The Annals of Mathematical Statistics
1970, Vol. 41, No. 2, 539–550

ERGODIC THEOREMS FOR INFINITE PROBABILISTIC TABLES[1]

BY A. PAZ

University of California at Los Angeles and Israel Institute of Technology

0. Summary. This paper is concerned with infinite-state probabilistic transition tables, representing the dynamical behavior of a probabilistic automaton and nonhomogeneous infinite-state Markov chains. Many known theorems are generalized from the finite-state case to the infinite-state case. A possible application to the problem of computing, approximately, products of infinite-state stochastic matrices is outlined.

1. Introduction and basic definitions. Let $P = [p_{ij}]$ be an infinite (countable) Markov matrix (i.e., $p_{ij} \geqq 0$, $i, j = 1, 2, \cdots$ and $\sum_j p_{ij} = 1$, $i = 1, 2, \cdots$). It is well known that Markov matrices (finite or infinite) are closed under ordinary multiplication of matrices, moreover multiplication is associative for infinite Markov matrices (as well as for finite Markov matrices). Markov matrices are sometimes called stochastic matrices and we shall use both terms as convenient.

Our work here is concerned with two notions derived from Markov matrices and defined as follows:

DEFINITION 1. *A probabilistic infinite table* (PIT) is a triple $(\Sigma, S, \{A(\sigma)\})$ where Σ is a finite set (representing an alphabet) S is a countable infinite set (representing an infinite set of states) and $\{A(\sigma)\}$ is a finite set of Markov matrices of infinite (countable) order ($A(\sigma)$ represents the transition probabilities from state to state related to the symbol σ) a matrix $A(\sigma)$ for each symbol $\sigma \in \Sigma$.

The PIT as defined above represents a device having a countable number of internal states and changing its state upon receiving external inputs, one change with every input symbol $\sigma \in \Sigma$. The change of states is probabilistic and is governed by the matrix $A(\sigma)$ associated with the input σ in a way such that $a_{ij}(\sigma)(A(\sigma) = [a_{ij}(\sigma)])$ is the probability of going from state i to state j when input σ is received.

NOTATION. If $x = \sigma_1 \cdots \sigma_k$ is a word (equals a sequence of symbols) in Σ^* (equals the set of all words over Σ) then $A(x)$ denotes the matrix $A(x) = A(\sigma_1) \cdots A(\sigma_k)$. It is clear that the entries in $A(x)$ represent the transition probabilities between states induced by the word x.

DEFINITION 2. *A nonhomogeneous infinite* Markov chain (NIMC) is a couple $(S, (P_i))$ where S is a countable set (the states of the system) and (P_i) is an infinite sequence of Markov matrices of infinite (countable) order.

Received January 3, 1969; revised July 23, 1969.

[1] Research sponsored in part by the Air Force Office of Scientific Research, Office of Aerospace Research, United States Air Force under AFOSR Grants AF-AFOSR-639-67, (at the University of California, Berkeley) and AF-AFOSR-68-1408 (at the University of California, Los Angeles).

As in the PIT case the matrix P_i represents the transition probabilities from state to state of the system at time $t = i$.

NOTATION. The product $\prod_{i=m+1}^{n} P_i$ is denoted by $H_{mn}(n > m)$. The ij entry in H_{mn} represents the probability that the system will enter the state j at time $t = n$ if it was at state i at time $t = m$.

Note that if all the P_i's are equal or if all $A(\sigma)$'s are equal in Definition 1, then both concepts reduce to an ordinary Markov (infinite) *homogeneous* chain.

In what follows we shall be concerned with the properties of the matrices $A(x)$ when $l(x)$ (equals the length of x, or the number of symbols in the word x) grows indefinitely for PIT's and with the properties of the matrices H_{mn} when n grows indefinitely and m fixed for NIMC's. The two concepts defined above have been the subject of intense study in the past 15 years (see the references) and many dynamical systems can be simulated in them. In particular the ùnderlying structure of a probabilistic automaton is a PIT. Most of the research on PIT or NIMC's in the past was devoted to the finite-state case (S finite). On the other hand there is a tendency, in the last few years, to extend the study of dynamical systems to the infinite-state case.

Our aim here is to extend some of the results proved previously for finite-state PIT's and NIMC's to the infinite-state case. Some of our results may have application to the problem of computing products of infinite-state stochastic matrices in real time and with the computation error kept under control. The main tools to be used will be defined in the next section.

2. Functionals over Markov matrices.

NOTATION. If a is a real number then a^+ denotes the number $a^+ = \max(a, 0)$ and a^- denotes the number $a^- = \min(a, 0)$.

DEFINITION 3. Let $A = [a_{ij}]$ be an infinite stochastic matrix. The functional $\delta(A)$ is defined as:

$$\delta(A) = \sup_{i_1, i_2} \sup_{\{n'\}} \sum_{j \in \{n'\}} (a_{i_1 j} - a_{i_2 j})$$

where $\{n'\}$ denotes a subset of the set of natural numbers.

LEMMA 1. *If A is a stochastic matrix then* $0 \leqq \delta(A) \leqq 1$ *and*

$$\delta(A) = \sup_{i_1, i_2} \sum (a_{i_1 j} - a_{i_2 j})^+.$$

PROOF. Straightforward and is left to the reader.

REMARK. If A is a stochastic matrix then

$$2\delta(A) = \sup_{i_1 i_2} \sum_j |a_{i_1 j} - a_{i_2 j}|. \tag{1}$$

PROOF. It follows from $\sum_j a_{i_1 j} - \sum_j a_{i_2 j} = 0$ that $\sum(a_{i_1 j} - a_{i_2 j})^+ = -\sum(a_{i_1 j} - a_{i_2 j})^-$. Thus $\sum_j |a_{i_1 j} - a_{i_2 j}| = 2\sum_j (a_{i_1 j} - a_{i_2 j})^+$ and (1) follows by Lemma 1.

DEFINITION 4. Let A be an infinite stochastic matrix. The functional $\gamma(A)$ is defined as

$$\gamma(A) = \inf_{i_1, i_2} \sum_{j=1}^{\infty} \min(a_{i_1 j}, a_{i_2 j}).$$

DEFINITION 5. If A is an arbitrary matrix (not necessarily square, vectors included) then the functional $||A||$ is defined as $||A|| = \sup_i \sum_j |a_{ij}|$ (in particular $||A||$ may assume an infinite value).

LEMMA 2. *If A and B are arbitrary matrices such that the product AB is defined then* $||AB|| \leqq ||A|| \cdot ||B||$.

PROOF. Straightforward and left to the reader.

DEFINITION 6. A stochastic matrix is called *constant* if all its rows are equal one to the other.

LEMMA 3. *Let A be an infinite stochastic matrix and let A_{i_0} be the matrix all the rows of which are equal to the i_0 row of A then $\delta(A) \geqq \frac{1}{2}||A - A_{i_0}||$ but for every $\varepsilon > 0$ there is an index i_0 such that $\delta(A) \leqq \frac{1}{2}||A - A_{i_0}|| + \varepsilon$.*

PROOF. The lemma follows directly from formula (1) above.

LEMMA 4. *Every stochastic matrix A can be represented in the form $A = E + Q$ where E is a constant stochastic matrix and $||Q|| \leqq 2\delta(A)$.*

PROOF. Let E be a matrix A_{i_0} as in Lemma 3 then $Q = A - A_{i_0}$ and $||Q|| = ||A - A_{i_0}|| \leqq 2\delta(A)$ by Lemma 3.

LEMMA 5. *Let $A_1 \cdots A_n$, $\bar{A}_1 \cdots \bar{A}_n$ be two sets of n matrices such that $||A_i|| \leqq 1$, $||\bar{A}_i|| \leqq 1$ and $||A_i - \bar{A}_i|| \leqq \varepsilon$ for $i = 1, \cdots, n$ and some $\varepsilon > 0$. Assume in addition that the products $\prod_{i=1}^{n} A_i$ and $\prod_{i=1}^{n} \bar{A}_i$ are defined. Then $|| \prod_{i=1}^{n} A_i - \prod_{i=1}^{n} \bar{A}_i || \leqq n\varepsilon$.*

PROOF. Straightforward and is left to the reader.

LEMMA 6. *If A is a stochastic matrix then $\delta(A) = 1 - \gamma(A)$.*

PROOF. Set $\gamma_{i_1 i_2}(A) = \sum_{j=1}^{\infty} \min(a_{i_1 j}, a_{i_2 j})$ and $\delta_{i_1 i_2}(A) = \sum_{j=1}^{\infty} (a_{i_1 j} - a_{i_2 j})^+$. Then by Lemma 1,

$$\delta(A) \geqq \delta_{i_1 i_2}(A) = \sum_{j=1}^{\infty} (a_{i_1 j} - a_{i_2 j})^+ = \sum_{j=1}^{\infty} [a_{i_1 j} - \min(a_{i_1 j}, a_{i_2 j})]$$
$$= 1 - \sum_{j=1}^{\infty} \min(a_{i_1 j}, a_{i_2 j}) = 1 - \gamma_{i_1 i_2}(A).$$

Thus $\delta(A) \geqq 1 - \gamma(A)$ for the inequality above does not depend on i_1 or i_2. Similarly $\delta_{i_1 i_2}(A) = 1 - \gamma_{i_1 i_2}(A) \leqq 1 - \gamma(A)$ by the definition of $\gamma(A)$ and the above sequence of equalities. This implies that $\delta(A) \leqq 1 - \gamma(A)$ as the inequality above does not depend on i_1 and i_2. It follows that $\delta(A) = 1 - \gamma(A)$ as required.

LEMMA 7. *If A and B are stochastic matrices then $\delta(AB) \leqq \delta(A)\delta(B)$.*

PROOF. This lemma has been proved in [7] page 779.

LEMMA 8. *Let A be a stochastic matrix and let ξ be a nonzero row vector of the same dimension as A (including the infinite case) such that $||\xi|| < \infty$ and $\sum_i \xi_i = 0$ $(\xi = (\xi_i))$. Then $||\xi A|| \leqq ||\xi||\delta(A)$.*

PROOF. Define the vectors $\zeta^1 = (\zeta_i^{\,1})$ and $\zeta^2 = (\zeta_i^{\,2})$ as follows:

$$\zeta_i^{\,1} = 2\xi_i^{\,+}/||\xi|| \quad \text{and} \quad \zeta_i^{\,2} = 2|\zeta_i^{\,-}|/||\xi||.$$

Using an argument similar to the one used in the proof of Lemma 3 one can easily prove that $2\sum_i \xi_i^{\,+} = \sum_i |\xi_i| = 2\sum_i |\xi_i^{\,-}|$ which implies that both ζ^1 and ζ^2 are stochastic vectors with

$$\zeta^1 - \zeta^2 = 2\xi/||\xi||. \tag{2}$$

Let B be a matrix of the same order as A such that its first row equals ζ^1 all its other rows being equal to ζ^2. Then:

$$2\delta(BA) = 2\sum_j(\sum_k(\zeta_k^{\,1} - \zeta_k^{\,2})a_{kj})^+ = \sum_j|\sum_k(\zeta_k^{\,1} - \zeta_k^{\,2})a_{kj}|$$

by Lemma 1 and using the argument used in the proof of Lemma 3. It follows that (using the relation (2))

$$2\delta(BA) = \sum_j |\sum_k (\zeta_k^{\,1} - \zeta_k^{\,2})a_{kj}| = 2\sum_j \left|\sum_k \frac{\xi_k}{||\xi||} a_{kj}\right| = \frac{2}{||\xi||}\sum_j |\sum_k \xi_k a_{kj}| = 2\frac{||\xi A||}{||\xi||}$$

But by Lemmas 7 and 1, $\delta(BA) \leqq \delta(B)\delta(A) \leqq \delta(A)$ with the consequence that $||\xi A||/||\xi|| = \delta(BA) \leqq \delta(A)$ as required.

COROLLARY 1. *Let A be a matrix $||A|| < \infty$ such that its rows have the properties of the vector ξ in Lemma 8, and let B be a stochastic matrix, A and B being of the same order. Then $||AB|| \leqq ||A||\delta(B)$.*

PROOF. Trivial, using the definitions and Lemma 8.

COROLLARY 2. *Let A and B be stochastic matrices of the same order. Then $||AB - B|| \leqq 2\delta(B)$. In particular if π is a stochastic vector and ζ is a row of the matrix B (π and ζ of the same order) then $||\pi B - \zeta|| \leqq 2\delta(B)$.*

PROOF. $||AB - B|| = ||(A - I)B|| \leqq ||A - I||\,\delta(B) \leqq (||A|| + ||I||)\,\delta(B) = 2\delta(B)$ by Corollary 1 with I the identity matrix.

REMARKS. (1) The lemmas and corollaries proved in this section will serve as main tools for the proofs of the theorems in the following sections, some of them however are of independent interest. All the above lemmas are true in the infinite-state as well as in the finite-state case. In the finite-state case, however, "sup" is to be replaced by "max" and "inf" is to be replaced by "min".

(2) Many authors, working on Markov chains, have used either the functional $\delta(A)$ which measures, in a certain sense, how different the rows of A are, or the functional $\gamma(A)$ as their main tool. Some writers have tried to establish some relation between the two functionals (see [1], [3], [6] and [7]). It was a surprise to find out that the two functionals are connected by the simple relation given in

Lemma 6, for their definition is quite different. It is to be mentioned, however, that Dobrushin [1] was very close to the proof of our Lemma 6 here but, as he was mainly interested in the functional $\gamma(A)$, he did not even define explicitly the functional $\delta(A)$. Dobrushin [1] also proved the inequality $1-\gamma(AB) \leqq (1-\gamma(A))(1-\gamma(B))$ for stochastic matrices A and B, which (by Lemma 6) is equivalent to Lemma 7 here, and proved implicitly the inequality $\|\xi A\| \leqq \|\xi\| (1-\gamma(A))$ for ξ and A as in our Lemma 8. This last inequality is equivalent (by Lemma 6) to our Lemma 8. On the other hand the proofs of Dobrushin are different and seem to be more complicated than our proofs here.

Before proceeding to the next sections we shall exhibit some uses of the above lemmas to the study of doubly stochastic matrices, where a doubly stochastic matrix is a matrix A such that both A and A^T (equals A transpose) are stochastic.

PROPERTY 1. Let A be a doubly stochastic matrix such that $\delta(A) = 0$ then A is of finite order, say n, and all the entries of A are equal to $1/n$.

PROOF. By $\delta(A) = 0$ we have that $a_{ij} = a_{kj} = \alpha_j$. Thus by $\sum_i a_{ij} = 1$ the order of A is finite (say n). Hence $na_j = 1$ and $a_{ij} = a_{kj} = \alpha_j = 1/n$.

PROPERTY 2. Let A be a doubly stochastic matrix of finite order n such that $\delta(A) < 1$, and let E be a square matrix of order n all the entries of which are equal to $1/n$. Then $\lim_{m\to\infty} \|A^m - E\| = 0$.

PROOF. If A is doubly stochastic then so is A^m for any m (this is easy to prove and well known). Furthermore it is easy to prove that for any doubly stochastic matrix B, $EB = E$ where E is as specified above (this is left to the reader to verify). Thus $\|A^m - E\| = \|A^m - EA^m\| \leqq 2\delta(A^m) \leqq 2[\delta(A)]^m \to 0$, by Corollary 2, Lemma 7 and by the requirement that $\delta(A) < 1$.

PROPERTY 3. Let A be a doubly stochastic matrix of infinite (countable) order, then $\delta(A) = 1$.

PROOF. Assume that $\delta(A) < 1$. Then for any $\frac{1}{2} > \varepsilon > 0$ there is n with $\delta(A^n) \leqq [\delta(A)]^n < \varepsilon$. As A is stochastic we have also that for any fixed i_0 there is an integer k such that

$$\sum_{j=1}^{k} a_{i_0 j}^{(n)} > 1 - \varepsilon \tag{3}$$

where ε is as above and $A^n = [a_{ij}^{(n)}]$. On the other hand, for any i_0 and i we have that $\left|\sum_{j=1}^{k} a_{i_0 j}^{(n)} - \sum_{j=1}^{k} a_{ij}^{(n)}\right| \leqq \delta(A^n) < \varepsilon$ implies, by the previous inequality (3) that

$$\sum_{j=1}^{k} a_{ij}^{(n)} \geqq 1 - 2\varepsilon \qquad i = 1, 2, \cdots.$$

Thus $\sum_{i=1}^{\infty} \sum_{j=1}^{k} a_{ij}^{(n)} = \infty$. But, as A^n is doubly stochastic, we have also that $\sum_{i=1}^{\infty} \sum_{j=1}^{k} a_{ij}^{(n)} = \sum_{j=1}^{k} \sum_{i=1}^{\infty} a_{ij}^{(n)} = k$, a contradiction. It follows that $\delta(A) = 1$.

3. Nonhomogeneous infinite Markov chains. Considering again the definitions and their interpretation, and remembering that $\delta(A)$ and $\gamma(A)$ provide us, in a certain sense, with a measure of the "distance" between two arbitrary rows of the stochastic matrix A, we shall distinguish between two cases for the long-range behavior of a given NIMC.

Case 1. $\lim_{n\to\infty}\delta(H_{mn}) = 0$ for $m = 1, 2, \cdots$ where $H_{mn} = \prod_{i=m+1}^{n} P_i$ (see Definition 2 and the notation after). In this case the chain is called weakly ergodic.

Case 2. For any given integer m there is a constant matrix Q (see Definition 6) such that $\lim_{n\to\infty} \|H_{mn} - Q\| = 0$ in this case the chain is called strongly ergodic.

In addition to the two above distinctions there may be other distinctions as well (e.g., the matrix Q in the second case may not be constant or the limit, in both cases, may exist only for some m but not for all m etc.) but, because of their restrictive nature, those distinctions will not be considered here.

Most of the theorems to be proved here are generalizations to the infinite-state case of theorems which have been proved previously for the finite-state case (see [3], [5] and [6]). In the finite-state case, however, a weaker norm for matrices has been used, namely $|A|$ instead of $\|A\|$ where $|A| = \max_{i,j}|a_{ij}|$, $A = [a_{ij}]$. Also a weaker measure $d(A)$ has been used instead of $\delta(A)$ where $d(A) = \max_j \max_{i_1 i_2} |a_{i_1 j} - a_{i_2 j}|$. It was first necessary to realize that these functionals, $d(A)$ and "$|A|$" although definable for infinite-state matrices too, will not suffice and must be replaced by the functionals $\delta(A)$ and "$\|A\|$" to make the generalizations possible.

THEOREM 1. *An* NIMC *is weakly ergodic if, and only if, there exists a subdivision of the chain into blocks of matrices* $\{H_{i_j i_{j+1}}\}$ *such that* $\sum_{j=1}^{\infty} \gamma(H_{i_j i_{j+1}})$ *diverges*, $(i_1 = 1)$.

PROOF. The condition is sufficient, for $\sum_{j=1}^{\infty} \gamma(H_{i_j i_{j+1}})$ diverges implies that for any j_0, $\lim_{n\to\infty} \prod_{j=j_0}^{n}(1-\gamma(H_{i_j i_{j+1}})) = 0$ and using Lemmas 7, 1, and 6 we have that:

$$\delta = (\prod_{i=m}^{m+n} P_i) \leqq \delta(\prod_{i_j > m}^{m+n} H_{i_j i_{j+1}}) \leqq \prod_{i_j > m}^{m+n} \delta(H_{i_j i_{j+1}}) = \prod_{i_j > m}^{m+n}(1-\gamma(H_{i_j i_{j+1}}))$$

where $i_j > m$ means that the product begins with the first index $i_j \geqq m$. Taking limits on both sides we get that

$$\lim_{n\to\infty} \delta(\prod_{i=m}^{m+n} P_i) \leqq \lim_{n\to\infty} \delta(\prod_{i_j > m}^{n} H_{i_j i_{j+1}}) = 0.$$

If $\lim_{n\to\infty} \delta(\prod_{i=m}^{n} P_i) = 0$ $m = 1, 2, \cdots$, then by Lemma 6

$$\lim_{n\to\infty} \gamma(\prod_{i=m}^{n} P_i) = \lim_{n\to\infty}(1-\delta(\prod P_i)) = (1-\lim_{n\to\infty}\delta(\prod P_i)) = 1.$$

Let $0 < \varepsilon < 1$ be a small constant, then it follows from the above inequalities that a sequence of blocks $H_{i_j i_{j+1}}$ can be found such that $\gamma(H_{i_j i_{j+1}}) > \varepsilon$ so that $\sum_{j=1}^{\infty} \gamma(H_{i_j i_{j+1}})$ diverges. $\square$

THEOREM 2. *A given* NIMC *is weakly ergodic if, and only if, for each m there is a sequence of constant Markov matrices* E_{mn} *such that* $\lim_{n\to\infty} \|H_{mn} - E_{mn}\| = 0$.

PROOF. Let $\varepsilon > 0$ be an arbitrary small number and let i_1, i_2 be two arbitrary indices. Let $H_{mn} = [a_{ij}]$ $E_{mn} = [e_{ij}]$ and suppose that n is so big that $\|H_{mn} - E_{mn}\| < \varepsilon$.

Then by (1) we have that

$$2\delta(H_{mn}) = \sup_{i_1 i_2} \sum_j |a_{i_1 j} - a_{i_2 j}| \leqq \sup_{j_1} \sum_j |a_{i_1 j} - e_{i_1 j}| + \sup \sum_j |a_{i_2 j} - e_{i_2 j}| \leqq 2\varepsilon$$

(for $e_{i_1 j} = e_{i_2 j}$, E_{mn} being constant). This proves that the condition is sufficient. That the condition is necessary follows trivially from (1).

THEOREM 3. *Let* $(S, (P_i))$ *be a given* NIMC *a and let* $P_i = E_i + R_i$ *with* E_i *a constant stochastic matrix. Then the given* NIMC *is weakly ergodic* iff $\lim_{n\to\infty} ||\prod_{i=m}^{m+n} R_i|| = 0$.

PROOF. One proves easily that if P_i is stochastic and E_i constant then $P_i E_i = E_i$ and $E_i P_i$ is constant. Thus $(P_1 - E_1)(P_2 - E_2) = P_1 P_2 - E_1 P_2$.
Hence, by induction

$$\prod_{m=1}^{m+n} R_i = \prod_{m=1}^{m+n} (P_i - E_i) = \prod_{i=m}^{m+n} P_i - E_m \prod_{i=m+1}^{m+n} P_i$$

where the second term on the right-hand side is constant. It thus follows that the condition of Theorem 3 implies the condition of Theorem 2, which implies weak ergodicity. On the other hand

$$||\prod_{i=m}^{m+n} P_i - E_m \prod_{i=m+1}^{m+n} P_i|| = ||(P_m - E_m) \prod_{i=m+1}^{m+n} P_i|| \leqq ||P_m - E_m||\, \delta(\prod_{i=m+1}^{m+n} P_i)$$

by Corollary 1 and therefore weak ergodicity implies the condition of our theorem. $\square$

The following theorem gives a characterization of strong ergodicity. It also confirms an intuitive feeling that strong ergodicity implies weak ergodicity.

THEOREM 4. *An* NIMC $(S, (P_i))$ *is strongly ergodic if, and only if, for every* m *there is a sequence of constant stochastic matrices* $\{E_{mn}\}$, *and a sequence of stochastic constant matrices* $\{E_m\}$ *such that*

(1) $\lim_{n\to\infty} ||H_{mn} - E_{mn}|| = 0$ *and*

(2) $\lim_{n\to\infty} ||E_{mn} - E_m|| = 0$.

PROOF. If (1) and (2) hold true then

$$\lim_{n\to\infty} ||H_{mn} - E_m|| \leqq \lim_{n\to\infty} ||H_{mn} - E_{mn}|| + ||E_{mn} - E_m|| = 0.$$

But if (1) and (2) hold true then E_m is independent on m. To prove this we note that $P_m H_{mn} = H_{m-1,n}$ and $P_m E_m = E_m$ (let the reader verify that if A is constant and B is stochastic then $BA = A$). Thus

$$\begin{aligned} ||E_{m-1} - E_m|| &\leqq ||E_{m-1} - H_{m-1,n}|| + ||P_m H_{mn} - P_m E_m|| \\ &= ||E_{m-1} - H_{m-1,n}|| + ||P_m(H_{mn} - E_m)|| \leqq ||E_{m-1} - H_{m-1,n}|| + ||P_m|| \cdot ||H_{mn} - E_m|| \end{aligned}$$

by Lemma 2. Using now the fact that for any stochastic matrix A, $||A|| = 1$ we conclude that

$$||E_{m-1} - E_m|| \leqq ||E_{m-1} - E_{m-1,n}|| + ||E_{m-1,n} - H_{m-1,n}|| + ||H_{mn} - E_{mn}|| + ||E_{mn} - E_m||.$$

Now allowing n to grow indefinitely in both sides of the inequality and using the conditions of the theorem we have that $||E_{m-1} - E_m|| = 0$ which is equivalent to

$E_{m-1} = E_m$, independently on m. Thus (1) and (2) imply that the chain is strongly ergodic. Conversely if the chain is strongly ergodic, then, setting $Q = E_{mn} = E_m$ for all m and n we have that (1) and (2) hold true.

COROLLARY 3. *A strongly ergodic chain is also weakly ergodic. A weakly ergodic chain which satisfies* (2) *is strongly ergodic.*

PROOF. Strong ergodicity implies the condition (1) in Theorem 4, which, by Theorem 2 implies weak ergodicity. Conversely, by Theorem 2 weak ergodicity implies (1) which together with (2) implies strong ergodicity by Theorem 4.

COROLLARY 4. *Conditions* (1) *and* (2) *in Theorem* 4 *and Corollary* 3 *can be replaced by the condition:*

$$(2') \qquad \lim_{n\to\infty} \|H_{mn} - E_m\| = 0.$$

PROOF. $\|H_{mn}-E_m\| \leqq \|H_{mn}-E_{mn}\| + \|E_{mn}-E_m\|$. *Thus* (1) *and* (2) *imply* (2′). *Conversely setting in* (1) *and* (2) $E_{mn} = E_m$ *for all* m *and* n *one sees that* (2′) *implies* (1) *and* (2).

COROLLARY 5. *Condition* (2) *in Theorem* 4 *and Corollary* 3 *can be replaced by the condition: there is a constant stochastic matrix* E *such that*

$$(2'') \qquad \lim_{n\to\infty} \|EH_{mn} - E\| = 0.$$

PROOF. $\|EH_{mn}-E\| \leqq \|EH_{mn}-H_{mn}\| + \|H_{mn}-E\| \leqq 2\delta(H_{mn}) + \|H_{mn}-E\|$ by Corollary 2. Condition (1) of Theorem 4 implies that $\delta(H_{mn}) \to 0$ and condition (2) in that theorem implies that $\|H_{mn}-E_m\| = \|H_{mn}-E\| \to 0$ (E_m is independent on m as proved in the proof of that theorem). Thus condition (2″) holds with $E = E_m$. Conversely if (1) and (2″) hold true, then let $E_m = E$. It follows that:

$$\|H_{mn}-E\| \leqq \|EH_{mn}-E\| + \|EH_{mn}-H_{mn}\| \leqq \|EH_{mn}-E\| + 2\,\delta(H_{mn}) \to 0$$

which is condition (2′).

THEOREM 5. *Let* $(S, (P_i))$ *and* $(S, (\bar{P}_i))$ *be two* NIMC*'s such that* $\sum_i \|P_i - \bar{P}_i\| < \infty$ *then for any* $\varepsilon > 0$ *there is an integer* m_0 *such that* $\|H_{mn} - \bar{H}_{mn}\| < \varepsilon$, *for all* $m \geqq m_0$ *and all* $n > m$ ($\bar{H}_{mn}$ *is the product of* $\bar{P}_i$*'s corresponding to* H_{mn}).

PROOF. Let $P_i - \bar{P}_i = E_i$ with $\|E_i\| = e_i$ then $H_{mn} = \prod(\bar{P}_i + E_i) = H_{mn} + R_{mn}$ where R_{mn} contains all possible products of $\bar{P}_i$ and E_i matrices. Using the facts that $\|E_i\| = e_i$ is finite for all i, $\|\bar{P}_i\| = 1$ for all i (P_i is stochastic), and Lemma 2 we have that

$$\|R_{mn}\| \leqq \sum e_i + \sum_{i,j} e_i e_j + \sum_{i,j,k} e_i e_j e_k + \cdots + \prod_{i=m+1}^{n} e_i = \prod_{i=m+1}^{n} (1+e_i) - 1.$$

Note that the e_i's are nonnegative. Now as $\sum e_i < \infty$, the product $\prod_{i=m+1}^{n} (1+e_i)$ converges and therefore for any ε there is m with $\|R_{mn}\| < \varepsilon$. The theorem is thus proved.

COROLLARY 6. *Let* $(S, (P_i))$ *be two* NIMC *satisfying the conditions of Theorem* 5. *If one of the chains is weakly ergodic then so is the other.*

PROOF. We must show that $\lim_{n\to\infty} \delta(\bar{H}_{mn}) = 0$ $m = 0, 1 \cdots$. It suffices to prove this for $m = m_0, m_0+1, \cdots$ as $\delta(\bar{H}_{in}) \leqq \delta(\bar{H}_{i,m_0-1})\delta(\bar{H}_{m_0n}) \leqq \delta(\bar{H}_{m_0n})$ by Lemmas 1 and 7. Let $\varepsilon > 0$ be a constant and choose m_0 such that $\|H_{mn}-\bar{H}_{mn}\| < \frac{1}{3}\varepsilon$ for all $m > m_0$, this is possible by the previous Theorem 5. Let m be a *fixed* but arbitrary integer $m > m_0$, and let $C^{i_0}_{mn}$ be the matrix all the rows of which are equal to the i_0 row of $\bar{H}_{mn}$. Choose i_0 such that $\delta(\bar{H}_{mn}) \leqq \frac{1}{2}\|\bar{H}_{mn}-\bar{C}^{i_0}_{mn}\|+\frac{1}{2}\varepsilon$ this is possible for fixed n by Lemma 4 and if we choose $n = m+1$ the above inequality will be true for all $n > m$ as $\delta(\bar{H}_{mn}) \leqq \delta(\bar{H}_{m,m+1})\delta(\bar{H}_{m+2,n}) \leqq \delta(\bar{H}_{m,m+1})$ by Lemma 1 and Lemma 7. We have thus shown that for any $m \geqq m_0$ there is an i_0 such that for any $n > m_0$ we have that $\delta(\bar{H}_{mn}) \leqq \frac{1}{2}\|\bar{H}_{mn}-\bar{C}^{i_0}_{mn}\|+\frac{1}{2}\varepsilon$. Now $\|\bar{H}_{mn}-\bar{C}^{i_0}_{mn}\| \leqq \|\bar{H}_{mn}-H_{mn}\|+\|H_{mn}-C^{i_0}_{mn}\|+\|C^{i_0}_{mn}-\bar{C}^{i_0}_{mn}\|$ where $C^{i_0}_{mn}$ is the matrix all the rows of which are equal to the i_0 row of H_{mn}. But $\|C^{i_0}_{mn}-\bar{C}^{i_0}_{mn}\| \leqq \|\bar{H}_{mn}-H_{mn}\|$ by definition, $\|\bar{H}_{mn}-H_{mn}\| \leqq \frac{1}{3}\varepsilon$ by the choice of m (for any $n > m$) and, as in the proof of Lemma 3, $\|H_{mn}-C^{i_0}_{mn}\| \leqq \delta(H_{mn})$. As the first chain is weakly ergodic, there is $n > m$ such that $\delta(H_{mn}) < \frac{1}{3}\varepsilon$ combining all these results together we have that for any $m \geqq m_0$ and $\varepsilon > 0$ there is $n > m$ such that $\delta(\bar{H}_{mn}) \leqq \frac{1}{2}(\frac{1}{3}\varepsilon+\frac{1}{3}\varepsilon+\frac{1}{3}\varepsilon)+\frac{1}{2}\varepsilon = \varepsilon$ which proves that also the chain $(S, (\bar{P}_i))$ is weakly ergodic. □

REMARK. NIMC's in general can be classified according to the following four types

Type	$\|H_{mn}-P_m\| \to 0$	$\delta(H_{mn}) \to 0$
Strongly ergodic	Yes	Yes
Weakly ergodic	No	Yes
Convergent	Yes	No
Oscillating	No	No

Where $\|H_{mn}-P_m\| \to 0$ means that for any m there is a matrix P_m (not necessarily constant) such that $\lim_{n\to\infty} \|H_{mn}-P_m\| = 0$. In Corollary 6 it is proved that if two chains satisfy the conditions of Theorem 5 and one of them is weakly ergodic then so is the other. It can be proved that the same is true for all the other three types of chains above.

4. Probabilistic infinite tables. The difference between PIT's to be dealt with in this section, and NIMC's considered in the previous section is that in the PIT model one studies the set of all possible products of Markov matrices taken from a (finite) given set of such matrices, while in the NIMC model one investigates a specific given infinite product of Markov matrices and its possible sub-products. The approach in this section is closer to the automaton concept where the set of all words over a given alphabet is studied with regard to the transitions induced on the states of the automaton by the different words.

The words correspond here to products of Markov matrices which induce a probabilistic transition between the states of the automaton. The reader is referred to Definition 1 and the notation after, for the following definitions and theorems.

DEFINITION 7. A PIT $(\Sigma, S, \{A(\sigma)\})$ is weakly ergodic if for any $\varepsilon > 0$ there is an integer $n = n(\varepsilon)$ such that $\delta(A(x)) \leqq \varepsilon$ for all words x such that $l(x) \geqq n(\varepsilon)$ where $l(x)$ denotes the length of the word x.

REMARK. If a Markov system is weakly ergodic then $\delta(A(x)) \to 0$ uniformly, the magnitude of $\delta(A(x))$ depending only on the length of x and not on the specific symbols contained in x. Such a requirement of uniformity will be too restrictive for the strong ergodicity and therefore strong ergodicity will not be dealt with for PIT's.

Note that $A(xy) = A(x)A(y)$ so that $\delta(A(xy)) \leqq \delta(A(x))\delta(A(y)) \leqq \delta(A(x))$ and therefore if $\delta(A(x)) \leqq \varepsilon$ for all x with $l(x) = n(\varepsilon)$ then $\delta(A(x)) \leqq \varepsilon$ for all x with $l(x) \geqq n(\varepsilon)$.

THEOREM 6. *A PIT is weakly ergodic iff there is an integer k such that $\delta(A(x)) < 1$ for all x such that $l(x) \geqq k$.*

PROOF. Necessity follows directly from the definition. To prove sufficiency set $\delta = \max_{l(x)=k} \delta(A(x)) < 1$ (there are only finitely many words x with $l(x) = k$ because Σ is finite). Let n_0 be an integer such that $\delta^{n_0} < \varepsilon$ for a given $\varepsilon > 0$. Let x be a word such that $l(x) \geqq kn_0$, then $x = y_1 \cdots y_{n_0} y$ where $l(y_1) = \cdots = l(y_{n_0}) = k$ and $l(y) \geqq 0$. Thus $\delta(A(x)) \leqq \delta(A(y_1)) \cdots \delta(A(y_{n_0})) \leqq \delta^{n_0} < \varepsilon$. It follows that the system is weakly ergodic.

REMARK. The theorem will remain true even if the alphabet Σ is infinite provided the requirement that $\delta(A(x)) < 1$ is replaced by the requirement that there is a real number $\delta < 1$ such that $\delta(A(x)) < \delta$ for all x with $l(x) \leqq k$.

THEOREM 7. *Let $(\Sigma, S, \{A(\delta)\})$ and $(\Sigma, S, \{\bar{A}(\sigma)\})$ be two PIT's such that the first is weakly ergodic and the second is arbitrary. There is $\varepsilon > 0$ such that if $\|A(\sigma) - \bar{A}(\sigma)\| \leqq \varepsilon$ for all $\sigma \in \Sigma$, then the second system is also weakly ergodic.*

PROOF. Using Theorem 6, we must prove that there is ε such that if $\|A(\sigma) - \bar{A}(\sigma)\| \leqq \varepsilon$ for all $\sigma \in \Sigma$, then there is n such that $\delta(A(x)) < 1$ for all x with $l(x) \geqq n$. Let $A_{i_0}(x)$ be the matrix such that all its rows are equal to the i_0 row of $A(x)$, then $\|A(x) - A_{i_0}(x)\| \leqq 2\delta(A(x))$ as in the proof of Lemma 3. As the first system is weakly ergodic, there is n_0 such that $\delta(A(x)) < \frac{1}{6}$ for all x with $l(x) \geqq n_0$ i.e., $\|A(x) - A_{i_0}(x)\| < \frac{1}{3}$ for all such x and any i_0. Let $\bar{x}$ be a fixed but arbitrary word with $l(\bar{x}) = n_0$ and choose i_0 so that $\delta(\bar{A}(\bar{x})) < \frac{1}{2}\|\bar{A}(\bar{x}) - \bar{A}_{i_0}(\bar{x})\| + \frac{1}{2}$. Such an i_0 exists by Lemma 4. Finally let ε be a number $0 < \varepsilon < \frac{1}{3}n_0$, and let $\|A(\sigma) - \bar{A}(\sigma)\| < \varepsilon$ for all $\sigma \in \Sigma$. Then by Lemma 5 we have that $\|A(\bar{x}) - \bar{A}(\bar{x})\| < \frac{1}{3}$ (for $l(\bar{x}) = n_0$). Thus $\delta(\bar{A}(\bar{x})) < \frac{1}{2} + \frac{1}{2}\|\bar{A}(\bar{x}) - \bar{A}_{i_0}(\bar{x})\| \leqq \frac{1}{2} + \frac{1}{2}(\|\bar{A}(\bar{x}) - A(\bar{x})\| + \|A(\bar{x}) - A_{i_0}(\bar{x})\| + \|A_{i_0}(\bar{x}) - \bar{A}_{i_0}(\bar{x})\|) \leqq \frac{1}{2} + \frac{1}{2}(\frac{1}{3} + \frac{1}{3} + \frac{1}{3}) = 1$, for $\|A_{i_0}(\bar{x}) - A_{i_0}(\bar{x})\| \leqq \|A(\bar{x}) - \bar{A}(\bar{x})\| < \frac{1}{3}$. But $\bar{x}$ is arbitrary and therefore we have that $\delta(\bar{A}(x)) < 1$ for all x with $l(x) = n_0$ provided that $\|A(\sigma) - \bar{A}(\sigma)\| < \frac{1}{3}n_0$ for all $\sigma \in \Sigma$, where n_0 is an integer such that $\delta(A(x)) < \frac{1}{6}$ for all x with $l(x) \geqq n_0$. To complete the proof we note that if $\delta(\bar{A}(x)) < 1$ for all x with $l(x) = n_0$ then $\delta(A(x)) < 1$ also for all x with $l(x) \geqq n_0$, as mentioned before.

THEOREM 8. *Let* $(\Sigma, S, \{A(\sigma)\})$ *and* $(\Sigma, S, \{\bar{A}(\sigma)\})$ *be two* PIT'*s such that the first is weakly ergodic. For any* $\delta > 0$ *there is* $\varepsilon > 0$ *such that if* $\|A(\sigma)-\bar{A}(\sigma)\| \leqq \varepsilon$ *for all* $\sigma \in \Sigma$ *then* $\|A(x)-\bar{A}(x)\| \leqq \delta$ *for all* $x \in \Sigma^*$.

PROOF. By the previous theorem, there is ε_1 such that $\|A(\sigma)-\bar{A}(\sigma)\| \leqq \varepsilon_1$, for all $\sigma \in \Sigma$, implies that both systems are weakly ergodic. Thus there is ε_1 such that there is n_0 with both $\delta(A(x)) \leqq \delta/6$ and $\delta(\bar{A}(x)) \leqq \delta/6$ for all x with $l(x) \geqq n_0$ and the given δ. For the number n_0 above there is ε_2 such that if $\|A(\sigma)-\bar{A}(\sigma)\| \leqq \varepsilon_2$ then $\|A(x)-\bar{A}(x)\| \leqq \delta/6$ for all x with $l(x) < n_0$ (this follows from Lemma 5). Let $\varepsilon = \min(\varepsilon_1, \varepsilon_2)$. Then for all x with $l(x) < n_0$, $\|A(x)-\bar{A}(x)\| \leqq \frac{1}{3}\delta < \delta$. If $x = yz$ with $l(z) = n_0$ and $l(y) \geqq 0$ i.e., if $l(x) \geqq n_0$ then, using Corollary 2 we have that:

$$\begin{aligned}\|A(x)-\bar{A}(x)\| &= \|A(yz)-\bar{A}(yz)\| \leqq \|A(y)A(z)-A(z)\|+\|\bar{A}(y)\bar{A}(z)-\bar{A}(z)\| \\ &\quad +\|A(z)-\bar{A}(z)\| \leqq 2\delta(A(z))+2\delta(\bar{A}(z))+\|A(z)-\bar{A}(z)\| \leqq \tfrac{1}{3}\delta+\tfrac{1}{3}\delta+\tfrac{1}{3}\delta = \delta. \quad \square\end{aligned}$$

REMARK. Theorem 8 provides an interesting application. Assume that a PIT $(\Sigma, S, \{A(\sigma)\})$ is given together with an initial distribution π over the states and it is required to compute the values of the vector $\pi A(x)$ for several words x. If the number of states is countable infinite then it will be impossible to compute the exact values of the entries of $\pi A(x)$. Assume now that $\pi A(x) = \pi A(\sigma_1)A(\sigma_2)\cdots A(\sigma_k)$ and assume that the system is weakly ergodic. Furthermore let $\delta > 0$ be a real number and let $\varepsilon = \varepsilon(\delta)$ be the ε related to that δ as in Theorem 8. Then one can find a new vector π' having only finitely many non-zero entries, say the first n ones, and such that $\|\pi-\pi'\| \leqq \varepsilon$. The product $\pi' A(\sigma_1)$ involves only the first n rows of $A(\sigma_1)$ and therefore one can find another matrix $A'(\sigma_1)$ such that its first n rows have non-zero entries only in the first n_2 columns and such that $\|A(\sigma_1)-A'(\sigma_1)\| < \varepsilon$. Proceeding this way one can replace the product $\pi A(\sigma_1)\cdots A(\sigma_k)$ by the product $\pi' A'(\sigma_1)A'(\sigma_2)\cdots A'(\sigma_k)$ such that the second product involves only finitely many arithmetical operations and $\|\pi-\pi'\| \leqq \varepsilon, \|A'(\sigma_1)-A(\sigma_1)\| \leqq \varepsilon \cdots \|A'(\sigma_k)-A(\sigma_k)\| \leqq \varepsilon$, this implying that $\|\pi A(x)-\pi' A'(x)\| \leqq \delta$. An infinite computation can thus be replaced by a finite computation and the resulting error can be kept under control. Theorem 8 may also be used for rounding off the entries in the individual matrices $A(\sigma)$ (in order to simplify the computation, or to make computation possible when the entries are irrational) and keeping the resulting error in long computations under control.

Because of the importance of Theorem 8 one is induced to ask whether the condition of that theorem is best (i.e., whether it is also a necessary condition for the theorem to hold true). The answer to this question is negative and the reader is referred to ([8] page 53) for full comments on this topic. It is also shown, however, in the above mentioned reference that Theorem 8 is not true in general (i.e., for PIT with no conditions imposed to them), and the findings of a necessary and sufficient condition for the consequences of Theorem 8 to hold true is still an open problem even in the finite case.

REFERENCES

[1] Dobrushin, R. L. (1956). Central limit theorem for nonstationary Markov chains I, II. *Theor. Probability Appl.* **1** 65–80 and 329–383, (English translation).

[2] Hajnal, J. (1956). The ergodic properties of nonhomogeneous finite Markov chains. *Proc. Cambridge Philos. Soc.* **52** 67–77.

[3] Hajnal, J. (1958). Weak ergodicity in nonhomogeneous Markov-chains. *Proc. Cambridge Philos. Soc.*, **54** 233–246.

[4] Kozniewska, J. (1958). Ergodicity of nonhomogeneous Markov chains with two-states. *Colloq. Math.* **5** 208–215.

[5] Kozniewska, J. (1962). Ergodicité et stationnarité des chaines de Markoff variables a un nombre fini d'etats possibles. *Colloq. Math.* **9** 333–346.

[6] Paz, A. (1965). Definite and quasidefinite sets of stochastic matrices. *Proc. Amer. Math. Soc* **16** 634–641.

[7] Paz, A. (1966). Some aspects of probabilistic automata. *Information and Control*, **9** 26–60.

[8] Paz, A. and Reichaw, M. (1967). Ergodic theorems for sequences of infinite stochastic matrices. *Proc. Cambridge Philos. Soc.* **63** 777–784.

[9] Rabin, M. O. (1964). Probabilistic automata. *Sequential Machines*, selected papers, ed. E. F. Moore. Addison-Wesley, Reading, 98–114.

[10] Sarymsakov, T. A. (1961). Inhomogeneous Markov chains. *Theor. Probability Appl.* **6** 196–201.

[11] Wolfowitz, J. (1963). Products of indecomposable, aperiodic, stochastic matrices. *Proc Amer. Math. Soc.* **14** 733–737.