THE STRONG LAW OF LARGE NUMBERS FOR A CLASS OF MARKOV CHAINS

By Leo Breiman[1]

University of California at Los Angeles

1. Introduction. The following problem has arisen in the study of Markov chains of the learning model type. (See [1] for definitions). Let the state space be, for example, the unit interval [0, 1] and let the chain have a unique invariant initial distribution $\pi(dx)$. Now let the chain be started at some point $x \varepsilon [0, 1]$; is it true that

$$\frac{1}{N}\sum_{n=1}^{N} X_n \rightarrow E_\pi X_1 \quad \text{a.s.?} \tag{1}$$

From the ergodic theorem we know that there is a set $S \subset [0, 1]$ such that $\pi(S) = 1$, and, if $x \varepsilon S$, then (1) holds. In learning models, however, π may be singular with respect to Lesbesgue measure, so a stronger result is desirable. We prove for a wide class of chains, including learning models, that (1) holds for every possible starting point. This result is well known for chains satisfying Doeblin's condition. Unfortunately, learning models do not.

2. The theorem. Let the state space Ω be a compact Hausdorf space, and $\mathcal{B}$ the Baire σ-field in Ω. The Markov transition probabilities $P(A \mid x)$ are assumed probabilities on $\mathcal{B}$ for fixed x, $\mathcal{B}$-measurable functions on Ω for fixed A, and such that there is a unique probability π on $\mathcal{B}$ satisfying

$$\pi(A) = \int P(A \mid x)\pi(dx), \qquad \text{all } A \varepsilon \mathcal{B}.$$

Let C be the class of all continuous functions on Ω, and add the final restriction that, if $f \varepsilon C$, so is $E(f(X_1) \mid X_0 = x)$. Let $\Omega^{(\infty)}$ be the infinite sequence space with coordinates in Ω. In the usual way, we construct a σ-field $\mathcal{B}^{(\infty)}$ in $\Omega^{(\infty)}$ and, using the initial distribution $X_0 = x$, a probability P_x on $\mathcal{B}^{(\infty)}$. Then

Theorem. *Let $\phi \varepsilon C$, Then, for any $x \varepsilon \Omega$,*

$$\frac{1}{N}\sum_{1}^{N} \phi(X_n) \rightarrow E_\pi \Phi(X_1) \quad \text{a.s. } P_x .$$

Proof. The proof of this theorem is a combination of the Kakutani-Yosida norms ergodic lemma and an argument concerning conditional probabilities.

3. The topological part. We prove first a proposition which summarizes the topological ergodic theorem we need. Define the operator T on C into C by $(T\phi)(x) = E(\phi(X_1) \mid X_0 = x)$, so that $(T^k\phi)(x) = E(\phi(X_k) \mid X_0 = x)$, and

Received October 16, 1959.

[1] This research was carried on during the tenure of a National Science Foundation Fellowship.

set $\bar{T}_N\phi = \sum_1^N T^n\phi/N$. Then

PROPOSITION 1. For any $\phi \varepsilon C$, $\bar{T}_N\phi$ converges uniformly to $E_\pi\phi(X_1)$.

PROOF. This proposition and its proof are well known in linear space theory. However, for completeness, we give a short demonstration. Let $\mathfrak{M}$ be the class of probability measures on $\mathcal{B}$, and consider the operators V, $\bar{V}_N$ on $\mathfrak{M}$ into $\mathfrak{M}$ defined by

$$(VQ)(A) = \int P(A \mid x)Q(dx), \qquad \bar{V}_N Q = \sum_1^N V^k Q/N.$$

By the Helly-Bray theorem, $\mathfrak{M}$ is closed and compact in the weak dual topology, so that there are plenty of convergent subsequences $\bar{V}_{N_k}Q$. But every limit point of $\bar{V}_N Q$ is invariant under V and hence is identified with π, so that $\bar{V}_N Q \to \pi$ in our topology. Therefore, for every $Q \varepsilon \mathfrak{M}$, and $\phi \varepsilon C$ we have

$$(\bar{T}_N\phi, Q) = (\phi, \bar{V}_N Q) \to (\phi, \pi)$$

and hence $\bar{T}_N\phi$ converges weakly to $E_\pi\phi$. Applying the Kakutani-Yosida norms ergodic lemma (see, for example, [2], pg. 441), we conclude that $\bar{T}_N\phi$ converges uniformly to $E_\pi\phi$.

4. The probabilistic part. Let $X_1, X_2, \cdots$ be distributed according to P_x, and define

$$Z_n^{(1)} = \begin{cases} \phi(X_n) - E(\phi(X_n) \mid X_{n-1}), & n > 1 \\ 0, & n \leq 1 \end{cases}$$

$$Z_n^{(k)} = \begin{cases} E(\phi(X_n) \mid X_{n-k+1}) - E(\phi(X_n) \mid X_{n-k}), & n > k \\ 0, & n \leq k. \end{cases}$$

PROPOSITION 2. $N^{-1}\sum_{n=1}^N Z_{(n)}^{(k)} \to 0$ a.s. P_x.

PROOF. We use the following result ([2], pg. 387). Let $Y_1, Y_2, \cdots$ be a sequence of random variables such that $E(Y_n \mid Y_{n-1}, \cdots, Y_1) = 0$ and

$$EY_n^2 \leq M < \infty,$$

all n. Then

$$\frac{1}{N}\sum_{n=1}^N Y_n \to 0 \quad \text{a.s.}$$

To apply this, note that

$$E(Z_n^{(k)} \mid Z_{n-1}^{(k)}, \cdots, Z_1^{(k)}) = E(E(Z_n^{(k)} \mid X_{n-k}, X_{n-k-1}, \cdots, X_1) \mid Z_{n-1}^{(k)}, \cdots, Z_1^{(k)}),$$

and that, since the $X_1, X_2, \cdots$ form a Markov chain,

$$E(Z_n^{(k)} \mid X_{n-k}, \ldots) = E(Z_n^{(k)} \mid X_{n-k}) = 0.$$

Further, $E(Z_n^{(k)})^2 \leq 2(\sup|\phi|)^2$, thus giving the proposition.

5. Conclusion of the proof. To complete the demonstration of the theorem, write

$$\phi(X_n) - E(\phi(X_n) \mid X_{n-k}) = Z_n^{(1)} + Z_n^{(2)} + \cdots + Z_n^{(k)}, \qquad n > k.$$

Thus, by proposition 2,

$$\lim_N \left| \frac{1}{N} \sum_{n=1}^{N} \phi(X_n) - \frac{1}{N} \sum_{n=k+1}^{N} E(\phi(X_n) \mid X_{n-k}) \right| = 0, \quad \text{a.s. } P_x.$$

Or, neglecting at most k terms,

$$\lim_N \left| \frac{1}{N} \sum_{n=1}^{N} \phi(X_n) - \frac{1}{N} \sum_{n=1}^{N} E(\phi(X_{n+k}) \mid X_n) \right| = 0, \quad \text{a.s. } P_x,$$

so that, for fixed M,

$$\lim_N \left| \frac{1}{N} \sum_{n=1}^{N} \phi(X_n) - \frac{1}{N} \sum_{k=1}^{N} \left[\frac{1}{M} \sum_{k=1}^{M} E(\phi(X_{n+k}) \mid X_n) \right] \right| = 0, \quad \text{a.s. } P_x.$$

By proposition 1, for any $\epsilon > 0$, we may choose M such that

$$\left| \frac{1}{M} \sum_{k=1}^{M} E(\phi(X_{n+k}) \mid X_n) - E_\pi \phi(X_1) \right| \leqq \epsilon,$$

and for such an M we have

$$\lim_N \left| \frac{1}{N} \sum_{n=1}^{N} \phi(X_n) - E_\pi \phi(X_1) \right| \leqq \epsilon \quad \text{a.s. } P_x)$$

proving the theorem.

REFERENCES

[1] Samuel Karlin, "Some random walks occurring in learning models," *Pacific J. Math.*, Vol. 3 (1933), 725–756.
[2] Michael Loève, *Probability Theory*, D. Van Nostrand, New York, 1955.

EMPTINESS IN THE FINITE DAM

By A. Ghosal

Central Fuel Research Institute, Dhanbad, India

1. Summary: The paper discusses the general problem of emptiness in the finite dam and considers the probability that, starting with an arbitrary storage, the dam dries up before it fills completely. Some exact results are given both for discrete and continuous inputs. An interesting relation between this probability and the asymptotic distribution function of the dam content has also been obtained.

Received April 9, 1959; revised February 20, 1960.