KŌDAI MATH. SEM. REP.
21 (1969), 426—447

MARKOV CHAINS WITH RANDOM TRANSITION MATRICES

BY YUKIO TAKAHASHI

Introduction.

Let P^t $(t=1, 2, 3, \cdots)$ be the transition matrix from epoch $t-1$ to t of a Markov chain with a finite state space S, and $\alpha^{(n)}$ $(n=0, 1, 2, \cdots)$ be the probability distribution at n. Then we have

$$\alpha^{(n)}=\alpha^{(0)}P^1\cdots P^n. \tag{*}$$

Now we assume that $\alpha^{(0)}$ and P^t are mutually independent random variables and that $\alpha^{(n)}$ is defined by (*). Then $\{\alpha^{(n)}\}$ is a Markov process on the space of probability distributions on S. ($\alpha^{(n)}$ represents the probability distribution at n, starting with the initial distribution $\alpha^{(0)}$ and following to the *random* transition matrices P^t.) Such a process will be called a "*Markov chain with random transition matrices*" (*M.C. with R.T.M.*).

The author intended to generalize ordinary Markov chains as briefly mentioned above, by the following reason.

Markov chains have been applied in many fields, and one of their applications is in the analysis or prediction of market shares. Many authors have worked with so-called Markov brand-switching models, in which $\alpha^{(n)}$ represents the market shares at epoch n and P^t represents the transition matrix from epoch $t-1$ to t. In many cases they assume that these Markov chains (they consider that $\alpha^{(n)}$ is the distribution of a Markov chain at step n) are stationary. However some other authors have given warnings of the failure of stationarity and of other defects of these models (see e.g. A. S. C. Ehrenberg [1]). The author thinks that one of the causes of the warnings is in the assumption that $\alpha^{(0)}$ and P^t are *a priori* given (known) and hence have no stochastic fluctuation. The transition matrix P^t reflects the choices of purchasers, and so it is essentially stochastic. Hence it seems to be natural to consider that P^t and $\alpha^{(n)}$ are random variables. The most simple stochastic model for market shares is the model using our M.C. with R.T.M.

In this paper, properties of M.C. with R.T.M., in particular moments of $\alpha^{(n)}$ and conditions for the convergence (in law) of $\alpha^{(n)}$, are given. And we classify stationary and irreducible M.C. with R.T.M. into three groups; ergodic chains, aperiodic and non-ergodic chains, and periodic chains. Finally we prove some ergodic theorems.

Received June 17, 1969.

1. Markov chains with random transition matrices.

Let $S=\{1, 2, \cdots, s\}$ be a finite state space, $\mathcal{A}$ be the set of all probability measures on S, and $\mathcal{P}$ be the set of all stochastic matrices with index sets S and S. We may be consider $\mathcal{A}$ and $\mathcal{P}$ as subsets of s- and s^2-dimensional Euclidean spaces respectively;

$$(1.1)\qquad \mathcal{A}=\left\{\alpha=(\alpha_1, \cdots, \alpha_s)\,\middle|\, \alpha_i\geqq 0 \quad i=1, \cdots, s, \quad \sum_{i=1}^{s}\alpha_i=1\right\},$$

$$(1.2)\qquad \mathcal{P}=\left\{P=\begin{pmatrix} p_{11}\cdots p_{1s} \\ \vdots \quad \vdots \\ p_{s1}\cdots p_{ss}\end{pmatrix}\,\middle|\, p_{ij}\geqq 0 \quad i, j=1, \cdots, s, \quad \sum_{j=1}^{s}p_{ij}=1\right\}.$$

Therefore we can define random variables which take values on $\mathcal{A}$ or $\mathcal{P}$.

Given a sequence of probability spaces $(\Omega^t, \mathcal{F}^t, \Pr^t)$ $(t=0, 1, 2, \cdots)$, a vector valued random variable

$$(1.3)\qquad \boldsymbol{\alpha}^{(0)}(\omega^0)=(\boldsymbol{\alpha}_i^{(0)}(\omega^0) \quad i=1, 2, \cdots, s)$$

on Ω^0 with values in $\mathcal{A}$, and matrix valued random variables

$$(1.4)\qquad \boldsymbol{P}^t(\omega^t)=(\boldsymbol{p}_{ij}^t(\omega^t) \quad i, j=1, 2, \cdots, s)$$

on Ω^t $(t=1, 2, 3, \cdots)$ with values in $\mathcal{P}$, then we define the product probability space $(\Omega, \mathcal{F}, \Pr)$ by

$$(1.5)\qquad (\Omega, \mathcal{F}, \Pr)=\prod_{t=0}^{\infty}(\Omega^t, \mathcal{F}^t, \Pr^t),$$

and random variables $\boldsymbol{\alpha}^{(0)}(\omega)$ and $\boldsymbol{P}^t(\omega)$ on Ω by

$$(1.6)\qquad \boldsymbol{\alpha}^{(0)}(\omega)=\boldsymbol{\alpha}^{(0)}(\omega^0) \quad \text{and} \quad \boldsymbol{P}^t(\omega)=\boldsymbol{P}^t(\omega^t) \qquad t=1, 2, 3, \cdots$$

respectively, where

$$(1.7)\qquad \Omega\ni\omega=(\omega^0, \omega^1, \omega^2, \cdots).$$

DEFINITION. A *Markov chain with random transition matrices* (*M.C. with R.T.M.*) $\{\boldsymbol{\alpha}^{(n)}(\omega)\}$ is a Markov process on Ω with values in $\mathcal{A}$ of the form

$$(1.8)\qquad \boldsymbol{\alpha}^{(n)}(\omega)=\boldsymbol{\alpha}^{(0)}(\omega)\boldsymbol{P}^1(\omega)\cdots\boldsymbol{P}^n(\omega).$$

In the following sections we mainly study the stationary case where all probability spaces $(\Omega^t, \mathcal{F}^t, \Pr^t)$ $(t=1, 2, 3, \cdots)$ but $(\Omega^0, \mathcal{F}^0, \Pr^0)$ are identical and random variables $\boldsymbol{P}^t(\omega^t)$ have a common distribution. We will refer to such a chain as a *stationary M.C. with R.T.M.*

We denote the n-step transition probability matrix by

$$(1.9)\qquad \boldsymbol{P}^{(n)}(\omega)=(\boldsymbol{p}_{ij}^{(n)}(\omega))=\boldsymbol{P}^1(\omega)\cdots\boldsymbol{P}^n(\omega).$$

Clearly $\{\boldsymbol{P}^{(n)}(\omega)\}$ is a Markov process and so we sometimes call it also a M.C. with R.T.M.

In the following sections, ω will be omitted when no confusion arises.

2. Moments.

We first calculate the moments of $\boldsymbol{\alpha}^{(n)}(\omega)$ and of $\boldsymbol{P}^{(n)}(\omega)$. We prepare some notations.

Let S_k $(k=1, 2, 3, \cdots)$ be the set of all ordered k-tuples $(i_1, \cdots, i_k)$ of states in S (e.g. when $S=\{1, 2\}$, $S_1=\{(1), (2)\}$, $S_2=\{(1, 1), (1, 2), (2, 1), (2, 2)\}, \cdots)$. The k-th moment of $\boldsymbol{\alpha}^{(n)}(\omega)$ is denoted by the row vector

$$\xi_k^{(n)}=(\xi_k^{(n)}(i_1, \cdots, i_k),\ (i_1, \cdots, i_k)\in S_k) \tag{2. 1}$$

where

$$\xi_k^{(n)}(i_1, \cdots, i_k)=E\{\boldsymbol{\alpha}_{i_1}^{(n)}(\omega)\cdots\boldsymbol{\alpha}_{i_k}^{(n)}(\omega)\}, \tag{2. 2}$$

and the k-th moment matrix of $\boldsymbol{P}^t(\omega)$ is denoted by

$$\Sigma_k^t=(\sigma_k^t(i_1, \cdots, i_k; j_1, \cdots, j_k),\ (i_1, \cdots, i_k; j_1, \cdots, j_k)\in S_k\times S_k) \tag{2. 3}$$

where

$$\sigma_k^t(i_1, \cdots, i_k; j_1, \cdots, j_k)=E\{\boldsymbol{p}_{i_1 j_1}^t(\omega)\cdots\boldsymbol{p}_{i_k j_k}^t(\omega)\}. \tag{2. 4}$$

In the stationary case Σ_k^t will be abbreviated to Σ_k. The same notations as the moments of $\boldsymbol{P}^t(\omega)$ will be used for the moments of $\boldsymbol{P}^{(n)}(\omega)$ with brackets on their shoulders.

We note that $\xi_k^{(n)}$ may be considered as a probability measure on S_k and that Σ_k^t and $\Sigma_k^{(n)}$ are stochastic matrices. In fact

$$\begin{aligned}\sum_{(i_1,\cdots,i_k)\in S_k}\xi_k^{(n)}(i_1, \cdots, i_k)&=\sum_{(i_1,\cdots,i_k)\in S_k}E\{\boldsymbol{\alpha}_{i_1}^{(n)}\cdots\boldsymbol{\alpha}_{i_k}^{(n)}\}\\&=E\left\{\left(\sum_{i_1=1}^{s}\boldsymbol{\alpha}_{i_1}^{(n)}\right)\cdots\left(\sum_{i_k=1}^{s}\boldsymbol{\alpha}_{i_k}^{(n)}\right)\right\}=1\end{aligned} \tag{2. 5}$$

which proves the first statement, and a similar calculation leads to the second. Furthermore, we may prove the following

THEOREM 1. *For a M.C. with R.T.M. we have*

$$\Sigma_k^{(n)}=\Sigma_k^1\cdots\Sigma_k^n \quad \textit{and} \quad \xi_k^{(n)}=\xi_k^{(0)}\Sigma_k^{(n)}=\xi_k^{(0)}\Sigma_k^1\cdots\Sigma_k^n. \tag{2. 6}$$

In particular, for a stationary chain we have

$$\Sigma_k^{(n)}=(\Sigma_k)^n \quad \text{and} \quad \xi_k^{(n)}=\xi_k^{(0)}(\Sigma_k)^n. \tag{2. 7}$$

Proof. We shall show that $\xi_k^{(n)}=\xi_k^{(0)}\Sigma_k^{(n)}$. By the independence of variables we have

$$
\begin{aligned}
\xi_k^{(n)}(i_1, \cdots, i_k) &= E\left\{\left(\sum_{l_1=1}^{s} \boldsymbol{a}_{l_1}^{(0)} \boldsymbol{p}_{l_1 i_1}^{(n)}\right)\cdots\left(\sum_{l_k=1}^{s} \boldsymbol{a}_{l_k}^{(0)} \boldsymbol{p}_{l_k i_k}^{(n)}\right)\right\} \\
&= \sum_{(l_1,\cdots,l_k)\in S_k} E\{\boldsymbol{a}_{l_1}^{(0)}\cdots\boldsymbol{a}_{l_k}^{(0)}\cdot\boldsymbol{p}_{l_1 i_1}^{(n)}\cdots\boldsymbol{p}_{l_k i_k}^{(n)}\} \\
&= \sum_{(l_1,\cdots,l_k)\in S_k} E\{\boldsymbol{a}_{l_1}^{(0)}\cdots\boldsymbol{a}_{l_k}^{(0)}\}E\{\boldsymbol{p}_{l_1 i_1}^{(n)}\cdots\boldsymbol{p}_{l_k i_k}^{(n)}\} \\
&= \sum_{(l_1,\cdots,l_k)\in S_k} \xi_k^{(0)}(l_1, \cdots, l_k)\cdot\sigma_k^{(n)}(l_1, \cdots, l_k; i_1, \cdots, i_k).
\end{aligned}
\tag{2. 8}
$$

The first relation in (2. 6) can be proved by a similar calculation. Q.E.D.

Now we consider the convergence problem of $\boldsymbol{a}^{(n)}(\omega)$. We need a lemma for the convergence of a sequence of random variables (e.g. see Feller [3] p. 244).

LEMMA 2. *Let $\{\boldsymbol{X}^n\}$ be a uniformly bounded sequence of random variables in r-dimensional Euclidean space and μ_k^n be the k-th moment vector of $\boldsymbol{X}^n$. $\boldsymbol{X}^n$ converges in law to a limit $\boldsymbol{X}$ if, and only if, for each k, μ_k^n converges to a limit vector μ_k. In this case μ_k is the k-th moment vector of $\boldsymbol{X}$.*

Applying this lemma to our chains, we obtain the following theorem.

THEOREM 3. *A M.C. with R.T.M. converges in law if, and only if, $\xi_k^{(n)}=\xi_k^{(0)}\Sigma_k^1\cdots\Sigma_k^n$ converges for each k.*

3. Classification of stationary M.C. with R.T.M. (I)—Periodicity.

The theory of ordinary Markov chains suggests us that stationary M.C. with R.T.M. could be classified by similar ideas. For convenience we consider $\boldsymbol{P}^{(n)}$-process instead of $\boldsymbol{a}^{(n)}$-process. Theorem 1 shows that for the expectation of $\boldsymbol{P}^{(n)}$ we can consider the ordinary Markov chain with transition matrix Σ_1. It seems to be natural to classify stationary M.C. with R.T.M. by the properties of M.C. Σ_1 (we refer to an ordinary stationary Markov chain by its transition matrix and write as M.C. Σ_1). We might define that a stationary M.C. with R.T.M. is *irreducible* (*reducible*) if M.C. Σ_1 is irreducible (reducible), and that a stationary and irreducible M.C. with R.T.M. is aperiodic (periodic) if M.C. Σ_1 is aperiodic (periodic). This definition of an irreducible chain is adequate in the sense that for each pair i, j ($\in S$) there is an n such that

$$
\Pr\{\boldsymbol{p}_{ij}^{(n)}(\omega)>0\}>0. \tag{3. 1}
$$

However Example 1 below shows that the above definition of an aperiodic chain does not seem to be adequate.

In the following sections (except in Theorem 8) we will consider stationary and irreducible M.C. with R.T.M. only, and sometimes the words "irreducible" and "stationary" will be omitted.

EXAMPLE 1. Let $s=3$ and the distribution of $\boldsymbol{P}^t$ be

$$\Pr\{\boldsymbol{P}^t=P_1\}=\Pr\{\boldsymbol{P}^t=P_2\}=\Pr\{\boldsymbol{P}^t=P_3\}=\frac{1}{3} \tag{3.2}$$

where

$$P_1=\begin{pmatrix}1&0&0\\0&0&1\\0&1&0\end{pmatrix},\quad P_2=\begin{pmatrix}0&0&1\\0&1&0\\1&0&0\end{pmatrix}\quad\text{and}\quad P_3=\begin{pmatrix}0&1&0\\1&0&0\\0&0&1\end{pmatrix}. \tag{3.3}$$

Then we have

$$\Sigma_1=\begin{pmatrix}1/3&1/3&1/3\\1/3&1/3&1/3\\1/3&1/3&1/3\end{pmatrix}. \tag{3.4}$$

Hence M.C. Σ_1 is aperiodic.

Now let

$$P_4=\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix},\quad P_5=\begin{pmatrix}0&0&1\\1&0&0\\0&1&0\end{pmatrix}\quad\text{and}\quad P_6=\begin{pmatrix}0&1&0\\0&0&1\\1&0&0\end{pmatrix}. \tag{3.5}$$

Since

$$\begin{aligned}&P_1P_1=P_2P_2=P_3P_3=P_4, &&P_4P_1=P_5P_2=P_6P_3=P_1,\\ &P_1P_2=P_2P_3=P_3P_1=P_5, &&P_4P_2=P_5P_3=P_6P_1=P_2,\\ &P_1P_3=P_2P_1=P_3P_2=P_6, &&P_4P_3=P_5P_1=P_6P_2=P_3,\end{aligned} \tag{3.6}$$

we have

$$\Pr\{\boldsymbol{P}^{(n)}=P_1\}=\Pr\{\boldsymbol{P}^{(n)}=P_2\}=\Pr\{\boldsymbol{P}^{(n)}=P_3\}=\frac{1}{3} \tag{3.7}$$

if n is odd, and

$$\Pr\{\boldsymbol{P}^{(n)}=P_4\}=\Pr\{\boldsymbol{P}^{(n)}=P_5\}=\Pr\{\boldsymbol{P}^{(n)}=P_6\}=\frac{1}{3} \tag{3.8}$$

if n is even. Hence we would rather say that this M.C. with R.T.M. has "period two".

Thus we must make a new definition of the period of a stationary M.C. with R.T.M. By Lemma 2, for the convergence of $\boldsymbol{P}^{(n)}$ we are enough to examine the convergence of its moments only. We shall show that there is an integer $r\geqq 1$ such that $\Sigma_k^{(nr+m)}=(\Sigma_k)^{nr+m}$ converges as $n\to\infty$ for each k and m $(m=0,1,2,\cdots,r-1)$. (Convergence of a sequence of matrices means element-wise convergence.)

By the theory of ordinary Markov chains, each state $(i_1, \cdots, i_k) \in S_k$ has its own period with respect to M.C. Σ_k. We define that $(j_1, \cdots, j_h) \in S_h$ and $(i_1, \cdots, i_k) \in S_k$ be *equivalent* if both consist of the same states in S. (For example, $(1, 1, 2)$ is equivalent to $(2, 1, 2, 2)$.) Let $(j_1, \cdots, j_h)$ be equivalent to $(i_1, \cdots, i_k)$. By the very definition we have

$$\sigma_k^{(n)}(i_1, \cdots, i_k; i_1, \cdots, i_k) = E\{\boldsymbol{p}_{i_1 i_1}^{(n)} \cdots \boldsymbol{p}_{i_k i_k}^{(n)}\} \tag{3.9}$$

and

$$\sigma_h^{(n)}(j_1, \cdots, j_h; j_1 \cdots, j_h) = E\{\boldsymbol{p}_{j_1 j_1}^{(n)} \cdots \boldsymbol{p}_{j_h j_h}^{(n)}\}. \tag{3.10}$$

So the equivalence of $(i_1, \cdots, i_k)$ and $(j_1, \cdots, j_k)$ implies that the values in both braces in the right sides of (3.9) and (3.10) vanish simultaneously. Therefore $\sigma_k^{(n)}(i_1, \cdots, i_k; i_1, \cdots, i_k) = 0$ if, and only if, $\sigma_h^{(n)}(j_1, \cdots, j_h; j_1, \cdots, j_h) = 0$. Thus equivalent states have the same period if they are periodic. And it is easily shown that if a state in S_k is transient (with respect to M.C. Σ_k), then each state, which is equivalent to it, is also transient. Therefore equivalent states have a common period.

Let r be the least common multiple of the periods of states in S_s (s is the number of states in S). Then $\Sigma_s^{(nr+m)} = (\Sigma_s)^{nr+m}$ converges as $n \to \infty$ for each m ($m = 0, 1, 2, \cdots, r-1$) and the limit matrices are different for different m's. We note that each state in S_k has an equivalent state in S_s. Hence r is also a common multiple of the periods of states in S_k and $\Sigma_k^{(nr+m)} = (\Sigma_k)^{nr+m}$ converges as $n \to \infty$ for each m ($m = 0, 1, 2, \cdots, r-1$). Therefore applying Lemma 2 to r requences $\{\boldsymbol{P}^{(nr+m)}\}$ ($m = 0, 1, 2, \cdots, r-1$) we obtain the following

THEOREM 4. *For a stationary and irreducible M.C. with R.T.M. there exists unique integer $r \geqq 1$ such that r sequences $\{\boldsymbol{P}^{(nr+m)}\}$ ($m = 0, 1, 2, \cdots, r-1$) converge in law as $n \to \infty$ and that their limit distributions are different from each other.*

DEFINITION. The *period* of a stationary and irreducible M.C. with R.T.M. is the number whose existence is assured in Theorem 4.

The discussion preceding Theorem 4 shows that the period of a M.C. with R.T.M. is the least common multiple of the periods of states in S_s. Turning to Example 1, the states $(1, 1, 1)$, $(2, 2, 2)$ and $(3, 3, 3)$ have period one with respect to M.C. Σ_3 and other states in S_3 have period two. Hence the chain has period two. Thus the new definition of the period seems to be adequate.

4. Classification of stationary M.C. with R.T.M. (II)—Ergodicity.

In the last section we classified stationary and irreducible M.C. with R.T.M. by their periods. However, there is another and more essential classification; ergodic chains or non-ergodic chains. We shall start with two examples.

EXAMPLE 2. Let $s = 2$ and the distribution of $\boldsymbol{P}^t$ be

$$\Pr\{\boldsymbol{P}^t=P_1\}=\Pr\{\boldsymbol{P}^t=P_2\}=\frac{1}{2} \tag{4. 1}$$

where

$$P_1=\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} \quad \text{and} \quad P_2=\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}. \tag{4. 2}$$

Then

$$\Sigma_1=\begin{pmatrix}1/2 & 1/2\\ 1/2 & 1/2\end{pmatrix}, \tag{4. 3}$$

and it is easily shown that this chain is aperiodic. Since

$$P_1P_1=P_2P_2=P_1 \quad \text{and} \quad P_1P_2=P_2P_1=P_2, \tag{4. 4}$$

we have

$$\Pr\{\boldsymbol{P}^{(n)}=P_1\}=\Pr\{\boldsymbol{P}^{(n)}=P_2\}=\frac{1}{2} \tag{4. 5}$$

for each n. Hence if $\boldsymbol{\alpha}^{(0)}=(p, 1-p)$ with probability one,

$$\Pr\{\boldsymbol{\alpha}^{(n)}=(p, 1-p)\}=\Pr\{\boldsymbol{\alpha}^{(n)}=(1-p, p)\}=\frac{1}{2}. \tag{4. 6}$$

Thus the distribution of $\boldsymbol{\alpha}^{(n)}$ is independent of n, and the limit distribution is given also by (4. 6). We note that the limit distribution depends on $\boldsymbol{\alpha}^{(0)}$.

EXAMPLE 3. Let S and the distribution of $\boldsymbol{P}^t$ be as in Example 2, but this time we put

$$P_1=\begin{pmatrix}1/2 & 1/2\\ 1 & 0\end{pmatrix} \quad \text{and} \quad P_2=\begin{pmatrix}1/2 & 1/2\\ 0 & 1\end{pmatrix}. \tag{4. 7}$$

Then Σ_1 is given by (4. 3) and the chain is aperiodic too. But the distribution of $\boldsymbol{\alpha}^{(n)}$ is not so simple as the preceding example. Direct calculation shows that

$$\Pr\left\{\boldsymbol{P}^{(n)}=\begin{pmatrix}p_{n,k} & 1-p_{n,k}\\ q_{n,k} & 1-q_{n,k}\end{pmatrix}\right\}=\Pr\left\{\boldsymbol{P}^{(n)}=\begin{pmatrix}1-p_{n,k} & p_{n,k}\\ 1-q_{n,k} & q_{n,k}\end{pmatrix}\right\}=\frac{1}{2^n} \tag{4. 8}$$

where

$$p_{n,k}=\frac{2k-1}{2^n} \quad \text{and} \quad q_{n,k}=\frac{k}{2^{n-1}} \qquad (k=1, 2, \cdots, 2^{n-1}). \tag{4. 9}$$

Therefore $\boldsymbol{P}^{(n)}$ converges in law to a random matrix

(4. 10)
$$\begin{pmatrix} \boldsymbol{q} & 1-\boldsymbol{q} \\ \boldsymbol{q} & 1-\boldsymbol{q} \end{pmatrix}$$

where $\boldsymbol{q}$ is a random variable following to the uniform distribution on the unit interval [0, 1]. Hence even if $\boldsymbol{\alpha}^{(0)}=(p, 1-p)$ with probability one, $\boldsymbol{\alpha}^{(n)}$ converges in law to

(4. 11)
$$(\boldsymbol{q},\ 1-\boldsymbol{q})$$

which does not depend on $\boldsymbol{\alpha}^{(0)}$.

Two chains in above have the same Σ_1, but their behaviors are quite different. Therefore we have to distinguish aperiodic chains into two types.

DEFINITION. A stationary and irreducible M.C. with R.T.M. is *ergodic* if it is aperiodic and its limit distribution does not depend on the initial variable $\boldsymbol{\alpha}^{(0)}$.

It is easily shown that a chain is ergodic if, and only if, $\boldsymbol{P}^{(n)}$ converges in law to a random variable

(4. 12)
$$\boldsymbol{Q}=\begin{pmatrix} \boldsymbol{q}_1 \cdots \boldsymbol{q}_s \\ \vdots \quad\ \vdots \\ \boldsymbol{q}_1 \cdots \boldsymbol{q}_s \end{pmatrix}$$

which has the same row vectors. In section 6 we shall obtain some necessary and sufficient conditions for ergodicity.

5. Dual processes.

We shall define the "dual process" which plays an important role in ergodic theorems, and introduce some notations.

We have denoted the n-step transition probability matrix by

(5. 1)
$$\boldsymbol{P}^{(n)}(\omega)=(\boldsymbol{p}_{ij}^{(n)}(\omega))=\boldsymbol{P}^1(\omega)\cdots\boldsymbol{P}^n(\omega), \qquad \omega\in\Omega.$$

If the order of multiplications in (5. 1) is reversed, the value of the matrix differs from (5. 1), so we denote it by

(5. 2)
$$\hat{\boldsymbol{P}}^{(n)}(\omega)=(\hat{\boldsymbol{p}}_{ij}^{(n)}(\omega))=\boldsymbol{P}^n(\omega)\cdots\boldsymbol{P}^1(\omega), \qquad \omega\in\Omega.$$

Clearly $\hat{\boldsymbol{P}}^{(n)}$ is a Markov process, and we will call it the *dual process* (of a chain $\boldsymbol{\alpha}^{(n)}$ or $\boldsymbol{P}^{(n)}$). Similarly, we denote the n-step transition probability matrix from epoch m $(m=1, 2, \cdots)$ by

(5. 3)
$${}^m\boldsymbol{P}^{(n)}(\omega)=({}^m\boldsymbol{p}_{ij}^{(n)}(\omega))=\boldsymbol{P}^{m+1}(\omega)\cdots\boldsymbol{P}^{m+n}(\omega), \qquad \omega\in\Omega,$$

and its dual by

(5. 4)
$${}^m\hat{\boldsymbol{P}}^{(n)}(\omega)=({}^m\hat{\boldsymbol{p}}_{ij}^{(n)}(\omega))=\boldsymbol{P}^{m+n}(\omega)\cdots\boldsymbol{P}^{m+1}(\omega), \qquad \omega\in\Omega.$$

For a stationary M.C. with R.T.M., random variables $\boldsymbol{P}^t(\omega)$ are mutually independent and have a common distribution. Hence the distributions of $\boldsymbol{P}^{(n)}$, $\hat{\boldsymbol{P}}^{(n)}$, ${}^m\boldsymbol{P}^{(n)}$ and ${}^m\hat{\boldsymbol{P}}^{(n)}$ coincide with each other. So, for any s^2-dimensional Borel set B

(5.5) $$\Pr\{\boldsymbol{P}^{(n)}\in B\}=\Pr\{\hat{\boldsymbol{P}}^{(n)}\in B\}=\Pr\{{}^m\boldsymbol{P}^{(n)}\in B\}=\Pr\{{}^m\hat{\boldsymbol{P}}^{(n)}\in B\}.$$

We denote the maximum and the minimum of the j-th column of $\boldsymbol{P}^{(n)}[\hat{\boldsymbol{P}}^{(n)}]$ by

(5.6) $$\boldsymbol{M}_j^{(n)}(\omega)=\operatorname*{Max}_{1\leqq i\leqq s}\boldsymbol{p}_{ij}^{(n)}(\omega)\qquad\left[\hat{\boldsymbol{M}}_j^{(n)}(\omega)=\operatorname*{Max}_{1\leqq i\leqq s}\hat{\boldsymbol{p}}_{ij}^{(n)}(\omega)\right]$$

and

(5.7) $$\boldsymbol{m}_j^{(n)}(\omega)=\operatorname*{Min}_{1\leqq i\leqq s}\boldsymbol{p}_{ij}^{(n)}(\omega)\qquad\left[\hat{\boldsymbol{m}}_j^{(n)}(\omega)=\operatorname*{Min}_{1\leqq i\leqq s}\hat{\boldsymbol{p}}^{(n)}(\omega)\right]$$

respectively. Since

(5.8) $$\hat{\boldsymbol{p}}_{ij}^{(n+1)}=\sum_{k=1}^{s}\boldsymbol{p}_{ik}^{n+1}\hat{\boldsymbol{p}}_{kj}^{(n)},$$

we have

(5.9) $$\hat{\boldsymbol{p}}_{ij}^{(n+1)}\leqq\hat{\boldsymbol{M}}_j^{(n)}\sum_{k=1}^{s}\boldsymbol{p}_{ik}^{n+1}=\hat{\boldsymbol{M}}_j^{(n)}$$

and similarly

(5.10) $$\hat{\boldsymbol{p}}_{ij}^{(n+1)}\geqq\hat{\boldsymbol{m}}_j^{(n)}.$$

Hence we obtain the following relation:

(5.11) $$0\leqq\hat{\boldsymbol{m}}_j^{(1)}\leqq\hat{\boldsymbol{m}}_j^{(2)}\leqq\cdots\leqq\boldsymbol{m}_j^{(n)}\leqq\cdots\leqq\hat{\boldsymbol{M}}_j^{(n)}\leqq\cdots\leqq\hat{\boldsymbol{M}}_j^{(2)}\leqq\hat{\boldsymbol{M}}_j^{(1)}\leqq 1.$$

Since $\{\hat{\boldsymbol{M}}_j^{(n)}(\omega)\}$ and $\{\hat{\boldsymbol{m}}_j^{(n)}(\omega)\}$ are bounded monotone sequences, there exist their limit variables $\hat{\boldsymbol{M}}_j(\omega)$ and $\hat{\boldsymbol{m}}_j(\omega)$ for each j:

(5.12) $$\lim_{n\to\infty}\hat{\boldsymbol{m}}_j^{(n)}(\omega)=\hat{\boldsymbol{m}}_j(\omega)\leqq\hat{\boldsymbol{M}}_j(\omega)=\lim_{n\to\infty}\hat{\boldsymbol{M}}_j^{(n)}(\omega).$$

Using these $\hat{\boldsymbol{M}}_j$ we define the matrix valued random variable $\hat{\boldsymbol{M}}$ with the same row vectors by

(5.13) $$\hat{\boldsymbol{M}}(\omega)=\begin{pmatrix}\hat{\boldsymbol{M}}_1(\omega)\cdots\hat{\boldsymbol{M}}_s(\omega)\\ \vdots\\ \hat{\boldsymbol{M}}_1(\omega)\cdots\hat{\boldsymbol{M}}_s(\omega)\end{pmatrix}.$$

6. Ergodic theorems.

In this section we shall obtain some ergodic theorems. Theorem 5 shows the fundamental relation between the ergodicity and the convergence of the dual process, and Theorem 6 gives us a good criterion for the ergodicity. Theorem 9 also gives a necessary and sufficient condition for the ergodicity, but it might be useful to study non-ergodic chains. Theorem 8 treats the non-stationary case and states that under some mild assumptions the effect of the initial variable vanishes in the long run.

THEOREM 5. *For a stationary and irreducible M.C. with R.T.M., the following four statements are equivalent*:

(a) *The chain is ergodic.*

(b) $\boldsymbol{P}^{(n)}(\omega)$ $[\hat{\boldsymbol{P}}^{(n)}(\omega)]$ *converges in low to* $\hat{\boldsymbol{M}}(\omega)$.

(c) $\hat{\boldsymbol{P}}^{(n)}(\omega)$ *converges in probability to* $\hat{\boldsymbol{M}}(\omega)$.

(d) $\hat{\boldsymbol{P}}^{(n)}(\omega)$ *converges with probability one to* $\hat{\boldsymbol{M}}(\omega)$.

(*We note that the expression in* (b) *is justified by the relation* (5. 5).)

Proof. As stated in section 5, (a) is equivalent to

(a′) $\boldsymbol{P}^{(n)}(\omega)$ $[\hat{\boldsymbol{P}}^{(n)}(\omega)]$ converges in law to a random variable

$$\boldsymbol{Q}(\omega')=\begin{pmatrix}\boldsymbol{q}_1(\omega')\cdots\boldsymbol{q}_s(\omega')\\ \vdots \qquad \vdots\\ \boldsymbol{q}_1(\omega')\cdots\boldsymbol{q}_s(\omega')\end{pmatrix}\qquad(\omega'\in\Omega')\tag{6. 1}$$

which has the same row vectors, where $(\Omega', \mathcal{F}', \Pr')$ is a certain probability space.

By the well known theorems for convergences of random variables, it is clear that (d) implies (c) and that (c) implies (b). Also it is obvious that (b) implies (a′). So we need only to show that (c) implies (d) and that (a′) implies (c).

Suppose that (c) is satisfied. Then there is a subsequence $\{\hat{\boldsymbol{P}}^{(n_k)}\}$ which converges with probability one to $\hat{\boldsymbol{M}}$, i.e., for every i, j

$$\hat{\boldsymbol{p}}_{ij}^{(n_k)}\to\hat{\boldsymbol{M}}_j\qquad(k\to\infty)\quad\text{w.p. 1.}\tag{6. 2}$$

Hence

$$\hat{\boldsymbol{m}}_j^{(n_k)}\to\hat{\boldsymbol{M}}_j\qquad(k\to\infty)\quad\text{w.p. 1.}\tag{6. 3}$$

By the monotonicity of $\{\hat{\boldsymbol{m}}_j^{(n)}\}$, (6. 3) implies that

$$\hat{\boldsymbol{m}}_j^{(n)}\to\hat{\boldsymbol{M}}_j\qquad(n\to\infty)\quad\text{w.p. 1.}\tag{6. 4}$$

Since for every i, j

$$\hat{\boldsymbol{m}}_j^{(n)}\leqq\hat{\boldsymbol{p}}_{ij}^{(n)}\leqq\hat{\boldsymbol{M}}_j^{(n)},\tag{6. 5}$$

we have

$$\hat{\boldsymbol{p}}_{ij}^{(n)}\to\hat{\boldsymbol{M}}_j\qquad(n\to\infty)\quad\text{w.p. 1.}\tag{6. 6}$$

which is the same as (d).

Now we show that (a′) implies (c). Let I be an interval of continuity for the distribution of $\boldsymbol{Q}$ (i.e., I is open and its boundary has probability zero. See Feller [3] p. 242.), then (a′) implies that

(6. 7) $$\Pr\{\hat{\boldsymbol{P}}^{(n)}\in I\}\to\Pr'\{\boldsymbol{Q}\in I\}\qquad (n\to\infty).$$

We can choose a finite set of points $\{a_\nu\}$ $(\nu=0, 1, \cdots, u)$ such that each a_ν is a point of continuity for each marginal distribution of $\boldsymbol{q}_j$ and that

(6. 8) $$a_0<0,\quad a_u>1\quad\text{and}\quad 0<a_\nu-a_{\nu-1}<\delta\qquad (\nu=1, \cdots, u)$$

for arbitrary given positive number δ. Let

(6. 9) $$\left.\begin{array}{l} I(\nu_1, \cdots, \nu_s)=(a_{\nu_1-1}, a_{\nu_1})\times\cdots\times(a_{\nu_s-1}, a_{\nu_s}) \\ \qquad\cdots\cdots\cdots \\ \qquad\times(a_{\nu_1-1}, a_{\nu_1})\times\cdots\times(a_{\nu_s-1}, a_{\nu_s}) \end{array}\right\} s \text{ times.}$$

Then $I(\nu_1, \cdots, \nu_s)$ $(\nu_j=1, \cdots, u)$ are intervals of continuity of the distribution of $\boldsymbol{Q}$ and they are mutually exclusive. From (6. 7) we have

(6. 10) $$\Pr\{\hat{\boldsymbol{P}}^{(n)}\in I(\nu_1, \cdots, \nu_s)\}\to\Pr'\{\boldsymbol{Q}\in I(\nu_1, \cdots, \nu_s)\}.$$

Summing up them with respect to $(\nu_1, \cdots, \nu_s)$, we obtain that

(6. 11) $$\begin{aligned}\Pr\left\{\hat{\boldsymbol{P}}^{(n)}\in\bigcup_{(\nu_1,\cdots,\nu_s)} I(\nu_1, \cdots, \nu_s)\right\}&=\sum_{(\nu_1,\cdots,\nu_s)}\Pr\{\hat{\boldsymbol{P}}^{(n)}\in I(\nu_1, \cdots, \nu_s)\}\\ &\to\sum_{(\nu_1,\cdots,\nu_s)}\Pr'\{\boldsymbol{Q}\in I(\nu_1, \cdots, \nu_s)\}.\end{aligned}$$

We first show that the right side of (6. 11) is equal to one. Let

(6. 12) $$\begin{array}{l} I^*(\nu_1, \cdots, \nu_s)=(a_{\nu_1-1}, a_{\nu_1})\times\cdots\times(a_{\nu_s-1}, a_{\nu_s}) \\ \qquad\left.\begin{array}{l}\times(a_0, a_u)\times\cdots\times(a_0, a_u) \\ \cdots\cdots\cdots \\ \times(a_0, a_u)\times\cdots\times(a_0, a_u)\end{array}\right\} s-1 \text{ times}\end{array}$$

and

(6. 13) $$\left.\begin{array}{l} I^*=(a_0, a_u)\times\cdots\times(a_0, a_u) \\ \qquad\cdots\cdots\cdots \\ \quad\times(a_0, a_u)\times\cdots\times(a_0, a_u) \end{array}\right\} s \text{ times.}$$

Since $\boldsymbol{Q}$ has the same row vectors, we have

(6. 14) $$\begin{aligned}&\Pr'\{\boldsymbol{Q}\in I(\nu_1, \cdots, \nu_s)\}\\ &=\Pr'\{a_{\nu_1-1}<\boldsymbol{q}_1<a_\nu, \cdots, a_{\nu_s-1}<\boldsymbol{q}_s<a_{\nu_s}\}\\ &=\Pr'\{\boldsymbol{Q}\in I^*(\nu_1, \cdots, \nu_s)\}.\end{aligned}$$

Hence we obtain the desired result as follows:

$$\begin{aligned}
&\sum_{(\nu_1,\cdots,\nu_s)} \Pr' \{\boldsymbol{Q}\in I(\nu_1,\cdots,\nu_s)\} \\
&=\sum_{(\nu_1,\cdots,\nu_s)} \Pr' \{\boldsymbol{Q}\in I^*(\nu_1,\cdots,\nu_s)\} \\
&=\Pr' \left\{\boldsymbol{Q}\in \bigcup_{(\nu_1,\cdots,\nu_s)} I^*(\nu_1,\cdots,\nu_s)\right\} \qquad (6.15)\\
&=\Pr' \{\boldsymbol{Q}\in I^*\} \\
&=\Pr' \{0\leqq \boldsymbol{q}_j\leqq 1 \quad \text{for all } j\} \\
&=1,
\end{aligned}$$

that is

$$\sum_{(\nu_1,\cdots,\nu_s)} \Pr' \{\boldsymbol{Q}\in I(\nu_1,\cdots,\nu_s)\}=1. \qquad (6.16)$$

Next we calculate the left side of (6. 11), then

$$\begin{aligned}
&\Pr\left\{\hat{\boldsymbol{P}}^{(n)}\in \bigcup_{(\nu_1,\cdots,\nu_s)} I(\nu_1,\cdots,\nu_s)\right\} \\
&=\Pr\left\{\bigcup_{(\nu_1,\cdots,\nu_s)} (a_{\nu_j-1}<\hat{\boldsymbol{p}}_{ij}^{(n)}<a_{\nu_j} \quad \text{for all } i,j)\right\} \\
&=\Pr\left\{\bigcup_{(\nu_1,\cdots,\nu_s)} (a_{\nu_j-1}<\hat{\boldsymbol{m}}_j^{(n)} \quad \text{and} \quad a_{\nu_j}>\hat{\boldsymbol{M}}_j^{(n)} \quad \text{for all } j)\right\} \qquad (6.17)\\
&\leqq\Pr\left\{\bigcup_{(\nu_1,\cdots,\nu_s)} (\hat{\boldsymbol{M}}_j^{(n)}-\hat{\boldsymbol{m}}_j^{(n)}<\delta \quad \text{for all } j)\right\} \\
&=\Pr \{\hat{\boldsymbol{M}}_j^{(n)}-\hat{\boldsymbol{m}}_j^{(n)}<\delta \quad \text{for all } j\} \\
&\leqq\Pr \{|\hat{\boldsymbol{p}}_{ij}^{(n)}-\hat{\boldsymbol{M}}_j^{(n)}|<\delta \quad \text{for all } i,j\}.
\end{aligned}$$

Combining (6. 16) and (6. 17) with (6. 11), we obtain that for any positive number δ

$$\Pr \{|\hat{\boldsymbol{p}}_{ij}^{(n)}-\hat{\boldsymbol{M}}_j^{(n)}|<\delta \quad \text{for all } i,j\}\to 1 \quad (n\to\infty). \qquad (6.18)$$

If we denote the length of a vector P in s^2-dimensional Euclidean space by $||P||$, we have

$$\begin{aligned}
&\Pr \{||\hat{\boldsymbol{P}}^{(n)}-\hat{\boldsymbol{M}}||>\delta\} \\
&\leqq \sum_{i,j=1}^{s} \Pr \left\{(\hat{\boldsymbol{p}}_{ij}^{(n)}-\hat{\boldsymbol{M}}_j)^2>\frac{\delta^2}{s^2}\right\} \qquad (6.19)\\
&=\sum_{i,j=1}^{s} \Pr \left\{|\hat{\boldsymbol{p}}_{ij}^{(n)}-\hat{\boldsymbol{M}}_j|>\frac{\delta}{s}\right\}.
\end{aligned}$$

For any positive number δ, each term in the last summation in (6. 19) tends to zero as $n\to\infty$, so we have proved that (a′) implies (c). Q.E.D.

THEOREM 6. *A stationary and irreducible M.C. with R.T.M. is ergodic if, and only if,*

$$\Pr\{\hat{\boldsymbol{m}}_j(\omega)>0\}>0 \quad \text{for some } j, \tag{6. 20}$$

or equivalently if, and only if, for some j and N

$$\Pr\{\hat{\boldsymbol{p}}_{ij}^{(N)}(\omega)>0 \quad \text{for all } i\}=\Pr\{\boldsymbol{p}_{ij}^{(N)}(\omega)>0 \quad \text{for all } i\}>0. \tag{6. 21}$$

Theorem 6 is an easy corollary of Theorem 8 or of Theorem 9, but the proof of Theorem 8 is complicated while the proof of Theorem 6 is rather simpler by using the dual processes. The structures of both proofs are similar to each other, hence we shall prove Theorem 6 first and then modify it for Theorem 8. To prove these theorems we need the following lemma.

LEMMA 7. *Let $P=(p_{ij})$, $Q=(q_{ij})$ and $R=QP=(r_{ij})$ be stochastic matrices (i.e., they are elements of $\mathcal{P}$). We denote the maximum and the minimum of the j-th column of P by*

$$M_j=\operatorname*{Max}_{1\leqq i\leqq s} p_{ij} \quad \text{and} \quad m_j=\operatorname*{Min}_{1\leqq i\leqq s} p_{ij} \tag{6. 22}$$

and similarly those of R by

$$M_j'=\operatorname*{Max}_{1\leqq i\leqq s} r_{ij} \quad \text{and} \quad m_j'=\operatorname*{Min}_{1\leqq i\leqq s} r_{ij}. \tag{6. 23}$$

If for some j_0 there is a number $\delta>0$ such that

$$q_{ij_0}>\delta \tag{6. 24}$$

for every i, then

$$M_j'-m_j'\leqq\left(1-\frac{\delta}{2}\right)(M_j-m_j) \tag{6. 25}$$

for every j.

Proof of Lemma 7. Through this proof, j is arbitrarily fixed. When $M_j=m_j$, (6. 25) is trivial, because $m_j\leqq m_j'\leqq M_j'\leqq M_j$ as in (5. 11). Hence we may assume that $M_j>m_j$. Let J be a subset of S defined by

$$J=\left\{i\,\middle|\, p_{ij}\leqq\frac{1}{2}(M_j+m_j)\right\}. \tag{6. 26}$$

Since

$$r_{ij}=\sum_{k=1}^{s} q_{ik}p_{kj}, \tag{6. 27}$$

we have for some i

$$
\begin{aligned}
M_j' = r_{ij} &= \sum_{k\in J} q_{ik} p_{kj} + \sum_{k\notin J} q_{ik} p_{kj} \\
&\leqq M_j \sum_{k\in J} q_{ik} + \frac{1}{2}(M_j + m_j) \sum_{k\notin J} q_{ik} \qquad (6.28)\\
&= M_j - \frac{1}{2}(M_j - m_j) \sum_{k\notin J} q_{ik}.
\end{aligned}
$$

Similarly, for some i' we have

$$m_j' = r_{i'j} \geqq m_j + \frac{1}{2}(M_j - m_j) \sum_{k\in J} q_{i'k}. \tag{6.29}$$

Subtracting (6. 29) from (6. 28) we obtain the desired result:

$$
\begin{aligned}
M_j' - m_j' &\leqq (M_j - m_j)\left(1 - \frac{1}{2}\sum_{k\in J} q_{ik} - \frac{1}{2}\sum_{k\notin J} q_{i'k}\right) \qquad (6.30)\\
&< (M_j - m_j)\left(1 - \frac{\delta}{2}\right),
\end{aligned}
$$

for if $j_0 \in J$ then $\sum_{k\in J} q_{i'k} > \delta$ and if $j_0 \notin J$ then $\sum_{k\notin J} q_{ik} > \delta$. Q.E.D.

Proof of Theorem 6. First we shall prove the necessity. By Theorem 5 we may suppose that

$$\hat{\boldsymbol{p}}_{ij}^{(n)} \to \hat{\boldsymbol{M}}_j \qquad (n \to \infty) \quad \text{w.p. 1.} \tag{6.31}$$

for every i, j and so we have

$$\hat{\boldsymbol{m}}_j = \hat{\boldsymbol{M}}_j \qquad \text{w.p. 1.} \tag{6.32}$$

Since

$$\sum_{j=1}^{s} \hat{\boldsymbol{p}}_{ij}^{(n)} = 1 \tag{6.33}$$

for every i and n, (6. 31) and (6. 32) implies that

$$\sum_{j=1}^{s} \hat{\boldsymbol{m}}_j = 1 \qquad \text{w.p. 1.} \tag{6.34}$$

and so we conclude (6. 20).

Next we shall prove the sufficiency by showing that (6. 21) implies (c) in Theorem 5. We may restate (6. 21) as follows; for some j there is an integer N and a positive number δ such that

$$\Pr\{\hat{\boldsymbol{p}}_{ij}^{(N)} > \delta \quad \text{for all i}\} > 0. \tag{6.35}$$

For these N and δ, let

(6. 36) $$A_k=\{\omega\in\Omega\mid {}^{kN}\hat{\boldsymbol{p}}_{ij}^{(N)}>\delta \quad \text{for all } i\}, \qquad (k=1,2,3,\cdots).$$

Then A_k are mutually independent events and moreover from (5. 5) and (6. 35) we have

(6. 37) $$\sum_{k=1}^{\infty}\Pr\{A_k\}=\infty.$$

Therefore using Borel-Cantelli lemma, it follows that

(6. 38) $$\Pr\{A_k \text{ occurs infinitely often}\}=1.$$

In other words, if the random variable $\boldsymbol{K}^{(n)}(\omega)$ denotes the number of occurrences within the n events $\{A_1, A_2, \cdots, A_n\}$, for each integer r

(6. 39) $$\Pr\{\boldsymbol{K}^{(n)}\geqq r\}\to 1 \qquad (n\to\infty).$$

We divide the whole space into 2^n events as

(6. 40) $$\Omega=\bigcup_{(i_1,\cdots,i_n)}\left\{\bigcap_{k=1}^{n} B_{ki_k}\right\}$$

where $i_k=0$ or 1, and $B_{k0}=A_k$, $B_{k1}=A_k^c$. Then the number of zeros in $\{i_k, k=1,\cdots,n\}$ is equal to the value of $\boldsymbol{K}^{(n)}$. If $\omega\in B_{k0}=A_k$ then by Lemma 7 we have

(6. 41) $$\hat{\boldsymbol{M}}_j^{((k+1)N)}(\omega)-\hat{\boldsymbol{m}}_j^{((k+1)N)}(\omega)\leqq\left(1-\frac{\delta}{2}\right)(\hat{\boldsymbol{M}}_j^{(kN)}(\omega)-\hat{\boldsymbol{m}}_j^{(kN)}(\omega)).$$

Because, we may replace P, Q and R in Lemma 7 by $\hat{\boldsymbol{P}}^{(kN)}$, ${}^{kN}\hat{\boldsymbol{P}}^{(N)}$ and $\hat{\boldsymbol{P}}^{((k+1)N)}$ respectively.

Now we use the inequality (6. 41) for $\omega\in B_{k0}$ and use the inequality

(6. 42) $$\hat{\boldsymbol{M}}_j^{((k+1)N)}(\omega)-\hat{\boldsymbol{m}}_j^{((k+1)N)}(\omega)\leqq\hat{\boldsymbol{M}}_j^{(kN)}(\omega)-\hat{\boldsymbol{m}}_j^{(kN)}(\omega)$$

for $\omega\in B_{k1}$, then for $\omega\in\{\boldsymbol{K}^{(n)}\geqq r\}$ we have

(6. 43) $$\hat{\boldsymbol{M}}_j^{((k+1)N)}(\omega)-\hat{\boldsymbol{m}}_j^{((k+1)N)}\leqq\left(1-\frac{\delta}{2}\right)^r.$$

For each $\delta'>0$ there is an integer r such that $(1-\delta/2)^r<\delta'$, and therefore for arbitrarily small δ' we have

(6. 44) $$\Pr\{\hat{\boldsymbol{M}}_j^{(nN)}-\hat{\boldsymbol{m}}_j^{(nN)}<\delta'\}\geqq\Pr\{\boldsymbol{K}^{(n)}\geqq r\}\to 1 \qquad (n\to\infty).$$

By the monotonicity of $\hat{\boldsymbol{M}}_j^{(n)}$ and $\hat{\boldsymbol{m}}_j^{(n)}$, (6. 44) implies that

(6. 45) $$\Pr\{\hat{\boldsymbol{M}}_j^{(n)}-\hat{\boldsymbol{m}}_j^{(n)}<\delta'\}\to 1 \qquad (n\to\infty).$$

So we have

(6. 46) $$\Pr\{|\hat{p}_{ij}^{(n)}-\hat{M}_j|<\delta'\}\to 1 \qquad (n\to\infty).$$

Repeating the discussion at the last paragraph in the proof of Theorem 5, we complete the proof.

Next we shall consider the non-stationary case where in general no limit exist. However, by generalizing Theorem 6, we can show that the effect of the initial variable vanishes in the long run.

THEOREM 8. *If there is an increasing sequence* $\{n_k\}$ *of integers and a positive number* δ *such that*

(6. 47) $$\sum_{k=1}^{\infty}\Pr\{{}^{n_k}p_{ij}^{(n_{k+1}-n_k)}>\delta \textit{ for some } j \textit{ and all } i\}=\infty,$$

then for any positive number ε

(6. 48) $$\Pr\{M_j^{(n)}(\omega)-m_j^{(n)}(\omega)<\varepsilon \textit{ for every } j\}\to 1 \qquad (n\to\infty).$$

Proof. For a non-stationary M.C. with R.T.M., the basic relation (5. 5) does not hold and so we cannot use the concept of the dual process. Hence we shall define a substitutional process with which this proof can be done in parallel with the proof of Theorem 6.

Select a large integer L and define random variables $\tilde{P}_L^t(\omega)$ $(t=1,\cdots,L)$ by

(6. 49) $$\tilde{P}_L^t(\omega)=P^{L+1-t}(\omega), \qquad \omega\in\Omega$$

and a process $\{\tilde{P}_L^{(n)}\}$ by

(6. 50) $$\tilde{P}_L^{(n)}(\omega)=\tilde{P}_L^n(\omega)\cdots\tilde{P}_L^1(\omega)={}^{L-n}P^{(n)}(\omega) \qquad (n=1,\cdots,L).$$

We will use the same symbols with tilde and L, instead of hat, for corresponding variables as those in the dual process.

Let us denote the events in the braces of (6. 47) by

(6. 51) $$C_k=\{\omega\in\Omega\,|\,{}^{L-n_{k+1}}\tilde{p}_{ij}^{(n_{k+1}-n_k)}(\omega)>\delta \text{ for some } j \text{ and all } i\}$$

for $n_{k+1}\leqq L$, then $\{C_k; n_{k+1}\leqq L\}$ are mutually independent. Hence by (6. 47), Borel-Cantelli lemma assures that

(6. 52) $$\Pr\{C_k \text{ occurs infinitely often}\}=1$$

or if the random variable $\tilde{K}^{(n)}(\omega)$ denotes the number of occurrences within the n events $\{C_1, C_2,\cdots,C_n\}$, then for each integer r

(6. 53) $$\Pr\{\tilde{K}^{(n)}\geqq r\}\to 1 \qquad (n\to\infty).$$

Next, let $k_0=\operatorname{Max}\{k\,|\,n_{k+1}\leqq L\}$ and divide Ω as

(6. 54) $$\Omega=\bigcup_{(i_1,\cdots,i_n)}\left\{\bigcap_{k=1}^{k_0}\tilde{B}_{ki_k}\right\}$$

where $i_k=0$ or 1, and $\tilde{B}_{k0}=C_k$, $\tilde{B}_{k1}=C_k^c$. For $\omega\in C_k$, by Lemma 7,

$$\tilde{\boldsymbol{M}}_{Lj}^{(L-n_k)}(\omega)-\tilde{\boldsymbol{m}}_{Lj}^{(L-n_k)}(\omega)\leqq\left(1-\frac{\delta}{2}\right)(\tilde{\boldsymbol{M}}_{Lj}^{(L-n_{k+1})}(\omega)-\tilde{\boldsymbol{m}}_{Lj}^{(L-n_{k+1})}(\omega)). \tag{6. 55}$$

Because, we may replace P, Q and R in Lemma 7 by $\tilde{\boldsymbol{P}}_L^{(L-n_{k+1})}$, ${}_{L-n_{k+1}}\tilde{\boldsymbol{P}}_L^{(n_{k+1}-n_k)}$ and $\tilde{\boldsymbol{P}}_L^{(L-n_k)}$ respectively. Therefore, if we use the inequality (6. 55) for $\omega\in\tilde{B}_{k0}$ and use the inequality

$$\tilde{\boldsymbol{M}}_{Lj}^{(L-n_k)}(\omega)-\tilde{\boldsymbol{m}}_{Lj}^{(L-n_k)}(\omega)\leqq\tilde{\boldsymbol{M}}_{Lj}^{(L-n_{k+1})}(\omega)-\tilde{\boldsymbol{m}}_{Lj}^{(L-n_{k+1})}(\omega) \tag{6. 56}$$

for $\omega\in\tilde{B}_{k1}$, then for $\omega\in\{\tilde{\boldsymbol{K}}^{(k_0)}\geqq r\}$ we have

$$\begin{aligned}\boldsymbol{M}_j^{(L)}(\omega)-\boldsymbol{m}_j^{(L)}(\omega)&=\tilde{\boldsymbol{M}}_{Lj}^{(L)}(\omega)-\tilde{\boldsymbol{m}}_{Lj}^{(L)}(\omega)\\ &\leqq(\tilde{\boldsymbol{M}}_{Lj}^{(L-n_{k_0})}(\omega)-\tilde{\boldsymbol{m}}_{Lj}^{(L-n_{k_0})}(\omega))\left(1-\frac{\delta}{2}\right)^r\\ &\leqq\left(1-\frac{\delta}{2}\right)^r.\end{aligned} \tag{6. 57}$$

For each $\delta'>0$, there is an integer r such that $(1-\delta/2)^r<\delta'$. Hence we have

$$\Pr\{\boldsymbol{M}_j^{(L)}-\boldsymbol{m}_j^{(L)}<\delta' \quad \text{for all } j\}\geqq\Pr\{\tilde{\boldsymbol{K}}^{(k_0)}\geqq r\}. \tag{6. 58}$$

By (6. 53), for each $\varepsilon>0$, there is an integer k' such that for $k\geqq k'$

$$\Pr\{\tilde{\boldsymbol{K}}^{(k')}\geqq r\}>1-\varepsilon. \tag{6. 59}$$

Therefore for any $L(\geqq n_{k'})$

$$\Pr\{\boldsymbol{M}_j^{(L)}-\boldsymbol{m}_j^{(L)}<\delta' \quad \text{for all } j\}>1-\varepsilon \tag{6. 60}$$

which shows (6. 48).

Now we shall prove one more theorem which is useful to examine the structure of a non-ergodic, stationary and irreducible chain.

THEOREM 9. *A stationary and irreducible M.C. with R.T.M. is ergodic if, and only if, for each pair of subsets I and K of S ($I\neq\phi$, $K\neq\phi$ and $I\cap K=\phi$)*

$$\Pr\{\textit{there is an } \boldsymbol{n}=\boldsymbol{n}(\omega) \textit{ such that } \hat{\boldsymbol{p}}_{ik}^{(\boldsymbol{n})}(\omega)=0 \textit{ for } i\in I,\ k\notin I \textit{ and for } i\in K,\ k\notin K\}<1. \tag{6. 61}$$

Proof. We first prove the sufficiency in several steps.

(i) Since the number of possible pairs (I, K) is finite, (6. 61) implies that there are positive numbers δ and γ such that for each pair (I, K)

$$\Pr\{\text{there is an } \boldsymbol{n}=\boldsymbol{n}(\omega) \text{ such that } \hat{\boldsymbol{p}}_{ik}^{(\boldsymbol{n})}(\omega)<\delta \text{ for } i\in I,\ k\notin I \text{ and for } i\in K,\ k\notin K\}<1-\gamma. \tag{6. 62}$$

For arbitrarily fixed j, let $\{q_{ij}\}$ $(i=1, \cdots, s)$ be a set of rational numbers satisfying the following conditions:

(6. 63) $$0 \leqq q_{ij} \leqq 1 \qquad i=1, \cdots, s$$

and

(6. 64) $$\operatorname*{Max}_{1 \leqq i \leqq s} q_{ij} = M > m = \operatorname*{Min}_{1 \leqq i \leqq s} q_{ij}.$$

We put

(6. 65) $$I = \{i \mid q_{ij} = M\}$$

and

(6. 66) $$K = \{i \mid q_{ij} = m\}$$

and define the positive numbers d and ε by

(6. 67) $$d = \operatorname{Min}\left[\operatorname*{Min}_{i \notin I} (M - q_{ij}),\ \operatorname*{Min}_{i \notin K} (q_{ij} - m)\right]$$

and

(6. 68) $$\varepsilon = \frac{d\delta}{2}.$$

(ii) From now on we concentrate on the j-th column of $\hat{\boldsymbol{P}}^{(n)}(\omega)$. We define the sequence of random times $\boldsymbol{n}_t(\omega)$ inductively by

(6. 69) $$\boldsymbol{n}_1(\omega) = \begin{cases} \operatorname{Min} \{n \mid |\hat{\boldsymbol{p}}_{ij}^{(n)}(\omega) - q_{ij}| < \varepsilon \quad \text{for all } i\}, \\ \infty \quad \text{if the set in above braces is empty,} \end{cases}$$

and

(6. 70) $$\boldsymbol{n}_t(\omega) = \begin{cases} \operatorname{Min} \{n > \boldsymbol{n}_{t-1}(\omega) \mid |\hat{\boldsymbol{p}}_{ij}^{(n)}(\omega) - q_{ij}| < \varepsilon \quad \text{for all } i\}, \\ \infty \quad \text{if } \boldsymbol{n}_{t-1}(\omega) = \infty \text{ or if } \boldsymbol{n}_{t-1}(\omega) < \infty \text{ and the} \\ \qquad \text{set in above braces is empty.} \end{cases}$$

Then if $\boldsymbol{n}_{t+1}(\omega) < \infty$, we have

(6. 71) $${}^{\boldsymbol{n}_t}\hat{\boldsymbol{p}}_{ik}^{(\boldsymbol{n}_{t+1} - \boldsymbol{n}_t)} < \frac{2\varepsilon}{d} = \delta$$

for $i \in I$, $k \notin I$ and for $i \in K$, $k \notin K$. In fact, since

(6. 72) $$\hat{\boldsymbol{p}}_{ij}^{(\boldsymbol{n}_{t+1})} = \sum_{k=1}^{s} {}^{\boldsymbol{n}_t}\hat{\boldsymbol{p}}_{ik}^{(\boldsymbol{n}_{t+1} - \boldsymbol{n}_t)} \hat{\boldsymbol{p}}_{kj}^{(\boldsymbol{n}_t)},$$

for $i\in I$ and $k_0\notin I$ we have

$$
\begin{aligned}
M-\varepsilon<\hat{p}_{ij}^{(n_{t+1})}&<\sum_{k=1}^{s}{}^{n_t}\hat{p}_{ik}^{(n_{t+1}-n_t)}(q_{kj}+\varepsilon)\\
&=\varepsilon+\sum_{k\neq k_0}q_{kj}\hat{p}_{jk}^{(n_{t+1}-n_t)}+q_{k_0j}{}^{n_t}\hat{p}_{ik_0}^{(n_{t+1}-n_t)}\\
&\leqq\varepsilon+M(1-{}^{n_t}\hat{p}_{ik_0}^{(n_{t+1}-n_t)})+q_{k_0j}{}^{n_t}\hat{p}_{ik_0}^{(n_{t+1}-n_t)}\\
&=\varepsilon+M-(M-q_{k_0j}){}^{n_t}\hat{p}_{ik_0}^{(n_{t+1}-n_t)}.
\end{aligned}
\tag{6.73}
$$

Hence for $i\in I$, $k_0\notin I$ we have

$$
{}^{n_t}\hat{p}_{ik_0}^{(n_{t+1}-n_t)}<\frac{2\varepsilon}{M-q_{k_0j}}\leqq\frac{2\varepsilon}{d}=\delta. \tag{6.74}
$$

Similarly the same relation holds for $i\in K$, $k_0\notin K$.

(iii) Next we shall show that

$$
\Pr\{n_{t+1}(\omega)<\infty\}<(1-\gamma)^t. \tag{6.75}
$$

We may divide the event $\{n_{t+1}(\omega)<\infty\}$ into disjoint events as

$$
\{n_{t+1}(\omega)<\infty\}=\bigcup_{\nu<\infty}\bigcup_{\mu<\infty}\{n_t(\omega)=\nu,\ n_{t+1}(\omega)=\nu+\mu\}. \tag{6.76}
$$

By the result obtained in (ii) we have

$$
\{n_t(\omega)=\nu,\ n_{t+1}(\omega)=\nu+\mu\}\subset\{{}^{\nu}\hat{p}_{ik}^{(\mu)}(\omega)<\delta\quad\text{for } i\in I,\ k\notin I \text{ and for } i\in K,\ k\notin K\}, \tag{6.77}
$$

while

$$
\{n_t(\omega)=\nu,\ n_{t+1}(\omega)=\nu+\mu\}\subset\{n_t(\omega)=\nu\}. \tag{6.78}
$$

Therefore

$$
\begin{aligned}
&\{n_t(\omega)=\nu,\ n_{t+1}(\omega)=\nu+\mu\}\\
&\subset\{{}^{\nu}\hat{p}_{ik}^{(\mu)}(\omega)<\delta \text{ for } i\in I,\ k\notin I \text{ and for } i\in K,\ k\notin K\}\cap\{n_t(\omega)=\nu\}.
\end{aligned}
\tag{6.79}
$$

The event $\{n_t=\nu\}$ depends only on the fraction $(\omega^1,\cdots,\omega^{\nu})$ of $\omega=(\omega^0,\omega^1,\omega^2,\cdots)$, and the event $\{{}^{\nu}\hat{p}_{ik}^{(\mu)}<\delta$ for $i\in I$, $k\notin I$ and for $i\in K$, $k\notin K\}$ depends on the fraction $(\omega^{\nu+1},\cdots,\omega^{\nu+\mu})$. So they are independent. Hence we have

$$
\begin{aligned}
&\Pr\{n_t=\nu,\ n_{t+1}=\nu+\mu\}\\
&\leqq\Pr\{{}^{\nu}\hat{p}_{ik}^{(\mu)}<\delta \text{ for } i\in I,\ k\notin I \text{ and for } i\in K,\ k\notin K\}\cdot\Pr\{n_t=\nu\}.
\end{aligned}
\tag{6.80}
$$

Therefore

$$\Pr\{\boldsymbol{n}_{t+1}<\infty\}=\sum_{\nu<\infty}\sum_{\mu<\infty}\Pr\{\boldsymbol{n}_t=\nu,\ \boldsymbol{n}_{t+1}=\nu+\mu\}$$

$$\leqq\sum_{\nu<\infty}\sum_{\mu<\infty}\Pr\{\boldsymbol{n}_t=\nu\}\Pr\{{}^{\nu}\hat{\boldsymbol{p}}_{ik}^{(\mu)}<\delta \text{ for } i\in I,\ k\notin I \text{ and for } i\in K,\ k\notin K\}$$

$$=\sum_{\nu<\infty}\Pr\{\boldsymbol{n}_t=\nu\}\cdot\Pr\{\text{there is an } \boldsymbol{n}(\omega)>\nu \text{ such that } {}^{\nu}\hat{\boldsymbol{p}}_{ik}^{(\boldsymbol{n})}<\delta$$

$$\text{for } i\in I,\ k\notin I \text{ and for } i\in K,\ k\notin K\} \tag{6. 81}$$

$$=\sum_{\nu<\infty}\Pr\{\boldsymbol{n}_t=\nu\}\cdot\Pr\{\text{there is an } \boldsymbol{n}(\omega)>0 \text{ such that } \hat{\boldsymbol{p}}_{ik}^{(\boldsymbol{n})}<\delta$$

$$\text{for } i\in I,\ k\notin I \text{ and for } i\in K,\ k\notin K\}$$

$$<(1-\delta)\sum_{\nu<\infty}\Pr\{\boldsymbol{n}_t=\nu\}=(1-\delta)\Pr\{\boldsymbol{n}_t<\infty\}.$$

Using this relation repeatedly we obtain the desired result (6. 75).

(iv) If $\hat{\boldsymbol{p}}_{ij}^{(n)}(\omega)$ $(i=1,\cdots,s)$ has an accumulation point in the interval $(q_{ij}-\varepsilon/2,\ q_{ij}+\varepsilon/2)$ $(i=1,\cdots,s)$, then $\hat{\boldsymbol{p}}_{ij}^{(n)}(\omega)$ visits the interval $(q_{ij}-\varepsilon,\ q_{ij}+\varepsilon)$ $(i=1,\cdots,s)$ infinitely often. Therefore if $A(q_{ij})$ denotes the event that $\hat{\boldsymbol{p}}_{ij}^{(n)}(\omega)$ has an accumulation point in the interval $(q_{ij}-\varepsilon/2,\ q_{ij}+\varepsilon/2)$ $(i=1,\cdots,s)$, then for every t

$$A(q_{ij})\subset\{\boldsymbol{n}_t<\infty\}. \tag{6. 82}$$

Hence

$$\Pr\{A(q_{ij})\}\leqq\lim_{t\to\infty}\Pr\{\boldsymbol{n}_t<\infty\}\leqq\lim_{t\to\infty}(1-\gamma)^{t-1}=0. \tag{6. 83}$$

(v) Now we shall consider the case in which the chain is not ergodic. Let B be the event that $\hat{\boldsymbol{P}}^{(n)}(\omega)$ does not converge to $\hat{\boldsymbol{M}}(\omega)$. For every $\omega\in B$, there exists some j and $\{q_{ij}\}$ such that $\omega\in A(q_{ij})$. Therefore

$$B\subset\bigcup_{j=1}^{s}\bigcup_{\{q_{ij}\}}A\{q_{ij}\}. \tag{6. 84}$$

Since q_{ij} are rational numbers, the number of possible $\{q_{ij}\}$ is countable. Therefore $\Pr\{B\}=0$. This completes the proof of the sufficiency.

Now we shall prove the necessity. Suppose that there exists a pair of subsets (I, K) of S $(I\neq\phi,\ K\neq\phi$ and $I\cap K=\phi)$ with which

$$\Pr\{\text{there is an } \boldsymbol{n}=\boldsymbol{n}(\omega) \text{ such that } \hat{\boldsymbol{p}}_{ik}^{(\boldsymbol{n})}(\omega)=0$$

$$\text{for } i\in I,\ k\notin I \text{ and for } i\in K,\ k\notin K\}=1. \tag{6. 85}$$

Let D $(\subset\mathcal{P})$ be the set of all $P=(p_{ij})$ such that $p_{ij}=0$ for $i\in I$, $k\notin I$ and for $i\in K$, $k\notin K$. Then the assumption is

$$\Pr\{\text{there is an } \boldsymbol{n}=\boldsymbol{n}(\omega) \text{ such that } \hat{\boldsymbol{P}}^{(\boldsymbol{n})}\in D\}=1, \tag{6. 86}$$

If $\hat{\boldsymbol{P}}^{(n)} \in D$ and ${}^{n}\hat{\boldsymbol{P}}^{(m)} \in D$, then it is easily shown that $\hat{\boldsymbol{P}}^{(n+m)} \in D$. Hence if we define that $\boldsymbol{t} = \text{Min}\{n \mid \hat{\boldsymbol{P}}^{(n)} \in D\}$ then

$$\begin{aligned} &\{\text{there exist at least two } \boldsymbol{n}\text{'s such that } \hat{\boldsymbol{P}}^{(\boldsymbol{n})} \in D\} \\ &= \bigcup_{\nu<\infty} \{\boldsymbol{t}=\nu, \text{ there exists an } \boldsymbol{n} \text{ such that } {}^{\nu}\hat{\boldsymbol{P}}^{(\boldsymbol{n})} \in D\}. \end{aligned} \tag{6.87}$$

Therefore we have

$$\begin{aligned} &\Pr\{\text{there exist at least two } \boldsymbol{n}\text{'s such that } \hat{\boldsymbol{P}}^{(\boldsymbol{n})} \in D\} \\ &= \sum_{\nu<\infty} \Pr\{\boldsymbol{t}=\nu\} \cdot \Pr\{\text{there exists an } \boldsymbol{n} \text{ such that } {}^{\nu}\hat{\boldsymbol{P}}^{(\boldsymbol{n})} \in D\} \\ &= \sum_{\nu<\infty} \Pr\{\boldsymbol{t}=\nu\} \cdot \Pr\{\text{there exists an } \boldsymbol{n} \text{ such that } \hat{\boldsymbol{P}}^{(\boldsymbol{n})} \in D\} \\ &= \sum_{\nu<\infty} \Pr\{\boldsymbol{t}=\nu\} \cdot 1 \\ &= \Pr\{\text{there exists an } \boldsymbol{n} \text{ such that } \hat{\boldsymbol{P}}^{(\boldsymbol{n})} \in D\} = 1. \end{aligned} \tag{6.88}$$

Similarly for each m we can show that

$$\Pr\{\text{there exist at least } m \text{ epochs } (\boldsymbol{n}\text{'s}) \text{ such that } \hat{\boldsymbol{P}}^{(\boldsymbol{n})} \in D\} = 1. \tag{6.89}$$

Therefore for every N, there is an $\boldsymbol{n} > N$ with probability one such that $\hat{\boldsymbol{P}}^{(\boldsymbol{n})} \in D$. Hence by the monotonicity of $\hat{\boldsymbol{m}}_j^{(n)}$ we have

$$\hat{\boldsymbol{m}}_j^{(N)} \leqq \hat{\boldsymbol{m}}_j^{(\boldsymbol{n})} \leqq \hat{\boldsymbol{p}}_{ij}^{(\boldsymbol{n})} = 0 \tag{6.90}$$

for $\boldsymbol{n} = \boldsymbol{n}(\omega) > N$ with which $\hat{\boldsymbol{P}}^{(\boldsymbol{n})} \in D$ and for $i \in I$, $j \notin I$ and $i \in K$, $j \notin K$. Since $I^c \cup K^c = S$, we have for each j

$$\hat{\boldsymbol{m}}_j^{(N)} = 0 \tag{6.91}$$

with probability one, and as N is arbitrary, this implies that

$$\hat{\boldsymbol{m}}_j = \lim_{N\to\infty} \hat{\boldsymbol{m}}_j^{(N)} = 0. \tag{6.92}$$

On the other hand, for every N and i

$$\sum_{j=1}^{s} \hat{\boldsymbol{M}}_j^{(N)} \geqq \sum_{j=1}^{s} \hat{\boldsymbol{p}}_{ij}^{(N)} = 1. \tag{6.93}$$

Therefore for each ω at least one $\hat{\boldsymbol{M}}_j = \lim_{N\to\infty} \hat{\boldsymbol{M}}_j^{(N)}$ is strictly positive, and this contradicts to (d) in Theorem 5.

ACKNOWLEDGEMENT. The author wishes to thank Professor H. Morimura for the suggestion and the guidance.

References

[1] Ehrenberg, A. S. C., An appraisal of Markov brand-switching models. Journ. of Marketing Research **2** (1965), 347–362.

[2] Feller, W., An introduction to probability theory and its applications. Vol. 1 (3rd ed.), John Wiley and Sons, Inc., New York (1968).

[3] Feller, W., An introduction to probability theory and its applications. Vol 2, John Wiley and Sons, Inc., New York (1966).

Department of Applied Physics,
Tokyo Institute of Technology.