RANDOM PAIRWISE AVERAGING
AND PRODUCTS OF RANDOM MATRICES

David Tanny

York University

and

Jon Wellner

University of Washington

December 1981

Revised May 1983

AMS 1980 subject classifications: Primary 60F15; Secondary 60G42, 60J05.

Key words and phrases: Almost sure convergence, branching processes, doubly stochastic matrices, finite Markov chains, majorization, mean vector, random environments, T-transforms.

ABSTRACT

Let $\{\underline{x}^{(n)}\}_{n=0}^{\infty}$ be the sequence of N-vectors obtained from a given vector $\underline{x} = \underline{x}^{(0)} \in R^N$ by averaging two 'randomly' chosen coordinates of $\underline{x}^{(n)}$ and replacing these coordinates by their average to get $\underline{x}^{(n+1)}$. This averaging process is formulated in terms of products of random doubly stochastic matrices, and, using martingale methods, it is shown that $\underline{x}^{(n)}$ converges to $\bar{\underline{x}}$ w.p. 1, where $\bar{\underline{x}}$ is the N-vector with all coordinates equal to $\bar{x}$, the average of the coordinates of $\underline{x} = \underline{x}^{(0)}$. An assertion of Bretagnolle, which arose in an investigation of the Kolmogorov statistic, follows as an easy corollary. Products of random doubly stochastic matrices are then studied more generally, and a condition is given which is both necessary and sufficient for convergence to the matrix with $1/N$ in all entries when the terms in the product form a stationary and ergodic sequence. These results are related to and illustrated by finite Markov chains in 'random environments', and a necessary and sufficient condition for a circular random walk to be accessible is provided.

0. Introduction. Let $\underline{x} = (x_1,\ldots,x_N) \in R^N$ be an arbitrary vector of real numbers, and consider the following averaging process: choose 'at random' two coordinates of $\underline{x}$, say x_i and x_j; form their average $\frac{1}{2}(x_i + x_j)$; replace both x_i and x_j by this average to get a new vector

$$\underline{x}^{(1)} = (x_1,\ldots,x_{i-1},\tfrac{1}{2}(x_i+x_j),x_{i+1},\ldots,x_{j-1},\tfrac{1}{2}(x_i+x_j),x_{j+1},\ldots,x_N) ;$$

and repeat this process. It is intuitively clear that the resulting sequence $\{\underline{x}^{(n)}\}_{n\geq 1}$ converges to $\underline{\bar{x}} = (\bar{x},\ldots,\bar{x})$ with all coordinates equal to $\bar{x} = N^{-1}(x_1+\ldots+x_N)$, the average of all the coordinates of the original vector $\underline{x}$. Our first aim here is to give a simple proof of this convergence; our second aim is to generalize this simple averaging model and show how these generalizations are related to finite Markov chains in 'random environments'.

In our first section we formulate the above simple averaging process in terms of products of random doubly stochastic matrices, relate it to some classical deterministic results conveniently summarized by Marshall and Olkin (1979) and an assertion of Bretagnolle (1980), and give a simple proof of convergence for this special case using martingale methods. In Section 2 we study products of random doubly stochastic matrices more generally, and give a condition which is both necessary and sufficient for convergence when the terms in the product form a stationary ergodic process.

The motivation for this theorem is the desire to understand the limiting behavior of a "Markov chain in a random environment (MCRE)".

In Section 3, this theorem is illustrated by a random walk in a random environment and branching processes in random environments, both of which are MCRE's. An important application of the random walk example is the construction of pseudo-random number generators with desirable properties, which is related to the work of Brown and Solomon (1976).

Since the set of doubly stochastic $N \times N$ matrices form a semi-group, our results can be viewed as strengthening (to almost sure convergence and to dependent sequences) much earlier results of Rosenblatt (1960) concerning convolutions of measures on compact topological semigroups in a special case. Moore (1981) has also obtained an "almost sure" extension of Rosenblatt's result; however, while Moore's result applies to a wider class of stochastic matrices that just the doubly stochastic matrices, his "almost sure" convergence is limited to a weaker Cesaro type averaging. Cogburn (1982) generalized Moore's results to MCRE's having either a finite or a σ-finite invariant measure. Diaconis and Shashahani (1981), proceeding in a different direction, have essentially shown that the limiting distribution of the product of random transpositions is the uniform measure on the permutation matrices (they also provide a rate for this convergence), where their limit is with respect to a variation norm.

There are clearly a variety of closely related questions which we will not address here, including: (i) other (deterministic) pairwise averaging schemes (e.g., at each stage average the largest and smallest coordinates of $\underline{x}$); (ii) convergence rates for such schemes; and (iii) central limit type theorems for random pairwise averaging and other products of random doubly stochastic matrices. Some of these questions have been examined since our first version of this paper by Proschen and Shaked (1982).

1. Random pairwise averaging. Let $\mathcal{P} = \{P_{ij}\}$ denote the collection of NxN permutation matrices, and let $\mathcal{T} = \{T_{ij}\}$ denote the collection of associated 'averaging matrices' defined by $T_{ij} = \frac{1}{2}(I + P_{ij})$. Thus $P_{ij}\underline{x} = (x_1,\ldots,x_{i-1},x_j,x_{i+1},\ldots,x_{j-1},x_i,x_{j+1},\ldots,x_N)'$ and $T_{ij}\underline{x} = (x_1,\ldots,x_{i-1},\frac{1}{2}(x_i+x_j),x_{i+1},\ldots,x_{j-1},\frac{1}{2}(x_i+x_j),x_{j+1},\ldots,x_N)'$ for any vector $\underline{x} \in R^N$. Note that both $\mathcal{P}$ and $\mathcal{T}$ contain $\binom{N}{2}$ distinct elements. Let M be a random matrix uniformly distributed over the $\binom{N}{2}$ T_{ij}'s in $\mathcal{T}$: $P(M = T_{ij}) = 1/\binom{N}{2}$ for all $T_{ij} \in \mathcal{T}$. Clearly $M\underline{x}$ represents the result of 'randomly' choosing two coordinates of $\underline{x}$, averaging them, and replacing both of the chosen coordinates by their average. Every T_{ij} is doubly stochastic, and hence M is doubly stochastic with probability one.

Letting $M_1, M_2, \ldots$ be random matrices independent and identically distributed as M, the averaging process described in the introduction is represented by the vector $M_n \cdots M_1\underline{x}$ after n steps of random pairwise averaging.

THEOREM 1. If $M_1, M_2, \ldots$ are iid as M (uniformly distributed over $\mathcal{T}$) then, for any $\underline{x} \in R^N$,

$$Y_n\underline{x} \equiv M_n \cdots M_1\underline{x} \to \underline{\bar{x}} \quad \text{a.s.} \qquad \text{as } n \to \infty$$

where $\underline{\bar{x}} = (\bar{x},\ldots,\bar{x})$, $\bar{x} = N^{-1}(x_1+\ldots+x_N)$. Equivalently

$$Y_n \equiv M_n \cdots M_1 \to \frac{1}{N} J \quad \text{a.s.} \qquad \text{as } n \to \infty$$

where J is the NxN matrix of ones.

PROOF. Let $\underline{x} \in R^N$. Since $\underline{y} = T_{ij}\underline{x}$ has $\bar{y} = N^{-1} \Sigma_{i=1}^N y_i = N^{-1} \Sigma_{i=1}^N x_i = \bar{x}$, and $T_{ij}\bar{\underline{x}} = \bar{\underline{x}}$, we may suppose that $\bar{x} = 0$ without loss of generality.

Now let $\underline{y} = \underline{x}^{(m)} \equiv M_m \ldots M_1 \underline{x}$ denote the vector obtained after m steps, and let $\underline{z} \equiv \underline{x}^{(m+1)} = M_{m+1} M_m \ldots M_1 \underline{x} = M_{m+1} \underline{x}^{(m)}$. Thus

(a) $z_j = y_j$ for $j \neq r$, $j \neq s$

$z_r = z_s = \frac{1}{2}(y_r + y_s)$

for some random r, s. Clearly $\bar{z} = \bar{y} = 0$. Further, from (a),

$$\text{(b)} \quad \sum_{j=1}^N z_j^2 - \sum_{j=1}^N y_j^2 = 2\left(\frac{y_r + y_s}{2}\right)^2 - y_r^2 - y_s^2$$

$$= -\tfrac{1}{2}(y_r - y_s)^2 \leq 0 .$$

Thus, letting $V_m \equiv \Sigma_{i=1}^N (x_i^{(m)})^2$, it follows from (b) that

$$\text{(c)} \quad V_{m+1} - V_m = -\tfrac{1}{2}(y_r - y_s)^2 \leq 0$$

so that the sequence $\{V_m : m \geq 1\}$ is almost surely decreasing and $V_\infty \equiv \lim_{m\to\infty} V_m$ exists with probability one. But from (c) we have

$$E(V_{m+1} - V_m \mid \underline{y}) = -\tfrac{1}{2} E\{(y_r - y_s)^2 \mid \underline{y}\} = -\frac{V_m}{N(N-1)} ,$$

which implies, letting $\varepsilon \equiv \frac{1}{N(N-1)}$,

$$E(V_{m+1}) = (1 - \varepsilon)V_m$$

and hence that

$$E(V_m) = (1 - \varepsilon)^m E(V_0) = (1 - \varepsilon)^m (\sum_{i=1}^N x_i^2) \to 0 \quad \text{as} \quad m \to \infty .$$

Therefore, by Fatou's lemma,

$$\text{(d)} \quad E(V_\infty) = E(\lim_m V_m) \le \liminf_{m \to \infty} E(V_m) = 0$$

which implies that $V_\infty = 0$ a.s. which in turn yields the conclusion of the theorem. □

Now we want to relate Theorem 1 to some material in Marshall and Olkin (1979) and an assertion of Bretagnolle (1980). As in Marshall and Olkin (1979), for any two vectors $\underline{x}, \underline{y} \in R^N$, $\underline{x} \prec \underline{y}$ if $\sum_{i=1}^k x_{[i]} \le \sum_{i=1}^k y_{[i]}$ $k = 1,\ldots,N-1$, and $\sum_{i=1}^n x_{[i]} = \sum_{i=1}^n y_{[i]}$ where $x_{[1]} \ge \ldots \ge x_{[N]}$ denote the coordinates of $\underline{x}$ in decreasing order (and we say that $\underline{y}$ majorizes $\underline{x}$). For any vector $\underline{x} \in R^N$, it is clear that $\overline{\underline{x}} \prec \underline{x}$. Let $T = \lambda I + (1-\lambda)P$ where $0 \le \lambda \le 1$, I is the NxN identity matrix, and $P \in \mathcal{P}$ is a permutation matrix be a 'T-transform'. A classical result (Marshall and Olkin (1979) page 21) says that if $\underline{x} \prec \underline{y}$ then there exists a <u>finite</u> number of T-transforms $T_1,\ldots,T_k$ such that $T_k \cdots T_1 \underline{y} = \underline{x}$. This implies, in particular, that for any vector $\underline{x} \in R^N$, $T_k \cdots T_1 \underline{x} = \overline{\underline{x}}$ for some finite number of T-transforms (depending on $\underline{x}$). Note that each T_i may involve a <u>different</u> number $\lambda = \lambda_i$ in this result.

In the (random) averaging scheme treated in Theorem 1, λ is forced to be 1/2 always. Consideration of the $N = 3$ deterministic case quickly shows that a finite number of such 'T-1/2' transforms will not yield $\overline{\underline{x}}$ in general (when $N = 2^r$ for some integer $r \ge 1$ a finite number will work). Thus the deterministic question becomes: Does there always exist a sequence of 'T-1/2 transforms' $T_1, T_2, \ldots$. such that $T_k \cdots T_1 \underline{x} \to \overline{\underline{x}}$ as $k \to \infty$?

Theorem 1 gives a probabilistic proof of the existence of many such sequences: take any ω in the set $A = \{\omega: Y_n(\omega)\underline{x} \to \bar{\underline{x}} \text{ as } n \to \infty\}$ which has $P(A) = 1$. Using this fact, we will now give a simple proof of an assertion of Bretagnolle (1980) which arose in the course of an investigation of the Kolmogorov statistic in the case of independent, non-identically distributed random variables.

<u>COROLLARY 1</u>. Let $\underline{F} = (F_1,\ldots,F_N)$ be n arbitrary distribution functions, with average of $\bar{F}$ given by $\bar{F}(x) = \frac{1}{N}\sum_{i=1}^N F_i(x)$, $x \in R$. Then there exists a sequence of $T - \frac{1}{2}$ transforms $\{T_m\}_{m=1}^{\infty}$ such that

$$\lim_{m\to\infty} T_m \ldots T_1 \underline{F}(x) = \bar{\underline{F}}(x) \quad \text{uniformly in } x \in R ;$$

i.e.

$$\|T_m \ldots T_1 \underline{F} - \bar{\underline{F}}\|_\infty \to 0 \quad \text{as} \quad m \to \infty .$$

<u>PROOF</u>. Since $T_m \ldots T_1 \equiv S_m \to \frac{1}{N}J$ as $m \to \infty$,

$$\begin{aligned}
\|S_m\underline{F} - \bar{\underline{F}}\|_\infty &= \max_{1\le i\le N} \sup \left|\sum_{j=1}^N (S_m(i,j) - \tfrac{1}{N}) F_j(x)\right| \\
&\le \max_{1\le i\le N} \sum_{j=1}^N \left|S_m(i,j) - \tfrac{1}{N}\right| \max_{1\le j\le N} \sup_x F_j(x) \\
&\le \max_{1\le i\le N} \sum_{j=1}^N \left|S_m(i,j) - \tfrac{1}{N}\right| \\
&\to 0 \quad \text{as} \quad m \to \infty . \qquad \square
\end{aligned}$$

2. Products of random doubly stochastic matrices. Let Q denote the transition matrix of a discrete time finite-state Markov chain. It is well-known (e.g. Karlin and Taylor (1975)) that if Q is aperiodic and irreducible, the $\lim_{n\to\infty} Q^n = \Pi$ where the matrix Π has identical rows, each row being the stationary distribution of the Markov chain. If Q is doubly stochastic, then it is easily seen that each element of Π equals $1/N$ where N is the number of states. Consider a "Markov chain in a random environment" (MCRE) described as follows: Let M^+ denote the space of nonnegative $N\times N$ stochastic matrices and let (Ω,A,P) be an appropriately large probability space. Suppose that $\{X_n\}_{n=1}^{\infty}$ is a stationary and ergodic process with $X_i\colon \Omega \to M^+$ for $i = 1,2,\ldots$. Then a process $\{H_n\}$ is called a MCRE with environment $\{X_n\}_{n=1}^{\infty}$ if, conditional on the environmental sequence $\{X_n\}_{n=1}^{\infty}$, $\{H_n\}_{n=1}^{\infty}$ is a time inhomogeneous Markov chain with transition matrix X_n at time n. It is then of interest to know the behaviour of $Y_n = X_1\cdot X_2\cdots X_n$, the n-step transition probability matrix. If the matrices $\{X_n\}_{n=1}$ are independent and identically distributed, rather than just stationary and ergodic, and if $E(X_1) \equiv Q$ is an irreducible aperiodic matrix, then it is clear that $\lim_{n\to\infty} E(Y_n) = \lim_{n\to\infty} Q^n = \Pi$ where the rows of Π are identical and each row is the unique stationary distribution for the stochastic matrix $E(X_1) = Q$. If the matrices $\{X_n\}_{n=1}$ are doubly stochastic (as well as iid), then $E(X_1)$

$= Q$ is doubly stochastic, from which it follows that each element of Π is $1/N$. Note that if one only assumes that $\{X_i\}_{i=1}^{\infty}$ are stationary and ergodic rather than independent and identically distributed, even these rather limited results are no longer obvious.

Our main result concerning the behaviour of the products $\{Y_n\}$ is the following theorem:

<u>THEOREM 2</u>. Let $\{X_n\}_{n=1}$ be a stationary ergodic sequence of random variables taking values in the space M^+ of $N \times N$ nonnegative doubly stochastic matrices. Let $Y_n = X_1 \cdot X_2 \cdots X_n$ for $n = 1,2,\ldots$. Suppose there exists a positive integer k such that

(A) $$P(\min_{1 \le i,j \le N} (Y_k)_{ij} > 0) > 0.$$

Then

(1) $$\lim_{n\to\infty} Y_n = \frac{1}{N} J \quad \text{a.s.}$$

and

(2) $$\lim_{n\to\infty} E(Y_n) = \frac{1}{N} J$$

where J is the $N \times N$ matrix of ones.

<u>PROOF of Theorem 1</u>. Let $V_n = X_n \cdot X_{n+1} \cdots X_{n+k-1}$ for $n = 1,2,\ldots$. By Theorem 6.6 of Breiman (1968), $\{V_n\}_{n=1}^{\infty}$ is a stationary ergodic process. Let $H = \{B \in M^+ : \min_{1 \le i,j \le N} B_{ij} > 0\}$. Then, by the Birkhoff Ergodic Theorem (e.g. Breiman (1968)),

(3) $$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} 1_H(V_j) = P(V_1 \in H) \quad \text{a.s.}$$

where $1_H(\cdot)$ is the indicator of the set H. Since

(4) $$P(V_1 \in H) = P(\min_{1 \le i,j \le N} (Y_k)_{ij} > 0) > 0$$

by (A), (3) implies that there exists w.p. 1 a positive integer $n_0 = n_0(\omega) > 1$ such that $V_{n_0} \in H$. Now

(5) $$\min_{1 \le i,j \le N}(Y_{n_0+k-1})_{ij} = \min_{1 \le i,j \le N} \sum_{s=1}^{N}(Y_{n_0-1})_{is}(V_{n_0})_{sj}$$

$$\ge \min_{1 \le i,j \le N}(V_{n_0})_{ij} \min_{1 \le i \le N} \sum_{s=1}^{N}(Y_{n_0-1})_{is}$$

But Y_{n_0-1} is stochastic since it is the product of stochastic matrices, and since $V_{n_0} \in H$, it follows from (5) that

(6) $$\min_{1 \le i,j \le N}(Y_{n_0+k-1})_{ij} \ge \min_{1 \le i,j \le N}(V_{n_0})_{ij} > 0.$$

Thus, for $n > n_0+k-1$,

(7) $$\min_{1 \le i,j \le N}(Y_n)_{ij} = \min_{1 \le i,j \le N} \sum_{s=1}^{N}(Y_{n_0+k-1})_{is}(X_{n_0+k}\cdots X_n)_{sj}$$

$$\ge \min_{1 \le i,j \le N}(Y_{n_0+k-1})_{ij} \sum_{s=1}^{N}(X_{n_0+k}\cdots X_n)_{sj}$$

and since $\{X_n\}_{n=1}^{\infty}$ are doubly stochastic, $X_{n_0+k}\cdots X_n$ is also doubly stochastic, so that (6) and (7) yield $Y_n \in H$ for $n > n_0+k-1$.

Now let $\mathbb{X}$ denote the process $\{X_n\}_{n=1}^{\infty}$ and let

(8) $$m_n(\mathbb{X}) = \min_{1 \le i,j \le N}(X_1\cdots X_n)_{ij}$$

and

(9) $$M_n(\mathbb{X}) = \max_{1 \le i,j \le N}(X_1\cdots X_n)_{ij} .$$

Then it is easily seen that $m_n(\mathbb{X})$ is an increasing sequence since

(10) $$m_{n+1}(\mathbb{X}) = \min_{1 \le i,j \le N} \sum_{s=1}^{N}(X_1\cdots X_n)_{is}(X_{n+1})_{sj}$$

$$\ge m_n(\mathbb{X}) \min_{1 \le j \le N} \sum_{s=1}^{N}(X_{n+1})_{sj} = m_n(\mathbb{X})$$

because X_{n+1} is doubly stochastic. A similar argument shows that $M_n(\mathbb{X})$ is a decreasing sequence. Since $m_n(\mathbb{X})$ and $M_n(\mathbb{X})$ are bounded below by zero and above by 1, the sequences converge. Let $m(\mathbb{X}) \equiv \lim_{n\to\infty} m_n(\mathbb{X})$ and $M(\mathbb{X}) \equiv \lim_{n\to\infty} M_n(\mathbb{X})$. Denote by T the shift operator on the process $\mathbb{X}$; i.e. $T\mathbb{X} = (X_2, X_3, \ldots)$. Then, for $n > 1$,

$$(11) \qquad m_n(\mathbb{X}) = \min_{1\le i,j\le N} \textstyle\sum_{s=1}^{N} (X_1)_{is}(X_2\cdots X_n)_{sj}$$

$$\ge m_{n-1}(T\mathbb{X}) \min_{1\le i\le N} \textstyle\sum_{s=1}^{N} (X_1)_{is} = m_{n-1}(T\mathbb{X}).$$

Taking limits in (11) as $n \to \infty$ yields

$$(12) \qquad m(\mathbb{X}) \ge m(T\mathbb{X}),$$

and a similar argument shows that

$$(13) \qquad M(\mathbb{X}) \le M(T\mathbb{X}).$$

Equations (12) and (13) imply that $m(\mathbb{X})$ and $M(\mathbb{X})$ are invariant random variables under the shift T, and since the process $\mathbb{X}$ is ergodic, Proposition 6.18 of Breiman (1968) implies that $m(\mathbb{X})$ and $M(\mathbb{X})$ are constants almost surely.

For any positive integer r, we can write

$$(14) \qquad (X_{r+1}\cdot X_{r+2}\cdots X_n)_{ij} = m_{n-r}(T^r\mathbb{X}) + \varepsilon_{ij}$$

where $\varepsilon_{ij} = \varepsilon_{ij}(T^r\mathbb{X}, n)$ is a random variable bounded between 0 and 1 for all $1 \le i,j \le N$. Then

$$(15) \qquad m_n(\mathbb{X}) = \min_{1\le i,j\le N} \textstyle\sum_{s=1}^{N} (X_1\cdots X_r)_{is}(X_{r+1}\cdots X_n)_{sj}$$

$$= \min_{1\le i,j\le N} \textstyle\sum_{s=1}^{N} (X_1\cdots X_r)_{is}(m_{n-r}(T^r\mathbb{X}) + \varepsilon_{sj})$$

$$= m_{n-r}(T^r \mathbb{X}) + \min_{1 \le i,j \le N} \sum_{s=1}^{N} (X_1 \cdots X_r)_{is} \varepsilon_{sj}$$

$$\ge m_{n-r}(T^r \mathbb{X}) + m_r(\mathbb{X}) \min_{1 \le j \le N} \sum_{s=1}^{N} \varepsilon_{sj} .$$

Since $(X_{r+1} \cdots X_n)$ is doubly stochastic, it follows that, for any $1 \le j \le N$,

(16) $$N m_{n-r}(T^r \mathbb{X}) + \sum_{s=1}^{N} \varepsilon_{sj} = 1,$$

and so, upon combining (15) and (16),

(17) $$m_n(\mathbb{X}) \ge m_{n-r}(T^r \mathbb{X}) + m_r(\mathbb{X}) \{1 - N m_{n-r}(T^r \mathbb{X})\}.$$

Also,

(18) $$M_{n-r}(T^r \mathbb{X}) + (N-1) m_{n-r}(T^r \mathbb{X}) \le 1,$$

so combining equations (17) and (18) yields

(19) $$m_n(\mathbb{X}) \ge m_{n-r}(T^r \mathbb{X}) + m_r(\mathbb{X}) \{ M_{n-r}(T^r \mathbb{X}) - m_{n-r}(T^r \mathbb{X})\}.$$

Taking limits in equation (19) as n tends to infinity (with r fixed) yields

(20) $$m(\mathbb{X}) \ge m(T^r \mathbb{X}) + m_r(\mathbb{X}) \{ M(T^r \mathbb{X}) - m(T^r \mathbb{X})\}$$

and since $m(\mathbb{X})$ and $M(\mathbb{X})$ are constants almost surely, equation (20) implies that

(21) $$m_r(\mathbb{X}) \{ M(\mathbb{X}) - m(\mathbb{X})\} \le 0 \quad \text{a.s.}.$$

But for r sufficiently large $m_r(\mathbb{X}) > 0$ a.s., and $M(\mathbb{X}) \ge m(\mathbb{X})$ and so it follows from (21) that $M(\mathbb{X}) = m(\mathbb{X})$ a.s. .

Since $(X_1 \cdots X_n)$ is doubly stochastic

(22) $$N\, M_n(\mathbb{X}) \geq 1 \quad \text{and} \quad N\, m_n(\mathbb{X}) \leq 1;$$

but this together with $m(\mathbb{X}) = M(\mathbb{X})$ forces $m(\mathbb{X}) = M(\mathbb{X}) = 1/N$ a.s., and this completes the proof of (1) of the theorem.

The second half of Theorem 2, (2), follows immediately from (1) and the Lebesgue dominated convergence theorem since all the entries of the matrices are bounded by 0 and 1. □

REMARK 1. Condition (A) is clearly a necessary condition for equation (1) to hold, so Theorem 1 is sharp. Some simple sufficient conditions for (A) will be given at the end of Section 3.

REMARK 2. If one regards X_n as the transition matrices of a Markov chain, then, conditional on $\{X_n\}_{n=1}^{\infty}$ one can construct a time-inhomogeneous Markov chain. It is well known (Orey (1971)) that the associated space-time process is time homogeneous; i.e. it has a fixed transition probability matrix $P = P(\mathbb{X})$. It is possible, therefore, that using standard results for the limit of the matrix P^n, one could "transfer" these results from the space-time process to the original process and thus obtain a different proof of Theorem 1.

REMARK 3. It should be noted that the limiting behavior of products has been studied extensively (see Senata (1973) for a nice account) and, that conditional on the given sequence of matrices, the random product becomes a deterministic one. Thus one is tempted to try to "transfer" the deterministic result to the ramdom product case. When $\{X_n\}_{n=1}^{\infty}$ are independent and identically distributed, Theorem 2 could in fact, be proved using classical results in Senata (1973) or related extensions in Kaijser (1975). Even the stationary ergodic case can probably be proved in this manner although this case poses a greater technical difficulty. Unfortunately, a complete proof along these lines would involve repetition of most of the proof of Theorem 2 given above, and hence no advantage is gained with this alternate approach in this case.

3. Examples.

EXAMPLE 3.1. Random walk in a random environment on a circular lattice.

Consider a particle executing a random walk in a random environment on the lattice $\{0, 1, \ldots, N-1\}$ modulo N; i.e. $\{(q_n, r_n, p_n)\}_{n=0}^{\infty}$ is a stationary and ergodic sequence of positive random variables such that

$$p_n + q_n + r_n = 1, \tag{23}$$

where, conditional on the environment $\{(p_n, q_n, r_n)\}_{n=0}^{\infty}$, p_n is the probability of a unit move in the counterclockwise direction, q_n is the probability of a unit move in the clockwise direction, and r_n is the probability of sitting still on the n^{th} move. If X_n denotes the transition probability matrix conditional on the environment, then

$$X_n = \begin{matrix} r_n & p_n & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & q_n \\ q_n & r_n & p_n & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & q_n & r_n & p_n & 0 & \cdots & 0 & 0 & 0 & 0 \\ \vdots & & & & & & & & & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & q_n & r_n & p_n \\ p_n & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & q_n & r_n \end{matrix} \tag{24}$$

where X_n is an $N \times N$ matrix. If H_n denotes the position of the particle at time n, and if $\overline{\xi}$ denotes the environment $\{(q_n, r_n, p_n)\}_{n=0}^{\infty}$, then Theorem 1 states that, for $0 \leq i, j \leq N-1$,

$$(25) \qquad \lim_{n\to\infty} P(H_n = j \mid H_0 = i, \overline{\xi}) = \frac{1}{N} \quad \text{a.s.}$$

if we assume that there exists a positive integer k independent of H_0 such that, starting at H_0, all states are accessible at time k. ("Accessible at time k" means that there is positive probability that the k-step transition probability is strictly positive; i.e. that condition (A) holds.) This result is related to the fact that the only possible limit distributions for the convolution of a measure on a compact group G is the uniform measure on a subgroup K (Heyer (1977), Thm. 2.17, page 90). Of possible greater interest is the use of examples of this type to construct pseudo-random number generators, and in this regard, to extend the results of Brown and Solomon (1976). We conjecture, in fact, that using a stationary ergodic process in place of a convolution measure to construct a pseudo-random number generator will allow the user greater freedom in specifying "desirable" properties of pseudo-random numbers generated.

We will now provide a necessary and sufficient condition for accessibility which can be easily verified in practice.

PROPOSITION 2. A random walk in a random environment on a circular lattice is accessible if and only if

(B) $P(0 < r_0 < 1, \text{ or } 0 < p_0 < 1 \text{ and } 0 < q_0 < 1) > 0$

if the number of states N is odd, or

(C) $P(0 < r_0 < 1) > 0$

if the number of states N is even.

PROOF. Let $J(i,n) = \{j: (Y_n)_{ij} > 0\}$ where $Y_n = X_0 \cdot X_1 \cdots X_n$ is the $n+1$ - step transition probability matrix, and let $\alpha(i,n)$ denote the length of the largest interval contained in $J(i,n)$. (For example $[N-1,0,1]$ is an interval of length 3.) It is easy to see that $\alpha(i,n)$ is a non-decreasing function of n. For if $0 < r_{n+1} \leq 1$, then $J(i,n) \subseteq J(i,n+1)$; and if $r_{n+1} = 0$, then either $p_{n+1} > 0$ or $q_{n+1} > 0$ which implies that either

(26) $\{(j+1) \bmod N: j \in J(i,n)\} \subseteq J(i,n+1)$

or

(27) $\{(j-1) \bmod N: j \in J(i,n)\} \subseteq J(i,n+1)$.

If $0 < r_{n+1} < 1$ then either $p_{n+1} > 0$ or $q_{n+1} > 0$ from which it follows that, if $\alpha(i,n) < N$ then

(28) $\alpha(i,n+1) \geq \alpha(i,n) + 1$.

Suppose $P(0 < r_0 < 1) > 0$. Since $\{\xi_n\}_{n=0}^{\infty}$ is stationary and ergodic, then w.p. 1 there exists a sequence $n_k = n_k(\bar{\xi})$ strictly increasing such that $0 < r_{n_k} < 1$. It follows that for $n \geq n_N$, $\alpha(i,n) = N$. Since n_N is finite w.p. 1, there exists an integer K such that $P(n_N = K) > 0$, and it follows immediately that

$$(29) \qquad P(\min_{1 \le i,j \le N}(Y_K)_{ij} > 0) \;>\; 0$$

so that the random walk on the circular lattice is accessible. Hence condition (C) is a sufficient condition for accessibility whether N is odd or even.

Case 1. N even. Suppose condition (C) fails. Then $P(r_0 = 0 \text{ or } r_0 = 1) = 1$. If $\xi_n = (0,1,0)$ then, for each i, $J(i,n-1) = J(i,n)$; i.e. there is no change. If $\xi_n = (0,0,1)$ then

$$(30) \qquad J(i,n) = \{(j+1) \bmod N : \; j \in J(i,n-1)\}$$

so that $\alpha(i,n) = \alpha(i,n-1)$. A similar argument holds for $\xi_n = (1,0,0)$. Now $|J(i,0)| \le 2$ where $|J(i,n)|$ denotes the cardinality of the set $J(i,n)$ and furthermore $\alpha(i,0) = 1$. If $\xi_n = (q_n,0,p_n)$ where $q_n > 0$ and $p_n > 0$, and if $|J(i,n-1)| \le N/2$, then $\alpha(i,n) = 1$ from which it follows that $|J(i,n)| \le N/2$ for all n, and hence the random walk in the circular lattice is not accessible.

Case 2. N odd. We will first show that $P(0 < p_0 < 1 \text{ and } 0 < q_0 < 1) > 0$ implies that the random walk is accessible. We may without loss of generality assume $P(r_0 = 0 \text{ or } r_0 = 1) = 1$ since otherwise we have already shown that the random walk is accessible. Now it is easy to see that if $\xi_n = (q_n,0,p_n)$ where $q_n > 0$ and $p_n > 0$ and if $|J(i,n-1)| < (N+1)/2$, then

$$(31) \qquad |J(i,n)| \;=\; |J(i,n-1)| \;+\; 1.$$

Since the process $\bar{\xi}$ is stationary and ergodic, there exists w.p. 1 a strictly increasing sequence $n_k = n_k(\bar{\xi})$ such that $p_{n_k} > 0$, $q_{n_k} > 0$, and $r_{n_k} = 0$. It follows that, for $n \ge n_N$, $|J(i,n)| \ge (N+1)/2$ which implies that $\alpha(i,n) \geq 2$. But it is easy to see that if

(32) $$2 \leq \alpha(i,n) < N$$

and if $p_{n+1} > 0$, $q_{n+1} > 0$, and $r_{n+1} = 0$, then

(33) $$\alpha(i,n+1) \geq \alpha(i,n) + 1$$

from which it follows that for $n \geq n_{2N}$, $\alpha(i,n) = N$. Since n_{2N} is finite w.p. 1, there exists an integer K such that $P(n_{2N} = K) > 0$, and so for this K

(34) $$P(\min_{1 \leq i,j \leq N}(Y_K)_{ij} > 0) > 0;$$

i.e. the random walk is accessible. It follows that condition (B) is a sufficient condition for the random walk to be accessible if N is odd.

Suppose condition (B) fails. Then $P(\xi_0 \in \{(0,1,0),(1,0,0),(0,0,1)\}) = 1$. It is easy to see that in this case $|J(i,n)| \equiv 1$ for all n and $i = 1,\ldots,N$ so that the random walk is not accessible. []

EXAMPLE 3.2. A sterile multi-type branching process in a random environment (MBPRE). (See Athreya and Ney (1970) or Karlin and Taylor (1975) for the definition of MBPRE). Consider a collection of particles with x_1 of type 1, x_2 of type 2, x_N of type N. Suppose that all the particles are sterile but that $p_{ij}(\xi_n) = (X_n)_{ij}$ equals the probability that an i-type particle in the n-th generation "mutates" to become a j-type particle in the (n+1)-th generation (conditional on the environment $\xi_n = X_n$). Suppose that it is equally likely for a j-type particle to mutate to a i-type particle as it is for an i-type particle to mutate to a j-type particle; i.e. $(X_n)_{ij} = p_{ij}(\xi_n) = p_{ji}(\xi_n) = (X_n)_{ji}$ for all $i,j = 1,\ldots,N$. Then the matrix $X_n = P_n = P(\xi_n) = (p_{ij}(\xi_n))$ is doubly stochastic. A necessary and sufficient condition for $X_0 \cdot X_1 \cdots X_n$ to converge to $(1/N)J$ a.s. as n tends to infinity is that there exists a positive integer $K \geq 0$ such that

(35) $$P(\min_{1 \leq i,j \leq N}(X_0 \cdots X_K)_{ij} > 0) > 0.$$

The interpretation of this result is that, assuming equation (35), the expected number of particles of each type is the same "in the long run" independent of the initial population.

EXAMPLE 3.3. Random pairwise averaging. Let $\mathcal{T} = \{T_{ij}\}$ be the collection of 'pairwise averaging matrices' described in Section 1. Note again that $\mathcal{T}$ contains $\binom{N}{2}$ elements and that $\mathcal{T} \subset \mathcal{M}^+$. Suppose that $\{X_n\}_{n=1}^{\infty}$ is a stationary ergodic sequence with values in $\mathcal{T}$ and that

(D) $$P(X_1 = T_{ij}) > 0 \qquad \text{for each } T_{ij} \in \mathcal{T}.$$

These conditions are clearly satisfied if, for example, the X_i's are independent and uniformly distributed over $\mathcal{T}$ as in Section 1. We will now show that (D) implies condition (A) of Theorem 2 and hence, by Theorem 2, $Y_n = X_1 \cdots X_n \to (1/N)J$ a.s. as $n \to \infty$.

Let $J(i,n) = \{j: (Y_n)_{ij} > 0\}$. Since $(X_k)_{jj} \in \{1,\frac{1}{2}\}$ for all $k \geq 1$ and all $j = 1,\ldots,N$, it follows that $J(i,n) \subseteq J(i,n+1)$ for all $n \geq 1$. Let $1 \leq k_0 \leq N$ be fixed. Let A denote the unique element of $\mathcal{T}$ such that $(A)_{ik_0} = 1/2$. Since $\{X_n\}_{n=1}^{\infty}$ is a stationary and ergodic sequence and since $P(X_1 = A) > 0$, there exists w.p. 1 an integer $n_0 = n_0(\omega)$ such that $X_{n_0(\omega)} = A$. Now

(37) $$(Y_{n_0})_{ik_0} \geq (Y_{n_0-1})_{ii}(X_{n_0})_{ik_0} = \frac{1}{2}(Y_{n_0-1})_{ii} > 0$$

since $i \in J(i,1)$ implies that $(Y_n)_{ii} > 0$. Hence $k_0 \in J(i,n)$ for $n \geq n_0$. Since $1 \leq k_0 \leq N$ was arbitrary, it follows that there exists a positive integer $n_i = n_i(\omega)$ such that, for all $n \geq n_i$, $J(i,n) = \{1,2,\ldots,N\}$. Let $n^* = n^*(\omega) = \sup_{1 \leq i \leq N} n_i(\omega)$. Then for $n \geq n^*$, the cardinality of $J(i,n)$ equals N for all i. Since $n^* < \infty$ w.p. 1, there exists a positive integer $K > 0$ such that $P(n^* = K) > 0$. But

(38) $$\min_{1 \leq i,j \leq N} (Y_{n^*})_{ij} > 0$$

and hence

(39) $$P(\min_{1 \leq i,j \leq N} (Y_K)_{ij} > 0) > 0;$$

i.e. condition (A) of Theorem 2 is satisfied.

By use of arguments similar to those used in Examples 3.1 and 3.3, it can be seen that the following are sufficient conditions for condition (A) of Theorem 2 to hold:

(A_1) $P(\min_{1 \leq i,j \leq N} (X_1)_{ij} > 0) > 0;$

i.e., there is positive probability that all the entries of X_1 are strictly positive.

(A_2) $P(\min_{1 \leq i \leq N} (X_1)_{ii} > 0) = 1$ and $\min_{1 \leq i,j \leq N} P((X_1)_{ij} > 0) > 0;$

i.e. all the diagonal elements of X_1 are strictly positive w.p. 1, and each element of X_1 has positive probability of being non-zero. (This condition handles Example 3.3.)

(A_3) $\{X_n\}_{n=1}^{\infty}$ are independent and identically distributed and take values in a finite subset $\mathcal{A}$ of $\mathcal{M}^+$ such that there exists $A \in \mathcal{A}$ which is an irreducible aperiodic matrix and $P(X_1 = A) > 0$.

Acknowledgements. Our original proof of Theorem 1 was based on a martingale argument suggested by David Ruppert; the present proof is due to J.H.B. Kemperman.

REFERENCES

Athreya, K.B., and Ney, P. (1970). Branching Processes. Springer-Verlag, Berlin.

Breiman, L. (1968). Probability. Addison-Wesley, Reading.

Bretagnolle, Jean (1980). Statistique de Kolmogorov-Smirnov pour un enchantillon non-equireparti. To appear, proceedings of the St. Flour Conf.

Brown, M. and Solomon, H. (1976). On a method of combining pseudo-random number generators. Tech. Report Number 233, Department of Statistics, Stanford University, Stanford, California.

Cogburn, Robert (1982). The ergodic theory of Markov chains in random environments. Preprint, University of New Mexico.

Diaconis, P. and Shahshahani, M. (1981). Generating a random permutation with random transpositions. Zeitschrift fur Wahrscheinlichkeitstheori[e] und verw. Gebiete 57 159-179.

Heyer, H. (1977). Probability Measures on Locally Compact Groups. Springer-Verlag, Berlin.

Karlin, S., and Taylor, H. (1975). A First Course in Stochastic Processes. Academic Press, New York.

Kaijser, Thomas (1975). A limit theorem for partially observed Markov chains. Ann. Probability 3 677-696.

References (cont.)

Marshall, Albert W., and Olkin, Ingram (1979). Inequalities: Theory of Majorization and Its Applications. Academic Press, New York.

Moore, J. (1981). Random Products of Stochastic Matrices. Ph.D. thesis, University of California, Los Angeles.

Orey, S. (1971). Limit Theorems for Markov Chain Transition Probabilities. Van Nostrand - Reinhold, London.

Proschan, F. and Shaked, M. (1982). Random averaging of vector elements. Florida State University Technical Report M638.

Rosenblatt, M. (1960). Limits of convolution sequences of measures on a compact topological semigroup. Jour. Math. Mech. **9** 293-305.

Seneta, E. (1973). Non-Negative Matrices, An Introduction to Theory Applications. Springer-Verlag, New York.

Tanny, D. (1981). On multi-type branching processes in a random environment. Adv. Appl. Prob. **13** 464-497.

York University
4700 Keele Street,
Downsview, Ontario
M3J 1P3 Canada

Department of Statistics
University of Washington
Seattle, Washington 98195
U.S.A.