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 $\epsilon > 0$, which we call the $\epsilon$-variable-length coding. The first-order optimum mean codeword length of $\epsilon$-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 $\epsilon$-variable-length codes. They have revealed that the second-order optimum mean codeword length of the $\epsilon$-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 $\epsilon$-variable-length codes.

Inspired by the result in [8], we also focus on the the $\epsilon$-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 $\epsilon$-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 $\epsilon$-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: nomru@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

0018-9448 © 2019 IEEE. Translations and content mining are permitted for academic research only. Personal use is also permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
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 \( X = \{ X^n = (X_1^{(n)}, X_2^{(n)}, \ldots, X_n^{(n)}) \}_{n=1}^{\infty} \) 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 \( \psi_{n}: \mathcal{X}^{\infty} \rightarrow \mathcal{U}^* \) and \( \eta_{n}: \mathcal{U}^{*} \rightarrow \mathcal{X}^{\infty} \) denote a variable-length encoder and a decoder respectively, where \( \mathcal{U} = \{1, 2, \ldots, K\} \) 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 \( x \in \mathcal{X}^n \) is denoted by \( l(\psi_{n}(x)) \) when we use the encoder \( \psi_{n} \).

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

**Definition 2.1:** The error probability of \( (\psi_{n}, \eta_{n}) \) and the overflow probability of \( (\psi_{n}, \eta_{n}) \) with threshold \( \eta_{n} \) are respectively defined as

\[
\varepsilon_{n} := \Pr \{ X^n \neq \psi_{n}(\eta_{n}(X^n)) \},
\]
\[
\delta_{n}(\eta_{n}) := \Pr \{ l(\psi_{n}(X^n)) > \eta_{n} \}.
\]

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 \( X \) and a threshold \( \eta_{n} \) we cannot minimize \( \varepsilon_{n} \) 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 \((\varepsilon, \delta)\)-achievable \((\varepsilon, \delta \in [0, 1])\), if there exists a sequence of variable-length code \((\psi_{n}, \eta_{n})\) such that

\[
\limsup_{n \to \infty} \varepsilon_{n} \leq \varepsilon, \quad \limsup_{n \to \infty} \delta_{n}(nR) \leq \delta.
\]

**Definition 2.3 (First-order \((\varepsilon, \delta)\)-optimum threshold):**

\[
R(\varepsilon, \delta | X) := \inf \{ R | R \text{ is } (\varepsilon, \delta)\text{-achievable} \}.
\]

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

**Definition 2.4:** Rate \( L \) is said to be \((\varepsilon, \delta, R)\)-achievable \((\varepsilon, \delta \in [0, 1], R \geq 0)\), if there exists a sequence of variable-length code \((\rho_{n}, \nu_{n})\) such that

\[
\limsup_{n \to \infty} \varepsilon_{n} \leq \varepsilon, \quad \limsup_{n \to \infty} \delta_{n}(nR + \sqrt{n}L) \leq \delta.
\]

**Definition 2.5 (Second-order \((\varepsilon, \delta, R)\)-optimum threshold):**

\[
L(\varepsilon, \delta, R | X) := \inf \{ L | L \text{ is } (\varepsilon, \delta, R)\text{-achievable} \}.
\]

**Remark 2.1:** It is not difficult to check that the condition \( \varepsilon + \delta \geq 1 \) yields the trivial result such as \( R(\varepsilon, \delta | X) = 0 \) or \( L(\varepsilon, \delta, R | X) = -\infty \). 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 \((\rho_{n}, \nu_{n})\), let \( D_n(\rho_{n}, \nu_{n}) \subset \mathcal{X}^n \) and \( T_n(\rho_{n}, \eta_{n}) \subset \mathcal{X}^n \) be defined as follows:

\[
D_n(\rho_{n}, \nu_{n}) := \{ x \in \mathcal{X}^n | x = \nu_{n}(\rho_{n}(x)) \},
\]
\[
T_n(\rho_{n}, \eta_{n}) := \{ x \in \mathcal{X}^n | l(\rho_{n}(x)) \leq \eta_{n} \}.
\]

Then, since any sequence \( x \in D_n(\rho_{n}, \nu_{n}) \) is correctly decodable, it holds that

\[
|D_n(\rho_{n}, \nu_{n}) \cap T_n(\rho_{n}, \eta_{n})| \leq \sum_{i=1}^{\eta_{n}} K^i < K^{\eta_{n}+1}.
\]

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 \((\varepsilon, \delta)\)-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

\[
H^+(X) := \min_{A \in \mathcal{X}^n} \{ \Pr[X \in A] \}^{1-\gamma} \log |A|.
\]

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

\[
R(\varepsilon, \delta | X) = \limsup_{n \to \infty} \frac{1}{n} H^{+\varepsilon+\delta+\nu}(X^n).
\]

**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 \( \varepsilon \) is called the \( \varepsilon \)-optimum coding rate. Uyematsu [12] has provided the general formula of the \( \varepsilon \)-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 ($\epsilon, \delta$)-optimum threshold (as well as the second-order ($\epsilon, \delta, R$)-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 thresholds 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 $a_n \geq 0$, $\eta_n \geq 1$ be arbitrary positive numbers. Then, for any $A_n \subset X^n$ satisfying $\Pr\{X^n \in A_n\} \geq 1 - \epsilon$, there exists a variable-length code $(\phi_n, \psi_n)$ such that

$$\epsilon_n \leq \Pr\{\phi_n(x) > \eta_n\}.$$  \hspace{1cm} (3.1)

**Proof:** We first construct the encoder and the decoder. [Encoder $\phi_n$]: For any fixed $A_n \subset X^n$ satisfying $\Pr\{X^n \in A_n\} \geq 1 - \epsilon$, we define the encoder as

$$\phi_n(x) = \begin{cases} f_n(x) \quad &x \in A_n, \\ 1 &\text{otherwise,} \end{cases}$$  \hspace{1cm} (3.2)

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

$$l(\phi_n(x)) = \begin{cases} \lceil -\log \frac{P_{X^n}(x)}{\Pr\{X^n \in A_n\}} \rceil \\ \text{otherwise,} \end{cases}$$  \hspace{1cm} (3.3)

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

$$\epsilon_n \leq \Pr\{X^n \notin A_n\} \leq \epsilon.$$  \hspace{1cm} (3.4)

To do so, let us define $S_n$ and $B_n$ as follows:

$$S_n := \{ x \in X^n \mid |(\phi_n(x)) > \eta_n \},$$  \hspace{1cm} (3.5)

$$B_n := \{ x \in X^n \mid a_n \frac{P_{X^n}(x)}{\Pr\{X^n \in A_n\}} \geq K^{-\eta_n} \}.$$  \hspace{1cm} (3.6)

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

$$\Pr\{\phi_n(x) > \eta_n\} = \Pr\{x \in A_n \mid x \notin S_n \}.$$  \hspace{1cm} (3.7)

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

$$\delta_n(\eta_n) = \Pr\{X^n \in S_n \cap B_n\} + \Pr\{X^n \in S_n \cap (B_n)^c\} \leq \Pr\{X^n \in A_n \cap B_n\} + \sum_{x \in (S_n \cap B_n)^c} \Pr\{X^n \in A_n\}$$

$$\leq \Pr\{X^n \in A_n \cap B_n\} + \sum_{x \in (B_n)^c} K^{-(\eta_n-1)} \Pr\{X^n \in A_n\}$$

$$\leq \Pr\{X^n \in A_n \cap B_n\} + |(B_n)^c| K^{-(\eta_n-1)} \Pr\{X^n \in A_n\}.$$  \hspace{1cm} (3.8)

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

$$1 \geq \sum_{x \notin B_n} P_{X^n}(x)$$

$$\geq \sum_{x \notin B_n} K^{-\eta_n} \frac{\Pr\{X^n \in A_n\}}{a_n}$$

$$= |(B_n)^c| K^{-\eta_n} \frac{1}{a_n} \Pr\{X^n \in A_n\}.$$  \hspace{1cm} (3.9)

from which it follows that

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

Plugging (3.10) into (3.8) yields (3.1). \hfill \square

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

$$\delta_n(\eta_n) \geq \Pr\{X^n \in D_n \mid X^n \in A_n \cap B_n\} - a_n K \Pr\{X^n \in D_n\}.$$  \hspace{1cm} (3.11)

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

$$\tilde{B}_n := \{ x \in X^n \mid a_n \frac{P_{X^n}(x)}{\Pr\{X^n \in A_n\}} \leq K^{-\eta_n} \}.$$  \hspace{1cm} (3.12)
and $S_n$ as in (3.4). Then, we have

$$\Pr \left[ \frac{P_{X^n}(X^n)}{\Pr \{ X^n \in D_n \}} \leq a_nK^{-\eta_n}, \; X^n \in D_n \right]$$

$$= \sum_{x \in D_n \cap \mathcal{B}_{\delta_n} \cap S_n} P_{X^n}(x) + \sum_{x \in D_n \cap \mathcal{B}_{\delta_n} \cap S_n^c} P_{X^n}(x) \leq \delta_n(\eta_n) + \sum_{x \in D_n \cap \mathcal{B}_{\delta_n} \cap S_n^c} a_nK^{-\eta_n} \Pr \{ X^n \in D_n \}$$

$$= \delta_n(\eta_n) + |D_n \cap (S_n^c)| a_nK^{-\eta_n} \Pr \{ X^n \in D_n \} \leq \delta_n(\eta_n) + a_nK \Pr \{ X^n \in D_n \},$$

(3.13)

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

\[ \square \]

IV. FIRST-ORDER ($\varepsilon, \delta$)-OPTIMUM THRESHOLD

A. General Formula

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

$$G_{\varepsilon, \delta}(X) := \inf \left\{ R \left| \lim \sup_{n \to \infty} \inf_{\mathcal{A}_n \subseteq \mathcal{X}^n: \Pr \{ X^n \in \mathcal{A}_n \} \geq 1 - \varepsilon - \nu} \Pr \left\{ \frac{1}{n} \log \frac{P_{X^n}(X^n)}{\Pr \{ X^n \in \mathcal{A}_n \}} \geq R, \; X^n \in \mathcal{A}_n \right\} \leq \delta \right. \right\}. \quad (4.1)$$

Then, we have the following theorem:

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

$$R(\varepsilon, \delta) = G_{\varepsilon, \delta}(X). \quad (4.2)$$

**Proof:** The proof consists of two parts.

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

$$\Pr \left\{ \frac{1}{n} \log \frac{P_{X^n}(X^n)}{\Pr \{ X^n \in \mathcal{A}_n \}} \geq R_0 + \gamma, \; X^n \in \mathcal{A}_n \right\} \leq \inf_{\mathcal{A}_n \subseteq \mathcal{X}^n: \Pr \{ X^n \in \mathcal{A}_n \} \geq 1 - \varepsilon - \nu} \Pr \left\{ \frac{1}{n} \log \frac{P_{X^n}(X^n)}{\Pr \{ X^n \in \mathcal{A}_n \}} \geq R_0 + \gamma, \; X^n \in \mathcal{A}_n \right\} + \lambda_n. \quad (4.3)$$

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

$$\varepsilon_n \leq \varepsilon + \nu, \quad (4.4)$$

and

$$\delta_n(nR) \geq \Pr \left\{ \frac{1}{n} \log \frac{P_{X^n}(X^n)}{\Pr \{ X^n \in \mathcal{D}_n \}} \leq R_0 + \gamma, \; X^n \in \mathcal{D}_n \right\} - K^{-\eta_n + 1}. \quad (4.41)$$

It follows from (4.3) that

$$\limsup_{n \to \infty} \Pr \left\{ \frac{1}{n} \log \frac{P_{X^n}(X^n)}{\Pr \{ X^n \in \mathcal{A}_n \}} \geq R_0 + 2\gamma \right\} \leq \limsup_{n \to \infty} \inf_{\mathcal{A}_n \subseteq \mathcal{X}^n: \Pr \{ X^n \in \mathcal{A}_n \} \geq 1 - \varepsilon - \nu} \Pr \left\{ \frac{1}{n} \log \frac{P_{X^n}(X^n)}{\Pr \{ X^n \in \mathcal{A}_n \}} \geq R_0 + \gamma, \; X^n \in \mathcal{A}_n \right\} \leq \delta, \quad (4.6)$$

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

$$\inf_{\mathcal{A}_n \subseteq \mathcal{X}^n: \Pr \{ X^n \in \mathcal{A}_n \} \geq 1 - \varepsilon - \nu} \Pr \left\{ \frac{1}{n} \log \frac{P_{X^n}(X^n)}{\Pr \{ X^n \in \mathcal{A}_n \}} \geq R_0 + \gamma, \; X^n \in \mathcal{A}_n \right\} \leq \delta. \quad (4.7)$$

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

(Converse Part:) We assume that $R$ is ($\varepsilon, \delta$)-achievable. Then, it holds that

$$\limsup_{n \to \infty} \varepsilon_n \leq \varepsilon, \quad (4.8)$$

$$\limsup_{n \to \infty} \delta_n(nR) \leq \delta. \quad (4.9)$$

By using Theorem 3.2 with $a_n = K^{-\eta_n}$ ($\forall \gamma > 0$) and $\eta_n = nR = n(R_0 + 2\gamma)$, we have

$$\delta_n(nR) \geq \Pr \left\{ \frac{1}{n} \log \frac{P_{X^n}(X^n)}{\Pr \{ X^n \in \mathcal{D}_n \}} \leq K^{-n(\gamma + R_0 + 2\gamma)}, \; X^n \in \mathcal{D}_n \right\} - K^{-n(\gamma + R_0 + 2\gamma)}. \quad (4.10)$$

where $\mathcal{D}_n = \{ x \in \mathcal{X}^n \mid x = \psi_n(\phi_n(x)) \}$. Here, (4.8) means that for any $\nu \in (0, 1 - \varepsilon)$ there exists $n_0$ such that $\Pr \{ X^n \in \mathcal{D}_n \} \geq 1 - \varepsilon - \nu$ ($\forall n > n_0$) holds. Thus, for any $\nu \in (0, 1 - \varepsilon)$ we have

$$\limsup_{n \to \infty} \delta_n(nR) \geq \limsup_{n \to \infty} \Pr \left\{ \frac{1}{n} \log \frac{P_{X^n}(X^n)}{\Pr \{ X^n \in \mathcal{D}_n \}} \geq R + \gamma, \; X^n \in \mathcal{D}_n \right\} \geq \limsup_{n \to \infty} \inf_{\mathcal{A}_n \subseteq \mathcal{X}^n: \Pr \{ X^n \in \mathcal{A}_n \} \geq 1 - \varepsilon - \nu} \Pr \left\{ \frac{1}{n} \log \frac{P_{X^n}(X^n)}{\Pr \{ X^n \in \mathcal{A}_n \}} \geq R + \gamma, \; X^n \in \mathcal{A}_n \right\}. \quad (4.11)$$
Substituting this inequality into (4.9), we obtain
\[
\delta \geq \limsup_{n \to \infty} \inf_{A_n \subseteq \mathcal{X}^n; \Pr[X^n \in A_n] \geq 1 - \epsilon - \nu} \Pr \left\{ \frac{1}{n} \log \frac{P_X(X^n)}{\Pr[X^n \in A_n]} \geq R + \gamma, \; X^n \in A_n \right\}. \tag{4.12}
\]
This means that
\[
R + \gamma \geq \inf_{\nu \geq 0} \limsup_{n \to \infty} \inf_{A_n \subseteq \mathcal{X}^n; \Pr[X^n \in A_n] \geq 1 - \epsilon - \nu} \Pr \left\{ \frac{1}{n} \log \frac{P_X(X^n)}{\Pr[X^n \in A_n]} \geq R, \; X^n \in A_n \right\} \geq \delta, \tag{4.13}
\]
which implies that the converse part holds.

**B. Another Expression of the General Formula**

The general formula of the \((\epsilon, \delta)\)-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 \(G_{\epsilon, \delta}(X)\) which enables us not only to compute the \((\epsilon, \delta)\)-optimum thresholds for tractable sources but also to understand the structure of the optimum overflow thresholds in the \(\epsilon\)-variable-length coding. As a result, the relationship with the \(\gamma\)-optimum coding in the fixed-length coding is revealed.

Set
\[
F(\epsilon, R) := \limsup_{n \to \infty} \inf_{A_n \subseteq \mathcal{X}^n; \Pr[X^n \in A_n] \geq 1 - \epsilon} \Pr \left\{ \frac{1}{n} \log \frac{P_X(X^n)}{\Pr[X^n \in A_n]} \geq R, \; X^n \in A_n \right\}. \tag{4.14}
\]
Then, the function \(F(\epsilon, R)\) is a monotonically nonincreasing function of \(\epsilon\) and \(R\). Then, we have the following lemma:

**Lemma 4.1:**
\[
G_{\epsilon, \delta}(X) = \inf_{\nu \geq 0} \left\{ R \mid \limsup_{n \to \infty} \Pr \left\{ \frac{1}{n} \log \frac{P_X(X^n)}{\Pr[X^n \in A_n]} \geq R + \nu, \; X^n \in A_n \right\} \leq \delta \right\}. \tag{4.15}
\]

**Proof:** We first fix \(v\) as \(v \in (0, 1 - \epsilon)\) and \(A_n \subseteq \mathcal{X}^n\) satisfying \(\Pr\{X^n \in A_n\} \geq 1 - \epsilon - \nu\). Then, for any \(x \in \mathcal{X}^n\) it holds that
\[
\left| \frac{1}{n} \log P_X(x) + \frac{1}{n} \log \frac{P_X(x)}{\Pr[X^n \in A_n]} \right| \leq \left| \frac{1}{n} \log P_X(x) + \frac{1}{n} \log \frac{1}{1 - \epsilon - \nu} \right| = \frac{1}{n} \log \frac{1}{1 - \epsilon - \nu}. \tag{4.16}
\]
Thus, noting that for any \(\gamma > 0\) it holds that \(\frac{1}{n} \log \frac{1}{1 - \epsilon - \nu} < \gamma (\forall n > n_0)\), which implies that the difference between \(-\frac{1}{n} \log P_X(x)\) and \(-\frac{1}{n} \log \frac{P_X(x)}{\Pr[X^n \in A_n]}\) becomes arbitrarily small, we obtain the lemma.

Here, we define two quantities
\[
\tilde{H}_{\epsilon, \delta}(X) := \inf \{ R \mid F(\epsilon, R) \leq \delta \}, \tag{4.17}
\]
and
\[
\overline{H}_{\gamma}(X) := \inf \left\{ R \mid \limsup_{n \to \infty} \Pr \left\{ \frac{1}{n} \log \frac{1}{P_X(X^n)} > R \right\} \leq \gamma \right\}. \tag{4.18}
\]
It should be noted that the \(\gamma\)-optimum coding rate in the fixed length coding is characterized by \(\overline{H}_{\gamma}(X)\) [3] (cf. Remark 4.1 below).

Then, the following theorem holds:

**Theorem 4.2:** For any \(\epsilon, \delta \in (0, 1)\) satisfying \(\epsilon + \delta < 1\), it holds that
\[
G_{\epsilon, \delta}(X) = \tilde{H}_{\epsilon, \delta}(X) = \overline{H}_{\epsilon + \delta}(X). \tag{4.19}
\]

□

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

From the theorem we immediately have:

**Corollary 4.1:** Fix \(\epsilon, \delta \in [0, 1)\) arbitrarily. Then, for any \(\epsilon_1, \epsilon_2 \geq 0\) satisfying \(\epsilon_1 + \epsilon_2 = \epsilon\) it holds that
\[
\tilde{H}_{\epsilon_1, \epsilon_2}(X) = \overline{H}_{\epsilon}(X). \tag{4.20}
\]

This equality reveals a deep relationship between the fixed-length coding and the (first-order) \(\epsilon\)-optimum threshold in the variable-length coding without error. From the above equality and Corollary 4.1, for any \(\epsilon_1, \epsilon_2 \geq 0\) satisfying \(\epsilon_1 + \epsilon_2 = \epsilon\), we obtain the following relation
\[
R(\epsilon, 0|X) = R(0, \epsilon|X) = \overline{H}_{\epsilon}(X). \tag{4.21}
\]
It should be emphasized that the relation (4.21) subsumes (4.20), because \(\epsilon_1\) and \(\epsilon_2\) in (4.21) may be arbitrary nonnegative numbers satisfying \(\epsilon_1 + \epsilon_2 = \epsilon\). This relation can also be obtained from the fact that the r.h.s. of (2.11) coincides with \(\overline{H}_{\epsilon + \delta}(X)\) [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
\[
\tilde{H}_{\epsilon + \nu, \delta}(X) \leq G_{\epsilon, \delta}(X) = \lim_{\nu \downarrow 0} \tilde{H}_{\epsilon + \nu, \delta}(X) \leq \tilde{H}_{\epsilon, \delta}(X). \tag{4.22}
\]

**Proof of Lemma 4.2:** Since \(\tilde{H}_{\epsilon, \delta}(X)\) is a monotonically nonincreasing function of \(\epsilon\), 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 $\varepsilon$ and $R$, $G_{\varepsilon, \delta}(X)$ can be expressed as
\begin{equation}
G_{\varepsilon, \delta}(X) = \inf_{v > 0} \left\{ R \mid F(\varepsilon + v, R) \leq \delta \right\}. \tag{4.23}
\end{equation}
For any $v > 0$, we have
\begin{equation}
G_{\varepsilon, \delta}(X) \geq \inf_{v > 0} \left\{ R \mid F(\varepsilon + v, R) \leq \delta \right\} = \bar{H}_{\varepsilon + v, \delta}(X), \tag{4.24}
\end{equation}
which implies
\begin{equation}
G_{\varepsilon, \delta}(X) \geq \sup_{v > 0} \bar{H}_{\varepsilon + v, \delta}(X) = \lim_{v \downarrow 0} \bar{H}_{\varepsilon + v, \delta}(X). \tag{4.25}
\end{equation}
On the other hand, for any $v > 0$, we have
\begin{equation}
\lim_{v \downarrow 0} \bar{H}_{\varepsilon + v, \delta}(X) \geq \inf_{v > 0} \left\{ R \mid F(\varepsilon + v, R) \leq \delta \right\} = G_{\varepsilon, \delta}(X), \tag{4.26}
\end{equation}
This inequality for an arbitrary $v > 0$ implies that
\begin{equation}
\lim_{v \downarrow 0} \bar{H}_{\varepsilon + v, \delta}(X) \geq G_{\varepsilon, \delta}(X), \tag{4.27}
\end{equation}
where the equality is due to (4.23). Combining (4.25) and (4.27) yields
\begin{equation}
G_{\varepsilon, \delta}(X) = \lim_{v \downarrow 0} \bar{H}_{\varepsilon + v, \delta}(X). \tag{4.28}
\end{equation}
This completes the proof of the lemma.

Proof of Theorem 4.2: From Lemma 4.2, it suffices to prove two inequalities:
\begin{equation}
G_{\varepsilon, \delta}(X) \geq \bar{H}_{\varepsilon + \delta}(X), \tag{4.29}
\end{equation}
\begin{equation}
H_{\varepsilon, \delta}(X) \leq \bar{H}_{\varepsilon + \delta}(X). \tag{4.30}
\end{equation}
(Proof of (4.29):) For any fixed $R > G_{\varepsilon, \delta}(X)$, we show that $R \geq \bar{H}_{\varepsilon + \delta}(X)$ holds.
Set $J_n \subseteq \chi^n$ as
\begin{equation}
J_n := \left\{ x \in \chi^n \mid \frac{1}{n} \log P_{X^n}(x) \geq R \right\}. \tag{4.31}
\end{equation}
Then, by the assumption $R > G_{\varepsilon, \delta}(X)$ and (4.15), it holds that
\begin{equation}
\delta \geq \limsup_{v \downarrow 0} \inf_{n \to \infty} \frac{1}{A_{n} \subseteq \chi^n} \Pr \left\{ X^n \in A_n \cap J_n \mid \Pr(X^n \in A_n) \geq 1 - \varepsilon - v \right\}, \tag{4.32}
\end{equation}
for any $v \in (0, 1 - \varepsilon)$. Moreover, we define a subset $K_n \subseteq \chi^n$ such that
\begin{equation}
\Pr \left\{ X^n \in K_n \right\} \geq 1 - \varepsilon - v \tag{4.33}
\end{equation}
and
\begin{equation}
\Pr \left\{ X^n \in J_n \cap K_n \right\} \leq \inf_{A_{n} \subseteq \chi^n} \Pr \left\{ X^n \in A_n \cap J_n \right\} + \varepsilon \tag{4.34}
\end{equation}
hold. Then, from (4.32) it holds that
\begin{equation}
\limsup_{n \to \infty} \Pr \left\{ X^n \in J_n \cap K_n \right\} \leq \delta + \varepsilon. \tag{4.35}
\end{equation}
On the other hand, from (4.33), we have
\begin{equation}
\Pr \left\{ X^n \in J_n \right\} \leq \Pr \left\{ X^n \in J_n \cap K_n \right\} + \Pr \left\{ X^n \in (K_n)^c \right\} \leq \Pr \left\{ X^n \in J_n \cap K_n \right\} + \varepsilon + v. \tag{4.36}
\end{equation}
This means that
\begin{equation}
\limsup_{n \to \infty} \Pr \left\{ X^n \in J_n \right\} \leq \limsup_{n \to \infty} \Pr \left\{ X^n \in J_n \cap K_n \right\} + \varepsilon + v \leq \delta + \varepsilon + 2v. \tag{4.37}
\end{equation}
Since $v \in (0, 1 - \varepsilon)$ is arbitrarily, we have
\begin{equation}
\limsup_{n \to \infty} \Pr \left\{ X^n \in J_n \right\} \leq \delta + \varepsilon, \tag{4.38}
\end{equation}
which implies $R \geq \bar{H}_{\varepsilon + \delta}(X)$.
(Proof of (4.30):) For any fixed $R > \bar{H}_{\varepsilon + \delta}(X)$, we show that $R > \bar{H}_{\varepsilon, \delta}(X)$ holds. We also use the set $J_n$ defined in (4.31). Since $R > \bar{H}_{\varepsilon + \delta}(X)$ holds, we have
\begin{equation}
\limsup_{n \to \infty} \Pr \left\{ X^n \in J_n \right\} \leq \varepsilon + \delta. \tag{4.39}
\end{equation}
Here, without loss of generality we assume that the elements of $X^n$ be ordered $[x_1, x_2, x_3, \cdots]$ with decreasing probabilities, that is, for any $i < j$, $P_{X^n}(x_i) \geq P_{X^n}(x_j)$ holds. We define a positive integer $i^*$ and $L_n = \left\{ x_1, x_2, \cdots, x_{i^*} \right\}$ such that
\begin{equation}
\Pr \left\{ X^n \in L_n \setminus \left\{ x_{i^*} \right\} \right\} = \sum_{i=1}^{i^*-1} P_{X^n}(x_i) < 1 - \varepsilon, \tag{4.40}
\end{equation}
\begin{equation}
\Pr \left\{ X^n \in L_n \right\} = \sum_{i=1}^{i^*} P_{X^n}(x_i) \geq 1 - \varepsilon. \tag{4.41}
\end{equation}
We then evaluate the probability
\begin{equation}
\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\}. \tag{4.42}
\end{equation}
By the definition of $J_n$, for all $n$ satisfying $\frac{1}{n} \log \frac{1}{P_{X^n}(x_{i^*})} < R$, it holds that
\begin{equation}
\Pr \left\{ X^n \in J_n \cap L_n \right\} = 0. \tag{4.43}
\end{equation}
On the other hand, for all $n$ satisfying $\frac{1}{n} \log \frac{1}{P_{X^n}(x_{i^*})} \geq R$,
\begin{equation}
P_{X^n}(x_{i^*}) \leq e^{-nR} \tag{4.44}
\end{equation}
and $(L_n)^c \subseteq J_n$ hold. Hence, we have
\begin{equation}
\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 \left\{ x_{i^*} \right\})^c \right\} - P_{X^n}(x_{i^*}) \geq e^{-nR}, \tag{4.45}
\end{equation}
where the last inequality is due to (4.40) and (4.44). Thus, from (4.42), (4.43), and (4.45) we have
\begin{equation}
\Pr \left\{ X^n \in J_n \cap L_n \right\} \leq \max \left\{ 0, \Pr \left\{ X^n \in J_n \right\} - \left( e^{-nR} \right) \right\}. \tag{4.46}
\end{equation}
for all n. Taking lim sup\(_{n \to \infty}\) on both sides yields
\[
\limsup_{n \to \infty} \Pr \left\{ X^n \in J_n \cap L_n \right\} \leq \max \left\{ 0, \limsup_{n \to \infty} \Pr \left\{ X^n \in J_n \right\} - \epsilon \right\} \leq \delta,
\]
where the last inequality is due to (4.39).

Since \( L_n \) satisfies (4.41) we obtain
\[
\limsup_{n \to \infty} \inf_{A_0 \subset X^n} \Pr \left\{ X^n \in A_0 \cap J_n \right\} \leq \delta. \tag{4.48}
\]
This implies \( R \geq H_{\epsilon, \delta}(X) \). \( \square \)

V. SECOND-ORDER \((\epsilon, \delta, R)\)-OPTIMUM THRESHOLD

In this section, we establish the general formula of the second-order \((\epsilon, \delta, R)\)-optimum threshold. We define the quantity \( G_{\epsilon, \delta}(R|X) \) as
\[
G_{\epsilon, \delta}(R|X) := \inf \left\{ L \left| \limsup_{n \to \infty} \inf_{\epsilon > 0} \Pr \left\{ \frac{1}{n} \log \frac{P_{X^n}(X^n)}{\Pr[X^n \in A_0]} \geq R + \frac{L}{\sqrt{n}} \right\} \leq \delta \right\}. \tag{5.1}
\]

Theorem 5.1 (Second-order \((\epsilon, \delta, R)\)-optimum threshold): For any \( \epsilon, \delta \in [0, 1) \) satisfying \( \epsilon + \delta < 1 \), it holds that
\[ L(\epsilon, \delta, R) = G_{\epsilon, \delta}(R|X). \]

Proof: The proof of the theorem proceeds in parallel with that of Theorem 4.1. \( \square \)

As in the first-order case (cf. Remark 4.1), \( G_{\epsilon, \delta}(R|X) \) can also be expressed as in another way. Let us define an information-spectrum quantity:
\[
\overline{H}_{\gamma}(R|X) := \inf \left\{ L \left| \limsup_{n \to \infty} \Pr \left\{ \frac{1}{n} \log \frac{1}{P_{X^n}(X^n)} \geq R + \frac{L}{\sqrt{n}} \right\} \leq \gamma \right\}. \tag{6.5}
\]

Theorem 5.2: For any \( \epsilon, \delta \in [0, 1) \) satisfying \( \epsilon + \delta < 1 \), it holds that
\[ G_{\epsilon, \delta}(R|X) = \overline{H}_{\epsilon + \delta}(R|X). \]

Proof: The proof is similar to that of Theorem 4.2. \( \square \)

The second-order \((\epsilon, R)\)-optimum coding rate in the fixed-length coding is characterized by \( \overline{H}_{\epsilon}(R|X) \) [13]. This means that in order to compute \( G_{\epsilon, \delta}(R|X) \) 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 \((\epsilon, \delta)\)-Optimum Thresholds

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

Theorem 6.1 (Steinberg and Verdù [15]): For any \( \gamma \in [0, 1) \), it holds that
\[
\overline{H}_{\gamma}(X) = H(X), \tag{6.1}
\]
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 \( \epsilon, \delta \in [0, 1) \) satisfying \( \epsilon + \delta < 1 \) it holds that
\[
R(\epsilon, \delta|X) = H(X) \quad (0 \leq \epsilon, \delta < 1). \tag{6.2}
\]

Thus, the \((\epsilon, \delta)\)-optimum overflow thresholds equals to the entropy of the source irrespective of \( \epsilon \) and \( \delta \).

B. Second-Order \((\epsilon, \delta, R)\)-Optimum Overflow Thresholds

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

Theorem 6.3 (Hayashi [13]):
\[
\overline{H}_{\gamma}(R|X) = \begin{cases} \gamma & \text{if } R < H(X) \leq \gamma \sqrt{V_X} \Phi^{-1}(\gamma) \\ \infty & \text{if } R > H(X) \end{cases} \tag{6.3}
\]
where
\[
\Phi(x) := \frac{1}{\sqrt{2\pi}} \int_{x}^{\infty} e^{-\frac{t^2}{2}} dt \tag{6.4}
\]
is the complementary standard Gaussian distribution function and
\[
V_X := \sum_{x \in A_x} P_X(x) (-\log P_X(x) - H(X))^2. \tag{6.5}
\]
denotes the variance of the self-information called \textit{varentropy} of the source [6]. \( \square \)

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

Case 1) Setting \( R \) as the first-order \((\epsilon, \delta)\)-optimum thresholds \( R_1 \):

In this case, from Theorem 4.1 \( R_1 = \overline{H}_{\epsilon + \delta}(X) \) holds. Thus, from Theorem 6.1 we set
\[
R_1 = H(X) \tag{6.6}
\]
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 \( \epsilon \)-variable-length codes has been first determined by Koga and Yamamoto [7] in the case that \( \epsilon \in [0, 1) \), while Han [3], [16] has derived it in the case of \( \epsilon = 0 \).
From the result in [7] we shall set

\[
R_2 = H(\varepsilon|X)
\]

\[
= \lim_{n \to \infty} \inf_{\nu \in \mathcal{A}^n} \frac{1}{n} \sum_{x \in \mathcal{X}^n} P_{X^n}(x) \log \frac{1}{P_X(x)}.
\]

(6.7)

In particular,

\[
R_2 = (1 - \varepsilon) H(X)
\]

(6.8)

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

Remark 6.1: The optimum second-order mean codeword length \( L_{[\varepsilon]}(H(\varepsilon|X)|X) \) 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

\[
L_{[\varepsilon]}(H(\varepsilon|X)|X) = \sqrt{\frac{V(X)}{2\pi}} e^{-\frac{(\varepsilon-1)(\varepsilon+1)}{2}}.
\]

(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 \( R_1 = H(X) \) 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

\[
\begin{align*}
L(\varepsilon, \delta, R_1|X) &= \sqrt{V_X} \Phi^{-1}(\varepsilon + \delta), \\
L(\varepsilon, \delta, R_2|X) &= \sqrt{V_X} \Phi^{-1}(\delta),
\end{align*}
\]

(6.10)

(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 \( L(\varepsilon, \delta, R_2|X) \) 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 \( \overline{R}_{\varepsilon+\delta}(R_2|X) \) it holds that

\[
L(\varepsilon, \delta, R_2|X) = \overline{R}_{\varepsilon+\delta}(R_2|X)
\]

\[
= \inf \left\{ \frac{1}{n} \log \frac{1}{P_X^n(x)} > R_2 + \frac{L}{\sqrt{n}} \right\} \leq \varepsilon + \delta
\]

\[
= \inf \left\{ \frac{1}{n} \log \frac{1}{P_X^n(x)} > (1 - \varepsilon) H(X) + \frac{L}{\sqrt{n}} \right\} \leq \varepsilon + \delta.
\]

(6.12)

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

\[
\lim_{n \to \infty} \frac{1}{n} \log \frac{1}{P_X^n(x)} > (1 - \varepsilon) H(X) + \frac{L}{\sqrt{n}} = 1,
\]

(6.13)

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

The above theorem shows that if we set \( R = R_2 \) (the optimum mean codeword length of \( \varepsilon \)-variable-length codes), the second-order \( (\varepsilon, \delta, R) \)-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 \( L(\varepsilon, \delta, R|X) = \infty \) for \( R < H(X) \), and \( L(\varepsilon, \delta, R|X) = -\infty \) for \( R > H(X) \). Hence, in order to analyze the second-order optimum threshold \( L \) in the \( \varepsilon \)-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 \( \varepsilon \)-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):

\[
\limsup_{n \to \infty} \epsilon_n \leq \epsilon, \quad \liminf_{n \to \infty} \delta_n(nR) \leq \delta.
\]

(7.1)

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

We have also clarified that the relationship between the \( \varepsilon \)-variable-length coding and the \( \gamma \)-fixed-length coding. In the \( \gamma \)-fixed-length coding problem, the \( \gamma \)-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 \( (\varepsilon, \delta) \)-optimum thresholds as well as the \( (\epsilon, \delta, R) \)-optimum thresholds in the \( \varepsilon \)-variable-length coding. Actually, in this paper we compute the optimum thresholds in the \( \varepsilon \)-variable-length coding for the stationary memoryless source by using the previous results for the \( \gamma \)-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. Verdú, “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.