Towards a Regular Hilberg Process

Łukasz Dębowski*

Abstract

A regular Hilberg process is a hypothetical stationary measure that satisfies both a power-law growth of block entropy and a hyperlogarithmic growth of maximal repetition. Such processes may arise in statistical modeling of natural language. A puzzling property of these processes is that the length of the Lempel-Ziv code is orders of magnitude larger than the block entropy. In the present paper, we construct a class of random hierarchical association (RHA) processes, some of which may give rise to regular Hilberg processes. Whereas we have been able to prove the most important results, a few questions remain to be open problems.

Keywords: block entropy, maximal repetition, redundancy ratio, asymptotically mean stationary processes

*Ł. Dębowski is with the Institute of Computer Science, Polish Academy of Sciences, ul. Jana Kazimierza 5, 01-248 Warszawa, Poland (e-mail: ldebowsk@ipipan.waw.pl).

I Introduction

Consider a stationary measure μ on the measurable space of infinite sequences $(\mathbb{A}^{\mathbb{N}}, \mathcal{A}^{\mathbb{N}})$ from a finite alphabet $\mathbb{A} \subset \mathbb{N}$. The random symbols will be denoted as $\xi_i : \mathbb{A}^{\mathbb{N}} \ni (x_i)_{i\in\mathbb{N}} \mapsto x_i \in \mathbb{A}$, whereas blocks of symbols will be denoted as $x_{k:l} = (x_i)_{i=k}^{l}$. The expectation with respect to μ is denoted as $\mathbf{E}_\mu$. We also use shorthand $\mu(x_{1:m}) = \mu(\xi_{1:m} = x_{1:m})$. The block entropy of measure μ is function

$$H_\mu(m) := \mathbf{E}_\mu \left[-\log \mu(\xi_{1:m}) \right], \tag{1}$$

and the entropy rate of μ is the limit

$$h_\mu := \inf_{m\in\mathbb{N}} \frac{H_\mu(m)}{m} = \lim_{m\to\infty} \frac{H_\mu(m)}{m}. \tag{2}$$

The subject of this paper is motivated by a linguistic application. It is often assumed that texts in natural language can be modeled by a stationary measure and given this assumption various parameters of this hypothetical measure are estimated from the empirical data. In particular, reanalyzing Shannon's famous estimates of the entropy of printed English obtained by universal text prediction with human subjects [1], Hilberg [2] stated a hypothesis that the entropy of a block of letters is roughly proportional to the square root of the block length,

$$H_\mu(m) = \Theta\left(m^\beta\right), \tag{3}$$

where $\beta \approx 0.5$. When extrapolated to infinity, this hypothesis implies that the entropy rate of natural language production is zero, $h_\mu = 0$, and consequently natural language production would be asymptotically infinitely compressible and deterministic. Such a condition may seem incredible. Therefore, in a few later papers [3, 4, 5, 6, 7, 8, 9, 10] a relaxed condition was rather considered, namely, that the mutual information between two adjacent blocks of letters is proportional to a power of the block length

$$2H_\mu(m) - H_\mu(2m) = \Theta\left(m^\beta\right), \tag{4}$$

where $\beta \in (0, 1)$. Some recent data concerning maximal repetition and subword complexity of texts in natural language [11, 12] indicate, however, that Hilberg's original hypothesis (3) may be true.

Although Hilberg's original hypothesis (3) implies that natural language production is asymptotically infinitely compressible and deterministic, it does not necessarily mean that the length of a universal code, such as the Lempel-Ziv code, can be of the same magnitude as the block entropy. This question is very important and requires some exposition. For a uniquely decodable code C let us denote its length for blocks of length m as $|C(\xi_{1:m})|$. The code is called here universal if $\mathbf{E}_\mu |C(\xi_{1:m})| - H_\mu(m) = o(m)$ for any stationary measure μ over a given alphabet $\mathbb{A}$. Universal codes exist if and only if the alphabet is finite [13, 14] but there is no universal sublinear bound on the redundancy $\mathbf{E}_\mu |C(\xi_{1:m})| - H_\mu(m)$ [15]. Consequently, we may ask what the typical value of a *redundancy ratio* $\mathbf{E}_\mu |C(\xi_{1:m})| / H_\mu(m)$ is for a universal code if Hilberg's hypothesis (3) is satisfied. As we will see, there is no universal sublinear bound for this ratio, either. For any uniquely decodable code, the expected length of code $\mathbf{E}_\mu |C(\xi_{1:m})|$ can exceed the block entropy $H_\mu(m)$ by orders of magnitude.

Some insight into this question may be given by studying maximal repetition. Maximal repetition $L(\xi_{1:m})$ is the maximal length of a repeated substring in block $\xi_{1:m}$ [16, 17, 18, 19], formally defined as

$$L(\xi_{1:m}) := \max \left\{k : \exists_{1 \le i < j \le m-k+1} \, \xi_{i:i+k-1} = \xi_{j:j+k-1}\right\}. \tag{5}$$

As observed in [20], the length of the Lempel-Ziv code [21], being the oldest known example of a universal code, satisfies inequality

$$|C(\xi_{1:m})| \ge \frac{m}{L(\xi_{1:m}) + 1} \log \frac{m}{L(\xi_{1:m}) + 1}. \tag{6}$$

This inequality stems from a simple observation that the length of the Lempel-Ziv code is greater than $V \log V$, where V is the number of Lempel-Ziv phrases, whereas the Lempel-Ziv phrases must be shorter than the maximal repetition plus 1. Moreover, motivated by natural language data [11], which seem to satisfy relationship

$$\mathbf{E}_\mu \, L(\xi_{1:m}) = \Theta\left((\log m)^{1/\beta}\right), \tag{7}$$

we have supposed in [20] that there exists a large class of stationary measures μ that obey both conditions (3) and (7). In that case, the expected length of the Lempel-Ziv code would satisfy inequality

$$\mathbf{E}_\mu \, |C(\xi_{1:m})| = \Omega\left(\frac{m}{(\log m)^{1/\beta - 1}}\right), \tag{8}$$

where the right term is orders of magnitude larger than the entropy (3)!

Hypothetical stationary processes with measures satisfying conditions (3) and (7) have been called regular Hilberg processes in [20]. For further discussion of Hilberg's hypothesis, it may be illuminating to exhibit some explicit examples of such processes. Let us recall that in [9, 22, 23], some stationary processes were constructed which satisfy condition (4) with an entropy rate $h_\mu > 0$. In contrast, in this paper, we will introduce a class of processes, called random hierarchical association (RHA) processes. These processes are parameterized by certain parameters k_n, called perplexities, which take values in natural numbers and are related to the entropies of blocks of lengths 2^n. In this way, we can control the value of block entropy and force the entropy rate to be zero.

Formally speaking, the RHA processes are not stationary, and we ignore whether they are asymptotically mean stationary (AMS). In spite of this we have been able to show that the RHA processes have a stationary mean μ with respect to blocks, a relaxed condition called here AMSB. Subsequently, we have been able to show the following result:

Theorem 1 *For perplexities*

$$k_n = \left\lfloor \exp\left(2^{\beta n}\right) \right\rfloor, \tag{9}$$

where $0 < \beta < 1$*, the stationary mean* μ *of the RHA process satisfies the following conditions:*

(i) The entropy rate is $h_\mu = 0$*.*

(ii) The block entropy is sandwiched by

$$C_1 m \left(\frac{1}{\log m}\right)^{1/\beta-1} \le H_\mu(m) \le C_2 m \left(\frac{\log\log m}{\log m}\right)^{1/\beta-1} . \tag{10}$$

(iii) The maximal repetition and the topological entropy

$$H_{top}(m|w) := \log \operatorname{card} \{x_{1:m} : x_{1:m} \textit{ is a subsequence of } w\} \tag{11}$$

μ-almost surely satisfy conditions

$$L(\xi_{1:m}) = \Theta\left((\log m)^{1/\beta}\right), \tag{12}$$

$$H_{top}(m|\xi_{1:\infty}) = \Theta\left(m^\beta\right), \tag{13}$$

where the lower bound for the maximal repetition and the upper bound for the topological entropy are uniform.

(iv) Moreover, the maximal repetition satisfies condition (7).

(v) The random ergodic measure $F = \mu(\cdot|\mathcal{I})$, where $\mathcal{I}$ is the shift-invariant algebra [24, 25], μ-almost surely F-almost surely satisfies conditions (12)–(13) and $H_F(m) \le H_{top}(m|\xi_{1:\infty})$.

(vi) The stationary mean μ is nonergodic and the entropy of the shift invariant algebra $H_\mu(\mathcal{I})$, as defined in [26], is infinite.

We suppose that the random ergodic measure F from Theorem 1(v) defines almost surely a regular Hilberg process. At least from what we have proved so far, it follows that for the random source F, the length of the Lempel-Ziv code is almost surely orders of magnitude larger than the block entropy $H_F(m)$. Moreover, by Theorem 1(ii) and (v), the expected length of any uniquely decodable code $\mathbf{E}_\mu \mathbf{E}_F |C(\xi_{1:m})| \ge H_\mu(m)$ is orders of magnitude larger than the expected block entropy $\mathbf{E}_\mu H_F(m)$ for the random ergodic measure. In other words, on average, there is no efficient way of learning how to compress or to predict the source F from the empirical data. We can easily generate a realization for a random measure F, but if we do not know the complete measure F a priori, we cannot predict this realization efficiently. Although measure F is asymptotically deterministic, on average it escapes our capability of prediction.

We hope that our example of the RHA processes may also inspire some progress in statistical modeling of language. We have shown that processes satisfying Hilberg-like conditions (12)–(13) arise in a very simple setting of random sampling of texts from a restricted random hierarchical pool. In the linguistic application, this pool of admissible texts need not be so random. In other words, the random ergodic measure for natural language may be concentrated on some subset of ergodic measures which is a null set with respect to the RHA process. Then efficient universal text prediction by humans might be feasible. Consequently, this might explain why the estimates of entropy obtained through universal text prediction experiments with human subjects, such as [1, 27], are so low and follow relationship (3). To verify our hypothesis of nonrandomness of the pool of admissible texts, it would be advisable to identify some mechanisms that select texts for the pool. We suppose that these mechanisms are related to the idea of memetic cultural evolution [28, 29, 11].

The further organization of this paper is as follows. In Section II, the RHA processes are constructed. In Section III, it is shown that the RHA processes have a stationary mean. In Section IV, the entropy and the maximal repetition for the RHA process and its stationary mean are related. Section V concerns some further auxiliary results, such as probabilities of no repeat and a link between the topological entropy and maximal repetition. In Section VI, block entropies of the RHA processes are discussed. In Section VII, Theorem 1 is proven. Two hypotheses concerning the RHA processes are formulated in the concluding Section VIII.

II Definition of the RHA process

Let $(k_n)_{n\in\{0\}\cup\mathbb{N}}$ be some sequence of strictly positive natural numbers that satisfy

$$k_{n-1} \le k_n \le k_{n-1}^2. \tag{14}$$

Next, for each $n \in \mathbb{N}$, let $(L_{nj}, R_{nj})_{j\in\{1,...,k_n\}}$ be an independent random combination of k_n pairs of numbers from the set $\{1, ..., k_{n-1}\}$ drawn without repetition. That is, we assume that each pair (L_{nj}, R_{nj}) is different, the elements of pairs may be identical ($L_{nj} = R_{nj}$), and the sequence $(L_{nj}, R_{nj})_{j\in\{1,...,k_n\}}$ is sorted lexicographically. Formally, we assume that random variables L_{nj} and R_{nj} are supported on some probability space $(\Omega, \mathcal{J}, P)$ and we have the uniform distribution

$$P((L_{n1}, R_{n1}, ..., L_{nk_n}, R_{nk_n}) = (l_{n1}, r_{n1}, ..., l_{nk_n}, r_{nk_n})) = \binom{k_{n-1}^2}{k_n}^{-1}. \tag{15}$$

The collection of random variables (L_{nj}, R_{nj}) will be denoted as

$$\mathcal{G} = (L_{nj}, R_{nj})_{n\in\mathbb{N},j\in\{1,...,k_n\}}. \tag{16}$$

We will also use notations

$$\mathcal{G}_{\le m} = (L_{nj}, R_{nj})_{n\le m,j\in\{1,...,k_n\}}, \tag{17}$$

$$\mathcal{G}_{>m} = (L_{nj}, R_{nj})_{n>m,j\in\{1,...,k_n\}}. \tag{18}$$

Subsequently we define random variables

$$Y_j^0 = j, \qquad j \in \{1, ..., k_0\}, \tag{19}$$

$$Y_j^n = Y_{L_{nj}}^{n-1} \times Y_{R_{nj}}^{n-1}, \qquad j \in \{1, ..., k_n\}, n \in \mathbb{N}, \tag{20}$$

where $a \times b$ denotes concatenation. Hence Y_j^n are random blocks of 2^n natural numbers, whereas collection $\mathcal{G}_{\le m}$ fully determines variables Y_j^m for a fixed m.

Variables Y_j^n will be the building blocks of yet another process. For this sake, let $(C_n)_{n\in\{0\}\cup\mathbb{N}}$ be a sequence of independent random variables, independent from $\mathcal{G}$, with uniform distribution

$$P(C_n = j) = 1/k_n, \qquad j \in \{1, ..., k_n\}. \tag{21}$$

Definition 1 *The random hierarchical association (RHA) process $\mathcal{X}$ with perplexities $(k_n)_{n\in\{0\}\cup\mathbb{N}}$ is defined as*

$$\mathcal{X} = Y^0_{C_0} \times Y^1_{C_1} \times Y^2_{C_2} \times \dots . \tag{22}$$

It is convenient to define a few more random variables for the RHA process. First, sequence $\mathcal{X}$ will be parsed into a sequence of numbers X_j, where

$$\mathcal{X} = X_1 \times X_2 \times X_3 \times \dots, \tag{23}$$

and, more generally, into a sequence of blocks X^n_j of length 2^n, where

$$\mathcal{X} = Y^1_{C_0} \times Y^1_{C_1} \times Y^2_{C_2} \times \dots \times Y^n_{C_n} (= X^n_1) \times X^n_2 \times X^n_3 \times \dots . \tag{24}$$

Let us also observe that there exist unique random variables K_{nj} such that

$$X^n_j = Y^n_{K_{nj}}. \tag{25}$$

Moreover, we also denote blocks starting at any position as

$$X_{k:l} = X_k \times X_{k+1} \times \dots \times X_l \tag{26}$$

and for $n \geq 1$ as

$$X^n_{k:l} = X^n_k \times X^n_{k+1} \times \dots \times X^n_l. \tag{27}$$

III Stationary mean

Let us introduce shift operation $T : \mathbb{A}^{\mathbb{N}} \ni (x_i)_{i\in\mathbb{N}} \mapsto (x_{i+1})_{i\in\mathbb{N}} \in \mathbb{A}^{\mathbb{N}}$ and shift invariant algebra $\mathcal{I} = \left\{A \in \mathcal{A}^{\mathbb{N}} : T^{-1}A = A\right\}$. We recall this definition:

Definition 2 *[30] A measure ν on $(\mathbb{A}^{\mathbb{N}}, \mathcal{A}^{\mathbb{N}})$ is called asymptotically mean stationary (AMS) if limits*

$$\mu(A) := \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \nu(T^{-i}A) \tag{28}$$

exist for every event $A \in \mathcal{A}^{\mathbb{N}}$.

For an AMS measure ν, function μ is a stationary measure on $(\mathbb{A}^{\mathbb{N}}, \mathcal{A}^{\mathbb{N}})$, called the stationary mean of ν. Moreover, measures μ and ν are equal on the shift invariant algebra, i.e., $\mu(A) = \nu(A)$ for all $A \in \mathcal{I}$.

Now, let $\mathbb{A}^+ = \bigcup_{n\in\mathbb{N}} \mathbb{A}^n$. There is a related relaxed condition:

Definition 3 *A measure ν on $(\mathbb{A}^{\mathbb{N}}, \mathcal{A}^{\mathbb{N}})$ is called asymptotically mean stationary with respect to blocks (AMSB) if limits*

$$\mu(x_{1:m}) := \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \nu(\xi_{i:i+m-1} = x_{1:m}) \tag{29}$$

exist for every block $x_{1:m} \in \mathbb{A}^+$.

For an AMSB measure ν over a finite alphabet $\mathbb{A}$, function μ, extended via $\mu(\xi_{1:m} = x_{1:m}) := \mu(x_{1:m})$, is also a stationary measure on $(\mathbb{A}^{\mathbb{N}}, \mathcal{A}^{\mathbb{N}})$. We shall continue to call this μ a stationary mean of ν. However, an AMSB measure need not be AMS, cf. [31, Example 6.3], and for an AMSB measure ν we need not have $\mu(A) = \nu(A)$ for shift invariant events $A \in \mathcal{I}$.

Subsequently, we will show that the RHA process is AMSB. First we will prove this property.

Theorem 2 *Variables K_{nj} are independent from $\mathcal{G}_{\le n}$ and satisfy*

$$P(K_{nj} = l, K_{n,j+1} = m) = 1/k_n^2, \qquad l, m \in \{1, ..., k_n\}, j \in \mathbb{N}. \tag{30}$$

Proof: Each K_{nj} is a function of C_q for some $q \ge n$ and $\mathcal{G}_{>n}$. Hence K_{nj} are independent from $\mathcal{G}_{\le n}$.

Now we will show by induction on j that (30) is satisfied.

The induction begins with $K_{n1} = C_n$ and $K_{n2} = L_{n+1,C_{n+1}}$. These two variables are independent by definition and by definition K_{n1} is uniformly distributed on $\{1, ..., k_n\}$. It remains to show that so is K_{n2}. Observe that $(L_{n+1,k}, R_{n+1,k})$ are independent of C_{n+1}. Hence for $l, m \in \{1, ..., k_n\}$ we obtain

$$\begin{aligned} P(K_{n2} = l, K_{n3} = m) &= \sum_{k=1}^{k_{n+1}} P(L_{n+1,k} = l, R_{n+1,k} = m) P(C_{n+1} = k) \\ &= \frac{1}{k_{n+1}} \sum_{k=1}^{k_{n+1}} P(L_{n+1,k} = l, R_{n+1,k} = m) \\ &= \frac{1}{k_{n+1}} \binom{k_n^2}{k_{n+1}}^{-1} \binom{k_n^2 - 1}{k_{n+1} - 1} = \frac{1}{k_{n+1}} \frac{k_{n+1}}{k_n^2} = \frac{1}{k_n^2}, \end{aligned}$$

so K_{n2} is uniformly distributed on $\{1, ..., k_n\}$.

The inductive step is as follows: (i) if $K_{n+1,j}$ is uniformly distributed on $\{1, ..., k_{n+1}\}$ then $(K_{n,2j}, K_{n,2j+1}) = (L_{n+1,K_{n+1,j}}, R_{n+1,K_{n+1,j}})$ is uniformly distributed on $\{1, ..., k_n\} \times \{1, ..., k_n\}$, and (ii) if $(K_{n+1,j}, K_{n+1,j+1})$ is uniformly distributed on $\{1, ..., k_{n+1}\} \times \{1, ..., k_{n+1}\}$ then $(K_{n,2j+1}, K_{n,2j+2}) = (R_{n+1,K_{n+1,j}}, L_{n+1,K_{n+1,j+1}})$ is uniformly distributed on $\{1, ..., k_n\} \times \{1, ..., k_n\}$. Now observe that $(L_{n+1,k}, R_{n+1,k})$ are independent of $K_{n+1,j}$. Hence, for $l, m \in \{1, ..., k_n\}$ we obtain

$$\begin{aligned} &P(K_{n,2j} = l, K_{n,2j+1} = m) \\ &= \sum_{k=1}^{k_{n+1}} P(L_{n+1,k} = l, R_{n+1,k} = m) P(K_{n+1,j} = k) \\ &= \frac{1}{k_{n+1}} \sum_{k=1}^{k_{n+1}} P(L_{n+1,k} = l, R_{n+1,k} = m) \\ &= \frac{1}{k_{n+1}} \binom{k_n^2}{k_{n+1}}^{-1} \binom{k_n^2 - 1}{k_{n+1} - 1} = \frac{1}{k_{n+1}} \frac{k_{n+1}}{k_n^2} = \frac{1}{k_n^2}, \end{aligned}$$

which proves claim (i). On the other hand, for $l, m \in \{1, ..., k_n\}$ we obtain

$$
\begin{aligned}
& P(K_{n,2j+1} = l, K_{n,2j+2} = m) \\
&= \sum_{p,q=1}^{k_{n+1}} P(R_{n+1,p} = l, L_{n+1,q} = m) P(K_{n+1,j} = p, K_{n+1,j+1} = q) \\
&= \frac{1}{k_{n+1}^2} \sum_{p,q=1}^{k_{n+1}} P(R_{n+1,p} = l, L_{n+1,q} = m) \\
&= \frac{1}{k_{n+1}^2} \sum_{p=1}^{k_{n+1}} P(R_{n+1,p} = l, L_{n+1,p} = m) \\
&\quad + \frac{1}{k_{n+1}^2} \sum_{p,q=1,\, p\neq q}^{k_{n+1}} P(R_{n+1,p} = l, L_{n+1,q} = m) \\
&= \frac{1}{k_{n+1}^2} \binom{k_n^2}{k_{n+1}}^{-1} \left(\binom{k_n^2 - 1}{k_{n+1} - 1} + (k_n^2 - 1) \binom{k_n^2 - 2}{k_{n+1} - 2} \right) \\
&= \frac{1}{k_{n+1}^2} \left(\frac{k_{n+1}}{k_n^2} + (k_n^2 - 1) \frac{k_{n+1}(k_{n+1} - 1)}{k_n^2 (k_n^2 - 1)} \right) = \frac{1}{k_n^2},
\end{aligned}
$$

which proves claim (ii). □

Now we demonstrate the main result of this section.

Theorem 3 *The RHA process is AMSB. In particular, for $m \le 2^n$ and $k \in \mathbb{N}$, the stationary mean is*

$$
\mu(x_{1:m}) = \frac{1}{2^n} \sum_{j=0}^{2^n - 1} P(X_{k2^n+j:k2^n+j+m-1} = x_{1:m}). \tag{31}
$$

Proof: Block $X_{k2^n+j:k2^n+j+m-1}$ is a subsequence of $X^n_{k:k+1}$ for $m \le 2^n$, $k \in \mathbb{N}$, and $0 \le j < 2^n$. In particular, there exist functions f_{mj} such that

$$
X_{k2^n+j:k2^n+j+m-1} = f_{mj}(X^n_{k:k+1}).
$$

Hence probabilities $P(X_{i:i+m-1} = x_{1:m})$ are periodic by Theorem 2. This implies the formula for $\mu(x_{1:m})$. □

We suppose that the RHA process is AMS but we could not prove it so far.

IV Bounds for the stationary mean

In this section we want to derive some bounds for the entropy and the maximal repetition of the stationary mean of the RHA process from the analogical bounds for blocks X^n_j. In the following we will denote $X^n_{kj} = X_{k2^n+j:k2^n+j+2^n-1}$. (We have $X^n_{k0} = X^n_k$.) Subsequently, for entropy $H(X) = \mathbf{E}\left[-\log P(X)\right]$, we obtain:

Theorem 4 *For the stationary mean μ of the RHA process, we have*

$$
H(X^{n-1}_j) \le H_\mu(2^n) \le H(X^{n+1}_j) + n \log 2. \tag{32}
$$

Proof: By the Jensen inequality for function $p \mapsto -p \log p$ and Theorem 3, we hence obtain

$$H_\mu(2^n) \geq \frac{1}{2^n} \sum_{j=0}^{2^n-1} H(X^n_{kj}). \tag{33}$$

Now we observe that for each $k \geq 1$ and j there exists a q such that X^{n-1}_q is a subsequence of X^n_{kj}. Thus we have $H(X^n_{kj}) \geq H(X^{n-1}_q)$. This combined with inequality (33) yields $H(X^{n-1}_j) \leq H_\mu(2^n)$. On the other hand, using inequality $\mu(x_{1:2^n}) \geq 2^{-n} P(X^n_{kj} = x_{1:2^n})$ and Theorem 3, we obtain

$$H_\mu(2^n) \leq \frac{1}{2^n} \sum_{j=0}^{2^n-1} H(X^n_{kj}) + n \log 2. \tag{34}$$

Now we observe that for each $k > 1$ and j there exists a q such that X^n_{kj} is a subsequence of X^{n+1}_q. Thus we have $H(X^n_{kj}) \leq H(X^{n+1}_q)$. This combined with inequality (34) yields $H_\mu(2^n) \leq H(X^{n+1}_j) + n \log 2$. $\square$

Analogically, we can bound the maximal repetition of the stationary mean. The result will be stated more generally. We will say that a function $\phi : \mathbb{A}^+ \to \mathbb{R}$ is increasing if for u being a subsequence of w, we have $\phi(u) \leq \phi(w)$. Examples of increasing functions include the maximal repetition $L(w)$, the topological entropy $H_{top}(m|w)$, and the indicator function $\mathbf{1}\{\phi(w) > k\}$, where ϕ is increasing.

Theorem 5 *For the stationary mean μ of the RHA process and an increasing function ϕ, we have*

$$\mathbf{E}_P\, \phi(X^{n-1}_j) \leq \mathbf{E}_\mu\, \phi(\xi_{1:2^n}) \leq \mathbf{E}_P\, \phi(X^{n+1}_j). \tag{35}$$

Proof: By Theorem 3,

$$\mathbf{E}_\mu\, \phi(\xi_{1:2^n}) = \frac{1}{2^n} \sum_{j=0}^{2^n-1} \mathbf{E}_P\, \phi(X^n_{kj}). \tag{36}$$

Now we observe that for each $k \geq 1$ and j there exists a q such that X^{n-1}_q is a subsequence of X^n_{kj}. Thus we have $\phi(X^n_{kj}) \geq \phi(X^{n-1}_q)$. This combined with equality (36) yields $\mathbf{E}_P\, \phi(X^{n-1}_j) \leq \mathbf{E}_\mu\, \phi(\xi_{1:2^n})$. On the other hand, for each $k > 1$ and j there exists a q such that X^n_{kj} is a subsequence of X^{n+1}_q. Thus we have $\phi(X^n_{kj}) \leq \phi(X^{n+1}_q)$. This combined with equality (36) yields $\mathbf{E}_\mu\, \phi(\xi_{1:2^n}) \leq \mathbf{E}_P\, \phi(X^{n+1}_j)$. $\square$

Hence, to obtain the desired bounds for the stationary mean, it suffices to investigate the distribution of blocks X^n_j.

V Further auxiliary results

We begin with a simple upper bound for the topological entropy and a lower bound for the maximal repetition of blocks X^n_j in terms of perplexity.

Theorem 6 *If $H_{top}(m|\xi_{1:\infty}) < \log(k-m+1)$ then $L(\xi_{j:j+k-1}) \geq m$ for any j.*

Proof: Cf. [11, Lemma 1]: String $\xi_{j:j+k-1}$ contains $k-m+1$ substrings of length m (on overlapping positions). Among them there can be at most $\exp(H_{top}(m|\xi_{1:\infty}))$ different substrings. Since $\exp(H_{top}(m|\xi_{1:\infty})) < k-m+1$, there must be some repeat of length n. Hence $L(\xi_{j:j+k-1}) \geq m$. □

Theorem 7 *For the RHA process, almost surely*

$$H_{top}(2^m|\mathcal{X}) \leq 2\log k_m. \tag{37}$$

Proof: For a given realization of the RHA process (i.e., for fixed Y_j^m), there are at most k_m different values of blocks X_j^m. Therefore, there are at most k_m^2 different values of blocks X_{kj}^m in sequence $\mathcal{X}$. □

Obtaining a lower bound for the entropy and an upper bound for the maximal repetition of blocks X_j^n is more involved. These topics will be discussed in the following sections. For this goal, we will consider events $A_{n,-1} := \emptyset$ and

$$A_{nm} := (X_1^n \text{ consists of } 2^{n-m} \text{ distinct blocks } X_j^m) \tag{38}$$

We have $P(A_{nn}) = 1$ and $A_{nm} \supset A_{n,m-1}$.

Theorem 8 *For the RHA process, we have $P(A_{nm}) = 0$ for $k_m < 2^{n-m}$, whereas for $k_m \geq 2^{n-m}$ and $m < n$ we have*

$$P(A_{nm}) = P(A_{n,m+1})\frac{k_m(k_m-1)\dots(k_m-2^{n-m}+1)}{k_m^2(k_m^2-1)\dots(k_m^2-2^{n-m-1}+1)}. \tag{39}$$

Proof: There are no more than k_m distinct blocks X_j^m in block X_1^n. Thus $P(A_{nm}) = 0$ for $k_m < 2^{n-m}$. Now assume $k_m \geq 2^{n-m}$. Introduce random variables D_{mi} such that $X_1^n = Y_{D_{m1}}^m \times ... \times Y_{D_{m2^{n-m}}}^m$. Consider probabilities $p_m = P(D_{m1} = d_1, ..., D_{m2^{n-m}} = d_{2^{n-m}})$, where d_i are distinct. It can be easily shown by induction on decreasing m that p_m do not depend on d_i and satisfy

$$p_m = p_{m+1}\binom{k_m^2}{k_{m+1}}^{-1}\binom{k_m^2 - 2^{n-m-1}}{k_{m+1} - 2^{n-m-1}}.$$

Moreover, since p_m do not depend on d_i, we obtain $P(A_{nm}) = p_m k_m(k_m - 1)\dots(k_m - 2^{n-m}+1)$. Hence the claim follows. □

VI Block entropy

For random variables X, Y and Z, where X is discrete whereas Y and Z need not be so, besides entropy $H(X) = \mathbf{E}[-\log P(X)]$, we define conditional entropy $H(X|Y) = \mathbf{E}[-\log P(X|Y)]$, mutual information $I(X;Y) := H(X) - H(X|Y)$, and conditional mutual information $I(X;Y|Z) := H(X|Z) - H(X|Y,Z)$. Given these objects, we will bound the entropies of blocks X_j^n.

Theorem 9 *We have*

$$H(X_j^n|\mathcal{G}_{\leq n}) = \log k_n \tag{40}$$

and $I(X_j^n; X_{j+1}^n|\mathcal{G}_{\leq n}) = 0$.

Proof: Given $\mathcal{G}_{\le n}$, the correspondence between X_j^n and K_{nj} is one-to-one. Hence $H(X_j^n|\mathcal{G}_{\le n}) = H(K_{nj}|\mathcal{G}_{\le n})$. From Theorem 2 we further obtain $H(K_{nj}|\mathcal{G}_{\le n}) = H(K_{nj}) = \log k_n$ and $H(K_{nj}, K_{n,j+1}|\mathcal{G}_{\le n}) = H(K_{nj}) + H(K_{n,j+1})$. □

Theorem 10 *We have*

$$H(\mathcal{G}_{\le n}) = \sum_{l=1}^{n} \log \binom{k_{l-1}^2}{k_l}. \tag{41}$$

Proof: The claim follows by chain rule $H(\mathcal{G}_{\le n}) = H(\mathcal{G}_{\le n-1}) + H(\mathcal{G}_{\le n}|\mathcal{G}_{\le n-1})$ from $H(\mathcal{G}_{\le 0}) = 0$ and $H(\mathcal{G}_{\le n}|\mathcal{G}_{\le n-1}) = \log \binom{k_{n-1}^2}{k_n}$. □

Theorem 11 *We have*

$$H(X_j^n) \le \min_{0\le l\le n} \left(H(\mathcal{G}_{\le l}) + 2^{n-l} \log k_l \right). \tag{42}$$

Proof: For any $0 \le l \le n$ we have $H(X_j^n) \le H(X_j^n, \mathcal{G}_{\le l}) = H(X_j^n|\mathcal{G}_{\le l}) + H(\mathcal{G}_{\le l})$, whereas $H(X_j^n|\mathcal{G}_{\le l}) \le 2^{n-l} H(K_{lj}|\mathcal{G}_{\le l}) = 2^{n-l} H(K_{lj}) = 2^{n-l} \log k_l$. □

Hence we may introduce an important parameter of the RHA process.

Definition 4 *The combinatorial entropy rate of the RHA process is*

$$h := \inf_{l\in\mathbb{N}} 2^{-l} \log k_l = \lim_{l\to\infty} 2^{-l} \log k_l. \tag{43}$$

Theorem 12 *We have*

$$\inf_{n\in\mathbb{N}} 2^{-n} H(X_j^n) = h. \tag{44}$$

Proof: On the one hand, by Theorem 9,

$$\inf_{n\in\mathbb{N}} 2^{-n} H(X_j^n) \ge \inf_{n\in\mathbb{N}} 2^{-n} H(X_j^n|\mathcal{G}_{\le n}) = \inf_{l\in\mathbb{N}} 2^{-l} \log k_l.$$

On the other hand, by Theorem 11,

$$\inf_{n\in\mathbb{N}} 2^{-n} H(X_j^n) \le \inf_{l\in\mathbb{N}} \inf_{n\in\mathbb{N}} \left(2^{-n} H(\mathcal{G}_{\le l}) + 2^{-l} \log k_l \right) = \inf_{l\in\mathbb{N}} 2^{-l} \log k_l.$$

□

The above result combined with Theorem 4 yields a bound for the entropy rate of the stationary mean of the RHA process.

Theorem 13 *For the stationary mean μ of the RHA process, we have*

$$h/2 \le h_\mu \le 2h. \tag{45}$$

Remark: In particular, the combinatorial entropy rate vanishes ($h = 0$) if and only if the entropy rate of the stationary mean vanishes ($h_\mu = 0$) as well.

Proof: Divide inequality (32) by 2^n and take the infimum. □

Inequality $H(X_j^n) \ge H(X_j^n|\mathcal{G}_{\le n}) = \log k_n$ gives a certain lower bound for the block entropy of the RHA process. For perplexities (9), this lower bound is orders of magnitude smaller than the upper bound (42). Concluding this section we would like to produce a lower bound which is of comparable order to (42).

Theorem 14 *We have*

$$H(X_j^n) \ge \max_{0\le l\le n} \left(\log \binom{k_{l-1}^2}{k_l} - \log \binom{k_{l-1}^2 - 2^{n-l}}{k_l - 2^{n-l}} \right) P(A_{nl}), \tag{46}$$

where $P(A_{nl})$ are the probabilities of no repeat (38).

Proof: We have

$$H(X_j^n) \ge I(X_j^n; \mathcal{G}_{\le l}|\mathcal{G}_{\le l-1}) = H(\mathcal{G}_{\le l}|\mathcal{G}_{\le l-1}) - H(\mathcal{G}_{\le l}|\mathcal{G}_{\le l-1}, X_j^n).$$

We have $H(\mathcal{G}_{\le n}|\mathcal{G}_{\le n-1}) = \log \binom{k_{n-1}^2}{k_n}$. As for $H(\mathcal{G}_{\le l}|\mathcal{G}_{\le l-1}, X_j^n)$, given X_j^n consisting of 2^{n-l} distinct blocks of length 2^l, tuple $(L_{lj}, R_{lj})_{j\in\{1,\dots,k_l\}}$ may assume at most $\binom{k_{l-1}^2 - 2^{n-l}}{k_l - 2^{n-l}}$ distinct values. Hence

$$H(\mathcal{G}_{\le l}|\mathcal{G}_{\le l-1}, X_j^n) \le P(A_{nl}) \log \binom{k_{l-1}^2 - 2^{n-l}}{k_l - 2^{n-l}},$$

from which the claim follows. □

VII Main result

Now we can demonstrate the main result.

Proof of Theorem 1:

(i) For perplexities (9) the combinatorial entropy rate is $h = 0$. Hence $h_\mu = 0$ by Theorem 13.

(ii) Expression (41) can be bounded as

$$H(\mathcal{G}_{\le n}) = \sum_{l=1}^{n} \log \binom{k_{l-1}^2}{k_l} \le \sum_{l=1}^{n} 2k_l \log k_{l-1} \le 2nk_n \log k_n.$$

Hence, from (42), for $0 \le l \le n$ we obtain an upper bound:

$$H(X_j^n) \le \left(2lk_l + 2^{n-l}\right) \log k_l.$$

If we choose $l = \left\lfloor \beta^{-1} \log_2 \left(\frac{n \log 2}{\log n}\right) \right\rfloor$ then for perplexities (9) we obtain

$$\begin{aligned} H(X_j^n) &\le \left[2\beta^{-1} \log_2 \left(\frac{n\log 2}{\log n}\right) 2^{n/\log n} + 2^n \left(\frac{n \log 2}{\log n}\right)^{-1/\beta} \right] \frac{n \log 2}{\log n} \\ &= \Theta\left(2^n \left(\frac{\log n}{n}\right)^{1/\beta - 1} \right). \end{aligned}$$

On the other hand, from (46) and (39), for $0 \le l \le n$ we have

$$H(X_j^n) \ge \left(\log \binom{k_{l-1}^2}{k_l} - \log \binom{k_{l-1}^2 - 2^{n-l}}{k_l - 2^{n-l}} \right) P(A_{nl}) \\ \ge 2^{n-l} \log \left(\frac{k_{l-1}^2 - 2^{n-l} + 1}{k_l - 2^{n-l} + 1} \right) P(A_{nl}),$$

where

$$P(A_{nl}) = \prod_{m=l}^{n-1} \frac{k_m (k_m - 1) \dots (k_m - 2^{n-m} + 1)}{k_m^2 (k_m^2 - 1) \dots (k_m^2 - 2^{n-m-1} + 1)} \\ \ge \prod_{m=l}^{n-1} \left(\frac{(k_m - 2^{n-m} + 2)(k_m - 2^{n-m} + 1)}{k_m^2 - 2^{n-m-1} + 1} \right)^{2^{n-m-1}} \\ \ge \left(\frac{(k_l - 2^{n-l} + 2)(k_l - 2^{n-l} + 1)}{k_l^2 - 2^{n-l-1} + 1} \right)^{\sum_{m=l}^{n-1} 2^{n-m-1}} \\ \ge \left(1 - \frac{k_l (2^{n-l+1} - 3) + 2}{k_l^2 - 2^{n-l-1} + 1} \right)^{2^n} \\ \ge \left(1 - 2^n \frac{k_l (2^{n-l+1} - 3) + 2}{k_l^2 - 2^{n-l-1} + 1} \right). \tag{47}$$

If we choose $l = \lceil \beta^{-1} \log_2(2n) \rceil$ then for perplexities (9) we obtain that $k_l > \exp(2n) > 2^{2n}$. Hence $P(A_{nl})$ is greater than a certain constant $\alpha > 0$ and

$$H(X_j^n) \ge \alpha 2^n (2n)^{-1/\beta} [2^{1-\beta} - 1] 2n = \Theta \left(2^n \left(\frac{1}{n} \right)^{1/\beta - 1} \right).$$

Hence, from Theorem 4, we obtain the desired sandwich bound for the entropy of the stationary mean.

(iii) By Theorem 7 and Theorem 5 we obtain

$$0 = \mathbf{E}_P \mathbf{1}\{H_{top}(2^m|\mathcal{X}) > 2 \log k_m\} \\ \ge \mathbf{E}_P \mathbf{1}\left\{H_{top}(2^m|X_j^{n+1}) > 2 \log k_m\right\} \\ \ge \mathbf{E}_\mu \mathbf{1}\{H_{top}(2^m|\xi_{1:2^n}) > 2 \log k_m\}.$$

Hence μ-almost surely $H_{top}(2^m|\xi_{1:\infty}) \le 2 \log k_m = 2^{\beta m + 1}$, which implies the upper bound $H_{top}(m|\xi_{1:\infty}) < C_1 m^\beta$ for a certain constant C_1. From this we obtain the lower bound $L(\xi_{1:m}) > C_2 (\log m)^{1/\beta}$ by Theorem 6.

As for the converse bounds, we have $L(X_1^n) \ge 2^l$ for A_{nl}^c, where A_{nl} are the events of no repeat (38). Hence by Theorem 5

$$\mathbf{E}_\mu \mathbf{1}\{L(\xi_{1:2^n}) \ge l\} \le \mathbf{E}_P \mathbf{1}\left\{L(X_l^{n+1}) \ge l\right\} \le 1 - P(A_{n+1,l}).$$

Now, if we choose $l = \lceil \beta^{-1} \log_2(2n) \rceil$ then for perplexities (9) we obtain that $k_l > \exp(2n) > 2^{2n}$. Hence, by (47), $\sum_{n=0}^{\infty} (1 - P(A_{n+1,l})) < \infty$. Consequently, by the Borel-Cantelli lemma $L(\xi_{1:2^n}) < l$ must hold for sufficiently large n μ-almost surely. Thus $L(\xi_{1:m}) < C_3(\log m)^{1/\beta}$ for sufficiently large m. From this we obtain the lower bound $H_{top}(m|\xi_{1:\infty}) > C_4 m^\beta$ for sufficiently large m by Theorem 6.

(iv) The lower bound follows from the uniform bound in (iii). As for the desired upper bound, we have $2^{m-1} \le L(X_1^n) < 2^{m+1}$ for $A_{nm} \setminus A_{n,m-1}$, where A_{nm} are the events of no repeat (38). Denoting

$$\lambda_n := \sum_{m=0}^{n} 2^m (P(A_{nm}) - P(A_{n,m-1})),$$

we hence obtain

$$\lambda_n/2 \le \mathbf{E}_P \, L(X_j^n) < 2\lambda_n.$$

To bound λ_n we will apply formula (39) in the following way. If we choose $l = \lceil \beta^{-1} \log_2(2n) \rceil$ then for perplexities (9) we have

$$\begin{aligned}
\lambda_n &\le \sum_{m=0}^{n} 2^m (1 - P(A_{n,m-1}|A_{nm})) \\
&\le \sum_{m=0}^{l-1} 2^m + \sum_{m=l}^{n} 2^m \left(1 - \frac{k_m(k_m-1)\dots(k_m - 2^{n-m}+1)}{k_m^2(k_m^2-1)\dots(k_m^2 - 2^{n-m-1}+1)} \right) \\
&\le 2^l + \sum_{m=l}^{n} 2^m \left(1 - \left(\frac{(k_m - 2^{n-m}+2)(k_m - 2^{n-m}+1)}{k_m^2 - 2^{n-m-1}+1} \right)^{2^{n-m-1}} \right) \\
&= 2^l + \sum_{m=l}^{n} 2^m \left(1 - \left(1 - \frac{k_m(2^{n-m+1}-3)+2}{k_m^2 - 2^{n-m-1}+1} \right)^{2^{n-m-1}} \right) \\
&\le 2^l + \sum_{m=l}^{n} 2^m 2^{n-m-1} \frac{k_m(2^{n-m+1}-3)+2}{k_m^2 - 2^{n-m-1}+1} \\
&\le 2^l + \sum_{m=l}^{n} 2^n \frac{2^{n-m+1}+1}{k_m - 1} \le 2^l + O(1),
\end{aligned}$$

since $k_l > \exp(2n) > 2^{2n}$. Hence $\mathbf{E}_P \, L(X_j^n) \le C n^{1/\beta}$ for a certain constant C, which implies the upper bound $\mathbf{E}_\mu \, L(\xi_1^m) = O\left((\log m)^{1/\beta}\right)$ by Theorem 5.

(v) We have $\mu = \int F d\mu$. Hence every event of full measure μ must be μ-almost surely an event of full measure F. Moreover, by the ergodic theorem, μ-almost surely we have

$$H_{top}(m|\xi_{1:\infty}) \ge \log \operatorname{card} \{x_{1:m} : F(x_{1:m}) > 0\} \ge H_F(m).$$

Thus the present claim follows from claim (iii).

(vi) The entropy of the shift-invariant algebra may be bounded by mutual information

$$\begin{aligned}
H_\mu(\mathcal{I}) = \lim_{m\to\infty} I_\mu(\mathcal{I};\xi_{1:m}) &= \lim_{m\to\infty} [H_\mu(\xi_{1:m}) - H_\mu(\xi_{1:m}|\mathcal{I})] \\
&= \lim_{m\to\infty} [H_\mu(m) - \mathbf{E}_\mu \, H_F(m)] = \infty.
\end{aligned}$$

Since the entropy of the shift-invariant algebra is positive, the measure μ is nonergodic.

□

VIII Open problems

We conclude this paper with the following hypotheses:

Conjecture 1 *The RHA processes are AMS.*

Conjecture 2 *For perplexities (9) where* $0 < \beta < 1$*, the stationary mean* μ *of the RHA process satisfies the following conditions:*

(i) the stationary mean μ *is strongly nonergodic, i.e., the shift-invariant algebra* $\mathcal{I}$ *is nonatomic w.r.t. the stationary mean,*

(ii) the random ergodic component $F = \mu(\cdot|\mathcal{I})$ *is almost surely a regular Hilberg process, i.e., apart from what was stated in Theorem 1, it* μ*-almost surely satisfies conditions*

$$\mathbf{E}_F\, L(\xi_{1:m}) = O\left((\log m)^{1/\beta}\right), \tag{48}$$

$$H_F(m) = \Omega\left(m^{\beta}\right). \tag{49}$$

References

[1] C. Shannon, "Prediction and entropy of printed English," *Bell Syst. Tech. J.*, vol. 30, pp. 50–64, 1951.

[2] W. Hilberg, "Der bekannte Grenzwert der redundanzfreien Information in Texten — eine Fehlinterpretation der Shannonschen Experimente?" *Frequenz*, vol. 44, pp. 243–248, 1990.

[3] W. Ebeling and G. Nicolis, "Entropy of symbolic sequences: the role of correlations," *Europhys. Lett.*, vol. 14, pp. 191–196, 1991.

[4] ——, "Word frequency and entropy of symbolic sequences: a dynamical perspective," *Chaos Sol. Fract.*, vol. 2, pp. 635–650, 1992.

[5] W. Ebeling and T. Pöschel, "Entropy and long-range correlations in literary English," *Europhys. Lett.*, vol. 26, pp. 241–246, 1994.

[6] W. Bialek, I. Nemenman, and N. Tishby, "Predictability, complexity and learning," *Neural Comput.*, vol. 13, p. 2409, 2001.

[7] ——, "Complexity through nonextensivity," *Physica A*, vol. 302, pp. 89–99, 2001.

[8] J. P. Crutchfield and D. P. Feldman, "Regularities unseen, randomness observed: The entropy convergence hierarchy," *Chaos*, vol. 15, pp. 25–54, 2003.

[9] Ł. Dębowski, "On the vocabulary of grammar-based codes and the logical consistency of texts," *IEEE Trans. Inform. Theory*, vol. 57, pp. 4589–4599, 2011.

[10] ——, "Hilberg exponents: New measures of long memory in the process," 2014, http://arxiv.org/abs/1403.1757.

[11] ——, "Maximal repetitions in written texts: Finite energy hypothesis vs. strong Hilberg conjecture," 2014, http://www.ipipan.waw.pl/~ldebowsk/.

[12] ——, "Estimation of entropy from subword complexity," 2014, http://www.ipipan.waw.pl/~ldebowsk/.

[13] L. D. Davisson, "Universal noiseless coding," *IEEE Trans. Inform. Theory*, vol. 19, pp. 783–795, 1973.

[14] L. Györfi, I. Páli, and E. C. van der Meulen, "There is no universal source code for infinite alphabet," *IEEE Trans. Inform. Theory*, vol. 40, pp. 267–271, 1994.

[15] P. C. Shields, "Universal redundancy rates don't exist," *IEEE Trans. Inform. Theory*, vol. IT-39, pp. 520–524, 1993.

[16] A. de Luca, "On the combinatorics of finite words," *Theor. Comput. Sci.*, vol. 218, pp. 13–39, 1999.

[17] P. C. Shields, "String matching: The ergodic case," *Ann. Probab.*, vol. 20, pp. 1199–1203, 1992.

[18] ——, "String matching bounds via coding," *Ann. Probab.*, vol. 25, pp. 329–336, 1997.

[19] R. Kolpakov and G. Kucherov, "On maximal repetitions in words," *J. Discr. Algor.*, vol. 1, pp. 159–186, 1999.

[20] Ł. Dębowski, "Hilberg's conjecture — a challenge for machine learning," 2015, http://www.ipipan.waw.pl/~ldebowsk/.

[21] J. Ziv and A. Lempel, "A universal algorithm for sequential data compression," *IEEE Trans. Inform. Theory*, vol. 23, pp. 337–343, 1977.

[22] Ł. Dębowski, "Mixing, ergodic, and nonergodic processes with rapidly growing information between blocks," *IEEE Trans. Inform. Theory*, vol. 58, pp. 3392–3401, 2012.

[23] ——, "On hidden Markov processes with infinite excess entropy," *J. Theor. Probab.*, vol. 27, pp. 539–551, 2014.

[24] O. Kallenberg, *Foundations of Modern Probability.* Springer, 1997.

[25] R. M. Gray and L. D. Davisson, "The ergodic decomposition of stationary discrete random processses," *IEEE Trans. Inform. Theory*, vol. 20, pp. 625–636, 1974.

[26] Ł. Dębowski, "A general definition of conditional information and its application to ergodic decomposition," *Statist. Probab. Lett.*, vol. 79, pp. 1260–1268, 2009.

[27] T. M. Cover and R. C. King, "A convergent gambling estimate of the entropy of English," *IEEE Trans. Inform. Theory*, vol. 24, pp. 413–421, 1978.

[28] R. Dawkins, *The Selfish Gene*. Oxford University Press, 1976.

[29] G. Chaitin, *Proving Darwin: Making Biology Mathematical*. New York: Random House, 2013.

[30] R. M. Gray and J. C. Kieffer, "Asymptotically mean stationary measures," *Ann. Probab.*, vol. 8, pp. 962–973, 1980.

[31] Ł. Dębowski, "Variable-length coding of two-sided asymptotically mean stationary measures," *J. Theor. Probab.*, vol. 23, pp. 237–256, 2010.