Optimum Overflow Thresholds in Variable-Length Source Coding Allowing Non-Vanishing Error Probability

Ryo Nomura, *Member, IEEE*, and Hideki Yagi, *Member, IEEE*

***Abstract*— The variable-length source coding problem allowing the error probability up to some constant is considered for general sources. In this problem, the optimum mean codeword length of variable-length codes has already been determined. On the other hand, in this paper, we focus on the overflow (or excess codeword length) probability instead of the mean codeword length. The infimum of overflow thresholds under the constraint that both of the error probability and the overflow probability are smaller than or equal to some constant is called the optimum overflow threshold. In this paper, we first derive finite-length upper and lower bounds on these probabilities so as to analyze the optimum overflow thresholds. Then, by using these bounds, we determine the general formula of the optimum overflow thresholds in both of the first-order and second-order forms. Next, we consider another expression of the derived general formula so as to reveal the relationship with the optimum coding rate in the fixed-length source coding problem. Finally, we apply the general formula derived in this paper to the case of stationary memoryless sources.**

***Index Terms*— Error probability, general source, overflow probability, variable-length source coding.**

I. Introduction

THE variable-length source coding is one of important problems from both practical and theoretical points of view. The performance of variable-length codes is evaluated by several criteria such as the mean codeword length, the overflow probability (or excess codeword length) and so on. Shannon [2] has first demonstrated that the infimum of the mean codeword length coincides with the source entropy for stationary memoryless sources. Han [3] has extended the results into the case of *general* sources. The overflow probability, which is defined as the probability of codeword length being above some threshold, has also been analyzed in several contexts [4]–[6]. Uchida and Han [5] have shown the infimum of achievable thresholds given the overflow probability *exponent* r for *general* sources. Kontoyiannis and Verdú [6] have investigated the optimum overflow threshold, which means the infimum of the overflow threshold under the constraint that the overflow probability is smaller than or equal to $\delta > 0$. They have considered the optimum codeword length *without prefix constraints* and shown that the relationship between the optimum overflow threshold in variable-length coding and the optimum error probability in fixed-length coding. All the results mentioned above are for the variable-length coding *without* error.

In this paper, on the other hand, we consider the variable-length coding allowing the error probability up to some constant $\varepsilon > 0$, which we call the ε-variable-length coding. The *first-order* optimum mean codeword length of ε-variable-length codes has been derived by Han [3], and Koga and Yamamoto [7]. Kostina *et al.* [8] have determined the *second-order* optimum mean codeword length of the ε-variable-length codes. They have revealed that the *second-order* optimum mean codeword length of the ε-variable-length codes has a completely different behavior with that of the variable-length codes *without error* [8]. Yagi and Nomura [9] have also characterized the *first-* and *second-order* optimum mean codeword *cost* of the ε-variable-length codes.

Inspired by the result in [8], we also focus on the the ε-variable-length coding problem and attempt to investigate the optimum overflow threshold in the problem. As we have mentioned previously, the *first- and second-order* optimum overflow thresholds in the variable-length coding *without error* have already been studied [6], [10]. We extend the problem setting to the case of ε-variable-length coding and derive the *general* formula of the *first- and second-order* optimum overflow thresholds. To this end, we first derive finite-length upper and lower bounds on the error probability and the overflow probability. Then, using these bounds, we determine the *general* formulas of the *first- and second-order* optimum overflow thresholds. We also provide another expression of our general formulas so as to reveal the relationship with the optimum coding rate in the fixed-length coding problem.

Related works include the work by Saito and Matsushima [11], in which the *first-order* optimum overflow threshold in ε-variable-length coding has been determined by using the smooth max entropy (or smooth Rényi entropy of order zero). The analyses here are based on information

Manuscript received October 6, 2018; revised May 18, 2019; accepted May 21, 2019. Date of publication June 3, 2019; date of current version November 20, 2019. This work was supported in part by JSPS KAKENHI under Grant JP16K06340, Grant JP18H01438, Grant JP17K00020, and Grant JP18K04150. This paper was presented at the 2017 IEEE Information Theory Workshop [1].

R. Nomura is with the Center for Data Science, Waseda University, Tokyo 162-0042, Japan (e-mail: nomu@waseda.jp).

H. Yagi is with the Department of Computer and Network Engineering, The University of Electro-Communications, Tokyo 184-8795, Japan (e-mail: h.yagi@uec.ac.jp).

Communicated by I. Kontoyiannis, Associate Editor At Large.

Digital Object Identifier 10.1109/TIT.2019.2920417

spectrum methods and the approach is different from that in [11].

This paper is organized as follows. In Section II, we describe the problem setting and give some definitions of the *first*- and *second*-order optimum overflow thresholds. In Section III, we derive the finite blocklength upper and lower bounds so as to investigate the optimum overflow threshold in the subsequent sections. In Section IV, we show the *general formula* of the optimum *first-order* overflow threshold. We also give another expression of the *general* formula and compare to the *first*-order optimum achievable rates in the *fixed-length* source coding. In Section V, we show the *general* formula of the *second*-order optimum thresholds. In Section VI, we compute the optimum thresholds for the stationary memoryless source by using *general formulas* given in the preceding sections. Finally, we provide some concluding remarks on our results in Section VII.

II. Variable-Length Coding Allowing Errors

A. Problem Setting

Let $\mathbf{X} = \left\{X^n = (X_1^{(n)}, X_2^{(n)}, \ldots, X_n^{(n)})\right\}_{n=1}^{\infty}$ denote a *general* source, where each $X_i^{(n)}$ takes a value in the finite or countably infinite alphabet $\mathcal{X}$. We use the term *general source* to denote a sequence of random variables X^n indexed by blocklength n and denote the probability distribution of X^n as P_{X^n}. We consider the variable-length codes characterized as follows. Let $\varphi_n : \mathcal{X}^n \to \mathcal{U}^*$ and $\psi_n : \mathcal{U}^* \to \mathcal{X}^n$ denote a variable-length encoder and a decoder respectively, where $\mathcal{U} = \{1, 2, \cdots, K\}$ is a code alphabet and $\mathcal{U}^*$ is the set of all finite-length strings over $\mathcal{U}$ excluding the null string. The codeword length for the source sequence $\mathbf{x} \in \mathcal{X}^n$ is denoted by $l(\varphi_n(\mathbf{x}))$ when we use the encoder φ_n.

In this setting, we are interested in the following two probabilities:

Definition 2.1: The error probability of (φ_n, ψ_n) and the overflow probability of (φ_n, ψ_n) with threshold η_n are respectively defined as

$$\varepsilon_n := \Pr\left\{X^n \neq \psi_n(\varphi_n(X^n))\right\}, \tag{2.1}$$

$$\delta_n(\eta_n) := \Pr\left\{l(\varphi_n(X^n)) > \eta_n\right\}. \tag{2.2}$$

Notice here that $\eta_n < 1$ always leads to $\delta_n(\eta_n) = 1$. Hence, without loss of generality we assume that $\eta_n \geq 1$. In particular, we consider two cases such as $\eta_n = nR$ and $\eta_n = nR + \sqrt{n}L$.

We next define the achievability considered in this paper. For a given source $\mathbf{X}$ and a threshold η_n we cannot minimize ε_n and $\delta_n(\eta_n)$ simultaneously, because there exists a trade-off relation between these two quantities in general. Instead, we focus on the first and second-order optimum overflow threshold as follows.

Definition 2.2: Rate R is said to be (ε, δ)-achievable ($\varepsilon, \delta \in [0, 1)$), if there exists a sequence of variable-length code (φ_n, ψ_n) such that

$$\limsup_{n\to\infty} \varepsilon_n \leq \varepsilon, \quad \limsup_{n\to\infty} \delta_n(nR) \leq \delta. \tag{2.3}$$

Definition 2.3 (First-order (ε, δ)-optimum threshold):

$$R(\varepsilon, \delta|\mathbf{X}) := \inf\left\{R | R \text{ is } (\varepsilon, \delta)\text{-achievable}\right\}. \tag{2.4}$$

The second-order optimum threshold is similarly defined as follows.

Definition 2.4: Rate L is said to be (ε, δ, R)-achievable ($\varepsilon, \delta \in [0, 1)$, $R \geq 0$), if there exists a sequence of variable-length code (φ_n, ψ_n) such that

$$\limsup_{n\to\infty} \varepsilon_n \leq \varepsilon, \quad \limsup_{n\to\infty} \delta_n\left(nR + \sqrt{n}L\right) \leq \delta. \tag{2.5}$$

Definition 2.5 (Second-order (ε,δ,R)-optimum threshold):

$$L(\varepsilon, \delta, R|\mathbf{X}) := \inf\{L | L \text{ is } (\varepsilon, \delta, R)\text{-achievable}\}. \tag{2.6}$$

Remark 2.1: It is not difficult to check that the condition $\varepsilon + \delta \geq 1$ yields the trivial result such as $R(\varepsilon, \delta|\mathbf{X}) = 0$ or $L(\varepsilon, \delta, R|\mathbf{X}) = -\infty$. Hence, in this paper we assume that $\varepsilon + \delta < 1$ holds.

In this paper, we consider the non-prefix variable-length code. We here derive the necessary condition for non-prefix variable-length codes allowing errors. For a variable-length code (φ_n, ψ_n), let $D_n(\varphi_n, \psi_n) \subset \mathcal{X}^n$ and $T_n(\varphi_n, \eta_n) \subset \mathcal{X}^n$ be defined as follows:

$$D_n(\varphi_n, \psi_n) := \left\{\mathbf{x} \in \mathcal{X}^n \,|\mathbf{x} = \psi_n(\varphi_n(\mathbf{x}))\right\}, \tag{2.7}$$

$$T_n(\varphi_n, \eta_n) := \left\{\mathbf{x} \in \mathcal{X}^n \,| l(\varphi_n(\mathbf{x})) \leq \eta_n\right\}. \tag{2.8}$$

Then, since any sequence $\mathbf{x} \in D_n(\varphi_n, \psi_n)$ is correctly decodable, it holds that

$$|D_n(\varphi_n, \psi_n) \cap T_n(\varphi_n, \eta_n)| \leq \sum_{i=1}^{\eta_n} K^i < K^{\eta_n + 1}. \tag{2.9}$$

We use (2.9) instead of Kraft's inequality as a condition for non-prefix variable-length codes in this paper. Throughout this paper, the logarithm is taken to the base K.

B. Previous Results

Saito and Matsushima [11] have derived the *first-order* (ε, δ)-optimum threshold by using the smooth max entropy.

Definition 2.6 (Smooth max entropy): For any given $\gamma \in [0, 1)$, the smooth max entropy (or smooth Rényi entropy of order zero) of the source is defined by

$$H^{\gamma}(X) := \min_{A \subset \mathcal{X} : \Pr\{X \in A\} \geq 1 - \gamma} \log |A|. \tag{2.10}$$

Theorem 2.1 (Saito and Matsushima [11]): For any $\varepsilon, \delta \in [0, 1)$ satisfying $\varepsilon + \delta < 1$, it holds that

$$R(\varepsilon, \delta|\mathbf{X}) = \lim_{\nu \downarrow 0} \limsup_{n\to\infty} \frac{1}{n} H^{\varepsilon + \delta + \nu}(X^n). \tag{2.11}$$

□

Remark 2.2: In the fixed-length source coding problem, the infimum of achievable rates under the constraint that the error probability is asymptotically up to ε is called the ε-optimum coding rate. Uyematsu [12] has provided the *general* formula of the ε-optimum coding rate also by using the smooth max entropy.

In the *zero-error* variable-length coding problem, Kontoyiannis and Verdú [6] have discussed the *optimum* codeword length without *prefix* constraints by considering the optimum variable-length codes. They have also pointed out that the minimum error probability of n-to-R fixed-length codes coincides with the minimum overflow probability of variable-length codes with the threshold $\eta_n = nR$. Hence, if we consider the *zero-error* variable-length coding problem, we can evaluate the optimum overflow threshold by investigating the optimum fixed-length coding rate.

One simple way to evaluate the first-order (ε, δ)-optimum threshold (as well as the second-order (ε, δ, R)-optimum threshold) is to extend this relation in [6] into the case of the variable-length coding allowing errors. In this paper, however, we employ another way to derive these optimum threshold called the information-spectrum methods developed by Verdú and Han.

III. Finite Blocklength Bounds

In this section, we derive the finite blocklength upper and lower bounds on the error probability and the overflow probability.

Theorem 3.1 (Finite blocklength upper bound): Let $a_n > 0$, $\eta_n \geq 1$ be arbitrary positive numbers. Then, for any $A_n \subset \mathcal{X}^n$ satisfying $\Pr\{X^n \in A_n\} \geq 1 - \varepsilon$, there exists a variable-length code (φ_n, ψ_n) such that

$$\begin{aligned} &\varepsilon_n \leq \varepsilon, \\ &\delta_n(\eta_n) \leq \Pr\left\{ a_n \frac{P_{X^n}(X^n)}{\Pr\{X^n \in A_n\}} \leq K^{-\eta_n},\ X^n \in A_n \right\} \\ &\qquad + a_n K. \end{aligned} \tag{3.1}$$

Proof: We first construct the encoder and the decoder.

[Encoder φ_n]: For any fixed $A_n \subset \mathcal{X}^n$ satisfying $\Pr\{X^n \in A_n\} \geq 1 - \varepsilon$, we define the encoder as

$$\varphi_n(\mathbf{x}) = \begin{cases} f_n(\mathbf{x}) & \mathbf{x} \in A_n, \\ 1 & \text{otherwise}, \end{cases} \tag{3.2}$$

where $f_n : A_n \to \mathcal{U}^*$ is an injection mapping which assigns the codeword whose length[1] is $\lceil -\log \frac{P_{X^n}(\mathbf{x})}{\Pr\{X^n \in A_n\}} \rceil$ to each $\mathbf{x} \in A_n$. It is not difficult to verify that there exists such an injection mapping. Actually, if we consider the probability distribution $P_{\overline{X}^n}(\mathbf{x}) = \frac{P_{X^n}(\mathbf{x})}{\Pr\{X^n \in A_n\}}$ over A_n, then there exists a prefix code for A_n without error, because $\sum_{\mathbf{x} \in A_n} K^{\log \frac{P_{X^n}(\mathbf{x})}{\Pr\{X^n \in A_n\}}} = 1$ holds. The codeword length $l(\varphi_n(\mathbf{x}))$ of this code is given by

$$l(\varphi_n(\mathbf{x})) = \begin{cases} \lceil -\log \frac{P_{X^n}(\mathbf{x})}{\Pr\{X^n \in A_n\}} \rceil & \mathbf{x} \in A_n, \\ 1 & \text{otherwise}. \end{cases} \tag{3.3}$$

[Decoder ψ_n]: The decoder ψ_n is arranged to be an inverse mapping of f_n. That is, for a received sequence $\mathbf{u} \in \mathcal{U}^*$, if there exists $\mathbf{x} \in A_n$ such that $\mathbf{u} = f_n(\mathbf{x})$, then the decoder outputs $\psi_n(\mathbf{u}) = \mathbf{x}$. If there does not exist such $\mathbf{x}$ (this would happen, for example, when $f_n(\mathbf{x}) \neq 1$ holds for all $\mathbf{x} \in A_n$), then the decoder declares an error.

Next, we evaluate the error probability and the overflow probability of this variable-length code. From the construction of the code, the error probability is given by $\varepsilon_n \leq \Pr\{X^n \notin A_n\} \leq \varepsilon$. Hence, it suffices to show (3.1). To do so, let us define S_n and B_n as follows:

$$S_n := \left\{ \mathbf{x} \in \mathcal{X}^n \,\middle|\, l(\varphi_n(\mathbf{x})) > \eta_n \right\}, \tag{3.4}$$

$$B_n := \left\{ \mathbf{x} \in \mathcal{X}^n \,\middle|\, a_n \frac{P_{X^n}(\mathbf{x})}{\Pr\{X^n \in A_n\}} \leq K^{-\eta_n} \right\}. \tag{3.5}$$

Then, since $\eta_n \geq 1$ holds, we have

$$(A_n)^c \subseteq (S_n)^c \tag{3.6}$$

from the construction of the code (φ_n, ψ_n), where c denotes the complement. Moreover, for any $\mathbf{x} \in S_n$ it holds that $-\log \frac{P_{X^n}(\mathbf{x})}{\Pr\{X^n \in A_n\}} > \eta_n - 1$. This means that for any $\mathbf{x} \in S_n$

$$P_{X^n}(\mathbf{x}) < K^{-(\eta_n - 1)} \Pr\{X^n \in A_n\}. \tag{3.7}$$

Thus, from (3.6) and (3.7) we have

$$\begin{aligned} \delta_n(\eta_n) &= \Pr\left\{X^n \in S_n \cap B_n\right\} + \Pr\left\{X^n \in S_n \cap (B_n)^c\right\} \\ &\leq \Pr\left\{X^n \in A_n \cap B_n\right\} + \sum_{\mathbf{x} \in S_n \cap (B_n)^c} P_{X^n}(\mathbf{x}) \\ &\leq \Pr\left\{X^n \in A_n \cap B_n\right\} \\ &\quad + \sum_{\mathbf{x} \in (B_n)^c} K^{-(\eta_n - 1)} \Pr\{X^n \in A_n\} \\ &\leq \Pr\left\{X^n \in A_n \cap B_n\right\} \\ &\quad + |(B_n)^c| K^{-(\eta_n - 1)} \Pr\{X^n \in A_n\}. \end{aligned} \tag{3.8}$$

Next, we evaluate the second term on the r.h.s. of (3.8). From the definition of B_n, we have

$$\begin{aligned} 1 &\geq \sum_{\mathbf{x} \notin B_n} P_{X^n}(\mathbf{x}) \\ &\geq \sum_{\mathbf{x} \notin B_n} \frac{K^{-\eta_n}}{a_n} \Pr\left\{X^n \in A_n\right\} \\ &= |(B_n)^c| \frac{K^{-\eta_n}}{a_n} \Pr\left\{X^n \in A_n\right\} \end{aligned} \tag{3.9}$$

from which it follows that

$$|(B_n)^c| \leq a_n K^{\eta_n} \frac{1}{\Pr\{X^n \in A_n\}}. \tag{3.10}$$

Plugging (3.10) into (3.8) yields (3.1). □

Theorem 3.2 (Finite blocklength lower bound): For an arbitrary fixed variable-length code (φ_n, ψ_n), we set $D_n = \{\mathbf{x} \in \mathcal{X}^n \,|\, \mathbf{x} = \psi_n(\varphi_n(\mathbf{x}))\}$. Then, for any $a_n > 0$ and any $\eta_n \geq 1$ it holds that

$$\begin{aligned} \delta_n(\eta_n) &\geq \Pr\left\{ \frac{P_{X^n}(X^n)}{\Pr\{X^n \in D_n\}} \leq a_n K^{-\eta_n},\ X^n \in D_n \right\} \\ &\quad - a_n K \Pr\left\{X^n \in D_n\right\}. \end{aligned} \tag{3.11}$$

Proof: Set $\tilde{B}_n$ as

$$\tilde{B}_n := \left\{ \mathbf{x} \in \mathcal{X}^n \,\middle|\, \frac{P_{X^n}(\mathbf{x})}{\Pr\{X^n \in D_n\}} \leq a_n K^{-\eta_n} \right\} \tag{3.12}$$

[1] When $\frac{P_{X^n}(\mathbf{x})}{\Pr\{X^n \in A_n\}} = 1$ holds, then $\lceil -\log \frac{P_{X^n}(\mathbf{x})}{\Pr\{X^n \in A_n\}} \rceil = 0$ holds. In this special case, we formally set $\varphi_n(\mathbf{x}) = 0$, and hence, $l(\varphi_n(\mathbf{x})) = 1$ holds.

and S_n as in (3.4). Then, we have

$$\begin{aligned}
&\Pr\left\{\frac{P_{X^n}(X^n)}{\Pr\{X^n \in D_n\}} \le a_n K^{-\eta_n},\ X^n \in D_n\right\}\\
&= \sum_{\mathbf{x}\in D_n\cap\tilde{B}_n\cap S_n} P_{X^n}(\mathbf{x}) + \sum_{\mathbf{x}\in D_n\cap\tilde{B}_n\cap(S_n)^c} P_{X^n}(\mathbf{x})\\
&\le \delta_n(\eta_n) + \sum_{\mathbf{x}\in D_n\cap(S_n)^c} a_n K^{-\eta_n}\Pr\left\{X^n \in D_n\right\}\\
&= \delta_n(\eta_n) + \left|D_n\cap(S_n)^c\right| a_n K^{-\eta_n}\Pr\left\{X^n \in D_n\right\}\\
&\le \delta_n(\eta_n) + a_n K \Pr\left\{X^n \in D_n\right\}, \qquad (3.13)
\end{aligned}$$

where the last inequality is due to (2.9). This completes the proof of the theorem. □

IV. First-Order (ε, δ)-Optimum Threshold

A. General Formula

In this section, we establish the general formula of the first-order (ε, δ)-optimum threshold by using Theorems 3.1 and 3.2. We define the quantity $G_{\varepsilon,\delta}(\mathbf{X})$ as

$$\begin{aligned}
&G_{\varepsilon,\delta}(\mathbf{X})\\
&:= \inf\left\{R \,\middle|\, \lim_{\nu\downarrow 0}\limsup_{n\to\infty} \inf_{\substack{A_n\subset\mathcal{X}^n:\\ \Pr\{X^n\in A_n\}\ge 1-\varepsilon-\nu}} \Pr\left\{-\frac{1}{n}\log\frac{P_{X^n}(X^n)}{\Pr\{X^n\in A_n\}} \ge R,\ X^n\in A_n\right\} \le \delta\right\}. \qquad (4.1)
\end{aligned}$$

Then, we have the following theorem:

Theorem 4.1 (First-order (ε, δ)-optimum threshold): For any $\varepsilon, \delta \in [0, 1)$ satisfying $\varepsilon + \delta < 1$, it holds that

$$R(\varepsilon, \delta|\mathbf{X}) = G_{\varepsilon,\delta}(\mathbf{X}). \qquad (4.2)$$

Proof: The proof consists of two parts.

(Direct Part:) Setting R_0 as $R_0 := G_{\varepsilon,\delta}(\mathbf{X})$, we show that for any $\gamma > 0$, $R = R_0 + 2\gamma$ is (ε, δ)-achievable. To do so, we arbitrarily fix $\nu \in (0, 1-\varepsilon]$ and use Theorem 3.1 with $a_n = K^{-n\gamma}$ and $\eta_n = nR = n(R_0 + 2\gamma)$. Let $\lambda_1 > \lambda_2 > \cdots \to 0$ be an arbitrary decreasing sequence. We choose $A_n \subseteq \mathcal{X}^n$ satisfying $\Pr\{X^n \in A_n\} \ge 1 - \varepsilon - \nu$ and

$$\begin{aligned}
&\Pr\left\{-\frac{1}{n}\log\frac{P_{X^n}(X^n)}{\Pr\{X^n\in A_n\}} \ge R_0+\gamma,\ X^n\in A_n\right\}\\
&\le \inf_{\substack{A_n\subset\mathcal{X}^n:\\ \Pr\{X^n\in A_n\}\ge 1-\varepsilon-\nu}} \Pr\left\{-\frac{1}{n}\log\frac{P_{X^n}(X^n)}{\Pr\{X^n\in A_n\}} \ge R_0+\gamma,\ X^n\in A_n\right\} + \lambda_n. \qquad (4.3)
\end{aligned}$$

Then, for this $A_n \subseteq \mathcal{X}^n$, from Theorem 3.1 there exists a variable-length code (φ_n, ψ_n) such that

$$\varepsilon_n \le \varepsilon + \nu, \qquad (4.4)$$

and

$$\begin{aligned}
\delta_n(nR) &= \Pr\left\{\frac{1}{n}l(\varphi_n(X^n)) > R_0 + 2\gamma\right\}\\
&\le \Pr\left\{-\frac{1}{n}\log\frac{P_{X^n}(X^n)}{\Pr\{X^n\in A_n\}} \ge R_0+\gamma,\ X^n\in A_n\right\}\\
&\quad + K^{-n\gamma+1}. \qquad (4.5)
\end{aligned}$$

It follows from (4.3) that

$$\begin{aligned}
&\limsup_{n\to\infty}\Pr\left\{\frac{1}{n}l(\varphi_n(X^n)) > R_0 + 2\gamma\right\}\\
&\le \limsup_{n\to\infty} \inf_{\substack{A_n\subset\mathcal{X}^n:\\ \Pr\{X^n\in A_n\}\ge 1-\varepsilon-\nu}} \Pr\left\{-\frac{1}{n}\log\frac{P_{X^n}(X^n)}{\Pr\{X^n\in A_n\}} \ge R_0+\gamma,\ X^n\in A_n\right\}\\
&\le \delta, \qquad (4.6)
\end{aligned}$$

where the last inequality follows immediately from the definition of R_0 because for any $\overline{\nu} < \nu$ it holds that

$$\begin{aligned}
&\inf_{\substack{A_n\subset\mathcal{X}^n:\\ \Pr\{X^n\in A_n\}\ge 1-\varepsilon-\nu}} \Pr\left\{-\frac{1}{n}\log\frac{P_{X^n}(X^n)}{\Pr\{X^n\in A_n\}} \ge R_0+\gamma,\ X^n\in A_n\right\}\\
&\le \inf_{\substack{A_n\subset\mathcal{X}^n:\\ \Pr\{X^n\in A_n\}\ge 1-\varepsilon-\overline{\nu}}} \Pr\left\{-\frac{1}{n}\log\frac{P_{X^n}(X^n)}{\Pr\{X^n\in A_n\}} \ge R_0+\gamma,\ X^n\in A_n\right\}\\
&\le \lim_{\overline{\nu}\downarrow 0} \inf_{\substack{A_n\subset\mathcal{X}^n:\\ \Pr\{X^n\in A_n\}\ge 1-\varepsilon-\overline{\nu}}} \Pr\left\{-\frac{1}{n}\log\frac{P_{X^n}(X^n)}{\Pr\{X^n\in A_n\}} \ge R_0+\gamma,\ X^n\in A_n\right\}. \qquad (4.7)
\end{aligned}$$

From (4.4), (4.5) and (4.6), the direct part has been proved.

(Converse Part:) We assume that R is (ε, δ)-achievable. Then, it holds that

$$\limsup_{n\to\infty} \varepsilon_n \le \varepsilon, \qquad (4.8)$$

$$\limsup_{n\to\infty} \delta_n(nR) \le \delta. \qquad (4.9)$$

By using Theorem 3.2 with $a_n = K^{-n\gamma}$ $(\forall\gamma > 0)$ and $\eta_n = nR$, we have

$$\begin{aligned}
\delta_n(nR) \ge \Pr&\left\{\frac{P_{X^n}(X^n)}{\Pr\{X^n\in D_n\}} \le K^{-n(R+\gamma)},\ X^n\in D_n\right\}\\
&-K^{-n\gamma+1}, \qquad (4.10)
\end{aligned}$$

where $D_n = \{\mathbf{x} \in \mathcal{X}^n \,|\, \mathbf{x} = \psi_n(\varphi_n(\mathbf{x}))\}$. Here, (4.8) means that for any $\nu \in (0, 1-\varepsilon)$ there exists n_0 such that $\Pr\{X^n \in D_n\} \ge 1 - \varepsilon - \nu$ $(\forall n > n_0)$ holds. Thus, for any $\nu \in (0, 1-\varepsilon)$ we have

$$\begin{aligned}
&\limsup_{n\to\infty} \delta_n(nR)\\
&\ge \limsup_{n\to\infty}\Pr\left\{-\frac{1}{n}\log\frac{P_{X^n}(X^n)}{\Pr\{X^n\in D_n\}} \ge R+\gamma,\ X^n\in D_n\right\}\\
&\ge \limsup_{n\to\infty} \inf_{\substack{A_n\subset\mathcal{X}^n:\\ \Pr\{X^n\in A_n\}\ge 1-\varepsilon-\nu}} \Pr\left\{-\frac{1}{n}\log\frac{P_{X^n}(X^n)}{\Pr\{X^n\in A_n\}} \ge R+\gamma,\ X^n\in A_n\right\}. \qquad (4.11)
\end{aligned}$$

Substituting this inequality into (4.9), we obtain

$$\delta \geq \limsup_{n\to\infty} \inf_{\substack{A_n \subset \mathcal{X}^n: \\ \Pr\{X^n \in A_n\} \geq 1-\varepsilon-\nu}} \Pr\left\{-\frac{1}{n}\log\frac{P_{X^n}(X^n)}{\Pr\{X^n \in A_n\}} \geq R+\gamma,\ X^n \in A_n\right\}. \quad (4.12)$$

This means that

$$R+\gamma \geq \inf\left\{R \,\middle|\, \lim_{\nu\downarrow 0}\limsup_{n\to\infty} \inf_{\substack{A_n \subset \mathcal{X}^n: \\ \Pr\{X^n \in A_n\} \geq 1-\varepsilon-\nu}} \Pr\left\{-\frac{1}{n}\log\frac{P_{X^n}(X^n)}{\Pr\{X^n \in A_n\}} \geq R, X^n \in A_n\right\} \leq \delta\right\}, \quad (4.13)$$

which implies that the converse part holds. □

B. Another Expression of the General Formula

The *general formula* of the (ε, δ)-optimum thresholds derived in the previous subsection seems complicated and hard to compute even for tractable sources such as stationary memoryless sources, Markov sources and so on. Hence in this subsection, we derive another expression of $G_{\varepsilon,\delta}(\mathbf{X})$ which enables us not only to compute the (ε, δ)-optimum thresholds for tractable sources but also to understand the structure of the optimum overflow thresholds in the ε-variable-length coding. As a result, the relationship with the γ-optimum coding rate in the fixed-length coding is revealed.

Set

$$F(\varepsilon, R) := \limsup_{n\to\infty} \inf_{\substack{A_n \subset \mathcal{X}^n: \\ \Pr\{X^n \in A_n\} \geq 1-\varepsilon}} \Pr\left\{-\frac{1}{n}\log P_{X^n}(X^n) \geq R,\ X^n \in A_n\right\}. \quad (4.14)$$

Then, the function $F(\varepsilon, R)$ is a monotonically nonincreasing function of ε and R. Then, we have the following lemma:

Lemma 4.1:

$$G_{\varepsilon,\delta}(\mathbf{X}) = \inf\left\{R \,\middle|\, \lim_{\nu\downarrow 0} F(\varepsilon+\nu, R) \leq \delta\right\}. \quad (4.15)$$

Proof: We first fix ν as $\nu \in (0, 1-\varepsilon)$ and $A_n \subset \mathcal{X}^n$ satisfying $\Pr\{X^n \in A_n\} \geq 1-\varepsilon-\nu$. Then, for any $\mathbf{x} \in \mathcal{X}^n$ it holds that

$$\begin{aligned}
&\left|-\frac{1}{n}\log P_{X^n}(\mathbf{x}) + \frac{1}{n}\log\frac{P_{X^n}(\mathbf{x})}{\Pr\{X^n \in A_n\}}\right| \\
&\leq \left|-\frac{1}{n}\log P_{X^n}(\mathbf{x}) + \frac{1}{n}\log\frac{P_{X^n}(\mathbf{x})}{1-\varepsilon-\nu}\right| \\
&= \frac{1}{n}\log\frac{1}{1-\varepsilon-\nu}. \qquad (4.16)
\end{aligned}$$

Thus, noting that for any $\gamma > 0$ it holds that $\frac{1}{n}\log\frac{1}{1-\varepsilon-\nu} < \gamma$ $(\forall n > n_0)$, which implies that the difference between $-\frac{1}{n}\log P_{X^n}(\mathbf{x})$ and $-\frac{1}{n}\log\frac{P_{X^n}(\mathbf{x})}{\Pr\{\ X^n \in A_n\}}$ becomes arbitrarily small, we obtain the lemma. □

Here, we define two quantities

$$\tilde{H}_{\varepsilon,\delta}(\mathbf{X}) := \inf\{R \,|\, F(\varepsilon, R) \leq \delta\}, \quad (4.17)$$

and

$$\overline{H}_\gamma(\mathbf{X}) := \inf\left\{R \,\middle|\, \limsup_{n\to\infty}\Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} > R\right\} \leq \gamma\right\}. \quad (4.18)$$

It should be noted that the γ-optimum coding rate in the fixed length coding is characterized by $\overline{H}_\gamma(\mathbf{X})$ [3] (cf. Remark 4.1 below).

Then, the following theorem holds:

Theorem 4.2: For any $\varepsilon, \delta \in [0, 1)$ satisfying $\varepsilon + \delta < 1$, it holds that

$$G_{\varepsilon,\delta}(\mathbf{X}) = \tilde{H}_{\varepsilon,\delta}(\mathbf{X}) = \overline{H}_{\varepsilon+\delta}(\mathbf{X}). \quad (4.19)$$

□

Since $\overline{H}_\gamma(\mathbf{X})$ is a right-continuous function of γ (see, [3]), the function $G_{\varepsilon,\delta}(\mathbf{X})$ and $\tilde{H}_{\varepsilon,\delta}(\mathbf{X})$ are also right-continuous functions of ε and δ.

From the theorem we immediately have:

Corollary 4.1: Fix $\varepsilon \in [0, 1)$ arbitrarily. Then, for any $\varepsilon_1, \varepsilon_2 \geq 0$ satisfying $\varepsilon_1 + \varepsilon_2 = \varepsilon$ it holds that

$$\tilde{H}_{\varepsilon_1,\varepsilon_2}(\mathbf{X}) = \overline{H}_\varepsilon(\mathbf{X}).$$

□

Remark 4.1: The ε-optimum coding rate in the fixed-length coding and the (first-order) ε-optimum threshold in the variable-length coding *without error* coincide with $R(\varepsilon, 0|\mathbf{X})$ and $R(0, \varepsilon|\mathbf{X})$, respectively. Then, the following relation has already been shown in [13], [14].

$$R(\varepsilon, 0|\mathbf{X}) = R(0, \varepsilon|\mathbf{X}) = \overline{H}_\varepsilon(\mathbf{X}). \quad (4.20)$$

This equality reveals a deep relationship between the fixed-length coding and the variable-length coding *without error*. From the above equality and Corollary 4.1, for any $\varepsilon_1, \varepsilon_2 \geq 0$ satisfying $\varepsilon_1 + \varepsilon_2 = \varepsilon$, we obtain the following relation

$$R(\varepsilon_1, \varepsilon_2|\mathbf{X}) = R(\varepsilon, 0|\mathbf{X}) = R(0, \varepsilon|\mathbf{X}). \quad (4.21)$$

It should be emphasized that the relation (4.21) subsumes (4.20), because ε_1 and ε_2 in (4.21) may be arbitrary nonnegative numbers satisfying $\varepsilon_1 + \varepsilon_2 = \varepsilon$. This relation can also be obtained from the fact that the r.h.s. of (2.11) coincides with $\overline{H}_{\varepsilon+\delta}(\mathbf{X})$ [12].

Remark 4.2: In [6], the optimum *zero-error* variable-length code has been discussed. We can consider the optimum variable-length code *allowing errors* as an immediate extension of the argument in [6]. Then, by using this optimum code and the argument developed in [6], we can also show (4.21).

Before proving Theorem 4.2, we show the following lemma.

Lemma 4.2: For any $\nu > 0$, it holds that

$$\tilde{H}_{\varepsilon+\nu,\delta}(\mathbf{X}) \leq G_{\varepsilon,\delta}(\mathbf{X}) = \lim_{\nu\downarrow 0}\tilde{H}_{\varepsilon+\nu,\delta}(\mathbf{X}) \leq \tilde{H}_{\varepsilon,\delta}(\mathbf{X}). \quad (4.22)$$

Proof of Lemma 4.2: Since $\tilde{H}_{\varepsilon,\delta}(\mathbf{X})$ is a monotonically nonincreasing function of ε, the first and second inequalities hold immediately from Lemma 4.1. Hence, it suffices to show the intermediate equality.

Since $F(\varepsilon, R)$ is a monotonically nonincreasing function of ε and R, $G_{\varepsilon,\delta}(\mathbf{X})$ can be expressed as

$$G_{\varepsilon,\delta}(\mathbf{X}) = \inf \bigcap_{\nu>0} \{R \,|\, F(\varepsilon+\nu, R) \le \delta\}. \tag{4.23}$$

For any $\nu > 0$, we have

$$\begin{aligned} G_{\varepsilon,\delta}(\mathbf{X}) &\ge \inf \{R \,|\, F(\varepsilon+\nu, R) \le \delta\} \\ &= \tilde{H}_{\varepsilon+\nu,\delta}(\mathbf{X}), \end{aligned} \tag{4.24}$$

which implies

$$G_{\varepsilon,\delta}(\mathbf{X}) \ge \sup_{\nu>0} \tilde{H}_{\varepsilon+\nu,\delta}(\mathbf{X}) = \lim_{\nu\downarrow 0} \tilde{H}_{\varepsilon+\nu,\delta}(\mathbf{X}). \tag{4.25}$$

On the otherhand, again for any $\nu > 0$, we have

$$\begin{aligned} \lim_{\nu\downarrow 0} \tilde{H}_{\varepsilon+\nu,\delta}(\mathbf{X}) &\ge \tilde{H}_{\varepsilon+\nu,\delta}(\mathbf{X}) \\ &= \inf \{R | F(\varepsilon+\nu, R) \le \delta\}. \end{aligned} \tag{4.26}$$

This inequality for an arbitrary fixed $\nu > 0$ implies that

$$\begin{aligned} \lim_{\nu\downarrow 0} \tilde{H}_{\varepsilon+\nu,\delta}(\mathbf{X}) &\ge \inf \bigcap_{\nu>0} \{R | F(\varepsilon+\nu, R) \le \delta\} \\ &= G_{\varepsilon,\delta}(\mathbf{X}), \end{aligned} \tag{4.27}$$

where the equality is due to (4.23). Combining (4.25) and (4.27) yields

$$G_{\varepsilon,\delta}(\mathbf{X}) = \lim_{\nu\downarrow 0} \tilde{H}_{\varepsilon+\nu,\delta}(\mathbf{X}). \tag{4.28}$$

This completes the proof of the lemma. □

Proof of Theorem 4.2: From Lemma 4.2, it suffices to prove two inequalities:

$$G_{\varepsilon,\delta}(\mathbf{X}) \ge \overline{H}_{\varepsilon+\delta}(\mathbf{X}), \tag{4.29}$$

$$\tilde{H}_{\varepsilon,\delta}(\mathbf{X}) \le \overline{H}_{\varepsilon+\delta}(\mathbf{X}). \tag{4.30}$$

(Proof of (4.29):) For any fixed $R > G_{\varepsilon,\delta}(\mathbf{X})$, we show that $R \ge \overline{H}_{\varepsilon+\delta}(\mathbf{X})$ holds.

Set $J_n \subseteq \mathcal{X}^n$ as

$$J_n := \left\{ \mathbf{x} \in \mathcal{X}^n \,\middle|\, -\frac{1}{n} \log P_{X^n}(\mathbf{x}) \ge R \right\}. \tag{4.31}$$

Then, by the assumption $R > G_{\varepsilon,\delta}(\mathbf{X})$ and (4.15), it holds that

$$\begin{aligned} \delta &\ge \lim_{\nu\downarrow 0} \limsup_{n\to\infty} \inf_{\substack{A_n \subset \mathcal{X}^n: \\ \Pr\{X^n \in A_n\} \ge 1-\varepsilon-\nu}} \Pr\left\{X^n \in A_n \cap J_n\right\} \\ &\ge \limsup_{n\to\infty} \inf_{\substack{A_n \subset \mathcal{X}^n: \\ \Pr\{X^n \in A_n\} \ge 1-\varepsilon-\nu}} \Pr\left\{X^n \in A_n \cap J_n\right\} \end{aligned} \tag{4.32}$$

for any $\nu \in (0, 1-\varepsilon)$. Moreover, we define a subset $K_n \subseteq \mathcal{X}^n$ such that

$$\Pr\left\{X^n \in K_n\right\} \ge 1 - \varepsilon - \nu \tag{4.33}$$

and

$$\begin{aligned} &\Pr\left\{X^n \in J_n \cap K_n\right\} \\ &\le \inf_{\substack{A_n \subset \mathcal{X}^n: \\ \Pr\{X^n \in A_n\} \ge 1-\varepsilon-\nu}} \Pr\left\{X^n \in A_n \cap J_n\right\} + \nu \end{aligned} \tag{4.34}$$

hold. Then, from (4.32) it holds that

$$\limsup_{n\to\infty} \Pr\left\{X^n \in J_n \cap K_n\right\} \le \delta + \nu. \tag{4.35}$$

On the other hand, from (4.33), we have

$$\begin{aligned} \Pr\left\{X^n \in J_n\right\} &\le \Pr\left\{X^n \in J_n \cap K_n\right\} + \Pr\left\{X^n \in (K_n)^c\right\} \\ &\le \Pr\left\{X^n \in J_n \cap K_n\right\} + \varepsilon + \nu. \end{aligned} \tag{4.36}$$

This means that

$$\begin{aligned} &\limsup_{n\to\infty} \Pr\left\{X^n \in J_n\right\} \\ &\le \limsup_{n\to\infty} \Pr\left\{X^n \in J_n \cap K_n\right\} + \varepsilon + \nu \\ &\le \delta + \varepsilon + 2\nu. \end{aligned} \tag{4.37}$$

Since $\nu \in (0, 1-\varepsilon)$ is arbitrarily, we have

$$\limsup_{n\to\infty} \Pr\left\{X^n \in J_n\right\} \le \delta + \varepsilon, \tag{4.38}$$

which implies $R \ge \overline{H}_{\varepsilon+\delta}(\mathbf{X})$.

(Proof of (4.30):) For any fixed $R > \overline{H}_{\varepsilon+\delta}(\mathbf{X})$, we show that $R \ge \tilde{H}_{\varepsilon,\delta}(\mathbf{X})$ holds. We also use the set J_n defined in (4.31). Since $R > \overline{H}_{\varepsilon+\delta}(\mathbf{X})$ holds, we have

$$\limsup_{n\to\infty} \Pr\left\{X^n \in J_n\right\} \le \varepsilon + \delta. \tag{4.39}$$

Here, without loss of generality we assume that the elements of X^n be ordered $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \cdots\}$ *with decreasing probabilities*, that is, for any $i < j$, $P_{X^n}(\mathbf{x}_i) \ge P_{X^n}(\mathbf{x}_j)$ holds. We define a positive integer i^* and $L_n = \{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_{i^*}\}$ such that

$$\Pr\left\{X^n \in L_n \setminus \{\mathbf{x}_{i^*}\}\right\} = \sum_{i=1}^{i^*-1} P_{X^n}(\mathbf{x}_i) < 1 - \varepsilon, \tag{4.40}$$

$$\Pr\left\{X^n \in L_n\right\} = \sum_{i=1}^{i^*} P_{X^n}(\mathbf{x}_i) \ge 1 - \varepsilon. \tag{4.41}$$

We then evaluate the probability

$$\begin{aligned} &\Pr\left\{X^n \in J_n \cap L_n\right\} \\ &= \Pr\left\{X^n \in J_n\right\} - \Pr\left\{X^n \in J_n \cap (L_n)^c\right\}. \end{aligned} \tag{4.42}$$

By the definition of J_n, for all n satisfying $\frac{1}{n} \log \frac{1}{P_{X^n}(\mathbf{x}_{i^*})} < R$, it holds that

$$\Pr\left\{X^n \in J_n \cap L_n\right\} = 0. \tag{4.43}$$

On the other hand, for all n satisfying $\frac{1}{n} \log \frac{1}{P_{X^n}(\mathbf{x}_{i^*})} \ge R$,

$$P_{X^n}(\mathbf{x}_{i^*}) \le e^{-nR} \tag{4.44}$$

and $(L_n)^c \subseteq J_n$ hold. Hence, we have

$$\begin{aligned} \Pr\left\{X^n \in J_n \cap (L_n)^c\right\} &= \Pr\left\{X^n \in (L_n)^c\right\} \\ &= \Pr\left\{X^n \in (L_n \setminus \{\mathbf{x}_{i^*}\})^c\right\} - P_{X^n}(\mathbf{x}_{i^*}) \\ &\ge \varepsilon - e^{-nR}, \end{aligned} \tag{4.45}$$

where the last inequality is due to (4.40) and (4.44). Thus, from (4.42), (4.43), and (4.45) we have

$$\begin{aligned} &\Pr\left\{X^n \in J_n \cap L_n\right\} \\ &\le \max\left\{0, \Pr\left\{X^n \in J_n\right\} - \left(\varepsilon - e^{-nR}\right)\right\} \end{aligned} \tag{4.46}$$

for all n. Taking $\limsup_{n\to\infty}$ on both sides yields

$$\limsup_{n\to\infty} \Pr\left\{X^n \in J_n \cap L_n\right\} \leq \max\left\{0, \limsup_{n\to\infty} \Pr\left\{X^n \in J_n\right\} - \varepsilon\right\} \leq \delta, \quad (4.47)$$

where the last inequality is due to (4.39).

Since L_n satisfies (4.41) we obtain

$$\limsup_{n\to\infty} \inf_{\substack{A_n \subset \mathcal{X}^n: \\ \Pr\{X^n \in A_n\} \geq 1-\varepsilon}} \Pr\left\{X^n \in A_n \cap J_n\right\} \leq \delta. \quad (4.48)$$

This implies $R \geq \tilde{H}_{\varepsilon,\delta}(\mathbf{X})$. □

V. Second-Order (ε, δ, R)-Optimum Threshold

In this section, we establish the general formula of the second-order (ε, δ, R)-optimum threshold. We define the quantity $G_{\varepsilon,\delta}(R|\mathbf{X})$ as

$$G_{\varepsilon,\delta}(R|\mathbf{X}) := \inf\left\{L \,\middle|\, \lim_{\nu\downarrow 0} \limsup_{n\to\infty} \inf_{\substack{A_n \subset \mathcal{X}^n: \\ \Pr\{X^n \in A_n\} \geq 1-\varepsilon-\nu}} \Pr\left\{-\frac{1}{n}\log\frac{P_{X^n}(X^n)}{\Pr\{X^n \in A_n\}} \geq R + \frac{L}{\sqrt{n}}, X^n \in A_n\right\} \leq \delta\right\}. \quad (5.1)$$

Theorem 5.1 (Second-order (ε, δ, R)-optimum threshold): For any $\varepsilon, \delta \in [0, 1)$ satisfying $\varepsilon + \delta < 1$, it holds that

$$L(\varepsilon, \delta, R|\mathbf{X}) = G_{\varepsilon,\delta}(R|\mathbf{X}).$$

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

As in the *first-order* case (cf. Remark 4.1), $G_{\varepsilon,\delta}(R|\mathbf{X})$ can also be expressed as in another way. Let us define an information-spectrum quantity:

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

Theorem 5.2: For any $\varepsilon, \delta \in [0, 1)$ satisfying $\varepsilon + \delta < 1$, it holds that

$$G_{\varepsilon,\delta}(R|\mathbf{X}) = \overline{H}_{\varepsilon+\delta}(R|\mathbf{X}).$$

Proof: The proof is similar to that of Theorem 4.2. □

The *second-order* (ε, R)-optimum coding rate in the fixed-length coding is characterized by $\overline{H}_\varepsilon(R|\mathbf{X})$ [13]. This means that in order to compute $G_{\varepsilon,\delta}(R|\mathbf{X})$ for some specified sources such as i.i.d. sources and Markov sources, we can use the similar technique as the one in [13] (see, Section VI).

VI. Application to Stationary Memoryless Sources

In this section we compute the optimum overflow thresholds for the stationary memoryless source with generic distribution X by using *general formulas* obtained in the preceding sections.

A. First-Order (ε, δ)-Optimum Thresholds

For a stationary memoryless source X, the following theorem is well-known.

Theorem 6.1 (Steinberg and Verdú [15]): For any $\gamma \in [0, 1)$, it holds that

$$\overline{H}_\gamma(\mathbf{X}) = H(X), \quad (6.1)$$

where $H(X)$ denotes the entropy of the source X.

From Theorems 4.1, 4.2, and 6.1, we immediately have

Theorem 6.2: For any $\varepsilon, \delta \in [0, 1)$ satisfying $\varepsilon + \delta < 1$ it holds that

$$R(\varepsilon, \delta|\mathbf{X}) = H(X) \quad (0 \leq \varepsilon, \delta < 1). \quad (6.2)$$

Thus, the (ε, δ)-optimum overflow thresholds equals to the entropy of the source irrespective of ε and δ.

B. Second-Order (ε, δ, R)-Optimum Overflow Thresholds

In the second-order coding rate analysis, it is well-known that:

Theorem 6.3 (Hayashi [13]):

$$\overline{H}_\gamma(R|X) = \begin{cases} -\infty & R > H(X) \\ +\infty & R < H(X) \\ \sqrt{V_X}\Phi^{-1}(\gamma) & R = H(X), \end{cases} \quad (6.3)$$

where

$$\Phi(x) := \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-\frac{t^2}{2}}dt \quad (6.4)$$

is the complementary standard Gaussian distribution function and

$$V_X := \sum_{x\in\mathcal{X}} P_X(x)\left(-\log P_X(x) - H(X)\right)^2. \quad (6.5)$$

denotes the variance of the self-information called *varentropy* of the source [6]. □

As is known from the above theorem, the setting of *first-order* constant R is quite important to analyze the second-order (ε, δ, R)-optimum overflow thresholds. In this paper, we consider the following two case:

Case 1) Setting R as the first-order (ε, δ)-optimum thresholds R_1:

In this case, from Theorem 4.1 $R_1 = \overline{H}_{\varepsilon+\delta}(\mathbf{X})$ holds. Thus, from Theorem 6.1 we set

$$R_1 = H(X) \quad (6.6)$$

for the stationary memoryless source with generic distribution X.

Case 2) Setting R as the optimum mean codeword length R_2:

The optimum mean codeword length of the ε-variable-length codes has been first determined by Koga and Yamamoto [7] in the case that $\varepsilon \in [0, 1)$, while Han [3], [16] has derived it in the case of $\varepsilon = 0$.

From the result in [7] we shall set

$$R_2 = H_{[\varepsilon]}(\mathbf{X}) := \lim_{\nu \downarrow 0} \limsup_{n\to\infty} \frac{1}{n} \inf_{\substack{A_n \subset \mathcal{X}^n: \\ \Pr\{X^n \in A_n\} \geq 1-\varepsilon-\nu}} \sum_{\mathbf{x}\in A_n} P_{X^n}(\mathbf{x}) \log \frac{1}{P_{X^n}(\mathbf{x})}. \tag{6.7}$$

In particular,

$$R_2 = (1-\varepsilon)H(X) \tag{6.8}$$

holds for the stationary memoryless source with generic distribution X [7], [9].

Remark 6.1: The optimum *second-order* mean codeword length $L_{[\varepsilon]}(H_{[\varepsilon]}(\mathbf{X})|\mathbf{X})$ for the stationary memoryless source X has been determined by Kostina *et al.* [8] as follows.

Assuming that the third absolute moment of $-\log P_X(X)$ is finite, then for any given $\varepsilon \in (0,1)$, it holds that

$$L_{[\varepsilon]}(H_{[\varepsilon]}(\mathbf{X})|\mathbf{X}) = -\sqrt{\frac{V(X)}{2\pi}} e^{-\frac{(\Phi^{-1}(\varepsilon))^2}{2}}. \tag{6.9}$$

The above result shows an interesting phenomenon in which the optimum *second-order* mean codeword length is always negative.

Summarizing up, we set R as $R_1 = H(X)$ and $R_2 = (1-\varepsilon)H(X)$.

Then, we obtain the following theorem.

Theorem 6.4: For any $\varepsilon, \delta \in [0,1)$ satisfying $\varepsilon + \delta < 1$, it holds that

$$L(\varepsilon, \delta, R_1|\mathbf{X}) = \sqrt{V_X}\Phi^{-1}(\varepsilon+\delta), \tag{6.10}$$

$$L(\varepsilon, \delta, R_2|\mathbf{X}) = \begin{cases} \sqrt{V_X}\Phi^{-1}(\delta) & \varepsilon = 0 \\ +\infty & \varepsilon \neq 0. \end{cases} \tag{6.11}$$

Proof: From Theorems 5.1, 5.2, and 6.3 we obtain (6.10) as well as (6.11) in the case of $\varepsilon = 0$. On the other hand, when we consider the case of $L(\varepsilon, \delta, R_2|\mathbf{X})$ with $\varepsilon > 0$, it holds that $R_2 = (1-\varepsilon)H(X) < H(X)$. Thus, from Theorems 5.1 and 5.2, and the definition of $\overline{H}_{\varepsilon+\delta}(R_2|\mathbf{X})$ it holds that

$$\begin{aligned} &L(\varepsilon, \delta, R_2|\mathbf{X}) \\ &= \overline{H}_{\varepsilon+\delta}(R_2|\mathbf{X}) \\ &= \inf\left\{L \,\middle|\, \limsup_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} > R_2 + \frac{L}{\sqrt{n}}\right\} \leq \varepsilon+\delta\right\} \\ &= \inf\left\{L \,\middle|\, \limsup_{n\to\infty} \right. \\ &\qquad \left. \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} > (1-\varepsilon)H(X) + \frac{L}{\sqrt{n}}\right\} \leq \varepsilon+\delta\right\}. \end{aligned} \tag{6.12}$$

Here, from the law of large numbers this case necessarily yields that

$$\lim_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} > (1-\varepsilon)H(X) + \frac{L}{\sqrt{n}}\right\} = 1, \tag{6.13}$$

for any constant $L < \infty$. Hence, in this case we set formally as $L(\varepsilon, \delta, R_2|\mathbf{X}) = +\infty$. □

The above theorem shows that if we set R as $R = R_2$ (the optimum mean codeword length of ε-variable-length codes), the second-order (ε, δ, R)-optimum thresholds is always infinity as long as $\varepsilon > 0$. This means that the error probability or the overflow probability cannot be less than or equal to the desired value irrespective with the second-order threshold L. Moreover, from the similar argument to the proof of Theorem 6.4, we observe that $L(\varepsilon, \delta, R|\mathbf{X}) = \infty$ for $R < H(X)$, and $L(\varepsilon, \delta, R|\mathbf{X}) = -\infty$ for $R > H(X)$. Hence, in order to analyze the *second-order* optimum threshold L in the ε-variable-length coding, the *first-order* rate R should be set as the first-order optimum threshold: $R_1 = H(X)$.

VII. Concluding Remarks

We have so far considered the *first-* and *second-order* achievability to evaluate the optimum overflow thresholds in the ε-variable-length coding problem. As shown in the proofs of this paper, the information spectrum approach is substantial in analyses. In particular, Theorems 3.1 and 3.2 enable us to analyze the *first-* and *second-order* optimum overflow thresholds by the unified approach. In addition, we can apply these theorems into the case of the optimistic coding scenario [17], [18]. For example, the first-order achievability in the *optimistic* scenario is defined by using the following conditions instead of (2.3):

$$\limsup_{n\to\infty} \varepsilon_n \leq \varepsilon, \quad \liminf_{n\to\infty} \delta_n(nR) \leq \delta. \tag{7.1}$$

Then, we can show that the first-order *optimistic* optimum overflow thresholds is characterized by $G^*_{\varepsilon,\delta}(\mathbf{X})$ in which the $\limsup_{n\to\infty}$ in the definition of $G_{\varepsilon,\delta}(\mathbf{X})$ (eq. (4.1)) is replaced by $\liminf_{n\to\infty}$. Analogous to Theorem 4.2, $G^*_{\varepsilon,\delta}(\mathbf{X})$ is equal to $\overline{H}^*_{\varepsilon,\delta}(\mathbf{X})$ in which the $\limsup_{n\to\infty}$ in the definition of $\overline{H}^*_{\varepsilon,\delta}(\mathbf{X})$ is again replaced by $\liminf_{n\to\infty}$.

We have also clarified that the relationship between the ε-variable-length coding and the γ-fixed-length coding. In the γ-fixed-length coding problem, the γ-optimum coding rate has already been derived for several tractable sources such as stationary memoryless sources, Markov sources and mixed sources [13], [19]. We can use these results to compute the (ε, δ)-optimum thresholds as well as the (ε, δ, R)-optimum thresholds in the ε-variable-length coding. Actually, in this paper we compute the optimum thresholds in the ε-variable-length coding for the stationary memoryless source by using the previous results for the γ-fixed-length coding.

Acknowledgement

The authors would like to thank Associate Editor, Prof. Ioannis Kontoyiannis and the anonymous reviewers for their valuable comments which have improved the content of this paper.

References

[1] R. Nomura and H. Yagi, "Overflow probability of variable-length codes allowing non-vanishing error probability," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Nov. 2017, pp. 589–593.

[2] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, Jul/Oct. 1948.

[3] T. S. Han, *Information-Spectrum Methods in Information Theory*. Springer, New York, NY, USA: 2003.
[4] N. Merhav, “Universal coding with minimum probability of codeword length overflow,” *IEEE Trans. Inf. Theory*, vol. 37, no. 3, pp. 556–563, May 1991.
[5] O. Uchida and T. S. Han, “The optimal overflow and underflow probabilities of variable-length coding for the general source,” *IEICE Trans. Fundam. Electron., Commun. Comput. Sci.*, vol. 84, no. 10, pp. 2457–2465, Oct. 2001.
[6] I. Kontoyiannis and S. Verdú, “Optimal lossless data compression: Non-asymptotics and asymptotics,” *IEEE Trans. Inf. Theory*, vol. 60, no. 2, pp. 777–795, Feb. 2014.
[7] H. Koga and H. Yamamoto, “Asymptotic properties on codeword lengths of an optimal FV code for general sources,” *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1546–1555, Apr. 2005.
[8] V. Kostina, Y. Polyanskiy, and S. Verdú, “Variable-length compression allowing errors,” *IEEE Trans. Inf. Theory*, vol. 61, no. 8, pp. 4316–4330, Aug. 2015.
[9] H. Yagi and R. Nomura, “Variable-length coding with cost allowing non-vanishing error probability,” *IEICE Trans. Fundam. Electron., Commun. Comput. Sci.*, vol. 100, no. 8, pp. 1683–1692, Aug. 2017.
[10] O. Kosut and L. Sankar, “Universal fixed-to-variable source coding in the finite blocklength regime,” in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2013, pp. 649–653.
[11] S. Saito and T. Matsushima, “Threshold of overflow probability in terms of smooth max-entropy for variable-length compression allowing errors,” in *Proc. Int. Symp. Inf. Theory Its Appl. (ISITA)*, Nov. 2016, pp. 21–25.
[12] T. Uyematsu, “A new unified method for fixed-length source coding problems of general sources,” *IEICE Trans. Fundam. Electron., Commun. Comput. Sci.*, vol. 93, no. 11, pp. 1868–1877, Nov. 2010.
[13] 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.
[14] R. Nomura and T. Matsushima, “On the overflow probability of fixed-to-variable length codes with side information,” *IEICE Trans. Fundam. Electron., Commun. Comput. Sci.*, vol. 94, no. 11, pp. 2083–2091, Nov. 2011.
[15] 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.
[16] T. S. Han, “Weak variable-length source coding,” *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1217–1226, Jul. 2000.
[17] S. Vembu, S. Verdu, and Y. Steinberg, “The source-channel separation theorem revisited,” *IEEE Trans. Inf. Theory*, vol. 41, no. 1, pp. 44–54, Jan. 1995.
[18] 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, Nov. 1999.
[19] R. Nomura and T. S. 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.

Ryo Nomura (M’00) received B.Eng., M.Eng., and D.Eng. degrees from Waseda University, Japan in 1996, 1998 and 2007, respectively. From 2000 to 2003 he was a Research Associate with School of Science and Engineering, Waseda University, Tokyo, Japan; from 2004 to 2010, Research Associate then Assistant Professor with College of Science and Engineering, Aoyama Gakuin University, Kanagawa, Japan; from 2012 to 2018, lecturer, Associate Professor and then Professor with School of Network and Information, Senshu University, Kanagawa, Japan. Since 2019, he has been a Professor at Center for Data Science, Waseda University, Tokyo, Japan. His research interests include information theory and its applications, especially source coding and Shannon Theory.

Hideki Yagi (S’03–M’05) received the B.E. degree, the M.E. degree, and the Ph.D. degree in industrial and management systems engineering from Waseda University, Tokyo, Japan in 2001, 2003, and 2005, respectively. He is currently an Associate Professor at the Department of Computer and Network Engineering, University of Electro-Communications, Tokyo, Japan. He was with Media Network Center, Waseda University as a Research Associate from 2005 to 2007, and an Assistant Professor from 2007 to 2008. In the winter of 2008 and from July, 2010 to January, 2011, he was a Visiting Fellow at Princeton University. His research interests include multi-user information and coding theory and information theoretic security.