Source Resolvability and Intrinsic Randomness: Two Random Number Generation Problems With Respect to a Subclass of f-Divergences

Ryo Nomura, *Member, IEEE*

***Abstract*— This paper deals with two typical random number generation problems in information theory. One is the source resolvability problem (resolvability problem for short) and the other is the intrinsic randomness problem. In the literature, optimum achievable rates in these two problems with respect to the variational distance as well as the Kullback-Leibler (KL) divergence have already been analyzed. On the other hand, in this study we consider these two problems with respect to f-divergences. The f-divergence is a general non-negative measure between two probabilistic distributions on the basis of a convex function f. The class of f-divergences includes several important measures such as the variational distance, the KL divergence, the Hellinger distance and so on. Hence, it is meaningful to consider the random number generation problems with respect to f-divergences. In this paper, we impose some conditions on the function f so as to simplify the analysis, that is, we consider a subclass of f-divergences. Then, we first derive general formulas of the first-order optimum achievable rates with respect to f-divergences. Next, we particularize our general formulas to several specified functions f. As a result, we reveal that it is easy to derive optimum achievable rates for several important measures from our general formulas. The second-order optimum achievable rates and optimistic optimum achievable rates have also been investigated.**

***Index Terms*— f-divergence, general source, intrinsic randomness, Kullback-Leibler divergence, information-spectrum methods, source resolvability, variational distance.**

I. Introduction

IN THIS paper, we consider two typical fixed-length random number generation problems: the source resolvability problem (the resolvability problem for short) and the intrinsic randomness problem. The *resolvability* problem is formulated as follows. Given an arbitrary general source $\mathbf{X} = \{X^n\}_{n=1}^{\infty}$ (the *target* random number), we approximate it by using a discrete *uniform* random number whose size is requested to be as small as possible. A degree of approximation is measured by several criteria. Han and Verdú [2], and Steinberg and Verdú [3] have determined the *first-order* optimum achievable rates for general sources with respect to the variational distance and the *normalized* Kullback-Leibler (KL) divergence. Nomura [4] has studied the *first-order* optimum achievable rates for general sources with respect to the *unnormalized* KL divergence. Uyematsu [5] has characterized the *first-order* optimum achievable rates with respect to the variational distance by using the smooth Rényi entropy. It should be emphasized that a close relation to the fixed-length source coding problem has been reported with respect to each criterion [2]–[4].

The *second-order* optimum achievable rates in the resolvability problem have also been studied by several researchers. Nomura and Han [6] have determined the *second-order* optimum achievable rates for general sources and computed it explicitly for mixed sources. They have considered the resolvability problem with respect to the *variational distance*, while Nomura [4] has shown the *second-order* optimum achievable rates with the KL divergence.

On the other hand, the *intrinsic randomness* problem, which is also one of typical random number generation problems, has also been studied. The intrinsic randomness problem is formulated as follows. By using a given arbitrary general source $\mathbf{X} = \{X^n\}_{n=1}^{\infty}$ (the *coin* random number), we approximate a discrete *uniform* random number whose size is requested to be as large as possible. Also in the intrinsic randomness problem, the optimum achievable rates with respect to several criteria have been shown. Vembu and Verdú [7] have considered the intrinsic randomness problem with respect to the variational distance as well as the normalized KL divergence and derived *general formulas* of the *first-order* optimum achievable rates (cf. Han [8]). Uyematsu and Kunimatsu [9] have characterized the *first-order* optimum achievable rates with respect to the variational distance by using the smooth Rényi entropy. Hayashi [10] has considered the *first-* and *second-order* optimum achievable rates with respect to the *unnormalized* KL divergence.

Related works include works given by Liu *et al.* [11], Yagi and Han [12], Kumagai and Hayashi [13], [14], and Yu and Tan [15]. In [11], the *channel* resolvability problem with respect to the E_γ-divergence has been considered. They have also particularized their results in the case of the *source* resolvability problem. Yagi and Han [16] have determined the optimum *variable-length* resolvability rates with respect to the variational distance as well as the KL divergence. Their results are based on the smooth Rényi entropy. Kumagai and Hayashi [13], [14] have determined the *first- and second- order* optimum achievable rates for stationary memoryless sources

Manuscript received August 1, 2019; revised March 21, 2020; accepted July 6, 2020. Date of publication July 14, 2020; date of current version November 20, 2020. This work was supported in part by JSPS KAKENHI under Grant JP18K04150 and in part by the Waseda University Grant for Special Research Projects under Project 2020C-528. This article was presented in part at the 2019 IEEE International Symposium on Information Theory.

The author is with the Center for Data Science, Waseda University, Tokyo 169-8050, Japan (e-mail: nomu@waseda.jp).

Communicated by F. Alajaji, Associate Editor for Shannon Theory.
Digital Object Identifier 10.1109/TIT.2020.3009208

in the *random number conversion* problem, which includes the resolvability and intrinsic randomness problems considered in this paper. In [13] and [14], an approximation measure related to the Hellinger distance has been used. Yu and Tan [15] have also considered the *random number conversion* problem with respect to the Rényi divergence.

As we have mentioned above, several approximation measures have been employed in the resolvability problem and the intrinsic randomness problem. In this paper, we focus on the f-divergence as a measure of approximation. The f-divergence is a general distance measure on the basis of a convex function f [17], [18]. The class of f-divergences includes several important measures such as the total variational distance, the KL-divergence, the Hellinger distance and so on. Hence, it is meaningful to consider these two problems with respect to f-divergences. In order to tackle these problems, we first impose some conditions on the function f. Then, in both of these two problems we derive *general formulas* of the *first- and second-order* optimum achievable rates. It should be emphasized that the subclass of f-divergences considered in this paper includes the half variational distance, the reverse KL-divergence, the Hellinger distance, and the E_γ-divergence. One of main contributions of the present paper is to provide the unified viewpoint to the analysis in the resolvability problem (or the intrinsic randomness problem) with respect to the subclass of f-divergences. In previous results, the analysis of the optimum achievable rate in the resolvability problem (or the intrinsic randomness problem) has been relied on the specified approximation measure. On the other hand, our analysis does not depend on the specified approximation measure. This is an advantage to consider the class of f-divergences. It will turn out that we can easily derive optimum achievable rates with respect specified measures from our general formulas. As a result, we establish the *general formulas* of the first- and second-order optimum achievable rates for several important measures that have not been considered yet. This is also one of significant contributions of this paper.

This paper is organized as follows. In Section II, we describe the problem setting and give some definitions of the optimum *first*-order achievable rates. The subclass of f-divergences considered in this paper has also been introduced. In Section III and IV, we show *general formulas* of the optimum *first-order* achievable rates in the resolvability problem and the intrinsic randomness problem, respectively. In Section V, we apply general formulas to some specified functions f and compute the optimum *first-order* achievable rates in each case. In Section VI, we show *general* formulas of the optimum *second*-order achievable rates. In Section VII, we discuss the resolvability and intrinsic randomness problems in the optimistic scenario. In Section VIII, we clarify the relationship to the fixed length source coding problem. Finally, we provide some concluding remarks on our results in Section IX.

II. Preliminaries

We consider the *general* source defined as an infinite sequence $\mathbf{X} = \left\{X^n = \left(X_1^{(n)}, X_2^{(n)}, \ldots, X_n^{(n)}\right)\right\}_{n=1}^{\infty}$ of n-dimensional random variables X^n, where each component random variable $X_i^{(n)}$ takes values in a countable set $\mathcal{X}$. Let $P_X(\cdot)$ denote the probability distribution of the random variable X. The uniform random number U_M is defined by

$$P_{U_M}(i) = \frac{1}{M}, \quad i \in \mathcal{U}_M := \{1, 2, \cdots, M\}. \tag{1}$$

The f-divergence between two probabilistic distributions P_Z and $P_{\overline{Z}}$ is defined as follows [17]. Let $f(t)$ be a convex function defined for $t > 0$ and $f(1) = 0$.

Definition 2.1 (f-Divergence [17]): Let P_Z and $P_{\overline{Z}}$ denote probability distributions over a finite or countably infinite set $\mathcal{Z}$. The f-divergence between P_Z and $P_{\overline{Z}}$ is defined by

$$D_f(Z||\overline{Z}) := \sum_{z\in\mathcal{Z}} P_{\overline{Z}}(z) f\left(\frac{P_Z(z)}{P_{\overline{Z}}(z)}\right), \tag{2}$$

where we set $0\ f\left(\frac{0}{0}\right) = 0$, $f(0) = \lim_{t\to 0} f(t)$, $0f(\frac{a}{0}) = \lim_{t\to 0} tf(\frac{a}{t}) = a \lim_{u\to\infty} \frac{f(u)}{u}$.

We give some examples of f-divergences [17], [18]:

- $f(t) = t\log t$: (Kullback-Leibler (KL) divergence)
$$D_f(Z||\overline{Z}) = \sum_{z\in\mathcal{Z}} P_Z(z) \log \frac{P_Z(z)}{P_{\overline{Z}}(z)} =: D(Z||\overline{Z}). \tag{3}$$
- $f(t) = -\log t$: (Reverse Kullback-Leibler divergence)
$$D_f(Z||\overline{Z}) = \sum_{z\in\mathcal{Z}} P_{\overline{Z}}(z) \log \frac{P_{\overline{Z}}(z)}{P_Z(z)} = D(\overline{Z}||Z). \tag{4}$$
- $f(t) = (t-1)^2$: (χ^2-divergence)
$$D_f(Z||\overline{Z}) = \sum_{z\in\mathcal{Z}} \frac{(P_Z(z) - P_{\overline{Z}}(z))^2}{P_{\overline{Z}}(z)}. \tag{5}$$
- $f(t) = 1 - \sqrt{t}$: (Hellinger distance)
$$D_f(Z||\overline{Z}) = 1 - \sum_{z\in\mathcal{Z}} \sqrt{P_Z(z) P_{\overline{Z}}(z)}. \tag{6}$$
- $f(t) = |t-1|$: (Variational distance)
$$D_f(Z||\overline{Z}) = \sum_{z\in\mathcal{Z}} |P_Z(z) - P_{\overline{Z}}(z)|. \tag{7}$$
- $f(t) = (1-t)^+ := \max\{1-t, 0\}$: (Half variational distance)
$$\begin{aligned} D_f(Z||\overline{Z}) &= \frac{1}{2}\sum_{z\in\mathcal{Z}} |P_Z(z) - P_{\overline{Z}}(z)| \\ &= \sum_{z\in\mathcal{Z}: P_Z(z) > P_{\overline{Z}}(z)} \left(P_Z(z) - P_{\overline{Z}}(z)\right). \end{aligned} \tag{8}$$
- $f(t) = (t-\gamma)^+$: (E_γ-divergence) For any given $\gamma \geq 1$,
$$\begin{aligned} D_f(Z||\overline{Z}) &= \sum_{z\in\mathcal{Z}: P_Z(z) > \gamma P_{\overline{Z}}(z)} \left(P_Z(z) - \gamma P_{\overline{Z}}(z)\right) \\ &=: E_\gamma(Z||\overline{Z}). \end{aligned} \tag{9}$$

The E_γ-divergence is a generalization of the half variational distance defined in (8), because $\gamma \geq 1$ is arbitrarily.

Remark 2.1: Since the relation

$$\sum_{z\in\mathcal{Z}} |P_Z(z) - P_{\overline{Z}}(z)| = 2 \sum_{z\in\mathcal{Z}: P_Z(z) > P_{\overline{Z}}(z)} \left(P_Z(z) - P_{\overline{Z}}(z)\right) \tag{10}$$

holds, the variational distance is expressed as an f-divergence using the function $f(t) = 2(1-t)^+$.

Remark 2.2: The E_γ-divergence between two probabilistic distributions P_Z and $P_{\overline{Z}}$ is defined as (9) [11], [18]. Since a representation of f-divergence is unique up to an additive term which is a constant multiple of $(t-1)$, the function which represents E_γ-divergence can be given as

$$\begin{aligned} f(t) &= (t-\gamma)^+ + (1-t) \\ &= (\gamma - t)^+ + (1-\gamma). \end{aligned} \tag{11}$$

This yields alternative expression of the E_γ-divergence as an f-divergence using the function:

$$f(t) = (\gamma - t)^+ + 1 - \gamma. \tag{12}$$

The following key property holds for the f-divergence from Jensen's inequality [17]:

$$\sum_{z \in \mathcal{Z}'} b(z) f\left(\frac{a(z)}{b(z)}\right) \geq \left(\sum_{z\in\mathcal{Z}'} b(z)\right) f\left(\frac{\sum_{z\in\mathcal{Z}'} a(z)}{\sum_{z\in\mathcal{Z}'} b(z)}\right). \tag{13}$$

Thus, together with the fact that $f(1) = 0$, we immediately have

$$D_f(Z||\overline{Z}) \geq 0. \tag{14}$$

As we have mentioned above, the f-divergence is a general distortion measure, which includes important measures. In this study, we assume the following conditions on the function f.

C1) The function $f(t)$ is a monotonically decreasing function of t. That is, for any pair of positive real numbers (a, b) satisfying $a < b$ it holds that

$$f(a) \geq f(b). \tag{15}$$

C2)

$$\lim_{u\to\infty} \frac{f(u)}{u} = 0. \tag{16}$$

C3) For any pair of positive real numbers (a, b), it holds that

$$\lim_{n\to\infty} \frac{f\left(e^{-nb}\right)}{e^{na}} = 0. \tag{17}$$

Remark 2.3: Notice here that functions $f(t) = -\log t$, $f(t) = 1 - \sqrt{t}$, and $f(t) = (1-t)^+$ satisfy the above conditions, while $f(t) = t\log t$ does not satisfy conditions C1) and C2). Moreover, it is not difficult to check that (12) satisfies these conditions.

Remark 2.4: From the definition of the f-divergence, C2) means

$$0 f\left(\frac{a}{0}\right) = 0, \tag{18}$$

for any $a \in (0, 1]$.

III. Source Resolvability Problem

First, we consider the *general formula* of the first-order optimum D-achievable rate in the resolvability problem. Let us begin with the definition of the achievability.

Definition 3.1: R is said to be D-achievable with the given f-divergence if there exists a sequence of mapping $\phi_n : \mathcal{U}_{M_n} \to \mathcal{X}^n$ such that

$$\limsup_{n\to\infty} D_f(X^n||\phi_n(U_{M_n})) \leq D, \tag{19}$$

$$\limsup_{n\to\infty} \frac{1}{n}\log M_n \leq R. \tag{20}$$

Definition 3.2 (First-Order Optimum Resolvability Rate):

$$S_r^{(f)}(D|\mathbf{X}) := \inf\{R\,|R \text{ is } D\text{-achievable with the given } f\text{-divergence}\}. \tag{21}$$

Remark 3.1: It should be noted that we does not employ $D_f(\phi_n(U_{M_n})||X^n)$ but $D_f(X^n||\phi_n(U_{M_n}))$ as the one of conditions in Def. 3.1. This is important to consider the *asymmetric* measure such as the KL-divergence.

Remark 3.2: In this paper, we only consider the case that D is in $[0, f(0))$ under the given f-divergence. This is because $D \geq f(0)$ means that there exists no restriction about the approximation error (for example, $f(0) = 1$ in the case of the half variational distance and $f(0) = \infty$ in the case of the KL divergence). This case leads the trivial result that the first-order optimum resolvability rate equals to 0. Hence, we focus on the case of $D \in [0, f(0))$. This remark is applicable throughout the paper.

In order to show the general formula of the *first-order* optimum resolvability rate, we introduce the information quantity on the basis of the function f given $0 \leq \varepsilon < f(0)$:

$$\overline{K}_f(\varepsilon|\mathbf{X}) := \inf\left\{R\left|\limsup_{n\to\infty} f\left(\Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \leq R\right\}\right) \leq \varepsilon\right.\right\}. \tag{22}$$

Remark 3.3: Because of the condition C1), the function $f\left(\Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \leq R\right\}\right)$ monotonically increases as R decreases. Hence, the above quantity is uniquely determined given ε.

Remark 3.4: From the condition C1) and the definition that $f(1) = 0$, it is not difficult to verify that

$$\overline{K}_f(\varepsilon|\mathbf{X}) \leq \overline{K}_f(0|\mathbf{X}) \leq \overline{H}(\mathbf{X}), \tag{23}$$

for any $0 \leq \varepsilon < f(0)$, where

$$\overline{H}(\mathbf{X}) := \inf\left\{R\left|\lim_{n\to\infty}\Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} > R\right\} = 0\right.\right\}, \tag{24}$$

is called the *spectral sup-entropy rate* of the source $\mathbf{X}$ [8]. Furthermore, if $\min\{a|f(a) = 0\} = 1$ holds, then it holds that

$$\overline{K}_f(0|\mathbf{X}) = \overline{H}(\mathbf{X}). \tag{25}$$

The following theorem addresses the general formula of the *first-order* optimum resolvability.

Theorem 3.1 (First-Order Optimum Resolvability Rate): Assuming that the function f satisfies conditions C1)–C3), then for any $0 \le D < f(0)$ it holds that

$$S_r^{(f)}(D|\mathbf{X}) = \overline{K}_f(D|\mathbf{X}). \tag{26}$$

Remark 3.5: This theorem shows that the first-order optimum resolvability rate with the given f-divergence does not depend on the behavior of the function $f(t)$ in $t > 1$. This means that for two functions f and g, if the restriction of f to $[0,1]$ is the same as that of g, then they will lead the same first-order optimum resolvability rate, that is, $S_r^{(f)}(D|\mathbf{X}) = S_r^{(g)}(D|\mathbf{X})$.

For example, if we set

$$f(t) = -\log t, \tag{27}$$

$$g(t) = (-\log t)^+ := \max\{-\log t, 0\}. \tag{28}$$

Then, it holds that

$$D_f(X^n||\phi_n(U_{M_n})) = \sum_{\mathbf{x}\in\mathcal{X}^n} P_{\tilde{X}^n}(\mathbf{x}) \log \frac{P_{\tilde{X}^n}(\mathbf{x})}{P_{X^n}(\mathbf{x})}, \tag{29}$$

$$D_g(X^n||\phi_n(U_{M_n})) = \sum_{\mathbf{x}\in\mathcal{X}^n : P_{\tilde{X}^n}(\mathbf{x}) > P_{X^n}(\mathbf{x})} P_{\tilde{X}^n}(\mathbf{x}) \log \frac{P_{\tilde{X}^n}(\mathbf{x})}{P_{X^n}(\mathbf{x})}, \tag{30}$$

where we denote $\tilde{X}^n = \phi_n(U_{M_n})$. Theorem 3.1 means that $S_r^{(f)}(D|\mathbf{X}) = S_r^{(g)}(D|\mathbf{X})$.

Proof of Theorem 3.1: The proof consists of two parts:

(Direct Part:) Letting $R_0 = \overline{K}_f(D|\mathbf{X})$, we shall show that $R = (R_0 + 2\gamma)$ is D-achievable with the given f-divergence for any $\gamma > 0$. To do so, first we construct the mapping $\phi_n : \mathcal{U}_{M_n} \to \mathcal{X}^n$.

Let $M_n = e^{nR} = e^{n(R_0+2\gamma)}$. We define the set S_n as follows

$$S_n := \left\{ \mathbf{x} \in \mathcal{X}^n \left| \frac{1}{n} \log \frac{1}{P_{X^n}(\mathbf{x})} \le R_0 + \gamma \right. \right\}. \tag{31}$$

Index the elements in S_n as $\{\mathbf{x}_1, \mathbf{x}_2, \dots \mathbf{x}_{|S_n|}\}$. Since,

$$1 \ge \sum_{\mathbf{x}\in S_n} P_{X^n}(\mathbf{x}) \ge |S_n| \, e^{-n(R_0+\gamma)}, \tag{32}$$

holds, we have

$$|S_n| \le e^{n(R_0+\gamma)}. \tag{33}$$

Furthermore, let $P_{\overline{X}^n}$ denote the probability distribution over S_n defined by

$$P_{\overline{X}^n}(\mathbf{x}) = \begin{cases} \frac{P_{X^n}(\mathbf{x})}{\Pr\{X^n \in S_n\}} & \mathbf{x} \in S_n, \\ 0 & \text{otherwise.} \end{cases} \tag{34}$$

For $\mathbf{x}_1$ set $k_0 = 0$ and determine k_1 such that

$$\frac{k_1}{M_n} \le P_{\overline{X}^n}(\mathbf{x}_1), \quad \frac{k_1+1}{M_n} > P_{\overline{X}^n}(\mathbf{x}_1). \tag{35}$$

Secondly, for $\mathbf{x}_2$ we determine k_2 such that

$$\frac{k_2-k_1}{M_n} \le P_{\overline{X}^n}(\mathbf{x}_2), \quad \frac{k_2-k_1+1}{M_n} > P_{\overline{X}^n}(\mathbf{x}_2). \tag{36}$$

In the similar way, we repeat this operation to choose k_i for $\mathbf{x}_i$ as long as possible. Suppose that this operation stops at $\mathbf{x}_{i_0}$. Because of the construction of the mapping, $i_0 = |S_n|$ holds.

Now, we define the mapping $\phi_n : \mathcal{U}_{M_n} \to \mathcal{X}^n$ by using i_0 and $k_i (1 \le i \le i_0)$ as follows

$$\phi_n(j) = \begin{cases} \mathbf{x}_i & k_{i-1}+1 \le j \le k_i, i < i_0 \\ \mathbf{x}_{i_0} & \text{otherwise} \end{cases} \tag{37}$$

and set $\tilde{X}^n = \phi_n(U_{M_n})$.

Next, we evaluate the performance of this mapping ϕ_n. From the setting of M_n, clearly we have

$$\limsup_{n\to\infty} \frac{1}{n} \log M_n \le R. \tag{38}$$

Thus, it suffices to show that

$$\limsup_{n\to\infty} D_f(X^n||\phi_n(U_{M_n})) \le D. \tag{39}$$

From the construction of the code, for all i satisfying $1 \le i \le i_0 - 1$ it holds that

$$P_{\tilde{X}^n}(\mathbf{x}_i) \le P_{\overline{X}^n}(\mathbf{x}_i) \tag{40}$$

and

$$P_{\overline{X}^n}(\mathbf{x}_i) - P_{\tilde{X}^n}(\mathbf{x}_i) \le \frac{1}{M_n}. \tag{41}$$

Furthermore, $P_{\tilde{X}^n}(\mathbf{x}_{i_0})$ can be evaluated as follows. From (41), it holds that

$$\begin{aligned} &P_{\tilde{X}^n}(\mathbf{x}_{i_0}) - P_{\overline{X}^n}(\mathbf{x}_{i_0}) \\ &= \left(1 - \sum_{i=1}^{i_0-1} P_{\tilde{X}^n}(\mathbf{x}_i)\right) - \left(1 - \sum_{i=1}^{i_0-1} P_{\overline{X}^n}(\mathbf{x}_i)\right) \\ &= \sum_{i=1}^{i_0-1} P_{\overline{X}^n}(\mathbf{x}_i) - \sum_{i=1}^{i_0-1} P_{\tilde{X}^n}(\mathbf{x}_i) \\ &= \sum_{i=1}^{i_0-1} (P_{\overline{X}^n}(\mathbf{x}_i) - P_{\tilde{X}^n}(\mathbf{x}_i)) \\ &\le \frac{|S_n|}{M_n} \\ &\le e^{-n\gamma}. \end{aligned} \tag{42}$$

Thus, noting that the condition C2) and Remark 2.4, the f-divergence between $P_{\tilde{X}^n}$ and P_{X^n} is evaluated as follows:

$$\begin{aligned} &D_f\left(X^n||\phi_n(U_{M_n})\right) \\ &= \sum_{i=1}^{i_0} P_{\tilde{X}^n}(\mathbf{x}_i) f\left(\frac{P_{X^n}(\mathbf{x}_i)}{P_{\tilde{X}^n}(\mathbf{x}_i)}\right) \\ &= \sum_{i=1}^{i_0} P_{\tilde{X}^n}(\mathbf{x}_i) f\left(\frac{P_{\overline{X}^n}(\mathbf{x}_i)\Pr\{X^n \in S_n\}}{P_{\tilde{X}^n}(\mathbf{x}_i)}\right) \\ &\le \sum_{i=1}^{i_0-1} P_{\tilde{X}^n}(\mathbf{x}_i) f\left(\frac{P_{\tilde{X}^n}(\mathbf{x}_i)\Pr\{X^n \in S_n\}}{P_{\tilde{X}^n}(\mathbf{x}_i)}\right) \\ &\quad + P_{\tilde{X}^n}(\mathbf{x}_{i_0}) f\left(\frac{P_{\overline{X}^n}(\mathbf{x}_{i_0})\Pr\{X^n \in S_n\}}{P_{\tilde{X}^n}(\mathbf{x}_{i_0})}\right) \\ &\le \sum_{i=1}^{i_0-1} P_{\tilde{X}^n}(\mathbf{x}_i) f\left(\Pr\{X^n \in S_n\}\right) \\ &\quad + P_{\tilde{X}^n}(\mathbf{x}_{i_0}) f\left(\frac{P_{\overline{X}^n}(\mathbf{x}_{i_0})\Pr\{X^n \in S_n\}}{P_{\overline{X}^n}(\mathbf{x}_{i_0}) + e^{-n\gamma}}\right), \end{aligned} \tag{43}$$

where the second equality is due to (34), the first and the last inequality is due to C1) and (42), respectively.

Here, let the second term on the right hand side of (43) denote A_n for short. Then, we can show that

$$A_n \leq P_{\tilde{X}^n}(\mathbf{x}_{i_0}) f\left(\Pr\left\{X^n \in S_n\right\}\right) + o(1). \tag{44}$$

The proof appears in the appendix.

Hence, we have

$$\limsup_{n\to\infty} D_f\left(X^n || \phi_n(U_{M_n})\right) \leq \limsup_{n\to\infty} f(\Pr\left\{X^n \in S_n\right\}). \tag{45}$$

Here, from the definition of R and S_n we obtain

$$\begin{aligned}&\limsup_{n\to\infty} f\left(\Pr\left\{X^n \in S_n\right\}\right)\\ &= \limsup_{n\to\infty} f\left(\Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \leq R_0 + \gamma\right\}\right)\\ &\leq D,\end{aligned} \tag{46}$$

which completes the proof of the Direct Part.

(Converse Part:) In the proof of this part, we do not use C3). Suppose that R is D-achievable with the given f-divergence, then there exists a mapping ϕ_n such that

$$\limsup_{n\to\infty} D_f(X^n || \phi_n(U_{M_n})) \leq D, \tag{47}$$

$$\limsup_{n\to\infty} \frac{1}{n}\log M_n \leq R. \tag{48}$$

We fix this mapping ϕ_n and define the probability distribution $P_{\tilde{X}^n}$ by $\tilde{X}^n = \phi_n(U_{M_n})$. Define the set S_n' as

$$S_n' := \left\{\mathbf{x} \in \mathcal{X}^n \left| \frac{1}{n}\log\frac{1}{P_{X^n}(\mathbf{x})} \leq R + 2\gamma\right.\right\}. \tag{49}$$

For any set S, let S^c denote its complement. Since for $\forall \mathbf{x} \in (S_n')^c$ it holds that

$$\frac{1}{n}\log\frac{1}{P_{X^n}(\mathbf{x})} > R + 2\gamma, \tag{50}$$

we obtain

$$P_{X^n}(\mathbf{x}) < e^{-n(R+2\gamma)} \quad (\forall \mathbf{x} \in (S_n')^c). \tag{51}$$

On the other hand, (48) means that $M_n \leq e^{n(R+\gamma)}$ holds for sufficiently large n. Here, define the set

$$B_n := \left\{\mathbf{x} \in \mathcal{X}^n | P_{\tilde{X}^n}(\mathbf{x}) > 0\right\}, \tag{52}$$

and index the element of B_n as $B_n = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{|B_n|}\}$. Then, from the property of the mapping ϕ_n, we obtain

$$|B_n| \leq M_n \leq e^{n(R+\gamma)}, \tag{53}$$

for sufficiently large n. Thus, from the condition C2) we obtain

$$\begin{aligned}&D_f\left(X^n || \phi_n(U_{M_n})\right)\\ &= \sum_{\mathbf{x}\in B_n} P_{\tilde{X}^n}(\mathbf{x}) f\left(\frac{P_{X^n}(\mathbf{x})}{P_{\tilde{X}^n}(\mathbf{x})}\right)\\ &\geq f\left(\Pr\left\{X^n \in B_n \cap S_n'\right\} + \Pr\left\{X^n \in B_n \cap (S_n')^c\right\}\right)\\ &\geq f\left(\Pr\left\{X^n \in S_n'\right\} + \sum_{\mathbf{x}\in B_n\cap(S_n')^c} e^{-n(R+2\gamma)}\right)\\ &= f\left(\Pr\left\{X^n \in S_n'\right\} + |B_n \cap (S_n')^c| e^{-n(R+2\gamma)}\right)\\ &\geq f\left(\Pr\left\{X^n \in S_n'\right\} + e^{-n\gamma}\right),\end{aligned} \tag{54}$$

for sufficiently large n, where the first inequality is due to (13) and the second inequality is due to (51) and C1), and the last inequality is due to (53) and C1).

Hence, from (47) we have

$$\begin{aligned}D &\geq \limsup_{n\to\infty} D_f\left(X^n || \phi_n(U_{M_n})\right)\\ &\geq \limsup_{n\to\infty} f\left(\Pr\left\{X^n \in S_n'\right\} + e^{-n\gamma}\right)\\ &= \limsup_{n\to\infty} f\left(\Pr\left\{X^n \in S_n'\right\}\right),\end{aligned} \tag{55}$$

from the continuity of the function f. Therefore, for the D-achievable rate R it must hold that

$$\limsup_{n\to\infty} f\left(\Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \leq R+2\gamma\right\}\right) \leq D. \tag{56}$$

This inequality means that the converse part holds. □

Remark 3.6: The proof of the direct part in Theorem 3.1 shows that the way of the construction of the optimum mapping is always same irrespective of the given f-divergence in the asymptotic sense. This observation is useful for the construction of the mapping.

IV. Intrinsic Randomness Problem

In this section, we consider the intrinsic randomness problem. We first define the achievable rate in this problem.

Definition 4.1: R is said to be Δ-achievable with the given f-divergence if there exists a sequence of mapping $\varphi_n : \mathcal{X}^n \to \mathcal{U}_{M_n}$ such that

$$\limsup_{n\to\infty} D_f(\varphi_n(X^n) || U_{M_n}) \leq D, \tag{57}$$

$$\liminf_{n\to\infty} \frac{1}{n}\log M_n \geq R. \tag{58}$$

Definition 4.2 (First-Order Optimum Intrinsic Randomness Rate):

$$\begin{aligned}&S_\iota^{(f)}(\Delta|\mathbf{X})\\ &:= \sup\left\{R \,|\, R \text{ is } \Delta\text{-achievable with the given } f\text{-divergence}\right\}.\end{aligned} \tag{59}$$

Remark 4.1: In this paper, we employ the f-divergence: $D_f(\varphi_n(X^n)||U_{M_n})$ instead of $D_f(U_{M_n}||\varphi_n(X^n))$ (cf. Remark 3.1).

In order to characterize $S_\iota^{(f)}(\Delta|\mathbf{X})$, we introduce the quantity which is an analogue of $\overline{K}_f(\varepsilon|\mathbf{X})$ defined in Section III.

$$\begin{aligned}&\underline{K}_f(\varepsilon|\mathbf{X})\\ &:= \sup\left\{R \left| \limsup_{n\to\infty} f\left(\Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \geq R\right\}\right) \leq \varepsilon\right.\right\}.\end{aligned} \tag{60}$$

Remark 4.2: From the condition C1), the function

$$f\left(\Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \geq R\right\}\right) \tag{61}$$

is a monotonically increasing function of R. Hence, $\underline{K}_f(\varepsilon|\mathbf{X})$ is uniquely determined given ε.

Remark 4.3: From the condition C1) and the definition that $f(1) = 0$, it is not difficult to verify that

$$\underline{K}_f(\varepsilon|\mathbf{X}) \geq \underline{K}_f(0|\mathbf{X}) \geq \underline{H}(\mathbf{X}), \tag{62}$$

for any $0 \le \varepsilon < f(0)$, where

$$\underline{H}(\mathbf{X}) = \sup\left\{ R \,\middle|\, \lim_{n\to\infty} \Pr\left\{ \frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \ge R \right\} = 1 \right\}, \tag{63}$$

is called the *spectral inf-entropy rate* [8]. Furthermore, if $\min\{a | f(a) = 0\} = 1$ holds, then it holds that

$$\underline{K}_f(0|\mathbf{X}) = \underline{H}(\mathbf{X}). \tag{64}$$

Then, we have the following theorem:

Theorem 4.1 (First-Order Optimum Intrinsic Randomness Rate): Assuming that the function f satisfies C1) and C2), then for any $0 \le \Delta < f(0)$ it holds that

$$S_\iota^{(f)}(\Delta|\mathbf{X}) = \underline{K}_f(\Delta|\mathbf{X}). \tag{65}$$

Proof: The proof is a modification of the proof given by Hayashi [10, Theorem 7]. The proof consists of two parts.

(Direct Part:) Setting $R_0 = \underline{K}_f(\Delta|\mathbf{X})$, we show that $R = (R_0 - 2\gamma)$ is Δ-achievable with the given f-divergence for any $\gamma > 0$. To do so, we define the mapping $\varphi_n : \mathcal{X}^n \to \mathcal{U}_{M_n}$ as follows.

Setting

$$T_n := \left\{ \mathbf{x} \in \mathcal{X}^n \,\middle|\, \frac{1}{n}\log\frac{1}{P_{X^n}(\mathbf{x})} \ge R_0 - \gamma \right\}, \tag{66}$$

$$B_n := \{ \mathbf{x} \in \mathcal{X}^n \,| P_{X^n}(\mathbf{x}) > 0 \}, \tag{67}$$

we define

$$M_n = e^{nR}\Pr\{X^n \in T_n\} = e^{n(R_0-2\gamma)}\Pr\{X^n \in T_n\}. \tag{68}$$

Since, for $\forall \mathbf{x} \in T_n$ it holds that

$$P_{X^n}(\mathbf{x}) \le e^{-n(R_0-\gamma)}. \tag{69}$$

we have

$$\begin{aligned} |T_n \cap B_n| &\ge \frac{\Pr\{X^n \in T_n\}}{e^{-n(R_0-\gamma)}} \\ &\ge \Pr\{X^n \in T_n\}e^{n(R_0-\gamma)} \\ &> M_n. \end{aligned} \tag{70}$$

Thus, from the definition of T_n, for any i $(1 \le i \le M_n)$, there exists a surjective mapping $\overline{\phi}_n : T_n \cap B_n \to \mathcal{U}_{M_n}$ such that:

$$\begin{aligned} \frac{P_{\overline{\phi}_n(X^n)}(i)}{\Pr\{X^n \in T_n\}} &> \frac{1}{M_n} - \frac{e^{-n(R_0-\gamma)}}{\Pr\{X^n \in T_n\}} \\ &= \frac{1}{M_n}\left(1 - e^{-n\gamma}\right). \end{aligned} \tag{71}$$

By using this mapping $\overline{\phi}_n$, we construct the mapping $\varphi_n : \mathcal{X}^n \to \mathcal{U}_{M_n}$ such that

$$\varphi_n(\mathbf{x}) = \begin{cases} \overline{\phi}_n(\mathbf{x}) & \mathbf{x} \in T_n \cap B_n \\ 1 & \text{otherwise,} \end{cases} \tag{72}$$

and set $\tilde{U}_{M_n} = \varphi_n(X^n)$.

We next evaluate the performance of φ_n. From the definition of M_n, we clearly obtain

$$\begin{aligned} \frac{1}{n}\log M_n &\ge R_0 - 2\gamma + \frac{1}{n}\log\Pr\{X^n \in T_n\} \\ &\ge R_0 - 3\gamma, \end{aligned} \tag{73}$$

for sufficiently large n.[1] Hence, in order to prove the direct part it suffices to show

$$\begin{aligned} \limsup_{n\to\infty} D_f(\varphi_n(X^n)||U_{M_n}) &= \limsup_{n\to\infty} D_f(\tilde{U}_{M_n}||U_{M_n}) \\ &\le \Delta. \end{aligned} \tag{74}$$

From the construction of $\tilde{U}_{M_n}$ and (71), it holds that

$$P_{\tilde{U}_{M_n}}(i) \ge \frac{1}{M_n}\left(1 - e^{-n\gamma}\right)\Pr\{X^n \in T_n\}. \tag{75}$$

Then, from (75) and the monotonicity of the function f (C1)), the f-divergence is upper-bounded by

$$\begin{aligned} &D_f(\tilde{U}_{M_n}||U_{M_n}) \\ &= \sum_{i=1}^{M_n} P_{U_{M_n}}(i) f\left(\frac{P_{\tilde{U}_{M_n}}(i)}{P_{U_{M_n}}(i)}\right) \\ &\le \sum_{i=1}^{M_n} \frac{1}{M_n} f\left(\frac{\frac{1}{M_n}\left(1 - e^{-n\gamma}\right)\Pr\{X^n \in T_n\}}{\frac{1}{M_n}}\right) \\ &= f\left((1 - e^{-n\gamma})\Pr\{X^n \in T_n\}\right) \\ &\le f\left(\Pr\{X^n \in T_n\} - e^{-n\gamma}\right). \end{aligned} \tag{76}$$

Since the function f is continuous, we obtain

$$\limsup_{n\to\infty} D_f(\tilde{U}_{M_n}||U_{M_n}) \le \limsup_{n\to\infty} f(\Pr\{X^n \in T_n\}). \tag{77}$$

Therefore, from the definition of R_0 we have

$$\begin{aligned} &\limsup_{n\to\infty} f\left(\Pr\{X^n \in T_n\}\right) \\ &= \limsup_{n\to\infty} f\left(\Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \ge R_0 - \gamma\right\}\right) \\ &\le \Delta, \end{aligned} \tag{78}$$

which completes the proof.

(Converse Part:) Suppose that R is Δ-achievable with the given f-divergence, then there exists a mapping φ_n satisfying

$$\limsup_{n\to\infty} D_f(\varphi_n(X^n)||U_{M_n}) \le \Delta, \tag{79}$$

$$\liminf_{n\to\infty} \frac{1}{n}\log M_n \ge R. \tag{80}$$

We fix this mapping φ_n and define the probability distribution $\tilde{U}_{M_n}$ by $\tilde{U}_{M_n} = \varphi_n(X^n)$. For any fixed $\gamma > 0$, define the set T_n' as:

$$T_n' := \left\{ \mathbf{x} \in \mathcal{X}^n \,\middle|\, \frac{1}{n}\log\frac{1}{P_{X^n}(\mathbf{x})} \ge R - 2\gamma \right\}. \tag{81}$$

Then, for $\forall \mathbf{x} \in (T_n')^c$ it holds that

$$\frac{1}{n}\log\frac{1}{P_{X^n}(\mathbf{x})} < R - 2\gamma. \tag{82}$$

Thus, we have

$$P_{X^n}(\mathbf{x}) > e^{-n(R-2\gamma)} \quad (\forall \mathbf{x} \in (T_n')^c), \tag{83}$$

from which it holds that

$$|(T_n')^c| < e^{n(R-2\gamma)}. \tag{84}$$

[1] From Remark 3.2, $\Pr\{X^n \in T_n\} > 0$ holds for sufficiently large n.

Here, define the set $\varphi_n((T'_n)^c)$ as

$$\varphi_n((T'_n)^c) := \{\varphi_n(\mathbf{x}) \,|\, \mathbf{x} \in (T'_n)^c\}. \tag{85}$$

Thus, the set $\varphi_n((T'_n)^c)$ is the set of i constructing from at least one $\mathbf{x} \in (T'_n)^c$. Then, the difference set $\mathcal{U}_{M_n} \setminus \varphi_n((T'_n)^c)$ is the set of index i constructing only from $\mathbf{x} \in T'_n$,

Then, from (84) and the definition of the mapping it holds that

$$|\varphi_n((T'_n)^c)| \le |(T'_n)^c| \le e^{n(R-2\gamma)}. \tag{86}$$

Since (80) means that $M_n \ge e^{n(R-\gamma)}$ holds for sufficiently large n, we obtain

$$\frac{|\varphi_n((T'_n)^c)|}{M_n} \le \frac{e^{n(R-2\gamma)}}{e^{n(R-\gamma)}} \le e^{-n\gamma}, \tag{87}$$

for sufficiently large n. Thus, from the above inequality and C2), we obtain

$$\lim_{n\to\infty} \frac{|\varphi_n((T'_n)^c)|}{M_n} f\left(\frac{M_n}{|\varphi_n((T'_n)^c)|}\right) = 0. \tag{88}$$

Hence, the f-divergence between $\tilde{U}_{M_n}$ and U_{M_n} is lower-bounded by

$$\begin{aligned}
&D_f\left(\tilde{U}_{M_n}||U_{M_n}\right)\\
&= \sum_{i=1}^{M_n} \frac{1}{M_n} f\left(\frac{P_{\tilde{U}_{M_n}}(i)}{\frac{1}{M_n}}\right)\\
&= \sum_{i\in\mathcal{U}_{M_n}\setminus\varphi_n((T'_n)^c)} \frac{1}{M_n} f\left(\frac{P_{\tilde{U}_{M_n}}(i)}{\frac{1}{M_n}}\right)\\
&\quad + \sum_{i\in\varphi_n((T'_n)^c),1\le i\le M_n} \frac{1}{M_n} f\left(\frac{P_{\tilde{U}_{M_n}}(i)}{\frac{1}{M_n}}\right)\\
&\ge \frac{|\mathcal{U}_{M_n}\setminus\varphi_n((T'_n)^c)|}{M_n} f\left(\frac{\sum_{i\in\mathcal{U}_{M_n}\setminus\varphi_n((T'_n)^c)} P_{\tilde{U}_{M_n}}(i)}{\frac{|\mathcal{U}_{M_n}\setminus\varphi_n((T'_n)^c)|}{M_n}}\right)\\
&\quad + \frac{|\varphi_n((T'_n)^c)|}{M_n} f\left(\frac{\sum_{i\in\varphi_n((T'_n)^c),1\le i\le M_n} P_{\tilde{U}_{M_n}}(i)}{\frac{|\varphi_n((T'_n)^c)|}{M_n}}\right)\\
&\ge \frac{|\mathcal{U}_{M_n}\setminus\varphi_n((T'_n)^c)|}{M_n} f\left(\frac{\Pr\{X^n\in T'_n\}}{\frac{|\mathcal{U}_{M_n}\setminus\varphi_n((T'_n)^c)|}{M_n}}\right)\\
&\quad + \frac{|\varphi_n((T'_n)^c)|}{M_n} f\left(\frac{1}{\frac{|\varphi_n((T'_n)^c)|}{M_n}}\right)\\
&= \left(1 - \frac{|\varphi_n((T'_n)^c)|}{M_n}\right) f\left(\frac{\Pr\{X^n\in T'_n\}}{1-\frac{|\varphi_n((T'_n)^c)|}{M_n}}\right)\\
&\quad + \frac{|\varphi_n((T'_n)^c)|}{M_n} f\left(\frac{M_n}{|\varphi_n((T'_n)^c)|}\right)\\
&\ge \left(1-e^{-n\gamma}\right) f\left(\frac{\Pr\{X^n\in T'_n\}}{1-e^{-n\gamma}}\right)\\
&\quad + \frac{|\varphi_n((T'_n)^c)|}{M_n} f\left(\frac{M_n}{|\varphi_n((T'_n)^c)|}\right)\\
&= \left(1-e^{-n\gamma}\right) f\left(\Pr\{X^n\in T'_n\}(1+\gamma')\right)\\
&\quad + \frac{|\varphi_n((T'_n)^c)|}{M_n} f\left(\frac{M_n}{|\varphi_n((T'_n)^c)|}\right),
\end{aligned} \tag{89}$$

for sufficiently large n, where we set $\gamma' = \frac{e^{-n\gamma}}{1-e^{-n\gamma}}$ and the first inequality is due to (13), the second and last inequalities are due to the condition C1).

Therefore, from (79) and (88) it holds that

$$\begin{aligned}
\Delta &\ge \limsup_{n\to\infty} D_f\left(\tilde{U}_{M_n}||U_{M_n}\right)\\
&\ge \limsup_{n\to\infty} f\left(\Pr\{X^n\in T'_n\}(1+\gamma')\right)\\
&\quad - \limsup_{n\to\infty} e^{-n\gamma} f\left(\Pr\{X^n\in T'_n\}(1+\gamma')\right)\\
&\quad + \liminf_{n\to\infty} \frac{|\varphi_n((T'_n)^c)|}{M_n} f\left(\frac{M_n}{|\varphi_n((T'_n)^c)|}\right)\\
&= \limsup_{n\to\infty} f\left(\Pr\{X^n\in T'_n\}\right),
\end{aligned} \tag{90}$$

where the last equality is due to the continuity of the function f.

This means that for any Δ-achievable rate R it holds that

$$\limsup_{n\to\infty} f\left(\Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \ge R-2\gamma\right\}\right) \le \Delta, \tag{91}$$

for any $\gamma > 0$. This completes the proof of the converse part. □

V. Particularization to Several Distance Measures

In previous sections, we have derived the general formula of the first-order optimum resolvability and intrinsic randomness rates with respect to f-divergences. In this section, we focus on the several specified functions f satisfying conditions C1)–C3) and compute these rates by using Theorems 3.1 and 4.1. We use the notation

$$D_f(X^n||\tilde{X}^n) := D_f(X^n||\phi_n(U_{M_n})), \tag{92}$$

$$D_f(\tilde{U}_{M_n}||U_{M_n}) := D_f(\varphi_n(X^n)||U_{M_n}), \tag{93}$$

for convenience.

Remark 5.1: Since the function $f(t) = t\log t$ (which indicates the KL divergence) does not satisfy C1) and C2), we can not apply Theorems 3.1 and 4.1 into the case of the KL divergence:

$$\begin{aligned}
D_f(X^n||\tilde{X}^n) &= D(X^n||\tilde{X}^n)\\
&= \sum_{\mathbf{x}\in\mathcal{X}^n} P_{X^n}(\mathbf{x}) \log\frac{P_{X^n}(\mathbf{x})}{P_{\tilde{X}^n}(\mathbf{x})},
\end{aligned} \tag{94}$$

$$\begin{aligned}
D_f(\tilde{U}_{M_n}||U_{M_n}) &= D(\tilde{U}_{M_n}||U_{M_n})\\
&= \sum_{1\le i\le M_n} P_{\tilde{U}_{M_n}}(i) \log\frac{P_{\tilde{U}_{M_n}}(i)}{P_{U_{M_n}}(i)}.
\end{aligned} \tag{95}$$

The resolvability problem with respect to the KL divergence of this direction has not been considered yet, while Nomura [4] has considered the problem with respect to $D(\tilde{X}^n||X^n)$ (which is indicated by $f(t) = -\log t$) and Steinberg and Verdú [3] have considered the problem with respect to the normalized KL divergence: $1/nD(\tilde{X}^n||X^n)$ (cf. Han [8]). On the other hand, in the intrinsic randomness problem, Hayashi [10, Theorem 7] has studied the problem with respect to $1/nD(\tilde{U}_{M_n}||U_{M_n})$ as well as $D(U_{M_n}||\tilde{U}_{M_n})$.

We introduce the following quantities given $0 \le \varepsilon < 1$ so as to express first-order optimum achievable rates.

$$\overline{H}(\varepsilon|\mathbf{X}) := \inf \left\{ R \left| \limsup_{n\to\infty} \Pr\left\{ \frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} > R \right\} \le \varepsilon \right. \right\}, \quad (96)$$

$$\underline{H}(\varepsilon|\mathbf{X}) := \sup \left\{ R \left| \limsup_{n\to\infty} \Pr\left\{ \frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} < R \right\} \le \varepsilon \right. \right\}. \quad (97)$$

These quantities are called *the ε-spectral sup-entropy rate* and *the ε-spectral inf-entropy rate*, respectively [8].

A. Half Variational Distance

We first consider the case of $f(t)$ as $f(t) = (1-t)^+$ which indicates

$$D_f(X^n||\tilde{X}^n) = \frac{1}{2}\sum_{\mathbf{x}\in\mathcal{X}^n} |P_{X^n}(\mathbf{x}) - P_{\tilde{X}^n}(\mathbf{x})|, \quad (98)$$

$$D_f(\tilde{U}_{M_n}||U_{M_n}) = \frac{1}{2}\sum_{1\le i\le M_n} \left| P_{\tilde{U}_{M_n}}(i) - P_{U_{M_n}}(i) \right|. \quad (99)$$

In this special case, we obtain the following corollary:

Corollary 5.1: For $f(t) = (1-t)^+$, it holds that

$$S_r^{(f)}(D|\mathbf{X}) = \overline{H}(D|\mathbf{X}), \quad (100)$$

$$S_\iota^{(f)}(D|\mathbf{X}) = \underline{H}(D|\mathbf{X}). \quad (101)$$

Proof: In the case of $f(t) = (1-t)^+$, $\overline{K}_f(D|\mathbf{X})$ reduces to

$$\begin{aligned}\overline{K}_f(D|\mathbf{X}) &= \inf \left\{ R \left| \limsup_{n\to\infty} \left(1 - \Pr\left\{ \frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \le R \right\} \right)^+ \le D \right. \right\} \\ &= \inf \left\{ R \left| \limsup_{n\to\infty} \Pr\left\{ \frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} > R \right\} \le D \right. \right\} \\ &= \overline{H}(D|\mathbf{X}). \end{aligned} \quad (102)$$

Similarly, we obtain

$$\underline{K}_f(\Delta|\mathbf{X}) = \underline{H}(\Delta|\mathbf{X}). \quad (103)$$

Hence, from Theorems 3.1 and 4.1 we obtain the corollary. □

The former result in the above corollary coincides with the result given by Steinberg and Verdú [3, Theorem 4] (see, also Han [8, Theorem 2.4.1]) while the latter one coincides with the result given by Vembu and Verdú [7, Theorem 1], and Han [8, Theorem 2.4.2].

B. Reverse Kullback-Leibler Divergence

Secondly, we consider the case of $f(t) = -\log t$, which indicates

$$\begin{aligned} D_f(X^n||\tilde{X}^n) &= D(\phi_n(U_{M_n})||X^n) \\ &= \sum_{\mathbf{x}\in\mathcal{X}^n} P_{\tilde{X}^n}(\mathbf{x}) \log\frac{P_{\tilde{X}^n}(\mathbf{x})}{P_{X^n}(\mathbf{x})}, \end{aligned} \quad (104)$$

$$\begin{aligned} D_f(\tilde{U}_{M_n}||U_{M_n}) &= D(U_{M_n}||\varphi_n(X^n)) \\ &= \sum_{1\le i\le M_n} P_{U_{M_n}}(i)\log\frac{P_{U_{M_n}}(i)}{P_{\tilde{U}_{M_n}}(i)}. \end{aligned} \quad (105)$$

In this case, we obtain the following corollary:

Corollary 5.2: For $f(t) = -\log t$, it holds that

$$\begin{aligned} S_r^{(f)}(D|\mathbf{X}) &= \overline{H}(1-e^{-D}|\mathbf{X}), \\ S_\iota^{(f)}(D|\mathbf{X}) &= \underline{H}(1-e^{-D}|\mathbf{X}). \end{aligned} \quad (106)$$

Proof: From Theorem 3.1, we have

$$\begin{aligned} \overline{K}_f(D|\mathbf{X}) &= \inf \left\{ R \left| \limsup_{n\to\infty} f\left(\Pr\left\{ \frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \le R \right\} \right) \le D \right. \right\} \\ &= \inf \left\{ R \left| \limsup_{n\to\infty} -\log \Pr\left\{ \frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \le R \right\} \le D \right. \right\} \\ &= \inf \left\{ R \left| \limsup_{n\to\infty} \Pr\left\{ \frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} > R \right\} \le 1 - e^{-D} \right. \right\} \\ &= \overline{H}\left(1-e^{-D}|\mathbf{X}\right). \end{aligned} \quad (107)$$

Similarly from Theorem 4.1, we have

$$\underline{K}_f(D|\mathbf{X}) = \underline{H}\left(1-e^{-D}|\mathbf{X}\right). \quad (108)$$

Hence, the corollary holds. □

The former result in the above corollary coincides with the result given by Nomura [4, Theorem 3.1] while the latter one coincides with the result given by Hayashi [10, Theorem 7].

C. Hellinger Distance

We consider the case of $f(t) = 1 - \sqrt{t}$, which indicates

$$D_f(X^n||\tilde{X}^n) = 1 - \sum_{\mathbf{x}\in\mathcal{X}^n} \sqrt{P_{X^n}(\mathbf{x})P_{\tilde{X}^n}(\mathbf{x})}, \quad (109)$$

$$D_f(\tilde{U}_{M_n}||U_{M_n}) = 1 - \sum_{1\le i\le M_n} \sqrt{P_{\tilde{U}_{M_n}}(i)P_{U_{M_n}}(i)}. \quad (110)$$

In this case, we have the corollary:

Corollary 5.3: For $f(t) = 1 - \sqrt{t}$, we have

$$S_r^{(f)}(D|\mathbf{X}) = \overline{H}(2D - D^2|\mathbf{X}), \quad (111)$$

$$S_\iota^{(f)}(D|\mathbf{X}) = \underline{H}(2D - D^2|\mathbf{X}). \quad (112)$$

Proof: From Theorem 3.1, we obtain

$$\begin{aligned} \overline{K}_f(D|\mathbf{X}) &= \inf \left\{ R \left| \limsup_{n\to\infty} f\left(\Pr\left\{ \frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \le R \right\} \right) \le D \right. \right\} \\ &= \inf \left\{ R \left| \limsup_{n\to\infty} \left(1 - \sqrt{\Pr\left\{ \frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \le R \right\}} \right) \le D \right. \right\} \\ &= \overline{H}\left(2D - D^2|\mathbf{X}\right). \end{aligned} \quad (113)$$

Similarly, we have from Theorem 4.1

$$\underline{K}_f(\Delta|\mathbf{X}) = \underline{H}\left(2\Delta - \Delta^2|\mathbf{X}\right). \quad (114)$$

Hence, the corollary holds. □

Remark 5.2: Kumagai and Hayashi [13], [14] have considered the first- and second- order optimum achievable rates in the random number conversion problem with respect to the approximation measure related to the Hellinger distance. It should be emphasized that they have focused on stationary memoryless sources while Corollary 5.3 is valid for *general sources*.

D. E_γ-Divergence

Finally, we consider the case of $f(t) = (\gamma - t)^+ + 1 - \gamma$, which indicates

$$D_f(X^n||\tilde{X}^n) = \sum_{\mathbf{x}\in\mathcal{X}^n: P_{X^n}(\mathbf{x})>\gamma P_{\tilde{X}^n}(\mathbf{x})} \left(P_{X^n}(\mathbf{x}) - \gamma P_{\tilde{X}^n}(\mathbf{x})\right). \quad (115)$$

$$D_f(\tilde{U}_{M_n}||U_{M_n}) = \sum_{1\le i\le M_n: P_{\tilde{U}_{M_n}}(i)>\gamma P_{U_{M_n}}(i)} \left(P_{\tilde{U}_{M_n}}(i) - \gamma P_{U_{M_n}}(i)\right). \quad (116)$$

In this case, we obtain the corollary:

Corollary 5.4: For $f(t) = (\gamma - t)^+ + 1 - \gamma$, we have

$$S_r^{(f)}(D|\mathbf{X}) = \overline{H}(D|\mathbf{X}),$$
$$S_\iota^{(f)}(D|\mathbf{X}) = \underline{H}(D|\mathbf{X}). \quad (117)$$

Proof: Noting that $\gamma \ge 1$, from Theorem 3.1, we obtain

$$\begin{aligned}\overline{K}_f(D|\mathbf{X}) &= \inf\left\{R \,\middle|\, \limsup_{n\to\infty} f\left(\Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \le R\right\}\right) \le D\right\}\\ &= \inf\left\{R \,\middle|\, \limsup_{n\to\infty}\left(\gamma - \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \le R\right\} + 1 - \gamma\right) \le D\right\}\\ &= \overline{H}(D|\mathbf{X}).\end{aligned} \quad (118)$$

Similarly, we have from Theorem 4.1

$$\underline{K}_f(\Delta|\mathbf{X}) = \underline{H}(\Delta|\mathbf{X}). \quad (119)$$

Hence, the corollary holds. □

Remark 5.3: The above corollary shows that both of optimum achievable rates with respect to the E_γ-divergence does not depend on γ, which means that these rates coincides with the optimum achievable rates with respect to the half variational distance (cf. Corollary 5.1).

Remark 5.4: Liu *et al.* [11] have dealt with the source resolvability problem with respect to the E_γ-divergence. However, our achievability (Definitions 3.1 and 3.2) differs with that in [11, Definition 14]. Hence, Corollary 5.4 and Remark 5.3 have not been provided in [11].

VI. Second-Order Optimum Achievable Rate

So far, we have considered the first-order optimum achievable rates in two random number generation problems. In this section, we consider the *second-order* optimum achievable rates in these problems.

A. General Formula

We first define the second-order achievability in the resolvability problem.

Definition 6.1: L is said to be (D, R)-achievable with the given f-divergence if there exists a sequence of mapping $\phi_n : \mathcal{U}_{M_n} \to \mathcal{X}^n$ such that

$$\limsup_{n\to\infty} D_f(X^n||\phi_n(U_{M_n})) \le D, \quad (120)$$

$$\limsup_{n\to\infty} \frac{1}{\sqrt{n}}\log\frac{M_n}{e^{nR}} \le L. \quad (121)$$

Definition 6.2 (Second-Order Optimum Resolvability Rate):

$$S_r^{(f)}(D, R|\mathbf{X}) := \inf\{L\,|L \text{ is } (D,R)\text{-achievable with the given } f\text{-divergence}\}. \quad (122)$$

In order to characterize the general formula of the second-order optimum resolvability rate $S_f(D, R|\mathbf{X})$, we define the information spectrum quantity (123), shown at the bottom of the next page, on the basis of the function f.

Then, the following theorem holds:

Theorem 6.1 (Second-Order Optimum Resolvability Rate): Assuming that the function f satisfies conditions C1)–C3), then for any $0 \le D < f(0)$ it holds that

$$S_r^{(f)}(D, R|\mathbf{X}) = \overline{K}_f(D, R|\mathbf{X}). \quad (125)$$

Proof: The proof is similar to the proof of Theorem 3.1. □

We next consider the case of the intrinsic randomness problem.

Definition 6.3: L is said to be (Δ, R)-achievable with the given f-divergence if there exists a sequence of mapping $\varphi_n : \mathcal{X}^n \to \mathcal{U}_{M_n}$ such that

$$\limsup_{n\to\infty} D_f(\varphi_n(X^n)||U_{M_n}) \le \Delta, \quad (126)$$

$$\liminf_{n\to\infty} \frac{1}{\sqrt{n}}\log\frac{M_n}{e^{nR}} \ge L. \quad (127)$$

Definition 6.4 (Second-Order Optimum Intrinsic Randomness Rate):

$$S_\iota^{(f)}(\Delta, R|\mathbf{X}) := \sup\{L\,|L \text{ is } (\Delta,R)\text{-achievable with the given } f\text{-divergence}\}. \quad (128)$$

In order to characterize $S_\iota^{(f)}(\Delta, R|\mathbf{X})$, we introduce the quantity (124), shown at the bottom of the next page, which is an analogue to $\overline{K}_f(\varepsilon, R|\mathbf{X})$ defined in (123).

Then, we have the theorem:

Theorem 6.2 (Second-Order Optimum Intrinsic Randomness Rate): Assuming that the function f satisfies C1) and C2), then for any $0 \le D < f(0)$ it holds that

$$S_\iota^{(f)}(\Delta, R|\mathbf{X}) = \underline{K}_f(\Delta, R|\mathbf{X}). \quad (129)$$

Proof: The proof proceeds in parallel with the proof of Theorem 4.1. □

B. Particularizations to Several Distance Measures

Analogously to Section V, we compute $S_r^{(f)}(D, R|\mathbf{X})$ and $S_\iota^{(f)}(\Delta, R|\mathbf{X})$ for the specified function f satisfying C1)-C3), by using Theorems 6.1 and 6.2. To do so, we define two information theoretic quantities as follows:

$$\overline{H}(\varepsilon, R|\mathbf{X}) := \inf \left\{ L \left| \limsup_{n\to\infty} \Pr\left\{ \frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} > R + \frac{L}{\sqrt{n}} \right\} \le \varepsilon \right. \right\}, \tag{130}$$

$$\underline{H}(\varepsilon, R|\mathbf{X}) := \sup \left\{ L \left| \limsup_{n\to\infty} \Pr\left\{ \frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} < R + \frac{L}{\sqrt{n}} \right\} \le \varepsilon \right. \right\}. \tag{131}$$

Remark 6.1: The *second-order* optimum coding rate in the fixed-length source coding is characterized by $\overline{H}(\varepsilon, R|\mathbf{X})$, while $\underline{H}(\varepsilon, R|\mathbf{X})$ is used to characterize the *second-order* optimum rate in the intrinsic randomness problem with respect to the variational distance [10].

We obtain the following corollary:

Corollary 6.1: It holds that

$$S_r^{(f)}(D, R|\mathbf{X}) = \begin{cases} \overline{H}(D, R|\mathbf{X}) & f(t) = (1-t)^+, \\ \overline{H}(1-e^{-D}, R|\mathbf{X}) & f(t) = -\log t, \\ \overline{H}(2D - D^2, R|\mathbf{X}) & f(t) = 1 - \sqrt{t}, \\ \overline{H}(D, R|\mathbf{X}) & f(t) = (\gamma - t)^+ + 1 - \gamma. \end{cases} \tag{132}$$

$$S_\iota^{(f)}(\Delta, R|\mathbf{X}) = \begin{cases} \underline{H}(\Delta, R|\mathbf{X}) & f(t) = (1-t)^+, \\ \underline{H}(1-e^{-\Delta}, R|\mathbf{X}) & f(t) = -\log t, \\ \underline{H}(2\Delta - \Delta^2, R|\mathbf{X}) & f(t) = 1 - \sqrt{t}, \\ \underline{H}(\Delta, R|\mathbf{X}) & f(t) = (\gamma - t)^+ + 1 - \gamma. \end{cases} \tag{133}$$

Proof: The proof is similar to the proof of Corollaries 5.1-5.4. □

The second-order optimum resolvability rates in the case of the variational distance and the KL-divergence have been already given by Nomura and Han [6, Theorem 3.1] and Nomura [4, Theorem 5.1], respectively. The second-order optimum intrinsic randomness rates in the case of the variational distance and the KL-divergence have been given by Hayashi [10, Theorems 3 and 7].

Remark 6.2: From these results, the information theoretic quantities $\overline{H}(\varepsilon, R|\mathbf{X})$ and $\underline{H}(\varepsilon, R|\mathbf{X})$ play essential roles to analyze the second-order optimum achievable rates. It should be emphasized that $\overline{H}(\varepsilon, R|\mathbf{X})$ as well as $\underline{H}(\varepsilon, R|\mathbf{X})$ have been computed for several tractable sources. For example, $\overline{H}(\varepsilon, R|\mathbf{X})$ and $\underline{H}(\varepsilon, R|\mathbf{X})$ have been explicitly calculated by Hayashi [10] for the stationary memoryless source and by Nomura and Han [6] for mixed sources.

VII. Optimistic Optimum Achievable Rate

A. Source Resolvability

In previous sections, we have considered the first- and second-order optimum resolvability and intrinsic randomness rates. In this section, we establish analogous theorems in the optimistic sense. The notion of the optimistic optimum rates has first been introduced by Vembu *et al.* [19] in the source-channel coding framework. Then, several researchers have developed the optimistic coding scenario in other information theoretic problems [10], [20]–[22]. In particular, Hayashi [10] has considered the first- and second-order optimum intrinsic randomness rates with respect to the variational distance and the KL divergence in the optimistic scenario. In this subsection, we develop the notion of the optimistic optimum rates to the resolvability problem with respect to f-divergences.

Definition 7.1: R is said to be optimistically D-achievable with the given f-divergence if there exists a sequence of mapping $\phi_n : \mathcal{U}_{M_n} \to \mathcal{X}^n$ such that for any $\nu > 0$

$$D_f(X^{n_i}||\phi_{n_i}(U_{M_{n_i}})) \le D + \nu, \tag{134}$$

$$\frac{1}{n_i}\log M_{n_i} \le R + \nu. \tag{135}$$

holds for some subsequence $n_1 < n_2 < \cdots$.

Definition 7.2 (Optimistic First-Order Optimum Resolvability Rate):

$$T_r^{(f)}(D|\mathbf{X}) := \inf\{R\,|R \text{ is optimistically } D\text{-achievable with the given } f\text{-divergence}\}. \tag{136}$$

We similarly define the second-order achievability in the optimistic scenario.

Definition 7.3: L is said to be optimistically (D, R)-achievable with the given f-divergence if there exists a sequence of mapping $\phi_n : \mathcal{U}_{M_n} \to \mathcal{X}^n$ such that for any $\nu > 0$

$$D_f(X^{n_i}||\phi_{n_i}(U_{M_{n_i}})) \le D + \nu, \tag{137}$$

$$\frac{1}{\sqrt{n_i}}\log\frac{M_{n_i}}{e^{n_i R}} \le L + \nu, \tag{138}$$

holds for some subsequence $n_1 < n_2 < \cdots$.

Definition 7.4 (Optimistic Second-Order Optimum Resolvability Rate):

$$T_r^{(f)}(D, R|\mathbf{X}) := \inf\{L\,|L \text{ is optimistically } (D,R)\text{-achievable with the given } f\text{-divergence}\}. \tag{139}$$

Remark 7.1: One may consider that we can define different quantities as follows:

$$\overline{K}_f(\varepsilon, R|\mathbf{X}) := \inf \left\{ L \left| \limsup_{n\to\infty} f\left(\Pr\left\{ \frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \le R + \frac{L}{\sqrt{n}} \right\} \right) \le \varepsilon \right. \right\}. \tag{123}$$

$$\underline{K}_f(\varepsilon, R|\mathbf{X}) := \sup \left\{ L \left| \limsup_{n\to\infty} f\left(\Pr\left\{ \frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \ge R + \frac{L}{\sqrt{n}} \right\} \right) \le \varepsilon \right. \right\}. \tag{124}$$

Definition 7.5: R is said to be type I D-achievable with the given f-divergence if there exists a sequence of mapping $\phi_n : \mathcal{U}_{M_n} \to \mathcal{X}^n$ such that

$$\liminf_{n\to\infty} D_f(X^n||\phi_n(U_{M_n})) \le D, \tag{144}$$

$$\limsup_{n\to\infty} \frac{1}{n}\log M_n \le R. \tag{145}$$

Definition 7.6:

$$T_r^{\dagger(f)}(D|\mathbf{X}) := \inf\{R\,|R \text{ is type I } D\text{-achievable with the given } f\text{-divergence}\}. \tag{146}$$

Definition 7.7: R is said to be type II D-achievable with the given f-divergence if there exists a sequence of mapping $\phi_n : \mathcal{U}_{M_n} \to \mathcal{X}^n$ such that

$$\limsup_{n\to\infty} D_f(X^n||\phi_n(U_{M_n})) \le D, \tag{147}$$

$$\liminf_{n\to\infty} \frac{1}{n}\log M_n \le R. \tag{148}$$

Definition 7.8:

$$T_r^{\ddagger(f)}(D|\mathbf{X}) := \inf\{R\,|R \text{ is type II } D\text{-achievable with the given } f\text{-divergence}\}. \tag{149}$$

Then, it is not difficult to check that

$$T_r^{(f)}(D|\mathbf{X}) = T_r^{\dagger(f)}(D|\mathbf{X}) = T_r^{\ddagger(f)}(D|\mathbf{X}). \tag{150}$$

The similar relationship also holds for the optimistic *second-order* optimum resolvability rates as well as the optimistic intrinsic randomness rates in the subsequent subsection. □

In order to show general formulas of the optimistic *first* and *second-order* optimum resolvability rates, we introduce two information quantities (140) and (141), shown at the bottom of the page, on the basis of the function f given $0 \le \varepsilon < f(0)$.

Then, we have the following theorem.

Theorem 7.1: Assuming that the function f satisfies conditions C1)-C3), then for any $0 \le D < f(0)$ it holds that

$$T_r^{(f)}(D|\mathbf{X}) = \overline{K}_f^*(D|\mathbf{X}), \tag{151}$$

$$T_r^{(f)}(D,R|\mathbf{X}) = \overline{K}_f^*(D,R|\mathbf{X}). \tag{152}$$

Proof: In view of (150), the proof proceeds in parallel with proofs of Theorems 3.1 and 6.1 in which $\limsup_{n\to\infty} D_f(X^n||\phi_n(U_{M_n}))$ is replaced by $\liminf_{n\to\infty} D_f(X^n||\phi_n(U_{M_n}))$. □

B. Intrinsic Randomness

We next consider the case of the intrinsic randomness problem in the optimistic scenario.

Definition 7.9: R is said to be optimistically Δ-achievable with the given f-divergence if there exists a sequence of mapping $\varphi_n : \mathcal{X}^n \to \mathcal{U}_{M_n}$ such that for any $\nu > 0$

$$D_f(\varphi_{n_i}(X^{n_i})||U_{M_{n_i}}) \le \Delta + \nu, \tag{153}$$

$$\frac{1}{n_i}\log M_{n_i} \ge R - \nu, \tag{154}$$

for some subsequence $n_1 < n_2 < \cdots$.

Definition 7.10 (Optimistic First-Order Optimum Intrinsic Randomness Rate):

$$T_\iota^{(f)}(\Delta|\mathbf{X}) := \sup\{R\,|R \text{ is } \Delta\text{-achievable with the given } f\text{-divergence}\}. \tag{155}$$

Definition 7.11: L is said to be optimistically (Δ, R)-achievable with the given f-divergence if there exists a sequence of mapping $\varphi_n : \mathcal{X}^n \to \mathcal{U}_{M_n}$ such that for any $\nu > 0$

$$D_f(\varphi_{n_i}(X^{n_i})||U_{M_{n_i}}) \le \Delta + \nu, \tag{156}$$

$$\frac{1}{\sqrt{n}}\log \frac{M_{n_i}}{e^{n_i R}} \ge L - \nu, \tag{157}$$

for some subsequence $n_1 < n_2 < \cdots$.

Definition 7.12 (Optimistic Second-Order Optimum Intrinsic Randomness Rate):

$$T_\iota^{(f)}(\Delta, R|\mathbf{X}) := \sup\{L\,|L \text{ is optimistically } (\Delta,R)\text{-achievable with the given } f\text{-divergence}\}. \tag{158}$$

In order to characterize $T_\iota^{(f)}(\Delta|\mathbf{X})$ and $T_\iota^{(f)}(\Delta,R|\mathbf{X})$, we introduce two quantities (142) and (143), shown at the bottom of the page. Then, we have the theorem.

Theorem 7.2: Assuming that the function f satisfies C1) and C2), then for any $0 \le \Delta < f(0)$ it holds that

$$T_\iota^{(f)}(\Delta|\mathbf{X}) = \underline{K}_f^*(\Delta|\mathbf{X}), \tag{159}$$

$$T_\iota^{(f)}(\Delta,R|\mathbf{X}) = \underline{K}_f^*(\Delta,R|\mathbf{X}). \tag{160}$$

Proof: The proof is similar to proofs of Theorems 4.1 and 6.2. □

Particularizations of Theorems 7.1 and 7.2 can be considered similarly in Sections V and VI (cf. Theorem 8.2 below).

$$\overline{K}_f^*(\varepsilon|\mathbf{X}) := \inf\left\{R \left| \liminf_{n\to\infty} f\left(\Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \le R\right\}\right) \le \varepsilon\right.\right\}. \tag{140}$$

$$\overline{K}_f^*(\varepsilon,R|\mathbf{X}) := \inf\left\{L \left| \liminf_{n\to\infty} f\left(\Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \le R + \frac{L}{\sqrt{n}}\right\}\right) \le \varepsilon\right.\right\}. \tag{141}$$

$$\underline{K}_f^*(\varepsilon|\mathbf{X}) := \sup\left\{R \left| \liminf_{n\to\infty} f\left(\Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \ge R\right\}\right) \le \varepsilon\right.\right\}. \tag{142}$$

$$\underline{K}_f^*(\varepsilon,R|\mathbf{X}) := \sup\left\{L \left| \liminf_{n\to\infty} f\left(\Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \ge R + \frac{L}{\sqrt{n}}\right\}\right) \le \varepsilon\right.\right\}. \tag{143}$$

VIII. Discussion

A. Alternative Expressions of Optimum Achievable Rates

From results in Section V, one may wonder whether first-order optimum achievable rates in the resolvability problem and the intrinsic randomness problem are always characterized by using $\overline{H}(\varepsilon|\mathbf{X})$ and $\underline{H}(\varepsilon|\mathbf{X})$. In order to give the positive answer to this question, we assume the following condition.

C4) The function f is a strictly decreasing function in $t \in (0, 1]$.

Then, we obtain:

Theorem 8.1: Assuming that the function f satisfies C1)–C4), it holds that

$$S_r^{(f)}(D|\mathbf{X}) = \overline{H}(1 - f^{-1}(D)|\mathbf{X}), \tag{161}$$

$$S_\iota^{(f)}(D|\mathbf{X}) = \underline{H}(1 - f^{-1}(D)|\mathbf{X}). \tag{162}$$

Proof: It is clear from Theorems 3.1 and 4.1, and the condition C4). □

This theorem shows a kind of *duality* between these two random number generation problems. It should be emphasized that in the case of the variational distance, this *duality* has already been reported [8]. It is not difficult to check that functions $f(t) = -\log t$, $f(t) = 1 - \sqrt{t}$, $f(t) = (1-t)^+$, and $f(t) = (\gamma - t)^+ + 1 - \gamma$ satisfy C4).

Furthermore, in the case of $D = 0$ we have

$$\overline{H}(1 - f^{-1}(D)|\mathbf{X}) = \overline{H}(0|\mathbf{X}) = \overline{H}(\mathbf{X}), \tag{163}$$

$$\underline{H}(1 - f^{-1}(D)|\mathbf{X}) = \underline{H}(0|\mathbf{X}) = \underline{H}(\mathbf{X}), \tag{164}$$

under the condition C4), where $\overline{H}(\mathbf{X})$ and $\underline{H}(\mathbf{X})$ are defined in (24) and (63), respectively. This means that, under condition C1)–C4) the optimum 0-achievable rates are equal irrespective of the function f.

On the other hand, for $0 \le \forall D < f(0)$ the following relations hold

$$\underline{H}(\mathbf{X}) \le \overline{H}(1 - f^{-1}(D)|\mathbf{X}) \le \overline{H}(\mathbf{X}), \tag{165}$$

$$\underline{H}(\mathbf{X}) \le \underline{H}(1 - f^{-1}(D)|\mathbf{X}) \le \overline{H}(\mathbf{X}). \tag{166}$$

If the strong converse property holds for the source $\mathbf{X}$, then $\underline{H}(\mathbf{X}) = \overline{H}(\mathbf{X})$ holds [8]. Thus, we have the following corollary.

Corollary 8.1: Assuming that the function f satisfies C1)–C4) and the source $\mathbf{X}$ has the strong converse property, it holds that

$$S_r^{(f)}(D|\mathbf{X}) = S_\iota^{(f)}(D|\mathbf{X}) = \overline{H}(\mathbf{X}). \quad (0 \le \forall D < f(0)) \tag{167}$$

We also obtain the second-order optimum achievable rates as follows:

Theorem 8.2: Assuming that the function f satisfies C1)–C4), it holds that

$$S_r^{(f)}(D, R|\mathbf{X}) = \overline{H}(1 - f^{-1}(D), R|\mathbf{X}), \tag{168}$$

$$S_\iota^{(f)}(D, R|\mathbf{X}) = \underline{H}(1 - f^{-1}(D), R|\mathbf{X}), \tag{169}$$

where $\overline{H}(\varepsilon, R|\mathbf{X})$ and $\underline{H}(\varepsilon, R|\mathbf{X})$ are defined in (130) and (131).

Proof: We obtain the theorem from Theorems 6.1 and 6.2, and the condition C4). □

Similarly, in the optimistic scenario, following quantities given $0 \le \varepsilon < 1$ have important roles.

$$\overline{H}^*(\varepsilon|\mathbf{X}) := \inf\left\{ R \,\middle|\, \liminf_{n\to\infty} \Pr\left\{ \frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} > R \right\} \le \varepsilon \right\}, \tag{170}$$

$$\underline{H}^*(\varepsilon|\mathbf{X}) := \sup\left\{ R \,\middle|\, \liminf_{n\to\infty} \Pr\left\{ \frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} < R \right\} \le \varepsilon \right\}, \tag{171}$$

$$\overline{H}^*(\varepsilon, R|\mathbf{X}) := \inf\left\{ L \,\middle|\, \liminf_{n\to\infty} \Pr\left\{ \frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} > R + \frac{L}{\sqrt{n}} \right\} \le \varepsilon \right\}, \tag{172}$$

$$\underline{H}^*(\varepsilon, R|\mathbf{X}) := \sup\left\{ L \,\middle|\, \liminf_{n\to\infty} \Pr\left\{ \frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} < R + \frac{L}{\sqrt{n}} \right\} \le \varepsilon \right\}. \tag{173}$$

Then, we obtain the following theorem by using the similar argument to proofs of Theorems 8.1 and 8.2.

Theorem 8.3: Assuming that the function f satisfies C1)–C4), it holds that

$$T_r^{(f)}(D|\mathbf{X}) = \overline{H}^*(1 - f^{-1}(D)|\mathbf{X}), \tag{174}$$

$$T_\iota^{(f)}(D|\mathbf{X}) = \underline{H}^*(1 - f^{-1}(D)|\mathbf{X}), \tag{175}$$

$$T_r^{(f)}(D, R|\mathbf{X}) = \overline{H}^*(1 - f^{-1}(D), R|\mathbf{X}), \tag{176}$$

$$T_\iota^{(f)}(D, R|\mathbf{X}) = \underline{H}^*(1 - f^{-1}(D), R|\mathbf{X}). \tag{177}$$

Notice here that $\overline{H}^*(\varepsilon|\mathbf{X}), \underline{H}^*(\varepsilon|\mathbf{X}), \overline{H}^*(\varepsilon, R|\mathbf{X})$, $\underline{H}^*(\varepsilon, R|\mathbf{X})$ have been computed explicitly for the stationary memoryless sources [10], [20].

B. Relation to ε-Fixed Length Source Coding

We consider the fixed length source coding as follows. Let $\xi_n : \mathcal{X}^n \to \mathcal{U}_{M_n}$, $\psi_n : \mathcal{U}_{M_n} \to \mathcal{X}^n$ be an encoder and a decoder, respectively, for the source $\mathbf{X}$. The decoding error probability is $\varepsilon_n := \Pr\{X^n \neq \psi_n(\xi_n(X^n))\}$. This code is called an (n, M_n, ε_n) code.

Definition 8.1: R is said to be ε-achievable if there exists an (n, M_n, ε) code such that

$$\limsup_{n\to\infty} \varepsilon_n \le \varepsilon, \quad \limsup_{n\to\infty} \frac{1}{n}\log M_n \le R. \tag{178}$$

Definition 8.2 (Optimum ε-Fixed Length Source Coding Rate):

$$R_{fl}(\varepsilon|\mathbf{X}) := \inf\{R | R \text{ is } \varepsilon\text{-achievable}\}. \tag{179}$$

Then, the following theorem is well-known.

Theorem 8.4 (Steinberg and Verdú [3]):

$$R_{fl}(\varepsilon|\mathbf{X}) = \overline{H}(\varepsilon|\mathbf{X}). \tag{180}$$

From the above theorem and Theorem 3.1 we immediately have:

Corollary 8.2: Assuming that the function f satisfies C1)–C4), it holds that

$$S_r^{(f)}(D|\mathbf{X}) = R_{fl}(1 - f^{-1}(D)|\mathbf{X}). \tag{181}$$

The above corollary, together with Corollary 8.3 below, reveals a relationship between the ε-fixed length source coding and the source resolvability with respect to the f-divergence.

Remark 8.1: The relationship between the fixed length source coding with the correct decoding exponent D and the resolvability with respect to the *normalized* KL divergence has been clarified by Steinberg and Verdú [3]. It should be noted that they have demonstrated this relationship, directly [3, Theorem 9]. That is, they have shown that the optimum fixed length code can be constructed from the optimum mapping in the resolvability problem with respect to the normalized KL divergence and vice versa. On the other hand, we have revealed the relationship in our setting, via the information quantity $\overline{H}(\varepsilon|\mathbf{X})$.

The similar relationship in the optimistic scenario can also be shown.

Definition 8.3: R is said to be optimistically ε-achievable if there exists an (n, M_n, ε) code such that for any $\nu > 0$

$$\varepsilon_{n_i} \le \varepsilon + \nu, \quad \frac{1}{n_i}\log M_{n_i} \le R + \nu, \tag{182}$$

for some subsequence $n_1 < n_2 < \cdots$.

Definition 8.4 (Optimistic Optimum ε-Fixed Length Source Coding Rate):

$$R^*_{fl}(\varepsilon|\mathbf{X}) := \inf\{R | R \text{ is optimistically } \varepsilon\text{-achievable}\}. \tag{183}$$

Then, it is known that

Theorem 8.5 (Chen and Alajaji [20], Hayashi [10]):

$$R^*_{fl}(\varepsilon|\mathbf{X}) = \overline{H}^*(\varepsilon|\mathbf{X}), \tag{184}$$

where $\overline{H}^*(\varepsilon|\mathbf{X})$ is defined in (170).

Thus, we have the following corollary.

Corollary 8.3: Assuming that the function f satisfies C1)–C4), it holds that

$$T_r^{(f)}(D|\mathbf{X}) = R^*_{fl}(1 - f^{-1}(D)|\mathbf{X}). \tag{185}$$

IX. Concluding Remarks

We have so far considered the first- and second-order optimum achievable rates in two random number generation problems with respect to a subclass of f-divergences. We have demonstrated the *general formulas* of the optimum achievable rates with respect to the given f-divergence by using the information spectrum approach. We have also shown that we can easily derive the results on specified functions f from our *general formulas*. In our analyses, four information quantities $\overline{H}(\varepsilon|\mathbf{X})$, $\overline{H}(\varepsilon, R|\mathbf{X})$, $\underline{H}(\varepsilon|\mathbf{X})$ and $\underline{H}(\varepsilon, R|\mathbf{X})$ have important roles. We can compute these values for several tractable sources by using previous results [3], [6], [8], [10].

In this paper, we have considered the f-divergence $D_f(X^n||\phi_n(U_{M_n}))$ in the case of the resolvability problem and $D_f(\varphi_n(X^n)||U_{M_n})$ in the case of intrinsic randomness problem. As a result, a kind of duality of these problems has been revealed. On the other hand, we can consider the resolvability problem with respect to $D_f(\phi_n(U_{M_n})||X^n)$ as well as the intrinsic randomness problem with respect to $D_f(U_{M_n}||\varphi_n(X^n))$. However, in order to treat these problems it seems we need some novel techniques, which remain to be studied. In actual, Hayashi [10] has shown that in the intrinsic randomness problem the optimum achievable rates with respect to the KL divergences $1/nD(\varphi_n(X^n)||U_{M_n})$ has completely different behavior to the optimum rates with respect to $1/nD(U_{M_n}||\varphi_n(X^n))$.

When we consider the practical situation, it is important to discuss how to construct the random number generation mapping. Proofs of Theorems 3.1 and 4.1 indicate that the way of the construction of the optimum mapping is always same irrespective of the given f-divergence in each of two problems. This observation is quite useful, because we can construct the mapping without considering the approximation measure. Furthermore, as we have mentioned in Remark 3.5, our results show that the first- and second-order optimum resolvability and intrinsic randomness rates with the given f-divergence does not depend on the behavior of the function $f(t)$ in $t > 1$. This is also interesting from the practical and theoretical points of view.

One of extensions of the setting in this paper is to consider the channel resolvability problem or the channel intrinsic randomness problem [2], [23]–[25]. When we consider the channel resolvability (or the intrinsic randomness) problem, the random coding technique is standard to show the existence of the mapping. It is interesting to discuss the random coding technique, when we consider the f-divergence as the approximation measure. This extension is one of our future works.

Finally, the condition C3) has only been needed to show Direct Part in the resolvability problem. To consider the necessity of this condition is also a future work.

Appendix
Proof of (44)

We shall show (44) in the proof of Theorem 3.1. First of all, we have

$$\begin{aligned}
&P_{\tilde{X}^n}(\mathbf{x}_{i_0}) f\left(\frac{P_{\overline{X}^n}(\mathbf{x}_{i_0})\Pr\{X^n \in S_n\}}{P_{\overline{X}^n}(\mathbf{x}_{i_0}) + e^{-n\gamma}}\right)\\
&\le \left(P_{\overline{X}^n}(\mathbf{x}_{i_0}) + e^{-n\gamma}\right) f\left(\frac{P_{\overline{X}^n}(\mathbf{x}_{i_0})\Pr\{X^n \in S_n\}}{P_{\overline{X}^n}(\mathbf{x}_{i_0}) + e^{-n\gamma}}\right)\\
&\le P_{\overline{X}^n}(\mathbf{x}_{i_0}) f\left(\frac{(1-e^{-n\gamma})P_{\overline{X}^n}(\mathbf{x}_{i_0})\Pr\{X^n \in S_n\}}{P_{\overline{X}^n}(\mathbf{x}_{i_0})}\right)\\
&\quad + e^{-n\gamma} f\left(\frac{e^{-n\gamma}P_{\overline{X}^n}(\mathbf{x}_{i_0})\Pr\{X^n \in S_n\}}{e^{-n\gamma}}\right)\\
&\le P_{\overline{X}^n}(\mathbf{x}_{i_0}) f\left((1-e^{-n\gamma})\Pr\{X^n \in S_n\}\right)\\
&\quad + e^{-n\gamma} f\left(P_{\overline{X}^n}(\mathbf{x}_{i_0})\Pr\{X^n \in S_n\}\right),
\end{aligned} \tag{186}$$

where the second inequality is due to (13). Here, noting that $\mathbf{x}_{i_0} \in S_n$ and (34), the second term of the right-hand side of the above inequality is upper bounded by

$$\begin{aligned}
e^{-n\gamma} f\left(P_{\overline{X}^n}(\mathbf{x}_{i_0})\Pr\{X^n \in S_n\}\right) &= e^{-n\gamma} f\left(P_{X^n}(\mathbf{x}_{i_0})\right)\\
&\le e^{-n\gamma} f\left(e^{-n(R_0+\gamma)}\right),
\end{aligned} \tag{187}$$

from which, together with (17), we obtain

$$\lim_{n\to\infty}\left(e^{-n\gamma}f\left(P_{\overline{X}^n}(\mathbf{x}_{i_0})\Pr\left\{X^n\in S_n\right\}\right)\right)=0. \tag{188}$$

This means that

$$P_{\tilde{X}^n}(\mathbf{x}_{i_0})f\left(\frac{P_{\overline{X}^n}(\mathbf{x}_{i_0})\Pr\left\{X^n\in S_n\right\}}{P_{\overline{X}^n}(\mathbf{x}_{i_0})+e^{-n\gamma}}\right)$$
$$\le P_{\overline{X}^n}(\mathbf{x}_{i_0})f\left(\left(1-e^{-n\gamma}\right)\Pr\left\{X^n\in S_n\right\}\right)+o(1) \tag{189}$$

holds.

Next, we evaluate the first term of the right-hand side of (189). From (15), we have

$$P_{\overline{X}^n}(\mathbf{x}_{i_0})f\left(\left(1-e^{-n\gamma}\right)\Pr\left\{X^n\in S_n\right\}\right)$$
$$\le P_{\overline{X}^n}(\mathbf{x}_{i_0})f\left(\Pr\left\{X^n\in S_n\right\}-e^{-n\gamma}\right). \tag{190}$$

Here, from the construction of the mapping it holds that

$$\begin{aligned}&P_{\overline{X}^n}(\mathbf{x}_{i_0})-P_{\tilde{X}^n}(\mathbf{x}_{i_0})\\&=\left(1-\sum_{i=1}^{i_0-1}P_{\overline{X}^n}(\mathbf{x}_i)\right)-\left(1-\sum_{i=1}^{i_0-1}P_{\tilde{X}^n}(\mathbf{x}_i)\right)\\&=\sum_{i=1}^{i_0-1}P_{\tilde{X}^n}(\mathbf{x}_i)-\sum_{i=1}^{i_0-1}P_{\overline{X}^n}(\mathbf{x}_i)\\&=\sum_{i=1}^{i_0-1}\left(P_{\tilde{X}^n}(\mathbf{x}_i)-P_{\overline{X}^n}(\mathbf{x}_i)\right)\le 0.\end{aligned} \tag{191}$$

Substituting (191) into (190) yields

$$\begin{aligned}&P_{\overline{X}^n}(\mathbf{x}_{i_0})f\left(\left(1-e^{-n\gamma}\right)\Pr\left\{X^n\in S_n\right\}\right)\\&\le P_{\tilde{X}^n}(\mathbf{x}_{i_0})f\left(\Pr\left\{X^n\in S_n\right\}-e^{-n\gamma}\right)\\&\le P_{\tilde{X}^n}(\mathbf{x}_{i_0})f\left(\Pr\left\{X^n\in S_n\right\}\right)+o(1),\end{aligned} \tag{192}$$

because of the continuity of the function f.

Therefore, from (189) and (192), we obtain

$$P_{\tilde{X}^n}(\mathbf{x}_{i_0})f\left(\frac{P_{\overline{X}^n}(\mathbf{x}_{i_0})\Pr\left\{X^n\in S_n\right\}}{P_{\overline{X}^n}(\mathbf{x}_{i_0})+e^{-n\gamma}}\right)$$
$$\le P_{\tilde{X}^n}(\mathbf{x}_{i_0})f\left(\Pr\left\{X^n\in S_n\right\}\right)+o(1). \tag{193}$$

This completes the proof. □

Acknowledgment

The author would like to thank Associate Editor, Prof. Fady Alajaji and the anonymous reviewers for their valuable comments which have improved the content of this article.

References

[1] R. Nomura, "Source resolvability problem with respect to a certain subclass of f-divergence," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2019, pp. 2234–2238.

[2] T. S. Han and S. Verdu, "Approximation theory of output statistics," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 752–772, May 1993.

[3] Y. Steinberg and S. Verdu, "Simulation of random processes and rate-distortion theory," *IEEE Trans. Inf. Theory*, vol. 42, no. 1, pp. 63–86, Jan. 1996.

[4] R. Nomura, "Source resolvability with kullback-leibler divergence," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2018, pp. 2042–2046.

[5] T. Uyematsu, "Relating source coding and resolvability: A direct approach," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2010, pp. 1350–1354.

[6] R. Nomura and T. Sun Han, "Second-order resolvability, intrinsic randomness, and fixed-length source coding for mixed sources: Information spectrum approach," *IEEE Trans. Inf. Theory*, vol. 59, no. 1, pp. 1–16, Jan. 2013.

[7] S. Vembu and S. Verdu, "Generating random bits from an arbitrary source: Fundamental limits," *IEEE Trans. Inf. Theory*, vol. 41, no. 5, pp. 1322–1332, Mar. 1995.

[8] T. S. Han, *Information-Spectrum Methods Information Theory*. New York, NY, USA: Springer, 2003.

[9] 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. 1–5.

[10] M. Hayashi, "Second-order asymptotics in fixed-length source coding and intrinsic randomness," *IEEE Trans. Inf. Theory*, vol. 54, no. 10, pp. 4619–4637, Oct. 2008.

[11] J. Liu, P. Cuff, and S. Verdú, "E_γ-resolvability," *IEEE Trans. Inf. Theory*, vol. 63, no. 5, pp. 2629–2658, May 2017.

[12] H. Yagi and T. Sun Han, "Variable-length resolvability for mixed sources and its application to variable-length source coding," 2018, *arXiv:1801.04439*. [Online]. Available: http://arxiv.org/abs/1801.04439

[13] W. Kumagai and M. Hayashi, "Second-order asymptotics of conversions of distributions and entangled states based on Rayleigh-normal probability distributions," *IEEE Trans. Inf. Theory*, vol. 63, no. 3, pp. 1829–1857, Mar. 2017.

[14] W. Kumagai and M. Hayashi, "Random number conversion and LOCC conversion via restricted storage," *IEEE Trans. Inf. Theory*, vol. 63, no. 4, pp. 2504–2532, Apr. 2017.

[15] L. Yu and V. Y. F. Tan, "Simulation of random variables under Rényi divergence measures of all orders," *IEEE Trans. Inf. Theory*, vol. 65, no. 6, pp. 3349–3383, Jun. 2019.

[16] H. Yagi and T. Sun Han, "Variable-length resolvability for general sources and channels," 2017, *arXiv:1701.08712*. [Online]. Available: http://arxiv.org/abs/1701.08712

[17] I. Csiszár and P. C. Shields, "Information theory and statistics: A tutorial," *Found. Trends Commun. Inf. Theory*, vol. 1, no. 4, pp. 417–528, 2004.

[18] I. Sason and S. Verdú, "F-divergence inequalities," *IEEE Trans. Inf. Theory*, vol. 62, no. 11, pp. 5973–6006, Nov. 2016.

[19] S. Vembu, S. Verdu, and Y. Steinberg, "The source-channel separation theorem revisited," *IEEE Trans. Inf. Theory*, vol. 41, no. 1, pp. 44–54, Feb. 1995.

[20] P.-N. Chen and F. Alajaji, "Optimistic Shannon coding theorems for arbitrary single-user systems," *IEEE Trans. Inf. Theory*, vol. 45, no. 7, pp. 2623–2629, Jun. 1999.

[21] H. Yagi, "Channel resolvability theorems for general sources and channels," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2017, pp. 2746–2750.

[22] H. Yagi and R. Nomura, "Variable-length coding with cost allowing non-vanishing error probability," *IEICE Trans. Fundamentals Electron., Commun. Comput. Sci.*, vol. E100.A, no. 8, pp. 1683–1692, 2017.

[23] M. Hayashi, "General nonasymptotic and asymptotic formulas in channel resolvability and identification capacity and their application to the wiretap channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1562–1575, Apr. 2006.

[24] M. Bloch, "Channel intrinsic randomness," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2010, pp. 2607–2611.

[25] M. R. Bloch and J. N. Laneman, "Strong secrecy from channel resolvability," *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 8077–8098, Dec. 2013.

Ryo Nomura (Member, IEEE) received the B.Eng., M.Eng., and D.Eng. degrees from Waseda University, Tokyo, Japan, in 1996, 1998, and 2007, respectively. From 2000 to 2003, he was a Research Associate with the School of Science and Engineering, Waseda University, a Research Associate and an Assistant Professor with the College of Science and Engineering, Aoyama Gakuin University, Japan, from 2004 to 2010, and a Lecturer, an Associate Professor, and a Professor with the School of Network and Information, Senshu University, Japan, from 2012 to 2018. Since 2019, he has been a Professor with the Center for Data Science, Waseda University. His research interests include information theory and its applications, especially source coding and Shannon theory.