1971

Asymptotic properties of super-positions of non-negative kernels

Richard William Madsen

Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/rtd

Part of the Statistics and Probability Commons

Recommended Citation

https://lib.dr.iastate.edu/rtd/4479

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Retrospective Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
MADSEN, Richard William, 1941-
ASYMPTOTIC PROPERTIES OF SUPERPOSITIONS OF
NON-NEGATIVE KERNELS.

Iowa State University, Ph.D., 1971
Statistics

University Microfilms, A XEROX Company, Ann Arbor, Michigan
Asymptotic properties of superpositions of non-negative kernels

by

Richard William Madsen

A Dissertation Submitted to the Graduate Faculty in Partial Fulfillment of The Requirements for the Degree of DOCTOR OF PHILOSOPHY

Major Subject: Statistics

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

For the Major Department

Signature was redacted for privacy.

For the Graduate College

Iowa State University
Ames, Iowa
1971
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>A. Statement of the Problem</td>
<td>1</td>
</tr>
<tr>
<td>B. Literature Review</td>
<td>6</td>
</tr>
<tr>
<td>II. PROPERTIES OF STOCHASTIC KERNELS</td>
<td>14</td>
</tr>
<tr>
<td>A. Properties of the Ergodic Coefficient</td>
<td>14</td>
</tr>
<tr>
<td>B. Types of Ergodic Behavior</td>
<td>17</td>
</tr>
<tr>
<td>C. Equivalence of Two Definitions of Weak and Strong Ergodicity</td>
<td>28</td>
</tr>
<tr>
<td>III. L₁ BEHAVIOR OF NON-NEGATIVE KERNELS</td>
<td>33</td>
</tr>
<tr>
<td>A. Normalization and Transformations</td>
<td>33</td>
</tr>
<tr>
<td>B. Weak Ergodicity Using Eigenfunctions</td>
<td>37</td>
</tr>
<tr>
<td>C. Weak Ergodicity Without Using Eigenfunctions</td>
<td>54</td>
</tr>
<tr>
<td>IV. POINTWISE BEHAVIOR OF NON-NEGATIVE KERNELS</td>
<td>58</td>
</tr>
<tr>
<td>A. Weak Ergodicity</td>
<td>58</td>
</tr>
<tr>
<td>B. Strong Ergodicity</td>
<td>72</td>
</tr>
<tr>
<td>C. Convergence Theorems for Eigenvalues and Eigenfunctions</td>
<td>92</td>
</tr>
<tr>
<td>V. APPLICATIONS TO STATISTICS</td>
<td>100</td>
</tr>
<tr>
<td>A. Sequential Probability Ratio Tests</td>
<td>100</td>
</tr>
<tr>
<td>B. Generalized Sequential Probability Ratio Tests</td>
<td>110</td>
</tr>
<tr>
<td>VI. LITERATURE CITED</td>
<td>115</td>
</tr>
<tr>
<td>VII. ACKNOWLEDGMENTS</td>
<td>117</td>
</tr>
</tbody>
</table>
I. INTRODUCTION

A. Statement of the Problem

This thesis is a study of properties of kernel superpositions by the use of the ergodic coefficient introduced by Dobrushin (1956). Patricia Conn (1969) also studied kernel superpositions, but the use of the ergodic coefficient allows many of the same results to be obtained under conditions weaker than she imposed. Conditions for types of asymptotic behavior different than those considered by Conn are also given.

The kind of asymptotic behavior in which we are interested is related to that studied in Volume I of Feller (1968) or in other texts of a similar nature. If \( P \) is a transition matrix for a Markov chain defined on a finite state space and if \( a_0 = (a_{01}, a_{02}, \ldots, a_{0N}) \) is a probability distribution over the states \( E_1, E_2, \ldots, E_N \), then one is interested in the behavior of \( a_n \) defined by

\[
a_n = a_{n-1}P = a_0P^n.
\]

It is well known that if the matrix \( P \) is irreducible and aperiodic, then there exist numbers \( u_k > 0 \) satisfying

\[
\sum_{k=1}^{N} u_k = 1 \quad \text{and} \quad u_k = \sum_{j=1}^{N} u_j P_{jk}.
\]

The vector \( u = (u_1, u_2, \ldots, u_N) \) is called an invariant probability measure and is, in fact, a left eigenvector of

...
the matrix $P$ corresponding to the eigenvalue 1. Further, the vector $a_n$ converges to the vector $u$ componentwise, i.e.,

$$a_{nk} \rightarrow u_k.$$  

This convergence is independent of the initial probability distribution $a_0$.

In general, if the asymptotic behavior of the probability vector $a_n$ is independent of the choice of the initial probability distribution $a_0$, the behavior will be called weakly ergodic, or equivalently, the chain will be called weakly ergodic. If, in addition to being weakly ergodic, the sequence $\{a_n\}$ converges to some vector, the behavior will be called strongly ergodic. Since, when weak ergodicity holds, the effect of the initial probability distribution is lost, we could say that "memory" has been lost. Using this terminology, if $P$ were irreducible and aperiodic we would say the Markov chain is strongly ergodic, or loses memory and converges.

The situation studied by Feller can be generalized in several ways. First, one could use a different transition matrix at each step. That is, the transition probabilities for the transition from state $j$ at time $n-1$ to state $k$ at time $n$ would be given by $P_{jk}^n$. A Markov chain determined in this way is said to be non-stationary or non-homogeneous.
In this case, one would consider products \( P_1 \cdot P_2 \cdots \cdot P_n \) rather than powers of one matrix, \( P^n \). One must also consider the possible effects of starting with an initial probability vector at time \( m-1 \) rather than at time zero. In this case, the relevant matrix products would be

\[
P_m \cdot P_{m+1} \cdots \cdot P_{m+n}.
\]

A second generalization would be to study Markov "chains" which are defined on an arbitrary state space rather than on a discrete state space. In this case, the term Markov "process" is usually used instead of "chain". Given a measure space \((S, \mathcal{B}, \mu)\) and a Markov process defined on that space, the transition probability function \( P(x,A) \) plays the role that the transition matrix played in the discrete state space case. In this work, we will consider only those processes for which a stochastic transition kernel \( P(x,y) \) exists. In order to be a stochastic transition kernel, \( P(x,y) \) must, by definition, satisfy

\[
P(x,y) \geq 0 \quad \text{for all } x, y,
\]

\[
\int_S P(x,y) \mu(dy) = 1 \quad \text{for all } x.
\]

Further, the relationship of the stochastic kernel to the probability function is that for all sets \( A \) in the \( \sigma \)-algebra \( \mathcal{B} \),
\[ P(x,A) = \int_A P(x,y) \mu(dy) . \]

With an arbitrary state space, the non-stationary case can also be studied. Instead of the matrix product as in (1.1.1), one would consider kernel superpositions defined by

\[ P_{m,m+n}(x,y) = \int \ldots \int P_m(x,z_1) P_{m+1}(z_1,z_2) \ldots P_{m+n}(z_n,y) \mu(dz_1) \ldots \mu(dz_n). \]

Note that, unless otherwise specified, the range of integration is to be over the whole space \( S \).

A final generalization would be to remove the requirement that \( P_n(x,y) \) be a stochastic kernel and study the asymptotic behavior of superpositions of non-negative kernels. Throughout this paper, we consider only those non-negative kernels \( M_n(x,y) \) which are measurable kernels defined on \( S \times S \) into the reals, where \((S, \mathcal{B}, \mu)\) is a \( \sigma \)-finite measure space, and for which superpositions \( M_{m,m+n}(x,y) \) defined by

\[ M_{m,m+n}(x,y) = \int \ldots \int M_m(x,z_1) M_{m+1}(z_1,z_2) \ldots M_{m+n}(z_n,y) \mu(dz_1) \ldots \mu(dz_n) \]  

exist for all \( m \) and \( n \).

In particular, we consider starting functions or initial
functions $f_0(x)$ which are non-negative functions satisfying
\[ 0 < \int f_0(x) \mu(dx) < \infty . \] (1.1.2)

These functions are analogous to the initial vectors $a_0$ defined earlier, and, in that same spirit, we consider the asymptotic behavior of
\[ f_n(y) = \int f_{n-1}(x) M_n(x,y) \mu(dx) . \] (1.1.3)

In addition to the assumption of the existence of superpositions of kernels as in (1.1.2), we assume throughout this paper that the kernels are sufficiently well-behaved to assure that for all non-negative functions $f(x)$ in $L_1(\mu)$, the function $g(y)$ defined by
\[ g(y) = \int f(x) M_n(x,y) \mu(dx) \]
is integrable for all $n$. Note that a sufficient condition for this to hold is that for each $n$, there exists a finite $B_n$ such that
\[ M_n(x,S) = \int M_n(x,y) \mu(dy) \leq B_n . \]

Note also that for stochastic or substochastic kernels, the number 1, is such a bound.

With suitable modifications, weakly and strongly ergodic behavior can be defined in the general case. The sense in which the modified functions $\{f_n(x)\}$ lose memory or converge
can, for example, be in $L_1(\mu)$ or pointwise. Sufficient conditions for behavior in the former sense are given in Chapter III and for behavior in the latter sense, in Chapter IV.

B. Literature Review

Harris (1963) considered ergodic behavior in the stationary case. A kernel $M(x,y)$ defined on a space $S$ of finite measure ($\mu(S) < \infty$) is said to be primitive if $M(x,S) = \int M(x,y)\mu(dy)$ is a bounded function of $x$ and if there exist constants $c$ and $d$ and an integer $n$ such that

$$0 < c \leq M^n(x,y) \leq d < \infty.$$

Harris gave the following theorem: If $M$ is primitive, $M$ has a positive eigenvalue $\lambda$, larger in magnitude than any other eigenvalue, which corresponds to right and left eigenfunctions $\phi(x)$ and $\psi(y)$ which are bounded above and below by positive constants. Furthermore, if $\phi$ and $\psi$ are normalized so that

$$\int \phi(x)\psi(x)\mu(dx) = 1 \quad \text{and} \quad \int \psi(x)\mu(dx) = 1$$

then

$$M^n(x,y) = \lambda^n \phi(x)\psi(y)[1 + O(\Delta^n)], \quad 0 < \Delta < 1, \quad n \to \infty \ (1.2.1)$$
where the bound $\Delta$ can be taken independently of $x$ and $y$.

In view of this result and using the sequence $\{f_n(y)\}$ as defined in (1.1.3) for the stationary case, it follows that

$$f_n(y) = \int f_0(x) \mu^n(x,y) \mu(dx) = \lambda^n \psi(y) \int f_0(x) \phi(x) \mu(dx).$$

In Chapter III we give some rationale for considering the modified sequence $\{f^*(y)\}$ defined as

$$f^*_n(y) = f_n(y)/\int f_n(y) \mu(dy). \quad (1.2.2)$$

If the modification or normalization is done in this way, then under the condition of primitivity,

$$f^*_n(y) = \frac{\lambda^n \psi(y) \int f_0(x) \phi(x) \mu(dx)}{\int \lambda^n \psi(y) \mu(dy) \int f_0(x) \phi(x) \mu(dx)} = \psi(y)$$

independently of the choice of $f_0$. Hence Harris has answered the question about strongly ergodic behavior in the stationary, primitive case.

Conn (1965) studied kernel superpositions from the non-stationary viewpoint under the following conditions. The space $(S, \mathcal{B}, \mu)$ was taken to be the real interval $[a,b]$, with the $\sigma$-algebra the Lebesgue sets, and $\mu$ Lebesgue measure. The kernels were assumed to be measurable on $[a,b] \times [a,b]$ and uniformly bounded above and below by positive constants. That is, it was assumed that there exist constants $\overline{m}$ and $\overline{M}$ such that for all $n$, 
Conn considered sequences of functions \( \{f_n(y)\} \) defined as in (1.1.3) and normalized these functions as in (1.2.2). She studied the behavior of the sequence \( \{f^*_n(y)\} \) and was able to prove the following results, where \( \phi_n \) and \( \psi_n \) represent positive right and left eigenfunctions of \( M_n(x,y) \):

1. If \( \int_a^b |\psi_n(y)-\psi_{n+1}(y)| \, dy \to 0 \), then \( |f^*_n(y)-\psi_n(y)| \to 0 \).
2. If \( |\psi_n(y)-\psi_{n+1}(y)| \to 0 \), then \( |f^*_n(y)-\psi_n(y)| \to 0 \).
3. If \( \int_a^b |\phi_n(y)-\phi_{n+1}(y)| \, dy \to 0 \), then there exists a sequence of functions \( \{q_n(y)\} \), independent of \( f_0(y) \), such that \( |f^*_n(y)-q_n(y)| \to 0 \).

It is clear that if any of the three hypotheses were satisfied, memory would be lost. From (2), it follows that if the left eigenfunctions converge, \( \{f^*_n(y)\} \) will also converge as well as lose memory.

In Chapter IV, we show that under conditions weaker than those given by Conn, results (1), (2), and (3) will still hold.

Dobrushin (1956) defined the ergodic coefficient \( \alpha \) for stochastic transition functions as
\[ \alpha(P) = 1 - \sup_{x,z,A \in \mathcal{B}} |P(x,A) - P(z,A)|. \]

If the transition function has a density, as we have assumed, then the ergodic coefficient is

\[ \alpha(P) = 1 - \sup_{x,z,A \in \mathcal{B}} \int_A [P(x,y) - P(z,y)] \mu(\text{dy})|. \] (1.2.3)

He defined a non-homogeneous chain to be ergodic if for all \( k \) and all \( x, z \in S \),

\[ \lim_{n \to \infty} |P_{k,n}(x,A) - P_{k,n}(z,A)| = 0 \]

uniformly for \( A \in \mathcal{B} \). He related the ergodic behavior of the Markov chain to the ergodic coefficients of the individual transition density functions making up the chain.

In this paper we use the ergodic coefficient as a tool to obtain conditions under which non-stochastic kernels (more precisely, non-negative kernels) demonstrate weakly or strongly ergodic behavior. Since the ergodic coefficient plays such an important role in this paper, Chapter II contains a discussion of its properties.

Dobrushin also showed that the ergodic coefficient assumes a simpler form when the Markov chain has a countable state space. In this case, the stochastic transition function is an infinite (or finite) stochastic matrix with ergodic coefficient

\[ \alpha(P) = \inf_{j,k} \sum_{m=1}^{\infty} \min(P_{j,m}, P_{k,m}). \] (1.2.4)

Mott (1957) and Hajnal (1958), although apparently un-
aware of Dobrushin's work, both implicitly required conditions in terms of the ergodic coefficient for a non-homogeneous finite Markov chain to be weakly ergodic. Mott stated that a finite chain will be weakly ergodic if at least one column of each $P_j$, not necessarily the same column in each case, has its elements bounded below by a positive constant, $c > 0$, where $c$ is independent of $j$. This clearly implies that the ergodic coefficient is positive, in fact since for each $k$ and $\ell$, 
$$
\sum_{m=1}^{N} \min(p_{k,m} P_{\ell,m}) \geq c,
$$
it follows that $\alpha(P_j) \geq c$ for all $j$.

Hajnal defined a scrambling matrix as follows: Given any two rows, say $i$ and $j$, there exists at least one column, say $k$, such that both $p_{i,k} > 0$ and $p_{j,k} > 0$. In view of Equation (1.2.4), for a finite scrambling matrix, $\alpha(P) > 0$. Hajnal, in fact, defined a measure of "scrambling power" of a matrix by

$$
\{P\} = \min_{i,j} \sum_{k=1}^{N} \min(p_{i,k} P_{j,k})
$$

which is exactly Dobrushin's "$\alpha$" for the finite case!

Hajnal pointed out that to study weak ergodicity for Markov chains, one needs to measure the accumulated scrambling effect on the product
He proved that a finite state non-homogeneous Markov chain is weakly ergodic if and only if there exists an increasing subsequence of integers \( \{i_j\} \) which partitions the individual transition matrices into blocks of matrices

\[
P_{i_j+1,i_{j+1}} = \prod_{k=i_j+1}^{i_{j+1}} P_k, \text{ such that } \sum_{j=1}^{\infty} \{P_{i_j+1,i_{j+1}}\}, \text{ or}
\]

equivalently

\[
\sum_{j=1}^{\infty} \alpha(P_{i_j+1,i_{j+1}}), \text{ diverges.}
\]

Paz (1970) extended the above theorem of Hajnal's to the case of infinite matrices and gave two other equivalent conditions for weak ergodicity. He also gave equivalent conditions for strong ergodicity.

Madsen (1971) gave a generalization of Paz' results to an arbitrary state space, hence these results are applicable when kernels exist. Since these results are used in Chapters III and IV, a more complete discussion of them is given in Chapter II.

When considering ergodic coefficients, it is sometimes convenient to define \( \delta(P) = 1 - \alpha(P) \). It is not hard to show that if \( P \) is stochastic, then

\[
\delta(P) = \sup_{x,z} \int \left[ P(x,y) - P(z,y) \right]^+ \mu(dy). \tag{1.2.5}
\]

Blum and Reichaw used (1.2.5) to define \( \delta \) for arbitrary kernels rather than just stochastic ones. They proved the
following inequality which reduces to Equation (2.1.1) when the kernels are chosen to be stochastic:

\[ \delta(LK) \leq \delta(L)\delta(K) + \bar{\alpha} \inf_{x} \int K(x,y)\mu(dy) \]

where \( L(x,y) \) and \( K(x,y) \) are non-negative, \( LK \) is the superposition of \( L \) with \( K \) and

\[ \bar{\alpha} = \sup_{x,z} \int [L(x,y) - L(z,y)]\mu(dy). \]

Returning to the finite state case, Sarymsakov (1953) obtained the following sufficient condition for weak ergodicity of a non-homogeneous finite Markov chain. Define the class of matrices \( G_2 \) to be:

\[ G_2 = \{ A : A \text{ is } N \times N, \text{ stochastic, and primitive and } AB \text{ is primitive whenever } B \text{ is primitive} \} . \]

Denote the elements of the \( k \)th matrix by \( p_{ij}(k) \) and the non-zero elements by \( p_{ij}^+(k) \). If \( \{P_n\} \) is a sequence of matrices all belonging to \( G_2 \) and \( \min_{1 \leq i, j \leq N} p_{ij}^+(k) \geq \lambda > 0 \) uniformly in \( k \), then the Markov chain determined by \( \{P_n\} \) is weakly ergodic.

It is not hard to see that in the stationary case, primitivity implies that for some \( n \),

\[ \alpha(P^n) = \alpha_0 > 0 , \quad (1.2.6) \]
and, using the results of Hajnal, this implies weak ergodicity. Primitivity is not necessary for (1.2.6) to hold, so we can define a matrix (or a kernel) to be $\alpha$-primitive if for some $n$ (1.2.6) holds.

Note that in the stationary case, a primitive matrix will determine a weakly ergodic chain, but Sarymsakov's work shows that in the non-stationary case each matrix being primitive is not sufficient for weak ergodicity. The same is clearly true if we replace "primitive" with "$\alpha$-primitive". The following, as far as we know, is an unanswered question. What are sufficient conditions for a sequence of $\alpha$-primitive matrices to be weakly ergodic?

Rosenblatt-Roth (1964, 1966) used the ergodic coefficient of Dobrushin to prove limit theorems of a different nature than discussed here. In particular he proved theorems on the strong law of large numbers for non-homogeneous Markov chains.

Dobrushin (1956) used the ergodic coefficient in central limit theorems for non-stationary Markov chains.
II. PROPERTIES OF STOCHASTIC KERNELS

A. Properties of the Ergodic Coefficient

In Chapter I we stated that Dobrushin's ergodic coefficient, defined for stochastic kernels by Equation (1.2.3), would play a significant role in the remainder of this work. The following lemmas give some of the important properties of the ergodic coefficient $\alpha$ and will be used in proving some of the later theorems. These lemmas give known results.

Lemma 2.1.1: If $P(x,y)$ is a stochastic kernel, then

$$\alpha(P) = 1 - \sup_{x,z} \int [P(x,y) - P(z,y)]^+ \mu(dy)$$

$$= 1 - \frac{1}{2} \sup_{x,z} \int |P(x,y) - P(z,y)| \mu(dy) .$$

Proof: In view of Equation (1.2.3), we consider for any $x$ and $z$ fixed, $\sup_A |\int_{A} [P(x,y) - P(z,y)] \mu(dy)|$. Clearly this is the larger of $\int [P(x,y) - P(z,y)]^+ \mu(dy)$ and $\int [P(z,y) - P(x,y)]^+ \mu(dy)$. However, since $P(x,y)$ is stochastic,

$$\int [P(x,y) - P(z,y)] \mu(dy) = 1 - 1 = 0 .$$

Hence

$$\int [P(x,y) - P(z,y)]^+ \mu(dy) = \int [P(x,y) - P(z,y)]^- \mu(dy)$$

$$= \int [P(z,y) - P(x,y)]^+ \mu(dy) .$$
Consequently

\[
\sup_{x,z,A} \left| \int [P(x,y) - P(z,y)] \mu(dy) \right|
\]

\[
= \sup_{x,z} \int [P(x,y) - P(z,y)]^+ \mu(dy)
\]

\[
= \frac{1}{2} \sup_{x,z} \int \left| P(x,y) - P(z,y) \right| \mu(dy).
\]

Lemma 2.1.2: If \( P(x,y) \) is a stochastic kernel, then

\[
0 \leq \alpha(P) \leq 1 \quad \text{and} \quad 0 \leq \delta(P) \leq 1.
\]

Proof: Since \( P(x,y) \) integrates over \( y \) to 1 for each \( x \), it must be that \( 0 \leq \int |P(x,y) - P(z,y)| \mu(dy) \leq 2 \) for all \( x \) and \( z \). Hence from Lemma 2.1.1, \( 0 \leq \alpha(P) \leq 1 \).

Since \( \delta(P) = 1 - \alpha(P) \), it follows that \( 0 \leq \delta(P) \leq 1 \).

Lemma 2.1.3: If \( P \) and \( Q \) are stochastic kernels and if \( PQ(x,y) = \int P(x,z)Q(z,y)\mu(dz) \) is the superposition of \( P \) and \( Q \), then

\[
\delta(PQ) \leq \delta(P)\delta(Q).
\]  \hspace{1cm} (2.1.1)

Proof: Paz and Reichaw (1967) gave a proof of this lemma for the countable matrix case and a very straightforward adaptation can be used for the kernel case. Dobrushin (1956) essentially proved this lemma for kernels, although the proof is not as easy as the one by Paz and Reichaw. A third alternative is the specialization of the inequality
proven by Blum and Reichaw (1971) to the stochastic case.]

A kernel will be called a constant kernel if \( P(x,y) = P(z,y) \) a.e. for all \( z \in S \). It is clear that \( \delta(P) = 0 \) if and only if \( P \) is a constant kernel.

Weakly ergodic behavior for the non-homogeneous case is defined in terms of \( \delta \) as follows.

**Definition 2.1.1:** A non-homogeneous Markov chain, denoted by \( \{P_n\} \), will be called weakly ergodic if for all \( m \),

\[
\lim_{n \to \infty} \delta(P_{m,n}) = 0.
\]

We require this to be true for all \( m \) so that no matter when the process starts, whether at time zero or time \( m-1 \), memory will be lost. In this way, loss of memory does not come about because of the effect of one kernel (or indeed a finite number of kernels) which happen to be in the sequence.

**Lemma 2.1.4:** A stationary chain, denoted by \( \{P\} \) is weakly ergodic if and only if \( P \) is \( \alpha \)-primitive.

**Proof:** If the chain is weakly ergodic, then, from Definition 2.1.1, \( \delta(P^n) \to 0 \). Hence for some \( n \), \( \delta(P^n) < 1 \), hence \( \alpha(P^n) > 0 \).

Conversely, if \( \alpha(P^n_0) > 0 \), then \( \delta(P^n_0) < 1 \). In view of
Lemma 2.1.3, $\delta(P^{n_0})^k \leq [\delta(P^{n_0})]^k$ which tends to zero as $k \to \infty$. Since from Lemmas 2.1.2 and 2.1.3 the sequence $\{\delta(P^n)\}$ is non-increasing, it follows that $\delta(P^n)$ tends to zero as $n \to \infty$. \]

**B. Types of Ergodic Behavior**

Dobrushin (1956) defined ergodic behavior to be that which we have called weakly ergodic. He defined a Markov chain $\{P^n\}$ to be "strongly ergodic" if every chain $\{Q_n\}$ which contains $\{P^n\}$ is also weakly ergodic. That is, if $\{P^n\}$ is embedded in any other chain, that chain will be weakly ergodic. We will refer to this type of ergodic behavior as "contagious". We can state a theorem of Dobrushin's in these terms.

**Theorem 2.2.1:** A Markov chain is contagious ergodic if and only if $\sum_{n=1}^{\infty} \alpha(P_n) = \infty$.

It is clear that a chain which is contagious ergodic is also weakly ergodic. That the converse fails will become clear when Theorem 2.2.2 is given.

The following theorem given by Madsen (1971) gives three equivalent conditions for weak ergodicity. We will here supply the proof of the fact that condition (b) is equivalent to weak ergodicity since the criterion will be used in later theorems. The norm used in the theorem is defined as
\[ ||K(x,y)|| = \sup_{x} \int |K(x,y)| \mu(dy). \]

Recall that as in Equation (1.1.2), the notation \( K_{m,n} \) indicates the superposition of kernels \( K_{m} \) through \( K_{n} \) inclusive.

**Theorem 2.2.2:** Let \( \{P_n\} \) be a non-homogeneous Markov chain. The following are equivalent:

(a) \( \{P_n\} \) is weakly ergodic.

(b) There exists a subsequence of integers \( \{i_j\} \) which partitions the chain into blocks of kernels \( \{P^i_{j+1},...P^i_{j+1}\} \)

such that \( \sum_{j=1}^{\infty} \alpha(P^i_{j+1},P^i_{j+1}) \) diverges.

(c) For each \( m \) there is a sequence of constant stochastic kernels \( \{E^m_n\} \) such that \( \lim_{n \to \infty} ||P^m_n - E^m_n|| = 0. \)

(d) If \( P_n = E_n + R_n \), where \( E_n \) is a constant stochastic kernel, then \( \lim_{n \to \infty} ||R^m_n|| = 0. \)

**Proof of (a) \( \iff \) (b):** Assume \( \{P_n\} \) is weakly ergodic. Then given any integer, \( i_j \) say, \( \lim_{n \to \infty} \delta(P^i_{j+1},n) = 0. \) Hence there exists some \( N=N(i_j) \) such that \( \delta(P^i_{j+1},N) < \frac{1}{2} \), say. But this is equivalent to saying that \( \alpha(P^i_{j+1},N) > \frac{1}{2} \). If we define \( i_1 = 0 \) and \( i_{j+1} = N(i_j), \) \( j = 1,2,..., \) then we have that

\[ \sum_{j=1}^{\infty} \alpha(P^i_{j+1},P^i_{j+1}) > \sum_{j=1}^{\infty} \frac{1}{2} = \infty. \]
On the other hand, assume that (b) holds. From well-known properties of infinite products (see for example Nehari (1952) page 284), the infinite product \( \prod_{j=1}^{\infty} (1-a_j) \) diverges to zero if and only if \( \sum_{j=1}^{\infty} a_j = \infty \) (for \( 0 \leq a_j < 1 \)). Hence \( \sum_{j=1}^{\infty} \alpha(P_{i_j+1,i_{j+1}}) \) diverges implies
\[
\prod_{j=1}^{\infty} (1-\alpha(P_{i_j+1,i_{j+1}})) = \prod_{j=1}^{\infty} \delta(P_{i_j+1,i_{j+1}}) = 0 . \quad (2.2.1)
\]

Now let \( m \) be arbitrary and \( n > m \). Then for some \( k \) and \( \ell \), \( i_{k-1} < m \leq i_k \) and \( i_\ell < n \leq i_{\ell+1} \). It follows from Lemma 2.1.2 and Lemma 2.1.3 that
\[
\delta(P_{m,n}) = \delta(P_{m,i_k}, P_{i_k+1,i_{k+1}}, \ldots P_{i_\ell+1,n}) \\
\leq \delta(P_{m,i_k}) \prod_{j=k}^{\ell-1} \delta(P_{i_j+1,i_{j+1}}) \delta(P_{i_\ell+1,n}) \\
\leq \prod_{j=k}^{\ell-1} \delta(P_{i_j+1,i_{j+1}}) .
\]

Since \( \ell \to \infty \) as \( n \to \infty \), it follows from Equation (2.2.1) that \( \delta(P_{m,n}) \to 0 \). \( \square \)

A third kind of ergodic behavior involves convergence of superpositions of stochastic kernels to a constant stochastic kernel.

**Definition 2.2.1:** A non-homogeneous Markov chain \( \{P_n\} \) will be called strongly ergodic if there exists a constant
A simple corollary to Theorem 2.2.3 is that every strongly ergodic chain is weakly ergodic. This can be seen as follows. Let $E_{mn}$ be the constant kernel described in the theorem. Since $E_{mn}$ is constant, it follows that

$$
\int |P_{m,n}(x,y) - P_{m,n}(z,y)| \mu(dy)
\leq \int |P_{m,n}(x,y) - E_{mn}(x,y)| \mu(dy) + \int |E_{mn}(z,y) - P_{m,n}(z,y)| \mu(dy)
\leq 2 \sup_x \int |P_{m,n}(x,y) - E_{mn}(x,y)| \mu(dy) = 2 |P_{m,n} - E_{mn}|.
$$

Since the upper bound given here is independent of $x$ and $z$,
it follows from (a) that $\sup_{x,z} \int |P_{m,n}(x,y) - P_{m,n}(z,y)| \mu(dy)^{n} \to 0$.

The three types of ergodic behavior for stochastic kernels discussed above all depend on the integral of some quantity getting small, i.e., on convergence in the $L_1$ sense. Nothing has been said about pointwise behavior to this point. The following theorem does relate to pointwise behavior.

**Theorem 2.2.4**: Let $\{P_n\}$ be a sequence of stochastic kernels such that for all $n$,

$$0 \leq P_n(x,y) \leq \Delta < \infty.$$

Then for any $m$,

$$\sup_{x} P_{m,n}(x,y) - \inf_{x} P_{m,n}(x,y) \leq \Delta \delta(P_{m,n-1}).$$

**Proof**: Define $A_{m,n}(y) = \sup_{x} P_{m,n}(x,y) - \inf_{x} P_{m,n}(x,y)$.

Then

$$A_{m,n}(y) = \sup_{x,z} \{P_{m,n}(x,y) - P_{m,n}(z,y)\}$$

$$= \sup_{x,z} \{\int [P_{m,n-1}(x,u)P_n(u,y) - P_{m,n-1}(z,u)P_n(u,y)]\mu(du)\}$$

$$= \sup_{x,z} \{[P_{m,n-1}(x,u) - P_{m,n-1}(z,u)]^+ - [P_{m,n-1}(x,u) - P_{m,n-1}(z,u)]^-\}P_n(u,y)\mu(du)$$
\[ \leq \sup_{x,z} \left\{ \sup_{u} P_{n}(u,y) \int \left[ P_{m,n-1}(x,u) - P_{m,n-1}(z,u) \right]^{+} \mu(du) \right\} - \inf_{u} P_{n}(u,y) \int \left[ P_{m,n-1}(x,u) - P_{m,n-1}(z,u) \right]^{-} \mu(du) \] 

However, since all the kernels are stochastic, 
\[ \int \left[ P_{m,n-1}(x,u) - P_{m,n-1}(z,u) \right] \mu(du) = 1 - 1 = 0, \] 
so the positive and negative parts must be equal. Therefore 

\[
A_{m,n}(y) \leq \sup_{x,z} \left\{ \left[ \sup_{u} P_{n}(u,y) - \inf_{u} P_{n}(u,y) \right] \int \left[ P_{m,n-1}(x,u) - P_{m,n-1}(z,u) \right]^{+} \mu(du) \right\}.
\]

In view of the bounds on \( P_{n} \) and Lemma 2.1.1, 

\[
A_{m,n}(y) \leq \Delta \sup_{x,z} \int \left[ P_{m,n-1}(x,u) - P_{m,n-1}(z,u) \right]^{+} \mu(du)
\]

\[ = \Delta \delta(P_{m,n-1}). \]

The next two corollaries follow immediately from Lemma 2.1.3.

**Corollary 2.2.1:** If \( \{P_{n}\} \) is a contagious ergodic sequence satisfying \( 0 \leq P_{n}(x,y) \leq \Delta < \infty \), then for all \( m \)

\[
\sup_{x} P_{m,n}(x,y) - \inf_{x} P_{m,n}(x,y) \leq \Delta \prod_{j=m}^{n-1} \delta(P_{j}) \rightarrow 0.
\]
Corollary 2.2.2: If \( \{P_n\} \) is a weakly ergodic sequence satisfying \( 0 \leq P_n(x,y) \leq \Delta < \infty \), then for all \( m \),

\[
\sup_x P_{m,n}(x,y) - \inf_x P_{m,n}(x,y) \xrightarrow{n \to \infty} 0.
\]

Proof: In the proof of Theorem 2.2.2, it was shown that

\[
\prod_{j=k}^{k-1} \delta(P_{i_j+1,i_j+1}) \leq \prod_{j=k}^{k-1} \delta(P_{i_j+1,i_j+1}).
\]

By weak ergodicity, a sequence \( \{i_j\} \) can be found such that this product will go to zero.

Corollary 2.2.3: If \( \{P_n\} \) is a sequence of stochastic kernels satisfying \( 0 \leq P_n(x,y) \leq \Delta < \infty \), and if for each \( n \)

\[
\delta(P_n) \leq 1 - \epsilon = \delta,
\]

then for all \( m \),

\[
\sup_x P_{m,n}(x,y) - \inf_x P_{m,n}(x,y) \leq \Delta \delta^{n-m}.
\]

Proof: This follows from Lemma 2.1.3 and the hypothesis of this corollary.

Note that Conn proved a theorem similar to Corollary 2.2.3 under the condition that \( P_n(x,y) \) be bounded below by a positive constant and that the kernels be defined on a finite space. These conditions imply a uniform positive lower bound on \( \alpha(P_n) \) or equivalently the existence of \( \delta < 1 \) such that \( \delta(P_n) \leq \delta \).

Using Corollary 2.2.3, we can show that in the stationary
case \( P_n(x,y) = P(x,y) \) for all \( n \), under suitable conditions, a weakly ergodic sequence is also strongly ergodic. In fact, the powers of \( P \) converge to a left eigenfunction of \( P \) corresponding to the eigenvalue 1.

**Corollary 2.2.4:** If \( \{P_n\} \) is a stationary weakly ergodic sequence of stochastic kernels with a uniform upper bound \( \Delta \), say, and if \( \sup_x P(x,y) \) is an integrable function of \( y \), then there exists a function \( q(y) \) such that if for all \( x \), \( Q(x,y) = q(y) \), then

(a) \( \|P^n(x,y) - Q(x,y)\| \rightarrow 0 \)

and

(b) \( \int q(x) P(x,y) \mu(dx) = q(y) \).

**Proof:** From Lemma 2.1.4, we know that a stationary sequence is weakly ergodic if and only if \( P \) is \( \alpha \)-primitive. Say \( \alpha(P^\infty) > 0 \), then \( \delta(P^\infty) = \delta < 1 \). Let \( \lceil \frac{n}{\delta} \rceil \) represent the greatest integer function. Applying Corollary 2.2.3, we have that

\[
\sup_{x} P^n(x,y) - \inf_{x} P^n(x,y) \leq \Delta(\delta)^{\lceil \frac{n}{\delta} \rceil}
\]  

(2.2.2)

which, of course, tends to zero as \( n \to \infty \).

Now consider the sequence of functions \( \{\sup_x P^n(x,y)\} \).

This is a non-increasing sequence since
\[
\sup_x P^{n+1}(x,y) = \sup_x \int P(x,z)P^n(z,y)\mu(dz)
\]

\[
\leq \sup_z [\sup_x P^n(z,y)] \int P(x,z)\mu(dz)
\]

\[
= \sup_z P^n(z,y).
\]

Hence for each \(y\), \(\{\sup_x P^n(x,y)\}\) is a non-increasing sequence bounded below, hence the pointwise limit \(q(y)\) exists.

Similarly, for each \(y\), \(\{\inf_x P^n(x,y)\}\) is a non-decreasing sequence which is bounded above by the number \(q(y)\), hence its pointwise limit also exists. In fact, in view of (2.2.2), \(q(y)\) must be the common limit.

Also, since \(Q(x,y) = q(y)\) for all \(x\),

\[
|P^n(x,y) - Q(x,y)| = |P^n(x,y) - q(y)|
\]

\[
\leq |\sup_x P^n(x,y) - \inf_x P^n(x,y)| \leq \Delta(\delta) \frac{[n]}{x} \rightarrow 0, \quad (2.2.3)
\]

uniformly in \(x\) and \(y\).

In order to show that (a) holds, we must show that

\[
\limsup_{k\to\infty} \int |P^k(x,y) - Q(x,y)|\mu(dy) = 0.
\]

Let \(h_k(x,y) = |P^k(x,y) - Q(x,y)|\). Then

\[
\limsup_{k\to\infty} \int h_k(x,y)\mu(dy) \leq \lim_{k\to\infty} \int \sup_x h_k(x,y)\mu(dy)
\]

\[
= \int \limsup_{k\to\infty} [\sup_x h_k(x,y)]\mu(dy).
\]
The interchange of limit and integral is justified by the Lebesque dominated convergence theorem since for any \( x \)

\[
h_k(x, y) = |P^k(x, y) - Q(x, y)| = |P^k(x, y) - q(y)|
\]

\[
\leq \max[P^k(x, y), q(y)] \leq \max[\sup P^k(x, y), q(y)]
\]

\[
= \sup P^k(x, y).
\]

The last equality holds since \( q(y) \) is the pointwise limit of the non-increasing sequence \( \{\sup P^k(x, y)\} \). Also, since it is non-increasing, \( \sup P^k(x, y) \leq \sup P(x, y) \) which is an integrable function by hypothesis. Hence, for any \( x \),

\[
|P^k(x, y) - Q(x, y)| \leq \sup P(x, y)
\]

i.e.,

\[
\sup_x |P^k(x, y) - Q(x, y)| \leq \sup_x P(x, y).
\]

Now it suffices to show that the integrand is zero, i.e., that \( \limsup_{k \to \infty} h_k(x, y) = 0 \). It follows from (2.2.3) that

\[
\sup_x h_k(x, y) \leq \Delta(\delta)^{[k]} \leq \Delta(\delta)^{[\frac{1}{k}]}\]

which tends to zero as \( k \to \infty \).

To prove (b), consider

\[
P^{k+1}(x, y) = \int P^k(x, z) P(z, y) \mu(dz).
\]
The integrand is dominated by an integrable function since

\[ P^k(x,z)P(z,y) \leq \Delta \sup_x P^k(x,z) \leq \Delta \sup_x P(x,z). \]

Further, we know from (2.2.3) that \( \lim_{k \to \infty} P^k(x,y) = q(y) \).

Hence

\[ q(y) = \lim_{k \to \infty} P^{k+1}(x,y) = \lim_{k \to \infty} \int P^k(x,z)P(z,y)\mu(dz) \]

\[ = \int \lim_{k \to \infty} P^k(x,z)P(z,y)\mu(dz) \]

\[ = \int q(z)P(z,y)\mu(dz). \]

Note that it is easy to show that \( Q(x,y) \) is in fact a stochastic kernel, and since it is constant, (a) of Corollary 2.2.4 is equivalent to strong ergodicity.

It is clear that on a finite measure space, the upper bound \( \Delta \) on \( P(x,y) \) implies that \( \sup_x P(x,y) \) is an integrable function of \( y \). Hence a stationary sequence on a finite measure space with a bounded kernel is weakly ergodic if and only if it is strongly ergodic.

In view of these last remarks, it is impossible to find a finite stationary Markov chain which is weakly ergodic but does not converge.

Doob (1953) proved a result analogous to Corollary 2.2.4 for the finite matrix case.
C. Equivalence of Two Definitions of Weak and Strong Ergodicity

In the study of Markov chains most authors (see, for example, Feller (1968) or Karlin (1966)) say that a Markov chain is determined by the stochastic transition function and an initial probability distribution over the state space. In the case of a finite state space, they would say the chain is determined by the transition matrices and an initial probability vector. With this approach, one studies the behavior of the function which gives the probability distribution over the state space at any time \( n \), namely

\[
f_n(y) = \int f_{n-1}(x) P_n(x,y) \mu(dx) = \int f_0(x) P_{1,n}(x,y) \mu(dx),
\]

(2.3.1)

where \( f_0(y) \) is the initial probability distribution, hence must satisfy

\[
f_0(y) \geq 0 \quad \text{and} \quad \int f_0(y) \mu(dy) = 1.
\]

It is clear that since \( P_n(x,y) \) is a stochastic kernel, \( f_n(y) \) will also satisfy these two properties. \( f_0(y) \) will be called a starting density.

As was stated after Definition 2.1.1, in the non-stationary case it is necessary to consider behavior for a chain starting at time \( m \) rather than time \( 1 \). To do this, we consider \( f_{m,n}(y) \) defined by
\[ f_{m,n}(y) = \int f_0(x) P_{m,n}(x,y) \mu(dx) . \]

For convenience, when \( m=1 \), we will write "\( f_n(y) \)" as in (2.3.1).

Weak and strong ergodicity can now be defined in terms of starting densities. Theorems 2.3.1 and 2.3.2 show that these definitions are in fact equivalent to Definitions 2.1.1 and 2.2.1.

**Definition 2.3.1:** A non-homogeneous Markov chain will be called weakly ergodic if for all \( m,\)

\[ \sup_{f_0,g_0} \int |f_{m,n}(y) - g_{m,n}(y)| \mu(dy) \xrightarrow{n} 0 \]

where \( f_0 \) and \( g_0 \) are starting densities.

**Definition 2.3.2:** A non-homogeneous Markov chain will be called strongly ergodic if there exists a starting density \( q(y) \) such that for all \( m,\)

\[ \sup_{f_0} \int |f_{m,n}(y) - q(y)| \mu(dy) \xrightarrow{n} 0 \]

where \( f_0 \) is a starting density.

**Theorem 2.3.1:** Definition 2.1.1 is equivalent to Definition 2.3.1.

**Proof:** Assume \( \{P_n\} \) is weakly ergodic in the sense of Definition 2.1.1. Let \( f_0 \) and \( g_0 \) be any two starting
densities and let \( m \) be given. Define
\[
Q(x,y) = \begin{cases} 
  f_0(y) & \text{if } x \in A \\
  g_0(y) & \text{if } x \in A^C
\end{cases}
\]
where \( A \in \mathcal{B} \) and \( 0 < \mu(A) < \mu(S) \). Then \( f_{m,n}(y) \) can be expressed
\[
f_{m,n}(y) = \int f_0(x) P_{m,n}(x,y) \mu(dx) 
= \int Q(z_1,x) P_{m,n}(x,y) \mu(dx) = QP_{m,n}(z_1,y)
\]
if \( z_1 \in A \). Similarly if \( z_2 \in A^C \), then \( g_{m,n}(y) = QP_{m,n}(z_2,y) \).

Now since \( f_0 \) and \( g_0 \) are starting densities, \( Q \) is a stochastic kernel, and Lemmas 2.1.1, 2.1.2, and 2.1.3 can be applied to give
\[
\int |f_{m,n}(y) - g_{m,n}(y)| \mu(dy) = \int |QP_{m,n}(z_1,y) - QP_{m,n}(z_2,y)| \mu(dy) 
= \sup_{x,z} \int |QP_{m,n}(x,y) - QP_{m,n}(z,y)| \mu(dy) 
= 2\delta(QP_{m,n}) \leq 2\delta(Q)\delta(P_{m,n}) \leq 2\delta(P_{m,n}).
\]

Now if Definition 2.1.1 holds, \( \delta(P_{m,n}) \nrightarrow 0 \). Further, this goes to zero independently of the choice of \( f_0 \) and \( g_0 \).

Conversely, assume \( \sup_{f_0,g_0} \int |f_{m,n}(y) - g_{m,n}(y)| \mu(dy) \nrightarrow 0 \) for all \( m \). Let \( m \) be given. Define
\[
f^X(y) = P_m(x,y).
\]
Then the collection \( \{f^X(y): x \in S\} \) is a subset of the set of all
starting densities. Also,

\[
f_{m+1,n}(y) = \int f(z) P_{m+1,n}(z,y) \mu(dz) \]

\[
= \int P_m(x,z) P_{m+1,n}(z,y) \mu(dz) = P_{m,n}(x,y) .
\]

Using this equality we get

\[
\delta(P_{m,n}) = \frac{1}{2} \sup_{x,z} \left| P_{m,n}(x,y) - P_{m,n}(z,y) \right| \mu(dy)
\]

\[
= \frac{1}{2} \sup_{x,z} \left| f_{m+1,n}(y) - f_{m+1,n}(y) \right| \mu(dy)
\]

\[
< \frac{1}{2} \sup_{f_0, g_0} \int \left| f_{m+1,n}(y) - g_{m+1,n}(y) \right| \mu(dy)
\]

and this last expression tends to zero by assumption. \(\Box\)

**Theorem 2.3.2:** Definition 2.2.1 is equivalent to definition 2.3.2.

**Proof:** Assume \(\{P_n\} \) is strongly ergodic in the sense of Definition 2.2.1. Let \(Q(x,y)\) be the constant kernel described in Definition 2.2.1 and let \(q(y) = Q(x_0,y)\) for some \(x_0 \in S\). Then

\[
\int |f_{m,n}(y) - q(y)| \mu(dy) = \int \left| \int f_0(x) P_{m,n}(x,y) \mu(dx) - q(y) \right| \mu(dy)
\]

\[
= \int \left| \int f_0(x) [P_{m,n}(x,y) - q(y)] \mu(dx) \right| \mu(dy)
\]

\[
\leq \int \int f_0(x) |P_{m,n}(x,y) - q(y)| \mu(dx) \mu(dy)
\]
\[
= \int \int f_0(x) |P_{m,n}(x,y) - q(y)| \mu(dy) \mu(dx) \tag{2.3.2}
\]

\[
= \int f_0(x) \left[ \int |P_{m,n}(x,y) - q(y)| \mu(dy) \right] \mu(dx)
\]

\[
\leq \sup_x \int |P_{m,n}(x,y) - q(y)| \mu(dy) \int f_0(x) \mu(dx)
\]

\[
= \sup_x \int |P_{m,n}(x,y) - Q(x_0,y)| \mu(dy)
\]

\[
= ||P_{m,n} - Q||.
\]

The last equality holds since for each \( x \), \( Q(x,y) = Q(x_0,y) \) a.e. Also the norm goes to zero by assumption, independently of the choice of \( f_0 \). Equation (2.3.2) follows by Fubini's Theorem.

Conversely, assume there exists a starting density which satisfies the conditions of Definition 2.3.2. Define \( Q(x,y) = q(y) \) for all \( x \in S \). As in Theorem 2.3.1, define \( f^X(y) = P_m(x,y) \) so that \( f_{m+1,n}(y) = P_{m,n}(x,y) \). Then

\[
\sup_x \int |P_{m,n}(x,y) - Q(x,y)| \mu(dy) = \sup_x \int |f_{m+1,n}(y) - q(y)| \mu(dy)
\]

\[
\leq \sup_{f_0} \int |f_{m+1,n}(y) - q(y)| \mu(dy).
\]

This last expression tends to zero by assumption, hence the conclusion follows.

In the following chapters we focus attention on the behavior of functions analogous to \( f_{m,n}(y) \).
III. \( L_1 \) BEHAVIOR OF NON-NEGATIVE KERNELS

A. Normalization and Transformations

As indicated in Chapter I, we consider only sequences of non-negative kernels \( \{M_n(x,y)\} \) which are defined on \( S \times S \), measurable, for which kernel superpositions as defined in (1.1.2) exist, and for which integrals as in (1.1.3) exist. We do, however, exclude kernels which are zero a.e. for each \( x \). We will consider behavior of a sequence \( \{f_n(y)\} \) analogous to that defined in (2.3.1). We pointed out at that time that if \( \mathbf{P}_n \) is stochastic and if \( f_0(y) \) is any density, then \( f_n(y) \) will integrate to 1 for all \( n \).

It is easy to see that this last property will not necessarily hold for \( M_n \) non-negative. For example, if \( M_1(x,y) \) is defined on \( [0,1] \times [0,1] \) by

\[
M_1(x,y) = \begin{cases} 2 & x \in [0,\frac{1}{3}] = A_1 \\ 3 & x \in [\frac{1}{3},\frac{2}{3}] = A_2 \\ 4 & x \in [\frac{2}{3},1] = A_3 
\end{cases}
\]

then by choosing \( f_0 \) to be a density function which has its support on \( A_1, A_2, \) or \( A_3 \) respectively, then \( f_1(y) \) will integrate to \( \frac{2}{3}, 1, \) or \( \frac{4}{3} \). We certainly could not expect the difference \( |f_n(y) - g_n(y)| \) to get small in the \( L_1 \) sense if \( f_n \) and \( g_n \) do not integrate to the same thing. One obvious solution which is used extensively in Chapter IV, is
to compare $f_n^*(y)$ with $g_n^*(y)$ where

$$f_n^*(y) = f_n(y) / \int f_n(y) \mu(dy).$$

In this case, we would say that $f_n(y)$ is normalized. More generally, we should consider the behavior of $f_{m,n}^*(y)$. Now, since $\int f_{m,n}(y) \mu(dy)$ is a normalizing factor which depends on $f_0, m,$ and $n$, a more general approach to the problem of normalizing would be to look for sequences of constants $k(f_0, m, n)$ and $k(g_0, m, n)$ such that $|f_{m,n}^*(y) - g_{m,n}^*(y)|$ converges to zero in some sense.

Since beginning with a density is not helpful in this chapter we define a "starting function" to be any $f_0(y)$ satisfying

$$f_0(y) \geq 0$$

$$0 < \int f_0(y) \mu(dy) < \infty.$$

Also for any starting function, we define

$$f_{m,n}(y) = \int f_0(x) M_{m,n}(x,y) \mu(dx).$$

Further, given a sequence of constants $k(f_0, m, n)$ we define

$$f_{m,n}^*(y) = f_{m,n}(y) / k(f_0, m, n).$$

**Definition 3.1.1:** A sequence of non-negative kernels $\{M_n\}$ will be called $L_1$ weakly ergodic if for each pair of starting functions $f_0$ and $g_0$ there exist sequences of
positive constants $k(f_0,m,n)$ and $k(g_0,m,n)$ such that for all $m$,

$$\sup_{f_0,g_0} \int |f_{m,n}^*(y) - g_{m,n}^*(y)| \mu(\text{dy}) \overset{n}{\to} 0 \quad (3.1.1)$$

while

$$\int f_{m,n}^*(y) \mu(\text{dy}) \to 0 \text{ and } \int g_{m,n}^*(y) \mu(\text{dy}) \not\to 0. \quad (3.1.2)$$

If $k(f_0,m,n)$ and $k(g_0,m,n)$ are such that (3.1.1) and (3.1.2) are satisfied they will be called "satisfactory norming sequences". Note that if (3.1.1) holds and either of the expressions of (3.1.2) holds, then so will the other.

From Conn's work, it appears that the eigenfunctions of the kernels $\{M_n(x,y)\}$ play a significant role in the analysis of ergodic behavior. In view of this and in order to use the ergodic coefficient of Dobrushin, it is reasonable to transform the non-negative kernels into stochastic kernels as follows. If such exist, let $\phi_n(x)$ and $\psi_n(y)$ be respectively right and left positive eigenfunctions of $M_n(x,y)$ corresponding to eigenvalue $\lambda_n$. Then

$$P_n(x,y) = M_n(x,y) \phi_n(y)/\lambda_n \phi_n(x) \quad (3.1.3)$$

and

$$Q_n(x,y) = \psi_n(y) M_n(y,x)/\lambda_n \psi_n(x) \quad (3.1.4)$$

are both stochastic kernels. $\phi_n(x)$ and $\psi_n(y)$ are required
to be positive for the following reasons. If $\phi_n$ is a positive function, then $P_n(x,y)$ is well defined, whereas if $\phi_n$ were allowed to assume the value zero on some set of points, then $P_n(x,y)$ would not be defined on that set. If $\phi_n$ were allowed to assume both positive and negative values, then $P_n(x,y)$ would assume negative values, in which case $P_n$ would not be stochastic. $\psi_n$ is required to be positive for the same reasons.

It is not true that all non-negative kernels possess positive eigenfunctions. There are, however, certain classes of kernels which do have positive eigenfunctions. In Chapter I, we noted that according to Harris (1963), the class of primitive kernels has this property.

Jentzsch (1912) gave sufficient conditions for an eigenfunction to be positive, although his results are a special case of a theorem given by Sarymsakov (1949). The sufficient condition given by Sarymsakov is that the non-negative kernel defined on $[a,b] \times [a,b]$ be continuous and that for any point $(x,t)$ in the square, there exists an iterate $M^m(x,t)$ which is positive.

Birkhoff (1957) obtained various extensions of Jentzsch's theorem. One of these pertains to positive valued kernels.

Since some non-negative kernels do not have positive eigenfunctions, it is possible to define another transformation which is more general than (3.1.3). Let $G_n(y)$ be
any non-negative function such that

$$I_n(x) = \int M_n(x,y)G_n(y)\mu(dy)$$

exists and is positive for all \( x \). Then

$$R_n(x,y) = \frac{M_n(x,y)G_n(y)}{I_n(x)}$$  \hspace{1cm} (3.1.5)

is a stochastic kernel. If \( G_n(y) \) is an eigenfunction, then (3.1.5) reduces to (3.1.3). A similar transformation could be made to generalize (3.1.4).

B. Weak Ergodicity Using Eigenfunctions

The following lemmas will be useful in the proof of Theorem 3.2.1 which gives sufficient conditions for weak ergodicity of non-negative kernels.

**Lemma 3.2.1:** Let \( g(x) \) be an integrable function and let \( h(x) \) satisfy \( 0 < 1 - \epsilon < h(x) < 1 + \epsilon \). Then

$$|\int g(x)h(x)\mu(dx)| \leq |\int g(x)\mu(dx)| + \epsilon \int |g(x)|\mu(dx).$$

**Proof:** Let \( g(x) = g^+(x) - g^-(x) \) and \( |g(x)| = g^+(x) + g^-(x) \). Assume without loss of generality that \( \int g(x)h(x)\mu(dx) > 0 \). Then

$$|\int g(x)h(x)\mu(dx)| = \int g(x)h(x)\mu(dx) = \int g^+(x)h(x)\mu(dx)$$

$$- \int g^-(x)h(x)\mu(dx)$$
\[
\leq (1+\varepsilon) \int g^+(x)u(dx) - (1-\varepsilon)\int g^-(x)u(dx)
\]
\[
= \left| \int g^+(x)u(dx) - \int g^-(x)u(dx) + \varepsilon \left[ \int g^+(x)u(dx) 
+ \int g^-(x)u(dx) \right] \right|
\]
\[
\leq \left| \int g(x)u(dx) \right| + \varepsilon \int |g(x)|u(dx) . \]

Note that Lemma 3.2.1 holds when \( x \) is taken to be a vector and \( u \) is taken to be a product measure.

The next lemma shows that under certain conditions it is possible to find sequences of normalizing constants, dependent on \( \gamma \), so that if \( n \) is taken large enough, \( \int |f^*_n(y) - g^*_n(Y)|u(dy) \) will be less than \( \gamma \).

**Lemma 3.2.2:** Let \( \{M_n(x,y)\} \) be a sequence of non-negative kernels with the following properties:

a) there exist right eigenfunctions \( \phi_n(x) \) such that

i) \( 0 < b \leq \phi_n(x) \leq B < \infty \)

ii) \( \varepsilon_n = \sup_x \left| \frac{\phi_n(x)}{\phi_{n-1}(x)} - 1 \right| \) satisfies \( \sum_{n=1}^{\infty} \varepsilon_n < \infty \)

b) the sequence of stochastic kernels \( \{P_n(x,y)\} \) defined by (3.1.3) is weakly ergodic.

Then given \( \gamma > 0 \), \( f_0 \) and \( g_0 \) starting functions, there are
sequences of normalizing constants \( \{d_n(\gamma)\} \) and \( \{e_n(\gamma)\} \) such that for \( n \geq N(\gamma) \),

\[
\left| f_n^*(\gamma) - g_n^*(\gamma) \right| \mu(\text{d}y) < \gamma .
\]

Note that since eigenfunctions are determined only up to multiplicative constants, given \( \phi_{n-1}(x) \) one should choose \( \phi_n(x) \) in such a way that \( \epsilon_n \) is as small as possible, subject to \( \phi_n(x) \) remaining within specified bounds.

**Proof:** Since \( \sum_{n=1}^{\infty} \epsilon_n < \infty \), it must be that \( \prod_{n=1}^{\infty} (1+\epsilon_n) \) converges, hence, given \( \epsilon \) such that \( 0 < \epsilon < 1 \), there exists \( N_1 = N_1(\epsilon) \) such that

\[
\prod_{n=N_1}^{M} (1+\epsilon_n) \leq \prod_{n=N_1}^{\infty} (1+\epsilon_n) < 1 + \epsilon . \tag{3.2.1}
\]

Also it is not hard to show that \( \prod_{n=N_1}^{M} (1-\epsilon_n) \geq \prod_{n=N_1}^{\infty} (1-\epsilon_n) > 1-\epsilon . \)

If \( \{p_n\} \) is weakly ergodic, given \( \epsilon > 0 \) and \( m \), there exists an \( N_2 = N_2(\epsilon, m) \) such that for \( n \geq N_2 \),

\[
\delta(p_{m,n}) < \epsilon . \tag{3.2.2}
\]

Let \( \epsilon = \gamma b/4 \) where \( \gamma \) is some given small number. Let \( m = N_1(\epsilon) + 1 \) and let \( N(\gamma) = N_2(\epsilon, m) \). If \( f_0 \) and \( g_0 \) are the given starting functions define \( \{d_n\} \) as follows

\[
d_n = \begin{cases} 
1 & \text{if } n \leq N_1 \\
\left[ \prod_{j=N_1+1}^{n} \lambda_j \right] & \text{if } n > N_1
\end{cases} \tag{3.2.3}
\]

\[
\int_{N_1}^{n} f_{N_1}(y) \phi_{N_1+1}(y) \mu(\text{d}y)
\]
and define \({\{e_n\}}\) similarly using \(g_{N_1}\) instead of \(f_{N_1}\).

For \(n > N_2\), consider \(f_n^*(y)\) and \(g_n^*(y)\). For convenience, take

\[
f_{N_1}^{**}(y) = f_{N_1}^*(y) \frac{\phi_{N_1+1}(y)}{\int f_{N_1}^*(y) \phi_{N_1+1}(y) \mu(dy)}.\]

Clearly, \(f_{N_1}^{**}(y)\) is a density function. Then \(f_n(y)\) can be written as follows for \(n > N_1\).

\[
f_n(y) = \int f_{n-1}(x) M_n(x,y) \mu(dx)
= \int \int f_{n-2}(x) M_{n-1}(x,z_1) M_n(z_1,y) \mu(dz_1) \mu(dx)
= \int \cdots \int f_{N_1}(z_1) M_{N_1+1}(z_1,z_2) \cdots M_n(z_v,y) \mu(dz_1) \cdots \mu(dz_v)
\]

where \(v = n-N_1\). Using (3.1.3) this becomes

\[
f_n(y) = \int \cdots \int f_{N_1}(z_1) \left[ \lambda_{n} \phi_{N_1+1}(z_1) P_{N_1+1}(z_1,z_2)/\phi_{N_1+1}(z_2) \right]
\cdots \left[ \lambda_{n} \phi_{n}(z_v) P_{n}(z_v,y)/\phi_{n}(y) \right] \mu(dz_1) \cdots \mu(dz_v)
= \int \cdots \int f_{N_1}(z_1) \phi_{N_1+1}(z_1) \frac{\prod_{j=N_1+1}^{n} \lambda_j P_{N_1+1}(z_1,z_2)}{\phi_{N_1+1}(z_2)}
\cdots P_n(z_v,y) \frac{\prod_{j=2}^{v} \phi_{N_1+j}(z_j)/\phi_{N_1+j-1}(z_j)}{\phi_{n}(y)} \mu(dz_1) \cdots \mu(dz_v)
\]

(3.2.4)

Using (3.2.4) and \(d_n\) as defined in (3.2.3), \(f_n^*(y)\) can be written as
\[ f_n^*(y) = \int \ldots \int f^{**}_N(z_1 P_{N_1 + 1}(z_1, z_2) \ldots P_n(z_\nu, y) \prod_{j=2}^{\nu} \frac{\phi_{N_1 + j}(z_j)}{\phi_{N_1 + j - 1}(z_j)}] \]
\[
= \frac{1}{\phi_n(y)} \mu(dz_1) \ldots \mu(dz_\nu) . \quad (3.2.5)
\]

If \( g_n^*(y) \) is defined in the same way as \( f_n^*(y) \) and if \( n > N = N_2(\varepsilon, m) \), then
\[
\int |f_n^*(y) - g_n^*(y)| \mu(dy)
= \int |\int \ldots \int h(z_1, \ldots, z_\nu, y) \phi(z_2, \ldots, z_\nu) \mu(dz_1) \ldots \mu(dz_\nu)| \frac{1}{\phi_n(y)} \mu(dy)
\]

where
\[ h(z_1, \ldots, z_\nu, y) = [f^{**}_N(z_1) - g^{**}_N(z_1)] P_{N_1 + 1}(z_1, z_2) \ldots P_n(z_\nu, y) \]

and
\[ \phi(z_2, \ldots, z_\nu) = \prod_{j=2}^{\nu} \frac{\phi_{N_1 + j}(z_j)}{\phi_{N_1 + j - 1}(z_j)} . \]

Now \( h \) can assume positive and negative values and it follows from (3.2.1) that \( 1 - \varepsilon \leq \phi \leq 1 + \varepsilon \), hence Lemma 3.2.1 can be applied to get
\[ \int |f_n^*(y) - g_n^*(y)| \mu(dy) \]

\[ \leq \int \left| \int [f_{N1}^*(z_1) - g_{N1}^*(z_1)] P_{N1+1,n}(z_1,y) \mu(dz_1) \right| \frac{1}{\phi_n(y)} \mu(dy) \]

\[ + \varepsilon \int \left| \int f_{N1}^*(z_1) - g_{N1}^*(z_1) | P_{N1+1,n}(z_1,y) \mu(dz_1) \right| \frac{1}{\phi_n(y)} \mu(dy) \]

\[ = (3.2.6) \]

Since \( f_{N1}^* \) and \( g_{N1}^* \) were constructed to be densities, define
\[ P_{N1}^*(x,y) = \begin{cases} 
  f_{N1}^*(y) & \text{if } x \in A \\
  g_{N1}^*(y) & \text{if } x \in A^c 
\end{cases} \]

where \( 0 < \mu(A) < \mu(S) \). Then if \( x \in A \) and \( z \in A^c \), the first term of (3.2.6) becomes
\[ \int |P_{N1}^* P_{N1+1,n}(x,y) - P_{N1}^* P_{N1+1,n}(z,y)| \frac{1}{\phi_n(y)} \mu(dy) \]

\[ \leq \frac{1}{b} \int |P_{N1}^* P_{N1+1,n}(x,y) - P_{N1}^* P_{N1+1,n}(z,y)| \mu(dy) \]

\[ = \frac{2}{b} \delta(P_{N1}^* P_{N1+1,n}) \leq \frac{2}{b} \delta(P_{N1+1,n}) \leq \frac{2\varepsilon}{b} \]

where the inequalities of (3.2.7) follow from the bounds on \( \phi_n \) and Equation (3.2.2).

Now consider the second term of (3.2.6).
\[ \varepsilon \int \left| \int f_{N1}^*(z_1) - g_{N1}^*(z_1) | P_{N1+1,n}(z_1,y) \mu(dz_1) \right| \frac{1}{\phi_n(y)} \mu(dy) \]
Combining (3.2.7) and (3.2.8) yields

\[ \int |f^*_n(y) - g^*_n(y)| \mu(dy) \leq \frac{2e}{b} + \frac{2e}{b} = \gamma. \]

It remains to show that the norming constants used are positive. That is we must show that

\[ d_n = \left[ \int f_{N_1}^n(y) \phi_{N_1+1}(y) \mu(dy) \right] \prod_{j=N_1+1}^{N} \lambda_j > 0. \]

Now if \( \phi(x) \) is an eigenfunction and \( 0 < b \leq \phi(x) \leq B \infty \), then

\[ \lambda B \geq \lambda \phi(x) = \int M(x,y) \phi(y) \mu(dy) \geq b \int M(x,y) \mu(dy) \text{ for all } x \quad \text{and} \quad \lambda b \leq \lambda \phi(x) = \int M(x,y) \phi(y) \mu(dy) \leq B \int M(x,y) \mu(dy) \text{ for all } x. \]

Hence

\[ \lambda \geq \frac{b}{B} \sup_x \int M(x,y) \mu(dy) > 0 \]

where, \( \sup_x \int M(x,y) \mu(dy) > 0 \) since we exclude kernels which are zero a.e.

Similarly, \( 0 < \frac{\lambda b}{B} \leq \int M(x,y) \mu(dy) \text{ for all } x \) implies that
\[
\int f_n(y) \mu(dy) = \iint f_{n-1}(x) M_n(x,y) \mu(dx) \mu(dy)
\]
\[
= \int f_{n-1}(x) [\int M_n(x,y) \mu(dy)] \mu(dx)
\]
\[
\geq \frac{\lambda_n b}{B} \int f_{n-1}(x) \mu(dx)
\]
\[
\geq \left[ \prod_{k=1}^{n} \frac{\lambda_k}{B} \right]^{b^n} \int f_0(x) \mu(dx) > 0. \quad (3.2.9)
\]

Hence \( d_n > 0 \). \( \square \)

It should be pointed out that although \( d_n \) and \( e_n \) depend on the choice of \( f_0 \) and \( g_0 \), \( N_1 \) and \( N \) do not depend on \( f_0 \) and \( g_0 \), hence the result holds uniformly for any choice of starting functions.

Later it will be necessary to know whether \( \int f_n^*(y) \mu(dy) \) tends to zero. Under the conditions of Lemma 3.2.2 and using Equation (3.2.5), it is easy to see, since the integrand is positive and the \( P_n \) are stochastic that

\[
\int f_n^*(y) \mu(dy) = \int \ldots \int f_{N_1}^{**}(z_1) P_{N_1+1}(z_1, z_2) \ldots P_n(z_v, y)
\]

\[
\left[ \prod_{j=2}^{N_1} \phi_{N_1+j-1}(z_j) \right] \frac{1}{\phi_n(y)} \mu(dz_1) \ldots \mu(dz_v) \mu(dy)
\]

\[
\geq \frac{1}{B} \left[ \prod_{j=2}^{N_1} (1-\varepsilon_{N_1+j}) \right] \int \ldots \int f_{N_1}^{**}(z_1) P_{N_1+1}(z_1, z_2) \ldots P_n(z_v, y) \mu(dz_1)
\]

\[
\ldots \mu(dz_v) \mu(dy)
\]

\[
= \left[ \prod_{j=2}^{N_1} (1-\varepsilon_{N_1+j}) \right] / B.
\]
This last term is positive since the product converges. In fact

\[
\prod_{j=2}^{\infty} (1-\varepsilon_{N_1+j})/B \geq \prod_{j=2}^{\infty} (1-\varepsilon_{N_1+j})/B > 0 .
\]

A similar result can be found for an upper bound, hence

\[
\prod_{j=2}^{\infty} (1-\varepsilon_{N_1+j})/B \leq \int f_n^*(y) u(dy) \leq \prod_{j=2}^{\infty} (1+\varepsilon_{N_1+j})/b \quad (3.2.10)
\]

for all \( n \geq N_1+1 \). Further, these bounds are independent of \( f_0 \).

**Theorem 3.2.1:** If \( \{M_n(x,y)\} \) is a sequence of non-negative kernels satisfying the conditions of Lemma 3.2.2, then \( \{M_n\} \) is \( L_1 \) weakly ergodic.

**Proof:** Following Definition 3.1.1, let \( f_0 \) and \( g_0 \) be any starting functions. It suffices to consider the case \( m=1 \), i.e., \( f_{1,n}(y) \equiv f_n(y) \), since for any other \( m \) the arguments are identical.

Let \( \{\gamma_i\} \) be a sequence of constants decreasing to zero. By Lemma 3.2.2, for each \( i \) there exist sequences of constants \( \{d_n(\gamma_i)\} \) and \( \{e_n(\gamma_i)\} \) such that

\[
n > N(\gamma_i) + \int |f_n^*(y) - q_n^*(y)| u(dy) < \gamma_i .
\]

Without loss of generality, assume that \( \{N(\gamma_i)\} \) forms an increasing sequence.

The coefficients \( \{d_n(\gamma_i)\} \) can be written in an array as follows:
A similar array can be written for \( \{e_n(y_i)\} \). It was pointed out earlier that the numbers \( N(y_i) \) are independent of \( f_0 \) and \( g_0 \), hence can be used for both the \( \{d_n\} \) and \( \{e_n\} \) sequences. Now define

\[
k(f_0, l, n) = \begin{cases} 
  d_n(y_1) & n \leq N(y_2) \\
  d_n(y_i) & N(y_i) < n < N(y_{i+1})
\end{cases}
\]

and define \( k(g_0, l, n) \) similarly.

These sequences of constants will satisfy the conditions of Definition 3.1.1, for if \( \varepsilon > 0 \) is given, there is some \( i \) such that \( y_i < \varepsilon \). Then for any \( n > N(y_1) \),

\[
\int |f_n^*(y) - g_n^*(y)\mu(dy) < y_i < \varepsilon
\]

independently of the choice of \( f_0 \) and \( g_0 \).

It remains to show \( \int f_n^*(y)\mu(dy) \rightarrow 0 \). In the proof of Lemma 3.2.2, \( \varepsilon \) was taken to be \( yb/4 \), so as \( y \rightarrow 0 \), \( \varepsilon \rightarrow 0 \) and \( N_1 = N_1(\varepsilon) \) can be assumed to go to infinity as \( \varepsilon \rightarrow 0 \). Since \( \varepsilon \)
is a function of \( y \), we could write \( N_1 = N_1(y) \).

Then since \( \mathbb{N}_1(y_1) = \mathbb{N}_1(y_{i+1}) \), we know from (3.2.10) that for all \( n > \mathbb{N}_1(y_1) + 1 \),

\[
\prod_{j=2}^{\infty} \frac{(1 - \varepsilon \mathbb{N}_1(y_1 + j))}{B} \leq \int_{\mathbb{N}_1(y_1 + j)} f_n^*(y) \mu(dy) \leq \prod_{j=2}^{\infty} \frac{(1 + \varepsilon \mathbb{N}_1(y_1 + j))}{B}.
\]

In fact we could show that

\[
1/B \leq \lim_{n \to \infty} \int f_n^*(y) \mu(dy) \leq \lim_{n \to \infty} \int f_n^*(y) \mu(dy) \leq 1/b.
\] (3.2.11)

Hence \( \int f_n^*(y) \mu(dy) \to 0 \). []

In Conn's work and in Chapter IV satisfactory norming sequences are obtained by using \( f_n^*(y) \mu(dy) \) (or more generally, \( f_{m,n}^*(y) \mu(dy) \)). The question arises as to whether this norming would be satisfactory under the conditions of Theorem 3.2.1. The answer is in the affirmative as is shown by the following corollary.

**Corollary 3.2.1:** If \( \{M_n(x,y)\} \) is a sequence of non-negative kernels satisfying the conditions of Lemma 3.2.2, then \( k'(f_0,m,n) = \int f_{m,n}(y) \mu(dy) \) and \( k'(g_0,m,n) = \int g_{m,n}(y) \mu(dy) \) are satisfactory norming sequences for showing that \( \{M_n\} \) is \( L_1 \) weakly ergodic.

**Proof:** Here again it suffices to show this for the cases of \( m = 1 \). Define
\[ f_{n}^{**}(y) = f_{n}^{*}(y) / \int f_{n}^{*}(y) \mu(dy). \] (3.2.12)

Then, since \( f_{n}^{*}(y) = f_{n}(y) / k(f_{0}, l, n) \), it follows that

\[ f_{n}^{**}(y) = f_{n}(y) / \int f_{n}(y) \mu(dy) = f_{n}(y) / k'(f_{0}, l, n) \] (3.2.13)

Let \( \varepsilon > 0 \) be given. We wish to show that there exists \( N'(\varepsilon) \), independent of the choice of \( f_{0} \) and \( g_{0} \), such that if \( n > N' \), then

\[ \left| \frac{f_{n}(y)}{\int f_{n}(y) \mu(dy)} - \frac{g_{n}(y)}{\int g_{n}(y) \mu(dy)} \right| < \varepsilon. \]

From Equation (3.2.11) there exists \( N_B \) such that \( n > N_B \) implies that

\[ \frac{1}{2B} \leq \int f_{n}^{*}(y) \mu(dy). \] (3.2.14)

Define \( \varepsilon_{1} = \varepsilon / 4B \). By Theorem 3.2.1, there exists an \( N(\varepsilon_{1}) \) such that \( n > N(\varepsilon_{1}) \) implies that

\[ \left| f_{n}^{*}(y) - g_{n}^{*}(y) \right| \mu(dy) < \varepsilon_{1}. \]

Now let \( n > N'(\varepsilon) \equiv \max\{N(\varepsilon_{1}), N_B\} \) and define

\[ A_{n}(y) = \left| \frac{f_{n}(y)}{\int f_{n}(y) \mu(dy)} - \frac{g_{n}(y)}{\int g_{n}(y) \mu(dy)} \right|. \]

Then from (3.2.12) and (3.2.13),
\[ A_n(y) = \left| \frac{f_n^*(y) - g_n^*(y)}{f_n^*} - \frac{g_n^*(y)}{g_n^*} \right| \]

\[ \leq \left| \frac{f_n^*(y)}{f_n^*} - \frac{f_n^*(y)}{g_n^*} \right| + \left| \frac{f_n^*(y) - g_n^*(y)}{g_n^*} - \frac{g_n^*(y)}{f_n^*} \right| \quad (3.2.15) \]

Considering the first term of (3.2.15), we find

\[ \left| \frac{f_n^*(y) - g_n^*(y)}{f_n^*} - \frac{g_n^*(y)}{g_n^*} \right| = f_n^*(y) \left| \int \frac{g_n^*(y) - f_n^*(y)]u(dy)}{f_n^*} \right| \]

\[ = \frac{f_n^*(y)}{f_n^*} \left| \int \frac{g_n^*(y) - f_n^*(y)]u(dy)}{g_n^*} \right| \]

\[ \leq 2B \frac{f_n^*(y)}{f_n^*} \left| \int [g_n^*(y) - f_n^*(y)]u(dy) \right|, \]

where the inequality holds from (3.2.14) as applied to \( g_n^*(y) \).

Applying the same inequality to the second term of (3.2.15) we get:

\[ \left| \frac{f_n^*(y) - g_n^*(y)}{f_n^*} - \frac{g_n^*(y)}{g_n^*} \right| = \frac{1}{g_n^*} \left| f_n^*(y) - g_n^*(y) \right| \leq 2B \left| f_n^*(y) - g_n^*(y) \right|. \]

Finally, since

\[ \left| \int [g_n^*(y) - f_n^*(y)]u(dy) \right| \leq \left| \int [f_n^*(y) - g_n^*(y)]u(dy) \right| < \varepsilon_1. \]

it follows that
\[ \int A_n(y) \mu(dy) \leq \left\{ \begin{array}{l}
\frac{f^*(y)}{f_n^*(y)} (2B) \left| \int [g_n^*(y) - f_n^*(y)] \mu(dy) \right| \\
+ (2B) \left| f_n^*(y) - g_n^*(y) \right| \mu(dy)
\end{array} \right\}
\leq 2B \left( \epsilon_1 \int \frac{f_n^*(y)}{f_n^*(y)} \mu(dy) + \int \left| f_n^*(y) - g_n^*(y) \right| \mu(dy) \right)
\leq 2B(\epsilon_1 + \epsilon_1) = \epsilon.
\]

It is clear that \( \int f_n(y)/k'(f_0,l,n) \mu(dy) = 1 \rightarrow 0 \).
Also it follows from (3.2.9) that \( k'(f_0,l,n) = \frac{f_n(y)\mu(dy)}{k_n(y)} > 0 \).

Theorem 3.2.1 and Corollary 3.2.1 show that under certain conditions, there is more than one way to find a satisfactory norming sequence. In fact, if \( k(f_0,m,n) \) and \( k(g_0,m,n) \) are satisfactory norming sequences, then for any constant \( c \), the sequences \( c \cdot k(f_0,m,n) \) and \( c \cdot k(g_0,m,n) \) will also be satisfactory norming sequences. We conclude this section with some lemmas concerning the relationship between different satisfactory norming sequences.

For notational simplicity, assume \( m=1 \) and write, for example, \( k(f_0,l,n) \) as \( k(f_0,n) \), etc.

**Lemma 3.2.3:** If \( r_n = \frac{k(f_0,n)}{k'(f_0,n)} = \frac{k(g_0,n)}{k'(g_0,n)} \) and if
\[ 0 < c_1 = \lim r_n \leq \lim r_n = c_2 < \infty, \]
then the sequences \(k(f_0, n), k(g_0, n)\) are satisfactory norming sequences if and only if \(k'(f_0, n)\) and \(k'(g_0, n)\) are.

**Proof:** Let \(I_n = \int \left| \frac{f_n(y)}{k(f_0, n)} - \frac{g_n(y)}{k(g_0, n)} \right| \mu(dy)\) and \(J_n = \int \frac{f_n(y)}{k(f_0, n)} \mu(dy)\) with \(I'_n\) and \(J'_n\) similarly defined. Assume \(k(f_0, n)\) and \(k(g_0, n)\) are satisfactory norming sequences (SNS). Then it follows that

\[
\overline{\lim} I_n = \lim I_n = 0
\]

\[
\overline{\lim} J_n = L_1 > 0.
\]

Note that

\[
J'_n = \int f_n(y)/k'(f_0, n) \mu(dy) = \int \frac{f_n(y)}{k(f_0, n)} \cdot \frac{k(f_0, n)}{k'(f_0, n)} \mu(dy)
\]

\[
= r_n J_n.
\]

Similarly, \(I'_n = r_n I_n\). Then since all terms involved are positive, we use properties of \(\overline{\lim}\) to obtain the following.

\[
\overline{\lim} I'_n = \overline{\lim} r_n I_n \leq \overline{\lim} r_n \overline{\lim} I_n
\]

\[
= c_2 \overline{\lim} I_n = 0.
\]

Also, \(\overline{\lim} J'_n > 0\), since if \(\overline{\lim} J'_n = 0\), then

\[
\overline{\lim} J_n = \overline{\lim} J'_n/r_n \leq \overline{\lim} J'_n \cdot \overline{\lim} 1/r_n
\]

\[
= \overline{\lim} J'_n/\overline{\lim} r_n = \overline{\lim} J'_n/c_1 = 0,
\]
which contradicts (3.2.16). Likewise it is easy to show that
\[ \int g_n(y)/k'(g_0,n) \mu(dy) \to 0. \] Hence \( k'(f_0,n) \) and \( k'(g_0,n) \) are SNS's.

The converse follows at once since \( r_n' = \frac{k(f_0,n)}{k'(f_0,n)} \) will satisfy \( 0 < \frac{1}{c_2} = \lim r_n' \leq \lim r_n' = \frac{1}{c_1} < \infty \). \( \square \)

Lemma 3.2.4: If \( r_n = \frac{k(f_0,n)}{k'(f_0,n)} \) and \( s_n = \frac{k(g_0,n)}{k'(g_0,n)} \)
and \( 0 < c_1 = \lim r_n < \lim r_n = c_2 < \infty \) and \( |r_n-s_n| \to 0 , \)
and \( \lim \int f_n(y)/k(f_0,n) \mu(dy) = L_1 < \infty \), then \( k(f_0,n) , \)
\( k(g_0,n) \) are SNS's if and only if \( k'(f_0,n) \) and \( k'(g_0,n) \) are.

Proof: Assume \( k(f_0,n) \) and \( k(g_0,n) \) are SNS's. The fact that \( \lim J_n' > 0 \) follows as in Lemma 3.2.3. Also,

\[ I_n' = \int \left| \frac{f_n(y)}{k'(f_0,n)} - \frac{g_n(y)}{k'(g_0,n)} \right| \mu(dy) \]

\[ = \int \left| \frac{f_n(y)}{k(f_0,n)} r_n - s_n \frac{g_n(y)}{k(f_0,n)} \right| \mu(dy) \]

\[ \leq \int |r_n-s_n| \frac{f_n(y)}{k(f_0,n)} \mu(dy) + \int s_n \left| \frac{f_n(y)}{k(f_0,n)} - \frac{g_n(y)}{k(g_0,n)} \right| \mu(dy) \]

\[ = |r_n-s_n| J_n + s_n I_n . \]

The hypotheses imply that \( \lim s_n = \lim r_n = c_2 , \) hence
\[
\lim I_n' \leq \lim \left[ |r_n - s_n| \cdot J_n \right] + \lim s_n I_n
\]
\[
\leq \lim |r_n - s_n| \cdot \lim J_n + \lim s_n \cdot \lim I_n
\]
\[
= L_1 \lim |r_n - s_n| + c_2 \lim I_n = 0 .
\]

As in Lemma 3.2.3, the converse is easy. [ ]

Note that since \( J_n' = J_n r_n \), \( \lim J_n \) being finite implies \( \lim J_n' \) is finite. Also, since \( K_n = \int g_n(y)/k(g_0,n) \mu(dy) \) satisfies

\[
|J_n - K_n| < I_n ,
\]

it follows that \( \lim K_n = \lim J_n = L_1 < \infty \). Hence under the conditions of Lemma 3.2.4, if any one of \( J_n, J_n', K_n, \) or \( K_n' \) has a finite \( \lim \sup \), then so will the other three.

The converse of Lemma 3.2.4 is not true. In fact even with the additional restriction that \( \lim J_n \) and \( \lim J_n' \) be finite, it is possible for \( \lim r_n \) to be zero or \( \lim r_n \) to be infinite. For example, if \( k(f_0,n) = \int f_n(y) \mu(dy) \) and \( k(g_0,n) = \int g_n(y) \mu(dy) \) are SNS's and if

\[
k'(f_0,n) = \begin{cases} 
k(f_0,n) & n \text{ odd} \\
k(f_0,n) & n \text{ even} \end{cases}
\]

then clearly the primed sequences will be SNS's. Further \( \lim J_n = \lim J_n' = 1 \), while \( \lim r_n = 0 \). Also by interchanging the roles of primed and unprimed sequences in this
example, it follows that its possible to have \( \lim r_n = \infty \).

C. Weak Ergodicity Without Using Eigenfunctions

At the end of Section A of this Chapter, we indicated that in the case where positive eigenfunctions do not exist, it is possible to try another transformation, namely that given in Equation (3.1.5). A generalized version of Theorem 3.2.1 can be given in terms of this alternate transformation.

If \( I_n(x) = \int M_n(x,y)G_n(y)\mu(dy) \), define

\[
H_n(x) = I_n(x)/\lambda_n
\]

where \( \lambda_n \) is some positive constant. While \( \lambda_n \) is not necessarily an eigenvalue, it will play a role similar to that of the eigenvalues in Lemma 3.2.1.

**Theorem 3.3.1:** Let \( \{M_n(x,y)\} \) be a sequence of non-negative kernels with the following properties:

a) there exist functions \( G_n(y) \) such that

i) \( 0 < b \leq G_n(y) \leq B < \infty \)

ii) \( I_n(x) = \int M_n(x,y)G_n(y)\mu(dy) \) exists and is positive for all \( x \)

iii) there exists a sequence of constants \( \{\lambda_n\} \) such that
\[ e_n = \sup_x \left| \frac{H_n(x)}{G_{n-1}(x)} - 1 \right| = \sup_x \left| \frac{I_n(x)}{\lambda_n G_{n-1}(x)} - 1 \right| \]

satisfies

\[ \sum e_n < \infty \]

b) the sequence of stochastic kernels \( \{R_n(x,y)\} \)

defined by (3.1.5) is weakly ergodic.

Then \( \{M_n\} \) is \( L_1 \) weakly ergodic.

The proof of this theorem follows the proofs of Lemma 3.2.2 and Theorem 3.2.1.

Note that if a sequence of functions \( \{G_n(y)\} \) such that (i) and (ii) are satisfied is given, a sequence of constants \( \{\lambda_n\} \) such that condition (iii) is satisfied exists if and only if

\[ e'_n = \inf_{\lambda_n > 0} \sup_x \left| \frac{I_n(x)}{\lambda_n G_{n-1}(x)} - 1 \right| \]

satisfies \( \sum e'_n < \infty \). In fact if \( \lambda_n \) is chosen such that

\[ e_n = \sup_x \left| \frac{I_n(x)}{\lambda_n G_{n-1}(x)} - 1 \right| < e'_n + \frac{1}{2^n}, \]

Then it is clear that \( \sum e'_n < \infty \) implies \( \sum e_n < \infty \).

We conclude this section with an example for which

\( L_1 \) weak ergodicity can be shown to hold by using Theorem 3.3.1.
but not by using Theorem 3.2.1. Theorem 3.2.1 cannot be applied since positive eigenfunctions may not exist.

For simplicity we consider a discrete state space and take $M_n$ to be a matrix. Let $\{a_n\}$ be a sequence of independent random numbers taken, say, from a uniform distribution over $(1/4, 1/2)$. For $n = 1, 2, \ldots$, define

$$M_n = \begin{pmatrix} 1 & 0 \\ 0 & b_n \end{pmatrix}$$

where

$$b_n = \frac{a_n}{1-a_n} \cdot \frac{1-2a_{n-1}}{a_{n-1}}.$$ 

Also define

$$G_n = \begin{pmatrix} a_n \\ 1-a_n \end{pmatrix}.$$ 

Clearly the elements of $G_n$ are bounded by $1/4$ and $3/4$, hence (i) of Theorem 3.3.1 is satisfied. Also

$$I_n = M_n G_n = \begin{pmatrix} 1 & 0 \\ 0 & \frac{a_n}{1-a_n} \cdot \frac{1-2a_{n-1}}{a_{n-1}} \end{pmatrix} \begin{pmatrix} a_n \\ 1-a_n \end{pmatrix}$$

$$= \begin{pmatrix} \frac{a_n}{a_n} \\ \frac{a_n(1-a_{n-1})}{a_{n-1}} \end{pmatrix} = \frac{a_n}{a_n-1} \begin{pmatrix} a_{n-1} \\ 1-a_{n-1} \end{pmatrix} = \frac{a_n}{a_n-1} G_{n-1}.$$ 

Hence $I_n$ is positive, and so (ii) also holds.
Now, choosing \( \lambda_n = \frac{a_n}{a_{n-1}} \), we find that \( I_n = \lambda_n G_{n-1} \), hence \( \varepsilon_n = 0 \) for all \( n \) and so \( \sum \varepsilon_n = 0 < \infty \), i.e. (iii) holds.

Finally, the stochastic matrix \( R_n \) defined by (3.1.5) can be shown to be

\[
R_n = \begin{pmatrix}
1 & 0 \\
\frac{a_{n-1}}{1-a_{n-1}} & 1-2a_{n-1} \\
\frac{1}{1-a_{n-1}} & \frac{1}{1-a_{n-1}}
\end{pmatrix}.
\]

Using Equation (1.2.4) to find the ergodic coefficient, we find

\[
\alpha(R_n) = \frac{a_{n-1}}{1-a_{n-1}} \geq \frac{1}{3},
\]

hence the sequence \( \{R_n\} \) is weakly ergodic. This shows that all conditions of Theorem 3.3.1 are satisfied.

On the other hand, it is easy to see that \( b_n \) satisfies \( 0 < b_n < 2 \), and it is not hard to show that for \( b_n > 1 \), there are no positive eigenvectors.
IV. POINTWISE BEHAVIOR OF NON-NEGATIVE KERNELS

A. Weak Ergodicity

In Chapter III we considered the behavior of \( \{f_n^*(y)\} \) and \( \{g_n^*(y)\} \) in the \( L_1 \) sense, i.e. \( \int |f_n^*(y) - g_n^*(y)| \mu(dy) \). In this chapter, we consider pointwise behavior, i.e.

\[
|f_n^*(y) - g_n^*(y)|
\]

and we look for conditions under which there will be loss of memory (weak ergodicity) or convergence and loss of memory (strong ergodicity). In Chapter III the space \( S \) was assumed to be \( \sigma \)-finite, but in this chapter it will be assumed finite. Under the assumption that \( \mu(S) < \infty \), uniform pointwise convergence implies \( L_1 \) convergence and since all of the principal results of this chapter are of the former type of convergence, they will also be of the latter type.

In addition to the usual assumptions for the sequence \( \{M_n\} \) of measurability, existence of sequential superpositions, etc. we will, for various theorems, impose some of the following conditions:

(i) \( 0 \leq M_n(x,y) \leq \Delta \) \hspace{1cm} (4.1.1)

(ii) \( M_n(x,S) = \int M_n(x,y) \mu(dy) \geq v > 0 \) for all \( x \) \hspace{1cm} (4.1.2)

(iii) there is an eigenvalue \( \lambda_n \) with corresponding right and left eigenfunctions \( \phi_n(x) \) and \( \psi_n(y) \)
satisfying
\[ 0 < b \leq \phi_n(x) \leq B < \infty \quad (4.1.3) \]
\[ 0 < d \leq \psi_n(y) \leq D < \infty \quad (4.1.4) \]

(iv) the sequence \( \{P_n\} \) defined by (3.1.3) satisfies for some \( \delta \)

\[ \delta(P_n) \leq \delta < 1 \quad (4.1.5) \]

(v) the sequence \( \{Q_n\} \) defined by (3.1.4) satisfies for some \( \delta \)

\[ \delta(Q_n) \leq \delta < 1. \quad (4.1.6) \]

Note that if condition i) is replaced by,

(i') \[ 0 < \Lambda \leq M_n(x,y) \leq \overline{\Lambda} < \infty, \]

then all of conditions (ii) through (v) are satisfied.

Condition (i') is the condition imposed by Conn and it is easy to give families of kernels for which (i) through (v) are satisfied even though (i') is not. Note also that condition (ii) in the matrix case would reduce to requiring that the row sums be bounded below by a positive number.

The bounds in conditions (i) through (v) allow other useful bounds to be derived. These bounds are derived in the next lemma.
Lemma 4.1.1:

(a) \((4.1.1)\) implies \(M_{n,n+m}(x,y) \leq \Delta^{m+1} \mu(S)^m\) \((4.1.7)\)

(b) \((4.1.1)\) and \((4.1.2)\) imply \(0 < \int f_n(y) \mu(dy) < \infty\)

and

\[ f_n^*(y) = f_n(y)/\int f_n(y) \mu(dy) \leq \Delta/\nu \] \((4.1.9)\)

(c) \((4.1.1)\), \((4.1.2)\), and \((4.1.4)\) imply

\[ 0 < \nu \leq \lambda_n \leq \Delta \mu(S) \] \((4.1.10)\)

(d) \((4.1.1)\), \((4.1.2)\), and \((4.1.4)\) imply

\[ Q_m(x,y) \leq \Delta D/\nu d \] \((4.1.11)\)

(e) \((4.1.1)\), \((4.1.2)\), \((4.1.3)\), and \((4.1.4)\) imply

\[ P_n(x,y) \leq \Delta B/\nu b \] \((4.1.12)\)

Proof: (a) When \(m=0\), the result is trivial since

\[ M_n(x,y) \leq \Delta. \]

Assume the result holds for \(m-1\). Then

\[ M_{n,n+m}(x,y) = \int M_{n,n+m-1}(x,z)M_{n+m}(z,y) \mu(dz) \]

\[ \leq \Delta^m \mu(S)^{m-1} \int M_{n+m}(z,y) \mu(dz) \]

\[ \leq \Delta^{m+1} \mu(S)^m \int \mu(dz) = \Delta^{m+1} \mu(S)^m. \]

(b) \[ \int f_n(y) \mu(dy) = \int f_{n-1}(x) M_n(x,y) \mu(dx) \mu(dy) \]

\[ = \int f_{n-1}(x) M_n(x,y) \mu(dy) \mu(dx) \geq \nu \int f_{n-1}(x) \mu(dx) \]

\[ \geq \nu^2 \int f_{n-2}(x) \mu(dx) \ldots \geq \nu^n \int f_0(x) \mu(dx) > 0. \]
The last inequality follows since \( f_0 \) is a starting function.

Likewise,

\[
\int f_n(y) \mu(dy) = \int \int f_{n-1}(x) M_n(x,y) \mu(dx) \mu(dy)
\]

\[
\leq \Delta \int \int f_{n-1}(x) \mu(dx) \mu(dy) = \Delta \mu(S) \int f_{n-1}(x) \mu(dx)
\quad (4.1.14)
\]

\[
\leq [\Delta \mu(S)]^2 \int f_{n-2}(x) \mu(dx) \ldots \leq [\Delta \mu(S)]^n \int f_0(x) \mu(dx) < \infty.
\]

Inequality (4.1.9) follows from (4.1.13) since

\[
f_n(y)/\int f_n(y) \mu(dy) \leq \frac{\int f_{n-1}(x) M_n(x,y) \mu(dx)}{\nu \int f_{n-1}(x) \mu(dx)}
\]

\[
\leq \frac{\Delta}{\nu} \frac{\int f_{n-1}(x) \mu(dx)}{\int f_{n-1}(x) \mu(dx)} = \frac{\Delta}{\nu}.
\]

(c) Since \( \psi_n(y) \) is a bounded measurable function defined on a finite measure space, we know \( \psi_n(y) \) is integrable, hence

\[
\lambda_n \int \psi_n(y) \mu(dy) = \int \int \psi_n(x) M_n(x,y) \mu(dx) \mu(dy)
\]

\[
\leq \Delta \int \int \psi_n(x) \mu(dx) \mu(dy) = \Delta \mu(S) \int \psi_n(x) \mu(dx)
\].

Therefore, \( \lambda_n \leq \Delta \mu(S) \). Similarly, using (4.1.2),
\[ \lambda_n \int \psi_n(y) u(dy) = \int \int \psi_n(x) M_n(x,y) u(dy) u(dx) \]

\[ = \int \psi_n(x) M_n(x,S) u(dx) \geq \int \psi_n(x) u(dx) . \]

Therefore, \( \lambda_n \geq v > 0 . \)

(d) and (e). These are trivial using (4.1.10). □

**Definition 4.1.1:** A sequence of non-negative kernels \( \{M_n\} \) will be called pointwise weakly ergodic if for each pair of starting functions \( f_0 \) and \( g_0 \) there exist sequences of positive constants \( k(f_0,m,n) \) and \( (k(g_0,m,n) \) such that for all \( m, \)

\[ \sup_{f_0,g_0} |f_{m,n}^*(y) - g_{m,n}^*(y)| \rightarrow 0 \]

uniformly in \( y, \) while

\[ \int f_{m,n}^*(y) u(dy) \rightarrow 0 \text{ and } \int g_{m,n}^*(y) u(dy) \rightarrow 0 . \]

Note that throughout this chapter none of the proofs in any way depend on the choice of \( m. \) Hence from this point on, we will proceed, without loss of generality, with \( m=1. \) Also throughout this chapter we take \( k(f_0,1,n) = \int f_n(y) u(dy). \) Hence we always have

\[ f_n^*(y) = f_n(y)/k(f_0,1,n) = f_n(y)/\int f_n(y) u(dy) . \]

Finally, with this convention and in view of (4.1.8), it
follows that the norming constants are positive and that
\[ \int f_n^*(y) \mu(dy) = 1 \rightarrow 0. \]

In order to summarize the necessary assumptions for the theorems to follow, we will define various conditions. Condition W is the first.

**Condition W:** A sequence of kernels satisfying (4.1.1), (4.1.2), (4.1.3), (4.1.4), and (4.1.5) will be said to satisfy Condition W.

The following theorem gives sufficient conditions for pointwise weak ergodicity. We point out here that the techniques used in proving this and subsequent theorems in this chapter are those used by Conn.

**Theorem 4.1.1:** Let \( \{M_n(x,y)\} \) be a sequence of kernels satisfying Condition W. If
\[ \int |\phi_n(x) - \phi_{n+1}(x)| \mu(dx) \overset{n}{\rightarrow} 0, \]
then there exists a sequence of functions \( \{q_n(x)\} \) independent of \( f_0(x) \), such that \( |f_n^*(x) - q_n(x)| \overset{n}{\rightarrow} 0 \) uniformly in \( x \) and \( f_0 \).

In order to prove this theorem, we will first need to prove the following lemmas.

**Lemma 4.1.2:** Under the conditions of Theorem 4.1.1, for all \( k \), the following holds uniformly in \( y \).
\[ \int \frac{M_{n+1,n+k}(x,y) \phi_{n+k}(y)}{\prod_{n+1} \lambda_i \phi_{n+k}(x)} - P_{n+1,n+k}(x,y) \mu(dx) = 0. \]

**Proof:** We proceed by induction. When \( k=1 \), the result is trivial since by (3.1.3) the integrand is zero.

Now assume the result is true for \( k \), and consider

\[ \int \frac{M_{n+1,n+k+1}(x,y) \phi_{n+k+1}(y)}{\prod_{n+1} \lambda_i \phi_{n+k+1}(x)} - P_{n+1,n+k+1}(x,y) \mu(dx) \]

\[ = \int \left\{ \frac{M_{n+1,n+k}(x,z) \phi_{n+k}(z)}{\prod_{n+1} \lambda_i} \frac{M_{n+k+1}(z,y) \phi_{n+k+1}(y)}{\phi_{n+k+1}(z)} \frac{\phi_{n+k+1}(z)}{\phi_{n+k+1}(x)} \right\} \mu(dz) \mu(dx) \]

\[ = \int \left\{ \frac{M_{n+1,n+k}(x,z)}{\prod_{n+1} \lambda_i} \left( \frac{\phi_{n+k+1}(z)}{\phi_{n+k+1}(x)} - \frac{\phi_{n+k}(z)}{\phi_{n+k}(x)} \right) + \frac{M_{n+1,n+k}(x,z)}{\prod_{n+1} \lambda_i} \frac{\phi_{n+k}(z)}{\phi_{n+k}(x)} \right\} \mu(dz) \mu(dx) \]

\[ - \int P_{n+1,n+k}(x,z) P_{n+k+1}(z,y) \mu(dz) \mu(dx) \]
\[ \Delta' \left\{ \int \int \frac{M_{n+1,n+k}(x,z)}{n+k \prod_{i=n+1}^{n+k} \lambda_i} \left| \frac{\phi_{n+k+1}(z)}{\phi_{n+k+1}(x)} - \frac{\phi_{n+k}(z)}{\phi_{n+k}(x)} \right| \mu(dz) \mu(dx) \right\} \]

\[ + \Delta' \left\{ \int \int \frac{M_{n+1,n+k}(x,z)}{n+k \prod_{i=n+1}^{n+k} \lambda_i} \left| \frac{\phi_{n+k}(z)}{\phi_{n+k}(x)} - P_{n+1,n+k}(x,z) \right| \mu(dz) \mu(dx) \right\} \]

(4.1.15)

where \( \Delta' = \Delta B/v_b \) and the inequality follows from (4.1.12).

Considering the first term of (4.1.15) and using bounds (4.1.7) and (4.1.10), we find

\[ \text{1st term} \leq \frac{\Delta' k \mu(S)^{k-1}}{v^k} \int \left\{ \left| \phi_{n+k+1}(z) - \phi_{n+k}(z) \right| \frac{\phi_{n+k+1}(z)}{\phi_{n+k+1}(x)} - \frac{\phi_{n+k}(z)}{\phi_{n+k}(x)} \right| \mu(dz)\mu(dx) \}. \]

(4.1.16)

Using the bounds in (4.1.3) the integrand in (4.1.16) can be expressed as follows, using \( j = n+k \)

\[ \left| \phi_{j+1}(z) - \phi_j(z) \right| \frac{\phi_{j+1}(z)}{\phi_j(z)} \]

\[ = \left| \frac{\phi_{j+1}(z) \phi_j(z) - \phi_j(z) \phi_{j+1}(z) + \phi_j(z) \phi_j(z) - \phi_j(z) \phi_{j+1}(z)}{\phi_{j+1}(x) \phi_j(x)} \right| \]

\[ \leq \frac{\phi_{j+1}(z) - \phi_j(z)}{\phi_{j+1}(x)} + \frac{\phi_j(z)}{\phi_{j+1}(x)} \phi_{j+1}(x) - \phi_j(z) \phi_{j+1}(x) \]

\[ \leq \frac{1}{B} |\phi_{j+1}(z) - \phi_j(z)| + \frac{B}{2} |\phi_j(z) - \phi_{j+1}(z)|. \]
Using this inequality in (4.1.16), we get

\[
\int \int | \frac{\phi_{j+1}(z)}{\phi_j(x)} - \frac{\phi_j(z)}{\phi_j(x)} | u(dx) u(dx) \leq \frac{1}{b} u(S) \int | \phi_{j+1}(z) - \phi_j(z) | u(dx)
\]

which tends to zero by the hypothesis of Theorem 4.1.1.

Now consider the second term of (4.1.15) and note that the integrand is bounded independently of \( n \). Also the space \( S \) is of finite measure, hence

\[
\int \int \frac{\mathcal{M}_{n+1,n+k}(x,z)}{\phi_{n+1}(x)} \frac{\phi_{n+k}(z)}{\phi_{n+k}(x)} - \mathcal{M}_{n+1,n+k}(x,z) | u(dx) u(dx) \]

\[
\frac{B}{b^2} u(S) \int | \phi_j(x) - \phi_{j+1}(x) | u(dx)
\]

tends to zero by the bounded convergence theorem and the induction hypothesis. \( \square \)

**Lemma 4.1.3**: Under the conditions of Theorem 4.1.1, given \( \varepsilon > 0 \), there exists a sequence of functions \( \{ t_n(y) \} \) such that for all \( n \geq N(\varepsilon), | f_n^*(y) - t_n(y) | < \varepsilon \), independently of the choice of \( f_0 \) or \( y \).

**Proof**: Define \( s_n^k(y) = \max_{x} P_{n+1,n+k}(x,y) \). Since \( P_n(x,y) \) is bounded for all \( n \) by \( \Delta' \), Corollary 2.2.3 can be applied, so that given \( \gamma > 0 \), for \( k \geq N(\gamma) \),
Now define

\[ r_n^k(y) = \frac{s_n^k(y)}{\phi_n^k(y)} \text{ and } R_n^k = \int r_n^k(y) \mu(dy). \]

We will show that for suitable \( k_0 \), the appropriate sequence of functions \( \{t_n(y)\} \) can be defined by

\[ t_{n+k_0}(y) = \frac{r_n^k(y)}{R_n^k}. \]

So consider, for any \( k \),

\[ |f^*_n(y) - \frac{r_n^k(y)}{R_n^k}|. \tag{4.1.18} \]

If we let \( (\phi, f) = \int \phi(x)f(x)\mu(dx) \), then (4.1.18) is

less than or equal to

\[
\left| f^*_n(y) - \frac{f^*_n(y) \int f_{n+k}(y) \mu(dy) \lambda_{n+k} \phi_{n+k} \frac{f_n(y)}{R_n^k}}{n+k \phi_{n+k} \frac{f_n(y)}{R_n^k}} \right| \\
\left| f^*_n(y) - \frac{f_n(y)}{R_n^k} \right| \\
\left| \sup_y f^*_n(y) \left| 1 - \frac{\int f_{n+k}(y) \mu(dy) \lambda_{n+k} \phi_{n+k} \frac{f_n(y)}{R_n^k}}{n+k \phi_{n+k} \frac{f_n(y)}{R_n^k}} \right| \right|
\]
Now if we let

\[ A^k_n(y) = \frac{f_{n+k}(y)}{n+k} \prod_{i=n+1}^{n+k} \frac{R^k_n(\phi_{n+k}, f_n)}{\lambda_i} - \frac{r^k_n(y)}{R_n^k} \] (4.1.19)

then (4.1.19) becomes

\[ \sup_y f^*_n(y) |A^k_n(y)| \mu(dy) + |A^k_n(y)| \] (4.1.20)

If we can show that \(|A^k_n(y)|\) can be made small, then it will be easy to show that the first term can also be made small. We first find lower bounds for \(R^k_n\) and \((\phi_{n+k}, f^*_n)\):

\[ R^k_n = \int [s^k_n(y)/\phi_{n+k}(y)] \mu(dy) \geq \frac{1}{B} \int s^k_n(y) \mu(dy) \]

\[ = \frac{1}{B} \sup_x P_{n+1,n+k}(x,y) \mu(dy) \geq \frac{1}{B} P_{n+1,n+k}(x_0,y) \mu(dy) \]

\[ = \frac{1}{B} \]

where \(x_0\) is any point in \(S\).

\[ (\phi_{n+k}, f^*_n) = \int \phi_{n+k}(y) f^*_n(y) \mu(dy) \]

\[ \geq b \int f^*_n(y) \mu(dy) = b. \]
\[ |A_n^k(y)| = \left| \frac{f_{n+k}(y)}{\prod_{i=1}^{n+k} \lambda_i \int f_n(y) \mu(dy)} - r_n^k(y) \frac{\phi_{n+k}(y)}{\phi_n(y)} \right| \]

\[ \leq \frac{B}{b} \left| \frac{f_{n+k}(y)}{\prod_{i=1}^{n+k} \lambda_i \int f_n(y) \mu(dy)} - r_n^k(y) \frac{\phi_{n+k}(y)}{\phi_n(y)} \right| \]

\[ = \frac{B}{b} \left| \prod_{i=1}^{n+k} \lambda_i \int f_n(y) \mu(dy) \mu(dx) \right| \]

\[ \leq \frac{B}{b} \sup_x f_n^*(x) \left| \frac{M_{n+1,n+k}(x,y)}{\prod_{i=1}^{n+k} \lambda_i} - r_n^k(y) \frac{\phi_{n+k}(x)}{\phi_n(y)} \right| \mu(dx) \]

\[ \leq \frac{B}{b} \sup_x f_n^*(x) \left| \frac{M_{n+1,n+k}(x,y)}{\prod_{i=1}^{n+k} \lambda_i} - s_n^k(y) \frac{\phi_{n+k}(x)}{\phi_n(y)} \right| \mu(dx) \]

\[ = \frac{B}{b} \sup_x f_n^*(x) \left| \frac{\phi_{n+k}(x)}{\phi_n(y)} \right| \left( \frac{M_{n+1,n+k}(x,y) \phi_{n+k}(y)}{\prod_{i=1}^{n+k} \lambda_i} - s_n^k(y) \frac{\phi_{n+k}(x)}{\phi_n(y)} \right) \mu(dx) \]

\[ - s_n^k(y) \mu(dx) \]
Applying Equation (4.1.9), this can be written

\[
\leq \frac{B^2}{b^2} \sup_x f_n'(x) \int \left| \frac{M_{n+1,n+k}(x,y) \phi_{n+k}(y)}{\prod_{n+1} \lambda_i \phi_{n+k}(x)} \right| \mu(dx)
\]

\[= P_{n+1,n+k}(x,y) + P_{n+1,n+k}(x,y) - s_{n}^{k}(y) \left| \mu(dx) \right|.
\]

Now, according to (4.1.17)

\[|s_{n}^{k}(y) - P_{n+1,n+k}(x,y)| < \Delta' \delta^{k-1},\]

so choose \(k_0\) so that \(\Delta' \delta^{k_0-1} \mu(S) < \gamma\), then

\[\int |P_{n+1,n+k_0}(x,y) - s_{n}^{k_0}(y)| \mu(dx) < \Delta' \delta^{k_0-1} \mu(S) < \gamma.
\]

Also, from Lemma 4.1.2, given \(k_0\), there exists \(N=N(k_0,\gamma)\) such that \(n \geq N(k_0,\gamma)\) implies

\[\int \left| \frac{M_{n+1,n+k_0}(x,y) \phi_{n+k_0}(y)}{\prod_{n+1} \lambda_i \phi_{n+k_0}(x)} \right| \mu(dx) < \gamma.
\]

Hence

\[\frac{B^2}{b^2} \sup_x f_n'(x) \int \left| \frac{M_{n+1,n+k}(x,y) \phi_{n+k}(y)}{\prod_{n+1} \lambda_i \phi_{n+k}(x)} \right| \mu(dx) < \gamma.
\]
\[ |A_n^0(y)| \leq \frac{b^2 \Delta}{v^2 b} (y+\gamma) = c_1 \gamma \]

where \( c_1 = \frac{2b^2 \Delta}{b^2 v} \).

Next consider \( |\int A_n^0(y) \mu(dy)| \), for \( n \geq N(k_0, \gamma) \)

\[ |\int A_n^0(y) \mu(dy)| \leq \int |A_n^0(y)| \mu(dy) \leq c_1 \mu(S) . \]

Consequently for \( n \geq N(k_0, \gamma) \), it follows from (4.1.20) that

\[ |f_{n+k_0}^* (y) - r_n^0(y)/R_n^0| \leq \frac{\Delta}{v} c_1 \mu(S) \gamma + c_1 \gamma = c_2 \gamma , \]

where \( c_2 = \frac{\Delta}{v} c_1 \mu(S) + c_1 \).

Finally, if \( \epsilon \) is given and if \( \gamma \) is chosen to be \( \epsilon/c_2 \), then if \( t_{n+k_0}^0(y) = r_n^0(y)/R_n^0 \) and if \( n \geq N(k_0, \gamma) \),

\[ |f_{n+k_0}^* (y) - t_{n+k_0}^0(y)| < \epsilon . \]

Using Lemma 4.1.3, it is an easy matter to prove

Theorem 4.1.1.

Proof of Theorem 4.1.1: Let \( \{\epsilon_j\} \) be a sequence of constants decreasing to zero. From Lemma 4.1.3, there are (increasing) sequences \( \{k_j\} \) and \( \{N(k_j, \epsilon_j)\} \) and sequences of functions \( \{t_n^{(j)}\} \) such that \( n \geq N_{-j} = N(k_j, \epsilon_j) \) implies

\[ |f_{n+k_j}^* (y) - t_n^{(j)} (y)| < \epsilon_j . \]
Define the sequence \( \{q_n(y)\} \) as follows:

\[
q_n(y) = \begin{cases} 
1 & \text{or any arbitrary function, for } n = 1, 2, \ldots, N_1 + k_1 \\
t_n(j) & n = N_j + k_j + 1, \ldots, N_{j+1} + k_{j+1}
\end{cases}
\]

By construction, the sequence \( \{q_n(y)\} \) satisfies the conclusion of Theorem 4.1.1. 

It is clear that as a result of Theorem 4.1.1, kernels satisfying the conditions of Theorem 4.1.1 will be pointwise weakly ergodic in the sense of Definition 4.1.1.

Note that if the sequence of kernels under consideration is stochastic, then \( \phi_n(x) = 1 \) is a positive bounded right eigenfunction for all \( n \), hence \( \int |\phi_n(x) - \phi_{n+1}(x)| \mu(dx) = 0 \). Also \( \int m_n(x, y) \mu(dy) = 1 \) for all \( n \) implies that condition (4.1.2) is satisfied. Hence, to apply Theorem 4.1.1, only conditions (4.1.1) and (4.1.5) need be verified.

B. Strong Ergodicity

In this section we give sufficient conditions for strongly ergodic behavior in the pointwise sense as described in the following definition.

**Definition 4.2.1:** A sequence of non-negative kernels will be called pointwise strongly ergodic if there exists a function \( q(y) \) and sequences of constants \( k(f_0, m, n) \) such
that for all \( m \),

\[
\sup_{f_0} \left| f_{m,n}^\star (y) - q(y) \right| \to 0 \text{ uniformly in } y
\]

where

\[
\int q(y) u(dy) > 0.
\]

As indicated after Definition 4.1.1, the proofs given in this chapter allow \( m \) to be taken to be 1 and

\[ k(f_0, 1, n) = \int f_n(y) u(dy). \]

We first prove some results relating to pointwise weak ergodicity and from these results we will be able to obtain sufficient conditions for pointwise strong ergodicity.

In Equation (1.1.2) we defined kernel superpositions and introduced the notation \( M_{m,n}(x,y) \) to mean the superposition of kernels \( M_m, M_{m+1}, \ldots, M_n \). We now extend the notation so that if \( m > n \), then \( M_{m,n} \) will mean the superposition of kernels \( M_m, M_{m-1}, \ldots, M_n \). We will use this notation in this section.

When positive left eigenfunctions exist for a non-negative kernel, we will define

\[
R(x,y) = \frac{\Psi(x) M(x,y)}{\lambda \Psi(y)}.
\]

In this notation, \( Q(x,y) \) defined in (3.1.4) would be \( R(y,x) \). Further, considering superpositions, it is clear that
In the proof of Lemma 4.1.1, we stated that \( \psi_n \) is integrable. Without loss of generality, we can take the version of \( \psi_n \) which integrates to 1, i.e., we will assume

\[
\int \psi_n(y) \mu(dy) = 1.
\]

We now summarize the assumptions necessary for the theorems to follow as Condition S. We then give some lemmas which will be used in proving Theorem 4.2.1.

**Condition S:** A sequence of kernels satisfying (4.1.1), (4.1.2), (4.1.4), and (4.1.6) will be said to satisfy Condition S.

**Lemma 4.2.1:** Let \( \{M_n(x,y)\} \) be a sequence of kernels satisfying Condition S. If

\[
\int |\psi_n(x) - \psi_{n+1}(x)| \mu(dx) \overset{n}{\rightarrow} 0,
\]

then for every \( k \), the following holds uniformly in \( y \).

\[
\int \left| \frac{M_{n+1,n+k}(x,y) \psi_{n+k}(x)}{n+k} - R_{n+1,n+k}(x,y) \right| \mu(dx) \overset{n}{\rightarrow} 0.
\]

**Proof:** The proof follows exactly that of Lemma 4.1.2. \( \square \)
In the next lemma, we give a bound which will be useful in this section.

**Lemma 4.2.2**: Let \( \{M_n(x, y)\} \) be a sequence of kernels satisfying Condition S. If we define

\[
s_n^k(x) = \sup_{y} Q_{n+k, n+1}(y, x),
\]

then

\[
s_n^k(x) \geq \left[ \frac{d\nu}{dA\mu(S)} \right]^k \cdot \frac{1}{\mu(S)} > 0.
\]

**Proof**: We use the bounds (4.1.4) and (4.1.10) to get

\[
s_n^k(x) = \sup_{y} \int \ldots \int Q_{n+k}(y, z_k) \ldots Q_{n+1}(z_2, x) \mu(dz_k) \ldots \mu(dz_2)
\]

\[
= \sup_{y} \int \ldots \int \frac{M_{n+k}(z_k, y) \ldots M_{n+1}(x, z_2)}{\prod_{i=n+1}^{n+k} \lambda_i} \frac{\psi_{n+k}(z_k)}{\psi_{n+k}(y)} \frac{\psi_{n+1}(x)}{\psi_{n+1}(z_2)} \mu(dz_k) \ldots \mu(dz_2)
\]

\[
\geq \left[ \frac{d}{d\Delta \mu(S)} \right]^k \frac{1}{\prod_{i=n+1}^{n+k} \lambda_i} \sup_{y} \int \ldots \int M_{n+1}(x, z_2) \ldots M_{n+k}(z_k, y) \mu(dz_2) \ldots \mu(dz_k)
\]

\[
\geq \left[ \frac{d}{d\Delta \mu(S)} \right]^k \sup_{y} M_{n+1, n+k}(x, y).
\]
In view of this last inequality, it suffices to show that \( \sup_{y} \frac{M_{n+1,n+k}(x,y)}{\mu(S)} \geq \frac{\nu^{k}}{\mu(S)} \). Consider \( M_{n+1,n+k}(x,S) \).

Using Fubini's theorem and (4.1.2), we get

\[
M_{n+1,n+k}(x,S) = \int M_{n+1,n+k}(x,z) \mu(dz)
\]

\[
= \int \int M_{n+1,n+k-1}(x,y) M_{n+k}(y,z) \mu(dy) \mu(dz)
\]

\[
= \int \int M_{n+1,n+k-1}(x,y) M_{n+k}(y,z) \mu(dz) \mu(dy)
\]

\[
= \int M_{n+1,n+k-1}(x,y) M_{n+k}(y,S) \mu(dy)
\]

\[
\geq \nu \int M_{n+1,n+k-1}(x,y) \mu(dy)
\]

\[
\geq \nu^{2} \int M_{n+1,n+k-2}(x,y) \mu(dy) \ldots \geq \nu^{k}.
\]

Now, since \( M_{n+1,n+k}(x,S) \geq \nu^{k} \), it must be that

\[
\sup_{y} \frac{M_{n+1,n+k}(x,y)}{\mu(S)} \geq \frac{\nu^{k}}{\mu(S)}, \quad \text{for if not, then} \quad \sup_{y} \frac{M_{n+1,n+k}(x,y)}{\mu(S)} < \frac{\nu^{k}}{\mu(S)}, \quad \text{which means}
\]

\[
M_{n+1,n+k}(x,S) = \int M_{n+1,n+k}(x,y) \mu(dy)
\]

\[
\leq \sup_{y} \frac{M_{n+1,n+k}(x,y)}{\mu(S)} < \nu^{k}
\]

which is a contradiction. \( \square \)
Theorem 4.2.1: Let \( \{M_n(x,y)\} \) be a sequence of kernels satisfying Condition S. If \( \int |\psi_n(x)-\psi_{n+1}(x)| \mu(dx) \xrightarrow{n} 0 \), then

\[
|f^*_n(x)-\psi_n(x)| \xrightarrow{n} 0 \text{ uniformly.}
\]

Proof: Define

\[ t^k_n(x) = s^k_n(x)f_n(x)/\psi_{n+k}(x) \quad \text{and} \quad T^k_n = \int t^k_n(x) \mu(dx). \]

Then

\[
|f^*_{n+k}(x)-\psi_{n+k}(x)| = |f^*_{n+k}(x) - \frac{f_{n+k}(x)}{\prod \lambda_i T^k_{n+1}} + \frac{f_{n+k}(x)}{\prod \lambda_i T^k_{n+1}} - \psi_{n+k}(x)|
\]

\[
\leq \left| f^*_{n+k}(x) - \frac{f_{n+k}(x)}{\prod \lambda_i T^k_{n+1}} \int f_{n+k}(x) \mu(dx) \right| + \left| \frac{f_{n+k}(x)}{\prod \lambda_i T^k_{n+1}} - \psi_{n+k}(x) \right|
\]

\[
\leq \sup_x f^*_{n+k}(x) \int f_{n+k}(x) \mu(dx) + |\psi_{n+k}(x)|. \quad (4.2.3)
\]
where $A_n^k(x) = \frac{f_{n+k}(x)}{n+k} - \psi_{n+k}(x)$. Note that here we use the fact that $\psi_{n+k}(x)$ integrates to 1. As in the proof of Theorem 4.1.1, we will show that for some $k_0$, $|A_n^0(x)|$ will be small for $n$ sufficiently large. Now by Lemma 4.2.2,

$$\int \frac{s_n^k(y) f_n^*(y)}{\psi_{n+k}(y)} u(dy) \geq \frac{1}{D} \int s_n^k(y) f_n^*(y) u(dy)$$

$$\geq \left[ \frac{dv}{D \Delta \mu(S)} \right]^k \frac{1}{\mu(S) D} \int f_n^*(y) u(dy) = \left[ \frac{dv}{D \Delta \mu(S)} \right]^k / \mu(S) D \equiv c_1.$$

We use this inequality in the following

$$|A_n^k(x)| = \frac{f_{n+k}(x) \int f_n(y) u(dy)}{n+k} \frac{T_n^k \int f_n(y) u(dy) \psi_{n+k}(x)}{\prod_{i=n+1}^{n+k} \lambda_i \int f_n(y) u(dy)}$$

$$= \frac{1}{\int s_n^k(y) f_n^*(y) \psi_{n+k}(y) u(dy)} \frac{f_{n+k}(x)}{n+k} \frac{\prod_{i=n+1}^{n+k} \lambda_i \int f_n(y) u(dy)}$$

$$- \int s_n^k(y) f_n^*(y) \psi_{n+k}(x) \psi_{n+k}(y) u(dy)$$

$$\leq \frac{1}{c_1} \frac{f_{n+k}(x)}{n+k \prod_{i=n+1}^{n+k} \lambda_i \int f_n(y) u(dy)} - \int s_n^k(y) f_n^*(y) \psi_{n+k}(x) \psi_{n+k}(y) u(dy)$$
Using inequalities (4.1.4) and (4.1.9), we can say that (4.2.4) is less than or equal to

$$\frac{\Delta D}{dV_{c_1}} \left\{ \left| \frac{M_{n+1,n+k}(y,x)}{n+k} \; \lambda_i \; \psi_{n+k}(y) \right| - s_n^k(y) \right\} \mu(dy)$$

$$\leq \frac{\Delta D}{dV_{c_1}} \left\{ \left| \frac{M_{n+1,n+k}(y,x)}{n+k} \; \lambda_i \; \psi_{n+k}(x) \right| - P_{n+1,n+k}(y,x) \right\} \mu(dy)$$

$$\quad + \left| P_{n+1,n+k}(y,x) - s_n^k(y) \right| \mu(dy) \right\}$$

(4.2.5)

Now, in view of (4.2.1) and (4.2.2), the second term of (4.2.5) is
Using the bound $\Delta' = \Delta D / \nu d$ from (4.1.11) and using Corollary 2.2.3, we know that given $\gamma > 0$, there exists $k_0$ such that

$$|Q_{n+k_0, n+1}(x, y) - \sup_x Q_{n+k_0, n+1}(x, y)| < \Delta' \delta^{k_0-1} \nu(S) \gamma / \mu(S)$$

Then for every $x$,

$$\int \left| Q_{n+k_0, n+1}(x, y) - \sup_x Q_{n+k_0, n+1}(x, y) \right| \mu(dy) < \Delta' \delta^{k_0-1} \nu(S) \gamma .$$

By Lemma 4.2.1, given $k_0$ and $\gamma$, there exists some $N = N(k_0, \gamma)$, such that for all $n > N$,

$$\int \left| M_{n+1, n+k_0}(y, x) \psi_{n+k_0}(y) \right| \lambda_i \psi_{n+k_0}(x) \mu(dy) < \gamma$$

Combining (4.2.6) and (4.2.7), we have that for $n > N(k_0, \gamma)$,

$$|A_{n}^{k_0}(x)| \leq 2\gamma \Delta D / \nu d \leq c_2 \gamma$$

and

$$\int |A_{n}^{k_0}(x)| \mu(dx) \leq \int |A_{n}^{k_0}(x)| \mu(dx) \leq c_2 \gamma \mu(S),$$

where $c_2 = 2\Delta D / \nu d$.

Finally, in view of (4.2.3), defining $c_3 = c_2 (\Delta \nu \mu(S) + 1)$,
we can say that given $\varepsilon < 0$, choose $\gamma < \varepsilon / c_3$. Then there exists $k_0$ and $N(k_0, \varepsilon)$ such that if $n > N(k_0, \varepsilon)$, then

$$|f^*_{n+k_0}(x) - \psi_{n+k_0}(x)| \leq \sup_x f^*_{n+k_0}(x) \int A_n^k(x) \mu(dx) + \frac{k_0}{n} \leq \frac{A}{V}[C\gamma \mu(S)] + C_2 \gamma = c_3 \gamma < \varepsilon.$$

Note that in this proof, $k_0$ and $N(k_0, \varepsilon)$ are chosen independently of $f_0$ and $x$.

In view of the fact that $\mu(S) < \infty$, the following corollary can be given.

**Corollary 4.2.1:** Let $\{M_n(x,y)\}$ be a sequence of kernels satisfying Condition S. If for each $x$,

$$|\psi_n(x) - \psi_{n+1}(x)| \to 0,$$

then for every starting function $f_0$,

$$|f^*_n(x) - \psi_n(x)| \to 0 \text{ uniformly.}$$

**Proof:** The pointwise condition (4.2.8) and the fact that $\psi_n$ is bounded allows the bounded convergence theorem to be applied so that

$$\int |\psi_n(x) - \psi_{n+1}(x)| \mu(dx) \to 0.$$ 

We can now state the following theorem giving sufficient conditions for pointwise strongly ergodic behavior.

**Theorem 4.2.2:** Let $\{M_n(x,y)\}$ be a sequence of kernels satisfying Condition S. If the sequence of left
eigenfunctions \{\psi_n(x)\} converges uniformly to \psi(x), say, then \{M_n\} is pointwise strongly ergodic.

**Proof:** Corollary 4.2.1 holds, so consider

$$|f^*_n(x) - \psi(x)| \leq |f^*_n(x) - \psi_n(x)| + |\psi_n(x) - \psi(x)|.$$  

The first term goes to zero at a rate independent of the choice of \(f_0\), and the second term goes to zero, uniformly in \(x\) by assumption. \(\square\)

Now, in view of Corollary 4.2.1, under Condition S, it is true that

$$|\psi_n(x) - \psi_{n+1}(x)| \to 0 \Rightarrow |f^*_n(x) - f^*_{n+1}(x)| \to 0. \quad (4.2.9)$$

We can show that under conditions similar to Condition S, the converse of (4.2.9) also holds.

**Condition T:** A sequence of kernels satisfying Condition S and such that the eigenvalue \(\lambda_n\), corresponding to eigenfunctions \(\psi_n(x)\) and \(\phi_n(y)\), is simple and such that \(\phi_n(y) \geq 0\) is integrable will be said to satisfy Condition T.

Note that by a simple eigenvalue we mean one for which there are unique left and right eigenfunctions, up to constant multiples. Also, if \(\phi_n(y)\) is integrable, so is \(\psi_n(y)\phi_n(y)\), so we can assume that

$$\int \psi_n(y)\phi_n(y)\mu(dy) = 1.$$
We remarked earlier in this chapter that the conditions we impose are weaker than those imposed by Conn. The explicit requirement that $\lambda_n$ be simple is satisfied, according to Harris (1963), for all primitive kernels, hence this requirement would be satisfied under Conn's conditions, and in fact would be met for a larger class of kernels.

**Lemma 4.2.3:** Let $\{M_n(x,y)\}$ be a sequence of kernels satisfying Condition T. For a given starting function $f_0$, define

$$\rho_n = \frac{\int f_{n-1}(y)\mu(dy)}{\int f_n(y)\mu(dy)}.$$  

If

$$\int |f_n^*(y) - f_{n+1}^*(y)| \mu(dy) \to 0,$$

then for all $k$,

$$|f_n^*(y) - \rho_n f_n^*(x)M_n^k(x,y)\mu(dx)| \to 0,$$

uniformly in $y$.

**Proof:** We first show that $\rho_n$ is bounded by positive constants. From Equations (4.1.13) and (4.1.14),

$$\int f_n(y)\mu(dy) \geq v \int f_{n-1}(y)\mu(dy),$$

$$\int f_n(y)\mu(dy) \leq \Delta \mu(S) \int f_{n-1}(y)\mu(dy),$$

hence

$$\frac{1}{\Delta \mu(S)} \leq \rho_n \leq \frac{1}{v}. \quad (4.2.10)$$
Now proceed by induction. When $k = 1,$

$$\left| f_n^*(y) - \rho_n \int f_n^*(x) M_n(x, y) \mu(dx) \right|$$

$$= \left| \int \frac{f_n(x) M_n(x, y)}{f_n(y) \mu(dy)} \mu(dx) - \rho_n \int f_n^*(x) M_n(x, y) \mu(dx) \right|$$

$$= \rho_n \left| \int [f_n^*(x) - f_n(x)] M_n(x, y) \mu(dx) \right|$$

$$\leq \Delta \rho \int |f_n^*(x) - f_n(x)| \mu(dx)$$

$$\leq \frac{\Delta}{V} \int |f_n^*(x) - f_n(x)| \mu(dx).$$

This last expression tends to zero by the hypothesis. This clearly goes to zero uniformly in $y$ since the last expression is independent of $y.$

Now assume that the result holds for $k$ and show that it holds for $k + 1.$

$$\left| \rho_n^{k+1} \int f_n^*(x) M_n^{k+1}(x, y) \mu(dx) - f_n^*(y) \right|$$

$$= \left| \rho_n \int \rho_n^k f_n^*(x) \int M_n^k(x, z) M_n(z, y) \mu(dz) \mu(dx) \right.$$
\[ \leq \frac{\Delta}{\nu} \int \left| \int \rho_n f_n^*(x) M_n(x, z) u(dx) - f_n^*(z) u(dz) \right| u(dz) \]

\[ + \int \left| f_n^*(z) - f_{n-1}^*(z) \right| u(dz) \]  

\[ (4.2.11) \]

By the induction hypothesis, the integrand of the first term of (4.2.11) goes to zero uniformly in \( z \). Since the space is of finite measure, the first term goes to zero. The second term of (4.2.11) goes to zero by the hypothesis of the lemma. Hence the expression (4.2.11) goes to zero with \( n \) independently of the choice of \( y \). \[ \square \]

In Chapter I, we stated the result (1.2.1) of Harris which holds for primitive kernels. The next lemma gives the same result under different conditions. We do not require primitivity, but we do require the kernel to be bounded and for the transformed kernel \( Q \) to have positive \( \alpha(Q) \). All of the other conditions we require are implied by primitivity. Note that in this lemma we consider the stationary case.

**Lemma 4.2.4:** Let \( M(x, y) \) be a kernel which satisfies Condition T. Then

\[ \left| \frac{M^k(x, y)}{\lambda_k} - \phi(x) \psi(y) \right| < \frac{D}{d} \Delta' \delta^{k-1} \]

where \( \Delta' = D\Delta/d\nu \) and \( 0 < \delta < 1 \).
Proof: We first comment that it is not necessary for \( \phi(x) \) to be bounded, only that it be integrable. As indicated earlier, this allows \( \phi \) and \( \psi \) to be chosen in such a way that
\[
\int \phi(x) \psi(x) \mu(dx) = 1.
\]

From (4.1.11) \( Q(x,y) \) is bounded by \( \Delta' = \frac{D\Delta}{dv} \). By assumption (4.1.6), \( \delta(Q) = \delta < 1 \), that is, \( \alpha(Q) > 0 \). Hence we can apply corollaries 2.2.3 and 2.2.4 and say that
\[
\sup_x Q^k(x,y) - \inf_x Q^k(x,y) \leq \Delta' \delta^{k-1}
\]
and that there exists \( q(y) = \lim_{k \to \infty} Q^k(x,y) = \lim_{k \to \infty} \frac{Q^k(x,y)}{\sum_{k=0}^{\infty} \frac{Q^k(x,y)}{\lambda}} \)
which is a left eigenfunction of \( Q(x,y) \) corresponding to the eigenvalue 1. We also know that \( q(y) \) integrates to 1, since if stochastic kernels converge, they converge to a constant stochastic kernel.

As part of the hypothesis, we have that \( \phi(x) \) is a right eigenfunction corresponding to the simple eigenvalue \( \lambda \). However, we now show that \( \frac{q(x)}{\psi(x)} \) is also a right eigenfunction for the same eigenvalue.

\[
\int M(x,y) \frac{q(y)}{\psi(y)} \mu(dy) = \int \frac{\lambda \psi(y) Q(y,x)}{\psi(x)} \frac{q(y)}{\psi(y)} \mu(dy)
\]
\[
= \frac{\lambda}{\psi(x)} \int q(y) Q(y,x) \mu(dy) = \lambda \frac{q(x)}{\psi(x)}.
\]
Now, since $\lambda$ was assumed to be simple, it must be that

$$\phi(x) = \frac{cq(x)}{\psi(x)}$$

i.e., $cq(x) = \phi(x)\psi(x)$. However, since both $q(x)$ and $\phi(x)\psi(x)$ integrate to 1, $c$ must be 1, hence

$$q(x) = \phi(x)\psi(x).$$

It is easy to see that $Q^k(y,x) = \frac{\psi(x)M^k(x,y)}{\lambda^k\psi(y)}$. From Corollary 2.2.3, $|Q^k(y,x) - q(x)| \leq A \delta^{k-1}$, hence

$$\left|\frac{\psi(x)M^k(x,y)}{\lambda^k\psi(y)} - \phi(x)\psi(x)\right| \leq A \delta^{k-1}$$

and so

$$\left|\frac{M^k(x,y)}{\lambda^k} - \phi(x)\psi(y)\right| \leq \frac{\psi(y)}{\psi(x)} A \delta^{k-1} \leq \frac{DA'}{D} \delta^{k-1}.$$ 

We note that if the conditions of Lemma 4.2.4 are changed so that the transformed $Q$ is required to be $\alpha$-primitive rather than being required to have positive $\alpha$, then it is still true that

$$\left|\frac{M^k(x,y)}{\lambda^k} - \phi(x)\psi(y)\right| \xrightarrow{k \to \infty} 0.$$
Theorem 4.2.3: Let \( \{M_n(x, y)\} \) be a sequence of kernels satisfying Condition T. For a given starting function \( f_0 \), if \( \int |f_n^*(y) - f_{n+1}^*(y)| u(dy) \to 0 \), then \( |f_n^*(y) - \psi_n(y)| \leq 0 \) uniformly.

Proof:
\[
|f_n^*(y) - \psi_n(y)| \\
\leq |f_n^*(y) - \frac{f_n^*(y)}{(\rho_n \lambda_n)^k(\phi_n, f_n)}| + \frac{f_n^*(y)}{(\rho_n \lambda_n)^k(\phi_n, f_n)}| - \psi_n(y)| \\
\leq \sup_y f_n^*(y) \left| 1 - \frac{1}{(\rho_n \lambda_n)^k(\phi_n, f_n)} \right| \\
+ \left| \frac{f_n^*(y)}{(\rho_n \lambda_n)^k(\phi_n, f_n)} - \psi_n(y) \right| \quad \text{(4.2.12)}
\]

Then since we chose \( \psi_n(y) \) to integrate to 1, if we define
\[
A_n^k(y) = \frac{f_n^*(y)}{(\rho_n \lambda_n)^k(\phi_n, f_n)} - \psi_n(y),
\]
it follows that (4.2.12) can be written as
\[
\sup_y f_n^*(y) \left| \int A_n^k(y) u(dy) \right| + |A_n^k(y)|.
\]

We will show that for some \( k_0 \) and \( n \) sufficiently large, \( |A_n^k(y)| \) can be made small.

\[
|A_n^k(y)| = \frac{1}{(\phi_n, f_n)} \left| \frac{f_n^*(y)}{(\rho_n \lambda_n)^k} - \psi_n(y)(\phi_n, f_n^*) \right| \\
\leq \frac{1}{(\phi_n, f_n)} \left| \frac{f_n^*(y)}{(\rho_n \lambda_n)^k} - \frac{f_n^*(x) M_n^k(x, y) u(dx)}{\lambda_n^k} \right| \quad \text{(4.2.13)}
\]
We now consider the various components of (4.2.13).

First, $(\phi_n, f_n^*) \geq b$, since

$$(\phi_n, f_n^*) = \int \phi_n(x) f_n^*(x) \mu(dx) \geq b \int f_n^*(x) \mu(dx) = b.$$ 

Also

$$\left| \int \frac{f_n^*(x) M_n^k(x, y) \mu(dx)}{\lambda_n^k} - \int f_n^*(x) \phi_n(x) \psi_n(y) \mu(dx) \right|$$

$$\leq \sup_x f_n^*(x) \left| \int \frac{M_n^k(x, y)}{\lambda_n^k} - \phi_n(x) \psi_n(y) \right| \mu(dx). \quad (4.2.14)$$

From Equation (4.2.14) and Lemma 4.2.4, given $\gamma > 0$, there exists $k_0$, such that for all $n$,

$$\left| \frac{M_n^k(x, y)}{\lambda_n^k} - \phi_n(x) \psi_n(y) \right| < \frac{D}{\Delta} \delta \frac{k_0 - 1}{\Delta \mu(S)}.$$ 

Hence (4.2.14) is less than

$$\frac{\Delta}{\nu} \int \frac{\psi(y)}{\Delta \mu(S)} \mu(dx) = \gamma.$$ 

Using the bounds in (4.2.10) for $\rho_n$, the bounds for $\lambda_n$, and Lemma 4.2.3, we know that for $k_0$ fixed and $\gamma > 0$ given, there exists an $N(k_0, \gamma)$ such that for $n \geq N$,

$$\left| f_n^*(y) - \rho_n \int f_n^*(x) M_n^k(x, y) \mu(dx) \right| < [\nu \Delta \mu(S)] k_0 \gamma.$$
Hence

\[
\left| \frac{f^*_n(y)}{(\rho_n \lambda_n)} - \frac{\int f^*_n(x) M^0_n(x,y) \mu(dx)}{\lambda_n} \right| = \frac{1}{(\rho_n \lambda_n)} \left| f^*_n(y) - \rho_n \int f^*_n(x) M^0_n(x,y) \mu(dx) \right|
\]

\[
< \frac{[vA\mu(S)]_{0,0}^0}{(\rho_n \lambda_n)} \gamma < \gamma.
\]

Now, using the above inequalities and choosing \( \gamma = \frac{eb}{2} \), we have that for \( n > N(k_0,y) \)

\[
|A^n_0(y)| \leq \frac{1}{(\phi_n, f^*_n)(y+y)} \leq \frac{2y}{b} = \varepsilon.
\]

Since (4.2.15) is independent of \( y \), it is true that

\[
\left| \int A^n_0(y) \mu(dy) \right| \leq \int |A^n_0(y)| \mu(dy) \leq \varepsilon \mu(S).
\]

Using (4.2.15) in Equation (4.2.12), we find for \( n > N(k_0,y) \),

\[
|f^*_n(y) - \psi_n(y)| \leq \sup_y f^*_n(y) \left| \int A^n_0(y) \mu(dy) \right| + |A^n_0(y)|
\]

\[
< \frac{A}{v} \varepsilon \mu(S) + \varepsilon = \frac{[\Delta \mu(S)]}{v} + 1 \varepsilon.
\]

We now state a corollary which is essentially the converse of (4.2.9).
Corollary 4.2.2: If Condition $T$ holds for $\{M_n(x,y)\}$ and if for some starting function $f_0$, $|f_n^*(y) - f_{n+1}^*(y)| \xrightarrow{n} 0$, then $|\psi_n(y) - \psi_{n+1}(y)| \xrightarrow{n} 0$ uniformly.

Proof: We know from Equation (4.1.9) that $f_n^*(y)$ is bounded, hence applying the bounded convergence theorem, $\int |f_n^*(y) - f_{n+1}^*(y)| u(\text{d}y)$ tends to zero. Thus Theorem 4.2.3 can be applied, so we know $|f_n^*(y) - \psi_n(y)| \xrightarrow{n} 0$. Hence

$$|\psi_n(y) - \psi_{n+1}(y)| \leq |\psi_n(y) - f_n^*(y)| + |f_n^*(y) - f_{n+1}^*(y)|$$

Since each of these terms goes to zero, the result follows.

The hypotheses of Corollary 4.2.2 require that for some $f_0(y)$, $|f_n^*(y) - f_{n+1}^*(y)| \xrightarrow{n} 0$. It is easy to see that if this condition does hold for some starting function, then it will hold for all starting functions $f_0(y)$. This follows from Corollary 4.2.2 and (4.2.9).

We can summarize these results by saying that under Condition $T$, $|\psi_n(y) - \psi_{n+1}(y)| \xrightarrow{n} 0$ if and only if $|f_n^*(y) - f_{n+1}^*(y)| \xrightarrow{n} 0$ for all starting functions $f_0(y)$. Further either condition implies $|f_n^*(y) - \psi_n(y)| \xrightarrow{n} 0$ uniformly. Hence the sequence $\{M_n(x,y)\}$ is pointwise strongly ergodic if and only if $\{\psi_n(y)\}$ converges uniformly.

In the next section, we find conditions which will
guarantee convergence of $|\psi_n(y) - \psi_{n+1}(y)|$ to zero. 

C. Convergence Theorems for Eigenvalues and Eigenfunctions

In this section we impose conditions which, although rather stringent, are weaker than those imposed by Conn.

For a given kernel $M(x,y)$, the dominant eigenvalue or dominant root is defined to be the eigenvalue which is simple, positive, and larger in absolute value than any other eigenvalue. Harris (1963) shows that if a kernel is primitive, it will possess a dominant root. From Frobenius's theorem (see, for example, Gantmacher (1959)), it is clear that non-negativity by itself is not sufficient to guarantee the existence of a dominant root. Since we wish to use a characterization of dominant roots given by Harris for primitive kernels, one of the assumptions in this section will be primitivity.

**Condition C:** A sequence of kernels satisfying Condition T and for which each kernel is primitive will be said to satisfy Condition C.

Note that Condition C is stronger than Condition T. Also, although primitivity implies the existence of positive bounds on the left eigenfunctions $\psi_n(y)$ corresponding to the dominant root $\lambda_n$, condition (4.1.4) imposes the additional restriction that there exist bounds which work
for all $n$.

**Theorem 4.3.1**: Let $\{M_n(x,y)\}$ be a sequence of kernels satisfying Condition C. If $\int |M_n(x,y)-M_{n+1}(x,y)| \mu(dx) \to 0$ uniformly in $y$, then $|\lambda_n - \lambda_{n+1}| \to 0$.

**Proof**: Harris (1964) gives the following characterization for the dominant root for primitive kernels. Let

$$S_n = \{\lambda_n > 0: \text{there exists a bounded, non-negative function } f(x) \text{ such that } \int f(x)M_n(x,y) \mu(dx) \geq \lambda_n f(y), \text{ but } \int f(x)M_n(x,y) \mu(dx) \neq \lambda_n f(y)\}.$$  

Then $\lambda_n = \operatorname{l.u.b.} \{S_n\}$.

Let $\varepsilon > 0$ be given. We will show that for $n$ sufficiently large, $(\lambda_{n+1} - \varepsilon, \varepsilon) \subset S_n$ and $(\lambda_n - \varepsilon, \varepsilon) \subset S_{n+1}$. If this is true, then $\lambda_{n+1} - \varepsilon \leq \lambda_n$ and $\lambda_n - \varepsilon \leq \lambda_{n+1}$, i.e., $|\lambda_n - \lambda_{n+1}| < \varepsilon$.

Choose $\gamma < d\varepsilon$. Then for all $n$, $\varepsilon \gamma_n(y) \geq d \gamma$. Now since $\int |M_n(x,y) - M_{n+1}(x,y)| \mu(dx) \to 0$, there exists $N = N(\gamma)$ such that if $n > N(\gamma)$, then for every $y$,

$$\int |M_n(x,y) - M_{n+1}(x,y)| \mu(dx) < \gamma/D.$$  

Then

$$\left|\int \psi_{n+1}(x)M_n(x,y) \mu(dx) - \int \psi_{n+1}(x)M_{n+1}(x,y) \mu(dx)\right|$$

$$\leq D \int |M_n(x,y) - M_{n+1}(x,y)| \mu(dx) < \gamma.$$  

Therefore,
\[
\int \psi_{n+1}(x)M_n(x,y)\mu(dx) \geq \int \psi_{n+1}(x)M_{n+1}(x,y)\mu(dx) - \gamma
\]

\[
= \lambda_{n+1}\psi_{n+1}(y) - \gamma > \lambda_{n+1}\psi_{n+1}(y) - \varepsilon\psi_{n+1}(y)
\]

\[
= (\lambda_{n+1} - \varepsilon)\psi_{n+1}(y)
\]

Since \(\psi_{n+1}(y)\) is a bounded non-negative function and

\[
\int \psi_{n+1}(x)M_n(x,y)\mu(dx) \geq (\lambda_{n+1} - \varepsilon)\psi_{n+1}(y),
\]

it follows that \((\lambda_{n+1} - \varepsilon)\in S_n\).

Likewise considering

\[
|\int \psi_n(x)M_n(x,y)\mu(dx) - \int \psi_n(x)M_{n+1}(x,y)\mu(dx)|
\]

it is easy to show that for \(n > N(y)\), \((\lambda_n - \varepsilon)\in S_{n+1}\), which is what we needed to show. \(\square\)

The following lemma will be useful in proving Theorem 4.3.2.

**Lemma 4.3.1:** Let \(\{M_n(x,y)\}\) be a sequence of kernels satisfying Condition C and such that

\[
\int |M_n(x,y) - M_{n+1}(x,y)|\mu(dx) \rightarrow 0
\]

uniformly in \(y\). Then for all \(k\), the following holds uniformly in \(y\).
\[
\int \left| \frac{M_n^k(x,y)}{\lambda_n^k} - \frac{M_n^{k+1}(x,y)}{\lambda_n^{k+1}} \right| \mu(dx) \xrightarrow{n \to 0} 0.
\]

**Proof:** For notational convenience, define

\[
K_n^k(x,y) = \frac{M_n^k(x,y)}{\lambda_n^k}.
\]

The proof follows by induction. Let \( k=1 \), then

\[
\int \left| \frac{M_n(x,y)}{\lambda_n} - \frac{M_n^1(x,y)}{\lambda_n} \right| \mu(dx)
\]

\[
= \int \left| \frac{M_n(x,y)}{\lambda_n} - \frac{M_n(x,y)}{\lambda_n} + \frac{M_n(x,y)}{\lambda_n} - \frac{M_n^1(x,y)}{\lambda_n} \right| \mu(dx)
\]

\[
\leq \int M_n(x,y) \left| \frac{1}{\lambda_n} - \frac{1}{\lambda_n} \right| \mu(dx)
\]

\[
+ \frac{1}{\lambda_n+1} \int |M_n(x,y) - M_n^1(x,y)| \mu(dx)
\]

\[
\leq \Delta \mu(S) \left| \frac{\lambda_n+1-\lambda_n}{\lambda_n+1} \right| + \frac{1}{\nu} \int |M_n(x,y) - M_n^1(x,y)| \mu(dx):
\]

\[(4.3.1)\]

Consider the first term of (4.3.1). By Theorem 4.3.1, \( |\lambda_{n+1} - \lambda_n| \xrightarrow{n \to 0} 0 \) and, from (4.1.10), \( 0 < \nu < \lambda_n \), hence the first term goes to zero. The second term goes to zero by the hypothesis.

Now assume that the result holds for \( k \) and show that it holds for \( k+1 \).
\[
\int |K_n^{k+1}(x,y) - K_n^{k+1}(x,y)| \mu(dx)
\]
\[
= \int \left| \int [K_n^k(x,z)K_n^k(z,y) - K_{n+1}^k(x,z)K_{n+1}^k(z,y)] \mu(dz) \right| \mu(dx)
\]
\[
\leq \int \left| \int |K_n^k(x,z)K_n^k(z,y) - K_{n+1}^k(x,z)K_{n+1}^k(z,y)| \mu(dz) \right| \mu(dx)
\]
\[
+ \int \left| \int |K_n^k(x,z)K_{n+1}^k(z,y) - K_{n+1}^k(x,z)K_{n+1}^k(z,y)| \mu(dz) \right| \mu(dx)
\]
\[\tag{4.3.2}\]

Now since \(K_n^k(x,z)\) is bounded (Equation (4.1.7)), the first term of (4.3.2) is less than or equal to
\[
[A^u(S)^{-v}] \int |K_n^k(z,y) - K_{n+1}^k(z,y)| \mu(dz) \mu(dx)
\]
\[
= [A^u(S)/v]^k \int |K_n^k(z,y) - K_{n+1}^k(z,y)| \mu(dz),
\]
and this last integral was shown to tend to zero. Further, applying the bounded convergence theorem and the induction hypothesis, the second term of (4.3.2) also tends to zero. \(\square\)

**Lemma 4.3.2:** Under the conditions of Lemma 4.3.1,
\[
|\psi_n(y) - \int \psi_n(x)K_n^k(x,y)\mu(dx)| \rightarrow 0 \text{ uniformly in } y
\]
for every \(k\).
Proof: Since \( \psi_n(y) = \int \psi_n(x)M_n(x,y) / \lambda_n \mu(dx) \), it follows that

\[
\psi_n(y) = \int \psi_n(x)M_n^k(x,y) / \lambda_n \mu(dx) = \int \psi_n(x)K_n(x,y) \mu(dx),
\]

for every \( k \). Hence for every \( y \)

\[
|\psi_n(y) - \int \psi_n(x)K_n^k(x,y) \mu(dx)| \leq D \int |K_n^k(x,y) - K_n^{k+1}(x,y)| \mu(dx)
\]

which tends to zero by Lemma 4.3.1. □

Theorem 4.3.2: If \( \{M_n(x,y)\} \) is a sequence of kernels satisfying Condition C and if \( \int |M_n(x,y) - M_{n+1}(x,y)| \mu(dx) \n \) uniformly in \( y \), then \( |\psi_n(y) - \psi_{n+1}(y)| \n \) uniformly in \( y \).

Proof: \( |\psi_n(y) - \psi_{n+1}(y)| \leq |\psi_n(y) - (\psi_n, \phi_{n+1}) \psi_{n+1}(y)| + |(\psi_n, \phi_{n+1}) \psi_{n+1}(y) - \psi_{n+1}(y)| \)

\[
= |\psi_n(y) - (\psi_n, \phi_{n+1}) \psi_{n+1}(y)| + \psi_{n+1}(y)|(\psi_n, \phi_{n+1}) - (\psi_n) - 1|.
\]

Let

\[
A_n(y) = \psi_n(y) - (\psi_n, \phi_{n+1}) \psi_{n+1}(y).
\]

Since \( \psi_n(y) \) integrates to 1, we can write
\[ |\psi_n(y) - \psi_{n+1}(y)| \leq |A_n(y)| + |\psi_{n+1}(y)| \int A_n(y) \mu(dy). \]

Since \( \psi_n(y) \) and \( A_n(y) \) are bounded, it suffices to show that \( |A_n(y)| \overset{n}{\to} 0 \).

So let \( \varepsilon > 0 \) be given.

\[ |A_n(y)| = |\psi_n(y) - (\psi_n, \phi_{n+1}) \psi_{n+1}(y)| \]
\[ \leq |\psi_n(y) - \int \psi_n(x) k_{n+1}^k(x,y) \mu(dx)| \]
\[ + |\int \psi_n(x) k_{n+1}^k(x,y) \mu(dx) - \int \psi_n(x) \phi_{n+1}(x) \psi_{n+1}(y) \mu(dx)| \]
\leq |\psi_n(y) - \int \psi_n(x) k_{n+1}^k(x,y) \mu(dx)| \]
\[ + D \int |k_{n+1}^k(x,y) - \phi_{n+1}(x)\psi_{n+1}(y)| \mu(dx). \]

We remarked earlier that Condition C is stronger than Condition T, and so Lemma 4.2.4 can be applied, hence

\[ |k_{n+1}^k(x,y) - \phi_{n+1}(x)\psi_{n+1}(y)| = \left| \frac{M_{n+1}^k(x,y)}{\lambda_{n+1}^k} - \phi_{n+1}(x)\psi_{n+1}(y) \right| \]
\[ < \frac{D}{d} \Delta^{k-1} \delta \]

Therefore it is possible to choose \( k_0 \) large enough so that

\[ \frac{D}{d} \Delta^{k_0-1} \delta < \varepsilon / [2\mu(S)D]. \]

In this case,
Finally Lemma 4.3.2 can be applied, so given $k_0$, there exists $N = N(k_0, \varepsilon)$ such that for $n \geq N$,

$$\left| \Psi_n(y) - \int \Psi_n(x) K_{n+1}^0(x, y) \mu(dx) \right| < \varepsilon/2.$$ 

Hence for $n \geq N(k_0, \varepsilon)$, $|A_n(y)| < \varepsilon$. 

It is possible to show, by arguments like those given in this section, that if $\{M_n\}$ and $M_0$ are kernels satisfying Condition C and if

$$\int |M_n(x, y) - M_0(x, y)| \mu(dx) \rightarrow 0 \quad (4.3.3)$$

uniformly in $y$, then $\lambda_n \rightarrow \lambda_0$ and $\Psi_n(y) \rightarrow \Psi_0(y)$ uniformly.

Hence, in view of Theorem 4.2.2, Equation (4.3.3) is sufficient for pointwise strong ergodicity.

We note that if $\{M_n\}$ and $\{N_n\}$ are two sequences of kernels satisfying Condition C and if

$$\int |M_n(x, y) - N_n(x, y)| \mu(dx) \rightarrow 0$$

uniformly in $y$, then results analogous to those given in this section can be obtained.

Finally, note that conditions analogous to those given in Theorem 4.3.2 can be given which guarantee that

$$|\phi_n(x) - \phi_{n+1}(x)| \not\rightarrow 0$$

uniformly.
V. APPLICATIONS TO STATISTICS

A. Sequential Probability Ratio Tests

We will discuss sequential probability ratio tests (SPRT's) from the point of view of random walks with absorbing barriers. Hence we will begin this section with a brief discussion of such random walks.

An example of a random walk with absorbing barriers is provided by the well-known "gambler's ruin" problem (see, for example Feller (1968) or Parzen (1962)). In this problem, the gambler, playing against an infinitely rich opponent, wins or loses one dollar with probability \( p \) or \( q = 1 - p \) respectively. The game terminates when the gambler's fortune is zero. Another version of the problem has the gambler playing against a finitely rich opponent and the game also stops if the gambler wins all of his opponent's money, say when his fortune reaches \( F \) dollars. The states 0 and \( F \) are called absorbing states and the transition matrix describing such a chain is given by

\[
\begin{pmatrix}
0 & 1 & 2 & 3 & \ldots & F-2 & F-1 & F \\
0 & 1 & 0 & 0 & 0 & \ldots & 0 & 0 & 0 \\
1 & q & 0 & p & 0 & \ldots & 0 & 0 & 0 \\
2 & 0 & q & 0 & p & \ldots & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\
F-1 & 0 & 0 & 0 & 0 & \ldots & q & 0 & p \\
F & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}
\]
If an initial probability distribution \( f_0 \) over the states is given, then the probability distribution after \( n \) trials is given by \( f_0 P^n \). If we are interested in the probability distribution over the non-absorbing states, this would be given by \( f^* \) where \( f_n = f_0 M^n \) and \( M \) is the matrix \( P \) with the rows and columns corresponding to the absorbing states deleted. Of course the long-run behavior of \( f_n^* \) could be predicted if one could show that the conditions of one of the pertinent theorems of Chapter III or IV are satisfied.

In a simple random walk, the only possible transitions are from a state \( k \), say, to one of its immediate neighbors. In a generalized random walk, transitions to other states are possible. Such a walk could arise from the following situation. Let \( \{X_n\} \) be a sequence of independent discrete random variables defined on some probability space. Given \( S_0 = 0 \), say, define \( S_n = \sum_{i=1}^{n} X_i \). If \( S_n = k \), we say the chain is in state \( k \) at time \( n \). The probabilities of transition from state \( k \) at time \( n-1 \) to \( k+j \) at time \( n \) are determined by the probability distribution of \( X_n \). The chain is homogeneous or not depending on whether the \( X_n \) are identically distributed or not.

If the same problem is formulated with \( S_n \) being the sum of independent continuous random variables, then the transitions from "state" \( x \) at time \( n-1 \) to "state" \( y \) at time \( n \) will still depend on the probability distribution of \( X_n \). In fact, if \( h_n(t) \) is the probability density function
for $X_n$, then the stochastic kernel $P_n(x,y)$ is given by

$$P_n(x,y) = h_n(y-x).$$

In this case $P_n(x,y)$ would be the transition kernel for a non-homogeneous Markov chain defined on $(S, \mathcal{B}, \mu)$ with $S = \text{Reals}$, $\mathcal{B} = \text{Lebesgue sets}$, and $\mu = \text{Lebesgue measure}$. If the real numbers $a$ and $b$ ($a<b$) are absorbing barriers, then

$$M_n(x,y) = P_n(x,y) \ (x,y) \in (a,b) \times (a,b)$$

is a kernel defined over the non-absorbing part of the space.

The probability distribution over the non-absorbing region, assuming of course that absorption has not yet taken place, is given by

$$f_n^*(y) = f_n(y)/\int_a^b f_n(y) \, dy$$

where

$$f_n(y) = \int_a^b f_0(x) \, M_{1,n}(x,y) \, dx .$$

We now show how the above discussion pertains to SPRT's. An SPRT is a method for testing hypotheses when the sample size is not fixed in advance, but is determined by the observations as they appear. The test is a likelihood-ratio procedure and a test of $H_0: f(x) = f_0(x)$ against $H_1: f(x) = f_1(x)$ is performed as follows: Assuming the observations are independent, define
\[ \lambda_n = \prod_{i=1}^{n} \frac{f_1(x_i)}{f_0(x_i)}. \]

If

\[ B < \lambda_n < A \]

continue sampling, if

\[ \lambda_n < B, \]

accept \( H_0 \), and if

\[ \lambda_n > A \]

reject \( H_0 \). The constants \( A \) and \( B \) are determined by the constants \( \alpha \) and \( \beta \), the probability of a type I and type II error respectively (see for example, Wald (1947)).

An equivalent test can be found by taking logs. In this case, continue sampling if

\[ \ln B < \sum_{i=1}^{n} \ln \frac{f_1(x_i)}{f_0(x_i)} < \ln A \] (5.1.1)

and accept or reject \( H_0 \) depending on whether

\[ \sum_{i=1}^{n} \ln \frac{f_1(x_i)}{f_0(x_i)} \leq \ln B \] or \[ \sum_{i=1}^{n} \ln \frac{f_1(x_i)}{f_0(x_i)} \geq \ln A. \]

If we let \( a = \ln B \), \( b = \ln A \), and \( z_i = \ln [f_1(x_i)/f_0(x_i)] \), then the \( z_i \) are independent if the \( x_i \) are (as long as \( f_1 \) and \( f_0 \) are Borel functions) and (5.1.1) becomes

\[ a < \sum_{i=1}^{n} z_i < b. \]
If the $Z_i$ are independent then $S_n = \sum_{i=1}^{n} Z_i$ will form a Markov chain with transition kernels determined by the densities of the $Z_i$, namely $P_i(x,y) = h_{Z_i}(y-x)$. If the $Z_i$ are identically distributed, as they are in the type of SPRT described above, the corresponding Markov chain will be homogeneous.

Knowledge of the probability distribution over the non-absorbing states may be useful in constructing truncation rules. A truncation rule is a rule for acceptance or rejection of $H_0$ at the $N^{th}$ trial if the original SPRT procedure did not call for a decision by time $N$. We will not discuss truncation rules in this work, but will give some examples of tests which lead to random walks such that some theorems from Chapter IV will apply.

We point out that Conn noted the relevance of kernel superpositions to SPRT's. However, her attention to positive kernels excluded the consideration of examples such as those given here.

**Example 5.1.1:** For the gamma distribution with $\alpha$ known, test

$$H_0: \beta = \beta_0 \quad \text{against} \quad H_1: \beta = \beta_1.$$  

Assume $\beta_1 > \beta_0$. In terms of a random walk with absorbing barriers, we consider $Z_i = \ln[f_1(x_i)/f_0(x_i)]$. Since
\[ f_1(x) = \frac{1}{\Gamma(a+1)\beta_1} x^{\alpha} e^{-x/\beta_1} \]

\[ f_0(x) = \frac{1}{\Gamma(a+1)\beta_0} x^{\alpha} e^{-x/\beta_0} \]

\[ = \left(\frac{\beta_0}{\beta_1}\right) e^{-x\left(\frac{1}{\beta_1} - \frac{1}{\beta_0}\right)} \]

\[ Z = (a+1) \ln\left(\frac{\beta_0}{\beta_1}\right) - x\left(\frac{1}{\beta_1} - \frac{1}{\beta_0}\right) \text{ for } x > 0. \]

The SPRT then requires that we continue sampling if

\[ \ln B < \sum_{i=1}^{n} [(a+1)\ln\left(\frac{\beta_0}{\beta_1}\right) - x_i\left(\frac{1}{\beta_1} - \frac{1}{\beta_0}\right)] < \ln A. \quad (5.1.2) \]

Since \( \frac{1}{\beta_0} - \frac{1}{\beta_1} > 0 \), if we define

\[ a = \ln B / (1/\beta_0 - 1/\beta_1) \]
\[ b = \ln A / (1/\beta_0 - 1/\beta_1) \]
\[ \gamma = (a+1)(\ln \beta_1 - \ln \beta_0) / (1/\beta_0 - 1/\beta_1) \]

then (5.1.2) becomes

\[ a < \sum_{i=1}^{n} (x_i - \gamma) < b. \]

If we define \( Z_i = (x_i - \gamma) \), then clearly the \( Z_i \) are independent if the \( x_i \) are, and \( Z_i \) have densities given by

\[ h_i(t) = \begin{cases} \frac{1}{\Gamma(a+1)\beta^{a+1}} (t+\gamma)^{\alpha} e^{-(t+\gamma)/\beta} & t > -\gamma \\ 0 & t \leq -\gamma \end{cases} \]
Hence

\[ P_n(x,y) = h_n(y-x) = \begin{cases} \frac{1}{\Gamma(a+1)\beta^a+1} (y-x+\gamma)^a e^{-(y-x+\gamma)/\beta} & \text{if } y-x > -\gamma \\ 0 & \text{if } y-x \leq -\gamma. \end{cases} \]

With \( a \) and \( b \) as defined in (5.1.3) we can define a kernel

\[ M(x,y) = P(x,y) \text{ for } (x,y) \in (a,b) \times (a,b). \]

This kernel does not necessarily have a positive lower bound, since it can be shown that \( M(x,y) \) will assume the value zero if \( y < b - a \), or equivalently if \( (\beta_1/\beta_0)^{a+1} < A/B \).

On the other hand, Condition S is satisfied as can be seen as follows. The kernels are continuous, hence they are measurable. An upper bound for \( M(x,y) \) is \( \Delta = \frac{a e^{-a}}{\Gamma(a+1)\beta} \), and \( M(x,S) \) can be shown to be bounded below by a positive number. It is not hard to see that \( M(x,y) \) is primitive, hence has a left eigenfunction bounded above and below by positive numbers. Finally, because of the exponential tail of \( M(x,y) \), it is easy to see that, for \( Q \) defined by (3.1.4), \( \alpha(Q) > 0 \). In addition, because of the stationarity of this sequence, \( \int_a^b |\psi_n(y) - \psi_{n+1}(y)| dy = 0 \), hence Theorem 4.2.1. applies, so \( |f_n(y) - \psi(y)| \to 0 \).

Although we know that a positive left eigenfunction exists for the kernel of this example, it is not easily found. In the next example, we can exhibit the eigenfunction of interest.
Example 5.1.2: For the negative exponential with $\beta$ known, test

$H_0: \alpha = \alpha_0$ against $H_1: \alpha = \alpha_1$.

For simplicity assume $\beta=0$ and assume $\alpha_0 > \alpha_1$. In this case,

\[
\frac{f_1(x)}{f_0(x)} = \frac{\alpha_1 e^{-\alpha_1 x}}{\alpha_0 e^{-\alpha_0 x}} = \frac{\alpha_1}{\alpha_0} e^{x(\alpha_0 - \alpha_1)}
\]

Hence

\[
\ln \left[ \frac{f_1(x)}{f_0(x)} \right] = \ln \left( \frac{\alpha_1}{\alpha_0} \right) + (\alpha_0 - \alpha_1)x.
\]

Since $(\alpha_0 - \alpha_1) > 0$, if we define

\[
a = \ln \frac{\beta}{(\alpha_0 - \alpha_1)} \\
b = \ln \frac{A}{(\alpha_0 - \alpha_1)} \\
\gamma = \frac{(\ln \alpha_0 - \ln \alpha_1)}{(\alpha_0 - \alpha_1)}
\]

then the SPRT requires that we continue sampling if

\[
a < \frac{n}{\sum_{i=1}^{n} (x_i - \gamma)} < b.
\]

Again, if $Z_i = x_i - \gamma$, then the $Z_i$ are independent if the $x_i$ are, and the $Z_i$ have densities given by

\[
h_i(t) = \begin{cases} 
\alpha e^{-\alpha(t+\gamma)} & \text{if } t \geq -\gamma \\
0 & \text{if } t < -\gamma
\end{cases}
\]

With $a$ and $b$ as defined in (5.1.4), the appropriate
kernel for this problem, for \((x,y) \in (a,b) \times (a,b)\), is

\[
M(x,y) = \begin{cases} 
  a e^{-a(y-x+\gamma)} & \text{if } y-x \geq -\gamma \\
  0 & \text{if } y-x < -\gamma .
\end{cases}
\]

As in the previous example, it can be shown that the conditions of Theorem 4.2.1 hold, so that \(|f_n^*(y) - \Psi(y)| \to 0\), where \(\Psi(y)\) is the positive bounded left eigenfunctions of \(M(x,y)\). We can in fact find the eigenfunction \(\Psi(y)\) for this problem.

We note that \(M(x,y) > 0\) whenever \(y-x \geq -\gamma\), that is when \(x \leq y+\gamma\). Now for \(y \in [b-\gamma, b)\), since \(y+\gamma, (b-\gamma) + \gamma = b\), \(M(x,y)\) is positive for all \(x \in (a,b)\). Hence, for \(y \in [b-\gamma, b)\),

\[
\lambda \Psi(y) = \int_a^b \Psi(x) M(x,y) \, dx = \int_a^b \Psi(x) a e^{-a(y-x+\gamma)} \, dx
\]

\[
= a e^{-a\gamma} e^{-a\gamma} \int_a^b \Psi(x) e^{ax} \, dx = c_0 K e^{-a\gamma}
\]

where \(K = a e^{-a\gamma}\) and \(c_0 = \int_a^b \Psi(x) e^{ax} \, dx\). Clearly, if \(\Psi(y)\) is to be an eigenfunction, it must be that

\[
\Psi(y) = (c_0 K/\lambda) e^{-a\gamma} . \tag{5.1.5}
\]

Note however that (5.1.5) is only for values of \(y \in [b-\gamma, b)\). If \(a < b-\gamma\), we must proceed as follows: for \(y \in (b-2\gamma, b-\gamma)\), \(b-\gamma < y + \gamma < b\), and since \(M(x,y) = 0\) for \(x > y + \gamma\), we have
\[ \lambda \psi(y) = \int_a^b \psi(x)M(x,y)dx = \int_a^{y+\gamma} \psi(x)e^{-\alpha(y-x+\gamma)}dx \]

\[ + \int_b^{y+\gamma} \psi(x)\cdot 0 \, dx \]

\[ = \int_a^{b-\gamma} \psi(x)Ke^{-\alpha y}e^{\alpha x}dx \]

\[ + \int_{y+\gamma}^{b-\gamma} [(c_0K/\lambda)e^{-\alpha x}]Ke^{-\alpha y}e^{\alpha x}dx . \]

Note that in the second integral,

\[ x \in (b-\gamma,y+\gamma) \subset (b-\gamma,b) , \]

so we can use (5.1.5) to write the functional form of \( \psi(x) \).

Now if we define

\[ c_n = \int_a^{b-n\gamma} \psi(x)e^{\alpha x} \, dx , \]

then for \( y \in [b-2\gamma,b-\gamma) \),

\[ \psi(y) = \left\{ \frac{c_1K}{\lambda} + \frac{c_0K^2}{\lambda^2} [y-(b-2\gamma)] \right\}e^{-\alpha y} . \]

Continuing in this way, we find that for \( y \in [b-n\gamma,b-(n-1)\gamma) \),

\[ \psi(y) = \sum_{r=1}^{n} \{c_{n-r}(\frac{K}{\lambda}) [y-(b-n\gamma)] \}^{r-1}/(r-1)!e^{-\alpha y} . \quad (5.1.6) \]

Equation (5.1.6) holds for \( n = 1,2,\ldots,N \), where \( N=(b-a)/\gamma \)

if this is an integer, or \( N = \lfloor \frac{b-a}{\gamma} \rfloor + 1 \), where \( \lfloor \cdot \rfloor \) represents the greatest integer function.
One would choose the constant multiple of $\Psi(y)$ as defined in (5.1.6) which gives the eigenfunction which integrates to 1. Note too, that the eigenfunction $\Psi(y)$ depends on the true value of the parameter $\alpha$ which is being tested by the SPRT.

**B. Generalized Sequential Probability Ratio Tests**

Weiss (1953) gives the following definition of a generalized sequential probability ratio test (GSPRT) for testing a simple hypothesis $H_0$ against a simple alternative $H_a$. Let $\{A_n\}$ and $\{B_n\}$ be two sequences of constants ($B_n < A_n$). Continue sampling if

$$B_n < \prod_{i=1}^{n} \left[ \frac{f_1(x_i)}{f_0(x_i)} \right] < A_n$$

and accept or reject $H_0$ if

$$\prod_{i=1}^{n} \left[ \frac{f_1(x_i)}{f_0(x_i)} \right] \leq B_n \text{ or } \prod_{i=1}^{n} \left[ \frac{f_1(x_i)}{f_0(x_i)} \right] \geq A_n$$

respectively. This test differs from the SPRT in that the decision boundaries are a function of $n$.

Note that Conn studied this problem with the condition that $\ln A_n - \ln B_n = C$ for all $n$. We do not impose this condition.

If we take logs and if $Z_i = \ln \left[ \frac{f_1(x_i)}{f_0(x_i)} \right]$ are independent, and if
a_n = \ln B_n \\
\quad b_n = \ln A_n ,

then the GSPRT can be thought of as a random walk with moving absorbing barriers. If we want to consider the distribution over the non-absorbing states, we must consider the sequence \( \{f_n(y)\} \), where

\[
f_n(y) = \int_{a_n}^{b_n} f_{n-1}(x)M(x,y)\,dx
= \int_{a_n}^{b_n} \cdots \int_{a_1}^{b_1} f_0(z_1)M(z_1,z_2)\ldots M(z_n,y)\,dz_1\ldots dz_n.
\]

(5.2.1)

Note that the range of integration on \( z_k \) is from \( a_k \) to \( b_k \). This does not follow the conditions in the first part of this work where the range of integration was the same for each kernel. We will give a transformation which, when applicable, will change this to a problem of different kernels defined on the same domain.

Define \( a = \inf a_n \) and \( b = \sup b_n \). If both \( a \) and \( b \) are finite, then for \( z \in (a_k, b_k) \), we make the linear transformation

\[ z = c_k + r_kw \]

where \( c_k = (a_kb-ab_k)/(b-a) \) and \( r_k = (b_k-a_k)/(b-a) \). Note that \( r_k \leq 1 \).
Then for \( y \epsilon (a_{n+1}, b_{n+1}) \), applying this transformation to (5.2.1) we find

\[
 f_n(y) = \int_a^b \cdots \int_a^b f_0(c_1+r_1w_1)M(c_1+r_1w_1, c_2+r_2w_2) \cdots M(c_n+r_nw_n, y)
\]

\[
 (r_1r_2 \cdots r_n)dw_1 \cdots dw_n .
\]

Now define

\[
 g_0(w) = f_0(c_1+r_1w)
\]

\[
 M_k(v,w) = r_kM(c_k+r_kv, c_{k+1}+r_{k+1}w)
\]

(5.2.2)

Then \( f_n(y) \) can be written as

\[
 f_n(y) = \int_a^b \cdots \int_a^b g_0(w_1)M_1(w_1, w_2)
\]

\[
 \cdots M_{n-1}(w_{n-1}, w_n) r_nM(c_n+r_nw_n, y) dw_1 \cdots dw_n.
\]

If \( w = (y-c_{n+1})/r_{n+1} \), then \( w \epsilon (a, b) \) and

\[
 f_n(c_{n+1}+r_{n+1}w) = g_n(w) = \int_a^b \cdots \int_a^b g_0(w_1)M_1(w_1, w_2)
\]

\[
 \cdots M_n(w_n, w) dw_1 \cdots dw_n .
\]

(5.2.3)

and this last equation is consistent with our previous notation. Hence the probability distribution over the non-absorbing states at time \( n \) can be found either by using \( f_n^* \) with state space \((a_{n+1}, b_{n+1})\) or by using \( g_n^* \) with state
We now give conditions for \( \int_a^b |M_n(v,w) - M_{n+1}(v,w)| \, dv \)
to tend to zero as \( n \to \infty \).

**Lemma 5.2.1:** Let \( \{M_n\} \) be a sequence of kernels
defined by (5.2.2) with \( M(x,y) \), the original kernel, non-
negative and uniformly continuous in both arguments for
\((x,y) \in (a,b) \times (a,b) \). If \( |b_n - b_{n+1}| \to 0 \) and \( |a_n - a_{n+1}| \to 0 \), then

\[
\int_a^b |M_n(v,w) - M_{n+1}(v,w)| \, dv \to 0
\]

uniformly in \( w \).

**Proof:** We can show that both \( |r_n - r_{n+1}| \) and \( |c_n - c_{n+1}| \)
tend to zero since

\[
|r_n - r_{n+1}| = \left| \frac{(b_n - a_n) - (b_{n+1} - a_{n+1})}{b-a} \right|
\]

\[
\leq \frac{|b_n - b_{n+1}| + |a_n - a_{n+1}|}{b-a} \to 0
\]

and

\[
|c_n - c_{n+1}| = \left| \frac{(ab_n - a_n b) - (ab_{n+1} - a_{n+1} b)}{b-a} \right|
\]

\[
\leq \left\{ |a| |b_n - b_{n+1}| + |b| |a_n - a_{n+1}| \right\} \to 0 .
\]

Then

\[
\int_a^b |M_n(v,w) - M_{n+1}(v,w)| \, dv
\]

\[
= \int_a^b |r_n M(c_n + r_n v, c_{n+1} + r_{n+1} w) - M_{n+1}(v,w)| \, dv
\]
Now the first term of (5.2.4) goes to zero since $M$ is bounded and since $|r_n - r_{n+1}| \to 0$. Since $r_{n+1} < 1$ and since $M$ is uniformly continuous, it follows that the second term also goes to zero. $\square$

**Corollary 5.2.1:** Under the conditions of Lemma 5.2.1, if the kernels $\{M_n\}$ satisfy Condition C, then

$$|g_n^*(w) - \psi_n(w)| \to 0$$

where $g_n^*(w)$ is defined in (5.2.3).

**Proof:** The proof follows immediately from Lemma 5.2.1, Theorem 4.3.2, and Corollary 4.2.1. $\square$
VI. LITERATURE CITED


VII. ACKNOWLEDGMENTS

The author sincerely thanks his co-major professors, Dr. Herb T. David and Dr. Dean L. Isaacson for their guidance in the research and writing of this thesis. They both were patient, able, and readily available to help with problems large and small.

Thanks also to my family for their confidence and encouragement during these last three years. Special thanks to my wife, Carole for her love, understanding, and patience during some difficult times.