STATIONARY ABSOLUTE DISTRIBUTIONS
FOR CHAINS OF INFINITE ORDER

Thesis by

Robert Jack Hemstead

In Partial Fulfillment of the Requirements

For the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1968

(Submitted April 1, 1968)

ACKNOWLEDGMENTS

I wish to thank Professor Henry A. Krieger for his advice and encouragement in the preparation of this thesis. His suggestions and criticisms have been invaluable. He has given freely of his time and has expressed a strong personal interest in my progress.

I wish also to thank Professor Richard A. Dean for numerous conversations which have encouraged me in this endeavor.

I am grateful to Professor Charles E. Bures for conversations about creativity and exposition which proved very helpful.

Also I am indebted to my typist, Mrs. Carolyn Einung, for her competent work and patient co-operation.

I wish to thank my parents and my uncle, Dr. Sidley K. Macfarlane, for their constant encouragement and personal support.

Finally, I acknowledge financial assistance from the Woodrow Wilson Foundation, the National Science Foundation, and the California Institute of Technology.

ABSTRACT

Let $\{Z_n\}_{n=-\infty}^{\infty}$ be a stochastic process with state space $S_1 = \{0,1,\ldots,D-1\}$. Such a process is called a chain of infinite order. The transitions of the chain are described by the functions

$$Q_i(i^{(0)}) = P(Z_n=i \mid Z_{n-1}=i_1^{(0)}, Z_{n-2}=i_2^{(0)}, \ldots) \quad (i \in S_1),$$

where $i^{(0)} = (i_1^{(0)}, i_2^{(0)}, \ldots)$ ranges over infinite sequences from S_1. If $i^{(n)} = (i_1^{(n)}, i_2^{(n)}, \ldots)$ for $n = 1,2,\ldots$, then $i^{(n)} \to i^{(0)}$ means that for each k, $i_k^{(n)} = i_k^{(0)}$ for all n sufficiently large.

Given functions $Q_i(i^{(0)})$ such that

(i) $0 \le Q_i(i^{(0)}) \le \xi < 1$

(ii) $\sum_{i=0}^{D-1} Q_i(i^{(0)}) \equiv 1$

(iii) $Q_i(i^{(n)}) \to Q_i(i^{(0)})$ whenever $i^{(n)} \to i^{(0)}$,

we prove the existence of a stationary chain of infinite order $\{Z_n\}$ whose transitions are given by

$$P(Z_n=i \mid Z_{n-1}, Z_{n-2}, \ldots) = Q_i(Z_{n-1}, Z_{n-2}, \ldots)$$

with probability 1. The method also yields stationary chains $\{Z_n\}$ for which (iii) does not hold but whose transition probabilities are, in a sense, "locally Markovian."

These and similar results extend a paper by T. E. Harris [_Pac. J. Math._, 5 (1955), 707-724].

Included is a new proof of the existence and uniqueness of a stationary absolute distribution for an Nth order Markov chain in which all transitions are possible. This proof allows us to achieve our main results without the use of limit theorem techniques.

TABLE OF CONTENTS

I. INTRODUCTION

A discrete-time stochastic process $\{Z_n\}_{n=-\infty}^{\infty}$ with infinite past and with a finite number of states is called a chain of infinite order. If the states of the chain are denoted by $0,1,\ldots,D-1$, where $D \geq 2$, then the transition probabilities have the form

$$Q_i^{(n)}(i_1,i_2,\ldots) = \text{Prob}\ (Z_n=i \mid Z_{n-1}=i_1, Z_{n-2}=i_2,\ldots).$$

Thus the future behavior of such a chain depends in general upon its entire past history.

We consider here only those chains of infinite order whose transition probabilities are temporally homogeneous—that is, those chains for which the transition probabilities

$$Q_i^{(n)}(i_1,i_2,\ldots) = Q_i(i_1,i_2,\ldots)$$

are independent of n. Given $Q_0,\ldots,Q_{D-1}$, it seems natural to ask whether there exists a chain of infinite order $\{Z_n\}$ with the Q_i's as transition probabilities, such that $\{Z_n\}$ has a stationary absolute distribution. More precisely, the question is the following: If $Q_0(i_1,i_2,\ldots),\ldots,$ $Q_{D-1}(i_1,i_2,\ldots)$ are non-negative functions such that

$$\sum_{i=0}^{D-1} Q_i \equiv 1,$$

does there exist a chain of infinite order $\{Z_n\}_{n=-\infty}^{\infty}$ such that

$$(*) \quad \text{Prob}\,(Z_n = i \mid Z_{n-1}, Z_{n-2}, \ldots) = Q_i(Z_{n-1}, Z_{n-2}, \ldots)$$

with probability 1 and such that the probabilities

$$\text{Prob}\,(Z_n = i) = p_i$$

are independent of n for $n=0,\pm 1,\ldots$ and $i=0,\ldots,D-1$? Such a chain $\{Z_n\}$ is called a stationary chain of infinite order. If $(\Omega, \mathcal{S}, \mathcal{P})$ is the underlying probability space on which the Z_n's are defined, the stationarity of $\{Z_n\}$ is equivalent to the validity of the invariance equation

$$(**) \quad \mathcal{P}(Z_n = i) = \int_\Omega Q_i(Z_n, Z_{n-1}, \ldots)\, d\mathcal{P}$$

for $n=0,\pm 1,\ldots$ and $i=0,\ldots,D-1$.

The problem of constructing a probability measure $\mathcal{P}$ such that (*) and (**) hold is in general quite difficult. Fortunately the problem can be simplified conceptually by representing $\{Z_n\}$ by a suitable Markov process on [0,1]. The representation makes use of a familiar device in the theory of stochastic processes: that of replacing the original state space of a process by the space of all probabilistically distinguishable "past histories" of the process. For example, if $\{Y_n\}$ is an Nth order Markov chain on the state space S, we can replace S by

$$S' = \{(i_1,\ldots,i_N) : i_1,\ldots,i_N \in S\}$$

and consider a new process $\{Y'_n\}$ on S' with transition probabilities

$$\text{Prob}\,(Y'_n = (j_0,\ldots,j_{N-1}) \mid Y'_{n-1} = (i_1,\ldots,i_N))$$

$$= \begin{cases} \text{Prob}\,(Y_n = j_0 \mid Y_{n-1} = i_1, \ldots\ Y_{n-N} = i_N) \\ \qquad\qquad\qquad \text{if } i_1 = j_1, \ldots, i_{N-1} = j_{N-1} \\ 0 \qquad\qquad\qquad \text{otherwise.} \end{cases}$$

Each past history of the Y_n-process is compressed into a single state of the Y'_n-process, with the result that the latter process is Markovian. No information is lost by means of this transformation; the principal disadvantage is that S' is larger than S.

In the case of a chain of infinite order with state space S, the space S' must consist of infinite sequences from S, due to the dependence of the transition probabilities on an infinite past. Thus if $S = \{0,\ldots,D-1\}$, then $S' = \{(i_1,i_2,\ldots) : i_1,i_2,\ldots \varepsilon\ S\}$ is actually uncountable. S' does have the advantage, however, of admitting a natural mapping

$$(i_1,i_2,\ldots) \to \sum_{k=1}^{\infty} i_k/D^k$$

onto the unit interval [0,1]. This mapping allows us to regard the elements of S' as simply real numbers in [0,1]. To the sequence $(i_1,i_2,\ldots)\ \varepsilon\ S'$ we associate the real number x with D-ary expansion $.i_1 i_2 \ldots$. Unfortunately the correspondence is not 1-1, since D-adic rational numbers have two D-ary expansions. However, this discrepancy turns out to be incidental, because a stationary chain of infinite order assigns zero probability to all sets of the form

$$\{Z_{n-1}=i_1,\ldots,Z_{n-k}=i_k,\ Z_{n-k-1}=Z_{n-k-2}=\ldots=0\}$$

where $i_k \neq 0$, and all sets of the form

$$\{Z_{n-1}=i_1,\ldots,Z_{n-k}=i_k,\ Z_{n-k-1}=Z_{n-k-2}=\ldots=D-1\}$$

where $i_k \neq D-1$. Thus in a natural way we can represent a stationary chain of infinite order $\{Z_n\}_{n=-\infty}^{\infty}$ on $\{0,\ldots,D-1\}$ by a Markov process $\{X_n\}_{n=-\infty}^{\infty}$ on $[0,1]$, and $\{X_n\}$ is itself stationary.

Corresponding to the transition

$$(i_1,i_2,\ldots) \rightarrow (i,i_1,i_2,\ldots)$$

on S' with probability $Q_i(i_1,i_2,\ldots)$ is the X_n-transition

$$x = .i_1i_2\ldots \rightarrow (i+x)/D = .ii_1i_2\ldots,$$

the probability of which we denote by $f_i(x)$. Thus the transition probabilities for the X_n-process are simply real functions $f_0,\ldots,f_{D-1}$ on $[0,1]$. For convenience we let $\vec{f} = (f_0,\ldots,f_{D-1})$. If

$$\phi_0(x) = (0+x)/D,\ldots,\phi_{D-1}(x) = (D-1+x)/D,$$

then the transition function for the X_n-process is given by

$$\text{Prob } (X_{n+1}\varepsilon B \mid X_n=x) = P_{\vec{f}}(x,B) = \sum_{i=0}^{D-1} f_i(x)\chi_B[\phi_i(x)].$$

Hence, for Borel-measurable f_i's, a probability measure μ on the Borel sets of [0,1] gives a stationary absolute distribution for $\{X_n\}$ if and only if

$$(***) \quad \mu(B) = \int_0^1 \mu(dx) P_{\vec{f}}(x,B)$$

for every Borel set B. In this manner, the problem of studying existence and uniqueness of stationary chains $\{Z_n\}_{n=-\infty}^{\infty}$ with prescribed $Q_0,\ldots,Q_{D-1}$ reduces to the problem of studying solutions μ of (***) for a given $\vec{f}$.

A standard method of obtaining stationary absolute distributions for a Markov process $\{X_n\}$ is the following: Start the process at time n = 0 with an arbitrary initial distribution

$$\mu_0(\cdot) = \text{Prob } (X_0 \in (\cdot));$$

let the process evolve in time; then observe whether the absolute distributions

$$\mu_n(\cdot) = \text{Prob } (X_n \in (\cdot))$$

tend (in the usual sense or in a suitable Cesaro sense) to a limiting distribution μ as $n \to \infty$. If they do, μ is a stationary distribution for the process. Furthermore, if μ is independent of μ_0, then μ is the unique stationary absolute distribution for $\{X_n\}$.

Using this technique, T. E. Harris proved the following theorem in [4, p. 712]: Let $f_0,\ldots,f_{D-1}:[0,1] \to [0,1]$

such that

$$\sum_{i=0}^{D-1} f_i \equiv 1$$

and such that one of the f_i's is bounded away from 0. Define the sequence $\varepsilon_1, \varepsilon_2, \ldots$ by setting

$$\varepsilon_m = \max_{0 \leq i \leq D-1} \sup_{(x \equiv y)_m} |f_i(x) - f_i(y)|,$$

where $(x \equiv y)_m$ if x and y have the same (terminating) D-ary expansion to at least m places. If

$$\sum_{m=1}^{\infty} \prod_{k=1}^{m} [1 - (D/2)\varepsilon_k] = \infty$$

then there is a unique stationary Markov process $\{X_n\}$ on [0,1] with transition function $P_{\vec{f}}(x,B)$.

This theorem generalized a theorem of Doeblin and Fortet (see [1]), who proved a similar result under the stronger hypothesis

$$\sum_{m=1}^{\infty} \varepsilon_m < \infty.$$

Harris' condition on the ε_m's seems somewhat unnatural. A more natural and less restrictive condition to impose would be that $\{\varepsilon_m\}$ form a null sequence. This is equivalent to requiring that the f_i's be right-continuous with a left-hand limit at each D-adic rational less than 1, and continuous everywhere else. In the context of chains of infinite order this condition becomes the following natural requirement of continuity: If $i^{(0)} = (i_1^{(0)}, i_2^{(0)}, \ldots)$, $i^{(1)} = (i_1^{(1)}, i_2^{(1)}, \ldots)$, $i^{(2)} = (i_1^{(2)}, i_2^{(2)}, \ldots)$, ... such that for each k, $i_k^{(n)} = i_k^{(0)}$ for all n sufficiently large, then

$$Q_i(i^{(n)}) \to Q_i(i^{(0)})$$

as $n \to \infty$. Roughly this says that the transition probabities of the process depend only slightly upon the remote past.

By a method quite different from that of Harris, we prove that $\varepsilon_m \downarrow 0$ implies the existence of a $P_{\vec{f}}$-invariant measure μ, provided the f_i's are bounded away from 1. Further, we show that even the condition $\varepsilon_m \downarrow 0$ is inessential, provided the discontinuities of the f_i's are sufficiently separated.

Our approach is as follows: We consider the relation A which associates to each transition function $P_{\vec{f}}$ those measures μ which satisfy (***). Then we study the question: In what sense is the graph of A closed? More precisely, suppose that a sequence of transition vectors $\vec{f}^{(n)}$ converges in some sense to a transition vector $\vec{f}$. Suppose further that measures μ_n exist such that μ_n and $P_{\vec{f}^{(n)}}$ are A-related for each n. If the μ_n's have a limit measure μ (in some sense), does it follow that μ and $P_{\vec{f}}$ are A-related? With suitable definitions of convergence, the answer is affirmative.

The organization of the sections of the thesis is generally as follows: In Section II we standardize our terminology. In Section III we introduce the general notions of a transition function P and of a measure μ invariant with respect to P, in the sense of (***). In Section IV we define the so-called D-ary transition functions $P_{\vec{f}}$ which reflect the structure of a chain of infinite order, and we classify the points at which $P_{\vec{f}}$-invariant measures can have positive mass. In Section V we discuss in detail the relationship between chains of infinite order on $\{0,\ldots,D-1\}$ and Markov processes on [0,1] with D-ary transition functions. These results are known [4, p. 710], but heretofore a detailed proof of them has not appeared. In Section VI we use a functional analytic fixed-point theorem to prove the existence of a $P_{\vec{f}}$-invariant measure for vectors $\vec{f}$ with continuous components. Then we prove our major existence theorem (Theorem 30) using the "closed graph" approach described earlier. In Section VII we give a new proof of the existence and uniqueness of a stationary absolute distribution for Nth order Markov chains in which all transitions are possible. The proof requires an application of Theorem 30, and the theorem is phrased in the terminology of the preceding sections. Finally in Section VIII we apply the theorems of Sections VI and VII to obtain wide classes of vectors $\vec{f}$ for which there exist $P_{\vec{f}}$-invariant measures. In corollaries we describe the corresponding results for chains of infinite order. In particular, we prove the existence of certain stationary chains of infinite order with "locally Markovian" transition probabilities.

II. PRELIMINARY NOTATIONS AND REMARKS

R denotes the collection of real numbers, and R_0 denotes the collection of non-negative real numbers. Z denotes the collection of integers, Z_0 denotes the collection of non-negative integers, and Z_1 denotes the collection of positive integers. D denotes a fixed but arbitrary integer greater than or equal to 2. For $N \in Z_0$, S_N denotes the set $\{0,\ldots,D^N-1\}$.

If Ω is a set and B is a subset of Ω, then χ_B denotes the indicator function of B. If C is a collection of subsets of Ω, then $\sigma(C)$ denotes the σ-algebra in Ω generated by C. If $\{Z_\alpha\}$ is a collection of real functions on Ω, then $\sigma(\{Z_\alpha\})$ denotes the smallest σ-algebra such that each Z_α is $\sigma(\{Z_\alpha\})$-measurable.

The notation "x" replaces the notation "{x}" for singleton x, whenever the context permits.

III. TRANSITION FUNCTIONS AND INVARIANT MEASURES

The probabilistic structure of a Markov chain $\{X_n\}_{n=-\infty}^{\infty}$ with stationary transition probabilities and a stationary absolute distribution is completely determined by the transition function P and the invariant absolute distribution μ of the process. Customarily $P(x,B)$ denotes the probability of a transition from the point x into the set B; that is,

$$P(x,B) = \text{Prob } (X_n \in B \mid X_{n-1} = x).$$

On the other hand, $\mu(B)$ denotes the absolute probability that the process lies in B, under no assumptions about the past; thus

$$\mu(B) = \text{Prob } (X_n \in B).$$

The quantities $P(x,B)$ and $\mu(B)$ are independent of n by the stationarity assumptions.

The requirement that μ be a stationary absolute distribution for the process is equivalent to the requirement that μ and P be related by a certain functional equation. This equation is introduced, and several of its features are briefly studied, in Section III.

Definition 1: Let $\mathcal{B}$ denote the collection of Borel sets in $[0,1]$. Let $\mathcal{F}$ denote the collection of bounded $\mathcal{B}$-measurable functions mapping $[0,1]$ into $\mathcal{R}$, and let $\mathcal{F}_0$ denote the collection of non-negative functions in $\mathcal{F}$. Let $\mathcal{M}$ denote the collection of probability measures on $\mathcal{B}$. If $\mu \in \mathcal{M}$, let F_μ denote the left-continuous distribution function of μ.

<u>Definition 2</u>: A function $P : [0,1] \times \mathcal{B} \to \mathcal{R}$ is called a transition function on $[0,1]$ if

(i) $P(\cdot,B) \in \mathcal{F}$ for each fixed $B \in \mathcal{B}$

and

(ii) $P(x,\cdot) \in \mathcal{M}$ for each fixed $x \in [0,1]$.

Let $\mathcal{T}$ denote the collection of transition functions on $[0,1]$, and let

$$\mathcal{A} = \{(\mu,P) \in \mathcal{M} \times \mathcal{T} : \int_0^1 \mu(dx)P(x,B) = \mu(B) \qquad (B \in \mathcal{B})\}.$$

If $(\mu,P) \in \mathcal{A}$, then μ is said to be <u>P-invariant</u>.

Given a probability measure μ and a transition function P, we may call a Borel set B (μ,P)-invariant if

$$\int_0^1 \mu(dx)P(x,B) = \mu(B).$$

Thus μ is a P-invariant measure if and only if all Borel sets B are (μ,P)-invariant. It is useful to note that the (μ,P)-invariant sets always form a monotone class, even when μ is not P-invariant.

<u>Lemma 3</u>: Let $\mu \in \mathcal{M}$ and $P \in \mathcal{T}$. Then the collection

$$\mathcal{C} = \{B \in \mathcal{B} : \int_0^1 \mu(dx)P(x,B) = \mu(B)\}$$

is a monotone class.

Proof: The verification is the same for increasing or decreasing sequences. If, for example, $\{C_n\}_{n=1}^{\infty}$ is an increasing sequence of sets in C, then

$$\int_0^1 \mu(dx)P(x, \bigcup_{n=1}^{\infty} C_n) = \int_0^1 \mu(dx) \lim_{n\to\infty} P(x,C_n)$$
$$= \lim_{n\to\infty} \int_0^1 \mu(dx)P(x,C_n)$$
$$= \lim_{n\to\infty} \mu(C_n) = \mu(\bigcup_{n=1}^{\infty} C_n)$$

by the countable additivity of $P(x,\cdot)$ and μ, and by monotone convergence.

In Section V we shall show how to construct a chain of infinite order from a set of transition probabilities with certain properties. This construction rests heavily upon Kolmogorov's Consistency Theorem, which is stated as Theorem 18. The following two lemmas, which depend only upon general properties of transition functions and probability measures and, in the case of Lemma 4, upon the invariance equation, are needed for the application of Kolmogorov's Theorem to our construction.

Lemma 4: If $(\mu,P) \in A$, then

$$\int_0^1 \int_{B_1} \cdots_{B_k} \mu(dx)P(x,dx_1)\ldots P(x_{k-1},dx_k)$$
$$= \int_{B_1} \cdots \int_{B_k} \mu(dx_1)P(x_1,dx_2)\ldots P(x_{k-1},dx_k)$$

for $B_1, \ldots, B_k \varepsilon \mathcal{B}$.

Proof: The $\mathcal{B}$-measurability of each function

$$f_j(x) = \int_{B_j} \cdots \int_{B_k} P(x, dx_j) \ldots P(x_{k-1}, dx_k) \qquad (j=1, \ldots, k)$$

follows from the $\mathcal{B}$-measurability of $P(\cdot, B)$. For $B \varepsilon \mathcal{B}$ set

$$H_B = \{f \varepsilon F : \int_0^1 \mu(dx) \int_B P(x, dy) f(y) = \int_B \mu(dy) f(y)\}.$$

Then $H_B \supseteq \{\chi_{B_0}\}_{B_0 \varepsilon \mathcal{B}}$ because

$$\int_0^1 \mu(dx) \int_B P(x, dy) \chi_{B_0}(y) = \int_0^1 \mu(dx) P(x, B \cap B_0)$$

$$= \mu(B \cap B_0) = \int_B \mu(dy) \chi_{B_0}(y).$$

Also H_B has the properties

(i) $f_1, f_2 \varepsilon H_B \rightarrow \alpha_1 f_1 + \alpha_2 f_2 \varepsilon H_B \qquad (\alpha_1, \alpha_2 \varepsilon R)$

and

(ii) $0 \leq f_n \varepsilon H_B,\ f_n \uparrow f \varepsilon F \rightarrow f \varepsilon H_B$.

By the usual approximation argument it follows that $H_B = F$. Thus $f_2 \varepsilon H_{B_1}$ and

$$\int_0^1 \int_{B_1} \cdots \int_{B_k} \mu(dx) P(x, dx_1) \ldots P(x_{k-1}, dx_k)$$

$$= \int_0^1 \mu(dx) \int_{B_1} P(x, dx_1) f_2(x_1) = \int_{B_1} \mu(dx_1) f_2(x_1)$$

$$= \int_{B_1} \cdots \int_{B_k} \mu(dx_1) P(x_1, dx_2) \ldots P(x_{k-1}, dx_k).$$

<u>Lemma 5</u>: Let $[0,1]^k = \mathop{\times}_{j=1}^{k} [0,1]$, and let $B^{(1)}, B^{(2)}, \ldots \subseteq [0,1]^k$ such that $\bigcup_{n=1}^{\infty} B^{(n)} = [0,1]^k$ and $B^{(n)} \cap B^{(m)} = \emptyset$ for $n \neq m$.

Let $B^{(n)} = \times_{j=1}^{k} B_j^{(n)}$, where $B_1^{(n)},\ldots,B_k^{(n)} \in \mathcal{B}$. Then

$$\sum_{n=1}^{\infty} \int_{B_1^{(n)}} \cdots \int_{B_k^{(n)}} \mu(dx_1)P(x_1,dx_2)\ldots P(x_{k-1},dx_k) = 1$$

for $\mu \in M$, $P \in T$.

Proof: Set

$$\rho = \sum_{n=1}^{\infty} \int_{B_1^{(n)}} \cdots \int_{B_k^{(n)}} \mu(dx_1)P(x_1,dx_2)\ldots P(x_{k-1},dx_k)$$

$$= \sum_{n=1}^{\infty} \int_0^1 \cdots \int_0^1 \mu(dx_1)P(x_1,dx_2)\ldots P(x_{k-1},dx_k) \prod_{j=1}^{k} \chi_{B_j^{(n)}}(x_j).$$

Repeated application of the monotone convergence theorem gives

$$\rho = \int_0^1 \cdots \int_0^1 \mu(dx_1)P(x_1,dx_2)\ldots P(x_{k-1},dx_k) \sum_{n=1}^{\infty}\prod_{j=1}^{k} \chi_{B_j^{(n)}}(x_j).$$

Since $\{B^{(n)}\}_{n=1}^{\infty}$ forms a partition of $[0,1]^k$, it follows that

$$\sum_{n=1}^{\infty}\prod_{j=1}^{k} \chi_{B_j^{(n)}}(x_j) \equiv \sum_{n=1}^{\infty} \chi_{B^{(n)}}(x_1,\ldots,x_k)$$

$$\equiv \chi_{\bigcup_{n=1}^{\infty} B^{(n)}}(x_1,\ldots,x_k) \equiv 1.$$

Thus

$$\rho = \int_0^1 \cdots \int_0^1 \mu(dx_1)P(x_1,dx_2)\ldots P(x_{k-1},dx_k) = 1.$$

IV. D-ARY TRANSITION FUNCTIONS

We turn now to studying the particular transition functions appropriate to our topic. For us the D-ary expansion of a real number $x \varepsilon [0,1]$ represents a possible past history in a chain of infinite order. Hence we study only those Markov chains $\{X_n\}_{n=-\infty}^{\infty}$ on $[0,1]$ in which the transition

$$X_n \to X_{n+1}$$

preserves the D-ary expansion of X_n. Thus if

$$X_n = \sum_{k=1}^{\infty} i_k/D^k,$$

then X_{n+1} must lie among the points

$$\phi_i(X_n) = i/D + \sum_{k=2}^{\infty} i_{k-1}/D^k \qquad (i \varepsilon S_1).$$

This restriction leads to the following definition, in which $f_i(x)$ represents the probability of a transition from the point x to the point $\phi_i(x)$.

<u>Definition 6</u>: Let

$$F^{(D)} = \{\vec{f} = (f_0,\ldots,f_{D-1}) : f_0,\ldots,f_{D-1} \varepsilon F_0;\ \sum_{i=0}^{D-1} f_i \equiv 1\}.$$

The elements of $F^{(D)}$ will be called <u>D-ary transition vectors</u>. For $\vec{f} \varepsilon F^{(D)}$ define $P_{\vec{f}} : [0,1] \times B \to R$ by setting

$$P_{\vec{f}}(x,B) = \sum_{i=0}^{D-1} f_i(x)\chi_B[\phi_i(x)] \qquad (x \in [0,1],\ B \in \mathcal{B}),$$

where

$$\phi_i(x) = (i+x)/D \qquad (x \in [0,1];\ i \in S_1).$$

<u>Proposition 7</u>: $P_{\vec{f}} \in \mathcal{T}$ for $\vec{f} \in \mathcal{F}^{(D)}$.

Proof: The $\mathcal{B}$-measurability of $P_{\vec{f}}(\cdot,B)$ follows from the $\mathcal{B}$-measurability of $f_0,\ldots,f_{D-1},\chi_B,\phi_0,\ldots,\phi_{D-1}$. For fixed $x \in [0,1]$, $P_{\vec{f}}(x,\cdot)$ is the discrete measure assigning probability $f_i(x)$ to the point $\phi_i(x)$, $i \in S_1$.

In Sections VI, VII, and VIII the question of ultimate concern to us will be the following: Given a D-ary transition vector $\vec{f}$, when does there exist a $P_{\vec{f}}$-invariant probability measure μ? Lemma 9 simplifies our procedure for testing whether a given measure μ is actually $P_{\vec{f}}$-invariant. First we make a useful definition.

<u>Definition 8</u>: For $N \in Z_0$ and $j \in S_N$ set

$$I(j,N) = \begin{cases} [j/D^N,\ (j+1)/D^N) & \text{if} \quad j < D^N-1 \\ [j/D^N,\ (j+1)/D^N] & \text{if} \quad j = D^N-1. \end{cases}$$

<u>Lemma 9</u>: Let $\mu \in \mathcal{M}$ and $\vec{f} \in \mathcal{F}^{(D)}$. Then $(\mu,P_{\vec{f}}) \in \mathcal{A}$ if and only if

$$(*) \qquad \int_{I(j,N)} f_i(x)\mu(dx) = \mu[I(iD^N+j,N+1)]$$

for $j \in S_N$, $N \in Z_0$, $i \in S_1$.

Proof: Since

$$\phi_k^{-1}[I(iD^N+j,N+1)] = \begin{cases} I(j,N) & \text{if} \quad k = i \\ \emptyset & \text{if} \quad k \neq i \end{cases}$$

it follows that

$$\int_0^1 \mu(dx)P_{\vec{f}}(x,I(iD^N+j,N+1)) = \int_{I(j,N)} f_i(x)\mu(dx).$$

Hence if $\mathcal{C}'$ is the collection

$$\mathcal{C}' = \{I(j,N) : j \in S_N,\ N \in Z_1\},$$

then (*) is equivalent to

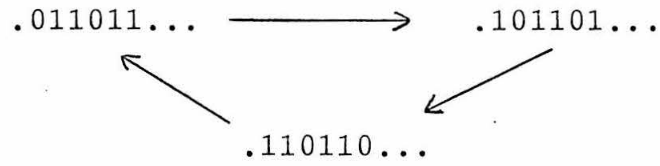
$$(**) \quad \int_0^1 \mu(dx)P_{\vec{f}}(x,C') = \mu(C')$$

for $C' \in \mathcal{C}'$. Thus the necessity of (*) is clear.

Suppose conversely that (**) holds. Assume, in other words, that the collection $\mathcal{C}$ of $(\mu,P_{\vec{f}})$-invariant sets contains the semi-algebra $\mathcal{C}'$. Since $\mathcal{C}$ contains all finite disjoint unions of its elements, this implies that $\mathcal{C}$ actually contains the algebra $a(\mathcal{C}')$ generated by $\mathcal{C}'$. But by Lemma 3, $\mathcal{C}$ is a monotone class; therefore $\mathcal{C}$ must contain the σ-algebra generated by $a(\mathcal{C}')$. Since $\sigma[a(\mathcal{C}')] = \mathcal{B} \supseteq \mathcal{C}$, it follows that $\mathcal{C} = \mathcal{B}$. Thus $(\mu,P_{\vec{f}}) \in \mathcal{A}$.

It is interesting to observe that a $P_{\vec{f}}$-invariant measure μ can have point masses only at certain points; for example, no D-adic rational other than 0 or 1 can have positive μ-measure. However, stationary Markov processes with

$P_{\vec{f}}$-transitions can have "orbits" of the type shown below:



in which case the chain can be decomposed into a chain consisting entirely of orbits and a chain whose absolute distribution is continuous. By belonging to such an orbit a point with cyclic D-ary expansion can have positive μ-measure. Interestingly enough, this is the only situation in which point masses can occur.

<u>Definition</u> <u>10</u>: Let $\mathcal{Q}^{(D)}$ denote the collection of D-adic rationals in $[0,1]$, and let $\mathcal{Q}_0^{(D)} = \mathcal{Q}^{(D)} \cap (0,1)$. Let $\mathcal{I}_0^{(D)} = [0,1] \sim \mathcal{Q}_0^{(D)}$.

<u>Lemma</u> <u>11</u>: Let $\mu \varepsilon \mathcal{M}$ and $\vec{f} \varepsilon \mathcal{F}^{(D)}$ such that $(\mu, P_{\vec{f}}) \varepsilon \mathcal{A}$. Then $\mu\left(\mathcal{Q}_0^{(D)}\right) = 0$.

Proof: Since

$$\phi_i^{-1}(0) = \begin{cases} 0 & \text{if} \quad i = 0 \\ \emptyset & \text{if} \quad 0 < i \leq D-1, \end{cases}$$

it follows that

$$\mu(0) = \sum_{i=0}^{D-1} \int_0^1 f_i(y) \chi_{\{0\}}[\phi_i(y)] \mu(dy) = f_0(0)\mu(0).$$

Hence $\mu(0) > 0$ implies $f_0(0) = 1$. Similarly $\mu(1) > 0$

implies $f_{D-1}(1) = 1$. Let

$$x = \sum_{k=1}^{n} i_k/D^k,$$

where $i_n > 0$. Clearly

$$\phi_i^{-1}\left(\sum_{k=1}^{n-j+1} i_{k+j-1}/D^k\right) = \begin{cases} \sum_{k=1}^{n-j} i_{k+j}/D^k & \text{if} \quad i = i_j \\ \emptyset & \text{if} \quad i \neq i_j \end{cases}$$

for $j=1,\ldots,n-1$; and

$$\phi_i^{-1}(i_n/D) = \begin{cases} 0 & \text{if} \quad i = i_n \\ 1 & \text{if} \quad i = i_n - 1 \\ \emptyset & \text{otherwise.} \end{cases}$$

Therefore

$$\mu(x) = \left[\prod_{j=1}^{n-1} f_{i_j}\left(\sum_{k=1}^{n-j} i_{k+j}/D^k\right)\right] [f_{i_n}(0)\mu(0) + f_{i_n-1}(1)\mu(1)]$$

by repeated application of the invariance equation for μ. But $\mu(0) > 0$ implies $f_{i_n}(0) \leq 1 - f_0(0) = 0$, and $\mu(1) > 0$ implies $f_{i_n-1}(1) \leq 1 - f_{D-1}(1) = 0$; so that

$$f_{i_n}(0)\mu(0) + f_{i_n-1}(1)\mu(1) = 0.$$

Thus $\mu(x) = 0$.

<u>Theorem 12</u>: Let $\mu \in \mathcal{M}$ and $\vec{f} \in \mathcal{F}^{(D)}$ such that $(\mu, P_{\vec{f}}) \in \mathcal{A}$. If $\mu(x_0) > 0$ for some $x_0 \in [0,1]$, then

$$x_0 = \sum_{m=0}^{\infty} \sum_{k=1}^{N} i_k/D^{mN+k}$$

for some finite sequence $(i_1,\ldots,i_N)$ of elements of S_1. Set $i_0 = i_N$ and $i_{N+1} = i_1$, $i_{N+2} = i_2,\ldots,i_{2N-1} = i_{N-1}$. If $x_1,\ldots,x_{N-1}$ are defined by

$$x_j = \sum_{m=0}^{\infty} \sum_{k=1}^{N} i_{k+j}/D^{mN+k} \qquad (j=1,\ldots,N-1),$$

then

$$\mu(x_0) = \mu(x_1) = \ldots = \mu(x_{N-1})$$

and

$$f_{i_j}(x_j) = 1 \qquad (j=0,\ldots,N-1).$$

Proof: Suppose $x_0 \in [0,1]$ such that $\mu(x_0) > 0$. Then $x_0 \in I_0^{(D)}$ by Lemma 11. The elements of $I_0^{(D)}$ have a unique D-ary expansion; hence the function $\psi : I_0^{(D)} \to I_0^{(D)}$ defined by setting

$$\psi\left(\sum_{k=1}^{\infty} i_k/D^k\right) = \sum_{k=1}^{\infty} i_{k+1}/D^k \qquad \left(\sum_{k=1}^{\infty} i_k/D^k \in I_0^{(D)}\right)$$

is well-defined. Note that

$$\phi_i^{-1}\left(\sum_{k=1}^{\infty} i_k/D^k\right) = \begin{cases} \psi\left(\sum_{k=1}^{\infty} i_k/D^k\right) & \text{if} \quad i = i_1 \\ \emptyset & \text{if} \quad i \neq i_1. \end{cases}$$

Therefore, if

$$x_0 = \sum_{k=1}^{\infty} i_k/D^k,$$

then

$$\mu(x_0) = f_{i_1}[\psi(x_0)]\,\mu[\psi(x_0)]$$

$$\leq \mu[\psi(x_0)]$$

$$= f_{i_2}[\psi^2(x_0)]\mu[\psi^2(x_0)]$$

$$\leq \mu[\psi^2(x_0)]$$

$$= \ldots .$$

In short,

$$0 < \mu(x_0) = \mu[\psi^0(x_0)] \leq \mu[\psi^1(x_0)] \leq \mu[\psi^2(x_0)] \leq \ldots .$$

Since μ is a finite measure, the set

$$\{\psi^0(x_0), \psi^1(x_0), \psi^2(x_0), \ldots\}$$

must be finite; so that

$$\psi^j(x_0) = \psi^{j+n}(x_0)$$

for some $j \in Z_0$ and $n \in Z_1$. Set

$$J = \min \{j \in Z_0 : \psi^j(x_0) = \psi^{j+n}(x_0) \text{ for some } n \in Z_1\},$$

and set

$$N = \min \{n \in Z_1 : \psi^J(x_0) = \psi^{J+n}(x_0)\}.$$

Suppose $J > 0$. Since $\psi^J(x_0) = \psi^{J+N}(x_0)$ but $\psi^{J-1}(x_0) \neq \psi^{J+N-1}(x_0)$, it follows that $i_J \neq i_{J+N}$. Now $f_{i_J}[\psi^J(x_0)] > 0$ because

$$0 < \mu[\psi^{J-1}(x_0)] = f_{i_J}[\psi^J(x_0)]\mu[\psi^J(x_0)].$$

Therefore

$$f_{i_{J+N}}[\psi^{J+N}(x_0)] = f_{i_{J+N}}[\psi^{J}(x_0)] \leq 1-f_{i_J}[\psi^J(x_0)] < 1.$$

But then

$$\begin{aligned}\mu[\psi^J(x_0)] &\leq \mu[\psi^{J+N-1}(x_0)] \\ &= f_{i_{J+N}}[\psi^{J+N}(x_0)]\mu[\psi^{J+N}(x_0)] \\ &< \mu[\psi^{J+N}(x_0)] \\ &= \mu[\psi^J(x_0)],\end{aligned}$$

which is contradictory. Thus J = 0 and

$$x_0 = \psi^N(x_0);$$

that is,

$$x_0 = \sum_{k=1}^{\infty} i_k/D^k = \sum_{k=1}^{\infty} i_{k+N}/D^k.$$

From the uniqueness of the expansion and a simple inductive proof it now follows that

$$i_{mN+k} = i_k$$

for $k = 1,\ldots,N$ and $m \in Z_1$. Thus

$$x_0 = \sum_{m=0}^{\infty} \sum_{k=1}^{N} i_k/D^{mN+k}.$$

Next, recall the definition of $x_1,\ldots,x_{N-1}$ in the statement of the theorem, and observe that

$$x_1 = \psi^1(x_0),\ldots,x_{N-1} = \psi^{N-1}(x_0).$$

Since

$$\mu(x_0) \le \mu[\psi^1(x_0)] \le \dots \le \mu[\psi^{N-1}(x_0)]$$
$$\le \mu[\psi^N(x_0)] = \mu(x_0),$$

it follows that

$$\mu(x_0) = \mu(x_1) = \dots = \mu(x_{N-1}).$$

Finally, set $x_{-1} = x_{N-1}$; and note that

$$0 < \mu(x_{j-1}) = f_{i_j}(x_j)\mu(x_j) = f_{i_j}(x_j)\mu(x_{j-1})$$

implies

$$f_{i_j}(x_j) = 1$$

for $j = 0,\dots,N-1$.

Since motion within an orbit is deterministic, we can guarantee the non-existence of orbits by requiring that all components of the D-ary transition vector be less than 1. With no orbits, the chain must then have a continuous absolute distribution.

Corollary 13: Let $\mu \in M$ and $\vec{f} \in F^{(D)}$ such that $(\mu, P_{\vec{f}}) \in A$. If $f_i(x) < 1$ for all $x \in [0,1]$ and $i \in S_1$, then F_μ is continuous.

Proof: If $x \in [0,1]$ such that $\mu(x) > 0$, then by Theorem 12 there is a finite sequence $(i_1,\dots,i_N)$ of elements of S_1 such that

$$x = \sum_{m=0}^{\infty} \sum_{k=1}^{N} i_k/D^k$$

and

$$f_{i_N}(x) = 1.$$

But $f_{i_N}(x) < 1$ by assumption. Therefore $\mu(x) = 0$ and F_μ is continuous.

V. CONSTRUCTION OF STATIONARY CHAINS OF INFINITE ORDER

In this section we define formally the concept of a chain of infinite order, and we give conditions under which a stationary chain of infinite order can be constructed with prescribed transition probabilities. The construction is carried out in detail in Theorem 19.

Definition 14: Let $(\Omega,\mathcal{S},P)$ be a probability space, and let $S = \{c_0,\ldots,c_{D-1}\}$ be a finite set. A doubly infinite sequence $\{Z_n\}_{n=-\infty}^{\infty}$ of random variables on $(\Omega,\mathcal{S},P)$ with values in S is called a chain of infinite order. The chain $\{Z_n\}$ is said to be stationary if the probability

$$P\{\omega \in \Omega : Z_n(\omega) = c_{i_0},\ldots,Z_{n+k}(\omega) = c_{i_k}\}$$

is independent of $n \in Z$ for each choice of $k \in Z_0$ and $i_0,\ldots,i_k \in S_1$.

The next definition makes precise the correspondence between possible past histories in a chain of infinite order and real numbers $x \in [0,1]$. Since D-adic rationals have two D-ary expansions, the correspondence must exclude either those past histories other than $(0,0,\ldots)$ which terminate in 0's or those other than $(D-1,D-1,\ldots)$ which terminate in $(D-1)$'s. Fortunately, as is shown in Lemma 17, a stationary chain of infinite order assigns zero probability to all such sequences, so that no difficulties arise from this discrepancy. We have arbitrarily chosen to exclude sequences terminating in (D-1)'s.

<u>Definition 15</u>: Let $\Omega_1 = \overset{\infty}{\underset{k=1}{\times}} S_1$, and let $\mathcal{S}_1 = \sigma(\{Z_k\}_{k=1}^{\infty})$, where $\{Z_k\}$ is the sequence of co-ordinate functions on Ω_1. Let $\nu : [0,1] \to \Omega_1$ be defined as follows:
$\nu(1) = (D-1, D-1, \ldots)$; and if $0 \leq x < 1$, then
$\nu(x) = (i_1, i_2, \ldots)$, where $.i_1 i_2 \ldots$ is the unique D-ary expansion of x which does not terminate in (D-1)'s.

By means of this correspondence between Ω_1 and [0,1], we may now regard the transition probabilities in a chain of infinite order either as functions Q_i on Ω_1 or as functions f_i on [0,1]. The latter point of view seems conceptually simpler. In either case the transition probabilities should be measurable functions on their domain. The next lemma shows that $\mathcal{S}_1$-measurability of the Q_i's is equivalent to $\mathcal{B}$-measurability of the f_i's.

<u>Lemma 16</u>: Let $Q : \Omega_1 \to \mathcal{R}$ and $f : [0,1] \to \mathcal{R}$ such that

$$f(x) = Q[\nu(x)] \qquad (x \in [0,1]).$$

Then f is $\mathcal{B}$-measurable if and only if Q is $\mathcal{S}_1$-measurable.

Proof: Since

$$\nu^{-1}\{\omega_1 \in \Omega_1 : Z_1(\omega_1) = i_1, \ldots, Z_k(\omega_1) = i_k\}$$

$$= I\left(\sum_{j=1}^{k} i_j D^{k-j},\ k\right) \in \mathcal{B},$$

it follows that ν is a measurable transformation from $([0,1],\mathcal{B})$ to $(\Omega_1,\mathcal{S}_1)$. Hence if Q is $\mathcal{S}_1$-measurable, then f is $\mathcal{B}$-measurable.

Suppose conversely that f is $\mathcal{B}$-measurable. Clearly $\sum_{k=1}^{\infty} Z_k/D^k$ is $\mathcal{S}_1$-measurable, so that $f\left[\sum_{k=1}^{\infty} Z_k/D^k\right]$ is $\mathcal{S}_1$-measurable. Note that

$$\{\omega_1 : f\left[\sum_{k=1}^{\infty} Z_k(\omega_1)/D^k\right] = Q(\omega_1)\} \supseteq A_1,$$

where

$$A_1 = \{(D-1,D-1,\ldots)\} \cup \overline{\lim_{k\to\infty}} \{\omega_1 : Z_k(\omega_1) \neq D-1\} \in \mathcal{S}_1$$

and

$$\Omega_1 \sim A_1 = \bigcup_{k=1} \bigcup_{\substack{i_1,\ldots,i_k \in S_1 \\ i_k \neq D-1}} \{(i_1,\ldots,i_k,D-1,D-1,\ldots)\}$$

For $\alpha \in \mathcal{R}$ set

$$G_\alpha(i_1,\ldots,i_k) = \begin{cases} \{(i_1,\ldots,i_k,D-1,D-1,\ldots)\} \\ \qquad \text{if } Q(i_1,\ldots,i_k,D-1,D-1,\ldots) \leq \alpha \\ \emptyset \quad \text{otherwise.} \end{cases}$$

Clearly $G_\alpha(i_1,\ldots,i_k) \in \mathcal{S}_1$. Hence

$$\{\omega_1 : Q(\omega_1) \leq \alpha\} = C_\alpha^{(1)} \cup C_\alpha^{(2)},$$

where

$$C_\alpha^{(1)} = \{\omega_1 \in A_1 : f\left[\sum_{k=1}^{\infty} Z_k(\omega_1)/D^k\right] \leq \alpha\} \in \mathcal{S}_1$$

and

$$C_{\alpha}^{(2)} = \bigcup_{k=1} \bigcup_{\substack{i_1,\ldots,i_k \varepsilon S_1 \\ i_k \neq D-1}} G_{\alpha}(i_1,\ldots,i_k) \;\varepsilon\; S_1.$$

Thus Q is S_1-measurable.

Lemma 17: Let $\{Z_n\}_{n=-\infty}^{\infty}$ be a stationary chain of infinite order on the probability space (Ω,S,P) with values in S_1. Then

$$P\left(\bigcup_{n=-\infty}^{\infty}\{\omega : Z_n(\omega)\neq D-1,\ Z_{n-1}(\omega)=Z_{n-2}(\omega)=\ldots=D-1\}\right) = 0.$$

Proof: For $i \varepsilon S_1$ let $Q_i = Q_i(i_1,i_2,\ldots)$ be a version of the conditional probability

$$P(Z_n=i \mid Z_{n-1},Z_{n-2},\ldots) = P(Z_n=i \mid Z_{n-1}=i_1,Z_{n-2}=i_2,\ldots).$$

Fix $N \varepsilon Z$, and let

$$C^{(N)} = \bigcap_{k=1}^{\infty} \{\omega : Z_{N-k}(\omega) = D-1\}.$$

Also let

$$C_i^{(N)} = \{\omega \varepsilon C^{(N)} : Z_N(\omega) = i\} \qquad (i \varepsilon S_1),$$

and note that

$$P\left(C_{D-1}^{(N)}\right) = \lim_{k\to\infty} P\{\omega : Z_N(\omega) = \ldots = Z_{N-k+1}(\omega) = D-1\}$$

$$= \lim_{k\to\infty} P\{\omega : Z_{N-1}(\omega) = \ldots = Z_{N-k}(\omega) = D-1\}$$

$$= P\left(C^{(N)}\right)$$

by the stationarity of $\{Z_n\}$. From (*) it follows that

$$(**) \quad P\left(C_i^{(N)}\right) = \int_{C^{(N)}} Q_i[Z_{N-1}(\omega), Z_{N-2}(\omega), \ldots] P(d\omega)$$

$$= \int_{C^{(N)}} Q_i(D-1, D-1, \ldots) P(d\omega)$$

$$= Q_i(D-1, D-1, \ldots) \cdot P\left(C^{(N)}\right);$$

in particular,

$$P\left(C^{(N)}\right) = P\left(C_{D-1}^{(N)}\right) = Q_{D-1}(D-1, D-1, \ldots) \cdot P\left(C^{(N)}\right).$$

Thus either $P(C^{(N)}) = 0$ or $Q_{D-1}(D-1,D-1,\ldots) = 1$; that is, either $P(C^{(N)}) = 0$ or $Q_i(D-1,D-1,\ldots) = 0$ for $i = 0,\ldots,D-2$. In either case, $P\left(C_i^{(N)}\right) = 0$ for $i = 0,\ldots,D-2$ by (**).

Since $N \in Z$ was arbitrary, it follows that

$$P\left(\bigcup_{N=-\infty}^{\infty} \bigcup_{i=0}^{D-2} C_i^{(N)}\right) = 0.$$

The next theorem, which we state without proof, is known as Kolmogorov's Consistency Theorem [5, p. 92]. With its aid we can construct, for any D-ary transition vector $\vec{f}$, a stationary Markov process $\{X_n\}_{n=-\infty}^{\infty}$ on [0,1] with D-ary transition function $P_{\vec{f}}$ and $P_{\vec{f}}$-invariant absolute distribution μ.

<u>Theorem</u> <u>18</u>: Let $\hat{\Omega} = \times_{n=-\infty}^{\infty} [0,1]$, and let $\{X_n\}_{n=-\infty}^{\infty}$ be the

sequence of co-ordinate functions on $\hat{\Omega}$. For $n \in Z$ and $k \in Z_0$ let $\hat{C}^n_{n+k}$ denote the semi-algebra of subsets of $\hat{\Omega}$ of the form

$$\{\hat{\omega} \in \hat{\Omega} : X_n(\hat{\omega}) \in B_n, \ldots, X_{n+k}(\hat{\omega}) \in B_{n+k}\},$$

where $B_n, \ldots, B_{n+k} \in \mathcal{B}$. Suppose there is defined on each $\hat{C}^n_{n+k}$ a set function $\hat{P}^n_{n+k}$ such that

$$\text{(i)} \quad \hat{P}^n_{n+k}(\emptyset) = 0; \ \hat{C} \in \hat{C}^n_{n+k} \to \hat{P}^n_{n+k}(\hat{C}) \geq 0; \ \hat{P}^n_{n+k}(\hat{\Omega}) = 1$$

and

$$\text{(ii)} \quad \hat{C}, \hat{C}_1, \hat{C}_2, \ldots \in \hat{C}^n_{n+k}, \ \hat{C} = \bigcup_{i=1}^{\infty} \hat{C}_i, \ \hat{C}_i \cap \hat{C}_j = \emptyset \ (i \neq j)$$

$$\to \hat{P}^n_{n+k}(\hat{C}) = \sum_{i=1}^{\infty} \hat{P}^n_{n+k}(\hat{C}_i).$$

Suppose further that

$$\text{(iii)} \quad \hat{P}^{n-1}_{n+k} \{\hat{\omega} : X_n(\hat{\omega}) \in B_n, \ldots, X_{n+k}(\hat{\omega}) \in B_{n+k}\}$$

$$= \hat{P}^n_{n+k} \{\hat{\omega} : X_n(\hat{\omega}) \in B_n, \ldots, X_{n+k}(\hat{\omega}) \in B_{n+k}\}$$

$$= \hat{P}^n_{n+k+1} \{\hat{\omega} : X_n(\hat{\omega}) \in B_n, \ldots, X_{n+k}(\hat{\omega}) \in B_{n+k}\}.$$

Then if

$$\hat{C} = \bigcup_{n=-\infty}^{\infty} \bigcup_{k=0}^{\infty} \hat{C}^n_{n+k},$$

there exists a unique probability measure $\hat{P}$ on $\sigma(\hat{C})$ such that

$$\hat{P}(\hat{C}) = \hat{P}^{n}_{n+k}(\hat{C})$$

for $\hat{C} \in \hat{C}^{n}_{n+k}$.

The question whether there exists a (unique) stationary chain of infinite order with prescribed transition probabilities $Q_0,\ldots,Q_{D-1}$ can now be reduced to the question whether there exists a (unique) $P_{\vec{f}}$-invariant measure μ, where $\vec{f} = (Q_0 \circ \nu,\ldots,Q_{D-1} \circ \nu)$.

Theorem 19: Let $Q_0,\ldots,Q_{D-1} : \Omega_1 \to \mathcal{R}_0$ be $\mathcal{S}_1$-measurable functions such that

$$\sum_{i=0}^{D-1} Q_i \equiv 1,$$

and let $\vec{f} = (f_0,\ldots,f_{D-1})$ be defined by setting

$$f_i(x) = Q_i[\nu(x)] \qquad (x \in [0,1];\ i \in S_1).$$

Suppose there exists $\mu \in \mathcal{M}$ such that $(\mu, P_{\vec{f}}) \in \mathcal{A}$. Then there is a probability space $(\Omega,\mathcal{S},\mathcal{P})$ and a chain of infinite order $\{Z_n\}_{n=-\infty}^{\infty}$ on $(\Omega,\mathcal{S},\mathcal{P})$ with values in S_1 such that

(i) $\mathcal{S} = \sigma(\{Z_n\})$

(ii) $\mathcal{P}(Z_n = i \mid Z_{n-1},Z_{n-2},\ldots) = Q_i(Z_{n-1},Z_{n-2},\ldots)$ almost surely on Ω

and

(iii) $\{Z_n\}$ is stationary.

Furthermore, if μ is the only $P_{\vec{f}}$-invariant measure in $\mathcal{M}$, then P is the only probability measure on S satisfying (ii) and (iii).

Proof: Note that $\vec{f} \in F^{(D)}$ by the properties of the Q_i's and by Lemma 16; thus $P_{\vec{f}} \in T$. Define $\hat{\Omega}$, $\{X_n\}$, $\hat{C}$, $\{\hat{C}^n_{n+k}\}$ as in Theorem 18. On $\hat{C}^n_{n+k}$ define $\hat{P}^n_{n+k}$ as follows: If $\hat{C} \in \hat{C}^n_{n+k}$ and

$$\hat{C} = \{\hat{\omega} : X_n(\hat{\omega}) \in B_n, \ldots, X_{n+k}(\hat{\omega}) \in B_{n+k}\},$$

set

$$\hat{P}^n_{n+k}(\hat{C}) = \int_{B_n} \cdots \int_{B_{n+k}} \mu(dx_n) P_{\vec{f}}(x_n, dx_{n+1}) \cdots P_{\vec{f}}(x_{n+k-1}, dx_{n+k}).$$

Clearly (i) of Theorem 18 holds for $\hat{P}^n_{n+k}$. Also (ii) holds, as can be seen from the following observation: $\hat{C}^n_{n+k}$ is a semi-algebra on $\hat{\Omega}$, and Lemma 5 states that $\hat{P}^n_{n+k}$ is countably additive on each countable partition of $\hat{\Omega}$ by sets in $\hat{C}^n_{n+k}$. By a simple and well-known result in measure theory, this implies that $\hat{P}^n_{n+k}$ is countably additive on $\hat{C}^n_{n+k}$. Finally, the first half of (iii) holds for $\{\hat{P}^n_{n+k}\}$ by Lemma 4, and the second half holds trivially. Thus, by Theorem 18, there exists a unique probability measure $\hat{P}$ on $\sigma(\hat{C})$ such that

$$\hat{P}(\hat{C}) = \hat{P}^n_{n+k}(\hat{C})$$

for $\hat{C} \in \hat{C}^n_{n+k}$.

Let $\Omega = \mathop{\times}\limits_{n=-\infty}^{\infty} S_1$, and let $\{Z_n\}_{n=-\infty}^{\infty}$ be the sequence of co-ordinate functions on Ω. Define $\Phi : \Omega \to \hat{\Omega}$ by setting

$$X_n[\Phi(\omega)] = \sum_{k=1}^{\infty} Z_{n-k}(\omega)/D^k \qquad (n \in Z,\ \omega \in \Omega).$$

Φ is 1-1, as the following argument shows: If ω_1, ω_2 are distinct elements of Ω, then

$$Z_{n-1}(\omega_1) \neq Z_{n-1}(\omega_2)$$

for some $n \in Z$. If

$$Z_{n-1}(\omega_1) < Z_{n-1}(\omega_2)$$

but

$$\sum_{k=1}^{\infty} Z_{n-k}(\omega_1)/D^k = \sum_{k=1}^{\infty} Z_{n-k}(\omega_2)/D^k,$$

then

$$Z_{n-1}(\omega_2) - Z_{n-1}(\omega_1) = 1$$

and

$$\sum_{k=1}^{\infty} Z_{n-k-1}(\omega_1)/D^k = \sum_{k=1}^{\infty} (D-1)/D^k = 1$$

$$\neq 0 = \sum_{k=1}^{\infty} 0/D^k = \sum_{k=1}^{\infty} Z_{n-k-1}(\omega_2)/D^k.$$

Thus either

$$X_n[\Phi(\omega_1)] \neq X_n[\Phi(\omega_2)]$$

or

$$X_{n-1}[\Phi(\omega_1)] \neq X_{n-1}[\Phi(\omega_2)],$$

so that

$$\Phi(\omega_1) \neq \Phi(\omega_2).$$

Similarly if $Z_{n-1}(\omega_1) > Z_{n-1}(\omega_2)$, then $\Phi(\omega_1) \neq \Phi(\omega_2)$. Thus Φ is 1-1.

Observe next that

$$\text{range } (\Phi) = \hat{\Omega}_\Phi,$$

where

$$\hat{\Omega}_\Phi = \bigcap_{n=-\infty}^{\infty} \bigcup_{i=0}^{D-1} \{\hat{\omega} : X_n(\hat{\omega}) = [i + X_{n-1}(\hat{\omega})]/D\}.$$

Indeed, since

$$(Z_{n-1}(\omega) + X_{n-1}[\Phi(\omega)])/D = Z_{n-1}(\omega)/D + \sum_{k=2}^{\infty} Z_{n-k}(\omega)/D^k$$
$$= X_n[\Phi(\omega)]$$

for $\omega \in \Omega$ and $n \in Z$, it is clear that range $(\Phi) \subseteq \hat{\Omega}_\Phi$. On the other hand, if $\hat{\omega} \in \hat{\Omega}_\Phi$, let $\omega \in \Omega$ have co-ordinates

$$Z_{n-1}(\omega) = i_{n-1},$$

where

$$X_n(\hat{\omega}) = [i_{n-1} + X_{n-1}(\hat{\omega})]/D.$$

Then

$$\sum_{k=1}^{\infty} Z_{n-k}(\omega)/D^k = \sum_{k=1}^{\infty} [X_{n-k+1}(\hat{\omega})/D^{k-1} - X_{n-k}(\hat{\omega})/D^k]$$

$$= X_n(\hat{\omega}),$$

so that $\hat{\omega} = \Phi(\omega)$. Thus range $(\Phi) \supseteq \hat{\Omega}_\Phi$.

From this it follows that range (Φ) has full $\hat{P}$-measure, since

$$\hat{\Omega}_\Phi = \bigcap_{n=-\infty}^{\infty} \bigcup_{i=0}^{D-1} \bigcap_{N=1}^{\infty} \bigcup_{j=0}^{D^N-1} \{\hat{\omega} : X_{n-1}(\hat{\omega}) \in I(j,N),$$
$$X_n(\hat{\omega}) \in I(iD^N+j,N+1)\}$$

and

$$P\left(\bigcup_{i=0}^{D-1} \bigcup_{j=0}^{D^N-1} \{\hat{\omega} : X_{n-1}(\hat{\omega}) \in I(j,N),\ X_n(\hat{\omega}) \in I(iD^N+j,N+1)\}\right)$$

$$= \sum_{i=0}^{D-1} \sum_{j=0}^{D^N-1} P\{\hat{\omega} : X_{n-1}(\hat{\omega}) \in I(j,N),\ X_n(\hat{\omega}) \in I(iD^N+j,N+1)\}$$

$$= \sum_{i=0}^{D-1} \sum_{j=0}^{D^N-1} \int_{I(j,N)} \mu(dx) P_{\vec{f}}\left(x, I(iD^N+j,N+1)\right)$$

$$= \sum_{i=0}^{D-1} \sum_{j=0}^{D^N-1} \sum_{k=0}^{D-1} \int_0^1 \mu(dx) \chi_{I(j,N)}(x) f_k(x) \chi_{I(iD^N+j,N+1)}[\phi_k(x)]$$

$$= \sum_{i=0}^{D-1} \sum_{j=0}^{D^N-1} f_i(x)\mu[I(j,N)] = 1.$$

Let $S = \sigma(\{Z_n\}_{n=-\infty}^{\infty})$. Then Φ maps S-measurable sets onto $\sigma(\hat{C})$-measurable sets because

$$\Phi(\{\omega \in \Omega : Z_{n-1}(\omega) = i_{n-1}, \ldots, Z_{n+k-1}(\omega) = i_{n+k-1}\})$$

$$= \{\hat{\omega} \in \hat{\Omega}_\Phi : X_n(\hat{\omega}) \in I(i_{n-1},1), \ldots, X_{n+k}(\hat{\omega}) \in I(i_{n+k-1},1)\},$$

which belongs to $\sigma(\hat{C})$. Therefore, since $\hat{P}(\hat{\Omega}_\Phi) = 1$ and since $\Phi : \Omega \to \hat{\Omega}_\Phi$ is 1-1 and onto, the set function $P : S \to R$ defined by setting

$$P(A) = \hat{P}[\Phi(A)] \qquad (A \in S)$$

is a probability measure on S.

Set $S^{(n)} = \sigma(\{Z_{n-k}\}_{k=1}^{\infty})$. Clearly the mapping $\omega \to (Z_{n-1}(\omega), Z_{n-2}(\omega), \ldots)$ is measurable from $(\Omega, S^{(n)})$ into (Ω_1, S_1), and by assumption the mapping $Q_i : \Omega_1 \to R_0$ is S_1-measurable. Therefore the mapping $Q_i(Z_{n-1}, Z_{n-2}, \ldots) : \Omega \to R_0$ is $S^{(n)}$-measurable.

Note that

$$\{\omega : f_i\left[\sum_{k=1}^{\infty} Z_{n-k}(\omega)/D^k\right] = Q_i\left[Z_{n-1}(\omega), Z_{n-2}(\omega), \ldots\right]\} \supseteq A,$$

where

$$A = \left(\bigcap_{k=1}^{\infty}\{\omega : Z_{n-k}(\omega) = D-1\}\right) \cup \left(\overline{\lim_{k\to\infty}}\{\omega : Z_{n-k}(\omega) \neq D-1\}\right)$$

and

$$P(\Omega \sim A) = P\left(\bigcup_{k=1}^{\infty} \bigcup_{\substack{i_1,\ldots,i_k \in S_1 \\ i_k \neq D-1}} \{\omega : Z_{n-1}(\omega) = i_1, \ldots, Z_{n-k}(\omega) = i_k, \right.$$

$$\left. Z_{n-k-1}(\omega) = Z_{n-k-2}(\omega) = \ldots = D-1\}\right)$$

$$= \hat{P}\left(\bigcup_{k=1}^{\infty} \bigcup_{\substack{i_1,\ldots,i_k \in S_1 \\ i_k \neq D-1}} \{\hat{\omega} : X_n(\hat{\omega}) = \sum_{j=1}^{k} i_j/D^j + 1/D^k\}\right)$$

$$= \mu\left(\mathcal{Q}_0^{(D)}\right) = 0$$

by Lemma 11. Hence if

$$C = \{\omega : Z_{n-1}(\omega) = i_1, \ldots, Z_{n-k}(\omega) = i_k\},$$

then

$$\int_C Q_i[Z_{n-1}(\omega), Z_{n-2}(\omega), \ldots] P(d\omega)$$

$$= \int_C f_i\left[\sum_{k=1}^{\infty} Z_{n-k}(\omega)/D^k\right] P(d\omega)$$

$$= \int_{\{\hat{\omega} : X_n(\hat{\omega}) \in I\left(\sum_{j=1}^{k} i_j D^{k-j}, k\right)\}} f_i[X_n(\hat{\omega})]\hat{P}(d\hat{\omega})$$

$$= \int_{I\left(\sum_{j=1}^{k} i_j D^{k-j}, k\right)} f_i(x)\mu(dx)$$

$$= \mu\left[I(iD^k + \sum_{j=1}^{k} i_j D^{k-j},\ k+1)\right]$$

$$= \hat{P}\{\hat{\omega} : X_{n+1}(\hat{\omega}) \in I(iD^k + \sum_{j=1}^{k} i_j D^{k-j},\ k+1)\}$$

$$= P\{\omega : Z_n(\omega) = i,\ Z_{n-1}(\omega) = i_1, \ldots, Z_{n-k}(\omega) = i_k\}$$

$$= P\{\omega \in C : Z_n(\omega) = i\}.$$

Since the sets of the form C generate $S^{(n)}$, it follows that

$$Q_i(Z_{n-1}, Z_{n-2}, \ldots) = P(Z_n = i \mid Z_{n-1}, Z_{n-2}, \ldots)$$

almost surely on Ω. Note also that the probability

$$P(C) = \mu\left[I\left(\sum_{j=1}^{k} i_j D^{k-j}, k\right)\right]$$

is independent of n, which shows that $\{Z_n\}$ is stationary. Thus P and $\{Z_n\}$ satisfy (ii) and (iii) of the theorem. By definition $\{Z_n\}$ satisfies (i).

It remains to prove that P is unique if μ is the only $P_{\vec{f}}$-invariant measure. Let P_1 and P_2 be distinct probability measures on S such that (ii) and (iii) hold with $P = P_1, P_2$. Since

$$\{\omega : Q_i[Z_{n-1}(\omega), Z_{n-2}(\omega), \ldots] \neq f\left[\sum_{k=1}^{\infty} Z_{n-k}(\omega)/D^k\right]\}$$

$$\subseteq \bigcup_{k=1}^{\infty} \{\omega : Z_{n-k}(\omega) \neq D-1,$$

$$Z_{n-k-1}(\omega) = Z_{n-k-2}(\omega) = \ldots = D-1\},$$

it follows from Lemma 17 that

$$P_i\{\omega : Q_i[Z_{n-1}(\omega), Z_{n-2}(\omega), \ldots] = f\left[\sum_{k=1}^{\infty} Z_{n-k}(\omega)/D^k\right]\} = 1$$

for i = 1,2. Now P_1 and P_2 induce probability measures $\hat{P}_1$ and $\hat{P}_2$ on $\sigma(\hat{C})$ by means of the formulas

$$\hat{P}_i(\hat{A}) = P_i[\Phi^{-1}(\hat{A})] \qquad (\hat{A} \in \sigma(\hat{C});\ i = 1,2).$$

For $i_1, \ldots, i_k \in S_1$, observe that the probabilities

$$\hat{P}_i\{\omega : X_n(\hat{\omega}) \in I\left(\sum_{j=1}^{k} i_j D^{k-j}, k\right)\}$$

$$= P_i\{\omega : Z_{n-1}(\omega) = i_1, \ldots, Z_{n-k}(\omega) = i_k\}$$

are independent of n because $\{Z_n\}$ is stationary on (Ω, S, P). Since the intervals $I\left(\sum_{j=1}^{k} i_j D^{k-j},\ k\right)$ generate $\mathcal{B}$, it follows that the probabilities

$$\hat{P}_i\{\hat{\omega} : X_n(\hat{\omega}) \in B\}$$

are independent of n for every $B \in \mathcal{B}$. Set

$$\mu_i(B) = \hat{P}_i\{\hat{\omega} : X_n(\hat{\omega}) \in B\} \qquad (B \in \mathcal{B};\ i = 1,2).$$

Clearly $\mu_1, \mu_2 \in M$. If $k \in S_1$ and $j = \sum_{\ell=1}^{N} i_\ell D^{N-\ell} \in S_N$, set

$$C = \{\omega : Z_{n-1}(\omega) = i_1, \ldots, Z_{n-N}(\omega) = i_N\}$$

and note that

$$\begin{aligned}
\int_{I(j,N)} f_k(x)\mu_i(dx) &= \int_{\{\hat{\omega} : X_n(\hat{\omega}) \in I(j,N)\}} f_k[X_n(\hat{\omega})]\hat{P}_i(d\hat{\omega}) \\
&= \int_C f_k\left[\sum_{\ell=1}^{\infty} Z_{n-\ell}(\omega)/D^\ell\right] P_i(d\omega) \\
&= \int_C Q_k[Z_{n-1}(\omega), Z_{n-2}(\omega), \ldots] P_i(d\omega) \\
&= P_i\{\omega \in C : Z_n(\omega) = k\} \\
&= \hat{P}_i\{\hat{\omega} : X_{n+1}(\hat{\omega}) \in I(kD^N + j, N+1)\} \\
&= \mu[I(kD^N + j, N+1)].
\end{aligned}$$

Hence $(\mu_i, P_{\vec{f}}) \in A$ by Lemma 9. Since P_1 and P_2 are distinct, there exist $i_1, \ldots, i_k \in S_1$ such that

$$P_1\{\omega : Z_{n-1}(\omega) = i_1, \ldots, Z_{n-k}(\omega) = i_k\}$$

$$\neq P_2\{\omega : Z_{n-1}(\omega) = i_1, \ldots, Z_{n-k}(\omega) = i_k\}.$$

But then

$$\mu_1[I\left(\sum_{j=1}^{k} i_j D^{k-j}, k\right)] \neq \mu_2[I\left(\sum_{j=1}^{k} i_j D^{k-j}, k\right)],$$

so that $\mu_1 \neq \mu_2$. Thus there are two distinct $P_{\vec{f}}$-invariant measures.

VI. EXISTENCE THEOREMS FOR $P_{\vec{f}}$-INVARIANT MEASURES

In Section V we showed that the question whether there exists a stationary chain of infinite order with prescribed transition probabilities $Q_0,\ldots,Q_{D-1}$ is equivalent to the question whether there exists a $P_{\vec{f}}$-invariant measure, where $\vec{f} = (Q_0 \circ \nu,\ldots,Q_{D-1} \circ \nu)$. In this section we give conditions on a D-ary transition vector $\vec{f}$ which ensure that there exists a $P_{\vec{f}}$-invariant measure.

We approach the problem from two standpoints, both of which rely upon the notion of weak convergence of measures. In the first approach we use several standard results from functional analysis. Given any D-ary transition vector $\vec{f}$, we can define an operator $\mathcal{U}$ from $\mathcal{M}$ into $\mathcal{M}$ by the formula

$$\mathcal{U}\mu(B) = \int_0^1 \mu(dx) P_{\vec{f}}(x,B) \qquad (B \in \mathcal{B}).$$

Clearly every fixed-point of $\mathcal{U}$ is a $P_{\vec{f}}$-invariant measure. Now $\mathcal{M}$ may be regarded as a convex weak*-compact subset of the space of all bounded linear functionals on C[0,1], where C[0,1] is the Banach space of continuous functions on [0,1] with the sup norm. As an operator on this space, $\mathcal{U}$ turns out to be weak*-continuous. By invoking a well-known fixed-point theorem, we conclude that $\mathcal{U}$ has a fixed point whenever $\vec{f}$ has continuous components.

The following lemma, a proof of which can be found in [2, p. 456], records the fixed-point theorem in the form in which we shall use it.

Lemma 20: Let K be a compact convex subset of a linear topological space X, and let τ be a continuous linear operator on X such that $\tau(K) \subseteq K$. Then there exists $L \in K$ such that $\tau(L) = L$.

Lemma 21: Let $\vec{f} \in F^{(D)}$. Let F be regarded as the Banach space of bounded $\mathcal{B}$-measurable functions on $[0,1]$ with the sup norm, and let T be the linear operator on F defined by

$$Tg(x) = \int_0^1 P_{\vec{f}}(x,dy)g(y) \qquad (x \in [0,1];\ g \in F).$$

Let H be a closed subspace of F such that $T(H) \subseteq H$, and let T_H be the restriction of T to H. Let H^* be the adjoint space of H, and let T_H^* be the adjoint operator of T_H. Then if K is any weak*-closed convex subset of the unit ball in H^* such that $T_H^*(K) \subseteq K$, there exists $L \in K$ such that $T_H^*(L) = L$.

Proof: Since

$$\sup_{0\le x\le 1} |Tg(x)| \le \sup_{0\le x\le 1} \int_0^1 P_{\vec{f}}(x,dy) \sup_{0\le t\le 1} |g(t)| = ||g||,$$

the operator T is a bounded linear operator on F; hence T_H is a bounded linear operator on H. The adjoint operator T_H^* of T_H is a weak*-continuous linear operator on H^*. By Alaoglu's Theorem [2, p. 424] K is weak*-compact in H^*. Thus the theorem follows from Lemma 20 by setting

$X = H^*$ with the weak*-topology and $\tau = T^*_H$.

<u>Theorem</u> <u>22</u>: Let $\vec{f} \in F^{(D)}$ such that each f_i is continuous on $[0,1]$. Then there exists $\mu \in M$ such that $(\mu, P_{\vec{f}}) \in A$.

Proof: Let H be the collection of continuous functions on $[0,1]$. Then H with the sup norm is a closed subspace of F, and H^* may be regarded as the collection of countably additive set functions on B with finite total variation. For $g \in H$ and $\mu \in H^*$ set

$$\langle g,\mu\rangle = \int_0^1 g(x)\mu(dx).$$

Let T be defined as in Lemma 21, and observe that

$$Tg(x) = \sum_{i=0}^{D-1} f_i(x) g[\phi_i(x)] \qquad (x \in [0,1];\ g \in H).$$

Since $f_0,\ldots,f_{D-1}, \phi_0,\ldots,\phi_{D-1} \in H$, it follows that $Tg \in H$ when $g \in H$, i.e., $T(H) \subseteq H$.

Let $K = M \subseteq H^*$. Clearly K is convex; also K is a subset of the unit ball in H^*, because $||\mu|| = 1$ for $\mu \in K$. Moreover, K is weak*-closed in H^*, since

$$K = (\{\mu \in H^* : \langle 1,\mu\rangle = 1\}) \cap \Big(\bigcap_{\substack{g \in H \\ g \geq 0}} \{\mu \in H^* : \langle g,\mu\rangle \geq 0\}\Big).$$

Now let T_H and T^*_H be defined as in Lemma 21. For $\mu \in K$ set

$$\mu_1(\cdot) = \int_0^1 \mu(dx)P(x,\cdot) \in K,$$

and observe for $\mu \in K$ and $g \in H$ that

$$\begin{aligned} \langle g, T_H^*\mu\rangle &= \langle T_H g, \mu\rangle \\ &= \int_0^1 \mu(dx) \int_0^1 P(x,dy)g(y) \\ &= \int_0^1 g(y)\mu_1(dy) = \langle g, \mu_1\rangle \end{aligned}$$

by a simple variant of Fubini's Theorem. Thus $T_H^*\mu = \mu_1 \in K$ if $\mu \in K$, i.e., $T_H^*(K) \subseteq K$.

By Lemma 21 there exists $\mu \in K$ such that $\mu = T_H^*\mu = \mu_1$. Hence

$$\mu(B) = \mu_1(B) = \int_0^1 \mu(dx)P_{\vec{f}}(x,B)$$

for every $B \in \mathcal{B}$, and $(\mu, P_{\vec{f}}) \in A$.

By Theorem 19, the result of Theorem 22 implies the existence of stationary chains of infinite order with transition probabilities satisfying certain continuity conditions. These chains are described in Corollary 24. First we define the natural notion of convergence in Ω_1.

<u>Definition 23</u>: Let $\{i^{(n)}\}_{n=1}^{\infty}$ be a sequence of elements of Ω_1, where $i^{(n)} = (i_1^{(n)}, i_2^{(n)}, \ldots)$ for each n. Then $\{i^{(n)}\}_{n=1}^{\infty}$ is said to <u>converge</u> to $i^{(0)} = (i_1^{(0)}, i_2^{(0)}, \ldots) \in \Omega_1$ (in symbols, <u>$i^{(n)} \to i^{(0)}$</u>) if, for each k, $i_k^{(n)} = i_k^{(0)}$ for all n sufficiently large.

Corollary 24: Let $Q_0,\ldots,Q_{D-1} : \Omega_1 \to \mathcal{R}_0$ such that

(i) $\sum_{i=0}^{D-1} Q_i \equiv 1$

(ii) $i^{(n)} \to i^{(0)}$ implies $Q_i(i^{(n)}) \to Q_i(i^{(0)})$ $(i \in S_1)$

(iii) $Q_i(i_1,\ldots,i_k,D\text{-}1,D\text{-}1,\ldots) = Q_i(i_1,\ldots,i_k+1,0,0,\ldots)$

$(i \in S_1;\ i_k \neq D\text{-}1)$.

Then there exists a stationary chain of infinite order $\{\mathcal{Z}_n\}_{n=-\infty}^{\infty}$ on a probability space $(\Omega,\mathcal{S},\mathcal{P})$ with values in S_1 such that

$$\mathcal{P}(\mathcal{Z}_n = i \mid \mathcal{Z}_{n-1},\mathcal{Z}_{n-2},\ldots) = Q_i(\mathcal{Z}_{n-1},\mathcal{Z}_{n-2},\ldots) \quad (i \in S_1)$$

almost surely on Ω.

Proof: Conditions (ii) and (iii) ensure that the vector $\vec{f} = (Q_0 \circ \nu,\ldots,Q_{D-1} \circ \nu)$ has continuous components; therefore, by Theorem 22 there exists a $P_{\vec{f}}$-invariant measure μ. Also by Lemma 16 the functions $Q_0,\ldots,Q_{D-1}$ are $\mathcal{S}_1$-measurable. Hence the corollary follows from Theorem 19.

Condition (ii) of Corollary 24 seems natural. If at time n two "past histories" in a chain of infinite order coincide back to time n-k, where k is large, it seems reasonable to assume that they have nearly the same transition probabilities. On the other hand, condition (iii) appears quite unnatural. In general there is no reason to assume that two distinct "past histories" should have exactly the same transition probabilities. Thus the class of D-ary transition vectors with continuous components corresponds to a somewhat restricted class of transition probabilities for chains of infinite order. The techniques of functional analysis do permit a generalization of Theorem 22 asserting the existence of an invariant measure for every D-ary transition function; however, in this generality the measure may be only finitely additive.

Theorem 25: For every $\vec{f} \in F^{(D)}$ there exists a finitely additive probability measure μ on $\mathcal{B}$ such that

$$\int_0^1 \mu(dx) P_{\vec{f}}(x,B) = \mu(B)$$

for $B \in \mathcal{B}$.

Proof: Let $H = F$, so that $H^* = F^*$ may be regarded as the collection of finitely additive set functions on $\mathcal{B}$ with finite total variation. Clearly $T(H) \subseteq H$. Let K be the collection of finitely additive probability measures on $\mathcal{B}$. As in the proof of Theorem 22, K can be shown to satisfy the hypotheses of Lemma 21; hence there exists $\mu \in K$ such that $T^*\mu = \mu$, i.e., such that

$$\int_0^1 \mu(dx) P_{\vec{f}}(x,B) = \mu(B)$$

for $B \in \mathcal{B}$.

A simple but interesting example of a D-ary transition vector $\vec{f}$ all of whose invariant measures are only finitely additive is the following: Let D = 2, and let $\vec{f} = (f_0, f_1)$, where $f_0 = \chi_{(0,1]}$ and $f_1 = \chi_{\{0\}}$. If μ is an invariant measure for $\vec{f}$, then

$$\mu(0) = f_0(0)\cdot\mu(0) = 0\cdot\mu(0) = 0;$$

hence for $B \in \mathcal{B}$ we have

$$\begin{aligned}\mu(B) &= \sum_{i=0}^{1} \int_0^1 f_i(x)\chi_B[\phi_i(x)]\mu(dx) \\ &= \int_0^1 \left[\chi_{\phi_0^{-1}(B)\cap(0,1]}(x) + \chi_{\phi_1^{-1}(B)\cap\{0\}}(x)\right]\mu(dx) \\ &= \mu[\phi_0^{-1}(B)].\end{aligned}$$

Now if $B \cap (0,\delta) = (0,\delta)$ for some $\delta \in (0,1)$, then $\phi_0^{-n}(B) \supseteq (0,1]$ for some $n \in Z_1$; so that

$$\mu(B) = \mu[\phi_0^{-1}(B)] = \ldots = \mu[\phi_0^{-n}(B)] \geq \mu((0,1]) = 1.$$

Regarding μ as a continuous linear functional on $\mathcal{F}$, we have that

$$\langle\chi_B,\mu\rangle = 1.$$

Since $\langle\cdot,\mu\rangle$ is linear, it readily follows that

$$\langle f,\mu\rangle = c$$

for every simple function $f \in \mathcal{F}$ such that $f(x) = c$ for x in some interval $(0,\delta)$. Using the continuity of $\langle\cdot,\mu\rangle$, we can then show that

$$\langle f,\mu\rangle = c$$

for every function $f \in F$ such that $f(x) \to c$ as $x \downarrow 0$. This shows how $\langle\cdot,\mu\rangle$ is defined on the subspace of F consisting of all functions $f \in F$ having a right-hand limit at 0. Clearly μ is not countably additive, since

$$\lim_{n\to\infty} \mu((0,1/n)) = \lim_{n\to\infty} \langle\chi_{(0,1/n)},\mu\rangle = 1 \neq 0 = \mu(\emptyset).$$

This completes our study of the existence of invariant measures by means of fixed-point theorems in functional analysis. Henceforth we adopt a different point of view. Roughly our approach is as follows: Suppose that $\vec{f}^{(n)}$ is a sequence of D-ary transition vectors converging in some sense to a D-ary transition vector $\vec{f}$. Suppose further that each vector $\vec{f}^{(n)}$ has a $P_{\vec{f}^{(n)}}$-invariant measure μ_n. Is there a subsequence of $\{\mu_n\}$ which converges in some sense to a $P_{\vec{f}}$-invariant measure μ? With suitable definitions of convergence in F and M, the answer is affirmative; and a general result of this type is given in Theorem 30.

The next definition specifies the type of convergence in M appropriate to our task. This is actually weak*-convergence in M considered as $C[0,1]^*$.

<u>Definition 26</u>: A sequence $\mu_1,\mu_2,\ldots \in M$ is said to <u>converge weakly</u> to $\mu \in M$ (in symbols, $\mu_n \xrightarrow{w} \mu$) if

$$\int_0^1 f(x)\mu_n(dx) \to \int_0^1 f(x)\mu(dx)$$

for every continuous function f on [0,1].

The next two propositions, which we state without proof (see [5, p. 116] and [3, p. 261] respectively), give familiar properties of weakly convergent sequences of measures.

Proposition 27: Let $\mu,\mu_1,\mu_2,\ldots \varepsilon\ M$. Then

$$\mu_n \xrightarrow{w} \mu$$

if and only if

$$F_{\mu_n}(x) \to F_\mu(x)$$

at every point $x\ \varepsilon\ [0,1]$ at which F_μ is continuous.

Proposition 28: If $\mu_1,\mu_2,\ldots \varepsilon\ M$, there exists a subsequence $\mu_{n_1},\mu_{n_2},\ldots$ and a measure $\mu\ \varepsilon\ M$ such that $\mu_{n_k} \xrightarrow{w} \mu$.

The following proposition concerning weak convergence of measures is less familiar, and we prove it here for the sake of completeness. It is quoted in [5, p. 119].

Proposition 29: Let $\mu,\mu_1,\mu_2,\ldots \varepsilon\ M$ such that $\mu_n \xrightarrow{w} \mu$. If $B\ \varepsilon\ \mathcal{B}$ such that $\mu(\partial B) = 0$, then

$$\int_B f(x)\mu_n(dx) \to \int_B f(x)\mu(dx)$$

for every continuous function f on [0,1].

Proof: Without loss of generality assume that $0 \le f \le 1$, and let $\varepsilon > 0$ be given. By the regularity of μ there exist a closed set H and an open set $\mathcal{O}$ in [0,1] such that

$$H \subseteq \mathring{B} \subseteq \overline{B} \subseteq \mathcal{O}$$

and

$$\max\{\mu(\mathring{B} \sim H),\ \mu(\mathcal{O} \sim \overline{B})\} < \varepsilon.$$

By Urysohn's Lemma [2, p. 15] there exist continuous functions $g_1, g_2 : [0,1] \to [0,1]$ such that

$$g_1(H) \equiv g_2(\overline{B}) \equiv 1$$

and

$$g_1([0,1] \sim \mathring{B}) \equiv g_2([0,1] \sim \mathcal{O}) \equiv 0.$$

Clearly

$$\chi_H \le g_1 \le \chi_B \le g_2 \le \chi_{\mathcal{O}}.$$

Note that

$$\int_{B\sim H} f(x)\mu(dx) \le \mu(B \sim H) \le \mu(\mathring{B} \sim H) + \mu(\partial B) < \varepsilon$$

and

$$\int_{\mathcal{O}\sim B} f(x)\mu(dx) \le \mu(\mathcal{O} \sim B) \le \mu(\mathcal{O} \sim \overline{B}) + \mu(\partial B) < \varepsilon.$$

Also

$$\int_0^1 f(x)g_i(x)\mu_n(dx) \to \int_0^1 f(x)g_i(x)\mu(dx) \qquad (i = 1,2)$$

because fg_1, fg_2 are continuous. Hence

$$\int_B f(x)\mu(dx) - \varepsilon < \int_H f(x)\mu(dx) \le \int_0^1 f(x)g_1(x)\mu(dx)$$
$$= \lim_{n\to\infty} \int_0^1 f(x)g_1(x)\mu_n(dx) \le \underline{\lim}_{n\to\infty} \int_B f(x)\mu_n(dx)$$

and

$$\int_B f(x)\mu(dx) + \varepsilon > \int_O f(x)\mu(dx) \ge \int_0^1 f(x)g_2(x)\mu(dx)$$
$$= \lim_{n\to\infty} \int_0^1 f(x)g_2(x)\mu_n(dx) \ge \overline{\lim}_{n\to\infty} \int_B f(x)\mu_n(dx).$$

Since $\varepsilon > 0$ was arbitrary, it follows that

$$\overline{\lim}_{n\to\infty} \int_B f(x)\mu_n(dx) \le \int_B f(x)\mu(dx) \le \underline{\lim}_{n\to\infty} \int_B f(x)\mu_n(dx).$$

Thus

$$\int_B f(x)\mu_n(dx) \to \int_B f(x)\mu(dx).$$

With these results we can prove the following general existence theorem for $P_{\vec{f}}$-invariant measures.

Theorem 30: Let $\mu_1,\mu_2,\ldots \varepsilon\ \mathcal{M}$ and $f^{(1)},f^{(2)},\ldots \varepsilon\ \mathcal{F}^{(D)}$ such that $(\mu_n, P_{\vec{f}^{(n)}})\ \varepsilon\ \mathcal{A}$ for each n. Suppose $\mu_n \xrightarrow{W} \mu\ \varepsilon\ \mathcal{M}$ such that F_μ is continuous, and suppose $B_1,B_2,\ldots \varepsilon\ \mathcal{B}$ are disjoint sets such that

(i) $\mu\left(\bigcup_{m=1}^{\infty} B_m\right) = 1$

and

(ii) $\mu(\partial B_1) = \mu(\partial B_2) = \ldots = 0.$

Suppose there exists $\vec{f} \in F^{(D)}$ such that

(iii) $f_i^{(n)} \to f_i$ uniformly on B_m $(m \in Z_1;\ i \in S_1)$,

and suppose for each m there exist functions $g_0^{(m)}, \ldots, g_{D-1}^{(m)}$ continuous on the set $\overline{B}_m$ such that

$$g_i^{(m)}(x) = f_i(x) \qquad (x \in B_m;\ i \in S_1).$$

Then $(\mu, P_{\vec{f}}) \in A$.

Proof: Note first that $g_0^{(m)}, \ldots, g_{D-1}^{(m)}$ can be extended to continuous functions $h_0^{(m)}, \ldots, h_{D-1}^{(m)}$ on [0,1] by Tietze's Extension Theorem [2, p. 15].

Let $\varepsilon > 0$ be given. By (i) there exists $M \in Z_1$ such that

$$\mu\left(\bigcup_{m=1}^{M} B_m\right) > 1 - \varepsilon/4.$$

Set $B^{(M)} = \bigcup_{m=1}^{M} B_m$, and observe that

$$\mu(\partial B^{(M)}) \leq \sum_{m=1}^{M} \mu(\partial B_m) = 0$$

by (ii). Thus

$$\mu_n(B^{(M)}) = \int_{B^{(M)}} 1\ \mu_n(dx) \to \int_{B^{(M)}} 1\ \mu(dx) = \mu(B^{(M)})$$

by Proposition 29. Hence there exists $M' \varepsilon Z_1$ such that

$$\mu_n(B^{(M)}) > 1 - \varepsilon/4 \qquad (n > M').$$

From (iii) it follows that $f_i^{(n)} \to f_i$ uniformly on $B^{(M)}$ for each $i \varepsilon S_1$. Let $M'' \varepsilon Z_1$ such that

$$\sup_{i\varepsilon S_1} \sup_{x\varepsilon B^{(M)}} | f_i(x) - f_i^{(n)}(x) | < \varepsilon/4 \qquad (n > M'').$$

Now if $N \varepsilon Z_0$ and $j \varepsilon S_N$, then

$$\mu[\partial I(j,N)] = 0$$

because F_μ is continuous; therefore

$$\mu[\partial(I(j,N) \cap B_m)] \leq \mu[\partial I(j,N)] + \mu[\partial B_m] = 0$$

for $m = 1,\ldots,M$ by (ii). Since $h_0^{(m)},\ldots,h_{D-1}^{(m)}$ are continuous on $[0,1]$, it follows that

$$\int_{I(j,N)\cap B_m} f_i(x)\mu_n(dx) = \int_{I(j,N)\cap B_m} h_i^{(m)}(x)\mu_n(dx)$$

$$\to \int_{I(j,N)\cap B_m} h_i^{(m)}(x)\mu(dx) = \int_{I(j,N)\cap B_m} f_i(x)\mu(dx)$$

by an application of Proposition 29. Thus there exist integers $N_1,\ldots,N_M \varepsilon Z_1$ such that

$$\sup_{i\varepsilon S_1} \left| \int_{I(j,N)\cap B_m} f_i(x)\mu(dx) - \int_{I(j,N)\cap B_m} f_i(x)\mu_n(dx) \right| < \varepsilon/4M \qquad (n > N_m)$$

for m = 1,...,M. For any $i \varepsilon S_1$ and $n > \max\{M', M'', N_1, \ldots, N_M\}$ it now follows that

$$\left| \int_{I(j,N)} f_i(x)\mu(dx) - \int_{I(j,N)} f_i^{(n)}(x)\mu_n(dx) \right|$$

$$\leq \left| \int_{I(j,N)\cap B^{(M)}} f_i(x)\mu(dx) - \int_{I(j,N)\cap B^{(M)}} f_i(x)\mu_n(dx) \right|$$

$$+ \left| \int_{I(j,N)\cap B^{(M)}} f_i(x)\mu_n(dx) - \int_{I(j,N)\cap B^{(M)}} f_i^{(n)}(x)\mu_n(dx) \right|$$

$$+ \left| \int_{I(j,N)\sim B^{(M)}} f_i(x)\mu(dx) - \int_{I(j,N)\sim B^{(M)}} f_i^{(n)}(x)\mu_n(dx) \right|$$

$$\leq \sum_{m=1}^{M} \left| \int_{I(j,N)\cap B_m} f_i(x)\mu(dx) - \int_{I(j,N)\cap B_m} f_i(x)\mu_n(dx) \right|$$

$$+ \int_{I(j,N)\cap B^{(M)}} \left| f_i(x) - f_i^{(n)}(x) \right| \mu_n(dx)$$

$$+ \int_{I(j,N)\sim B^{(M)}} f_i(x)\mu(dx) + \int_{I(j,N)\sim B^{(M)}} f_i^{(n)}(x)\mu_n(dx)$$

$$< \sum_{m=1}^{M} (\varepsilon/4M) + \int_{B^{(M)}} \left[\sup_{x\varepsilon B^{(M)}} |f_i(x) - f_i^{(n)}(x)|\right] \mu_n(dx)$$

$$+ \int_{[0,1]\sim B^{(M)}} 1\ \mu(dx) + \int_{[0,1]\sim B^{(M)}} 1\ \mu_n(dx)$$

$$< \varepsilon/4 + (\varepsilon/4) \cdot \sup_{n\varepsilon Z_1} \mu_n(B^{(M)})$$

$$+ \mu([0,1] \sim B^{(M)}) + \sup_{n>M'} \mu_n([0,1] \sim B^{(M)})$$

$$< \varepsilon/4 + \varepsilon/4 + \varepsilon/4 + \varepsilon/4 = \varepsilon.$$

This shows that

$$\int_{I(j,N)} f_i^{(n)}(x) \mu_n(dx) \to \int_{I(j,N)} f_i(x) \mu(dx).$$

But $\mu[\partial I(iD^N+j,N+1)] = 0$ and $(\mu_n, P_{\vec{f}(n)}) \varepsilon A$ for each n, so that

$$\mu[I(iD^N+j,N+1)] = \lim_{n\to\infty} \mu_n[I(iD^N+j,N+1)]$$

$$= \lim_{n\to\infty} \int_{I(j,N)} f_i^{(n)}(x) \mu_n(dx)$$

$$= \int_{I(j,N)} f_i(x) \mu(dx).$$

Hence $(\mu, P_{\vec{f}}) \in A$ by Lemma 9.

One drawback of Theorem 30 is that the limit measure μ must have a continuous distribution F_μ. Corollary 32 gives general assumptions on the $\vec{f}^{(n)}$'s which ensure that F_μ will be continuous. The corollary depends on the following proposition.

Proposition 31: Let $1/D \leq \xi < 1$, and let $\{\vec{f}^{(\alpha)}\}_{\alpha \varepsilon A}$ be a family of D-ary transition vectors such that

$$\text{range } (f_i^{(\alpha)}) \subseteq [0,\xi] \qquad (i \in S_1;\ \alpha \in A).$$

If $\{\mu_\alpha\}_{\alpha \varepsilon A}$ is a family of measures in M such that

$$(\mu_\alpha, P_{\vec{f}^{(\alpha)}}) \in A \qquad (\alpha \in A),$$

then the family of distributions $\{F_{\mu_\alpha}\}_{\alpha \varepsilon A}$ is uniformly equicontinuous.

Proof: For $N \in Z_0$ let $P(N)$ be the proposition that

$$\mu_\alpha[I(j,N)] \leq \xi^N \qquad (j \in S_N;\ \alpha \in A).$$

Clearly $P(0)$ holds. If $P(N)$ holds and $k \in S_{N+1}$, then $k = iD^N + j$ for some $i \in S_1$ and $j \in S_N$; and

$$\mu_\alpha[I(k,N+1)] = \int_{I(j,N)} f_i(x)\mu_\alpha(dx)$$

$$\leq \xi \cdot \mu_\alpha[I(j,N)] \leq \xi^{N+1},$$

so that $P(N+1)$ holds. By induction $P(N)$ holds for all $N \in Z_0$.

Given $\varepsilon > 0$, set $\delta = 1/D^N$, where $2\xi^N < \varepsilon$. If $x,y \in [0,1]$ such that $0 \leq y - x < \delta$, then

$$x,y \in I(j,N) \cup I(j+1,N)$$

for some $0 \leq j < D^N-1$; and

$$\begin{aligned} 0 &\leq F_{\mu_\alpha}(y) - F_{\mu_\alpha}(x) \\ &\leq \mu_\alpha[I(j,N)] + \mu_\alpha[I(j+1,N)] \\ &\leq 2\xi^N < \varepsilon \end{aligned}$$

for all $\alpha \in A$. Thus $\{F_{\mu_\alpha}\}_{\alpha\varepsilon A}$ is uniformly equicontinuous.

<u>Corollary 32</u>: Let $1/D \leq \xi < 1$, and let $\vec{f}^{(1)}, \vec{f}^{(2)}, \ldots \varepsilon\ F^{(D)}$ such that

$$\text{range } (f_i^{(n)}) \subseteq [0,\xi] \qquad (i \in S_1;\ n \in Z_1).$$

Let $\mu, \mu_1, \mu_2, \ldots \varepsilon\ M$ such that

$$(\mu_n, P_{\vec{f}(n)}) \in A \qquad (n \in Z_1)$$

and such that $\mu_n \xrightarrow{W} \mu$. Then F_μ is uniformly continuous.

Proof: By Proposition 31 the family $\{F_{\mu_n}\}_{n=1}^{\infty}$ is uniformly equicontinuous; hence, given $\varepsilon > 0$, there exists $\delta > 0$ such that

$$| F_{\mu_n}(x) - F_{\mu_n}(y) | < \varepsilon \qquad (n \in Z_1)$$

whenever $x,y \in [0,1]$ such that $|x-y| < \delta$. By Proposition 27 and the monotonicity of F_μ there is a countable set $B \subseteq [0,1]$ such that

$$F_{\mu_n}(x) \to F_\mu(x) \qquad (x \in [0,1] \sim B).$$

If $x,y \in B^c = [0,1] \sim B$ such that $|x-y| < \delta$, it follows that

$$| F_\mu(x) - F_\mu(y) | = \lim_{n\to\infty} | F_{\mu_n}(x) - F_{\mu_n}(y) | \leq \varepsilon.$$

Now B^c is dense in $[0,1]$ because B is countable. Therefore, if $x,y \in [0,1]$ such that $0 \leq y - x < \delta$, there exist points $x_1, y_1 \in B^c$ such that $x_1 \leq x \leq y \leq y_1$ and $y_1 - x_1 < \delta$; and

$$0 \leq F_\mu(y) - F_\mu(x) \leq F_\mu(y_1) - F_\mu(x_1) \leq \varepsilon.$$

Thus F_μ is uniformly continuous.

We conclude Section VI by making a simple application of Theorem 30. This corollary will be useful in the next section, where we give a new proof of the existence and uniqueness of stationary absolute distributions for a wide class of Nth order Markov chains.

<u>Corollary 33</u>: Let $1/D \leq \xi < 1$ and $N \in Z_0$, and let

$\vec{f}^{(0)}, \vec{f}^{(1)}, \vec{f}^{(2)}, \ldots \varepsilon\ \mathcal{F}^{(D)}$ such that

$$\text{range}\ (f_i^{(n)}) \subseteq [0,\xi] \qquad (i \in S_1;\ n \in Z_0)$$

and such that each function $f_i^{(n)}$ is constant on the intervals $I(j,N)$, where $j = 0,\ldots,D^N-1$. Suppose that

$$f_i^{(n)}(x) \to f_i^{(0)}(x) \qquad (x \in [0,1];\ i \in S_1),$$

and suppose there exist $\mu_1,\mu_2,\ldots \varepsilon\ \mathcal{M}$ such that $(\mu_n, P_{\vec{f}(n)}) \in A$ for $n \in Z_1$. Then there exists $\mu_0 \in \mathcal{M}$ such that $(\mu_0, P_{\vec{f}(0)}) \in A$.

Proof: By Proposition 28 there exists a subsequence $\{\mu_{n_k}\}_{k=1}^{\infty}$ of $\{\mu_n\}_{n=1}^{\infty}$ such that $\mu_{n_k} \xrightarrow{w} \mu_0$ for some $\mu_0 \in \mathcal{M}$. Observe that F_{μ_0} is continuous by Corollary 32. For $m=1,\ldots,D^N$ set $B_m = I(m-1,N)$. Then

$$\mu\left(\bigcup_{m=1}^{D^N} B_m\right) = \mu([0,1]) = 1;$$

also

$$\mu(\partial B_1) = \ldots = \mu(\partial B_{D^N}) = 0$$

because F_{μ_0} is continuous. Now the functions $f_i^{(0)}, f_i^{(n_1)}, f_i^{(n_2)}, \ldots$ are step functions with the same intervals of constancy; hence $f_i^{(n_k)} \to f_i^{(0)}$ pointwise on $[0,1]$ implies that $f_i^{(n_k)} \to f_i^{(0)}$ uniformly on $[0,1]$. In particular, $f_i^{(n_k)} \to f_i^{(0)}$ on each set B_m. Furthermore, each function $f_i^{(0)}$, $i \in S_1$, can be extended to a constant (hence continuous) function $g_i^{(m)}$ on B_m by setting

$$g_i^{(m)}(x) \equiv f_i^{(0)}[(m-1)/D^N] \qquad (x \in \overline{B}_m).$$

Thus $(\mu_0, P_{\vec{f}(0)}) \in A$ by Theorem 30.

VII. STATIONARY ABSOLUTE DISTRIBUTIONS FOR NTH ORDER MARKOV CHAINS

If S is a denumerable set and $\{X_n\}$ is an Nth order Markov chain on S, then $\{X_n\}$ can be regarded in a natural way as a Markov chain $\{X'_n\}$ on the set S' consisting of all N-tuples from S. The transition probabilities

$$P'(X'_n = (i_0,\ldots,i_{N-1}) \mid X'_{n-1} = (i_1,\ldots,i_N))$$

in the new process are merely equated to the transition probabilities

$$P(X_n = i_0 \mid X_{n-1} = i_1,\ldots,X_{n-N} = i_N)$$

in the original process, and all other X'_n-transitions are assigned zero probability. If S is finite and if all transitions are possible in the X_n-process, the X'_n-process is an irreducible aperiodic chain in which all states are persistent. (For definitions see [6, p. 52].) Then from the Erdös-Pollard-Feller Theorem [6, p. 57] it follows that $\{X'_n\}$ (hence $\{X_n\}$) has a unique stationary absolute distribution.

In this section we give a new proof of the fact that a finite-state Nth order Markov chain in which all transitions are possible has a unique stationary absolute distribution. Such a chain may be regarded as a chain of infinite order in which the transition probabilities

$Q_0, \ldots, Q_{D-1}$ depend upon only their first N arguments. The functions $f_0 = Q_0 \circ \nu, \ldots, f_{D-1} = Q_{D-1} \circ \nu$ are then constant on each of the intervals $I(0,N), \ldots, I(D^N-1,N)$. As is proved in Lemma 37, a $P_{\vec{f}}$-invariant measure μ for such a vector $\vec{f} = (Q_0 \circ \nu, \ldots, Q_{D-1} \circ \nu)$ is completely determined by its values on the intervals $I(0,N), \ldots, I(D^N-1,N)$; hence the problem of establishing the existence and uniqueness of μ reduces to the problem of proving the non-singularity of a certain $D^N \times D^N$ matrix $A(\vec{f})$. In Lemma 36 we demonstrate the linear independence of the first D^N-1 rows of $A(\vec{f})$; from this the non-singularity of $A(\vec{f})$ follows readily. The main result of the section is contained in Corollary 39.

__Definition 34__: A function $f \in F$ is said to be a __D-ary step function of order N__ ($N \in Z_1$) if f is constant on each of the intervals $I(0,N), \ldots, I(D^N-1,N)$. A vector $\vec{f} \in F^{(D)}$ such that each f_i is a D-ary step function of order N is called a __Markovian transition vector of order N__. Let $K_N^{(D)}$ be the collection of those vectors in $F^{(D)}$ which are Markovian transition vectors of order N. If $\vec{f} \in K_N^{(D)}$, the constant value of f_i on $I\left(\sum_{k=1}^{N} i_k D^{N-k}, N\right)$ will be denoted by $f(i; i_1, \ldots, i_N)$.

__Definition 35__: For $N \in Z_1$ and $i_1, \ldots, i_N \in S_1$ set

$$\lambda(i_1, \ldots, i_N) = \sum_{k=1}^{N} i_k D^{N-k} + 1.$$

Then for each $\vec{f} \in K_N^{(D)}$ let $A(\vec{f})$ be the $D^N \times D^N$ matrix whose rows

$$\vec{R}_1 = \vec{R}_{\lambda(0,\ldots,0)}, \ \ldots, \ \vec{R}_{D^N} = \vec{R}_{\lambda(D-1,\ldots,D-1)}$$

are given by

$$\vec{R}_{\lambda(i_1,\ldots,i_N)} = \vec{\varepsilon}_{\lambda(i_1,\ldots,i_N)} - \sum_{i=0}^{D-1} f(i_1;\ i_2,\ldots,i_N,\ i)\vec{\varepsilon}_{\lambda(i_2,\ldots,i_N,\ i)}$$

$$(1 \le \lambda(i_1,\ldots,i_N) < D^N)$$

and

$$\vec{R}_{\lambda(D-1,\ldots,D-1)} = \sum_{i_1,\ldots,i_N \in S_1} \vec{\varepsilon}_{\lambda(i_1,\ldots,i_N)},$$

where $\vec{\varepsilon}_1,\ldots,\vec{\varepsilon}_{D^N}$ are the unit co-ordinate vectors in $\mathcal{R}^{D^N}$.

<u>Lemma</u> <u>36</u>: Let $\vec{f} \in K_N^{(D)}$ such that

$$\text{range } (f_i) \subseteq (0,1] \qquad (i \in S_1).$$

Then the rows $\vec{R}_1,\ldots,\vec{R}_{D^N-1}$ of $A(\vec{f})$ are linearly independent.

Proof: Suppose $c_1,\ldots,c_{D^N-1} \in \mathcal{R}$ such that

$$\sum_{j=1}^{D^N-1} c_j\vec{R}_j = \theta$$

where θ is the zero vector in $\mathcal{R}^{D^N}$. Equating components gives

$$c_{\lambda(i_1,\ldots,i_N)} - \sum_{i=0}^{D-1} f(i;i_1,\ldots,i_N)c_{\lambda(i,i_1,\ldots,i_{N-1})} = 0$$

$$(1 \le \lambda(i_1,\ldots,i_N) \le D^N-D),$$

$$c_{\lambda(i_1,\ldots,i_N)} - \sum_{i=0}^{D-2} f(i;i_1,\ldots,i_N)c_{\lambda(i,i_1,\ldots,i_{N-1})} = 0$$

$$(D^N-D < \lambda(i_1,\ldots,i_N) < D^N),$$

and

$$- \sum_{i=0}^{D-2} f(i;D-1,\ldots,D-1)c_{\lambda(i,D-1,\ldots,D-1)} = 0.$$

Thus if $c_{D^N} = c_{\lambda(D-1,\ldots,D-1)}$ is defined to be 0, then

$$c_{\lambda(i_1,\ldots,i_N)} = \sum_{i=0}^{D-1} f(i;i_1,\ldots,i_N)c_{\lambda(i,i_1,\ldots,i_{N-1})}$$

$$(1 \le \lambda(i_1,\ldots,i_N) \le D^N).$$

Since

$$f(i;i_1,\ldots,i_N) > 0$$

and

$$\sum_{i=0}^{D-1} f(i;i_1,\ldots,i_N) = 1,$$

it follows that

$$(*) \quad \min_{i \varepsilon S_1} c_{\lambda(i,i_1,\ldots,i_{N-1})}$$

$$= \sum_{i=0}^{D-1} f(i;i_1,\ldots,i_N)\left[\min_{i \varepsilon S_1} c_{\lambda(i,i_1,\ldots,i_{N-1})}\right]$$

$$\leq \sum_{i=0}^{D-1} f(i;i_1,\ldots,i_N)\, c_{\lambda(i,i_1,\ldots,i_{N-1})}$$

$$= c_{\lambda(i_1,\ldots,i_N)},$$

with equality if and only if

$$(**) \quad c_{\lambda(i,i_1,\ldots,i_{N-1})} = c_{\lambda(i_1,\ldots,i_N)} \qquad (i \in S_1).$$

Let $V = \{c_j\}_{j=1}^{D^N}$. Let $c_{\lambda(j_1,\ldots,j_N)} = \min\{c : c \in V\}$, and let $c_{\lambda(k_1,\ldots,k_N)}$ be arbitrary in V. By (*)

$$\min_{k \in S_1} c_{\lambda(k,j_1,\ldots,j_{N-1})} \leq c_{\lambda(j_1,\ldots,j_N)},$$

and equality holds by definition of $c_{\lambda(j_1,\ldots,j_N)}$. Therefore, by (**),

$$c_{\lambda(k_N,j_1,\ldots,j_{N-1})} = c_{\lambda(j_1,\ldots,j_N)}.$$

Repetition of the argument gives

$$c_{\lambda(k_{N-1},k_N,j_1,\ldots,j_{N-2})} = c_{\lambda(k_N,j_1,\ldots,j_{N-1})}$$

$$\vdots$$

$$c_{\lambda(k_1,\ldots,k_N)} = c_{\lambda(k_2,\ldots,k_N,j_1)}.$$

Thus

$$c_{\lambda(k_1,\ldots,k_N)} = c_{\lambda(j_1,\ldots,j_N)}.$$

Since $c_{\lambda(k_1,\ldots,k_N)}$ was arbitrary in V, it follows that all elements of V are equal. Their common value must be 0, because $c_{D^N} = 0$.

<u>Lemma</u> <u>37</u>: Let $\vec{f} \in K_N^{(D)}$ such that

$$\text{range}\,(f_i) \subseteq (0,1] \qquad (i \in S_1),$$

and let $M(\vec{f})$ be the collection of $P_{\vec{f}}$-invariant measures $\mu \in M$. Let L be the collection of $D^N \times 1$ column vectors of real numbers, and let $L_0 \subseteq L$ be the collection of positive solutions X to the matrix equation

$$A(\vec{f})\cdot X = Y,$$

where Y is the $D^N \times 1$ column vector with final entry 1 and other entries 0. Then the function $\Phi : M(\vec{f}) \to L$ given by

$$\Phi(\mu) = \begin{bmatrix} \mu[I(0,N)] \\ \vdots \\ \mu[I(D^N-1,N)] \end{bmatrix} \qquad (\mu \in M(\vec{f}))$$

is a 1-1 mapping of $M(\vec{f})$ onto L_0.

Proof: To see that Φ maps $M(\vec{f})$ into L_0, let $\mu \in M(\vec{f})$ and set $X = \Phi(\mu)$. The positivity of X follows readily from the invariance equation and the positivity of the step functions $f_0,\ldots,f_{D-1}$. If X has components $x_1,\ldots,x_{D^N}$, then

$$\vec{R}_{\lambda(i_1,\ldots,i_N)}\cdot X$$

$$= x_{\lambda(i_1,\ldots,i_N)} - \sum_{i=0}^{D-1} f(i_1;i_2,\ldots,i_N,i)\, x_{\lambda(i_2,\ldots,i_N,i)}$$

$$= \mu[I(\sum_{k=1}^{N} i_k D^{N-k}, N)]$$

$$- \sum_{i=0}^{D-1} f(i_1;i_2,\ldots,i_N,i)\,\mu[I(\sum_{k=1}^{N-1} i_{k+1} D^{N-k} + i, N)]$$

$$= \mu[I(\sum_{k=1}^{N} i_k D^{N-k}, N)]$$

$$- \sum_{i=0}^{D-1} \int_{I(\sum_{k=1}^{N-1} i_{k+1} D^{N-k} + i,\, N)} f_{i_1}(x)\,\mu(dx)$$

$$= \mu[I(\sum_{k=1}^{N} i_k D^{N-k}, N)]$$

$$- \sum_{i=0}^{D-1} \mu[I(\sum_{k=1}^{N} i_k D^{N+1-k} + i, N+1)]$$

$$= \mu[I(\sum_{k=1}^{N} i_k D^{N-k}, N)] - \mu[I(\sum_{k=1}^{N} i_k D^{N-k}, N)]$$

$$= 0 \qquad (1 \le \lambda(i_1,\ldots,i_N) < D^N)$$

and

$$\vec{R}_{\lambda(D-1,\ldots,D-1)} \cdot X = \sum_{j=1}^{D^N} x_j = \sum_{j=0}^{D^N-1} \mu[I(j,N)] = 1.$$

Thus $\Phi(\mu) \varepsilon L_0$. This shows that Φ maps $M(\vec{f})$ into L_0.

To see that Φ maps $M(\vec{f})$ onto L_0, let $X \varepsilon L_0$ with components $x_1,\ldots,x_{D^N}$. Note that

$$\sum_{1\le\lambda(i_1,\ldots,i_N)\le D^N}\ \sum_{i=0}^{D-1} f(i_1;i_2,\ldots,i_N,i)\,x_{\lambda(i_2,\ldots,i_N,i)}$$

$$= \sum_{1\le\lambda(i_2,\ldots,i_N,i)\le D^N} \left[\sum_{i_1=0}^{D-1} f(i_1;i_2,\ldots,i_N,i)\right] \cdot x_{\lambda(i_2,\ldots,i_N,i)}$$

$$= \sum_{1\le\lambda(i_2,\ldots,i_N,i)\le D^N} x_{\lambda(i_2,\ldots,i_N,i)}$$

$$= \sum_{1\le\lambda(i_1,\ldots,i_N)\le D^N} x_{\lambda(i_1,\ldots,i_N)}.$$

Since $X \in L_0$,

$$\sum_{i=0}^{D-1} f(i_1;i_2,\ldots,i_N,i)\,x_{\lambda(i_2,\ldots,i_N,i)} = x_{\lambda(i_1,\ldots,i_N)}$$

$$(1 \le \lambda(i_1,\ldots,i_N) < D^N).$$

Adding these last D^N-1 equations and subtracting the result from the preceding equation gives

$$\sum_{i=0}^{D-1} f(D-1;D-1,..,D-1,i)\,x_{\lambda(D-1,..,D-1,i)} = x_{\lambda(D-1,..,D-1)}.$$

Thus

$$x_{\lambda(i_1,\ldots,i_N)} = \sum_{i=0}^{D-1} f\,(i_1;i_2,\ldots,i_N,i)\,x_{\lambda(i_2,\ldots,i_N,i)}$$

$$(1 \le \lambda(i_1,\ldots,i_N) \le D^N).$$

Now let C be the collection of intervals $I(j,N+k)$, where

$k \in Z_0$ and $j \in S_{N+k}$; and let $\hat{\mu}$ be the set function on C defined by setting

$$\hat{\mu}[I(\sum_{j=1}^{N+k} i_j D^{N+k-j}, N+k)] = [\prod_{j=1}^{k} f(i_j; i_{j+1}, \ldots, i_{N+j})] x_{\lambda(i_{k+1}, \ldots, i_{N+k})},$$

where an empty product (in case $k = 0$) is defined to have the value 1. Clearly $\hat{\mu}$ is non-negative on C. Also $\hat{\mu}$ is additive on C, as follows from the equations

$$\sum_{i_{N+k+1}, \ldots, i_{N+k+\ell} \varepsilon S_1} \hat{\mu}[I(\sum_{j=1}^{N+k+\ell} i_j D^{N+k+\ell-j}, N+k+\ell)]$$

$$= \sum_{i_{N+k+1}, \ldots, i_{N+k+\ell} \varepsilon S_1} [\prod_{j=1}^{k+\ell} f(i_j; i_{j+1}, \ldots, i_{N+j})] \cdot x_{\lambda(i_{k+\ell+1}, \ldots, i_{N+k+\ell})}$$

$$= \sum_{i_{N+k+1}, \ldots, i_{N+k+\ell-1} \varepsilon S_1} [\prod_{j=1}^{k+\ell-1} f(i_j; i_{j+1}, \ldots, i_{N+j})] \cdot [\sum_{i_{N+k+\ell} \varepsilon S_1} f(i_{k+\ell}; i_{k+\ell+1}, \ldots, i_{N+k+\ell}) \cdot x_{\lambda(i_{k+\ell+1}, \ldots, i_{N+k+\ell})}]$$

$$= \sum_{i_{N+k+1}, \ldots, i_{N+k+\ell-1} \varepsilon S_1} [\prod_{j=1}^{k+\ell-1} f(i_j; i_{j+1}, \ldots, i_{N+j})] \cdot x_{\lambda(i_{k+\ell}, \ldots, i_{N+k+\ell-1})}$$

$$= \cdots = \sum_{i_{N+k+1} \varepsilon S_1} [\prod_{j=1}^{k+1} f(i_j; i_{j+1}, \ldots, i_{N+j})] \cdot x_{\lambda(i_{k+2}, \ldots, i_{N+k+1})}$$

$$= [\prod_{j=1}^{k} f(i_j; i_{j+1}, \ldots, i_{N+j})] \cdot [\sum_{i_{N+k+1} \varepsilon S_1} f(i_{k+1}; i_{k+2}, \ldots, i_{N+k+1}) \cdot x_{\lambda(i_{k+2}, \ldots, i_{N+k+1})}]$$

$$= [\prod_{j=1}^{k} f(i_j; i_{j+1}, \ldots, i_{N+j})] x_{\lambda(i_{k+1}, \ldots, i_{N+k})}$$

$$= \hat{\mu}[I(\sum_{j=1}^{N+k} i_j D^{N+k-j}, N+k)].$$

Two other important properties of $\hat{\mu}$ are

$$(*) \quad \sum_{j \varepsilon S_N} \hat{\mu}[I(j,N)] = \sum_{j=1}^{D^N} x_j = 1$$

and

$$(**) \quad \hat{\mu}[I(\sum_{j=1}^{N+k} i_j D^{N+k-j}, N+k)] \le \xi^k,$$

where

$$\xi = \max \{f_i(x) : 0 \le x \le 1; i \varepsilon S_1\}.$$

Property (**) follows from the definition of $\hat{\mu}$ and the fact that each $x_j \le 1$. Since no f_i is ever 0, no f_i is ever 1; hence $\xi < 1$ because the f_i's are step functions.

Now let $F_{\hat{\mu}} : \mathcal{Q}^{(D)} \to \mathcal{R}$ be defined by setting $F_{\hat{\mu}}(0) = 0$ and $F_{\hat{\mu}}(1) = 1$ and

$$F_{\hat{\mu}}(k/D^N) = \sum_{j=0}^{k-1} \hat{\mu}[I(j,N)]$$

whenever $1 \le k < D^N$ such that $k \nmid D$. From the non-negativity and additivity of $\hat{\mu}$ on $\mathcal{C}$, and from (*) and (**), it follows that $F_{\hat{\mu}}$ is non-negative, non-decreasing, and uniformly continuous on $\mathcal{Q}^{(D)}$; hence $F_{\hat{\mu}}$ has a unique extension to a continuous distribution function F_μ on [0,1]. Let μ be the unique measure in $\mathcal{M}$ such that

$$\mu([0,x)) = F_\mu(x) \qquad (x \in [0,1]).$$

By construction

$$\mu[I(j,N)] = \hat{\mu}[I(j,N)] = x_{j+1} \qquad (j \in S_N),$$

so that $\Phi(\mu) = X$. Also if $i \in S_1$ and

$$j = \sum_{\ell=1}^{N+k} i_\ell D^{N+k-\ell} \quad \varepsilon \; S_{N+k},$$

then

$$\begin{aligned}
&\mu[I(iD^{N+k}+j,N+k+1)] \\
&\quad = f(i;i_1,\ldots,i_N) \\
&\qquad \cdot \left[\prod_{\ell=1}^{k} f(i_\ell;i_{\ell+1},\ldots,i_{N+\ell})\right] \cdot x_{\lambda(i_{k+1},\ldots,i_{N+k})} \\
&\quad = f(i;i_1,\ldots,i_N)\mu[I(j,N+k)] \\
&\quad = \int_{I(j,N+k)} f_i(x)\mu(dx),
\end{aligned}$$

so that $\mu \in M(\vec{f})$ by a slight variant of Lemma 9. Thus Φ maps $M(\vec{f})$ onto L_0.

It remains to show that Φ is 1-1. If μ_1 and μ_2 are elements of $M(\vec{f})$ which agree on the intervals $I(0,N),\ldots,I(D^N-1,N)$, then they agree on all of C (hence on all of $\mathcal{B}$) because they are both invariant measures for a vector $\vec{f}$ whose components are D-ary step functions of order N. Thus $\Phi(\mu_1) = \Phi(\mu_2)$ implies $\mu_1 = \mu_2$, which shows that Φ is 1-1.

<u>Theorem</u> <u>38</u>: Let $\vec{f} \in K_N^{(D)}$ such that

$$\text{range } (f_i) \subseteq (0,1] \qquad (i \in S_1).$$

Then there exists a unique $\mu \in M$ such that $(\mu, P_{\vec{f}}) \in A$.

Proof: The collection $H_N^{(D)}$ of vectors $\vec{f} = (f_0,\ldots,f_{D-1})$ such that each f_i is a D-ary step function of order N is a Banach space under the natural norm

$$||\vec{f}|| = \max \{||f_i|| : i \in S_1\},$$

where $||f_i||$ denotes the sup norm of f_i. Let $\hat{K} = \hat{K}_N^{(D)}$ denote the collection of vectors $\vec{f} \in K_N^{(D)}$ such that range $(f_i) \subseteq (0,1]$ for $i \in S_1$. Clearly $\hat{K}$ is connected. Let K' denote the collection of those vectors $\vec{f} \in \hat{K}$ for which there exists a unique $P_{\vec{f}}$-invariant measure $\mu \in M$. The theorem asserts that $K' = \hat{K}$. Since $\hat{K}$ is connected, it suffices to show that K' is non-empty and open and closed in

$\hat{K}$. If $\vec{f} \in \hat{K}$ is the vector whose components are identically equal to 1/D, it is easy to see that Lebesgue measure is the only $P_{\vec{f}}$-invariant measure in M. Thus $K' \neq \emptyset$.

To see that K' is open in $\hat{K}$, let $\vec{f} \in K'$. By Lemma 37 there is a unique positive column vector $X \in L$ such that $A(\vec{f}) \cdot X = Y$. Then $[A(\vec{f})]^{-1}$ exists and $X = [A(\vec{f})]^{-1} \cdot Y$. By the continuity of the determinant function and of matrix multiplication there is a neighborhood N of $\vec{f}$ in $\hat{K}$ such that $\vec{g} \in N$ implies that $[A(\vec{g})]^{-1}$ exists and that $[A(\vec{g})]^{-1} \cdot Y$ has positive components. By Lemma 37, $N \subseteq K'$. Thus K' is open in $\hat{K}$.

To see that K' is closed in $\hat{K}$, let $\vec{f}^{(1)}, \vec{f}^{(2)}, \ldots$ be a sequence of vectors in K' converging to a vector $\vec{f} \in \hat{K}$. Eventually the vectors $\vec{f}^{(n)}$ must have components bounded uniformly away from 1; hence by Corollary 33 there exists $\mu \in M$ such that $(\mu, P_{\vec{f}}) \in A$. Thus there exists

$$X^{(0)} = \begin{pmatrix} x_1^{(0)} \\ \cdot \\ \cdot \\ \cdot \\ x_{D^N}^{(0)} \end{pmatrix} \in L_0$$

by Lemma 37. Since by Lemma 36 the rows $R_1, \ldots, R_{D^N-1}$ of $A(\vec{f})$ are linearly independent, the solution set $\{X\}$ of the matrix equation

$$\begin{pmatrix} \vec{R}_1 \\ \cdot \\ \cdot \\ \cdot \\ \vec{R}_{D^N-1} \end{pmatrix} \cdot X = \begin{pmatrix} 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{pmatrix}$$

is a line L through the origin in R^{D^N}. Clearly $(x_1^{(0)},\ldots,x_{D^N}^{(0)}) \in L$. But if $(\lambda x_1^{(0)},\ldots,\lambda x_{D^N}^{(0)})$ is any point of L whose co-ordinates sum to 1, then

$$\lambda = \lambda \sum_{j=1}^{D^N} x_j^{(0)} = \sum_{j=1}^{D^N} \lambda x_j^{(0)} = 1;$$

so that $(\lambda x_1^{(0)},\ldots,\lambda x_{D^N}^{(0)}) = (x_1^{(0)},\ldots,x_{D^N}^{(0)})$. This shows that $X^{(0)}$ is the unique solution X of the equation $A(\vec{f})\cdot X = Y$. Thus L_0 has exactly one element, so that $\vec{f}$ has exactly one $P_{\vec{f}}$-invariant measure by Lemma 37. Hence $\vec{f} \in K'$, and K' is closed in $\hat{K}$.

<u>Corollary 39</u>: Let $Q_0,\ldots,Q_{D-1} : \Omega_1 \to R_0 \sim \{0\}$ such that

$$\sum_{i=0}^{D-1} Q_i \equiv 1$$

and such that each function Q_i depends upon only a finite number of its arguments. Then there exists a measurable space (Ω,S) and a sequence of S-measurable functions $\{Z_n\}_{n=-\infty}^{\infty}$ on Ω with $S = \sigma(\{Z_n\})$ such that there is exactly one probability measure P on S for which

(i) $\{Z_n\}$ is stationary

and

(ii) $P(Z_n = i \mid Z_{n-1}, Z_{n-2}, \ldots) = Q_i(Z_{n-1}, Z_{n-2}, \ldots)$

almost surely on Ω $(i \in S_1)$.

Proof: By assumption on the Q_i's there exists $N \in Z_1$ such that the components f_i of the vector $\vec{f} = (Q_0 \circ \nu, \ldots, Q_{D-1} \circ \nu)$ are D-ary step functions of order N satisfying range $(f_i) \subseteq (0,1]$. By Theorem 38 there exists a unique $P_{\vec{f}}$-invariant measure. Also the functions $Q_0, \ldots, Q_{D-1}$ are S_1-measurable by Lemma 16. Hence the corollary follows from Theorem 19.

VIII. ADDITIONAL EXISTENCE THEOREMS FOR $P_{\vec{f}}$-INVARIANT MEASURES

In this section we apply the results of Sections VI and VII to obtain further existence theorems for $P_{\vec{f}}$-invariant measures. As a consequence, we obtain further existence theorems for stationary chains of infinite order. In particular, we prove the existence of a stationary chain of infinite order for every set $Q_0,\ldots,Q_{D-1}$ of transition probabilities satisfying the natural continuity condition (ii) of Corollary 24, provided the Q_i's are bounded away from 1. We prove a similar result for Q_i's which are, in a sense, "locally Markovian."

We begin the section by introducing, for each D-ary vector $\vec{f}$, a sequence $\varepsilon_1(\vec{f}),\varepsilon_2(\vec{f}),\ldots$ which gives a sort of modulus of continuity for the components of $\vec{f}$.

<u>Definition 40</u>: For each vector $\vec{f} = (f_0,\ldots,f_{D-1})$ with components in F let $\varepsilon_1(\vec{f}),\varepsilon_2(\vec{f}),\ldots$ be the sequence of real numbers defined by

$$\varepsilon_m(\vec{f}) = \max_{i\varepsilon S_1} \max_{j\varepsilon S_m} \sup_{x,y\varepsilon I(j,m)} | f_i(x) - f_i(y) | \qquad (m \varepsilon Z_1).$$

In [4, p. 712] T. E. Harris showed that a D-ary transition vector $\vec{f}$ with at least one component bounded away from 0 has a $P_{\vec{f}}$-invariant measure whenever

$$\sum_{m=1}^{\infty} \prod_{k=1}^{m} [1 - (D/2)\varepsilon_k(\vec{f})] = \infty.$$

Under the slightly different assumption that every f_i is bounded away from 1, we relax Harris' condition to the weaker condition

$$\varepsilon_m(\vec{f}) \downarrow 0.$$

Of considerable use is the following proposition, which characterizes those D-ary transition vectors $\vec{f}$ for which $\{\varepsilon_m(\vec{f})\}$ forms a null sequence.

Proposition 41: For $\vec{f} \in F^{(D)}$ the following three statements are equivalent:

(i) The functions $f_0,\ldots,f_{D-1}$ are continuous on $I_0^{(D)}$ and right-continuous on $Q_0^{(D)}$, with a left-hand limit at each point of $Q_0^{(D)}$;

(ii) $\varepsilon_m(\vec{f}) \downarrow 0$ as $m \to \infty$;

(iii) There is a sequence of vectors $\vec{f}^{(1)} \in K_1^{(D)}$, $\vec{f}^{(2)} \in K_2^{(D)},\ldots$ such that $f_i^{(n)} \to f_i$ uniformly on $[0,1]$ for each $i \in S_1$.

Proof: Clearly the sequence $\varepsilon_1(\vec{f}),\varepsilon_2(\vec{f}),\ldots$ is always non-increasing. To see that (i) $\to$ (ii), assume that

$$\varepsilon_m(\vec{f}) \downarrow \varepsilon > 0.$$

Then for some $i \in S_1$ there is a sequence of intervals

$I(j_1,1), I(j_2,2), \ldots$ with $j_k \in S_k$ such that

$$\sup_{x,y \in I(j_k,k)} | f_i(x) - f_i(y) | \geq \varepsilon \qquad (k \in Z_1).$$

Let x_0 be any cluster point of the sequence $j_1/D, j_2/D^2, \ldots,$ and let $0 < \varepsilon' < \varepsilon$. If f_i had right- and left-hand limits at x_0, there would exist intervals $J_1 = (x_0 - \delta, x_0) \cap [0,1]$ and $J_2 = (x_0, x_0 + \delta) \cap [0,1]$ such that

$$x,y \in J_k \rightarrow | f_i(x) - f_i(y) | < \varepsilon' \qquad (k = 1,2).$$

But by definition of x_0 there is an interval $I(j_N,N)$ contained in either J_1 or J_2; and for this interval

$$\sup_{x,y \in I(j_N,N)} | f_i(x) - f_i(y) | \geq \varepsilon > \varepsilon'.$$

Thus either $\lim_{x \uparrow x_0} f_i(x)$ or $\lim_{x \downarrow x_0} f_i(x)$ does not exist. By contraposition (i) $\rightarrow$ (ii).

To see that (ii) $\rightarrow$ (iii), let $\varepsilon_m(\vec{f}) \downarrow 0$ and let $\vec{f}^{(1)}, \vec{f}^{(2)}, \ldots$ be defined by

$$f_i^{(N)}(x) = f_i(j/D^N) \qquad (x \in I(j,N);\ j \in S_N,\ N \in Z_1).$$

Clearly $\vec{f}^{(N)} \in K_N^{(D)}$ for each N. Furthermore if $\varepsilon > 0$ be given and $\varepsilon_{N_0}(\vec{f}) < \varepsilon$, then for $n > N_0$

$$\max_{i \in S_1} \sup_{0 \leq x \leq 1} | f_i(x) - f_i^{(n)}(x) |$$

$$= \max_{i \in S_1} \max_{j \in S_n} \sup_{x \in I(j,n)} | f_i(x) - f_i(j/D^n) |$$

$$\leq \varepsilon_n \leq \varepsilon_{N_0} < \varepsilon;$$

thus $f_i^{(n)} \to f_i$ uniformly on $[0,1]$ for each $i \in S_1$.

To see that (iii) → (i), assume that (iii) holds for $\vec{f}$, and let $x_0 = j/D^N$ for some $N \in Z_1$ and $0 \leq j < D^N$. Given $\varepsilon > 0$, let $N' > N$ such that

$$\max_{i \in S_1} \sup_{0 \leq x \leq 1} | f_i(x) - f_i^{(N')}(x) | < \varepsilon/2.$$

Then $x_0 = j'/D^{N'}$ for some $0 \leq j < D^{N'}$; and if $i \in S_1$ and $x \in I(j',N')$,

$$\begin{aligned} | f_i(x_0) - f_i(x) | &\leq | f_i(x_0) - f_i^{(N')}(x_0) | \\ &\quad + | f_i^{(N')}(x_0) - f_i^{(N')}(x) | \\ &\quad + | f_i^{(N')}(x) - f_i(x) | \\ &< \varepsilon/2 + 0 + \varepsilon/2 = \varepsilon. \end{aligned}$$

It follows that each f_i is right-continuous on $Q_0^{(D)}$. The other assertions in (i) are verified in a similar fashion.

<u>Theorem</u> <u>42</u>: Let $1/D \leq \xi < 1$, and let $\vec{f} \in F^{(D)}$ such that

$$\text{range}(f_i) \subseteq [0,\xi] \qquad (i \in S_1)$$

and such that one of the conditions (i), (ii), (iii) of Proposition 41 holds for $\vec{f}$. Then there exists $\mu \in M$ such that $(\mu, P_{\vec{f}}) \in A$.

Proof: By Proposition 41 there is a sequence of vectors $\vec{f}^{(1)} \in K_1^{(D)}, \vec{f}^{(2)} \in K_2^{(D)}, \ldots$ such that $f_i^{(n)} \to f_i$ uniformly on $[0,1]$ for each $i \in S_1$. If necessary, the $\vec{f}^{(n)}$'s may be modified so that

$$\text{range } (f_i^{(n)}) \subseteq (0,\xi] \qquad (i \in S_1;\ n \in Z_1).$$

By Theorem 38 there is a sequence $\mu_1, \mu_2, \ldots \in M$ such that $(\mu_n, P_{\vec{f}^{(n)}}) \in A$ for each n. By Proposition 28 some subsequence $\{\mu_{n_k}\}$ of $\{\mu_n\}$ converges weakly to a measure $\mu \in M$, and F_μ is continuous by Corollary 32. In order to show that $(\mu, P_{\vec{f}}) \in A$, it suffices (by Lemma 9 and the closing remarks in the proof of Theorem 30) to show that

$$\int_{I(j,N)} f_i^{(n_k)}(x)\mu_{n_k}(dx) \to \int_{I(j,N)} f_i(x)\mu(dx)$$

$$(j \in S_N,\ N \in Z_0;\ i \in S_1).$$

Fix $i \in S_1$, $N \in Z_0$, $j \in S_N$, and let $\varepsilon > 0$ be given. Choose $K \in Z_1$ such that

$$\sup_{0 \le x \le 1} \mid f_i(x) - f_i^{(n_k)}(x) \mid < \varepsilon/3 \qquad (k \ge K).$$

Note that this implies

$$\sup_{0 \le x \le 1} \mid f_i^{(n_K)}(x) - f_i^{(n_k)}(x) \mid < 2\varepsilon/3 \qquad (k > K).$$

For some $L \in Z_0$ there is a decomposition of $I(j,N)$ into intervals $I(jD^L,N+L),\ldots,I(jD^L+D^L-1,N+L)$ on which $f_i^{(n_K)}$ is constant. On each of these intervals $f_i^{(n_K)}$ can be extended to a constant (hence continuous) function on [0,1]. Since F_μ is continuous, these intervals have boundaries with zero μ-measure. Therefore, by Proposition 29,

$$\int_{I(j,N)} f_i^{(n_K)}(x)\mu_{n_k}(dx) = \sum_{\ell=0}^{D^L-1} \int_{I(jD^L+\ell,N+L)} f^{(n_K)}(x)\mu_{n_k}(dx)$$

$$\to \sum_{\ell=0}^{D^L-1} \int_{I(jD^L+k,N+L)} f^{(n_K)}(x)\mu(dx) = \int_{I(j,N)} f^{(n_K)}(x)\mu(dx).$$

Consequently,

$$\overline{\lim_{k\to\infty}} \left| \int_{I(j,N)} f_i(x)\mu(dx) - \int_{I(j,N)} f_i^{(n_k)}(x)\mu_{n_k}(dx) \right|$$

$$\le \overline{\lim_{k\to\infty}} \left| \int_{I(j,N)} f_i(x)\mu(dx) - \int_{I(j,N)} f_i^{(n_K)}(x)\mu(dx) \right|$$

$$+ \overline{\lim_{k\to\infty}} \left| \int_{I(j,N)} f_i^{(n_K)}(x)\mu(dx) - \int_{I(j,N)} f_i^{(n_K)}(x)\mu_{n_k}(dx) \right|$$

$$+ \overline{\lim_{k\to\infty}} \left| \int_{I(j,N)} f_i^{(n_K)}(x)\mu_{n_k}(dx) - \int_{I(j,N)} f_i^{(n_k)}(x)\mu_{n_k}(dx) \right|$$

$$\le \int_{I(j,N)} \left| f_i(x) - f_i^{(n_K)}(x) \right| \mu(dx)$$

$$+ \ 0$$

$$+ \overline{\lim_{k\to\infty}} \int_{I(j,N)} | \, f_i^{(n_K)}(x) - f_i^{(n_k)}(x) \, | \, \mu_{n_k}(dx)$$

$$< \varepsilon/3 + 2\varepsilon/3 = \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, it follows that

$$\int_{I(j,N)} f_i^{(n_k)}(x)\mu_{n_k}(dx) \to \int_{I(j,N)} f_i(x)\mu(dx).$$

<u>Corollary</u> <u>43</u>: Let $1/D \le \xi < 1$, and let $Q_0,\dots,Q_{D-1} : \Omega_1 \to \mathcal{R}_0$ such that

(i) $\sum_{i=0}^{D-1} Q_i \equiv 1$

(ii) range $(Q_i) \subseteq [0,\xi]$ $\qquad (i \in S_1)$

and

(iii) $i^{(n)} \to i^{(0)}$ implies $Q_i(i^{(n)}) \to Q_i(i^{(0)})$ $\qquad (i \in S_1)$.

Then there exists a stationary chain of infinite order $\{Z_n\}_{n=-\infty}^{\infty}$ on a probability space $(\Omega, \mathcal{S}, \mathcal{P})$ with values in S_1 such that

$$\mathcal{P}(Z_n = i \mid Z_{n-1}, Z_{n-2}, \dots) = Q_i(Z_{n-1}, Z_{n-2}, \dots) \quad (i \in S_1)$$

almost surely on Ω.

Proof: Conditions (ii) and (iii) ensure that the vector $\vec{f} = (Q_0 \circ \nu, \ldots, Q_{D-1} \circ \nu)$ satisfies the conditions of Theorem 42. (With regard to (i), (ii), (iii) of Proposition 41, it is easiest to see that $\vec{f}$ satisfies (i).) Therefore, there exists a $P_{\vec{f}}$-invariant measure $\mu \in \mathcal{M}$. Clearly the f_i's are $\mathcal{B}$-measurable, so that the Q_i's are $\mathcal{S}_1$-measurable by Lemma 16. Hence the corollary follows from Theorem 19.

<u>Theorem</u> <u>44</u>; Let $\mathcal{I}$ be the collection of intervals of the form $I(j,N)$, where $N \in Z_0$ and $j \in S_N$. Let $\mathcal{I}'$ be a pairwise disjoint subcollection of $\mathcal{I}$ whose union is $[0,1]$. Let $1/D \leq \xi < 1$, and let $\vec{f} \in \mathcal{F}^{(D)}$ such that

$$\text{range }(f_i) \subseteq [0,\xi] \qquad (i \in S_1)$$

and such that $f_0, \ldots, f_{D-1}$ are uniformly continuous on each interval of $\mathcal{I}'$. Then there exists $\mu \in \mathcal{M}$ such that $(\mu, P_{\vec{f}}) \in \mathcal{A}$.

Proof: For $I \in \mathcal{I}'$ the uniform continuity of $f_0, \ldots, f_{D-1}$ on I implies the existence of a left-hand limit for $f_0, \ldots, f_{D-1}$ at the right endpoint of I; hence by Proposition 41 there exist vectors $\vec{f}^{(1)} \in \mathcal{K}_1^{(D)}, \vec{f}^{(2)} \in \mathcal{K}_2^{(D)}, \ldots$ such that $f_i^{(n)} \to f_i$ uniformly on $[0,1]$ for each $i \in S_1$. The $\vec{f}^{(n)}$'s may be assumed to satisfy

$$\text{range }(f_i^{(n)}) \subseteq (0,\xi] \qquad (i \in S_1;\ n \in Z_1).$$

As in the proof of Theorem 42, there is a sequence $\mu_1, \mu_2, \ldots \in \mathcal{M}$ such that $(\mu_n, P_{\vec{f}(n)}) \in A$ for each n; further, there is a subsequence $\mu_{n_1}, \mu_{n_2}, \ldots$ converging weakly to a measure $\mu \in \mathcal{M}$ with F_μ continuous. By assumption

$$\mu\Big(\bigcup_{I \in \mathcal{I}'} I\Big) = \mu([0,1]) = 1;$$

and since F_μ is continuous,

$$\mu(\partial I) = 0 \qquad\qquad (I \in \mathcal{I}').$$

Now $f_i^{(n)} \to f_i$ uniformly on each $I \in \mathcal{I}'$ because $f_i^{(n)} \to f_i$ uniformly on [0,1]. Also, if $I \in \mathcal{I}'$ has right endpoint b, then f_i on I can be extended to a continuous function $g_i^{(I)}$ on $\overline{I}$ by setting

$$g_i^{(I)}(x) = \begin{cases} f_i(x) & \text{if} \quad x \in I \\ \lim_{y \uparrow b} f_i(y) & \text{if} \quad x = b. \end{cases}$$

By Theorem 30 it follows that $(\mu, P_{\vec{f}}) \in A$.

<u>Corollary 45</u>: Let $\mathcal{C} = \{C_m\}_{m=1}^{\infty}$ be a disjoint collection of subsets of Ω_1 of the form

$$C_m = \{\omega_1 \in \Omega_1 : Z_1(\omega_1) = i_1^{(m)}, \ldots, Z_{k_m}(\omega_1) = i_{k_m}^{(m)}\}$$

such that

$$\bigcup_{m=1}^{\infty} C_m = \Omega_1.$$

Let $1/D \leq \xi < 1$, and let $Q_0, \ldots, Q_{D-1} : \Omega_1 \to R_0$ such that

$$\sum_{i=0}^{D-1} Q_i \equiv 1$$

and

$$\text{range } (Q_i) \subseteq [0, \xi] \qquad (i \in S_1).$$

Suppose further that

$$(*) \quad i^{(n)} \to i^{(0)} \text{ implies } Q_i(i^{(n)}) \to Q_i(i^{(0)})$$

whenever $i^{(0)}, i^{(1)}, i^{(2)}, \ldots \in C_m$ for some m. Then there exists a stationary chain of infinite order with transition probabilities $Q_0, \ldots, Q_{D-1}$.

Proof: Condition (*) implies that each function $f_i = Q_i \circ \nu$ is continuous on each interval $\nu^{-1}(C_m)$, with a left-hand limit at the right endpoint of $\nu^{-1}(C_m)$. Hence each f_i is uniformly continuous on each interval $\nu^{-1}(C_m)$. By the usual argument the corollary follows from Theorem 44.

Theorem 44 affirms the existence of a D-ary transition vector $\vec{f}$ with a $P_{\vec{f}}$-invariant measure such that

$$\varepsilon_m(f) \not\to 0.$$

In fact, if

$$1/2 \geq \eta_1 \geq \eta_2 \geq \ldots \geq 0$$

is any non-increasing sequence of non-negative terms, let D = 2 and define

$$f_0(x) = \begin{cases} 1/4 + \sum_{k=1}^{m} (-1)^{k+1} \varepsilon_k & \text{if } (2^m-1)/2^{m+1} \leq x < (2^{m+1}-1)/2^{m+2} \quad (m \in Z_0) \\ 1/2 & \text{if } 1/2 \leq x \leq 1 \end{cases}$$

and

$$f_1(x) = 1 - f_0(x) \qquad (0 \leq x \leq 1).$$

It is easy to check that

$$\varepsilon_m(\vec{f}) = \eta_m \qquad (m \in Z_1)$$

and that $\vec{f} = (f_0, f_1)$ satisfies the hypotheses of Theorem 44.

This example of a vector $\vec{f}$ whose components are step functions with countably many values suggests an interesting application of Corollary 45 to a class of chains of infinite order. The application concerns a generalization of the notion of an Nth order Markov chain. By definition the transition probabilities in an Nth order Markov chain depend upon only a fixed finite segment of the past. For a chain of infinite order with transition probabilities which are not Markovian of any order, it may nevertheless be the case that in some neighborhood of each sequence $(i_1, i_2, \ldots)$ the functions $Q_0, \ldots, Q_{D-1}$ depend upon only a finite number of their arguments. In this case we may say that the chain has "locally Markovian" transition probabilities.

<u>Definition</u> <u>46</u>: Let $Q_0, \ldots, Q_{D-1} : \Omega_1 \to R_0$ such that

$$\sum_{i=0}^{D-1} Q_i \equiv 1.$$

Then the functions $Q_0,\ldots,Q_{D-1}$ are said to be locally Markovian if every sequence

$$(i_1,i_2,\ldots) \in \Omega_1$$

belongs to a cylinder set of the form

$$\{\omega_1 \in \Omega_1 : Z_1(\omega_1) = i_1,\ldots,Z_k(\omega_1) = i_k\}$$

on which $Q_0,\ldots,Q_{D-1}$ are constant.

Corollary 47: Let $1/D \leq \xi < 1$, and let $Q_0,\ldots,Q_{D-1} : \Omega_1 \to R_0$ such that

$$\sum_{i=0}^{D-1} Q_i \equiv 1$$

and such that

$$\text{range } (Q_i) \subseteq [0,\xi] \qquad (i \in S_1).$$

If $Q_0,\ldots,Q_{D-1}$ are locally Markovian, there exists a stationary chain of infinite order with transition probabilities $Q_0,\ldots,Q_{D-1}$.

Proof: Since the Q_i's are locally Markovian, the cylinder sets on which they are constant form a (countable) cover of Ω_1. It is easy to select a subcover C of sets which are pairwise disjoint. On each set of C the continuity condition (*) of Corollary 45 is trivially satisfied; hence

Corollary 47 follows from Corollary 45.

REFERENCES

1. Doeblin, W., and R. Fortet, "Sur des chaînes à liaisons complètes," Bull. Soc. Math. France, 65 (1937), 132-148.

2. Dunford, Nelson, and Jacob T. Schwartz, Linear Operators, Part I: General Theory. Interscience, New York, 1958.

3. Feller, William, An Introduction to Probability Theory and Its Applications, Vol. II. John Wiley, New York, 1966.

4. Harris, T. E., "On chains of infinite order," Pac. J. Math., 5 (1955), 707-724.

5. Krickeberg, Klaus, Probability Theory. Addison-Wesley, Reading, Mass., 1965.

6. Rosenblatt, Murray, Random Processes. Oxford University Press, New York, 1962.