Contributions to Random Iterations and Dynamical Systems

By

Santanu Chakraborty

Thesis submitted to the Indian Statistical Institute
in partial fulfilment of the requirements for the award
of the degree of Doctor of Philosophy

CALCUTTA

2000

Dedicated to

My father Late Satya Ranjan Chakraborty

Acknowledgements

First of all, I am thankful to my supervisor Professor B.V.Rao whose help and constant encouragement has made it possible for me to complete this work. Inspite of his busy schedule, he gave me so much time on a regular basis. He has been very affectionate and more like a father figure to me.

I thank Professors R.N.Bhattacharya and A.Mukherjea for their suggestions. They have also kindly provided me copies of their preprints related to the problems discussed in the thesis which were helpful. I thank Professors A.Goswami, J.C.Gupta, R.L.Karandikar, A.R.Rao, Probal Choudhury, Arindam sengupta, Ujjwal Bhattacharya and Sreela Gangopadhyay for their help at different stages of the work. I thank all my teachers at the Institute who shaped my thinking.

I am thankful to Professor S.B.Rao, Director of the Indian Statistical Institute and Professor S.C.Bagchi, Professor-in-charge of the Stat-Math Division for providing me all the facilities to carry out this work. I appreciate their offer of a visiting position at the Institute while I was on leave from the Reserve Bank of India . Their advice and encouragement are gratefully acknowledged.

I pay my tribute to my beloved mother, who shouldered all the household responsibilities so that I fully concentrate on my research work. She encouraged me throughout the period I was engaged in writing my thesis.

I am thankful to the Reserve Bank of India, especially to Dr. R.B. Burman, Principal Adviser, DESACS, R.B.I., Mumbai who granted me one year leave to complete the research. I thank Dr. D. Ray, Adviser, DESACS, R.B.I., Mr. P. Maria, Director, DESACS, R.B.I., Calcutta and Mr. S. Majumdar, Assistant Adviser, DESACS, R.B.I. for their encouragement.

I thank my friends Mr. Punam Kr. Saha, Mr. Shubhashis Roy, Mr. Mrinal Kanti Mukherjee, Mr. Parthasarathi Bhattacharya, Mr. Arnab Kr. Laha, Dr. Sujay Datta for the motivation and useful tips they provided me at several stages of this work.

I thank all my friends in the Stat-Math Division for their encouragement. Last but not the least, I thank the Stat-Math Division, the Dean's Office, the Director's Office, the Library and the Reprography units for the assistance that they provided me during my research period.

Contents

Preface

In recent years random iterations of maps on Polish spaces has gained prominence. They are nice examples of Markov processes whose invariant measures can be used in Computer imaging (see Berger [1]. They also arise as random perturabations of deterministic dynamical systems.

Let S be a Polish space with its Borel σ-field. Let Γ be a collection of Borel maps from S to S. Let P be a probability on Γ. Then starting with a point x in S, we choose a map $\gamma_1 \in \Gamma$ according to the law P and move to $\gamma_1(x)$. Then we choose $\gamma_2 \in \Gamma$ again according to the law P and move to $\gamma_2 \gamma_1(x)$ and so on. This gives rise to a Markov process with state space S. One is interested in the existence and uniqueness of invariant measures for the process. Also when there is a unique invariant measure, one is interested in the nature of invariant measure like its support, whether it is absolutely continuous with respect to another given probability etc. Eventhough there were earlier works, L.E. Dubins and D.A. Freedman [12] for the first time made a systematic study when the Polish space S is real line. A part of their work was generalized to higher dimensions by R.N. Bhattacharya and his co-authors. The problems dealt with in this thesis are either directly or indirectly connected with random iterations. The thesis has four chapters. Each chapter starts with a summary of its own. We briefly describe the main contents below.

In chapter I, we discuss the problem of completeness of a metric – introduced by R.N. Bhattacharya and O. Lee [3] – on the space of probabilities on $\mathbb{R}^k$. This metric was introduced by them in generalizing the works of Dubins and Freedman [12] regarding existence of invariant measures for Markov processes generated by random iterations of monotone maps. They obtained positive results bypassing the problem of completeness of the metric. They suggested that if the metric could be proved to be complete, then a fixed point theorem will make the arguments simpler. We carry out this programme. To generalize these results from $\mathbb{R}^k$ to appropriate subsets S of $\mathbb{R}^k$, it is necessary to know for which subsets S of $\mathbb{R}^k$, the class of probabilities on S, say, $\mathcal{P}(S)$ is complete under the metric. However, we do not know the full answer to this question.

In chapters II and III, we study a problem whose origins go back to the works of M. Rosenblatt [29]. Given a probability μ on S_d, the space of stochastic matrices of order d – which is a semigroup under multiplication – find conditions for the convolution sequence μ^n to converge. Several conditions in the general context of compact groups and semigroups were already available in Rosenblatt [29]. See A. Mukherjea and G. Hognas [17] for a thorough and uptodate treatment. The question however is to find some simply verifiable conditions on μ so that μ^n converges. When $d = 2$, this was treated by A. Mukherjea [22]. His theorem reads as follows. If μ is probability on the space of 2×2 stochastic matrices, then μ^n converges if and oly if μ is not the point mass at the matrix $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$.

In chapter II, we generalise this result to the case $d = 3$. The main theorem was also obtained independently using algebraic methods by S. Dhar and A. Mukherjea [11]. Roughly speaking, μ^n converges unless some kind of periodicities are present. When $d = 2$, and support of μ has at least one matrix with an entry different from 0 and 1, then the limit of μ^n (which exists) is concentrated on rank-one stochastic matrices. Thus we can regard the limiting measure as a probability on $[0,1]$ – identify $0 \leq c \leq 1$ with the matrix $\begin{bmatrix} c & 1-c \\ 1-c & c \end{bmatrix}$. It makes sense to ask for conditions under which this limiting probability is absolutely continuous or singular (w.r.t. Lebesgue measure). This problem was already raised by Rosenblatt [29] and some partial results were obtained by A. Mukherjea and his co-authors. We conclude chapter II by establishing a connection between a simple case of this problem and Bernoulli convolutions.

In chapter III, we give necessary and sufficient conditions for μ^n to converge when μ is a probability on S_d – the set of $d \times d$ stochastic matrices, generalizing results of chapter II. Here is the idea. Suppose $\mathcal{S}$ is the closed semigroup generated by the support of μ. A structure theorem for kernel of $\mathcal{S}$ was already outlined in Rosenblatt [29]. We extend this to obtain a structure theorem for $\mathcal{S}$ itself and obtain a map from $\mathcal{S}$ to an appropriate permutation group (on at most d symbols). With the help of this map, we transfer μ to a probability $\tilde{\mu}$ on this permutation group. Our main theorem says that μ^n converges if and only if $\tilde{\mu}^n$ converges.

In chapter IV, we study a problem whose origin is in the work of R.N. Bhattacharya and A. Goswami [2]. They considered the following problem. Let X_0 be a strictly positive random variable and let $(Z_n)_{n\geq 1}$ be an i.i.d. sequence of random variables independent of X_0 and each taking only two values 0 and θ. Define a Markov process $(X_n)_{n\geq 1}$ by $X_{n+1} = Z_{n+1} + \dfrac{1}{X_n}$ for $n \geq 0$. By using the Gauss map on $[0,1)$ and its properties, they showed that when $\theta = 1$, the unique invariant measure is singular. We generalize Gauss map as follows. Fix $0 < \theta \leq 1$. Define $T : [0,\theta) \mapsto [0,\theta)$ by $T(x) = \theta(\frac{1}{\theta x} - [\frac{1}{\theta x}])$ for $x > 0$ and $T(0) = 0$. We were unable to see if it is conjugate to the Gauss map. We study this map and show that for several values of θ, it admits an absolutely continuous invariant probability. Moreover, like the Gauss map, the successive averages of almost all orbits diverge to infinity. We do hope that these results yield some information about the Markov process mentioned above. We conclude this chapter with an alternative proof of the theorem of Bhattacharya and Goswami mentioned earlier.

CHAPTER - I

Bhattacharya metric on the space of probabilities

Summary.

This chapter has five sections. In section 1, we start with a brief introduction to random iterations and after recalling relevant definitions, we introduce the Bhattacharya metric d_1 on the space of probabilities on $\mathbb{R}^k$. In sections 2 and 3, we discuss the completeness of the metric d_1 for the cases $k = 1$ and $k > 1$ respectively. Section 4 gives an application of the result proved in section 3 by recalling an argument of Bhattacharya and Lee on the existence of invariant measures. We conclude with some interesting remarks in section 5.

Section 1 : Introduction

Consider the closed unit interval $[0,1]$. Suppose Γ is the collection of continuous monotone non-decreasing functions of the interval to itself. Suppose P is a probability on the Borel σ-field of Γ. This gives rise to a Markov Process (X_n) with state space $[0,1]$ as follows : If we are at x, we select a $\gamma \in \Gamma$ according to the law P and move to $\gamma(x)$. Suppose there is an x_0 and an integer $m \geq 1$ such that

$$P^m\{(\gamma_1, \cdots, \gamma_m) : Range(\gamma_m \cdots \gamma_1) \subset [0,x] > 0\}$$

and

$$P^m\{(\gamma_1, \cdots, \gamma_m) : Range(\gamma_m \cdots \gamma_1) \subset [x,1] > 0\}$$

Then, Dubins and Freedman [12] showed that the Markov Process has a unique invariant distribution. They named this condition "splitting". Such systems as these are now a days called iterated function systems. We must add that though Dubins and Freedman [12] made a systematic analysis, they were not the first to consider these problems. Motivated by problems in the theory of learning, R.R. Bush and F. Mosteller; S.Karlin and others considered such systems earlier (We are not intending to survey this vast area and give only references relevant to us). J. Yahav [36] considered this problem by removing the restriction of continuity of the maps. However, both Dubins-Freedman and Yahav treated only the case of compact subintervals of the real line - more precisely, the underlying functions of the system are defined on a fixed closed bounded subinterval of the real line.

Motivated by problems in Time Series and Economic models, Bhattacharya and Lee [3] provided a generalization of this set up to higher dimensions. They considered a Borel subset $S \subset \mathbb{R}^k$. For points $x, y \in S$, say that $x \leq y$ if the inequality holds for every co-ordinate. Say that a Borel function $f : S \mapsto S$ is monotone non-decreasing if $f(x) \leq f(y)$ whenever $x \leq y$. Let Γ be a collection of such maps. Assume that Γ has a σ-field $\mathcal{F}$ such that the evaluation map $(\gamma, x) \mapsto \gamma(x)$ is a jointly measurable map. Suppose that P is a probability on $\mathcal{F}$. This gives rise, as earlier, to a Markov Process with state space S as follows : If we are at x, select a $\gamma \in \Gamma$ according to P and move to $\gamma(x)$. Among other things, Bhattacharya and Lee [3] considered in this set up the existence of invariant measures for this Markov Process. We shall return to this in section 4. For now, it suffices to say that in this context, they were lead to introduce the following metric on $\mathcal{P}(\mathbb{R}^k)$, the space of probabilities on the Borel σ-field of $\mathbb{R}^k$. Say that $f : \mathbb{R}^k \mapsto \mathbb{R}$ is monotone non-decreasing if $f(x) \leq f(y)$ whenever $x \leq y$.

Also, we let $\mathcal{G}_1$ denote the class of all non-decreasing maps on $\mathbb{R}^k$ to the unit interval [0, 1]. We define Bhattacharya metric d_1 on $\mathcal{P}^k$ as follows :

$$d_1(\mu, \nu) = sup\{\mid \int f d\mu - \int f d\nu \mid : f \in \mathcal{G}_1\}$$

Section 2 : Case $k = 1$

For a probability μ on $(\mathbb{R}, \mathcal{B})$, let $F_\mu(x)$ denote the distribution function of μ. Define the usual supremum metric

$$\rho(\mu, \nu) = \sup_x \mid F_\mu(x) - F_\nu(x) \mid .$$

We make a series of observations.

1°. $\rho(\mu, \nu) \leq d_1(\mu, \nu)$.

This is obvious by noting that for each x, the indicator function $1_{(x,\infty)}$ belongs to $\mathcal{G}$.

$\underline{2}^o$. If $\mu_n \xrightarrow{\rho} \mu$ then $\mu_n \xrightarrow{d_1} \mu$.

Indeed, let $\mu_n \xrightarrow{\rho} \mu$ and $\epsilon > 0$. Choose an integer k such that $\frac{k-1}{8}\epsilon < 1 < \frac{k}{8}\epsilon$. Choose integer N such that for each $n \geq N$ and for each x, $| F_{\mu_n}(x) - F_\mu(x) | < \frac{\epsilon}{16k}$. This in turn implies that for any interval J (open, closed or semi open), $| \mu_n(J) - \mu(J) | < \frac{\epsilon}{8k}$. Let now $f \in \mathcal{G}$. For $i = 0, 1, 2, \cdots, J_i$ be the interval $f^{-1}[\frac{i}{8}\epsilon, \frac{i+1}{8}\epsilon)$ and g be the function defined by $g(x) = f(\frac{i}{8}\epsilon)$ for $x \in J_i$. Then direct calculation shows that $| \int g d\mu_n - \int g d\mu | \leq \frac{\epsilon}{8}$ for $n \geq N$. Since $| \int f d\mu_n - \int g d\mu_n | < \frac{\epsilon}{8}$ and $| \int f d\mu - \int g d\mu | < \frac{\epsilon}{8}$. We conclude that $| \int f d\mu_n - \int f d\mu | < \epsilon$ for each $n \geq N$. This being true for any $f \in \mathcal{G}$ it follows that $d_1(\mu_n, \mu) < \epsilon$ for $n \geq N$, completing the proof.

$\underline{3}^o$. $\mathcal{P}^1$ is complete under ρ.

This is clear.

$\underline{4}^o$. $\mathcal{P}^1$ is complete under d_1.

In fact, if $\{\mu_n\}$ is d_1 Cauchy then 1^o implies that it is ρ Cauchy and hence by 3^o there is a $\mu \in \mathcal{P}^1$ such that $\mu_n \xrightarrow{\rho} \mu$. 2^o implies that $\mu_n \xrightarrow{d_1} \mu$.

Now consider any Borel subset $S \subset I\!R$ with its Borel σ field. Let $\mathcal{P}(S)$ be the collection of all probabilities on S. Since $\mathcal{P}(S) \subset \mathcal{P}^1$ (with abuse of notation) we can restrict d_1 and ρ to $\mathcal{P}(S)$ and still denote them by d_1 and ρ respectively. We can also define for $\mu, \nu \in \mathcal{P}(S)$.

$\overline{d_1}(\mu, \nu) = \sup\{| \int f d\mu - \int f d\nu | : f \uparrow, \text{ Borel }, \text{on } S \text{ to } [0,1]\}$

$\overline{\rho}(\mu, \nu) = \sup_{x \in S} | F_\mu(x) - F_\nu(x) |$.

Then, it is easy to see that

$\underline{5}^o$. For $\mu, \nu \in \mathcal{P}(S)$, $\overline{\rho}(\mu,\nu) = \rho(\mu,\nu)$ and $\overline{d_1}(\mu,\nu) = d_1(\mu,\nu)$.

$\underline{6}^o$. $\mathcal{P}(S)$ is complete under d_1 iff $\overline{S} - S$ is at most a countable set.

To see this first assume that $\overline{S} - S$ is a countable set. Let $\mu_n \in \mathcal{P}(S)$ for $n \geq 1$. If $\{\mu_n\}$ is Cauchy under d_1 then there is a $\mu \in \mathcal{P}^1$ so that $\mu_n \xrightarrow{d_1} \mu$ in $\mathcal{P}^1$. By 1^o, $\mu_n \xrightarrow{\rho} \mu$ and in particular $\mu_n \to \mu$ weakly. This implies that $\mu(\overline{S}) = 1$, since $\mu_n(S) = 1$ for each n. Now $\mu_n \xrightarrow{\rho} \mu$ implies that for each $x \in R$, $\mu_n\{x\} \to \mu\{x\}$. As a consequence $\mu\{x\} = 0$ for each $x \in \overline{S} - S$. We conclude that $\mu(\overline{S} - S) = 0$. In other words $\mu \in \mathcal{P}(S)$ and $\mu_n \xrightarrow{d_1} \mu$.

To prove the converse, let $\overline{S} = S$ be uncountable and hence we can fix a homeomorphism h from the coin tossing space $C = \{0,1\}^\omega$ into $\overline{S} - S$. Fix $n \geq 1$. Let s_n be the set of all sequences of 0's and 1's of length n. For $s \in s_n$, let U_s be the set of all points in C whose initial segment is s. For each $s \in s_n$ we can fix an open set V_s so that

$$h(U_s) \subset V_s \subset \text{ Ball of radius } \frac{1}{n} \text{ around } h(U_s).$$

Moreover $\{h(U_s) : s \in s_n\}$ being disjoint compact sets we can assume that $\{V_s, s \in s_n\}$ are disjoint. As $h(U_s) \subset \overline{S}$ for each $s \in s_n$ we can fix a point $\zeta_s \in S \cap V_s$. Let μ_n be the probability putting mass $\frac{1}{2^n}$ at ζ_s for $s \in s_n$. Let λ be the fair coin tossing measure on C and $\mu = \lambda h^{-1}$. Then it is not difficult to show that in $\mathcal{P}^1$, $\mu_n \xrightarrow{d_1} \mu$. We conclude that $\{\mu_n\}$ is d_1 Cauchy in $\mathcal{P}(S)$ but does not converge to an element of $\mathcal{P}(S)$. This shows that $\mathcal{P}(S)$ is not complete, as required.

For example if S is closed or if S is a finite disjoint union of open intervals then $\mathcal{P}(S)$ is complete under d_1. However, if S is the set of rationals (which is a countable union of closed sets) or if S is the complement of the Cantor Set in the unit interval (which is a countable union of open intervals) $\mathcal{P}(S)$ is not complete (cf. Remark 2.5.1 of [3]).

Section 3 : Case $k > 1$

We consider only the case $k = 2$. This is done for two reasons. Firstly, for notational convenience – the same idea works for $k > 2$ as well. Secondly, there is perhaps a simpler proof which we are missing and there is no point in burdening the reader with the details for the general case. Accordingly, we denote by $\mathcal{P}$ the set of probabilities on $(\mathbb{R}^2, \mathcal{B}^2)$. For $\mu \in \mathcal{P}$ we denote its distribution function by $F_\mu(x,y)$ and as in the case of $k=1$ we can define

$$\rho(\mu,\nu) = \sup_{(x,y)\in\mathbb{R}^2} \mid F_\mu(x,y) - F_\nu(x,y) \mid .$$

As in the case of $k = 1$, we have,

<u>1</u>^{o}. $\rho(\mu,\nu) \leq d_1(\mu,\nu)$.

However, unlike the previous case,

<u>2</u>^{o}. ρ and d_1 are not equivalent.

Indeed if μ_n is the Lebesgue measure on the line segment $\{(x+\frac{1}{n}, -x) : 0 \leq x \leq 1\}$ and μ is the Lebesgue measure on $\{(x,-x) : 0 \leq x \leq 1\}$ then it is easy to see that $\rho(\mu_n,\mu) < \frac{1}{n}$ so that $\mu_n \xrightarrow{\rho} \mu$. However $d_1(\mu_n,\mu_m) = 1$ for $n \neq m$.

<u>3</u>^{o}. If μ_n is d_1 Cauchy then there is a unique μ so that $\mu_n \xrightarrow{\rho} \mu$.

In fact, 1^o implies that μ_n is ρ Cauchy and hence F_{μ_n} converges in supremum metric to some F, which is necessarily F_μ for some $\mu \in \mathcal{P}$.

For the rest of the section we fix a d_1-Cauchy sequence $\{\mu_n\}$. Let μ be as above. We shall show that $\mu_n \xrightarrow{d_1} \mu$, there by showing that $\mathcal{P}$ is complete with the metric d_1.

We start with a definition.
<u>Definition</u> : A Borel Set $L \subset \mathbb{R}^2$ is a *Left set* if $(x,y) \in L$ and $(x',y') \leq (x,y)$ then $(x',y') \in L$.
Here is a reduction of the problem.

<u>4</u>^{o}. If $\mu_n(L) \to \mu(L)$ for each left set L then $\mu_n \xrightarrow{d_1} \mu$.

To see this, fix any increasing Borel measurable f on $\mathbb{R}^2$ to [0,1]. Let $\epsilon > 0$. Then the set $L_j = F^{-1}[j\epsilon, (j+1)\epsilon)$ is the difference of two left sets so that $\mu_n(L_j) \to \mu(L_j)$. Let g be the simple function $g(x, y) = j\epsilon$ for $(x, y) \in L_j$ for each j. The previous observation implies that $\int g d\mu_n \to \int g d\mu$. Since for each $(x, y), \mid f(x, y) - g(x, y) \mid < \epsilon$ we conclude in a routine way that $\mid \int f d\mu_n - \int f d\mu \mid < 3\epsilon$ for sufficiently large n. This shows that $\int f d\mu_n \to \int f d\mu$ for each $f \in \mathcal{G}^2$. To show $\mu_n \xrightarrow{d_1} \mu$ we still have to establish that the above convergence holds uniformly in $f \in \mathcal{G}^2$. But this is a standard argument as follows : Fix $\epsilon > 0$. Since $\{\mu_n\}$ is d_1-Cauchy, fix N such that for $n, m \geq N$ $d_1(\mu_n, \mu_m) \leq \epsilon$. For any $f \in \mathcal{G}^2$ and $n, m \geq N$ we have $\mid \int f d\mu_n - \int f d\mu_m \mid \leq \epsilon$. Letting $m \to \infty$, and using the fact proved above we conclude that $\mid \int f d\mu_n - \int f d\mu \mid \leq \epsilon$ for each $n \geq N$ and for each $f \in \mathcal{G}^2$. In other words $d_1(\mu_n, \mu) \leq \epsilon$ for $n \geq N$.

To describe left sets, we need a definition.

Definition : Let ϕ be nonincreasing function on $\mathbb{R}$ taking values in $[-\infty, \infty]$. Put

$$G_\phi^- = \{(x, y) \in \mathbb{R}^2 : y < \phi(x+)\}$$

$$G_\phi^+ = \{(x, y) \in \mathbb{R}^2 : y > \phi(x-)\}$$

$$\text{and , } B_\phi = \{(x, y) \in \mathbb{R}^2 : \phi(x+) \leq y \leq \phi(x-)\}$$

A Borel subset $A \subset B_\phi$ will be called a **boundary set** if
$(x, y) \in A$, $(x, y') \in B_\phi$, $y' \leq y$ implies $(x, y') \in A$;
and
$(x, y) \in A$, $(x', y) \in B_\phi$, $x' \leq x$ implies $(x', y) \in A$.

5°. Let ϕ be a nonincreasing function on $\mathbb{R}$ to $[-\infty, \infty]$. Then, the following facts are easy to verify :-

1. G_ϕ^- is a left set and an open set, i.e., an open left set.
2. $G_\phi^- \cup B_\phi$ is a left set and a closed set i.e. a closed left set. Indeed the closure of G_ϕ^- is $G_\phi^- \cup B_\phi$.
3. If A is a boundary set contained in B_ϕ then $G_\phi^- \cup A$ is a left set.

We also have,

6°. Let L be a left set. Then there is a nonincreasing function ϕ on $\mathbb{R}$ to $[-\infty, \infty]$ and a boundary set $A \subset B_\phi$ such that $L = G_\phi^- \cup A$.

In fact $\phi(x) = \sup\{y : (x, y) \in L\}$ will serve the purpose.

$\underline{7}^o$. If L is an open left set then there are open left sets L_k such that $L_k \uparrow L$ and $\mu(\partial L_k) = 0$ for each k.

To see this fix an integer $k \geq 1$. Fix increasing sequences $\{x_n : -\infty < n < \infty\}$, $\{y_n : -\infty < n < \infty\}$ so that $\mu(\{x_n\} \times I\!R) = 0$, $\mu(I\!R \times \{y_n\}) = 0$, $\mid x_n - x_{n+1} \mid < \frac{1}{2^k}$ and $\mid y_n - y_{n+1} \mid < \frac{1}{2^k}$ for each n. Consider the tilling of the plane given by the rectangles with corners (x_i, y_j); $-\infty < i < \infty$, $-\infty < j < \infty$. Let L_k be the union of these closed rectangles which are contained in L. Then L_k is a left set, ∂L_k consists of vertical segments at $x_n s$ and horizontal segments at $y_n s$ so that $\mu(\partial L_k) = 0$. If we moreover take the k^{th} partition to be refinement of the corresponding $(k-1)^{th}$ partition then it is easy to see that $L_k \uparrow L$ completing the proof.

$\underline{8}^o$. If L is an open left set then $\mu_n(L) \to \mu(L)$.

In fact, fix L_k as in 7^o. Fix an $\epsilon > 0$. Let N be such that for $n, m \geq N$, $d_1(\mu_n, \mu_m) \leq \epsilon$. In particular for each k, $\mid \mu_n(L_k) - \mu_m(L_k) \mid \leq \epsilon$ for $n, m \geq N$ (Note that indicator of the complement of L_k is in $\mathcal{G}^2$). Since $\mu_m \to \mu$ weakly and $\mu(\partial L_k) = 0$ we conclude that $\mid \mu_n(L_k) - \mu(L_k) \mid \leq \epsilon$ for $n \geq N$. This being true for each k, we let $k \to \infty$ to see that $\mid \mu_n(L) - \mu(L) \mid \leq \epsilon$ for each $n \geq N$. This completes the proof.

$\underline{9}^o$. If M is a closed left set then there are closed left sets M_k such that $M_k \downarrow M$ and $\mu(\partial M_k) = 0$ for each k.

This is done by taking a tilling of $I\!R^2$ as in 7^o and taking M_k to be all those rectangles that intersect M. However M_k is not by itself a left set. By adjoining some more rectangles we get a closed left set M_k. We leave the details to the reader.

$\underline{10}^o$. If M is a closed left set then $\mu_n(M) \to \mu(M)$.

This is a consequence of 9^o, just as 8^o was a consequence of 7^o.

$\underline{11^o}$. Let ϕ be a nonincreasing function on $I\!R$ to $[-\infty, \infty]$. Then $\mu_n(G_\phi^-) \to \mu(G_\phi^-)$; $\mu_n(G_\phi^+) \to \mu(G_\phi^+)$; $\mu_n(B_\phi) \to \mu(B_\phi)$.

This is a consequence of 5^o, 8^o and 10^o.

$\underline{12^o}$. For any vertical line segment V, $\mu_n(V) \to \mu(V)$. V may be finite or infinite; open, semi open or closed.

For example let V be the vertical segment $(x = x_o, a < y \leq b)$. Fix $x_k < x_o$, $x_k \uparrow x_o$ and $a_k > a$, $a_k \uparrow a$. Define the sets $A_k = (x \leq x_o, y \leq b)$; $B_k = (x \leq x_o, y \leq a_k)$; $C_k = (x \leq x_k, y \leq b)$; $D_k = (x \leq x_k, y \leq a_k)$; $E_k = (x_k < x \leq x_o, a_k < y \leq b)$. Then clearly $\mu_n(E_k) = \mu_n(A_k) - \mu_n(B_k) - \mu_n(C_k) + \mu_n(D_k)$ with a similar equation holding for μ. Now fix $\epsilon > 0$ and choose N so that for $n, m \geq N$, $d_1(\mu_n, \mu_m) < \epsilon/4$. As a consequence for $n, m \geq N, \mid \mu_n(E_k) - \mu_m(E_k) \mid < \epsilon$. Since A_k, B_k, C_k, D_k are closed left sets, 10^o implies that $\mu_m(E_k) \to \mu(E_k)$ as $m \to \infty$. This implies that for any $k \geq 1$ and $n \geq N, \mid \mu_n(E_k) - \mu(E_k) \mid \leq \epsilon$. Taking limit over k we observe that $\mid \mu_n(V) - \mu(V) \mid \leq \epsilon$ for $n \geq N$, completing the proof. Other cases of vertical segments are treated in the same way. One could also use the fact that F_{μ_n} converges to F_μ uniformly to give an alternative proof.

Exactly as above we can show the following :

$\underline{13^o}$. For any horizontal line segment H, $\mu_n(H) \to \mu(H)$.

From now on we fix a nonincreasing function ϕ on $I\!R$ to $[-\infty, \infty]$.

$\underline{14^o}$. Let K be a compact boundary set contained in B_ϕ. Then $\mu_n(K) \to \mu(K)$.

In order to show this, fix an integer $k \geq 1$. As in the proof of 7^o fix increasing sequences, $\{x_i, -\infty < i < \infty\}$ $\{y_j, -\infty < j < \infty\}$ and R_{ij} be the rectangle $[x_i, x_{i+1}] \times [y_j, y_{j+1}]$. Let L'_k be the union of all those rectangles which are either contained in G_ϕ^- or which intersect K. However L'_k may not already be a left set, we add enough of these rectangles to get a left set L_k. The following properties of the sets L_k can easily be verified : (a) If $(x, y) \in G_\phi^-$ then $(x, y) \in L_k$ for all large k. (b) If $(x, y) \in K$ then $(x, y) \in L_k$ for all k. (c) If $(x, y) \notin G_\phi^- \cup K$ then for sufficiently large k, $(x, y) \notin L_k$.

Indeed (a) is a consequence of the fact that G_ϕ^- is open. (b) is by construction. (c) is observed as follows : If $(x,y) \notin G_\phi^- \cup K$ then either $(x,y) \in G_\phi^+$ or $(x,y) \in B_\phi - K$. In the first case G_ϕ^+ being open, for sufficiently large k, one of the rectangles $R_{ij} \subset G_\phi^+$ and includes the point (x,y). As a consequence this rectangle does not appear in the formation of L_k'. Since $G_\phi^- \cup B_\phi$ is a left set none of the rectangles above R_{ij} appear in the formation of L_k. Thus R_{ij} does not appear in the formation of L_k as well showing that $(x,y) \notin L_k$ for sufficiently large k. Assume now that $(x,y) \in B_\phi - K$. K being compact for sufficiently large k one of the rectangles R_{ij} includes the point (x,y) and is disjoint with K. Clearly R_{ij} can not appear in the formation of L_k'. If k is so large that length of the sides of the rectangle is smaller than $x - \sup\{x' < x : \text{ for some } y', \ (x',y') \in K\}$ then none of the rectangles above R_{ij} can appear in the formation of L_k so that it can not appear in the formation of L_k either. Now fix $\epsilon > o$. Fix N so large that for $n,m \geq N$; $d(\mu_n,\mu_m) < \epsilon$. In particular for each k and $m,n \geq N$; $| \mu_n(L_k) - \mu_m(L_k) | < \epsilon$. L_k being a closed left set 10^o implies that $| \mu_n(L_k) - \mu(L_k) | \leq \epsilon$ for each k and $n \geq N$. In view of the properties (a), (b), (c) observed above, taking limit over k. We get that $| \mu_n(G_\phi^- \cup K) - \mu(G_\phi^- \cup K) | \leq \epsilon$ for each $n \geq N$. But by 11^o, $\mu_n(G_\phi^-) \to \mu(G_\phi^-)$ and G_ϕ^- and K are disjoint. These imply that for sufficiently large n, $| \mu_n(K) - \mu(K) | < 2\epsilon$. This yields the result.

Note that $B_\phi = L \cup M$ where L is the union of closed nondegenerate Horizontal and Vertical segments contained in B_ϕ and $M = B_\phi - L$. Clearly these are Borel sets. A moment's reflection shows that every Borel subset A of M is a boundary set.

15^o. For any Borel set $A \subset M$, $\underline{\lim}\ \mu_n(A) \geq \mu(A)$.

Indeed, for any compact set $K \subset A$, $\mu_n(A) \geq \mu_n(K)$ and hence by 14^o, $\underline{\lim}\ \mu_n(A) \geq \underline{\lim}\ \mu_n(K) = \mu(K)$. This being true for all compact subsets K of A we deduce the result. In particular we have,

$\underline{16}^o$. $\underline{\lim}\ \mu_n(M) \geq \mu(M)$.

$\underline{17}^o$. $\underline{\lim}\ \mu_n(L) \geq \mu(L)$.

Indeed, L is a countable union of horizontal and vertical segments and for the union F of finitely many of them, we have in view of 12^o and 13^o, $\underline{\lim}\ \mu_n(L) \geq \underline{\lim}\ \mu_n(F) = \mu(F)$. Consequently we have $\underline{\lim}\ \mu_n(L) \geq \mu(L)$ as required.

<u>18</u>^{o}. $\mu_n(L) \to \mu(L)$ and $\mu_n(M) \to \mu(M)$.

This follows from 16^o and 17^o and the fact that by 11^o, $\mu_n(B_\phi) \to \mu(B_\phi)$.

<u>19</u>^{o}. For any Borel set $A \subset M$, $\mu_n(A) \to \mu(A)$.

Indeed, as remarked earlier any Borel set $A \subset M$ is a boundary set and hence by 15^o, $\underline{\lim}\ \mu_n(A) \geq \mu(A)$ and $\underline{\lim}\ \mu_n(M - A) \geq \mu(M - A)$. But by 18^o, $\mu_n(M) \to \mu(M)$. This shows that $\mu_n(A) \to \mu(A)$.

<u>20</u>^{o}. If $A \subset L$ is a boundary set then $\mu_n(A) \to \mu(A)$.

In fact, such a set A is a union of countably many horizontal and vertical segments and so is $L - A$. Using arguments as in 17^o and 19^o the proof is completed.

<u>21</u>^{o}. For any boundary set $A \subset B_\phi$, $\mu_n(A) \to \mu(A)$.

This is because any such set A is a union of two disjoint sets $A_1 \subset L$ and $A_2 \subset M$ where A_1 is a boundary set. 19^o applies for A_2 and 20^o applies for A_1 to complete the proof.

<u>22</u>^{o}. For any Left set $L, \mu_n(L) \to \mu(L)$.

This is immediate from 6^o, 11^o and 21^o.

Combining 3^o, 4^o and 22^o we finally obtain,

Theorem : $(\mathcal{P}^2, d_1)$ is a complete metric space.

Section 4 : An application to invariant measures

In this section, we shall briefly reproduce an argument of Bhattacharya and Lee [3] to obtain invariant measures for certain iterated function systems, using the theorem of previous section.

Let S be a Borel subset of $\mathbb{R}^k$ and let $\mathcal{S}$ be its Borel σ-field. Let Γ be a family of non-decreasing maps from S into itself. Let $\mathcal{F}$ be a σ-field on Γ making the evaluation map $(\gamma, x) \mapsto \gamma(x)$ jointly measuarble. Let P be a probability on Γ. Consider $(\Gamma^\infty, \mathcal{F}^\infty, P^\infty)$. Here is the Markov process X_n with state space S defined on Γ^∞.

$$X_1 = \alpha_1 X_0,$$
$$X_2 = \alpha_2\alpha_1 X_0,$$
$$\cdots\cdots$$
$$\cdots\cdots$$
$$\cdots\cdots$$
$$X_n = \alpha_n X_{n-1} = \alpha_n\alpha_{n-1}\cdots\alpha_1 X_0.$$

where $\alpha_1,\ \alpha_2,\ \cdots$ is a sequence of i.i.d. random maps with common distribution P and X_0 is a random variable independent of this sequence.

Then clearly, the transition function of this process is given by ,

$$p(x, B) = P(\gamma \in \Gamma : \gamma(x) \in B) = p(x, B) \quad \forall x \in S \quad \text{and} \quad \forall B \in \mathcal{S}$$

Then p defines the usual operator T^* on $\mathcal{P}(S)$ by

$$(T^*\mu)(B) = \int p(x, B)\mu(dx), \quad \mu \in \mathcal{P}(S)$$
$$= \int\int I_A(\gamma(x))dP(\gamma)\mu(dx).$$

so that for any bounded measurable function f on S,

$$\int f(x)T^*\mu(dx) = \int\int f(\gamma(x))dP(\gamma)\mu(dx).$$

Note that $f \in \mathcal{G}_1$ so that $\int f(\gamma(x))dP(\gamma)$ also belongs to $\mathcal{G}_1$. This immediately leads to the following :

$$d_1(T^*\mu, T^*\nu) \leq d_1(\mu, \nu) \quad \forall \mu, \nu \in \mathcal{P}(S) \tag{4.1}$$

Let $x_0 \in S$ and fix a positive integer m.
Let

$$\Gamma_1 = \{(\gamma_1, \cdots, \gamma_m) \in \Gamma^m : \gamma_m \cdots \gamma_1 x \leq x_0 \quad \forall x\} \tag{4.2}$$

$$\Gamma_2 = \{(\gamma_1, \cdots, \gamma_m) \in \Gamma^m : \gamma_m \cdots \gamma_1 x \geq x_0 \quad \forall x\} \tag{4.3}$$

Let $\delta = \max\{1 - P^m(\Gamma_1), 1 - P^m(\Gamma_2)\}$
Then, we have

Theorem [3] : With the above notation,

$$d_1(T^{*n}\mu, T^{*n}\nu) \leq \delta^{[n/m]} d_1(\mu, \nu) \quad \forall \mu, \nu \in \mathcal{P}(S) \tag{4.4}$$

<u>Proof</u>:
Let $f \in \mathcal{G}_1$. Note that,

$$\int f(x) dT^{*m}\mu = \int h(x) d\mu(x).$$

where

$$h(x) = \int f(\gamma_m \cdots \gamma_1 x) P^m(d\gamma_1 \cdots d\gamma_m)$$

Let h_1, h_2, h_3, h_4 be the functions defined like h but the integral on the right hand side is taken over $\Gamma_1 - \Gamma_1 \cap \Gamma_2$, $\Gamma_2 - \Gamma_1 \cap \Gamma_2$, $\Gamma - \Gamma_1 \cup \Gamma_2$ and $\Gamma_1 \cap \Gamma_2$ respectively. Clearly, these are non-negative functions, monotone non-decreasing and their sum equals h. Set

$$a_1 = P^m(\Gamma_1) - P^m(\Gamma_1 \cap \Gamma_2)$$

$$a_2 = P^m(\Gamma_2) - P^m(\Gamma_1 \cap \Gamma_2)$$

and

$$a_3 = 1 - P^m(\Gamma_1 \cup \Gamma_2)$$

Clearly, $h_1 \leq a_1 f(x_0)$ so that $\frac{h_1}{a_1 f(x_0)} \in \mathcal{G}_1$. Suimilarly, $\frac{a_2 - h_2}{a_2[1 - f(x_0)]}$ and $\frac{h_3}{a_3}$ are also in $\mathcal{G}_1$. Clearly, h_4 is a constant. Thus,

$$
\begin{aligned}
&| \int h d\mu - \int h d\nu | \\
&\leq | \int h_1 d\mu - \int h_1 d\nu | + | \int (a_2 - h_2) d\mu - \int (a_2 - h_2) d\nu | \\
&\qquad\qquad + | \int h_3 d\mu - \int h_3 d\nu | \\
&\leq [a_1 f(x_0) + a_2(1 - f(x_0)) + a_3] \; d_1(\mu, \nu) \\
&\leq (| a_1 - a_2 | + a_2 + a_3) d_1(\mu, \nu) \; \leq \; \delta \; d_1(\mu, \nu)
\end{aligned}
$$

That is,

$$| \int f dT^{*m}\mu - \int f dT^{*m}\nu | \leq \; \delta_1 \, d_1(\mu, \nu)$$

This being true for all $f \in \mathcal{G}_1$, $\quad d_1(T^{*m}\mu, T^{*m}\nu) \; \leq \; \delta \, d_1(\mu, \nu)$

From (4.1), $\quad d_1(T^*\mu, T^*\nu) \; \leq \; d_1(\mu, \nu)$

Combining these two inequalities, the result is immediate.

Now assume that P satisfies the following condition :

(4.6) $\qquad \exists x_0 \quad \text{and} \quad m \quad \text{such that} \quad P^m(\Gamma_1) > 0, \; P^m(\Gamma_2) > 0$

where Γ_1, Γ_2 are as defined in (4.2) and (4.3).

Theorem : . Under (4.6), T^{*m} is a strict contraction on $\mathcal{P}(S)$ with the metric d_1.

Proof. Use (4.4) and the fact that (4.6) implies $\delta < 1$.

If we now assume that S is such that $\mathcal{P}(S)$ is complete then the contraction mapping principle gives us a unique $\pi \in \mathcal{P}(S)$ such that $T^{*m}\pi = \pi$. We can use (4.4) to show that $T^*\pi = \pi$. Indeed, taking $\nu = \pi$ and $\mu = T^*\pi$ in

(4.4), we get

$$d_1(T^{*mn}\mu, T^{*mn}\nu) = d_1(T^{*(mn+1)}\pi, \pi) \longrightarrow 0$$

Similarly, $d_1(T^{*(mn+k)}\pi, \pi) \longrightarrow 0$ for each $k = 1, 2, \cdots, m-1$. This is enough to conclude that $T^{*n}\pi \longrightarrow \pi$. In particular, $T^*\pi = \pi$. Actually, for any μ, $T^{*n}\mu$ converges to π. Moreover, π is the unique invariant probability for the Markov Process.

Thus the Markov process X_n has a unique invariant measure. In particular, when $S = I\!R^k$, $\mathcal{P}(S)$ is complete by the theorem of the previous section and hence the earlier argument applies. Some more cases of sets S for which $\mathcal{P}(S)$ is complete are given in the next section.

Section 5 : Concluding Remarks

We conclude this chapter with a few remarks.

Remark 1. Clearly, $(\mathcal{P}(I\!R^k), d_1)$ is not a separable metric space. Indeed, d_1 restricted to the set of point masses gives rise to the discrete topology. As is well known, under weak convergence, $\mathcal{P}(I\!R^k)$ is separable.

Remark 2. If we have S_1 and S_2 so that $\mathcal{P}(S_1)$ and $\mathcal{P}(S_2)$, both are complete with respect to d_1, one can argue out that $\mathcal{P}(S_1 \cap S_2)$ is complete with respect to d_1. Same is true for countable intersections.

Remark 3. We do not have a characterization of Borel sets $S \subset I\!R^2$ for which $\mathcal{P}(S) \subset \mathcal{P}^2$ is complete under the metric, d_1. If S is a closed set then clearly $\mathcal{P}(S)$ is closed in $\mathcal{P}^2$ and is hence complete. If S is a rectangle, open or semiopen, with sides parallel to the axes, then also, $\mathcal{P}(S)$ is complete. Also, if S is a left set, then $(\mathcal{P}(S), d_1)$ is complete. This is because, whenever $\mu_n \xrightarrow{d_1} \mu$, we have $\mu_n(S) \to \mu(S)$.

Remark 4. Even, for some half spaces $\mathcal{P}(S)$ is not complete. This can be seen as follows :

Let us take $S \subset I\!R^2$ to be the region $x > y$ in the $I\!R^2$ plane. Let μ_n to be the linear Lebesgue measure on the line joining the points $(0, -\frac{1}{n})$ and $(1, 1-\frac{1}{n})$ for each $n \geq 1$. Let μ be the linear Lebesgue measure on the line joining the points $(0,0)$ and $(1,1)$.

Then we claim that $\mu_n \xrightarrow{d_1} \mu \in \mathcal{P}^2$. For this, let $\epsilon > 0$. Choose N so large that $n > N$ implies distance between the points $(x, x - \frac{1}{n})$ and $(x + \frac{1}{n}, x)$, is less than ϵ for any x in $[0, 1]$, that is, $\frac{\sqrt{2}}{n}$ is less than ϵ. Take any left set L. Then if x_0 is the largest number with $0 \leq x_0 \leq 1$ so that (x_0, x_0) is in the support of μ, we should have $| \mu_n(L) - \mu(L) | \leq \frac{\sqrt{2}}{n}$ (which is the distance between the points $(x_0, x_0 - \frac{1}{n})$ and $(x_0 + \frac{1}{n}, x_0)) < \epsilon$. Thus, our claim is proved.

Clearly, $\mu_n(S) = 1$ for all $n \geq 1$ but $\mu(S) = 0$. Hence, $\mathcal{P}(S)$ is not complete.

Remark 5. In a different direction, Bhattacharya and Lee [3] studied yet another metric. Let $\mathcal{A}$ denote the class of all sets of the form $\{x \in I\!R^k : \gamma(x) \leq c\}$ where $\gamma : I\!R^k \mapsto I\!R^k$ is continuous and non-decreasing, and $c \in I\!R^k$. Recall that $\leq$ is co-ordinatewise. Let $\overline{\mathcal{A}}$ denote the class of all sets in $\mathcal{A}$, together with limits of sequences in $\mathcal{A}$. Thus, $\overline{\mathcal{A}}$ is closed under finite unions and intersections. Then, they set

$$d(\mu, \nu) = sup_{A \in \overline{\mathcal{A}}} \quad | \mu(A \cap S) - \nu(A \cap S) |$$

They showed that for a large class $\mathcal{S}$ of subsets S of $I\!R^k$, $\mathcal{P}(S)$ is complete under d.

Remark 6. More recently, with a wide variety of applications in mind, Bhattacharya and Majumder [4] generalized the results of Bhattacharya and Lee [3] to a more abstract set up. But that is a different story.

CHAPTER - II

Convolution powers of probabilities on Stochastic matrices of order 3

Summary.

This chapter has nine sections. In section 1, we introduce the problem: Given a probability μ on the set of $d \times d$ stochastic matrices, to find necessary and sufficient conditions for the convergence of its convolution powers (μ^n). In section 2, we make the basic observation (which is perhaps not new) that in case the closed subsemigroup generated by the support of μ includes a rank one matrix, then μ^n indeed converges. In the next five sections, we consider only the case $d = 3$. In section 3, we classify 3×3 stochastic matrices – depending on classical terminology of recurrence and transience. This is mostly to fix up our notations and help to deal with the convergence case by case depending on the support of μ. In section 4, some elementary cases are discussed where the convergence or non-convergence is clear by inspection. In section 5, some special cases are discussed where the convergence needs involved arguments. Section 6 gives the case by case analysis using the results of the previous two sections and winds up the discussion. The final outcome is stated as the main theorem in section 7 using the concept of cyclicity. In the last two sections, we discuss the case $d = 2$. In section 8, we consider the problem of convergence of convolutions of a given sequence of probabilities on S_2. Our results here are not complete. In section 9, we still consider the case $d = 2$, but return to the i.i.d. case. We show that in some special cases of convergence, the absolute continuity of the limiting distribution of μ^n is related to Bernoulli Convolutions. We conclude this section with a selective and brief survey of the Bernoulli Convolutions which we found interesting.

Section 1 : Introduction

There is a vast amount of literature on convergence of convolution powers of probabilities on the space of matrices. In this and the next chapter we shall be interested in the following problem : Suppose that μ is a probability on S, the set of stochastic matrices of a fixed order,say of order d . Find easily verifiable conditions on μ so that the n-th convolution power of μ, i.e., μ^n converges. We are using here the topology of weak convergence on the space of probabilities.

Our interest in the problem stems from various points of view. Firstly, each μ on S gives rise to a Markov Process on $\mathbb{R}^d$ via random iterations as follows : If we are at $x \in \mathbb{R}^d$, we select a matrix A according to μ and move to Ax (see Marc Berger [1] for more on such matters). Secondly, in the classical theory of Markov chains with d states, one knows all about the limiting behaviour of powers of the transition matrix. However, if the transition matrix is selected according to some probability law, at each step, one would like to know if the classical result still holds in some form. Thirdly, it is natural to enquire if the neat proposition of A. Mukherjea [22] quoted below for the case of 2×2 stochastic matrices admits a neat generalization.

It should be remarked that the space of stochastic matrices being already compact, tightness criteria by Mukherjea [23] for the sequence $\{\mu^n\}_{n \geq 1}$ are of little help. It should be noted that a necessary and sufficient condition for convergence of convolution powers is given in Lemma 3, p.151 of Rosenblatt [29]. However this condition involves determining the kernel of the closure of $\{\mu^n, n \geq 1\}$. In fact origins of the present problem can be traced to notes 5.4, p.159 of Rosenblatt [29]. This problem was subsequently dealt with in some detail in [11], [22], and [25].

Here is the beautiful result of A. Mukherjea [22], mentioned above, for probabilities on 2×2 stochastic matrices:

Theorem 1.1 (A. Mukherjea [22]) :- Let μ be a probability on S, the set of stochastic matrices of order 2 and μ^n denote the nth convolution power of μ. Then the sequence μ^n converges weakly to a probability iff μ is not the point mass at the matrix $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. Further, if $S(\mu)$, the closed support of μ contains a matrix other than $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ then the limit probability is concentrated on K – the set of all matrices of the form $\begin{bmatrix} a & 1-a \\ a & 1-a \end{bmatrix}$, where $0 \leq a \leq 1$.

In this chapter the problem mentioned earlier will be considered for probabilities on 3×3 stochastic matrices, and in the next chapter for probabilities on general $d \times d$ stochastic matrices. Even though our study of probabilities on general $d \times d$ stochastic matrices will be self-contained, we present the investigations for the case of 3×3 stochastic matrices separately because of two reasons. Firstly, it gives more transparent picture about the support of μ when μ^n converges. The calculations are less algebraic in nature compared to the general case. Secondly, this analysis would perhaps help in determining whether the limit of μ^n , when exists, is singular or absolutely continuous with respect to the Lebesgue measure in an appropriate parametrization of the problem.

We should mention that S. Dhar and A.Mukherjea [11] have independently obtained Theorem 7.1 by different techniques. Their arguments are mainly algebraic in nature and depend on earlier results of Mukherjea and his coauthors. Our argument is necessarily lengthy because we consider case by case and is more probabilistic in nature with explicit computations in some cases where convergence actually occurs.

Before concluding the section we present a brief proof of the theorem mentioned above, essentially the same as in [22] :

So let μ be a probability on S, the space of 2×2 stochastic matrices. Let I stand for the identity matrix and T stand for the matrix $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$.

Clearly if μ is the point mass at the matrix T then μ^n does not converge. Conversely suppose that μ is not point mass at T. The case when μ is concentrated at I alone is trivial. If μ is concentrated only at I and T – and hence giving positive mass to each of them – then a direct calculation shows that μ^n converges to the probability giving equal mass to each of these matrices. If this is not the case then the closed support of μ includes a matrix P which corresponds to either an absorbing chain or an aperiodic recurrent chain and one knows that in such a case the powers of P converge to a rank one matrix. Thus the closed semigroup generated by the closed support of μ contains a rank one matrix. By a result of Rosenblatt [29], to be stated in the next section, any limit point of μ^n is concentrated on rank one matrices. For any two rank one matrices A and B we have $AB = B$. Thus if ν_1 and ν_2 are two limit points then $\nu_2 = \nu_1 * \nu_2 = \nu_2 * \nu_1 = \nu_1$. This shows that there is only one limit point for the sequence μ^n. The compactness of the space of probabilities now completes the proof.

Section 2 : A Basic Observation

Here we introduce some convenient notations. S_d denotes the set of all $d \times d$ stochastic matrices with usual topology. S_d is a semigroup with identity under multiplication. For any probability μ on S_d, $S_d(\mu)$ denotes the closed support of μ and $\mathcal{S}_d$ denotes the closed semigroup generated by $S_d(\mu)$. For two probabilities μ, ν on S_d; $\mu * \nu$ denotes their convolution and μ^n denotes the nth convolution power of μ.

Though our main interest is the semigroup S_d, it is not always the case that the closed support of μ generates S_d. Thus we shall be interested in subsemigroups of S_d as well. We now recall a fundamental result of Rosenblatt [29] which plays a crucial role in the analysis. So let S be a compact metric semigroup. Recall that Kernel K of S is the smallest non-empty two-sided ideal of S. Suppose that μ is a probability on (the Borel σ-field of) S.

Lemma (Rosenblatt) : Every limit point of μ^n is concentrated on K.

This is Lemma 3, p.141 of [29]. Of course there the result is stated for more general compact Hausdorff semigroups, but we will have no occasion to use in this generality.

In our case, K_d, the Kernel of S_d, consists of all $d \times d$ stochastic matrices with identical rows. If $\mathcal{S}_d \cap K_d \neq \emptyset$, then by the above result, every limit point ν of μ^n is concentrated on $\mathcal{S}_d \cap K_d$. Observe that $xy = y$ holds for all $x, y \in K_d$. This implies that if ν_1 and ν_2 are two probabilities concentrated on K_d then $\nu_1 * \nu_2 = \nu_2$. As a consequence, if $\mathcal{S}_d \cap K_d \neq \emptyset$ and ν_1, ν_2 are two limit points of $\{\mu^n\}_{n \geq 1}$, then $\nu_2 = \nu_1 * \nu_2 = \nu_2 * \nu_1 = \nu_1$. Thus we get

Lemma 2.1 :- Let $d \geq 2$. If $\mathcal{S}_d \cap K_d \neq \emptyset$ then μ^n converges.

Of course, even when $\mathcal{S}_d \cap K_d = \emptyset$, μ^n may converge. For example if $d = 2$, there are only three cases when $\mathcal{S}_d \cap K_d = \emptyset$, namely :-

$$S_d(\mu) = \left\{ \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \right\}, \text{ or } S_d(\mu) = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right\},$$

$$\text{or } S_d(\mu) = \left\{ \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right\}.$$

These cases have been already discussed in the previous section.

But even for $d = 3$, there are a large number of interesting cases where $\mathcal{S}_d \cap K_d = \emptyset$. We deal with them in the following sections.

In what follows, we omit the subscripts 'd' from S_d, K_d, $\mathcal{S}_d$ and $S_d(\mu)$ and write them simply as $S, K, \mathcal{S}$ and $S(\mu)$ respectively with the understanding that $d = 3$.

Section 3 : Classifying stochastic matrices of order 3:

As stated earlier, our analysis of convergence of μ^n is done case by case , depending on which matrices are in the support of μ. To facilitate this, we divide S into certain subsets according to – following classical terminology – the number of recurrent and transient classes :-

(1) All three states are recurrent and they form disjoint classes :- Identity matrix is the only matrix in this subset. Let us call this subset to be S_1. Then S_1 is closed in S.

(2) Two recurrent classes and no transient class :- There are three such subsets, namely,

$$S_{21} = \left\{ \begin{pmatrix} 1 & 0 & 0 \\ 0 & \alpha & 1-\alpha \\ 0 & \beta & 1-\beta \end{pmatrix} : 0 \leq \alpha < 1, \quad 0 < \beta \leq 1 \right\}$$

$$S_{22} = \left\{ \begin{pmatrix} 1-\beta & 0 & \beta \\ 0 & 1 & 0 \\ 1-\alpha & 0 & \alpha \end{pmatrix} : 0 \leq \alpha < 1, \quad 0 < \beta \leq 1 \right\}$$

$$S_{23} = \left\{ \begin{pmatrix} \alpha & 1-\alpha & 0 \\ \beta & 1-\beta & 0 \\ 0 & 0 & 1 \end{pmatrix} : 0 \leq \alpha < 1, \quad 0 < \beta \leq 1 \right\}$$

Call $S_2 = S_{21} \cup S_{22} \cup S_{23}$. Of course S_2 is not closed in S.

(3) Two recurrent and one transient classes :- There are, once again, three such subsets, namely,

$$S_{31} = \left\{ \begin{pmatrix} 1-\alpha-\beta & \alpha & \beta \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} : \alpha, \beta \geq 0, 0 < \alpha + \beta \leq 1 \right\}$$

$$S_{32} = \left\{ \begin{pmatrix} 1 & 0 & 0 \\ \beta & 1-\alpha-\beta & \alpha \\ 0 & 0 & 1 \end{pmatrix} : \alpha, \beta \geq 0, 0 < \alpha + \beta \leq 1 \right\}$$

$$S_{33} = \left\{ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \alpha & \beta & 1-\alpha-\beta \end{pmatrix} : \alpha, \beta \geq 0, 0 < \alpha + \beta \leq 1 \right\}$$

Call $S_3 = S_{31} \cup S_{32} \cup S_{33}$.

(4) One recurrent class and no transient class :- This class consists of the irreducible matrices, namely,

(i) Irreducible, aperiodic matrices forming the subset S_4^1

(ii) Irreducible period-two matrices : -

$$S_{41}^2 = \left\{ \begin{pmatrix} 0 & \alpha & 1-\alpha \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} : 0 < \alpha < 1 \right\}$$

$$S_{42}^2 = \left\{ \begin{pmatrix} 0 & 1 & 0 \\ 1-\alpha & 0 & \alpha \\ 0 & 1 & 0 \end{pmatrix} : 0 < \alpha < 1 \right\}$$

$$S_{43}^2 = \left\{ \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ \alpha & 1-\alpha & 0 \end{pmatrix} : 0 < \alpha < 1 \right\}$$

Denote $S_4^2 = S_{41}^2 \cup S_{42}^2 \cup S_{43}^2$.

(iii) Irreducible period-three matrices : -

$$S_{41}^3 = \left\{ \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \right\}, \quad S_{42}^3 = \left\{ \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \right\}$$

Call $S_4^3 = S_{41}^3 \cup S_{42}^3$.

(5) One recurrent class having two states and one transient class :-

S_{51} is the set of all matrices

$$\left\{ \begin{pmatrix} 1-\gamma-\delta & \gamma & \delta \\ 0 & \alpha & 1-\alpha \\ 0 & \beta & 1-\beta \end{pmatrix} \right\}$$

where

$$0 \leq \alpha < 1, \ 0 < \beta \leq 1, \ \gamma, \delta \geq 0, \ 0 < \gamma + \delta \leq 1$$

S_{52} is the set of all matrices

$$\left\{ \begin{pmatrix} 1-\beta & 0 & \beta \\ \delta & 1-\gamma-\delta & \gamma \\ 1-\alpha & 0 & \alpha \end{pmatrix} \right\}$$

where

$$0 \leq \alpha < 1,\ 0 < \beta \leq 1,\ \gamma, \delta \geq 0,\ 0 < \gamma + \delta \leq 1$$

S_{53} is the set of all matrices

$$\left\{ \begin{pmatrix} \alpha & 1-\alpha & 0 \\ \beta & 1-\beta & 0 \\ \gamma & \delta & 1-\gamma-\delta \end{pmatrix} \right\}$$

where

$$0 \leq \alpha < 1,\ 0 < \beta \leq 1,\ \gamma, \delta \geq 0,\ 0 < \gamma + \delta \leq 1$$

Call $S_5 = S_{51} \cup S_{52} \cup S_{53}$.

(6) One recurrent class with one state and other states are transient :

$$S_{61} = \left\{ \begin{pmatrix} 1 & 0 & 0 \\ \alpha & 1-\alpha-\beta & \beta \\ \gamma & \delta & 1-\gamma-\delta \end{pmatrix} : 0 \leq \alpha, \beta, \gamma, \delta; \alpha+\beta > 0, \gamma+\delta > 0 \right\}$$

$$S_{62} = \left\{ \begin{pmatrix} 1-\gamma-\delta & \gamma & \delta \\ 0 & 1 & 0 \\ \beta & \alpha & 1-\alpha-\beta \end{pmatrix} : 0 \leq \alpha, \beta, \gamma, \delta; \alpha+\beta > 0, \gamma+\delta > 0 \right\}$$

$$S_{63} = \left\{ \begin{pmatrix} 1-\alpha-\beta & \beta & \alpha \\ \delta & 1-\gamma-\delta & \gamma \\ 0 & 0 & 1 \end{pmatrix} : 0 \leq \alpha, \beta, \gamma, \delta; \alpha+\beta > 0, \gamma+\delta > 0 \right\}$$

Call $S_6 = S_{61} \cup S_{62} \cup S_{63}$.

Now if we study carefully the closures of the above sets, we can make the following conclusions : -

Remark 1. Matrices from (2), (3) and (5) above have come from one of the following large subclasses of S :-

$$T_1 = \left\{ \begin{pmatrix} 1-\gamma-\delta & \gamma & \delta \\ 0 & \alpha & 1-\alpha \\ 0 & \beta & 1-\beta \end{pmatrix} : 0 \leq \gamma, \delta, \gamma+\delta, \alpha, \beta \leq 1 \right\}$$

$$T_2 = \left\{ \begin{pmatrix} 1-\beta & 0 & \beta \\ \delta & 1-\gamma-\delta & \gamma \\ 1-\alpha & 0 & \alpha \end{pmatrix} : 0 \leq \gamma, \delta, \gamma+\delta, \alpha, \beta \leq 1 \right\}$$

$$T_3 = \left\{ \begin{pmatrix} \alpha & 1-\alpha & 0 \\ \beta & 1-\beta & 0 \\ \gamma & \delta & 1-\gamma-\delta \end{pmatrix} : 0 \leq \gamma, \delta, \gamma+\delta, \alpha, \beta \leq 1 \right\}$$

Then,

$$T_1 = \overline{S_{21}} \cup \overline{S_{31}} \cup \overline{S_{51}}, \quad T_2 = \overline{S_{22}} \cup \overline{S_{32}} \cup \overline{S_{52}}, \quad T_3 = \overline{S_{23}} \cup \overline{S_{33}} \cup \overline{S_{53}}$$

Observe that an appropriate renaming of the states leads from one of the large subclasses above to the others .

Remark 2. Closures of each of the subsets in (2),(3),(5) and (6) contains S_1, that is, their closures contain the identity matrix.

Remark 3. We have,

$$N_1 \in \overline{S_{21}} \cap S_{32}, \text{ where } N_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

$$N_2 \in \overline{S_{21}} \cap S_{33}, \text{ where } N_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

Similarly, $\overline{S_{22}} \cap S_{33}, \overline{S_{22}} \cap S_{31}, \overline{S_{23}} \cap S_{31}, \overline{S_{23}} \cap S_{32}$ contain the zero-one matrices N_3, N_4, N_5, N_6 respectively defined in an appropriate way.

Remark 4.

Apart from the identity matrix e_o , there are five more permutation matrices, namely,

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$

and $\begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$.

We denote them by $\boldsymbol{e_1}, \boldsymbol{e_2}, \boldsymbol{e_3}, \boldsymbol{e_4}$ and $\boldsymbol{e_5}$ respectively. Then $\boldsymbol{e_1} \in S_{21}, \boldsymbol{e_2} \in S_{22}, \boldsymbol{e_3} \in S_{23}$ and $S_{41}^3 = \{\boldsymbol{e_4}\}, S_{42}^3 = \{\boldsymbol{e_5}\}, S_1 = \{\boldsymbol{e_0}\}$. Let us call P to be the set of all these permutation matrices and let $P_1 = \{\boldsymbol{e_1}, \boldsymbol{e_2}, \boldsymbol{e_3}\}$ and $P_2 = \{\boldsymbol{e_0}, \boldsymbol{e_4}, \boldsymbol{e_5}\}$.

Remark 5. Define $S_{51}^0, S_{52}^0, S_{53}^0$,subsets of S_{51}, S_{52}, S_{53} respectively, as follows : -

$$S_{51}^0 = \left\{ \begin{pmatrix} 1-\gamma-\delta & \gamma & \delta \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} : \gamma, \delta \geq 0, 1 \geq \gamma + \delta > 0 \right\}$$

$$S_{52}^0 = \left\{ \begin{pmatrix} 0 & 0 & 1 \\ \delta & 1-\gamma-\delta & \gamma \\ 1 & 0 & 0 \end{pmatrix} : \gamma, \delta \geq 0, 1 \geq \gamma + \delta > 0 \right\}$$

$$S_{53}^0 = \left\{ \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ \gamma & \delta & 1-\gamma-\delta \end{pmatrix} : \gamma, \delta \geq 0, 1 \geq \gamma + \delta > 0 \right\}$$

Then $\boldsymbol{e_j} \in \overline{S_{5j}^0}$, for $j = 1, 2, 3$.

Call $S_5^0 = S_{51}^0 \cup S_{52}^0 \cup S_{53}^0$.

Remark 6. We have, $\overline{S_{41}^2} \cap S_{52}, \overline{S_{41}^2} \cap S_{53}, \overline{S_{42}^2} \cap S_{53}, \overline{S_{42}^2} \cap S_{51}, \overline{S_{43}^2} \cap S_{51}, \overline{S_{43}^2} \cap S_{52}$ are singletons containing the matrices $M_1, M_2, M_3, M_4, M_5, M_6$ respectively, where,

$$M_1 = \left\{ \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \right\} \quad M_2 = \left\{ \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \right\} \quad M_3 = \left\{ \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \right\}$$

$$M_4 = \left\{ \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \right\} \quad M_5 = \left\{ \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \right\} \quad M_6 = \left\{ \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \right\}.$$

Section 4 : Some Elementary Cases

From now on, μ is a probability on S. We start with some preliminary observations :-

1. If $S(\mu) \subseteq \overline{S_{21}}$, then state 1 being fixed, 2×2 case of Theorem 1.1 shows that μ^n converges iff $\mu \neq \delta_{e_1}$. Similar conclusions hold if $S(\mu) \subseteq \overline{S_{22}}$ or $S(\mu) \subseteq \overline{S_{23}}$.

2. If $S(\mu) \cap S_4^1 \neq \emptyset$, then for each matrix $A \in S_4^1$, we know that A^n converges to a matrix with identical rows. Thus, $\mathcal{S} \cap K \neq \emptyset$ which implies, by lemma 2.1, that μ^n converges.

3. If $S(\mu) \cap (S_5 - S_5^0) \neq \emptyset$, then $\mathcal{S}$ contains a matrix for which one column will be zero, one column will consist of all a's and the remaining column will consist of all $(1-a)$'s for some $0 < a < 1$. This once again implies that, $\mathcal{S} \cap K \neq \emptyset$ and μ^n converges.

4. If $S(\mu) \cap S_6 \neq \emptyset$ then $\mathcal{S}$ has a matrix which has one column consisting of ones implying that $\mathcal{S} \cap K \neq \emptyset$ and hence μ^n converges.

5. If $S(\mu) \subseteq \{N_1, N_2\}$, or $S(\mu) \subseteq \{N_3, N_4\}$, or $S(\mu) \subseteq \{N_5, N_6\}$, we have, $\mu^n = \mu$ for all $n \geq 1$ so that μ^n converges to μ itself. On the other hand, if $S(\mu) \subseteq \{M_1, M_2\}$, or $S(\mu) \subseteq \{M_3, M_4\}$, or $S(\mu) \subseteq \{M_5, M_6\}$, then μ^n does not converge.

6. If $S(\mu) \subseteq \overline{S_{41}^2}$ or $S(\mu) \subseteq \overline{S_{42}^2}$ or $S(\mu) \subseteq \overline{S_{43}^2}$, then the supports of μ^k for k even and k odd are disjoint so that μ^n does not converge.

7. If $S(\mu) \subseteq \overline{S_{51}^0}$ or $S(\mu) \subseteq \overline{S_{52}^0}$ or $S(\mu) \subseteq \overline{S_{53}^0}$, then, once again, the supports of μ^k for k even and k odd are disjoint so that μ^n does not converge.

8. If $S(\mu) = \{N_i, M_i\}$ for $1 \leq i \leq 6$, then μ^n converges to the measure putting mass $1/2$ to each of the two matrices.

9. If $S(\mu) \subseteq \{N_1, N_2, M_1, M_2\}$ and includes $\{N_1, M_2\}$ or $\{N_2, M_1\}$, then μ^n converges to the uniform probability on $\{N_1, N_2, M_1, M_2\}$.

   If $S(\mu) \subseteq \{N_3, N_4, M_3, M_4\}$ and includes $\{N_3, M_4\}$ or $\{N_4, M_3\}$, then μ^n converges to the uniform probability on $\{N_3, N_4, M_3, M_4\}$.

If $S(\mu) \subseteq \{N_5, N_6, M_5, M_6\}$ and includes $\{N_5, M_6\}$ or $\{N_5, M_6\}$, then μ^n converges to the uniform probability on $\{N_5, N_6, M_5, M_6\}$.

10. If $S(\mu) \subseteq P$, the set of permutation matrices (see Remark 4 of previous section for notations), then we have the following conclusions.

    (a) In case $\mu = \delta_{e_i}, 1 \leq i \leq 5$, clearly, μ^n does not converge.

    (b) If $S(\mu) \subseteq P_1$, supports of μ^n and μ^{n+1} are disjoint so that μ^n does not converge.

    (c) If $S(\mu) \subseteq P_2$ and is not a singleton, then μ^n converges to the limit putting equal masses at e_0, e_4, e_5.

    (d) If $S(\mu) \cap P_1 \neq \emptyset$ and $S(\mu) \cap P_2 \neq \emptyset$, then μ^n converges to the limit which is uniform having support P.

11. If $S(\mu) \cap S_4^3 \neq \emptyset$ and either of the following holds, then $S \cap K \neq \emptyset$: -

    (a)$S(\mu) \cap (\overline{S_5^0} - \{e_1, e_2, e_3\}) \neq \emptyset$,

    (b)$S(\mu) \cap \overline{S_4^2} \neq \emptyset$,

    (c)$S(\mu) \cap S_3 \neq \emptyset$,

    (d)$S(\mu) \cap (S_2 - \{e_1, e_2, e_3\}) \neq \emptyset$.

Remark 1. Because of the above, we shall make the following assumptions for sections 5 and 6 :

i)$S(\mu) \cap S_4^1 = \emptyset, \quad S(\mu) \cap (S_5 - S_5^0) = \emptyset, \quad S(\mu) \cap S_6 = \emptyset$. See 2, 3 and 4 above.

ii)$S(\mu) - \overline{S_{2i}} \neq \emptyset, 1 \leq i \leq 3$. See 1 above.

iii)$S(\mu) - \overline{S_{4i}^2} \neq \emptyset, 1 \leq i \leq 3$. See 6 above.

iv)$S(\mu) - \overline{S_{5i}^0} \neq \emptyset, 1 \leq i \leq 3$. See 7 above.

v)$S(\mu) - P \neq \emptyset$. See 10 above.

vi)$S(\mu) \cap S_4^3 = \emptyset$. See 11 above.

vii)$S(\mu) - \{N_i, M_i, N_{i+1}, M_{i+1}\} \neq \emptyset$ for $i = 1, 3, 5$. See 5, 8, and 9 above.

Remark 2. Under the above assumptions, it follows that

$$S(\mu) \subseteq S_1 \cap \overline{S_2} \cap \overline{S_3} \cap \overline{S_4^2} \cap \overline{S_5^0}.$$

We shall now divide the remaining cases into two broad divisions according as $\mathcal{S} \cap K = \emptyset$ and $\mathcal{S} \cap K \neq \emptyset$ and discuss them in sections 5 and 6 respectively.

Section 5 : Some Special Cases

In the next section we discuss the convergence or otherwise of μ^n in all cases. In order not to interrupt that discussion we consider four special cases in this section. In all these cases, we find that $\mathcal{S} \cap \mathcal{K} = \emptyset$. However we succeed in constructing the Kernel of the semigroup $\mathcal{S}$. This by Rosenblatt's lemma [29] allows us to infer where the limit points of μ^n are concentrated. But an extra argument is still needed to show that there is only one limit. We shall do this thereby showing that in all these four cases μ^n does indeed converge.

Case I : First of all, we consider the case when $S(\mu) \subseteq S_{31}$ or $S(\mu) \subseteq S_{32}$ or $S(\mu) \subseteq S_{33}$.

To be specific, we consider :-

$$S(\mu) \subseteq S_{31} = \left\{ \begin{pmatrix} 1-\alpha-\beta & \alpha & \beta \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} : 0 \leq \alpha,\ \beta \leq 1,\ \alpha+\beta > 0 \right\}.$$

Then, $\overline{S_{31}}$ is the closed semigroup generated by S_{31} , that is, $\mathcal{S} \subseteq \overline{S_{31}}$ and the kernel K_{31} of $\overline{S_{31}}$ is given by :-

$$K_{31} = \left\{ \begin{pmatrix} 0 & \alpha & 1-\alpha \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} : 0 \leq \alpha \leq 1 \right\}$$

Then, clearly, $\mathcal{S} \cap K_{31} \neq \emptyset$. So, by Rossenblatt's result mentioned in section 2, any cluster point of $\{\mu^n\}_{n\geq 1}$ will have support $\subseteq K_{31}$. Since $xy = x$ for all $x, y \in K_{31}$, we see as in section 2 that if ν_1 and ν_2 are two cluster points, then $\nu_1 = \nu_2$ so that μ^n converges.

Similarly, in case $S(\mu) \subseteq S_{32}$ or $S(\mu) \subseteq S_{33}$, analogous arguments will conclude that μ^n converges.

Case II : Next we consider the following set of cases :-

$$(a) N_1 \in S(\mu), \quad S(\mu) - \{N_1, M_1\} \subseteq S^2_{41}$$
$$(b) N_2 \in S(\mu), \quad S(\mu) - \{N_2, M_2\} \subseteq S^2_{41}$$
$$(c) N_3 \in S(\mu), \quad S(\mu) - \{N_3, M_3\} \subseteq S^2_{42}$$
$$(d) N_4 \in S(\mu), \quad S(\mu) - \{N_4, M_4\} \subseteq S^2_{42}$$
$$(e) N_5 \in S(\mu), \quad S(\mu) - \{N_5, M_5\} \subseteq S^2_{43}$$
$$(f) N_6 \in S(\mu), \quad S(\mu) - \{N_6, M_6\} \subseteq S^2_{43}$$

We shall discuss the case (d). Other cases are similar.
Now,

$$S^2_{42} = \left\{ \begin{pmatrix} 1-\alpha & 0 & \alpha \\ 0 & 1 & 0 \\ 1-\alpha & 0 & \alpha \end{pmatrix} : 0 < \alpha < 1 \right\}$$

The closed semigroup generated by S^2_{42} is given by

$$\mathcal{S}_4 = \left\{ \begin{pmatrix} 1-\alpha & 0 & \alpha \\ 0 & 1 & 0 \\ 1-\alpha & 0 & \alpha \end{pmatrix}, \begin{pmatrix} 0 & 1 & 0 \\ 1-\beta & 0 & \beta \\ 0 & 1 & 0 \end{pmatrix} : 0 \leq \alpha,\ \beta \leq 1 \right\}$$

$= \mathcal{S}_{41} \cup \mathcal{S}_{42}$ (say)

where, $\mathcal{S}_{41} = \left\{ \begin{pmatrix} 1-\alpha & 0 & \alpha \\ 0 & 1 & 0 \\ 1-\alpha & 0 & \alpha \end{pmatrix} : 0 \leq \alpha \leq 1 \right\}$

and $\mathcal{S}_{42} = \left\{ \begin{pmatrix} 0 & 1 & 0 \\ 1-\beta & 0 & \beta \\ 0 & 1 & 0 \end{pmatrix} : 0 \leq \beta \leq 1 \right\}$

Thus in this case $\mathcal{S}$, the closed semigroup generated by the support of μ is contained in $\mathcal{S}_4$. Then, observe that N_3, M_3, N_4, M_4 all belong to $\mathcal{S}_4$, where,

$$N_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}, N_4 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

$$M_3 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, M_4 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}.$$

The Kernel of the semigroup K_4 is $\mathcal{S}_4$ itself.

Then the case under consideration is covered by the following claim which we shall prove :-

<u>Claim</u> : If $\mu(\mathcal{S}_4) = 1$ then, μ^n converges as soon as $\mu(\mathcal{S}_{41}) > 0$ and $\mu(\mathcal{S}_{42}) > 0$.

To do this, we shall indeed show that for every Borel subset A of $\mathcal{S}_4$, $\mu^n(A)$ converges – in particular μ^n converges weakly. We define a map $\phi : \mathcal{S}_4 \longrightarrow \mathcal{S}_4$ by

$$\phi\begin{pmatrix} 1-\alpha & 0 & \alpha \\ 0 & 1 & 0 \\ 1-\alpha & 0 & \alpha \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 1-\alpha & 0 & \alpha \\ 0 & 1 & 0 \end{pmatrix}, \quad 0 \leq \alpha \leq 1$$

and

$$\phi\begin{pmatrix} 0 & 1 & 0 \\ 1-\beta & 0 & \beta \\ 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 1-\beta & 0 & \beta \\ 0 & 1 & 0 \\ 1-\beta & 0 & \beta \end{pmatrix}, \quad 0 \leq \beta \leq 1$$

Then ϕ is a bijection, $\phi = \phi^{-1}$, $\phi(\mathcal{S}_{41}) = \mathcal{S}_{42}$ and $\phi(\mathcal{S}_{42}) = \mathcal{S}_{41}$.
We start by observing that

$$\begin{aligned} x \in \mathcal{S}_{41}, y \in \mathcal{S}_{41} &\implies xy = y \in \mathcal{S}_{41} \\ x \in \mathcal{S}_{41}, y \in \mathcal{S}_{42} &\implies xy = y \in \mathcal{S}_{42} \\ x \in \mathcal{S}_{42}, y \in \mathcal{S}_{41} &\implies xy = \phi(y) \in \mathcal{S}_{42} \\ x \in \mathcal{S}_{42}, y \in \mathcal{S}_{42} &\implies xy = \phi(y) \in \mathcal{S}_{41} \end{aligned}$$

Note that, for any two probabilities μ_1 and μ_2 on $\mathcal{S}_4$,

$$\mu_1 * \mu_2(A) = \int \mu_2[y : xy \in A] \quad \mu_1(dx)$$

$$\begin{aligned}
&= \int_{S_{41}} \mu_2[y : xy \in A] \quad \mu_1(dx) + \int_{S_{42}} \mu_2[y : xy \in A] \quad \mu_1(dx) \\
&= \int_{S_{41}} \mu_2[y : y \in A] \quad \mu_1(dx) + \int_{S_{42}} \mu_2[y : \phi(y) \in A] \quad \mu_1(dx) \\
&= \mu_2(A)\mu_1(S_{41}) + \mu_2\phi^{-1}(A)\mu_1(S_{42}) \\
&= \mu_2(A)\mu_1(S_{41}) + \mu_2\phi(A)\mu_1(S_{42}) \qquad [\text{ since } \phi = \phi^{-1}] - -(*)
\end{aligned}$$

Now let μ be any probability on S_4 with $\mu(S_{41}) = c$, $0 < c < 1$. Let $A \subseteq S_{41}$. Let for $n \geq 1$, α_n and β_n denote $\mu^n(A)$ and $\mu^n(\phi(A))$ respectively.

Then, from $(*)$, for any Borel set B,

$$\mu^{n+1}(B) = \mu * \mu^n(B) = \mu^n(B)\mu(S_{41}) + \mu^n(\phi(B))\mu(S_{42})$$

In particular, setting $B = A$, we get,

$$\begin{aligned}
\alpha_{n+1} = \mu^{n+1}(A) &= \mu^n(A)\mu(S_{41}) + \mu^n(\phi(A))\mu(S_{42}) \\
&= \alpha_n c + \beta_n(1-c)
\end{aligned}$$

and setting $B = \phi(A)$, we get,

$$\begin{aligned}
\beta_{n+1} = \mu^{n+1}(\phi(A)) &= \mu^n(\phi(A))\mu(S_{41}) + \mu^n(A)\mu(S_{42}) \\
&= \beta_n c + \alpha_n(1-c)
\end{aligned}$$

So, $$\begin{pmatrix} \alpha_{n+1} \\ \beta_{n+1} \end{pmatrix} = \begin{pmatrix} c & 1-c \\ 1-c & c \end{pmatrix} \begin{pmatrix} \alpha_n \\ \beta_n \end{pmatrix} = \begin{pmatrix} c & 1-c \\ 1-c & c \end{pmatrix}^n \begin{pmatrix} \alpha_1 \\ \beta_1 \end{pmatrix}$$

$$\longrightarrow \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \beta_1 \end{pmatrix} \text{ as } n \to \infty$$

as $0 < c < 1$.

So $\alpha_n \to \frac{\alpha_1+\beta_1}{2}$ and $\beta_n \to \frac{\alpha_1+\beta_1}{2}$ as $n \to \infty$. Hence, $\mu^n(A)$ and $\mu^n(\phi(A))$ converges completing the proof.

Cases (a),(b),(c),(e) and (f) can be disposed of in a similar fashion.

<u>Case III</u> :

Here we consider the following three cases depending on $i = 1, 2$ or 3:
$S(\mu) \cap S_{3i} \neq \emptyset$, $\quad S(\mu) \cap \overline{S_{5i}^0} \neq \emptyset$ and $\quad S(\mu) \subseteq S_{3i} \cup \overline{S_{5i}^0}$.

We shall only consider the case corresponding to $i = 1$. Then the closed semigroup generated by S_{31} and $\overline{S^0_{51}}$ is given by $\mathcal{S}_{35} = \mathcal{S}_{33} \cup \mathcal{S}_{55}$ where

$$\mathcal{S}_{33} = \left\{ \begin{pmatrix} a & b & c \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} : a, b, c \geq 0, \ \ a + b + c = 1 \right\}$$

$$\mathcal{S}_{55} = \left\{ \begin{pmatrix} a & b & c \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} : a, b, c \geq 0, \ \ a + b + c = 1 \right\}.$$

Thus in this case the closed semigroup $\mathcal{S}$ generated by the support of μ is contained in $\mathcal{S}_{35}$. The kernel of the semigroup $\mathcal{S}_{35}$ is $K_{35} = K_{33} \cup K_{55}$ where K_{33} and K_{55} consist of all matrices in $\mathcal{S}_{33}$ and $\mathcal{S}_{55}$ respectively with $a = 0$. The case under consideration is covered by the following claim which we shall prove

Claim : If $\mu(\mathcal{S}_{35}) = 1$, $\mu(\mathcal{S}_{33}) > 0$, $\mu(\mathcal{S}_{55}) > 0$, then μ^n converges.

From now on we assume that μ is as in the claim and we put $c = \mu(\mathcal{S}_{33})$ so that $0 < c < 1$.

If Q is any limit point of μ^n then by Rosenblatt's result, $Q(K_{35}) = 1$. We now argue that $Q(K_{33}) = Q(K_{55}) = \frac{1}{2}$. First note that for $x, y \in \mathcal{S}_{35}$ we have $xy \in \mathcal{S}_{33}$ iff either both x, y are in $\mathcal{S}_{33}$ or both x, y are in $\mathcal{S}_{55}$. As a consequence if we let $\alpha_n = \mu^n(\mathcal{S}_{33})$ then

$$\begin{aligned} \alpha_{n+1} &= \int \mu^n(y : xy \in \mathcal{S}_{33}) d\mu(x) \\ &= \alpha_n c + (1 - \alpha_n)(1 - c) \end{aligned}$$

Thus

$$\begin{aligned} \alpha_{n+1} &= \alpha_1 (2c - 1)^n + (1 - c) \sum_{k=0}^{n-1} (2c - 1)^k \\ &\to \frac{1}{2} \qquad \text{as } \ 0 < c < 1. \end{aligned}$$

Let ϕ be the map from $\mathcal{S}_{35}$ to $\mathcal{S}_{35}$ which interchanges the last two columns. Then ϕ is a bijection, $\phi = \phi^{-1}$; $\phi(\mathcal{S}_{33}) = \mathcal{S}_{55}$; $\phi(K_{33}) = K_{55}$. Moreover for $x, y \in K_{35}$ we have $x.y = x$ or $\phi(x)$ according as $y \in K_{33}$ or $y \in K_{55}$. Thus the premultiplier matrix x determines the entries of the product.

As a consequence for two probabilities Q_1, Q_2 supported on K_{35} it is easy to see that

$$Q_1 * Q_2(A) = \frac{Q_1(A) + Q_1(\phi(A))}{2}$$

Then, since by Lemma 2.1 proved in section 2 of chapter 3, the set of limit points of $(\mu^n)_{n\geq 1}$ is a group, let R be the identity of the group. Then, for any other limit point Q, we have, $Q * R = R * Q = Q$. So, taking $Q_2 = R$ above, we get,

$$Q_1(A) = \frac{Q_1(A) + Q_1(\phi(A))}{2}$$

This implies that $Q_1(A) = Q_1(\phi(A))$ for any Borel subset of $\mathcal{S}_{35}$ and consequently it follows that $Q_1 = Q_1 * Q_2$.

Thus if Q_1, Q_2 are two limit points of (μ^n) then using the fact that $Q_1 * Q_2 = Q_2 * Q_1$ we get

$$Q_1 = Q_1 * Q_2 = Q_2 * Q_1 = Q_2.$$

So, any two limit points are same or μ^n converges.

Alternatively, if one does not wish to use the lemma, one can argue as follows :

Let us consider a sequence of i.i.d. matrices $X_1, X_2, \cdots$ each having distribution μ so that $Y_n = X_1 \cdots X_n$ has distribution μ^n. First observe that if μ is concentrated on $\left\{ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \right\}$ giving positive mass to both the matrices, then μ^n converges to Q which puts mass $\frac{1}{2}$ at each of these two matrices – a fact already pointed out in section 4. From now on we assume that this is not the case. In other words if Z denotes the first entry of the random matrix X_1 then $\mu(Z < 1) > 0$. Note that as a consequence if $p = E(Z)$ then $0 \leq p < 1$.

To make the arguments transparent we shall first consider the case $\mu(K_{35}) > 0$ – though the proof for the general case applies here too. In this case we show that for every Borel set $A \subset S_{35}$, $\mu^n(A)$ converges. Since $\mu(K_{35}) > 0$, almost surely $X_N \in K_{35}$ for some random integer N and then of course for all $n > N$, $Y_n \in K_{35}$. As a consequence if $A = S_{35} - K_{35}$ then $\mu^n(A) \to 0$ (as it should). Now fix any Borel set $A \subset K_{35}$. We show that $\{\mu^n(A)\}$ is a cauchy sequence. To this end fix $\epsilon > 0$. Choose an integer k so that $P(N < k) > 1 - \epsilon/4$ and also that $|\alpha_n - \frac{1}{2}| < \epsilon/4$ for $n \geq k$. Recall that $\alpha_n = \mu^n(S_{35}) \to 1/2$. Now for any $n > 2k$

$$\begin{aligned}
&\mu^n(A)\\
&= P(Y_n \in A)\\
&= P(Y_k \in A,\ \textstyle\prod_{i=k+1}^n X_i \in S_{33}) + P(Y_k \in \phi(A);\ \prod_{i=k+1}^n X_n \in S_{55})\\
&+P(Y_k \notin K_{35};\ Y_n \in A)\\
&= \alpha_{n-k}\mu^k(A) + (1-\alpha_{n-k})\mu^k(\phi(A)) + P(Y_k \notin K_{35},\ Y_n \in A).
\end{aligned}$$

Since $|\alpha_{n-k} - \frac{1}{2}| < \epsilon/4$ and $P(Y_k \notin K_{35}) < \epsilon/4$ we get that for $n > 2k$,

$$|\mu^n(A) - \frac{\mu^k(A) + \mu^k(\phi(A))}{2}| < \epsilon/2$$

showing that for $n, m > 2k$ $|\mu^n(A) - \mu^m(A)| < \epsilon$ to complete the proof.

We shall now consider the general case. It suffices to show that for every (bounded) continuous function f on S_{35} with bounded first derivatives $\int f d\mu^n$ converges — or that it is a cauchy sequence. Define the numbers

$$a_n = E[f(Y_n)1_{Y_n \in S_{33}}],\quad b_n = E[f(\phi(Y_n))1_{Y_n \in S_{55}}]$$
$$c_n = E[f(Y_n)1_{Y_n \in S_{55}}],\quad d_n = E[f(\phi(Y_n))1_{Y_n \in S_{33}}]$$

and the matrix M_n by

$$M_n = \begin{pmatrix} a_n & b_n \\ c_n & d_n \end{pmatrix}$$

Let Z_n denote the first entry of Y_n. Explicit calculations show that if $X_{n+1} \in S_{33}$ then $Y_n X_{n+1} - Y_n$ has second and third rows null while each entry in first row is smaller then Z_n in modulus. On the other hand, if $X_{n+1} \in S_{55}$ then $Y_n X_{n+1} - \phi(Y_n)$ has second and third rows null while each entry in the first row is smaller than Z_n in modulus. This fact combined with the meanvalue theorem yields that $|f(Y_{n+1}) - f(Y_n)| \leq kZ_n$ when $X_{n+1} \in S_{33}$ and $|f(Y_{n+1}) - f(\phi(Y_n))| \leq kZ_n$ when $X_{n+1} \in S_{55}$ where k is a constant depending on the first derivates of f which were assumed bounded. Observing that

$$a_{n+1} = E[f(Y_{n+1})1_{Y_n \in S_{33}}1_{X_{n+1} \in S_{33}}] + E[f(Y_{n+1})1_{Y_n \in S_{55}}1_{X_{n+1} \in S_{55}}]$$

we obtain

$$|a_{n+1} - [ca_n + (1-c)b_n]| \leq kE(Z_n)$$

Observing that Z_n is nothing but the product of the first entries of $X_1, \cdots, X_n$ we have $E(Z_n) = p^n$. Recall that $p = E(Z_1)$ and $0 \leq p < 1$. Thus

$$|a_{n+1} - [ca_n + (1-c)b_n]| \leq kp^n$$

Letting C be the matrix $\begin{pmatrix} c & 1-c \\ 1-c & c \end{pmatrix}$ and U be the matrix $\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$, a similar calculation with b_n, c_n, d_n gives us

$$M_nC - p^nU \leq M_{n+1} \leq M_nC + p^nU$$

entrywise. Noting that C^n converges to the matrix with all entries $\frac{1}{2}$ and the fact that Σp^n converges it is not difficult to show that entries of M_n form cauchy sequences. Observing that $E[f(Y_n)] = a_n + c_n$ we conclude that it is a cauchy sequence to complete the proof. Incidentally, notice that $E(f(Y_n))$ and $E(f(\phi(Y_n)))$ have the same limit.

This completes the proof of the claim.

Case IV :-

Lastly, we have three cases depending on $i = 1,\ 2,$ or 3 are :

$S(\mu) \cap \overline{S_{2i}} \neq \emptyset, \quad S(\mu) \cap \overline{S_{4i}^2} \neq \emptyset$ and $S(\mu) \subseteq \overline{S_{2i}} \cup \overline{S_{4i}^2}$.

We shall only discuss the case $i = 1$. The other cases can be similarly disposed of.

So. for $i = 1$, the closed semigroup generated by $\overline{S_{21}}$ and $\overline{S_{41}^2}$ is given by,

$$\mathcal{S}_{24} = \left\{ \begin{pmatrix} 1 & 0 & 0 \\ 0 & \alpha & 1-\alpha \\ 0 & \beta & 1-\beta \end{pmatrix}, \begin{pmatrix} 0 & \delta & 1-\delta \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} : 0 \leq \alpha, \beta, \delta \leq 1 \right\}$$

Thus in this case the closed semigroup $\mathcal{S}$ generated by the support of μ is contained in $\mathcal{S}_{24}$. The kernel of $\mathcal{S}_{24}$ is given by

$$K_{24} = \left\{ \begin{pmatrix} 1 & 0 & 0 \\ 0 & \gamma & 1-\gamma \\ 0 & \gamma & 1-\gamma \end{pmatrix}, \begin{pmatrix} 0 & \delta & 1-\delta \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} : 0 \leq \gamma, \delta, \leq 1 \right\}$$

This is the same as the kernel K_4 in Case II mentioned earlier which we just get by renaming the states (1,2,3) as (2,3,1). But unlike Case II, here the closed semigroup is not itself the kernel. In that sense, it is rather like Case III.

Let us write $\mathcal{S}_{24} = \mathcal{S}_{22} \cup \mathcal{S}_{44}$

where

$$\mathcal{S}_{22} = \left\{ \begin{pmatrix} 1 & 0 & 0 \\ 0 & \alpha & 1-\alpha \\ 0 & \beta & 1-\beta \end{pmatrix} : 0 \le \alpha, \beta \le 1 \right\}$$

$$\mathcal{S}_{44} = \left\{ \begin{pmatrix} 0 & \delta & 1-\delta \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} : 0 \le \delta \le 1 \right\}$$

Also write $K_{24} = K_{22} \cup K_{44}$.
where

$$K_{22} = \left\{ \begin{pmatrix} 1 & 0 & 0 \\ 0 & \gamma & 1-\gamma \\ 0 & \gamma & 1-\gamma \end{pmatrix} : 0 \le \gamma \le 1 \right\}$$

$$K_{44} = \left\{ \begin{pmatrix} 0 & \delta & 1-\delta \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} : 0 \le \delta \le 1 \right\}$$

Then note :

(1) $\mathcal{S}_{22}$ is the closed semigroup $\overline{S_{21}}$ and K_{22} is its corresponding kernel. So, if $S_\mu \subseteq \mathcal{S}_{22}$, then the 2×2 case implies that μ^n converges unless $\mu = \delta \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$.

(2) $\mathcal{S}_{44} = K_{44}$.

Now, let $\phi : K_{24} \longrightarrow K_{24}$ be defined by

$$\phi \begin{pmatrix} 1 & 0 & 0 \\ 0 & \gamma & 1-\gamma \\ 0 & \gamma & 1-\gamma \end{pmatrix} = \begin{pmatrix} 0 & \gamma & 1-\gamma \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

and

$$\phi \begin{pmatrix} 0 & \gamma & 1-\gamma \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \gamma & 1-\gamma \\ 0 & \gamma & 1-\gamma \end{pmatrix}$$

Then ϕ is a bijection, $\phi = \phi^{-1}$ and $\phi(K_{22}) = K_{44}$.

Also, observe that if $x, y \in K_{24}$, $xy = y$ or $\phi(y)$ according as $x \in K_{22}$, or K_{44}. Hence, unlike Case III, here the matrix which post multiplies determines the entries of the product matrix. The case under consideration is covered by the following claim which we shall prove.

<u>Claim</u> : If $\mu(S_{24}) = 1$, $\mu(S_{22}) > 0$, $\mu(S_{44}) > 0$ then μ^n converges.

From now on, we assume that μ is as in the claim and we put $c = \mu(S_{33})$ so that $0 < c < 1$.

If Q is any limit point of μ^n, then once again, by Rosenblatt's result, $Q(K_{24}) = 1$. Now, denoting $\mu^n(S_{22})$ by α_n, we can prove exactly as in Case III that $\alpha_n \to \frac{1}{2}$ as $n \to \infty$. This explains : $Q(K_{22}) = Q(K_{44}) = \frac{1}{2}$.

Again , $\quad Q_1{*}Q_2(A) = \dfrac{Q_2(A) + Q_2(\phi(A))}{2} \quad$ for any Borel subset A of K_{24}.

Now, once again, using the Lemma 2.1 proved in section 2 of chapter 3 that the set of limit points of $(\mu^n)_{n\geq 1}$ is a group, let R be the identity of the group. Then, taking $Q_1 = R$ above, we get $Q_2(A) = Q_2(\phi(A))$ for any Borel subset A of S_{24}.

Thus if Q_1, Q_2 are two limit points of (μ^n) then using the fact that $Q_1 * Q_2 = Q_2 * Q_1$ we get

$$Q_2 = Q_1 * Q_2 = Q_2 * Q_1 = Q_1.$$

So, any two limit points are equal.

Alternatively, without using the lemma also, the above claim can be proved arguing in a similar fashion as in Case III.

This completes the proof of the claim.

Section 6 : Completion of the discussion

Having paved the way by discussing several cases in the previous two sections we now start our book keeping and discuss the convergence or otherwise of the sequence μ^n. At the outset let us recall that we can and shall assume that condition A below holds. When A fails section 4 already discusses the cases of convergence (See Remark 1 and Remark 2 of section 4).

(A) $$S(\mu) \subset S_1 \cup \overline{S_2} \cup \overline{S_3} \cup \overline{S_4^2} \cup \overline{S_5^0}.$$

In paticular,

(A i) $S(\mu) \cap S_4^1 = \emptyset, S(\mu) \cap (S_5 - S_5^0) = \emptyset,\ S(\mu) \cap S_6 = \emptyset.$

(A ii) $S(\mu) - \overline{S_{2i}} \neq \emptyset, 1 \leq i \leq 3.$

(A iii) $S(\mu) - \overline{S_{4i}^2} \neq \emptyset, 1 \leq i \leq 3.$

(A iv) $S(\mu) - \overline{S_{5i}^0} \neq \emptyset, 1 \leq i \leq 3.$

(A v) $S(\mu) - P \neq \emptyset.$

(A vi) $S(\mu) \cap S_4^3 = \emptyset.$

(A vii) $S(\mu) - \{N_i, M_i, N_{i+1}, M_{i+1}\} \neq \emptyset$ for $i = 1, 3, 5.$

1. If $S(\mu) \subseteq S_5^0$, then because of (A iii) and (A iv), $\mathcal{S} \cap K \neq \emptyset$.
2. If $S(\mu) \subseteq S_5^0 \cup S_4^2$, then because of (A iii), $\mathcal{S} \cap K \neq \emptyset$.

<u>Remark</u>. 1 and 2 above complete the discussion of convergence of $\{\mu^n\}$ when $S(\mu)$ is contained in $\overline{S_5^0} \cup \overline{S_4^2}$.

3. $S(\mu) \subseteq S_5^0 \cup S_4^2 \cup S_3$, then under asssumptions i)-vii) in Remark 1 of Section 4, we see that $\mathcal{S} \cap K \neq \emptyset$ unless one of the following four cases (a), (b), (c) or (d) holds:-

(a) $S(\mu) \subseteq S_{31}$ or $S(\mu) \subseteq S_{32}$ or $S(\mu) \subseteq S_{33}$.

This has been done in section 5, case (I).

(b)

i)$S(\mu) \cap S_{31} = \{N_4\}$ and $S(\mu) - S_{31} \subseteq S_{42}^2$,

ii)$S(\mu) \cap S_{31} = \{N_5\}$ and $S(\mu) - S_{31} \subseteq S_{43}^2$, iii)$S(\mu) \cap S_{32} = \{N_6\}$ and $S(\mu) - S_{32} \subseteq S^2_{43}$,

iv)$S(\mu) \cap S_{32} = \{N_1\}$ and $S(\mu) - S_{32} \subseteq S^2_{41}$,

v)$S(\mu) \cap S_{33} = \{N_2\}$ and $S(\mu) - S_{33} \subseteq S^2_{41}$,

vi)$S(\mu) \cap S_{33} = \{N_3\}$ and $S(\mu) - S_{33} \subseteq S^2_{42}$.

These cases have already been discussed in section 5, case(II).

(c)$S(\mu) \cap S_{3i} \neq \emptyset, S(\mu) \cap S^0_{5i} \neq \emptyset$ and $S(\mu) \subseteq S_{3i} \cup S^0_{5i}$ for some i, $i = 1, 2, 3$.

These cases have been discussed in section 5, case(III).

(d) $N_{2i-1}, M_{2i-1} \in S(\mu)$ and $S(\mu) - \{N_{2i-1}, M_{2i-1}\} \subseteq S^2_{4i}$ for some i, $i = 1, 2, 3$ or,

$N_{2i}, M_{2i} \in S(\mu)$ and $S(\mu) - \{N_{2i}, M_{2i}\} \subseteq S^2_{4i}$ for some i, $i = 1, 2, 3$.

These cases have been considered in section 5, case(II).

Remark. 1,2 and 3 above complete the discussion of convergence of $\{\mu^n\}$ when $S(\mu)$ is contained in $\overline{S^0_5} \cup \overline{S^2_4} \cup \overline{S_3}$.

4. $S(\mu) \subseteq S^0_5 \cup S^2_4 \cup S_3 \cup S_2$, then we can see that under (A i-vii) , $\mathcal{S} \cap K \neq \emptyset$ unless one of the following two cases (a) or (b) hold :-

(a)$\mathbf{e}_i \in S(\mu)$ and $S(\mu) - \{e_i\} \subseteq S_{3i} \cup S^0_{5i}$ and $S(\mu) \cap S_{3i} \neq \emptyset$ for some i, $i = 1, 2, 3$.

This case has also been done in section 5, case(III).

(b)$S(\mu) \cap S_{2i} \neq \emptyset$, $S(\mu) \cap \overline{S^2_{4i}} \neq \emptyset$ and $S(\mu) \subseteq \overline{S_{2i}} \cup \overline{S^2_{4i}}$.

This case has been done in section 5, case (IV).

Remark. 1,2,3 and 4 above complete the discussion of convergence of $\{\mu^n\}$ when $S(\mu)$ is contained in $\overline{S^0_5} \cup \overline{S^2_4} \cup \overline{S_3} \cup \overline{S_2}$.

5. $e_o \in S(\mu)$.

   In this case μ^n always converges as can be seen by going through all the previous cases successively. Firstly, in the four cases considered in section 5, e_o is already allowed in Cases I, III and IV and allowing it in Case II does not cause any problem. Secondly in all the cases considered above whenever convergence holds, it continues to hold even if e_o is present in $S(\mu)$. Finally in the few cases above where convergence failed, including e_o leads to convergence either by direct computation or by observing that $S \cap K \neq \emptyset$ or by appealing to the cases in section 5.

Remark. The arguments in sections 4, 5 and 6 above conclude the discussion of convergence of μ^n.

Section 7 : Main Theorem

From our discussions so far, it is clear that μ^n does not converge iff one of the following conditions hold :-

1. $S(\mu) \subset \overline{S^0_{5i}}$, for some $i = 1, 2, 3$
2. $S(\mu) \subset \overline{S^2_{4i}}$, for some $i = 1, 2, 3$
3. $S(\mu) = S^3_{41} = \{e_4\}$ or $S(\mu) = S^3_{42} = \{e_5\}$.
4. $S(\mu) \subset P_1$.

A clear picture will emerge if we make the following definition :-

Suppose S is a set of stochastic matrices of order 3. We say that S is a cyclic family if there are $S_1, \cdots, S_m \longrightarrow$ pairwise disjoint subsets of $\{1, 2, 3\}$ so that for any $1 \leq l \leq m$, for all $i \in S_l$, $\sum_{j \in S_{l+1}} p_{ij} = 1$ [Treat $m + 1$ as 1]. Here $\cup_1^m S_i$ need not be equal to $\{1, 2, 3\}$.

Condition (1) mentioned at the beginning of this section corresponds to $S_1 = \{2\}$, $S_2 = \{3\}$. Similar construction of S_1, S_2 hold for the other analogues. In condition (2), $S_1 = \{1\}$, $S_2 = \{2, 3\}$. Similarly, we can write down for the other analogues. In condition (3), if $S(\mu) = \{e_4\}$, $S_1 = \{1\}$, $S_2 = \{2\}$, $S_3 = \{3\}$. & if $S(\mu) = \{e_5\}$, $S_1 = \{1\}$, $S_2 = \{3\}$, $S_3 = \{2\}$. However in condition (4) mentioned above there is no such cyclic family.

The conclusion mentioned at the beginning of the section may now be succinctly stated as follows :-

Theorem 7.1 :- Suppose μ is a probability on the set of stochastic matrices of order 3. Then μ^n does not converge to a limit if either $S(\mu)$ is cyclic or $S(\mu) \subset \{\boldsymbol{e_1}, \boldsymbol{e_2}, \boldsymbol{e_3}\}$.

It is interesting to note that in all the cases of nonconvergence supports of μ and μ^2 are disjoint. But of course the converse is clearly false.

We conclude this section with a few remarks :

Remark 1 :- Following a suggestion in (p.160 of [29]), it would be interesting to find conditions for the limit of (μ^n) – when it exists – to be discrete, singular or absolutely continuous, under suitable parametrization.

Remark 2 :- In all the four cases of non-convergence mentioned in section 7, it is easy to see that we have finitely many limit points for the sequence μ^n (see Theorem 3.4 in [22]). In fact, except case (3), we have only two limit points for the other cases. In case of (3), we have three limit points for each of the subcases :- $\mu = \delta_{\boldsymbol{e_4}}$ or $\mu = \delta_{\boldsymbol{e_5}}$.

Remark 3 :- It is clear from Remark 2 that in any case, $\frac{1}{n}\sum_1^n \mu^k$ converges. This is of course well known [29].

Remark 4 :- When $d = 2$ the only case when $\{\mu^n\}_{n\geq 1}$ does not converge is given by $\mu = \delta_{\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}}$. In that case, $S_1 = \{1\}$ and $S_2 = \{2\}$ form the two cyclically moving subclasses and this is the only case when $S(\mu)$ is cyclic.

Remark 5 : For μ on $S_d\,(d \geq 2)$, if $S_d(\mu) \cap K_d \neq \emptyset$, μ^n converges weakly. This follows from our discussions in section 2.

Section 8 : Non-i.i.d. Case (2×2)

In this section, we concentrate on S_2 – the 2×2 stochastic matrices and consider a sequence of probabilities, $(\mu_n)_{n\geq 1}$ on S_2. Let $\nu_n = \mu_1 * \cdots * \mu_n$. We are interested in conditions for the convergence of the sequence $(\nu_n)_{n\geq 1}$. This problem was raised by A. Mukherjea. We do not have a complete answer. We shall be content with making some remarks. In our discussion, I denotes the identity matrix $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and T denotes the matrix $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$.

First, we make some simplifications by removing uninteresting cases. We can and shall assume that for no n, $\mu_n = \delta_I$ because such μ_n's can be deleted without affecting the discussion of convergence (and the limit too). If after some stage, say, $n \geq N$, we have $\mu_n = \delta_T$, then convergence holds iff ν_n is symmetric under permutation of the two columns and in that case, $\nu_n = \nu_N$ for all $n \geq N$. Thus, we can and shall assume that μ_n is not eventually δ_T. Again, if for infinitely many n, $\mu_n = \delta_T$, then we can delete all such μ_n's and call the resulting sequence $\tilde{\mu}_n$. Denote $\tilde{\nu}_n = \tilde{\mu}_1 * \tilde{\mu}_2 * \cdots * \tilde{\mu}_n$. It is easy to see that ν_n converges iff $\tilde{\nu}_n$ converges to some λ which is invariant under the permutation of the columns. Thus, we can and shall assume that $\mu_n = \delta_T$ for only finitely many n. A little reflection shows that we can as well assume that for no n, $\mu_n = \delta_T$. Since δ_T appears finitely many times, only for values of $n \leq N$ say,we can replace the sequence $(\mu_n)_{n \geq 1}$ by $\nu_N, \mu_{N+1}, \cdots$ as far as the convergence question is concerned.

Thus, we shall assume from now on that for no n, $\mu_n = \delta_I$ and for no n, $\mu_n = \delta_T$. Suppose each μ_n is concentrated only on $\{I, T\}$. Let $\mu_i(I) = a_i$ and $\mu_i(T) = 1 - a_i$. Also, let $\nu_i(I) = \alpha_i$ and $\nu_i(T) = 1 - \alpha_i$.

Then,

$$\alpha_{i+1} = \alpha_i a_{i+1} + (1 - \alpha_i)(1 - a_{i+1}).$$

Setting $\epsilon_i = \alpha_i - \frac{1}{2}$ for all i, we get,

$$\frac{1}{2} + \epsilon_{i+1} = (\frac{1}{2} + \epsilon_i)a_{i+1} + (\frac{1}{2} - \epsilon_i)(1 - a_{i+1})$$

which implies $\epsilon_{i+1} = \epsilon_i[2a_{i+1} - 1]$ for all i. Now, since $0 < a_i < 1$ for all i, we have, $| 2a_{i+1} - 1 | < 1$ for all i. So,

$$| \epsilon_{i+1} | \leq | \epsilon_i | \quad \forall i.$$

Hence, $| \epsilon_i |$ decreases. Let it decrease to c.

In case $c = 0$, $\alpha_n \to \frac{1}{2}$ and consequently, ν_n converges to the measure putting mass $\frac{1}{2}$ at each of I and T. One can verify that if a_n's are bounded away from 0 and 1, then $c = 0$.

In case $c > 0$, two cases arise. The first case is when $a_n > \frac{1}{2}$ after some stage, say, N. Then clearly ϵ_n converges to $+c$ or $-c$ depending on the sign of ϵ_N. Thus, ν_n does converge. The second case is when $a_n < \frac{1}{2}$ for infinitely many n. Then, the signs of ϵ_n also change infinitely many times and hence

ϵ_n has two limit points $+c$ and $-c$. In this case, ν_n does not converge, but it has exactly two limit points.

Finally, we wish to make a comment regarding the periodic case.

Since we are assuming that none of our probabilities is δ_T, clearly, $\mu_1 * \cdots * \mu_k \neq \delta_T$. Hence, $\lim_n(\mu_1 * \cdots * \mu_k)^n$ exists. Denote it by λ. Obviously, $\lim_n \nu_n$ exists iff we have $\lambda * \mu_i = \lambda$ for $1 \leq i \leq k$. Under our assumptions, $\lim_n \mu_i{}^n$ exists. Let us denote it by λ_i. The condition $\lambda * \mu_i = \lambda$ implies, in particular, that $\lambda * \lambda_i = \lambda$.

An interesting situation obtains if we assume that none of the μ_i's are concentrated on $\{I, T\}$ alone. Then, each of the λ_i's as well as λ are concentrated on the Kernel. Consequently, $\lambda * \lambda_i = \lambda_i$. Thus, in this situation, if ν_n converges, then we must necessarily have, $\lambda_i = \lambda$ for all i.

Before concluding this section, we must mention that Hognas and Mukherjea [17] provided several conditions for the convergence of ν_n.

Section 9 : Bernoulli Convolutions

As in the previous section, we consider S_2 – the space of 2×2 stochastic matrices. We know that if μ is a probability on S_2 and μ is not the point mass at the matrix $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, then μ^n converges (see Theorem 1.1). We now specialize to μ concentrated on the subset of S_2 consisting of matrices $\begin{bmatrix} a & 1-a \\ b & 1-b \end{bmatrix}$, with $a > b$. Let D_* be the subset of the unit square consisting of points below the diagonal. More precisely, $D_* = \{(a, b) : 0 \leq b < a \leq 1\}$. Clearly, we can think of D_* as the above set of matrices by identifying (a, b) with $\begin{bmatrix} a & 1-a \\ b & 1-b \end{bmatrix}$. Thus μ is a probability on D_*. μ^n converges to a probability λ concentrated on the Kernel K, namely, the set of matrices $\begin{bmatrix} x & 1-x \\ x & 1-x \end{bmatrix}$. With this identification in mind, the equation $\lambda * \mu = \lambda$ can be written as

$$\lambda[0, x] = \int_{D_*} \lambda[0, \frac{x-b}{a-b}] \, d\mu(a, b) \tag{9.1}$$

We shall consider the iterated function system on $[0, 1]$ given by (D_*, μ) where we identify $(a, b) \in D_*$ with the function $f_{a,b} : [0, 1] \mapsto [0, 1]$ given by $f_{a,b}(x) = b + (a - b)x, \quad x \in [0, 1]$. Note that $f_{a,b}(0) = b$ and $f_{a,b}(1) = a$ so that the function $f_{a,b}$ determines the point (a, b). More precisely, we have the

Markov process outlined in chapter 1 : If we are at $x \in [0,1]$ then we select $(a,b) \in D_*$ according to μ and move to $f_{a,b}(x)$.

Theorem : The iterated function system (D_*, μ) on $[0,1]$ has unique invariant measure and it is the probability λ described above.

Proof.

The splitting condition holds and hence the system has a unique invariant probability – call it ν. So ν satisfies

$$\nu(B) = \int p(y,B)d\nu(y)$$

where B is a Borel set and the transition function of the Markov process is

$$p(y,B) = \mu\{(a,b) \in D_* : f_{a,b}(y) \in B\}$$

Note that $f_{a,b}(y) \leq x$ iff $y \leq \dfrac{x-b}{a-b}$. Thus, we have for every x with $0 \leq x \leq 1$;

$$\nu[0,x] = \int \mu\{(a,b) \in D_* : f_{a,b}(y) \leq x\} d\nu(y)$$

$$= \nu \otimes \mu\{(y,(a,b)) : f_{a,b}(y) \leq x\}$$

$$= \nu \otimes \mu\{(y,(a,b)) : y \leq \frac{x-b}{a-b}\}$$

$$= \int \nu[0, \frac{x-b}{a-b}] \; d\mu(a,b)$$

Conversely, any ν satisfying this equation for every $x \in [0,1]$ is an invariant measure. By uniqueness of invariant measure and equation (9.1) above, we get the result.

Let us now further specialize μ to a probability concentrated at two points (a_1,b_1) and (a_2,b_2). The problem discussed in [22] is the nature of λ. From Dubins and Freedman [12] or Mukherjea [22], λ has to be pure, either singular or absolutely continuous.

Now fix $(a_1,b_1),(a_2,b_2) \in D_*$ and consider $f_1(x) = b_1 + (a_1 - b_1)x$ and $f_2(x) = b_2 + (a_2 - b_2)x$ on $[0,1]$. f_1 and f_2 have fixed points $\alpha = \dfrac{b_1}{1-(a_1-b_1)}$

and $\beta = \dfrac{b_2}{1-(a_2-b_2)}$ respectively. It is also easy to see that for $x < \alpha$, the orbit $f_1^n(x)$ increases to α whereas for $x > \alpha$, the orbit $f_1^n(x)$ decreases to α. Similar remark holds for f_2 and β. Let us now assume that for specificness, $0 \leq \alpha \leq \beta \leq 1$. If $\alpha = \beta$, then for every point x, the orbit of x under any applications of f_1 or f_2 at each stage, converges to α and the limiting probability λ is point mass at α.

Let us now assume that $0 \leq \alpha < \beta \leq 1$. Then orbits being as described above, any invariant measure is concentrated on the interval $[\alpha, \beta]$, which is left invariant by both f_1 and f_2. Thus, λ is absolutely continuous iff the invariant measure for the iterated function system on the interval $[\alpha, \beta]$ given by

$$f_1(x) = b_1 + (a_1 - b_1)x \quad \text{w.p.} \quad p$$

$$f_2(x) = b_2 + (a_2 - b_2)x \quad \text{w.p.} \quad 1-p$$

is absolutely continuous.

Let us define $\phi : [0,1] \mapsto [0,1]$ by $\phi(x) = \alpha + (\beta - \alpha)x$. Then f_1, f_2 on $[\alpha, \beta]$ are conjugate to g_1 and g_2 on $[0,1]$ respectively where

$$g_1(x) = (a_1 - b_1)x$$

$$g_2(x) = (a_2 - b_2)x + [1 - (a_2 - b_2)]$$

Thus, we have,

Theorem : λ is absolutely continuous iff the (unique) invariant measure for the iterated function system

$$g_1(x) = (a_1 - b_1)x \quad \text{w.p.} \quad p$$

$$g_2(x) = (a_2 - b_2)x + [1 - (a_2 - b_2)] \quad \text{w.p.} \quad 1-p$$

on $[0,1]$ is absolutely continuous.

Let us now further specialize to the case $p = \frac{1}{2}$ and $a_1 - b_1 = a_2 - b_2 = t$, say. Thus our system consists of the two functions tx and $tx + (1-t)$, each with probability $\frac{1}{2}$. As noted already in Dubins and Freedman, the invariant measure for this system is nothing but the distribution of $(1-t)\sum_{n=0}^{\infty} t^n \eta_n$ where η_n's are i.i.d. taking values 0 and 1, with probability $\frac{1}{2}$ each. Thus, we have,

Theorem : Let μ be the probability giving equal mass to the matrices

$$\begin{bmatrix} a_1 & 1-a_1 \\ b_1 & 1-b_1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} a_2 & 1-a_2 \\ b_2 & 1-b_2 \end{bmatrix}$$

where $a_1 > b_1, a_2 > b_2$ and $a_1 - b_1 = a_2 - b_2 = t$, say. Then, the limit of μ^n is absolutely continuous (where the limit on the Kernel is identified as probability on $[0,1]$) iff the law of $\sum_{n=0}^{\infty} t^n \epsilon_n$ is absolutely continuous where ϵ_n's are i.i.d. $+1$ or -1 with equal probabilities.

Proof. Just note that $\eta_n = \frac{1}{2}(\epsilon_n + 1)$.

So, let us denote $\sum_{n=0}^{\infty} t^n \epsilon_n$ by X_t, where ϵ_n's are, as above, i.i.d. ± 1 with equal probabilities. The problem of deciding whether X_t has an absolutely continuous distribution is known in the literature as the problem of Bernoulli Convolutions or Erdos problem. This has a long history. Since we are enthused about it, we decided to give some references and review the status of the problem. In Kreshner and Wintner [20], they observed that the range of X_t is either a nowhere dense perfect set of Lebesgue measure zero or the range consists of the entire interval $[\frac{-1}{1-t}, \frac{1}{1-t}]$, according as $t < \frac{1}{2}$ or $t \geq \frac{1}{2}$. In particular,for $t < \frac{1}{2}$, X_t is singular. For $t = \frac{1}{2}$, it is easy to see that X_t is uniform $[-2, 2]$.

In what follows, we shall restrict our attention to $\frac{1}{2} < t < 1$. In Pisot [26] and Vijayraghavan [34], certain algebraic numbers were studied which were named as P.V. numbers by Salem [30]. A number $\lambda > 1$ is a P.V. number if it is the real root of a polynomial with integer co-efficients and leading co-efficient unity whose other roots are strictly smaller than one in modulus. Erdos [13] showed that if t is reciprocal of a P.V. number then the characteristic function of X_t does not vanish at infinity and hence (by purity) X_t is singular. In Salem [31], it was shown that if the characteristic function of X_t does not vanish at infinity, then t must be the reciprocal of a P.V. number. In fact the only known examples of singular X_t are when t is reciprocal of a P.V. number. In Siegel [32], a detailed study of the P.V. numbers is made and several such numbers in $(\frac{1}{2}, 1)$ can be found. In Pisot and Dufresnoy [27], it is shown that the smallest limit point of P.V. numbers is the root of $\alpha^2 - \alpha - 1 = 0$. Reciprocal of the P.V. number given by the root of the above equation belongs to $(\frac{1}{2}, \frac{3}{4})$. This shows that there are infinitely many reciprocals of P.V. numbers in $(\frac{1}{2}, 1)$. Thus there are infinitely many $t > \frac{1}{2}$ for which X_t is singular.

In the other direction, Wintner [35] showed, among other things, that X_t is absolutely continuous for $t = \frac{1}{2^{\frac{1}{2}}}, \frac{1}{2^{\frac{1}{3}}}, \frac{1}{2^{\frac{1}{4}}}, \cdots$. This is easy to verify directly also. Erdos [14] showed that for almost every t in an interval near 1, X_t is absolutely continuous. Finally, Solomyak [33] showed that for a.e. $t \in (\frac{1}{2}, 1)$, X_t has an L^2 density thus settling the problem. Kahane and Salem [18] and Garcia [16] provided criteria for absolutely continuity in a more general context $(\sum r_n \epsilon_n)$.

Inspite of all this, so far the only specific examples of singularity are when t is reciprocal of a P.V. number and those of absolute continuity are when t is an integral root of $\frac{1}{2}$. One should perhaps try t which satisfy an equation $t^n + t^{n+1} = 1$ or rational roots of $\frac{1}{2}$. We had no success.

In Mukherjea and Tserpes[25], Mukherjea[22] or Mukherjea and Ratti [24], several results regarding absolute continuity/singularity of the limit of μ^n are given . The problem of deciding absolute continuity or singularity of the limiting measure, of course, goes back to Rosenblatt (p.160 of [29]). What we have done above is to relate the problem with random iterations. Further the Rosenblatt problem when μ puts equal masses at the matrices $\begin{bmatrix} a_1 & 1-a_1 \\ b_1 & 1-b_1 \end{bmatrix}$ and $\begin{bmatrix} a_2 & 1-a_2 \\ b_2 & 1-b_2 \end{bmatrix}$, where $a_1 > b_1, a_2 > b_2$ and $a_1 - b_1 = a_2 - b_2$, is shown to be equivalent to that of Erdos problem on Bernoulli Convolutions. Thus all results there can be translated to the Rosenblatt problem. If $a_1 - b_1 \neq a_2 - b_2$, or if we have unequal probabilities, then we have to go beyond symmetric Bernoulli Convolutions and that is a different story.

CHAPTER - III

Convolution powers of probabilities on Stochastic matrices

Summary

This chapter has five sections. In the first section, we introduce the problem of convergence of convolution powers of a probability on the space of $d \times d$ stochastic matrices. In section 2, we state and prove a wellknown theorem on the set of limit points of a convolution sequence that will be needed later. In the third section we fix our notation and recall the structure of kernels of subsemigroups of $d \times d$ stochastic matrices following Rosenblatt. We slightly modify this to obtain a decomposition of the kernel which extends to the subsemigroup itself. In the following section, we state and prove our main theorem, namely the sequence of convolution powers of a given probability converges if and only if the sequence of convolution powers of an associated probability on a finite permutation group converges. We conclude with some remarks in the last section.

Section 1 : Introduction

In this chapter, we shall look at the behaviour of convolution powers of probabilities on general $d \times d$ stochastis matrices.

Let S_d denote the semigroup of $d \times d$ stochastic matrices, with usual topology. Let μ be a probability on S_d and μ^n denote the n-fold convolution of μ. We shall provide a necessary and sufficient condition for the convergence of the sequence μ^n. Roughly speaking μ^n converges unless some kind of periodicities are present. In a different direction, Dhar and Mukherjea [11] showed the following : Let μ be a probability on the multiplicative semigroup of $d \times d$ matrices with non-negative entries. Suppose that the sequence $(\mu^n)_{n \geq 1}$ is tight and that the support of μ contains a matrix with at least $d-1$ positive diagonal entries. Then μ^n coverges.

Section 2 : A Well Known Lemma

The following lemma is well known – (see Hognas and Mukherjea [17], p.91-92, Theorem 2.13(ii)). Since their result is for more general semigroups and their proof uses the machinery of semigroup theory, we decided to provide a proof for the situation we have in mind.

Lemma 2.1.

Let $\mu \in P(S_d)$. Let $\mathcal{G} = \{\lambda : \lambda \text{ is a limit point of } (\mu^n)\}$. Then $\mathcal{G}$ is a commutative group under multiplication.

Proof.

If $\mu^{n_i} \to \lambda_1$ and $\mu^{m_i} \to \lambda_2$ then by continuity of convolution, $\mu^{n_i+m_i} \to \lambda_1 * \lambda_2$ and $\mu^{m_i+n_i} \to \lambda_2 * \lambda_1$, showing that $\mathcal{G}$ is closed under convolution and also that convolution is commutative on $\mathcal{G}$.

If $\lambda_1, \lambda_2 \in \mathcal{G}$, we can get a $\lambda_3 \in \mathcal{G}$ so that $\lambda_1 = \lambda_3 * \lambda_2$. To see this, let $\lambda_1 = \lim_i \mu^{n_i}$ and $\lambda_2 = \lim_i \mu^{m_i}$. By taking a subsequence of n_i if necessary, we can and shall assume that $n_i - m_i \uparrow \infty$. By taking subsequences of both n_i and m_i, we can and shall assume that $\mu^{n_i-m_i}$ converges to say λ_3. This serves our purpose. Thus it shows that for any $\lambda \in \mathcal{G}$, we have, $\mathcal{G} * \lambda = \lambda * \mathcal{G} = \mathcal{G}$.

Fix $\lambda \in \mathcal{G}$. Using $\lambda * \mathcal{G} = \mathcal{G}$, get $\eta \in \mathcal{G}$ so that $\lambda * \eta = \lambda$. Now, for any $\alpha \in \mathcal{G}$, using $\mathcal{G} * \lambda = \mathcal{G}$, we get $\beta \in \mathcal{G}$ so that $\beta * \lambda = \alpha$ to see that $\alpha * \eta = \beta * \lambda * \eta = \beta * \lambda = \alpha$. Thus η is the identity element. The observation of earlier paragraph gives inverse elements. This completes the proof.

Section 3 : Kernel Structure

Let $\mathcal{S}$ be a closed subsemigroup of S_d. Then, S_d being compact, $\mathcal{S}$ will be compact. So, $\mathcal{S}$ has a kernel $\mathcal{K}$ which is a minimal two sided ideal and matrices in $\mathcal{K}$ are of minimal rank in $\mathcal{S}$. For example, if $\mathcal{S} = S_d$ itself, then $\mathcal{K}$ precisely consists of all rank one matrices. Not only that, if $\mathcal{S}$ contains a rank one matrix from S_d, then also, all the matrices in $\mathcal{K}$ will be of rank one.

To formulate our theorem, we need to know the structure of kernels. To this end we fix some notation. So let $\mathcal{S}$ be a closed subsemigroup of S_d with kernel $\mathcal{K}$. Let the rank of matrices in $\mathcal{K}$ be r. We shall now give a description of matrices in the kernel.

Suppose that $C_1, C_2, \cdots, C_r, T$ is a partition of $\{1, 2, \cdots, d\}$ with $|C_i| = d_i$ for $1 \leq i \leq r$ and $|T| = d_0$. Here $|C|$ denotes the number of elements in the set C. Clearly $\sum d_i = d$. To avoid notational complications we assume that $C_1 = \{j : 1 \leq j \leq d_1\}$ and in general $C_i = \{j : \sum_1^{i-1} d_l < j \leq \sum_1^i d_l\}$ and $T = \{j : \sum_1^r d_l < j \leq d\}$. Suppose that $v_1, v_2, \cdots, v_r$ are probability vectors (column) where v_i is of length d_i. By $K(v_1, v_2, \cdots, v_r)$ we denote the block diagonal matrix for which the $C_i \times C_i$ block consists of identical rows each row being v_i'. If π is a permutation of $\{1, 2, \cdots, r\}$ we denote by πK the block matrix with $C_i \times C_j$ block being 0 if $i \neq \pi(j)$ while the $C_{\pi(j)} \times C_j$ block consists of $d_{\pi(j)}$ identical rows each row being v_j'. Thus, for example, if $r = 2$, $d_1 = 2$, and $d_2 = 3$ and v_1, v_2 are probability vectors of size 2 and 3 respectively and π is the permutation interchanging 1 and 2, then

$$K(v_1, v_2) = \begin{pmatrix} v_1' & 0 \\ v_1' & \\ & v_2' \\ 0 & v_2' \\ & v_2' \end{pmatrix} \qquad \text{and} \qquad \pi K = \begin{pmatrix} 0 & v_2' \\ & v_2' \\ v_1' & \\ v_1' & 0 \\ v_1' & \end{pmatrix}$$

Let W be a stochastic matrix of order $d_0 \times (d - d_0)$ for which the columns other than $1, d_1 + 1, d_1 + d_2 + 1, \cdots, d_1 + \cdots + d_{r-1} + 1$ are zero columns. Such matrices will be called weight matrices.

Here then is the structure of the matrices in the kernel as developed by Rosenblatt [29]. After a suitable renaming of rows and columns , if necessary – there is a partition $C_1, C_2, \cdots, C_r, T$ as above such that every matrix in the kernel of $\mathcal{S}$ has the form $\begin{pmatrix} \pi K & 0 \\ W\pi K & 0 \end{pmatrix}$ for some K and weight matrix W and some permutation π of $\{1, 2, \cdots, r\}$ as above. Since, the block matrix πK consists of r blocks, each of rank 1 and since the rows corresponding to the states in T are convex linear combinations of the rows above it, this justifies that the matrices of above description are indeed of rank r. In this description of the kernel, T is allowed to be empty. In such a case, weight matrices do not appear and thus every matrix in the kernel is of the form πK.

Thus if $\mathcal{K}$ is the kernel of $\mathcal{S}$ we can partition $\mathcal{K}$ as $\mathcal{K} = \bigcup \mathcal{K}_\pi$ where $\mathcal{K}_\pi$ consists of all those matrices in $\mathcal{K}$ that have the form $\begin{pmatrix} \pi K & 0 \\ W\pi K & 0 \end{pmatrix}$ as above. Note that since K is block diagonal, πK uniquely determines π. Of course some of the $\mathcal{K}_\pi$ may be empty. Note that if $M_1 \in \mathcal{K}_\pi$ and $M_2 \in \mathcal{K}_\sigma$ then $M_1 M_2 \in \mathcal{K}_{\pi\sigma}$. This shows that those π for which $\mathcal{K}_\pi \neq \emptyset$ is a subgroup of the permutation group on $\{1, 2, \cdots, r\}$. This subgroup will be denoted by H. The identity permutation will be denoted by e.

Given $\mathcal{S}$, the partition stated above is not uniquely determined. For instance, let $d = 2$, and $\mathcal{S}$ consist of the single matrix $\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$. Thus , in this case, $\mathcal{S} = \mathcal{K}$. It is easy to see that $C_1 = \{1\}$, $T = \{2\}$ as well as $C_1 = \{1, 2\}, T = \emptyset$ are possible choices for the partition mentioned above. To avoid this ambiguity (and to get a nice description of $\mathcal{S}$ later on) we shall fix our partition as follows. First observe that if $m \in T$ then for every matrix in $\mathcal{K}_e$ the row corresponding to m is of the form $\langle w_1 v_1', w_2 v_2', \cdots, w_r v_r', 0\rangle$ where v_i' s are the vectors appearing in the $C_i \times C_i$ block, and $\langle w_1, w_2, \cdots, w_r\rangle$ is a probability vector. If it so happens that there is an i with $1 \leq i \leq r$ such that the row corresponding to m of every matrix in $\mathcal{K}_e$ is of the above form with $w_i = 1$ (and other w_j s being necessarily 0) then we shall incorporate this state m in the class C_i itself. From now on we assume that such a partition is fixed in the description of the kernel. Note that with this choice we have the following: Given any $m \in T$ there is a matrix in $\mathcal{K}_e$ such that the row corresponding to m is of the above type with no single w_j being one. For future reference we refer to this property as $(*)$. With such a description of the kernel stated as above, we now proceed to observe that matrices in $\mathcal{S}$ have similar block structure.

Firstly, given $A \in \mathcal{S}$ there is a permutation $\pi \in H$ such that $C_i \times C_j$ block of A is 0 for $i \neq \pi(j)$. Indeed fix $M \in \mathcal{K}_e$. Suppose that $AM \in \mathcal{K}_\pi$, say. Thus for $i \neq \pi(j)$ the $C_i \times C_j$ block in AM is 0. The structure of M now implies that the $C_i \times C_j$ block in A itself must be 0.

Secondly, for any $A \in \mathcal{S}$, $C_i \times T$ block is 0. Indeed, take any state $s \in C_i$ and any state $t \in T$. We need to show that $a_{st} = 0$. Fix a matrix M in $\mathcal{K}_e$ having property $(*)$ above corresponding to the state t, say $w_p \neq 0$ and $w_q \neq 0$. Either $\pi(p) \neq i$ or $\pi(q) \neq i$. To fix ideas let us say that $\pi(p) \neq i$. In particular, in the row of M corresponding to t there is an index $u \in C_p$ such that $m_{tu} \neq 0$. As a consequence, if $a_{st} > 0$ then $a_{st}\, m_{tu} > 0$ implying that $C_i \times C_p$ block of AM is nonzero leading to a contradiction.

Thirdly ,if $A \in \mathcal{S}$, $M \in \mathcal{K}_e$ and $AM \in \mathcal{K}_\pi$ then, $MA \in \mathcal{K}_\pi$ as well. To see this if $MA \in \mathcal{K}_\sigma$, then MAM should be in both $\mathcal{K}_\sigma$ and $\mathcal{K}_\pi$ leading to the conclusion that $\sigma = \pi$.

Finally, let us also observe that for any $A \in \mathcal{S}$ the permutation π mentioned in the first observation, above does not depend on the choice of the matrix from $\mathcal{K}_e$. If M_1 and M_2 are in $\mathcal{K}_e$ and $AM_1 \in \mathcal{K}_\pi$, then a similar argument as above using M_1AM_2 shows that $AM_2 \in \mathcal{K}_\pi$.

This last observation allows us to define a map Π on $\mathcal{S}$ as follows: For $A \in \mathcal{S}$, $\Pi(A)$ is the unique permutation from H obtained above. Moreover if $M \in \mathcal{K}_\sigma$ and $\Pi(A) = \pi$,then $AM \in \mathcal{K}_{\pi\sigma}$. This has the interesting and useful consequence that $\Pi(AB) = \Pi(A)\Pi(B)$. Thus $\mathcal{S}$ is also partitioned into $\mathcal{S} = \bigcup_{\pi \in H} \mathcal{S}_\pi$ in a natural way.

For those familiar with the Rees-Suschekewitsch decomposition, the above conclusion can be alternatively restated as follows: If $E \times H \times F$ is the Rees-Suschekewitsch decomposition of the kernel $\mathcal{K}$, where, as usual, H is the group factor, then the homomorphism Π defined on $\mathcal{K}$ to H can be extended to all of the semigroup S.

Section 4 : Main Theorem

Now suppose that μ is a probability on S_d. Let $\mathcal{S}$ be the closed semigroup generated by the support of μ. From now on the partition $C_1, C_2, \cdots, C_r, T$ as well as the group H refer to this semigroup. Let $\tilde{\mu}$ be the probability on H induced by μ via the map Π. In other words, $\tilde{\mu}(\pi) = \mu(\mathcal{S}_\pi)$ for $\pi \in H$. As in chapter 2, convergence of probabilities is understood as weak convergence. Here is the main theorem.

Theorem: μ^n converges on S_d iff $\tilde{\mu}^n$ converges on H.

<u>Proof</u>:

First observe that the map Π has the property that $\Pi(A)\Pi(B) = \Pi(AB)$. Since $\tilde{\mu} = \mu\Pi^{-1}$, it immediately follows that $\tilde{\mu}^n = \mu^n\Pi^{-1}$. As a consequence – Π being a continuous map – the *only if* part of the theorem follows.

We shall now prove the *if* part of the theorem. So let us assume that $\tilde{\mu}^n$ converges. Then a simple argument shows that for each $\pi \in H$, $\tilde{\mu}^n(\pi) \rightarrow 1/|H|$. We proceed to show that μ^n converges. We use, in what follows, the fact that any limit point of this sequence is concentrated on the kernel $\mathcal{K}$, (see Rosenblatt's Lemma in section 2 of chapter 2). Since the space S_d is compact, we only need to show that there is only one limit point.

Case 1: $T = \emptyset$.

First suppose that H consists of only one element namely the identity element. A simple calculation shows that, in this case, for any two matrices M_1 and M_2 in the kernel we have $M_1M_2 = M_2$. As a consequence, for any two limit points ν_1 and ν_2, we have $\nu_1 = \nu_2 * \nu_1$ and $\nu_1 * \nu_2 = \nu_2$. Since $\nu_1 * \nu_2 = \nu_2 * \nu_1$, we conclude that $\nu_1 = \nu_2$.

Next assume that $|H| = m > 1$. As remarked earlier, the hypothesis implies that $\mu^n(\mathcal{S}_\pi) \to 1/m$ for each $\pi \in H$. As a consequence for any limit point ν we have $\nu(\mathcal{K}_\pi) = 1/m$ for each $\pi \in H$. Observe that H acts on $\mathcal{K}$ in an obvious way, namely, if $\pi \in H$ then we have the map $\phi_\pi : \mathcal{K} \to \mathcal{K}$ defined by $\phi_\pi(M) = \pi M$, that is, if $M = \sigma K$ for $K \in \mathcal{K}_e$ then $\phi_\pi(M) = \pi\sigma K$. This is easily seen to be a group action, that is, $\phi_\pi \, o \, \phi_\sigma = \phi_{\pi\,\sigma}$. Also observe that if $x \in \mathcal{K}_\pi$ and $y \in \mathcal{K}_\sigma$ then the product xy does not depend on x and, in fact, $xy = \phi_\pi(y)$. As a consequence, if ν_1 and ν_2 are two limit points of the sequence μ^n, then we have for any Borel $B \subset \mathcal{K}$,

$$\nu_1 * \nu_2\,(B) = \int_{\mathcal{K}} \nu_2\{y : xy \in B\}\,\nu_1(dx) = \frac{1}{m}\sum_{\pi\in H} \nu_2\,(\phi_\pi^{-1}(B)) \qquad (1)$$

In particular, for any $\sigma \in H$ we have

$$\nu_1 * \nu_2\,(\phi_\sigma^{-1}(B)) = \frac{1}{m}\sum_{\pi\in H} \nu_2\,(\phi_{\sigma\pi}^{-1}(B)) = \frac{1}{m}\sum_{\pi\in H} \nu_2\,(\phi_\pi^{-1}(B)) = \nu_1 * \nu_2\,(B)$$

Now taking ν_1 to be the identity of the group of limit points (see Lemma 2.1, section 2) we get that,

$$\nu_2\,(B) = \nu_2\,(\phi_\sigma^{-1}(B)) \quad \text{for any} \quad \sigma \in H$$

Using this in equation (1) above, we see that $\nu_1 * \nu_2 = \nu_2$ for any two limit points and so we once again have $\nu_1 = \nu_2 * \nu_1 = \nu_1 * \nu_2 = \nu_2$.

Case 2: $T \neq \emptyset$.

We again start with the special case when H is singleton. Consider a sequence $(X_k)_{k\geq 1}$ of i.i.d random matrices with common distribution μ. Set $Y_k = \prod_1^k X_i$. Thus the distribution of Y_k is μ^k. The structure of $\mathcal{S}$ now implies (recall that H is a singleton) that X_k and Y_k have the form

$$X_k = \begin{pmatrix} X_k^1 & 0 & . & . & . & 0 & 0 \\ 0 & X_k^2 & 0 & . & . & 0 & 0 \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ 0 & 0 & . & . & . & X_k^r & 0 \\ U_k^1 & U_k^2 & . & . & . & U_k^r & U_k^{r+1} \end{pmatrix}; Y_k = \begin{pmatrix} Y_k^1 & 0 & . & . & . & 0 & 0 \\ 0 & Y_k^2 & 0 & . & . & 0 & 0 \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ 0 & 0 & . & . & . & Y_k^r & 0 \\ V_k^1 & V_k^2 & . & . & . & V_k^r & V_k^{r+1} \end{pmatrix}$$

For $1 \leq i \leq r$, X_k^i , Y_k^i are of order $d_i \times d_i$; U_k^i , V_k^i are of order $d_0 \times d_i$; U_k^{r+1} , V_k^{r+1} are of order $d_0 \times d_0$.

Denoting by e^i the $d_i \times 1$ column vector consisting of ones, and noting that $X_k^i e^i = e^i$, we get

$$V_{k+1}^i e^i = V_k^i e^i + V_k^{r+1} U_{k+1}^i e^i \quad \text{for} \quad 1 \leq i \leq r$$

This shows that $V_k^i e^i$ increases (entrywise)with k and has an almost sure limit W^i, say. Since we have $\sum_i V_k^i e^i = 1$ (entrywise) for every k it now follows that $V_k^{r+1} e^{r+1}$ decreases (entrywise) as k increases and hence has a limit , say, W^{r+1}. Since the sequence μ^k has convergent subsequences, and since every limit point is concentrated on $\mathcal{K}$, implying that there is a subsequence of $V_k^{r+1} e^{r+1}$ which converges to the zero vector in distribution. Since the sequence has been shown to be convergent we conclude that $V_k^{r+1} e^{r+1}$ must converge to the zero vector a.e., which immediately implies that V_k^{r+1} converges a.e. to the zero matrix. It also follows that $\sum_{i=1}^r W^i$ equals one entrywise.

Define the map $\psi(X) = (\psi_1(X), \psi_2(X))$ on $\mathcal{S}$ as follows: For each matrix X of the above form, $\psi_1(X)$ is the first $(d-d_0) \times (d-d_0)$ minor of X. $\psi_2(X)$ is a matrix of order $d_0 \times (d-d_0)$ whose 1st, (d_1+1)-th,$\cdots$, $(d_1+d_2+\cdots+d_{r-1}+1)$-th columns are U^1e^1, U^2e^2, $\cdots$, U^re^r respectively and the other columns are zero. Going back to the earlier notation, Case 1 applies to conclude that $\psi_1(Y_k)$ converges in distribution. The argument above shows that $\psi_2(Y_k)$ converges almost surely. Thus by Slutsky, $\psi(Y_k)$ converges in distribution , say, to λ.

Now let us take a convergent subsequence of μ^k , say, μ^{k_i} converging to ν. We know that ν is concentrated on $\mathcal{K}$. In the present case – H being a singleton – observe that $\mathcal{K}$ consists of matrices of the form $\begin{pmatrix} K & \mathbf{0} \\ WK & \mathbf{0} \end{pmatrix}$. As a consequence ν is determined by the joint distribution of (K, W). Thus, ν is determined by $\nu\psi^{-1}$. But ψ being a continuous map, we have $\mu^{k_i}\psi^{-1} \to \nu\psi^{-1}$. But, as argued above $\mu^k\psi^{-1} \to \lambda$, so that $\nu\psi^{-1} = \lambda$. Thus μ^k has only one limit point and hence it converges, as was to be proved.

Finally, we consider the case of general H. In this case, neither the first $(d-d_0)\times(d-d_0)$ minor of a matrix in $\mathcal{S}$ is block diagonal nor the $V_k^i e^i$ defined above converge a.e. However, only a slight modification is needed, which will be explained now. As earlier start with a sequence X_k of i.i.d random matrices with common law μ and set $Y_k = \prod_1^k X_i$. Define the map $\psi(X) = (\psi_1(X), \psi_2(X))$ on $\mathcal{S}$ as follows. $\psi_1(X)$ is as earlier the first $(d-d_0)\times(d-d_0)$ minor of X. If $X \in \mathcal{S}_\pi$, then $\psi_2(X)$ is the matrix of order $d_0 \times (d-d_0)$ whose 1st, (d_1+1)th,$\cdots$, $(d_1+d_2+\cdots+d_{r-1}+1)$th columns are $U^{\pi^{-1}(1)}e^{\pi^{-1}(1)}$, $\cdots$, $U^{\pi^{-1}(r)}e^{\pi^{-1}(r)}$. Here U^i is the $T \times C_i$ block of X and as earlier, e^i is $d_i \times 1$ column vector with all entries 1. With this change, a simple calculation shows, that $\psi_2(Y_k)$ converges almost surely. The remaining part of the proof is as earlier and is hence omitted.

This completes the proof.

Here is an *alternative proof* of the *if* part of the theorem :

Assume that $\tilde{\mu}^n$ converges. It is easy to see that for each $\pi \in H$, $\tilde{\mu}^n(\pi) \to \frac{1}{|H|}$. In particular, we can get an integer N such that for all $n \geq N$, $\mu^n(\mathcal{S}_e) > 0$. Let $M \in \mathcal{K}_e$ and W be an open set containing M. For any $A \in \mathcal{S}_e$, we have $MAM = M$ so that we can get an open set U containing M such that $U\mathcal{S}_eU \subset W$. Compactness of $\mathcal{S}_e$ is used here. But U being open and $\mathcal{S}$ being the closure of $\cup_{i\geq 1}\mathcal{S}_{\mu^i}$ we can get a j such that $\mu^j(U) > 0$. Thus $\mu^{2j+n}(W) > 0$ for all $n \geq N$. In otherwords the set L is nonempty where

$$L = \{M : \forall \text{ open } W \text{ with } M \in W, \quad \exists N, \quad \forall n \geq N, \mu^n(W) > 0\}$$

Now we use the following theorem to complete the proof:

Theorem.

Let $\mu \in P(S_d)$. The sequence μ^n converges weakly iff $\lim_n \inf S(\mu^n)$ is non-empty, where $\lim_n \inf S(\mu^n)$ is defined as $\{x \in S_d :$ for every open set $V(x)$ containing x, there exists a positive integer N such that $n \geq N \Rightarrow V(x) \cap S(\mu^n) \neq \emptyset\}$.

This is Theorem 2.13 (*iv*), page 91-92, in Hognas and Mukherjea [17]. Of course, the theorem is true for general semigroups, however we stated it for S_d.

Section 5 : Concluding Remarks

We conclude the section with a few remarks :

Remark 1: Since H is a subgroup of the permutation group, the convergence of μ^n is reduced to convergence of probabilities on a finite set.

Remark 2: Suppose G is a finite group and $\tilde{\mu}$ is a probability on G. If $\tilde{\mu}^n$ converges, then the limit being idempotent, must be the uniform distribution on a subgroup and hence, more specifically, it must be the uniform distribution on the subgroup generated by the support of $\tilde{\mu}$. Denote this subgroup by H.

Further, since we are dealing with convergence of probabilities on a finite set, pointwise convergence holds. Thus, if $\tilde{\mu}^n$ converges, then after some stage, the support of $\tilde{\mu}^n$ must be all of H. Conversely, if for some N, support of $\tilde{\mu}^N$ is all of H, then $\tilde{\mu}^n$ does indeed converge to the uniform distribution on H. To see this, run the Markov chain with state space H and initial distribution $\tilde{\mu}^N$ and transition probability as follows : if we are at x, choose y according to law μ and move to yx (perhaps there are other ways of seeing it).

In particular it follows that if $\tilde{\mu}^n$ converges then for some i, $S_{\tilde{\mu}^i} \cap S_{\tilde{\mu}^{i+1}} \neq \emptyset$. Conversely, if for some i, $S_{\tilde{\mu}^i} \cap S_{\tilde{\mu}^{i+1}} \neq \emptyset$ then indeed $\tilde{\mu}^n$ does converge. This can be argued as follows. First notice that if $x \in S_{\tilde{\mu}^i} \cap S_{\tilde{\mu}^{i+1}}$ and $y \in S_{\tilde{\mu}}$, then $xy \in S_{\tilde{\mu}^{i+1}} \cap S_{\tilde{\mu}^{i+2}}$.Thus in general, $S_{\tilde{\mu}^n} \cap S_{\tilde{\mu}^{n+1}} \neq \emptyset$ for all $n \geq i$. By considering order of any element in support of $\tilde{\mu}$, we can get a j such that $e \in S_{\tilde{\mu}^j}$. This immediately implies that $S_{\tilde{\mu}^{rj}}$ increases with r and hence (being in a finite set up), do stabilize after some stage. Thus we get a p such that $S_{\tilde{\mu}^{nj}}$ remains same for all $n \geq p$. One can argue out easily that $S_{\tilde{\mu}^{nj}}$ is a subgroup for each $n \geq p$. Indeed, for any $n \geq p$, $S_{\tilde{\mu}^{nj}} = S_{\tilde{\mu}^{2nj}}$. So, firstly for $a, b \in S_{\tilde{\mu}^{nj}}$, $ab \in S_{\tilde{\mu}^{2nj}}$ which is same as $S_{\tilde{\mu}^{nj}}$. Secondly, if $g_1, g_2, \cdots, g_l$ are all the distinct elements in $S_{\tilde{\mu}^{nj}}$, then for any $a \in S_{\tilde{\mu}^{nj}}$, $ag_1, ag_2, \cdots, ag_l$ are distinct and they are in $S_{\tilde{\mu}^{2nj}}$. But $|S_{\tilde{\mu}^{2nj}}|$ being also equal to l, they are the only elements in $S_{\tilde{\mu}^{2nj}}$ which is same as $S_{\tilde{\mu}^{nj}}$. Now one of these elements is e. So the inverse of a is in $S_{\tilde{\mu}^{nj}}$. Thus, for any $n \geq p$, $S_{\tilde{\mu}^{nj}}$ is a group. Consider $n > p$ such that $nj > i$. Then $S_{\tilde{\mu}^{nj}} \cap S_{\tilde{\mu}^{nj+1}} \neq \emptyset$. But $S_{\tilde{\mu}^{nj}}$ being a subgroup and as $|S_{\tilde{\mu}^{nj+1}}| \geq |S_{\tilde{\mu}^{nj}}|$, we conclude that $S_{\tilde{\mu}^{nj}} = S_{\tilde{\mu}^{nj+1}}$. Hence, $\tilde{\mu}^n$ converges.

One sufficient condition for $\tilde{\mu}^n$ to converge is that g.c.d. of the orders of the elements in $S_{\tilde{\mu}}$ is 1. This can be shown as follows. Suppose that for $1 \leq i \leq k$, $o(x_i) = m_i$. Then $e \in S_{\tilde{\mu}^{m_i}}$ $i = 1, 2, \cdots, k$. So $e \in S_{\tilde{\mu}^n}$ where n is any positive integer linear combination of the m_i's. But by Euler's theorem, there exists two such linear combinations a and b with $a - b = 1$. In other words, $e \in S_{\tilde{\mu}^b}$ as well as $e \in S_{\tilde{\mu}^{b+1}}$. From what was argued in the previous paragraph, it follows that $\tilde{\mu}^n$ converges.

These comments are perhaps not new. They are mentioned here only because they provide algorithms to decide when $\tilde{\mu}^n$ converges.

Remark 3: *In case $d = 2$, the only case of non-convergence occurs when* $r = 2$, $T = \emptyset$ and $\tilde{\mu}$ is concentrated on the non-identity permutation. This corresponds to the theorem of Mukherjea [22] discussed in section 2.

Remark 4: In case $d = 3$, the only cases of non-convergence are the following:

1. $r = 2$; T, C_1, and C_2 *each consists of one element and $\tilde{\mu}$ is supported* on the non-identity permutation.

2. $r = 2$; $T = \emptyset$, C_1 consists of one element and C_2 consists of two elements and $\tilde{\mu}$ is supported on the non-identity permutation.

3. $r = 3$; (hence) $T = \emptyset$ while C_1, C_2, C_3 consist of one element each and $\tilde{\mu}$ is concentrated EITHER on the permutation $\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$ OR *on the* permutation $\begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}$ OR on the set containing the three permutations $\begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}$ $\begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}$ and $\begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}$.

These correspond to the cases mentioned in Chakraborty and Rao [8] and – in a different order – these were the cases mentioned in Dhar and Mukherjea [11]. Of course as d increases there are more and more cases of non-convergence but they all amount to investigating convolution powers on the permutation group and Remark 2 above applies.

Remark 5: If we look at Theorem 7.1 of chapter 2 closely it appears that the non-convergence of the convolution sequence (μ^n) happens because of periodicity of the supports of μ^n. This periodicity may arise due to the cyclicity of the states. It may arise even otherwise also. It is interesting to note that in such a case μ was concentrated on zero-one matrices. We do not know if there is any analogue of this in the general case.

Remark 6: It is interesting to note that most of the analysis depended only on the structure of the matrices and not on the exact values of the entries (apart from the knowledge of whether the entry is zero or not). There exists a large amount of literature on probabilities on graphs. Consider a set of d vertices. Let Γ be the set of all directed graphs with loops allowed. If G_1, G_2 are two such graphs, then $G_1 * G_2$ is the graph where (u, v) is an edge iff for some vertex x, (u, x) is an edge in G_1 and (x, v) is an edge in G_2. Thus, Γ is a semigroup. One could fix a probability μ on Γ and find conditions for convergence of the sequence μ^n. The same ideas as discussed here would perhaps go through. Since we are not familiar with the literature in this area, we have not undertaken this exercise. In any case, it is a different story.

CHAPTER - IV

Random Continued Fraction Expansions

Summary

In this chapter, our interest is in the existence of absolutely continuous invariant measure and its properties for the transformation $T(x) = \theta(\frac{1}{\theta x} - [\frac{1}{\theta x}])$ on the interval $[0, \theta]$ where $0 < \theta < 1$. This chapter has five sections. In section 1, we introduce the problem of random continued fraction expansions studied by Bhattacharya and Goswami. In section 2, we fix our notations. We define continued fraction expansions with respect to θ and discuss their properties. We relate these to the map T described above. In section 3, we study absolutely continuous invariant measures for such maps when θ^2s are reciprocals of positive integers. In section 4, we study this problem for more general θ. The main interest revolves around the following two properties that hold for usual continued fraction expansions : (i) The limit of the averages of the digits in the expansion is almost surely infinity and (ii) the limit of $\frac{1}{n}\log q_n$ is almost surely finite where p_n/q_n is the n-th convergent. Finally in the last section we return to the Markov Processes that motivated this study and make some comments.

Section 1 : Introduction

Motivated by problems in random number generation, R.N.Bhattacharya and A.Goswami [2] considered the following problem :

Suppose $(Z_n)_{n\geq 1}$ is a sequence of i.i.d. non-negative random variables not identically zero. Let X_0 be a strictly positive random variable independent of $(Z_n)_{n\geq 1}$. Let $(X_n)_{n\geq 0}$ be the Markov Process defined by

$$X_{n+1} = Z_{n+1} + \frac{1}{X_n} \quad \text{for} \quad n \geq 0 \tag{2}$$

Clearly this process arises as an iterated function system in the following way. For each non-negative number u let f_u be the function on $(0, \infty)$ defined by $f_u(x) = u + \frac{1}{x}$. Let $\mathcal{F}$ be the collection of functions so obtained. If μ is the distribution of Z_1 then we can transfer it to a probability, to be denoted by ν on $\mathcal{F}$. Here is another way of describing the above Markov Process. If we are at state x, we select a function $f \in \mathcal{F}$ according to the law ν and then move to $f(x)$.

The case when Z_1 has gamma distribution was studied by Letac and Seshadri [21]. Bhattacharya and Goswami [2] showed that the Markov Process converges in distribution to the random continued fraction $[Z_1; Z_2, \cdots]$. They further showed that the limiting distribution, which is the unique invariant probability π of (X_n), is non-atomic. They then considered the special case

$$Z_1 = \begin{cases} 0 & \text{w.p.} \quad \alpha, \\ \theta & \text{w.p.} \quad 1-\alpha \end{cases}$$

where $0 \leq \alpha < 1, \quad \theta > 0.$

If Z_1 is degenerate at θ, then clearly the Markov Process is essentially deterministic (its only randomness comes from X_0). It converges to the Dirac measure at $[\theta; \theta, \cdots]$. Bhattacharya and Goswami showed that the support of π is all of $(0, \infty)$ in case $\theta \leq 1$ and is a Cantor set if $\theta > 1$. Moreover when $\theta = 1$, they showed that π is singular and gave explicit computation of the distribution function. The main ingredients in their proof were the following two important facts : Suppose that for any number x in $[0, 1)$ we denote by $[a_1, a_2, \cdots]$ its usual continued fraction expansion and $\frac{p_n}{q_n}$ denotes the n-th convergent of the number x. Then,

$$(*) \qquad \frac{a_1 + \cdots + a_n}{n} \longrightarrow \infty \quad \text{a.e.}$$

and

$$(**) \qquad \lim_{n\to\infty} \tfrac{1}{n} \log q_n = \gamma \quad \text{a.e. for some finite constant} \quad \gamma.$$

Here almost everywhere is w.r.t the Lebesgue measure on $[0, 1)$. But the crucial role is played by the Ergodic theorem and the fact that the Gauss measure $d\mu(x) = \frac{1}{\log 2} \frac{1}{1+x} dx$ on $[0, 1)$ is invariant for the Gauss map U given by,

$$U(x) = \begin{cases} \frac{1}{x} - [\frac{1}{x}] & \text{if } x \neq 0, \\ 0 & \text{if } x = 0 \end{cases}$$

In this chapter, we define the generalised version of the Gauss map for values of θ strictly smaller than one . For certain values of θ, we obtain an absolutely continuous invariant probability for this generalised map. Then we prove (*) and (**) also hold. Our notation and exposition follow Billingsley [5], Khinchine [19] and Bhattacharya and Goswami [2] closely.

Section 2 : Preliminaries

Throughout our discussion we fix a θ with $0 < \theta < 1$. To understand the nature of the generalised Gauss map, we start with a discussion of continued fraction expansion w.r.t θ, analogous to the usual expansion which corresponds to the case $\theta = 1$.

Let $x > 0$. Let $a_0 = \max\{n \geq 0 : n\theta \leq x\}$. If x already equals $a_0\theta$, we write $x = [a_0\theta]$. Otherwise, define r_1 by $x = a_0\theta + \frac{1}{r_1}$ where $0 < \frac{1}{r_1} < \theta$. Then $r_1 > \frac{1}{\theta} \geq \theta$ and let $a_1 = \max\{n \geq 0 : n\theta \leq r_1\}$. If $r_1 = a_1\theta$, then we write $x = [a_0\theta, a_1\theta]$, i.e., $x = a_0\theta + \frac{1}{a_1\theta}$. If $a_1\theta < r_1$, define r_2 by $r_1 = a_1\theta + \frac{1}{r_2}$ where $0 < \frac{1}{r_2} < \theta$. So, $r_2 > \frac{1}{\theta} \geq \theta$ and let $a_2 = \max\{n \geq 0 : n\theta \leq r_2\}$. Proceeding in this way, either the process terminates at, say, n steps or it continues indefinitely. In the former case, we write $x = [a_0\theta; a_1\theta, \cdots, a_n\theta]$ and we call this the continued fraction expansion of x with respect to θ terminating at the n-th stage. In the latter case, we write $x = [a_0\theta; a_1\theta, a_2\theta, \cdots]$ and it is called the infinite or non-terminating continued fraction expansion of x with respect to θ. We shall later justify this notation by showing that the infinite expansion does actually converge, in an appropriate sense, to x.

From now on, unless otherwise mentioned, we refer to this expansion as the continued fraction expansion of a number in $(0, \infty)$. Since during any discussion a particular value of θ is fixed, we shall omit the phrase 'w.r.t θ'.

As in the case of usual continued fraction expansion, we shall define the

n-th convergent of a number $x \in (0, \infty)$ as

$$\frac{p_n}{q_n} = [a_0\theta; a_1\theta, \cdots, a_n\theta], \quad n \geq 0.$$

In case x has terminating expansion, say, $x = [a_0\theta; a_1\theta, \cdots, a_k\theta]$,then clearly $\frac{p_k}{q_k} = x$. We make the usual convention that in this case

$$\frac{p_n}{q_n} = x \quad \text{for} \quad n \geq k.$$

In the non-terminating case, arguing as in Khinchine [19] or Bhattacharya and Goswami [2], one can show that $\frac{p_n}{q_n}$ converges to x as $n \to \infty$. We shall illustrate this for $x < \theta$.

When $x < \theta$, we have $a_0 = 0$ and instead of writing $x = [0; a_1\theta, a_2\theta, \cdots]$, we write $x = [a_1\theta, a_2\theta, \cdots]$ which is same as writing, in the usual notation

$$x = \cfrac{1}{a_1\theta + \cfrac{1}{a_2\theta + \cdots}}.$$

As in the case of usual continued fraction expansion, we define a map from $[0, \theta)$ to $[0, \theta)$ as follows

$$T(x) = \begin{cases} \frac{1}{x} - \theta[\frac{1}{\theta x}] & \text{if} \quad x \neq 0, \\ 0 & \text{if} \quad x = 0 \end{cases} \tag{2.1}$$

This is what we call the generalised Gauss map.

With the help of this map T, we can neatly describe $\frac{1}{r_n}$ that appears as reminder at the n-th stage in the definition of the continued fraction expansion of x as simply $T^n(x)$.

Let $0 < x < \theta, \quad x = [a_1\theta, a_2\theta, \cdots]$. When we need to show the dependence of a_n, p_n and q_n on x we write $a_n(x)$, $p_n(x)$ and $q_n(x)$ respectively. The following are routine to verify (The stated identities hold for all n in case x has non-terminating expansion and they hold for $n \leq k$ in case x has expansion terminating at the k-th stage):

$$a_n(x) = a_1(T^{n-1}x), \quad n = 1, 2, \cdots \tag{2.2}$$

For $n \geq 1$,

(2.3(i)) $$p_n(x) = a_n(x)\theta p_{n-1}(x) + p_{n-2}(x)$$
and
(2.3(ii)) $$q_n(x) = a_n(x)\theta q_{n-1}(x) + q_{n-2}(x)$$

(Here we use the convention followed in the usual continued fraction expansion, namely, $p_{-1}(x) = 1, p_0(x) = 0, q_{-1}(x) = 0, q_0(x) = 1$.)

As with usual expansion,

(2.4) $$p_{n-1}(x)q_n(x) - p_n(x)q_{n-1}(x) = (-1)^n \quad \text{for} \quad n \geq 0$$

Let $n \geq 1$. Let $x = [a_1(x)\theta, a_2(x)\theta, \cdots, a_n(x)\theta + \frac{1}{r_n}]$. Then a little algebra shows that as in the case of usual expansion,

(2.5) $$x = \frac{p_n(x) + \frac{1}{r_n}p_{n-1}(x)}{q_n(x) + \frac{1}{r_n}q_{n-1}(x)}$$

But since, $\frac{1}{r_n} = T^n(x)$ we get,

(2.6) $$x = \frac{p_n(x) + T^n(x)p_{n-1}(x)}{q_n(x) + T^n(x)q_{n-1}(x)}$$

Therefore,

$$\left| x - \frac{p_n(x)}{q_n(x)} \right| = \left| \frac{p_n(x) + T^n(x)p_{n-1}(x)}{q_n(x) + T^n(x)q_{n-1}(x)} - \frac{p_n(x)}{q_n(x)} \right|$$

$$= \left| \frac{T^n(x)(p_{n-1}(x)q_n(x) - q_{n-1}(x)p_n(x))}{q_n(x)(q_n(x) + T^n(x)q_{n-1}(x))} \right|$$

$$= \frac{1}{q_n(x)((T^n(x))^{-1}q_n(x) + q_{n-1}(x))}$$

or,

$$(2.7) \qquad \left| x - \frac{p_n(x)}{q_n(x)} \right| = \frac{1}{q_n(x)((T^n(x))^{-1}q_n(x) + q_{n-1}(x))}$$

Now, $a_{n+1}(x)\theta \leq (T^n(x))^{-1} \leq (a_{n+1}(x)+1)\theta$. Using these estimates in (2.7) and noting (2.3(i)),(2.3(ii)), we get,

$$(2.8) \qquad \frac{1}{q_n(x)(q_{n+1}(x) + \theta q_n(x))} \leq \left| x - \frac{p_n(x)}{q_n(x)} \right| \leq \frac{1}{q_n(x)q_{n+1}(x)}$$

Since $q_0 = 1 \geq \theta$ and $q_1 = a_1\theta \geq \theta$, we have by using (2.3(ii)), $q_n \geq \theta \quad \forall n \geq 1$. Further, (2.3(ii)) also gives

$$q_n = a_n\theta q_{n-1} + q_{n-2} \geq \theta^2 + q_{n-2} \quad \forall n \geq 2$$

Using this inequality, we have, by induction on n, $q_n \geq [\frac{n}{2}]\theta^2 \to \infty$. This was already observed in Bhattacharya and Goswami [2]. So, from (2.8), $| x - \frac{p_n}{q_n} | \to 0$ as $n \to \infty$. Again, from (2.3(ii)), $q_n \geq (\theta^2+1)q_{n-2}$ for $n \geq 1$. Now using induction on n, we get, for even n,

$$q_n \geq (\theta^2+1)^{\frac{n}{2}} q_0 = (\theta^2+1)^{\frac{n}{2}}$$

and for odd n,

$$q_n \geq (\theta^2+1)^{\frac{n-1}{2}} q_1 \geq (\theta^2+1)^{\frac{n-1}{2}}\theta.$$

Combining these, one can write,

$$q_n \geq (\theta^2+1)^{[\frac{n}{2}]}\theta \quad \forall n.$$

Now, we shall discuss briefly as to when a given sequence $[a_1\theta, a_2\theta, \cdots]$ arises as the continued fraction expansion of a number x smaller than θ. It is easy to see that a sequence $[a_0\theta; a_1\theta, a_2\theta, \cdots]$ arises as the continued fraction expansion of a number if and only if $[a_1\theta, a_2\theta, \cdots]$ arises as the expansion of a number smaller than θ.

To understand the idea, note that, for the usual continued fraction expansion (case $\theta = 1$), $[0; 2, 1]$ does not arise as the expansion of any number whereas $[0; 3]$ does arise. In fact in the usual case this is the only restriction.

More precisely $[a_1, a_2, \cdots]$ arises as the usual expansion of a number smaller than one iff (i) each $a_i \geq 1$; (ii) in case it is terminating the last a_k is strictly larger than one.

Let us start with the most simple case $\theta = \frac{1}{n\theta}$ for some n. (Roughly speaking $\frac{1}{\theta}$ is an integer w.r.t θ, because $\frac{1}{\theta} = n\theta$) In this case $[a_1\theta, a_2\theta, \cdots]$ arises as the expansion of a number smaller than θ iff (i) each $a_i \geq n$; (ii) in case it is terminating the last a_k is strictly larger than n. This can be seen as follows. Suppose $[a_1\theta, a_2\theta, \cdots, a_k\theta]$ is continued fraction expansion of a number $x < \theta$. Then $n\theta = \frac{1}{\theta} < \frac{1}{x}$ shows that $a_1 \geq n$. Now $\frac{1}{x} = a_1\theta + \frac{1}{r_2}$ where $0 < \frac{1}{r_2} < \theta$ implying as earlier $a_2 \geq n$. Proceeding this way, we get that $a_i \geq n$ for all $i < k$. Since $r_k = a_k\theta > \frac{1}{\theta} = n\theta$, we get $a_k > n$ as claimed. Conversely, suppose we have integers a_i for $1 \leq i \leq k$ so that $a_i \geq n$ for $i < k$ and $a_k > n$. Then, since $a_k > n$, we have $a_k\theta > n\theta$ or $\frac{1}{a_k\theta} < \theta$. Also, $a_{k-1} \geq n$ implies that $a_{k-1}\theta + \frac{1}{a_k\theta} > n\theta$ or $[a_{k-1}\theta, a_k\theta] < \theta$. Proceeding this way, we can show $[a_i\theta, \cdots, a_k\theta] < \theta$ for $1 \leq i \leq k$. Thus if we define $x = [a_1\theta, \cdots, a_k\theta]$ then $x < \theta$ and indeed $[a_1\theta, \cdots, a_k\theta]$ is the continued fraction expansion of x. Similar but simpler argument applies to show that in the non-terminating case, it is necessary and sufficient to have each $a_i \geq n$.

To consider a slightly more general case suppose that $\frac{1}{\theta} = [n_1\theta; n_2\theta]$. Thus we have

$$\frac{1}{(n_1+1)\theta} < \theta < \frac{1}{n_1\theta}$$

and

$$\theta = \frac{1}{n_1\theta + \frac{1}{n_2\theta}}$$

It should be observed that in such a case $n_2 > (n_1 + 1)$. In this case $[a_1\theta, a_2\theta, \cdots]$ arises as the expansion of a number smaller than θ iff the following conditions hold:

(i) Each $a_i \geq n_1$;

(ii) In case an $a_i = n_1$ then $a_{i+1} < n_2$;

(iii) In the terminating case the last a_k must satisfy $a_k > n_1$.

This can be seen as follows. Suppose $[a_1\theta, \cdots, a_k\theta]$ is the continued fraction expansion of a number $x < \theta$. If $k = 1$, it trivially follows that $a_1 > n_1$ as claimed. Let us assume that $k \geq 2$. Arguing as in the previous case, we obtain that $a_i \geq n_1$ for each i and $a_k > n_1$. Suppose that for some $i < k$, $a_i = n_1$. Then $[a_i\theta, a_{i+1}\theta, \cdots, a_k\theta] < \theta$ implies that $a_i\theta + [a_{i+1}\theta, \cdots, a_k\theta] > n_1\theta + \frac{1}{n_2\theta}$. But since $a_i = n_1$, this immediately implies $[a_{i+1}\theta, \cdots, a_k\theta] > \frac{1}{n_2\theta}$ so that $a_{i+1}\theta + [a_{i+2}\theta, \cdots, a_k\theta] < n_2\theta$ and consequently, $a_{i+1} < n_2$. Conversely, suppose that $a_1, \cdots, a_k$ are integers satisfying the conditions of the claim. As in the previous case, $a_k > n_1$ implies $[a_k\theta] < \theta$. Now if $a_{k-1} > n_1$, $a_{k-1}\theta + \frac{1}{a_k\theta} > n_1\theta + \frac{1}{n_2\theta}$ implying that $[a_{k-1}\theta, a_k\theta] < \theta$. On the other hand, if $a_{k-1} = n_1$, then using the hypothesis that $a_k < n_2$, we get $a_{k-1}\theta + \frac{1}{a_k\theta} = n_1\theta + \frac{1}{a_k\theta} > n_1\theta + \frac{1}{n_2\theta}$. So, $[a_{k-1}\theta, a_k\theta] < \theta$. One can now proceed as in the earlier case and show that $[a_1\theta, \cdots, a_k\theta]$ is indeed the continued fraction expansion of a number $x < \theta$. The non-terminating case is dealt with in an analogous manner.

Now consider the case $\frac{1}{\theta} = [n_1\theta; n_2\theta, \cdots, n_m\theta]$. (Roughly speaking $\frac{1}{\theta}$ is rational w.r.t θ as it has a terminating expansion w.r.t θ, or equivalently it is ratio of two polynomial expressions in θ). In this case $[a_1\theta, a_2\theta, \cdots]$ arises as the expansion of a number smaller than θ iff the following conditions hold:

(i) Each $a_i \geq n_1$;

(ii) In case for some $i \geq 1$ and $p < m$, $\langle a_{i+1}, \cdots, a_{i+p}\rangle = \langle n_1, \cdots, n_p\rangle$ then we should have $a_{i+p+1} \leq n_{p+1}$ if $p+1$ is even while $a_{i+p+1} \geq n_{p+1}$ if $p+1$ is odd.

Moreover if m is even and $p+1$ equals m, then $a_{i+p+1} < n_{p+1}$.

(iii) In the terminating case the last a_k must satisfy $a_k > n_1$ and further if for some even $p < m$, (i.e, $p+1$ is odd) $\langle a_{k-p}, \cdots, a_{k-1}\rangle = \langle n_1, \cdots, n_p\rangle$ then moreover $a_k > n_{p+1}$.

The same ideas as in the earlier two cases, but executed with a little care, will lead to a proof of this claim. We shall not go into the details.

Finally we assume that $\frac{1}{\theta}$ has non-terminating expansion, say,

$$\frac{1}{\theta} = [n_1\theta; n_2\theta, \cdots].$$

In this case $[a_1\theta, a_2\theta, \cdots]$ arises as the expansion of a number smaller than θ iff (i) each $a_i \geq n_1$; (ii) in case for some $i \geq 1$ and $p \geq 1$, $\langle a_{i+1}, \cdots, a_{i+p}\rangle = \langle n_1, \cdots, n_p\rangle$ then $a_{i+p+1} \leq n_{p+1}$ if $p+1$ is even while $a_{i+p+1} \geq n_{p+1}$ if $p+1$ is odd; (iii) In the terminating case the last a_k must satisfy $a_k > n_1$ and further if for some even $p \geq 1$, (i.e, $p+1$ is odd) $\langle a_{k-p}, \cdots, a_{k-1}\rangle = \langle n_1, \cdots, n_p\rangle$ then moreover $a_k > n_{p+1}$.

The above discussion gives necessary and sufficient criteria for an expression $[a_0\theta; a_1\theta, \cdots]$, to be actually the continued fraction expansion of a number. However those conditions depend on the expansion of $\frac{1}{\theta}$. More precisely the conditions depended on the sequence of integers $n_1, n_2, \cdots$ where $\frac{1}{\theta} = [n_1\theta; n_2\theta, \cdots]$,

It is natural to enquire as to how the expansion of $\frac{1}{\theta}$ itself looks like. We shall not go into the details except to make the following comment. One can easily observe that in such a case each n_i must be at least as large as n_1 and in the terminating case the last n_m must be indeed strictly larger than n_1. However this is not a sufficient condition. For example we can not have $\frac{1}{\theta} = [2\theta; 3\theta]$, a simple algebra shows that the correct expansion is $\frac{1}{\theta} = [3\theta;]$.

Before proceeding further, we mention that in the literature, there exist several generalizations of the usual continued fraction expansions. For example, see Bissinger [6], Everett [15] and Renyi [28].

As in Billingsley[5], we introduce the sets $\Delta_{a_1,a_2,\cdots,a_n}$ and the maps $\psi_{a_1,a_2,\cdots,a_n} : [0, \theta) \to [0, \theta)$. These will be needed later to show, in some cases, ergodicity of the map T with respect to an appropriate measure. $\Delta_{a_1,a_2,\cdots,a_n}$ is the set of all x such that $a_i(x) = a_i$ for $i = 1, 2, \cdots, n$. Obviously, because of the above discussion, $\Delta_{a_1,a_2,\cdots,a_n}$ may be empty for some n-tuples $(a_1, a_2, \cdots, a_n)$. In what follows we assume that $\Delta_{a_1,a_2,\cdots,a_n}$ is non-empty for the n-tuple $(a_1, a_2, \cdots, a_n)$. $\psi_{a_1,a_2,\cdots,a_n}$ is given by,

$$\psi_{a_1,a_2,\cdots,a_n}(t) = \cfrac{1}{a_1\theta + \cfrac{1}{a_2\theta + \cfrac{1}{\cdots + \cfrac{1}{a_n\theta + t}}}}, \quad t \in [0, \theta).$$

Then $\Delta_{a_1,a_2,\cdots,a_n}$ is the image of $[0,\theta)$ under $\psi_{a_1,a_2,\cdots,a_n}$. One can show that $\psi_{a_1,a_2,\cdots,a_n}(t) = \dfrac{p_n + tp_{n-1}}{q_n + tq_{n-1}}$ for $t \in [0,\theta)$ just like in (2.5). Also $\psi_{a_1,a_2,\cdots,a_n}(t)$ is decreasing for odd n and increasing for even n. So,

$$\Delta_{a_1,a_2,\cdots,a_n} = \begin{cases} \left[\dfrac{p_n}{q_n}, \dfrac{p_n + \theta p_{n-1}}{q_n + \theta q_{n-1}} \right] & \text{if } n \text{ even}, \\ \left[\dfrac{p_n + \theta p_{n-1}}{q_n + \theta q_{n-1}}, \dfrac{p_n}{q_n} \right] & \text{if } n \text{ odd}. \end{cases}$$

Using (2.4), we see,

$$\lambda(\Delta_{a_1,a_2,\cdots,a_n}) = \frac{\theta}{q_n(q_n + \theta q_{n-1})} \tag{2.9}$$

where λ, as usual, denotes Lebesgue measure.

Before concluding this section, we remark the following. One can define a map on $[0,\theta)$ to itself by putting $U_1(x) = \theta(\frac{x}{\theta} - [\frac{x}{\theta}])$ and one can also define a map on $[0,\frac{1}{\theta})$ to itself by putting $U_2(x) = \frac{1}{\theta}(\frac{1}{\theta x} - [\frac{1}{\theta x}])$. Obviously, these maps are conjugate to the Gauss map U on $[0,1)$. However, the map T that we defined above is different from U_1 and U_2 and this map T is relevent for our discussion. We could not see if this is conjugate to the Gauss map U. Perhaps it is not.

Section 3 : Invariant Measure for T when $\frac{1}{\theta^2} \in \mathbb{N}$

In this section we assume that $\dfrac{1}{\theta^2} \in \mathbb{N}$. Thus for some integer, say l, $\dfrac{1}{\theta} = l\theta$. Thus $\dfrac{1}{\theta}$ has a continued fraction expansion terminating at the first stage itself, or in the notation of the previous section $\dfrac{1}{\theta} = [l\theta]$. Throughout this section θ and hence the integer l is fixed.

First recall that the Gauss measure $d\mu(x) = \dfrac{1}{\log 2}\dfrac{1}{1+x}dx$ on $[0,1)$ is invariant for the transformation associated with the usual continued fraction expansion , namely,

$$U(x) = \begin{cases} \frac{1}{x} - [\frac{1}{x}] & \text{if } x \neq 0, \\ 0 & \text{if } x = 0 \end{cases}$$

The analogue of this transformation for the θ expansion is the transformation T defined on $[0, \theta)$ as follows :

$$T(x) = \begin{cases} \frac{1}{x} - \theta[\frac{1}{\theta x}] & \text{if } x \neq 0, \\ 0 & \text{if } x = 0 \end{cases}$$

We shall now extend the usual argument (see Billingsley [5]) to get an absolutely continuous invariant measure for the above transformation.
In fact, we claim that

$$dP(x) = \frac{1}{\log \frac{l+1}{l}} \frac{1}{\sqrt{l} + x} dx$$

which is same as saying

$$dP(x) = \frac{1}{\log(1+\theta^2)} \frac{\theta}{1+\theta x} dx$$

is the required invariant measure for T. In the present case, we are lucky enough to explicitly write down the invariant measure which we will not be able to do for the more general cases to be considered later.

Since we could not see any direct way of connecting these two transformations T and U we shall verify the above claim by carrying the same steps as in Billingsley referred to above. There is no new idea involved. In order to show that T preserves P, it is enough to show that $P[0, \theta u) = P(T^{-1}[0, \theta u))$ for all $u \in [0, 1)$.

Since $T^{-1}[0, \theta u) = \bigcup_{k=l}^{\infty} (\frac{1}{(k+u)\theta}, \frac{1}{k\theta})$ (equality is upto a set of Lebesgue measure zero), it is enough to verify the following :

$$\int_0^{\theta u} \frac{\theta}{1+\theta x} dx = \sum_{k=l}^{\infty} \int_{\frac{1}{(k+u)\theta}}^{\frac{1}{k\theta}} \frac{\theta}{1+\theta x} dx.$$

$$\sum_{k=l}^{\infty} [\log(1 + \frac{1}{k}) - \log(1 + \frac{1}{k+u})] = \sum_{k=l}^{\infty} [\log(\frac{k+u}{k}) - \log(\frac{k+1+u}{k+1})].$$

Thus the last sum is a telescopic sum and equals $\log\left(\frac{l+u}{l}\right)$ which is same as the left hand side.

We now show that T is ergodic under this P. For notational convenience, let us denote $\Delta_{a_1,a_2,\cdots,a_n}$ and $\psi_{a_1,a_2,\cdots,a_n}$ defined in section 2 by Δ_n and ψ_n respectively. Here we fix $a_1, a_2, \cdots, a_n$. Then Δ_n has length $|\,\psi_n(\theta)-\psi_n(0)\,|$. Also,for $0 \le x < y \le \theta$, the interval $\{\omega : x \le T^n(\omega) < y\} \cap \Delta_n$ has length $|\,\psi_n(y) - \psi_n(x)\,|$.

So, using the notation, $\lambda(A|B) = \lambda(A \cap B) \,/\, \lambda(B)$, we have,

$$\lambda(T^{-n}[x,y)|\Delta_n) = \frac{\psi_n(y) - \psi_n(x)}{\psi_n(\theta) - \psi_n(0)}.$$

where, on simplification, the absolute value of the numerator is given by $\frac{y-x}{(q_n + xq_{n-1})(q_n + yq_{n-1})}$ and that of the denominator is given by $\frac{\theta}{q_n(q_n + \theta q_{n-1})}$.

Therefore, using a little algebra,

$$\frac{\psi_n(y) - \psi_n(x)}{\psi_n(\theta) - \psi_n(0)} = \frac{y-x}{\theta}\,\frac{1}{1 + y\frac{q_{n-1}}{q_n}}\,\frac{1}{1 - \frac{(\theta - x)q_{n-1}}{q_n + \theta q_{n-1}}}. \tag{3.1}$$

Now $\frac{q_n}{q_{n-1}} \ge \theta$ and hence, the right hand side of (3.1) $\ge \frac{y-x}{2\theta}$.

Again, $\frac{q_{n-1}}{q_n + \theta q_{n-1}} \le \frac{1}{2\theta}$ so that $1 - \frac{(\theta - x)q_{n-1}}{q_n + \theta q_{n-1}} \ge \frac{1}{2}$ and hence the right hand side of (3.1) is $\le \frac{2(y-x)}{\theta}$.

So,

$$\frac{y-x}{2\theta} \le \frac{\psi_n(y) - \psi_n(x)}{\psi_n(\theta) - \psi_n(0)} \le \frac{2(y-x)}{\theta}.$$

Or,

$$\frac{y-x}{2\theta} \le \lambda(T^{-n}[x,y)|\Delta_n) \le \frac{2(y-x)}{\theta}.$$

Hence, for any Borel set A also, we have,

$$\frac{\lambda(A)}{2\theta} \le \lambda(T^{-n}(A)|\Delta_n) \le \frac{2\lambda(A)}{\theta}. \tag{3.2}$$

Now, since $0 \le x < \theta$,

$$\frac{1}{\log(1+\theta^2)}\frac{\theta}{1+\theta^2} \le \frac{1}{\log(1+\theta^2)}\frac{\theta}{1+\theta x} \le \frac{\theta}{\log(1+\theta^2)}.$$

Hence, for any Borel set M, we have,

$$\frac{1}{\log(1+\theta^2)}\frac{\theta}{1+\theta^2}\lambda(M) \le P(M) \le \frac{\theta}{\log(1+\theta^2)}\lambda(M). \tag{3.3}$$

So, $\lambda(M) \le \frac{1+\theta^2}{\theta}\log(1+\theta^2)P(M)$ and $\lambda(M) \ge \frac{\log(1+\theta^2)}{\theta}P(M)$.

Therefore, using these inequalities together with (3.2) and (3.3), we get the following :

$$C_1(\theta)P(A) \le P(T^{-n}(A)|\Delta_n) \le C_2(\theta)P(A)$$

where C_1, C_2 are constants depending on θ only. Now if A is invariant, the above inequality becomes

$$C_1(\theta)P(A) \le P(A|\Delta_n) \le C_2(\theta)P(A).$$

Assuming $P(A) > 0$, we get,

$$C_1(\theta)P(\Delta_n) \le P(\Delta_n|A) \le C_2(\theta)P(\Delta_n).$$

Hence, for any Borel set E,

$$C_1(\theta)P(E) \le P(E|A) \le C_2(\theta)P(E).$$

Taking $E = A^c$, one gets $P(A^c) = 0$ so that $P(A) = 1$. Therefore, T is ergodic under P.

We now proceed towards proving (*) and (**) in our case. As mentioned earlier already, we verify that the same steps as in the usual case go through.

Recall that (*) and (**) are :

$$\frac{a_1 + \cdots + a_n}{n} \longrightarrow \infty \quad \text{a.e.} \tag{*}$$

$$\lim_{n\to\infty}\frac{1}{n}\log q_n = \gamma \quad \text{a.e. for some finite number} \quad \gamma, \tag{**}$$

By ergodic theorem, if f is any non-negative function on $[0,\theta]$, integrable or not, we have,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1} f(T^k(\omega)) = \frac{1}{\log(1+\theta^2)}\int_0^\theta \frac{\theta f(x)}{1+\theta x}dx \quad \text{a.e.} \quad [P].$$

Taking $f = a_1$, the right hand side becomes,

$$\frac{1}{\log(1+\theta^2)}\int_0^\theta \frac{\theta a_1(x)}{1+\theta x}dx = \sum_{k=l}^{\infty}\frac{1}{\log(1+\theta^2)}\int_{\frac{1}{(k+1)\theta}}^{\frac{1}{k\theta}}\frac{k\theta}{1+\theta x}dx$$

$$= \frac{1}{\log(1+\theta^2)}\sum_{k=l}^{\infty} k\log(1+\frac{1}{k^2+2k}) = \infty.$$

The last equality follows from the fact that , $\log(1+x)$ is like x for x near 0 (This can be made precise). So, left hand side becomes

$$\lim_{n\to\infty}\sum_{k=0}^{n-1} a_1(T^k(\omega)) = \lim_{n\to\infty}\sum_{k=1}^{n} a_k(\omega) = \infty \quad \text{a.e.} \quad [P].$$

This proves (*).

To prove (**), first of all notice that,

$$\frac{1}{q_n(\omega)} = \prod_{k=1}^{n}\frac{p_{n+1-k}(T^{k-1}(\omega))}{q_{n+1-k}(T^{k-1}(\omega))}. \tag{3.4}$$

Also, from (2.8),

$$|\, x - \frac{p_n}{q_n} \,| \leq \frac{1}{q_n(x)q_{n+1}(x)}$$

Or,

$$|\, \frac{x}{\frac{p_n}{q_n}} - 1 \,| \leq \frac{1}{p_n(x)q_{n+1}(x)} \leq \frac{1}{(1+\theta^2)^n}$$

Or,

$$|\log(\frac{x}{\frac{p_n}{q_n}})\,| \leq \log(1+\frac{1}{(1+\theta^2)^n}) \leq \frac{1}{(1+\theta^2)^n} \tag{3.5}$$

So, using (3.4) and (3.5),

$$\begin{aligned}
&\mid \log[a_k(\omega)\theta, a_{k+1}(\omega)\theta, \cdots] - \log[a_k(\omega)\theta, a_{k+1}(\omega)\theta, \cdots, a_n(\omega)\theta] \mid \\
&=\mid \log(T^{k-1}(\omega)) - \log[a_1(T^{k-1}(\omega))\theta, a_2(T^{k-1}(\omega))\theta, \cdots, a_{n-k+1}(T^{k-1}(\omega))\theta] \mid \\
&=\mid \log(T^{k-1}(\omega)) - \frac{p_{n+1-k}(T^{k-1}(\omega))}{q_{n+1-k}(T^{k-1}(\omega))} \mid \leq \frac{1}{(1+\theta^2)^{n-k+1}}.
\end{aligned}$$

Then, summing over k from 1 to n and dividing by n, we get,

$$\begin{aligned}
&\frac{1}{n}\log\frac{1}{q_n(\omega)} = \frac{1}{n}\log\prod_{k=1}^{n}\frac{p_{n+1-k}(T^{k-1}(\omega))}{q_{n+1-k}(T^{k-1}(\omega))}. \\
&= \frac{1}{n}\sum_{k=1}^{n}\log(T^{k-1}(\omega)) + \frac{1}{n}\sum_{k=1}^{n}\frac{\zeta_{n,k}}{(1+\theta^2)^{n-k+1}}
\end{aligned}$$

for some numbers $\zeta_{n,k}$ which are smaller than one in modulus.

The second summand on the right hand side converges to zero since $\sum_{i=1}^{\infty}\frac{1}{(1+\theta^2)^i}$ is finite. The function $\log x$ being integrable on $[0,\theta)$ (w.r.t. P), the ergodic theorem implies that the first summand on the right hand side converges to

$$\frac{1}{1+\theta^2}\int_0^\theta \frac{\theta\log x}{1+\theta x}dx.$$

Hence, we have,

$$\frac{1}{n}\log\frac{1}{q_n(\omega)} \longrightarrow \frac{1}{1+\theta^2}\int_0^\theta \frac{\theta\log x}{1+\theta x}dx. \tag{3.6}$$

Hence, the limit is finite.

This proves (**).

Thus we have the following :

Theorem : Let $\theta = \frac{1}{\sqrt{l}}$, $l \in I\!N$. Then μ given by

$$d\mu(x) = \frac{1}{\log(1+\theta^2)}\frac{\theta}{1+\theta x}dx$$

is invariant and ergodic for the generalised Gauss map T on $[0,\theta)$ defined by

$$T(x) = \begin{cases} \frac{1}{x} - \theta[\frac{1}{\theta x}] & \text{if} \quad x \neq 0, \\ 0 & \text{if} \quad x = 0 \end{cases}$$

Moreover, for a.e. $x = [a_1\theta, a_2\theta, \cdots]$ we have,

$$\lim_{n\to\infty} \frac{a_1 + \cdots + a_n}{n} = \infty.$$

Further, there is a finite number γ, such that if $\frac{p_n}{q_n}$ denotes the n-th convergent of x, then for a.e. x, we have,

$$\lim_{n\to\infty} \frac{1}{n} \log q_n = \gamma.$$

Section 4 : Case of general θ's

In case θ^2 is not the reciprocal of a positive integer, $\frac{1}{\theta}$ has either finite or infinite continued fraction expansion with respect to θ. We shall mainly concentrate on θ's for which $\frac{1}{\theta}$ has finite continued fraction expansion. At the end of the section, we make few comments regarding θ's where the corresponding expansion is non-terminating. From now on, whenever we shall write a set as the union of a finitely or infinitely many sets, the equality will be upto a set of Lebesgue measure zero.

When $\theta^2 = \frac{1}{l}$ for some integer l then $[0,\theta) = \bigcup_{k\geq l}(\frac{1}{(k+1)\theta}, \frac{1}{k\theta})$ where each of the intervals in the union is mapped onto $[0,\theta)$ by T. This gave rise to a description of the invariant probability explicitly. In the general case, it is not so. For example, if $\frac{1}{\theta} = [n_1\theta; n_2\theta]$, then all the intervals $(\frac{1}{(k+1)\theta}, \frac{1}{k\theta})$ for $k \geq n_1 + 1$ are contained in $[0,\theta)$ and each of these is mapped onto $[0,\theta)$ by T. However, these intervals alone do not make up $[0,\theta)$, there is a leftover interval, namely, $(\frac{1}{(n_1+1)\theta}, \theta)$. At first sight it appears that we can partition this into intervals as earlier so that each of them is mapped onto $[0,\theta)$ by T^2.

Unfortunately this is not so. Even if θ is general —- $\frac{1}{\theta}$ not necessarily having expansion terminating at the second stage —- the leftover interval mentioned above remains the same. A little care is needed. What we do is to partition $(\frac{1}{(n_1+1)\theta}, \theta)$ into countably many intervals such that each of these intervals is mapped onto $[0, \theta)$ by an appropriate power of T (the power depending on the interval of the partition). We shall then define a new map T^* by putting it to be the corresponding power of T on each of these intervals separately. All this will be made precise in what follows.

This naturally brings us to the setup of sections 2 and 3, chapter V of [10] which we shall now briefly explain (Renyi [28] was perhaps the first to discuss these problems, though he did not introduce the set up formally) . In order to do that, we first of all, need the definition of a Markov map.

Definition 4.1. Let I be a bounded interval. Let $T : I \mapsto I$ be a C^1 map. Then T is called a **Markov** map if there exists a finite or countable family I_i of disjoint open intervals in I such that the following hold

(a) $I - \bigcup_i I_i$ has Lebesgue measure zero. There exist $C > 0$ and $\delta > 0$ so that for each $n \in I\!N$, and each interval J such that $T^j(J)$ is contained in one of the intervals I_i for each $j = 0, 1, 2, \cdots, n$ one has

$$\left|\frac{DT^n(x)}{DT^n(y)} - 1\right| \leq C \mid T^n(x) - T^n(y) \mid ^{\delta} \quad \forall x, y \in J.$$

(b) If $T(I_k) \cap I_j \neq \emptyset$ then $T(I_k) \supset I_j$.

(c) There exists $d > 0$ such that $|T(I_i)| \geq d \quad \forall i$.

Remark 4.1. The assumption (a) in the above definition can be replaced by the following two conditions (see [10], p.351):

(i) There exist $C, \delta > 0$ such that $T|I_i$ is a $C^{1+\delta}$-diffeomorphism for each i and that for each i and for each $x, y \in I_i$, the following Holder condition is satisfied :

$$\left|\frac{DT(x)}{DT(y)} - 1\right| \leq C \mid T(x) - T(y) \mid ^{\delta}.$$

(ii) T is expanding in the following sense : There exist $K > 0, \beta > 1$ so that $| DT^n(x) | \geq K\beta^n \quad \forall n \in I\!N$ and $x \in I$ for which $T^j(x) \in \bigcup_i I_i \quad \forall\, 0 \leq j \leq n$.

Having defined Markov map, we are interested in the existence of absolutely continuous invariant measures for such maps. The following theorem guarantees that :

Theorem 4.1. Let $T : I \rightarrow I$ be a Markov map and let $\bigcup_i I_i$ be the corresponding partition.Then there exists a T-invariant absolutely continuous probability measure μ, on the Borel subsets of I and it has the following properties :

(a) Its density $\frac{d\mu}{d\lambda}$ is uniformly bounded and Holder continuous. Moreover, for each i the density is either 0 on I_i or uniformly bounded away from 0.

(b) If for every i and j there is an $n \geq 1$ such that one has $T^n(I_j) \supset I_i$ then the measure is unique and ergodic, its density $\frac{d\mu}{d\lambda}$ is strictly positive.

(c) If $T(I_i) = I$ for each i, then the density of μ is also uniformly bounded from below.

In our case, T is not a Markov map. So, we need to bring in another idea – namely, that of an induced map. Suppose I is a bounded interval and $T : I \mapsto I$ be a C^1 map. Suppose we have a partition $(I_i,\, i \geq 1)$ of I and natural numbers k_i for $i \geq 1$. Define $T^* : I \mapsto I$ by putting $T^*(x) = T^{k_i}(x)$ if $x \in I_i$. Then T^* is called a map induced by T, more precisely it is the map induced by T and (I_i, k_i), $i \geq 1$. In fact it is not necessary to ensure that the intervals I_i cover all of I, it is enough if they cover upto a set of Lebesgue measure zero. We do need in this generality for the applications we have in mind.

Theorem 4.2. Suppose T^* is a map induced by T and (I_i, k_i), $i \geq 1$. Suppose T^* is a Markov map. Let ν be absolutely continuous invariant measure for T^*. If

$$\sum_{i=1}^{\infty} k_i\, \nu(I_i) < \infty \qquad (4.1)$$

then T has an absolutely continuous invariant probability measure. If

(4.2) $$\sum_{i=1}^{\infty} k_i |I_i| < \infty.$$

then (4.1) holds.

At this point, we would like to make a simple observation, which is not explicitly stated in [10]. To prove the theorem quoted above, one simply takes

$$\mu(A) = \sum_{n=1}^{\infty} \sum_{i=0}^{k_n-1} \nu(T^{-i}A \cap I_n)$$

for all Borel sets A and verifies that μ is indeed absolutely continuous finite measure which is invariant for T. We wish to now observe that if ν is ergodic for T^* (as we usually have from Theorem 4.1) , then the μ given above for T is also ergodic. Indeed if A is invariant for T, i.e., $T^{-1}A = A$ then $\mu(A) = \sum_{n=1}^{\infty} k_n \nu(A \cap I_n)$. If moreover, $\mu(A) > 0$ then for some n we must have $\nu(A \cap I_n) > 0$ and hence $\nu(A) > 0$ as well. Since A is invariant for T^* as well (at any point x, T^* is simply an appropriate power of T), we conclude that $\nu(A^c) = 0$ under the assumption that ν is ergodic for T^*. But then $\mu(A^c) = \sum_{n=1}^{\infty} k_n \nu(A^c \cap I_n) = 0$, showing that μ is ergodic for T.

Our plan now is as follows : We shall implement Theorem 4.2 in our set up by a careful choice of T^* thereby getting an absolutely continuous invariant probability measure for T wherever possible. Next we show the analogues of the two crucial facts (*) and (**) in our present setup. We proceed to execute the plan now.

(A) The case when expansion of $1/\theta$ terminates at the second stage :

To understand the execution without much notational complications, we begin our discussion with the case $\frac{1}{\theta} = [n_1\theta; n_2\theta]$. Then $T(\theta) = [n_2\theta] = \frac{1}{n_2\theta}$. Because of our discussion in section 2, $n_2 > n_1 + 1$. As explained above, we shall concentrate on the interval $(\frac{1}{(n_1+1)\theta}, \theta)$ where we are to replace T by suitable powers of T towards getting an induced map T^*. Denote $J = (\frac{1}{(n_1+1)\theta}, \theta)$.

Now, $T(J) = (T(\theta), \theta) = (\frac{1}{n_2\theta}, \theta) = (\frac{1}{n_2\theta}, \frac{1}{(n_1+1)\theta}) \cup (\frac{1}{(n_1+1)\theta}, \theta)$. Or equivalently, $T(J) = (\frac{1}{n_2\theta}, \frac{1}{(n_1+1)\theta}) \cup J$.

Therefore, $J = T_J^{-1}(\frac{1}{n_2\theta}, \frac{1}{(n_1+1)\theta}) \cup T_J^{-1}(J)$ where, for any A, we set $T_J^{-1}(A) = \{x \in J : T(x) \in A\}$. Since

$$(\frac{1}{n_2\theta}, \frac{1}{(n_1+1)\theta}) = \bigcup_{k=1}^{n_2-n_1-1} (\frac{1}{(n_1+k+1)\theta}, \frac{1}{(n_1+k)\theta}),$$

each of the intervals in the union on the right hand side is mapped by T onto $(0,\theta)$, so $T_J^{-1}(\frac{1}{n_2\theta}, \frac{1}{(n_1+1)\theta})$ is mapped by T^2 onto $(0,\theta)$.

Let us call $V = T_J^{-1}(\frac{1}{n_2\theta}, \frac{1}{(n_1+1)\theta})$. Then,

$$J = V \cup T_J^{-1}J = V \cup T_J^{-1}V \cup T_J^{-2}V = \cdots$$

We now argue that indeed

$$J = \bigcup_{i=0}^{\infty} T_J^{-i}V.$$

A simple calculation shows that

$$V = ([n_1\theta, (n_1+1)\theta], \theta)$$

$$T_J^{-1}V = ([(n_1+1)\theta], [n_1\theta, n_1\theta, (n_1+1)\theta])$$

$$T_J^{-2}V = ([n_1\theta, n_1\theta, n_1\theta, (n_1+1)\theta], [n_1\theta, (n_1+1)\theta])$$

In general

$$T_J^{-2n}V = (a_{2n}, b_{2n}) \quad \text{where}$$

$$a_{2n} = [n_1\theta \text{ repeated } (2n+1) \text{ times }, (n_1+1)\theta]$$

$b_{2n} = [n_1\theta \text{ repeated } (2n-1) \text{ times }, (n_1+1)\theta]$

$T_J^{-(2n+1)}V = (a_{2n+1}, b_{2n+1})$ where

$a_{2n+1} = [n_1\theta \text{ repeated } 2n \text{ times }, (n_1+1)\theta]$

$b_{2n+1} = [n_1\theta \text{ repeated } (2n+2) \text{ times }, (n_1+1)\theta]$

Their lengths add up, in a telescpoic manner, to $\theta - [(n_1+1)\theta]$ which is the length of J as desired.

Notice that V is a subinterval of J and for each $i \geq 1$, $T_J^{-i}V$ is also a subinterval of J. T is a one to one map on J. Note also, for future use, that even though T^2 maps V onto $(0, \theta)$, in order to talk about the C^1 properties of T^2 on V, we need to split the interval V as follows :

$$V = \bigcup_{k=1}^{n_2-n_1-1} T_J^{-1}\left(\frac{1}{(n_1+k+1)\theta}, \frac{1}{(n_1+k)\theta}\right),$$

where each of the sets in the union on the right hand side are again intervals on each of which T^2 is a C^1 map. Similar considerations apply for $T_J^{-i}V$ as well for $i \geq 1$.

All this can equivalently be described as follows. Let T_J be the restriction of T to J. Then $V = \{x \in J : T_J x \notin J\}$ and in general, $T_J^{-i}V = \{x \in J : T_J x, \cdots, T_J^i x \in J \text{ but } T_J^{i+1}x \notin J\}$.

So, our induced map T^* is as follows :

$$(4.3) \qquad T^* = T^{i+2} \quad \text{on} \quad T^{-i}V, \quad i \geq 0.$$

As noted earlier, $J = \bigcup_{i=0}^{\infty} T_J^{-i}V$ and hence (4.3) above defines T^* on whole of J. We extend T^* to all of $[0, \theta)$ by defining $T^* = T$ on $[0, \theta) - J$.

It is not difficult to show that T is not a Markov map. However, T^* is a Markov map as shown below.

Lemma 4.1. T^* is a Markov map.

Proof. The intervals I_i are as follows. For each $k \geq n_1+1$, $\left(\frac{1}{(k+1)\theta}, \frac{1}{k\theta}\right)$

is an I_i. For $1 \leq k \leq n_2 - n_1 - 1$, $T_J^{-1}(\frac{1}{(n_1+k+1)\theta}, \frac{1}{(n_1+k)\theta})$ is an I_i. and in general, for each $m \geq 1$, and $1 \leq k \leq n_2 - n_1 - 1$, $T_J^{-m}(\frac{1}{(n_1+k+1)\theta}, \frac{1}{(n_1+k)\theta})$ is an I_i. These are the only intervals. As seen above, these intervals cover $[0, \theta)$ except for a Lebesgue null set.

To verify condition (a) of Definition (4.1), we show that conditions (i) and (ii) of Remark 4.1 hold. First let us observe these conditions for T itself. If $x, y \in I_i$ for some $i \geq 1$, then after simplification, it follows that

$$\left|\frac{DT(x)}{DT(y)} - 1\right| \leq | x + y | \cdot \frac{y}{x} \cdot | T(x) - T(y) | \leq 4\theta \, | T(x) - T(y) | .$$

Thus choosing $C = 4\theta$ and $\delta = 1$ verifies condition (i) of Remark 4.1. Again,

$$| DT^n(x) | = \frac{1}{| T^{n-1}(x) T^{n-2}(x) \cdots T(x) x |^2} \geq \frac{1}{\theta^{2n}} \quad \forall x.$$

Thus choosing $K = 1$ and $\beta = \frac{1}{\theta^2}$ verifies (ii) of Remark 4.1.

So, (a) is satisfied for T. So, there exist $C, \delta > 0$ such that for each $n \in \mathbb{N}$ and each interval I such that $T^j(I)$ is contained in one of the intervals I_i for each $j = 0, 1, \cdots, n$, one has for all $x, y \in I$,

$$\left|\frac{DT^n(x)}{DT^n(y)} - 1\right| \leq C \, | T^n(x) - T^n(y) |^\delta .$$

We shall now verify (i) and (ii) of Remark 4.1 for T^*.

Regarding (i), observe that if $x, y \in I_i$ and $T^* = T^{k_i}$ on I_i and $T^j(I_i)$ is contained in $T^{j+1}(I_i)$ for $j = 0, 1, \cdots, k_i$ (note that it is enough to consider k_i's larger than one). Hence,applying condition (a) for T we have,

$$\left|\frac{DT^*(x)}{DT^*(y)} - 1\right| = \left|\frac{DT^{k_i}(x)}{DT^{k_i}(y)} - 1\right| \leq C \, | T^{k_i}(x) - T^{k_i}(y) |^\delta = C \, | T^*(x) - T^*(y) |^\delta .$$

Thus, condition (i) follows. Condition (ii) follows exactly as for T. The fact $T^*(I_i) = (0, \theta)$ for each i, shows that the conditions (b) and (c) of Definition 4.1 hold.

Hence, T^* is a Markov map.

Therefore, by Theorem (4.1), it does admit an absolutely continuous invariant probability measure ν, the density of which is uniformly bounded. Since $T^*(I_i) = (0, \theta)$, the density $p(x)$ is bounded below. Now if condition (4.2) is satisfied, T also admits an absolutely continuous invariant probability measure by Theorem (4.2). This is not always the case. We get a sufficient condition now.

Because of (4.3), we conclude that (4.2) holds if $\sum_{i=0}^{\infty}(i+2)\,|T_J^{-i}(V)| < \infty$. Recalling the description of V, $T_J^{-1}V$, $T_J^{-2}V, \cdots$ given earlier we note the following. If two points u, v are in V, then $u = \dfrac{1}{n_1\theta + x}$, $v = \dfrac{1}{n_1\theta + y}$ for some x, $y \in [0, \theta)$ so that

$$| u - v | \leq \frac{| x - y |}{(n_1\theta)^2} \leq \frac{\theta}{(n_1\theta)^2}$$

If u, $v \in T_J^{-1}V$, then $u = \dfrac{1}{n_1\theta + x}$, $v = \dfrac{1}{n_1\theta + y}$ for some x, $y \in V$ so that

$$| u - v | \leq \frac{| x - y |}{(n_1\theta)^2} \leq \frac{\theta}{(n_1\theta)^4}$$

Thus length of $V \leq \dfrac{\theta}{(n_1\theta)^2}$, length of $T_J^{-1}V \leq \dfrac{\theta}{(n_1\theta)^4}$ and in general, length of $T_J^{-i}V \leq \dfrac{\theta}{(n_1\theta)^{(2i+2)}}$.

Thus

$$\sum_{i=0}^{\infty}(i+2)\,|T_J^{-i}V| \leq \sum_{i=0}^{\infty}(i+2)\frac{\theta}{(n_1\theta)^{(2i+2)}}$$

As a consequence, (4.2) holds if $n_1\theta > 1$. For example, if $\dfrac{1}{\theta} - \theta > 1$ then this condition holds because $\dfrac{1}{\theta} - \theta < n_1\theta$. But $\dfrac{1}{\theta} - \theta > 1$ is same as saying $0 < \theta < \dfrac{\sqrt{5}-1}{2}$.

We now proceed to prove (*) and (**) for the situation mentioned above. For proving (*), in view of the Ergodic theorem it is sufficient to show that $\int a_1(y)\lambda(dy) = \infty$. We shall indeed show that $\int_A a_1(y)\lambda(dy) = \infty$ for all invariant sets A with positive probability. The reason is the following. Even though for certain values of θ we obtained an absolutely continuous ergodic measure with nice properties, it is conceivable (we do not know) that one may be able to obtain for certain other values of θ an absolutely continuous invariant measure with nice properties but not necessarily ergodic. In such a case, this argument shows that the conditional expectation of a_1 given the invariant σ-field equals infinity almost surely. Ergodic theorem would then yield that even in such a situation $(*)$ holds.

Thus , let A be an invariant set with positive probability. Let $J' = [0, \theta) - J$. Clearly, $A \cap J'$ also has positive probability. To see this just note that almost every point of J is taken to J' by some power of T. Fix a $k \geq n_1 + 1$. Let $A_k = (\frac{1}{(k+1)\theta}, \frac{1}{k\theta}) \cap A$ so that $A \cap J' = \bigcup_{k \geq n_1+1} A_k$ and hence not all A_k can have measure zero.Let $S = \{\alpha \in (0,1) : \frac{1}{(k+\alpha)\theta} \in A_k\}$. By invariance of A it is not difficult to see that S does not depend on k (as long as $k \geq n_1 + 1$). Moreover, since A_k has positive probability, S has positive Lebesgue measure.

Now on J', the density of μ is same as that of ν (upto a normalizing constant) as shown in the construction of μ. So, by Theorem 4.1, it is uniformly bounded below by some constant $c > 0$. Thus, it is enough to show that $\int_{A \cap J'} a_1(y)\lambda(dy) = \infty$. We show this as follows :

$$\int_{A \cap J'} a_1(y)\lambda(dy) = \sum_{k=n_1+1}^{\infty} \int_{A_k} a_1(y)\lambda(dy) = \sum_{k=n_1+1}^{\infty} \int_S \frac{k}{((k+y)\theta)^2}\lambda(dy)$$

$$= \int_S \sum_{k=n_1+1}^{\infty} \frac{k}{((k+y)\theta)^2}\lambda(dy).$$

Now, for each $y \in S$, the integrand is infinity and so, the integral is also infinity, as required.

To verify $(**)$, we proceed exactly as in section 3 and we appear at the following step :

$$\lim_{n \to \infty} \frac{1}{n} \log(\frac{1}{q_n(\omega)}) = E(\log x | \mathcal{I}) \quad \text{a.e.}[\mu].$$

To show that this limit is finite a.e., we argue that the function $\log x$ is integrable w.r.t. μ. On J, the integral is finite because μ is a probability and $\log x$ is bounded on J. As regards J', firstly, the densities of μ and ν are same (upto a normalizing constant). Secondly, by Theorem 4.1, the density of ν is bounded. Thirdly, the function $\log x$ is integrable w.r.t. Lebesgue measure. These three observations show that the function $\log x$ is integrable on J' as well, completing the proof.

Thus we have the following :

Theorem : Let $0 < \theta < \frac{\sqrt{5}-1}{2}$ and that $\frac{1}{\theta}$ has expansion terminating at the second stage. Then the generalised Gauss map on $[0, \theta)$ admits an ergodic invariant probability which is equivalent to the Lebesgue measure. Moreover, for a.e. $x = [a_1\theta, a_2\theta, \cdots]$ we have,

$$\lim_{n\to\infty} \frac{a_1 + \cdots + a_n}{n} = \infty.$$

Further, there is a finite number γ, such that if $\frac{p_n}{q_n}$ denotes the n-th convergent of x, then for a.e. x, we have,

$$\lim_{n\to\infty} \frac{1}{n} \log q_n = \gamma.$$

(B) The general terminating case :

Now we consider a little more general case. First we consider the case when $\frac{1}{\theta}$ has finite continued fraction expansion w.r.t. θ, say terminating at the m-th stage given by

$$\frac{1}{\theta} = [n_1\theta, \cdots, n_m\theta].$$

Since there is a difference between the odd and even cases of m, we assume from now on that m is even. The difference being minor, we shall not consider the other case. We shall also assume that $n_2, \cdots, n_m$ are at least as large as $n_1 + 1$.

The case considered earlier corresponds to $m = 2$. The basic idea is same even for the present case. Only the partitioning of $[0, \theta)$ gets complicated which we are going to explain. Earlier each of the intervals $(\frac{1}{(k+1)\theta}, \frac{1}{k\theta})$ was an I_i. It will be so even now. But earlier, towards getting the remaining I_is in the partition, as a first step, we decomposed $J = (\frac{1}{(n_1+1)\theta}, \frac{1}{n_1\theta})$ into two parts : one part consisting of points which go to $(0, \frac{1}{(n_1+1)\theta})$ and the other part remaining within J after an application of the map T. It so happened that the image under T of the points that leave J is a union of some of the intervals $(\frac{1}{(k+1)\theta}, \frac{1}{k\theta})$, $k \geq n_1 + 1$. We defined T^* to be T^2 on these points of J. The image under T of the points in the second part was whole of J. So, we again split the set and continuing the process, we were able to get a partition of J into intervals each of which is mapped onto $[0, \theta)$ by some power of T. In the present case, that is when $m > 2$, the range of T on those points of J that leave J is not only made up of some full intervals of the form $(\frac{1}{(k+1)\theta}, \frac{1}{k\theta})$, $(k \geq n_1+1)$ but also a partial interval. To be a little more precise, unlike the previous case, in the first step of the decomposition, we now have three parts : the first part that leaves J under an application of T and consists of some intervals of the form $(\frac{1}{(k+1)\theta}, \frac{1}{k\theta})$, $(k \geq n_1 + 1)$; the second part also leaves J under an application of T and consists of a partial interval; and the third part remains within J after an application of T. So, here in the next step, we have to split both the parts, the second and the third part. This is done by applying T^2 on these points. We proceed this way. But now, another complication arises. While carrying out the process, it may so happen that a point returns back to J. So, we have to go on splitting until we partition J into intervals which are mapped onto $[0, \theta)$ by some integral power of T. This is achieved as follows :

As earlier, let $J = (\frac{1}{(n_1+1)\theta}, \theta)$. Then, $TJ = (T\theta, \theta)$. Note that $T\theta = [n_2\theta, \cdots, n_m\theta]$. So, we have our first step decomposition,

$$J = R_1 \cup J_1 \cup V_1, \text{ say },$$

where

$$R_1 = \{x \in J : Tx \in J\}$$

$$J_1 = \{x \in J : Tx \in (T\theta, \frac{1}{n_2\theta})\}$$

$$V_1 = \{x \in J : Tx \in (\frac{1}{n_2\theta}, \frac{1}{(n_1+1)\theta})\}$$

Each of these sets on the right side is a subinterval of J. R_1 is the set of points that come back to J. We leave them untouched at this stage. TV_1 is a union of intervals of the form $(\frac{1}{(k+1)\theta}, \frac{1}{k\theta})$, $(k \geq n_1+1)$ and so we need not do anything for the points in TV_1. We have to handle J_1 further. We shall split this as follows. Note that $T^2J_1 = (0, T^2\theta)$ and $T^2\theta = [n_3\theta, \cdots, n_m\theta]$. So, we have the second step decomposition,

$$J_1 = J_2 \cup V_2, \text{ say .}$$

where

$$J_2 = \{x \in J_1 : T^2x \in (\frac{1}{(n_3+1)\theta}, T^2\theta)\}$$

$$V_2 = \{x \in J_1 : T^2x \in (0, \frac{1}{(n_3+1)\theta})\}$$

At this stage no point returned to J on application of T^2 on J_1 and so we do not have analogue of R_1 or equivalently, we treat (for notational convenience) $R_2 = \emptyset$. As before, T^2V_2 is a union of intervals of the form $(\frac{1}{(k+1)\theta}, \frac{1}{k\theta})$, $(k \geq n_1+1)$ and we need not do anything for these points. We handle and split J_2. Note that $T^3J_2 = (T^3\theta, \theta)$ and $T^3\theta = [n_4\theta, \cdots, n_m\theta]$. So, we have the third step decomposition,

$$J_2 = R_3 \cup J_3 \cup V_3 \text{ say ,}$$

where

$$R_3 = \{x \in J_2 : T^3x \in J\}$$

$$J_3 = \{x \in J_2 : T^3x \in (T^3\theta, \frac{1}{n_4\theta})\}$$

$$V_3 = \{x \in J_2 : T^3x \in (\frac{1}{n_4\theta}, \frac{1}{(n_1+1)\theta})\}$$

In general, if p is odd and less than m, then $T^pJ_{p-1} = (T^p\theta, \theta)$ and $T^p\theta = [n_{p+1}\theta, \cdots, n_m\theta]$ so that p-th step decompostion is,

$$J_{p-1} = R_p \cup J_p \cup V_p \text{ say },$$

where

$$R_p = \{x \in J_{p-1} : T^px \in J\}$$

$$J_p = \{x \in J_{p-1} : T^px \in (T^p\theta, \frac{1}{n_{p+1}\theta})\}$$

$$V_p = \{x \in J_{p-1} : T^px \in (\frac{1}{n_{p+1}\theta}, \frac{1}{(n_1+1)\theta})\}$$

On the other hand, if p is even and less than m, then $T^pJ_{p-1} = (0, T^p\theta)$ and $T^p\theta = [n_{p+1}\theta, \cdots, n_m\theta]$ so that p-th step decomposition is,

$$J_{p-1} = J_p \cup V_p, \text{ say }.$$

$$J_p = \{x \in J_{p-1} : T^px \in (\frac{1}{(n_{p+1}+1)\theta}, T^p\theta)\}$$

$$V_p = \{x \in J_{p-1} : T^px \in (0, \frac{1}{(n_{p+1}+1)\theta})\}$$

For the same reasons as mentioned earlier in the second step decomposition, we do not have R_p in this case. however for notational convenience we shall take $R_p = \emptyset$.

Thus apart from a Lebesgue null set of points, J is written as union of intervals,

$$J = \bigcup_{p=1}^{m-1} V_p \cup \bigcup_{p=1}^{m-1} R_p.$$

Of course, for p even, $R_p = \emptyset$. Not only that, note that for each odd p, $T^p R_p = J$ and for $q < p$, $T^q R_p$ does not cover any full interval of the form $(\frac{1}{(k+1)\theta}, \frac{1}{k\theta})$, $(k \geq n_1 + 1)$. For odd $j \leq m-1$, let T_j denote T restricted to R_j. Then, for each such j, we again have,

$$R_j = T_j^{-j} J = \bigcup_{p=1}^{m-1} T_j^{-j} V_p \cup \bigcup_{p=1}^{m-1} T_j^{-j} R_p$$

where the second union extends only over odd integers $p < m$.

Then, we can write J as follows :

$$J = \bigcup_{p=1}^{m-1} V_p \cup \bigcup_{j_1=1}^{m-1} \bigcup_{p=1}^{m-1} T_{j_1}^{-j_1} V_p \cup \bigcup_{j_1=1}^{m-1} \bigcup_{j_2=1}^{m-1} T_{j_1}^{-j_1} R_{j_2}.$$

Here j_1, j_2 range over odd integers less than m and p ranges from 1 to $m-1$. We shall denote T_{j_1,j_2} to be T restricted to the set $\{x : T_{j_1}^{j_1} x \in R_{j_2}\}$. Since the domain of T_{j_1} is already R_{j_1} this set is actually a subset of R_{j_1}. In general, for odd integers $j_1, \cdots, j_r$ less than m, let $T_{\{j_1,j_2,\cdots,j_r\}}$ denote T restricted to $\{x : T_{\{j_1,j_2,\cdots,j_{r-1}\}}^{j_1+j_2+\cdots+j_{r-1}} x \in R_{j_r}\}$. This is defined by induction on r. Then,

$$J = \bigcup_{p=1}^{m-1} \bigcup_{r=0}^{\infty} \bigcup_{(j_1,\cdots,j_r)} T_{\{j_1,j_2,\cdots,j_r\}}^{-(j_1+j_2+\cdots+j_r)} V_p.$$

The above union is upto a null set. Here $j_1, j_2, \cdots$ range over odd integers less than m. When $r = 0$, $\langle j_1, \cdots, j_r \rangle$ simply means empty sequence and in that case $T_{\{j_1,j_2,\cdots,j_r\}}^{-(j_1+j_2+\cdots+j_r)} V_p$ simply means V_p. With this understanding, we describe the partition of $[0, \theta)$. Each of the intervals $(\frac{1}{(k+1)\theta}, \frac{1}{k\theta})$ for $k \geq n_1 + 1$ is an I_i. And on such an interval $T^* = T$. Each of the sets V_p is a union of certain intervals each of which is mapped onto $[0, \theta)$ by T^{p+1}. These subintervals are also I_i s and on each of them, we put $T^* = T^{p+1}$. Then $T_{j_1}^{-j_1} V_p$ is similarly a union of certain intervals each of which is mapped onto $[0, \theta)$ by T^{j_1+p+1}. Each of these intervals is again an I_i and we put $T^* = T^{j_1+p+1}$

on them. In general, $T^{-(j_1+j_2+\cdots+j_r)}_{\{j_1,j_2,\cdots,j_r\}}V_p$ is in a similar way, a union of certain intervals each of which is mapped onto $[0,\theta)$ by $T^{j_1+\cdots+j_r+p+1}$ and we put $T^* = T^{j_1+\cdots+j_r+p+1}$ on them.

Exactly, as in the case of $m = 2$, we can show that T^* is Markov map. Thus T^* admits an absolutely continuous invariant probability. We shall now proceed to obtain conditions for (4.2) to hold, so that T also admits an absolutely continuous invariant probability. If $x, y \in T^{-(j_1+j_2+\cdots+j_r)}_{\{j_1,j_2,\cdots,j_r\}}V_p$, then,

$$| x - y | \leq \frac{1}{(n_1\theta)^{2(j_1+\cdots+j_r)}}.$$

This is because of the assumption that n_is are at least n_1+1 for $2 \leq i \leq m-1$. Here, $j_1, \cdots, j_r$ range only over odd integers less than m. In this case, (4.2) reduces to showing that the following sum is finite:

$$\sum_{p=1}^{m-1} |V_p| \left[p + 1 + \sum_{r=1}^{\infty} \sum_{(i_1,\cdots,i_r)} \frac{p+1+\sum_{j=1}^{r}(2i_j - 1)}{(n_1\theta)^{2\sum_{j=1}^{r}(2i_j-1)}} \right].$$

Here we denoted $\frac{m}{2}$ by k (note that m is even). Further, $i_1, \cdots, i_r$ range from 1 to k so that $2i_j - 1$ range over odd integers less than m. This change is made to facilitate in the calculation of the sums involved.

Now, to find conditions for the above sum to be finite we proceed as follows. Since outer sum consists of finitely many terms, it is enough to get conditions when

$$\sum_{r=1}^{\infty} \sum_{(i_1,\cdots,i_r)} \frac{p+1+\sum_{j=1}^{r}(2i_j - 1)}{(n_1\theta)^{2\sum_{j=1}^{r}(2i_j-1)}} < \infty.$$

The above sum

$$\leq \sum_{r=1}^{\infty} \sum_{s=r}^{kr} \sum_{i_1+\cdots+i_r=s} \frac{p+1+2s-r}{(n_1\theta)^{4s-2r}}$$

$$= \sum_{r=1}^{\infty} \sum_{t=0}^{(k-1)r} \sum_{i_1+\cdots+i_r=t} \frac{p+1+2t+r}{(n_1\theta)^{4t+2r}}$$

$$\leq \sum_{r=1}^{\infty} \sum_{t=0}^{\infty} \sum_{i_1+\cdots+i_r=t} \frac{p+1+2t+r}{(n_1\theta)^{4t+2r}}$$

$$= \sum_{r=1}^{\infty} \sum_{t=0}^{\infty} \binom{t+r-1}{r} \frac{p+1+2t+r}{(n_1\theta)^{4t+2r}}$$

$$= \sum_{r=1}^{\infty} \frac{2}{(n_1\theta)^{2r}} \sum_{t=0}^{\infty} \binom{t+r-1}{r} \frac{t}{(n_1\theta)^{4t}} \ +$$

$$\sum_{r=1}^{\infty} \frac{r+p+1}{(n_1\theta)^{2r}} \sum_{t=0}^{\infty} \binom{t+r-1}{r} \frac{1}{(n_1\theta)^{4t}}$$

Clearly if $n_1\theta \leq 1$ the above sums do not converge. So let us assume that from now on $n_1\theta > 1$. Note that the second sum in the second term is $\frac{1}{(1-\frac{1}{(n_1\theta)^4})^r}$-times the sum of the negative binomial probabilities with parameters $1 - \frac{1}{(n_1\theta)^4}$ and r, while the second sum in the first term is $\frac{1}{(1-\frac{1}{(n_1\theta)^4})^r}$-times the expectation of the same negative binomial distribution.

So, the above sum ,on simplification,becomes

$$\sum_{r=1}^{\infty} \frac{1}{(n_1\theta)^{2r}(1-\frac{1}{(n_1\theta)^4})^r} \left[\frac{2r}{(n_1\theta)^4 - 1} + r + p + 1\right].$$

This sum is finite iff $(n_1\theta)^2 - \frac{1}{(n_1\theta)^2} > 1$. Solving, we get, $n_1\theta > (\frac{\sqrt{5}+1}{2})^{\frac{1}{2}}$. Now, $n_1\theta > \frac{1}{\theta} - \theta$. So, if $\frac{1}{\theta} - \theta > (\frac{\sqrt{5}+1}{2})^{\frac{1}{2}}$, then $n_1\theta$ will be so. But the last inequality implies that if $\theta < \frac{\sqrt{c+4}-\sqrt{c}}{2}$ where $c = \frac{\sqrt{5}+1}{2}$, then (4.2) is satisfied in this case and an absolutely continuous invariant probability measure exists. But this is only a sufficient condition. Better bounds should exist for θ. Once an absolutely continuous invariant probability measure exists, the conditions (*) and (**) can be checked exactly in the same way as done for the case $m = 2$. (Note that, there we did not even use that θ has a terminating continued fraction expansion w.r.t. itself.)

Thus we have the following :

Theorem : Let $0 < \theta < \frac{\sqrt{c+4}-\sqrt{c}}{2}$ where $c = \frac{\sqrt{5}+1}{2}$. Also let

$$\frac{1}{\theta} = [n_1\theta, n_2\theta, \cdots, n_m\theta].$$

Assume that $n_2, \cdots, n_m$ are at least as large as $n_1 + 1$. Then the generalised Gauss map on $[0, \theta)$ admits an ergodic invariant probability which is equivalent to the Lebesgue measure. Moreover, for a.e. $x = [a_1\theta, a_2\theta, \cdots]$ we have,

$$\lim_{n\to\infty} \frac{a_1 + \cdots + a_n}{n} = \infty.$$

Further, there is a finite number γ, such that if $\frac{p_n}{q_n}$ denotes the n-th convergent of x,then for a.e. x, we have,

$$\lim_{n\to\infty} \frac{1}{n} \log q_n = \gamma.$$

(C) The non-terminating case :

Finally, we consider the case when $\frac{1}{\theta}$ has non-terminating continued fraction expansion , say,

$$\frac{1}{\theta} = [n_1\theta, n_2\theta, \cdots].$$

Then for any $j \geq 1$,

$$T^j(\theta) = [n_{j+1}\theta, n_{j+2}\theta, \cdots].$$

We shall only briefly indicate the arguments without going into the details. Proceeding as in the case of terminating expansion, we can define J_p's, V_p's for $p \geq 1$ and R_p's for odd positive integers. Of course now we have these sets indexed by p running over all the integers. Thus, we have, the following:

If p is odd, then $T^pJ_{p-1} = (T^p\theta, \theta)$ and

$$T^p\theta = [n_{p+1}\theta, n_{p+2}\theta, \cdots]$$

so that the p-th step decompostion is,

$$J_{p-1} = R_p \cup J_p \cup V_p\,, \text{ say },$$

$$R_p = \{x \in J_{p-1} : T^p x \in J\}$$

$$J_p = \{x \in J_{p-1} : T^p x \in (T^p\theta, \frac{1}{n_{p+1}\theta})\}$$

$$V_p = \{x \in J_{p-1} : T^p x \in (\frac{1}{n_{p+1}\theta}, \frac{1}{(n_1+1)\theta})\}$$

On the other hand, if p is even, then $T^p J_{p-1} = (0, T^p\theta)$ and

$$T^p\theta = [n_{p+1}\theta, n_{p+2}\theta, \cdots]$$

so that the p-th step decomposition is,

$$J_{p-1} = J_p \cup V_p\,, \text{ say }.$$

$$J_p = \{x \in J_{p-1} : T^p x \in (\frac{1}{(n_{p+1}+1)\theta}, T^p\theta)\}$$

$$V_p = \{x \in J_{p-1} : T^p x \in (0, \frac{1}{(n_{p+1}+1)\theta})\}$$

We do not write R_p in this case and treat $R_p = \emptyset$. Thus apart from a countable set of points, J is written as union of intervals,

$$J = \bigcup_{p=1}^{\infty} V_p \cup \bigcup_{p=1}^{\infty} R_p.$$

Note that for each odd p, $T^p R_p = J$ and for $q < p$, $T^q R_p$ does not cover any full interval of the form $(\frac{1}{(k+1)\theta}, \frac{1}{k\theta})$, $(k \geq n_1 + 1)$.

Then proceeding exactly as in the terminating case, for odd integers $j_1, \cdots, j_r \geq 1$, let $T_{\{j_1,j_2,\cdots,j_r\}}$ denote T restricted to $\{x : T^{j_1+j_2+\cdots+j_r-1}_{\{j_1,j_2,\cdots,j_{r-1}\}} x \in R_{j_r}\}$. This is defined by induction on r. Then,

$$J = \bigcup_{p=1}^{\infty} \bigcup_{r=0}^{\infty} \bigcup_{(j_1,\cdots,j_r)} T^{-(j_1+j_2+\cdots+j_r)}_{\{j_1,j_2,\cdots,j_r\}} V_p\,.$$

The above union is upto a Lebesgue null set. Here $j_1, j_2, \cdots$ range over odd positive integers. When $r = 0$, $\langle j_1, \cdots, j_r \rangle$ simply means empty sequence and in that case $T^{-(j_1+j_2+\cdots+j_r)}_{\{j_1,j_2,\cdots,j_r\}} V_p$ simply means V_p. With this understanding, we describe the partition of $[0, \theta)$. Each of the intervals $(\frac{1}{(k+1)\theta}, \frac{1}{k\theta})$ for $k \geq n_1 + 1$ is an I_i. And on such an interval $T^* = T$. Each of the sets V_p is a union of certain intervals each of which is mapped onto $[0, \theta)$ by T^{p+1}. These subintervals are also I_i s and on each of them, we put $T^* = T^{p+1}$. Then $T^{-j_1}_{j_1} V_p$ is similarly a union of certain intervals each of which is mapped onto $[0, \theta)$ by T^{j_1+p+1}. Each of these intervals is again an I_i and we put $T^* = T^{j_1+p+1}$ on them. In general, $T^{-(j_1+j_2+\cdots+j_r)}_{\{j_1,j_2,\cdots,j_r\}} V_p$ is in a similar way, a union of certain intervals each of which is mapped onto $[0, \theta)$ by $T^{j_1+\cdots+j_r+p+1}$ and we put $T^* = T^{j_1+\cdots+j_r+p+1}$ on them.

Exactly, as in the case of $m = 2$, we can show that T^* is a Markov map. Thus T^* admits an absolutely continuous invariant probability. We shall now proceed to obtain conditions for (4.2) to hold, so that T also admits an absolutely continuous invariant probability. If $x, y \in T^{-(j_1+j_2+\cdots+j_r)}_{\{j_1,j_2,\cdots j_r\}} V_p$, then,

$$| x - y | \leq \frac{1}{(n_1\theta)^{2(j_1+\cdots+j_r)}}.$$

This is because of the assumption that n_is are at least $n_1 + 1$ for $i \geq 2$. Here, $j_1, \cdots, j_r$ range only over odd positive integers.

In the present case, (4.2) reduces to showing that the following sum is finite:

$$\sum_{p=1}^{\infty} |V_p| \left[p + 1 + \sum_{r=1}^{\infty} \sum_{(i_1,\cdots,i_r)} \frac{p + 1 + \sum_{j=1}^{r}(2i_j - 1)}{(n_1\theta)^{2\sum_{j=1}^{r}(2i_j-1)}} \right].$$

After a slight change of notation (writing $j_r = 2i_r - 1$ for $r \geq 1$) the above summands are attained. This change is made to facilitate in the calculation of the sums involved.

So, the sum corresponding to (4.2) in our case is less than or equal to,

$$\sum_{p=1}^{\infty} |V_p| \left[p + 1 + \sum_{r=1}^{\infty} \sum_{(i_1,\cdots,i_r)} \frac{p + 1 + \sum_{j=1}^{r}(2i_j - 1)}{(n_1\theta)^{2\sum_{j=1}^{r}(2i_j-1)}} \right]. \tag{4.4}$$

We now get conditions for the above sum to be finite. Here, unlike the previous case, the outer sum consists of infinitely many terms.

Then, first of all arguing exactly as in the previous case, we show,

$$\sum_{r=1}^{\infty} \sum_{(i_1,\cdots,i_r)} \frac{p+1+\sum_{j=1}^{r}(2i_j-1)}{(n_1\theta)^{2\sum_{j=1}^{r}(2i_j-1)}} < \infty.$$

provided $\theta < \dfrac{\sqrt{c+4}-\sqrt{c}}{2}$ where $c = \dfrac{\sqrt{5}+1}{2}$.

Call this finite sum to be $D(\theta)+(p+1)E(\theta)$. Then, the sum in (4.4) equals,

$$\sum_{p=1}^{\infty} |V_p|[p+1+D(\theta)+(p+1)E(\theta)] = \sum_{p=1}^{\infty} |V_p|[D(\theta)+(p+1)E^*(\theta)].$$

where, $E^*(\theta) = E(\theta)+1$.

Now, one can show that,

$$|V_p| \leq \frac{2}{(n_1\theta)^{2p+1}}, \quad p = 1, 2, \cdots.$$

So the above sum is less than or equal to

$$\sum_{p=1}^{\infty} [D(\theta)+(p+1)E^*(\theta)] \frac{2}{(n_1\theta)^{2p+1}}.$$

which is finite iff $n_1\theta > 1$ which is already satisfied as $\theta < \dfrac{\sqrt{c+4}-\sqrt{c}}{2}$ where $c = \dfrac{\sqrt{5}+1}{2}$.

Thus under this assumption, (4.2) is satisfied and an absolutely continuous invariant probability measure exists. Once again, this is only a sufficient condition. But once an absolutely continuous invariant probability measure exists, $(*)$ and $(**)$ can be proved here also exactly as in the case $m = 2$.

Thus we have the following :

Theorem : Let $0 < \theta < \dfrac{\sqrt{c+4}-\sqrt{c}}{2}$ where $c = (\dfrac{\sqrt{5}+1}{2})$. Also let

$$\frac{1}{\theta} = [n_1\theta, n_2\theta, \cdots].$$

Assume that $n_2, n_3, \cdots$ are at least as large as $n_1 + 1$. Then the generalised Gauss map on $[0, \theta)$ admits an ergodic invariant probability which is equivalent to the Lebesgue measure. Moreover, for a.e. $x = [a_1\theta, a_2\theta, \cdots]$ we have,

$$\lim_{n\to\infty} \frac{a_1 + \cdots + a_n}{n} = \infty.$$

Further, there is a finite number γ, such that if $\frac{p_n}{q_n}$ denotes the n-th convergent of x, then for a.e x we have,

$$\lim_{n\to\infty} \frac{1}{n} \log q_n = \gamma.$$

Note that the above discussion is under the assumption $n_i \geq n_1 + 1$ for all i larger than one. Without this assumption, the problem becomes cumbersome and we shall not attempt to do that.

Section 5 : Back to Markov Processes

We now return to the Markov process mentioned at the beginning of the chapter. To recall, we have a sequence $(Z_n)_{n\geq 1}$ of i.i.d. random variables each taking values 0 and θ with probabilities α and $1 - \alpha$ respectively. Here $0 < \theta \leq 1$. X_0 is a strictly positive random variable independent of the sequence $(Z_n)_{n\geq 1}$. The Markov process $(X_n)_{n\geq 0}$ is given by,

$$X_{n+1} = Z_{n+1} + \frac{1}{X_n} \quad \text{for} \quad n \geq 0.$$

As explained in section 1, to show that the invariant distribution F for the process is singular when $\theta = 1$, the route followed in Bhattacharya and Goswami [2] is the following : They first obtained an explicit formula for F. They used conditions (*) and (**) (see section 1) to show that the derivative of F is zero a.s. Lebesgue Differentiation Theorem now implies that F must be singular.

For general θ, that is for $0 < \theta < 1$, it may not be possible to obtain any explicit formula for F on all of $\mathbb{R}$. Since F is known to be pure, in order to show that F is singular for certain value of θ, it suffices to show that F has a singular part. Thus if one could get some formula for F on a suitable subinterval of $\mathbb{R}$, then perhaps one can repeat their calculations.

With this in mind, we developed the machinery of earlier sections. However, we have not been able to get F on any subinterval and thus the envisaged program could not be carried out. We give below another proof of the result of Bhattacharya and Goswami [2] where (*) again plays a crucial role.

We are assuming $\theta = 1$. Thus we have $(Z_i)_{i\geq 0}$ i.i.d. random variables taking values 0 and 1 with probabilities α and $1-\alpha$. It is known that the invariant probability is nothing but the distribution of the random continued fraction $[Z_0; Z_1, Z_2, \cdots]$. Let this random variable be denoted by Y. Because of the presence of zeroes, $[Z_0(\omega); Z_1(\omega), Z_2(\omega), \cdots]$ is NOT the continued fraction expansion of the number $Y(\omega)$. But because each Z_i takes only two values 0 and 1, it is not difficult to discover the continued fraction expansion of $Y(\omega)$. This is what we obtain now.

Let us assume that $Z_0(\omega) = 1$ or equivalently, consider the set $\Omega_1 = \{\omega : Z_0(\omega) = 1\}$. Define the stopping times for the process $(Z_i)_{i\geq 1}$ as follows :

$\tau_0(\omega) =$ First odd integer i such that $Z_i(\omega) \neq 0$

$\tau_1(\omega) =$ First even integer $i > \tau_0$ such that $Z_i(\omega) \neq 0$

$\tau_2(\omega) =$ First odd integer $i > \tau_1$ such that $Z_i(\omega) \neq 0$

etc.

Let us now define,

$a_0(\omega) = \sum_{0\leq i<\tau_0(\omega)} Z_i(\omega)$

$a_1(\omega) = \sum_{\tau_0(\omega)\leq i<\tau_1(\omega)} Z_i(\omega)$

etc.

Then, a simple calculation shows that for a.e. $\omega \in \Omega_1$,

$$Y(\omega) = [a_0(\omega); a_1(\omega), a_2(\omega), \cdots].$$

Denote by S_i the partial sums of the Z_i sequence. More precisely, $S_0(\omega) = Z_0(\omega)$, and in general, $S_k(\omega) = \sum_{0\leq i\leq k} Z_i(\omega)$. Then it follows that,

$$a_0 = S_{\tau_0-1}; \, a_1 = S_{\tau_1-1} - S_{\tau_0-1}; \cdots$$

Thus

$$\text{(A)} \qquad \frac{1}{k+1}\sum_{i=0}^{k} a_i(\omega) = \frac{S_{\tau_k - 1}}{k+1} = \frac{S_{\tau_k-1}}{\tau_k - 1}.\frac{\tau_k - 1}{k+1}$$

By the strong law of large numbers, we conclude that, $\lim_{k\to\infty} \frac{S_{\tau_k-1}}{\tau_k - 1}$ converges to $1-\alpha$ for a.e. ω. It is easy to see that $\tau_0, \tau_1 - \tau_0, \tau_2 - \tau_1, \cdots$ are i.i.d. random variables and they take values $1, 3, 5, \cdots$ with probabilities $1-\alpha$; $\alpha(1-\alpha)$; $\alpha^2(1-\alpha), \cdots$. Thus they have the common expectation $\frac{3-\alpha}{1-\alpha}$. So, again by the strong law of large numbers $\lim_{k\to\infty} \frac{\tau_k - 1}{k+1} = \frac{3-\alpha}{1-\alpha}$. As a consequence, from (A), we conclude that

$$\lim_{k\to\infty} \frac{1}{k+1}\sum_{i=0}^{k} a_i(\omega) = 3-\alpha \qquad \text{for a.e. } \omega \in \Omega_1.$$

Thus for almost every $\omega \in \Omega_1$, the average of the digits in the (continued fraction) expansion of $Y(\omega)$ has a finite limit. Actually, we can carefully repeat this argument on each of the sets $\Omega_{01} = \{\omega : Z_0(\omega) = 0 \text{ and } Z_1(\omega) = 1\}; \Omega_{001} = \{\omega : Z_0(\omega) = 0, Z_1(\omega) = 0, Z_2(\omega) = 1\}$ etc. to conclude that for almost every $\omega \in \Omega$, the average of the digits in the expansion of $Y(\omega)$ converges to the finite number, $3-\alpha$. Here,

$$\Omega = \Omega_1 \cup \Omega_{01} \cup \Omega_{001} \cup \cdots.$$

In other words, in view of (*), the range of Y is a Lebesgue null set (we have a_0 terms also here. (*) was stated only for numbers between 0 and 1 and here we are using it for all numbers in $(0, \infty)$. But this makes no difference.). This shows that Y has singular distribution as desired.

In fact this proof gives more information. If the distribution of $Y = [Z_0; Z_1, Z_2, \cdots]$ (where Z_is are i.i.d. taking values 0 and 1 with probabilities α and $1-\alpha$) is denoted by P_α on $(\mathbb{R}, \mathcal{B})$ then for $0 < \alpha < 1$, the probabilities P_α are all singular w.r.t. Lebesgue measure and moreover, this family (P_α) is a uniformly singular family.

The difficulty in generalizing to the case when $\theta < 1$ is the following: we can collapse the zeros as we did earlier and end up with $Y(\omega) = [a_0(\omega)\theta; a_1(\omega)\theta, a_2(\omega)\theta, \cdots]$. Unfortunately, this may still be NOT the expansion of $Y(\omega)$ in view of the validity problems mentioned in section 2. As a result, we have not been able to extend this argument any further.

References

1. Berger, Marc A. [1992]: Random affine iterated function systems : mixing and encoding. *Diffusion processes and related problems in Analysis II*, p.315–346, Editors: M. Pinsky et al.,Birkhausser.

2. Bhattacharya, R.N. and Goswami, A. [1999] : A Markovian algorithm for random continued fractions and a computation of their singular equillibria. *Proc. Third Calcutta Triennial Symp.* (To appear)

3. Bhattacharya, R.N. and Lee, O. [1988] : Asymptotics of a class of Markov process which are not in general irreducible. *Ann. Prob.*, **16**, p.1333-1347. Correction note, ibid [1997] : **25**, p.1541-1543.

4. Bhattacharya, R.N. and Majumdar, M. [1999] : On a Theorem of Dubins and Freedman. Preprint.

5. Billingsley, P. [1965] : *Ergodic Theory and Information.* Wiley, New York.

6. Bissinger, B.H. [1944] : Generalization of continued fractions. *Bull. Amer. Math. Soc.*, **50**, p.868-876.

7. Chakraborty, S. and Rao, B.V. [1998]: Completeness of the Bhattacharya metric on the space of probabilities. *Stat. and Prob. Letters*, **36**, p.321-326.

8. Chakraborty, S. and Rao, B.V. [1998]: Convolution powers of probabilities on stochastic matrices of order 3. *Sankhya A*, **60**, p.151-170.

9. Chakraborty, S. and Rao, B.V. [2000]: Convolution powers of probabilities on stochastic matrices. To appear in *J. Th. Prob.*

10. De Melo, W. and Strein, S. van [1993] : *One Dimensional Dynamics.* Springer-Verlag, New York.

11. Dhar, S. and Mukherjea, A. [1997]: Convergence in Distribution of products of I.I.D Nonnegative Matrices. *J. Th. Prob.*, **10**, p.375–393.

12. Dubins, L.E. and Freedman, D.A. [1966] : Invariant probabilities for certain Markov processes. *Ann. Math. Stat.* **37**, p.837-847.

13. Erdos, P. [1939] : On a family of Symmetric Bernoulli convolitions. *Amer. J. Math.*, **61**, p.974-976.

14. Erdos, P. [1940] : On the smoothness properties of a Bernoulli convolutions. *Amer. J. Math.*, **62**, p.180-186.

15. Everett, C.J. [1946] : Representations for real numbers. *Bull.Amer. Math. Soc.*, **52**, p.861-869.

16. Garcia, A.M. [1962] : Arithmetic properties of Bernoulli convolutions. *Trans. Amer. Math. Soc.*, **102**, p.409-432.

17. Hognas, G. and Mukherjea, A. [1995]: *Probability measures on semigroups.* Plenum Press, N.Y.

18. Kahane, J.P. and Salem, R. [1958] : Sur la Convolution d'une infinite de distributions de Bernoulli. *Colloq. Math.*, **VI**, p.193-202.

19. Khinchin, A.Ya [1964]: *Continued fractions.* English translation of third Russian edition, Univ.Chicago Press, Chicago.

20. Kreshner, R. and Wintner, A. [1935] : On symmetric Bernoulli convolutions. *Amer. J. Math.*, **57**, p.541-548.

21. Letac, G. and Seshadri, V. [1995]: A random continued fraction in $I\!R^{d+1}$ with an inverse Gaussian distribution. *Bernoulli*, **1**, p. 381-394.

22. Mukherjea, A. [1979] : Limit Theorems : Stochastic matrices, ergodic markov chains and measures on semigroups. *Probability Analysis and Related Topics.* **2**, p.143-203, Ed. A.T. Bharucha Reid, Academic Press, N.Y.

23. Mukherjea, A. [1991]: Tightness of products of i.i.d. random matrices. *Probability Theory and Related Fields*, **87**, p.389-401.

24. Mukherjea, A. and Ratti, J.S. [1997] : Continuous Singularity of the Weak Limit of Convolution Powers of a Discrete Probability Measure on 2×2 Stochastic Matrices. *J. Th. Prob.*, **10**, p.499-506.

25. Mukherjea, A. and Tserpes, N.A. [1976] : *Measures on Topological Semigroups.* LNM, **547**, Springer-Verlag.

26. Pisot, C. [1938] : La repartition modulo un et les nombres algebriques. *Ann. Scuola Norm. Sup. Pisa*, **2**, p.205-248.

27. Pisot, C. and Dufresnoy, J. [1953] : Sur un ensemble ferme de nombres algebriques. *Ann. Sci. Ecole Norm. Sup.*, **98**, p.105-133.

28. Renyi, A [1957] : Representations for real numbers and their Ergodic properties. *Acta Math. Acad. Sei. Hung*, 8, p.477-495.

29. Rosenblatt, M. [1971]: *Markov processes : Structure and Asymptotic Behaviour.* Springer-Verlag, Berlin and New York.

30. Salem, R. [1941] : Sets of uniqueness and sets of multiplicity. *Trans. Amer. Math. Soc.*, 54, p.218-228.

31. Salem, R. [1944] : A remarkable class of algebraic integers, proof of a conjecture of Vijayaraghavan. *Duke Math. J.*, **11**, p.103-108.

32. Siegel, C.L. [1944] : Algebraic integers whose conjugates lie in the unit circle. *Duke Math. J.*, 11, p.597-602.

33. Solomyak, Boris [1995]: On the random series $\sum \pm\lambda^n$ (an Erdos problem). *Ann. Math.*, 142, p.611-625.

34. Vijayaraghavan, T. [1941] : On the fractional parts of the powers of a number. *Proc. Camb. Philos. Soc.*, **37**, p.349-357.

35. Wintner, A. [1935] : On Convergent Poisson Convolutions. *Amer. J. Math*, **57**, p.827-838.

36. Yahav, J.A. [1975] : On a fixed point theorem and its stochastic equivalent. *J. Appl. Probab.*, 12, p.605-611.