Optimum Self-Random Number Generation Rate and Its Application to the Rate-Distortion-Perception-Problem

Ryo Nomura
Center for Data Science
Waseda University
Tokyo, Japan
nomu@waseda.jp

Abstract

In this paper, we consider the rate-distortion-perception (RDP) problem with respect to f-divergences from the viewpoint of information-theoretic random number generation. First, we address the self-random number generation problem, which is a subproblem of the RDP problem, and derive the general formula for the optimum achievable rate. Then, we apply our findings to the RDP problem.

1 Introduction

Rate–distortion (RD) theory reveals the tradeoff between the information rate and distortion level [Shannon, 1948, Cover and Thomas, 1991], but lower distortion does not necessarily imply higher perceptual quality in a realistic situation, such as image processing. Blau and Michaeli [2019] formulated this perceptual quality of reconstructed data as the variational distance between the probability distribution of the original information source and that of the reconstructed information, and they showed the existence of a tradeoff between perception and distortion. Matsumoto [2018] was the first to attempt to incorporate this perceptional quality into RD theory. He introduced the tradeoff among information rate, distortion, and perceptual quality and derived the general formula for the rate-distortion-dispersion (RDP) function. It should be emphasized that Matsumoto [2018] treated the RD problem as a subproblem of the RDP problem. A coding theorem for more general settings has been given by Theis and Wagner [2021].

Meanwhile, the self-random number generation (SRNG) problem has been considered in information theory. The main objective of the SRNG problem is to approximate the source $\mathbf{X} = \{X^n\}_{n=1}^{\infty}$ by using that source efficiently while keeping the approximation error smaller than or equal to some given constant. To formulate this problem, let $d(X^n, Y^n)$ denote some approximation measure between two probability distributions P_{X^n} and P_{Y^n} (e.g., the variational distance). Then, given an arbitrary general source $\mathbf{X} = \{X^n\}_{n=1}^{\infty}$, we seek a mapping $\phi_n(X^n)$ with $\limsup_{n\to\infty} d(X^n, \phi_n(X^n)) \le D$, where the rate $\log|\phi_n|$ should be as small as possible. This problem can also be considered as a subproblem of the RDP problem, so analyzing the SRNG problem is a beneficial step toward a comprehensive understanding of the RDP problem. In this paper, we derive the optimal achievable rate for the SRNG problem and then extend our findings to address the broader scope of the RDP problem.

Information-Theoretic Principles in Cognitive Systems Workshop at the 37th Conference on Neural Information Processing Systems (NeurIPS 2023).

2 SRNG problem

In this paper, we consider the *general* source [Han, 2003] 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 finite or infinite countable set $\mathcal{X}$. Here, we discuss how to approximate the source $\mathbf{X}$ by using $\mathbf{X}$ itself efficiently. Consider two mappings $\varphi_n : \mathcal{X}^n \to \mathcal{M}_n := \{1, 2, \ldots, M_n\}$ and $\psi_n : \mathcal{M}_n \to \mathcal{X}^n$, and set $\tilde{X}^n = \psi_n(\varphi_n(X^n))$. We consider the f-divergence as an approximation measure between X^n and $\tilde{X}^n$. The f-divergence between two probabilistic distributions P_Z and $P_{\overline{Z}}$ is defined as follows [Csiszár and Shields, 2004], letting $f(t)$ be a convex function defined for $t > 0$ and $f(1) = 0$.

Definition 2.1. 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

$$D_f(Z||\overline{Z}) := \sum_{z \in \mathcal{Z}} P_{\overline{Z}}(z) f\left(\frac{P_Z(z)}{P_{\overline{Z}}(z)}\right), \tag{1}$$

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

The following are some examples of f-divergences [Csiszár and Shields, 2004, Sason and Verdú, 2016]:

- $f(t) = t \log t$: (Kullback–Leibler divergence)
  $D_f(Z||\overline{Z}) = \sum_{z \in \mathcal{Z}} P_Z(z) \log \frac{P_Z(z)}{P_{\overline{Z}}(z)} =: D(Z||\overline{Z})$;
- $f(t) = -\log t$: (reverse Kullback–Leibler divergence) $D_f(Z||\overline{Z}) = D(\overline{Z}||Z)$;
- $f(t) = (1-t)^+ := \max\{1-t, 0\}$: (variational distance)
  $D_f(Z||\overline{Z}) = \frac{1}{2} \sum_{z \in \mathcal{Z}} |(P_Z(z) - P_{\overline{Z}}(z))|$.

The optimum SRNG rate with respect to the given f-divergence is defined as follows.

Definition 2.2. Rate R is said to be Δ-achievable with the given f-divergence if there exists a sequence of mappings (φ_n, ψ_n) such that

$$\limsup_{n \to \infty} D_f\left(X^n||\tilde{X}^n\right) \le \Delta \text{ and } \limsup_{n \to \infty} \frac{1}{n} \log M_n \le R, \tag{2}$$

where $\tilde{X}^n = \psi_n(\varphi_n(X^n))$.

Definition 2.3 (Optimum SRNG rate)**.**

$$S_f(\Delta|\mathbf{X}) = \inf\left\{R \quad | \quad R \text{ is } \Delta\text{-achievable}\right\}. \tag{3}$$

The optimum SRNG rate has been considered in the case of the variational distance (i.e., $f(t) = (1-t)^+$) and $\Delta = 0$, and Han [2003] showed the following theorem.

Theorem 2.1 (Han [2003])**.** For $f(t) = (1-t)^+$, it holds that

$$S_f(0|\mathbf{X}) = \inf\left\{R \left| \lim_{n \to \infty} \Pr\left\{\frac{1}{n} \log \frac{1}{P_{X^n}(X^n)} > R\right\} = 0\right.\right\} =: \overline{H}(\mathbf{X}), \tag{4}$$

where $\overline{H}(\mathbf{X})$ is called the *spectral sup-entropy rate* of the source $\mathbf{X}$ [Han, 2003].

The spectral sup-entropy rate $\overline{H}(\mathbf{X})$ is a generalization of the entropy defined by Shannon [1948] for an i.i.d. source.

To derive $S_f(\Delta|\mathbf{X})$, we assume the following three conditions on the function f.

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

C2 For any pair of positive real numbers (a, b), it holds that $\lim_{n \to \infty} e^{-na} f\left(e^{-nb}\right) = 0$.

C3 For any positive number $a \in [0, 1]$, it holds that $0f\left(\frac{a}{0}\right) = 0$.

Remark 2.1. Notice here that $f(t) = -\log t$ and $f(t) = (1-t)^+$ satisfy these conditions, while $f(t) = t \log t$ does not.

3 Optimum SRNG rate

3.1 Fundamental lemmas

Before considering optimum achievable rates in the SRNG problem with f-divergences, we show two useful lemmas. Proofs of these lemmas can be found in Nomura [2023].

Lemma 3.1 (Nomura [2023])**.** Assuming that the function f satisfies conditions C1 and C3, for any M_n and $\gamma > 0$, there exists a pair of mappings (φ_n, ψ_n) satisfying

$$D_f(X^n||\psi_n(\varphi_n(X^n))) \le f\left(\Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \le \frac{1}{n}\log M_n - \gamma\right\} - e^{-n\gamma}\right) + e^{-n\gamma} f\left(\frac{1}{M_n}\right). \tag{5}$$

Remark 3.1. In the direct part of the proof of Theorem 2.1, Han used a pair of mappings that is essentially the same as the optimum fixed-length source code. However, in the proof of the above lemma, one may wonder whether we can use a pair of mappings such as

$$\varphi'_n(\mathbf{x}_i) = \begin{cases} i & \mathbf{x} \in S_n, \\ 1 & \text{otherwise} \end{cases} \tag{6}$$

and $\psi'_n(i) = \mathbf{x}_i$, where S_n is a high-probability set satisfying $|S_n| \approx M_n$ (see the proof of the lemma in the Appendix). This pair (φ'_n, ψ'_n) has often been used to show the source coding theorem, but unfortunately it is difficult to derive a similar bound by using (φ'_n, ψ'_n) in the case of f-divergences.

Lemma 3.2 (Nomura [2023])**.** Assuming that the function f satisfies conditions C1 and C3, for any pair of mappings (φ_n, ψ_n) it holds that

$$D_f\left(X^n||\psi_n(\varphi_n(X^n))\right) \ge f\left(\Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \le \frac{1}{n}\log M_n + \gamma\right\} + e^{-n\gamma}\right) \tag{7}$$

for any $\gamma > 0$.

3.2 General formula

By using the previous two lemmas, we show here the general formula for the SRNG problem with f-divergences. To express the general formula for the optimum SRNG rate, we define the following quantity that depends on the function f:

$$K_f(\Delta|\mathbf{X}) := \inf\left\{R \,\middle|\, \limsup_{n\to\infty} f\left(\Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \le R\right\}\right) \le \Delta\right\}. \tag{8}$$

Then, we have the following theorem.

Theorem 3.1. Assuming that the function f satisfies conditions C1–C3, then for any $0 \le \Delta < f^{-1}(0)$ it holds that

$$S_f(\Delta|\mathbf{X}) = K_f(\Delta|\mathbf{X}). \tag{9}$$

Proof. See Nomura [2023]. □

Next, we show a particular example of our theorem by considering only the case of $f(t) = (1-t)^+$, which indicates the variational distance. From Theorem 3.1, we obtain the following corollary.

Corollary 3.1. For $f(t) = (1-t)^+$, it holds that

$$S_f(\Delta|\mathbf{X}) = \inf\left\{R \,\middle|\, \limsup_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} > R\right\} \le \varepsilon\right\} =: \overline{H}(\Delta|\mathbf{X}). \tag{10}$$

This corollary is a generalization of Theorem 2.1, so the quantity $K_f(\Delta|\mathbf{X})$ in (8) is a form of generalization of the entropy.

4 RDP function

Next, we consider the RDP problem. Let $\phi_n : \mathcal{X}^n \to \mathcal{M}_n$ and $\xi_n : \mathcal{M}_n \to \mathcal{X}^n$ denote a lossy source encoder and a decoder, respectively. A general distortion function is defined by a mapping $g_n : \mathcal{X}^n \times \mathcal{X}^n \to [0, +\infty)$, where $g_n(\mathbf{x}, \mathbf{x}) = 0$ for any $\mathbf{x} \in \mathcal{X}^n$.

Below, we define the RDP function with respect to f-divergences.

Definition 4.1. A triplet (R, D, Δ) is said to be achievable with the given f-divergence if there exists a sequence of (ϕ_n, ξ_n) such that

$$\limsup_{n\to\infty} \frac{1}{n} \log M_n \le R, \quad \limsup_{n\to\infty} \frac{1}{n} \mathbb{E}\left[g_n\left(X^n, \xi_n(\phi_n(X^n))\right)\right] \le D, \tag{11}$$

$$\limsup_{n\to\infty} D_f(X^n || \xi_n(\phi_n(X^n))) \le \Delta. \tag{12}$$

Definition 4.2 (RDP function with the given f-divergence)**.**

$$R_f(D, \Delta) = \inf\left\{R | (R, D, \Delta) \text{ is achievable}\right\}. \tag{13}$$

Then, from Theorem 3.1 we immediately have the following theorem.

Theorem 4.1.

$$R_f(D, \Delta) \ge \max\{r(D|\mathbf{X}), K_f(\Delta|\mathbf{X})\}, \tag{14}$$

where $r(D|\mathbf{X})$ is the general RD function (see Steinberg and Verdú [1996] and Han [2003]) and $K_f(\Delta|\mathbf{X})$ is defined in (8).

Proof. The theorem is obvious from Theorem 3.1 and the result for the RD function in a general setting [Steinberg and Verdú, 1996, Han, 2003]. □

The above theorem shows the lower bound of the RDP function with the given f-divergence. Unfortunately, the upper bound in the general case is difficult to derive. Here, we define two quantities $\overline{g_n} := \max_{(\mathbf{x}, \mathbf{x}')} g_n(\mathbf{x}, \mathbf{x}')$ and $D_{threshold} := \frac{1}{n}\overline{g_n} \cdot \Pr\{-\log P_{X^n}(X^n) \ge K_f(\Delta|\mathbf{X})\}$ given Δ, the function f, and g_n.

Theorem 4.2. For $D \ge D_{threshold}$, we have

$$R_f(D, \Delta) \le K_f(\Delta|\mathbf{X}). \tag{15}$$

Proof. We can prove this theorem by using the pair of mappings used in the proof of Lemma 3.1. □

Intuitively, relaxing the constraints on distortion levels leads to the prominence of the condition stated in (12) regarding perceptual quantity. This phenomenon is highlighted by the condition $D \ge D_{threshold}$ in the above theorem.

Originally, the RDP function was defined with respect to the variational distance instead of the f-divergence in (12) [Matsumoto, 2019], and Matsumoto showed the following theorem.

Theorem 4.3 (RDP function [Matsumoto, 2019])**.** For $f(t) = (1-t)^+$, it holds that

$$R_f(D, \Delta) = \max\{r(D|\mathbf{X}), \overline{H}(\Delta|\mathbf{X})\}, \tag{16}$$

where $\overline{H}(\Delta|\mathbf{X})$ is defined in (10).

Note here that Theis and Wagner [2021] considered the RDP function in a more general setting, considering a stochastic encoder and a decoder. It is not straightforward to derive the upper bound (direct part) of the RDP function in general. However, we believe that Lemma 3.1 and its proof give useful insights into how to construct a suitable code for the RDP problem.

Acknowledgement

The author would like to thank anonymous reviewers for valuable comments. This work was supported in part by JSPS KAKENHI Grant Number JP22K04111, the Kayamori Foundation of Informational Science Advancement.

References

Y. Blau and T. Michaeli. Rethinking lossy compression: The rate-distortion-perception tradeoff. In *Proc. of the 36th International Conference on Machine Learning*, pages 675 – 685, 2019.

T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley, 1991.

I. Csiszár and P. C Shields. Information theory and statistics: A tutorial. *Foundations and Trends® in Communications and Information Theory*, 1(4):417–528, 2004.

T. S. Han. *Information-Spectrum Methods in Information Theory*. Springer, New York, 2003.

R. Matsumoto. Introducing the perception-distortion tradeoff into the rate-distortion theory of general information sources. *IEICE Communications Express*, 7(11):427–431, 2018. doi: 10.1587/comex.2018XBL0109.

R. Matsumoto. Rate-distortion-perception tradeoff of variable-length source coding for general information sources. *IEICE Communications Express*, 8(2):38–42, 2019. doi: 10.1587/comex.2018XBL0139.

R. Nomura. Optimum self-random number generation rate and its application to rdp function. In *2023 IEEE International Symposium on Information Theory (ISIT)*, pages 460–465, 2023. doi: 10.1109/ISIT54713.2023.10206993.

I. Sason and S. Verdú. f-divergence inequalities. *IEEE Trans. Inf. Theory*, 62(11):5973–6006, 2016.

C. E. Shannon. A mathematical theory of communication. *Bell Syst. Tech. J.*, 27:379–423, 623–656, 1948.

Y. Steinberg and S. Verdú. Simulation of random processes and rate-distortion theory. *IEEE Trans. Inf. Theory*, 42(1):63–86, 1996.

L. Theis and A. B. Wagner. A coding theorem for the rate-distortion-perception function. In *Ninth International Conference on Learning Representations neural compression workshop*, 2021.