

## RECURRENCE DENSITIES FOR STATIONARY PROCESSES

MICHAEL HOCHMAN

The purpose of this note is to present in simple manner one of my main results in [1], which unfortunately was obscured by the generality and length of that paper.

Let  $(X_n)_{n=1}^\infty$  be a stationary and ergodic process on a finite alphabet. For  $i \leq j$  the block of symbols from  $i$  to  $j$  in the sequence  $(X_n)$  is denoted  $X_i^j = X_i X_{i+1} \dots X_j$ . The first recurrence time of the initial  $n$ -block is the first time the block  $X_1^n$  repeats in the sample  $X_1^\infty$ , i.e.

$$R_n = \min\{k > 0 : X_0^n = X_k^{k+n}\}$$

The Wyner-Ziv/Ornstein-Weiss recurrence time theorem [3, 2] states that with probability 1,

$$\frac{1}{n} \log R_n \rightarrow h$$

where  $h = h((X_n)_{n=1}^\infty)$  is the entropy of the process (by convention, all logarithms are in base 2).

Our main concern here will be another quantity, the frequency of recurrences of the initial  $n$ -block in longer block of the sample. More precisely, given  $N \geq n$  we define

$$F_{n,N} = \frac{1}{n} \#\{1 \leq k \leq N : X_0^n = X_k^{k+n}\}.$$

By the ergodic theorem, with probability one we have

$$F_{n,\infty} := \lim_{N \rightarrow \infty} F_{n,N} = P(X_0^n)$$

and, since by the Shannon-McMillan-Breiman theorem  $-\frac{1}{n} \log P(X_0^n) \rightarrow h$  as  $n \rightarrow \infty$ , we also have

$$\lim_{n \rightarrow \infty} -\frac{1}{n} \log F_{n,\infty} = h$$

Notice that

$$F_{n,R_n} = \frac{2}{R_n}$$

(and  $F_{n,R_n-1} = 1/R_n$ ), so by the WZ/OW theorem we also have

$$\lim_{n \rightarrow \infty} -\frac{1}{n} \log F_{n,R_n} = h$$

To summarize, as  $n \rightarrow \infty$  the frequency of the initial block  $X_0^n$  is asymptotically correct both in the relatively short block  $X_1^{R_n}$  (which is essentially in fact the shortest block in which it has a chance of being correct), and also in the infinite

sample  $X_1^\infty$ . It is now natural to ask about the behavior of these frequencies for intermediate sample sizes, i.e.  $F_{n,m}$  where  $R_n < m < \infty$ .

**Theorem 1.** *If  $(X_n)$  is an ergodic, stationary process on a finite symbol set then*

$$\lim_{n \rightarrow \infty} \left( \sup_{R_n < i, j < \infty} \left| \frac{1}{n} \log F_{n,i} - \frac{1}{n} \log F_{n,j} \right| \right) = 0$$

consequently if  $i(n)$  is any sequence such that  $2^{-hn}i(n) \rightarrow \infty$  then

$$-\frac{1}{n} \log F_{n,i(n)} \rightarrow h = \text{entropy of } (X_n)$$

The proof splits in two parts, an upper bound and a lower bound. Define

$$U_n^+ = \sup_{i > R_n} -\frac{1}{n} \log F_{n,i}$$

and

$$U_n^- = \inf_{i > R_n} -\frac{1}{n} \log F_{n,i}$$

The theorem is then equivalent to the equalities  $U_n^\pm \rightarrow h$ . Clearly  $U_n^- \leq U_n^+$  so we must show that  $\liminf U_n^- \geq h$  and  $\limsup U_n^+ \leq h$ . The first of these inequalities is easier, and requires only classical entropy theory. The second is not much harder, but relies on the more sophisticated WZ/OW result. We shall prove each of these separately, following some preliminaries in the next section.

## 1. PRELIMINARIES

An interval  $[m, n]$  shall always be understood to be an interval of integers, i.e.  $m \leq n$  are integers and  $[m, n] = \{i \in \mathbb{N} : m \leq i \leq n\}$ .

**Lemma 2.** *Let  $\{[x_i, x_i + r_i]\}_{i=1}^\infty$  be a (finite or infinite) collection of intervals. Then there is a subset  $I \subseteq \mathbb{N}$  such that*

- (1) *The intervals  $[x_i, x_i + r_i]$ ,  $i \in I$ , are pairwise disjoint,*
- (2) *The intervals  $[x_i, x_i + r_i]$ ,  $i \in I$ , cover all the  $x_j$ 's.*

*Proof.* We may assume each integer occurs at most once as an  $x_i$ , and that the intervals are ordered so that  $x_1 < x_2 < \dots < x_n$ . Choose  $1 \in I$ , and inductively add  $i$  to  $I$  if  $x_i \notin \cup_{j < i} [x_j, x_j + r_j]$ . The resulting collection is seen to have the desired property.  $\square$

This rather trivial lemma has a big brother called the Besicovitch covering lemma, which holds in  $\mathbb{Z}^d$  when intervals are replaced by symmetric cubes and, more generally, by balls induced by a fixed norm in  $\mathbb{R}^d$ . In the higher-dimensional version (and also the one-dimensional one, when one uses intervals in which  $x_i$  is at the center) the formulation involves a constant which does not appear above because it equals 1 in our special case. Note that one-sided cubes, i.e. cubes in

which the “base point” is taken to be one of the corners, do not satisfy the Besicovitch property in higher dimensions; dimension 1 is special in this regard. Most of the complications in [1] came from the attempt to find suitable analogs to this lemma in other groups; there are not many groups that possess such sequences that are also Følner sequences, though some weaker conditions were proposed in [1] that can serve as partial substitutes for the purposes we have in mind. We remark that the Besicovitch property has other applications in ergodic theory. For example, it has been used to prove the maximal inequality for commuting non-singular actions, and plays a role in my recent proof of the ratio ergodic theorem for such actions.

The main application of this lemma comes in two forms, one combinatorial and the other ergodic:

**Proposition 3.** *Let  $A \subseteq \mathbb{N}$  be a set. Define  $s_n : \mathbb{N} \rightarrow \mathbb{R}$  by*

$$s_n(i) = \frac{|A \cap [i, i+n]|}{n+1}$$

and  $m : \mathbb{N} \rightarrow \mathbb{R}$  by

$$m(i) = \inf_n s_n(i)$$

Then

$$\liminf_{N \rightarrow \infty} \frac{|\{0 \leq i \leq N : i \in A \text{ and } m(i) < \varepsilon\}|}{N+1} \leq \varepsilon$$

*Proof.* Let  $I \subseteq A$  be the set of  $i$  with  $m(i) < \varepsilon$ , and for each such  $i$  let  $r_i$  satisfy  $\frac{1}{r_i+1}|A \cap [i, i+r_i]| < \varepsilon$ . Either  $I$  is unbounded or there is nothing to prove.

For each integer  $k$  consider  $I_k = I \cap [0, k]$ , let

$$E_k = [0, \max_{i \in I_k} (i+r_i)]$$

and

$$I'_k = I \cap E_k$$

Fixing  $k$  and applying the lemma to the collection  $[i, i+r_i], i \in I_k$  we get a disjoint collection of intervals covering  $I \cap E_k$ , and in each interval the relative density of  $A$  is  $< \varepsilon$ . Thus

$$\frac{|I'_k|}{|U_k|} < \varepsilon$$

This ratio appears as a term in the lim inf we are trying to estimate. The result follows by taking  $k \rightarrow \infty$ . □

Here is the dynamical (or analytic) consequence that we shall use.

**Proposition 4.** *Let  $(\Omega, \mathcal{B}, \mu, T)$  be a measure preserving system and  $B \subseteq \Omega$  measurable. Let*

$$s_n = \frac{1}{n} \sum_{i=0}^n T^i \circ 1_B$$

and let

$$m = \inf_n s_n$$

Then

$$\mu\{\omega \in B : m < \varepsilon\} < \varepsilon$$

*Proof.* Fix a typical  $\omega \in \Omega$  set  $A = \{i \geq 0 : m(T^i\omega) < \varepsilon\}$ . Then  $m(T^j\omega) = m(j)$ , where in the right hand side we use the notation of the previous theorem, and  $m$  defined as in that theorem with respect to our choice of  $A$ . By the previous theorem the  $\liminf$  is, which is bounded above by  $\varepsilon$ , and by the ergodic theorem it is a limit and is equal to the measure we are trying to estimate.  $\square$

## 2. ERGODIC SETUP AND NOTATION

For notational convenience we assume henceforth that  $X_n$  arises from an ergodic measure-preserving transformation  $T$  on a probability space  $(\Omega, \mathcal{B}, \mu)$ , so  $X_n = X \circ T^n$  for some measurable finite-valued function  $X$  defined on  $\Omega$ . The random variables  $R_n, F_{n,N}$  and  $U_n^\pm$  are thus defined on  $\Omega$ . Let  $h$  denote the entropy of the process.

Let  $\Sigma$  be the finite set of values attained by the process, and  $\Sigma^n$  the set of blocks of length  $n$  over  $\Sigma$ . For a block  $a \in \Sigma^n$ , the cylinder set defined by  $a$  is

$$[a] = \{\omega \in \Omega : X_0^{n-1}(\omega) = a\}$$

More generally we write  $[\Phi] = \cup_{a \in \Phi} [a]$  for  $\Phi \subseteq \Sigma^*$ .

By the Shannon-McMillan-Breiman theorem, we may choose a sequence of sets  $\Phi_n \subseteq \Sigma^n$  such that

- (1) Almost every  $\omega \in \Omega$  is eventually in  $[\Phi_n]$ .
- (2)  $-\frac{1}{n} \log \mu([a_n]) \rightarrow h$  for any sequence  $a_n \in \Phi_n$ .
- (3)  $\frac{1}{n} \log |\Phi_n| \rightarrow h$  (this follows from (1) and (2)).

## 3. PROOF THAT $\limsup_{n \rightarrow \infty} U_n^+ \leq h$

Fix  $\varepsilon > 0$  and let  $A = A_\varepsilon$  be the event

$$A = \{\omega \in \Omega : \limsup_n U_n^+ > h + \varepsilon\}$$

It suffices to show that  $\mu(A) = 0$ .

Choose  $N$  large enough so that for every  $n > N$  we have  $|\Phi_n| < 2^{(h+\varepsilon/2)n}$  and for every  $a \in \Phi_n$ ,

$$2^{-(h+\varepsilon/2)n} < \mu(a) < 2^{-(h-\varepsilon/2)n}$$

For such  $n, a$  let

$$\begin{aligned} A_{n,a} &= \{\omega \in [a] : U_n^+ > h + \varepsilon\} \\ &= \{\omega \in [a] : \inf_{i > R_n} F_{n,i} < 2^{-(h+\varepsilon)n}\} \end{aligned}$$

For  $\omega \in [a]$  the quantity  $F_{n,i}$  is just an ergodic average of the function  $1_{[a]}$ , so by proposition 4,

$$\mu(A_{n,a}) < 2^{-(h+\varepsilon)n}$$

Hence

$$\mu\left(\bigcup_{a \in \Phi_n} A_{n,a}\right) \leq \sum_{a \in \Phi_n} 2^{-(h+\varepsilon)n} \leq |\Phi_n| \cdot 2^{-(h+\varepsilon)n} \leq 2^{-\varepsilon n/2}$$

so, writing

$$A_N = \bigcup_{n > N} \bigcup_{a \in \Phi_n} A_{n,a}$$

we have

$$\mu(A_N) \leq \sum_{n > N} \mu\left(\bigcup_{a \in \Phi_n} A_{n,a}\right) = O_\varepsilon(1) \cdot 2^{-\varepsilon N/2}$$

This is summable, so almost every  $\omega \in \Omega$  belongs only to finitely many of the sets  $A_N$ . Since a.e.  $\omega \in \Omega$  is in all but finitely many  $[\Phi_n]$ , it follows that  $A$  is precisely the set of  $\omega$  that visit infinitely many  $A_N$ 's; hence  $\mu(A) = 0$  as required.

#### 4. PROOF THAT $\liminf_{n \rightarrow \infty} U_n^- \geq h$

Fix  $\varepsilon > 0$  and let  $B = B_\varepsilon$  be the set

$$B = \{\omega \in \Omega : \liminf U_n^- < h - \varepsilon\}$$

We must show that  $\mu(B) = 0$ . Fix  $N$  so that for  $n > N$  we have  $|\Phi_n| < 2^{(h+\varepsilon/100)n}$ . For  $n > N$  and  $a \in \Sigma^n$  let

$$B_{n,a} = \{\omega \in [a] : U_n^- < h - \varepsilon\}$$

It suffices to show that a.e.  $\omega$  belongs to finitely many of the  $B_{n,a}$ .

Split  $B_{n,a}$  in two:

$$B'_{n,a} = \{\omega \in B : R_n < 2^{(h-\varepsilon/100)n}\}$$

and

$$B''_{n,a} = B_{n,a} \setminus B'_{n,a}$$

By the WZ/OW theorem, a.e.  $\omega$  belongs to only finitely many  $B'_{n,a}$ 's; so we must only show the same  $B''_{n,a}$ .

Let  $\omega \in B''_{n,a}$  and let  $i$  be such that  $-\frac{1}{n} \log F_{n,i} < 2^{-(h-\varepsilon)n}$ . By definition of  $B''_{n,a}$  we know that  $i > 2^{(h-\varepsilon/100)n}$ . Let  $J = J(\omega) \subseteq [0, i]$  be the set

$$J = \{0 \leq j \leq i : R_n(T^j \omega) > 2^{(h-\varepsilon/2)n}\}$$

thus for  $j \in J$  we may set  $r_j = 2^{(h-\varepsilon/2)n}$  and we have  $F_{n,r_j}(T^j\omega) = 2^{-(h-\varepsilon/2)n}$ . From proposition 3 we deduce that

$$\frac{|J|}{[0, i + 2^{(h-\varepsilon/2)n}]} < 2^{-(h-\varepsilon/2)n}$$

since  $i > 2^{(h-\varepsilon/100)n}$  we conclude that for large enough  $n$ ,

$$\frac{|J|}{i} < 2 \cdot 2^{-(h-\varepsilon/2)n}$$

Consider now a typical  $\omega$  and the set  $I \subseteq \mathbb{N}$  so that  $T^i\omega \in B''_{n,a}$  for  $i \in I$ . For  $i \in I$  there is by definition an  $r_i > R_n(T^i\omega)$  so that  $F_{n,r_i}(T^i\omega) > 2^{-(h-\varepsilon)n}$ . Using lemma 2 we find a subcollection  $I_0 \subseteq I$  so that  $I \subseteq \cup_{i \in I_0} [i, i + r_i]$ , and the intervals  $[i, i + r_i]$ ,  $i \in I_0$ , are pairwise disjoint. For each  $i \in I_0$  we have seen that

$$\frac{1}{r_i} \#\{j \in [i, i + r_i] : R_n(T^j\omega) > 2^{(h-\varepsilon/2)n}\} < 2 \cdot 2^{-(h-h/2)n}$$

Summing these densities over all of  $i \in I_0$  and applying the ergodic theorem to assure convergence, we have that

$$\lim_{N \rightarrow \infty} \frac{1}{N} \#\{0 \leq j \leq N : T^j\omega \in B''_{n,a} \text{ and } R_n(T^j\omega) > 2^{(h-\varepsilon/2)n}\} < O(1)2^{-(h-\varepsilon/2)n}$$

so by the ergodic theorem again,

$$\mu\{\omega \in B''_{n,a} : R_n(\omega) > 2^{(h-\varepsilon/2)n}\} < O(1) \cdot 2^{-(h-\varepsilon/2)n}$$

We claim that these probabilities are summable for  $a \in \Phi_n$ , because

$$\sum_{n > N} \sum_{a \in \Phi_n} \mu(\dots) < \sum_{n > N} |\Phi_n| \mu(\dots) < \sum_{n > N} 2^{(h+\varepsilon/100)n} \cdot O(1) \cdot 2^{-(h-\varepsilon/2)n} < \infty$$

Hence  $\omega$  is in infinitely many  $B''_{n,a}$ 's it must either satisfy  $\omega \notin [\Phi_n]$  infinitely often, which is impossible by choice of  $\Phi_n$ , or  $R_n(\omega) < 2^{(h-\varepsilon)n}$  infinitely often, which contradicts the WZ/OW theorem again. This completes the proof.

#### REFERENCES

- [1] Michael Hochman. Return times, recurrence densities and entropy for actions of some discrete amenable groups. *J. Anal. Math.*, 100:1–51, 2006.
- [2] Donald Samuel Ornstein and Benjamin Weiss. Entropy and data compression schemes. *IEEE Trans. Inform. Theory*, 39(1):78–83, 1993.
- [3] Aaron D. Wyner and Jacob Ziv. Some asymptotic properties of the entropy of a stationary ergodic data source with applications to data compression. *IEEE Trans. Inform. Theory*, 35(6):1250–1258, 1989.