ENTROPY OF FIRST RETURN PARTITIONS

OF A MARKOV CHAIN

by

E. M. Klimko and James Yackel*

Department of Statistics

Division of Mathematical Sciences

Mimeograph Series No. 175

December 1968

---

*The work of the second author was supported in part by National Science Foundation Grant GP7631.

ENTROPY OF FIRST RETURN PARTITIONS OF A MARKOV CHAIN

by

E. M. Klimko and James Yackel *

Abstract

We consider the first return time distributions for each state in a Markov chain and show that finiteness of entropy of these distributions is a class property for both recurrent and transient classes.

1. Introduction.

In this note, we answer affirmatively the question raised in [2] concerning the finiteness of entropy for the first return time distributions of Markov chains as a class property. The interest of our result lies in the null recurrent and transient classes since it is known that the finite mean return time of a positive recurrent state implies that the first return distribution has finite entropy. On the other hand, it is easy to construct Markov chains whose first return distribution to a given state has infinite entropy; indeed, it is possible to construct a chain with any given first return distribution to a fixed state, c.f. [1] p. 64.

In section 2, we derive some bounds on entropy which are applied in section 3 to prove our probabilistic result.

*The work of the second author was supported in part by National Science Foundation Grant GP 7631.

2. Preliminaries.

Let $(\Omega, \mathcal{B}, \mu)$ be a σ-finite measure space. To a partition $\mathcal{C} = \{A_i\}_{i=0}^{\infty}$ of a set A, $\mu(A) < \infty$, is associated a sequence $f = \{f_i\}_{i=0}^{\infty}$ with $f_i = \mu(A_i)$ $i = 0, 1, \cdots$. The entropy of f is the entropy of $\mathcal{C}$

(1)
$$H(f) = H(\mathcal{C}) = -\sum_{i=1}^{\infty} f_i \log f_i \ .$$

(The base of the logarithm is usually taken to be 2; $0 \log 0 = 0$; there are no difficulties in definition (1) since at most a finite number of terms can be negative). The norm $|f| = \Sigma f_i$; the convolution of f and g is $f*g$, i.e., $(f*g)_n = \sum_{i=0}^{n} f_{n-i} g_i$ and f^{*k} is the k-fold convolution of f with itself.

Lemma 1. Let f, g be sequences. Then there is a constant C, depending only on $|f|$, $|g|$ and the base of the logarithm such that

(2)
$$\text{Max}\,(H(f), H(g)) - C \leq H(f + g) \leq H(f) + h(g),$$

in particular

$H(f + g) < \infty$ if and only if $H(f) < \infty$ and $H(g) < \infty$.

Proof. The function $-\log x$ is decreasing and $-x \log x$ is increasing for $x \in [0, 1/e]$. For each n,

$$\begin{aligned} -(f_n + g_n) \log (f_n + g_n) &= -f_n \log (f_n + g_n) - g_n \log (f_n + g_n) \\ &\leq -f_n \log f_n - g_n \log g_n \ , \end{aligned}$$

while for n sufficiently large, $f_n + g_n \in [0, 1/e]$ and

$$-\max\,(f_n, g_n) \log \max\,(f_n, g_n) \leq -(f_n + g_n) \log (f_n + g_n).$$

Lemma 2. If f and g are sequences, then

$$\max\,(f_{i_0} H(g),\ g_{j_0} H(f)) - C \leq H(f*g) \leq |f|H(g) + |g|H(f). \tag{3}$$

where f_{i_0}, g_{j_0} are arbitrary non zero elements of f, g. In particular, we conclude $H(f*g) < \infty$ if and only if $H(f) < \infty$ and $H(g) < \infty$.

Proof. By the monotonicity of the loagrithm function,

$$H(f*g) = -\sum_{n=0}^{\infty} \left(\sum_{i=0}^{n} f_i g_{n-i}\right) \log \left(\sum_{i=0}^{n} f_i g_{n-i}\right)$$

$$\leq \sum_{n=0}^{\infty} \left(\sum_{i=0}^{n} f_i g_{n-i} \log f_i g_{n-i}\right).$$

Now interchanging the order of summation yields

$$-\sum_{i=0}^{\infty} \sum_{n=i}^{\infty} f_i g_{n-i}\,(\log f_i + \log g_{n-i}) = |g|\,H(f) + |f|H(g).$$

The lower bound is obtained by noting that for i_0, n sufficiently large .

$$-\left(\sum_{i=0}^{n} f_i g_{n-i}\right) \log \left(\sum_{i=0}^{n} f_i g_{n-i}\right) \geq -f_{i_0} g_{n-i_0} \log f_{i_0} g_{n-i_0}$$

$$\geq -f_{i_0} g_{n-i_0} \log g_{n-i_0}$$

which summed on n gives us $f_{i_0} H(g) - C$. We similarly can obtain a bound involving $H(f)$.

Lemma 3. If we consider the k-fold convolution of f, f^{*k}, then

$$H(f^{*k}) \leq k|f|^{k-1}H(f). \tag{4}$$

Proof. We prove (4) by induction, noting that for $k=1$ we have equality. If (4) holds for some positive integer k, it follows from Lemma 2 that

$$\begin{aligned} H(f^{*(k+1)}) &= H(f*f^{*k}) \leq |f^{*k}|H(f) + |f|H(f^{*k}) \\ &= (k+1)\ |f|^{k}H(f). \end{aligned}$$

Lemma 4. If $|f| < 1$, then $H(f) < \infty$ implies

$$H(\sum_{k=0}^{\infty} f^{*k}) \leq \frac{H(f)}{(1-|f|)^2} < \infty \quad .$$

Proof. The result follows from (2), (4) and $\Sigma ka^{k-1} = (1-a)^{-2}$.

3. Main Result.

We now apply the preceeding lemmas to obtain the following proposition. (We follow the standard notation and terminology of [1]).

Proposition. The finiteness of the entropy of first return distributions $f_{kk} = \{f^n{}_{kk}\}_{n=1}^{\infty}$ is a class property for Markov chains.

Proof. Let the states i and j communicate. It is easily verified probabilistically that for any two states h, k

$$f^n{}_{kk} = {}_hf^n{}_{kk} + ({}_kf_{kh} * f_{hk})(n) = {}_hf^n{}_{kk} + ({}_kf_{kh} * (\sum_{m=0}^{\infty} {}_kf_{hh}{}^{*m}) * {}_hf_{hk})(n).$$

If $H(f_{ii}) < \infty$, our lemmas imply that (i) $H({}_jf_{ii}) < \infty$, (ii) $H({}_if_{ij}) < \infty$, (iii) $H({}_if_{jj}) < \infty$ and (iv) $H({}_jf_{ji}) < \infty$. Since i and j communicate, we assert that $|{}_jf_{ii}| < 1$. From Lemma 4 we conclude that

$$H\left(\sum_{m=0}^{\infty} {}_jF_{ii}^{*m}\right) < \infty.$$

This together with (ii) and (iv) implies

$$H\left({}_jf_{ji} * \left(\sum_{m=0}^{\infty} {}_jf_{ii}^{*m}\right) * {}_if_{ij}\right) < \infty,$$

which together with (iii) completes the proof.

References

1. Chung, K. L. : Markov Chains with Stationary Transition probabilities. Berlin-Heidelberg-New York: Springer, 1967.

2. Klimko, E. M. and Sucheston, Louis : On Convergence of Information in Spaces with Infinite Invariant measure. Z. Wahrscheinlichkeitstheorie verw. Geb. 10, 226-235 (1968).