Relationship among Complexities of Individual Sequences over Countable Alphabet

Shigeaki KUZUOKA and Tomohiko UYEMATSU

Dept. of Communications and Integrated Systems
Tokyo Institute of Technology
Ookayama, Meguro-ku, Tokyo 152-8552, Japan
Phone:+81-3-5734-3243, Fax:+81-3-5734-2905
Email: {kuzuoka,uematsu}@it.ss.titech.ac.jp

Abstract

This paper investigates some relations among four complexities of sequence over countably infinite alphabet, and shows that two kinds of empirical entropies and the self-entropy regarding a finite state source are asymptotically equal and lower bounded by the muximun number of phrases in distinct parsing of the sequence. Some connections with source coding theorems are also investigated. Further, we consider the empirical entropies with fidelity criterion.

1. INTRODUCTION

In information theory, various properties of sequences over finite alphabet have been investigated. These researches of sequences are not only of theoretic interest, but also have many applications such as universal source coding, hypothesis testing, and prediction (see, e.g., [1], [2] and subsequent papers). Especially, following four complexities of sequences play important roles in variable-rate universal source coding problem: (i) the overlapping empirical entropy $\hat{H}_k(x_1^n)$ of x_1^n gives asymptotically the optimal compression rate attainable by finite-state information-lossless (IL) encoders, (ii) the non-overlapping empirical entropy $\tilde{H}_k(x_1^n)$ of x_1^n gives a lower bound of the compression rate attainable by k-block encoders, (iii) the self-entropy $-(1/n)\log\mu(x_1^n)$ of x_1^n regarding a finite state source characterized by the measure μ is the codeword length of x_1^n associated with the Shannon code which minimizes the average codeword length, and (iv) the muximun number $c(x_1^n)$ of phrases in arbitrary distinct parsing of x_1^n appears in a lower bound of the compression rate attainable by IL encoders.

On the other hand, only little is known about sequences over countably infinite alphabet. In this paper, we dealt with sequences over countably infinite alphabet and investigate the relations among the aforementioned four quantities. Our main result shows that two kinds of empirical entropies and the infinum of the self-entropy over finite state sources are asymptotically equal, and are asymptotically lower bounded by $(\log n/n)c(x_1^n)$, provided that given sequence x is finitely encodable. Some connections with source coding theorems are also investigated. In addition to main results, we consider the empirical entropies with fidelity criterion. Although it is known that the non-overlapping empirical entropy with fidelity criterion plays an important role in the lossy coding of individual sequences [3], the overlapping empirical entropy with fidelity criterion has not been investigated so far. We show that two kinds of empirical entropies with fidelity criterion are asymptotically equal.

2. Main Results

Let $\mathcal{X}$ be a countably infinite alphabet and $\mathcal{X}^\infty$ be a set of infinite sequences over $\mathcal{X}$. Let x_n^m denote the subsequence $x_n x_{n+1}\cdots x_m$ of a sequence $x = x_1x_2\cdots\in\mathcal{X}^\infty$. In this paper, we consider the following four quantities $\hat{H}(x)$, $\tilde{H}(x)$, $h_{fs}(x)$ and $C(x)$ of x.

Definition 1 (Overlapping empirical entropy) *The overlapping empirical distribution $p_k(\cdot|x_1^n)$ of x is defined as*

$$p_k(a_1^k|x_1^n) \triangleq \frac{|\{i : 0\le i\le n-k, x_{i+1}^{i+k} = a_1^k\}|}{n-k+1},$$

for each $a_1^k\in\mathcal{X}^k$, where $|\cdot|$ denotes the cardinality of the set. Let

$$\hat{H}_k(x_1^n) \triangleq -\frac{1}{k}\sum_{a_1^k\in\mathcal{X}^k} p_k(a_1^k|x_1^n)\log p_k(a_1^k|x_1^n),$$

and

$$\hat{H}_k(x) \triangleq \limsup_{n\to\infty}\hat{H}_k(x_1^n).$$

Then, the overlapping empirical entropy $\hat{H}(x)$ of $x\in\mathcal{X}^\infty$ is defined as $\hat{H}(x)\triangleq\inf_k\hat{H}_k(x)$.

Definition 2 (Non-overlapping empirical entropy) *The non-overlapping empirical distribution is defined as*

$$q_k(a_1^k|x_1^n) \triangleq \frac{|\{i : 0 \le i < m, x_{ki+1}^{ki+k} = a_1^k\}|}{m},$$

for each $a_1^k \in \mathcal{X}^k$*, and* $n = mk + r$ $(0 \le r < k)$*. Let*

$$\tilde{H}_k(x_1^n) \triangleq -\frac{1}{k} \sum_{a_1^k \in \mathcal{X}^k} q_k(a_1^k|x_1^n) \log q_k(a_1^k|x_1^n),$$

and

$$\tilde{H}_k(x) \triangleq \limsup_{n\to\infty} \tilde{H}_k(x_1^n).$$

Then, the non-overlapping empirical entropy $\tilde{H}(x)$ *of* $x \in \mathcal{X}^\infty$ *is defined as* $\tilde{H}(x) \triangleq \inf_k \tilde{H}_k(x)$.

Definition 3 *For all* $x \in \mathcal{X}^\infty$*, let*

$$h_{fs}(x) \triangleq \lim_{S\to\infty} \inf_{\mu \in \mathcal{P}_{fs}(S)} \limsup_{n\to\infty} \min_{s_0 \in \mathcal{S}} -\frac{1}{n} \log \mu(x_1^n|s_0),$$

where $\mathcal{P}_{fs}(S)$ *denotes the set of finite state sources with the number of states* S *over the alphabet* $\mathcal{X}$*, and* $\mathcal{S}$ *denotes the state set of* μ.

Definition 4 *For all* $x \in \mathcal{X}^\infty$*, let*

$$C(x) \triangleq \limsup_{n\to\infty} \frac{\log n}{n} c(x_1^n),$$

where $c(x_1^n)$ *denotes the maximum number of phrases in arbitrary distinct parsing of* x_1^n.

As stated in the introduction, these four quantities have a connection with source coding problem of individual sequences. For convenience, we consider length functions instead of coding procedures. For any set A, a non-negative real function σ on A is called a *length function* on A, if σ satisfies the Kraft inequality $\sum_{a\in A} \exp\{-\sigma(a)\} \le 1$. It is known that the codeword lengths of any uniquely decodable code must satisfy the Kraft inequality. Conversely the Kraft inequality is a sufficient condition for the existence of a codeword set with the specified set of codeword lengths (see, for example, [4]).

Before stating our main result, we define finitely encodable sequences.

Definition 5 $x \in \mathcal{X}^\infty$ *is finitely encodable (f.e.) if there exists a length function* σ *on* $\mathcal{X}$ *such that*

$$\lim_{u\to\infty} \sup_n \sum_{a:\sigma(a)\ge u} p_1(a|x_1^n)\sigma(a) = 0. \quad (1)$$

It should be pointed out that if $x \in \mathcal{X}^\infty$ is f.e., then $\hat{H}_1(x) < \infty$ and $\tilde{H}_1(x) < \infty$, and thus,

$$\hat{H}(x) = \inf_k \hat{H}_k(x) < \infty, \text{ and } \tilde{H}(x) = \inf_k \tilde{H}_k(x) < \infty.$$

Here, we state our main result.

Theorem 1 *For any f.e. sequence* $x \in \mathcal{X}^\infty$*, we have*

$$\hat{H}(x) = \lim_{k\to\infty} \hat{H}_k(x), \quad \tilde{H}(x) = \lim_{k\to\infty} \tilde{H}_k(x). \quad (2)$$

and

$$C(x) \le h_{fs}(x) = \hat{H}(x) = \tilde{H}(x). \quad (3)$$

Remark 1 *Since every sequence* x *over finite alphabet* $\mathcal{X}$ *is f.e., Theorem 1 contains finite alphabet case. In finite alphabet case,* $C(x) \le \hat{H}(x)$*,* $C(x) \le h_{fs}(x)$ *and* $h_{fs}(x) = \hat{H}(x)$ *has been known ([2], [5] and [6]). To the best of authors' knowledge,* $\hat{H}(x) = \tilde{H}(x)$ *has not been shown even if* $|\mathcal{X}| < \infty$.

Theorem 1 can be proved by the following lemmas 2 - 5.

Lemma 1 *For any finite state source* μ *with state set* $\mathcal{S}$*, any* $x_1^n \in \mathcal{X}^n$ *and arbitrary parsing* $x_1^n = w_1 w_2 \cdots w_t$ *of* x_1^n $(w_i \in \mathcal{X}^*)$*, we have*

$$\sum_{i=1}^t \log \frac{t}{\eta(w_i)} \le \min_{s_0\in\mathcal{S}} \{-\log \mu(x_1^n|s_0)\} + n\delta_n, \quad (4)$$

where $\eta(w_i) \triangleq |\{j : 1 \le j \le t, w_j = w_i\}|$*, and* $\delta_n = (t/n)\log(n/t) + (t/n)\log(eS)$ $(S = |\mathcal{S}|)$ *and thus,* $\delta_n \to 0$ *as* $n \to \infty$ *provided that* $t/n \to 0$.

Proof: For $l = 1, 2, \cdots$, let $P^{(l)} : \mathcal{X} \to [0, 1]$ be

$$P^{(l)}(x_1^l) = \max_{s_0\in\mathcal{S}} \mu(x_1^l|s_0), \quad \forall x_1^l \in \mathcal{X}^l.$$

Then, $\{P^{(l)}\}_{l=1}^\infty$ satisfies

$$1 \le \sum_{x_1^l\in\mathcal{X}^l} P^{(l)}(x_1^l) \le S, \quad l = 1, 2, \cdots,$$

and for all l, m and $x_1^{l+m} \in \mathcal{X}^{l+m}$,

$$P^{(l+m)}(x_1^{l+m}) \le P^{(l)}(x_1^l) P^{(m)}(x_{l+1}^{l+m}),$$

For $\{P^{(l)}\}_{l=1}^\infty$, we can define the sequence $\{\beta_l\}_{l=1}^\infty$ such that (i) $1 \le \beta_l \le S$, and (ii) $\bar{P}^{(l)}(x_1^l) \triangleq P^{(l)}(x_1^n)/\beta_l$ is the probability distribution on $\mathcal{X}^l$ for all l.

Fix $x_1^n \in \mathcal{X}^n$ and assume that x_1^n is parsed as $x_1^n = w_1 \cdots w_t$. Let l_i be the length $\|w_i\|$ of the i-th word w_i, and $t_l = |\{i : l_i = l\}|$. Then, we have

$$\begin{aligned} -\log P^{(n)}(x_1^n) &\geq \sum_{i=1}^{t}\{-\log P^{(l_i)}(w_i)\} \\ &= \sum_{i=1}^{t}\{-\log \bar{P}^{(l_i)}(w_i)\} - \sum_{i=1}^{t}\log \beta_{l_i} \\ &\geq \sum_{l}\sum_{i:l_i=l}\{-\log \bar{P}^{(l)}(w_i)\} - t\log S \\ &\overset{(a)}{\geq} \sum_{l}\sum_{i:l_i=l}\left\{\log \frac{t_l}{\eta(w_i)}\right\} - t\log S \\ &= \sum_{i=1}^{t}\log \frac{t}{\eta(w_i)} - \sum_{l} t_l \log \frac{t}{t_l} - t\log S \\ &\overset{(b)}{\geq} \sum_{i=1}^{t}\log \frac{t}{\eta(w_i)} - t\log \frac{n}{t} - t\log(eS). \end{aligned}$$

The inequality (a) follows from that

$$\begin{aligned} &\sum_{l}\sum_{i:l_i=l}\{-\log \bar{P}^{(l)}(w_i)\} - \sum_{l}\sum_{i:l_i=l}\left\{\log \frac{t_l}{\eta(w_i)}\right\} \\ =\ &\sum_{l} t_l D(Q^{(l)}\|\bar{P}^{(l)}), \end{aligned}$$

where $Q^{(l)}(w_i) \triangleq \eta(w_i)/t_l$ is the probability distribution on $\mathcal{X}^l$. The inequality (b) follows from that

$$\sum_{l} t_l \log \frac{t}{t_l} \leq t\{\log(EL) + \log(e)\} = t\log \frac{n}{t} + t\log(e),$$

where L is the random variable determined by the distribution $\Pr\{L = l\} = t_l/t$. □

Lemma 2 *For any f.e. sequence $x \in \mathcal{X}^\infty$ and any $\epsilon > 0$, there is an integer $N = N(x,\epsilon)$ such that*

$$c(x_1^n) \leq \frac{n}{\log n}\left(h_{fs}(x) + \epsilon\right), \quad \forall n \geq N.$$

Proof: Since x is f.e., there exists a length function σ on $\mathcal{X}$ which satisfies

$$\limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}\sigma(x_i) < \infty. \tag{5}$$

By [7, Lemma 1], we have

$$\frac{c(x_1^n)}{n} \leq \frac{(1/n)\sum_{i=1}^{n}\sigma(x_i)}{\log(c(x_1^n)/4)} \to 0. \tag{6}$$

On the other hand, for any distinct parsing $x_1^n = w_1 \cdots w_c$,

$$\sum_{i=1}^{t}\log \frac{t}{\eta(w_i)} = c\log c = c\log n - c\log(n/c).$$

By Lemma 1 and (6), we have

$$c\log n \leq \min_{s_0\in\mathcal{S}}\{-\log \mu(x_1^n|s_0)\} + n(\epsilon/2),$$

for any finite state source μ and sufficiently large n. □

Lemma 3 *For any $x_1^n \in \mathcal{X}^n$, any integer S and any $\gamma > 0$, there exists an integer $K = K(x,S,\gamma)$ such that for all $k \geq K$,*

$$\tilde{H}_k(x) \leq \inf_{\mu\in\mathcal{P}_{fs}(S)} \limsup_{n\to\infty} \min_{s_0\in\mathcal{S}}\left\{-\frac{1}{n}\log \mu(x_1^n|s_0)\right\} + \gamma.$$

Proof: Apply Lemma 1 to the parsing $x_1^n = w_1 \cdots w_t$ of x_1^n such that $\|w_i\| = k$ for each $i = 1,\cdots,t$. Then, the right hand side of (4) equals to $\tilde{H}_k(x_1^n)$. □

Lemma 4 *For any f.e. sequence $x_1^n \in \mathcal{X}^n$, any integer S and any $\gamma > 0$, there exists an integer $K = K(x,S,\gamma)$ such that for all $k \geq K$,*

$$\inf_{\mu\in\mathcal{P}_{fs}(S)} \limsup_{n\to\infty} \min_{s_0\in\mathcal{S}}\left\{-\frac{1}{n}\log \mu(x_1^n|s_0)\right\} \leq \hat{H}_k(x) + \gamma.$$

Proof: This lemma can be proved by a similar manner as [6, proof of Theorem 1] in which finite alphabet sequences are considered. What we have to prove is

$$\left|H_k^\circ(x_1^n) - \hat{H}_k(x_1^n)\right| \to 0 \ (n \to \infty), \tag{7}$$

where

$$\begin{aligned} H_k^\circ(x_1^n) &\triangleq -\frac{1}{k}\sum_{a_1^k\in\mathcal{X}^k} p_k^\circ(a_1^k|x_1^n)\log p_k^\circ(a_1^k|x_1^n), \\ p_k^\circ(a_1^k|x_1^n) &\triangleq \frac{|\{i : 0 \leq i < n, \breve{x}_{i+1}^{i+k} = a_1^k\}|}{n}, \end{aligned}$$

and $\breve{x}_1^n = x_1^{n+k-1}$ is defined as $\breve{x}_1^n = x_1^n$ and $\breve{x}_{n+1}^{n+k-1} = x_1^{k-1}$. By the definition of p_k°, there exists a subset B_k of $\mathcal{X}^k$ such that $|B_k| \leq k$ and

$$p_k^\circ(a_1^k|x_1^n) = \frac{n-k+1}{n}p_k(a_1^k|x_1^n), \quad \forall a_1^k \notin B_k,$$

and thus, for $a_1^k \notin B_k$,

$$\begin{aligned} &-\sum_{a_1^k\notin B_k} p_k^\circ(a_1^k|x_1^n)\log p_k^\circ(a_1^k|x_1^n) \\ &= -\frac{n-k+1}{n}\sum_{a_1^k\notin B_k} p_k(a_1^k|x_1^n)\log p_k(a_1^k|x_1^n) \\ &\quad -\frac{n-k+1}{n}\sum_{a_1^k\notin B_k} p_k(a_1^k|x_1^n)\log \frac{n-k+1}{n}. \end{aligned}$$

On the other hand, since

$$|p_k^\circ(a_1^k|x_1^n) - p_k(a_1^k|x_1^n)| \le \frac{2k}{n-k+1}, \quad \forall a_1^k \in \mathcal{X}^k,$$

we have

$$\begin{aligned}&\Bigg| -\frac{1}{k}\sum_{a_1^k\in B_k} p_k^\circ(a_1^k|x_1^n)\log p_k^\circ(a_1^k|x_1^n) \\ &\qquad +\frac{1}{k}\sum_{a_1^k\in B_k} p_k(a_1^k|x_1^n)\log p_k(a_1^k|x_1^n)\Bigg| \\ &\le J_{n,k}\log|B_k| - J_{n,k}\log J_{n,k} \\ &\le J_{n,k}\log k - J_{n,k}\log J_{n,k},\end{aligned}$$

where

$$J_{n,k} \triangleq \sum_{a_1^k\in B_k} |p_k^\circ(a_1^k|x_1^n) - p_k(a_1^k|x_1^n)|.$$

According to $J_{n,k} \le (2k^2)/(n-k+1)$, we have

$$\begin{aligned}&\left|H_k^\circ(x_1^n) - \hat{H}_k(x_1^n)\right| \\ &\le \frac{k-1}{n}\tilde{H}_k(x_1^n) - \frac{n-k+1}{n}\log\frac{n-k+1}{n} \\ &\quad +J_{n,k}\log k - J_{n,k}\log J_{n,k}.\end{aligned}$$

Since x is f.e., $\hat{H}_k(x) < \infty$, and thus (7) was proved. □

Lemma 5 *For any f.e. sequence $x \in \mathcal{X}^\infty$, any integer k and any $\epsilon > 0$, there is an integer $M = M(x,k,\epsilon)$ such that for all $m \ge M$,*

$$\hat{H}_m(x) \le \tilde{H}_k(x) + \epsilon.$$

Proof: Let N be a large number such that,

$$\tilde{H}_k(x_1^n) \le \tilde{H}_k(x) + \epsilon/2, \quad \forall n \ge N.$$

Choose n and m such that $n \ge N$ and $n \ge m \ge k$ (specified later). Let $\tilde{\sigma}_k$ be a length function on $\mathcal{X}^k$ such that

$$\tilde{\sigma}_k(a_1^k) = -\log q_k(a_1^k|x_1^n), \quad \forall a_1^k \in \mathcal{X}^k,$$

and σ_x be a length function which satisfies (1). Using $\tilde{\sigma}_k$ and σ_x, we can construct a length function $\hat{\sigma}_m$ on $\mathcal{X}^m$ which satisfies

$$\hat{\sigma}_m(x_i^{i+m-1}) \le \sum_{\substack{j\equiv 1 \pmod{k}\\ i\le j\le i+m-k}} \tilde{\sigma}_k(x_j^{j+k-1}) + m\delta_{m,i},$$

for each $i = 1, 2, \cdots$, where

$$\delta_{m,i} = \frac{1}{m}\left\{\sum_{j=i}^{i-1+s_1}\sigma_x(x_j) + \sum_{j=i+m-s_2}^{i+m-1}\sigma_x(x_j) + \log k\right\},$$

and s_1 is the minimum nonnegative integer such that $i+s_1 \equiv 1 \pmod{k}$ and s_2 is the minimum nonnegative integer such that $i+m-s_2 \equiv 1 \pmod{k}$. Since $s_1 + s_2 \le 2k$ and σ_x satisfies (5),

$$\frac{1}{n}\sum_{i=1}^{n-m+1}\delta_{m,i} \le \epsilon/2,$$

if n and m are large enough. Hence, we have,

$$\begin{aligned}&mn\hat{H}_m(x_1^n) \\ &\le \sum_{i=1}^{n-m+1}\hat{\sigma}_m(x_i^{i+m-1}) \\ &\le m\sum_{\substack{i\equiv 1 \pmod{k}\\ 1\le i\le n-k+1}} \tilde{\sigma}_k(x_i^{i+k-1}) + m\sum_{i=1}^{n-m+1}\delta_{m,i} \\ &\le mn\tilde{H}_k(x_1^n) + m\sum_{i=1}^{n-m+1}\delta_{m,i} \\ &\le mn\left\{\tilde{H}_k(x) + \epsilon/2\right\} + m\sum_{i=1}^{n-m+1}\delta_{m,i},\end{aligned}$$

and thus, for sufficiently large m and n,

$$\hat{H}_m(x_1^n) \le \tilde{H}_k(x) + \epsilon.$$

Therefore, we have

$$\hat{H}_m(x) = \limsup_{n\to\infty}\hat{H}_m(x_1^n) \le \tilde{H}_k(x) + \epsilon.$$

□

Next, we state some consequences of Theorem 1.

Corollary 1 *For any f.e. sequence $x \in \mathcal{X}^\infty$, and any k and any length function σ_k on $\mathcal{X}^k$,*

$$h_{fs}(x) = \hat{H}(x) = \tilde{H}(x) \le \limsup_{m\to\infty}\frac{1}{km}\sum_{i=0}^{m-1}\sigma_k(x_{ik}^{(i+1)k}). \quad (8)$$

Now suppose $x \in \mathcal{X}^\infty$ is compressed by separately encoding each non-overlapping k-block $x_{ik}^{(i+1)k}$ in x into a codeword of length $\sigma_k(x_{ik}^{(i+1)k})$. Corollary 1 shows that for any σ_k, the compression rate is lower bounded by $\hat{H}(x)$ provided that x is f.e. On the other hand, the next corollary demonstrates the existence of a compression scheme, for which the compression rate asymptotically tends to $\hat{H}(x)$.

Corollary 2 *For any length function σ on $\mathcal{X}$, let $\mathcal{X}^\infty(\sigma)$ be the subset of $\mathcal{X}^\infty$ such that $x \in \mathcal{X}^\infty(\sigma)$ iff (1) holds.*

Consider countable length functions $\sigma^{(1)}, \sigma^{(2)}, \cdots$ on $\mathcal{X}$. There exists a sequence of length functions $\{\sigma_n^*\}_{n=1}^{\infty}$ ($\sigma_n^* : \mathcal{X}^n \to \mathbf{R}$) such that

$$\limsup_{n\to\infty} \frac{\sigma_n^*(x_1^n)}{n} \le h_{fs}(x) = \hat{H}(x) = \tilde{H}(x),$$
$$\forall x \in \bigcup_{i=1}^{\infty} \mathcal{X}^{\infty}(\sigma^{(i)}). \tag{9}$$

Proof: Let $\phi_\sigma^{(LZ)}$ be a variation of Lempel-Ziv 78 encoder [2] which uses $\sigma(a)$ bits to encode the last symbol $a \in \mathcal{X}$ of each phrase. $\phi_\sigma^{(LZ)}$ satisfies

$$\limsup_{n\to\infty} \frac{1}{n} \left\| \phi_\sigma^{(LZ)}(x_1^n) \right\| \le h_{fs}(x),\ \forall x \in \mathcal{X}^{\infty}(\sigma).$$

We can construct an length function σ_n^* such that

$$\sigma_n^*(x_1^n) = \lceil \log n \rceil + \min_{1\le i\le n} \|\phi_{\sigma_i}^{(LZ)}(x_1^n)\|.$$

It is easy to see that σ_n^* satisfies (9). □

3. Empirical entropies with fidelity criterion

Let $\mathcal{Y}$ be a finite alphabet. Let $\rho : \mathcal{X} \times \mathcal{Y} \to [0, \infty)$ be a function and $\{\rho_k\}_{k=1}^{\infty}$ be the single-letter fidelity criterion generated by ρ, that is, for each $k = 1, 2, \cdots$, $\rho_k : \mathcal{X}^k \times \mathcal{Y}^k \to [0, \infty)$ is defined as

$$\rho_k(x_1^k, y_1^k) = \frac{1}{k} \sum_{i=1}^{k} \rho(x_i, y_i), \quad \forall x_1^k \in \mathcal{X}^k \text{ and } y_1^k \in \mathcal{Y}^k.$$

A partition $\Pi^{(k)}$ of $\mathcal{X}^k$is *D-admissible* if for all $\pi \in \Pi^{(k)}$, there exists $y_1^k \in \mathcal{Y}^k$ such that

$$\rho_k(x_1^k, y_1^k) \le D, \quad \forall x_1^k \in \pi.$$

In the following, we assume that the distortion measure ρ and fidelity criterion D satisfy that there is a D-admissible partition of $\mathcal{X}^k$ for all $k \ge 1$.

The empirical entropies based on the partition $\Pi^{(k)}$ of $\mathcal{X}^k$ are defined as

$$\hat{H}_k(\Pi^{(k)}|x_1^n) \stackrel{\triangle}{=} -\frac{1}{k} \sum_{\pi\in\Pi^{(k)}} p_k(\pi|x_1^n) \log p_k(\pi|x_1^n),$$

and

$$\tilde{H}_k(\Pi^{(k)}|x_1^n) \stackrel{\triangle}{=} -\frac{1}{k} \sum_{\pi\in\Pi^{(k)}} q_k(\pi|x_1^n) \log q_k(\pi|x_1^n),$$

where $p_k(\pi|x_1^n) = \sum_{a_1^k\in\pi} p_k(a_1^k|x_1^n)$ and $q_k(\pi|x_1^n) = \sum_{a_1^k\in\pi} q_k(a_1^k|x_1^n)$.

Above two quantities have a connection with fixed-distortion variable-length lossy compression of individual sequences. Yang and Kieffer [3] considered the coding schemes, under which $x \in \mathcal{X}^{\infty}$ is encoded by separately encoding each non-overlapping k-block, and showed that the optimal coding rate tends to $\inf_k \limsup_{n\to\infty} \inf_{\Pi^{(k)}} \tilde{H}_k(\Pi^{(k)}|x_1^n)$. Our next theorem characterizes this quantity.

Theorem 2

$$\begin{aligned}
&\inf_k \limsup_{n\to\infty} \inf_{\substack{\Pi^{(k)}:\\ D\text{-admissible}}} \hat{H}_k(\Pi^{(k)}|x_1^n) \\
=\ &\lim_{k\to\infty} \limsup_{n\to\infty} \inf_{\substack{\Pi^{(k)}:\\ D\text{-admissible}}} \hat{H}_k(\Pi^{(k)}|x_1^n) \\
=\ &\inf_k \limsup_{n\to\infty} \inf_{\substack{\Pi^{(k)}:\\ D\text{-admissible}}} \tilde{H}_k(\Pi^{(k)}|x_1^n) \\
=\ &\lim_{k\to\infty} \limsup_{n\to\infty} \inf_{\substack{\Pi^{(k)}:\\ D\text{-admissible}}} \tilde{H}_k(\Pi^{(k)}|x_1^n)
\end{aligned}$$

where $\inf_{\Pi^{(k)}}$ are taken over all D-admissible partitions of $\mathcal{X}^k$.

Theorem 2 can be shown by the following Lemma 6 and 7.

Lemma 6 *For any k and any $\epsilon > 0$, there exists an integer $M = M(k, \epsilon)$ such that for all $m \ge M$ and sufficiently large n,*

$$\inf_{\substack{\Pi^{(m)}:\\ D\text{-admissible}}} \hat{H}_m(\Pi^{(m)}|x_1^n) \le \inf_{\substack{\Pi^{(k)}:\\ D\text{-admissible}}} \tilde{H}_k(\Pi^{(k)}|x_1^n) + \epsilon.$$

Proof: Let $\bar{f} : \mathcal{X} \to \mathcal{Y}$ be the function defined as

$$\bar{f}(a) = \arg\min_{b\in\mathcal{Y}} \rho(a, b), \quad \forall a \in \mathcal{X},$$

and $\bar{\ell}$ be the length function on $\mathcal{Y}$ such that $\bar{\ell}(b) = \lceil \log |\mathcal{Y}| \rceil$ ($\forall b \in \mathcal{Y}$).

Choose D-admissible partition $\tilde{\Pi}^{(k)}$ of $\mathcal{X}^k$ such that

$$\tilde{H}_k(\tilde{\Pi}^{(k)}|x_1^n) \le \inf_{\substack{\Pi^{(k)}:\\ D\text{-admissible}}} \tilde{H}_k(\Pi^{(k)}|x_1^n) + \epsilon/2.$$

By $\tilde{\Pi}^{(k)}$, we can define D-admissible length function $\tilde{\sigma}_k$ such that

$$\tilde{\sigma}_k(a_1^k) = -\log q_k(\pi|x_1^n), \quad \text{if } a_1^k \in \pi \in \tilde{\Pi}^{(k)}.$$

Choose n and m such that $n \ge m \ge k$ (specified later). Using $\bar{f}, \bar{\ell}$ and $\tilde{\sigma}_k$, we can construct a D-admissible length function $\hat{\sigma}_m$ on $\mathcal{X}^m$ such that for all $i = 1, 2, \cdots$,

$$\hat{\sigma}_m(x_i^{i+m-1}) \le \sum_{\substack{j\equiv 1 \pmod k\\ i\le j\le i+m-k}} \tilde{\sigma}_k(x_j^{j+k-1}) + m\delta_m,$$

where $\delta_m = (\log k + 2\lceil k \log |\mathcal{Y}|\rceil)/m$. We have

$$
\begin{aligned}
nm & \inf_{\substack{\Pi^{(m)}: \\ D\text{-admissible}}} \hat{H}(\Pi^{(m)}|x_1^n) \\
& \leq nm \sum_{a_1^m \in \mathcal{X}^m} p_k(a_1^m|x_1^n)\hat{\sigma}_m(a_1^m) \\
& = \sum_{i=1}^{n-m+1} \hat{\sigma}_m(x_i^{i+m-1}) \\
& \leq m \sum_{\substack{j\equiv 1 \pmod k \\ 1\leq j\leq n-k+1}} \tilde{\sigma}_k(x_j^{j+k-1}) + nm\delta_m \\
& \leq nm\tilde{H}_k(\tilde{\Pi}^{(k)}|x_1^n) + nm\delta_m \\
& \leq nm\left\{ \inf_{\substack{\Pi^{(k)}: \\ D\text{-admissible}}} \tilde{H}_k(\Pi^{(k)}|x_1^n) + \epsilon/2 \right\} + nm\delta_m.
\end{aligned}
$$

Since $\delta_m \leq \epsilon/2$ if m is large enough, we have the lemma. □

Lemma 7 *For any k and any $\epsilon > 0$, there exists an integer $M = M(k,\epsilon)$ such that for all $m \geq M$ and sufficiently large n,*

$$
\inf_{\substack{\Pi^{(m)}: \\ D\text{-admissible}}} \tilde{H}_m(\Pi^{(m)}|x_1^n) \leq \inf_{\substack{\Pi^{(k)}: \\ D\text{-admissible}}} \hat{H}_k(\Pi^{(k)}|x_1^n) + \epsilon.
$$

Proof: Let $\bar{f} : \mathcal{X} \to \mathcal{Y}$ be the function defined as

$$
\bar{f}(a) = \arg\min_{b\in\mathcal{Y}} \rho(a,b), \quad \forall a \in \mathcal{X},
$$

and $\bar{\ell}$ be the length function on $\mathcal{Y}$ such that $\bar{\ell}(b) = \lceil \log |\mathcal{Y}| \rceil$ $(\forall b \in \mathcal{Y})$.

Choose D-admissible partition $\hat{\Pi}^{(k)}$ of $\mathcal{X}^k$ such that

$$
\hat{H}_k(\hat{\Pi}^{(k)}|x_1^n) \leq \inf_{\substack{\Pi^{(k)}: \\ D\text{-admissible}}} \hat{H}_k(\Pi^{(k)}|x_1^n) + \epsilon/2.
$$

By $\hat{\Pi}^{(k)}$, we can define D-admissible length function $\hat{\sigma}_k$ such that

$$
\hat{\sigma}_k(a_1^k) = -\log p_k(\pi|x_1^n), \quad \text{if } a_1^k \in \pi \in \hat{\Pi}^{(k)}.
$$

Choose n and m so that $n \geq m \geq k$ (specified later). Using $\bar{f}, \bar{\ell}$ and $\hat{\sigma}_k$, we can construct a D-admissible length function $\tilde{\sigma}_m$ on $\mathcal{X}^m$ such that

$$
\begin{aligned}
& \tilde{\sigma}_m(a_1^m) \\
& = \log k + 2\lceil k\log|\mathcal{Y}|\rceil + \min_{0\leq t\leq k-1} \left\{ \sum_{\substack{i\equiv t \pmod k \\ 1\leq i\leq m-k}} \hat{\sigma}_k(a_i^{i+k}) \right\} \\
& \leq \log k + 2\lceil k\log|\mathcal{Y}|\rceil + \frac{1}{k}\sum_{i=1}^{m-k} \hat{\sigma}_k(a_i^{i+k}).
\end{aligned}
$$

$$
\begin{aligned}
n & \inf_{\substack{\Pi^{(m)}: \\ D\text{-admissible}}} \tilde{H}(\Pi^{(m)}|x_1^n) \\
& \leq \frac{n}{m} \sum_{a_1^m\in\mathcal{X}^m} q_k(a_1^m|x_1^n)\tilde{\sigma}_m(a_1^m) \\
& = \sum_{i=0}^{n/m-1} \tilde{\sigma}_m(x_{mi+1}^{m(i+1)}) \\
& \leq \sum_{i=0}^{n/m-1} \frac{1}{k}\sum_{j=1}^{m-k} \hat{\sigma}_k(x_{mi+j}^{mi+j+k}) + n\delta_m \\
& \leq n\hat{H}_k(\hat{\Pi}^{(k)}|x_1^n) + n\delta_m \\
& \leq n\left\{ \inf_{\substack{\Pi^{(k)}: \\ D\text{-admissible}}} \hat{H}_k(\Pi^{(k)}|x_1^n) + \epsilon/2 \right\} + n\delta_m,
\end{aligned}
$$

where $\delta_m = (\log k + 2\lceil k\log|\mathcal{Y}|\rceil)/m$ and $\delta_m \leq \epsilon/2$ if m is large enough. Hence, we have the lemma. □

References

[1] A. Lempel and J. Ziv, "On the complexity of finite sequences," *IEEE Trans. Inform. Theory*, vol. IT-22, pp. 75-81, January 1976.

[2] J. Ziv and A. Lempel, "Compression of individual sequences via variable-rate coding," *IEEE Trans. Inform. Theory*, vol. IT-24, pp. 530-536, September 1978.

[3] E. Yang and J. C. Kieffer, "Simple universal lossy data compression schemes derived from the Lempel-Ziv algorithm," *IEEE Trans. Inform. Theory,* vol. IT-42, pp.239-245, January 1996.

[4] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley, New York, 1991.

[5] E. Plotnik, M. J. Weinberger, and J. Ziv, "Upper bounds on the probability of sequences emitted by finite-state sources and on the redundancy of the Lempel-Ziv algorithm," *IEEE Trans. Inform. Theory*, vol. IT-38, pp. 66-72, January 1992.

[6] T. Uyematsu and F. Kanaya, "Relations among three optimal criteria of universal coding of individual sequences", *Proc. of STW2003,* pp. 15-20, September 2003.

[7] T. Uyematsu and F. Kanaya, "Asymptotical Optimality of Two Variations of Lempel-Ziv Codes for Sources with Countable Infinite Alphabet", *Proc. of SITA2001*, pp. 481-484, December 2001.