A STUDY ON PROBABILISTIC MACHINE

Dissertation
submitted by

DIBYENDU DAS

to

The
Department of Mathematics
VISVA-BHARATI
Santiniketan

In partial fulfillment of the requirement for award of the Degree of

MASTER OF SCIENCE
IN
MATHEMATICS

Under the guidance and supervision of
Professor Swapan Raha
Department of Mathematics
***Visva-Bharati**, Santiniketan,* I N D I A

September 2020

A STUDY ON PROBABILISTIC MACHINE

Dissertation
submitted by

DIBYENDU DAS

Class: M.Sc. Semester - IV
Roll No.: M.Sc (Sem - IV) Math – 11
Registration No.: VB -0259 of 2018 - 2019

Department of Mathematics
Siksha-Bhavana (Institute of Sciences)
Visva-Bharati
Santiniketan

September 2020

Approved	Approved
Teacher-in-charge	Head of the Department

C E R T I F I C A T E

I, Sri Dibyendu Das, do hereby declare that the dissertation work on

'A study on probabilistic machine'

is a bonafide record of work done by me at the Department of Mathematics, Siksha-Bhavana, Visva-Bharati, Santiniketan. This dissertation is submitted in partial fulfillment of the requirements for the award of the Degree of M.Sc. in Mathematics.

Dibyendu Das

(Dibyendu Das)

Contents

List of Symbols

$\mathbb{N}$	the set of natural numbers
$\mathbb{R}$	the set of real numbers
Σ	alphabet
σ	input
Σ^*	set of all strings
β	behaviour of a automata

Acknowledgements

I am obligated to Prof. Prashanta Kumar Mandal, Head of the Department of Mathematics, for the kind interest he took in the preparation of this dissertation. This is my first acquintance with finite state machines. I would like to thank all my teachers who gave so generously of their time, knowledge and advice in connection with this dissertation. They have taught me much of what I know during the last two years at Santiniketan.

It is needless to say that without the support, encouragement, and friendly care of a number of people this dissertation would not have been possible. I am, in particular, grateful to my supervisor Prof. Swapan Raha for the constant support and supervision he had been providing throughout by means of introducing me to the subject, rejuvenating my thoughts with new ideas from seemingly noninteresting branches of Mathematics and thus guiding me through to the successful completion of this dissertation.

I am also thankful to my friends in Santiniketan for their unconditional support, inspiration and cooperation. They made my studies more enjoyable, and who could always be consulted with whenever in doubt. It has been a pleasure to be with them. Many thanks to all of them. Wishing them good luck with their future endeavours.

My deepest gratitude is to my parents for their love, affection, encouragement and unfailing emotional support during the period of my study in Santiniketan. Many thanks to them for the unwavering support and love. It has given me strength to continue my studies.

This dissertation has been primarily supported by the UGC SAP-DRS Phase-III program under the Department of Mathematics at Visva-Bharati. UGC's financial support is highly appreciated.

Chapter 1

Introduction

Finite automata are mathematical models for systems capable of a finite number of states which admit at discrete time intervals certain inputs (incoming signals) and emit certain outputs. If the system is in state s and the input is σ then the system will move into a new state s_i which depends only on s and σ and will have an output which depends only (is a function of) on s_i. Thus the system will transform a sequence of inputs into a sequence of outputs and the relevant aspect of the system is this transformation. Sequential circuits, and even whole digital computers, provided the computer operates using only internal memory or just a fixed amount of tape, are systems which behave like finite automata. There is an extensive literature on finite automata. In this dissertation we follow the notations and use some of the results on automata contained in the paper by Rabin and Scott (1959). In particular the formulation given there amounts to assuming that the set of outputs contains just two elements. This is a convenient restriction which we follow also here but the results immediately extend to the general case of more than two outputs. Because of the restriction to two-valued outputs automata can be viewed as defining sets of sequences of inputs (tapes) and this point of view is adopted throughout this dissertation.

Finite automata exhibit a deterministic behavior. The state s and input σ determine the next state of the automaton. It is quite natural to consider automata with stochastic behavior. The idea is that the automaton, when in state s and when the input is σ, can move into any state s_i and the probability for moving into state s_i is a function $p_i(s, \sigma)$ of s and σ.

A practical motivation for considering probabilistic automata is that even the sequential circuits which are intended to be deterministic exhibit stochastic behavior because of random malfunctioning of components. This situation was first taken up by yon Neumann (1956) who considered schemes for organizing combinatorial (and to some extent also sequential) circuits constructed with specific components so as to increase their reliability.

Chapter 2

Automata Theory

In this section we give a brief resume of the basic definitions and some basic results which will be used in the sequel, from the theory of finite (deterministic) automata. The exposition follows closely that in Rabin and Scott (1959). By automaton we shall mean, throughout this section, deterministic automaton.

Let Σ be a finite nonempty set, to be called the **alphabet**. Letter σ (with subscripts) will usually denote elements of Σ. The set of all finite sequences of elements of Σ will be denoted by Σ^*. The elements of Σ^* will be called **tapes**. The letters x, y, z, u, v (with subscripts) will always denote tapes.The empty tape (i.e., the sequence of length zero) will be denoted by Λ.

Definition 2.1. A **sequential machine without output** is a four tuple $\mathbb{S} = <S;\Sigma_k;M;a>$ where, S is a non-empty set called the set of internal states,$\Sigma_k = (\sigma_0, \sigma_1, ..., \sigma_k)$ is the input alphabet, M is the function from $S \times \Sigma_k \to S$,called the direct transition function or the next state function and a, a given element of S, called the initial state.

M can be extended to a function from $S \times \Sigma_k \to S$ to S by $M(s, \Lambda) =$ s, $M(s, x\sigma) = M(M(s,x), \sigma)$ $(s \in S, x \in \Sigma^*, \sigma \in \Sigma)$. $M(s,x)$ is the state in which $\mathbb{S}$ *gets off* the tape x if it started on x in state s.

Definition 2.2. An automaton or a recognition device is a five-tuple $\mathbb{S} = <S;\Sigma_k;M;a;F>$ where $<S;\Sigma_k;M;a>$ is a sequential machine without output and F is the given subset of S called the set of final states or output state.

Definition 2.3. The response function of a sequential machine $\mathbb{S} = <S;\Sigma_k;M;a;F>$ defined by rp_s is a function from $\Sigma_k^* \to S$ defined by

$$\forall x \in \Sigma_k^* rp_s(x) = M(a,x)$$

Definition 2.4. The behavior of a machine $\mathbb{S} =< S;\Sigma_k;M;a;F >$ is the set of all words recognised by $\mathbb{S}$ and denoted by

$$\beta_s = (x \in \Sigma_k^* : rp_s(x) \in F)$$

Definition 2.5. An event $\beta \subset \Sigma_k^*$ is called a regular event if for some finite automaton $\mathbb{S}$, $\beta_s = \beta$. Every finite event is regular. If U and V are regular so are $U \cap V$, U U V and $\Sigma_K^* - U$. (see Rabin and Scott, 1959). In Rabin and Scott (1959) a necessary and sufficient condition for an event to be regular was given in terms of right equivalence relations.

Definition 2.6. R is called a right congruence relation on a sequential machine $\mathbb{S} =< S;\Sigma_k;M;a;F >$ iff R is an equivalance reltion on S which has the following substitution property(s.p)

$$\forall (u,v) \in S \ \forall \sigma \in \Sigma_k \ u \ R \ v \implies M(u,\sigma) R M(v,\sigma)$$

Definition 2.7. Let R be a equivalence relation on a set S .Then R is said to refine a set $F \subset S$ iff

$$\forall (u,v) \in S \ u \ R \ v \implies (u \in F \iff v \in F)$$

2.1 Quotient Machine

Definition 2.8. Let R be a right congruence relation on a machine $\mathbb{S} =< S;\Sigma_k;M;a;F >$ s.t R refines F the quotient machine of $\mathbb{S}$ modulo R denoted by

$$\mathbb{S}/R =< T;\Sigma_k;N;b;G >$$

where $< T;\Sigma_k;N;b >=< S;\Sigma_k;M;a > /R$ and $G = \{R[u] : u \in F\}$. and $T = \{R[s] : s \in S\}$, and $(\forall s \in S)$ $(\forall \sigma \in \Sigma_K)$ $N(R(s),\sigma) = R[M(s,\sigma)]$ and $b = R[a]$.

Theorem 2.1. $\mathbb{S}/R$ *is well-defined.*

Theorem 2.2. $\beta_{\frac{S}{R}} = \beta_s$

2.2 Free Automaton

Definition 2.9. The right Free sequential machine without output over Σ_K, denoted by F_k,is defined by

$$F_k = < \Sigma_K^*;\Sigma_k;M;\Lambda >$$

where $(\forall x \in \Sigma_k^*)$ $(\forall \sigma \in \Sigma_k)$ $M(x,\sigma) = x\sigma$.The left Free machine may be similarly defined by $(\forall x \in \Sigma_k^*)$ $(\forall \sigma \in \Sigma_k)$ $M(x,\sigma) = \sigma x$.

Definition 2.10. Let β be any subset of Σ_k^*. The Free-automaton

$$F_k(\beta) =< \Sigma_k^*; \Sigma_k; M; \Lambda; \beta >$$

where $F_k \;=< \Sigma_k^*; \Sigma_k; M; \Lambda >$ is the free machine without output.

Definition 2.11. Let R be a right congruence relation on Σ_k^* and $\beta \subset \Sigma_k^*$ such that R refines β.The quotient sequential machine with output modulo R and parameter β denoted by $T(R,\beta)$ is defined by

$$T(R,\beta) \;=\; F_k(\beta)/R$$

i.e. $T(R,\beta) \;= < T; \Sigma_k; N, b, G >$ Where $< T; \Sigma_k; N; b > = \; F_k/R \;=\; T(R)$
and $G \;=\; \{R[U] : U \in \beta\}$

Theorem 2.3. *$T(R,\beta)$ is well defined.*

Theorem 2.4. $\beta_{T(R,\beta)} \;=\; \beta$.

2.3 Minimisation of the Free-automaton

Let β be any given subset of Σ_k^*.Let us try to find ,if possible, a finite state machine whose behaviour is β. We know that β is the behaviour of the free automaton $F_k(\beta)$ which has infinite number of states .But if by minimising $F_k(\beta)$ we obtain a finite machine then this provides an answer to our problem.

Definition 2.12. The congruence relation on $F_k(\beta) \;= < \Sigma_K^*; \Sigma_k; M, A, \beta >$ induced by β is called the right congruence relation on Σ_k^* induced by β and denoted by R_β i.e

$$(\forall x, y \in \Sigma_k^*)\; xR_\beta y \iff (\forall z \in \Sigma_k^*)(xz \in \beta \iff yz \in \beta)$$

R_β is often called ***Nerodes equivalence relation*** .

Theorem 2.5. *R_β is a Right congruence relation on Σ_K^*.*

Theorem 2.6. *$R_(\beta)$ refines β.*

Theorem 2.7. *R_β is the largest right congruence relation on Σ_k^* which refines β.*

Definition 2.13. The minimal machine associated with $F_k(\beta)$ will be denoted by $M(\beta)$ i.e.

$$M(\beta) \;=\; (F_k(\beta))^M \;= F_k(\beta)/R_\beta \;=\; T(R_\beta, \beta)$$

i.e. $M(\beta)$ is a machine $M(\beta) \;= < T; \Sigma_k; N; b; G >$, where $T \;=\; \{R_\beta[x] : x \in \Sigma_k^*\}$.
and $(\forall x \in \Sigma_K^*)$ $(\forall \sigma \in \Sigma_K)$ $N(R_\beta[x], \sigma) \;= R_\beta[x\sigma]$, $b \;=\; R_\beta[\Lambda]$, $G \;=\; \{R_\beta[u] : u \in \beta\}$.

Theorem 2.8. *$M(\beta)$ is well-defined.*

Theorem 2.9. $\beta_{M(\beta)} = \beta$

Theorem 2.10. ***(Rabin and Scott, 1959).***
A set $\beta \subset \Sigma_k^*$ is a regular event if and only if the number of equivalence classes of Σ_k^* by the equivalence relation R_β finite. If the number of equivalence classes is $\alpha < \infty$ then for a suitable automata $\mathbb{S}$, $\beta_s = \beta$ where the automaton $\mathbb{S}$ has α states. No automaton with fewer than α states defines β.

Chapter 3

Probabilistic Automata

We shall now define the basic concept of this investigation, namely the concept of probabilistic automata. It will be seen that probabilistic automata are like the usual automata except that now the transition table M assigns to each pair $(s, \sigma) \in S \times \Sigma$ certain transition probabilities.

Definition 3.1. A probabilistic sequential machine is a five tuple $\bar{\mathbb{S}} = < S; \Sigma_2; P; s_1; F >$ where$S = \{s_1, s_2, ..., s_n\}$ is a nonempty finite set called the set of internal states, Σ_2 the input alphabet,s_1 a given element of S called the initial state,F a given subset of S called the set of final states or the output set and $P: \ S \times \Sigma_2 \rightarrow [0,1]^n$ such that

$$\forall \sigma \in \Sigma_2 \ P(s_i, \sigma) = (p_1(s_i, \sigma), p_2(s_i, \sigma), ..., p_n(s_i, \sigma))$$

where $0 \leq p_j(s_i, \sigma) \leq 1$ (i,j=1,2,....n).and $\sum_{j=1}^{n} p_j(s_i, \sigma) = 1$ for all i,j. thus

$$P(\sigma) = \begin{bmatrix} P(s_1, \sigma) \\ P(s_2, \sigma) \\ . \\ . \\ P(s_n, \sigma) \end{bmatrix} = (p_j(s_i, \sigma)).$$

is a stochastic matrix. $p_j(s_i, \sigma)$ is usually called the probability of transition from state $s_i \rightarrow s_j$ when input σ occurs. And the matrix $P(\sigma)$ is called the matrix of transition probabilities or the transition matrix corresponding to input σ.

Remark: In the interpretation wenote that the transition probability from s_i to s_j when input σ occurs is $p_j(s_i, \sigma)$ which is independent of previous state or previous input.

Definition 3.2. For $\sigma \in \Sigma$ and $x = \sigma_1 \sigma_2 .....\sigma_n$ define the n+1 matrices $P(\sigma)$ and $P(x)$ by
$P(\sigma) = [p_j(s_i, \sigma)]_{0 \leq i \leq n, 0 \leq j \leq n}$
$P(x) = P(\sigma_1) P(\sigma_2) .... P(\sigma_n) = [p_j(s_i . \sigma)]_{0 \leq i \leq n, 0 \leq j \leq n}$

Remark: An easy calculation (involving induction on n) will show the (i+1,j+1) element $p_j(s_i, \sigma)$ is the probability of $\bar{\mathbb{S}}$ for moving from state s_i to state s_j by the input sequence x.

Definition 3.3. If $\bar{\mathbb{S}} = < S; \Sigma_2; P; s_0; F >$ and $F = \{s_{i_0}, s_{i_1}, ...s_{i_r}\}$, $I = i_0, i_1, ..., i_r$, define

$$p(x) = \sum_{i \in I} p_i(s_0, x)$$

.

p(x) clearly is the probability for $\bar{\mathbb{S}}$, when started in s_0, to enter into a state which is member of F by the input sequence x.

3.1 Sets of tapes defined by probabilistic automata

A p.a. $\bar{\mathbb{S}}$ may be used to define sets of tapes in a manner similar to that of deterministic automata except that now the set of tapes will depend not just on $\bar{\mathbb{S}}$ but also on a parameter λ.

Definition 3.4. Let $\bar{\mathbb{S}}$ be p.a. and λ be a real number, $0 \leq \lambda < 1$. The set of tapes $\beta_{\bar{S}}(\lambda)$ is defined by

$$\beta_{\bar{S}}(\lambda) = \{x : x \in \Sigma^*, \ \lambda < p(x)\}$$

If $x \in \beta_{\bar{S}}(\lambda)$ we say that x is accepted by $\bar{\mathbb{S}}$ with cut-point λ.

3.2 Relation between ordinary deterministic automata and probabilistic automata

Definition 3.5. $\mathbb{S} = < S; \Sigma_2; M; s_1; F >$ be an ordinary finite sequential machine where $S = \{s_1, s_2, ..., s_n\}$ then the probabilistic machine $\bar{\mathbb{S}} = < S; \Sigma_2; P; s_1; F >$ will called probabilistic machine associated with $\mathbb{S}$ where

$$p_j(s_i, \sigma) = \begin{cases} 1 & \text{if } M(s_i, \sigma) = s_j \\ 0 & \text{if } M(s_i, \sigma) \neq s_j \end{cases}$$

Theorem 3.1. *A probabilistic machine associated with an ordinary machine $\mathbb{S}$ uniquely determines $\bar{\mathbb{S}}$.*

Theorem 3.2. *Every Regular set is the behaviour of some probabilistic machine.*

Theorem 3.3. *There exists a probabilistic machine $\bar{\mathbb{S}}$ and a cut point λ $(0 \le \lambda < 1)$ such that $\beta_s(\lambda)$ is not regular.*

Proof: Consider a particular probabilistic machine $\bar{\mathbb{S}} = < S; \Sigma_2; P; s_0; F >$ where $S = \{s_1, s_2\}$ $F = \{s_2\}$ and $P(\sigma_0) = \begin{bmatrix} 1 & 0 \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$;$P(\sigma_1) = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ 0 & 1 \end{bmatrix}$

For any word $x = \sigma_{i_1}\sigma_{i_2}...\sigma_{i_n}$, if $p_n = .i_n i_{n-1}....i_1$ then $P(x) = \begin{bmatrix} 1-p_n & p_n \\ 1-p_n-2^{-n} & p_n+2^{-n} \end{bmatrix}$ and $p(x) = p_n$, i.e $p(\sigma_{i_1}\sigma_{i_2}....\sigma_{i_n}) = .i_n i_{n-1}....i_1$(writtenin binary scale) let $0 \le \lambda < \lambda^* < 1$. There exists a rational number having finite binary representation of the form $.i_n i_{n-1}...i_1$ such that $\lambda < .i_n i_{n-1}....i_1 < \lambda^*$ or $\lambda < p(\sigma_{i_1}\sigma_{i_2}.....\sigma_{i_n}) < \lambda^*$, so that $\sigma_{i_1}\sigma_{i_2}......\sigma_{i_n} \in \beta_S(\lambda)$ but not to $\beta_S(\lambda^*)$. Hence $\beta_S(\lambda)$ contains $\beta_S(\lambda^*)$ and $\beta_S(\lambda) \neq \beta_S(\lambda^*)$.

For every $\lambda \in [0,1)$ there corresponds a unique set $\beta_S(\lambda)$ and the correspondence is one to one since [0,1] is non-enumerable and $\{\beta_S(\lambda) : \lambda \in [0,1)\}$ is also non-enumerable but since the set of regular sets is enumerable there exists a $\lambda \in [0,1)$ s.t. $\beta_S(\lambda)$ is not regular.

3.3 Isolated cut-point

Definition 3.6. A cut point λ $(0 \le \lambda < 1)$ is said to be an isolated cut point of a probabilistic machine $\bar{\mathbb{S}} = < S; \Sigma_2; P; s_1; F >$ iff λ is neither a point nor a limit point of the set $\{p(x) : x \in \Sigma_2^*\}$.i.e. $\exists \delta > 0$ $s.t.$ $(\forall x \in \Sigma_2^*)$ $|p(x) - \lambda| \ge \delta$.

3.3.1 Motivation for introducing the isolated cut-point:

The following consideration applies the motivation for introducing the isolated cut-point. Let $\bar{\mathbb{S}}$ be a probabilistic machine having cut-point λ and x, a given word for which p(x) is unknown. We try to find out an experimental procedure to determine if $x \in \beta_{\bar{S}}(\lambda)$.Let E_n be the random experiment of making input x n-times succesively and noticing every time whether x is accepted or not.If the random variable X_n denote the number of time x is accepted in E_n, then we know $\frac{X_n}{n}$ converging to p(x) as $n \to \infty$. Given ϵ>0(0<ϵ<1) we can find an $n(\epsilon, \eta)$ s.t $P(|\frac{X_n}{n} - p(x)| < \eta) \ge 1 - \epsilon$. or $P(p(x) - \eta < \frac{X_n}{n} < p(x) + \eta) \ge 1 - \epsilon$.
If $x \in \beta_{\bar{S}}(\lambda)$ then $p(x) \ge \lambda$. choose $\eta = p(x) - \lambda$. then

$$P(\frac{X_n}{n} > \lambda) \ge P(\lambda < \frac{X_n}{n} < 2p(x) - \lambda) \ge 1 - \epsilon$$

If n is large enough $p(x) \approx \frac{X_n}{n}$ and on the basis of E_n we propose to take $x \in \beta_{\bar{S}}(\lambda)$ if we observe $\frac{X_n}{n} > \lambda$.The above lines state that the probability of making a correct dicision is not less than $1 - \epsilon$.

Now n depends on η which in this case is $p(x)-\lambda$ and hence n depends on p(x), which is unknown . Hence the performance of E_n bege a prior knowledge of p(x) which is impossible.The above difficulty may be avoided , if λ is an isolated cut-point . In this case there exist $\delta > 0$ s.t.$\forall x \in \Sigma_2^* \ |p(x)-\lambda| \ \geq \delta$,determine $n = n(\epsilon, \delta)$ s.t. $P(|\frac{X_n}{n} - p(x)| < \delta) \ \geq \ 1-\epsilon$. If $p(x) > \lambda$ then $p(x) \geq \lambda + \delta$ and $P(\frac{X_n}{n} > \lambda) \ \geq P(|\frac{X_n}{n} - p(x)| < p(x) - \lambda) \geq P(|\frac{X_n}{n} - p(x)| < \delta) \geq 1-\epsilon$. In this case n depends on ϵ *and* δ and not on p(x) so that the random experiment E_n can be performed.

3.4 The Reduction theorem

Theorem 3.4. *Let $\bar{\mathbb{S}}$ be a probabilistic automaton and λ be an isolated cut-point. Then there exists a deterministic automaton $\mathbb{S}$ such that $\beta_{\bar{S}}(\lambda) \ = \ \beta_S(\lambda)$. If $\bar{\mathbb{S}}$ has n states and F consists of just one state then $\mathbb{S}$ can be chosen to have α states where $\alpha \leq [1+\frac{1}{\delta}]^{\frac{n-1}{2}}$*

Proof:Let the set of states S be $\{s_0, s_1, ..., s_{n-1}\}$ and $F \ = \ \{s_{n-1}\}$. For every tape x, P(x) is an $n \times n$ matrix and p(x) is the upper left corner element of P(x).
Let $x_1, ..., x_k$ be tapes which are pairwise in-equivalent by R_β. thus for every $i \leq k, \ j \leq k, \ i \neq j$, there exists a tape y s.t.

$$x_i y \ \in \beta_{\bar{S}}(\lambda), \ \ x_j y \ \notin \ \beta_{\bar{S}}(\lambda) \ .....(1)$$

or vice versa.Let the first row of $P(x_i)$, $1 \leq i \leq n$, be $(\xi_1^i, ..., \xi_n^i)$ and the last column of P(y), for the particular y appearing above, be $(\eta_1, ...\eta_n)$. From $P(x_i y) \ = \ P(x_i)P(y)$ and $P(x_j y) \ = \ P(x_i)P(y)$ it follows that
$p(x_i y) \ = \ \xi_1^i \eta_1 \ + \ .... + \ \xi_n^i \eta_n$, $p(x_j y) \ = \ \xi_1^j \eta_1 \ + ... + \ \xi_n^j \eta_n$ ......(2)
Combining (1) and (2) we get $\lambda < \ \xi_1^i \eta_1 \ + \ .... + \ \xi_n^i \eta_n$ and $\xi_1^j \eta_1 \ + ... + \ \xi_n^j \eta_n \leq \lambda$ ....(3).
since λ is a isolated cut-point and $\delta \leq |p(x) - \lambda|$ for $x \in \Sigma^*$,(3) implies
$2\delta \leq \ (\xi_1^i - \xi_1^i)\eta_1 \ + ... + (\xi_n^i - \xi_n^i)\eta_n$. ....(4)
Taking absolute values and observing that the η_i, as elements of a stochastic matrix, satisfy $|\eta_i|$ (4) leads to
$2\delta \leq \ |\xi_1^i - \xi_1^i| \ + ... + |\xi_n^i - \xi_n^i|$........(5)
An argument involving volumes in n-dimensional space will now be used to infer from (5) a bound on k. The n-tuples $(\xi_1, ..., \xi_n)$ will be considered as points of Euclidean n-space. Let σ_i, $1 \leq i \leq k$ be the set

$$\sigma_i \ = \ \{(\xi_1, ..., \xi_n)| \ \xi_j^i \leq \xi_j, \ 1 \leq j \leq n, \ \sum_j (\xi_j - \xi_j^i) \ = \ \delta\}$$

.

Each σ_i is a translate of the set

$$\sigma \ = \ \{(\xi_1, ..., \xi_n)| \ 0 \leq \xi_j, \ 1 \leq j \leq n, \ \sum_j \xi_j \ = \ \delta\}$$

.

The set σ is readily seen to be an (n-1)-dimensional simplex which is a subset of the hyperplane $x_1 + x_2 + ... + x_n = \delta$. The n - 1 dimensional volume $V_{n-1}(\sigma)$ of σ, expressed as a function of δ ,is $c\delta^{n-1}$ where c is some constant not depending on δ.
From $\sum_j \xi_j^i = 1$ it follows that $(\xi_1, ..., \xi_n) \in \sigma_i$ implies

$$\sum_j \xi_j = 1 + \delta, \quad 0 \leq \xi_j, \ 1 \leq j \leq n$$

Thus $\sigma_i \subseteq \tau$ where

$$\tau = \{(\xi_1, ..., \xi_n) | \sum_j \xi_j = 1 + \delta, \ 0 \leq \xi_j, 1 \leq j \leq n\}$$

.

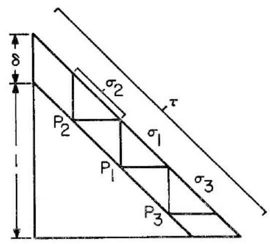

Fig. 1. The sets σ_i and τ for $n = 2$ and $k = 3$

A point $(\xi_1, ..., \xi_n) \in \sigma_i$ is an interior point of σ_i iff $\xi_p - \xi_p^i > 0$ for $1 \leq p \leq n$. because of (5) σ_i and σ_j , $i \neq j$,have no interior points in common . For otherwise ,if $(\xi_1, ..., \xi_n)$ is interior to both σ_i and σ_j , we would have $\xi_p - \xi_p^i > 0$, $\xi_p - \xi_p^j > 0$ and hence

$$|\xi_p^i - \xi_p^j| < |\xi_p - \xi_p^i| + |\xi_p - \xi_p^j|, \quad 1 \leq p \leq n.$$

Hence

$$\sum_p |\xi_p^i - \xi_p^j| < \sum_p |\xi_p - \xi_p^i| + \sum_p |\xi_p - \xi_p^j| = \delta + \delta,$$

contradicting (5).
Thus for $i \neq j$, σ_i and σ_j have no interior point in common.This implies

$$kc\delta^{n-1} = V_{n-1}(\sigma_1) + ... + V_{n-1}(\sigma_k) \leq V_{n-1}(\tau) = c(1+\delta)^{n-1}.$$

Hence $k \leq [1 + \frac{1}{\delta}]^{n-1}$.Thus the number of equivalence classes of the relation R_β is at most $[1 + \frac{1}{\delta}]^{n-1}$.Thus here required minimal deterministic machine is $\mathbb{S} = M(\beta)$ such that

$$\beta_{\bar{S}}(\lambda) = \beta_S(\lambda).$$

3.5 Saving of states

From the proof of the Reduction Theorem, it seems possible that in passing from a p.a. $\bar{\mathbb{S}}$ to an equivalent deterministic automaton we may have to increase the number of states. In other words, the p.a. is more economical in terms of number of states. The following theorem shows that this does in fact happen in certain cases.

Theorem 3.5. *There exists an automaton $\bar{\mathbb{S}}$ with just two states and a sequence λ_n , $1 \leq n < \infty$, of isolated cut-points such that for each n, the automaton $\mathbb{S}$ with the least number of states which satisfies $\beta_{\bar{S}}(\lambda_n) = \beta_S(\lambda_n)$ has at least n states.*

Proof: Let $\Sigma = \{0, 2\}$, $S = \{s_0, s_1\}$ and $F = \{s_1\}$. Let the transition probabilities be such that
P(0)=$\begin{bmatrix} 1 & 0 \\ \frac{2}{3} & \frac{1}{3} \end{bmatrix}$ and P(2)=$\begin{bmatrix} \frac{1}{3} & \frac{2}{3} \\ 0 & 1 \end{bmatrix}$.
It is easy to verify that if $x = \delta_1\delta_2......\delta_n \in \Sigma^*$ then

$$p(x) = \frac{\delta_n}{3} + \frac{\delta_{n-1}}{3^2} + ... + \frac{\delta_1}{3^{n-1}}.$$

Remembering that $\delta_i \in \{0, 2\}$ we see that the topological closure C of the set $P = \{p(x)|x \in \Sigma^*\}$ is precisely Cantor's discontinuum.
Thus all the points λ, $0 \leq < 1$,which satisfy $\lambda \notin C$ are isolated cut-points for $\bar{\mathbb{S}}$. Consider now the real number (written in ternary notation) $\lambda_n = .22.....211$ where the number of digits is n+1. For $x \in \Sigma^*$ to satisfy $\lambda_n < p(x)$ it is necessary and sufficient that x have the form $x = x_1 22....2$ where $x_1 \in \Sigma^*$ and the number of 2's is at least n.Thus the set $\beta_{\bar{S}}(\lambda_n)$ is nonempty and if $x \in \beta_{\bar{S}}(\lambda_n)$ then $n \leq l(x)$ .It follows from elementary theory of automata (see Rabin and Scott,1959, Theorem 7) that the minimal deterministic automata $\mathbb{S}$ for which $\beta_{\bar{S}}(\lambda_n) = \beta_S(\lambda_n)$ has at least n+1 states.

Chapter 4

Actual Automata

In certain actual situations it is natural to assume about an automaton $\bar{\mathbb{S}}$ that all transitions between states have strictly positive (though sometimes very small) probabilities. This motivates the following definition.

Definition 4.1. A p.a, $\bar{\mathbb{S}}$ is called an actual automaton if for all $s \in S$, $s_i \in S$, and $\sigma \in \Sigma$ the transition probability $p_i(s,\sigma)$ of moving from state s to state s_i under input σ satisfies $p_i(s,\sigma) \ > \ 0$.

4.1 Products of positive stochastic matrices

It turns out that actual automata have very special properties. To study them we need some results about products of strictly positive stochastic matrices. The following Lemma 6 is a restatement, in our notation, of Theorem 4.1.3 of Kemeny and Snell (1960); the proof is included for the sake of completeness. Corollary 7 and Lemma 8 are closely related to Theorems 4.1.4-4.1.6 of IKemeny and Snell (1960) except that we treat products of several matrices instead of powers of a single matrix. The possibility of this generalization was pointed out by A. Paz.

Definition 4.2. If $\alpha \ = \ [a_i]_{1\leq i\leq n}$ is a column vector then $||\alpha|| \ = \ max_i a_i - min_i a_i$.If A is an $n\times n$ matrix having columns $\alpha_1,...,\alpha_n$ then $||A||$ is defined by $||A|| \ = \ max_i \ ||\alpha_i||$.

Lemma 4.1. *If $P \ = \ [p_{ij}]_{1\leq i,j\leq n}$ is a $n\times n$ stochastic matrix and $\Delta \ = \ min_{i,j} \ p_{ij}$ and if $\alpha \ = \ [a_i]_{1\leq i\leq n}$ is a column vector then $||P\alpha|| \leq (1-2\Delta)||\alpha||$.*

Corollary 4.1. *If $H \ = \ \{P_1,....,P_k\}$ where the matrices P_i, $1\leq i\leq k$ are stochastic and all elements of the P_i are greater than $\Delta > 0$ then for any $1\leq i_1,.....,i_m \leq k$,*

$$||P_{i_1}P_{i_2}....P_{i_m}|| \leq (1-2\Delta)^{m-1}.$$

For any $m \times n$ matrix $A = [a_{ij}]$ we define $|A| = max_{i,j} \; |a_{ij}|$. This |A| clearly has the usual property of norm.

Lemma 4.2. *If P is a stochastic $n \times n$ matrix and $\alpha = [a_i]_{1 \leq i \leq n}$ is a column vector then*

$$|P\alpha - \alpha| \; \leq \; ||\alpha||.$$

Corollary 4.2. *If P is a stochastic $n \times n$ matrix and A is an $n \times n$ matrix then $|PA - A| \leq ||A||$.*

4.2 Definite events

It will turn out that the sets accepted by actual automata are just those described in the following.

Definition 4.3. A set $\beta \subseteq \Sigma^*$ is called a definite event if for some integer k the following holds. If $l(x) \geq k$ then $x \in \beta$ if and only if $x = yz$ where $l(z) = k$ and $z \in \beta$.

In (Perles, Rabin, and Shamir, 1963) the properties of definite sets and the (deterministic) automata defining them are studied in detail.

4.3 Actual automaton and definite set

Theorem 4.3. *If $\breve{\mathbb{S}}$ is an actual automaton and λ is an isolated cut-point then $\beta_{\breve{\mathbb{S}}}(\lambda)$ is a definite set. Conversely, every definite set is definable by some actual automaton with isolated cut-point.*

Proof: Let $\breve{\mathbb{S}} = < S; \Sigma_2; P; s_1; F >$ be the machine where the state set $S = \{s_1, s_2, ..., s_n\}$ and the output state $F = \{s_{l_1}, s_{l_2}, ..., s_{l_r}\}$.
Since $\breve{\mathbb{S}}$ is actual we can find some positive number $\Delta(0 < \Delta < 1/2)$ s.t

$$\forall \sigma \in \Sigma_2 \;\; p_j(s_i, \sigma) \geq \Delta \;\; (i,j = 1,2,..,n)$$

Since λ is an isolated cut-point,$\exists \; \delta > 0$ s.t. for every $x \in \Sigma_2^*$, $|p(x) - \lambda| \geq \delta$.
Choose a positive integer q such that $(1-2\Delta)^{q-1} < \frac{2\delta}{r}$. (r=no.of states of F)
Let $x \in \Sigma_2^*$ s.t. $l(x) = q$. If $x = \sigma_{i_1}\sigma_{i_2}....\sigma_{i_q}$, Then

$$||P(x)|| = ||P(\sigma_{i_1})P(\sigma_{i_2}).....P(\sigma_{i_n})|| \leq (1-2\Delta)^{q-1} \; by \; corollary(4.1)$$

Now $p(x) = \sum_{i=1}^{r} p_{l_i}(s_i,x)$ (probability of acceptance of x)
and $p(yx) = \sum_{i=1}^{r} p_{l_i}(s_i,yx)$ for all $y \in \Sigma_2^*$
Then

$$| p(yx) - p(x) | \leq \sum_{i=1}^{r} |p_{l_i}(s_1,yx) - p_{l_i}(s_1,x)| \leq r|P(yx) - P(x)| = r|P(y)P(x) - P(x)|$$

$$\leq r||P(x)|| < 2\delta \quad By\ [lemma\ 4.2]$$

Now $p(x) > \lambda \wedge p(yx) \leq \lambda \implies p(x) \geq \lambda + \delta \wedge p(yx) \leq \lambda - \delta \implies | p(yx) - p(x) | \geq 2\delta$ which is not true. Similarly $p(x) \leq \lambda \wedge p(yx) > \lambda$ is not true.
Hence we have

$$p(x) > \lambda \iff p(yx) > \lambda$$

Thus

$$\forall\ x,y \in \Sigma_2^*\ \ lg(x) = q \implies (x \in \beta_{\breve{S}}(\lambda) \iff yx \in \beta_{\breve{S}}(\lambda))$$

i.e. $\beta_{\breve{S}}(\lambda)$ is a weakly q-definite set. And so $\beta_{\breve{S}}(\lambda)$ is a definite set.

Converse part

Let α be a q definite set so that $\alpha = \alpha_1 \cup \Sigma_2^* \alpha_2$ where α_1 is the set of all words of α of length less than q and α_2 is the set of all words of α of length equal to q.
Write $\alpha_1 \cup \alpha_2 = \{x_1, x_2, ..., x_r\}$. We construct the probabilistic machine $\breve{\mathbb{S}} = < S; \Sigma_2; P; s_1; F >$ as follows ,
Let $c \neq 0\ or\ 1$, S is taken to be the set of all q-tuples of the form $(c, c, c, ...c, i_1, i_2, ..., i_m)$ where $i_1, i_2, ..., i_m = 0\ or\ 1$ and $0 \leq m \leq q$, Write $s_1 = (c, c, ..., c)$ (q-tuple).Thus $S = \{s_1, s_2, ..., s_n\}$ where $n = \sum_{m=0}^{q} 2^m = 2^{q+1} - 1$.
Now the mapping $\phi : (c, c, c, ...c, i_1, i_2, ..., i_m) \to \sigma_{i_1}\sigma_{i_2}.....\sigma_{i_m}$ is an one-to-one mapping from the set S onto the set of all words of length less than equal to q.
Now we construct the set of final states,let $F = \{s_{l_1}, ..., s_{l_r}\}$ where $\phi^{-1}(x_i) = s_{l_i}$ (i=1,2,...,r).
If $s_i = (\tau_1, \tau_2, ..., \tau_q)$,we define

$$p_j(s_i, \sigma_k) = \begin{cases} 1 - \epsilon & if \quad s_j = (\tau_2, .., \tau_q, k) \\ \frac{\epsilon}{n-1} & if \quad s_j \neq (\tau_2, .., \tau_q, k) \end{cases}$$

Where ϵ $(0 < \epsilon < 1)$ is to be chosen later on . Clearly, $\breve{\mathbb{S}}$ is an actual probabilistic machine.
If $x \in \Sigma_2^*$ is a word $x = \sigma_{i_1}\sigma_{i_2}....\sigma_{i_r}$,(r<q) and s_i is the set $(c, c, c, ..., c)$ then $p_j(s_i, \sigma_{i_1}\sigma_{i_2}....\sigma_{i_r}) = (1-\epsilon)^r$ for $s_j = (c, c, ..., c, i_1, i_2, .., i_r)$.If $x \in \Sigma_2^*$ is a word $x = \sigma_{i_1}\sigma_{i_2}....\sigma_{i_q}$ and s_i is any state then, $p_j(s_i, \sigma_{i_1}\sigma_{i_2}....\sigma_{i_q}) = (1-\epsilon)^q$ for $s_j = (i_1, .., i_q)$. If $x, y \in \Sigma_2^*$ are words such that $x = \sigma_{i_1}\sigma_{i_2}....\sigma_{i_q}$, then for $s_j = (i_1, .., i_q)$ and for any state s_i

$$p_j(s_i, yx) = \sum_{k=1}^{n} p_k(s_i, y)\ p_j(s_k, x) > (1-\epsilon)^q \sum_{k=1}^{n} p_k(s_i, y) = (1-\epsilon)^q \times 1 = (1-\epsilon)^q.$$

If $x \in \alpha_1$, then the corresponding word $x = \sigma_{i_1}\sigma_{i_2}....\sigma_{i_m}$ $(0 \leq m < q)$,And then $s_j = (c, c, ..., c, i_1, i_2, .., i_m) = \phi^{-1}(x) \in F$ so that

$$p(x) \geq p_j(s_1, \sigma_{i_1}\sigma_{i_2}....\sigma_{i_m}) > (1-\epsilon)^q$$

If $y \in \alpha_2$; $y = \sigma_{i_1}\sigma_{i_2}....\sigma_{i_q}$, So $s_j = (i_1, .., i_q) = \phi^{-1}(y) \in F$. And hence for any word $x \in \Sigma_2^*$

$$p(xy) \geq\ p_j(s_1, xy) > (1-\epsilon)^q.$$

Thus if $x \in \alpha$, $p(x) > (1-\epsilon)^q$.
If x is such that $lg(x) < q$ but $x \notin \alpha_1$ $x = \sigma_{i_1}\sigma_{i_2}....\sigma_{i_m}$ $(0 \leq m < q)$
$s_j = (c, c, .., c, i_1, .., i_m) = \phi^{-1}(x) \in S - F$, So that $p(x) \leq 1 - p_j(s_1, x) < 1 - (1-\epsilon)^q$,
Also if y is word such that $lg(y) = q$, $y \notin \alpha_2$ and $y = \sigma_{i_1}...\sigma_{i_q}$.
Then $s_j = (i_1, i_2, .., i_q) = \phi^{-1}(y) \in S - F$
So that for any $x \in \Sigma_2^*$ $p(xy) \leq\ 1 - p_j(s_1, xy) < 1 - (1-\epsilon)^q$.
Hence of $x \notin \alpha$,$p(x) < 1 - (1-\epsilon)^q$.
Now chose ϵ s.t $(1-\epsilon)^q > \frac{3}{4}$.For $x \in \alpha$ $p(x) > \frac{3}{4}$ and for $x \notin \alpha$ $p(x) < \frac{1}{4}$. Setting $\lambda = \frac{1}{2}$ we find λ is an isolated cut-point and it follows that $\alpha = \beta_{\breve{S}}(\lambda)$.(proved)

4.4 The stability theorem

Consider a probabilistic automata $\bar{\mathbb{S}}$ and an isolated cut-point λ. It is natural to ask whether the set $\beta_{\bar{S}}(\lambda)$ remains unchanged (stable) under small perturbations of the transition probabilities of $\bar{\mathbb{S}}$. Results along this line we shall call stability theorems.

Theorem 4.4. *Let $\breve{\mathbb{S}} =< S, M, s_0, F >$ be an actual automaton and λ be an isolated cut-point. There exists an $\epsilon > 0$ such that for every automaton $\breve{\mathbb{S}}' =< S, M', s_0, F >$with transition probabilities differing from those of by less than ϵ, λ is an isolated cut-point of $\breve{\mathbb{S}}'$ and $\beta_{\breve{S}}(\lambda) = \beta_{\breve{S}'}(\lambda)$.*

Chapter 5

Conclusion

Though the generalization from the abstract deterministic automata to the abstract probabilistic automata lies near at hand, there are no general results about probabilistic automata in the literature. In particular, it was not even known whether probabilistic automata can do more than deterministic automata. In this dissertation, we develop a general theory of probabilistic automata and answer some of the basic questions about them.

It turns out that, in general, probabilistic automata are stronger than deterministic automata. We introduce, however, a new concept of isolated cut-point and prove the fundamental **reduction theorem** that every probabilistic automata with isolated cut-point is equivalent to a suitable deterministic automaton.

Further, we define actual automata which are automata such that all their transition probabilities are strictly positive. These automata define a very limited class of regular events.

Bibliography

[1] Kemeny, J. G. AND SNELL, J. L. (1960), "Finite Markov Chains." Van Nostrand, New York • .

[2] PERLES, M., RABIN, M. O. AND SHAMIR, E. (1963) *The theory of definite automata* IRE Transactions on Computers.

[3] RABIN, M. O. AND SCOTT, D. (1959) *Finite automata and their decision problems* IBM J. Research Develop 3, 114-125.

[4] VON NEUMANN, J. (1956) Probabilistic logics and the synthesis of reliable organisms from unreliable components *Annals of Mathematics Studies,* Vol. 34, SHANNON C. E. AND McCARTHY, J., eds, pp. 43-98. Princeton Univ. Press, Princeton, New Jersey