### Four Limits in Probability and Their Roles in Source Coding

**Koga Hiroki**

*IEICE transactions on fundamentals of electronics, communications and computer sciences*  
**Volume E94.A**  
**Number 11**  
**Page range 2073-2082**  
**Year 2011-11**

© 2011 The Institute of Electronics, Information and Communication Engineers

URL [http://hdl.handle.net/2241/117318](http://hdl.handle.net/2241/117318)

Four Limits in Probability and Their Roles in Source Coding*

Hiroki KOGA†, 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 \(X = \{X^n\}_{n=1}^{\infty}\) 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 \(\varphi_n\) 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 \(\psi_n\). We are interested in the optimum value of the coding rate \(\frac{1}{n} \log M_n\) that is asymptotically achievable for the pair \((\varphi_n, \psi_n), n \geq 1\), satisfying \(e_n \to 0\) as \(n \to \infty\), where \(\log(\cdot) = \log_2(\cdot)\) throughout this paper and \(e_n = \Pr[\hat{X}^n \neq X^n]\) denotes the decoding error probability. It is known that the optimal rate coincides with the spectral-sup-entropy rate \(\overline{H}(X)\) of the source \(X\), which is defined as the limit superior in probability of \(\frac{1}{n} \log \frac{1}{P_{in}(X^n)}\) [4].

Roughly speaking, \(\overline{H}(X)\) can be interpreted as the right endpoint of the distribution of \(\frac{1}{n} \log \frac{1}{P_{in}(X^n)}\) (called the entropy-spectrum) in an asymptotic sense.

However, \(\overline{H}(X)\) actually means a tight upper bound on the limit superior of the coding rates \(\frac{1}{n} \log M_n, n \geq 1\), at which fixed-to-fixed coding with \(e_n \to 0\) 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 \(\overline{H}(X)\) and attains a sufficiently small decoding error probability. Vembu and Verdú [16] defined the optimistic coding rate as the coding rate at which \(e_n\) 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, \(\underline{H}(X)\) is defined as the limit inferior in probability of \(\frac{1}{n} \log \frac{1}{P_{in}(X^n)}\) [4]. This \(\underline{H}(X)\) corresponds to the left endpoint of the entropy-spectrum in an asymptotic sense. Han [3] gives the strong converse theorem which claims that \(e_n\) goes to one for any fixed-to-fixed length code with the coding rate less than \(\overline{H}(X)\) if and only if \(\overline{H}(X) = \overline{H}(X)\). In addition, \(\overline{H}(X)\) itself is known as the supremum achievable rate of the intrinsic randomness problem [16].

In this paper we newly define \(\overline{H}(X)\) and \(\overline{H}(X)\) as variants of \(\overline{H}(X)\) and \(\overline{H}(X)\), respectively, and discuss roles of \(\overline{H}(X)\), \(\overline{H}(X)\), \(\overline{H}(X)\) and \(\overline{H}(X)\) in coding of a general source \(X\). In fact, \(\overline{H}(X)\) and \(\overline{H}(X)\) are defined as the right and left endpoints of the entropy-spectrum in different asymptotic senses, respectively. The roles of \(\overline{H}(X)\) and \(\overline{H}(X)\) become clear if we consider the infimum achievable rate \(R_e(X)\) of fixed-to-fixed length coding with \(e\)-error in ordinary senses, where the decoding error probability \(e_n\) is required to satisfy lim sup\(_{n \to \infty}\) \(e_n\leq e\) for a constant \(e\in[0,1)\). On the other hand, the role of \(\overline{H}(X)\) and \(\overline{H}(X)\) becomes more understandable if we consider the supremum unachievable rate \(U_e(X)\) with \(e\)-error such that any fixed-to-fixed length code with rate less than \(U_e(X)\) satisfies lim inf\(_{n \to \infty}\) \(e_n > e\). In particular, \(U_e(X)\) with \(e = 0\) turns out to be equal to the infimum achievable rate \(R_e(X)\) with the vanishing decoding error probability in optimistic sense. We can regard the coding theorem on \(U_e(X)\) as a dual counterpart of the ordinary coding theorem on \(R_e(X)\). We should note that, while \(\overline{H}(X)\) and \(\overline{H}(X)\) coincide with \(U_1\) and \(U_0\) 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 \(\overline{H}(X)\) and \(\overline{H}(X)\) but also \(\overline{H}(X)\) and \(\overline{H}(X)\) 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 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 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
We unveil a necessary and sufficient sup- (resp., inf-) entropy rate \([3]\). In addition, we also show that the four quantities are expressed in terms of the smooth Rényi entropy rate \([10, 11]\) of order zero, where the expressions for \(\overline{H}(X)\) and \(\underline{H}(X)\) are given in \([14]\).

The rest of this paper is organized as follows. In Sect. 2 we formally define \(\overline{H}(X)\) and \(\underline{H}(X)\) as well as \(\overline{H}(X)\) and \(\underline{H}(X)\) and investigate their fundamental relationships. We unveil a necessary and sufficient condition under which \(\overline{H}(X) \leq \overline{H}(X) \leq \overline{H}(X) \leq \overline{H}(X)\) 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 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 \(\overline{H}(X), \overline{H}(X), \overline{H}(X)\) and \(\overline{H}(X)\).

2. Four Limits in Probability

Let \(X\) be finite or countably infinite alphabet. For each \(n \geq 1\) let \(X^n\) be a random variable taking values in \(X^n\). Denote by \(P_{X^n}\) the probability distribution of \(X^n\). The probability of \(X^n = x^n\) for an \(x^n \in X^n\) is denoted by \(P_{X^n}(x^n)\), where throughout this paper random variables and their realizations are expressed in the case upper-case and lower-case letters, respectively. We do not require that \(P_{X^n}\), \(n \geq 1\), satisfy the consistency condition. We call \(X = \{X^n\}_{n=1}^{\infty}\) 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 \(X = \{X^n\}_{n=1}^{\infty}\) we define:

\[
\overline{H}(X) = \inf \left\{ \alpha : \liminf_{n \to \infty} \Pr \left( \frac{1}{n} \log P_{X^n}(x^n) \leq \alpha \right) = 1 \right\},
\]

\[
\underline{H}(X) = \inf \left\{ \alpha : \limsup_{n \to \infty} \Pr \left( \frac{1}{n} \log P_{X^n}(x^n) \leq \alpha \right) = 1 \right\},
\]

\[
\overline{H}(X) = \sup \left\{ \beta : \liminf_{n \to \infty} \Pr \left( \frac{1}{n} \log P_{X^n}(x^n) \geq \beta \right) = 1 \right\},
\]

\[
\underline{H}(X) = \sup \left\{ \beta : \limsup_{n \to \infty} \Pr \left( \frac{1}{n} \log P_{X^n}(x^n) \geq \beta \right) = 1 \right\},
\]

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

Note that \(\overline{H}(X)\) (resp., \(\underline{H}(X)\)) is equal to the ordinary spectrum sup- (resp., inf-) entropy rate \([3]\). In addition, \(\overline{H}(X)\) and \(\underline{H}(X)\) coincide with \(\overline{H}_1\) and \(\underline{H}_0\) in \([2, p. 204]\), respectively. While it is known that \(0 \leq \overline{H}(X) \leq \overline{H}(X) \leq \infty [3]\), we assume that \(\overline{H}(X) < \infty\) 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 \(H(X), H(X), \overline{H}(X)\) and \(\underline{H}(X)\) 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:

\[
\overline{H}(X) = \sup \left\{ \beta : \limsup_{n \to \infty} \Pr \left( \frac{1}{n} \log P_{X^n}(x^n) \geq \beta \right) > 0 \right\},
\]

\[
\overline{H}(X) = \sup \left\{ \beta : \limsup_{n \to \infty} \Pr \left( \frac{1}{n} \log P_{X^n}(x^n) \geq \beta \right) > 0 \right\},
\]

\[
\overline{H}(X) = \inf \left\{ \alpha : \limsup_{n \to \infty} \Pr \left( \frac{1}{n} \log P_{X^n}(x^n) \leq \alpha \right) > 0 \right\},
\]

\[
\overline{H}(X) = \inf \left\{ \alpha : \limsup_{n \to \infty} \Pr \left( \frac{1}{n} \log P_{X^n}(x^n) \leq \alpha \right) > 0 \right\}.
\]

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

\[
\limsup_{n \to \infty} \Pr \left( \frac{1}{n} \log P_{X^n}(x^n) \leq \overline{H}(X) + \gamma \right) = 1, \quad (1)
\]

\[
\limsup_{n \to \infty} \Pr \left( \frac{1}{n} \log P_{X^n}(x^n) \leq \overline{H}(X) - \gamma \right) < 1 \quad (2)
\]

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

\[
\liminf_{n \to \infty} \Pr \left( \frac{1}{n} \log P_{X^n}(x^n) \geq \overline{H}(X) - \gamma \right) < 0, \quad (3)
\]

\[
\liminf_{n \to \infty} \Pr \left( \frac{1}{n} \log P_{X^n}(x^n) \geq \overline{H}(X) + \gamma \right) = 0 \quad (4)
\]

for any constant \(\gamma > 0\). The relationships similar to (1)–(4) hold for \(\overline{H}(X), \overline{H}(X)\) and \(\overline{H}(X)\) 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 \(X = \{X^n\}_{n=1}^{\infty}\) is the stationary memoryless source specified by a probability distribution \(P_X\) on \(X\) 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 \(x^n = x_1, x_2, \ldots, x_n\) for this source. Then, it holds that

\[
\overline{H}(X) = \overline{H}(X) = \overline{H}(X) = \overline{H}(X) = \overline{H}(X)
\]

owing to the weak law of large numbers.

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

\[
\overline{H}(X) = \overline{H}(X) = \overline{H}(X) = \overline{H}(X) = \overline{H}(X)
\]
Then, it holds that

\[ \overline{H}(X) = \underline{H}(X) = H(P_X), \quad H(X) = \overline{H}^*(X) = H(P_X). \]

In particular, \( \overline{H}(X) = H(P_X) \) is verified from

\[
\liminf_{n \to \infty} \Pr \left( \frac{1}{n} \log \frac{1}{P_X(x^n)} \geq H(P_X) - \gamma \right) \geq \tau > 0, \quad (5)
\]

\[
\liminf_{n \to \infty} \Pr \left( \frac{1}{n} \log \frac{1}{P_X(x^n)} \geq H(P_X) + \gamma \right) = 0 \quad (6)
\]

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

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

\[
P_X(x^n) = \begin{cases} P_{X'}(x^n), & \text{if } n \text{ is odd}, \\ P_{X''}(x^n), & \text{if } n \text{ is even}. \end{cases}
\]

Then, it holds that

\[ \overline{H}(X) = H^*(X) = H(P_X) \quad \text{and} \quad H(X) = \overline{H}(X) = H(P_X). \]

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

\[
\liminf_{n \to \infty} \Pr \left( \frac{1}{n} \log \frac{1}{P_X(x^n)} \geq H(P_X) - \gamma \right) = 1 > 0, \quad (7)
\]

\[
\liminf_{n \to \infty} \Pr \left( \frac{1}{n} \log \frac{1}{P_X(x^n)} \geq H(P_X) + \gamma \right) = 0 \quad (8)
\]

for any constant \( \gamma > 0 \). Hence, \( \overline{H}(X) = H(P_X) \) follows.

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

\[
P_X^{(1)}(x^n) = \begin{cases} (1 - \tau)P_X(x^n) + \tau P_{X_1}(x^n), & \text{if } n \text{ is odd}, \\ (1 - \tau')P_X(x^n) + \tau' P_{X_1}(x^n), & \text{if } n \text{ is even}, \end{cases}
\]

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

\[
P_X^{(2)}(x^n) = \begin{cases} (1 - \tau)P_X(x^n) + \tau P_{X_2}(x^n), & \text{if } n \text{ is odd}, \\ (1 - \tau')P_X(x^n) + \tau' P_{X_2}(x^n), & \text{if } n \text{ is even}, \end{cases}
\]

and denote by \( X^{(2)} \) the general source corresponding to \( P_X^{(2)}, n \geq 1 \). We can verify that \( \overline{H}(X^{(i)}) = H(P_X), \overline{H}(X^{(i)}) = H(P_X), \overline{H}(X^{(i)}) = H(P_X) \) for \( i = 1, 2 \) (see Fig. 1).

In general, the following relationships hold.

**Proposition 2.2:** For a general source \( X \) it holds that

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

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

\[ \overline{H}(X) = \overline{H}(X) = H(P_X), \quad H(X) = \overline{H}^*(X) = H(P_X). \]

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

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

**Definition 2.2:** We say a general source \( X = \{X^n\}_{n=1}^{\infty} \) is canonical if there exists a constant \( \alpha_0 \) such that for any \( \gamma > 0 \) both the following two inequalities hold:

\[
\liminf_{n \to \infty} \Pr \left( \frac{1}{n} \log \frac{1}{P_X(x^n)} \geq \alpha_0 - \gamma \right) > 0, \quad (9)
\]

\[
\liminf_{n \to \infty} \Pr \left( \frac{1}{n} \log \frac{1}{P_X(x^n)} \geq \alpha_0 + \gamma \right) > 0. \quad (10)
\]

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 \( \alpha_0 \) satisfying \( H(P_X) \leq \alpha_0 \leq H(P_X) \) 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 \( \alpha_0 \) so that both (9) and (10) are satisfied.

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

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

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

\[
\liminf_{n \to \infty} \Pr \left( \frac{1}{n} \log \frac{1}{P_X(x^n)} \leq H^*(X) + \gamma \right) > 0. \quad (11)
\]

On the other hand, \( \overline{H}^*(X) \) in Proposition 2.1 implies that
\[
\liminf_{n \to \infty} \Pr \left\{ \frac{1}{n} \log \frac{1}{P_{X^n}(X^n)} \geq \overline{H}(X) - \gamma \right\} > 0,
\]
which, together with \( H'(X) \leq \overline{H}(X) \), leads to
\[
\liminf_{n \to \infty} \Pr \left\{ \frac{1}{n} \log \frac{1}{P_{X^n}(X^n)} \geq H'(X) - \gamma \right\} > 0. \tag{12}
\]

Then, (11) and (12) guarantee that \( X \) is canonical because (9) and (10) are satisfied with \( \alpha_0 = H'(X) \).

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

3. Operational Meanings

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

We describe the ordinary \( \varepsilon \)-source coding problem as follows. Given a general source \( X = \{X^n\}_{n=1}^{\infty} \), let an encoder \( \varphi_n : X^n \to M_n \overset{\text{def}}{=} \{1, 2, \ldots, M_n\} \) and a decoder \( \psi_n : M_n \to X^n \) be deterministic mappings. Define the decoding error probability by
\[
\varepsilon_n \overset{\text{def}}{=} \Pr[\varphi_n(\varphi_n(X^n)) \neq X^n].
\]
We are interested in the infimum of the coding rates at which \( \varepsilon_n \) 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 \( \varepsilon \)-achievable if there exists a sequence \( \{\varphi_n, \psi_n\}_{n=1}^{\infty} \) of pairs of an encoder and a decoder satisfying
\[
\limsup_{n \to \infty} \frac{1}{n} \log M_n \leq R \tag{13}
\]
and
\[
\limsup_{n \to \infty} \varepsilon_n \leq \varepsilon. \tag{14}
\]
Denote by \( R_\varepsilon(X) \) the infimum of the \( \varepsilon \)-achievable rates.

For simplifying notations we define
\[
\overline{J}_\varepsilon(X) = \inf\{R : \overline{F}(R) \leq \varepsilon\}, \tag{15}
\]
\[
\underline{J}_\varepsilon(X) = \inf\{R : \underline{F}(R) \leq \varepsilon\}, \tag{16}
\]
where
\[
\overline{F}(R) = \limsup_{n \to \infty} \Pr \left\{ \frac{1}{n} \log \frac{1}{P_{X^n}(X^n)} \geq R \right\},
\]
\[
\underline{F}(R) = \liminf_{n \to \infty} \Pr \left\{ \frac{1}{n} \log \frac{1}{P_{X^n}(X^n)} \geq R \right\}.
\]
Clearly, it holds that \( \overline{J}_\varepsilon(X) = \overline{H}(X) \) and \( \underline{J}_\varepsilon(X) = \underline{H}(X) \).
The following theorem gives a formula of \( R_\varepsilon(X) \).

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

**Proof:** Since the first equality in (17) is obvious from the operational meaning of \( \overline{H}(X) \) in fixed-length coding [4], we prove the second equality in (17). In view of Theorem 3.1, we have only to prove \( H'(X) = \lim_{\varepsilon \downarrow 1} \overline{J}_\varepsilon(X) \). We note that \( \overline{F}(R) \) is a monotone decreasing function of \( R \) and \( \overline{J}_\varepsilon(X) \) is a monotone decreasing function of \( \varepsilon \). In particular, the expression of \( H'(X) \) in Definition 2.1 guarantees that \( \overline{F}(R) = 1 \) if \( R < H'(X) \) and \( \overline{F}(R) < 1 \) if \( R > H'(X) \). Hence, letting \( \gamma > 0 \) be an arbitrarily small constant, for any \( \varepsilon \in (0, 1) \) it holds that \( \overline{J}_\varepsilon(X) \geq H'(X) - \gamma \) because \( \overline{F}(H'(X) - \gamma) = 1 - \varepsilon \). By letting \( \varepsilon \downarrow 1 \), we have
\[
\lim_{\varepsilon \downarrow 1} \overline{J}_\varepsilon(X) \geq H'(X) - \gamma. \tag{18}
\]
In addition, since \( \overline{F}(H'(X) + \gamma) < 1 \), there exists a positive number \( \delta_0 \) such that \( \overline{F}(H'(X) + \gamma) \leq 1 - \delta_0 \). Therefore, owing to the definition of \( \underline{J}_\varepsilon(X) \) it holds that
\[
\lim_{\varepsilon \downarrow 1} \underline{J}_\varepsilon(X) \leq \underline{J}_{1-\delta_0}(X) \leq H'(X) + \gamma, \tag{19}
\]
where the first inequality follows because \( \overline{J}_\varepsilon(X) \) is a monotone decreasing function of \( \varepsilon \) and the second inequality follows from the definition of \( \underline{J}_{1-\delta_0}(X) \). Since \( \gamma > 0 \) is arbitrary in (18) and (19), \( H'(X) = \lim_{\varepsilon \downarrow 1} \overline{J}_\varepsilon(X) \) follows. \( \square \)

Next, let us consider the operational meanings of \( H'(X) \) and \( \overline{H}(X) \). 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 \( \varepsilon \)-unachievable in the strong sense if
\[
\liminf_{n \to \infty} \varepsilon_n > \varepsilon \tag{20}
\]
holds for any sequence \( \{\varphi_n, \psi_n\}_{n=1}^{\infty} \) of pairs of an encoder and a decoder satisfying (13). Denote by \( U_\varepsilon(X) \) the supremum of the \( \varepsilon \)-unachievable rates in the strong sense.

It is obvious from the definition of \( U_\varepsilon(X) \) that for any \( R > U_\varepsilon(X) \) there exists a sequence \( \{\varphi_n, \psi_n\}_{n=1}^{\infty} \) of pairs of an encoder \( \varphi_n \) and a decoder \( \psi_n \) satisfying (13) and
Hence, \( U_e(X) \) equals to the infimum achievable coding rate the codes satisfying (13) and (21). Thus, the formula for \( U_e(X) \) coincides with the formula of \( R^*(\varepsilon, \hat{p}) \) in [5, Theorem 1]. However, since \( U_e(X) \) is closely related to the infimum \( \varepsilon \)-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 \( \varepsilon \)-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

\[
\frac{1}{n_i} \log M_n \leq R + \gamma \quad \text{and} \quad \varepsilon_{n_i} \leq \varepsilon + \gamma \quad \text{for all} \quad i \geq 1. \tag{22}
\]

Denote by \( R_e^*(X) \) the infimum of the \( \varepsilon \)-achievable rates in the optimistic sense.

The following theorem is obtained from simple observations on \( U_e(X) \) and \( R_e^*(X) \).

**Theorem 3.2:** \( R_e^*(X) = U_e(X) \) for all \( \varepsilon \in [0, 1) \).

**Proof:** If \( R < U_e(X) \), 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 \( R_e^*(X) \geq U_e(X) \). On the other hand, if \( R > U_e(X) \), 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 \( R_e^*(X) \leq U_e(X) \).

Theorem 3.2 leads to the following formula of \( U_e(X) \).

**Theorem 3.3** ([2], [5], [12]): \( U_e(X) = J_e(X), \varepsilon \in [0, 1) \).

Similarly to Corollary 3.1, we can obtain the following operational meaning of \( H^\varepsilon(X) \) and \( H(X) \), where the first equality in (23) is mentioned in [2].

**Corollary 3.2:** It holds that

\[
U_0(X) = H^\varepsilon(X) \quad \text{and} \quad \lim_{\varepsilon \uparrow 1} U_e(X) = H(X). \tag{23}
\]

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. Walfowitz defined that a source satisfies the strong converse property if for any coding rate \( R < R_0(X) \) any sequence \( \{ (\varphi_n, \psi_n) \}_{n=1}^{\infty} \) of pairs of an encoder and a decoder meeting the rate constraint (13) satisfies \( \varepsilon_n \rightarrow 1 \) as \( n \rightarrow \infty \) [18]. Han shows that a general source \( X \) satisfies the strong converse property if and only if \( H(X) = H(X) \) [3]. Note that, in view of Proposition 2.2, \( H(X) = H(X) \) actually means that \( H(X) = \overline{H}(X) = H(X) \). 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 \( \varepsilon \)-strong converse property.

**Definition 4.1:** Let \( \varepsilon \in [0, 1) \) be an arbitrary constant. A general source \( X \) is said to satisfy the \( \varepsilon \)-strong converse property if for any coding rate \( R < R_e(X) \) 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 \( \varepsilon \)-strong converse property with \( \varepsilon = 0 \).

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

The proof of Theorem 4.1 is immediate from Theorems 3.1 and 3.3. Note that \( U_e(X) \leq R_0(X) \) from their definitions and recall that any \( \{ (\varphi_n, \psi_n) \}_{n=1}^{\infty} \) satisfying (13) meets (20) if \( R < U_e(X) \). Hence, if \( U_e(X) = R_e(X) \), \( X \) satisfies the \( \varepsilon \)-strong converse property. Conversely, suppose that \( X \) satisfies the \( \varepsilon \)-strong converse property. If \( U_e(X) < R_e(X) \), then for any \( R \) satisfying \( U_e(X) < R < R_e(X) \) there exists a \( \{ (\varphi_n, \psi_n) \}_{n=1}^{\infty} \) satisfying (13) and (21). This conflicts the assumption that \( X \) satisfies the \( \varepsilon \)-strong converse property.

This simple argument on the \( \varepsilon \)-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 \( X \) is said to satisfy the optimistic strong converse property if for any coding rate \( R < U_0(X) = \overline{H}(X) \) any \( \{ (\varphi_n, \psi_n) \}_{n=1}^{\infty} \) meeting (13) satisfies \( \varepsilon_n \rightarrow 1 \).

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

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

The definition of the optimistic strong converse property (24) can be weakened to \( \lim \sup_{n \rightarrow \infty} \varepsilon_n = 1 \). Obviously, \( \overline{H}(X) = \overline{H}(X) \) is the necessary and sufficient condition for the weakened optimistic strong converse property.

**Example 4.1:** Consider the source \( X^{(1)} \) in Example 2.4. This source satisfies the zero-strong converse property if and only if \( H(P_{X_1}) = H(P_{X_1}) \). Note that we asymptotically have a mass at \( R = H(P_{X_1}) \) in this case. This source satisfies the optimistic strong converse property if and only if \( H(P_{X_1}) = H(P_{X_1}) \). Obviously, if \( R > H(P_{X_1}) \), 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 $H(P_{X_n}) = H(P_{X_1})$.

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 $\varepsilon$-source coding problem [10], [11]. In particular, the smooth Rényi entropy of order zero for a random variable $X^n \in X^n$ can be written as

$$K_0(X^n) = \min_{P_X \in \mathcal{A}_n} \frac{1}{n} \log |\mathcal{A}_n|,$$

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

$$\lim_{\delta \to 0} \limsup_{n \to \infty} \frac{1}{n} K_0(X^n) = \overline{H}(X),$$

$$\lim_{\delta \to 1} \liminf_{n \to \infty} \frac{1}{n} K_0(X^n) = \underline{H}(X).$$

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

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

$$\lim_{\delta \to 0} \liminf_{n \to \infty} \frac{1}{n} K_\delta(X^n) = \overline{H}'(X),$$

$$\lim_{\delta \to 1} \limsup_{n \to \infty} \frac{1}{n} K_\delta(X^n) = \underline{H}'(X).$$

For simplifying the notations we define

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

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

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

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

$$\Pr\{X^n \in \mathcal{V}_n\} \geq 1 - \delta - 2^{-n\gamma}$$

for all $n \geq 1$, where

$$\mathcal{V}_n = \left\{x^n \in X^n : \frac{1}{n} \log \frac{1}{P_{X^n}(x^n)} \leq \frac{1}{n} K_\delta(X^n) + \gamma \right\}.$$  (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 $\mathcal{A}_n \subset X^n$ satisfying

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

Then, it follows that

$$\Pr\{X^n \in \mathcal{A}_n\} = \Pr\{X^n \in \mathcal{A}_n \cap \mathcal{V}_n\} + \Pr\{X^n \in \mathcal{A}_n \cap \mathcal{V}_n^c\} \leq \Pr\{X^n \in \mathcal{V}_n\} + \Pr\{X^n \in \mathcal{A}_n \cap \mathcal{V}_n^c\},$$

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

$$\Pr\{X^n \in \mathcal{A}_n \cap \mathcal{V}_n^c\} \leq \sum_{x^n \in \mathcal{A}_n \cap \mathcal{V}_n^c} P_{X^n}(x^n) \leq \frac{|\mathcal{A}_n \cap \mathcal{V}_n^c|}{|\mathcal{A}_n|^{2^{n\gamma}}} \leq 2^{-n\gamma}.$$  (32)

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

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

$$\lim_{n \to \infty} \Pr\left\{ \frac{1}{n} \log \frac{1}{P_{X^n}(x^n)} \leq K_\delta(X) + 2\gamma \right\} = 1$$

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

$$\mathcal{U}_n = \left\{x^n \in X^n : \frac{1}{n} \log \frac{1}{P_{X^n}(x^n)} \leq K_\delta(X) + 2\gamma \right\}.$$  (34)

Since

$$\frac{1}{n} K_\delta(X^n) \leq \overline{K}_0(X) + \gamma$$

infinitely often from the definition of $\overline{K}_0(X)$, it holds that

$$\Pr\{X^n \in \mathcal{V}_n\} \leq \Pr\{X^n \in \mathcal{U}_n\}$$

infinite often.  (35)

In addition, since $K_\delta(X)$ is monotone decreasing with respect to $\delta$, we have

$$\Pr\left\{ \frac{1}{n} \log \frac{1}{P_{X^n}(x^n)} \leq K_\delta(X) + 2\gamma \right\} \geq \Pr\{X^n \in \mathcal{U}_n\}$$

for all $n \geq 1$. (36) Then, the combination of (36), (37) and Lemma 5.1 yields

$$\Pr\left\{ \frac{1}{n} \log \frac{1}{P_{X^n}(x^n)} \leq K_\delta(X) + 2\gamma \right\} \geq 1 - \delta - 2^{-n\gamma}$$

infinite often.  (38)

This guarantees

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

Since $\delta \in (0, 1)$ in (39) is arbitrary, we have (34). In order to establish (a-2), we will establish

$$\lim_{n \to \infty} \inf \Pr\left\{ \frac{1}{n} \log \frac{1}{P_{X^n}(x^n)} \leq \overline{K}_0(X) + 2\gamma \right\} > 0$$

(40)
for any \( \delta \in (0, 1) \) and \( \gamma > 0 \). This, together with the expression of \( H'(X) \) in Proposition 2.1, implies that \( H'(X) \leq K_\delta(X) \) for any \( \delta \in (0, 1) \). Inequality (a-2) can be obtained by letting \( \delta \uparrow 1 \). Hereinafter, we establish (40). Define

\[ U'_n = \left\{ x^n \in X^n : \frac{1}{n} \log \frac{1}{P_{X^n}(x^n)} \leq K_\delta(X) + 2\gamma \right\}. \quad (41) \]

From the definition of \( K_\delta(X) \), for any \( \delta \in (0, 1) \) and \( \gamma > 0 \) there exists an integer \( n_0 = n_0(\gamma, \delta) \) satisfying

\[ \frac{1}{n} K_\delta(x^n) \leq K_\delta(X) + \gamma \quad \text{for all } n \geq n_0. \]

Since \( V_n \subset U'_n \) for all \( n \geq n_0 \), it holds that

\[ \Pr\{x^n \in U'_n\} \geq \Pr\{x^n \in V'_n\} \geq 1 - \delta - 2^{-n\gamma} \quad \text{for all } n \geq n_0, \quad (42) \]

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

\[ \liminf_{n \to \infty} \Pr\left( \frac{1}{n} \log \frac{1}{P_{X^n}(x^n)} \leq K_\delta(X) + 2\gamma \right) \geq 1 - \delta, \]

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

Next, we prove (b-1) \( K_\delta(X) \leq H'(X) \) and (b-2) \( K_1(X) \leq H'(X) \). Inequalities (b-1) and (b-2) are proved in almost the same way. We begin with the proof of (b-1). Define

\[ S_n = \left\{ x^n \in X^n : \frac{1}{n} \log \frac{1}{P_{X^n}(x^n)} \leq H'(X) + \gamma \right\}. \quad (43) \]

Since \( P_{X^n}(x^n) \geq 2^{-n\overline{H}(X)+\gamma} \) for all \( x^n \in S_n \), we can easily verify that

\[ |S_n| \leq 2^{n\overline{H}(X)+\gamma} \quad \text{for all } n \geq 1. \quad (44) \]

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

\[ \Pr\{x^n \in S_n\} \geq 1 - \delta \quad \text{infinitely often}. \quad (45) \]

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

\[ \frac{1}{n} K_\delta(x^n) \leq H'(X) + \gamma \quad \text{infinitely often}, \]

which yields

\[ K_1(X) = \liminf_{n \to \infty} \frac{1}{n} K_\delta(x^n) \leq H'(X) + \gamma. \]

We can establish \( K_\delta(X) \leq H'(X) + \gamma \) by letting \( \delta \downarrow 0 \). Since \( \gamma > 0 \) is arbitrary, \( K_\delta(X) \leq H'(X) \) is established.

Finally, we prove (b-2). In view of the expression of \( H'(X) \) in Proposition 2.1, for any \( \gamma > 0 \) there exists a constant \( \epsilon_0 > 0 \) such that

\[ \Pr\left( \frac{1}{n} \log \frac{1}{P_{X^n}(x^n)} \leq H'(X) + \gamma \right) \geq \epsilon_0 \quad \text{for all } n \geq n_0. \quad (46) \]

Define \( S_n \) by (43), where \( \overline{H}(X) \) is replaced with \( H'(X) \). Then, we have \( |S_n| \leq 2^{n\overline{H}(X)+\gamma} \) for all \( n \geq 1 \) and

\[ \frac{1}{n} K_{1-\epsilon_0}(x^n) \leq H'(X) + \gamma \quad \text{for all } n \geq n_0 \]

from the definition of \( K_{1-\epsilon_0}(X^n) \). By taking the limit superior of both sides, it holds that

\[ \overline{K}_{1-\epsilon_0}(X) = \limsup_{n \to \infty} \frac{1}{n} K_{1-\epsilon_0}(x^n) \leq H'(X) + \gamma. \quad (47) \]

Since \( \overline{K}_\delta(X) \) is monotone decreasing with respect to \( \delta \), we have \( \overline{K}_1(X) \leq \overline{K}_{1-\epsilon_0}(X) \). Therefore, (47) leads to

\[ \overline{K}_1(X) \leq H'(X) + \gamma. \quad (48) \]

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

**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 \( \overline{H}(X) \) is assumed to be finite. Then, we can define

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

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

\[ \mathcal{G} = \left\{ (a_n, b_n)_{n=1}^{\infty} : \text{for any constant } \gamma > 0 \right\} \]

\[ \liminf_{n \to \infty} \Pr\left( \frac{1}{n} \log \frac{1}{P_{X^n}(x^n)} \in (a_n - \gamma, b_n + \gamma) \right) = 1. \]

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 \( W(X) \) was defined in [6] in the context of a fixed-length homomorphic coding. Quite recently, Arimura and Iwata showed that \( W(X) \) coincides with the infimum achievable redundancy of fixed-to-fixed length coding [1].

We also define

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

where

\[ \mathcal{G}^* = \left\{ (a_n, b_n)_{n=1}^{\infty} : \text{for any constant } \gamma > 0 \right\} \]

\[ \limsup_{n \to \infty} \Pr\left( \frac{1}{n} \log \frac{1}{P_{X^n}(x^n)} \in (a_n - \gamma, b_n + \gamma) \right) = 1. \]

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

The following proposition gives bounds on \( W(X) \) and
Proposition 6.1: For a general source $X$ we have:

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

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

$$ \liminf_{n \to \infty} (-p_n) = - \limsup_{n \to \infty} p_n, $$
$$ \limsup_{n \to \infty} (-p_n) = - \liminf_{n \to \infty} p_n, $$
$$ \liminf_{n \to \infty} p_n + \limsup_{n \to \infty} q_n \leq \liminf_{n \to \infty} (p_n + q_n), $$
$$ \liminf_{n \to \infty} p_n + \limsup_{n \to \infty} q_n \leq \limsup_{n \to \infty} p_n + \limsup_{n \to \infty} q_n. $$

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

$$ \Pr \left\{ \frac{1}{n} \log \frac{1}{P_{X^n}(x^n)} \in (H(X) - \gamma, \overline{H}(X) + \gamma) \right\} $$
$$ = 1 - \left[ \Pr \left\{ \frac{1}{n} \log \frac{1}{P_{X^n}(x^n)} \leq H(X) - \gamma \right\} + \Pr \left\{ \frac{1}{n} \log \frac{1}{P_{X^n}(x^n)} \geq \overline{H}(X) + \gamma \right\} \right]. $$

Therefore, (53) leads to

$$ \liminf_{n \to \infty} \Pr \left\{ \frac{1}{n} \log \frac{1}{P_{X^n}(x^n)} \in (H(X) - \gamma, \overline{H}(X) + \gamma) \right\} $$
$$ \geq 1 - \limsup_{n \to \infty} \Pr \left\{ \frac{1}{n} \log \frac{1}{P_{X^n}(x^n)} \leq H(X) - \gamma \right\} - \limsup_{n \to \infty} \Pr \left\{ \frac{1}{n} \log \frac{1}{P_{X^n}(x^n)} \geq \overline{H}(X) + \gamma \right\}, $$

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 $H(X)$ and $\overline{H}(X)$ 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 $((\overline{H}(X), \underline{H}(X)))_{n=1}^{\infty} \in G$. Since $W(X)$ is defined as the infimum with respect to elements of $G$, $W(X) \leq \overline{H}(X) - \underline{H}(X)$ follows.

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

$$ \limsup_{n \to \infty} \Pr \left\{ \frac{1}{n} \log \frac{1}{P_{X^n}(x^n)} \in (\overline{H}'(X) - \gamma, \overline{H}(X) + \gamma) \right\} $$
$$ \geq 1 - \liminf_{n \to \infty} \Pr \left\{ \frac{1}{n} \log \frac{1}{P_{X^n}(x^n)} \leq \overline{H}'(X) - \gamma \right\} - \limsup_{n \to \infty} \Pr \left\{ \frac{1}{n} \log \frac{1}{P_{X^n}(x^n)} \geq \overline{H}(X) + \gamma \right\}, $$

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 $\overline{H}'(X)$ and $\overline{H}(X)$ 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 $((\overline{H}'(X), \overline{H}(X)))_{n=1}^{\infty} \in G'$. We now have $W^*(X) \leq \overline{H}(X) - \overline{H}'(X)$ in view of the definition of $W^*(X)$.

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

$$ b_n - a_n \leq \overline{H}'(X) - \overline{H}'(X) - 4\gamma_0 $$
for all $n \geq n_0$. (56)

We define the four intervals as follows:

$$ E_n^{(1)} = \{ t \in R : t \leq a_n - \gamma_0 \}, $$
$$ E_n^{(2)} = \{ t \in R : t \geq b_n + \gamma_0 \}, $$
$$ F_n^{(1)} = \{ t \in R : t \leq \overline{H}'(X) + \gamma_0 \}, $$
$$ F_n^{(2)} = \{ t \in R : t \geq \overline{H}(X) - \gamma_0 \}, $$

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

$$ \Pr \left\{ \frac{1}{n} \log \frac{1}{P_{X^n}(x^n)} \in E_n^{(1)} \cup E_n^{(2)} \right\} $$
$$ \geq \min_{i=1,2} \Pr \left\{ \frac{1}{n} \log \frac{1}{P_{X^n}(x^n)} \in F_n^{(i)} \right\} $$
for all $n \geq n_0$, (61)

which guarantees

$$ \liminf_{n \to \infty} \Pr \left\{ \frac{1}{n} \log \frac{1}{P_{X^n}(x^n)} \in E_n^{(1)} \cup E_n^{(2)} \right\} > 0 $$
for all $n \geq n_0$, (62)
owing to the expressions of $H'(X)$ and $\overline{H}'(X)$ 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 $W(X) \geq \overline{H}(X) - H'(X)$ and $W(X) \geq \overline{H}'(X) - H(X)$. We show only $W(X) \geq \overline{H}(X) - H'(X)$ below because $W(X) \geq \overline{H}'(X) - H(X)$ can be proved in the same way. Similarly to the proof of (d), we prove $W(X) \geq \overline{H}(X) - H'(X)$ by contradiction. Suppose that $\overline{H}(X) - H'(X)$ and assume that there exist a constant $\gamma_0 > 0$ and an integer $n_0$ satisfying

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

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

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

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 $H'(X)$ and $\overline{H}'(X)$ 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 $\{(a_n, b_n)\}_{n=1}^\infty \in \mathcal{G}$. \hfill \Box

Example 6.1: Let us consider the general source $X^{(1)}$ in Example 2.4. In view of the definitions of $W(X)$ and $W'(X)$, we have $W(X^{(1)}) = H(P_{X^{(1)}}) - H(P_{X^{(1)}}) = \overline{H}(X^{(1)}) - H'(X^{(1)})$ and $W'(X^{(1)}) = H(P_{X^{(1)}}) - H(P_{X^{(1)}}) = \overline{H}'(X^{(1)}) - H'(X^{(1)})$. That is, the bounds in (b) and (d) are satisfied with equality. On the other hand, for the general source $X^{(2)}$ in Example 2.4, if $H(P_{X^{(2)}}) - H(P_{X^{(2)}}) - H(P_{X^{(2)}})$ is assumed, we have $W(X^{(2)}) = H(P_{X^{(2)}}) - H(P_{X^{(2)}}) = \overline{H}(X^{(2)}) - H'(X^{(2)})$ and $W'(X^{(2)}) = H(P_{X^{(2)}}) - H(P_{X^{(2)}}) = \overline{H}'(X^{(2)}) - H'(X^{(2)})$ 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 $\overline{H}'(X)$ and $H'(X)$ 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 $\overline{H}'(X)$ and $H'(X)$ as new terminology 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

Appendix: Proof of Proposition 2.1

We prove here that $\overline{H}(X)$ 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 $\overline{H}(X)$, $\overline{H}(X)$, and $\overline{H}(X)$.

Define

$\mathcal{A} = \left\{ \alpha \in R : \liminf_{n \to \infty} \Pr \left( 1/n \log \frac{1}{P_X(X^n)} \leq \alpha \right) = 1 \right\}$,

$\mathcal{B} = \left\{ \beta \in R : \limsup_{n \to \infty} \Pr \left( 1/n \log \frac{1}{P_X(X^n)} \geq \beta \right) > 0 \right\}$.

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 $\overline{H}(X)$ in Definition 2.1 and Proposition 2.1, respectively.

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

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

for all $n \geq 1$, which implies

$$1 = \liminf_{n \to \infty} \Pr \left( \frac{1}{n} \log \frac{1}{P_X(X^n)} \leq \alpha \right) \leq \liminf_{n \to \infty} \Pr \left( \frac{1}{n} \log \frac{1}{P_X(X^n)} < \beta \right) \leq 1,$$

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

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

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

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

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

This immediately means

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

which can be written as

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

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