IEICE TRANS. FUNDAMENTALS, VOL.E96–A, NO.12 DECEMBER 2013

PAPER *Special Section on Information Theory and Its Applications*

Redundancy-Optimal FF Codes for a General Source and Its Relationships to the Rate-Optimal FF Codes*

Mitsuharu ARIMURA[†a)], **Hiroki KOGA**[††b)], *Senior Members*, *and* **Ken-ichi IWATA**[†††c)], *Member*

SUMMARY In this paper we consider fixed-to-fixed length (FF) coding of a general source $\boldsymbol{X}$ with vanishing error probability and define two kinds of optimalities with respect to the coding rate and the redundancy, where the redundancy is defined as the difference between the coding rate and the symbolwise ideal codeword length. We first show that the infimum achievable redundancy coincides with the asymptotic width $W(\boldsymbol{X})$ of the entropy spectrum. Next, we consider the two sets $C_{\overline{H}}(\boldsymbol{X})$ and $C_W(\boldsymbol{X})$ and investigate relationships between them, where $C_{\overline{H}}(\boldsymbol{X})$ and $C_W(\boldsymbol{X})$ denote the sets of all the optimal FF codes with respect to the coding rate and the redundancy, respectively. We give two necessary and sufficient conditions corresponding to $C_{\overline{H}}(\boldsymbol{X}) \subseteq C_W(\boldsymbol{X})$ and $C_W(\boldsymbol{X}) \subseteq C_{\overline{H}}(\boldsymbol{X})$, respectively. We can also show the existence of an FF code that is optimal with respect to both the redundancy and the coding rate.
key words: *fixed-to-fixed length source coding, information-spectrum methods, general sources, coding rate, redundancy*

1. Introduction

Information-spectrum methods, which are described in detail in [4], originate from a seminal paper by Han and Verdú [5]. Information-spectrum methods provide a methodology to analyze performance of coding of general sources, where the class of general sources includes vast classes of sources such as stationary memoryless sources, stationary ergodic sources, stationary sources and nonstationary and/or nonergodic sources. Given a general source $\boldsymbol{X}$, it is fundamental to characterize the infimum achievable coding rate of fixed-to-fixed length (FF) codes subject to a criterion on the error probability. If we require that the error probability asymptotically vanishes, the infimum achievable coding rate coincides with the spectral sup-entropy rate $\overline{H}(\boldsymbol{X})$ of the source [5].

In this paper we consider redundancy of FF coding of a general source $\boldsymbol{X}$. The redundancy introduced in this paper can be regarded as one of variations of the worst-case redundancy of fixed-to-variable length lossless data compression codes [8], [10]–[12] and can be used as another measure of performance of FF codes. We define the redundancy of FF codes as the difference between the symbolwise ideal codeword length and the coding rate. It is shown that the infimum achievable redundancy coincides with the asymptotic width $W(\boldsymbol{X})$ of the entropy spectrum of $\boldsymbol{X}$. The asymptotic width $W(\boldsymbol{X})$ was first defined by one of the authors in [6], [7] in the context of homophonic coding. The obtained result indicates that $W(\boldsymbol{X})$ has another operational meaning in a simpler problem of the redundancy of FF codes.

Next, we define the class $C_W(\boldsymbol{X})$ of all the optimal FF codes with respect to the redundancy and investigate relationships between $C_W(\boldsymbol{X})$ and another class $C_{\overline{H}}(\boldsymbol{X})$ of the optimal codes with respect to the coding rate. We obtain two necessary and sufficient conditions corresponding to $C_{\overline{H}}(\boldsymbol{X}) \subseteq C_W(\boldsymbol{X})$ and $C_W(\boldsymbol{X}) \subseteq C_{\overline{H}}(\boldsymbol{X})$, respectively. More precisely, we show that $C_{\overline{H}}(\boldsymbol{X}) \subseteq C_W(\boldsymbol{X})$ if and only if $W(\boldsymbol{X}) = \overline{H}(\boldsymbol{X}) - \underline{H}(\boldsymbol{X})$ while $C_W(\boldsymbol{X}) \subseteq C_{\overline{H}}(\boldsymbol{X})$ if and only if $W(\boldsymbol{X}) = \overline{H}(\boldsymbol{X}) - \underline{H}^*(\boldsymbol{X})$, where $\underline{H}(\boldsymbol{X})$ is the spectral inf-entropy rate [4] and $\underline{H}^*(\boldsymbol{X})$ is a quantity defined in [2], [9]. These results immediately imply that $C_{\overline{H}}(\boldsymbol{X}) = C_W(\boldsymbol{X})$ if and only if $\underline{H}(\boldsymbol{X}) = \underline{H}^*(\boldsymbol{X})$, which means that the left endpoint of the entropy spectrum converges to a constant. In addition, we show that the intersection of $C_{\overline{H}}(\boldsymbol{X})$ and $C_W(\boldsymbol{X})$ is always nonempty. That is, there exists an FF code which is asymptotically optimal with respect to both the coding rate and the redundancy.

This paper is organized as follows. Section 2 is devoted to definitions of information-theoretic quantities that are used throughout this paper. In Sect. 3, we define the infimum achievable redundancy $R_{red}(\boldsymbol{X})$ of an FF code and show that $R_{red}(\boldsymbol{X})$ coincides with $W(\boldsymbol{X})$. The two classes of the optimal codes are defined in Sect. 4. Relationships between the two classes are analyzed in detail.

2. Preliminaries

Let $\mathbb{N}$ be the set of all the positive integers. For each $n \in \mathbb{N}$ let $X^n = X_1X_2\cdots X_n$ be a random variable representing n outputs from a source, where each X_i takes values in a finite or countable set $\mathcal{X}$. The probability distribution of X^n is denoted by P_{X^n}. The probability of $X^n = x^n$ is expressed as $P_{X^n}(x^n)$. We call $\boldsymbol{X} = \{X^n\}_{n\in\mathbb{N}}$ a general source [5]. We do not impose the consistency condition on $P_{X^n}, n \in \mathbb{N}$. The probability distribution of $\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)}$, the self information

Manuscript received February 8, 2013.
Manuscript revised June 25, 2013.
[†]The author is with the Department of Applied Computer Sciences, Shonan Institute of Technology, Fujisawa-shi, 251-8511 Japan.
[††]The author is with the Graduate School of Systems and Information Engineering, University of Tsukuba, Tsukuba-shi, 305-8571 Japan.
[†††]The author is with the Graduate School of Engineering, University of Fukui, Fukui-shi, 910-8507 Japan.
*This paper was partly presented at 2010 International Symposium on Information Theory and its Applications, Taichung, Taiwan.
a) E-mail: arimura@m.ieice.org
b) E-mail: koga@iit.tsukuba.ac.jp
c) E-mail: k-iwata@u-fukui.ac.jp
DOI: 10.1587/transfun.E96.A.2332

per symbol, is called the *entropy spectrum*.

For a general source $\boldsymbol{X}$, we define four limits concerning the entropy spectrum.

Definition 2.1 (Han-Verdú [5]):

$$\overline{H}(\boldsymbol{X}) = \inf\left\{\alpha : \liminf_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \le \alpha\right\} = 1\right\}, \tag{1}$$

$$\underline{H}(\boldsymbol{X}) = \sup\left\{\beta : \liminf_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \ge \beta\right\} = 1\right\}. \tag{2}$$

Definition 2.2 (Chen-Alajaji [2]):

$$\overline{H}^*(\boldsymbol{X}) = \inf\left\{\alpha : \limsup_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \le \alpha\right\} = 1\right\},$$
$$\underline{H}^*(\boldsymbol{X}) = \sup\left\{\beta : \limsup_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \ge \beta\right\} = 1\right\}. \tag{3}$$

Throughout this paper, the bases of logarithmic and exponential functions are assumed to be 2, and any source $\boldsymbol{X}$ is assumed to satisfy $\overline{H}(\boldsymbol{X}) < \infty$.

Next we define the asymptotic width of the entropy spectrum of a source.

Definition 2.3 (Koga [6], [7]):

$$W(\boldsymbol{X}) = \inf_{\mathcal{G}} \limsup_{n\to\infty} (b_n - a_n), \tag{4}$$

where

$$\mathcal{G} = \left\{\{(a_n, b_n)\}_{n\in\mathbb{N}} : \text{for any constant } \gamma > 0 \right.$$
$$\left. \lim_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \in (a_n - \gamma, b_n + \gamma)\right\} = 1\right\}$$

is a set of sequences of intervals and throughout this paper we consider sequences $\{(a_n, b_n)\}_{n\in\mathbb{N}}$ of intervals satisfying $a_n \le b_n$ for all $n \ge 1$.

It is known that $W(\boldsymbol{X})$ has the following upper and lower bounds [6], [7], [9]:

$$W(\boldsymbol{X}) \le \overline{H}(\boldsymbol{X}) - \underline{H}(\boldsymbol{X}), \tag{5}$$
$$W(\boldsymbol{X}) \ge \overline{H}(\boldsymbol{X}) - \underline{H}^*(\boldsymbol{X}), \tag{6}$$
$$W(\boldsymbol{X}) \ge \overline{H}^*(\boldsymbol{X}) - \underline{H}(\boldsymbol{X}).$$

In the following, we show several examples of sources. First we give two sources such that the equalities hold in both (5) and (6).

Example 2.1: Let $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ be stationary and memoryless sources with probability distributions P_1 and P_2, respectively. The entropies of P_1 and P_2 are written as $H(P_1)$ and $H(P_2)$, respectively. Assume that $H(P_1) < H(P_2)$. Let $\boldsymbol{X} = \{X^n\}_{n\in\mathbb{N}}$ be the mixed source of $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ with probability distribution

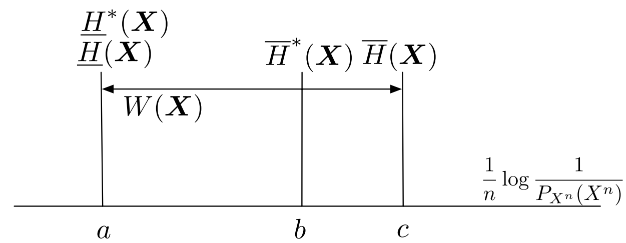

Fig. 1 Example of a source such that the equalities hold in both (5) and (6).

$$P_{X^n}(x^n) = (1-\alpha)\prod_{i=1}^{n} P_1(x_i) + \alpha\prod_{i=1}^{n} P_2(x_i),$$

where α is a constant satisfying $0 < \alpha < 1$. Then $\overline{H}(\boldsymbol{X}) = \overline{H}^*(\boldsymbol{X}) = H(P_2)$ and $\underline{H}(\boldsymbol{X}) = \underline{H}^*(\boldsymbol{X}) = H(P_1)$. For this source, $W(\boldsymbol{X}) = \overline{H}(\boldsymbol{X}) - \underline{H}(\boldsymbol{X}) = \overline{H}(\boldsymbol{X}) - \underline{H}^*(\boldsymbol{X}) = H(P_2) - H(P_1)$.

Example 2.2: Consider the three probability distributions P_{X_1}, $P_{X_{2a}}$ and $P_{X_{2b}}$ on $\mathcal{X}$ satisfying $H(P_{X_1}) = a$, $H(P_{X_{2a}}) = b$ and $H(P_{X_{2b}}) = c$ for some constants $a < b < c$. For all $n \ge 1$ define the probability distributions on $\mathcal{X}^n$ by

$$P_{X^n}(x^n) = \begin{cases} \frac{1}{2}P_{X_1^n}(x^n) + \frac{1}{2}P_{X_{2a}^n}(x^n) & \text{if } n \text{ is odd,} \\ \frac{1}{2}P_{X_1^n}(x^n) + \frac{1}{2}P_{X_{2b}^n}(x^n) & \text{if } n \text{ is even,} \end{cases}$$

for all $x^n \in \mathcal{X}^n$, where $P_{X_1^n}(x^n) = \prod_{i=1}^n P_{X_1}(x_i)$, $P_{X_{2a}^n}(x^n) = \prod_{i=1}^n P_{X_{2a}}(x_i)$ and $P_{X_{2b}^n}(x^n) = \prod_{i=1}^n P_{X_{2b}}(x_i)$.

Figure 1 depicts the entropy spectrum of this source for sufficiently large n. Concerning (5) and (6), this source satisfies $\underline{H}(\boldsymbol{X}) = \underline{H}^*(\boldsymbol{X})$ and

$$W(\boldsymbol{X}) = \overline{H}(\boldsymbol{X}) - \underline{H}^*(\boldsymbol{X}) = \overline{H}(\boldsymbol{X}) - \underline{H}(\boldsymbol{X}) = c - a. \tag{7}$$

Next we give an example of $\boldsymbol{X}$ such that the inequality strictly holds in (5) and the equality holds in (6).

Example 2.3: Let $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ be the sources defined in Example 2.1. Let $\boldsymbol{X} = \{X^n\}_{n\in\mathbb{N}}$ be a nonstationary source defined as

$$P_{X^n}(x^n) = \begin{cases} \prod_{i=1}^{n} P_1(x_i) & \text{if } n \text{ is odd,} \\ \prod_{i=1}^{n} P_2(x_i) & \text{if } n \text{ is even.} \end{cases}$$

Then it holds that $\overline{H}(\boldsymbol{X}) = H(P_2)$ and $\underline{H}(\boldsymbol{X}) = H(P_1)$. For this source, $\overline{H}(\boldsymbol{X}) - \underline{H}(\boldsymbol{X}) = H(P_1) - H(P_2)$. On the other hand, this source satisfies $W(\boldsymbol{X}) = 0$ because the information spectrum concentrates to one point as $n \to \infty$ due to the law of large numbers.

Finally, we give an example such that the inequalities strictly hold in both (5) and (6).

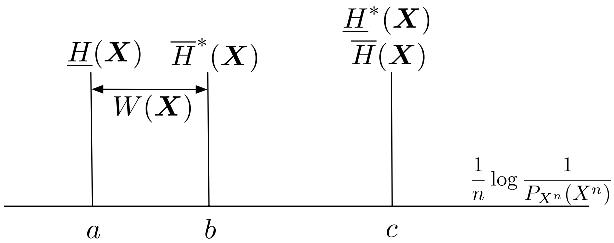

Fig. 2 Example of a source such that the inequalities strictly hold in both (5) and (6).

Example 2.4: Consider three probability distributions $P_{X_{1a}}$, $P_{X_{1b}}$ and P_{X_2} on $\mathcal{X}$ satisfying $H(P_{X_{1a}}) = a$, $H(P_{X_{1b}}) = b$ and $H(P_{X_2}) = c$ for some constants $a < b < c$. Define

$$P_{X^n}(x^n) = \begin{cases} \frac{1}{2}P_{X^n_{1a}}(x^n) + \frac{1}{2}P_{X^n_{1b}}(x^n) & \text{if } n \text{ is odd,} \\ P_{X^n_2}(x^n) & \text{if } n \text{ is even,} \end{cases}$$

for all $x^n \in \mathcal{X}^n$ and $n \geq 1$, where $P_{X^n_{1a}}(x^n) = \prod_{i=1}^n P_{X_{1a}}(x_i)$, $P_{X^n_{1b}}(x^n) = \prod_{i=1}^n P_{X_{1b}}(x_i)$ and $P_{X^n_2}(x^n) = \prod_{i=1}^n P_{X_2}(x_i)$.

Figure 2 depicts the entropy spectrum of this source for sufficiently large n. Concerning (5) and (6), this source satisfies $W(\boldsymbol{X}) = \overline{H}^*(\boldsymbol{X}) - \underline{H}(\boldsymbol{X})$ and therefore we have

$$W(\boldsymbol{X}) > \overline{H}(\boldsymbol{X}) - \underline{H}^*(\boldsymbol{X}) = 0, \tag{8}$$

$$W(\boldsymbol{X}) < \overline{H}(\boldsymbol{X}) - \underline{H}(\boldsymbol{X}). \tag{9}$$

3. Infimum Achievable Redundancy of FF Codes

In this section, we study the redundancy of an FF code for a general source $\boldsymbol{X}$. An output $x^n \in \mathcal{X}^n$ from the source is encoded by an encoder $\varphi_n : \mathcal{X}^n \to \mathcal{M}_n = \{1, 2, \ldots, M_n\}$ to a codeword $\varphi_n(x^n)$, where $M_n < \infty$. The codeword $\varphi_n(x^n)$ is decoded by a decoder $\psi_n : \mathcal{M}_n \to \mathcal{X}^n$ to $\psi_n(\varphi_n(x^n))$. The coding rate is given by $\frac{1}{n}\log M_n$. Since φ_n is not one-to-one, decoding error occurs for some x^n. Define the probability of the decoding error (error probability) as

$$\varepsilon_n = \Pr\{X^n \notin \mathcal{D}_n\},$$

where $\mathcal{D}_n = \{x^n \in \mathcal{X}^n : \psi_n(\varphi_n(x^n)) = x^n\}$. Let $\boldsymbol{C}$ denote a sequence $\{(\varphi_n, \psi_n)\}_{n\in\mathbb{N}}$ of the pairs of an encoder φ_n and a decoder ψ_n. We call $\boldsymbol{C}$ a *code* for simplicity.

Han and Verdú define the infimum achievable coding rate of an FF code as follows.

Definition 3.1 (Han-Verdú [5]): A rate R is called *achievable coding rate* for a source $\boldsymbol{X}$ if there exists a code $\boldsymbol{C} = \{(\varphi_n, \psi_n)\}_{n\in\mathbb{N}}$ satisfying

$$\lim_{n\to\infty} \varepsilon_n = 0, \tag{10}$$

$$\limsup_{n\to\infty} \frac{1}{n}\log M_n \leq R. \tag{11}$$

The infimum of achievable coding rate R for $\boldsymbol{X}$ is called the *infimum achievable coding rate* and is denoted by $R_{rate}(\boldsymbol{X})$.

The infimum achievable coding rate for $\boldsymbol{X}$ is given in the following theorem.

Theorem 3.1 (Han-Verdú [5]):

$$R_{rate}(\boldsymbol{X}) = \overline{H}(\boldsymbol{X}).$$

Next we introduce the infimum achievable redundancy of an FF code.

Definition 3.2: A redundancy R is called an *achievable redundancy* for $\boldsymbol{X}$ if there exists a code $\boldsymbol{C} = \{(\varphi_n, \psi_n)\}_{n\in\mathbb{N}}$ satisfying (10) and

$$\lim_{n\to\infty} \Pr\left\{\frac{1}{n}\log M_n - \frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \leq R + \gamma\right\} = 1 \quad \text{for any } \gamma > 0. \tag{12}$$

The infimum of the achievable redundancy R for $\boldsymbol{X}$ is called the *infimum achievable redundancy* and is denoted by $R_{red}(\boldsymbol{X})$.

A general formula of the infimum achievable redundancy for $\boldsymbol{X}$ is given in the following theorem.

Theorem 3.2:

$$R_{red}(\boldsymbol{X}) = W(\boldsymbol{X}).$$

Proof: Letting $\gamma > 0$ be an arbitrary constant, we first prove that $W(\boldsymbol{X})$ is achievable redundancy. To this end, we note that the definition of $\mathcal{G}$ guarantees the existence of a sequence of intervals $\{(a_n, b_n)\}_{n\in\mathbb{N}} \in \mathcal{G}$ satisfying

$$\limsup_{n\to\infty} (b_n - a_n) \leq W(\boldsymbol{X}) + \gamma, \tag{13}$$

$$\lim_{n\to\infty} \Pr\{X^n \in \mathcal{A}_n\} = 0, \tag{14}$$

$$\lim_{n\to\infty} \Pr\{X^n \in \mathcal{B}_n\} = 0, \tag{15}$$

where $\mathcal{A}_n$ and $\mathcal{B}_n$ are defined as

$$\mathcal{A}_n = \left\{x^n \in \mathcal{X}^n : \frac{1}{n}\log\frac{1}{P_{X^n}(x^n)} \leq a_n - \gamma\right\}, \tag{16}$$

$$\mathcal{B}_n = \left\{x^n \in \mathcal{X}^n : \frac{1}{n}\log\frac{1}{P_{X^n}(x^n)} \geq b_n + \gamma\right\}. \tag{17}$$

Define $M_n = \exp(\lceil n(b_n+\gamma)\rceil)$. Then, from [4, Lemma 1.3.1], there exists a code $\boldsymbol{C}$ satisfying

$$\varepsilon_n \leq \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \geq \frac{1}{n}\log M_n\right\} \leq \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \geq b_n + \gamma\right\}. \tag{18}$$

In view of (15), (17) and (18), the error probability of $\boldsymbol{C}$ satisfies

$$\lim_{n\to\infty} \varepsilon_n = 0.$$

Next, we evaluate the redundancy of $\boldsymbol{C}$. Due to (13), there exists an integer $n_0 = n_0(\gamma)$ satisfying

$$b_n - a_n \leq W(\boldsymbol{X}) + 2\gamma$$

for all $n \geq n_0$. Then, for all $n \geq \max\{n_0, 1/\gamma\}$ the pointwise redundancy of each $x^n \notin \mathcal{A}_n$ is bounded as follows:

$$\begin{aligned}\frac{1}{n}\log M_n - \frac{1}{n}\log\frac{1}{P_{X^n}(x^n)} &\leq b_n + \gamma + \frac{1}{n} - (a_n - \gamma)\\ &\leq W(\boldsymbol{X}) + 5\gamma.\end{aligned}$$

Therefore, together with (14), we obtain

$$\lim_{n\to\infty}\Pr\left\{\frac{1}{n}\log M_n - \frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \leq W(\boldsymbol{X}) + 5\gamma\right\} = 1.$$

Since $\gamma > 0$ is arbitrary, this establishes that $W(\boldsymbol{X})$ is the achievable redundancy.

Hereinafter, we prove that $R \geq W(\boldsymbol{X})$ always holds if R is an achievable redundancy. From the assumption, there exists a code satisfying

$$\lim_{n\to\infty}\varepsilon_n = 0, \tag{19}$$

$$\lim_{n\to\infty}\Pr\left\{\frac{1}{n}\log M_n - \frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \leq R + \gamma\right\} = 1 \tag{20}$$

for any $\gamma > 0$. Recall here that any code $\boldsymbol{C}$ satisfies

$$\varepsilon_n \geq \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \geq \frac{1}{n}\log M_n + \gamma\right\} - \exp(-n\gamma)$$

for all $n \geq 1$ and any $\gamma > 0$ [4, Lemma 1.3.2]. Then (19) implies that

$$\lim_{n\to\infty}\Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \geq \frac{1}{n}\log M_n + \gamma\right\} = 0. \tag{21}$$

Define a_n and b_n by

$$a_n = \frac{1}{n}\log M_n - R \quad\text{and}\quad b_n = \frac{1}{n}\log M_n.$$

Then, it follows from (20) and (21) that

$$\lim_{n\to\infty}\Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \in (a_n - \gamma, b_n + \gamma)\right\} = 1$$

for any $\gamma > 0$, which shows that $\{(a_n, b_n)\}_{n\in\mathbb{N}} \in \mathcal{G}$. Therefore, the definition of $W(\boldsymbol{X})$ implies that

$$\limsup_{n\to\infty}(b_n - a_n) = R \geq W(\boldsymbol{X}).$$

Q.E.D.

Example 3.1: Consider the nonstationary source in Example 2.3 with a finite alphabet $\mathcal{X}$. Let $\boldsymbol{C}$ be a code satisfying $M_n = \exp(\lceil n(H(P_2) + \gamma_n)\rceil)$ for all $n \in \mathbb{N}$ and $\varepsilon_n \to 0$ as $n \to \infty$, where $\{\gamma_n\}_{n\in\mathbb{N}}$ is a sequence of positive numbers satisfying $\gamma_n \to 0$ and $\sqrt{n}\gamma_n \to \infty$ as $n \to \infty$. The existence of such $\boldsymbol{C}$ is easily verified by using the weak law of large numbers. It is obvious that $\boldsymbol{C}$ is optimal with respect to the coding rate, i.e., $\boldsymbol{C}$ satisfies (10) and (11) with $R = \overline{H}(\boldsymbol{X})$. However, $\boldsymbol{C}$ is not optimal with respect to the redundancy because (12) is not satisfied with $R = W(\boldsymbol{X}) = 0$.

On the other hand, let $\boldsymbol{C}'$ be another code with M'_n codewords satisfying $\varepsilon_n \to 0$ as $n \to \infty$, where

$$M'_n = \begin{cases}\exp(\lceil n(H(P_1) + \gamma_n)\rceil) & \text{if } n \text{ is odd},\\ \exp(\lceil n(H(P_2) + \gamma_n)\rceil) & \text{if } n \text{ is even}.\end{cases}$$

The existence of $\boldsymbol{C}'$ is also verified by using the weak law of large numbers. Since this $\boldsymbol{C}'$ satisfies (10) and (11) with $R = \overline{H}(\boldsymbol{X})$ and (12) with $R = W(\boldsymbol{X}) = 0$, $\boldsymbol{C}'$ is optimal with respect to both the coding rate and the redundancy.

In Example 3.1 we can say that $\boldsymbol{C}'$ is more efficient than $\boldsymbol{C}$ because $M'_n \leq M_n$ for all $n \geq 1$ and M'_n is much smaller than M_n for odd n. This means that the optimality with respect to the rate does not always ensure the efficiency of codes for finite n, while a certain property on the source should be reflected in the construction of the optimal code with respect to the redundancy. Introducing the nonconventional notion of the optimality can unveil new aspects of the FF coding of $\boldsymbol{X}$ as are discussed in the following section.

4. Relationships between the Two Classes of Optimal Codes

4.1 Definitions of the Two Classes

In this section we discuss differences between the two kinds of optimalities defined based on Definitions 3.1 and 3.2, respectively. First, we introduce the class $\mathcal{C}_{\overline{H}}(\boldsymbol{X})$ of the optimal codes as follows.

Definition 4.1: A code $\boldsymbol{C}$ is said to be *$\overline{H}$-optimal* (or *rate-optimal*) for $\boldsymbol{X}$ if $\boldsymbol{C}$ satisfies (10) and (11) with $R = \overline{H}(\boldsymbol{X})$.

Definition 4.2: The class $\mathcal{C}_{\overline{H}}(\boldsymbol{X})$ of the $\overline{H}$-optimal codes is the set of all the $\overline{H}$-optimal codes for $\boldsymbol{X}$.

Next we define another class $\mathcal{C}_W(\boldsymbol{X})$ of the optimal codes as follows.

Definition 4.3: A code $\boldsymbol{C}$ is said to be *W-optimal* (or *redundancy-optimal*) for $\boldsymbol{X}$ if $\boldsymbol{C}$ satisfies (10) and (12) with $R = W(\boldsymbol{X})$.

Definition 4.4: The class $\mathcal{C}_W(\boldsymbol{X})$ of the W-optimal codes is the set of all the W-optimal codes for $\boldsymbol{X}$.

We investigate relationships between the two classes $\mathcal{C}_W(\boldsymbol{X})$ and $\mathcal{C}_{\overline{H}}(\boldsymbol{X})$. Table 1 summarizes all the relationships between $\mathcal{C}_W(\boldsymbol{X})$ and $\mathcal{C}_{\overline{H}}(\boldsymbol{X})$, where the dependency on $\boldsymbol{X}$ is omitted. Note that all the conditions are given in the form whether the equalities are satisfied or not in (5) and (6). In addition, we can show that all the conditions are necessary and sufficient.

4.2 Condition for $\mathcal{C}_{\overline{H}}(\boldsymbol{X}) \subseteq \mathcal{C}_W(\boldsymbol{X})$

In this subsection, we investigate the condition for $\mathcal{C}_{\overline{H}}(\boldsymbol{X}) \subseteq \mathcal{C}_W(\boldsymbol{X})$. We show that $\mathcal{C}_{\overline{H}}(\boldsymbol{X}) \subseteq \mathcal{C}_W(\boldsymbol{X})$ if and only if the

Table 1 The relationships between C_W and $C_{\overline{H}}$.

	$W(\boldsymbol{X}) = \overline{H}(\boldsymbol{X}) - \underline{H}^*(\boldsymbol{X})$	$W(\boldsymbol{X}) > \overline{H}(\boldsymbol{X}) - \underline{H}^*(\boldsymbol{X})$
$W(\boldsymbol{X}) = \overline{H}(\boldsymbol{X}) - \underline{H}(\boldsymbol{X})$	$C_W = C_{\overline{H}}$	$C_W \supsetneq C_{\overline{H}}$
$W(\boldsymbol{X}) < \overline{H}(\boldsymbol{X}) - \underline{H}(\boldsymbol{X})$	$C_W \subsetneq C_{\overline{H}}$	$C_W \cap C_{\overline{H}} \neq \emptyset$, $C_W \backslash C_{\overline{H}} \neq \emptyset$, $C_{\overline{H}} \backslash C_W \neq \emptyset$

equality holds in (5). The "if" and "only if" parts are established separately in Theorems 4.1 and 4.2, respectively.

Theorem 4.1: If $\boldsymbol{X}$ satisfies $W(\boldsymbol{X}) = \overline{H}(\boldsymbol{X}) - \underline{H}(\boldsymbol{X})$, then $C_{\overline{H}}(\boldsymbol{X}) \subseteq C_W(\boldsymbol{X})$.

Proof: Fix an $\overline{H}$-optimal code $\boldsymbol{C}$ arbitrarily. This code satisfies (10) and

$$\limsup_{n\to\infty} \frac{1}{n} \log M_n \leq \overline{H}(\boldsymbol{X}). \tag{22}$$

Fix a constant $\gamma > 0$ arbitrarily. Then (22) guarantees the existence of an integer $n_0 = n_0(\gamma)$ satisfying

$$\frac{1}{n} \log M_n \leq \overline{H}(\boldsymbol{X}) + \gamma \tag{23}$$

for all $n \geq n_0$. On the other hand, with defining a set $\mathcal{L}_n$ by

$$\mathcal{L}_n = \left\{ x^n \in \mathcal{X}^n : \frac{1}{n} \log \frac{1}{P_{X^n}(x^n)} \leq \underline{H}(\boldsymbol{X}) - \gamma \right\},$$

it holds from (2) and (10) that

$$\lim_{n\to\infty} \Pr\{X^n \in \mathcal{D}_n \cap \mathcal{L}_n^c\} = 1, \tag{24}$$

where $\mathcal{L}_n^c$ is the complement of $\mathcal{L}_n$. From the assumption of the theorem and (23), any $x^n \in \mathcal{D}_n \cap \mathcal{L}_n^c$ satisfies

$$\begin{aligned} \frac{1}{n} \log M_n - \frac{1}{n} \log \frac{1}{P_{X^n}(x^n)} &\leq \overline{H}(\boldsymbol{X}) + \gamma - (\underline{H}(\boldsymbol{X}) - \gamma) \\ &= W(\boldsymbol{X}) + 2\gamma \end{aligned} \tag{25}$$

for all $n \geq n_0$. Combining (24) and (25), we have

$$\begin{aligned} &\Pr\left\{ \frac{1}{n} \log M_n - \frac{1}{n} \log \frac{1}{P_{X^n}(X^n)} \leq W(\boldsymbol{X}) + 2\gamma \right\} \\ &\quad \geq \Pr\{X^n \in \mathcal{D}_n \cap \mathcal{L}_n^c\} \to 1 \quad \text{as } n \to \infty. \end{aligned} \tag{26}$$

Since $\boldsymbol{C}$ satisfies (10) and $\gamma > 0$ is arbitrary in (26), $\boldsymbol{C}$ turns out to be W-optimal. Q.E.D.

Theorem 4.2: If $C_{\overline{H}}(\boldsymbol{X}) \subseteq C_W(\boldsymbol{X})$, then $\boldsymbol{X}$ satisfies $W(\boldsymbol{X}) = \overline{H}(\boldsymbol{X}) - \underline{H}(\boldsymbol{X})$.

Proof: Since (5) holds for any $\boldsymbol{X}$, it suffices to establish $W(\boldsymbol{X}) \geq \overline{H}(\boldsymbol{X}) - \underline{H}(\boldsymbol{X})$. Suppose that a code $\boldsymbol{C}$ is $\overline{H}$-optimal and satisfies (10),

$$\frac{1}{n} \log M_n > \overline{H}(\boldsymbol{X}) \text{ for all } n \geq 1 \text{ and}$$

$$\lim_{n\to\infty} \frac{1}{n} \log M_n = \overline{H}(\boldsymbol{X}). \tag{27}$$

Note that the existence of such a code $\boldsymbol{C}$ can be proved by the diagonal line argument as follows. Let $\{\gamma_k\}_{k\in\mathbb{N}}$ and $\{\varepsilon_k\}_{k\in\mathbb{N}}$ be arbitrary sequences satisfying

$$\gamma_1 > \gamma_2 > \cdots > \gamma_k > \cdots > 0 \text{ and } \lim_{k\to\infty} \gamma_k = 0, \tag{28}$$

$$1 > \varepsilon_1 > \varepsilon_2 > \cdots > \varepsilon_k > \cdots > 0 \text{ and } \lim_{k\to\infty} \varepsilon_k = 0, \tag{29}$$

respectively. For each k define $M_n^{(k)} = \exp(\lceil n(\overline{H}(\boldsymbol{X}) + \gamma_k) \rceil)$. From [4, Lemma 1.3.1], there exists a code $\{(\varphi_n^{(k)}, \psi_n^{(k)})\}_{k\in\mathbb{N}}$ satisfying

$$\Pr\{X^n \notin \mathcal{D}_n^{(k)}\} \leq \Pr\left\{ \frac{1}{n} \log \frac{1}{P_{X^n}(X^n)} \geq \overline{H}(\boldsymbol{X}) + \gamma_k \right\}, \tag{30}$$

where $\mathcal{D}_n^{(k)} = \{x^n \in \mathcal{X}^n : \psi_n^{(k)}(\varphi_n^{(k)}(x^n)) = x^n\}$. Since the right hand side of (30) converges to 0 as $n \to \infty$ for each $k \in \mathbb{N}$ from the definition of $\overline{H}(\boldsymbol{X})$ in (1) and the fact that $\gamma_k > 0$, $\Pr\{X^n \notin \mathcal{D}_n^{(k)}\} \leq \varepsilon_k$ for all sufficiently large n. Set

$$\mathcal{N}_k = \left\{ n \in \mathbb{N} : \Pr\left\{ \frac{1}{n} \log \frac{1}{P_{X^n}(X^n)} \geq \overline{H}(\boldsymbol{X}) + \gamma_k \right\} \leq \varepsilon_k \right\}$$

and define a sequence $\{N_k\}_{k=0}^{\infty}$ by $N_0 = 1$ and

$$N_k = \min\left\{ N > N_{k-1} : n \in \mathcal{N}_k \text{ for all } n \geq N \right\} \tag{31}$$

for $k \geq 1$. Notice that for any $k \in \mathbb{N}$ there exists an N such that $n \in \mathcal{N}_k$ for all $n \geq N$ since (30) holds for *all* sufficiently large n. Clearly, $\{N_k\}_{k=0}^{\infty}$ is strictly monotone increasing and satisfies $N_k \to \infty$ as $k \to \infty$. Then we can define the encoder and the decoder by $\varphi_n = \varphi_n^{(k)}$ and $\psi_n = \psi_n^{(k)}$, respectively, where k is the nonnegative integer satisfying $N_k \leq n < N_{k+1}$. Clearly, this code $\boldsymbol{C} = \{(\varphi_n, \psi_n)\}_{n\in\mathbb{N}}$ satisfies (10) and (27) and therefore is $\overline{H}$-optimal.

From the assumption of the theorem, any $\overline{H}$-optimal code is W-optimal. Therefore, letting $\gamma > 0$ be an arbitrary constant, the code $\boldsymbol{C}$ satisfies

$$\lim_{n\to\infty} \Pr\left\{ \frac{1}{n} \log M_n - \frac{1}{n} \log \frac{1}{P_{X^n}(X^n)} \geq W(\boldsymbol{X}) + \gamma \right\} = 0. \tag{32}$$

Then, it follows from (27) and (32) that

$$\lim_{n\to\infty} \Pr\left\{ \frac{1}{n} \log \frac{1}{P_{X^n}(X^n)} \leq \overline{H}(\boldsymbol{X}) - W(\boldsymbol{X}) - \gamma \right\} = 0.$$

Notice that due to the definition of $\underline{H}(\boldsymbol{X})$ in (2), it holds that $\overline{H}(\boldsymbol{X}) - W(\boldsymbol{X}) - \gamma \leq \underline{H}(\boldsymbol{X})$, i.e., $W(\boldsymbol{X}) \geq \overline{H}(\boldsymbol{X}) - \underline{H}(\boldsymbol{X}) - \gamma$. Since $\gamma > 0$ can be arbitrarily small, $W(\boldsymbol{X}) \geq \overline{H}(\boldsymbol{X}) - \underline{H}(\boldsymbol{X})$ is established. Q.E.D.

4.3 Condition for $\mathcal{C}_W(\boldsymbol{X}) \subseteq \mathcal{C}_{\overline{H}}(\boldsymbol{X})$

Next, we investigate the condition for $\mathcal{C}_W(\boldsymbol{X}) \subseteq \mathcal{C}_{\overline{H}}(\boldsymbol{X})$. We show that $\mathcal{C}_W(\boldsymbol{X}) \subseteq \mathcal{C}_{\overline{H}}(\boldsymbol{X})$ if and only if the equality holds in (6).

Theorem 4.3: If $\boldsymbol{X}$ satisfies $W(\boldsymbol{X}) = \overline{H}(\boldsymbol{X}) - \underline{H}^*(\boldsymbol{X})$, then $\mathcal{C}_W(\boldsymbol{X}) \subseteq \mathcal{C}_{\overline{H}}(\boldsymbol{X})$.

Proof: Fix a constant $\gamma > 0$ arbitrarily. Let $\boldsymbol{C}$ be an arbitrary W-optimal code. Then, $\boldsymbol{C}$ satisfies (10) and

$$\lim_{n\to\infty} \Pr\left\{\frac{1}{n}\log M_n - \frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \le W(\boldsymbol{X}) + \gamma\right\} = 1$$

for any $\gamma > 0$. Due to the assumption of the theorem, we have

$$\lim_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \ge \frac{1}{n}\log M_n + \underline{H}^*(\boldsymbol{X}) - \overline{H}(\boldsymbol{X}) - \gamma\right\} = 1. \quad (33)$$

Define $A = \limsup_{n\to\infty} \frac{1}{n}\log M_n$. Then, it holds that

$$\frac{1}{n}\log M_n \ge A - \gamma \quad \text{infinitely often.} \quad (34)$$

By using (33) and (34), it is not hard to verify that

$$\limsup_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \ge A + \underline{H}^*(\boldsymbol{X}) - \overline{H}(\boldsymbol{X}) - 2\gamma\right\} = 1.$$

In view of the definition of $\underline{H}^*(\boldsymbol{X})$ in (3), we have

$$A + \underline{H}^*(\boldsymbol{X}) - \overline{H}(\boldsymbol{X}) - 2\gamma \le \underline{H}^*(\boldsymbol{X}),$$

which yields

$$\limsup_{n\to\infty} \frac{1}{n}\log M_n \le \overline{H}(\boldsymbol{X}) + 2\gamma.$$

Since $\gamma > 0$ can be arbitrarily small, it holds that

$$\limsup_{n\to\infty} \frac{1}{n}\log M_n \le \overline{H}(\boldsymbol{X}). \quad (35)$$

Since $\boldsymbol{C}$ satisfies both (10) and (35), the code turns out to be $\overline{H}$-optimal. Q.E.D.

Theorem 4.4: If $\mathcal{C}_W(\boldsymbol{X}) \subseteq \mathcal{C}_{\overline{H}}(\boldsymbol{X})$, then $\boldsymbol{X}$ satisfies $W(\boldsymbol{X}) = \overline{H}(\boldsymbol{X}) - \underline{H}^*(\boldsymbol{X})$.

We actually establish the following Proposition 4.1 instead of proving Theorem 4.4 directly. This is because the combination of the contraposition of Proposition 4.1 with (6) lead to the claim of Theorem 4.4.

Proposition 4.1: If $\boldsymbol{X}$ satisfies $W(\boldsymbol{X}) > \overline{H}(\boldsymbol{X}) - \underline{H}^*(\boldsymbol{X})$, then there exists a code $\boldsymbol{C}$ satisfying $\boldsymbol{C} \in \mathcal{C}_W(\boldsymbol{X})$ and $\boldsymbol{C} \notin \mathcal{C}_{\overline{H}}(\boldsymbol{X})$.

In the following, we construct the code $\boldsymbol{C}$ in the claim of Proposition 4.1. Before describing the construction, we give four lemmas for clarifying the key ideas in the construction.

Fix sequences $\{\gamma_k\}_{k\in\mathbb{N}}$ and $\{\varepsilon_k\}_{k\in\mathbb{N}}$ satisfying (28) and (29) arbitrarily. From the definition of $\underline{H}^*(\boldsymbol{X})$ in (3), it holds for any constant $\gamma > 0$ and $\varepsilon \in (0,1)$ that

$$\Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \le \underline{H}^*(\boldsymbol{X}) - \gamma\right\} \le \varepsilon \quad \text{infinitely often.}$$

Defining $\mathcal{N}_k$ by

$$\mathcal{N}_k = \left\{n \in \mathbb{N} : \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \le \underline{H}^*(\boldsymbol{X}) - \gamma_k\right\} \le \varepsilon_k\right\}$$

for $k \in \mathbb{N}$, we have Lemmas 4.1–4.3.

Lemma 4.1: $\mathcal{N}_{k+1} \subseteq \mathcal{N}_k$ for all $k \in \mathbb{N}$.

Proof: Fix $k \in \mathbb{N}$ arbitrarily and assume that $n \in \mathcal{N}_{k+1}$. Then it follows from (28) and (29) that

$$\Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \le \underline{H}^*(\boldsymbol{X}) - \gamma_k\right\} \le \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \le \underline{H}^*(\boldsymbol{X}) - \gamma_{k+1}\right\} \le \varepsilon_{k+1} \le \varepsilon_k,$$

which means $n \in \mathcal{N}_k$. Therefore, $\mathcal{N}_{k+1} \subseteq \mathcal{N}_k$ for all $k \in \mathbb{N}$ follows. Q.E.D.

Lemma 4.2: There exists a strictly monotone increasing sequence $\{n_k\}_{k\in\mathbb{N}}$ of positive integers satisfying $\{n_j : j \ge k\} \subseteq \mathcal{N}_k$ for all $k \in \mathbb{N}$.

Proof: Define $n_0 = 0$ and

$$n_k = \min\{n \in \mathcal{N}_k : n > n_{k-1}\} \quad \text{for } k \ge 1. \quad (36)$$

Note that, since $\mathcal{N}_k$ is a countably infinite set, for each $k \ge 1$ n_k is well-defined. That is, if there is no $n \in \mathcal{N}_k$ satisfying $n > n_{k-1}$, $\mathcal{N}_k$ turns out to be a finite set, which is a contradiction. Therefore, the claim of this lemma follows because $n_k \in \mathcal{N}_k$ and $\mathcal{N}_{k+1} \subseteq \mathcal{N}_k$ for $k \in \mathbb{N}$. Q.E.D.

Lemma 4.3: The sequence defined by (36) satisfies

$$\lim_{k\to\infty} \Pr\left\{\frac{1}{n_k}\log\frac{1}{P_{X^{n_k}}(X^{n_k})} \le \underline{H}^*(\boldsymbol{X}) - \gamma\right\} = 0. \quad (37)$$

Proof: Letting $\varepsilon \in (0,1)$ be an arbitrary constant, we prove

$$\Pr\left\{\frac{1}{n_k}\log\frac{1}{P_{X^{n_k}}(X^{n_k})} \le \underline{H}^*(\boldsymbol{X}) - \gamma\right\} \le \varepsilon$$

for all sufficiently large k. In view of the definition of $\{\varepsilon_k\}_{k\in\mathbb{N}}$, for any $\varepsilon \in (0,1)$ we can define k_0 as the minimum integer $k \geq 1$ satisfying $\varepsilon_k \leq \varepsilon$. This implies that

$$\Pr\left\{\frac{1}{n_k}\log\frac{1}{P_{X^{n_k}}(X^{n_k})} \leq \underline{H}^*(\boldsymbol{X}) - \gamma\right\} \leq \varepsilon_k \leq \varepsilon_{k_0} \leq \varepsilon$$

for all $k \geq k_0$. Q.E.D.

The following lemma plays a key role in the construction of a code $\boldsymbol{C}$ satisfying $\boldsymbol{C} \in \mathcal{C}_W(\boldsymbol{X})$ and $\boldsymbol{C} \notin \mathcal{C}_{\overline{H}}(\boldsymbol{X})$.

Lemma 4.4: There exists a sequence of intervals $\{(a_n, b_n)\}_{n\in\mathbb{N}} \in \mathcal{G}$ and a monotone increasing function $\kappa : \mathbb{N} \to \mathbb{N}\cup\{0\}$ such that for any $\gamma > 0$ and any $\{\gamma_k\}_{k\in\mathbb{N}}$ satisfying (28) it holds that

$$\limsup_{n\to\infty}(b_n - a_n) \leq W(\boldsymbol{X}) + \frac{\gamma}{3}, \tag{38}$$

$$\lim_{n\to\infty}\Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \leq a_n - \gamma\right\} = 0, \tag{39}$$

$$\lim_{n\to\infty}\Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \geq b_n + \gamma_{\kappa(n)}\right\} = 0, \tag{40}$$

$$\lim_{n\to\infty}\kappa(n) = \infty. \tag{41}$$

Proof: Fix $\gamma > 0$, $\{\gamma_k\}_{k\in\mathbb{N}}$ satisfying (28) and $\{\varepsilon_k\}_{k\in\mathbb{N}}$ satisfying (29) arbitrarily. It is obvious from (4) that there exists $\{(a_n, b_n)\}_{n\in\mathbb{N}} \in \mathcal{G}$ satisfying (38). Since we have

$$\lim_{n\to\infty}\Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \in (a_n - \gamma, b_n + \gamma)\right\} = 1,$$

(39) clearly holds. In the following, (40) and (41) are proved by using an argument similar to the proof of Theorem 4.2.

Since it holds that $\{(a_n, b_n)\}_{n\in\mathbb{N}} \in \mathcal{G}$, for each $k \geq 1$

$$\Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \in (a_n - \gamma_k, b_n + \gamma_k)\right\} \geq 1 - \varepsilon_k \tag{42}$$

is satisfied for all sufficiently large n. Set

$$\mathcal{N}_k = \left\{n \in \mathbb{N} : \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \in (a_n - \gamma_k, b_n + \gamma_k)\right\} \geq 1 - \varepsilon_k\right\}$$

and define a sequence $\{N_k\}_{k=0}^{\infty}$ by $N_0 = 1$ and (31) for $k \geq 1$. Notice that $\{N_k\}_{k=0}^{\infty}$ is strictly monotone increasing and satisfies $N_k \to \infty$ as $k \to \infty$. Then, for each $n \in \mathbb{N}$ we can find $k \geq 0$ satisfying $N_k \leq n < N_{k+1}$. Define $\kappa(n)$ by $\kappa(n) = k$. Then we have (41). Setting $k = \kappa(n)$ in (42), it holds that

$$\Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \in (a_n - \gamma_{\kappa(n)}, b_n + \gamma_{\kappa(n)})\right\} \geq 1 - \varepsilon_{\kappa(n)} \quad \text{for all } n \in \mathbb{N},$$

which yields (40) because the right hand side of the above inequality converges to 1 as $n \to \infty$. Q.E.D.

Now we are ready to prove Proposition 4.1.

Proof of Proposition 4.1: In this proof we construct a code $\boldsymbol{C}$ which is W-optimal but not $\overline{H}$-optimal under the assumption of $W(\boldsymbol{X}) > \overline{H}(\boldsymbol{X}) - \underline{H}^*(\boldsymbol{X})$. We use $\{(a_n, b_n)\}_{n\in\mathbb{N}} \in \mathcal{G}$ in Lemma 4.4.

(a) Construction of a code: Define

$$\frac{1}{n}\log M_n = \begin{cases} \underline{H}^*(\boldsymbol{X}) + W(\boldsymbol{X}) & \text{if } n \in \mathcal{N}, \\ b_n + \gamma_{\kappa(n)} & \text{if } n \notin \mathcal{N}, \end{cases}$$

where $\mathcal{N} = \{n_k : k \in \mathbb{N}\}$ is the set of integers defined in Lemma 4.2. Then, [4, Lemma 1.3.1] guarantees the existence of a code $\boldsymbol{C} = \{(\varphi_n, \psi_n)\}_{n\in\mathbb{N}}$ satisfying

$$\varepsilon_n \leq \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \geq \frac{1}{n}\log M_n\right\}. \tag{43}$$

Hereinafter, we focus on the case where not only $\mathcal{N}$ but also $\mathbb{N}\backslash\mathcal{N}$ is a countably infinite set. As is obvious from the proof below, the proof becomes simpler if $\mathbb{N}\backslash\mathcal{N}$ is a finite set.

(b) Error probability: Since $W(\boldsymbol{X}) > \overline{H}(\boldsymbol{X}) - \underline{H}^*(\boldsymbol{X})$ from the assumption of the proposition, there exists a constant $\delta_0 > 0$ satisfying $W(\boldsymbol{X}) = \overline{H}(\boldsymbol{X}) - \underline{H}^*(\boldsymbol{X}) + \delta_0$. In the following, we evaluate the error probability using (43). Let $\varepsilon \in (0, 1)$ be an arbitrary constant. If $n \in \mathcal{N}$, we have

$$\begin{aligned}\varepsilon_n &\leq \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \geq \underline{H}^*(\boldsymbol{X}) + W(\boldsymbol{X})\right\} \\ &= \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \geq \overline{H}(\boldsymbol{X}) + \delta_0\right\},\end{aligned}$$

which is smaller than ε for sufficiently large $n \in \mathcal{N}$ from (1). On the other hand, if $n \notin \mathcal{N}$, we have

$$\varepsilon_n \leq \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \geq b_n + \gamma_{\kappa(n)}\right\},$$

which is smaller than ε for sufficiently large $n \notin \mathcal{N}$ from (40). Combining both cases, the error probability is bounded by ε for all sufficiently large n. Since $\varepsilon \in (0, 1)$ is arbitrary, this means that the error probability of this code converges to 0 as $n \to \infty$.

(c) Redundancy: Fix $\gamma > 0$ and $\varepsilon \in (0, 1)$ arbitrarily.

In the case of $n \in \mathcal{N}$,

$$\begin{aligned}&\Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \leq \frac{1}{n}\log M_n - W(\boldsymbol{X}) - \gamma\right\} \\ &= \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \leq \underline{H}^*(\boldsymbol{X}) - \gamma\right\}\end{aligned}$$

is satisfied. It is proved in Lemma 4.3 that this value is bounded by ε for all sufficiently large $n \in \mathcal{N}$.

In the case of $n \notin \mathcal{N}$, since $\{(a_n, b_n)\}_{n\in\mathbb{N}} \in \mathcal{G}$ satisfies (38) for any $\gamma > 0$, there exists an integer n_0 such that

$$b_n - a_n \leq W(\boldsymbol{X}) + \frac{2\gamma}{3} \quad \text{for all } n \geq n_0.$$

For the same n, we have

$$\begin{aligned}&\Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \leq \frac{1}{n}\log M_n - W(\boldsymbol{X}) - \gamma\right\} \\ &= \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \leq b_n + \gamma_{\kappa(n)} - W(\boldsymbol{X}) - \gamma\right\}\end{aligned}$$

$$\le \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \le a_n + \gamma_{\kappa(n)} - \frac{\gamma}{3}\right\}$$
$$\le \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \le a_n - \frac{\gamma}{6}\right\}, \tag{44}$$

where the last inequality is obtained from $\gamma_{\kappa(n)} < \gamma/6$ for all sufficiently large n because of $\gamma_{\kappa(n)} \to 0$ as $n \to \infty$. It is clear from (39) that the right hand side of (44) is bounded by ε for all sufficiently large n. Combining both cases, it is proved for any $\varepsilon > 0$ and sufficiently large n that

$$\Pr\left\{\frac{1}{n}\log M_n - \frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \ge W(\boldsymbol{X}) + \gamma\right\} \le \varepsilon.$$

From the above arguments on the error probability and redundancy, the code $\boldsymbol{C}$ is proved to be W-optimal.

(d) Coding Rate: From the definition of M_n, we have

$$\limsup_{n\to\infty}\frac{1}{n}\log M_n \ge \limsup_{k\to\infty}\frac{1}{n_k}\log M_{n_k}$$
$$= \underline{H}^*(\boldsymbol{X}) + W(\boldsymbol{X}). \tag{45}$$

Note that Lemma 4.2 guarantees that $\mathcal{N}$ is a countably infinite set. Since the assumption of the theorem means that $W(\boldsymbol{X}) = \overline{H}(\boldsymbol{X}) - \underline{H}^*(\boldsymbol{X}) + \delta_0$ for some $\delta_0 > 0$, it follows from (45) that

$$\limsup_{n\to\infty}\frac{1}{n}\log M_n \ge \overline{H}(\boldsymbol{X}) + \delta_0 > \overline{H}(\boldsymbol{X}),$$

which means that the code $\boldsymbol{C}$ is not $\overline{H}$-optimal. Q.E.D.

4.4 Condition for $C_W(\boldsymbol{X}) = C_{\overline{H}}(\boldsymbol{X})$

From Theorems 4.1, 4.3, 4.2, and 4.4, we can immediately obtain the following corollary.

Corollary 4.1: $C_W(\boldsymbol{X}) = C_{\overline{H}}(\boldsymbol{X})$ if and only if $\underline{H}(\boldsymbol{X}) = \underline{H}^*(\boldsymbol{X})$.

Proof: If $\underline{H}(\boldsymbol{X}) = \underline{H}^*(\boldsymbol{X})$, we have $W(\boldsymbol{X}) = \overline{H}(\boldsymbol{X}) - \underline{H}(\boldsymbol{X}) = \overline{H}(\boldsymbol{X}) - \underline{H}^*(\boldsymbol{X})$. Then, Theorems 4.1 and 4.3 guarantee that $C_W(\boldsymbol{X}) = C_{\overline{H}}(\boldsymbol{X})$.

Conversely, if $C_W(\boldsymbol{X}) = C_{\overline{H}}(\boldsymbol{X})$, Theorems 4.2 and 4.4 tell us that $W(\boldsymbol{X}) = \overline{H}(\boldsymbol{X}) - \underline{H}(\boldsymbol{X}) = \overline{H}(\boldsymbol{X}) - \underline{H}^*(\boldsymbol{X})$, which immediately yields $\underline{H}(\boldsymbol{X}) = \underline{H}^*(\boldsymbol{X})$. Q.E.D.

4.5 Optimal Code with Respect to Both the Coding Rate and the Redundancy

In this subsection, given a source $\boldsymbol{X}$ satisfying $\overline{H}(\boldsymbol{X}) < \infty$, we show that there exists a code which is both W-optimal and $\overline{H}$-optimal. Before proving the theorem, we give a lemma used in the proof of the theorem.

Lemma 4.5: For any element $\{(a_n, b_n)\}_{n\in\mathbb{N}}$ of $\mathcal{G}$, define

$$\tilde{a}_n = \min\{a_n, \overline{H}(\boldsymbol{X})\} \text{ and } \tilde{b}_n = \min\{b_n, \overline{H}(\boldsymbol{X})\}.$$

Then, it holds that $\{(\tilde{a}_n, \tilde{b}_n)\}_{n\in\mathbb{N}} \in \mathcal{G}$ and

$$\tilde{b}_n - \tilde{a}_n \le b_n - a_n \quad \text{for all } n \in \mathbb{N}.$$

Proof: Since $\{(a_n, b_n)\}_{n\in\mathbb{N}} \in \mathcal{G}$ is satisfied, it holds that $a_n \le b_n$ for all $n \ge 1$. It is easily verified that $\tilde{a}_n \le \tilde{b}_n$ and $\tilde{b}_n - \tilde{a}_n \le b_n - a_n$ for each $n \in \mathbb{N}$. In fact, we have only to treat the three cases $\overline{H}(\boldsymbol{X}) \le a_n \le b_n$, $a_n \le \overline{H}(\boldsymbol{X}) \le b_n$ and $a_n \le b_n \le \overline{H}(\boldsymbol{X})$ separately.

In the following, $\{(\tilde{a}_n, \tilde{b}_n)\}_{n\in\mathbb{N}} \in \mathcal{G}$ is proved. Let $\{(a_n, b_n)\}_{n\in\mathbb{N}} \in \mathcal{G}$ be an arbitrary sequence of intervals, and $\gamma > 0$ an arbitrary constant. Define the sets $\mathcal{A}_n$ and $\mathcal{B}_n$ by (16) and (17), respectively. Since $\{(a_n, b_n)\}_{n\in\mathbb{N}} \in \mathcal{G}$, we have

$$\lim_{n\to\infty}\Pr\{X^n \in \mathcal{A}_n\} = \lim_{n\to\infty}\Pr\{X^n \in \mathcal{B}_n\} = 0. \tag{46}$$

With defining $\mathcal{H}_n$ by

$$\mathcal{H}_n = \left\{x^n \in \mathcal{X}^n : \frac{1}{n}\log\frac{1}{P_{X^n}(x^n)} \ge \overline{H}(\boldsymbol{X}) + \gamma\right\},$$

it holds from the definition of $\overline{H}(\boldsymbol{X})$ in (1) that

$$\lim_{n\to\infty}\Pr\{X^n \in \mathcal{H}_n\} = 0. \tag{47}$$

Furthermore, define

$$\widetilde{\mathcal{A}}_n = \left\{x^n \in \mathcal{X}^n : \frac{1}{n}\log\frac{1}{P_{X^n}(x^n)} \le \tilde{a}_n - \gamma\right\}, \tag{48}$$
$$\widetilde{\mathcal{B}}_n = \left\{x^n \in \mathcal{X}^n : \frac{1}{n}\log\frac{1}{P_{X^n}(x^n)} \ge \tilde{b}_n + \gamma\right\}. \tag{49}$$

Since $\tilde{a}_n \le a_n$ from the definition, we have $\widetilde{\mathcal{A}}_n \subseteq \mathcal{A}_n$. Thus, it follows from (46) that

$$\lim_{n\to\infty}\Pr\{X^n \in \widetilde{\mathcal{A}}_n\} \le \lim_{n\to\infty}\Pr\{X^n \in \mathcal{A}_n\} = 0. \tag{50}$$

In addition, since either $\widetilde{\mathcal{B}}_n = \mathcal{B}_n$ or $\widetilde{\mathcal{B}}_n = \mathcal{H}_n$ is satisfied from the definition of $\tilde{b}_n$, it holds that $\widetilde{\mathcal{B}}_n \subseteq \mathcal{B}_n \cup \mathcal{H}_n$. Hence,

$$\lim_{n\to\infty}\Pr\{X^n \in \widetilde{\mathcal{B}}_n\} \le \lim_{n\to\infty}\Pr\{X^n \in \mathcal{B}_n \cup \mathcal{H}_n\}$$
$$\le \lim_{n\to\infty}\Pr\{X^n \in \mathcal{B}_n\} + \lim_{n\to\infty}\Pr\{X^n \in \mathcal{H}_n\}$$
$$= 0, \tag{51}$$

where the second inequality follows from the union bound, and the equality is obtained from (46) and (47). Using (50), (51) and the union bound, we have

$$\lim_{n\to\infty}\Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \in (\tilde{a}_n - \gamma, \tilde{b}_n + \gamma)\right\}$$
$$= 1 - \lim_{n\to\infty}\Pr\{X^n \in \widetilde{\mathcal{A}}_n \cup \widetilde{\mathcal{B}}_n\}$$
$$\ge 1 - \lim_{n\to\infty}\Pr\{X^n \in \widetilde{\mathcal{A}}_n\} - \lim_{n\to\infty}\Pr\{X^n \in \widetilde{\mathcal{B}}_n\} = 1,$$

which implies that $\{(\tilde{a}_n, \tilde{b}_n)\}_{n\in\mathbb{N}} \in \mathcal{G}$. Q.E.D.

Using this lemma, we can prove the following theorem.

Theorem 4.5: There exists an FF code which is both W-optimal and $\overline{H}$-optimal.

Proof: Fix $\gamma > 0$ arbitrarily. First we show the existence of a code $\boldsymbol{C}$ satisfying

$$\limsup_{n\to\infty} \frac{1}{n} \log M_n \le \overline{H}(\boldsymbol{X}) + \gamma,$$

$$\lim_{n\to\infty} \Pr\left\{\frac{1}{n} \log M_n - \frac{1}{n} \log \frac{1}{P_{X^n}(X^n)} \le W(\boldsymbol{X}) + 5\gamma\right\} = 1$$

$$\text{for any } \gamma > 0,$$

$$\lim_{n\to\infty} \varepsilon_n = 0.$$

From the definition of $W(\boldsymbol{X})$, there exists a sequence of intervals $\{(a_n, b_n)\}_{n\in\mathbb{N}} \in \mathcal{G}$ satisfying

$$\limsup_{n\to\infty} (b_n - a_n) \le W(\boldsymbol{X}) + \gamma.$$

We construct another sequence of intervals $\{(\tilde{a}_n, \tilde{b}_n)\}_{n\in\mathbb{N}} \in \mathcal{G}$ from $\{(a_n, b_n)\}_{n\in\mathbb{N}} \in \mathcal{G}$ in the same manner as in Lemma 4.5. Then, it holds that

$$\limsup_{n\to\infty} (\tilde{b}_n - \tilde{a}_n) \le W(\boldsymbol{X}) + \gamma, \tag{52}$$

$$\limsup_{n\to\infty} \tilde{b}_n \le \overline{H}(\boldsymbol{X}), \tag{53}$$

$$\limsup_{n\to\infty} \Pr\{X^n \in \widetilde{\mathcal{A}}_n\} = 0, \tag{54}$$

$$\limsup_{n\to\infty} \Pr\{X^n \in \widetilde{\mathcal{B}}_n\} = 0, \tag{55}$$

where $\widetilde{\mathcal{A}}_n$ and $\widetilde{\mathcal{B}}_n$ are defined in (48) and (49), respectively. Define $M_n = \exp(\lceil n(\tilde{b}_n+\gamma)\rceil)$. Then, from [4, Lemma 1.3.1], there exists a code such that the coding rate is equal to $\frac{1}{n} \log M_n$ and the error probability satisfies (43). The limit of the error probability is bounded as follows:

$$\begin{aligned}\limsup_{n\to\infty} \varepsilon_n &\le \limsup_{n\to\infty} \Pr\left\{\frac{1}{n} \log \frac{1}{P_{X^n}(X^n)} \ge \frac{1}{n} \log M_n\right\} \\ &\le \limsup_{n\to\infty} \Pr\{X^n \in \widetilde{\mathcal{B}}_n\} = 0,\end{aligned}$$

where the equality follows from (55). In addition, by using (53), the coding rate is bounded as

$$\limsup_{n\to\infty} \frac{1}{n} \log M_n \le \limsup_{n\to\infty} \left(\tilde{b}_n + \gamma + \frac{1}{n}\right) \le \overline{H}(\boldsymbol{X}) + \gamma.$$

Furthermore, since the individual redundancy of $x^n \notin \widetilde{\mathcal{A}}_n$ is evaluated as

$$\begin{aligned}\frac{1}{n} \log M_n - \frac{1}{n} \log \frac{1}{P_{X^n}(X^n)} &\le \left(\tilde{b}_n + \gamma + \frac{1}{n}\right) - (\tilde{a}_n - \gamma) \\ &= (\tilde{b}_n - \tilde{a}_n) + 3\gamma\end{aligned}$$

for all sufficiently large n, the combination of (52) with (54) yields

$$\lim_{n\to\infty} \Pr\left\{\frac{1}{n} \log M_n - \frac{1}{n} \log \frac{1}{P_{X^n}(X^n)} \le W(\boldsymbol{X}) + 5\gamma\right\} = 1.$$

Now we apply the diagonal line argument to obtain sharp bounds on the coding rate and the redundancy. Let $\{\gamma_k\}_{k\in\mathbb{N}}$ and $\{\varepsilon_k\}_{k\in\mathbb{N}}$ be arbitrary sequences satisfying (28) and (29), respectively. For each k define $M_n^{(k)} = \exp(\lceil n(\overline{H}(\boldsymbol{X})+\gamma_k)\rceil)$. Then there exists a code $\{(\varphi_n^{(k)}, \psi_n^{(k)})\}_{k\in\mathbb{N}}$ satisfying (30). Notice here that we can choose a strictly monotone increasing sequence $\{N_k\}_{k=0}^{\infty}$ in the same way as in the proof of Lemma 4.4. Thus for each $n \in \mathbb{N}$ we can define an encoder and a decoder by $\varphi_n = \varphi_n^{(k)}$ and $\psi_n = \psi_n^{(k)}$, where k is a unique integer satisfying $N_k \le n < N_{k+1}$. The obtained code $\boldsymbol{C}' = \{(\varphi_n, \psi_n)\}_{n\in\mathbb{N}}$ satisfies

$$\limsup_{n\to\infty} \frac{1}{n} \log M_n \le \overline{H}(\boldsymbol{X}),$$

$$\lim_{n\to\infty} \varepsilon_n = 0,$$

and

$$\lim_{n\to\infty} \Pr\left\{\frac{1}{n} \log M_n - \frac{1}{n} \log \frac{1}{P_{X^n}(X^n)} \le W(\boldsymbol{X}) + \gamma\right\} = 1$$

for any $\gamma > 0$, which establishes that $\boldsymbol{C}'$ is both W-optimal and $\overline{H}$-optimal. Q.E.D.

4.6 Examples of Optimal Codes

In this subsection, we show examples of codes which are optimal under either one of the criteria and a code which is optimal under both criteria. All of these codes exist if the source $\boldsymbol{X}$ satisfies both $W(\boldsymbol{X}) > \overline{H}(\boldsymbol{X}) - \underline{H}^*(\boldsymbol{X})$ and $W(\boldsymbol{X}) < \overline{H}(\boldsymbol{X}) - \underline{H}(\boldsymbol{X})$, which corresponds to the lower-right cell of Table 1.

Consider the source defined in Example 2.4. Then, (8) and (9) imply that there exist constants $\delta_1 > 0$, $\delta_2 > 0$ satisfying

$$W(\boldsymbol{X}) = \delta_1 \text{ and } W(\boldsymbol{X}) = \overline{H}(\boldsymbol{X}) - \underline{H}(\boldsymbol{X}) - \delta_2.$$

For this source, we give three kinds of codes below. The first one is W-optimal but not $\overline{H}$-optimal, the second one is $\overline{H}$-optimal but not W-optimal, and the last one is both $\overline{H}$-optimal and W-optimal. Note that since an example which is neither $\overline{H}$-optimal nor W-optimal code is trivial, we do not show it.

Since $P_{X_{1a}^n}$, $P_{X_{1b}^n}$, and $P_{X_2^n}$ in Example 2.4 are stationary and memoryless, due to the law of large numbers there exists a sequence $\{\gamma_n\}_{n\in\mathbb{N}}$ satisfying $\gamma_n \to 0$, $\sqrt{n}\gamma_n \to \infty$,

$$\Pr\left\{\frac{1}{n} \log \frac{1}{P_{X^n}(X^n)} \in (\underline{H}(\boldsymbol{X}) - \gamma_n, \overline{H}^*(\boldsymbol{X}) + \gamma_n)\right\} \to 1 \quad \text{for odd } n, \tag{56}$$

and

$$\Pr\left\{\frac{1}{n} \log \frac{1}{P_{X^n}(X^n)} \in (\underline{H}^*(\boldsymbol{X}) - \gamma_n, \overline{H}(\boldsymbol{X}) + \gamma_n)\right\} \to 1 \quad \text{for even } n. \tag{57}$$

Example 4.1 (W-optimal but not $\overline{H}$-optimal code): For γ_n satisfying (56) and (57), define M_n by

$$\frac{1}{n}\log M_n = \begin{cases} \overline{H}^*(\boldsymbol{X}) + \gamma_n & \text{if } n \text{ is odd,} \\ \overline{H}(\boldsymbol{X}) + W(\boldsymbol{X}) + \gamma_n & \text{if } n \text{ is even,} \end{cases}$$

and construct a code such that all x^n satisfying $P_{X^n}(x^n) \geq 1/M_n$ are correctly decoded for all $n \in \mathbb{N}$. This code satisfies (43).

Error probability: From (43),

$$\varepsilon_n \leq \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \geq \overline{H}^*(\boldsymbol{X}) + \gamma_n\right\} \to 0$$

is satisfied for odd n. For even n, it holds that

$$\begin{aligned}\varepsilon_n &\leq \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \geq \overline{H}(\boldsymbol{X}) + W(\boldsymbol{X}) + \gamma_n\right\} \\ &\leq \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \geq \overline{H}(\boldsymbol{X}) + \gamma_n\right\} \to 0.\end{aligned}$$

Therefore, we have $\varepsilon_n \to 0$ as $n \to \infty$.

Redundancy: For odd n, it holds that

$$\begin{aligned}&\Pr\left\{\frac{1}{n}\log M_n - \frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \geq W(\boldsymbol{X}) + 2\gamma_n\right\} \\ &\quad = \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \leq \underline{H}(\boldsymbol{X}) - \gamma_n\right\} \to 0.\end{aligned}$$

On the other hand, for even n we have

$$\begin{aligned}&\Pr\left\{\frac{1}{n}\log M_n - \frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \geq W(\boldsymbol{X}) + 2\gamma_n\right\} \\ &\quad = \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \leq \underline{H}^*(\boldsymbol{X}) - \gamma_n\right\} \to 0.\end{aligned}$$

Combining with the evaluation of the error probability, this code is proved to be W-optimal.

Coding Rate: For even n, this code satisfies

$$\frac{1}{n}\log M_n = \overline{H}(\boldsymbol{X}) + \delta_1 + \gamma_n.$$

Since $\delta_1 > 0$, this code is not $\overline{H}$-optimal.

Example 4.2 ($\overline{H}$-optimal but not W-optimal code): For γ_n satisfying (56) and (57), define M_n by

$$\frac{1}{n}\log M_n = \overline{H}(\boldsymbol{X}) + \gamma_n, \tag{58}$$

and construct a code such that any x^n satisfying $P_{X^n}(x^n) \geq 1/M_n$ are correctly decoded for all $n \in \mathbb{N}$. This code satisfies (43).

Error probability: From (43) it holds that

$$\varepsilon_n \leq \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \geq \overline{H}(\boldsymbol{X}) + \gamma_n\right\} \to 0.$$

Coding Rate: From (58), we have

$$\limsup_{n\to\infty}\frac{1}{n}\log M_n = \overline{H}(\boldsymbol{X}).$$

Thus, this code is proved to be $\overline{H}$-optimal.

Redundancy: For odd n, it holds that

$$\begin{aligned}&\Pr\left\{\frac{1}{n}\log M_n - \frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \geq W(\boldsymbol{X}) + 2\gamma_n\right\} \\ &\quad = \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \leq \underline{H}(\boldsymbol{X}) + \delta_2 - \gamma_n\right\}.\end{aligned}$$

Note that $\delta_2 - \gamma_n > 0$ is satisfied for sufficiently large n. Applying (2), the right hand side is positive for infinitely many n, which implies that this code is not W-optimal.

Example 4.3 (W-optimal and $\overline{H}$-optimal code): For γ_n satisfying (56) and (57), define M_n by

$$\frac{1}{n}\log M_n = \begin{cases} \overline{H}^*(\boldsymbol{X}) + \gamma_n, & \text{if } n \text{ is odd,} \\ \overline{H}(\boldsymbol{X}) + \gamma_n, & \text{if } n \text{ is even,} \end{cases}$$

and construct a code such that all x^n satisfying $P_{X^n}(x^n) \geq 1/M_n$ are correctly decoded for all $n \in \mathbb{N}$. This code satisfies (43).

Error probability: From (43),

$$\varepsilon_n \leq \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \geq \overline{H}^*(\boldsymbol{X}) + \gamma_n\right\} \to 0$$

is satisfied for odd n. For even n, it holds that

$$\varepsilon_n \leq \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \geq \overline{H}(\boldsymbol{X}) + \gamma_n\right\} \to 0.$$

Therefore, we have $\varepsilon_n \to 0$.

Redundancy: For odd n, it holds that

$$\begin{aligned}&\Pr\left\{\frac{1}{n}\log M_n - \frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \geq W(\boldsymbol{X}) + 2\gamma_n\right\} \\ &\quad = \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \leq \underline{H}(\boldsymbol{X}) - \gamma_n\right\} \to 0.\end{aligned}$$

On the other hand, it is satisfied for even n that

$$\begin{aligned}&\Pr\left\{\frac{1}{n}\log M_n - \frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \geq W(\boldsymbol{X}) + 2\gamma_n\right\} \\ &\quad \leq \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \leq \underline{H}^*(\boldsymbol{X}) - \gamma_n\right\} \to 0.\end{aligned}$$

Combining with the evaluation of the error probability, this code is proved to be W-optimal.

Coding Rate: For odd n, this code satisfies

$$\frac{1}{n}\log M_n = \overline{H}^*(\boldsymbol{X}) + \gamma_n.$$

For even n, this code satisfies

$$\frac{1}{n}\log M_n = \overline{H}(\boldsymbol{X}) + \gamma_n.$$

With the evaluation of the error probability, this code is proved to be $\overline{H}$-optimal.

5. Concluding Remarks

In this paper we have considered fixed-to-fixed length (FF) coding of a general source $\boldsymbol{X}$ satisfying $\overline{H}(\boldsymbol{X}) < \infty$ and investigated relationships of the two classes $C_{\overline{H}}(\boldsymbol{X})$ and $C_W(\boldsymbol{X})$ of the optimal FF codes, where $C_{\overline{H}}(\boldsymbol{X})$ and $C_W(\boldsymbol{X})$ denote the sets of the optimal codes in terms of the coding rate and the redundancy, respectively. The relationships are characterized by the asymptotic width $W(\boldsymbol{X})$ of the entropy spectrum of the source that satisfies $\overline{H}(\boldsymbol{X}) - \underline{H}^*(\boldsymbol{X}) \leq W(\boldsymbol{X}) \leq \overline{H}(\boldsymbol{X}) - \underline{H}(\boldsymbol{X})$ in general. It is shown that $C_W(\boldsymbol{X}) \subseteq C_{\overline{H}}(\boldsymbol{X})$ if and only if $W(\boldsymbol{X})$ coincides with the upper bound, while $C_{\overline{H}}(\boldsymbol{X}) \subseteq C_W(\boldsymbol{X})$ if and only if $W(\boldsymbol{X})$ coincides with the lower bound. These results immediately implies that $\underline{H}(\boldsymbol{X}) = \underline{H}^*(\boldsymbol{X})$ is a necessary and sufficient condition for $C_W(\boldsymbol{X}) = C_{\overline{H}}(\boldsymbol{X})$. It has also proved that $C_{\overline{H}}(\boldsymbol{X}) \cap C_W(\boldsymbol{X}) \neq \emptyset$ for general sources satisfying $\overline{H}(\boldsymbol{X}) < \infty$.

Since the necessary and sufficient condition $\underline{H}(\boldsymbol{X}) = \underline{H}^*(\boldsymbol{X})$ means that the left endpoint of the entropy spectrum of $\boldsymbol{X}$ converges, we are interested in the property of FF codes that corresponding to $\overline{H}(\boldsymbol{X}) = \overline{H}^*(\boldsymbol{X})$ as well, where $\overline{H}(\boldsymbol{X}) = \overline{H}^*(\boldsymbol{X})$ means that the right endpoint the entropy spectrum converges. It is shown [1] that $\overline{H}(\boldsymbol{X}) = \overline{H}^*(\boldsymbol{X})$ is actually a necessary and sufficient condition under which the coding rate of all the codes in $C_{\overline{H}}(\boldsymbol{X})$ converges to $\overline{H}(\boldsymbol{X})$.

Acknowledgments

The authors wish to thank Prof. Shigeaki Kuzuoka in Wakayama University who gave constructive comments on the primal version of this paper presented in IEICE Technical Committee Meeting on Information Theory held in Iizuka, Fukuoka on May 25, 2012. The authors express their thanks to the associate editor and the two referees who carefully read the manuscript and gave valuable comments. This research is supported in part by MEXT Grant-in-Aid for Scientific Research(C) 24560482.

References

[1] M. Arimura, H. Koga, and K. Iwata, "A characterization of optimal FF coding rate using a new optimistically optimal code," IEICE Trans. Fundamentals, vol.E96–A, no.12, pp.2443–2446, Dec. 2013.
[2] P.-N. Chen and F. Alajaji, "Optimistic Shannon coding theorem for arbitrary single-user systems," IEEE Trans. Inf. Theory, vol.45, no.7, pp.2623–2629, Nov. 1999.
[3] T.M. Cover and J.A. Thomas, Elements of Information Theory, 2nd ed., Wiley, 2006.
[4] T.S. Han, Information-Spectrum Methods in Information Theory, Springer, 2003.
[5] T.S. Han and S. Verdú, "Approximation theory of output statistics," IEEE Trans. Inf. Theory, vol.39, no.3, pp.752–772, May 1993.
[6] H. Koga, "A coding theorem on the fixed-length homophonic coding for a general source," Proc. SITA2000, pp.133–136, Aso, 2000. (in Japanese)
[7] H. Koga, "A coding theorem on the fixed-length homophonic coding for a general source," Proc. ISIT2001, p.217, Washington, DC, June 2001.
[8] H. Koga and H. Yamamoto, "Asymptotic properties on codeword lengths of an optimal FV code for general sources," IEEE Trans. Inf. Theory, vol.51, no.4, pp.1546–1555, April 2005.
[9] H. Koga, "Four limits in probability and their roles in source coding," IEICE Trans. Fundamentals, vol.E94–A, no.11, pp.2073–2082, Nov. 2011.
[10] 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. Inf. Theory, vol.38, no.1, pp.66–72, Jan. 1992.
[11] F.M.J. Willems, Y.M. Shtarkov, and T.J. Tjalkens, "The context-tree weighting method: Basic properties," IEEE Trans. Inf. Theory, vol.41, no.3, pp.653–664, May 1995.
[12] E.-h. Yang and J.C. Kieffer, "Efficient universal lossless data compression algorithms based on a greedy sequential grammar transform — Part one: Without context models," IEEE Trans. Inf. Theory, vol.46, no.3, pp.755–777, May 2000.

Mitsuharu Arimura received B.E., M.E. and Ph.D. degrees from University of Tokyo, in 1994, 1996 and 1999, respectively. From 1999 to 2004, he was a Research Associate in the Graduate School of Information Systems at the University of Electro-Communications, Tokyo, Japan. Since 2004, he has been with Shonan Institute of Technology, where he is currently a Lecturer of Faculty of Engineering. His research interests include Shannon theory and data compression algorithms. Dr. Arimura is a member of the IEEE.

Hiroki Koga received B.E., M.E. and D.E. degrees from University of Tokyo, in 1990, 1992 and 1995, respectively. From 1995 to 1999, he was a Research Associate in Graduate School of Engineering, University of Tokyo. Since 1999, he has been with University of Tsukuba, where he is currently an Associate Professor of Graduate School of Systems and Information Engineering. His research interests are in Shannon theory and information theory. Dr. Koga is a member of the IEEE.

Ken-ichi Iwata received B.Ed. degree from Wakayama University in 1993, and M.Sc. degree in Information Science from Japan Advanced Institute of Science and Technology in 1995, and D.E. degree from the University of Electro-Communications in 2006. Since 2008, he has been with University of Fukui as an associate professor of Graduate School of Engineering. He is a member of the IEEE.