Z. Wahrscheinlichkeitstheorie verw. Gebiete
37, 211 – 226 (1977)

Zeitschrift für
Wahrscheinlichkeitstheorie
und verwandte Gebiete

A Martingale Approach to the Convergence of the Iterates of a Transition Function

F. Papangelou

Department of Mathematics, The University, Manchester M 13 9 PL, England

§ 1. Introduction

The problem of the convergence of the iterates $p^{(n)}(x, B)$ of the transition function of a Markov process defined on a "continuous" state space received early attention in the development of the theory of Markov processes. Results establishing the cyclic behaviour and uniform convergence (along the appropriate subsequences) of iterates to certain "periodic" distributions were already obtained in the thirties and we refer the reader to p. 628 of [6] for references. Doeblin's work in particular ([3, 4]) was very influential on most subsequent literature on the subject. These early results were mostly based on two different conditions, the "Doeblin condition" (which is probabilistic in nature) and the "Kryloff-Bogoliouboff condition" (a functional analytic one). It was Yosida and Kakutani [19] who obtained Doeblin type convergence results operator-theoretically under the Kryloff-Bogoliouboff condition and showed that the two approaches were essentially equivalent. (See also [16, p. 167], [12, p. 45].)

In [5] Doob assumed the existence of a stationary distribution and, under a condition weaker than those mentioned above, established the cyclic behaviour and convergence of the appropriate subsequences of $p^n(x, B)$ for "every x" and "every B", but not necessarily uniformly. Another insight into the convergence problem was given by Jacobs in [11]. Harris introduced his recurrence condition for Markov processes in [9] and proved that under this condition there is an invariant σ-finite measure (unique to within a constant factor). Orey [17] then showed that a Harris recurrent Markov process exhibits the familiar cyclic behaviour and, for processes with finite invariant measure, strengthened Doob's "simple" convergence to convergence in total variation.

An alternative proof of Orey's theorem (and a more general version of it) was given in [14], where Jamison and Orey make use of the space-time harmonic functions to show that the tail field of the process is finite for any initial distribution. Jain and Jamison [13] established limit theorems under broader assumptions, by reducing them to Orey's theorem. Foguel (see [7] and [8]) and S. Horowitz [10] gave a functional analytic formulation and proof of Orey's theorem. Ornstein and Sucheston [18] derived the Jamison-Orey theorem from their "zero-two law" and there have been a host of other papers exploring its L_1-operator-theoretic implications and alternative approaches to it.

In the present paper we are exclusively interested in Markov processes whose cycle has length one (the "aperiodic" case). Our purpose is to indicate an alternative probabilistic approach to the problem of convergence to asymptotic stationarity, which is based on the theory of martingales and involves no hard technicalities other than the martingale convergence theorem. In Theorem 2.1 below we assume the existence of an invariant probability measure π and formulate a necessary and sufficient condition for the convergence (in total variation) of $p^{(n)}(x, \cdot)$ to $\pi(\cdot)$ for π-almost every x. The novelty lies in the method of proof rather than the statement itself, which can easily be derived from results of [13] or of [5] (with the required "total variation twist" provided by Theorem 7.3 below).

The condition given in the theorem has an advantage over some others that have been used in the literature in that it is readily verifiable in many examples encountered by the working probabilist. The theorem is also well adapted to situations where there are many invariant probability measures. To give an idea of the spirit of the approach we outline here the main line of the proof.

Suppose $\ldots, X_{-1}, X_0, X_1, \ldots$ is a Markov process with transition function $p(x, B)$ and such that $P(X_n \in B) = \pi(B)$ for all n (where π is the invariant probability measure). If $p^{(n)}(x, dy) = f_n(x, y)\,\pi(dy) + v_n(x, dy)$ is the decomposition of $p^{(n)}(x, \cdot)$ into absolutely continuous and singular parts with respect to π, then, for every x, the process $f_n(x, X_n)$, $n = 1, 2, \ldots$ is a backward supermartingale (Lemma 3.2). It follows among other things that for fixed x the functions $f_n(x, \cdot)$, $n \geqq 1$, are uniformly integrable under π and hence the sequence is relatively compact in the $\sigma(L_1, L_\infty)$-topology, i.e. every subsequence contains a sub-subsequence converging to some function g in the $\sigma(L_1, L_\infty)$-topology. Using the fact that backward supermartingales converge in the mean we prove in Lemma 4.2 that there are functions $g_1, g_2, \ldots$ such that $g(X_0) = g_1(X_1) = g_2(X_2) = \ldots$ almost surely. The condition of Theorem 2.1 is then shown to imply that g equals some constant almost surely and hence that $f_n(x, \cdot)$ converges ($\sigma(L_1, L_\infty)$) to this constant (Theorem 4.1). In § 6 it is proved that this constant is 1 for almost all x (Theorem 6.1) and this means that $\lim_{n \to \infty} p^{(n)}(x, B) = \pi(B)$ for all measurable B (Theorem 6.3).

In § 7 we give a simple and apparently hitherto unnoticed way of strengthening this type of convergence to convergence in total variation: This consists in showing (Lemma 7.1, Theorem 7.3) that $\liminf_{n \to \infty} f_n(x, y) \geqq 1$ for $\pi \otimes \pi$-almost all (x, y), if the functions $f_n(x, y)$ are chosen to be jointly measurable in (x, y). This easily implies the L_1-convergence of $f_n(x, \cdot)$, $n \geqq 1$, to 1 and hence the convergence in total variation of $p^{(n)}(x, \cdot)$ to π (Theorem 7.2).

Orey mentions in [17] that the assertion "$\lim_{n \to \infty} p^{(n)}(x, B) = \pi(B)$ for all x and B" for an aperiodic Harris process follows from Doob's Theorem 5 in [5]. It is interesting that the argument of Theorem 7.3 makes Orey's stronger assertion ("$\lim_{n \to \infty} \|p^{(n)}(x, \cdot) - \pi(\cdot)\| = 0$ for all x") a simple consequence of Doob's theorem.

Theorem 7.4, which seems to us to be new, asserts the same type of convergence for the measure $\rho^{(n)}(B, y) = \int_B f_n(x, y)\,\pi(dx)$. This is connected with the fact that the process $f_n(X_{-n}, y)$, $n = 1, 2, \ldots$, is a backward supermartingale (Lemma 7.5).

Section § 5 indicates that Kolmogorov's classical theorem on the convergence of the iterates of a positive-recurrent, irreducible, aperiodic transition matrix on a countable state space already follows from Theorem 4.1. Finally in the appendix we show how Orey's theorem for the case of an aperiodic Harris process admitting an invariant probability measure can be obtained from 2.1. For the major theorems it is assumed that the σ-field of the state space is contained in the completion (with respect to π) of a countably generated σ-subfield. The reader interested in achieving greater generality is referred to [14, §4] for methods of removing such assumptions.

Note Added in May 1976. The author has recently received a preprint of a paper by Y. Derriennic entitled "Lois zero ou deux pour les processus de Markov. Applications aux marches aleatoires", giving a striking martingale approach to the "zero-two law" from which, as mentioned above, Ornstein and Sucheston derived Orey's theorem.

§ 2. Statement of the Theorem

Let $(S, \mathscr{B})$ be a measurable space. If λ, μ are finite measures on $(S, \mathscr{B})$ we shall denote below by $\lambda \otimes \mu$ their product measure on $(S \times S, \mathscr{B} \otimes \mathscr{B})$ and by $\|\lambda - \mu\|$ the total variation norm of $\lambda - \mu$, i.e. $\|\lambda - \mu\| = (\lambda - \mu)^+(S) + (\lambda - \mu)^-(S)$ where $(\lambda - \mu)^+$ and $(\lambda - \mu)^-$ are the positive and negative parts of $\lambda - \mu$ in its Jordan decomposition. Recall that the measures λ, μ are called singular with respect to each other (denoted $\lambda \perp \mu$) if there is a set $B \in \mathscr{B}$ such that $\lambda(B) = 0$ and $\mu(B^c) = 0$. At the other end, λ is said to be absolutely continuous with respect to μ (denoted $\lambda \ll \mu$) if $\mu(B) = 0$ implies $\lambda(B) = 0$. Given μ, every λ can be decomposed into a sum $\lambda_1 + \lambda_2$, where $\lambda_1 \ll \mu$ and $\lambda_2 \perp \mu$.

As is well-known a probability transition function on $(S, \mathscr{B})$ is a function $p(x, B)(x \in S, B \in \mathscr{B})$ such that for fixed x, $p(x, \cdot)$ is a probability measure on $\mathscr{B}$ and for fixed B, $p(\cdot, B)$ is a $\mathscr{B}$-measurable function on S. The iterates $p^{(n)}(x, B)$ of p are defined inductively as follows:

$$p^{(1)}(x, B) = p(x, B), \qquad p^{(n+1)}(x, B) = \int_{y \in S} p^{(n)}(x, dy)\, p(y, B).$$

A probability measure π on $(S, \mathscr{B})$ is said to be stationary or invariant under $p(\cdot, \cdot)$ if

$$\pi(B) = \int_{x \in S} \pi(dx)\, p(x, B) \qquad (B \in \mathscr{B}). \tag{1}$$

Throughout the present paper it will be assumed that π is a fixed stationary probability.

For many of the results the following hypothesis will be made.

Hypothesis (H). *There is a countably generated σ-field $\mathscr{G} \subset \mathscr{B}$ such that $\mathscr{B}$ is contained in the completion of $\mathscr{G}$ with respect to π.*

(2.1) **Theorem.** *Let π be a probability measure on $(S, \mathscr{B})$, stationary under $p(\cdot, \cdot)$, and suppose hypothesis* (H) *holds. The following conditions are then equivalent:*

(i) *For $\pi\otimes\pi$-almost all $(x,y)\in S\times S$ there is an n (depending on (x,y)) such that the measures $p^{(n)}(x,\cdot)$ and $p^{(n)}(y,\cdot)$ are not singular with respect to each other.*

(ii) *For π-almost all $x\in S$*

$$\lim_{n\to\infty}\|p^{(n)}(x,\cdot)-\pi(\cdot)\|=0.$$

The implication (ii) $\Rightarrow$ (i) is immediate since (ii) implies that, for $\pi\otimes\pi$-almost all (x,y), $\lim_{n\to\infty}\|p^{(n)}(x,\cdot)-p^{(n)}(y,\cdot)\|=0$ and two probability measures λ, μ with $\|\lambda-\mu\|<2$ cannot be singular. Our aim is to prove the converse implication (i) $\Rightarrow$ (ii). In the text, (i) and (ii) will be referred to as "condition (i)" and "condition (ii)" respectively.

All subsets of S appearing below will be assumed, without explicit mention, to be members of $\mathscr{B}$. Since the proof of Theorem 2.1 depends on probabilistic considerations we introduce the class of Markov processes with transition function $p(x,B)$. First we define the stationary process.

Let $\Omega=\cdots\times S\times S\times S\times\cdots$, $\mathscr{F}=\cdots\otimes\mathscr{B}\otimes\mathscr{B}\otimes\mathscr{B}\otimes\cdots$ and for each $\omega=(\ldots,x_{-1},x_0,x_1,\ldots)\in\Omega$ let $X_n(\omega)=x_n$. There is a probability P_π on $(\Omega,\mathscr{F})$ under which the process $\ldots,X_{-1},X_0,X_1,\ldots$ is a stationary Markov process with transition function $p(x,B)$ and such that each X_n has distribution π. In other words

(a) For any k, any $n\geqq 1$ and any $B\in\mathscr{B}$, $p^{(n)}(x,B)$ is a version of $P_\pi(X_{k+n}\in B\,|\,X_k=x)$.

(b) For any n and any $B\in\mathscr{B}$, $P_\pi(X_n\in B)=\pi(B)$.

This P_π is uniquely determined by its finite-dimensional marginals, defined as follows

$$P_\pi(X_i\in B_0, X_{i+1}\in B_1,\ldots,X_{i+k}\in B_k)$$
$$=\int_{x_0\in B_0}\pi(dx_0)\int_{x_1\in B_1}p(x_0,dx_1)\ldots\int_{x_k\in B_k}p(x_{k-1},dx_k),$$

with the formula valid for arbitrary $i\,(-\infty<i<\infty)$. Expectations with respect to P_π will be denoted by E_π.

The unilateral sequence $X_0,X_1,\ldots$ can be made into a Markov process with the same transition function but with an arbitrary initial distribution λ. More specifically, for every probability measure λ on $(S,\mathscr{B})$ there is a probability measure P_λ on $(\Omega,\hat{\mathscr{F}})$, where $\hat{\mathscr{F}}$ is the σ-subfield of $\mathscr{F}$, generated by $X_0,X_1,X_2,\ldots$, such that X_0 has distribution λ and (a) above holds with P_π replaced by P_λ and with the additional restriction $k\geqq 0$. In this case

$$P_\lambda(X_0\in B_0, X_1\in B_1,\ldots,X_k\in B_k)$$
$$=\int_{x_0\in B_0}\lambda(dx_0)\int_{x_1\in B_1}p(x_0,dx_1)\ldots\int_{x_k\in B_k}p(x_{k-1},dx_k).$$

§ 3. The Backward Supermartingale

For the results of sections § 3 and § 4 hypothesis (H) will not be needed.

If λ is a finite measure on $(S,\mathscr{B})$ we define the measure $\lambda\circ p^{(n)}$ $(n=1,2,\ldots)$ by $(\lambda\circ p^{(n)})(B)=\int_S\lambda(dx)\,p^{(n)}(x,B)$. Note that if λ is a probability measure then $(\lambda\circ p^{(n)})(B)=P_\lambda(X_n\in B)$.

(3.1) *Definitions.* Let λ be a probability measure on $(S, \mathscr{B})$, to be called the "initial distribution". We set $\lambda^0 = \lambda$, $\lambda^n = \lambda \circ p^{(n)} (n \geqq 1)$ and for each $n \geqq 0$ we let μ_n and ν_n be the absolutely continuous and singular parts, respectively, of λ^n with respect to π, i.e. $\lambda^n = \mu_n + \nu_n$, $\mu_n \ll \pi$, $\nu_n \perp \pi$. For each $n \geqq 0$ we choose a version of the Radon-Nikodym density $\frac{d\mu_n}{d\pi}$ and denote it by $f_n(\lambda, \cdot)$, i.e.

$$\mu_n(B) = \int_B f_n(\lambda, x)\, \pi(dx) \qquad (B \in \mathscr{B}). \tag{2}$$

It follows easily from (1) that $\mu_n \circ p \ll \pi$ and hence

$$\mu_n \circ p \leqq \mu_{n+1} \tag{3}$$

for every $n \geqq 0$. This in turn implies

$$\mu_n(S) \leqq \mu_{n+1}(S), \qquad \nu_n(S) \geqq \nu_{n+1}(S).$$

We define $c(\lambda) = \lim_{n \to \infty} \mu_n(S)$

(3.2) **Lemma.** *For each* $n \geqq 1$

$$E_\pi(f_{n-1}(\lambda, X_{n-1}) | X_n, X_{n+1}, \ldots) \leqq f_n(\lambda, X_n) \qquad P_\pi\text{-a.s.}$$

If $\mathscr{F}_n$ *denotes the* σ*-field* $\mathscr{F}(X_n, X_{n+1}, \ldots)$ *generated by* $X_n, X_{n+1}, \ldots$ *then* $\{f_n(\lambda, X_n), \mathscr{F}_n, n \geqq 0\}$ *is a backward supermartingale under* P_π.

Note. A "backward supermartingale" is a process which becomes a supermartingale after a reversal of time. Thus the present lemma asserts that the process

$$\ldots, f_2(\lambda, X_2), f_1(\lambda, X_1), f_0(\lambda, X_0)$$

is a supermartingale.

Since λ and π are held fixed in several sections we shall occasionally write $f_n(x)$ instead of $f_n(\lambda, x)$ and P, E instead of P_π and E_π, though we shall continue to use the latter in most statements of theorems.

Proof of Lemma (3.2). Suppose $\Delta \in \mathscr{F}_n$ and denote by χ_Δ the indicator function of Δ. Then

$$\begin{aligned}
&\int_\Delta E(f_{n-1}(X_{n-1}) | \mathscr{F}_n)\, dP \\
&\quad = \int_\Delta f_{n-1}(X_{n-1})\, dP = \int_\Omega f_{n-1}(X_{n-1})\, \chi_\Delta\, dP = \int_\Omega f_{n-1}(X_{n-1})\, P(\Delta | X_{n-1})\, dP \\
&\quad = \int_S f_{n-1}(x)\, P(\Delta | X_{n-1} = x)\, \pi(dx) = \int_S P(\Delta | X_{n-1} = x)\, \mu_{n-1}(dx) \qquad \text{by (2).}
\end{aligned}$$

Now the Markov property implies that

$$P(\Delta | X_{n-1} = x) = \int_{y \in S} p(x, dy)\, P(\Delta | X_n = y),$$

so the last integral above is equal to

$$\int_{x \in S} \mu_{n-1}(dx) \int_{y \in S} p(x, dy)\, P(\Delta | X_n = y) = \int_{y \in S} (\mu_{n-1} \circ p)(dy)\, P(\Delta | X_n = y).$$

By (3) this is

$$\leq \int_{y\in S} \mu_n(dy)\, P(\Delta | X_n = y) = \int_{y\in S} \pi_n(dy)\, f_n(y)\, P(\Delta | X_n = y)$$

$$= \int_{\Omega} f_n(X_n)\, P(\Delta | X_n)\, dP = \int_{\Delta} f_n(X_n)\, dP$$

which proves the theorem. The following corollary is now a consequence of [6, p. 329].

(3.3) **Corollary.** *The random variables $f_n(X_n)$, $n \geq 0$, are uniformly integrable with respect to P_π and there is a random variable Y on Ω, measurable with respect to the tail σ-field $\bigcap_{n=0}^{\infty} \mathscr{F}_n$ of $X_0, X_1, X_2, \ldots$ and such that $\lim_{n\to\infty} f_n(X_n) = Y$ almost surely and in the mean, with respect to P_π.*

§ 4. Convergence to $c(\lambda)$ in the $\sigma(L_1, L_\infty)$-Topology

(4.1) **Theorem.** *If condition* (i) *holds, then the sequence $\{f_n(\lambda, \cdot), n \geq 0\}$ converges to the constant $c(\lambda)$ in the $\sigma(L_1, L_\infty)$-topology and hence*

$$\lim_{n\to\infty} \mu_n(B) = c(\lambda)\, \pi(B) \quad \textit{for all } B \in \mathscr{B}. \tag{4}$$

(This theorem will be strengthened below, under hypothesis (H), to convergence in total variation.)

Proof. The uniform integrability of $f_n(X_n)$, $n \geq 0$, on $(\Omega, \mathscr{F}, P)$ implies that of f_n, $n \geq 0$, on $(S, \mathscr{B}, \pi)$. By [15, p. 20] the set $\{f_n, n \geq 0\}$ is sequentially relatively compact in $L_1(S, \mathscr{B}, \pi)$ with respect to the $\sigma(L_1, L_\infty)$-topology and hence for every subsequence of $\{f_n\}$ there is a sub-subsequence $\{f_{n_k}\}$ and a non-negative function $g \in L_1(S, \mathscr{B}, \pi)$ such that

$$\lim_{k\to\infty} \int_S f_{n_k} h\, d\pi = \int_S g\, h\, d\pi \tag{5}$$

for all bounded $\mathscr{B}$-measurable functions h. It is now sufficient to prove that any such g equals $c(\lambda)$ π-almost everywhere. This will follow from Lemma 4.3 below.

First note that (5) implies that for every bounded random variable Z on $(\Omega, \mathscr{F})$

$$\int_\Omega g(X_0)\, Z\, dP = \lim_{k\to\infty} \int_\Omega f_{n_k}(X_0)\, Z\, dP \tag{6}$$

since

$$\begin{aligned}\int_\Omega g(X_0)\, Z\, dP &= \int_\Omega g(X_0)\, E(Z | X_0)\, dP \\ &= \int_S g(x)\, E(Z | X_0 = x)\, \pi(dx) \\ &= \lim_{k\to\infty} \int_S f_{n_k}(x)\, E(Z | X_0 = x)\, \pi(dx) \qquad \text{by (5)} \\ &= \lim_{k\to\infty} \int_\Omega f_{n_k}(X_0)\, Z\, dP.\end{aligned}$$

Let now $g_n(x)$ be a version of $E(g(X_0)|X_n=x)$, $n \geqq 1$.

(4.2) **Lemma.** *P_π-almost surely*

$$g(X_0)=g_1(X_1)=g_2(X_2)=\dots. \tag{7}$$

To prove this fix $m \geqq 1$ and consider the equality $g(X_0)=g_m(X_m)$. It is sufficient to prove that for every $\Delta \in \mathscr{F}$

$$\int_\Delta g(X_0)\,dP=\int_\Delta g_m(X_m)\,dP. \tag{8}$$

By (6)

$$\begin{aligned}
&\Big|\int_\Delta g(X_0)\,dP-\int_\Delta g_m(X_m)\,dP\Big| \\
&\quad=\Big|\int_\Delta g(X_0)\,dP-\int_\Delta E(g(X_0)|X_m)\,dP\Big| \\
&\quad=\Big|\int_\Omega g(X_0)\,\chi_\Delta\,dP-\int_\Omega g(X_0)\,P(\Delta|X_m)\,dP\Big| \\
&\quad=\lim_{k\to\infty}\Big|\int_\Omega f_{n_k}(X_0)\,\chi_\Delta\,dP-\int_\Omega f_{n_k}(X_0)\,P(\Delta|X_m)\,dP\Big| \\
&\quad=\lim_{k\to\infty}\Big|\int_\Delta f_{n_k}(X_0)\,dP-\int_\Delta E(f_{n_k}(X_0)|X_m)\,dP\Big| \\
&\quad\leqq \lim_{k\to\infty} E\,|f_{n_k}(X_0)-E(f_{n_k}(X_0)|X_m)| \\
&\quad\leqq \lim_{k\to\infty} E\,|f_{n_k}(X_0)-f_{n_k+m}(X_m)|+\lim_{k\to\infty} E\,|E(f_{n_k}(X_0)|X_m)-f_{n_k+m}(X_m)| \\
&\quad\leqq 2\lim_{k\to\infty} E\,|f_{n_k}(X_0)-f_{n_k+m}(X_m)|.
\end{aligned}$$

The last limit is zero by Corollary 3.3 and the fact that, under P_π, the joint distribution of (X_0, X_m) is the same as that of (X_{n_k}, X_{n_k+m}). The Lemma is thus proved.

Since the distribution of the pair (X_0, X_m) is given by $P(X_0\in dx, X_m\in dy)=\pi(dx)\,p^{(m)}(x,dy)$, (8) together with Fubini's theorem imply that for every m there is a π-null set C_m in S such that if $x\notin C_m$, then

$$g(x)=g_m(y) \qquad \text{for } p^{(m)}(x,\cdot)\text{-almost all } y\in S. \tag{9}$$

Define $D=\left(\bigcup_{m=1}^{\infty} C_m\right)^c$. Then $\pi(D)=1$ and we have:

(4.3) **Lemma.** *If condition* ($\bar{\text{i}}$) *holds, then $g(x_1)=g(x_2)$ for $\pi\otimes\pi$-almost all (x_1,x_2).*

In fact, for $\pi\otimes\pi$-almost all $(x_1,x_2)\in D\times D$ there is an m such that $p^{(m)}(x_1,\cdot)$ and $p^{(m)}(x_2,\cdot)$ are not singular with respect to each other. By (9) the sets

$$E_1=\{y\colon g(x_1)=g_m(y)\},$$
$$E_2=\{y\colon g(x_2)=g_m(y)\}$$

carry the measures $p^{(m)}(x_1,\cdot)$ and $p^{(m)}(x_2,\cdot)$ respectively. Since these two measures are not singular, we must have $E_1\cap E_2\neq\emptyset$ and choosing $y\in E_1\cap E_2$ we deduce

$$g(x_1)=g_m(y)=g(x_2).$$

It follows trivially that g equals some constant π-almost everywhere and it is easily seen that the constant is $\lim_{n\to\infty}\mu_n(S)$. This completes the proof of Theorem 4.1.

(4.4) **Corollary.** *If condition* (i) *holds and if* $c(\lambda)=1$ *then*

$$\lim_{n\to\infty}\lambda^n(B)=\pi(B) \quad \textit{for all } B\in\mathscr{B}.$$

This is true in particular if $\lambda\ll\pi$.

In fact, in the latter case the singular parts ν_n are all zero and hence $\lambda^n=\mu_n$.

We end this section with a sufficient condition for $c(\lambda)=1$, though we will not make use of this below. First some notation. If $x\in S$, we denote by δ_x the probability measure with unit mass at x and we write P_x instead of P_{δ_x}. Clearly

$$\delta_x^n(\cdot)=p^{(n)}(x,\cdot).$$

(4.5) **Proposition.** *Suppose* λ *is such that if* $\pi(B)=1$ *then* $P_x(X_n\in B$ *for some* $n)=1$ *for* λ*-almost all* x. *Then* $c(\lambda)=1$.

Proof. As before let $\lambda^n=\mu_n+\nu_n$, $\mu_n\ll\pi$, $\nu_n\perp\pi$. For each n choose a set D_n such that $\pi(D_n)=1$, $\nu_n(D_n)=0$ and define $D=\bigcap_{n=0}^{\infty}D_n$. Then

$$\pi(D)=1, \qquad \nu_n(D)=0 \quad \text{for all } n. \tag{10}$$

Observe now that the set

$$C=\{x\in D: p(x,D^c)>0\}$$

is null with respect to all the measures $\pi, \lambda, \lambda^1, \lambda^2, \ldots$, for the equality

$$0=\pi(D^c)=\int_S \pi(dx)\,p(x,D^c)$$

implies $\pi(C)=0$ and this in turn implies $\mu_n(C)=0$ for all n. However, we also have $\nu_n(C)=0$ since $C\subset D$.

We now prove that for any $A\in\mathscr{B}$

$$\nu_n(A)=P_\lambda(X_0\in D^c, X_1\in D^c, \ldots, X_{n-1}\in D^c, X_n\in A\cap D^c). \tag{11}$$

In fact

$$\begin{aligned}\nu_n(A)&=\lambda_n(A\cap D^c)=P_\lambda(X_n\in A\cap D^c)\\ &=P_\lambda(X_{n-1}\in D^c, X_n\in A\cap D^c)+P_\lambda(X_{n-1}\in D, X_n\in A\cap D^c).\end{aligned}$$

The second term is 0 because it is equal to

$$\int_D \lambda^{n-1}(dx)\,p(x,A\cap D^c)\leqq\int_D \lambda^{n-1}(dx)\,p(x,D^c)=\int_C \lambda^{(n-1)}(dx)\,p(x,D^c)=0$$

by what was shown above. So

$$\nu_n(A)=P_\lambda(X_n\in A\cap D^c)=P_\lambda(X_{n-1}\in D^c, X_n\in A\cap D^c)$$

and proceeding inductively we can show that the latter is

$$=P_\lambda(X_0\in D^c, X_1\in D^c, \ldots, X_{n-1}\in D^c, X_n\in A\cap D^c)$$

which proves (11). This now implies

$$\lim_{n\to\infty} \nu_n(S) = \lim_{n\to\infty} P_\lambda(X_i \in D^c \text{ for } i=0,1,\ldots,n)$$
$$= P_\lambda(X_i \in D^c \text{ for all } i \geqq 0)$$
$$= \int_S \lambda(dx) P_x(X_i \in D^c \text{ for all } i \geqq 0)$$

which is 0 by (10) and the hypothesis. Hence $c(\lambda)=1$ and the proposition is proved.

§5. The Countable Case

If $\lambda \perp \pi$ then the constant $c(\lambda)$ may be zero, i.e. μ_n may be 0 for all n. In such a case Theorem 4.1 asserts nothing. However this theorem already implies the classical Kolmogorov limit theorem for irreducible, aperiodic, positive-recurrent Markov chains on a countable state space I. In this case the stationary distribution is determined by $\pi(i) = E_i(T_i)^{-1}$ $(i \in I)$, where T_i is the first recurrence time for the state i (see [1, p. 144]). Since $\pi(i) > 0$ for all i, every probability distribution $\{\lambda(i)\}$ on I is absolutely continuous with respect to $\{\pi(i)\}$, so Corollary 4.4 implies

$$\lim_{n\to\infty} \sum_{i\in I} \lambda(i) p_{ij}^{(n)} = \pi(j) \qquad \text{for all } j \in I$$

and (since $\sum_j \pi(j) = 1$) also

$$\lim_{n\to\infty} \sum_{j\in I} \Big| \sum_{i\in I} \lambda(i) p_{ij}^{(n)} - \pi(j) \Big| = 0$$

or equivalently

$$\lim_{n\to\infty} \|\lambda^n - \pi\| = 0.$$

Here $p_{ij}^{(n)} = P(X_n = j \mid X_0 = i)$ and $\lambda^n(j) = \sum_{i\in I} \lambda(i) p_{ij}$. Theorem 3.2 reduces in the countable case to the assertion that the sequence $\left\{\dfrac{\lambda^n(X_n)}{\pi(X_n)}, n \geqq 0\right\}$ is a backward martingale under P_π, i.e.

$$E\left(\frac{\lambda^n(X_n)}{\pi(X_n)} \,\middle|\, X_{n+1}, X_{n+2}, \ldots\right) = \frac{\lambda^{n+1}(X_{n+1})}{\pi(X_{n+1})}.$$

§6. The Case $\lambda = \delta_x$

From now on we assume hypothesis (H). The purpose of the present section is to prove Theorem 6.1 below. The proof given here is not the shortest possible but is probabilistically simple.

(6.1) **Theorem.** *Under hypothesis* (H), *if condition* (i) *holds, then* $c(\delta_x) = 1$ *for* π*-almost all* $x \in S$.

Let $p^{(n)}(x,\cdot) = \mu_n(x,\cdot) + \nu_n(x,\cdot)$ be the decomposition of $p^{(n)}(x,\cdot)$ into absolutely continuous and singular parts with respect to π $(\mu_n(x,\cdot) \ll \pi,\ \nu_n(x,\cdot) \perp \pi)$. Under hypothesis (H), $\mu_n(x,B)$ and $\nu_n(x,B)$ are $\mathscr{B}$-measurable functions of x for fixed B ([6, p. 616–7]) and it is easily seen that

$$\nu_{n+m}(x,\cdot)\leqq \int_S \nu_n(x,dy)\,\nu_m(y,\cdot). \tag{12}$$

(6.2) **Lemma.** *$c(\delta_x)>0$ for π-almost all $x\in S$.*

Proof. This follows from the fact that $p^{(n)}(x,\cdot)\perp\pi(\cdot)$ implies $p^{(n)}(x,\cdot)\perp p^{(n)}(y,\cdot)$ for π-almost all $y\in S$. To see this suppose B is a set in $\mathscr{B}$ such that $p^{(n)}(x,B)=0$, $\pi(B)=1$. By the stationarity of π we have $\int_S \pi(dy)\,p^{(n)}(y,B)=1$, so $p^{(n)}(y,B)=1$ for π-almost all y and the lemma is proved.

This lemma implies that for π-almost every $x\in S$ there is an $n\geqq 1$ such that $\nu_n(x,S)<1$. For each $j=1,2,\dots$ define

$$F_j=\left\{x\in S\colon\ \nu_j(x,S)\leqq 1-\frac{1}{j}\right\}.$$

By (3),

$$F_1\subset F_2\subset\dots \quad\text{and}\quad \pi\left(\bigcup_{j=1}^{\infty}F_j\right)=1. \tag{13}$$

Now fix a j and consider the following two Markov processes.

(M_1) The Markov process $X_0, X_j, X_{2j},\dots$ with state space S and transition function $p^{(j)}(x,B)$.

(M_2) A Markov process $Y_0, Y_1,\dots$ whose state space is obtained by adjoining a new element ζ to S (and enlarging in an obvious way the σ-field $\mathscr{B}$), and whose transition function $q(x,B)$ is determined as follows: If $x\in S$ and $B\subset S$ then $q(x,B)=\nu_j(x,B)$, while $q(x,\{\zeta\})=1-\nu_j(x,S)=\mu_j(x,S)$ and finally $q(\zeta,\{\zeta\})=1$, so that ζ is an absorbing state.

Since π is a stationary distribution for the Markov process M_1, by the recurrence theorem we have for π-almost all $x\in F_j$.

$$P_x(X_{nj}\in S\setminus F_j \text{ eventually})=0. \tag{14}$$

If we denote by Q_x the probability measure for $Y_0, Y_1,\dots$ determined by the initial distribution δ_x and the transition function q, then (14) implies that for π-almost all $x\in F_j$

$$Q_x(Y_n\in S\setminus F_j \text{ eventually})=0$$

since $q(y,B)\leqq p^{(j)}(y,B)$ for all $y\in S$ and all $B\subset S$. Hence for π-almost all $x\in F_j$

$$Q_x(Y_n=\zeta \text{ eventually})+Q_x(Y_n\in F_j \text{ for infinitely many } n)=1.$$

However, Proposition 7 in [2] and the fact that $q(x,\{\zeta\})\geqq 1/j$ for all $x\in F_j$ now imply that for arbitrary x

$$Q_x(Y_n\in F_j \text{ for infinitely many } n)=0.$$

We deduce that

$$Q_x(Y_n = \zeta \text{ eventually}) = 1$$

for π-almost all $x \in F_j$ and by (12)

$$v_{nj}(x, S) \leqq \int_{x_1 \in S} q(x, dx_1) \int_{x_2 \in S} q(x_1, dx_2) \dots \int_{x_n \in S} q(x_{n-1}, dx_n) \leqq Q_x(Y_n \neq \zeta)$$

which implies $v_{nj}(x, S) \searrow 0$ as $n \to \infty$.

It follows that for π-almost all $x \in F_j$ $\lim_{k \to \infty} v_k(x, S) = 0$ and this combined with (13) establishes Theorem 6.1.

(6.3) **Theorem.** *Under hypothesis (H), if condition* (i) *holds, then for π-almost every $x \in S$ the following is true*

$$\lim p^{(n)}(x, B) = \pi(B) \quad \textit{for all } B \in \mathscr{B}. \tag{15}$$

Proof. This follows from Theorem 6.1 and Corollary 4.4.

Note that (15) can also be deduced from Theorem 6.1 and condition (i) via Theorem 6 in [5]. The present approach is however free from spectral considerations.

§7. Convergence in Total Variation

If the initial distribution λ in §3 is δ_x, we shall denote the function $f_n(\delta_x, y)$ by $f_n(x, y)$ for simplicity. Thus

$$p^{(n)}(x, dy) = f_n(x, y)\, \pi(dy) + v_n(x, dy).$$

Under hypothesis (H) the functions $f_n(x, y)$ can be chosen to be $\mathscr{B} \otimes \mathscr{B}$-measurable in the pair (x, y) (see [6, pp. 616–7]) and in the sequel we shall assume that they have been so chosen. It is easy to see that (3) implies that for every $n \geqq 1$, every $m \geqq 1$ and every initial distribution λ

$$f_{n+m}(\lambda, y) \geqq \int_S f_n(\lambda, u) f_m(u, y)\, \pi(du) \tag{16}$$

for π-almost all $y \in S$.

(7.1) **Lemma.** *Under hypothesis* (H), *if condition* (i) *holds, then for every initial distribution λ*

$$\liminf_{n \to \infty} f_n(\lambda, y) \geqq c(\lambda) \tag{17}$$

for π-almost all $y \in S$. In particular

$$\liminf_{n \to \infty} f_n(x, y) \geqq 1 \tag{18}$$

for $\pi \otimes \pi$-almost all $(x, y) \in S \times S$.

Proof. Let $\psi(\lambda, y) = \liminf_{k\to\infty} f_k(\lambda, y)$. By (16), for fixed m, arbitrary $r>0$ and π-almost all $y \in S$

$$\psi(\lambda, y) = \liminf_{n\to\infty} f_{n+m}(\lambda, y) \geqq \liminf_{n\to\infty} \int_S f_n(\lambda, u) f_m(u, y)\, \pi(du)$$

$$\geqq \liminf_{n\to\infty} \int_S f_n(\lambda, u) \min\{f_m(u, y), r\}\, \pi(du) = c(\lambda) \int_S \min\{f_m(u, y), r\}\, \pi(du)$$

by Theorem 4.1. Letting $r \to \infty$ we obtain

$$\psi(\lambda, y) \geqq c(\lambda) \int_S f_m(u, y)\, \pi(du) \quad \pi\text{-a.e.}$$

Integrating over an arbitrary $B \in \mathscr{B}$ leads to

$$\int_{y\in B} \psi(\lambda, y)\, \pi(dy) \geqq c(\lambda) \int_{u\in S} \int_{y\in B} f_m(u, y)\, \pi(dy)\, \pi(du) = c(\lambda) \int_{u\in S} \mu_m(u, B)\, \pi(du)$$

and letting $m \to \infty$ we see by Theorems 4.1 and 6.1 that

$$\int_{y\in B} \psi(\lambda, y)\, \pi(dy) \geqq c(\lambda)\, \pi(B)$$

which implies (17). (18) follows easily from this, by Theorem 6.1.

We are now ready to strengthen Theorem 4.1.

(7.2) **Theorem.** *Under hypothesis* (H), *if condition* (i) *holds, then for every initial distribution* λ *the sequence* $\{f_n(\lambda, \cdot), n \geqq 0\}$ *converges to the constant* $c(\lambda)$ *in* L_1*-norm:*

$$\lim_{n\to\infty} \int |f_n(\lambda, y) - c(\lambda)|\, \pi(dy) = 0 \tag{19}$$

and hence

$$\lim_{n\to\infty} \|\mu_n - c(\lambda)\, \pi\| = 0.$$

If $c(\lambda) = 1$, *then*

$$\lim_{n\to\infty} \|\lambda^n - \pi\| = 0. \tag{20}$$

In particular

$$\lim_{n\to\infty} \|p^{(n)}(x, \cdot) - \pi(\cdot)\| = 0 \tag{21}$$

for π*-almost every* $x \in S$.

Proof. (19) follows from (17) and the fact that

$$\int_S f_n(\lambda, y)\, \pi(dy) = \mu_n(S) \to c(\lambda).$$

Theorem 6.1 then implies (21).

This completes the proof of the implication (i) $\Rightarrow$ (ii) in Theorem 2.1. Note that the proof of Lemma 7.1 in fact establishes the following general Theorem:

(7.3) **Theorem.** *If hypothesis* (H) *holds and if for* π*-almost every* $x \in S$

$$\lim_{n\to\infty} \mu_n(x, B) = \pi(B) \qquad \textit{for all } B \in \mathscr{B},$$

then $\liminf_{n\to\infty} f_n(x, y) \geqq 1$ *for* $\pi \otimes \pi$*-almost all* $(x, y) \in S \times S$ *and hence*

$$\lim_{n\to\infty} \|p^{(n)}(x, \cdot) - \pi(\cdot)\| = 0$$

for π*-almost all* $x \in S$.

There is a "dual" to Theorem 7.2 *which reads as follows.*

(7.4) **Theorem.** *Assume hypothesis* (H) *and condition* (i). *If for each* $n \geqq 1$ *we define the measure* $\rho_n(\cdot, y)$ *by*

$$\rho_n(B, y) = \int_B f_n(x, y)\, \pi(dx) \qquad (B \in \mathscr{B})$$

then for π*-almost every* $y \in S$

$$\lim_{n\to\infty} \|\rho_n(\cdot, y) - \pi(\cdot)\| = 0. \tag{22}$$

The following proof reveals the analogy with the forward transition probabilities. Denote by $\mathscr{F}_n^*$ $(n \geqq 0)$ the σ-field $\mathscr{F}(\ldots, X_{-n-1}, X_{-n})$ generated by $\ldots, X_{-n-1}, X_{-n}$. The following lemma parallels 3.2.

(7.5) **Lemma.** *Under hypothesis* (H), *for* π*-almost every* $y \in S$, $\{f_n(X_{-n}, y), \mathscr{F}_n^*, n \geqq 1\}$ *is a backward supermartingale under* P_π.

Proof. Note first that (12) implies $\mu_{n+m}(x, \cdot) \geqq \int_S p^{(n)}(x, du)\, \mu_m(u, \cdot)$ from which it follows in turn that for every $n \geqq 1$, every $m \geqq 1$ and every $x \in S$

$$f_{n+m}(x, y) \geqq \int_S p^{(n)}(x, du) f_m(u, y)$$

for π-almost all $y \in S$. By Fubini's theorem there is a set $C \subset S$ with $\pi(C) = 1$ and such that if $y \in C$ then

$$f_n(x, y) \geqq \int_S p(x, du) f_{n-1}(u, y)$$

for all $n \geqq 2$ and π-almost all $x \in S$. The right-hand side is a version of

$$E(f_{n-1}(X_{-n+1}, y) \mid X_{-n} = x)$$

and hence, for all $y \in C$ and $n \geqq 2$,

$$f_n(X_{-n}, y) \geqq E(f_{n-1}(X_{-n+1}, y) \mid X_{-n}) = E(f_{n-1}(X_{-n+1}, y) \mid X_{-n}, X_{-n-1}, \ldots)$$

P-a.s., where the last equality follows from the Markov property. This proves the lemma.

(7.6) **Lemma.** *Under hypothesis* (H), *if condition* (i) *holds, then for* π*-almost every* $y \in S$

$$\lim_{n\to\infty} \int_B f_n(x, y)\, \pi(dx) = \pi(B) \qquad (B \in \mathscr{B}), \tag{23}$$

i.e. the sequence $f_n(\cdot, y)$ *converges to* 1 *in the* $\sigma(L_1, L_\infty)$*-topology.*

Proof. One shows first, as in the proof of Theorem 4.1, that every subsequence of $\{f_n(\cdot, y)\}$ contains a sub-subsequence $\{f_{n_k}(\cdot, y)\}$ which converges to some $\phi \in L_1(S, \mathscr{B}, \pi)$ in the $\sigma(L_1, L_\infty)$-topology. We shall prove that $\phi = 1$.

By assumption

$$\lim_{k\to\infty} \int_A f_{n_k}(x, y)\, \pi(dx) = \int_A \phi(x)\, \pi(dx) \qquad (A \in \mathscr{B}). \tag{24}$$

If we take $\lambda = \pi$ in (16) we see that for π-almost all $y \in S$, $\int_S f_m(u, y)\, \pi(du) \leqq 1$, since $f_n(\pi, u) = 1$. Thus the convergence in (24) is bounded and Lebesgue's theorem implies that for arbitrary $B \in \mathscr{B}$

$$\begin{aligned} \pi(B) \int_A \phi(x)\, \pi(dx) &= \lim_{k\to\infty} \int_{y\in B} \int_{x\in A} f_{n_k}(x, y)\, \pi(dx)\, \pi(dy) \\ &= \lim_{k\to\infty} \int_{x\in A} \int_{y\in B} f_{n_k}(x, y)\, \pi(dy)\, \pi(dx) \\ &= \lim_{k\to\infty} \int_{x\in A} \mu_{n_k}(x, B)\, \pi(dx) = \int_{x\in A} \pi(B)\, \pi(dx) = \pi(A)\, \pi(B). \end{aligned}$$

It follows that $\phi = 1$ π-a.e.

Combining now (23) (with $B = S$) and (18) we obtain

$$\lim_{n\to\infty} \int_S |f_n(x, y) - 1|\, \pi(dx) = 0$$

which is equivalent to (22).

§8. Appendix: The Harris Recurrence Condition and Orey's Theorem

In the present appendix we indicate briefly the connection of Theorem 2.1 with Orey's theorem ([17, 14]). The following theorem shows how Orey's result for the case under consideration can be made to follow from 2.1.

(8.1) **Theorem.** *Under hypothesis* (H), *conditions* (i) *and* (ii) *there and conditions* (iii) *and* (iv) *below are equivalent to each other.*

(iii) *For π-almost every $x \in S$ the following is true: If $\pi(B) > 0$ ($B \in \mathscr{B}$), then there is an N (depending on x and B) such that, $p^{(n)}(x, B) > 0$ for all $n \geqq N$.*

(iv) *For $\pi \otimes \pi$-almost every $(x, y) \in S \times S$ the following is true: If $\pi(B) > 0$ ($B \in \mathscr{B}$), then there is an n (depending on (x, y) and B) such that $p^{(n)}(x, B) > 0$, $p^{(n)}(y, B) > 0$.*

Proof. The theorem will follow if we complete the chain
(i) $\Rightarrow$ (ii) $\Rightarrow$ (iii) $\Rightarrow$ (iv) $\Rightarrow$ (i). The first implication was established above, while the next two are trivial. Only the implication (iv) $\Rightarrow$ (i) remains to be proved.

First note that there are versions of $f_n(x, y)$ such that

$$f_{n+m}(x, y) \geqq \int f_n(x, u) f_m(u, y)\, \pi(du) \tag{25}$$

for all x, y, n, m (see [6, p. 196]). If (iv) is true, then Lemma 2 in [13] implies that there are two sets $A, B \subset S$ and a $k \geqq 1$ such that $\pi(A) > 0$, $\pi(B) > 0$ and $f_k(x, y) > 0$ for all $x \in A$, $y \in B$. The basic idea of this lemma has been used in the majority of papers on the subject since Doeblin (cf. [6, Lemma 5.3, p. 200] and the references cited in [13, p. 23]) and it is interesting that it was not needed in the proof of Theorem 2.1 above. We will use it however for the implication (iv) $\Rightarrow$ (i) and in order to make our discussion self-contained, we outline here a proof of the above assertion, simplifying somewhat in the present context the argument in [13].

First note that (iv) implies that for π-almost all $x \in S$, $\pi(\{y \in S: f_m(x,y)>0$ for some $m\})=1$ and that consequently the set $\{(x,y) \in S^2: f_m(x,y)>0$ for some $m\}$ has $\pi \otimes \pi$-measure one. It follows from this that there are $i \geqq 1$ and $j \geqq 1$ such that the set $Q=\{(x,y,z) \in S^3: f_i(x,y)>0, f_j(y,z)>0\}$ has positive π^3-measure (where $\pi^3=\pi \otimes \pi \otimes \pi$). Since π^3-almost every point of Q is a "point of density 1" for Q, there is a "rectangle" $J \times K \times L$ in S^3 such that $\pi^3(Q \cap (J \times K \times L)) > \frac{3}{4} \pi^3(J \times K \times L)$ (cf. [6, pp. 199 and 201]). If we define A to be the set of all $x \in J$ such that $\pi(\{y \in K: f_i(x,y)>0\}) > \frac{3}{4} \pi(K)$, then it is easily seen that $\pi(A)>0$. Similarly, if we define B to be the set of all $z \in L$ such that $\pi(\{y \in K: f_j(y,z)>0\}) > \frac{3}{4}\pi(K)$, then $\pi(B)>0$. For arbitrary $x \in A$, $z \in B$ we then have $\pi(\{y \in K: (x,y,z) \in Q\}) > \frac{1}{2} \pi(K) > 0$ and so $f_{i+j}(x,z) \geqq \int_K f_i(x,y) f_j(y,z)\, \pi(dy)>0$ by (25) and all we have to do is set $k=i+j$.

Returning to the implication (iv) $\Rightarrow$ (i), if $x \in S$, $y \in S$ and $n \geqq 1$ are such that $p^{(n)}(x,A)>0$, $p^{(n)}(y,A)>0$, then by the definition of A and B, for every subset C of B with $\pi(C)>0$ we have $p^{(n+k)}(x,C)>0$, $p^{(n+k)}(y,C)>0$. It follows easily that $p^{(n+k)}(x,\cdot)$ and $p^{(n+k)}(y,\cdot)$ cannot be singular with respect to each other and hence (iv) implies (i). This completes the proof of the theorem.

Recall now that the Markov process is said to be recurrent in the sense of Harris with respect to π if it satisfies the following condition: If $\pi(A)>0$, then $P_x(X_n \in A \text{ i.o.})=1$ for all $x \in S$. If this is the case and if the period d of the process is 1 (see [17] for this) then condition (iii) as well as Theorem 6.1 follow trivially (for the latter see [17, top of p. 813]) so that in fact our section §3 becomes in this case redundant. Since in this case $c(\delta_x)=1$ for every x, and not just almost every x, (20) implies that (21) is true for all x, and this is Orey's theorem for the present case.

In the introduction we remarked that, conversely, Theorem 2.1 can be reduced to Orey's theorem by arguments given in [13].

References

1. Breiman, L.: Probability. Reading, Mass.: Addison-Wesley 1968
2. Chung, K.L.: The general theory of Markov processes according to Doeblin. Z. Wahrscheinlichkeitstheorie verw. Gebiete **2**, 230–254 (1964)
3. Doeblin, W.: Sur les propriétés asymptotiques de mouvement régis par certains types de chaines simples. Bull. Math. Soc. Sci. Math. R.S. Roumanie **39**, **1**, 57–115; **2**, 3–61 (1937)
4. Doeblin, W.: Éléments d'une théorie générale des chaines simples constantes de Markoff. Ann. Sci. École Norm. Sup. (3) **57**, No. 2, 61–111 (1940)
5. Doob, J.L.: Asymptotic properties of Markoff transition probabilities. Trans. Amer. Math. Soc. **63**, 393–421 (1948)
6. Doob, J.L.: Stochastic processes. New York: Wiley and Sons 1953
7. Foguel, S.R.: The ergodic theorem for Markov processes. Israel J. Math. **4**, 11–22 (1966)
8. Foguel, S.R.: The ergodic theory of Markov processes. New York: van Nostrand 1969
9. Harris, T.E.: The existence of stationary measures for certain Markov Processes. Proc. 3rd Berkeley Sympos. Math. Statist. Probab., **II**, 113–124 (1956)
10. Horowitz, S.: Some limit theorems for Markov processes. Israel J. Math. **6**, 107–118 (1968)

11. Jacobs, K.: Zur Theorie der Markoffschen Prozesse. Math. Ann. **133**, 375–399 (1957)
12. Jacobs, K.: Neuere Methoden und Ergebnisse der Ergodentheorie. Berlin-Göttingen-Heidelberg: Springer-Verlag 1960
13. Jain, N., Jamison, B.: Contributions to Doeblin's theory of Markov processes. Z. Wahrscheinlichkeitstheorie verw. Gebiete **8**, 19–40 (1967)
14. Jamison, B., Orey, S.: Markov chains recurrent in the sense of Harris. Z. Wahrscheinlichkeitstheorie verw. Gebiete **8**, 41–48 (1967)
15. Meyer, P.A.: Probability and Potentials. Waltham, Mass.: Blaisdell 1966
16. Neveu, J.: Bases mathématiques du calcul des probabilités. Paris: Masson 1964
17. Orey, S.: Recurrent Markov chains. Pacific J. Math. **9**, 805–827 (1959)
18. Ornstein, D., Sucheston, L.: An operator theorem on L_1 convergence to zero with applications to Markov kernels. Ann. Math. Statist. **41**, 1631–1639 (1970)
19. Yosida, K., Kakutani, S.: Operator-theoretical treatment of Markoff's process and mean ergodic theorem. Ann. Math. **42**, 188–228 (1941)

Received September 9, 1975