On the Conditional Smooth Rényi Entropy and its Applications in Guessing and Source Coding

Shigeaki Kuzuoka, Member, IEEE

Abstract—A novel definition of the conditional smooth Rényi entropy, which is different from that of Renner and Wolf, is introduced. It is shown that our definition of the conditional smooth Rényi entropy is appropriate for providing lower and upper bounds on the optimal guessing moment in a guessing problem where the guesser is allowed to stop guessing and declare an error. Further a general formula for the optimal guessing exponent is presented. In particular, a single-letterized formula for a mixture of i.i.d. sources is obtained. It is also shown that our definition is appropriate to characterize the optimal exponential moment of the codeword length in the problem of source coding with common side-information available at the encoder and decoder under a constraint on the probability of a decoding error.

Index Terms—Guessing, information-spectrum method, side information, source coding, conditional smooth Rényi entropy.

I. INTRODUCTION

Let us consider the problem of guessing the value of a random variable $X$ by asking questions of the form “Is $X$ equal to $x$?”. This guessing game was introduced by Massey [1], where the expectation $\mathbb{E}[G(X)]$ of the number $G(x)$ of guesses required to reveal $X$ was investigated. Subsequently Arikan [2] investigated the $\rho$-th moment $\mathbb{E}[G(X)^\rho]$ of the number of guesses for $\rho \geq 0$ and demonstrated that the Rényi entropy [3] fully characterizes the moment (up to a factor $\log |X|$). He also investigated the problem of guessing $X$ with the side-information $Y$, where the conditional Rényi entropy [4] characterizes the moment of the number of guesses.

In this paper, we consider the variation of the problem of guessing $X$ with the side-information $Y$ where the guesser is allowed to stop guessing and declare an error. We evaluate the $\rho$-th moment of the number of guesses under a constraint on the probability of the error. To do this, we introduce the conditional smooth Rényi entropy.

The concept of a “smoothed” version of the Rényi entropy was introduced by Renner and Wolf [5], [6]. They defined the conditional $\varepsilon$-smooth Rényi entropy $\tilde{H}_\varepsilon^\alpha(X|Y)$ of order $\alpha$, and showed the significance of two special cases of $\tilde{H}_\varepsilon^\alpha(X|Y)$ in coding problems; generally speaking, they demonstrated (i) $\tilde{H}_\varepsilon^\alpha(X|Y) \triangleq \lim_{\varepsilon \to 0} \tilde{H}_\varepsilon^\alpha(X|Y)$ characterizes the minimum codeword length in the source coding problem of $X$ with the side-information $Y$ available at the decoder under the constraint that the probability of a decoding error is at most $\varepsilon$, and (ii) $\tilde{H}_\varepsilon^\alpha(X|Y) \triangleq \lim_{\varepsilon \to 0} \tilde{H}_\varepsilon^\alpha(X|Y)$ characterizes the amount of uniform randomness that can be extracted from $X$.

Observing the results of Arikan [2] along with those of Renner and Wolf [5], [6], it is natural to expect that we can use $\tilde{H}_\varepsilon^\alpha(X|Y)$ to characterize the moment of the number of guesses under a constraint on the probability of the error. However, the definition of $\tilde{H}_\varepsilon^\alpha(X|Y)$ is not appropriate to be used in the analysis of the guessing problem. In this paper, we introduce another “smoothed” version $H_\varepsilon^\alpha(X|Y)$ of the conditional Rényi entropy. Thereafter, by using $H_\varepsilon^\alpha(X|Y)$, we provide lower and upper bounds on the minimum cost of guessing the value of $X$ with the side-information $Y$ under the constraint that the guessing error probability is at most $\varepsilon$. Further we demonstrate another application of $H_\varepsilon^\alpha(X|Y)$ in the variable-length lossless coding problem of the source $X$ with the common side-information $Y$ available at the encoder and decoder. Our contributions are summarized as follows.

A. Contributions

First we introduce a novel definition of the conditional $\varepsilon$-smooth Rényi entropy $H_\varepsilon^\alpha(X|Y)$ of the order $\alpha > 0$, and then investigate its properties. In our definition, similar to that of Renner and Wolf, the minimization over the set of non-negative functions satisfying a particular condition is involved (our definition is given in (1)–(2), and that of Renner and Wolf is given in (8)–(9)). Our first contribution, Theorem 1, characterizes the non-negative function $Q$ attaining the minimum in the definition of $H_\varepsilon^\alpha(X|Y)$ for $\alpha \in (0, 1)$. This characterization is useful in the proof of our theorems in guessing. Further, by using the information spectrum method [7], we investigate the asymptotic behavior of $(1/n) H_\varepsilon^\alpha(X^n|Y^n)$ for sequences $X^n = (X_1, X_2, \ldots, X_n)$ and $Y^n = (Y_1, Y_2, \ldots, Y_n)$ of random variables as the length $n$ of sequences tends to infinity. Particularly, in Theorem 2, we show that the asymptotic value $H_\varepsilon^\alpha(X|Y)$ of the conditional smooth Rényi entropy for the mixture of i.i.d. sources is determined by the conditional entropy $H(X_i|Y_i)$ of a component of the mixture. This result allows us to present a single-letterized formula in guessing and source coding.
Next we investigate the problem of “guessing allowing error”, i.e., the problem of guessing $X$ with the side-information $Y$ where the guesser can, after each guess, stochastically choose (i) to give up and declare an error or (ii) to continue guessing. The cost of guessing is evaluated in the same way as done by Arikan [2]; the cost is $i^p$ for $\rho > 0$ if the value is correctly guessed at the $i$-th step. We consider the minimization of the expected value $\bar{C}_\rho$ of the guessing cost under the constraint that the error probability $p_e$ is at most $\varepsilon$. Our results, Theorems 3 and 4, provide lower and upper bounds on the minimum cost in terms of $H_{1/(1+\rho)}^e(X|Y)$. Further, a general formula for the exponent of the optimal guessing cost is derived; see Theorem 5. In particular, a single-letterized formula is obtained for the mixture of i.i.d. sources. Moreover, our result for the i.i.d. sources demonstrates that allowing a vanishing error (i.e., $p_e \to 0$ as $n \to \infty$) drastically changes the problem from Arikan’s original problem, where the zero-error constraint (i.e., $p_e = 0$ for all $n$) is imposed.

The last contribution of this paper is the illustration of the significance of our conditional smooth Rényi entropy $H_\alpha^e(X|Y)$ in the problem of source coding. We consider the variable-length lossless coding problem of the source $X$ with the common side-information $Y$ available at the encoder and decoder. We allow the decoder to make a decoding error with probability at most $\varepsilon$. Then, we evaluate the exponential moment $M_\rho$ of the codeword length (Refer (48) for the precise definition of $M_\rho$). In a similar manner as in the guessing problem, our results show that $H_\alpha^e(X|Y)$ can be used to characterize the minimum value of $M_\rho$: Theorems 6 and 7 provide lower and upper bounds on the minimum value of $M_\rho$ in terms of $H_{\rho/(\rho+1)}^e(X|Y)$, and then Theorem 8 presents a general formula for the exponent of the minimum value of $M_\rho$.

B. Related Work

As mentioned above, the concept of smooth Rényi entropy was first introduced by Renner and Wolf [5], [6]. Properties of the smooth Rényi entropy were investigated by Koga [9] using majorization theory. One of the results in [9] can be obtained as a corollary of our Theorem 1.

It is known that two special cases, $\alpha = 0$ and $\alpha = \infty$, of the smooth Rényi entropy have clear operational meaning in the fixed-length source coding [5], [6], [10] and intrinsic randomness problem [5], [6], [11] respectively. Similarly, the smooth Rényi divergence of order zero and order $\infty$ (see, e.g. Definitions 1 and 7 of [13]) finds applications in several coding problems as well [13], [14]. To the author’s best knowledge, this is the first paper that gives clear operational meaning of the conditional smooth Rényi entropy of order $\alpha \in (0, 1)$ in guessing and source coding.

1 By general formula, we imply that we consider the sequences of guessing problems and do not place any underlying structure such as stationarity, memorylessness, and ergodicity on the source [7], [8].

2 The smooth Rényi entropy of order zero (resp. order $\infty$) is known as the smooth max (resp. min) Rényi entropy as well. Properties of the max and min smooth Rényi entropies were also investigated in the quantum setting by Datta and Renner [12].

As mentioned above, Arikan [2] demonstrated the significance of Rényi entropy in the problem of guessing. Recently, tighter bounds on guessing moments were given by Sason and Verdú [15] where the Rényi entropy is used as well. The guessing problem has been studied in various contexts such as the problem of guessing subjected to distortion [16], [17], investigation of large deviation perspective of guessing [18], [19], and guesswork in multi-user systems [20], etc.; see, e.g., [15] and the references therein.

Campbell [21] proposed the exponential moment of the codeword length to serve as an alternative to the average codeword length as a criterion for variable-length lossless source coding, and provided upper and lower bounds on the exponential moment in terms of the Rényi entropy. The problem of variable-length source coding allowing errors was investigated under the criterion of the average codeword length by Koga and Yamamoto [23] and Kostina et al. [24], [25].

A generalization of Campbell’s result to the case where a decoding error is allowed was presented in [26]. Recently, a similar result without the prefix condition of codewords was given by Sason and Verdú [15]. Our results in Section IV can be seen as a generalization of [26], as the result of [26] is obtained by setting $|\mathcal{Y}| = 1$ in our results.

In this paper, two problems of guessing and source coding are investigated. The relation between the limiting guessing exponent and limiting exponent of the moment generating function of codeword lengths in source coding were pointed out by Arikan and Merhav [16]; refer [18]. Similar result was also obtained by Sason and Verdú [15, Lemma 7] without the prefix condition of codewords. More recently Beirami et al. [27] illustrated an interesting connection between guessing and data compression from the geometric perspective. As expected, Theorems 5 and 8 below reveal the equivalence between the optimal guessing exponent and optimal exponential moment of codeword lengths.

C. Paper Organization

The rest of the paper is organized as follows. In Section II, the conditional $\varepsilon$-smooth Rényi entropy $H_\alpha^e(X|Y)$ of order $\alpha$ is defined, and its properties are investigated. The problems of guessing and source coding are investigated in Section III and Section IV, respectively. Concluding remarks and directions for future work are provided in Section V. To ensure that the main ideas are seamlessly communicated in the main text, we relegate all proofs to the appendices.

II. CONDITIONAL SMOOTH RÉNYI ENTROPY

Let $\mathcal{X}$ and $\mathcal{Y}$ be finite or countably infinite sets. For $\varepsilon \in [0, 1)$ and a probability distribution $P_{XY}$ on $\mathcal{X} \times \mathcal{Y}$, let $\mathcal{B}(P_{XY})$ be the set of non-negative functions $Q$ with domain $\mathcal{X} \times \mathcal{Y}$ such that $Q(x, y) \leq P_{XY}(x, y)$, for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$, and $\sum_{x,y} Q(x, y) \geq 1 - \varepsilon$. Then, for $\alpha \in (0, 1) \cup (1, \infty)$, the conditional $\varepsilon$-smooth Rényi entropy of order $\alpha$ is defined
as

\[ H^\varepsilon_a(X|Y) \triangleq \frac{a}{1 - a} \log r^\varepsilon_a(X|Y) \] (1)

where

\[ r^\varepsilon_a(X|Y) \triangleq \inf_{Q \in \mathcal{B}^s(P_{XY})} \left[ \sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}} [Q(x, y)]^a \right]^{1/a} . \] (2)

In the following, we assume that \( 0 < a < 1 \). Hence, \( H^\varepsilon_a(X|Y) \) can be rewritten as

\[ H^\varepsilon_a(X|Y) = \inf_{Q \in \mathcal{B}^s(P_{XY})} \frac{a}{1 - a} \log \sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}} [Q(x, y)]^a \] \[ \quad = \inf_{Q \in \mathcal{B}^s(P_{XY})} \frac{a}{1 - a} \log \sum_{y \in \mathcal{Y}} P_y(y) \left[ \sum_{x \in \mathcal{X}} [Q(x, y) / P_y(y)]^a \right]^{1/a} . \] (3)

In the case of \( \varepsilon = 0 \), \( H^\varepsilon_a(X|Y) \) is equivalent to the conditional R\'enyi entropy of order \( a \):\(^6\)

\[ H^0_a(X|Y) = \frac{a}{1 - a} \log \sum_{y \in \mathcal{Y}} \left[ \sum_{x \in \mathcal{X}} [P_{XY}(x, y)]^a \right]^{1/a} \] (5)

In the case of \( \mathcal{Y} = 1 \), \( H^\varepsilon_a(X) \) is equivalent to the \( \varepsilon \)-smooth R\'enyi entropy of order \( a \), which is defined as

\[ H^\varepsilon_a(X) \triangleq \inf_{Q \in \mathcal{B}^s(P_X)} \frac{a}{1 - a} \log \sum_{x \in \mathcal{X}} [Q(x)]^a \] (6)

where \( \mathcal{B}^s(P_X) \) is defined in the same manner as \( \mathcal{B}^s(P_{XY}) \).

It should be emphasized that our definition (1) of \( H^\varepsilon_a(X|Y) \) is slightly different from that of Renner and Wolf [6]. In [6] the conditional smooth R\'enyi entropy is defined as

\[ \tilde{H}^\varepsilon_a(X|Y) \triangleq \frac{1}{1 - a} \log \tilde{r}^\varepsilon_a(X|Y) \] (8)

where

\[ \tilde{r}^\varepsilon_a(X|Y) \triangleq \inf_{Q \in \mathcal{B}^s(P_{XY})} \max_{P_y(y) > 0} \sum_{x \in \mathcal{X}} \left[ \frac{Q(x, y)}{P_y(y)} \right]^a . \] (9)

To note the difference, rewrite \( \tilde{r}^\varepsilon_a(X|Y) \) for \( 0 < a < 1 \) as

\[ \tilde{H}^\varepsilon_a(X|Y) \]

\[ = \inf_{Q \in \mathcal{B}^s(P_{XY})} \frac{1}{1 - a} \log \max_{P_y(y) > 0} \sum_{x \in \mathcal{X}} \left[ \frac{Q(x, y)}{P_y(y)} \right]^a \] (10)

\[ = \inf_{Q \in \mathcal{B}^s(P_{XY})} \frac{a}{1 - a} \log \max_{P_y(y) > 0} \sum_{x \in \mathcal{X}} \left[ \frac{Q(x, y)}{P_y(y)} \right]^a \] (11)

Comparing (11) with (4), we can observe that the average of \[ \left[ \sum_y [Q(x, y) / P_y(y)]^a \right]^{1/a} \] is taken in \( H^\varepsilon_a(X|Y) \), while the maximum is taken in \( \tilde{H}^\varepsilon_a(X|Y) \). Hence, it is apparent that

\[ H^\varepsilon_a(X|Y) \leq \tilde{H}^\varepsilon_a(X|Y) \] (12)

and equality does not hold in general.

**Remark 1.** The author assumes that the availability of the side-information \( Y \) causes the difference between \( H^\varepsilon_a(X|Y) \) and \( \tilde{H}^\varepsilon_a(X|Y) \). In the problem of guessing considered in Section III, the guesser can change the strategy according to the given \( y \in \mathcal{Y} \). Similarly, in the problem of source coding with common side-information considered in Section IV, the encoder can choose the encoding function according to the given \( y \in \mathcal{Y} \). Hence, in these problems, “the average with respect to \( Y \)” has significance in the coding theorems. On the other hand, in the problems considered in [6], the encoder (or the extractor) cannot access \( Y \) and has to prepare for the worst case. Hence, “the maximum with respect to \( Y \)” has significance in [6].

Now, we show several properties of \( H^\varepsilon_a(X|Y) \). First, we investigate \( Q \in \mathcal{B}^s(P_{XY}) \) attaining the inf in the definition of \( H^\varepsilon_a(X|Y) \). To do this, we introduce the following notation: For each \( y \in \mathcal{Y} \) and \( i = 1, 2, \ldots \), let \( x_i \) be the \( i \)-th probable \( x \in \mathcal{X} \) given \( y \); i.e., \( (x_i)_{i=1}^\infty \) is defined such that

\[ P_{XY}(x_i^y|y) \geq P_{XY}(x_{i+1}^y|y) \geq \cdots . \] (13)

Then, for each \( y \in \mathcal{Y} \) and a given \( \varepsilon_y \) satisfying \( 0 \leq \varepsilon_y < 1 \), let \( i_y^* = i_y^*(\varepsilon_y) \) be the minimum integer such that

\[ \sum_{i=1}^{i_y^*} P_{XY}(x_i^y|y) \geq 1 - \varepsilon_y \] (14)

and let

\[ Q_{i_y^*}(x_i^y|y) \]

\[ = \begin{cases} P_{XY}(x_i^y|y), & i = 1, 2, \ldots, i_y^* - 1, \\ 1 - \varepsilon_y - \sum_{i=1}^{i_y^*} P_{XY}(x_i^y|y), & i = i_y^*, \\ 0, & i > i_y^*. \end{cases} \] (15)

For \( \varepsilon_y = 1 \), let \( i_y^*(1) = \infty \) and \( Q_1^*(x|y) = P_{XY}(x|y) \) for all \( x, y \in \mathcal{X} \times \mathcal{Y} \).

**Theorem 1.** Using notations introduced above, we have

\[ H^\varepsilon_a(X|Y) \]

\[ = \inf_{(\varepsilon_y) \in \mathcal{E}_0(\varepsilon)} \frac{a}{1 - a} \log \sum_{y \in \mathcal{Y}} \left[ \sum_{i=1}^{i_y^*} [Q_{i_y^*}(x_i^y|y)]^a \right]^{1/a} \] (16)

where \( \mathcal{E}_0(\varepsilon) \) is the set of \( (\varepsilon_y)_{y \in \mathcal{Y}} \) satisfying \( 0 \leq \varepsilon_y \leq 1 \), for all \( y \in \mathcal{Y} \), and \( \sum_y \varepsilon_y P_y(y) = \varepsilon \).

Theorem 1 will be proved in Appendix A. As a corollary, we have a known property of \( H^\varepsilon_a(X) \), which is proved in (A) of Theorems 1 and 2 of [9].

\(^4\)Throughout this paper, \( \log \) denotes the natural logarithm.

\(^5\)As it is shown in Sections III and IV, the conditional \( \varepsilon \)-smooth R\'enyi entropy of order \( a = 1/(1 + \rho) \in (0, 1) \) plays an important role in guessing and source coding.

\(^6\)It was introduced by Arimoto [4].
Corollary 1. Assume that $x^1, x^2, \ldots$, are sorted such that $P_X(x^1) \geq P_X(x^2) \geq \ldots$ and let $i^*$ be the minimum integer such that $\sum_{i=1}^{i^*} P_X(x^i) \geq 1 - \varepsilon$. Then

$$Q^i_\alpha(x^i) \triangleq \begin{cases} P_X(x^i), & i = 1, 2, \ldots, i^* - 1, \\ 1 - \varepsilon - \sum_{i=1}^{i^*} P_X(x^i), & i = i^*, \\ 0, & i > i^* \end{cases}$$

attains the infimum in the definition (7) of $H^\alpha_n(X)$; i.e.,

$$H^\varepsilon_\alpha(X) = \frac{1}{1 - \alpha} \sum_{i=1}^{i^*} \left[ Q^i_\alpha(x^i) \right]^n. \quad (18)$$

Next, we investigate the asymptotic behavior of the conditional $\varepsilon$-smooth Rényi entropy by using the information spectrum method [7]. Let us consider a pair of correlated general sources $(X, Y) = \{(X^n, Y^n) \}_{n=1}^{\infty}$, which is a sequence of pairs $(X^n, Y^n)$ of correlated random variables $X^n$ on the $n$-fold Cartesian product $\mathcal{X}^n$ of $\mathcal{X}$ and $\mathcal{Y}^n$ on $\mathcal{Y}^n$. The joint distribution of $(X^n, Y^n)$ is denoted by $P_{X^nY^n}$, which is not required to satisfy the consistency condition.\footnote{The consistency condition requires that, for all $x^n \in \mathcal{X}^n$ and $y^n \in \mathcal{Y}^n$, $P_{X^nY^n}(x^n, y^n) = \sum_{\hat{x}^n} P_{X^{n+1}Y^{n+1}}(x^n \circ \hat{x}, y^n \circ \hat{y})$, where $\circ$ means the concatenation.}

Given $(X, Y)$, $\alpha \in (0, 1)$, and $\varepsilon \in [0, 1]$, let us define $H^\varepsilon_\alpha(X|Y)$ as

$$H^\varepsilon_\alpha(X|Y) \triangleq \lim_{\delta \to 0} \lim_{n \to \infty} \frac{1}{n} H^\varepsilon_{\alpha + \delta}(X^n|Y^n). \quad (19)$$

As shown in the following sections, this quantity plays an important role in the general formulas of guessing and source coding.

Here it is worth to note that $H^\varepsilon_\alpha(X|Y)$ is non-negative for all $\alpha \in (0, 1)$ and $\varepsilon \in [0, 1]$. Indeed, we can prove a stronger fact that

$$\lim_{n \to \infty} \frac{1}{n} H^\varepsilon_\alpha(X^n|Y^n) \geq 0, \quad \alpha \in (0, 1), \varepsilon \in [0, 1]. \quad (20)$$

We will prove (20) in Appendix B.

To give a single-letterized form of $H^\varepsilon_\alpha(X|Y)$, we consider a mixture of i.i.d. sources. Let us consider $m$ distributions $P_{X_iY_i}$ $(i = 1, 2, \ldots, m)$ on $\mathcal{X} \times \mathcal{Y}$. A general source $(X, Y)$ is said to be a mixture of $P_{X_1Y_1}, P_{X_2Y_2}, \ldots, P_{X_mY_m}$ if there exists $(a_1, a_2, \ldots, a_m)$ satisfying (i) $\sum_{i=1}^{m} a_i = 1$, (ii) $a_i > 0$ $(i = 1, \ldots, m)$, and (ii) for all $n = 1, 2, \ldots, x^n = (x_1, x_2, \ldots, x_n) \in \mathcal{X}^n$ and $y^n = (y_1, y_2, \ldots, y_n) \in \mathcal{Y}^n$,

$$P_{X^nY^n}(x^n, y^n) = \sum_{i=1}^{m} a_i P_{X^n_iY^n_i}(x^n, y^n) \quad (21)$$

$$= \sum_{i=1}^{m} a_i \prod_{k=1}^{n} P_{X_iY_i}(x_i, y_i). \quad (22)$$

For later use, let

$$A_i \triangleq \sum_{j=1}^{i-1} a_j, \quad i = 1, 2, \ldots, m, \quad (23)$$

and $A_{m+1} \triangleq 1$. Further, to simplify the analysis, we assume that

$$H(X_1|Y_1) > H(X_2|Y_2) > \cdots > H(X_m|Y_m) \quad (24)$$

where $H(X_i|Y_i)$ is the conditional entropy determined by $P_{X_iY_i}$ [28]. Then, $H^\varepsilon_\alpha(X|Y)$ of the mixture $(X, Y)$ is characterized in the following theorem.

Theorem 2. Let $(X, Y)$ be a mixture of i.i.d. sources satisfying (24). Fix $\alpha \in (0, 1)$, $i$, and $\varepsilon \in [A_i, A_{i+1})$, where $A_i$ is defined in (23). Then, we have

$$H^\varepsilon_\alpha(X|Y) = H(X_i|Y_i). \quad (25)$$

Theorem 2 will be proved in Appendix C.

Remark 2. Although Theorem 2 assumes that the components are i.i.d. sources, this assumption is not crucial. Indeed, the property of i.i.d. sources used in the proof of the theorem is only that the AEP [28] holds, i.e.,

$$\lim_{n \to \infty} \Pr \left[ \left| \frac{1}{n} \log \frac{1}{P^\varepsilon_{X^n}(X^n|Y^n)} - H(X_i|Y_i) \right| > \zeta \right] = 0 \quad (26)$$

for all $i = 1, 2, \ldots, m$ and any $\zeta > 0$. Hence, it is straightforward to extend the theorem such that it can be applied for the mixture of stationary and ergodic sources.

III. GUESSING

In this section, we assume that the alphabet $\mathcal{X}$ is finite; we assume that $|\mathcal{X}| = K$ and $\mathcal{Y} = \{1, 2, \ldots, K\}$.

A guessing strategy $G = ((\sigma_y, \pi_y))_{y \in \mathcal{Y}}$ for $X$ given $Y$ is defined by a collection of pairs $(\sigma_y, \pi_y)$, for each $y$, of (i) a permutation $\sigma_y$ on $\mathcal{X}$ and (ii) a vector $\pi_y = (\pi_y(i))_{i=1}^{K}$ satisfying $0 \leq \pi_y(i) \leq 1$ for all $i = 1, 2, \ldots, K$. Given the side information $y \in \mathcal{Y}$, the “guesser” corresponding to the strategy $G$ guesses the value of $X$ in the following manner. At the $i$th step ($i = 1, 2, \ldots, K$), the guesser determines whether to “give up” or not; the guesser gives up and stops guessing with probability $\pi_y(i)$ and a guessing error is declared. If the guesser does not give up then the guesser asks “Is $X = \sigma_y^{-1}(i)$”? The guessing continues until the guesser gives up or when the value of $X$ is correctly guessed (i.e., $\sigma_y^{-1}(i) = X$ at the $i$th step). It should be noted here that the guessing function studied in [2] corresponds to the guessing strategy that $\pi_y(i) = 0$ for all $y \in \mathcal{Y}$ and $i = 1, 2, \ldots, K$.

In this paper, we evaluate the “cost” of guessing as follows. If the guessing is stopped before the value of $X$ is correctly guessed then a constant cost $c_y \geq 0$ is incurred as “penalty”. Otherwise, the cost of guessing is given by $i^p$ when the value of $X$ is correctly guessed at the $i$th step, where $\rho \geq 0$ is a constant. For each $y \in \mathcal{Y}$, let

$$\lambda_y(i) \triangleq \prod_{j=1}^{i} (1 - \pi_y(j)), \quad i = 1, 2, \ldots, K. \quad (27)$$

Then we note that, given $y \in \mathcal{Y}$, the conditional probability of the event “the value of $X$ is correctly guessed at the $i$-th step before giving up” is

$$\lambda_y(i) P_X|Y(\sigma_y^{-1}(i)|y) \quad (28)$$
and thus, the conditional probability of the event “the guesser gives up before guessing the value of X correctly” is
\[
1 - \sum_{i=1}^{K} \lambda_y(i) P_{X|Y}(\sigma_y^{-1}(i)|y) = 1 - \sum_{x \in \mathcal{X}} \lambda_y(\sigma_y(x)) P_{X|Y}(x|y). \tag{29}
\]
Hence, the error probability \( p_e = p_e(G|X, Y) \), i.e., the average probability such that the guessing is stopped before the value of X is correctly guessed, is given by
\[
p_e = \sum_{y \in \mathcal{Y}} P_Y(y) \left( 1 - \sum_{x \in \mathcal{X}} \lambda_y(\sigma_y(i)) P_{X|Y}(x|y) \right)
= 1 - \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} \lambda_y(\sigma_y(x)) P_{X|Y}(x,y), \tag{30}
\]
and the expected value \( \tilde{C}_\rho' = \tilde{C}_\rho'(G|X, Y) \) of the cost is given by
\[
\tilde{C}_\rho' = \sum_{y \in \mathcal{Y}} P_Y(y) \left[ \sum_{i=1}^{K} \lambda_y(i) P_{X|Y}(\sigma_y^{-1}(i)|y)i^\rho \right] + p_e c_e. \tag{32}
\]
For some applications, it may be natural to simply minimize the cost \( \tilde{C}_\rho' \). In general, however, it may not be easy to determine the precise value of the penalty for stopping the guessing.\(^8\) In such a situation, we may consider the cost of guessing and penalty separately, and minimize
\[
\tilde{C}_\rho(G|X, Y)
= \sum_{y \in \mathcal{Y}} P_Y(y) \left[ \sum_{i=1}^{K} \lambda_y(i) P_{X|Y}(\sigma_y^{-1}(i)|y)i^\rho \right]
= \sum_{y \in \mathcal{Y}} P_Y(y) \left[ \sum_{x \in \mathcal{X}} \lambda_y(\sigma_y(x)) P_{X|Y}(x|y)\sigma_y(x) \right]
= \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} \lambda_y(\sigma_y(x)) P_{X|Y}(x,y)\sigma_y(x)^\rho \tag{33}
\]
under the constraint on the probability \( p_e \) of stopping the guessing. Further, if the minimum value \( \tilde{C}_\rho^*(\epsilon) \) of \( \tilde{C}_\rho \) under the constraint that \( p_e \leq \epsilon \) is known, it is not hard to optimize \( \tilde{C}_\rho' \); the optimal value can be written as \( \inf_{\epsilon} [\tilde{C}_\rho^*(\epsilon) + \epsilon c_e] \). So, we study the minimizing problem of \( \tilde{C}_\rho \) under the constraint that \( p_e \leq \epsilon \) for a given constant \( 0 \leq \epsilon \leq 1 \); the results are summarized in the following theorems:

**Theorem 3.** Fix \( \rho > 0 \) and \( \epsilon \in [0, 1] \). For any guessing strategy \( G \) satisfying \( p_e(G|X, Y) \leq \epsilon \), the expected value \( \tilde{C}_\rho = \tilde{C}_\rho(G|X, Y) \) of the cost must satisfy
\[
\tilde{C}_\rho \geq (1 + K)^{-\rho} \exp \left\{ \rho H^c_1(X|Y) \right\}. \tag{36}
\]

**Theorem 4.** Fix \( \rho > 0 \) and \( \epsilon \in [0, 1] \). There exists a guessing strategy \( G \) such that the error probability satisfies \( p_e(G|X, Y) \leq \epsilon \) and the expected value \( \tilde{C}_\rho = \tilde{C}_\rho(G|X, Y) \) of the cost satisfies
\[
\tilde{C}_\rho \leq \exp \left\{ \frac{\rho H^c_1(X|Y)}{1-\rho} \right\}. \tag{37}
\]

**Remark 3.** The latter part of Theorem 4 indicates the fact that “the optimal guesser throws a dice at most once.” In other words, the optimal strategy guesses in decreasing order of the conditional probability \( P_{X|Y} \) until the weight of the remaining guesses is at most \( \epsilon \). Randomness is used to guarantee that the probability of error is exactly \( \epsilon \).

Theorems 3 and 4 will be proved in Appendix D. Now, let us consider the asymptotic behavior of the cost of guessing. Particularly we investigate the asymptotic behavior of the exponent of the cost. Thus, we define an achievable guessing exponent as follows.

**Definition 1.** Given a constant \( \rho > 0 \) and a general source \((X, Y)\), a guessing exponent \( E_g \) is said to be \( \epsilon \)-achievable if there exists a sequence \( \{G_n\}_{n=1}^{\infty} \) of strategies satisfying
\[
\limsup_{n \to \infty} p_e(G_n|X^n, Y^n) \leq \epsilon \tag{39}
\]
and
\[
\limsup_{n \to \infty} \frac{1}{n} \log \tilde{C}_\rho(G_n|X^n, Y^n) \leq E_g. \tag{40}
\]

The infimum of \( \epsilon \)-achievable guessing exponents is denoted by \( E_g(\rho, \epsilon|X, Y) \).

Then we have the following theorem, where \( H^c_1(X|Y) \) defined in \( (19) \) characterizes the optimal guessing exponent.

**Theorem 5.** For any \( \rho > 0 \) and \( \epsilon \in [0, 1] \),
\[
E_g(\rho, \epsilon|X, Y) = \rho H^c_1(X|Y). \tag{41}
\]

The theorem will be proved in Appendix D-C. Combining Theorem 5 with Theorem 2, we can obtain the single-letterized characterization of \( E_g(\rho, \epsilon|X, Y) \) for a mixed source \((X, Y)\).

**Corollary 2.** Let \((X, Y)\) be a mixture of i.i.d. sources satisfying \( (24) \). Assume that \( \epsilon \) satisfies \( \epsilon \in [A_i, A_{i+1}) \), where \( A_i \) is defined in \( (23) \). Then, for any \( \rho > 0 \),
\[
E_g(\rho, \epsilon|X, Y) = \rho H(X_i|Y). \tag{42}
\]

In particular, let us consider a special case of guessing for an i.i.d. source (i.e., \( m = 1 \)) under the constraint \( \epsilon = 0 \). Corollary 2 shows
\[
E_g(\rho, 0|X, Y) = \rho H(X|Y). \tag{43}
\]
In other words, $E_2(\rho, 0|X, Y)$ is determined by the parameter $\rho$ and the conditional entropy $H(X|Y)$ of the source. On the other hand, Proposition 5 of [2] shows that the exponent of the optimal guessing for an i.i.d. source is $\rho H_{1/(1+\rho)}(X|Y)$. It may seem to be a contradiction, however it is not. The constraint $\varepsilon = 0$ in our problem requires that $p_e \to 0$ as $n \to \infty$; i.e., a vanishing error is allowed. On the other hand, Arikian’s original guessing problem [2] imposes the zero-error constraint; i.e., it is required that $p_e = 0$ for all $n$. Our result shows that allowing a vanishing error changes the problem drastically.

**Remark 4.** It is well known that, in the channel coding problem, the zero-error capacity [29] is quite different from the conventional capacity. The above argument demonstrates that an analogous observation holds true in the guessing problem as well.

### IV. Source Coding

A variable-length source code $\Phi = (\varphi, \psi, C)$ for $X$ with the common-side-information $Y$ is determined by a collection of triplets $(\varphi, \psi, C)$, for each $y \in \mathcal{Y}$, of (i) a set $\mathcal{C}_y \subset \{0, 1\}^*$ of finite-length binary strings, (ii) a stochastic encoder mapping $\varphi_y : \mathcal{X} \to \mathcal{C}_y$, and (iii) a decoder mapping $\psi_y : \mathcal{C}_y \to \mathcal{X}$. Without loss of generality, we assume that $\mathcal{C}_y = \{\varphi_y(x) : x \in \mathcal{X}\}$. Further, we assume that $\mathcal{C}_y$ is prefix-free for each $y \in \mathcal{Y}$.

Note that we allow the encoder mapping $\varphi_y$ to be stochastic. Let $W_{\varphi_y}(c|x)$ be the probability that $x \in \mathcal{X}$ is encoded to $c \in \mathcal{C}_y$ by $\varphi_y$. Then, the error probability $p_e = p_e(\Phi|X, Y)$ of the code $\Phi$ is defined as

$$p_e \triangleq \sum_{y \in \mathcal{Y}} P_Y(y) \Pr\{X \neq \psi_y(\varphi_y(X))\}$$

$$= \sum_{y \in \mathcal{Y}} P_Y(y) \left( \sum_{x \in \mathcal{X}} P_{X|Y}(y|x) \sum_{c \in \mathcal{C}_y \setminus \varphi_y(c)} W_{\varphi_y}(c|x) \right)$$

$$= \sum_{(x, y) \in \mathcal{X} \times \mathcal{Y}} P_{XY}(x, y) \sum_{c \in \mathcal{C}_y \setminus \varphi_y(c)} W_{\varphi_y}(c|x).$$

The length of the codeword $\varphi_y(x)$ of $x$ (in bits) is denoted by $L(\varphi_y(x))$. Let $\ell(\cdot|y)$ be the length function (in nats):

$$\ell(x|y) \triangleq L(\varphi_y(x)) \log 2.$$ 

In this study, we focus on the exponential moment of the length function. For a given $\rho > 0$, let us define an exponential moment

$$M_\rho = \mathbb{E}_{P_{XY}} \left[ \exp(\rho \ell(X|Y)) \right]$$

$$= \sum_{y \in \mathcal{Y}} P_Y(y) \left( \sum_{x \in \mathcal{X}} P_{X|Y}(y|x) \right)$$

$$\times \sum_{c \in \mathcal{C}_y} W_{\varphi_y}(c|x) \exp(\rho L(c) \log 2)$$

$$= \sum_{(x, y) \in \mathcal{X} \times \mathcal{Y}} P_{XY}(x, y) \sum_{c \in \mathcal{C}_y} W_{\varphi_y}(c|x) \exp(\rho L(c) \log 2),$$

where $\mathbb{E}_P$ denotes the expectation with respect to the distribution $P$.

**Remark 5.** Without loss of optimality we can assume that the decoder mapping $\psi_y$ is deterministic for all $y \in \mathcal{Y}$. Indeed, for a given $W_{\psi_y}$, we can choose $\psi_y$ such that

$$\psi_y(c) = \arg \max_{x \in \mathcal{X}} W_{\psi_y}(c|x) P_{X|Y}(x|y).$$

We consider the problem of minimizing $M_\rho$ under the constraint that $p_e \leq \varepsilon$ for a given constant $0 \leq \varepsilon \leq 1$; the results are summarized in the following theorems:

**Theorem 6.** Fix $\rho > 0$ and $\varepsilon \in [0, 1)$. For any code $\Phi$ satisfying $p_e(\Phi|X, Y) \leq \varepsilon$, the moment $M_\rho = M_\rho(\Phi|X, Y)$ must satisfy

$$M_\rho \geq \exp \left( \frac{\rho H_{1/(1+\rho)}(X|Y)}{1+\rho} \right).$$

**Theorem 7.** Fix $\rho > 0$ and $\varepsilon \in [0, 1)$. There exists a code $\Phi$ such that $p_e(\Phi|X, Y) \leq \varepsilon$ and $M_\rho = M_\rho(\Phi|X, Y)$ satisfies

$$M_\rho \leq 2^{\rho \varepsilon} \exp \left( \frac{\rho H_{1/(1+\rho)}(X|Y)}{1+\rho} \right) + 2^\rho.$$ 

Theorems 6 and 7 will be proved in Appendix E.

**Remark 6.** While we allow the encoder mapping $\varphi$ to be stochastic in Theorem 7, we can see the fact that “the optimal encoder throws a dice at most once”; cf. Remark 3. In other words, for each $y \in \mathcal{Y}$, there exists at most one symbol $x_y \in \mathcal{X}$ such that $x_y$ is encoded stochastically and other symbols $x \neq x_y$ are encoded deterministically (for $x \neq x_y$, there exists $c \in \mathcal{C}_y$ satisfying $W_{\varphi_y}(c|x) = 1$). Hence, it is not hard to modify the theorem for the case where only deterministic encoder mappings are allowed. We ignore the details but see Proposition 1 of [26] for the case of $|\mathcal{Y}| = 1$.

Now, let us consider the asymptotic behavior of the exponential moment of the length function. As in the guessing problem, we define an achievable codeword-length exponent as follows.

**Definition 2.** Given a constant $\rho > 0$ and a general source $(X, Y)$, a codeword-length exponent $E_\varepsilon$ is said to be $\varepsilon$-achievable if there exists a sequence $\{\Phi_n\}_{n=1}^\infty$ of variable-length codes satisfying

$$\limsup_{n \to \infty} p_e(\Phi_n|X^n, Y^n) \leq \varepsilon$$

and

$$\limsup_{n \to \infty} \frac{1}{n} \log M_\rho(\Phi_n|X^n, Y^n) \leq E_\varepsilon.$$ 

The infimum of $\varepsilon$-achievable codeword-length exponents is denoted by $E_\varepsilon(\rho, \varepsilon|X, Y)$.

Then we have the following general formula, which will be proved in Appendix E-C.

**Theorem 8.** For any $\rho > 0$ and $\varepsilon \in [0, 1)$,

$$E_\varepsilon(\rho, \varepsilon|X, Y) = \rho H_{1/(1+\rho)}(X|Y).$$
Combining Theorem 8 with Theorem 2, we can obtain the single-letterized characterization of \( E_s(\rho, \epsilon | X, Y) \) for a mixed source \((X, Y)\).

**Corollary 3.** Let \((X, Y)\) be a mixture of i.i.d. sources satisfying (24). Assume that \( \epsilon \) satisfies \( \epsilon \in [A_i, A_{i+1}) \), where \( A_i \) is defined in (23). Then, for any \( \rho > 0 \),

\[
E_s(\rho, \epsilon | X, Y) = \rho H(X_i | Y_i) .
\]

(57)

\[ \]

**V. CONCLUDING REMARKS**

In this paper, a novel definition of the conditional smooth Rényi entropy was introduced, and its significance in the problems of guessing and source coding was demonstrated.

Although properties of \( H_\alpha^c(X|Y) \) and \( H_\alpha^c(\cdot|X) \) are investigated in Section II, we consider only the case of \( \alpha \in (0, 1) \). An important future work is to investigate the properties of \( H_\alpha^c(X|Y) \) and \( H_\alpha^c(\cdot|X) \) of order \( \alpha > 1 \). On the other hand, in the coding theorems in Sections III and IV, it is sufficient to consider the conditional smooth Rényi entropy of order \( \alpha = 1/(1 + \rho) \in (0, 1) \). Another important future work is to determine the operational meaning of \( H_\alpha^c(X|Y) \) of order \( \alpha > 1 \).

In Section IV, we assume that the common side-information \( Y \) is available at the encoder and decoder. As mentioned in Remark 1, the availability of \( Y \) at the encoder is very important. The author conjectures that, in case \( Y \) is only available at the decoder, \( H_\alpha^c( \cdot | X ) \) instead of \( H_\alpha^c(X|Y) \) characterizes the exponential moment of codeword lengths. We leave it as a future research work.

**APPENDIX A**

**PROOF OF THEOREM 1**

Before proving the theorem, we provide several lemmata in Subsection A-A. The proof of the theorem is given in Subsection A-B.

**A. Lemmata**

**Lemma 1.** Let \( k \) be a positive integer. Assume that \( p = (p_i)_{i=1}^k \) and \( q = (q_i)_{i=1}^k \) satisfy \( p \leq 0, q_i \geq 0 \) \( (i = 1, 2, \ldots, k) \),

\[
\sum_{i=1}^j p_i \leq \sum_{i=1}^j q_i, \quad j = 1, 2, \ldots, k-1,
\]

and

\[
\sum_{i=1}^k p_i = \sum_{i=1}^k q_i.
\]

(58)

(59)

Then, for all \( \alpha \in (0, 1) \), we have

\[
\sum_{i=1}^k p_i^\alpha \geq \sum_{i=1}^k q_i^\alpha.
\]

(60)

The lemma is a consequence of the Schur concavity of the function \( f(p) = \sum_{i=1}^k p_i^\alpha \) for \( \alpha \in (0, 1) \); \( f \) is Schur concave and the assumption of the lemma means \( p \prec q \) (i.e., \( q \) majorizes \( p \)), hence we have \( f(p) \geq f(q) \). Refer to [30] for more details. It should be noted that Lemma 1 is essentially the same as (A) of Theorem 1 of [9].

By modifying Lemma 1, we have the following key lemma, which is essentially the same as (A) of Theorem 2 of [9]. However, we provide proof for comprehensiveness as only a proof sketch is given in [9].

**Lemma 2.** Fix \( \alpha \in (0, 1) \), \( \epsilon \in (0, 1) \) and \( p = (p_i)_{i=1}^\infty \) satisfying (i) \( p_i \geq 0 \) for all \( i = 1, 2, \ldots \), (ii) \( p_1 \geq p_2 \geq p_3 \geq \ldots \), and (iii) \( \sum_{i=1}^\infty p_i = 1 \). Let \( i^* \) be the least integer such that \( \sum_{i=i^*}^\infty p_i \geq 1 - \epsilon \) and let

\[
q_i^* = \begin{cases} 
 1 - \epsilon - \sum_{i=1}^{i^*-1} p_i, & i = i^*, \\
 0, & i > i^*.
\end{cases}
\]

(61)

Then, we have

\[
\sum_{i=1}^\infty (q_i^*)^\alpha = \inf_{q \in B^c(\epsilon)} \sum_{i=1}^\infty q_i^\alpha
\]

where \( B^c(\epsilon) \) is the set of \( q = (q_i)_{i=1}^\infty \) satisfying \( 0 \leq q_i \leq p_i \) \((i = 1, 2, \ldots)\) and \( \sum_{i=1}^\infty q_i \geq 1 - \epsilon \).

**Proof:** Let

\[
B_0^c(\epsilon) \triangleq \{ q \in B^c(\epsilon) : \sum_{i=1}^\infty q_i = 1 - \epsilon \}
\]

and

\[
B^c_{0, \text{finite}}(\epsilon) \triangleq \{ q \in B_0^c(\epsilon) : \exists k \text{ s.t. } q_i = 0 \text{ for all } i > k \}.
\]

(62)

(63)

(64)

For any \( q \in B^c(\epsilon) \), there exists \( \hat{q} \in B^c_{0, \text{finite}}(\epsilon) \) satisfying \( 0 \leq \hat{q}_i < q_i \) for all \( i = 1, 2, \ldots \). Since \( \sum_{i=1}^\infty \hat{q}_i^\alpha \leq \sum_{i=1}^\infty q_i^\alpha \), we have

\[
\inf_{q \in B^c_{0, \text{finite}}(\epsilon)} \sum_{i=1}^\infty q_i^\alpha \leq \inf_{q \in B^c(\epsilon)} \sum_{i=1}^\infty q_i^\alpha.
\]

(65)

On the other hand, Lemma 1 guarantees that

\[
\sum_{i=1}^\infty (q_i^*)^\alpha = \inf_{q \in B^c_{0, \text{finite}}(\epsilon)} \sum_{i=1}^\infty q_i^\alpha.
\]

(66)

Thus, to prove the lemma, it is sufficient to prove the following fact: for any \( q \) satisfying \( q \in B^c_{0, \text{finite}}(\epsilon) \) and \( q \notin B^c(\epsilon) \), there exists \( \hat{q} \in B^c_{0, \text{finite}}(\epsilon) \) such that \( \sum_{i=1}^\infty \hat{q}_i^\alpha \leq \sum_{i=1}^\infty q_i^\alpha \).

Indeed, the fact can be proved as follows. Since \( q \in B^c_{0, \text{finite}}(\epsilon) \) and \( q \notin B^c(\epsilon) \), we can choose finite integers \( j \) and \( k \) such that \( q_j < p_j, k > j, \) and \( \sum_{i=k+1}^\infty q_i < p_j - q_j \). Then, choose \( \hat{q} \) so that

\[
\hat{q}_i = \begin{cases} 
 q_i, & i = 1, 2, \ldots, j - 1, j + 1, \ldots, k, \\
 q_j + \sum_{i=j+1}^\infty q_i, & i = j, \\
 0, & i > k.
\end{cases}
\]

(67)

As \( i^* \geq t \) for \( 0 \leq t \leq 1 \), we have

\[
\frac{(q_i)^\alpha}{(q_j)^\alpha} + \sum_{i'=k+1}^\infty \frac{(q_{i'})^\alpha}{(q_j)^\alpha} \geq \frac{q_j}{q_j} + \sum_{i'=k+1}^\infty \frac{q_{i'}}{q_j} = 1
\]

(68)
and thus
\[ q_j^a + \sum_{i=k+1}^{\infty} q_i^a \geq \hat{q}_j^a. \] (69)

Hence, we have \[ \sum_{i=1}^{\infty} \hat{q}_i^a \leq \sum_{i=1}^{\infty} q_i^a. \]

\[ \square \]

**B. Proof of Theorem 1**

Given \( Q(x, y) \in \mathcal{B}^s(P_{XY}) \), we note that \( \gamma_{xy} \triangleq Q(x, y)/P_{XY}(x, y) \) satisfies (i) \( 0 \leq \gamma_{xy} \leq 1 \) for each \( (x, y) \in \mathcal{X} \times \mathcal{Y} \) and (ii) \( \sum_{x, y} \gamma_{xy} P_{XY}(x, y) \geq 1 - \varepsilon \). Hence, we can rewrite (4) as

\[
H^s_a(X|Y) = \inf_{\gamma_{xy} \in \mathcal{B}^s(P_{XY})} \left\{ \sum_{x \in \mathcal{X}} \left[ \sum_{y \in \mathcal{Y}} \frac{\gamma_{xy} P_{XY}(x, y)}{P_Y(y)} \right]^a \right\}^{1/a}
\]

where \( \inf \) is taken over all \( (\gamma_{xy})_{(x,y) \in \mathcal{X} \times \mathcal{Y}} \) such that (i) \( 0 \leq \gamma_{xy} \leq 1 \) for each \( (x, y) \in \mathcal{X} \times \mathcal{Y} \) and (ii) \( \sum_{x, y} \gamma_{xy} P_{XY}(x, y) \geq 1 - \varepsilon \).

Now, suppose that \( (\gamma_{xy})_{(x,y) \in \mathcal{X} \times \mathcal{Y}} \) satisfies (i) and (ii) above. Then, as \( \sum_{x, y} \gamma_{xy} P_{XY}(x, y) \leq \sum_{x, y} P_{XY}(x, y) = P_Y(y) \), we can see that \( \gamma_{xy} \triangleq \sum_{y \in \mathcal{Y}} \gamma_{xy} P_{XY}(x, y)/P_Y(y) \) satisfies (i) \( 0 \leq \gamma_{xy} \leq 1 \) for all \( y \in \mathcal{Y} \) and (ii) \( \sum_{x, y} \gamma_{xy} P_{XY}(x, y) \geq 1 - \varepsilon \). In other words, \( \varepsilon_y \triangleq 1 - \gamma_y \) and \( Q_{X|Y}(x|y) \triangleq \gamma_{xy} P_{XY}(x, y) \) satisfy (i) \( 0 \leq \varepsilon_y \leq 1 \), (ii) \( \sum_{y \in \mathcal{Y}} \varepsilon_y P_Y(y) \leq \varepsilon \), (iii) \( 0 \leq Q_{X|Y}(x|y) \leq P_{X|Y}(x|y) \) and (iv) \( \sum_{x \in \mathcal{X}} Q_{X|Y}(x|y) = 1 - \varepsilon_y \). On the other hand, given \( \varepsilon_y \in \mathcal{Y} \) and \( Q_{X|Y} \) satisfying (i–iv), we can determine the corresponding \( (\gamma_{xy})_{(x,y) \in \mathcal{X} \times \mathcal{Y}} \). This observation brings another representation of (71):

\[
H^s_a(X|Y) = \inf_{(\varepsilon_y) \in \mathcal{E}(\varepsilon)} \left\{ \sum_{y \in \mathcal{Y}} \frac{\alpha}{1 - \alpha} \log \sum_{x \in \mathcal{X}} P_{XY}(x, y) \right\}^{1/a}
\]

where (i) \( \mathcal{E}(\varepsilon) \) is the set of \( (\varepsilon_y)_{y \in \mathcal{Y}} \) satisfying \( 0 \leq \varepsilon_y \leq 1 \) and \( \sum_{y \in \mathcal{Y}} \varepsilon_y P_Y(y) \leq \varepsilon \), and (ii) \( \mathcal{B}^s_0(P_{X|Y}) \) is the set of non-negative functions \( Q_{X|Y} \) on \( \mathcal{X} \) such that \( \tilde{Q}_{X|Y}(x|y) \leq P_{X|Y}(x|y) \) and \( \sum_{x \in \mathcal{X}} \tilde{Q}_{X|Y}(x|y) = 1 - \varepsilon_y \). From Lemma 2, we can see that, for all \( y \in \mathcal{Y} \),

\[
\sum_{i=1}^{\mathcal{I}(\varepsilon_y)} \left[ \tilde{Q}^s_{\varepsilon_y}(x_i^a|y) \right]^a \geq \inf_{Q_{X|Y} \in \mathcal{B}^s_0(P_{X|Y})} \left[ \sum_{x \in \mathcal{X}} \left[ Q_{X|Y}(x|y) \right]^a \right] \]

where \( \mathcal{I}(\varepsilon_y) \) and \( \tilde{Q}^s_{\varepsilon_y} \) are defined as in (14) and (15). Further, for any \( (\varepsilon_y)_{y \in \mathcal{Y}} \in \mathcal{E}(\varepsilon) \), there exists \( (\tilde{\varepsilon}_y)_{y \in \mathcal{Y}} \in \mathcal{E}(\varepsilon) \) such that \( \varepsilon_y \leq \tilde{\varepsilon}_y \) and thus

\[
\sum_{i=1}^{\mathcal{I}(\varepsilon_y)} \left[ \tilde{Q}^s_{\varepsilon_y}(x_i^a|y) \right]^a \geq \sum_{i=1}^{\mathcal{I}(\varepsilon_y)} \left[ \tilde{Q}^s_{\tilde{\varepsilon}_y}(x_i^a|y) \right]^a, \quad y \in \mathcal{Y}. \] (74)

Hence, we have (16). \( \square \)

**APPENDIX B**

**Proof of (20)**

From (4), we have

\[
H^s_a(X^n|Y^n) = \inf_{Q_{X|Y} \in \mathcal{B}^s_0(P_{X|Y})} \left\{ \frac{1}{n} \log \sum_{y \in \mathcal{Y}} \left[ \sum_{x \in \mathcal{X}} Q_{X^n|Y^n}(x^n|y^n) \right]^a \right\}^{1/a}. \]

As

\[
\frac{Q_{X^n}(x^n, y^n)}{P_{Y^n}(y^n)} = \frac{P_{X^n|Y^n}(x^n, y^n)}{P_{Y^n}(y^n)} \leq P_{X^n|Y^n}(x^n|y^n) \leq 1, \]

we have \( \frac{Q_{X^n}(x^n, y^n)}{P_{Y^n}(y^n)} \leq \left[ \frac{Q_{X^n}(x^n, y^n)}{P_{Y^n}(y^n)} \right]^a \) for \( a \in (0, 1) \), and thus,

\[
H^s_a(X^n|Y^n) \leq \inf_{Q_{X|Y} \in \mathcal{B}^s_0(P_{X|Y})} \left\{ \frac{1}{n} \log \sum_{y \in \mathcal{Y}} \left[ \sum_{x \in \mathcal{X}} Q_{X^n|Y^n}(x^n|y^n) \right]^a \right\}^{1/a}. \]

where (a) follows from Jensen’s inequality. Hence,

\[
\frac{1}{n} H^s_a(X^n|Y^n) \geq \frac{1}{n} \log(1 - \varepsilon). \]

Taking the inferior limit of both sides, we have (20). \( \square \)

**APPENDIX C**

**Proof of Theorem 2**

Before proving the theorem, we provide several lemmata in Subsection C-A. The proof of the theorem is given in Subsection C-B.
Lemma 3. Fix $\zeta > 0$. Then, there exists an integer $n_0$ such that for all $n \geq n_0$ and all $i = 1, 2, \ldots, m$,

$$\Pr \left\{ \frac{1}{n} \log \frac{1}{P_{X^n|Y^n}(X^n|Y^n)} \geq H(X_i|Y_i) - \zeta \right\} \geq A_{i+1} - \frac{\zeta}{2},$$

(82)

Proof: For each $k = 1, 2, \ldots, m$, let

$$S_k^n \triangleq \left\{ (x^n, y^n) : \frac{1}{n} \log \frac{1}{P_{X^n|Y^n}(x^n|y^n)} \geq H(X_k|Y_k) - \zeta \right\}.$$ 

(83)

As i.i.d. sources satisfy the AEP [28], we can choose $n_1$ such that

$$\sum_{(x^n, y^n) \in S_k^n} P_{X^n|Y^n}(x^n, y^n) \geq 1 - \frac{\zeta}{4}, \quad \forall n \geq n_1, \forall k = 1, 2, \ldots, m.$$ 

(84)

Moreover, we can choose $n_0 \geq n_1$ such that

$$\frac{1}{n} \log \frac{\zeta}{4} \leq \frac{\zeta}{2}, \quad \forall n \geq n_0.$$ 

(85)

Then, for all $n \geq n_0$, any $i = 1, 2, \ldots, m$, and any $k = 1, 2, \ldots, i$,

$$\tilde{S}_i^n \triangleq \left\{ (x^n, y^n) : \frac{1}{n} \log \frac{1}{P_{X^n|Y^n}(x^n|y^n)} \geq H(X_i|Y_i) - \zeta \right\}.$$ 

(86)

and

$$\tilde{T}_i^n \triangleq \left\{ (x^n, y^n) : P_{X^n|Y^n}(x^n|y^n) \leq \frac{\zeta}{4} P_{X^n|Y^n}(x^n|y^n) \right\}.$$ 

(87)

satisfy that

$$\tilde{S}_i^n \cup \tilde{T}_i^n \supseteq \left\{ (x^n, y^n) : \frac{1}{n} \log \frac{\zeta/4}{P_{X^n|Y^n}(x^n|y^n)} \geq H(X_i|Y_i) - \zeta \right\} \geq H(X_i|Y_i) - \zeta - \frac{1}{n} \log \frac{\zeta}{4}.$$ 

(88)

As i.i.d. sources satisfy the AEP [28], we can choose $n_1$ such that

$$\sum_{(x^n, y^n) \in \tilde{S}_k^n} P_{X^n|Y^n}(x^n, y^n) \geq 1 - \frac{\zeta}{4}, \quad \forall n \geq n_1, \forall k = 1, 2, \ldots, m.$$ 

(100)

Moreover, we can choose $n_0 \geq n_1$ such that

$$\frac{1}{n} \log \alpha_k \leq \frac{\zeta}{3}, \quad \forall n \geq n_0, \forall k = 1, 2, \ldots, m.$$ 

(101)

Lemma 4. Fix $\zeta > 0$. Then, there exists an integer $n_0$ such that for all $n \geq n_0$ and all $i = 1, 2, \ldots, m$,

$$\Pr \left\{ \frac{1}{n} \log \frac{1}{P_{X^n|Y^n}(X^n|Y^n)} \leq H(X_i|Y_i) + \zeta \right\} \geq 1 - A_i - \frac{\zeta}{2}.$$ 

(98)

Proof: For each $k = 1, 2, \ldots, m$, let

$$S_k^n \triangleq \left\{ (x^n, y^n) : \frac{1}{n} \log \frac{1}{P_{X^n|Y^n}(x^n|y^n)} \leq H(X_k|Y_k) + \zeta \right\}.$$ 

(99)

As i.i.d. sources satisfy the AEP [28], we can choose $n_1$ such that

$$\sum_{(x^n, y^n) \in S_k^n} P_{X^n|Y^n}(x^n, y^n) \leq 1 - \frac{\zeta}{4}, \quad \forall n \geq n_1, \forall k = 1, 2, \ldots, m.$$ 

(100)

Moreover, we can choose $n_0 \geq n_1$ such that

$$\frac{1}{n} \log \alpha_k \leq \frac{\zeta}{3}, \quad \forall n \geq n_0, \forall k = 1, 2, \ldots, m.$$ 

(101)
and
\[ \frac{1}{n} \log \frac{\zeta}{4} \leq \frac{\zeta}{3} \quad \forall n \geq n_0. \] (102)

Then, for all \( n \geq n_0 \) and any \( i = 1, 2, \ldots, m \), and any \( k = i, i + 1, \ldots, m \),
\[ \mathcal{S}_k^n \triangleq \left\{ (x^n, y^n) : \frac{1}{n} \log \frac{1}{P_{X^n|Y^n}(x^n|y^n)} \leq H(X_i|Y_i) + \zeta \right\} \] (103)
and
\[ \mathcal{T}_k^n \triangleq X^n \times \left\{ y^n : P_{Y_k^n}(y^n) \leq \frac{\zeta}{4} P_{Y^n}(y^n) \right\} \] (104)
satisfy that
\[ \mathcal{S}_k^n \cup \mathcal{T}_k^n \]
\[ = \left\{ (x^n, y^n) : \frac{1}{n} \log \frac{1}{P_{X^n|Y^n}(x^n, y^n)} - \frac{1}{n} \log \frac{1}{P_{Y^n}(y^n)} \leq H(X_i|Y_i) + \zeta \right\} \cup \mathcal{T}_k^n \] (105)
\[ \supseteq \left\{ (x^n, y^n) : \frac{1}{n} \log \frac{1}{\alpha_k P_{X_k^n|Y_k^n}(x^n, y^n)} - \frac{1}{n} \log \frac{1}{P_{Y^n}(y^n)} \leq H(X_i|Y_i) + \zeta \right\} \cup \mathcal{T}_k^n \] (106)
\[ \supseteq \left\{ (x^n, y^n) : \frac{1}{n} \log \frac{\zeta/4}{P_{Y_k^n}(y^n)} \leq H(X_i|Y_i) + \zeta \right\} \] (107)
\[ \supseteq \left\{ (x^n, y^n) : \frac{1}{n} \log \frac{1}{P_{X_k^n|Y_k^n}(x^n|y^n)} \leq H(X_i|Y_i) + \frac{\zeta}{3} \right\} \] (108)
\[ \supseteq \mathcal{S}_k^n. \] (109)

where the last inclusion follows from the fact that \( H(X_k|Y_k) \leq H(X_i|Y_i) \). Thus, we have
\[ \sum_{(x^n, y^n) \in \mathcal{S}_k^n} P_{X^n|Y^n}(x^n, y^n) \]
\[ \geq \sum_{k=i}^{m} \alpha_k \sum_{(x^n, y^n) \in \mathcal{S}_k^n} P_{X_k^n|Y_k^n}(x^n, y^n) \] (110)
\[ \geq \sum_{k=i}^{m} \alpha_k \sum_{(x^n, y^n) \in \mathcal{S}_k^n} P_{X_k^n|Y_k^n}(x^n, y^n) \]
\[ - \sum_{k=i}^{m} \alpha_k \sum_{(x^n, y^n) \in \mathcal{T}_k^n} P_{X_k^n|Y_k^n}(x^n, y^n) \] (111)
\[ = \sum_{k=i}^{m} \alpha_k \sum_{(x^n, y^n) \in \mathcal{S}_k^n} P_{X_k^n|Y_k^n}(x^n, y^n) \]
\[ - \sum_{k=i}^{m} \alpha_k \sum_{(x^n, y^n) \in \mathcal{T}_k^n} P_{Y_k^n}(y^n) P_{X_k^n|Y_k^n}(x^n|y^n) \] (112)
\[ \geq (1 - A_i) \left( 1 - \frac{\zeta}{4} \right) \]
\[ - \sum_{k=1}^{i} \alpha_k \sum_{(x^n, y^n) \in \mathcal{T}_k^n} \frac{\zeta}{4} P_{Y^n}(y^n) P_{X_k^n|Y_k^n}(x^n|y^n) \] (113)
\[ \geq (1 - A_i) \left( 1 - \frac{\zeta}{4} \right) - \sum_{k=1}^{i} \alpha_k \sum_{y^n} \frac{\zeta}{4} P_{Y^n}(y^n) \] (114)
\[ \geq (1 - A_i) \left( 1 - \frac{\zeta}{4} \right) - \frac{\zeta}{4} \] (115)
\[ \geq 1 - A_i - \frac{\zeta}{2}. \] (116)

\[ \Box \]

**Lemma 5.** Fix \( \zeta > 0 \) such that \( H(X_i|Y_i) - \zeta > H(X_{j+1}|Y_{j+1}) + \zeta \) for all \( j = 1, 2, \ldots, m - 1 \). Then, for sufficiently large \( n \) and any \( i = 1, 2, \ldots, m \),
\[ \alpha_i - \zeta \]
\[ \leq \Pr \left\{ \frac{1}{n} \log \frac{1}{P_{X^n|Y^n}(x^n|y^n)} - H(X_i|Y_i) \leq \zeta \right\} \]
\[ \leq \alpha_i + \zeta. \] (117)

**Proof:** From Lemmata 3 and 4, we have
\[ \Pr \left\{ \frac{1}{n} \log \frac{1}{P_{X^n|Y^n}(x^n|y^n)} - H(X_i|Y_i) \leq \zeta \right\} \]
\[ = \Pr \left\{ \frac{1}{n} \log \frac{1}{P_{X^n|Y^n}(x^n|y^n)} \leq H(X_i|Y_i) + \zeta \right\} \]
\[ - \Pr \left\{ \frac{1}{n} \log \frac{1}{P_{X^n|Y^n}(x^n|y^n)} < H(X_i|Y_i) - \zeta \right\} \] (118)
\[ \geq \{1 - A_i - \zeta/2\} - \{1 - (A_{i+1} - \zeta/2)\} \] (119)
\[ = \alpha_i - \zeta \] (120)

and
\[ \Pr \left\{ \frac{1}{n} \log \frac{1}{P_{X^n|Y^n}(x^n|y^n)} - H(X_i|Y_i) \leq \zeta \right\} \]
\[ = \Pr \left\{ \frac{1}{n} \log \frac{1}{P_{X^n|Y^n}(x^n|y^n)} \leq H(X_i|Y_i) + \zeta \right\} \]
\[ - \Pr \left\{ \frac{1}{n} \log \frac{1}{P_{X^n|Y^n}(x^n|y^n)} < H(X_i|Y_i) - \zeta \right\} \] (121)
\[ \leq \Pr \left\{ \frac{1}{n} \log \frac{1}{P_{X^n|Y^n}(x^n|y^n)} < H(X_{i+1}|Y_{i+1}) - \zeta \right\} \]
\[ - \Pr \left\{ \frac{1}{n} \log \frac{1}{P_{X^n|Y^n}(x^n|y^n)} \leq H(X_{i+1}|Y_{i+1}) + \zeta \right\} \] (122)
\[ \leq \{1 - (A_i - \zeta/2)\} - \{1 - (A_{i+1} - \zeta/2)\} \] (123)
\[ = \alpha_i + \zeta. \] (124)

\[ \Box \]

**B. Proof of Theorem 2**

To prove the proposition, it is sufficient to show that, for \( \varepsilon \) satisfying \( A_i < \varepsilon < A_{i+1} \),
\[ \limsup_{n \to \infty} \frac{1}{n} H_\varepsilon^n(X^n|Y^n) \leq H(X_i|Y_i) \] (125)
and
\[
\liminf_{n \to \infty} \frac{1}{n} H_n^c(X^n|Y^n) \geq H(X_i|Y_i). \tag{126}
\]

**Proof of (125):** Fix \(\zeta > 0\) to be sufficiently small such that \(H(X_i|Y_i) - \zeta > H(X_{i+1}|Y_{i+1}) + \zeta\) for all \(j = 1, 2, \ldots, m - 1\) and that \(A_i + m\zeta < \varepsilon\). For \(j = 1, 2, \ldots, m\), let
\[
T_n(j) = \left\{ (x^n, y^n) : \frac{1}{n} \log \frac{1}{P_{X^n|Y^n}(x^n|y^n)} - H(X_j|Y_j) \leq \zeta \right\} \tag{127}
\]
and for each \(y^n \in \mathcal{Y}^n\) let
\[
T_n(j|y^n) = \left\{ x^n : (x^n, y^n) \in T_n(j) \right\}. \tag{128}
\]
Note that \(T_n(j) \cap T_n(j') = \emptyset (j \neq j')\). Further, from Lemma 5, we have
\[
\Pr \left\{ (X^n, Y^n) \in \bigcup_{j=1}^m T_n(j) \right\} = \sum_{j=1}^m \Pr \left\{ (X^n, Y^n) \in T_n(j) \right\} \geq \sum_{j=1}^m (\alpha_j - \zeta) \geq 1 - A_i - m\zeta \geq 1 - \varepsilon. \tag{130}
\]
From (132), we can see that
\[
Q_n(x^n, y^n) \triangleq \begin{cases} P_{X^n|Y^n}(x^n, y^n), & \text{if } (x^n, y^n) \in \bigcup_{j=1}^m T_n(j) \\ 0, & \text{otherwise} \end{cases} \tag{133}
\]
satisfies \(Q_n \in \mathcal{B}^c(P_{X^n|Y^n})\). Thus, from the definition of \(r_n^c(X^n|Y^n)\),
\[
r_n^c(X^n|Y^n) \leq \sum_{y^n \in \mathcal{Y}^n} \left[ \sum_{x^n \in \mathcal{X}^n} [Q_n(x^n, y^n)]^a \right]^{1/a} \leq \sum_{y^n \in \mathcal{Y}^n} \left[ \sum_{j=1}^m \sum_{x^n \in T_n(j)|y^n} [P_{X^n|Y^n}(x^n, y^n)^a \right]^{1/a} \leq \sum_{y^n \in \mathcal{Y}^n} P_{Y^n}(y^n) \left[ \sum_{j=1}^m \sum_{x^n \in T_n(j)|y^n} [P_{X^n|Y^n}(x^n|y^n)^a \right]^{1/a} \leq \sum_{y^n \in \mathcal{Y}^n} P_{Y^n}(y^n) \times \left[ \sum_{j=1}^m [T_n(j)|y^n] \exp \{-\alpha n (H(X_j|Y_j) - \zeta)\} \right]^{1/a} \tag{137}
\]

\[
\leq \sum_{y^n \in \mathcal{Y}^n} P_{Y^n}(y^n) \left[ \sum_{j=1}^m \exp \{n(H(X_j|Y_j) + \zeta)\} \times \exp \{-\alpha n (H(X_j|Y_j) - \zeta)\} \right]^{1/a} \tag{138}
\]
\[
= \sum_{y^n \in \mathcal{Y}^n} P_{Y^n}(y^n) \times \left[ \sum_{j=1}^m \exp \{n[(1 - \alpha) H(X_j|Y_j) + (1 + \alpha)\zeta]\} \right]^{1/a} \tag{139}
\]
\[
\leq \sum_{y^n \in \mathcal{Y}^n} P_{Y^n}(y^n) \times \left[ m \exp \{n[(1 - \alpha) H(X_j|Y_j) + (1 + \alpha)\zeta]\} \right]^{1/a} = m^{1/\alpha} \exp \left\{ n \left[ \frac{1}{\alpha} (1 - \alpha) H(X_j|Y_j) + (1 + \alpha)\zeta / \alpha \right] \right\}. \tag{140}
\]
Hence, we have
\[
\frac{1}{n} H_n^c(P_{X^n}) \leq H(X_i|Y_i) + \frac{1 + \alpha}{1 - \alpha} \zeta + \frac{1}{n(1 - \alpha)} \log m \tag{142}
\]
and thus
\[
\limsup_{n \to \infty} \frac{1}{n} H_n^c(P_{X^n}) \leq H(X_i|Y_i) + \frac{1 + \alpha}{1 - \alpha} \zeta. \tag{143}
\]
As we can choose \(\zeta \to 0\) arbitrarily small, we have (125). \(\square\)

**Proof of (126):** If \(H(X_i) = 0\) then (126) is apparent, since (20) holds. Thus, we assume \(H(X_i) > 0\). Fix \(\zeta > 0\) sufficiently small such that \(H(X_j) - \zeta > H(X_j + 1) + \zeta\) for all \(j = 1, 2, \ldots, m - 1\) and that \(A_i + 6m\zeta < \varepsilon < A_{i+1} - 6m\zeta\). We assume that \(n\) is sufficiently large such that \(\exp \{-n[H(X_i|Y_i) - \zeta]\} \leq m\zeta\).

In this proof, we use \(x\) (resp. \(y\)) instead of \(x^n \in \mathcal{X}^n\) (resp. \(y^n \in \mathcal{Y}^n\)) to simplify the notation; \(n\) should be apparent from the context. From Theorem 1, we can choose \((\varepsilon_y)_{y \in \mathcal{Y}^n} \in \mathcal{E}_0(\varepsilon), Q_{i_y}^+(\cdot|y)\), and \(i_y^+ = i_y^-(\varepsilon_y)\) such that
\[
H_n^c(X^n|Y^n) + \zeta \geq \frac{\alpha}{1 - \alpha} \log \sum_{y^n \in \mathcal{Y}^n} P_{Y^n}(y) \left[ \sum_{i=1}^{i_y^+} Q_{i_y}^+(x_i^n|y)^a \right]^{1/a} \tag{144}
\]
where \(x_1^+, x_2^+, x_3^+, \ldots\) are sorted such that
\[
P_{X^n|Y^n}(x_1^n|y) \geq P_{X^n|Y^n}(x_2^n|y) \geq P_{X^n|Y^n}(x_3^n|y) \geq \cdots. \tag{145}
\]
Let \(A_n(y) \triangleq \{x_i^n : i = 1, 2, \ldots, i_y^+ - 1\}\) and \(A_{n}^+(y) \triangleq A_n(y) \cup \{x_{i_y^+}^n\}. Then, letting\(A_n^+ \triangleq \bigcup_{y \in \mathcal{Y}^n} A_{n}^+(y) \times \{y\}, \tag{146}\)
we have
\[
\Pr \{(X^n, Y^n) \in A_{n}^+\} \geq 1 - \varepsilon. \tag{147}\]
On the other hand, let us define $T_n(j)$ and $T_n(j|y)$ as in (127) and (128). Then, from Lemma 5, we have

$$\Pr \left\{ (X^n, Y^n) \notin \bigcup_{j=1}^{m} T_n(j) \right\} \leq m\zeta$$  \hspace{1cm} (148)

and

$$\Pr \left\{ (X^n, Y^n) \in \bigcup_{j=1}^{m} T_n(j) \right\} \leq \sum_{j=1}^{m} \sum (a_j + \zeta)$$  \hspace{1cm} (149)

$$\leq 1 - A_i - 1 + m\zeta$$  \hspace{1cm} (150)

Thus, from (147), (148), and (151), we have

$$\Pr \left\{ (X^n, Y^n) \in A_i^+ \cap \bigcup_{j=1}^{i} T_n(j) \right\} \geq 4m\zeta.$$  \hspace{1cm} (152)

In other words,

$$\sum_{y \in \mathcal{Y}^n} P_{Y^n}(y) \sum_{x \in A_i^+(y) \cap \bigcup_{j=1}^{i} T_n(j)} P_{X^n|Y^n}(x|y) \geq 4m\zeta.$$  \hspace{1cm} (153)

Hence, for all $y \in \mathcal{U}_n$,

$$\sum_{x \in A_i(y) \cap \bigcup_{j=1}^{i} T_n(j)} P_{X^n|Y^n}(x|y) \geq 4m\zeta.$$  \hspace{1cm} (154)

Now fix $y \in \mathcal{U}_n$. If $x^n \notin \bigcup_{j=1}^{i} T_n(j|y)$ then

$$\sum_{x \in A_i(y) \cap \bigcup_{j=1}^{i} T_n(j|y)} P_{X^n|Y^n}(x|y) \geq 2m\zeta.$$  \hspace{1cm} (155)

and thus,

$$\sum_{y \in \mathcal{U}_n} P_{Y^n}(y^n) \geq \frac{2m\zeta}{1 - 2m\zeta}.$$  \hspace{1cm} (156)

If $x^n \in T_n(j|y)$ for some $j = 1, \ldots, i$ then

$$P_{X^n|Y^n}(x^n|y) \leq \exp(-n|H(X_j|Y_j) - \zeta|) \leq \exp(-n|H(X_j|Y_j) - \zeta|) \leq m\zeta.$$  \hspace{1cm} (157)

Hence, we have

$$\sum_{x \in \mathcal{A}_n(y) \cap \bigcup_{j=1}^{i} T_n(j|y)} P_{X^n|Y^n}(x|y) \geq \sum_{x \in \mathcal{A}_n(y) \cap \bigcup_{j=1}^{i} T_n(j|y)} P_{X^n|Y^n}(x|y) - m\zeta \geq m\zeta.$$  \hspace{1cm} (159)

Further, let $p_j(y) \triangleq \sum_{x \in A_i(y) \cap T_n(j|y)} P_{X^n|Y^n}(x|y)$ for $j = 1, 2, \ldots, i$. Then, since $T_n(j|y) \cap T_n(j'|y) = \emptyset$ $(j \neq j')$, we have

$$\sum_{j=1}^{i} p_j(y) \geq m\zeta,$$  \hspace{1cm} (160)

and

$$|A_i(y) \cap T_n(j|y)| \geq p_j(y) \exp[n|H(X_j|Y_j) - \zeta|].$$  \hspace{1cm} (161)

Now, recall (144). From (156) and (167), we have

$$H_{\alpha}^{\zeta}(X^n|Y^n) + \zeta$$

$$\geq \frac{a}{1 - \alpha} \log \sum_{y \in \mathcal{U}_n} P_{Y^n}(y) \sum_{i=1}^{t_j} \frac{1}{Q_{i,y}} 1/a$$  \hspace{1cm} (168)

$$\geq \frac{a}{1 - \alpha} \log \sum_{y \in \mathcal{U}_n} P_{Y^n}(y) \sum_{i=1}^{t_j-1} \frac{1}{Q_{i,y}} 1/a$$  \hspace{1cm} (169)

$$\geq \frac{a}{1 - \alpha} \log \sum_{y \in \mathcal{U}_n} P_{Y^n}(y) \sum_{x \in \mathcal{A}_n(y)} \left[ P_{X^n|Y^n}(x|y) \right]^a 1/a$$  \hspace{1cm} (170)
\[
\times \left[ \sum_{x \in \mathcal{X}} P_{X|x} (x|y)^{\alpha} \right]^{1/\alpha} \geq \frac{a}{1 - \alpha} \log \sum_{y \in \mathcal{Y}} P_{Y|X}(y) \\
\times (m \zeta)^{1/\alpha} \exp \left\{ n \left[ \frac{1 - \alpha}{\alpha} H(X|Y) - \frac{(1 + \alpha) \zeta}{\alpha} \right] \right\} \]  
\]  
(172)

and thus
\[
\liminf_{n \to \infty} \frac{1}{n} H_e^c(X^n|Y^n) \geq H(X|Y) - \frac{1 + \alpha}{1 - \alpha}. \tag{175}
\]

As we can choose \( \zeta > 0 \) arbitrarily small, we have (126). \( \square \)

**APPENDIX D**

**PROOF OF THEOREMS 3, 4, AND 5**

Proof of Theorems 3, 4 and 5 is given in Subsections D-A, D-B, and D-C, respectively.

**A. Proof of Theorem 3**

The proof of Theorem 3 is essentially the same as that of Theorem 1 of Arikan [2]. To prove Theorem 3, we use Lemma 1 of [2]:

**Lemma 6** (Lemma 1 of [2]). For nonnegative numbers \( a_i \) and \( q_i \) (\( i = 1, 2, \ldots, K \)), and any \( 0 < \lambda < 1 \), we have
\[
\sum_{i=1}^{K} a_i q_i \geq \left[ \sum_{i=1}^{K} a_i \right]^{1/\lambda} \left[ \sum_{i=1}^{K} q_i \right]^{\lambda}. \tag{176}
\]

Given a guessing strategy \( G \), let
\[
Q(x, y) \triangleq \lambda_y(\sigma_y(x)) P_{XY}(x, y), \quad (x, y) \in \mathcal{X} \times \mathcal{Y}
\]
where \( \lambda_y \) is defined as in (27). From (31), we have
\[
\sum_{(x, y) \in \mathcal{X} \times \mathcal{Y}} Q(x, y) = 1 - p_x \geq 1 - \epsilon.
\]

Further, it is apparent that \( 0 \leq \lambda_y(\sigma_y(x)) \leq 1 \) for all \( (x, y) \in \mathcal{X} \times \mathcal{Y} \), and hence
\[
Q \in \mathcal{B}^c(P_{XY}). \tag{176}
\]

Fix \( y \in \mathcal{Y} \). Letting \( q_i = Q(\sigma_y^{-1}(i), y)/P_Y(y), a_i = i^\rho \), and \( \lambda = 1/(1 + \rho) \) in Lemma 6, we have
\[
\sum_{i=1}^{K} \frac{Q(\sigma_y^{-1}(i), y)}{P_Y(y)} i^\rho \geq \left[ \sum_{i=1}^{K} \frac{Q(\sigma_y^{-1}(i), y)}{P_Y(y)} \right]^{1/\lambda} \left[ \sum_{i=1}^{K} i^{\lambda} \right]^{\lambda-1} \tag{177}
\]

and thus
\[
\sum_{x \in \mathcal{X}} \lambda_y(\sigma_y(x)) P_{XY}(x, y) \sigma_y(x)^\rho \geq \left[ \sum_{i=1}^{K} \frac{Q(\sigma_y^{-1}(i), y)}{P_Y(y)} \right]^{1/\lambda} \left[ \sum_{i=1}^{K} i^{\lambda} \right]^{\lambda-1} \tag{178}
\]

Taking the summation overall \( y \in \mathcal{Y} \) and recalling (35), we have
\[
\hat{C}_\rho \geq \left( 1 + \log K \right)^{-\rho} \sum_{y \in \mathcal{Y}} \left[ \sum_{x \in \mathcal{X}} Q(x, y) \right]^{1/\lambda}. \tag{179}
\]

Further, from the fact that
\[
\sum_{i=1}^{K} i \leq 1 + \log K, \tag{180}
\]
we have
\[
\hat{C}_\rho \geq (1 + \log K)^{-\rho} \sum_{y \in \mathcal{Y}} \left[ \sum_{x \in \mathcal{X}} Q(x, y) \right]^{1/\lambda}. \tag{181}
\]

Hence, from the definition of \( H^c_e(Y|X) \) and (176), we have (36). \( \square \)

**B. Proof of Theorem 4**

To prove Theorem 4, we construct a guessing strategy \( G \) as follows. Recall Theorem 1: for any \( \zeta > 0 \), we can choose \( (e_y)_{y \in \mathcal{Y}}' \in \mathcal{E}_0(\epsilon), i^* = i^*_y(e_y) \), and \( Q^*_y \), satisfying
\[
\rho H_e^c(Y|X) + \gamma = \log \sum_{y \in \mathcal{Y}} P_Y(y) \left[ \sum_{i=1}^{K} Q^*_y(\sigma_y^{-1}(i), y) \right]^{1/\lambda}. \tag{182}
\]

Using this notation, for each \( y \in \mathcal{Y} \), let us define \( \sigma_y \) and \( \pi_y \) such that
\[
\sigma_y(\pi_y) = i \tag{183}
\]

and
\[
\pi_y(i) = \begin{cases} 0 : i < i^*_y, \\ 1 - \frac{Q^*_y(\sigma_y^{-1}(i), y) / P_X(y) \sigma_y^{-1}(i), y)}{P_X(\sigma_y^{-1}(i), y)} : i = i^*_y, \\ 1 : i > i^*_y, \end{cases} \tag{184}
\]

From the definition of \( Q^*_y \) in (15) and the definition of \( \lambda_y(i) \) in (27), we note that
\[
Q^*_y(\sigma_y^{-1}(i), y) = \lambda_y(i) P_{X|Y}(\sigma_y^{-1}(i), y), \quad y \in \mathcal{Y}, i = 1, 2, \ldots, K. \tag{185}
\]
Hence, we can upper bound $\tilde{C}_\rho = \tilde{C}_\rho(G|X, Y)$ as follows.\(^9\)
For each $y \in \mathcal{Y}$,
\[
\sigma_y(x) = \sum_{\hat{x} : \sigma(x) \leq \sigma_y(x)} 1
\leq \sum_{\hat{x} : \sigma(x) \leq \sigma_y(x)} \left[ \frac{Q_{x\hat{x}}^*(\hat{x}|y)}{Q_{\hat{x}\hat{x}}^*(\hat{x}|y)} \right]^{\frac{1}{1+\rho}}
\leq \sum_{\hat{x} \in \mathcal{X}} \left[ \frac{Q_{x\hat{x}}^*(\hat{x}|y)}{Q_{\hat{x}\hat{x}}^*(\hat{x}|y)} \right]^{\frac{1}{1+\rho}}
\approx \frac{1}{K} \sum_{j=1}^{K} \frac{Q_{x\hat{x}}^*(\hat{x}|y)}{Q_{\hat{x}\hat{x}}^*(\hat{x}|y)}
= \frac{1}{K} \sum_{j=1}^{K} \frac{Q_{x\hat{x}}^*(\hat{x}|y)}{Q_{\hat{x}\hat{x}}^*(\hat{x}|y)}
= \sum_{j=1}^{K} \frac{Q_{x\hat{x}}^*(\hat{x}|y)}{Q_{\hat{x}\hat{x}}^*(\hat{x}|y)}
\]
and thus, for each $i$,
\[
i = \sigma_y(x_i^j)
\leq \sum_{j=1}^{i^\varepsilon} \left[ \frac{Q_{x\hat{x}}^*(\hat{x}|y)}{Q_{\hat{x}\hat{x}}^*(\hat{x}|y)} \right]^{\frac{1}{1+\rho}}.
\]
Hence, we have
\[
\tilde{C}_\rho
= \sum_{y \in \mathcal{Y}} P_Y(y) \left\{ \sum_{i=1}^{i^\varepsilon} Q_{x\hat{x}}^*(\hat{x}|y) i^\rho \right\}
= \sum_{y \in \mathcal{Y}} P_Y(y) \left\{ \sum_{i=1}^{i^\varepsilon} Q_{x\hat{x}}^*(\hat{x}|y) i^\rho \right\}
\leq \sum_{y \in \mathcal{Y}} P_Y(y) \left\{ \sum_{i=1}^{i^\varepsilon} Q_{x\hat{x}}^*(\hat{x}|y) \left[ \frac{Q_{x\hat{x}}^*(\hat{x}|y)}{Q_{\hat{x}\hat{x}}^*(\hat{x}|y)} \right]^{\frac{1}{1+\rho}} \right\}
= \sum_{y \in \mathcal{Y}} P_Y(y) \left\{ \sum_{i=1}^{i^\varepsilon} Q_{x\hat{x}}^*(\hat{x}|y) \left[ \frac{Q_{x\hat{x}}^*(\hat{x}|y)}{Q_{\hat{x}\hat{x}}^*(\hat{x}|y)} \right]^{\frac{1}{1+\rho}} \right\}
= \sum_{y \in \mathcal{Y}} P_Y(y) \left[ \sum_{i=1}^{i^\varepsilon} Q_{x\hat{x}}^*(\hat{x}|y) \right]^{1+\rho}
\exp \left\{ \rho H_{\frac{1}{1+\rho}}(X|Y) + \gamma \right\}
\]
where the last equality follows from (182). As we can choose $\varepsilon > 0$ arbitrarily small, we have (37).\(^9\)

C. Proof of Theorem 5

Direct Part: Fix $\delta > 0$ such that $0 < \varepsilon + \delta < 1$. From Theorem 4, there exists $\{G_n\}_{n=1}^\infty$ such that, for $n = 1, 2, \ldots$,
\[
pe(G_n|X^n, Y^n) \leq \varepsilon + \delta
\]
and
\[
\tilde{C}_\rho(G_n|X^n, Y^n) \leq \exp \left\{ \rho H_{\frac{1}{1+\rho}}(X^n|Y^n) \right\}.
\]
Hence,
\[
\limsup_{n \to \infty} \frac{1}{n} \log \tilde{C}_\rho(G_n|X^n, Y^n)
\leq \rho \limsup_{n \to \infty} \frac{1}{n} H_{\frac{1}{1+\rho}}(X^n|Y^n).
\]
By using the diagonal line argument (see [7]), we can conclude that $\rho H_{\frac{1}{1+\rho}}(X|Y)$ is $\varepsilon$-achievable.\(^9\)

Converse Part: Assume that $E_g$ is $\varepsilon$-achievable and fix $\delta > 0$. Then there exists $\{G_n\}_{n=1}^\infty$ such that, for sufficiently large $n$,
\[
pe(G_n|X^n, Y^n) \leq \varepsilon + \delta
\]
and
\[
\limsup_{n \to \infty} \frac{1}{n} \log \tilde{C}_\rho(\Phi_n|X^n, Y^n) \leq E_g.
\]
On the other hand, from Theorem 3, for sufficiently large $n$ such that (202) holds,
\[
\tilde{C}_\rho(G_n|X^n, Y^n)
\geq (1 + \log K)^{-\rho} \exp \left\{ \rho H_{\frac{1}{1+\rho}}(X^n|Y^n) \right\}.
\]
Combining (203) with (204), we have
\[
E_g \geq \rho \limsup_{n \to \infty} \frac{1}{n} H_{\frac{1}{1+\rho}}(X^n|Y^n).
\]
As $\delta > 0$ is arbitrary, letting $\delta \downarrow 0$, we have $E_g \geq \rho H_{\frac{1}{1+\rho}}(X|Y)$.\(^9\)

\section*{Appendix E}

Proof of Theorems 6, 7, and 8

Proof of Theorems 6, 7, and 8 is given in Subsections E-A, E-B, and E-C, respectively.

A. Proof of Theorem 6

Fix a code $\Phi$ such that $pe(\Phi) \leq 1 - \varepsilon$.\(^9\)
Fix $y \in \mathcal{Y}$. Recall that we allow $\varphi_y$ to be stochastic. Let $W_y(c|x) \triangleq W_{\varphi_y}(c|x)$ be the probability such that $x \in \mathcal{X}$ is mapped to $c \in \mathcal{C}$ by $\varphi_y$. Let
\[
\gamma_{xy} \triangleq \{ c \in \mathcal{C} : W_y(c|x) > 0, x = \varphi_y(c) \}
\]
and
\[
\gamma_{xy} \triangleq \sum_{c \in \gamma_{xy}} W_y(c|x).
\]
Note that, since $pe(\Phi) \leq \varepsilon$, we have
\[
\sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} \gamma_{xy} P_{XY}(x,y) \geq 1 - \varepsilon.
\]
\(^9\)The following step is essentially the same as that of Proposition 4 of Arikan [2].
Further, using Jensen’s inequality, we have

\[
\sum_{c \in \Gamma_y} W_y(c|x) \exp[\rho L(c) \log 2] \geq \sum_{c \in \Gamma_y(x)} W_y(c|x) \exp[\rho L(c) \log 2]
\]

(209)

\[
\geq \gamma_{xy} \exp \left\{ \rho \sum_{c \in \Gamma_y(x)} \frac{W_y(c|x)}{\gamma_{xy}} L(c) \log 2 \right\}
\]

(210)

\[
\geq \gamma_{xy} \exp \left\{ \rho \sum_{c \in \Gamma_y(x)} \frac{W_y(c|x)}{\gamma_{xy}} \tilde{\ell}(x|y) \right\}
\]

(211)

\[
= \gamma_{xy} \exp \{\rho \tilde{\ell}(x|y)\}
\]

(212)

where

\[
\tilde{\ell}(x|y) \triangleq \min_{c \in \Gamma_y(x)} L(c) \log 2.
\]

(213)

Substituting (212) into (50), we have

\[
M_\rho(\Phi|X, Y) \geq \sum_{(x,y) \in A^r \times \mathcal{Y}} \gamma_{xy} P_{XY}(x,y) \exp \{\rho \tilde{\ell}(x|y)\}.
\]

(214)

Let \(Q(x,y) = \gamma_{xy} P_{XY}(x,y)\). Then, from (208), we have \(Q \in B^\rho(P_{XY})\). Further, let \(Q(x,y) \triangleq \frac{Q(x,y)}{P_Y(y)}\) and \(A_y \triangleq \{x : Q(x|y) > 0\}\) for each \(y \in \mathcal{Y}\). Then, (214) can be written as

\[
M_\rho \geq \sum_{y \in \mathcal{Y}} P_Y(y) \sum_{x \in A_y} Q(x|y) \exp \{\rho \tilde{\ell}(x|y)\}.
\]

(215)

Now, fix \(y \in \mathcal{Y}\). From the definition of the set \(\Gamma_y(x)\), we can see that \(\Gamma_y(x) \cap \Gamma_y(x') = \emptyset\) for all \(x, x' \in A_y\) such that \(x \neq x'\). Thus, from the prefix condition, we have

\[
\sum_{x \in A_y} \exp (-\tilde{\ell}(x|y)) \leq 1.
\]

(216)

Then, let us consider the problem of minimizing \(\sum_{x \in A_y} Q(x|y) \exp \{\rho \tilde{\ell}(x|y)\}\) subject to (216). As shown in Section 3.1 of [22], the minimum is achieved by

\[
\tilde{\ell}(x|y) = -\log \frac{[Q(x|y)]^{1/(1+\rho)}}{\sum_{x' \in A_y} [Q(x'|y)]^{1/(1+\rho)}}, \ x \in A_y.
\]

(217)

Applying the above argument for each \(y \in \mathcal{Y}\), we can rewrite (215) as

\[
M_\rho \geq \sum_{y \in \mathcal{Y}} P_Y(y) \sum_{x \in A_y} Q(x|y) \exp \{\rho \tilde{\ell}(x|y)\}
\]

(218)

\[
\geq \sum_{y \in \mathcal{Y}} P_Y(y) \left[ \sum_{x \in A_y} Q(x|y) \exp \left\{ -\rho \log \frac{[Q(x|y)]^{1/(1+\rho)}}{\sum_{x' \in A_y} [Q(x'|y)]^{1/(1+\rho)}} \right\} \right]
\]

(219)

\[
= \sum_{y \in \mathcal{Y}} P_Y(y) \sum_{x \in A_y} Q(x|y) \exp \left\{ -\rho \log \frac{[Q(x|y)]^{1/(1+\rho)}}{\sum_{x' \in A_y} [Q(x'|y)]^{1/(1+\rho)}} \right\}
\]

(220)

\[
= \sum_{y \in \mathcal{Y}} \sum_{x \in A_y} Q(x, y) \exp \left\{ -\rho \log \frac{[Q(x, y)]^{1/(1+\rho)}}{\sum_{x' \in A_y} [Q(x', y)]^{1/(1+\rho)}} \right\}
\]

(221)

\[
= \sum_{y \in \mathcal{Y}} \sum_{x \in A_y} Q(x, y) \left[ \frac{[Q(x, y)]^{1/(1+\rho)}}{\sum_{x' \in A_y} [Q(x', y)]^{1/(1+\rho)}} \right]^{-\rho}
\]

(222)

\[
= \sum_{y \in \mathcal{Y}} \sum_{x \in A_y} \left[ \frac{[Q(x, y)]^{1/(1+\rho)}}{\sum_{x' \in A_y} [Q(x', y)]^{1/(1+\rho)}} \right] \rho
\]

(223)

\[
= \sum_{y \in \mathcal{Y}} \left[ \sum_{x \in A_y} [Q(x, y)]^{1/(1+\rho)} \right]^{1+\rho}
\]

(224)

\[
= \sum_{y \in \mathcal{Y}} \left[ \sum_{x \in A_x} [Q(x, y)]^{1/(1+\rho)} \right]^{1+\rho}
\]

(225)

\[
\geq r_{1/(1+\rho)}^Y(X|Y)
\]

(226)

where the last inequality follows from the fact \(Q \in B^\rho(P)\).

By the definition of the conditional smooth Rényi entropy, we have (52).

\[
\square
\]

B. Proof of Theorem 7

Fix \(\zeta > 0\) and choose \(Q \in B^\rho(P_{XY})\) such that

\[
\log \sum_{y \in \mathcal{Y}} \left[ \sum_{x \in A_x} [Q(x, y)]^{1/(1+\rho)} \right]^{1+\rho} \leq \rho H_{1/(1+\rho)}^Y(X|Y) + \zeta.
\]

(227)

Letting \(Q(x,y) \triangleq Q(x,y)/P_Y(y)\), we have

\[
\log \sum_{y \in \mathcal{Y}} P_Y(y) \left[ \sum_{x \in A_x} [Q(x, y)]^{1/(1+\rho)} \right]^{1+\rho} \leq \rho H_{1/(1+\rho)}^Y(X|Y) + \zeta.
\]

(228)

Fix \(y \in \mathcal{Y}\). Let \(A_y \triangleq \{x \in A : Q(x|y) > 0\}\) and

\[
\hat{Q}^{(\rho)}(x|y) \triangleq \frac{[Q(x|y)]^{1/(1+\rho)}}{\sum_{x' \in A_y} [Q(x'|y)]^{1/(1+\rho)}}.
\]

(229)

As

\[
\sum_{x \in A_y} 2^{-\log \hat{Q}^{(\rho)}(x|y)} \leq 1
\]

(230)

holds true, we can construct a variable-length code \((\hat{\phi}_y, \psi_y, \hat{\gamma}_y, \zeta_y)\) such that (i) \(\hat{\gamma}_y \triangleq \hat{\gamma}_y(x) : x \in A_y\) is prefix free, (ii) \(\hat{\phi}_y : A_y \to \hat{\gamma}_y\) satisfies

\[
L(\hat{\phi}_y(x)) = -\log_2 \hat{Q}^{(\rho)}(x|y),
\]

(231)

and, (iii) \(\hat{\psi}_y : \hat{\gamma}_y \to A_y\) satisfy \(x = \psi_y(\hat{\phi}_y(x))\) for all \(x \in A_y\); e.g., we can use the Shannon code for the
distribution $\tilde{Q}^{(p)}(x|y)$. Further, for each $x \in \mathcal{X}$, let $\gamma_{xy} = Q(x|y)/P_{X|Y}(x|y)$. Note that $0 \leq \gamma_{xy} \leq 1$ and $\gamma_{xy} = 0$ for all $x \notin A_y$. Then, we construct a stochastic encoder for $X$ given $y$ as follows:

$$\varphi_y(x) = \begin{cases} 0 & \text{with probability } \gamma_{xy} \\ 1 & \text{with probability } 1 - \gamma_{xy} \end{cases}$$

(232)

where $\circ$ denotes the concatenation. That is, $x$ is encoded to “0” following $\varphi_y(x)$ with probability $\gamma_{xy}$, and “1” with probability $1 - \gamma_{xy}$. We can construct the corresponding decoder $\psi_y$ such that $x = \psi_y(\varphi_y(x))$ for all $x \in \mathcal{A}_y$ if $x$ is encoded to 0 or $\varphi_y(x)$. The length function $\ell(x|y) = L(\varphi_y(x))$ log 2 satisfies that, if $x$ is encoded to “0” following $\varphi_y(x)$,

$$\ell(x|y) \leq -\log \tilde{Q}^{(p)}(x|y) + 2 \log 2$$

(233)

and otherwise $\ell(x|y) = \log 2$.

By applying the above argument for each $y \in \mathcal{Y}$, we have

$$M(p) = \mathbb{E}_{P_{XY}} \left[ \exp\{p \ell(X|Y)\} \right]$$

$$\leq \sum_{y \in \mathcal{Y}} P_y(y) \sum_{x \in \mathcal{X}} P_x|y(x|y) \times \gamma_{xy} \exp\left\{ \rho \left( -\log \tilde{Q}^{(p)}(x|y) + 2 \log 2 \right) \right\} + \sum_{y \in \mathcal{Y}} P_y(y) \sum_{x \in \mathcal{X}} P_x|y(x|y)(1 - \gamma_{xy}) \exp(\rho \log 2)$$

$$\leq 2^{2\rho} \sum_{y \in \mathcal{Y}} P_y(y) \sum_{x \in \mathcal{A}_y} Q(x|y) \exp\left\{ -\rho \log \tilde{Q}^{(p)}(x|y) \right\} + \varepsilon 2^p$$

(236)

$$= 2^{2\rho} \sum_{y \in \mathcal{Y}} P_y(y) \left\{ \sum_{x \in \mathcal{A}_y} \left[ Q(x|y) \right]^{1/(1+\rho)} \right\}^{1/(1+\rho)} + \varepsilon 2^p$$

(237)

$$\leq 2^{2\rho} \exp\left\{ \rho H_{1/(1+\rho)}^\zeta(P) + \zeta \right\} + \varepsilon 2^p$$

(238)

where the inequality (a) follows from the fact that, since $Q \in \mathcal{B}_\zeta(P)$,

$$\sum_{y \in \mathcal{Y}} P_y(y) \sum_{x \in \mathcal{X}} P_x|y(x|y) \gamma_{xy} = \sum_{x, y} Q(x, y) \geq 1 - \varepsilon$$

(239)

and the inequality (b) follows from (228).

As we can choose $\zeta > 0$ arbitrarily small, we have (53).

C. Proof of Theorem 8

Direct Part: Fix $\delta > 0$ such that $0 < \varepsilon + \delta < 1$. From Theorem 7, there exists $\{\Phi_n\}_{n=1}^\infty$ such that, for $n = 1, 2, \ldots$,

$$p_e(\Phi_n|X^n, Y^n) \leq \varepsilon + \delta$$

(240)

and

$$M(p|X^n, Y^n) \leq 2 \max\left\{ 2^{2p} \exp\left\{ \rho H_{1/(1+\rho)}^\zeta(X^n|Y^n) \right\}, (\varepsilon + \delta)2^p \right\}.$$  

(241)

From (20) and (241), we can see

$$\limsup_{n \to \infty} \frac{1}{n} \log M(p|X^n, Y^n) \leq \rho \limsup_{n \to \infty} \frac{1}{n} H_{1/(1+\rho)}^\zeta(X^n|Y^n).$$

(242)

By using the diagonal line argument (refer [7]), we can conclude that $p H_{1/(1+\rho)}^\zeta(X|Y)$ is $\varepsilon$-achievable.

Converse Part: Suppose that $E_s$ is $\varepsilon$-achievable and fix $\delta > 0$. Then there exists $\{\Phi_n\}_{n=1}^\infty$ such that, for sufficiently large $n$,

$$p_e(\Phi_n|X^n, Y^n) \leq \varepsilon + \delta$$

(243)

and

$$\limsup_{n \to \infty} \frac{1}{n} \log M(p|X^n, Y^n) \leq E_s.$$  

(244)

On the other hand, from Theorem 6, for sufficiently large $n$ such that (243) holds,

$$M(p|X^n, Y^n) \geq \exp\left\{ \rho H_{1/(1+\rho)}^\zeta(X^n|Y^n) \right\}.$$  

(245)

Combining (244) with (245), we have

$$E_s \geq \rho \limsup_{n \to \infty} \frac{1}{n} H_{1/(1+\rho)}^\zeta(X^n|Y^n).$$

(246)

As $\delta > 0$ is arbitrary, letting $\delta \downarrow 0$, we have $E_s \geq \rho H_{1/(1+\rho)}^\zeta(X|Y).$

ACKNOWLEDGMENT

The author would like to thank anonymous reviewers for their valuable comments, which helped in improving the presentation of this article.

REFERENCES

[1] J. L. Massey, “Guessing and entropy,” in Proc. IEEE Int. Symp. Inf. Theory, Jun./Jul., 1994, p. 204.

[2] E. Arian, “An inequality on guessing and its application to sequential decoding,” IEEE Trans. Inf. Theory, vol. 42, no. 1, pp. 99–105, Jan. 1996.

[3] A. Rényi, “On measures of entropy and information,” in Proc. 4th Berkeley Symp. Math. Statist. Probab., 1961, pp. 547–561.

[4] S. Arimoto, “Information measures and capacity of order a for discrete memoryless channels,” in Proc. 2nd Colloq. Topics Inf. Theory, Keszthely, Hungary, vol. 16, 1975, pp. 41–52.

[5] R. Renner and S. Wolf, “Smooth Rényi entropy and applications,” in Proc. IEEE ISIT, Jun./Jul., 2004, p. 232.

[6] R. Renner and S. Wolf, “Simple and tight bounds for information reconciliation and privacy amplification,” in Advances in Cryptology—ASIACRYPT 2005. Berlin, Germany: Springer, 2005, pp. 199–216.

[7] T. S. Han, Information-Spectrum Method in Information Theory. New York, NY, USA: Springer-Verlag, 2002.

[8] S. Watanabe, S. Kuzuoka, and Y. F. Tan, “Nonasymptotic and second-order achievability bounds for coding with side-information,” IEEE Trans. Inf. Theory, vol. 61, no. 4, pp. 1574–1605, Apr. 2015.

[9] H. Koga, “Characterization of the smooth Rényi entropy using majorization,” in Proc. IEEE Inf. Theory Workshop, Seville, Spain, Sep. 2013, pp. 604–608.

[10] 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, 2010.

[11] T. Uyematsu and S. Kunimatsu, “A new unified method for intrinsic randomness problems of general sources,” in Proc. IEEE Inf. Theory Workshop (ITW), Sep. 2013, pp. 624–628.

[12] N. Datta and R. Renner, “Smooth entropies and the quantum information spectrum;” IEEE Trans. Inf. Theory, vol. 55, no. 6, pp. 2807–2815, Jun. 2009.
[13] N. A. Warsi, “One-shot bounds for various information theoretic problems using smooth min and max Rényi divergences,” in Proc. IEEE Inf. Theory Workshop (ITW), Sep. 2013, pp. 1–5.

[14] L. Wang, R. Colbeck, and R. Renner, “Simple channel coding bounds,” in Proc. IEEE Int. Symp. Inf. Theory (ISIT), Jun./Jul. 2009, pp. 1804–1808.

[15] I. Sason and S. Verdú, “Improved bounds on lossless source coding and guessing moments via Rényi measures,” IEEE Trans. Inf. Theory, vol. 64, no. 6, pp. 4323–4346, Jun. 2018.

[16] E. Arikan and N. Merhav, “Guessing subject to distortion,” IEEE Trans. Inf. Theory, vol. 44, no. 3, pp. 1041–1056, May 1998.

[17] S. Saito and T. Matsushima, “Non-asymptotic fundamental limits of guessing subject to distortion,” Aug. 2018, arXiv:1808.06190. [Online]. Available: https://arxiv.org/abs/1808.06190.

[18] M. K. Hanawal and R. Sundaresan, “Guessing revisited: A large deviations approach,” IEEE Trans. Inf. Theory, vol. 57, no. 1, pp. 70–78, Jan. 2011.

[19] M. M. Christiansen and K. R. Duffy, “Guesswork, large deviations, and Shannon entropy,” IEEE Trans. Inf. Theory, vol. 59, no. 2, pp. 796–802, Feb. 2013.

[20] M. M. Christiansen, K. R. Duffy, F. du Pin Calmon, and M. Médard, “Multi-user guesswork and brute force security,” IEEE Trans. Inf. Theory, vol. 61, no. 12, pp. 6876–6886, Dec. 2015.

[21] L. L. Campbell, “A coding theorem and Rényi’s entropy,” Inf. Control, vol. 8, no. 4, pp. 423–429, Aug. 1965.

[22] N. Merhav, “On optimum strategies for minimizing the exponential moments of a loss function,” Commun. Inf. Syst., vol. 11, no. 4, pp. 343–368, 2011.

[23] H. Koga and H. Yamamoto, “Asymptotic properties on codeword lengths of an optimal FV code for general sources,” IEEE Trans. Inf. Theory, vol. 51, no. 4, pp. 1546–1555, Apr. 2005.

[24] V. Kostina, Y. Polyanskiy, and S. Verdú, “Variable-length compression allowing errors,” in Proc. IEEE Int. Symp. Inf. Theory (ISIT), Jun./Jul. 2014, pp. 2679–2683.

[25] V. Kostina, Y. Polyanskiy, and S. Verdú, “Variable-length compression allowing errors,” Feb. 2014, arXiv:1402.0608. [Online]. Available: https://arxiv.org/abs/1402.0608.

[26] S. Kuzuoka, “On the smooth Rényi entropy and variable-length source coding allowing errors,” in Proc. IEEE Int. Symp. Inf. Theory (ISIT), Barcelona, Spain, Jul. 2016, pp. 745–749. [Online]. Available: http://arxiv.org/abs/1512.06499.

[27] A. Beirami, A. R. Calderbank, M. M. Christiansen, K. Duffy, and M. Médard, “A characterization of guesswork on swiftly tilting curves,” Jan. 2018, arXiv:1801.09021. [Online]. Available: https://arxiv.org/abs/1801.09021.

[28] T. M. Cover and J. A. Thomas, Elements of Information Theory, 2nd ed. Hoboken, NJ, USA: Wiley, 2006.

[29] C. E. Shannon, “The zero error capacity of a noisy channel,” IRE Trans. Inf. Theory, vol. 2, no. 3, pp. 8–19, Sep. 1956.

[30] A. W. Marshall, I. Olkin, and B. C. Arnold, Inequalities: Theory of Majorization and Its Applications, 2nd ed. New York, NY, USA: Springer-Verlag, 2010.

Shigeaki Kuzuoka (S’05–M’07) received the B.E., M.E., and Ph.D. degrees from Tokyo Institute of Technology in 2002, 2004, and 2007 respectively. He was an assistant professor from 2007 to 2009, a lecturer from 2009 to 2015, and has been an associate professor since 2015 in the Faculty of Systems Engineering at Wakayama University. His current research interests are in the areas of information theory, particularly Shannon theory, source coding, and multi-terminal information theory.