THE ERGODIC THEOREM FOR MARKOV PROCESSES(*)

BY

S. R. FOGUEL

ABSTRACT

Most of the material in Sections 4–5–6–8–11 has been published in [4]–[10]. We shall deal with the asymptotical behavior of the iterates of a Markov transition function. Our aim is to generalize the results about the "cyclic" convergence of the iterates of a Markov matrix. Throughout the paper functional analytic methods are used and not probabilistic arguments. The report is self contained, modulo standart results from functional analysis, except for the decomposition into conservative and dissipative parts. Also we assume the existence of an invariant σ finite measure on the conservative part. This has been proved, under some restrictions, by several authors using probabilistic methods.

1. **Definitions and notation.** Let (X, Σ, ν) be a measure space. By a measure we shall mean a finite positive measure unless otherwise mentioned (e.g. signed measure, σ finite positive measures and finitely additive measures). Let $P(x,A)$ be a Markov subtransition function on it, i.e., a function on $X \times \Sigma$ which is, for each $x \in X$ a measure of total mass $\leqq 1$ and, for each $A \in \Sigma$ a measurable function. The subtransition function induces an operator on bounded measurable functions and on signed measures by

$$(1.1) \qquad (Pf)(x) = \int f(y)P(x,dy)$$

$$(1.2) \qquad (\mu P)(A) = \int P(x,A)\,\mu(dx)$$

Thus if 1_{A_0} denotes the characteristic function of $A_0 \in \Sigma$ and δ_{x_0} the Dirac measure at x_0 then

$$(P\,1_{A_0})(x) = P(x, A_0)$$

$$(\delta_{x_0}P)(A) = P(x_{,0} A).$$

Equation 1.2 will be occasionally used for σ finite positive measures and for finitely additive measures too.

Received June 21, 1966

(*) The research reported in this document has been sponsored by the Air Force Office of Scientific Research under Grant AF EOAR 66–18, through the European Office of Aerospace Research (OAR) United States Air Force.

The two operators are related by

$$\int (Pf)(x)\mu(dx) = \int f(x)(\mu P)(dx). \tag{1.3}$$

The measure ν is assumed to satisfy

$$\nu P \prec \nu \tag{1.4}$$

(νP is absolutely continuous with respect to ν). Hence if $\nu(A) = 0$ then $P(x, A) = 0$ a.e.ν. Equation 1.4. can be always achieved if one replaces ν by $\Sigma 2^{-n}\nu P^n$.

The iterates of P are defined inductively by

$$P^n(x, A) = \int P^{n-k}(x,dy)P^k(y, A) \quad 0 < k < n. \tag{1.5}$$

This definition corresponds to the notion of powers of the operator P considered either on bounded measurable functions or on signed measures.

2. **The operator** P **on** L_1 (X, Σ, ν). Let us consider the action of P on the signed measures which are absolutely continuous with respect to ν (weaker than ν). If $\mu \prec \nu$ then $d\mu = f d\nu$ where $f \in L_1(X, \Sigma, \nu)$ is the Radon Nikodym derivative of μ with respect to ν. Let $d\mu_k = f_k d\nu$ where $f_k(x) = f(x)$ if $|f(x)| \leqq k$ and $f_k(x) = 0$ if $|f(x)| > k$. Then

$$(\mu_k P)(A) = \int P(x, A)\mu_k(dx) \leqq k \int P(x, A) d\nu.$$

Thus if $\nu(A) = 0$ then $(\mu_k P)(A) = 0$. Since $\mu_k \to \mu$ in the norm of signed measures (total variation) it follows that

$$\int P(x, A)\mu_k(dx) \to \int P(x, A)\mu(dx).$$

Therefore if $\mu \prec \nu$ then $\mu P \prec \nu$, or P leaves the subspace, consisting of signed measures that are weaker than ν, invariant.

For this section only let us denote

$$fP = g \text{ iff whenever } d\mu = f d\nu \text{ then } g = \frac{d\mu P}{d\nu} \tag{2.1}$$

This can be written as:

$$fP = g \quad iff \int_A g(x)\nu(dx) = \int P(x, A)f(x)\nu(dx). \tag{2.2}$$

Note that P on $L_1(\nu)$ is the restriction of 1.2 *not* of 1.1. Now the operator P on signed measures is a contraction operator (of norm $\leqq 1$) and maps positive measures to positive measures. Thus P on L_1 (ν) is a contraction and if $f \geqq 0$ a.e. ν then $fP \geqq 0$ a.e. ν. On the other hand we can not apply to P the classical

Ergodic Theorem since $P\,1 \neq 1$ usually. Equality would mean that ν is an invariant measure. This situation has been studied by the Chacon-Orenstein Theorem and related results. We shall cite only one result that will be used later:

The spacce X is the disjoint union of its conservative part C and its dissipative part D. These sets satisfy:

(2.3) *If $\mu \prec \nu$ then $\Sigma \mu P^n$ is σ finite on D.*

(2.4) *If $\mu \prec \nu$ then $\Sigma(\mu P^n)(A) = \infty$, unless $\Sigma(\mu P^n)(A) = 0$, whenever $A \subset C$ and $\nu(A) > 0$.*

(2.5) $$P^n 1_C = P^n(x, C) = 1_C \text{ a.e.}$$

See [12 Proposition V.5.2].

3. **Convergence on** D. Let D_j be disjoint sets whose union is D such that $\Sigma_n(\nu P^n)(D_j) < \infty$. Such sets exist by Equation 2.3.

THEOREM 1. *If $\mu \prec \nu$ then $\lim_{n\to\infty}(\mu P^n)(D_j) = 0$.*

Proof. Let $d\mu = f d\nu$ where $0 \leqq f \in L_1(\nu)$. Now $\int \Sigma_n P^n(x, D_j)\,\nu(dx) < \infty$ hence $\Sigma_n P^n(x, D_j) < \infty$ a.e. ν. Thus a.e. ν $f(x)P^n(x, D_j)_{n\to\infty} \to 0$ and the assertion follows by the Lebesgue Dominated Convergence Theorem since $P^n(x, D_j) \leqq 1$.

Most of our report will be concerned with $(\mu P^n)(A)$ where $A \subset C$. For this we shall need some results on operators in Hilbert spaces and the assumption of existence of an invariant measure on C.

4. **Processes with an invariant measure.** In the rest of this report we shall assume the existence of an invariant measure.

ASSUMPTION I. *There exists a σ finite measure λ which is equivalent to the restriction of ν to C, and $\lambda P = \lambda$.*

We shall not deal here with the problem of finding such measure under suitable conditions. Let us just mention that this is done in [3] and additional references are given there. In what will follow it will be seen that this assumption is essential. Finally in [2] it is shown that if one assumes that $\lambda P \leqq \lambda$ (subvinariance) then invariance follows.

The purpose of this section is to establish that P is a contraction operator on $L_p(C, \Sigma, \lambda)$ for every $1 \leqq p \leqq \infty$. This is well known and is given here just for completeness sake.

We shall take real spaces though all the results are valid for complex spaces as well, see [7].

Let $f = 0$ a.e. λ (hence a.e. ν) then:

$$|(Pf)(x)| \leqq (P|f|)(x) = \lim_{k \to \infty} P(\min(|f|, k))(x)$$

now

$$P(\min(|f|, k))(x) = \int P(x, dy) \min(|f(y)|, k) \leqq k P(x, \{y: f(y) \neq 0\}) = 0$$

a.e. by 1.4.

Thus P operates on $L_\infty(\lambda)$ and if $|f(x)| \leqq M$ (we may drop the a.e. part by the above argument) then $|(Pf)(x)| \leqq \int P(x, dy)|f(y)| \leqq MP(x, C) = M$.

On the other hand if $f = \Sigma c_i 1_{A_i}$ where the sets A_i are measurable and contained in C then

$$\int |(Pf)(x)| \lambda(dx) \leqq \Sigma |c_i| \int P(x, A_i) \lambda(dx) = \Sigma |c_i| \lambda(A_i).$$

Since step functions are dense in L_1 (C, Σ, λ) it follows that $\| P \|_1 \leqq 1$.

Thus by Riesz Convexity Theorem the operator P, as defined by Equation 1.1., is a contraction operator on $L_p(C, \Sigma, \lambda)$ for every $1 \leqq p \leqq \infty$.

5. **Some results on contractions in a Hilbert space.**

The next two theorems will deal with a contraction operators in a Hilbert space.

Theorem 2 has been proved in [11] by an extensive use of the authors' Dilations Theory. We shall give an elementary proof of it.

Let T be a conraction operator on the Hilbert space H. Define

$$K = \{x: \| T^n x \| = \| T^{*n} x \| = \| x \| \qquad n = 1, 2, \cdots\} \tag{5.1}$$

Now

$$\| Tx \| = \| x \| \quad \text{iff} \quad T^*Tx = x. \tag{5.2}$$

Because:

$$\| x \|^2 = \langle Tx, Tx \rangle = \langle T^*Tx, x \rangle \leqq \| T^*Tx \| \| x \| \leqq \| x \|^2$$

Where $\langle x, y \rangle$ denotes the inner product, thus equality holds in the Schwartz's inequality, this is possible only if T^*Tx is proportional to x and it follows that $T^*Tx = x$. Thus K is a subspace of H. Now if $x \in H$ then

$$\| T^n Tx \| = \| T^{n+1} x \| = \| x \| \text{ and } \| T^{*n} Tx \| = \| T^{*n-1}(T^*Tx) \|$$
$$= \| T^{*n-1} x \| = \| x \|$$

by 5.2. Thus

THEOREM 2. *The subspace K in invariant under T and T^* and T restricted to K is a unitary operator. Furthermore:*

$$\text{(5.3)} \qquad \textit{If } x \perp K \textit{ then } \operatorname{weak\,lim} T^n x = \operatorname{weak\,lim} T^{*n} x = 0.$$

Proof. Only 5.3 requires a proof since the other parts follow from the preceeding remarks.

For every $x \in H$

$$\text{(a)} \qquad \| T^{*k}T^kT^nx - T^nx \|^2 \leqq 2\| T^nx \| - 2\,\mathrm{Re}\langle T^{*k}T^kT^nx, T^nx\rangle$$
$$= 2(\| T^nx \|^2 - \| T^{n+k}x \|^2) \to 0.$$
$$n \to \infty.$$

Also

$$\text{(b)} \qquad \| T^kT^{*k}T^nx - T^nx \|^2 \leqq \| T^{*k}T^kT^{n-k}x - T^{n-k}x \|^2 \to 0.$$

Let $x \perp K$ and $y = \text{weak}\lim T^{n_i}x$ for some subsequence of the integers, n_i. Then from (a) and (b) $T^{*k}T^ky = T^kT^{*k}y = y$ or $y \in K$. But $K^{\perp}$ is invariant under T and x belongs to it hence $y \in K \cap K^{\perp} = \{0\}$. Since the Hilbert space H is weakly sequentially compact the sequence T^nx itself converges weakly to zero. The result on $T^{*n}x$ follows by symmetry.

THEOREM 3. *Let T be a contraction operator on the Hilbert space H. Then* weak lim $T^nx = 0$ *iff* lim $\langle T^nx, x\rangle = 0$.

Proof. Let $x = u + v$ where $u \in K$ and $v \perp K$. Then $\langle T^nx,x\rangle = \langle T^nu,u\rangle + \langle T^nv,v\rangle$ and $\langle T^nv,v\rangle \to 0$ always by Theorem 2. Now if $\text{weak}\lim T^nx = 0$ then clearly $\langle T^nx,x\rangle \to 0$. Conversely if $\langle T^nx,x\rangle \to 0$ then $\langle T^nu,u\rangle \to 0$ and it is enough to show that $\text{weak}\lim T^nu = 0$ since $\text{weak}\lim T^n v = 0$.

Now $u \in K$ and on K the operator T is unitary. As in the previous proof it is enough to show that if $\text{weak}\lim T^{n_i}u = w$ then $w = 0$. But $\langle w, T^ku\rangle = \lim_{n_i}\langle T^{n_i}u, T^ku\rangle = \lim_{n_i}\langle T^{n_i-k}u,u\rangle = 0$. On the other hand w is in the subspace generated by T^ku and so must be zero.

6. **The structure of** K. Let us return to P acting on $L_2(C,\Sigma,\lambda)$. Thus from now on we denote

$$\text{(6.1)} \qquad K = \{f : f \in L_2(\lambda),\ \|P^nf\| = \|P^{*n}f\| = \|f\| \quad n = 1,2,\cdots\}.$$

Notice that we do not have an explicit expression for P^* the L_2 adjoint to P. Put

$$\text{(6.2)} \qquad \Sigma_1 = \textit{the } \sigma\textit{-field generated by sets } A \textit{ with } 1_A \in K.$$

THEOREM 4. *$K = L_2(C,\Sigma_1,\lambda)$ equivalently $f \in K$ iff $f \in L_2(C,\Sigma,\lambda)$ and is Σ_1 measurable.*

Proof. We shall divide the proof into several steps.

(a) If $f \in K$ so does $|f|$:

$|P^k f| \leqq P^k|f|$ since P is order preserving hence $\|f\| = \|P^k f\| \leqq \|P^k|f|\| \leqq \|f\|$ and equality holds. In order to apply this to P^* we only need to show that P^* is order preserving as well. Assume, to the contrary, that for some $g \geqq 0$ $P^*g < 0$ on a set A of positive measure. Take A to have finite λ measure and $0 > \int_A P^*g d\lambda = \int gP1_A d\lambda \geqq 0$ a contradiction.

(b) If f and g belong to K so do $\max(f,g)$ and $\min(fg)$:

$$\max(f,g) = \tfrac{1}{2}(|f-g| + f + g)$$

$$\min(f,g) = \tfrac{1}{2}(f + g - |f-g|).$$

(c) If A and B belong to Σ_1, so do $A \cap B$ and $A \cup B$: $1_{A\cup B} = \max(1_A, 1_B)$, $1_{A\cap B} = \min(1_A, 1_B)$.

(d) If $0 \leqq f \in K$ and c is a positive constant then $\min(f,c) \in K$:

Notice first that if $0 \leqq g \leqq c$ then $(Pg)(x) = \int g(y) P(x,dy) \leqq c$. Also $P^*g \leqq c$ a.e. for otherwise if $P^*g > c$ on a set A of positive finite λ measure then

$$c\lambda(A) < \int_A P^*g d\lambda = \int g(x) P(x,A) \lambda(dx) \leqq c \int P(x,A) \lambda(dx) = c\lambda(A).$$

Thus

$$P^k P^{*k}(\min(f,c)) \leqq P^k P^{*k} f = f$$

$$P^{*k} P^k(\min(f,c) \leqq P^{*k} P^k f = f$$

Also

$$P^k P^{*k}(\min(f,c)) \leqq c, \quad P^{*k} P^k(\min(f,c)) \leqq c.$$

Therefore

$$P^k P^{*k}(\min(f,c)) \leqq \min(f,c), \quad P^{*k} P^k(\min(f,c)) \leqq \min(f,c).$$

Hence

$$P^{*k} P^k(f - \min(f,c)) = f - P^{*k} P^k(\min(f,c)) \geqq f - \min(f,c)$$

$$P^k P^{*k}(f - \min(f,c)) \geqq f - \min(f,c)$$

but $\|P^{*k} P^k\| \leqq 1$ and $\|P^k P^{*k}\| \leqq 1$ while inequality would mean that these operators have norm greater than one.

(e) If $f \in K$ then the characteristic function of $\{x : f(x) > c > 0\}$ belongs to K

Let $f_+ = \max(f, 0) \in K$.

Let $g = c^{-1}\min(f_+, c) \in K$. Then $0 \leqq g \leqq 1$ and put $h_\varepsilon = \varepsilon^{-1}\min(\varepsilon g, f_+ - \min(f_+, c))$ for every $\varepsilon > 0$. Now $0 \leqq h_\varepsilon \leqq 1$ and $h_\varepsilon \in K$. Also if $f_+(x) \geqq c + \varepsilon$ then $\varepsilon g(x) \leqq f_+(x) - c$ and $h_\varepsilon(x) = g(x)$ but $f_+(x) \geqq c$ implies that $g(x) = 1$. Thus if $f_+(x) \geqq c + \varepsilon$ then $h_\varepsilon(x) = 1$. On the other hand if $f_+(x) \leqq c$ then $f_+ - \min(f_+, c) = 0$ and $h_\varepsilon(x) = 0$.

Therefore as $\varepsilon \to 0$ h_ε tends to $1_{\{x,f(x)>c\}}$.

(f) Let us now prove the theorem by contradiction:

If $f \in K$ and is orthogonal to Σ_1 then for every positive c $\int_{\{x,f(x)>c\}} f d\lambda = 0$, by part e, hence $\lambda\{x:f(x)>c\} = 0$ and $f_+ = 0$ a.e. Apply this to $-f$ to get that $f = 0$ a.e.

THEOREM 5. *If $A \in \Sigma_1$ and is of finite λ measure then $P1_A$ and P^*1_A ($= P^{-1}1_A$ since on K P is unitary) are both characteristic functions of sets in Σ_1.*

Proof. Let us prove the theorem for $P1_A$ since the proof for P^*1_A is identical. Put $f = P1_A$ then $0 \leqq f \leqq 1$ (see part d of the preceding proof). Let $B \in \Sigma_1$ be such that on B $0 < f(x) < 1$ (f is Σ_1-measurable) put $g = (1-f)1_B$. Thus $0 \leqq g \leqq 1$, $g + f \leqq 1$ and $g \in K$. Hence $P^*(f+g) \leqq 1$ while $P^*f = 1_A$. Therefore $(P^*g)(x) = 0$ if $x \in A$ or $0 = \langle 1_A, P^*g\rangle = \langle P^*f, P^*g\rangle = \langle f, g\rangle$ but on B both f and g are positive which implies that $f = 0$ or $f = 1$ almost everywhere.

The last two theorems can be summarized: The space K is an L_2 space on which P acts as a measure preserving transformation.

7. **The non atomic part of Σ_1.** The atoms of Σ_1, are mapped into atoms by P and P^{-1}. Let Σ_2 be the purely non atomic part of Σ_1. Then P and P^{-1} are measure preserving automorphisms of Σ_2, For every $A \in \Sigma_2$ $P^n(x,A)$ assumes only the values zero or one a.e. On Σ introduce the decomposition

$$P^n(x,\cdot) = Q_n(x,\cdot) + R_n(x,\cdot) \tag{7.1}$$

$$Q_n(,\cdot) \prec \lambda \qquad R_n(x,\cdot) \perp \lambda.$$

We shall use here the argument of [3 Theorem 1]. If $A \in \Sigma_2$ then $A = \bigcup_{j=1}^{2^n} A_j$, where $\lambda(A_{j,n}) = \frac{1}{2}{}^n\lambda(A)$ and $A_{j,n} \in \Sigma_2$ and each $A_{k,n+1}$ is contained in some $A_{j,n}$, since Σ_2 is non atomic. Now for a fixed m $P^m(x,A_{j,n})$ is either zeto or one a.e. Disregarding the exceptional set $P^m(x,A_{j,n}) = 0$ except possibly for one value of j. Thus $Q_m(x,A_{j,n})$ vanishes too for every value of j except, at most one. Thus outside of a set of measure zero

$$Q_m(x,A) = Q_m(x,A_{j(x),n})$$

and the sets $A_{j(x),n}$ form a decreasing sequence and $\lambda(A_{j(x),n}) \to 0$. Because $Q_m(x,\cdot) \prec \lambda$ it follows that $Q_m(x,A) = 0$. In conclusion:

THEOREM 6. *If $A \in \Sigma_2$ and $0 < \lambda(A) < \infty$ then:*

$$\lambda\{x: Q_m(x,A) > 0\} = 0 \quad \textit{for every } m, \tag{7.2}$$

$$\lambda\{x: R_k(x,A) = 1\} \neq 0 \quad \textit{for every } k. \tag{7.3}$$

Proof. It is enough to mention that since $Q_k(x,A) = 0$ a.e. $R_k(x,A) = P^k(x,A)$ a.e. and $P^k(x,A) = (P1_A)(x)$ from which 6.3. follows by Theorem 5.

It should be noted that in [3] the negation of 7.2. was shown to be sufficient for the existence of an invariant measure. Of course there λ in 7.2. is replaced by ν but as they are equivalent, on C, it does not matter.

8. **Assume Σ_1 atomic.** In the rest of this paper we shall assume

ASSUMPTION II. *The set Σ_1 is atomic.*

By Theorem 6 it is enough to assume that there are no sets, A, of arbitrary small measure such that $\nu\{Q_m(x,A)>0\}=0$ and $\nu\{R_k(x,A)=1\}\neq 0$ for every m and k. Another way of putting it is to say that the process does not contain a deterministic subprocess. Where a subprocess is obtained by taking a subfield of Σ and is called deterministic if the transition function and its iterates assume the values zero-one a.e. From Theorem 4 follows that Σ_1 is generated by sets of finite measure and since λ is σ finite we have $\Sigma_1=\{W_i\}_{i=1}^{\infty}$. For each i the sets whose characteristic functions are $P^n 1_{W_i}$ are atoms of Σ_1. Let us denote these atoms by $P^n W_i$.

THEOREM 7. *For every atom W there exists an integer k such that $P^k W = W$.*

Proof. If $P^n W$ are not disjoint then $P^n W = P^m W$ for some $m<n$ but since P is a unitary operator on $L_2(C,\Sigma_1,\lambda)$ $P^{n-m}W=W$.

If $P^n W$ are all disjoint and μ=restriction of λ to W then $\Sigma(\mu P^n)(W)=\mu(W)=\lambda(W)$ which contradicts 2.4.

$\{W\cup PW\cup\cdots\cup P^{k-}W\}$ *is called a cycle. The integer k is called the order of W.*

Also define

(8.1) $$C_1 = \text{union of all atoms}.$$

(8.2) $$C_2 = C - C_1 = C - \cup\Sigma_1.$$

9. **The limit theorem for measures weaker than ν.** Let us study μP^n where μ is a measure on C weaker than ν and hence $\mu\prec\lambda$.

THEOREM 8. *Let $\mu\prec\lambda$ and A a set of finite λ measure.*

(a) *If $A\subset C_2$ then* $\lim_{n\to\infty}(\mu P^n)(A)=0$

(b) *If $A\subset W$ where W is an atom of order k then*

$$\lim_{n\to\infty}(\mu P^{nk+r})(A)=\lambda(W)^{-1}\lambda(A)(\mu P^r)(W).$$

Proof. Let $d\mu=fd\lambda$ where $0\leqq f\in L_1(C,\Sigma,\lambda)$. Let us first prove the result under the additional assumption that $f\in L_2(C,\Sigma,\lambda)$. Since every function in L_1 can be approximated, in the L_1 norm, by L_2 functions and P is a contraction operator on L_1 our result will follow.

(a) If A is disjoint to Σ_1, then 1_A is orthogonal to K and by Theorem 4

$$(\mu P^n)(A) = \langle P^n 1_A, f\rangle \to 0$$

(b) Let $g = 1_A - \lambda(W)^{-1}\lambda(A)1_W$ then g is supported in W hence is orthogonal to any other atom of Σ_1. On the other hand

$$\langle g, 1_W\rangle = \lambda(A \cap W) - \lambda(W)^{-1}\lambda(A)\lambda(W) = \lambda(A \cap W) - \lambda(A)$$

but $A \subset W$ hence $\langle g, 1_W\rangle = 0$. Thus g is orthogonal to K and

$$(\mu P^{nk+r})(A) = \int P^{nk+r}(x, A)f(x)\lambda(dx) = \langle P^{nk+r}1_A, f\rangle$$
$$= \langle P^{nk+r}g, f\rangle + \lambda(W)^{-1}\lambda(A)\langle P^{nk+r}1_W, f\rangle$$

The first term tends to zero by Theorem 3 while the second term is equal to $\langle P^r 1_W, f\rangle = \int P^r(x, A)f(x)\lambda(dx) = (\mu P^r)(A)$.

THEOREM 9. *Let $\mu \prec \nu$ and $\mu(C) = 0$. Let A be a set of finite λ measure.*
(a) *If $A \subset C_2$ then* $\lim(\mu P^n)(A) = 0$.
(b) *If $A \subset W$, where W is an atom of order k, then the limit of $(\mu P^{nk+r})(A)$ exists for every $0 \leqq r < k$.*

Proof. Let τ_n be the restriction of μP^n to D and σ_n its restriction to C. Thus $\mu P^n = \tau_n + \sigma_n$ and $\mu P^{n+j} = \tau_{n+j} + \sigma_{n+j} = \tau_n P^j + \sigma_n P^j$. Now

$$(\sigma_n P^j)(D) = \int_C P^j(x,D)\sigma_n(dx \leqq \int_C (1 - P^j(x, C))\sigma_n(dx) = 0 \text{ since } P(x,c) = 1_C \text{ a.e.}$$

Thus only $\tau_n P^j$ can contribute to τ_{n+j} and $\tau_{n+j} \leqq \tau_n P^j$. Thus $\tau_{n+j}(D) \leqq \int P^j(x,D)\tau_n(dx)$ $\leqq \tau_n(X) = \tau_n(D)$. Also

$$\tau_{n+j}(D) \leqq \int_D P^j(x,D)\tau_n(dx) = \int_D (1 - P^j(x, C))\tau_n(dx) = \tau_n(D) - (\tau_n P^j)(C)$$

and

$$(\tau_n P^j)(C) \leqq \tau_n(D) - \tau_{n+j}(D)$$

The right-hand side is monotonically decreasing and for every $\varepsilon > 0$ there exists an N such that $(\tau_n P^j)(C) < \varepsilon$ whenever $n \geqq N$. Therefore

(a) $$(\mu P^n)(A) = (\tau_N P^{n-N})(A) + (\sigma_n P^{n-N})(A)$$

the first term is smaller than ε since $A \subset C$ and the second term tends to zero as $n \to \infty$ by part (a). of Theorem 8.

(b) $$|(\mu P^{nk+r})(A) - (\mu P^{mk+r})(A)| \leqq$$

$$(\tau_{Nk}P^{(n-N)k+r})(A) + (\tau_{Nk}P^{(m-N)k+r})(A) + |(\sigma_{Nk}P^{(n-N)k+r})(A) - (\sigma_{Nk}P^{(m-N)k+r})(A)|$$

The first two terms are smaller than ε since $A \subset C$ and the third term tends to zero as $n,m \to \infty$ by part (b) of Theorem 8.

10. Existence of invariant measures weaker than ν.

Let $\mu = \mu P$ and $\mu \prec \nu$.

Then for each D_j

$$\mu(D_j) = (\mu P^n)(D_j) \underset{n \to \infty}{\to} 0 \quad \text{by Theorem 1.}$$

Also for every set A in C_2 of finite λ measure $\mu(A) = (\mu P^n)(A) \to 0$ by Theorem 9 part (b). Thus μ is supported by C_1. If $A \in W$ where W is an atom of order k then

$$\mu(A) = (\mu P^{nk+r})(A) \to \lambda(W)^{-1}(\mu P^r)(W)\lambda(A) = \frac{\mu(W)}{\lambda(W)}\lambda(A).$$

Now the multiplicative constant $\mu(W)/\lambda(W)$ is the same on the cycle of W by invariance of μ and λ, hence on this cycle μ is proportional to λ.

Conversely every such measure is clearly invariant.

Note that our method fails if we wish to consider σ finite invariant measures.

11. The limit theorem. To obtain the asymptotical behavior of μP^n for *every* μ we shall need a stronger assumption. Throughout the rest of the paper we shall assume.

Assumption III. *There exists an integer d such that if $\nu(A) = 0$ then* $\sup\{P^d(x,A): x \in X\} < 1$.

Let us compare this with Theorem 6:

For every x

$$R_d(x, X) = R_d(x, A_x) \text{ with } \nu(A_x) = 0$$

Thus $R_d(x,X) = P^d(x,A_x) < 1$ and by Theorem 6 the collection Σ_1 is atomic.

Thus Assumption III *implies Assumption* II.

Also this Assumption is weaker than the classical Doeblin Condition (see [1 p. 192 hypothesis D]). There one assumes the conclusion whenever $\nu(A) \leqq \varepsilon$ for some fixed $\varepsilon > 0$; also uniformity in the sets A is assumed in Doeblin Condition.

Theorem 10. *Let μ be a given measure and $\mu P^n = \tau_n + \sigma_n$ where $\tau_n \prec \nu$ and $\sigma_n \perp \nu$ then* $\lim \sigma_n(X) = 0$.

Proof. Since $\tau_{n+1} + \sigma_{n+1} = \tau_n P + \sigma_n P$ and $\tau_n P \prec \nu$ by Section 2, only $\sigma_n P$ contributes to σ_{n+1} and $\sigma_{n+1} \leqq \sigma_n P$. In particular

$$\sigma_{n+1}(X) \leqq (\sigma_n P)(X) = \int P(x, X)\sigma_n(dx) \leqq \sigma_n(X).$$

Let us assume, to the contrary, that $\lim \sigma_n(X) \neq 0$. Since σ_n are functionals over the space of bounded measurable functions and the sequence is bounded there exists a weak * limit point, σ, to σ_n where σ is a positive *finitely* additive measure. Let $Y \in \Sigma$ be such that $v(Y) = 0$ and $\sigma_n(X - Y) = 0$. Given $\varepsilon > 0$ choose n so that

$$\left|(\sigma P^d)(Y) - (\sigma_n P^d)(Y)\right| < \varepsilon$$

then

$$(\sigma P^d)(Y) \geqq (\sigma_n P^d)(Y) - \varepsilon \geqq \sigma_{n+d}(Y) - \varepsilon \geqq \lim \sigma_m(X) - \varepsilon = \sigma(X) - \varepsilon.$$

Thus

$$\sigma(X) \leqq (\sigma P^d)(Y) = \int P^d(x, Y)\sigma(dx) \leqq \sup\{P^d(x, Y): x \in X\}\sigma(X) < \sigma(X).$$

By Assumption III.

THEOREM 11. *Let μ be a given measure. Let A be a set of finite λ measure.*

(a) $\lim_{n\to\infty} (\mu P^n)(D_j) = 0$

(b) *If $A \subset C_2$ then* $\lim (\mu P^n)(A) = 0$.

(c) *If $A \subset W$ where W is an atom of order k then the limit of* $(\mu P^{nk+r})(A)$ *exists for every* $0 \leqq r < k$.

Proof. Let $\varepsilon > 0$ be given and choose N so that $\mu P^N = \sigma_N + \tau_N$ as in Theorem 10 and $\sigma_n(X) < \varepsilon$ $n \geqq N$.

(a) $$((\mu P^N)P^m)(D_j) = (\sigma_N P^m)(D_j) + (\tau_N P^m)(D_j)$$

the first term is smaller than $\sigma_N(X) < \varepsilon$ while the second term tends to zero when $m \to \infty$ by Theorem 1.

(b) Again choose as in (a) and

$$((\mu P^N)P^m)(A) = (\sigma_N P^m)(A) + (\tau_N P^m)(A)$$

The first term is bounded by $\sigma_N(X)$ and the second tends to zero when $m \to \infty$ by Theorem 9 part (a) and Theorem 8 part (a).

(c) Let N be as in part (a) Then

$$\left|(\mu P^{nk+r})(A) - (\mu P^{mk+r})(A)\right| \leqq (\sigma_{Nk} P^{(n-N)k+r})(A) +$$
$$(\sigma_{Nk} P^{(m-N)k+r})(A) + \left|(\tau_{Nk} P^{(n-N)k+r}))(A) - (\tau_{Nk} P^{(m-N)k+r})(A)\right|$$

The sum of the first two terms is smaller than $2\sigma_{Nk}(X) < 2\varepsilon$ while the third term tends to zero as $n, m \to \infty$ by Theorems 8–9 part b.

If we choose $\mu = \delta_x$ we get:

COROLLARY. *For every*

$$x \in X \text{ and } A \text{ with } \lambda(A) < \infty:$$

(a) $\lim_{n\to\infty} P^n(x, D_j) = 0$.

(b) *If* $A \subset C_2$ *then* $\lim P^n(x, A) = 0$.

(c) *If* $A \subset W$ *where* W *is an atom of order* k *then the limit of* $P^{nk+r}(x, A)$ *exists for every* $0 \leqq r < k$.

REMARK. Let μ be an invariant measure. Then $\mu = \mu P^n$ and by Theorem 10 $\mu \prec \nu$. Thus all the result of Section 10 are valid.

REFERENCES

1. L. J. Doob, *Stochastic processes*, Wiley, New York 1953.

2. J. Feldman, *Subinvariant measures for Markov operators*, Duke Math. J. **29** (1962), 71–98.

3. ———, *Integral kernels and invariant measures for Markov transition functions*, Annals Math. Stat. **36** (1965), 517–523.

4. S. R. Foguel, *Weak and strong convergence for Markov processes*, Pacific J. Math. **10** (1960), 1221–1234.

5. ———, *Strongly continuous Markov processes*, Pacific J. Math. **11** (1961), 879–888.

6. ———, *Markov processes with stationary measure*, Pacific J. Math. **12** (1962), 505–510.

7. ———, *On order preserving contractions*, Israel J. Math. **1** (1963), 54–59.

8. ———, *Powers of a contraction in Hilbert space*, Pacific J. Math. **13** (1963), 551–562.

9. ———, *An L_p theory for a Markov process with a subinvariant measure*, Proc. Amer. Math. Soc. **16** (1965), 398–406.

10. ———, *Limit theorems for Markov processes*. Trans. Amer. Math. Soc. **121** (1966), 200–209.

11. B. Sz-Nagy and C. Foias, *Sur les contractions de l'espace de Hilbert* IV, Acta Sci. Math. **21** (1960), 251–259.

12. J. Neveu, *The calculus of probability*, Holden-Day, San Francisco, 1965.

THE HEBREW UNIVERSITY
OF JERUSALEM