_i \in F \\ 0 & s_i \notin F \end{cases}$.

We define the probability that tape x is accepted by A as $p(x) = \pi(x)\eta^F = \pi_0 M(x)\eta^F$.

As with finite automata, probabilistic automata may be used to define sets of tapes. Because of the more general nature of probabilistic automata, these sets of tapes will depend not only on A, but also on an additional parameter, λ, called the cut-point.

<u>Definition 1.3.5</u>. Let A be a PA and λ be a real number, $\lambda \in [0,1)$. The set of tapes $T(A,\lambda)$ is defined by $T(A,\lambda) = \{x | x \in \Sigma^*, \lambda < p(x)\}$. If $x \in T(A,\lambda)$, we say that x is accepted with cut-point λ. $T(A,\lambda)$ is called the set defined by A with cut-point λ.

<u>Definition 1.3.6</u>. Let $U \subset \Sigma^*$. U is a probabilistic cut-point event (PCE) if $U = T(A,\lambda)$ for some PA A and some $\lambda \in [0,1)$.

For the case of any finite automaton A, $p(x) = 1$ if and only if $x \in T(A)$. If $x \notin T(A)$, then $p(x) = 0$. Hence, $T(A) = T(A,\lambda)$ for any $\lambda \in [0,1)$. Thus, every regular event is a PCE. Rabin constructed a particular probabilistic automaton A to demonstrate that the converse is not true. For this A, $\{T(A,\lambda) | \lambda \in [0,1)\}$ forms an uncountable pairwise different collection of sets. There are, however, only a countable number of regular events. Hence, this probabilistic automaton A defines a set which is not defined by any finite automaton. Thus, the class of PCE's is a proper superset of T; that is, the class of probabilistic automata is more general than the class of finite automata. The class of PCE's is uncountable, but Paz showed that not every subset of Σ^* is a PCE.

For any probabilistic automaton A, the state distribution function π leads to a right invariant equivalence relation. We say, for any $x,y \in \Sigma^*$ xEy iff $\pi(x) = \pi(y)$. Clearly, $\pi(xz) = \pi(yz)$ $\forall z \in \Sigma^*$; thus E is right invariant. E, in the case of probabilistic automata, may have up to a countable number of equivalence classes. However, in a manner similar to Theorem 1.2.1, any set of the form $T(A,\lambda)$ can be represented as the union of the equivalence classes of E satisfying the condition that $\pi(x)\eta^F > \lambda$.

Under certain conditions we can guarantee that there exists a finite automaton with the same acceptance set as a probabilistic automaton with a specific cut-point.

Definition 1.3.7. A cut-point λ is called isolated with respect to A if there exists a $\delta > 0$ such that $|p(x) - \lambda| \geq \delta \quad \forall x \in \Sigma^*$.

Theorem 1.3.1. Let A be a PA and λ an isolated cut-point for A. Then there exists an FA B such that $T(A,\lambda) = T(B)$. If A has n states, then B can be chosen to have n' states, where $n' \leq (1 + \frac{1}{2\delta})^{n-1}$.

It seems that in passing from a probabilistic automaton to an equivalent finite automaton we may have to increase the number of states. In many cases a probabilistic automaton equivalent to a finite automaton can be constructed which is more economical in the state saving sense.

Paz found other conditions for the equivalence of probabilistic and finite automata. His analysis in this area also introduced other interesting mathematical results.

Definition 1.3.8. Two PA A and B are strongly equivalent iff $T(A,\lambda) = T(B,\lambda) \quad \forall \lambda \in [0,1)$.

Definition 1.3.9. A probabilistic table (PT) is a system $\theta = (S,M)$, where S and M are as in definition 1.3.2. A discriminative probabilistic table (DPT) is a system $T = (S,M,F)$, where S, M, and F are as in definition 1.3.2.

Definition 1.3.10. Let T be a DPT and π_0 and ρ_0 any initial distributions for T. A tape x distinguishes between π_0 and ρ_0 if

$\pi_0 M(x)\eta^F \neq \rho_0 M(x)\eta^F$. Two initial distributions are k-equivalent if they are not distinguishable by any tape $x \in \Sigma^*$ such that $L(x) \leq k$.

These definitions lead immediately to results of equivalence of probabilistic automata.

Theorem 1.3.3. Let T be a DPT. If two initial distributions are distinguishable, then there exists $x \in \Sigma^*$ such that $L(x) \leq n-2$ which distinguishes between them, where n is the cardinality of the state set S.

Theorem 1.3.4. Given two PA A_1 and A_2 over Σ, if $p_1(x) = p_2(x)$ $\forall x \in \Sigma^*$ such that $L(x) \leq n_1 + n_2 - 2$, then A_1 and A_2 are strongly equivalent.

Paz extended the concept of definiteness for finite automata to construct a body of results for automata equivalence and approximation. The k-suffix of any tape x of length at least k is the tape of the last k symbols of x.

Definition 1.3.11. Let $k \geq 0$ and $U \subset \Sigma^*$. U is a weakly k-definite set if for any $x \in \Sigma^*$ such that $L(x) \geq k$, then $x \in U$ iff the k-suffix of x is in U. U is k-definite if it is weakly k-definite but not weakly (k-1)-definite. A set U is definite if it is k-definite for some k.

Clearly, every definite set is definable by some finite automaton. Such finite automata are called definite automata. Paz sought to generalize this concept to form a class of definite probabilistic automata.

Definition 1.3.12. A PT $\theta = (S,M)$ is weakly k-definite iff, for any $x \in \Sigma^*$ such that $L(x) \geq k$ and for any initial distributions π_0 and ρ_0, the following holds: $(\pi_0 - \rho_0)M(x) = \underset{\sim}{0}$.

This, indeed, does generalize the notion of weak k-definiteness. Given any final set F and initial distribution π_0 for a weakly k-definite probabilistic table, then for all $x \in \Sigma^*$ such that $L(x) \geq k$, $p(x) = p(\text{k-suffix of } x)$.

Definition 1.3.13. A PT is k-definite if it is weakly k-definite but not weakly (k-1) definite. A PT is definite if it is k-definite for some k.

Paz extended the result of Perles, Rabin, and Shamir [15] to include his probabilistic notion of definiteness.

Theorem 1.3.5. Let $\theta = (S,M)$ be a PT. If θ is definite, then it is (n-1)-definite, where n is the cardinality of S.

Note that any probabilistic automaton with a definite probabilistic table defines a definite event and, hence, must have an equivalent definite automaton. Paz introduced an even weaker notion of definiteness, i.e., quasidefiniteness, to define a larger class of automata. Members of this class do not, in general, have equivalent finite automata; however, this class possesses some interesting approximation properties.

Definition 1.3.14. Given any arbitrary n-dimensional vector ξ, we define $|\xi| = \max_i |\xi_i|$. Also, for any arbitrary $n \times n$ matrix Q we define $|Q| = \max_{i,j} |q_{ij}|$.

Definition 1.3.15. A PT $\theta = (S,M)$ is quasidefinite iff, for every $\varepsilon > 0$ there is a number $N(\varepsilon)$ such that for any $x \in \Sigma^*$ such that $L(x) \geq N(\varepsilon)$ and for any possible initial distributions π_0 and ρ_0, then $|(\pi_0 - \rho_0)M(x)| < \varepsilon/n$.

Hence, for any probabilistic automaton A with quasidefinite table let us fix ε; if $x \in \Sigma^*$ is any tape such that $L(x) \geq N(\varepsilon)$, then $|p(x) - p(N(\varepsilon)\text{-suffix of } x)| < \varepsilon$. Thus we have a looser, approximation notion of definiteness.

Definition 1.3.16. Let $M \in \mathcal{M}_n$. M is called scrambling if for every pair of states i_1 and i_2, there is a state j such that $m_{i_1 j} > 0$ and $m_{i_2 j} > 0$. A finite subset $\{M_1,\ldots,M_k\}$ of $\mathcal{M}_n$ satisfies the H_4 condition of order r if there is an r such that any product of the M's of length at least r is scrambling. For a PA, any tape $x \in \Sigma^*$ such that $M(x)$ is scrambling is called a scrambling tape.

Theorem 1.3.6. Let $\theta = (S,M)$ be a PT. θ is quasidefinite iff the set of matrices $\{M(\sigma)|\sigma \in \Sigma\}$ satisfies the H_4 condition.

Theorem 1.3.7. Let $\theta = (S,M)$ be any PT. There is an effective procedure which decides in a finite number of steps whether θ is quasidefinite.

Definition 1.3.17. Let $U \subset \Sigma^*$ and $U^c = \Sigma^* - U$. The set of tapes U ε-approximates the set of tapes defined by a PA with cut-point λ if

$$(T(A,\lambda) - U) \cup (T(A,\lambda)^c - U^c) \subset \{x \,|\, |p(x) - \lambda| < \varepsilon\} \quad .$$

Theorem 1.3.8. Let A be a PA with quasidefinite table. Then for every $\varepsilon > 0$ there exists a definite automaton A_ε such that $T(A_\varepsilon)$ ε-approximates $T(A,\lambda)$ for any given λ.

Theorem 1.3.9. If A is a PA with quasidefinite table and if λ is an isolated cut-point, then $T(A,\lambda)$ is a definite set.

Rabin defined the more specific concept of an actual automaton to derive a converse for Theorem 1.3.9. Note that every actual automaton has a quasidefinite table.

Definition 1.3.18. A PA A is an actual automaton iff $m_{ij}(\sigma) > 0 \quad \forall\sigma\in\Sigma \quad \forall i,j=1,\ldots,n$.

The use of actual automata is natural in simulating certain systems for which all of the transition probabilities are positive, no matter how small they are.

Theorem 1.3.10. Every definite set is definable by some actual automaton with isolated cut-point.

Paz also extended the theorem of Rabin concerning the effects of small perturbations of the transition probabilities of an automaton. Results in this realm are useful in considering the stability of properties of systems subject to perturbation and approximation.

Theorem 1.3.11. Let $A = (S,M,\pi_0,F)$ be a PA with quasidefinite table. Given any $\delta > 0$, there exists an $\varepsilon = \varepsilon(\delta,A) > 0$ such that for every PA $A_\varepsilon = (S,M',\pi_0',F)$ with $|M(\sigma) - M'(\sigma)| < \varepsilon \quad \forall\sigma\in\Sigma$ and $|\pi_0 - \pi_0'| < \varepsilon$, then $|p(x) - p_\varepsilon(x)| < \delta \quad \forall x\in\Sigma^*$.

<u>Corollary 1.3.12</u>. Let A and A_ε be PA's as in Theorem 1.3.11 and let λ be an isolated cut-point for A. For any sufficiently small $\varepsilon > 0$, λ is an isolated cut-point for A_ε and $T(A,\lambda) = T(A_\varepsilon,\lambda)$.

CHAPTER II

Environments

2.1 The Concept of Environments for Automata

In discussing machines subject to error or probabilistic transition, we have considered those systems whose transition probabilities are only related to the present state and latest input symbol and are independent of any external factors. Now, however, we shall construct automata models to be more general and more realistic. The viewpoint taken is that every machine is located within an environment and the transition function of that machine is not only related to the input but also to the configuration of the environment.

We denote the set of possible values that the environment may assume by E. Presently, we do not make any stipulations as to the origin, form, structure, or cardinality of E. As each new input symbol is read by the machine, the configuration of the environment is observed and the transition ensues as a function of the input symbol and the condition of the environment. We also view the initial distribution as a function of the environment. Thus, we shall consider the total environment for a machine. The total environment is the sequence of environments that consists of the initial and the subsequent environments as each symbol is read.

Definition 2.1.1. Let E^∞ denote the cartesian product of a countable number of copies of the environment set E. A total environment, or environment sequence, is a mapping $\underline{e}(k) = e_k \epsilon E$ for $k=0,1,2,\ldots$. We may also denote $\underline{e}$ by the sequence $(e_0,e_1,e_2,\ldots)$.

2.2 Automata in Deterministic Environments

Here we shall consider the case in which the environment is completely specified. That is, for the initial distribution and subsequent transitions we are given the precise condition of the environment. In such a case we say that the environments for the initial distribution and subsequent transitions are obtained by a deterministic rule, the total environment.

Definition 2.2.1. An automaton in a deterministic environment (ADE) is a system $A = (\Sigma,S,G,\pi_0,F,E)$, where Σ is a finite input alphabet, $S = \{s_1,\ldots,s_n\}$ is the finite state set, G is a mapping $G\colon \Sigma\times E \to M_n$ (the basic matrix transition function), $\pi_0\colon E \to V_n$ (the initial distribution function), $F\subset S$ (the set of final states), and E is the set of environments.

The matrix transition function M is defined on $\Sigma^*\times E^\infty$ by the inductive recursion rule:

let $x\epsilon\Sigma^*$ such that $L(x) = k$ and $\underline{e}\epsilon E^\infty$,

then $M(x\sigma,\underline{e}) = M(x,\underline{e})G(\sigma,e_{k+1})$.

Hence, for all $x\epsilon\Sigma^*$ and $\underline{e}\epsilon E^\infty$, $M(x,\underline{e}) = \prod_{i=1}^{k} G(\sigma_i,e_i)$, where $x = \sigma_1 \ldots \sigma_k$ and $\underline{e} = (e_0,e_1,\ldots)$. In considering the function π_0 we shall use the notation $\pi_0(e_0)$ and $\pi_0(\underline{e})$ interchangeably, where $\underline{e} = (e_0,e_1,\ldots)$.

As with probabilistic automata we are able to compute the state distribution of A after input x. Now, however, this will depend on the total environment $\underline{e}$. We denote the state distribution of A after input x in total environment $\underline{e}$ by $\pi(x,\underline{e}) = \pi_0(e_0)M(x,\underline{e})$. Similarly, we let $p(x,\underline{e}) = \pi(x,\underline{e})\eta^F$, where η^F is as in definition 1.3.4, denote the probability that A is in a state in F after input x in total environment $\underline{e}$.

<u>Definition 2.2.2</u>. Let A be an ADE and λ a real number, $\lambda\in[0,1)$. The set of tapes $T(A,\underline{e},\lambda)$ defined by

$$T(A,\underline{e},\lambda) = \{x \mid x\in\Sigma^*,\ \lambda<p(x,\underline{e})\}$$

is called the set of tapes accepted, or defined, by A in total environment $\underline{e}$ with cut-point λ.

<u>Definition 2.2.3</u>. Let $Q(e_0,\ldots,e_k) = \{\underline{e}\in E^\infty \mid \underline{e}(i) = e_i \text{ for } i=0,1,\ldots k\}$ and $Q'(e_1,\ldots,e_k) = \{\underline{e}\in E^\infty \mid \underline{e}(i) = e_i \text{ for } i=1,2,\ldots k\}$.

<u>Lemma 2.2.1</u>. Let A be an ADE. For any $k \geq 0$ and any fixed $e_0,e_1,\ldots,e_k\in E$, the following functions are constant in $Q(e_0,\ldots,e_k)$: $M(x,\cdot)$, $\pi(x,\cdot)$, and $p(x,\cdot)$ for each $x\in\Sigma^*$ such that $L(x) \leq k$.

<u>Proof</u>: Let $k \geq 0$; fix $e_0,e_1,\ldots,e_k\in E$ and fix $x = \sigma_1,\ldots,\sigma_r$ such that $L(x) = r \leq k$. Let $\underline{e}'\in Q(e_0,\ldots,e_k)$. Then $\underline{e}' = (e_0,e_1,\ldots,e_k,e'_{k+1},e'_{k+2},\ldots)$, where $e'_{k+1},e'_{k+2},\ldots\in E$. So $M(x,\underline{e}') = \prod_{i=1}^{r} G(\sigma_i,e_i)$. Thus, $M(x,\cdot)$ is constant in $Q(e_0,\ldots,e_k)$. Similarly, $\pi(x,\underline{e}') = \pi_0(e_0)\prod_{i=1}^{r} G(\sigma_i,e_i)$ and $p(x,\underline{e}') = \pi_0(e_0)\left(\prod_{i=1}^{r} G(\sigma_i,e_i)\right)\eta^F$ are also constant in $Q(e_0,\ldots,e_k)$. Note that $M(x,\cdot)$ is also constant in $Q'(e_1,\ldots,e_k)$. □

2.3 Mathematical Generality of the ADE Model

Let $A' = (S',M',\pi_0',F')$ be any probabilistic automaton over the alphabet Σ'. Let E be any nonempty set; let us fix $e \in E$. We can now construct an ADE $A = (\Sigma,S,G,\pi_0,F,E)$ such that $T(A',\lambda) = T(A,(e,e,\ldots),\lambda)$ for all $\lambda \in [0,1)$. We simply let $\Sigma=\Sigma'$, $S=S'$, $F=F'$, $\pi_0(e)=\pi_0'$, and $G(\sigma,e) = M'(\sigma)\ \forall\ \sigma\in\Sigma$. Thus, $p(x,\underline{e}) = \pi_0(e)M(x,(e,e,\ldots))\eta^F = \pi_0'\left(\prod_{i=1}^{k} G(\sigma_i,e)\right)\eta^F = \pi_0'\left(\prod_{i=1}^{k} M'(\sigma)\right)\eta^F = p'(x)$. Therefore, for any $\lambda\in[0,1)$ $x\in T(A',\lambda)$ iff $x\in T(A,(e,e,\ldots),\lambda)$ and, hence, every PCE can be defined by an ADE with some total environment $\underline{e}\in E^\infty$.

Let $\Sigma = \{\sigma\}$. Paz [14] demonstrated that there exists $U\subset\Sigma^*$ such that U is not a PCE. Let us now consider a two state ADE A over Σ^* where $E = \{0,1\}$, $\pi_0(e) = (1-e,e)$, $G(\sigma,e) = \begin{bmatrix} 1-e & e \\ 1-e & e \end{bmatrix}\ \forall e\in E$, and $\eta^F = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. Let us denote a tape of k repetitions of the symbol σ by $\sigma^k = \sigma\,\sigma\ldots\sigma$. Note that $\sigma^0 = \Lambda$, the empty tape. Thus, for our particular situation all of the elements of Σ^* are of the form σ^k for some $k \geq 0$. Let U be any subset of Σ^*; then $U = \{\sigma^{k_1},\sigma^{k_2},\ldots\}$. We construct $\underline{e}(k) = \begin{cases} 1 & k=k_i \text{ for some } i\geq 0 \\ 0 & \text{elsewhere} \end{cases}$

Such an $\underline{e}$ is an element of E^∞. Now, for $k\geq 1$ we have

$$M(\sigma^k,\underline{e}) = \begin{cases} \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} & \text{if } k = k_i \text{ for some } i \\ \\ \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} & \text{elsewhere .} \end{cases}$$

Therefore, $p(\sigma^k,\underline{e}) = \pi_0(e_0)M(\sigma^k,\underline{e})\eta^F = \begin{cases} 1 & \text{if } k=k_i \text{ for some } i \\ 0 & \text{elsewhere.} \end{cases}$

When $k = 0$, $p(\sigma^0,\underline{e}) = p(\Lambda,\underline{e}) = \pi_0(e_0)\eta^F = \begin{cases} 1 & \text{if } k_i=0 \text{ for some } i \\ 0 & \text{elsewhere .} \end{cases}$

Let $\lambda = 0$, then $U = T(A,\underline{e},\lambda)$. Since U is arbitrary, every subset of our particular Σ^* is definable by A under some environment. But there exists $U \subset \Sigma^*$ which is not definable by a probabilistic automaton. Thus, the class of automata in deterministic environments properly contains the class of probabilistic automata.

Let A be an ADE and let $\underline{e} \in E^\infty$. For any $x, y \in \Sigma^*$ such that $\pi(x,\underline{e}) = \pi(y,\underline{e})$, then $\pi(xz,\underline{e}) = \pi(yz,\underline{e})$ for all $z \in \Sigma^*$. Hence, for any $\underline{e} \in E^\infty$, the state distribution function $\pi(\circ,\underline{e})$ is a right invariant equivalence relation on Σ^*. As with probabilistic automata, for any fixed $\underline{e} \in E^\infty$ there may be up to a countable number of equivalence classes of the relation $\pi(\circ,\underline{e})$. For any $\lambda \in [0,1)$, $T(A,\underline{e},\lambda)$ is the union of the equivalence classes satisfying the condition that $\pi(x,\underline{e})\eta^F > \lambda$. The greater generality of the ADE model manifests itself in the fact that the equivalence relation and, hence, the equivalence classes may depend on $\underline{e}$. Depending on A, there may be an uncountable number of unique equivalence relations. The above two state ADE has an uncountable number of unique equivalence relations.

2.4 Finite Environment Set

Let A be an ADE with a finite environment set. We shall construct an effective procedure to find a probabilistic automaton which simulates the operation of A.

Definition 2.4.1. A PA A' over alphabet Σ' simulates the ADE $A = (\Sigma, S, G, \pi_0, F, E)$ if there is a relation R between Σ'^* and $\Sigma^* \times E^\infty$ such that

1) for any $(x,\underline{e}) \in \Sigma^* \times E^\infty$ there exists $x' \in \Sigma'^*$ such that $[x',(x,\underline{e})] \in R$;

2) if $[x',(x,\underline{e})]\in R$, then for each $\lambda\in[0,1)$, $x'\in T(A',\lambda)$ iff $x\in T(A,\underline{e},\lambda)$;

3) if for some $\lambda\in[0,1)$ $x'\in T(A',\lambda)$, then there exists $(x,\underline{e})\in\Sigma^*\times E^\infty$ such that $[x',(x,\underline{e})]\in R$.

If a PA A' simulates an ADE A, given any input x for A under total environment $\underline{e}$, we have a rule R to find an input tape x' for A' such that x' has the same acceptance properties as x for A under total environment $\underline{e}$. A tape $x'\in\Sigma'^*$ is admissible if and only if there is an input tape x for A under some total environment $\underline{e}$ from which the rule R will yield x'. By condition 3 of definition 2.4.1 we see that tapes which are not admissible are not accepted by A' for any $\lambda\in[0,1)$; that is, if x' is not admissible, then the probability of acceptance of x' by A' is zero. By lemma 2.2.1 we see that the probability of acceptance of any tape $x\in\Sigma^*$ depends only on the environments $e_0,e_1,\ldots,e_k$, where $k = L(x)$. Hence, the relation R need only depend on x and $e_0,e_1,\ldots,e_k$, where $k = L(x)$. Thus, all pairs of the form $(x,\underline{e})$, where $\underline{e}\in Q(e_0,\ldots,e_k)$ and $k = L(x)$, can be considered equivalent.

<u>Theorem 2.4.1</u>. Let $A = (\Sigma,S,G,\pi_0,\mathcal{F},E)$ be an ADE such that $\#(E) < \infty$[1]. There exists a PA A' over some finite alphabet Σ' which simulates A.

<u>Proof</u>: Let $m = \#(\Sigma)$ and $E = \#(E)$. Without loss of generality we shall let $\Sigma = \{1,2,\ldots,m\}$ and $E = \{0,1,\ldots,E-1\}$. Let us consider $\Sigma' = \{0,1,\ldots,E(m+1)-1\}$. Let ϕ be any symbol such that $\phi\notin\Sigma$. We now define a mapping $g: (\Sigma\cup\{\phi\})\times E \to \Sigma'$ as follows:

---

[1] $\#(E)$ denotes the cardinality of the set E.

$$g(\sigma,e) = e + \sigma E \qquad \forall\ \sigma\epsilon\Sigma \text{ and } \forall e\epsilon E$$

$$g(\phi,e) = e \qquad \forall\ e\epsilon E\ .$$

This mapping is a one-to-one correspondence. We extend g to $((\Sigma\cup\{\phi\})\times E)^*$, the set of all finite sequences of symbols of $(\Sigma\cup\{\phi\})\times E$, by component-wise application of the mapping and concatenation of the results. That is, $a_1,a_2,\ldots,a_k\epsilon(\Sigma\cup\{\phi\})\times E$, then $a_1\ a_2 \ldots a_k \epsilon((\Sigma\cup\{\phi\})\times E)^*$ and $g(a_1a_2\ldots a_k) = g(a_1)g(a_2)\ldots g(a_k)$. g remains a one-to-one correspondence between $((\Sigma\cup\{\phi\})\times E)^*$ and Σ'^* Let b be any nonempty element of $((\Sigma\cup\{\phi\})\times E)^*$, then b is isomorphic to the form $(y,(e_0,\ldots,e_k))$, where for some $k \geq 0$ $y\epsilon(\Sigma\cup\{\phi\})^*$, $L(y) = k+1$, and $e_0,\ldots,e_k\epsilon E$. We define the relation F between Σ'^* and $\Sigma^*\times E^\infty$ to be the set of all elements of the form: $[g(\phi x,(e_0,\ldots,e_k)),(x,\underline{e})]$, where $x\epsilon\Sigma^*$, $L(x) = k$, and $\underline{e} = (e_0,\ldots,e_k,\ldots)$. Note that when $x = \Lambda\epsilon\Sigma^*$, then $L(\Lambda) = 0$ and $\phi\Lambda = \phi$.

We shall now construct a PA A' to simulate A. Let $A' = (S',M',\pi_0',F')$ be a PA over Σ'. Let $S' = S\cup\{s_{n+1},s_{n+2}\}$, where $S = \{s_1,\ldots,s_n\}$ is the state set of A. For any $\sigma'\epsilon E\subset\Sigma'$, then $0 \leq \sigma' \leq E-1$. So for any σ' such that $0 \leq \sigma' \leq E-1$, we define

$$M'(\sigma') = \begin{pmatrix} & & & . & 0 & 1 \\ & 0 & & \vdots & \vdots & \vdots \\ \cdots & \cdots & \cdots & . & 0 & 1 \\ \pi_0^{(1)}(\sigma') & \cdots & \pi_0^{(n)}(\sigma') & . & 0 & 0 \\ 0 & \cdots & 0 & . & 0 & 1 \end{pmatrix}$$

where $\pi_0^{(i)}(\sigma')$ is the i-th component of $\pi_0(\sigma')$. Note that $0 \leq \sigma' \leq E-1$ implies $\sigma'\epsilon E$. These values of σ' are used to incorporate the initial distribution function of A into the PA model A'. The values

$\sigma'\epsilon\Sigma'$ such that $E \le \sigma \le E(m+1)-1$ are used in order that A' can imitate the basic matrix transition function of A. The set $\{\sigma' | E \le \sigma' \le E(m+1)-1\}$ is in one-to-one correspondence with the set $\Sigma\times E$ under the mapping g. So for $E \le \sigma' \le E(m+1)-1$, define

$$M'(\sigma') = \begin{pmatrix} & G(g^{-1}(\sigma')) & & \begin{matrix} . \\ \vdots \\ . \end{matrix} & \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix} & \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix} \\ \cdots & \cdots & \cdots & . & & \\ 0 & \cdots & 0 & . & 0 & 1 \\ 0 & \cdots & 0 & . & 0 & 1 \end{pmatrix}$$

where G is the basic matrix transition function of A. Also, let $\pi'_0 = (0,\ldots,0,1,0)$ and $F' = F$. Therefore $\eta^{F'} = \begin{pmatrix} \eta^F \\ 0 \\ 0 \end{pmatrix}$.

It is clear from the definition of R that for any $(x,\underline{e})\epsilon\Sigma\times E^\infty$ there exists a unique $x' = g(\phi x,(e_0,\ldots,e_k))$, where $k = L(x)$, such that $[x',(x,\underline{e})]\epsilon R$.

Suppose $[x',(\Lambda,\underline{e})]\epsilon R$. Therefore, $x' = g(\phi,e_0) = e_0$. Let $p'(x')$ be the probability that A' accepts x' and $p(x,\underline{e})$ be the probability that A accepts x under total environment $\underline{e}$. Hence,

$$p'(x') = p'(e_0) = \pi'_0 M'(e_0)\eta^{F'}$$

$$= (0,\ldots,0,1,0) \begin{pmatrix} & 0 & & \begin{matrix} . \\ \vdots \\ . \end{matrix} & \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix} & \begin{matrix} 1 \\ \vdots \\ 1 \end{matrix} \\ \cdots & \cdots & \cdots & . & & \\ \pi_0^{(1)}(e_0) & \cdots & \pi_0^{(n)}(e_0) & . & 0 & 0 \\ 0 & \cdots & 0 & . & 0 & 1 \end{pmatrix} \begin{pmatrix} \eta^F \\ 0 \\ 0 \end{pmatrix}$$

$$= \pi_0(e_0)\eta^F = \pi_0(e_0)M(\Lambda,\underline{e})\eta^F = p(\Lambda,\underline{e}) .$$

Thus, for each $\lambda\epsilon[0,1)$, $p'(x') = p'(e_0) > \lambda$ iff $p(\Lambda,\underline{e}) > \lambda$. That is, if $[x',(\Lambda,\underline{e})]\epsilon R$, then for each $\lambda\epsilon[0,1)$, $x'\epsilon T(A',\lambda)$ iff $\Lambda\epsilon T(A,\underline{e},\lambda)$.

Suppose $[x',(x,\underline{e})]\epsilon R$, where $x \neq \Lambda$. Hence, $L(x) > 0$. By the definition of R we have that $x' = g(\phi x,(e_0,\ldots,e_k))$, where $x = \sigma_1,\ldots,\sigma_k$ has length k. So $x' = g(\phi,e_0)g(\sigma_1,e_1),\ldots,g(\sigma_k,e_k)$ $= e_0\sigma_1',\ldots,\sigma_k'$. Thus, $p'(x') = p'(e_0\sigma_1',\ldots,\sigma_k') = \pi_0'M'(e_0\sigma_1' \ldots \sigma_k')\eta^{F'}$. But

$$\prod_{i=1}^{k} M'(\sigma_i') = \prod_{i=1}^{k} \left(\begin{array}{ccc|cc} & G(g^{-1}(\sigma_i')) & & 0 & 0 \\ & & & \vdots & \vdots \\ & & & 0 & 0 \\ \cdot & \cdots & \cdot & & \\ 0 & \cdots & 0 & 0 & 1 \\ 0 & \cdots & 0 & 0 & 1 \end{array}\right)$$

$$= \prod_{i=1}^{k} \left(\begin{array}{ccc|cc} & G(\sigma_i,e_i) & & 0 & 0 \\ & & & \vdots & \vdots \\ & & & 0 & 0 \\ \cdot & \cdots & \cdot & & \\ 0 & \cdots & 0 & 0 & 1 \\ 0 & \cdots & 0 & 0 & 1 \end{array}\right)$$

$$= \left(\begin{array}{ccc|cc} & \prod_{i=1}^{k} G(\sigma_i,e_i) & & 0 & 0 \\ & & & \vdots & \vdots \\ & & & 0 & 0 \\ \cdot & \cdots & \cdot & & \\ 0 & \cdots & 0 & 0 & 1 \\ 0 & \cdots & 0 & 0 & 1 \end{array}\right)$$

$$= \left(\begin{array}{ccc|cc} & M(x,\underline{e}) & & 0 & 0 \\ & & & \vdots & \vdots \\ & & & 0 & 0 \\ \cdot & \cdots & \cdot & & \\ 0 & \cdots & 0 & 0 & 1 \\ 0 & \cdots & 0 & 0 & 1 \end{array}\right)$$

Hence, we have

$$p'(x') = \pi_0' M'(e_0) \begin{pmatrix} M(x,\underline{e}) & \begin{matrix} 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{matrix} \\ \cdots & \cdots \\ 0. \; . \; .0 & 0 \;\; 1 \\ 0. \; . \; .0 & 0 \;\; 1 \end{pmatrix} \eta^{F'}$$

$$= (0,\ldots,0,1,0) \begin{pmatrix} & 0 & & \begin{matrix} 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{matrix} \\ \cdot & \cdots & \cdot & \cdots \\ \pi_0^{(1)}(e_0) & \cdots & \pi_0^{(n)}(e_0) & 0 \;\; 0 \\ 0 & & 0 & 0 \;\; 0 \end{pmatrix} \begin{pmatrix} M(x,\underline{e}) & \begin{matrix} 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{matrix} \\ \cdots & \cdots \\ 0 \; . \; . \; . \; 0 & 0 \;\; 1 \\ 0 \; . \; . \; . \; 0 & 0 \;\; 1 \end{pmatrix} \begin{pmatrix} \eta^F \\ 0 \\ 0 \end{pmatrix}$$

$$= \pi_0(e_0) M(x,\underline{e}) \eta^F = p(x,\underline{e}).$$

So, if $[x',(x,\underline{e})] \in R$ and $x \neq \Lambda$, then for each $\lambda \in [0,1)$ $p'(x') > \lambda$ iff $p(x,\underline{e}) > \lambda$. Thus, $x' \in T(A',\lambda)$ iff $x \in T(A,\underline{e},\lambda)$.

We have verified that $\forall\, x \in \Sigma^*$ and $\forall\, \underline{e} \in E^\infty$ if $[x',(x,\underline{e})] \in R$, then for each $\lambda \in [0,1)$, $x' \in T(A',\lambda)$ iff $x \in T(A,\underline{e},\lambda)$.

Let $x' \in \Sigma'^*$ such that there does not exist $x \in \Sigma^*$ and $\underline{e} \in E^\infty$ so that $[x',(x,\underline{e})] \in R$. Hence, there does not exist $x \in \Sigma^*$ and $\underline{e} \in E^\infty$ such that $x' = g(\phi x,(e_0,\ldots,e_k))$, where $k = L(x)$. Therefore, such an x' is not admissible. However, $x' = g(b)$ for some $b \in ((\Sigma \cup \{\phi\}) \times E)^*$ since g is a one-to-one correspondence between $((\Sigma \cup \{\phi\}) \times E)^*$ and Σ'^*. Λ', the empty tape for A', is the image of Λ^0, the empty tape in $((\Sigma \cup \{\phi\}) \times E)^*$, under g. Clearly, Λ' is not admissible. $p'(\Lambda') = \pi_0' \eta^{F'} = (0,\ldots,0,1,0) \begin{pmatrix} \eta^F \\ 0 \\ 0 \end{pmatrix} = 0$. Hence $\Lambda' \notin T(A',\lambda)$ for any $\lambda \in [0,1)$.

Let $x' \in \Sigma'^*$ be any tape which is not admissible and $x' \neq \Lambda'$. $x' = \sigma_0' \ldots \sigma_k'$ is the image under g of some element $b \in ((\Sigma \cup \{\phi\}) \times E)^*$, where b has the form $(y,(e_0,\dots,e_k))$ and $L(x') = L(y) = k+1$. Note that $y\epsilon(\Sigma\cup\{\phi\})^*$. Let $y = \tau_0\ \tau_1 \dots \tau_k$, where $\tau_i\epsilon\Sigma\cup\{\phi\}$ for $i=0,1,\dots,k$. x' is not admissible iff $\tau_0 \neq \phi$ or $\tau_i \notin \Sigma$ for some $i=1,2,\dots,k$. If $\tau_0 \neq \phi$, then $\tau_0\epsilon\Sigma$. $\sigma_0' = g(\tau_0,e_0)$ and, hence $E \le \sigma_0' \le E(m+1)-1$. By our definition of M' we see that A' under input $\sigma_0'\ \sigma_1' \dots \sigma_k'$ enters state s_{n+2} at the first transition. But s_{n+2} is an absorbing state and $s_{n+2} \notin F=F'$. Hence, $p'(x') = 0$. So $x' \notin T(A',\lambda)$ for any $\lambda\epsilon[0,1)$. If $\tau_i \notin \Sigma$ for some $i=1,2,\dots,k$; say $\tau_j \notin \Sigma$. Then $\tau_j = \phi$. $\sigma_j' = g(\tau_j,e_j)$ and, hence, $0 \le \sigma_j' \le E-1$. By our definition of M' we see that A' under input $\sigma_0'\ \sigma_1' \dots \sigma_k'$ enters state s_{n+2} at the $(j+1)$th transition. Again, s_{n+2} is an absorbing state and $s_{n+2} \notin F=F'$. Hence, $p'(x') = 0$ and $x'\epsilon T(A',\lambda)$ for any $\lambda\epsilon[0,1)$.

Thus, if for some $\lambda\epsilon[0,1)$ $x'\epsilon T(A',\lambda)$, then x' is admissible; that is, there exists $(x,\underline{e})\epsilon\Sigma^*\times E^\infty$ such that $[x',(x,\underline{e})]\epsilon R$.

Consequently, A', as constructed, simulates A. □

The results of section 2.3 and theorem 2.4.1 seem to present a paradox. We have found for a particular Σ^* that there exists $U\subset\Sigma^*$ which is not a PCE; that is, for any $\lambda\epsilon[0,1)$ there is no probabilistic automaton A_p such that $U=T(A_p,\lambda)$. However, we showed that there exists an ADE with finite environment set over Σ such that $U=T(A,\underline{e},\lambda)$ for some $\underline{e}\epsilon E^\infty$ and $\lambda\epsilon[0,1)$. By theorem 2.4.1 there exists a probabilistic automaton A' over an expanded alphabet Σ' such that A' simulates A. Let $U' = \{g(\phi x,\underline{e})\,|\,x\epsilon U\}$, where $\underline{e}$ is the total environment whereby $U = T(A,\underline{e},\lambda)$. $U' = T(A',\lambda)$ and, hence, is a PCE. The paradox is resolved by noting that we not only have expanded Σ to Σ'

but also enriched the alphabet with a structure that had been included in the environment. The problem for $U \subset \Sigma^*$ remains because U' has a different structure within Σ'^*. In fact, we can find a subset of Σ'^* which is not a PCE.

<u>Example 2.4.1</u>. As an example of the application of theorem 2.4.1, let us consider the ADE discussed in section 2.3. A is a two-state ADE over $\Sigma = \{\sigma\}$ with $E = \{0,1\}$. $\pi_0(e) = (1-e,e)$, $G(\sigma,e) = \begin{pmatrix} 1-e & e \\ 1-e & e \end{pmatrix}$, and $\eta^F = \binom{0}{1}$. By theorem 2.4.1 A' has alphabet $\Sigma' = \{0,1,2,3\}$ and M' is defined on Σ' as follows

$$M'(0) = \begin{bmatrix} 0&0&0&1 \\ 0&0&0&1 \\ 1&0&0&0 \\ 0&0&0&1 \end{bmatrix} \qquad M'(1) = \begin{bmatrix} 0&0&0&1 \\ 0&0&0&1 \\ 0&1&0&0 \\ 0&0&0&1 \end{bmatrix}$$

$$M'(2) = \begin{bmatrix} 1&0&0&0 \\ 1&0&0&0 \\ 0&0&0&1 \\ 0&0&0&1 \end{bmatrix} \qquad M'(3) = \begin{bmatrix} 0&1&0&0 \\ 0&1&0&0 \\ 0&0&0&1 \\ 0&0&0&1 \end{bmatrix}$$

The mapping g is described by:

$$g(\phi,e) = e \quad \text{and} \quad g(\sigma,e) = 2+e .$$

With $\pi'_0 = (0,0,1,0)$ and $\eta^F = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$, it is easily verified by the same procedure as the proof that for any $\lambda \in [0,1)$ $T(A',\lambda) = \{g(\phi x,\underline{e}) \mid \underline{e} \in E^\infty$ and $x \in T(A,\underline{e},\lambda)\}$ and that A' simulates A.

<u>Corollary 2.4.2</u>. Let $A = (\Sigma,S,G,\pi_0,F,E)$ be an ADE. For any fixed $e \in E^\infty$ such that $E = \{e_i \mid i=0,1,\ldots\}$ is a finite set, then there exists a PA A' over some alphabet Σ' such that for any $x \in \Sigma^*$

and $\lambda\epsilon[0,1)$, $x\epsilon T(A,\underline{e},\lambda)$ iff $g_E(x,\underline{e})\epsilon T(A',\lambda)$, where g_E is defined for the ADE $A_E = (\Sigma,S,G,\pi_0,F,E)$ as the function g in theorem 2.4.1.

2.5 Reducing the Environment Set and the Induced Metric

Let $A = (\Sigma,S,G,\pi_0,F,E)$ be an ADE and let e and e' be two distinct configurations of the environment. If $\pi_0(e) = \pi_0(e')$ and $G(\sigma,e) = G(\sigma,e')$ for all $\sigma\epsilon\Sigma$, then e and e' have identical effects on the transition functions. That is, the configurations e and e' are equivalent in the sense that for any $x\epsilon\Sigma^*$ and $\underline{e}\epsilon E^\infty$ $\pi(x,\underline{e})$ remains constant if we substitute e for e' whenever e' occurs in $\underline{e}$. Thus, we do not need to include e' in the environment set since it does not augment the structure of A. Moreoever, we can reduce E so that no two distinct environments are equivalent in the above sense.

Definition 2.5.1. Let $A = (\Sigma,S,G,\pi_0,F,E)$ be an ADE, then we say E is irreducible if for all $e, e'\epsilon E$ $e = e'$ iff $\pi_0(e) = \pi_0(e')$ and $G(\sigma,e) = G(\sigma,e')$ for all $\sigma\epsilon\Sigma$.

Let us assume now that E is an irreducible environment set for the ADE $A = (\Sigma,S,G,\pi_0,F,E)$. Recalling definition 1.3.14, we construct a metric on E. Consider the function $d(e,e') = \max(|\pi_0(e)-\pi_0(e')|, \max_{\sigma\epsilon\Sigma}|G(\sigma,e) - G(\sigma,e')|)$, where $e, e'\epsilon E$. Clearly, $d(e,e') = d(e',e)$ and $d(e,e') \geq 0$ for all $e,e'\epsilon E$. For any $e,e', e''\epsilon E$, then

$$\begin{aligned} d(e,e') &= \max(|\pi_0(e)-\pi_0(e')|, \max_{\sigma\epsilon\Sigma}|G(\sigma,e)-G(\sigma,e')|) \\ &\leq \max(|\pi_0(e)-\pi_0(e'')| + |\pi_0(e'')-\pi_0(e')|, \\ &\qquad \max_{\sigma\epsilon\Sigma}(|G(\sigma,e)-G(\sigma,e'')| + |G(\sigma,e'')-G(\sigma,e)|) \\ &\leq \max(|\pi_0(e)-\pi_0(e'')|, \max_{\sigma\epsilon\Sigma}|G(\sigma,e)-G(\sigma,e'')|) \end{aligned}$$

$$+ \max(|\pi_0(e')-\pi_0(e'')|, \max_{\sigma\in\Sigma}|G(\sigma,e')-G(\sigma,e'')|)$$

$$= d(e,e'') + d(e'',e').$$

So d satisfies the triangle property. Also, if $e = e'$, then $\pi_0(e) = \pi_0(e')$ and $G(\sigma,e) = G(\sigma,e')$ for all $\sigma\in\Sigma$. So $d(e,e') = 0$ when $e = e'$. Since E is irreducible, if $e \neq e'$, then by definition $d(e,e') > 0$. So the function d as defined above is a metric on E.

Theorem 2.5.1. If E is an irreducible environment for an ADE A, then the function $d(e,e') = \max(|\pi_0(e)-\pi_0(e')|, \max_{\sigma\in\Sigma}|G(\sigma,e)-G(\sigma,e')|)$ is defined for all $e,e'\in E$ and is a metric on E.

Since irreducibility is needed only to show that if $e \neq e'$, then $d(e,e') \neq 0$, the function d is a pseudo-metric on any arbitrary environment set.

Let us now consider an arbitrary ADE $A = (\Sigma,S,G,\pi_0,F,E)$ for which the environment set is reducible to a finite set $E^\dagger$. That is, we can define an onto mapping $h\colon E \to E^\dagger \subset E$, where if $h(e) = e^\dagger$, then $d(e,e^\dagger) = 0$. The mapping h is unique only when $E^\dagger$ is irreducible. h can be extended to $h\colon E^\infty \to (E^\dagger)^\infty$ by component-wise application of h. That is, $h(\underline{e}) = (h(e_0),h(e_1),\ldots)$. So for all $e\in E$, $d(e,h(e)) = 0$ and, hence, we know that $\pi_0(e) = \pi_0(h(e))$ and $G(\sigma,e) = G(\sigma,h(e))$ for all $\sigma\in\Sigma$. So by extension, for all $\underline{e}\in E^\infty$, $M(x,\underline{e}) = M(x,h(\underline{e}))$, $\pi(x,\underline{e}) = \pi(x,h(\underline{e}))$, and $p(x,\underline{e}) = p(x,h(\underline{e}))$ for all $x\in\Sigma^*$.

Theorem 2.5.2. Let $A = (\Sigma,S,G,\pi_0,F,E)$ be any ADE, then $T(A,\underline{e},\lambda) = T(A,h(\underline{e}),\lambda)$ for all $\underline{e}\in E^\infty$ and $\lambda\in[0,1)$. Furthermore, if $\#(h(E)) < \infty$, then A can be simulated by a PA.

Proof: Let $\underline{e} \in E^\infty$. We have seen that $p(x,\underline{e}) = p(x,h(\underline{e}))$ for all $x \in \Sigma^*$. So for any $\lambda \in [0,1)$ $p(x,\underline{e}) > \lambda$ iff $p(x,h(\underline{e})) > \lambda$. Hence, $x \in T(A,\underline{e},\lambda)$ iff $x \in T(A,h(\underline{e}),\lambda)$; that is, $T(A,\underline{e},\lambda) = T(A,h(\underline{e}),\lambda)$.

Let $E^\dagger = h(E)$. So if $\#(E^\dagger) < \infty$, then by theorem 2.4.1 there exists a PA A' over alphabet Σ' which simulates the ADE $A^\dagger = (\Sigma,S,G,\pi_0,F,E^\dagger)$. Thus, there exists a relation $R^\dagger$ between Σ'^* and $((\Sigma\cup\{\phi\})\times E^\dagger)^*$ satisfying the conditions of definition 2.4. We now define a relation R between Σ'^* and $((\Sigma\cup\{\phi\})\times E)^*$ by the rule: $[x',(x,\underline{e})] \in R$ iff $[x',(x,h(\underline{e}))] \in R^\dagger$. Clearly, for all $(x,\underline{e}) \in \Sigma^* \times E^\infty$, there exists $x' \in \Sigma'^*$, such that $[x',(x,\underline{e})] \in R$. If $[x',(x,\underline{e})] \in R$, then $[x',(x,h(\underline{e}))] \in R^\dagger$. Hence, for any $\lambda \in [0,1)$ $x' \in T(A',\lambda)$ iff $x \in T(A^\dagger,h(\underline{e}),\lambda)$. But $T(A,\underline{e},\lambda) = T(A^\dagger,h(\underline{e}),\lambda)$ for all $\lambda \in [0,1)$. Therefore, if $[x',(x,\underline{e})] \in R$, then for each $\lambda \in [0,1)$ $x' \in T(A',\lambda)$ iff $x \in T(A,\underline{e},\lambda)$. If for any $x' \in \Sigma'^*$ there does not exist $(x,\underline{e}) \in \Sigma^* \times E^\infty$ such that $[x',(x,\underline{e})] \in R$, then since h is an onto mapping, there does not exist $(x,\underline{e}^\dagger) \in \Sigma^* \times (E^\dagger)^*$ such that $[x',(x,\underline{e}^\dagger)] \in R^\dagger$, and conversely. That is, x' is not admissible under the relation R iff x' is not admissible under the relation $R^\dagger$. Hence, A' simulates A via the relation R as defined. □

Corollary 2.5.3. Let $A = (\Sigma,S,G,\pi_0,F,E)$ be an ADE. For any fixed $\underline{e} \in E^\infty$ such that $E = \{e_i | i=0,1,\ldots\}$ is reducible to a finite set, then there exists a PA A' over some alphabet Σ' such that for any $x \in \Sigma^*$ and $\lambda \in [0,1)$, $x \in T(A,\underline{e},\lambda)$ iff $g(x,h(\underline{e})) \in T(A',\lambda)$, where g is defined in theorem 2.4.1 and h is defined in theorem 2.5.2.

2.6 The Environment Set as a Semi-group

Definition 2.6.1. The pair $(W,\circ)$, where W is any set and $\circ$ is any binary operation, is a semi-group if W is closed under the operation $\circ$ and $\circ$ is associative.

The set M_n of all $n\times n$ stochastic matrices and the multiplication operation forms a semi-group. In Section 1.2 we observed that any Σ^* with the operation of concatenation forms a semi-group. Let A be any PA and $x,y\epsilon\Sigma^*$, then $M(xy) = M(x)M(y)$. Since M_n is associative under the operation of multiplication, A defines a subsemi-group $\{M(x)|x\epsilon\Sigma^*\}$. We shall now generalize this result.

Lemma 2.6.1. Let $(W,\circ)$ be any semi-group. $\{M(w)\epsilon M_n|w\epsilon W\}$ is a semi-group iff the mapping $h: w \rightarrow M(w)$ is a homomorphism.

Proof: Obvious from the definition of homomorphism and the fact that M_n is a semi-group.

Suppose we have an ADE $A = (\Sigma,S,G,\pi_0,F,E)$, where E and some binary operation $\circ$ form a semi-group. For each $\sigma\epsilon\Sigma$ let the mapping $G(\sigma,\cdot) : E \rightarrow M_n$ be a homomorphism. By lemma 2.6.1 for each $\sigma\epsilon\Sigma$ the set $G(\sigma) = \{G(\sigma,e)|e\epsilon E\}$ is a semi-group under the operation of multiplication. That is, for any $e_1,e_2\epsilon E$, then $e_1\circ e_2\epsilon E$ and $G(\sigma,e_1\circ e_2) = G(\sigma,e_1)G(\sigma,e_2)$. Also, since $G(\sigma) \subset M_n$, multiplication of elements of $G(\sigma)$ is associative. Under certain conditions an ADE on an environment set which is a semi-group may be simulated by a probabilitic automaton.

<u>Theorem 2.6.1.</u> Let $A = (\Sigma, S, G, \pi_0, F, E)$ be an ADE such that:

1. $E = S(E^+)$, where E^+ is any finite set and $S(E^+)$ denotes the semi-group generated by E^+ and the binary operation $\circ$;
2. for any $\sigma \epsilon \Sigma$ the mapping $G_\sigma : E \to M_n$, defined as $G_\sigma(e) = G(\sigma, e)$, is a homomorphism;
3. $\pi_0(\cdot)$ is any fixed constant vector in V_n for all $e \epsilon E$.

Then there exists a PA A' which simulates A.

<u>Proof</u>: $E = S(E^+)$ implies $E^+ \subset E$. Also, any $e \epsilon E$ can be expressed as a finite product of elements of E^+. That is, $e = e_1^+ \circ \ldots \circ e_k^+$, where $e_i^+ \epsilon E^+$ $i=1,\ldots,k$ and k is finite. Since for any $\sigma \epsilon \Sigma$ G_σ is a homomorphism, then $G(\sigma, e) = \prod_{i=1}^{k} G(\sigma, e_i^+)$. Clearly, if $e = e_1^{+\prime} \circ \ldots \circ e_{k'}^{+\prime}$, where $e_i^{+\prime} \epsilon E^+$ $i = 1,\ldots,k'$ and k' is finite, is any other decomposition of e, then $\prod_{i=1}^{k} G(\sigma, e_i^+) = \prod_{i=1}^{k'} G(\sigma, e_i^{+\prime})$.

Let $x \epsilon \Sigma^*$ and $\underline{e} \epsilon E^\infty$. Suppose $L(x) = r$. For each $e_i \epsilon E$ $i = 1,\ldots,r$ e_i has a finite decomposition $e_i = e_{i1}^+ \circ \ldots \circ e_{ik_i}^+$, where $e_{ij}^+ \epsilon E^+$ for $j = 1,\ldots,k_i$. Thus, $M(x,\underline{e}) = \prod_{i=1}^{r} G(\sigma_i, e_i) = \prod_{i=1}^{r} \prod_{j=1}^{k_i} G(\sigma_i, e_{ij}^+)$. Let $x^+ = \sigma_1^{k_1} \ldots \sigma_r^{k_r} \epsilon \Sigma^+$, and $\underline{e}^+$ be any element of $Q'(e_{11}^+,\ldots,e_{1k_1}^+, e_{21}^+,\ldots,e_{rk_r}^+)$. Then $p(x,\underline{e}) = \pi_0 M(x,\underline{e}) \eta^F = \pi_0 (\prod_{i=1}^{r} \prod_{j=1}^{k_i} G(\sigma_i, e_{ij}^+)) \eta^F = p(x^+, \underline{e}^+)$. Hence, for all $\lambda \epsilon [0,1)$, $x \epsilon T(A,\underline{e},\lambda)$ iff $x^+ \epsilon T(A,\underline{e}^+,\lambda)$ for all $\underline{e}^+ \epsilon Q'(e_{11}^+,\ldots,e_{1k_1}^+,\ldots,e_{21}^+,\ldots,e_{rk_r}^+)$. Let $e_i = e'_{i1} \circ \ldots \circ e'_{ik'_i}$ be some other decomposition of e_i $i = 1,\ldots,r$. Let $x' = \sigma_1^{k'_1},\ldots,\sigma_r^{k'_r} \epsilon \Sigma^*$ and let $\underline{e}' \epsilon Q'(e'_{11},\ldots,e'_{1k'_1}, e'_{21},\ldots,e'_{k'_r})$. But again, for all $\lambda\epsilon[0,1)$, $x\epsilon T(A,\underline{e},\lambda)$ iff $x'\epsilon T(A,\underline{e}',\lambda)$ for all $\underline{e}'\epsilon Q'(e'_{11},\ldots,e'_{1k_1},e'_{21},\ldots,e'_{rk_r})$. Thus, for all $\lambda\epsilon[0,1)$, $x\epsilon T(A,\underline{e},\lambda)$ iff $x^+\epsilon T(A,\underline{e}^+,\lambda)$ for all $\underline{e}^+\epsilon Q'(e^+_{11},\ldots,e^+_{1k_1},e^+_{21},\ldots,e^+_{rk_r})$ and for any decomposition of e_i $i=1,\ldots,r$. Note that x^+ depends on x and the decomposition of e_i $i = 1,\ldots,r$.

Let us now consider an ADE $A^+ = (\Sigma,S,G,\pi_0 F,E^+)$. For any $x\epsilon\Sigma^*$, where $L(x) = r$, we now restrict the choice of $\underline{e}^+\epsilon Q'(e^+_{11},\ldots,e^+_{1k_1},e^+_{21},\ldots e^+_{rk_r}) \cap (E^+)^\infty$. For any $x\epsilon\Sigma^*$ and $\underline{e}\epsilon E^\infty$, such an $\underline{e}^+$ must exist since $e^+_{ij}\epsilon E^+$ for $i = 1,\ldots,r$, $j = 1,\ldots,k_i$. By the above we see that for any $\lambda\epsilon[0,1)$ $x\epsilon T(A,\underline{e},\lambda)$ iff $x^+\epsilon T(A,\underline{e}^+,\lambda)$, where e^+ is any appropriate decomposed total environment and x^+ is the derived input tape. But $\#(E^+) < \infty$. Hence, there is a PA A' which simulates A^+. So we have a relation R^+ between Σ'^* and $\Sigma^*\times(E^+)^\infty$ satisfying the properties of definition 2.4.1.

We now define a relation R between Σ'^* and $\Sigma^*\times E^\infty$ as follows: $[x',(x,\underline{e})]\epsilon R$ iff for any appropriate decomposed total environment $\underline{e}^+$ and derived input tape x^+, $[x',(x^+,\underline{e}^+)]\epsilon R^+$. For any $(x,\underline{e})\epsilon\Sigma^*\times E^\infty$ we can find an associated decomposed total environment and derived input tape x^+. Hence, for any $(x,\underline{e})\epsilon\Sigma^*\times E^\infty$ there exists $x'\epsilon\Sigma'^*$ such that $[x',(x,\underline{e})]\epsilon R$. Also, we have for all $\lambda\epsilon[0,1)$ that $x\epsilon T(A,\underline{e},\lambda)$ iff $x\epsilon T(A^+,\underline{e}^+,\lambda)$. Hence, if $[x',(x,\underline{e})]\epsilon R$, then for any $\lambda\epsilon[0,1)$ $x'\epsilon T(A',\lambda)$ iff $x\epsilon T(A,\underline{e},\lambda)$. And, clearly, x' is not admissible by R^+ iff x' is not admissible by R. Thus, we have the desired result that A' simulates A. □

Example 2.6.1. As an example of theorem 2.6.1 let us consider the ADE $A_h = (\Sigma,S,G,\pi_0,F,E_h)$, where π_0 is a constant and

$E_h = \{h, 2h, \ldots\}$ for some constant $h > 0$. Clearly, E_h and the operation of addition form a semi-group. Let G be such that for any $\sigma\epsilon\Sigma$ and $e\epsilon E_h$ $G(\sigma,e) = (G(\sigma,h))^{e/h}$. Thus for any $\sigma\epsilon\Sigma$ and $e_1, e_2 \epsilon E_h$, $G(\sigma,e_1)G(\sigma,e_2) = (G(\sigma,h))^{e_1/h}(G(\sigma,h))^{e_2/h} = (G(\sigma,h))^{(e_1+e_2)/h}$ $= G(\sigma,e_1+e_2)$. Hence, G is a homomorphism for each $\sigma\epsilon\Sigma$. Clearly, $E_h = S(\{h\})$. For any $e\epsilon E_h$, $e = h+\ldots+h$; there are e/h occurrences of h in the decomposition of e. Let $x\epsilon\Sigma^*$ such that $L(x) = r$ and $\underline{e}\epsilon E_h^\infty$. Consequently, $x^+ = \sigma_1^{e_1/h}\ldots\sigma_r^{e_r/h}$ and $\underline{e}^+ = (h,h,\ldots)$. Thus, $M(x,\underline{e}) = \prod_{i=1}^{r} G(\sigma_i,e_i) = \prod_{i=1}^{r}(G(\sigma_i,h))^{e_i/h} = \prod_{i=1}^{r}\prod_{j=1}^{e_i/h} G(\sigma_i,h) = M(x^+,\underline{e}^+)$.

Now define a PA $A' = (S',M',\pi_0',F')$ over Σ' as in theorem 2.4.1 to simulate $A_h^+ = (\Sigma,S,G,\pi_0,F,\{h\})$. Hence, $p(x,\underline{e}) = \pi_0 M(x,\underline{e})\eta^F =$ $\pi_0(\prod_{i=1}^{r}\prod_{j=1}^{e_i/h} G(\sigma_i h))\eta^F = \pi_0' M'(\phi\tau_1^{e_1/h}\ldots\tau_r^{e_r/h})\eta^{F'} = p'(\phi\tau_1^{e_1/h}\ldots\tau_r^{e_r/h})$ where $\tau_1,\ldots,\tau_r\epsilon\Sigma'$ and $g(\sigma_i,h) = \tau_i$. Consequently, $x\epsilon T(A,\underline{e},\lambda)$ iff $x' = \phi\tau^{e_1/h}\ldots\tau_r^{e_r/h}\epsilon T(A',\lambda)$ for all $\lambda\epsilon[0,1)$.

By our definition we may regard probabilistic automata as only being able to change state within a discrete time scale $t = 1,2,\ldots$. The concept of the continuous time probabilistic automata as introduced by Knast [8] is similar to the Markov chain with a continuous parameter. The next state function of such an automaton is defined for $t\epsilon[0,\infty)$ as well as $\sigma\epsilon\Sigma$. A continuous time probabilistic automaton may change state at any time.

Definition 2.6.2. A continuous time probabilistic automaton (CTPA) is a system $A = (\Sigma,S,G,\pi_0,F,T)$, where Σ,S,π_0, and F are as in definition 1.3.2, $T = [0,\infty)$, and $G(\sigma,t)$ is a matrix transition function satisfying:

1. $G(\sigma,t)\epsilon M_n$ for all $\sigma\epsilon\Sigma$ and $t\epsilon T$.
2. $G(\sigma,t_1+t_2) = G(\sigma,t_1)G(\sigma,t_2)$ for any $\sigma\epsilon\Sigma$ and $t_1,t_2\epsilon T$.
3. Every g_{ij} is a measurable function in $(0,\infty)$ and $\lim_{t\to 0} g_{ii}(\sigma,t) = 1$ for all $\sigma\epsilon\Sigma$ and $i = 1,\ldots,n$ and $j = 1,\ldots,n$.

The element $g_{ij}(\sigma,t)$ is the probability of being in state s_j at time $t_0 + t$ when the input is σ given that the system is in state s_i at time t_0 for any $t_0\epsilon T$. Clearly, the transition properties depend on the time interval for input σ.

Definition 2.6.3. An infinitesimal matrix of the semi-group $(\{G(\sigma,t)|t\epsilon T\}$, multiplication) is the limit $A(\sigma) = \lim_{t\to 0}(G(\sigma,t)-I_n)/t$, where I_n denotes the $n\times n$ identity matrix. The elements $a_{ij}(\sigma)$ of $A(\sigma)$ are called the rates of transition from state s_i to state s_j via input symbol σ.

Knast proved that the transition function for every CTPA has the canonical form $G(\sigma,t) = e^{A(\sigma)t} = \sum_{i=0}^{\infty} \frac{(tA(\sigma))^i}{i!}$.

It is readily seen from the definition of the CTPA that it is a special case of the ADE model with a semi-group environment set whose structure is preserved by the homomorphisms G_σ for all $\sigma\epsilon\Sigma$. Since there does not exist a finite set E^+ such that $T=S(E^+)$, theorem 2.5.1 may not be used to show the existence of a PA A' to simulate a CTPA. Knast showed, however, that under certain conditions a CTPA can be approximated by an ADE of the form A_h as above. For further results concerning this special case of the ADE model see Knast [8].

CHAPTER III

Stability Results

3.1 The Stability Problem

The stability problem is a natural consequence of the inevitable impreciseness of the measurement of the environment configuration. We shall consider the effects of small perturbations of the environmental condition on the acceptance sets and transition functions.

Implicit in the notion of perturbation of the environment is consideration of a natural metric on the environment set by which the size of the perturbation is measured. We shall denote the natural metric on the environment set by d^N. For example, let us consider a system whose transitions are related to temperature. That is, suppose $E = [0,100]$; $d^N(e,e') = |e-e'|$, where $e,e' \in E$, may be a useful natural metric to consider perturbations of the environment. The salient point concerning d^N is that it is computed directly from the environments without necessarily considering the transition functions.

In Section 2.5 we considered the function

$$d(e,e') = \max(|\pi_0(e)-\pi_0(e')|, \max_{\sigma\in\Sigma}|G(\sigma,e) - G(\sigma,e')|)$$

based on the ADE $A = (\Sigma,S,G,\pi_0,F,E)$, where $e,e' \in E$. The function d was, in general, found to be a pseudo-metric on the environment set E. We shall now establish the notion of continuity of the transition functions with respect to the natural metric.

Definition 3.1.1. Let $A = (\Sigma,S,G,\pi_0,F,E)$ be an ADE and d^N be a natural metric on the environment set E. The basic matrix transition function G is continuous at e with respect to d^N if for every $\varepsilon > 0$ there exists $\delta(e,\varepsilon) > 0$ such that $\max_{\sigma\in\Sigma}|G(\sigma,e) - G(\sigma,e')| < \varepsilon$ whenever $d^N(e,e') < \delta(e,\varepsilon)$. G is uniformly continuous with respect to d^N if G is continuous at each $e\in E$ and δ can be chosen independent of e. Similarly, π_0 is continuous at e with respect to d^N if for every $\varepsilon > 0$ there exists $\delta(e,\varepsilon) > 0$ such that $|\pi_0(e) - \pi_0(e')| < \varepsilon$ whenever $d^N(e,e') < \delta(e,\varepsilon)$. π_o is uniformly continuous with respect to d^N if π_0 is continuous at each $e\in E$ and δ can be chosen independent of e.

Continuity conditions will allow us to bound the perturbations of the transition probabilities given the perturbation of the environment with respect to the natural metric.

Lemma 3.1.1. Let $A = (\Sigma,S,G,\pi_0,F,E)$ be an ADE with natural metric d^N. If G and π_0 are continuous at e, given any $\varepsilon > 0$, we can find $\delta(e,\varepsilon) > 0$ such that for any $e'\in E$, $d(e,e') < \varepsilon$ whenever $d^N(e,e') < \delta$.

Proof: Clear from the definition of continuity and d.

Thus, if G and π_0 are continuous at e, for small perturbations of e with respect to d^N we can bound the perturbation of G and π_0. Furthermore, by making the perturbation of e sufficiently small we can make the perturbation of G and π_0 as small as desired. Also, if G and π_0 are uniformly continuous, then the choice of δ can be made independently of e.

For notational convenience, let $J = \{0,1,2,\ldots\}$ and $J_+ = \{1,2,\ldots\}$.

Definition 3.1.2. An ADE $A = (\Sigma,S,G,\pi_0,F,E)$ with natural metric d^N on E and cut-point λ is tape-acceptance stable (a-stable) at $\underline{e} = (e_0,e_1,\ldots)$ iff there exists $\delta > 0$ such that $\sup_{i\in J} d^N(e_i,e_i') < \delta$ implies $T(A,\underline{e},\lambda) = T(A,\underline{e}',\lambda)$.

In other words, an ADE with natural metric d^N and with cut-point λ is a-stable if the set of tapes defined by A under total environment $\underline{e}$ with cut-point λ is not changed by sufficiently small perturbations of the environments.

Definition 3.1.3. An ADE $A = (\Sigma,S,G,\pi_0,F,E)$ with natural metric d^N on E is strictly stable (s-stable) at $\underline{e} = (e_0,e_1,\ldots)$ iff given any $\varepsilon > 0$ there exists $\delta > 0$ such that $\sup_{i\in J} d^N(e_i,e_i') < \delta$ implies $|\pi(x,\underline{e}) - \pi(x,\underline{e}')| < \varepsilon \quad \forall x\in\Sigma^*$, where π is the state distribution function.

Theorem 3.1.1. Let $A = (\Sigma,S,G,\pi_0,F,E)$ be an ADE with natural metric d^N on E. If A is s-stable at $\underline{e}$, then given any $\gamma > 0$ there exists $\delta > 0$ such that $\sup_{i\in J} d^N(e_i,e_i') < \delta$ implies $|p(x,\underline{e}) - p(x,\underline{e}')| < \gamma \quad \forall x\in\Sigma^*$, where p is the acceptance probability function.

Proof: $p(x,\underline{e}) = \pi(x,\underline{e})\eta^F$, where η^F is as in definition 1.3.4. For any $x\in\Sigma^*$, $|p(x,\underline{e}) - p(x,\underline{e}')| = |(\pi(x,\underline{e}) - \pi(x,\underline{e}'))\eta^F| \le |\pi(x,\underline{e}) - \pi(x,\underline{e}')|\#(F)$. Given any $\gamma > 0$ let $\varepsilon = \gamma/\#(F)$. Since A is s-stable at $\underline{e}$, there exists $\delta > 0$ such that $\sup_{i\in J} d^N(e_i,e_i') < \delta$

implies $|\pi(x,\underline{e}) - \pi(x,\underline{e}')| < \varepsilon = \gamma/\#(F)$. Thus, $|p(x,\underline{e}) - p(x,\underline{e}')| < \varepsilon\#(F) = \gamma \quad \forall x\in\Sigma^*$. □

Thus, s-stability at $\underline{e}$ guarantees a bound on the perturbation of the acceptance probability function.

Definition 3.1.4. A cut-point λ is called isolated with respect to an ADE A at $\underline{e}$ if there exists $\gamma > 0$ such that $|p(x,\underline{e})-\lambda| \geq \gamma$ $\forall x\in\Sigma^*$.

Theorem 3.1.2. Let $A = (\Sigma,S,G,\pi_0,F,E)$ be an ADE with natural metric d^N on E. If λ is an isolated cut-point for A at $\underline{e}$ and if A is s-stable at $\underline{e}$, then A with cut-point λ is a-stable at $\underline{e}$.

Proof: λ is an isolated cut-point for A at $\underline{e}$. So there exists $\gamma > 0$ such that $|p(x,\underline{e})-\lambda| \geq \gamma$ $\forall x\in\Sigma^*$. Since A is s-stable at $\underline{e}$, choose ε such that $0 < \varepsilon < \gamma$. Then by theorem 3.1.1 there exists $\delta > 0$ such that $\sup_{i\in J} d^N(e_i,e_i') < \delta$ implies $|p(x,\underline{e}) - p(x,\underline{e}')| < \varepsilon < \gamma$ $\forall x\in\Sigma$.

If $x\in T(A,\underline{e},\lambda)$, then $p(x,\underline{e}) > \lambda$. But λ is isolated, so $p(x,\underline{e}) \geq \lambda + \gamma$. Thus, $p(x,\underline{e}') \geq \lambda + \gamma - \varepsilon > \lambda$. Hence $x\in T(A,\underline{e}',\lambda)$.

If $x \notin T(A,\underline{e},\lambda)$, then $p(x,\underline{e}) \leq \lambda$. Again, λ is isolated, so $p(x,\underline{e}) \leq \lambda - \gamma$. Thus, $p(x,\underline{e}') \leq \lambda - \gamma + \varepsilon \leq \lambda$. And, hence, $x \notin T(A,\underline{e},\lambda)$.

Consequently, there exists $\delta > 0$ such that $\sup_{i\in J} d^N(e_i,e_i') < \delta$ implies $T(A,\underline{e},\lambda) = (A,\underline{e}',\lambda)$. That is, A with cut-point λ is a-stable. □

The concept of a-stability differs from s-stability in that a-stability allows a large perturbation of the acceptance probability function as long as the cut-point is not crossed.

3.2 Stability with Tapes of Bounded Length

Definition 3.2.1. If $\xi = (\xi_i)$ is an arbitrary vector and P is an arbitrary matrix, then $||\xi|| = \sum_i |\xi_i|$ and $||P|| = \max_i \sum_j |p_{ij}|$.

Lemma 3.2.1. a) Let $B,B',C,C' \epsilon M_n$, then $||BC-B'C'|| \leq ||B-B'|| + ||C-C'||$. b) Let $v,v' \epsilon V_n$, then $||vB-v'B'|| \leq ||v-v'|| + ||B-B'||$.

Proof: We must first show that $||\cdot||$ satisfies the triangle property on the set of all $n \times n$ matrices.

$$\begin{aligned} ||H-H'|| &= \max_i \sum_j |h_{ij}-h'_{ij}| \\ &= \max_i \sum_j |h_{ij}-h''_{ij}+h''_{ij}-h'_{ij}| \\ &\leq \max_i \sum_j \{|h_{ij}-h''_{ij}| + |h''_{ij}-h_{ij}|\} \\ &\leq \max_i \sum_j |h_{ij}-h''_{ij}| + \max_i \sum_j |h''_{ij}-h'_{ij}| \\ &= ||H-H''|| + ||H''-H'|| \quad \text{for any } n \times n \text{ matrices} \end{aligned}$$

H,H', and H''. Let us now consider $||BC-B'C'|| = ||BC-B'C + B'C - B'C'|| \leq ||(B-B')C|| + ||B'(C-C')||$. But

$$\begin{aligned} ||(B-B')C|| &= \max_i \sum_j | \sum_k (b_{ik}-b'_{ik})c_{kj}| \\ &\leq \max_i \sum_j \sum_k c_{kj}|b_{ik}-b'_{ik}| \\ &= \max_i \sum_k \sum_j c_{kj}|b_{ik}-b'_{ik}| \\ &= \max_i \sum_k |b_{ik}-b'_{ik}| \\ &= ||B-B'||. \end{aligned}$$

Similarly,

$$
\begin{aligned}
||B'(C-C')|| &= \max_i \sum_j \Big|\sum_k b'_{ik}(c_{kj}-c'_{kj})\Big| \\
&\le \max_i \sum_j \sum_k b'_{ik}|c_{kj}-c'_{kj}| \\
&= \max_i \sum_k b'_{ik} \sum_j |c_{kj}-c'_{kj}| \\
&\le \max_i \sum_k b'_{ik}||C-C'|| \\
&= ||C-C'|| \ .
\end{aligned}
$$

Thus, $||BC-B'C'|| \le ||B-B'|| + ||C-C'||$.

For b) $||\cdot||$ similarly satisfies the triangle property on the set of all n-dimensional vectors. Thus,

$$
\begin{aligned}
||vB-v'B'|| &= ||vB-v'B+v'B-v'B'|| \\
&\le ||(v-v')B|| + ||v'(B-B')|| \ .
\end{aligned}
$$

We see that

$$
\begin{aligned}
||(v-v')B|| &= \sum_j \Big|\sum_i (v_i-v'_i)b_{ij}\Big| \\
&\le \sum_j \sum_i b_{ij}|v_i-v'_i| \\
&= \sum_i \sum_j b_{ij}|v_i-v'_i| \\
&= \sum_i |v_i-v'_i| \\
&= ||v-v'|| \ .
\end{aligned}
$$

Also,

$$
\begin{aligned}
||v'(B-B')|| &= \sum_j \Big|\sum_i v'_i(b_{ij}-b'_{ij})\Big| \\
&\le \sum_j \sum_i v'_i|b_{ij}-b'_{ij}|
\end{aligned}
$$

$$= \sum_i v'_i \sum_j |b_{ij} - b'_{ij}|$$

$$\leq \sum_i v_i ||B - B'||$$

$$= ||B - B'|| \ .$$

Hence, $||vB - v'B'|| \leq ||v - v'|| + ||B - B'||$. □

Lemma 3.2.2. For any integer $k > 0$, let $B_1, \ldots, B_k$ and $C_1, \ldots, C_k$ all belong to M_n and let $v, w \in V_n$, then

$$||v \prod_{i=1}^{k} B_i - w \prod_{i=1}^{k} C_i|| \leq ||v - w|| + \sum_{i=1}^{k} ||B_i - C_i||.$$

Proof: By induction on the results of lemma 3.2.1.

Lemma 3.2.3. Let $B, C \in M_n$ and $v, w \in V_n$, then $|B-C| \leq ||B-C|| \leq n|B-C|$ and $|v-w| \leq ||v-w|| \leq n|v-w|$.

Proof: $|B-C| = \max_{i,j} |b_{ij} - c_{ij}|$. Let i^* and j^* be the indices where the maximum is obtained. Thus, $|B-C| = |b_{i^*j^*} - c_{i^*j^*}|$. But $|b_{i^*j^*} - c_{i^*j^*}| \leq \sum_j |b_{i^*j} - c_{i^*j}| \leq \max_i \sum_j |b_{ij} - c_{ij}| = ||B-C||$. And also, $\max_i \sum_j |b_{ij} - c_{ij}| \leq \max_i \sum_j |b_{i^*j^*} - c_{i^*j^*}| = n|B-C|$. And similarly for the result concerning vectors. □

Keeping in mind the above results, let us now consider an ADE $A = (\Sigma, S, G, \pi_0, F, E)$ with natural metric d^N on E. For any fixed K, let $x \in \Sigma^*$ be an arbitrary input tape such that $L(x) = k \leq K$. If x is input under some environment $\underline{e} = (e_0, e_1, \ldots)$, we evaluate $p(x, \underline{e}) = \pi_0(e_0)\left(\prod_{i=1}^{k} G(\sigma_i, e_i)\right)\eta^F$. Let us now consider the acceptance probability when $e_0, \ldots, e_K$ are perturbed with respect to d^N by less than δ. That is, $\max_{i=0,\ldots,K} d^N(e_i, e'_i) < \delta$. If we assume that π_0 is continuous at e_0 and G is continuous at each e_i, $i=1,\ldots,K$, then $|\pi_0(e_0)-\pi_0(e_0')| < \varepsilon_1(e_0)$ and $\max_{\sigma\in\Sigma} |G(\sigma,e_i) - G(\sigma,e_i')| < \varepsilon_2(e_i)$, $i = 1,\ldots,k$. Thus, if we let $\varepsilon^* = \max(\varepsilon_1(e_0), \max_{i=1,\ldots,K} \varepsilon_2(e_i))$, then

$\max_{i=0,\ldots,K} d^N(e_i,e_i') < \delta$ implies $|\pi_0(e_0) - \pi_0(e_0')| < \varepsilon^*$ and

$\max_{\sigma\in\Sigma}|G(\sigma,e_i) - G(\sigma,e_i')| < \varepsilon^*$, $i = 1,\ldots,k$. Using lemma 3.2.3 we see that $\max_{i=0,\ldots,K} d^N(e_i,e_i') < \delta$ implies $||\pi_0(e_0) - \pi_0(e_0')|| < n\varepsilon^*$ and $\max_{\sigma\in\Sigma} ||G(\sigma,e_i) - G(\sigma,e_i')|| < n\varepsilon^*$, $i = 1,\ldots,K$.

By lemma 3.2.2 we have

$$\begin{aligned}
||\pi(x,\underline{e}) - \ (x,\underline{e}')|| &= ||\pi_0(e_0)\prod_{i=1}^{k} G(\sigma_i,e_i) - \pi_0(e_0')\prod_{i=1}^{k} G(\sigma_i,e_i')|| \\
&\le ||\pi_0(e_0) - \pi_0(e_0')|| + \sum_{i=1}^{k} ||G(\sigma_i,e_i) - G(\sigma_i,e_i')|| \\
&\le n\varepsilon^* + kn\varepsilon^* \\
&\le (K+1)n\varepsilon^* \quad \forall x\in\Sigma^* \text{ such that } L(x) \le K
\end{aligned}$$

and $\forall \underline{e}'\in E^\infty$ such that $\max_{i=0,\ldots,K} d^N(e_i,e_i') < \delta$.

Now, consider

$$\begin{aligned}
|p(x,\underline{e}) - p(x,\underline{e}')| &= |(\pi(x,\underline{e}) - \pi(x,\underline{e}'))\eta^F| \\
&\le \sum_i |\pi^{(i)}(x,\underline{e}) - \pi^{(i)}(x,\underline{e}')| \\
&= ||\pi(x,\underline{e}) - \pi(x,\underline{e}')|| \,.
\end{aligned}$$

Thus, $\max_{i=0,\ldots,k} d^N(e_i,e_i') < \delta$ implies $|\pi(x,\underline{e}) - \pi(x,\underline{e}')| \le ||\pi(x,\underline{e}) - \pi(x,\underline{e}')|| \le (K+1)n\varepsilon^*$ and $|p(x,\underline{e}) - p(x,\underline{e}')| \le (K+1)n\varepsilon^*$ for any $x\in\Sigma^*$ such that $L(x) \le k$.

Since K and n are fixed finite numbers, given any $\varepsilon > 0$, we let $0 < \varepsilon^* < \varepsilon/(K+1)n$. By the continuity of π_0 at e_0 and G at each e_i $i = 1,\ldots,K$ there exists $\delta > 0$ such that

$\max_{i=0,\ldots,K} d^N(e_i,e_i') < \delta$ implies $|\pi_0(e_0) - \pi_0(e_0')| < \varepsilon^*$ and

$\max_{\sigma\in\Sigma} |G(\sigma,e_i) - G(\sigma,e_i')| < \varepsilon^*$ for $i = 1,\ldots,K$. In turn, this implies that $|p(x,\underline{e}) - p(x,\underline{e}')| \le (K+1)n\varepsilon^* < \varepsilon$ $\forall x\in\Sigma^*$ such that $L(x) \le K$.

Theorem 3.2.1. Let $A = (\Sigma,S,G,\pi_0,F,E)$ be an ADE with natural metric d^N on E. For any fixed $\underline{e} = (e_0,e_1,\ldots)\in E^\infty$ and finite non-negative integer K, if

1. π_0 is continuous with respect to d^N at e_0;
2. G is continuous with respect to d^N at each e_i $i = 1,\ldots,K$

then given any $\varepsilon > 0$ there exists $\delta > 0$ such that $\max_{i=0,\ldots,K} d^N(e_i,e_i') < \delta$ implies $|p(x,\underline{e}) - p(x,\underline{e}')| < \varepsilon$ for all $x\in\Sigma^*$ such that $L(x) \le K$.

When the continuity conditions are satisfied, we can apply theorem 3.2.1 to say that for any K and any $x\in\Sigma^*$ such that $L(x) \le K$, $p(x,\underline{e}') \to p(x,\underline{e})$ as $\max_{i=0,\ldots,K} d^N(e_i,e_i') \to 0$. Thus, $p(x,\underline{e}')$ can be used to approximate $p(x,\underline{e})$ and we have a bound on the maximum error of the approximation as a function of the largest perturbation of $e_0,e_1,\ldots,e_K$.

Example 3.2.1. Consider the ADE A with $\Sigma = \{0,1,\ldots,m-1\}$, $S = \{s_1,s_2,s_3\}$, and $E = [0,1]$. The basic matrix transition function is

$$G(\sigma,e) = \begin{pmatrix} 1/m & \frac{m-\sigma e-1}{m} & \frac{\sigma e}{m} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

and the initial distribution function is $\pi_0(e) = (e,1-e,0)$. Let $F = \{s_3\}$; hence, $\eta^F = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$. Also let $\underline{e} = (1,1,\ldots)$.

Suppose Euclidean distance is the natural metric on E. Thus, we have $|\pi_0(e') - \pi_0(1)| = |1-e'| = d^N(1,e')$ and $\max_{\sigma\in\Sigma}|G(\sigma,e') - G(\sigma,1)| \le |1-e'| = d^N(1,e')$. So then π_0 and g are continuous at $e = 1$.

Let K be any nonnegative integer and $x\in\Sigma^*$ such that $L(x) \le K$. Given any $\varepsilon > 0$, we can find $\delta > 0$ such that for every $\underline{e}'\in E^\infty$ that satisfies $\max_{i=0,\ldots,K} d^N(1,e_i) = \max_{i=0,\ldots,K} |1-e_i'| < \delta$ we have $|p(x,\underline{e}') - p(x,(1,1,\ldots))| < \varepsilon$. In fact, we can see that for $L(x) = k \le K$, $p(x,\underline{e}') = \pi_0(e_0')\left(\prod_{i=1}^{k} G(\sigma_i e_i')\right)\eta^F = e_0' \sum_{i=1}^{k} \frac{\sigma_i e_i'}{m^i}$.

And similarly, $p(x,(1,1,\ldots)) = \sum_{i=1}^{k} \frac{\sigma_i}{m^i}$. So we choose $\delta < \varepsilon/(K+1)$ and we have $|p(x,\underline{e}') - p(x,(1,1,\ldots))| = \left| e_0 \sum_{i=1}^{k} \frac{\sigma_i e_i'}{m^i} - \sum_{i=1}^{k} \frac{\sigma_i}{m^i} \right| \le \sum_{i=0}^{k} |e_i' - 1|$. This last inequality follows from lemma 3.2.2. And clearly, $\sum_{i=0}^{k} |e_i'-1| \le \delta(K+1) < \varepsilon$.

If we consider A operating under the environment $(1,1,\ldots)$, we see that such a system is a probabilistic automaton $A^1 = (S,M,\pi_0,F)$ over Σ, where S,F, and Σ are in A and $\pi_0 = \pi_0(1)$ and $M(\sigma) = G(\sigma,1)$. So for any $\lambda\in[0,1)$, $\mathcal{T}(A^1,\lambda) = \mathcal{T}(A,(1,1,\ldots),\lambda)$ since $p^1(x) = p(x,(1,1,\ldots))\ \forall x\in\Sigma^*$.

Hence, we can use the probabilistic automaton A^1 to approximate the acceptance probability of any input for A. Moreover, for any nonnegative integer K, any $\lambda \in [0,1)$, and any $x \in \Sigma^*$, $p(x,\underline{e}(\delta,K)) \le p(x,\underline{e}') \le p^1(x)$, where $\max_{i=0,\dots,K} d^N(e_i',1) < \delta$ and $\underline{e}(\delta,K) = (\overbrace{1-\delta,\dots,1-\delta}^{K+1 \text{ times}},0,0,\dots)$. Hence, $T(A,\underline{e}(\delta,K),\lambda) \subset T(A,\underline{e}',\lambda) \subset T(A^1,\lambda)$. But $T(A,\underline{e}(\delta,K),\lambda)$ is strictly increasing as $\delta \to 0$ and bounded above by $T(A^1,\lambda)$. So the limit exists.

$$\lim_{\delta\to 0} T(A,\underline{e}(\delta,K),\lambda) = \{x\in\Sigma^* \mid x\in T(A^1,\lambda) \text{ and } L(x) \le K\} .$$

Also, $\lim_{\delta\to 0} T(A,\underline{e}(\delta,K),\lambda)$ is strictly increasing as $K \to \infty$ and bounded above by $T(A^1,\lambda)$; hence, this limit exists.

$$\lim_{K\to\infty} \lim_{\delta\to 0} T(A,\underline{e}(\delta,K),\lambda)$$

$$= \lim_{K\to\infty} \{x\in\Sigma^* \mid x\in T(A^1,\lambda) \text{ and } L(x) \le K\}$$

$$= T(A^1,\lambda) .$$

Similarly, $\lim_{\delta\to 0} \lim_{K\to\infty} T(A,\underline{e}(\delta,K),\lambda) = T(A^1,\lambda)$.

Now we can say that

$$\lim_{K\to\infty} \lim_{\max_{i=0,\dots,K} d^N(e_i',1)\to 0} T(A,\underline{e}',\lambda) = T(A^1,\lambda)$$

for all $\lambda\in[0,1)$.

3.3 Ergodic Conditions for s-Stability

In the previous section we produced sufficient conditions for a stability result on sets of input tapes of bounded length. We shall now consider the asymptotic properties of long products of stochastic matrices to generate sufficient conditions for s-stability. That is, under such conditions for some fixed $e\epsilon E^{\infty}$, given any $\varepsilon > 0$, we can find $\delta > 0$ such that if the largest perturbation of the environment is less than δ, then the largest perturbation of the state distribution function is less than ε for any input tape.

Definition 3.3.1. For any real number a, let $a^+ = \max(a,0)$ and $a^- = \min(a,0)$. For any $P\epsilon M_n$, we define

$$\partial(P) = \max_{i_1,i_2} \sum_j (p_{i_1 j} - p_{i_2 j})^+ .$$

Lemma 3.3.1. For any $p,q\epsilon V_n$, $\sum_{i=1}^{n}(p_i-q_i)^+ = -\sum_{i=1}^{n}(p_i-q_i)^-$ $= \frac{1}{2}\sum_{i=1}^{n}|p_i-q_i|$.

Proof: Let $N' = \{i|p_i-q_i \geq 0\}$ and $N'' = \{i|p_i-q_i < 0\}$.

$N'\cup N'' = \{1,2,\ldots,n\}$ and $N'\cap N'' = \phi$

$$0 = \sum_{i=1}^{n} p_i - \sum_{i=1}^{n} q_i = \sum_{i=1}^{n}(p_i-q_i) = \sum_{i\epsilon N'}(p_i-q_i) + \sum_{i\epsilon N''}(p_i-q_i)$$

$$= \sum_{i=1}^{n}(p_i-q_i)^+ + \sum_{i=1}^{n}(p_i-q_i)^- .$$

Hence, $\sum_{i=1}^{n}(p_i-q_i)^+ = -\sum_{i=1}^{n}(p_i-q_i)^-$. Also,

$$\frac{1}{2}\sum_{i=1}^{n}|p_i-q_i| = \frac{1}{2}\sum_{i=1}^{n}(p_i-q_i)^+ - \frac{1}{2}\sum_{i=1}^{n}(p_i-q_i)^-$$

$$= \frac{1}{2}\sum_{i=1}^{n}(p_i-q_i)^+ + \frac{1}{2}\sum_{i=1}^{n}(p_i-q_i)^+$$

$$= \sum_{i=1}^{n}(p_i-q_i)^+ = -\sum_{i=1}^{n}(p_i-q_i)^- . \qquad \square$$

Lemma 3.3.2. For any $A,B \in M_n$,

1. $0 \le \partial(A) \le 1$

2. $\partial(AB) \le \partial(A)\partial(B)$.

Proof: Let us consider any fixed indices i_1 and i_2 , then $0 \le \sum_j (a_{i_1 j} - a_{i_2 j})^+ \le \sum_j a_{i_1 j} = 1$. Hence $0 \le \partial(A) = \max_{i_1,i_2} \sum_j (a_{i_1 j} - a_{i_2 j})^+ \le 1$.

Again, let us consider any fixed indices i_1 and i_2. Let $N' = \{j \mid \sum_k (a_{i_1 k} - a_{i_2 k}) b_{kj} \ge 0\}$. Now we have

$$\sum_j \left(\sum_k (a_{i_1 k} - a_{i_2 k}) b_{kj}\right)^+ = \sum_{j \in N'} \sum_k (a_{i_1 k} - a_{i_2 k}) b_{kj}$$

$$= \sum_k (a_{i_1 k} - a_{i_2 k}) \sum_{j \in N'} b_{kj}$$

$$\le \sum_k (a_{i_1 k} - a_{i_2 k})^+ \max_k \sum_{j \in N'} b_{kj}$$

$$+ \sum_k (a_{i_1 k} - a_{i_2 k})^- \min_k \sum_{j \in N'} b_{kj} .$$

By lemma 3.3.1

$$\sum_j \left(\sum_k (a_{i_1 k} - a_{i_2 k}) b_{kj}\right)^+ \le \sum_k (a_{i_1 k} - a_{i_2 k})^+ \left(\max_k \sum_{j \in N'} b_{kj} - \min_k \sum_{j \in N'} b_{kj}\right)$$

Note that the elements of N' depend on i_1 and i_2, but

$$\max_k \sum_{j \in N'} b_{kj} - \min_k \sum_{j \in N'} b_{kj} = \max_{k_1,k_2} \sum_{j \in N'} (b_{k_1 j} - b_{k_2 j})$$

$$\le \max_{k_1,k_2} \sum_{j \in N'} (b_{k_1 j} - b_{k_2 j})^+$$

$$\le \max_{k_1,k_2} \sum_j (b_{k_1 j} - b_{k_2 j})^+ = \partial(B)$$

which is independent of i_1 and i_2. Therefore,

$$\sum_j \Big(\sum_k (a_{i_1k}-a_{i_2k})b_{kj}\Big)^+ \le \sum_k (a_{i_1k}-a_{i_2k})^+ \partial(B) \ .$$

Now,

$$\begin{aligned}\partial(AB) &= \max_{i_1,i_2} \sum_k \Big(\sum_k (a_{i_1k}-a_{i_2k})b_{kj}\Big)^+ \\ &\le \max_{i_1,i_2} \sum_k (a_{i_1k}-a_{i_2k})^+ \partial(B) \\ &= \partial(A)\partial(B) \ . \qquad \square\end{aligned}$$

<u>Lemma 3.3.3</u>. For any $B\in M_n$ and any n-dimensional vector ξ such that $\sum_i \xi_i = 0$, then $||\xi B|| \le ||\xi||\partial(B)$.

<u>Proof</u>: If ξ is the zero vector, then $0 = ||\xi B|| = ||\xi||\partial(B)$ and the result is true.

Let ξ be any nonzero vector such that $\sum_i \xi_i = 0$. Define the vectors ζ^1 and ζ^2 as

$$\zeta_i^1 = \frac{2\xi_i^+}{||\xi||} \text{ and } \zeta_i^2 = \frac{-2\xi_i^-}{||\xi||} \ .$$

Since $\sum_i \xi_i = 0$, we have that $\sum_i \xi_i^+ = -\sum_i \xi_i^- = 1/2 \sum|\xi_i|$. Hence $\zeta^1,\zeta^2 \in V_n$ and $\zeta^1 - \zeta^2 = \dfrac{2\xi}{||\xi||}$.

Let $H\in M_n$ be such that its first row is ζ^1 and all of its other rows are ζ^2. So $2\partial(HB) = 2 \sum_j\big(\sum_k(\zeta_k^1 - \zeta_k^2)b_{kj}\big)^+$. By lemma 3.3.1,

$$2\partial(HB) = \sum_j |\sum_k (\zeta_k^1 - \zeta_k^2) b_{kj}| = 2 \sum_j |\sum_k \frac{\xi_k b_{kj}}{||\xi||}| = \frac{2}{||\xi||} ||\sum_k \xi_k b_{kj}|| = 2 \frac{||\xi B||}{||\xi||} .$$

Thus, $\frac{||\xi B||}{||\xi||} = \partial(HB) \leq \partial(H)\partial(B) \leq \partial(B)$ by lemma 3.3.2. And therefore, the result $||\xi B|| \leq ||\xi||\partial(B)$. □

Lemma 3.3.4. For any $B \in M_n$ and any $n \times n$ matrix A such that all of its rows satisfy the condition for ξ in lemma 3.3.3, then $||AB|| \leq ||A||\partial(B)$.

Proof: Let α^i denote the i-th row of A.

$$||AB|| = \max_i \sum_k |\sum_j a_{ij} b_{jk}| = \max_i \sum_k |\alpha_j^i b_{jk}| = \max_i ||\alpha^i B||$$

$$\leq \partial(B) \max_i ||\alpha^i|| = \partial(B) \max_i \sum_j |\alpha_j^i|$$

$$= \partial(B) \max_i \sum_j |a_{ij}| = ||A||\partial(B) .$$ □

Lemma 3.3.5. Let $A,B \in M_n$, then $||AB-B|| \leq 2\partial(B)$.

Proof: $||AB-B|| = ||(A-I_n)B||$. Each row of $A-I_n$ sums to 0 and $||A-I_n|| \leq 2$. Hence, by lemma 3.3.4 $||AB-B|| \leq 2\partial(B)$. □

Definition 3.3.2. An ADE $A = (\Sigma,S,G,\pi_0,F,E)$ is weakly ergodic (w-ergodic) iff for any $\varepsilon > 0$ there exists $N(\varepsilon)$ such that $\partial(M(x,\underline{e})) < \varepsilon$ $\forall \underline{e} \in E^\infty$ and $\forall x \in \Sigma^*$ such that $L(x) \geq N(\varepsilon)$.

The weakly ergodic condition is a generalization of quasidefiniteness for probabilistic automata. That is, for any two stochastic vectors

π and ρ, we have $|(\pi-\rho)M(x,\underline{e})| \le ||(\pi-\rho)M(x,\underline{e})|| \le 2\partial(M(x,\underline{e}))$ $\le 2\varepsilon$ $\forall x\in\Sigma^*$ such that $L(x) \ge N(\varepsilon)$. If we have an ADE that is weakly ergodic, then $\partial(M(x,\underline{e})) \to 0$ as $L(x)\to\infty$ uniformly with respect to $\underline{e}$.

Definition 3.3.3. Let $\underline{e} = (e_0,e_1,\ldots)\in E^\infty$. The total environment $\underline{e}_k = (e_k,e_{k+1},\ldots)\in E^\infty$ is called the k-tail of $\underline{e}$. $E^*(\underline{e}) = \{\underline{e}_k | k\in J\}$ is called the set of all tails of $\underline{e}$.

Theorem 3.3.1. An ADE A is w-ergodic iff there exists $k\in J$ and $0 < \delta < 1$ such that $\partial(M(x,\underline{e})) < \delta$ $\forall \underline{e}\in E^\infty$ and $\forall x\in\Sigma^*$ such that $L(x) = k$.

Proof: If A is w-ergodic, then for any $\varepsilon > 0$, there exists $N(\varepsilon)$ such that $\partial(M(x,\underline{e})) < \varepsilon$ $\forall \underline{e}\in E^\infty$ and $\forall x\in\Sigma^*$ such that $L(x) \ge N(\varepsilon)$. Choose any $0 < \varepsilon < 1$, then let $\delta = \varepsilon$ and $k = N(\varepsilon)$.

For the converse, suppose there exist $k\in J$ and $0 < \delta < 1$ such that $\partial(M(x,\underline{e})) < \delta$ $\forall \underline{e}\in E^\infty$ and $\forall x\in\Sigma^*$ such that $L(x) = k$. For any $\varepsilon > 0$ we can find n_0 such that $\delta^{n_0} < \varepsilon$. Let x be any tape such that $L(x) \ge kn_0$. Then x has the form $x = y_1,\ldots y_{n_0} y$, where $L(y_i) = k$ and $L(y) \ge 0$. Thus, for any fixed, but arbitrary $\underline{e}\in E^\infty$, $M(x,\underline{e}) = (\prod_{i=1}^{n_0} M(y_i,\underline{e}_{k(i-1)}))M(y,\underline{e}_{n_0k})$. And by lemma 3.3.2,

$$\partial(M(x,\underline{e})) \le (\prod_{i=1}^{n_0} \partial(M(y_i,\underline{e}_{k(i-1)})))\partial(M(y,\underline{e}_{n_0k})) \le \delta^{n_0}\partial(M(y,\underline{e}_{n_0k})) .$$

Also by lemma 3.3.2, $0 \le \partial(M(y,\underline{e}_{n_0k})) \le 1$, so $\partial(M(x,\underline{e})) \le \delta^{n_0} < \varepsilon$. Thus, given any $\varepsilon > 0$, choose $N(\varepsilon) = kn_0$. □

Suppose we have an ADE A such that for each $\underline{e}\in E^{\infty}$, given any $\varepsilon > 0$, we can find $N(\varepsilon,\underline{e})$ such that $\partial(M(x,\underline{e})) < \varepsilon \quad \forall x\in\Sigma^*$ such that $L(x) \geq N(\varepsilon,\underline{e})$. We present the following example to demonstrate that such an ADE need not be w-ergodic.

Example 3.3.1. Let A with $\Sigma = \{\sigma\}$, $E = (0,1)$, and $G(\sigma,e) = \begin{pmatrix} 1-e & e \\ e & 1-e \end{pmatrix}$. The matrix function $G(\sigma,e)$ can be written as $G(\sigma,e) = U_1 + (1-2e)U_2$, where $U_1 = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}$ and $U_2 = \begin{pmatrix} 1/2 & -1/2 \\ -1/2 & 1/2 \end{pmatrix}$.

But $U_1^2 = U_1$, $U_2^2 = U_2$, $U_1U_2 = U_2U_1 = 0$. So by induction we have that

$$\prod_{i=1}^{k} G(\sigma,e_i) = U_1 + U_2\Big(\prod_{i=1}^{k}(1 - 2e_i)\Big).$$

Thus

$$\partial\Big(\prod_{i=1}^{k} G(\sigma,e_i)\Big) = \partial\Big(U_1 + U_2\Big(\prod_{i=1}^{k}(1 - 2e_i)\Big)\Big) = \prod_{i=1}^{k} |1 - 2e_i| .$$

Now, for any $\underline{e} = (e_0,e_1,\ldots)\in E^{\infty}$ and any $x\in\Sigma^*$ such that $L(x) = k$, $M(x,\underline{e}) = \prod_{i=1}^{k} G(\sigma,e_i)$. Hence, $\partial(M(x,\underline{e})) = \prod_{i=1}^{k} |1 - 2e_i|$, where $L(x) = k$. But $|1-2e| < 1 \quad \forall \underline{e}\in E$. So given any $\underline{e}\in E^{\infty}$ and any $\varepsilon > 0$, we can find $N(\varepsilon)$ such that $\prod_{i=1}^{N(\varepsilon)} |1-2e_i| < \varepsilon$. Thus, for this $\underline{e}\in E^{\infty}$, $\partial(M(x,\underline{e})) = \prod_{i=1}^{k} |1-2e_i| \leq \prod_{i=1}^{N(\varepsilon)} |1-2e_i| < \varepsilon \quad \forall x\in\Sigma^*$ such that $L(x) = k \geq N(\varepsilon)$. However, this choice of $N(\varepsilon)$ depends on $\underline{e}$. And for any $N(\varepsilon)$, we can find $\underline{e}'\in E^{\infty}$ such that $\prod_{i=1}^{N(\varepsilon)} |1-2e_i'|$ is arbitrarily close to 1. So for some $0 < \varepsilon < 1$ and any $N(\varepsilon)$, there is $\underline{e}'\in E^{\infty}$ and $x\in\Sigma^*$ such that $L(x) \geq N(\varepsilon)$ whereby $\partial(M(x,\underline{e}')) > \varepsilon$. Hence, A is not w-ergodic.

Definition 3.3.4. An ADE $A = (\Sigma, S, G, \pi_0 F, E)$ is ever-ergodic at $\underline{e} \epsilon E^\infty$ iff for any $\varepsilon > 0$ there exists $N(\varepsilon)$ such that $\partial(M(x,\underline{e}_k)) < \varepsilon \;\; \forall \underline{e}_k \epsilon E^*(\underline{e})$ and $\forall x \epsilon \Sigma^*$ such that $L(x) \geq N(\varepsilon)$.

Theorem 3.3.2. If A is an ADE that is w-ergodic, then for each $\underline{e} \epsilon E^\infty$, A is ever-ergodic at $\underline{e}$.

Proof: If A is w-ergodic, then given any $\varepsilon > 0$, there exists $N(\varepsilon)$ such that $\partial(M(x,\underline{e})) < \varepsilon \;\; \forall \underline{e} \epsilon E^\infty$ and $\forall x \epsilon \Sigma^*$ such that $L(x) \geq N(\varepsilon)$. But for any $\underline{e} \epsilon E^\infty$, we have $E^*(\underline{e}) \subset E^\infty$. Then, clearly, given any $\varepsilon > 0$, choose $N(\varepsilon)$ as above; and hence, $\partial(M(x,\underline{e}_k)) < \varepsilon \;\; \forall \underline{e}_k \epsilon E^*(\underline{e})$ and $\forall x \epsilon \Sigma^*$ such that $L(x) \geq N(\varepsilon)$. □

The w-ergodic property is global in E^∞ while the ever-ergodic property is a pointwise property. If the choice of $N(\varepsilon)$ is not dependent on $\underline{e}$ and if A is ever-ergodic at each point of E^∞, then A is w-ergodic.

Lemma 3.3.6. If A is an ADE that is ever-ergodic at $\underline{e}$, then A is ever-ergodic at $\underline{e}_k$ $\forall \underline{e}_k \epsilon E^*(\underline{e})$. Further, if there exists $k \epsilon J$ such that A is ever-ergodic at $\underline{e}_k$, then A is ever-ergodic at $\underline{e}$.

Proof: Let A be ever-ergodic at $\underline{e}$. For all $k \epsilon J$, $E^*(\underline{e}_k) \subset E^*(\underline{e})$. Clearly, this implies A is ever-ergodic at $\underline{e}_k$ $\forall k \epsilon J$. Also, given any $\varepsilon > 0$, the choice $N(\varepsilon)$ need not depend on k.

Suppose there exists $k \epsilon J$ such that A is ever-ergodic at $\underline{e}_k$. Therefore, given any $\varepsilon > 0$, there exists $N_k(\varepsilon)$ such that $\partial(M(\underline{e}_{k+j})) < \varepsilon \;\; \forall j \epsilon J$ and $\forall x \epsilon \Sigma^*$ such that $L(x) \geq N_k(\varepsilon)$. For this ε

to prove ever-ergodicity at $\underline{e}$ we simply choose $N(\varepsilon) = N_k(\varepsilon) + k$. Let $x \in \Sigma^*$ such that $L(x) \geq N(\varepsilon) = N_k(\varepsilon) + k$. So $x = x_k y$, where $L(x_k) = k$ and $L(y) \geq N_k(\varepsilon)$. Thus, $M(x,\underline{e}) = M(x_k,\underline{e})M(y,\underline{e}_k)$. Now by lemma 3.3.2

$$\begin{aligned}\partial(M(x,\underline{e})) &\leq \partial(M(x_k,\underline{e}))\partial(M(y,\underline{e}_k)) \\ &\leq \partial(M(y,\underline{e}_k)) < \varepsilon .\end{aligned}$$ □

Lemma 3.3.7. An ADE A is ever-ergodic at $\underline{e}$ iff there exists $0 < \delta < 1$ and an increasing sequence of nonnegative integers $(k_0,k_1,\ldots)$, where $k_0 = 0$ and $k_{i+1} - k_i \leq K$ for some K, such that $\partial(M(x_i,\underline{e}_{k_i})) \leq \delta$ $\forall x_i \in \Sigma^*$ such that $L(x_i) = k_{i+1} - k_i$.

Proof: If A is ever-ergodic at $\underline{e}$, then given any $0 < \varepsilon < 1$, there exists $N(\varepsilon)$ such that $\partial(M(x,\underline{e}_j)) < \varepsilon$ $\forall \underline{e}_j \in E^*(\underline{e})$ and $\forall x \in \Sigma^*$ such that $L(x) \geq N(\varepsilon)$. Choose $\delta = \varepsilon$, $K = N(\varepsilon)$, and $k_i = iK$.

Suppose now we have $0 < \delta < 1$ and a sequence $(k_0,k_1,\ldots)$ as above. Therefore given any $\varepsilon > 0$, there exists n_0 such that $\delta^{n_0} < \varepsilon$. We wish to examine the behavior of M at an arbitrary tail of $\underline{e}$, $\underline{e}_j$, where j is any fixed, but arbitrary positive integer. We can find $j^* \in J$ such that $k_{j^*} \geq j$ and $k_{j^*-1} \leq j$. Let $x \in \Sigma^*$ such that $L(x) \geq (n_0+1)K$. Thus, x has the form $x = y_1 x_1 \ldots x_{n_0} y_2$, where $L(x_i) = k_{j^*+i} - k_{j^*+i-1} \leq K$, $L(y_1) = k_{j^*} - j \leq K$, and $L(y_2) \geq 0$. Hence,

$$M(x,\underline{e}_j) = M(y_1,\underline{e}_j)\left(\prod_{i=j^*}^{j^*+n_0} M(x_{i-j^*+1},\underline{e}_{k_i})\right)M(y_2,\underline{e}_{k_{j^*+n_0+1}}) .$$

So by lemma 3.3.2 we deduce

$$\partial(M(x,\underline{e}_j) \le \partial(M(y_1,\underline{e}_j))\left[\prod_{i=j^*}^{j^*+n_0} \partial(M(x_{i-j^*+1},\underline{e}_{k_i}))\right]\partial(M(y_2,\underline{e}_{k_{j^*+n_0+1}}))$$

$$\le \delta^{n_0} < \varepsilon .$$

Since j is an arbitrary positive integer, given any $\varepsilon > 0$, we can find $N(\varepsilon) = (n_0+1)K$ and n_0 such that $\partial(M(x,\underline{e}_j)) < \varepsilon \;\; \forall \underline{e}_j$, $j=1,2,\ldots$ and $\forall x\in\Sigma^*$ such that $L(x) \ge N(\varepsilon)$. That is, A is ever-ergodic at $\underline{e}_1$. Hence, by lemma 3.3.6 A is ever-ergodic at $\underline{e}$. □

Definition 3.3.5. For some $\delta > 0$, a matrix $A\in M_n$ is δ-scrambling iff for any pair of rows i_1 and i_2, there is a column j such that $a_{i_1j} > \delta$ and $a_{i_2j} > \delta$.

Definition 3.3.6. An ADE $A = (\Sigma,S,G,\pi_0,F,E)$ is said to satisfy the ever-$H_4(\delta)$ condition at $\underline{e}$ iff there exists $N(\delta)$ such that $\forall x\in\Sigma^*$, where $L(x) \ge N(\delta)$, $M(x,\underline{e}_k)$ is δ-scrambling $\forall \underline{e}_k\in E^*(\underline{e})$.

These definitions are extensions of the concepts concerning probabilistic automata. Note that if $A\in M_n$ is δ-scrambling, then $\partial(A) < 1-\delta$. We now extend theorem 1.3.6.

Theorem 3.3.3. A is ever-ergodic at $\underline{e}$ iff for some $\delta > 0$, A satisfies the ever-$H_4(\delta)$ condition at $\underline{e}$.

Proof: If A is ever-ergodic at $\underline{e}$, then given any $\varepsilon > 0$, there exists $N(\varepsilon)$ such that $\partial(M(x,\underline{e}_k)) < \varepsilon \;\; \forall \underline{e}_k\in E^*(\underline{e})$ and $\forall x\in\Sigma^*$ such that $L(x) \ge N(\varepsilon)$. For some $k\in J$, let i_1 be the index of an arbitrary row $M(x,\underline{e}_k)$. Since $M(x,\underline{e}_k)\in M_n$, then there exists

$s_j \epsilon S$ such that $m_{i_1 j}(x,\underline{e}_k) \geq 1/n$. But $\partial(M(x,\underline{e}_k)) < \varepsilon$, so for any $i_2 \neq i_1$, $|m_{i_1 j}(x,\underline{e}_k) - m_{i_2 j}(x,\underline{e}_k)| < \varepsilon$. Hence, if we choose $\varepsilon = 1/2n$, then $m_{ij}(x,\underline{e}_k) > 1/2n$ for any i. So, choose $\delta = 1/2\ n$ and $N(1/2\,n) = N(1/2n\ )$. The choice of k is arbitrary. Consequently, $M(x,\underline{e}_k)$ is $1/2n$-scrambling $\forall \underline{e}_k \epsilon E^*(\underline{e})$ and $\forall x \epsilon \Sigma^*$ such that $L(x) \geq N(1/2n)$.

Suppose now that we have $\delta > 0$ such that A satisfies the ever-$H_4(\delta)$ condition at $\underline{e}$. Thus, there exists $N(\delta)$ such that $\forall x \epsilon \Sigma^*$ such that $L(x) \geq N(\delta)$ implies $M(x,\underline{e}_k)$ is δ-scrambling $\forall \underline{e}_k \epsilon E^*(\underline{e})$. Hence, $\partial(M(x,\underline{e}_k)) \leq 1 - \delta < 1$ $\forall \underline{e}_k \epsilon E^*(\underline{e})$ and $\forall x \epsilon \Sigma^*$ such that $L(x) \geq N(\delta)$. Let $k_i = iN(\delta)$ and $0 < \delta^* = 1-\delta < 1$. By lemma 3.3.7 A is ever-ergodic at $\underline{e}$. □

We shall now see that under a certain continuity condition, if A is ever-ergodic at $\underline{e}$, then A is ever-ergodic at each $\underline{e}'$ in some neighborhood of $\underline{e}$ with respect to the pseudo-metric $\sup_{i \epsilon J_+} d^N(e_i, e_i')$ on E^∞.

<u>Theorem 3.3.4</u>. If A is ever-ergodic at $\underline{e} \epsilon E^\infty$, then there exists $\varepsilon > 0$ such that for any $\underline{e}' \epsilon E^\infty$, A is ever-ergodic at $\underline{e}'$ if $\max_{\sigma \epsilon \Sigma} ||G(\sigma,e_i) - G(\sigma,e_i')|| < \varepsilon$ for all $i \epsilon J_+$.

<u>Proof</u>: Let $k \epsilon J$ be fixed, but arbitrary. Let $M^{i_0}(x,\underline{e}_k)$ denote the matrix in M_n all of whose rows are equal to the i_0-th row of $M(x,\underline{e}_k)$. Then by lemma 3.3.1 $||M(x,\underline{e}_k) - M^{i_0}(x,\underline{e}_k)|| =$

$$\max_i \sum_j |m_{ij} - m_{i_0 j}| = 2 \max_i \sum_j (m_{ij} - m_{i_0 j})^+ \text{, where } M(x,\underline{e}_k) = (m_{ij}).$$

Hence,

$$||M(x,\underline{e}_k)-M^{i_0}(x,\underline{e}_k)|| \le 2 \max_{i_1,i_2} \sum_j (m_{i_1j}-m_{i_2j})^+ = 2\partial(M(x,\underline{e}_k)) .$$

Since A is ever-ergodic at $\underline{e}$, for any $0 < \delta < 1$ there exists n_0 independent of k such that $\partial(M(x,\underline{e}_k)) < \delta/6$, $\forall \underline{e}_k \in E^*(\underline{e})$ and $\forall x \in \Sigma^*$ satisfying $L(x) \ge n_0$. Thus, by the above proved inequality, we have $||M(x,\underline{e}_k) - M^{i_0}(x,\underline{e}_k)|| < \delta/3$ $\forall s_{i_0} \in S$ and $\forall x \in \Sigma^*$ such that $L(x) \ge n_0$.

Let $\dot{x}$ be a fixed, but arbitrary tape such that $L(\dot{x}) = n_0$. Hence, $||M(\dot{x},\underline{e}_k) - M^{i_0}(\dot{x},\underline{e}_k)|| < \delta/3$ $\forall s_{i_0} \in S$.

Now let $0 < \varepsilon < \delta/3n_0$. Hence, for any $s_{i_0} \in S$

$$||M^{i_0}(\dot{x},\underline{e}_k) - M^{i_0}(\dot{x},\underline{e}'_k)|| \le ||M(\dot{x},\underline{e}_k) - M(\dot{x},\underline{e}'_k)||$$

$$= || \prod_{i=1}^{n_0} G(\sigma_i,e_{k+i}) - \prod_{i=1}^{n_0} G(\sigma_i,e'_{k+i})||$$

$$\le \sum_{i=1}^{n_0} ||G(\sigma_i,e_{k+i}) - G(\sigma_i,e'_{k+i})||$$

by lemma 3.2.2

$$< \sum_{i=1}^{n_0} \delta/3n_0 = \delta/3 .$$

For any $s_{i_0} \in S$,

$$\partial(M(\dot{x},\underline{e}'_k)) = \max_{i_1,i_2} \sum_j (m'_{i_1j}-m'_{i_2j})^+ = 1/2 \max_{i_1,i_2} \sum_j |m'_{i_1j}-m'_{i_2j}|$$

$$\le 1/2 \max_{i_1,i_2} \sum_j [|m'_{i_1j}-m'_{i_0j}| + |m'_{i_2j}-m'_{i_0j}|]$$

$$\leq 1/2 \max_{i_1} \sum_j |m'_{i_1 j} - m'_{i_0 j}| + 1/2 \max_{i_2} \sum |m'_{i_2 j} - m'_{i_0 j}|$$

$$= 1/2 ||M(\dot{x},\underline{e}'_k) - M^{i_0}(\dot{x},\underline{e}'_k)|| + 1/2 ||M(\dot{x},\underline{e}'_k) - M^{i_0}(\dot{x},\underline{e}_k)||$$

$$= \quad ||M(\dot{x},\underline{e}'_k) - M^{i_0}(\dot{x},\underline{e}'_k)|| \ .$$

Thus, for any $s_{i_0} \in S$

$$\partial(M(\dot{x},\underline{e}_k)) \leq ||M(\dot{x},\underline{e}'_k) - M^{i_0}(\dot{x},\underline{e}'_k))||$$

$$\leq ||M(\dot{x},\underline{e}'_k) - M(\dot{x},\underline{e}_k)||$$

$$+ ||M(\dot{x},\underline{e}_k) - M^{i_0}(\dot{x},\underline{e}_k)||$$

$$+ ||M^{i_0}(\dot{x},\underline{e}_k) - M^{i_0}(\dot{x},\underline{e}'_k)||$$

$$< \delta/3 + \delta/3 + \delta/3 = \delta \ .$$

But $\dot{x}$ and k are arbitrary. Therefore, $\partial(M(x,\underline{e}'_k)) < \delta < 1$ $\forall \underline{e}_k \in E^*(e)$ and $\forall x \in \Sigma^*$ such that $L(x) = n_0$. Let $k_i = in_0$. Hence, by lemma 3.3.7, A is ever-ergodic at $\underline{e}'$. □

<u>Definition 3.3.7</u>. Let $A = (\Sigma,S,G,\pi_0,F,E)$ be an ADE with natural metric d^N and $E(\underline{e}) = \{e_j | j \in J_+\}$. G is uniformly continuous on $E(\underline{e})$ iff given any $\varepsilon > 0$, there exists $\delta > 0$ such that

$\max_{\sigma \in \Sigma} |G(\sigma,e) - G(\sigma,e')| < \varepsilon$ whenever $d^N(e,e') < \delta$ for any $e \in E(\underline{e})$.

The choice of δ is independent of $e \in E(\underline{e})$.

<u>Corollary 3.3.5</u>. If A is ever-ergodic at $\underline{e} \in E^\infty$ and G is uniformly continuous on $E(\underline{e})$, then there exists $\delta > 0$ such that for

any $\underline{e}'\in E^{\infty}$, where $\sup_{i\in J_+} d^N(e_i,e_i') < \delta$, A is ever-ergodic at $\underline{e}'$.

Proof: By theorem 3.3.4 there exists $\varepsilon^* > 0$ such that if

$\max_{\sigma\in\Sigma} ||G(\sigma,e_i) - G(\sigma,e_i')|| < \varepsilon^*$ $\forall i\in J_+$ and A is ever-ergodic at $\underline{e}$, then A is ever-ergodic at $\underline{e}'$.

Uniform continuity of G on $E(\underline{e})$ implies that given any $\varepsilon > 0$, there exists $\delta > 0$ such that when $\sup_{i\in J_+} d^N(e_i,e_i') < \delta$, we have

$\max_{\sigma\in\Sigma} |G(\sigma,e_i) - G(\sigma,e_i')| < \varepsilon/n$ $\forall i\in J_+$. Hence, by lemma 3.2.3

$$\max_{\sigma\in\Sigma} ||G(\sigma,e_i) - G(\sigma,e_i')|| \le n \max_{\sigma\in\Sigma} |G(\sigma,e_i) - G(\sigma,e_i')| < n(\varepsilon/n) = \varepsilon \quad \forall i\in J_+.$$

Thus, if we choose $0 < \varepsilon < \varepsilon^*$, we can find $\delta > 0$ such that

$\sup_{i\in J_+} d^N(e_i,e_i') < \delta$ implies $\max_{\sigma\in\Sigma} ||G(\sigma,e_i) - G(\sigma,e_i')|| < \varepsilon^*$ $\forall i\in J_+$.

By theorem 3.3.4 A is ever-ergodic at $\underline{e}'$.

Theorem 3.3.6. Let $A = (\Sigma,S,G,\pi_0,F,E)$ be an ADE that is ever-ergodic at $\underline{e}$. For any $\varepsilon > 0$, there exists $\varepsilon^* > 0$ such that $||M(x,\underline{e}) - M(x,\underline{e}')|| < \varepsilon$ $\forall x\in\Sigma^*$ whenever $\max_{\sigma\in\Sigma} ||G(\sigma,e_i) - G(\sigma,e_i')|| < \varepsilon^*$ $\forall i\in J_+$.

Proof: By theorem 3.3.4 there exists $\varepsilon_1 > 0$ such that A is ever-ergodic at $\underline{e}'$ whenever $\max_{\sigma\in\Sigma} ||G(\sigma,e_i) - G(\sigma,e_i)|| < \varepsilon_1$, $\forall i\in J_+$. Hence, A is ever-ergodic at both $\underline{e}$ and $\underline{e}'$. So, given any $\varepsilon > 0$, we can find n_0 such that $\partial(M(x,\underline{e}_k)) < \varepsilon/6$ and $\partial(M(x,\underline{e}_k')) < \varepsilon/6$ $\forall k\in J$ and $\forall x\in\Sigma^*$ such that $L(x) \ge n_0$.

Choose $0 < \varepsilon_2 < \varepsilon/3n_0$. By lemma 3.2.2 we see that for any $k \in J$ $||M(x,\underline{e}_k) - M(x,\underline{e}_k')|| < \varepsilon/3$ $\forall x \in \Sigma^*$ such that $L(x) \le n_0$ whenever $\max_{\sigma\in\Sigma}||G(\sigma,e_i) - G(\sigma,e_i')|| < \varepsilon_2$ $\forall i \in J_+$.

Now let $\varepsilon^* = \min(\varepsilon_1,\varepsilon_2)$. The above results remain true when $\max_{\sigma\in\Sigma}||G(\sigma,e_i) - G(\sigma,e_i')|| < \varepsilon^*$ $\forall i \in J_+$.

Since $\underline{e}_0 = \underline{e}$ and $\underline{e}_0' = \underline{e}'$, we have $||M(x,\underline{e}) - M(x,\underline{e}')|| < \varepsilon/3 < \varepsilon$ $\forall x \in \Sigma^*$ such that $L(x) \le n_0$ whenever $\max_{\sigma\in\Sigma}||G(\sigma,e_i) - G(\sigma,e_i')|| < \varepsilon^*$ $\forall i \in J_+$. Thus, we have the desired result $\forall x \in \Sigma^*$ such that $L(x) < n_0$.

Now let $x \in \Sigma^*$ such that $L(x) \ge n_0$. We can represent $x = \dot{x}y$, where $L(y) = n_0$ and $L(\dot{x}) = k \ge 0$ for some $k \in J$. Hence

$$
\begin{aligned}
||M(x,\underline{e}) - M(x,\underline{e}')|| &= ||M(\dot{x},\underline{e})M(y,\underline{e}_k) - M(\dot{x},\underline{e}')M(y,\underline{e}'_k)|| \\
&\le ||M(\dot{x},\underline{e})M(y,\underline{e}_k) - M(y,\underline{e}_k)|| \\
&\quad +||M(\dot{x},\underline{e}')M(y,\underline{e}_k') - M(y,\underline{e}_k')|| \\
&\quad +||M(y,\underline{e}_k) - M(y,\underline{e}_k')|| \\
&\le 2\partial(M(y,\underline{e}_k)) + 2\partial(M(y,\underline{e}_k')) \\
&\quad +||M(y,\underline{e}_k) - M(y,\underline{e}_k')|| \quad \text{by lemma 3.3.5.}
\end{aligned}
$$

Since $L(y) = n_0$, then $\partial(M(y,\underline{e}_k)) < \varepsilon/6$ and $\partial(M(y,\underline{e}_k')) < \varepsilon/6$. Also $||M(y,\underline{e}_k) - M(y,\underline{e}_k')|| < \varepsilon/3$. Thus

$$||M(x,\underline{e}) - M(x,\underline{e}')|| < 2(\varepsilon/6) + 2(\varepsilon/6) + \varepsilon/3 = \varepsilon$$

$\forall x \in \Sigma^*$ such that $L(x) \ge n_0$ whenever $\max_{\sigma\in\Sigma}||G(\sigma,e_i) - G(\sigma,e_i')|| < \varepsilon$ $\forall i \in J_+$.

Hence, the result is true $\forall x \in \Sigma^*$. □

Theorem 3.3.6 is a useful stability result. Moreover, under certain continuity conditions we have s-stability.

Corollary 3.3.7. Let $A = (\Sigma, S, G, \pi_0, F, E)$ be an ADE with natural metric d^N on E. If for some $\underline{e} \in E^\infty$,

1. A is ever-ergodic at $\underline{e}$
2. G is uniformly continuous on $E(\underline{e}) = \{e_i | i \in J_+\}$
3. π_0 is continuous at e_0 ,

then A is s-stable at $\underline{e}$.

Proof: By theorem 3.3.6, given any $\varepsilon > 0$, there exists $\varepsilon^* > 0$ such that $||M(x,\underline{e}) - M(x,\underline{e}')|| < \varepsilon/2 \quad \forall x \in \Sigma^*$ whenever $\max_{\sigma \in \Sigma} ||G(\sigma, e_i) - G(\sigma, e_i')|| < \varepsilon^* \quad \forall i \in J_+$. Since G is uniformly continuous on $E(\underline{e})$, then there exists $\delta_1 > 0$ such that $\max_{\sigma \in \Sigma} ||G(\sigma, e_i) - G(\sigma, e_i)|| < \varepsilon^*$ whenever $\sup_{i \in J_+} d^N(e_i, e_i') < \delta_1$.

The initial distribution function π_0 is continuous at e_0, so there exists $\delta_2 > 0$ such that $|\pi_0(e_0) - \pi_0(e_0')| < \varepsilon/2n$ when $d^N(e_0, e_0') < \delta_2$. Hence by lemma 3.2.3 $||\pi_0(e_0) - \pi_0(e_0')|| < \varepsilon/2$ when $d^N(e_0, e_0') < \delta_2$.

Choose $\delta = \min(\delta_1, \delta_2)$. Hence, whenever $\sup_{i \in J} d^N(e_i, e_i') < \delta$,

$$|\pi(x,\underline{e}) - \pi(x,\underline{e}')| \le ||\pi(x,\underline{e}) - \pi(x,\underline{e}')||$$

by lemma 3.2.3 $$= ||\pi_0(e_0)M(x,\underline{e}) - \pi_0(e_0')M(x,\underline{e})||$$

by lemma 3.2.1 $$\le ||\pi_0(e_0) - \pi_0(e_0')|| + ||M(x,\underline{e}) - M(x,\underline{e}')||$$

$$< \varepsilon/2 + \varepsilon/2 = \varepsilon \qquad \forall x \in \Sigma^* .$$

Hence, A is s-stable at $\underline{e}$. □

Example 3.3.1. Let $A = (\Sigma, S, G, \pi_0, F, E)$ be an ADE with $\Sigma = \{\sigma\}$, $S = \{s_1, s_2, s_3\}$ and $E = [-1/2, 1/2]$. Let Euclidean distance be the natural metric on E; that is, for any $e, e' \in E$, $d^N(e,e') = |e-e'|$. The function G is defined

$$G(\sigma,e) = \begin{pmatrix} 1/3+e^5 & 0 & 2/3-e^5 \\ e^4/2 & e^4/2 & 1-e^4 \\ (3+e^5)/5 & (2-e^5)/5 & 0 \end{pmatrix} .$$

We define $\pi_0(e) = (1/6+e^4,\ 1/3+e^4,\ 1/2-2e^4)$.

Let $\underline{e} \in E^\infty$ such that $\liminf_{i\to\infty} |e_i| \geq \gamma$ for some $\gamma > 0$. Hence, there exists N such that $|e_i| \geq \gamma/2 > 0$ $\forall i \geq N$. But $G(\sigma,e)$ is $\gamma^4/32$-scrambling $\forall e \in E$ such that $|e| \geq \gamma/2$. Hence, for $N(\gamma^4/32) = 1$, A satisfies the ever-$H_4(\gamma^4/32)$ condition at $\underline{e}_N$. By theorem 3.3.3 A is ever-ergodic at $\underline{e}$.

The components of G are continuous at each $e \in E$. But E is a compact set with respect to the natural metric topology. Hence, G is uniformly continuous. Clearly, G is uniformly continuous on $E(\underline{e})$ for any $\underline{e} \in E^\infty$. The function π_0 is also continuous at any $e \in E$.

Hence, given any $\varepsilon > 0$, we can find $\delta > 0$ such that $|\pi(x,\underline{e}) - \pi(x,\underline{e}')| < \varepsilon$ $\forall x \in \Sigma^*$ whenever $\sup_{i\in J} d^N(e_i, e_i') = \sup_{i \in J} |e_i - e_i'| < \delta$.

Let $A = (\Sigma, S, G, \pi_0, F, E)$ be an ADE with natural metric d^N on E. Let E be a compact set with respect to the natural metric topology. If the components of $G(\sigma,e)$ are continuous at each $e \in E$, then G is uniformly continuous with respect to d^N. Moreover, if there exists $\delta > 0$ such that all of the components of $G(\sigma,e)$ are bounded below by

γ, then A is w-ergodic and, hence, ever-ergodic at each $\underline{e}\in E^\infty$. An ADE satisfying the above conditions is s-stable at every $\underline{e}\in E^\infty$. Furthermore, the choice of δ need not depend on $\underline{e}$.

3.4 Response Intervals and a-Stability

Now we shall establish sufficient conditions for an ADE to be tape-acceptance stable under environment sequence $\underline{e} = (e_0, e_1, \ldots)$. We shall consider the problem when we have that λ is an isolated cut-point for A under total environment $\underline{e}$. Recall that a cut-point λ for A under total environment $\underline{e}$ is isolated if there exists $\gamma > 0$ such that $|p(x,\underline{e}) - \lambda| \geq \gamma$ $\forall x\in\Sigma^*$. In view of theorem 3.1.2 and corollary 3.3.7, we shall not require the automaton to be s-stable or ever-ergodic at $\underline{e}$. We shall tolerate perturbations of the transition functions as long as the acceptance probability function does not cross the cut-point.

Let us consider an ADE $A = (\Sigma, S, G, \pi_0, F, E)$ and some total environment $\underline{e}\in E^\infty$. Define the response intervals R_i, $i = 1,2,\ldots,$ as

$$R_i = [0,1] \cap \left\{ \bigcup_{\sigma\in\Sigma} [\sum_{j\in F} \Delta(\sigma,i,j), \sum_{j\in F} \nabla(\sigma,i,j)] \right\},$$

where $\Delta(\sigma,i,j) = \min_k\{g_{kj}(\sigma,e_i)\}$ and $\nabla(\sigma,i,j) = \max_k\{g_{kj}(\sigma,e_i)\}$. The values $g_{kj}(\sigma,e_i)$ are the components of $G(\sigma,e_i)$. We also define

$$R_0 = \{p(\Lambda,\underline{e})\} = \{\pi_0(e_0)\eta^F\} .$$

<u>Theorem 3.4.1</u>. Let $A = (\Sigma, S, G, \pi_0, F, E)$ be an ADE. For any $\underline{e}\in E^\infty$ and $\lambda\in[0,1)$, $T(A,\underline{e},\lambda) = T(A,\underline{e}',\lambda)$ whenever

1. $\lambda \notin \bigcup_{i\in J} R_i$.

2. $\inf_{r_i \in R_i} \frac{|\lambda - r_i|}{\#(F)} > |G(\sigma,e_i) - G(\sigma,e_i')| \quad \forall\sigma\in\Sigma$ and $\forall i\in J_+$.

3. $|\lambda - p(\Lambda,\underline{e})| > |p(\Lambda,\underline{e}) - p(\Lambda,\underline{e}')|$,

where R_i, $i=0,1,\ldots$, are defined for A under total environment $\underline{e}$ as above.

<u>Proof</u>: Consider Λ, the empty tape. $R_0 = \{p(\Lambda,\underline{e})\}$. $\lambda \notin \bigcup_{i\in J} R_i$ implies $\lambda \neq p(\Lambda,\underline{e})$.

If $\Lambda\in T(A,\underline{e},\lambda)$, then $p(\Lambda,\underline{e}) > \lambda \geq 0$. Now $p(\Lambda,\underline{e}') = p(\Lambda,\underline{e}) + p(\Lambda,\underline{e}') - p(\Lambda,\underline{e}) \geq p(\Lambda,\underline{e}) - |p(\Lambda,\underline{e}) - p(\Lambda,\underline{e}')|$. Hence, by condition 3, $p(\Lambda,\underline{e}') > p(\Lambda,\underline{e}) - |\lambda - p(\Lambda,\underline{e})| \geq \lambda$. Thus, $p(\Lambda,\underline{e}') > \lambda$; consequently $\Lambda\in T(A,\underline{e}',\lambda)$.

If $\Lambda \notin T(A,\underline{e},\lambda)$, then $0 \leq p(\Lambda,\underline{e}) < \lambda$. Now

$$\begin{aligned} p(\Lambda,\underline{e}') &= p(\Lambda,\underline{e}) + p(\Lambda,\underline{e}') - p(\Lambda,\underline{e}) \\ &\leq p(\Lambda,\underline{e}) + |p(\Lambda,\underline{e}) - p(\Lambda,\underline{e}')| \\ &< p(\Lambda,\underline{e}) + |\lambda - p(\Lambda,e)| \quad \text{by condition 3} \\ &\leq \lambda . \end{aligned}$$

Thus $p(\Lambda,\underline{e}') \leq \lambda$; consequently $\Lambda\notin T(A,\underline{e}',\lambda)$.

Now consider any other element of Σ^*. We can write it in the form $x\sigma\in\Sigma^*$, where $x\in\Sigma^*$ and $\sigma\in\Sigma$. Since $\sigma\in\Sigma$, then it is impossible for $x\sigma=\Lambda$.

Let v and w be arbitrary vectors in V_n. The functions Δ and ∇ are defined so that for any $v\in V_n$ and $i = L(x) + 1$, we have

$$\sum_{j\in F} \Delta(\sigma,i,j) \le vG(\sigma,e_i)\eta^F \le \sum_{j\in F} \nabla(\sigma,i,j) .$$

Condition 2 implies that for any $w\in V_n$ that

$$\sum_{j\in F} \Delta(\sigma,i,j) - \delta_i{}^{\#}(F) \le wG(\sigma,e_i')\eta^F \le \sum_{j\in F} \nabla(\sigma,i,j) + \delta_i{}^{\#}(F) ,$$

where $\delta_i = \inf_{r_i\in R_i} \frac{|\lambda-r_i|}{{}^{\#}(F)} > 0 \quad \forall i\in J_+$.

Now let $v = \pi(x,\underline{e})$ and $w = \pi(x,\underline{e}')$. So $vG(\sigma,e_i)\eta^F = p(x\sigma,\underline{e})$ and $wG(\sigma,e_i')\eta^F = p(x\sigma,\underline{e}')$.

Since $\lambda \notin R_i$, then $p(x\sigma,\underline{e}) \ne \lambda$. If $x\sigma\in T(A,\underline{e},\lambda)$, then $p(x\sigma,\underline{e}) > \lambda$. But by condition 2, $\lambda < \sum_{j\in F} \Delta(\sigma,i,j) - \delta_i{}^{\#}(F) \le p(x\sigma,\underline{e}')$. Hence, $p(x\sigma,\underline{e}') > \lambda$; therefore $x\sigma\in T(A,\underline{e}',\lambda)$.

If $x\sigma\notin T(A,\underline{e},\lambda)$, then $p(x\sigma,\underline{e}) < \lambda$. Again by condition 2, $p(x\sigma,\underline{e}) \le \sum_{j\in F} \nabla(\sigma,i,j) + \delta_i{}^{\#}(F) < \lambda$. So $p(x\sigma,\underline{e}') < \lambda$; hence, $x\sigma\notin T(A,\underline{e}',\lambda)$.

Therefore, $x\sigma\in T(A,\underline{e},\lambda)$ iff $x\sigma\in T(A,\underline{e}',\lambda)$. But $x\sigma$ is any arbitrary, nonempty tape. Together with the considerations of the empty tape, we have $T(A,\underline{e},\lambda) = T(A,\underline{e}',\lambda)$. □

<u>Corollary 3.4.2</u>. Let $A = (\Sigma,S,G,\pi_0,F,E)$ be an ADE with natural metric d^N on E. If for some $\underline{e}\in E^\infty$ and $\lambda\in[0,1)$,

1. $\lambda \notin \overline{\bigcup_{i\in J} R_i}$ ($\overline{\quad}$ denotes closure),
2. G is uniformly continuous on $E(\underline{e})$,
3. π_0 is continuous at e_0,

then A with cut-point λ is a-stable at $\underline{e}$.

Proof: We are interested in finding $\delta > 0$ such that $\sup_{i\epsilon J} d^N(e_i,e_i') < \delta$ implies the conditions of theorem 3.4.1 hold.

Clearly $\lambda \notin \overline{\bigcup_{i\epsilon J} R_i}$ implies $\lambda \notin \bigcup_{i\epsilon J} R_i$. Also, $\lambda \notin \overline{\bigcup_{i\epsilon J} R_i}$ implies there exists $\gamma_1 > 0$ such that $\inf_{i\epsilon J_+} \inf_{r_i\epsilon R_i} |\lambda - r_i| = \gamma_1$. Now let $\varepsilon_1 = \gamma_1/\#(F)$. By the uniform continuity of G on $E(\underline{e})$, there exist $\delta_1 > 0$ such that

$$\max_{\sigma\epsilon\Sigma} |G(\sigma,e_i) - G(\sigma,e_i')| < \varepsilon_1 = \gamma_1/\#(F) \le \inf_{r_i\epsilon R_i} \frac{|\lambda - r_i|}{\#(F)} ,$$

$\forall i\epsilon J_+$ whenever $\sup_{i\epsilon J_+} d^N(e_i,e_i') < \delta_1$.

Again, $\lambda \notin \overline{\bigcup_{i\epsilon J} R_i}$ implies $\lambda \notin R_0$. Hence $\lambda \neq p(\Lambda,\underline{e})$. So let $\gamma_2 = |\lambda - p(\Lambda,\underline{e})| > 0$. Let $\varepsilon_2 = \gamma_2/\#(F)$. By the continuity of π_0 at e_0, there exists $\delta_2 > 0$ such that $|\pi_0(e_0) - \pi_0(e_0')| < \varepsilon_2$ whenever $d^N(e_0,e_0') < \delta_2$. Therefore

$$|p(\Lambda,\underline{e}) - p(\Lambda,\underline{e})| \le |\pi_0(e_0) - \pi_0(e_0')|\#(F) < \varepsilon_2\#(F) = \gamma_2$$
$$= |\lambda - p(\Lambda,\underline{e})|$$

whenever $d^N(e_0,e_0') < \delta_2$.

Now choose $\delta = \min(\delta_1,\delta_2)$. The conditions of theorem 3.4.1 are verified. Hence, there exists $\delta > 0$ such that $\forall \underline{e}'\epsilon E^\infty$, where $\sup_{i\epsilon J} d^N(e_i,e_i') < \delta$, then $T(A,\underline{e},\lambda) = T(A,\underline{e}',\lambda)$. That is, A with cut-point λ is a-stable at $\underline{e}$. □

Corollary 3.4.3. Let $A = (\Sigma,S,G,\pi_0,F,E)$ be an ADE. For any $\underline{e}\epsilon E^\infty$, λ is an isolated cut-point for A under total environment $\underline{e}$

if $\lambda \notin \overline{\bigcup_{i\in J} R_i}$.

<u>Proof</u>: Clearly, $\{p(x,\underline{e}) | x\in\Sigma^*\} \subset \overline{\bigcup_{i\in J} R_i}$. But $\lambda \notin \overline{\bigcup_{i\in J} R_i}$ and $\overline{\bigcup_{i\in J} R_i}$ is a closed set. Hence, there exists $\gamma > 0$ such that $|p(x,\underline{e})-\lambda| \geq \gamma \quad \forall x\in\Sigma^*$. □

<u>Example 3.4.1</u>. Let $A = (\Sigma,S,G,\pi_0,F,E)$ be an ADE with $\Sigma = \{\sigma_1,\sigma_2\}$, $S = \{s_1,s_2,s_3,s_4\}, F = \{s_2,s_3\}$, and $E = [-1/2,1/2]$. Let Euclidean distance be the natural metric on E. We define G for σ_1 as

$$G(\sigma_1,e) = \begin{pmatrix} 11/12-e^4 & 1/12+e^4 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1/6+e^4 & 5/6-e^4 \end{pmatrix}$$

and for σ_2 as

$$G(\sigma_2,e) = \begin{pmatrix} 1/12 & 1/12-e^4 & 5/6+e^4 & 0 \\ 0 & 1/4+e^4 & 3/4-e^4 & 0 \\ 0 & 0 & 1 & 0 \\ 2e^4 & 0 & 3/4-2e^4 & 1/4 \end{pmatrix} .$$

Define $\pi_0(e) = (e_4, 1/4-3e^4, e^4, 3/4+e^4)$.

E is a compact set with respect to the natural metric topology. Since for each $\sigma\in\Sigma$ the components of $G(\sigma,e)$ are continuous at each $e\in E$, then G is uniformly continuous with respect to d^N. For any $e\in E$, π_0 is also continuous at e with respect to d^N.

Let $\underline{e} = (e_0, e_1, \ldots) \epsilon E^\infty$. Thus $R_0 = \{1/4 - 2e_0^4\}$ and $R_i = [0, 1/4 + 2e_i^4] \cup [3/4 - 2e_i^4, 1]$. Clearly, $\overline{\bigcup_{i \epsilon J} R_i} \subset [0, 3/8] \cup [5/8, 1]$. Hence, for any $\lambda \epsilon (3/8, 5/8)$, A with cut-point λ is a-stable at any $\underline{e} \epsilon E^\infty$. Note that any $\lambda \epsilon (3/8, 5/8)$ is an isolated cut-point.

For any $\delta > 0$ and any $\underline{e} \epsilon E^\infty$, there does not exist an integer $k > 0$ such that $M(\sigma_1^k, \underline{e}) = \prod_{i=1}^{k} G(\sigma_1, e_i)$ is δ-scrambling. Hence, A does not satisfy the ever-$H_4(\delta)$ condition at any $\underline{e} \epsilon E^\infty$ for any $\delta > 0$. Thus, A is not ever-ergodic at any $\underline{e} \epsilon E^\infty$.

3.5 Round Off Considerations

Suppose we have an ADE $A = (\Sigma, S, G, \pi_0, F, E)$ with natural metric d^N on E. We, however, are not able to measure the environment configuration exactly; the environment can only be measured correctly up to a certain approximation and the remainder is subject to round off. That is, if $e \epsilon E$ is the true condition of the environment, we can only observe a unique round off estimate $f_\gamma(e)$ such that $d^N(e, f_\gamma(e)) \le \gamma$ for some $\gamma > 0$. Thus, we shall assume that we have a subset $F_\gamma = \{f_\gamma(e) | e \epsilon E\}$ that is γ-dense in E. That is, for any $e \epsilon E$, there exists a unique $f \epsilon F_\gamma$ such that $d^N(e, f) \le \gamma$.

For notational convenience, we shall let $f_\gamma(\underline{e})$ denote the component-wise rounded value of $\underline{e}$. Thus $f_\gamma(\underline{e}) = (f_\gamma(e_0), f_\gamma(e_1), \ldots)$. Let $F_\gamma^\infty = \{(f_0, f_1, \ldots) | f_i \epsilon F_\gamma\}$. Clearly, $F_\gamma^\infty = \{f_\gamma(\underline{e}) | \underline{e} \epsilon E^\infty\}$. For any $e \epsilon E^\infty$, $\sup_{i \epsilon J} d^N(e_i, f(e_i)) \le \gamma$. Hence, F_γ^∞ is γ-dense in E^∞ with respect to the metric $\sup_{i \epsilon J} d^N$.

Now we can use the stability results of the previous sections to find round off schemes which maintain stability of an acceptance set or stability of the acceptance probability function.

Suppose we have an ADE A which is a-stable at some cut-point λ for some environment sequence $\underline{f}\in F_\gamma^\infty$. There exists $\delta > 0$ such that $T(A,\underline{e},\lambda) = T(A,\underline{f},\lambda)$ whenever $\sup_{i\in J} d^N(e_i,f_i) < \delta$. Clearly, if $\gamma < \delta$, then $\forall \underline{e}\in E^\infty$ such that $f_\gamma(\underline{e}) = \underline{f}$, we have $\sup_{i\in J} d^N(f_i,e_i) = \sup_{i\in J} d^N(f_\gamma(e_i),e_i) \leq \gamma < \delta$. Thus for sufficiently small $\gamma > 0$, $T(A,\underline{f},\lambda) = T(A,\underline{e},\lambda)$ $\forall \underline{e}\in E^\infty$ such that $f_\gamma(\underline{e}) = \underline{f}$. So if we have a-stability for some cut-point λ at the rounded value of the environment sequence and if the round off error is sufficiently small, then the set defined by the ADE with cut-point λ remains unchanged by the round off procedure. We also obtain this result under the conditions of corollary 3.4.2 when γ is sufficiently small.

Now suppose that A is s-stable at some $\underline{f}\in F_\gamma^\infty$. Hence, given any $\varepsilon > 0$, there exists $\delta(\varepsilon)$ such that $|\pi(x,\underline{f}) - \pi(x,\underline{e})| < \varepsilon$ $\forall x\in\Sigma^*$ whenever $\sup_{i\in J} d^N(f_i,e_i) < \delta(\varepsilon)$. Clearly, if $\gamma < \delta(\varepsilon)$, then $\forall \underline{e}\in E^\infty$ such that $f_\gamma(\underline{e}) = \underline{f}$, we have $|\pi(x,f_\gamma(\underline{e})) - \pi(x,\underline{e})| < \varepsilon$ $\forall x\in\Sigma^*$. For γ sufficiently small, the conditions of corollary 3.3.7 also imply this result.

The round off procedure is also applicable to approximation of the acceptance probability function for sets of tapes of bounded length. In this case we need only consider $f_0,f_1,\ldots,f_k$, for some $k\in J$. For any $\varepsilon > 0$, by theorem 3.2.1 if γ is sufficiently small, then $|p(x,\underline{e})-p(x,\underline{f})| < \varepsilon$ $\forall x\in\Sigma^*$ such that $L(x)\leq k$, whenever $f_\gamma(e_i) = f_i$ for $i = 1,\ldots,k$.

CHAPTER IV

Automata in Random Environments

4.1. Basic Notions.

We shall now consider the case of automata operating within random environments. In contrast to the case of the ADE, where we suppose we can control the environment, we shall assume that the realization of the environment is governed by some probabilistic structure. This formulation is useful when we are only able to make certain probabilistic assumptions about the occurrence of any environment configuration or when the environment configuration can only be measured by statistical techniques.

Let $\mathcal{A} = (\Sigma, S, G, \pi_0, F, \mathcal{E})$ be an automaton in a deterministic environment. Let $(\Omega, \mathcal{B}, P)$ be a probability space. For each $j \in \{0,1,\ldots\} = J$, let z_j be a measurable function from $(\Omega, \mathcal{B})$ to $(\mathcal{E}, \mathcal{B}')$, where $\mathcal{B}'$ is a σ-field of subsets of $\mathcal{E}$. We shall use z_0 to denote the random variable for the initial configuration of the environment and z_j, $j=1,2,\ldots$, to denote the random variable for the configuration of the environment for the j-th input symbol. The family of random variables $Z = \{z_j \mid j \in J\}$ is an environmental stochastic process (ESP). When we wish to emphasize the probability structure, we shall use $(\Omega, \mathcal{B}, P, Z)$ to denote the process.

Definition 4.1.1. An automaton in a random environment (ARE) is a system (A,Z), where A is an ADE and Z is an ESP.

Let $A = (\Sigma,S,G,\pi_0,F,E)$ be an ADE and (Ω,B,P,Z) be an ESP. Note that $G: \Sigma\times E\to M_n$ and $z_j: \Omega\to E$ for all $j\in J_+ = \{1,2,\ldots\}$. Now for all $\sigma\in\Sigma$, we define $G_\sigma(e) = G(\sigma,e)$ and assume $G_\sigma(e)$ is measurable. Thus, $G_\sigma\circ z_j: \Omega\to M_n$ such that $G_\sigma\circ z_j(\omega) = G(\sigma,z_j(\omega))$ is measurable. For convenience we shall call $G_\sigma\circ z_j$ by the name $G(\sigma,z_j)$. Similarly, $z_0:\Omega\to E$ and we assume that $\pi_0: E\to V_n$ is also measurable. Hence, $\pi_0\circ z_0:\Omega\to V_n$ such that $\pi_0\circ z_0(\omega) = \pi_0(z_0(\omega))$ is measurable. We shall call $\pi_0\circ z_0$ by the name $\pi_0(z_0)$. Note that $G(\sigma,z_j)$ for all $j\in J_+$ and all $\sigma\in\Sigma$ is a random stochastic matrix and $\pi_0(z_0)$ is a random stochastic vector. We shall, however, encounter no ambiguity in using the notation G and π_0 in both the ADE and ARE contexts.

Let (Ω,B,P,Z) be an ESP with environment set E. Define $z: \Omega \to E^\infty$ by $z(\omega) = (z_0(\omega),z_1(\omega),\ldots)$. The mapping z is measurable map from (Ω,B) to (E^∞, B'^∞), where B'^∞ is the smallest σ-field generated by the measurable cylinders. We can use this to define the random matrix transition function. Let $x\in\Sigma^*$ such that $L(x) = k$. Now, $M(x,z(\omega)) = \prod_{i=1}^{k} G(\sigma_i,z_i(\omega))$. Similarly, we have $\pi(x,z(\omega)) = \pi_0(z_0(\omega))M(x,z(\omega))$ and $p(x,z(\omega)) = \pi(x,z(\omega))\eta^F$. Furthermore, $T(A,z,\lambda)$ for any $\lambda\in[0,1)$ is a set function of $\omega\in\Omega$ taking values in the set of all subsets of Σ^*.

The functions $\pi_0(z_0)$ and $\pi(x,z)$ for any $x\in\Sigma^*$ take values in V_n. It is not difficult to see that V_n is a convex set; thus, $E\ \pi_0(z_0)\in V_n$ and $E\ \pi(x,z)\in V_n$ for any $x\in\Sigma^*$. Similarly, M_n is a convex set; hence $E\ G(\sigma,z_j)\in M_n$ for all $\sigma\in\Sigma$ and for all $j\in J_+$ and

$EM(x,z)\in\mathcal{M}_n$ for all $x\in\Sigma^*$. Note also that [0,1] is a convex set; so $Ep(x,z)\in[0,1]$ for all $x\in\Sigma^*$.

Definition 4.1.2. Let (A,Z) be an ARE and λ a real number, $\lambda\in[0,1)$. The set of tapes $ET(A,z,\lambda)$ defined by $ET(A,z,\lambda) = \{x \mid x\in\Sigma^*, \lambda<Ep(x,z)\}$ is called the expected set of tapes accepted, or defined, by (A,Z) with cut-point λ.

Suppose $z^{(1)}, z^{(2)}, \ldots$ are independent identically distributed (IID) sequences of random environments. Thus, for any fixed, but arbitrary $x\in\Sigma^*$, $\{p(x,z^{(i)})\}_{i=1}^{\infty}$ is a sequence of IID random variables since $|p(x,z^{(i)})| \le 1$ a.s. for all $i=1,2,\ldots$, we may apply the strong law of large numbers to obtain $\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} p(x,z^{(i)}) = Ep(x,z^{(1)})$ a.s. We can obtain similar almost sure convergence for the components of π and M.

Let $A = (\Sigma,S,G,\pi_0,F,E)$ be an ADE. Let us consider the probability space $([0,1], \mathcal{B}([0,1]),\mu)$, where $\mathcal{B}([0,1])$ is the relative Borel field and μ is Lebesgue measure. We define the random variable $\psi(\theta) = \theta$, $\theta\in[0,1]$, on this probability space. For any fixed, but arbitrary $x\in\Sigma^*$ and $\underline{e}\in E^{\infty}$, we define the random variables

$$I(x,\underline{e},\psi(\theta)) = \begin{cases} 1 & \text{if } \psi(\theta) \le p(x,\underline{e}) \\ 0 & \text{otherwise .} \end{cases}$$

Clearly, for any $x \in \Sigma^*$ and $\underline{e} \in E^\infty$, $I(x,\underline{e},\psi)$ is measurable and $\mu(\theta \mid I(x,\underline{e},\psi(\theta)) = 1) = p(x,\underline{e})$. Also, $E\, I(x,\underline{e},\psi) = \int_{[0,1]} I(x,\underline{e},\psi(\theta))d\mu(\theta) = p(x,\underline{e})$. The random variable $I(x,\underline{e},\psi)$ has the same relevant probability structure as an indicator which is 1 when A is in a state in F after input x under total environment $\underline{e}$ and 0 otherwise.

Let $(\Omega,\mathcal{B},P,Z)$ be an ESP with environment set E. We have defined $z\colon \Omega \to E^\infty$. Hence, we can consider the composition $I(x,z(\cdot),\psi(\cdot))$ as a function defined on the product space $[0,1] \times \Omega$, where

$$I(x,z(\omega),\psi(\theta)) = \begin{cases} 1 & \text{if} \quad \psi(\theta) \le p(x,z(\omega)) \\ 0 & \text{otherwise} . \end{cases}$$

So for any fixed, but arbitrary $x \in \Sigma^*$, $I(x,z,\psi)$ is a random variable defined on the product probability space. We now obtain $E\, I(x,z,\psi)$ with respect to the product measure.

$$E\, I(x,z,\psi) = \int_{[0,1]\times\Omega} I(x,z(\omega),\psi(\theta))d(\mu \times p) .$$

By Fubini's theorem,

$$\begin{aligned} E\, I(x,z,\psi) &= \int_\Omega \int_{[0,1]} I(x,z(\omega),\psi(\theta))d\mu(\theta)dp(\omega) \\ &= \int_\Omega p(x,z(\omega))dp(\omega) = E\, p(x,z) , \end{aligned}$$

where the last expectation is taken with respect to $(\Omega,\mathcal{B},P)$.

Suppose $\{(z^{(i)},\psi^{(i)})\}_{i=1}^\infty$ are independent copies of (z,ψ). Hence $\{I(x,z^{(i)},\psi^{(i)})\}_{i=1}^\infty$ is a sequence of random variables which are IID according to the product measure $\mu \times P$. By the strong law of large numbers,

$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} I(x,z^{(i)},\psi^{(i)}) = E\ p(x,z)$ a.s. $\mu\times P$. Thus, if we repeatedly input tape x under IID random environment sequences, the relative frequency of acceptance of x is a strongly consistent, unbiased estimator of $E\ p(x,z)$.

4.2 Random Environment Sequences of Independent Random Variables

Let (A,Z) be an ARE such that $z = (z_0,z_1,\ldots)$ is a sequence of independent random variables. Thus, for any $x\epsilon\Sigma^*$ such that

$$L(x) = k,\ Ep(x,z) = E\{\pi_0(z_0)\Big(\prod_{i=1}^{k} G(\sigma_i z_i)\Big)\eta^F\}$$

$$= (E\pi_0(z_0))\Big(\prod_{i=1}^{k} EG(\sigma_i,z_i)\Big)\eta^F$$

And, similarly, the expectations of the other automata functions are products of the expectations of their factors.

Let $Z(z) = \{z_j | j\epsilon J\}$ be the set of all mappings in the sequence z. For our purposes here, we shall consider two environmental random variables to be equal if their distribution functions are equal. Let E' be a set which is in one-to-one correspondence with $Z(z)$. That is, we have a functional $g\colon Z(z) \to E'$, which is a one-to-one correspondence. For any $z\epsilon Z(z)$ and $\sigma\epsilon\Sigma$, define $\pi_0'(g(z)) = E\pi_0(z)$ and $G'(\sigma,g(z)) = EG(\sigma,z)$. Thus, $A' = (\Sigma,S,G',\pi_0,F,E')$ is an ADE. Also, for any $x\epsilon\Sigma^*$,

$$p'(x,g(z)) = \pi_0'(g(z_0))\Big(\prod_{i=1}^{k} G(\sigma_i,g(z_i))\Big)\eta^F$$

$$= (E\pi_0(z_0))\Big(\prod_{i=1}^{k} EG(\sigma_i,z_i)\Big)\eta^F$$

$$= Ep(x,z).$$

Therefore, for any $\lambda\epsilon[0,1)$, $T(A',g(z),\lambda) = ET(A,z,\lambda)$. Furthermore, if $Z(z)$ contains only a finite number of random variables with distinct distributions, then E' is a finite set. Hence, A' can be simulated by a PA. In the particular case when z is a sequence of IID random variables, then $\#(E') = 1$ and we can find a PA A'' over Σ such that $T(A'',\lambda) = ET(A,z,\lambda)$ for all $\lambda\epsilon[0,1)$.

Suppose we have an ARE (A,Z) where $\Sigma = \{\sigma\}$ and $z_1,z_2,\ldots$ is a sequence of IID random variables. The following results hold for state sets of A of any finite cardinality, but for the sake of simplicity of exposition we shall restrict consideration to $S = \{s_1,s_2\}$. Let $\sigma^N\epsilon\Sigma^*$ denote the tape of N successive inputs of σ. We wish to find the asymptotic moments of $M(\sigma^N,z)$. Let (α_N,β_N) be the first row of $M(\sigma^N,z)$. Clearly, for every $\omega\epsilon\Omega$, $(\alpha_N,\beta_N)\epsilon V_2$ for all $N = 1,2,\ldots$. Clearly, $(\alpha_{N+1},\beta_{N+1}) = (\alpha_N,\beta_N)G(\sigma,z_{N+1})$. By the IID property, $(E\alpha_{N+1},E\beta_{N+1}) = (E\alpha_N,E\beta_N)EG(\sigma,z_{N+1}) = (E\alpha_1,E\beta_1)(EG(\sigma,z_1))^N$. We say a matrix $A\epsilon M_n$ is ergodic if it can be considered as a transition matrix of an irreducible ergodic Markov chain as defined in Feller [6]. Thus, if $EG(\sigma,z_1)$ is an ergodic matrix, then $(EG(\sigma,z_1))^N$ converges to a limit matrix L_1. The rows of L_1 are identical; thus $(E\alpha_N,E\beta_N)$ converges to $(E\alpha_1,E\beta_1)L_1 = \ell_1$, where ℓ_1 is any row of L_1.

Let k be any fixed positive integer. For every $\omega\epsilon\Omega$ and $N = 1,2,\ldots$, $(\alpha_N+\beta_N)^k = 1$. Thus $(\alpha_N^k,\binom{k}{1}\alpha_N^{k-1}\beta_N,\ldots,\beta_N^k)\epsilon V_{k+1}$ for every $\omega\epsilon\Omega$ and $N = 1,2,\ldots$. From the above, we know that

$$\alpha_{N+1} = \alpha_N g_{11}(\sigma,z_{N+1}) + \beta_N g_{21}(\sigma,z_{N+1})$$
$$\beta_{N+1} = \alpha_N g_{12}(\sigma,z_{N+1}) + \beta_N g_{22}(\sigma,z_{N+1}) \;,$$

where $g_{ij}(\sigma,z_{N+1})$ is the (i,j) component of $G(\sigma,z_{N+1})$. By taking the appropriate products we can derive an expression for each component of $(\alpha_{N+1}^{k},\binom{k}{1}\alpha_{N+1}^{k-1}\beta_{N+1},\ldots,\beta_{N+1}^{k})$ in terms of the components of $(\alpha_{N}^{k},\binom{k}{1}\alpha_{N}^{k-1}\beta_{N},\ldots,\beta_{N}^{k})$ and $G(\sigma,z_{N+1})$. This reduces to the relation $(\alpha_{N+1}^{k},\binom{k}{1}\alpha_{N+1}^{k-1}\beta_{N+1},\ldots,\beta_{N+1}^{k}) = (\alpha_{N}^{k},\binom{k}{1}\alpha_{N}^{k-1}\beta_{N},\ldots,\beta_{N}^{k})G(\sigma,z_{N+1})_{[k]}$, where $G(\sigma,z_{n+1})_{[k]}$ is the k-th Kronecker power of $G(\sigma,z_{N+1})$ as defined in Bellman [3]. Observe that $G(\sigma,z_{N+1})_{[k]}\in\mathcal{M}_{k+1}$. Again, by the IID property

$$\begin{aligned}
&(E\alpha_{N+1}^{k},E\binom{k}{1}\alpha_{N+1}^{k-1}\beta_{N+1},\ldots,E\beta_{N+1}^{k}) \\
&\quad = (E\alpha_{N}^{k},E\binom{k}{1}\alpha_{N}^{k-1}\beta_{N},\ldots,E\beta_{N}^{k})(EG(\sigma,z_{N+1})_{[k]}) \\
&\quad = (E\alpha_{1}^{k},E\binom{k}{1}\alpha_{1}^{k-1}\beta_{1},\ldots,E\beta_{1}^{k})(EG(\sigma,z_{1})_{[k]})^{N} .
\end{aligned}$$

If $EG(\sigma,z_1)_{[k]}$ is an ergodic matrix, $(EG(\sigma,z_1)_{[k]})^N$ converges to a limit L_k with identical rows. Thus as before, $(E\alpha_{N}^{k},E\binom{k}{1}\alpha_{N}^{k-1}\beta_{N},\ldots,E\beta_{N}^{k})$ converges to ℓ_k, a row of L_k. So ℓ_k is a vector of asymptotic moments of the first row of $M(\sigma^N,z)$.

4.3 Limit Results for Sequences of Environment Processes

Let $A = (\Sigma,S,G,\pi_0,F,E)$ be an ADE and $(\Omega,\mathcal{B},P)$ be a probability space. Let $\ell_\alpha(\Omega,\mathcal{B},P;E)$ be the space of all environment valued sequences defined on Ω such that for any

$$z = (z_0,z_1,\ldots)\in\ell_\alpha(\Omega,\mathcal{B},P;E), \quad \sum_{j\in J} E|z_j(\omega)|^\alpha < \infty$$

for almost every $\omega\in\Omega$. For the sake of generality, we do not give an explicit expression for the norm of the environment configuration.

For $z^{(1)}, z^{(2)} \in \ell_\alpha(\Omega,\mathcal{B},P;\mathcal{E})$, if we consider $z^{(1)} = z^{(2)}$ iff $z^{(1)}(\omega) = z^{(2)}(\omega)$ a.s., then $\ell_\alpha(\Omega,\mathcal{B},P;\mathcal{E})$ is a metric space.

The metric for this space when $\alpha\in(0,1]$ is

$$d_\alpha(z^{(1)},z^{(2)}) = \sum_{j\in J} E|z_j^{(1)}-z_j^{(2)}|^\alpha ,$$

where $z^{(1)} = (z_0^{(1)},z_1^{(1)},\ldots)$ and $z^{(2)} = (z_0^{(2)},z_1^{(2)},\ldots)$ are arbitrary elements of $\ell_\alpha(\Omega,\mathcal{B},P;\mathcal{E})$.

Let G be the basic transition function of A. We say G satisfies the Lipschitz condition of order α if there exists a positive constant K such that $\max_{\sigma\in\Sigma}||G(\sigma,e) - G(\sigma,e')|| \le K|e-e'|^\alpha$ for all $e,e'\in\mathcal{E}$. We shall only consider $\alpha\in(0,1]$ since other values imply that $G(\sigma,\circ)$ is not continuous or that $G(\sigma,\circ)$ is a constant for each $\sigma\in\Sigma$.

Let $z^{(1)},z^{(2)},\ldots$ be a sequence of environment sequences in $\ell_\alpha(\Omega,\mathcal{B},P;\mathcal{E})$ that converges to z in $\ell_\alpha(\Omega,\mathcal{B},P;\mathcal{E})$. That is, $d_\alpha(z^{(i)},z) \to 0$ as $i \to \infty$. Let us now suppose that G satisfies the Lipschitz condition of order α. For any $x\in\Sigma^*$ and any $\omega\in\Omega$, by lemma 3.2.2, we have

$$||M(x,z^{(i)}(\omega)) - M(x,z(\omega))|| \le \sum_{j=1}^{k} ||G(\sigma_j,z_j^{(i)}(\omega)) - G(\sigma_j,z_j(\omega))||.$$

However, by the Lipschitz condition

$$\sum_{j=1}^{k} ||G(\sigma_j,z_j^{(i)}(\omega)-G(\sigma_j,z_j(\omega))|| \le K\sum_{j=1}^{k} |z_j^{(i)}(\omega)-z_j(\omega)|^\alpha .$$

Thus

$$E||M(x,z^{(i)})-M(x,z)|| = \int_\Omega ||M(x,z^{(i)}(\omega))-M(x,z(\omega)||dP(\omega)$$

$$\le \int_\Omega K\sum_{j=1}^{k} |z_j^{(i)}(\omega)-z_j(\omega)|^\alpha dP(\omega)$$

$$= K \sum_{j=1}^{k} \int_{\Omega} |z_j^{(i)}(\omega)-z_j(\omega)|^{\alpha} dP(\omega)$$

$$\leq K \sum_{j\in J} \int_{\Omega} |z_j^{(i)}(\omega)-z_j(\omega)|^{\alpha} dP(\omega)$$

$$= K\, d_{\alpha}(z^{(i)},z) \ .$$

But $d_{\alpha}(z^{(i)},z) \to 0$ as $i \to \infty$. Thus, $E||M(x,z^{(i)})-M(x,z)|| \to 0$ as $i \to \infty$ for all $x\in\Sigma^*$. Hence, $M(x,z^{(i)}) \to M(x,z)$ in probability uniformly for all $x\in\Sigma^*$. Since $z^{(i)} \to z$ in $\mathcal{L}_{\alpha}(\Omega,\mathcal{B},P;E)$, we have that $z_0^{(i)} \to z_0$ in probability. If π_0 is continuous, then $\pi_0(z_0^{(i)}) \to \pi_0(z_0)$ in probability. Thus, by the fact that the transition functions are bounded, we have for all $x\in\Sigma^*$ that $\pi(x,z^{(i)}) = \pi_0(z_0^{(i)})M(x,z^{(i)}) \to \pi_0(z_0)M(x,z) = \pi(x,z)$ in probability and $p(x,z^{(i)} = \pi(x,z^{(i)})\eta^F \to \pi(x,z)\eta^F = p(x,z)$ in probability as $i \to \infty$. Consequently, we have a stability and approximation result under the above conditions.

It is also possible to obtain an ever-ergodic type of stability for automata in random environments.

Definition 4.3.1. An ARE (A,Z) is ever-ergodic a.s. iff for any $\varepsilon > 0$, there exists $W\subset\Omega$, where $P(W) = 0$, and there exists $N(\varepsilon)$ such that $\partial(M(x,z_k(\omega))) < \varepsilon \ \ \forall z_k(\omega)\in E^*(z(\omega))$ whenever $\omega\notin W$ and $L(x)\geq N(\varepsilon)$.

Note that $z_k(\omega)$ is the k-tail of $z(\omega)$. Hence, for every $\omega\notin W$, A is ever-ergodic at $z(\omega)$ and the choice of $N(\varepsilon)$ is independent of ω.

Suppose there exists $0 < \varepsilon_1 < 1$ such that there exists $W_{\varepsilon_1} \subset\Omega$, where $P(W_{\varepsilon_1}) = 0$, and there exists $N(\varepsilon_1)$ whereby $\partial(M(x,z_k(\omega))) < \varepsilon$ $\forall z_k(\omega)) \in E^*(z(\omega))$ whenever $\omega \notin W_{\varepsilon_1}$ and $L(x) \geq N(\varepsilon_1)$. For any ε such that $0 < \varepsilon < \varepsilon_1$, we can find n_0 such that $\varepsilon_1^{n_0} < \varepsilon$. By lemma 3.3.2 we have that $\partial(M(x,z_k(\omega))) < \varepsilon_1^{n_0} < \varepsilon$ $\forall z_k(\omega) \in E^*(z(\omega))$ whenever $\omega \notin W_{\varepsilon_1}$ and $L(x) \geq n_0 N(\varepsilon_1)$. Thus, we can choose W to be independent of ε.

Let (A,Z) be an ARE. If there exists $\delta > 0$ such that each component of $G(\sigma,z_i)$ $\forall \sigma \in \Sigma$ and $\forall i \in J_+$ is greater than $\delta/2$ for almost every $\omega \in \Omega$, then $\partial(G(\sigma,z_i)) < 1-\delta$ a.s. $\forall \sigma \in \Sigma$ and $\forall i \in J_+$. By application of lemma 3.3.1 we see that $\partial(M(x,z_k)) < (1-\delta)^{n_0}$ a.s. $\forall z_k \in E^*(z)$ and $\forall x \in \Sigma^*$ such that $L(x) \geq n_0$. Thus, we have generated a sufficient condition for (A,Z) to be ever-ergodic a.s.

<u>Theorem 4.3.1</u>. Let (A,Z) be an ARE with natural metric d^N on E. If (A,Z) is ever-ergodic a.s., G is uniformly continuous and π_0 is uniformly continuous, then given any $\varepsilon > 0$, there exists $\delta > 0$ such that for any ARE (A,Z'), where Z' is defined on (Ω,B,P), satisfying $\sup_{i \in J} d^N(z_i(\omega),z_i'(\omega)) < \delta$ for almost every $\omega \in \Omega$, we have $|\pi(x,z(\omega))-\pi(x,z'(\omega))| < \varepsilon$ $\forall x \in \Sigma^*$ and for almost every $\omega \in \Omega$.

The proof of theorem 4.3.1 follows in a similar way as corollary 3.3.7 and its antecedents except that the results apply uniformly for almost every $\omega \in \Omega$. Furthermore, if we have a sequence of $z^{(1)},z^{(2)},\ldots$ of random environment sequences such that $z^{(i)} \to z$ a.s., then under the continuity conditions of theorem 4.3.1 we obtain that

$\pi(x,z^{(i)}) \to \pi(x,z)$ a.s. $\forall x \in \Sigma^*$. And, consequently, $p(x,z^{(i)}) \to p(x,z)$ a.s. $\forall x \in \Sigma^*$.

4.4 A Canonical Representation for Automata in Random Environments

In section 4.1 we encountered the dual nature of the randomness involved in automata in random environments. We have that the environment sequence is random and that for each value the environment sequence assumes, the probabilistic state transitions are defined. Hence, the state of the machine after an input tape under a random environment sequence is a function defined on a product probability space. We shall show that for any ARE (A,Z) there exists an ARE (A_D,Y) over the same alphabet whose transition probabilities are equivalent in the mean to those of (A,Z), but the state of (A_D,Y) obtained after an input under a random sequence is only related to the value the random environment sequence obtains. To accomplish this we need that the initial distribution function and the basic matrix transition functions of A_D take on values which, in addition to being stochastic vectors and stochastic matrices, respectively, must also have components that are either 0 or 1. Thus, the only randomness in the system (A_D,Y) is the randomness of the environment sequence. We shall also construct (A_D,Y) to have a finite environment set.

<u>Theorem 4.4.1</u>. Let (A,Z) be an ARE, where $A = (\Sigma,S,G,\pi_0,F,E)$ is an ADE and $Z = \{z_j | j \in J\}$ is an ESP defined on $(\Omega,\mathcal{B},P)$ taking values in E. There exists an ARE (A_D,Y), where $A_D = (\Sigma,S,H,\rho_0,F,E')$ is an

ADE and $Y = \{y_j | j \epsilon J\}$ is an ESP defined on some $(\Omega', \mathcal{B}', P')$ taking values in $\mathcal{E}'$, such that (A_D, Y) satisfies the following properties:

1. $H: \Sigma \times \Omega' \to D(\mathcal{M}_n)$
2. $\rho_0 : \Omega' \to D(\mathcal{V}_n)$
3. $\#(\mathcal{E}') < \infty$
4. $E\pi(x, z) = E\rho(x, y) \quad \forall x \epsilon \Sigma$, where $y = (y_0, y_1, \ldots)$
5. $ET(A, z, \lambda) = ET(A_D, y, \lambda) \quad \forall \lambda \epsilon [0,1)$.

Here we are using π and ρ to denote the state distribution functions of (A, Z) and (A_D, Y), respectively. Recall from definition 1.3.3 that $D(\mathcal{M}_n)$ is the set of all $n \times n$ stochastic matrices with components 0 or 1 and that $D(\mathcal{V}_n)$ is the set of all n-dimensional stochastic vectors with components 0 or 1. Note that n is the cardinality of the state set S. This property implies that knowledge of the input and environment sequence determines the state of (A_D, Y) obtained.

Proof: For any $\sigma \epsilon \Sigma$ and $e \epsilon \mathcal{E}$ we may decompose $G(\sigma, e)$ into a finite convex linear combination of elements of $D(\mathcal{M}_n)$ by the following algorithm:

1. Set $G(\sigma, e) = G_1$ and set $i = 1$.
2. Let g^i_{jk} be the smallest nonzero component of G_i, where j and k are its coordinates. If there is a tie, choose any one of the contenders.
3. Set $\alpha_{i\sigma}(e) = g^i_{jk}$.
4. Construct $B_{i\sigma}(e) \epsilon D(\mathcal{M}_n)$ by the rule:

   a. $(B_{i\sigma}(e))_{jk} = 1$.

b. For $r \neq j$, find the largest component of the r-th row of G_i, say g_{rt}, then set $(B_{i\sigma}(e))_{rt} = 1$. If there is a tie within any row, choose any one of the contenders.

c. Set all of the remaining components of $B_{i\sigma}(e)$ equal to zero.

5. Set $G_{i+1} = G_i - \alpha_{i\sigma}(e)B_{i\sigma}(e)$.

6. If G_{i+1} is a matrix of all 0 components, stop. If not, increase the value of i by 1 and go to step 2.

This algorithm must terminate in at most $p=n^2-n+1$ iterations regardless of e and σ because at each iteration at least one component of G_i is zeroed and the rows contain nonnegative components summing to $1 - \sum_{j=1}^{i} \alpha_{j\sigma}(e)$. Clearly for each i, $B_{i\sigma}(e) \epsilon D(M_n)$. If for some σ and e this process terminates before $i = p$, we define $\alpha_{i\sigma}(e) = 0$ for the remaining values of i and assign the corresponding $B_{i\sigma}(e)$ any arbitrary value in $D(M_n)$.

So we now have $\sum_{i=1}^{p} \alpha_{i\sigma}(e) = 1$ and $\sum_{i=1}^{p} \alpha_{i\sigma}(e)B_{i\sigma}(e) = G(\sigma,e)$ for each $\sigma \epsilon \Sigma$ and each $e \epsilon E$. Hence for each $e \epsilon E$, we have generated the following arrays of numbers and matrices:

$$\begin{Bmatrix} \alpha_{1\sigma_1}(e) & \cdots & \alpha_{p\sigma_1}(e) \\ \vdots & & \vdots \\ \alpha_{1\sigma_m}(e) & \cdots & \alpha_{p\sigma_m}(e) \end{Bmatrix} \qquad \begin{Bmatrix} B_{1\sigma_1}(e) & \cdots & B_{p\sigma_1}(e) \\ \vdots & & \vdots \\ B_{1\sigma_m}(e) & \cdots & B_{p\sigma_m}(e) \end{Bmatrix}$$

where m is the cardinality of Σ. For any $i_1,\ldots,i_m=1,\ldots,p$, let

$a_e(i_1,\ldots,i_m) = \prod_{j=1}^{m} \alpha_{i_j\sigma_j}(e)$. We now consider $A(e) = \{a_e(i_1,\ldots,i_m) \mid i_1,\ldots,i_m=1,\ldots,p\}^L$ as a list of numbers. A list differs from a set in that redundancy is preserved. That is, the lists $\{0\}^L$ and $\{0,0\}^L$ are not equal because 0 occurs once in the first list and twice in the second list. Clearly for any $e\epsilon E$ and any $i_1,\ldots,i_m=1,\ldots,p$, $a_e(i_1,\ldots,i_m) \geq 0$. Now for any $e\epsilon E$,

$$\sum_{i_1} \cdots \sum_{i_m} a_e(i_1,\ldots,i_m) = \sum_{i_1} \cdots \sum_{i_m} \prod_{j=1}^{m} \alpha_{i_j\sigma_j}(e)$$

$$= \sum_{i_1} \cdots \sum_{i_{m-1}} \prod_{j=1}^{m-1} \alpha_{i_j\sigma_j}(e) \sum_{i_m} \alpha_{i_m\sigma_m}(e)$$

$$= \sum_{i_1} \cdots \sum_{i_{m-1}} \prod_{j=1}^{m-1} \alpha_{i_j\sigma_j}(e) = \ldots = 1.$$

For each $\sigma_k \epsilon \Sigma$, we may partition $A(e)$ into p sublists $A_{1\sigma_k}(e),\ldots,A_{p\sigma_k}(e)$, where

$$A_{i\sigma_k}(e) = \{ a_e(i_1,\ldots,i_{k-1},i,i_{k+1},\ldots,i_m) \\ \mid i_1,\ldots,i_{k-1},i_{k+1},\ldots,i_m = 1,\ldots,p\}^L .$$

We obtain the sum of the elements of the list $A_{i\sigma_k}(e)$

$$\sum_{i_1} \cdots \sum_{i_{k-1}} \sum_{i_{k+1}} \cdots \sum_{i_m} a_e(i_1,\ldots,i_{k-1},i,i_{k+1},\ldots,i_m)$$

$$= \sum_{i_1} \cdots \sum_{i_{k-1}} \sum_{i_{k+1}} \cdots \sum_{i_m} \alpha_{i\sigma_k}(e) \prod_{j\neq k} \alpha_{i_j\sigma_j}(e)$$

$$= \alpha_{i\sigma_k}(e) \sum_{i_1} \cdots \sum_{i_{k-1}} \sum_{i_{k+1}} \cdots \sum_{i_m} \prod_{j\neq k} \alpha_{i_j\sigma_j}(e) = \alpha_{i\sigma_k}(e) .$$

There are p^m elements in the list $A(e)$. We now construct the list $A'(e) = \{a_i'(e) | i=1,\ldots,p^m\}^L$. For $i = 1 + \sum_{j=1}^{m} (i_j-1)p^{j-1}$, set $a_i'(e) = a_e(i_1,\ldots,i_m)$. This uniquely defines each element of $A'(e)$ and $A'(e) = A(e)$. For each $\sigma_k \epsilon \Sigma$, we partition $A'(e)$ into p sublists $A_{1\sigma_k}'(e),\ldots,A_{p\sigma_k}'(e)$, where $a_i'(e) \epsilon A_{j\sigma_k}'(e)$ iff $a_e(i_1,\ldots,i_{k-1},j,i_{k+1},\ldots,i_m) \epsilon A_{j\sigma_k}(e)$ and $i = 1 + (j-1)p^{k-1} + \sum_{r \neq k} (i_r-1)p^{r-1}$.

Actually, $A_{j\sigma_k}'(e) = A_{j\sigma_k}(e)$. Thus, we have reordered the set $A(e)$ and maintained the properties of the partitioning.

We now construct a list $B'(e) = \{B_{i\sigma}'(e) | i=1,\sigma\epsilon\Sigma,\ldots,p^m\}^L$ by the rule:

1. Set $i = 1$ and $k = 1$.
2. $a_i'(e)$ must belong to at least one of the sublists $A_{1\sigma_k}'(e),\ldots,$ $A_{p\sigma_k}'(e)$, say $A_{j\sigma_k}'(e)$. Set $B_{i\sigma}'(e) = B_{j\sigma_k}$.
3. Delete one occurrence of $a_i'(e)$ from the list $A_{j\sigma_k}'(e)$, but let the amended list retain the name $A_{j\sigma_k}'(e)$.
4. Increase the value of i by 1.
5. If $i \leq p^m$, go to step 2. If not, go to step 6.
6. If $k < m$, then set $i=1$, increase the value of k by 1, and go to step 2. If $k \geq m$, stop.

This algorithm must terminate because for each $\sigma\epsilon\Sigma$, $A_{1\sigma}'(e),\ldots,A_{p\sigma}'(e)$ is a partition of the p^m elements of $A'(e)$. Furthermore for each $\sigma\epsilon\Sigma$ each of these sublists will be emptied by the deletion process.

As a result of the construction of $A'(e)$ and $B'(e)$ we obtain for each $e\epsilon E$

1. $a_i'(e) \geq 0 \quad \forall i=1,\ldots,p^m, \sum_{i=1}^{p_m} a_i'(e) = 1.$

2. For each $\sigma \in \Sigma$, $A'(e)$ can be partitioned into p sublists $A_{1\sigma}'(e),\ldots,A_{p\sigma}'(e)$ such that $\sum_{a_i'(e) \in A_{j\sigma}'(e)} a_i'(e) = \alpha_{j\sigma}(e)$.

3. $B_{i\sigma}'(e) \in D(M_n) \quad \forall i=1,\ldots,p^m$ and $\forall \sigma \in \Sigma$ and $B_{i\sigma}'(e) = B_{j\sigma}(e)$ when i is an index of an element in $A_{j\sigma}'(e)$.

Also, for any $\sigma \in \Sigma$ and $e \in E$,

$$\sum_{i=1}^{p_m} a_i'(e) B_{i\sigma}'(e) = \sum_{j=1}^{p} \sum_{a_i'(e) \in A_{j\sigma}'(e)} B_{j\sigma}(e) a_i'(e)$$

$$= \sum_{j=1}^{p} B_{j\sigma}(e) \sum_{a_i'(e) \in A_{j\sigma}'(e)} a_i'(e)$$

$$= \sum_{j=1}^{p} B_{j\sigma}(e) \alpha_{j\sigma}(e) = G(\sigma,e) \; .$$

Hence, for each $e \in E$ we have the following arrangement

$$\begin{array}{lll} a_1'(e) & B_{1\sigma_1}'(e) \;\ldots & B_{1\sigma_m}'(e) \\ \vdots & \vdots & \vdots \\ a_{p^m}'(e) & B_{p^m\sigma_1}'(e) \;\ldots & B_{p^m\sigma_m}'(e) \end{array} \; .$$

Consider now the finite set $D_{n,m} = D(M_n) \times \ldots \times D(M_n)$, the Cartesian product of $D(M_n)$ with itself m times. Since $D(M_n)$ is a finite set with n^n distinct elements, then $D_{n,m}$ is a finite set with n^{nm} distinct elements. Since $D_{n,m}$ is a finite set, we can order it in a systematic way. For each $k = 1,\ldots,n^{nm}$, take the k-th element of

$D_{n,m}$ by our ordering and call it D_k. Also, for each $e \in E$ and $k = 1,\ldots,n^{nm}$, define $I_k(e) = \{i \mid (B'_{i\sigma_1}(e),\ldots,B'_{i\sigma_m}(e)) = D_k\}$ Note that $I_{k_1}(e) \cap I_{k_2}(e) = \emptyset$ for $k_1 \neq k_2$ and $\bigcup_{k=1}^{n} I_k(e) = \{1,\ldots,p^m\}$.

Thus, we set

$$\beta_k(e) = \begin{cases} \sum_{i \in I_k(e)} a'_i(e) & \text{if } I_k(e) \neq \emptyset \\ 0 & \text{if } I_k(e) = \emptyset \end{cases}$$

Hence, $\beta_k(e) \geq 0$ for $k = 1,\ldots,n^{nm}$ and for all $e \in E$. Also for each $e \in E$,

$$\sum_{k=1}^{n^{nm}} \beta_k(e) = \sum_{k=1}^{n^{nm}} \sum_{i \in I_k(e)} a'_i(e) = \sum_{i=1}^{p^m} a'_i(e) = 1$$

and

$$\sum_{k=1}^{n^{nm}} \beta_k(e) D_k = \sum_{k=1}^{n^{nm}} \sum_{i \in I_k(e)} a'_i(e)(B'_{i\sigma_1}(e),\ldots,B'_{i\sigma_m}(e))$$

$$= \sum_{i=1}^{p^m} a'_i(e)(B'_{i\sigma_1}(e),\ldots,B'_{i\sigma_m}(e))$$

$$= (G(\sigma_1,e),\ldots,G(\sigma_m,e)) .$$

We are now ready to define (A_D,Y). We retain Σ, S, and F from (A,Z). Let $E' = \{1,\ldots,n^{nm}\}$. Define $H(\sigma_i,k)$ to be the i-th matrix in D_k. For $k = 1,\ldots,n$, let $\rho_0(k)$ be an n-dimensional row vector of all zeroes except for a 1 in the k-th component. For $n+1 \leq k \leq n^{nm}$, let $\rho_0(k)$ be the n-dimensional vector $(1,0,\ldots,0)$. Hence, H is defined for all $\sigma \in \Sigma$ and all $e' \in E'$ and ρ_0 is defined for all $e' \in E$.

Define $\gamma_k(e) = \begin{cases} \pi_0^{(k)}(e) & k = 1,\ldots,n \\ 0 & k = n+1,\ldots,n^{nm} \end{cases}$

where $\pi_0^{(k)}(e)$ is the k-th component of $\pi_0(e)$. By Kolmogorov's Existence Theorem we define the process $Y = \{y_j | j \in J\}$ by the finite dimensional distributions

$$P(y_{j_1}=k_1,\ldots,y_{j_r}=k_r) = \int_\Omega \beta_{k_1}(z_{j_1})\ldots\beta_{k_r}(z_{j_r})dP ,$$

where $j_1,\ldots,j_r \in J$ are all distinct. If some $j_i = 0$, replace $\beta_{k_i}(z_{j_i})$ by $\gamma_{k_i}(z_0)$. Note that the integral is taken with respect to the probability space defining Z.

Thus, for any $r \in J$,

$$P(y_0=k_0,y_1=k_1,\ldots,y_r=k_r) = \int_\Omega \gamma_{k_0}(z_0)\beta_{k_1}(z_1)\ldots\beta_{k_r}(z_r)dP .$$

Let $x \in \Sigma^*$ be an input tape of arbitrary length, say $x = \sigma_1\ \sigma_2\ \ldots\ \sigma_r$. Hence

$$E\rho(x,y) = \sum_{k_0}\ldots\sum_{k_r} \rho_0(k_0)H(\sigma_1,k_1)\ldots H(\sigma_r,k_r)P(y_0=k_0,y_1=k_1,\ldots,y_r=k_r)$$

where each sum is taken from 1 to n^{nm}. By our construction of Y,

$$E\rho(x,y) = \sum_{k_j}\ldots\sum_{k_r} \rho_0(k_0)H(\sigma_1,k_1)\ldots H(\sigma_r k_r)\int_\Omega \gamma_{k_0}(z_0)\beta_{k_1}(z_1)\ldots$$
$$\beta_{k_r}(z_r)dP$$

$$= \int_\Omega \sum_{k_0}\ldots\sum_{k_r} \rho_0(k_0)H(\sigma_1,k_1)\ldots H(\sigma_r,k_r)\,\gamma_{k_0}(z_0)\beta_{k_1}(z_1)\ldots$$
$$\beta_{k_r}(z_r)dP$$

$$= \int_\Omega \Big(\sum_{k_0} \rho_0(k_0)\,\gamma_{k_0}(z_0)\Big)\Big(\sum_{k_1} H(\sigma_1,k_1)\beta_{k_1}(z_1)\Big)\ldots$$

$$\Big(\sum_{k_r} H(\sigma_r,k_r)\beta_{k_r}(z_r)\Big)dP$$

$$= \int_\Omega \pi_0(z_0)G(\sigma_1,z_1)\ldots G(\sigma_r,z_r)dP \quad \text{by construction}$$

$$= \int_\Omega \pi(x,z)dP = E\pi(x,z) \;.$$

Moreover, if we denote the probability of acceptance of x by (A_D,Y) under random environment sequence y by $p'(x,y)$, we have

$$Ep'(x,y) = E\rho(x,y)\eta^F = (E\rho(x,y))\eta^F$$

$$= (E\pi(x,z))\eta^F = E\pi(x,z)\eta^F = Ep(x,z) \;.$$

Thus, since the expected probability of acceptance remains unchanged, $ET(A_D,y,\lambda) = ET(A,z,\lambda) \;\; \forall\lambda\epsilon[0,1)$. □

Example 4.4.1. Let $A = (\Sigma,S,G,\pi_0,F,E)$, where $\Sigma = \{\sigma_1,\sigma_2\}$, $S = \{s_1,s_2\}$, and $E = [0,1]$. Let G be defined as

$$G(\sigma_1,e) = \begin{pmatrix} e & 1-e \\ 1-e^2 & e^2 \end{pmatrix} \qquad G(\sigma_2,e) = \begin{pmatrix} e^2 & 1-e^2 \\ \sin e & 1-\sin e \end{pmatrix}$$

Clearly, $G(\sigma_1,e) = (e-e^2)\begin{pmatrix}1&0\\1&0\end{pmatrix} + (1-e)\begin{pmatrix}0&1\\1&0\end{pmatrix} + e^2\begin{pmatrix}1&0\\0&1\end{pmatrix} \;\; \forall e\epsilon E$

and
$$G(\sigma_2,e) = \begin{cases} (1-e^2)\begin{pmatrix}0&1\\0&1\end{pmatrix} + \sin e\begin{pmatrix}1&0\\1&0\end{pmatrix} + (e^2-\sin e)\begin{pmatrix}1&0\\0&1\end{pmatrix} \;\; \forall e\epsilon E \text{ such that } e \ge e_0 \\ (1-\sin e)\begin{pmatrix}0&1\\0&1\end{pmatrix} + e^2\begin{pmatrix}1&0\\1&0\end{pmatrix} + (\sin e-e^2)\begin{pmatrix}1&0\\0&1\end{pmatrix} \;\; \forall e\epsilon E \text{ such that } e < e_0 \end{cases},$$

where $e_0 \epsilon E$ such that $e_0 \neq 0$ and $\sin e_0 - e_0^2 = 0$. The approximate value of e_0 is .88.

We now define $a_i'(e)$ for $i=1,\ldots,6$ to be the functions

	$0 \le e < e_0$	$e_0 \le e \le 1$
$a_1'(e)$	e^2	$\sin e$
$a_2'(e)$	$\sin e - e^2$	0
$a_3'(e)$	$e - \sin e$	$e-e^2$
$a_4'(e)$	$1-e$	0
$a_5'(e)$	0	$e^2 - \sin e$
$a_6'(e)$	0	$1-e$

Hence for all $e \epsilon E$, $a_i'(e) \ge 0$ and $\sum_{i=1}^{6} a_i'(e) = 1$. We now define $B_{i\sigma}'(e)$ for $i = 1,\ldots,6$ and $\sigma \epsilon \Sigma$

$$B_{1\sigma_1}'(e) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad B_{1\sigma_2}'(e) = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$$

$$B_{2\sigma_1}'(e) = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \qquad B_{2\sigma_2}'(e) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$B_{3\sigma_1}'(e) = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \qquad B_{3\sigma_2}'(e) = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}$$

$$B_{4\sigma_1}'(e) = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} \qquad B_{4\sigma_2}'(e) = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}$$

$$B_{5\sigma_1}'(e) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad B_{5\sigma_2}'(e) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

$$B_{6\sigma_1}'(e) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad B_{6\sigma_2}'(e) = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} .$$

Clearly, $\sum_{i=1}^{6} a_i'(e) B_{i\sigma}'(e) = G(\sigma,e)$ for all $\sigma \epsilon \Sigma$ and $e \epsilon E$.

CHAPTER V

Implications

5.1 Limit Results for Markov Systems.

Automata theory is generally concerned with input tapes of finite length. Because of its application to Markov systems in environments, we shall now allow input tapes that are countable sequences of elements of Σ. We shall let $\hat{\Sigma}$ denote the set of all countable sequences of elements of Σ. If $\hat{x} = \sigma_1\sigma_2\ldots\epsilon\hat{\Sigma}$, then let $\hat{x}_k = \sigma_1\sigma_2\ldots\sigma_k$ be the tape of the first k symbols of $\hat{x}$.

<u>Definition 5.1.1</u> Let C_n denote the class of all $n\times n$ stochastic matrices for which there is an eigenvector corresponding to the eigenvalue 1 that is a stochastic vector.

Thus, if $A\epsilon C_n$, there exists $v\epsilon V_n$ such that $vA = v$. If A is an ergodic matrix as defined in section 4.2, then $A\epsilon C_n$.

Suppose we have a countable input sequence $\hat{x}\epsilon\hat{\Sigma}$ and an environment sequence $\underline{e}\epsilon E$. The following theorems give conditions for $M(\hat{x}_k,\underline{e})$ to have a limit as $k \to \infty$.

<u>Theorem 5.1.1.</u> Let $A = (\Sigma,S,G,\pi_0,F,E)$ be an ADE. Let $\hat{x} = \sigma_1\sigma_2\ldots\epsilon\hat{\Sigma}$ and $\underline{e} = (e_0,e_1,\ldots)\epsilon E^\infty$. If

1. A is ever-ergodic at $\underline{e}$
2. $G(\sigma_i,e_i)\epsilon C_n$ for all $i = 1,2,\ldots$
3. $\sum_{i\epsilon j} ||v_{i+1}-v_i|| < \infty$, where $v_i\epsilon V_n$ such that $v_i\, G(\sigma_i,e_i) = v_i$,

then there exists $M(\hat{x},\underline{e})$ such that $\partial(M(\hat{x},\underline{e})) = 0$ and

$\lim_{k\to\infty} ||M(\hat{x}_k,\underline{e}) - M(\hat{x},\underline{e})|| = 0.$

Proof. The space of n-dimensional vectors with norm $||v|| = \sum_{i=1}^{n} |v_i|$ is complete. So by condition 3, there exists an n-dimensional vector v such that $||v_i - v|| \to 0$ as $i \to \infty$. Since V_n is a closed subset of the space of n-dimensional vectors and since each $v_i \in V_n$, then $v \in V_n$. Let $V_i \in M_n$ be the matrix each of whose rows is v_i. Similarly, let $M(\hat{x},\underline{e})$ be the matrix each of whose rows is v. Clearly, $\partial(M(\hat{x},\underline{e})) = 0$. Note that for any $A \in M_n$ with identical rows, $||A|| = ||a||$, where a is any row of A.

For notational convenience, we shall denote $\prod_{i=i_1}^{i_2} G(\sigma_i,e_i)$ by $G_{i_1}^{i_2}$ Thus,

$$\begin{aligned} ||M(\hat{x}_k,\underline{e})-M(\hat{x},\underline{e})|| &= ||G_1^k - M(\hat{x},\underline{e})|| \\ &\le ||G_1^k - V_k|| + ||V_k - M(\hat{x},\underline{e})|| \\ &\le ||G_1^k - V_j G_j^k|| + ||V_j G_j^k - V_k|| + ||V_k - M(\hat{x},\underline{e})|| . \end{aligned}$$

Note that
$$\begin{aligned} V_j G_j^k &= V_j G_{j+1}^k \\ &= V_k G(\sigma_k,e_k) + \sum_{i=j}^{k-1} (V_i - V_{i+1}) G_{i+1}^k . \end{aligned}$$

But $V_k G(\sigma_k,e_k) = V_k$. Therefore,

$$\begin{aligned} ||V_j G_j^k - V_k|| &= ||\sum_{i=j}^{k-1} (V_i - V_{i+1}) G_{i+1}^k|| \\ &\le \sum_{i=j}^{k-1} ||v_i - v_{i+1}||\,\partial(G_{i+1}^k) \quad \text{by lemma 3.3.3} \\ &\le \sum_{i=j}^{k-1} ||v_i - v_{i+1}|| \le \sum_{i=j}^{\infty} ||v_{i+1} - v_i||. \end{aligned}$$

It follows from condition 3 that for $j \geq N(\varepsilon)$ this last expression can be made less than $\varepsilon/3$.

Also, $||G_1^k - v_j G_j^k|| = ||(G_1^{j-1} - v_j)G_j^k|| \leq ||G_1^{j-1} - v_j|| \partial(\prod_{i=j}^{k} G(\sigma_i, e_i))$ by lemma 3.3.3. Since A is ever-ergodic at $\underline{e}$, for fixed j there exists $N_1(\varepsilon)$ such that the above expression can be made less than $\varepsilon/3$ for $k \geq N_1(\varepsilon)$.

And finally $||v_k - M(\hat{x}, \underline{e})|| = ||v_k - v||$. By condition 3, this expression can be made less than $\varepsilon/3$ for $k \geq N_2(\varepsilon)$.

Hence, $||M(\hat{x}_k, \underline{e}) - M(\hat{x}, \underline{e})||$ can be made less than ε for sufficiently large k. □

Note that theorem 5.1.1 remains true if we replace the ever-ergodic condition by the weaker condition: for $\hat{x} = \sigma_1\sigma_2 \ldots \epsilon \hat{\Sigma}$ and $\underline{e} \epsilon E^{\infty}$, given any $\varepsilon > 0$, there exists $N(\varepsilon)$ such that $\partial(\prod_{i=j}^{j+N(\varepsilon)} G(\sigma_i, e_i)) < \varepsilon \quad \forall_j \epsilon J_+$. We may also weaken condition 2 to require that $G(\sigma_i, e_i) \notin C_n$ for only a finite number of times.

Theorem 5.1.2. Let $A = (\Sigma, S, G, \pi_0, F, E)$ be an ADE. Let $\hat{x} = \sigma_1\sigma_2 \ldots \epsilon \hat{\Sigma}$ and $\underline{e} = (e_0, e_1, \ldots) \epsilon E^{\infty}$. If

1. $G(\sigma_i, e_i) \epsilon C_n$ for all $i = 1, 2, \ldots$
2. there exists $v \epsilon V_n$ such that $||v_i - v|| \to 0$ as $i \to \infty$, where $v_i \epsilon V_n$ and $v_i G(\sigma_i, e_i) = v_i$
3. there exists a constant γ such that $\sum_{j=1}^{k} \partial(G_j^k) \leq \gamma$ for all $k = 1, 2, \ldots,$

then there exists $M(\hat{x},\underline{e})$ such that $\partial(M(\hat{x},\underline{e})) = 0$ and $\lim_{k\to\infty} ||M(\hat{x}_k,\underline{e}) - M(\hat{x},\underline{e})|| = 0$.

<u>Proof</u>. Construct V_i and $M(\hat{x},\underline{e})$ as in theorem 5.1.1. Since $\sum_{j=1}^{k} \partial(G_j^k) \le \gamma$, we can make $\partial(G_{k-1}^k)$ less than ε for sufficiently large k. But by lemma 3.3.2, $\partial(G_j^k) \le \partial(G_j^{k-1})\partial(G_{k-1}^k) \le \partial(G_{k-1}^k)$. Hence, $\partial(G_j^k) \to 0$ as $k \to \infty$ for all $j\in J_+$.

Now,

$$||M(\hat{x}_k,\underline{e})-M(\hat{x},\underline{e})|| = ||G_1^k-M(\hat{x},\underline{e})||$$
$$\le ||G_1^k-V_k|| + ||V_k-M(\hat{x},\underline{e})||.$$

But,

$$||G_1^k-V_k|| \le ||(G(\sigma_1,e_1)-V_1)G_2^k + \sum_{i=2}^{k}(V_{i-1}-V_i)G_i^k||$$
$$\le ||G(\sigma_1,e_1)-V_1||\partial(G_2^k) + \sum_{i=2}^{k}||V_{i-1}-V_i||\partial(G_i^k) .$$

But $||G(\sigma_1,e_1)-V_1|| \le 2$ and $\partial(G_2^k) \to 0$ as $k \to \infty$. Hence, we can make $||G(\sigma_1,e_1)-V_1||\partial(G_2^k)$ less than $\varepsilon/4$ for $k\ge N_1(\varepsilon)$.

By condition 2, $||V_{i-1}-V_i|| \to 0$ as $i \to \infty$. So given $\varepsilon > 0$, there exists $\mathcal{k}(\varepsilon) > 1$ such that $||V_{i-1}-V_i|| \le \varepsilon/8\gamma$ for $i\ge\mathcal{k}(\varepsilon)$. For all $k\ge\mathcal{k}(\varepsilon) + 1$, $\sum_{i=\mathcal{k}+1}^{k}||V_{i-1}-V_i||\partial(G_i^k) \le \varepsilon/8$. Also, we have seen earlier that $\partial(G_i^k) \to 0$ as $k \to \infty$ for all i. Hence, given $\varepsilon > 0$, there exists $N_2(\varepsilon)$ such that $\sum_{i=2}^{\mathcal{k}} \partial(G_i^k) \le \varepsilon/16$. But $||V_{i-1}-V_i|| \le 2$; so we have $\sum_{i=2}^{\mathcal{k}}||V_{i-1}-V_i||\partial(G_i^k) \le \varepsilon/8$. Hence, for $k\ge N_2(\varepsilon)$ we have $\sum_{i=2}^{k}||V_{i-1}-V_i||\partial(G_i^k) \le \varepsilon/4$.

Hence, $||G_1^k-V_k|| \le \varepsilon/2$ for $k \ge \max(N_1(\varepsilon),N_2(\varepsilon))$. By condition 2, $||V_k-M(\hat{x},\underline{e})|| \le \varepsilon/2$ for $k \ge N_3(\varepsilon)$. So given any $\varepsilon > 0$, $||M(\hat{x}_k,\underline{e})-M(\hat{x},\underline{e})|| \le \varepsilon$ for $k \ge \max(N_1(\varepsilon),N_2(\varepsilon),N_3(\varepsilon))$. □

We may again weaken condition 1 to require that $G(\sigma_i, e_i) \notin C_n$ for only a finite number of times.

By these theorems under the appropriate conditions, we have demonstrated a consistent way to define the transition function of $\hat{x} \epsilon \hat{\Sigma}$ under the environment sequence $\underline{e} \epsilon E$.

5.2 Applications.

Consideration of automata in environments has allowed us to generalize the finite-state system model to incorporate the various external factors which may influence the behavior of the system. A probabilistic automaton may be constructed to account for the effects of the environments only when the system can be simulated by a probabilistic automaton. Thus, the interesting case for application of the new automaton in environments model is the case that the system cannot be simulated by a probabilistic automaton. In such a case the environment cannot be reduced to a finite set by the induced pseudo-metric or does not have the appropriate semi-group properties. Moreover, we may incorporate a random structure to the occurrence of the environments. This formulation is particularly useful where there is no control over the occurrence of the environments and we can only make probabilistic assumptions about the occurrences of the environments.

A traditional motive for automata models has been the behavior of digital computers. These systems are indeed situated in environments which affect their mechanical and electrical properties and, hence, their output. Lightning storms, power fluctuations, earth tremors, and late evening personnel are often blamed as the source of errors. We admit that our sophisticated technology has produced digital computer systems which rarely have transitions other than the ones anticipated.

However, there are many less reliable systems which have probabilistic transitions affected to a greater extent by the configuration of the environment. Information transmission systems may be modeled by automata in environments. The inclusion of the environments will allow us to consider the various types of noise which may occur nonhomogeneously during a transmission.

Another application for automata in environments is modeling learning and pattern recognition. The input represents a stimulus which will encourage some desired behavior. The environment may be taken to be some external circumstances. An alternative interesting possibility for the environments would be their use as records of penalties for previous behavior. The transition functions may then incorporate this history to modify the transition probabilities due to reinforcement of desired behavior.

Decision procedures may be modeled by automata with the input symbols representing the data collected and with the environments added to account for the cost of various actions of the system or to account for certain external phenomena which may be controlled or subject to random variation.

See Ash [1], Chandrasekaran and Shen [5], Krylov and Tsetlin [9], Tsetlin [19], Varshavskii and Vorontsova [21], and Yakowitz [22] for references appropriate to this discussion.

5.3 Conclusions and Summary.

We have taken the viewpoint that every system is situated within environments which affect its behavior. Accordingly, we have generalized the Rabin-Paz model of a probabilistic automaton to account for the effects of the environments. The initial state distribution is no longer considered as a constant, but rather as a function of the initial configuration of the environments. Similarly, the subsequent probabilistic transitions become functions not only of the present input, but also the present configuration of the environment.

We first discussed the case of automata with deterministic environment rules $\underline{e} \epsilon E^{\infty}$. That is, $\underline{e} = (e_0, e_1, \ldots)$ specifies the configuration of the environment for the initial distribution and the subsequent transitions. Such a system is called an automaton in deterministic environments (ADE). We might have considered the environments as a nuisance, but actually they enable us to construct a model with greater capacity and flexibility.

It is not always necessary to consider the ADE model for systems with environments. Theorem 2.4.1 demonstrates that a system may be simulated by a probabilistic automaton when the environment set is finite. This involves construction of a probabilistic automaton over a larger input alphabet which incorporates the role of the environments. Theorem 2.4.1 led to further simulation results when the environment set can be reduced to a finite set or when the environment has the appropriate semi-group structure preserved by the system. In section 2.3 we find that there are sets of tapes which can be defined by some

ADE but which cannot be defined by any probabilistic automaton. However, it is possible that such ADE's may be simulated by probabilistic automata.

In Chapter III we tackled the stability problem. It is important because of possible perturbations of the environments or some impreciseness of the measurements of the environments. In section 3.3 we discussed sufficient conditions for an ADE to be s-stable for general perturbations of the environments or when the perturbations are measured with respect to a natural metric. We needed to extend the notion of quasidefiniteness to derive the ever-ergodic condition for the proofs. Theorem 3.4.1 and corollary 3.4.2 give sufficient conditions for a-stability without requiring s-stability or the ever-ergodic condition. In section 3.5 we explored the application of the stability results when round off procedures were used in measuring the configuration of the environment.

Finally, we considered the case of automata operating in random environments (ARE). In section 4.1 we defined the new model and discussed procedures to estimate the expected properties of the system. In section 4.2 we discovered that when the random environments are independent then the properties in the mean can be found by consideration of a derived ADE. When $\Sigma = \{\sigma\}$ and $z = (z_0, z_1, \ldots)$ is a sequence of IID random variables, this derived ADE is actually a probabilistic automaton; moreover, we found sufficient conditions to compute the asymptotic moments of the components of $M(\sigma^N, z)$.

We observed the dual nature of the randomness of automata in random environments. However, theorem 4.4.1 demonstrates that there is a mean equivalent canonical representation which eliminates the randomness due to probabilistic transition; for any ARE, we can find another ARE with finite environment set, deterministic assignment of the initial state, and deterministic transitions which has a state distribution equivalent in the mean. Since the ADE is a special case of the ARE, we find that every ADE has a canonical representation as an ARE with the above properties.

There are some interesting and still open problems which merit further investigation. At various points we have assumed that we could determine whether or not a certain cut-point was isolated. Beyond corollary 3.4.3 we have not given any conditions to locate isolated cut-points. In certain regions of the environment set it may be that the transition functions $G(\sigma,\circ)$ have certain components which remain 0 for the entire region. In such cases there may be weaker conditions needed for stability. There are also many opportunities for application of the new models to various systems as suggested in section 5.2 .

BIBLIOGRAPHY

[1] Ash, R. (1967). *Information Theory.* Wiley, New York.

[2] Bachman, G. and Narici, L. (1966). *Functional Analysis.* Academic Press, New York.

[3] Bellman, R. (1954). Limit theorems for non-commutative operations I. *Duke Mathematical Journal 21*, 491-500.

[4] Booth, T.L. (1967). *Sequential Machines and Automata Theory.* Wiley, New York.

[5] Chandrasekaran, B. and Shen, D.W.C. (1969). Stochastic automata games. *IEEE Transactions on Systems Science and Cybernetics 5*, 145-149.

[6] Feller, W. (1968, 1966). *An Introduction to Probability Theory and its Applications, 1,2.* Wiley, New York.

[7] Flachs, G.M. (1967). Stability and cut points of probabilistic automata. Ph.D. dissertation, Michigan State University.

[8] Knast, R. (1969). Continuous-time probabilistic automata. *Information and Control 15*, 333-352.

[9] Krylov, V. Yu. and Tsetlin, M.L. (1963). Games between automata. *Automation and Remote Control 24*, 975-987.

[10] Madsen, R.W. and Isaacson, D.L. (1973). Strongly ergodic behavior for non-stationary Markov processes. *Annals of Probability 1*, No. 2, 329-335.

[11] Minsky, M.L. (1967). *Computation: Finite and Infinite Machines.* Prentice Hall, Englewood Cliffs, N.J.

[12] Paz, A. (1965). Definite and quasi-definite sets of stochastic matrices. *Proceedings of the American Mathematical Society 16*, 634-641.

[13] Paz, A. (1966). Some aspects of probabilistic automata. *Information and Control 9*, 26-60.

[14] Paz, A. (1970). *Probabilistic Automata.* Academic Press, New York.

[15] Perles, M., Rabin, M.O., and Shamir, E. (1963). The theory of definite automata. *IEEE Transactions on Electronic Computers EC-12*, 233-243.

[16] Rabin, M.O. (1963). Probabilistic automata. *Information and Control 6*, 230-245.

[17] Rabin, M.O. and Scott, D. (1959). Finite automata and their decision problems. *I.B.M. Journal of Research and Development 3*, 114-125.

[18] Salomaa, A. (1969). *Theory of Automata.* Pergamon Press, New York.

[19] Tsetlin, M.L. (1961). On the behavior of finite automata in a random media. *Automation and Remote Control 22*, 1345-1354.

[20] Turing, A.M. (1936). On computable numbers, with an application to the entscheidungsproblem. *Proceedings of the London Mathematical Society 42*, 230-265.

[21] Varshavskii, V.I. and Vorontsova, I.P. (1963). On the behavior of stochastic automata with a variable structure. *Automation and Remote Control 24*, 353-360.

[22] Yakowitz, S. (1974). Multiple hypothesis testing by finite memory algorithms. *Annals of Statistics 2*, No. 2, 323-336.

INDEX

NOTATION