PAPER *Special Section on Information Theory and Its Applications*

Four Limits in Probability and Their Roles in Source Coding*

Hiroki KOGA[†a)], ***Senior Member***

SUMMARY In information-spectrum methods proposed by Han and Verdú, quantities defined by using the limit superior (or inferior) in probability play crucial roles in many problems in information theory. In this paper, we introduce two nonconventional quantities defined in probabilistic ways. After clarifying basic properties of these quantities, we show that the two quantities have operational meaning in the ε-coding problem of a general source in the ordinary and optimistic senses. The two quantities can be used not only for obtaining variations of the strong converse theorem but also establishing upper and lower bounds on the width of the entropy-spectrum. We also show that the two quantities are expressed in terms of the smooth Rényi entropy of order zero.

key words: *coding theorem, information-spectrum methods, optimistic coding, strong converse property, smooth Rényi entropy*

1. Introduction

In information-spectrum methods originating from a seminal paper by Han and Verdú [4], quantities defined by the limit inferior or superior in probability have operational meanings in coding problems of general sources and channels [3]. Given a general source $\boldsymbol{X} = \{X^n\}_{n=1}^{\infty}$ with a discrete alphabet $\mathcal{X}$, let us consider the situation where a source output X^n of blocklength n is encoded to a codeword $W_n = \varphi_n(X^n)$ by a fixed-to-fixed length code φ_n with M_n codewords. Here, X^n is regarded as a random variable taking values in $\mathcal{X}^n$. Suppose also that the codeword W_n is decoded to $\hat{X}^n$ by a decoder ψ_n. We are interested in the optimum value of the coding rate $\frac{1}{n}\log M_n$ that is asymptotically achievable for the pair (φ_n, ψ_n), $n \geq 1$, satisfying $\varepsilon_n \to 0$ as $n \to \infty$, where $\log(\cdot) = \log_2(\cdot)$ throughout this paper and $\varepsilon_n = \Pr\{\hat{X}^n \neq X^n\}$ denotes the decoding error probability. It is known that the optimal rate coincides with the spectral sup-entropy rate $\overline{H}(\boldsymbol{X})$ of the source $\boldsymbol{X}$, which is defined as the limit superior in probability of $\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)}$ [4]. Roughly speaking, $\overline{H}(\boldsymbol{X})$ can be interpreted as the right endpoint of the distribution of $\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)}$ (called the *entropy-spectrum*) in an asymptotic sense.

However, $\overline{H}(\boldsymbol{X})$ actually means a tight upper bound of the limit superior of the coding rates $\frac{1}{n}\log M_n$, $n \geq 1$, at which fixed-to-fixed coding with $\varepsilon_n \to 0$ as $n \to \infty$ is realizable. Hence, for some blocklength n there may exist a fixed-to-fixed length code such that its coding rate $\frac{1}{n}\log M_n$ is less than $\overline{H}(\boldsymbol{X})$ and attains a sufficiently small decoding error probability. Vembu and Verdú [16] defined the optimistic coding rate as the coding rate at which ε_n becomes sufficiently small for infinitely many n. Chen and Alajaji [2] give an upper and a lower bounds on the minimum attainable optimistic coding rate.

Similarly, $\underline{H}(\boldsymbol{X})$ is defined as the limit inferior in probability of $\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)}$ [4]. This $\underline{H}(\boldsymbol{X})$ corresponds to the left endpoint of the entropy-spectrum in an asymptotic sense. Han [3] gives the strong converse theorem which claims that ε_n goes to one for any fixed-to-fixed length code with the coding rate less than $\overline{H}(\boldsymbol{X})$ if and only if $\underline{H}(\boldsymbol{X}) = \overline{H}(\boldsymbol{X})$. In addition, $\underline{H}(\boldsymbol{X})$ itself is known as the supremum achievable rate of the intrinsic randomness problem [16].

In this paper we newly define $\overline{H}^*(\boldsymbol{X})$ and $\underline{H}^*(\boldsymbol{X})$ as variants of $\overline{H}(\boldsymbol{X})$ and $\underline{H}(\boldsymbol{X})$, respectively, and discuss roles of $\overline{H}(\boldsymbol{X})$, $\underline{H}(\boldsymbol{X})$, $\overline{H}^*(\boldsymbol{X})$ and $\underline{H}^*(\boldsymbol{X})$ in coding of a general source $\boldsymbol{X}$. In fact, $\overline{H}^*(\boldsymbol{X})$ and $\underline{H}^*(\boldsymbol{X})$ are defined as the right and left endpoints of the entropy-spectrum in different asymptotic senses, respectively. The roles of $\overline{H}(\boldsymbol{X})$ and $\underline{H}^*(\boldsymbol{X})$ become clear if we consider the infimum achievable rate $R_\varepsilon(\boldsymbol{X})$ of fixed-to-fixed length coding with ε-error in ordinary senses, where the decoding error probability ε_n is required to satisfy $\limsup_{n\to\infty}\varepsilon_n \leq \varepsilon$ for a constant $\varepsilon \in [0, 1)$. On the other hand, the role of $\overline{H}^*(\boldsymbol{X})$ and $\underline{H}(\boldsymbol{X})$ becomes more understandable if we consider the supremum *unachievable* rate $U_\varepsilon(\boldsymbol{X})$ with ε-error such that any fixed-to-fixed length code with rate less than $U_\varepsilon(\boldsymbol{X})$ satisfies $\liminf_{n\to\infty}\varepsilon_n > \varepsilon$. In particular, $U_\varepsilon(\boldsymbol{X})$ with $\varepsilon = 0$ turns out to be equal to the infimum achievable rate $R^*(\boldsymbol{X})$ with the vanishing decoding error probability in optimistic sense. We can regard the coding theorem on $U_\varepsilon(\boldsymbol{X})$ as a *dual* counterpart of the ordinary coding theorem on $R_\varepsilon(\boldsymbol{X})$. We should note that, while $\overline{H}^*(\boldsymbol{X})$ and $\underline{H}^*(\boldsymbol{X})$ coincide with $\overline{U}_{1^-}$ and $\underline{U}_0$ in [2], the main results in [2] are restricted to the formulas of the optimum attainable coding rates in the optimistic sense.

Using not only $\overline{H}(\boldsymbol{X})$ and $\underline{H}(\boldsymbol{X})$ but also $\overline{H}^*(\boldsymbol{X})$ and $\underline{H}^*(\boldsymbol{X})$ enriches terminologies for discussions of coding theorems in information-spectrum methods. We can easily obtain variations of the strong converse theorem of a general source [3]. While in this paper we can focus on general sources where two or three out of the four quantities coincide, the strong converse theorem in [3] corresponds to the case where all the four quantities coincide. Furthermore, we can discuss lower and upper bounds on the width

Manuscript received February 18, 2011.

Manuscript revised June 10, 2011.

†The author is with the Graduate School of Systems and Information Engineering, University of Tsukuba, Tsukuba-shi, 305-8573 Japan.

*The results of this paper were first presented in [8].

a) E-mail: koga@iit.tsukuba.ac.jp

DOI: 10.1587/transfun.E94.A.2073

$W(\boldsymbol{X})$ of the entropy-spectrum by using these four quantities [6]. In particular, the lower bound $W(\boldsymbol{X}) \geq \max\{\overline{H}(\boldsymbol{X}) - \underline{H}^*(\boldsymbol{X}), \overline{H}^*(\boldsymbol{X}) - \underline{H}(\boldsymbol{X})\}$ is first obtained in this paper. We can also show that the four quantities are expressed in terms of the smooth Rényi entropy [10], [11] of order zero, where the expressions for $\overline{H}(\boldsymbol{X})$ and $\underline{H}(\boldsymbol{X})$ are given in [14].

The rest of this paper is organized as follows. In Sect. 2 we formally define $\overline{H}^*(\boldsymbol{X})$ and $\underline{H}^*(\boldsymbol{X})$ as well as $\overline{H}(\boldsymbol{X})$ and $\underline{H}(\boldsymbol{X})$ and investigate their fundamental relationships. We unveil a necessary and sufficient condition under which $\underline{H}(\boldsymbol{X}) \leq \underline{H}^*(\boldsymbol{X}) \leq \overline{H}^*(\boldsymbol{X}) \leq \overline{H}(\boldsymbol{X})$ is satisfied. Section 3 is devoted to observation of the operational meanings of these four quantities. We see that their operational meanings are explained in the ε-source coding problems in the ordinary and optimistic senses. In Sect. 4 we discuss variations of the strong converse theorem [3]. The variations are obtained as easy consequences of the results in Sect. 3. We give different interpretations of the four quantities in terms of the smooth Rényi entropy of order zero in Sect. 5. In Sect. 6 we consider two kinds of width of the entropy-spectrum of a general source and establish their upper and lower bounds expressed in terms of $\overline{H}(\boldsymbol{X})$, $\underline{H}(\boldsymbol{X})$, $\overline{H}^*(\boldsymbol{X})$ and $\underline{H}^*(\boldsymbol{X})$.

2. Four Limits in Probability

Let $\mathcal{X}$ be a finite or countably infinite alphabet. For each $n \geq 1$ let X^n be a random variable taking values in $\mathcal{X}^n$. Denote by P_{X^n} the probability distribution of X^n. The probability of $X^n = x^n$ for an $x^n \in \mathcal{X}^n$ is denoted by $P_{X^n}(x^n)$, where throughout this paper random variables and their realizations are expressed italic upper-case and lower-case letters, respectively. We do not require that P_{X^n}, $n \geq 1$, satisfy the consistency condition. We call $\boldsymbol{X} = \{X^n\}_{n=1}^\infty$ a *general source* [3]. The class of general sources includes vast classes of sources such as memoryless sources, Markov sources, stationary ergodic sources and stationary sources.

In this paper we observe basic properties of the following four quantities:

Definition 2.1: Given a general source $\boldsymbol{X} = \{X^n\}_{n=1}^\infty$ we define:

$$\overline{H}(\boldsymbol{X}) = \inf\left\{\alpha \,:\, \liminf_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \leq \alpha\right\} = 1\right\},$$
$$\overline{H}^*(\boldsymbol{X}) = \inf\left\{\alpha \,:\, \limsup_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \leq \alpha\right\} = 1\right\},$$
$$\underline{H}(\boldsymbol{X}) = \sup\left\{\beta \,:\, \liminf_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \geq \beta\right\} = 1\right\},$$
$$\underline{H}^*(\boldsymbol{X}) = \sup\left\{\beta \,:\, \limsup_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \geq \beta\right\} = 1\right\},$$

where throughout this paper $\Pr\{\cdot\}$ means the probability with respect to P_{X^n} unless stated otherwise.

Note that $\overline{H}(\boldsymbol{X})$ (resp., $\underline{H}(\boldsymbol{X})$) is equal to the ordinary spectrum sup- (resp., inf-) entropy rate [3]. In addition, $\overline{H}^*(\boldsymbol{X})$ and $\underline{H}^*(\boldsymbol{X})$ coincide with $\overline{U}_{1^-}$ and $\underline{U}_0$ in [2, p. 2024], respectively. While it is known that $0 \leq \underline{H}(\boldsymbol{X}) \leq \overline{H}(\boldsymbol{X}) \leq \infty$ [3], we assume that $\overline{H}(\boldsymbol{X}) < \infty$ throughout this paper so that we can simplify arguments and clearly see essential contributions of this paper. Note that this assumption implies that all of $\underline{H}(\boldsymbol{X})$, $\underline{H}^*(\boldsymbol{X})$, $\overline{H}^*(\boldsymbol{X})$ and $\overline{H}(\boldsymbol{X})$ are finite due to Proposition 2.2 appearing afterwards.

We show that each of the four quantities can be expressed in a different form. These expressions will be used in the following sections. We give formal proofs in Appendix A for completeness of this paper.

Proposition 2.1: The four quantities in Definition 2.1 can equivalently be written as follows, respectively:

$$\overline{H}(\boldsymbol{X}) = \sup\left\{\beta \,:\, \limsup_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \geq \beta\right\} > 0\right\},$$
$$\overline{H}^*(\boldsymbol{X}) = \sup\left\{\beta \,:\, \liminf_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \geq \beta\right\} > 0\right\},$$
$$\underline{H}(\boldsymbol{X}) = \inf\left\{\alpha \,:\, \limsup_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \leq \alpha\right\} > 0\right\},$$
$$\underline{H}^*(\boldsymbol{X}) = \inf\left\{\alpha \,:\, \liminf_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \leq \alpha\right\} > 0\right\}.$$

It is important to notice that the definition of $\overline{H}^*(\boldsymbol{X})$ in Definition 2.1 is equivalent to

$$\limsup_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \leq \overline{H}^*(\boldsymbol{X}) + \gamma\right\} = 1, \quad (1)$$
$$\limsup_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \leq \overline{H}^*(\boldsymbol{X}) - \gamma\right\} < 1 \quad (2)$$

for any constant $\gamma > 0$. Similarly, the expression of $\overline{H}^*(\boldsymbol{X})$ in Proposition 2.1 is equivalent to

$$\liminf_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \geq \overline{H}^*(\boldsymbol{X}) - \gamma\right\} > 0, \quad (3)$$
$$\liminf_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \geq \overline{H}^*(\boldsymbol{X}) + \gamma\right\} = 0 \quad (4)$$

for any constant $\gamma > 0$. The relationships similar to (1)–(4) hold for $\overline{H}(\boldsymbol{X})$, $\underline{H}(\boldsymbol{X})$ and $\underline{H}^*(\boldsymbol{X})$ as well.

We give several general sources for which we can easily find values of the four quantities in Definition 2.1.

Example 2.1: Consider the case where $\boldsymbol{X} = \{X^n\}_{n=1}^\infty$ is the stationary memoryless source specified by a probability distribution P_X on $\mathcal{X}$ satisfying $H(P_X) < \infty$, where $H(P_X)$ denotes the entropy of P_X. We have $P_{X^n}(x^n) = \prod_{k=1}^n P_X(x_k)$ for all $n \geq 1$ and $x^n = x_1 x_2 \cdots x_n$ for this source. Then, it holds that

$$\overline{H}(\boldsymbol{X}) = \overline{H}^*(\boldsymbol{X}) = \underline{H}(\boldsymbol{X}) = \underline{H}^*(\boldsymbol{X}) = H(P_X)$$

owing to the weak law of large numbers.

Example 2.2: Next, consider the case where $\boldsymbol{X} = \{X^n\}_{n=1}^\infty$ is the mixed source [3] specified by two probability distributions P_{X_1} and P_{X_2} on $\mathcal{X}$ satisfying $H(P_{X_1}) < H(P_{X_2}) < \infty$. Define $P_{X_1^n}(x^n) = \prod_{k=1}^n P_{X_1}(x_k)$ and $P_{X_2^n}(x^n) = \prod_{k=1}^n P_{X_2}(x_k)$ for each $n \geq 1$ and $x^n \in \mathcal{X}^n$ and set

$$P_{X^n}(x^n) = (1-\tau)P_{X_1^n}(x^n) + \tau P_{X_2^n}(x^n) \quad \text{for all } n \geq 1,$$

where $\tau \in (0,1)$ is a constant. Then, it holds that

$$\overline{H}(\boldsymbol{X}) = \overline{H}^*(\boldsymbol{X}) = H(P_{X_2}) \text{ and } \underline{H}(\boldsymbol{X}) = \underline{H}^*(\boldsymbol{X}) = H(P_{X_1}).$$

In particular, $\overline{H}^*(\boldsymbol{X}) = H(P_{X_2})$ is verified from

$$\liminf_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \geq H(P_{X_2}) - \gamma\right\} \geq \tau > 0, \quad (5)$$

$$\liminf_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \geq H(P_{X_2}) + \gamma\right\} = 0 \quad (6)$$

for any constant $\gamma > 0$ and the expression of $\overline{H}^*(\boldsymbol{X})$ in Proposition 2.1.

Example 2.3: Consider the following perturbing source. Let $P_{X_1^n}(x^n)$ and $P_{X_2^n}(x^n)$ be the probability distributions given in Example 2.2 and define P_{X^n} by

$$P_{X^n}(x^n) = \begin{cases} P_{X_1^n}(x^n), & \text{if } n \text{ is odd,} \\ P_{X_2^n}(x^n), & \text{if } n \text{ is even.} \end{cases}$$

Then, it holds that

$$\overline{H}(\boldsymbol{X}) = \underline{H}^*(\boldsymbol{X}) = H(P_{X_2}) \text{ and } \underline{H}(\boldsymbol{X}) = \overline{H}^*(\boldsymbol{X}) = H(P_{X_1}).$$

Note that, unlike Example 2.2, the left side of (5) coincide with zero for any $\gamma \in (0, H(P_{X_2}) - H(P_{X_1}))$. Instead of (5), it actually holds that

$$\liminf_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \geq H(P_{X_1}) - \gamma\right\} = 1 > 0, \quad (7)$$

$$\liminf_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \geq H(P_{X_1}) + \gamma\right\} = 0 \quad (8)$$

for any constant $\gamma > 0$. Hence, $\overline{H}^*(\boldsymbol{X}) = H(P_{X_1})$ follows.

Example 2.4: We can also consider combinations of Examples 2.2 and 2.3. Letting $P_{X_1}, P_{X_2}, P_{X_3}$ and P_{X_4} be probability distributions on $\mathcal{X}$ satisfying $H(P_{X_1}) \leq H(P_{X_2}) \leq H(P_{X_3}) \leq H(P_{X_4}) < \infty$, define $P_{X_i^n}(x^n) = \prod_{k=1}^n P_{X_i}(x_k)$ for each $n \geq 1$ and $i = 1,2,3,4$. We define $P_{X^n}^{(1)}$ by

$$P_{X^n}^{(1)}(x^n) = \begin{cases} (1-\tau)P_{X_1^n}(x^n) + \tau P_{X_4^n}(x^n), & \text{if } n \text{ is odd,} \\ (1-\tau')P_{X_2^n}(x^n) + \tau' P_{X_3^n}(x^n), & \text{if } n \text{ is even,} \end{cases}$$

where $\tau, \tau' \in (0,1)$ are constants. Denote by $\boldsymbol{X}^{(1)}$ the general source specified by $P_{X^n}^{(1)}$, $n \geq 1$. In addition, define $P_{X^n}^{(2)}$ by

$$P_{X^n}^{(2)}(x^n) = \begin{cases} (1-\tau)P_{X_1^n}(x^n) + \tau P_{X_3^n}(x^n), & \text{if } n \text{ is odd,} \\ (1-\tau')P_{X_2^n}(x^n) + \tau' P_{X_4^n}(x^n), & \text{if } n \text{ is even,} \end{cases}$$

and denote by $\boldsymbol{X}^{(2)}$ the general source corresponding to $P_{X^n}^{(2)}$, $n \geq 1$. We can verify that $\underline{H}(\boldsymbol{X}^{(i)}) = H(P_{X_1})$, $\underline{H}^*(\boldsymbol{X}^{(i)}) = H(P_{X_2})$, $\overline{H}^*(\boldsymbol{X}^{(i)}) = H(P_{X_3})$ and $\overline{H}(\boldsymbol{X}^{(i)}) = H(P_{X_4})$ for $i = 1,2$ (see Fig. 1).

In general, the following relationships hold.

Proposition 2.2: For a general source $\boldsymbol{X}$ it holds that

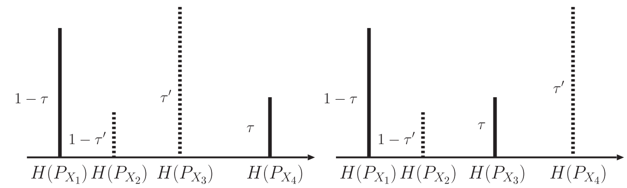

Fig. 1 Entropy spectrum of general sources $\boldsymbol{X}^{(1)}$ (left) and $\boldsymbol{X}^{(2)}$ (right).

(a) $\underline{H}(\boldsymbol{X}) \leq \overline{H}^*(\boldsymbol{X}) \leq \overline{H}(\boldsymbol{X})$,
(b) $\underline{H}(\boldsymbol{X}) \leq \underline{H}^*(\boldsymbol{X}) \leq \overline{H}(\boldsymbol{X})$.

Proof: We only prove (a). The definitions of $\overline{H}^*(\boldsymbol{X})$ and $\overline{H}(\boldsymbol{X})$ in Definition 2.1 clearly imply $\overline{H}^*(\boldsymbol{X}) \leq \overline{H}(\boldsymbol{X})$. In addition, $\underline{H}(\boldsymbol{X}) \leq \overline{H}^*(\boldsymbol{X})$ immediately follows from comparison of the definition of $\underline{H}(\boldsymbol{X})$ in Definition 2.1 with the expression of $\overline{H}^*(\boldsymbol{X})$ in Proposition 2.1. We can prove (b) in the same way. □

We have seen that the mixed source in Example 2.2 satisfies $\underline{H}^*(\boldsymbol{X}) < \overline{H}^*(\boldsymbol{X})$. On the other hand, the perturbing source in Example 2.3 satisfies $\overline{H}^*(\boldsymbol{X}) < \underline{H}^*(\boldsymbol{X})$. These examples suggest that $\underline{H}^*(\boldsymbol{X}) \leq \overline{H}^*(\boldsymbol{X})$ does not hold in general. In order to clarify a relationship between $\underline{H}^*(\boldsymbol{X})$ and $\overline{H}^*(\boldsymbol{X})$, we introduce the following class of sources.

Definition 2.2: We say a general source $\boldsymbol{X} = \{X^n\}_{n=1}^\infty$ is *canonical* if there exists a constant α_0 such that for any $\gamma > 0$ both the following two inequalities hold:

$$\liminf_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \leq \alpha_0 + \gamma\right\} > 0, \quad (9)$$

$$\liminf_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \geq \alpha_0 - \gamma\right\} > 0. \quad (10)$$

Clearly, the stationary memoryless source in Example 2.1 is canonical. That is, the law of large numbers guarantees that both (9) and (10) are satisfied by $\alpha_0 = H(P_X)$. In addition, the mixed source in Example 2.2 is also canonical. In fact, any choice of α_0 satisfying $H(P_{X_1}) \leq \alpha_0 \leq H(P_{X_2})$ meets (9) and (10). The two sources in Example 2.4 are also canonical. However, the perturbing source in Example 2.3 is not canonical because we cannot choose α_0 so that both (9) and (10) are satisfied.

The following theorem claims that the canonicality of a source $\boldsymbol{X}$ is a key to judge whether $\underline{H}^*(\boldsymbol{X}) \leq \overline{H}^*(\boldsymbol{X})$ holds or not.

Theorem 2.1: A general source $\boldsymbol{X}$ satisfies $\underline{H}^*(\boldsymbol{X}) \leq \overline{H}^*(\boldsymbol{X})$ if and only if $\boldsymbol{X}$ is canonical.

Proof: First, we prove that $\boldsymbol{X}$ is canonical if $\underline{H}^*(\boldsymbol{X}) \leq \overline{H}^*(\boldsymbol{X})$. Let $\gamma > 0$ be an arbitrary constant. In view of $\underline{H}^*(\boldsymbol{X})$ in Proposition 2.1 we have

$$\liminf_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \leq \underline{H}^*(\boldsymbol{X}) + \gamma\right\} > 0. \quad (11)$$

On the other hand, $\overline{H}^*(\boldsymbol{X})$ in Proposition 2.1 implies that

$$\liminf_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \geq \overline{H}^*(\boldsymbol{X}) - \gamma\right\} > 0,$$

which, together with $\underline{H}^*(\boldsymbol{X}) \leq \overline{H}^*(\boldsymbol{X})$, leads to

$$\liminf_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \geq \underline{H}^*(\boldsymbol{X}) - \gamma\right\} > 0. \quad (12)$$

Then, (11) and (12) guarantee that $\boldsymbol{X}$ is canonical because (9) and (10) are satisfied with $\alpha_0 = \underline{H}^*(\boldsymbol{X})$.

Next, we prove that $\underline{H}^*(\boldsymbol{X}) \leq \overline{H}^*(\boldsymbol{X})$ if $\boldsymbol{X}$ is canonical. Since $\boldsymbol{X}$ is canonical, there exists a constant α_0 such that both (9) and (10) hold for an arbitrarily fixed $\gamma > 0$. Then, it follows from (9) and Proposition 2.1 that $\underline{H}^*(\boldsymbol{X}) \leq \alpha_0 + \gamma$. Similarly, owing to (10) and Proposition 2.1 we have $\overline{H}^*(\boldsymbol{X}) \geq \alpha_0 - \gamma$. Thus, it holds that $\underline{H}^*(\boldsymbol{X}) - \gamma \leq \alpha_0 \leq \overline{H}^*(\boldsymbol{X}) + \gamma$, which implies $\underline{H}^*(\boldsymbol{X}) \leq \overline{H}^*(\boldsymbol{X}) + 2\gamma$. Since $\gamma > 0$ is arbitrary, $\underline{H}^*(\boldsymbol{X}) \leq \overline{H}^*(\boldsymbol{X})$ follows. □

3. Operational Meanings

In this section we consider the operational meanings of $\overline{H}(\boldsymbol{X})$, $\overline{H}^*(\boldsymbol{X})$, $\underline{H}(\boldsymbol{X})$ and $\underline{H}^*(\boldsymbol{X})$. It is well-known that $\overline{H}(\boldsymbol{X})$ means the infimum achievable coding rate of fixed-length codes with the vanishing decoding error probability [4], while $\underline{H}(\boldsymbol{X})$ corresponds to the supremum achievable rate of the intrinsic randomness problem [16]. We will see that all the four quantities can be explained in the context of the ε-source coding problem [2].

We describe the ordinary ε-source coding problem as follows. Given a general source $\boldsymbol{X} = \{X^n\}_{n=1}^{\infty}$, let an encoder $\varphi_n : \mathcal{X}^n \to \mathcal{M}_n \stackrel{\text{def}}{=} \{1, 2, \ldots, M_n\}$ and a decoder $\psi_n : \mathcal{M}_n \to \mathcal{X}^n$ be deterministic mappings. Define the decoding error probability by

$$\varepsilon_n \stackrel{\text{def}}{=} \Pr\{\psi_n(\varphi_n(X^n)) \neq X^n\}.$$

We are interested in the infimum of the coding rates at which ε_n is asymptotically bounded by an arbitrarily given constant $\varepsilon \in [0, 1)$.

Definition 3.1: Let $\varepsilon \in [0, 1)$ be an arbitrary constant. A rate R is called *ε-achievable* if there exists a sequence $\{(\varphi_n, \psi_n)\}_{n=1}^{\infty}$ of pairs of an encoder and a decoder satisfying

$$\limsup_{n\to\infty} \frac{1}{n}\log M_n \leq R \quad (13)$$

and

$$\limsup_{n\to\infty} \varepsilon_n \leq \varepsilon. \quad (14)$$

Denote by $R_\varepsilon(\boldsymbol{X})$ the infimum of the ε-achievable rates.

For simplifying notations we define

$$\overline{J}_\varepsilon(\boldsymbol{X}) = \inf\{R : \overline{F}(R) \leq \varepsilon\}, \quad (15)$$
$$\overline{J}_\varepsilon^*(\boldsymbol{X}) = \inf\{R : \overline{F}^*(R) \leq \varepsilon\}, \quad (16)$$

where

$$\overline{F}(R) = \limsup_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \geq R\right\},$$
$$\overline{F}^*(R) = \liminf_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \geq R\right\}.$$

Clearly, it holds that $\overline{J}_0(\boldsymbol{X}) = \overline{H}(\boldsymbol{X})$ and $\overline{J}_0^*(\boldsymbol{X}) = \overline{H}^*(\boldsymbol{X})$.

The following theorem gives a formula of $R_\varepsilon(\boldsymbol{X})$.

Theorem 3.1 ([2], [3], [13]): $R_\varepsilon(\boldsymbol{X}) = \overline{J}_\varepsilon(\boldsymbol{X})$, $\varepsilon \in [0, 1)$.

Theorem 3.1 indicates the following operational meanings of $\overline{H}(\boldsymbol{X})$ and $\underline{H}^*(\boldsymbol{X})$. Notice that the equality on $\overline{H}(\boldsymbol{X})$ is mentioned in [2], [3].

Corollary 3.1: It holds that

$$R_0(\boldsymbol{X}) = \overline{H}(\boldsymbol{X}) \quad \text{and} \quad \lim_{\varepsilon\uparrow 1} R_\varepsilon(\boldsymbol{X}) = \underline{H}^*(\boldsymbol{X}). \quad (17)$$

Proof: Since the first equality in (17) is obvious from the operational meaning of $\overline{H}(\boldsymbol{X})$ in fixed-length coding [4], we prove the second equality in (17). In view of Theorem 3.1, we have only to prove $\underline{H}^*(\boldsymbol{X}) = \lim_{\varepsilon\uparrow 1} \overline{J}_\varepsilon(\boldsymbol{X})$. We note that $\overline{F}(R)$ is a monotone decreasing function of R and $\overline{J}_\varepsilon(\boldsymbol{X})$ is a monotone decreasing function of ε. In particular, the expression of $\underline{H}^*(\boldsymbol{X})$ in Definition 2.1 guarantees that $\overline{F}(R) = 1$ if $R < \underline{H}^*(\boldsymbol{X})$ and $\overline{F}(R) < 1$ if $R > \underline{H}^*(\boldsymbol{X})$. Hence, letting $\gamma > 0$ be an arbitrarily small constant, for any $\varepsilon \in (0, 1)$ it holds that $\overline{J}_\varepsilon(\boldsymbol{X}) \geq \underline{H}^*(\boldsymbol{X}) - \gamma$ because $\overline{F}(\underline{H}^*(\boldsymbol{X}) - \gamma) = 1 > \varepsilon$. By letting $\varepsilon \uparrow 1$, we have

$$\lim_{\varepsilon\uparrow 1} \overline{J}_\varepsilon(\boldsymbol{X}) \geq \underline{H}^*(\boldsymbol{X}) - \gamma. \quad (18)$$

In addition, since $\overline{F}(\underline{H}^*(\boldsymbol{X}) + \gamma) < 1$, there exists a positive number δ_0 such that $\overline{F}(\underline{H}^*(\boldsymbol{X}) + \gamma) \leq 1 - \delta_0$. Therefore, owing to the definition of $\overline{J}_\varepsilon(\boldsymbol{X})$ it holds that

$$\lim_{\varepsilon\uparrow 1} \overline{J}_\varepsilon(\boldsymbol{X}) \leq \overline{J}_{1-\delta_0}(\boldsymbol{X}) \leq \underline{H}^*(\boldsymbol{X}) + \gamma, \quad (19)$$

where the first inequality follows because $\overline{J}_\varepsilon(\boldsymbol{X})$ is a monotone decreasing function of ε and the second inequality follows from the definition of $\overline{J}_{1-\delta_0}(\boldsymbol{X})$. Since $\gamma > 0$ is arbitrary in (18) and (19), $\underline{H}^*(\boldsymbol{X}) = \lim_{\varepsilon\uparrow 1} \overline{J}_\varepsilon(\boldsymbol{X})$ follows. □

Next, let us consider the operational meanings of $\underline{H}(\boldsymbol{X})$ and $\overline{H}^*(\boldsymbol{X})$. To this end, we take a nonconventional approach and define the supremum of unachievable rates in the strong sense.

Definition 3.2: Let $\varepsilon \in [0, 1)$ be an arbitrary constant. A rate R is called *ε-unachievable in the strong sense* if

$$\liminf_{n\to\infty} \varepsilon_n > \varepsilon \quad (20)$$

holds for any sequence $\{(\varphi_n, \psi_n)\}_{n=1}^{\infty}$ of pairs of an encoder and a decoder satisfying (13). Denote by $U_\varepsilon(\boldsymbol{X})$ the supremum of the ε-unachievable rates in the strong sense.

It is obvious from the definition of $U_\varepsilon(\boldsymbol{X})$ that for any $R > U_\varepsilon(\boldsymbol{X})$ there exists a sequence $\{(\varphi_n, \psi_n)\}_{n=1}^{\infty}$ of pairs of an encoder φ_n and a decoder ψ_n satisfying (13) and

$$\liminf_{n\to\infty} \varepsilon_n \le \varepsilon. \tag{21}$$

Hence, $U_\varepsilon(\boldsymbol{X})$ equals to the infimum achievable coding rate the codes satisfying (13) and (21). Thus, the formula for $U_\varepsilon(\boldsymbol{X})$ coincides with the formula of $R^\dagger(\varepsilon|\bar{p})$ in [5, Theorem 1]. However, since $U_\varepsilon(\boldsymbol{X})$ is closely related to the infimum ε-achievable coding rate in the optimistic sense, which originates from [17] and is discussed in [2], [12], [14], we take the following approach.

Definition 3.3 ([2], [12], [17]): Let $\varepsilon \in [0, 1)$ be an arbitrary constant. A rate R is called ε-achievable in the optimistic sense if there exists a sequence $\{(\varphi_n, \psi_n)\}_{n=1}^\infty$ of pairs of an encoder and a decoder such that for any $\gamma > 0$ there exists a subsequence $\{n_i\}_{i=1}^\infty$ satisfying

$$\frac{1}{n_i} \log M_{n_i} \le R + \gamma \quad \text{and} \quad \varepsilon_{n_i} \le \varepsilon + \gamma \quad \text{for all } i \ge 1. \tag{22}$$

Denote by $R^*_\varepsilon(\boldsymbol{X})$ the infimum of the ε-achievable rates in the optimistic sense.

The following theorem is obtained from simple observations on $U_\varepsilon(\boldsymbol{X})$ and $R^*_\varepsilon(\boldsymbol{X})$.

Theorem 3.2: $R^*_\varepsilon(\boldsymbol{X}) = U_\varepsilon(\boldsymbol{X})$ for all $\varepsilon \in [0, 1)$.

Proof: If $R < U_\varepsilon(\boldsymbol{X})$, then any $\{(\varphi_n, \psi_n)\}_{n=1}^\infty$ satisfies (20) and therefore cannot satisfy the second inequality in (22) for any $\gamma > 0$. This implies that $R^*_\varepsilon(\boldsymbol{X}) \ge U_\varepsilon(\boldsymbol{X})$. On the other hand, if $R > U_\varepsilon(\boldsymbol{X})$, there exists a $\{(\varphi_n, \psi_n)\}_{n=1}^\infty$ satisfying (13) and (21). Such a $\{(\varphi_n, \psi_n)\}_{n=1}^\infty$ obviously satisfies the two inequalities in (22) for some subsequence $\{n_i\}_{i=1}^\infty$. This establishes $R^*_\varepsilon(\boldsymbol{X}) \le U_\varepsilon(\boldsymbol{X})$. □

Theorem 3.2 leads to the following formula of $U_\varepsilon(\boldsymbol{X})$.

Theorem 3.3 ([2], [5], [12]): $U_\varepsilon(\boldsymbol{X}) = \overline{J}^*_\varepsilon(\boldsymbol{X})$, $\varepsilon \in [0, 1)$.

Similarly to Corollary 3.1, we can obtain the following operational meaning of $\overline{H}^*(\boldsymbol{X})$ and $\underline{H}(\boldsymbol{X})$, where the first equality in (23) is mentioned in [2].

Corollary 3.2: It holds that

$$U_0(\boldsymbol{X}) = \overline{H}^*(\boldsymbol{X}) \quad \text{and} \quad \lim_{\varepsilon\uparrow 1} U_\varepsilon(\boldsymbol{X}) = \underline{H}(\boldsymbol{X}). \tag{23}$$

The proof of Corollary 3.2 is omitted because the proof is essentially the same as the proof of Corollary 3.1.

4. Variations of the Strong Converse Theorem

In this section we discuss variations of the strong converse theorem using the four quantities in Definition 2.1. Wolfowitz defined that a source satisfies the strong converse property if for any coding rate $R < R_0(\boldsymbol{X})$ any sequence $\{(\varphi_n, \psi_n)\}_{n=1}^\infty$ of pairs of an encoder and a decoder meeting the rate constraint (13) satisfies $\varepsilon_n \to 1$ as $n \to \infty$ [18]. Han shows that a general source $\boldsymbol{X}$ satisfies the strong converse property if and only if $\overline{H}(\boldsymbol{X}) = \underline{H}(\boldsymbol{X})$ [3]. Note that, in view of Proposition 2.2, $\overline{H}(\boldsymbol{X}) = \underline{H}(\boldsymbol{X})$ actually means that $\overline{H}(\boldsymbol{X}) = \overline{H}^*(\boldsymbol{X}) = \underline{H}^*(\boldsymbol{X}) = \underline{H}(\boldsymbol{X})$. This motivates us to investigate the source coding problems in which different kinds of equalities, say, two or three out of the above four quantities being equal, are obtained as necessary and sufficient conditions.

We first define the ε-strong converse property.

Definition 4.1: Let $\varepsilon \in [0, 1)$ be an arbitrary constant. A general source $\boldsymbol{X}$ is said to satisfy the *ε-strong converse property* if for any coding rate $R < R_\varepsilon(\boldsymbol{X})$ any $\{(\varphi_n, \psi_n)\}_{n=1}^\infty$ meeting (13) satisfies (20). In particular, we say that a source satisfies the *zero-strong converse property* if the source satisfies the ε-strong converse property with $\varepsilon = 0$.

Theorem 4.1: A general source $\boldsymbol{X}$ satisfies the ε-strong converse property if and only if $R_\varepsilon(\boldsymbol{X}) = U_\varepsilon(\boldsymbol{X})$. In particular, a source satisfies the zero-strong converse property if and only if $\overline{H}(\boldsymbol{X}) = \overline{H}^*(\boldsymbol{X})$.

The proof of Theorem 4.1 is immediate from Theorems 3.1 and 3.3. Note that $U_\varepsilon(\boldsymbol{X}) \le R_\varepsilon(\boldsymbol{X})$ from their definitions and recall that any $\{(\varphi_n, \psi_n)\}_{n=1}^\infty$ satisfying (13) meets (20) if $R < U_\varepsilon(\boldsymbol{X})$. Hence, if $U_\varepsilon(\boldsymbol{X}) = R_\varepsilon(\boldsymbol{X})$, $\boldsymbol{X}$ satisfies the ε-strong converse property. Conversely, suppose that $\boldsymbol{X}$ satisfies the ε-strong converse property. If $U_\varepsilon(\boldsymbol{X}) < R_\varepsilon(\boldsymbol{X})$, then for any R satisfying $U_\varepsilon(\boldsymbol{X}) < R < R_\varepsilon(\boldsymbol{X})$ there exists a $\{(\varphi_n, \psi_n)\}_{n=1}^\infty$ satisfying (13) and (21). This conflicts the assumption that $\boldsymbol{X}$ satisfies the ε-strong converse property.

This simple argument on the ε-strong converse property explains one of the reasons why we adopt a nonconventional definition in Definition 3.2. The author discussed coding of a general source with the zero strong converse property in [7].

We can define another strong converse property.

Definition 4.2: A general source $\boldsymbol{X}$ is said to satisfy the *optimistic strong converse property* if for any coding rate $R < U_0(\boldsymbol{X}) = \overline{H}^*(\boldsymbol{X})$ any $\{(\varphi_n, \psi_n)\}_{n=1}^\infty$ meeting (13) satisfies

$$\liminf_{n\to\infty} \varepsilon_n = 1. \tag{24}$$

Theorem 4.2: A general source $\boldsymbol{X}$ satisfies the optimistic strong converse property if and only if $\overline{H}^*(\boldsymbol{X}) = \underline{H}(\boldsymbol{X})$.

Theorem 4.2 is obvious from Corollary 3.2. If $\boldsymbol{X}$ is canonical, Theorem 2.1 tells us that the optimistic strong converse property implies $\overline{H}^*(\boldsymbol{X}) = \underline{H}^*(\boldsymbol{X}) = \underline{H}(\boldsymbol{X})$. This means that for any $R > \overline{H}^*(\boldsymbol{X})$ there is no $\{(\varphi_n, \psi_n)\}_{n=1}^\infty$ satisfying (13) and $\varepsilon_{n_i} \to 1$ as $i \to \infty$ for a subsequence $\{n_i\}_{i=1}^\infty$.

The definition of the optimistic strong converse property (24) can be weakened to $\limsup_{n\to\infty} \varepsilon_n = 1$. Obviously, $\overline{H}^*(\boldsymbol{X}) = \underline{H}^*(\boldsymbol{X})$ is the necessary and sufficient condition for the weakened optimistic strong converse property.

Example 4.1: Consider the source $\boldsymbol{X}^{(1)}$ in Example 2.4. This source satisfies the zero-strong converse property if and only if $H(P_{X_3}) = H(P_{X_4})$. Note that we asymptotically have a mass at $R = H(P_{X_3})$ in this case. This source satisfies the optimistic strong converse property if and only if $H(P_{X_1}) = H(P_{X_3})$. Obviously, if $R > H(P_{X_3})$, we can construct $\{(\varphi_n, \psi_n)\}_{n=1}^\infty$ with arbitrarily small decoding error probability for even n. Similarly, the source satisfies the weakened optimistic strong converse property if and only if $H(P_{X_2}) = H(P_{X_3})$.

5. Relationship to the Smooth Rényi Entropy of Order Zero

Renner and Wolf defined the smooth Rényi entropy and pointed out a relationship between the smooth Rényi entropy and the ε-source coding problem [10], [11]. In particular, the smooth Rényi entropy of order zero for a random variable $X^n \in \mathcal{X}^n$ can be written as

$$K_\delta(X^n) = \min_{\substack{\mathcal{A}_n \subset \mathcal{X}^n : \\ \Pr\{X^n \in \mathcal{A}_n\} \geq 1-\delta}} \log |\mathcal{A}_n|,$$

where $\delta \in (0, 1)$ is an arbitrary constant. Quite recently, Uyematsu [14] discussed coding of a general source in terms of the smooth Rényi entropy of order zero to and showed that

$$\lim_{\delta \downarrow 0} \limsup_{n\to\infty} \frac{1}{n} K_\delta(X^n) = \overline{H}(\boldsymbol{X}), \tag{25}$$

$$\lim_{\delta \uparrow 1} \liminf_{n\to\infty} \frac{1}{n} K_\delta(X^n) = \underline{H}(\boldsymbol{X}). \tag{26}$$

The following theorem shows that not only $\overline{H}(\boldsymbol{X})$ and $\underline{H}(\boldsymbol{X})$ but also $\overline{H}^*(\boldsymbol{X})$ and $\underline{H}^*(\boldsymbol{X})$ are characterized by using the smooth Rényi entropy of order zero.

Theorem 5.1: For a general source $\boldsymbol{X} = \{X^n\}_{n=1}^\infty$ we have:

$$\lim_{\delta \downarrow 0} \liminf_{n\to\infty} \frac{1}{n} K_\delta(X^n) = \overline{H}^*(\boldsymbol{X}), \tag{27}$$

$$\lim_{\delta \uparrow 1} \limsup_{n\to\infty} \frac{1}{n} K_\delta(X^n) = \underline{H}^*(\boldsymbol{X}). \tag{28}$$

For simplifying the notations we define

$$\underline{K}_\delta(\boldsymbol{X}) = \liminf_{n\to\infty} \frac{1}{n} K_\delta(X^n), \quad \overline{K}_\delta(\boldsymbol{X}) = \limsup_{n\to\infty} \frac{1}{n} K_\delta(X^n)$$

for $\delta \in (0, 1)$ and

$$\underline{K}_0(\boldsymbol{X}) = \lim_{\delta \downarrow 0} \underline{K}_\delta(\boldsymbol{X}), \quad \overline{K}_1(\boldsymbol{X}) = \lim_{\delta \uparrow 1} \overline{K}_\delta(\boldsymbol{X}).$$

We need the following lemma in the proof of Theorem 5.1.

Lemma 5.1: For any $\gamma > 0$ and $\delta \in (0, 1)$ it holds that

$$\Pr\{X^n \in \mathcal{V}_n\} \geq 1 - \delta - 2^{-n\gamma} \quad \text{for all } n \geq 1, \tag{29}$$

where

$$\mathcal{V}_n = \left\{ x^n \in \mathcal{X}^n \ : \ \frac{1}{n} \log \frac{1}{P_{X^n}(x^n)} < \frac{1}{n} K_\delta(X^n) + \gamma \right\}. \tag{30}$$

Proof: Fix $n \geq 1$ and $\delta \in (0, 1)$ arbitrarily. From the definition of $K_\delta(X^n)$, for each $n \geq 1$ there exists a subset $\mathcal{A}_n \subset \mathcal{X}^n$ satisfying

$$K_\delta(X^n) = \log |\mathcal{A}_n| \ \text{ and } \ \Pr\{X^n \in \mathcal{A}_n\} \geq 1 - \delta. \tag{31}$$

Then, it follows that

$$\begin{aligned} \Pr\{X^n \in \mathcal{A}_n\} &= \Pr\{X^n \in \mathcal{A}_n \cap \mathcal{V}_n\} + \Pr\{X^n \in \mathcal{A}_n \cap \mathcal{V}_n^c\} \\ &\leq \Pr\{X^n \in \mathcal{V}_n\} + \Pr\{X^n \in \mathcal{A}_n \cap \mathcal{V}_n^c\}, \end{aligned} \tag{32}$$

where $\mathcal{V}_n^c$ denotes the complement of $\mathcal{V}_n$. By using $P_{X^n}(x^n) \leq \frac{1}{|\mathcal{A}_n| 2^{n\gamma}}$ for all $x^n \in \mathcal{V}_n^c$, we have

$$\begin{aligned} \Pr\{X^n \in \mathcal{A}_n \cap \mathcal{V}_n^c\} &= \sum_{x^n \in \mathcal{A}_n \cap \mathcal{V}_n^c} P_{X^n}(x^n) \\ &\leq \frac{|\mathcal{A} \cap \mathcal{V}_n^c|}{|\mathcal{A}_n| 2^{n\gamma}} \\ &\leq 2^{-n\gamma}. \end{aligned} \tag{33}$$

The claim of the lemma follows from the combination of (31), (32) and (33). □

Proof of Theorem 5.1: We first prove (a-1) $\overline{H}^*(\boldsymbol{X}) \leq \underline{K}_0(\boldsymbol{X})$ and (a-2) $\underline{H}^*(\boldsymbol{X}) \leq \overline{K}_1(\boldsymbol{X})$. Let us consider (a-1). In view of the expression of $\overline{H}^*(\boldsymbol{X})$ in Definition 2.1 it suffices to prove

$$\limsup_{n\to\infty} \Pr\left\{ \frac{1}{n} \log \frac{1}{P_{X^n}(X^n)} \leq \underline{K}_0(\boldsymbol{X}) + 2\gamma \right\} = 1 \tag{34}$$

for any $\gamma > 0$. Fix $\gamma > 0$ and $\delta \in (0, 1)$ arbitrarily. Define $\mathcal{U}_n$ by

$$\mathcal{U}_n = \left\{ x^n \in \mathcal{X}^n \ : \ \frac{1}{n} \log \frac{1}{P_{X^n}(x^n)} \leq \underline{K}_\delta(\boldsymbol{X}) + 2\gamma \right\}. \tag{35}$$

Since

$$\frac{1}{n} K_\delta(X^n) \leq \underline{K}_\delta(\boldsymbol{X}) + \gamma \quad \text{infinitely often}$$

from the definition of $\underline{K}_\delta(\boldsymbol{X})$, it holds that

$$\Pr\{X^n \in \mathcal{V}_n\} \leq \Pr\{X^n \in \mathcal{U}_n\} \quad \text{infinitely often}. \tag{36}$$

In addition, since $\underline{K}_\delta(\boldsymbol{X})$ is monotone decreasing with respect to δ, we have

$$\Pr\left\{ \frac{1}{n} \log \frac{1}{P_{X^n}(X^n)} \leq \underline{K}_0(\boldsymbol{X}) + 2\gamma \right\} \geq \Pr\{X^n \in \mathcal{U}_n\} \quad \text{for all } n \geq 1. \tag{37}$$

Then, the combination of (36), (37) and Lemma 5.1 yields

$$\Pr\left\{ \frac{1}{n} \log \frac{1}{P_{X^n}(X^n)} \leq \underline{K}_0(\boldsymbol{X}) + 2\gamma \right\} \geq 1 - \delta - 2^{-n\gamma} \quad \text{infinitely often}. \tag{38}$$

This guarantees

$$\limsup_{n\to\infty} \Pr\left\{ \frac{1}{n} \log \frac{1}{P_{X^n}(X^n)} \leq \underline{K}_0(\boldsymbol{X}) + 2\gamma \right\} \geq 1 - \delta. \tag{39}$$

Since $\delta \in (0, 1)$ in (39) is arbitrary, we have (34).

In order to establish (a-2), we will establish

$$\liminf_{n\to\infty} \Pr\left\{ \frac{1}{n} \log \frac{1}{P_{X^n}(X^n)} \leq \overline{K}_\delta(\boldsymbol{X}) + 2\gamma \right\} > 0 \tag{40}$$

for any $\delta \in (0, 1)$ and $\gamma > 0$. This, together with the expression of $\underline{H}^*(\boldsymbol{X})$ in Proposition 2.1, implies that $\underline{H}^*(\boldsymbol{X}) \leq \overline{K}_\delta(\boldsymbol{X})$ for any $\delta \in (0, 1)$. Inequality (a-2) can be obtained by letting $\delta \uparrow 1$. Hereinafter, we establish (40). Define

$$\mathcal{U}'_n = \left\{ x^n \in \mathcal{X}^n \ : \ \frac{1}{n} \log \frac{1}{P_{X^n}(x^n)} \leq \overline{K}_\delta(\boldsymbol{X}) + 2\gamma \right\}. \quad (41)$$

From the definition of $\overline{K}_\delta(\boldsymbol{X})$, for any $\delta \in (0, 1)$ and $\gamma > 0$ there exists an integer $n_0 = n_0(\gamma, \delta)$ satisfying

$$\frac{1}{n} K_\delta(X^n) \leq \overline{K}_\delta(\boldsymbol{X}) + \gamma \quad \text{for all } n \geq n_0.$$

Since $\mathcal{V}_n \subset \mathcal{U}'_n$ for all $n \geq n_0$, it holds that

$$\begin{aligned} \Pr\{X^n \in \mathcal{U}'_n\} &\geq \Pr\{X^n \in \mathcal{V}_n\} \\ &\geq 1 - \delta - 2^{-n\gamma} \quad \text{for all } n \geq n_0, \end{aligned} \quad (42)$$

where the second inequality follows from Lemma 5.1. By taking the limit inferior of (42), we have

$$\liminf_{n\to\infty} \Pr\left\{ \frac{1}{n} \log \frac{1}{P_{X^n}(X^n)} \leq \overline{K}_\delta(\boldsymbol{X}) + 2\gamma \right\} \geq 1 - \delta,$$

which immediately yields (40) because $\delta \in (0, 1)$.

Next, we prove (b-1) $\underline{K}_0(\boldsymbol{X}) \leq \overline{H}^*(\boldsymbol{X})$ and (b-2) $\overline{K}_1(\boldsymbol{X}) \leq \underline{H}^*(\boldsymbol{X})$. Inequalities (b-1) and (b-2) are proved in almost the same way. We begin with the proof of (b-1). Define

$$\mathcal{S}_n = \left\{ x^n \in \mathcal{X}^n \ : \ \frac{1}{n} \log \frac{1}{P_{X^n}(X^n)} \leq \overline{H}^*(\boldsymbol{X}) + \gamma \right\}. \quad (43)$$

Since $P_{X^n}(x^n) \geq 2^{-n(\overline{H}^*(\boldsymbol{X})+\gamma)}$ for all $x^n \in \mathcal{S}_n$, we can easily verify that

$$|\mathcal{S}_n| \leq 2^{n(\overline{H}^*(\boldsymbol{X})+\gamma)} \quad \text{for all } n \geq 1. \quad (44)$$

In addition, due to the expression of $\overline{H}^*(\boldsymbol{X})$ in Definition 2.1, for any $\delta \in (0, 1)$ it holds that

$$\Pr\{X^n \in \mathcal{S}_n\} \geq 1 - \delta \quad \text{infinitely often.} \quad (45)$$

Then, in view of (44), (45) and the definition of $K_\delta(X^n)$, we have

$$\frac{1}{n} K_\delta(X^n) \leq \overline{H}^*(\boldsymbol{X}) + \gamma \quad \text{infinitely often,}$$

which yields

$$\underline{K}_\delta(\boldsymbol{X}) = \liminf_{n\to\infty} \frac{1}{n} K_\delta(X^n) \leq \overline{H}^*(\boldsymbol{X}) + \gamma.$$

We can establish $\underline{K}_0(\boldsymbol{X}) \leq \overline{H}^*(\boldsymbol{X}) + \gamma$ by letting $\delta \downarrow 0$. Since $\gamma > 0$ is arbitrary, $\underline{K}_0(\boldsymbol{X}) \leq \overline{H}^*(\boldsymbol{X})$ is established.

Finally, we prove (b-2). In view of the expression of $\underline{H}^*(\boldsymbol{X})$ in Proposition 2.1, for any $\gamma > 0$ there exists a constant $\varepsilon_0 > 0$ such that

$$\Pr\left\{ \frac{1}{n} \log \frac{1}{P_{X^n}(X^n)} \leq \underline{H}^*(\boldsymbol{X}) + \gamma \right\} \geq \varepsilon_0 \quad \text{for all } n \geq n_0. \quad (46)$$

Define $\mathcal{S}_n$ by (43), where $\overline{H}^*(\boldsymbol{X})$ is replaced with $\underline{H}^*(\boldsymbol{X})$. Then, we have $|\mathcal{S}_n| \leq 2^{n(\underline{H}^*(\boldsymbol{X})+\gamma)}$ for all $n \geq 1$ and

$$\frac{1}{n} K_{1-\varepsilon_0}(X^n) \leq \underline{H}^*(\boldsymbol{X}) + \gamma \quad \text{for all } n \geq n_0$$

from the definition of $K_{1-\varepsilon_0}(X^n)$. By taking the limit superior of both sides, it holds that

$$\overline{K}_{1-\varepsilon_0}(\boldsymbol{X}) = \limsup_{n\to\infty} \frac{1}{n} K_{1-\varepsilon_0}(X^n) \leq \underline{H}^*(\boldsymbol{X}) + \gamma. \quad (47)$$

Since $\overline{K}_\delta(\boldsymbol{X})$ is monotone decreasing with respect to δ, we have $\overline{K}_1(\boldsymbol{X}) \leq \overline{K}_{1-\varepsilon_0}(\boldsymbol{X})$. Therefore, (47) leads to

$$\overline{K}_1(\boldsymbol{X}) \leq \underline{H}^*(\boldsymbol{X}) + \gamma. \quad (48)$$

which establishes (b-2) because $\gamma > 0$ is arbitrary. □

Remark: Relationship similar to (25)–(28) hold for the smooth Rényi entropy of order infinity as well. See [8], [15] for more details.

6. Bounds on the Width of the Entropy-Spectrum

This section is devoted to investigation of bounds on the width of the entropy-spectrum, where the entropy-spectrum means the distribution of $\frac{1}{n} \log \frac{1}{P_{X^n}(X^n)}$ [3]. Recall that $\overline{H}(\boldsymbol{X})$ is assumed to be finite. Then, we can define

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

where $\mathcal{G}$ denotes the set of sequences of intervals defined by

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

Here, we require that any sequence of intervals $\{(a_n, b_n)\}_{n=1}^\infty$ must satisfy $a_n \leq b_n$ for each $n \geq 1$ (we regard the interval (a_n, b_n) as the empty set if $a_n = b_n$). This $W(\boldsymbol{X})$ was defined in [6] in the context of a fixed-length homophonic coding. Quite recently, Arimura and Iwata showed that $W(\boldsymbol{X})$ coincides with the infimum achievable redundancy of fixed-to-fixed length coding [1].

We also define

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

where

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

We can show that $W^*(\boldsymbol{X})$ is the infimum achievable worst-case redundancy of fixed-to-fixed length coding in the optimistic sense [9].

The following proposition gives bounds on $W(\boldsymbol{X})$ and

$W^*(\boldsymbol{X})$. Note that Proposition 6.1-(b) is mentioned in [6] without proof and the proof given below is much simpler than the derivation in [1]. Lower bounds on $W(\boldsymbol{X})$ and $W^*(\boldsymbol{X})$ are first obtained in this paper.

Proposition 6.1: For a general source $\boldsymbol{X}$ we have:

(a) $0 \le W^*(\boldsymbol{X}) \le W(\boldsymbol{X})$.
(b) $W(\boldsymbol{X}) \le \overline{H}(\boldsymbol{X}) - \underline{H}(\boldsymbol{X})$.
(c) $W^*(\boldsymbol{X}) \le \min\{\overline{H}(\boldsymbol{X}) - \underline{H}^*(\boldsymbol{X}), \overline{H}^*(\boldsymbol{X}) - \underline{H}(\boldsymbol{X})\}$.
(d) $W^*(\boldsymbol{X}) \ge \overline{H}^*(\boldsymbol{X}) - \underline{H}^*(\boldsymbol{X})$ if $\boldsymbol{X}$ is canonical.
(e) $W(\boldsymbol{X}) \ge \max\{\overline{H}(\boldsymbol{X}) - \underline{H}^*(\boldsymbol{X}), \overline{H}^*(\boldsymbol{X}) - \underline{H}(\boldsymbol{X})\}$.

Proof: Since (a) is obvious, we will establish (b)–(e) below. Recall that for arbitrary two sequences $\{p_n\}_{n=1}^\infty$ and $\{q_n\}_{n=1}^\infty$ of real numbers it holds that

$$\liminf_{n\to\infty}(-p_n) = -\limsup_{n\to\infty} p_n, \tag{49}$$

$$\limsup_{n\to\infty}(-p_n) = -\liminf_{n\to\infty} p_n, \tag{50}$$

$$\liminf_{n\to\infty} p_n + \liminf_{n\to\infty} q_n \le \liminf_{n\to\infty}(p_n + q_n) \le \liminf_{n\to\infty} p_n + \limsup_{n\to\infty} q_n, \tag{51}$$

$$\liminf_{n\to\infty} p_n + \limsup_{n\to\infty} q_n \le \limsup_{n\to\infty}(p_n + q_n) \le \limsup_{n\to\infty} p_n + \limsup_{n\to\infty} q_n. \tag{52}$$

First we prove (b). Recall that $\underline{H}(\boldsymbol{X}) \le \overline{H}(\boldsymbol{X})$ from their definitions. We can prove $\{(\underline{H}(\boldsymbol{X}), \overline{H}(\boldsymbol{X}))\}_{n=1}^\infty \in \mathcal{G}$ in the following way. Letting $\gamma > 0$ be an arbitrary constant, for each $n \ge 1$ it holds that

$$\begin{aligned}
&\Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \in (\underline{H}(\boldsymbol{X}) - \gamma, \overline{H}(\boldsymbol{X}) + \gamma)\right\} \\
&= 1 - \left[\Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \le \underline{H}(\boldsymbol{X}) - \gamma\right\}\right. \\
&\qquad \left. + \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \ge \overline{H}(\boldsymbol{X}) + \gamma\right\}\right].
\end{aligned} \tag{53}$$

Therefore, (53) leads to

$$\begin{aligned}
&\liminf_{n\to\infty}\Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \in (\underline{H}(\boldsymbol{X}) - \gamma, \overline{H}(\boldsymbol{X}) + \gamma)\right\} \\
&\ge 1 - \limsup_{n\to\infty}\Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \le \underline{H}(\boldsymbol{X}) - \gamma\right\} \\
&\qquad - \limsup_{n\to\infty}\Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \ge \overline{H}(\boldsymbol{X}) + \gamma\right\},
\end{aligned} \tag{54}$$

where the inequality follows from (49) and (52). Note that the second and third terms on the right side of (54) are equal to 0 due to the expressions of $\underline{H}(\boldsymbol{X})$ and $\overline{H}(\boldsymbol{X})$ in Proposition 2.1. In addition, note that the left side of (54) is less than or equal to one. Hence, we can conclude that the left side of (54) coincides with one and therefore $\{(\underline{H}(\boldsymbol{X}), \overline{H}(\boldsymbol{X}))\}_{n=1}^\infty \in \mathcal{G}$. Since $W(\boldsymbol{X})$ is defined as the infimum with respect to elements of $\mathcal{G}$, $W(\boldsymbol{X}) \le \overline{H}(\boldsymbol{X}) - \underline{H}(\boldsymbol{X})$ follows.

Next, we prove (c). It suffices to prove both $W^*(\boldsymbol{X}) \le \overline{H}(\boldsymbol{X}) - \underline{H}^*(\boldsymbol{X})$ and $W^*(\boldsymbol{X}) \le \overline{H}^*(\boldsymbol{X}) - \underline{H}(\boldsymbol{X})$. We only prove $W^*(\boldsymbol{X}) \le \overline{H}(\boldsymbol{X}) - \underline{H}^*(\boldsymbol{X})$ here because $W^*(\boldsymbol{X}) \le \overline{H}^*(\boldsymbol{X}) - \underline{H}(\boldsymbol{X})$ can be proved similarly. Recall that $\underline{H}^*(\boldsymbol{X}) \le \overline{H}(\boldsymbol{X})$ from Proposition 2.2. Thus, we can consider the sequence of intervals $\{(\underline{H}^*(\boldsymbol{X}), \overline{H}(\boldsymbol{X}))\}_{n=1}^\infty$. Then, for any constant $\gamma > 0$ it follows that

$$\begin{aligned}
&\limsup_{n\to\infty}\Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \in (\underline{H}^*(\boldsymbol{X}) - \gamma, \overline{H}(\boldsymbol{X}) + \gamma)\right\} \\
&\ge 1 - \liminf_{n\to\infty}\Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \le \underline{H}^*(\boldsymbol{X}) - \gamma\right\} \\
&\qquad - \limsup_{n\to\infty}\Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \ge \overline{H}(\boldsymbol{X}) + \gamma\right\},
\end{aligned} \tag{55}$$

where the inequality follows from (50) and (51). Note that the second and third terms on the right side of (55) are equal to 0 due to the expressions of $\underline{H}^*(\boldsymbol{X})$ and $\overline{H}(\boldsymbol{X})$ in Proposition 2.1. In addition, note that the left side of (55) is less than or equal to one. Hence, we can conclude that the left side of (55) coincides with one and therefore $\{(\underline{H}^*(\boldsymbol{X}), \overline{H}(\boldsymbol{X}))\}_{n=1}^\infty \in \mathcal{G}^*$. We now have $W^*(\boldsymbol{X}) \le \overline{H}(\boldsymbol{X}) - \underline{H}^*(\boldsymbol{X})$ in view of the definition of $W^*(\boldsymbol{X})$.

Now we prove (d). Suppose that $\boldsymbol{X}$ is canonical. Then, $\underline{H}^*(\boldsymbol{X}) \le \overline{H}^*(\boldsymbol{X})$ holds from Theorem 2.1. Without loss of generality we can assume that $\underline{H}^*(\boldsymbol{X}) < \overline{H}^*(\boldsymbol{X})$ (otherwise, the bound in (d) is trivial). We prove $W^*(\boldsymbol{X}) \ge \overline{H}^*(\boldsymbol{X}) - \underline{H}^*(\boldsymbol{X})$ by contradiction. Assume that there exists an $\{(a_n, b_n)\}_{n=1}^\infty \in \mathcal{G}^*$ satisfying $\limsup_{n\to\infty}(b_n - a_n) < \overline{H}^*(\boldsymbol{X}) - \underline{H}^*(\boldsymbol{X})$. Then, there exist a constant $\gamma_0 > 0$ and an integer n_0 satisfying

$$b_n - a_n \le \overline{H}^*(\boldsymbol{X}) - \underline{H}^*(\boldsymbol{X}) - 4\gamma_0 \quad \text{for all } n \ge n_0. \tag{56}$$

We define the four intervals as follows:

$$E_n^{(1)} = \{t \in \boldsymbol{R} : t \le a_n - \gamma_0\}, \tag{57}$$

$$E_n^{(2)} = \{t \in \boldsymbol{R} : t \ge b_n + \gamma_0\}, \tag{58}$$

$$F^{(1)} = \{t \in \boldsymbol{R} : t \le \underline{H}^*(\boldsymbol{X}) + \gamma_0\}, \tag{59}$$

$$F^{(2)} = \{t \in \boldsymbol{R} : t \ge \overline{H}^*(\boldsymbol{X}) - \gamma_0\}, \tag{60}$$

where $\boldsymbol{R}$ denotes the set of real numbers. It is important to notice that at least one of $F^{(1)} \subset E_n^{(1)}$ and $F^{(2)} \subset E_n^{(2)}$ holds for all $n \ge n_0$. That is, if $a_n - \gamma_0 \ge \underline{H}^*(\boldsymbol{X}) + \gamma_0$, we have $F^{(1)} \subset E_n^{(1)}$. Otherwise, we have $F^{(2)} \subset E_n^{(2)}$ because it follows from (56) and $a_n \le \underline{H}^*(\boldsymbol{X}) + 2\gamma_0$ that $b_n \le a_n + \overline{H}^*(\boldsymbol{X}) - \underline{H}^*(\boldsymbol{X}) - 4\gamma_0 \le \overline{H}^*(\boldsymbol{X}) - 2\gamma_0$, i.e., $b_n + \gamma_0 \le \overline{H}^*(\boldsymbol{X}) - \gamma_0$ for all $n \ge n_0$. Therefore, it holds that

$$\begin{aligned}
&\Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \in E_n^{(1)} \cup E_n^{(2)}\right\} \\
&\ge \min_{i=1,2}\Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \in F^{(i)}\right\} \quad \text{for all } n \ge n_0,
\end{aligned} \tag{61}$$

which guarantees

$$\liminf_{n\to\infty}\Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \in E_n^{(1)} \cup E_n^{(2)}\right\} > 0 \tag{62}$$

owing to the expressions of $\underline{H}^*(\boldsymbol{X})$ and $\overline{H}^*(\boldsymbol{X})$ in Proposition 2.1 and the definitions of $F^{(1)}$ and $F^{(2)}$. Now we can prove

$$\limsup_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \in (a_n - \gamma_0, b_n + \gamma_0)\right\}$$
$$= 1 - \liminf_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \in E_n^{(1)} \cup E_n^{(2)}\right\} < 1,$$

where the equality follows from (50) and the inequality is guaranteed by (62). This means that $\{(a_n, b_n)\}_{n=1}^{\infty} \notin \mathcal{G}^*$, which is a contradiction.

For the proof of (e) it suffices to establish $W(\boldsymbol{X}) \geq \overline{H}(\boldsymbol{X}) - \underline{H}^*(\boldsymbol{X})$ and $W(\boldsymbol{X}) \geq \overline{H}^*(\boldsymbol{X}) - \underline{H}(\boldsymbol{X})$. We show only $W(\boldsymbol{X}) \geq \overline{H}(\boldsymbol{X}) - \underline{H}^*(\boldsymbol{X})$ below because $W(\boldsymbol{X}) \geq \overline{H}^*(\boldsymbol{X}) - \underline{H}(\boldsymbol{X})$ can be proved in the same way. Similarly to the proof of (d), we prove $W(\boldsymbol{X}) \geq \overline{H}(\boldsymbol{X}) - \underline{H}^*(\boldsymbol{X})$ by contradiction. Suppose that $\overline{H}(\boldsymbol{X}) > \underline{H}^*(\boldsymbol{X})$ and assume that there exist a constant $\gamma_0 > 0$ and an integer n_0 satisfying

$$b_n - a_n \leq \overline{H}(\boldsymbol{X}) - \underline{H}^*(\boldsymbol{X}) - 4\gamma_0 \quad \text{for all } n \geq n_0. \quad (63)$$

We define $E_n^{(1)}$, $E_n^{(2)}$ and $F^{(1)}$ by (57), (58) and (59), respectively and replace the definition of $F^{(2)}$ in (60) with

$$F^{(2)} = \{t \in \boldsymbol{R} : t \geq \overline{H}(\boldsymbol{X}) - \gamma_0\}.$$

Then, we can show that one of $F^{(1)} \subset E_n^{(1)}$ and $F^{(2)} \subset E_n^{(2)}$ holds for all $n \geq n_0$. This fact leads to (61), which guarantees

$$\limsup_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \in E_n^{(1)} \cup E_n^{(2)}\right\} > 0 \quad (64)$$

owing to the expressions of $\underline{H}^*(\boldsymbol{X})$ and $\overline{H}(\boldsymbol{X})$ in Proposition 2.1. Then, we can prove

$$\liminf_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \in (a_n - \gamma_0, b_n + \gamma_0)\right\}$$
$$= 1 - \limsup_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \in E_n^{(1)} \cup E_n^{(2)}\right\} < 1,$$

where the equality follows from (49) and the inequality is guaranteed by (64). This means that $\{(a_n, b_n)\}_{n=1}^{\infty} \notin \mathcal{G}$ and contradicts the assumption of $\{(a_n, b_n)\}_{n=1}^{\infty} \in \mathcal{G}$. □

Example 6.1: Let us consider the general source $\boldsymbol{X}^{(1)}$ in Example 2.4. In view of the definitions of $W(\boldsymbol{X})$ and $W^*(\boldsymbol{X})$, we have $W(\boldsymbol{X}^{(1)}) = H(P_{X_4}) - H(P_{X_1}) = \overline{H}(\boldsymbol{X}^{(1)}) - \underline{H}(\boldsymbol{X}^{(1)})$ and $W^*(\boldsymbol{X}^{(1)}) = H(P_{X_3}) - H(P_{X_2}) = \overline{H}^*(\boldsymbol{X}^{(1)}) - \underline{H}^*(\boldsymbol{X}^{(1)})$. That is, the bounds in (b) and (d) are satisfied with equality. On the other hand, for the general source $\boldsymbol{X}^{(2)}$ in Example 2.4, if $H(P_{X_3}) - H(P_{X_1}) \leq H(P_{X_4}) - H(P_{X_2})$ is assumed, we have $W(\boldsymbol{X}^{(2)}) = H(P_{X_4}) - H(P_{X_2}) = \overline{H}(\boldsymbol{X}^{(2)}) - \underline{H}^*(\boldsymbol{X}^{(2)})$ and $W^*(\boldsymbol{X}^{(2)}) = H(P_{X_3}) - H(P_{X_1}) = \overline{H}^*(\boldsymbol{X}^{(2)}) - \underline{H}(\boldsymbol{X}^{(2)})$ for this source, i.e., the bounds in (c) and (e) are satisfied with equality.

7. Conclusion

In this paper we have introduced two nonconventional quantities $\overline{H}^*(\boldsymbol{X})$ and $\underline{H}^*(\boldsymbol{X})$ and have investigated their roles in coding of a general source. We have discussed the operational meanings, variations of the strong converse theorem, relationships to the smooth Rényi entropy of order zero, and bounds on the width of the entropy-spectrum. These results show importance of $\overline{H}^*(\boldsymbol{X})$ and $\underline{H}^*(\boldsymbol{X})$ as new terminologies in information-spectrum methods.

Acknowledgement

The author would like to express his thanks to Prof. Mitsuharu Arimura in Shonan Institute of Technology and Prof. Ken-ichi Iwata in University of Fukui for their comments on the primary version of Sect. 6. Constructive comments from anonymous referees helped to improve readability of this paper. This research is supported in part by MEXT grant-in-aid for scientific research (C) 20500004.

References

[1] M. Arimura and K. Iwata, "The minimum achievable redundancy rate of fixed-to-fixed length source codes for general sources," Proc. ISITA 2010, pp.595–600, Taiwan, 2010.
[2] P.-N. Chen and F. Alajaji, "Optimistic Shannon coding theorems for arbitrary single-user systems," IEEE Trans. Inf. Theory, vol.45, no.7, pp.2623–2629, 1999.
[3] T.S. Han, Information-Spectrum Methods in Information Theory, Springer, 2003.
[4] T.S. Han and S. Verdú, "Approximation theory of output statistics," IEEE Trans. Inf. Theory, vol.39, no.3, pp.752–772, 1993.
[5] M. Hayashi, "Second order asymptotics in fixed-length source coding and intrinsic randomness," IEEE Trans. Inf Theory, vol.54, no.10, pp.4619–4637, 2008.
[6] H. Koga, "A coding theorem on the fixed-length homophonic coding for a general source," Proc. SITA 2000, Aso, pp.133–136, 2000. (in Japanese)
[7] H. Koga, "Two generalizations of a coding theorem for a (2, 2)-threshold scheme with a cheater," Proc. ISITA 2010, pp.678–683, Taiwan, 2010.
[8] H. Koga, "Four limits in probability and coding of a general source," IEICE Technical Report, IT2010-60, Jan. 2011.
[9] H. Koga, M. Arimura, and K. Iwata, "The width of information-spectrum and the worst-case redundancy of fixed-length coding," IEICE Technical Report, IT2010-87, March 2011.
[10] R. Renner and S. Wolf, "Smooth Rényi entropy and applications," Proc. ISIT 2004, p.232, Chicago, 2004.
[11] R. Renner and S. Wolf, Simple and Tight Bounds for Information Reconciliation and Privacy Amplification, Advance in Cryptology — ASIACRYPT 2005, LNCS 3788, pp.199–216, Springer-Verlag, 2005.
[12] N. Sato and H. Koga, "New results in optimistic source coding," IEICE Trans. Fundamentals, vol.E87-A, no.10, pp.2577–2580, Oct. 2004.
[13] Y. Steinberg and S. Verdú, "Simulation of random processes and rate-distortion theory," IEEE Trans. Inf. Theory, vol.41, no.1, pp.63–86, 1996.
[14] T. Uyematsu, "A new unified method for fixed-length source coding problems of general sources," IEICE Trans. Fundamentals, vol.E93-A, no.11, pp.1868–1877, Nov. 2010.
[15] T. Uyematsu, "Smooth Rényi entropy and intrinsic randomness," IEICE Technical Report, IT2010-3, 2010.
[16] S. Vembú and S. Verdú, "Generating random bits from an arbitrary source: Fundamental limits," IEEE Trans. Inf. Theory, vol.41, no.5, pp.1322–1332, 1995.

[17] S. Vembú, S. Verdú, and Y. Steinberg, "The source-channel transmission theorem revisited," IEEE Trans. Inf. Theory, vol.41, no.1, pp.44–54, 1995.
[18] J. Wolfowitz, Coding Theorems of Information Theory, 3rd ed., Springer-Verlag, NY, 1978.

Appendix: Proof of Proposition 2.1

We prove here that $\overline{H}(\boldsymbol{X})$ in Definition 2.1 can be expressed in the form of Proposition 2.1. We can use the same method for verification of the equivalence of the two forms of $\overline{H}^*(\boldsymbol{X})$, $\underline{H}^*(\boldsymbol{X})$ and $\underline{H}(\boldsymbol{X})$.

Define

$$\mathcal{A} = \left\{\alpha \in \boldsymbol{R} \,:\, \liminf_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \le \alpha\right\} = 1\right\},$$
$$\mathcal{B} = \left\{\beta \in \boldsymbol{R} \,:\, \limsup_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \ge \beta\right\} > 0\right\}.$$

Set $A = \inf_{\alpha\in\mathcal{A}}\alpha$ and $B = \sup_{\beta\in\mathcal{B}}\beta$. Clearly, A and B correspond to the two expressions of $\overline{H}(\boldsymbol{X})$ in Definition 2.1 and Proposition 2.1, respectively.

We begin with the proof of $B \le A$. To this end, let $\alpha \in \mathcal{A}$ and $\beta \in \mathcal{B}$ be arbitrary elements. We prove $\beta \le \alpha$ by contradiction in the following way. Suppose that $\alpha < \beta$ is true. Then, we have

$$\Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \le \alpha\right\} \le \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} < \beta\right\}$$

for all $n \ge 1$, which implies

$$\begin{aligned}1 &= \liminf_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \le \alpha\right\}\\ &\le \liminf_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} < \beta\right\} \le 1,\end{aligned}$$

where the equality follows from $\alpha \in \mathcal{A}$. Thus, it holds from (49) that

$$\limsup_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \ge \beta\right\} = 0,$$

which contradicts $\beta \in \mathcal{B}$. Therefore, $\beta \le \alpha$ must be satisfied. This establishes $B \le A$ because $\alpha \in \mathcal{A}$ and $\beta \in \mathcal{B}$ are arbitrary.

Next, we prove $A \le B$. Note that $A - \gamma \notin \mathcal{A}$ for any constant $\gamma > 0$ and therefore it holds that

$$\liminf_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \le A - \gamma\right\} < 1.$$

This immediately means

$$\liminf_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} < A - \gamma\right\} < 1,$$

which can be written as

$$\limsup_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \ge A - \gamma\right\} > 0$$

and shows $A - \gamma \in \mathcal{B}$. Now we have $A - \gamma \le B$ due to the definition of B. Since $\gamma > 0$ is arbitrary, $A \le B$ follows.

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 security. Dr. Koga is a member of IEEE information theory society.