

## COVERING FINITE SETS BY ERGODIC IMAGES

BY  
J. MICHAEL STEELE

**ABSTRACT.** For any ergodic transformation  $T$  a set  $A$  of measure less than  $\varepsilon$  is constructed with the property that for every finite set  $F$  there is a  $j = j(F)$  such that  $F \subset T^{-j}A$ . The basic tool used to prove this is a purely combinatorial result which says there is a small subset of  $\{1, 2, \dots, n\}$  which can be shifted a small amount to cover any  $k$  set in  $\{j: \delta n \leq j \leq n\}$ . Applications are given to the theory of combinatorial entropy.

**I. Introduction.** The primary objective of this paper is to prove the following:

If  $T$  is an ergodic transformation, then there is a set  $A$  of measure less than  $\varepsilon$  such that for every finite set  $F$  there is a  $j$  such that  $F \subset T^{-j}A$ .

In the first place the very definition of ergodicity provides us with the initial introduction to the theorem. For an ergodic  $T$  and *any* set  $A$  of positive measure, one has that the images  $T^{-j}A$  are sufficient to “catch” almost any *single* element set. The crux of the present result is that there is *some* (arbitrarily small)  $A$  whose images  $T^{-j}A$  are sufficient to “catch” any *finite* element set.

As this relation to ergodicity suggests, this type of finite covering property relates in several ways to the taxonomy of measure preserving transformations. In fact the *aperiodicity* of a transformation is equivalent in a sense described in the third section to the possession of the finite covering property.

The method of proof employed here also has by-products which may prove to be of wider application. In particular the proof proceeds by first establishing some purely combinatorial results, and then using Rohlin’s Theorem as a means of applying the combinatorial facts to ergodic theory problems. The second section contains only combinatorial work and can be read independently of the remainder.

Finally, we give several applications of the finite covering property. Particularly, the relationship of Kolmogorov’s entropy to the combinatorial entropy of Vapnik and Chervonenkis is considered.

**II. A combinatorial lemma.** In the following, the notation  $[0, n]$  will indicate the set  $\{j: 0 \leq j \leq n\}$ . Similarly, we will use  $[0, r)$  for  $\{j: 0 \leq j < r\}$  and

---

Received by the editors October 22, 1976 and, in revised form, July 14, 1977.

*Key words.* Ergodic, aperiodic, entropy, combinatorial method.

$[0, \delta n]$  for  $\{j: 0 \leq j \leq \delta n\}$ , even when  $\delta n$  is not an integer. As usual  $|S|$  denotes the number of elements of the set  $S$ .

We now have the main result of this section.

LEMMA II.1. *Let  $k$  be a positive integer and  $\delta, \varepsilon_1$  and  $\varepsilon_2$  be positive reals. Then there is an arbitrarily large  $n$  and a set  $S \subset [0, n]$  such that the following conditions hold:*

- (1)  $|S|/n \leq \varepsilon_1$
- (2) For any  $k$  element set  $F \subset [\delta n, n]$  there is a natural number  $t$  such that  $F \subset S + t$  with  $t \leq \varepsilon_2 n$ .

**Proof.** We first write  $n = \tau!$  and  $\tau$  will be chosen in the course of the proof. The set  $S$  will be given as the union over  $j = 1, 2, \dots, k$  of the sets  $S_j$  defined by

$$S_j = \{r : r \in (s(\tau - 2j + 1)! - (\tau - 2j)!), s(\tau - 2j + 1)!\} \text{ for some } s \in [1, \tau! / (\tau - 2j + 1)!]$$

Now let  $n_1 < n_2 < n_3 \dots < n_k$  be the elements of the  $k$  set  $F$ . We need to check that there is a  $t \leq \varepsilon_2 n$  such that  $S + t$  contains  $F$ . The integer  $t$  will be given as the sum of  $k$  integers  $t_1, t_2, \dots, t_k$  which will be constructed to have the properties that

$$n_j - (t_1 + t_2 + \dots + t_j) = s_j(\tau - 2j + 1)!, \quad s_j > 0,$$

and

$$t_{j+1} + t_{j+2} + \dots + t_k \in [0, (\tau - 2j)!].$$

By the definition of the  $S_j$  we therefore have that  $n_j - t \in S_j$  for  $j = 1, 2, \dots, k$  and hence  $F \subset S + t$ .

We now construct the  $t_j$ . First by choosing  $\tau \geq 3/\delta$  we have  $\delta n \geq 3(\tau - 1)!$  and hence from the fact that  $n_1 \geq \delta n$  we have

$$s_1(\tau - 1)! \leq n_1 < (s_1 + 1)(\tau - 1)! \quad \text{with} \quad s_1 \geq 3.$$

We then define

$$t_1 = n_1 - s_1(\tau - 1)!$$

and note that  $0 \leq t < (\tau - 1)!$  Now for any  $j$  we write inductively that

$$s_j(\tau - 2j + 1)! \leq n_j - t_1 - t_2 - \dots - t_{j-1} < (s_j + 1)(\tau - 2j + 1)!$$

and define  $t_j$  by

$$t_j = n_j - t_1 - t_2 - \dots - t_{j-1} - s_j(\tau - 2j + 1)!.$$

We now note that

$$0 \leq t_j < (\tau - 2j + 1)!$$

and hence  $t_{j+1} + t_{j+2} + \dots + t_k < k(\tau - 2j - 1)!$ . By taking  $\tau \geq 3k$  we have for  $j = 1, 2, \dots, k - 1$

$$t_{j+1} + \dots + t_k < (\tau - 2j)!$$

Thus we have  $t_1 + t_2 + \dots + t_k < (\tau - 1)! + (\tau - 2)!$  so  $n_j - t_1 - t_2 - \dots - t_{j-1} > n_1 - (\tau - 1)! - (\tau - 2)! > (\tau - 1)!$ , and hence  $s_j \geq (\tau - 1)! / (\tau - 2j + 1)! > 0$ , for  $j = 2, 3, \dots, k$ . We have thus verified that  $S + t$  contains  $F$ . This has been done with  $t < (\tau - 1)! + (\tau - 2)!$  so  $t < \varepsilon_2 n$  provided  $\tau$  is chosen with  $\tau > 2/\varepsilon_2$ . Property (2) is thus completely verified.

For the fact that  $|S|/n < \varepsilon$  we note that for  $j = 1, 2, \dots, k$  we have  $|S_j| \leq (\tau - 2j)! \tau! / (\tau - 2j + 1)! \leq (\tau - 2k)! \tau! / (\tau - 2k + 1)!$  so  $|S| (\tau!)^{-1} \leq k / (\tau - 2k + 1)$ . The last requirement we place on  $\tau$  is that it be so large that  $k / (\tau - 2k + 1) < \varepsilon_1$  and the proof of the lemma is complete.

The difficulty of Lemma II.1. is tied to the finiteness of the interval  $[0, n]$ . For the infinite interval one can obtain analogous results more easily. Only Lemma II.1 will be used in the sequel, but we given an example of what a more precise, yet easier, result on  $[0, \infty)$ .

**THEOREM II.2.** *There is a subset  $S \subset [0, \infty)$  with the following properties:*

- (1)  $\lim_{n \rightarrow \infty} S \cap [0, n] / n = 0$ , i.e.  $S$  has zero density.
- (2) For any disjoint finite sets  $F$  and  $G$  there is a positive  $t$  such that  $F \subset S - t$  and  $G \cap (S - t) = \emptyset$ .

**Proof.** Let  $C_1, C_2, \dots, C_f$ , where  $f = 2^k$ , be a list of the  $2^k$  district strings of 0 and 1 of length  $k$ . Let  $C_0$  be a string of  $2^{2k}$  zeros, and  $l_k$  the string of 0's and 1's obtained by writing  $C_0$  followed by  $C_1, C_2$  etc. Finally let  $l$  be the infinite string obtained by writing  $l_2, l_3$ , etc. The set  $S$  is now given by

$$S = \{r \in [0, \infty) : \text{The } r\text{th element of } l \text{ is a } 1\}.$$

It is easily checked that  $S$  has zero density due to the large blocks of zeros  $C_0$ . Now given any finite disjoint sets  $F$  and  $G$  there is an  $h$  such that  $F \cup G \subset [0, h]$ , and since

$$c = (a_1, a_2, \dots, a_h), \quad a_i = \begin{cases} 1 & \text{if } i \in F \\ 0 & \text{otherwise,} \end{cases}$$

is equal to one of the blocks in  $l$  there is a  $t$  such that  $S - t$  contains  $F$  and is disjoint from  $G$ . Hence the construction is complete.

With these combinatorial observations, we can proceed to the main result.

**III. Aperiodic transformations and the main result.** We will be concerned with transformations  $T$  on a probability space  $(\Omega, \Sigma, \mu)$  which are aperiodic. That is, we assume that for any  $n$  and any  $B \in \Sigma$  with  $\mu(B) > 0$  there is a  $B' \subset B$

such that  $\mu(B' \Delta T^{-n}B') > 0$ . It will also be required that  $T$  is measure preserving, i.e. for  $A \in \Sigma$  we have  $T^{-1}A \in \Sigma$  and  $\mu(T^{-1}A) = \mu(A)$ . (Here, of course,  $A \Delta B$  denotes  $A \setminus B \cup B \setminus A$ , the symmetric difference.)

Any ergodic transformation is aperiodic, and for an example of an aperiodic transformation which is not ergodic one can take the transformation on  $[0, 1] \times [0, 1]$  defined by  $T(x, y) = ((x + \alpha) \bmod 1, y)$  where  $\alpha$  is irrational. The only fact which we use about aperiodic transformations is the following version of Rohlin's Theorem.

**LEMMA III.1.** *Let  $T$  be an aperiodic transformation on a non-atomic probability space. Given any  $\varepsilon > 0$  and any integer  $n \geq 1$ , there is an  $E \in \Sigma$  such that  $E, T^{-1}E, T^{-2}E, \dots, T^{-n}E$  are disjoint and  $\mu(\bigcup_{i=0}^n T^{-i}E) \geq 1 - \varepsilon$ .*

The proof of this result is essentially contained in [1, p. 71] and is also a special case of [2, p. 282]. Our main reason for focusing on the class of aperiodic transformations is the availability of Lemma III.1. We now have the main result of this paper.

**THEOREM III.1.** *Let  $T$  be an aperiodic, bimeasurable, measure preserving transformation on a non-atomic probability space. There is a set  $A$  of measure less than  $\varepsilon > 0$  such that for every finite set  $F$  there is a  $j = j(F)$  such that  $F \subset T^{-j}A$ .*

**Proof.** By the combinatorial lemma we select  $S_k \subset [0, n_k]$  such that

$$(4.1) \quad |S_k|/n_k \leq \varepsilon 2^{-k}, \text{ and}$$

$$(4.2) \quad \text{For any } k \text{ element set } F \subset [n2^{-k}, n] \text{ there is a positive integer } t \text{ such that } F \subset S_k + t \text{ where } t \leq n2^{-k}, n = n_k.$$

Now by Rohlin's theorem we obtain a set  $E_k$  such that  $E_k, T^{-1}E_k, T^{-2}E_k, \dots, T^{-n}E_k$  are disjoint and whose union has measure greater than  $1 - 2^{-k}$ . The basic building block of the construction is then defined by

$$(4.3) \quad A_k = \bigcup_{j \in S_k} T^{-j}E_k$$

We first observe that  $\mu(A_k) \leq |S_k|/n_k \leq \varepsilon 2^{-k}$  so one has  $\mu(\bigcup_{k=2}^{\infty} A_k) \leq \varepsilon$ . Next we let  $B(k)$  be defined by

$$(4.4) \quad B(k) = \left( \bigcup_{j=0}^n T^{-j}E_k \right)^c \cup \left( \bigcup_{j=0}^{n2^{-k}} T^{-j}E_k \right).$$

Since  $\mu(B(k)) \leq 2^{-k+1}$  we have for  $B = \bigcap_{j=2}^{\infty} \bigcup_{k=j}^{\infty} B(k)$  that  $\mu(B) = 0$ . The set  $A$  required by the theorem can now be taken to be

$$(4.5) \quad A = \left( \bigcup_{k=2}^{\infty} A_k \right) \cup \left( \bigcup_{j=0}^{\infty} T^j B \right).$$

By the previous estimates  $\mu(A) \leq \varepsilon$  so it remains only to show that  $A$  can

“catch” any finite set  $F$ . First let  $F_1 = B \cap F$  and note that  $B \subset T^{-j}A$  for all  $j$  so  $F_1 \subset T^{-j}A$  for all  $j$  also. Next let  $F_2 = F \cap B^c$ . Since  $F_2 \subset B^c$  there is a  $k_0$  such that  $F_2 \subset B(k)^c$  for all  $k \geq k_0$ . Now for  $F_2 \subset B(k)^c$  we define a subset of  $[n2^{-k}, n]$  by  $F' = \{m : (T^{-m}E) \cap F_2 \neq \emptyset\}$  and recall there is a  $t, 0 \leq t \leq n2^{-k}$  such that  $F' \subset S_k + t$ . Consequently we have

$$(4.6) \quad F_2 \subset \bigcup_{m \in F'} T^{-m}E \subset \bigcup_{m \in S_k+t} T^{-m}E \subset T^{-t}A_k.$$

For this choice of  $t$  we therefore have  $F = F_1 \cup F_2 \subset T^{-t}A$  so the theorem is complete.

An easy consequence of the preceding result is that the finite covering property actually characterizes the aperiodic transformations.

**THEOREM III.2.** *A bimeasurable, measure preserving transformation on a non-atomic probability space is aperiodic if and only if for  $\epsilon > 0$  there is an  $A$  with  $\mu(A) < \epsilon$  such that for every finite  $F$  there is a  $j = j(F)$  such that  $F \subset T^{-j}A$ .*

We have already proved that aperiodicity yields the finite covering property and the converse is a consequence of the following elementary lemma.

**LEMMA III.2.** *If for any  $\epsilon > 0$  there is a  $A$  with  $\mu(A) < \epsilon$  such that  $\mu(\bigcup_{i=0}^{\infty} T^{-i}A) = 1$  then  $T$  is aperiodic.*

**Proof.** Suppose  $T$  is not aperiodic and let  $B \in \Sigma$  denote an invariant set with  $0 < \mu(B) \leq 1$  for which there is an  $n$  such that  $\mu(B' \Delta T^{-n}B') = 0$  for all  $B' \subset B, B' \in \Sigma$ . If  $A$  is chosen as in the hypothesis, then  $\mu(B \Delta \bigcup_{i=0}^{\infty} T^{-i}(A \cap B)) = 0$  since  $B$  is invariant. But  $\mu(A \cap (B \Delta T^{-n}(A \cap B))) = 0$  so  $\mu(\bigcup_{i=0}^{\infty} T^{-i}(A \cap B)) \leq n\mu(A \cap B)$ , and consequently  $\mu(B) \leq n\mu(A \cap B) \leq n\mu(A)$ . Since  $A$  can be chosen independently of  $n$  and arbitrarily small, we have a contradiction to  $\mu(B) \neq 0$ . This contradiction proves the lemma.

**IV. Applications to combinatorial entropy.** If  $S$  is a class of subsets of  $\Omega$  and  $X_1, X_2, \dots, X_n, \dots$  is a sequence of independent identically distributed random variables which take values in  $\Omega$  we first define a new random variable as follows:

$$\Delta^S(X_1, X_2, \dots, X_n) = |\{B : B = A \cap \{X_1, X_2, \dots, X_n\}, A \in S\}|$$

i.e.  $\Delta^S(X_1, X_2, \dots, X_n)$  is the number of distinct subsets of the sample  $X_1, X_2, \dots, X_n$  obtained by intersection with an element of  $S$ .

The Vapnik–Chervonenkis entropy of the class  $S$  with respect to the random variables  $X_i$  is defined by the limit

$$h(S) = \lim_{n \rightarrow \infty} \frac{1}{n} \log E(\Delta^S(X_1, X_2, \dots, X_n)).$$

This limit always exists, as proved by Vapnik and Chervonenkis [4] and [5].

The most important consequence of this definition is that the random variables  $D_n = \sup_{A \in \mathcal{S}} |n^{-1} \sum_{i=1}^n 1_A(X_i) - P(X_1 \in A)|$  converges to zero in probability if and only if  $h(S) = 0$ . (Here  $1_A$  denotes the indicator of the set  $A$ .) This result was obtained in [4], and in [3] it was proved that, in fact,  $D_n$  converges to zero a.s. if and only if  $h(S) = 0$ .

Now, given a transformation  $T$  it is especially reasonable to consider the class  $\mathcal{S}$  defined by  $\mathcal{S} = \{T^{-i}A\}_{i=0}^{\infty}$ . Once this is done there is the hope of comparing the Vapnik–Chervonenkis entropy of  $\mathcal{S}$  with the Kolmogorov entropy of  $T$  and the partition  $(A, A^c)$ . One aspect of these hopes, that zero Kolmogorov entropy implies zero V-C entropy, is dashed by the following.

**THEOREM IV.1.** *If  $T$  is any ergodic transformation there is an  $A$  such that the V-C entropy of  $\mathcal{S} = \{T^{-i}A\}_{i=1}^{\infty}$  is positive with respect to the i.i.d. random variables  $X_i$   $i = 1, 2, \dots$  with uniform distribution (i.e.  $P(X_i \in A) = \mu(A)$ ).*

**Proof.** One simply notes that by Theorem III.2. there is an  $A$  such that  $\mathcal{S} = \{T^{-i}A\}_{i=0}^{\infty}$  covers almost every finite sample  $\{X_1, X_2, \dots, X_n\}$ . If  $A$  is also chosen so that  $\mu(A) \leq \frac{1}{2}$  we have

$$(4.1) \quad D_n = \sup_{A \in \mathcal{S}} \left| n^{-1} \sum_{i=1}^n 1_A(X_i) - P(X_i \in A) \right| \geq \frac{1}{2}$$

and hence by the main theorem of Vapnik and Chervonenkis we have  $h(S) \neq 0$ . To complete the comparison of Kolmogorov entropy and Vapnik–Chervonenkis entropy we add two observations due to M. Ellis (personal communication). The first is that a transformation with positive Kolmogorov entropy need not have positive V-C entropy for all  $A$ , and the second is that a transformation with completely positive entropy (or  $K$ -automorphism) must have positive V-C entropy for all  $A$ .

To prove the first observation let  $S$  be a transformation on  $[0, 1]$  with positive  $K$ -entropy. If  $I$  is the identity on  $[0, 1]$ , let  $T$  be defined by  $T = S \times I$ .  $T$  can be checked to have positive  $K$ -entropy and be aperiodic, but  $[0, 1] \times [0, \frac{1}{2}]$  is a fixed set under  $T$  which hence has V-C entropy 0.

The transformation constructed above naturally has a factor of zero entropy, and if one rules out this pathology one has that all non-trivial  $A$  have positive V-C entropy. Specifically, if  $T$  has no factors of zero entropy it is known that  $T$  must be mixing [1, 121]. But for  $T$  mixing one has the  $d$ -fold product  $S = T \times T \times \dots \times T$  is ergodic, so for any  $A$  with  $0 < \mu(A) < 1$  one has for  $B = A \times A \times \dots \times A$  one has  $\bigcup_{j=1}^{\infty} S^{-j}B$  has product measure 1. By the criterion of (4.1) one then sees  $A$  has positive V-C entropy. This completes the proof of the second observation.

#### REFERENCES

1. P. Billingsley, *Ergodic Theory and Information*, Wiley, New York, 1965.
2. P. R. Halmos, *Lectures on Ergodic Theory*, Chelsea Publishing, New York, 1956.

3. L. K. Jones and U. Krengel, *On transformations without invariant measure*, *Adv. in Math.*, **12** (1974), 275–295.
4. J. M. Steele, *Combinatorial Entropy and Uniform Limit Laws*, Stanford University (Thesis), 1975.
5. V. N. Vapnik and A. Ya. Chervonenkis, *On the uniform convergence of relative frequencies of events to their probabilities*, *Theory Prob. and Appl.*, **16** (1971), 264–280.
6. —, *Teoriya Raspoznavaniya Obrazov*, Nauka, Moscow, 1974.

DEPARTMENT OF MATHEMATICS  
UNIVERSITY OF BRITISH COLUMBIA  
2075 WESTBROOK PLACE  
VANCOUVER, B.C., V6T 1W5