Requests for additional copies by agencies of the Department of Defense, their contractors, and other Government agencies should be directed to the:

Defense Documentation Center (DDC)
Cameron Station
Alexandria, Virginia 22314

Department of Defense contractors must be established for DDC services or have their "need-to-know" certified by the cognizant military agency of their project or contract.

All other persons and organizations should apply to the:

Clearinghouse for Federal Scientific
and Technical Information (CFSTI)
Sills Building
5285 Port Royal Road
Springfield, Virginia 22151

THE UNIVERSITY OF MICHIGAN

COLLEGE OF LITERATURE, SCIENCE, AND THE ARTS
Communication Sciences Program

Technical Report

EQUIVALENCES BETWEEN PROBABILISTIC SEQUENTIAL MACHINES

Carl V. Page

ORA Projects 03105, 07363, and 07537

Supported by:

DEPARTMENT OF THE NAVY
OFFICE OF NAVAL RESEARCH
CONTRACT NO. Nonr-1224(21)
WASHINGTON, D.C.

and

U. S. ARMY RESEARCH OFFICE (DURHAM)
GRANT NO. DA-ARO(D)-31-124-G665
DURHAM, NORTH CAROLINA

and

NATIONAL INSTITUTES OF HEALTH
PUBLIC HEALTH SERVICE
GRANT NO. GM-12236-02
BETHESDA, MARYLAND

administered through:

OFFICE OF RESEARCH ADMINISTRATION ANN ARBOR

November 1965

This report was also a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in The University of Michigan, 1965.

ERRATA

Page	Line (B--from bottom T--from top)	Corrections
23	Figure 1.1 at state ③	5 should be deleted and replaced with 1.
26	Figure 1.2	Arrowheads left out ③ ⟶ ⑤ and ⑦ ⟶ ②.
37	Equation (2)	$\mu_i^A(x) = \mu_i^A(y) \Longleftrightarrow IA(x)(F^j) =$
39	1 T	... all symbols σ and ...
41	Figure 2.1	Arrowheads left out $\boxed{1}$ ⟶ ② and ⑦ ⟶ ②
79	Figure 3.1	A: arrowhead ① ⟶ ②.
145	4 T	$b_{\ell-1}$
168	8 B	Period after also.

ACKNOWLEDGEMENTS

I would like to express my gratitude and appreciation to Professors John Holland, Arch Naylor, Betzalel Peleg, Eugene Lawler and Dr. Richard Karp for serving on my doctoral committee. I would like to thank Steve Hedetniemi and Dr. Richard Arnold for their many important questions during the formative stage of this work.

Many aspects of this work would not have been developed without the interdisciplinary nature of the Communication Sciences Program. I would like to express my gratitude to the many members of the University of Michigan faculty who have been responsible for the existence of such a program.

My wife Gloria deserves special thanks for her loyalty and understanding without which this work might not have been possible.

Finally, I would like to acknowledge the aid of the members of the Logic of Computers Group, especially Kuniko Misawa, who typed the final manuscript and several earlier drafts, and the financial support of the Office of Naval Research (Nonr-1224(21)), the U.S. Army Research Office, Durham (DA-ARO(D)-31-124-G665) and the National Institutes of Health (GM 12236-02).

RESEARCH PROGRESS REPORT

Title: "Equivalences Between Probabilistic Sequential Machines," by C.V. Page, The University of Michigan Technical Report

Background: The Logic of Computers Group of the Communication Sciences Program of The University of Michigan is investigating the application of logic and mathematics to the theory of the design of computing automata.

Condensed Report Contents: This report introduces new definitions of behavioral equivalence and shows relationships among the various notions of behavioral equivalence between probabilistic machines. Four basic problems are discussed for several different behavioral equivalences between machines. In what follows we use the symbol "$\equiv$" as a variable to denote one of the several behavioral equivalences considered in this report. Given an arbitrary probabilistic sequential machine A: (1) Is there an input-state calculable machine A' (a machine with deterministic switching and random outputs) such that $A \equiv A'$? (2) What are the machines A' with minimal number of states such that $A \equiv A'$? (3) How does one obtain all stable modifications of A, i.e., all machines A' such that the switching probabilities of A' and A differ but $A \equiv A'$? (4) Is there a finite bound on the length of experiments required to establish whether $A \equiv A'$ holds? The four basic problems are solved completely for some equivalences between machines and are left open for other equivalences. Some applications are made to optimal control problems.

For Further Information: The complete report is available in the major Navy technical libraries and can be obtained from the Defense Documentation Center. A few copies are available for distribution by the author.

TABLE OF CONTENTS

TABLE OF CONTENTS (Concluded)

LIST OF TABLES

LIST OF ILLUSTRATIONS

LIST OF SYMBOLS AND TERMS

A, A', B, B'	Probabilistic sequential machines.
$A^{(i)}$	The i^{th} machine in a sequence of machines.
$\longrightarrow$	$A \longrightarrow B$ denotes the machine formed by plugging the output of A into the input of B. Def. 3.1, 43.
$\longleftarrow$	$(\ldots(S_1 \longleftarrow S_2) \longleftarrow \ldots S_n)$ state reduction in sequence. Def. 8.4, 124.
$\oplus$	$A \oplus A'$: the abstract join of machines A and A'. Def. 4.4, 58.
$\equiv$	A behavioral equivalence variable denoting a set of behavioral equivalences fixed in context. Subscripts denote particular behavioral equivalences.
$\equiv(S)$	State equivalence of machines variable. Def. 8.1, 121.
$\equiv$-stable perturbation	Def. 6.1, 90.
$\equiv_E$	Expectation equivalence. Def. 1.2, 9.
$\equiv_N$	N-moment equivalence. Def. 2.5, 30.
$\equiv_D$	Distribution equivalence. Def. 2.2, 28.
$\equiv_T$	Tape acceptance equivalence. Def. 3.3, 43.
$\equiv_I$	Indistinguishability. Def. 3.9, 49.
$\equiv_r$	Where r is an integer or integer valued variable: $\equiv_N$ with $N = r$.
$\equiv_{\lambda IE}$	λ-input event equivalence. Def. 9.2, 139.

LIST OF SYMBOLS AND TERMS (Continued)

$\equiv_{IE}$	Input event equivalence. Def. 9.3, 139.
$\equiv_{\lambda IO}$	λ-input-output equivalence. Def. 9.5, 142.
$\equiv_{IO}$	Input-output equivalence. Def. 9.6, 142.
$\sim$, $\overset{A}{\sim}$	An equivalence of distributions (or states) (A variable equivalence associated with the behavioral equivalence variable $\equiv$.) Def. 8.2, 122.
$\sim_E$, $\overset{A}{\sim}_E$	Equivalence of distributions (or states) of machine A associated with $\equiv_E$. Def. 4.2, 56.
$\sim_N$, $\overset{A}{\sim}_N$	Equivalence of distributions (or states) associated with $\equiv_N$. Def. 8.2, 122.
$\sim_I$, $\overset{A}{\sim}_I$	Equivalence of distributions (or states) associated with $\equiv_I$. Def. 8.2, 122.
$\overset{A}{\sim}_T$	Equivalence of distributions (or states) for $\equiv_T$. Def. 8.6, 130.
$\sim_{KE}$, $\overset{A}{\sim}_{KE}$	Equivalence of distributions (or states) for all strings of length K or less for $\equiv_E$. Def. 3, 56.
$A(\sigma)$	Symbol matrix for machine A and input σ. Def. 0.1, 2.
$A(y_i/\sigma)$	Conditional probability transition matrix for input symbol σ and output symbol y_i. Def. 3.7, 47.
$A[\Pi(S_i)]$	Reduced machine in which state S_i is replaced by distribution Π. Def. 8.3, 123.
$\langle \quad \rangle$	Meaning depends on context, i.e. denotes a tuple of things as in machine definition: (1) $A = \langle r, I, \Sigma, A(0), \ldots, A(k-1), F, 0_A \rangle$ Def. 0.1, 1. (2) The machine A started in initial distribution I: (a variant of (1)): $\langle I, A \rangle$ Def. 6.6, 103.

	(3) $\langle V \rangle$ where V is a set of vectors means the span of V. Def. 1.11, 17. (4) $\langle A(\Sigma^*),\cdot\rangle$ means the monoid of matrices $A(x) : x \in \Sigma^*$ with matrix multiplication, 90.
$\#\beta$	Where β is a finite set; the number of elements of β.
Backward determinism	Def. 7.2, 107.
Balance property	Def. 5.8, 77.
$C(A,\lambda)$	Set of critical tapes for machine A with cutpoint λ. Def. 8.7, 132.
$C_{ij}(k)$	Cost of transition from state S_i to S_j caused by input symbol k, 144.
Complete set of invariants for an equivalence	Def. 4.1, 52.
Canonical distribution pair	Def. 5.10, 85.
$D(A,\lambda)$	Set of critical distributions of machine A with cutpoint λ. Def. 8.8, 133.
$\dim(L_j)$	Dimension of space L_j.
δ	Separation of isolated cutpoint, Def. 5.1, 66 or parameter in stability near the cutpoint, Def. 5.9, 81.
E	Expectation in the sense of probability theory, 29.
$E_A(x)$	Expectation of output of input sequence x of machine A. Def. 1.1, 9.

LIST OF SYMBOLS AND TERMS (Continued)

$E(n)$	n: dimensional identity matrix.
E_x	Error matrix of input string x, 91.
e^i	The i^{th} stochastic vector in n-dimensional space in an indexed set of such stochastic vectors.
F, F^A	Output vector of machine A. Def. 0.1, 2.
(F^k)	Column vector such that the i^{th} entry is just the i^{th} entry of F raised to the k^{th} power. Def. 2.3, 28.
ϕ	The empty set.
$h_r(i)$, $h_{k\#}(i)$	Index merging function. Def. 8.5, 123.
H_x	Error matrix of input string x (sometimes written E_x), 91.
H_σ	An invariant error matrix for symbol σ. Def. 6.2, 97. An eigenerror matrix for symbol σ. Def. 6.3, 98.
I, I_A, I^A	Initial probability distribution over the states of machine A.
I_i	Probability of being in state S_i initially for machine A.
$IO_y(A,\lambda)$	Input-output event for output sequence y of machine A with cutpoint λ. Def. 9.4, 141.
Isolated cutpoint	Def. 5.1, 66.
$K(A,F)$	Where A is an n x n matrix and F an n dimensional column vector is the set of all column vectors ϕ

	such that $A\phi \geq F$, 109.
Kern.(T_r)	Where T_r is a linear transformation is the set of all vectors α such that $\alpha \cdot T_r = 0$.
$K^*(A,F)$	The polar set of $K(A,F)$. Def. 7.3, 117.
k_A	The value of the Rabin-Paz bound for machine A, 86.
Λ	Symbol with identity symbol matrix, i.e. $A(\Lambda) = E(n)$.
Λ_Σ	Input semigroup identity: $\Lambda_\Sigma \cdot x = x \quad \forall x \in \Sigma^*$, 4.
$\ell g.(x)$	The length of input sequence x, 4.
$\ell g.(y)$	The length of input (or output) sequence y.
λ	If real number, cutpoint. Remark 3, 3. (In Chapters 3 and 4 sometimes used as an initial distribution).
$M(S,\sigma)$	Transition function for a deterministic machine M.
Minority property of pairs of machines	Def. 5.7, 77.
Monoid	A semigroup with identity.
n, n_A	The number of states of machine A.
n_F	The independence number of machine A (with output vector F). Def. 4.5, 60.
n_D	Number of states of deterministic machine D.
Nonsingular machine	pg. 108.

LIST OF SYMBOLS AND TERMS (Continued)

$0, 0_A$	Output function for machine A. Def. 0.1, 2.
$0_A^*(x)$	The output random variable for input string x in machine $A = 0_A(rp_A(x))$. Def. 2.1, 27.
$0_A^R(S)$	Random output attached to machine A depending on state S.
$\mu_i^A(x)$	i^{th} central moment of output for input string x of machine A. Def. 2.3, 29.
P( ), Prob.( )	The probability of the event in the parenthesis.
$P_\pi^A(y/x)$	The probability that output sequence y will be observed given input sequence x in machine A started from initial distribution π. Def. 3.5, 46 and Remark 3.7, 48.
$\mathbb{P}$	Permutation of the integers.
π	Depending on contex either (1) A permutation of the integers. (2) An initial distribution of machine A. (3) A policy (Chapter 9). pg. 145(13).
π_x	Permutation matrix depending on string x. Theorem 7.3, 111.
Π	Partition on the set of input sequences Σ^*.
$\pi(S_i)$	State S_i is replaced by distribution π. Def. 8.3, 123.
$\pi^*(y/x)$	Terminal distribution for input sequence x and output sequence y. Def. 3.8, 41.
$\pi(1)$, $\pi(2)$, $\pi(3)$	Extreme points of Example 8.1, 134.
$\pi'(1)$, $\pi'(3)$	Projections of $\pi(1)$, and $\pi(3)$. Fig. 8.2, 134.

$q(y/ux,v)$	The probability of observing output string y from machine A given that input string u produced string v as output and that x is the next input sequence, 141.
$q^{\pi}_{S_i}(x,t)$	Probability that input string x occurs at time t because of policy π for a machine (specified in context) with initial state S_i, 158.
R	Depending on context either (1) The set of real numbers. (2) An arbitrary relation (usually a right congruence relation) on input strings.
$R[x_1]$	The equivalence class of x_1 in equivalence relation R. Def. 1.4, 11. (Note that $\overset{A}{\sim}_E[S_{1\#}]$ means the equivalence class of $S_{1\#}$ in equivalence relation $\overset{A}{\sim}_E$.)
R_E	Reduction relation for $\equiv_E$. Def. 1.3, 11.
R_N	Reduction relation for $\equiv_N$. Def. 2.5, 34.
$\sim R$, $\not R$	Negation of the relation R.
R_β	Where β is a set of input strings: the *right congruence relation induced by* β. i.e. $x_1 R_\beta x_2 \Longleftrightarrow [x_1 z \in \beta \Longleftrightarrow x_2 z \in \beta]\ \forall z \in \Sigma^*;\ x_1 x_2 \in \Sigma^*$
$R_S(A,A')$	The *stability congruence relation* (right congruence relation induced by S(A,A')). Def. 5.4, 74.
$rp_A(x)$	Response of machine A to input x; i.e. the state of the machine after input x. Def. 1.6, 13.
Relation R *refines* a set β	Def. 1.5, 12.
S(A,A')	*Stable set* of input tapes for machines A and A'. Def. 5.3, 74.

LIST OF SYMBOLS AND TERMS (Continued)

S, S_A	Set of state vectors of machine A.
S_i	State $S_i = (\overbrace{0, \ldots, 0, 1}^{i}, 0, \ldots, 0)$.
S^+	Convex closure of the state set, i.e. all probability distributions for an n state machine. Def. 1.8, 15.
Σ	Input alphabet set of machine A, 2.
σ	An arbitrary input symbol,
Σ^*	Set of all finite input sequences.
Σ^k, $(\Sigma)^k$	Set of all input sequences of length k.
Stability near the cutpoint	Def. 5.9, 81.
$T(A,\lambda)$	The set of input tapes accepted by probabilistic sequential machine A with cutpoint λ. Remark 3, 3.
$T_y(A,\lambda)$	y-input event. Def. 9.1, 139.
$T(D)$	Set of tapes accepted by deterministic machine D, i.e. those tapes which lead to designated final states.
$T(A) = A'$	Stability transformation of machine A giving machine A', 90. (Also written $T : A \longrightarrow A'$)
$T(R)$	Where R is a right congruence relation: the quotient machine Σ^*/R, 11.
$T(A^{(1)}, \ldots, A^{(r)}, \lambda)$	Experiment on the set of machines $A^{(1)}, \ldots, A^{(r)}$ using the input set T. Def. 5.5, 76.

LIST OF SYMBOLS AND TERMS (Concluded)

Tape indistinguishability	(for a class of machines) Def. 3.4, 44.
U	The n dimensional column vector of ones, $\begin{pmatrix}1\\ \cdot\\ \cdot\\ \cdot\\ 1\end{pmatrix}$, or a subspace depending on context.
$V(A)$	$\{IA(x): x \varepsilon \Sigma^*\}$, the set of distributions accessible by A, 15.
v_σ	An eigenstate for symbol σ. Def. 6.5, 103.
$W(S_4(S_2))$	Wedge shaped volume in S^+ affected by mapping S_2 onto S_4, 135.
x	Arbitrary input string ($x \varepsilon \Sigma^*$).
$X_i(\pi)$	Expected cost of policy π from initial state S_i, 145.
$X_i(\pi, t)$	Expected cost at time t of policy π started from state S_i.
$Y = \bigcup_{i=1}^{n} \{F_i\}$	Set of output codes for a machine A. Remark 2, 3.
Y^*	Set of all finite sequences of symbols from Y.
$(Y)^r$	Set of all sequences of Y of length r.
$Z(A)$	Set of singular switching matrices of A.

CHAPTER 0

INTRODUCTION

The notion of behavioral equivalence is a fundamental part of the study of automata theory. Two superficially different definitions of behavioral equivalence occur in the literature for deterministic machines. One, discussed by Burks [1961], which we will write $\equiv_I$, calls two machines behaviorally equivalent if they define the same function from input strings to output strings. The other, part of Rabin-Scott [1959] automata theory, which we will write $\equiv_T$, calls two machines behaviorally equivalent if they accept the same set of tapes. For any deterministic machines D and D' with the same input alphabets and binary outputs, D $\equiv_I$ D' holds if and only if D $\equiv_T$ D holds. However, for the generalizations of $\equiv_I$ (Carlyle [1961]) and $\equiv_T$ (Rabin [1964]) to probabilistic machines A and A', we observe that A $\equiv_T$ A' does *not* imply A $\equiv_I$ A'.

This thesis is concerned with properties of several behavioral equivalences between probabilistic machines. In order to gain insight into the kinds of equivalences which will be studied, two models of probabilistic sequential machines will be presented later in this chapter.

0.1 THE CONCEPT OF PROBABILISTIC SEQUENTIAL MACHINE

By a probabilistic sequential machine is meant a system which satisfies one of the following definitions:

Definition 0.1: A (Moore-type) probabilistic sequential machine A is a system $A = \langle n, I, S, \Sigma, A(\sigma) : \sigma \in \Sigma, F, 0 \rangle$

where

n: a natural number, the number of states

I: an n-dimensional stochastic vector, the initial distribution

S: set of state vectors = $\{S_1 = (1, 0, \ldots, 0), \ldots, S_n = (0, \ldots, 0, 1)\}$

Σ: alphabet set. Usually $\Sigma = \{0, 1, 2, \ldots, k-1\}$

$A(\sigma)$: $\sigma \varepsilon \Sigma$ n x n switching matrix for input symbol σ. $A(\sigma)_{\ell m}$ is the probability of a transition from state ℓ to state m via symbol σ. We frequently call $A(\sigma)$ a "symbol matrix".

F: output vector, a n-dimensional column vector whose entries are real numbers. F_i is the output from state S_i.

O: output function $O(S_i) = S_i \cdot F = F_i$

When clear from context, some of the parts of the formal definition may be omitted for the definition of a particular machine.

Definition 0.2: A (Mealy-type) probabilistic sequential machine.

$A = \langle n, I, S, \Sigma, A(\sigma) : \sigma \varepsilon \Sigma, W, P\rangle$

where n, I, S, Σ, $A(\sigma) : \sigma \varepsilon \Sigma$ are as in Definition 0.1 and where the output function P satisfies

$$P(S_i, \sigma_j) = W_{ij} \qquad S_i \varepsilon S, \sigma_j \varepsilon \Sigma$$

It is an easy matter to show that Definition 0.1 and Definition 0.2 are equivalent in the following sense: For every Moore-type probabilistic sequential machine there is a Mealy-type sequential machine which is indistinguishable in an intuitive sense and vice-versa. Consequently, we will be concerned only with the properties of Moore-type probabilistic sequential machines which from now on will be called "probabilistic sequential machines" or just "machines".

There seem to be many systems like probabilistic sequential machines in fields of study not historically associated with automata theory. Brains and Svechinsky discuss a system like Definition 0.1 in their paper "Matrix Structure in Simulation of Learning" [1962]. If one takes the cartesian product of machines of Definition 0.2, one gets Markov processes with rewards and alternatives as studied in sequential decision theory as presented by Howard [1960]. Matrix games as discussed by Thrall [1957] can be considered as instances of Definition 0.1 in which I and F are strategy vectors and game matrix A(x) is defined by a string x. A simple correspondence shows that the noisy discrete channel of Shannon [1948] is equivalent to the system of Definition 0.2. Someday probabilistic sequential machines may become a unifying concept, organizing and providing results for diverse fields.

Probabilistic sequential machines were devised as slight generalizations of the notion of Rabin [1964] of probabilistic automata. If one restricts I to elements of S and $F_i = 0$ or 1 for $i = 1, 2, \ldots, n$ then Definition 0.1 defines probabilistic automata. Following Rabin, we describe how the transition matrices for sequences of inputs are generated by the symbol matrices.

<u>Remark 1</u>: Let $x = i_1 \ldots i_r$, $i_j \in \Sigma$, $j = 1, \ldots, r$. Then $A(x) = A(i_1) \ldots A(i_r)$ i.e., the switching matrix for a string x is found by multiplying the matrices for the symbols of x together in order.

<u>Remark 2</u>: Sometimes the real numbers which are the outputs of a probabilistic sequential machine will be regarded as codes for symbolic outputs.

<u>Remark 3</u>: The <u>expectation of output</u> for input string x of machine A is just $E_A(x) = IA(x)F$. For any real number λ, the set of <u>tapes accepted by A with cutpoint λ</u>, $T(A,\lambda) = \{x \in \Sigma^*: E_A(x) \geq \lambda\}$

0.2 NOTATIONAL CONVENTIONS

In what follows it will be convenient to use certain notational conventions. With regard to subscripts, note that state S_i is identified with the vector $(\overbrace{0, \ldots, 1}^{i}, 0, \ldots, 0)$. We use $A(x)_i$ to mean the i-th row of the matrix $A(x)$. Usually $A(x)^i$ will mean the i-th power of the matrix $A(x)$. On rare occasions when clear in context, $A(x)^i$ may mean the i-th column of $A(x)$. The following notational identities are used frequently:

$$S_i \cdot A(x) \cdot F = A(x)_i \cdot F = (A(x) \cdot F)_i$$

When necessary, identifying superscripts will be added to basic symbols, e.g. F^B is the output vector of machine B.

If Γ is an alphabet, all sequences of length r are denoted by Γ^r.

Following the conventions of automata theory, given an alpahbet Γ, Γ^* denotes the set of all finite sequences (or strings or tapes as they are often called) of symbols from Γ. It will always be assumed that Γ^* contains an identity string Λ_Γ so that $\Lambda_\Gamma x = x \quad x \in \Gamma^*$

Furthermore, we will assume that the input alphabet set Σ contains a symbol Λ such that $A(\Lambda) = E(n)$: the n-dimensional identity matrix. We do not require that $\Lambda x = x$ although $A(\Lambda x) = A(x)$.

The length of a string x is the length of the sequence of symbols which it denotes and will be written $\text{lg.}(x)$

Contrary to normal conventions we regard Λ as having length 1

$$\text{lg.}(\Lambda) = 1$$

whereas in general $\text{lg.}(\Lambda_\Gamma) = 0$.

Concatenation between strings will be indicated by juxtaposition. Multiplication of matrices will be indicated by either juxtaposition or other customary notation. Hence if x and y are strings

$$\text{lg.}(xy) = \text{lg.}(x) + \text{lg.}(y)$$

The exponential notation on strings will be used to indicate repetition i.e. $x^n = \overbrace{x \ldots x}^{n \text{ times}}$ so that $\text{lg.}(x^n) = n \cdot \text{lg.}(x)$. An abstract machine (a machine with no initial state specified) will be indicated by leaving a blank in the definition of the machine e.g.:

$$A = \langle n, \ , \Sigma, A(\sigma) : \sigma \in \Sigma, F, 0 \rangle$$

0.3 MODELS OF PROBABILISTIC SEQUENTIAL MACHINES

Two models will be considered, one of which is probabilistic and one of which is deterministic, although both fall within the axiomatic framework of probabilistic sequential machines.

Example 0.1: Probabilistic internal operation: A slot-machine.

A simple model of a probabilistic sequential machine is a slot-machine. The static position of the dials represents the present state of the machine. Usually there are 20 different positions on the dial and 3 dials for a total of 8,000 states. The input consists of putting in a coin and pulling a lever, causing the machine to travel transiently through many states until it settles down in one state. An output is associated with each state. Nothing (which is associated with 0) comes out unless the dials all display the same object. In that case, some change tumbles out (which is associated with the corresponding real number) usually dependent only on the kind of object being displayed, i.e., the state of the machine. Such a machine whose output is controlled by its states is known as a "Moore machine" [1956]. Each state can be associated with a number between 1 and 8,000, and the output for each state can be tabulated in a column vector or 8,000 x 1 matrix. In the formalism, this column vector will be called the "output vector" and designated by the symbol "F". The output for state i will be written as "F_i".

The enormous number of distinct ways the lever can be pulled are prevented from significantly influencing the outcome by spring loading. Hence a normal pull of the lever L produces only one kind of state transition law which could in principle be determined and tabulated in a "switching" matrix A(L). The behavior of a slot machine A could be described using a finite state markov chain with rewards and transition matrix A(L) but for the fact that various nonstandard but repeatable inputs have been developed by players of such machines. A more complete description requires some finite number of additional transition laws associated with the nonstandard inputs to the machine. We associate such inputs with additional input symbols.

Consider how the dials of the machine might be found initially. If the dials can be completely observed, the initial state S_i is represented by a vector I (or a 1 x 8,000 matrix) with a 1 in the i'th component and zeroes elsewhere. On the other hand, the dials may not be completely visible, we may wish to specify the average behavior of a large number of machines run simultaneously, or we may wish to consider the average return from playing one machine only when it is left by other players in one of a set of preferred states. In any one of these cases, I can be a stochastic vector $(I_1, \ldots, I_{8,000})$ where I_i is the probability of being in state S_i at time t_0.

In the general case, the next state probabilities starting with an initial state vector I and an input string x are given by $I \cdot A(x)$. Hence the expected value of output of a machine A starting with initial state distribution I and output vector F after a string x of inputs has occurred is just

$$E_A(x) = I \cdot A(x) \cdot F$$

which is a bilinear form in I and F with form matrix A(x). The variance in output and other higher moments can be defined analogously.

Example 0.2: Deterministic internal structure: Chemical production cell.

Suppose a chemical tank A is divided into several isolated compartments $A_1, \ldots, A_n$ by partitions which are interconnected by an electronically controlled system of pumps and valves. Suppose that there is a finite set of controls $\Sigma = \{0, 1, \ldots, K-1\}$ and that for each control c a fixed fraction of the chemical in compartment A_i, v_{ij}^c, is pumped into compartment A_j. For all controls c in Σ, the full influence on redistribution of liquid in the tank can be described in a n x n matrix A(c) with v_{ij}^c being $A(c)_{ij}$. Furthermore, suppose that the liquid being pumped between compartments is a catalyst which causes production of a desired end product in each compartment with a different efficiency, i.e., if the mass fraction of catalyst in A_i is P_i and F_i is the efficiency of A_i, then the output of end product is P_iF_i. Note that it is assumed that the output of the compartment depends linearly on the catalyst present.

The initial state I is an n component vector with the i'th component I_i being the mass fraction of catalyst in compartment i. Note $\sum_{i=1}^{n} I_i = 1$ since the tank is a closed system as far as the catalyst is concerned. The distribution of mass fractions of catalyst over the compartments after a sequence of controls $x = i_1 \ldots i_m$ is just

$$I \circ A(i_1) \circ \ldots \circ A(i_m) = I \circ A(x)$$

That is, $(I \circ A(x))_i$ is the mass fraction of catalyst in compartment i after starting with initial distribution I of catalyst fractions over compartments and the string of control inputs $x = i_1 \ldots i_m$.

The total end product from the tank is the sum of the outputs from each compartment: $\sum_{i=1}^{n} (I \circ A(x))_i F_i$ which can be written $I \circ A(x) \circ F$ in matrix notation. This expression has the same form as the expectation of output for the probabilistic slot-machine, but there are no overt probabilities involved here. The mass fractions of catalyst play the same role as the probabilities in the first example. However, the output will still be written like an expectation as $E_A(x)$.

The total end product accumulated, T_x, for the string of controls x from time t_0 to time $t_0 + m$ is given by adding the output from each substring, i.e.,

$$T_x = E_A(i_1) + E_A(i_1 i_2) + \ldots + E_A(i_1 i_2 \ldots i_m)$$

CHAPTER 1

DETERMINING WHETHER A PROBABILISTIC SEQUENTIAL MACHINE IS EXPECTATION EQUIVALENT TO A FINITE DETERMINISTIC MACHINE

1.1 THE CONCEPT OF EXPECTATION EQUIVALENCE

In the two models discussed in the introduction, the expected value of output, $E_A(x)$, played an important role in the physical interpretations. Let us repeat the definition of the expected value of output.

Definition 1.1: The expected value of output for an input string x of a probabilistic sequential machine A is given by

$$E_A(x) = I \cdot A(x) \cdot F \quad \text{for } x \text{ in } \Sigma^*$$

Definition 1.2: Machines A and A' are expectation equivalent, written $A \equiv_E A'$; if

$$E_A(x) = E_{A'}(x) \quad \text{for all } x \text{ in } \Sigma^*$$

Recall from Example 0.2 that $E_A(x)$ was the actual output of the chemical cell and not an expectation. Hence the basic concept of expectation equivalence is analogous to the definition of behavioral equivalence $\equiv_I$ for Example 0.2. However for Example 0.1, the slot-machine, expectation equivalence is not the generalization of this kind of behavioral equivalence. Instead, the concept of indistinguishability discussed in Chapter 3 seems to be the appropriate generalization.

Example 1.1. Machines A and A' which are expectation equivalent:

$$IA(x)F = I'A'(x)F' \quad \forall x \in \Sigma^*$$

$$A = \langle I, A(0), A(1), F\rangle \text{ and } A' = \langle I, A'(0), A'(1), F'\rangle$$

$$A(0) = \begin{pmatrix} 1 & 0 & 0 \\ 1/2 & 1/4 & 1/4 \\ 1/4 & 0 & 3/4 \end{pmatrix} \qquad A(1) = \begin{pmatrix} 3/5 & 1/5 & 1/5 \\ 1/5 & 4/5 & 0 \\ 4/5 & 1/5 & 0 \end{pmatrix}$$

$$A'(0) = \begin{pmatrix} 1 & 0 & 0 \\ 5/8 & 0 & 3/8 \\ 0 & 1/2 & 1/2 \end{pmatrix} \qquad A'(1) = \begin{pmatrix} 7/10 & 0 & 3/10 \\ 3/5 & 0 & 2/5 \\ 9/10 & 0 & 1/10 \end{pmatrix}$$

$$F = F' = \begin{pmatrix} 7 \\ 5 \\ 3 \end{pmatrix}$$

These machines are expectation equivalent from any initial probability distribution, I, over the states.

The previous example shows that two machines can have very different switching matrices and still be expectation equivalent. Some graph theoretic properties of the transition matrices which are important to markov theory, such as the accessibility of a state, depend on the location of the zeroes. This example shows that the location of the zeroes is not the only relevant factor in the study of expectation equivalence. Consideration of the interplay between the state transitions and the real number outputs attached to states requires the use of elementary linear algebra.

1.2 THE REDUCTION RELATION R_E

In this section a congruence relation on input sequences, R_E, will be defined so that a quotient machine can be constructed. If the rank of R_E happens to be finite, the constructed machine has a finite number of states. States of the quotient machine will correspond to values of expectation which occur for input strings. By attaching a deterministic output device to each state of the constructed machine, an expectation equivalent deterministic machine is obtained.

If the rank of R_E is finite, some class of the relation must contain infinitely many strings. A necessary condition for R_E to be finite in rank is that it be non-trivial, i.e, at least two different strings are contained in some class. This weak necessary condition requires the symbol matrices to satisfy certain strong conditions.

Definition 1.3: The reduction relation R_E is given by

$$x \; R_E \; y \text{ iff } E_A(x) = E_A(y) \; \& \; E_A(xz) = E_A(yz) \; \forall z \in \Sigma^* \; \forall I \in S$$

If Σ^* contains Λ, a semigroup identity, the definition reduces to

$$x \; R_E \; y \text{ iff } E_A(xz) = E_A(yz) \; \forall z \in \Sigma^*, \; \forall I \in S.$$

R_E is a right congruence relation on Σ^* because of the reflexivity, transitivity and symmetry of "=" and the substitution property in its definition.

It follows that strings x and y which are in the same class of the relation R_E will have equal expectations from any initial state of the machine and will continue to have equal expectations for any finite input continuation z. As far as expectation of output is concerned, the behavior of the machine A is the same after either string x or string y.

1.3 CONSTRUCTION OF THE QUOTIENT MACHINE

Definition 1.4: The equivalence class of x' of R, an equivalence relation, is given by

$$R[x'] = \{x : x \; R \; x'\}$$

It is a well known result from Rabin and Scott [1959] that given a right congruence relation R on Σ^*, one can construct a quotient automaton with no output T(R)

$$T(R) = \langle a, S, M\rangle$$

where

$$a = R[\Lambda]$$

$$S = \{R[x] : x \in \Sigma^*\}$$

M is a function from S x Σ into S such that

$$M(R[x],\sigma) = R[x\sigma] \quad x \in \Sigma^*; \sigma \in \Sigma$$

Definition 1.5: Let $\beta \subseteq \Sigma^*$. A congruence R refines β if

$$x \mathrel{R} y \Longrightarrow x \in \beta \text{ iff } y \in \beta$$

Theorem 1.1 Rabin and Scott [1959]

Let β be a subset of Σ^*. β is the behavior of a finite (deterministic) automaton $A = \langle T(R), \mathcal{F}\rangle$ over Σ where $\mathcal{F} = \{R[x] : x \in \beta\}$ iff there exists a right congruence relation R of finite rank which refines β.

Theorem 1.2

If the congruence relation R_E has finite rank, then for any λ there is a finite deterministic automaton A' such that the tapes accepted by A' are $T(A,\lambda)$.

Proof: Let $\beta = T(A,\lambda) = \{x : E_A(x) \geq \lambda\}$. Note that R_E refines β i.e. $x \mathrel{R_E} y \Longrightarrow x \in T(A,\lambda)$ iff $y \in T(A,\lambda)$. If R_E has finite rank, by definition $R_E[x]$ has a finite number of members. Using Theorem 1.1 we construct

$$T(R_E) = \langle a, S, M\rangle \quad \text{and}$$

$$A' = \langle a, S, M, \mathcal{F}\rangle \text{ which accepts } T(A,\lambda)$$

Q. E. D.

1.4 CONSTRUCTION OF AN EXPECTATION EQUIVALENT FINITE DETERMINISTIC MACHINE

The quotient machine construction will be used to obtain a sufficient condition for the reduction of a probabilistic sequential machine into an expectation equivalent finite deterministic machine whose output function is either a constant $C(s)$ for each state s or a random device $O_A^R(s)$ with expectation $E(O_A^R(s)) = C(s)$.

Definition 1.6: $rp_A(x)$ is the response of A to input string x. If A is deterministic, $rp_A(x)$ is the state of A after an input of x. If A is probabilistic, $rp_A(x)$ is a random variable taking on values which are states with distribution $I\circ A(x)$.

Theorem 1.3

The reduction relation R_E defined by a probabilistic machine A has finite rank if and only if there exists a finite deterministic machine A' with a deterministic output $O_{A'}$ such that $O_{A'}(rp_{A'}(x)) = E_A(x) \quad \forall x \in \Sigma^*$

Proof (sufficiency): By Theorem 1.1 let $A' = \langle a, S, M, \phi\rangle$ where ϕ is the empty set. Note any congruence R refines ϕ vacuously. We attach an output function $O_{A'}$ to elements of S,

$$O_{A'}(s) = E_A(x) \qquad s = R_E[x]$$

For a deterministic machine, M is extended to M* which operates on strings rather than symbols by

$$M^*(s,\sigma) = M(s,\sigma) \qquad s \in S \qquad \sigma \in \Sigma$$

$$M^*(s,\sigma x) = M^*(M^*(s,\sigma),x) \qquad x \in \Sigma^*$$

We note that $M^*(a,x) = rp_{A'}(x)$ so we need to show only that $rp_{A'}(x) = R_E[x]$. Let $x = i_1 i_2 \ldots i_m$ for $i_j \in \Sigma$; $j = 1, 2, \ldots, m$.

$$rp_{A'}(x) = M^*(\Lambda,x) = M^*(M^*(a,i_1),i_2 \ldots i_m)$$

$$= M^*(M(a,i_1),i_2 \ldots i_m)$$

$$= M^*(M(R_E[\Lambda], i_1), i_2 \ldots i_m)$$
$$= M^*(R_E[\Lambda i_1], i_2 \ldots i_m)$$
$$= M^*(M(R_E[i_1], i_2), i_3 \ldots i_m)$$
$$\vdots \qquad \vdots$$
$$= R_E[i_1 i_2 \ldots i_m] = R_E[x]$$

Hence the constructed sequential machine is $A' = \langle a, S, M, O_{A'} \rangle$.

(necessity)

Given $A' = \langle a, S, M, O_{A'} \rangle$ such that

$$O_{A'}(rp_{A'}(x)) = E_A(x) \qquad \forall x \in \Sigma^*$$
$$O_{A'}(rp_{A'}(xz)) = E_A(xz) \qquad \forall z \in \Sigma^*$$

Let $rp_{A'}(x) = S_x \;\; x \in \Sigma^*$

Define

$$S_x \, R_O \, S_y \quad \text{iff} \quad x \, R_E \, y$$

Let n' be the cardinality of S — finite.

$$\text{rank } R_O = \text{rank } R_E$$
$$\text{rank } R_O \leq n'$$

Hence rank R_E is finite.

Q. E. D.

Corollary 1.3

The reduction relation R_E defined by a probabilistic machine A has finite rank $\Longleftrightarrow$ there exists a finite deterministic machine A' such that $A \equiv_E A'$

Proof: The machine A' of Theorem 1.3 meets the condition of the corollary since

$$E_{A'}(x) = O_{A'}(rp_{A'}(x)) = E_A(x) \qquad \forall x \in \Sigma^*$$

Q. E. D.

Instead of the deterministic function $O_{A'}$, a random device $O^R_{A'}(s)$ such that $E(O^R_{A},(s)) = E_A(x)$ could have been used in the construction.

1.5 THE PARTITION OF THE SET OF ACCESSIBLE STATE DISTRIBUTIONS INDUCED BY R_E

Definition 1.7: $V(A) = \{IA(x) : x \in \Sigma^*\}$ — the set of all stochastic vectors which can occur as distributions over the states of A. We sometimes call V(A) the "state vectors accessible in A".

Definition 1.8: A set of vectors $V = \{v_1, v_2, \ldots\}$ is convex if for any finite set of indices J, real numbers $c_j \geq 0 \quad j \in J$ and $\sum_{j\in J} c_j = 1 \Longrightarrow \sum_{j\in J} c_j v_j \in V$. The convex closure of a set of vectors V, written $V^+ = \{v' : v' = \sum_{j\in J} c_j v_j, \sum_{j\in J} c_j = 1, c_j \geq 0 \text{ and } v_j \in V\}$. It is clear that $V(A) \subseteq S^+$.

Theorem 1.4

If R_E has finite rank r, there exists a partition $\Pi = (\Pi_1, \ldots, \Pi_r)$ on V(A) and an integer valued function $g(\ell,m)$ such that

$$\Pi_i A(\sigma) \subset \Pi_{g(i,\sigma)} \qquad i = 1, \ldots, r; \quad \sigma \in \Sigma$$

Proof: R_E induces an equivalence on the set of stochastic vectors accessible by the machine.

Since R_E has finite rank, form a set of an arbitrary distinct representative from each congruence class, say $x_1, \ldots, x_r$ where $x_i \neq x_j$ $i = 1, 2, \ldots, r; \ j < i$.

Define $\Pi_j = \bigcup_{x \in R_E[x_j]} \{IA(x)\}$

We show that $(\Pi_1, \ldots, \Pi_r)$ is a partition of V(A)

Let $W = \bigcup_{i=1}^{r} \Pi_i$

$$IA(x') \in W \Longrightarrow IA(x') \in V(A)$$

$$IA(x') \; \varepsilon \; V(A) \Longrightarrow x' \; \varepsilon \; R_E[x_k] \quad \text{for some } k = 1, \ldots, r$$

$$\Longrightarrow IA(x') \; \varepsilon \; \Pi_k \text{ for some } k = 1, \ldots, r$$

$$\Longrightarrow IA(x') \; \varepsilon \; W$$

Hence

$$W = \bigcup_{i=1}^{n} \Pi_i = V(A)$$

We show

$$\Pi_i \cap \Pi_j = \phi \qquad i \neq j$$

suppose that

$$IA(y) \; \varepsilon \; \Pi_i \cap \Pi_j$$

$$IA(y) \; \varepsilon \; \Pi_i \Longrightarrow y \; \varepsilon \; R_E[x_i] \Longrightarrow y \; R_E \; x_i$$

$$IA(y) \; \varepsilon \; \Pi_j \Longrightarrow y \; \varepsilon \; R_E[x_j] \Longrightarrow y \; R_E \; x_j$$

Hence we get

$$y \; R_E \; x_i \Longrightarrow x_i \; R_E \; y \text{ by symmetry}$$

and transitivity of R_E gives

$$x_i \; R_E \; x_j \Longrightarrow x_i \; \varepsilon \; R_E[x_j]$$

But x_i and x_j are representatives and there is only one representative from each class

$$x_i = x_j \qquad i \neq j$$

which is a contradiction.

Finally we show there exists an integer valued function g(i,σ) such that

$$\Pi_i A(\sigma) \subseteq \Pi_{g(i,\sigma)} \qquad \sigma \; \varepsilon \; \Sigma$$

$$v_1 \; \varepsilon \; \Pi_i \Longrightarrow v_1 = IA(w_1) \quad \text{for some } w_1 \; \varepsilon \; \Sigma^*$$

$$v_1 A(\sigma) = IA(w_1)A(\sigma) = IA(w_1\sigma) \; \varepsilon \; \Pi_j$$

for some j as has been shown above.

$$v_2 \; \varepsilon \; \Pi_i \Longrightarrow v_2 = IA(w_2) \quad \text{for some } w_2 \; \varepsilon \; \Sigma^*$$

$$v_2 A(\sigma) = IA(w_2\sigma) \; \varepsilon \; \Pi_j$$

since elements of R_E have the substitution property, i.e.

$$w_1 \; R_E \; x_i \Longrightarrow w_1\sigma \; R_E \; x_i\sigma \qquad \sigma \varepsilon \Sigma$$

$$w_2 \; R_E \; x_i \Longrightarrow w_2\sigma \; R_F \; x_i\sigma \qquad \sigma \varepsilon \Sigma$$

$x_i\sigma$ is an element of a class with representative x_j for some j and depends only on x_i and σ. So there is a function $g(\ell,m)$ such that

$$g(i,\sigma) = j \qquad \sigma \varepsilon \Sigma$$

Q. E. D.

1.6 NECESSARY AND SUFFICIENT CONDITIONS THAT STRINGS BE IN THE SAME R_E CLASS

The relation R_E has occupied an important place in the development of this theory. The structure of the matrices of strings which are in the same R_E class will now be studied.

Definition 1.9: A relation R is non-trivial if there exist x and y in the domain of R with $x \neq y$ such that $x \; R \; y$.

Definition 1.10: The kernel of F = Kern. (F)

$$= \{v \; \varepsilon \; R^n \; : \; v \circ F = 0\}$$

where R is the set of reals.

Definition 1.11: The span of a set of vectors $\{v_1, \dots, v_r\}$ is denoted by

$$\langle\{v_1, \dots, v_r\}\rangle = \left\{\sum_{i=1}^{r} c_i v_i \;\; \forall c_i \; \varepsilon \; R\right\}$$

Theorem 1.5

A necessary and sufficient condition for x and y to be in the same class of the reduction relation R_E is:

$x \; R_E \; y \Longleftrightarrow$ there exists a subspace U of Kern. (F) such that

(i) $UA(z) \subset \text{Kern.}(F) \qquad \forall z \in \Sigma^*$

(ii) $A(x) = A(y) + \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}$ with $u_i \in U \quad i = 1, \ldots, n$

Proof:

$$x \, R_E \, y \iff IA(xz)F = IA(yz)F \quad \forall I \in S \quad \forall z \in \Sigma^*$$

hence

$$A(x)F = A(y)F$$

since

$$S = \{(1, 0, \ldots, 0), \ldots, (0, \ldots, 0, 1)\} \text{ and } \Lambda \in \Sigma^*$$

By elementary linear algebra the solution of the above is a particular solution and a kernel.

$$A(x) = A(y) + \begin{vmatrix} h_1 \\ \vdots \\ h_n \end{vmatrix} \quad \text{where } h_i \in \text{Kern.}(F) \quad i = 1, 2, \ldots, n$$

multiplying by A(z)

$$A(x)A(z) = A(y)A(z) + \begin{vmatrix} h_1 \\ \vdots \\ h_n \end{vmatrix} A(z) \quad \forall z \in \Sigma^*$$

$$A(xz) = A(yz) + \begin{vmatrix} h_1 \\ \vdots \\ h_n \end{vmatrix} A(z)$$

Multiplying by an arbitrary distribution I and output vector F

$$IA(xz)F = IA(yz)F + I \circ \begin{vmatrix} h_1 \\ \vdots \\ h_n \end{vmatrix} \circ A(z)F \quad I \in S^+$$

But since x and y are in R_E

$$IA(xz)F = IA(yz)F \quad \forall z \in \Sigma^* \quad I \in S^+$$

Hence

$$I \cdot \begin{vmatrix} h_1 \\ \vdots \\ h_n \end{vmatrix} \cdot A(z)F = 0 \Longrightarrow h_i \; A(z) \in \text{Kern.}(F), \quad i = 1, 2, \ldots, n.$$

Let $U = \langle \{h_1, \ldots, h_n\} \rangle$

We get $UA(z) \subset \text{Kern.}(F) \quad \forall z \in \Sigma^*$

Let $H = \begin{vmatrix} h_1 \\ \vdots \\ h_n \end{vmatrix}$ where $h_i \in U \subset \text{Kern.}(F)$, $i = 1, 2, \ldots, n.$

$$A(x) = A(y) + H \tag{1}$$

Multiplying by I on the left and F on the right gives

$$IA(x)F = IA(y)F + I \cdot \begin{pmatrix} h_1 \cdot F \\ \vdots \\ h_n \cdot F \end{pmatrix} \quad I \in S^+$$

but $h_iF = 0$ since $h_i \in \text{Kern.}(F)$ $i = 1, \ldots, n$

$$IA(x)F = IA(y)F$$

using (1) again and the same argument gives

$$A(xz) = A(yz) + HA(z)$$

$$IA(xz)F = IA(yz)F + IHA(z)F$$

$$= IA(yz)F$$

since $HA(z) = \begin{vmatrix} u_1 \\ \vdots \\ u_n \end{vmatrix}$ where $u_i \in \text{Kern.}(F)$, $i = 1, 2, \ldots, n$

Q. E. D.

Part (i) of Theorem 1.5 will be restricted to the finite class of symbols rather than the unbounded class of strings.

Theorem 1.6

Let $U = \left\langle \bigcup_{x \in \Sigma^*} \{A(x)_i - A(y)_i\} \; ; \; i = 1, \ldots, n \text{ for } x, y \text{ such that } x \, R_E \, y \right\rangle$

then

$$U \circ A(z) \subset \mathrm{Kern.}\,(F) \Longleftrightarrow [\exists V \text{ a subspace of } R^n;$$

$$\text{(i)} \quad UA(\sigma) \subset V \; ; \; \forall \sigma \in \Sigma$$

$$\text{(ii)} \quad VA(\sigma) \subset V \subset \mathrm{Kern.}\,(F) \qquad \forall \sigma \in \Sigma]$$

Proof:

$UA(z) \subset \mathrm{Kern.}\,(F)$

Let $V = \langle \{u \circ A(z); \quad u \in U, \quad z \in \Sigma^*\} \rangle$

$VA(\sigma) = \{uA(z)A(\sigma); \quad u \in U, \quad z \in \Sigma^*\}$

$= V$

Consider an arbitrary $v \in V$. There must be some set of indexes J and constants c_j such that:

$$v = \sum_{j \in J} c_j u_j A(z_j) \text{ by definition of } V.$$

$$v \circ F = \left(\sum_{j \in J} c_j u_j A(z_j) \right) \circ F$$

$$= \sum_{j \in J} c_j (u_j A(z_j) F$$

But

$$u_j A(z_j) F = 0 \text{ by } UA(z) \subset \mathrm{Kern.}\,(F)$$

so

$$v \circ F = 0$$

Hence

$$V \subset \mathrm{Kern.}\,(F)$$

$$UA(z) \subset V \text{ already shown}$$

$$\therefore UA(z) \subset \mathrm{Kern.}\,(F)$$

Q. E. D.

1.7 INVARIANT SUBSPACE CHARACTERIZATION OF R_E

Definition 1.12: A subspace V is invariant under a set of linear transformations $\{T_i : i = 1, 2, \ldots, m\}$ if

$$V \circ T_i \subset V \quad i = 1, 2, \ldots, m$$

Theorems 1.5 and 1.6 yield the following directly:

Theorem 1.7

Strings x and y are in the same class of R_E if and only if there exists a subspace V of Kern.(F) such that

(i) V is invariant under $\{A(\sigma); \forall\sigma \in \Sigma\}$

(ii) $A(x) = A(y)+H$ where $H_i \in V \quad i = 1, \ldots, n$

1.8 NECESSARY AND SUFFICIENT CONDITIONS THAT R_E BE NON-TRIVIAL

A very weak necessary condition that R_E have finite rank is that it is at least non-trivial. From Theorem 1.7 it is immediate that:

Corollary 1.8

The reduction relation R_E is non-trivial $\Longleftrightarrow$ there exists a subspace V of Kern.(F) such that

(i) V is invariant under $\{A(\sigma); \forall\sigma \in \Sigma\}$

(ii) $A(x) = A(y)+H$ where $H_i \in V \quad i = 1, \ldots, n$

(iii) $x \neq y$.

Hence we now know that given strings x and y in the same class of R_E, the difference between the rows of the matrices A(x) and A(y) must be elements of a subspace V which has special properties. Namely V must be invariant under all symbol matrices and contained in the kernel of the output vector.

Theorem 1.9

A necessary condition that R_E be nontrivial is that $A(\sigma)$: $\forall\sigma \in \Sigma$ be reducible for the same change of basis. In other words, there exists a subspace V and a linear transformation W of the state vectors S to a basis for V such that

$$W^{-1}A(\sigma)W = \begin{bmatrix} \overbrace{A_1^\sigma}^{\text{basis for V}} & 0 \\ A_2^\sigma & A_3^\sigma \end{bmatrix}$$

Where 0 denotes a submatrix of zeroes and A_1^σ, A_2^σ, and A_3^σ are submatrices which for all σ in Σ have the same number of columns and rows.

Proof: By Theorem 1.7 and standard matrix theory [see Jacobson, Lectures in Abstract Algebra, V. II: Linear Algebra, Van Nostrand, New York, 1952, pp. 116-117].

Theorem 1.9 gives us a strong matrix reformulation of the statement that R_E be nontrivial.

Example 1.2. We construct an expectation equivalent finite deterministic machine from a probabilistic sequential machine A illustrating Theorem 1.3, Corollary 1.3 and Theorem 1.7.

$$A = \langle I, A(0), A(1), F\rangle$$

where

$$I = (8/10, \quad 1/10, \quad 1/10, \quad 0, \quad 0, \quad 0)$$

$$A(0) = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 3/4 & 0 & 1/4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \qquad F = \begin{pmatrix} 10 \\ 5 \\ 1 \\ 2 \\ 1 \\ 2 \end{pmatrix}$$

$$A(1) = \begin{pmatrix} 0 & 0 & 1/8 & 0 & 7/8 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 4/8 & 0 & 4/8 & 0 \\ 0 & 0 & 3/8 & 0 & 5/8 & 0 \\ 0 & 0 & 2/8 & 0 & 6/8 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

The state diagram for A is shown in Figure 1.1.

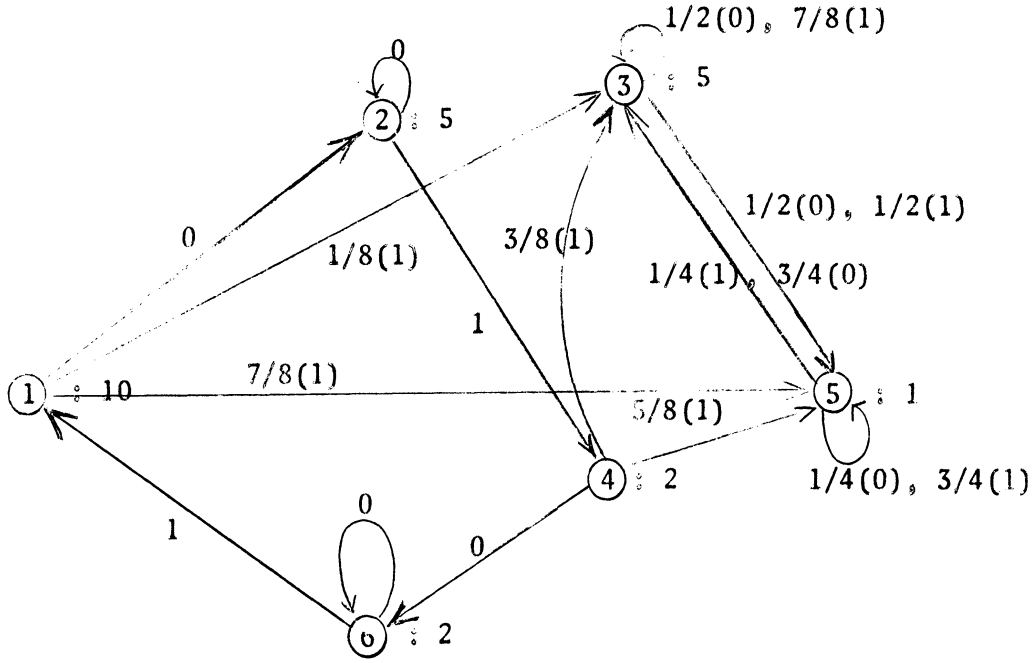


Figure 1.1 State Diagram for the Machine of Example 1.2.

The following labeling conventions are used:

$p(K)$: p $[0,1]$; $K \varepsilon \Sigma$ means probability of transition of p via symbol K.

ℓ : F_ℓ : Output of F_ℓ occurs when the machine is in state ℓ.

$$0 \underset{P_2(K_2)}{\overset{P_1(K_1)}{\rightrightarrows}} 0 \text{ : is replaced by } 0 \xrightarrow{P_1(K_1),\ P_2(K_2)} 0$$

It will now be demonstrated that for machine A

$$00R_E0$$

$$A(00) = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 3/4 & 0 & 1/4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 3/4 & 0 & 1/4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 5/8 & 0 & 3/8 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 9/16 & 0 & 7/16 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

which gives

$$(A(00) - A(0))F = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & +1/8 & 0 & -1/8 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -3/16 & 0 & 3/16 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} 10 \\ 5 \\ 1 \\ 2 \\ 1 \\ 2 \end{pmatrix} = (0,0,0,0,0,0)$$

Hence $A(00)F = A(0)F$ or $IA(00)F = IA(0)F$ for all I.

Furthermore, for all $P \varepsilon [0,1]$

$$(0, 0, P, 0, 1-P, 0)A(0) = (0, 0, P, 0, 1-P, 0)$$

$$(0, 0, P, 0, 1-P, 0)A(1) = (0, 0, P, 0, 1-P, 0)$$

that is

$$W = \langle\{(0, 0, P, 0, 1-P, 0)\}\rangle$$

is invariant under the symbol matrices $A(0)$ and $A(1)$.

$$V = \langle\{(0, 0, P, 0, -P, 0)\}\rangle \subset W \text{ and } \begin{array}{l} VA(0) = V \\ VA(1) = V \end{array}$$

By Theorem 1.7 we know $00R_E0$ but let us verify this fact.

For $z \varepsilon \Sigma^*$

$$(A(00) - A(0))A(z) = C_z \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & +1/8 & 0 & -1/8 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -3/16 & 0 & 3/16 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} = D$$

where C_z is a constant depending on the string z

and

$$(A(00) - A(0))A(z)F = DF = 0$$

consequently $\forall z \epsilon \Sigma^*$ $\qquad \forall I \epsilon S^+$

$$IA(00)A(z)F = IA(0)A(z)F$$

or

$$E_A(00z) = E_A(0z), \text{ which shows } 00R_E0.$$

By the same method one can show that

$$10\ R_E\ 1 \quad 011\ R_E\ 11 \quad 01011\ R_E\ 11 \quad 111\ R_E\ 11 \quad 01010\ R_E\ 0$$

so all strings are in the classes

$$R_E(\Lambda), \quad R_E[0], \quad R_E[1], \quad R_E[11], \quad R_E[01], \quad R_E[010], \quad R_E[0101]$$

which means that R_E has finite rank.

Following Theorem 1.3, we compute the expectations and construct the expectation equivalent deterministic machine A'. Note that the values of expectation depend on the initial state I.

$$E_A(\Lambda) = IA(\Lambda)F = IF = 8.6$$

$$E_A(0) = (0, 9/10, 1/20, 0, 1/20, 0)F = 4.6$$

$$E_A(1) = (0, 0, 3/20, 2/20, 15/20, 0)F = 1.1$$

$$E_A(01) = (0, 0, 3/80, 72/80, 5/80, 0)F = 1.9$$

$$E_A(10) = (0, 0, 3/20, 0, 15/20, 2/20)F = 1.1 = E_A(1) \text{ (since } 10R_E1)$$

$$E_A(11) = (0, 0, 9/40, 0, 31/40, 0)F = 1.0$$

$$E_A(010) = 1.9$$

$$E_A(0101) = 9.1$$

The expectation equivalent deterministic machine of Corollary 1.3 is shown in Figure 1.2.

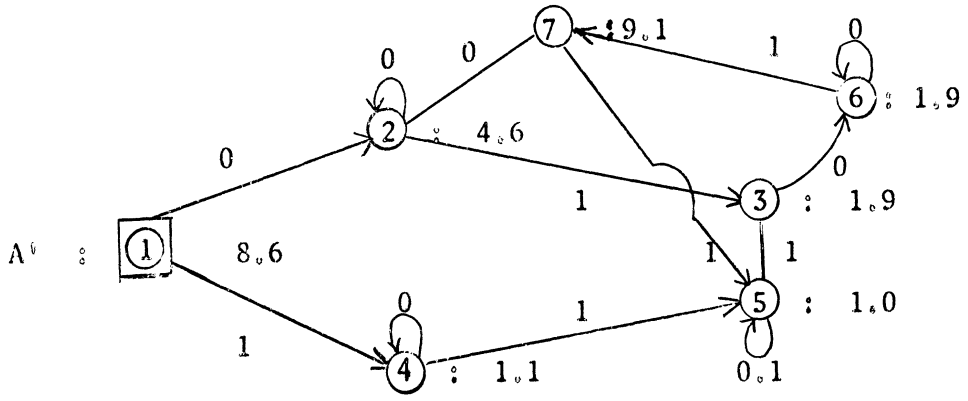


Figure 1.2 $A' = \Sigma^*/R_E$ for Example 1.1

We note that A' has 7 states while A has just 6 states. The deterministic cycle 0101 appears in both machines.

CHAPTER 2

DETERMINING WHETHER A PROBABILISTIC SEQUENTIAL MACHINE IS N-MOMENT EQUIVALENT TO AN INPUT-STATE CALCULABLE MACHINE

2.0 INTRODUCTION

In this chapter the concept of expectation equivalence is generalized to N-moment equivalence. A congruence relation R_N is defined which partitions the set of input strings into classes. All members of a particular class produce the same expectation and first N-1 central moments for the machine defining R_N. If R_N has finite rank, a finite quotient machine can be constructed which is deterministic with each state corresponding to a congruence class. Each state can be connected to a random device having the same expectation and N-1 moments as the class represented by the state, giving a deterministic machine with random outputs. The constructed input-state calculable machine is N-moment equivalent to the probabilistic machine.

After the first theorem concerning a necessary and sufficient condition that two strings be in the same R_N class, a simple substitution gives generalizations of some results of Chapter 1. Hence the generalizations are presented in this chapter without proofs.

2.1 DISTRIBUTION EQUIVALENCE: $\equiv_D$

The random variable structure of probabilistic sequential machines will be investigated in this section.

Definition 2.1: $0^*_A(x)$: the output random variable of the machine A after a string x has occurred as input. Using Definition 1.6 we note that

$$0^*_A(x) = 0(rp_A(x))$$

Definition 2.2: A and A' are distribution equivalent, written $A \equiv_D A'$, if for $J_A = \{j : (IA(x)_j F_j \neq 0\}$ there is a 1-1 map h between $J_{A'}$ and J_A such that

$$IA(x)_{h(j)} = I'A'(x)_j \qquad j \in J_{A'} \qquad x \in \Sigma^*$$

$$F_{h(j)} = F'_j \qquad j \in J_{A'}$$

Distritution equivalence corresponds to the conventional definition of equivalence for discrete random variables except for random variables $F_i \neq F_j$ for $i \neq j$.

Referring back to Example 0.2, two chemical cells are distribution equivalent if (1) We neglect those partitioned areas which have either zero efficiency or a zero fraction of the catalyst. (2) Of the remaining partitioned areas there is a correspondence between the partitioned areas of one cell and the other such that corresponding areas have the same fraction of catalyst regardless of the sequence of controls entering the cells. (3) Corresponding partioned areas have the same efficiencies.

2.2 MOMENTS OF THE OUTPUT RANDOM VARIABLE

Definition 2.3: Let

$$F = \begin{pmatrix} F_1 \\ \vdots \\ F_n \end{pmatrix} \qquad F_i \in R \qquad i = 1, 2, \ldots, n$$

call

$$(F^k) = \begin{pmatrix} F_1^k \\ \vdots \\ F_n^k \end{pmatrix}$$

Then the <u>i'th central moment of $O^*_A(x)$</u> is

$$\mu^A_i(x) = E[(O^*_A(x) - E_A(x))^i] \qquad i = 2, 3, \ldots$$

<u>Theorem 2.1</u>

$$\mu^A_i(x) = \sum_{k=0}^{i} \binom{i}{k}(-1)^k IA(x)(F^{i-k})E_A(x)^k \qquad i = 2, 3, \ldots$$

<u>Proof</u>: By the binominal theorem

$$\mu^A_i(x) = E[\sum_{k=0}^{i} (-1)^k O^*_A(x)^{i-k} E_A(x)^k \binom{i}{k}]$$

To compute the expectation of the discrete random variable $O^*_A(x)^{i-k}$ note that it has the same distribution as $O^*_A(x)$ but takes on values $F_1^{i-k}, \ldots, F_n^{i-k}$ for $i \neq k$

$$\mu^A_i(x) = \sum_{k=0}^{i-1} (-1)^k \binom{i}{k} E[O^*_A(x)^{i-k}]E_A(x)^k + (-1)^i E_A(x)^i$$

$$= \sum_{k=0}^{i-1} (-1)^k \binom{i}{k} \cdot IA(x)(F^{i-k})E_A(x)^k + (-1)^i E_A(x)^i$$

Q. E. D.

2.3 SPECIAL PROPERTIES OF RABIN PROBABILISTIC AUTOMATA

<u>Definition 2.4</u>: A <u>Rabin probabilistic automaton</u> [1964] is a probabilistic sequential machine such that $I \in S$ and $F_i = 0$ or $F_i = 1 \quad i = 1, 2, \ldots, n.$

Rabin probabilistic automata have rather special features as far as the random variable of the output is concerned.

<u>Corollary 2.1</u>: For a Rabin probabilistic automaton A

$$\mu_i^A(x) = \sum_{k=0}^{i-1} (-1)^k \binom{i}{k} E_A(x)^{k+1} + (-1)^i E_A(x)^i \qquad i = 2, 3, \ldots$$

Proof: $F_i = 0$ or 1 hence

$$(F^{i-k}) = F \quad \text{for} \quad i \neq k$$

and the result from Theorem 2.1.

Corollary 2.2: If $E_A(x) = E_A(y)$ for some Rabin probabilistic automaton A, then all central moments for x and y are equal also, i.e.

$$\mu_i^A(x) = \mu_i^A(y) \quad \text{for} \quad i = 2, 3, \ldots$$

Note: for $i = 2$ we get the variances of the outputs are equal.

Corollary 2.3: If two Rabin probabilistic automaton A and A' are expectation equivalent then

$$\mu_i^A(x) = \mu_i^{A'}(x) \qquad i = 2, 3, \ldots \qquad \forall x \in \Sigma^*$$

2.4 THE CONCEPT OF N-MOMENT EQUIVALENCE: $\equiv_N$

Even if two machines are expectation equivalent, the statistics of their behavior may be so different that for many purposes we would not want to consider the machines behaviorally equivalent. Returning to Example 0.1, two slot-machines can be expectation equivalent, meaning that the average payoff is the same for both, but one can be much more desirable than the other for a player of limited resources. For a player with limited resources might have a far longer average time until "gambler's ruin" on one machine than the other. Hence in order to associate machines in the same class whose statistics of behavior are somewhat alike, the notion of N-moment equivalence will be introduced.

Definition 2.5: Probabilistic sequential machines A and A' are N-moment equivalent, written $A \equiv_N A'$ if

$$E_A(x) = E_{A'}(x)$$

$$\mu_i^A(x) = \mu_i^{A'}(x) \quad i = 2, \ldots, N \quad \text{for all } x \text{ in } \Sigma^*$$

Example 2.1. Probabilistic sequential machines A and A' such that $A \equiv_N A'$ for any initial distribution I, i.e.

$$E_A(x) = E_{A'}(x)$$

and

$$\mu_i^A(x) = \mu_i^{A'}(x) \quad \forall x \in \Sigma^* \quad i = 2, 3, \ldots \quad \forall I \in S^+$$

$$A(0) = \begin{vmatrix} 1 & 0 & 0 \\ 1/2 & 1/4 & 1/4 \\ 1/4 & 0 & 3/4 \end{vmatrix} \qquad A(1) = \begin{vmatrix} 3/5 & 1/5 & 1/5 \\ 1/5 & 4/5 & 0 \\ 4/5 & 1/5 & 0 \end{vmatrix}$$

$$A'(0) = \begin{vmatrix} 0 & 0 & 1 \\ 0 & 1/4 & 3/4 \\ 0 & 0 & 1 \end{vmatrix} \qquad A'(1) = \begin{vmatrix} 4/5 & 1/5 & 0 \\ 0 & 4/5 & 1/5 \\ 4/5 & 1/5 & 0 \end{vmatrix}$$

For both machines

$$F = \begin{vmatrix} F_1 \\ F_2 \\ F_1 \end{vmatrix} \quad \text{for } F_1, F_2 \text{ arbitrary real numbers.}$$

2.5 THE RELATIONSHIP BETWEEN $\equiv_D$ AND $\equiv_N$

Theorem 2.2

For probabilistic sequential machines A and A'

$$A \equiv_D A' \Longrightarrow A \equiv_N A' \text{ for all finite N}$$

Proof: Distribution equivalence means there exists an h such that

$$F_{h(i)} = F'_i$$

$$(IA(x))_{h(i)} = (I'A'(x))_i \qquad \forall x \in \Sigma^*$$

when

$$(I'A'(x))_i F_i \neq 0$$

Hence

$$\sum_{i=1}^{n} (IA(x))_{h(i)}F_{h(i)} = \sum_{i=1}^{n} (I'A'(x))_{i}F_{i}$$

or

$$E_{A}(x) = E_{A'}(x)$$

which is expectation equivalence. For any finite N

$$F_{h(i)}^{N} = (F_{i}')^{N}$$

The fact that

$$\mu_{N}^{A}(x) = \mu_{N}^{A'}(x)$$

comes from inspection of Theorem 2.1. Symbolically, we have shown $A \equiv_{D} A' \Rightarrow A \equiv_{N} A'$ for any N. How close one can come to a converse to Theorem 2.2 depends on the form of the entries of F.

Lemma 2.1 (Gantmacher [1959])

Given a sequence $S_{o}, S_{1}, \ldots$ of real numbers S, if one determines positive numbers

$$r_{1} > 0, r_{2} > 0, \ldots, r_{n} > 0$$

$$\infty > V_{m} > V_{m-1}, \ldots, V_{1} > 0$$

such that the following equations hold

$$S_{p} = \sum_{j=1}^{m} r_{j}V_{j}^{p} \ (p = 0, 1, 2, \ldots) \qquad (*)$$

then the solution to (*) is unique. We can apply the lemma to get the following partial converse.

Theorem 2.3

If machines A and A' meet the following requirements (Letting $h(i) = i$ W. L. G.)

(i) $(IA(x))_i F_i = 0$ iff $(I'A'(x))_i F_i' = 0 \quad i = 1, 2, \ldots, n.$

(ii) All states in a given machine have distinct output symbols.

(iii) $E_A(x) = E_{A'}(x) \qquad \forall x \in \Sigma^*$

$$\mu_i^A(x) = \mu_i^{A'}(x) \qquad i = 2, 3, \ldots$$

Then A and A' are distribution equivalent.

Proof: We use Lemma 2.1.

Since the central moments and expected values of output are equal for any string, the moments of $0_A^*(x)$ and $0_{A'}^*(x)$ about zero are equal for any string.

$$s_0 = \Sigma_i[(IA(x))_i \text{ such that } F_i \neq 0]$$

$$s_1 = E_A(x) = E_{A'}(x)$$

$$s_2 = \mu_2^A(x) + E_A(x)^2 = \mu_2^{A'}(x) + E_{A'}(x)^2$$

$$\vdots$$

We discard those components whose contribution to the moment is zero and relabel the non-zero components by the index j. Let

$$J = \{i : IA(x)_i F_i \neq 0\}$$

Because of assumption (i) we also have

$$J = \{i : I'A'(x)_i F_i' \neq 0\}$$

Hence

$$s_P = \sum_{j \in J} (IA(x))_j (F_j)^P \qquad P = 0, 1, 2, \ldots$$

$$= \sum_{j \in J} (I'A'(x))_j (F_j')^P \qquad P = 0, 1, 2, \ldots$$

By the lemma the solution is unique.

$$(IA(x))_j = (I'A'(x))_j \qquad j \in J$$

$$F_j = F_j'$$

Therefore A and A' are distribution equivalent.

Example 2.2

The condition (ii) of Theorem 2.3 is necessary as shown by the following:

$$IA(x) = (.5, .3, .2) \qquad F = F' = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$$

$$I'A'(x) = (.5, .4, .1)$$

$$E_A(x) = IA(x)F = .5$$

$$E_{A'}(x) = I'A'(x)F' = .5$$

Since A and A' are Rabin automata, by Corollary 2.3

$$\mu_i^A(x) = \mu_i^{A'}(x) \qquad i = 2, 3, \ldots$$

However, A and A' have different distributions over states for the string x.

2.6 THE N-MOMENT REDUCTION RELATION

Definition 2.5: The N-moment reduction relation R_N: xR_Ny if for all I in S

$$E_A(xz) = E_A(yz) \quad \text{and} \quad \mu_i^A(xz) = \mu_i^A(yz) \quad \forall z \in \Sigma^*, \quad i = 2, 3, \ldots, N$$

The relation R_N is a congruence relation and $R_E = R_N$ for $N = 1$.

2.7 INPUT-STATE CALCULABLE MACHINES

A probabilistic sequential machine has randomness associated with its switching, or state transitions, and a deterministic output function O. For some problems it is convenient to view a randomly behaving machine as having deterministic switching but a random output device. We study now connections between these two viewpoints.

Definition 2.6: (Carlyle [1965]) A machine A is input-state calculable if knowing the state at time t and input at time t, the state at time t+1 can be calculated by a deterministic function.

As Carlyle has pointed out, the class of finite input-state calculable machines consists of exactly those machines which have finite deterministic switching and random outputs.

Definition 2.7: A random output depending on the state S_i with parameters of $r_1, \ldots, r_k$ will be written

$$S_i : \boxed{r_1, \ldots, r_k}$$

2.8 CHARACTERIZATION OF INPUT-STATE CALCULABLE MACHINES EQUIVALENT BY $\equiv_N$ TO PROBABILISTIC SEQUENTIAL MACHINES

We obtain a generalization of Theorem 1.3.

Theorem 2.4

Let R_N be the N-moment reduction relation defined by a probabilistic sequential machine A.

Rank $|R_N| = r$ finite $\Longleftrightarrow$ there exists an input-state calculable machine A' such that $A \equiv_N A'$

Proof: Using the quotient construction of Theorem 1.1, obtain an $A'' = \Sigma^*/R_N$ where

$$A'' = \langle R_N[\Lambda], R_N[x], M[R_N[x], \sigma] \rangle$$

and M is analogous to the function M in Theorem 1.3. Elements in the same congruence class of R_N have expectations and the first N-1 central moments equal. Hence the machine A'' can have random devices attached to the states (which are $R_N[x]$) such that the first N-1 central moments and expectation of each device is the same as the congruence class represented by the state. The resulting machine A' has deterministic switching and random output functions and is equivalent by $\equiv_N$ to the probabilistic machine defining R_N.

The details of the proof parallel the proof of Theorem 1.3.

Q. E. D.

2.9 A NECESSARY AND SUFFICIENT CONDITION FOR THE N-MOMENT REDUCTION RELATION TO HOLD

In the previous section we have seen the importance of the N-moment reduction relation R_N in characterizing those probabilistic sequential machines for which there is an input-state calculable machine equivalent by $\equiv_N$. Let us now obtain invariant subspace conditions for strings to be in the same class, analogous to those of the theorems of Chapter 1.

Theorem 2.5

$$x \, R_N \, y \Longleftrightarrow A(x) = A(y) + \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix}$$

where $\langle\{h_1, \ldots, h_n\}\rangle \subset \bigcap_{i=1}^{N} \text{Kern.}(F^i)$

and $\langle\{h_1, \ldots, h_n\}\rangle \, A(z) \subset \bigcap_{i=1}^{N} \text{Kern.}(F^i) \qquad \forall z \in \Sigma^*$

Proof: Suppose that R_N holds for x and y

$$E_A(x) = E_A(y)$$

$$\Longleftrightarrow IA(x)F = IA(y)F \qquad \forall I \in S$$

$$\Longleftrightarrow A(x) = A(y) + \begin{vmatrix} r_1 \\ \vdots \\ r_n \end{vmatrix} \qquad r_j \in \text{Kern.}(F) \quad j = 1, 2, \ldots, n$$

$$\mu_2^A(x) = IA(x)(F^2) - E_A(x)^2$$

$$\mu_2^A(y) = IA(y)(F^2) - E_A(y)^2$$

$$IA(x)(F^2) = IA(y)(F^2) \qquad \forall I \in S$$

$$A(x) = A(y) + \begin{vmatrix} r_1 \\ \vdots \\ r_n \end{vmatrix} \quad r_j \varepsilon \text{ Kern.}(F^2) \quad j = 1, 2, \ldots, n$$

For any i, $\mu_i^A(x)$ can be written as a recursive function of $IA(x)(F^i)$ and smaller powers of F, i.e.,

$$\mu_i^A(x) = IA(x)(F^i) + \sum_{k=1}^{i-1} (-1)^k \binom{i}{k} IA(x)(F^{i-k})E_A(x)^k + (-1)^i E_A(x)^i$$

Hence by induction we assume

$$IA(x)(F^k) = IA(y)(F^k) \quad k = 1, 2, \ldots, i-1; \quad \forall I \varepsilon S \qquad (1)$$

Hence

$$\mu_i^A(x) = IA(x)(F^i) + \beta$$

$$\mu_i^A(y) = IA(y)(F^i) + \beta$$

$$\mu_i^A(x) = \mu_i^A(y) \qquad IA(x)(F^j) = IA(y)(F^j) \quad \forall I \varepsilon S^+ \quad j \leq i \qquad (2)$$

$$\mu_i^A(x) = \mu_i^A(y) \Longleftrightarrow A(x) = A(y) + \begin{pmatrix} r_1 \\ \vdots \\ r_n \end{pmatrix} \text{ where } r_j \varepsilon \text{ Kern.}(F^i) \quad j = 1, 2, \ldots, n$$

which completes the induction.

The rest of the proof is analogous to Theorem 1.5.

Q. E. D.

If we substitute R_N for R_E and $\bigcap_{i=1}^{N}$ Kern.(F^i) for Kern.(F), the proofs of Theorems 1.4, 1.6, 1.8, and 1.9 go through exactly as before and we state the dual theorems which are obtained.

Theorem 1.4D

If R_N has finite rank r there exists a partition $\pi = (\pi_1, \ldots, \pi_r)$

on $V(A)$ and an integer valued function $g(i,m)$ such that $\pi_i A(\sigma) \subset \pi_{g(i,\sigma)}$

$i = 1, 2, \dots, r \qquad \sigma \in \Sigma.$

Theorem 1.6D

Let $U = \left\langle \bigcup_{x \in \Sigma^*} \{A(x)_i - A(y)_i\} \quad i = 1, 2, \dots, n \quad \text{and } x \, R_N \, y \right\rangle$ then

for any $z \in \Sigma^*$

$$UA(z) \subset \bigcap_{i=1}^{N} \text{Kern.}(F^i)$$ and there exists V a subspace of R^n

such that for any $\sigma \in \Sigma$

(i) $UA(\sigma) \subset V$

(ii) $VA(\sigma) \subset V \subset \bigcap_{i=1}^{N} \text{Kern.}(F^i)$

Theorem 1.8D

R_N is non-trivial $\Longleftrightarrow$ $(\exists V)$ a subspace of R^n such that

(i) $V \subset \bigcap_{i=1}^{N} \text{Kern.}(F^i)$

(ii) V is invariant under $A(\sigma) \qquad \forall \sigma \in \Sigma$

(iii) $A(x) = A(y) + H$ where $H_i \in V$

some $H_i \neq 0$

Theorem 1.9D

R_N is non-trivial $\Longrightarrow$ there exists a subspace V such that the symbol matrices $A(\sigma) : \sigma \in \Sigma$ be reducible for the same change of basis for V i.e. there exists a linear transformation W from the state basis S to a basis for V such that

$$W^{-1}A(\sigma)W = \begin{vmatrix} \overbrace{A_1^{\sigma}}^{\text{basis for } V} & 0 \\ & \\ A_2^{\sigma} & A_3^{\sigma} \end{vmatrix}$$

where 0 denotes a block of all zeroes the same size for all symbols i and

$$V \subset \bigcap_{i=1}^{N} \mathrm{Kern.}(F^{i})$$

Example 2.3.

We extend Example 1.2 to illustrate Theorems 2.4, 2.5 and 1.8D.

$$\langle\{(0, 0, p, 0, -p, 0)\}\rangle \in \bigcap_{n=1}^{N} \mathrm{Kern.}\begin{pmatrix}10^n\\ 5^n\\ 1^n\\ 2^n\\ 1^n\\ 2^n\end{pmatrix}$$

for any finite N.

Note that in this case that the classes of R_E are also the classes of R_N. Hence we can replace the output from any state of the machine in Figure 1.2 with a random device possessing the same first N central moments as the probabilistic sequential machine. Let us compute the variances.

$$\overset{\Lambda}{\mu}_2(\Lambda) = (8/10, 1/10, 1/10, 0, 0, 0)\begin{pmatrix}100\\ 25\\ 1\\ 4\\ 1\\ 4\end{pmatrix} - (8.6)^2$$

$$= 8.84$$

Likewise, we get

$$\overset{\Lambda}{\mu}_2(0) = 1.44$$

$$\overset{\Lambda}{\mu}_2(1) = .09$$

$$\overset{\Lambda}{\mu}_2(01) = .09$$

$$\overset{\Lambda}{\mu}_2(10) = .09$$

$$\mu_2^A(11) = (0, 0, 9/40, 0, 31/40, 0) \begin{pmatrix} 100 \\ 25 \\ 1 \\ 4 \\ 1 \\ 4 \end{pmatrix} - (1.0)^2$$

$$= 0.0$$

$$\mu_2^A(010) = (0, 0, 21/320, 0, 11/320, 72/80) \begin{pmatrix} 100 \\ 25 \\ 1 \\ 4 \\ 1 \\ 4 \end{pmatrix} - 3.6.$$

$$= 0.0$$

$$\mu_2^A(0101) = (72/80, 0, 53/1280, 0, 75/1280, 0) \begin{pmatrix} 100 \\ 25 \\ 1 \\ 4 \\ 1 \\ 4 \end{pmatrix} - (9.1)^2$$

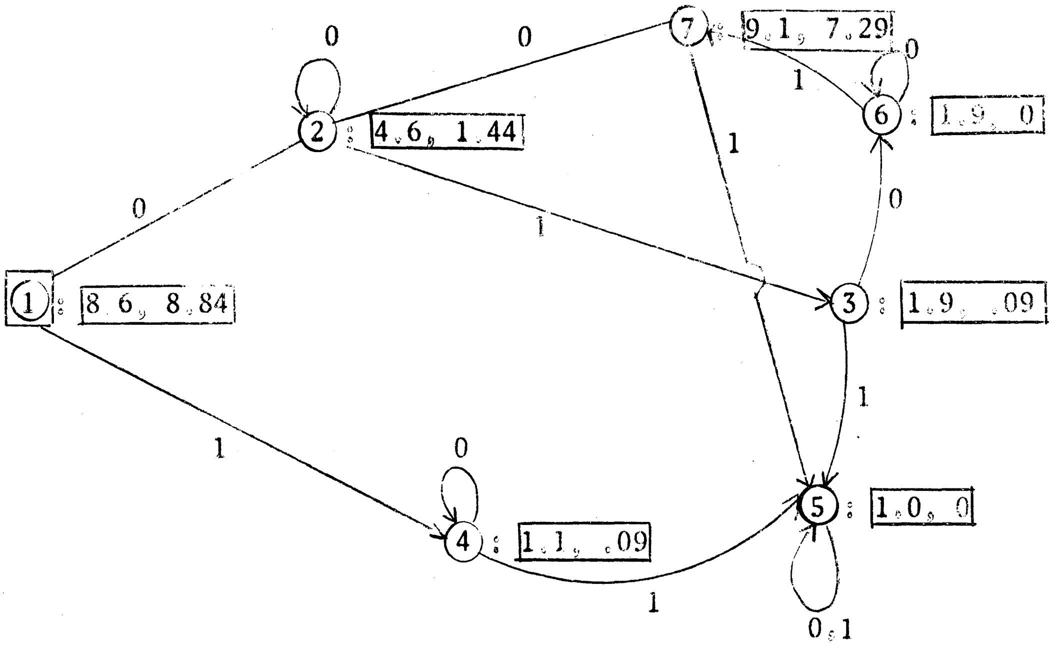
$$= 7.29$$

A machine A' which has the same expectation and variance for each string and deterministic switching will be constructed using random output devices symbolized by

$$S' : \boxed{e, V}$$

attached to states S' which supply random numbers with mean e and variance V.

The machine A', shown in Figure 2.1, is the machine of Example 1.2 with the outputs connected to random devices such as the above rather than deterministic outputs.



① is the initial state of A'

Figure 2.1 Input-state calculable machine A' which has the same expectation and variance for all strings as probabilistic machine A of Example 1.2.

CHAPTER 3

THE NOTION OF INDISTINGUISHABILITY AS A CRITERION OF BEHAVIORAL EQUIVALENCE

Suppose probabilisitic sequential machines A and A' are behaviorally equivalent in an intuitive sense. Taking into consideration how machines are built and repaired, one would expect them to be interchangeable as submachines of any larger machine. Indistinguishability of two machines in any machine into which they can be plugged is a strong criterion, the ramifications of which will be investigated. The following example of Arnold [1964] illustrates how the notion of distribution equivalence, $\equiv_D$, fails to meet the interchangeability requirement.

3.1 EXAMPLE OF TWO DISTRIBUTION EQUIVALENT MACHINES WHICH ARE NOT INTERCHANGEABLE AS COMPONENTS OF A MACHINE

$$A_1 = \langle I_1, A_1(0), A_1(1), F_1 \rangle$$

$$A_2 = \langle I_2, A_2(0), A_2(1), F_2 \rangle$$

where $I_2 = I_1$, $F_2 = F_1$

$$A_1(0) = A_1(1) = \begin{pmatrix} 0 & 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$$F_1 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 0 \end{pmatrix} \qquad I_1 = (1, 0, 0, 0, 0)$$

$$A_2(0) = A_2(1) = \begin{pmatrix} 0 & 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Machines A_1 and A_2 happen to be independent of the input i.e. are Markov provesses since $A_1(0) = A_1(1)$ and $A_2(0) = A_2(1)$.

TABLE 3.1

COMPARISON OF MACHINES A_1 AND A_2

x	$E_{A_1}(x)$	$I_1A_1(x)$	$E_{A_2}(x)$	$I_2A_2(x)$
	0	(1, 0, 0, 0, 0)	0	(1, 0, 0, 0, 0)
0 or 1	1/2	(0, 1/2, 1/2, 0, 0)	1/2	(0, 1/2, 1/2, 0, 0)
00, 01, 10 or 11	1/2	(0, 0, 0, 1/2, 1/2)	1/2	(0, 0, 0, 1/2, 1/2)
all x: $\ell g(x) \geq 3$	0	(0, 0, 0, 0, 0)	0	(0, 0, 0, 0, 0)

Table 3.1 establishes that $A_1 \equiv_D A_2$. Later a machine will be shown which behaves differently with A_1 as a submachine than it does with A_2 as a submachine even though the state behaviors of A_1 and A_2 are Markov processes.

Definition 3.1: $A \longrightarrow B$ denotes the machine obtained from plugging the output of A into the input of B, subject to the provision that the input symbols of B include the output symbols of A.

Definition 3.2: The set of tapes accepted by machine A with cutpoint λ written $T(A,\lambda)$: $T(A,\lambda) = \{x : E_A(x) \geq \lambda\}$

Definition 3.3: A and A' are tape equivalent machines, written $A \equiv_T A'$ if for some specified λ_1 and λ_2

$$T(A,\lambda_1) = T(A',\lambda_2)$$

Definition 3.4: A and A' are tape indistinguishable for a class of machines if

$$T(A\to C,\lambda) = T(A'\to C,\lambda)$$

for all λ and C ε $\mathcal{C}$.

The class $\mathcal{C}$ could be something more special than finite deterministic or probabilistic automata, e.g. the class of definite automata.

Theorem 3.1

If probabilistic sequential machines A and A' are distribution equivalent they are not necessarily tape-indistinguishable for the class of finite deterministic automata.

Proof: (by example): Let C be a finite deterministic machine which accepts 01, 10 with probability 1 and all other tapes with probability 0. We tabulate the expectations of $A_1 \to C$ and $A_2 \to C$ in Table 3.2.

TABLE 3.2

EXPECTATION OF $A_1 \to C$ AND $A_2 \to C$
FOR STRINGS x OF LENGTH 2.

y	$P_{I_1}^{A_1}(y/x)$	$E_{A_1} \to C^{(x)}$	$P_{I_2}^{A_2}(y/x)$	$E_{A_2} \to C^{(x)}$
00	0	0	1/4	0
01	1/2	1/2	1/4	1/4
10	1/2	1/2	1/4	1/4
11	0	0	1/4	0

Hence $T(A_1\to C,\lambda) \neq T(A_2\to C,\lambda)$ for any λ ε (1/2, 1/4). The reason for this difference is that the conditional probabilities of output random variables differ for A_1 and A_2. For example,

$$\text{Prob. } \{0^*_{A_1}(01) = 1\} = 1 \text{ given } 0^*_{A_1}(1) = 0$$

While

$$\text{Prob. } \{0^*_{A_2}(01) = 1\} = 1/2 \text{ given } 0^*_{A_2}(1) = 0$$

Theorem 3.2

For probabilistic sequential machines A and A', if for all finite deterministic machines C and any cutpoint λ:

$$T(A\to C,\lambda) = T(A'\to C,\lambda)$$

$$\Longrightarrow A \equiv_E A'$$

Proof: Suppose $E_A(x) \neq E_{A'}(x)$ for some tape x of length k. Without loss of generality pick $E_A(x) > E_{A'}(x)$. Since the rationals are dense in the reals, let λ_c be a rational such that $E_A(x) > \lambda_c > E_{A'}(x)$. Let C be a deterministic machine which beginning at time k computes the number $i_k-\lambda_c$ where i_k is the input at time k. Since λ_c is rational C needs only a finite number of states. C accepts the string x iff $i_k-\lambda_c \geq 0$, which can be done in a finite number of steps.

$$x \in T(B\to C,\lambda_c) \text{ iff } E_{B\to C}(x) \geq \lambda_c$$

but since C is deterministic

$$x \in T(B\to C,\lambda_c) \text{ iff } E_B(x) \geq \lambda_c$$

hence let B = A and B = A':

$$x \in T(A\to C,\lambda_c) \quad \text{and} \quad x \notin T(A'\to C,\lambda_c)$$

so

$$T(A\to C,\lambda_c) \neq T(A'\to C,\lambda_c)$$

By logical equivalence we have shown for the class C of finite deterministic machines

$$(\lambda)(C)[T(A\to C,\lambda) = T(A'\to C,\lambda)] \Longrightarrow (x)[E_A(x) = E_{A'}(x)]$$

Q. E. D.

By the example presented in Theorem 3.1 we know the converse is not true.

3.2 A MORE SATISFACTORY TECHNICAL NOTION OF INDISTINGUISHABILITY

The example at the beginning of this section shows that machine equivalences such as $\equiv_T$ equivalence and even distribution equivalence, $\equiv_D$, break down under composition of machines.

To obtain a more satisfactory definition of behavioral equivalence, the conditional probability structure of probabilistic sequential machines must be explored. A stronger concept of equivalence, called indistinguishability, based upon equality for the two machines of the probabilities of all possible output strings given all possible input strings will be formulated, following the development of Carlyle [1961].

In what follows it is assumed that Σ contains a string of one symbol Λ so that $A(\Lambda) = E(n)$, the n-dimensional matrix identity.

<u>Definition 3.5</u>: The conditional probability for a sequence of outputs $y = y_1 y_2 \ldots y_m$ given a sequence of inputs $x = \sigma_1 \ldots \sigma_m$ starting from an initial distribution $\Pi = (\Pi_1, \Pi_2, \ldots, \Pi_n)$ of a machine A will be written

$$P_{\Pi}^{A}(y/x)$$

or if the machine involved is clear from context, just $P_{\Pi}(y/x)$. Table 3.2 shows how machines A_1 and A_2 differ with respect to Definition 3.5.

The symbols of the output alphabet are real numbers which occur as components of the output column vector F, i.e. the output alphabet Y can be written

$$Y = \bigcup_{i=1}^{n} \{F_i\}$$

As usual, the set of all finite sequences of symbols from Y will be denoted by Y^*.

Definition 3.6: The probability of a sequence of transitions $S_{i_1} \to S_{i_2} \to \ldots \to S_{i_j}$ with output sequence y because of input sequence x will be written

$$P_{S_{i_1} \to \ldots \to S_{i_j}}(y/x)$$

Definition 3.7: The conditional probability transition matrix $A(y_i/\sigma)$ is formed from $A(\sigma)$ by zeroing out all columns except those corresponding to states with output y_i. More formally,

Let

$$J_{y_i} = \{j : F_j = y_i\} \quad y_i \in Y$$

and let Q^{y_i} be the matrix with $[Q^{y_i}]_{j,j} = 1$ for $j \in J_{y_i}$ and $[Q^{y_i}]_{k,j} = 0$ otherwise. Then $A(y_i/\sigma) = A(\sigma)Q^{y_i}$ $y_i \in Y$, $\sigma \in \Sigma$. Note that $[A(y_k/\sigma)]_{i,j}$ is just $P_{S_i \to S_j}(y_k/\sigma)$.

Remark 3.6: Let $y \in Y^*$, $x \in \Sigma^*$, $y_i \in Y$, $\sigma \in \Sigma$ such that $\ell g.(y) = \ell g.(x)$. Then

$$A(yy_i/x\sigma) = A(y/x)A(y_i/\sigma)$$

By definition $[A(yy_i/x\sigma)]_{\ell,m}$ is $P_{S_\ell \to S_m}(yy_i/x\sigma)$

For any state S_k

$$P_{S_\ell \to S_m}(yy_i/x\sigma) = P_{S_\ell \to S_k}(y/x)P_{S_k \to S_m}(y_i/\sigma)$$

Since transitions to different states S_k are mutually exclusive events

$$P_{S_\ell \to S_m}(yy_i/x\sigma) = \sum_{k=1}^{n} P_{S_\ell \to S_k}(y/x)P_{S_k \to S_m}(y_i/\sigma)$$

using the definitions again

$$[A(yy_i/x\sigma)]_{\ell,m} = \sum_{k=1}^{n} [A(y/x)]_{\ell,k}[A(y_i/\sigma)]_{k,m}$$

or in matrix form

$$A(yy_i/x\sigma) = A(y/x)A(y_i/\sigma)$$

Hence the conditional probability transition matrices for output strings given input strings can be generated by the conditional probability transition matrices for output symbols given input symbols, analogous to the case for the transition matrices $A(x)$.

Remark 3.7: Given initial distribution over states Π, the probability of getting output string y from input string x is just

$$P^A_\Pi(y/x) = \sum_{j=1}^{n} \sum_{i=1}^{n} \Pi_i [A(y/x)]_{i,j}$$

with $U = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$ we can write

$$P^A_\Pi(y/x) = \Pi A(y/x)U$$

Remark 3.8: We note the following identity

$$P^A_\Pi(y/x) = \sum_{y_i \varepsilon Y} P^A_\Pi(yy_i/x\sigma) \text{ for all } \sigma \varepsilon \Sigma$$

since

$$\sum_{y_i \varepsilon Y} P^A_\Pi(yy_i/x\sigma) = \sum_{y_i \varepsilon Y} \Pi A(y/x)A(y_i/\sigma)U$$

$$= \Pi A(y/x) \sum_{y_i \varepsilon Y} A(y_i/\sigma)U = \Pi A(y/x)A(\sigma)U$$

But for any n x n stochastic row matrix C

$$CU = U$$

Hence

$$\Pi A(y/x)A(\sigma)U = \Pi A(y/x)U = P^A_\Pi(y/x)$$

Definition 3.8: The terminal distribution $\Pi^*(y/x)$ for a sequence of outputs y given inputs x (assuming $P^A_\Pi(y/x) > 0$)

$$\Pi^*(y/x) = \frac{\Pi A(y/x)}{\Pi A(y/x)U}$$

The i'th component of $\Pi^*(y/x)$ is the probability of being in state i after input string x has occurred and output string y has been observed.

The following identity holds whenever $P^A_\Pi(y/x) > 0$.

$$P^A_\Pi(yy_i/x\sigma) = P^A_\Pi(y/x)P^A_{\Pi^*(y/x)}(y_i/\sigma) \qquad y_i \in Y, \ \sigma \in \Sigma \quad x \in \Sigma^*, \ y \in Y^*$$

Definition 3.9: Machines A and A' are indistinguishable written $A \equiv_I A'$ if

$$P^A_\Pi(y/x) = P^{A'}_{\Pi'}(y/x) \qquad \forall x \in \Sigma^*, \quad \forall y \in Y^*$$

The concept of indistinguishability for machines depends on observable identity when both machines are started from their initial state distributions.

Definition 3.10: Machines A and A' are k-indistinguishable if

$$P^A_\Pi(y/x) = P^{A'}_{\Pi'}(y/x) \qquad x \in (\Sigma)^m, \quad y \in (Y)^m \qquad \text{for} \quad m = 0, 1, \ldots, k.$$

Definition 3.11: In a machine A, two initial state distributions Π and λ are indistinguishable if

$$P^A_\Pi(y/x) = P^A_\Pi(y/x) \qquad \forall y \in Y^*, \quad \forall x \in \Sigma^*$$

Definition 3.12: In a machine A, two initial state distributions Π and λ are k-indistinguishable if

$$P^A_\Pi(y/x) = P^A_\lambda(y/x) \qquad \forall x \in (\Sigma)^k, \quad \forall y \in (Y)^k$$

Checking whether the indistinguishability definition (3.9) for machines or for initial distributions (3.11) holds using only the definitions involves calculation of an unbounded sequence of conditional probabilities. In the next section is shown a bound for the length of strings whose probabilities need to be calculated. If n is the number of states, then only strings of length n-1 or less need be considered in establishing indistinguishability.

3.3 THE RELATIONSHIP BETWEEN THE INTUITIVE AND TECHNICAL CONCEPTS OF INDISTINGUISHABILITY

We have yet to relate the intuitive notion of indistinguishability to the technical Definition 3.9. The next theorem shows that two machines indistinguishable in the technical sense are indeed indistinguishable when plugged into C, any finite state probabilistic or deterministic machine. Since C has a finite number of states, it is assumed that finite strings of $Z = C(Y^*)$, the random variable taking on values of strings of outputs of C given strings of inputs from the random variable Y, depend only on finite strings Y^*.

Theorem 3.4

Let C^* be the class of finite state probabilistic and deterministic sequential machines. For any $C \varepsilon C^*$

$$\text{If} \quad A \equiv_I A' \quad \text{then } A \rightarrow C \equiv_I A' \rightarrow C$$

Proof: For any fixed value y of the output string random variable of A, Y_A

$$P_{\Pi}^{A\rightarrow C}(z = C(y)/x) = P_{\Pi}^{A}(y/x)P^{C}(z = C(y)/y)$$

since the occurence of different y are disjoint events, for all $y \varepsilon Y^*$: $\ell g(y) = \ell g(x)$.

$$P_{\Pi'}^{A\rightarrow C}(z = C(y)/x) = \sum_{y \varepsilon (Y) \ell g(x)} P_{\Pi'}^{A'}(y/x)P^{C}(z' = C(y)/y)$$

Since Z and Z' range over the same set and the indistinguishability of A and A'

$$P^{A}_{\Pi}(y/x) = P^{A'}_{\Pi'}(y/x)$$

So for all $x \in \Sigma^*$ and all $z \in Y^*_C$

$$P^{A\to C}_{\Pi}(z = C(y)/x) = P^{A'\to C}_{\Pi'}(z = C(y)/x)$$

which means $A\to C$ and $A'\to C$ are indistinguishable.

Q. E. D.

Since the machine C might ignore its inputs, it is clear that the converse to Theorem 3.4 does not hold.

The criterion of interchangeability as a submachine has lead us to $\equiv_I$ as a behavioral equivalence for probabilistic sequential machines. The equivalence $\equiv_I$ is well known as an equivalence between communication channels. The other kinds of equivalences discussed are equally valid for channels with numerically coded outputs.

The relationship between the equivalences $\equiv_T$, $\equiv_E$, $\equiv_N$, $\equiv_D$, and $\equiv_I$ can be summarized in the following schematic way:

$$\begin{tikzcd}
A \equiv_D A' \arrow[dr, Rightarrow] & & & \\
& A \equiv_N A' \arrow[r, Rightarrow] & A \equiv_E A' \arrow[r, Rightarrow] & A \equiv_T A' \\
A \equiv_I A' \arrow[ur, Rightarrow] & & &
\end{tikzcd}$$

As we have seen in previous chapters, the concepts of behavioral equivalence for probabilistic machines analogous to those of deterministic machine theory depend on the device being modeled. Consequently, applications of probabilistic sequential machines to new domains are likely to suggest new kinds of behavioral equivalences.

CHAPTER 4

FINITE COMPLETE SETS OF INVARIANTS FOR THE BEHAVIORAL EQUIVALENCES $\equiv_E$, $\equiv_N$, AND $\equiv_I$ AND THE REDUCTION CONGRUENCE RELATIONS R_E AND R_N

The results of the previous chapters involve relations defined over all finite strings of the input alphabet. In this section are found bounds for the length of strings necessary to consider in order to decide whether two elements of the domains of the relations are in the same class.

Definition 4.1: A set of functions $\mathcal{F} = \{f_1, \ldots, f_m\}$ is a set of invariants for the relation R if for all x and y in the domain of R

$$xRy \Longrightarrow f_i(x) = f_i(y) \qquad i = 1, \ldots, m$$

The set of functions $\mathcal{F}$ is a complete set of invariants if $xRy \Longleftrightarrow f_i(x) = f_i(y) \quad i = 1, \ldots, m$

We exhibit sets of functions which are invariants for the above relations. A set of functions which are invariant over R_E and R_N are:

$$\begin{aligned} &f_{(A,1,z)}(x) = E_A(xz) \\ &\quad\vdots \qquad\qquad\qquad \text{for all } z;\ \ell g(z) \leq i, \quad \text{for all } I \in S \\ &f_{(A,N,z)}(x) = \mu_N^A(xz) \end{aligned}$$

While for the relation $\equiv_I$, the set of functions below is a set of invariants:

$$g_{(x,y)}(A) = P_\Pi^A(y/x) \quad \text{for all x and y; } \ell g(x) = \ell g(y) \leq i$$

Likewise the set

$$h_{(x,1)}(A) = E_A(x) \quad \text{for all x: } \ell g(x) \leq i$$

$$h_{(x,r)}(A) = \mu_r^A(x) \quad \text{for } r = 2, \ldots, N$$

is a set of invariants for the relations $\equiv_E$ and $\equiv_N$.

It is clear that for an unbounded i, the above are complete sets of invariants. However, in what follows a finite value of i will be found for each of these cases. In the case of $\equiv_E$ the bound will be the same as the well known Moore bound for deterministic automata but in the case of $\equiv_N$ it will be lower for most machines. The main tool used in finding the various values of i is the following simple lemma.

4.1 THE FUNDAMENTAL LEMMA

Lemma 4.1

Given an n-dimensional vector space V, a finite set $T = \{T_i\}$ where each $T_i \varepsilon V \times V$ is a linear transformation on V and some finite set of vectors $V_0 \subset V$ such that $\dim \langle V_0 \rangle = r \geq 1$.

Define

$$M_0 = V_0$$

$$M_1 = \{v_0 \cdot T_i : T_i \varepsilon T, \forall v_0 \varepsilon V_0\}$$

$$\vdots$$

$$M_k = \{v_0 \cdot T_{i_1} \ldots T_{i_k} : T_{i_1}, \ldots, T_{i_k} \varepsilon T, v_0 \varepsilon V_0\}$$

and let

$$L_i = \left\langle \bigcup_{j=0}^{i} M_j \right\rangle$$

Then there exists an integer J(T) such that

(i) $L_{J(T)} = L_{J(T)+1}$

(ii) $L_{k-1} \subsetneqq L_k$ for $k \leq J(T)$

(iii) $J(T) \leq n-r$

Proof: $L_0 \subset L_1 \subset \ldots \subset L_i \subset \ldots \subset L_k$ as a consequence of the definition. The sequence $\{\dim L_j\}_{j=0}^{\infty}$ is bounded above by n, the dimension of V.

Call $J(T)$ the smallest index k such that $L_{k+1} = L_k$. Showing that the sequence $\{\dim L_j\}_{j=0}^{J(T)}$ is strictly increasing requires that for all $j+1 < J(T)$

$$L_{j+1} \neq L_{j+2} \Longrightarrow L_{j+1} \neq L_j$$

which is logically equivalent to

$$L_{j+1} = L_j \Longrightarrow L_{j+2} = L_{j+1}$$

Hence it is sufficient to show

$$L_{j+1} \subset L_j \Longrightarrow L_{j+2} \subset L_{j+1}$$

Assume

$$L_{j+1} \subset L_j$$

W.L.G. pick

$$v = v_0 \cdot T_{i_1} \ldots T_{i_{j+2}} \quad \varepsilon \; L_{j+2} = (v_0 \cdot T_{i_1} \ldots T_{i_{j+1}}) \cdot T_{i_{j+2}}$$

But

$$w = v_0 \cdot T_{i_1} \ldots T_{i_{j+1}} \quad \varepsilon \; L_{j+1}$$

So there is a finite set of indices $I = \{i\}$ of a spanning set U' for L_j

$$U' = \{v_0 \cdot T_{B^i_1} \cdot T_{B^i_2} \ldots T_{B^i_{r_i}} : i \; \varepsilon \; I\}$$

such that $r_i \leq j$ and constants c_i

$$w = \sum_{i \varepsilon I} c_i (v_0 T_{B^i_1} \ldots T_{B^i_{r_i}})$$

so

$$v = w \cdot T_{i_{j+2}} = \sum c_i (v_0 T_{B^i_1} \ldots T_{B^i_{r_i}} \cdot T_{i_{j+1}}) \; \varepsilon \; L_{j+1}$$

i.e.

$$L_{j+2} \subset L_{j+1}$$

Now consider the sequence of dimensions

$$\dim L_0, \dim L_1, \ldots, \dim L_{j(T)}$$

since

$$L_k \subsetneqq L_{k+1} \quad \text{for} \quad k+1 \leq J(T), \quad \dim L_k < \dim L_{k+1} \quad \text{for} \quad k+1 \leq J(T)$$

Noting that

$$\dim L_0 = r, \quad \dim L_0 + J(T) \leq \dim L_{J(T)} \leq n$$

which gives

$$J(T) \leq n - r$$

Q. E. D.

4.2 A BOUND FOR TESTING FOR MEMBERSHIP IN $\equiv_I$

Theorem 4.1

If A is a probabilistic sequential machine with n states, then (n-1)-indistinguishability of initial distributions π and π' is sufficient to guarantee indistinguishability of initial distributions π and π'.

Proof: Using Lemma 4.1 let

$$V_0 = \{U\} = \left\{ \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \right\} \text{ and } \dim \langle V_0 \rangle = 1$$

$$T = \{A(y_i/\sigma) : y_i \in Y, \quad \sigma \in \Sigma\}$$

$$V_0 \circ T_i = A(y_i/\sigma)U$$

by the lemma.

For any string $x = i_1 \ldots i_{r'}$; for r' finite, $A(y/x)U$ can be expressed as

$$A(y/x)U = \sum_{i \in I} c_i A(y_{B i_1} \ldots y_{B i_{r_i}} / \sigma_{j i_1} \ldots \sigma_{j i_{r_i}})U \qquad (*)$$

with

$$r_i \leq n - 1 \quad \text{for} \quad i \in I, \quad y_{B^i_k} \in Y, \quad \sigma_{j^i_k} \in \Sigma \qquad (\text{for } k = 1, \ldots, r_i)$$

Hence for initial distributions π and π'

$$P^A_\Pi(y/x) = \Pi A(y/x)U = \sum_{i \in I} c_i \Pi A(y_{B^i_1} \cdots y_{B^i_{r_i}} / \sigma_{j^i_1} \cdots \sigma_{j^i_{r_1}})U$$

Let

$$y^i = y_{B^i_1} \cdots y_{B^i_{r_i}} \quad \text{and} \quad x^i = \sigma_{j^i_1} \cdots \sigma_{j^i_{r_1}}$$

$$P^A_\Pi(y/x) = \sum_{i \in I} c_i P^A_\Pi(y^i/x^i) \text{ with } \ell g(y^i) = \ell g(x^i) \leq n - 1$$

multiplying (*) by π' gives

$$P^A_{\Pi'}(y/x) = \sum_{i \in I} c_i P^A_{\Pi'}(y^i/x^i)$$

By the assumption of (n-1)-indistinguishability for π and π'

$$P^A_\Pi(y^i/x^i) = P^A_{\Pi'}(y^i/x^i) \quad \ell g(x^i) = \ell g(y^i) \leq n - 1$$

Hence

$$P^A_\Pi(y/x) = P^A_{\Pi'}(y/x)$$

Q. E. D.

4.3 EQUIVALENCE OF DISTRIBUTIONS IN ONE MACHINE

Using Lemma 4.1, we can make effective the definition of the relations R_E and R_N of Chapter 2. A bound will be obtained on the lengths of strings needed for deciding whether x and y are in the same congruence class.

Definition 4.2: Distributions π and λ are expectation equivalent for a machine A, written $\pi \overset{A}{\underset{E}{\sim}} \lambda$, if $\pi A(x)F = \lambda A(x)F \quad x \in \Sigma^*$

Definition 4.3: Distributions π and λ are K-expectation equivalent for a machine A, written $\pi \overset{A}{\underset{KE}{\sim}} \lambda$, if

$$\Pi A(x)F = \lambda A(x)F \qquad x \in \Sigma^*;\ 0 \leq \ell g(x) \leq K$$

Theorem 4.2 (Generalization of the result of Paz [1964])

If A is a probabilistic sequential machine with n states and if π and λ are n-2 equivalent distributions of A then

$$\pi \overset{A}{\underset{E}{\sim}} \lambda$$

Proof: We use the elementary fact that for any constant c

$$\Pi A(x)F = \lambda A(x)F \Longleftrightarrow \Pi A(x)F + c = \lambda A(x)F + c \qquad \forall x \in \Sigma^*$$

Since $\Pi A(x)$ and $\lambda A(x)$ are stochastic

$$\Longleftrightarrow \quad \Pi A(x)[F + c\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}] = \lambda A(x)[F + c\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}] \qquad \forall x \in \Sigma^*$$

If all the entries of F are equal, then all distributions are expectation equivalent. If at least two of the entries of F are not equal, then $\langle F\rangle \neq \langle F + c\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}\rangle$. Let A' be a machine differing from A only in that $F' = F + c\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}$. Suppose we experiment with A and A' simultaneously. No new information is obtained, i.e. distributions are expectation equivalent for A iff they are expectation equivalent for A'. However, if we compute a bound for the two machine experiment using Lemma 4.1, the bound will be lower than would have been obtained from an experiment on A alone. Since the results of the two experiments are identical, the lower bound applies to A also.

$$V_0 = \{F, F + c\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}\} \qquad \dim \langle V_0\rangle = 2$$

$$T = \{A(\sigma): \sigma \in \Sigma\} \qquad V_0 \cdot T_i = A(i) \cdot V_0$$

By Lemma 4.1, there is a finite set of indices J of vectors $A(x_j)V^j$ with $V^j \varepsilon V_0$ with $\ell g(x_j) \leq n - 2$ such that for an arbitrary $x \in \Sigma^*$ there are constants c_j so that

$$A(x)F = \sum_{j \varepsilon J} c_j A(x_j)V^j$$

which reduces to

$$= \sum_{j \varepsilon J} c_j A(x_j)F + c' \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \quad \text{where } c' \text{ is a constant.}$$

Multiplying by the initial distributions gives:

$$\Pi A(x)F = \sum_{j \varepsilon J} c_j \Pi A(x_j)F + c'$$

$$\lambda A(x)F = \sum_{j \varepsilon J} c_j \lambda A(x_j)F + c'$$

(n-2)-expectation equivalence gives

$$\Pi A(x_j)F = \lambda A(x_j)F$$

so $$\Pi A(x)F = \lambda A(x)F$$

Q. E. D.

4.4 BOUNDS FOR TESTING FOR MEMBERSHIP IN $\equiv_E$ AND R_E

Definition 4.4: The *abstract join* of probabilistic sequential machines $A = \langle \pi, A(0), \ldots A(k-1), F \rangle$ with n states and $A' = \langle \lambda, A'(0), \ldots A'(k-1), F' \rangle$ with n' states is the abstract n + n' state machine $A^{\oplus}$ written

$$A^{\oplus} = A \oplus A' = \langle \;, A^{\oplus}(0), \ldots A^{\oplus}(k-1), F^{\oplus} \rangle$$

where

$$A^{\oplus}(i) = \left|\begin{array}{c|c} A(i) & 0 \\ \hline 0 & A'(i) \end{array}\right|$$

and

$$F^{\oplus} = \left|\begin{array}{c} F \\ \hline F' \end{array}\right|$$

π and λ can be embedded in the n + n' dimensional space as

$$\pi^{\oplus} = (\pi, \overbrace{0, \ldots, 0}^{n' \text{ zeroes}}) \qquad \lambda^{\oplus} = (\overbrace{0, \ldots, 0}^{n \text{ zeroes}}, \lambda)$$

The problem of deciding whether two machines A and A' are expectation equivalent:

$$\pi A(x)F = \lambda A'(x)F \qquad \forall x \in \Sigma^*$$

is logically equivalent to deciding when $\pi^{\oplus}$ and $\pi^{\oplus}$ are equivalent in $A \oplus A'$, i.e. whether

$$\pi^{\oplus} \overset{A\oplus A'}{\underset{E}{\sim}} \lambda^{\oplus}$$

Hence following Caryle [1961], we use Theorem 4.2 to state

Remark 4.1:

$$\pi^{\oplus} \overset{A\oplus A'}{\underset{E}{\sim}} \lambda^{\oplus} \Longleftrightarrow \pi^{\oplus} \overset{A\oplus A'}{\underset{KE}{\sim}} \lambda \qquad \text{where } K = n + n' - 2$$

which gives the following theorem.

Theorem 4.3

Let A and A' be probabilistic sequential machines having n and n' states respectively.

Then a necessary and sufficient condition that A and A' are expectation equivalent:

$$[\pi A(z)F = \lambda A'(z)F' \quad \forall z \in \Sigma^*] \Longleftrightarrow [\pi A(x)F = \lambda A'(x)F' \quad \forall x\colon \ell g(x) \leq n+n'-2]$$

Theorem 4.3 makes the experimental determination of expectation equivalence possible provided the number of states of each machine is known. Furthermore, it gives a bound on the process of finding whether two strings are in the same equivalence class under the reduction relation R_E of Chapter 1. This result is summarized in the following theorem.

Theorem 4.4

Strings x and y are in the same equivalence class under the reduction relation R_E of an n state probabilistic sequential machine A if and only if

$E_A(xz) = E_A(yz)$ for all strings z: $\ell g(z) \leq n - 2$ and all $I \in S$.

Proof:

$$x \, R_E \, y \Longleftrightarrow E_A(xz) = E_A(yz) \quad \text{for all } z \in \Sigma^*, \text{ for all } I \in S$$

$$\Longrightarrow IA(x)A(z)F = IA(y)A(z)F$$

Let $\pi = IA(x)$ and $\lambda = IA(y)$

$$\Longrightarrow \pi A(z)F = \lambda A(z)F \qquad \forall z \in \Sigma^*$$

By Theorem 4.2 and its obvious converse, we get

$$IA(x) \overset{A}{\underset{n-2E}{\sim}} IA(y)$$

which gives the theorem.

4.5 BOUNDS FOR TESTING FOR MEMBERSHIP IN $\equiv_N$ AND R_N

Definition 4.5: n_F = the independence number of an n state machine A with output vector F:

$$n_F = \dim \langle \{(F^i) : i = 1, 2, \ldots, n\} \rangle$$

It follows from vector space arguments that

$$n_F = \# \{F_k : F_k \neq 0\} \quad \text{where } \# \text{ is the cardinality operator on sets.}$$

The independence number is just the dimension of the space generated by powers of the components of the output vector F. For a Rabin automata $n_F = 1$ and all central moments reduce to polynominals in what we may consider the first "central moment" $E_A(x)$. In general, if the independence number is n_F, then for all x in Σ^*, the (n_F+1)'st central moment $\mu^A_{n_F+1}(x)$ reduces to a polynomial in the lower central moments since

$$\mu^A_{n_F+1}(x) = IA(x)(F^{n_F+1}) + Q(x)$$

where Q(x) is a polynomial in which $IA(x)(F^i)$, $i = 1, \ldots, n_F$ occur.

Hence

$$\mu^A_{n_F+1}(x) = IA(x) \sum_{i=1}^{n_F} c_i(F^i) + Q(x)$$

since n_F is the dimension of the space $\langle (F^i) : i = 1, 2, \ldots, n \rangle$

$$= \sum_{i=1}^{n_F} c_i IA(x)(F^i) + Q(x)$$

Theorem 4.5

Let A be a probabilistic sequential machine with output vector F and n states. Then for any $r \leq n_F$ and strings x and y in Σ^*:

$$\left\{ \left[\begin{matrix} E_A(xz) = E_A(yz) \\ \mu_2^A(xz) = \mu_2^A(yz) \\ \vdots \\ \mu_r^A(xz) = \mu_r^A(yz) \end{matrix} \quad \forall z \in \Sigma^* \right] \right\} \Longleftrightarrow \left\{ \left[\begin{matrix} E_A(xz') = E_A(yz') \\ \mu_2^A(xz') = \mu_2^A(yz') \\ \vdots \\ \mu_r^A(xz') = \mu_r^A(yz') \end{matrix} \quad \forall z' : \ell g(z') < n-r-p \right] \right\}$$

where $p = 0$ if $\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \varepsilon \langle \{F, (F^2), \ldots, (F^r)\} \rangle$, $p = 1$ otherwise.

Proof:

Let $V_0' = \{F, (F^2), \ldots, (F^r)\}$

$$V_0 = \begin{cases} V_0' \text{ if} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \varepsilon \langle V_0' \rangle \\ \\ V_0' \cup \left\{ \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \right\} \text{ otherwise} \end{cases}$$

The reason for the inclusion of $\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$ in V_0 here is the same as in Theorem 4.2, i.e. for $i = 1, 2, \ldots, r$

$$IA(x)(F^i) = IA(y)(F^i) \Longleftrightarrow IA(x)[(F^i) + c\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}] = IA(y)[(F^i) + c\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}]$$

Hence the dimension of $\langle V_0 \rangle$ is either r or r+1. We will call it r+p where p is defined in the statement of the theorem

$$\dim \langle V_0 \rangle = r+p \quad \text{where } r \leq n_F$$

$$\{T_i\} = \{A(i) \; ; \; i \; \varepsilon \; \Sigma\}$$

For any $v_0 \; \varepsilon \; \langle V_0 \rangle$ there exist constants c_k such that (defining $(F^o) = \left| \begin{matrix} 1 \\ \vdots \\ 1 \end{matrix} \right|$)

$$v_0 \circ T_i = A(i)v_0 = \sum_{k=p}^{r} c_k A(i)(F^k)$$

Consider any string

$$z: \ell g(z) = m' \text{ finite}$$

Then there exists a spanning set $A(x_j)v_0$ with $j \; \varepsilon \; J$ and constants $c_j(v_0)$ so that

$$A(z)v_0 = \sum_{j \varepsilon J} c_j(v_0) \; A(x_j)v_0 \; : \; \ell g(x_j) \leq n - r - p$$

Let v_0 range over the (F^i) $i = p, \ldots, r$ and multiply by π and λ

$$\pi A(z)(F^i) = \sum_{j \varepsilon J} c_j((F^i))\pi A(x_j)(F^{n_j}) \qquad n_j \leq r$$

$$\lambda A(z)(F^i) = \sum_{j \varepsilon J} c_j((F^i))\lambda A(x_j)(F^{n_j}) \qquad n_j \leq r$$

Since

$$\pi A(x_j)(F^{n_j}) = \lambda A(x_j)(F^{n_j}) \text{ by assumption}$$

$$\pi A(x)(F^i) = \lambda A(y)(F^i)$$

That is, the moments about zero from π and λ are equal if they are equal for all strings of length $\leq n-r-p$. Let $\pi = IA(x)$ and $\lambda = IA(y)$. Then we have for any z and any initial distribution

$$IA(xz)(F^i) = IA(yz)(F^i) \quad i = p, \ldots, r$$

holds if and only if for $i = p, \ldots, r$

$$IA(xz')(F^i) = IA(yz')(F^i)$$

for all strings z' of length less than or equal to $n-r-p$. Noting by Theorem 2.4 (equation (2)) that any central moment $\mu_m^A(x)$ is a recursive function of IA(x)(F), ..., IA(x)(F^m) the result is established.

Q. E. D.

Corollary 4.5 (Bound for testing the relation R_N)

Let A be a probabilistic sequential machine with n states and with $N \leq n_F$. $x \, R_N \, y \iff$ for all strings z' : $\ell g(z') \leq n-N-p$

$$\left[\begin{array}{l} E_A(xz') = E_A(yz') \\ \mu_2^A(xz') = \mu_2^A(yz') \\ \vdots \\ \mu_N^A(xz') = \mu_N^A(yz') \end{array}\right] \quad \text{for all } I \in S$$

Where $p = 0$ if $\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \in \langle\{F, (F^2), \ldots, (F^N)\}\rangle$, $p = 1$ otherwise.

Theorem 4.6 (Finite set of invariants for $\equiv_N$)

Suppose A and A' are probabilistic sequential machines having n and n' states respectively

$$\text{Let } p = \begin{cases} 0 \text{ if } \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \in \langle\{F^{\oplus}, (F^{\oplus^2}), \ldots, (F^{\oplus^r})\}\rangle \\ \quad \text{where} \\ \qquad r \leq n_F + n_{F'} - \# \{y : y \in Y \cap Y' \text{ and } y \neq 0\} \\ 1 \text{ otherwise} \end{cases}$$

For any initial distribution π of A and any initial distribution λ of A':

$$A \equiv_r A' \iff \left[\begin{array}{l} E_A(x') = E_{A'}(x') \\ \mu_2^A(x') = \mu_2^{A'}(x') \quad \forall x' : \ell g(x') \leq n+n'-r-p \\ \vdots \\ \mu_r^A(x') = \mu_r^{A'}(x') \end{array}\right]$$

Proof: Construct $A^{\oplus} = A \oplus A'$ and let V_0 in Lemma 4.1 be

$$V_0' = \{\begin{pmatrix}F\\F'\end{pmatrix}, \ldots, \begin{pmatrix}F^{n+n'}\\F'^{n+n'}\end{pmatrix}\} \text{ if } \begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix} \varepsilon \langle V_0' \rangle \text{ or else } V_0' \cup \left\{\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}\right\} \text{ otherwise.}$$

$$n_F^{\oplus} = \dim \langle V_0' \rangle$$

$$n_F^{\oplus} = \{\hat{y} : (\hat{y} \varepsilon Y \text{ or } \hat{y} \varepsilon Y') \text{ and } \hat{y} \neq 0 \text{ and } \hat{y} \notin Y \cap Y'\}$$

$$= n_F + n_{F'} - \#\{\hat{y} : \hat{y} \varepsilon Y \cap Y' \text{ and } \hat{y} \neq 0\}$$

Using Lemma 4.1 and an argument like the one in Theorem 4.4 establishes the theorem.

Q. E. D.

4.6 DISCUSSION OF THE GENERALIZATION OF THE MOORE BOUND

Corollary 4.6

Let A and A' be n-state deterministic machines with two-valued output alphabet $Y = Y' = \{1, 2\}$. Then A and A' are indistinguishable for all strings if they are indistinguishable for all strings of length at most $2n-2$.

Proof: In Theorem 4.6 we have $n_F^{\oplus} = 2+2-2 = 2$ and $p = 0$ so that $r \leq 2$. For deterministic machines, indistinguishability reduces to $E_A(x) = E_{A'}(x)$ for all $x \varepsilon \Sigma^*$ and also

$$E_A(x) = E_{A'}(x) \Longrightarrow \mu_2^A(x) = \mu_2^{A'}(x)$$

Hence the right side of Theorem 4.6 gives the result.

Q. E. D.

Theorem 4.6 can be regarded as a generalization of the Moore result [1956] to probabilistic machines with arbitrary rather than binary output alphabets. Note that Moore's bound is $2n-1$ since he considers the initial output as part of the experiment. We consider the initial outputs when considering strings of length 1 since the symbol Λ has identity symbol matrix.

The role of the zero output symbol in Theorem 4.6 is a significant departure from Moore's deterministic results. In order to get $p = 0$ in Corollary 4.6 we used a two-valued output set $\{1,2\}$ rather than $\{0,1\}$ with the implicit assumption that such recoding of output symbols cannot affect indistinguishability between deterministic machines. Without the recoding, $p = 1$ and the bound is still the Moore bound.

CHAPTER 5

ON A FINITE COMPLETE SET OF INVARIANTS FOR TAPE EQUIVALENCE $\equiv_T$

5.0 INTRODUCTION

Rabin [1964] and Paz [1964] have shown that some probabilistic automata can accept nonregular sets. Consequently an experiment unbounded in length may be needed to determine whether two arbitrary probabilistic sequential machines accept the same set of tapes. Only for certain classes of machines do the results of a bounded experiment provide a finite complete set of invariants for $\equiv_T$.

In this chapter will be found finite experiments for determining whether $\equiv_T$ holds and hence a finite complete set of invariants for two special classes of machines. Part I deals with machines with isolated cutpoints. Rabin has proven that an isolated cutpoint machine accepts the same set of tapes as some finite deterministic machine. This property makes such machines very special and in Part II we seek a more generally defined class of machines

Part II contains a sufficient condition for the existence of a finite bound on experiments for deciding whther $\equiv_T$ holds. A class of probabilistic machines will be presented which have a larger bound on experimentation than isolated cutpoint machines.

PART I. ISOLATED CUTPOINT MACHINES

Definition 5.1: For a probabilistic sequential machine A, a cutpoint λ is isolated if there exists a $\delta > 0$, called the separation such that

$$|E_A(x) - \lambda| \geq \delta \qquad \text{for all } x \text{ in } \Sigma^*$$

The concept of isolated cutpoint was developed by Rabin [1964] to resolve the problem of being able to decide statistically whether a machine with random switching accepts a given tape. The number of sample runs necessary to determine whether a tape x is accepted (for any fixed probability of being correct) depends on the difference between the expectation of the tape $E_A(x)$ and the cutpoint λ. But the expectations $E_A(x)$ are not known beforehand. However the assumption of a uniform bound for the difference between the expectation of any tape and the cutpoint provides a bound on the number of trial runs dependent only on the probability that the decision be correct.

No finite characterization of isolated cutpoint machines has appeared in the literature. However it is extremely easy to construct examples of such machines.

5.1 THE RABIN REDUCTION THEOREM

Rabin proved [1964] that any set of tapes accepted by an n-state probabilistic automaton with isolated cutpoint λ and separation $\delta > 0$ also can be accepted by some deterministic machine D with no more than n_D states where

$$n_D \leq (1 + \frac{1}{\delta})^{n-1}$$

Paz [1964] sharpened the bound to

$$n_D \leq (1 + \frac{1}{2\delta})^{n-1}$$

A similar situation holds for probabilistic sequential machines. In order to prepare results for the next section, the bound will be established using two theorems. The proof of the first theorem is exactly analogous to a portion of the proof of the Rabin reduction theorem, (Theorem 3, Rabin [1964]) and the improvement made by Paz [1964].

Theorem 5.1

Let A be a probabilistic sequential machine with n states. If

$E = \{e^j ; i = 1, \ldots, k\}$ is any set of distributions of A such that there exist strings z_{ij} so that

$$|e^j A(z_{ij})F - e^i A(z_{ij})F| \geq \delta$$

$$\text{for all } i \neq j;\ i, j = 1, 2, \ldots, k$$

then

$$k \leq (1 + \frac{d}{\delta})^{n-1} \quad \text{where } d = \max_i(F_i) - \min_i(F_i)$$

Proof: Let e^i and e^j be in E. Write

$$A(z_{ij})F = \begin{pmatrix} r_1 \\ \vdots \\ r_n \end{pmatrix}$$

hence

$$\delta \leq |(e^j - e^i)A(z_{ij})F|$$

$$\delta \leq |(e^j_1 - e^i_1)r_1 + \ldots + e^j_n - e^i_n)r_n| \qquad (1)$$

Following Paz [1964], we rewrite the sum as two parts:

$$\sum_{k=1}^{n} (e^j_k - e^i_k)r_k = \sum_{k \varepsilon K^+} (e^j_k - e^i_k)r_k + \sum_{k \varepsilon K^-} (e^j_k - e^i_k)r_k$$

where the sets of indices are

$$K^+ = \{k : (e^j_k - e^i_k) \geq 0\}$$

$$K^- = \{k : (e^j_k - e^i_k) < 0\}$$

We note that

$$\sum_{k \varepsilon K^+} (e^j_k - e^i_k) + \sum_{k \varepsilon K^-} (e^j_k - e^i_k) = \sum_{k=1}^{n} e^j_k - \sum_{k=1}^{n} e^i_k$$

But since e^j and e^i are stochastic

$$= 1 - 1 = 0$$

Hence

$$\sum_{k \varepsilon K^+} (e^j_k - e^i_k) = - \sum_{k \varepsilon K^-} (e^j_k - e^i_k) \qquad (2)$$

Since all terms on the left are positive and all on the right are negative

$$\sum_{k\varepsilon K^+} |e_k^j - e_k^i| = \sum_{k\varepsilon K^-} |e_k^j - e_k^i|$$

From which we get

$$\sum_{k=1}^{n} |e_k^j - e_k^i| = 2\sum_{k\varepsilon K^+} |e_k^j - e_k^i| \tag{3}$$

Using (1) and the above rewriting we get

$$\delta \leq |\sum_{k\varepsilon K^+} (e_k^j - e_k^i) \cdot r_k + \sum_{k\varepsilon K^-} (e_k^j - e_k^i) r_k|$$

Letting $F_{max} = \max_i(F_i)$, $F_{min} = \min_i(F_i)$

since $A(z_{ij})$ is a stochastic matrix

$$F_{max} \geq r_k \geq F_{min}$$

$$\delta \leq |\sum_{k\varepsilon K^+} (e_k^j - e_k^i) F_{max} + \sum_{k\varepsilon K^-} (e_k^j - e_k^i) F_{min}|$$

using (2)

$$\delta \leq |\sum_{k\varepsilon K^+} (e_k^j - e_k^i)(F_{max} - F_{min})|$$

$$\leq |\sum_{k\varepsilon K^+} (e_k^j - e_k^i)| \cdot |F_{max} - F_{min}|$$

$$\leq \sum_{k\varepsilon K^+} |e_k^j \cdot e_k^i| \cdot (F_{max} - F_{min})$$

so we get (excluding the trivial case where $F_{max} = F_{min}$)

$$\sum_{k\varepsilon K^+} |e_k^j - e_k^i| \geq \frac{\delta}{(F_{max} - F_{min})}$$

By (3)

$$2\sum_{k\varepsilon K^+} |e_k^j - e_k^i| = \sum_{k=1}^{n} |e_k^j - e_k^i| \geq \frac{2\delta}{(F_{max} - F_{min})} \tag{4}$$

From this point on the proof of Rabin [1964] will be paraphrased. Note that the "δ" used in (1) is the same as "2δ" for Rabin so our results are superficially different. Using an argument involving volumes of sets in n-dimensional space, a bound on k will be deduced. Let d = $F_{max} - F_{min}$

Define the set σ_i $i = 1, \ldots, k$ by

$$\sigma_i = \{(e_1, \ldots, e_n) \mid e_j^i \le e_j, \; j = 1, \ldots, n; \; \sum_{j=1} (e_j - e_j^i) = \frac{\delta}{d}\}$$

Each σ_i is the set σ translated by e^i : $i = 1, \ldots, k$ where:

$$\sigma = \{(e_1', \ldots, e_n') \mid 0 \le e_j', \; j = 1, \ldots, n, \sum_j e_j' = \frac{\delta}{d}\}$$

The set σ is an n-1 dimensional simplex contained in the hyperplane $x_1 + \ldots + x_n = \frac{\delta}{d}$. The n-1 dimensional volume, $V_{n-1}(\sigma)$ of σ, is

$$V_{n-1}(\sigma) = c(\frac{\delta}{d})^{n-1}$$ where c is a constant with respect to δ.

Since the e^i : $i = 1, \ldots, k$ are all stochastic, we get

$$(e_1, \ldots, e_n) \varepsilon \sigma_i \Longrightarrow \sum_{i=1}^{n} e_i = 1 + \frac{\delta}{d}, \qquad 0 \le e_j \qquad j = 1, \ldots, n$$

Hence each $\sigma_i \subseteq \tau$ where

$$\tau = \{(e_1'', \ldots, e_n'') \mid \sum_{i=1}^{n} e_i'' = 1 + \frac{\delta}{d}, \qquad 0 \le e_i'' \qquad i = 1, \ldots, n\}$$

We show that the σ_i have no interior points in common.

A point $e = (e_1, \ldots, e_n) \varepsilon \sigma_i$ is an interior point of σ_i (in the topology of the hyperplane $x_1 + \ldots + x_n = 1 + \frac{\delta}{d}$) if and only if

$$0 < e_p - e_p^i \qquad p = 1, \ldots, n$$

Suppose σ_i and σ_j have a point e in common:

$$0 < e_p - e_p^i, \qquad 0 < e_p - e_p^j \qquad p = 1, \ldots, n \qquad i \neq j$$

$$e_p^i - e_p^j = (e_p - e_p^j) - (e_p - e_p^i) \qquad p = 1, \ldots, n$$

$$|e_p^i - e_p^j| < |e_p - e_p^j| + |e_p - e_p^i| \qquad p = 1, \ldots, n$$

$$\sum_{p=1}^{n} |e_p^i - e_p^j| < \sum_{p=1}^{n} |e_p - e_p^j| + \sum_{p=1}^{n} |e_p - e_p^i| = \frac{2\delta}{d}$$

which contradicts equation (4). Hence σ_i and σ_j $(i \neq j)$ have no interior points in common.

Consequently

$$V_{n-1}(\sigma) + \dots + V_{n-1}(\sigma_k) \leq V_{n-1}(\tau)$$

Hence

$$ck(\frac{\delta}{d})^{n-1} \leq c\ (1 + \frac{\delta}{d})^{n-1}$$

So

$$k \leq (1 + \frac{d}{\delta})^{n-1}$$

Q. E. D.

We are now ready to prove the adaptation of the Rabin reduction theorem to probabilistic sequential machines.

Theorem 5.2 (Reduction theorem for probabilistic sequential machines)

Let A be an n state probabilistic machine having isolated cutpoint λ with separation δ. Then there exists a deterministic machine D with n_D states such that

$$T(A,\lambda) = T(D)$$

and

$$n_D \leq (1 + \frac{d}{2\delta})^{n-1} \text{ where } d = (\max_i(F_i) - \min_i(F_i))$$

Proof: We define R_T, the congruence relation induced by $T(A,\lambda)$ by:

$$x\ R_T\ y \text{ if } xz \in T(A,\lambda) \Longleftrightarrow yz \in T(A,\lambda) \qquad \forall z \in \Sigma^*$$

Let $x_1, \dots, x_r, \dots$ be a set of distinct representatives of the classes of R_T.

Since $x_i \not R_T\, x_j$ $\quad i \neq j$ (where $\not R_T$ is the negation of the relation R_T) there exists $z \in \Sigma^*$ (W.L.G.) such that

$$IA(x_i z)F \geq \lambda; \quad IA(x_j z)F < \lambda$$

Because λ is an isolated cutpoint of A

$$IA(x_i z)F \geq \lambda + \delta; \quad IA(x_j z)F \leq \lambda - \delta$$

Since A is a machine

$$IA(x_i)A(z)F \geq \lambda + \delta; \quad IA(x_j)A(z)F \leq \lambda - \delta$$

hence for each pair of classes i and j there is a z so that

$$|(IA(x_i) - IA(x_j))A(z)F| \geq 2\delta$$

By Theorem 5.1, there are at most k distinct distributions $IA(x_i)$ where $k \leq (1 + \frac{d}{2\delta})^{n-1}$. If two representatives of classes of R_T have the same distribution it follows immediately that the representatives are in the same R_T class. Consequently there are at most $(1 + \frac{d}{2\delta})^{n-1}$ classes of R_T. Since R_T has finite rank and is almost by definition a right congruence relation, Theorem 2 of Rabin & Scott [1959] applies and there exists a finite deterministic machine D such that

$$T(D) = T(A, \lambda).$$

Q. E. D.

5.2 A FINITE SET OF INVARIANTS FOR $\equiv_T$ FOR PROBABILISTIC SEQUENTIAL MACHINES HAVING ISOLATED CUTPOINTS.

It is now possible to write down a bound on the length of strings required for determining whether $\equiv_T$ holds between two probabilistic sequential machines having isolated cutpoints.

Theorem 5.3

Let A and A' be probabilistic sequential machines having n and n' states, isolated cutpoints λ and λ', and separations δ and δ', respectively, then $A \equiv_T A'$, i.e.,

$$T(A,\lambda) = T(A',\lambda) \text{ if and only if } x \in T(A,\lambda) \Longleftrightarrow x \in T(A',\lambda)$$

$$\forall x \in \Sigma^* : \ell g(x) \leq p$$

where

$$p \leq (1 + \frac{d}{2\delta})^{n-1} \circ (1 + \frac{d'}{2\delta'})^{n'-1} \text{ with } d = |\max_i F_i - \min_i F_i|$$

$$\text{and } d' = |\max_i F'_i - \min_i F'_i|$$

Proof: By Theorem 5.2 there are deterministic machines D and D' such that

$$T(A,\lambda) = T(D)$$

$$T(A',\lambda') = T(D')$$

and

$$n_D \leq (1 + \frac{d}{2\delta})^{n-1}$$

$$n_{D'} \leq (1 + \frac{d'}{2\delta'})^{n'-1}$$

By Theorem 10 of Rabin and Scott [1959], two deterministic automata are inequivalent by $\equiv_T$ if and only if there is a tape x of length less than the product of the number of internal states of the two automata accepted by one machine but not by the other.

Hence

$$T(D) = T(D') \text{ iff for all } x \text{ in } \Sigma^* \text{ such that } \ell g(x) \leq n_D \cdot n_{D'}$$

$$x \in T(D) \Longleftrightarrow x \in T(D')$$

Substitution of T(A,λ) and T(A',λ') into the above gives the theorem.

Q. E. D.

PART II. A CLASS OF MACHINES CONTAINING ISOLATED CUTPOINT MACHINES

5.3 THE STABILITY CONGRUENCE RELATION

<u>Definition 5.2</u>: An input string x is <u>stable</u> for a pair of machines A and A' if $x \in T(A,\lambda) \Longleftrightarrow x \in T(A',\lambda)$.

<u>Definition 5.3</u>: The <u>stable set</u> for A and A', written S(A,A'):

$$S(A,A') = \{x : x \text{ is a stable tape for A and A'}\}$$

Note that $S(A,A') = T(A,\lambda) \cap T(A',\lambda) \cup [\Sigma^* - T(A,\lambda)] \cap [\Sigma^* - T(A',\lambda)]$.

<u>Lemma 5.1</u>
$$S(A,A') = \Sigma^* \Longleftrightarrow A \equiv_T A'$$

<u>Proof</u>: Suppose $S(A,A') = \Sigma^*$. Then for all $x \in \Sigma^*$:

$x \in T(A,\lambda) \Longleftrightarrow x \in T(A',\lambda)$ which is just the definition of $A \equiv_T A'$.

If $A \equiv_T A'$, for all $x \in \Sigma^*$: $x \in T(A,\lambda) \Longleftrightarrow x \in T(A',\lambda)$ which implies $x \in S(A,A')$.

<u>Definition 5.4</u>: The <u>stability right congruence relation</u> $R_s(A,A')$:

$$x \; R_s(A,A') \; y \text{ iff } \forall z \in \Sigma^*: xz \in S(A,A') \Longleftrightarrow yz \in S(A,A')$$

We remark that $R_s(A,A')$ is the right congruence relation induced by $S(A,A')$. If it is clear in context, $R_s(A,A')$ will be abbreviated to R_s. R_s has only one class if $A \equiv_T A'$, since by Lemma 5.1 the stable set consists of all possible finite input sequences and is hence closed under extensions. However, if $A \equiv_T A'$ does <u>not</u> hold, R_s may have several congruence classes which represent, intuitively, distinct ways that the machines A and A' can be unstable.

5.4 A SUFFICIENT CONDITION FOR A FINITE COMPLETE SET OF INVARIANTS FOR $\equiv_T$

<u>Theorem 5.4</u> (Sufficient condition for a finite experiment for deciding whether $\equiv_T$ holds.)

If the rank of $R_s(A,A') = r$, finite, then

$$A \equiv_T A' \Longleftrightarrow [\forall x \in \Sigma^*: \ell g.(x) \leq r;\; x \in S(A,A'')]$$

<u>Proof</u>: We first show that R_s refines the set $\Sigma^* - S(A,A')$. Since $\Lambda \in \Sigma^*$, we set $z = \Lambda$ which gives

$$x \; R_s \; y \Longrightarrow [x \; \varepsilon \; S(A,A') \Longleftrightarrow y \; \varepsilon \; S(A,A')]$$

Logical equivalence gives

$$x \; R_s \; y \Longrightarrow [x \not\varepsilon \; S(A,A') \Longleftrightarrow y \not\varepsilon \; S(A,A')]$$

which means that R_s refines $\Sigma^* - S(A,A')$, i.e.,

$$x \; R_s \; y \Longrightarrow [x \; \varepsilon \; \Sigma^* - S(A,A') \Longleftrightarrow y \; \varepsilon \; \Sigma^* - S(A,A')]$$

Using Theorem 2 of Rabin and Scott [1959], $\Sigma^*/R_s = D$ is a finite deterministic machine such that T(D), the set of tapes accepted by D, is just the set $\Sigma^* - S(A,A')$ i.e. $D = \langle R_s[\Lambda], \{R_s[x]\}, M, \mathcal{F} \rangle$

Where

$R_s[\Lambda]$ is the initial state of D

$R_s[x]$ is the state set of D. $\#\{R_s[x]\} = \text{Rank } R_s = r$

M is the transition function of D: $M(R_s[x],\sigma) = R_s[x\sigma]$

$\mathcal{F}$ is the set of final states of D: $\mathcal{F} = \{R_s[x] \; : \; x \; \varepsilon \; \Sigma^* - S(A,A')\}$

Note that

$$x \not\varepsilon \; T(D) \Longleftrightarrow x \; \varepsilon \; S(A,A') \qquad (*)$$

Theorem 7 of Rabin and Scott [1959] tells us that T(D) is empty iff D rejects all strings whose lengths are at most the number of states of D.

Symbolically we have:

$$T(D) = \Phi \Longleftrightarrow [\forall x \; \varepsilon \; \Sigma^*: \; \ell g.(x) \le r; \; x \not\varepsilon \; T(D)]$$

Using (*) we obtain

$$T(D) = \Phi \Longleftrightarrow [\forall x \; \varepsilon \; \Sigma^* \; : \; \ell g.(x) \le r; \; x \; \varepsilon \; S(A,A')] \qquad (**)$$

But $T(D) = \Sigma^* - S(A,A') = \Phi \Longleftrightarrow \Sigma^* = S(A,A')$.

Hence by (**)

$$\Sigma^* = S(A,A') \Longleftrightarrow [\forall x \; \varepsilon \; \Sigma^*: \; \ell g.(x) \le r; \; x \; \varepsilon \; S(A,A')]$$

From Lemma 5.1 and the above equivalence

$$A \equiv_T A' \Longleftrightarrow [\forall x \; \varepsilon \; \Sigma^*: \; \ell g.(x) \le r; \; x \; \varepsilon \; S(A,A')]. \qquad \text{Q. E. D.}$$

The converse to Theorem 5.4 does not hold. Even though the rank of R_s is infinite, there might be a finite string which shows that $A \equiv_T A'$ does not hold. Let us now seek properties of pairs of machines which cause the rank of $R_s(A,A')$ to be finite.

5.5 EXPERIMENTS WITH MACHINES

Let us formalize what is usually meant by experiments with a machine to make later definitions more precise.

Definition 5.5: An *experiment* on the r machines $A^{(1)}, \ldots, A^{(r)}$ using the input set $T = \{x_1, \ldots, x_k, \ldots\}$ written $T(A^{(1)}, \ldots, A^{(r)}, \lambda)$ is a table:

$$T \left\{ \begin{array}{c|ccccc} & x_1 & x_2 & \ldots & x_k & \ldots \\ \hline A^{(1)} & b^1_1 & b^1_2 & \ldots & b^1_k & \ldots \\ A^{(2)} & b^2_1 & b^2_2 & \ldots & b^2_k & \ldots \\ \vdots & \vdots & & & & \\ A^{(r)} & b^r_1 & b^r_2 & \ldots & b^r_k & \ldots \end{array} \right.$$

$$\begin{array}{lll} \text{where } b^s_j = 1 \text{ if } I^{(s)}A^{(s)}(x_j)F^{(s)} \geq \lambda & & x_j \in T \\ \phantom{\text{where } b^s_j} = 0 \text{ if } I^{(s)}A^{(s)}(x_j)F^{(s)} < \lambda & & s = 1, \ldots, r \end{array}$$

An experiment is finite if T is finite.

Definition 5.6: A *simultaneous experiment* on the machines $A^{(1)}, \ldots, A^{(r)}$ using the input sets of tapes $T_1, \ldots, T_m$, written

$$T_1 \times T_2 \times \ldots \times T_m(A^{(1)}, \ldots, A^{(r)}, \lambda)$$

is a table of m experiments

$$\begin{array}{c} T_1(A^{(1)}, \ldots, A^{(r)}, \lambda) \\ \vdots \\ T_m(A^{(1)}, \ldots, A^{(r)}, \lambda) \end{array}$$

In what follows we will not make use of the concept of simultaneous experiment, but have included it here for the sake of generality. In particular, we will consider an experiment with two tapes xz and yz on the two machines A and A' which will be written according to the notation of Definition 5.5 as $\{xz, yz\}(A, A', \lambda)$.

5.6 THE MINORITY PROPERTY

Definition 5.7: *A machine A is in the minority* for an experiment $\{xz,yz\}(A,A',\lambda)$ if the experiment yields

$$\begin{array}{ccc} & xz & yz \\ A: & 0 & 1 \\ A': & 0 & 0 \end{array} \qquad \text{or} \qquad \begin{array}{ccc} & xz & yz \\ A: & 1 & 0 \\ A': & 0 & 0 \end{array}$$

or else a complement event of the above occurs.

$$\begin{array}{ccc} & xz & yz \\ A: & 1 & 0 \\ A': & 1 & 1 \end{array} \qquad \text{or} \qquad \begin{array}{ccc} & xz & yz \\ A: & 0 & 1 \\ A': & 1 & 1 \end{array}$$

5.7 THE BALANCE PROPERTY

A property of two machines will be defined which leads to a finite test for $\equiv_T$.

Definition 5.8: Machines A and A' have the *balance property* if for all $x \in \Sigma^*$ and those $y \in \Sigma^*$ such that $\sim x \; R_s(A,A') \; y$ either:

(i) There are strings z and z' such that A is in the minority for the experiment $\{xz,yz\}(A,A',\lambda)$ and A' is in the minority for the experiment $\{xz',yz'\}(A,A',\lambda)$.

or else

(ii) There is an experiment $\{xz_2,yz_2\}(A,A',\lambda)$ which yields

$$\begin{array}{ccc} & xz_2 & yz_2 \\ A: & 0 & 1 \\ A': & 1 & 0 \end{array}$$

or the complement

$$\begin{array}{ccc} & xz_2 & yz_2 \\ A: & 1 & 0 \\ A': & 0 & 1 \end{array}$$

Note that condition (i) specifically rules out the case when only one machine ever is in the minority; i.e., there are at least two distinct z's such that $xz \in S(A,A') \not\Leftrightarrow yz \in S(A,A')$. Condition (ii) allows just one z_k to cause A and A' to accept one continuation and reject the other for both x and y.

<u>Example 5.1</u>: Machines A and A' which have the balance property.

Suppose we consider a mod 2 counter and a mod 3 counter connected together by means of a 3 state buffer. Two of the states of the buffer serve as a channel for carrying information between the counters. However, one of the states in the buffer is a sink. Each time the buffer is entered or left there is probability ϵ that the machine will enter the sink state which has output equal to the cutpoint. Consequently we view the entry to the sink as a breakdown of the machine which causes it to accept all subsequent tapes.

In general the cutpoints for the machines will not be isolated. Figure 5.1 shows the machines A and A' which are nearly isomorphic as far as state transitions are concerned. The 0 symbol of the counters and most other loops have been deleted from the diagram.

A and A' are each started in an initial state rather than a distribution over the states, so that input strings cause distributions with only two positive components. One component will be the probability of being in the sink or absorbing state and the other (initially much larger) will be the probability of being in some other state. Tape acceptance depends only on this later state since we will assume $F_{(sink)} = 0$ and the cutpoint $\lambda = 0$. Hence in this case it is sufficient to consider all pairs of states rather than all pairs of accessible distributions in an experiment to find out whether the balance property holds. With a minor change in the notation for an experiment to point out this simplification, we summarize a sequence of experiments in Table 5.1

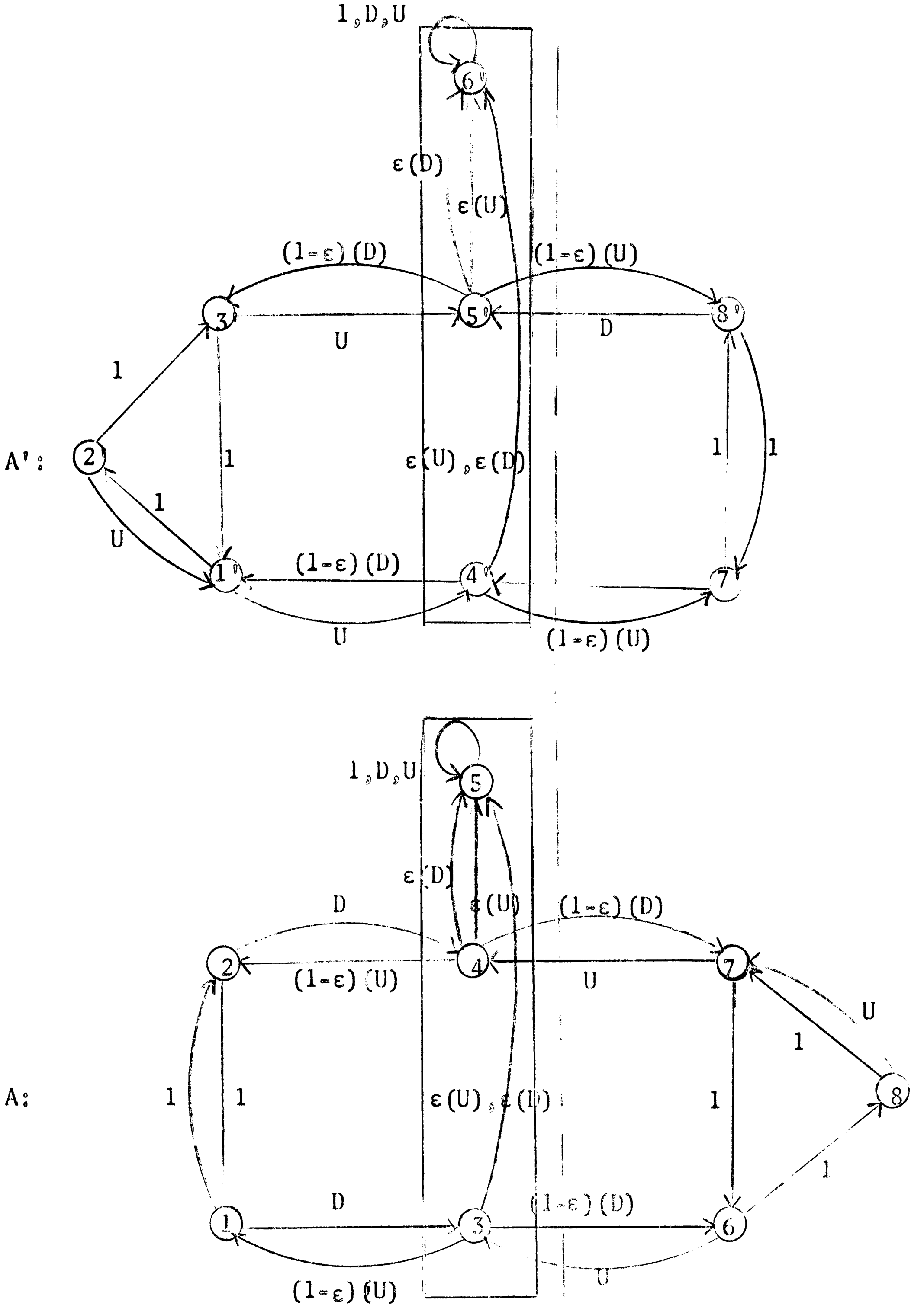


Figure 5.1: Transition diagrams for machines A and A' which have the balance property. States left of the center line have outputs greater than or equal to the cutpoint; states on the right have outputs less than the cutpoint. Most loops are omitted from the above graphs to simplify them; hence if the graph leaves the transition of a state undefined for a particular symbol the reader may assume a loop has been deleted. The symbols are read as: U—up, D—down, 1—one. Note that "up" has a different orientation for the two machines.

Table 5.1 A sequence of experiments which shows that the balance property holds for A and A'. The strings x and y of the definition of the balance property are replaced by states (which are the responses to the inputs x and y other than the sink state).

Accepted Strings:

	Machine states		Z_i	Z_j
E_1:	A		D D 1 1 U	U U
		(1)	1	1
		(2)	0	1
	A'			
		(1')	1	0
		(2')	1	1
E_2:	A		1 D D 1 1 U	1 U U
		(1)	1	1
		(2)	0	1
	A'	(1')	1	1
		(3')	1	0
E_3:	A		D D 1 1 U	1 1 U U
		(1)	1	1
		(2)	0	1
	A'	(2')	1	0
		(3')	1	1
E_4:	A	(3)	using substrings of those for (1) and (2)	using same strings as for (1) and (2) or with the initial symbol removed
		(4)		
	A'	Any accepted states		

Rejected strings:

	Machine states		Z_i	Z_j
E_5:	A		1 U U	D D 1 U U
		(6)	0	1
		(7)	1	1
	A'	(7')	0	1
		(8')	0	0
E_6:	A		U	D D 1 1 U U
		(6)	1	1
		(8)	0	1
	A'	(7')	0	0
		(8')	0	1

5.8 THE CONCEPT OF STABILITY NEAR THE CUTPOINT

As we have seen above, it is a straightforward task to find a finite bound on the length of experiments needed to determine whether two machines are equivalent by $\equiv_T$ when the machines both have isolated cutpoints. There is an apparently large class of machines which may have non-isolated cutpoints for which a bound can be found. Although the cutpoints are not necessarily isolated for this class of machines, special properties will be required of them.

Definition 5.9: Machines A and A' are stable when closer than δ to the cutpoint λ, if for any x in Σ^*:

Either $E_A(x)$ or $E_{A'}(x)$ in the open interval $(\lambda-\delta,\lambda+\delta) \Longrightarrow x \in T(A,\lambda) \Longleftrightarrow x \in T(A',\lambda)$.

When Definition 5.9 holds, we sometimes will just say that "A and A' are stable near the cutpoint" with the δ understood in context.

Clearly isolated cutpoint machines are special cases of those machines stable close to the cutpoint. But what properties of isolated cutpoint machines would we expect machines stable close to the cutpoint to have? In general it is not true that machines stable near the cutpoint are tape equivalent to deterministic machines, nor even that there is a finite set of invariants for $\equiv_T$. However, for the class of machines stable near the cutpoint and with the balance property, there is a bound on experimentation for $\equiv_T$ analogous to the isolated cutpoint case.

Example 5.2: A class of machines stable near the cutpoint but with non-isolated cutpoints.

We shall assume that the state descriptions of machines M and M' can be decomposed into two parts B_1 and B_2 and B_1' and B_2' respectively.

$$M(\sigma) = \begin{array}{cc} & \begin{array}{ccc} & S_2 & \end{array} \\ \begin{array}{c} \\ S_0 \\ \\ \end{array} & \left(\begin{array}{c|c|c} B_1(\sigma) & & 0 \\ \hline 0 \;\ldots\; 0 & 1 & 0 \;\ldots\; 0 \\ \hline 0 & & B_2(\sigma)\end{array}\right)\end{array} \qquad \sigma \neq R \qquad \sigma \in \Sigma$$

while

$$M(R) = \begin{array}{cc} & \begin{array}{cc} \begin{array}{c} S_{in} \\ \downarrow \end{array} & S_2 \end{array} \\ \begin{array}{c} \\ S_0 \\ \\ \end{array} & \left(\begin{array}{ccc|c} \multicolumn{3}{c|}{B_1(R)} & 0 \\ \hline 0 & \ldots & 0 & 1.\;\ldots\; 0 \\ \hline 0 & \begin{array}{|c|} 1 \\ \vdots \\ 1 \end{array} & \multicolumn{2}{c}{0} \end{array}\right)\end{array}$$

The analogous situation holds for M'. Figure 5.2 shows the state decomposition.

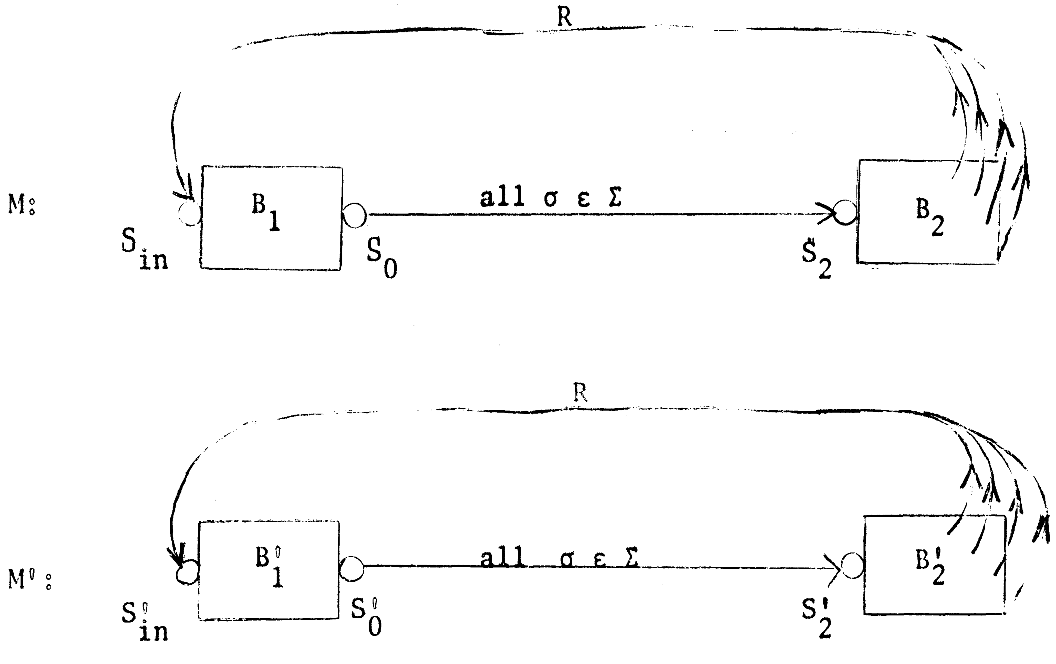


Figure 5.2: Machines of Example 5.2.

Special properties will be imposed on the submachines so that M and M' will be stable when near the cutpoint. Any state s of a probabilistic sequential machine has an output weight F_s. However, this fact will be ignored temporarily in order to define a particular property of M and M'.

(i) Let S_{in} and S'_{in} be the initial states of the machines under consideration.

(ii) Let B_1 and B'_1 be deterministic.

(iii) The only way state transitions can occur from B_1 to B_2 is via state S_0 to S_2. Likewise, the only way state transitions can occur from B'_1 to B'_2 is from S'_0 to S'_2. We will require B_2 and B'_2 to be started simultaneously in what follows. To obtain simultaneous arrivals at S_2 and S'_2, we ignore the outputs and regard B_1 and B'_1 as Rabin-Scott automata with final states S_0 and S'_0 respectively: Then we need only

$$T(B_1) = T(B'_1)$$

Note that we are <u>not</u> requiring that

$$T(B_1,\lambda) = T(B'_1,\lambda)$$

because we have not used the normal outputs.

(iv) The submachines B_2 and B'_2 satisfy

$$E_{B_2}(x) = E_{B'_2}(x) \qquad \text{for all } x \text{ not containing an } R$$

That is,

$$B_2 \equiv_E B'_2 \text{ for the alphabet } \Sigma - \{R\}.$$

Furthermore, B_2 and B'_2 are selected so that cutpoint λ is not isolated.

(v) Lastly, the entries of F_M and $F_{M'}$ have the following property: There is a $\delta > 0$ so that

$F_i > \lambda + \delta$ or $F_i < \lambda - \delta$ where S_i is in B_1.

$F'_i > \lambda + \delta$ or $F'_i < \lambda - \delta$ where S'_i is in B'_1.

$\lambda - \delta \le F_i \le \lambda - \delta$ all S_i in B_2:

$$\lambda - \delta \leq F'_i < \lambda - \delta \quad \text{all } S'_i \text{ in } B'_2$$

Whenever the state activity of M is in B_1, the state activity of M' is in B'_1. The expectations for the submachines B_1 and B'_1 are just the weights attached to the states since these submachines are deterministic; hence

$$E_{M_1}(x) \not\in (\lambda-\delta,\lambda+\delta)$$

$$E_{M'_1}(x) \not\in (\lambda-\delta,\lambda+\delta)$$

The only other possible mode of operation is for the state activity of M to be in B_2 and the state activity of M' to be in B'_2 for which we have:

$$(\lambda-\delta) \leq E_{M_1}(x) = E_{M'_1}(x) \leq (\lambda+\delta)$$

Hence M and M' are stable when close to the non-isolated cutpoint λ. More elaborate examples can be constructed by weakening the above restrictions. In particular (ii) and (iii) can be modified easily.

5.9 A FINITE TEST FOR $\equiv_T$ FOR MACHINES STABLE NEAR THE CUTPOINT AND WHICH HAVE THE BALANCE PROPERTY

Theorem 5.5

Let A and A' be probabilistic sequential machines having n_A and $n_{A'}$ states respectively. If A and A' are stable near the cutpoint λ (when closer than δ) and have the balance property then

$$A \equiv_T A' \text{ if } \forall x \in (\Sigma)^p \qquad [x \in T(A,\lambda) \Longleftrightarrow x \in T(A',\lambda)]$$

where

$$p \leq (1 + \frac{d}{\delta})^{n_A-1} \cdot (1 + \frac{d'}{\delta})^{n_{A'}-1} \quad \text{and } d = (\max_i(F_i) - \min_i(F_i))$$

$$d' = (\max_i(F'_i) - \min_i(F'_i))$$

Proof: Let us pick a set of representatives $x_1, \ldots, x_r, \ldots$ from the congruence classes of the stability congruence relation R_s. Then $\sim x_i R_s x_j$ for $i \neq j$ $i = 1, 2, \ldots$

Definition 5.10: Given a representative x_i of a congruence class of $R_s(A,A')$, the pair of distributions $(IA(x_i), I'A'(x_i))$ will be called a canonical distribution pair.

Of course, there is at least one canonical distribution pair for each representative of a class of R_s. Subject to the hypothesis of the theorem, it will be shown that there are only a finite number of canonical pairs.

The definition of the balance property (Definition 5.8) leads us to consider two cases.

Case (i): Suppose part (i) of Definition 5.8 is satisfied. Then for any representatives of the classes of R_s, x_i and x_j, there are strings z and z' such that A is in the minority for the experiment $\{x_iz, x_jz\}(A,A',\lambda)$ and A' is in the minority for the experiment $\{x_iz', x_jz'\}(A,A',\lambda)$.

Without loss of generality, assume that $x_iz \notin S(A,A')$ in the experiment in which A is in the minority. The expectations $E_A(x_iz)$ and $E_{A'}(x_iz)$ must lie on opposite sides of the interval $(\lambda-\delta, \lambda+\delta)$ since A and A' are stable near the cutpoint. Furthermore, $E_A(x_jz)$ must lie on the opposite side of λ from $E_A(x_iz)$ because A is in the minority. The value of $E_A(x_jz)$ may approach λ but the value of $E_A(x_iz)$ must be no closer than δ to λ. Hence

$$|E_A(x_iz) - E_A(x_jz)| \geq \delta$$

which gives by definition

$$|I \circ A(x_iz) \circ F - I \circ A(x_jz) \circ F| \geq \delta$$

Using the property of machines

$$|I \circ A(x_i) \circ A(z) \circ F - I \circ A(x_j) \circ A(z) \circ F| \geq \delta \qquad (5)$$

a similar argument gives:

$$|I'A'(x_i)A'(z')F' - I'A'(x_j)A'(z')F'| \geq \delta \qquad (6)$$

Case (ii): Without loss of generality, assume that $IA(x_iz_2)F \geq \lambda$. Then we must have by definition $IA(x_jz_2)F < \lambda$, $I'A'(x_iz_2)F' < \lambda$ and $I'A'(x_jz_2)F' \geq \lambda$. Since A and A' are stable near the cutpoint, all expectations above are outside the interval $(\lambda-\delta,\lambda+\delta)$.

Hence

$$|IA(x_iz_2)F - IA(x_jz_2)F| \geq 2\delta > \delta \qquad (5)'$$

$$|I'A'(x_iz_2)F' - I'A'(x_jz_2)F'| \geq 2\delta > \delta \qquad (6)'$$

Consequently, either (5) and (6) hold or (5)' and (6)' hold.

Letting $e^i = IA(x_i)$

$e^j = IA(x_j)$ and $z_{ij} = z$ or z_2 we observe that Theorem 5.1 applies and the maximal number of distinct distributions $IA(x_i)$ of A is

$$d_A \leq (1 + \frac{d}{\delta})^{n_A-1} = k_A$$

and letting

$$e^i = I'A'(x_i)$$
$$e^j = I'A'(x_j)$$

and $z_{ij} = z'$ or z_2

Using Theorem 5.1 we get

$$d_{A'} \leq (1 + \frac{d'}{\delta})^{n_A-1} = k_{A'}$$

Consequently the number of distinct canonical distribution pairs is finite and equal to $d_A \cdot d_{A'}$. We note that $d_Ad_{A'} \leq k_Ak_{A'}$.

To complete the proof it will be shown that x_i and x_j distinct representatives implies that

$$(IA(x_i), I'A'(x_i))$$

and

$$(IA(x_j), I'A'(x_j))$$

are distinct pairs of distributions.

Assume to the contrary that $IA(x_i) = IA(x_j)$ and $I'A'(x_i) = I'A'(x_j)$. Multiplying the former equality by $A(z)F$ and the latter by $A'(z)F'$ for an arbitrary z gives:

$$I \circ A(x_i) \circ A(z)F = IA(x_j) \circ A(z)F$$

$$I'A'(x_i) \circ A'(z)F' = I'A'(x_j) \circ A'(z)F'$$

Since A and A' are machines we obtain

$$IA(x_i z)F = IA(x_j z)F$$

$$I'A'(x_i z)F' = I'A'(x_j z)F'$$

Suppose

$$IA(x_i z)F \geqq \lambda \Longleftrightarrow I'A'(x_i z)F' \geqq \lambda$$

Substituting equals for equals we get

$$IA(x_j z)F \geqq \lambda \Longleftrightarrow I'A'(x_j z)F' \geqq \lambda$$

Which means we have shown that

$$[IA(x_i z)F \geqq \lambda \Longleftrightarrow I'A'(x_i z)F' \geqq \lambda] \Longrightarrow [IA(x_j z)F \geqq \lambda \Longleftrightarrow I'A'(x_j z)F' \geqq \lambda]$$

Using Definition 5.3 we get

$$x_i z \; \varepsilon \; S(A,A') \Longrightarrow x_j z \; \varepsilon \; S(A,A') \qquad (*)$$

Reversing the roles of x_i and x_j in the previous argument gives the converse of (*) so we have for an arbitrary z

$$x_i z \; \varepsilon \; S(A,A') \Longleftrightarrow x_j z \; \varepsilon \; S(A,A')$$

Hence by Definition 5.4

$$x_i \; R_s(A,A') x_j$$

which is a contradiction.

Consequently there are $d_A \circ d_{A'}$ classes of $R_s(A,A')$. Theorem 5.4 applies:

$$A \equiv_T A' \Longleftrightarrow [\forall x \; \varepsilon \; (\Sigma)^{d_A \circ d_{A'}} : x \; \varepsilon \; S(A,A')]$$

But $d_A \circ d_{A'} \leqq k_A \circ k_{A'}$ and transitivity of implication yields the theorem.

Q. E. D.

5.10 CONCLUSIONS

If the properties of being stable near the cutpoint and having the balance property imply that a machine has an isolated cutpoint, then the bound of Theorem 5.5 is not as good as the bound of Theorem 5.3. If this were the case, instead of a finite set of invariants for $\equiv_T$ for certain probabilistic machines, we would have obtained a very peculiar behavioral characterization of finite deterministic machines. However the great richness of the class of probabilistic machines leads the author to believe that there probably are many such probabilistic machines whose cutpoints are non-isolated. The lack of an adequate characterization of isolated cutpoint machines adds to the difficulty of proving whether the class of machines of Part I is property contained in the class of Part II.

CHAPTER 6

STABILITY PROBLEMS FOR $\equiv_I$, $\equiv_N$ AND $\equiv_E$

6.1 ENUMERATION OF THE BEHAVIORAL EQUIVALENCE CLASS OF A MACHINE

In general a machine may be behaviorally equivalent to a machine with fewer states, with the same number of states, or with more states. Some machines behaviorally equivalent to machines with fewer states will be discussed in Chapter 8. The systematic expansion of a machine into behaviorally equivalent machines with more states does not appear in the literature of automata theory. Although the expansions of a machine are relevant to certain problems such as the decomposition of machines, they will not be discussed here. This chapter and Chapter 7 will deal with the remaining case; machines behaviorally equivalent to a given machine and having the same number of states as the given machine.

In order to avoid large differences between machines masked by the initial distributions, only those equivalences defined for all initial states will be considered. To justify this strong restriction consider digital computers to be probabilistic machines. An automaton state model may compress all the intricacies of a stored program into a single initial state, e.g., different stored programs correspond to different initial states. Using such a model and defining behavioral equivalence for two machines from particular initial states makes heterogeneous kinds of computers behaviorally equivalent. For example, a computer could be in a behavior class with all other computers for which there are stored programs to do some particular computation. But behavioral equivalence from any initial states requires that the behavior class contains only those machines which have sets of stored programs making them interchangeable.

6.2 THE GENERAL STABILITY PROBLEM

Our goal in this section is to study changes in the switching of a machine which do not change the behavior class of the machine. Therefore the output vector F will be kept fixed. Usually we will think of A as a satisfactory machine and A' as the machine which A becomes after some permanent perturbation in switching. Of interest are those A' which are stable changes, i.e., behaviorally equivalent to A.

Definition 6.1: A' is a $\equiv$-stable perturbation of A if

(i) A and A' have the same number of states.

(ii) F = F', i.e., A and A' have identical output vectors.

(iii) A $\equiv$ A' for all I ε S (where $\equiv$ is a machine equivalence).

Rabin [1964] formulated a stability problem for $\equiv_T$. However, the Rabin stability problem requires equivalence from fixed initial states rather than all states. Very few results concerning this problem have appeared in the literature since it was published more than two years ago. Progress has been made by Paz [1964] for certain special cases. Some of the results of Chapter 7 apply to the Rabin problem.

Some of the general properties of the transformations T which map the machine A into a stable perturbation A' will be studied. In particular, those transformations T which map the monoid $\langle A(\Sigma^*), \cdot \rangle$ into the monoid $\langle A'(\Sigma^*), \cdot \rangle$ such that A $\equiv$ A' for some machine equivalence $\equiv$ will be called stability transformations.

When a stability transformation T is applied to symbol matrices, we obtain by definition:

$$T(A(i)) = A'(i) \qquad i \varepsilon \Sigma$$

Let $x = i_1 \ldots i_p$ $\qquad i_j \varepsilon \Sigma: j = 1, 2, \ldots, p$

But

$$A'(x) = A'(i_1)A'(i_2) \dots A'(i_p)$$

since A' is a machine. Hence T must be a monoid homorphism preserving the monoid operation matrix multiplication.

$$T(A(i_1 \dots i_p)) = T(A(i_1) \cdot A(i_2) \cdot \dots \cdot A(i_p))$$

$$= T(A(i_1)) \cdot T(A(i_2)) \cdot \dots \cdot T(A(i_n))$$

Theorem 6.1

The most general form of stability transformation is

$$T(A(x)) = A(x) + E_x \qquad x \in \Sigma^*$$

where E_x is an n x n "error matrix" with $\sum_{k=1}^{n} (E_x)_{ik} = 0$ and

$1 \geq A(x)_{ik} + (E_x)_{ik} \geq 0$

Proof: Given a machine A and any stable perturbation A' = T(A), start each machine in the states $S_1 = (1, 0, \dots, 0), \dots, S_n = (0, \dots, 0, 1)$

Let $S_1 \cdot A(x) - S_1(TA(x)) = -(e^x_{11}, \dots, e^x_{1n})$

$$\vdots \qquad \vdots \qquad \vdots \qquad \vdots \qquad \vdots$$

$$S_n \cdot A(x) - S_n(TA(x)) = -(e^x_{n1}, \dots, e^x_{nn})$$

Any initial state I is a linear combination of the above basis state vectors. Let us call $E_x = ||-e^x_{ij}||$ which is uniquely determined. The error pattern for distributions over the states from any initial distribution I can be obtained by $I(-E_x) = IA(x) - IT(A(x))$.

Both T(A(x)) and A(x) are stochastic, so summing over the rows

$$\sum_{j=1}^{n} [T(A(x))]_{ij} = \sum_{j=1}^{n} [A(x)]_{ij} + \sum_{j=1}^{n} [E_x]_{ij}$$

$$1 = 1 + \sum_{j=1}^{n} e^x_{ij}$$

$$0 = \sum_{j=1}^{n} e_{ij}^{x}$$

Since $1 \geq [A'(x)]_{ij} = A(x)_{ij} + e_{ij}^{x} \geq 0$ the other part is established,

Q. E. D.

Computation of the error matrix for a string x from the error matrices of the symbols of x is given by the next result.

Theorem 6.2

Let $T : A \longrightarrow A'$ be a stability transformation and let

$$x = i_1 \ldots i_k \qquad i_j \in \Sigma; \qquad j = 1, \ldots, k$$

$$E_x = \prod_{j=1}^{k} (A(i_j) + E_{i_j}) - A(x)$$

Proof:

$$\begin{aligned} E_x &= A'(x) - A(x) \\ &= A'(i_1 \ldots i_k) - A(x) \\ &= A'(i_1) \ldots A'(i_k) - A(x) \\ &= \prod_{j=1}^{k} (A(i_j) + E_{i_j}) - A(x) \end{aligned}$$

Corollary 6.2

Let $T : A \longrightarrow A'$ be a stability transformation. Then the error matrix for $x\sigma$ is specified by the error matrices for x and σ by

$$E_{x\sigma} = E_x A(\sigma) + A(x)E_\sigma + E_x E_\sigma \qquad x \in \Sigma^* \qquad \sigma \in \Sigma$$

The previous two theorems taken together tell us that the most general form of stability transformation is uniquely determined by the error matrices of the symbols.

Error matrices for some of the equivalences defined in previous chapters will be characterized.

6.3 ENUMERATION OF STABLE MACHINES FOR $\equiv_I$, $\equiv_N$ AND $\equiv_E$

Let us first consider indistinguishability $\equiv_I$. As one might expect, those machines which are indistinguishable from every initial state have closely related symbol matrices.

Theorem 6.3

A' is a $\equiv_I$-stable perturbation of A iff there exists a subspace

$$V \subset \text{Kern.} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \quad \text{such that}$$

(i) $A'(y/\sigma) = A(y/\sigma) + H_{(y/\sigma)}$ where $(H_{(y/\sigma)})_i \; \varepsilon \; V, \; i = 1, 2, \ldots, n$

(ii) $V \, A(y/\sigma) \subset V$ for all $y \; \varepsilon \; Y, \;\; \sigma \; \varepsilon \; \Sigma$

Proof: (necessity)

$$S_i A(y/\sigma) \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = S_i A'(y/\sigma) \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \qquad y \; \varepsilon \; Y, \; \sigma \; \varepsilon \; \Sigma \qquad i = 1, 2, \ldots, n$$

Hence

$$A(y/\sigma) \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = A'(y/\sigma) \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \qquad y \; \varepsilon \; Y, \; \sigma \; \varepsilon \; \Sigma$$

The solution of the above equation is a particular solution plus a kernel:

$$A'(y/\sigma) = A(y/\sigma) + H_{(y/\sigma)} \quad \text{where } H_{(y/\sigma)} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$$

For strings of length 2 we have

$$A(y_1y_2/\sigma_1\sigma_2) \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = A'(y_1y_2/\sigma_1\sigma_2) \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \quad y_1, y_2 \; \varepsilon \; Y, \quad \sigma_1, \sigma_2 \; \varepsilon \; \Sigma$$

Hence

$$A'(y_1y_2/\sigma_1\sigma_2) = A(y_1y_2/\sigma_1\sigma_2) + H_2$$

where

$$H_2 \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$$

But

$$A'(y_1y_2/\sigma_1\sigma_2) = [A(y_1/\sigma_1) + H_{(y_1/\sigma_1)}][A(y_2/\sigma_2) + H_{(y_2/\sigma_2)}]$$

$$= A(y_1y_2/\sigma_1\sigma_2) + H_{(y_1/\sigma_1)}A(y_2/\sigma_2) + H'$$

where

$$H' = A(y_1/\sigma_1)H_{(y_2/\sigma_2)} + H_{(y_1/\sigma_1)}H_{(y_2/\sigma_2)}$$

Hence

$$H_2 \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = H_{(y_1/\sigma_1)}A(y_2/\sigma_2) \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} + H' \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

$$0 = H_{(y_1/\sigma_1)}A(y_2/\sigma_2) \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} + 0$$

i.e., $H_{(y_1/\sigma_1)}A(y_2/\sigma_2) \in \text{Kern}\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \qquad \forall y_1, y_2 \in Y \qquad \forall \sigma_1, \sigma_2 \in \Sigma$

(Sufficiency)

Let $y = y_1 \ldots y_r : y_i \in Y$ be any sequence of outputs and $x = x_1 \ldots x_r : x_i \in \Sigma$ any sequence of inputs

$$A'(y/x) = \prod_{i=1}^{r} [A(y_i/x_i) + H_{(y_i/x_i)}]$$

$$= A(y/x) + A(y_1 \ldots y_{r-1}/x_1 \ldots x_{r-1})H_{(y_r/x_r)} + \ldots + H_{(y_1/x_1)}A(y_2 \ldots y_r/x_2 \ldots x_r)$$

$$= A(y/x) + \ldots + (\ldots(H_{(y_1/x_1)}A(y_2/x_2)) \ldots A(y_r/x_r)$$

$$= A(y/x) + H' \text{ where } (H')_i \varepsilon \text{ Kern.}\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}$$

Hence

$$A'(y/x)\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix} = A(y/x)\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}$$

By substituting $A(x)$ for $A(y/x)$ and $\text{Kern.}(F^r)$, $r = 1, \ldots, N$, for $\text{Kern}\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}$ in the proof of Theorem 6.3, a proof (similar to the proof of Theorem 2.5) of an analogous result is obtained for $\equiv_N$. Table 6.1 summarizes these results.

	Stability Class	Condition on Error Matrices
$\equiv_I$	$A'(y/\sigma) = A(y/\sigma) + H_{(y/\sigma)}$	$V \supset \langle\{H_{(y/\sigma)_i} : i = 1, 2, \ldots, n, y\varepsilon Y, \sigma\varepsilon\Sigma\}\rangle$ $V \circ A(y/\sigma) \subset V \subset \text{Kern.}\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}$
$\equiv_N$	$A'(\sigma) = A(\sigma) + H_\sigma$	$V \supset \langle\{(H_\sigma)_i : i = 1, 2, \ldots, n; \sigma\varepsilon\Sigma\rangle$ $V\ A(\sigma) \subset V \subset \bigcap_{r=1}^{N} \text{Kern.}(F^r)$
$\equiv_E$	$A'(\sigma) = A(\sigma) + H_\sigma$	$V \supset \langle\{(H_\sigma)_i : i = 1, 2, \ldots, n; \sigma\varepsilon\Sigma\rangle$ $V\ A(\sigma) \subset V \subset \text{Kern.}(F)$

Table 6.1. Characterization of the error matrices for $\equiv_I$, $\equiv_N$, and $\equiv_E$

Example 6.1

We take the following example from the literature of decomposition of probabilistic machines. Bacon [1964] has a notion of decomposition which is related to the error matrices of $\equiv_I$. The following example is due to Bacon but is expressed in our notation.

$$M = \langle 4, \quad , \{u_1, u_2\}, M(u_1), M(u_2), F_M, 0_M \rangle$$

where

$$F_M = \begin{pmatrix} F_1 \\ F_1 \\ F_3 \\ F_3 \end{pmatrix}$$

$$M(u_1) = \begin{pmatrix} .2 & .2 & .3 & .3 \\ 0 & .4 & 0 & .6 \\ .4 & .1 & .4 & .1 \\ .25 & .25 & .25 & .25 \end{pmatrix} \qquad M(u_2) = \begin{pmatrix} .3 & 0 & .7 & 0 \\ 0 & .3 & 0 & .7 \\ .08 & .12 & .32 & .48 \\ .06 & .14 & .24 & .56 \end{pmatrix}$$

Hence by definition $Y = \{F_1, F_3\}$

$$M(F_1/u_1) = \begin{pmatrix} .2 & .2 & 0 & 0 \\ 0 & .4 & 0 & 0 \\ .4 & .1 & 0 & 0 \\ .25 & .25 & 0 & 0 \end{pmatrix} \qquad M(F_3/u_1) = \begin{pmatrix} 0 & 0 & .3 & .3 \\ 0 & 0 & 0 & .6 \\ 0 & 0 & .4 & .1 \\ 0 & 0 & .25 & .25 \end{pmatrix}$$

$$M(F_1/u_1) = \begin{pmatrix} .3 & 0 & 0 & 0 \\ 0 & .3 & 0 & 0 \\ .08 & .12 & 0 & 0 \\ .06 & .14 & 0 & 0 \end{pmatrix} \qquad M(F_3/u_2) = \begin{pmatrix} 0 & 0 & .7 & 0 \\ 0 & 0 & 0 & .7 \\ 0 & 0 & .32 & .48 \\ 0 & 0 & .24 & .56 \end{pmatrix}$$

The subspace $\langle \{(1, -1, 0, 0), (0, 0, 1, -1)\} \rangle = V$ is invariant under all $M(F_i/u_j)$ $i = 1, 2;$ $j = 1, 2,$ since:

$$(1, -1, 0, 0)\, M(F_1/u_1) = .2\,(1, -1, 0, 0)$$
$$(0, 0, 1, 1)\, M(F_1/u_1) = .15(1, -1, 0, 0)$$
$$(1, -1, 0, 0)\, M(F_3/u_1) = .3\,(0, 0, 1, -1)$$
$$(0, 0, 1, -1)\, M(F_3/u_1) = .15(0, 0, 1, -1)$$
$$(1, -1, 0, 0)\, M(F_1/u_2) = .3\,(1, -1, 0, 0)$$
$$(0, 0, 1, -1)\, M(F_1/u_2) = .15(1, -1, 0, 0)$$
$$(1, -1, 0, 0)\, M(F_3/u_2) = .7\,(0, 0, 1, -1)$$
$$(0, 0, 1, -1)\, M(F_3/u_2) = .08(0, 0, 1, -1)$$

Furthermore we note that $V \subset \mathrm{Kern}.(F)$

$$(1,\ -1,\ 0,\ 0) \begin{pmatrix} F_1 \\ F_1 \\ F_3 \\ F_3 \end{pmatrix} = 0$$

$$(0,\ 0,\ 1,\ -1) \begin{pmatrix} F_1 \\ F_1 \\ F_3 \\ F_3 \end{pmatrix} = 0$$

There seems to be a close relationship between Bacon's definition of lumpability of a matrix and ours of error matrices of a matrix. We shall take up this matter again in Chapter 10.

6.4 INVARIANT ERROR MATRICES

Table 6.1 shows that the rows of each error matrix come from a particular invariant subspace. In the next section, some general results will be proved about such error matrices. Hence we make the following definition in order to discuss error matrices without considering the kernel in which the invariant subspace is contained.

Definition 6.2: H_σ is an invariant error matrix for symbol σ of machine A if there is a stochastic matrix $A'(\sigma)$ such that

(i) $A'(\sigma) = A(\sigma) + H_\sigma$

(ii) There is a subspace V such that

$$V\,A(\sigma) \subset V \quad \forall \sigma \in \Sigma \quad \text{and } (H_\sigma)_i \in V \quad i = 1, 2, \ldots, n$$

If there is some symbol σ' such that $H_{\sigma'} \neq 0$ then there is a machine A' with symbol matrices $\{A'(\sigma) : \sigma \in \Sigma\}$ such that $A' \neq A$.

The simplest invariant subspaces are those of dimension one. Most of the results of this section will deal with such one dimensional invariant subspaces. Recall that a vector v is an eigenvector of matrix

A with eigenvalue λ if $v \cdot A = \lambda v$. It is clear that a one dimensional subspace invariant under a linear transformation is just all the multiples of an eigenvector of the transformation.

Definition 6.3: H_σ is an eigenerror matrix for symbol σ of machine A if Definition 6.2 holds with the subspace V of part (ii) being one dimensional.

Some symbol matrices give rise to non-zero eigenerrors while other symbol matrices have only the trivial zero eigenerror. A general theory which characterizes the existence or non-existence of nontrivial eigenerrors from elementary properties of the symbol matrices does not exist. The next few results deal with certain types of machines which help to show the scope required of a general theory.

Theorem 6.4

Probabilistic sequential machine A has nontrivial eigenerror matrices for every symbol σ if there is a vector $v \neq 0$ such that

(i) $vA(\sigma) = \lambda_\sigma v \qquad \forall \sigma \in \Sigma$

(ii) $\sum_{i=1}^{n} v_i = 0$

(iii) There exist constants $K(i,\sigma)$ not all equal to zero for any σ such that $1 \geq A(\sigma)_{ij} + K_{(i,\sigma)} \cdot v_j \geq 0$

Proof: Construct

$$H_\sigma = \left\| \begin{matrix} K_{(1,\sigma)}v \\ \vdots \\ K_{(n,\sigma)}v \end{matrix} \right\|$$

$$A(\sigma)_{ij} + (H_\sigma)_{ij} = A'(\sigma)_{ij}$$

$A'(\sigma)$ is a stochastic matrix because of (ii) and (iii) above:

Furthermore:

$$H_{\sigma_2} \cdot A(\sigma_1) = \left\| \begin{matrix} \lambda_{\sigma_1} \cdot K_{(1,\sigma_2)} v \\ \vdots \\ \lambda_{\sigma_1} \cdot K_{(n,\sigma_2)} v \end{matrix} \right\| \qquad \text{all } \sigma_2, \sigma_1 \in \Sigma$$

i.e., $(H_{\sigma_2})_i \in V = \langle v \rangle$ and $[(H_{\sigma_2})A(\sigma_1)]_i \in V$ $\quad i = 1, 2, \ldots, n$

Consequently, by (iii) $H_\sigma \neq 0$ for any σ. All $H_\sigma : \sigma \in \Sigma$ are invariant error matrices and are one dimensional by construction.

Q. E. D.

Theorem 6.4 provides a simple method for constructing stable perturbations of some machines. Indeed, it was this result which was used to construct most of the examples of Chapter 1 and Chapter 2.

Cycles are one of the most important features of the state behavior of machine. There is a relationship for deterministic machines between cycles in the state behavior and invariant subspaces of the transition matrices. Hence machines with cycles will now be considered. The next theorem, with the aid of the following lemma, shows that the only eigenerror of a symbol which causes a deterministic cycle of all the states is the trivial zero eigenerror. On the other hand, Example 6.2 shows that symbols causing deterministic cycles which do not involve all the states may have nontrivial eigenerrors.

Lemma 6.1:

Let P be an n x n permutation matrix of the cyclic permutation Π of the indices of the states. The only real eigenvectors of P are

$v = (v_1, \ldots, v_1)$ if $\lambda_P = +1$ and

$v = (v_1, -v_1, \ldots, v_1, -v_1)$ if $\lambda_P = -1$ and n is even.

Proof: Since P is a permutation matrix, the real eigenvalues of

P are +1 and -1.

Case (i) $\lambda_P = +1$

$$v(P)^r = v \qquad r = 1, 2, \ldots, n+1$$

which gives

$$v_i = v_{\pi(i)} = v_{\pi^2(i)} = \ldots = v_{\pi^n(i)} = v_i$$
$$i = 1, 2, \ldots, n$$

For each k = 1, 2, ..., n there is some j such that $\Pi^j(i) = k$ hence

$$v_1 = v_2 = \ldots v_n$$

Case (ii) $\lambda_P = -1$

$$vP^r = (-1)^r v \qquad r = 1, 2, \ldots, n+1$$

i.e., $v_i = v_{\Pi(i)} = v_{\Pi^2(i)} = \ldots (-1)^n v_{\Pi^n(i)} = v_i$

If n is odd we get $v_i = -v_i = 0$. But if n is even

$$v = (v_1, -v_1, \ldots, v_1, -v_1)$$

Theorem 6.5

Let A be a machine with 3 states or more such that there is a symbol σ so that A(σ) is a cyclic permutation of all the states. Then A(σ) has no (nontrivial) eigenerror matrix.

Proof: Suppose to the contrary that A(σ) has eigenerror matrix H_σ. The rows of H_σ must be multiples of some eigenvector v i.e.

$$H_\sigma = \begin{pmatrix} d_1 & 0 & \ldots & 0 \\ \vdots & \ddots & & \vdots \\ & & \ddots & \vdots \\ 0 & & & d_n \end{pmatrix} \begin{pmatrix} v \\ \vdots \\ v \end{pmatrix}$$

In order for H_σ to be nonzero, we require some $d_i \neq 0$ and $v \neq (0, \ldots, 0)$.

Since $(H_\sigma)_i$ is invariant under each symbol matrix, it is also invariant under A(σ):

$$(H_\sigma)_i A(\sigma) = d_i(vA(\sigma)) = d_i \lambda_\sigma \cdot v$$

That is, v is an eigenvector of $A(\sigma)$. But by Lemma 6.1, the only non-trivial real eigenvectors of $A(\sigma)$ are

$$v = (v_1,\ldots,v_1) \text{ if } \lambda_\sigma = 1$$

$$v' = (v_1,-v_1,\ldots,v_1,-v_1) \text{ if } \lambda_\sigma = -1 \text{ and n is even.}$$

The vector v can not serve as an eigenerror since for constant c

$$\sum_{j=1}^{n} cv_j = cnv_1 \neq 0 \text{ for } v_1 \neq 0$$

The vector $d_{i'}v'$ has n/2 negative components if $v_1 \neq 0$. Hence adding this nonzero vector to row i' of $A(\sigma)$ causes at least n/2-1 negative entries in the corresponding row of $A'(\sigma)$ so that $A'(\sigma)$ is not stochastic, which is a contradiction.

Q. E. D.

Example 6.2

A symbol which causes cycles whose lengths are shorter than the number of states may possess non-trivial eigenerrors.

$$\text{Let } A(K) = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

$$\lambda_1 = 1,\ \lambda_2 = 1,\ \lambda_3 = -1$$

Let the subspace associated with 1 be U_1, with -1 be U_{-1}.

$$U_1 = \{v : v = (v_1,-v_2,+v_1)\}$$

$$U_{-1} = \{v : v = (v_1,\ 0,-v_1)\}$$

For rows 1 and 3 there are no eigenerrors from the space U_1. However, there can be eigenerrors from the space U_{-1}.

$$v_1 = (p_1, 0, -p_1) \qquad 1 \geq p_1 \geq 0 \qquad \text{is by definition}$$

an eigenerror for row 1 and $-c \cdot v_1$ is an eigenerror for row 3 for $0 \leq c \leq 1$.

On the other hand,

$$v_2 \in U_{-1} : v_2 = (+p_2, -2 \cdot p_2, p_2) \qquad 1/2 \geq p_2 \geq 0$$

is an eigenerror pattern for row 2.

However $U_1 \cap U_{-1} = \{(0, 0, 0)\}$ as is the case in general for the spaces associated with eigenvectors of different eigenvalues.

Hence we can have eigenerror patterns of either

$$A'(K) = A(K) + \begin{pmatrix} p_1 & 0 & -p_1 \\ 0 & 0 & 0 \\ -cp_1 & 0 & cp_1 \end{pmatrix}$$

or

$$A'(K) = A(K) + \begin{pmatrix} 0 & 0 & 0 \\ p_2 & -2 \cdot p_2 & p_2 \\ 0 & 0 & 0 \end{pmatrix} \qquad 1/2 \geq p_2 \geq 0$$

Note that the fixed point S_2 has the eigenerror pattern of the positive eigenvalue. It is possible that this is true in general, but neither a general proof nor a counter-example has been found.

6.5 STABILITY TRANSFORMATIONS WHICH PRESERVE THE BEHAVIOR OF EIGENSTATES

Suppose $T : A \rightarrow A'$ is a stability transformation which maps the stochastic eigenvector v_σ of $A(\sigma)$ into v'_σ, i.e., $v_\sigma T(A(\sigma)) = v'_\sigma$. Thus while machine A stays in distribution v_σ for input sequences σ^r, $r = 1, 2, \ldots$, machine A' will travel through a trajectory of distributions $v_\sigma(TA(\sigma))^r$, $r = 1, 2, \ldots$ Hence only for certain types of equivalences between machines can A and A' be equivalent; i.e., the behavior of the trajectory $v_\sigma A'(\sigma^r) : r = 1, 2, \ldots$ must be equivalent to that of the distribution v_σ. Let us formalize this concept.

Definition 6.4: A stochastic vector v_σ is an eigenstate for input σ of machine A if $v_\sigma A(\sigma) = v_\sigma$ for some $\sigma \varepsilon \Sigma$.

Definition 6.5: The machine A started in initial distribution I will be written $\langle I, A \rangle$.

Definition 6.6: A stability transformation $T : A \longrightarrow A'$ is eigenstate behavior preserving for behavioral equivalence $\equiv$ if for all eigenstates v_σ of A

$$\langle v_\sigma, A' \rangle \equiv \langle v_\sigma A'(\sigma^r), A' \rangle \qquad r = 1, 2, \ldots$$

Theorem 6.6

If any stability transformation $T : A \longrightarrow A'$ preserves $\equiv_N$, i.e.

$$T(A(\sigma)) = A'(\sigma) = A(\sigma) + H_\sigma \qquad \sigma \varepsilon \Sigma$$

where $H = \{(H)_i : i = 1, 2, \ldots, n: \quad \forall \sigma \varepsilon \Sigma\}$

and there is a space V such that

$$H \subset V \subset \bigcap_{k=1}^{N} \text{Kern.}(F^k)$$

$$V \cdot A(\sigma) \subset V$$

then T is eigenstate behavior preserving for $\equiv_N$.

Proof:

$$v_\sigma A(\sigma) = v_\sigma$$

$$v_\sigma A(\sigma)^r = v_\sigma = v_\sigma A(\sigma^r)$$

By Table 6.1 and Theorem 6.2

$$v_\sigma A'(\sigma^r x) = v_\sigma [A(\sigma^r x) + H_{(\sigma^r x)}]$$

where

$$[H_{(\sigma^r x)}]_i \varepsilon V \qquad i = 1, 2, \ldots, n$$

$$= v_\sigma A(\sigma^r x) + V_\sigma \cdot H_{(\sigma^r x)}$$

multiplying by (F^j) : $j \leq k \leq N$ gives

$$v_\sigma \cdot A'(\sigma^r x)(F^j) = v_\sigma \cdot A(\sigma^r x)(F^j) + v_\sigma(H_{(\sigma^r x)}(F^j))$$

since $H \subset V \subset \bigcap_{k=1}^{N} \text{Kern.}(F^k)$

$$= v_\sigma \; A(\sigma^r x)(F^j)$$

$$= (v_\sigma \; A(\sigma^r))A(x)(F^j)$$

$$= v_\sigma \; A(x)(F^j)$$

Since $\mu_k^A(x) = \mu_k^{A'}(x)$ by Table 6.1, we get for any $j \leq k$

$$= v_\sigma \; A'(x)(F^j)$$

we have

$$(v_\sigma \; A'(\sigma^r))A'(x)(F^j) = v_\sigma \; A'(x)(F^j) \quad r = 1, 2, \ldots \qquad j \leq k \leq N$$

which gives

$$\langle v_\sigma \; A'(\sigma^r), A'\rangle \equiv_N \langle v_\sigma, A'\rangle$$

Q. E. D.

Theorem 6.7

If any stability transformation $T : A \longrightarrow A'$ preserves $\equiv_I$, i.e.

$$T(A(y/\sigma)) = A'(y/\sigma) = A(y/\sigma) + H_{(y/\sigma)} \qquad \sigma \in \Sigma, \; y \in Y$$

where $H = \langle\{(H_{(y/\sigma)})_i : i = 1, 2, \ldots, n; \;\; \forall\sigma \in \Sigma, \;\; \forall y \in Y\}\rangle$

and there is a space V such that

$$H \subset V \subset \text{Kern.}\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}$$

$$V \; A(y/\sigma) \subset V$$

Then T is eigenstate behavior preserving for $\equiv_I$.

Proof: Suppose $\qquad v_\sigma \; A(\sigma) = v_\sigma$

Let y range over all output sequences of length r; i.e. $y \in (Y)^r$.

$v_\sigma A'(yw/\sigma^r x) = v_\sigma[A(yw/\sigma^r x) + H']$ for all $w \varepsilon y^*$, $x \varepsilon \Sigma^*$ and some H' depending on yw and $\sigma^r x$.

Summing over each side

$$\sum_{y\varepsilon(Y)^r} v_\sigma A'(yw/\sigma^r x) = \sum_{y\varepsilon(Y)^r}[v_\sigma A(yw/\sigma^r x) + v_\sigma H']$$

$$\sum_{y\varepsilon(Y)^r} v_\sigma A'(y/\sigma^r)A'(w/x) = \sum_{y\varepsilon(Y)^r}[v_\sigma A(y/\sigma^r)A(w/x) + v_\sigma H']$$

Noting that

$$\sum_{y\varepsilon(Y)^r} A(y/\sigma^r) = A(\sigma^r)$$

since if $y = y_p y_i \quad y_i \varepsilon Y \quad y_p \varepsilon (Y)^{r-1}$

then

$$\sum_{y_p y_i \varepsilon(Y)^r} A(y_p y_i/\sigma^r) = \sum_{y_p\varepsilon(Y)^{r-1}} A(y_p/\sigma^{r-1}) \sum_{y_i\varepsilon Y} A(y_i/\sigma)$$

$$= \sum_{y_p\varepsilon(Y)^{r-1}} A(y_p/\sigma^{r-1})A(\sigma)$$

Continuing the process (or using a formal induction) we get the result.

Hence we obtain

$$(v_\sigma A'(\sigma^r))A'(w/x) = v_\sigma A(\sigma^r)A(w/x) + v_\sigma H''$$

where $H'' = \sum_{y\varepsilon(Y)^r} H'$

$$= v_\sigma A(\sigma)^r A(w/x) + v_\sigma H''$$

$$= v_\sigma A(w/x) + v_\sigma H''$$

multiplying by $\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}$ and noting that H is closed under addition we get

$$(v_\sigma A'(\sigma^r))A'(w/x)\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix} = v_\sigma A(w/x)\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix} + v_\sigma H''\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}$$

$$= v_\sigma \; A(w/x)\begin{pmatrix}1\\ \vdots \\ 1\end{pmatrix}$$

By definition

$$P^{A'}_{(v_\sigma \; A'(\sigma^r))}(w/x) = P^A_{v_\sigma}(w/x)$$

But the transformation T is a stability transformation for any initial state — which here implies any initial distribution.

$$P^{A'}_{v_\sigma}(w/x) = P^A_{v_\sigma}(w/x)$$

So by transitivity of equality we get

$$P^{A'}_{(v_\sigma \; A'(\sigma^r))}(w/x) = P^{A'}_{v_\sigma}(w/x)$$

which means

$$\langle v_\sigma \; A'(\sigma^r), A'\rangle \equiv_I \langle v_\sigma, A'\rangle$$

which shows that the stability transformation T is eigenstate behavior preserving.

Q. E. D.

CHAPTER 7

ON THE STABILITY PROBLEM FOR TAPE EQUIVALENCE $\equiv_T$

7.1 INTRODUCTION

Suppose malfunctions cause permanent changes in the symbol matrices of a machine A, producing a machine A' where

$$A'(\sigma) = A(\sigma) + E_\sigma \qquad \sigma \in \Sigma$$

As in Chapter 6, E_σ will be called the "error matrix" or "error pattern" for the symbol σ. The stability problem for $\equiv_T$ considered here involves characterizing those E_σ so that for any initial state distribution, the set of tapes accepted is unchanged. Results which are true for any initial state are certainly true for particular initial states; hence the results of this chapter imply certain facts relevant to the Rabin problem. Usually such facts will be presented as corollaries.

Definition 7.1: A' is a λ-stable perturbation of A if Definition 6.1 holds for $\equiv_T$ and $\lambda = \lambda'$.

Although we are interested in λ-stable perturbations of a machine, when a result holds for unequal cutpoints, i.e., $\lambda \neq \lambda'$, it will be proved for the general cutpoint case.

7.2 FARKAS' LEMMA AND TAPE ACCEPTANCE FOR NONSINGULAR MACHINES

Definition 7.2: A string x of inputs to a probabilistic sequential machine A is backward deterministic if knowing x begins at time $t-\ell g(x)$ and the distribution at time t, then the distribution at time $t-\ell g(x)$ can be deduced.

The notion of backward determinism has been discussed by Burks [1957] for the deterministic case. This definition is equivalent to requiring the matrices of the symbols which occur in x to be nonsingular.

If all $\sigma \varepsilon \Sigma$ are backward deterministic, the machine A will be called "backward deterministic" or more frequently "nonsingular".

Farkas' lemma is widely used in the study of convex sets and linear programming. For backward deterministic input strings for probabilistic sequential machines, it gives a necessary and sufficient condition that they will be accepted. A nonhomogeneous form of Farkas' lemma (Thrall and Tornheim, <u>Vector Spaces and Matrices</u>, pp. 291-292, Theorems 11.4 C and D) will be used. The theorem will be quoted from Thrall and Tornheim and then restricted to our application using notation appropriate to the stability problem.

"Theorem 11.4 D. Let A be an m-by-n matrix, let δ be a vector in V_m^c (m-dimensional column space), let k be a scalar, and suppose that there is at least one vector ϕ in V_n^r (n-dimensional row space) such that

$$A\phi \geqq \delta \qquad (1)$$

Then a vector α in V_n^r will satisfy the condition $\alpha\phi \geq k$ for all ϕ that satisfy (1) if and only if there exists a vector $\gamma \geq 0$ in V_n^r such that $\alpha = \gamma A$ and $\gamma\delta \geq k$."

To apply Farkas' lemma to probabilistic machines, the above notation will be changed and simplifications made. Restricing α to the set of n-dimensional stochastic vectors, we replace it by I. F replaces δ while λ replaces k. Only matrices A whose inverses are stochastic matrices describing the state transitions of a machine will be considered, hence A is replaced by $A(x)^{-1}$.

Note that $A(x)^{-1}\phi \geqq F$ has the solution $\phi = A(x)F$ which permits simplification of the statement of the lemma. In addition, $\gamma = IA(x)$ is stochastic since I and A(x) are stochastic. Farkas' lemma can now be restated about nonsingular switching matrices of probabilistic machines which gives a necessary and sufficient condition for tape acceptance.

Theorem 7.1

Let A(x) be an n-by-n nonsingular stochastic matrix. Let F and ϕ be n-dimensional column vectors, let λ be a scalar (such that $\lambda \leq \max_i(F_i)$ to avoid trivial cases). Then an n-dimensional stochastic vector I will satisfy the condition $I\phi \geq \lambda$ for all ϕ such that $A(x)^{-1}\phi \geq F$ if and only if there exists a stochastic vector γ such that $\gamma = IA(x)$ and $IA(x)F \geq \lambda$.

The theorem can be restated in the following symbolic form:

$$I \varepsilon \bigcap_{\substack{\phi: \\ \phi \varepsilon K(A(x)^{-1}, F)}} \{I : I\phi \geq \lambda\} \Longleftrightarrow IA(x)F \geq \lambda \Longleftrightarrow x \ \varepsilon \ T(A, \lambda)$$

where

$$K(A(x)^{-1}, F) = \{\phi : A(x)^{-1}\phi > F\} \text{ which is a convex set.}$$

Example 7.1 The convex set $K(A(x)^{-1}, F)$

$$\text{Let } A(x) = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \\ 0 & 0 & 1 \end{pmatrix} \qquad F = \begin{pmatrix} F_1 \\ F_2 \\ F_3 \end{pmatrix}$$

$$A(x)^{-1} = \begin{pmatrix} 2 & -2 & 1 \\ 0 & 2 & -1 \\ 0 & 0 & 1 \end{pmatrix}$$

$K[A(x)^{-1}, F]$ = all $\phi = (\phi_1, \phi_2, \phi_3)$ such that

$$2 \cdot \phi_1 - 2 \cdot \phi_2 + \phi_3 \geq F_1$$

$$2 \cdot \phi_2 - \phi_3 \geq F_2$$

$$\phi_3 \geq F_3$$

Solving the above inequalities gives:

$$\phi_1 \geq F_1 + F_2$$

$$\phi_2 \geq \frac{F_2 + F_3}{2}$$

$$\phi_3 \geq F_3$$

Figure 7.1 shows $K(A(x)^{-1},F)$ for this case.

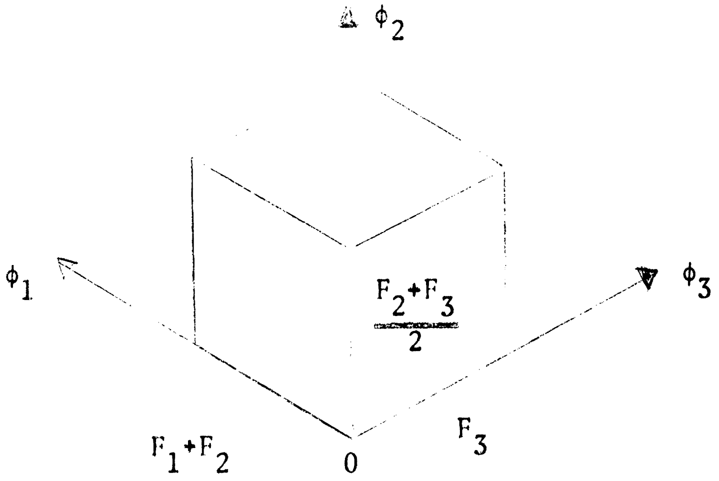


Figure 7.1. $K(A(x)^{-1},F)$ of Example 7.1 is the positive orthant with the rectangular solid removed.

7.3 A NECESSARY AND SUFFICIENT CONDITION FOR λ-STABILITY FOR NON-SINGULAR MACHINES

Theorem 7.2 (Stability result)

Let A and A' be nonsingular probabilistic sequential machines both with initial distribution I.

$$[T(A,\lambda) = T(A',\lambda')\ \forall I \in S^+] \text{ iff } \bigcap_{\substack{\phi: \\ \phi \in K(A(x)^{-1},F)}} \{I : I\phi \geq \lambda\} = \bigcap_{\substack{\phi: \\ \phi \in K(A'(x)^{-1},F')}} \{I : I\phi \geq \lambda'\}\ \forall x \in \Sigma^*$$

Proof: Using the symbolic formulation of Theorem 7.1

$$x \in T(A,\lambda) \Longleftrightarrow I \in \bigcap_{\substack{\phi: \\ \phi \in K(A(x)^{-1},F)}} \{I : I\phi > \lambda\}$$

$$x \in T(A',\lambda') \Longleftrightarrow I \in \bigcap_{\substack{\phi: \\ \phi \in K(A(x)^{-1},F')}} \{I : I\phi \geq \lambda'\}$$

Assuming that

$$T(A,\lambda) = T(A',\lambda') \qquad \forall I \in S^+$$

gives the necessary and sufficient condition

$$\forall x \in \Sigma^* : \bigcap_{\substack{\phi: \\ \phi \in K(A(x)^{-1},F)}} \{I : I\phi \geq \lambda\} = \bigcap_{\substack{\phi: \\ \phi \in K(A'(x)^{-1},F')}} \{I : I\phi \geq \lambda'\}$$

Q. E. D.

Corollary 7.2 (Application to the Rabin stability problem)

Let A and A' be nonsingular probabilistic sequential machines with initial distributions I and I' respectively.

$$T(A,\lambda) = T(A',\lambda) \Longleftrightarrow [I \in \bigcap_{\substack{\phi: \\ \phi \in K(A(x)^{-1},F)}} \{I : I\phi \geq \lambda\} \Longleftrightarrow I' \in \bigcap_{\substack{\phi: \\ \phi \in K(A'(x)^{-1},F')}} \{I : I\phi \geq \lambda\} \text{ for all } x \in \Sigma^*]$$

7.4 A SPECIAL CLASS OF STABILITY TRANSFORMATIONS FOR NONSINGULAR MACHINES

Theorem 7.2 is identically true when $K(A(x)^{-1},F) = K(A'(x)^{-1},F')$. Let us investigate the constraints imposed upon the stability transformation $T : A \to A'$ by this equality, assuming as well that $F = F'$.

Theorem 7.3

$K(A(x)^{-1},F) = K(A'(x)^{-1},F)$ iff $A'(x) = A(x)\pi_x$

where (i) π_x is a permutation matrix

(ii) $(\pi_x)F = F$

Proof: We establish the theorem by showing a series of lemmas.

Lemma 7.1

Let A and B be n x n matrices

$$K(A,F) = \{\phi : A\phi \geq F\}$$

$$K(B,F) = \{\phi : B\phi \geq F\}$$

$K(A,F) = K(B,F)$ iff there exist non-negative matrices Γ_A and Γ_B such that

$$\text{(i)}\quad A = \Gamma_B B$$

$$B = \Gamma_A A$$

$$\text{(ii)}\quad \Gamma_A F \geq F$$

$$\Gamma_B F \geq F$$

Proof: If $K(A,F) \subseteq K(B,F)$ then for all ϕ in $K(A,F)$: $[A\phi \geq F \Rightarrow B\phi \geq F]$

Let B_i = i'th row of B $\quad F_i$ = i'th row of F $\quad i = 1, 2, \ldots, n$

Then the above becomes:

$$[A\phi \geq F \Longrightarrow B_i\phi \geq F_i] \text{ for all } \phi \in K(A,F) \quad i = 1, 2, \ldots, n$$

We use Farkas' lemma for each i getting:

$$\exists \gamma_i \varepsilon V_n^r : \gamma_i \geq 0 \text{ and } \begin{array}{ll} (1) & \gamma_i F \geq F_i \\ (2) & B_i = \gamma_i A \end{array} \quad i = 1, 2, \ldots, n$$

Rewriting the above in matrix notation where

γ_i = the i'th row of Γ_A we obtain:

There is a matrix Γ_A such that $(\Gamma_A)_{ij} \geq 0$

$$\text{(i)}'\quad B = \Gamma_A A$$

$$\text{(ii)}'\quad \Gamma_A F \geq F$$

Now assume $K(B,F) \subseteq K(A,F)$. By the same argument we get there is a matrix Γ_B with $(\Gamma_B)_{ij} \geq 0$ such that

$$\text{(i)}''\quad A = \Gamma_B B$$

$$\text{(ii)}''\quad \Gamma_B F \geq F$$

By (i)', (ii)', (i)", and (ii)" we obtain the necessary part. To obtain the sufficient part of the lemma, multiply the defining inequality for $K(A,F)$ by Γ_A, for $K(B,F)$ by Γ_B and use (2).

Q. E. D.

Lemma 7.2

If A and B are nonsingular

$\Gamma_A \Gamma_B = E_n$: the n-dimensional matrix identity.

Proof:

$$B = \Gamma_A A = \Gamma_A(\Gamma_B B) \Longrightarrow \Gamma_A \Gamma_B = E_n$$

Lemma 7.3

For any n x n nonsingular matrix A such that

$$\sum_{k=1}^{n} A_{ik} = \ell \qquad i = 1, 2, \ldots, n$$

then

$$\sum_{k=1}^{n} (A^{-1})_{ik} = 1/\ell \qquad i = 1, 2, \ldots, n$$

Proof:

$$E_n = A^{-1}A$$

$$= \sum_{k=1}^{n} (A^{-1})_{ik}(A)_{kj}$$

$$1 = \sum_{j=1}^{n} (E_n)_{ij} = \sum_{j=1}^{n} \sum_{k=1}^{n} ((A^{-1})_{ik})(A)_{kj}$$

$$= \sum_{k=1}^{n} (A^{-1})_{ik} \sum_{j=1}^{n} (A)_{kj}$$

$$= \sum_{k=1}^{n} (A^{-1})_{ik} \ell$$

hence

$$\sum_{k=1}^{n} (A^{-1})_{ik} = 1/\ell$$

Lemma 7.4

Let $A = A(x)^{-1}$ $\quad B = A'(x)^{-1}$

then

(i) $\Gamma_A = \Gamma_B^{-1}$

(ii) Γ_A and Γ_B are stochastic matrices

Proof: By Lemma 7.1

$\exists\Gamma_A \geq 0$ so that

$$A'(x)^{-1} = \Gamma_A A(x)^{-1}$$

$$[A'(x)^{-1}]_{ij} = \sum_{k=1}^{n} (\Gamma_A)_{ik}[A(x)^{-1}]_{kj}$$

$$\sum_{j=1}^{n}[A'(x)^{-1}]_{ij} = \sum_{j=1}^{n}\sum_{k=1}^{n} (\Gamma_A)_{ik}[A(x)^{-1}]_{kj}$$

$$= \sum_{k=1}^{n}\sum_{j=1}^{n} (\Gamma_A)_{ik}[A(x)^{-1}]_{kj}$$

$$= \sum_{k=1}^{n} [\Gamma_A]_{ik} \sum_{j=1}^{n} [A(x)^{-1}]_{kj}$$

By the preceding Lemma 7.3

$$\sum_{j=1}^{n} [A(x)^{-1}]_{kj} = 1 \quad \text{since } A(x) \text{ is stochastic}$$

$$\sum_{j=1}^{n} [A'(x)^{-1}]_{kj} = 1 \quad \text{since } A'(x) \text{ is stochastic}$$

hence

$$1 = \sum_{k=1}^{n} (\Gamma_A)_{ik} \qquad i = 1, 2, \ldots, n$$

since $\Gamma_A \geq 0$, by definition Γ_A is a stochastic matrix.

We know by Lemma 7.1 that there exists $\Gamma_B \geq 0$ such that

$$A(x)^{-1} = \Gamma_B A'(x)^{-1}$$

The same argument gives that

Γ_B is stochastic.

By Lemma 7.2 we conclude that

$$\Gamma_A\Gamma_B = E_n \text{ or } \Gamma_A = \Gamma_B^{-1}$$

Q. E. D.

Lemma 7.5

The only stochastic matrices whose inverses are stochastic are permutation matrices.

Proof: Let X be an n x n matrix which is stochastic such that X^{-1} is stochastic.

If X is not a permutation matrix, then some state vector S_i must be mapped by X into a vector with at least two non-zero components. Assume that (W.L.G.)

$$(1, 0, \ldots, 0)X = (a_1, a_2, \ldots, 0)$$

Where $a_1 + a_2 = 1 \qquad a_1 > 0 \qquad a_2 > 0$

$$(a_1, a_2, 0, \ldots, 0)X^{-1} = (1, 0, \ldots, 0) \qquad (*)$$

Let $X^{-1} = ||x'_{ij}||$

The above equation (*) gives

$$a_1x'_{11} + a_2x_{21} = 1$$

$$a_1x'_{1i} + a_2x'_{2i} = 0 \qquad i = 2, \ldots, n$$

But $1 \geq x'_{ij} \geq 0$ because x^{-1} is stochastic.

Hence the only solution to the previous equations

$$x'_{21} = x'_{11} = 1, \ x'_{1i} = x'_{2i} = 0 \qquad i = 2, \ldots, n$$

But this requires that X^{-1} be of the form

$$X^{-1} = \begin{pmatrix} 1 & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ 1 & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ x'_{31} & x'_{32} & \cdot & \cdot & \cdot & \cdot & 0 \\ \vdots & \vdots & & & & & \vdots \\ x'_{n1} & x'_{n2} & \cdot & \cdot & \cdot & \cdot & x'_{nn} \end{pmatrix}$$

which is singular since it has two rows equal, contradicting the non-singularity of X. Hence X must be a permutation matrix. It is well known also that the inverse of a permutation matrix is a permutation matrix.

Q. E. D.

We are now able to prove the theorem.

Proof: (of Theorem 7.3) By Lemma 7.1 there exists Γ_A and Γ_B such that

$$A(x)^{-1} = \Gamma_A A'(x)^{-1}$$

$$A'(x)^{-1} = \Gamma_B A(x)^{-1}$$

and $\Gamma_B F \geq F \qquad \Gamma_A F \geq F$

By Lemma 7.4 $\Gamma_B = \Gamma_A^{-1}$ and Γ_B and Γ_A are stochastic. Multiplication by a stochastic matrix preserves a vector inequality, hence

$$\Gamma_A(\Gamma_B F) \geq \Gamma_A F$$

$$(\Gamma_A \Gamma_B)F = (\Gamma_A \Gamma_A^{-1})F = F$$

$$F \geq \Gamma_A F$$

but $\qquad F \leq \Gamma_A F$

consequently $\qquad F = \Gamma_A F$

By the same argument $\qquad F = \Gamma_B F$

Q. E. D.

Corollary 7.3a

If A and A' are nonsingular probabilistic sequential machines then

$$\forall x \in \Sigma^* \quad K(A(x)^{-1},F) = K(A'(x)^{-1},F) \Longrightarrow \text{for all } I \in S^+ : A \equiv_D A'$$

Proof: Immediate from Theorem 7.3 and the definition of distribution equivalence.

Corollary 7.3b

Let A and A' be probabilistic sequential machines with string x backward deterministic for both machines

$$K(A(x)^{-1},F) = K(A'(x)^{-1},F) \Longrightarrow A'(x) = A(x) + H$$

$$\text{where } H_i \in \text{Kern.}(F) \quad i = 1, 2, \ldots, n$$

Proof:

$F = \Gamma_A F \quad \Gamma_A = E_n + H'$ where $H_i' \varepsilon$ Kern.(F) $i = 1, 2, \ldots, n$

$$A(x)^{-1} = \Gamma_A A'(x)^{-1}$$

$$A'(x) = A(x)\Gamma_A$$

$$A'(x) = A(x)(E_n + H')$$

$$A'(x) = A(x) + A(x)H'$$

But $(A(x)H')F = A(x)(H'F) = 0$ Hence $A(x)H' = H$ where $H_i \varepsilon$ Kern.(F).

Q. E. D.

Hence the result that $\forall x \varepsilon \Sigma^* : K(A(x)^{-1},F) = K(A'(x)^{-1},F)$ guarantees stability implies the result that A and A' are λ-stable if they are distribution equivalent. Let us try to find more general conditions on nonsingular machines A and A' such that $A \equiv_T A'$.

7.5 POLAR SETS OF VECTORS

If u and v are vectors in the same space, the inner product of u and v will be written (u,v). Of course, regarding u and v as 1 x n matrices we have

$$(u,v) = u \cdot v^T = v \cdot u^T$$

Definition 7.3: Let K be a convex set. The polar set of K, written K*, is defined as:

$$K^* = \{u \; ; \; (v,u) \geq 0 \quad \text{for all } v \varepsilon K\}$$

In some definitions of the polar set, ≤ is used in place of ≥ and the constant may be one rather than zero (Radström [1949]). The following identity holds for any of these definitions of the polar set.

Lemma 7.6

If K_1 and K_2 are sets of points in a vector space V, then

$$K_1^* \cap K_2^* = (K_1 \cup K_2)^*$$

Proof: (W.L.G. assume K_1 and K_2 are sets of column vectors)

$$\alpha\varepsilon(K_1 \cup K_2)^* \Rightarrow \alpha \cdot k \geq 0 \quad \text{for all } k \; \varepsilon \; K_1 \cup K_2$$

$$\Rightarrow \alpha \cdot k \geq 0 \quad \text{for } k \; \varepsilon \; K_1 \text{ and } \alpha \cdot k \geq 0 \text{ for } k \; \varepsilon \; K_2$$

$$\Rightarrow \alpha\varepsilon K_1^* \cap K_2^*$$

The argument reverses to give the opposite containment.

Q. E. D.

Example 7.2

Let K be the triangular region shown in Figure 7.2

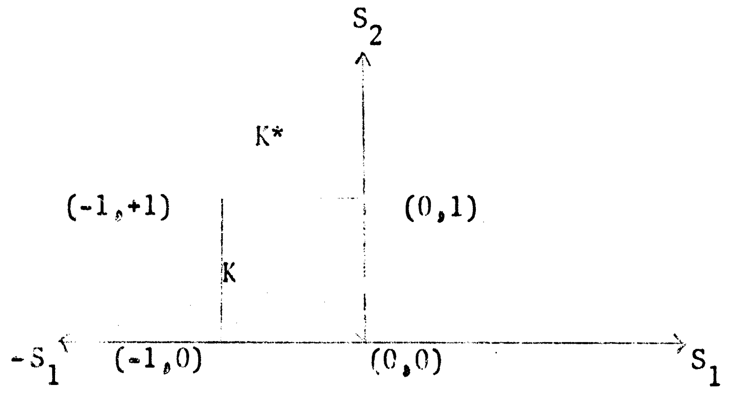


Figure 7.2. Polar set example.

The set K is the region with extreme points (-1,+1), (-1,0) and (0,1). K* is the infinite region bounded by the S_2 axis and the line through (-1,+1) and (0,0).

7.6 TAPE ACCEPTANCE EXPRESSED IN POLAR SET NOTATION

In this section we will show how Theorem 7.2 can be rephrased in the notation of convex sets using the polar set concept.

Theorem 7.4

Let x be a backward deterministic string in both A and A':

$$[x\varepsilon T(A,\lambda) \Longleftrightarrow x\varepsilon T(A',\lambda') \ \forall I\varepsilon S^{+}] \Longleftrightarrow [K[A(x)^{-1}, F-\lambda\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}]^{*}\cap S^{+} = K[A'(x)^{-1}, F'-\lambda'\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}]^{*}\cap S^{+}]$$

Proof:

$$x \ \varepsilon \ T(A,\lambda) \Longleftrightarrow IA(x)F \geqq \lambda$$

since $IA(x)$ is stochastic for all x

$$IA(x)\circ\lambda\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix} = \lambda\bullet IA(x)\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix} = \lambda$$

Hence

$$IA(x)F \geqq \lambda \Longleftrightarrow IA(x)[F - \lambda\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}] \geq 0$$

which gives

$$x \ \varepsilon \ T(A,\lambda) \Longleftrightarrow IA(x)[F - \lambda\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}] \geq 0$$

Likewise we get for A'

$$x \ \varepsilon \ T(A',\lambda') \Longleftrightarrow I'A'(x)[F' - \lambda'\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}] \geq 0$$

In order for the above equations to hold for any initial distribution I, we use Theorem 7.2 getting

$$\bigcap_{\substack{\phi:\\ \phi\varepsilon K[A(x)^{-1}, F - \lambda\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}]}} \{I : I\phi \geq 0\} = \bigcap_{\substack{\phi:\\ \phi\varepsilon K[A'(x)^{-1}, F' - \lambda'\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}]}} \{I' : I'\phi \geq 0\}$$

But if we regard I and ϕ as vectors in the same space

$$I\phi \geq 0 \qquad (I,\phi) \geq 0$$

Hence translation of the above gives

$$K[A(x)^{-1}, F - \lambda\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}]^{*}\cap S^{+} = K[A'(x)^{-1}, F' - \lambda'\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}]^{*}\cap S^{+}$$

Q. E. D.

7.7 STABILITY FOR CERTAIN MACHINES WHICH HAVE SINGULAR SWITCHING MATRICES

Definition 7.4: The set of singular switching matrices of a machine A will be written

$$Z(A) = \{A(\sigma) : \sigma \varepsilon \Sigma \text{ and determ.}(A(\sigma)) = 0\}$$

For some probabilistic machines A with singular switching matrices Z(A), the methods of Chapter 6 can be used to obtain a machine A' such that

$$Z(A') = \phi$$

while

$$A \equiv_E A'$$

so that

$$x \varepsilon T(A,\lambda) \Longleftrightarrow x \varepsilon T(A',\lambda)$$

Indeed, Example 1.1 is of this type if we interchange A and A' i.e., the machine A' has a singular matrix A'(1) while A(1) is nonsingular.

Hence applying Farkas' lemma to the nonsingular machine A' we obtain

$$x \varepsilon T(A,\lambda) = I\varepsilon \bigcap_{\substack{\Phi: \\ \phi\varepsilon K(A'(x)^{-1},F)}} \{I : I\phi > \lambda\}$$

Therefore $A'(x)^{-1}$ plays the role of the inverse of the singular matrix A(x). Unfortunately, finding conditions for the existence of a nonsingular machine A' which is expectation equivalent to an arbitrary machine A contains subproblems concerning invariant error patterns for A left open in Chapter 6.

CHAPTER 8

MINIMAL PROBABILISTIC SEQUENTIAL MACHINES

This chapter will contrast the properties of $\equiv_E$, $\equiv_N$, and $\equiv_I$ with those of $\equiv_T$. The former equivalences lead to minimal machine results which are similar to those of deterministic machine theory. On the other hand, the later equivalence requires such disciplines as convex set theory to discuss minimality. Consequently, such results which can be obtained represent a departure from deterministic machine theory.

8.1 STATE EQUIVALENCE OF MACHINES

In previous chapters, the same symbolism was used to represent both abstract, or arbitrary initial distribution machines, and machines with a fixed initial distribution. In what follows we use $\langle I,A\rangle$ to mean the machine A started in initial distribution I. We let $\equiv$ be a machine equivalence variable ranging over the above equivalences.

Definition 8.1: Machines A and A' are state equivalent for $\equiv$, written $A \equiv(s) A'$. If there is a (possibly multi-valued) function h such that:

$$\langle S_i,A\rangle \equiv \langle S'_{h(i)},A'\rangle \qquad i = 1, 2, \ldots, n$$

and

$$\langle S_{h^{-1}(i)},A\rangle \equiv \langle S'_{i'},A'\rangle \qquad i' = 1, 2, \ldots, n'$$

Definition 8.1 requires that for every initial state of A there is an initial state of A' such that the machines are equivalent and vice-versa.

8.2 EQUIVALENT STATES OF A MACHINE

In order to discuss reduction of a machine to a smaller machine, an equivalence relation on states is needed. They symbol "$\sim$" will be used as a state equivalence variable. When "$\sim$" is used in a context we will mean that any one of a designated class of state equivalences can be substituted for "$\sim$" in the context.

Definition 8.2: States S_i and S_j of a machine A are equivalent by $\sim$ if

$$\langle S_i, A\rangle \equiv \langle S_j, A\rangle$$

Hence for $\equiv_T$, $\equiv_N$, and $\equiv_I$ we have states S_i and S_j are equivalent for:

(a) $\equiv_E$: $S_i \underset{E}{\sim} S_j \Longleftrightarrow S_iA(z)F = S_jA(z)F \qquad \forall z \in \Sigma^*$

(b) $\equiv_N$: $S_i \underset{N}{\sim} S_j \Longleftrightarrow S_iA(z)(F^k) = S_jA(z)(F^k) \qquad \forall z \in \Sigma^* \quad k = 1, 2, \ldots, N$

(c) $\equiv_I$: $S_i \underset{I}{\sim} S_j \Longleftrightarrow S_iA(y/x)\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix} = S_jA(y/x)\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix} \qquad \forall y \in Y^*, \quad \forall x \in \Sigma^*$

8.3 STATE AND DISTRIBUTION REDUCTION OF A MACHINE

Merging some of the states of a machine to get a new (perhaps nondeterministic) machine is a notion from automata theory. For probabilistic machines, Carlyle [1961] has noted that a state S_i can be "merged" with a distribution $\pi = (\pi_1, \ldots, \pi_{i-1}, 0, \pi_{i+1}, \ldots, \pi_n)$ over the other states of the machine. That is, when the machine attempts to enter state S_i, it can be routed to the other states with distribution π. If the i'th row of every symbol matrix happens to be the same convex sum of the other rows, this possibility of reduction holds for every state equivalence, i.e.,

$$[S_iA(\sigma) = \sum_{j\neq i} \pi_j S_j A(\sigma) \qquad \forall \sigma \in \Sigma] \Longrightarrow \langle S_i, A\rangle \equiv \langle \pi, A\rangle$$

As Carlyle has pointed out, a probabilistic sequential machine with a minimal number of states by state reduction may be reduced further by the use of such equivalent distributions. For any real model of a probabilistic machine, such minimizing of states would cause the machine to operate more slowly. Hence the results will be concerned only with equivalent state reduction although the definitions will be general enough for equivalent distribution reduction.

Definition 8.3: The reduced machine $A' = A[\pi(S_i)]$ where distribution π replaces state S_i of machine A is obtained by:

(i) $I'_k = I_k + I_i\pi_k \quad k = 1, 2, \ldots, n; \quad k \neq i$

(ii) The i-th row is deleted in F

(iii) The symbol matrices of $A(\pi(S_i))$ are given by:

$$A[\pi(S_i)](\sigma)_{j,k} = A(\sigma)_{j,k} + A(\sigma)_{j,i}\pi_k$$

$$k = 1, 2, \ldots, n \quad k \neq i$$

$$j = 1, 2, \ldots, n \quad j \neq i$$

This notation for the reduced machine has been selected for intuitive rather than esthetic reasons. If π is activated when S_i is specified, then π depends on S_i, i.e., $\pi(S_i)$.

8.4 REDUCTION IN SEQUENCE

Let $\alpha_1, \ldots, \alpha_m$ be distributions and $S_1, \ldots, S_m$ be states. The replacement of the states by the corresponding distributions will be formalized. Implicit in what follows is the assumption that $(\alpha_r)_r = 0$.

Definition 8.4: The sequentially reduced machine

$A[\ldots[\alpha_1(S_1)]\alpha_2(S_2)]\ldots\alpha_m(S_m)]$ is defined recursively by:

$$A[\ldots[\alpha_1(S_1)]\alpha_2(S_2)]\ldots\alpha_m(S_m)] = A^{m-1}[\alpha_m(S_m)]$$

where

$$A^{r+1} = A^r[\alpha_r(S_r)] \qquad r = 1, 2, \ldots, m-1 \text{ and } A^1 = A$$

If $\alpha_1, \alpha_2, \ldots, \alpha_m$ happen to be states and $\alpha_r = S_{r+1}$ $r = 1, \ldots, m-1$ then the above defines state reduction in sequence. For this special case we use the notation

$$A[\ \ldots\ [S_1 \leftarrow S_2] \leftarrow S_3] \leftarrow \ldots\ ] \leftarrow S_m] = A^{m-1}[S_m(S_{m-1})]$$

Definition 8.5: The index merging function $h_r(i)$ for state reduction in sequence

$$h_r(i) = \begin{cases} r \text{ if } i \leq r \\ i \text{ otherwise} \end{cases}$$

More generally, if $S_{1\#}$ is a representative of a $\sim$ class:

$$h_{1\#}(i) = \begin{cases} 1\# \text{ if } S_i \in \sim[S_{1\#}] \\ i \text{ otherwise} \end{cases}$$

Let $\mathcal{P}$ be a permutation on the integers $1, \ldots, m$. Then we obtain the following results.

Theorem 8.1

Let $\sim$ range over $\sim_E$, $\sim_N$ and $\sim_I$.

If $S_1 \sim S_2$, $S_2 \sim S_3$, $\ldots$, $S_{m-1} \sim S_m$ then

$$A \equiv(s)\ A[\ldots\ [S_1 \leftarrow S_2] \leftarrow \ldots] \leftarrow S_m] \equiv(s)\ A[\ldots\ [S_{\mathcal{P}(1)} \leftarrow S_{\mathcal{P}(2)}] \leftarrow \ldots\ ] \leftarrow S_{\mathcal{P}(m)}]$$

Proof: The proof will branch into parts (a), (b) and (c) corresponding to $\equiv_E$, $\equiv_N$ and $\equiv_I$.

For $r = 1$ $\quad A \equiv(s)\ A^1$. Suppose by induction that

$$A^r = A^{r-1}[S_r(S_{r-1})] \equiv(s)\ A \text{ for } r < m-1$$

Then

$$A^{r+1} = A^r[S_{r+1}(S_r)]$$

and

$$(A^r[S_{r+1}(S_r)](\sigma))_{j,k} = \begin{cases} A^r(\sigma)_{j,r} + A^r(\sigma)_{j,r+1} \text{ for } k = r; \\ \qquad j \neq r, \\ \qquad j = 1, 2, \ldots, n_A r = n' \\ A^r(\sigma)_{j,k} \text{ for } k \neq r, k \neq r+1 \end{cases}$$

(a) We show that for any initial state S_i of A, there is a state $S_{h_{r+1}(i)}$ of $A^r[S_{r+1}(S_r)]$ such that

$$E_A(\sigma) = E_{A^r[S_{r+1}(S_r)]}(\sigma) \qquad \forall \sigma \in \Sigma$$

Let $k = h_{r+1}(i)$

Call

$$A^r[S_{r+1}(S_r)](x) = A'(x) \text{ for } x \in \Sigma^*$$

$$A^r[S_{r+1}(S_r)](\sigma) = A'(\sigma) \text{ for } \sigma \in \Sigma$$

$$A^{r-1}[S_r(S_{r-1})](\sigma) = ||b_{ij}|| \quad i,j = 1, 2, \ldots, n_A r = n'$$

Then

$$A'(\sigma)F' = \begin{pmatrix} b_{1,1} & \cdots & b_{1,r-1} & (b_{1,r} + b_{1,r+1}) & \cdots & b_{1,n'} \\ \vdots \\ b_{r-1,1} \\ b_{r+1,1} \\ \vdots \\ b_{n',1} & \cdots & b_{n',r-1} & (b_{n',r} + b_{n',r+1}) & \cdots & b_{n'n'} \end{pmatrix} \begin{pmatrix} F_1 \\ F_2 \\ \vdots \\ F_{r-1} \\ F_{r+1} \\ \vdots \\ F_{n'} \end{pmatrix}$$

$$(A'(\sigma)F')_k = \sum_{\substack{j=1 \\ j\neq r, r+1}}^{n'} b_{k,j}F_j + (b_{k,r} + b_{k,r+1})F_{r+1}$$

But since $S_r \overset{A}{\sim} S_{r+1}$ and $\Lambda \in \Sigma^*$

$$S_r A(\Lambda)F = S_{r+1}A(\Lambda)F \Longrightarrow F_r = F_{r+1}$$

$$(A'(\sigma)F')_k = \sum_{\substack{j\neq r,r+1\\ j=1}}^{n'} b_{k,j}F_j + b_{k,r}F_r + b_{k,r+1}F_{r+1}$$

$$= \sum_{j=1}^{n'} b_{k,j}F_j = (A^{r-1}[S_r(S_{r-1})](\sigma))_{h_{r+1}(i)} \cdot \begin{pmatrix} F_1 \\ \vdots \\ F_r \\ F_{r+1} \\ \vdots \\ F_{n'} \end{pmatrix}$$

By the induction hypothesis on states

$$A'(\sigma)F'_k = (A(\sigma))_i F = A^{r-1}[S_r(S_{r-1})](\sigma)_{h_{r+1}(i)} \cdot \begin{pmatrix} F_1 \\ \vdots \\ F_r \\ F_{r+1} \\ \vdots \\ F_{n'} \end{pmatrix}$$

By definition

$$= (A^r[S_{r+1}(S_r)](\sigma)F')_{h_{r+1}(i)}$$

which shows that there are states such that A and $A^r[S_{r+1}(S_r)]$ have the same expectation for any string of length one. Hence we have set up the initial condition for an induction on strings $x \in \Sigma^*$ to prove that A and $A^r[S_{r+1}(S_r)]$ are state equivalent for expectation equivalence.

Suppose

$$(A(x)F)_i = (A^r[S_{r+1}(S_r)](x)F')_{h_{r+1}(i)} \text{ for all } x \in (\Sigma)^d \text{ for d finite.}$$

Let $F'' = A'(x)F'$ and call $h_{r+1}(i) = k$. By the induction hypothesis on strings

$$F''_k = (A'(x)F')_k = (A(x)F)_i \qquad (*)$$

$$(A'(\sigma)F'')_k = \sum_{\substack{j=1\\ j\neq r,r+1}}^{n'} b_{k,j}F''_j + b_{k,r+1}F''_r + b_{k,r}F''_r$$

By $(*)$ $F''_r = F''_{r+1}$ and letting $F^* = \begin{pmatrix} F_1 \\ \vdots \\ F_r \\ F_{r+1} \\ \vdots \\ F_{n'} \end{pmatrix}$

we obtain:
$$\begin{aligned} &= (A'(\sigma)A'(x)F'')_k = (A'(\sigma x)F'')_k \\ &= \sum_{j=1}^{n'} b_{k,j}F''_j \\ &= (A^{r-1}[S_r(S_{r-1})](\sigma x))F^*)_k \end{aligned}$$

Noting that F^* is the output vector of $A^{r-1}[S_r(S_{r-1})]$ and using the induction hypothesis on states

$$(A^{r-1}(S_r(S_{r-1})](\sigma x)F^*)_k = (A(\sigma x)F_i$$

Hence

$$(A'(\sigma x)F')_k = (A^r[S_{r+1}(S_r)](\sigma x)F^*)_k = (A(\sigma x)F_i$$

and the length of σx is $d+1$, which completes the induction on strings.

Hence

$$(A^r[S_{r+1}(S_r)](z))_{h_{r+1}(i)}F' = (A(z)F)_i \text{ for all } i \text{ and all } z \in \Sigma^*$$

By definition

$$A^{r+1} \equiv_E(s)\ A$$

which completes the induction on states.

Consequently,

$$A[\ \ldots\ [S_1 \leftarrow S_2] \leftarrow \ldots\ ] \leftarrow S_m] \equiv_E(s)\ A$$

Given any permutation $\mathbb{P}$ of the integers $\{1, 2, \ldots, m\}$ by the transitivity and symmetry of $\overset{A}{\sim}_E$ we can obtain (subject to the hypothesis of the theorem)

$$S_i \overset{A}{\sim}_E S_j \Longrightarrow S_{\mathbb{P}(i)} \overset{A}{\sim}_E S_{\mathbb{P}(j)}$$

Letting $S_{i'} = S_{\wp(i)}$, the proof proceeds exactly as before so that:

$$A[\ \ldots\ [S_1 \leftarrow S_2] \leftarrow \ldots\ ] \leftarrow S_m] \equiv_E (s)\ A[\ \ldots [S_{\wp(1)} \leftarrow S_{\wp(2)}] \leftarrow \ldots\ ] \leftarrow S_{\wp(m)}].$$

(b) N-moment equivalence $\equiv_N$.

It must be shown that for each initial state S_i of A there is a state $S_{h_{r+1}(i)}$ of $A^r[S_{r+1}(S_r)]$ such that for any string x

$$(A(x)F)_i = (A^r[S_{r+1}(S_r)]_{h_{r+1}(i)}(x)F_{A^{r+1}}$$

$$\vdots$$

$$(A(x)(F^N))_i = (A^r[S_{r+1}(S_r)]_{h_{r+1}(i)}(x)(F^N_{A^{r+1}})$$

By substituting (F^k), $k = 1, 2, \ldots, N$ for F in part (a), the proof proceeds in the same way.

(c) Indistinguishability $\equiv_I$:

Define a new alphabet $\Sigma' = \{y/\sigma\colon y \in Y,\ \sigma \in \Sigma\}$ with the string operation o defined by

$$y_1/\sigma_1 \circ y_2/\sigma_2 = y_1 \circ y_2/\sigma_1 \circ \sigma_2.$$

Let F be replaced by $\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$ and the proof proceeds as part (a).

In all cases we have $A \equiv (s)\ A^m$. Hence for any initial distribution I, the i'th equality defining $\overset{A}{\sim}$ for equivalent states can be multiplied by $\sum_{j \in \overset{A}{\sim}[S_i]} I_j$ and adding over i gives $A \equiv A^m$ for any initial distribution.

Q. E. D.

8.5 MINIMAL MACHINES FOR $\equiv_E$, $\equiv_N$ AND $\equiv_I$

The formalism necessary for proving a fundamental result has now been established. In the theorem which follows $\equiv$ will be restricted to $\{\equiv_I, \equiv_E, \equiv_N\}$ and $\sim$ to $\{\sim_I, \sim_E, \sim_N\}$. The reader may suspect that the result will hold for any equivalence relation on states which also has the properties of the classes of the partition of V(A) discussed in Theorem 1.4. However, as we shall see in the next section, for $\sim_T$, Theorem 8.1 does not hold. Consequently the basic result which follows is proved only for the above equivalences.

Theorem 8.2

Let A be a probabilistic sequential machine. There is a machine A' with rank $|\overset{A}{\sim}|$ states such that

$$A \equiv(s)\ A'$$

and any machine B, such that B $\equiv$(s) A has at least rank $|\overset{A}{\sim}|$ states.

Proof: Partition the states of A by $\overset{A}{\sim}$ into classes w_1, ..., w_r having representatives $S_{1\#}$, $S_{2\#}$, ..., $S_{r\#}$. Use Theorem 8.1 to reduce A to A' by collapsing all states in the class $w_1 = \overset{A}{\sim}[S_{1\#}]$.

By Theorem 8.1, regardless of the order of collapsing, we have

$$A' \equiv(S)\ A$$

where we let $h_{1\#}(i)$ be the index merging function for A' and A, i.e.,

$$(A'(x)F')_{h_{1\#}(i)} \equiv (A(x)F)_i \text{ for all } S_i \varepsilon S, \text{ for all } x \varepsilon \Sigma^*.$$

Proceed to the class $\overset{A}{\sim}[S_{2\#}]$. Suppose $S_i \varepsilon \overset{A}{\sim}[S_{2\#}]$ and $S_j \varepsilon \overset{A}{\sim}[S_{2\#}]$. Clearly S_i and S_j are not in $\overset{A}{\sim}[S_{1\#}]$ so $h_{1\#}(i) = i$ and $h_{1\#}(j) = j$. Hence we get

$$S'_j A'(x)F' = S_j A(x)F \qquad \forall x \ \varepsilon\ \Sigma^*$$

$$S'_i A'(x)F' = S_i A(x)F \qquad \forall x \ \varepsilon\ \Sigma^*$$

But $S_i \overset{A}{\sim} S_j \Longrightarrow S'_i \overset{A'}{\sim} S'_j$. Therefore we use Theorem 8.1 on A' (noting that $\overset{A'}{\sim}[S'_{2\#}]$ has the same number of members as $\overset{A}{\sim}[S_{2\#}]$). By substituting $k^\#$ and $(k+1)^\#$ for $1^\#$ and $2^\#$ above, an inductive argument is obtained. A machine with rank $|\overset{A}{\sim}|$ states is obtained at the termination of the induction. The particular machine obtained depends on the representatives selected from the classes of $\overset{A}{\sim}$ and is hence not unique.

However if A'' and A''' are two minimal state machines then A'' $\equiv(s)$ A''' since each is individually equivalent by $\equiv(s)$ to A.

If there were some machine B with fewer states than rank $|\overset{A}{\sim}|$ then for some i and all $x \in \Sigma^*$

$$S_{h(i)}B(x)F^B = S_iA(x)F^A$$

$$S_{h(i)}B(x)F^B = S_kA(x)F^A$$

for $S_i \overset{A}{\not\sim} S_k$, i.e., S_i and S_k in different classes under $\overset{A}{\sim}$. Therefore there must be a x' such that

$$S_iA(x')F^A \neq S_jA(x')F^A$$

which gives the contradiction

$$S_{h(i)}B(x')F^B \neq S_{h(i)}B(x')F^B$$

Q. E. D.

The previous two theorems are formalizations and generalizations of remarks made by Carlyle [1961] concerning $\equiv_I$. Let us turn attention now to $\equiv_T$ to see why it does not satisfy Theorem 8.1.

8.6 MINIMAL MACHINES FOR $\equiv_T$

We make a definition of $\sim_T$ analogous to 8.2 (a), (b) and (c).

Definition 8.6: Equivalence of states for $\equiv_T$.

$$S_i \overset{A}{\sim}_T S_j \text{ if } S_iA(z)F \geq \lambda \iff S_jA(z)F \geq \lambda$$

As in the definition of $\equiv_T$, the above depends on the cutpoint λ. And as before it will be assumed that λ is defined with A and remains fixed in a context. The relation $\overset{A}{\sim}_T$ is an equivalence relation on states. Furthermore $S_i \overset{A}{\sim}_T S_j \Longrightarrow S_iA(x) \overset{A}{\sim}_T S_jA(x)$. However while multiples of equalities can be added, expressions involving $\geq$ and $\leq$ cannot be manipulated in this way. This elementary fact leads to the conclusion that neither equivalence in Theorem 8.1 holds.

Remark 8.3: There is a machine A such that for $S_i \overset{A}{\sim}_T S_j$:

(a) $A[S_i(S_j)] \equiv_T A$ while $A \not\equiv_T A[S_j(S_i)]$

Furthermore, there are classes of machines such that:

(b) $S_i \overset{A}{\sim}_T S_j \not\Longrightarrow A[S_i(S_j)] \equiv_T A$.

Demonstration (of the remark)

Suppose Definition 8.6 is satisfied. It is not necessary for $F_i = F_j$, i.e.,

$$F_i \ge \lambda \Longleftrightarrow F_j \ge \lambda$$

So specify that $F_i \ne F_j$. In particular let

$A = \langle 3, (.1, .3, .6), \Sigma, A(\Sigma), F = \begin{pmatrix}1\\2.1\\4\end{pmatrix}, 0_A\rangle$ and $\lambda = 2$. Let

$\Sigma = \{\Lambda, 1\}$ and $A(1) = \begin{pmatrix}0 & 0 & 1\\0 & 0 & 1\\0 & 0 & 1\end{pmatrix}$

$$IA(\Lambda)F = (.1, .3, .6)\begin{pmatrix}1\\2.1\\4\end{pmatrix} = 3.13 \text{ so } \Lambda \in T(A,\lambda)$$

For all $z \in \Sigma^*$: $S_2A(z)F \ge 2$ $\quad S_3A(z)F > 2$ so $S_2 \overset{A}{\sim}_T S_3$.

$$A' = A[S_2(S_3)] = \langle 2, (.1, .9), \Sigma, A'(1), \begin{pmatrix}1\\2.1\end{pmatrix}, 0_A\rangle$$

$I'A'(\Lambda)F' = I'F' = (.1, .9)\begin{pmatrix}1\\2.1\end{pmatrix} = 1.99$ so $\Lambda \notin T(A',\lambda)$ Hence $A[S_2(S_3)] \not\equiv_T A$

On the other hand

$$A'' = A[S_3(S_2)] = \langle 2, (.1, .9), \Sigma, A''(1), \begin{pmatrix}1\\4\end{pmatrix}, 0_{A''}\rangle$$

$$I''A''(\Lambda)F'' = (.1, .9)\begin{pmatrix}1\\4\end{pmatrix} = 3.7 \text{ so } \Lambda \in T(A'',\lambda)$$

For all $z'' \in \Sigma, I''A''(z'')F'' > 2$ so $T(A,2) = T(A'',2)$, i.e., $A[S_3(S_2)] \equiv_T A$.

So even though $S_3 \overset{A}{\sim}_T S_2$

$$A[S_3(S_2)] \not\equiv_T A[S_2(S_3)]$$

Of course much more elaborate machines can be constructed to show that this property is not dependent only on the identity but occurs for a large class of probabilistic machines.

Following in the same reasoning, a few other observations are immediate.

8.7 THE CRITICAL SET OF TAPES

Definition 8.7: The set of all tapes x such that $IA(x)F = \lambda$ will be called the critical tapes for A, i.e.,

$$C(A,\lambda) = \{x_c : IA(x_c)F = \lambda\}$$

Theorem 8.3

If $A[S_i(S_j)] \equiv_T A$ and $A[S_j(S_i)] \equiv_T A$ while $F_j > F_i$ and if the cutpoint λ is not isolated for A, i.e., $C(A,\lambda) \neq \phi$ for all $x_c \in C(A,\lambda)$: $IA(x_c)_j = 0$

Proof: By assumption $F_j > F_i$, then $E_{A[S_j(S_i)]}(x) \geq E_A(x)F \quad \forall x \in \Sigma^*$ since the probabilities of S_i will be shifted to state S_j whose output weight is greater. Equality can occur only when $IA(x)_i = 0$. Likewise

$$E_A[x] \geq E_{A[S_i(S_j)]}(x) \qquad \forall x \in \Sigma^*$$

with equality only when $IA(x)_j = 0$. Suppose the cutpoint is not isolated. Then for $x_c \in C(A,\lambda)$ we have

$$E_{A[S_j(S_i)]}(x_c) \geq \lambda \geq E_{A[S_i(S_j)]}(x_c)$$

Suppose $IA(x)_j \neq 0$; then $x_c \in T(A[S_j(S_i)],\lambda)$ and $x_c \notin T(A[S_i(S_j)],\lambda)$. Hence $A[S_i(S_j)] \not\equiv_T A[S_j(S_i)]$ contrary to the assumption.

Q. E. D.

8.8 GEOMETRIC INTERPRETATION OF S^+

The set of distributions over the states of an n-state machine, S^+, is isomorphic to a set in n-1 dimensional space because only n-1 of the probabilities can be specified independently. For four state machines S^+ can be depicted using a tetrahedron which has unit altitudes. Each vertex of the tetrahedron represents a state. Label the vertices S_1, S_2, S_3, S_4 and let a_i be the altitude from S_i; $i = 1, \ldots, 4$. The distribution $v = (p_1, p_2, p_3, p_4)$ is represented by coordinates on the altitudes measured from the base. Figure 8.1 shows this convention for the first component.

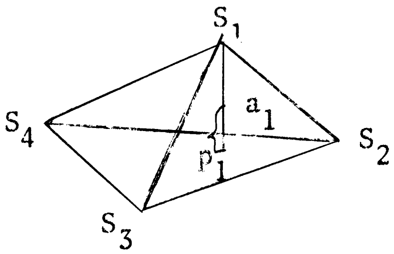


Figure 8.1. A coordinate of a distribution in the tetrahedron isomorphic to S^+ for n=4.

8.9 THE CRITICAL DISTRIBUTIONS OF S^+

Let us generalize the notion of critical tapes (Definition 8.7) to distributions.

Definition 8.8: A distribution vector π is critical for machine A if $\pi F = \lambda$. The set of critical distributions $D(A,\lambda) = \{\pi \in S^+ : \pi F = \lambda\}$.

From the definitions it is clear that

$$\{I \cdot A(x_c) : x_c \in C(A,\lambda)\} \subset D(A,\lambda)$$

Definition 8.9: A distribution π is said to be accepted if $\pi F \geq \lambda$. A distribution π is rejected if $\pi F < \lambda$.

The set of critical distributions $D(A,\lambda)$ is a plane in S^+ which separates the accepted distributions from the rejected distributions. The following example illustrates these notions.

Example 8.1

Let $F = \begin{pmatrix} 4 \\ 1/2 \\ 2 \\ 1 \end{pmatrix}$ and $\lambda = 2$.

The solution set of the equation $\pi F = 2$ where π is stochastic is a convex set determined by its extreme points. Using the well known method of Kemeny et.al [1959], we find the extreme points

$$\pi(1) = (3/7, 4/7, 0, 0)$$

$$\pi(2) = (0, 0, 1, 0) = S_3$$

$$\pi(3) = (1/3, 0, 0, 2/3)$$

The plane $D(A,\lambda)$ is shown in Figure 8.2.

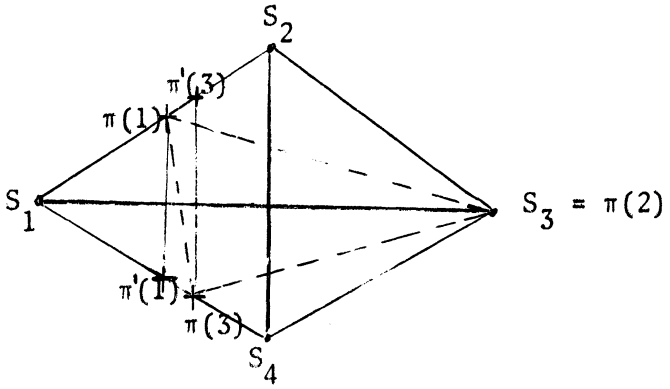


Figure 8.2. $D(A,\lambda)$ is the area enclosed by the dotted lines, i.e. the triangle $\pi(1)$-$\pi(2)$-$\pi(3)$.

8.10 GEOMETRIC INTERPRETATION OF STATE MERGING

All the distributions on the "S_1 side" of $D(A,\lambda)$ in Figure 8.2 are accepted, while all distributions on the "S_2-S_4 side" are rejected. Clearly, identifying S_1 with either S_2 or S_4 would change the acceptance of the empty string (and probably many other strings) into rejection and hence for any machine A with output vector $F = \begin{pmatrix} 4 \\ 1/2 \\ 2 \\ 1 \end{pmatrix}$

$$T(A,2) \neq T(A[S_4(S_1)],2)$$

On the other hand, suppose S_2 is identified with S_4. Figure 8.3 shows the resulting homomorphism of the tetrahedron.

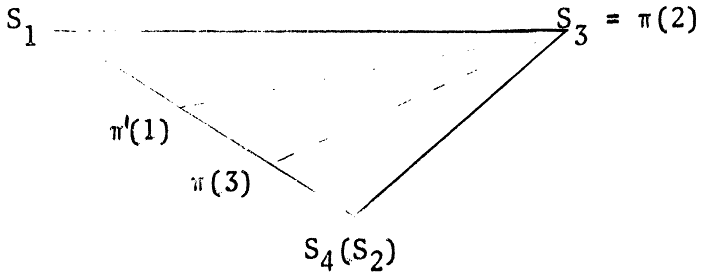


Figure 8.3. The set of distributions when S_2 is identified with S_4, i.e. $S_4(S_2)$.

From Figure 8.3 we observe that $\pi'(1)$ (the projection of $\pi(1)$ on S_1-S_4) lies inside the accepted region and is no longer an extreme point. Furthermore, all the distributions in the wedge-shaped volume (of Figure 8.2) bounded by $\pi(3)-\pi(1)-\pi'(3)-\pi(2)$ will be called the $S_4(S_2)$ wedge and written $W(S_4(S_2))$. While the distributions of $W(S_4(S_2))$ are rejected, their images after merging are accepted. Hence if any accessible distribution $I \circ A(x) \varepsilon V(A)$ for any machine A with output F and cutpoint 2 falls in $W(S_4(S_2))$ we have

$$T(A,2) \neq T(A[S_4(S_2)],2)$$

Likewise we find that if $I \circ A(x) \varepsilon V(A)$ and $I \circ A(x) \varepsilon W(S_2(S_4))$ then

$$T(A,2) \neq T(A[S_2(S_4)],2)$$

We summarize these remarks in the following theorem. The first part of which will not be proved since the substitution of general variables in the above remarks sketches a proof.

Theorem 8.4

Let A be a four state probabilistic sequential machine and let $S_i \underset{T}{\sim} S_j$:

$$A \equiv_T A[S_i(S_j)] \Leftrightarrow W(S_i(S_j)) \cap V(A) = \phi \text{ or } W(S_i(S_j)) = D(A,\lambda)$$

Proof: Sufficiency is sketched in the example above. For necessity we note that

$$W(S_i(S_j)) \cap V(A) = \phi \Longrightarrow V(A) \subset (S^+ - W(S_i(S_j)))$$

But any distribution in $S^+ - W(S_i(S_j))$ does not have its acceptance changed by the identification of S_j with S_i. Likewise if $W(S_i(S_j)) = D(A,\lambda)$, $D(A,\lambda)$ is parallel to S_i-S_j and merging S_i with S_j does not change the acceptance of any distribution. Hence if $x \varepsilon T(A,\lambda)$, then $x \varepsilon T(A[S_i(S_j)],\lambda)$ and vice versa.

Q. E. D.

Corollary 8.4

If $F_i = F_j$ for a four state machine A and $S_i \underset{T}{\sim} S_j$ then

$$A[S_i(S_j)] \equiv_T A \equiv_T A[S_j(S_i)]$$

Proof: For $F_i = F_j$ and four states the extreme points on the lines S_i-S_k and S_j-S_k (where S_k is adjacent to S_i and S_j) are permutations of each other, and the line joining them is hence parallel to S_i-S_j. Therefore, identifying S_i and S_j causes $D(A,\lambda)$ to be compressed without sweeping out any volume; i.e. $W(S_i(S_j)) = D(A,\lambda)$ and $W(S_j(S_i)) = D(A,\lambda)$. Theorem 8.4 hence gives the result.

Q. E. D.

The proofs of this section concerning the four state machine admittedly make use of many implicit facts of the geometry of three dimensions. It seems likely that analogous arguments (except for Corollary 8.4) hold in n-dimensions, but the author has no proofs for the general case.

8.11 COMPARISON OF STATE REDUCTION FOR $\equiv_I$, $\equiv_N$ AND $\equiv_E$ WITH $\equiv_T$

The last paragraph of Theorem 8.1 shows the essential difference between the former equivalences and $\equiv_T$. If $S_i \overset{A}{\underset{T}{\sim}} S_j$, $S_i \overset{A}{\underset{N}{\sim}} S_j$ or $S_i \overset{A}{\underset{E}{\sim}} S_j$ then simple multiplication by constants and addition of equalities shows that for any initial distribution I and distribution I' (which is I with S_i and S_j merged) we have

$$A \equiv (S)A[S_i(S_j)] \Longrightarrow [\langle I,A\rangle \equiv \langle I', A[S_i(S_j)]\rangle]$$

However for $\equiv_T(S)$, a system of inequalities is obtained which may not be able to be added and $S_i \underset{T}{\sim} S_j$ is not enough to guarantee this property.

$$A \equiv_T (S)A[S_i(S_j)] \not\Longrightarrow [\langle I,A\rangle \equiv_T \langle I', A[S_i(S_j)]\rangle]$$

The results of the previous sections concerning convex sets show that merging of equivalent states does not preserve tape equivalence. The status of the vectors $I \circ A(x)$: $x \in \Sigma^*$ in the set S^+ relative to $D(A, \lambda)$ must be considered.

CHAPTER 9

APPLICATIONS TO OPTIMAL CONTROL PROBLEMS

9.0 INTRODUCTION

The close connection between the theory of sequential machines and the theory of optimal control has been pointed out by various authors. Arbib [1964] has attempted to make a "rapprochement" between the two areas of study. The purpose of this chapter is to show how some of the notions of previous chapters apply to the optimal control problem as defined by Eaton and Zadeh [1962]. Before applying some of the previous theorems to certain optimal control problems, let us consider some basic observable events of a probabilistic system.

9.1 INPUT EVENTS

The most fundamental conditional probability of a machine is the probability which relates sequences of outputs to sequences of inputs. As has been observed in Chapter 3, such a joint probability is defined recursively from the joint probability of output symbol y and states S_j given input symbol σ and initial state S_i. The "Moore model" which has been used throughout requires y to be a function only of S_j. Hence this basic probability can be expressed as (using the notation of Chapter 3):

$$P(y,S_j/\sigma,S_i) = A(y/\sigma)_{ij} = \begin{cases} A(\sigma)_{ij} \text{ if } 0(S_j) = y \\ 0 \text{ otherwise} \end{cases}$$

From the above conditional probabilities the system probabilities or stochastic input-output relation can be computed. As in previous chapters, the probability of observing the output sequence $y_1 \ldots y_r$ if the machine A is started in initial distribution π and the input sequence $\sigma_1 \ldots \sigma_r$ is applied will be designated $P_{\pi}^{A}(y_1 \ldots y_r/\sigma_1 \ldots \sigma_r)$.

Control engineers frequently need to use all the information available in order to control a process. The following concepts, systematized by Carlyle [1965], formalize other conditional probabilities which can be determined by experiments on a probabilistic system.

Definition 9.1: A *y-input event* for a machine A, written $T_y(A,\lambda)$, is the set of input strings which produce $y \in Y^*$ as the last sequence of outputs with probability greater than or equal to λ.

$$T_y(A,\lambda) = \{x \in \Sigma^* : \sum_{y_p \in (Y)^{\ell g(x)-\ell g(y)}} I' A(y_p y/x) \cdot \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \geq \lambda\}$$

We note the connection between y-input events and the concept of tape acceptance:

$$T_1(A,\lambda) = T(A,\lambda)$$

$$T_0(A,\lambda) = \Sigma^* - T(A,\lambda)$$

9.2 INPUT EVENT EQUIVALENCE: $\equiv_{IE}$

Definition 9.2: Machines A and A' are *λ-input event equivalent*, written $A \equiv_{\lambda IE} A'$, if for a *fixed* real number λ and all y in the output *alphabet* Y:

$$T_y(A,\lambda) = T_y(A',\lambda)$$

Following Carlyle [1965] and Paz [1964] (who calls it "strong equivalence"), let us define a very strong equivalence for input events.

Definition 9.3: Machines A and A' are *input event equivalent*, written $A \equiv_{IE} A'$, if for *all* λ and all y in the output alphabet Y:

$$T_y(A,\lambda) = T_y(A',\lambda)$$

Note that $\equiv_{IE}$ is just the universal quantification over λ of $\equiv_{\lambda IE}$.

It has been shown (Carlyle [1965], Theorem 7) that

$$A \equiv_I A' \Rightarrow A \equiv_{IE} A'$$

Let us relate $\equiv_{IE}$ to the other previously defined machine equivalences.

Theorem 9.1

$$A \equiv_D A' \Rightarrow A \equiv_{IE} A' \Rightarrow A \equiv_N A'$$

Proof:

$A \equiv_{IE} A' \Rightarrow \text{Prob.}\{0^*_A(x) = y\} = \text{Prob.}\{0^*_{A'}(x) = y\} \quad \forall y \in Y, \forall x \in \Sigma^*$

By definition

$$\text{Prob.}\{0^*_A(x) = y\} = \sum_{\substack{i: \\ 0(S_i)=y}} I\cdot A(x)_i$$

Distribution equivalence implies

$$\sum_{\substack{i: \\ 0(S_i)=y}} I\cdot A(x)_i = \sum_{\substack{i': \\ 0'(S_{i'})=y}} I'\cdot A'(x)_{i'}, \quad \text{for } y \neq 0 \text{ and not with probability zero of occurring.}$$

Hence taking sums over those $y \neq 0$

$$\sum_{\substack{i: \\ 0(S_i)=y\neq 0}} I\cdot A(x)_i = \sum_{\substack{i': \\ 0'(S_{i'})=y\neq 0}} I'\cdot A'(x)_{i'} \Rightarrow \sum_{\substack{i: \\ 0(S_i)=0}} I\cdot A(x)_i = \sum_{\substack{i': \\ 0'(S_{i'})=0}} I'\cdot A'(x)_{i'}$$

which gives the first part.

For the second implication, regrouping of the terms of Theorem 2.1 shows that $\mu^A_N(x)$ depends only on $\sum_{\substack{i: \\ 0(S_i)=y}} I\cdot A(x)_i$ and (F^k), $k = 1, \ldots, N$.

But from the above and noting that $F = F'$, we see that the arguments of the function $\mu^A_N(x)$ are all equal for machines A and A'.

Q. E. D.

Immediately from the definition we obtain:

$$A \equiv_{IE} A' \Rightarrow A \equiv_{\lambda IE} A'$$

In general, there is no relationship between $\equiv_{\lambda IE}$ and $\equiv_T$. However, for the special case of Rabin automata, the inequalities involved are degenerate so that: $A \equiv_{\lambda IE} A' \Longleftrightarrow A \equiv_T A'$.

9.3 INPUT-OUTPUT EVENTS

We now consider events which require information about the outcome of initial portions of the experiment. Generalizing the definition of Carlyle [1965], we let $q(y/ux,v)$ be the probability of observing output string y as a consequent of input string x, having already observed v as output because of the input string u. The fundamental definition of conditional probabilities gives us that:

$$q(y/ux,v) = \frac{I'A(vy/ux)\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}}{I'A(v/u)\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}}$$

<u>Definition 9.4</u>: The <u>input-output event</u> for $y \in Y^*$ of machine A and cutpoint λ, written $IO_y(A,\lambda)$ is given by:

$$IO_y(A,\lambda) = \{(ux,v) : q(y/ux,v) \geq \lambda\}$$

Here we let x and y be arbitrary strings while Carlyle restricts them to symbols. Let us interpret one kind of input-output event for a computer. Suppose the input sequences are programs, the output sequences are results and λ is a reliability. The input-output event for a particular result y consists of all programs ux and "intermediate results" v which guarantee that y will occur as an output sequence with reliability at least as high as λ.

9.4 INPUT-OUTPUT EVENT EQUIVALENCE: $\equiv_{IO}$

Analogous with the concept of tape equivalence, a fixed cutpoint equivalence will be defined for input-output events.

Definition 9.5: Machines A and A' are λ-input output event equivalent, written $A \equiv_{\lambda IO} A'$, if:

$$IO_y(A,\lambda) = IO_y(A',\lambda) \quad \text{for all } y \in Y$$

By universally quantifying the cutpoint of $\equiv_{\lambda IO}$, the definition of Carlyle [1965] is obtained.

Definition 9.6: Machines A and A' are input-output event equivalent, written $A \equiv_{IO} A'$, if:

$$A \equiv_{IO} A' \quad \text{for all real } \lambda$$

The equivalence $\equiv_{IO}$ amounts to the same thing as $\equiv_I$; i.e. as has been noted by Carlyle, the definitions yield:

Theorem 9.2

$$A \equiv_I A' \Leftrightarrow A \equiv_{IO} A'$$

Let us show the relationships among the machine equivalences defined thus far.

$$\begin{tikzcd}
A \equiv_D A' \arrow[dr, Rightarrow] & & & & \\
\boxed{\begin{array}{c} A \equiv_I A' \\ \text{Shannon} \end{array}} \arrow[r, Rightarrow] \arrow[dd, Leftrightarrow] & \boxed{\begin{array}{c} A \equiv_{IE} A' \\ \text{Paz} \end{array}} \arrow[r, Rightarrow] \arrow[dr, Rightarrow] & A \equiv_N A' \arrow[r, Rightarrow] & A \equiv_E A' \arrow[r, Rightarrow] & \boxed{\begin{array}{c} A \equiv_T A' \\ \text{Rabin} \end{array}} \\
 & & A \equiv_{\lambda IE} A' & & \\
\boxed{\begin{array}{c} A \equiv_{IO} A' \\ \text{Carlyle} \end{array}} \arrow[r, Rightarrow] & A \equiv_{\lambda IO} A' & & &
\end{tikzcd}$$

Figure 9.1. Relationships among the machine equivalences.

9.5 OPTIMAL CONTROL PROBLEMS

An optimal control problem is essentially a probabilistic sequential machine with costs associated with the action of inputs. A certain state, or set of states (usually absorbing) is designated as the "target" and the problem involves getting the state activity to the target with minimal costs. Usually information obtained from the outcome of an input or sequence of inputs is used to evaluate some function which specifies the next input or sequence of inputs. Following Eaton and Zadeh [1962], we shall assume that after each input the resulting state is known. The kinds of next input functions or policies which will be considered here usually depend only on the state. The outputs of the probabilistic machine will not be used as outputs in the usual sense but will be used to express the costs of the actions on the control system. Hence the expectation of "output" $E_A(x)$ will actually be the expected cost at the time step after the input of the string x into control system A. It is helpful to think of the output vector F as being expanded into a n x 2 matrix, the first column having arbitrary but unequal components which are used to identify the state. The second column functions as the output in the sense of the formal definition but contains the costs associated with the states. A minor difficulty occurs because many optimal control problems have costs associated with the state transitions rather than states. However, we shall see in Lemma 9.1 that an expanded machine can be built for any such control problem with the costs associated only with the states. Lemma 9.1 also translates between the terminology of probabilistic machines and that of the previously mentioned paper of Eaton and Zadeh.

Lemma 9.1 Translation between probabilistic sequential machines and optimal control problems:

The following makes a correspondence between the terminology and assumptions of Eaton and Zadeh [1962] in "Optimal Pursuit Strategies in Discrete State Probabilistic Systems," and those of probabilistic sequential machines.

Eaton & Zadeh	First level translation: Machine A	Second level translation: expanded Machine B
(1) Transitions take place at integral values of t starting at $t = 0$	Same	Same
(2) Input u_t takes on values $\{\alpha_1,\ldots,\alpha_m\}$	$\Sigma = \{\sigma_1,\sigma_2,\ldots,\sigma_m\}$ sometimes abbreviated $1,2,\ldots,m$	$\Sigma = \{\sigma_1,\sigma_2,\ldots,\sigma_m\}$
(3) Inputs referred to as "commands" or "actions"	Inputs called "input symbols"	Inputs called "input symbols"
(4) States $q_1,\ldots,q_n,q_{n+1}$	States $S = \{s_1,\ldots,s_n,s_{n+1}\}$	States (s_i,s_j,k) where $s_i,s_j \varepsilon S$ and $k \varepsilon \Sigma$
(5) The state q_{n+1} is the <u>target</u> state	s_{n+1} is a sink or absorbing state	If the state (s_{n+1},s_j,k) is accessible then $s_j = s_{n+1}$
(6) $P_{ij}(k)$ prob. of one step transition from q_i to q_j when command α_k is applied	$A(k)_{i,j}$	$B(k)_{(m,i,u),(\ell,j,v)}$ $= A(k)_{ij}$ where $v = k$ and $i = \ell$ $= 0$ otherwise
(7) $P_{ij}(k)$ is stochastic	$A(k)_{i,j}$ is stochastic	all $\sum_{(\ell,j,v)} B(k)_{(m,i,u),(\ell,j,v)}$ $= \sum_j A(k)_{i,j} = 1$ so $B(k)$ is stochastic
(8) Associated with each non-zero transition $s_i \rightarrow s_j$ by command α_k is a positive cost $C_{ij}(k)$	$C(k)$ is the cost matrix for $A(k)$ (no way to express this as a single output vector in machine A, but can be expressed as the output vectors of a set of machines having the same transitions as A).	$F^B_{(i,j,k)} = C_{ij}(k)$ i.e., the output vector of B contains all costs.

(9) The cost of going to the target is 0	$C(k)_{i,n+1} = 0$	$F^B_{(i,n+1,k)} = 0$
(10) By an ℓ-step transition sequence is a sequence of states $(\alpha_0,\alpha_1,\alpha_{\ell-1};s_0,\ldots,s_\ell)$	Same as E. & Z.	A sequence of ℓ-states $b_0 = (s_0,s_1,\alpha_0)$ $b_1 = (s_1,s_2,\alpha_1),\ldots$ $b_{-1} = (s_{\ell-1},s_\ell,\sigma_{\ell-1})$
(11) A transition sequence is possible if the conditional prob. of observing the sequence given the initial state q_0 is positive	... if $F^A_i \neq F^A_j$ for $i \neq j$ $p^A_{s_0}(F_0\ldots F_\ell/\sigma_0\ldots$ $\sigma_{\ell-1}) > 0$	... if $B(\sigma_0)_{b_0,b_1} \cdot B(\sigma_1)_{b_1,b_2} \cdot$ $\ldots \cdot B(\sigma_{\ell-1})_{b_{\ell-1},b_\ell} > 0$
(12) Cost of the sequence of length ℓ is the sum of the one step costs. $C = \sum_{r=0}^{\ell-1} c_{i_r j_{r+1}}(k_r)$	Same as E. & Z. but outside normal P.S.M. framework.	$C = \sum_{r=0}^{\ell-1} F_{(s_r,s_{r+1},\sigma_r)}$
(13) A strategy or a policy is a function π which assigns a command to every state $\mu_t = \pi(q_t)$ write $\pi = (\pi_1,\ldots,\pi_n)$	Same as E. & Z.	π' is a projective function defined by $\pi'(s_{t-1},s_t,u_{t-1})$ $= \pi(s_t)$ write $\pi' = (\sigma_{\pi_1},\ldots,\sigma_{\pi_n})$
(14) The policy matrix $P_{ij}(\pi_i)$ where π_i is the optimal control for state q_i	$A(\pi) = \begin{pmatrix} s_1A(\pi_1) \\ \vdots \\ s_{n+1}A(\pi_{n+1}) \end{pmatrix}$	$B(\pi)_{(s_\ell,s_i,\sigma),(s_i,s_j,\sigma_{\pi_i})}$ $= A(\pi)_{ij}$ $= 0$ otherwise
(15) The cost matrix for policy π $C(\pi) = C_{ij}(\pi_i)$	$C(\pi) = \begin{pmatrix} s_1C(\pi_1) \\ \vdots \\ s_{n+1}C(\pi_{n+1}) \end{pmatrix}$	Costs are embedded in the output vector F^B $C(\pi)_{i,j} = F^B_{(s_i,s_j,\sigma_{\pi_i})}$
(16) $X_i(\pi)$ is the expected cost of policy from initial state q_i	Same as E. & Z.	Analogous to E. & Z. i.e., $X'_{(m,i,\sigma_{\pi_i})}(\pi)$
(17) A policy π is proper if $x_i(\pi) < \infty$ $i=1,2,\ldots,r$	Same as E. & Z.	$X'_{(m,i,\sigma_{\pi_i})}(\pi) < \infty$ $i=1,2,\ldots,n$

Lemma 9.2

$$X'(\pi) = \sum_{i=1}^{\infty} B(\pi)^i F^B \text{ if } \pi \text{ is a proper policy.}$$

Proof: The basic recursive equation of Eaton and Zadeh will be formulated using the expanded machine B

$$X_i(\pi) = E\{C_{ij}(\pi_i) + X_j(\pi)\}$$

$$= \sum_{j=1}^{n+1} P_{ij}(\pi_i)[c_{ij}(\pi_i) + X_j(\pi)]$$

i.e.

$$X(\pi) = C(\pi) + P(\pi) \cdot X(\pi) \qquad (1)$$

where the one step average cost from state i is:

$$C(\pi)_i = \sum_{j=1}^{n} c_{ij}(\pi_i)\hat{P}_{ij}(\pi_i) \text{ where } \hat{P}_{ij}(\pi_i) = ||P_{ij}(\pi_i) \; i,j = 1, 2, \ldots, n||$$

If the target can be reached from any state with positive probability by the policy then $\hat{P}(\pi)$ is not stochastic, i.e., the matrix $\hat{P}(\pi)$ has all eigenvalues less than unity in modulus and (1) can be solved as

$$X(\pi) = (E - \hat{P}(\pi))^{-1}C(\pi)$$

equivalently

$$X(\pi) = \sum_{i=0}^{\infty} P(\pi)^i C(\pi)$$

Equation (1) can be expressed using machine B and definition (16) of Lemma 9.1. $X'(\pi) = B(\pi)F^B + B(\pi)X'(\pi)$ which gives, if a unique solution exists (i.e., π is proper):

$$X'(\pi) = \sum_{i=1}^{\infty} B(\pi)^i \; F^B \qquad (2)$$

which is a sum of expectations of B.

Q. E. D.

Lemma 9.3

$$X(\pi) = X(\pi') \Longleftrightarrow X'(\pi) = X'(\pi')$$

Proof: From (16) of Lemma 9.1 we see that the elements of $X(\pi)$ are rewritten m∘n times in $X'(\pi)$.

9.6 EQUIVALENCE OF POLICIES

Theorem 9.1

Let π and π' be policies of an optimal control problem. Let $A = \langle n+1, \ , S, \Sigma = \{\pi,\pi'\}, A(\pi), A(\pi'), F, 0\rangle$ be a probabilistic sequential machine having input symbols π and π' and switching matrices $P(\pi) = A(\pi)$ and $P(\pi') = A(\pi')$.

Let the cost matrix C be independent of the input. Furthermore let the column space of C be generated by powers of the components of F:

$$(C^T)_i \varepsilon \langle\{(F), (F^2), \ldots, (F^N)\}\rangle$$

Then for $R_N(A)$, the N-moment reduction relation defined by A

$$\pi R_N(A)\pi' \Longrightarrow X(\pi) = X(\pi')$$

Proof: There must be a set of constants $\{d_k^i\}$ so

$$(C^T)_i = \sum_{k=1}^{N} d_k^i(F^k)$$

$$X(\pi)_i = \sum_{j=1}^{\infty} s_i A(\pi)^j \cdot (C^T)_i$$

$$= \sum_{j=1}^{\infty} s_i A(\pi)^j \cdot \sum_{k=1}^{N} d_k^i(F^k)$$

$$= \sum_{j=1}^{\infty} \sum_{k=1}^{N} d_k^i \cdot s_i \cdot A(\pi)^j (F^k) \qquad (*)$$

since $\pi R_N \pi'$ and $A(\pi)$ and $A(\pi')$ are assumed to be distinct; Theorem 1.8D gives:

$$A(\pi) = A(\pi') + H$$

where $H_i \in V \quad i = 1, 2, \ldots, n$ and some $H_i \neq 0$

while

$$VA(\sigma) \subset V \qquad \sigma \in \{\pi, \pi'\}$$

$$V \subset \bigcap_{i=1}^{N} \text{Kern.}(F^i)$$

Hence

$$\begin{aligned} A(\pi)^j &= A(\pi')^j + HA(\pi')^{j-1} + \ldots + H^j \\ &= A(\pi')^j + H' + \ldots + H'' \end{aligned}$$

where H_i' and H_i'' are in V for $i = 1, 2, \ldots, n$

multiplying by $(F^k) \quad 1 \leq k \leq N$

$$\begin{aligned} A(\pi)^j(F^k) &= A(\pi')^j(F^k) + H'(F^k) + \ldots + H''(F^k) \\ &= A(\pi')^j(F^k) + 0 + \ldots + 0 \end{aligned}$$

Substitution in (*) gives for $i = 1, 2, \ldots, n$

$$X(\pi)_i = \sum_{j=1}^{\infty} \sum_{k=1}^{N} d_k^i s_i \cdot A(\pi')^j (F^{(k)}) = X(\pi')_i$$

Q. E. D.

Example 9.1

We illustrate Theorem 9.1 with the following optimal control problem. Using the notation of Eaton and Zadeh:

Let

$Q = \{q_1, q_2, q_3\}$ where q_3 is the target state

$\alpha = \{0, 1\}$ input alphabet.

$$A(0) = \begin{pmatrix} 0 & 3/8 & 5/8 \\ 0 & 3/4 & 1/4 \\ 0 & 0 & 1 \end{pmatrix}$$

$$A(1) = \begin{pmatrix} 1/4 & 1/4 & 1/2 \\ 1/2 & 1/2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

The cost matrix C is independent of input:

$$C(1) = C(0) = C = \begin{pmatrix} 1 & 2 & 0 \\ 3 & 6 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Let

$$F = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} \quad \text{and } N = 1$$

<u>Claim</u>: The policies of $\pi = (1, 0)$ and $\pi' = (0, 1)$ are equivalent, i.e., $X(\pi) = X(\pi')$.

Note that π and π' are proper:

$$A(\pi) = \begin{pmatrix} 1/4 & 1/4 & 1/2 \\ 0 & 3/4 & 1/4 \\ 0 & 0 & 1 \end{pmatrix}$$

$$A(\pi') = \begin{pmatrix} 0 & 3/8 & 5/8 \\ 1/2 & 1/2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

For $N = 1$ $\quad R_N(A) = R_E(A)$

From Theorem 4.4:

$$\pi R_E \pi' \Longleftrightarrow A(\pi z)F = A(\pi' z)F \qquad \forall z \in \{\pi, \pi'\}^* \text{ such that } \ell g.(z) \leq 1$$

Hence

$$A(\pi)F = A(\pi')F = \begin{pmatrix} 3/4 \\ 3/2 \\ 0 \end{pmatrix}$$

Likewise the reader may verify that

$$A(\pi\pi')F = A(\pi'\pi')F$$

$$A(\pi\pi) = A(\pi'\pi)F$$

which shows that $\pi R_E \pi'$.

By Theorem 9.1 we know

$$X(\pi) = X(\pi')$$

However, in the next example we will construct the expanded machine B from A in order to make this fact more transparent.

Example 9.2 Construction of the expanded machine B from machine A.

In writing down the matrices the following ordering is used for the states of B; i.e., the r'th row or column of the matrix corresponds to the r'th state in the ordering:

$S_{110}, S_{111}, S_{120}, S_{121}, S_{130}, S_{131}, S_{210}, S_{211}, S_{220}, S_{221}, S_{230}, S_{231}, S_{310}, S_{311}, S_{320}, S_{321}, S_{330}, S_{331}$

Hence

$$F^B = \begin{pmatrix} 1 \\ 1 \\ 2 \\ 2 \\ 0 \\ 0 \\ 3 \\ 3 \\ 6 \\ 6 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

while $B(\pi)$ =

$$\begin{array}{c|cccccccccccccccccc}
 & 110 & 111 & 120 & 121 & 130 & 131 & 210 & 211 & 220 & 221 & 230 & 231 & 210 & 311 & 320 & 321 & 330 & 331 \\
\hline
110 & & 1/4 & & 1/4 & & 1/2 & & & & & & & & & & & & \\
111 & & 1/4 & & 1/4 & & 1/2 & & & & & & & & & & & & \\
120 & & & & & & & 0 & & 3/4 & & 1/4 & & & & & & & \\
121 & & & & & & & 0 & & 3/4 & & 1/4 & & & & & & & \\
130 & & & & & & & & & & & & & & & & & 1 & \\
131 & & & & & & & & & & & & & & & & & 1 & \\
210 & & 1/4 & & 1/4 & & 1/2 & & & & & & & & & & & & \\
211 & & 1/4 & & 1/4 & & 1/2 & & & & & & & & & & & & \\
220 & & & & & & & 0 & & 3/4 & & 1/4 & & & & & & & \\
221 & & & & & & & 0 & & 3/4 & & 1/4 & & & & & & & \\
230 & & & & & & & & & & & & & & & & & 1 & \\
231 & & & & & & & & & & & & & & & & & 1 & \\
310 & & 1/4 & & 1/4 & & 1/2 & & & & & & & & & & & & \\
311 & & 1/4 & & 1/4 & & 1/2 & & & & & & & & & & & & \\
320 & & & & & & & 0 & & 3/4 & & 1/4 & & & & & & & \\
321 & & & & & & & 0 & & 3/4 & & 1/4 & & & & & & & \\
330 & & & & & & & & & & & & & & & & & 1 & \\
331 & & & & & & & & & & & & & & & & & 1 &
\end{array}$$

and $B(\pi') =$

$$\left(\begin{array}{c|cccccccccccccccccc}
 & 110 & 111 & 120 & 121 & 130 & 131 & 210 & 211 & 220 & 221 & 230 & 231 & 310 & 311 & 320 & 321 & 330 & 331 \\
110 & 0 & & 3/8 & & 5/8 & & & & & & & & & & & & & \\
111 & 0 & & 3/8 & & 5/8 & & & & & & & & & & & & & \\
120 & & & & & & & & 1/2 & & 1/2 & & 0 & & & & & & \\
121 & & & & & & & & 1/2 & & 1/2 & & 0 & & & & & & \\
130 & & & & & & & & & & & & & & & & & 1 & \\
131 & & & & & & & & & & & & & & & & & 1 & \\
210 & 0 & & 3/8 & & 5/8 & & & & & & & & & & & & & \\
211 & 0 & & 3/8 & & 5/8 & & & & & & & & & & & & & \\
220 & & & & & & & & 1/2 & & 1/2 & & 0 & & & & & & \\
221 & & & & & & & & 1/2 & & 1/2 & & 0 & & & & & & \\
230 & & & & & & & & & & & & & & & & & 1 & \\
231 & & & & & & & & & & & & & & & & & 1 & \\
310 & 0 & & 3/8 & & 5/8 & & & & & & & & & & & & & \\
311 & 0 & & 3/8 & & 5/8 & & & & & & & & & & & & & \\
320 & & & & & & & & 1/2 & & 1/2 & & 0 & & & & & & \\
321 & & & & & & & & 1/2 & & 1/2 & & 0 & & & & & & \\
330 & & & & & & & & & & & & & & & & & 1 & \\
331 & & & & & & & & & & & & & & & & & 1 &
\end{array}\right)$$

We note that

$$B(\pi')F^B = B(\pi)F^B = \begin{pmatrix} 3/4 \\ 3/4 \\ 9/2 \\ 9/2 \\ 0 \\ 0 \\ 3/4 \\ 3/4 \\ 9/2 \\ 9/2 \\ 0 \\ 0 \\ 3/4 \\ 3/4 \\ 9/2 \\ 9/2 \\ 0 \\ 0 \end{pmatrix}$$

which means

$$B(\pi') - B(\pi) = H \qquad (1)$$

where $H_i F^B = 0 \qquad i = 1, 2, \ldots, n \qquad (2)$

Call

$$\langle\{H_i, \quad i = 1, 2, \ldots, 18\}\rangle = V_B$$

It follows that

$$V_B = \langle\{1/8(0,2,-3,2,-5,4,0,0,0,0,0,0,0,0,0,0,0,0),\ 1/4(0,0,0,0,0,0,0,-2,3/4,-1/2,3/4,-1/2,1/4,0,0,0,0,0,0)\}\rangle$$

It can be easily verified that the subspace V_B is invariant under both $B(\pi)$ and $B(\pi')$ i.e.,

$$V_B \cdot B(\pi) \subset V_B \qquad (3)$$

$$V_B \cdot B(\pi') \subset V_B$$

From (2) we see that $V_B F^B = 0$ i.e., $V_B \subset \text{Kern.}(F^B)$

Hence for any r (using (2) and (3))

$$B(\pi')^r = (B(\pi) + H)^r = B(\pi)^r + H'$$

where $H_i' \subset \text{Kern.}(F^B)$ $\quad i = 1, 2, \ldots, n_B$

Consequently,

$$B(\pi')^r F^B = B(\pi)^r F^B$$

Summing over r and making use of the fact that π and π' are proper policies

$$X'(\pi') = \sum_{r=1}^{\infty} B(\pi')^r F^B = \sum_{r=1}^{\infty} B(\pi)^r F^B = X'(\pi)$$

which means by Lemma 9.3

$$X(\pi) = X(\pi')$$

It is clear from the example that the simple expression for the expected cost provided by machine B is obtained at the expense of the large size and high redundancy of the matrices of B. Each row of the policy matrix $A(\pi)$ is copied $(n+1)\cdot m$ times in the matrix $B(\pi)$ which has order $(n+1)^2 m$.

One can imagine very obscure situations in which this redundancy could be used for error correction. For instance, the formal construct of expanded machine B might be useful in studying certain high reliability biological systems for which the large number of states is not a prohibitive restriction.

Still considering those optimal control problems for which the cost matrix is independent of the input symbols we obtain a more general and easily testable version of Theorem 9.1.

Theorem 9.2

Let $C(\pi) = C(\pi')$ for proper policies π and π'.

If $A(\pi)_i - A(\pi')_i \varepsilon V \subset \text{Kern.}(C^T)_i$ and $VA(\pi') \subset V \quad i = 1, 2, \ldots, n$

then $X(\pi) = X(\pi')$

Proof: $A(\pi)_i = A(\pi')_i + H_i$ where $H_i \varepsilon V \qquad i = 1, 2, \ldots, n$

For any $j \quad A(\pi)^j = (A(\pi')+H)^j \qquad (*)$

But $H_i A(\pi') \varepsilon V \qquad i = 1, 2, \ldots, n$

Hence $A(\pi)^j = A(\pi')^j + H'$ where $H'_i \varepsilon V \qquad i = 1, 2, \ldots, n$

Multiplying by S_i and $(C^T)_i$

$$S_i A(\pi)^j (C^T)_i = S_i A(\pi')^j (C^T)_i + H'_i (C^T)_i$$

$$= S_i A(\pi')^j (C^T)_i \qquad i = 1, 2, \ldots, n$$

Summing over j we obtain

$$X(\pi) = X(\pi')$$

Q. E. D.

We note that Theorem 9.2 can not be generalized by the substitution of V_i for V in the statement of the theorem. Inasmuch as the terms of $(*)$ of the form $H_i \cdot H$ need not be in $\text{Kern.}(C^T)_i$, the result does not follow.

Theorem 9.3

Given an optimal control problem with n+1 states, actions $\alpha_1, \ldots, \alpha_m$ and cost matrices $C(\alpha_1), \ldots, C(\alpha_m)$.

Suppose the expected costs at times $t = 1, 2, \ldots, 2n$ for policies π and π' are equal, i.e.,

$$X(\pi,t) = X(\pi',t) \qquad t = 1, 2, \ldots, 2n$$

Then

$$X(\pi) = X(\pi')$$

Proof: We construct a family of probabilistic sequential machines and test them in pairs for expectation equivalence.

Consider the machines

$$A^i = \langle n+1, S_i, S, \{\pi\}, A(\pi), (C(\pi)_i)^T, 0_{A^i} \rangle$$

$$A'^i = \langle n+1, S_i, S, \{\pi'\}, A(\pi'), (C(\pi')_i)^T, 0'_{A^i} \rangle$$

$$i = 1, 2, \ldots, n+1$$

From the definitions it follows that

$$A^i \equiv_E A'^i \Longrightarrow X(\pi)_i = X(\pi')_i \qquad i = 1, 2, \ldots, n+1$$

From Theorem 4.3 we get that

$$[S_iA(\pi)^t \cdot (C(\pi)_i)^T = S_iA(\pi')^t(C(\pi')_i)^T \quad t = 1, 2, \ldots, 2(n+1)-2] \Rightarrow A^i \equiv_E A'^i$$

But by definition

$$S_iA(\pi)^t \cdot (C(\pi)_i)^T = X_i(\pi,t)$$

Hence by the transitivity of implication

$$\begin{bmatrix} X(\pi,t) = X(\pi',t) \\ t = 1, 2, \ldots, 2n \end{bmatrix} \Longrightarrow X(\pi) = X(\pi')$$

Q. E. D.

The assumption that costs are independent of input does not yield a lower bound than Theorem 9.3.

9.7 CHARACTERIZATION OF EQUIVALENT PROPER POLICIES

Theorem 9.4

Let π and π' be proper policies with reduced (target omitted) matrices $B(\pi)$ and $B(\pi')$

$$X(\pi) = X(\pi') \Longleftrightarrow D[E - B(\pi)]^{-1}F^B = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} \text{ where } D = B(\pi) - B(\pi')$$

Proof: By Lemma 9.3 we have

$$X(\pi) = X(\pi') \Longleftrightarrow X'(\pi) = X'(\pi')$$

By Lemma 9.2

$$X'(\pi) = X'(\pi') \Longleftrightarrow [E - B(\pi)]^{-1}B(\pi)F^B = [E - B(\pi')]^{-1}B(\pi')F^B$$

$X(\pi) = X(\pi')$ if and only if

$$[E - B(\pi)]^{-1}B(\pi) = [E - B(\pi')]^{-1}B(\pi') + H$$

where $H_i \varepsilon$ Kern. (F^B). But since π is proper

$$[E - B(\pi)]^{-1}B(\pi) = B(\pi)[E - B(\pi)]^{-1}$$

$$B(\pi)[E - B(\pi)]^{-1} = [E - B(\pi')]^{-1}B(\pi') + H$$

Clearing this expression of inverses we get

$$[E - B(\pi')]B(\pi) = B(\pi')[E - B(\pi)] + [E - B(\pi')]H[E - B(\pi)]$$

Hence

$$B(\pi) = B(\pi') + [E - B(\pi')]H[E - B(\pi)]$$

or

$$D = [B(\pi) - B(\pi')] = [E - B(\pi')]H[E - B(\pi)]$$

$$D[E - B(\pi)]^{-1} = [E - B(\pi')]H$$

Multiplying by F^B

$$D[E - B(\pi)]^{-1}F^B = [E - B(\pi')](HF^B)$$

$$= \begin{pmatrix} 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{pmatrix}$$

which establishes sufficiency.

Necessity follows from the fact that

$$D[E - B(\pi)]^{-1}F^B = 0 \Longrightarrow D[E - B(\pi)]^{-1} = H'$$

where $H_i' \varepsilon$ Kern. (F^B) $\quad i = 1, 2, \ldots, n_B$

Write

$$H' = [E - B(\pi')]([E - B(\pi')]^{-1}H')$$

$$= [E - B(\pi')]H$$

and since

$$H_i F^B = [E - B(\pi')]_i^{-1} H' F^B = 0$$

$$i = 1, 2, \ldots, n_B$$

Reversing the previous argument for sufficiency gives the theorem.

Q. E. D.

Theorem 9.4 shew that two policies are equivalent if and only if one can be interchanged on the first time step for the other without affecting the total expected cost, i.e.,

$$\sum_{i=0}^{\infty} B(\pi) B(\pi')^i F^B = X(\pi') \qquad (*)$$

and

$$X(\pi) = \sum_{i=0}^{\infty} B(\pi') B(\pi)^i F^B$$

Reversing the role of π and π' in Theorem 9.4 was required to get (*) above.

9.8 EXPERIMENTATION ON THE EXPANDED MACHINE

Given an n+1 state optimal control problem, the expanded machine B has $(n+1)^2 m$ states where m is the number of possible input actions. Were we to use the methods of Theorem 9.3 to obtain a bound on the length of the sequence of costs $\{X_i'(\pi,t)\}_{t=1}^{t^*}$, the bound would be $2(n+1)^2 m-2$. However, Theorem 9.3 can be translated directly into machine B and we get:

Theorem 9.5

If for the expanded machine B associated with an n+1 state optimal control problem with proper policies π and π'

$$S_i B(\pi)^t F^B = S_i B(\pi')^t F^B \quad \text{for } t = 1, 2, \ldots, 2n$$

$$i = (j_1, 1, \pi_1), \ldots, (j_n, n, \pi_n)$$

$$\text{for arbitrary } j_1, \ldots, j_n$$

then

$$X(\pi) = X(\pi')$$

The difference between the bound $2(n+1)^2m-2$ for arbitrary matrices and 2n for the matrices of B clearly illustrates the special character of the matrices of the expanded machines.

9.9 OPTIMAL CONTROL PROBLEMS WITH OPTIMAL POLICIES OF SEQUENCES OF INPUTS

This section differs from previous sections in two ways. All costs will depend only on the states. But more important, the concept of policy will be extended to a function which assigns to states strings of elementary actions, i.e., strings of symbols rather than individual symbols. We might have done this in Lemma 9.1 when establishing the correspondence between optimal control problems and probabilistic sequential machines. However, treating the elementary actions as symbols seemed to be the most natural procedure. The consequence of this difference in viewpoint is that the states that the machine passes through while the string comes in are not used in the policy function.

Theorem 9.6

Let $B = \langle n+1, \; , \Sigma, B(\sigma): \forall\sigma \in \Sigma, F^B, 0_B\rangle$ be an abstract probabilistic sequential machine constructed from the n+1 state optimal control problem with set of actions Σ and costs F^B associated only with the states. Suppose B has a proper policy π^*.

Let $R_E(B)$ be the expectation equivalence reduction relation defined by B.

If rank $|R_E(B)| = r$ (finite) there is a deterministic policy of strings $\pi_D = (x_1^\circ, \ldots, x_n^\circ)$ such that $X(\pi_D) \leq X(\pi)$ for any stationary policy π.

Proof: We modify Theorem 1.3 by defining $0_{A'}(R_E(x))$ to be a vector of outputs A(x)F. In fact, Example 1.3 illustrates this point. The initial state I is not used in showing that R_E has finite rank, but only in computing the expectation, IA(x)F, serving as output for the "state" $R_E[x]$. Instead, the vector A(x)F could have been used as the output.

Hence if R_E has finite rank r there is a deterministic machine $\Sigma^*/R_E = A'$ with r states such that $0_{A'}(R_E(x)) = A(x)F$. Schematically we have (for $\Sigma = \{0,1\}$ W.L.G.):

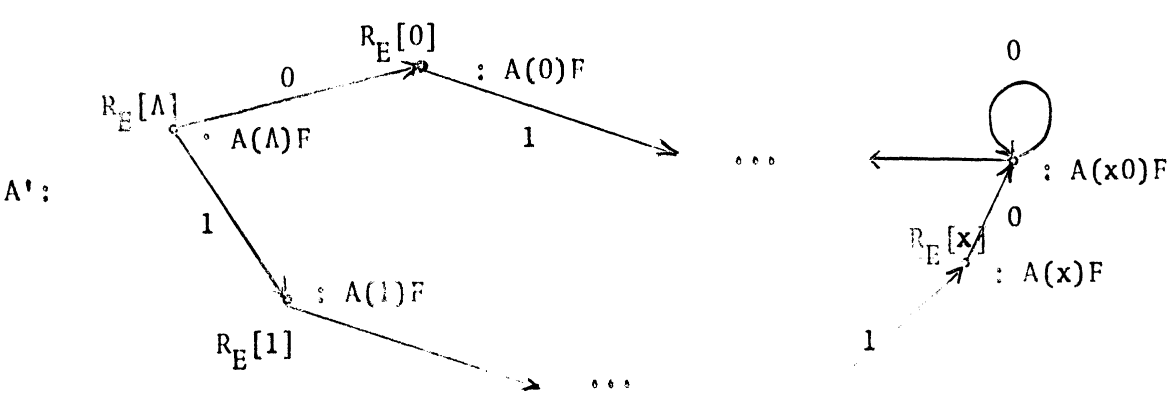


Figure 9.2. The constructed finite deterministic machine with vector outputs.

Any stationary policy π is defined by a stationary stochastic process. Started from initial state S_i at time t, the process causes a distribution of probabilities $q^{\pi}_{S_i}(x,t)$ over all strings x of length t.

We attach the stochastic process which provides the set of distributions $\{q^{\pi}_{S_i}(x,t)\}_{t=0}^{\infty}$ over input sequences to the input of $B' = \Sigma^*/R_E(B)$ which has vectorial outputs $0_{B'}(R_E[x]) = B(x)F^B$. Let $x_1, \ldots, x_r$ be arbitrary representatives of the class of R_E. Every string x produced by the stochastic process falls into some class of R_E, indeed a probability distribution is induced over the classes of R_E which are just the states of B'. Call this probability $P^{\pi}_i(R_E[x_j],t)$

for the class with representative x_j. The occurrence of different strings of the same length are disjoint events. Hence:

$$P_i^\pi(R_E[x_j],t) = \sum_{\substack{x \\ x\varepsilon(\Sigma)^t \cap R_E[x_j]}} q_{s_i}^\pi(x,t)$$

We now show that the expected cost of the t'th step is given by:

$$X_i(\pi,t) = \sum_{j=1}^{r} P_i^\pi(R_E[S_j],t)\cdot B(x_j)_i F^B$$

Using the notation of 2.1 we have:

(where S(0) is the state at time t = 0)

$$X_i(\pi,t) = E\{\sum_{x\varepsilon(\Sigma)^t} 0_B^*(x)/S(0) = S_i\}$$

$$= \sum_{x\varepsilon(\Sigma)^t} q_{s_i}^\pi(x,t)S_i B(x)F^B$$

Since each x has a representative x_j

$$= P_i^\pi(R_E[x_1],t)\cdot B(x_1)_i F^B + \dots + P_i^\pi(R_E[x_r],t)\cdot B(x_r)_i F^B$$

which establishes the above.

Consequently, the expected cost of the t'th step for policy π can be viewed as an expectation of the distribution $(P_i^\pi(R_E[x_1],t), \dots, P_i^\pi(R_E[x_r],t))$ over the states of B'.

Since we have assumed the existence of a proper policy π^* $X(\pi^*) = \sum_{t=0}^{\infty} X_i(\pi^*,t)$ must be finite for all i = 1, 2, ..., n

But

$$X_i(\pi^*) = (\sum_{t=0}^{\infty} P_i^{\pi^*}(R_E[x_1],t))\cdot B(x_1)_i F^B + \dots + (\sum_{t=0}^{\infty} P_i^{\pi^*}(R_E[x_r],t))\cdot B(x_r)_i F^B$$

Since $\sum_{j=1}^{r} P_i^{\pi^*}(x_j,t) = 1$, and the costs are non-negative, at least one of the terms, say the x_i^* diverges

$$\sum_{t=0}^{\infty} P_i^{\pi^*}(R_E[x_i^*],t) \longrightarrow \infty$$

for any stationary stochastic process. In order for $X_i(\pi^*)$ to be finite, it follows that at least (W.L.G.)

$$B(x_i^*)_i F^B = 0$$

Consider the class of stings $G^*(i)$:

$$G^*(i) = \{x : x \varepsilon R_E[x_i^*]\}$$

Given the machine $B' = \Sigma^*/R_E(B)$; $G^*(i)$ consists of all paths from $R_E[\Lambda]$ to $R_E[x_i^*]$. The expected total cost of the deterministic sequence $x = i_1 \ldots i_v$ is just the sum of the i'th components of the outputs of the states of B'

$$E(\text{total cost of } x) = \sum_{j=1}^{v} B(i_1 \ldots i_j)_i F^B$$

The shortest path from $R_E[\Lambda]$ to $R_E[x_i^*]$ is a good candidate for minimality of cost. Of course, depending on the values attached to states, a longer one might have lower cost.

Suppose x_i^o is a deterministic string with minimal total cost of all strings in $R_E[x_i^*]$. We show any stationary policy is no less costly than the policy $\pi_D = (x_1^o, \ldots, x_n^o)$

Let π be a stationary proper policy. As we have shown

$$X_i(\pi,t) = \sum_{j=1}^{r} P_i^\pi(R_E[x_j],t) B(x_j)_i F^B$$

In order for $X_i(\pi)$ to be finite

$$\lim_{t \to \infty} P_i^\pi(R_E[x_i^*],t) = 1$$

That is, all strings produced with positive probability by the stochastic process defining π are ultimately in $R_E[x_i^*]$.

If a string z has occurred, then its substrings must have occurred also. Hence the total cost of the policy π from state S_i at time t is just:

$$X_i(\pi,t) = \sum_{\substack{\text{all } z \varepsilon \Sigma^*: \\ \ell g(z)=t}} \text{Prob.}(z \text{ occurs}) \cdot E(\text{total cost of } z)$$

Hence

$$X_i(\pi) = \lim_{\ell g(z) \longrightarrow \infty} \sum_{\substack{z\varepsilon\Sigma^* \\ \ell g(z)=t}} \text{Prob.}(z \text{ occurs})\cdot E(\text{total cost of } z)$$

Since all z with positive probability of occurring are in $R_E[x_i^*]$:

$$X_i(\pi) = \lim_{\ell g(z) \longrightarrow \infty} \sum_{\text{all } z\varepsilon R_E[x_i^*]} q_{s_i}^{\pi}(z, g(z))\cdot E(\text{total cost of } z)$$

But by definition

$X_i(\pi_D) = E(\text{total cost of } x_i^o) \leq E(\text{total cost of } z)$ for all $z\varepsilon R_E[x_i^*]$

Substituting we get

$$X_i(\pi) \geq \lim_{\ell g(z) \longrightarrow \infty} \Big(\sum_{z\varepsilon R_E[x_i^*]} q_{s_i}^{\pi}(z,\ell g(z))\Big)\cdot X_i(\pi_D)$$

But since all z which occur with positive probability are in $R_E[x_i^*]$

as $\ell g(z) \longrightarrow \infty$

$$\lim_{\ell g(z) \longrightarrow \infty} \Big(\sum_{z\varepsilon R_E[x_i^*]} q_{s_i}^{\pi}(z,\ell g(z))\Big) = 1$$

which gives

$$X_i(\pi) \geq X_i(\pi_D)$$

Q. E. D.

Derman [1962] has shown that the approach of Eaton and Zadeh cannot be improved upon by collecting added information. That is, a policy consisting of deterministic selection of input symbols based upon knowledge of the state alone has as low a cost as any other type of policy. In Theorem 9.6 we have shown that for the class of machines which have rank of R_E finite, the state information is not always needed. An optimal policy consisting of sequences of inputs can be found which disregards intermediate states.

CHAPTER 10

SUMMARY AND OPEN PROBLEMS

10.1 SUMMARY

Let us summarize the most important original aspects of the work. Chapter 1 contains a method of constructing a finite deterministic machine (if one exists) which is expectation equivalent to a given probabilistic sequential machine (Theorem 1.3 and Theorem 1.7). In Chapter 2 the method is extended giving invariant subspace conditions for the existence of an input-state calculable (deterministic switching but random outputs) machine which is N-moment equivalent to a given probabilistic sequential machine (Theorems 2.4 and 2.5).

Chapter 3 raises the criterion of interchangeability as submachines as a condition of behavioral equivalence. Indistinguishability is the only equivalence which meets the interchangeability condition (Theorem 3.4). It is also proven that the class of finite deterministic machines can distinguish between two tape equivalent machines which are not expectation equivalent (Theorem 3.2).

Chapter 4 provides bounds on the length of strings necessary for deciding $\equiv_E$ (Theorem 4.3), $\equiv_N$ (Theorem 4.5), the reduction relations R_E (Theorem 4.4) and R_N (Corollary 4.5).

Chapter 5 extends the Rabin reduction theorem for probabilistic automata to probabilistic sequential machines (Theorem 5.2). A finite set of invariants exists for $\equiv_T$ for isolated cutpoint machines (Theorem 5.3). Using the notion of stability congruence relation, a sufficient condition is found for the existence of a finite set of invariants for $\equiv_T$ (Theorem 5.4). Lastly, two new properties are defined, the balance property and stability near the cutpoint, which lead to a class of machines with a finite set of invariants for $\equiv_T$ (Theorem 5.5).

Chapter 6 considers stability problems for $\equiv_E$, $\equiv_N$, and $\equiv_I$ simultaneously. Properties of general stability transformations are obtained (Theorem 6.1 and Theorem 6.2) in terms of error matrices. Properties of error matrices for the three cases above are summarized in Table 6.1. The notions of invariant error matrices and eigenerrors are introduced to allow the three equivalences to be studied together. A sufficient condition for the existence of eigenerrors is found (Theorem 6.4). A symbol matrix of a cyclic permutation of all the states ($n \geq 3$) has no nontrivial eigenerrors (Theorem 6.5). Finally, it is shown that stability transformations associated with invariant error matrices preserve the behavior of eigenstates, the stochastic vectors mapped into themselves by the symbol matrices (Theorem 6.6 and Theorem 6.7).

Chapter 7 brings out the essential difference between stability problems for $\equiv_T$ and the machine equivalences of Chapter 6 by translating the tape acceptance of a machine into the notation of convex sets. For nonsingular machines, necessary and sufficient conditions are obtained (Theorem 7.2 and Theorem 7.4). A sufficient condition for nonsingular machines to be equivalent by $\equiv_D$ is found as a by-product (Theorem 7.3). Extensions to certain singular cases are possible.

Chapter 8 shows the essential difference between minimization of states for $\equiv_T$ and the equivalences $\equiv_E$, $\equiv_N$, and $\equiv_I$. It is proven formally that a minimal state machine can be obtained via state reduction for $\equiv_E$, $\equiv_N$, and $\equiv_I$ (Theorem 8.1 and Theorem 8.2). Examples are presented showing the failure of state reduction methods for $\equiv_T$ (Remark 8.3). A necessary condition is provided for state reduction to be commutative for $\equiv_T$ (Theorem 8.3). Finally, machines with four states are discussed geometrically to exhibit fundamental properties of state reduction for $\equiv_T$ (Theorem 8.4).

Chapter 9 connects probabilistic sequential machines with optimal control theory. Several equivalences having experimental interpretations are defined for input and output events. All machine equivalences in this thesis are related by Figure 9.1. A translation is made between the terminology of probabilistic machines and that of optimal control (Lemma 9.1). The subject of equivalent policies is considered for certain special cases (Theorems 9.1, 9.2 and 9.3). A characterization of equivalent proper policies for machines with costs associated with the states is obtained (Theorem 9.4). The reduction relation R_E of Chapter 1 is used to obtain a sufficient condition for the existence of optimal policies consisting of strings rather than symbols (Theorem 9.6).

10.2 OPEN PROBLEMS

A great many open problems have been suggested by this research. There are a rather large number of new concepts contained in this thesis which seem to deserve further study.

In Chapter 2, there is no bound relating the number of states of an input-state calculable machine to the number of states of an N-moment equivalent probabilistic sequential machine. Perhaps the method of Theorem 5.1 could be used to obtain such a bound. In addition, the methods of Chapter 2 may be able to be extended to given an input-state calculable machine equivalent by $\equiv_T$ to a probabilistic sequential machine

It would be interesting to know bounds similar to those of Chapter 4 for deciding whether the new equivalences of Chapter 9 hold between arbitrary machines.

One of the principal results of Chapter 5 is not satisfactory in that it is either a characterization of certain isolated cutpoint machines or else is a characterization of a broader class than isolated cutpoint machines (Theorem 5.5). Finding out which is the case involves either finding a very fortunate example or additional results concerning the basic concepts involved. The discovery of any classes of non-isolated cutpoint machines which have a finite set of invariants for $\equiv_T$ would be very enlightening.

The stability problems of Chapter 6 and Chapter 7 seem to provide a very rich area of study. The concept of invariant error matrix, when fully developed theoretically, may provide insight into decompositions of probabilistic machines.. The notion of lumpability used by Bacon leads to one kind of invariant error matrix and to a decomposition which is equivalent to the original machine by $\equiv_I$. Other kinds of invariant error matrices may lead to decompositions equivalent by different machine equivalences. The theorey of convex sets seems especially relevant to the equivalences $\equiv_T$, $\equiv_{\lambda IE}$ and $\equiv_{\lambda IO}$. The machine interpretation of additional basic notions of convex sets and the convex set interpretation of additional machine concepts would make the connection shown in Chapter 7 more important.

An interesting class of problems arises if the probabilities of the probabilistic sequential machines are identified with physical quantities. Suppose we construct a voltage analog of the expectation of a probabilistic sequential machine using n voltage storage devices. The output values F_i correspond to amplification factors. Let us pass its analog output (corresponding to $E_A(x)$) through a threshold device which emits a 1 when the voltage is greater than or equal to λ

and a 0 otherwise. Consequently, the constructed device is able to compute $T(A,\lambda)$ by analog means. But how is the size of such a device related to the size of a deterministic machine which also recognizes the set $T(A,\lambda)$? If we have a deterministic machine D which recognizes a set $T(D)$, how do we find the smallest (by number of states which are voltage storage devices in this example) probabilistic automata A' which recognizes $T(A',\lambda) = T(D)$?

A partial answer to the first question is provided by the generalization of the Rabin-Paz bound (Theorem 5.2). We let n_D be the number of states of the deterministic machine, n the number of voltage storage devices of the analog tape recognizing machine:

$$n_D \leq (1 + \frac{d}{2\delta})^{n-1} \quad \text{where } d = F_{max} - F_{min}$$

The term 2δ can be interpreted as the minimum separation in expectation that can be allowed between inequivalent input sequences. In this example, δ is related to the precision of voltage measurement. Unfortunately no characterization of the class of machines for which the bound is met exactly is known. The Appendix contains probabilistic machines for which no tape equivalent deterministic machine meets the bound.

It is easy to show that the following relationship holds between two state "cores" of D and the voltage storage devices.

$$\text{Number of voltage storage devices} \geq \frac{1}{\log_2(1 + \frac{d}{2\delta})} \cdot \text{Number of "cores"} + 1.$$

The answer to the second question, providing an algorithm for going from a set of tapes accepted to a minimal state probabilistic machine which accepts the same set of tapes would provide interesting results for the representation theory of abstract algebras.

APPENDIX

We exhibit a class of machines for which the Rabin-Paz bound is non-minimal. Each member of the class will accept the same set of tapes as some finite deterministic machine, but the bound will be too high by a factor of at most e.

Let the class be $C = \{A^{(p)} : p \in [0,1]\}$ where

$$A^{(p)} = \langle 2, (1, 0), \{0\}, A(0), \begin{pmatrix} 1 \\ 0 \end{pmatrix}, 0_A \rangle$$

and

$$A(0) = \begin{pmatrix} 1-p & p \\ 0 & 1 \end{pmatrix}$$

$$[IA(0^r)]_1 = [(1,0)\begin{pmatrix} 1-p & p \\ 0 & 1 \end{pmatrix}^r]_1$$

$$= (1-p)^r$$

$$E_{A^{(p)}}(0^r) = (1-p)^r$$

pick some finite r' and let

$$(1-p)^{r'} - (1-p)^{r'+1} = 2\delta$$

Hence

$$2\delta = (1-p)^{r'}p$$

Pick $\lambda = (1-\varepsilon)^{r'} - \delta$ which makes it an isolated cutpoint with separation δ.

It is clear that

$$T(A^{(p)},\lambda) = \{0^t : t \leq r'\}$$

Consequently there is a deterministic machine D with r'+2 states (and no less) such that

$$T(A^{(p)},\lambda) = T(D)$$

The Rabin-Paz bound gives

$$n_D \leq (1 + \frac{1}{(1-p)^{r'}p})^1$$

Note that for any r', the function

$$f(p) = \frac{1}{(1-p)^{r'}p}$$

is continuous with respect to p and has minimum value for $p = \hat{p} = \frac{1}{r'+1}$.
Hence

$$f(p) \geq f(\hat{p}) = \frac{r'+1}{(1 - \frac{1}{r+1})^{r'}}$$

As we have seen above, $n_D = r'+2$. Substituting the values in the bound we get

$$r'+2 \leq 1 + \frac{r'+1}{(1 - \frac{1}{r'+1})^{r'}}$$

We observe that

$$\lim_{r' \to \infty} \frac{1}{(1 - \frac{1}{r'+1})^{r'}} = e = 2.71828....$$

Furthermore, by checking a few terms, it can be seen that

$$\frac{1}{(1 - \frac{1}{r'+1})^{r'}} < e \text{ for any finite } r'$$

However, for these probabilistic machines, a good approximation for n_D is obtained by dividing the Rabin-Paz bound by e. For the following case the bound can be divided by e also

It is clear that if the n state "probabilistic" machine of the bound is actually a deterministic automaton, we obtain

$$n_D = n \leq (1 + \frac{1}{1})^{n-1}$$

$$n \leq 2^{n-1}$$

$$n \leq \frac{1}{e} 2^{n-1} \text{ for } n \geq 5$$

It is not known whether there are other classes of machines for which the Rabin-Paz bound is too high by the factor of e.

BIBLIOGRAPHY

1. Arbib, M. [1964], "Automata Theory and Control Theory — A Rapprochement' (unpublished manuscript to appear in SIAM).

2. Arnold, R. F. [1964], unpublished communication.

3. Bacon, G. C. [1964], "The Decomposition of Stochastic Automata," Information and Control, 7, September 1964, pp. 320-329.

4. Braines, S. N. and Svechincky, V. B. [1962], "Matrix Structure in Simulation of Learning" I.R.E. Transactions on Information Theory Vol. IT-S, No. 5, Sept., 1962, pp. 186-190.

5. Burks, A. W. [1961], "Computation, Behavior and Structure in Fixed and Growing Automata," Behavioral Science, Vol. 6, No. 1, January 1961.

6. Burks, A. W. and Wang, H. [1957], "The Logic of Automata — Part II," JACM, Vol. 4, No. 3, July 1957, p. 286.

7. Carlyle, J. W. [1961], "Equivalent Stochastic Sequential Machines," Institute of Engineering Research Report Series, No. 60, Issue No. 415, Electronics Research Laboratory, University of California, Berkeley, California.

8. Carlyle, J. W. [1965], "State-Calculable Stochastic Sequential Machines, Equivalences, and Events," 1965 IEEE Conference Record on Switching Theory and Logical Design, pp. 258-263.

9. Derman, C. [1962], "On Sequential Decisions and Markov Chains," Management Sci. Vol. 9, p. 16.

10. Eaton, J. H. and Zadeh, L. A. [1962], "Optimal Pursuit Strategies in Discrete State Probabilistic Systems," Trans. ASME (Journal of Basic Engineering), March 1962, pp. 23-29.

11. Gantmacher, F. R. [1959], The Theory of Matrices, Vol. 2, Chelsea Publishing Co., N. Y., N. Y., pp. 236-237.

12. Howard, R. A. [1960], Dynamic Programming and Markov Processes M.I.T. Press and John Wiley and Sons, Inc., New York, (Chapter 3).

13. Kemeny, J. G. et al [1959], Finite Mathematical Structures, Prentice Hall, Inc., Englewood Cliffs, N. J., 1959, pp. 345-346.

14. Moore, E. F. [1956], "Gedanken-experiments on Sequential Machines", Automata Studies, C. E. Shannon and J. McCarthy eds., Princeton University Press, pp. 129-153.

15. Paz, A. [1964], "Probabilistic Automata, Part I and Part II," (to appear in Information and Control).

16. Rabin, M. O. [1964], "Probabilistic Automata," Sequential Machines, Selected Papers, Edited by E. F. Moore, Addison-Wesley Publishing Co., Reading, Mass., pp. 98-114.

17. Rabin, M. O., and Scott, D [1959], "Finite Automata and Their Decision Problems," IBM Journal Res. and Dev., 3, 1959, pp. 114-125.

18. Rådström, H. [1949], "Polar Reciprocity," Institute for Advanced Study Seminar on Convex Sets, 1949-1950, (U. of Michigan Library QA-248.P95), pp. 27-29.

19. Shannon, C. E. and Weaver, W. [1948], The Mathematical Theory of Communication, University of Illinois Press, Urbana, pp. 34.

20. Thrall, R. M. and Tornheim, L. [1957], Vector Spaces and Matrices, John Wiley and Sons, Inc., New York, pp. 298-300.

DISTRIBUTION LIST

(One copy unless otherwise noted)

Technical Library
Director Defense Res. & Eng.
Room 3C-128, The Pentagon
Washington, D.C. 20301

Defense Documentation Center 20
Cameron Station
Alexandria, Virginia 22314

Chief of Naval Research 2
Department of the Navy
Washington 25, D.C.
Attn: Code 437, Information
Systems Branch

Director, Naval Research Laboratory 6
Technical Information Officer
Washington 25, D.C.
Attention: Code 2000

Commanding Officer 10
Office of Naval Research
Navy 100, Fleet Post Office Box 39
New York, New York 09599

Commanding Officer
ONR Branch Office
207 West 24th Street
New York 11, New York

Office of Naval Research Branch
Office
495 Summer Street
Boston, Massachusetts 02110

Naval Ordnance Laboratory
White Oaks, Silver Spring 19
Maryland
Attn: Technical Library

David Taylor Model Basin
Washington,D.C. 20007
Attn: Code 042, Technical Library

Naval Electronics Laboratory
San Diego 52, California
Attn: Technical Library

Dr. Daniel Alpert, Director
Coordinated Science Laboratory
University of Illinois
Urbana, Illinois

Air Force Cambridge Research Labs
Laurence C. Hanscom Field
Bedford, Massachusetts
Attn: Research Library, CRMXL R

U. S. Naval Weapons Laboratory
Dahlgren, Virginia 22448
Attn: G. H. Gleissner, Code K4
Asst. Dir. for Computation

National Bureau of Standards
Data Processing Systems Division
Room 239, Building 10
Washington 25, D.C.
Attn: A. K. Smilow

George C. Francis
Computing Laboratory, BRL
Aberdeen Proving Ground, Maryland

Office of Naval Research
Branch Office Chicago
219 South Dearborn Street
Chicago, Illinois 60604

Commanding Officer
ONR Branch Office
1030 E. Green Street
Pasadena, California

Commanding Officer
ONR Branch Office
1000 Geary Street
San Francisco 9, California

DISTRIBUTION LIST (Concluded)

The University of Michigan
Department of Philosophy
Attn: Professor A. W. Burks

National Physical Laboratory
Teddington, Middlesex, England
Attn: Dr. A. M. Uttley, Supt.
Autonomics Division

Commanding Officer
Harry Diamond Laboratories
Washington, D.C. 20438
Attn: Library

Commanding Officer and Director
U. S. Naval Training Device Center
Port Washington
Long Island, New York
Attn: Technical Library

Department of the Army
Office of the Chief of Research
and Development
Pentagon, Room 3D442
Washington 25, D.C.
Attn: Mr. L. H. Geiger

National Security Agency
Fort George G. Meade, Maryland
Attn: Librarian, C-332

Lincoln Laboratory
Massachusetts Institute of Technology
Lexington 73, Massachusetts
Attn: Library

Office of Naval Research
Washington 25, D.C.
Attn: Code 432

Kenneth Krohn
6001 Dunham Springs Road
Nashville, Tennessee

Mr. Laurence J. Fogel
General Dynamics/Astronautics
Division of General Dynamics Corp.
San Diego, California

UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)*
Logic of Computers Group
The University of Michigan
Ann Arbor, Michigan 48104

2a. REPORT SECURITY CLASSIFICATION
UNCLASSIFIED

2b. GROUP

3. REPORT TITLE

EQUIVALENCES BETWEEN PROBABILISTIC SEQUENTIAL MACHINES

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*
Technical Report

5. AUTHOR(S) *(Last name, first name, initial)*

Page, Carl V.

6. REPORT DATE	**7a. TOTAL NO. OF PAGES**	**7b. NO. OF REFS**
November, 1965	178	20

8a. CONTRACT OR GRANT NO.
Nonr 1224(21)

b. PROJECT NO.

c.

d.

9a. ORIGINATOR'S REPORT NUMBER(S)
03105-41-T

9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*

10. AVAILABILITY/LIMITATION NOTICES

Qualified requesters may obtain copies of this report from DDC.

11. SUPPLEMENTARY NOTES

12. SPONSORING MILITARY ACTIVITY
Office of Naval Research
Department of the Navy
Washington 25, D.C.

13. ABSTRACT

This report introduces new definitions of behavioral equivalence and shows relationships among the various notions of behavioral equivalence between probabilistic machines. Four basic problems are discussed for several different behavioral equivalences between machines. In what follows we use the symbol "$\equiv$" as a variable to denote one of the several behavioral equivalences considered in this report. Given an arbitrary probabilistic sequential machine A: (1) Is there an input-state calculable machine A' (a machine with deterministic switching and random outputs) such that $A \equiv A'$? (2) What are the machines A' with minimal number of states such that $A \equiv A'$? (3) How does one obtain all stable modifications of A, i.e., all machines A' such that the switching probabilities of A' and A differ but $A \equiv A'$? (4) Is there a finite bound on the length of experiments required to establish whether $A \equiv A'$ holds? (U)

The four basic problems are solved completely for some equivalences between machines and are left open for other equivalences. Some applications are made to optimal control problems. (U)

DD FORM 1 JAN 64 1473

UNCLASSIFIED
Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
1. sequential machine						
2. automata theory						
3. communications channel						
4. stochastic machines						
5. discrete stochastic processes						
6. markov processes with rewards						
7. congruence relations						
8. probabilistic automata						

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization *(corporate author)* issuing the report.

2*a*. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2*b*. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year; or month, year. If more than one date appears on the report, use date of publication.

7*a*. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7*b*. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8*a*. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8*b*, 8*c*, & 8*d*. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9*a*. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9*b*. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers *(either by the originator or by the sponsor)*, also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through
____."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through
____."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through
____."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring *(paying for)* the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as *(TS)*, *(S)*, *(C)*, or *(U)*.

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, rules, and weights is optional.

UNIVERSITY OF MICHIGAN

3 9015 03695 2292