INVARIANT PROBABILITIES FOR CERTAIN MARKOV PROCESSES

By Lester E. Dubins[1] and David A. Freedman[2]

University of California, Berkeley

1. Introduction. This paper, though self-contained, is concerned with a class of Markov operators closely related to those studied in (Bush and Mosteller, 1953), (Harris, 1952), and (Karlin, 1953).

Throughout this paper, Ω is a set, Γ is a set of mappings of Ω into itself, and P is a probability on Γ. Each P determines a Markov process with Ω for state space and this transition mechanism: when at $\omega \varepsilon \Omega$, choose a $\gamma \varepsilon \Gamma$ according to P, and move to the new state $\gamma(\omega)$. If ω itself is random with distribution μ, then the distribution of the new state is $P\mu$. If $P\mu = \mu$, then μ is *P-invariant.*

Here are some sample results when Ω is compact metric, and Γ is a finite set of uniformly strict, one-to-one contractions of Ω. There is one and only one invariant probability, μ. If P assigns positive mass to each $\gamma \epsilon \Gamma$, then: μ is continuous unless there is a common fixed point for all the $\gamma \varepsilon \Gamma$, in which case μ plainly assigns probability 1 to that point; and the support of μ (that is, the smallest closed set of μ-probability 1) is all of Ω if and only if each point of Ω is in the range of some $\gamma \varepsilon \Gamma$. If m is a probability on Ω, and for all $\gamma \varepsilon \Gamma$, $m\gamma^{-1} \ll m$ (that is, the distribution of γ under m is absolutely continuous with respect to m), then μ is either absolutely continuous or purely singular with respect to m.

In Section 6, Ω is the closed unit interval, and Γ is the set of all linear functions. For this special case, the results are rather complete, and are summarized at the beginning of the section. Some applications are in (Dubins and Freedman, 1966).

The assertions in each section presuppose the hypotheses on Ω, Γ, and P given at the beginning of that section.

2. Each $\gamma \varepsilon \Gamma$ is measurable. Let Ω be a set, $\mathfrak{F}$ be a σ-field of subsets of Ω, and N be the set of all non-negative, finite measures on $\mathfrak{F}$. If N_0 and N_1 are subsets of N, then $N_0 + N_1$ is the set of all $\mu_0 + \mu_1$ for $\mu_0 \varepsilon N_0$ and $\mu_1 \varepsilon N_1$. If $\mu_i \varepsilon N_i$, $\nu_i \varepsilon N_i$, and $\mu_0 + \mu_1 = \nu_0 + \nu_1$ imply $\mu_0 = \nu_0$, then the notation $N_0 \oplus N_1$ is used instead of $N_0 + N_1$. If μ, $\nu \varepsilon N$, then $\mu \leqq \nu$ means $\mu(A) \leqq \nu(A)$ for all $A \varepsilon \mathfrak{F}$. If, for $\sigma \varepsilon N$, $\sigma \leqq \mu$ and $\sigma \leqq \nu$ imply $\sigma = 0$, then $\mu \perp \nu$ (that is, μ is singular with respect to ν). For $M \subset N$, $M^{\perp}$ is the set of all $\nu \varepsilon N$ with $\nu \perp \mu$ for all $\mu \varepsilon M$.

(2.1) Lemma. *$N = M \oplus M^{\perp}$ if and only if $M = M^{\perp\perp}$.*

Proof. Easy. □

(2.2) Lemma. *Let: $M = M^{\perp\perp}$; T be a mapping of N into itself, with $TM \subset M$,*

Received 30 September 1965.

[1] This research was supported in part by the Information Systems branch of ONR under Contract 222(53).

[2] Supported in part by National Science Foundation Grant GP 2593 and by a grant from the Sloan Foundation.

$(T\nu)(\Omega) = \nu(\Omega)$, and $T(\nu + \nu') = T\nu + T\nu'$, for all $\nu, \nu' \varepsilon N$; $\mu \varepsilon N$, $T\mu = \mu$, and $\mu = \mu_M + \mu_\perp$ for $\mu_M \varepsilon M$ and $\mu_\perp \varepsilon M^\perp$. Then $T\mu_M = \mu_M$ and $T\mu_\perp = \mu_\perp$.

PROOF. $\mu = T\mu = T\mu_M + T\mu_\perp$. But $\mu \geqq T\mu_M \varepsilon M$, so $\mu_M \geqq T\mu_M$. Finally, $\mu_M(\Omega) = (T\mu_M)(\Omega)$, so $\mu_M = T\mu_M$. □

Let Γ be a non-empty set of $\mathfrak{F}$-measurable mappings of Ω into itself. Let Γ be endowed with a σ-field Σ such that $(\gamma, \omega) \rightarrow \gamma(\omega)$ is measurable from $(\Gamma \times \Omega, \Sigma \times \mathfrak{F})$ to $(\Omega, \mathfrak{F})$. Let P be a probability on Σ.

(2.3) DEFINITION. For $\gamma \varepsilon \Gamma$ and $\mu \varepsilon N$, let $(\mu\gamma^{-1})(A) = \mu(\gamma^{-1}A)$ and $(P\mu)(A) = \int_\Gamma \mu(\gamma^{-1}A)P(d\gamma)$ for all $A \varepsilon \mathfrak{F}$. Plainly, $\mu\gamma^{-1}$ and $P\mu$ are in N. If $P\mu = \mu$, then μ is *P-invariant.*

(2.4) REMARK. For $m \varepsilon N$ and $\gamma \varepsilon \Gamma$, these two conditions are equivalent:

(i) $\nu \ll m$ implies $\nu\gamma^{-1} \ll m$;

(ii) $m\gamma^{-1} \ll m$.

The proof of the following theorem was suggested by (Karlin, 1953, Theorem 40).

(2.5) THEOREM. *Let $m \varepsilon N$. Suppose: either, for all $\gamma \varepsilon \Gamma$, $\nu \ll m$ implies $\nu\gamma^{-1} \ll m$, or, Γ is countable and for all $\gamma \varepsilon \Gamma$, $\nu \perp m$ implies $\nu\gamma^{-1} \perp m$. If $\mu \varepsilon N$ is P-invariant, so is that part of μ which is absolutely continuous with respect to m, as well as that part of μ which is singular with respect to m. In particular, if there is only one P-invariant probability, it must be either purely singular or absolutely continuous with respect to m.*

PROOF. Apply (2.2). □

To avoid trivial complications, suppose each one-point subset $\{\omega\}$ of Ω is in $\mathfrak{F}$.

(2.6) DEFINITION. $\nu \varepsilon N$ is *continuous* if $\nu\{\omega\} = 0$ for each $\omega \varepsilon \Omega$, and is *discrete* if $\sum_{\omega\varepsilon\Omega} \nu\{\omega\} = \nu(\Omega)$.

(2.7) THEOREM. *Suppose: either, for each $\gamma \varepsilon \Gamma$ and $\omega \varepsilon \Omega$, $\gamma^{-1}\{\omega\}$ is a countable set, or, Γ is countable. If μ is P-invariant, so is the continuous part of μ, as well as its discrete part. In particular, if there is only one P-invariant probability, it must be either continuous or discrete.*

PROOF. Use (2.2). □

(2.8) DEFINITION. $A \subset \Omega$ is *P-invariant* if $\gamma A \subset A$ for P-almost all γ.

This theorem supplements (2.7):

(2.9) THEOREM. *Suppose all $\gamma \varepsilon \Gamma$ are one-to-one, $\mu \varepsilon N$ is discrete and P-invariant. Then $\mu = \sum_i \mu_i$, where each $\mu_i \varepsilon N$ is uniform on a finite invariant subset F_i of Ω.*

PROOF. Let $s = \max\{\mu\{\omega\}: \omega \varepsilon \Omega\}$ and $S = \{\omega: \omega \varepsilon \Omega \text{ and } \mu\{\omega\} = s\}$. Then S is finite. Once it is seen that S is also P-invariant, the theorem follows. If $\omega \varepsilon S$, then $s = \mu\{\omega\} = \int_\Gamma \mu\gamma^{-1}\{\omega\}P(d\gamma)$. Since $\gamma^{-1}\{\omega\}$ is either empty or a one-point set, $\mu\gamma^{-1}\{\omega\} \leqq s$. Therefore, P assigns measure 1 to the set G of all $\gamma \in \Gamma$ such that, for each $\omega \varepsilon S$, $\gamma^{-1}\{\omega\}$ is a one-point subset of S. For $\gamma \varepsilon G$, $\gamma S \subset S$, because γ is one-to-one and S is finite. □

(2.10) THEOREM. *Suppose all $\gamma \varepsilon \Gamma$ are one-to-one, and there is a unique P-invariant probability μ. Then either μ is continuous, or there is a unique finite P-invariant set on which μ is uniform.*

PROOF. (2.7) and (2.9) apply. □

(2.11) REMARK. (2.2) and its proof apply to finitely additive measures on fields. Therefore, results like the ones of this section hold for finitely additive processes.

3. Each $\gamma \varepsilon \Gamma$ is continuous. Let Ω be a compact metric space, Γ be the set of continuous mappings of Ω into itself in the topology of uniform convergence, and P be a probability on Γ.

(3.1) LEMMA. *If A is a non-empty compact and P-invariant subset of Ω, there is a P-invariant probability μ with $\mu(A) = 1$.*

PROOF. Apply the Markov-Kakutani fixed-point theorem (Dunford and Schwartz, 1958, p. 456). □

(3.2) LEMMA. *If γ is in the support of P, and ω is in the support of μ, then $\gamma(\omega)$ is in the support of $P\mu$.*

PROOF. $(\gamma, \omega) \to \gamma(\omega)$ is continuous. □

(3.3) THEOREM. *If there is a unique P-invariant probability μ, the support of μ is the smallest non-empty compact P-invariant subset of Ω.*

PROOF. Clear from (3.1) and (3.2). □

If K is a subset of Γ, and A is a subset of Ω, then A is *K-invariant* if $\omega \varepsilon A$ and $\gamma \varepsilon K$ implies $\gamma(\omega) \varepsilon A$. Let K be the support of P. Theorem (3.3) may be restated: If there is a unique P-invariant probability μ, the support of μ is the smallest non-empty compact K-invariant subset of Ω. In particular, the support of μ depends only on K and is non-decreasing with K.

Let $\pi\Gamma$ (respectively, $\pi\Omega$) be the set of probabilities on Γ (respectively, Ω). Endow these sets with the weak* topologies ($P_n \varepsilon \pi\Gamma$ converges to $P \varepsilon \pi\Gamma$ if and only if $\int f\, dP_n$ converges to $\int f\, dP$ for each bounded, continuous f on Γ). Of course, $\pi\Gamma$ is complete, separable, and metric, because Γ is. Let $2^{\pi\Omega}$ be the set of non-empty, closed subsets of $\pi\Omega$, in the usual (Hausdorff, 1957, Section 28) topology. For $P \varepsilon \pi\Gamma$, let $\Sigma(P) \varepsilon 2^{\pi\Omega}$ be the set of P-invariant $\mu \varepsilon \pi\Omega$.

(3.4) THEOREM. *Σ is upper semi-continuous from $\pi\Gamma$ to $2^{\pi\Omega}$, in the sense of* (Kuratowski, 1932).

PROOF. $(P, \mu) \to P\mu$ *is continuous from* $\pi\Gamma \times \pi\Omega$ *to* $\pi\Omega$. □

(3.5) COROLLARY. *Σ is continuous at all P for which there is only one P-invariant μ.*

4. Each $\gamma \varepsilon \Gamma$ is a contraction.

(4.1) DEFINITION. Let (Ω, ρ) be a metric space, and γ be a mapping of Ω into itself. If $\rho(\gamma x, \gamma y) \leqq \rho(x, y)$ for all $x, y \varepsilon \Omega$, then γ is a *contraction* of Ω. If $\rho(\gamma x, \gamma y) < \rho(x, y)$ for all $x, y \varepsilon \Omega$ with $x \neq y$, then γ is *strict*. If there is a $c < 1$ such that $\rho(\gamma x, \gamma y) \leqq c\rho(x, y)$ for all $x, y \varepsilon \Omega$, then γ is *uniformly strict*.

(4.2) ASSUMPTION. Throughout this section, Ω is a compact metric space, Γ is the set of contractions of Ω, and P is a probability on Γ whose support contains a strict contraction.

The distance from γ_1 to γ_2 in Γ is $\max\{\rho(\gamma_1\omega, \gamma_2\omega) : \omega \varepsilon \Omega\}$.

(4.3) LEMMA. *Let γ be a strict contraction of Ω. Then $\bigcap_{j=1}^{\infty} \gamma^j\Omega$ is a single point.*

Let $\gamma_1 \cdots \gamma_n$ be contractions of Ω, of which k consecutive ones are within ϵ of γ. Then the diameter of $\gamma_1\gamma_2 \cdots \gamma_n\Omega$ does not exceed the diameter of $\gamma^k\Omega$ by more than $2k\epsilon$.

PROOF. Easy. □

(4.4) THEOREM. *There is one and only one P-invariant probability μ_P. For each probability ν on Ω, $P^n\nu$ converges to μ_P in the weak* topology.*

PROOF. Let $\gamma_1, \gamma_2, \cdots$ be independent random elements of Γ, with common distribution P. If ω is a random element of Ω, independent of $\gamma_1, \gamma_2, \cdots$, and ω has distribution ν, then $\gamma_1\gamma_2 \cdots \gamma_n\omega$ has distribution $P^n\nu$. However, $\gamma_1\gamma_2 \cdots \gamma_n\Omega$ shrinks almost surely to a random point in Ω, by (4.3). □

(4.5) THEOREM. *If P-almost all γ are one-to-one, then μ_P is continuous unless all γ in the support of P have a common fixed point, in which case μ_P assigns probability 1 to that point.*

PROOF. Use (4.4), (2.10), and (4.3). □

REMARK. (4.3) to (4.5) continue to hold (with essentially the same proof) if Γ is any compact semigroup of continuous mappings of Ω into itself, provided there is a γ in the support of P for which $\bigcap_{n=1}^{\infty} \gamma^n\Omega$ is a single point.

(4.6) QUESTION. If P-almost all γ are finite (respectively, countable) to one, then the support of μ_P is either finite (respectively, countable) or perfect. For the set of limit points (respectively, points of condensation) of the support is P-invariant; (3.3) applies. Moreover, μ_P is either discrete or continuous, by (2.7). Can μ_P be discrete, yet have perfect support? If $\Omega = [0, 1]$ and the support of P is a finite number of uniformly strict two-to-one contractions?

(4.7) DEFINITION. Let $\Gamma(P)$ be the smallest compact sub-semigroup of Γ having P-measure 1.

(4.8) REMARK. From the proof of (4.4), P^n converges in the weak* topology to a probability $\hat{P}$ on $\Gamma(P)$, whose support is the set of constant mappings in $\Gamma(P)$. The projection of $\hat{P}$ on Ω (by assigning to each constant mapping its value) is μ_P. Therefore, the support of μ_P is all of Ω if and only if $\Gamma(P)$ contains all constant mappings. For closely related material, see (Grenander, 1963, Chap. 2).

(4.9) THEOREM. *Let A and B be disjoint sets whose union is Ω. Suppose each point of A is in the range of a strict contraction in $\Gamma(P)$, and each point of B is in the support of μ_P. Then the support of μ_P is all of Ω.*

PROOF. Suppose that the support S of μ_P were a proper subset of Ω. Let α be a point of Ω whose distance from S, namely $\rho(\alpha, S)$, is maximal. Plainly, $\alpha \in A$. Let $\gamma \in \Gamma(P)$ be strict, and $\alpha = \gamma\alpha'$, with $\alpha' \in \Omega$. Let β' be the point of S closest to α'. By (3.3), $\gamma(\beta') \in S$. But $\rho(\alpha, \gamma\beta') < \rho(\alpha', \beta') \leqq \rho(\alpha, S)$, a contradiction. □

(4.10) COROLLARY. *Suppose each γ in the support K of P is a strict contracton. Then these two conditions are equivalent:*

(i) *the support of μ_P is all of Ω;*

(ii) $\bigcup\{\gamma\Omega: \gamma \in K\} = \Omega$.

PROOF. It is clear that (i) $\rightarrow$ (ii). By (4.9), (ii) $\rightarrow$ (i). □

If K is a subset of Γ, and A is a subset of Ω, let K^n be the set of $\gamma_1\gamma_2 \cdots \gamma_n$ with all $\gamma_i \in K$, and let KA be the set of $\gamma(\omega)$ with $\gamma \in K$ and $\omega \in A$.

Here is a generalization of (4.10).

(4.11) COROLLARY. *Suppose each γ in the support K of P is a strict contraction. Then the support of μ_P is $\bigcap_{n=1}^{\infty} K^n\Omega$.*

PROOF. Plainly, $\Omega_\infty = \bigcap_{n=1}^{\infty} K^n\Omega$ is compact and $\Omega_\infty = K\Omega_\infty$. Apply (4.10), with Ω_∞ for Ω. □

If, for example, $\Omega = [0, 1]$, and the support of P is finite but contains a contraction which is not strict, we do not know how to decide whether the support of μ_P is all of Ω.

The following simple fact will be used in Section 6.

(4.12) FACT. Unless all γ in the support of P have a common fixed point, there is a positive integer n and disjoint open spheres U and V such that P^n assigns positive measure both to $\{\gamma : \gamma \in \Gamma$ and $\gamma\Omega \subset U\}$ and to $\{\gamma : \gamma \in \Gamma$ and $\gamma\Omega \subset V\}$.

5. Each $\gamma \in \Gamma$ is continuous and monotone. Throughout this section, Γ is the set of continuous, monotone mappings of the closed unit interval Ω into itself. In order to state (5.4), the first theorem of this section, some notation is needed. Δ is the set of distribution functions on Ω. Each $F \in \Delta$ corresponds to one and only one probability $|F|$ on Ω, with $F(x) = |F|[0, x]$ for all $x \in \Omega$. For a bounded, real-valued function f on Ω, $\|f\| = \sup\{|f(x)| : x \in \Omega\}$. The distance from γ_1 to γ_2 in Γ is $\|\gamma_1 - \gamma_2\|$, and the distance from F_1 to F_2 in Δ is $\|F_1 - F_2\|$. Of course, Γ is a complete, separable, metric semigroup under composition, and Δ is complete. (The topology induced on Δ by this metric is stronger than weak*.) To each $\gamma \in \Gamma$ there corresponds a mapping γ^* of Δ into itself, with $|\gamma^* F| = |F|\gamma^{-1}$. Let P be a probability on Γ. The mapping P^* of Δ into itself is determined by: $|P^*F| = P|F|$. Of course,

$$P^*F = \int_\Gamma (\gamma^*F)P(d\gamma). \tag{5.1}$$

(5.2) DEFINITION. For each subset A of Ω, let RA be the set of $\gamma \in \Gamma$ with $\gamma\Omega \subset A$.

(5.3) DEFINITION. A probability on Γ *splits* if there is an $x \in \Omega$ with $R[0, x]$ and $R[x, 1]$ both of positive probability.

(5.4) THEOREM. *If P splits, then P^* is a uniformly strict contraction of Δ.*

PROOF. Choose x to satisfy (5.3). If P assigns positive probability to the constant function with value x, the result is clear. Otherwise, delete this function from $R[0, x]$ and from $R[x, 1]$, leaving R_0 and R_1 respectively. Let R_2 be the complement of $R_0 \cup R_1$ in Γ. Let F and G be in Δ. By (5.1), $P^*F - P^*G = I_0 + I_1 + I_2$, where

$$I_j = \int_{R_j} (\gamma^*F - \gamma^*G)P(d\gamma).$$

Because γ^* is a contraction of Δ, $\|I_j\| \leqq P(R_j)\|F - G\|$. Because I_0 vanishes on $[x, 1]$, and I_1 vanishes on $[0, x)$, $\|I_0 + I_1\| \leqq \max\{\|I_0\|, \|I_1\|\}$. So

$$\|P^*F - P^*G\| \leqq (\max\{P(R_0), P(R_1)\} + P(R_2))\|F - G\|. \qquad □$$

(5.5) COROLLARY. *If some convolution power of P splits, there is one and only one invariant probability μ; and, for each probability ν on Ω, the distribution function of $P^n\nu$ converges uniformly to the distribution function of μ.*

(5.6) LEMMA. *Suppose that for each $x \varepsilon \Omega$, P assigns measure 0 to $\{\gamma: \gamma \varepsilon \Gamma$ and $\gamma^{-1}\{x\}$ is uncountable$\}$. Then for each continuous probability ν on Ω, $P\nu$ is continuous.*

PROOF. Clear. □

(5.7) COROLLARY. *If some convolution power of P splits, and for each $x \varepsilon \Omega$, P assigns measure 0 to $\{\gamma: \gamma \varepsilon \Gamma$ and $\gamma^{-1}\{x\}$ is uncountable$\}$, then the unique P-invariant probability is continuous.*

PROOF. Apply (5.5) and (5.6). □

Even when P does not split, P^* may be uniformly strict:

(5.8) EXAMPLE. Let P be the image of Lebesgue measure on Ω under the mapping:

$$b \rightarrow \langle \tfrac{1}{2}, b \rangle \quad \text{for} \quad 0 \leqq b \leqq \tfrac{1}{2},$$

$$b \rightarrow \langle -\tfrac{1}{2}, b \rangle \quad \text{for} \quad \tfrac{1}{2} < b \leqq 1,$$

where $\langle a, b \rangle$ is the linear function sending x to $ax + b$.

DISCUSSION OF (5.8). Plainly, P does not split. Let F, $G \varepsilon \Delta$, $y \varepsilon \Omega$, $H = F - G$. Then

$$(5.9) \qquad (P^*F)(y) - (P^*G)(y) = \tfrac{1}{2}\textstyle\int_J H(x)\, dx - \tfrac{1}{2}\int_K H(x)\, dx,$$

where J and K are sub-intervals of Ω that vary with y. Clearly, the absolute value of the right side of (5.9) does not exceed $\frac{1}{2}\|H\|$, so P^* contracts distances by a factor of $\frac{1}{2}$ or less. □

In some problems, P assigns probability 1 to the set of linear functions with non-negative slope (Dubins and Freedman, 1966, Section 9). Under this condition, if P^* is a uniformly strict contraction, then P splits. It takes only a little extra effort to prove the more general Theorem (5.10). To state (5.10), let Γ^+ (respectively, Γ^-) be the set of $\gamma \varepsilon \Gamma$ with $\gamma(1) - \gamma(0)$ positive (respectively, negative). Say P is *essentially of one sign* if there is a countable subset K of Γ, such that either $P(\Gamma^+ \setminus K) = 0$ or $P(\Gamma^- \setminus K) = 0$, where $A \setminus B$ is the set of points in A but not in B.

(5.10) THEOREM. *Suppose P is essentially of one sign. Then P^* is a uniformly strict contraction of Δ if and only if P splits.*

PROOF. The "if" part is contained in (5.4). For "only if", suppose that P does not split. The ideas of the proof that P^* is not uniformly strict are brought out more clearly by first considering three special cases.

Case 1. P assigns probability 1 to a finite subset K of $\Gamma^+ \cup \Gamma^-$.

Plainly, there is a non-empty open interval J, with $J \subset \gamma\Omega$ for all $\gamma \varepsilon K$. So, there is a $y \varepsilon J$, such that: for $\gamma \varepsilon K$, $\gamma^{-1}\{y\}$ is a onepoint subset of $(0, 1)$, say $\{\gamma^{-1} y\}$, and for $\alpha \varepsilon K \cap \Gamma^+$, $\beta \varepsilon K \cap \Gamma^-$, $\alpha^{-1}y \neq \beta^{-1}y$. Then there are F, $G \varepsilon \Delta$, with $F \neq G$, and

$$F(\gamma^{-1}y) - G(\gamma^{-1}y) = \|F - G\| \quad \text{for all} \quad \gamma \varepsilon K \cap \Gamma^+,$$
$$= -\|F - G\| \quad \text{for all} \quad \gamma \varepsilon K \cap \Gamma^-.$$

Clearly, $(P^*F)(y) - (P^*G)(y) = \|F - G\|$, completing the proof in Case 1.

Case 2. P assigns probability 1 to a countable set $K \subset \Gamma^+ \cup \Gamma^1$.

For any $\epsilon > 0$, there is a finite subset K_ϵ of K, with $P(K_\epsilon) \geqq 1 - \epsilon$. The construction of Case 1, with K_ϵ for K, yields $F, G \varepsilon \Delta$, with $F \neq G$, and

$$\|P^*F - P^*G\| \geqq (1 - 2\epsilon)\|F - G\|,$$

completing the proof in Case 2.

(5.11) LEMMA. *For any probability Q on Γ which does not split, and any $\epsilon > 0$, there is a non-empty open interval J, such that for each $y \varepsilon J$, the Q-probability of*

$$\{\gamma : \gamma \varepsilon \Gamma \textit{ and } \gamma^{-1}\{y\} \textit{ is a non-empty subset of } (0, 1)\}$$

exceeds $1 - \epsilon$.

PROOF. For $x \varepsilon \Omega$, let $L(x) = Q\{\gamma : \gamma\Omega \subset [0, x]\}$ and $U(x) = Q\{\gamma : \gamma\Omega \subset [x, 1]\}$. Clearly, L is non-decreasing, $L(0) = 0$, U is non-increasing, and $U(1) = 0$. Let

$$L^* = \sup\{x : L(x) = 0\} \text{ and } U_* = \inf\{x : U(x) = 0\}.$$

Because Q does not split, for each $x \varepsilon \Omega$, either $U(x) = 0$ or $L(x) = 0$. Consequently, $U_* \leqq L^*$. If $U_* < L^*$, let $J = (U_*, L^*)$, and (5.11) holds for any $\epsilon > 0$. If $U_* = L^* = z$, then for some open interval J of y's with end-point z, $U(y) + L(y) \leqq \epsilon$, because U is continuous from the left and L from the right. This completes the proof of (5.11).

Case 3. $P(\Gamma^+) = 1$.

Let $\epsilon > 0$. Use (5.11) to find $y \varepsilon \Omega$ and $\delta > 0$ for which the P-probability of

$$\{\gamma : \gamma \varepsilon \Gamma^+ \text{ and } \gamma^{-1}\{y\} \text{ is a non-empty subset of } (\delta, 1 - \delta)\}$$

exceeds $1 - \epsilon$. Choose $F, G \varepsilon \Delta$ with $F \neq G$ and $F(x) - G(x) = \|F - G\|$ for all x satisfying $\delta < x < 1 - \delta$. Then

$$(P^*F)(y) - (P^*G)(y) \geqq (1 - 2\epsilon)\|F - G\|,$$

completing the proof in Case 3.

The general case.

Suppose there is a countable subset K of Γ^- such that $P(\Gamma^- \backslash K) = 0$, the other situation being quite similar. Let $\epsilon > 0$. Choose K_ϵ, a finite subset of K, so that $P(\Gamma^- \backslash K_\epsilon) \leqq \epsilon$, and $\gamma \varepsilon K_\epsilon$ implies $P\{\gamma\} > 0$. For $\gamma \varepsilon \Gamma^-$, let $f(\gamma)$ be the union of those non-degenerate closed intervals on which γ is constant. Let $e(\gamma)$ be the set of $x \varepsilon \Omega \backslash f(\gamma)$ for which

$$P\{g : g \varepsilon \Gamma^+ \text{ and } g(x) = \gamma(x)\} > 0.$$

Plainly, $e(\gamma)$ is countable. Let $v(\gamma)$ be the image of $e(\gamma) \cup f(\gamma)$ under γ. Plainly, $v(\gamma)$ is countable. Use (5.11) to find a $y \varepsilon \Omega$ satisfying these three conditions:
(5.12) the P-probability of
$\{\gamma : \gamma \varepsilon \Gamma^+$ and $\gamma^{-1}\{y\}$ is a non-empty subset of $(0, 1)\}$
exceeds $P(\Gamma^+) - \epsilon$;
(5.13) $\gamma \varepsilon K_\epsilon$ implies $\gamma^{-1}\{y\}$ is a non-empty subset of $(0, 1)$;
and
(5.14) $y \not\varepsilon \bigcup\{v(\gamma) : \gamma \varepsilon K_\epsilon\}$.
In view of (5.13) and (5.14), if $\gamma \varepsilon K_\epsilon$, then $\gamma^{-1}\{y\} = \{\gamma^{-1}y\}$ with $0 < \gamma^{-1}y < 1$. For each $\delta > 0$, let V_δ be the set of all x satisfying $\delta < x < 1 - \delta$ and $|\gamma^{-1}y - x| \geqq \delta$ for all $\gamma \varepsilon K_\epsilon$. Using (5.12) to (5.14), find $\delta > 0$ such that the P-probability of

$$\{\gamma : \gamma \varepsilon \Gamma^+ \text{ and } \gamma^{-1}\{y\} \text{ is a non-empty subset of } V_\delta\}$$

exceeds $P(\Gamma^+) - \epsilon$.

Choose distinct $F, G \varepsilon \Delta$ for which

$$F(x) - G(x) = \|F - G\| \quad \text{for all} \quad x \varepsilon V_\delta$$

and

$$F(\gamma^{-1}y) - G(\gamma^{-1}y) = -\|F - G\| \quad \text{for all} \quad \gamma \varepsilon K_\epsilon .$$

Since $P(\Gamma^+) + P(\Gamma^-) = 1$,

$$(P^*F)(y) - (P^*G)(y) \geqq (1 - 4\epsilon)\|F - G\|. \square$$

Say $x \varepsilon \Omega$ is *P-fixed* if $\gamma(x) = x$ for P-almost all γ.

(5.15) THEOREM. *Suppose P-almost all γ are strictly increasing. Then the following three conditions are equivalent*:

(i) *some power of P^* is a uniformly strict contraction of Δ*;

(ii) *some convolution power of P splits*;

(iii) *there is no P-fixed point, and there is only one minimal closed non-empty P-invariant interval.*

PROOF. (i) $\rightarrow$ (ii) by (5.10).

(ii) $\rightarrow$ (iii). If x is P-fixed, it is P^n-fixed, so P^n cannot split. There is always at least one minimal closed P-invariant interval. If there are several, they are necessarily disjoint, and each carries a P-invariant probability by (3.1). Apply (5.5).

(iii) $\rightarrow$ (i). Let $[a, b]$ be the minimal closed P-invariant interval. Let a^* be the infimum on n of the P^n-essential infimum of the function $\gamma \rightarrow \gamma(1)$. Since $\gamma(b) \geqq a$ for P^n-almost all γ, $a^* \geqq a$. Because $[a^*, 1]$ is P-invariant, it includes $[a, b]$; so $a^* = a$. Plainly, $a < b$, for $a = b$ would be a P-fixed point. Consequently, there is a positive integer n_1, and a strictly increasing γ_1 in the support of P^{n_1}, such that $\gamma_1(1) < (a + b)/2$. Similarly, there is a positive integer n_0, and a strictly increasing γ_0 in the support of P^{n_0}, such that $\gamma_0(0) > (a + b)/2$.

Plainly, ${\gamma_0}^{n_1}$ and ${\gamma_1}^{n_0}$ are in the support of $P^{n_0+n_1}$, ${\gamma_0}^{n_1}(0) > (a+b)/2$, and ${\gamma_1}^{n_0}(1) < (a+b)/2$; so $P^{n_0+n_1}$ splits. ∎

Suppose P-almost all γ are non-decreasing and strict contractions. Then the unique P-invariant probability has Ω for its support if and only if each $y \varepsilon \Omega$ is in the range of some γ in the support K of P, and there is a strict contraction in K whose fixed point is arbitrarily close to 0, and one whose fixed point is arbitrarily close to 1. With only a little more work, it is possible to establish the more general Theorem (5.17), whose proof uses

(5.16) LEMMA. *Let J be a closed sub-interval of Ω, and let $\gamma \varepsilon \Gamma$ be a contraction of Ω.*

(i) *If γ is non-decreasing and $\gamma(x) = x$ for some $x \varepsilon J$, then $\gamma J \subset J$.*

(ii) *If J is symmetric about $\frac{1}{2}$, γ is non-increasing, and $\gamma(x) = 1 - x$ for some $x \varepsilon J$, then $\gamma J \subset J$.*

PROOF. Easy. ∎

For any subset K of Γ, let K_s be the set of all $\gamma \varepsilon K$ which are strict contractions. Say K *covers* Ω if $\bigcup\{\gamma\Omega : \gamma \varepsilon K\} = \Omega$. Say K *spans* Ω if at least one of the following four conditions holds:

(i) there is an f in K_s whose fixed point is arbitrarily close to 0, and a g in K_s whose fixed point is arbitrarily close to 1;

(ii) there is an f in K_s with fixed point arbitrarily close to 0, and a g in K with $g(0) = 1$;

(iii) there is an f in K with $f(1) = 0$, and a g in K_s with fixed point arbitrarily close to 1;

(iv) there is an f and a g in K with $f(1) = 0$ and $g(0) = 1$, and for an x in Ω arbitrarily close to the two-point set $\{0, 1\}$, there is a γ_x in K_s with $\gamma_x(x) = 1 - x$.

(5.17) THEOREM. *Suppose P-almost all γ are strict contractions. Then the unique P-invariant probability μ has Ω for its support if and only if the support K of P covers Ω and spans Ω.*

PROOF. The "if" part. If (i), (ii), or (iii) holds, then 0 and 1 are in the support of μ, by (3.3). Suppose (iv) holds. There are two cases.

Case 1. There is an $\alpha < 1$ and a $\beta > 0$ such that $f(x) = 1 - x$ for $\alpha \leqq x \leqq 1$ and $g(x) = 1 - x$ for $0 \leqq x \leqq \beta$.

Suppose without loss of generality that $\gamma_n(x_n) = 1 - x_n$ for $\gamma_n \varepsilon K_s$ and $x_n \to 0$. Then for large n, $f\gamma_n \varepsilon (K^2)_s$ has fixed point x_n, and Condition (ii) holds for P^2, implying 0 and 1 are in the support of μ.

Case 2. Either $f(x) < 1 - x$ for all x with $0 \leqq x < 1$, or $g(x) > 1 - x$ for all x with $0 < x \leqq 1$.

Let $\gamma_0 = fg$ and $\gamma_1 = gf$. Plainly, either ${\gamma_0}^n(1) \downarrow 0$ or ${\gamma_1}^n(0) \uparrow 1$, so 0 and 1 are in the support of μ.

Consequently, if (i), (ii), (iii), or (iv) holds, then 0 and 1 are in the support of μ. If $y \varepsilon \Omega$, $\gamma \varepsilon K$, $y = \gamma(0)$ or $\gamma(1)$, then y is in the support of μ, by (3.2). If $y \varepsilon \Omega$ but y is $\gamma(0)$ or $\gamma(1)$ for no $\gamma \varepsilon K$, then y is in the range of some $\gamma \varepsilon K_s$. By (4.9), the support of μ is all of Ω, completing the "if" part of the proof.

The "only if" part. If $y \varepsilon \Omega$ is in the range of no $\gamma \varepsilon K$, then y is not in the support of μ, by (3.3). Therefore, suppose (i) through (iv) are all false. Since (i) is false, there are three cases.

Case 1. There is an α with $0 < \alpha \leqq 1$, and α no greater than the fixed point of each $\gamma \varepsilon K_s$, but there is a g in K_s with fixed point arbitrarily close to 1.

By (5.16)(i), $\gamma[\alpha^*, 1] \subset [\alpha^*, 1]$ for all $\alpha^* \leqq \alpha$ and non-decreasing $\gamma \varepsilon K_s$. Because (iii) is false, there is a $\delta > 0$ such that $f(1) \geqq \delta$ for all $f \varepsilon K$. So, $\gamma[\delta^*, 1] \subset [\delta^*, 1]$ for all $\delta^* \leqq \delta$ and non-increasing $\gamma \varepsilon K$. Consequently, $[\min\{\alpha, \delta\}, 1]$ is P-invariant. Apply (3.3) to complete the proof in Case 1.

Case 2. There is a β with $0 \leqq \beta < 1$, and β no less than the fixed point of each $\gamma \varepsilon K_s$, but there is an f in K_s with fixed point arbitrarily close to 0.

The proof is similar to that for Case 1.

Case 3. There is an α and β, $0 < \alpha \leqq \beta < 1$, with the fixed point of each $\gamma \varepsilon K_s$ in $[\alpha, \beta]$.

Because (iv) is false, there are three sub-cases.

Case 3a. There is an α' with $0 < \alpha' \leqq 1$ and $f(1) \geqq \alpha'$ for all $f \varepsilon K$. Proceed as in Case 1.

Case 3b. There is a β' with $0 \leqq \beta' < 1$ and $g(0) \leqq \beta'$ for all $g \varepsilon K$. The reasoning for Case 1 applies here too.

Case 3c. There is a closed proper sub-interval J of Ω such that for each $\gamma \varepsilon K_s$ and $x \varepsilon \Omega$, $\gamma(x) = 1 - x$ implies $x \varepsilon J$. Choose $J \supset [\alpha, \beta]$ and symmetric about $\frac{1}{2}$, and apply (5.16). □

The condition that P-almost all γ are strict contractions cannot be relaxed completely:

(5.18) EXAMPLE. Let $f(x) = x/2$. Let $g(x) = \frac{1}{2}$, $0 \leqq x \leqq \frac{1}{4}$; $= 2x$, $\frac{1}{4} \leqq x \leqq \frac{1}{2}$; $= 1$, $\frac{1}{2} \leqq x \leqq 1$. Let the support of P be the two-point set $\{f, g\}$.

DISCUSSION OF (5.18). f and g are non-decreasing. Each $x \varepsilon \Omega$ is in the range of f or in the range of g. But the unique P-invariant probability has $\{0, \cdots, \frac{1}{4}, \frac{1}{2}, 1\}$ for its support, by (3.3). □

6. Each $\gamma \varepsilon \Gamma$ is linear. Throughout this section, Γ is the set of linear mappings of the closed unit interval Ω into itself. Plainly, Γ is a compact semigroup of contractions in the sup metric, and each element of Γ is a uniformly strict contraction, except for these two: $x \to x$ and $x \to 1 - x$.

ASSUMPTION. The probability P on Γ assigns positive measure to the set of uniformly strict contractions.

(6.1) SPECIALIZATION OF PREVIOUS RESULTS. For each probability ν on Ω, $P^n\nu$ converges to the unique P-invariant probability μ_P, by (4.4). The support of μ_P is non-decreasing with the support of P, by (3.3); and μ_P depends continuously on P, by (3.5). There are two cases, an exceptional one and a typical one.

In the exceptional one, there is an $x_P \varepsilon \Omega$ such that for P-almost all γ, $\gamma(x_P) = x_P$. Then $\mu_P\{x_P\} = 1$. Unless P assigns positive measure to the constant mapping $x \to x_P$, no power of P^* is a uniformly strict contraction of Δ (or even of the continuous $F \varepsilon \Delta$) in the sup metric, from (5.6). It is not difficult to construct discrete F and G in Δ with $\|P^*F - P^*G\| = \|F - G\|$.

In the typical case, there is no P-fixed point x_P. Then some convolution power of P is a uniformly strict contraction of Δ, by (4.12) and (5.4) (or by (5.15)), so the distribution function of $P^n\nu$ converges uniformly to the distribution function of μ_P. Moreover, μ_P is continuous if and only if P assigns measure 0 to each constant mapping, from (5.7). If, for example, P assigns probability 1 to the set of non-decreasing functions, then P^* is a uniformly strict contraction if and only if P splits, by (5.10).

Let $\hat{P}$ be the conditional P-distribution of γ, given that γ is not the function $x \rightarrow x$. Then μ_P is $\hat{P}$-invariant, and the convolution of $\hat{P}$ with itself assigns measure 0 to the function $x \rightarrow 1 - x$. Consequently, after suitably modifying P, (5.17) gives a necessary and sufficient condition for the support of μ_P to be all of Ω. Suppose, for example, that P assigns probability 1 to the set of non-decreasing functions. Let K_s be the set of uniformly strict contractions in the support of P. Then the support of μ_P is all of Ω if and only if each interior point of Ω is in the range of some $\gamma \varepsilon K_s$, and there are $\gamma \varepsilon K_s$ with fixed points arbitrarily close to 0 and to 1.

Unless P assigns positive measure to the set of constant mappings, μ_P is either absolutely continuous or purely singular, by (2.5), but it is difficult to decide which. For $0 \leqq t < 1$, let P_t assign probability $\frac{1}{2}$ to each of the two functions $x \rightarrow tx$ and $x \rightarrow (1 - t) + tx$. Then F_t, the distribution function of the unique P_t-invariant probability, is the distribution function of $(1 - t)\sum_{n=0}^{\infty} t^n \varepsilon_n$, where ε_n are independent and 0 or 1 with probability $\frac{1}{2}$ each. Deciding whether F_t is purely singular or absolutely continuous for $t > \frac{1}{2}$ is a famous open question; for a discussion, see (Garsia, 1962). Of course, $F_t(x) = 1 - F_t(1 - x)$; F_0 assigns mass $\frac{1}{2}$ to 0 and to 1; $F_{\frac{1}{2}}$ is the uniform distribution; F_{1-} assigns mass 1 to $\frac{1}{2}$. We guess that, for each $x \geqq \frac{1}{2}$, $F_t(x)$ increases with t.

The next theorem relies on the special nature of Γ.

(6.2) THEOREM. *Suppose Q and R are probabilities on Γ. Then $Q^* = R^*$ implies $Q = R$.*

PROOF. Let $\gamma \varepsilon \Gamma$ have slope $\sigma(\gamma)$, and let $\tilde{Q}$ be the distribution under Q of the random vector $\gamma \rightarrow (\sigma(\gamma), \gamma(0))$. Since $\tilde{Q}$ determines Q, it suffices to prove that Q^* determines $\tilde{Q}$. For each n, let

$$q_n(x) = x^n \int_\Gamma [\sigma(\gamma)]^n Q(d\gamma) + x^{n-1} \int_\Gamma [\sigma(\gamma)]^{n-1} \gamma(0) Q(d\gamma) + \cdots + \int_\Gamma [\gamma(0)]^n Q(d\gamma).$$

The coefficients of the polynomial q_n are the nth order moments of $\tilde{Q}$. Since $\tilde{Q}$ is determined by its moments, it suffices to prove Q^* determines $q_n(x)$ for each $x \varepsilon \Omega$. Let $F_x \varepsilon \Delta$ assign measure 1 to $x \varepsilon \Omega$. Then

$$\int_\Omega y^n (Q^* F_x)(dy) = \int_\Omega [\gamma(x)]^n Q(d\gamma) = q_n(x). \square$$

REFERENCES

BUSH, ROBERT R. and MOSTELLER, FREDERICK (1953). A stochastic model with applications to learning. *Ann. Math. Statist.* **24** 559–585.

DUBINS, LESTER E. and FREEDMAN, DAVID A. (1966). Random distribution functions. To appear in *Fifth Berkeley Symp. Math. Statist. Prob.*, Univ. of California Press.

DUNFORD, NELSON and SCHWARTZ, JACOB T. (1958). *Linear Operators, part I: General Theory.* Wiley, New York.

GARSIA, A. M. (1962). Arithmetic properties of Bernoulli convolutions. *Trans. Amer. Math. Soc.* **102** 409–432.

GRENANDER, ULF (1963). *Probabilities on Algebraic Structures.* Wiley, New York.

HARRIS, T. E. (1952). A method for limit theorems in Markov chains. *Ann. Math. Statist.* **23** 141.

HAUSDORFF, FELIX (1957). *Set Theory.* Chelsea, New York.

KARLIN, SAMUEL (1953). Some random walks arising in learning models, I. *Pacific J. Math.* **3** 725–756.

KURATOWSKI, CASIMIR (1932). Les fonctions semi-continues dans l'espace des ensembles fermés. *Fund. Math.* **18** 148–159.