The Annals of Probability
1975, Vol. 3, No. 4, 677–696

A LIMIT THEOREM FOR PARTIALLY OBSERVED MARKOV CHAINS

BY THOMAS KAIJSER

Linköping University

Let $\{X_n, n = 1, 2, \cdots\}$ be a Markov chain with finite state space $S = \{1, 2, \cdots, d\}$, transition probability matrix P and initial distribution p. Let g be a function with S as domain and define $Y_n = g(X_n)$. Define

$$Z_n^i = \Pr[X_n = i \mid Y_1, Y_2, \cdots, Y_n],$$
$$Z_n = (Z_n^1, Z_n^2, \cdots, Z_n^d),$$

and let μ_n denote the probability distribution of the vector Z_n. In this paper we prove that if $\{X_n, n = 1, 2, \cdots\}$ is ergodic and if P and g satisfy a certain condition then μ_n converges to a limit and this limit is independent of the initial distribution p.

1. Introduction. Let $\{X_n, n = 1, 2, \cdots\}$ be a Markov chain with finite state space $S = \{1, 2, \cdots, d\}$, (stationary) transition probability matrix (tr p m) $P = (p_{i,j})$, and initial distribution $p = (p_1, p_2, \cdots, p_d)$ where

$$p_i = \Pr[X_1 = i], \qquad i = 1, 2, \cdots, d.$$

Let g be a function with domain S and define

$$Y_n = g(X_n), \qquad n = 1, 2, \cdots.$$

A process $\{Y_n\}$ constructed in this way is usually called a *partially observed* Markov chain. Put

$$Z_n^i = \Pr[X_n = i \mid Y_1, Y_2, \cdots, Y_n], \qquad i = 1, 2, \cdots, d, n = 1, 2, \cdots$$

and

$$Z_n = (Z_n^1, Z_n^2, \cdots, Z_n^d).$$

The purpose of this paper is to prove that if $\{X_n\}_1^\infty$ is ergodic and if P and g satisfy a certain condition A—specified below—then

(i) the probability measure of Z_n converges weakly to a limit measure, and
(ii) the limit measure is independent of the initial distribution p.

We shall also give an example (see Section 10) showing that the second of these results does not hold if we merely assume ergodicity, thus contradicting a conjecture made by D. Blackwell (see [2] pages 17–18).

In order to give a precise statement of our result we first need some notations.

Received June 8, 1974; revised September 6, 1974.

AMS 1970 *subject classifications.* Primary 60J10; Secondary 60J05, 60F99.

Key words and phrases. Partially observed Markov chains, random systems with complete connection, products of random matrices.

Let

$$K = \{(x_1, x_2, \cdots, x_d) \in R^d : x_i \geqq 0, \sum x_i = 1\}$$
$$\mathscr{E} = \text{the collection of Borel sets of } K$$
$$C[K] = \text{the collection of real-valued, continuous functions on } K$$
$$\nu_{n,p}(E) = \Pr[Z_n \in E], \qquad E \in \mathscr{E}$$

where the subscript p indicates that the distribution of X_1 is taken to be p. Further let

$$A = \{g(i) : i \in S\}$$
$$S(a) = \{i : g(i) = a\}, \qquad a \in A$$

and define a matrix $M(a) = (m_{i,j}(a))$, for $a \in A$, by

$$m_{i,j}(a) = p_{ij} \quad \text{if } g(j) = a$$
$$= 0 \quad \text{otherwise.}$$

We also need a notion for matrices which we shall call subrectangularity.

DEFINITION 1.1. Let $M = (m_{i,j})$ denote a $d \times d$ matrix. If $m_{i_1,j_1} \neq 0$ and $m_{i_2,j_2} \neq 0$ implies that also $m_{i_1,j_2} \neq 0$ and $m_{i_2,j_1} \neq 0$ then we call M a *subrectangular* matrix.

We now state the theorem to be proved in this paper.

THEOREM A. *Suppose the Markov chain $\{X_n\}_1^\infty$ is ergodic. Suppose further that P and g are such that the following condition holds*:

CONDITION A. *There exists a finite sequence $a_1, a_2, \cdots, a_m$ of elements belonging to A such that the matrix product $M(a_1)M(a_2) \cdots M(a_m)$ is a nonzero subrectangular matrix.*

Then there exists a unique probability measure ν on $(K, \mathscr{E})$ such that $\{\nu_{n,p}\}_{n=1}^\infty$ converges weakly to ν for all p; i.e., if $u \in C[K]$ then

$$\lim_{n\to\infty} \int_K u(y)\nu_{n,p}(dy) = \int_K u(y)\nu(dy) .$$

REMARK. In many special cases it is easy to verify Condition A. For example, if the tr p m P itself is subrectangular or if there exists an $a \in A$ corresponding to exactly one element of S then Condition A holds trivially.

The basic idea in the proof is to use the fact that Z_n can be represented as a normalized product of random matrices, an observation due to M. Rudemo (see [8] page 587). Then by using an estimate for products of nonnegative subrectangular matrices, essentially due to H. Furstenberg and H. Kesten (see [4] Lemma 3) and using the fact that $\{Z_n\}_1^\infty$ is a Markov chain we prove the theorem with methods similar to those used in [6].

REMARK. It is worth observing that the "set-up" given above also includes the seemingly more general situation, when the Y_n are *random* functions of the X_n in the sense that

$$(1.1) \qquad \Pr[Y_n = a \mid X_n = i, X_m, Y_m, 1 \leqq m < n] = \Pr[Y_n = a \mid X_n = i] = q_{i,a}$$

where a belongs to a finite set A and $\sum_A q_{i,a} = 1$ for $i \in S$. (If we only allow $q_{i,a}$ to take the values 1 or 0 we have the deterministic case.) For, defining

$$X_n' = (X_n, Y_n)$$

we note from (1.1) that $\{X_n'\}_1^\infty$ is also a Markov chain and then by defining

$$g(X_n') = Y_n$$

we have reduced this "random" case to a "deterministic" case.

2. Some representation formulas. We denote the i, jth element of a matrix M by $(M)_{i,j}$ and similarly we denote (when convenient) the ith component of a vector x by $(x)_i$. Now for each $a \in A$ we define a matrix $I(a)$ by

$$\begin{aligned} (I(a))_{i,j} &= 1 \qquad \text{if} \quad i = j \quad \text{and} \quad g(i) = a \\ &= 0 \qquad \text{otherwise.} \end{aligned}$$

We also define

$$||x|| = \textstyle\sum_1^d |x_i|, \; x \in R^d .$$

Clearly

$$\Pr[Y_1 = a] = \textstyle\sum_{i:g(i)=a} p_i = ||pI(a)||$$

and

$$\Pr[X_1 = i \mid Y_1 = a] = \frac{(pI(a))_i}{||pI(a)||} .$$

Hence

$$Z_1 = \frac{pI(Y_1)}{||pI(Y_1)||}$$

Generalizing we have

LEMMA 2.1. (See Åström [1] pages 182–183 and Rudemo [8] page 586.)

(a)

$$\begin{aligned} \Pr[Y_1 = a_1, Y_2 = a_2, \cdots, Y_n = a_n] \\ = ||pI(a_1)M(a_2)M(a_3) \cdots M(a_n)|| \qquad n = 1, 2, \cdots \end{aligned}$$

(b)

$$Z_n = \frac{pI(Y_1)M(Y_2)M(Y_3) \cdots M(Y_n)}{||pI(Y_1)M(Y_2)M(Y_3) \cdots M(Y_n)||}, \qquad n = 1, 2, \cdots .$$

REMARK. Observe that the denumerator in (b) is equal to zero with probability zero because of (a).

PROOF. Both formulas are simple consequences of Bayes rule and the following obvious fact:

LEMMA 2.2. *Let M_1 and M_2 be two $d \times d$ matrices and let $y \in R^d$ be such that $||yM_1|| \neq 0$ and $||yM_1M_2|| \neq 0$. Then*

$$\frac{\frac{yM_1}{||yM_1||} \cdot M_2}{\left\|\frac{yM_1}{||yM_1||} \cdot M_2\right\|} = \frac{yM_1M_2}{||yM_1M_2||} .$$

Next let $\mathscr{A}$ denote the subsets of A and define $Q : K \times \mathscr{A} \to [0, 1]$ by

$$Q(x, B) = \textstyle\sum_{a \in B} \|xM(a)\| , \qquad x \in K, B \in \mathscr{A} . \tag{2.1}$$

Clearly $Q : K \times \mathscr{A} \to [0, 1]$ defines a transition probability function (tr p f). Further let

$$D_1 = \{(x, a) \in K \times A : \|xM(a)\| > 0\} ,$$

define $h : D_1 \to K$ by

$$h(x, a) = \frac{xM(a)}{\|xM(a)\|} \tag{2.2}$$

and let

$$A(h^{-1}(x, E)) = \{a \in A : h(x, a) \in E\} .$$

LEMMA 2.3. (See Åström [1] page 187 or Blackwell [2] pages 14, 15). *The process* $\{Z_n\}_1^\infty$ *is a Markov chain with state space* $(K, \mathscr{E})$ *initial distribution* $\nu_{1,p}$ *and* tr p f $R : K \times \mathscr{E} \to [0, 1]$ *defined by*

$$R(x, E) = Q(x, A(h^{-1}(x, E))) . \tag{2.3}$$

PROOF. The lemma follows easily from (b) of Lemma 2.1, Lemma 2.2 and the definitions of $Q(x, B)$, $h(x, a)$ and $A(h^{-1}(x, E))$.

Next let x be an arbitrary element of K, let $\{Z_n(x)\}_{n=0}^\infty$ denote the Markov chain which starts at x ($Z_0(x) = x$) and has tr p f $R(\cdot, \cdot)$, and let $\mu_{n,x}$ denote the probability distribution of the vector $Z_n(x)$. The following corollary is an the definitions immediate consequence of Lemma 2.3.

COROLLARY 2.1. *If for all* $x \in K$, $\{\mu_{n,x}\}_{n=0}^\infty$ *converges weakly to a limit which is independent of* x *then also* $\{\nu_{n,p}\}_{n=1}^\infty$ *converges weakly to the same limit for all initial distributions* p.

3. A random system with complete connection. From Corollary 2.1 of the previous section we note that in order to prove Theorem A, if suffices to study the Markov chains $\{Z_n(x)\}_{n=0}^\infty$, $x \in K$ with tr p f $R(\cdot, \cdot)$ defined by (2.3).

We now show how the Markov chain $\{Z_n(x)\}_1^\infty$ is obtained. We start at $Z_0(x) = x$. Then we pick an element $Y_1(x) \in A$ according to $Q(x, \cdot)$ (defined by (2.1)) and take $Z_1(x) = h(x, Y_1(x))$ where $h(x, a)$ is defined by (2.2). Next we pick an element $Y_2(x) \in A$ according to $Q(Z_1(x), \cdot)$ and take $Z_2(x) = h(Z_1(x), Y_2(x))$ etc. The mathematical objects involved in this procedure—namely the two measurable sets $(K, \mathscr{E})$ and $(A, \mathscr{A})$, the tr p f $Q : K \times \mathscr{A} \to [0, 1]$ and the function $h : D_1 \to K$—constitute a set $\{(K, \mathscr{E}), (A, \mathscr{A}), Q, h\}$ called a *random system with complete connection*. (See Iosifescu-Theoderescu [5] Chapter 2, especially Section 2.3.3.1.). At least essentially. There is namely one point at which the set $\{(K, \mathscr{E}), (A, \mathscr{A}), Q, h\}$ does not quite satisfy the definition of a random system with complete connection and that point concerns the function $h : D_1 \to K$ which only is defined on a subset D_1 of $K \times A$ and not on all of $K \times A$. Therefore we shall need some extra concepts and notations which are usually not needed when studying random systems with complete connection.

We shall next justify the construction of $\{Z_n(x)\}_0^\infty$ given above. Let A^n denote the n-product set of A and let A^∞ denote the infinite product set of A. Let $\mathscr{A}^n$ denote the subsets of A^n and let $\mathscr{A}^\infty$ denote the σ-algebra on A^∞ generated by the cylinder-sets. Let $q : K \times A \to [0, 1]$ be defined by

$$q(x, a) = ||xM(a)||$$

($q(x, a)$ can be regarded as the density function of $Q(x, B)$.). Further let $a^n = (a_1, a_2, \cdots, a_n)$ denote an element of A^n and let

$$D_n = \{(x, a^n) \in K \times A^n : ||xM(a_1)M(a_2) \cdots M(a_n)|| > 0\} .$$

Define $q_n : K \times A^n \to [0, 1]$ and $h_n : D_n \to K$, $n = 1, 2, \cdots$ by

$$q_n(x, a^n) = ||xM(a_1)M(a_2) \cdots M(a_n)||$$

and

$$h_n(x, a^n) = \frac{xM(a_1)M(a_2) \cdots M(a_n)}{||xM(a_1)M(a_2) \cdots M(a_n)||} .$$

By applying Lemma 2.2 we observe

LEMMA 3.1. *If* $(x, a^{n+m}) \in D_{n+m}$, $n, m \geqq 1$, *then*

$$q_{n+m}(x, a^{n+m}) = q_n(x, a^n)q_m(h_n(x, a^n), {}^n a^{n+m}) \tag{3.1}$$

and

$$h_{n+m}(x, a^{n+m}) = h_m(h_n(x, a^n), {}^n a^{n+m}) , \tag{3.2}$$

where

$${}^n a^{n+m} = (a_{n+1}, a_{n+2}, \cdots, a_{n+m}) .$$

Next for each $x \in K$ and $n \geqq 1$ we define $Q^n(x, \cdot) : \mathscr{A}^n \to [0, 1]$ by

$$Q^n(x, B) = \textstyle\sum_{a^n \in B} q_n(x, a^n) .$$

From the definition of $q_n(x, a^n)$ we observe that

$$Q^n(x, A) = \textstyle\sum_A ||xM(a_1) \cdots M(a_n)|| = ||xP^n|| = 1$$

for all $n \geqq 1$. Thus $\{Q^n(x, \cdot)\}_1^\infty$ constitutes a sequence of probability measures. Moreover

$$Q^{n+1}(x, B \times A) = Q^n(x, B) , \qquad B \in \mathscr{A}^n$$

and from Lemma 3.1 we also have

$$Q^{n+m}(x, B_1^n \times B_2^m) = \textstyle\sum_{a^n \in B_1^n} q_n(x, a^n) \sum_{a^m \in B_2^m} q_m(h_n(x, a^n), a^m)$$

if $a^{n+m} \in A_x^{n+m}$, where

$$A_x^n = \{a^n \in A^n : ||xM(a_1) \cdots M(a_n)|| > 0\} \tag{3.3}$$
$$(= \text{ the support of } Q^n(x, \cdot)) .$$

Hence, applying a theorem due to Ionescu Tulcea (see [7] Section V.1) we have

EXISTENCE THEOREM. *To every* $x \in K$ *there exists a probability space* $(A^\infty, \mathscr{A}^\infty,$

Q_x^∞) *and a sequence of random variables* $\{Y_n(x)\}_1^\infty$ *on* $(A^\infty, \mathscr{A}^\infty)$ *and with values in* A *such that*

$$Q_x^\infty(Y_1(x) \in B) = Q(x, B), \qquad B \in \mathscr{A}$$

and

$$Q_x^\infty(Y_{n+1}(x) \in B \mid Y_m(x), m = 1, 2, \cdots, n) = Q(h_n(x, Y^n(x)), B), Q_x^\infty - \text{a.s.}, \qquad B \in \mathscr{A},$$

where

$$Y^n(x) = (Y_1(x), Y_2(x), \cdots, Y_n(x)).$$

From this existence theorem and Lemma 3.1 it is easy to see that we also have

LEMMA 3.2. *For* $n \geqq 1$

$$\Pr[Y_1(x) = a_1, Y_2(x) = a_2, \cdots, Y_n(x) = a_n] = ||xM(a_1)M(a_2)\cdots M(a_n)||$$

and

$$\Pr[Z_1(x) \in E_1, Z_2(x) \in E_2, \cdots, Z_n(x) \in E_n] = \Pr[h_1(x, Y^1(x)) \in E_1, h_2(x, Y^2(x)) \in E_2, \cdots, h_n(x, Y^n(x)) \in E_n],$$

where $E_i \in \mathscr{E}$, $i = 1, 2, \cdots, n$. (Compare also Lemma 2.1.)

4. The transition operator. We have already observed that in order to prove Theorem A is suffices to prove that $\{\mu_{n,x}\}_0^\infty$ converges weakly to a limit which is independent of x, where $\mu_{n,x}$ denotes the probability measure corresponding to $Z_n(x)$. Now let $B[K]$ = collection of real-valued, bounded, Borel functions on K and define the *transition operator* $T: B[K] \to B[K]$ by

$$Tu(x) = \int_K u(y)R(x, dy), \qquad u \in B[K]. \tag{4.1}$$

From the definition of $R(\cdot, \cdot)$ (see (2.3)) we note that

$$Tu(x) = \sum_{a:||xM(a)||>0} u\left(\frac{xM(a)}{||xM(a)||}\right)||xM(a)|| = \sum_{A_x^1} u(h(x, a))q(x, a).$$

Next let $R^n(\cdot, \cdot)$ denote the n-step tr p f of $R(\cdot, \cdot)$. By the Chapman–Kolmogorov equality we have

$$T^n u(x) = \int_K u(y)R^n(x, dy)$$

and introducing the notation

$$u_n(x) = T^n u(x), \qquad u \in B[K]$$

we also have

$$u_{n+m}(x) = T^n u_m(x). \tag{4.2}$$

Furthermore it is not hard to see that

$$u_n(x) = \sum_{A_x^n} u(h_n(x, a^n))q_n(x, a^n). \tag{4.3}$$

Now from (4.1) and (4.2) it follows that

$$\sup_{x \in K} u_n(x) \geqq \sup_{x \in K} u_{n+1}(x) \tag{4.4}$$

and

$$\inf_{x \in K} u_n(x) \leqq \inf_{x \in K} u_{n+1}(x) . \tag{4.5}$$

Therefore defining

$$\varphi(u) = \sup_{x \in K} u(x) - \inf_{x \in K} u(x) , \qquad u \in B[K] ,$$

we have

$$\varphi(u_{n+1}) \leqq \varphi(u_n) .$$

LEMMA 4.1. *If for each* $u \in C[K]$

$$\lim_{n \to \infty} \varphi(u_n) = 0 \tag{4.6}$$

then there exists a unique probability measure μ *on* $(K, \mathscr{E})$ *such that* $\{\mu_{n,x}\}_0^\infty$ *converges weakly to* μ *for all* $x \in K$.

PROOF. The lemma is a simple consequence of (4.4), (4.5) and Riesz representation theorem. (Compare [3] Chapter 8, page 243 and page 266.)

Next we define

$$||u|| = \sup_{x \in K} |u(x)| , \qquad u \in B[K] .$$

LEMMA 4.2. *The set* $\{u \in B[K] : \lim_{n \to \infty} \varphi(u_n) = 0\}$ *is closed under the supremum norm topology.*

PROOF. Standard. (See [6] Chapter 1, for details).

From Lemma 4.1, Lemma 4.2 and Corollary 2.1 we see that Theorem A will be proved if we can prove (4.6) for a set of functions which is dense in $C[K]$, for example the set of Lipschitz functions. That is the set

$$\mathrm{Lip}\,[K] = \left\{ u \in C[K] : \sup_{x \neq y} \frac{|u(x) - u(y)|}{||x - y||} < \infty \right\} .$$

The following property of Lip $[K]$ will be used later.

LEMMA 4.3.

$$u \in \mathrm{Lip}\,[K] \Rightarrow Tu \in \mathrm{Lip}\,[K] .$$

PROOF. Let $x = (x_1, x_2, \cdots, x_d)$ and $y = (y_1, y_2, \cdots, y_d)$ be two fixed but arbitrary vectors of K and let $\{e_i : i = 1, 2, \cdots, d\}$ denote the set of base-vectors of R^d. Define

$$S_1 = \{i \in S : x_i > 0 \text{ and } y_i > 0\}$$
$$S_2 = \{i \in S : x_i > 0 \text{ and } y_i = 0\}$$
$$S_3 = \{i \in S : x_i = 0 \text{ and } y_i > 0\}$$

and

$$A_i = \{a \in A : ||e_i M(a)|| > 0\} .$$

Now using the fact that

$$q(x, a) = \sum_i x_i q(e_i, a)$$

it is not very difficult to convince oneself that for any $u \in B[K]$

$$
\begin{aligned}
&|Tu(x) - Tu(y)| \\
&\quad = |\textstyle\sum_{S_1} x_i \sum_{A_i} u(h(x, a))q(e_i, a) + \sum_{S_2} x_i \sum_{A_i} u(h(x, a))q(e_i, a) \\
&\qquad - \textstyle\sum_{S_1} y_i \sum_{A_i} u(h(y, a))q(e_i, a) - \sum_{S_3} y_i \sum_{A_i} u(h(y, a))q(e_i, a)| \\
&\quad \leqq |\textstyle\sum_{S_1} (x_i - y_i) \sum_{A_i} u(h(x, a))q(e_i, a) \qquad (4.7)\\
&\qquad + \textstyle\sum_{S_2} (x_i - y_i) \sum_{A_i} u(h(x, a))q(e_i, a) \\
&\qquad + \textstyle\sum_{S_3} (x_i - y_i) \sum_{A_i} u(h(y, a))q(e_i, a)| \\
&\qquad + |\textstyle\sum_{S_1} y_i \sum_{A_i} (u(h(x, a)) - u(h(y, a)))q(e_i, a)| = I_1 + I_2 \quad \text{say.}
\end{aligned}
$$

Now clearly

$$
I_1 \leqq \textstyle\sum_S |x_i - y_i| \, ||u|| \sum_{A_i} q(e_i, a) = ||x - y|| \, ||u|| \tag{4.8}
$$

since

$$
\textstyle\sum_{A_i} q(e_i, a) = 1 \, .
$$

Furthermore if $i \in S_1$ and $a \in A_i$ then both $||xM(a)||$ and $||yM(a)||$ are larger than zero, and by simple calculations we obtain

$$
\left|\left| \frac{xM(a)}{||xM(a)||} - \frac{yM(a)}{||yM(a)||} \right|\right| \leqq \frac{2||xM(a) - yM(a)||}{||yM(a)||} \, .
$$

Therefore if $u \in \mathrm{Lip}\,[K]$ and we define

$$
|u|_L = \inf \left\{ \gamma : \sup_{x \neq y} \frac{|u(x) - u(y)|}{||x - y||} \leqq \gamma \right\} , \qquad u \in \mathrm{Lip}\,[K]
$$

we obtain

$$
\begin{aligned}
I_2 &\leqq \textstyle\sum_{S_1} y_i \sum_{A_i} |u|_L ||h(x, a) - h(y, a)|| q(e_i, a) \\
&\leqq |u|_L \textstyle\sum_{S_1} y_i \sum_{A_i} \dfrac{2||xM(a) - yM(a)||q(e_i, a)}{||yM(a)||} \\
&\leqq 2|u|_L \textstyle\sum_{S_1} \sum_{A_i} \dfrac{y_i q(e_i, a)}{||yM(a)||} ||xM(a) - yM(a)|| \leqq 2|u|_L \sum_A ||xM(a) - yM(a)|| \\
&\leqq 2|u|_L \textstyle\sum_A \sum_i |x_i - y_i| \sum_j (M(a))_{i,j} = 2|u|_L \sum_i |x_i - y_i| \sum_j (P)_{i,j} \\
&= 2|u|_L \, ||x - y|| \, .
\end{aligned}
$$

Thus

$$
I_2 \leqq 2|u|_L ||x - y|| \tag{4.9}
$$

and by combining (4.7), (4.8) and (4.9) we obtain, for $u \in \mathrm{Lip}\,[K]$,

$$
|Tu(x) - Tu(y)| \leqq (||u|| + 2|u|_L)||x - y|| \, .
$$

The lemma is proved.

5. A coupling device. From Lemma 4.1, Lemma 4.2, Corollary 2.1 and (4.3) we see that what we want to prove is that

$$
\lim_{n \to \infty} \sup_{x, y \in K} |\textstyle\sum_{a^n \in A_x{}^n} u(h_n(x, a^n))q_n(x, a^n) - \sum_{a^n \in A_y{}^n} u(h_n(y, a^n)q_n(y, a^n)| = 0
$$

for all $u \in \text{Lip}[K]$. One difficulty that arises when trying to estimate the quantity

$$|\sum_{a^n \in A_x^n} u(h_n(x, a^n)q_n(x, a^n) - \sum_{a^n \in A_y^n} u(h_n(y, a^n)q_n(y, a^n)|$$

is caused by the fact that in general $A_x^n \neq A_y^n$. The purpose of this section is to show how this difficulty can be overcome.

We start with some notations. For $x \in K$, we define

$$S(x) = \{i \in S : (x)_i > 0\} .$$

Further if S' denotes a subset of S we define

$$K(S') = \{x \in K : S(x) = S'\}$$

and if $0 < \alpha \leqq 1$, we define

$$K(\alpha) = \{x \in K : (x)_i \geqq \alpha \quad \text{if} \quad i \in S(x)\}$$

and

$$K(\alpha, S') = K(\alpha) \cap K(S') .$$

LEMMA 5.1. *There exist constants α_0 and β_0, $0 < \alpha_0, \beta_0 \leqq 1$, an integer n_0 and a set $S_0 \subset S$ such that for all $x \in K$*

$$\Pr[Z_{n_0}(x) \in K(\alpha_0, S_0)] \geqq \beta_0 . \tag{5.1}$$

Since the proof is rather long, we postpone it until Section 9.

REMARK. The lemma is false without Condition A.

Next we introduce an equivalence relation $\sim$ by

$$x \sim y \qquad \text{if} \quad S(x) = S(y) , \qquad x, y \in K ,$$

and define

$$\tilde{K}(\alpha) = \{(x, y) \in K(\alpha) \times K(\alpha) : x \sim y\} .$$

LEMMA 5.2. *If $x \sim y$ then $A_x^n = A_y^n$, for all $n \geqq 1$.*

PROOF. Follows from the definition of A_x^n (see (3.3)).

Now let x and y be two arbitrary elements of K. Let $\{(Z_n'(x), Z_n''(y))\}_{n=0}^\infty$ be a sequence of two d-dimensional random variables such that $\{Z_n'(x)\}_0^\infty$ and $\{Z_n''(y)\}_0^\infty$ are two independent Markov chains both generated by $R(\cdot, \cdot)$ and such that $Z_0'(x) = x$ and $Z_0''(y) = y$. Define

$$N_{x,y} = \min\{n : (Z_n'(x), Z_n''(y)) \in \tilde{K}(\alpha_0)\}$$

where α_0 is the constant of Lemma 5.1.

LEMMA 5.3. *There exist constants C_0 and ρ_0, $0 < \rho_0 < 1$ such that for all $x, y \in K$*

$$\Pr[N_{x,y} > n] \leqq C_0 \rho_0^n , \qquad n = 0, 1, 2, \cdots$$

PROOF. Follows easily from Lemma 5.1.

LEMMA 5.4. *Let $n > m$. Then for any $u \in B[K]$ we have*

$$\begin{aligned} &\sup_{x,y \in K} |u_n(x) - u_n(y)| \\ &\quad \leqq \max_{0 \leq k \leq m} \sup_{(x,y) \in \tilde{K}(\alpha_0)} |u_{n-k}(x) - u_{n-k}(y)| + \varphi(u) C_0 \rho_0^m . \end{aligned} \tag{5.2}$$

PROOF. Let $u \in B[K]$. We have

$$\begin{aligned} |u_n(x) - u_n(y)| &= |Eu(Z_n(x)) - Eu(Z_n(y))| \\ &= |E[u(Z_n'(x)) - u(Z_n''(y))]| \\ &\leqq \textstyle\sum_{k=0}^{m} |E[u(Z_n'(x) - u(Z_n''(y)) : N_{x,y} = k]| \\ &\quad + |E[u(Z_n'(x)) - u(Z_n''(x)) : N_{x,y} > m]| , \end{aligned} \tag{5.3}$$

where $E[\ :B]$ denotes integration over the set B.

From Lemma 5.3 follows that

$$E[u(Z_n'(x)) - u(Z_n''(y)) : N_{x,y} > m] \leqq \varphi(u) C_0 \rho^m . \tag{5.4}$$

Moreover, from the Markov property follows that

$$\begin{aligned} &|E[u(Z_n'(x)) - u(Z_n''(y)) : N_{x,y} = k]| \\ &\quad \leqq \lambda_k \sup_{(x,y)\in \tilde{K}(\alpha_0)} |E[u(Z'_{n-k}(x)) - u(Z''_{n-k}(y))]| \\ &\quad = \lambda_k \sup_{(x,y)\in \tilde{K}(\alpha_0)} |u_{n-k}(x) - u_{n-k}(y)| , \end{aligned} \tag{5.5}$$

where

$$\lambda_k = \Pr [N_{x,y} = k] . \tag{5.6}$$

The inequality (5.2) now follows from (5.3), (5.4), (5.5) and (5.6).

We also have

LEMMA 5.5. *Let* $0 < \alpha \leqq 1$, *and* $(x, y) \in \tilde{K}(\alpha)$. *Then*

$$\begin{aligned} |u_n(x) - u_n(y)| &\leqq (1 - \alpha)\varphi(u) \\ &\quad + |u|_L \textstyle\sum_{A_x^n} ||h_n(x, a^n) - h_n(y, a^n)|| q_n(y, a^n) \end{aligned} \tag{5.7}$$

for all $u \in \mathrm{Lip}\,[K]$.

PROOF. For any $x, y \in K$ we have

$$\textstyle\sum_{A^n} q_n(x, a^n) = \sum_{A^n} q_n(y, a^n) = 1$$

and

$$\textstyle\sum_{A^n} |q_n(x, a^n) - q_n(y, a^n)| \leqq ||x - y|| .$$

Moreover if $(x, y) \in \tilde{K}(\alpha)$ then $||x - y|| \leqq 2 \cdot (1 - \alpha)$ and $A_x^n = A_y^n$. Therefore if $(x, y) \in \tilde{K}(\alpha)$ and $u \in \mathrm{Lip}\,[K]$ we have

$$\begin{aligned} |u_n(x) - u_n(y)| &\leqq |\textstyle\sum_{A_y^n} u(h_n(x, a^n))(q_n(x, a^n) - q_n(y, a^n))| \\ &\quad + |\textstyle\sum_{A_y^n} (u(h_n(x, a^n) - u(h_n(y, a^n))q_n(y, a^n)| \\ &\leqq \varphi(u)(1 - \alpha) + |u|_L \textstyle\sum_{A_x^n} ||h_n(x, a^n) - h_n(y, a^n)|| q_n(y, a^n) , \end{aligned}$$

which was to be proved.

6. Two lemmas on products of matrices.

On the right-hand side of (5.7) the quantity

$$\begin{aligned} &||h_n(x, a^n) - h_n(y, a^n)|| \\ &\qquad \left(= \left\| \frac{xM(a_1) \cdots M(a_n)}{||xM(a_1) \cdots M(a_n)||} - \frac{yM(a_1) \cdots M(a_n)}{||yM(a_1) \cdots M(a_n)||} \right\| \right) \end{aligned}$$

occurs. The purpose of this section is to prove two inequalities for products of nonnegative, nonzero subrectangular matrices. (See Definition 1.1 for the definition of a subrectangular matrix.)

We first need some notations. Let M denote a nonnegative $d \times d$ matrix. We define

$$S_1(M) = \{i : (M)_{i,j} > 0\,, \text{ some } j\}$$
$$S_2(M) = \{j : (M)_{i,j} > 0\,, \text{ some } i\}$$
$$(M)_i = \sum_j (M)_{i,j}$$

and

$$||M|| = \max_i (M)_i\,.$$

The following lemma is a slight generalization of Lemma 3 in [4].

LEMMA 6.1. *Let $M_1, M_2, \cdots, M_n$, $n \geqq 1$, be nonnegative, nonzero, subrectangular matrices such that*

$$\max_{i,j\in S} (M_m)_{i,j} \leqq 1\,, \qquad m = 1, 2, \cdots, n\,. \tag{6.1}$$

Let $U = M_1 M_2 \cdots M_n$ and assume that

$$||U|| > 0\,. \tag{6.2}$$

Let

$$\delta_m = \min_{i,j} \{(M_m)_{i,j} : (M_m)_{i,j} > 0\}\,, \qquad m = 1, 2, \cdots, n\,. \tag{6.3}$$

Then if i_1 and $i_2 \in S_1(U)$ we have

$$\left|\frac{(U)_{i_1,j}}{(U)_{i_1}} - \frac{(U)_{i_2,j}}{(U)_{i_2}}\right| \leqq \prod_{m=1}^{n} (1 - \delta_m{}^2\delta_n) \tag{6.4}$$

for all $j \in S$.

PROOF. The proof is essentially based on the same ideas as used by H. Furstenberg and H. Kesten in their proof of Lemma 3 in [4].

First we observe that since i_1 and $i_2 \in S_1(U)$ we have $(U)_{i_1} > 0$ and $(U)_{i_2} > 0$ and hence the left-hand side of (6.4) is well defined. Next we state

PROPOSITION 6.1. *Let M and N denote two nonnegative $d \times d$ matrices. If M is subrectangular so are MN and NM.*

PROOF. Trivial.

From this proposition follows that U is subrectangular.

PROPOSITION 6.2. *If $j_1, j_2 \in S_2(U)$ and $i_1, i_2 \in S_1(U)$ then*

$$\delta_1\delta_n \leqq \frac{(U)_{i_1,j_2}}{(U)_{i_2,j_2}} \leqq (\delta_1\delta_n)^{-1}\,. \tag{6.5}$$

PROOF. Since $j_1, j_2 \in S_2(U)$ and $i_1, i_2 \in S_1(U)$ we obtain from the subrectangularity of U $(U)_{i_1,j_1} > 0$ and $(U)_{i_2,j_2} > 0$. Next let $n \geqq 3$ and denote $V \doteq M_2 M_3 \cdots M_{n-1}$. Then

$$(U)_{i_1,j_1} = \sum_{r,k} (M_1)_{i_1,r}(V)_{r,k}(M_n)_{k,j}.$$

and

$$(U)_{i_2,j_2} = \sum_{r,k} (M_1)_{i_2,r}(V)_{r,k}(M_n)_{k,j_2}\,.$$

Using the subrectangularity of M_1, V and M_n we obtain from (6.1), (6,2) and (6.3)

$$0 < \delta_1\delta_n \sum_{r,k}(V)_{r,k} \leqq (U)_{i,j} \leqq \sum_{r,k}(V)_{r,k}\,, \qquad \text{if } i = i_1, i_2; j = j_1, j_2$$

from which (6.5) follows.

That (6.5) also holds when $n = 1$ or 2 is evident.

From Proposition 6.2 clearly follows that

$$\delta_1\delta_n(U)_{i_1} \leqq (U)_{i_2} \leqq (\delta_1\delta_n)^{-1}(U)_{i_1}\,, \qquad i_1, i_2 \in S_1(U) \tag{6.6}$$

and by combining (6.6) and (6.5) we see that (6.4) holds if $n = 1$. Next let $n \geqq 2$ and denote $W = M_2 M_3 \cdots M_n$. Then if $i \in S_1(U)$ we have

$$\frac{(U)_{i,j}}{(U)_i} = \sum_{S_1(W)} \frac{(M_1)_{i,k}(W)_k}{(U)_i} \cdot \frac{(W)_{k,j}}{(W)_k}\,.$$

Thus if we denote

$$\alpha_k = \frac{(M_1)_{i_1,k}(W)_k}{(U)_{i_1}}\,, \qquad k = 1, 2, \cdots, d$$

and

$$\beta_k = \frac{(M_1)_{i_2,k}(W)_k}{(U)_{i_2}}\,, \qquad k = 1, 2, \cdots, d$$

we have

$$\frac{(U)_{i_1,j}}{(U)_{i_1}} - \frac{(U)_{i_2,j}}{(U)_{i_2}} = \sum_{S_1(W)} (\alpha_k - \beta_k)\,\frac{W_{k,j}}{(W)_k}\,. \tag{6.7}$$

Now let $k \in S_2(M_1)$. Then applying (6.3) and (6.6) we obtain

$$\frac{(M_1)_{i_2,k}(W)_k}{(U)_{i_2}} \geqq \delta_1^2\,\delta_n\,\frac{(M_1)_{i_1,k}(W)_k}{(U)_{i_1}}\,.$$

Hence if $k \in S_2(M_1)$

$$\beta_k \geqq \delta_1^2\delta_n\,\alpha_k \tag{6.8}$$

and since M_1 is subrectangular the inequality holds trivially if $k \notin S_2(M_1)$. Furthermore using the fact that

$$\sum_{S_1(W)} (M_1)_{i,k}(W)_k = (U)_i$$

we have

$$\sum_{S_1(W)} \alpha_k = \sum_S \alpha_k = 1 = \sum_S \beta_k = \sum_{S_1(W)} \beta_k\,. \tag{6.9}$$

Therefore, defining $S^+ = \{k : \alpha_k \geqq \beta_k\}$ we obtain from (6.7), (6.8) and (6.9)

$$\left|\frac{(U)_{i_1,j}}{(U)_{i_1}} - \frac{(U)_{i_2,j}}{(U)_{i_2}}\right| \leqq \left(\max_{S_1(W)} \frac{(W)_{k,j}}{(W)_k} - \min_{S_1(W)} \frac{(W)_{k,j}}{(W)_k}\right) \sum_{S^+} (\alpha_k - \beta_k)$$

$$\leqq \left(\max_{S_1(W)} \frac{(W)_{k,j}}{(W)_k} - \min_{S_1(W)} \frac{(W)_{k,j}}{(W)_k}\right)(1 - \delta_1^2\delta_n)\,.$$

(6.4) then follows by induction.

LEMMA 6.2. *Let $M_1, M_2, \cdots, M_n$ and U be as in Lemma 6.1. Let x and $y \in K$, and suppose that $||xU|| > 0$ and $||yU|| > 0$. Then*

$$\left\|\frac{xU}{||xU||} - \frac{yU}{||yU||}\right\| \leqq d \cdot (\textstyle\prod_1^n (1 - \delta_m{}^2\delta_n)) .$$

PROOF. From the definition of the norm we have

$$\left\|\frac{xU}{||xU||} - \frac{yU}{||yU||}\right\| = \sum_j \left|\sum_i \frac{x_i(U)_{i,j}}{||xU||} - \sum_i \frac{y_i(U)_{i,j}}{||yU||}\right| .$$

But

$$\sum_i \frac{x_i(U)_{i,j}}{||xU||} = \sum_{S_1(U)} \frac{x_i(U)_{i,j}}{||xU||} = \sum_{S_1(U)} \frac{(U)_{i,j}}{(U)_i} \frac{x_i(U)_i}{||xU||}$$

and similarly

$$\sum_i \frac{y_i(U)_{i,j}}{||yU||} = \sum_{S_1(U)} \frac{(U)_{i,j}}{(U)_i} \cdot \frac{y_i(U)_i}{||yU||} .$$

Therefore

$$\left\|\frac{xU}{||xU||} - \frac{yU}{||yU||}\right\| = \sum_j \left|\sum_{S_1(U)} \frac{(U)_{i,j}}{(U)_i} (\alpha_i - \beta_i)\right|$$

where we have defined

$$\alpha_i = \frac{x_i(U)_i}{||xU||} \quad \text{and} \quad \beta_i = \frac{y_i(U)_i}{||yU||}, \quad i = 1, 2, \cdots, d .$$

But $\sum x_i(U)_i = ||xU||$ and hence $\sum \alpha_i = 1$. Similarly we obtain $\sum \beta_i = 1$. Therefore

$$\left|\sum_{S_1(U)} \frac{(U)_{i,j}}{(U)_i} (\alpha_i - \beta_i)\right|$$
$$\leqq \left(\max_{i \in S_1(U)} \frac{(U)_{i,j}}{(U)_i} - \min_{i \in S_1(U)} \frac{(U)_{i,j}}{(U)_i}\right) \cdot \tfrac{1}{2} \sum_i |\alpha_i - \beta_i| .$$

Then applying Lemma 6.1 we obtain

$$\left\|\frac{xU}{||xU||} - \frac{yU}{||yU||}\right\| \leqq d(\textstyle\prod_{m=1}^n (1 - \delta_m{}^2\delta_n)) ,$$

since $\frac{1}{2} \sum |\alpha_i - \beta_i| \leqq 1$. The lemma is proved.

7. The subrectangular case. In this section we prove the assertion of Theorem A under the extra assumption that the tr p m P is subrectangular. First we observe that

$$P \ \text{subrectangular} \Rightarrow M(a) \ \text{subrectangular} \quad \text{for all} \quad a \in A .$$

Next let $(x, y) \in \tilde{K}(\alpha_0)$ and let $a^n \in A_x{}^n$. Then $||xM(a^n)|| > 0$ and $||yM(a^n)|| > 0$. Therefore, if we denote

$$\delta = \min_{i,j} \{(P)_{i,j} : (P)_{i,j} > 0\}$$

we obtain from Lemma 6.2

$$(7.1) \qquad ||h_n(x, a^n) - h_n(y, a^n)|| \leqq d(1 - \delta^3)^n .$$

Hence applying Lemma 5.5, for $u \in \text{Lip}[K]$, we obtain

$$(7.2) \qquad \sup_{(x,y)\in \tilde{K}(\alpha_0)} |u_n(x) - u_n(y)| \leqq (1 - \alpha_0)\varphi(u) + |u|_L d\rho_1^n$$

where $\rho_1 = (1 - \delta^3)$.

Then combining (7.2) with Lemma 5.4 we obtain, for $n, m \geqq 1$,

$$(7.3) \qquad \varphi(u_{n+m}) \leqq (1 - \alpha_0)\varphi(u) + |u|_L d\rho_1^n + \varphi(u)C_0\rho_0^m .$$

Thus if we choose n and m sufficiently large, say $n = n_1$ and $m = m_1$, then

$$\varphi(u_{n_1+m_1}) \leqq (1 - \alpha_1)\varphi(u)$$

where $\alpha_1 = \alpha_0/2$.

But because of Lemma 4.3 we have that

$$u \in \text{Lip}[K] \Rightarrow u_n \in \text{Lip}[K] , \qquad n = 1, 2, \cdots$$

Therefore we can apply (7.3) to the function $u_{n_1+m_1}$. We then obtain

$$\begin{aligned}\varphi(u_{n_1+m_1+n+m}) &= (\text{by } (4.2)) = \varphi((u_{n_1+m_1})_{n+m}) \\ &\leqq (1 - \alpha_0)\varphi(u_{n_1+m_1}) + |u_{n_1+m_1}|_L d\rho_1^n + \varphi(u_{n_1+m_1}) \cdot C_0\rho_0^m \\ &\leqq (1 - \alpha_0)(1 - \alpha_1)\varphi(u) + |u_{n_1+m_1}|_L d\rho_1^n + \varphi(u_{n_1+m_1})C_0\rho_0^m .\end{aligned}$$

Again choosing n and m sufficiently large (probably much larger this time since $|u_{n_1+m_1}|_L$ might be very large), say $n = n_2$ and $m = m_2$, we obtain

$$\varphi(u_{n_1+n_2+m_1+m_2}) \leqq (1 - \alpha_1)^2\varphi(u) .$$

Repeating this procedure and using the fact that $\{\varphi(u_n)\}$ is a nonincreasing sequence we see that if $u \in \text{Lip}[K]$ then $\lim_{n\to\infty} \varphi(u_n) = 0$, which together with Lemma 4.2, Lemma 4.1, and Corollary 2.1 proves that the assertion of Theorem A is true if we assume that P is a subrectangular matrix.

8. The final step. In order to complete the proof of Theorem A all that remains to do (besides proving Lemma 5.1) is to prove the following lemma.

LEMMA 8.1. *Assume Condition* A *holds. Then, if* $\alpha > 0$,

$$(8.1) \qquad \lim_{n\to\infty} \sup_{\tilde{K}(\alpha)} \textstyle\sum_{A_x^n} ||h_n(x, a^n) - h_n(y, a^n)||q_n(x, a^n) = 0 .$$

For, having proved this lemma, we by the same arguments as used in the previous section can prove

$$u \in \text{Lip}[K] \Rightarrow \lim_{n\to\infty} \varphi(u_n) = 0$$

form which Theorem A follows.

PROOF OF LEMMA 8.1. Since we assume that Condition A holds, there exists an integer m_0 and a sequence $\{b_0, b_1, b_2, \cdots, b_{m_0}\}$ such that the matrix product $M(b_0)M(b_1) \cdots M(b_{m_0})$ is a nonzero subrectagular matrix.

Now let $\{Y_n(x)\}_1^\infty$, $x \in K$ be the stochastic process introduced in the existence theorem of Section 3. Denote

$$(Y_1(x), Y_2(x), \cdots, Y_n(x)) = Y^n(x)$$

and

$$(Y_{n+1}(x), Y_{n+2}(x), \cdots, Y_{n+m}(x)) = {}^nY^m(x) .$$

From Lemma 3.2 it is clear that if $x \sim y$ then

$$\begin{aligned} \sum_{A_x{}^n} \|h_n(x, a^n) - h_n(y, a^n)\| q_n(x, a^n) \\ = E[\|h_n(x, Y^n(x)) - h_n(y, Y^n(x))\|] . \end{aligned} \tag{8.2}$$

Define

$$n_1 = \min \{n : \min_{i,j} (P^n)_{i,j} > 0\} .$$

Such an n_1 exists since $\{X_n\}_1^\infty$ is assumed to be ergodic. Denote

$$M(Y_1(x))\, M(Y_2(x)) \cdots M(Y_n(x)) = M^n(x)$$

and

$$M(Y_{n+1}(x))\, M(Y_{n+2}(x)) \cdots M(Y_{n+m}(x)) = M_n{}^m(x) .$$

LEMMA 8.2. *There exists a constant γ_0 such that for all $x \in K$*

$$\Pr [M^{n_1+m_0}(x) \text{ is subrectangular}] \geqq \gamma_0 . \tag{8.3}$$

PROOF. Because of Proposition 6.1

$$\begin{aligned} \Pr [M^{n_1+m_0}(x) & \text{ is subrectangular}] \\ & \geqq \Pr [M_{n_1}^{m_0}(x) \text{ is subrectangular}] \\ & \geqq \Pr [Y_{n_1+k}(x) = b_k, k = 1, 2, \cdots, m_0] . \end{aligned}$$

Now since the product $M(b_0)\, M(b_1) \cdots M(b_{m_0})$ is nonzero we have $\|e_{i_0} M(b_0) \cdots M(b_{m_0})\| \geqq \gamma_1 > 0$ for some vector e_{i_0} where $\{e_i\}_1{}^d$, as before, denotes the set of base-vectors of R^d. Moreover, using the ergodicity of the matrix P it is not very difficult to prove that there exists $\gamma_2 > 0$ such that for all $x \in K$

$$\Pr [(Z_{n_1}(x))_{i_0} \geqq \gamma_2] \geqq \delta_1$$

where

$$\delta_1 = \min_{i,j} \{(P^{n_1})_{i,j}\} .$$

From (3.1) then follows that

$$\Pr [Y_{n_1+k}(x) = b_k, k = 1, 2, \cdots, m_0] \geqq \delta_1 \gamma_2 \gamma_1$$

and hence by taking $\gamma_0 = \delta_1 \gamma_2 \gamma_1$ we obtain (8.3), and hence Lemma 8.2 is proved.

Next for each $x \in K$ define a sequence of positive, integer-valued stochastic variables $\{N_k(x)\}_{k=1}^\infty$ by

$$N_1(x) = \min \{n : M(Y_1(x))\, M(Y_2)) \cdots M(Y_n(x)) \text{ is subrectangular}\}$$

$$N_{k+1}(x) = \min \{n : M(Y_{k'}(x))\, M(Y_{k'+1}(x)) \cdots M(Y_{k'+n}(x)) \text{ is subrectangular}\} \quad k = 1, 2 \cdots$$

where k' denotes $N_k(x) + 1$.

From Lemma 8.2, Lemma 3.1 and Lemma 3.2 it is not difficult to convince oneself that the following result holds:

LEMMA 8.3. *There exist constants C_2 and ρ_2, $0 < \rho_2 < 1$ such that for all $x \in K$ and for all choices of $m, k_1, k_2, k_m, n_1, n_2, \cdots, n_m$*

$$\Pr[N_{k_1}(x) \geqq n_1, N_{k_2}(x) \geqq n_2, \cdots, N_{k_m}(x) \geqq n_m] \leqq C_2\rho_2^{n_1} C_2\rho_2^{n_2} \cdots C_2\rho_2^{n_m} .$$

We omit the proof.

To simplify notations we from now on write N_k instead of $N_k(x)$. Next define

$$L = L(n) = \max\{k : N_1 + N_2 + \cdots + N_k \leqq n\}, \quad \text{if } N_1 \leqq n,$$
$$= 0 \quad \text{otherwise.}$$

We have

$$\begin{aligned} &E[\|h_n(x, Y^n(x)) - h_n(y, Y^n(x))\|] \\ (8.4) \qquad &\leqq E[\|h_n(x, Y^n(x)) - h_n(y, Y^n(x))\| : L(n) \geqq 2] \\ &+ 2\Pr[L(n) \leqq 1] = I_1(n) + I_2(n) \quad \text{say.} \end{aligned}$$

From Lemma 8.3 follows

$$(8.5) \qquad \lim_{n\to\infty} I_2(n) = 0 .$$

Next denote

$$\mu_k = N_1 + N_2 + \cdots + N_k , \qquad k = 1, 2, \cdots ,$$

and

$$G_1 = \prod_1^{N_1} M(Y_m(x))$$
$$G_{k+1} = \prod_{\mu_k+1}^{\mu_{k+1}} M(Y_m(x)) , \qquad k = 1, 2, \cdots .$$

On $\{L(n) \geqq 2\}$ define

$$G_L' = \prod_{m=\mu_{L-1}+1}^{n} M(Y_m(x)) .$$

By definition G_k, $k = 1, 2, \cdots$ are subrectangular and

$$\min_{i,j}\{(G_k)_{i,j} : (G_k)_{i,j} > 0\} \geqq \delta^{N_k} ,$$

where as before

$$(8.6) \qquad \delta = \min\{(P)_{i,j} : (P)_{i,j} > 0\} .$$

Also, by Proposition 6.1, we note that G_L' is subrectangular and that

$$\min_{i,j}\{(G_L')_{i,j} : (G_L')_{i,j} > 0\} \geqq \delta^{N'}$$

where $N' = n - \mu_{L-1}$. Therefore, by Lemma 6.2, we have

$$(8.7) \qquad I_1(n) \leqq E[(\prod_{m=1}^{L-1}(1 - \delta^{2N_m+N'}))(1 - \delta^{3N'}) : L(n) \geqq 2] .$$

Now, by using Lemma 8.3, it is elementary but somewhat tedious to prove that

$$(8.8) \qquad \lim_{n\to\infty} E[(\prod_1^{L-1}(1 - \delta^{2N_m+N'}))(1 - \delta^{3N'}) : L(n) \geqq 2] = 0 .$$

(8.1) then follows by combining (8.2), (8.5), (8.7) and (8.8), and hence Lemma 8.1 is proved.

9. Proof of Lemma 5.1. In this section we prove Lemma 5.1. First, since we assume that Condition A holds, we know that there exists a sequence

$a_1, a_2, \cdots, a_m$ such that

$$M = M(a_1)\, M(a_2) \cdots M(a_m)$$

is a nonzero subrectangular matrix. In agreement with the notation in Section 6 we let $S_1(M)$ and $S_2(M)$ denote respectively the nonzero rows and nonzero columns of M. Now let $i_0 \in S_1(M)$. Since $\{X_n\}_1^\infty$ is ergodic there exists an integer n_1 and sequences $\{k_0(i), k_1(i), \cdots, k_{n_1}(i)\}$, $i \in S$, such that $k_0(i) = i$, $k_{n_1}(i) = i_0$ and

$$\prod_{n=1}^{n_1} p(k_{n-1}(i), k_n(i)) > 0 ,$$

where $p(i, j)$ denotes the i, jth element of the matrix P. Define

$$\begin{aligned} b_n(i) &= g(k_n(i)) , && 1 \leqq n \leqq n_1 \\ a_n(i) &= b_n(i) , && 1 \leqq n \leqq n_1 \\ &= a_{n-m} , && n_1 + 1 \leqq n \leqq n_1 + m , \end{aligned}$$

put

$$\begin{aligned} M_i &= M(b_1(i))\, M(b_2(i)) \cdots M(b_{n_1}(i)) \\ n_2 &= n_1 + m \\ a^{n_2}(i) &= \{a_1(i), a_2(i), \cdots, a_{n_2}(i)\} , \end{aligned}$$

and let e_i, $i \in S$, denote the ith base-vector of R^d. It is easy to see that

$$S_2(M_i M) = S_2(M)$$

and therefore, for all $i \in S$, we have

$$(h_{n_2}(e_i, a^{n_2}(i)))_j > 0 \qquad \text{if } j \in S_2(M)$$

and

$$(h_{n_2}(e_i, a^{n_2}(i)))_j = 0 \qquad \text{if } j \notin S_2(M) .$$

Defining

$$\begin{aligned} \alpha_i' &= \min \{(h_{n_2}(e_i, a^{n_2}(i)))_j, j \in S_2(M)\} , \\ \alpha' &= \min \{\alpha_i', i \in S\} , \\ \delta &= \min \{(P)_{i,j} : (P)_{i,j} > 0\} , \qquad \text{and} \\ \beta' &= \delta^{n_2} , \end{aligned}$$

we obtain

$$\begin{aligned} \Pr [Z_{n_2}(e_i) \in K(\alpha', S_2(M))] &\geqq \Pr [Z_{n_2}(e_i) = h_{n_2}(e_i, a^{n_2}(i))] \\ &\geqq q_{n_2}(e_i, a^{n_2}(i)) \geqq \delta^{n_2} = \beta' , \end{aligned}$$

and hence (5.1) holds for $x = e_i$, $i \in S$, if we take $n_0 = n_2$, $\alpha_0 = \alpha'$, $\beta_0 = \beta'$ and $S_0 = S_2(M)$. Now let x be an arbitrary element of K. Since we can write $x = \sum x_i e_i$ and since $\sum x_i = 1$, at least one of the terms x_i, $i = 1, 2, \cdots, d$, is larger than d^{-1}. We may of course assume that $x_1 \geqq d^{-1}$. Then if $j \in S_2(M)$, we have

$$(h_{n_2}(x, a^{n_2}(1)))_j = \frac{(xM_1M)_j}{||xM_1M||} \geqq \frac{x_1(e_1M_1M)_j}{||xM_1M||} \geqq d^{-1}\alpha'\beta'$$

and

$$q_{n_2}(x, a^{n_2}(1)) \geqq x_1 q_{n_2}(e_1, a^{n_2}(1)) \geqq d^{-1}\beta' .$$

But if $j \notin S_2(M)$ we have

$$(h_{n_2}(x, a^{n_2}(i)))_j = 0$$

for all $i \in S$, since $S_2(M_i M) = S_2(M)$.

Hence if we take $n_0 = n_2$, $\alpha_0 = d^{-1}\alpha'\beta'$, $\beta_0 = d^{-1}\beta'$ and $S_0 = S_2(M)$ we can conclude that (5.1) holds for all $x \in K$. The lemma is proved.

REMARK. I want to thank the referee for helping me to find a correct proof of this lemma.

10. A counter example. As we mentioned in the introduction the conclusion of Theorem A does not hold if we merely assume that the process $\{X_n\}_1^\infty$ is ergodic. We shall show this by an example.

Let the state space consist of four elements $S = \{1, 2, 3, 4\}$ and let the function g be such that

$$g(1) = g(2) = a, g(3) = g(4) = b\,, \qquad a \neq b\,. \tag{10.1}$$

Let the transition probability matrix P be given by

$$P = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}. \tag{10.2}$$

It is easy to prove that P is ergodic.

Now let $p = (p_1, p_2, p_3, p_4)$ be the initial distribution for the Markov chain $\{X_n\}_1^\infty$ generated by P and let us for simplicity assume that $p_3 = p_4 = 0$.

Denote

$$\alpha_1 = (p_1, p_2, 0, 0)\,, \qquad \alpha_2 = (p_2, p_1, 0, 0)$$
$$\alpha_3 = (0, 0, p_1, p_2)\,, \qquad \alpha_4 = (0, 0, p_2, p_1)\,.$$

We shall show below that

$$\lim_{n\to\infty} \Pr[Z_n = \alpha_i] = \tfrac{1}{4}\,, \qquad i = 1, 2, 3, 4\,. \tag{10.3}$$

From (10.3) then follows that if P and g are given by (10.2) and (10.1), the distribution function of Z_n converges to a limit, but this limit *does* depend on the initial distribution p. To prove (10.3) let us first observe that

$$q(x, a) = ||xM(a)|| = \tfrac{1}{2} = ||q(x, b)|| = ||xM(b)||\,, \qquad x \in K$$

and that

$$h(x, a) = (x_1 + x_3, x_2 + x_4, 0, 0)\,, \qquad x \in K$$
$$h(x, b) = (0, 0, x_1 + x_4, x_2 + x_3)\,, \qquad x \in K\,.$$

Therefore

$$q(\alpha_i, a) = q(\alpha_i, b) = \tfrac{1}{2}\,, \qquad i = 1, 2, 3, 4 \tag{10.4}$$

and

$$h(\alpha_1, a) = \alpha_1, \quad h(\alpha_2, a) = \alpha_2$$
$$(10.5) \quad h(\alpha_3, a) = \alpha_2, \quad h(\alpha_4, a) = \alpha_1$$
$$h(\alpha_1, b) = \alpha_3, \quad h(\alpha_2, b) = \alpha_4$$
$$h(\alpha_3, b) = \alpha_4, \quad h(\alpha_3, b) = \alpha_3 .$$

From (10.5) we observe that in this very special case the state space of the Markov chain $\{Z_n\}_1^\infty$ only consists of the four points α_1, α_2, α_3 and α_4. Now let $B = (b_{i,j})$ be the tr p m associated to the process. By definition $b_{i,j} = \Pr[Z_{n+1} = \alpha_j \mid Z_n = \alpha_i]$ and from (10.4) and (10.5) we obtain

$$\Pr[Z_{n+1} = \alpha_i \mid Z_n = \alpha_1] = \tfrac{1}{2}, \qquad i = 1, 3$$
$$\Pr[Z_{n+1} = \alpha_i \mid Z_n = \alpha_2] = \tfrac{1}{2}, \qquad i = 2, 4$$
$$\Pr[Z_{n+1} = \alpha_i \mid Z_n = \alpha_3] = \tfrac{1}{2}, \qquad i = 2, 4$$
$$\Pr[Z_{n+1} = \alpha_i \mid Z_n = \alpha_4] = \tfrac{1}{2}, \qquad i = 1, 3 .$$

Hence the tr p m B is also given by (10.2) and since this matrix is ergodic and double-stochastic, (10.3) follows.

REMARK 1. The above example is the same as the one used by Blackwell in [2] to show that the entropy of the $\{X_n\}$-process need not to be larger than the entropy of the $\{Y_n\}$-process.

REMARK 2. Suppose that we in (10.2) take $p_{11} = \frac{1}{2} - \varepsilon$, $0 < \varepsilon < \frac{1}{2}$ instead of $\frac{1}{2}$ and take $p_{12} = \varepsilon$ instead of 0. Then the product $M(a)M(a)M(b)M(a)M(a)M(b)M(b)$ is a nonzero subrectangular matrix; hence Condition A is satisfied and Theorem A applies.

REMARK 3. Suppose that we in (10.2) again take $p_{11} = \frac{1}{2} - \varepsilon$ $(0 < \varepsilon < \frac{1}{2})$, but now take $p_{13} = \frac{1}{2} + \varepsilon$ instead of $\frac{1}{2}$. It is then not very hard to convince oneself that Condition A does *not* hold and hence Theorem A cannot be applied. However it is my belief that the conclusion of Theorem A still holds.

Acknowledgment. I am very grateful to Dr. Urban Hjorth for drawing my attention to the problem studied in this paper. I also want to thank Dr. Mats Rudemo for the reference [2] and finally, and in particular, I want to thank Professor Harry Kesten for many valuable suggestions and comments, and for his kind and stimulating interest in my work.

REFERENCES

[1] ÅSTRÖM, K. J. (1965). Optimal control of Markov processes with inclompete state information. *J. Math. Anal. Appl.* **10** 174–205.

[2] BLACKWELL, D. (1957). The entropy of functions of finite-state Markov chains. *Trans. First Prague Conference Information Theory, Statistical Decision Functions, Random Processes*, Prague, 13–20.

[3] FELLER, W. (1966). *An Introduction to Probability Theory and Its Applications* **2**. Wiley, New York.

[4] FURSTENBERG, H. and KESTEN, H. (1960). Products of random matrices. *Ann. Math. Statist.* **31** 457–469.

[5] IOSIFESCU, M. and THEODERESCU, R. (1969). *Random Processes and Learning.* Springer-Verlag, Berlin.

[6] KAIJSER, T. (1972). Some limit theorems for Markov chains with applications to learning models and products of random matrices. *Report Institute Mittag-Leffler*, Djursholm, Sweden.

[7] NEVEU, J. (1965). *Mathematical Foundations of the Calculus of Probability.* Holden-Day, San Francisco.

[8] RUDEMO, M. (1973). State estimation for partially observed Markov chains. *J. Math. Anal. Appl.* **44** 581–611.

DEPARTMENT OF MATHEMATICS
LINKÖPING UNIVERSITY
S-581 83 LINKÖPING, SWEDEN