IEICE TRANS. FUNDAMENTALS, VOL.E89–A, NO.12 DECEMBER 2006

PAPER

Properties of a Word-Valued Source with a Non-prefix-free Word Set

Takashi ISHIDA[†a)], **Masayuki GOTO**[††], **Toshiyasu MATSUSHIMA**[†], ***Members***, ***and*** **Shigeichi HIRASAWA**[†], ***Fellow***

SUMMARY Recently, a word-valued source has been proposed as a new class of information source models. A word-valued source is regarded as a source with a probability distribution over a word set. Although a word-valued source is a nonstationary source in general, it has been proved that an entropy rate of the source exists and the Asymptotic Equipartition Property (AEP) holds when the word set of the source is prefix-free. However, when the word set is not prefix-free (non-prefix-free), only an upper bound on the entropy density rate for an i.i.d. word-valued source has been derived so far. In this paper, we newly derive a lower bound on the entropy density rate for an i.i.d. word-valued source with a finite non-prefix-free word set. Then some numerical examples are given in order to investigate the behavior of the bounds.

key words: *word-valued source, word set, non-prefix-free, entropy rate, entropy density rate, Asymptotic Equipartition Property (AEP)*

1. Introduction

Recently, *a word-valued source* has been proposed as a new class of information source models [14]. A word-valued source is regarded as a source which has a probability distribution over *a word set*, where *a word* is defined as a finite sequence over a finite alphabet. Data sequences to be compressed are possibly emitted by word unit from the source, so it is natural to assume that the source model has the probability distribution over the word set. Although a word-valued source is a nonstationary source in general [14], it was shown that an entropy rate of the source exists and that the Asymptotic Equipartition Property (AEP) holds under some assumptions [6], [14].

M. Nishiara and H. Morita [14] have studied *an i.i.d. word-valued source* which is defined as a pair of an independently and identically distributed (i.i.d.) source with a countable alphabet $\mathcal{Y}$ and a mapping function ϕ which maps each element $y \in \mathcal{Y}$ to a word w over a finite alphabet $\mathcal{W}$. They have derived an upper bound on an entropy density rate of an i.i.d. word-valued source. Furthermore, they have shown that if ϕ is *prefix-free*, that is, any word is not a prefix of other words, then the entropy rate is obtained by a simple expression and the AEP holds. M. Goto et al. [6] generalized an i.i.d. word-valued source into an ergodic word-valued source which is defined by a pair of an ergodic source with a countable alphabet and a mapping ϕ from each symbol to a word. They have derived an entropy rate of the source when ϕ is prefix-free. Moreover, the recurrence time theorem and the universality of LZ77 code [16] for the source were shown. T. Ishida et al. [9], [10] considered the source which emits data sequences by block unit (the block stationary source), and discussed the universality of LZ78 code [17] and Bayes code [12].

In these arguments, prefix-free property plays an important role to derive an entropy rate of the source. This property makes analysis easy for the word-valued sources. However, it is also important to analyze the non-prefix-free cases to reflect more of the probability structure of the actual data sequences. For example, Japanese sentences may be regarded as the sequence which is emitted from a word-valued source with non-prefix-free word set.

A word-valued source has been proved to be equivalent to the recurrent source [14]. For the non-prefix-free cases (non-prefix-free word-valued source), it is not generally clear whether the source has an entropy rate. Only an upper bound of the entropy rate has been derived in [14] until now. There still remain many issues which are not yet investigated about the property of an entropy rate for the non-prefix-free word-valued source.

The purpose of our study is to clarify the property of an entropy rate of a non-prefix-free word-valued source. As the first step of the study, in this paper, we derive a lower bound on the entropy density rate for the non-prefix-free case theoretically, and then we show some numerical experiments in order to verify the bound.

This paper is organized as follows: Sect. 2 defines the i.i.d. word-valued source and reviews the previous studies. In Sect. 3, we describe the property of the non-prefix-free word-valued source. Section 4 provides a lower bound on an entropy density rate of the source, and we present some numerical experiments in Sect. 5. Finally, concluding remarks are given in Sect. 6.

2. Word-Valued Source

2.1 Definition of an i.i.d. Word-Valued Source

At first, the definition of an i.i.d. word-valued source by

Manuscript received May 18, 2006.
Manuscript revised July 28, 2006.
Final manuscript received August 17, 2006.

[†]The authors are with the Department of the Industrial and Management Systems Engineering, School of Science and Engineering, Waseda University, Tokyo, 169-8555 Japan.
[††]The author is with the Faculty of Environmental and Information Studies, Musashi Institute of Technology, Yokohama-shi, 224-0015 Japan.
a) E-mail: ishida@hirasa.mgmt.waseda.ac.jp

DOI: 10.1093/ietfec/e89–a.12.3710

Nishiara and Morita is described below.

Definition 1: (i.i.d. word-valued source [14]) Let $\boldsymbol{Y} = Y_1Y_2Y_3\cdots$ be an i.i.d. source with a countable alphabet $\mathcal{Y}$. Here, Y_i is a random variable. Let $\mathcal{X}$ be a finite alphabet with $\|\mathcal{X}\|$ symbols[†] and $\mathcal{X}^*$ be a set of all finite sequences over $\mathcal{X}$. That is, $\mathcal{X}^* = \cup_{i=0}^{\infty}\mathcal{X}^i$. Here, $\mathcal{X}^i$ is a product space of $\mathcal{X}$, i.e., $\mathcal{X}^i = \underbrace{\mathcal{X}_1 \times \mathcal{X}_2 \cdots \mathcal{X}_i}_{i}$. A mapping ϕ is given by $\phi : \mathcal{Y} \to \mathcal{X}^*$. $\phi(\boldsymbol{Y})$ is the sequence of the random variable X obtained by a concatenation of the sequences $\phi(Y_1)$, $\phi(Y_2)$, $\phi(Y_3), \cdots$ for $\boldsymbol{Y} = Y_1Y_2Y_3\cdots$. Then *an i.i.d. word-valued source* $\boldsymbol{X} = X_1X_2X_3\cdots$ is defined as follows:

$$\boldsymbol{X} \stackrel{\text{def}}{=} \phi(\boldsymbol{Y}) = X_1X_2X_3\cdots. \tag{1}$$

□

Goto et al. [6] have extended $\boldsymbol{Y}$ to an ergodic source. Such a case is called an ergodic word-valued source.

The mapping $\phi : \mathcal{Y} \to \mathcal{X}^*$ is said to be *prefix-free* if $\phi(y)$ is not a prefix of $\phi(y')$ for any $y \neq y'$.

Next, we define a word and a word set as follows:

Definition 2: (Word and Word set) Let $W = \phi(Y)$ be a random variable over a countable alphabet $\mathcal{W} \subseteq \mathcal{X}^*$. We denote a sequence of $W_1 = \phi(Y_1)$, $W_2 = \phi(Y_2)$, $W_3 = \phi(Y_3)$ $\cdots$ by $\boldsymbol{W} = W_1W_2W_3\cdots$. Let $w = \phi(y) \in \mathcal{W}$ be the realization value of W for $y \in \mathcal{Y}$. We call w *a word*, and $\mathcal{W}$ *a word set*. And the word set $\mathcal{W}$ is said to be *a prefix-free word set* if and only if ϕ is prefix-free. On the contrary, if ϕ is not prefix-free, then $\mathcal{W}$ is said to be *a non-prefix-free word set*.

□

For each finite number $n = 1, 2, \cdots$, we denote the sequence of X with length n by $X^n = X_1X_2X_3\cdots X_n$, and its realization value by $x^n = x_1x_2x_3\cdots x_n$ respectively. Similarly, for each finite number $m = 1, 2, \cdots$, we use the notation such that $Y^m = Y_1Y_2Y_3\cdots Y_m$, $y^m = y_1y_2y_3\cdots y_m$, $W^m = W_1W_2W_3\cdots W_m$, and $w^m = w_1w_2w_3\cdots w_m$. Furthermore, for each finite number $m = 1, 2, \cdots$, we denote by $\phi(Y^m)$ a sequence of X obtained by a concatenation of $W_1 = \phi(Y_1)$, $W_2 = \phi(Y_2)$, $\cdots$, $W_m = \phi(Y_m)$ for $Y^m = Y_1Y_2\cdots Y_m$.

Word-valued source emits a sequence of words $w^m = \phi(y_1)\phi(y_2)\cdots\phi(y_m)$. A sequence y^m can not be observed. We can only recognize a sequence x^n which is obtained by concatenation of words $w = \phi(y)$. That is, $x^n = \phi(y^m)$ and $n = |\phi(y_1)| + |\phi(y_2)| + \cdots + |\phi(y_m)| = |w_1| + |w_2| + \cdots + |w_m|$ for given m[††].

Example 1 (word and word set): Let $\mathcal{Y}$ be $\mathcal{Y} = \{1, 2, 3, 4\}$ and $\mathcal{X}$ be $\mathcal{X} = \{0, 1\}$. And mapping $\phi : \mathcal{Y} \to \mathcal{W} \subset \mathcal{X}^*$ is given by $\phi(1) = 0$, $\phi(2) = 01$, $\phi(3) = 101$, $\phi(4) = 111$. Considering $y^4 = y_1y_2y_3y_4 = 2431$, the words are $w_1 = \phi(y_1) = 01$, $w_2 = \phi(y_2) = 111$ and so on. In this example, the word set $\mathcal{W} = \{0, 01, 101, 111\}$ corresponds to a non-prefix-free case. Word-valued source emits the sequence word by word, that is, $w^4 = w_1w_2w_3w_4 = \phi(y_1)\phi(y_2)\phi(y_3)\phi(y_4) =$ 01 111 101 0. However we can only observe the sequence x^9 which is obtained by concatenating the words, that is, $x^9 = \phi(y^4) = 011111010$. We never see the sequence y^4. Here $n = |w_1| + |w_2| + |w_3| + |w_4| = 2 + 3 + 3 + 1 = 9$. □

Hereafter, we call w^m (in some cases also y^m) "*a word sequence*," and x^n "*a symbol sequence*" respectively.

2.2 Probability Distribution of an i.i.d. Word-Valued Source

Let $\mathcal{Y}^i = \mathcal{Y}_1 \times \mathcal{Y}_2 \times \cdots \mathcal{Y}_i$ be a product space of $\mathcal{Y}$. And let $\mathcal{Y}^\infty = \mathcal{Y}_1 \times \mathcal{Y}_2 \times \cdots$ be a sample space with each elementary event $y^\infty = y_1y_2\cdots \in \mathcal{Y}^\infty$. Random variables Y^m and W^m are defined as the functions on $\mathcal{Y}^\infty$, $Y^m : \mathcal{Y}^\infty \to \mathcal{Y}^m$ and $W^m : \mathcal{Y}^\infty \to \mathcal{W}^m$[†††]. Then, the probability distributions of Y^m and W^m are defined as follows:

$$P_{Y^m}(y^m) \stackrel{\text{def}}{=} \Pr\left(\{y^\infty \in \mathcal{Y}^\infty | Y^m(y^\infty) = y^m\}\right), \tag{2}$$

$$P_{W^m}(w^m) \stackrel{\text{def}}{=} \Pr\left(\{y^\infty \in \mathcal{Y}^\infty | W^m(y^\infty) = w^m\}\right). \tag{3}$$

Throughout this paper, {A} denotes a set of elements satisfying the condition A.

Especially, if the mapping $\phi : \mathcal{Y} \to \mathcal{X}^*$ is one-to-one, then

$$P_{W^m}(w^m) = P_{Y^m}(y^m), \tag{4}$$

holds for $w^m = \phi(y_1)\phi(y_2)\cdots\phi(y_m)$. If $m = 1$, we simply use the notation $P_Y(y) = P_{Y^1}(y^1)$ and $P_W(w) = P_{W^1}(w^1)$.

Let π be the prefix operator such that

$$y^m\pi = y^{m-1}, \tag{5}$$

where y^0 means the null sequence λ with length 0. And let $\pi\{x^n\}$ be the set of all prefixes of the sequence x^n including λ and x^n itself. Denoting the set of all finite sequences of y by $\mathcal{Y}^* = \cup_{i=0}^{\infty}\mathcal{Y}^i$, we define $\Gamma_\phi(x^n) \subset \mathcal{Y}^*$ as the set of y^* such that x^n is a prefix of $\phi(y^*)$ but not so of $\phi(y^*\pi)$:

$$\Gamma_\phi(x^n) \stackrel{\text{def}}{=} \{y^* \in \mathcal{Y}^* | \, (x^n \in \pi\{\phi(y^*)\}) \wedge (x^n \notin \pi\{\phi(y^*\pi)\})\}. \tag{6}$$

Let Y^* be a random variable over $\mathcal{Y}^*$, then, the probability distribution of X^n is given by

$$\begin{aligned} P_{X^n}(x^n) &\stackrel{\text{def}}{=} \Pr(\{y^\infty \in \mathcal{Y}^\infty | Y^*(y^\infty) = y^* \in \Gamma_\phi(x^n)\}) \\ &= \sum_{y^* \in \Gamma_\phi(x^n)} P_{Y^*}(y^*). \end{aligned} \tag{7}$$

When $n = 1$, we simply use the notation $P_X(x) = P_{X^1}(x^1)$.

[†] $\|\mathcal{A}\|$ means a cardinality of a set $\mathcal{A}$.

[††] $|\phi(y)|$ $(= |w|)$ means the length of sequence $\phi(y)$ $(= w)$.

[†††] More accurately, Y^m and W^m should be written as $Y^m(y^\infty)$ and $W^m(y^\infty)$ respectively. However, if there is no likelihood of confusion, we adopt the notations omitting "(y^∞)" for brevity. The same applies to X^n.

2.3 Entropy Rate of a Word-Valued Source

An entropy density rate is defined by $-\frac{1}{n}\log P_{X^n}(X^n)$ in [7]. If there exists a limit of an expectation of the entropy density rate for $\boldsymbol{X}$;

$$H(\boldsymbol{X}) = \lim_{n\to\infty} E_{P_{X^n}}\left[-\frac{1}{n}\log P_{X^n}(X^n)\right], \tag{8}$$

then $H(\boldsymbol{X})$ is called *an entropy rate* of the word-valued source $\boldsymbol{X}$ [6], [14]. Here, $E_P[\cdot]$ means an expectation over the probability distribution P. We assume the base of logarithm is 2 throughout this paper.

Let $H(Y)$ and $H(W)$ be the entropy of i.i.d. sources $\boldsymbol{Y}$ and $\boldsymbol{W}$ respectively, which are given by

$$H(Y) = -\sum_{y\in\mathcal{Y}} P_Y(y)\log P_Y(y), \tag{9}$$

and

$$H(W) = -\sum_{w\in\mathcal{W}} P_W(w)\log P_W(w). \tag{10}$$

If the mapping $\phi : \mathcal{Y} \to \mathcal{W}$ is one-to-one, then $H(Y) = H(W)$ holds. *An expected word length* $E[|W|]$ is given by

$$E[|W|] = \sum_{w\in\mathcal{W}} |w| \cdot P_W(w). \tag{11}$$

Nishiara and Morita [14] have shown an entropy rate and the AEP for an i.i.d. word-valued source.

Lemma 1: (Entropy rate and the AEP of an i.i.d. word-valued source [14]) Let $\boldsymbol{Y}$ be an i.i.d. source with countable alphabet $\mathcal{Y}$. When $\boldsymbol{X} = \phi(\boldsymbol{Y})$, $H(Y) < \infty$, and $E[|W|] < \infty$, then we have following formulas[†].

$$\limsup_{n\to\infty} E_{P_{X^n}}\left[-\frac{1}{n}\log P_{X^n}(X^n)\right] \le \frac{H(Y)}{E[|W|]}, \tag{12}$$

and

$$\limsup_{n\to\infty}\left[-\frac{1}{n}\log P_{X^n}(X^n)\right] \le \frac{H(Y)}{E[|W|]}, \quad a.s. \tag{13}$$

Furthermore, if ϕ is prefix-free, then

$$H(\boldsymbol{X}) = \frac{H(Y)}{E[|W|]}, \tag{14}$$

and

$$\lim_{n\to\infty}\left[-\frac{1}{n}\log P_{X^n}(X^n)\right] = \frac{H(Y)}{E[|W|]}, \quad a.s. \tag{15}$$

hold. □

In previous studies, the word-valued source have been discussed mainly in the case that the mapping $\phi : \mathcal{Y} \to \mathcal{W}$ is prefix-free, i.e., $\mathcal{W}$ is a prefix-free word set. When ϕ is prefix-free, the existence of an entropy rate and the AEP of the source have been already proved for an i.i.d. word-valued source (Lemma 1 [14]) and for an ergodic word-valued source (Goto et al. [6]). It has also been proved that LZ77 code [16] is universal for an ergodic word-valued source [6], and that LZ78 code [17] is universal for an ergodic word-valued source [9], [10]. In [9], it was shown that the Bayes code [12] can be constructed for a block-wise word-valued source which has the word set with unknown fixed word length $h \ge 1$.

3. Non-prefix-free Word-Valued Source

Prefix-free property plays an important role in analyses of the word-valued sources. In order to derive an entropy rate, the property makes analyses easy for the word-valued sources.

When ϕ is prefix-free, the mapping from the word sequences $\mathcal{W}^m$ to the symbol sequences $\mathcal{X}^*$ becomes one-to-one. From an observed symbol sequence x^n, if we know the word set $\mathcal{W}$, we can uniquely determine a certain word sequence w^m which is actually emitted from the source when $n = \sum_{i=1}^{m} |w_i|$. That is, we can see that where the unobserved "gaps" between each word in the symbol sequence x^n are. By determining where the gaps are in given symbol sequence x^n, we can specify one certain word sequence w^m.

When ϕ is not prefix-free, on the other hand, the mapping from the word sequences $\mathcal{W}^*$ to the symbol sequences $\mathcal{X}^n$ is generally a many-to-one mapping.

Example 2 (many-to-one mapping of $\mathcal{W}^*$ to $\mathcal{X}^n$):
Figure 1 shows an example of the relation $\mathcal{W}^*$ and $\mathcal{X}^n$. It is found that some word sequences w^* are mapped to one symbol sequence $x^9 = 001111000$. In this case, the elements of $\Gamma_\phi(x^9) = \Gamma_\phi(001111000)$ are shown in Table 1. Here, $\|\Gamma_\phi(001111000)\| = 21$. And then, $P_{X^n}(001111000)$ is obtained by the summation of $P_{Y^*}(y^*)$ for all $y^* \in \Gamma_\phi(001111000)$. □

Generally, the appearance probability $P_{X^n}(x^n)$ has a

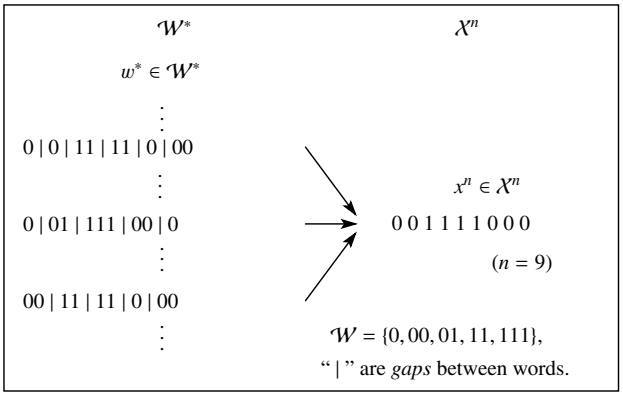


Fig. 1 Mapping from $\mathcal{W}^*$ to $\mathcal{X}^n$ in the case that ϕ is not prefix-free.

[†]We use the notation "$f_n \le g_n$, $a.s.$ when $n \to \infty$" to represent "$\Pr\{f_n > g_n, \text{infinitely often } n\} = 0$."

Table 1 Elements of $\Gamma_\phi(001111000)$ in Fig. 1.

0|0|11|11|0|0|0	00|11|11|00|0
0|0|11|11|0|0|00	00|11|11|00|00
0|0|11|11|0|0|01	00|11|11|00|01
0|0|11|11|0|00	0|01|111|0|0|0
0|0|11|11|00|0	0|01|111|0|0|00
0|0|11|11|00|00	0|01|111|0|0|01
0|0|11|11|00|01	0|01|111|0|00
00|11|11|0|0|0	0|01|111|00|0
00|11|11|0|0|00	0|01|111|00|00
00|11|11|0|0|01	0|01|111|00|01
00|11|11|0|00	

complicated structure depending on the mapping from $\mathcal{W}^*$ to $\mathcal{X}^n$. This is one of the reasons why the analysis of the non-prefix-free cases is difficult. If the word-valued source has the non-prefix-free mapping, then we call it a non-prefix-free word-valued source.

Nishiara and Morita [14] have stated that the word-valued source is equivalent to recurrent irreducible countable-states source with a fixed initial state and that it is not stationary in general. No explicit single letter expression of the entropy rate for the model is known [4]. In [14], the upper bound on the entropy density rate (Eq. (12)) has been shown. However, there is no argument about the behavior of the actual value of the entropy density rate or its upper bound. As to a lower bound, only the trivial bound, that is $H(\boldsymbol{X}) \geq 0$, has been known until now.

4. Main Results

We newly derive a lower bound on an entropy density rate of the non-prefix-free word-valued source for the purpose of clarification of the source. In our analysis, we restrict the case of an i.i.d. word-valued source with a finite word set in order to obtain the bound.

Theorem 1 shows a lower bound on an entropy density rate for a non-prefix-free i.i.d. word-valued source with a finite word set. At first, we give a definition of the model for the theorem.

4.1 Model

Definition 3: (Non-prefix-free i.i.d. word-valued source with a finite word set) Let $\boldsymbol{Y}$ be an i.i.d. source with *a finite* alphabet $\mathcal{Y}$, and $\mathcal{X}$ be a finite alphabet with $\|\mathcal{X}\|$ symbols. Let ϕ be a one-to-one mapping $\phi : \mathcal{Y} \to \cup_{s=1}^{K} \mathcal{X}^s$ $(= \mathcal{W})$, where K is the maximum length of the words. Then a non-prefix-free i.i.d. word-valued source with a finite word set is given by $\boldsymbol{X} = \phi(\boldsymbol{Y})$. □

Here, ϕ is restricted as one-to-one mapping. This restriction is given in order to simplify the relation between $P_Y(y)$ and $P_W(w)$. Even if ϕ is one-to-one, the mapping from $\mathcal{W}^*$ to $\mathcal{X}^n$ is generally a many-to-one mapping because $\mathcal{W}$ is not prefix-free word set. As mentioned in section 3, the essence of the problem of the non-prefix-free word-valued source is in the many-to-one mapping from $\mathcal{W}^*$ to $\mathcal{X}^n$.

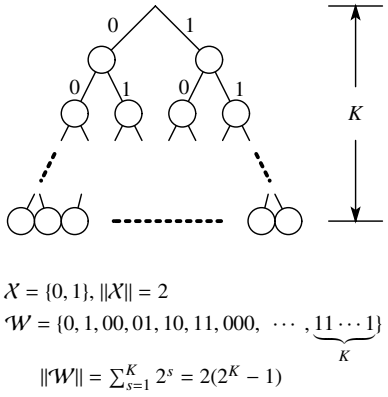


Fig. 2 An example of *a word tree*.

Since ϕ is a one-to-one mapping, $P_Y(y) = P_W(w)$ when $w = \phi(y)$ and $H(W) = H(Y)$ hold. An entropy rate of $\boldsymbol{W}$ is given by

$$H(W) = -\sum_{w \in \mathcal{W}} P_W(w) \log P_W(w). \tag{16}$$

The word set $\mathcal{W}$ defined above is represented by a complete $\|\mathcal{X}\|$-ary tree called *a word tree*, where the depths of all leaf nodes are K (Fig. 2). Each branch in the word tree is labeled by a symbol $x \in \mathcal{X}$. Each of the nodes represents a word $w \in \mathcal{W}$ which corresponds to a sequence of the labels on the path from the root to each node.

We can construct the various models depending on the probability distribution $P_W(w)$ over a word set $\mathcal{W}$. It also happens that the model described above becomes equivalent to the prefix-free case by setting $P_W(w) = 0$ to some words appropriately. However, in these arguments, we consider only the non-prefix-free cases except the prefix-free cases. If the word set is prefix-free, we need not consider the bounds on the entropy rate because $H(\boldsymbol{X})$ is obtained by Lemma 1.

From the same argument as [14], it can be easily shown that the i.i.d. word-valued source with a finite non-prefix-free word set defined above is equivalent to recurrent irreducible finite-states source with a fixed initial state. The entropy rate of the source also has not been shown until now, and only the upper bound was given by [14].

4.2 Main Theorem

Now we show a lower bound on an entropy density rate of the non-prefix-free i.i.d. word-valued source with a finite word set.

Theorem 1: (The lower bound on an entropy density rate of a non-prefix-free i.i.d. word-valued source with a finite word set) For a non-prefix-free i.i.d. word-valued source with a finite word set with the maximum word length K,

$$\liminf_{n\to\infty} \left[-\frac{1}{n} \log P_{X^n}(X^n) \right] \geq \frac{H(Y)}{E[|W|]} - \frac{H(S)}{E[|W|]}, \quad a.s. \tag{17}$$

holds. Here $H(S)$ and $P_S(s)$ are given by

$$H(S) = -\sum_{s=1}^{K} P_S(s) \log P_S(s), \tag{18}$$

$$P_S(s) = \sum_{\{w \in \mathcal{W}: |w|=s\}} P_W(w), \quad (s \le K), \tag{19}$$

where S is a random variable representing the word length $s = |w|$, and $P_S(s)$ is a probability distribution of S.

Proof: See Appendix. □

From Lemma 1 and Theorem 1 we obtain the following results.

Corollary 1: If there exists $H(\boldsymbol{X})$ of an i.i.d. non-prefix-free word-valued source with a finite word set,

$$\frac{H(W)}{E[|W|]} - \frac{H(S)}{E[|W|]} \le H(\boldsymbol{X}) \le \frac{H(W)}{E[|W|]}, \tag{20}$$

holds.

Proof: From the Fatou's Lemma [3], we have

$$E_{P_{X^n}}\left[\liminf_{n\to\infty}\left[-\frac{1}{n}\log P_{X^n}(X^n)\right]\right] \le \liminf_{n\to\infty} E_{P_{X^n}}\left[-\frac{1}{n}\log P_{X^n}(X^n)\right]. \tag{21}$$

Here, if there exists $H(\boldsymbol{X})$,

$$\liminf_{n\to\infty} E_{P_{X^n}}\left[-\frac{1}{n}\log P_{X^n}(X^n)\right] = H(\boldsymbol{X}), \tag{22}$$

is satisfied. By substituting them into Lemma 1 and Theorem 1, we have

$$\frac{H(Y)}{E[|W|]} - \frac{H(S)}{E[|W|]} \le H(\boldsymbol{X}) \le \frac{H(Y)}{E[|W|]}. \tag{23}$$

Because ϕ is a one-to-one mapping, we can rewrite $H(Y)$ by $H(W)$. Then Corollary 1 is obtained. □

Remark 1: From Eq. (20), it is found that the difference between the upper bound and the lower bound is given by $H(S)/E[|W|]$. The source with the smaller value of $H(S)$ or the larger value of $E[|W|]$ has the smaller difference. □

Remark 2: Because the joint probability of W and S, denoting $P_{WS}(w, s)$, satisfies

$$P_{WS}(w, s) = \begin{cases} P_W(w) & s = |w|, \\ 0 & \text{otherwise}, \end{cases} \tag{24}$$

we have the following equation:

$$\sum_{s=1}^{K} P_{WS}(w, s) \log P_{WS}(w, s) = P_W(w) \log P_W(w). \tag{25}$$

Denoting the conditional probability and the conditional entropy of W given S by $P_{W|S}(w|s)$ and $H(W|S)$ respectively, then,

$$\begin{aligned} H(W|S) &= -\sum_{s=1}^{K}\sum_{w\in\mathcal{W}} P_{WS}(w, s) \log P_{W|S}(w|s) \\ &= -\sum_{s=1}^{K}\sum_{w\in\mathcal{W}} P_{WS}(w, s) \log \frac{P_{WS}(w, s)}{P_S(s)} \\ &= -\sum_{w\in\mathcal{W}}\sum_{s=1}^{K} P_{WS}(w, s) \log P_{WS}(w, s) \\ &\quad + \sum_{s=1}^{K}\sum_{w\in\mathcal{W}} P_{WS}(w, s) \log P_S(s) \\ &= -\sum_{w\in\mathcal{W}} P_W(w) \log P_W(w) + \sum_{s=1}^{K} P_S(s) \log P_S(s) \\ &= H(W) - H(S) \end{aligned} \tag{26}$$

holds. From Eq. (26), we can rewrite the lower bound (left hand side of Eq. (20)) as

$$\frac{H(W)}{E[|W|]} - \frac{H(S)}{E[|W|]} = \frac{H(W|S)}{E[|W|]}. \tag{27}$$

It is found that the value of the lower bound depends on the dispersion in the distribution $P_W(w)$ when $P_S(s)$ was given. □

Remark 3: The trivial bounds on the entropy rate of $H(\boldsymbol{X})$ is given by

$$0 \le H(\boldsymbol{X}) \le \log \|\mathcal{X}\|. \tag{28}$$

The lower bound in Eq. (20) always satisfies

$$\frac{H(W)}{E[|W|]} - \frac{H(S)}{E[|W|]} = \frac{H(W|S)}{E[|W|]} \ge 0. \tag{29}$$

On the contrary, the upper bound in Eq. (20) does not always satisfy

$$\frac{H(W)}{E[|W|]} \le \log \|\mathcal{X}\|. \tag{30}$$

When the length of the words $w \in \mathcal{W}$ satisfies Kraft's inequality, namely when

$$\sum_{\{w\in\mathcal{W}: P_W(w)>0\}} \|\mathcal{X}\|^{-|w|} \le 1, \tag{31}$$

holds, the expected word length is lower-bounded by the entropy. That is,

$$\begin{aligned} -\sum_{w\in\mathcal{W}} P_W(w) \log_{\|\mathcal{X}\|} P_W(w) &= \frac{H(W)}{\log \|\mathcal{X}\|} \\ &\le \sum_{w\in\mathcal{W}} P_W(w) \cdot |w| = E[|W|], \end{aligned} \tag{32}$$

is satisfied, and then, Eq. (30) is ensured. However, in the non-prefix-free cases, the Kraft's inequality does not always hold.

Therefore the upper bound on an entropy rate of a non-prefix-free word-valued source is given by

$$\min\left\{\frac{H(W)}{E[|W|]},\ \log \|\mathcal{X}\|\right\}. \tag{33}$$

□

Remark 4: If $H(S) = H(W)$, the left hand side of Eq. (20) is equal to 0, so the lower bound becomes trivial. Obviously, this situation corresponds to the case that the lengths of all words in $\mathcal{W}$ are different from one another. We notice that this may also happen when the word set is prefix-free. It is because that in order to derive a lower bound on the entropy rate of a non-prefix-free word-valued source, we have directed our attention to only the length of words. However, when the word set is prefix-free, we can use Lemma 1 for verifying the entropy rate. Our theorem should be applied only for the non-prefix-free word-valued source. □

Remark 5: When $H(S) = 0$, the lower bound completely corresponds to the upper bound. If all of the words have the same length h, then $H(S) = 0$ holds and $H(\boldsymbol{X}) = H(W)/E[|W|] = H(W)/h$. This result was referred in [9] and it is the special case when ϕ is prefix-free. □

5. Numerical Experiments

In this section, focusing only upon the non-prefix-free word-valued source, some numerical experiments are shown in order to investigate the behavior of the value of the entropy density rate and the bounds.

5.1 Conditions of Experiments

We set the word set $\mathcal{W}$ to be the finite non-prefix-free word set described in Definition 3 in the case of $\mathcal{X} = \{0, 1\}$ and $K = 5$, that is,

$$\mathcal{W} = \{0, 1, 00, 01, 10, 11, \cdots, 11110, 11111\},$$
$$\|\mathcal{W}\| = 2 + 4 + 8 + 16 + 32 = 62.$$

We compute the average of the value of an entropy density rate $-\frac{1}{n}\log P_{X^n}(x^n)$ of some sequences with sufficiently large n actually emitted from the source. Then it is compared with the value of a lower bound (calculated by left hand side (l.h.s.) of Eq. (20)) and an upper bound (calculated by right hand side (r.h.s.) of Eq. (20)).

The property of word-valued source models are dependent on the probability distributions $P_W(w)$ over $\mathcal{W}$. In this experiment, we assume $P(w) > 0 (\neq 0)$ for all $w \in \mathcal{W}$. Here, the word set $\mathcal{W}$ is surely non-prefix-free.

5.2 Experiment I: Convergence of the Entropy Density Rate

It has not been clarified whether the entropy rate of a non-prefix-free word-valued source exists or not. At first we investigate the convergence of the entropy density rate of the non-prefix-free word-valued source.

We fix one specific non-prefix-free word-valued source by giving the randomly generated probability distribution $P_W(w)$ on the word set $\mathcal{W}$. Then we calculate the entropy density rate for the 200 sequences with length of $n = 40000$ actually emitted from the source. Figure 3 shows the convergence process of the average (solid line) and variance (broken line) of the entropy density rate.

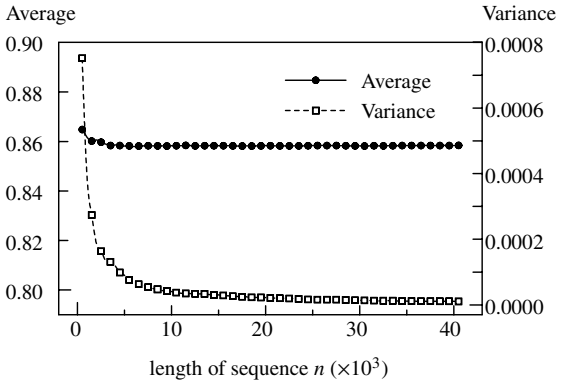


Fig. 3 The convergence process of entropy density rate.

Table 2 The conditions of $P_S(s)$ of word-valued source.

	$P_S(1)$	$P_S(2)$	$P_S(3)$	$P_S(4)$	$P_S(5)$
(i)	0.01	0.04	0.05	0.10	0.80
(ii)	0.50	0.20	0.15	0.10	0.05
(iii)	0.80	0.10	0.05	0.04	0.01

	$H(S)$	$E[\|W\|]$
(i)	1.06	4.64
(ii)	1.92	2.00
(iii)	1.06	1.36

Figure 3 shows that the average of the entropy density rate seems to converge to a constant value as n becomes large, and it may be expected that the entropy rate exists also for the non-prefix-free word-valued source. Furthermore, the variance appears to approach 0 as n becomes large. It is also expected that the AEP holds. Figure 3 is the result of one specific model, however, the similar results were obtained for some other models generated by the same method.

We assume that the entropy rate of the non-prefix-free word-valued source exists in the following experiments. For each source, we calculate the average of the entropy density rate of 200 sequences with length of 40000, and we consider it as the estimator of the entropy rate of the source, denoted by $\hat{H}(\boldsymbol{X})$.

5.3 Experiment II: The Behavior of $\hat{H}(\boldsymbol{X})$ and Its Bounds

As in Experiment I, generating the probability distribution $P_W(w)$ randomly, we compute the $\hat{H}(\boldsymbol{X})$ and its bounds, and then compare them.

In Experiment II, we assume that $P_S(s)$ is fixed and $P_{W|S}(w|s)$ is randomly given. $P_W(w)$ is obtained by the product of $P_S(s)$ and $P_{W|S}(w|s)$. We consider three cases of $P_S(s)$ ($s = 1, 2, \cdots, 5$) as shown in Table 2. 200 word-valued source models are generated randomly in respective cases of (i)–(iii). We calculate $\hat{H}(\boldsymbol{X})$ and its bounds for each model.

Figures 4, 5 and 6 show the calculation results of $\hat{H}(\boldsymbol{X})$ (denoted by "+" as each randomly generated model), the upper bound (denoted by UB) and the lower bound (denoted by LB) under the cases (i), (ii) and (iii). The horizontal axis

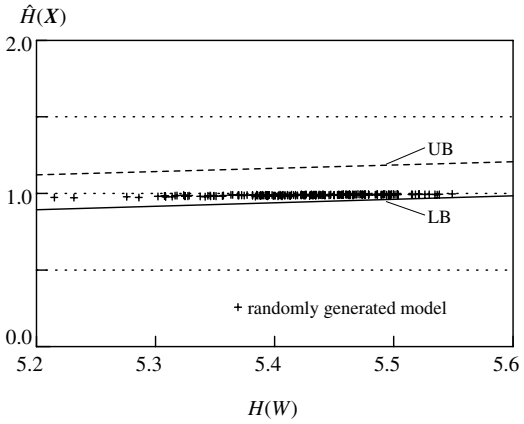


Fig. 4 The result of case (i) in Experiment II.

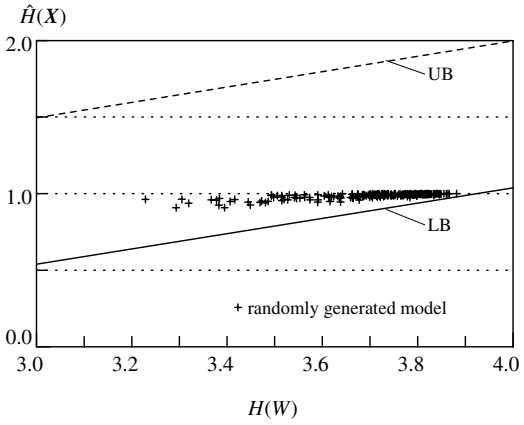


Fig. 5 The result of case (ii) in Experiment II.

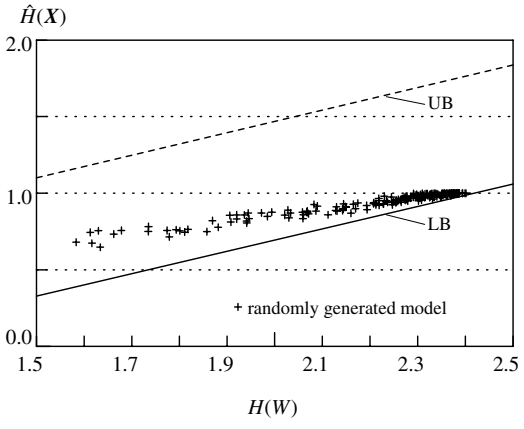


Fig. 6 The result of case (iii) in Experiment II.

represents $H(W)$. Since $P_S(s)$ is fixed in each case, $H(S)$ and $E[|W|]$ are constant. Therefore, the upper bound and the lower bound are given as a straight line with the same gradient in the figures.

From the results, at first, we can see the large difference in the range of value which $H(W)$ takes depending on $P_S(s)$. Deviation of the probability distribution $P_W(w)$ becomes large by giving small probability to words with long length, and giving large probability to words with short length because the number of longer words is more than that of shorter ones. And then, $H(W)$ tends to take smaller values. It corresponds to the case (iii).

In case (i) and (ii), it is found that $H(W)$ takes relatively larger values, and both $\hat{H}(\boldsymbol{X})$ and the lower bound approach very close to 1 which is the trivial upper bound (Figs. 4 and 5). Trivial lower bound ($H(\boldsymbol{X}) \geq 0$) do not have a meaning at all.

In case (iii), in the region where $H(W)$ is smaller, the lower bound is far from 1 and $\hat{H}(\boldsymbol{X})$ tends to take the value in the middle of them (Fig. 6).

From above results, although they are still qualitative, we may say that there is a meaning in the proposed lower bound as the first step to discuss the entropy rate of a non-prefix-free word-valued source.

More detailed numerical experiments under various conditions including the case with $P_W(w) = 0$ for some $w \in \mathcal{W}$ are shown in [11].

Remark 6: Since we assume here that $P(w) > 0$ for all $w \in \mathcal{W}$, Kraft's inequality (Eq. (31)) does not hold. As stated in Remark 3, therefore, it may happen that the value of the upper bound becomes larger than the trivial upper bound $\log \|\mathcal{X}\| = 1$ here. Figure 4, 5 and 6 show that the upper bounds are larger than the trivial bound in all models in this experiment. This is because $H(W)$ probably tends to take a larger value than $E[|W|]$ under the assumption. To the contrary, if $P_W(w) = 0$ for some words w, $H(W)$ tends to get a smaller value. At this time, the upper bound becomes smaller than the trivial bound, and it might be expected that there are many cases that the upper bound is effective. □

6. Concluding Remarks

In this paper, we have newly derived a lower bound on an entropy density rate of a non-prefix-free i.i.d. word-valued source with a finite word set. The lower bound satisfies $\frac{H(Y)}{E[|W|]} - \frac{H(S)}{E[|W|]} \geq 0$. Then we have investigated the behavior of the entropy density rate and its bounds for some case by numerical computation. From the results, we have obtained a suggestion about the effectiveness of the lower bound. Although only qualitative evaluation has been provided this time, we must clarify the effectiveness theoretically in future.

As a future study, we try to derive a tighter bound. Moreover, the existence of the entropy rate and the AEP of a non-prefix-free word-valued source should be theoretically investigated. An analysis of the bounds on the entropy rate for *an ergodic* word-valued source with a non-prefix-free word set is also a future work. Up to now, there is only the discussion that Lemma 1 (upper bound) can be extended to the case where $\boldsymbol{Y}$ is an ergodic source by the present authors [11].

Acknowledgement

The authors would like to thank anonymous reviewers for their constructive comments to clearly prove Theorem 1. The authors also wish to acknowledge Profs. M. Nishiara and H. Morita for their useful comments and introduction to their papers which lead to their study. One of the authors, T. Ishida, wishes to thank Dr. M. Kobayashi for very helpful discussion.

This work was partially supported by Japan Society for the Promotion of Science under Grant-in-Aid for Scientific Research No.18700160 and Waseda University Grant for Special Research Project No.2006K-108.

References

[1] P. Billingsley, Ergodic Theory and Information, John Wiley & Sons, New York, 1965.
[2] L. Breiman, "The individual ergodic theorems of information theory," Ann. Math. Stat., no.28, pp.809–811, 1957.
[3] K.L. Chung, A Course in Probability Theory, Academic Press, New York, 1974.
[4] Y. Ephraim and N. Merhav, "Hidden Markov processes," IEEE Trans. Inf. Theory, vol.48, no.6, pp.1518–1569, 2002.
[5] W. Feller, An Introduction to Probability Theory and Its Applications, vol.1, 3rd ed., John Wiley & Sons, 1968.
[6] M. Goto, T. Matsushima, and S. Hirasawa, "A source model with probability distribution over word set and recurrence time theorem," IEICE Trans. Fundamentals, vol.E86-A, no.10, pp.2517–2525, Oct. 2003.
[7] T.S. Han, Information-Spectrum Methods in Information Theory, Springer-Verlag, 2002.
[8] T.S. Han and K. Kobayashi, Mathematics of Information and Coding (Translations of Mathematical Monographs), Amer Mathematical Society, 2002.
[9] T. Ishida, M. Goto, and S. Hirasawa, "On universality of both Bayes codes and Ziv-Lempel codes for sources which emit data sequence by block unit," IEICE Trans. Fundamentals (Japanese Edition), vol.J84-A, no.9, pp.1167–1178, Sept. 2001.
[10] T. Ishida, M. Goto, and S. Hirasawa, "On universality of LZ78 codes for sources which emit data sequence with word unit," The 24th Symposium on Information Theory and Its Applications (SITA2001), pp.243–246, Sept. 2001.
[11] T. Ishida, M. Goto, T. Matsushima, and S. Hirasawa, "Properties of a word-valued source with a non-prefix-free word set," IEICE Technical Report, IT2003-5, 2003.
[12] T. Matsushima, H. Inazumi, and S. Hirasawa, "A class of distortionless codes designed by Bayes decision theory," IEEE Trans. Inf. Theory, vol.37, no.5, pp.1288–1293, Sept. 1991.
[13] B. McMillan, "The basic theorems of information theory," Ann. Math. Stat., no.24, pp.196–219, 1953.
[14] M. Nishiara and H. Morita, "On the AEP of word-valued sources," IEEE Trans. Inf. Theory, vol.46, no.3, pp.1116–1120, 2000.
[15] C.E. Shannon, "A mathematical theory of communication," Bell Syst. Tech. J., vol.27, pp.379–423, and pp.623–656, 1948.
[16] J. Ziv and A. Lempel, "A universal algorithm for sequential data compression," IEEE Trans. Inf. Theory, vol.23, no.3, pp.337–343, 1977.
[17] J. Ziv and A. Lempel, "Compression of individual sequences via variable-rate coding," IEEE Trans. Inf. Theory, vol.24, no.5, pp.530–536, 1978.

Appendix: Proof of Theorem 1

First, we give the outline of the proof by asymptotical arguments in order to clearly describe the abstract of deriving the lower bound. Next, some definitions and lemmas for the proof of Theorem 1 are prepared. Finally we show the complete proof with strict and careful arguments.

A.1 Outline of the Proof

We also use N_m and M_n as [6], [14] for an i.i.d. source $\boldsymbol{Y}$.

$$N_m \stackrel{\text{def}}{=} \sum_{i=1}^{m} |\phi(Y_i)|, \tag{A·1}$$

$$M_n \stackrel{\text{def}}{=} \min_{m\geq 1}\{m|N_m \geq n\}. \tag{A·2}$$

Here, N_m means the total length of $\phi(Y^m)$, and M_n means the minimum length of Y^m such that $N_m \geq n$. Both N_m and M_n are random variables. The following lemmas have been proved.

Lemma 2 (Nishiara and Morita [14]): For all sample sequences,

$$\lim_{n\to\infty} M_n = \infty, \tag{A·3}$$

holds. □

Lemma 3 (Goto et al. [6]): For an i.i.d. source $\boldsymbol{Y}$[†],

$$\lim_{n\to\infty} \frac{M_n}{n} = \frac{1}{E[|\phi(Y)|]}, \quad a.s. \tag{A·4}$$

holds. □

In order to derive a lower bound on entropy density rate of the non-prefix-free word-valued source, we need to evaluate the appearance probability of a symbol sequence x^n. Considering the property of AEP [15] on the measure of $\boldsymbol{Y}$, we find that actually occurring y^m has asymptotically equal probability when $m \to \infty$. Therefore, we can obtain $P_{X^n}(x^n)$ by counting up the total number of the word sequences $w^* \in \mathcal{W}^*$ (i.e., $y^* \in \mathcal{Y}^*$) mapped to the same symbol sequence x^n, which is equal to the cardinality of $\Gamma_\phi(x^n)$. However, as we mentioned in Sect. 3, $P_{X^n}(x^n)$ has a complicated structure and it depends on a many-to-one mapping from $\mathcal{W}^*$ to $\mathcal{X}^n$. In the outline of the proof, we are concerned only with the typical sequence [8] on $\boldsymbol{Y}$ because non typical y^* does not occur when $m \to \infty$.

Denoting $p^{\max}(x^n)$ as the maximum probability of y^{M_n} such as $y^{M_n} \in \Gamma_\phi(x^n)$ and typical, $P_{Y^{M_n}}(y^{M_n}) \leq p^{\max}(x^n)$ holds for all $y^{M_n} \in \Gamma_\phi(x^n)$ when n is sufficiently large (i.e., M_n is sufficiently large by Lemma 2). We can rewrite Eq. (7) as

[†]In [6], Goto et al. have shown this lemma in the case that $\boldsymbol{Y}$ is an ergodic source.

$$P_{X^n}(X^n) = \sum_{y^{M_n} \in \Gamma_\phi(X^n)} P_{Y^{M_n}}(y^{M_n}) \le p^{\max}(X^n) \cdot \left\|\Gamma_\phi(X^n)\right\|, \quad a.s. \tag{A·5}$$

when $n \to \infty$.

Then, we obtain

$$\liminf_{n\to\infty}\left[-\frac{1}{n}\log P_{X^n}(X^n)\right] \ge \liminf_{n\to\infty}\left[-\frac{1}{n}\log p^{\max}(X^n)\right] + \liminf_{n\to\infty}\left[-\frac{1}{n}\log\left\|\Gamma_\phi(X^n)\right\|\right], \quad a.s. \tag{A·6}$$

In Eq. (A·6), if we can show that the first term of the right hand side (r.h.s.) is equal to $\frac{H(Y)}{E[|W|]}$, and that the second term corresponds to $-\frac{H(S)}{E[|W|]}$, then the proof of the theorem is concluded.

To evaluate the first term, we can use the AEP [15] of sequence Y^m. Because we now consider the typical sequence on the measure of $\boldsymbol{Y}$ and $M_n \to \infty$ when $n \to \infty$,

$$\lim_{n\to\infty}\left[-\frac{1}{M_n}\log p^{\max}(X^n)\right] = H(Y), \quad a.s. \tag{A·7}$$

holds. From Lemma 3, replacing $\phi(Y)$ by W here, we have

$$\liminf_{n\to\infty}\left[-\frac{1}{n}\log p^{\max}(X^n)\right] = \lim_{n\to\infty}\left[-\frac{M_n}{n}\frac{1}{M_n}\log p^{\max}(X^n)\right] = \frac{H(Y)}{E[|W|]}, \quad a.s. \tag{A·8}$$

Next we consider the second term of the r.h.s. of Eq. (A·6). If we can count up the number of word sequences mapped to the same sequence X^n, the second term can be evaluated.

We introduce the idea such that the mapping becomes one-to-one by assuming the gaps between words in x^n are fixed. Determining a certain word sequence w^* from a given symbol sequence x^n is equivalent to separating x^n into words. Each different word sequence with the same separation pattern of words is never mapped to the same symbol sequence x^n. We can evaluate the number of the word sequences mapped to the same symbol sequence x^n by counting up the number of separation patterns of x^n.

Let $N(w|W^m)$ be the number of the word $w \in \mathcal{W}$ which appears in the sequence W^m. We define $L_m(s)$ as the occurrence number of the word w with length s ($s = 1, 2, \cdots, K$) in the word sequence W^m. And $P_S(s)$ is defined as the probability distribution of s. That is,

$$L_m(s) \stackrel{\text{def}}{=} \sum_{w:\ |w|=s} N(w|W^m), \tag{A·9}$$

$$P_S(s) \stackrel{\text{def}}{=} \sum_{w:\ |w|=s} P_W(w), \tag{A·10}$$

where S denotes the random variable of s.

We consider $y^m = y_1 y_2 \cdots y_m$ and $y'^m = y'_1 y'_2 \cdots y'_m$, and also $w^m = \phi(y_1)\phi(y_2)\cdots\phi(y_m)$, $w'^m = \phi(y'_1)\phi(y'_2)\cdots\phi(y'_m)$, $x^n = \phi(y^m)$ and $x'^n = \phi(y'^m)$. When m and $L_m(s)$ are fixed and $n = \sum_{i=1}^m |W_i|$ holds, the following facts are indicated:

1. If $w^m \ne w'^m$ and $|w_i| = |w'_i|$ ($i = 1, 2, \cdots, m$), then $x^n \ne x'^n$ surely holds. That is, each different w^m with the same separation pattern is certainly mapped to the different x^n.
2. Even if there exists i ($i = 1, 2, \cdots, m$) such that $|w_i| \ne |w'_i|$, then, of course $w^m \ne w'^m$, there exist the case that $x^n = x'^n$. That is, it may happen that each w^m with different separation pattern is mapped to the same x^n.

It is found that, therefore, the number of W^m which mapped to a certain X^n is bounded from above by the number of the total separation pattern of X^n into words V_m;

$$V_m \stackrel{\text{def}}{=} \frac{m!}{L_m(1)!L_m(2)!\cdots L_m(K)!}. \tag{A·11}$$

V_m is also a random variable and is equal to the total number of arrangement of the m words which contains $L_m(s)$ words with length s each.

When n is sufficiently large, $M_n \sim \frac{n}{E[|W|]}$ holds[†] from Lemma 3, that is, X^n includes almost $\frac{n}{E[|W|]}$ words. Paying attention to above arguments, although M_n is a random variable, we can obtain

$$\left\|\Gamma_\phi(X^n)\right\| \le V_{M_n}, \quad a.s. = \frac{M_n!}{L_{M_n}(1)!L_{M_n}(2)!\cdots L_{M_n}(K)!}, \tag{A·12}$$

when $n \to \infty$.

From the law of large number [5] and the definition of $L_m(s)$,

$$\lim_{m\to\infty}\frac{L_m(s)}{m} = P_S(s), \quad a.s., \tag{A·13}$$

holds for every s ($s = 1, 2, \cdots, K$). This means that the word sequences with length M_n includes $L_{M_n}(s) \sim M_n P_S(s)$ words of each length s when n is sufficiently large. From $M_n \sim \frac{n}{E[|W|]}$ and $L_{M_n}(s) \sim M_n P_S(s)$, we can rewrite Eq. (A·12) using Stirling's formula [5], $\log n! \sim n\log n$, as

$$\begin{aligned} -\frac{1}{n}\log\left\|\Gamma_\phi(X^n)\right\| &\ge -\frac{1}{n}\log\frac{M_n!}{\prod_{s=1}^K (M_n P_S(s))!} \\ &\sim -\frac{M_n}{n}\left(-\sum_{s=1}^K P_S(s)\log P_S(s)\right) \\ &\sim -\frac{H(S)}{E[|W|]}, \end{aligned} \tag{A·14}$$

when n is sufficiently large. Here,

[†]Here we use the notation "$F(n) \sim G(n)$" to represent "$\lim_{n\to\infty}\frac{F(n)}{G(n)} = 1$."

$$H(S) = -\sum_{s=1}^{K} P_S(s) \log P_S(s). \tag{A· 15}$$

The lower bound on the entropy rate is provided by assigning Eqs. (A· 8) and (A· 14) for Eq. (A· 6).

$$\liminf_{n\to\infty} \left[-\frac{1}{n} \log P_{X^n}(X^n) \right] \geq \frac{H(Y)}{E[|W|]} - \frac{H(S)}{E[|W|])}, \quad a.s. \tag{A· 16}$$

We present the strict proof of Theorem 1 as follows. At first, in next section, we give some definitions and lemmas in (1)–(4) for the preparation for the complete proof.

A.2 Preparation

(1) Typical sequence

As mentioned in the outline of the proof, $N(w|W^m)$ denotes the number of the word $w \in \mathcal{W}$ which appears in the sequence W^m. Because W^m is i.i.d. process and $\mathcal{W}$ is finite in Definition 3, we have the following lemma by the law of iterated logarithm [5].

Lemma 4:

$$P_Y(y) - \delta_{m,w} < \frac{N(w|W^m)}{m} < P_Y(y) + \delta_{m,w}, \quad a.s. \tag{A· 17}$$

when $m \to \infty$. Here,

$$\delta_{m,w} = O\left(\sqrt{\frac{\log\log m}{m}} \right). \tag{A· 18}$$

□

We define $\mathcal{E}_Y^{(m)} \subset \mathcal{Y}^\infty$ as

$$\mathcal{E}_Y^{(m)} \stackrel{\text{def}}{=} \left\{ y^\infty \in \mathcal{Y}^\infty \middle| \left| \frac{N(w|W^m)}{m} - P_W(w) \right| < \delta_{m,w} \right\}. \tag{A· 19}$$

$\mathcal{E}_Y^{(m)}$ represents the event such that $Y^m(y^\infty) = y^m$ is a typical sequence with length m on the measure of $\boldsymbol{Y}$. Furthermore, $\mathcal{E}_Y$ is defined as

$$\mathcal{E}_Y \stackrel{\text{def}}{=} \bigcup_{k=1}^{\infty} \bigcap_{m=k}^{\infty} \mathcal{E}_Y^{(m)}, \tag{A· 20}$$

then Lemma 4 suggests

$$\Pr(\mathcal{E}_Y) = 1. \tag{A· 21}$$

(2) The AEP of Y^m

We consider the sequence Y^m from an i.i.d. source with probability distribution $P_Y(y)$ ($y \in \mathcal{Y}$). For an arbitrary $\varepsilon > 0$, we define

$$\overline{p_{(m,\varepsilon)}} \stackrel{\text{def}}{=} 2^{-m(H(Y)-\varepsilon)}, \tag{A· 22}$$

$$\underline{p_{(m,\varepsilon)}} \stackrel{\text{def}}{=} 2^{-m(H(Y)+\varepsilon)}, \tag{A· 23}$$

Then we have the following lemma.

Lemma 5: For arbitrary small $\varepsilon > 0$

$$\underline{p_{(M_n,\varepsilon)}} < P_{Y^{M_n}}(Y^{M_n}) < \overline{p_{(M_n,\varepsilon)}}, \quad a.s. \tag{A· 24}$$

holds on $\mathcal{E}_Y$ when $n \to \infty$.

Proof: From the definition of $\mathcal{E}_Y$,

$$\lim_{m\to\infty} \frac{N(\phi(y)|W^m)}{m} = P_Y(y), \tag{A· 25}$$

holds for any $y^\infty \in \mathcal{E}_Y$ where we notice that ϕ is one-to-one. Then, we obtain

$$\begin{aligned} &\lim_{m\to\infty} -\frac{1}{m} \log P_{Y^m}(Y^m) \\ &= \lim_{m\to\infty} -\frac{1}{m} \log \prod_{y\in\mathcal{Y}} P_Y(y)^{N(\phi(y)|W^m)} \\ &= \lim_{m\to\infty} \left[-\sum_{y\in\mathcal{Y}} \frac{N(\phi(y)|W^m)}{m} \log P_Y(y) \right] \\ &= H(Y). \end{aligned} \tag{A· 26}$$

Replacing m by M_n on Eq. (A· 26), we have

$$\lim_{n\to\infty} -\frac{1}{M_n} \log P_{Y^{M_n}}(Y^{M_n}) = H(Y), \quad a.s. \tag{A· 27}$$

because $M_n \to \infty$ as $n \to \infty$ holds by Lemma 2. Eq. (A· 27) implies Eq. (A· 24). □

(3) Number of the words included in X^n

Here, we investigate the asymptotic property of M_n defined by Eq. (A· 2). For a given n and arbitrary $\varepsilon' > 0$, we define

$$\overline{m_{(n,\varepsilon')}} \stackrel{\text{def}}{=} n\left(\frac{1}{E[|W|]} + \varepsilon' \right), \tag{A· 28}$$

$$\underline{m_{(n,\varepsilon')}} \stackrel{\text{def}}{=} n\left(\frac{1}{E[|W|]} - \varepsilon' \right), \tag{A· 29}$$

Then we obtain the following lemma.

Lemma 6: For arbitrary small $\varepsilon' > 0$

$$\underline{m_{(n,\varepsilon')}} < M_n < \overline{m_{(n,\varepsilon')}}, \quad a.s. \tag{A· 30}$$

holds on $\mathcal{E}_Y$ when $n \to \infty$.

Proof: It is obviously ensured that Lemma 3 also holds on $\mathcal{E}_Y$ because Eq. (A· 25) holds for any $y^\infty \in \mathcal{E}_Y$ and because $\mathcal{W}$ is finite. See the Appendix A in [6] for the details of the proof. Lemma 3 implies Eq. (A· 30). □

(4) Evaluation of the separation patterns of x^n into words

Next, we investigate the asymptotic property of V_m given by Eq. (A· 11). We consider $L_m(s)$ and $P_S(s)$ given by Eq. (A· 9) and Eq. (A· 10). Summing up the each term of Eq. (A· 17) for the words with same length $s = |w|$,

$$\begin{aligned} &m\left(P_S(s) - \delta'_{m,s} \right) < L_m(s) \\ &\quad < m\left(P_S(s) + \delta'_{m,s} \right), \quad a.s. \end{aligned} \tag{A· 31}$$

holds for each s ($s = 1, 2, \cdots, K$) when $m \to \infty$. In Eq. (A· 31), from the following relationship;

$$\sum_{\{w \in \mathcal{W}:|w|=s\}} \delta_{m,w} < \|\mathcal{X}\|^K \cdot \max_w \delta_{m,w}, \tag{A· 32}$$

we use $\delta'_{m,s}$ such that

$$\delta'_{m,s} = \|\mathcal{X}\|^K \cdot \max_w \delta_{m,w}, \tag{A· 33}$$

Here, $\delta'_{m,s} = O\left(\sqrt{\frac{\log\log m}{m}}\right)$.

From the definition of V_m, it is found that V_m is a random variable. V_m means the total number of different separation patterns of word sequences when m and $L_m(s)$ are fixed. To evaluate V_m, we define $\overline{V_m}$ and $\underline{V_m}$ as follows:

$$\overline{V_m} \stackrel{\text{def}}{=} \frac{m!}{\prod_{s=1}^K \lfloor m\left(P_S(s) - \delta'_{m,s}\right)\rfloor!}, \tag{A· 34}$$

$$\underline{V_m} \stackrel{\text{def}}{=} \frac{m!}{\prod_{s=1}^K \lceil m\left(P_S(s) + \delta'_{m,s}\right)\rceil!}. \tag{A· 35}$$

where $\lfloor z \rfloor$ means the maximum integer smaller than or equal to z, and $\lceil z \rceil$ means the minimum integer larger than or equal to z.

We obtain the following lemma from Lemma 2 and Eq. (A· 31).

Lemma 7:

$$\underline{V_m} < V_m < \overline{V_m}, \quad a.s. \tag{A· 36}$$

holds on $\mathcal{E}_Y$ when $m \to \infty$.

Proof: Eq. (A· 36). By substituting Eq. (A· 31) for the definition of V_m (Eq. (A· 11)), we obtain Eq. (A· 36). □

All definitions and lemmas used in the proof have been prepared above. Finally, we give the complete proof of Theorem 1 based on the above preparations.

A.3 Complete Proof of Theorem 1

Letting $\mathcal{E}_Y^C$ be a complementary event of $\mathcal{E}_Y$, we can rewrite Eq. (7), by using $\Pr(\{\mathcal{E}_Y^C\}) = 0$, as

$$\begin{aligned} &P_{X^n}(x^n) \\ &\quad = \Pr(\{y^\infty \in \mathcal{Y}^\infty | Y^*(y^\infty) = y^* \in \Gamma_\phi(x^n)\}) \\ &\quad = \Pr(\{y^\infty \in \mathcal{E}_Y | Y^*(y^\infty) = y^* \in \Gamma_\phi(x^n)\}) \\ &\qquad + \Pr(\{y^\infty \in \mathcal{E}_Y^C | Y^*(y^\infty) = y^* \in \Gamma_\phi(x^n)\}) \\ &\quad = \sum_{y^* \in \Gamma_\phi(x^n)} \Pr(\{y^\infty \in \mathcal{E}_Y | Y^*(y^\infty) = y^*\}). \end{aligned} \tag{A· 37}$$

We define

$$p^{\max}_{\mathcal{E}_Y}(x^n) \stackrel{\text{def}}{=} \max_{y^* \in \Gamma_\phi(x^n)} \Pr(\{y^\infty \in \mathcal{E}_Y | y^* = Y^*(y^\infty)\}), \tag{A· 38}$$

then,

$$P_{Y^{M_n}}(y^{M_n}) \le p^{\max}_{\mathcal{E}_Y}(x^n), \tag{A· 39}$$

holds for all $y^{M_n} \in \Gamma_\phi(x^n)$. Here, obviously, $P_{Y^{M_n}}(y^{M_n})$ for y^{M_n} which gives $p^{\max}_{\mathcal{E}_Y}(x^n)$ satisfies Eq. (A· 39) with equality. And Lemma 5 holds for such $P_{Y^{M_n}}(y^{M_n})$. Then we have

$$\underline{p_{(M_n,\varepsilon)}} < p^{\max}_{\mathcal{E}_Y}(X^n) < \overline{p_{(M_n,\varepsilon)}}, \quad a.s. \tag{A· 40}$$

when $n \to \infty$.

In Eq. (A· 37), replacing x^n by X^n,

$$\begin{aligned} &P_{X^n}(X^n) \\ &\quad = \sum_{y^* \in \Gamma_\phi(X^n)} \Pr(\{y^\infty \in \mathcal{E}_Y | Y^*(y^\infty) = y^*\}) \\ &\quad \le p^{\max}_{\mathcal{E}_Y}(X^n) \cdot \left\|\Gamma_\phi(X^n)\right\|, \end{aligned} \tag{A· 41}$$

holds.

Consequently, we can obtain

$$\begin{aligned} &\liminf_{n\to\infty}\left[-\frac{1}{n}\log P_{X^n}(X^n)\right] \\ &\quad \ge \liminf_{n\to\infty}\left[-\frac{1}{n}\log p^{\max}_{\mathcal{E}_Y}(X^n)\right] \\ &\qquad + \liminf_{n\to\infty}\left[-\frac{1}{n}\log\left\|\Gamma_\phi(X^n)\right\|\right]. \end{aligned} \tag{A· 42}$$

First, we evaluate the first term of the r.h.s. of Eq. (A· 42). From Eq. (A· 40), we have

$$\begin{aligned} &\liminf_{n\to\infty}\left[-\frac{1}{n}\log p^{\max}_{\mathcal{E}_Y}(X^n)\right] \\ &\quad \ge \liminf_{n\to\infty}\left[-\frac{1}{n}\log \overline{p_{(M_n,\varepsilon)}}\right], \quad a.s. \end{aligned} \tag{A· 43}$$

And from Lemma 3,

$$\begin{aligned} &\liminf_{n\to\infty}\left[-\frac{1}{n}\log \overline{p_{(M_n,\varepsilon)}}\right] \\ &\quad = \liminf_{n\to\infty}\left[-\frac{1}{n}\log(2^{-M_n(H(Y)-\varepsilon)})\right] \\ &\quad = \liminf_{n\to\infty}\left[\frac{M_n}{n}(H(Y)-\varepsilon)\right] \\ &\quad = \frac{H(Y)}{E[|W|]} - \frac{\varepsilon}{E[|W|]} \quad a.s. \end{aligned} \tag{A· 44}$$

holds. Then from Eq. (A· 43) and Eq. (A· 44), we have

$$\begin{aligned} &\liminf_{n\to\infty}\left[-\frac{1}{n}\log p^{\max}_{\mathcal{E}_Y}(X^n)\right] \\ &\quad \ge \frac{H(Y)}{E[|W|]} - \frac{\varepsilon}{E[|W|]} \quad a.s. \end{aligned} \tag{A· 45}$$

Next, we evaluate the second term of the r.h.s. in Eq. (A· 42). For fixed m and $L_{M_n}(s)$ such that $m = \sum_{s=1}^K L_{M_n}(s)$, as mentioned in outline, the number of the sequence W^m which is mapped to a certain X^n is upper bounded by V_{M_n}. Because, for given n, M_n and $L_{M_n}(s)$ are also random variables, we need to pay attention in the following arguments.

Now we decompose $\Gamma_\phi(X^n)$ by the length m as follows:

$$\Gamma_\phi(X^n) = \bigcup_m \left\{\Gamma_\phi(X^n) \cap \mathcal{Y}^m\right\}. \tag{A·46}$$

Because of Lemma 7 and $\lim_{n\to\infty} M_n = \infty$, we have

$$\|\Gamma_\phi(X^n) \cap \mathcal{Y}^m\| \le \overline{V_m}, \ \ a.s. \tag{A·47}$$

when $n \to \infty$ for any $y^\infty \in \mathcal{E}_Y$ and $m \ge 1$. Considering the set

$$\mathcal{M}^n_{\varepsilon'} \stackrel{\text{def}}{=} \left\{ m \ \middle|\ \underline{m_{(n,\varepsilon')}} < m < \overline{m_{(n,\varepsilon')}} \right\}, \tag{A·48}$$

we have

$$\begin{aligned}\|\Gamma_\phi(X^n)\| &= \sum_m \|\Gamma_\phi(X^n) \cap \mathcal{Y}^m\| \\ &= \sum_{m\in\mathcal{M}^n_{\varepsilon'}} \|\Gamma_\phi(X^n) \cap \mathcal{Y}^m\| \\ &\quad + \sum_{m\notin\mathcal{M}^n_{\varepsilon'}} \|\Gamma_\phi(X^n) \cap \mathcal{Y}^m\|. \end{aligned}\tag{A·49}$$

Noticing Lemma 6, we find that the first term of the r.h.s. of Eq. (A·49) satisfies

$$\sum_{m\in\mathcal{M}^n_{\varepsilon'}} \|\Gamma_\phi(X^n) \cap \mathcal{Y}^m\| \le \sum_{m\in\mathcal{M}^n_{\varepsilon'}} \overline{V_m}, \ \ a.s. \tag{A·50}$$

when $n \to \infty$ from Eq. (A·47). And the second term of it satisfies

$$\lim_{n\to\infty} \sum_{m\notin\mathcal{M}^n_{\varepsilon'}} \|\Gamma_\phi(X^n) \cap \mathcal{Y}^m\| = 0, \ \ a.s. \tag{A·51}$$

We have therefore

$$\|\Gamma_\phi(X^n)\| \le \sum_{m\in\mathcal{M}^n_{\varepsilon'}} \overline{V_m}, \ \ a.s. \tag{A·52}$$

when $n \to \infty$. Defining $\tilde{M}_n$ for a given n as

$$\tilde{M}_n = \operatorname*{argmax}_{m\in\mathcal{M}^n_{\varepsilon'}} \overline{V_m}, \tag{A·53}$$

the following is obviously satisfied:

$$\lim_{n\to\infty} \tilde{M}_n = \infty. \tag{A·54}$$

Because $\overline{V_m} \le \overline{V_{\tilde{M}_n}}$ holds for all $m \in \mathcal{M}^n_{\varepsilon'}$, we have

$$\sum_{m\in\mathcal{M}^n_{\varepsilon'}} \overline{V_m} \le \|\mathcal{M}^n_{\varepsilon'}\| \cdot \overline{V_{\tilde{M}_n}}. \tag{A·55}$$

Consequently, from Eq. (A·52) and Eq. (A·55),

$$\begin{aligned}&\liminf_{n\to\infty}\left[-\frac{1}{n}\log\left\|\Gamma_\phi(X^n)\right\|\right] \\ &\ge \liminf_{n\to\infty}\left[-\frac{1}{n}\log\left(\|\mathcal{M}^n_{\varepsilon'}\| \cdot \overline{V_{\tilde{M}_n}}\right)\right], \ \ a.s. \\ &\ge \liminf_{n\to\infty}\left[-\frac{1}{n}\log\left((2n\varepsilon'+1) \cdot \overline{V_{\tilde{M}_n}}\right)\right], \ \ a.s. \\ &= \liminf_{n\to\infty}\left[-\frac{1}{n}\log(2n\varepsilon'+1) - \frac{1}{n}\log\overline{V_{\tilde{M}_n}}\right], \ \ a.s. \\ &= \liminf_{n\to\infty}\left[-\frac{1}{n}\log\overline{V_{\tilde{M}_n}}\right], \ \ a.s. \end{aligned}\tag{A·56}$$

From the definition of $\overline{V_{M_n}}$, we have

$$-\frac{1}{n}\log\overline{V_{\tilde{M}_n}} = -\frac{1}{n}\log\frac{\tilde{M}_n!}{\prod_{s=1}^K \lfloor \tilde{M}_n\left(P_S(s) - \delta_{\tilde{M}_n,s}\right)\rfloor!}. \tag{A·57}$$

Here we introduce a parameter τ such that $\lfloor z \rfloor = \tau z$ in order to rewrite Eq. (A·57). $\tau_{\tilde{M}_n,s}$ satisfies

$$\begin{aligned}&\lfloor \tilde{M}_n\left(P_S(s) - \delta_{\tilde{M}_n,s}\right)\rfloor \\ &\quad = \tau_{\tilde{M}_n,s} \cdot \tilde{M}_n\left(P_S(s) - \delta_{\tilde{M}_n,s}\right). \end{aligned}\tag{A·58}$$

Because

$$0 \le \tilde{M}_n\left(P_S(s) - \delta_{\tilde{M}_n,s}\right) - \lfloor \tilde{M}_n\left(P_S(s) - \delta_{\tilde{M}_n,s}\right)\rfloor < 1, \tag{A·59}$$

we can obtain

$$1 - \frac{1}{\tilde{M}_n\left(P_S(s) - \delta_{\tilde{M}_n,s}\right)} < \tau_{\tilde{M}_n,s} \le 1, \tag{A·60}$$

then $\tau_{\tilde{M}_n,s} \to 1$ when $n \to \infty$.

From Stirling's formula [5],

$$m! = \sqrt{2\pi m}\left(\frac{m}{e}\right)^m e^{\theta_m}, \tag{A·61}$$

where θ_m is the term which satisfies $\theta_m \to 0$ when $m \to \infty$, we have

$$\begin{aligned}&-\frac{1}{n}\log\frac{\tilde{M}_n!}{\prod_{s=1}^K \lfloor \tilde{M}_n\left(P_S(s) - \delta_{\tilde{M}_n,s}\right)\rfloor!} \\ &\quad = -\frac{1}{n}\log\frac{\tilde{M}_n!}{\prod_{s=1}^K \left\{\tau_{\tilde{M}_n,s} \cdot \tilde{M}_n\left(P_S(s) - \delta_{\tilde{M}_n,s}\right)\right\}!} \\ &\quad = \frac{\tilde{M}_n}{n}\Big(\sum_{s=1}^K \tau_{\tilde{M}_n,s}(P_S(s) - \delta_{\tilde{M}_n,s}) \\ &\qquad \cdot \log\tau_{\tilde{M}_n,s}(P_S(s) - \delta_{\tilde{M}_n,s})\Big) + \gamma_{\tilde{M}_n}, \end{aligned}\tag{A·62}$$

where

$$\begin{aligned}\gamma_{\tilde{M}_n} &= \frac{(K-1)\log\sqrt{2\pi\tilde{M}_n}}{n} \\ &\quad + \frac{\left(\sum_{s=1}^K \theta_{\tau_{\tilde{M}_n,s}\tilde{M}_n(P_S(s)-\delta_{\tilde{M}_n,s})} - \theta_{\tilde{M}_n}\right)\log e}{n} \\ &\quad + \frac{\sum_{s=1}^K \log\sqrt{\tau_{\tilde{M}_n,s}(P_S(s) - \delta_{\tilde{M}_n,s})}}{n} \\ &\quad - \frac{\tilde{M}_n}{n}\left(\sum_{s=1}^K \tau_{\tilde{M}_n,s}\delta_{\tilde{M}_n,s}\right)\log\frac{\tilde{M}_n}{e} \\ &\quad - \frac{\tilde{M}_n}{n}\left(1 - \sum_{s=1}^K \tau_{\tilde{M}_n,s}P_S(s)\right)\log\frac{\tilde{M}_n}{e}. \end{aligned}\tag{A·63}$$

Considering $\tilde{M}_n \to \infty, \theta_{M_n} \to 0, \frac{M_n}{n} \to \frac{1}{E[|W|]}$ $a.s., \tau_{\tilde{M}_n,s} \to 1$ and $\delta_{M_n,s} = O\left(\sqrt{\frac{\log\log M_n}{M_n}}\right)$ when $n \to \infty$, we can obtain

$\gamma_{\tilde{M}_n} \to 0$, a.s. when $n \to \infty$. Here, the last term in Eq. (A· 63) $\to 0$ holds because, rewriting Eq. (A· 60), we have,

$$0 \le 1 - \sum_{s=1}^{K} \tau_{\tilde{M}_n,s} P_S(s) < \sum_{s=1}^{K} \frac{P_S(s)}{\tilde{M}_n(P_S(s) - \delta_{\tilde{M}_n,s})}, \tag{A· 64}$$

and then

$$\begin{aligned} 0 &\le \frac{\tilde{M}_n}{n}\left(1 - \sum_{s=1}^{K} \tau_{\tilde{M}_n,s} P_S(s)\right) \log \frac{\tilde{M}_n}{e} \\ &< O\left(\frac{\log \tilde{M}_n}{\sqrt{\tilde{M}_n \log\log \tilde{M}_n}}\right). \end{aligned} \tag{A· 65}$$

From Lemma 6, $\frac{\tilde{M}_n}{n} < \frac{1}{E[|W|]} + \varepsilon'$, $a.s.$ is satisfied when $n \to \infty$. Then we obtain by rewriting Eq. (A· 62),

$$\begin{aligned} &\liminf_{n\to\infty}\left[-\frac{1}{n}\log \frac{\tilde{M}_n!}{\prod_{s=1}^{K} \lfloor \tilde{M}_n\left(P_S(s) - \delta_{\tilde{M}_n,s}\right)\rfloor!}\right] \\ &= \liminf_{n\to\infty}\Big[\frac{\tilde{M}_n}{n}\Big(\sum_{s=1}^{K} \tau_{\tilde{M}_n,s}(P_S(s) - \delta_{\tilde{M}_n,s}) \\ &\quad \cdot \log(P_S(s) - \delta_{\tilde{M}_n,s})\Big)\Big], \quad a.s. \\ &> -\frac{H(S)}{E[|W|]} - \varepsilon' H(S), \quad a.s. \end{aligned} \tag{A· 66}$$

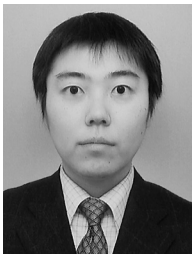
From Eq. (A· 56), Eq. (A· 57) and Eq. (A· 66), we can complete the evaluation of the second term of the r.h.s. in Eq. (A· 42).

$$\begin{aligned} &\liminf_{n\to\infty}\left[-\frac{1}{n}\log \left\|\Gamma_\phi(X^n)\right\|\right] \\ &\ge -\frac{H(S)}{E[|W|]} - \varepsilon' H(S), \quad a.s. \end{aligned} \tag{A· 67}$$

Finally we complete the proof of Theorem 1. From Eq. (A· 45) and Eq. (A· 67), we have

$$\begin{aligned} &\liminf_{n\to\infty}\left[-\frac{1}{n}\log P_{X^n}(X^n)\right] \\ &\ge \frac{H(Y)}{E[|W|]} - \frac{H(S)}{E[|W|]} - \varepsilon'', \quad a.s. \end{aligned} \tag{A· 68}$$

where $\varepsilon'' = \frac{\varepsilon}{E[|W|]} + \varepsilon' H(S)$. Because ε and ε' are arbitrarily small, we can set ε'' be an arbitrarily small positive integer. Consequently the proof of Theorem 1 was provided. □



Takashi Ishida was born in Tokyo, Japan, on Nov. 30, 1975. He received the B.E. degree and M.E. degree in Industrial and Management Systems Engineering from Waseda University, Tokyo, Japan, in 1999 and 2001, respectively. Since 2005, he has been a research associate of Industrial and Management Systems Engineering, Waseda University. His research interests are source coding and statistics. He is a member of IEEE, and the Society of Information Theory and Its Applications.

Masayuki Goto was born in Tokyo, Japan, on Jan. 1, 1969. He received the B.E. and M.E. degrees from Musashi Institute of Technology, Tokyo, Japan, in 1992 and 1994, respectively. From 1997 to 1999, he was a research associate of School of Science and Engineering, Waseda University, Tokyo, Japan. He received the Dr.E. degree in Industrial Engineering and Management from Waseda University. From 2000 to 2002, he was a research associate of School of Engineering, The University of Tokyo, Tokyo, Japan. He is now an associate professor of Faculty of Environmental and Information Studies, Musashi Institute of Technology, Yokohama, Japan. His research interests include information theory, model selection, machine learning theory, Bayesian statistics, industrial information engineering, and business model. He is a member of IEEE, the Society of Information Theory and Its Applications, the Japan Industrial Management Association, the Japan Society for Artificial Intelligence, the Business Model Association, and the Operations Research Society of Japan.

Toshiyasu Matsushima was born in Tokyo, Japan, on Nov. 26, 1955. He received the B.E. degree, M.E. degree and Dr.E. degree in Industrial and Management Systems Engineering from Waseda University, Tokyo, Japan, in 1978, 1980, and 1991, respectively. From 1980 to 1986, he was with Nippon Electric Corporation, Kanagawa, Japan. From 1986 to 1992, he was a lecturer at Department of Management Information, Yokohama College of Commerce. From 1993, he was an associate professor and since 1996 has been a professor of School of Science and Engineering, Waseda University, Tokyo, Japan. From 2001 to 2002, he was a Visiting Professor of the University of Hawaii at Manoa, U.S. From 2005, he has been the Chairperson of the Technical Group on Information Theory of IEICE. His research interests are information theory and its application, statistics, and artificial intelligence. He is a member of the IEEE, the Society of Information Theory and its Applications, the Japan Society for Quality Control, the Japan Industrial Management Association, and the Japan Society of Artificial Intelligence.

Shigeichi Hirasawa was born in Kobe, Japan, on Oct. 2, 1938. He received the B.S. degree in mathematics and the B.E. degree in electrical communication engineering from Waseda University, Tokyo, Japan, in 1961 and 1963, respectively, and the Dr.E. degree in electrical communication engineering from Osaka University, Osaka, Japan, in 1975. From 1963 to 1981, he was with the Mitsubishi Electric Corporation, Hyogo, Japan. Since 1981, he has been a professor of the School of Science and Engineering, Waseda University, Tokyo, Japan. In 1979, he was a Visiting Scholar in the Computer Science Department at the University of California, Los Angeles (CSD, UCLA), CA. He was a Visiting Researcher at the Hungarian Academy of Science, Hungary, in 1985, and at the University of Trieste, Italy, in 1986. In 2002, he was also a Visiting Faculty at CSD, UCLA. From 1987 to 1989, he was the Chairman of the Technical Group on Information Theory of IEICE. He received the 1993 Achievement Award and the 1993 Kobayashi-Memorial Achievement Award from IEICE. In 1996, he was the President of the Society of Information Theory and Its Applications (Soc. of ITA). His research interests are information theory and its applications, and information processing systems. He is an IEEE Fellow, and a member of Soc. of ITA, the Information Processing Society of Japan, and the Japan Industrial Management Association.