Z. Wahrscheinlichkeitstheorie verw. Geb. 15, 51–56 (1970)

A Uniform Ergodic Theorem for Certain Markov Operators on Lipschitz Functions on Bounded Metric Spaces*

M. Frank Norman

Summary. Some sufficient conditions are given for uniform convergence of the iterates of transition operators associated with random contractions of a bounded metric space.

1. Introduction

Let (T, d) be a metric space, Λ its Borel subsets. For any complex valued function f on T, let

$$|f| = \sup_{t\in T} |f(t)|,$$

$$m(f) = \sup_{s\neq t} \frac{|f(s)-f(t)|}{d(s,t)},$$

and

$$\|f\| = m(f) + |f|.$$

The class $L = \{f\colon \|f\| < \infty\}$ of bounded Lipschitz functions is a Banach space with respect to the norm $\|\cdot\|$. If g maps T into T, let

$$W(g) = \sup_{s\neq t} \frac{d(g(s), g(t))}{d(s,t)}.$$

Let (E, Γ) be a measurable space. For every $t \in T$ let $p(t, \cdot)$ be a probability on Γ such that, for some $M < \infty$ and all $G \in \Gamma$,

$$m(p(\cdot, G)) \leqq M.$$

For any $x \in E$, let u_x map T into T, and for $x_1, \ldots, x_n$ in E let $u_{x_1\ldots x_n} = u_{x_n} \circ \cdots \circ u_{x_1}$. Assume that $\{(t,x)\colon u_x(t) \in A\} \in \Lambda \times \Gamma$ if $A \in \Lambda$, from which it follows that

$$\{(t, x_1, \ldots, x_n)\colon u_{x_1\ldots x_n}(t) \in A\} \in \Lambda \times \Gamma^n \qquad \text{if } A \in \Lambda.$$

Finally, suppose that there is a Γ measurable real valued function v and an $a < \infty$ such that

$$W(u_x) \leqq v(x)$$

for all $x \in E$ and

$$\int_E p(t, dx)\, v(x) \leqq a$$

for all $t \in T$.

* Supported by National Science Foundation grant GP-7335.

For every $f \in L$, Uf is the complex valued function

$$U f(t) = \int_E p(t, dx) f(u_x(t)).$$

Clearly $U f(t) = \int_T K(t, ds) f(s)$, where $K(t, A) = p(t, \{x: u_x(t) \in A\})$ is a Markov transition probability function in T. The Markov operator U is linear with $|Uf| \leqq |f|$. The equality

$$\begin{aligned} &U f(t) - U f(s) \\ &\quad = \int_E p(t, dx)\big(f(u_x(t)) - f(u_x(s))\big) + \int_E \big(p(t, dx) - p(s, dx)\big) f(u_x(s)) \end{aligned}$$

yields

$$m(Uf) \leqq a\, m(f) + R\, |f|, \tag{1}$$

where $R = 2M$, from which it follows that $Uf \in L$ and that U is bounded with $\|U\| \leqq \max(a, R+1)$.

A theorem of Ionescu Tulcea [1, Theorem 1] implies that, if (a) (T, d) is bounded (i.e. $b = \sup_{s, t \in T} d(s, t) < \infty$), (b) $W(u_x) \leqq r$ for some $r < 1$ and all $x \in E$, and (c) $p(t, G) \geqq \lambda \mu(G)$ for some $\lambda > 0$ and probability μ, and all $t \in T$ and $G \in \Gamma$, then there are constants $U^\infty f$, β and $\alpha < 1$ such that

$$|U^n f - U^\infty f| \leqq \beta\, \alpha^{\sqrt{n}}\, \|f\| \tag{2}$$

for all $n \geqq 1$. For no generality is lost in assuming that f is real valued with $\|f\| = 1$, and in that case (2) follows from Ionescu Tulcea's theorem on taking $a_n = M\, b\, r^n$, $c_n = m(f)\, b\, r^n$, and $s = [\sqrt{n}]$ (the largest integer that does not exceed $\sqrt{n}$), and noting that $\|f\|^+ = 1$. Ionescu Tulcea's $\|\cdot\|$ is my $|\cdot|$. Ionescu Tulcea [1, Section 8] further showed that

$$\|U^n - U^\infty\| \leqq \gamma\, \zeta^n \tag{3}$$

for constants γ and $\zeta < 1$ under the additional assumption that (T, d) is compact.

Lemma 1 below shows that (3) follows from (1) with $a = r < 1$ and (2), so that the compactness assumption is superfluous. Furthermore, Theorems 1 and 2 indicate that (b) and (c), respectively, can be considerably relaxed. One of the implications of Theorem 2 is that (3) obtains if, in addition to (a) and (b), (c) holds on a subset T' of T that is invariant under all u_x. Some applications of these results to learning models are presented in Section 4.

Since, in the case of compact T, there is a well developed theory of Markov operators satisfying (1) with $a < 1$ (see [2], [5, Parts II.B and II.C], and [6]) one naturally seeks an appropriate compactification of T. We note in this connection that boundedness of (T, d) does not insure that d can be extended to a metric on the Stone-Čech compactification. Thus the reduction to the case of compact T in [3, Paragraph 1 of the proof of Lemma 1] is defective.

2. Uniform Convergence of U^n

Lemma 1 is slightly more general than our immediate needs require. For any bounded linear operator V on L, $\rho(V)$ is its spectral radius.

Lemma 1. *Let U be a bounded linear operator on L with $|Uf| \leqq |f|$, and for every $f \in L$ let $U^\infty f$ be a bounded complex valued function on T. If there are real numbers R and $r<1$, a positive integer k, and a sequence $\{\delta_n\}$ with limit 0 such that*

$$m(U^k f) \leqq r\, m(f) + R\,|f| \tag{4}$$

and

$$|U^n f - U^\infty f| \leqq \delta_n \|f\| \tag{5}$$

for all $n \geqq 1$ and $f \in L$, then there is a bounded linear operator V on L such that

$$U^n = U^\infty + V^n \tag{6}$$

for all $n \geqq 1$, and

$$\rho(V) < 1. \tag{7}$$

Proof. Iteration of (4) yields

$$m(U^{jk} f) \leqq r^j m(f) + R\,\frac{1-r^j}{1-r}\,|f|. \tag{8}$$

Thus letting $j \to \infty$ on the right in

$$\begin{aligned}|U^\infty f(s) - U^\infty f(t)| \leqq\, & m(U^{jk} f)\, d(s,t) \\ & + |U^\infty f(s) - U^{jk} f(s)| + |U^{jk} f(t) - U^\infty f(t)|\end{aligned}$$

and using (5) we obtain

$$m(U^\infty f) \leqq \frac{R}{1-r}\,|f|. \tag{9}$$

Since (5) implies that $|U^\infty f| \leqq |f|$, U^∞ is a bounded linear operator on L.

Since $U^{n+1} f$ converges uniformly to $U^\infty U f$, $U^\infty f$, and $U U^\infty f$, it follows that $U^\infty U = U^\infty = U U^\infty$. And since $U^n U^n f$ converges uniformly to $U^\infty U^\infty f$ and $U^\infty f$, $U^{\infty\,2} = U^\infty$. Let $V = U - U^\infty$. Then, by induction, $V^n = U^n - U^\infty$, and (6) is proved.

Replacing f by $f - U^\infty f$ in (8) and using (9) on the right hand side we obtain

$$m(V^{jk} f) \leqq r^j m(f) + R'\,|f|, \tag{10}$$

where $R' = R(2+r)/(1-r)$. Eq. (5) yields

$$|V^{jk} f| \leqq \delta_{jk}\, m(f) + \delta_{jk}\,|f|. \tag{11}$$

Changing f to $V^{jkn} f$ in (10) and (11) and letting

$$a_{j,n} = \begin{pmatrix} m(V^{jkn} f) \\ |V^{jkn} f| \end{pmatrix} \quad \text{and} \quad P_j = \begin{pmatrix} r^j & R' \\ \delta_{jk} & \delta_{jk} \end{pmatrix}$$

leads to the vector inequality $a_{j,n+1} \leqq P_j\, a_{j,n}$, hence to $a_{j,n} \leqq P_j^n\, a_{j,0}$. Adding coordinates on the right and left of this equation we obtain

$$\|V^{jkn}\| \leqq |P_j^n|,$$

where $|P_j^n|$ is the maximum of the column sums of P_j^n. Now $\|V^{jkn}\|^{1/jkn} \to \rho(V)$ and $|P_j^n|^{1/n} \to \rho_j$ as $n \to \infty$, where ρ_j is the spectral radius of P_j. Thus

$$\rho(V) \leqq \rho_j^{1/jk}.$$

Since $\rho_j < 1$ for j sufficiently large, the proof is complete.

Lemma 1 finishes the derivation of (3) from (a), (b), and (c). Theorem 1 reaches the same conclusion on the basis of assumptions more general than (b). Let $p_1 = p$, and, for $n > 1$, $t \in T$, and $A \in \Gamma^n$ let

$$p_n(t, A) = \int_E p(t, dx_1) \int_E \cdots \int_E p(u_{x_1 \ldots x_{n-1}}(t), dx_n)\, I_A(x_1, \ldots, x_n),$$

where I_A is the indicator function of A. Let μ^n be the nth Cartesian power of the probability μ on Γ.

Theorem 1. *Suppose that, in addition to* (a) *and* (c), *the following conditions obtain:* (b 1) *There is an $r < 1$, a positive integer k, and a Γ^k measurable real valued function h such that $W(u_{x_1 \ldots x_k}) \leqq h(x_1, \ldots, x_k)$ for all $x_1, \ldots, x_k \in E$, and*

$$\int_E p_k(t, d(x_1, \ldots, x_k))\, h(x_1, \ldots, x_k) \leqq r$$

for all $t \in T$. (b 2) *There is a positive integer k' and a $\Gamma^{k'}$ measurable real valued function h' such that*

$$W(u_{x_1 \ldots x_{k'}}) \leqq h'(x_1, \ldots, x_{k'}) \qquad \text{for all } x_1, \ldots, x_{k'} \in E,$$

and

$$\int_{E^{k'}} \mu^{k'}(d(x_1, \ldots, x_{k'}))\, h'(x_1, \ldots, x_{k'}) < 1.$$

Then there are bounded linear operators U^∞ and V on L such that $U^\infty f$ is constant for any $f \in L$, $\rho(V) < 1$, and $U^n = U^\infty + V^n$ for all $n \geqq 1$.

One proof of this theorem obtains (4) directly from (b 1), and (5), with $U^\infty f$ constant, by a routine modification of the proof of Ionescu Tulcea's Theorem 1. We omit details.

3. u_x Invariant Subsets of T

Let T' be a Borel subset of T such that $u_x(t) \in T'$ for all $x \in E$ and $t \in T'$. The assumptions of the second paragraph of Section 1, as well as (a) and (b), carry over to the restrictions d', p', and u' of d, p, and u to T'. Hence if p' satiesfies (c), (12) below holds, where U' is the operator on L' induced by p'. Thus Theorem 2 is applicable and (3) follows, as was claimed in Section 1.

Theorem 2. *Under* (a) *and* (b 1), *if there is a null sequence γ_n, and for every $f \in L'$ there is a constant $U'^\infty f$ such that*

$$|U'^n f - U'^\infty f|' \leqq \gamma_n \|f\|' \tag{12}$$

for all $n \geqq 1$, then $U^n = U^\infty + V^n$, where $U^\infty f = U'^\infty f'$ and V is a bounded linear operator on L with $\rho(V) < 1$.

Proof. The function $\psi(t)=\inf_{s\in T'} d(t,s)$ is bounded and Lipschitz. For any $s\in T'$

$$\psi\big(u_{x_1\ldots x_k}(t)\big)\leqq d\big(u_{x_1\ldots x_k}(t),\, u_{x_1\ldots x_k}(s)\big),$$

so

$$\psi\big(u_{x_1\ldots x_k}(t)\big)\leqq h(x_1,\ldots,x_k)\,\psi(t).$$

Thus $U^k\psi(t)\leqq r\,\psi(t)$. Iterating and using the positivity of U we obtain $U^{jk}\psi(t)\leqq r^j\psi(t)$, so that

$$U^{jk}\psi(t)\leqq b\,r^j \tag{13}$$

for all $j\geqq 0$.

For any $f\in L$, $t\in T$ and $s\in T'$,

$$|f(t)|\leqq|f(t)-f(s)|+|f(s)|,$$

so that

$$|f(t)|\leqq m(f)\,\psi(t)+|f'|',$$

where f' is the restriction of f to T'. Therefore

$$|U^{jk}f(t)|\leqq m(f)\,U^{jk}\psi(t)+|f'|',$$

or, using (13),

$$|U^{jk}f|\leqq b\,r^j\,m(f)+|f'|'. \tag{14}$$

Replacing f by $U^i f-U^\infty f$ in (14) we obtain

$$|U^{jk+i}f-U^\infty f|\leqq b\,r^j\,m(U^i f)+|U'^i f'-U'^\infty f'|'. \tag{15}$$

It has already been observed that (b1) implies (4), and the latter implies, in turn, that $m(U^i f)\leqq J\,\|f\|$ for some constant J and all i. This inequality along with (12) and (15) yields

$$|U^{jk+i}f-U^\infty f|\leqq(b\,J\,r^j+\gamma_i)\,\|f\|. \tag{16}$$

Letting $j=[n/2k]$ and $i=n-jk$ we obtain (5). An application of Lemma 1 completes the proof.

4. Applications

In many psychological learning experiments, a subject makes a sequence of choices from a measurable space (Y,Δ), and each of these is followed by an outcome from a measurable space (Z,Σ). Let $(E,\Gamma)=(Y,\Delta)\times(Z,\Sigma)$. The experimenter prescribes the probability distribution $\xi(y,\cdot)$ over Σ after a choice y. We assume that $\xi(\cdot,B)$ is Δ measurable for each $B\in\Sigma$. In many learning models, the subject is characterized at the time of a choice by a probability distribution t over Y. T is the set of choice distributions considered (e.g. the probabilities absolutely continuous with respect to some fixed measure on Δ), and d is the total variation metric. Obviously (T,d) is bounded. The joint distribution of response and outcome is

$$p(t,G)=\int_Y t(dy)\int_Z \xi(y,dz)\,I_G(y,z).$$

Clearly $m\big(p(\cdot,G)\big)\leqq 1$ for all $G\in\Gamma$. Occurrence of the response-outcome pair $x=(y,z)$ effects the transformation $t\to u_x(t)$ of the choice distribution. In linear models $u_x(t)=(1-\theta_x)\,t+\theta_x\,v_x$, for measurable mappings θ and v of (E,Γ) into the Borel subsets of $[0,1]$ and T, respectively. We assume convexity of T to insure that $u_x(T)\subseteqq T$, and separability of T to insure joint measurability of u. Clearly $W(u_x)=1-\theta_x\leqq 1$, and (b) is satisfied if $\theta_x\geqq\delta>0$ for all $x\in E$.

We now consider two sets of additional assumptions that permit us to conclude that $U^n = U^\infty + V^n$, where $U^\infty f$ is constant for all $f \in L$ and V is bounded on L with $\rho(V) < 1$. Case A is applicable to Suppes' [7] model, case B to Norman's [4].

A. The functions θ and v depend only on z, and $\xi(y, B) \geqq \lambda\,\mu(B)$ for some $\lambda > 0$ and probability μ on Σ. Since u doesn't depend on y, U is the same as the operator U^* corresponding to $E^* = Z$, $p^*(t, B) = p(t, Y \times B)$, and $u_z^* = u_{(y,z)}$. But $p^*(t, B) \geqq \lambda\,\mu(B)$, so p^* satisfies (c).

B. $v_x(B) \geqq \lambda\,\omega(B)$ for some $\lambda > 0$ and probability ω on Δ. Clearly the convex hull T' of the range of v is u_x invariant for all $x \in E$, and $t(B) \geqq \lambda\,\omega(B)$ for all $t \in T'$. Hence $p(t, G) \geqq \lambda\,\mu(G)$ for all $t \in T'$ and $G \in \Gamma$, where μ is the probability

$$\mu(G) = \int_Y \omega(dy) \int_Z \xi(y, dz)\, I_G(y, z)$$

on Γ.

References

1. Ionescu Tulcea, C.T.: On a class of operators occurring in the theory of chains of infinite order. Canadian J. Math. **11**, 112–121 (1959).
2. – Marinescu, G.: Theorie ergodique pour des classes d'operations non completement continues. Ann. of Math. II. Ser. **52**, 140–147 (1950).
3. Isaac, R.: Markov processes and unique stationary probability measures. Pacific J. Math. **12**, 273–286 (1962).
4. Norman, M.F.: An approach to free-responding on schedules that prescribe reinforcement probability as a function of interresponse time. J. math. Psychology **3**, 235–268 (1966).
5. – Mathematical learning theory. In G.B. Dantzig and A.F. Veinott (Eds.), Mathematics of the decision sciences, Part 2. p. 283–313. Providence, R. I.: Amer. Math. Soc. 1968.
6. – Compact Markov processes. Technical Report No 2, NSF grant GP-7335, August 1968.
7. Suppes, P.: A linear model for a continuum of responses. In R.R. Bush and W.K. Estes (Eds.), Studies in mathematical learning theory. P. 400–414. Stanford, Calif.: Stanford Univ. Press 1959.

M. Frank Norman
University of Pennsylvania
Department of Psychology
3813–15 Walnut Street
Philadelphia, Pa. 19104
USA

(Received January 15, 1969)