The Annals of Probability
1973, Vol. 1, No. 2, 329–335

STRONGLY ERGODIC BEHAVIOR FOR NON-STATIONARY MARKOV PROCESSES

BY RICHARD W. MADSEN AND DEAN L. ISAACSON

University of Missouri-Columbia and Iowa State University

This paper considers ergodic behavior of those non-stationary Markov processes which can be represented by a sequence of stochastic kernels, $\{P_n(x, y)\}$, defined on a σ-finite measure space $(S, \mathscr{F}, \mu)$. In particular, the convergence of the superpositions, $P_1 P_2 P_3 \cdots P_n$, of these kernels is related to the convergence of their corresponding left eigenfunctions, ψ_n, where $\psi_n(y) = \int \psi_n(x) P_n(x, y)\mu(dx)$ and $\int \psi_n(y)\mu(dy) = 1$.

It is then shown how these results can easily be extended to the general case where densities are not assumed.

0. Introduction and summary. It is well known that under certain conditions, powers of a stochastic matrix P will converge to a matrix, say Q, which has all of its rows the same. (See, for example, Feller [4].) Such a matrix is called ergodic. It is also true that the rows of Q are left eigenvectors of P corresponding to the eigenvalue 1. Hence, in the stationary matrix case, there is a relationship between ergodic behavior and left eigenvectors.

In generalizing from the stationary case of ergodic behavior to the non-stationary case, new questions arise. First of all we must distinguish between weak and strong ergodicity, (Definitions 1.2 and 1.3). Secondly, since each P_n in a non-stationary sequence of stochastic matrices has its own left eigenvector, ψ_n, corresponding to $\lambda = 1$, we ask whether or not the convergence of the $\{\psi_n\}$ is related to strong ergodicity of the sequence. Rather than restricting this discussion to stochastic matrices and eigenvectors we will give all results in terms of stochastic kernels and eigenfunctions (see [6]). We give conditions under which convergence of the left eigenfunctions implies strong ergodicity. We also give conditions under which strong ergodicity implies convergence of the left eigenfunctions. Finally we show how these results can easily be extended to the case where densities may not exist.

1. Assumptions and basic definitions. We consider a σ-finite measure space $(S, \mathscr{F}, \mu)$ and a sequence of stochastic kernels $\{P_n(x, y)\}$ defined on $S \times S$ which are sufficiently well-behaved so that superpositions defined by

$$P_{n,n+m}(x, y) = \int_S \cdots \int_S P_n(x, z_1)P_{n+1}(z_1, z_2) \cdots P_{n+m}(z_m, y)\mu(dz_1) \cdots \mu(dz_m)$$

exist for all m and n. We assume the above to be true throughout this paper.

Received July 24, 1972.

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

Key words and phrases. Weak and strong ergodicity, left eigenfunctions, superposition, stochastic kernel, ergodic coefficient.

We will use the following norm for arbitrary integrable kernels defined on $S \times S$:

$$||K|| = \sup_x \int |K(x, y)|\mu(dy) ,$$

in which we introduce the notation that if the range of integration is unspecified, we take it to be S.

For two kernels K and L we define their superposition, "KL," by $KL(x, y) = \int K(x, z)L(z, y)\mu(dz)$.

In our case, where we assume the existence of kernels, the ergodic coefficient of Dobrushin [3] takes the following form.

DEFINITION 1.1. If $P(x, y)$ is a stochastic kernel, then

$$\begin{aligned}\alpha(P) &= 1 - \sup_{x,z} \int [P(x, y) - P(z, y)]^+\mu(dy) \\ &= 1 - \tfrac{1}{2} \sup_{x,z} \int |P(x, y) - P(z, y)|\mu(dy) .\end{aligned}$$

For convenience we define $\delta(P) = 1 - \alpha(P)$. For properties of $\delta(P)$ and the norm $|| \ ||$, the reader is referred to [6]. The following definitions of weakly and strongly ergodic behavior are also given in [5], but are given here for convenience.

DEFINITION 1.2. A sequence of stochastic kernels $\{P_n\}$ is said to be weakly ergodic if $\delta(P_{m,n}) \rightarrow_n 0$ for all m.

DEFINITION 1.3. A sequence of stochastic kernels $\{P_n\}$ is said to be strongly ergodic if there exists a kernel $Q(x, y)$ with the property that $Q(x, y) = Q(z, y)$ for all x, y and z and $||P_{m,n} - Q|| \rightarrow_n 0$ for all m.

We now state two lemmas which follow from two inequalities of Blum and Reichaw [1].

LEMMA 1. *If* $\int R(x, y)\mu(dy) = 0$ *for all* x *and if* P *is stochastic, then* $||RP|| \leqq ||R||\delta(P)$.

LEMMA 2. *If* P *and* Q *are stochastic kernels, then* $\delta(PQ) \leqq \delta(P)\delta(Q)$. (This is also done by Dobrushin [3].)

The final definition of this section is given for notational purposes.

DEFINITION 1.4. Let $\mathscr{A}$ denote the class of kernels, P, for which the eigenvalue 1 has a nonnegative, integrable left eigenfunction, ψ. Take ψ to be normalized so that $\int \psi(y)\mu(dy) = 1$.

Given a stochastic kernel, P, in $\mathscr{A}$ we will take such a left eigenfunction as the left eigenfunction associated with P.

2. Conditions for strongly ergodic behavior. The following theorems give conditions under which convergence of left eigenfunctions implies strongly ergodic behavior.

THEOREM 2.1. *Let* $\{P_n(x, y)\}$ *be a weakly ergodic sequence of stochastic kernels in* $\mathscr{A}$. *If the corresponding left eigenfunctions,* ψ_n, *satisfy*

$$\sum_{j=1}^{\infty} ||\psi_{j+1}(y) - \psi_j(y)|| < \infty \tag{2.1}$$

then $\{P_n(x, y)\}$ *is strongly ergodic.*

PROOF. Since $L_1(\mu)$ is complete, let $||\psi_n(y) - \psi(y)|| \to 0$ as $n \to \infty$. Define $Q(x, y) = \psi(y)$ for all x. Then

$$(2.2)\qquad \begin{aligned} ||P_{m,n} - Q|| &\leqq ||P_{m,n} - \psi_n|| + ||\psi_n - \psi|| \\ &\leqq ||P_{m,k-1}P_{k,n} - \psi_k P_{k,n}|| + ||\psi_k P_{k,n} - \psi_n|| + ||\psi_n - \psi|| . \end{aligned}$$

By choosing $n \geqq N_1(\varepsilon)$, the third term of (2.2) can be made less than $\varepsilon/3$. Now consider the second term of (2.2). Since

$$\begin{aligned} \psi_k P_{k,n} &= \psi_k P_k P_{k+1,n} = (\psi_k P_{k+1,n} - \psi_{k+1} P_{k+1,n}) + \psi_{k+1} P_{k+1,n} \\ &= \cdots = (\psi_k - \psi_{k+1})P_{k+1,n} + (\psi_{k+1} - \psi_{k+2})P_{k+2,n} + \cdots + \psi_n P_n , \end{aligned}$$

it follows that

$$\begin{aligned} ||\psi_k P_{k,n} - \psi_n|| &= ||\textstyle\sum_{j=k}^{n-1} (\psi_j - \psi_{j+1})P_{j+1,n}|| \leqq \sum_{j=k}^{n-1} ||\psi_j - \psi_{j+1}||\delta(P_{j+1,n}) \\ &\leqq \textstyle\sum_{j=k}^{n-1} ||\psi_j - \psi_{j+1}|| \leqq \sum_{j=k}^{\infty} ||\psi_j - \psi_{j+1}|| . \end{aligned}$$

It follows from (2.1) that for $k \geqq K(\varepsilon)$, this last expression can be made less than $\varepsilon/3$.

Finally, since $\int P_{m,k-1}(x, y)\mu(dy) = \int \psi_k(y)\mu(dy) = 1$ for all x and k, Lemma 1 implies that

$$||P_{m,k-1}P_{k,n} - \psi_k P_{k,n}|| \leqq ||P_{m,k-1} - \psi_k||\delta(P_{k,n})$$

which for fixed k can be made less than $\varepsilon/3$ for $n \geqq N_2(\varepsilon)$ by the assumption of weak ergodicity. Hence (2.2) can be made less than ε for all n sufficiently large. □

THEOREM 2.2. *Let $\{P_n(x, y)\}$ be a sequence of stochastic kernels in $\mathscr{A}$. If the eigenfunctions converge in the sense that $||\psi_n - \psi|| \to 0$, and if there exists a constant D such that*

$$\textstyle\sum_{j=1}^{n} \delta(P_{j,n}) \leqq D \qquad \textit{for all } n$$

then $\{P_n(x, y)\}$ is strongly ergodic.

PROOF. First note that for all m, $\sum_{j=m}^{n} \delta(P_{j,n}) \leqq \sum_{j=1}^{n} \delta(P_{j,n}) \leqq D$. It is not hard to show that this implies that $\delta(P_{m,n}) \to_n 0$ for all m.

We now show that for all n sufficiently large, $||P_{m,n} - Q|| < \varepsilon$, where $Q(x, y) = \psi(y)$ for all x. Now

$$(23)\qquad ||P_{m,n} - Q|| \leqq ||P_{m,n} - \psi_n|| + ||\psi_n - \psi||$$

and clearly $||\psi_n - \psi||$ can be made less than $\varepsilon/2$ for $n \geqq N_1(\varepsilon)$. Next note that Lemma 1 can be applied in the following.

$$\begin{aligned} ||P_{m,n} - \psi_n|| &= ||P_{m,n-1}P_n - \psi_n P_n|| \\ &\leqq ||P_{m,n-1}P_n - \psi_{n-1}P_n|| + ||(\psi_{n-1} - \psi_n)P_n|| \\ &\leqq ||P_{m,n-2}P_{n-1,n} - \psi_{n-1}P_{n-1,n}|| + ||\psi_{n-1} - \psi_n||\delta(P_n) . \end{aligned}$$

Continuing in this way we can show that

$$(2.4)\qquad ||P_{m,n} - \psi_n|| \leqq ||P_m - \psi_m||\delta(P_{m+1,n}) + \textstyle\sum_{j=m+1}^{n} ||\psi_{j-1} - \psi_j||\delta(P_{j,n}) .$$

Since $\delta(P_{m+1,n}) \to_n 0$, the first term of the right-hand side of (2.4) can be made less than $\varepsilon/4$ for $n \geqq N_2(\varepsilon)$.

Now since $||\psi_{j-1} - \psi_j|| \to 0$, given $\varepsilon > 0$, there exists $M(\varepsilon)$ such that for $j \geqq M$, $||\psi_{j-1} - \psi_j|| \leqq \varepsilon/(8D) = \gamma$. Assume $M > m$. (The case where $M \leqq m$ is easier.) Then, for all n,

$$\sum_{j=M+1}^{n} ||\psi_{j-1} - \psi_j||\delta(P_{j,n}) \leqq \gamma \sum_{j=M+1}^{n} \delta(P_{j,n}) \leqq \varepsilon/8 .$$

Finally, since $||\psi_{j-1} - \psi_j|| \leqq 2$, and since given M, there exists $N_3(\varepsilon)$ such that for $n \geqq N_3$ and all $j \leqq M$,

$$\delta(P_{j,n}) \leqq \delta(P_{M,n}) \leqq \varepsilon/[16(M - m)] .$$

Hence $\sum_{j=m+1}^{M} ||\psi_{j-1} - \psi_j||\delta(P_{j,n}) \leqq 2 \sum_{j=m+1}^{M} \delta(P_{j,n}) \leqq \varepsilon/8$. Therefore, for $n \geqq \max\{N_1, N_2, N_3\}$, $||P_{m,n} - Q|| < \varepsilon$. □

COROLLARY 2.1. *If $\{P_n\}$ in $\mathscr{A}$ is stationary weakly ergodic, then $\{P_n\}$ is strongly ergodic.*

PROOF. This follows immediately from Theorem 2.1. Also since stationary weakly ergodic implies that for some β, $\delta(P^r) \leqq \beta > 1$ for some $r \geqq 1$, this corollary follows from Theorem 2.2 as well.

COROLLARY 2.2. *Let $\{P_n\}$ be a sequence of stochastic kernels in $\mathscr{A}$. If $||\psi_n - \psi|| \to 0$ as $n \to \infty$ and if for some β, $\delta(P_n) \leqq \beta < 1$ for all n, then $\{P_n\}$ is strongly ergodic.*

PROOF. Straightforward using Lemma 2 and Theorem 2.2.

3. Conditions for convergence of left eigenfunctions. The following theorem gives conditions under which strongly ergodic behavior implies convergence of left eigenfunctions.

THEOREM 3.1. *Let $\{P_n(x, y)\}$ be a strongly ergodic sequence of stochastic kernels in $\mathscr{A}$. If there exists an integer, k, and a real number, β, such that*

$$\delta(P_n^{\,k}) \leqq \beta < 1$$

for all n, then $\psi_n(y)$ converges (in norm) to $Q(x, y) = \lim_{n\to\infty} P_{1,n}(x, y)$.

Note: $\delta(P_n^{\,k}) < 1$ implies P_n has a unique left eigenfunction corresponding to $\lambda = 1$ with $\int \psi_n(y)\mu(dy) = 1$

PROOF. Define

$$E_n(x, y) = P_{1,n}(x, y) - P_{1,n-1}(x, y)$$

and

$$\Delta_n(x, y) = \psi_n(y) - P_{1,n-1}(x, y) .$$

It follows by strong ergodicity that $||E_n|| \to_n 0$.

Since $||P_{1,n-1} - Q|| \to_n 0$ by strong ergodicity, it suffices for us to show that $||\psi_n - P_{1,n-1}|| = ||\Delta_n|| \to_n 0$.

First note that

$$\int \Delta_n(x, y)P_n(y, z)\mu(dy) = \psi_n(z) - P_{1,n}(x, z) = \Delta_n(x, z) - E_n(x, z) ,$$

and

$$\int [\int \Delta_n(x, w)P_n(w, y)\mu(dw) + E_n(x, y)]P_n(y, z)\mu(dy) + E_n(x, z) = \Delta_n(x, z) ,$$

i.e.

$$\Delta_n(x, z) = E_n(x, z) + \int E_n(x, y)P_n(y, z)\mu(dy) + \int\int \Delta_n(x, w)P_n(w, y)P_n(y, z)\mu(dw)\mu(dy) .$$

Now Fubini's theorem can be applied to this expression to give

$$\Delta_n(x, z) = E_n(x, z) + \int E_n(x, y)P_n(y, z)\mu(dy) + \int \Delta_n(x, w)P^2(w, z)\mu(dw) .$$

If iteration is continued in this way, we find

$$\Delta_n(x, z) = E_n(x, z) + \sum_{j=1}^{M-1} \int E_n(x, y)P_n{}^j(y, z)\mu(dy) + \int \Delta_n(x, y)P_n{}^M(y, z)\mu(dy) .$$

Hence,

$$||\Delta_n|| \leqq ||E_n|| + \sum_{j=1}^{M-1} ||E_n P_n{}^j|| + ||\Delta_n P_n{}^M|| .$$

Now since $\int E_n(x, y)\mu(dy) = \int \Delta_n(x, y)\mu(dy) = 0$, it follows from Lemma 1 that $||E_n P_n{}^j|| \leqq ||E_n||\delta(P_n{}^j)$ and $||\Delta_n P_n{}^M|| \leqq 2\delta(P_n{}^M)$. Hence

(3.1) $$||\Delta_n|| \leqq ||E_n||\{1 + \sum_{j=1}^{M-1} \delta(P_n{}^j)\} + 2\delta(P_n{}^M) .$$

Now

$$\sum_{j=1}^{M-1} \delta(P_n{}^j) \leqq \sum_{j=1}^{\infty} \delta(P_n{}^j) = \sum_{j=1}^{k} \delta(P_n{}^j) + \sum_{j=k+1}^{2k} \delta(P_n{}^j) + \cdots \leqq k + k\beta + \cdots = k/(1 - \beta) .$$

Therefore, since $||E_n|| \to_n 0$, the first term of the right-hand side of (3.1) can be made less than $\varepsilon/2$ for $n \geqq N_1(\varepsilon)$. It is also true that $\delta(P_n{}^M)$ will be less than $\varepsilon/2$ for M sufficiently large. □

COROLLARY 3.1. *Let $\{P_n\}$ be any sequence of stochastic kernels in $\mathscr{A}$ such that for some β, $\delta(P_n) \leqq \beta < 1$. Then $\{P_n\}$ is strongly ergodic if and only if $\{\psi_n(y)\}$ converges.*

PROOF. Use the fact that Theorem 3.1 holds when $k = 1$ and for the converse use Corollary 2.2.

This corollary generalizes a result obtained by Conn [2] where she assumes $S = [a, b]$ and $0 < m \leqq P_n(x, y) \leqq M < \infty$.

4. The general case. The authors are grateful to the referee for pointing out that the above results extend easily to the case where densities for the transition probability functions do not exist. This type of extension is done in [5] and the reader is referred to this article for some of the details. Rather than reprove the above theorems, we will simply state them and mention some of the changes needed to cover this more general situation. In this case we consider transition probability functions, $P(x, B)$ where $P(x, \cdot)$ is a probability on $\mathscr{F}$ for each x and $P(\cdot, B)$ is an $\mathscr{F}$ measurable function for each $B \in \mathscr{F}$. There is no problem defining the ergodic coefficient for such a transition function since this was done by Dobrushin [3]. In particular, $\alpha(P) = 1 - \sup|P(x', B) - P(x'', B)|$ where the sup is taken over all $x', x'' \in S$ and $B \in \mathscr{F}$. Rather than defining the norm of a kernel as $\sup_x \int |k(x, y)|\mu(dy)$ we define the norm of a transition function, $k(x, B)$ to be $||k|| = \sup_x$ [total variation of $k(x, \cdot)$]. The composition of transition probability functions is given by $P_{k,k+m}(x, B) = \int \cdots \int P_k(x, dx_{k+1}) P_{k+1}(x_{k+1}, dx_{k+2}) \cdots P_{k+m}(x_{k+m}, B)$. In this case Lemma 1 says if $k(x, A)$ is a signed transition function with $k(x, S) = 0$ for all x then $||kP|| \leqq ||k|| \cdot \delta(P)$ where $kP(x, A) = \int_S k(x, dy)P(y, A)$. Lemma 2 is the same and is given in [5] for this case. We now consider the class, $\mathscr{A}$, of transition probability functions, P_n, which have an invariant probability measure, μ_n. (i.e., $\mu_n(S) = 1$ and $\mu_n(B) = \int \mu_n(dx)P_n(x, B)$.) Now a sequence of transition probability functions, $\{P_n\}$ will be called strongly ergodic if there exists a probability measure μ such that $||P_{m,n} - \mu|| \to 0$ as $n \to \infty$ for all m.

THEOREM 2.1′. *Let $\{P_n(x, B)\}$ be a weakly ergodic sequence of transition probability functions in $\mathscr{A}$. If the corresponding invariant probability measures, μ_n, satisfy $\sum_{j=1}^{\infty} ||\mu_{j+1} - \mu_j|| < \infty$, then $\{P_n(x, B)\}$ is strongly ergodic.*

PROOF. The proof follows that of Theorem 2.1 except that in this case one uses the completeness of the space of signed measures with variation norm.

THEOREM 2.2′. *Let $\{P_n(x, B)\}$ be a sequence of transition probability functions in $\mathscr{A}$. If the invariant probability measures converge in the sense that $||\mu_n - \mu|| \to 0$ and if there exists a constant D such that $\sum_{j=1}^{n} \delta(P_{j,n}) \leqq D$ for all n, then $\{P_n(x, B)\}$ is strongly ergodic.*

PROOF. The proof follows that of Theorem 2.2 with measures replacing eigenfunctions.

THEOREM 3.1′. *Let $\{P_n(x, B)\}$ be a strongly ergodic sequence of transition probability functions in $\mathscr{A}$. If there exists an integer, k, and a real number, β, such that $\delta(P_n{}^k) \leqq \beta < 1$ for all n, then μ_n converges in norm to $\mu = \lim_{n\to\infty} P_{1,n}(x, \cdot)$.*

PROOF. The proof follows that of Theorem 3.1. In order to show

$$\int\int \Delta_n(x, dw)P_n(w, dy)P_n(y, B) = \int \Delta_n(x, dw)P_n{}^2(w, B)$$

start with $P_n(y, B)$ an indicator function and then use the standard arguments to get the result for any nonnegative measurable function $P_n(y, B)$.

Acknowledgments. The authors wish to thank H. T. David and B. C. Arnold for some helpful discussions on this problem. The authors also thank the referee for his comments.

REFERENCES

[1] BLUM, J. R. and REICHAW, M. (1971). Two integral inequalities. *Israel J. Math.* **9** 20–26.
[2] CONN, PATRICIA (1969). Ph. D. Dissertation, Iowa State Univ.
[3] DOBRUSHIN, R. L. (1956). Central limit theorems for non-stationary Markov chains I, II. *Theor. Probability Appl.* **1** 65–80 and 329–383 (Eng. tr.).

[4] FELLER, WILLIAM (1968). *An Introduction to Probability Theory and Its Applications*, 3rd ed. Wiley, New York.

[5] IOSIFESCU, MARIUS (1972). Ergodicity in nonhomogeneous Markov chains. *Ann. Math. Statist.* **43** 1732–1736.

[6] MADSEN, RICHARD W. (1971). A note on some ergodic Theorems of A. Paz. *Ann. Math. Statist.* **42** 405–408.

DEPARTMENT OF STATISTICS
SNEDECOR HALL
IOWA STATE UNIVERSITY
AMES, IOWA 50010