Entropy estimate by a randomness criterion

(to appear in Ergodic Theory and Dynamical Systems)

Teturo KAMAE*

Abstract

We propose a new criterion for randomness of a word $x_1x_2\cdots x_n \in \mathbb{A}^n$ over a finite alphabet $\mathbb{A}$ defined by

$$\Xi^n(x_1x_2\cdots x_n) = \sum_{\xi\in\mathbb{A}^+} \psi(|x_1x_2\cdots x_n|_\xi),$$

where $\mathbb{A}^+ = \cup_{k=1}^\infty \mathbb{A}^k$ is the set of nonempty finite words over $\mathbb{A}$, for $\xi \in \mathbb{A}^k$,

$$|x_1x_2\cdots x_n|_\xi = \#\{i;\ 1 \le i \le n-k+1,\ x_ix_{i+1}\cdots x_{i+k-1} = \xi\},$$

and for $t \ge 0$, $\psi(0) = 0$ and $\psi(t) = t\log t$ $(t > 0)$. This value represents how random is the the word $x_1x_2\cdots x_n$ from the viewpoint of the block frequency. In fact, we define a randomness criterion as

$$Q(x_1x_2\cdots x_n) = (1/2)(n\log n)^2/\Xi^n(x_1x_2\cdots x_n).$$

Then,

$$\lim_{n\to\infty}(1/n)Q(X_1X_2\cdots X_n) = h(X)$$

holds with probability 1 if $X_1X_2\cdots$ is an ergodic, stationary process over $\mathbb{A}$ either with a finite energy or $h(X) = 0$, where $h(X)$ is the entropy of the process. Another criterion for randomness was proposed before with t^2 instead of $t\log t$ in [1]. Our new criterion fits better with the entropy than it. We also claim that not only our criterion represents the entropy asymptotically but also it represents well how random fixed finite words are.

1 Introduction

We are interested in finding a nonnegative valued function of finite words over a finite alphabet $\mathbb{A}$ which measures quantitatively how random they are. For example, we want to know

$$\text{Which is more random, 00110 or 00101?} \tag{1.1}$$

*Advanced Mathematical Institute, Osaka City University, 558-8585 Japan (kamae @apost.plala.or.jp)

The Kolmogorov-Chaitin complexity [6] is a criterion of randomness from the algorithmic point of view, which is theoretically perfect, but has two practical shortcomings. First is that it is not a computable function, second is that it has an ambiguity up to adding an arbitrary constant.

Forgetting the arithmetical point of view, we focus on the uniformity of the block frequency in the finite words, which is also an important factor of the randomness. From this point of view, the ideal finite words are equi-distributed ones, that is, $x_1x_2\cdots x_n \in \mathbb{A}^n$ is said to be *equi-distributed* [1] if for any $k = 1, 2, \cdots, n$ and ξ, η in $\mathbb{A}^k$, we have

$$||x_1x_2\cdots x_n|_\xi - |x_1x_2\cdots x_n|_\eta| \le 1,$$

where

$$|x_1x_2\cdots x_n|_\xi = \#\{i;\ 1 \le i \le n-k+1,\ x_ix_{i+1}\cdots x_{i+k-1} = \xi\}.$$

It is still an open problem whether an equi-distributed word of length n exists for any n or not, while we know that it exists for infinitely many n. To evaluate the uniformity of the block frequency by a single value, we take a strictly convex function $\phi : [0,\infty) \to \mathbb{R}$ and calculate $\sum_{\xi\in\mathbb{A}^+} \phi(|x_1x_2\cdots x_n|_\xi)$, where $\mathbb{A}^+ = \sum_{k=1}^\infty \mathbb{A}^k$. Then, this value is smaller if the word has more uniform block frequency, and hence, is more random in some restricted sense among the words with the same length. Assuming that the equi-distributed word of length n exists, this value becomes minimum if and only if $x_1x_2\cdots x_n$ is equi-distributed. For example, in (1.1) with $\mathbb{A} = \{0,1\}$, 00110 is equi-distributed, but 00101 is not, since $|00101|_{01} = 2$ while $|00101|_{11} = 0$. Hence, 00110 is more random than 00101. In fact, for $\phi(t) = t^2$,

$$\sum_{\xi\in\{0,1\}^+} |00110|_\xi^2 = 23 < 25 = \sum_{\xi\in\{0,1\}^+} |00101|_\xi^2.$$

The author with Xue [1], choosing $\phi(t) = t^2$, proposed a criterion

$$\Sigma^n(x_1x_2\cdots x_n) = \sum_{\xi\in\{0,1\}^+} |x_1x_2\cdots x_n|_\xi^2$$

for binary words $x_1x_2\cdots x_n \in \{0,1\}^n$. This criterion has merits that it is easy to calculate and easy to construct infinite words $x_1x_2\cdots \in \{0,1\}^\infty$ attaining the asymptotical minimum value, that is,

$$\lim_{n\to\infty} (1/n^2)\Sigma^n(x_1x_2\cdots x_n) = 3/2,$$

by choosing $x_{n+1} = 0$ or 1 according to $\Sigma^{n+1}(x_1x_2\cdots x_n, 0) \le \Sigma^{n+1}(x_1x_2\cdots x_n 1)$ or not starting from any initial finite word. These words with the asymptotically minimum value, say Σ-*random words*, have probability 1 with respect to the coin-tossing process and are proved to be normal numbers in the sense of E. Borel, but not only that, they satisfy some long range recurrence properties which are not necessarily satisfied by normal numbers. In fact, these properties together characterize the Σ-randomness [1]. The other extreme, the eventual periodicity of $x_1x_2\cdots \in \{0,1\}^\infty$ is also characterized by this value as

$$\lim_{n\to\infty}(1/n^3)\Sigma^n(x_1x_2\cdots x_n) \text{ exists and } > 0 \text{ (Kamae \& Kim [2]).}$$

All these results for $\{0,1\}$ can be easily generalized to a general alphabet $\mathbb{A}$. A shortcoming of this criterion is a lack of relation with the entropy.

In this paper, we propose another criterion taking $\psi(t) = t\log t$ $(t \geq 0)$ where $\psi(0) = 0$ for ϕ and a general alphabet $\mathbb{A}$ with $2 \leq A := \#\mathbb{A} < \infty$. That is,

$$\Xi^n(x_1x_2\cdots x_n) = \sum_{\xi\in\mathbb{A}^+}\psi(|x_1x_2\cdots x_n|_\xi).$$

Let $x = x_1x_2\cdots x_n \in \mathbb{A}^n$. For $k = 1,2,\cdots,n$ and $\xi \in \mathbb{A}^k$, let

$$p_\xi = p_\xi(x) := \frac{|x_1x_2\cdots x_n|_\xi}{n-k+1} \quad \text{and} \quad \mathcal{H}_k(x_1x_2\cdots x_n) = H(p_\xi;\ \xi\in\mathbb{A}^k),$$

where $H(p_1,\cdots k_l) := -\sum_{i=1}^l p_i\log p_i$ is the Shannon's entropy of a probability vector $(p_1,\cdots,p_l)$. Then, it holds that

$$\begin{aligned}
\Xi^n(x_1x_2\cdots x_n) &= \sum_{\xi\in\mathbb{A}^+}\psi(|x_1x_2\cdots x_n|_\xi) = \sum_{k=1}^n\sum_{\xi\in\mathbb{A}^k}\psi((n-k+1)p_\xi)\\
&= \sum_{k=1}^n\sum_{\xi\in\mathbb{A}^k}(n-k+1)p_\xi(\log(n-k+1)+\log p_\xi)\\
&= \sum_{k=1}^n(n-k+1)\left(\log(n-k+1) - \mathcal{H}_k(x_1x_2\cdots x_n)\right). \qquad (1.2)
\end{aligned}$$

It is clear that $\mathcal{H}_k(x_1x_2\cdots x_n) \leq \log(n+k-1)\wedge\log A^k$, Hence, we have

$$\Xi^n(x_1x_2\cdots x_n) \geq \sum_{k=1}^n(n-k+1)\{(\log(n-k+1) - k\log A)\vee 0\}.$$

Therefore, the main part of $\Xi^n(x_1x_2\cdots x_n)$ can be estimated from below by the area with horizontal stripes in Figure 1 multipled by n, that is, $n(\log n)^2(1+o(1))/2\log A$ as $n\to\infty$. Thus, we have

Fact 1. *For any $x_1x_2\cdots \in \mathbb{A}^\infty$, we have*

$$\liminf_{n\to\infty}\frac{\Xi^n(x_1x_2\cdots x_n)}{n(\log n)^2} \geq \frac{1}{2\log A}.$$

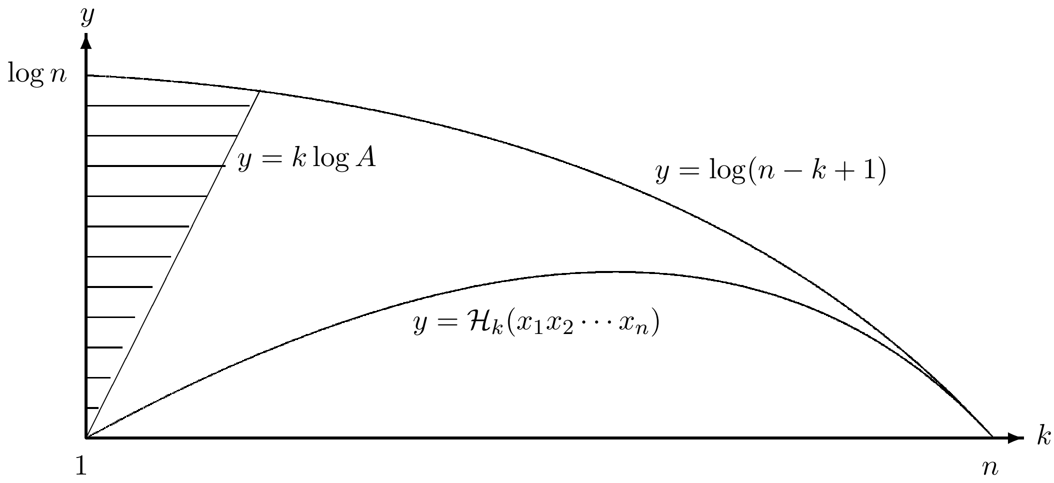

Figure 1: low estimate of $\Xi^n(x_1x_2\cdots x_n)$

Let $\mathcal{S}(\mathbb{A})$ be the set of stationary processes $X = X_1X_2\cdots$ over $\mathbb{A}$. Let

$$h(X) = \lim_{n\to\infty}(1/n)H(X_1X_2\cdots X_n)$$

be the entropy of X and P_X be the distribution of X on $\mathbb{A}^\infty$. Let $\mathcal{E}(\mathbb{A})$ be the subset of $\mathcal{S}(\mathbb{A})$ consisting of all the ergodic processes $X \in \mathcal{S}(\mathbb{A})$. We are interested in consistent estimators T_n $(n = 1, 2, \cdots)$ of $h(X)$ for $X \in \mathcal{E}(\mathbb{A})$ in the sense that $\lim_{n\to\infty} T_n(X_1X_2\cdots X_n) = h(X)$ holds P_X-almost surely.

The following theorem was proved by Shields [5], though our own proof is given to focus on the uniformity of the convergence.

Theorem 1. *Let $X \in \mathcal{E}(\mathbb{A})$ with $h(X) = h$. Take any fixed $\tilde{h}$ with $h < \tilde{h}$. Then, for P_X-almost all $x_1x_2\cdots \in \mathbb{A}^\infty$,*

$$\lim_{n\to\infty}(1/k_n)\mathcal{H}_{k_n}(x_1x_2\cdots x_n) = h$$

holds for any k_n $(n = 1, 2, \cdots)$ such that $c_n \le k_n \le (\log n)/\tilde{h}$ $(n = 1, 2, \cdots)$, where (c_n) is a fixed sequence with $\lim_{n\to\infty} c_n = \infty$. Moreover, the convergence is uniform in $(k_n;\ n = 1, 2, \cdots)$.

Let $X \in \mathcal{E}(\mathbb{A})$ with $h = h(X) > 0$. Consider the graph of $y = \mathcal{H}_k(x_1x_2\cdots x_n)$ as a function of $k = 1, 2, \cdots, n$ for P_X-almost all $x_1x_2\cdots \in \mathbb{A}^\infty$ and a sufficiently large n. By Theorem 1, $\mathcal{H}_k(x_1x_2\cdots x_n)$ increases almost linearly with slope h until $k = (\log n)/(h+\eta)$ with a sufficiently small $\eta > 0$, where it almost touch the graph $y = \log(n+k-1)$ (see Figure 2). It stays near $y = \log(n+k-1)$ within distance $\varepsilon \log n$ until $k = L\log n$ for any fixed $L > 1/h$ and $\varepsilon > 0$ (Lemma 12). Therefore, the main part of

$$\begin{aligned}
&\sum_{k=1}^{L\log n}\sum_{\xi\in\mathbb{A}^k}\psi(|x_1x_2\cdots x_n|_\xi)\\
&= \sum_{k=1}^{L\log n}(n-k+1)\left(\log(n-k+1) - \mathcal{H}_k(x_1x_2\cdots x_n)\right)
\end{aligned}$$

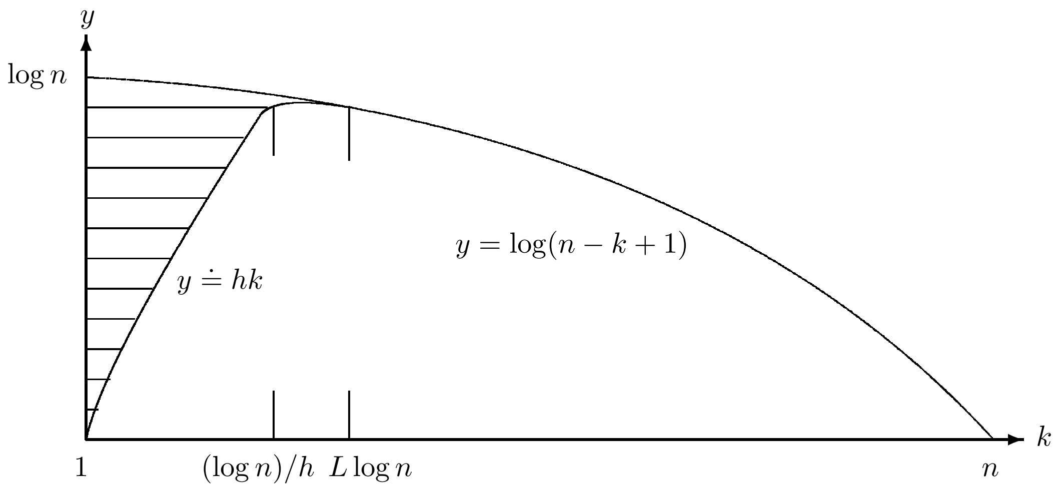

Figure 2: $y = \mathcal{H}_k(x_1 x_2 \cdots, x_n)$ as a function of k

is the area with horizontal stripes in Figure 2 multiplied by n, and hence, $(1/(2h))n(\log n)^2$. Moreover, if X has a finite energy (Definition 1), then

$$\mathcal{H}_k(x_1 x_2 \cdots x_n) = \log(n+k-1)$$

holds P_X-almost surely after $k = L \log n$ for some L (Lemma 13). Hence in this case, the main part of

$$\Xi^n(x_1 x_2 \cdots x_n) = \sum_{k=1}^{n} (n-k+1)\left(\log(n-k+1) - \mathcal{H}_k(x_1 x_2 \cdots x_n)\right)$$

is $(1/(2h))n(\log n)^2$.

Definition 1. (Shields [5]) $X \in \mathcal{E}(\mathbb{A})$ is said to have a *finite energy*, if there exists $0 < C < 1$ and K such that

$$P(X_{t+1} X_{t+2} \cdots X_{t+k} = \xi \mid X_1 X_2 \cdots X_t = \eta) < KC^k \tag{1.3}$$

holds for any nonnegative integers k, t and $\xi \in \mathbb{A}^k, \eta \in \mathbb{A}^t$.

Definition 2. For any $n = 1, 2, \cdots$ and $x_1 x_2 \cdots x_n \in \mathbb{A}^n$, define

$$Q(x_1 x_2 \cdots x_n) = (1/2)(n \log n)^2 / \Xi^n(x_1 x_2 \cdots x_n).$$

Definition 3. We call $x_1 x_2 \cdots \in \mathbb{A}^\infty$ a Ξ-*random word* if

$$\lim_{n\to\infty} (1/n) Q(x_1 x_2 \cdots x_n) = \log A \tag{1.4}$$

holds.

The above statement implies the following Theorem 2 and 3.

Theorem 2. *Let $X \in \mathcal{E}(\mathbb{A})$ with $h(X) = h > 0$. Then, for P_X-almost all $x_1x_2\cdots \in \mathbb{A}^\infty$ and any constant $L > 1/h$, we have*

$$\sum_{1\le k\le L\log n}\ \sum_{\xi\in\mathbb{A}^k} \psi(|x_1x_2\cdots x_n|_\xi) = (1/(2h))n(\log n)^2(1+o(1))$$

as $n \to \infty$.

Theorem 3. (1) *For any $x_1x_2\cdots \in \mathbb{A}^\infty$, we have*

$$\limsup_{n\to\infty}(1/n)Q(x_1x_2\cdots x_n) \le \log A.$$

(2) *Let $X \in \mathcal{E}(\mathbb{A})$. Then,*

$$\limsup_{n\to\infty}(1/n)Q(X_1X_2\cdots X_n) \le h(X)$$

holds P_X-almost surely.
(3) *Let $X \in \mathcal{E}(\mathbb{A})$. Assume that either X has a finite energy or $h(X) = 0$. Then,*

$$\lim_{n\to\infty}(1/n)Q(X_1X_2\cdots X_n) = h(X)$$

holds P_X-almost surely.
(4) *Let $X = X_1X_2\cdots$ be the i.i.d. process over $\mathbb{A}$ with $P(X_1 = a) = 1/A$ $(\forall a \in \mathbb{A})$. Then,*

$$\lim_{n\to\infty}(1/n)Q(X_1X_2\cdots X_n) = \log A$$

holds P_X-almost surely. That is, P_X-almost all $x_1x_2\cdots \in \mathbb{A}^\infty$ are Ξ-random words.

Theorem 4. *For $x = x_1x_2\cdots \in \mathbb{A}^\infty$, assume that there exists $X \in \mathcal{S}(\mathbb{A})$ and a subset $\{t_1 < t_2 < \cdots\} \subset \{1, 2, \cdots\}$ such that*

$$\lim_{n\to\infty}(1/t_n)\sum_{i=1}^{t_n}\delta_{\sigma^i x} = P_X$$

holds, where δ_y is the unit measure at $y \in \mathbb{A}^\infty$, σ is the shift on $\mathbb{A}^\infty$ and the limit implies the w^*-limit. Then, we have

$$\limsup_{n\to\infty}(1/t_n)Q(x_1x_2\cdots x_{t_n}) \le h(X).$$

Corollary 1. *If $x = x_1x_2\cdots \in \mathbb{A}^\infty$ is a Ξ-random word, then x is a normal number over $\mathbb{A}$. On the other hand, there exists a normal number over $\mathbb{A}$ such that*

$$\lim_{n\to\infty}(1/n)Q(x_1x_2\cdots x_n) = 0.$$

We consider the same algorithm to construct random words as in the case of Σ-random words. Starting from any finite word $x_1 \cdots x_m \in \mathbb{A}^m$, construct an infinite word $x_1 \cdots x_m x_{m+1} \cdots \in \mathbb{A}^\infty$ so that

$$\Xi^{n+1}(x_1 \cdots x_n x_{n+1}) = \min_{a \in \mathbb{A}} \Xi^{n+1}(x_1 \cdots x_n a) \tag{1.5}$$

for any $n = m, m+1, \cdots$. Then, we can prove the following theorem.

Theorem 5. *For any $x_1 x_2 \cdots \in \mathbb{A}^\infty$ satisfying* (1.5)*, we have*

$$\liminf_{n \to \infty} (1/n) Q(x_1 x_2 \cdots x_n) \geq (1/2) \log(6/5).$$

Remark 1. Not only the asymptotical property related to the entropy, the function Q behaves well as a function measuring randomness of finite words. In Example 2, we obtain the graph of $Q(x_1 x_2 \cdots x_n)$ with respect to n for the binary random numbers, Rudin-Shapiro sequence, Thue-Morse sequence and Fibonacci sequence. Except for the first one, they have 0-entropy. But, we can discriminate their degree of randomness by Q.

Remark 2. Several consistent estimators of the entropy $h(X)$ of $X \in \mathcal{E}(\mathbb{A})$ are known. We listed 3 of them.

(I) Kolmogorov-Chaitin compexity K. That is, let l be the minimum length of the binary codes outputting $x_1 x_2 \cdots x_n \in \mathbb{A}^n$ by a universal machine. Then, $K(x_1 x_2 \cdots x_n)$ is defined to be $l \log 2$. It is known that $K(x_1 x_2 \cdots x_n)/n$ is a consistent estimtor of the entropy [6].

(II) Lempel-Ziv universal coding [3, 5] also gives a consistent estimator of the entropy. That is,
(1) For $x^0 := x_1 x_2 \cdots x_n \in \mathbb{A}^n$, define $\xi^0 = \epsilon$, where ϵ is the empty word.
(2) For $k = 0, 1, \cdots$, assume that $x^0, x^1, \cdots, x^l \in \mathbb{A}^+ \cup \{\epsilon\}$ and $\xi^0, \xi^1, \cdots, \xi^l \in \mathbb{A}^+ \cup \{\epsilon\}$ are defined. If $x^l = \epsilon$, then define $LZ(x_1 x_2 \cdots x_n) = l$. If $x^l \in \mathbb{A}^+$, then let ξ^i be the longest prefix of x^l among $\xi^0, \xi^1, \cdots, \xi^l$ and let $x^l = \xi^i y$.
(3) If $y = \epsilon$, then define $LZ(x_1 x_2 \cdots x_n) = l + 1$.
(4) If $y \in \mathbb{A}^+$ and $y = ay'$ with $a \in \mathbb{A}$, then let $x^{l+1} = y'$ and $\xi^{l+1} = \xi^i a$. Go to (2) with $l + 1$ for l.
(5) Then, $LZ(x_1 x_2 \cdots x_n) \log n/n$ is a consistent estimtors of the entropy.

(III) Ornstein-Weiss statistic. Ornstein and Weiss [4] define another consisitent estimate of the entropy, say $OW(x_1 x_2 \ldots x_n)$ as follows.
(1) For $n = 1, 2, \cdots$, take a positive integer $k = k(n)$ so that $k / \log_A n \to 1$ as $n \to \infty$. Take a constant C with $0 < C < 1$. Given $x_1 x_2 \cdots x_n \in \mathbb{A}^n$.
(2) Take the minimum l such that there exists $S \subset \mathbb{A}^k$ with the properties that $\#S = l$ and $\sum_{\xi \in S} |x_1 x_2 \cdots x_n|_\xi > Cn$.
(3) Define $OW(x_1 x_2 \ldots x_n) = \log l/k$.

Remark 3. Just like the case of the Σ-randomness, we conjecture that (1.5) implies the Ξ-randomness (see Example 4), though we can only prove that (1.5) implies a positive Q-value per letter (Theorem 5).

Remark 4. One might conjecture that

$$\lim_{n\to\infty}(1/n)Q(X_1X_2\cdots X_n)=0$$

holds P_X-almost surely if $X\in\mathcal{E}(\mathbb{A})$ doesn't have a finite energy. In fact, as is proved in Corollary 1, if P_X admits a long range repetition almost surely, this is true.

2 Proof of Theorem 1

Lemma 1. *Let $\mathbb{B}$ be an alphabet with $2\le\#\mathbb{B}<\infty$. Let $(n_b;\ b\in\mathbb{B})$ be a vector of nonnegative integers such that $\sum_{b\in\mathbb{B}}n_b=n$. Then, we have*

$$\begin{aligned}&\#\{x_1x_2\cdots x_n\in\mathbb{B}^n;\ \#\{1\le i\le n;\ x_i=b\}=n_b \textit{ for any } b\in\mathbb{B}\}\\ &\le\exp\{nH(n_b/n;\ b\in\mathbb{B})+\log(n+1)\}.\end{aligned}$$

Proof Let $\mathbb{B}_0=\{b\in\mathbb{B};\ n_b>0\}$. It holds that

$$\begin{aligned}&\#\{x_1x_2\cdots x_n\in\mathbb{B}^n;\ \#\{1\le i\le n;\ x_i=b\}=n_b \text{ for any } b\in\mathbb{B}\}\\ &=\frac{n!}{\prod_{b\in\mathbb{B}_0}n_b!}\le\exp(\int_1^{n+1}\log x dx-\sum_{b\in\mathbb{B}_0}\int_1^{n_b}\log x dx)\\ &=\exp\{(n+1)\log(n+1)-\sum_{b\in\mathbb{B}_0}n_b\log n_b-\#\mathbb{B}_0\}\\ &=\exp\{\sum_{b\in\mathbb{B}_0}n_b(\log n-\log n_b)+(n+1)\log(n+1)-n\log n-\#\mathbb{B}_0\}\\ &\le\exp\{\sum_{b\in\mathbb{B}_0}n_b(\log n-\log n_b)+\log(n+1)\}\\ &=\exp\{nH(n_b/n;\ b\in\mathbb{B})+\log(n+1)\}.\end{aligned}$$

□

Lemma 2. *Let $\mathbb{B}$ be an alphabet with $2\le B=\#\mathbb{B}<\infty$. Then, for any Δ, we have*

$$\begin{aligned}&\#\{x_1x_2\cdots x_n\in\mathbb{B}^n;\ \mathcal{H}_1(x_1x_2\cdots x_n)\le\Delta\}\\ &\le\exp\{n\Delta+\log(n+1)\}\binom{n+B-1}{B-1}\\ &\le\exp\{n\Delta+2\log(n+B)\\ &\qquad+(n+B-1)H((B-1)/(n+B-1),\ n/(n+B-1))\}.\end{aligned}$$

Proof The number of vectors $(n_b;\ b \in \mathbb{B})$ such that the entries are non-negative integers with $\sum_{b\in\mathbb{B}} n_b = n$ is $\begin{pmatrix} n+B-1 \\ B-1 \end{pmatrix}$. On the other hand, if the vector $(n_\xi;\ \xi \in \mathbb{B})$ satisfies that $H(n_\xi/n;\ \xi \in \mathbb{B}) \le \Delta$, then by Lemma 1,

$$\begin{aligned} &\#\{x_1x_2\cdots x_n \in \mathbb{B}^n;\ \#\{1 \le i \le n;\ x_i = b\} = n_b \text{ for any } b \in \mathbb{B}\} \\ &\le \exp\{n\Delta + \log(n+1)\}. \end{aligned}$$

Hence,

$$\begin{aligned} &\#\{x_1x_2\cdots x_n \in \mathbb{B}^n;\ \mathcal{H}_1(x_1x_2\cdots x_n) \le \Delta\} \\ &\le \exp\{n\Delta + \log(n+1)\}\begin{pmatrix} n+B-1 \\ B-1 \end{pmatrix} \\ &\le \exp\{n\Delta + \log(n+1) + \log(n+B) \\ &\qquad +(n+B-1)H((B-1)/(n+B-1),\ n/(n+B-1))\}. \end{aligned}$$

□

Lemma 3. *Let h and C be positive constants. Let n and k be positive integer valued variables satisfying that $e^{kh} \le Cn$. Let $N = e^{kh},\ M = n/k$. Then, $(N+M)H(N/(N+M),\ M/(N+M)) = o(n)$ holds uniformly in k as $n \to \infty$.*

Proof Let $\lambda_1(n),\ \lambda_2(n)$ be the positive valued functions of n such that

$$\lambda_1(n)^2 e^{\lambda_1(n)h} = n\ ,\ \lambda_2(n)^{1/2} e^{\lambda_2(n)h} = n$$

for any $n = 1, 2, \cdots$. Then, it is easy to see that both $\lambda_1(n)$ and $\lambda_2(n)$ are increasing function of n which tend to the infinity as $n \to \infty$. Also, $\lambda_1(n) \le \lambda_2(n)$ holds. Moreover, $k \le \lambda_1(n)$ implies $M/N \ge \lambda_1(n)$ and $k \ge \lambda_2(n)$ implies $N/M \ge \lambda_2(n)^{1/2}$. Furthermore, $\lambda_1(n) \le k \le \lambda_2(n)$ implies

$$(N+M)/n \le \lambda_1(n)^{-1} + \lambda_2(n)^{-1/2}.$$

Let

$$\begin{aligned} \varepsilon(n) = \max\ \{\ & (C+1)H(1/(\lambda_1(n)+1),\ \lambda_1(n)/(\lambda_1(n)+1)), \\ & (C+1)H(1/(\lambda_2(n)^{1/2}+1),\ \lambda_2(n)^{1/2}/(\lambda_2(n)^{1/2}+1)), \\ & (\lambda_1(n)^{-1} + \lambda_2(n)^{-1/2})\log 2\ \ \}. \end{aligned}$$

Then,

$$(N+M)H(N/(N+M),\ M/(N+M)) \le \varepsilon(n)n$$

and $\varepsilon(n) \to 0$ hold as $n \to \infty$. Thus,

$$(N+M)H(N/(N+M),\ M/(N+M)) = o(n)$$

holds uniformly in k as $n \to \infty$. □

Lemma 4. *For any k, assume that $A^k \leq n$. Then for any Δ, we have*

$$\#\{x_1x_2\cdots x_n \in \mathbb{A}^n;\ \mathcal{H}_k(x_1x_2\cdots x_n) \leq k\Delta\} \leq \exp\{n(\Delta + o(1))\}$$

as $n \to \infty$, where this $o(1)$ is uniform in k and Δ.

Proof For $j = 1, 2, \cdots, k$, let

$$y^j = (x_j \cdots x_{j+k-1})(x_{j+k} \cdots x_{j+2k-1}) \cdots (x_{j+(l_j-1)k} \cdots x_{j+l_jk-1}) \in (\mathbb{A}^k)^{l_j},$$

where $l_j = \lfloor (n-j+1)/k \rfloor$. For any $\xi \in \mathbb{A}^k$, it holds that

$$|x_1x_2\cdots x_n|_\xi = \sum_{j=1}^k |y^j|_\xi,$$

and hence,

$$p_\xi(x_1x_2\cdots x_n) = \sum_{j=1}^k (l_j/(n-k+1))p_\xi(y^j).$$

This implies that

$$\mathcal{H}_k(x_1x_2\cdots x_n) \geq \sum_{j=1}^k (l_j/(n-k+1))\mathcal{H}_1(y^j).$$

Hence, there exists j such that $\mathcal{H}_1(y^j) \leq \mathcal{H}_k(x_1x_2\cdots x_n)$. Then, we have

$$\begin{aligned}
&\#\{x_1x_2\cdots x_n \in \mathbb{A}^n;\ \mathcal{H}_k(x_1x_2\cdots x_n) \leq k\Delta\} \\
&\leq \#\{x_1x_2\cdots x_n \in \mathbb{A}^n;\ \text{there exists } j \text{ such that } \mathcal{H}_1(y^j) \leq k\Delta\} \\
&\leq kA^{2(k-1)}\#\{y \in (\mathbb{A}^k)^{l-1} \cup (\mathbb{A}^k)^l;\ \mathcal{H}_1(y) \leq k\Delta\},
\end{aligned}$$

where $l = \lfloor n/k \rfloor$. Let $\mathbb{B} = \mathbb{A}^k$ and $B = A^k$. Applying Lemma 2, we have

$$\begin{aligned}
&\#\{x_1x_2\cdots x_n \in \mathbb{A}^n;\ \mathcal{H}_k(x_1x_2\cdots x_n) \leq k\Delta\} \\
&\leq 2kA^{2(k-1)}\exp\{lk\Delta \\
&\qquad +(l+B-1)H((B-1)/(l+B-1),\ l/(l+B-1)) + 2\log(l+B)\} \\
&\leq \exp\{n(\Delta + D) + (l+B-1)H((B-1)/(l+B-1),\ l/(l+B-1))\} \\
&= \exp\{n(\Delta + o(1))\}
\end{aligned}$$

where

$$D = (1/n)(\log(2kA^{2(k-1)}) + 2\log(l+B)) \leq 5(\log n)/n$$

holds for large n, and $(l+B-1)H((B-1)/(l+B-1),\ l/(l+B-1)) = o(n)$ holds uniformly in k and Δ by Lemma 3. Therefore, the above $o(1)$ is uniform in k and Δ, which completes the proof. $\square$

We can improve Lemma 4 as the following Lemma 6 using Lemma 5.

Lemma 5. (Ornstein and Weiss [4]) *Let $X \in \mathcal{E}(\mathbb{A})$ with $h(X) = h$. Then, for P_X-almost all $x_1x_2\cdots \in \mathbb{A}^\infty$ and any $\varepsilon > 0$, if k and n/k are sufficiently large, then there exists $\mathbb{B}_k \subset \mathbb{A}^k$ with $\#\mathbb{B}_k < e^{(h+\varepsilon)k}$, such that*

$$\begin{aligned}&\#\{1 \le i \le n-k+1;\ i \equiv j \ (mod\ k)\ and\ x_ix_{i+1}\cdots x_{i+k-1} \in \mathbb{B}_k\}\\&> (1-\varepsilon)n/k \quad (j = 1, 2, \cdots, k).\end{aligned} \tag{2.1}$$

Moreover, this $\mathbb{B}_k$ can be chosen independently of n and $x_1x_2\cdots \in \mathbb{A}^\infty$.

Lemma 6. *Let $X \in \mathcal{E}(\mathbb{A})$ with $h(X) = h$. Take any fixed $\tilde{h} > h$. Then, for any $\varepsilon > 0$, there exist n_0 and $\Omega \subset \mathbb{A}^\infty$ with $P_X(\Omega) > 1-\varepsilon$ such that for any $n \ge n_0$, k with $e^{\tilde{h}k} \le n$ and Δ,*

$$\#\{x_1x_2\cdots x_n \in \Omega_n;\ \mathcal{H}_k(x_1x_2\cdots x_n) \le k\Delta\} \le \exp\{n(\Delta+\varepsilon)\}$$

holds, where $\Omega_n \subset \mathbb{A}^n$ is the projection of Ω to $\mathbb{A}^n$.

Proof Since Lemma 4 implies Lemma 6 in the case $A^k \le n$, it is sufficint to prove Lemma 6 with the assumption that

$$(1/\log A)\log n \le k \le (1/\tilde{h})\log n. \tag{2.2}$$

Take any small $\varepsilon > 0$ such that $h+\varepsilon < \tilde{h}$. There exist $\Omega \subset \mathbb{A}^\infty$ with $P_X(\Omega) > 1-\varepsilon$ and n_0 such that for any $n \ge n_0$ and k satisfying (2.2), (2.1) holds for any $x_1x_2\cdots x_n \in \Omega_n$. Let $B_k = \#\mathbb{B}_k$. Then, $B_k < e^{\tilde{h}k}$. By the same argument as in the proof of Lemma 4,

$$\begin{aligned}&\#\{x_1x_2\cdots x_n \in \mathbb{A}^n;\ \mathcal{H}_k(x_1x_2\cdots x_n) \le k\Delta\}\\&\le kA^{2(k-1)}\#\{y_1y_2\cdots y_{l'} \in (\mathbb{A}^k)^{l-1} \cup (\mathbb{A}^k)^l;\ \mathcal{H}_1(y_1y_2\cdots y_{l'}) \le k\Delta\}\end{aligned}$$

holds, where $l = \lfloor n/k \rfloor$ and

$$y_i = x_{j+(i-1)k}x_{j+(i-1)k+1}\cdots x_{j+ik-1} \in \mathbb{A}^k \ (i = 1, 2, \cdots, l';\ l' = l-1 \text{ or } l)$$

for some $j = 1, 2, \cdots, k$. By (2.1), at least $(1-\varepsilon)n/k$ number of i satisfy that $y_i \in \mathbb{B}_k$ for any $y_1y_2\cdots y_{l'} \in \Lambda$, where Λ is the set of $y_1y_2\cdots y_{l'}$ coming from $x_1x_2\cdots x_n \in \Omega_n$ as above. Let the set of i such that $y_i \in \mathbb{B}_k$ be $\mathbb{I}$ and $\#\mathbb{I} = I$. By taking some subset of $\mathbb{I}$ if necessary, we may assume that exactly $I = \lfloor (1-\varepsilon)l \rfloor$ holds. Then,

$$\begin{aligned}&\mathcal{H}_1(y_1y_2\cdots y_{l'})\\&\ge (I/l')\mathcal{H}_1(y_i;\ i \in \mathbb{I}) + ((l'-I)/l')\mathcal{H}_1(y_i;\ i \notin \mathbb{I})\\&\ge (1-\varepsilon)\mathcal{H}_1(y_i;\ i \in \mathbb{I}),\end{aligned}$$

and hence,

$$\begin{aligned}&\#\{y_1y_2\cdots y_{l'} \in \Lambda;\ \mathcal{H}_1(y_1y_2\cdots y_{l'}) \le k\Delta\}\\&\le \binom{l}{l-I}(A^k)^{l-I}\#\{(y_i;\ i \in \mathbb{I}_0) \in \mathbb{B}_k^{\mathbb{I}_0};\ \mathcal{H}_1(y_i;\ i \in \mathbb{I}_0) \le (1-\varepsilon)^{-1}k\Delta\}\end{aligned}$$

with $\mathbb{I}_0 = \{1, 2, \cdots, I\}$. Applying Lemma 2, we have

$$
\begin{aligned}
&\#\{x_1x_2\cdots x_n \in \Omega_n;\ \mathcal{H}_k(x_1x_2\cdots x_n) \le k\Delta\} \\
&\le 2kA^{2(k-1)}\#\{y_1y_2\cdots y_{l'} \in \Lambda;\ \mathcal{H}_1(y_1y_2\cdots y_{l'}) \le k\Delta\} \quad (l' = l-1 \text{ or } l) \\
&\le 2kA^{2(k-1)}\binom{l}{l-I}(A^k)^{l-I}\#\{(y_i;\ i \in \mathbb{I}_0) \in \mathbb{B}_k^{\mathbb{I}_0};\ \mathcal{H}_1(y_i;\ i \in \mathbb{I}_0) \le 2k\Delta(1-\varepsilon)^{-1}\} \\
&\le 2kA^{2(k-1)}\binom{l}{l-I}(A^k)^{l-I}\exp\{Ik\Delta(1-\varepsilon)^{-1} + \log(I+1)\}\binom{I+B_k-1}{B_k-1} \\
&\le \exp\{n(\Delta + \varepsilon\log A + o(1)) \\
&\qquad\qquad +(I+B_k-1)H((B_k-1)/(I+B_k-1),\ I/(I+B_k-1))\} \\
&\le \exp\{n(\Delta + \varepsilon\log A + o(1)) + (N+M)H(N/(N+M),\ M/(N+M))
\end{aligned}
$$

as $n \to \infty$, where the above $o(1)$ is uniform in $k,\ \Delta$, and $N = e^{k\tilde{h}},\ M = n/k$. By (2.2), $\varepsilon > 0$ can be chosen arbitraray small if just $n \to \infty$. By Lemma 3, $(N+M)H(N/(N+M),\ M/(N+M)) = o(n)$ with $o(n)$ which is uniform in $k,\ \Delta$. Thus, the last line in the above inequalities can be written as $\exp\{n(\Delta + o(1))$ with $o(1)$ which is uniform in k and Δ, which completes the proof. $\square$

Lemma 7. *Let $X \in \mathcal{E}(\mathbb{A})$ with $h(X) = h$. Take any fixed $\tilde{h}$ with $h < \tilde{h}$. Then, for P_X-almost all $x_1x_2\cdots \in \mathbb{A}^\infty$,*

$$\liminf_{n\to\infty}(1/k_n)\mathcal{H}_{k_n}(x_1x_2\cdots x_n) \ge h$$

holds for any k_n $(n = 1, 2, \cdots)$ with $k_n \le (\log n)/\tilde{h}$. Moreover, the convergence is uniform in $(k_n;\ n = 1, 2, \cdots)$.

Proof For any $\varepsilon > 0$, by Lemma 6, there exist n_0 and $\Omega \subset \mathbb{A}^\infty$ with $P_X(\Omega) > 1 - \varepsilon$ such that

$$\#\{x_1x_2\cdots x_n \in \Omega_n;\ \mathcal{H}_k(x_1x_2\cdots x_n) \le k\Delta\} \le \exp\{n(\Delta + \varepsilon)\}$$

for any $n \ge n_0,\ k \le (\log n)/\tilde{h}$ and Δ. It follows from the Shannon-McMillan-Breiman theorem ([5]) that for any $\varepsilon > 0$, there exist n_1 and $\Lambda \subset \mathbb{A}^\infty$ with $P_X(\Lambda) > 1 - \varepsilon$ such that for any $n \ge n_1$ and $x_1x_2\cdots \in \Lambda$, $P_X(\langle x_1x_2\cdots x_n\rangle) < e^{n(-h+\varepsilon)}$ holds, where $\langle\cdot\rangle$ implies the cylinder set corresponding to a finite word or a set of finite words. Therefore, if $\Delta = h - 3\varepsilon$, then $P_X(\langle S_{n,k}\rangle \cap \Lambda) < e^{-n\varepsilon}$ holds for any sufficiently large $n \ge n_0$, where

$$S_{n,k} = \{x_1x_2\cdots x_n \in \Omega_n;\ \mathcal{H}_k(x_1x_2\cdots x_n) \le k\Delta\}.$$

Let

$$S_n = \bigcup_{k;\ k \le (\log n)/\tilde{h}} S_{n,k}.$$

Then, $P_X(S_n \cap \Lambda) < e^{-n\varepsilon}(\log n)/\tilde{h}$ and $\sum_{n=1}^{\infty} e^{-n\varepsilon}(\log n)/\tilde{h} < \infty$, we have

$$
\begin{aligned}
0 &= P_X(\cap_{m=1}^{\infty} \cup_{n=m}^{\infty} S_n \cap \Lambda) \\
&= P_X(\{x_1x_2\cdots \in \Omega \cap \Lambda;\ (1/k)\mathcal{H}_k(x_1x_2\cdots x_n) \le h - 3\varepsilon \text{ holds} \\
&\qquad\qquad \text{for some } k \text{ with } k \le (\log n)/\tilde{h} \text{ for infinitely many } n\} \\
&= P_X(x_1x_2\cdots \in \Omega \cap \Lambda;\ \liminf_{n\to\infty}(1/k_n)\mathcal{H}_{k_n}(x_1x_2\cdots x_n) < h - 3\varepsilon \\
&\qquad\qquad \text{holds for some } k_n\ (n = 1, 2, \cdots) \text{ with } k_n \le (\log n)/\tilde{h}\}.
\end{aligned}
$$

Since $P_X(\Omega) = 1 - o(1)$, $P_X(\Lambda) > 1 - \varepsilon$ and $\varepsilon > 0$ is arbitrary,

$$\liminf_{n\to\infty}(1/k_n)\mathcal{H}_{k_n}(x_1x_2\cdots x_n) \ge h$$

holds for any k_n $(n = 1, 2, \cdots)$ with $k_n \le (\log n)/\tilde{h}$ with probability 1. Moreover, it is clear from the above argument that this convergence is uniform in $(k_n;\ n = 1, 2, \cdots)$. $\square$

Lemma 8. *Let $X \in \mathcal{E}(\mathbb{A})$ with $h(X) = h$. Then, for P_X-almost all $x_1x_2\cdots \in \mathbb{A}^{\infty}$,*

$$\limsup_{n\to\infty}(1/k_n)\mathcal{H}_{k_n}(X_1X_2\cdots X_n) \le h$$

holds for any k_n $(n = 1, 2, \cdots)$ such that $c_n \le k_n \le n/c_n$, where (c_n) is a fixed sequence with $\lim_{n\to\infty} c_n = \infty$. Moreover, the convergence is uniform in $(k_n;\ n = 1, 2, \cdots)$.

Proof By Lemma 5, for any P_X-almost all $x_1x_2\cdots$ and $\varepsilon > 0$,

$$\mathcal{H}_{k_n}(x_1x_2\cdots x_n) \le H(\varepsilon, 1-\varepsilon) + \varepsilon k_n \log A + (1-\varepsilon)k_n(h+\varepsilon).$$

holds if k_n and n/k_n are sufficiently large, and hence, if $c_n \le k_n \le n/k_n$ and n is sufficiently large, which proves our claim. $\square$

Lemma 7 and 8 imply Theorem 1.

3 Proof of Theorem 2 and 3

Definition 4. *Let $x = x_1x_2\cdots x_n \in \mathbb{A}^n$. For $k = 1, 2, \cdots$ and $\xi \in \mathbb{A}^k$ let $\tilde{p_\xi} = \tilde{p_\xi}(x) := (1/n)|x_1x_2\cdots x_nx_{n+1}\cdots x_{k-1}|_\xi$, where we define $x_i := x_j$ if $i \equiv j \pmod n$ and $j \in \{1, 2, \cdots, n\}$. Define*

$$\tilde{\mathcal{H}}_k(x_1x_2\cdots x_n) = H(\tilde{p}_\xi;\ \xi \in \mathbb{A}^k).$$

Lemma 9. *For any $x = x_1x_2\cdots x_n \in \mathbb{A}^n$ and $k, l \ge 1$, we have*

$$\tilde{\mathcal{H}}_k(x_1x_2\cdots x_n) \le \tilde{\mathcal{H}}_{k+l}(x_1x_2\cdots x_n) \le \tilde{\mathcal{H}}_k(x_1x_2\cdots x_n) + \tilde{\mathcal{H}}_l(x_1x_2\cdots x_n).$$

Proof Since $\tilde{p}_\xi = \sum_{\zeta \in \mathbb{A}^l} \tilde{p}_{\xi\zeta}$ and $\tilde{p}_\eta = \sum_{\zeta \in \mathbb{A}^k} \tilde{p}_{\zeta\eta}$ hold for any $\xi \in \mathbb{A}^k$ and $\eta \in \mathbb{A}^l$, we have

$$\tilde{\mathcal{H}}_k(x_1x_2\cdots x_n) \leq \tilde{\mathcal{H}}_{k+l}(x_1x_2\cdots x_n)$$

and

$$\begin{aligned}
&\tilde{\mathcal{H}}_{k+l}(x_1x_2\cdots x_n) = H(\tilde{p}_{\xi\eta};\ \xi \in \mathbb{A}^k,\ \eta \in \mathbb{A}^l)\\
&\leq H(\tilde{p}_\xi \times \tilde{p}_\eta;\ \xi \in \mathbb{A}^k,\ \eta \in \mathbb{A}^l) = H(\tilde{p}_\xi;\ \xi \in \mathbb{A}^k) + H(\tilde{p}_\eta;\ \eta \in \mathbb{A}^l)\\
&= \tilde{\mathcal{H}}_k(x_1x_2\cdots x_n) + \tilde{\mathcal{H}}_l(x_1x_2\cdots x_n).
\end{aligned}$$

□

Lemma 10. *For any $x = x_1x_2\cdots x_n \in \mathbb{A}^n$ and $k = 1, 2, \cdots, n$, we have*

$$\begin{aligned}
&\frac{n-k+1}{n}\mathcal{H}_k \leq \tilde{\mathcal{H}}_k\\
&\qquad \leq \frac{n-k+1}{n}\mathcal{H}_k + \frac{k-1}{n}\log A + H\left(\frac{n-k+1}{n},\ \frac{k-1}{n}\right),
\end{aligned}$$

where we denote $\mathcal{H}_k = \mathcal{H}_k(x_1x_2\cdots x_n)$ and $\tilde{\mathcal{H}}_k = \tilde{\mathcal{H}}_k(x_1x_2\cdots x_n)$.

Proof Let $\{1, 2, \cdots, n\}$ be a probability space with the normalized counting measure and $X,\ Y$ be random variables on it defined by

$$X(i) = x_i x_{i+1}\cdots x_{i+k-1} \in \mathbb{A}^k$$

and

$$Y(i) = \begin{cases} 1 & i \leq n-k+1 \\ 0 & i \geq n-k \end{cases}.$$

Using the standard notations of Shannon's entropy, we have

$$\tilde{\mathcal{H}}_k = H(X)\ ,\ \mathcal{H}_k = H(X|Y=1).$$

Hence,

$$\begin{aligned}
&\tilde{\mathcal{H}}_k = H(X) \leq H(X,Y) = H(Y) + H(X|Y)\\
&= H\left(\frac{n-k+1}{n},\ \frac{k-1}{n}\right) + P(Y=0)H(X|Y=0) + P(Y=1)H(X|Y=1)\\
&\leq H\left(\frac{n-k+1}{n},\ \frac{k-1}{n}\right) + \frac{k-1}{n}\log A + \frac{n-k+1}{n}\mathcal{H}_k.
\end{aligned}$$

On the other hand,

$$\begin{aligned}
&\tilde{\mathcal{H}}_k = H(X) \geq H(X|Y)\\
&= P(Y=0)H(X|Y=0) + P(Y=1)H(X|Y=1)\\
&\geq P(Y=1)H(X|Y=1) = \frac{n-k+1}{n}\mathcal{H}_k.
\end{aligned}$$

□

Lemma 11. *Let L be any fixed number. Then, for any $\varepsilon > 0$, there exists n_0 such that for any $n \geq n_0$, $x_1x_2\cdots x_n \in \mathbb{A}^n$ and k with $k \leq L\log n$, we have $|\mathcal{H}_k - \tilde{\mathcal{H}}_k| \leq \varepsilon$, where we denote $\mathcal{H}_k = \mathcal{H}_k(x_1x_2\cdots x_n)$ and $\tilde{\mathcal{H}}_k = \tilde{\mathcal{H}}_k(x_1x_2\cdots x_n)$.*

Proof By Lemma 10, we have

$$|\mathcal{H}_k - \tilde{\mathcal{H}}_k| \leq \frac{k-1}{n}(\mathcal{H}_k + \log A) + H\left(\frac{n-k+1}{n}, \frac{k-1}{n}\right).$$

Since $k \leq L\log n$ and $\mathcal{H}_k \leq \log n$, the righthand side tends to 0 as $n \to \infty$, which completes the proof. □

Lemma 12. *Let $X \in \mathcal{E}(\mathbb{A})$ with $h(X) = h > 0$. Let $L > 1/h$ be any fixed number. For any $\varepsilon > 0$ and P_X-almost all $x_1x_2\cdots \in \mathbb{A}^\infty$, there exists $\eta > 0$ and n_0 such that for any $n \geq n_0$ and k with $(\log n)/(h+\eta) \leq k \leq L\log n$,*

$$(1-\varepsilon)\log(n-k+1) < \mathcal{H}_k(x_1x_2\cdots x_n) \leq \log(n-k+1).$$

Proof It is clear that $\mathcal{H}_k(x_1x_2\cdots x_n) \leq \log(n-k+1)$. For a small $\eta > 0$ which is specified later, let k_0 be the intger-valued function of n such that $(\log n)/(h+\eta) - 1 < k_0 \leq (\log n)/(h+\eta)$. For P_X-almost all $x_1x_2\cdots \in \mathbb{A}^\infty$, by Theorem 1, there exists n_1 such that for any $n \geq n_1$,

$$\mathcal{H}_{k_0}(x_1x_2\cdots x_n) \geq (h-\eta)k_0 \geq \frac{h-\eta}{h+\eta}\log n - h + \eta.$$

By Lemma 11 with η for ε and n_2 for n_0, we have

$$|\mathcal{H}_k(x_1x_2\cdots x_n) - \tilde{\mathcal{H}}_k(x_1x_2\cdots x_n)| \leq \eta$$

for any $n \geq n_2$ and k with $k \leq L\log n$. Then, for any $n \geq n_1 \vee n_2$ and k with $(\log n)/(h+\eta) \leq k \leq L\log n$, we have

$$\begin{aligned}&\mathcal{H}_k(x_1x_2\cdots x_n) \geq \tilde{\mathcal{H}}_k(x_1x_2\cdots x_n) - \eta\\ &\geq \tilde{\mathcal{H}}_{k_0}(x_1x_2\cdots x_n) - \eta \geq \frac{h-\eta}{h+\eta}\log n - h.\end{aligned}$$

Hence, for any $\varepsilon > 0$, we can choose $\eta > 0$ and $n_0 \geq n_1 \vee n_2$ so that

$$\begin{aligned}&\mathcal{H}_k(x_1x_2\cdots x_n) \geq \frac{h-\eta}{h+\eta}\log n - h\\ &\geq (1-\varepsilon)\log n \geq (1-\varepsilon)\log(n-k+1)\end{aligned}$$

for any $n \geq n_0$ and k with $(\log n)/(h+\eta) \leq k \leq L\log n$. □

Proof of Theorem 2:

Let $X \in \mathcal{E}(\mathbb{A})$ with $h(X) = h > 0$. Let L be any constants with $L > 1/h$. By Lemma 12, for any $\varepsilon > 0$ and P_X-almost all $x_1x_2\cdots \in \mathbb{A}^\infty$, there exists

$\eta > 0$ and n_0 such that for any $n \geq n_0$ and k with $(\log n)/(h+\eta) \leq k \leq L \log n$,

$$(1-\varepsilon)\log(n-k+1) < \mathcal{H}_k(x_1 x_2 \cdots x_n) \leq \log(n-k+1).$$

By Theorem 1, P_X-almost all $x_1 x_2 \cdots \in \mathbb{A}^\infty$ and $k \leq (\log n)/(h+\eta)$,

$$\mathcal{H}_k(x_1 x_2 \cdots x_n) = kh(1+o(1))$$

holds as $n \to \infty$ with $o(1)$ which is uniform in k. Therefore,

$$\begin{aligned}
&\sum_{k \leq L(\log n)} \sum_{\xi \in \mathbb{A}^k} \psi(|x_1 x_2 \cdots x_n|_\xi) \\
&= \sum_{k \leq L(\log n)} (n-k+1)(\log(n-k+1) - \mathcal{H}_k(x_1 x_2 \cdots x_n)) \\
&= n(1+o(1)) \sum_{k \leq L(\log n)} (\log(n-k+1) - \mathcal{H}_k(x_1 x_2 \cdots x_n))
\end{aligned}$$

$$\begin{aligned}
&= n(1+o(1)) \left(\sum_{k \leq (\log n)/\tilde{h}} (\log(n-k+1) - \mathcal{H}_k(x_1 x_2 \cdots x_n)) + \varepsilon D \right) \\
&= n \left(\frac{(\log n)^2}{\tilde{h}} - \frac{(\log n)^2 h}{2\tilde{h}^2} + \varepsilon D \right)(1+o(1))
\end{aligned}$$

as $n \to \infty$, where $\tilde{h} = h + \eta$ and $|D| \leq L(\log n)^2$. Since we can let $\eta \to 0$, we have

$$\sum_{h \leq L(\log n)} \sum_{\xi \in \mathbb{A}^k} \psi(|x_1 x_2 \cdots x_n|_\xi) = (1/(2h)) n (\log n)^2 (1+o(1)) + \varepsilon n D.$$

Since $\varepsilon \to 0$ can be chosen arbitrary small if just $n \to \infty$, we have $\varepsilon n D = o(n(\log n)^2)$, which completes the proof. $\square$

Lemma 13. *Let $X \in \mathcal{E}(\mathbb{A})$. Assume that X has a finite energy. Then, there exists L such that for P_X-almost all $x_1 x_2 \cdots \in \mathbb{A}^\infty$, there exists n_0 such that $\mathcal{H}_k(x_1 x_2 \cdots x_n) = \log(n+k-1)$ holds for any $n \geq n_0$ and $k \geq L \log n$.*

Proof Let (1.3) hold for our X. For a positive integer l with $0 < l < k$, let

$$\mathbb{A}_l^k = \{\xi \in \mathbb{A}^k;\ \text{there exists } \eta \in \mathbb{A}^l \text{ such that } \xi \text{ is a prefix of } \eta^\infty\}.$$

Then, $\#\mathbb{A}_l^k = A^l$ and

$$P(|X_1 X_2 \cdots X_n|_\xi \geq 1 \text{ for some } \xi \in \cup_{i=1}^{l-1} \mathbb{A}_i^k) \leq n A^l C^k$$

holds. On the other hand, let $U = \mathbb{A}^k \setminus \cup_{i=1}^{l-1} \mathbb{A}_i^k$. Then, we have

$$
\begin{aligned}
&P(|X_1X_2\cdots X_n|_\xi \geq 2 \text{ for some } \xi \in U) \\
&\leq \sum_{\substack{1\leq i<j\leq n-k+1 \\ j-i\geq l}} \sum_{\xi\in U} P(X_jX_{j+1}\cdots X_{j+k-1} = X_iX_{i+1}\cdots X_{i+k-1} = \xi) \\
&\leq \sum_{\substack{1\leq i<j\leq n-k+1 \\ j-i\geq l}} \sum_{\xi\in U} P(X_jX_{j+1}\cdots X_{j+k-1} = \xi \mid X_iX_{i+1}\cdots X_{i+k-1} = \xi) \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \times P(X_iX_{i+1}\cdots X_{i+k-1} = \xi) \\
&\leq \sum_{1\leq i<j\leq n-k+1} \sum_{\xi\in\mathbb{A}^k} KC^l P(X_iX_{i+1}\cdots X_{i+k-1} = \xi) \leq Kn^2C^l.
\end{aligned}
$$

Let $\beta = \dfrac{\log(1/C)}{\log A + \log(1/C)}$ and choose a positive integer l so that $|l - \beta k| < 1$. Then, $0 < \beta < 1$ and $A^{\beta k}C^k = C^{\beta k}$ hold. Hence, $A^lC^k < AC^{\beta k}$ holds. Moreover, since $Kn^2C^l < C^{\beta k}Kn^2C$, we have

$$P(|X_1X_2\cdots X_n|_\xi \geq 2 \text{ for some } \xi \in \mathbb{A}^k) \leq C^{\beta k}(A + Kn^2C) \leq \hat{K}n^2\hat{C}^k$$

with some constants $\hat{K}$ and $0 < \hat{C} < 1$. Hence,

$$\sum_{n=1}^{\infty} \sum_{L\log n\leq k\leq n} \hat{K}n^2\hat{C}^k \leq \sum_{n=1}^{\infty} \hat{K}n^3\hat{C}^{L\log n} = \hat{K}\sum_{n=1}^{\infty} n^{L\log\hat{C}+3} < \infty$$

holds if $L\log\hat{C} < -4$. This implies our Lemma. □

Proof of Theorem 3:

(1) follows from Fact 1.

(2) follows from Theorem 2 since $\sum_{k>L\log n}\sum_{\xi\in\mathbb{A}^k}\psi(|x_1x_2\cdots x_n|_\xi) \geq 0$.

(3) By Lemma 13, the equality holds in the case that $h(X) > 0$ and X has a finite energy. Let $X \in \mathcal{E}(\mathbb{A})$ with $h(X) = 0$. Then by Lemma 8, for P_X-almost all $x_1x_2\cdots \in \mathbb{A}^\infty$ and $\varepsilon > 0$, there exists K such that for any $k,\ n$ with $K \leq k \leq n/K$, we have

$$\mathcal{H}_k(x_1x_2\cdots x_n) \leq \varepsilon k.$$

Hence, if n is sufficiently large, we have

$$
\begin{aligned}
&\Xi^n(x_1x_2\cdots x_n) = \sum_{k=1}^{n}(n-k+1)(\log(n-k+1) - \mathcal{H}_k(x_1x_2\cdots x_n)) \\
&\geq \sum_{K\leq k<(\log n)/(2\varepsilon)} (n-k+1)(\log(n-k+1) - \varepsilon k) \vee 0 \\
&\geq \sum_{K\leq k<(\log n)/(2\varepsilon)} (1/3)n\log n \geq n(\log n)^2/(8\varepsilon).
\end{aligned}
$$

Thus,

$$\limsup_{n\to\infty}(1/n)Q(x_1x_2\cdots x_n) \le 4\varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, we have $\lim_{n\to\infty}(1/n)Q(x_1x_2\cdots x_n) = 0$.

(4) follows from (3) since the i.i.d. process over $\mathbb{A}$ has a finite energy. □

4 Proof of Theorem 4

For $x = x_1x_2\cdots \in \mathbb{A}^\infty$, assume that there exists $X \in \mathcal{S}(\mathbb{A})$ and a subset $\{t_1 < t_2 < \cdots\} \subset \{1, 2, \cdots\}$ such that

$$\lim_{n\to\infty}(1/t_n)\sum_{i=1}^{t_n}\delta_{\sigma^i x} = P_X.$$

Then, it holds that

$$\lim_{n\to\infty}\mathcal{H}_k(x_1x_2\cdots x_{t_n}) = H(X_1X_2\cdots X_k).$$

For any $\varepsilon > 0$, take k_0 such that $(1/k)H(X_1X_2\cdots X_k) < h(X) + \varepsilon$ for any $k \ge k_0$. Take $k_1 \ge k_0$ and n_0 such that

$$\mathcal{H}_{k_1}(x_1x_2\cdots x_{t_n}) < H(X_1X_2\cdots X_{k_1}) + \varepsilon < k_1h(X) + (k_1+1)\varepsilon$$

for any $n \ge n_0$. Then for any $n \ge n_0$ and k, by Lemma 9 and 10, we have

$$\begin{aligned}
&\mathcal{H}_k(x_1x_2\cdots x_{t_n}) \le \tilde{\mathcal{H}}_k(x_1x_2\cdots x_{t_n}) \le \lceil k/k_1\rceil\tilde{\mathcal{H}}_{k_1}(x_1x_2\cdots x_{t_n})\\
&\le \lceil k/k_1\rceil(\mathcal{H}_{k_1}(x_1x_2\cdots x_{t_n}) + (k_1/t_n)\log A + \log 2)\\
&\le \lceil k/k_1\rceil(k_1h(X) + (k_1+1)\varepsilon + (k_1/t_n)\log A + \log 2).
\end{aligned}$$

Hence, for any $\varepsilon > 0$, there exists K such that

$$\mathcal{H}_k(x_1x_2\cdots x_{t_n}) < k(h(X) + \varepsilon)$$

holds for any k, n with $K \le k \le t_n/K$. Then, for any sufficiently large n, we have

$$\begin{aligned}
&\Xi^{t_n}(x_1x_2\cdots x_{t_n}) = \sum_{k=1}^{t_n}(t_n - k + 1)(\log(t_n - k + 1) - \mathcal{H}_k(x_1x_2\cdots x_{t_n}))\\
&\ge \sum_{K\le k\le(\log t_n)/(h(X)+2\varepsilon)}(t_n - k + 1)(\log(t_n - k + 1) - k(h(X) + \varepsilon))\\
&\ge (1/(2h(X) + 4\varepsilon))t_n(\log t_n)^2.
\end{aligned}$$

By Definition 2,

$$\limsup_{n\to\infty}(1/t_n)Q(x_1x_2\cdots x_{t_n}) \le h(X) + 2\varepsilon,$$

which completes the proof since $\varepsilon > 0$ is arbitrary $\square$

Proof of Corollary 1:

Assume that $x = x_1x_2\cdots \in \mathbb{A}^\infty$ is not a normal number over $\mathbb{A}$. Then, there exists a subset $\{t_1 < t_2 < \cdots\} \subset \{1, 2, \cdots\}$ and $X \in \mathcal{S}(\mathbb{A})$ with $P_X \neq \lambda$ such that

$$\lim_{n\to\infty}(1/t_n)\sum_{i=1}^{t_n}\delta_{\sigma^i x} = P_X,$$

where λ is the product of the uniform probability measure on $\mathbb{A}$. Since $h(P_X) < \log A$, by Theorem 4, we have

$$\limsup_{n\to\infty}(1/t_n)Q(x_1x_2\cdots x_{t_n}) \leq h(X) < \log A,$$

which contradicts (1.4) and completes one half of the proof.

Let $x = x_1x_2\cdots \in \mathbb{A}^\infty$ be any normal number over $\mathbb{A}$. For $k = 1, 2, \cdots$, define $\xi^k = x_1x_2\cdots x_{2^k} \in \mathbb{A}^{2^k}$. Define $y = \xi^1\xi^2\cdots \in \mathbb{A}^\infty$. Then, it is eeasy to see that y is a normal number over $\mathbb{A}$. We prove that

$$\lim_{n\to\infty}(1/n)Q(y_1y_2\cdots y_n) = 0.$$

For any large n, let $2^{l+1} - 2 \leq n < 2^{l+2} - 2$. Then, $y_1y_2\cdots y_n$ contains $\xi^{l-1}\xi^{l-1}$ as a factor. Let

$$S_h = \{x_ix_{i+1}\cdots x_{i+h-1} \in \mathbb{A}^h;\ i = 1, 2, \cdots, 2^{l-1} - h + 1\}$$

for $h = 1, 2, \cdots, 2^{l-1}$. Then, $|y_1y_2\cdots y_n|_\eta \geq 2$ holds for any $\eta \in S_h$. This implies that

$$\sum_{\eta\in S_h}\psi(|y_1y_2\cdots y_n|_\eta) \geq \psi(2)\#S_h = 2(\log 2)(2^{l-1} - h + 1).$$

Therefore,

$$\begin{aligned}\Xi^n(y_1y_2\cdots y_n) &\geq \sum_{h=1}^{2^{l-1}} 2(\log 2)(2^{l-1} - h + 1)\\ &= (\log 2)2^{l-1}(2^{l-1} + 1) \geq (1/100)n^2\end{aligned}$$

Thus, by Definition 2,

$$\lim_{n\to\infty}(1/n)Q(y_1y_2\cdots y_n) = 0.$$

$\square$

5 Proof of Theorem 5

Let $\tau(t) = \psi(t) - \psi(t-1)$ $(t \geq 1)$. Then, for any $t > 1$,

$$\tau'(t) = \log t - \log(t-1) > 0 \quad \text{and} \quad \tau''(t) = \frac{1}{t} - \frac{1}{t-1} < 0$$

hold, and hence, $\tau(t)$ $(t \geq 1)$ is a concave function. Let $S_2(x)$ be the set of all suffixes ξ of x such that $|x|_\xi \geq 2$. For any $x \in \mathbb{A}^n$, denote

$$D(x) = \sum_{\xi \in S_2(x)} \tau(|x|_\xi).$$

Since if $\xi \in \mathbb{A}^+ \setminus S_2(xa)$, then either $|xa|_\xi = |x|_\xi$ or $\psi(|xa|_\xi) = \psi(|x|_\xi) = 0$, we have

$$\begin{aligned}
&\Xi^{n+1}(xa) - \Xi^n(x) \\
&= \sum_{\xi \in S_2(xa)} (\psi(|xa|_\xi) - \psi(|x|_\xi)) + \sum_{\xi \in \mathbb{A}^+ \setminus S_2(xa)} (\psi(|xa|_\xi) - \psi(|x|_\xi)) \\
&= \sum_{\xi \in S_2(xa)} (\psi(|xa|_\xi) - \psi(|x|_\xi)) = \sum_{\xi \in S_2(xa)} (\psi(|xa|_\xi) - \psi(|xa|_\xi - 1)) \\
&= \sum_{\xi \in S_2(xa)} \tau(|xa|_\xi) = D(xa).
\end{aligned}$$

Moreover, since τ is a concave function such that $\tau(t) \leq \log t + 1$ and $\sum_{a \in \mathbb{A}} |xa|_a = n + A$, we have

$$\begin{aligned}
&(1/A) \sum_{a \in \mathbb{A}} D(xa) - D(x) = \sum_{\xi \in S_2(x)} M(\xi) + \frac{\sum_{a \in \mathbb{A}} \tau(|xa|_a)}{A} \\
&\leq \sum_{\xi \in S_2(x)} M(\xi) + \tau((n+A)/A) \leq \sum_{\xi \in S_2(x)} M(\xi) + \log((n+A)/A) + 1 \\
&\leq \sum_{\xi \in S_2(x)} M(\xi) + \log n - \log A + 1 + (A/n),
\end{aligned}$$

where we put

$$M(\xi) = \frac{\sum_{a \in \mathbb{A}} \tau(|xa|_{\xi a})}{A} - \tau(|x|_\xi).$$

Since $\sum_{a \in \mathbb{A}} |xa|_{\xi a} = |x|_\xi + A - 1$ and $\tau(t)$ $(t \geq 1)$ is a concave function, we have

$$\sum_{a \in \mathbb{A}} \tau(|xa|_{\xi a}) \leq (A - r)\tau(q) + r\tau(q+1) =: B(u, A)$$

holds, where

$$u = |x|_\xi\ , \quad q = \lfloor (u + A - 1)/A \rfloor \quad \text{and} \quad r = \{(u + A - 1)/A\}A.$$

Therefore, since

$$B(2, A) = 2 \log 2 \quad \text{and} \quad B(3, A) = 4 \log 2,$$

if $u = 2$ or 3, then

$$M(\xi) \le (\max_{A \ge 3}(2/A) \log 2 - \tau(2)) \vee (\max_{A \ge 3}(4/A) \log 2 - \tau(3)) = -\log(27/16).$$

On the other hand, since $\log(t-1) + 1 \le \tau(t) \le \log t + 1$ $(t \ge 1)$, if $u \ge 4$, then

$$\begin{aligned} M(\xi) &\le \tau((u + A - 1)/A) - \tau(u) \le \log((u + A - 1)/A) - \log(u - 1) \\ &= \log\left(\frac{1}{u-1} + \frac{1}{A}\right) \le \log(1/3 + 1/2) = -\log(6/5). \end{aligned}$$

Thus, $M(\xi) \le -\log(6/5)$ for any $\xi \in S_2(x)$, and hence,

$$\begin{aligned} &(1/A) \sum_{a \in \mathbb{A}} D(xa) - D(x) \\ &\le -\#S_2(x) \log(6/5) + \log n - \log A + 1 + (A/n). \end{aligned}$$

Let $x_1 \cdots x_m x_{m+1} \cdots \in \mathbb{A}^\infty$ satisfy (1.5). Let $\varepsilon > 0$ and $n_0 \ge 1$ satisfy $-\log A + 1 + (A/n_0) < \varepsilon \log n_0$. Then for any $n \ge m \vee n_0$, we have

$$D(x_1 \cdots x_n x_{n+1}) \le D(x_1 \cdots x_n) + (1 + \varepsilon) \log n - \#S_2(x_1 \cdots x_n) \log(6/5). \tag{5.1}$$

On the other hand, we have

$$D(x_1 \cdots x_n) = \sum_{\xi \in S_2(x_1 \cdots x_n)} \tau(|x_1 \cdots x_n|_\xi) \le \#S_2(x_1 \cdots x_n)(\log n + 1).$$

Let $C_0 := 1/\log(6/5)$ and $n_1 := n_0 \vee \exp(\varepsilon^{-1} + 1)$. Then for any $n \ge n_1$,

$$\#S_2(x_1 \cdots x_n) \ge C_0(1 + 2\varepsilon) \log n$$

holds if

$$D(x_1 \cdots x_n) \ge C_0(1 + 4\varepsilon)(\log n)^2. \tag{5.2}$$

Hence,

$$D(x_1 \cdots x_n x_{n+1}) \le D(x_1 \cdots x_n) - \varepsilon \log n$$

holds if (5.2) is satisfied. Hence, there exists $N \ge n_1 \vee m$ such that

$$D(x_1 \cdots x_N) < C_0(1 + 4\varepsilon)(\log N)^2.$$

Then, we can prove by the induction that for any $n \ge N$, we have

$$D(x_1 \cdots x_n) < C_0(1 + 4\varepsilon)(\log n)^2 + (1 + \varepsilon) \log n. \tag{5.3}$$

For $n = N$, (5.3) is satisfied. Assume that (5.3) is satisfied for n with $n \geq N$. If (5.2) holds, then

$$\begin{aligned} &D(x_1 \cdots x_n x_{n+1}) \leq D(x_1 \cdots x_n) - \varepsilon \log n \\ &< C_0(1+4\varepsilon)(\log n)^2 + (1+\varepsilon)\log n - \varepsilon \log n \\ &< C_0(1+4\varepsilon)(\log(n+1))^2 + (1+\varepsilon)\log(n+1). \end{aligned}$$

If (5.2) is not satisfied, then by (5.1),

$$\begin{aligned} &D(x_1 \cdots x_n x_{n+1}) \leq D(x_1 \cdots x_n) + (1+\varepsilon)\log n \\ &< C_0(1+4\varepsilon)(\log n)^2 + (1+\varepsilon)\log n \\ &< C_0(1+4\varepsilon)(\log(n+1))^2 + (1+\varepsilon)\log(n+1). \end{aligned}$$

Thus, (5.3) holds for any $n \geq N$.

Since

$$\Xi^{n+1}(x_1 \cdots x_n x_{n+1}) - \Xi^n(x_1 \cdots x_n) = D(x_1 \cdots x_n x_{n+1}),$$

(5.3) implies that for any $n \geq N$,

$$\begin{aligned} &\Xi^n(x_1 \cdots x_n) < \Xi^N(x_1 \cdots x_N) + \sum_{k=N+1}^{n} (C_0(1+4\varepsilon)(\log k)^2 + (1+\varepsilon)\log k) \\ &= C_0(1+4\varepsilon)n(\log n)^2(1+o(1)) \quad (\text{as } n \to \infty). \end{aligned}$$

Since $\varepsilon > 0$ is arbitrary, it follows from this that

$$\liminf_{n\to\infty}(1/n)Q(x_1x_2\cdots x_n) \geq 1/(2C_0) = (1/2)\log(6/5),$$

which complete the proof of Theorem 5.

6 Examples

Example 1. We draw the graph of $\mathcal{H}_k(x_1x_2\cdots x_{2^8})$ for $1 \leq k \leq 2^8$ with respect to the binary random numbers, the random numbers with distribution $(1/8, 7/8)$ for the values 0 and 1, and Fibonacci sequence together with the graph of $\log(2^8 - k + 1)$ in Figure 3.

Example 2. We draw the graph of $Q(x_1x_2\cdots x_n)$ for $1 \leq n \leq 2^9$ with respect to the binary random numbers, Rudin-Shapiro sequence, Thue-Morse sequence and Fibonacci sequence in Figure 4. Except for the binary random numbers, they have entropy 0. Even though, there are differences in the complexities, which are discriminated by the Q-values. It is also interesting to see that the Q-values for Rudin-Shapiro sequence is not monotone in n.

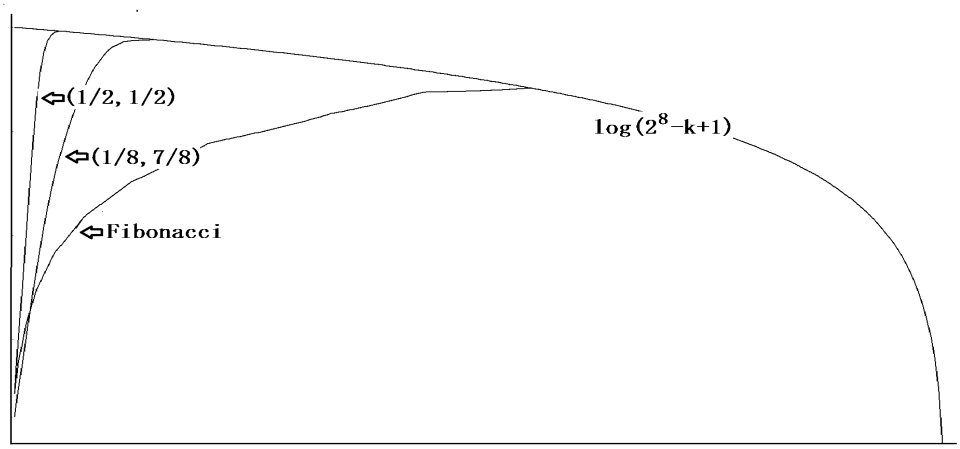

Figure 3: $\mathcal{H}_k(x_1x_2\cdots x_{2^8})$ for $1 \le k \le 2^8$

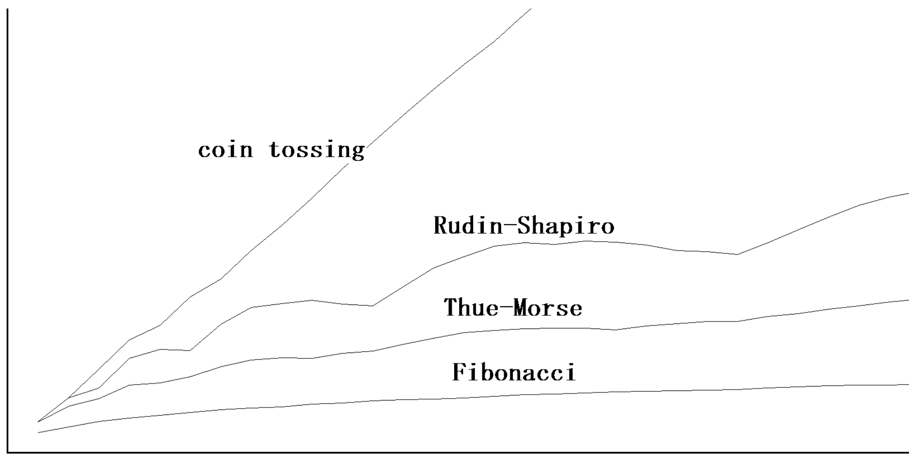

Figure 4: $Q(x_1x_2\cdots x_n)$ for $1 \le n \le 2^9$

Example 3. Consider $Q(x_1x_2\cdots x_{14})/14$ with the logarithmic base 2 for binary words of length 14. The smallest values and the words attaining them are as follows:

0.308	:	00000000000000 and its symmetric variants
0.367	:	00000000000001 and its symmetric variants
0.435	:	00000000000010 and its symmetric variants
0.441	:	00000000000011 and its symmetric variants
0.448	:	01010101010101 and its symmetric variants
0.507	:	00000000000100 and its symmetric variants
0.522	:	00000000000101, 01000000000001 and their symmetric variants

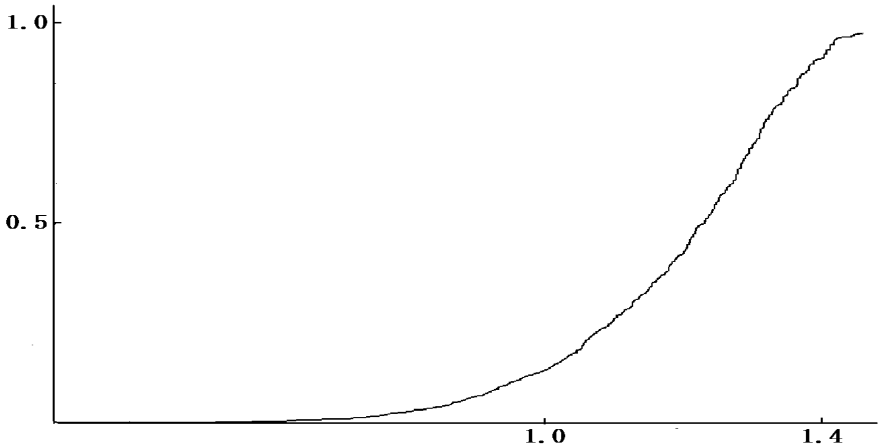

Figure 5: The distribution of $Q/14$ with the logarithmic base 2 for the binary words of length 14

The biggest value 1.46 are attained by the equi-distributed words like

$$\begin{array}{llll}
11100010011010, & 11100010100110, & 11100010110100, & 11100011010010 \\
11100100001101, & 11100100011010, & 11100101000011, & 11100101000110 \\
11100101101000, & 11100110100001, & 11100110100010, & 11101000010110 \\
11101000011001, & 11101000101100, & 11101000110010, & 11101001011000 \\
11101001100001, & 11101001100010, & 11101011000010, & 11101011000100 \\
11101011001000, & 11101100001001, & 11101100001010, & 11101100010100 \\
11101100100001, & 11101100101000, & 11100101000011, & \cdots\cdots\cdots\cdots
\end{array}$$

We draw the distribution function of the $Q/14$ for all the binary words of length 14 (Figure 5).

Example 4. Let $x = x_1x_2\cdots x_{200} \in \{0,1\}^{200}$ be such that $x_1x_2\cdots x_{20} = 0^{20}$ and for $i = 21, 22, \cdots, 200$, x_i is chosen so that $x_i = 0$ if and only if

$$\Xi^i(x_1x_2\cdots x_{i-1}0) \le \Xi^i(x_1x_2\cdots x_{i-1}1).$$

In fact,

$$\begin{array}{rl}
x = & 00000000000000000000111010110010100111110111001101 \\
& 00010010111100011000101010111011011001110000100011 \\
& 11010010011001000011011111110010110101000010111000 \\
& 10000010100011010111110000011001100001111110101010
\end{array}$$

Then, Figure 6 is the graph of $Q(x_1x_2\cdots x_i)$ (with the logarithmic base 2) with respect to $i = 1, 2, \cdots, 200$. The infinite words defined like this looks like having the full Q-value 1 per letter, although we can prove only a positive Q-value per letter in Theorem 5.

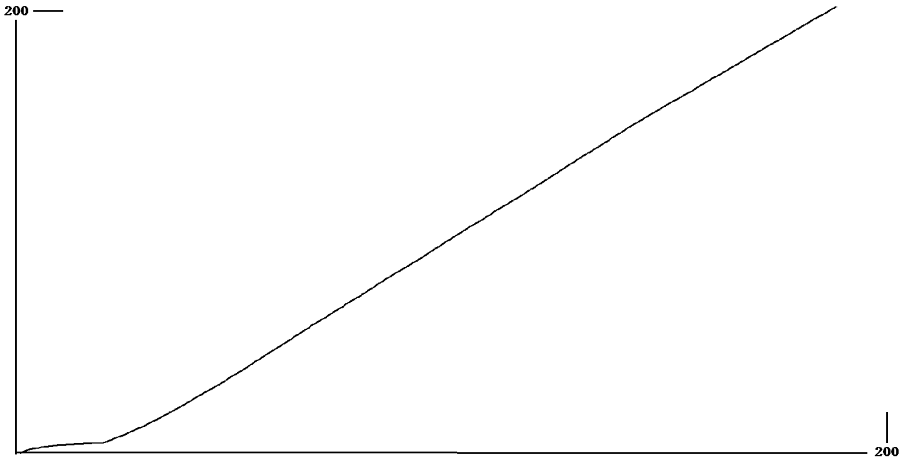

Figure 6:

References

[1] Teturo Kamae and Yu-Mei Xue, An easy criterion for randomness Sankhya A 77-1 (2015), pp.126-152.

[2] Teturo Kamae and Dong Han Kim, A characterization of eventually periodicity, Theoretical Computer Science 581 (2015), pp.1-8.

[3] J. Ziv and A. Lempel, A universal algorithm for sequential data compression, IEEE Trans. Inform. Theory IT-23 (1977), pp.337-343.

[4] Donald Ornstein and Benjamin Weiss, How samples reveals a process, The Annals of Probability 18-3 (1990), pp.905-930.

[5] P. C. Shields, The Ergodic Theory of Discrete Sample Paths, GSM vol. 13, Amer. Math. Soc., Providence, 1996.

[6] M. Li and P. Vitányi, An introduction to Kolmogorov complexity and Its applications. Springer, New York, third edition, 2008.