A NOTE ON LIMIT THEOREMS FOR THE ENTROPY OF MARKOV CHAINS[1]

By Harry Dym

Massachusetts Institute of Technology

Let X_j, $j = 1, 2, \cdots$ be a stationary ergodic m-step Markov chain defined on a probability space (Ω, Q, P) and having for its state space the finite set of integers $\{0, 1, \cdots, D - 1\}$. Here Ω, the sample (path) space, is equal to the set of all sequences $(\omega_1, \omega_2, \cdots)$ with $\omega_n = X_n(\omega) \varepsilon \{0, 1, \cdots, D - 1\}$. Q is the Borel field generated by the cylinder sets of Ω and P is a stationary probability measure on Ω which for all $n \geqq m + 1$ satisfies the relation

$$P\{X_n = i_n/X_{n-1} = i_{n-1}, \cdots, X_1 = i_1\} = P\{X_{m+1} = i_n/X_m = i_{n-1}, \cdots, X_1 = i_{n-m}\}$$

where $i_K \varepsilon \{0, 1, \cdots, D - 1\}$, $K = 1, \cdots, n$.

For $\omega \varepsilon \Omega$ let $[\omega]_n$ denote the cylinder set $\{u \varepsilon \Omega: u_1 = \omega_1, \cdots, u_n = \omega_n\}$ and correspondingly let $P([\cdot]_n)$ denote the random variable whose value at ω is $P([\omega]_n) = P\{u \varepsilon \Omega: u \varepsilon [\omega]_n\}$. In this note we establish a law of the iterated logarithm for the sequence of random variables $\{-\log P([\cdot]_n)\}$:

Theorem 1.

$$P\{\omega: \limsup_{n\to\infty} [(-\log P([\omega]_n) - nH)/(2Bn \log \log n)^{\frac{1}{2}}] = 1\} = 1$$

where H denotes the entropy rate of the process X_n, i.e.,

$$H = \lim_{n\to\infty} [E(-\log P([\omega]_n))/n]$$

and

$$B = \lim_{n\to\infty} [E\{(-\log P([\omega]_n) - nH)^2\}/n].$$

E denotes the expectation operator relative to the measure P.

The proof of Theorem 1, to be presented below, depends essentially upon the observation that there exists a function f and a one-step Markov chain $Z_j(\omega)$, $j = m + 1, m + 2, \cdots, \omega \varepsilon \Omega$, such that

$$(1) \qquad -\log P([\omega]_n) = \sum_{j=m+1}^{n} f(Z_j(\omega)) + O(1).$$

The proof is then completed by referring to a version of the law of the iterated logarithm theorem which is applicable to functionals of a Markov chain (Chung, [3], Theorem 5, p. 101). It is worth noting that (1) may be used in conjunction with other established limit theorems for functionals of a Markov chain, (Chung,

Received 22 September 1965.

[1] This work has been supported by the Mitre Corp., Bedford, Massachusetts, under Air Force Contract.

[3], Theorem 1, p. 94 and Theorem 2, p. 87) to derive, for example, the following theorems:

THEOREM 2.

$$\lim_{n\to\infty} P\{[-\log P([\omega]_n) - nH]/(Bn)^{\frac{1}{2}} \leqq a\} = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{a} e^{-v^2/2}\, dv.$$

THEOREM 3.

$$\lim_{n\to\infty} [-\log P([\omega]_n)/n] = H \textit{ with probability one.}$$

As it happens, both Theorem 2, the central limit theorem, and Theorem 3, the Shannon-McMillan theorem, are known to hold in a more general setting. For details concerning Theorem 2 see Ibragimov ([5], Theorem 2.6, p. 376); for Theorem 3 see Breiman [1] and [2] and further extensions by Chung [4].

PROOF OF THEOREM 1. Let $Y_j = \sum_{k=0}^{m-1} X_{j-k} D^{-k-1}$, $j = m, m+1, \cdots$. Then Y_j is an irreducible simple (one-step) Markov chain on Ω with state space $\{iD^{-m}\}$, $i = 0, 1, \cdots, D^m - 1$. For $s, v = 0, 1, \cdots, D^m - 1$, let P_{sv} denote the probability of a transition (in one step) from the state sD^{-m} to the state vD^{-m}, let $\pi_s (= P\{Y_k = sD^{-m}\})$ denote the corresponding stationary initial probabilities and let $N_{sv}([\omega]_n)$, $n \geqq m + 1$, be the number of one-step transitions from the state sD^{-m} to the state vD^{-m} in the sequence $Y_m(\omega), \cdots, Y_n(\omega)$. (Note that $\sum_{s,v} N_{sv}([\omega]_n) = n - m$.) Then

$$-\log P([\omega]_n) = -\log P([\omega]_m) - \sum_{s,v} N_{sv}([\omega]_n) \log P_{sv}.$$

Now, the sequence of random variables

$$Z_j = Y_j + D^{-m} Y_{j-1}, \quad j = m+1, m+2, \cdots,$$

form a stationary irreducible Markov chain on Ω and

$$N_{sv}([\omega]_n) = \sum_{j=m+1}^{n} \chi_{sv}(Z_j(\omega))$$

where

$$\chi_{sv}(a) = 1 \quad \text{if} \quad a = (D^{-m}s + v)D^{-m}$$
$$= 0 \quad \text{if} \quad a \neq (D^{-m}s + v)D^{-m}.$$

Hence, setting $f = -\sum_{s,v} (\log P_{sv})\chi_{sv}$ and ${}_mS_n = \sum_{j=m+1}^{n} f(Z_j)$ $(n \geqq m+1)$, we conclude that $-\log P([\omega]_n) = -\log P([\omega]_m) + {}_mS_n(\omega)$.

It is easily checked that

$$E[f(Z_j)] = -\sum \pi_s P_{sv} \log P_{sv} = H.$$

Thus

$$\limsup_{n\to\infty} [(-\log P([\omega]_n) - nH)/(2Bn \log\log n)^{\frac{1}{2}}]$$
$$= \limsup_{n\to\infty} [({}_mS_n - E({}_mS_n) - mH - \log P([\omega]_m))/(2Bn \log\log n)^{\frac{1}{2}}]$$
$$= \limsup_{n\to\infty} [({}_mS_n - E({}_mS_n))/(2Bn \log\log n)^{\frac{1}{2}}].$$

Furthermore, since Z_j, $j \geqq m + 1$, is a finite irreducible (and hence recurrent) stationary chain it follows that (Chung, [3], Corollary, p. 32)

$$\begin{aligned}\lim_{n\to\infty} n^{-1} \textstyle\sum_{j=m+1}^{n} E\{(f(Z_j) - H) \log P([\omega]_m)\} \\ &= E(f(Z_{m+1}) - H)E(\log P([\omega]_m)) \\ &= 0.\end{aligned}$$

Consequently

$$\begin{aligned}\lim_{n\to\infty} E\{({}_mS_n - (n - m)H) \log P([\omega]_m)/n\} \\ &= \lim_{n\to\infty} n^{-1} \textstyle\sum_{j=m+1}^{n} E\{(f(Z_j) - H) \log P([\omega]_m)\} \\ &= 0\end{aligned}$$

and

$$B = \lim_{n\to\infty} [E(-\log P([\omega]_n) - nH)^2/n] = \lim_{n\to\infty} [E({}_mS_n - E({}_mS_n))^2/n].$$

The desired result now follows by Theorem 3, Theorem 5 and the ensuing remark of Chung ([3], pp. 97, 101, 102).

REFERENCES

[1] Breiman, L. (1957). The individual ergodic theorem of information theory. *Ann. Math. Statist.* **28** 809–811.

[2] Breiman, L. (1960). Correction to the individual ergodic theorem of information theory. *Ann. Math. Statist.* **31** 809–810.

[3] Chung, K. L. (1960). *Markov Chains with stationary transition probabilities.* Springer, Berlin.

[4] Chung, K. L. (1961). A note on the ergodic theorem of information theory. *Ann. Math. Statist.* **32** 612–614.

[5] Ibragimov, I. A. (1962). Some limit theorems for stationary processes. *Theor. Prob. Appl.* **7** 349–382.