SHIFT INVARIANT MARKOV MEASURES

by

David Bryan Colwell

B.Sc. (Honours), Simon Fraser University, 1982

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in the Department

of

Mathematics and Statistics

SIMON FRASER UNIVERSITY

March, 1986

APPROVAL

Name: David Bryan Colwell

Degree: Master of Science

Title of Thesis: Shift Invariant Markov Measures

Chairman: Dr. A.R. Freedman

Dr. C. Kim
Senior Supervisor

Dr. R. Lockhart

Dr. J.J. Sember

Dr. G. Bojadziev
External Examiner
Professor
Department of Mathematics and Statistics
Simon Fraser University

Date approved: April 10, 1986

Title of Thesis/Project/Extended Essay

Shift Invariant Markov Measures

Author:

(signature)

David B. Colwell

(name)

10 April, 1986

(date)

Abstract

The purpose of this paper is to discuss various topics regarding probability measures on the space Ω of sequences taking the values $0,1,\ldots,r-1$ (where r is assumed fixed), together with its Borel subsets, $\mathcal{B}(\Omega)$. We emphasize shift invariant Markov measures, as well as the space of all shift invariant Markov measures, denoted $M(\Omega,T)$, where T is the (one-sided) shift on Ω. Throughout, we take an approach to the matrix representation of Markov measures that is a slight improvement over previous approaches. Also, we give direct proofs, for this setting, to some known results.

Our main results are the characterization of atomic ergodic measures as periodic orbit measures, and the characterization of atomic ergodic Markov measures as Markov measures induced by cyclic permutation matrices. Other results appear in the discussion of entropy, where we give necessary and sufficient conditions for a Markov measure to have entropy equal to zero or $\log r$. We also prove that the entropy map is continuous on $M(\Omega,T)$. Elsewhere, we prove that $M(\Omega,T)$ is not convex. In addition, as a result of our discussion of nonatomic measures on $(\Omega,\mathcal{B}(\Omega))$, we give a simple method for constructing examples of (strictly) increasing continuous functions on the unit interval.

ACKNOWLEDGEMENTS

I would like to thank Dr. C.W. Kim for his generous assistance and encouragement during my graduate studies in general, and during the writing of this thesis in particular. I would also like to thank the Department of Mathematics and Statistics, Simon Fraser University, for giving me either a job or financial assistance during the length of my graduate studies. Finally, many thanks to Cindy Lister for the excellent job of typing this thesis.

TABLE OF CONTENTS

Introduction

The purpose of this paper is to discuss various topics regarding the sequence space $\Omega = \prod_{0}^{\infty} \{0,1,\ldots,r-1\}$ (where r is considered fixed) endowed with the product topology, and the space $P(\Omega)$, of all probability measures on $(\Omega,\mathcal{B}(\Omega))$, where $\mathcal{B}(\Omega)$ is the σ-algebra of Borel subsets of Ω. We shall present some new results for shift invariant Markov measures and ergodic Markov measures, as well as for the space of all shift invariant Markov measures, denoted $M(\Omega,T)$, where T is the shift on Ω.

We begin, in sections 1 to 3, with a description of the setting that we will be working with throughout. In section 1, we discuss the sequence space Ω, a topological measurable space, and the shift T (which some authors call the one-sided shift). In section 2, we give a direct proof to the Kolmogorov existence theorem. In section 3 we begin discussing the space $P(\Omega)$ as well as the subspace, $P(\Omega,T)$, of all shift invariant probability measures on $(\Omega,\mathcal{B}(\Omega))$. We then look at ergodicity, presenting some known results for ergodic measures, as well as for the space of all ergodic measures, denoted $E(\Omega,T)$.

In section 4 we look at Markov measures. In our presentation of the existence theorem, (4.1), S denotes the index set for the probability distribution p and the stochastic matrix P that induce the measure μ, and E denotes the state space for the Markov chain on $(\Omega,\mathcal{B}(\Omega),\mu)$. Here we avoid assuming that $E = S$, that is, we allow for the possibility that $E \subsetneq S$. Because of this approach, we have, in (4.13), a slight improvement over the usual characterization of ergodic Markov measures. In another theorem, (4.11), we give a simple proof to a theorem of Doob [2]. We close this section by proving the nonconvexity of three subspaces of $P(\Omega)$: the subspace of Markov measures, $M(\Omega)$; the subspace of shift invariant Markov measures, $M(\Omega,T)$; and the subspace of Bernoulli measures, $B(\Omega,T)$.

Our major results appear in sections 6 and 7, where we consider atomic measures on $(\Omega,\mathcal{B}(\Omega))$. Section 5 begins the discussion with a look at periodic points, orbits, and orbit measures. The main result in section 6 is the characterization of atomic ergodic measures as periodic orbit measures. In section 7, we characterize atomic ergodic Markov measures as Markov measures induced by cyclic permutation matrices. In addition, (7.5) characterizes non-Markovian atomic ergodic measures, that is, non-Markovian periodic orbit measures.

In sections 8 to 10, we discuss the nonatomic case, where it is natural to consider the unit interval I, with its Borel subsets, $\mathcal{B}(I)$. In section 8 we look at distribution functions, f_ν, that correspond to measures ν on $(I,\mathcal{B}(I))$. Section 9 then presents some known relationships between nonatomic measures on $(\Omega,\mathcal{B}(\Omega))$ and nonatomic measures on $(I,\mathcal{B}(I))$, using the natural mapping φ from Ω onto I, where $\varphi(\omega) = \sum_{n=0}^{\infty} \frac{\omega_n}{r^{n+1}}$ for any $\omega = (\omega_n)_{n=0}^{\infty}$ in Ω. This section closes with (9.19) which furnishes a direct proof of the isomorphism of the probability spaces $(\Omega,\mathcal{B}(\Omega),\mu)$ and $(I,\mathcal{B}(I),m)$, where $\mu \in P(\Omega)$ is nonatomic, and m is Lebesgue measure. In section 10, a main result is the calculation of distribution functions. For any shift invariant Markov measure μ, (10.5) gives a simple method for calculating the distribution function f of the random variable φ on (Ω,μ). Theorem (10.6) gives the corresponding result for Bernoulli measures. It is known that if a probability measure μ is ergodic, and not equal to $(\frac{1}{r}, \ldots, \frac{1}{r})$-Bernoulli measure, then the distribution function f is singular. Thus the above method is a mechanism for constructing examples of (strictly) increasing continuous singular functions on the unit interval; for example, Lebesgue's singular function is generated by $(\frac{1}{2}, 0, \frac{1}{2})$-Bernoulli measure.

In section 11, we discuss entropy on $P(\Omega,T)$, beginning with basic definitions and some known results. One of our main results in this section is (11.10): for any shift invariant Markov measure μ, the entropy of T relative to μ, $h_\mu(T)$, equals zero iff μ is induced by a permutation matrix, i.e., μ is an ergodic atomic (Markov) measure. We also show, in (11.13), that for each measure $\mu \in M(\Omega,T)$, $h_\mu(T) = \log r$ iff μ is the $(\frac{1}{r}, \ldots, \frac{1}{r})$-Bernoulli measure. We close this section by proving that the entropy map, $\mu \mapsto h_\mu(T)$ is continuous on $M(\Omega,T)$.

1. The sequence space Ω

(1.1) Definitions. Throughout, r will denote a fixed but arbitrary positive integer such that $r \geq 2$. Let $S = \{0,1,\ldots,r-1\}$ be a finite set of r points with the discrete topology. Define the sequence space Ω by $\Omega = \prod_{n=0}^{\infty} S_n$, where $S_n = S$, that is

$$\Omega = \{(\omega_n)_{n\geq 0} : \omega_n \in S\} .$$

Endow the set Ω with the product topology.

For each $n \geq 0$, let $x_n : \Omega \to S$ be defined by $x_n(\omega) = \omega_n$. A subset of Ω of the form

$$\{\omega \in \Omega : x_0(\omega) = i_0,\ldots,x_n(\omega) = i_n\}, \quad \text{where } i_0,\ldots,i_n \in S, \quad n \geq 0,$$

denoted by $Z(i_0,\ldots,i_n)$, is called a cylinder set. Let $\mathcal{S}$ denote the collection of all cylinder sets, together with the empty set ϕ.

Recall that a topological space is zero-dimensional iff every point in the space has a neighbourhood base consisting of clopen sets, ie. sets that are both open and closed.

(1.2) Theorem. The sequence space Ω with the product topology is a compact, zero-dimensional Hausdorff space with countable base $\mathcal{S}$.

Proof. Clearly Ω is a compact Hausdorff space in the product topology. It is easily seen that the collection S is a countable base for the product topology of Ω and that each cylinder set is clopen. ◫

(1.3) Theorem. The sequence space Ω with the product topology is metrizable.

Proof. Define $d : \Omega \times \Omega \to [0,1]$ by

$$d(\omega,\eta) = \sum_{n=0}^{\infty} \frac{1}{2^{n+1}} \cdot \frac{|\omega_n - \eta_n|}{1 + |\omega_n - \eta_n|}$$

Then we see readily that d is a metric on Ω. Let

$$U(\omega,\varepsilon,N) = \{\eta\in\Omega : |\eta_n - \omega_n| < \varepsilon, \quad 0\le n\le N\} \quad \text{where} \quad \omega\in\Omega, \ \varepsilon>0, \ N\ge 0.$$

Recall that, on the real line $\mathbb{R}$, the two metrics $\rho(x,y) = |x-y|$ and $\rho'(x,y) = \frac{|x-y|}{1 + |x-y|}$ are uniformly equivalent. Hence, given $\varepsilon > 0$, there exists $\delta > 0$ such that $\rho(x,y) < \varepsilon$ whenever $\rho'(x,y) < \delta$. Suppose $d(\eta,\omega) < 2^{-(N+1)}\delta$. Then we have, for $0 \le n \le N$,

$$2^{-(n+1)}\rho'(\eta_n,\omega_n) < 2^{-(N+1)}\delta \quad \text{or}$$

$$\rho'(\eta_n,\omega_n) < 2^{-(N-n)}\delta \le \delta,$$

so that $\rho(\eta_n,\omega_n) < \varepsilon$. Thus the d-topology is stronger than the product topology.

Now consider the set $\{\eta \in \Omega : d(\eta,\omega) < \varepsilon\}$, where $\omega \in \Omega$ is fixed. Choose an $N > 2$ such that $\Sigma_{n\geq N} 2^{-n} < \varepsilon/2$. Choose $\delta > 0$ such that $\rho'(x,y) < \varepsilon/2$ whenever $\rho(x,y) < \delta$, for any $x,y \in \mathbb{R}$. Then we obtain, for each $\eta \in U(\omega,\delta,N-1)$,

$$d(\eta,\omega) \leq \sum_{n=0}^{N-1} 2^{-(n+1)} \rho'(\eta_n,\omega_n) + \Sigma_{n\geq N} 2^{-n}$$

$$< \varepsilon/2 + \varepsilon/2 = \varepsilon ,$$

so that the d-topology coincides with the product topology. ◫

A collection C of subsets of Ω is called a <u>semialgebra</u> if the following three conditions hold: (i) $\phi \in C$; (ii) if $A,B \in C$, then $A \cap B \in C$; (iii) if $A \in C$, then $\Omega - A$ is a finite union of pairwise disjoint sets in C.

Let $S^{n+1} = \prod_{i=0}^{n} S_i$, where $S_i = S$, $n \geq 0$.

(1.4) <u>Lemma</u>. The collection S is a semialgebra.

<u>Proof</u>. By definition, $\phi \in S$. For any $Z(i_0,\ldots,i_n)$ and $Z(j_0,\ldots,j_n)$ we have

$$Z(i_0,\ldots,i_n) \cap Z(j_0,\ldots,j_n) = Z(i_0,\ldots,i_n) \quad \text{or} \quad \phi$$

according as $(i_0,\ldots,i_n) = (j_0,\ldots,j_n)$ or $(i_0,\ldots,i_n) \neq (j_0,\ldots,j_n)$. Given $Z(i_0,\ldots,i_m)$ and $Z(j_0,\ldots,j_n)$ where $m < n$, we have

$$Z(i_0,\ldots,i_m) = \bigcup_{i_{m+1}} \cdots \bigcup_{i_n} Z(i_0,\ldots,i_m,i_{m+1},\ldots,i_n)$$

so that

$$Z(i_0,\ldots,i_m) \cap Z(j_0,\ldots,j_n) = Z(j_0,\ldots,j_n) \quad \text{or} \quad \phi$$

according as $(i_0,\ldots,i_m) = (j_0,\ldots,j_m)$ or $(i_0,\ldots,i_m) \neq (j_0,\ldots,j_m)$.

For any cylinder $Z(i_0,\ldots,i_n)$ we have

$$\Omega - Z(i_0,\ldots,i_n) = \cup\{Z(j_0,\ldots,j_n) : (j_0,\ldots,j_n) \in S^{n+1},$$
$$(j_0,\ldots,j_n) \neq (i_0,\ldots,i_n)\}$$

which is a union of $r^{n+1}-1$ disjoint cylinders in $\mathcal{S}$. We also have $\Omega = \bigcup_i Z(i)$. □

The algebra generated by the semialgebra $\mathcal{S}$ is denoted by $A(\mathcal{S})$.

(1.5) Lemma. The algebra $A(\mathcal{S})$ consists of sets of the form

$$\{\omega \in \Omega : (\omega_0,\ldots,\omega_n) \in E\} = \bigcup_{(i_0,\ldots,i_n) \in E} Z(i_0,\ldots,i_n)$$

where $E \subset S^{n+1}$, and $n \geq 0$.

Proof. It is plain that $A(\mathcal{S})$ consists of all finite disjoint unions of sets in the semialgebra $\mathcal{S}$. Thus $A \in A(\mathcal{S})$ iff $A = \{\omega: (\omega_0,\ldots,\omega_n) \in E\}$ where $E \subset S^{n+1}$, $n \geq 0$. □

Let $\mathcal{B}(\Omega)$ denote the σ-algebra generated by all open subsets of Ω and let $\sigma(\mathcal{S})$ denote the σ-algebra generated by the semialgebra $\mathcal{S}$. Then we obtain readily that $\mathcal{B}(\Omega) = \sigma(\mathcal{S})$ (since $\mathcal{S}$ is a countable base for Ω). Let $C(\Omega)$ denote the set of all real-valued continuous functions on Ω. Also, for each $E \subset \Omega$, we denote the indicator (characteristic) function of E by 1_E.

(1.6) Lemma. Let $\mathcal{F}$ be the family of all linear combinations of indicator functions of sets in the semialgebra $\mathcal{S}$. Then $\mathcal{F}$ is dense in $C(\Omega)$ in the uniform topology.

Proof. We have $\mathcal{F} = \{\sum_{i=1}^{n} c_i 1_{E_i} : E_i \in \mathcal{S},\ c_i \in \mathbb{R},\ n \geq 1\}$. Since each cylinder set E is a clopen subset of Ω, 1_E is in $C(\Omega)$, so that $\mathcal{F}$ is a subspace of the vector space $C(\Omega)$. It is easily seen that each $f \in \mathcal{F}$ can be written as $f = \sum_{i=1}^{n} c_i 1_{E_i}$ where $c_i \in \mathbb{R}$ and $E_1,\ldots,E_n$ are pairwise disjoint sets in $\mathcal{S}$, so that $|f| = \sum_{i=1}^{n} |c_i| 1_{E_i} \in \mathcal{F}$. Thus $\mathcal{F}$ is also a sublattice of the vector lattice $C(\Omega)$. Since $\mathcal{F}$ is a separating family of functions on Ω and contains the function 1, $\mathcal{F}$ is dense in $C(\Omega)$ by the Stone-Weierstrass theorem (see Hewitt-Stromberg [1]). □

(1.7) Definition. The transformation T defined on the sequence space Ω by $(T\omega)_n = \omega_{n+1}$, where $\omega \in \Omega$ and $n \geq 0$, is called the shift.

(1.8) Lemma. The shift T is a continuous surjection from Ω onto itself. In particular, T is a measurable transformation from the measurable space $(\Omega,\mathcal{B}(\Omega))$ onto itself.

Proof. By definition, $T\omega \in \Omega$ for each $\omega \in \Omega$. Given $\omega = (\omega_n)_{n\geq 0} \in \Omega$, we obtain $T^{-1}\{\omega\} = \{(i,\omega_0,\omega_1,\ldots) : i \in S\}$ so that T is a surjection but not an injection. For each cylinder set $Z(i_0,\ldots,i_n)$, we obtain

$$T^{-1}Z(i_0,\ldots,i_n) = T^{-1}\{\omega : x_0(\omega) = i_0,\ldots,x_n(\omega) = i_n\}$$

$$= \{\omega : x_1(\omega) = i_0,\ldots,x_{n+1}(\omega) = i_n\}$$

$$= \bigcup_i Z(i,i_0,\ldots,i_n)$$

so that T is continuous. □

(1.9) Remark. Since for any $n \geq 0$ the mapping $x_n : \Omega \to S$ is continuous, it is also a measurable function from the measurable space $(\Omega,\mathcal{B}(\Omega))$ onto the measurable space $(S,\mathcal{B}(S))$, where $\mathcal{B}(S)$ denotes the power set of S.

We state without proof the following unique extension theorem.

(1.10) Theorem. Let μ and ν be probability measures on $(\Omega,\mathcal{B}(\Omega))$. If $\mu(E) = \nu(E)$ for each E in the semialgebra $\mathcal{S}$, then $\mu = \nu$ on $\mathcal{B}(\Omega)$.

See Blumenthal and Getoor [1], Halmos [1], Royden [1] for the proof.

2. The Kolmogorov extension theorem

Let μ be a probability measure on the measurable space $(\Omega,\mathcal{B}(\Omega))$. Consider the sequence of random variables, or the stochastic process $\{x_n\}_{n\geq 0}$ on the probability space $(\Omega,\mathcal{B}(\Omega),\mu)$. For each $n \geq 0$, define the function $p_n : S^{n+1} \to [0,1]$ by

$$p_n(i_0,i_1,\ldots,i_n) = \mu(x_0=i_0,\ x_1=i_1,\ \ldots,\ x_n=i_n),$$

where $\{x_0=i_0,\ \ldots,\ x_n=i_n\} = \{\omega\in\Omega : x_0(\omega)=i_0,\ \ldots,\ x_n(\omega)=i_n\}$. Then the following consistency conditions hold:

$$\begin{cases} 0 \leq p_n(i_0,\ldots,i_n) \leq 1, \\ \sum_{i\in S} p_0(i) = 1, \\ p_n(i_0,\ldots,i_n) = \sum_{i\in S} p_{n+1}(i_0,\ldots,i_n,i). \end{cases}$$

Using a probability vector and a stochastic matrix, we may generate functions p_n, $n\geq 0$ satisfying the consistency conditions.

(2.1) Example. Let $p = (p_i)_{i\in S}$ be any probability vector on the set S. For each $n \geq 0$, define

$$p_n(i_0,i_1,\ldots,i_n) = p_{i_0}\, p_{i_1} \cdots p_{i_n}, \quad \text{where } i_0,i_1,\ldots,i_n \in S.$$

Then $\{p_n\}_{n\geq 0}$ satisfy the consistency conditions.

(2.2) Example. Let $p = (p_i)_{i \in S}$ be any probability vector on S and let $P = (p_{ij})_{i,j \in S}$ be any $r \times r$ stochastic matrix, i.e., $0 \le p_{ij} \le 1$ and $\sum_{k \in S} p_{ik} = 1$ for each $i,j \in S$. For each $n \ge 0$, the function defined by

$$p_n(i_0,i_1,\ldots,i_n) = p_{i_0} \, p_{i_0 i_1} \cdots p_{i_{n-1} i_n} ,$$

where $i_0,i_1,\ldots,i_n \in S$,

satisfies the consistency conditions.

There are functions $\{p_n\}_{n \ge 0}$ satisfying the consistency conditions that are not induced by any pair of a probability vector and a stochastic matrix.

(2.3) Example. Let $p = (p_i)$ and $q = (q_i)$ be two probability vectors on the set $S = \{0,1,\ldots,r-1\}$ such that $p_i = \frac{1}{r}$ for all i, and $q_0 = \frac{1}{2r}$, $q_1 = \frac{3}{2r}$, $q_i = \frac{1}{r}$ for $i \ge 2$. For each $n \ge 0$, define the function $p_n(i_0,\ldots,i_n)$ by

$$p_n(i_0,\ldots,i_n) = \tfrac{1}{2}(p_{i_0} \cdots p_{i_n} + q_{i_0} \cdots q_{i_n})$$

where $i_0,\ldots,i_n \in S$.

It is easily seen that the functions $\{p_n\}_{n \ge 0}$ satisfy the consistency conditions. In particular we obtain

$$p_0(0) = \frac{3}{4r}, \quad p_1(0,0) = \frac{5}{8r^2}, \quad p_2(0,0,0) = \frac{9}{16r^3}.$$

Suppose that there exist a probability vector $u = (u_i)$ on S and an $r \times r$ stochastic matrix $A = (a_{ij})$ such that

$$p_n(i_0,\ldots,i_n) = u_{i_0} a_{i_0 i_1} \ldots a_{i_{n-1} i_n}$$

for all $i_0,\ldots,i_n \in S$ and all $n \geq 0$. Then we obtain

$$u_0 = p_0(0) = \frac{3}{4r}, \quad a_{00} = p_1(0,0)/u_0 = \frac{5}{6r} \quad \text{and}$$

$$\frac{9}{16r^3} = p_2(0,0,0) = u_0 a_{00} a_{00} = \frac{3}{4r} \cdot \frac{5}{6r} \cdot \frac{5}{6r} = \frac{25}{48r^3}$$

so that $\frac{9}{16} = \frac{25}{48}$, a contradiction. ◫

(2.4) <u>The Kolmogorov extension theorem.</u> Let $S = \{0,1,..,r-1\}$, for some $r \geq 2$. Let $\{p_n\}_{n \geq 0}$ be a sequence of functions satisfying the consistency conditions, where p_n has domain S^{n+1}. Then there exists a unique probability measure μ on $(\Omega, \mathcal{B}(\Omega))$ such that

$$\mu(x_0 = i_0, \ldots, x_n = i_n) = p_n(i_0,\ldots,i_n) \quad \text{for all}$$

$$i_0,\ldots i_n \in S \text{ and all } n \geq 0.$$

<u>Proof.</u> Define the set function μ on the semialgebra $\mathcal{S}$ by

$$\mu(Z(i_0,\dots,i_n)) = p_n(i_0,\dots,i_n), \quad \mu(\phi) = 0.$$

We want to show the following:

(i) If a set A in S is a finite union of the disjoint sets $A_1,\dots,A_n$ in S, then $\mu A = \sum_{i=1}^{n} \mu A_i$.

(ii) If a set A in S is a countable union of the disjoint sets $A_1,A_2,\dots$ in S, then $\mu A \le \sum_{i\ge 1} \mu A_i$.

For any cylinder set $Z(i_0,\dots,i_n)$ and any $k \ge 1$, we have

$$Z(i_0,\dots,i_n) = \bigcup_{i_{n+1}\in S} \cdots \bigcup_{i_{n+k}\in S} Z(i_0,\dots,i_n,i_{n+1},\dots,i_{n+k}),$$

and

$$\begin{aligned}\mu(Z(i_0,\dots,i_n)) &= p_n(i_0,\dots,i_n) \\ &= \sum_{i_{n+1}} \cdots \sum_{i_{n+k}} p_{n+k}(i_0,\dots,i_n,i_{n+1},\dots,i_{n+k}) \\ &= \sum_{i_{n+1}} \cdots \sum_{i_{n+k}} \mu(Z(i_0,\dots,i_n,i_{n+1},\dots,i_{n+k})).\end{aligned}$$

Let $A = Z(i_0,\dots,i_m)$, be an arbitrary cylinder set. Suppose $A = \bigcup_{t=1}^{u} A_t$ where $A_1,\dots,A_u$ are pairwise disjoint cylinder sets. For each A_t, there exist $j_{t,1},\dots,j_{t,n_t}$ in S, $n_t \ge 1$, such that

$$A_t = Z(i_0,\ldots,i_m,j_{t,1},\ldots,j_{t,n_t}).$$

Let $n = \max\{n_t : 1 \le t \le u\}$. Then, as above, we obtain

$$A_t = \bigcup_{k_{n_t+1}} \cdots \bigcup_{k_n} Z(i_0,\ldots,i_m,j_{t,1},\ldots,j_{t,n_t},k_{n_t+1},\ldots,k_n)$$

so that

$$A = \bigcup_{t=1}^{u} A_t = \bigcup_t \left[\bigcup_{k_{n_t+1}} \cdots \bigcup_{k_n} Z(i_0,\ldots,i_m,j_{t,1},\ldots,j_{t,n_t},k_{n_t+1},\ldots,k_n)\right]$$

$$= \bigcup_{i_1'} \cdots \bigcup_{i_n'} Z(i_0,\ldots,i_m,i_1',\ldots,i_n').$$

Using the preceding results, we obtain

$$\sum_{t=1}^{u} \mu(A_t) = \sum_t \left(\sum_{k_{n_t+1}} \cdots \sum_{k_n} \mu(Z(i_0,\ldots,i_m,j_{t,1},\ldots,j_{t,n_t},k_{n_t+1},\ldots,k_n))\right)$$

$$= \sum_{i_1'} \cdots \sum_{i_n'} \mu(Z(i_0,\ldots,i_m,i_1',\ldots,i_n'))$$

$$= \mu(A)$$

Thus (i) holds.

To prove the σ-additivity of μ on S, suppose that a cylinder set A is the union of a countable collection $\mathcal{U} = \{A_n\}_{n\ge1}$ of disjoint sets in $\mathcal{S}$. From (1.2) we see that $\mathcal{U}$ is an open covering of the compact set A, so that there is a finite sub-collection of cylinder sets $\mathcal{U}' = \{A_{n_1}, \ldots, A_{n_k}\} \subset \mathcal{U}$ such that $A = \bigcup_{i=1}^{k} A_{n_i}$. Note that, since the sets in $\mathcal{U}$ are disjoint, we have $A_n = \phi$ whenever $A_n \in \mathcal{U} - \mathcal{U}'$. Thus (ii) holds. Therefore, by a well-known theorem (see Neveu [1], Royden [1]), μ has a unique extension to a probability measure on the σ-algebra $\sigma(\mathcal{S}) = \mathcal{B}(\Omega)$. □

Let μ be a probability measure on $(\Omega, \mathcal{B}(\Omega))$. By (1.8) T is measurable, and so we define the probability measure $T\mu$ by the formula

$$(T\mu)(E) = \mu(T^{-1}E) \quad \text{where} \quad E \in \mathcal{B}(\Omega).$$

A probability measure μ with $T\mu = \mu$ is called <u>T-invariant.</u>

Let μ be a T-invariant probability measure on $(\Omega, \mathcal{B}(\Omega))$, and let $p_n(i_0, i_1, \ldots, i_n) = \mu(x_0 = i_0,\ x_1 = i_1,\ \ldots,\ x_n = i_n)$ where $n \geq 0$ and $i_0, i_1, \ldots, i_n \in S$. Then the functions $\{p_n\}$ satisfy the consistency conditions, together with the shift invariance conditions:

$$\begin{aligned} p_n(i_0, \ldots, i_n) &= (T\mu)(x_0 = i_0,\ \ldots,\ x_n = i_n) \\ &= \mu(x_1 = i_0,\ \ldots,\ x_{n+1} = i_n) \\ &= \Sigma_{i \in S}\ \mu(x_0 = i,\ x_1 = i_0,\ \ldots,\ x_{n+1} = i_n) \\ &= \Sigma_{i \in S}\ p_{n+1}(i, i_0, \ldots, i_n). \end{aligned}$$

(2.5) Corollary. If functions $\{p_n\}_{n\geq 0}$ satisfy the consistency conditions, together with the shift invariance condition, then there exists a unique T-invariant measure μ on $(\Omega, \mathcal{B}(\Omega))$ such that

$$\mu(x_0=i_0, \ldots, x_n=i_n) = p_n(i_0,\ldots,i_n) \quad \text{for all}$$

$$i_0,\ldots,i_n \in S \quad \text{and all} \quad n \geq 0.$$

Proof. By the Kolmogorov extension theorem, there is a unique probability measure μ such that $\mu(x_0=i_0, \ldots, x_n=i_n) = p_n(i_0,\ldots,i_n)$ for all $n \geq 0$ and all $i_0,\ldots,i_n \in S$. Using the shift invariance condition, we obtain

$$\begin{aligned}
\mu(x_0=i_0, \ldots, x_n=i_n) = p_n(i_0,\ldots,i_n) &= \Sigma_{i\in S}\, p_{n+1}(i,i_0,\ldots,i_n) \\
&= \Sigma_{i\in S}\, \mu(x_0=i, x_1=i_0, \ldots, x_{n+1}=i_n) \\
&= \mu(x_1=i_0, \ldots, x_{n+1}=i_n) \\
&= \mu(T^{-1}(x_0=i_0, \ldots, x_n=i_n)) \\
&= (T\mu)(x_0=i_0, \ldots, x_n=i_n)
\end{aligned}$$

so that $\mu(E) = (T\mu)(E)$ for all $E \in \mathcal{S}$. By the unique extension theorem (1.10) we have $\mu = T\mu$. □

(2.6) Remark. The functions defined in Examples (2.1) and (2.3) satisfy the shift invariance conditions.

(2.7) <u>Example</u>. Let $P = (p_{ij})_{i,j\in S}$ be an irreducible stochastic matrix and let $p = (p_i)_{i\in S}$ be the stationary distribution for P, that is, $pP = p$. Let $p_n(i_0,i_1,\ldots,i_n) = p_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} i_n}$ for $n \geq 0$ and $i_0,\ldots,i_n \in S$. Then the functions $\{p_n\}_{n\geq 0}$ satisfy both the consistency conditions and the shift invariance conditions.

3. Probability measures and ergodic measures

(3.1) Definition. Let $P(\Omega)$ denote the set of all probability measures on $(\Omega, \mathcal{B}(\Omega))$. For each $\omega \in \Omega$, define the probability measure ε_ω by $\varepsilon_\omega(A) = 1_A(\omega)$ for $A \subset \Omega$. Let $E(\Omega) = \{\varepsilon_\omega : \omega \in \Omega\}$.

It is well-known that $P(\Omega)$ is a convex set and that each μ in $P(\Omega)$ is regular, that is, for each $A \in \mathcal{B}(\Omega)$,

$$\begin{aligned}\mu(A) &= \sup\{\mu(C) : C \text{ compact}, C \subset A\} \\ &= \inf\{\mu(G) : G \text{ open}, A \subset G\}.\end{aligned}$$

By the Riesz representation theorem, $P(\Omega)$ can be canonically mapped bijectively onto the set of all positive linear forms J on $C(\Omega)$ such that $J(1) = 1$. We shall assume that $P(\Omega)$ is endowed with the weak* topology. A base for the weak* topology on $P(\Omega)$ is given by the collection of all sets of the form

$$U(\mu; f_1, \ldots, f_k; \varepsilon) = \{\nu \in P(\Omega) : |\int f_i d\nu - \int f_i d\mu| < \varepsilon, 1 \le i \le k\},$$

where $\mu \in P(\Omega)$, $f_i \in C(\Omega)$, $k \ge 1$ and $\varepsilon > 0$. See Parthasarathy [1], Walters [1] for details.

(3.2) Lemma. Let $\{E_n\}_{n \ge 1}$ be an enumeration of the cylinder sets in Ω. Define

$$d(\mu,\nu) = \sum_{n=1}^{\infty} \frac{|\mu(E_n)-\nu(E_n)|}{2^n(1+|\mu(E_n)-\nu(E_n)|)} \quad \text{where} \quad \mu,\nu \in P(\Omega).$$

Then d is a metric on $P(\Omega)$ and the d-topology coincides with the weak* topology of $P(\Omega)$.

Proof. By the argument given in the proof of (1.3), the d-topology is equivalent to the topology having as a subbase the collection of all sets of the form $\{\nu \in P(\Omega) : |\nu(E_n)-\mu(E_n)| < \varepsilon\}$, where $\mu \in P(\Omega)$, $n \geq 1$ and $\varepsilon > 0$. By (1.6), this topology is weaker than the weak* topology. On the other hand, given $f \in C(\Omega)$ and $\varepsilon > 0$, there is, by (1.6) a function $g \in C(\Omega)$ of the form $g = \sum_{i=1}^{n} c_i 1_{E_i}$ such that $\|f-g\| < \varepsilon/4$. Let $c = \max\{|c_i| : 1 \leq i \leq n\}$. Suppose $|\mu(E_i)-\nu(E_i)| < \varepsilon/2nc$ for $1 \leq i \leq n$. Then we obtain

$$|\int f d\mu - \int f d\nu| \leq |\int f d\mu - \int g d\mu| + |\int g d\mu - \int g d\nu| + |\int g d\nu - \int f d\nu|$$

$$\leq 2\|f-g\| + \sum_{i=1}^{n} |c_i||\mu(E_i)-\nu(E_i)| < \varepsilon. \quad \square$$

(3.3) Corollary. Let $\mu, \mu_1, \mu_2, \ldots$ be probability measures in $P(\Omega)$. Then the following are equivalent:

(i) $\{\mu_n\}_{n\geq 1}$ converges to μ in the weak* topology.

(ii) $\lim_{n\to\infty} \mu_n E = \mu(E)$ for each cylinder set E.

(3.4) Theorem. $P(\Omega)$ is a compact convex set in the weak* topology.

Proof. By (3.2), the weak* topology is metrizable. Thus, to prove the compactness of $P(\Omega)$, it suffices to show that $P(\Omega)$ is sequentially compact. Let $\{\mu_n\}_{n\geq 1}$ be a sequence in $P(\Omega)$, and let $\{E_k\}_{k\geq 1}$ be an enumeration of the cylinder sets in Ω. By Cantor's diagonal procedure, we obtain a subsequence $\{\mu_{n_i}\}_i$ of $\{\mu_n\}_n$ such that the sequence of real numbers $\{\mu_{n_i}(E_k)\}_i$ converges for each cylinder set E_k. Define the set function ν on S by $\nu(E_k) = \lim_{i\to\infty} \mu_{n_i}(E_k)$ for each cylinder set E_k, and $\nu(\phi) = 0$. We see readily that ν is finitely additive on S. By an argument given in the proof of (2.4), ν is σ-additive on S, so that it is uniquely extended to a probability measure on Ω, denoted by μ. By (3.3), we obtain $\mu_{n_i} \to \mu$. □

(3.5) Definition. Let K be a subset of a real vector space X. The convex hull of K, denoted by $ch(K)$, is defined by $ch(K) =$ $\{\sum_{i=1}^{n} \alpha_i x_i : x_i \in K,\ 0 \leq \alpha_i \leq 1,\ \sum_{i=1}^{n} \alpha_i = 1,\ n \geq 1\}$. A point x in K is called an extreme point of K if whenever $x = \alpha y + (1-\alpha)z$, $0 < \alpha < 1$, $x,y \in K$, then $x=y=z$. The set of all extreme points of K is denoted by ext K.

(3.6) Theorem. $\text{ext}(P(\Omega)) = \{\varepsilon_\omega : \omega \in \Omega\}$.

Proof. Suppose $\varepsilon_\omega = p\mu + (1-p)\ $, where $\mu,\nu \in P(\Omega)$, $0 < p < 1$, and $\omega \in \Omega$ with $x_n(\omega) = i_n$, $n \geq 0$. Then we have

$$1 = \varepsilon_\omega(Z(i_0,\ldots,i_n)) = p\mu(Z(i_0,\ldots,i_n)) + (1-p)\nu(Z(i_0,\ldots,i_n))$$

for all $n \geq 0$,

so that $\mu(Z(i_0,\ldots,i_n)) = \nu(Z(i_0,\ldots,i_n)) = 1$ for all $n \geq 0$. Since $\{\omega\} = \bigcap_{n=0}^{\infty} Z(i_0,\ldots,i_n)$, we obtain $\mu(\{\omega\}) = \nu(\{\omega\}) = 1$, that is, $\mu = \nu = \varepsilon_\omega$.

On the other hand, suppose $\mu \in P(\Omega)$ is such that $\mu \neq \varepsilon_\omega$ for all $\omega \in \Omega$. Then

(i) There is a point $\omega \in \Omega$ such that $\mu(U) > 0$ for every open nhd (neighbourhood) U of ω.

If not, we choose, for each $\omega \in \Omega$, an open nhd U_ω of ω with $\mu(U_\omega) = 0$. By the compactness of Ω, there are points $\omega_1,\ldots,\omega_n \in \Omega$ such that $\Omega = \bigcup_{i=1}^{n} U_{\omega_i}$, and thus $0 < \mu(\Omega) = \sum_{i=1}^{n} \mu(U_{\omega_i}) = 0$, a contradiction.

(ii) There is an open nhd V of ω with $0 < \mu(\overline{V}) < 1$, where $\overline{V}$ denotes the closure of V.

By (i), we have $0 < \mu(U) \leq \mu(\overline{U}) \leq 1$ for every open nhd U of ω. Suppose $\mu(\overline{U}) = 1$ for some open nhd U of ω. Since $\mu \neq \varepsilon_\omega$, we have $0 < \mu(\overline{U}-\{\omega\}) \leq 1$. By the regularity of μ, there is a compact set $C \subset \overline{U} - \{\omega\}$ such that $0 < \mu(C) \leq \mu(\overline{U}-\{\omega\})$. Since Ω is a compact metrizable space, there is an open set V such that $\omega \in V \subset \overline{V} \subset \overline{U} - C$. Using (i) and since $\mu(C) > 0$, we obtain $0 < \mu(V) \leq \mu(\overline{V}) < 1$.

(iii) Let V be as in (ii), and let $p = \mu(\overline{V})$, $q = 1-p$. Define the measures μ_1 and μ_2 by

$$\mu_1(E) = \frac{1}{p}\,\mu(\overline{V}\cap E), \quad \mu_2(E) = \frac{1}{q}\,\mu((\Omega-\overline{V})\cap E)$$

for any $E \in \mathcal{B}(\Omega)$.

Then μ_1 and μ_2 are distinct elements in $P(\Omega)$, and $\mu = p\mu_1 + q\mu_2$, so that $\mu \notin \text{ext}(P(\Omega))$. ◫

We shall give a simple proof of the Krein-Milman theorem for $P(\Omega)$.

(3.7) <u>Theorem.</u> $P(\Omega)$ is the closed convex hull of $E(\Omega)$ in the weak* topology, that is, $P(\Omega) = \text{cch}(E(\Omega))$.

<u>Proof.</u> Since $P(\Omega)$ is convex and $E(\Omega) \subset P(\Omega)$, we obtain $\text{ch}(E(\Omega)) \subset P(\Omega)$. By (3.4), we also obtain $\text{cch}(E(\Omega)) \subset P(\Omega)$. To prove $P(\Omega) \subset \text{cch}(E(\Omega))$, let $V(\mu,n,\varepsilon) = \{\nu \in P(\Omega) : |\nu(Z(i_0,\ldots,i_n)) - \mu(Z(i_0,\ldots,i_n))| < \varepsilon,\ i_0,\ldots,i_n \in S\}$ where $\mu \in P(\Omega)$, $\varepsilon > 0$, and $n \geq 0$. Let $\{A_k\}_{1\leq k\leq r^{n+1}}$ be an enumeration of the partition $\{Z(i_0,\ldots,i_n) : i_0,\ldots,i_n \in S\}$ of Ω. Choose arbitrary points $\omega_k \in A_k$, for $1 \leq k \leq r^{n+1}$. Define the measure $\nu = \sum_{k=1}^{r^{n+1}} \mu(A_k)\cdot\varepsilon_{\omega_k}$. We see readily that $\nu \in \text{ch}(E(\Omega))$ and $\nu(A_k) = \mu(A_k)$ for all k, so that $\nu \in V(\mu,n,\varepsilon)$. □

We now consider the mapping T on $P(\Omega)$, introduced in section 2. Recall that this mapping is derived from the shift, T, on Ω, introduced in section 1. We have the following theorem.

(3.8) Theorem. The mapping $T : P(\Omega) \to P(\Omega)$ defined by $(T\mu)(A) = \mu(T^{-1}A)$, for any $A \in \mathcal{B}(\Omega)$, is continuous and affine, that is, $T(p\mu+(1-p)\nu) = pT\mu + (1-p)T\nu$, for any $\mu,\nu \in P(\Omega)$, $p \in [0,1]$.

Proof. Suppose $\mu_n \to \mu$ in $P(\Omega)$. Let $f \in C(\Omega)$. By (1.8), $f\circ T \in C(\Omega)$, so that

$$\int fd(T\mu_n) = \int f\circ Td\mu_n \to \int f\circ Td\mu = \int fd(T\mu).$$

Thus $T\mu_n \to T\mu$.

Let $\sigma = p\mu + (1-p)\nu$ where $\mu,\nu \in P(\Omega)$ and $0 \le p \le 1$. Then $\sigma \in P(\Omega)$ and $(T\sigma)(A) = \sigma(T^{-1}A) = p\mu(T^{-1}A) + (1-p)\nu(T^{-1}A) =$ $= p(T\mu)(A) + (1-p)(T\nu)(A)$ for each $A \in \mathcal{B}(\Omega)$, so that $T(p\mu+(1-p)\nu) =$ $= p(T\mu) + (1-p)(T\nu)$. □

(3.9) Definition. Let $P(\Omega,T)$ denote the set of those measures μ in $P(\Omega)$ such that $T\mu = \mu$.

Examples (2.1), (2.3) and (2.7), together with theorem (3.8), show that $P(\Omega,T)$ is a nonempty convex subset of $P(\Omega)$.

(3.10) Theorem. $P(\Omega,T)$ is a compact convex nowhere dense subset of $P(\Omega)$.

Proof. For any sequence $\{\mu_n\} \subset P(\Omega,T) \subset P(\Omega)$, there is, by (3.4), a subsequence $\{\mu_{n_i}\}$ such that $\mu_{n_i} \to \mu \in P(\Omega)$. By (3.8), we have

$\mu_{n_i} = T\mu_{n_i} \to T\mu$, so that $T\mu = \mu \in P(\Omega,T)$. Thus $P(\Omega,T)$ is a compact (closed) subset of $P(\Omega)$.

To prove that $P(\Omega,T)$ is nowhere dense in $P(\Omega)$, let μ be a measure in $P(\Omega,T)$ and let $\omega \in \Omega$ be such that $T\omega \neq \omega$. Define the sequence $\{\mu_n\}$ by $\mu_n = \frac{1}{n}\varepsilon_\omega + (1 - \frac{1}{n})\mu$ for all $n = 1,2,\ldots$. Then $\mu_n \in P(\Omega) - P(\Omega,T)$ for all $n \geq 1$, and $\mu_n \to \mu$. Thus the interior of $P(\Omega,T)$ is empty, and since it is closed, $P(\Omega,T)$ is nowhere dense. □

For each i in S, let $[i]$ denote the point $\omega = (\omega_n)_{n\geq 0}$ in the sequence space Ω such that $\omega_n = i$ for all $n \geq 0$.

(3.11) Theorem. $P(\Omega,T) \cap \text{ext}\, P(\Omega) = \{\varepsilon_{[i]} : i \in S\} \underset{\neq}{\subset} \text{ext}\, P(\Omega,T)$.

Proof. By (3.6), we have $P(\Omega,T) \cap \text{ext}\, P(\Omega) = P(\Omega,T) \cap \{\varepsilon_\omega : \omega \in \Omega\}$. Suppose $\varepsilon_\omega \in P(\Omega,T)$. Then $\varepsilon_\omega = T\varepsilon_\omega = \varepsilon_{T\omega}$, so that $T\omega = \omega$, or equivalently, $\omega = [i]$ for some $i \in S$. It is plain that $\{\varepsilon_{[i]} : i \in S\} \subset P(\Omega,T) \cap \{\varepsilon_\omega : \omega \in \Omega\}$, so that $\{\varepsilon_{[i]} : i \in S\} = P(\Omega,T) \cap \{\varepsilon_\omega : \omega \in \Omega\}$.

Suppose $\varepsilon_{[i]} = p\mu + (1-p)\nu$ for some $i \in S$, where $\mu,\nu \in P(\Omega,T) \subset P(\Omega)$ and $0 < p < 1$. By (3.6), we obtain $\varepsilon_{[i]} = \mu = \nu$, so that $\varepsilon_{[i]} \in \text{ext}\, P(\Omega,T)$ for all $i \in S$. It will be shown in sections 4 and 7 that $\{\varepsilon_{[i]} : i \in S\} \underset{\neq}{\subset} \text{ext}\, P(\Omega,T)$. □

(3.12) Remark. By a theorem of Oxtoby [1], the set of extreme points of any compact convex metric space is a G_δ-set. Thus both ext $P(\Omega)$ and ext $P(\Omega,T)$ are G_δ sets.

(3.13) Definition. A measure μ in $P(\Omega,T)$ is called T-ergodic or simply ergodic if whenever $T^{-1}A = A$, $A \in \mathcal{B}(\Omega)$, then either $\mu(A) = 0$ or $\mu(A) = 1$. Let $E(\Omega,T)$ denote the set of all T-ergodic measures in $P(\Omega,T)$.

(3.14) Remark. For each μ in $P(\Omega,T)$, let $\mathcal{I}_o(\mu,T) = \{A\in\mathcal{B}(\Omega) : A = T^{-1}A\}$ and $\mathcal{I}(\mu,T) = \{A\in\mathcal{B}(\Omega) : \mu(A\Delta T^{-1}A) = 0\}$. Then $\mathcal{I}_o(\mu,T)$ and $\mathcal{I}(\mu,T)$ are sub-σ-algebras of $\mathcal{B}(\Omega)$ such that $\mathcal{I}_o(\mu,T) \subset \mathcal{I}(\mu,T)$. It is known that $\mu \in P(\Omega,T)$ is ergodic iff $\mu(A) = 0$ or $\mu(A) = 1$ for each $A \in \mathcal{I}(\mu,T)$. (See Billingsley [1], Phelps [1], Walters [1]).

We state without proof the following useful result (see Brown [1], Walters [1], Denkers et al [1]).

(3.15) Theorem. Let μ be a measure in $P(\Omega,T)$. Then the following are equivalent:

(i) $\mu \in E(\Omega,T)$.

(ii) For each A,B in $\mathcal{S}$, $\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu(T^{-i}A \cap B) = \mu(A)\mu(B)$.

(iii) For each $f \in L_1(\Omega,\mathcal{B}(\Omega),\mu)$, $\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i\omega) = \int f d\mu$ μ-almost everywhere.

By definition, for measures $\mu,\nu \in P(\Omega)$, μ is <u>singular</u> with respect to ν, denoted $\mu \perp \nu$, if there exist sets $A,B \in \mathcal{B}(\Omega)$ such that $A \cap B = \phi$ and $\mu(A) = 1 = \nu(B)$. Also, ν is <u>absolutely continuous</u> with respect to μ, denoted $\nu << \mu$, if $\nu(B) = 0$ whenever $\mu(B) = 0$, $B \in \mathcal{B}(\Omega)$.

(3.16) <u>Theorem</u>. Let $\mu,\nu \in P(\Omega,T)$. Then

(i) If $\mu,\nu \in E(\Omega,T)$, then either $\mu = \nu$ or $\mu \perp \nu$.

(ii) If $\nu << \mu$, with $\mu \in E(\Omega,T)$, then $\mu = \nu$.

<u>Proof</u>. (i) (Billingsley [1]): Suppose $\mu \neq \nu$. Then there exists a set $A \in \mathcal{B}(\Omega)$ such that $\mu(A) \neq \nu(A)$. By (3.15), we have

$$\lim_n \frac{1}{n} \sum_{i=0}^{n-1} 1_A(T^i\omega) = \mu(A) \quad \mu\text{-a.e,}$$

and

$$\lim_n \frac{1}{n} \sum_{i=0}^{n-1} 1_A(T^i\omega) = \nu(A) \quad \nu\text{-a.e..}$$

If we let $E = \{\omega : \lim_n \frac{1}{n} \sum_{i=0}^{n-1} 1_A(T^i\omega) = \mu(A)\}$ and

$F = \{\omega : \lim_n \frac{1}{n} \sum_{i=0}^{n-1} 1_A(T^i\omega) = \nu(A)\}$, then $\mu(E) = 1 = \nu(F)$ and $E \cap F = \phi$, so that $\mu \perp \nu$.

(ii): Let $E \in \mathcal{B}(\Omega)$ be arbitrary. It follows from (3.15) that

$\lim_n \frac{1}{n} \sum_{i=0}^{n-1} 1_E(T^i\omega) = \mu(E)$ μ-a.e.. Since $\nu << \mu$, it follows that

$\lim_n \frac{1}{n} \sum_{i=0}^{n-1} 1_E(T^i\omega) = \mu(E)$ ν-a.e.. Thus, from the Lebesgue dominated

convergence theorem, we obtain

$$\mu(E) = \int \mu(E) d\nu = \lim_n \frac{1}{n} \sum_{i=0}^{n-1} \int (1_E \circ T^i) d\nu = \nu(E). \quad \square$$

We now show that $E(\Omega,T) = \text{ext } P(\Omega,T)$.

(3.17) Theorem. For each $\mu \in P(\Omega,T)$, $\mu \in E(\Omega,T)$ iff $\mu \in \text{ext } P(\Omega,T)$.

Proof. ($\Rightarrow$) Suppose $\mu \in E(\Omega,T)$ and $\mu = p\mu_1 + (1-p)\mu_2$, for some $\mu_1, \mu_2 \in P(\Omega,T)$, $0 < p < 1$. Then we have $\mu_i << \mu$, $i = 1,2$, so that, by (3.16), $\mu = \mu_1 = \mu_2$. Thus $\mu \in \text{ext } P(\Omega,T)$.

($\Leftarrow$) Let $\mu \in P(\Omega,T)$, and suppose $\mu \notin E(\Omega,T)$. Then there is a set $A \in \mathcal{B}(\Omega)$ such that $T^{-1}A = A$ and $0 < \mu(A) < 1$. Let $B = \Omega - A$. Define two measures μ_1 and μ_2 by

$$\mu_1(E) = \mu(E \cap A)/\mu(A) \quad \text{and} \quad \mu_2(E) = \mu(E \cap B)/\mu(B)$$

for any $E \in \mathcal{B}(\Omega)$.

Then we obtain $\mu(T^{-1}E \cap A) = \mu(T^{-1}E \cap T^{-1}A) = \mu(T^{-1}(E \cap A)) = \mu(E \cap A)$ so that $(T\mu_1)(E) = \mu_1(E)$. Thus $\mu_1 \in P(\Omega,T)$. Similarly, we also obtain

$\mu_2 \in P(\Omega,T)$. Note that $\mu_1(A) = 1$ and $\mu_2(A) = 0$, so that $\mu_1 \neq \mu_2$. Thus, since $\mu = \mu(A)\mu_1 + \mu(B)\mu_2$, we have that $\mu \notin \text{ext } P(\Omega,T)$. □

(3.18) Remark. By (3.10), the set $P(\Omega,T)$ is a compact convex set so that, by the Krein-Milman theorem, $P(\Omega,T) = \text{cch } E(\Omega,T)$. Since both $P(\Omega)$ and $P(\Omega,T)$ are metrizable compact convex sets, we can apply the Choquet representation theorem (recall that $E(\Omega) = \text{ext } P(\Omega)$ and $P(\Omega) = \text{cch } E(\Omega)$): For each μ in $P(\Omega)$ (or $P(\Omega,T)$), there is a unique probability measure τ defined on the Borel subsets of the compact metrizable space $P(\Omega)$ (or $P(\Omega,T)$) such that $\tau(E(\Omega)) = 1$ (or $\tau(E(\Omega,T)) = 1$) and

$$\int_\Omega f d\mu = \int_E \left(\int_\Omega f d\nu\right) d\tau(\nu)$$

for each $f \in C(\Omega)$, where $E = E(\Omega)$ or $E = E(\Omega,T)$. See Phelps [1].

4. Markov measures and ergodic Markov measures

Note that, given a measure μ in $P(\Omega)$ and sets A,B in $P(\Omega)$, the conditional probability of A given B, denoted $\mu(A|B)$, is defined by $\mu(A|B) = \frac{\mu(A\cap B)}{\mu(B)}$, provided that $\mu(B) > 0$.

On a probability space $(\Omega,\mathcal{B}(\Omega),\mu)$, the sequence $\{x_n\}_{n\geq 0}$ is called a Markov chain with state space S and transition matrix $P = (p_{ij})_{i,j\in S}$ if the following conditions hold:

(i) for each $i\in S$, $\mu(x_n=i) > 0$ for some $n \geq 0$,

(ii) The Markov property: $\mu(x_{n+1}=i_{n+1}|x_0=i_0,\ldots,x_n=i_n)$ $= \mu(x_{n+1}=i_{n+1}|x_n=i_n)$ for any $n \geq 0$ and any $i_0,\ldots,i_{n+1}$ in S such that $\mu(x_0=i_0,\ldots,x_n=i_n) > 0$

(iii) The stationary property: $\mu(x_{m+1}=j|x_m=i) = \mu(x_{n+1}=j|x_n=i)$ for any $0 \leq m < n$ and any i,j such that $\mu(x_m=i) > 0$ and $\mu(x_n=i) > 0$. The common value of $\mu(x_{n+1}=j|x_n=i)$ for all n such that $\mu(x_n=i) > 0$ is denoted by p_{ij}, and is called the one-step transition probability from state i to state j.

The probability vector $p = (p_i)_{i\in S}$ defined by $p_i = \mu(x_0=i)$ is called the initial distribution for the chain.

Suppose $\{x_n\}_{n\geq 0}$ is a Markov chain on a probability space $(\Omega,\mathcal{B}(\Omega),\mu)$ as above. Define the n-step transition probabilities

$p_{ij}^{(n)}$, $n \geq 0$, as follows:

$$p_{ij}^{(0)} = \delta_{ij} \text{ where } \delta_{ij} \text{ denotes the Kronecker symbol,}$$

$$p_{ij}^{(1)} = p_{ij} ,$$

$$p_{ij}^{(n+1)} = \sum_{k \in S} p_{ik}^{(n)} p_{kj} \quad \text{for } n \geq 1.$$

Thus if we denote the n-step transition matrix by $\left(p_{ij}^{(n)}\right)_{i,j \in S}$, then $P^n = \left(p_{ij}^{(n)}\right)_{i,j \in S}$. Note that the transition matrices are stochastic matrices. Recall that a square matrix $(a_{ij})_{1 \leq i,j \leq m}$ is stochastic if $0 \leq a_{ij} \leq 1$ for each i,j, and $\sum_{j=1}^{m} a_{ij} = 1$ for each i. If a stochastic matrix $(a_{ij})_{1 \leq i,j \leq m}$ also satisfies $\sum_{i=1}^{m} a_{ij} = 1$ for each i, it is called doubly stochastic. A probability vector $p = (p_i)_{i \in S}$ is called a stationary distribution for a stochastic matrix $P = (p_{ij})_{i,j \in S}$ if $pP = p$, that is, $\sum_{i \in S} p_i p_{ij} = p_j$ for each j in S.

(4.1) The Existence Theorem. Let $p = (p_i)_{i \in S}$ be any probability vector and let $P = (p_{ij})_{i,j \in S}$ be any stochastic matrix. Then there exist a unique measure $\mu \in P(\Omega)$ and a unique nonempty set $E \subset S$ such that

(i) $\mu(x_0 = i_0, \ldots, x_n = i_n) = p_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} i_n}$ for any $n \geq 0$ and any $i_0, \ldots, i_n \in S$,

(ii) the sequence $\{x_n\}_{n\geq 0}$ defined on $(\Omega, \mathcal{B}(\Omega), \mu)$ is a Markov chain with state space E, initial distribution $(p_i)_{i\in E}$ and transition matrix $(p_{ij})_{i,j\in E}$.

<u>Proof</u>. (i) Define $p_n(i_0,\ldots,i_n) = p_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} i_n}$ for any $n \geq 0$ and any $i_0,\ldots,i_n \in S$. Then we have that $0 \leq p_n(i_0,\ldots,i_n) \leq 1$, $\sum_{i\in S} p_0(i) = \sum_{i\in S} p_i = 1$ and $\sum_{i_{n+1}\in S} p_{n+1}(i_0,\ldots,i_n,i_{n+1})$
$= \sum_{i_{n+1}\in S} p_{i_0} p_{i_0 i_1} \cdots p_{i_n i_{n+1}} = p_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} i_n}$. By the Kolmogorov extension theorem (2.4), there exists a unique $\mu \in P(\Omega)$ with

$$\mu(x_0=i_0, \ldots, x_n=i_n) = p_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} i_n} .$$

(ii) If $\mu(x_n=i) > 0$, then

$$\mu(x_{n+1}=j \mid x_n=i) = \mu(x_n=i,\ x_{n+1} = j)/\mu(x_n=i)$$

$$= \sum_{i_0,\ldots,i_{n-1}} \mu(x_0=i_0, \ldots, x_{n-1}=i_{n-1}, x_n=i, x_{n+1}=j) /$$

$$\sum_{i_0,\ldots,i_{n-1}} \mu(x_0=i_0, \ldots, x_{n-1}=i_{n-1}, x_{n}=i)$$

$$= \sum_{i_0,\ldots,i_{n-1}} p_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} i} p_{ij} / \sum_{i_0,\ldots,i_{n-1}} p_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} i} = p_{ij}.$$

Suppose $\mu(x_0=i_0, \ldots, x_n=i_n) > 0$. Then we also obtain

$$\mu(x_{n+1}=j \mid x_0=i_0, \ldots, x_n=i_n)$$

$$= \mu(x_0=i_0, \ldots, x_n=i_n, x_{n+1}=j)/\mu(x_0=i_0, \ldots, x_n=i_n)$$

$$= p_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} i_n} p_{i_n j} / p_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} i_n}$$

$$= p_{i_n j} = \mu(x_{n+1}=j \mid x_n=i_n)$$

Let $E = \{i \in S : \mu(x_n=i) > 0$ for some $n \geq 0\}$. By definition, $p_i = \mu(x_0=i)$ and $\sum_{i\in S} p_i = 1$, so that $E \supset \{i \in S : p_i > 0\} \neq \phi$. We also have $1 = \sum_{i\in S} p_i = \sum_{i\in E} p_i$. Let $i \in E$. Then, $\mu(x_n=i) > 0$ for some $n \geq 0$ and $\mu(x_{n+1}=j) = 0$ for all $j \in S - E$, so that

$$1 = \sum_{j\in S} p_{ij} = \sum_{j\in S} \mu(x_{n+1}=j \mid x_n=i)$$

$$= \sum_{j\in E} \mu(x_{n+1}=j \mid x_n=i) = \sum_{j\in E} p_{ij} . \qquad \square$$

(4.2) Definition. The measure $\mu \in P(\Omega)$ defined by means of the probability vector $p = (p_i)_{i\in S}$ and stochastic matrix $P = (p_{ij})_{i,j\in S}$ as in (4.1) is called the (p,P)-Markov measure.

(4.3) Corollary. For any two probability vectors $p = (p_i)_{i\in S}$ and $q = (q_i)_{i\in S}$ there exists a unique probability measure μ in $P(\Omega)$, called the (q,p)-Markov measure, such that

$$\mu(x_0=i_0, \ldots, x_n=i_n) = q_{i_0} p_{i_1} \cdots p_{i_n}$$

for any $n \geq 0$ and any $i_0,\ldots,i_n$ in S. Moreover, $\{x_n\}_{n\geq 0}$ is a sequence of independent random variables on $(\Omega,\mathcal{B}(\Omega),\mu)$ with

$$\mu(x_0=i) = q_i \quad \text{and} \quad \mu(x_n=i) = p_i ,$$

for any $i \in S$ and any $n \geq 1$.

<u>Proof</u>. Define the stochastic matrix $P = (p_{ij})_{i,j\in S}$ by $p_{ij} = p_j$ for all $i,j \in S$. By (4.1), there exists a unique $\mu \in P(\Omega)$ with $\mu(x_0=i_0, \ldots, x_n=i_n) = q_{i_0} p_{i_1} \cdots p_{i_n}$. Clearly $\mu(x_0=i) = q_i$. We also obtain, for $n \geq 1$ and $i \in S$,

$$\mu(x_n=i) = \sum_{i_0,\ldots,i_{n-1}} \mu(x_0=i_0, \ldots, x_{n-1}=i_{n-1}, x_n=i)$$

$$= \sum_{i_0,\ldots,i_{n-1}} q_{i_0} p_{i_1} \cdots p_{i_{n-1}} p_i = p_i .$$

Therefore, $\mu(x_0=i_0, \ldots, x_n=i_n) = q_{i_0} p_{i_1} \cdots p_{i_n} = \mu(x_0=i_0)\mu(x_1=i_1)\ldots$ $\ldots \mu(x_n=i_n)$. □

As an immediate consequence of (4.3) we have

(4.4) <u>Corollary</u>. For any probability vector $p = (p_i)_{i\in S}$, there exists a unique probability measure $\mu \in P(\Omega)$, called the p-Bernoulli measure, such that

$$\mu(x_0=i_0, \ldots, x_n=i_n) = p_{i_0} p_{i_1} \cdots p_{i_n}$$

for any $n \geq 0$ and any $i_0,\ldots,i_n \in S$. Moreover, $\{x_n\}_{n\geq 0}$ is a sequence of i.i.d. (independent identically distributed) random variables on $(\Omega, \mathcal{B}(\Omega), \mu)$ with common distribution p.

(4.5) Theorem. Let μ be the (p,P)-Markov measure. Then the following are equivalent:

(i) p is a stationary distribution for P, i.e. $pP = p$.

(ii) μ is T-invariant, i.e. $\mu \in P(\Omega,T)$.

(iii) $\{x_n\}_{n\geq 0}$ is a stationary process on $(\Omega, \mathcal{B}(\Omega), \mu)$, that is,

$$\mu(x_k=i_0, x_{k+1}=i_1, \ldots, x_{k+n}=i_n) = \mu(x_0=i_0, x_1=i_1, \ldots, x_n=i_n)$$

for any $n \geq 0$, $k \geq 1$ and any $i_0,\ldots,i_n$ in S.

Proof. Let $p = (p_i)_{i\in S}$ and $P = (p_{ij})_{i,j\in S}$.

(i) $\Longrightarrow$ (ii): Suppose (i) holds. Let $p_n(i_0,i_1,\ldots,i_n) = p_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} i_n}$ for any $n \geq 0$ and any $i_0,\ldots,i_n$ in S. By (i), we obtain

$$\sum_{i\in S} p_{n+1}(i,i_0,\ldots,i_n) = \sum_i p_i p_{ii_0} p_{i_0 i_1} \cdots p_{i_{n-1} i_n} = p_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} i_n} =$$

$= p_n(i_0,i_1,\ldots,i_n)$, so that, by (2.5), $\mu \in P(\Omega,T)$.

(ii) ⟹ (iii): Suppose (ii) holds. Then we have

$$\mu(x_k=i_0,\ x_{k+1}=i_1,\ \ldots,\ x_{k+n}=i_n) = \mu T^{-k}(x_0=i_0,\ \ldots,\ x_n=i_n)$$

$$= \mu(x_0=i_0,\ \ldots,\ x_n=i_n).$$

(iii) ⟹ (i): If (iii) holds, then for each j in S, we have

$$p_j = \mu(x_0=j) = \mu(x_1=j) = \sum_{i\in S} \mu(x_0=i,\ x_1=j) = \sum_i p_i p_{ij}\ . \quad \square$$

(4.6) <u>Definition</u>. Let $M(\Omega)$ denote the set of all (p,P)-Markov measures and let $M(\Omega,T)$ denote the set of those (p,P)-Markov measures with $pP = p$. Let $B(\Omega,T)$ denote the set of all Bernoulli measures.

From (4.4) and (4.5) we have $\phi \neq B(\Omega,T) \subset M(\Omega,T) \subset P(\Omega,T)$. We shall characterize the set $M(\Omega,T) \cap E(\Omega,T)$ of ergodic Markov measures. We begin with a lemma.

(4.7) <u>Lemma</u>. Let $P = (p_{ij})_{i,j\in S}$ be a stochastic matrix. Then

(i) The Césaro limit $\lim\limits_{n\to\infty} \frac{1}{n} \sum\limits_{k=1}^{n} p_{ij}^{(k)} = q_{ij}$ exists for each i and j.

(ii) The matrix $Q = (q_{ij})_{i,j\in S}$ is a stochastic matrix such that $Q = \lim\limits_{n\to\infty} \frac{1}{n} \sum\limits_{k=1}^{n} P^k$, $QP = PQ = Q$, and $Q^2 = Q$.

(iii) Each row of the matrix Q is a stationary distribution for P.

Proof. (i), (ii): Let $Q_n = \frac{1}{n} \sum_{m=1}^{n} P^m$, where $n \geq 1$. If we write $Q_n = \left(q_{ij}^{(n)}\right)_{i,j \in S}$, then $q_{ij}^{(n)} = \frac{1}{n} \sum_{m=1}^{n} p_{ij}^{(m)}$ so that $0 \leq q_{ij}^{(n)} \leq 1$, for each i and j. Therefore, there is a subsequence $\{Q_{n_k}\}$ of $\{Q_n\}$ such that $\lim_{n_k \to \infty} q_{ij}^{(n_k)} = q_{ij}$ for each i,j in S. Let $Q = (q_{ij})_{i,j \in S}$. Then we have

$$\sum_{j \in S} q_{ij}^{(n_k)} = \sum_j \frac{1}{n_k} \sum_{m=1}^{n_k} p_{ij}^{(m)} = \frac{1}{n_k} \sum_{m=1}^{n_k} \sum_j p_{ij}^{(m)} = 1$$

so that

$$\sum_j q_{ij} = \sum_j \lim_{n_k \to \infty} q_{ij}^{(n_k)} = \lim_{n_k \to \infty} \sum_j q_{ij}^{(n_k)} = 1.$$

Thus Q is stochastic. On the other hand, we have

$$PQ_{n_k} = Q_{n_k} P = \frac{1}{n_k} \sum_{m=1}^{n_k} P^{m+1} = Q_{n_k} + \frac{1}{n_k} (P^{n_k+1} - P)$$

and so, taking the limit as $n_k \to \infty$, $PQ = QP = Q$.

Suppose $\{Q_{m_k}\}$ is a subsequence of $\{Q_n\}$ such that $\lim_{m_k \to \infty} Q_{m_k} = Q'$. Then we obtain

$$Q = Q\left(\frac{1}{m_k} \sum_{i=1}^{m_k} P^i\right) = QQ_{m_k} = Q_{m_k} Q$$

So that $Q = QQ' = Q'Q$. By interchanging the roles of Q and Q' we have

$Q' = QQ' = Q'Q = Q$. Therefore, $Q = Q^2$ and

$$Q = \lim_{n\to\infty} Q_n = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} P^k, \quad \text{that is,}$$

$$q_{ij} = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} p_{ij}^{(k)} \quad \text{for each } i,j \in S.$$

(iii) Let π_i be the i-th row of the matrix Q, that is, $\pi_i = (q_{ij})_{j\in S}$, and let $\pi_i(k)$ be the k-th coordinate of π_i, that is $\pi_i(k) = q_{ik}$. Since $QP = Q$, we get for each j,

$$\sum_k \pi_i(k)p_{kj} = \sum_k q_{ik}p_{kj} = q_{ij} = \pi_i(j). \quad \square$$

See Chung [1] and Hoel et al. [1] for probabilistic proofs of the above lemma.

(4.8) <u>Remarks</u>. Let $P = (p_{ij})_{i,j\in S}$ be a stochastic matrix. A state i in S is called <u>recurrent</u> (or <u>P-recurrent</u>) if $\sum_{n=1}^{\infty} p_{ii}^{(n)} = \infty$ and is called <u>transient</u> (or <u>P-transient</u>) if $\sum_{n=1}^{\infty} p_{ii}^{(n)} < \infty$. For each transient state j, we have $\sum_{n=1}^{\infty} p_{ij}^{(n)} < \infty$ for all $i \in S$. Let C and D denote the set of recurrent states and the set of transient states, respectively. Then $S = C \cup D$. A nonempty subset E of S is called a <u>closed</u> (or <u>P-closed</u>) set if $\sum_{j\in E} p_{ij} = 1$ for each $i \in E$. A closed set E is called <u>irreducible</u> if, for any $i,j \in E$, $p_{ij}^{(m)} > 0$ for some $m \geq 1$ and

$p_{ji}^{(n)} > 0$ for some $n \geq 1$. It is well-known that the set C is a nonempty closed set and is the union of a finite number of disjoint irreducible closed sets $C_1, \ldots, C_m$ and that the set D is not closed. Therefore, the state space S can be partitioned uniquely into the set D of transient states and a finite collection $\{C_k\}_{1 \leq k \leq m}$ of irreducible closed sets (of recurrent states). The matrix P is called irreducible if S is irreducible. See Chung [1], Doob [1], Feller [1], Hoel et al. [1] for details.

(4.9) Lemma. Let $P = (p_{ij})_{i,j \in S}$ be a stochastic matrix and let $Q = (q_{ij})_{i,j \in S}$ where $Q = \lim_n \frac{1}{n} \sum_{k=1}^{n} P^k$. Suppose that the set C of P-recurrent states is P-irreducible and that the set D of P-transient states is not empty. Then

(i) $q_{ij} = 0$ for all $(i,j) \in S \times D$,

(ii) $q_{ij} = q_{jj} > 0$ for all $(i,j) \in S \times C$,

(iii) P has a unique stationary distribution $p = (p_i)_{i \in S}$ given by $p_i = q_{ii}$.

Proof. (i): For each $j \in D$, we have $\sum_{n=1}^{\infty} p_{ij}^{(n)} < \infty$ for all $i \in S$, so that $q_{ij} = \lim_n \frac{1}{n} \sum_{k=1}^{n} p_{ij}^{(n)} = 0$ for all $i \in S$.

(ii): Let i be an arbitrary state in S and let $E_i = \{j \in C : q_{ij} > 0\}$. We claim that $E_i = C$. By (4.7), together with (i), we get

$$1 = \sum_{j\in S} q_{ij} = \sum_{j\in C} q_{ij}$$

so that $q_{ij_o} > 0$ for some $j_o \in C$. Let k be any state in C. Since C is irreducible, we have $p^{(n)}_{j_ok} > 0$ for some $n \geq 1$, so that by (4.7), $q_{ik} = \sum_{m\in S} q_{im} p^{(n)}_{mk} \geq q_{ij_o} p^{(n)}_{j_ok} > 0$. Thus $k \in E_i$, and so $E_i = C$. Since i was arbitrary, it follows that $q_{ij} > 0$ for each $(i,j) \in S\times C$.

It remains to show that $q_{jj} = q_{ij}$ for each $i \in S$, $j \in C$. For an arbitrary $j \in C$, let $q_j = \max\{q_{ij} : i \in S\}$. Suppose $q_j > q_{kj}$ for some $k \in S$. Since $Q = Q^2$, we obtain $q_{ij} = \sum_m q_{im}q_{mj} < \left(\sum_m q_{im}\right)\cdot q_j = q_j$, for all i so that $q_j < q_j$, a contradiction. Thus $q_{jj} = q_{ij} > 0$ for all $i \in S$, $j \in C$.

(iii) By (4.7), together with (i) and (ii), the probability vector $p = (p_i)_{i\in S}$ defined by $p_i = q_{ii}$ is a stationary distribution of P. Suppose $\pi = (\pi_i)_{i\in S}$ is a stationary distribution of P. Then we have

$$\pi_j = \sum_i \pi_i p^{(n)}_{ij} \quad \text{for each } j \text{ and each } n \geq 1,$$

so that

$$\pi_j = \sum_i \pi_i \left(\frac{1}{n} \sum_{k=1}^{n} p^{(k)}_{ij}\right) \quad \text{for each } j \text{ and each } n \geq 1.$$

Letting $n \to \infty$, we obtain

$$\pi_j = \sum_i \pi_i q_{ij} = \sum_i \pi_i q_{jj} = q_{jj} \quad \text{for all } j. \quad \square$$

(4.10) <u>Corollary</u>. Let $P = (p_{ij})_{i,j\in S}$ be an irreducible stochastic matrix and let $Q = (q_{ij})_{i,j\in S}$ where $Q = \lim_n \frac{1}{n} \sum_{k=1}^{n} P^k$. Then P has a unique positive stationary distribution $p = (p_i)_{i\in S}$ given by $p_i = q_{ii} > 0$, $i \in S$.

We prove the following version of a theorem of Doob [1].

(4.11) <u>Theorem</u>. Let $P = (p_{ij})_{i,j\in S}$ be a stochastic matrix and $Q = (q_{ij})_{i,j\in S}$ be given by $Q = \lim_n \frac{1}{n} \sum_{k=1}^{n} P^k$. Then the state space S can be partitioned uniquely into the set D of P-transient states and a finite collection $\{C_1,\ldots,C_m\}$ of P-irreducible closed sets such that

(i) for each $i \in S$ and each $j \in D$, $q_{ij} = 0$,

(ii) for each C_k, $q_{ij} = q_{jj}$ for all $i,j \in C_k$, and $\pi_k = (q_{jj})_{j\in C_k}$ is a unique stationary distribution of the stochastic matrix $(p_{ij})_{i,j\in C_k}$,

(iii) for distinct C_s and C_t, $q_{ij} = 0$ for each $i \in C_s$ and $j \in C_t$,

(iv) for each $i \in D$ and each $j \in C_k$, $q_{ij} = \left(\sum_{t\in C_k} q_{it}\right) q_{jj}$.

Furthermore, we have

(v) $p = (p_i)_{i \in S}$ is a stationary distribution for P iff $p = \sum_{k=1}^{m} c_k \sigma_k$, where σ_k is the probability distribution on S such that $\sigma_k(i) = \pi_k(i)$ for each $i \in C_k$, and $0 \le c_k \le 1$ with $\sum_{k=1}^{m} c_k = 1$.

Proof. The existence of the partition of S into $\{D, C_1, \ldots, C_m\}$ follows from (4.8). Clearly (i) holds. (ii) follows from (4.10).

(iii): Suppose $C_s \cap C_t = \phi$. Let $i \in C_s$, $j \in C_t$. Since both C_s and C_t are P-closed sets, we have $p_{ij}^{(n)} = 0$ for all $n \ge 0$, and thus $q_{ij} = 0$.

(iv): Let $i \in D$ and $j \in C_k$. It follows from (4.7), together with the preceding results, that

$$q_{ij} = \sum_{t \in S} q_{it} q_{tj} = \sum_{t \in C_k} q_{it} q_{tj} = \left(\sum_{t \in C_k} q_{it} \right) q_{jj}.$$

(v) ($\Rightarrow$) Suppose $p = (p_i)_{i \in S}$ is a stationary distribution of P. Fix C_k, where $1 \le k \le m$. Then we have, for each $j \in C_k$,

$$p_j = \sum_{i \in S} p_i \, p_{ij}^{(t)} = \sum_{i \in C_k} p_i \, p_{ij}^{(t)} \quad \text{for } t = 1, 2, \ldots,$$

so that

$$p_j = \sum_{i \in C_k} p_i \left(\frac{1}{n} \sum_{t=1}^{n} p_{ij}^{(t)} \right) \quad \text{for } n = 1, 2, \ldots.$$

Letting $n \to \infty$, we obtain,

$$p_j = \sum_{i \in C_k} p_i q_{ij} = \left(\sum_{i \in C_k} p_i \right) q_{jj} = c_k q_{jj}, \quad \text{for each } j \in C_k,$$

where $c_k = \sum_{i \in C_k} p_i$. By similar reasoning, we have $p_j = \sum_{i \in S} p_i \, q_{ij} = 0$ for all $j \in D$. It follows that $\sum_{k=1}^{m} c_k = 1$. Define the probability vector σ_k on S by $\sigma_k(j) = q_{jj}$ for $j \in C_k$ and $\sigma_k(j) = 0$ otherwise. Then we obtain $\sigma_k(i) = q_{ii} = \pi_k(i)$ for each $i \in C_k$ and $p = \sum_{k=1}^{m} c_k \sigma_k$.

(<=): It is straightforward to show that any probability vector p of the form $p = \sum_{k=1}^{m} c_k \sigma_k$, where c_k and σ_k are defined as in (v), is a stationary distribution of P. □

(4.12) Remark. Let $\mu \in M(\Omega,T)$ be a (p,P)-Markov measure with $P = (p_{ij})_{i,j \in S}$ and $p = (p_i)_{i \in S}$. By (4.1) and (4.5), the stationary process $\{x_n\}_{n \geq 0}$ defined on $(\Omega, \mathcal{B}(\Omega), \mu)$ has $p_i = \mu(x_n = i)$ for each $i \in S$ and each $n \geq 0$. Let $E = \{i \in S : p_i > 0\}$. Then, on $(\Omega, \mathcal{B}(\Omega), \mu)$, $\{x_n\}_{n \geq 0}$ is a Markov chain with state space E, transition matrix $P' = (p_{ij})_{i,j \in E}$ and stationary initial distribution $p' = (p_i)_{i \in E}$. By definition, the chain $\{x_n\}_{n \geq 0}$ is called a recurrent chain iff the state space E is irreducible, or equivalently, the matrix P' is irreducible. In this case, all states in E are called recurrent as well.

From the definition of the n-step transition probabilities we see by induction that, for any $\mu \in M(\Omega)$ with stochastic matrix $P = (p_{ij})_{i,j \in S}$ and initial distribution $p = (p_i)_{i \in S}$,

$$\mu(x_0 = i_0, \ldots x_m = i_m, x_{m+k} = j_0, \ldots, x_{m+k+n} = j_n)$$

$$= p_{i_0} p_{i_0 i_1} \cdots p_{i_{m-1} i_m} \cdot p^{(k)}_{i_m j_0} \cdot p_{j_0 j_1} \cdots p_{j_{n-1} j_n},$$

for any $i_0, \ldots, i_m, j_0, \ldots, j_n \in S$, and any $m, n \geq 0$, $k \geq 1$.

(4.13) Theorem. For each (p,P)-Markov measure μ in $M(\Omega,T)$, the following are equivalent:

(i) μ is ergodic.

(ii) there is a unique P-irreducible closed subset C of S such that $p_i > 0$ for each $i \in C$, and $\sum_{i \in C} p_i = 1$.

(iii) the stationary process $\{x_n\}_{n \geq 0}$ defined on $(\Omega, \mathcal{B}(\Omega), \mu)$ is a recurrent chain.

Proof. Notation is as in (4.11).

(i) $\Rightarrow$ (ii): Suppose μ is ergodic. Then $p_i p_j = p_i q_{ij}$ for all $i,j \in S$, for we have from (3.15) that

$$p_i p_j = \mu(x_0=i)\mu(x_0=j) = \lim_n \frac{1}{n} \sum_{k=1}^{n} \mu((x_0=i) \cap T^{-k}(x_0=j))$$

$$= \lim_n \frac{1}{n} \sum_{k=1}^{n} p_i \, p_{ij}^{(k)} = p_i q_{ij} .$$

By part (v) of (4.11), $p = \sum_{k=1}^{m} c_k \sigma_k$, where $0 \le c_k \le 1$, $\sum_{k=1}^{m} c_k = 1$. Thus $c_k > 0$ for some k. Let $i \in C_k$, where $c_k > 0$. Then $p_i = c_k \sigma_k(i) = c_k q_{ii} > 0$. But by above, $p_i = q_{ii}$, and thus we have $c_k = 1$. Therefore, $p = \sigma_k$, so that $p_j = q_{jj} > 0$ for all $j \in C_k$, and $\sum_{j \in C_k} p_j = \sum_{j \in C_k} q_{jj} = 1$.

(ii) $\Rightarrow$ (i): We may assume without loss of generality that $p = \sigma_1$. Let $A = \{x_0=i_0, \ldots, x_m=i_m\}$ and $B = \{x_0=j_0, \ldots, x_t=j_t\}$ where $m,t \ge 0$ and $i_0, \ldots, i_m, j_0, \ldots, j_t \in S$. Then we have, for each $k \geq m$,

$$\mu(A \cap T^{-k}B) = \mu(x_0=i_0, \ \ldots, \ x_m=i_m, \ x_k=j_0, \ \ldots, \ x_{k+t}=j_t)$$

$$= p_{i_0} p_{i_0 i_1} \cdots p_{i_{m-1} i_m} \, p^{(k-m)}_{i_m j_0} \, p_{j_0 j_1} \cdots p_{j_{t-1} j_t}$$

so that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \mu(A \cap T^{-k}B) = p_{i_0} p_{i_0 i_1} \cdots p_{i_{m-1} i_m} q_{i_m j_0} p_{j_0 j_1} \cdots p_{j_{t-1} j_t} .$$

We also have

$$\mu(A)\mu(B) = p_{i_0} p_{i_0 i_1} \cdots p_{i_{m-1} i_m} p_{j_0} p_{j_0 j_1} \cdots p_{j_{t-1} j_t} .$$

We want to show $\lim_n \frac{1}{n} \sum_{k=1}^{n} \mu(A \cap T^{-k}B) = \mu(A)\mu(B)$. If $\{i_0,\ldots,i_m\} \subset C_1$, then since $q_{i_m j_0} = p_{j_0}$ by (4.11), the equality holds. Otherwise, we have $\lim_n \frac{1}{n} \sum_{k=1}^{n} \mu(A \cap T^{-k}B) = 0$ and $\mu(A) = 0$, so that again the equality holds. By (3.15), we obtain (i). The equivalence (ii) <=> (iii) follows from (4.12). □

(4.14) Corollary. If $P = (p_{ij})_{i,j\in S}$ is an irreducible stochastic matrix and $p = (p_i)_{i\in S}$ is the positive stationary distribution of P, then the (p,P)-Markov measure is ergodic.

(4.15) Corollary. Every Bernoulli measure is ergodic.

(4.16) Definition. A measure $\mu \in P(\Omega,T)$ shall be called mixing (some authors use strong-mixing) if $\lim_{n\to\infty} \mu(A \cap T^{-n}B) = \mu(A)\mu(B)$ for each $A,B \in \mathcal{B}(\Omega)$, or equivalently, for each $A,B \in \mathcal{S}$.

It is clear that every mixing measure μ in $P(\Omega,T)$ is ergodic.

(4.17) Definition. A stochastic matrix $P = (p_{ij})_{i,j\in S}$ is called primitive if there is an $n \geq 1$ such that P^n is positive, that is, $p_{ij}^{(n)} > 0$ for all i,j.

Every primitive stochastic matrix is irreducible.

(4.18) Theorem. Let $P = (p_{ij})_{i,j\in S}$ be an irreducible stochastic matrix and let $p = (p_i)_{i\in S}$ be the positive stationary distribution of P. Let μ be the (p,P)-Markov measure. Then the following are equivalent:

(i) μ is mixing.

(ii) $\lim_{n\to\infty} p_{ij}^{(n)} = p_j$ for each i and j.

(iii) P is primitive.

Proof. (i) $\Rightarrow$ (ii): Suppose μ is mixing. Let $A = \{x_0=i\}$ and $B = \{x_0=j\}$. Then we have $p_i p_{ij}^{(n)} = \mu(x_0=i,\ x_n=j) = \mu(A\cap T^{-n}B) \to \mu(A)\mu(B)$ $= p_i p_j$ as $n \to \infty$. Since $p_i > 0$, we obtain (ii).

(ii) $\Rightarrow$ (i): Suppose (ii) holds. Let $A = \{x_0=i_0,\ldots,x_k=i_k\}$ and $B = \{x_0=j_0,\ldots,x_m=j_m\}$. Then, for $n > k$, we have

$$\begin{aligned}\mu(A\cap T^{-n}B) &= \mu(x_0=i_0,\ldots,x_k=i_k,\ x_n=j_0,\ldots,x_{n+m}=j_m)\\ &= p_{i_0} p_{i_0 i_1} \cdots p_{i_{k-1} i_k}\, p_{i_k j_0}^{(n-k)}\, p_{j_0 j_1} \cdots p_{j_{m-1} j_m}\end{aligned}$$

so that $\lim_n \mu(A\cap T^{-n}B) = p_{i_0} p_{i_0 i_1} \cdots p_{i_{k-1} i_k}\, p_{j_0} p_{j_0 j_1} \cdots p_{j_{m-1} j_m}$ $= \mu(A)\mu(B)$. Hence μ is mixing.

(ii) $\Rightarrow$ (iii): Suppose (ii) holds. For each j, there is an $n_j \geq 1$ such that $p_{ij}^{(n)} > 0$ for all $n \geq n_j$, $i \in S$. Let $n = \max_j n_j$. Then P^n is positive.

The implication (iii) ⟹ (ii) is the Markov-Kolmogorov theorem. See Chung [1], Feller [1]. □

(4.19) Theorem. Every p-Bernoulli measure, with p positive, is mixing.

We close this section by showing that $B(\Omega,T)$, $M(\Omega,T)$ and $M(\Omega)$ are compact nonconvex sets.

(4.20) Theorem. $B(\Omega,T)$ is a compact nonconvex subset of $P(\Omega,T)$.

Proof. To prove $B(\Omega,T)$ is compact, it is enough to show that it is sequentially compact. Suppose $\{\mu_n\}_{n\geq 1}$ is a sequence of p_n-Bernoulli measures, where $p_n = (p_n(i))_{i\in S}$ is a probability vector for each $n \geq 1$. Then there exists a subsequence $\{p_{n_k}\}_{n_k}$ of $\{p_n\}_n$ such that

$\lim_{n_k\to\infty} p_{n_k}(i) = p_i$ for each i in S. If we define $p = (p_i)_{i\in S}$, then p is a probability vector. Let μ be the p-Bernoulli measure. By (3.3) and (4.4), we obtain $\mu_{n_k} \to \mu$.

To prove $B(\Omega,T)$ is not convex, let $p = (p_i)_{i\in S}$ and $q = (q_i)_{i\in S}$ be two positive probability vectors (that is, $p_i > 0$, $q_i > 0$ for all $i \in S$) such that $p \neq q$. We may assume without loss of generality that $p_0 \neq q_0$, and $p_1 \neq q_1$. Let μ and ν be the p-Bernoulli measure and the q-Bernoulli measure, respectively. Clearly $c\mu + (1-c)\nu \in P(\Omega,T)$ for each $c \in [0,1]$. We claim that $c\mu + (1-c)\nu \in P(\Omega,T) - M(\Omega,T)$ for each $c \in (0,1)$. Let $\tau = c\mu + c'\nu$, where $c' = 1-c$. Suppose $\tau \in M(\Omega,T)$ for some $c \in (0,1)$. If we let $t_i = \tau(x_0=i)$ for $i \in S$, then

$$t_i = c\mu(x_0=i) + c'\nu(x_0=i) = cp_i + c'q_i > 0, \quad \text{and}$$

$$\tau(x_0=i,\ x_1=j) = cp_ip_j + c'q_iq_j > 0 \quad \text{for each} \quad i,j \in S.$$

Let $t_{ij} = \tau(x_1=j|x_0=i)$ for $i,j \in S$. Then $t = (t_i)_{i\in S}$ is a positive stationary distribution of the stochastic matrix $(t_{ij})_{i,j\in S}$. Therefore, the sequence $\{x_n\}_{n\geq 0}$ defined on $(\Omega,\mathcal{B}(\Omega),\tau)$ is a Markov chain with state space S, transition matrix $(t_{ij})_{i,j\in S}$, and stationary initial distribution t, and by (4.5), $\{x_n\}_{n\geq 0}$ is a stationary process on $(\Omega,\mathcal{B}(\Omega),\tau)$. Now we have

$$\tau(x_0=0,\ x_1=1,\ x_2=1) = t_0t_{01}t_{11} = (t_0t_{01})(t_1t_{11})/t_1$$

$$= (cp_0p_1 + c'q_0q_1)(cp_1^2 + c'q_1^2)/(cp_1 + c'q_1), \quad \text{and}$$

$$\tau(x_0=0,\ x_1=1,\ x_2=1) = c\mu(x_0=0,\ x_1=1,\ x_2=1) + c'\nu(x_0=0,\ x_1=1,\ x_2=1)$$

$$= cp_0p_1^2 + c'q_0q_1^2,$$

so that

$$(cp_0p_1 + c'q_0q_1)(cp_1^2 + c'q_1^2) = (cp_0p_1^2 + c'q_0q_1^2)(cp_1 + c'q_1).$$

A simple calculation yields $p_0p_1q_1(q_1-p_1) = p_1q_0q_1(q_1-p_1)$ and thus either $p_0 = q_0$, or $p_1 = q_1$ a contradiction. □

(4.21) <u>Theorem</u>. $M(\Omega)$ is a compact nonconvex subset of $P(\Omega)$.

Proof. By the proof of (4.20), $M(\Omega)$ is not convex. It remains to show that $M(\Omega)$ is sequentially compact. Suppose $\{\mu_n\}_{n\geq 1}$ is a sequence of (p_n, P_n)-Markov measures, where $p_n = (p_n(i))_{i\in S}$ is a probability vector and $P_n = (p_n(i,j))_{i,j\in S}$ is a stochastic matrix for each $n = 1,2,\ldots$. Then each measure μ_n is identified with a vector in I^k, where $I = [0,1]$ and $k = r + r^2$, so that there exists a subsequence $\{\mu_{n_k}\}_{n_k}$ such that

$$\lim_{k\to\infty} p_{n_k}(i) = p_i \geq 0 \quad\text{and}\quad \lim_{k\to\infty} p_{n_k}(i,j) = p_{ij} \geq 0,$$

where $i,j \in S$.

Observe that

$$\sum_{i\in S} p_i = \lim_{k\to\infty} \sum_{i\in S} p_{n_k}(i) = 1 \quad\text{and}\quad \sum_{j\in S} p_{ij} = \lim_{k\to\infty} \sum_{j\in S} p_{n_k}(i,j) = 1$$

for each i.

If we let $p = (p_i)_{i\in S}$ and $P = (p_{ij})_{i,j\in S}$, then p is a probability vector, and P is a stochastic matrix. Let μ denote the (p,P)-Markov measure. Then we obtain

$$\begin{aligned}
\mu(x_0=i_0,\ldots,x_m=i_m) &= p_{i_0} p_{i_0 i_1} \cdots p_{i_{m-1} i_m} \\
&= \lim_{k\to\infty} (p_{n_k}(i_0)\cdot p_{n_k}(i_0,i_1)\cdot\ldots\cdot p_{n_k}(i_{m-1},i_m)) \\
&= \lim_{k\to\infty} \mu_{n_k}(x_0=i_0,\ldots,x_m=i_m)
\end{aligned}$$

for any $m \geq 0$ and any $i_0,\ldots,i_m$ in S, and thus, by (3.3), $\mu_{n_k} \to \mu$. □

(4.22) Theorem. $M(\Omega,T)$ is a compact nonconvex subset of $P(\Omega)$.

Proof. By the proof of (4.20), $M(\Omega,T)$ is not convex. Since $M(\Omega,T) = P(\Omega,T) \cap M(\Omega)$, the set $M(\Omega,T)$ is compact by (3.10) and (4.21). □

(4.23) Remark. Given the probability distributions $p,p_1,p_2,\ldots$ and the stochastic matrices $P,P_1,P_2,\ldots,$ let μ be the (p,P)-Markov measure and let μ_n be the (p_n,P_n)-Markov measure, for $n = 1,2,\ldots.$ If $\lim_{n\to\infty} p_n(i) = p_i$ and $\lim_{n\to\infty} p_n(i,j) = p_{ij}$ for each i and j, then $\mu_n \to \mu$, by (3.3).

Suppose $\mu_n \to \mu$. Then

$$p_i = \mu(x_0=i) = \lim_{n\to\infty} \mu_n(x_0=i) = \lim_{n\to\infty} p_n(i) \quad \text{for each } i,$$

and

$$p_{ij} = \mu(x_1=j\,|\,x_0=i) = \lim_{n\to\infty} \mu_n(x_1=j\,|\,x_0=i) = \lim_{n\to\infty} p_n(i,j)$$

for each j provided $p_i > 0$.

Thus, if $p = (p_i)_{i\in S}$ is a positive probability vector, then $p_{ij} = \lim_{n\to\infty} p_n(i,j)$ for all $i,j \in S$. However, if p is not positive, then there may be stochastic matrices $Q = (q_{ij})_{i,j\in S}$ such that $P \neq Q$, and the pairs (p,Q) also represent the measure μ, that is,

$$\mu(x_0=i_0,\ldots,x_m=i_m) = p_{i_0} q_{i_0 i_1} \cdots q_{i_{m-1} i_m}$$

for any $m \geq 0$ and any $i_0,\ldots,i_m$ in S.

For example, let $\Omega = \prod_0^\infty \{0,1\}$ and $S = \{0,1\}$. Define

$$P = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ 0 & 1 \end{bmatrix}, \qquad Q_c = \begin{bmatrix} c & 1-c \\ 0 & 1 \end{bmatrix}, \quad \text{where } 0 \leq c \leq 1, \quad c \neq \frac{1}{2},$$

$$P_n = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{n} & 1 - \frac{1}{n} \end{bmatrix},$$

$$p = (0,1), \quad p_n = (\frac{2}{2+n}, \frac{n}{2+n}), \quad n = 1,2,\ldots.$$

Let μ_n denote the (p_n,P_n)-Markov measure and let μ denote the (p,P)-Markov measure. Note that each pair (p,Q_c) also represents the measure μ and $\mu_n \to \mu$, although $P_n \not\to Q_c$.

5. Periodic points and orbits

(5.1) Definition. An $\omega \in \Omega$ is called a periodic point of T (or a T-periodic point) with period $n \geq 1$ if $T^n\omega = \omega$ and $T^m\omega \neq \omega$ for all $m \in \{1,2,\ldots,n-1\}$. Let $P^n(T)$ denote the set of all periodic points of T with period n, and let $\tilde{P}(T)$ denote the set of all periodic points of T, i.e., $\tilde{P}(T) = \cup_{n\geq 1} P^n(T)$.

For each $n \geq 1$, let $F(T^n)$ denote the set of all fixed points of T^n, i.e., $F(T^n) = \{\omega : T^n\omega = \omega\}$. For any states $i_0,\ldots,i_{n-1} \in S$, let $[i_0,\ldots,i_{n-1}]$ denote the point $\omega \in \Omega$ such that $\omega_{qn+m} = i_m$ for all $q \geq 0$ and all $m \in \{0,1,\ldots,n-1\}$.

(5.2) Lemma. Let $\omega \in \Omega$ and $n \geq 1$. The following are equivalent:

(i) $T^n\omega = \omega$.

(ii) $\omega_k = \omega_{n+k}$ for all $k \geq 0$.

(iii) $\omega = [i_0,\ldots,i_{n-1}]$ for some $i_0,\ldots,i_{n-1}$ in S.

Proof. (i) $\Rightarrow$ (ii): If $T^n\omega = \omega$, then, for each $k \geq 0$,

$$\omega_k = x_k(\omega) = x_k(T^n\omega) = x_{n+k}(\omega) = \omega_{n+k}.$$

(ii) $\Rightarrow$ (iii): Suppose (ii) holds. Let $i_m = \omega_m$ for $m \in \{0,1,\ldots,n-1\}$. We obtain

$$i_m = \omega_m = \omega_{n+m} = \omega_{2n+m} = \dots = \omega_{qn+m} \quad \text{where} \quad q \in \mathbb{N},$$

so that (iii) holds. It is plain that (iii) $\Rightarrow$ (i). $\square$

(5.3) Lemma. Let n be any positive integer. Then

(i) $P^1(T) = F(T) \subset F(T^n)$, $P^n(T) \subset F(T^n)$.

(ii) $F(T^n) \cap F(T^{n+1}) = F(T)$.

(iii) $F(T^n) = \cup\{P^d(T) : d|n\}$.

(iv) $\tilde{P}(T) = \cup_{n\geq 1} F(T^n)$.

(v) $\tilde{P}(T)$ is a countably infinite dense subset of Ω.

Proof. (i) is obvious.

(ii): By (i) we obtain $F(T) \subset F(T^n) \cap F(T^{n+1})$. To prove (ii), let $\omega = [i_0,\dots,i_n] \in F(T^{n+1})$. If $\omega \in F(T^n)$, then, by (5.2),

$$i_0 = \omega_0 = \omega_{qn} = \omega_{qn-(q-1)(n+1)} = \omega_{n-(q-1)} = i_{n-(q-1)}, \quad 1 \leq q \leq n,$$

so that $\omega = [i_0] \in F(T)$. Thus (ii) holds.

(iii): Suppose $\omega \in F(T^n)$. If we define $d = \min\{m : T^m\omega = \omega\}$, then $1 \leq d \leq n$ and $\omega \in P^d(T)$. Let $n = qd + s$, where $q \geq 0$, $0 \leq s \leq d-1$. It follows that $\omega = T^n\omega = T^s(T^{qd}\omega) = T^s\omega$, so that $s = 0$. Thus $F(T^n) \subset \cup\{P^d(T) : d|n\}$. If $\omega \in P^d(t)$ and $d|n$, then $\omega \in F(T^n)$. Thus (iii) holds.

(iv): If $\omega \in \tilde{P}(T)$, then, by (5.1) together with (i), $\omega \in P^d(T) \subset F(T^d)$ for some $d \geq 1$. On the other hand, if $\omega \in F(T^n)$, then by (iii), $\omega \in P^d(T)$ for some $d \geq 1$ such that $d|n$. Thus (iv) holds.

(v): By (5.2) we have $\text{card } F(T^n) = r^n$ for each $n \geq 1$. It follows from (iv) that $\tilde{P}(T)$ is a countably infinite subset of Ω. Given $\omega \in \Omega$ and $n \geq 1$, let $\eta = [i_0, \ldots, i_{n-1}]$ where $i_k = \omega_k$, $0 \leq k \leq n-1$. Then we have $\eta \in F(T^n)$ and $\eta_k = \omega_k$, $1 \leq k \leq n-1$. Thus $\tilde{P}(t)$ is dense in Ω. □

(5.4) Definition. A point $\omega \in \Omega$ is called T-wandering if there is an open neighbourhood U of ω such that the sets $T^{-n}U$, $n \geq 0$ are pairwise disjoint. Let $W(T)$ denote the set of all T-wandering points. The set $\Omega - W(T)$ is called the non-wandering set for T.

(5.5) Theorem. $\tilde{P}(T) \subset \Omega - W(T) = \Omega$.

Proof. It is easy to see that $\omega \in \Omega - W(T)$ iff for each open neighbourhood of ω, $U \cap T^{-n}U \neq \phi$ for some $n \geq 1$. Suppose $\omega \in \tilde{P}(T)$. By (5.3 iv), $\omega \in F(T^n)$ for some $n \geq 1$. If U is any open neighbourhood of ω, then $T^n\omega = \omega \in U$ and $\omega \in T^{-n}T^n\omega \subset T^{-n}U$, so that $\omega \in U \cap T^{-n}U$. Thus we obtain $\tilde{P}(T) \subset \Omega - W(T)$.

It is plain that $W(T)$ is an open subset of Ω so that $\Omega - W(T)$ is a closed subset of Ω which contains the dense subset $\tilde{P}(T)$. Thus $\Omega - W(T) = \Omega$. □

(5.6) Definition. The set $O_T(\omega) = \{T^n\omega : n \geq 0\}$ is called the T-orbit, or simply the orbit of $\omega \in \Omega$. We shall write $O(\omega)$ for $O_T(\omega)$ and $\overline{O(\omega)}$ for the closure of the orbit $O(\omega)$. The orbit $O(\omega)$ is called a periodic orbit if ω is a periodic point of T.

(5.7) Lemma. For each $\omega \in \Omega$, the orbit $O(\omega)$ is an infinite set iff $T^m\omega \neq T^n\omega$ whenever $m \neq n$ in $\mathbb{Z}_+ = \{0,1,2,\ldots\}$.

Proof. Given $\omega \in \Omega$, define the map $f : \mathbb{Z}_+ \to \Omega$ by $f(n) = T^n\omega$. If f is an injection, then $O(\omega)$ is infinite. Suppose f is not an injection. Then there exists an $m \in \mathbb{Z}_+$ such that $f(m) = f(m+k)$, i.e., $T^m\omega = T^k(T^m\omega)$ for some $k \geq 1$. Let $n = \min\{m \in \mathbb{Z}_+ : T^m\omega \in \tilde{P}(T)\}$ and let $T^n\omega$ have period d. It follows that $T^i\omega \notin \tilde{P}(T)$, $0 \leq i \leq n-1$, and $O(\omega) = \{T^i\omega : 0 \leq i \leq n+d-1\}$. □

From the proof of (5.7) we obtain immediately,

(5.8) Lemma. For each $\omega \in \Omega$, the orbit $O(\omega)$ is a finite set iff there exists a unique $n \in \mathbb{Z}_+$ such that $T^i\omega \notin \tilde{P}(T)$, for $0 \leq i \leq n-1$, and $T^n\omega \in \tilde{P}(T)$. In this case $O(\omega) = \{T^i\omega : 0 \leq i \leq n+d-1\}$ where $T^n\omega \in P^d(T)$.

(5.9) Lemma. For any two periodic points ω, ω' of T, either $O(\omega) = O(\omega')$ or $O(\omega) \cap O(\omega') = \phi$.

Proof. Let $\omega \in P^d(T)$ and $\omega' \in P^{d'}(T)$. If $O(\omega) \cap O(\omega') \neq \phi$, then $T^i\omega = T^j\omega'$ for some i,j, where $0 \leq i \leq d-1$, $0 \leq j \leq d'-1$. It follows that, for each $n \geq 0$, $T^{i+n}\omega = T^{j+n}\omega' \in O(\omega')$ so that

$O(\omega) \subset O(\omega')$. Similarly we obtain $O(\omega') \subset O(\omega)$. Thus $O(\omega) = O(\omega')$ and $d = d'$. □

By (5.3.v) and (5.9) we obtain

(5.10) Theorem. $\tilde{P}(T)$ is a countably infinite union of pairwise disjoint periodic orbits.

(5.11) Lemma. If $\omega \in \tilde{P}(T)$ and $\omega' \notin \tilde{P}(T)$, then either $O(\omega) \cap O(\omega') = \phi$ or $O(\omega')$ is finite and $O(\omega) \subset O(\omega')$.

Proof. Let $\omega \in P^d(T)$. Suppose $O(\omega) \cap O(\omega') \neq \phi$. Then we have $T^i\omega = T^j\omega'$ where $0 \leq i \leq d-1$, $j \geq 1$, so that $O(\omega) \subset O(\omega')$. Also, we obtain $T^d(T^j\omega') = T^d(T^i\omega) = T^i\omega = T^j\omega'$, so that $O(\omega') = \{\omega', T\omega', \ldots, T^{j-1}\omega'\} \cup O(\omega)$. ◫

(5.12) Example. Define three points ω, ω', ω'' in Ω by

$$\omega = (0,1,0,0,1,0,0,1,\ldots),$$

$$\omega' = (1,1,0,0,1,0,0,1,\ldots),$$

$$\omega'' = (0,1,1,1,\ldots).$$

It is plain that ω, ω', ω'' are not periodic points of T. We see readily that $T\omega = T\omega'$ so that

$$O(\omega) = \{\omega\} \cup \{T^n\omega : n \geq 1\}, \quad O(\omega') = \{\omega'\} \cup \{T^n\omega : n \geq 1\},$$

$$O(\omega) \,\Delta\, O(\omega') = \{\omega,\omega'\}.$$

Observe also that $O(\omega'') = \{\omega'',[1]\}$ and $O(\omega) \cap O(\omega'') = O(\omega') \cap O(\omega'') = \phi$.

If the orbit of $\omega \in \Omega$ is finite, then $\overline{O(\omega)} = O(\omega) \subsetneq \Omega$, so that $O(\omega)$ is not dense in Ω. We shall show

(5.13) Theorem. $\{\omega : \overline{O(\omega)} = \Omega\}$ is a dense G_δ.

Proof. Let $\mathcal{U} = \{U_m\}_{m\geq 1}$ denote the countably infinite family of all cylinder sets $Z(i_0,\ldots,i_{n-1})$. Recall that $\mathcal{U}$ is a countable base for Ω. Note that $\overline{O(\omega)} = \Omega$ iff $O(\omega) \cap U_m \neq \phi$ for each m. It is straightforward to show that

$$\{\omega : \overline{O(\omega)} = \Omega\} = \cap_{m\geq 1} \cup_{n\geq 0} T^{-n}U_m.$$

Let $U_m, U_k \in \mathcal{U}$ be arbitrary and let

$$U_m = Z(i_0,\ldots,i_{s-1}), \quad U_k = Z(j_0,\ldots,j_{t-1}).$$

If $G_m = \cup_{n\geq 0} T^{-n} U_m$, then for each $n' \geq t$,

$$U_k \cap G_m \supset U_k \cap T^{-n'}U_m = \{x_0 = j_0,\ldots, x_{t-1} = j_{t-1}, x_{n'} = i_0,\ldots$$
$$\ldots, x_{n'+s-1} = i_{s-1}\} \neq \phi,$$

so that G_m is a dense open subset of Ω. By the Baire category theorem, the set $\{\omega : \overline{O(\omega)} = \Omega\}$ is a dense G_δ. □

6. Atomic ergodic measures

(6.1) Definition. Let μ be a measure in $P(\Omega)$. A point $\omega \in \Omega$ is called an atom of μ if $\mu(\{\omega\}) > 0$. The measure μ is called purely atomic if $\mu(A(\mu)) = 1$, where $A(\mu)$ denotes the set of all atoms of μ. For any $\omega \in \Omega$ a purely atomic measure $\mu \in P(\Omega)$ such that $A(\mu) = O(\omega)$ is called a T-orbit measure, or simply an orbit measure of ω. An orbit measure of $\omega \in \Omega$ is called a periodic orbit measure if ω is a periodic point of T.

Note that the set $A(\mu)$ can easily be seen to be countable, since $\Sigma_{\omega \in A(\mu)} \mu(\{\omega\}) = 1 < \infty$, with $\mu(\{\omega\}) > 0$ for all $\omega \in A(\mu)$. Also, it is plain that there exists a surjection from the set of all purely atomic measures in $P(\Omega)$ onto the set of all probability vectors $p = (p_n)_{n\geq 1}$.

(6.2) Lemma. For each $\omega \in \Omega$, μ is an orbit measure of ω iff there exists a unique positive probability vector $p = (p_n)_{n\geq 0}$ such that $\mu = \Sigma_n p_n \varepsilon_{T^n\omega}$, where the $T^n\omega$ are distinct.

Proof. Suppose μ is an orbit measure of ω. Then for any $B \in \mathcal{B}(\Omega)$ we have

$$\mu(B) = \mu(B \cap O(\omega)) = \Sigma_{T^n\omega \in B} \mu(T^n\omega) = \Sigma_{n\geq 0} \mu(T^n\omega) \cdot \varepsilon_{T^n\omega}(B),$$

and so $\mu = \Sigma_{n\geq 0} \mu(T^n\omega) \cdot \varepsilon_{T^n\omega}$. If we define $p_n = \mu(T^n\omega)$ for $n \geq 0$, then $p = (p_n)_{n\geq 0}$ is a positive probability vector, and $\mu = \Sigma_n p_n \varepsilon_{T^n}$.

On the other hand, for each positive (infinite or finite) probability vector,

$p = (p_n)_{n\geq 0}$, the formula $\mu = \Sigma_n p_n \varepsilon_{T^n\omega}$ defines the orbit measure μ of ω such that $\mu(T^n\omega) = p_n > 0$ for all n. □

(6.3) Lemma. If μ is an orbit measure of $\omega \in \Omega$ for which the orbit $O(\omega)$ is finite, then there exist $p \in (0,1]$, a purely atomic measure μ_1 with a finite number of atoms, and a periodic orbit measure μ_2 such that $\mu = (1-p)\mu_1 + p\mu_2$ and $\mu_1 \perp \mu_2$. The number p and measures μ_1, μ_2 are uniquely determined.

Proof. By (5.8) there exists a unique $n \in \mathbb{Z}_+$ such that $T^i\omega \notin \tilde{P}(T)$ for $i < n$ and $T^n\omega \in \tilde{P}(T)$. Let $T^n\omega \in P^d(T)$ where $d \geq 1$. Let $p_k = \mu(T^k\omega)$ for $0 \leq k \leq n+d-1$ and $p = \Sigma_{k=n}^{n+d-1} p_k$. Note that $0 < p \leq 1$ and $p = 1$ iff $n = 0$. Define

$$\mu_1 = \frac{1}{1-p} \Sigma_{k=0}^{n-1} p_k \varepsilon_{T^k\omega} \quad \text{if } p < 1, \text{ and}$$

$$\mu_2 = \frac{1}{p} \Sigma_{k=n}^{n+d-1} p_k \varepsilon_{T^k\omega} .$$

Then μ_2 is a periodic orbit measure of $T^n\omega$ and $\mu = \mu_2$ when $p = 1$. If $p < 1$, then $\mu = (1-p)\mu_1 + p\mu_2$. Since $A(\mu_1) \cap A(\mu_2) = \phi$, we have $\mu_1 \perp \mu_2$. It is straightforward to show the uniqueness of the decomposition of μ. □

(6.4) Lemma. If $\mu \in P(\Omega)$ is purely atomic, then $T\mu$ also is purely atomic and $A(T\mu) = TA(\mu)$.

Proof. If $A(\mu)$ is finite, say $A(\mu) = \{\omega_i : 1 \le i \le n\}$, then $\mu = \sum_{i=1}^{n} p_i \cdot \varepsilon_{\omega_i}$ where $p_i = \mu(\omega_i)$. Since T is an affine map on $P(\Omega)$ and $T\varepsilon_{\omega_i} = \varepsilon_{T\omega_i}$, we obtain $T\mu = \sum_{i=1}^{n} p_i \, \varepsilon_{T\omega_i}$.

Suppose $A(\mu) = \{\omega_i : i \ge 1\}$ is infinite. Then $\mu = \Sigma_{i\ge 1} p_i \, \varepsilon_{\omega_i}$, where $p_i = \mu(\omega_i)$. Define

$$q_n = \Sigma_{i>n} p_i \ , \qquad \mu_n = \frac{1}{q_n} (\Sigma_{i>n} p_i \, \varepsilon_{\omega_i})$$

where $n = 1,2,\ldots$. It is easily seen that $\mu_n \in P(\Omega)$ and $\mu = \sum_{i=1}^{n} p_i \, \varepsilon_{\omega_i} + q_n \mu_n$ for each $n \ge 1$, so that $T\mu = \sum_{i=1}^{n} p_i \, \varepsilon_{T\omega_i} + q_n \, T\mu_n$. For each $f \in C(\Omega)$, we obtain

$$|\langle f, T\mu\rangle - \langle f, \sum_{i=1}^{n} p_i \, \varepsilon_{T\omega_i}\rangle| = q_n |\langle f, T\mu_n\rangle| \le q_n \|f\| \to 0 \quad \text{as} \quad n \to \infty,$$

where $\langle f,\mu\rangle = \int f d\mu$. Thus $T\mu = \Sigma_{i\ge 1} p_i \, \varepsilon_{T\omega_i}$. □

(6.5) Lemma. Let $\mu \in P(\Omega, T)$ be atomic. If $\omega \in \Omega$ is an atom of μ, then ω is a periodic point of T, $\mu(\omega) = \mu(T^n\omega)$ for all $n \ge 1$, and $\mu(O(\omega)) = d\mu(\omega)$ where d denotes the period of ω.

Proof. We have, for each $n \ge 0$, $T^n\omega \in T^{-1}T(T^n\omega) = T^{-1}(T^{n+1}\omega)$ so that $\mu(T^n\omega) \le \mu(T^{n+1}\omega)$ and $\{\mu(T^n\omega)\}_{n\ge 0}$ is a non-decreasing sequence of positive numbers. Note that $T^n\omega$ is an atom of μ for any $n \ge 0$.

If $O(\omega)$ is infinite, then, by (5.7), $\mu(O(\omega)) = +\infty$, a contradiction. Thus $O(\omega)$ is finite. By (5.8) there exists a unique $m \geq 0$ such that $T^i\omega \notin \tilde{P}(T)$, $0 \leq i \leq m-1$, and $T^m\omega \in \tilde{P}(T)$. Let $T^m\omega \in P^d(T)$. If $m = 0$, then we are done. Suppose $m \geq 1$. We obtain easily that

$$\mu(T^m\omega) = \mu(T^{m+n}\omega) \quad \text{for all} \quad n \in \{1,2,\ldots,d\}; \quad \text{and}$$

$$\mu(O(\omega)) = \sum_{n=0}^{m-1} \mu(T^n\omega) + d\mu(T^m\omega).$$

Since $TO(\omega) \subset O(\omega)$, we also obtain $O(\omega) \subset T^{-1}TO(\omega) \subset T^{-1}O(\omega)$ and thus

$$\sum_{n=0}^{m-1} \mu(T^n\omega) + d\mu(T^m\omega) = \mu(O(\omega)) = \mu(TO(\omega))$$

$$= \sum_{n=1}^{m-1} \mu(T^n\omega) + (d+1)\mu(T^m\omega),$$

so that $\mu(\omega) = \mu(T^m\omega)$. Since $\omega \notin \tilde{P}(T)$, we have $\omega \neq T^d\omega$. It follows that $\{\omega, T^d\omega\} \subset T^{-m}T^m\omega = T^{-m}T^m(T^d\omega)$ and

$$\mu(\omega) < \mu(\omega) + \mu(T^d\omega) \leq \mu(T^m\omega) = \mu(\omega),$$

a contradiction. □

(6.6) Lemma. If $\mu \in P(\Omega,T)$ is atomic, then the set $A(\mu)$ of all atoms of μ is a countable union of pairwise disjoint periodic orbits.

Proof. By (6.5) each $\omega \in A(\mu)$ is a periodic point of T and $A(\mu) = \cup\{O(\omega) : \omega \in A(\mu)\}$. The result follows from (5.9). □□

(6.7) Theorem. Let $\omega \in \Omega$. Then

(i) ω is a periodic point of T iff there exists an orbit measure of ω in $P(\Omega,T)$.

(ii) If an orbit measure of ω exists in $P(\Omega,T)$, then it is the periodic orbit measure of ω in $P(\Omega,T)$.

Proof. Suppose $\omega \in P^d(T)$. If $\mu = \frac{1}{d}\sum_{i=0}^{d-1} \varepsilon_{T^i\omega}$, then μ is an orbit measure of ω and $T\mu = \frac{1}{d}\sum_{i=1}^{d} \varepsilon_{T^i\omega} = \mu$. By (6.5) we obtain the theorem. □

(6.8) Lemma. Let $\mu \in P(\Omega)$ be purely atomic. The following are equivalent:

(i) $\mu = T\mu$;

(ii) The restriction of T to $A(\mu)$ is a bijection of $A(\mu)$ onto itself and $\mu(\omega) = \mu(T\omega)$ for each $\omega \in A(\mu)$.

Proof. (i) $\Rightarrow$ (ii): Suppose $\mu = T\mu$. By (6.5) we show readily that T is a surjection of $A(\mu)$ onto itself and $\mu(\omega) = \mu(T\omega)$ for each $\omega \in A(\mu)$. Suppose $T\omega = T\omega'$ for some $\omega \neq \omega'$ in $A(\omega)$. It follows from (6.6) that $O(\omega) = O(\omega')$, and both ω and ω' have the same period d, so that, by (5.2), $\omega = [i_0,\ldots,i_{d-1}]$ and $\omega' = [j_0,\ldots,j_{d-1}]$, where $i_0,\ldots,i_{d-1},j_0,\ldots,j_{d-1} \in S$. Using $T^n\omega = T^n\omega'$ for $n \geq 1$, we obtain $\omega = \omega'$.

(ii) $\Rightarrow$ (i): Suppose (ii) holds. Since μ is purely atomic, we obtain $\mu = \Sigma_{n\geq 1}\, p_n\, \varepsilon_{\omega_n}$ where $A(\mu) = \{\omega_n : n \geq 1\}$ and $p_n = \mu(\omega_n) > 0$. By (6.4),

$T\mu$ is also purely atomic and $A(T\mu) = TA(\mu) = A(\mu)$. Given $\omega_n \in A(\mu)$, there is a unique $\omega_m \in A(\mu)$ such that $T\omega_m = \omega_n$. It follows that

$$T\mu(\omega_n) = \mu(T^{-1}\{\omega_n\}) = \mu(T^{-1}\{\omega_n\} \cap A(\mu))$$

$$= \mu(\omega_m) = \mu(T\omega_m) = \mu(\omega_n).$$

so that $T\mu = \mu$. □

(6.9) Theorem. Let $\mu \in P(\Omega,T)$. The following are equivalent:

(i) μ is a purely atomic ergodic measure.

(ii) μ is an atomic ergodic measure.

(iii) μ is a periodic orbit measure.

(iv) $\mu = \frac{1}{n}\sum_{i=0}^{n-1} \varepsilon_{T^i\omega}$ for some $\omega \in \Omega$ and $n \geq 1$.

Proof. It is plain that (i) ⟹ (ii). To prove (ii) ⟹ (iii), suppose μ is an atomic ergodic measure. Let $\omega \in A(\mu)$. We have from (6.5) that $\omega \in P^d(T)$ for some $d \geq 1$, $O(\omega) \subset A(\mu)$, and $\mu(O(\omega)) = d\mu(\omega) > 0$. It is easy to see that $O(\omega) \subset T^{-1}O(\omega)$, so that $\mu(T^{-1}O(\omega)\Delta O(\omega)) = 0$. By (3.14) and the ergodicity of μ, we have $\mu(O(\omega)) = 1$, and $O(\omega) = A(\mu)$.
(iii) ⟹ (iv): Suppose $A(\mu) = O(\omega)$ where $\omega \in P^d(T)$. By (6.5) we obtain $1 = \mu(O(\omega)) = d\mu(\omega)$ and $\mu(\omega) = \mu(T^i\omega)$, $0 \leq i \leq d-1$ so that

$$\mu = \frac{1}{d}\sum_{i=0}^{d-1} \varepsilon_{T^i\omega}.$$

(iv) $\Rightarrow$ (i): Suppose $\mu = \frac{1}{n}\sum_{i=0}^{n-1} \varepsilon_{T^i\omega}$, where $\omega \in \Omega$ and $n \geq 1$. Since $\mu = T\mu$, we obtain $T^n\omega = \omega$. If we denote the period of ω by d, then $d|n$ and $\mu = \frac{1}{d}\sum_{i=0}^{d-1} \varepsilon_{T^i\omega}$. Note that μ is purely atomic. To prove the ergodicity of μ, suppose $E = T^{-1}E$ and $\mu(E) > 0$. Then $T^i\omega \in E$ for some $i \in \{0,1,\ldots,d-1\}$. Since $TT^{-1}E = E$ (T is onto by (1.8)), we obtain $TE = E$, so that by induction, $T^nE = E$ for all $n \geq 1$. It follows that $O(\omega) \subset E$ and $\mu(E) = 1$. □

(6.10) <u>Corollary</u>. If $\mu = \frac{1}{d}\sum_{n=0}^{d-1} \varepsilon_{T^n\omega}$ for some $\omega \in P^d(T)$, then $T\mu = \mu$.

(6.11) <u>Theorem</u>. μ is a purely atomic measure in $P(\Omega,T)$ iff μ is a countable convex combination of purely atomic ergodic measures in $P(\Omega,T)$.

<u>Proof</u>. ($\Rightarrow$): Suppose μ is a purely atomic nonergodic measure in $P(\Omega,T)$. By (6.5) and (6.6) there exists a countable set $\{\omega_n\} \subset A(\mu)$ such that ω_n has period d_n and $A(\mu) = \cup O(\omega_n)$ where $O(\omega_n)$ are pairwise disjoint with $\mu(O(\omega_n)) = d_n\mu(\omega_n) > 0$. If we put $p_n = d_n\mu(\omega_n)$, then $0 < p_n$ and $\Sigma_{n\geq 1}\, p_n = 1$. For each n, define the purely atomic ergodic measure $\mu_n = \frac{1}{d_n}\sum_{i=0}^{d_n-1} \varepsilon_{T^i\omega_n}$. Then we have

$$\mu = \Sigma_{n\geq 1}\, \mu(\omega_n)\Big(\sum_{i=0}^{d_n-1} \varepsilon_{T^i\omega_n}\Big) = \Sigma_{n\geq 1}\, p_n\, \mu_n.$$

($\Leftarrow$): Suppose $\mu = \Sigma_{n\geq 1} p_n \mu_n$ where $p_n > 0$, $\Sigma_{n\geq 1} p_n = 1$ and μ_n are purely atomic ergodic measures in $P(\Omega,T)$. By (6.9) there exists a sequence $\{\omega_n\}_{n\geq 1}$ such that $\omega_n \in P^{d_n}(T)$ and $A(\mu_n) = O(\omega_n)$ so that $A(\mu) = \cup_{n\geq 1} A(\mu_n) = \cup_{n\geq 1} O(\omega_n)$. It is clear that μ is a purely atomic measure in $P(\Omega)$. Note also that for all m,n, either $O(\omega_m) = O(\omega_n)$ or $O(\omega_m) \cap O(\omega_n) = \phi$. We can see readily that the restriction of T to $A(\mu)$ is a bijection of $A(\mu)$ onto itself. By (6.8) it remains to show that $\mu(\omega) = \mu(T\omega)$ for each $\omega \in A(\mu)$. Given $\omega \in A(\mu)$, let $\omega \in O(\omega_m)$ and $E = \{n : \omega \in A(\mu_n)\}$. Then we have $A(\mu_n) = O(\omega_m)$ for all $n \in E$, so that $\mu_n = \mu_m$ for all $n \in E$. It follows that

$$\mu_n(\omega) = \mu_m(\omega) = \frac{1}{d_m} \quad \text{for all } n \in E.$$

Since $T\omega \in O(\omega_m)$, we also have

$$\mu_n(T\omega) = \mu_m(T\omega) = \frac{1}{d_m} \quad \text{for all } n \in E.$$

Thus we have

$$\mu(\omega) = \Sigma_{n\in E}\, p_n \mu_n(\omega) = \frac{1}{d_m} \Sigma_{n\in E}\, p_n = \mu(T\omega),$$

so that $\mu = T\mu$. Alternatively, we may show $\mu = T\mu$ by a minor modification of the proof of (6.4). □

7. Atomic ergodic Markov measures

The main result of this section is the following.

(7.1) Theorem. The following are equivalent:

(i) μ is a purely atomic ergodic Markov measure.

(ii) $\mu = \frac{1}{d}\sum_{n=0}^{d-1} \varepsilon_{T^n\omega}$ for some $\omega = [i_0,\ldots,i_{d-1}] \in P^d(T)$

where $i_0,\ldots,i_{d-1}$ are distinct states.

(iii) μ is the Markov measure induced by a cyclic permutation matrix $(p_{jk})_{j,k\in E}$, where $E \subset S$, and the uniform probability vector on E.

The proof is based on the following lemmas.

(7.2) Lemma. Let $E = \{i_0,\ldots,i_{d-1}\}$ be a subset of S and let $p = (p_j)_{j\in E}$ be the uniform probability vector on E. Let $P = (p_{jk})_{j,k\in E}$ be the cyclic permutation matrix such that $p_{i_n j} = \delta_{i_{n+1} j}$ for $0 \le n \le d-1$, $j \in E$, where $i_d = i_0$. If μ is the (p,P)-Markov measure, then

$$\mu = \frac{1}{d}\sum_{n=0}^{d-1} \varepsilon_{T^n\omega} \quad \text{where} \quad \omega = [i_0,\ldots,i_{d-1}].$$

In particular, μ is a purely atomic ergodic Markov measure.

Proof. Clearly $pP = p$, so that $\mu \in P(\Omega,T)$ by (4.5). Let $\omega = [i_0,\ldots,i_{d-1}]$. Then ω has period d, since the states $i_0,\ldots,i_{d-1}$ are distinct. We obtain, for each $n \ge 1$,

$$\mu(x_0=i_0,\ \ldots,\ x_{d-1}=i_{d-1},\ x_d=i_0,\ \ldots,\ x_{nd-1}=i_{d-1})$$

$$= p_{i_0}(p_{i_0 i_1} \cdots p_{i_{d-2} i_{d-1}})^n = p_{i_0} = \frac{1}{d},$$

so that $\mu(\omega) = \frac{1}{d}$. Similarly, we also obtain

$$\mu(T^n\omega) = \frac{1}{d} \text{ for all } n \in \{1,2,\ldots,d-1\}.$$

Thus $\mu(O(\omega)) = 1$. It follows from (6.9) that $\mu = \frac{1}{d}\sum_{n=0}^{d-1} \varepsilon_{T^n}$ and that μ is a purely atomic ergodic Markov measure. □

Note that if μ is the (p,P)-Markov measure, where P contains no cyclic permutation matrix, then μ is nonatomic. Consider the product $p_{i_0 i_1} \cdots p_{i_{r-1} i_r}$ for any $i_0,\ldots,i_r \in S$. The states $i_0,\ldots,i_r$ are not distinct and thus we have $\prod_{\nu=1}^{r} p_{i_{\nu-1} i_\nu} < 1$, for otherwise, P would contain a cyclic permutation matrix. Let $p' = \max_{j_0,\ldots,j_r \in S} p_{j_0 j_1} \cdots p_{j_{r-1} j_r}$. Let $\omega \in \Omega$ be arbitrary, where $x_n(\omega) = i_n$, $n \geq 0$. We obtain

$$\mu(\omega) = \lim_{n\to\infty} \mu(Z(i_0,\ldots,i_{n\cdot r})) = \lim_{n\to\infty} p_{i_0} \prod_{\nu=1}^{n\cdot r} p_{i_{\nu-1} i_\nu}$$

$$\leq \lim_{n\to\infty} p_{j_0} (p')^n = 0.$$

(7.3) <u>Lemma</u>. If $\omega = [i_0,\dots,i_{d-1}]$ where $i_0,\dots,i_{d-1}$ are distinct states, then $\mu = \frac{1}{d}\sum_{n=0}^{d-1} \varepsilon_{T^n\omega}$ is a purely atomic ergodic Markov measure.

<u>Proof</u>. Clearly $T\mu = \mu$. It follows from (6.9) that μ is a purely atomic ergodic measure. It remains to show that μ is Markov. Consider the process $\{x_n\}_{n\geq 0}$ on the probability space $(\Omega,\mathcal{B}(\Omega),\mu)$. It is plain that $\{x_n\}_{n\geq 0}$ is stationary. Let $E = \{i_0,\dots,i_{d-1}\}$ and let $i_q = i_t$ for $q \equiv t(\mathrm{mod}\ d)$. We obtain, for each $j \in E$,

$$\mu(x_0=j) = \sum_{n=0}^{d-1} \mu((x_0=j) \cap T^n\omega) = \frac{1}{d}\sum_{n=0}^{d-1} \delta_{i_n j} = \frac{1}{d}\ .$$

Define the probability vector $p = (p_j)_{j\in E}$ by $p_j = \frac{1}{d}$. We obtain, for each $i_m \in E$,

$$\mu(x_0=i_m,\ x_1=k) = \sum_{n=0}^{d-1} \mu((x_0=i_m,\ x_1=k) \cap T^k\omega)$$

$$= \frac{1}{d}\sum_{n=0}^{d-1} \delta_{i_n i_m}\ \delta_{i_{n+1} k} = \frac{1}{d}\ \delta_{i_{m+1} k}$$

for all $k \in E$ so that

$$\mu(x_1=k \mid x_0=i_m) = \delta_{i_{m+1} k}$$

for all $k \in E$. Define the stochastic matrix $P = (p_{jk})_{i,k\in E}$ by

$$p_{jk} = \mu(x_1=k \mid x_0=j) \ .$$

Note that, for each $i_m \in E$,

$$p_{i_m k} = \delta_{i_{m+1} k} \quad \text{for all} \quad k \in E,$$

that is, P is a $d\times d$ cyclic permutation matrix. It is easily seen that $pP = p$. We show readily that, for each $i_m \in E$ and $n \geq 1$,

$$\mu(x_{n+1}=k \mid x_0=i_m, \ x_1=i_{m+1}, \ \ldots, \ x_n=i_{m+n})$$

$$= \delta_{i_{m+n+1} k} = p_{i_{m+n} k}$$

for all $k \in E$. Thus $\{x_n\}_{n\geq 0}$ is a Markov chain with state space E, transition matrix P and stationary initial distribution p, so that μ is the (p,P)-Markov measure. □

Proof of the theorem. In view of (7.2) and (7.3), it remains to show the implication (i) ⟹ (ii). Suppose that μ is a purely atomic ergodic Markov measure, or equivalently, by (6.9),

$$\mu = \frac{1}{d} \sum_{n=0}^{d-1} \varepsilon_{T^n \omega} \quad \text{for some} \quad \omega \in P^d(T) \ .$$

We see readily that if $d=1$ or 2, then $\omega = [i]$ for some $i \in S$, or $\omega = [i,j]$ for some $i,j \in S$ with $i \neq j$. Thus it is enough to show that if $\omega = [i_0,\ldots,i_{d-1}]$ where $d \geq 3$ and $i_0,\ldots,i_{d-1}$ are not distinct, then the measure μ is not Markov. Let E be the set of distinct states in $[i_0,\ldots,i_{d-1}]$. Then card $E \geq 2$. Let $i_q = i_t$ for $q \equiv t \pmod d$. For each $j \in E$ define

$$m_j = \sum_{n=0}^{d-1} \delta_{ji_n} .$$

It follows that $1 \leq m_j \leq d-1$ for each $j \in E$, $\sum_{j\in E} m_j = d$ and $2 \leq m_j \leq d-1$ for some $j \in E$.

Consider the process $\{x_n\}_{n\geq 0}$ on the probability space $(\Omega,\mathcal{B}(\Omega),\mu)$. Since $T\mu = \mu$, the process $\{x_n\}_{n\geq 0}$ is stationary. We obtain, for each $j \in E$,

$$\mu(x_0=j) = \sum_{n=0}^{d-1} \mu((x_0=j) \cap T^n\omega) = \frac{1}{d}\sum_{n=0}^{d-1} \delta_{ji_n} = \frac{m_j}{d}$$

and $\mu(x_0\in E) = 1$. Note that, for all $n \geq 1$, $\mu(x_n=j) = \mu(x_0=j)$ for each $j \in E$, and $\mu(x_n\in E) = 1$. Define the probability vector $p = (p_j)_{j\in E}$ by $p_j = \frac{m_j}{d}$.. For any $j,k \in E$ we obtain

$$\mu(x_0=j,\ x_1=k) = \sum_{n=0}^{d-1} \mu((x_0=j,\ x_1=k) \cap T^n\omega)$$

$$= \frac{1}{d}\sum_{n=0}^{d-1} \delta_{ji_n}\,\delta_{ki_{n+1}} ,$$

so that

$$\mu(x_1=k \mid x_0=j) = \frac{d}{m_j}\,\mu(x_0=j,\ x_1=k)$$

$$= \frac{1}{m_j} \sum_{n=0}^{d-1} \delta_{ji_n}\,\delta_{ki_{n+1}}\ ,$$

$$\sum_{k \in E} \mu(x_1=k \mid x_0=j) = \frac{d}{m_j}\,\mu(x_0=j) = 1.$$

Define the stochastic matrix $P = (p_{jk})_{j,k \in E}$ by

$$p_{jk} = \mu(x_1=k \mid x_0=j)\ .$$

Note that $pP = p$ and that $p_{jk} = \mu(x_{n+1}=k \mid x_n=j)$ for all $n \geq 0$ and all $j,k \in E$.

To prove that μ is not Markov, it is enough to show that the process $\{x_n\}_{n \geq 0}$ does not have the Markov property. For each $j \in E$, let $A_j = \{n : j = i_n = i_{n+1},\ 0 \leq n \leq d-1\}$ and let

$$\tau_j = \begin{cases} \min A_j & \text{if } A_j \neq \phi, \\ \infty & \text{if } A_j = \phi. \end{cases}$$

Suppose that there is a state $j \in E$ such that $\tau_j < \infty$, i.e., $0 \leq \tau_j \leq d-1$. (Note that this condition is always satisfied when $d=3$.) We obtain $p_{jj} \geq \frac{1}{m_j} > 0$ and

$$\mu(x_0=x_1=\ldots=x_{d-1}=j) = 0 < \frac{m_j}{d}\left(\frac{1}{m_j}\right)^{d-1} \le p_j(p_{jj})^{d-1}$$

so that μ is not Markov.

Suppose that $\tau_j = \infty$ for all $j \in E$. (In this case $d \ge 4$.) We may assume without loss of generality that

$$i_0 \ne i_{d-1} \quad \text{and} \quad m_j \le m_{i_0} \quad \text{for all} \quad j \in E.$$

Let $m = m_{i_0}$ and let $\{t_k\}_{1\le k\le m+1}$ be such that

$$0 = t_1 < t_2 < \ldots < t_m < t_{m+1} = d \quad \text{and}$$

$$i_0 = i_{t_k} \quad \text{for all} \quad k.$$

Define $s = \min\{t_{k+1}-t_k : 1 \le k \le m\}$. Then $2 \le s \le \frac{d}{2}$ and $4 \le ms \le d$. Let $s = t_{k_o+1}-t_{k_o}$ where $1 \le k_o \le m$. If we write

$$j_\nu = i_{t_{k_o}+\nu} \quad \text{for} \quad 0 \le \nu \le s-1,$$

then $j_0=i_0$ and

$$\prod_{\nu=0}^{s-1} p_{j_\nu} p_{j_{\nu+1}} > 0 .$$

We have then,

$$\mu(x_{qs+\nu}=j_\nu,\quad 0 \le q \le m-1,\quad 0 \le \nu \le s-1;\quad x_{ms}=j_0)$$

$$= 0 < p_{i_0}\left(\prod_{\nu=0}^{s-1} p_{j_\nu j_{\nu+1}}\right)^m ,$$

(otherwise, either $\omega \in P^s(T)$, $s < d$, or $m_{i_0} = m+1$) so that μ is not Markov. ◻

(7.4) Lemma. For any $d > r$ and any $\omega \in P^d(T)$, $\mu = \frac{1}{d}\sum_{n=0}^{d-1} \varepsilon_{T^n\omega}$ is a purely atomic ergodic non-Markov measure.

Proof. Suppose that $d > r$ and $\omega \in P^d(T)$. It follows that $\omega = [i_0,\ldots,i_{d-1}]$ where $i_0,\ldots,i_{d-1}$ are not distinct states. Using (6.9) and (7.1) we obtain the result. ◻

We obtain, from (7.1) and (7.4), the following theorem.

(7.5) Theorem. Let μ be a purely atomic ergodic measure. Then the following are equivalent:

(i) μ is non-Markov.

(ii) μ is the periodic orbit measure of some $\omega \in P^d(T)$ provided that either $d > r$ or $3 \le d \le r$ and $\omega = [i_0,\ldots,i_{d-1}]$ where $i_0,\ldots,i_{d-1}$ are not distinct states.

By (7.1) we also have the following theorem.

(7.6) Theorem. Let μ be the (p,P)-Markov measure. Then the following are equivalent:

(i) μ is nonatomic ergodic.

(ii) μ is ergodic and P is not a cyclic permutation.

8. Nonatomic measures on the unit interval

We now state and prove some well-known results (see Billingsley [2], Halmos [1], Hewitt-Stromberg [1], Royden [1]) that will be used in the next section. Let I denote the closed unit interval and m, Lebesgue measure on I.

(8.1) Lemma. Let $f : I \to I$ be a continuous nondecreasing function. Then f is a surjection iff $f(0) = 0$ and $f(1) = 1$.

The proof is simple and is omitted.

(8.2) Lemma. Let $f : I \to I$ be a continuous nondecreasing function such that $f(0) = 0$ and $f(1) = 1$. Then

(i) For each $[a,b] \subset I$ with $a < b$, $f^{-1}([a,b])$ is a nondegenerate interval in I, i.e., $f^{-1}([a,b]) = [s,t]$ for some $s,t \in I$, $s < t$.

(ii) For $x \in I$, $f^{-1}(\{x\})$ is either a singleton or a non-degenerate closed interval in I.

(iii) The set $E = \{x \in I : f^{-1}(\{x\})$ is a nondegenerate interval$\}$ is either empty or countable.

(iv) The map $f : I - f^{-1}(E) \to I - E$, where E is the set defined in (iii), is a strictly increasing homeomorphism.

Proof. (i) Let $0 \le a < b \le 1$. By (8.1) there exist points x,y in I such that $f(x) = a < b = f(y)$, so that $x < y$ and $x,y \in f^{-1}([a,b])$. For any $u,v \in f^{-1}([a,b])$ with $u < v$, we have

$$a \le f(u) \le f(z) \le f(v) \le b \quad \text{for all} \quad z \in [u,v]$$

so that $[u,v] \subset f^{-1}([a,b])$. Thus $f^{-1}([a,b])$ is a connected subset of I. Since $f^{-1}([a,b])$ is also closed, $f^{-1}([a,b])$ is a nondegenerate closed subinterval of I.

(ii): By the argument used above we see that, for each $x \in I$, $f^{-1}(\{x\})$ is a nonempty closed connected subset of I, so that it is either a singleton or a nondegenerate closed subinterval of I.

(iii): Suppose $E \neq \phi$. It follows from (ii) that, for any $x,y \in E$ with $x \neq y$, $f^{-1}(\{x\})$ and $f^{-1}(\{y\})$ are disjoint nondegenerate closed subintervals of I. We show readily that $f^{-1}(E)$ consists of countable pairwise disjoint nondegenerate closed subintervals of I so that E must be countable.

(iv): By (iii), both $I - f^{-1}(E)$ and $I - E$ are nonempty Borel subsets of I. Since f is a surjection on I, we obtain

$$f(I - f^{-1}(E)) = f\, f^{-1}(I-E) = I - E.$$

Suppose that $f(x) = f(y)$ for some $x,y \in I - f^{-1}(E)$, $x < y$. Then we have $f(x)$, $f(y) \in I - E$ so that $x = f^{-1}f(x) = f^{-1}f(y) = y$, a contradiction. Thus $f : I - f^{-1}(E) \to I - E$ is a strictly increasing bijection. It is plain that f is continuous on $I - f^{-1}(E)$. To prove that f is an open map on $I - f^{-1}(E)$, let $(a,b) \cap (I - f^{-1}(E)) = A$, $a < b$. Then we have $f(A) = (f(a),f(b)) \cap I - E$, so that (iv) holds. □

Let $P(I)$ be the set of all probability measures on $(I,\mathcal{B}(I))$ where $\mathcal{B}(I)$ denotes the σ-algebra of Borel sets in I.

(8.3) <u>Lemma</u>. Let $\nu \in P(I)$ and let f_ν be the <u>distribution function</u> of ν defined by $f_\nu(x) = \nu([0,x))$, $x \in I$. Then

(i) f_ν is a nondecreasing function on I with $f_\nu(0) = 0$, $f_\nu(1) \leq 1$ and is continuous on the left in $(0,1]$.

(ii) f is continuous on I with $f_\nu(0) = 0$, $f_\nu(1) = 1$ iff ν is nonatomic.

<u>Proof</u>. Let $f = f_\nu$. It is plain that $f(0) = 0 \leq f(x) \leq f(y) \leq 1$ for $0 \leq x \leq y \leq 1$.

(i): To prove the left continuity of f, let $x \in (0,1]$ and let $\{x_n\}_{n\geq 1}$ be a nondecreasing sequence in $(0,1]$ such that $\lim_n x_n = x$. It follows that

$$f(x) = \nu(\bigcup_n [0,x_n)) = \lim_n \nu([0,x_n)) = \lim_n f(x_n).$$

(ii): Let $x \in [0,1]$ and let $\{x_n\}_{n\geq 1}$ be a nonincreasing sequence in $[0,1)$ such that $\lim_n x_n = x$. Then we have $\nu(\{x\}) = \nu(\bigcap_n [x,x_n)) =$ $= \lim_n \nu([x,x_n)) = \lim_n (f(x_n)-f(x)) = f(x^+) - f(x)$, so that $\nu(\{x\}) = 0$ iff f is continuous at x. On the other hand we also have, $f(1) = 1$ iff $\nu(\{1\}) = 0$. □

(8.4) *Theorem.* Let $\nu \in P(I)$ be nonatomic and let $f_\nu(x) = \nu([0,x))$, $x \in I$. Then f_ν is a measure-preserving continuous nondecreasing surjection from $(I,\mathcal{B}(I),\nu)$ onto $(I,\mathcal{B}(I),m)$.

Proof. Let $f = f_\nu$. By the preceding lemmas, f is a continuous nondecreasing surjection of I onto itself, so that it is Borel measurable. It remains to show that $f\nu = m$, i.e., $\nu(f^{-1}(A)) = m(A)$ for each Borel set A. It follows from (8.2.i) that, for each $[a,b] \subset I$ with $a < b$,

$$f^{-1}([a,b]) = [s,t] \quad \text{and} \quad [a,b] = [f(s),f(t)] \quad \text{for some} \quad 0 \le s < t \le 1,$$

so that $m([a,b]) = f(t) - f(s) = ([s,t)) = \nu([s,t]) = \nu(f^{-1}[a,b])$. By a general form of the unique extension theorem (see Blumenthal and Getoor [1], Halmos [1], Royden [1]) we obtain $f\nu = m$. □

Let f be a nondecreasing function, $f : I \to I$. f is called *singular* if $f' = 0$ m-a.e.. By a well-known theorem, f is absolutely continuous iff f is an indefinite integral with respect to m. We quote without proof the following. (See Billingsley [2], Hewitt-Stromberg [1].)

(8.5) *Theorem.* Let $\nu \in P(I)$ and let $f_\nu(x) = \nu([0,x))$. Then f_ν is absolutely continuous (f_ν is singular) iff $\nu \ll m$ ($\nu \perp m$).

(8.6) *Theorem.* Let $\nu \in P(I)$ be nonatomic, $f_\nu(x) = \nu([0,x))$, $x \in I$. Then

(i) There exist two uniquely determined nondecreasing continuous functions g_1 and g_2 on I whose sum is f_ν, such that $g_1(0) = g_2(0) = 0$,

g_1 is absolutely continuous and g_2 is singular.

(ii) Let ν_1 and ν_2 be the Borel measures defined by

$$\nu_1([0,x)) = g_1(x), \quad \nu_2([0,x)) = g_2(x), \quad \text{where} \quad x \in I.$$

Then $\nu = \nu_1 + \nu_2$ with $\nu_1 \ll m$, $\nu_2 \perp m$. This decomposition is the Lebesgue decomposition of ν.

Proof. (i): Let $f = f_\nu$. It is well-known (see Hewitt-Stromberg [1], Royden [1]) that f' exists m-a.e., and $0 \leq \int_x^y f'(t)dt \leq f(y) - f(x)$ whenever $x < y$ in I. If we define the functions g_1 and g_2 by

$$g_1(x) = \int_0^x f'(t)dt, \quad g_2(x) = f(x) - g_1(x)$$

where $x \in I$, then g_1 and g_2 have the desired properties.

To prove the uniqueness of the decomposition of f, let $f = g_1 + g_2$ and $f = h_1 + h_2$ be two such decompositions of f. Then $g_1 - h_1$ is absolutely continuous with $g_1' - h_1' = 0$ m-a.e., so that $g_1(x) - h_1(x) = g_1(0) - h_1(0) = 0$ for all $x \in I$. Thus we obtain $g_1 = h_1$ and $g_2 = h_2$.

(ii): We have that $\nu_1 \ll m$ and $\nu_2 \perp m$, by (8.5).

(8.7) Corollary. Let $\nu \in P(I)$ and $f_\nu(x) = \nu([0,x))$, $x \in I$. Then

$$\nu = m \text{ iff } f_\nu(x) = x \text{ for all } x \in I.$$

(8.8) Theorem. For each nonatomic $\nu \in P(I)$, there exist Borel sets I_1 and I_2 in I such that $\nu(I_1) = m(I_2) = 1$ and a measure-preserving homeomorphism from $(I_1,\mathcal{B}(I_1),\nu)$ onto $(I_2,\mathcal{B}(I_2),m)$. (In particular, $(I,\mathcal{B}(I),\nu)$ and $(I,\mathcal{B}(I),m)$ are isomorphic. See section 9.)

Proof. Let $f(x) = \nu([0,x))$ for $x \in I$. By (8.4), f is a measure-preserving continuous nondecreasing surjection from $(I,\mathcal{B}(I),\nu)$ onto $(I,\mathcal{B}(I),m)$. Let $E = \{x \in I : f^{-1}(x) \text{ is a nondegenerate interval}\}$, $I_1 = I - f^{-1}(E)$ and $I_2 = I - E$. It follows from (8.2) and (8.4) that both I_1 and I_2 are Borel sets and $\nu(I_1) = m(I_2) = 1$. Denote by g the restriction of f to I_1. By (8.2), g is a strictly increasing measure-preserving homeomorphism from $(I_1,\mathcal{B}(I_1),\nu)$ onto $(I_2,\mathcal{B}(I_2),m)$. ◻

9. Relationship between nonatomic measures on Ω and nonatomic measures on I

(9.1) Definition. Define the subsets Ω_0, Ω_{r-1} and Ω^* of Ω by

$$\Omega_0 = \{\omega : \omega \neq [0],\ \exists n \geq 1 \text{ such that } \omega_m = 0 \text{ for all } m \geq n.\}$$

$$\Omega_{r-1} = \{\omega : \omega \neq [r-1],\ \exists n \geq 1 \text{ such that } \omega_m = r-1 \text{ for all } m \geq n.\}$$

$$\Omega^* = \Omega - \Omega_{r-1}.$$

Let I_o be the set of all r-adic rational numbers in $(0,1)$, i.e.,

$$I_o = \{x : 0 < x < 1,\ x = \frac{a}{r^n} \text{ where } a,n \in \mathbb{N}\}.$$

Let $\varphi : \Omega \to I$ be defined by

$$\varphi(\omega) = \sum_{n=0}^{\infty} \frac{\omega_n}{r^{n+1}},$$

and let ψ be the restriction of the map φ to Ω^*.

(9.2) Lemma. (i) The map φ is a continuous surjection from Ω onto I.

(ii) The restriction of the map φ to $\Omega - (\Omega_0 \cup \Omega_{r-1})$ is a homeomorphism of $\Omega - (\Omega_0 \cap \Omega_{r-1})$ onto $I - I_o$.

Proof. (i): It is plain that $0 \le \varphi(\omega) \le 1$ for all $\omega \in \Omega$, $\varphi(\omega) = 0$ iff $\omega = [0]$, and $\varphi(\omega) = 1$ iff $\omega = [r-1]$. Suppose $\frac{a}{r^n} \in I_o$ where $r \nmid a$. Applying the division algorithm n times, we obtain a unique sequence $(i_0, i_1, \ldots, i_{n-1})$ in S with $i_{n-1} \ge 1$ such that

$$a = i_0 r^{n-1} + i_1 r^{n-2} + \ldots + i_{n-1}$$

so that $\frac{a}{r^n} = \frac{i_0}{r} + \frac{i_1}{r^2} + \ldots + \frac{i_{n-1}}{r^n}$. Define $\omega = (\omega_m)$ and $\omega' = (\omega'_m)$ by

$$\omega_m = i_m \quad \text{for} \quad 0 \le m \le n-1, \quad \omega_m = 0 \quad \text{for all} \quad m \ge n$$

and

$$\omega'_m = i_m \quad \text{for} \quad 0 \le m \le n-2, \quad \omega'_{n-1} = i_{n-1} - 1,$$

$$\omega'_m = r-1 \quad \text{for all} \quad m \ge n.$$

Then we have $\omega \in \Omega_0$, $\omega' \in \Omega_{r-1}$ and $\varphi(\omega) = \varphi(\omega') = \frac{a}{r^n}$.

Suppose $x \in (0,1) - I_o$. Since $[0,1) = \bigcup_{i=0}^{r-1} [\frac{i}{r}, \frac{i+1}{r})$, there is a unique $i_0 \in S$ such that

$$x \in \left(\frac{i_0}{r}, \frac{i_0+1}{r}\right) \quad \text{or} \quad 0 < x - \frac{i_0}{r} \in [0, \frac{1}{r}) = \bigcup_{i=0}^{r-1} \left[\frac{i}{r^2}, \frac{i+1}{r^2}\right).$$

Next, there exists a unique $i_1 \in S$ such that

$$x - \frac{i_0}{r} \in \left[\frac{i_1}{r^2}, \frac{i_1+1}{r^2}\right) \quad \text{or} \quad 0 < x - \frac{i_0}{r} - \frac{i_1}{r^2} \in [0, \frac{1}{r^2})$$

By a repetition of the argument just used, we obtain a unique sequence $\{i_n\}_{n\geq 0}$ in S such that

$$0 < x - \sum_{k=0}^{n} \frac{i_k}{r^{k+1}} < \frac{1}{r^{n+1}} \quad \text{for all} \quad n > 0,$$

so that $x = \sum_{n=0}^{\infty} \frac{i_n}{r^{n+1}}$. If we define $\omega = (i_n)_{n\geq 0}$, then $\omega \in \Omega-(\Omega_0 \cup \Omega_{r-1})$ and $\varphi(\omega) = x$. We see readily that $\varphi(\omega) \neq \varphi(\omega')$ whenever $\omega \neq \omega'$ in $\Omega-(\Omega_0 \cup \Omega_{r-1})$.

Recall that, for each $n \geq 0$, the map $x_n : \Omega \to S$ defined by $x_n(\omega) = \omega_n$ is continuous. It follows from the Weierstrass M-test that the map $\varphi(\omega) = \sum_{n=0}^{\infty} \frac{x_n(\omega)}{r^{n+1}}$ is continuous on Ω. Thus (i) holds.

(ii): The preceding argument shows that the restriction of the map φ to $\Omega-(\Omega_0 \cup \Omega_{r-1})$ is a continuous bijection of $\Omega-(\Omega_0 \cup \Omega_{r-1})$ onto $I - I_o$. We see readily that, for any $i_0, i_1, \ldots, i_n \in S$,

$$\varphi(\{x_0=i_0,\ldots,x_n=i_n\} \cap (\Omega-(\Omega_0 \cup \Omega_{r-1}))$$

$$= (u, u + \frac{1}{r^{n+1}}) \cap (I - I_o)$$

where $u = \sum_{k=0}^{n} \frac{i_k}{r^{k+1}}$, so that $\varphi : \Omega-(\Omega_0 \cup \Omega_{r-1}) \to I - I_o$ is an open map.

Thus (ii) holds. □

(9.3) Lemma. The map $\psi : \Omega^* \to I$ is a continuous bijection such that ψ^{-1} is Borel measurable.

Proof. By (9.2) the map ψ is a continuous bijection from Ω^* onto I. Since we have, for any $i_0,\dots,i_n \in S$,

$$\psi(x_0=i_0,\dots,x_n=i_n) = \varphi(\{x_0=i_0,\dots,x_n=i_n\} \cap (\Omega-\Omega_{r-1}))$$

$$= [u, u + \frac{1}{r^{n+1}})$$

where $u = \sum_{k=0}^{n} \frac{i_k}{r^{k+1}}$, the map $\psi^{-1} : I \to \Omega^*$ is Borel measurable. □

(9.4) Definition. Let $X_i = (X_i, \mathcal{B}_i, \mu_i)$, $i = 1,2$, be probability spaces, each with a measure-preserving transformation $T_i : X_i \to X_i$.

For any measurable transformation $f : X_1 \to X_2$, define the probability measure $f\mu_1$ on X_2 by $(f\mu_1)(B) = \mu_1(f^{-1}(B))$ for each $B \in \mathcal{B}_2$.

The probability spaces X_1 and X_2 are said to be isomorphic if there exists $Y_i \in \mathcal{B}_i$ with $\mu_i(Y_i) = 1$, where $i = 1,2$, and an invertible measure-preserving transformation $f : Y_1 \to Y_2$ (i.e., f is bijective, and f, f^{-1} are measurable and measure-preserving). The space Y_i is assumed to be equipped with the σ-algebra $Y_i \cap \mathcal{B}_i$ and the restriction of the measure μ_i to this σ-algebra.

The transformation T_1 is said to be <u>isomorphic</u> to T_2 if there exist $Y_i \in \mathcal{B}_i$ with $\mu_i(Y_i) = 1$ such that $T_i(Y_i) \subset Y_i$, $i = 1,2$, and an invertible measure-preserving transformation $f : Y_1 \to Y_2$ such that $fT_1(x) = T_2 f(x)$ for each $x \in X_1$.

(9.5) <u>Definition</u>. Let $P(\Omega^*) = \{\mu \in P(\Omega) : \mu(\Omega^*) = 1$, or equivalently, $\mu(\Omega_{r-1}) = 0\}$, $P(\Omega^*,T) = \{\mu \in P(\Omega^*) : T\mu = \mu\}$ and $E(\Omega^*,T) =$ $= \{\mu \in E(\Omega,T) : \mu(\Omega^*) = 1\}$.

(9.6) <u>Lemma</u>. Let φ and ψ be as in (9.1) and let $\mu \in P(\Omega)$. Then we have

(i) $\varphi\mu \in P(I)$, $\psi\mu$ is a measure on $(I,\mathcal{B}(I))$, and $\psi\mu \le \varphi\mu$,

(ii) $\psi\mu = \varphi\mu$ iff $\mu \in P(\Omega^*)$.

<u>Proof</u>. (i): It is plain that $\varphi\mu \in P(I)$ and $\psi\mu$ is a measure on $(I,\mathcal{B}(I))$. For each $B \in \mathcal{B}(I)$, we have

$$\varphi^{-1}(B) = (\varphi^{-1}(B) \cap \Omega^*) \cup (\varphi^{-1}(B) \cap \Omega_{r-1}) \supset \varphi^{-1}(B) \cap \Omega^* = \psi^{-1}(B)$$

so that $\mu(\varphi^{-1}(B)) \ge \mu(\psi^{-1}(B))$.

(ii) If $\mu \in P(\Omega^*)$, then $\mu(\varphi^{-1}(B)) = \mu(\varphi^{-1}(B) \cap \Omega^*) + \mu(\varphi^{-1}(B) \cap \Omega_{r-1})$ $= \mu(\varphi^{-1}(B) \cap \Omega^*) = \mu(\psi^{-1}(B))$, for each $B \in \mathcal{B}(I)$, so that $\varphi\mu = \psi\mu$. If $\mu \in P(\Omega) - P(\Omega^*)$, then $\mu(\Omega_{r-1}) > 0$ and $\mu(\varphi^{-1}(I)) = \mu(\Omega^*) + \mu(\Omega_{r-1})$ $> \mu(\Omega^*) = \mu(\psi^{-1}(I))$, so that $\psi\mu \ne \varphi\mu$. □

(9.7) Theorem. The mapping $\psi : P(\Omega^*) \to P(I)$ defined by

$$(\psi\mu)(B) = \mu(\psi^{-1}(B)), \quad \text{where} \quad \mu \in P(\Omega^*), \quad B \in \mathcal{B}(I),$$

is a bijection.

Proof. Let $\nu \in P(I)$. By (9.3) we have $\psi^{-1}\nu \in P(\Omega^*)$, so that $\psi(\psi^{-1}\nu) \in P(I)$. Since $\psi(\psi^{-1}(B)) = B$ for each $B \in \mathcal{B}(I)$, we also have

$$[\psi(\psi^{-1}\nu)](B) = \nu(\psi(\psi^{-1}(B)) = \nu(B)$$

for each $B \in \mathcal{B}(I)$ so that $\psi(\psi^{-1}\nu) = \nu$.

On the other hand, if $\mu \in P(\Omega^*)$, then $\psi^{-1}(\psi\mu) \in P(\Omega^*)$ and $[\psi^{-1}(\psi\mu)](A) = \mu(\psi^{-1}(\psi(A))) = \mu(A)$ for each $A \in \mathcal{B}(\Omega^*)$, so that $\psi^{-1}(\psi\mu) = \mu$. Suppose that $\psi\mu = \psi\mu'$ where $\mu,\mu' \in P(\Omega^*)$. It follows that $\mu = \psi^{-1}(\psi\mu) = \psi^{-1}(\psi\mu') = \mu'$. □

(9.8) Definition. Let $NA(\Omega) = \{\mu \in P(\Omega) : \mu$ is nonatomic$\}$ and $NA(I) = \{\nu \in P(I) : \nu$ is nonatomic$\}$.

(9.9) Theorem. Let φ and ψ be as in (9.1). Then we have

(i) $NA(\Omega) \subset P(\Omega^*)$,

(ii) the restriction of the mapping ψ to $NA(\Omega)$ is a bijection between $NA(\Omega)$ and $NA(I)$, and $\varphi\mu = \psi\mu$ for all $\mu \in NA(\Omega)$.

Proof. (i): Suppose $\mu \in NA(\Omega)$. Since the set Ω_{r-1} is a countably infinite subset of Ω, we obtain $\mu(\Omega_{r-1}) = 0$ so that $\mu \in P(\Omega^*)$.

(ii): Let $\mu \in NA(\Omega)$. For each $x \in I$, $\varphi^{-1}(\{x\})$ is either a one-point set in $\Omega-(\Omega_0 \cup \Omega_{r-1})$ or a two-point set in $\Omega_0 \cup \Omega_{r-1}$ so that $(\varphi\mu)(\{x\}) = \mu(\varphi^{-1}(\{x\})) = 0$, i.e., $\varphi\mu \in NA(I)$. By (9.6) we obtain $\varphi\mu = \psi\mu$. Given $\nu \in NA(I)$, we have, by (9.7), $\psi^{-1}\nu \in P(\Omega^*)$ and

$$(\psi^{-1}\nu)(\{\omega\}) = \nu(\{\psi(\omega)\}) = 0 \quad \text{for each} \quad \omega \in \Omega^*$$

so that $\psi^{-1}\nu \in NA(\Omega)$. Thus (ii) follows from (9.7). □

(9.10) Definition. Define the r-adic transformation T' on I by

$$T'(x) = rx \pmod 1 \quad \text{for} \quad 0 \le x < 1 \quad \text{and} \quad T'(1) = 1.$$

Let $P(I,T') = \{\nu \in P(I) : T'\nu = \nu\}$.

(9.11) Lemma. $\psi(T\omega) = T'(\psi(\omega))$ for each $\omega \in \Omega^*$.

Proof. It is clear that $T(\Omega^*) = \Omega^*$ and $T'(I) = I$. For each $\omega \in \Omega^*$ we obtain

$$\psi(T\omega) = \sum_{n=1}^{\infty} \frac{\omega_n}{r^n}, \quad T'(\psi(\omega)) = T'\left(\sum_{n=0}^{\infty} \frac{\omega_n}{r^{n+1}}\right).$$

It is readily seen that $\Omega^* = \bigcup_{i=0}^{r-1} \{x_0=i\} \cap \Omega^*$, $I = \bigcup_{i=0}^{r-2} [\frac{i}{r}, \frac{i+1}{r}) \cup [\frac{r-1}{r}, 1]$, and

$$\psi(\{x_0=i\} \cap \Omega^*) = [\frac{i}{r}, \frac{i+1}{r}) \quad \text{for} \quad 0 \le i \le r-2,$$

$$\psi(\{x_0=r-1\} \cap \Omega^*) = [\frac{r-1}{r}, 1] .$$

Thus, if $\omega \in \{x_0=i\} \cap \Omega^*$, $0 \le i \le r-1$, we obtain

$$T'(\psi(\omega)) = r\psi(\omega) - i = r\left(\frac{i}{r} + \sum_{n=1}^{\infty} \frac{\omega_n}{r^{n+1}}\right) - i$$

$$= \sum_{n=1}^{\infty} \frac{\omega_n}{r^n} = \psi(T\omega). \quad \square$$

(9.12) <u>Lemma</u>. $P(\Omega,T) \cap NA(\Omega) = P(\Omega^*,T) \cap NA(\Omega)$.

The proof is simple and is omitted.

(9.13) <u>Lemma</u>. Let ψ be the bijection between $P(\Omega^*)$ and $P(I)$ as defined in (9.7). Then the restriction of the mapping ψ to $P(\Omega^*,T)$ is a bijection between $P(\Omega^*,T)$ and $P(I,T')$.

<u>Proof</u>. It is enough to show that ψ is a surjection from $P(\Omega^*,T)$ onto $P(I,T')$. Let $\mu \in P(\Omega^*,T)$. It follows from (9.7) and (9.11) that $\psi\mu \in P(I)$ and

$$[T'(\psi\mu)](B) = \mu(\psi^{-1}(T'^{-1}(B))) = \mu(T^{-1}(\psi^{-1}(B)))$$

$$= [\psi(T\mu)](B) = (\psi\mu)(B) \quad \text{for each} \quad B \in \mathcal{B}(I),$$

so that $\psi\mu \in P(I,T')$. Similarly we obtain, for each $\nu \in P(I,T')$,

$$T(\psi^{-1}\nu) = (T\psi^{-1})\nu = (\psi^{-1}T')\nu = \psi^{-1}(T'\nu) = \psi^{-1}\nu$$

so that $\psi^{-1}\nu \in P(\Omega^*,T)$. □

We obtain from (9.9), (9.12) and (9.13) the following:

(9.14) Theorem. Let ψ be the bijection between $P(\Omega^*,T)$ and $P(I,T')$ as defined in (9.13). The restriction of the mapping ψ to $P(\Omega,T) \cap NA(\Omega)$ is a bijection between $P(\Omega,T) \cap NA(\Omega)$ and $P(I,T') \cap NA(I)$.

(9.15) Lemma. $E(\Omega,T) = E(\Omega^*,T) \subset P(\Omega^*,T)$.

Proof. By definition, $E(\Omega^*,T) = E(\Omega,T) \cap P(\Omega^*) \subset P(\Omega^*,T)$. Let μ be an atomic ergodic measure, i.e. $\mu \in E(\Omega,T) \cap A(\Omega)$. We see readily that no points in Ω_{r-1} are periodic points of T, so that, by (6.9), $\mu(\Omega_{r-1}) = 0$. Thus we obtain $E(\Omega,T) \cap A(\Omega) \subset E(\Omega^*,T)$. By (9.9) we also have $E(\Omega,T) \cap NA(\Omega) \subset E(\Omega,T) \cap P(\Omega^*) = E(\Omega^*,T)$. Using (6.9) we obtain

$$E(\Omega,T) = (E(\Omega,T) \cap A(\Omega)) \cup (E(\Omega,T) \cap NA(\Omega)) \subset E(\Omega^*,T)$$

so that $E(\Omega,T) = E(\Omega^*,T)$. □

(9.16) Theorem. Let ψ be the bijection between $P(\Omega^*,T)$ and $P(I,T')$ as defined in (9.13). Then the restriction of the mapping ψ to $E(\Omega,T)$ is a bijection between $E(\Omega,T)$ and $E(I,T')$.

Proof. Let $\mu \in E(\Omega,T)$ and $\nu = \psi\mu$. It follows from (9.15) and (9.13) that $\nu \in P(I,T')$. Suppose $T'^{-1}B = B$ where $B \in \mathcal{B}(I)$. By (9.11), we have $\psi^{-1}(B) = \psi^{-1}(T'^{-1}(B)) = T^{-1}(\psi^{-1}(B))$ so that $\nu(B) = \mu(\psi^{-1}(B)) = 0$ or 1. Thus $\nu \in E(I,T')$.

Let $\nu \in E(I,T')$ and $\mu = \psi^{-1}\nu$. Note that $\mu \in P(\Omega^*,T)$. Suppose $T^{-1}A = A$ where $A \in \mathcal{B}(\Omega)$. We readily see that

$$T\Omega^* \subset \Omega^* \subset T^{-1}T\Omega^* \subset T^{-1}\Omega^* ,$$

$$A \cap \Omega^* = T^{-1}A \cap \Omega^* \subset T^{-1}A \cap T^{-1}\Omega^* = T^{-1}(A\cap\Omega^*) .$$

Set $E = A \cap \Omega^*$. Then we have $E \subset T^{-1}E$ and $\psi(E) \subset \psi(T^{-1}(E)) = T'^{-1}(\psi(E))$ so that $\nu(T'^{-1}(\psi(E)) \,\Delta\, \psi(E)) = 0$. Therefore we have

$$\mu(E) = \nu(\psi(E)) = 0 \text{ or } 1, \text{ i.e. } \mu(A\cap\Omega^*) = 0 \text{ or } 1.$$

Since $\mu(A) = \mu(A\cap\Omega^*)$, we have $\mu(A) = 0$ or 1, and thus $\mu \in E(\Omega,T)$. □

Using (9.14) and (9.16) we obtain the following:

(9.17) Theorem. Let ψ be the bijection between $E(\Omega,T)$ and $E(I,T')$ as defined in (9.16). Then the restriction of the mapping ψ to $E(\Omega,T) \cap NA(\Omega)$ is a bijection between $E(\Omega,T) \cap NA(\Omega)$ and $E(I,T') \cap NA(I)$.

Let λ denote the $p = (p_i)_{i\in S}$ Bernoulli measure where $p_i = \frac{1}{r}$.

(9.18) Theorem. The probability spaces $\Omega = (\Omega, \mathcal{B}(\Omega), \lambda)$ and $I = (I, \mathcal{B}(I), m)$ are isomorphic, and the Bernoulli shift T on Ω and the r-adic transformation T' on I are isomorphic.

Proof. Since $\lambda \in E(\Omega,T) \cap NA(\Omega)$, we obtain, by (9.17), $\psi\lambda \in E(I,T') \cap NA(I)$. By (9.3) and (9.11), it remains to show that $\psi\lambda = m$. It is easily seen that, for each $n \geq 1$,

$$\lambda(\psi^{-1}[0, \frac{1}{r^n})) = \lambda(\{x_k=0,\ 0 \leq k \leq n-1\} \cap \Omega^*)$$

$$= \lambda(x_k=0,\ 0 \leq k \leq n-1) = \frac{1}{r^n},$$

$$\lambda(\psi^{-1}[1 - \frac{1}{r^n}, 1]) = \lambda(\{x_k=r-1,\ 0 \leq k \leq n-1\} \cap \Omega^*)$$

$$= \lambda(x_k=r-1,\ 0 \leq k \leq n-1) = \frac{1}{r^n},$$

$$\lambda(\psi^{-1}[u, u + \frac{1}{r^n})) = \lambda(\{x_k=i_k,\ 0 \leq k \leq n-1\} \cap \Omega^*)$$

$$= \lambda(x_k=i_k,\ 0 \leq k \leq n-1) = \frac{1}{r^n},$$

where $u := \sum_{k=0}^{n-1} \frac{i_k}{r^{k+1}} \in I_o$. By the unique extension theorem, we obtain $\psi\lambda = m$. □

The following result is a generalization of (8.8).

(9.19) Theorem. Let $\mu \in P(\Omega)$ be nonatomic. Then the probability spaces $(\Omega,\mathcal{B}(\Omega),\mu)$ and $(I,\mathcal{B}(I),m)$ are isomorphic.

Proof. Let φ and ψ be the mappings as defined in (9.1), let $\mu \in NA(\Omega)$, and let $\nu = \psi\mu$. By (9.9), $\nu \in NA(I)$. Let $f(x) = \nu([0,x))$ for $x \in I$, $E = \{x \in I : f^{-1}(\{x\})$ is a nondegenerate interval in $I\}$, $I_1 = I - f^{-1}(E)$ and $I_2 = I - E$. By (8.8), we have $\nu(I_1) = m(I_2) = 1$ and the mapping $f : (I_1,\mathcal{B}(I_1),\nu) \to (I_2,\mathcal{B}(I_2),m)$ is a measure-preserving homeomorphism. In particular, the two probability spaces $(I, \mathcal{B}(I),\nu)$ and $(I,\mathcal{B}(I),m)$ are isomorphic.

Define $\Omega' = \psi^{-1}(I_1)$. Then we have $\Omega' = \varphi^{-1}(I_1) \cap \Omega^* \in \mathcal{B}(\Omega)$ and $\mu(\Omega') = \mu(\psi^{-1}(I_1)) = \nu(I_1) = 1$. It is easily seen from (9.3) that the restriction of the mapping ψ to Ω' is an invertible measure-preserving map from $(\Omega',\mathcal{B}(\Omega'),\mu)$ onto $(I_1,\mathcal{B}(I_1),\nu)$. Note that $(\Omega,\mathcal{B}(\Omega),\mu)$ and $(I,\mathcal{B}(I),\nu)$ are isomorphic.

Define the mapping g by $g(\omega) = f(\psi(\omega))$ for each $\omega \in \Omega'$. It is straightforward to show that the mapping g is an invertible measure-preserving map from $(\Omega',\mathcal{B}(\Omega'),\mu)$ onto $(I_2,\mathcal{B}(I_2),m)$ so that the theorem follows. □

10. Continuous singular distribution functions

In this section, notation is as in section 9. Our point of departure is the following lemma.

(10.1) Lemma. Let $\mu_1, \mu_2 \in P(\Omega^*)$. Then we have

(i) $\mu_1 \perp \mu_2$ iff $\psi\mu_1 \perp \psi\mu_2$,

(ii) $\mu_1 \ll \mu_2$ iff $\psi\mu_1 \ll \psi\mu_2$.

Proof. By (9.7) we have $\psi\mu_i \in P(I)$, $i = 1,2$.

(i): Suppose that $\Omega^* = A_1 \cup A_2$, $A_1 \cap A_2 = \phi$ and $\mu_i(A_i) = 1$, $i = 1,2$. It follows from (9.3) that $I = \psi(\Omega^*) = \psi(A_1) \cup \psi(A_2)$, $\psi(A_1) \cap \psi(A_2) =$ $= \psi(A_1 \cap A_2) = \phi$, and $\psi^{-1}(\psi(A_i)) = A_i$ so that $\psi\mu_i(\psi(A_i)) = \mu_i(A_i) = 1$, $i = 1,2$. Thus $\psi\mu_1 \perp \psi\mu_2$.

Suppose that $I = B_1 \cup B_2$, $B_1 \cap B_2 = \phi$ and $(\psi\mu_i)(B_i) = 1$, $i = 1,2$. Then we have $\Omega^* = \psi^{-1}(I) = \psi^{-1}(B_1) \cup \psi^{-1}(B_2)$, $\psi^{-1}(B_1) \cap \psi^{-1}(B_2) = \phi$ and $\mu_i(\psi^{-1}(B_i)) = (\psi\mu_i)(B_i) = 1$, $i = 1,2$ so that $\mu_1 \perp \mu_2$.

(ii): Suppose that $\mu_1 \ll \mu_2$. If $(\psi\mu_2)(B) = \mu_2(\psi^{-1}(B)) = 0$, then $\mu_1(\psi^{-1}(B)) = (\psi\mu_1)(B) = 0$ so that $\psi\mu_1 \ll \psi\mu_2$.

Suppose that $\psi\mu_1 \ll \psi\mu_2$ and $\mu_2(A) = 0$ where $A \in \mathcal{B}(\Omega^*)$. By (9.3) we have $\psi^{-1}(\psi(A)) = A$ and so $\mu_2(A) = (\psi\mu_2)(\psi(A)) = 0$, so that $(\psi\mu_1)(\psi(A)) = \mu_1(A) = 0$. Thus $\mu_1 \ll \mu_2$. □

Using (9.16), (10.1) and (3.16), we obtain the following result.

(10.2) <u>Theorem</u>. Let $\mu,\nu \in P(I,T')$. Then

(i) if $\mu,\nu \in E(I,T')$, then either $\mu = \nu$ or $\mu \perp \nu$,

(ii) if $\nu \ll \mu$, $\mu \in E(I,T')$, then $\mu = \nu$.

(10.3) <u>Discussion and Definition</u>. For any $\nu \in P(I,T')$, let f be the distribution function of ν defined by $f(x) = \nu([0,x))$, $x \in I$. Since $T'\nu = \nu$, we obtain easily that $f(x) = \sum_{i=0}^{r-1} \{f(\frac{i}{r} + \frac{x}{r}) - f(\frac{i}{r})\}$, $x \in I$. Let $D(I,T')$ denote the set of all continuous singular nondecreasing functions f on I with the property $f(x) = \sum_{i=0}^{r-1} \{f(\frac{i}{r} + \frac{x}{r}) - f(\frac{i}{r})\}$, $x \in I$. It is plain that the set $D(I,T')$ is convex and is identified with the set of those nonatomic $\nu \in P(I,T')$ such that $\nu \perp m$.

(10.4) <u>Theorem</u>. Let $\mu \in P(\Omega,T)$, $\nu = \psi\mu$ and $f(x) = \nu([0,x))$, $x \in I$, and assume that $\nu \neq m$. Then the following are equivalent:

(i) μ is nonatomic T-ergodic, i.e., $\mu \in E(\Omega,T) \cap NA(\Omega)$.

(ii) ν is nonatomic T'-ergodic, i.e., $\nu \in E(I,T') \cap NA(I)$.

(iii) f is an extreme point of $D(I,T')$

<u>Proof</u>. By (9.17), we obtain (i) $\Longleftrightarrow$ (ii).

(ii) $\Longrightarrow$ (iii): Suppose $\nu \in E(I,T') \cap NA(I)$ and $f(x) = \nu([0,x))$, $x \in I$. Assume that $f = pg + qh$, where $g,h \in D(I,T')$, $0 < p,q < 1$, $p+q = 1$. Let σ and τ denote the measure on I induced by the distribution g

and h, respectively. It follows that $\sigma,\tau \in P(I,T') \cap NA(I)$ and $\nu = p\sigma + q\tau$. Since ν is an extreme point of $P(I,T')$ (as is easily seen by (3.17) and (9.17)), we must have $\nu = \sigma = \tau$, so that $f = g = h$.

(iii) $\Rightarrow$ (ii): Suppose f is an extreme point of $D(I,T')$. Let ν be the measure induced by f. Assume that $\nu = p\sigma + q\tau$ where $\sigma,\tau \in P(I,T')$, $0 < p,q < 1$, $p+q = 1$. Let g and h be the distribution functions of σ and τ, respectively. Then we have $f(x) = pg(x) + qh(x)$, $x \in I$. Since ν is nonatomic, both σ and τ are nonatomic, or equivalently, both g and h are continuous. We also have $f'(x) = pg'(x) + qh'(x) = 0$ m-a.e. so that $g'(x) = h'(x) = 0$ m-a.e., since $g' \geq 0$ and $h' \geq 0$ m-a.e. Consequently, we have $f = pg + qh$ where $g,h \in D(I,T')$. Since f is an extreme point of $D(I,T')$, we must have $f = g = h$, so that $\nu = \sigma = \tau$. □

(10.5) Theorem. Let μ be a (p,P)-Markov measure such that $pP = p$, and let f be the distribution function of the random variable φ on (Ω,μ), i.e. $f(x) = \mu(\omega : \varphi(\omega) < x)$. Then

(i) $$f\left(\sum_{k=1}^{n} \frac{i_{k-1}}{r^k}\right) = \sum_{j<i_0} p_j + p_{i_0}\left(\sum_{j<i_1} p_{i_0 j}\right) + \cdots$$

$$\cdots + p_{i_0 i_1} \cdots p_{i_{n-3} i_{n-2}} \left(\sum_{j<i_{n-1}} p_{i_{n-2} j}\right).$$

(ii) f is a strictly increasing (continuous) function iff P is positive.

Proof. (i) For any $i_0,\dots,i_{n-1} \in S$ with $i_{n-1} \geq 1$, we obtain

$$f\left(\sum_{k=1}^{n} \frac{i_{k-1}}{r^k}\right) = \mu\left(\omega : \varphi(\omega) < \sum_1^n \frac{i_{k-1}}{r^k}\right)$$

$$= \mu\left(x_0=j_0,\dots,x_{n-1}=j_{n-1} : \sum_1^n \frac{j_{k-1}}{r^k} < \sum_1^n \frac{i_{k-1}}{r^k}\right)$$

$$= \Sigma\left\{p_{j_0} p_{j_0 j_1} \cdots p_{j_{n-2} j_{n-1}} : \sum_1^n \frac{j_{k-1}}{r^k} < \sum_1^n \frac{i_{k-1}}{r^k}\right\}$$

$$= \sum_{j<i_0} p_j + p_{i_0}\left(\sum_{j<i_1} p_{i_0 j}\right) + \dots + p_{i_0} p_{i_0 i_1} \cdots p_{i_{n-3} i_{n-2}} \cdot$$

$$\cdot \left(\sum_{j<i_{n-1}} p_{i_{n-2} j}\right) .$$

(ii) Suppose P is positive. Then the probability p also is positive and the measure μ is nonatomic ergodic so that f is continuous. To prove that f is strictly increasing, let $0 \leq x < x' \leq 1$. There exist two r-adic rationals u and v such that $x \leq u < v \leq x'$, with

$$u = \sum_{k=1}^{n} \frac{i_{k-1}}{r^k}, \quad v = \sum_{k=1}^{n+m} \frac{i_{k-1}}{r^k}, \quad \text{where } i_0,\dots,i_{n+m-1} \in S, \ i_{n+m-1} \geq 1,$$

$n \geq 0$, $m \geq 1$. It follows that

$$f(x') - f(x) \geq f(v) - f(u)$$

$$\geq \sum_{j<i_{n+m-1}} \mu(x_0=i_0,\ldots,x_{n+m-2}=i_{n+m-2},\ x_{n+m-1}=j)$$

$$= p_{i_0} p_{i_0 i_1} \cdots p_{i_{n+m-3} i_{n+m-2}} \left(\sum_{j<i_{n+m-1}} p_{i_{n+m-2} j} \right) > 0 .$$

Suppose P is not positive. Then $p_{i_o j_o} = 0$ for some $i_o, j_o \in S$. It follows that $f\left(\frac{i_o}{r} + \frac{j_o}{r^2}\right) = \sum_{j<i_o} p_j + p_{i_o}\left(\sum_{j<j_o} p_{i_o j}\right) = f\left(\frac{i_o}{r} + \frac{j_o+1}{r^2}\right)$, so that f is not strictly increasing. □

(10.6) Theorem. Let μ be a p-Bernoulli measure where $p = (p_i)_{i\in S}$ is a positive probability, and let f be the distribution function of the random variable φ on (Ω,μ). Then

(i) if p is uniform, $p_i = \frac{1}{r}$ for all i, then $f(x) = x$ for all $x \in I$,

(ii) if p is not uniform, then f is a strictly increasing continuous singular function such that

$$f(x) = \sum_{j=0}^{i-1} p_j + p_i f(rx-1) \quad \text{for} \quad \frac{i}{r} \leq x < \frac{i+1}{r}, \quad 0 \leq i \leq r-1.$$

Proof. (i): It follows from the proof of (10.5) that, for each r-adic rational u of the form $u = \sum_{k=0}^{n-1} \frac{i_k}{r^{k+1}}$, $f(u) = u$. Since r-adic rationals are dense in I and f is continuous, $f(x) = x$ for all $x \in I$.

(ii): Suppose p is not uniform. It follows from (10.5) together with (10.2) that f is a strictly increasing continuous singular function. Suppose $\frac{i}{r} \le t \le \frac{i+1}{r}$, $0 \le i \le r-1$. Since $\{x_n\}_{n\ge 0}$ are i.i.d. random variables with a common distribution, $(p_i)_{i\in S}$, we obtain for $\nu = \psi\mu$,

$$f(t) = \sum_{j=0}^{i-1} \nu([\tfrac{j}{r}, \tfrac{j+1}{r})) + \nu([\tfrac{i}{r}, t))$$

$$= \sum_{j=0}^{i-1} \mu(x_0=j) + \mu(x_0=i, \tfrac{i}{r} + \sum_{n=1}^{\infty} \frac{x_n}{r^{n+1}} < t)$$

$$= \sum_{j=0}^{i-1} p_j + \mu(x_0=i)\mu(\sum_{n=1}^{\infty} \frac{x_n}{r^n} < rt-i)$$

$$= \sum_{j=0}^{i-1} p_j + p_i\mu(\omega : \varphi(\omega) < rt-i)$$

$$= \sum_{j=0}^{i-1} p_j + p_i f(rt-i). \quad \square$$

(10.7) Example. Let μ be the $(p_0,0,p_2)$-Bernoulli measure on $\Omega = \prod_0^{\infty} \{0,1,2\}$ where $0 < p_0,p_2 < 1$, $p_0 + p_2 = 1$. Let f be the distribution function of the random variable φ on $(\Omega,\mathcal{B}(\Omega),\mu)$. Since μ is nonatomic ergodic, f is a continuous singular function and is not strictly increasing by (10.5).

Define the open interval

$$I(i_0,i_1,\ldots,i_{n-2}) = \left(\sum_{k=0}^{n-2} \frac{i_k}{3^{k+1}} + \frac{1}{3^n} , \sum_{k=0}^{n-2} \frac{i_k}{3^{k+1}} + \frac{2}{3^n}\right) ,$$

where $i_k \in \{0,2\}$, $0 \le k \le n-2$, $n \ge 1$. Let

$$V_n = \cup\{I(i_0,\ldots,i_{n-2}) : i_0,\ldots,i_{n-2} \in \{0,2\}\} .$$

Then the Cantor set K is defined by $K = I - \bigcup_{n=1}^{\infty} V_n$. Using (10.5), we obtain, for $i_0,\ldots,i_{n-2} \in \{0,2\}$,

$$f\left(\sum_{k=0}^{n-2} \frac{i_k}{3^{k+1}} + \frac{1}{3^n}\right) = f\left(\sum_{k=0}^{n-2} \frac{i_k}{3^{k+1}} + \frac{2}{3^n}\right)$$

$$= \sum_{j<i_0} p_j + p_{i_0}\left(\sum_{j<i_1} p_j\right) + \cdots + p_{i_0}p_{i_1} \cdots p_{i_{n-3}}\left(\sum_{j<i_{n-2}} p_j\right)$$

$$+ p_{i_0}p_{i_1} \cdots p_{i_{n-2}}p_0$$

$$= p_0 j_0 + p_0 p_{i_0} j_1 + \cdots + p_0 p_{i_0} p_{i_1} \cdots p_{i_{n-3}} j_{n-2} +$$

$$+ p_0 p_{i_0} p_{i_1} \cdots p_{i_{n-2}}$$

where $j_k = \frac{i_k}{2}$, $0 \le k \le n-2$. Observe that

$$f(\tfrac{1}{3}) = f(\tfrac{2}{3}) = p_0 ,$$

$$f(\frac{1}{3^2}) = f(\frac{2}{3^2}) = p_0^2 , \quad f(\frac{2}{3} + \frac{1}{3^2}) = f(\frac{2}{3} + \frac{2}{3^2}) = p_0 + p_0 p_2 ,$$

$$f(\frac{1}{3^3}) = f(\frac{2}{3^3}) = p_0^3 , \quad f(\frac{2}{3^2} + \frac{1}{3^3}) = f(\frac{2}{3^2} + \frac{2}{3^3}) = p_0^2 + p_0^2 p_2 ,$$

$$f(\frac{2}{3} + \frac{1}{3^3}) = f(\frac{2}{3} + \frac{2}{3^3}) = p_0 + p_0^2 p_2 ,$$

$$f(\frac{2}{3} + \frac{2}{3^2} + \frac{1}{3^3}) = f(\frac{2}{3} + \frac{2}{3^2} + \frac{2}{3^3}) = p_0 + p_0 p_2 + p_0 p_2^2 .$$

Suppose $p_0 = p_1 = \frac{1}{2}$. In this case, the function f is called the Cantor function or Lebesgue's singular function. We obtain from the preceding result that

$$f(\tfrac{1}{3}) = f(\tfrac{2}{3}) = \tfrac{1}{2} ,$$

$$f(\frac{1}{3^2}) = f(\frac{2}{3^2}) = \frac{1}{2^2} , \quad f(\frac{2}{3} + \frac{1}{3^2}) = f(\frac{2}{3} + \frac{2}{3^2}) = \frac{1}{2} + \frac{1}{2^2} ,$$

$$f(\frac{1}{3^3}) = f(\frac{2}{3^3}) = \frac{1}{2^3} , \quad f(\frac{2}{3^2} + \frac{1}{3^3}) = f(\frac{2}{3^2} + \frac{2}{3^3}) = \frac{1}{2^2} + \frac{1}{2^3} ,$$

$$f(\frac{2}{3} + \frac{1}{3^3}) = f(\frac{2}{3} + \frac{2}{3^3}) = \frac{1}{2} + \frac{1}{2^3} ,$$

$$f(\frac{2}{3} + \frac{2}{3^2} + \frac{1}{3^3}) = f(\frac{2}{3} + \frac{2}{3^2} + \frac{2}{3^3}) = \frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3} , \ldots$$

and that, for $i_0, \ldots, i_{n-2} \in \{0,2\}$,

$$f\left(\sum_{k=0}^{n-2} \frac{i_k}{3^{k+1}} + \frac{1}{3^n}\right) = f\left(\sum_{k=0}^{n-2} \frac{i_k}{3^{k+1}} + \frac{2}{3^n}\right) = \sum_{k=0}^{n-2} \frac{j_k}{2^{k+1}} + \frac{1}{2^n},$$

where $j_k = \frac{i_k}{2}$.

Define $U = I - \bigcup_{n=1}^{\infty} \overline{V}_n$. The set $\bigcup_{n=1}^{\infty} \overline{V}_n$ is a union of countably infinite pairwise disjoint closed intervals J_k such that $f(J_k) = c_k$ (constant). We shall show that f is strictly increasing on the set U. It is straightforward to show that, for each $x \in U$, there is a unique $\omega = (\omega_n) \in \Omega$ such that $\omega_n \in \{0,2\}$, $n \geq 0$, and $x = \sum_{k=0}^{\infty} \frac{\omega_k}{3^{k+1}}$. Suppose $x,y \in U$, $x < y$. Let

$$x = \sum_{0}^{\infty} \frac{i_k}{3^{k+1}}, \quad y = \sum_{0}^{\infty} \frac{j_k}{3^{k+1}} \quad \text{where } i_k, j_k \in \{0,2\}.$$

Since $x < y$, there is an $n \geq 0$ such that

$$i_k = j_k \text{ for } k < n \text{ and } i_n \neq j_n, \text{ i.e. } i_n = 0,\ j_n = 2$$

Let $u = \sum_{0}^{n-1} \frac{i_k}{3^{k+1}}$. Then we obtain

$$u < x < u + \frac{1}{3^{n+1}} < u + \frac{2}{3^{n+1}} < y = u + \frac{2}{3^{n+1}} + \sum_{n+1}^{\infty} \frac{j_k}{3^{k+1}}.$$

Since not all j_k, $k \geq n+1$, are zero (for otherwise, $y \in \overline{V}_{n+1}$), there is an $m \geq n+1$ such that $j_m = 2$, so that

$$x < u + \frac{2}{3^{n+1}} < u + \frac{2}{3^{n+1}} + \sum_{n+1}^{m} \frac{j_k}{3^{k+1}} < y .$$

Thus we obtain $f(y) - f(x) \geq f\left(u + \frac{2}{3^{n+1}} + \sum_{n+1}^{m} \frac{j_k}{3^{k+1}}\right) - f\left(u + \frac{2}{3^{n+1}}\right) > 0$.

Note that $m(\{x : f'(x) = 0\}) = m(\bigcup_1^\infty V_n) = 1$. □

We shall give an example of a class of nonatomic non-Markov ergodic measures.

(10.8) Example. Let $S = 0,1,\ldots,r-1$ and $S' = 0,1,\ldots,r$ where $r \geq 2$. Let $q = (q_i)_{0\leq i\leq r}$ be a positive probability vector such that $q_0 \neq q_1$, $\max\{q_i\} < \frac{1}{2}$. Define the doubly stochastic matrix $P = (p_{ij})_{i,j\in S'}$ by

$$P = \begin{bmatrix} q_0 & q_1 & \cdots & q_r \\ q_1 & q_2 & \cdots & q_0 \\ \vdots & & & \\ q_r & q_0 & \cdots & q_{r-1} \end{bmatrix}$$

Clearly P is irreducible and the uniform probability vector p on S' is the stationary distribution for P. Let μ denote the (p,P)-Markov measure defined on $\Omega' = \prod_0^\infty S'$. Then μ is a nonatomic T-ergodic Markov measure on $(\Omega',\mathcal{B}(\Omega'))$. For each $n > 0$, define $y_n(\omega) = \omega_n$ for $\omega \in \Omega'$. Then $\{y_n\}_{n\geq 0}$ is a Markov chain with the state space S', the stationary transition matrix P and the stationary initial distribution p.

Let $f : S' \to S$ be such that $f(i) = i$ for all $i \in S$ and $f(r) = 0$. Define the stochastic process $\{x_n\}_{n\geq 0}$ on $(\Omega',\mathcal{B}(\Omega'),\mu)$ by

$$x_n(\omega) = f(y_n(\omega)), \quad \omega \in \Omega'.$$

It follows that, for any $i_0,\ldots,i_k \in S$, $k \geq 0$, $n \geq 1$,

$$\mu(x_0=i_0,\ldots,x_k=i_k) = \mu(y_0\in f^{-1}(i_0),\ldots,y_k\in f^{-1}(i_k))$$

$$= \mu(y_n\in f^{-1}(i_0),\ldots,y_{n+k}\in f^{-1}(i_k)) = \mu(x_n=i_0,\ldots,x_{n+k}=i_k)$$

so that $\{x_n\}_{n\geq 0}$ is stationary with state space S. However $\{x_n\}_{n\geq 0}$ is not a Markov chain on $(\Omega',\mathcal{B}(\Omega'),\mu)$, for we have

$$\mu(x_1=1|x_0=0) = (p_0p_{01} + p_rp_{r1})/(p_0+p_r) = \tfrac{1}{2}(q_1+q_0) \ ,$$

$$\mu(x_2=1|x_0=1,x_1=0) = p_1(p_{10}p_{01} + p_{1r}p_{r1})/p_1(p_{10}+p_{1r})$$

$$= \left(\frac{q_1}{q_1+q_0}\right) q_1 + \left(\frac{q_0}{q_1+q_0}\right) q_0 \ ,$$

whose only real solution is $q_1 = q_0$, a contradiction.

It is plain that the mapping $\xi : (\Omega',\mathcal{B}(\Omega')) \to (\Omega,\mathcal{B}(\Omega))$ defined by $(\xi(\omega))_n = x_n(\omega)$ for $\omega \in \Omega'$, $n \geq 0$ is a measurable surjection and that $\Omega \subset \Omega'$, $\mathcal{B}(\Omega) = \mathcal{B}(\Omega') \cap \Omega$. Define the probability ν on $(\Omega,\mathcal{B}(\Omega))$ by $\nu = \xi\mu$. It can be shown that ν is nonatomic. We see readily that $x_n(\omega) = \omega_n$ for each $\omega \in \Omega$ and $n \geq 0$, so that

$$\nu(x_0=i_0,\dots,x_n=i_n) = \mu(x_0=i_0,\dots,x_n=i_n),$$

where $i_0,\dots,i_n \in S$. It follows that $\nu(x_1=1|x_0=0) \neq \nu(x_2=1|x_0=1,x_1=0)$ so that ν is not Markov. Denote also by T the restriction of the shift T to Ω. Then we obtain

$$(\xi(T\omega))_n = x_n(T\omega) = f(y_n(T\omega)) = f(y_{n+1}(\omega))$$

$$= x_{n+1}(\omega) = (T\xi(\omega))_n ,$$

where $\omega \in \Omega'$, $n \geq 0$, so that $\xi T = T\xi$. To prove that ν is T-ergodic, suppose $T^{-1}E = E$, where $E \in \mathcal{B}(\Omega)$. It follows that $\xi^{-1}E = \xi^{-1}T^{-1}E = T^{-1}\xi^{-1}E$, where $\xi^{-1}E \in \mathcal{B}(\Omega')$, so that, by the ergodicity of μ, $\nu(E) = \mu(\xi^{-1}E) = 0$ or 1. □

11. Entropy

(11.1) Definition. A partition ξ of $(\Omega,\mathcal{B}(\Omega))$ is a finite disjoint collection $\xi = \{A_1,\ldots,A_k\}$ of measurable sets A_i such that $\Omega = \bigcup_{i=1}^{k} A_i$. For any two partitions $\xi = \{A_1,\ldots,A_k\}$ and $\eta = \{B_1,\ldots,B_n\}$ of $(\Omega,\mathcal{B}(\Omega))$, the join of ξ and η, denoted by $\xi \vee \eta$, is the partition defined by $\xi \vee \eta = \{A_i \cap B_j : 1 \le i \le k,\ 1 \le j \le m\}$. For each partition $\xi = \{A_1,\ldots,A_k\}$ of $(\Omega,\mathcal{B}(\Omega))$ and for each positive integer n, the join $\bigvee_{j=0}^{n-1} T^{-j}\xi$ of the partitions $\xi, T^{-1}\xi,\ldots,T^{-(n-1)}\xi$ is defined by

$$\bigvee_{j=0}^{n-1} T^{-j}\xi = \{A_{i_0} \cap T^{-1}A_{i_1} \cap \ldots \cap T^{-(n-1)}A_{i_{n-1}} : A_{i_0},\ldots,A_{i_{n-1}} \in \xi\}.$$

(11.2) Definition. Let $\xi = \{A_1,\ldots,A_k\}$ be a partition of $(\Omega,\mathcal{B}(\Omega))$ and let $\mu \in P(\Omega,T)$. The entropy $H_\mu(\xi)$ of the partition ξ relative to μ is defined by

$$H_\mu(\xi) = -\sum_{i=1}^{k} \mu(A_i)\log \mu(A_i)$$

where $\log \mu(A_i) = \log_e \mu(A_i)$ and $0\log 0 = 0$. The entropy $h_\mu(T,\xi)$ of the shift T relative to ξ and μ is defined by

$$h_\mu(T,\xi) = \overline{\lim}_n \frac{1}{n} H_\mu\left(\bigvee_{i=0}^{n-1} T^{-i}\xi\right).$$

The entropy (or the measure-theoretic entropy) $h_\mu(T)$ of the shift T relative to μ is defined by

$$h_\mu(T) = \sup\{h_\mu(T,\xi) : \xi \text{ is a partition of } (\Omega,\mathcal{B}(\Omega))\}$$

We state without proof the following result. (See Walters [1].)

(11.3) Lemma. Let $\eta : [0,1] \to [0, 1/e]$ be the function defined by

$$\eta(x) = \begin{cases} -x \log x & \text{if } 0 < x \le 1 \\ 0 & \text{if } x = 0. \end{cases}$$

Then

(i) η is strictly concave, that is,

$$c\eta(x) + (1-c)\eta(y) \le \eta(cx+(1-c)y) \quad \text{for all } x,y,c \in [0,1],$$

and equality holds iff $x=y$ or $c=0$ or $c=1$.

(ii) $\sum_{i=1}^{k} c_i\eta(x_i) \le \eta\left(\sum_{i=1}^{k} c_i x_i\right)$ for any $k \ge 1$ and any

$x_1,\dots,x_k$, $c_1,\dots,c_k \in [0,1]$ such that $\sum_{i=1}^{k} c_i = 1$. Equality holds iff all the x_i's corresponding to nonzero c_i are equal.

(11.4) Corollary. For each $\{x_i\}_{i=1}^{r} \subset [0,1]$ with $\sum_{i=1}^{r} x_i = 1$, $r \ge 2$,

$$\sum_{i=1}^{r} \eta(x_i) \le \log r.$$

The equality holds iff $x_i = \frac{1}{r}$ for all i.

Proof. By (11.3), we obtain

$$\sum_{i=1}^{r} \frac{1}{r} \eta(x_i) \le \eta\left(\sum_{i=1}^{r} \frac{1}{r} x_i\right) = \eta(\tfrac{1}{r}) = \frac{1}{r} \log r ,$$

so that $\sum_{i=1}^{r} \eta(x_i) \le \log r$. Again by (11.3), equality holds iff $x_1 = x_i$ for all i, or equivalently, $x_i = \frac{1}{r}$ for all i. ▫

(11.5) Remark. Let $\mu \in P(\Omega,T)$ and let ξ be the partition of $(\Omega,\mathcal{B}(\Omega))$ defined by $\xi = \{\{x_0=i\} : i \in S\}$. By (11.4), we obtain

$$H_\mu(\xi) = -\sum_{i=0}^{r-1} \mu(x_0=i) \log \mu(x_0=i) \le \log r.$$

It is easily seen that $H_\mu(T^{-i}\xi) = H_\mu(\xi)$ for each $i \ge 0$ and

$$H_\mu(\bigvee_{i=0}^{n-1} T^{-i}\xi) \le \sum_{i=0}^{n-1} H_\mu(T^{-i}\xi) \quad \text{for each } n \ge 1$$

so that $H(\bigvee_{i=0}^{n-1} T^{-i}\xi) \le n\, H_\mu(\xi) \le n \log r$ for each $n \ge 1$. It is well-known (see Billingsley [1], Walters [1]) that the sequence

$\{\frac{1}{n} H_\mu(\bigvee_{i=0}^{n-1} T^{-i}\xi)\}_{n\ge 0}$ is nonincreasing, so that

$$0 \le h_\mu(T,\xi) = \lim_n \frac{1}{n} H_\mu(\bigvee_{i=0}^{n-1} T^{-i}\xi) \le \log r.$$

We state the following version of the Kolmogorov-Sinai theorem.

(11.6) Theorem. Let $\xi = \{\{x_0=i\} : i \in S\}$. Then

$$0 \leq h_\mu(T) = h_\mu(T,\xi) \leq \log r \quad \text{for each} \quad \mu \in P(\Omega,T).$$

See Billingsley [1] and Walters [1] for the proof.

(11.7) Definition. The mapping $\mu \mapsto h_\mu(T)$ from $P(\Omega,T)$ into $[0, \log r]$ is called the entropy map of the shift.

Entropy was introduced into communication theory by Shannon [1] in 1948. In 1958, Kolmogorov [1] defined the entropy of the general measure-preserving transformation, and a basic contribution was made by Sinai [1], in 1959. We shall compute the entropy $h_\mu(T)$ for Markov measures and Bernoulli measures and investigate properties of the entropy map.

(11.8) Theorem. For each (p,P)-Markov measure μ in $M(\Omega,T)$,

$$h_\mu(T) = -\sum_{i \in S} \sum_{j \in S} p_i p_{ij} \log p_{ij} .$$

Proof. Let $E = \{i \in S : p_i > 0\}$ and $F = S - E$. Clearly $E \neq \phi$. It is easily seen that $\sum_{i \in E} p_i = 1$ and $p_{ij} = 0$ for each $i \in E$, $j \in F$, so that $\sum_{j \in E} p_{ij} = 1$ for each $i \in E$. We also have $p_j = \sum_{i \in E} p_i p_{ij}$ for each $j \in S$. Let $\xi = \{\{x_0=i\} : i \in S\}$. Then we have, for each $n > 1$,

$$\bigvee_{k=0}^{n-1} T^{-k}\xi = \{\{x_0=i_0,\ldots,x_{n-1}=i_{n-1}\} : i_0,\ldots,i_{n-1} \in S\}$$

and

$$H_\mu(\bigvee_{0}^{n-1} T^{-k}\xi) = -\Sigma\{\mu(x_0=i_0,\ldots,x_{n-1}=i_{n-1}) \cdot$$

$$\cdot \log \mu(x_0=i_0,\ldots,x_{n-1}=i_{n-1}) : i_0,\ldots,i_{n-1} \in S\}$$

$$= -\Sigma\{p_{i_0} p_{i_0 i_1} \cdots p_{i_{n-2} i_{n-1}} \log(p_{i_0} p_{i_0 i_1} \cdots p_{i_{n-2} i_{n-1}}) : i_0,\ldots,i_{n-1} \in S\}$$

$$= -\Sigma\{p_{i_0} p_{i_0 i_1} \cdots p_{i_{n-2} i_{n-1}} \log(p_{i_0} p_{i_0 i_1} \cdots p_{i_{n-2} i_{n-1}}) :$$

$$p_{i_0 i_1} \cdots p_{i_{n-2} i_{n-1}} > 0, \quad i_0,\ldots,i_{n-1} \in E\}$$

$$= -\Sigma\{p_{i_0} p_{i_0 i_1} \cdots p_{i_{n-2} i_{n-1}} (\log p_{i_0} + \sum_{j=1}^{n-1} \log p_{i_{j-1} i_j}) :$$

$$p_{i_0 i_1} \cdots p_{i_{n-2} i_{n-1}} > 0, \quad i_0,\ldots,i_{n-1} \in E\}$$

$$= -\Sigma\{p_{i_0} p_{i_0 i_1} \cdots p_{i_{n-2} i_{n-1}} \log p_{i_0} : i_0,\ldots,i_{n-1} \in E\}$$

$$- \sum_{j=1}^{n-1} \Sigma\{p_{i_0} p_{i_0 i_1} \cdots p_{i_{n-2} i_{n-1}} \log p_{i_{j-1} i_j} : p_{i_{j-1} i_j} > 0 ,$$

$$i_0,\ldots,i_{n-1} \in E\}$$

$$= -\sum_{i_0 \in E} p_{i_0} \log p_{i_0} - \sum_{j=1}^{n-1} \Sigma\{p_{i_{j-1}} (\log p_{i_{j-1} i_j}) p_{i_{j-1} i_j} \cdots p_{i_{n-2} i_{n-1}} : \; i_{j-1}, \ldots, i_{n-1} \in E\}$$

$$= -\sum_{i \in S} p_i \log p_i - (n-1) \sum_{i \in E} \sum_{j \in E} p_i p_{ij} \log p_{ij}$$

$$= -\sum_{i \in S} p_i \log p_i - (n-1) \sum_{i \in S} \sum_{j \in S} p_i p_{ij} \log p_{ij} ,$$

Since $0 \le -\sum_{i \in S} p_i \log p_i \le \log r$, we obtain

$$h_\mu(T,\xi) = \lim_{n\to\infty} \frac{1}{n} H_\mu(\bigvee_0^{n-1} T^{-k}\xi) = -\sum_{i \in S} \sum_{j \in S} p_i p_{ij} \log p_{ij} .$$

By the Kolmogorov-Sinai theorem, $h_\mu(T) = h_\mu(T,\xi) = -\sum_{i \in j} p_i p_{ij} \log p_{ij}$. □

(11.9) <u>Theorem</u>. For each p-Bernoulli measure μ, $h_\mu(T) = -\sum_{i \in S} p_i \log p_i$.

<u>Proof</u>. Each p-Bernoulli measure μ is the (p,P)-Markov measure where $p_{ij} = p_j$ for all $i,j \in S$. By (11.8), we obtain

$$h_\mu(T) = -\sum_{i \in S} \sum_{j \in S} p_i p_{ij} \log p_{ij} = -\sum_{i \in S} \sum_{j \in S} p_i p_j \log p_j$$

$$= \sum_{j \in S} p_j \log p_j . \quad \square$$

(11.10) Theorem. For each $\mu \in M(\Omega,T)$, the following are equivalent:

(i) $h_\mu(T) = 0$.

(ii) μ is a (p,P)-Markov measure such that $P = (p_{ij})_{i,j\in S}$ is a permutation matrix and $p = (p_i)_{i\in S}$ is a stationary distribution of P.

Proof. (ii) $\Rightarrow$ (i): Suppose (ii) holds. Since P is a permutation matrix, there exists a bijection φ of S onto itself such that

$$p_{ij} = \delta_{\varphi(i)j} \quad \text{for all } i,j \in S.$$

It follows that, for each $i \in S$,

$$\sum_j p_{ij} \log p_{ij} = \sum_j \delta_{\varphi(i)j} \log \delta_{\varphi(i)j} = 0$$

so that, by (11.9), $h_\mu(T) = 0$.

(i) $\Rightarrow$ (ii): Suppose (i) holds. Since $\mu \in M(\Omega,T)$, there exist a stochastic matrix $P = (p_{ij})_{i,j\in S}$ and a stationary distribution $p = (p_i)_{i\in S}$ of P

such that μ is the (p,P)-Markov measure. Suppose P is not a permutation matrix. Define $E = \{i \in S : p_i > 0\}$ and $F = S - E$. Then $E \neq \phi$. As we have shown in the proof of (11.8), $p' = (p_i)_{i \in E}$ is a positive probability vector on E, the matrix $P' = (p_{ij})_{i,j \in E}$ is stochastic, and $p'P' = p'$. Note that $p_{ij} = 0$ for each $i \in E$, $j \in F$.

We shall show that P' is a permutation matrix. Since $h_\mu(T) = \sum_{i \in E} (-\sum_{j \in S} p_{ij} \log p_{ij}) p_i = 0$, we obtain

$$-\sum_{j \in S} p_{ij} \log p_{ij} = 0 \quad \text{for each} \quad i \in E,$$

so that, for each $i \in E$,

$$p_{ij} = 0 \text{ or } 1 \quad \text{for all} \quad j \in S,$$

or equivalently, $p_{ij} = 0$ or 1 for all $j \in E$. Since $\sum_{j \in E} p_{ij} = 1$ for each $i \in E$, there is, for each $i \in E$, a unique $\varphi(i) \in E$ such that

$$p_{ij} = \delta_{\varphi(i)j} \quad \text{for all} \quad j \in E.$$

It remains to show that the mapping φ is a surjection of E onto itself. If not, there is a $k \in E$ such that $p_{ik} = 0$ for all $i \in E$. Then we obtain $p_k = \sum_{i \in E} p_i p_{ik} = 0$, a contradiction. Since E is a finite set, the mapping φ is a bijection of E onto itself so that P' is a permutation matrix. Note that

$$p_{ij} = \delta_{\varphi(i)j} \quad \text{for each } i \in E \text{ and each } j \in S.$$

We extend the permutation matrix $P' = (p_{ij})_{i,j\in E}$ to a permutation matrix $Q = (q_{ij})_{i,j\in S}$ by letting

$$q_{ij} = p_{ij} \quad \text{for } (i,j) \in E \times S; \quad q_{ij} = \delta_{ij}$$

$$\text{for } (i,j) \in F \times S.$$

It follows that

$$\sum_{i\in S} p_i q_{ij} = \sum_{i\in E} p_i q_{ij} = \sum_{i\in E} p_i p_{ij} = p_j$$

for each $j \in S$, so that $pQ = p$. Let ν denote the (p,Q)-Markov measure. To complete the proof, it is enough to show that for each $n \geq 0$,

$$\mu(x_0=i_0,\ldots,x_n=i_n) = \nu(x_0=i_0,\ldots,x_n=i_n)$$

or equivalently, $p_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} i_n} = p_{i_0} q_{i_0 i_1} \cdots q_{i_{n-1} i_n}$ for any states $i_0,\ldots,i_n$. We see at once that the above equality holds for $n = 0,1$. Suppose equality holds for some $n \geq 1$. Let $i_0,\ldots,i_{n+1}$ be any states. If $\mu(x_0=i_0,\ldots,x_n=i_n) = 0$, then $\mu(x_0=i_0,\ldots,x_{n+1}=i_{n+1}) = 0 =$ $= \nu(x_0=i_0,\ldots,x_{n+1}=i_{n+1})$. If $\mu(x_0=i_0,\ldots,x_n=i_n) > 0$, then all states $i_0,\ldots,i_n$ are in E, so that $p_{i_n i_{n+1}} = q_{i_n i_{n+1}}$ and $\mu(i_0=i_0,\ldots,x_{n+1}=i_{n+1}) =$ $= \nu(x_0=i_0,\ldots,x_{n+1}=i_{n+1})$. By induction, $\mu = \nu$. □

From (7.1) and (11.10) we obtain the following result.

(11.11) Corollary. $h_\mu(T) = 0$ for each atomic ergodic Markov measure μ.

(11.12) Theorem. For each $\mu \in M(\Omega,T)$, the following are equivalent:

(i) $h_\mu(T) = \log r$

(ii) μ is the $(\frac{1}{r}, \frac{1}{r}, \ldots, \frac{1}{r})$-Bernoulli measure.

Proof. (i) $\Rightarrow$ (ii): Suppose (i) holds. Let μ be the (p,P)-Markov measure in $M(\Omega,T)$. By (11.8), we obtain $0 \le h_\mu(T) = \sum_i f(i)p_i = \log r$, where $f(i) = -\sum_j p_{ij} \log p_{ij}$, $i \in S$. Since $0 \le f(i) \le \log r$ for each i, if $f(j) < \log r$ for some $j \in S$ with $p_j > 0$, then $h_\mu(T) = \sum_i f(i)p_i \le \log r$, a contradiction. Thus we have $f(i) = \log r$ for all i, so that, by (11.4), $p_{ij} = \frac{1}{r}$ for all i,j with $p_i > 0$. Also, we have, for each j,

$$p_j = \sum_i p_i p_{ij} = \sum_i p_i \frac{1}{r} = \frac{1}{r},$$

so that μ is the $(\frac{1}{r}, \ldots, \frac{1}{r})$-Bernoulli measure.

(ii) $\Rightarrow$ (i): If μ is the $(\frac{1}{r}, \ldots, \frac{1}{r})$-Bernoulli measure, then, by (11.9), $h_\mu(T) = -\sum_i \frac{1}{r} \log \frac{1}{r} = \log r$. ◫

We shall show that the entropy map is affine upper semicontinuous. We begin with a lemma.

(11.13) Lemma. Let $\xi = \{A_i\}_{i=1}^{k}$, $k \geq 2$, be a partition of $(\Omega, \mathcal{B}(\Omega))$. Then the mapping $\mu \mapsto H_\mu(\xi)$ from $P(\Omega)$ into $[0,\infty)$ is a bounded concave function.

Proof. By (11.4), we have $0 \leq H_\mu(\xi) = -\sum_{i=1}^{k} \mu(A_i)\log h(A_i) \leq \log k$ for each $\mu \in P(\Omega)$. Let $\mu,\nu \in P(\Omega)$ and $0 \leq c \leq 1$. Then, by (11.3),

$$H_{c\mu+(1-c)\nu}(\xi) = \sum_{i=1}^{k} \eta(c\mu(A_i) + (1-c)\nu(A_i))$$

$$\geq c \sum_i \eta(\mu(A_i)) + (1-c)\sum_i \eta(\nu(A_i))$$

$$= cH_\mu(\xi) + (1-c)H_\nu(\xi). \quad \square$$

(11.14) Lemma. Let $\xi = \{\{x_0=i\} : i \in S\}$. Then the mapping $\mu \mapsto h_\mu(T,\xi)$ is an affine mapping from $P(\Omega,T)$ into $[0, \log r]$.

Proof. Let $\mu,\nu \in P(\Omega,T)$ and $0 \leq c \leq 1$. By (11.13) we obtain

$$H_{c\mu+(1-c)\nu}(\bigvee_{i=0}^{n-1} T^{-i}\xi) \geq cH_\mu(\bigvee_{0}^{n-1} T^{-i}\xi)$$

$$+ (1-c)H_\nu(\bigvee_{0}^{n-1} T^{-i}\xi)$$

for each $n \geq 1$ so that

$$h_{c\mu+(1-c)\nu}(T,\xi) \geq ch_\mu(T,\xi) + (1-c)h_\mu(T,\xi).$$

Let α be any partition of $(\Omega, \mathcal{B}(\Omega))$ and let $A \in \alpha$. Put $p = \mu(A)$ and $q = \nu(A)$. Then

$$0 \le \eta(c\mu(A) + (1-c)\nu(A)) - c\eta(\mu(A)) - (1-c)\eta(\nu(A))$$

$$= -(cp + (1-c)q)\log(cp + (1-c)q) + cp \log p + (1-c)q \log q$$

$$= -cp[\log(cp + (1-c)q) - \log(cp)] - pc[\log cp - \log p]$$
$$-(1-c)q[\log(cp + (1-c)q) - \log(1-c)q]$$
$$-q(1-c)[\log(1-c)q - \log q]$$

$$\le -pc \log c - q(1-c)\log(1-c)$$

$$= \mu(A)\eta(c) + \nu(A)\eta(1-c).$$

By summing the above inequality over all $A \in \alpha$, we obtain

$$0 \le H_{c\mu+(1-c)\nu}(\alpha) - cH_\mu(\alpha) - (1-c)H_\nu(\alpha)$$
$$\le \eta(c) + \eta(1-c) \le \log 2,$$

for any partition α. Thus we have, for each $n \ge 1$,

$$0 \le H_{c\mu+(1-c)\nu}\Big(\bigvee_{i=0}^{n-1} T^{-i}\xi\Big) - cH_\mu\Big(\bigvee_{0}^{n-1} T^{-i}\xi\Big)$$
$$- (1-c)H_\nu\Big(\bigvee_{0}^{n-1} T^{-i}\xi\Big) \le \log 2,$$

so that

$$h_{c\mu+(1-c)\nu}(T,\xi) = ch_\mu(T,\xi) + (1-c)h_\nu(T,\xi).$$

By (11.5), the proof is complete. □

(11.15) Theorem. The entropy map $h_\mu(T)$ is a bounded affine function on $P(\Omega,T)$.

The proof follows from (11.6) and (11.14).

(11.16) Theorem. The entropy map $h_\mu(T)$ is a bounded affine upper semi-continuous function on $P(\Omega,T)$.

Proof. Let $\xi = \{\{x_0=i\} : i \in S\}$. By the Kolmogorov-Sinai theorem, $h_\mu(T) = h_\mu(T,\xi)$ for each $\mu \in P(\Omega,T)$. By (11.15), it remains to show that $h_\mu(T,\xi)$ is upper semicontinuous on $P(\Omega,T)$.

Suppose $\mu_k \to \mu$ in $P(\Omega,T)$. Let n be a fixed positive integer. We have then, $\bigvee_0^{n-1} T^{-j}\xi = \{Z(i_0,\ldots,i_{n-1}) : i_0,\ldots,i_{n-1} \in S\}$ and $\mu_k(Z(i_0,\ldots,i_n)) \to \mu(Z(i_0,\ldots,i_n))$ for each $Z(i_0,\ldots,i_n)$, so that $\lim_{k\to\infty} H_{\mu_k}(\bigvee_0^{n-1} T^{-j}\xi) = H_\mu(\bigvee_0^{n-1} T^{-j}\xi)$. Therefore for each fixed n, $H_\mu(\bigvee_0^{n-1} T^{-j}\xi)$ is a continuous function of μ. Since, by (11.5), $h_\mu(T,\xi) = \lim_{n\to\infty} \downarrow \frac{1}{n} H_\mu(\bigvee_0^{n-1} T^{-j}\xi)$, $h_\mu(T,\xi)$ is an upper semicontinuous function of μ. □

(11.17) Remark. We shall construct an example to show that the entropy mapping of T is not continuous on $P(\Omega,T)$. For each $n \geq 1$, consider the set of fixed points of T^n, $F(T^n)$. By (5.2), card $F(T^n) = r^n$. Let μ_n be defined by

$$\mu_n = \frac{1}{r^n} \sum_{\omega \in F(T^n)} \varepsilon_\omega , \quad n \geq 1 .$$

By (6.8), we have that $\mu_n \in P(\Omega,T)$, $n \geq 1$. For any cylinder set $Z(i_0,\ldots,i_m)$, we obtain, for $n \geq m$,

$$\mu_n(Z(i_0,\ldots,i_{m-1})) = \frac{1}{r^m} = \lambda(Z(i_0,\ldots,i_{m-1})),$$

where λ is $(\frac{1}{r},\ldots,\frac{1}{r})$-Bernoulli measure, $\lambda \in P(\Omega,T)$. Thus $\mu_n \to \lambda$. However $h_{\mu_n}(T) = 0$ for all $n \geq 1$, because μ_n is concentrated on a finite number of points, while $h_\lambda(T) = \log r$, by (11.12). Thus $h_{\mu_n}(T) \not\to h_\lambda(T)$.

(11.18) Theorem. The entropy map $\mu \mapsto h_\mu(T)$ is continuous on $M(\Omega,T)$.

Proof. Suppose $\mu_n \to \mu$ in $M(\Omega,T)$, where μ_n are the (p_n,P_n)-Markov measures and μ is the (p,P)-Markov measure. By (4.23), we have $\lim_{n\to\infty} p_n(i) = p_i$ for each $i \in S$, and $\lim_{n\to\infty} p_n(i,j) = p_{ij}$ for each $j \in S$, provided $p_i > 0$.

If $p_i = 0$, then since $0 \leq -p_n(i)p_n(i,j)\log p_n(i,j) \leq p_n(i)/e$ for each j, we have $\lim_{n\to\infty} p_n(i)p_n(i,j)\log p_n(i,j) = 0$ for each j.

It follows from (11.8) that

$$\lim_{n\to\infty} h_{\mu_n}(T) = -\lim_{n\to\infty} \sum_{i\in S} \sum_{j\in S} p_n(i)p_n(i,j)\log p_n(i,j)$$

$$= -\sum_i \sum_j p_i p_{ij} \log p_{ij} = h_\mu(T). \quad \square$$

(11.19) Remark. It follows from the Choquet representation theorem (3.18) together with (11.16) that, for each $\mu \in P(\Omega,T)$, there is a unique probability measure τ defined on $P(\Omega,T)$ such that $\tau(E(\Omega,T)) = 1$ and $h_\mu(T) = \int_{E(\Omega,T)} h_\nu(T)d\tau(\nu)$.

REFERENCES

P. Billingsley

[1] Ergodic Theory and Information. Wiley, New York, 1965.

[2] Probability and Measure. Wiley, New York, 1979.

R.M. Blumenthal and R.K. Getoor

[1] Markov Processes and Potential Theory. Academic Press, New York, 1968.

J.R. Brown

[1] Ergodic Theory and Topological Dynamics. Academic Press, New York, 1976.

K.L. Chung

[1] Markov Chains with Stationary Transition Probabilities. Springer-Verlag New York Inc., 1967.

M. Denker, C. Grillenberger, K. Sigmund

[1] Ergodic Theory on Compact Spaces. Springer-Verlag, New York, 1976.

J.L. Doob

[1] Stochastic Processes. Wiley, New York, 1953.

W. Feller

[1] An Introduction to Probability Theory and its Applications. Wiley, New York, 1950.

P. Halmos

[1] Measure Theory. Van Nostrand, New York, 1950.

E. Hewitt and K. Stromberg

[1] Real and Abstract Analysis. Springer-Verlag New York Inc., 1965.

P.G. Hoel, S.C. Port, and C.J. Stone

[1] Introduction to Stochastic Processes. Houghton Mifflin, Boston, 1972.

A.N. Kolmogorov

[1] "A new invariant for transitive dynamical systems". Dokl. Akad. Nauk SSSR 119, 1959.

J. Neveu

[1] Mathematical foundations of the calculus of probability. Holden-Day, 1965.

J.C. Oxtoby

[1] On two theorems of Parthasarathy and Kakutani. Proc. Int. Symp. on Ergodic Theory, Academic Press, N.Y., 1963, pp. 203-215.

K.R. Parthasarathy
[1] Probability measures on metric spaces. Academic Press, New York, 1967.

R.R. Phelps,
[1] Lectures on Choquet's Theorem. Van Norstrand, New York, 1966.

H.L. Royden
[1] Real Analysis. MacMillan, New York, 1963.

C. Shannon
[1] "A mathematical theory of communication". Bell System Tech. J.27, 1948.

Ya. G. Sinai
[1] "On the notion of the entropy of a dynamical system". Dokl. Akad. Nauk SSSR 124, 1959.

P. Walters
[1] An Introduction to Ergodic Theory. Springer-Verlag New York Inc., 1982.