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

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

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.