MARKOV PROCESSES AND UNIQUE STATIONARY PROBABILITY MEASURES

RICHARD ISAAC

1. Introduction. Let $X_k, k = 0, 1, 2, \cdots$ be a Markov process defined on a measurable space (Ω, Σ) with stationary transition probabilities $P^k(t, E)$. A stationary probability measure (SPM) R for the X_k process satisfies

$$\int P(t, E)R(dt) = R(E), t \in \Omega, E \in \Sigma, P(t, E) = P^1(t, E) . \tag{1.1}$$

We pose the following problem: determine some useful conditions that will ensure the uniqueness of an SPM. Section 2 investigates this problem from several angles in a general setting. Section 3 applies the results to learning processes (defined in § 3) and finally we conclude with an example where $P(t, E)$ has a continuous density.

2. Theorems yielding uniqueness. Define $1/n \sum_{k=1}^{n} P^k(t, E) = Q^n(t, E)$. In general, in the following, if $P^k(\cdot)$ is a function of some variables depending upon the positive integers k, then $Q^n(\cdot) = 1/n \sum_{k=1}^{n} P^k(\cdot)$. A simple sufficient condition shall be employed to conclude that an SPM, if it exists, is unique. Let $\Sigma_0 \subseteqq \Sigma$ be a determining class of sets for Σ, i.e., Σ is the minimal σ-field generated by the class Σ_0. Suppose $\lim_{n\to\infty} Q^n(t, E) = P_0(t, E)$ exists for each $t \in \Omega, E \in \Sigma_0$. Let $R(\cdot)$ be a SPM. Then

$$R(E) = \int Q^n(t, E)R(dt) \text{ for all } n = 1, 2, \cdots \tag{2.1}$$

This implies

$$R(E) = \lim_{n\to\infty} \int Q^n(t, E)R(dt) = \int P_0(t, E)R(dt) . \tag{2.2}$$

If $P_0(t, E) = P_0(E)$ is independent of t for each $E \in \Sigma_0$, $R(E) = P_0(E)$ on Σ_0. If $S(\cdot)$ is another SPM, the same reasoning shows $R(E) = S(E)$ on Σ_0 and so R and S are identical on Σ. Consequently all theorems in this section will have as object to show

$$P_0(t, E) = P_0(E) \text{ independent of } t , \tag{2.3}$$

for all $E \in \Sigma_0$, where $\Sigma_0 \subseteqq \Sigma$ and is a determining class of sets.

Although the primary concern in this paper is with uniqueness

Received March 6, 1961. Research supported in part by National Science Foundation Grant NSF-G14146 at Columbia University.

problems without regard to the question of existence, there are two important cases in which it would be worthwhile to mention that existence is assured. Suppose $\lim_{n\to\infty} Q^n(t, E) = P_0(t, E)$ exists for each $t \in \Omega$, $E \in \Sigma$. Then it is true that $P_0(t, \cdot)$ is an SPM for each t. The other situation concerns the case when Ω is a compact Hausdorff space and Σ consists of the Borel sets of Ω. Suppose the linear transformation $Tf(\cdot) = \int P(\cdot, dy)f(y)$ carries the space of real-valued continuous functions on Ω into itself. Then T^*, the adjoint transformation, maps the space of regular countably additive finite signed measures into itself by the relation: $T^*\mu = \int P(t, \cdot)\mu(dt)$. Since 1 is a proper value of T and $|T| = 1$, 1 is a proper value of T^*. Thus there is a finite signed measure μ with $\mu(\cdot) = \int P(t, \cdot)\mu(dt)$. μ must, in fact, be a measure. For if μ has negative values on Σ, a Hahn decomposition yields a Borel set H with $\mu(\Omega) < \mu(H)$. But then $\mu(H) = \int P(t, H)\mu(dt) \leqq \mu(\Omega)$ yields a contradiction. Thus $\mu(\cdot)/\mu(\Omega)$ is an SPM. This example will be applicable in § 3 when the learning process is discussed.

In the first case above when $\lim_{n\to\infty} Q^n(t, E) = P_0(t, E)$ exists for $t \in \Omega$, $E \in \Sigma$, there will be a decomposition of Ω into ergodic sets with the usual properties as discussed in [8]. Then (2.3) says that there is only one ergodic class or that the process is metrically transitive. The first theorem shows how this characterization may be employed.

THEOREM 1. *Let $\lim_{n\to\infty} Q^n(t, E) = P_0(t, E)$ for $t \in \Omega$, $E \in \Sigma$, and let Σ be a strictly separable σ-field (i.e. generated by a countable family of sets). If there exists a point $t_0 \in \Omega$ such that, for each $t \in \Omega$, there is an integer $n(t)$ and a number $\varepsilon(t) > 0$ such that $P^{n(t)}(t, \{t_0\}) \geqq \varepsilon(t)$ then $P_0(t, E) = P_0(E)$, for $E \in \Sigma$.*

Proof. According to Theorem 2 in [8], since $P_0(t, E)$ is appropriately defined and Σ is strictly separable, there is a decomposition $\Omega = F + \sum_\alpha A_\alpha$ into disjoint sets where the A_α are ergodic and F is a null set. If there were two distinct nonempty ergodic sets A_1 and A_2, the hypothesis implies that $t_0 \in A_1$ and $t_0 \in A_2$ because each A_α is closed. However, $A_1 \cap A_2 = \phi$ and thus the decomposition reduces to $\Omega = F + A$. Then $P_0(t, E) = P_0(E)$ independent of t for $t \in A$. For $t \in F$ we have

$$P_0(t, E) = \int P_0(y, E)P_0(t, dy) = \int_A P_0(y, E)P_0(t, dy) = P_0(E) .$$

Theorem 1 is a generalization of a theorem stated in [7] where $n(t)$ and $\varepsilon(t)$ are chosen independently of t. However, under such uniformity restrictions, one obtains $\lim_{n\to\infty} P^n(t, E) = P_0(E)$ uniformly in t.

In case each point does not have positive probability of leading to a distinguished point t_0, it may be that each point does behave well enough with regard to some set containing t_0 to ensure independence of t. The following theorems will assume that Ω is a metric space and all mention of continuity on Ω refers to the topology of this metric. It should be noted that each theorem postulates the existence of a point t_0 having a certain relationship with regard to all $t \in \Omega$, as is the case in Theorem 1. Thus, although the methods differ from one theorem to another, the intuitive content of the hypotheses remains the same: to tie up the behavior of each t intimately enough with some distinguished point t_0. Henceforth Σ refers to the σ-field generated by the open sets under the metric topology. For the remainder of this section it will be assumed without further mention that $\lim_{n\to\infty} Q^n(t, E) = P_0(t, E)$ exists for all $t \in \Omega$, $E \in \Sigma_0$, where Σ_0 determines Σ. Our object will be to show that (2.3) holds under various conditions, and so there is then at most one SPM for the process.

In the following, it will be helpful to consider the usual space Ω of sequences ξ: $(\omega_0, \omega_1, \cdots)$, $\omega_i \in \Omega$, with the usual infinite product probability $P(\cdot)$ and conditional probability $P(\cdot \mid \cdot)$ defined on Ω (see [4], p. 190). Statements such as (2.4) to follow should be referred to this background.

DEFINITION Let $S_\varepsilon(t)$ be the open ε-sphere about t as center. A point t_0 is called attractive if, for every $\varepsilon > 0$, the probability that the process enters $S_\varepsilon(t_0)$ infinitely often, starting from any initial position, is 1. In symbols

$$P(X_n \in S_\varepsilon(t_0)\ i.o. \mid X_0 = t) = 1,\ \varepsilon > 0,\ t \in \Omega\ . \tag{2.4}$$

Another way of saying this is that the conditional probability of the process entering any open set containing t_0 infinitely often is 1.

THEOREM 2. *A condition sufficient to ensure that $P_0(t, E) = P_0(E)$ independent of t for a fixed $E \in \Sigma_0$ is that there should exist an attractive point t_0 with $P_0(\cdot, E)$ continuous at t_0. If there exists a unique* SPM *for the process and Ω is separable (i.e., the metric topology has a countable base) then the condition is necessary provided Σ_0 is an open base for the topology of Ω.*

Thus, if there is, for each $E \in \Sigma_0$, an attractive point $t_0(E)$ satisfying the above condition, $P_0(t, E) = P_0(E)$ for all $E \in \Sigma_0$, and there is at most one SPM.

Proof. First the necessity is proved. Let R_0 be the unique SPM. $P_0(t, E) = R_0(E)$ by a result of Doob (TAMS vol. 63, (1948) p. 400, theorem 1 (d)) and from our assumption $P_0(t, E) = P_0(E)$. We show that there is an attractive point. (since $P_0(E)$ is constant, continuity is trivial). There must be a point t_0 with $R_0(S_\varepsilon(t_0)) > 0$ for all $\varepsilon > 0$. For, if not, then for each $t \in \Omega$, there is $\varepsilon_t > 0$ with $R_0(S_{\varepsilon_t}(t)) = 0$. $\bigcup_{t \in \Omega} S_{\varepsilon_t}(t) = \Omega$ and is an open cover. By Lindelof's theorem there is a countable subcover $\bigcup_{t_i \in \Omega} S_{\varepsilon_{t_i}}(t_i)$ and so

$$0 = \sum_{i=1}^{\infty} R_0(S_{\varepsilon_{t_i}}(t_i)) \geqq R_0(\bigcup_{t_i \in \Omega} S_{\varepsilon_{t_i}}(t_i)) = R_0(\Omega) = 1$$

a contradiction, so that t_0 must exist as asserted. Then

$$\limsup_{n\to\infty} Q^n(t, S_\varepsilon(t_0)) \geqq \lim_{n\to\infty} Q^n(t, E_\varepsilon) = R(E_\varepsilon) \geqq R(S_{\varepsilon'}(t_0)) \geqq \delta(\varepsilon) > 0 \tag{2.5}$$

for $t \in \Omega$, $\varepsilon > 0$, where $S_{\varepsilon'}(t_0) \subset E_\varepsilon \subset S_\varepsilon(t_0)$ and $E_\varepsilon \in \Sigma_0$. This implies that for each t there are infinitely many n (where n depends upon t and ε) with $P^n(t, S_\varepsilon(t_0)) > \delta(\varepsilon)/2 > 0$ for each $\varepsilon > 0$. Since $P(X_n \in S_\varepsilon(t_0)$ for some $n \mid X_0 = t) \geqq P(X_n \in S_\varepsilon(t_0) \mid X_0 = t) = P^n(t, S_\varepsilon(t_0))$ for each n, one obtains $\inf_t P(X_n \in S_\varepsilon(t_0)$ for some $n \mid X_0 = t) > \delta(\varepsilon)/2 > 0$ and by a theorem of Doeblin (See [2]), $P(X_n \in S_\varepsilon(t_0)\ i.o.) = 1$. Since this result holds with an arbitrary distribution on X_0, choose $P(X_0 = t) = 1$ and so, under this assumption, $P(X_n \in S_\varepsilon(t_0)\ i.o. \mid X_0 = t) = 1$. However, this conditional probability only depends upon the transition probabilities, so the statement is true for arbitrary distributions of X_0, each $t \in \Omega$ and all $\varepsilon > 0$. Thus, t_0 is attractive.

To prove the sufficiency, define $P^k(t, E, A) =$ probability of attaining E on the kth step after having passed through A sometime on or before the kth step, starting from t. Set $P_1^k(t, E, A) = P^k(t, E) - P^k(t, E, A) =$ probability of attaining E on the kth step without ever having visited A on or before the kth step, starting from t. In addition, (if E^c is the complement of a set E)

$$P(X_2 \in B, X_1 \notin A \mid X_0 = t) = \int_{A^c} P(y, B) P(t, dy)\ . \tag{2.6}$$

Let $P_A^2(t, B)$ be the integral (2.6). Define $P_A^k(t, B)$ by recursion by

$$P_A^k(t, B) = \int_{A^c} P(y, B) P_A^{k-1}(t, dy)\ , \qquad k > 2 \tag{2.7}$$

Set $P_A^1(t, B) = P(t, B)$. It is clear that $P(X_k \in B, X_n \notin A$, all $n < k \mid X_0 = t) = P_A^k(t, B)$ for $k > 1$ and that, fixing $E \in \Sigma_0$

$$\begin{aligned} P^k(t, E, A) = &\int_A P^{k-1}(y, E) P(t, dy) + \int_A P^{k-2}(y, E) P_A^2(t, dy) \\ &+ \cdots + \int_A P(y, E) P_A^{k-1}(t, dy) + P_A^k(t, A \cap E) \end{aligned} \tag{2.8}$$

$$= \sum_{i=1}^{k-1} \int_A P^{k-i}(y, E) P_A^i(t, dy) + P_A^k(t, A \cap E) .$$

Let k_0 be fixed, and let $k > k_0$, then

$$\begin{aligned} P^k(t, E) &= \sum_{i=1}^{k-1} \int_A P^{k-i}(y, E) P_A^i(t, dy) + P_A^k(t, A \cap E) \\ &\quad + P_1^k(t, E, A) \leqq \sum_{i=1}^{k_0} \int_A P^{k-i}(y, E) P_A^i(t, dy) \\ &\quad + \delta + P_A^k(t, A \cap E) + P_1^k(t, E, A) , \end{aligned} \tag{2.9}$$

where

$$\delta \leqq \sum_{k=k_0+1}^{\infty} P_A^k(t, A) . \tag{2.10}$$

Using the truncation at k_0, we may sum terms, and have, for $n > k_0$

$$\begin{aligned} \sum_{k=1}^{n} P^k(t, E) &\leqq \sum_{k=1}^{k_0} \int_A \left[\sum_{i=1}^{n-k} P^i(y, E) \right] P_A^k(t, dy) \\ &\quad + (n - k_0)\delta + 1 + \sum_{k=1}^{n} P_1^k(t, E, A) . \end{aligned} \tag{2.11}$$

Dividing by n and taking the limit yields

$$P_0(t, E) \leqq \sum_{k=1}^{k_0} \int_A P_0(y, E) P_A^k(t, dy) + \delta + \limsup_n Q_1^n(t, E, A) . \tag{2.12}$$

Observe that $\sum_{k=1}^{\infty} P_A^k(t, A)$ is the probability of entering A at least once, starting from t. Since $\sum_{k=1}^{\infty} P_A^k(t, A) \geqq P(X_n \in A \ i.o. \mid X_0 = t)$, if we place $A = S_\varepsilon(t_0)$ for any $\varepsilon > 0$ for t_0 attractive satisfying the continuity hypothesis with regard to E, we have

$$\sum_{k=1}^{\infty} P_{S_\varepsilon(t_0)}^k(t, S_\varepsilon(t_0)) = 1 \tag{2.13}$$

(2.13) implies $\limsup_n Q_1^n(t, E, S_\varepsilon(t_0)) = 0$ and if $k_0 \to \infty$, (2.10) implies $\delta \to 0$. This yields

$$P_0(t, E) \leqq \sum_{k=1}^{\infty} \int_{S_\varepsilon(t_0)} P_0(y, E) P_{S_\varepsilon(t_0)}^k(t, dy) . \tag{2.14}$$

A similar argument gives the opposite inequality in (2.14) (for arbitrary A), and proves

$$P_0(t, E) = \sum_{k=1}^{\infty} \int_{S_\varepsilon(t_0)} P_0(y, E) P_{S_\varepsilon(t_0)}^k(t, dy) . \tag{2.15}$$

$P_0(\cdot, E)$ is continuous at t_0, so for ε small,

$$(P_0(t_0, E) - h) \sum_{k=1}^{\infty} P_{S_\varepsilon(t_0)}^k(t, S_\varepsilon(t_0)) \leqq P_0(t, E)$$

$$\leqq (P_0(t_0, E) + h) \sum_{k=1}^{\infty} P^k_{S_\varepsilon(t_0)}(t, S_\varepsilon(t_0)) .$$

Let $\varepsilon \to 0$; continuity and (2.13) yield (2.3) and the proof is complete.

In case Ω is compact an interesting sufficient condition for uniqueness may be phrased. Call t_0 an accumulation point of the process if

$$(2.16) \qquad P(X_n \in S_\varepsilon(t_0), \text{ for some } n \mid X_0 = t) = \delta(\varepsilon, t) > 0$$

for every $\varepsilon > 0$. Notice that an attractive point is always an accumulation point.

THEOREM 3. *At most one SPM exists under the following conditions:*
(2.17) *Ω is compact.*
(2.18) *$P_0(\cdot, E)$ is continuous on Ω for each $E \in \Sigma_0$.*
(2.19) *The process has an accumulation point t_0.*

Proof. Let $E \in \Sigma_0$ be fixed. $P_0(\cdot, E)$ is continuous on Ω compact, so there are two points t_M, t_m such that $P_0(t_M, E)$ is a maximum and $P_0(t_m, E)$ is a minimum. Suppose A is a set with $P^k(t_M, A) > 0$ for some k. Then for some $y' \in A$, $P_0(y', E) = P_0(t_M, E)$. For, if not, $P_0(y, E) < P_0(t_M, E)$ for all $y \in A$ and

$$(2.20) \quad P_0(t_M, E) = \int_A P_0(y, E) P^k(t_M, dy) + \int_{A^c} P_0(y, E) P^k(t_M, dy)$$
$$< P_0(t_M, E) P^k(t_M, A) + P_0(t_M, E) P^k(t_M, A^c) = P_0(t_M, E) .$$

(2.19) implies there is a point t_0 with $P(X_n \in S_\varepsilon(t_0)$, for some $n \mid X_0 = t_M) = \delta(\varepsilon, t_M) > 0$ for each $\varepsilon > 0$. By the preceding, this means that the spheres $S_{\varepsilon_i}(t_0)$, $\varepsilon_i \to 0$, give rise to a sequence of points $\{y_i\}$ with $P_0(y_i, E) = P_0(t_M, E)$. But $\lim_{i\to\infty} y_i = t_0$ and by continuity $P_0(t_0, E) = P_0(t_M, E)$. A similar discussion involving t_m shows that $P_0(t_0, E) = P_0(t_m, E)$ and therefore $P_0(\cdot, E)$ is independent of t.

THEOREM 4. *At most one SPM exists under the following conditions:*
(2.21) *Ω is complete.*
(2.22) *Is A is the set of attractive points of the process, then A has a nonvoid interior.*
(2.23) *The transition probabilities $P^k(\cdot, E)$ are continuous on Ω for $E \in \Sigma_0$.*

Proof. The proof is a category argument. A residual set is a set whose complement is a set of category one, i.e., a set which is the union of a countable class of nowhere dense sets. It is known (see [10] p. 70, problem p) that on a complete metric space any function which is the pointwise limit of continuous functions is itself continuous on a residual set. Hence, by (2.21) and (2.23), $\lim_{n\to\infty} Q^n(\cdot, E) = P_0(\cdot, E)$ is at least continuous on a residual set R. Since R^c is of the first category, it has a vacuous interior. Therefore, there exists a point $t_0 \in A \cap R$ by (2.22). Theorem 2 concludes the proof.

REMARK. Theorem 4 is a special case of the simple principle: If more objects have characteristic B than characteristic A, there is some object with characteristics B and A^c. This principle may be applied in more general ways in conjunction with Theorem 2.

3. **Application to learning processes.** Let there be given $2N$ continuous functions from $[0, 1]$ into itself: $f_1, \cdots, f_N; p_1, \cdots, p_N$. The process is defined by a random walk on the unit interval where a point initially at t moves to $f_i(t)$ with probability $p_i(t)$. One requires $\sum_{i=1}^N p_i(t) = 1$ for $t \in [0, 1]$. The transition probability is defined by

$$P(t, E) = \sum_{i \ni f_i(t) \in E} p_i(t),\ E, \text{ a Borel set. } 0 \leqq t \leqq 1 . \tag{3.1}$$

N shall be assumed finite throughout the discussion. Such Markov processes with a continuum of possible states, but only a countable number of possible states given a starting position, arise often and have sometimes been designated as learning processes because of their occurrence in psychological learning model studies. For some discussions of these processes see, for example, [1] and [9]. Notice that the operator $Tf(\cdot) = \sum_{i=1}^N p_i(\cdot)\ f(f_i(\cdot))$ takes the space of continuous functions f on $[0, 1]$ into itself. The remarks preceding Theorem 1 thus guarantee the existence of an SPM for the learning process.

THEOREM 5. *A unique* SPM *exists for a learning process satisfying the following conditions*:

$$\sum_{i=1}^N p_i(t_1)\,|f_i(t_1) - f_i(t_2)| \leqq \alpha\,|t_1 - t_2| \text{ for some } \alpha < 1, \textit{ for all } t_1, t_2 \tag{3.2}$$

$$|p_i(t_1) - p_i(t_2)| \leqq \beta\,|t_1 - t_2| \textit{ for some } \beta > 0 \textit{ for all } i, t_1, t_2. \tag{3.3}$$

(3.4) *There is an attractive point for the process.*

Under these conditions $\lim_{n\to\infty} Q^n(t, E)$ *converges uniformly in* t *to a limit* $P_0(E)$ *for each* $E \in \Sigma$.

Proof. The theorem is proved by means of two general lemmas. We use the expression "learning-type process" to refer to a learning process as described above except that the state space may be an arbitrary bounded metric space Ω rather than $[0, 1]$.

LEMMA 1. *Let Ω be an arbitrary bounded metric space on which a learning type process is defined where the f_i are bounded uniformly continuous functions on Ω into itself. Let* (3.2) *and* (3.3) *be satisfied where absolute value is to be interpreted as the metric distance when such an interpretation is appropriate.*

Then $\lim_{n\to\infty} Q^n(t, E)$ exists for $E \in \Sigma$, $t \in \Omega$ unifomly (in t) to a limit $P_0(t, E)$.

Proof of Lemma 1. By a corollary to the Stone-Weierstrass theorem, ([5], p. 276) Ω can be densely embedded in a compact Hausdorff space Ω_1 such that the p_i have unique continuous extensions to continuous functions p_i^* on Ω_1. Ω_1 is metrizable because it contains a metric space as a dense subset, and since Ω_1 is complete the uniformly continuous functions $f_i: \Omega \to \Omega_1$ can be extended by continuity to unique uniformly continuous functions $f_i^*: \Omega_1 \to \Omega_1$. (3.2) and (3.3) continue to hold on Ω_1. Σ_1, the class of Borel sets of Ω_1, includes Σ as a subclass. Proving the lemma for the process on Ω_1 defined by p_i^* and f_i^* clearly imply its truth for the original process. It is therefore no loss of generality to assume at the outset that Ω is a compact metric space, which we shall do.

Consider the transformation, T, given by:

$$Tg = \sum_{i=1}^{N} p_i(t)g(f_i(t)) \tag{3.5}$$

where T is defined on the space S of continuous real or complex valued functions on Ω satisfying

$$\sup_{t_1, t_2 \in \Omega} \frac{|g(t_1) - g(t_2)|}{|t_1 - t_2|} = m(g)$$

where $m(g)$ is a finite constant. Set $\max_{t\in\Omega} |g(t)| = M(g)$. Then S becomes a Banach space under the norm:

$$\|g\| = M(g) + m(g)\ . \tag{3.6}$$

We also have:

$$\begin{aligned} |(Tg)(t_1) - (Tg)(t_2)| &= \Big|\sum_{i=1}^{N} p_i(t_1)g(f_i(t_1)) \\ &\quad - \sum_{i=1}^{N} p_i(t_2)g(f_i(t_2))\Big| \leqq \Big|\sum_{i=1}^{N} p_i(t_1)[g(f_i(t_1) - g(f_i(t_2))]\Big| \\ &\quad + \Big|\sum_{i=1}^{N} [p_i(t_1) - p_i(t_2)]g(f_i(t_2))\Big| \\ &\leqq m(g) \sum_{i=1}^{N} p_i(t_1)\,|f_i(t_1) - f_i(t_2)| + N\beta M(g)\,|t_1 - t_2| \end{aligned} \tag{3.7}$$

$$\leqq \alpha m(g) \mid t_1 - t_2 \mid + N\beta M(g) \mid t_1 - t_2 \mid < \left(1 + \frac{N\beta}{1-\alpha}\right) \| g \| \mid t_1 - t_2 \mid$$

and

$$M(Tg) \leqq M(g) \tag{3.8}$$

(3.7) and (3.8) yield:

$$\begin{aligned} \| Tg \| &= M(Tg) + m(Tg) \leqq M(g) + \left(1 + \frac{N\beta}{1-\alpha}\right) \| g \| \\ &\leqq \| g \| + \left(1 + \frac{N\beta}{1-\alpha}\right) \| g \| \leqq \left(2 + \frac{N\beta}{1-\alpha}\right) \| g \| . \end{aligned} \tag{3.9}$$

(3.9) assures the continuity of T. In fact $\| T^n \|$ is uniformly bounded since, assuming the inequality (3.7) for $n - 1$ iterations and proceeding by induction:

$$\begin{aligned} &\mid (T^n g)(t_1) - (T^n g)(t_2) \mid = \mid T[(T^{n-1}g)](t_1) - T[(T^{n-1}g)](t_2) \mid \\ &\leqq \mid \sum_{i=1}^{N} p_i(t_1)[g_{n-1}(f_i(t_1)) - g_{n-1}(f_i(t_2))] \mid \\ &+ \mid \sum_{i=1}^{N} [p_i(t_1) - p_i(t_2)] g_{n-1}(f_i(t_2)) \mid \leqq \alpha\left(1 + \frac{N\beta}{1-\alpha}\right) \| g \| \mid t_1 - t_2 \mid \\ &+ N\beta M(g) \mid t_1 - t_2 \mid = \alpha\left(1 + \frac{N\beta}{1-\alpha}\right)[M(g) + m(g)] \mid t_1 - t_2 \mid \\ &+ N\beta M(g) \mid t_1 - t_2 \mid = \left[\left(\alpha + \alpha\frac{N\beta}{1-\alpha} + N\beta\right) M(g) \right. \\ &\left. + \left(\alpha + \alpha \frac{N\beta}{1-\alpha}\right) m(g)\right] \mid t_1 - t_2 \mid \leqq \left(1 + \frac{N\beta}{1-\alpha}\right) \| g \| \mid t_1 - t_2 \mid . \end{aligned} \tag{3.10}$$

Thus $\| T^n \| \leqq 2 + N\beta/(1-\alpha)$ for all $n = 1, 2, \cdots$. Doeblin and Fortet in [3], p. 142 ff. sketch a proof that (3.3) and an assumption slightly less general than (3.2) imply that T is a quasi-compact (sometimes called quasi-completely continuous) operator in S. (See [11], for example, for definition). [7] deals with the general case of the operator T. The same arguments in [3] apply here with no change since $\| T^n \|$ is uniformly bounded in n as shown above. The work in [3] essentially analyses the spectrum of T. The general idea is the following: It is observed that there are a finite number of proper values on $\mid z \mid = 1$, each defining a projection $E(\lambda_i)$ onto a finite dimensional subspace (therefore $E(\lambda_i)$ is a compact operator). Moreover, the proper values do not accumulate to any point on $\mid z \mid = 1$. Lemma 4 in [3] shows that T has no continuous spectrum, for if it had, there would exist vectors f_n with $\| f_n \| = 1$ and $\lim_{n\to\infty} \| (\lambda I - T) f_n \| = 0$. But Lemma 4 asserts that there exists a constant C independent of g so that $(\lambda I - T)f = g$ implies $\| f \| / C \leqq \| g \|$.

The residual spectum cannot accumulate to the circle $|z| = 1$, since the residual spectrum of T is in the set of proper values of T^*, the adjoint operator, and the argument used to show that the proper values of T do not accumulate to $|z| = 1$ goes through here with little variation working with T^*. The final conclusion is that there are a finite number of spectral points on $|z| = 1$, each defining a compact projection $E(\lambda_i)$ in S, and there is a number $0 < s < 1$ such that, if z is in the spectrum of T and $|z| \neq 1$, then $|z| \leqq s$. This implies that $T = B + \sum_{i=1}^{k} E(\lambda_i)$ where $|B^n| \to 0$ geometrically in n. This shows that T is quasi-compact. (A good discussion of some of these concepts may be found in [5], especially Chapter 7 and Chapter 8, § 8, ff. Also see [11].)

Since T is quasi-compact, so is T^*. Each countably additive regular real or complex-valued signed measure μ is in S^*. By Fubini's theorem $T^*\mu = \int P(t, \cdot)\mu(dt)$. It is a standard fact that if T^* is quasi-compact, then $1/n \sum_{k=1}^{n} (T^*)^k$ converges in the uniform operator topology to a projection on the manifold of fixed points under T^*. By looking at the kernels of the integrals this means that $1/n \sum_{k=1}^{n} P^k(t, \cdot) = Q^n(t, \cdot)$ converges uniformly to a limit $P_0(t, \cdot)$ and so $\lim_{n\to\infty} Q^n(t, E) = P_0(t, E)$ for $t \in \Omega$, $E \in \Sigma$, proving Lemma 1.

LEMMA 2. *Let Ω be an arbitrary bounded metric space on which a learning type process is defined* (*the functions p_i and f_i are continuous, but do not necessarily satisfy* (3.2) *and* (3.3)). *Suppose that the conclusions of Lemma* 1 *above hold and that the process has an attractive point t_0. Then there is a unique* SPM.

Proof of Lemma 2. Existence of an SPM is assured by comments preceding Theorem 1. By virtue of Theorem 2, it is sufficient to show that $P_0(\cdot, E)$ is continuous at t_0 for all open sets E in some base for the metric topology in Ω. Let A_n be the finite set of points with $P(X_n \in A_n \mid X_0 = t_0) = 1$ and A_n is minimal. Set $A = \bigcup_{n=1}^{\infty} A_n$. Let $\Sigma_0(t)$ be the class of all open spheres $S_\varepsilon(t)$ about t such that $[A \cap$ boundary of $S_\varepsilon(t)] = \phi$. A is a countable set so Σ_0 (t) is a local base at t and $\bigcup_{t\in\Omega} \Sigma_0(t) = \Sigma_0$ is a base for the metric topology. If $S_\varepsilon(t') \in \Sigma_0$, then $P^k(t_0, S_\varepsilon(t'))$ is continuous at t_0 for each k since a discontinuity could only occur if X_k should take values on the boundary of $S_\varepsilon(t')$ which cannot happen. We then have:

$$\begin{aligned} P_0(t_0, S_\varepsilon(t')) &= \lim_{n\to\infty} Q^n(t_0, S_\varepsilon(t')) = \lim_{n\to\infty} \lim_{t\to t_0} Q^n(t, S_\varepsilon(t')) \\ &= \lim_{t\to t_0} \lim_{n\to\infty} Q^n(t, S_\varepsilon(t')) = \lim_{t\to t_0} P_0(t, S_\varepsilon(t')) \,. \end{aligned}$$

The interchange of order in taking limits is justified since the conclusion of Lemma 1 asserts that $\lim_{n\to\infty} Q^n(t, E) = P_0(t, E)$ uniformly in t. Since

$P_0(\cdot, S_\varepsilon(t'))$ is continuous at t_0 for all spheres in Σ_0, Theorem 2 concludes the proof of the lemma.

Theorem 5 follows as a special case of Lemmas 1 and 2.

COROLLARY 1. (*Blackwell* [1]) *If a learning type process is defined on a bounded metric space Ω such that*:

$$(3.11) \qquad |f_i(t_1) - f_i(t_2)| \leqq \alpha |t_1 - t_2| , \qquad \textit{for } \alpha < 1 .$$

$$(3.12) \qquad |p_i(t_1) - p_i(t_2)| \leqq \beta |t_1 - t_2| , \qquad \textit{for } \beta > 0 .$$

$$(3.13) \qquad p_i(t) \geqq \varepsilon > 0 \,[1] , \qquad t \in \Omega .$$

for all i, t_1, t_2 in Ω, then there is at most one SPM.

Proof. (3.11) is a special case of (3.2) so Lemma 1 holds. (This, incidentally, proves existence) It again is no loss of generality to assume that Ω is compact and so complete. Let $f_i^{(n)}$ denote composition of the function f_i with itself n times. The boundedness of Ω and (3.11) imply that the diameter of $f_1^{(n)}(\Omega)$, say, converges to zero, so by the completeness of Ω, there is a point $t_0 = \lim_{n\to\infty} f_1^{(n)}(\Omega)$. (3.11) and (3.13) make it clear that t_0 is indeed attractive by using the useful theorem of Doeblin quoted in the proof of the necessity in Theorem 2.

COROLLARY 2. *Let a process be defined on an arbitrary metric space Ω and let the operator $T\mu = \int P(t, \cdot)\mu(dt)$ be defined on some Banach space of measures into itself where $P(t, E)$ is the transition probabiliy of the process. Suppose that T defines a quasi-compact transformation on this Banach space. In addition, let there exist an attractive point t_0 and a countable set $A \subset \Omega$ with $P(X_n \in A \mid X_0 = t_0) = 1$ for each n. Then there is a unique* SPM.

Proof. Clear from Lemmas 1 and 2.

EXAMPLE 1.
$$f_1(t) = \sigma t \; ; \qquad f_2(t) = 1 - \alpha + \alpha t$$
$$p_1(t) = t \; ; \qquad p_2(t) = 1 - t$$

where the random walk is on $[0, 1]$, $0 < \sigma < 1$, $0 < \alpha < 1$. It is not difficult to check that 0 is an attractive point even though (3.13) is not satisfied. For any interval $[0, s]$ about 0, $\inf_t P(X_k \in [0, s]$ for some $k \mid X_0 = t) > 0$ and so Doeblin's theorem is applicable. Therefore there is a unique SPM by theorem 5.

[1] It suffices, clearly, to assume 3.13 only for some i in the proof given here dependent, upon Theorem 5. We state the theorem, however, as it appears in [**1**].

Example 2. $$f_1(t) = \sigma t; f_2(t) = 1 - \alpha + \alpha t$$
$$p_1(t) = 1 - t; p_2(t) = t\ .$$

This is the case of absorbing barriers at 0 and 1. There is no unique SPM in this case. From our point of view this occurs because there is no attractive point.

Examples 1 and 2 where discussed at length in [9].

EXAMPLE 3. Let (a_{ij}) be a 4×4 Markov matrix and write:

$$\begin{aligned}
r &= (a_{11} - a_{21})t + a_{21} \\
s &= (a_{11} + a_{12} - a_{21} - a_{22})t + a_{21} + a_{22} \\
u &= (a_{13} - a_{23})t + a_{23} \\
v &= (a_{13} + a_{14} - a_{23} - a_{24})t + a_{23} + a_{24}\ .
\end{aligned}$$

Set

$$f_1(t) = \frac{r}{s}\ ; \qquad p_1(t) = s$$

$$f_2(t) = \frac{u}{v}\ ; \qquad p_2(t) = v \tag{3.14}$$

whenever these functions are defined from $[0, 1]$ into itself. Whenever they are defined $p_i(t) \geqq \varepsilon > 0$ for ε chosen appropriately and all t. f_1 and f_2 are linear fractional transformations, either monotone increasing or decreasing with exactly one fixed point. These transformations have the property that $\lim_{n\to\infty} f_i^n(t) = t_{oi}$ exists where t_{oi} is the fixed point of f_i, as can be easily shown. Therefore, there is an attractive point for the process by the Doeblin Theorem. (3.3) is clearly satisfied. For (3.2) to be satisfied, we consider

$$\begin{aligned}
&\sum_{i=1}^{2} p_i(t)\,|f_i'(t)| \\
&= \left|\frac{a_{11}a_{22} - a_{21}a_{12}}{(a_{11} + a_{12} - a_{21} - a_{22})t + a_{21} + a_{22}}\right| \\
&\qquad + \left|\frac{a_{13}a_{24} - a_{14}a_{23}}{(a_{13} + a_{14} - a_{23} - a_{24})t + a_{23} + a_{24}}\right|\ .
\end{aligned} \tag{3.15}$$

whenever (3.15) is less than 1, Theorem 5 can be applied. The process of (3.14) is a particular example arising from the study of entropy of functions of finite state Markov chains as discussed in [1] when it is important to be able to assert uniqueness. Theorem 5 can be applied to the general category of processes of this nature considered in [1].

The example which follows satisfies Doeblin's condition (D), [4], p. 192, and so much more can be said about it than we do. We simply mention it to show the applicability of the preceding ideas to a well-known case.

4. Example of transition probability with density. Let $p_1(t, s)$ be defined on $R \times R$ into R, $R = (-\infty, +\infty)$ and let λ be Lebesgue measure. Suppose $p_1(t, s) \geqq 0$ and continuous on $R \times R$ and $\int_{-\infty}^{+\infty} p_1(t, s)\lambda(ds) = 1$ for each t where the integral converges uniformly in t. Set $(-\infty, x] = E_x$ and $P(t, E_x) = \int_{-\infty}^{x} p_1(t, s)\lambda(ds)$. The uniform convergence of the integral assures that $P(\cdot, E_x)$ is continuous on R for each fixed $x \in R$. Moreover, it is seen that $P^n(t, E_x) = \int_{-\infty}^{x} p_n(t, s)\lambda(ds)$ where $p_n(t, s)$ is defined inductively by $p_n(t, u) = \int_{-\infty}^{+\infty} p_{n-1}(t, s)p_1(s, u)\lambda(ds)$. The integrals $\int_{-\infty}^{+\infty} p_n(t, s)\lambda(ds) = 1$ converge uniformly in t since:

$$\begin{aligned}\int_{-N}^{+N} p_n(t, s)\lambda(ds) &= \int_{-N}^{+N}\int_{-\infty}^{+\infty} p_{n-1}(t, u)p_1(u, s)\lambda(du)\lambda(ds)\\ &= \int_{-\infty}^{+\infty} p_{n-1}(t, u)\left\{\int_{-N}^{+N} p_1(u, s)\lambda(ds)\right\}\lambda(du)\\ &\geqq (1-\varepsilon)\int_{-\infty}^{+\infty} p_{n-1}(t, u)\lambda(du) = 1 - \varepsilon\ .\end{aligned}$$

$P^n(\cdot, E_x)$ is therefore continuous for each n. Suppose that $p_1(t, s) \geqq \delta > 0$ for all t and all s in an open interval S. Then:

$$P(t, S_\varepsilon(s')) = \int_{S_\varepsilon(s')} p_1(t, s)\lambda(ds) \geqq \delta\lambda(S_\varepsilon(s')) > 0$$

for any fixed $s' \in S$ and all ε sufficiently small. Doeblin's theorem is then applicable to show that s' is attractive and so S consists of attractive points. Theorem 4 then asserts that there is at most one SPM.

To close with a specific example, let $\Omega = [0, \infty)$ and set $P(t, [0, x]) = 1 - e^{-(t+1)x} = \int_0^x (t+1)e^{-(t+1)s}\lambda(ds)$. The integral over Ω converges uniformly in t to 1 and $\inf_t P(t, [0, x]) = 1 - e^{-x} > 0$ for $x > 0$. The above discussion applies to show there is at most one SPM.

Acknowledgement. This paper is an elaboration and revision of part of a doctoral dissertation written at the University of California, Berkeley, under the direction of Professor David Blackwell. The author takes great pleasure in expressing his thanks to Dr. Blackwell for suggesting the problem, and for advice and encouragement during the preparation of the thesis. The author also thanks the referee for some valuable suggestions.

REFERENCES

1. D. Blackwell, *The entropy of functions of finite state Markov chains*, Trans. of the first Prague conference on information theory, stat. decision functions, random processes, Prague, 1957.
2. K. L. Chung and C. Derman, *Non-recurrent random Walks*, Pacific J. Math., **6** (1956), 441-447.
3. W. Doeblin and R. Fortet, *Sur des chaines à liaisons complètes*, Bulletin de la Société Mathématique de France, **65** (1937), 133-148.
4. J. L. Doob, *Stochastic processes*, Wiley and Sons, New York, 1953.
5. N. Dunford and J.T. Schwartz, *Linear operators, I*, Interscience, New York, 1958.
6. W. Feller, *An introduction to probability theory and its applications*, Volume 1, Wiley and sons, New York, 1950.
7. Ionescu-Tulcea and G. Marinescu, *Sur certaines chaînes à liaisons complètes*, C.R. Acad. Sci. Paris, **227** (1948), 667-669.
8. R. Isaac, *Some generalizations of Doeblin's decomposition*, Pacific J. Math. Summer 1961.
9. S. Karlin, *Some random walks arising in learning models I*, Pacific J. Math., **3** (1953), 725-756.
10. M. E. Munroe, *Introduction to measure and integration*, Addison-Wesley, 1953.
11. K. Yosida and S. Kakutani, *Operator theoretical treatment of Markov's process and mean ergodic theorem*, Ann. of Math., **42** (1941).

FORDHAM UNIVERSITY AND
COLUMBIA UNIVERSITY