

A CONVERGENCE THEOREM FOR A CERTAIN CLASS OF MARKOFF PROCESSES

MAURICE KENNEDY

1. Introduction. The object of this paper is to generalize, by means of an approach due to S. Karlin [9], a theorem originally obtained by Bellman, Harris and Shapiro [1] which may be stated in the following way:

A system is considered whose state may be described by a point t in the interval $[0, 1]$. A probability measure μ is given for the initial state of the system. At the end of each unit interval of time, one of the transformations A_0, A_1 is applied to the state t with probabilities $\phi_0(t), \phi_1(t)$ respectively, where $\phi_0(t) + \phi_1(t) = 1$. The transformations are defined by

$$(1.1) \quad A_0 t = \lambda_0 t, \quad A_1 t = \lambda_1 t + (1 - \lambda_1), \quad 0 \leq \lambda_0, \lambda_1 < 1^1.$$

The assumption is made that

$$(1.2) \quad \phi_0(t) = 1 - t, \quad \phi_1(t) = t.$$

It is clear that (1.1) and (1.2) ensure that the end-points of the interval $[0, 1]$ are absorbing, that is, if the state of the system is either 0 or 1, it remains so. Let $T\mu$ be the probability measure at the end of the first unit interval. It is then proved that as $n \rightarrow \infty$, $T^n\mu$ (that is, the probability distribution for the state of the system at time n) converges in distribution to a distribution concentrated at the points 0, 1 and the form of this limiting distribution which depends on μ is obtained.

The motivation for the consideration of such a system arose from certain learning models introduced by Bush and Mosteller. These are described in detail in their recent book [2]. (Condition (1.2) means that the state of the system may be identified with the probability of applying A_1).

The methods used in [1] to obtain the convergence of $T^n\mu$ are probabilistic. Karlin [9] considers the space of continuous functions on the unit interval and obtains a bounded operator U on this space whose adjoint is T . A convergence theorem is obtained for U^n and the result is translated into the adjoint space (that is, the space of measures) to

Received June 5, 1956. This paper generalizes the main theorem of the author's thesis presented at the California Institute of Technology 1954, being partial requirements for the degree of Doctor of Philosophy. The author wishes to express his sincere thanks to Professor Samuel Karlin for his help and guidance in the preparation of that thesis.

¹ Karlin also considers boundary cases where λ_0, λ_1 may be 1.

obtain the required result.

Karlin [9] also considers cases where (1.2) no longer holds and obtains for a wide class of non-absorbing models the convergence of $T^m \mu$ to a distribution which is independent of μ . These do not concern us here as the object is to consider only a class of absorbing problems, where of course the final distribution depends on the initial distribution.

We conclude this section by stating a well-known theorem [8].

THEOREM 1.1. *Let Ω be a compact Hausdorff space and let $\mathfrak{C}(\Omega)$ denote the Banach space of real-valued continuous functions $x(t)$ defined on Ω with*

$$(1.3) \quad \|x\| = \max_{t \in \Omega} |x(t)| .$$

Let $\mathfrak{M}(\Omega)$ denote the space of all real-valued completely additive regular set functions $\mu(E)$ defined for all Borel sets E of Ω , with

$$(1.4) \quad \|\mu\| = \sup_{E \subset \Omega} \mu(E) - \inf_{E \subset \Omega} \mu(E) .$$

Then $\mathfrak{M}(\Omega)$ is isometric (and lattice isomorphic) to the conjugate space of $\mathfrak{C}(\Omega)$, the correspondence being given by

$$(1.5) \quad (x, \mu) = \int_{\Omega} x(t) d\mu(t) .$$

2. Description of the process. Let Ω be a compact metric space with metric ρ . Since Ω satisfies the second axiom of countability, the concepts of Baire and Borel measures coincide, and thus since the former are always regular [5], we have that the set $\mathfrak{M}(\Omega)$ of Theorem 1.1 consists of all the completely-additive (finite) set functions defined on the Borel sets of Ω .

Let $\{\tau_i\}$ be a countable sets of points in Ω and $\{A_i\}$ a corresponding set of continuous transformations of Ω into itself with the following properties

$$(2.1) \quad A_i S_i \subset S_i \quad i=1, 2, \dots$$

where S_i is any open sphere with centre τ_i , and

$$(2.2) \quad \lim_{n \rightarrow \infty} A_i^n t = \tau_i \quad i=1, 2, \dots$$

for each $t \in \Omega$;

that is, repeated applications of the transformation A_i transforms t in the limit into τ_i and moreover every open sphere with centre τ_i is mapped by A_i into itself. The points $\{\tau_i\}$ will be referred to as boundary

points. It follows from (2.2) and the continuity of A_i , that

$$(2.3) \quad A_i \tau_i = \tau_i \quad i=1, 2, \dots .$$

Consider a system whose state may be described by a point t in Ω . Let $\{\phi_i(t)\}$ be a countable family of continuous functions defined on Ω with the property that

$$(2.4) \quad 0 \leq \phi_i(t) \leq 1, \quad i=1, 2, \dots$$

and

$$(2.5) \quad \sum_1^\infty \phi_i(t) = 1 .$$

Let $\mu(E)$ be a probability measure defined on the Borel sets of Ω , giving the probability distribution of the initial state of the system. Our process consists in applying at every unit interval of time one of the transformations $\{A_i\}$, A_i being applied with probability $\phi_i(t)$, where $t \in \Omega$ represents the state of the system.

Let

$$(2.6) \quad T\mu(E) = \sum_{i=1}^\infty \int_{A_i^{-1}E} \phi_i(t) d\mu(t) .$$

It is easily seen that $T\mu(E)$ is a Borel measure. It represents the probability measure for the state of the system after unit time. $T\mu$ is defined by (2.6) for any Borel measure μ and

$$(2.7) \quad \|T\mu\| = T\mu(\Omega) = \mu(\Omega) = \|\mu\| .$$

More generally, if

$$\mu \in \mathfrak{M}(\Omega), \quad \mu = \mu^+ - \mu^- ,$$

(2.6) defines $T\mu$ and

$$(2.8) \quad T\mu = T\mu^+ - T\mu^- .$$

T is a linear transformation of $\mathfrak{M}(\Omega)$ into itself and

$$\|T\mu\| = \|T\mu^+ - T\mu^-\| \leq \|T\mu^+\| + \|T\mu^-\| = \|\mu^+\| + \|\mu^-\| = \|\mu\| .$$

Thus we obtain

LEMMA 2.1. *T is a positive linear transformation of $\mathfrak{M}(\Omega)$ into itself of norm 1.*

Now consider $x(t) \in \mathfrak{C}(\Omega)$ (cf. Theorem 1.1). A function $Ux(t)$ is defined on Ω by

$$(2.9) \quad Ux(t) = \sum_{i=1}^{\infty} \phi_i(t)x(A_it) .^2$$

Each term of this series is continuous on Ω and $|x(A_it)| \leq \|x\|$.

Since $\sum \phi_i(t) = 1$ the convergence being uniform (by Dini's theorem), the series (2.9) is uniformly convergent and hence $Ux(t) \in \mathfrak{C}(\Omega)$. Clearly U is a linear transformation of $\mathfrak{C}(\Omega)$ into itself and $\|Ux\| \leq \|x\|$. Thus, since the functions which are constant on Ω , are fixed points of U , we have the following.

LEMMA 2.2. *U is a bounded positive linear transformation of $\mathfrak{C}(\Omega)$ into itself, for which the constant functions are fixed points. Moreover $\|U^n\| = 1$ for all positive integers n .*

Theorem 1.1 connects $\mathfrak{C}(\Omega)$ and $\mathfrak{M}(\Omega)$. We now prove

LEMMA 2.3. *T is the adjoint of U , that is,*

$$(2.10) \quad (Ux, \mu) = (x, T\mu), \text{ for each } x \in \mathfrak{C}(\Omega) \text{ and } \mu \in \mathfrak{M}(\Omega).$$

Since $\mu = \mu^+ - \mu^-$, it is clearly sufficient to prove (2.10) for the case $\mu \geq 0$. Let

$$\nu_i(E) = \int_E \phi_i(t) d\mu(t)$$

It is easy to see that

$$T\mu = \sum_1^{\infty} \nu_i A_i^{-1},$$

the convergence being in the sense of $\mathfrak{M}(\Omega)$. Hence

$$\begin{aligned} (x, T\mu) &= \sum_1^{\infty} (x, \nu_i A_i^{-1}) = \sum_1^{\infty} \int x(t) d\nu_i A_i^{-1}(t) \\ &= \sum_1^{\infty} \int x(A_it) d\nu_i(t) \\ &= \sum_1^{\infty} \int \phi_i(t)x(A_it) d\mu(t) \\ &= (Ux, \mu) \end{aligned}$$

since the series (2.9) converges uniformly³.

² Operators of the type U have been considered, and both convergence and $\mathfrak{C}-1$ convergence theorems for the iterates U^n obtained by Ocinescu, Mihoc, Doebelin, Fortet, Ionescu Tulcea and Marinescu [10, 3, 4, 6, 7].

³ This adjointness lemma expresses the fact that if t_1, t_2, \dots represents the process then $E\{E\{x(t_2)|t_1\}\} = E\{x(t_2)\}$.

3. **Absorption assumptions.**⁴ The first additional assumption to be made is that each of a finite number of the boundary points is an absorbing point, that is, we assume

$$(3.1) \quad \phi_i(\tau_i)=1 \quad i=1, 2, \dots, m.$$

This together with (2.3) ensure that τ_i ($i=1, 2, \dots, m$) are absorbing points. (Since $\sum \phi_i(t)=1$ and Ω is compact, it is not possible to extend the assumption (3.1) to an infinite number of the boundary points τ_i). The assumption (3.1) is strengthened as follows :

We assume that about each absorbing point τ_i ($1 \leq i \leq m$), an open sphere \sum_i may be drawn with centre τ_i on which the infinite product

$$(3.2) \quad \phi_i(t)\phi_i(A_it)\cdots\phi_i(A_i^n t)\cdots$$

converges uniformly (the convergence being in the sense of infinite products that is, the limit is nonzero).

Clearly assumption (3.2) together with (2.2) imply (3.1). Finally, the assumption is made that for each $t \in \Omega - \bigcup_1^m \sum_i$ there is a finite sequence of transformations

$$A_{j_1}, A_{j_2}, \dots, A_{j_n} \quad (1 \leq j_k < \infty)$$

where n, j_1, j_2, \dots, j_n depend on t , such that $A_{j_n}A_{j_{n-1}}\cdots A_{j_1}t$ is in one of the spheres \sum_i ($1 \leq i \leq m$) and such that each term of the sequence

$$(3.3) \quad \phi_{j_1}(t), \phi_{j_2}(A_{j_1}t), \dots, \phi_{j_n}(A_{j_{n-1}}\cdots A_{j_1}t)$$

is greater than zero.

Assumptions (3.2) and (3.3) imply that no matter what the initial state of the system there is always positive probability of reaching an absorbing point after an infinite number of steps. We conclude this section with the following lemma which is a consequence of (3.1).

LEMMA 3.1. *U preserves the values at the absorbing points, that is,*

$$(3.4) \quad Ux(\tau_i)=x(\tau_i), \quad i=1, 2, \dots, m$$

where $x(t) \in \mathbb{C}(\Omega)$.

Proof. Since

$$\sum_1^\infty \phi_i(t)=1 \quad \text{and} \quad \phi_i(t) \geq 0$$

⁴ Cf. discussion Bush and Mosteller, [2, pp. 167-169].

for each l , we have by assumption (3.1) that

$$(3.5) \quad \phi_i(\tau_i) = 0 \quad l \neq i, \quad 1 \leq i \leq m, \quad 1 \leq l < \infty.$$

The result follows by (2.3) from the definition (2.9) of U .

4. Examples.

EXAMPLE 4.1. Let $\{A_i\}$ be a countable set of transformations of Ω into itself with the property that

$$(4.1) \quad \rho(A_i t, A_i s) \leq \lambda \rho(t, s) \quad i = 1, 2, \dots$$

for all pairs of points $t, s \in \Omega$, where λ is a constant such that $0 \leq \lambda < 1$.

It follows from (4.1) that the transformations $\{A_i\}$ are continuous and moreover there exist points $\{\tau_i\}$ such that (2.1) and (2.2) are satisfied.

Let $\{\phi_i(t)\}$ be a family of continuous functions on Ω satisfying the conditions (2.4), (2.5) and the first absorption assumption (3.1). Suppose also for each i ($1 \leq i \leq m$) that there exists an open sphere \sum_i with centre τ_i and radius r_i on which $\phi_i(t) > 0$ and satisfies a uniform Lipschitz condition⁵ that is,

$$(4.2) \quad |\phi_i(t) - \phi_i(s)| \leq k \rho(t, s) \quad t, s \in \sum_i$$

Finally the assumption is made that one of the probability functions, say $\phi_1(t)$, satisfies

$$(4.3) \quad \phi_1(t) > 0 \quad \text{except at the points } \tau_i \quad (2 \leq i \leq m).$$

LEMMA 4.1. *The process just described satisfies the absorption assumptions of § 3.*

Proof. We first observe that (3.1) is satisfied by hypothesis. To establish (3.2), let $t \in \sum_i$ ($1 \leq i \leq m$) and $\psi_i(t) = 1 - \phi_i(t)$.

$$\begin{aligned} \psi_i(A_i^n t) &= 1 - \phi_i(A_i^n t) \\ &= \phi_i(A_i^n \tau_i) - \phi_i(A_i^n t) \quad \text{by (2.3) and (3.1)} \\ &\leq k \rho(A_i^n \tau_i, A_i^n t) \quad \text{by (4.2)} \\ &\leq k \lambda^n \rho(\tau_i, t) \quad \text{by (4.1)} \\ &\leq (k r_i) \lambda^n. \end{aligned}$$

Since $0 \leq \lambda < 1$, $\sum \lambda^n$ converges and hence, by a theorem on infinite

⁵ These assumptions link up with those given by other authors [10, 3, 4, 6, 7].

products $\prod_{n=0}^{\infty} (1 - \psi_i(A_i^n t))$ converges uniformly on \sum_i . Thus $\prod_{n=0}^{\infty} \phi_i(A_i^n t)$ converges uniformly on \sum_i and the assumption (3.2) is verified.

It remains to verify (3.3). Let $t \in \Omega - \bigcup_1^m \sum_i$. Since $A_1^n t \rightarrow \tau_1$, there exists n_0 such that $A_1^{n_0} t \in \sum_1$. By (4.3), $\phi_1(t) > 0$. Hence we take A_1 as our first transformation. If $A_1 t \in \bigcup_1^m \sum_i$, then (3.3) is already verified. If not $\phi_1(A_1 t) > 0$ and we take A_1 as our second transformation. Proceeding in this manner a finite sequence of A_1 's (of length $\leq n_0$) is obtained which satisfies the assumption (3.3). Hence the lemma is proved.

EXAMPLE 4.2. The example described in § 1 is a particular case of the example just given.

EXAMPLE 4.3. We now consider a generalization from 1 to N dimensions of the learning model considered by Karlin (cf. Bush and Mosteller [2]).

Let Ω be a simplex in E_N (Euclidean space of N dimensions). Any point of Ω is given by its barycentric coordinates $t = (t_1, t_2, \dots, t_{N+1})$ where $t_i \geq 0$ and $\sum_1^{N+1} t_i = 1$. The vertices e^i ($i = 1, \dots, N+1$) have coordinates $e_j^i = \delta_j^i$ (Kronecker delta). Let I denote the $(N+1) \times (N+1)$ unit matrix, and B_i ($1 \leq i \leq N+1$) denote the $(N+1) \times (N+1)$ projection matrix where each element of the i th row is unity and all other elements are zero. Clearly $B_i t = e^i$ for each $t \in \Omega$. Consider the family $\{A_i\}$ of transformations on Ω into itself defined as follows

$$(4.4) \quad A_i = \lambda_i I + (1 - \lambda_i) B_i, \quad 0 \leq \lambda_i < 1, \quad i = 1, 2, \dots, N+1,$$

that is, for $t \in \Omega$

$$A_i t = (\lambda_i t_1, \lambda_i t_2, \dots, \lambda_i t_i + 1 - \lambda_i, \lambda_i t_{i+1}, \dots, \lambda_i t_{N+1})$$

Clearly A_i represents a transformation which carries a point P into a point P' on the line PV_i where V_i is the vertex e^i and

$$P'V_i = \lambda_i(PV_i).$$

The transformations $\{A_i\}$ are continuous and satisfy the conditions (2.1) and (2.2) where $\tau_i = e^i$. For the probabilities $\phi_i(t)$ we take

$$(4.5) \quad \phi_i(t) = t_i \quad i = 1, \dots, N+1.$$

The conditions (2.4) and (2.5) are clearly satisfied. It remains to verify the absorption assumptions of § 3. Since $\phi_i(e^i) = e_i^i = 1$, the condition (3.1) is satisfied. To verify (3.2) we first note that since $B_i^2 = B_i$

$$A_i^n = \lambda_i^n I + (1 - \lambda_i^n) B_i$$

so that

$$\phi_i(t) = t_i, \quad \phi_i(A_i^n t) = \lambda_i^n t_i + (1 - \lambda_i^n)$$

If $t_i > \epsilon$, it is easily seen that the infinite product $\prod_0^\infty \phi_i(A_i^n t)$ converges uniformly.

Condition (3.3) is seen to be satisfied by noting that for any point t , one at least of the coordinates is nonzero, say t_i , and hence the i th coordinate of $A_i^n t$ ($n=1, 2, \dots$) is also nonzero.

5. Returning to the general absorption process described in §§ 2 and 3, we establish by means of the assumption (3.2), the equicontinuity of the family of functions $\{U^n x(t)\}$ at each of the absorbing points τ_i ($i=1, 2, \dots, m$).

LEMMA 5.1. *Let $x(t) \in \mathcal{C}(\Omega)$ be such that it vanishes at one of the absorbing points τ_i ($1 \leq i \leq m$), then for each $\epsilon > 0$, there exists a sphere $S_i(\epsilon)$ with centre τ_i , such that*

$$|U^n x(t)| < \epsilon \qquad n=1, 2, \dots$$

for $t \in S_i(\epsilon)$.

Proof. Without loss of generality we consider the case where $i=1$. Let

$$f_n(t) = \phi_1(t) \phi_1(A_1 t) \dots \phi_1(A_1^{n-1} t).$$

$\{f_n(t)\}$ form a nonincreasing sequence of functions which by assumption (3.2) converges uniformly on \sum_1 to a function $f(t)$. It follows that $f(t)$ is continuous, and thus since $f_n(\tau_1) = 1$ (by (2.3) and (3.1)) and therefore $f(\tau_1) = 1$, we have that given any positive number δ ($0 < \delta < 1$), there exists a neighbourhood V of τ_1 (contained in \sum_1) on which $f(t) > \delta$, which implies $f_n(t) > \delta$ for all n .

Choose $\delta > 1 - \epsilon / \|x\|$ and let q be a positive integer such that

$$(5.1) \qquad \frac{q-1}{q} \delta > 1 - \epsilon / \|x\|.$$

Since $x(\tau_1) = 0$ by hypothesis, there exists a neighbourhood V' of τ_1 such that for $t \in V'$

$$(5.2) \qquad |x(t)| \leq \frac{1}{q} \|x\|.$$

Let $S_1(\epsilon)$ be an open sphere with centre τ_1 and such that $S_1(\epsilon) \subset V \cap V'$.

By (2.1) if $t \in S_1(\epsilon)$, $A_1^n t \in S_1(\epsilon)$ for all positive integers n . Hence for $t \in S_1(\epsilon)$

$$(5.3) \quad |x(A_1^n t)| \leq \frac{1}{q} \|x\| \quad n=1, 2, \dots,$$

and

$$(5.4) \quad f_n(t) > \delta \quad n=1, 2, \dots$$

Now

$$U^n x(t) = \sum_{i_1, i_2, \dots, i_n} \phi_{i_1}(t) \phi_{i_2}(A_{i_1} t) \dots \phi_{i_n}(A_{i_{n-1}} \dots A_{i_1} t) x(A_{i_n} \dots A_{i_1} t),$$

$$|U^n x(t)| \leq \phi_1(t) \phi_1(A_1 t) \dots \phi_1(A_1^{n-1} t) |x(A_1^n t)|$$

$$+ \|x\| \sum' \phi_{i_1}(t) \phi_{i_2}(A_{i_1} t) \dots \phi_{i_n}(A_{i_{n-1}} \dots A_{i_1} t),$$

where \sum' denotes the summation omitting the term corresponding to

$$i_1=1, i_2=1, \dots, i_n=1.$$

Replacing $|x(A_1^n t)|$ by $-(\|x\| - |x(A_1^n t)|) + \|x\|$ we obtain

$$|U^n x(t)| \leq -f_n(t)(\|x\| - |x(A_1^n t)|)$$

$$+ \|x\| \sum \phi_{i_1}(t) \phi_{i_2}(A_{i_1} t) \dots \phi_{i_n}(A_{i_{n-1}} \dots A_{i_1} t),$$

or

$$(5.5) \quad |U^n x(t)| \leq -f_n(t)(\|x\| - |x(A_1^n t)|) + \|x\|,$$

since $\sum \phi_i(t) = 1$. Now let $t \in S_1(\epsilon)$. Then

$$|U^n x(t)| \leq \|x\| - \delta(\|x\| - |x(A_1^n t)|) \quad \text{by (5.4)}$$

$$\leq \|x\| - \delta \left(\|x\| - \frac{1}{q} \|x\| \right) \quad \text{by (5.3)}$$

$$= \|x\| \left(1 - \frac{q-1}{q} \delta \right)$$

$$< \epsilon \quad \text{by (5.1)}$$

Hence the lemma is proved.

THEOREM 5.1. *If $x(t) \in \mathfrak{C}(\Omega)$, then $\{U^n x(t)\}$ form an equicontinuous family of functions at each of the absorbing points $\{\tau_i\}$ ($i=1, 2, \dots, m$).*

Proof. Without loss of generality, we prove the theorem for the point τ_1 . It is required to prove that given $\varepsilon > 0$, there exists a sphere $S_1(\varepsilon)$ with centre τ_1 such that for $t \in S_1(\varepsilon)$

$$|U^n x(t) - U^n x(\tau_1)| < \varepsilon \quad n=0, 1, 2, \dots,$$

or equivalently by Lemma 3.1

$$(5.5) \quad |U^n x(t) - x(\tau_1)| < \varepsilon \quad n=0, 1, 2, \dots,$$

for $t \in S_1(\varepsilon)$.

Let $z(t) = x(t) - x(\tau_1)$. $z(t) \in \mathfrak{C}(\Omega)$ and $z(\tau_1) = 0$. Hence Lemma 5.1 may be applied to obtain a sphere $S_1(\varepsilon)$ with centre τ_1 on which

$$|U^n z(t)| < \varepsilon \quad n=0, 1, 2, \dots,$$

but since U preserves constant functions (by Lemma 2.2)

$$U^n z(t) = U^n x(t) - x(\tau_1).$$

Hence (5.5) is established and the theorem is proved.

6. The convergence theorem in $\mathfrak{C}(\Omega)$. In this section, the assumption (3.3) is applied in conjunction with Theorem 5.1 to obtain the convergence of $U^n x(t)$ in $\mathfrak{C}(\Omega)$.

LEMMA 6.1. *Let $\{S_i\}$ be spheres with centres $\{\tau_i\}$ such that $\overline{S_i} \subset \sum_i$ ($i=1, 2, \dots, m$). Then there exists a positive integer n_0 and a number δ , ($0 < \delta \leq 1$) such that for each $t \in \Omega$, there exists a sequence of n_0 transformations $A_{i_1}, A_{i_2}, \dots, A_{i_{n_0}}$ (depending on t) which, when applied consecutively, transform t into a point in $\bigcup_{i=1}^m S_i$, and such that the probability of the application of each transformation of the sequence is $\geq \delta$, that is, each term in the finite sequence*

$$\phi_{i_1}(t), \phi_{i_2}(A_{i_1}t), \dots, \phi_{i_{n_0}}(A_{i_{n_0-1}} \dots A_{i_1}t)$$

is $\geq \delta$.

Proof. By assumption (3.2), it is clear that

$$(6.1) \quad \phi_i(t) > 0 \quad \text{on } \sum_i \text{ (and hence on } \overline{S_i}) \quad i=1, 2, \dots, m.$$

and thus by the continuity of $\phi_i(t)$, there exists δ_0 such that $0 < \delta_0 \leq 1$ and

$$(6.2) \quad \phi_i(t) \geq \delta_0 \quad \text{for } t \in \bar{S}_i \quad i=1, 2, \dots, m$$

Let $t \in \Omega$. If $t \in \Omega - \bigcup_1^m S_i$, we have by the assumption (3.3) together with (2.1), (2.2) and (6.1) that there exists a finite chain of transformations $A_{j_1}, A_{j_2}, \dots, A_{j_n}$, which when applied consecutively, transform t into a point in one of the spheres, say S_l ($1 \leq l \leq m$). Moreover each term in the sequence $\phi_{j_1}(t), \phi_{j_2}(A_{j_1}t), \dots, \phi_{j_n}(A_{j_{n-1}} \cdots A_{j_1}t)$ is > 0 . If $t \in \bigcup_1^m S_i$, the same result holds for then $t \in S_l$ ($1 \leq l \leq m$) and thus by (2.1) and (6.1), it is sufficient to take a chain consisting of the single transformation A_l .

Consider $A_{j_n}^{-1}S_l$. This is an open set containing $A_{j_{n-1}} \cdots A_{j_1}t$. Since $\phi_{j_n}(A_{j_{n-1}} \cdots A_{j_1}t) > 0$, there exists an open set U_n such that

$$A_{j_{n-1}} \cdots A_{j_1}t \in U_n \subset A_{j_n}^{-1}S_l.$$

and on which $\phi_{j_n}(t) > 0$. By the regularity of Ω , there exists an open set V_n such that

$$A_{j_{n-1}} \cdots A_{j_1}t \in V_n \subset \bar{V}_n \subset U_n \subset A_{j_n}^{-1}S_l.$$

\bar{V}_n is compact and therefore there exists a positive number δ_n such that $\phi_{j_n}(t) \geq \delta_n$ on \bar{V}_n and hence in particular on V_n .

Now consider $A_{j_{n-1}}^{-1}V_n$. Proceeding as above, we obtain an open set V_{n-1} , such that

$$A_{j_{n-2}} \cdots A_{j_1}t \in V_{n-1} \subset \bar{V}_{n-1} \subset A_{j_{n-1}}^{-1}V_n$$

and a positive number δ_{n-1} , such that $\phi_{j_{n-1}}(t) \geq \delta_{n-1}$ on V_{n-1} . Proceeding in this manner, we arrive at an open set V_1 which is such that

$$t \in V_1 \subset \bar{V}_1 \subset A_{j_1}^{-1}V_2$$

and such that $\phi_{j_1}(t) \geq \delta_1$ on V_1 , where $\delta_1 > 0$.

Hence, the open set V_1 containing t has the property that each point in it is transformed by the sequence $A_{j_1}, A_{j_2}, \dots, A_{j_n}$ into a point of S_l ($\subset \bigcup_1^m S_i$) and the conditional probabilities of each of the successive transformations being applied are $\geq \delta_1, \delta_2, \dots, \delta_n$ respectively.

This process is repeated for every $t \in \Omega$. For each t an open set corresponding to V_1 is obtained. By the compactness of Ω , we have that Ω is covered by a finite number of open sets Ω_i ($i=1, 2, \dots, k$), where each set Ω_i has the property that there is a finite chain of transformations of length n_i (that is, n_i =the number of transformations in the chain) which when applied successively transform each point of Ω_i into one of the sphere S_i ($i=1, 2, \dots, m$) say S_i , and which has the

property that the conditional probabilities of applying the transformations of the chain are respectively $\geq \delta_{i_1}, \delta_{i_2}, \dots, \delta_{i_{n_l}}$, where each of these numbers is greater than zero.

Let $n_0 = \max_{1 \leq l \leq k} n_l$. The length of the chain n_l for each Ω_l may be extended to n_0 preserving the above properties. For if $t \in S_l$ ($1 \leq l \leq m$), $A, t \in S_l$ by (2.1) and $\phi_i(t) \geq \delta_0$ by (6.2).

Let

$$\delta = \min_{\substack{1 \leq i \leq n_l \\ 1 \leq l \leq k}} (\delta_{i_l}, \delta_0).$$

With these values of n_0 and δ , the lemma is established.

LEMMA 6.2. *Let $\{x_p(t)\}$ be a sequence of functions in $\mathfrak{C}(\Omega)$ with the following properties*

$$(6.3) \quad \|x_p\| \leq H \quad p=0, 1, 2, \dots$$

where H is a constant.

$$(6.4) \quad x_p(\tau_i) = 0 \text{ for all } p \quad i=1, 2, \dots, m$$

and

(6.5) *the family of functions $\{U^n x_p(t)\}$ ($n, p=0, 1, 2, \dots$) is equicontinuous at each of the absorbing points τ_i ($i=1, 2, \dots, m$).*

Then, under these conditions

$$\lim_{n \rightarrow \infty} \|U^n x_p\| = 0$$

where the convergence is uniform with respect to p .

Proof. Given $\epsilon > 0$, there exist by (6.5) spheres $S_i(\epsilon)$ with centres τ_i ($i=1, 2, \dots, m$) such that for $t \in S_i(\epsilon)$

$$|U^n x_p(t) - U^n x_p(\tau_i)| < \epsilon/2 \quad \text{all } n, p.$$

Hence by (6.4) and Lemma (3.1)

$$(6.6) \quad |U^n x_p(t)| < \epsilon/2, \quad t \in \bigcup_1^m S_i(\epsilon), \text{ all } n, p.$$

There is no loss in generality in assuming the spheres $S_i(\epsilon)$ so chosen that

$$(6.7) \quad \overline{S_i(\epsilon)} \subset \Sigma_i \quad i=1, 2, \dots, m.$$

Thus the spheres $S_i(\epsilon)$ ($i=1, 2, \dots, m$) satisfy the hypothesis of Lemma 6.1. The positive integer n_0 and the positive number δ obtained

in the lemma depend here on ε . Let

$$(6.8) \quad \alpha = 1 - \frac{1}{2} \delta^{n_0}.$$

Since $0 < \delta \leq 1$, it follows that $0 < \alpha < 1$. We now show that for all p

$$(6.9) \quad \|U^{kn_0} x_p\| \leq \mu_k \quad k=0, 1, 2, \dots$$

where $\mu_k = \max(\alpha^k H, \varepsilon)$.

We prove (6.9) by induction. Clearly by (6.3) it is true for $k=0$. Suppose it is true for k .

$$(6.10) \quad \begin{aligned} U^{(k+1)n_0} x_p(t) &= \sum_{i_1 i_2 \dots i_{n_0}} \phi_{i_1}(t) \phi_{i_2}(A_{i_1} t) \dots \phi_{i_{n_0}}(A_{i_{n_0-1}} \dots A_{i_1} t) \\ &\quad \times U^{kn_0} x_p(A_{i_{n_0}} \dots A_{i_1} t). \end{aligned}$$

Consider t fixed. By Lemma 6.1, there is associated with t a finite sequence of n_0 transformations $A_{j_1}, A_{j_2}, \dots, A_{j_{n_0}}$ (depending on t) which when applied consecutively transform t into $\bigcup_{i=1}^m S_i(\varepsilon)$ and such that each term of the finite sequence $\phi_{j_1}(t), \phi_{j_2}(A_{j_1} t), \dots, \phi_{j_{n_0}}(A_{j_{n_0-1}} \dots A_{j_1} t)$ is $\geq \delta$, that is,

$$(6.11) \quad A_{j_{n_0}} A_{j_{n_0-1}} \dots A_{j_1} t \in \bigcup_1^m S_i(\varepsilon)$$

and

$$(6.12) \quad \phi_{j_1}(t) \phi_{j_2}(A_{j_1} t) \dots \phi_{j_{n_0}}(A_{j_{n_0-1}} \dots A_{j_1} t) \geq \delta^{n_0}.$$

In (6.10) we take inequalities with absolute values and separate out the term corresponding to the above sequence and proceed as in § 5 (between the relations (5.4) and (5.5)) to obtain by the induction hypothesis

$$(6.13) \quad \begin{aligned} |U^{(k+1)n_0} x_p(t)| &\leq -\phi_{j_1}(t) \phi_{j_2}(A_{j_1} t) \dots \phi_{j_{n_0}}(A_{j_{n_0-1}} \dots A_{j_1} t) \\ &\quad \times (\mu_k - |U^{kn_0} x_p(A_{j_{n_0}} \dots A_{j_1} t)|) + \mu_k. \end{aligned}$$

(6.6), (6.11) and (6.12) give

$$|U^{(k+1)n_0} x_p(t)| \leq -\delta^{n_0}(\mu_k - \varepsilon/2) + \mu_k.$$

Since $\mu_k \geq \varepsilon$, we have $\mu_k - \varepsilon/2 \geq 1/2\mu_k$. Hence

$$|U^{(k+1)n}x_p(t)| \leq \mu_k \left(1 - \frac{1}{2}\delta^{n_0}\right) = \alpha\mu_k \quad \text{by (6.8).}$$

Therefore

$$\|U^{(k+1)n}x_p\| \leq \alpha\mu_k \leq \max(\alpha^{k+1}H, \epsilon) = \mu_{k+1}.$$

Hence (6.9) is established. Clearly there exists k_0 sufficiently large such that $\mu_{k_0} \leq \epsilon$. Then, since $\|U^n\|=1$, all n (Lemma 2.2), we have that for $n \geq n_0k_0$

$$\|U^n x_p\| \leq \epsilon \quad \text{all } p.$$

Hence the lemma is proved.

THEOREM 6.1. *U^n converges strongly on $\mathfrak{C}(\Omega)$, that is, there exists a continuous transformation U_∞ of norm 1 of $\mathfrak{C}(\Omega)$ into itself (which preserves constant function) such that*

$$(6.14) \quad \lim_{n \rightarrow \infty} \|U^n x - U_\infty x\| = 0$$

for each function $x \in \mathfrak{C}(\Omega)$.

Proof. Given $x(t) \in \mathfrak{C}(\Omega)$, let

$$(6.15) \quad x_p(t) = U^p x(t) - x(t) \quad p=0, 1, 2, \dots$$

Clearly $x_p(t) \in \mathfrak{C}(\Omega)$ and $\|x_p\| \leq 2\|x\|$. Moreover by Lemma 3.1, $x_p(\tau_i) = 0$ ($i=1, 2, \dots, m$). Hence the hypothesis (6.3) and (6.4) of Lemma (6.2) are verified for the family $\{x_p(t)\}$. It remains to verify (6.5).

Given $\epsilon > 0$, we have by Theorem 5.1 that there exists spheres $S_i(\epsilon)$ with centres τ_i ($i=1, 2, \dots, m$) such that for $n=0, 1, 2, \dots$ we have

$$|U^n x(t) - x(\tau_i)| < \epsilon/2, \quad t \in S_i(\epsilon), \quad i=1, 2, \dots, m.$$

Hence for $t \in S_i(\epsilon)$ ($1 \leq i \leq m$) and all n, p

$$|U^{n+p}x(t) - U^n x(t)| < \epsilon$$

or

$$(6.16) \quad |U^n x_p(t)| < \epsilon \quad \text{all } n, p, t \in \bigcup_1^m S_i(\epsilon).$$

Since $x_p(\tau_i) = 0$, $U^n x_p(\tau_i) = 0$ (Lemma 3.1) and thus it is clear from (6.16) that the hypothesis (6.5) of Lemma (6.2) is verified.

Hence applying Lemma 6.2 to the family $x_p(t)$ as defined by (6.15) we have that given $\epsilon > 0$, there exists n such that

$$\|U^n x_p\| < \epsilon \quad p=0, 1, 2, \dots,$$

or

$$\|U^{n+p}x - U^n x\| < \epsilon \text{ for all positive integers } p.$$

Hence since the space $\mathfrak{C}(\Omega)$ is complete, there exists an element $U_\infty x \in \mathfrak{C}(\Omega)$ such that

$$\lim_{n \rightarrow \infty} \|U^n x - U_\infty x\| = 0.$$

U_∞ is clearly a linear transformation of $\mathfrak{C}(\Omega)$ into itself. Since

$$\|U_\infty x\| = \lim_{n \rightarrow \infty} \|U^n x\| \leq \|x\|,$$

it follows that U_∞ is continuous and $\|U_\infty\| \leq 1$. However since U preserves the constant functions on Ω , it is clear that U_∞ does likewise and hence $\|U_\infty\| = 1$. Hence the theorem is proved.

7. The form of $U_\infty x$. The following lemma is a direct consequence of Lemma 3.1 and Theorem 6.1.

LEMMA 7.1. U_∞ preserves the values at the absorbing points, that is,

$$U_\infty x(\tau_i) = x(\tau_i) \quad i=1, 2, \dots, m$$

where $x(t) \in \mathfrak{C}(\Omega)$.

LEMMA 7.2. If $x(t)$ is a fixed point of U in $\mathfrak{C}(\Omega)$ having the value zero at each of the absorbing points τ_i ($i=1, 2, \dots, m$), then $x(t) \equiv 0$. Two continuous fixed points of U which are equal at each τ_i ($i=1, 2, \dots, m$) are identical.

Proof. Let $x(t)$ be a fixed point of U with $x(\tau_i) = 0$ ($i=1, 2, \dots, m$). We apply Lemma 6.2 to the family of functions consisting of the single function $x(t)$. Since $U^n x = x$ all n , the conditions of the lemma are trivially satisfied and hence $\lim_{n \rightarrow \infty} \|U^n x\| = 0$, that is, $\|x\| = 0$. Therefore the first part of the lemma is proved.

If $x(t), y(t)$ are two fixed points in $\mathfrak{C}(\Omega)$ such that $x(\tau_i) = y(\tau_i)$ ($i=1, 2, \dots, m$) then, applying the first part of the lemma to the function $z(t) = x(t) - y(t)$, we obtain $z(t) \equiv 0$. Hence the lemma is proved.

LEMMA 7.3. Let $\psi_i(t) = U_\infty \phi_i(t)$ ($1 \leq i < \infty$). Then $\psi_i(t)$ is a fixed point of U in $\mathfrak{C}(\Omega)$. If $i > m$, $\psi_i(t) \equiv 0$. If $i \leq m$, $\psi_i(t)$ is the unique fixed point of U having the value 1 at τ_i and the value zero at each of the other absorbing points τ_j ($j \neq i, 1 \leq j \leq m$). Moreover

$$(7.1) \quad \sum_1^m \psi_i(t) = 1.$$

Proof.
$$U\psi_i = UU_\infty\phi_i = U_\infty\phi_i = \psi_i \quad (1 \leq i < \infty)$$

since $UU_\infty = U_\infty$ by Theorem 6.1. Hence ψ_i is a fixed point of U . For $i > m$, $\phi_i(t)$ has the value zero at each of the absorbing points τ_j ($1 \leq j \leq m$) (by 3.5) and hence by Lemma 7.1 $\psi_i(t)$ has the same property and thus, by Lemma 7.2, is identically zero. If $1 \leq i \leq m$, then since $\phi_i(t)$ has the value 1 at τ_i (by (3.1)) and the value zero at each τ_j ($j \neq i, 1 \leq j \leq m$) by (3.5), we have by Lemma 7.1 that ψ_i has the same properties and hence since ψ_i is a fixed point of U , by Lemma 7.2 it is the unique fixed point with these values at the vertices.

By (2.5) $\sum_1^\infty \phi_i(t) = 1$. By Dini's theorem, the convergence is uniform so that we have in the sense of $\mathfrak{C}(\Omega)$, $\sum_1^\infty \phi_i = 1$. Hence since U_∞ is continuous $\sum_1^\infty \psi_i = 1$, and since $\psi_i(t) = 0$ ($i > m$), $\sum_1^m \psi_i(t) = 1$ and the lemma is established.

THEOREM 7.1. *If $x \in \mathfrak{C}(\Omega)$ then*

$$U_\infty x = \sum_{i=1}^m x(\tau_i)\psi_i.$$

Proof. Let

$$y(t) = \sum_{i=1}^m x(\tau_i)\psi_i(t).$$

Clearly by Lemma 7.3, y is a fixed point of U such that $y(\tau_i) = x(\tau_i)$ ($i = 1, \dots, m$). By Theorem 6.1, $U_\infty x$ is a fixed point of U and by Lemma 7.1 $U_\infty x(\tau_i) = x(\tau_i)$ ($i = 1, \dots, m$). Hence by Lemma 7.2, $y = U_\infty x$ and the theorem is proved.

8. The convergence theorem in $\mathfrak{M}(\Omega)$.

THEOREM 8.1. *Let*

$$\mu \in \mathfrak{M}(\Omega),$$

then

$$T^n \mu \rightharpoonup T_\infty \mu$$

where the half-arrow denotes weak-star convergence, that is,

$$(8.1) \quad \lim_{n \rightarrow \infty} (x, T^n \mu) = (x, T_\infty \mu), \quad x \in \mathfrak{C}(\Omega)$$

where T_∞ is a positive continuous linear transformation of norm 1 of $\mathfrak{M}(\Omega)$ into itself. T_∞ is the adjoint of U_∞ and

$$(8.2) \quad T_\infty \mu = \sum_1^m \left(\int_\Omega \phi_i(t) d\mu(t) \right) \delta_i$$

where δ_i is the probability measure with all its measure concentrated at the point τ_i .

Proof. Theorem 6.1 gives that for $x \in \mathfrak{C}(\Omega)$, we have $\lim_{n \rightarrow \infty} U^n x = U_\infty x$, the convergence being in the sense of $\mathfrak{C}(\Omega)$. It follows that $\lim_{n \rightarrow \infty} (U^n x, \mu) = (U_\infty x, \mu)$ for $\mu \in \mathfrak{M}(\Omega)$. Let T_∞ be the adjoint of U_∞ that is, $(U_\infty x, \mu) = (x, T_\infty \mu)$. Hence by (2.10) $\lim_{n \rightarrow \infty} (x, T^n \mu) = (x, T_\infty \mu)$ and (8.1) is established. T_∞ being the adjoint of U_∞ is a continuous linear transformation of $\mathfrak{M}(\Omega)$ into itself of norm 1 and it is clearly positive. By Theorem 7.1 $U_\infty x = \sum_1^m x(\tau_i) \phi_i$. Hence for $\mu \in \mathfrak{M}(\Omega)$, $(U_\infty x, \mu) = \sum_1^m x(\tau_i) (\phi_i, \mu)$. Let $T' \mu = \sum_1^m (\phi_i, \mu) \delta_i$. $T' \mu$ is an element of $\mathfrak{M}(\Omega)$ and it is clear that $(U_\infty x, \mu) = (x, T' \mu)$ for all $x \in \mathfrak{C}(\Omega)$. Hence $T' = U_\infty^* = T_\infty$. Thus $T_\infty \mu = \sum_1^m (\phi_i, \mu) \delta_i$ and the theorem is proved.

9. Probability interpretation of $\phi_i(t)$. It is easy to see from the definition (2.9) of U , that $U^n \phi_i(t)$ represents the probability, that given the initial state of the system is t , that at the end of the $(n+1)$ st unit time interval the transformation A_i is applied. $\phi_i(t) = \lim_{n \rightarrow \infty} U^n \phi_i(t)$ thus represents the limiting probability of applying A_i , given that initially the state of the system is t .

Another point of view is obtained from (8.2). If δ_{t_0} is the probability measure concentrated at the single point t_0 , then $T_\infty \delta_{t_0} = \sum_1^m \phi_i(t_0) \delta_i$, so that $\phi_i(t_0)$ gives the probability that if the initial state is t_0 , the limiting state is τ_i .

To sum up, we have two probability interpretations for $\phi_i(t)$:

- (1) Limiting probability as $n \rightarrow \infty$ that at the n th step in the process, the transformation A_i is applied, given that the initial state is t .
- (2) Probability that the limiting state is τ_i , given that the initial state is t .

I wish to express my thanks to the referee for some useful comments.

REFERENCES

1. R. Bellman, T. Harris and H. N. Shapiro. *Studies on functional equations occurring in decision processes*, RM 878, Rand corporation July, 1952.
2. R. R. Bush and F. Mosteller, *Stochastic models for learning*, Wiley 1955.
3. W. Doeblin and R. Fortet, *Sur les chaînes à liaisons complètes*, Bull. Soc. Math. France **65** (1937), 132–148.
4. R. Fortet, (Thèse), *Sur l'iteration des substitutions algébrique linéaires à une infinité de variables*, Revista, Nos. 424–426, Año 40, Lima, 1938.
5. P. R. Halmos, *Measure theory*, Van Nostrand 1950.
6. C. T. Ionescu Tulcea and G. Marinescu, *Sur certaines chaînes à liaisons complètes*, C. R. Acad. Sci. Paris, **227** (1948), 667–669.
7. ———, *Theorie ergodique pour des classes d'operations non-complètement continues*, Ann. of Math. **52** (1950), 140–147.
8. S. Kakutani, *Concrete representation of abstract (M)-spaces*, Ann. of Math. **42** (1941), 994–1023.
9. S. Karlin, *Some random walks arising in learning models I*, Pacific J. Math. **3** (1953), 725–756.
10. O. Onicescu and G. Mihoc, *Sur les chaînes de variables statistiques*, Bull. Sci. Math. **2**, **59** (1935), 174–192.

UNIVERSITY COLLEGE, DUBLIN
NATIONAL UNIVERSITY OF IRELAND