ERGODIC DECOMPOSITIONS INDUCED
BY CERTAIN MARKOV OPERATORS

BY

BENTON JAMISON[1]

0. We consider transition probability operators $P(x,E)$ on a compact metric space which satisfy a rather strong condition (uniform stability in mean). §1 is devoted largely to notation and terminology. In §2 an ergodic decomposition is obtained. It is essentially a sharpening of a decomposition of Yosida's valid under less stringent conditions [17]. In §3 we discuss the behavior of sample paths of the Markov processes themselves relative to the ergodic decomposition obtained in §2. In §4 we give some examples of transition probability operators which are uniformly stable in mean.

Most of the results of §2 and some of the results of §4 appeared in the author's thesis, written under the guidance of Professor Blackwell. The results of §3 had to await Breiman's discovery of a new version of the strong law of large numbers. Our Theorem 3.2 is a minor generalization of Breiman's theorem and our proof a slight modification of his.

1. S is a compact metric space. Its metric is denoted by d. Σ is the σ-field generated by the open subsets of S. $C(S)$ is the Banach space of real-valued continuous functions on S with the uniform norm. The dual $C^*(S)$ consists of all the finite countably additive measures on Σ. The value of a member μ of $C^*(S)$ at a member f of $C(S)$ will be denoted by either (f,μ) or $\int f d\mu$. A countably additive measure μ on any σ-field over any set Ω is called a *probability measure* if it is non-negative and if $\mu(\Omega) = 1$. We sometimes refer to probability measures in $C^*(S)$ as *distributions*.

A function P on $S \times \Sigma$ is called a transition probability operator if

(i) $P(x, \cdot)$ is a probability distribution for each $x \in S$,

(ii) $P(\cdot, E)$ is a Σ-measurable function for each $E \in \Sigma$.

Unless otherwise stated, we will from now on assume that P is a fixed transition probability operator. We will abuse the language by using '$P(x,E)$' rather than 'P' to denote not only a value of P but P itself. Associated with $P(x,E)$ are operators T and U, the first defined on bounded Σ-measurable functions f by $(Tf)(x) = \int f(y)P(x,dy)$, $x \in S$, the second defined on $C^*(S)$ by $(U\mu)(E) = \int P(x,E)\mu(dx)$, $E \in \Sigma$. Functions fixed under T and measures fixed under U are called *invariant*.

We will assume that T takes $C(S)$ into itself. The restriction of T to $C(S)$, also denoted by T, is a non-negative linear operator of norm 1, with U as its adjoint T^*. Henceforth we use 'T^*' to denote U. Iterates of $P(x,E)$ are defined inductively by $P^{n+1}(x,E) = \int P^n(y,E)P(x,dy)$, $n = 1,2,\cdots$, where $P^1(x,E) = P(x,E)$. $P^n(x,E)$ is a transition probability operator for each n, and we have $(T^n f)(x) = \int f(y)P^n(x,dy)$, $(T^{*n}\mu)(E) = \int P^n(x,E)\mu(dx)$.

If $A \subset S$, I_A is the function on S equal to 1 on A and to 0 on A^c. If U is a nonempty open subset of S, there is in $C(S)$ a sequence $0 \leq f_n \uparrow I_U$. It follows that for each $k = 1,2,\cdots$, $P^k(x,U)$ is a lower semi-continuous function of x, and hence that $\{x : P^k(x,U) > 0\}$ is open whenever U is open while $\{x : P^k(x,F) = 1\}$ is closed whenever F is closed.

Let $\Omega = \prod_{n=1}^{\infty} S_n$, where $S_n = S$ for each $n = 1,2\cdots$. Let $\mathscr{A}$ be the product σ-field $\prod_{n=1}^{\infty}\Sigma_n$, where $\Sigma_n = \Sigma$ for each $n = 1,2,\cdots$. Then, for each n, X_n is the S-valued function on Ω defined for each $\omega = (\omega_1,\omega_2,\cdots, \omega_n,\cdots)$ in Ω by $X_n(\omega) = \omega_n$. Each X_n is $(\mathscr{A} - \Sigma)$-measurable (that is, $X_n^{-1}(E) \in \mathscr{A}$ for each $E \in \Sigma$). Let μ be any probability measure on S. Then there is a unique probability measure P_μ on $\mathscr{A}$ such that for each $E \in \Sigma$

(i) $P_\mu(X_1 \in E) = \mu(E)$,

(ii) $P_\mu(X_{n+1} \in E \mid X_n = x) = P(x,E)$,

for $n = 1,2,\cdots$ and μ-almost all $x \in S$. (For a discussion of conditional probabilities expectations and distributions, see [14, Chapter 7]. For details on the construction and uniqueness of P_μ, see [14, pp. 362–366].) The sequence $\{X_n, n \geq 1\}$ referred to the probability space $(\Omega, \mathscr{A}, P_\mu)$ is called *the process with initial distribution* μ. The subscript μ of 'P_μ' will usually be omitted in discussions where μ is fixed. We then have $E(f(X_{n+k}) \mid X_n) = (T^k f)(X_n)$ P_μ-almost surely, and $(T^k f)(x) = E(f(X_{n+k}) \mid X_n = x)$ μ-almost everywhere.

We shall require that $P(x,E)$ satisfy a rather strong condition, of which there are several equivalent formulations.

1.1. THEOREM. *The following conditions are equivalent*:

(a) *For each* $f \in C(S)$, *the sequence* $\{(1/n)\sum_{k=1}^{n} T^k f\}$ *converges pointwise to an* $\bar{f} \in C(S)$.

(b) *For each* $f \in C(S)$, *the sequence* $\{(1/n)\sum_{k=1}^{n} T^k f\}$ *converges uniformly to a function* $\bar{f}$, *and* $T\bar{f} = \bar{f}$.

(c) *For each* $f \in C(S)$, *the sequence* $\{(1/n)\sum_{k=1}^{n} T^k f\}$ *is equicontinuous*.

(d) *For each* $x \in S$, *the sequence* $\{(1/n)\sum_{k=1}^{n} P^k x, \cdot)\}$ *converges in the weak* topology of $C^*(S)$ to a (necessarily invariant) limiting probability distribution* $= \phi_x$, *and the map* $x \to \phi_x$ *of S into $C^*(S)$ is continuous relative to the weak* topology on $C^*(S)$.

Proof. Since $\| T^n \| = 1$ for all n, it follows from 2 and 3 on p. 662 of [5] that $\{(1/n)\sum_{k=1}^{n} T^k f\}$ converges in norm to an invariant limit $\bar{f}$ if it is weakly sequentially compact (hence of course if it is strongly compact). But weak con-

vergence in $C(S)$ is bounded pointwise convergence, so (a) implies (b). If (c) holds, the sequences $\{(1/n) \sum_{k=1}^{n} T^k f\}$ are strongly compact by virtue of Ascoli's theorem [5 p. 266]. Hence (c) implies (b). Since (b) implies both (a) (immediate) and (c) (Ascoli's theorem), (a), (b) and (c) are equivalent.

Suppose (b) holds. For each $x \in S$, the map $f \to \bar{f}(x)$ is a bounded linear functional on $C(S)$, hence corresponds to a measure ϕ_x. The function 1 is mapped into the number 1, and non-negative functions are mapped into non-negative numbers. Thus ϕ_x is a probability measure. Since $(f, \phi_x) \leftarrow (1/n) \sum_{k=1}^{n} (T^k f)(x) = ((1/n) \sum_{k=1}^{n} T^k f, \delta_x) = (f, (1/n) \sum_{k=1}^{n} T^{*k} \delta_x)$, ϕ_x is the weak* limit of $(1/n) \sum_{k=1}^{n} T^{*k} \delta_x = (1/n) \sum_{k=1}^{n} P^k(x, \cdot)$. (Here δ_x denotes the measure concentrating unit mass at x.) It is easily shown that ϕ_x is invariant. To state that $x \to \phi_x$ is continuous relative to the weak* topology on $C^*(S)$ is to state that $\bar{f}$ is continuous, which is part of our assumption (b). Thus (d) holds, so (b) implies (d). The argument can be reversed to show that (d) implies (a), which completes the proof of the theorem.

1.2. DEFINITION. Both $P(x, E)$ and the associated operator T are called *uniformly stable in mean* if any (hence all) of conditions (a)–(d) of the preceding theorem are satisfied.

From now on we shall assume that $P(x, E)$ is uniformly stable in mean. We shall adhere to the notation used in 1.1. Thus for each $x \in S$, ϕ_x denotes the weak* limit of the sequence $\{(1/n) \sum_{k=1}^{n} P^k(x, \cdot)\}$, while, for each $f \in C(S)$, $\bar{f}$ denotes the uniform limit of the sequence $\{(1/n) \sum_{k=1}^{n} T^k f\}$.

In the next lemma we collect several facts for future reference,

1.3. LEMMA. (a) *For each* $x \in S$ *and* $f \in C(S)$.

$$\bar{f}(x) = \int f d\phi_x = \int \bar{f} d\phi_x.$$

(b) *For each* $E \in \Sigma$, *the map* $x \to \phi_x(E)$ *is a* Σ-*measurable function.*

(c) *If* ϕ *is an invariant probability distribution, then*

$$\phi(E) = \int \phi_x(E) \phi(dx)$$

for each $E \in \Sigma$.

Proof. (a) The first equation is really the definition of ϕ_x, for in the proof of 1.1, ϕ_x was introduced as the measure corresponding to the linear functional $f \to \bar{f}(x)$. Since $\bar{f}$ and ϕ_x are invariant, we have $(\bar{f}, \phi_x) = (\bar{f}, (1/n) \sum_{k=1}^{n} T^{*k} \phi_x) = ((1/n) \sum_{k=1}^{n} T^k \bar{f}, \phi_x)$; letting $n \to \infty$, the second equation follows.

(b) Let Σ' be the subclass of Σ consisting of sets E for which (b) holds. Let U be any open set. Let $\{f_n\} \subset C(S)$ be such that $f_n \uparrow I_U$. Then $\phi_x(U) = \int I_U d\phi_x = \int \lim f_n d\phi_x = \lim \int f_n d\phi_x = \lim \bar{f}_n(x)$, hence $x \to \phi_x(U)$ is Σ-measurable. Thus Σ' contains all open sets. Suppose U and V are both open. Then $\phi_x(U - V)$

$= \phi_x(U) - \phi_x(UV)$ is a Σ-measurable function of x. Since Σ' is clearly closed under finite disjoint unions, it contains the subfield of Σ consisting of finite disjoint unions of differences of open sets. Since Σ' is clearly closed under nondecreasing unions and nonincreasing intersections, it follows that $\Sigma' = \Sigma$.

(c) Again, it suffices to prove (c) for E open. First, if $f \in C(S)$ and ϕ is invariant, it is clear that $\int \bar{f} d\phi = \int \check{f} d\phi$. Thus if $E = U$, where U is open, and continuous $f_n \uparrow I_U$ we have $\phi(U) = \lim \int \check{f}_n(x) \phi(dx)$. But $\check{f}_n(x) \int f_n d\phi_x$ by (a), so, by a double application of the bounded convergence theorem we have

$$\phi(U) = \int \left\{ \int I_U d\phi_x \right\} \phi(dx) = \int \phi_x(U) \phi(dx).$$

2. We now proceed with the ergodic decomposition of S. We follow Yosida [17], but our hypotheses are stronger than his, and we obtain more and sharper conclusions.

2.1. DEFINITION. $\sim$ is the equivalence relation defined for $(x,y) \in S \times S$ by either

$$x \sim y \Leftrightarrow \check{f}(x) = \check{f}(y) \quad \text{for all } f \in C(S)$$

or

$$x \sim y \Leftrightarrow \phi_x = \phi y.$$

The partition of S induced by $\sim$ is denoted by $\mathscr{D}$. If $D \in \mathscr{D}$, the ϕ_x common to all $x \in D$ is denoted by ϕ_D.

The reason that the two definitions of $\sim$ are equivalent is that ϕ_x is the measure corresponding to the functional $f \to \check{f}(x)$. Since S is compact metric, $C(S)$ contains a dense subsequence, and if $f_1, f_2, \cdots$ is any sequence dense in $C(S)$, then clearly $x \sim y$ if and only if $\check{f}_n(x) = \check{f}_n(y)$ for each $n = 1, 2, \cdots$.

We say that a subset A of S is *stochastically closed* if $A \in \Sigma$, $A \neq \varnothing$, and $P(x,A) = 1$ for all $x \in A$. Then we have $P^n(x,A) = 1$ for every $x \in A$ and $n = 1, 2, \cdots$. A simple computation shows that if ϕ is an invariant distribution and A is stochastically closed, then $\phi \{x : x \in A^c \text{ and } P(x,A) > 0\} = 0$.

2.2. LEMMA. (a) *All members of $\mathscr{D}$ are closed. Let $D \in \mathscr{D}$. Then:*

(b) *if ϕ is an invariant distribution, then $\phi(D) = 1$ implies $\phi = \phi_D$. If A is a stochastically closed subset of D, then:*

(c) $\phi_D(\bar{A}) = 1$,

(d) *if $\phi_D(A) > 0$, then $\phi_D(A) = 1$.*

Proof. (a) $D = \{x : x \in S \text{ and } \phi_x = \phi_D\}$. Thus D is the inverse image of $\{\phi_D\}$ under the map $x \to \phi_x$. But this is continuous relative to the weak* topology on $C^*(S)$, so D is closed.

(b) This is an immediate consequence of 1.3(c).

(c) Let $x \in A$. If $f \in C(S)$ has support in the complement of $\bar{A}$, then clearly

$(f, T^{*n}\delta_x) = \int f(y)P^n(x,dy) = 0$. Taking averages and letting $n \to \infty$, we have $(f, \phi_x) = 0$. Since this is true for any f with support in $(\bar{A})^c$, it follows that $(\bar{A})^c$ has ϕ_x-measure zero, so $\phi_x(\bar{A}) = 1$. But $\phi_D = \phi_x$.

(d) Suppose that $0 < \phi_D(A) < 1$. Then the measure ψ on Σ defined by $\psi(E) = \phi_D(AE)/\phi_D(A)$, $E \in \Sigma$ is easily seen to be an invariant distribution distinct from ϕ_D for which $\psi(D) = 1$. But this contradicts (b).

We will now show that, for a given $D \in \mathscr{D}$, either $\phi(D) = 0$ for all invariant distributions ϕ or else D contains a topologically and stochastically closed subset A for which $\phi_D(A) = 1$. The totality of points contained in the second sort of D turns out to be a closed set of maximal probability (we say that a subset B of S is *of maximal probability* if $B \in \Sigma$ and $\phi(B) = 1$ for all invariant distributions ϕ).

2.3. Lemma. *Suppose $D \in \mathscr{D}$ and $\phi_D(D) = 1$. Then D contains a topologically and stochastically closed subset A such that $\phi_D(A) = 1$.*

Proof. Let ϕ be any invariant probability distribution. Then for any $E \in \Sigma$, $\phi(E) = \int P(x,E)\phi(dx)$. If $\phi(E) = 1$, we have $\phi(E) = \int_E P(x,E)\phi(dx)$. It follows that $\phi\{x: x \in E \text{ and } P(x,E) < 1\} = 0$. Also, if E is (topologically) closed, so is $\{x: x \in E \text{ and } P(x,E) = 1\}$. Suppose $D \in \mathscr{D}$ and $\phi_D(D) = 1$. Let $A_0 = D$, and define $A_{n+1} = \{x: x \in A_n \text{ and } P(x,A_n) = 1\}$, $n = 1, 2, \cdots$. Since A_0 is closed and of ϕ-measure 1, the same is true of all the A_n's. Setting $A = \bigcap_{n=1}^{\infty} A_n$, we see easily that A is stochastically closed, of ϕ-measure 1, and topologically closed. This completes the proof.

2.4. Definition. $\mathscr{E} = \{D: D \in \mathscr{D} \text{ and } \phi_D(D) = 1\}$. We will write $\Sigma\mathscr{E}$ for $\bigcup\mathscr{E}$. The members of $\mathscr{E}$ are called *ergodic sets*.

2.5. Theorem. $\Sigma\mathscr{E}$ *is a closed set of maximal probability.*

Proof. Let ϕ be any invariant distribution. If g is an invariant function, then $\int\int(g(y) - g(x))^2\phi_x(dy)\phi(dx) = \int\int g^2(y)\phi_x(dy)\phi(dx) - 2\int g(x)\int g(y)\phi_x(dy)\phi(dx) + \int\int g^2(x)\phi_x(dy)\phi(dx)$. But $\int\int g^2(y)\phi_x(dy)\phi(dx) = \int g^2(x)\phi(dx)$ by 1.3(c). $\int g(y)\phi_x(dy) = g(x)$ by 1.3(a) (for $\bar{g} = g$), while $\int\int g^2(x)\phi_x(dy)\phi(dx) = \int g^2(x)\phi(dx)$ simply because $\phi_x(S) = 1$. It follows that

$$\int\int (g(y) - g(x))^2\phi_x(dy)\phi(dx) = 0.$$

Now let $f_1, f_2, \cdots, f_n, \cdots$ be dense in $C(S)$. For $i = 1, 2, \cdots$, let $H_i = \{x: \phi_x\{y: \bar{f_i}(y) = \bar{f_i}(x)\} = 1\} = \{x: \int(\bar{f_i}(y) - \bar{f_i}(x))^2\phi_x(dy) = 0\}$. Taking $g = \bar{f_i}$ in the above, we have that $\phi(H_i) = 1$. We claim also that H_i is closed. For H_i is the inverse image of $\{0\}$ under the map $x \to \int(\bar{f_i}(y) - \bar{f_i}(x))^2\phi_x(dy)$, and expansion of the integrand followed by application of 1.1(d) yields the continuity of said map. It follows that $\bigcap_{i=1}^{\infty}H_i$ is closed and of ϕ-measure 1. Since ϕ is an arbitrary invariant distribution. $\bigcap H_i$ is of maximal probability. But it is easy to see that $\bigcap H_i = \Sigma\mathscr{E}$, so the theorem is proved.

Thus, for a given $D \in \mathscr{D}$, either $D \in \mathscr{E}$, in which case, by 2.3, D contains a stochastically and topologically closed subset A for which $\phi_D(A) = 1$, or else $D \notin \mathscr{E}$, in which case $D \subset (\sum \mathscr{E})^c$, whence by 2.4, $\phi(D) = 0$ for all invariant distributions ϕ (including, of course, ϕ_D). Thus it is impossible that $0 < \phi_D(D) < 1$ for some $D \in \mathscr{D}$.

Suppose $S = [0,1]$, and $P(x,E) = I_E(\phi(x))$, where $\phi(x) = x/2$. Then $Tf = f \circ \phi$, $f \in C(S)$. $T^n f \to f(0)$ uniformly as $n \to \infty$, and so $\phi_x = \delta_0$ for all $x \in S$. Thus $\mathscr{D} = \mathscr{E} = \{[0,1]\}$. Any closed set A which contains 0 and has the property that $x/2 \in A$ if $x \in A$ is a topologically and stochastically closed set. Thus there may be multitudes of stochastically and topologically closed sets in a member D of $\mathscr{E}$. It is true, of course, that such sets are determined up to sets of ϕ_D-measure zero, but this may not tell us very much (suppose, for instance, that ϕ_D concentrates on one point). In the present example it is quite evident that there is a minimal topologically and stochastically closed subset of S, namely $\{0\}$. We next show that each $E \in \mathscr{E}$ contains such a set. We say that a subset F of S is an *ergodic kernel* if it is stochastically and topologically closed and contains no proper subset with both these properties.

2.6. THEOREM. *Every $E \in \mathscr{E}$ contains one and only one ergodic kernel F, and $\phi_E(F) = 1$. Every ergodic kernel is contained in some member of $\mathscr{E}$.*

Proof. Let $E \in \mathscr{E}$. Let $\mathscr{F}$ be the class of all topologically and stochastically closed subsets of E. By 2.3, $\mathscr{F}$ is nonvoid. By 2.2 (c), any member of $\mathscr{F}$ has ϕ_E-measure 1; thus $\mathscr{F}$ has the finite intersection property. It follows that $F = \bigcap \mathscr{F}$ is nonvoid. The compact metric space S satisfies the second axiom of countability; that is, any collection $\mathscr{G}$ of open sets has a countable subcollection $\{G_1, G_2, \cdots\}$ such that $\bigcup \mathscr{G} = \bigcup_j G_j$. Taking complements, we see that we may select a sequence $\{F_j\}$ from $\mathscr{F}$ such that $F = \bigcap_j F_j$. Since $\phi_E(F_j) = 1$ for each j, $\phi_E(F) = 1$. Similarly $P(x,F) = 1$ for each $x \in F$. F is of course closed. The fact that the intersection F of all stochastically and topologically closed subsets of E is stochastically and topologically closed implies not only that F is an ergodic kernel but also that it is the only one contained in E.

Suppose F is an ergodic kernel. Let $x \in F$. Then $\phi_x(F) = 1$; since we also have $\phi_x(\sum \mathscr{E}) = 1$ it follows that F intersects some ergodic set E. Let G be the ergodic kernel contained in E. Then $\phi_x(G) = \phi_E(G) = 1$, so $F \cap G$, being of ϕ_x-measure 1, is nonvoid. But two ergodic kernels are either disjoint or identical, so $F = G \subset E$. This completes the proof of the theorem.

2.7. LEMMA. *Let F be a topologically and stochastically closed set. Then F is an ergodic kernel if and only if, for every $x \in F$ and nonempty U open in F, there is a positive integer n with $P^n(x,U) > 0$.*

Proof. Suppose F is a topologically and stochastically closed set with the

property of the lemma. Let G be a closed set properly contained in F. Then $F - G$ is nonempty and open in F. Let $x \in G$. Then there is an n such that $P^n(x, F - G) > 0$, so G is not stochastically closed. Thus F is minimal with respect to the property of being both topologically and stochastically closed, hence is an ergodic kernel.

Suppose F is an ergodic kernel. Suppose there is a nonempty set U open in F and an $x \in F$ such that $P^n(x, U) = 0$ for all $n = 1, 2, \cdots$. Let $U = F \cap V$, where V is open in S. Then $P^n(x, V) = 0$ for all $n = 1, 2, \cdots$. Let $H = \{y : P^n(y, V) = 0$ for all $n = 1, 2, \cdots\} = \bigcap_{n=1}^{\infty} \{y : P^n(y, V) = 0\} = \bigcap_{n=1}^{\infty} \{y : P^n(y, V^c) = 1\}$. H is closed. Letting $G = F \cap H$, we see that G is a topologically closed subset of F which is nonempty (since $x \in G$). Assume for the moment that $P(y, G) = 1$ for all $y \in G$. Then G is a stochastically and topologically closed subset of the ergodic kernel F, hence $G = F$. In other words, $P(y, U) = 0$ for all $y \in F$. If U were not equal to F then $F \sim U$ would be a proper topologically and stochastically closed subset of F, which is again a contradiction. So we must have $U = F$; but then $P(y, F) = 0$ for all $y \in F$, which once more contradicts the assumption that F is an ergodic kernel. Thus there cannot be an $x \in F$ and a U open in F for which $P^n(x, U) > 0$ for some $n = 1, 2, \cdots$.

It remains only to prove that $P(y, G) = 1$ for all $y \in G$. Suppose there is a $y \in G$ with $P(y, G^c) > 0$. Then there is an n for which $P(y, \{z : P^n(y, V) = 0\}^c) > 0$. Let $E = \{z : P^n(y, V) = 0\}^c = \{z : P^n(y, V) > 0\}$. Then $P^{n+1}(y, V) \geq \int_E P(y, dz) P^n(z, V) > 0$. Contradiction. So $P(y, G)$ is indeed equal to 1 for all $y \in G$, and the lemma is proved.

2.8. LEMMA. *Let K be the ergodic kernel contained in the ergodic set E. Then, if U is any open set whose intersection with K is nonvoid, we have $\phi_E(U) > 0$.*

Proof. Let K be the ergodic kernel contained in $E \in \mathscr{E}$. Since K is topologically closed it is a compact metric space. If we define $P_K(x, B) = P(x, B)$ for $x \in K$ and members B of Σ which are subsets of K, P_K is a transition probability operator with state space K. Thus to prove the lemma it suffices (by virtue of the preceding lemma) to show that, if for every $x \in S$ and nonempty open subset U of S there is an n with $P^1(x, U) > 0$, then any invariant distribution ϕ assigns positive measure to every open set. Let $U \neq \varnothing$ be open. For each $n = 1, 2, \cdots$, let $S(U, n) = \{x : P^n(x, U) > 0\}$. $S(U, n)$ is open for each n. By hypothesis, the collection $\{S(U, n) : n = 1, 2, \cdots\}$ covers S. Since S is compact, there are positive integers $n_1, \cdots, n_k$ such that $S = \bigcup_{j=1}^{k} S(U, n_j)$. Let $N = \max\{n_1, \cdots, n_k\}$. Then $\sum_{n=1}^{N} P^n(x, U)$ is positive for each $x \in S$. But $P^n(\cdot, U)$ is lower semi-continuous for each n, so $\sum_{n=1}^{N} P^n(\cdot, U)$ is also lower semi-continuous. A lower semi-continuous function on a compact set assumes its infimum, so there is a $\delta > 0$ such that $\sum_{n=1}^{N} P^n(x, U) \geq \delta$ for all $x \in S$. Let ϕ be any invariant distribution. Then $\phi(U) = \int (1/N) \sum_{n=1}^{N} P^n(x, U) \phi(dx) \geq \delta/N > 0$. This completes the proof of the lemma.

Let α be the collection of invariant distributions. It is obvious that α is convex, and a simple argument shows that it is compact relative to the weak* topology on $C^*(S)$. If A is a convex set in a linear space, then $x \in A$ is called an *extremal point* of A if x is not a convex combination of two members of A distinct from x. Since α is convex and weak* compact, it is, by virtue of the Krein-Milman theorem [5, p. 440], the weak* closure of the convex hull of its extremal points. We conclude this section by identifying the extremal points of α with the distributions $\phi_E, E \in \mathscr{E}$.

If $\phi \in \alpha$, we say that ϕ is *indecomposable* if $\phi(A)$ is equal to 0 or 1 for any stochastically closed subset A of S.

2.9. Theorem. $\phi \in \alpha$ is indecomposable if and only if $\phi = \phi_E$ for some $E \in \mathscr{E}$.

Proof. Suppose that $\phi \in \alpha$ is indecomposable. Let $\{X_n, n \geq 1\}$ be the process with initial distribution ϕ and transition probability $P(x, E)$. Since $\phi \in \alpha$, the process is stationary, and if $f \in C(S)$ the process $\{f(X_n), n \geq 1\}$ is also stationary. It follows from the ergodic theorem for stationary process [4, Theorem 2.1, p. 465] that $(1/n) \sum_{k=1}^{n} f(X_k)$ converges almost surely to an invariant limit Y as $n \to \infty$. It follows from the Markovian nature of the process $\{X_n\}$ that Y is ϕ-almost everywhere measurable with respect to the σ-field generated by X_1 [4, Theorem 1.1, p. 460]. But since ϕ is indecomposable and Y invariant it follows that Y is almost surely constant, and therefore must equal $\int f d\phi$. If we now form the conditional expectation of $(1/n) \sum_{k=1}^{n} f(X_k)$ with respect to X_1, let $n \to \infty$, and then interchange lim and $E\{\cdot \mid X_0\}$ (see [14, p. 348]) we obtain $(1/n) \sum_{k=1}^{n} (T^k f)(X_0) = (1/n) \sum_{k=1}^{n} E\{f(X_{k+1}) \mid X_0\} \to \int f d\phi$ almost surely. Thus $\phi\{x: (1/n) \sum_{k=1}^{n} (T^k f)(x) \to \int f d\phi\} = 1$ for any $f \in C(S)$. It easily follows from the separability of $C(S)$ that $\phi\{x: (1/n) \sum_{k=1}^{n} (T^k f)(x) \to \int f d\phi$ for all $f \in C(S)\} = 1$. However, $\lim (1/n) \sum_{k=1}^{n} (T^k f)(x) = \hat{f}(x) = \int f d\phi_x$ (1.3(a)) so $\{x: (1/n) \sum_{k=1}^{n} (T^k f)(x) \to \int f d\phi$ for all $f \in C(S)\} = \{x: \phi_x = \phi\}$ belongs to $\mathscr{D}$, in fact to $\mathscr{E}$ since it has ϕ-measure 1. Thus $\phi = \phi_E$ for some $E \in \mathscr{E}$. The converse follows immediately from 2.2(d), and the theorem is proved.

2.10. Theorem. ϕ is an extremal point of α if and only if ϕ is indecomposable.

Proof. Suppose first that $\phi \in \alpha$ is not indecomposable. Then there is a stochastically closed set F with $0 < \phi(F) < 1$. Let ϕ_1 and ϕ_2 be defined for $E \in \Sigma$ by $\phi_1(E) = \phi(EF)/\phi(F)$, $\phi_2(E) = \phi(EF^c)/\phi(F^c)$. Both ϕ_1 and ϕ_2 are invariant distributions. But $\phi = \phi(F)\phi_1 + \phi(F^c)\phi_2$; so ϕ is not an extremal point of α.

Suppose $\phi \in \alpha$ is indecomposable. If ϕ is not an extremal point of α, there are distinct ϕ_1 and ϕ_2 in α together with positive c_1, c_2 with $c_1 + c_2 = 1$ such that $\phi = c_1\phi_1 + c_2\phi_2$. If ϕ_1 is indecomposable, $\phi_1 = \phi_E$ for some $E \in \mathscr{E}$, by the preceding theorem, thus concentrates unit mass on the ergodic kernel contained in

$\mathscr{E}$. If ϕ_1 is not indecomposable, there is a stochastically closed A for which $0 < \phi_1(A) < 1$; in either case, there is a stochastically closed A for which $0 < \phi_1(A) \leqq 1$. Since $0 < c_1 < 1$ it follows that $0 < \phi(A) < 1$, contradicting the indecomposability of ϕ. Thus ϕ is an extremal point of α. This completes the proof (the argument is essentially that given in [2]).

3. In this section we study the behavior of the sample paths of the processes $\{X_n, n = 1, 2, \cdots\}$ and the processes $\{f(X_n), n = 1, 2, \ldots\}$ derived from them by composition with f's in $C(S)$. If μ is a probability distribution in $C^*(S)$, it follows from the definition of uniform stability that $\lim (1/n) \sum_{k=1}^{n} T^{*k}\mu$ is a probability distribution, which we will denote by $\bar{\mu}$. (Thus $\bar{\delta}_x = \phi_x$.) $\bar{\mu}$ is invariant for each $\mu \in C^*(S)$. The process with initial distribution $\bar{\mu}$ will be denoted by $\{\bar{X}_n, n = 1, 2, \cdots\}$. Since $\bar{\mu}$ is invariant, $\{\bar{X}_n, n = 1, 2, \cdots\}$ is stationary. We will on occasion write '$\bar{X}$' for '$\bar{X}_1$'.

3.1. LEMMA. *Suppose that $f \in C(S)$ is invariant. Then, for any process $\{X_n, n = 1, 2, \cdots\}$, the process $\{f(X_n), n = 1, 2, \cdots\}$ converges a.s. to a limit distributed as $f(\bar{X})$.*

Proof. Let μ be the initial distribution of $\{X_n, n = 1, 2, \cdots\}$. Using the Markov property of $\{X_n, n = 1, 2, \cdots\}$ and the invariance of f, we see that $E(f(X_n) \mid X_1, \cdots, X_{n-1}) = E(f(X_n) \mid X_{n-1}) = (Tf)(X_{n-1}) = f(X_{n-1})$ a.s.; thus $\{f(X_n), n = 1, 2, \cdots\}$ is a martingale. Since its members are bounded by $\|f\|$, it converges a.s. to a limit Y [14, p. 393]. If $g \in C(S)$, $g(f(X_n)) \to g(Y)$ a.s. Hence $(1/n) \sum_{k=1}^{n} E\{g(f(X_k))\} \to E\{g(Y)\}$. But $(1/n) \sum_{k=1}^{n} E\{g(f(X_k))\}$ $= (1/n) \sum_{k=1}^{n} \int E\{g(f(X_k)) \mid X_1 = x\}\mu(dx) = (1/n) \sum_{k=1}^{n} \int T^{k-1}(g \circ f)(x)\mu(dx)$ $= ((1/n) \sum_{k=1}^{n} T^{k-1} (g \circ f), \mu) = (g \circ f, (1/n) \sum_{k=1}^{n} T^{*k-1}\mu) \to (g \circ f, \bar{\mu})$ $= E\{g(f(\bar{X}))\}$. Thus $E\{g(Y)\} = E\{g(f(\bar{X}))\}$ for all $g \in C(S)$, and it follows that Y and $f(\bar{X})$ have the same distribution. This completes the proof of the lemma.

The following is the strong law of large numbers for uniformly stable in mean transition probability operators. It was discovered by Breiman [1], who proved it in the case of a unique invariant distribution. If the invariant distribution is unique, all the f's are constant, and so all the averages $(1/n) \sum_{j=1}^{n} T^j f$ converge uniformly to constants. But it is the uniformity of the convergence rather than the constancy of the limit which is important in Breiman's proof, and so his proof holds in the case of uniform stability in mean provided one line of additional argument is added. However, we give here a reorganization of Breiman's proof in which certain of Breiman's auxiliary notation is avoided. (This reorganization was suggested by another proof of Breiman's theorem due to Furstenberg [7].)

3.2. THEOREM. *Let $\{X_n, n = 1, 2, \cdots\}$ be any process. Then, for almost all $\omega \in \Omega$, the sequences $(1/n) \sum_{k=1}^{n} f(X_k(\omega))$ and $\bar{f}(X_k(\omega))$ have the same limit for all $f \in C(S)$. This common limit is distributed as $\bar{f}(\bar{X})$.*

Proof. The proof is in three parts.

(a) *Let $x_1, x_2, \cdots$ be a sequence in S for which*

$$(1/n) \sum_{k=1}^{n} \{(Tf)(x_k) - f(x_{k+1})\} \to 0 \text{ as } n \to \infty \text{ for all } f \in C(S).$$

Then

$$\tilde{f}(x_n) - (1/n) \sum_{k=1}^{n} f(x_k) \to 0 \text{ as } n \to \infty \text{ for all } f \in C(S).$$

To prove (a), assume the hypothesis, fix f, then replace f successively by Tf, $T^2f, \cdots$. This yields $(1/n) \sum_{k=1}^{n}\{(T^jf)(x_k) - f(x_{k+1})\} \to 0$ as $n \to \infty$ for each $j = 1, 2, \cdots$, hence $(1/n) \sum_{k=1}^{n}\{((1/m) \sum_{j=1}^{m} T^j f)(x_k) - f(x_{k+1}) \to 0$ as $n \to \infty$ for each $m = 1, 2, \cdots$. Since $(1/m) \sum_{j=1}^{m} T^j f$ converges uniformly to $\tilde{f}$, we have $\tilde{f}(x_n) - (1/n) \sum_{k=1}^{n} f(x_{k+1}) \to 0$ as $n \to \infty$.

(b) *For almost all $\omega \in \Omega$, $(1/n) \sum_{k=1}^{n}\{(Tf)(X_k(\omega)) - f(X_{k+1}(\omega))\} \to 0$ as $n \to \infty$ for each $f \in C(S)$.*

For a fixed $f \in C(S)$ the a.s. convergence to zero of

$$(1/n) \sum_{k=1}^{n} \{(Tf)(X_k) - f(X_{k+1})\}$$

follows from E on p. 387 of [14]. Now let f run through a sequence $f_1, f_2, \cdots$ dense in $C(S)$. For almost all $\omega \in \Omega$, $(1/n) \sum_{k=1}^{n}\{(Tf_i)(X_k(\omega)) - f_i(X_k(\omega))\}$ converges to zero for each $i = 1, 2, \cdots$. It follows from the density of $\{f_i\}$ in $C(S)$ that, for these ω's, $(1/n) \sum_{k=1}^{n}\{(Tf)(X_k(\omega)) - f(X_k(\omega))\}$ converges to zero for each $f \in C(S)$.

(c) *For almost all $\omega \in \Omega, \tilde{f}(X_n(\omega))$ converges as $n \to \infty$ for each $f \in C(S)$.*

This follows from 3.1 upon successive replacement of f by members $f_1, f_2, \cdots$ of a sequence dense in $C(S)$.

The theorem now follows from (a), (b), and (c).

Let $x_1, x_2, \cdots$ be a sequence in S, and A a subset of S. We say that $\{x_n\}$ *approaches* A if $d(x_n, A) \to 0$. Clearly this happens if and only if all the cluster points of $\{x_n\}$ belong to the closure of A. We say that $\{x_n\}$ *approaches A in density* if given any neighborhood V of A, $(1/n) \sum_{k=1}^{n} I_V(x_k) \to 1$ as $n \to \infty$.

3.3. THEOREM. *Let $\{X_n, n = 1, 2, ...\}$ be any process. For almost all $\omega \in \Omega$, the sample path $\{X_n(\omega)\}$ approaches an ergodic set and approaches in density the ergodic kernel contained in that ergodic set.*

Proof. We first show that for almost all ω, there is an $E \in \mathscr{E}$ approached by $\{X_n(\omega)\}$.

(a) *For almost all $\omega \in \Omega$, $\{X_n(\omega)\}$ approaches some $D \in \mathscr{D}$.*

Since $\mathscr{D}$ is a partition of S it suffices to show that for almost all $\omega, \{X_n(\omega)\}$

does not have cluster points in two distinct members of $\mathcal{D}$. Suppose $\{X_n(\omega)\}$ does have cluster points $x \in D$ and $y \in D'$, where D and D' are distinct members of $\mathcal{D}$. By definition of $\mathcal{D}$, there is an $f \in C(S)$ such that $\tilde{f}$ is constantly equal to h on D and to h' on D' where $h \neq h'$. By virtue of the continuity of $\tilde{f}$, $\{\tilde{f}(X_n(\omega))\}$ has as cluster points both h and h'. But if we exclude a P-null set, $\{\tilde{f}(X_n(\omega))\}$ converges for all $f \in C(S)$. It follows that for these ω's, $\{X_n(\omega)\}$ cannot have cluster points in distinct members of $\mathcal{D}$. This proves (a).

(b) *For almost all ω, $\{X_n(\omega)\}$ has at least one cluster point in $\sum \mathscr{E}$.*

Let m be a positive integer. Let $V = \{x: d(x, \sum \mathscr{E}) < 1/m\}$. Let $f \in C(S)$ with range in $[0,1]$ be equal to 0 on V^c and to 1 on $\sum \mathscr{E}$. Then $\tilde{f}(x) = \int f d\phi_x = \int_{\Sigma \mathscr{E}} f d\phi_x = \mu(\sum \mathscr{E}) = 1$ for all $x \in S$. Thus $\tilde{f}(X_n)$ is identically equal to 1, whence $(1/n)\sum_{k=1}^{n} f(X_k) \to 1$ a.s. as $n \to \infty$. It follows that if we exclude a P-null set N_m of ω's, $\{X_n(\omega)\}$ has a cluster point in $\{x: d(x, \sum \mathscr{E}) \leq 1/m\}$. If we exclude the P-null set $N = \bigcup_{m=1}^{\infty} N_m$, $\{X_n(\omega)\}$ has a cluster point in $\{x: d(x, \sum \mathscr{E}) \leq 1/m\}$ for each $m = 1, 2, \cdots$. Since the set of cluster points of a sequence is closed, each such $\{X_n(\omega)\}$ has a cluster point in $\sum \mathscr{E}$. This establishes (b).

That almost all sample paths $\{X_n(\omega)\}$ approach some $E \in \mathscr{E}$ now follows from (a) and (b). We now exclude from Ω the null set of ω's for which either $\{X_n(\omega)\}$ does not approach some $E \in \mathscr{E}$ or for which the assertion of Theorem 3.2 does not hold. Assume that $\{X_n(\omega)\}$ approaches $E \in \mathscr{E}$. Let K be the ergodic kernel contained in E. Let V be an open neighborhood of K. Let $f \in C(S)$ and into $[0,1]$ be equal to zero on V^c and to 1 on K. Then, if $x \in K$, $\tilde{f}(x) = \int f d\phi_x = \int_1 f d\phi_x = \phi_x(K) = 1$. But since $\tilde{f} = 1$ on K, $\tilde{f} = 1$ throughout E. Since $\{X_n(\omega)\}$ approaches E, $f(X_n(\omega)) \to 1$. But then $(1/n)\sum_{k=1}^{n} I_V(X_k(\omega)) \geq (1/n)\sum_{k=1}^{n} f(X_k(\omega)) \to \lim \tilde{f}(X_n(\omega)) = 1$. Thus $\{X_k(\omega)\}$ approaches K in density. This completes the proof of the theorem.

Let Z be the integers with the discrete topology, and let S be the one point compactification $Z \cup \{\infty\}$ of Z. Let $P(n, \{n+1\}) = P(n, \{n-1\}) = 1/2$ for all $n \in Z$, and let $P(\infty, \{\infty\}) = 1$. It is well known that $\phi_x = \delta_\infty$ for all $x \in S$, so $P(x,E)$ is uniformly stable in mean. $\mathscr{E} = \mathscr{D} = \{S\}$, while $\{\infty\}$ is the only ergodic kernel. However, almost all sample paths of any process visit all points in $\mathscr{Z}$ infinitely often [6, p. 288]. This example shows that although sample paths converge to ergodic sets, they need not converge to ergodic kernels. Thus the qualifying phrase "in density" is essential in the statement of Theorem 3.3.

We now examine the set $\sum \mathscr{K} = \bigcup \{K: K \in \mathscr{K}\}$. We denote by $C^+(S)$ the collection of non-negative members of $C(S)$.

3.4. THEOREM. $x \in \sum \mathscr{K}$ *if and only if* $f(x) > 0 \Rightarrow \tilde{f}(x) > 0$ *for all* $f \in C^+(S)$.

Proof. Suppose $x \in K$ for some $K \in \mathscr{K}$. If $f(x) > 0$, then $f(x) \geq \delta > 0$ in some neighborhood V of x. Since $\phi_x(V) > 0$ by 2.8, $\tilde{f}(x) = \int f d\phi_x \geq \delta \phi_x(V) > 0$.

Now assume that $x \in S$ has the property that $f(x) > 0$ implies $\tilde{f}(x) > 0$ for all $f \in C^+(S)$. To begin with, x must belong to some ergodic set. For suppose $x \notin \sum \mathscr{E}$.

Let f be any non-negative continuous function positive at x and zero on the closed set $\Sigma \mathscr{E}$. Then, since $\Sigma \mathscr{E}$ is of maximal probability, $\check{f}(x) = \int f d\phi_x = 0$. Let $\{X_n, n = 1,2,\ldots\}$ be the process starting at x. Let V be any open neighborhood of x. Then almost all sample paths of $\{X_n\}$ return to V with positive relative frequency. To show this, we proceed as follows. Let U be a closed neighborhood of x contained in V. There is an $f \in C(S)$ with range in $[0,1]$ which is 1 on U and zero on V^c. Then $(1/n)\sum_{k=1}^{n} I_V(X_k) \geqq (1/n)\sum_{k=1}^{n} f(X_k) \to Y$ almost surely as $n \to \infty$, where Y is a random variable distributed as $\check{f}(\hat{X})$. We must show that $P[Y > 0] = 1$. Since $f(x) > 0$, we have $\check{f}(x) > 0$. The distribution of $\hat{X}$ is ϕ_x, so $P[Y > 0] = P[\check{f}(\hat{X}) > 0] = \phi_x\{y : \check{f}(y) > 0\} \geqq \phi_x\{y : \check{f}(y) = \check{f}(x)\} \geqq \phi_x(E)$, where E is the ergodic set containing x. But $\phi_x(E) = 1$ by definition of "ergodic set." Since x has a countable neighborhood base it follows that almost all sample paths of $\{X_n\}$ return to all neighborhoods of x with positive relative frequency. We can now conclude that x belongs to some ergodic kernel K. For suppose not. For each $K \in \mathscr{K}$, let Ω_K be the set of ω's in Ω for which $\{X_n(\omega)\}$ approaches K in density. By Theorem 3.3, almost all ω's belong to $\bigcup\{\Omega_K : K \in \mathscr{K}\}$. Now fix K. Since $x \notin K$, there are disjoint neighborhoods V and U of x and K respectively. If $\omega \in \Omega_K$, $\lim (1/n)\sum_{k=1}^{n} I_U(X_k(\omega)) = 1$, so $\lim (1/n)\sum_{k=1}^{n} I_V(X_k(\omega)) = 0$. Letting K range over $\mathscr{K}$, we see that for each $\omega \in \bigcup\{\Omega_K : K \in \mathscr{K}\}$ there is a neighborhood of x to which $\{X(\omega)\}$ returns with zero relative frequency. Contradiction. x must belong to some ergodic kernel. This completes the proof of the theorem.

3.5. COROLLARY. $\Sigma \mathscr{K}$ is a G_δ.

Proof. There is in $C^+(S)$ a sequence $f_1, f_2, \cdots, f_n, \cdots$ such that for each $x \in S$ and open neighborhood U of x there is an f_n with $f_n(x) = 1$ and $f_n = 0$ on U^c. Using the preceding theorem, we then obtain the identity

$$\Sigma \mathscr{K} = \bigcap_{n=1}^{\infty} \{x : f_n(x) > 0 \Rightarrow \check{f}_n(x) > 0\}.$$

This last set, in turn, can be written as $\bigcap_{n=1}^{\infty}(\{x : f_n(x) = 0\} \cup \{x : \check{f}_n(x) > 0\})$. But for each n, $\{x : f_n(x) = 0\}$ is closed, hence a G_δ. Since the class of G_δ's is closed under finite union and countable intersection, it follows that $\Sigma \mathscr{K}$ is a G_δ.

We close this section with an example for which $\Sigma \mathscr{K}$ is a nonclosed G_δ. Let $S = [0,1] \times [0,1]$. Let v be a function on $[0,1]$ to $C^*[0,1]$ satisfying the following conditions: (i) $v(0) = \delta_0$, (ii) for each $x \in (0,1]$, $v(x)$ is a probability measure assigning positive measure to every open subset of $[0,1]$, (iii) v is continuous relative to the usual topology on $[0,1]$ and the weak* topology on $C^*[0,1]$. (It is easy to construct such v's.) For each $x \in [0,1]$, let $L_x = \{(x,y) : 0 \leqq y \leqq 1\}$. For each $x \in [0,1]$ and $E \subset S$, let $E_x = \{y : (x,y) \in E\}$. For each $(x,y) \in S$ and $E \in \Sigma$, let $P((x,y),E) = v(x)(E_x)$. It is easy to see that $P((x,y),E)$ is a transition probability operator and that the corresponding operator T takes $C(S)$ into

itself. It is also clear that $T^n = T$ for all $n = 1, 2, \cdots$; thus T is uniformly stable in mean. For each $(x, y) \in S$, the invariant measure $\phi_{(x,y)}$ associated with (x, y) is simply $P((x, y), \cdot)$. It then follows that $\mathscr{E} = \{L_x : 0 \leqq x \leqq 1\}$, while $\mathscr{K} = \{(0,0)\} \cup \{L_x : 0 < x \leqq 1\}$. Thus $\sum \mathscr{K} = \{(0,0)\} \cup \{(x,y) : 0 < x \leqq 1, 0 \leqq y < 1\}$ is not closed.

4. T is said to be quasi-weakly (-strongly) compact if $\| T^m - K \| < 1$ for some positive integer m and weakly (strongly) compact operator K. For operators in $C(S)$ the two notions coincide. To prove this we need the following facts. The weakly and strongly compact operators on a Banach space B form two-sided ideals in the linear space of all bounded operators on B. If $B = C(S)$, then the product of two weakly compact operators is strongly compact [5, pp. 484, 486, and 494]. It follows that the product of a finite number of operators on $C(S)$ is strongly compact if at least two of them are weakly compact. Now let $T^m = K + V$, where K is weakly compact and $\| V \| < 1$. Then

$$T^{mn} = K_n + KV^{n-1} + VKV^{n-2} + \cdots + V^{n-1}K + V^n,$$

where K_n is strongly compact. Since $\| KV^{n-1} + VKV^{n-2} + \cdots + V^{n-1}K + V^n \| \leqq n \| K \| \| V \|^{n-1} + \| V \|^n$, n may be chosen so that $\| T^{mn} - K_n \| < 1$, and the quasi-strong compactness of T follows.

The results of Yosida and Kakutani [18] imply that if T is a quasi-strongly compact operator with $\| T \| \leqq 1$, the averages $(1/n) \sum_{k=1}^n T^k$ converge in the uniform operator topology [5, p. 711, Corollary 4]. It follows, of course, that T is uniformly stable in mean. Our results are much weaker in all respects than those of Yosida and Kakutani, and this paper adds absolutely nothing to the theory of quasi-strongly compact operators. However, the condition of quasi-compactness excludes all but the least interesting of the Markov operators induced by point transformations and random point transformations. We will now be more explicit about this last concept.

Let N be a fixed positive integer. Let $\phi_1, \phi_2, \cdots, \phi_N$ be continuous maps of S into itself. Let $p_1, p_2, \cdots, P_N$ be continuous function on S into $[0,1]$ such that $\sum_{n=1}^N p_n = 1$. There is a Markov operator which corresponds to the following random point transformation: A particle at the point x at time n is moved to $\phi_1(x)$ with probability $p_i(x)$ at time $t = n + 1$, $i = 1, \cdots, N$. Thus for $x \in S$ and $E \in \Sigma$, $P(x, E) = \sum \{p_i(x) : \phi_i(x) \in E\}$, while, for $f \in C(S)$, $Tf = \sum_{i=1}^N p_i \cdot (f \circ \phi_i)$. The case $N = 1$ has been and still is the object of much study. We refer the interested reader to [15] and [8]. ([15] contains most of our results for the case $N = 1$.) Conditions on the p_i's and ϕ_i's which imply that T is uniformly stable in mean are hard to come by. We shall give three such conditions.

The first is that each p_i be constant and for each x, y in S and $i = 1, \cdots, n$ we have $d(\phi_i(x), \phi_i(y)) \leqq d(x, y)$. Given $f \in C(S)$ it is immediate that $\{T^n f\}$, hence $\{(1/n) \sum_{k=1}^n T^k f\}$, is equicontinuous, so T is uniformly stable in mean.

If $d(\phi_i(x), \phi_i(y)) = d(x, y)$ for all $i = 1, \cdots, N$ and x, y in S it seems appropriate to say that the processes is one of *random isometries*. The set-up is very similar to that in random ergodic theory [13]. In the latter case we have instead of a metric space (S, d) a measure space (S, Σ, μ), and the functions ϕ_i are required to be not isometries but measure preserving transformations: $\mu(\phi_i^{-1}E) = \mu(E)$ for each $i = 1, \cdots, N$ and $E \in \Sigma$. Ulam and von Neumann showed (see [13]) that $(1/n) \sum_{i=1}^n T^i f$ converges both in L_1-norm and almost everywhere for each $f \in L_1(S, \Sigma, \mu)$, whereas in our case we obtain the uniform convergence of the same averages for $f \in C(S)$.

The fact that T arises from a process of random isometries imposes no upper bound on the number of ergodic kernels. (Of course there cannot be more ergodic kernels than points in S, and there are at most $2^{\aleph_0}$ of them.) Simply take $N = 1$, thus forcing p_1 to equal 1, and take for ϕ_1 the identity map of S. Then each point is an ergodic kernel. This is a rather degenerate case. For suppose that there is more than one ergodic kernel. Let K_1 and K_2 be distinct ergodic kernels, and let $\delta = d(K_1, K_2) > 0$. Assume that none of $p_1, \cdots, p_N$ is equal to 0. Let $x \in K_1$ and $y \in K_2$. Then, given any pair of finite sequences $i_1, \cdots, i_m$ and $j_1, \cdots, j_n$ from $1, \cdots, N$, it follows from the stochastically closed nature of K_1 and K_2 that

$$d(\phi_{i_m}(\phi_{i_{m-1}} \cdots \phi_{i_1}(x) \cdots), \phi_{j_n}(\phi_{j_{n-1}} \cdots \phi_{j_1}(y) \cdots)) \geqq \delta.$$

Therefore, if for any x, y in S and $\delta > 0$ there are sequences $\{i_k\}$ and $\{j_l\}$ with $d(\phi_{i_m}(\phi_{i_{m-1}} \cdots \phi_{i_1}(x) \cdots), \phi_{j_n}(\phi_{j_{n-1}} \cdots \phi_{j_1}(y) \cdots)) < \delta$, there is exactly one ergodic kernel.

We now give another condition on the ϕ_i's and p_i's which implies uniform stability in mean. We assume that there is an $r < 1$ and a $k < \infty$ such that $d(\phi_i(x), \phi_i(y)) \leqq rd(x, y)$ and $|p_i(x) - p_i(y)| \leqq k d(x, y)$ for all x, y in S and $i = 1, \cdots, N$. We then say that the process is one of random contractions. The corresponding operators T were studied in detail by Doeblin and Fortet [3]. We summarize their main result. Let B be the subcollection of (complex) $C(S)$ consisting of functions f for which

$$m(f) = \sup_{x, y \in S; x \neq y} |f(x) - f(y)| / d(x, y)$$

is finite. B is a linear subspace of $C(S)$, and it is dense in $C(S)$ relative to the topology induced by the uniform norm on $C(S)$, which we now denote by $\| \cdot \|_\infty$. The equation $\|f\| = \|f\|_\infty + m(f)$ defines a new norm on B relative to which B is a Banach space. It is easy to verify that, under the assumptions just made, T maps B into itself, and the restriction of T to B is an operator of norm 1 relative to the norm $\| \|$. Doeblin and Fortet show that, as an operator on B, T is quasi-strongly compact (see also [10] and [11]). This implies that $(1/n) \sum_{k=1}^n T^k f$ converges in $\| \|$-norm for each $f \in B$, hence in $\| \cdot \|_\infty$-norm for each $f \in B$, hence

in $\| \cdot \|_\infty$ norm for each $f \in C(S)$. Thus T (as an operator in $C(S)$) is uniformly stable in mean.

It is not necessary to use the full strength of the analysis of Doeblin and Fortet to show that T is uniformly stable in mean. A simple computation shows that for any $f \in B$ we have $m(Tf) \leq rm(f) + Nk\|f\|_\infty$ (see [3, p. 143]). It follows by induction that $m(T^n f) \leq r^n m(f) + Nk((1-r^n)/(1-r))\|f\|_\infty$, for each $n = 1, 2, \cdots$ and $f \in B$. Fixing f, we see that the functions $T^n f$ satisfy a Lipschitz condition uniformly in n. So, therefore, do the functions $(1/n)\sum_{k=1}^n T^k f$. Thus $\{(1/n)\sum_{k=1}^n T^k f\}$ is equicontinuous, hence compact, relative to the uniform topology in $C(S)$. It then follows (see the proof of Theorem 1.1) that $(1/n)\sum_{k=1}^n Tf^k$ converges uniformly. This is true for each $f \in B$, but since B is dense in $C(S)$ follows that it holds for each $f \in C(S)$. Thus T is uniformly stable.

We now show that there can be only a finite number of ergodic kernels in the random contraction case. Suppose not. Then there is a sequence $\{K_j\}$ of ergodic kernels such that $0 < d(K_j, K_{j+1}) \to 0$ as $n \to \infty$. Let $f_j(x) = [1/d(K_j, K_{j+1})] \inf(d(x, K_j), d(K_j, K_{j+1}))$. It is easily verified that f_j is into $[0,1]$, is 0 on K_j and 1 on K_{j+1}, and that $f_j \in B$ (in fact, $m(f_j) = 1/d(K_j, K_{j+1})$). Since K_j and K_{j+1} are ergodic kernels we must have $\hat{f}_j$ also equal to 0 on K_j and 1 on K_{j+1}. Since $(1/n)\sum_{k=1}^n T^k f_j$ converges in the norm of B, $\hat{f}_j \in B$. Since $\|f_j\|_\infty = 1$, we see that $m(\hat{f}_j) \geq 1/d(K_j, K_{j+1}) \to \infty$ as $j \to \infty$. But $m(\hat{f}_j) = m(T^n \hat{f}_j) \leq r^n m(\hat{f}_j) + Nk((1-r^n/(1-r))\|f\|_\infty$ for each $n = 1, 2, \cdots$, and this implies that $m(\hat{f}_j) \leq Nk/1-r$ for each j. Thus the assumption that there are infinitely many ergodic kernels leads to a contradiction.

Isaac [11] has observed that the Doeblin-Fortet analysis of the operator T continues to hold provided that $\sum_{i=1}^N p_i(x)d(\phi_i(x), \phi_i(y)) \leq rd(x,y)$ for some $r < 1$ and all x, y in S (where it is assumed as before that $|p_i(y) - p_i(y)| \leq kd(x,y)$ for some $k < \infty$ and all $x, y \in S$).

We now specialize S in order to discuss chains of infinite order (see [9] for references to the literature). Let N be an arbitrary positive integer. Let $I = \{1, \cdots, N\}$ and let $S = \prod_{i=0}^\infty I_i$, where $I_i = I$ for each $i = 0, 1, \cdots$. Give S the product topology induced by the discrete topology on I. Then S is compact and metrizable; in fact the product topology is that of the metric $d(x,y) = \sum_{n=0}^\infty |x_n - y_n|/2^{n+1}$, where $x = (x_0, x_1, \cdots)$ and $y = (y_0, y_1, \cdots)$ are in S. For each $i = 1, \cdots, N$, ϕ_i is the function which sends $(i_0, i_1, i_2, \cdots)$ into $(i, i_0, i_1, \cdots)$. Then $d(\phi_i(x), \phi_i(y)) \leq (1/2)d(x,y)$. Now let $p_1, \cdots, p_N$ be real-valued continuous functions with $p_i \geq 0$ and $p_1 + \cdots + p_N = 1$. The p_i's may be thought of as specifying the laws of evolution of a stationary discrete-parameter random system with N states. If n is any positive integer, and $x = (x_0, x_1, \cdots)$ is in S, we interpret $p_i(x)$ as the conditional probability that the system is in state i at time $t = n + 1$ given that it is in state x_0 at time $t = n, x_1$ at time $t = n - 1, \cdots, x_k$ at time $t = n - k, \cdots$. Thus each $x \in S$ represents a possible past history of the system at a given time, and $p_i(x)$ is the probability that a system with past history x jumps to state i.

The successive past histories constitute a Markov chain: If the system has past history x at time $t = n$, it will have past history $\phi_i(x)$ at time $t = n + 1$ with probability $p_i(x)$.

If $x = (x_0, x_1, \cdots)$ and $y = (y_0, y_1, \cdots)$ are in S, $(x = y)_m$ means that $x_i = y_i$ for each $i = 1, \cdots, m$. The sets $U_m = \{(x, y) : x, y \in S, (x = y)_m\}$ for $m = 0, 1, \cdots$ constitute a base for a uniform structure $\mathcal{U}$ on S which generates the product topology on S. For each $m = 0, 1, \cdots$ let $\varepsilon_m = \sup_{1 \leq i \leq N} \sup_{(x=y)_m} |p_i(x) - (p_i y)|$. To say that $p_1, \cdots, p_N$ are continuous is to say that $\varepsilon_m \to 0$ as $m \to \infty$. One way of obtaining uniform stability in mean is to demand that ε_m approach zero at a certain rate. For instance, if one requires that $\varepsilon_m \leq k/2^m$, one obtains $|p_i(x) - p_i(y)| \leq k d(x, y)$, and so the Doeblin-Fortet analysis mentioned above applies to T. A weaker condition, also introduced in [3,] is that $\sum \varepsilon_m < \infty$. T is uniformly stable in mean in this case also. (This follows easily from the remarks and computations on p. 136 of [3].) Recently C.T. Ionescu Tulcea has sharpened and generalized the results obtained by Doeblin and Fortet in the case $\sum \varepsilon_m < \infty$[9].

We now let $S = G$, a compact separable group. Let μ be any probability measure on the Borel sets of G. Let H be the support of G (the smallest closed subset of G with μ-measure 1). Let $[\mu]$ denote the smallest closed set containing H which is closed under multiplication in G. It is easy to see that $[\mu]$ is a closed subgroup of G. Let λ denote Haar measure normalized so that $\lambda(G) = 1$. The transition probability operator we considered is the one which sends $g \in G$ into $g'g$, where g' is an element selected at random from G in accordance with the probability measure μ. Such a process is called one of random (left) multiplication. The corresponding operators T and T^* are given as follows:

$$(Tf)(g) = \int f(g'g) d\mu(g'), \quad f \in C(G), \quad g \in G,$$

$$T^* v = \mu * v, \quad v \in C^*(G),$$

where $\mu * v$ is the convolution of μ and v. It is well known that the topology of a compact separable group is given by a left-invariant metric (see [12] for references). If we endow G with this metric we easily verify that T is uniformly stable in mean. It is well known (see [12]) that if $[\mu] = G$, then λ is the unique invariant distribution. If $[\mu]$ is a proper subgroup, each member of $G/[\mu]$ is an ergodic kernel, the invariant distribution on $[\mu]$ is normalized Haar measure on $[\mu]$, and the invariant distribution on any other member of $G/[\mu]$ is obtained in the natural way from that on $[\mu]$. (None of this is new. In fact, our assumption that G is separable is unnecessary. For references, see the bibliography of [12].)

If not only $(1/n) \sum_{k=1}^{n} T^k f$ but also $\{T^n f\}$ is equicontinuous for each $f \in C(S)$ we say that T is *uniformly stable*. In this case, more is known about the behavior of $\{T^n f\}$ than its uniform Cesaro convergence. For example, in [12] we obtain, under the hypothesis that T is uniformly stable, necessary and sufficient conditions

for the uniform convergence of the sequences $\{T^n f\}$ themselves rather than their Cesaro averages.

The present paper overlaps [16] to some extent. In this work, Rosenblatt obtains an ergodic decomposition equivalent to ours. He characterizes $\sum \mathscr{E}$ as the closure of the union of the supports of the ϕ_x's as x ranges over S. The ergodic kernel contained in a given member E of $\mathscr{E}$ is characterized as the support of ϕ_E. He allows the state space S to be a (possibly nonmetrizable) compact Hausdorff space. This enables him to apply his results to the study of random multiplication on a compact semi-group. This application renders more interesting the question of whether the results of §3 hold when S is not metric. It is natural to suppose that they do, but the condition that $C(S)$ is separable (i.e., that S is metrizable) is used in an essential way in the proof of part (c) of 3.2, and 3.3 in turn requires the full strength of 3.2 for its proof. Metrizability of S is not required to show that for each $f \in C(S)$. the sequences $\hat{f}(X_n(\omega))$ and $(1/n) \sum_{k=1}^n f(X_k(\omega))$ converge to a common limit for almost all $\omega \in \Omega$. If $C(S)$ is not separable, however, it does not immediately follow that for almost all $\omega \in \Omega$ the sequences $\{\hat{f}(X_n(\omega))\}$ and $\{(1/n) \sum_{k=1}^n f(X_k(\omega))\}$ converge to a common limit for all $f \in C(S)$. This last assertion may not even hold when $C(S)$ is not separable. (If S is compact Hausdorff, $C(S)$ is separable if and only if S is metrizable: see [5, p. 340].) This leaves us willing to conjecture but unable to prove that 3.3 holds even when S is nonmetrizable. The motion within an ergodic kernel is studied in [16] under the hypothesis that T is uniformly stable ("equicontinuous" is the term used in [16]. In a future paper we will present further results on the motion within an ergodic kernel.

The referee has called our attention to a recent paper of S. P. Lloyd [19]. From the properties of the mapping $f \to \hat{f}$ of $C(S)$ into itself, Lloyd deduces a number of properties of the ϕ_x's (t_x's in his paper). He also studies the properties of the decomposition $\mathscr{D}$ induced by the equivalence relation $x \sim y$ (recall that $x \sim y$ iff $\phi_x = \phi_y$). He points out that $\mathscr{D}$ is upper semi-continuous, and that $\sum \mathscr{E}$ is closed. Thus there is considerable overlapping between his paper and the first two sections of ours.

BIBLIOGRAPHY

1. L. Breiman, *The strong law of large numbers for a class of Markov chains*, Ann. Math. Statist. **31** (1960), 801–803.

2. ———, *On achieving channel capacity in finite memory channels*, Illinois J. Math. **4** (1960), 246–252.

3. W. Doeblin and R. Fortet, *Sur les chaînes à liaisons complètes*, Bull. Soc. Math. France **65** (1937), 132–148.

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

5. N. Dunford and J. Schwartz, *Linear operators*, Vol. I, Interscience, New York, 1958.

6. W. Feller, *An introduction to probability theory and its applications*, 2nd ed., Wiley, New York, 1957.

7. Harry Furstenberg, *Noncommuting random products*, Trans. Amer. Math. Soc. **108** (1963), 377–428.

8. ———, *Strict ergodicity and transformation of the torus*, Amer. J. Math. **83** (1961), 573–601.

9. C. T. Ionescu Tulcea, *On a class of operators occurring in the theory of chains of infinite order*, Canad. J. Math. **11** (1959), 112–121.

10. C. T. Ionescu Tulcea and G. Marinescu, *Théorie ergodique pour les classes d'opérations non complètement continues*, Ann. of Math. (2) **52** (1950), 140–147.

11. R. Isaac, *Markov processes and unique stationary probability measures*, Pacific J. Math. **12** (1962), 273–286.

12. B. Jamison, *Asymptotic behavior of iterates of continuous functions under a Markov operator*, J. Math. Anal. Appl. **9** (1964), 203–214.

13. S. Kakutani, *Ergodic theory*, Proceedings of the International Congress of Mathematicians (Cambridge, Mass., 1950) Vol. 2, pp. 136–139, Amer. Math. Soc., Providence, R. I., 1952.

14. M. Loève, *Probability theory*, 2nd ed., Van Nostrand, Princeton, N. J., 1960.

15. J. Oxtoby, *Ergodic sets*, Bull. Amer. Math. Soc. **58** (1952), 116–135.

16. M. Rosenblatt, *Equicontinuous Markov operators*, Teor. Verojatnost. i Primenen. **9** (1964), 205–222.

17. K. Yosida, *Simple Markov process with a locally compact phase space*, Math. Japon. **1** (1948), 99–103.

18. K. Yosida and S. Kakutani, *Operator-theoretical treatment of Markoff's process and mean ergodic theorem*, Ann. of Math. (2) **42** (1941), 188–228.

19. S. P. Lloyd, *On certain projections in spaces of continuous functions*, Pacific J. Math. **13** (1963), 171–175.

UNIVERSITY OF MINNESOTA,
      MINNEAPOLIS, MINNESOTA