Billingsley dimension on shift spaces

C.-E. Pfister† and W.G. Sullivan‡

†Ecole Polytechnique Fédérale, Institut de Mathématiques, CH-1015 Lausanne, Switzerland

‡University College, Department of Mathematics, Belfield, Dublin 4, Ireland, and Dublin Institute of Technology, Rathmines Road, Dublin 6, Ireland

E-mail: `Charles.Pfister@epfl.ch`

E-mail: `Wayne.Sullivan@dit.ie`

Abstract. We consider a class of subshifts Σ over a finite alphabet, including sofic shifts. For a large class of metrics we determine the Hausdorff dimension of sets of points of Σ defined by their limit point set under the empirical measure. Our approach to computing Billingsley dimension of saturated sets is fundamentally different and applies to more general shift spaces and measures than the technique of Billingsley, which was significantly developed by Cajar and recently extended by others. One main feature of our approach is an algorithmic construction of a large (in the sense of dimension theory) subset of a saturated set. This generalizes similar constructions of subsets of normal or generic points.

AMS classification scheme numbers: 11K55, 37C35

Submitted to: *Nonlinearity*

1. Introduction

In this paper we develop a new approach to compute Billingsley dimension of saturated sets. Basic references about Billingsley dimension are [B1], [B2], where this dimension is introduced, [W] and Cajar's monograph [C], which is devoted to this topic. Billingsley introduced this dimension in a probability space (Σ, ν) in analogy with Hausdorff dimension in a metric space. However, for spaces of sequences based on a finite alphabet $\mathtt{A}$ and when the probability measure ν is non-atomic, which is the case considered here, there exists a semi-metric d_ν on Σ so that the Billingsley dimension of a subset B of the probability space (Σ, ν) is equal to the Hausdorff dimension of B in the semi-metric space (Σ, d_ν) [W]. Dimension theory plays an important role in the modern theory of dynamical systems. A good reference is the monograph [P], which however does not mention the work of Billingsley. See also [T1].

Let Σ denote a compact metric space. Let $S\colon \Sigma \to \Sigma$ be a continuous mapping. For $\omega \in \Sigma$ we use δ_ω to denote the Dirac measure, which is concentrated on the single point ω. The empirical measure (of order n) is

$$T_n(\omega) := \frac{1}{n}\sum_{k=0}^{n-1} \delta_{S^k(\omega)}\,. \tag{1.1}$$

The sequence $\{T_n\}_n$ has limit points, which are S-invariant probability measures. The set of limit points of this sequence is a compact connected set (weak* topology). Saturated sets are subsets of Σ of the form

$$G_F := \{\omega \in \Sigma\colon \text{the set of all limit points of } \{T_n(\omega)\}_n \text{ is } F\}\,,$$

where F is compact and connected, or an arbitrary union of sets of the above form. If $F = \{\alpha\}$, then $\omega \in G_{\{\alpha\}}$ is called a generic point for the stationary probability measure α [DGS]. Note that if F is not a singleton, then G_F contains *no* generic points. It is known that $\alpha[G_{\{\alpha\}}] = 1$ for all ergodic α, and $\alpha[G_{\{\alpha\}}] = 0$ otherwise. The fundamental problem considered here is how "large" is G_F for specified F. In this paper Σ is a S-invariant compact subset of $\mathtt{A}^{\mathbf{N}}$, where $\mathtt{A} = \{0, 1, \ldots, r-1\}$ is a finite alphabet with r characters, and S is the shift operator. A special and interesting case is when $\Sigma = \mathtt{A}^{\mathbf{N}}$ and $\nu = \beta$, with β the probability measure on $\mathtt{A}^{\mathbf{N}}$, which is product over $\mathbf{N}$ of the probability giving equal weight $1/r$ to all symbols of $\mathtt{A}$. The mapping,

$$\omega = (\omega_1, \omega_2, \ldots) \mapsto I(\omega) := \sum_{k=1}^{\infty} \omega_k r^{-k}\,,$$

maps $\mathtt{A}^{\mathbf{N}}$ to $[0,1]$. Eggleston [Eg] showed that the Hausdorff dimension of

$$I(G_{\{\rho\}}) = \{x \in [0,1]\colon x = I(\omega)\,,\, \omega \in G_{\{\rho\}}\}\,,$$

as subset of the metric space $[0,1]$ with metric $d(x,y) = |x-y|$, is equal to the Shannon entropy of ρ computed using $\log_r$, where ρ is the product of probability distributions on $\mathtt{A}$ which give unequal weights to the points of $\mathtt{A}$. Eggleston's result can be extended to the case in which ρ is ergodic, and Colebrook [Co] showed that the Hausdorff dimension of $I(G_F)$, where G_F is a saturated set of $\mathtt{A}^{\mathbf{N}}$, is equal to the minimum Shannon entropy of the measures in F. The relevance of these results in our context is that the Hausdorff dimension of $I(G_F) \subset [0,1]$ is equal to the Billingsley dimension of G_F, as subset of the probability space $(\mathtt{A}^{\mathbf{N}}, \nu)$, when $\nu = \beta$ (see e.g. section 14 in [B3]). In [C] Cajar investigated the case when ν is an ergodic Markov chain. Central to Cajar's approach is the study of limits of the form

$$\liminf_n \frac{\log \nu([\omega_1^n])}{\log \rho([\omega_1^n])}.$$

The power of this approach is shown by the fact that is applicable to the dimension of uncountable unions. A short partial list of recent works about Billingsley dimension is [R], [O1], [O2], [Cha], [ChaO]. In [T1] and [T2] closely related dimensions are studied (see comment 4 in subsection 2.3). In [ChaO] the results of Cajar are extended to the case of g-measure on aperiodic subshifts of finite type by approximating the g-measure by n-step Markov chains and using Cajar's results for the n-step Markov chains.

Our approach to compute Billingsley dimension of saturated sets is fundamentally different from [C] or from the above references. It is inspired by large deviations ideas‡ and energy/entropy arguments of Statistical Mechanics. It includes all cases considered by Cajar, where ν is an ergodic (possibly periodic) Markov chains, as well as the generalizations in [ChaO]. But it also includes cases which are not considered by Cajar or Chazottes and Olivier, for example when ν is the maximal entropy measure of an irreducible sofic shift§. Sofic shifts form an important class of symbolic dynamical systems, which also play a fundamental role in finite-state automata theory. See [LM]. Our main result is theorem 2.2, which gives an explicit formula for the Billingsley dimension $\dim_\nu(G_F)$ of a saturated set G_F,

$$\dim_\nu(G_F) = \inf_{\alpha\in F} \frac{h_{\rm Sh}(\alpha)}{\langle\, e_\nu, \alpha\,\rangle}\,.$$

In this formula $h_{\rm Sh}(\alpha)$ is the Shannon entropy of α and e_ν is a continuous function associated with the reference measure ν; it is defined in hypothesis H3 of subsection 2.3. When ν is the measure of maximal entropy of an irreducible sofic shift Σ, the above formula becomes

$$\dim_\nu(G_F) = \inf_{\alpha\in F} \frac{h_{\rm Sh}(\alpha)}{h_{\rm top}(\Sigma)}\,,$$

where $h_{\rm top}(\Sigma)$ is the topological entropy of the shift space Σ. The topological entropy $h_{\rm top}(\Sigma)$ is equal to $h_{\rm Sh}(\nu)$. Although we cannot apply our approach directly to uncountable unions, we can find a set which contains the union and has the desired dimension (see theorem 2.3). The core and difficult part of our method is an explicit (algorithmic) construction of a subset of a given saturated set, which is large from the point of view of dimension theory, in the sense that the dimension of the subset is equal to the dimension of the saturated set. This construction is given in subsection 4.1 and the results are summarized in proposition 4.1. This result is already new and important by itself. It is a non trivial generalization of a long list of results about construction of generic points or normal numbers, starting with the seminal paper of Champernowne [Ch]. See also the interesting paper of Lebesgue [L]. We make hypotheses (see section 2) on the reference probability measure ν and the shift space corresponding to the support of ν. A central assumption is that the shift space Σ is such that for any neighbourhood U of any shift-invariant probability measure α on Σ (weak* topology) and for any $\varepsilon > 0$, there exists an ergodic probability measure $\alpha' \in U$, such that the Shannon entropy $h_{\rm Sh}(\alpha') \geq h_{\rm Sh}(\alpha) - \varepsilon$. We prove in section 5 that irreducible sofic shifts have this property and that theorem 2.2 applies to the measure ν of maximal entropy of an irreducible sofic shift; one can show that theorem 2.2 also applies to an equilibrium state relative to ν of an absolutely convergent potential.

‡ A large deviation result is proved in [ChaO], but their approach to Billingsley dimension uses Cajar's results.
§ Transitive sofic shifts in the terminology of [F1].

2. Definitions, notations and main results

2.1. Notations

The cardinality of any finite set C is denoted by $|C|$. Let $\mathtt{A} := \{0, 1, \ldots, r-1\}$ (with discrete topology) and $\Omega := \mathtt{A}^{\mathbf{N}}$ be the product space. We use the notations $[i,j] := \{k \in \mathbf{N}\colon i \leq k \leq j\}$ $(i \leq j)$, $X_{[i,j]}$ for the projection from Ω to $\mathtt{A}^{[i,j]}$, and the abbreviation $X_n := X_{[1,n]}$. Elements of $\mathtt{A}^{[i,j]}$ are denoted by ω_i^j. The shift operator $S\colon \Omega \to \Omega$, $(S\omega)_n := \omega_{n+1}$, $n \in \mathbf{N}$, acts on functions by $(Sf)(\omega) := f \circ S(\omega)$.

Let $f\colon \Omega \to \mathbf{R}$. The sup-norm of f is $\|f\| := \sup_\omega |f(\omega)|$. A function f is local if there exists $1 \leq i \leq j < \infty$ such that $f(\omega) = f(\eta)$ whenever $X_{[i,j]}(\omega) = X_{[i,j]}(\eta)$. The σ-algebra generated by X_n is denoted by $\mathcal{F}_n$; the same symbol also denotes the set of local functions which are $\mathcal{F}_n$-measurable. The family of cylinder sets‖ $\mathcal{S}$ is the collection of all sets

$$[b_1^n] := \{\omega \in \Omega\colon \omega_1^n = b_1^n\}\,,\ b_1^n \in \mathtt{A}^{[1,n]}\,.$$

The space of Borel probability measures on Ω is $\mathcal{M}$, and $\mathcal{M}_S$ is the subset of shift-invariant probability measures. The integral of f with respect to α is denoted by $\langle f, \alpha \rangle$. The shift S acts on measures by $\langle f, S\alpha \rangle := \langle Sf, \alpha \rangle$. The empirical measure $T_n(\omega)$ is defined in (1.1). The Shannon entropy of $\alpha \in \mathcal{M}_S$ is

$$h_{\mathrm{Sh}}(\alpha) := \lim_{n\to\infty} \frac{1}{n} H_n(\alpha) \text{ with } H_n(\alpha) := -\sum_{\omega_1^n \in \mathtt{A}^{[1,n]}} \alpha([\omega_1^n]) \log \alpha([\omega_1^n])\,.$$

The topology on $\mathcal{M}$, which coincides with the weak* topology, is given by the norm

$$\|\rho\| := \sum_{n=1}^{\infty} 2^{-n} \|\rho\|_{\mathrm{TVn}}\,, \quad \text{where} \quad \|\rho\|_{\mathrm{TVn}} := \sup_{f \in \mathcal{F}_n,\, \|f\| \leq 1} |\langle f, \rho \rangle|\,.$$

It is also convenient to use the semi-norms

$$\|\rho\|_M := \sum_{n=1}^{M} 2^{-n} \|\rho\|_{\mathrm{TVn}}.$$

2.2. Billingsley dimension

Let $\nu \in \mathcal{M}_S$ be a non-atomic probability measure. For any $s \in \mathbf{R}^+$, $\delta > 0$ and $B \subset \Omega$, set

$$\mathcal{C}_\delta^s(B) := \inf \left\{ \sum_j \nu[U_j]^s\colon \{U_j\} \text{ countable cover of } B\,,\ U_j \in \mathcal{S}\,,\ \nu[U_j] \leq \delta \right\}$$

and

$$\mathcal{C}^s(B) := \lim_{\delta \downarrow 0} \mathcal{C}_\delta^s(B)\,.$$

The Billingsley dimension $\dim_\nu(B)$ of B is

$$\dim_\nu(B) := \inf\{s\colon \mathcal{C}^s(B) = 0\} = \sup\{s\colon \mathcal{C}^s(B) = \infty\}\,.$$

‖ Same convention as in [B1].

Since $\mathcal{C}^s$ is an outer measure, for a countable family of sets B_k, $k \in \mathbf{N}$,

$$\dim_\nu\Big(\bigcup_{k\in\mathbf{N}} B_k\Big) = \sup_{k\in\mathbf{N}} \dim_\nu(B_k)\,.$$

There is a natural way of expressing the Billingsley dimension as a Hausdorff dimension [W]. Let d_ν be the semi-metric

$$d_\nu(\omega,\eta) := \begin{cases} 0 & \text{if } \omega=\eta;\\ 1 & \text{if } \omega \text{ and } \eta \text{ have no common prefix;}\\ \nu[b_1^k] & \text{if } b_1^k \text{ is the largest common prefix in } \omega \text{ and } \eta. \end{cases}$$

One has $\mathrm{diam}([b_1^n]) = \nu([b_1^n])$; hence $\dim_\nu(B)$ is equal to the Hausdorff dimension of B as subset of the semi-metric space (Ω, d_ν). It is natural to introduce the shift-invariant compact subset Σ^ν,

$$\Sigma^\nu := \mathrm{support}(\nu) := \{\omega\in\Omega\colon \nu([\omega_1^n])>0\,,\ \forall\, n\in\mathbf{N}\}\,.$$

The topology of the metric space (Σ^ν, d_ν) coincides with the induced topology on $\Sigma^\nu \subset \Omega$ (see e.g. [E]). We have $\dim_\nu(B) = \dim_\nu(B\cap\Sigma^\nu)$, since $\Omega\backslash\Sigma^\nu$ can be covered by a countable union of sets $A\in\mathcal{S}$ with $\nu(A)=0$. Therefore $\dim_\nu(B)$ is equal to the Hausdorff dimension of $B\cap\Sigma^\nu$ as subset of the metric space (Σ^ν, d_ν).

2.3. Main results

Let $\nu\in\mathcal{M}_S$ be a given non-atomic probability measure. The metric space (Σ^ν, d_ν) is defined in subsection 2.2, $\mathcal{M}^\nu$ denotes the set of Borel probability measures on (Σ^ν, d_ν), and $\mathcal{M}_S^\nu$ the shift-invariant probability measures on Σ^ν. We also denote by $[b_1^n]$ the set $\{\omega\in\Sigma^\nu\colon \omega_1^n = b_1^n\}$. A finite string of length $n\geq 1$, $b_1^n\in\mathtt{A}^n$, is a word of length n of Σ^ν if and only if there exists $\omega\in\Sigma^\nu$ such that $X_{[1,n]}(\omega)=b_1^n$. The set of words of length n of Σ^ν is $\mathcal{L}_n^\nu$; $\mathcal{L}_0^\nu := \{\epsilon\}$, where ϵ is the empty word of length 0. Notice that for any $b_1^n\in\mathcal{L}_n^\nu$ and any $i\in\mathbf{N}$, there exists $\omega\in\Sigma^\nu$, such that $\omega_{i+k-1}=b_k$ for all $k\in[1,n]$. The shift space Σ^ν is defined by the forbidden words b_1^n such that $\nu([b_1^n])=0$, $n\geq 1$. Its language is $\mathcal{L}^\nu = \cup_{n\geq 0}\mathcal{L}_n^\nu$. Our main hypotheses are formulated for convenience as follows.

H 1 *For any neighbourhood U of $\alpha\in\mathcal{M}_S^\nu$, and for any $\varepsilon>0$, there exists an ergodic $\alpha'\in U\cap\mathcal{M}_S^\nu$ such that $h_{\mathrm{Sh}}(\alpha')\geq h_{\mathrm{Sh}}(\alpha)-\varepsilon$.*

H 2 *There exist two functions $g_l\colon \mathcal{L}^\nu\to\mathbf{N}_0$ and $g_r\colon \mathcal{L}^\nu\to\mathbf{N}_0$, $\mathbf{N}_0 := \mathbf{N}\cup\{0\}$, and a non-decreasing function $g\colon \mathbf{N}_0\to\mathbf{N}_0$ so that*

$$g_l(w)\leq g(|w|)\ ,\ g_r(w)\leq g(|w|) \quad \text{and} \quad \lim_{n\to\infty}\frac{g(n)}{n}=0\,;$$

if $w^1\in\mathcal{L}^\nu$ and $w^2\in\mathcal{L}^\nu$, then there exists $v\in\mathcal{L}^\nu$, such that the string $w:=w^1vw^2\in\mathcal{L}^\nu$ and $|v|\leq\max\{g_r(w^1), g_l(w^2)\}$; furthermore, $g_l(w)\leq g_l(w^1)$ and $g_r(w)\leq g_r(w^2)$.

H 3 *The given reference probability measure ν is shift-invariant and non-atomic. There exists a continuous nonnegative function* e_ν *on* Σ^ν *so that*

$$\limsup_n \sup_{\omega\in\Sigma^\nu} \left|\frac{1}{n}\log\nu[\omega_1^n] + \langle\, e_\nu, T_n(\omega)\,\rangle\right| = 0 \tag{2.1}$$

and

$$\exists\, C_\nu > 0 \textit{ so that } \langle\, e_\nu, \rho\,\rangle \geq C_\nu \;\forall\, \rho\in\mathcal{M}_S^\nu. \tag{2.2}$$

Comments.

1) Hypothesis H1 is an important condition which has been used in various contexts including large deviations and dynamical systems. A general result which implies H1 is theorem B of [EKW], where the context is a $\mathbf{Z}^d$ action on a compact metric space which satisfies a specification property. This theorem implies that the shift space of an irreducible aperiodic subshift of finite type satisfies H1, but does not cover the periodic case. Proposition 5.1 shows that the shift space of an irreducible sofic shift satisfies H1. This includes all irreducible subshifts of finite type. Note that it does not suffice to approximate α by ergodic α'; in $\mathtt{A}^{\mathbf{N}}$, every neighbourhood of $\alpha\in\mathcal{M}^S$ contains ergodic measures with zero entropy concentrated on periodic points.
2) Hypothesis H2 is similar to the concept of specification, which has been extensively studied (see [KH]). Specification implies topological mixing, while the weaker H2 only implies topological transitivity. In section 5 we show that irreducible (possibly periodic) sofic shifts satisfy H2. Here is an example of a shift space Σ^* neither sofic nor periodic which satisfies H2. Let $\mathtt{A}=\{\mathbf{0},\mathbf{1}\}$. We represent $\omega\in\mathtt{A}^{\mathbf{N}}$ in terms of its runs of $\mathbf{0}$'s and $\mathbf{1}$'s, $(m_0;n_1,m_1;n_2,m_2;\ldots)$. The number of $\mathbf{0}$'s before the first $\mathbf{1}$ is m_0; if $\omega_1=\mathbf{1}$, then $m_0=0$, but all other defined m_i, n_i are strictly positive. The length the run of $\mathbf{1}$'s following the first m_0 $\mathbf{0}$'s is n_1; the length the run of $\mathbf{0}$'s following the first run of $\mathbf{1}$'s is m_1; then n_2 $\mathbf{1}$'s, m_2 $\mathbf{0}$'s and so on. Possibly $m_i=\infty$ or $n_i=\infty$, meaning ω_k takes a fixed value for large enough k; all further m_j, n_j are *undefined*. The point ω is in Σ^* if and only if whenever n_i and n_k are defined and $i<k$, there exists j, $i\leq j<k$ so that $m_j^2\geq\max\{n_i,n_k\}$. In particular if $\omega\in\Sigma^*$ has a defined $n_i=\infty$, then $i=1$. Hypothesis H2 holds for Σ^* with $g(n):=\lceil\sqrt{n}\,\rceil$; $g_l(w)=g(n_l)$, where $n_l\geq 0$ is the length of the leftmost run of $\mathbf{1}$'s in w and $g_r(w)=g(n_r)\geq 0$ is the length of the rightmost. Concatenate words using runs of $\mathbf{0}$'s.
3) In lemma 3.1 below we show that if ν is non-atomic and possesses a regular conditional probability kernel¶ $\nu(\omega_1|\omega_2,\omega_3,\ldots)$ that is continuous and strictly positive, then

   $$e_\nu(\omega) := -\log\nu(\omega_1|\omega_2,\omega_3,\ldots)$$

   satisfies (2.1). Non-atomic ergodic (possibly periodic) Markov chains satisfy H1, H2 and H3. g-measures on irreducible aperiodic subshifts of finite type also satisfy

¶ See [B] p.387 for a definition.

H1, H2 and H3. In section 5 we show that the maximal entropy measure of an irreducible sofic shift satisfies H3, even though $-\log\nu(\omega_1|\omega_2,\omega_3,\ldots)$ is not continuous. Therefore these measures also satisfy H1, H2 and H3.

4) It is not necessary that ν is a probability measure. Our method also applies if hypothesis H3 is stated as follows: for each ω of the shift space $\nu[\omega_1^n] > 0$, is decreasing in n and $\lim_{n\to\infty}\nu[\omega_1^n] = 0$. There exists a continuous nonnegative function e_ν on Σ^ν so that (2.1) and (2.2) hold.
   In [T1] and [T2] metrics d_ν of that type are considered for the special case $\mathrm{e}_\nu(\omega)\equiv\gamma$, $\gamma > 0$ a constant. In the rest of the paper we however always suppose that ν is a shift-invariant non-atomic probability measure.

Theorem 2.1 *Let $\alpha \in \mathcal{M}_S$. Then there exists a sequence $\{\Gamma_n \subset \mathtt{A}^n : n \in \mathbf{N}\}$ with the following two properties.*

$$\lim_n \frac{1}{n}\log|\Gamma_n| = h_{\mathrm{Sh}}(\alpha)\,;$$

for any open neighbourhood U of α, there exists $N(U)$, such that if $n \geq N(U)$, then

$$T_n(\omega) \in U \quad \forall\,\omega \in [b_1^n]\;,\;\forall\, b_1^n \in \Gamma_n\,. \tag{2.3}$$

Furthermore, if α is ergodic, then one can choose the sets Γ_n so that

$$\lim_n \sum_{b_1^n\in\Gamma_n} \alpha([b_1^n]) = 1\,.$$

Theorem 2.1 (see [LPRS] for proof) is formulated for the full shift space Ω. If $\alpha \in \mathcal{M}_S^\nu$ is ergodic, then the second part of theorem 2.1 implies that one can choose $\Gamma_n \subset \mathcal{L}_n^\nu$. In general one cannot have simultaneously property (2.3) and $\Gamma_n \subset \mathcal{L}_n^\nu$ if $\alpha \in \mathcal{M}_S^\nu$. However, this is true if hypothesis H1 is verified.

Corollary 2.1 *Let $\alpha \in \mathcal{M}_S^\nu$ and Σ^ν be a shift space verifying hypothesis H1. Then there exists a sequence $\{\Gamma_n \subset \mathcal{L}_n^\nu \colon n \in \mathbf{N}\}$ with the following properties. Given $\varepsilon > 0$ and a neighbourhood U of α, there exists $N(U,\varepsilon)$ such that for all $n \geq N(U,\varepsilon)$,*

$$\log|\Gamma_n| \geq n\,(h_{\mathrm{Sh}}(\alpha)-\varepsilon) \quad \text{and} \quad T_n(\omega) \in U\;,\;\forall\,\omega \in X_n^{-1}(\Gamma_n)\,.$$

Proof: Let U_k, $k \in \mathbf{N}$, be a decreasing sequence of neighbourhoods of α such that $\cap_k U_k = \{\alpha\}$. In each U_k choose an ergodic α_k such that $h_{\mathrm{Sh}}(\alpha_k) \geq h_{\mathrm{Sh}}(\alpha) - 1/k$. For each k there exist N_k and $\Gamma_{n,k} \subset \mathcal{L}_n^\nu$, such that for all $n \geq N_k$,

$$\frac{1}{n}\log|\Gamma_{n,k}| \geq h_{\mathrm{Sh}}(\alpha_k) - \frac{1}{k}\,,$$

and

$$T_n(\omega) \in U_k \quad \forall\,\omega \in [b_1^n]\;,\;\forall\, b_1^n \in \Gamma_{n,k}\,.$$

Define

$$\Gamma_n := \begin{cases} \mathcal{L}_n^\nu & \text{if } 1 \leq n < N_1; \\ \Gamma_{n,k} & \text{if } N_k \leq n < N_{k+1}. \end{cases}$$

□

Definition 2.1 *Let $F \subset \mathcal{M}_S^\nu$ be a closed connected subset. We define*

$$^F G := \{\omega \in \Sigma^\nu \colon \{T_n(\omega)\}_n \text{ has a limit point in } F\}\,, \tag{2.4}$$

and

$$G_F := \{\omega \in \Sigma^\nu \colon \text{the set of all limit points of } \{T_n(\omega)\}_n \text{ is } F\}\,. \tag{2.5}$$

The set $^F G \supset G_F$ is in general much larger then G_F. One can show that $\omega \in {}^F G$ if and only if there exists a closed connected set F' with $\omega \in G_{F'}$ and $F' \cap F \neq \emptyset$.

Proposition 2.1 *Let $\nu \in \mathcal{M}_S$ be a probability measure verifying hypothesis H3. Let $F \subset \mathcal{M}_S^\nu$ be a nonempty closed subset. Then*

$$\dim_\nu({}^F G) \leq s^* := \sup_{\alpha \in F} \frac{h_{\rm Sh}(\alpha)}{\langle\, e_\nu, \alpha \,\rangle}.$$

Proposition 2.2 *Let $\nu \in \mathcal{M}_S$ be a probability measure verifying hypotheses H1, H2 and H3. Let $F \subset \mathcal{M}_S^\nu$ be a nonempty closed connected subset. Then*

$$\dim_\nu(G_F) \geq s_* := \inf_{\alpha \in F} \frac{h_{\rm Sh}(\alpha)}{\langle\, e_\nu, \alpha \,\rangle}.$$

Remark: The restriction to $F \subset \mathcal{M}_S^\nu$ is not severe, because, if $\{T_n(\omega)\}$ has a limit-point which does not belong to $\mathcal{M}_S^\nu$, then $\omega \notin \Sigma^\nu$. Our bounds for ν-Billingsley dimension depend on the ratio $h_{\rm Sh}(\alpha)/\langle\, e_\nu, \alpha \,\rangle$; it must be computed using the same base for the logarithm in $h_{\rm Sh}$ and e_ν; the resulting ratio is independent of the choice of base. Note that 1 is an upper bound both for s_* and s^*. Results below show that our hypotheses imply that $G_{\{\nu\}}$ has ν-Billingsley dimension 1, because $\langle\, e_\nu, \nu \,\rangle = h_{\rm Sh}(\nu)$. If $\mathcal{M}_S^\nu \neq \{\nu\}$, all points of the set G_{F_a}, where $F_a := \{(1-t)\nu + t\alpha : 0 \leq t \leq a\}, 0 < a \leq 1$ and $\alpha \in \mathcal{M}_S^\nu \setminus \{\nu\}$, are nongeneric; as $a \downarrow 0$, the ν-Billingsley dimension of G_{F_a} goes to 1.

Propositions 2.1 and 2.2 imply our main result which is the following.

Theorem 2.2 *Let $\nu \in \mathcal{M}_S$ be a probability measure verifying hypotheses H1, H2 and H3. If $F \subset \mathcal{M}_S^\nu$ is a nonempty closed subset, then*

$$\dim_\nu({}^F G) = \sup_{\alpha \in F} \frac{h_{\rm Sh}(\alpha)}{\langle\, e_\nu, \alpha \,\rangle}\,.$$

If, in addition, F is connected, then

$$\dim_\nu(G_F) = \inf_{\alpha \in F} \frac{h_{\rm Sh}(\alpha)}{\langle\, e_\nu, \alpha \,\rangle}\,.$$

Proof: Let $\{\alpha\} \subset F$. Then

$$G_{\{\alpha\}} \subset {}^F G\,.$$

Hence

$$\sup_{\alpha \in F} \frac{h_{\rm Sh}(\alpha)}{\langle\, e_\nu, \alpha \,\rangle} \geq \dim_\nu({}^F G) \geq \frac{h_{\rm Sh}(\alpha)}{\langle\, e_\nu, \alpha \,\rangle} \quad \forall\, \alpha \in F\,.$$

Similarly

$$G_F \subset {}^{\{\alpha\}} G\,.$$

Hence

$$\inf_{\alpha \in F} \frac{h_{\mathrm{Sh}}(\alpha)}{\langle e_\nu, \alpha \rangle} \leq \dim_\nu(G_F) \leq \frac{h_{\mathrm{Sh}}(\alpha)}{\langle e_\nu, \alpha \rangle} \quad \forall\, \alpha \in F\,.$$

□

The saturated sets are unions of sets of the form G_F. Since Billingsley dimension is based on countably additive measures, it is easy to apply our results to countable unions. In the next theorem we handle uncountable unions.

Theorem 2.3 *Under hypotheses H1, H2 and H3, let* $\{F_t\colon t \in \mathcal{T}\}$ *be a collection of closed connected subsets of* $\mathcal{M}_S^\nu$ *such that* $\bigcup_{\mathcal{T}} F_t \neq \emptyset$. *Then*

$$\dim_\nu(\bigcup_{t \in \mathcal{T}} G_{F_t}) = \widehat{s} := \sup_{t \in \mathcal{T}} \inf_{\alpha \in F_t} \frac{h_{\mathrm{Sh}}(\alpha)}{\langle e_\nu, \alpha \rangle}\,.$$

Proof: That $\dim_\nu(\cup_{t \in \mathcal{T}} G_{F_t}) \geq \widehat{s}$ follows from the definition of Billingsley dimension and proposition 2.2. Each $\omega \in \bigcup_{\mathcal{T}} G_{F_t}$ has the property that $\{T_n(\omega)\}$ has a limit point α with $h_{\mathrm{Sh}}(\alpha)/\langle e_\nu, \alpha \rangle \leq \widehat{s}$. We could then consider instead of the given union, the set of ω corresponding to all such α, but this set of $\alpha \in \mathcal{M}^\nu$ need not be closed, so we cannot apply our upper bound of proposition 2.1. However, we prove in lemma 2.1 below, that for s, $0 < s < 1$, there exists a sequence of closed sets $\{W_n^s \subset \Sigma^\nu\}$ such that $\dim_\nu(W_n^s) \leq s$ and $\bigcup W_n^s$ contains all $\omega \in \Sigma^\nu$ such that $\{T_n(\omega)\}$ has a limit point α with $h_{\mathrm{Sh}}(\alpha)/\langle e_\nu, \alpha \rangle < s$. This shows that

$$\dim_\nu(\bigcup_{t \in \mathcal{T}} G_{F_t}) \leq s,\ \forall\, s > \widehat{s}\,.$$

□

Lemma 2.1 *Assume H1, H2 and H3. For* s, $0 < s < 1$, *there exists a sequence of sets* $\{W_n^s\colon n \in \mathbf{N}\}$ *in* Σ^ν *such that* $\dim_\nu W_n^s \leq s$*; if* $\{T_k(\omega)\}$, $\omega \in \Sigma^\nu$, *has limit point* α, *then* $\omega \in \bigcup W_n^s$, *provided that*

$$\frac{h_{\mathrm{Sh}}(\alpha)}{\langle e_\nu, \alpha \rangle} < s\,. \tag{2.6}$$

Proof: Define

$$D_n^s := \{\alpha \in \mathcal{M}^\nu\colon \frac{1}{n} H_n(\alpha) \leq s \langle e_\nu, \alpha \rangle\}\,,$$

and let $W_n^s \subset \Sigma_\nu$ be those points ω for which $\{T_n(\omega)\}$ has a limit point in D_n^s. The set D_n^s is closed because $H_n(\,\cdot\,)$ and $\langle e_\nu, \cdot \rangle$ are continuous. If $D_n^s = \emptyset$, then $\dim_\nu(W_n^s) = 0$; otherwise proposition 2.1 and the fact that

$$\frac{H_n(\alpha)}{n} \downarrow h_{\mathrm{Sh}}(\alpha)$$

imply that $\dim_\nu(W_n^s) \leq s$. Now let $\alpha \in \mathcal{M}^\nu$ satisfy (2.6). Then for some n $H_n(\alpha)/n < s \langle e_\nu, \alpha \rangle$, so $\alpha \in W_n^s$. □

3. Proof of proposition 2.1

3.1. Preliminary results

Lemma 3.1 *Let $\nu \in \mathcal{M}_S$ be a probability measure which possesses a continuous strictly positive regular conditional probability $\nu(\omega_1|\omega_2,\omega_3,\ldots)$. Then*

$$e_\nu(\omega) := -\log \nu(\omega_1|\omega_2,\omega_3,\ldots)$$

satisfies (2.1). If ν satisfies (2.1), then for each $\delta > 0$ there exist $m_\delta, N_\delta \in \mathbf{N}$ and $f_\delta \in \mathcal{F}_{m_\delta}$ so that for all $n \geq N_\delta$, for all $\omega \in \Sigma^\nu$, $|e_\nu(\omega) - f_\delta(\omega)| \leq \delta$ and

$$|\langle f_\delta, T_n(\omega) \rangle + \frac{1}{n}\log \nu([\omega_1^n])| < \delta\,. \tag{3.1}$$

Proof: When there exists e_ν satisfying (2.1), the bound (3.1) follows from the fact that local functions are dense in the set of continuous functions. For the conditional probability distribution case to simplify the notation we set

$$h(\omega) \;:=\; \begin{cases} e_\nu(\omega) = -\log \nu(\omega_1|\,\omega_2,\omega_3,\ldots) & \text{if } \omega \in \Sigma^\nu\,; \\ +\infty & \text{otherwise.} \end{cases}$$
$$h_n(\omega) := \; -\log \nu(\omega_1|\,\omega_2,\ldots,\omega_n)\,.$$

We have

$$h_n(\omega) = -\log \int \exp(-h(\omega))\, d\nu(\omega_{n+1},\omega_{n+2},\ldots\;|\,\omega_2,\ldots,\omega_n)\,.$$

It follows that

$$\|h_n\|_\nu := \sup_{\omega\in\Sigma^\nu} h_n(\omega) \leq \sup_{\omega\in\Sigma^\nu} h(\omega) =: \|h\|_\nu\,.$$

By hypothesis h is continuous on Σ^ν. Therefore, for each $\delta > 0$ there exist $m_\delta \in \mathbf{N}$ and a local function $\widehat{f}_\delta \in \mathcal{F}_{m_\delta}$, so that

$$|h(\omega) - \widehat{f}_\delta(\omega)| < \delta/2 \quad \forall \omega \in \Sigma^\nu\,.$$

Since $h \geq 0$, this inequality is maintained if any negative value of $\widehat{f}_\delta$ is replaced by 0. $\widehat{f}_\delta$ is explicitly defined on $X_{m_\delta}(\Sigma^\nu)$; we set

$$f_\delta(\omega) := \begin{cases} \widehat{f}_\delta(\omega) & \text{if } X_{m_\delta}(\omega) \in \mathcal{L}^\nu_{m_\delta} \text{ and } \widehat{f}_\delta(\omega) \geq 0\,; \\ 0 & \text{otherwise.} \end{cases}$$

We have, for $\omega \in \Sigma^\nu$,

$$\exp(-f_\delta(\omega) - \delta/2) < \nu(\omega_1|\,\omega_2,\ldots) < \exp(-f_\delta(\omega) + \delta/2)\,.$$

For $n \geq m_\delta$, using the measure $d\nu(\omega_{n+1},\ldots|\,\omega_2,\ldots,\omega_n)$ to integrate the above and noting that $f_\delta(\omega)$ depends only on $\omega_1,\ldots,\omega_{m_\delta}$, we have

$$\exp(-f_\delta(\omega) - \delta/2) < \nu(\omega_1|\,\omega_2,\ldots,\omega_n) < \exp(-f_\delta(\omega) + \delta/2)\,;$$

hence for $n \geq m_\delta$

$$|h_n(\omega) - f_\delta(\omega)| < \delta/2 \quad \forall \omega \in \Sigma^\nu\,.$$

Slightly abusing function notation, we note that

$$-\log \nu[\omega_1^k] = h_1(\omega_k) + h_2(\omega_{k-1}, \omega_k) + \cdots + h_k(\omega) = \sum_{j=1}^{k} S^{k-j} h_j(\omega)\,.$$

Therefore,

$$\Big| -\frac{1}{k} \log \nu[\omega_1^k] - \langle\, f_\delta, T_k(\omega)\,\rangle \Big| \leq \Big| \frac{1}{k} \sum_{j=1}^{k} S^{k-j}\,(h_j(\omega) - f_\delta(\omega))\,\Big|$$

$$\leq \frac{k+1-m_\delta}{2k}\delta + \frac{m_\delta - 1}{k}(\|h\|_\nu + \|f_\delta\|)\,.$$

□

Corollary 3.1 *Let $\nu \in \mathcal{M}_S$ be a probability measure verifying hypothesis H3. For $\alpha \in \mathcal{M}^\nu$ we have $|\langle\, e_\nu, \alpha\,\rangle - \langle\, f_\delta, \alpha\,\rangle| \leq \delta$ and*

$$\lim_{n\to\infty} -\frac{1}{n} \sum_{\omega_1^n \in \mathcal{L}_n^\nu} \alpha[\omega_1^n] \log \nu[\omega_1^n] = \langle\, e_\nu, \alpha\,\rangle\,.$$

3.2. Proof of proposition 2.1

Since ν verifies hypothesis H3, the conclusions of lemma 3.1 hold. Let $F \subset \mathcal{M}_S^\nu$ be a closed subset. With s^* as defined in the proposition, assume that $s^* < s < 1$. It is sufficient to show that there exists a collection of cylinder sets $\{B_n\colon n \in \mathbf{N}\}$ such that

$$\Omega^F \subset \bigcup_n B_n, \quad \sum_n \nu[B_n]^s < \varepsilon\,.$$

For $\delta > 0$ (to be specified more precisely below) and $a > 0$ define $F^{a,\delta}$ by

$$F^{a,\delta} := \{\alpha \in F\colon |\langle\, f_\delta, \alpha\,\rangle - a| \leq \delta\}\,,$$

with f_δ specified in lemma 3.1. Suppose that $F^{a,\delta} \neq \emptyset$; the closed neighbourhoods

$$U_{M,\varepsilon} := \{\alpha \in \mathcal{M}\colon \exists \alpha' \in F^{a,\delta}\,,\ \|\alpha - \alpha'\|_M \leq \varepsilon\}\,,$$

are such that

$$\bigcap_{M,\varepsilon} U_{M,\varepsilon} = F^{a,\delta}\,.$$

Since h_{Sh} is upper semi-continuous on $\mathcal{M}_S$, one can find a closed neighbourhood, say $\widehat{F}^{a,\delta} := U_{M,\varepsilon}$, so that $M \geq m_\delta$, $|\langle\, f_\delta, \alpha\,\rangle - a| \leq 2\delta$ for all $\alpha \in \widehat{F}^{a,\delta}$, and if $\alpha \in \widehat{F}^{a,\delta} \cap \mathcal{M}_S$, then $h_{\mathrm{Sh}}(\alpha) \leq \sup_{\alpha \in F^{a,\delta}} h_{\mathrm{Sh}}(\alpha) + \delta$. Define

$$\Gamma_n^{a,\delta} := X_{n+M}\left[\{\omega \in \Sigma^\nu\colon T_n(\omega) \in \widehat{F}^{a,\delta}\}\right]\,.$$

Since we cover ${}^F G$ by such cylinder sets, we need an upper bound for

$$\sum_{\omega_1^{n+M} \in \Gamma_n^{a,\delta}} \nu([\omega_1^{n+M}])^s\,.$$

For $\omega \in \Sigma^\nu$, such that $X_{n+M}(\omega) \in \Gamma_n^{a,\delta}$, we have for $n \geq N_\delta$ (lemma 3.1),

$$|\langle\, f_\delta, T_n(\omega)\,\rangle - a| \leq 2\delta \quad \text{and} \quad \Big| \frac{1}{n+M} \log \nu([X_{n+M}(\omega)]) + a \Big| < 3\delta\,.$$

Hence

$$\log \sum_{\omega_1^{n+M} \in \Gamma_n^{a,\delta}} \nu([\omega_1^{n+M}])^s \leq \log |\Gamma_n^{a,\delta}| + s\Big(\max_{\omega_1^{n+M} \in \Gamma_n^{a,\delta}} \log \nu([\omega_1^{n+M}])\Big)$$
$$\leq \log |\Gamma_n^{a,\delta}| + s(n+M)(-a+3\delta)\,.$$

To estimate $\log |\Gamma_n^{a,\delta}|$, we use a standard large deviations result for the empirical measure T_n on the probability space (Ω,β), where β is the uniform product probability: for any compact subset $K \subset \mathcal{M}$

$$\limsup_n \frac{1}{n} \log \beta(\{\omega \in \Omega \colon T_n(\omega) \in K\}) \leq \sup_{\alpha \in K \cap \mathcal{M}_S} (h_{\mathrm{Sh}}(\alpha) - \log r)\,.$$

(if $K \cap \mathcal{M}_S = \emptyset$, then the supremum is equal to $-\infty$). We apply this result with $K = \widehat{F}^{a,\delta}$; the set $\{\omega \colon T_n(\omega) \in \widehat{F}^{a,\delta}\}$ is $\mathcal{F}_{n+M}$-measurable. Therefore

$$\beta(\{\omega \in \Omega \colon T_n(\omega) \in \widehat{F}^{a,\delta}\}) = \frac{|X_{n+M}(\{\omega \in \Omega \colon T_n(\omega) \in \widehat{F}^{a,\delta}\})|}{r^{n+M}}\,.$$

Hence, for n large enough,

$$\log |\Gamma_n^{a,\delta}| = \log |X_{n+M}(\{\omega \in \Omega \colon T_n(\omega) \in \widehat{F}^{a,\delta}\})|$$
$$\leq n\Big(\sup_{\alpha \in \widehat{F}^{a,\delta} \cap \mathcal{M}_S} (h_{\mathrm{Sh}}(\alpha) + \delta)\Big)\,.$$

By the choice of $\widehat{F}^{a,\delta}$, definition of s^*, $F^{a,\delta} \subset F$ and corollary 3.1, we then have

$$\sup_{\alpha \in \widehat{F}^{a,\delta} \cap \mathcal{M}_S} h_{\mathrm{Sh}}(\alpha) \leq \sup_{\alpha \in F^{a,\delta}} h_{\mathrm{Sh}}(\alpha) + \delta \leq \sup_{\alpha \in F^{a,\delta}} s^*(\delta + \langle f_\delta, \alpha \rangle) + \delta \leq s^*\, a + 3\delta\,.$$

Therefore

$$\log \sum_{\omega_1^{n+M} \in \Gamma_n^{a,\delta}} \nu([\omega_1^{n+M}])^s \leq (n+M)[a(s^* - s) + 7\delta]\,.$$

The constant C_ν of hypothesis H3 is strictly positive. If $F^{a,\delta} \neq \emptyset$, then we have (corollary 3.1) $a \geq C_\nu - 2\delta$. We choose $\delta > 0$, so that $C_\nu - 3\delta > 0$ and for all $a \geq C_\nu - 3\delta$, $a(s^* - s) + 7\delta < 0$. Since F is compact, we can cover F by finitely many sets of the form $\widehat{F}^{a_j,\delta}$, $j = 1, \ldots, k$, with $a_j \geq C_\nu - 3\delta$. For such $\{a_j\}$ and δ, we can find M so that the sum

$$\sum_{n \geq m} \sum_{j=1}^{k} \sum_{\omega_1^{n+M} \in \Gamma_n^{a_j,\delta}} \nu([\omega_1^{n+M}])^s$$

can be made arbitrarily small by taking m sufficiently large. If $\omega \in \Sigma^\nu$ is such that $\{T_n(\omega)\}$ has a limit point in F, then ω is in

$$\bigcup_{n=m}^{\infty} \bigcup_{j=1}^{k} \bigcup_{\omega_1^{n+M} \in \Gamma_n^{a_j,\delta}} X_{n+M}^{-1}\{\omega_1^{n+M}\}$$

for arbitrary large m. Hence, for arbitrary large m,

$$\{[\omega_1^{n+M}] \colon \omega_1^{n+M} \in \Gamma_n^{a_j,\delta}\,,\, 1 \leq j \leq k\,,\, n \geq m\}$$

is a cover of ${}^F G$. □

4. Proof of proposition 2.2

Let $F \subset \mathcal{M}_S^\nu$ be a nonempty, closed and connected subset. We do not try to deal directly with G_F but instead generate a subset $B \subset \Sigma^\nu$ whose limit points are exactly the points of the closed connected set F, and so that we can compute its Billingsley dimension. The construction of B is the main part of the proof of proposition 2.2. It is done in subsection 4.1, and the main results are summarized in proposition 4.1

4.1. Construction of $B \subset G_F$

The idea of the construction of B is simple. We construct a dense sequence in F, $\{\alpha'_n\}$, an increasing sequence of integers $\{L_j\}$ and a sequence of subsets $\Gamma_{\alpha'_j} \subset \mathcal{L}^\nu_{L_j}$, in such a way that for ω, with $X_{L_j}(\omega) \in \Gamma_{\alpha'_j}$, we have $T_{L_j}(\omega) \approx \alpha'_j$ (i.e., in a neighbourhood of α') and $\log|\Gamma_{\alpha'_k}| \approx L_j h_{\rm Sh}(\alpha'_j)$. This is possible because of hypotheses H1 and H3. We then use hypothesis H2 in order to construct a set B, whose elements are obtained by concatenation of the words of the sets $\Gamma_{\alpha'_j}$. The sets $\Gamma_{\alpha'_j}$ are chosen to make F the set of limit points of $\{T_n(\omega)\}$ for each $\omega \in B$. We first describe the construction of the set B from the sets $\Gamma_{\alpha'_k}$, and then the construction of the sequence $\{\Gamma_{\alpha'_k}\}$.

We define iteratively a sequence of subsets of $\mathcal{L}^\nu$, $B_1, B_2, \ldots$, so that each word of B_k is a prefix of a word of B_{k+1}, so that we can define B as the limit of these sets,

$$B := \bigcap_k \bigcup_{w \in B_k} [w]\,.$$

Let $B_1 := \Gamma_{\alpha'_1} \subset \mathcal{L}^\nu_{L_1}$. On B_1 we define a function

$$b_1(w) := |w|\,.$$

For each $w \in B_1$ we have $b_1(w) = L_1$ and $g_r(w) \leq g(L_1)$, and for each $w \in \Gamma_{\alpha'_2}$ we have $g_l(w) \leq g(L_2)$. According to H2, for each $w^1 \in B_1$ and each $w^2 \in \Gamma_{\alpha'_2}$, there exists at least one word v^{12} of length $|v^{12}| \leq g(L_2)$, so that $w^1 v^{12} w^2 \in \mathcal{L}^\nu$. In this way we obtain for each $w^1 \in B_1$ a set $E'(w^1)$ of $|\Gamma_{\alpha'_2}|$ new words, which have the same prefix w^1, but not necessarily the same length. For lemma 4.1 it is important that all words of the form $w^1 v^{12} w^2$, with the same prefix w^1, have the same length. In order to achieve this we select a subset $E(w^1) \subset E'(w^1)$. We partition $E'(w^1)$ into at most $g(L_2)+1$ subsets, so that each word in a given element of the partition has the same length. We choose for $E(w^1)$ an element of the partition of maximal cardinality. By definition

$$B_2 := \bigcup_{w^1 \in B_1} E(w^1)\,.$$

On B_2 we define a function b_2,

$$b_2(w) := |w|\,.$$

In general b_2 is not constant. Notice that b_1 can be defined also on B_2. We have $b_1(w) + L_2 \leq b_2(w) \leq b_1(w) + L_2 + g(L_2)$ for all $w \in B_2$, and

$$|B_2| \geq \frac{|B_1| \cdot |\Gamma_{\alpha'_2}|}{g(L_2)+1}\,. \tag{4.1}$$

We say that B_2 is a concat-product of B_1 and $\Gamma_{\alpha'_2}$. Of course, a concat-product is in general not unique, since it depends on the specific choices of the words v^{12} in $w^1v^{12}w^2$ and of $E(w^1)$, but it is well-defined. By hypothesis H2, for all $w \in B_2$, $g_r(w) \leq g(L_2)$. We can iterate this procedure, and define B_k as a concat-product of B_{k-1} and $\Gamma_{\alpha'_k}$, and on B_k the function $b_k(w) := |w|$, as well as the extensions of the functions b_j, $j+1,\ldots,k-1$, on B_k are well-defined. Because of the prefix property the set B is well-defined. It is convenient to define b_0 on B by $b_0(\omega) := 0$. The next lemma shows that the functions b_k, for all $k \in \mathbf{N}$, are well-defined on B.

Lemma 4.1 *Each $\omega \in B$ has a unique decomposition into*

$$\omega = w^1v^{12}w^2v^{23}w^3\cdots \quad \text{with} \quad w^j \in \Gamma_{\alpha'_j}\,,\ j \in \mathbf{N}\,.$$

Let ω, $\widehat{\omega} \in B$. Then $\omega = \widehat{\omega}$ if and only if

$$b_k(\omega) = b_k(\widehat{\omega}) \quad \forall\, k \in \mathbf{N} \quad \text{and} \quad \omega_{b_j-L_j+1}^{b_j} = \widehat{\omega}_{b_j-L_j+1}^{b_j} \quad \forall\, j \geq 1\,.$$

Proof: Let $\omega \in B$. Suppose that

$$\omega = w^1v^{12}w^2v^{23}w^3\cdots = \widehat{w}^1\widehat{v}^{12}\widehat{w}^2\widehat{v}^{23}\widehat{w}^3\cdots.$$

Since $b_1(\omega) = L_1$ for all ω, we must have $w^1 = \widehat{w}^1$. This implies that $w^1v^{12}w^2 \in E(w^1)$ and $w^1\widehat{v}^{12}\widehat{w}^2 \in E(w^1)$, and therefore $|v^{12}w^2| = |\widehat{v}^{12}\widehat{w}^2|$. But $|w^2| = |\widehat{w}^2|$, and therefore $w^2 = \widehat{w}^2$ and $v^{12} = \widehat{v}^{12}$. Hence $b_2(\omega)$ is well-defined. By induction this proves the first statement. The second statement follows from the fact that at each step of the construction of B, the choice of $v^{n\,n+1}$ depends only on $w^1v^{12}w^2v^{23}w^3\cdots w^n$ and w^{n+1}. □

We describe the construction of the sequence $\{\Gamma_{\alpha'_k}\}$ in detail. For each $\varepsilon > 0$ there exists a finite sequence of points $\alpha_1,\ldots,\alpha_n$ in F such that each point of F is within ε of some α_j. Because F is connected, possibly repeating some α_j, we can choose this sequence so that $\alpha_1,\ldots,\alpha_{n'}$ is not more than ε away from any point of F and $\|\alpha_j - \alpha_{j+1}\| < 2\varepsilon$ for each j. Extending this argument we deduce the following.

Lemma 4.2 *Let F be a nonempty closed connected set of the compact metric space $(\mathcal{M}_S^\nu, \|\cdot\|)$. Then there exists a sequence $\alpha_1, \alpha_2, \ldots$ in F so that the closure of $\{\alpha_j\colon j \in \mathbf{N},\ j > n\}$ for each $n \in \mathbf{N}$ equals F and*

$$\lim_{j\to\infty} \|\alpha_j - \alpha_{j+1}\| = 0.$$

Let $\varepsilon < 1$ and $\alpha \in \mathcal{M}_S^\nu$. By lemma 3.1 there exists N_ε, m_ε and $f_\varepsilon \in \mathcal{F}_{m_\varepsilon}$, so that for all $n \geq N_\varepsilon$,

$$\Big|\,\langle\, f_\varepsilon, T_n(\omega)\,\rangle + \frac{1}{n}\log\nu([\omega_1^n])\,\Big| < \frac{\varepsilon}{3} \text{ and } |\, f_\varepsilon(\omega) - e_\nu(\omega)\,| < \frac{\varepsilon}{3}\ \forall\, \omega \in \Sigma^\nu\,.$$

Define $M(\varepsilon) \in \mathbf{N}$ by

$$M(\varepsilon) := \max\Big\{m_\varepsilon, \min\{m \in \mathbf{N}\colon 2^{1-m} \leq \varepsilon\}\Big\}\,.$$

For $\rho, \rho' \in \mathcal{M}$, since $\|\rho - \rho'\|_{TV_n} \leq 2$, we have

$$\|\rho - \rho'\|_{M(\varepsilon)} < \varepsilon \Longrightarrow \|\rho - \rho'\| < 2\varepsilon\,.$$

For fixed α and $\varepsilon > 0$, by corollary 2.1, there exist sets $\Gamma(\alpha, n) \subset \mathcal{L}_n^\nu$ and $N(\alpha,\varepsilon) \geq N_\varepsilon$, so that for all $n \geq N(\alpha,\varepsilon)$,

$$\left.\begin{aligned}
&\|T_n(\omega) - \alpha\|_{M(\varepsilon)} && < \frac{\varepsilon}{3} \quad \forall\, \omega \in \Sigma^\nu \;,\; X_n(\omega) \in \Gamma(\alpha, n)\,;\\
&\frac{g(n)}{n} && < \varepsilon\,;\\
&\log|\Gamma(\alpha,n)| && > n\left(h_{\mathrm{Sh}}(\alpha) + \frac{\log(g(n)+1)}{n} - \varepsilon\right);\\
&|\,\langle\, e_\nu, \alpha\,\rangle + \frac{1}{n}\log\nu([\omega_1^n])\,| < \varepsilon \quad \forall\, \omega_1^n \in \Gamma(\alpha,n)\,.
\end{aligned}\right\} \tag{4.2}$$

We choose a dense sequence $\{\alpha_n\}$, as in lemma 4.2, and a sequence $\{\varepsilon_n \downarrow 0\}$; we assume that $\varepsilon_1 < \varepsilon$, $0 < \varepsilon < 1$ arbitrary but fixed, and set $N_k := N(\alpha_k, \varepsilon_k)$, so that conditions (4.2) hold for all $n \geq N_k$, with $\alpha = \alpha_k$ and $\varepsilon = \varepsilon_k$. From these inputs we construct a new sequence $\{\alpha'_k\}$, a sequence of integers $\{L_k\}$ and sets $\Gamma_{\alpha'_k}$. The construction of $\{\alpha'_k\}$ is done by stretching the sequence $\{\alpha_k\}$. This amounts to define an increasing diverging sequence of integers $\{K_m\}$ and to define the stretched sequence $\{\alpha'_k\}$ by

$$\alpha'_k := \alpha_m \quad \text{if} \quad K_m \leq k < K_{m+1}\,.$$

We write $\alpha'_k \equiv \alpha_{j_k}$. We also stretch the sequence $\{\varepsilon_k\}$ in the same way. α'_k and L_k are defined iteratively.

To start we set $j_1 := 1$, $\alpha'_1 := \alpha_{j_1}$ and $L_1 := N_{j_1}$, so that for all $n \geq L_1$ conditions (4.2) obtain; set $\Gamma_{\alpha'_1} := \Gamma(\alpha_{j_1}, L_1)$ and $B_1 := \Gamma_{\alpha'_1}$. Assume that $\alpha'_m \equiv \alpha_{j_m}$ and L_m have been chosen for $1 \leq m \leq k$.

A1 Check whether

$$\|T_{b_k(\omega)}(\omega) - \alpha_{j_k}\| < 5\varepsilon_{j_k}\,, \tag{4.3}$$

fails to hold for any $\omega \in B_k$. If this is the case, then go to **A2**. If this is not the case, i.e. (4.3) holds for all $\omega \in B_k$, go to **A3**.

A2 Set

$$j_{k+1} := j_k \quad , \quad L_{k+1} := L_k + 1 \quad , \quad \widehat{L}_{k+1} := L_{k+1} + g(L_{k+1})\,.$$

Define $\Gamma_{\alpha'_{k+1}} := \Gamma(\alpha_{j_{k+1}}, L_{k+1})$ (see (4.2)). Construct B_{k+1}, a concat-product of B_k and $\Gamma_{\alpha'_{k+1}}$. Then start again at **A1**.

The new strings of B_{k+1} have length $b_{k+1}(\omega)$,

$$b_k(\omega) + L_{k+1} \leq b_{k+1}(\omega) \leq b_k(\omega) + \widehat{L}_{k+1}\,.$$

We continue with the same α_{j_k} and tolerance ε_{j_k} until $T_{b_k(\omega)}(\omega)$ is within $5\varepsilon_{j_k}$ of α_{j_k} for each $\omega \in B_k$. Estimate (4.5) below shows this eventually occurs. This assures that for $\omega \in B$, the sequence $\{T_n(\omega)\}$ gets close to each α_j, the criterion of closeness approaching 0 as $j \to \infty$. Note that we increment L_k by one at each step. This is convenient, but not essential.

A3 Check whether $N_{j_k+1} > L_k + 1$. If this is the case, then go to **A2**. If this is not the case, i.e. $N_{j_k+1} \leq L_k + 1$, then go to **A4**.

A4 Set

$$j_{k+1} := j_k + 1 \quad , \quad L_{j+1} := L_j + 1 \quad , \quad \widehat{L}_{k+1} := L_{k+1} + g(L_{k+1})\,.$$

Define $\Gamma_{\alpha'_{k+1}} := \Gamma(\alpha_{j_{k+1}}, L_{k+1})$. Construct B_{k+1}, a concat-product of B_k and $\Gamma_{\alpha'_{k+1}}$. Then start again at **A1**.

Using this algorithm we construct the sets $\Gamma_{\alpha'_k}$. Set $\varepsilon'_k := \varepsilon_{j_k}$. Because of the rules, we have by definition

$$L_n = L_1 + (n-1)$$

and

$$\sum_{j=1}^n L_j = nL_1 + \sum_{j=1}^n (k-1) = n\left(L_1 + \frac{(n-1)}{2}\right)\,.$$

Since $\varepsilon_j < 1$ and (4.2) holds,

$$\sum_1^n L_j \leq b_n(\omega) \leq \sum_1^n \widehat{L}_j < 2\sum_1^n L_j\,.$$

The next estimates are important,

$$\lim_{n\to\infty} \frac{\widehat{L}_n}{b_{n-1}} = 0 \quad \text{and} \quad \lim_{n\to\infty} \frac{n}{b_{n-1}} = 0\,.$$

To simplify the notations we drop from now on the prime index, and write b_j for $b_j(\omega)$. We prove estimate (4.5) and show that $B \subset G_F$. Let $\omega \in B$, $n = b_m + k$ with $0 \leq k < b_{m+1} - b_m$, and $0 \leq q < m$. We have

$$T_n(\omega) = \frac{b_q}{b_m+k} T_{b_q}(\omega) + \sum_{j=q}^{m-1} \frac{b_{j+1}-b_j}{b_m+k} T_{b_{j+1}-b_j}(S^{b_j}\omega) + \frac{k}{b_m+k} T_k(S^{b_m}\omega)\,. \tag{4.4}$$

Using (4.2) and

$$\|T_n(\omega) - T_m(\omega)\| \leq \frac{2(n-m)}{n} \quad \text{if } n > m\,,$$

we get for $\omega \in B$, with $q = 0$,

$$\Big\| T_{b_m+k}(\omega) - \sum_{j=0}^{m-1} \frac{b_{j+1}-b_j}{b_m+k} T_{L_{j+1}}(S^{b_j}\omega) \Big\| \leq \frac{2k}{b_m} + 2\sum_{j=0}^{m-1} \frac{b_{j+1}-b_j}{b_m} \varepsilon_{j+1}\,,$$

and

$$\Big\| T_{b_m+k}(\omega) - \sum_{j=0}^{m-1} \frac{b_{j+1}-b_j}{b_m} \alpha_{j+1} \Big\| \leq \frac{3k}{b_m} + 3\sum_{j=0}^{m-1} \frac{b_{j+1}-b_j}{b_m} \varepsilon_{j+1}\,. \tag{4.5}$$

Suppose that $\lim_p T_{n_p}(\omega) = \alpha$. We write $n_p = b_{m_p} + k_p$ with $0 \leq k_p < b_{m_p+1}$. According to the construction of B, if p is large enough, there exists $\alpha_j \in F$, and a decomposition of m_p into $m_p = q_p + t_p$, with

$$\|T_{b_{q_p}}(\omega) - \alpha_j\| \leq 5\varepsilon_{q_p} \quad \text{and} \quad \|T_{L_{q_p+i}}(S^{b_{q_p+i-1}}\omega) - \alpha_{j+1}\| \leq \frac{\varepsilon_{j+1}}{3}$$

for all $i = 1, \ldots, t_p$. Using (4.4) we get that

$$\|T_{b_{m_p}}(\omega) - \alpha_j\| \leq 5\varepsilon_j + \frac{\varepsilon_{j+1}}{3} + \|\alpha_j - \alpha_{j+1}\| + \frac{2k_p}{b_{m_p}}.$$

This shows that $\alpha \in F$. Similarly, one shows that each $\alpha \in F$ is a limit-point of F. Finally, by (4.1) and (4.2)

$$\log |B_N| \geq \sum_{j=1}^{N} L_j(h_{\mathrm{Sh}}(\alpha_j) - \varepsilon_j).$$

We summarize the results of subsection 4.1 as follows.

Proposition 4.1 *Let ε, $0 < \varepsilon < 1$. Let $F \subset \mathcal{M}_S$ be a nonempty closed and connected set. Then there exist a sequence of subsets $B_n \subset \mathcal{L}^\nu$, $n \geq 1$, an increasing diverging sequence of integers L_n, $n \geq 1$, a decreasing sequence ε_n, $n \geq 1$, such that $\varepsilon_n \leq \varepsilon$ and $\lim_n \varepsilon_n = 0$, and a sequence $\alpha_n \in F$, $n \geq 1$ with the following properties.*
1. The closure of $\{\alpha_j : j \geq n\}$ is F, for all $n \geq 1$.
2. $\log |B_n| \geq \sum_{j=1}^{n} L_j(h_{\mathrm{Sh}}(\alpha_j) - \varepsilon_j)$.
3. Each word of B_n is a prefix of a word of B_{n+1}, and

$$B := \bigcap_k \bigcup_{w:\, w \in B_k} [w] \subset G_F.$$

4.2. Proof of proposition 2.2

Proposition 2.2 is trivial if $s_* = 0$. We therefore assume that $0 < s < s_*$. We construct B as in proposition 4.1, and show that $\mathcal{C}^s(B) = \infty$. To compute $\mathcal{C}^s_\delta(B)$ we can consider finite covers since B is compact. Let $\mathcal{D}$ be a finite cover of B. Each cylinder of $\mathcal{D}$ is labeled by a word in $\mathcal{L}^\nu$, and we also denote the set of these words by $\mathcal{D}$. Two cylinders labeled by two different words, say w^1 and w^2, are either disjoint or one is a subset of the other. The latter case occurs if and only if one of the word is the prefix of the other. For each $n \in \mathbf{N}$ we set

$$y_n := \min\{\nu([w]) \colon w \in B_n\}.$$

The main part of the proof of $\mathcal{C}^s(B) = \infty$ is lemma 4.3.

Lemma 4.3 *Let $0 < s \leq 1$. Let $\mathcal{D}$ be a cover of B, so that $\nu([w]) < y_N$, for all $w \in \mathcal{D}$. Then there exists $n \geq N$ so that*

$$\sum_{w \in \mathcal{D}} \nu([w])^s \geq \exp\Big(\sum_{j=1}^{n} L_j(h_{\mathrm{Sh}}(\alpha_j) - \varepsilon_j)\Big)\, y_{n+1}^s.$$

Proof: Since $\nu([w]) < y_N$, each $w \in \mathcal{D}$ is of the form $w = vu$, with $v \in B_N$. Moreover, any $v \in B_N$ appears as prefix, since $\mathcal{D}$ is a cover of B. We write

$$\sum_{w \in \mathcal{D}} \nu([w])^s = \sum_{v \in B_N} \sum_{u:\, vu \in \mathcal{D}} \nu([vu])^s.$$

Let $v_* \in B_N$ so that for all $v \in B_N$,

$$\sum_{u:\, vu \in \mathcal{D}} \nu([vu])^s \geq \sum_{u:\, v_*u \in \mathcal{D}} \nu([v_*u])^s.$$

Then, by proposition 4.1,

$$\sum_{w\in\mathcal{D}} \nu([w])^s \geq |B_N| \sum_{u\colon v_* u\in\mathcal{D}} \nu([v_* u])^s \geq \exp\big(\sum_{j=1}^{N} L_j(h_{\mathrm{Sh}}(\alpha_j)-\varepsilon_j)\big) \sum_{u\colon v_* u\in\mathcal{D}} \nu([v_* u])^s\,.$$

Either

$$\sum_{w\in\mathcal{D}} \nu([w])^s \geq \exp\big(\sum_{j=1}^{N} L_j(h_{\mathrm{Sh}}(\alpha_j)-\varepsilon_j)\big)\, y_{N+1}^s\,,$$

or

$$\sum_{w\in\mathcal{D}} \nu([w])^s < \exp\big(\sum_{j=1}^{N} L_j(h_{\mathrm{Sh}}(\alpha_j)-\varepsilon_j)\big)\, y_{N+1}^s\,.$$

In the latter case, we have

$$\sum_{u\colon v_* u\in\mathcal{D}} \nu([v_* u])^s < y_{N+1}^s\,,$$

so that $\nu([v_* u]) < y_{N+1}$ for all u. Let $P_{N+1} := \{v \in B_{N+1}\colon v_* \text{ is a prefix of } v\}$. For each u such that $v_* u \in \mathcal{D}$, we can write $v_* u = v'u'$, with $v' \in P_{N+1}$. Since $\mathcal{D}$ is a cover of B, all prefixes $v' \in P_{N+1}$ occur in the decompositions of $v_* u = v'u'$. By the construction of B_{N+1} and (4.2), we have

$$|P_{N+1}| \geq \exp\big(L_{N+1}(h_{\mathrm{Sh}}(\alpha_{N+1})-\varepsilon_{N+1})\big)\,.$$

Choose $v'_* \in P_{N+1}$ so that for all $v' \in P_{N+1}$,

$$\sum_{u'\colon v'u'\in\mathcal{D}} \nu([v'u'])^s \geq \sum_{u'\colon v'_* u'\in\mathcal{D}} \nu([v'_* u'])^s\,.$$

Therefore we get

$$\begin{aligned}\sum_{w\in\mathcal{D}} \nu([w])^s &\geq \exp\big(\sum_{j=1}^{N} L_j(h_{\mathrm{Sh}}(\alpha_j)-\varepsilon_j)\big) \sum_{u\colon v_* u\in\mathcal{D}} \nu([v_* u])^s \\ &= \exp\big(\sum_{j=1}^{N} L_j(h_{\mathrm{Sh}}(\alpha_j)-\varepsilon_j)\big) \sum_{v'\in P_{N+1}} \sum_{u'\colon v'u'\in\mathcal{D}} \nu([v'u'])^s \\ &\geq \exp\big(\sum_{j=1}^{N+1} L_j(h_{\mathrm{Sh}}(\alpha_j)-\varepsilon_j)\big) \sum_{u'\colon v'_* u'\in\mathcal{D}} \nu([v'_* u'])^s\,.\end{aligned}$$

We can repeat the above argument; since the cover $\mathcal{D}$ is finite, there exits $n \geq N$ such that

$$\sum_{w\in\mathcal{D}} \nu([w])^s \geq \exp\big(\sum_{j=1}^{n} L_j(h_{\mathrm{Sh}}(\alpha_j)-\varepsilon_j)\big)\, y_{n+1}^s\,.$$

□

We complete the proof of proposition 2.2. By lemma 3.1 and (4.5) we have

$$\lim_n \sup_{w\in B_{n+1}} \left| \frac{1}{b_{n+1}} \log\nu([w]) + \sum_{j=0}^{n} \frac{b_{j+1}-b_j}{b_{n+1}} \langle\, e_\nu, \alpha_{j+1}\,\rangle \right| = 0\,. \tag{4.6}$$

For $\delta > 0$ there exists N_δ so that for $n \geq N_\delta$ we deduce from (4.6), (4.1), (4.2) and the inequality $h_{\rm Sh}(\alpha_j)/s_* \geq \langle e_\nu, \alpha_j \rangle \geq C_\nu$ that

$$\sum_{j=1}^n L_j(h_{\rm Sh}(\alpha_j) - \varepsilon_j) + s \log y_{n+1} \geq$$

$$\sum_{j=1}^n h_{\rm Sh}(\alpha_j)(L_j - \frac{s}{s_*}(b_j - b_{j-1})) - s(b_{n+1} - b_n)\langle e_\nu, \alpha_{n+1} \rangle - \sum_{j=1}^n L_j\varepsilon_j - \delta s b_{n+1}.$$

Taking $\delta > 0$ so that $s\,\delta < (s_* - s)\,C_\nu$ and noting that $\lim_n \sum_1^n L_j/b_{n+1} = 1$, we deduce

$$\liminf_n \exp\,(\sum_{j=1}^n L_j(h_{\rm Sh}(\alpha_j) - \varepsilon_j))\, y_{n+1}^s = \infty\,,$$

and therefore $\mathcal{C}^s(B) = \infty$ by lemma 4.3. □

5. Application to irreducible sofic shifts

It is known that irreducible, aperiodic sofic shifts satisfy a specification condition, so theorem B of [EKW] implies that H1 obtains. We give another proof here which does not require aperiodicity, and then remark that H2 obtains. We also show that the measure of maximal entropy on an irreducible sofic shift verifies H3, so that our results apply for such a measure. Let $\nu \in \mathcal{M}_S$ be a non-atomic probability measure, Σ^ν the corresponding shift space. We follow [LM] for the treatment of sofic shifts+.

Proposition 5.1 *Suppose that the shift space Σ^ν is an irreducible sofic shift, and let $\alpha \in \mathcal{M}_S^\nu$. Then for any neighbourhood U of α, and for any $\varepsilon > 0$, there exists an ergodic $\alpha' \in U \cap \mathcal{M}_S^\nu$ such that $h_{\rm Sh}(\alpha') \geq h_{\rm Sh}(\alpha) - \varepsilon$.*

We first prove the following

Lemma 5.1 *Let $\mathcal{G} = (G, \mathcal{L})$ be an irreducible, fixed labeled graph, which is a minimal right-resolving presentation of Σ^ν. Let $G = (\mathcal{V}, \mathcal{E})$ be the underlying graph, with set of vertices $\mathcal{V} := \{A_1, \ldots, A_p\}$ and set of edges $\mathcal{E}$. Assume that there exists A_j with at least two distinct outward edges. Then there exists a prefix-code with p^2 synchronizing words $\mathcal{T} := \{w^{j,k}\colon j,k = 1, \ldots, p\}$ such that $w^{j,k}$ has a presentation with initial vertex A_j and terminal vertex A_k.*

Proof: We suppose that A_1 is the vertex with two outward edges. For every A_j there exists a synchronizing word, which has a presentation by a path from A_j to A_1 (see proposition 3.3.16 in [LM]). For each A_j we choose such a word of minimal length and denote it u^j. Note that possibly $u^i = u^j$ for $i \neq j$. Because of the minimality property the set of these words is a prefix code. Next choose two nonempty distinct words w^0 and w^1 so that they have presentations with initial and terminal vertices A_1 and in these

+ Chapter 3 of [LM] is written for two-sided sofic shifts. Our case corresponds to the one-sided shifts as defined in [LM] p. 461.

presentations A_1 does not otherwise appear. We construct a prefix-code with exactly p^2 words by setting

$$v^{j,k} := \underbrace{w^0 \cdots w^0}_{j \text{ times}} \underbrace{w^1 \cdots w^1}_{k \text{ times}} w^0 \quad j,k = 1,\ldots,p\,.$$

These words are used to encode the end-points of the paths presenting the words of $\mathcal{T}$. Finally let z^k, $k = 1,\ldots,p$ be p words, such that z^k has a presentation by a path from A_1 to A_k. We set

$$w^{j,k} := u^j v^{j,k} z^k \quad j,k = 1,\ldots,p \quad \text{and} \quad \mathcal{T} := \{w^{j,k}\colon j,k = 1,\ldots,p\}\,.$$

□

Proof of proposition 5.1: If the minimal right-resolving presentation of Σ^ν does not have a vertex with two distinct outward edges, then Σ^ν is finite and α is ergodic, so there is nothing to prove. It remains to consider the case when the assumptions of lemma 5.1 are satisfied. Let $\alpha \in \mathcal{M}_S^\nu$. For each n, $\mathcal{L}_n^\alpha \subset \mathcal{L}_n^\nu$. We partition $\mathcal{L}_n^\alpha$ into p subsets $\mathcal{L}_n^\alpha(i)$, $i = 1,\ldots,p$, so that each $y \in \mathcal{L}_n^\alpha(i)$ has a presentation by a path starting at A_i. The partition is not in general unique; we make a specific choice. For $1 \leq j \leq p$, we construct a set $\overline{\mathcal{W}}_n(j)$. Let r be the maximal length of the words in $\mathcal{T}$. Let $y \in \mathcal{L}_n^\alpha(i)$. We concatenate $w^{j,i}$ and y, and extend the word $w^{j,i}y$ to a word in $\mathcal{L}_{n+r}^\nu$ by a specified choice if necessary. We do this construction for all $y \in \mathcal{L}_n^\alpha$; this defines $\overline{\mathcal{W}}_n(j)$ and a bijection between $\mathcal{L}_n^\alpha$ and $\overline{\mathcal{W}}_n(j)$, since we can recover $y := \psi(w)$ from w by first decoding the word of $\mathcal{T}$, which is the prefix of y in w, and then reading the next n letters. Notice that ψ is well-defined on $\overline{\mathcal{W}}_n := \bigcup \overline{\mathcal{W}}_n(j)$. Each $w \in \overline{\mathcal{W}}_n$ is synchronizing; hence the terminal vertex, denoted by $t(w)$, of any presentation of w is unique. For $w \in \overline{\mathcal{W}}_n$ we use $s(w)$ to denote the initial vertex j of the unique $w^{j,k} \in \mathcal{T}$ which appears at the start of w. Next we introduce the $p \times p$ stochastic matrix

$$M^n = (M_{i,j}^n)\,,\; M_{i,j}^n := \sum_{w \in \overline{\mathcal{W}}_n(i),\, t(w) = A_j} \alpha(\psi(w))\,.$$

Let Y_n be the product space $\mathcal{L}_n^{\alpha\mathbf{N}}$, whose elements are written $y = (y^1, y^2, y^3 \cdots)$, where $y^j \in \mathcal{L}_n^\alpha$. Given a vertex $A_k \in \mathcal{V}$, we construct an injective map $\Phi_n^k\colon Y \to \Sigma^\nu$. $\Phi_n^k(y) := w^1 w^2 w^3 \cdots$ is the concatenation of w^1, w^2, w^3, where $w^1 \in \overline{\mathcal{W}}_n(k)$ and $\psi(w^1) = y^1$, $w^2 \in \overline{\mathcal{W}}_n(t(w^1))$ and $\psi(w^2) = y^2$, and so on.

For each $n \in \mathbf{N}$ we construct a probability measure $\alpha^n \in \mathcal{M}^\nu$. Let k^* be a recurrent state of the Markov chain defined by M^n on the state space $\{1,\ldots,p\}$. Let $K^* \subset \{1,\ldots,p\}$ be the communicating class of k^*. Let $M_{i,j}^*$ denote the restriction of $M_{i,j}^n$ to K^*. K^* and $M_{i,j}^*$ may depend on n, but for simplicity n is not included in the notation. $(M_{i,j}^*)$ is the stochastic matrix of a stationary ergodic Markov chain on K^*. Let $\{m(j)\colon j \in K^*\}$ be the stationary distribution:

$$m(j) > 0, \quad \sum_{j \in K^*} m(j) = 1, \quad \sum_{i \in K^*} M_{i,j}^* m(i) = m(j).$$

We define α^n on the algebra $\mathcal{F}_{k(n+r)}$. Each atom of $\mathcal{F}_{k(n+r)}$ is labeled by a word of $\mathcal{L}_{k(n+r)}^\nu$, which is uniquely decomposed into k blocks of length $n+r$, $\omega_1^{k(n+r)} \equiv w^1 w^2 \cdots w^k$. For

$i \in K^*$, $w_1 \in \overline{\mathcal{W}}_n(i)$ and $w^2, \ldots, w^k$ satisfying $w_j \in \overline{\mathcal{W}}_n(t(w_{j-1}))$, $j = 2, \ldots, k$,

$$\alpha_i^n([w^1 \cdots w^k]) := \prod_{j=1}^{k} \alpha([\psi(w^j)]),$$

$$\alpha^n([w^1 \cdots w^k]) := \sum_{i \in K^*} m(i) \alpha_i^n([w^1 \cdots w^k]),$$

with $\alpha^n([w^1 \cdots w^k]) = 0$ otherwise. α^n is defined consistently on the increasing family of algebras $\mathcal{F}_{k(n+r)}$, $k \in \mathbf{N}$. Kolmogorov's theorem implies the existence of α^n. Let

$$\overline{\mathcal{W}}_n^* := \Big\{ w \in \overline{\mathcal{W}}_n\colon\ s(w) \in K^* \Big\} .$$

It is not difficult to verify that α^n is a stationary Markov chain with state space $\overline{\mathcal{W}}_n^*$. The stochastic matrix entry for (w, w') equals zero unless $s(w') = t(w)$, in which case it equals $\alpha([\psi(w')]) > 0$. For $i, j \in K^*$, it is possible to go from the vertex A_i to the vertex A_j because K^* is a communicating class. It follows that α^n is a stationary ergodic Markov chain on $\overline{\mathcal{W}}_n^*$. Define the sequence $\{\rho^n\}$ on Σ^ν by

$$\rho^n := \frac{1}{n+r} \sum_{i=0}^{n+r-1} S^i \alpha^n \tag{5.1}$$

Lemma 5.2 *Let $\alpha \in \mathcal{M}_S^\nu$. Then ρ^n is a stationary ergodic probability measure on Σ^ν. We have*

$$\lim_n \rho_n = \alpha$$

and

$$\liminf_n h_{\mathrm{Sh}}(\rho_n) \geq h_{\mathrm{Sh}}(\alpha).$$

Proof: Since $S^{n+r}\alpha^n = \alpha^n$, ρ is stationary. Let f and g be two local positive functions. To prove the ergodicity of ρ^n, we show that

$$\lim_k \frac{1}{k} \sum_{j=0}^{k-1} \langle\, f\, S^j g, \rho^n \,\rangle = \langle\, f, \rho^n \,\rangle \, \langle\, g, \rho^n \,\rangle \, . \tag{5.2}$$

Elementary estimates show that it suffices to do so for the case that k a multiple of $n+r$.

$$\sum_{j=0}^{k(n+r)} \langle\, f\, S^j g, \rho^n \,\rangle = \sum_{j=0}^{k} \sum_{i=0}^{n+r-1} \langle\, f\, S^{j(n+r)+i} g, \rho^n \,\rangle.$$

By the ergodicity of α^n relative to S^{n+r},

$$\begin{aligned}\lim_{k\to\infty} \frac{1}{k(n+r)} \sum_{j=0}^{k(n+r)} \langle\, f\, S^j g, \rho^n \,\rangle &= \frac{1}{n+r} \sum_{i=0}^{n+r-1} \langle\, f, \rho^n \,\rangle \langle\, S^i g, \rho^n \,\rangle \\ &= \langle\, f, \rho^n \,\rangle \langle\, g, \rho^n \,\rangle,\end{aligned}$$

since $S\rho^n = \rho^n$. This implies (5.2).

We estimate the entropy of ρ^n. $H_m(\,\cdot\,)$ is continuous and concave. Therefore

$$H_m(\rho^n) = H_m\Big(\frac{1}{n+r} \sum_{j=0}^{n+r-1} S^j \alpha^n\Big) \geq \frac{1}{n+r} \sum_{j=0}^{n+r-1} H_m(S^j \alpha^n)\,.$$

Recall that $H_m(\alpha) \equiv H_{\mathcal{F}_{[1,m]}}(\alpha)$ is the entropy of the restriction of α on the algebra $\mathcal{F}_{[1,m]}$, which is generated by the projection $X_{[1,m]}$. Denote by $\mathcal{F}_{[j,k]}$ the algebra generated by $X_{[j,k]}$. We have

$$H_{\mathcal{F}_{[1,m]}}(S^j\alpha^n) = H_{\mathcal{F}_{[j,m+j]}}(\alpha^n)\,.$$

Let now $m \equiv k(n+r)$, $k \geq 2$. Let k_1 be the smallest integer such that $j \leq k_1(n+r)$, and k_2 the largest integer such that $k_2(n+r) \leq m+j$. We decompose the interval $[j, m+j]$ in $[j, k_1(n+r)] \cup [k_1(n+r)+1, k_2(n+r)] \cup [k_2(n+r)+1, m+j]$. By monotonicity of the entropy and definition of α^n, $\quad H_{\mathcal{F}_{[j,m+j]}}(\alpha^n) \geq$

$$H_{\mathcal{F}_{[k_1(n+r)+1,k_2(n+r)]}}(\alpha^n) = H_{\mathcal{F}_{[1,(k_2-k_1)(n+r)]}}(\alpha^n) = H_{(k-1)(n+r)}(\alpha^n)\,,$$

and

$$H_{(k-1)(n+r)}(\alpha^n) = (k-1)H_n(\alpha)\,.$$

Therefore,

$$h_{\mathrm{Sh}}(\rho^n) = \lim_k \frac{1}{k(n+r)} H_{k(n+r)}(\rho^n) \geq \lim_k \frac{k-1}{k(n+r)} H_n(\alpha) = \frac{1}{n+r} H_n(\alpha)\,.$$

Given $\varepsilon > 0$, if n is large enough, then $h_{\mathrm{Sh}}(\rho^n) \geq h_{\mathrm{Sh}}(\alpha) - \varepsilon$.

Next we compare $\langle\, f, \rho^n \,\rangle$ and $\langle\, f, \alpha \,\rangle$ for $f \in \mathcal{F}_q$. Let $1 \leq q < n$ and define $\psi_1(\omega) := \psi(\omega_1^{n+r})$. The function f is defined on Σ^ν but the value of $S^j f(\omega)$ depends only on ω_{j+1}^{j+q}. Let $y := \psi_1(\omega)$. Then

$$\left| \sum_{j=0}^{n+r-1} S^j f(\omega) - \sum_{j=0}^{n-q} S^j f(y) \right| \leq (r+q-1)\|f\|\,. \tag{5.3}$$

By the shift-invariance of α

$$\sum_{j=0}^{n-q} \langle\, S^j f(\psi_1(\omega)), \alpha^n \,\rangle = (n-q+1)\langle\, f, \alpha \,\rangle\,.$$

Inequality (5.3) implies

$$\left| \langle\, f, \rho^n \,\rangle - \frac{n+1-q}{n+r} \langle\, f, \alpha \,\rangle \right| \leq \frac{r+q-1}{n+r} \|f\|\,;$$

hence $\lim_n \langle\, f, \rho^n \,\rangle = \langle\, f, \alpha \,\rangle$. □

Remark 1: That an irreducible sofic shift Σ^ν satisfies H2 is easily seen as follows. Consider a connected graph giving a minimal right resolving presentation. For each pair of vertices A, B, let $m_{A,B}$ denote the length of the shortest path from A to B. Let m be the maximum of $m_{A,B}$ over pairs of distinct vertices. Then for pair of words $w^1, w^2 \in \mathcal{L}^\nu$ there is a word $v \in \mathcal{L}^\nu$ of length not greater than m so that $w^1 v w^2 \in \mathcal{L}^\nu$; hence $g(n) = g_l(w) = g_r(w) \equiv m$ suffices.

Remark 2: The measure ν of maximal entropy of an irreducible sofic shift satisfies H3. Let ν' be the Parry measure of an edge shift Markov chain whose projection is a minimal right resolving presentation of ν (see [LM] for details). Let $A(\omega_1, \omega_2)$ be the

01-matrix corresponding to ν'. Let λ be the Perron eigenvalue and $u(\omega_1)$ and $v(\omega_1)$ be the left and right Perron eigenvectors normalized so that $\sum_{\omega_1} u(\omega_1)v(\omega_1) = 1$. Then

$$\nu'[\omega_1^n] = u(\omega_1)v(\omega_1) \prod_{i=1}^{n-1} \frac{v(\omega_{i+1})A(\omega_i, \omega_{i+1})}{\lambda\, v(\omega_i)} = \lambda^{1-n} u(\omega_1)v(\omega_n),$$

hence $e_{\nu'} :\equiv \log \lambda$ satisfies the corresponding (2.1). Since the unique measure of maximal entropy ν on the sofic shift is the image of ν' under a finite to one map (see theorem 5 in [F2]), $e_\nu :\equiv \log \lambda$ satisfies H3 if ν is non-atomic. One can show that if $\rho \in \mathcal{M}_S^\nu$ is an equilibrium state of an absolutely summable potential (see [LPS]) relative to ν, then ρ satisfies H3 when ν does.

Remark 3: A simple example of a periodic sofic shift is the double even shift: for $\omega \in \Sigma \subset \{0,1\}^{\mathbf{N}}$, whenever $\omega_j \neq \omega_{j+1} = \omega_{j+2} = \cdots = \omega_{j+k} \neq \omega_{j+k+1}$, k is even. A run of odd length could appear at the start of ω. Let ν denote the measure of maximal entropy for this shift. Then $\nu(\omega_1|\omega_2, \omega_3, \ldots)$ is not continuous at the two points $\omega \in \Sigma^\nu$ where $\omega_i = \omega_j$ for all i, j. Other sofic shifts have sets with nonzero entropy on which the corresponding conditional probability kernel is not continuous.

[B] H. Bauer, Probability Theory, Walter de Gruyter, Berlin, New-York (1996).

[B1] P. Billingsley, Hausdorff dimension in probability theory, *Ill. J. Mat.* **4**, 187–209 (1960).

[B2] P. Billingsley, Hausdorff dimension in probability theory, *Ill. J. Mat.* **5**, 291–298 (1961).

[B3] P. Billingsley, Ergodic Theory and Information, John Wiley & Sons, New York (1965).

[C] H. Cajar, Billingsley Dimension in Probability Spaces, Springer LMN 892, Berlin (1980).

[Ch] D.G. Champernowne, The construction of decimal normal in the scale of ten, *J. London Math. Soc.* **8**, 254-260 (1933).

[Cha] J.-R. Chazottes, Dimensions and Waiting Times for Gibbs Measures, *J. Stat. Phys.* **98**, 305-320 (2000).

[ChaO] J.-R. Chazottes, E. Olivier, Relative entropy, dimensions and large deviations for *g*-measures, *J. Phys. A: Math. Gen.* **33**, 675-689 (2000).

[Co] C.M. Colebrook, The Hausdorff dimension of certain sets of nonnormal numbers, *Michigan Math. J.* **17** 103–116 (1970).

[DGS] M. Denker, C. Grillenberger and K. Sigmund, Ergodic Theory on Compact Spaces, Springer LMN 527, Berlin (1976).

[EKW] A. Eizenberg, Y. Kifer and B. Weiss, Large Deviations for $\mathbf{Z}^d$-actions, *Commun. Math. Phys.* **164** 433-454 (1994).

[E] G.A. Edgar, Measure, Topology and Fractal Geometry, Springer, Berlin (1990).

[Eg] H.G. Eggleston, The fractional dimension of a set defined by decimal properties, *Quart. J. Math. Oxford Ser.* **20**, 31–36 (1949).

[F1] R. Fisher, Sofic systems and graphs, *Mh. Math.* **80**, 179-186 (1975).

[F2] R. Fisher, Graphs and symbolic dynamics, *Colloquia Mathematica Soc. J. Bolyai* **16**, 229-244 (1975).

[KH] A. Katok and B. Hasselblatt, Introduction to the Modern Theory of Dynamical Systems, Cambridge University Press, Cambridge (1995).

[L] M.H. Lebesgue, Sur certaines démonstrations d'existence, *Bull. Soc. Math. France* **45**, 182-144 (1917).

[LM] D. Lind, B. Marcus, Symbolic Dynamics and Coding, Cambridge University Press, Cambridge (1995).

[LPRS] J.T. Lewis, C.-E. Pfister, R. Russel, W.G. Sullivan, Reconstruction sequences and equipartition measures: an examination of the asymptotic equipartition property, *IEEE Information Theory* **43**, 1935-1947 (1997).

[LPS] J.T. Lewis, C.-E. Pfister and W.G. Sullivan, Generic points for stationary measures via large deviation theory *Markov Processes and Related Fields* **5**, 235–267 (1999).

[M] B. Marcus, A note on periodic points for ergodic toral automorphismus *Mh. Math.* **89**, 121-129 (1980).

[O1] E. Olivier, Dimension de Billingsley d'ensembles saturés, *C. R. Acad. Sci. Paris* **328**, 13-16 (1999).

[O2] E. Olivier, Multifractal analysis in symbolic dynamics and distribution of pointwise dimension for g-measures, *Nonlinearity* **12**, 1571-1585 (1999).

[P] Y.E. Pesin, Dimension Theory in Dynamical Systems, Contemporary Views and Applications, Chicago Lectures in Mathematics, Univerity of chicago Press, Chicago (1997).

[R] J.-M. Rey, The role of Billingsley dimensions in computing fractal dimensions on Cantor-like spaces, *Proc. A.M.S.* **128**, 561-572 (2000).

[T1] A.A. Tempelman, Dimension of Random Fractals in Metric Spaces, *Teor. Veroyatnost. i Primenen.* **44**, 589-616 (1999); translation in *Theory Probab. Appl.* **44**, 537-557 (2000).

[T2] A.A. Tempelman, Multifractal Analysis of Ergodic Averages: A Generalization of Eggleston's Theorem, *J. Dynam. Control Systems* **7**, 535-551 (2001).

[W] H. Wegmann, Über den Dimensionsbegriff in Wahrsheinlichkeitsräumen von P. Billingsley I and II, *Z. Wahrscheinlichkeitstheorie verw. Geb.* **9**, 216-221 and 222-231 (1968).