Overflow Probability of Variable-Length Codes With Codeword Cost
Ryo Nomura, Member, IEEE

Abstract—Lossless variable-length source coding with codeword cost is considered for general sources. The problem setting, where we impose on unequal costs on code symbols, is called the variable-length coding with codeword cost. In this problem, the infimum of average codeword cost have already been determined for general sources. On the other hand, the overflow probability, which is defined as the probability of codeword cost being above a threshold, have not been considered yet. In this paper, we first determine the infimum of achievable threshold in the first-order sense and the second-order sense for general sources with additive memoryless codeword cost. Then, we compute it for some special sources such as i.i.d. sources and mixed sources. A generalization of the codeword cost is also discussed.

Index Terms—Codeword cost, general source, information-spectrum, overflow probability, variable-length coding.

I. INTRODUCTION

LOSSLESS variable-length coding problem is quite important not only from the theoretical viewpoint but also from the viewpoint of its practical applications. To evaluate the performance of variable-length codes, several criteria have been proposed. The most fundamental criterion is the average codeword length proposed by Shannon [2]. Many variable-length codes have been proposed and its performance has been evaluated by using the average codeword length [3]. The overflow probability of codeword length is one of other criteria, which denotes the probability of codeword length per symbol being above a threshold \( R > 0 \). Merhav [4] has first determined the optimal exponent of the overflow probability given \( R \) for unifilar sources. Uchida and Han [5] have shown the infimum of achievable threshold \( R \) under constraints that the overflow probability vanishes with given exponent \( r \). Their analyses are based on the information-spectrum methods and the results are valid for general sources. Nomura and Matsushima [6] have computed the infimum of \( \varepsilon \)-achievable overflow threshold for general sources. Here, \( \varepsilon \)-achievable overflow threshold means that there exists a variable-length code, whose overflow probability is smaller than or equal to \( \varepsilon \). All the results mentioned here are in the meaning of the codeword length.

On the other hand, as is well known, if we impose unequal costs on code symbols, it makes no sense to use the codeword length as a measure. In this setting, we have to consider the codeword cost, instead of the codeword length. The average codeword cost, which is a generalization of the average codeword length, have been first analyzed also by Shannon [2]. Moreover, several researchers have studied on the average codeword cost [7]–[9]. Karp [10] has given the variable-length code, which minimizes the average codeword cost. Golin and Rote [11], and Golin and Li [12] have proposed the efficient algorithm for constructing the optimal variable-length code for i.i.d. sources while Uyematsu et al. [13] have proposed the asymptotic optimal universal variable-length code for stationary ergodic sources, with respect to the average codeword cost. The infimum of the average codeword cost has been determined by Krause [9] for i.i.d. sources and extended to general sources by Han and Uchida [14]. Recently, Yagi and Nomura [15] have determined the general formula of the infimum of the average codeword cost in the variable-length code allowing non-vanishing error probability. Among others, Uchida and Han [5] have proposed the overflow probability of the codeword cost. They have considered the overflow probability as the probability of the codeword cost per symbol being above a threshold. Then, they have shown the infimum of achievable threshold, where the achievable threshold means that there exists a variable-length code whose overflow probability of the codeword cost decreases with given exponent \( r \).

In this paper, we also deal with the overflow probability of the codeword cost. In particular, we consider the \( \varepsilon \)-achievable threshold, which means that there exists a variable-length code, whose overflow probability is smaller than or equal to \( \varepsilon \). We first determine the infimum of first-order and second-order achievable thresholds for general sources. The finer evaluation of the achievable rate, called the second-order achievable rate, has been investigated in several contexts. In the variable-length source coding problem, Kontoyiannis [16] has established the second-order source coding theorem on the codeword length for i.i.d. sources and Markov sources. In the channel coding problem, Strassen [17] (see, Csiszár and Körner [8]), Hayashi [18], and Verdú et al. [19] have determined the second-order capacity rate. Hayashi [20] has also shown the second-order achievability theorems for the fixed-length source coding problem for general sources and compute the optimal second-order achievable rates for i.i.d. sources by using the asymptotic normality. Nomura and Han [21] have also computed the optimal second-order achievable rates in...
fixed-length source coding for mixed sources by using the two-peak asymptotic normality.

Analogously to these settings, we define the second-order achievable threshold on the overflow probability and derive the infimum of the second-order achievable thresholds. Notice here that Nomura and Matsushima [6] have already considered the first-order and the second-order achievements with respect to the overflow probability of the codeword length. One of the contributions of this paper is to extend the results in [6] to the case of the codeword cost. Our analysis is based on the information-spectrum methods and hence our results are valid for general sources. Furthermore, we apply our results to i.i.d. sources and mixed sources as special cases and compute the infimum of the first- and second-order achievable thresholds for these important sources.

A generalization of the codeword cost is also considered. The codeword cost considered in the above is the memoryless cost, which means that the component cost function is fixed and the cost capacity is uniquely determined. This property helps us to derive fundamental limits of the first- and second-order achievable thresholds explicitly. On the other hand, we also consider the case that the cost function may vary according to the previous sequence. This is a setting inheritance in the variable-length coding with cost, and hence it is meaningful to discuss. We try to extend our results to this general cost function and demonstrate the upper and lower bounds of the infimum of the first-order achievable thresholds.

Related works include works by Kontoyiannis and Verdú [22], and Kosut and Sankar [23], [24]. They have also considered the similar problem with the overflow probability. Kontoyiannis and Verdú [22] have derived the fundamental limit of this quantity without the prefix conditions. Kosut and Sankar [23], [24] have also derived the upper-bound of the overflow probability in universal setting. It should be emphasized that they have considered the overflow probability of codeword length for some special sources and derived bounds up to the third-order. On the other hand, in this paper we consider the overflow probability of codeword cost for general sources and address the fundamental limit of the achievable threshold up to the second-order.

This paper is organized as follows. In Section II, we state the problem settings and define two achievements treated in this paper. In Section III, we prove two lemmas which play the key role in the subsequent analysis. In Section IV, we determine the infimum of the first-order achievable thresholds. In Section V, we derive the infimum of the second-order achievable threshold and compute it for some special sources. In Section VI, we extend our results to the conditional cost. Finally, in Section VII, we conclude our results.

II. OVERFLOW PROBABILITY OF VARIABLE-LENGTH CODING WITH COST

A. Variable-Length Codes With Codeword Cost for General Source

The general source is defined as an infinite sequence

$\mathbf{X} = \left\{ X^n = (X_1^{(n)} X_2^{(n)} \ldots X_n^{(n)}) \right\}_{n=1}^\infty$ of $n$-dimensional random variables $X^n$, where each component random variable $X_i^{(n)}$ takes values in a countable set $\mathcal{X}$. It should be noted that each component of $X^n$ may change depending on block length $n$. This implies that even consistency condition, which means that for any integers $m$, $n$ ($m < n$), $X_i^{(m)} = X_i^{(n)}$ holds, may not hold.

Variable-length codes are characterized as follows. Let

$\varphi_n : \mathcal{X}^n \rightarrow \mathcal{U}^*$, $\psi_n : \{\varphi_n(x)\}_{x \in \mathcal{X}^n} \rightarrow \mathcal{X}^n$,

be a variable-length encoder and a decoder, respectively, where $\mathcal{U} = \{1, 2, \ldots, K\}$ is called the code alphabet and $\mathcal{U}^*$ is the set of all finite-length strings over $\mathcal{U}$ excluding the null string.

We consider the situation that there are unequal costs on code symbols. Let us define the cost function over $\mathcal{U}$ considered in this paper. Each code symbol $u \in \mathcal{U}$ is assigned the corresponding cost $c(u)$ such that $0 < c(u) < \infty$, and the additive cost $c(u)$ of $u = (u_1, u_2, \ldots, u_k)$ in $\mathcal{U}^k$ is defined by

$c(u) = \sum_{i=1}^k c(u_i)$.

In particular, we denote $c_{\text{max}} = \max_{u \in \mathcal{U}} c(u)$ for short. This cost function is called the memoryless cost function in this paper. A generalization of this cost function is discussed in Section VI.

We only consider variable-length codes satisfying prefix condition. It should be noted that every variable-length code with prefix condition over unequal costs, satisfies

$$\sum_{x \in \mathcal{X}^n} K^{-\alpha_c c(\varphi_n(x))} \leq 1, \quad (1)$$

where $\alpha_c$ is called cost capacity and defined as the positive unique root $\alpha$ of the equation [8]:

$$\sum_{u \in \mathcal{U}} K^{-\alpha c(u)} = 1. \quad (2)$$

Throughout this paper, the logarithm is taken to the base $K$.

B. Overflow Probability of Codeword Cost

The Overflow Probability of the Codeword Length is Defined as follows:

Definition 2.1: [4] Given a threshold $R$, the overflow probability of the variable-length encoder $\varphi_n$ is defined by

$$\varepsilon_n(\varphi_n, R) = \Pr \left\{ \frac{1}{n} l(\varphi(X^n)) > R \right\}, \quad (3)$$

where $l(\cdot)$ denotes the length function.

In this paper, we generalize the above overflow probability not only to the case for unequal costs on code symbols but also for the finer evaluation of the overflow probability. To this end, we consider the overflow probability of codeword cost as follows:

Definition 2.2 (Overflow Probability of Codeword Cost): Given some sequence $\{\eta_n\}_{n=1}^\infty$, where $0 < \eta_n < \infty$ for each $n = 1, 2, \ldots$, the overflow probability of variable-length encoder $\varphi_n$ is defined by

$$\varepsilon_n(\varphi_n, \eta_n) = \Pr \left\{ c(\varphi(X^n)) > \eta_n \right\}. \quad (4)$$
Remark 2.1: Nomura and Matsushima [6] have considered the overflow probability with respect to the codeword length, that is, \( \Pr \{|(\varphi(X^n)) > \eta_n\} \) and derived the achievability of the first-order and the second-order sense. Kosut and Sankar [23] have also defined the similar probability in the case of codeword length and derived the upper bound in universal setting.

Since \( \{\eta_n\}_{n=1}^{\infty} \) is an arbitrary sequence, the above definition is general. In particular, we shall consider the following two types of overflow probability in this paper:

1) \( \eta_n = nR \),
2) \( \eta_n = na + \sqrt{n}L \).

Remark 2.2: If we set \( \eta_n = nR \) for all \( n = 1, 2, \ldots \), that is, in the first case, the overflow probability can be written as

\[
epsilon_n(\varphi_n, nR) = \Pr \{ c(\varphi_n(X^n)) > nR \} = \Pr \left\{ \frac{1}{n} c(\varphi_n(X^n)) > R \right\}.
\]

Thus, in the case of \( \eta_n = nR \), the overflow probability defined by (4) means the probability that the codeword cost per symbol exceeds some constant \( R \). This is a natural extension of the overflow probability of the codeword length to the overflow probability of the codeword cost defined by (3).

On the other hand, in the analysis of the fixed-length coding problem, Hayashi [20] has shown the second-order asymptotics, which enables us a finer evaluation of the achievable coding rate. A coding theorem from the viewpoint of the second-order asymptotics have also been analyzed by Kontoyiannis [16]. Analogously to their results, we evaluate the overflow probability in the second-order sense. To do so, we consider the second case: \( \eta_n = na + \sqrt{n}L \) for all \( n = 1, 2, \ldots \). Hereafter, if we consider the overflow probability in the case of \( \eta_n = na + \sqrt{n}L \), we call it the second-order overflow probability given \( a \), while in the first case it is called the first-order overflow probability. The second-order overflow probability given \( a \) of the variable-length encoder \( \varphi_n \) with threshold \( L \) is written as

\[
epsilon_n(\varphi_n, na + L \sqrt{n}) = \Pr \{ c(\varphi_n(X^n)) > na + L \sqrt{n} \} = \Pr \left\{ \frac{c(\varphi_n(X^n))}{\sqrt{n}} > n a + L \right\}.
\]

It should be noted that since we assume that \( \eta_n \) satisfies \( 0 < \eta_n < \infty \), \( R \) is a positive number while \( L \) can be negative.

In the first-order case, we are interested in the infimum of the threshold \( R \) that we can achieve. This is formalized as follows.

**Definition 2.3:** Given \( 0 \leq \varepsilon < 1 \), \( R \) is called an \( \varepsilon \)-achievable overflow threshold for the source if there exists a sequence of variable-length code \( (\varphi_n, \psi_n) \) such that

\[
\limsup_{n \to \infty} \varepsilon_n(\varphi_n, nR) \leq \varepsilon.
\]

**Definition 2.4 (\( \varepsilon \)-achievable overflow threshold):**

\( R(\varepsilon | X) := \inf \{ R | R \text{ is an } \varepsilon \text{-achievable overflow threshold} \} \).

Also, in the analysis of the second-order overflow probability, we define the achievability:

**Definition 2.5:** Given \( 0 \leq \varepsilon < 1 \) and \( 0 < a < \infty \), \( L \) is called an \( (\varepsilon, a) \)-achievable overflow threshold for the source, if there exists a sequence of variable-length code \( (\varphi_n, \psi_n) \) such that

\[
\limsup_{n \to \infty} \varepsilon_n(\varphi_n, na + L \sqrt{n}) \leq \varepsilon.
\]

**Definition 2.6 ((\( \varepsilon, a \))-achievable overflow threshold):**

\[
L(\varepsilon, a | X) := \inf \{ L | L \text{ is an } (\varepsilon, a) \text{-achievable overflow threshold} \}.
\]

**III. Finite Blocklength Bounds**

Our objective in this paper is to demonstrate the infimum of the \( \varepsilon \)-achievable overflow thresholds as well as the infimum of the \((\varepsilon, a)\)-achievable overflow thresholds. To do so, in this section we derive two lemmas.

**Lemma 3.1:** For any general sources \( X \) and any sequence of positive number \( \{\eta_n\}_{n=1}^{\infty} \), there exists a variable-length encoder \( \varphi_n \) that satisfies

\[
\varepsilon_n(\varphi_n, \eta_n) < \Pr \{ z_n P_{X^n}(X^n) \leq K^{-a_c \eta_n} \} + z_n K^{a_c \eta_n + 1},
\]

for \( n = 1, 2, \ldots \), where \( \{z_n\}_{n=1}^{\infty} \) is a given sequence of an arbitrary number satisfying \( z_i > 0 \) for \( i = 1, 2, \ldots \) and \( a_c \) denotes the cost capacity defined in (2).

**Proof:** Here, we use the code proposed by Han and Uchida [14]. Notice here that from the property of the code, it holds that

\[
c(\varphi^*_n(x)) \leq -\frac{1}{a_c} \log P_{X^n}(x) + \frac{\log 2}{a_c} + c_{\max},
\]

for all \( n = 1, 2, \ldots \), where \( \varphi^*_n \) denotes the encoder of the code. Furthermore, we set the decoder as the inverse mapping of \( \varphi^*_n \) that is, \( \psi_n = \varphi^*_n \). Please note that the code is a uniquely decodable variable-length code for general sources with countably infinite source alphabet.

Next, we shall evaluate the overflow probability of this code. Set

\[
A_n = \{ x \in X^n | z_n P_{X^n}(x) \leq K^{-a_c \eta_n} \},
\]

\[
S_n = \{ x \in X^n | c(\varphi^*_n(x)) > \eta_n \}.
\]

The overflow probability is given by

\[
\varepsilon_n(\varphi_n, \eta_n) = \Pr \{ X^n \in S_n \} = \sum_{x \in S_n} P_{X^n}(x) = \sum_{x \in S_n \cap A_n} P_{X^n}(x) + \sum_{x \in S_n \cap \bar{A}_n} P_{X^n}(x) \leq \Pr \{ X^n \in A_n \} + \sum_{x \in S_n \cap \bar{A}_n} P_{X^n}(x),
\]

where \( \bar{A} \) denotes the complement set of the set \( A \).

Since (6) holds, for \( \forall x \in S_n \), we have

\[
-\frac{1}{a_c} \log P_{X^n}(x) + \frac{\log 2}{a_c} + c_{\max} > \eta_n.
\]

Thus, we have

\[
P_{X^n}(x) < K^{-a_c (\eta_n - c_{\max}) + \log 2},
\]

1 In Section VI, we have introduced the modified version of this code. For the derivation of this inequality, see also the proof of Lemma 6.2.
for $\forall x \in S_n$. Substituting the above inequality into (7), we have
\[
\varepsilon_n(\varphi_n, \eta_n) < \Pr \{ X^n \in A_n \} + \sum_{x \in S_n \cap A'_n} K^{-\alpha_n(\eta_n - c_{\text{max}})} + \log 2 \\
= \Pr \{ X^n \in A_n \} + |S_n \cap A'_n| K^{-\alpha_n(\eta_n - c_{\text{max}})} + \log 2.
\]

Here, from the definition of $A_n$, for $\forall x \in A'_n$, it holds that
\[
P_{X^n}(x) > \frac{K^{-\alpha_n \eta_n}}{z_n}.
\]
Thus, we have
\[
1 \geq \sum_{x \in X^n} P_{X^n}(x) \geq \sum_{x \in A'_n} P_{X^n}(x) > \sum_{x \in A'_n} \frac{K^{-\alpha_n \eta_n}}{z_n} = \frac{|A'_n|}{z_n}.
\]
This means that
\[
|S_n \cap A'_n| \leq |A'_n| < z_n K^{\alpha_n \eta_n}.
\]
Substituting (9) into (8), we obtain
\[
\varepsilon_n(\varphi_n, \eta_n) < \Pr \{ X^n \in A_n \} + z_n K^{\alpha_n \eta_n} K^{-\alpha_n(\eta_n - c_{\text{max}})} + \log 2 \\
\leq \Pr \{ X^n \in A_n \} + z_n K^{\alpha_n c_{\text{max}} + 1},
\]
because $\log 2 \leq 1$ holds. Therefore, we have proved the lemma.

**Lemma 3.2:** For any variable-length code and any sequence $\{\eta_n\}_{n=1}^{\infty}$, it holds that
\[
\varepsilon_n(\varphi_n, \eta_n) \geq \Pr \{ X^n \in A_n \} + z_n K^{\alpha_n \eta_n} - z_n
\]
for $n = 1, 2, \ldots$, where $\{z_n\}_{n=1}^{\infty}$ is a given sequence of an arbitrary number satisfying $z_i > 0$ for $i = 1, 2, \ldots$.

**Proof:** Let $\varphi_n$ be an encoder of the variable-length code. Set
\[
B_n = \{ x \in X^n \mid P_{X^n}(x) \leq z_n K^{\alpha_n \eta_n} \},
\]
\[
S_n = \{ x \in X^n \mid c(\varphi_n(x)) > \eta_n \}.
\]
Then, we have
\[
\Pr \{ P_{X^n}(X^n) \leq z_n K^{\alpha_n \eta_n} \}
= \sum_{x \in B_n} P_{X^n}(x)
= \sum_{x \in B_n \cap S_n} P_{X^n}(x) + \sum_{x \in B_n \cap S'_n} P_{X^n}(x)
\leq \sum_{x \in S_n} P_{X^n}(x) + \sum_{x \in B_n \cap S'_n} P_{X^n}(x)
\leq \varepsilon_n(\varphi_n, \eta_n) + \sum_{x \in B_n \cap S'_n} P_{X^n}(x).
\]
On the other hand, for $\forall x \in B_n$, it holds that
\[
P_{X^n}(x) \leq z_n K^{\alpha_n \eta_n}.
\]
Thus, we have
\[
\Pr \{ P_{X^n}(X^n) \leq z_n K^{\alpha_n \eta_n} \}
\leq \varepsilon_n(\varphi_n, \eta_n) + \sum_{x \in B_n \cap S'_n} P_{X^n}(x)
\leq \varepsilon_n(\varphi_n, \eta_n) + \sum_{x \in B_n \cap S'_n} z_n K^{\alpha_n \eta_n} + \log 2
\]
\[
= \varepsilon_n(\varphi_n, \eta_n) + |B_n \cap S'_n| K^{\alpha_n \eta_n}.
\]
This means that
\[
|B_n \cap S'_n| \leq |S'_n| \leq K^{\alpha_n \eta_n}.
\]
Hence, substituting (12) into (11), we have
\[
\Pr \{ P_{X^n}(X^n) \leq z_n K^{\alpha_n \eta_n} \}
\leq \varepsilon_n(\varphi_n, \eta_n) + |B_n \cap S'_n| K^{\alpha_n \eta_n}
\leq \varepsilon_n(\varphi_n, \eta_n) + K^{\alpha_n c_{\text{max}}} + z_n K^{\alpha_n \eta_n}
\]
\[
= \varepsilon_n(\varphi_n, \eta_n) + z_n.
\]
Therefore, we have proved the lemma.

**IV. INFIMUM OF $\varepsilon$-ACHIEVABLE OVERFLOW THRESHOLD**

**A. General Formula**

In this subsection, we determine $R(\varepsilon|X)$ for general sources. Before showing the theorem, we define the function $F(R)$ as follows:
\[
F(R) := \lim_{n \to \infty} \sup \frac{1}{n \varepsilon} \log \frac{1}{\Pr \{ X^n \leq z_n K^{\alpha_n \eta_n} \}}.
\]
The following theorem is one of our main results:

**Theorem 4.1:** For $0 \leq \varepsilon < 1$, it holds that
\[
R(\varepsilon|X) = \inf \{ R \mid F(R) \leq \varepsilon \}.
\]

**Proof:** The proof consists of two parts. (Direct Part) Let $R_0$ be as
\[
R_0 = \inf \{ R \mid F(R) \leq \varepsilon \},
\]
for short. Then, in this part we show that
\[
R(\varepsilon|X) \leq R_0 + \gamma,
\]
for any $\gamma > 0$ by showing that $R_0$ is an $\varepsilon$-achievable overflow threshold for the source. Let $\eta_n$ be as $\eta_n = n(R_0 + \gamma)$, then from Lemma 3.1 there exists a variable-length code $(\varphi_n, \psi_n)$ that satisfies
\[
\varepsilon_n(\varphi_n, n(R_0 + \gamma)) < \Pr \{ z_n P_{X^n}(X^n) \leq K^{-\alpha_n c_{\text{max}}(R_0 + \gamma)} \}
\]
\[
+ z_n K^{\alpha_n c_{\text{max}} + 1},
\]

for $n = 1, 2, \ldots$. Thus, we have

$$e_n(\psi_n, n(R_0 + \gamma))$$

$$< \Pr \left\{ \frac{1}{n} \log \frac{1}{P_X^n(x^n)} \geq R_0 + \frac{\gamma}{n\alpha_c} \right\} + z_n K - n \alpha_c c_{\max} + 1$$

$$= \Pr \left\{ \frac{1}{n} \log \frac{1}{P_X^n(x^n)} \geq R_0 + \gamma + \frac{1}{n\alpha_c} \log z_n \right\} + z_n K - n \alpha_c c_{\max} + 1,$$

for $n = 1, 2, \ldots$. Notice here that $z_n > 0$ is an arbitrary number. Setting $z_n = K^{-\gamma}$, we have

$$e_n(\psi_n, n(R_0 + \gamma))$$

$$< \Pr \left\{ \frac{1}{n\alpha_c} \log \frac{1}{P_X^n(x^n)} \geq R_0 + \gamma - \sqrt{n\alpha_c} \right\}$$

$$+ K^{-\gamma} + n \alpha_c c_{\max} + 1,$$

for sufficiently large $n$, because $\frac{\gamma}{n\alpha_c}$ as $n \to \infty$. Thus, since $\alpha_c$ and $c_{\max}$ are positive constants, by taking $\limsup_{n \to \infty}$, we have

$$\limsup_{n \to \infty} e_n(\psi_n, n(R_0 + \gamma))$$

$$\leq \limsup_{n \to \infty} \Pr \left\{ \frac{1}{n\alpha_c} \log \frac{1}{P_X^n(x^n)} \geq R_0 + \gamma - \sqrt{n\alpha_c} \right\}.$$

Since $\gamma > 0$ is an arbitrary constant, from the definition of $R_0$ we have

$$\limsup_{n \to \infty} e_n(\psi_n, n(R_0 + \gamma)) \leq \varepsilon.$$

Noting that $\gamma > 0$ is arbitrarily small, the direct part has been proved.

(Converse Part)

Assuming that $R_1$ satisfying

$$R_1 < \inf \{ R \mid F(R) \leq \varepsilon \}, \quad (14)$$

is an $\varepsilon$-achievable overflow threshold, then we shall show a contradiction.

Let $\eta_n$ be as $\eta_n = nR_1$. Then, from Lemma 3.2 for any sequence $\{z_n\}_{n=1}^{\infty}$ $(z_i > 0, i = 1, 2, \ldots)$ and any variable-length code it holds that

$$e_n(\psi_n, nR_1) > \Pr \left\{ P_X^n(x^n) \leq z_n K - n\eta_n R_1 \right\} - z_n,$$

for $n = 1, 2, \ldots$. Thus, for any variable-length code it holds that

$$e_n(\psi_n, nR_1)$$

$$> \Pr \left\{ P_X^n(x^n) \leq z_n K - n\eta_n R_1 \right\} - z_n$$

$$= \Pr \left\{ \frac{1}{n\alpha_c} \log \frac{1}{P_X^n(x^n)} \geq R_1 - \frac{1}{n\alpha_c} \log z_n \right\} - z_n.$$

Set $z_n = K^{-\gamma}$, where $\gamma > 0$ is a small constant that satisfies

$$R_1 + \frac{\gamma}{\alpha_c} < \inf \{ R \mid F(R) \leq \varepsilon \}. \quad (15)$$

Since we assume that (14) holds, it is obvious that there exists $\gamma > 0$ that satisfies the above inequality. Then, we have

$$e_n(\psi_n, nR_1) > \Pr \left\{ \frac{1}{n\alpha_c} \log \frac{1}{P_X^n(x^n)} \geq R_1 + \frac{\gamma}{\alpha_c} \right\}$$

$$- K^{-\gamma}.$$

Hence, we have

$$\limsup_{n \to \infty} e_n(\psi_n, nR_1)$$

$$\geq \limsup_{n \to \infty} \Pr \left\{ \frac{1}{n\alpha_c} \log \frac{1}{P_X^n(x^n)} \geq R_1 + \frac{\gamma}{\alpha_c} \right\} > \varepsilon,$$

where the last inequality is derived from (15) and the definition of $F(R)$.

On the other hand, since we assume that $R_1$ is an $\varepsilon$-achievable overflow threshold, it holds that

$$\limsup_{n \to \infty} e_n(\psi_n, nR_1) \leq \varepsilon.$$

This is a contradiction. Therefore, the proof of the converse part has been completed.

From the above theorem, we can show a corollary. Before describing the corollary, we define the spectral sup-entropy rate [25]:

$$\overline{H}(X) = \limsup_{n \to \infty} - \frac{1}{n} \log P_X^n(x^n).$$

Then, the following corollary holds.

**Corollary 4.1:**

$$R(0|X) = \frac{1}{\alpha_c} \overline{H}(X). \quad (16)$$

**B. Strong Converse Property**

Strong converse property is one of important properties for the source in the fixed-length source coding problem [25]. When we consider the second-order achievability, we have to give an appropriate first-order term. In many cases the first-order term is determined by considering the strong converse property and hence the strong converse property has an important meaning in the analysis of the second-order achievability.

Analogously to the fixed-length coding problem, we can consider the strong converse property in the meaning of the overflow probability in variable-length codes. In this subsection, we establish the strong converse theorem on the overflow probability of the variable-length coding with codeword cost. Let us begin with the definition of the strong converse property treated in this paper.

**Definition 4.1:** Source $X$ is said to satisfy the strong converse property, if any variable-length code $(\psi_n, \psi_n)$ with the
overflow probability $\epsilon_n(\phi_n, nR)$, where $R$ is an arbitrary rate satisfying $R < R(0|X)$, necessarily yields

$$\lim_{n \to \infty} \epsilon_n(\phi_n, nR) = 1.$$  

In order to state the strong converse theorem, we define the dual quantity of $\mathcal{P}(X)$ as

$$H(X) = \lim_{n \to \infty} \frac{1}{n} \log \frac{1}{P_{X^n}(X^n)},$$

which is called the spectral inf-entropy rate [25]. Then, we have the following theorem on the strong converse property.

**Theorem 4.2:** Source $X$ satisfies the strong converse property if and only if

$$\mathcal{P}(X) = H(X)$$  \hspace{1cm} (17)  

holds.

**Proof:** This theorem can be proved by using the similar argument with the proof of [25, Theorem 1.5.1]. For the completeness of the paper we give the proof in Appendix A. Theorem reveals that the strong converse property only depends on the source $X$ and is independent on the cost function.

**Remark 4.1:** For an i.i.d. source, the following relationship holds [25],

$$H(X) = \mathcal{P}(X) = H(X),$$

where $H(X)$ denotes the entropy of the source. Thus, any i.i.d. source satisfies the strong converse property. This means that the infimum of $\epsilon$-achievable overflow thresholds $R(\epsilon|X)$ is constant and is independent of $\epsilon$.

V. INFIMUM OF $(\epsilon, a)$-ACHIEVABLE OVERFLOW THRESHOLD

A. General Formula

So far, we have considered the first-order achievable threshold. In this section, we consider the second-order achievability. In the second-order case, the infimum $(\epsilon, a)$-achievable overflow threshold for general sources is also determined by using Lemma 3.1 and Lemma 3.2.

We define the function $F_\epsilon(a)$ given $a$ as follows, which is analogous to the function $F(R)$ in the first-order case,

$$F_\epsilon(a) := \lim_{n \to \infty} \sup \Pr \left\{ \frac{-\log P_{X^n}(X^n) - na}{\sqrt{na}} \geq L \right\}.$$  

Then, we have

**Theorem 5.1:** For $0 \leq \epsilon < 1$, it holds that

$$L(\epsilon, a|X) = \inf \{ L | F_\epsilon(a) \leq \epsilon \}.$$  

**Proof:** The proof is similar to the proof of Theorem 4.1. (Direct Part) Let $L_0$ be as

$$L_0 = \inf \{ L | F_\epsilon(a) \leq \epsilon \},$$

for short. Then, in this part we shall show that

$$L(\epsilon, a|X) \leq L_0 + \gamma,$$  \hspace{1cm} (18)  

for any $\gamma > 0$ by showing that $L_0$ is an $\epsilon$-achievable overflow threshold for the source. Let $a_n$ be as $a_n = n(a + \sqrt{n}(L_0 + \gamma))$, then from Lemma 3.1, for any sequence $(z_n)_{n=1}^\infty$ ($z_i > 0$, $i = 1, 2, \ldots$) there exists a variable-length encoder $\phi_n$ that satisfies

$$\epsilon_n(\phi_n, na + \sqrt{n}(L_0 + \gamma)) < \Pr \left\{ z_n P_{X^n}(X^n) \leq K^{-a_n}(na + \sqrt{n}(L_0 + \gamma)) \right\}$$

$$= \Pr \left\{ \frac{1}{\sqrt{na}} \log \frac{1}{P_{X^n}(X^n)} \geq \sqrt{na} L_0 + \gamma \right\}$$

$$+ \frac{1}{\sqrt{na}} \log \frac{1}{P_{X^n}(X^n)} \geq \sqrt{na} L_0 + \gamma + \frac{1}{\sqrt{na}} \log z_n \right\}$$

$$+ \frac{1}{\sqrt{na}} \log \frac{1}{P_{X^n}(X^n)} \geq \sqrt{na} L_0 + \gamma + \frac{1}{\sqrt{na}} \log z_n \right\}$$

$$+ \frac{1}{\sqrt{na}} \log \frac{1}{P_{X^n}(X^n)} \geq \sqrt{na} L_0 + \gamma + \frac{1}{\sqrt{na}} \log z_n \right\}$$

Let $z_n$ be as $z_n = K^{-\gamma n}$, then we have

$$\epsilon_n(\phi_n, na + \sqrt{n}(L_0 + \gamma))$$

$$< \Pr \left\{ \frac{1}{\sqrt{na}} \log \frac{1}{P_{X^n}(X^n)} \geq \sqrt{na} L_0 + \gamma \right\}$$

$$+ \frac{1}{\sqrt{na}} \log \frac{1}{P_{X^n}(X^n)} \geq \sqrt{na} L_0 + \gamma + \frac{1}{\sqrt{na}} \log z_n \right\}$$

$$+ \frac{1}{\sqrt{na}} \log \frac{1}{P_{X^n}(X^n)} \geq \sqrt{na} L_0 + \gamma + \frac{1}{\sqrt{na}} \log z_n \right\}$$

for sufficiently large $n$, because $\frac{\gamma}{2} > \frac{\gamma}{n}A$ holds for sufficiently large $n$.

By taking $\limsup_{n \to \infty}$, we have

$$\limsup_{n \to \infty} \epsilon_n(\phi_n, na + \sqrt{n}(L_0 + \gamma))$$

$$\leq \limsup_{n \to \infty} \Pr \left\{ \frac{-\log P_{X^n}(X^n) - na}{\sqrt{na}} \geq L_0 + \gamma \right\}.$$  

Hence, from the definition of $L_0$ we have

$$\limsup_{n \to \infty} \epsilon_n(\phi_n, na + \sqrt{n}(L_0 + \gamma)) \leq \epsilon.$$  

This means that (18) holds. Therefore, the direct part has been proved.

**Converse Part**

Assuming that $L_1$ satisfying

$$L_1 < \inf \{ L | F_\epsilon(a) \leq \epsilon \},$$  \hspace{1cm} (19)  


is an \((\epsilon, a)\)-achievable second-order overflow threshold, we shall show a contradiction.

From Lemma 3.2 for any sequence \(\{z_n\}_{n=1}^{\infty}\) \((z_i > 0, i = 1, 2, \ldots)\) and any variable-length encoder, it holds that
\[
\epsilon_n \left(\phi_n, na + \sqrt{nL_1}\right) \geq \Pr \left\{ P_{X^n}(X^n) \leq z_n K^{-a_c(na + \sqrt{nL_1})} \right\} - z_n,
\]
for \(n = 1, 2, \ldots\). Thus, for any variable-length encoder, we have
\[
\epsilon_n \left(\phi_n, na + \sqrt{nL_1}\right) \geq \Pr \left\{ P_{X^n}(X^n) \leq z_n K^{-a_c(na + \sqrt{nL_1})} \right\} - z_n
\]
\[
= \Pr \left\{ \frac{1}{\sqrt{nL_1}} \log P_{X^n}(X^n) + \sqrt{nL_1} \geq -\log z_n + \frac{1}{\sqrt{nL_1}} \right\} - z_n
\]
\[
= \Pr \left\{ \frac{-\log P_{X^n}(X^n) - na_c a}{\sqrt{nL_1}} \geq L_1 + \frac{1}{a_c} \right\} - z_n.
\]
Set \(z_n = K^{-\gamma}\), where \(\gamma > 0\) is a small constant that satisfies
\[
L_1 + \frac{\gamma}{a_c} < \inf \{ L \mid F_a(L) \leq \epsilon \}. \tag{20}
\]
Here, since we assume (19), it is obvious that there exists \(\gamma > 0\) satisfying the above inequality. Then, we have
\[
\epsilon_n \left(\phi_n, na + \sqrt{nL_1}\right) \geq \Pr \left\{ \frac{-\log P_{X^n}(X^n) - na_c a}{\sqrt{nL_1}} \geq L_1 + \frac{\gamma}{a_c} \right\} - K^{-\gamma}.\]
This implies that
\[
\limsup_{n \to \infty} \epsilon_n \left(\phi_n, na + \sqrt{nL_1}\right) \geq \limsup_{n \to \infty} \Pr \left\{ \frac{-\log P_{X^n}(X^n) - na_c a}{\sqrt{nL_1}} \geq L_1 + \frac{\gamma}{a_c} \right\} > \epsilon,
\]
where the last inequality is derived from (20) and the definition of \(F_a(L)\).

On the other hand, since we assume that \(L_1\) is \((\epsilon, a)\)-achievable, it holds that
\[
\limsup_{n \to \infty} \epsilon_n \left(\phi_n, na + \sqrt{nL_1}\right) \leq \epsilon.
\]
This is a contradiction. Therefore, the proof of converse part has been completed.

**Remark 5.1:** As we have shown in the above, the proof of Theorem 5.1 proceeds in parallel with that of Theorem 4.1 by using Lemma 3.1 and Lemma 3.2. We observe that we can use the same argument for other sequence \(\{\eta_n\}_{n=1}^{\infty}\) such that \(\eta_n \to \infty\). Thus, these lemmas are of great use to analyze the overflow probability of the variable-length code with codeword cost.

**B. Computation for i.i.d. Sources**

Theorem 5.1 is a quite general result, because there is no restriction about the probability structure for the source. However, to compute the function \(L \left(\epsilon, \frac{1}{a_c} H(X)\right)\) is hard in general. Next, we consider a simple case such as an i.i.d. source with countably infinite alphabet and we address the above quantity explicitly.

For an i.i.d. source, from Remark 4.1, we are interested in \(L \left(\epsilon, \frac{1}{a_c} H(X)\right)\). To specify this quantity for an i.i.d. source, we need to introduce the variance of the self-information as follows:
\[
\sigma^2 := E \left(\left(\log P_X(X) - H(X)\right)^2\right).
\]
Here, we assume that the above variance exists and \(\sigma^2 > 0\). Then, from Theorem 5.1 we obtain the following theorem.

**Theorem 5.2:** For any i.i.d. source, it holds that
\[
L \left(\epsilon, \frac{1}{a_c} H(X)\right) = \frac{1}{a_c} \sigma \Phi^{-1}(1 - \epsilon),
\]
where \(\Phi^{-1}\) denotes a inverse function of \(\Phi\) and \(\Phi(T)\) is the Gaussian cumulative distribution function with mean 0 and variance 1, that is, \(\Phi(T)\) is given by
\[
\Phi(T) = \int_{-\infty}^{T} \frac{1}{\sqrt{2\pi}} \exp \left(\frac{-y^2}{2}\right) dy. \tag{21}
\]
**Proof:** From the definition of \(F_a(L)\), we have
\[
F_{H(X)/a_c}(L) = \limsup_{n \to \infty} \Pr \left\{ \frac{-\log P_{X^n}(X^n) - nH(X)}{\sqrt{nL_1}} \geq L \right\}
\]
\[
= \limsup_{n \to \infty} \Pr \left\{ \frac{-\log P_{X^n}(X^n) - nH(X)}{\sqrt{nL_1}} \geq L \right\}.
\]
On the other hand, since we consider the i.i.d. source, from the asymptotic normality (due to the central limit theorem) it holds that
\[
\limsup_{n \to \infty} \Pr \left\{ \frac{-\log P_{X^n}(X^n) - nH(X)}{\sqrt{nL_1}} \leq U \right\} = \int_{-\infty}^{U} \frac{1}{\sqrt{2\pi}} \exp \left(\frac{-y^2}{2}\right) dy.
\]
This means that
\[
F_{H(X)/a_c}(L) = \int_{-\infty}^{U} \frac{1}{\sqrt{2\pi}} \exp \left(\frac{-y^2}{2}\right) dy.
\]
Thus, \(L \left(\epsilon, \frac{1}{a_c} H(X)\right)\) is given by
\[
L \left(\epsilon, \frac{1}{a_c} H(X)\right) = \inf \left\{ L \mid \int_{-\infty}^{U} \frac{1}{\sqrt{2\pi}} \exp \left(\frac{-y^2}{2}\right) dy \leq \epsilon \right\}
\]
\[
= \inf \left\{ L \mid 1 - \Phi \left(\frac{L a_c}{\sigma}\right) \leq \epsilon \right\}.
\]
Since \(\Phi \left(\frac{L a_c}{\sigma}\right)\) is a continuous function and monotonically increases as \(L\) increases, we have
\[
L \left(\epsilon, \frac{1}{a_c} H(X)\right) = \Phi^{-1}(1 - \epsilon).
\]
Therefore, the proof has been completed.

**Remark 5.2:** As shown in the proof, the derivation of Theorem 5.2 is based on of the asymptotic normality of the self-information. This means that the similar argument is valid for any source for which the asymptotic normality of self-information holds such as Markov sources (see, Hayashi [20]).
C. Computation for Mixed Sources

In this subsection we consider mixed sources. The class of mixed sources is very important, because all of stationary sources can be regarded as forming mixed sources obtained by mixing stationary ergodic sources with respect to appropriate probability measures. Notice here that, in general, the mixed source does not have the asymptotic normality of the self-information. So, we cannot simply apply Theorem 5.2. The second-order achievable rates for mixed sources has been first considered by Nomura and Han [21] in the fixed-length source coding problem. The result in this subsection is a variable-length coding counterpart of the result in [21].

We consider the mixed source consisting of two stationary memoryless sources \( X_i = \{X_{ni}\}_{n=1}^\infty \) with \( i = 1, 2 \). Then, the mixed source \( X = \{X^n\}_{n=1}^\infty \) is defined by

\[
P_{X^n}(x) = w(1) P_{X_1^n}(x) + w(2) P_{X_2^n}(x),
\]

where \( w(i) \) are constants satisfying \( w(1) + w(2) = 1 \) and \( w(i) > 0 \) \( (i = 1, 2) \). Since two i.i.d. sources \( X_i \) \( (i = 1, 2) \) are completely specified by giving just the first component \( X_i \) \( (i = 1, 2) \), we may write simply as \( X_i = \{X_i\} \) \( (i = 1, 2) \) and define the variances:

\[
\sigma_i^2 := E \left( \log \frac{1}{P_{X_i}(X_i)} - H(X_i) \right)^2 (i = 1, 2),
\]

where we assume that these variances exist and \( \sigma_i^2 > 0 \) \( (i = 1, 2) \).

Before showing second-order analysis we shall consider the first-order case. Without loss of generality, we assume that \( H(X_1) \geq H(X_2) \) holds.

**Theorem 5.3:** For any mixed source defined by (22), we have

\[
R(\epsilon|X) = \begin{cases} 
\frac{H(X_1)}{\alpha_1} & \text{if } 0 \leq \epsilon < w(1), \\
\frac{H(X_2)}{\alpha_2} & \text{if } w(1) \leq \epsilon < 1.
\end{cases}
\]

**Proof:** This theorem can be obtained as an immediate consequence of Theorem 4.1.

By using the above theorem we shall compute the second-order case. As we have mentioned in the above, the asymptotic normality of self-information does not hold for mixed sources. However, since we consider the case where \( X_i = \{X_i\} \) \( (i = 1, 2) \) is an i.i.d. source, the following asymptotic normality holds for each component i.i.d. source:

\[
\lim_{n \to \infty} \Pr \left\{ \frac{1}{\sqrt{n \sigma_i}} \left( -\log P_{X_i^n}(X_i^n) - n H(X_i) \right) \leq U \right\} = \Phi(U).
\]

The following lemma plays the key role in dealing with mixed sources in the proof of Theorem 5.4.

**Lemma 5.1 (Han [25], (see also [21, Lemma 4.1]):** Let \( \{z_n\}_{n=1}^\infty \) be any real-valued sequence. Then for the mixed source \( X \) it holds that, for \( i = 1, 2 \),

\[
\Pr \left\{ \frac{-\log P_{X_i^n}(X_i^n)}{\sqrt{n \sigma_i}} \geq z_n \right\} \geq \Pr \left\{ \frac{-\log P_{X_i^n}(X_i^n)}{\sqrt{n \sigma_i}} \geq z_n + \gamma_n \right\} - e^{-\sqrt{n \gamma_n}},
\]

\[
\Pr \left\{ \frac{-\log P_{X_i^n}(X_i^n)}{\sqrt{n \sigma_i}} \geq z_n \right\} \leq \Pr \left\{ \frac{-\log P_{X_i^n}(X_i^n)}{\sqrt{n \sigma_i}} \geq z_n - \gamma_n \right\},
\]

where \( \gamma_n > 0 \) satisfies \( \gamma_1 > \gamma_2 > \cdots > 0 \), \( \gamma_n \to 0 \), \( \sqrt{n \gamma_n} \to \infty \).

In the sequel, we also assume that \( H(X_1) \geq H(X_2) \) holds without loss of generality. Then, given \( \epsilon \) \( (0 \leq \epsilon < 1) \) we classify the problem into three cases:

- **Case I** \( H(X_1) = H(X_2) \) holds.
- **Case II** \( H(X_1) > H(X_2) \) and \( w(1) > \epsilon \) hold.
- **Case III** \( H(X_1) > H(X_2) \) and \( w(1) < \epsilon \) hold.

See Remark 5.3, in the case where \( H(X_1) > H(X_2) \) and \( w(1) = \epsilon \) hold. In Case I, we shall compute \( L(\epsilon, \frac{1}{\alpha_c} H(X_1)|X) \) (this is equal to \( L(\epsilon, \frac{1}{\alpha_c} H(X_2)|X) \)).

In Case II and Case III, we shall show \( L(\epsilon, \frac{1}{\alpha_c} H(X_1)|X) \) and \( L(\epsilon, \frac{1}{\alpha_c} H(X_2)|X) \), respectively. Then, from Theorem 5.1 we obtain the following theorem:

**Theorem 5.4:** For any mixed source, it holds that

**Case I** \[
L\left(\epsilon, \frac{1}{\alpha_c} H(X_1)\big|X\right) = T_1,
\]

where \( T_1 \) is specified by

\[
\epsilon = 1 - \sum_{i=1}^{2} w(i) \Phi\left( \frac{\alpha_i T_1}{\sigma_i} \right).
\]

**Case II** \[
L\left(\epsilon, \frac{1}{\alpha_c} H(X_1)\big|X\right) = T_2,
\]

where \( T_2 \) is specified by

\[
\epsilon = w(1) \left( 1 - \Phi\left( \frac{\alpha_c T_2}{\sigma_1} \right) \right).
\]

**Case III** \[
L\left(\epsilon, \frac{1}{\alpha_c} H(X_2)\big|X\right) = T_3,
\]

where \( T_3 \) is specified by

\[
\epsilon = w(1) + w(2) \left( 1 - \Phi\left( \frac{\alpha_c T_3}{\sigma_2} \right) \right).
\]

**Proof:** This theorem can be shown substantially same with [21, Theorem 5.1]. We only show the proof of Case I in Appendix B.

**Remark 5.3:** In the case where \( H(X_1) > H(X_2) \) and \( w(1) = \epsilon \) hold, it is not difficult to check that \( T_2 = -\infty \).
and \( T_3 = +\infty \) (cf. [21, Remark 5.2]). Hence, in this case we obtain
\[
L \left( \alpha, \frac{1}{\alpha_c} H(X_1) \bigg| X \right) = -\infty, \\
L \left( \alpha, \frac{1}{\alpha_c} H(X_2) \bigg| X \right) = +\infty.
\]

**Example 5.1:** We here give numerical examples of Theorem 5.4. Let us consider the mixed sources
\[
P_{X^n}(x) = 0.5 P_{X_1^n}(x) + 0.5 P_{X_2^n}(x)
\]
with alphabets \( \mathcal{X} = \{0, 1, 2\} \) and \( \mathcal{U} = \mathcal{X} \). Moreover, we consider the cost function satisfying \( c(0) = 1, c(1) = 1, \) and \( c(2) = 2 \). In this setting, the cost capacity \( \alpha_c \) is given by the unique solution of
\[
3^{-\alpha} + 3^{-\alpha} + 3^{-2\alpha} = 1.
\]
Thus, we have \( \alpha_c = \log_3(1 + \sqrt{2}) \).

First, we consider Case I, that is, \( H(X_1) = H(X_2) = 0.8 \) with \( P_{X_1}(0) = 0.2, P_{X_1}(1) \approx 0.145, P_{X_1}(2) \approx 0.655, P_{X_2}(0) = 0.4, P_{X_2}(1) \approx 0.065 \) and \( P_{X_2}(2) \approx 0.535 \). Then, we have \( \sigma_1^2 \approx 0.334 \) and \( \sigma_2^2 \approx 0.214 \). Hence, we have
\[
H(X_1) \quad \frac{H(X_2)}{\alpha_c} \approx 0.997.
\]
Moreover, let us consider the case \( n = 200 \). In this case it holds that
\[
\frac{\sigma_1}{\alpha_c \sqrt{n}} \approx 0.051, \quad \frac{\sigma_2}{\alpha_c \sqrt{n}} \approx 0.041.
\]

Fig. 1 illustrates this case with \( \varepsilon = 0.06 \). From Theorem 5.4, \( T_1 \approx 0.07 \) is computed by using two Gaussian distributions with mean 0.997 and \( \sigma = 0.051 \), and mean 0.997 and \( \sigma = 0.041 \) such that the probability of the weighted sum of shaded regions equals to \( \varepsilon = 0.06 \). We also observe that the infimum of the second-order achievable rate may be negative, when \( \varepsilon > 0.5 \).

Secondly, we shall give examples of Case II and Case III. We assume that the probability of each source symbol is given by \( P_{X_1}(0) = 0.5, P_{X_1}(1) = 0.3, P_{X_1}(2) = 0.2, P_{X_2}(0) = 0.7, P_{X_2}(1) = 0.2, P_{X_2}(2) = 0.1 \), respectively.

In this case, we have \( H(X_1) \approx 0.937, \) \( H(X_2) \approx 0.730, \) \( \sigma_1^2 \approx 0.11, \) and \( \sigma_2^2 \approx 0.41 \). Hence, we have
\[
\frac{H(X_1)}{\alpha_c} \approx 1.169, \quad \frac{H(X_2)}{\alpha_c} \approx 0.910.
\]
Moreover, let us consider the case \( n = 300 \). Then, it holds that
\[
\frac{\sigma_1}{\alpha_c \sqrt{n}} \approx 0.023, \quad \frac{\sigma_2}{\alpha_c \sqrt{n}} \approx 0.046.
\]

Fig. 2 and Fig. 3 illustrate Case II and Case III in this numerical example. It should be noted that the area of shaded regions is equal to \( \varepsilon = 0.025 \) (Fig. 2) and \( \varepsilon = 0.52 \) (Fig. 3). Then, from Theorem 5.4, \( T_2 \approx 0.038 \) and \( T_3 \approx 0.071 \) are computed by using two Gaussian distributions with mean 1.169 and \( \sigma = 0.023 \), and mean 0.910 and \( \sigma = 0.046 \).

We also observe that the infimum of the second-order achievable rate may be negative, when \( 0.25 < \varepsilon < 0.5 \) or \( 0.75 < \varepsilon < 1 \) holds.

**Remark 5.4:** In [21], the countably infinite mixture of i.i.d. sources and the general mixture of i.i.d. sources are treated. We can also obtain the infimum of \( (\varepsilon, \alpha) \)-achievable overflow threshold in these cases by using the similar argument.
VI. GENERALIZATION OF THE COST FUNCTION

So far, we have considered the additive cost consisting of the fixed memoryless cost function. In this section, we will extend our results to more general cost function, that is, the conditional cost function. The conditional cost function is the function that the cost of the code symbol $u_i$ depends on the previous sequence $u_{i-1}$. In the application such as the video coding or the coding on magnetic data storage, constrained codes have been utilized [26]–[28]. In particular, the symmetric fix-free code, the run-length constraint codes are representative. The symmetric constraint is the constraint that each codeword is required to be a palindrome. The run-length constraint code is the code in which the number of consecutive entries in a codeword is restricted. In each of these cases, it can be considered that the cost of $u_i$ depends on the previous sequence $u_{i-1}$. Furthermore, the conditional cost function considered in this section includes the regular cost function (see, Remark 6.1 for the definition of the regular cost function). The relationship between the variable-length coding with regular cost function and the simulating Markov process has been clarified [14]. Hence, the conditional cost function is important and meaningful to discuss from both the practical and theoretical viewpoints.

The conditional cost function treated in this section is defined as follows. The codeword cost $c(u')$ of a codeword $u' \in \mathcal{U}^d$ is defined by

$$c(u') = \sum_{i=1}^{l} c(u_i | u_{i-1}'),$$

where $c(u_i | u_{i-1}')$ is called the conditional cost of $u_i$ given $u_{i-1}'$ such that $0 < c_{\min} \leq c(u_i | u_{i-1}') \leq c_{\max} < \infty$ ($\forall i \in \mathcal{U}$, $\forall u_i \in \mathcal{U}$, $\forall u_{i-1}' \in \mathcal{U}^{l-1}$), where $c_{\min}$ and $c_{\max}$ are some constants. Here, we set $u_0'$ is the null string. We call this the conditional cost in this paper.

The conditional cost capacity $a_c(u_{i-1}')$ given $u_{i-1}'$ is defined by the positive unique root $\alpha$ of the equation

$$\sum_{u_i \in \mathcal{U}} K^{-\alpha c(u_i | u_{i-1}')} = 1,$$

for $u_{i-1}' \in \mathcal{U}^{l-1}$. It should be noted that $a_c(u_{i-1}')$ may vary according to $u_{i-1}'$. Moreover, we set

$$\overline{a}_{\alpha} = \sup_{u_i \in \mathcal{U}} \alpha_c(u), \quad \underline{a}_{\alpha} = \inf_{u_i \in \mathcal{U}} \alpha_c(u),$$

and assume that $0 < \underline{a}_{\alpha} \leq \overline{a}_{\alpha} < \infty$ holds.

In order to establish the coding theorem in this setting, we first have the following lemma instead of (1).

**Lemma 6.1:** The variable-length code $D = \{u_1, u_2, \ldots\}$ with the prefix condition over the conditional cost satisfies

$$\sum_{u_i \in \mathcal{U}} K^{-\sum_{i=1}^{m} \alpha_c(u_{i-1}') c(u_i | u_{i-1}')} \leq 1,$$

where $I(u)$ denotes the length of the codeword $u$.

**Proof:** We define

$$q(u) := K^{-\sum_{i=1}^{m} \alpha_c(u_{i-1}') c(u_i | u_{i-1}')}$$

for any $u \in D$. Then, the prefix condition and (29) assure that

$$\sum_{u \in D} q(u) \leq 1$$

holds (cf. [14, Lemma 1]).

Secondly, we have the following lemmas which are generalizations of Lemmas 3.1 and 3.2.

**Lemma 6.2:** For any general sources $X$ and any sequence of positive number $\{\eta_n\}_{n=1}^{\infty}$, there exists a variable-length encoder $\phi_n$ that satisfies

$$c_n(\phi_n, \eta_n) \leq \Pr \{z_n P_X^*(X^n) \leq K^{-\alpha_n} \eta_n \} + z_n K^{\alpha_n} c_{\max} + 1,$$

for $n = 1, 2, \ldots$, where $\{z_n\}_{n=1}^{\infty}$ is a given sequence of an arbitrary number satisfying $z_i > 0$ for $i = 1, 2, \ldots$.

**Proof:** First, we extend the code $\phi_n$ proposed by Han and Uchida [14] (see also, Yagi and Nomura [15]) to the cost function treated in this section. To do so, we assume that all the elements in $X^n$ are indexed as $(x_1, x_2, \ldots)$. We define

$$P_k := \sum_{j=1}^{k} P_X^*(x_j),$$

$$Q_k := P_k + \frac{1}{2} P_X^*(x_k),$$

for $k = 1, 2, \ldots$, where $P_1 = 0$. Moreover set

$$\beta(u) := \sum_{u' \prec u} q(u'), \quad \gamma(u) := \beta(u) + q(u),$$

and

$$I(u) := [\beta(u), \gamma(u)],$$

where $\prec$ denotes the lexicographic order on the set $\mathcal{U}^d$. Now, to each $x_k$ we uniquely assign $u_k$ such as

$$u_k = \arg \min_{u \in \mathcal{U}^d} I(u),$$

where $\mathcal{U}^d$ denotes the set of $u \in \mathcal{U}^d$ such that the interval $I(u)$ includes $Q_k$ but neither $P_k$ nor $P_{k+1}$. Here, we can verify the existence $u_k$ for $\forall x_k \in X^n$ as follows. From (29) we obtain

$$\sum_{u \in \mathcal{U}^d} q(u) = \sum_{u \in \mathcal{U}^d} K^{-\sum_{i=1}^{m} \alpha_c(u_{i-1}') c(u_i | u_{i-1}')} = 1,$$

for all $m = 1, 2, \ldots$. Furthermore, from the assumption that $\underline{a}_{\alpha} > 0$ and $c_{\min} > 0$, it holds that $I(u_{i-1}') \subseteq I(u_i')$ for any $u_{i-1}' \in \mathcal{U}^d$. Hence, for $\forall x_k \in X^n$ there exists $u_k$ satisfying (32).

Then, clearly $(P_k, P_{k+1}) \supseteq I(u)$ holds and each interval $I(u_1), I(u_2), \ldots$ is disjoint. This implies that $(u_1, u_2, \ldots)$ satisfies the prefix condition. Thus, we can define the prefix encoder

$$\phi_n^\ast(x_k) = u_k.$$
Finally, we have the following theorem, which reveals the

\[ \frac{P_{X^n}(x_k)}{2} < |I(\mathbf{u}_k)| \]

\[ \leq K^{-\sum_{i=1}^{n} a_i(u_{i-1}) c(u_i | u_{i-1})} \]

\[ \leq K^{-\sum_{i=1}^{n} a_i(u_{i-1}) c(u_i | u_{i-1})} \]

Only, we obtain

\[ c(\mathbf{u}_k) = -\frac{1}{\log_2 \mathcal{X}} \log P_{X^n}(x_k) + \frac{\log 2}{\log_2 \mathcal{X}}. \] (33)

Hence, we obtain

\[ c(\varphi_n(x_k)) \leq c(\mathbf{u}_k) + c_{max} \]

\[ c(\varphi_n(x_k)) \leq -\frac{1}{\log_2 \mathcal{X}} \log P_{X^n}(x_k) + \frac{\log 2}{\log_2 \mathcal{X}} + c_{max}. \] (34)

Now, we can prove the lemma by using the similar argument as that in the proof of Lemma 3.1. Note that we use the inequality (34) instead of (6).

**Lemma 6.3:** For any variable-length code and any sequence \( \{\eta_n\}_{n=1}^{\infty} \), it holds that

\[ e_n(\varphi_n, \eta_n) \geq \Pr \left\{ P_{X^n}(x^n) \leq z_n K^{-\varphi_n} \right\} - z_n, \]

for \( n = 1, 2, \ldots \), where \( \{z_n\}_{n=1}^{\infty} \) is a given sequence of an arbitrary number satisfying \( z_i > 0 \) for \( i = 1, 2, \ldots \).

**Proof:** From Lemma 6.1, for any prefix code \( D \) we obtain

\[ 1 \geq \sum_{x \in \mathcal{X}^n} K^{-\sum_{i=1}^{n} a_i(u_{i-1}) c(u_i | u_{i-1})} \]

\[ \geq \sum_{x \in \mathcal{X}^n} K^{-\eta_n} \sum_{i=1}^{n} c(u_i | u_{i-1}) \]

\[ = \sum_{x \in \mathcal{X}^n} K^{-\eta_n} c(u). \]

This means that

\[ 1 \geq \sum_{x \in \mathcal{X}^n} K^{-\eta_n} c(\varphi_n(x)). \] (35)

Then, the proof of the lemma proceeds in parallel with the proof of Lemma 3.2. Note that we use the inequality (35) instead of (1).

Finally, we have the following theorem, which reveals the upper and lower bounds of the infimum of the first-order achievable overflow threshold.

**Theorem 6.1:** For \( 0 \leq \forall \epsilon < 1 \), it holds that

\[ \inf \left\{ R \left| F(R) \leq \epsilon \right. \right\} \leq R(\epsilon | X) \leq \inf \left\{ R \left| F^*(R) \leq \epsilon \right. \right\}, \]

where \( F(R) \) and \( F^*(R) \) are respectively defined by

\[ F(R) := \lim_{n \to \infty} \Pr \left\{ \frac{1}{n \log_2 \mathcal{X}} \log \frac{1}{P_{X^n}(x^n)} \geq R \right\}, \]

\[ F^*(R) := \lim_{n \to \infty} \Pr \left\{ \frac{1}{n \log_2 \mathcal{X}} \log \frac{1}{P_{X^n}(x^n)} \geq R \right\}. \]

**Proof:** By using Lemmas 6.2 and 6.3 it suffices to proceed in parallel with the arguments as made in the proof of Theorem 4.1.

We immediately have the following corollary, which subsumes Theorem 4.1.

**Corollary 6.1:** Assuming that \( \varphi_c = \alpha_c \) holds, then we have

\[ R(\epsilon | X) = \inf \left\{ R \left| \lim_{n \to \infty} \Pr \left\{ \frac{1}{n \log_2 \mathcal{X}} \log \frac{1}{P_{X^n}(x^n)} \geq R \right\} \leq \epsilon \right. \right\}, \]

for \( 0 \leq \epsilon < 1. \)

**Remark 6.1:** As a special case of this conditional cost, we can consider the case that the conditional cost capacity \( a_c(u) = \alpha \) holds for all \( u \in \mathcal{U}^n \). Such a class of cost functions, said to be regular, has been first considered in [29], Han and Uchida [14] also have treated the regular cost function. This is a typical example where \( \varphi_c = \alpha_c \) holds.

**Remark 6.2:** In this section, we have only discussed the infimum of the first-order achievable threshold. It should be noted that we can also derive the upper and lower bounds of the infimum of the second-order achievable threshold by using Lemmas 6.2 and 6.3.

**VII. CONCLUDING REMARKS**

We have so far dealt with the overflow probability of the variable-length coding with codeword cost. As shown in the present paper, the information-spectrum approach is substantial in the analysis of the overflow probability of the variable-length coding.

In particular, Lemmas 3.1 and 3.2 are quite fundamental, because the first-order coding theorem (Theorem 4.1) and the second-order coding theorem (Theorem 5.1) have been established by invoking these lemmas. It should be emphasized that these theorems can be proved by using the similar argument. In the channel coding problem, the third-order optimal achievable rate has been discussed [30], [31]. Also in the variable-length coding regime, the optimal achievable overflow threshold with respect to the codeword length has been discussed in detail in [22]. If we set \( \eta_n \) appropriately in Lemmas 3.1 and 3.2, we can also derive the general formula of the optimal third- or higher-order overflow threshold for the variable-length coding with codeword cost.

Finally, we have extended our results for the memoryless cost to the conditional cost. Another typical but important cost is the finite-state cost [7] which admits \( c_{min} = 0 \). It is an interesting problem to extend our result to this finite-state cost.

**APPENDIX A**

**Proof of Theorem 4.2**

The proof consists of two parts.

(1) We assume that (17) holds. Set \( R = R(0 | X) - \frac{2\gamma}{\alpha_c} \), where \( \gamma > 0 \) is an arbitrary constant. Then, from Corollary 4.1 it holds that

\[ R = \frac{1}{\alpha_c} \mathcal{H}(X) - \frac{2\gamma}{\alpha_c}. \]
On the other hand, from Lemma 3.2 with \( \eta_n = nR \), it holds that

\[
\epsilon_n(\varphi_n, nR) > \Pr \left\{ \frac{1}{n \alpha_c} \log \frac{1}{P_{X^n}(X^n)} \geq R - \frac{\gamma}{\alpha_c} \right\} - z_n,
\]

for any sequence \( \{z_n\}_{n=1}^{\infty} \) (\( z_i > 0, i = 1, 2, \ldots \)). Let \( z_n \) be as \( K^{-n \gamma} \), then we have

\[
\epsilon_n(\varphi_n, nR) > \Pr \left\{ \frac{1}{n \alpha_c} \log \frac{1}{P_{X^n}(X^n)} \geq R + \frac{\gamma}{\alpha_c} \right\} - K^{-n \gamma},
\]

\[
= \Pr \left\{ \frac{1}{n \alpha_c} \log \frac{1}{P_{X^n}(X^n)} \geq \frac{1}{\alpha_c} \overline{H}(X) - \frac{\gamma}{\alpha_c} \right\} - K^{-n \gamma},
\]

Noting that \( \gamma > 0 \) is a constant, from the definition of \( \overline{H}(X) \), we have

\[
\lim_{n \to \infty} \epsilon_n(\varphi_n, nR) = 1.
\]

Therefore, the sufficiency has been proved.

\textbf{APPENDIX B}

\textbf{PROOF OF THEOREM 5.4}

We only show Case I. The proofs of Case II and Case III are similar to that of Case I and [21, Theorem 5.1]

The last equality is derived from the definition of the mixed source. Thus, from Lemma 5.1 we have

\[
\sum_{i=1}^{2} \lim_{n \to \infty} \Pr \left\{ \frac{-\log P_{X^n}(X^n) - nH(X_1)}{\sqrt{n \alpha_c}} \geq L - \gamma_n \right\} w(i) \geq F_{H(X_1)/\alpha_c}(L),
\]

where \( \gamma_n \) is specified in Lemma 5.1.

Then, noting that \( H(X_1) = H(X_2) \) holds, from the asymptotic normality, we have

\[
\lim_{n \to \infty} \Pr \left\{ \frac{-\log P_{X^n}(X^n) - nH(X_1)}{\sqrt{n \alpha_c}} \geq \frac{L\alpha_c}{\sigma_i} \right\} = \int_{\frac{L\alpha_c}{\sigma_i}}^{\infty} \frac{1}{\sqrt{2\pi}} \exp \left[ -\frac{z^2}{2} \right] dz = 1 - \Phi \left( \frac{L\alpha_c}{\sigma_i} \right),
\]

for \( i = 1, 2 \). Noting that \( \gamma_n \to 0 \) as \( n \to \infty \) and the continuity of the normal distribution function, we have

\[
\sum_{i=1}^{2} \lim_{n \to \infty} \Pr \left\{ \frac{-\log P_{X^n}(X^n) - nH(X_1)}{\sqrt{n \alpha_c}} \geq L - \gamma_n \right\} w(i) = 1 - \frac{2}{\alpha_c} \Phi \left( \frac{L\alpha_c}{\sigma_i} \right).
\]

Similarly, the last term in (B.1) is evaluated as

\[
\sum_{i=1}^{2} \lim_{n \to \infty} \Pr \left\{ \frac{-\log P_{X^n}(X^n) - nH(X_1)}{\sqrt{n \alpha_c}} \geq L + \gamma_n \right\} w(i) = 1 - \frac{2}{\alpha_c} \Phi \left( \frac{L\alpha_c}{\sigma_i} \right).
\]

Hence, we have proved Case I of the theorem. 

\textbf{ACKNOWLEDGMENT}

The author is grateful to Prof. Toshiyasu Matsushima of Waseda University for his valuable discussions and comments. The author also would like to thank Associate Editor, Prof. Yasutada Oohama and the anonymous reviewers for their valuable comments which have improved the content of this paper.
REFERENCES

[1] R. Nomura and T. Matsushima, “Information spectrum approach to overflow probability of variable-length codes with conditional cost function,” in Proc. IEEE Int. Symp. Inf. Theory, Jul. 2012, pp. 1102–1106.
[2] C. E. Shannon, “A mathematical theory of communication,” Bell Syst. Tech. J., vol. 27, no. 3, pp. 379–423 and 623–656, Jul./Oct. 1948.
[3] T. M. Cover and J. A. Thomas, Elements of Information Theory. Hoboken, NJ, USA: Wiley, 1991.
[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., vol. E84-A, no. 10, pp. 2457–2465, 2001.
[6] R. Nomura and T. Matsushima, “On the overflow probability of fixed-to-variable length codes with side information,” IEICE Trans. Fundam., vol. E94-A, no. 11, pp. 2083–2091, 2011.
[7] I. Csiszár, “Simple proofs of some theorems on noiseless channels,” Inf. Control, vol. 14, no. 3, pp. 285–298, 1969.
[8] I. Csiszár and J. Körner, Information Theory: Coding Theorems for Discrete Memoryless Systems. New York, NY, USA: Academic, 1981.
[9] R. M. Karp, “Channels which transmit letters of unequal duration,” Inf. Control, vol. 5, pp. 13–24, Mar. 1962.
[10] R. M. Karp, “Minimum-redundancy coding for the discrete noiseless channel,” IEEE Trans. Inf. Theory, vol. 7, pp. 27–38, Jan. 1961.
[11] M. J. Golin and J. Li, “More efficient algorithms and analyses for unequal letter cost prefix-free coding,” IEEE Trans. Inf. Theory, vol. 45, no. 4, pp. 1389–1399, Sep. 1999.
[12] 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.
[13] 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.
[14] 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.
[15] C. E. Shannon, “A mathematical theory of communication,” Bell Syst. Tech. J., vol. 27, no. 3, pp. 379–423 and 623–656, Jul./Oct. 1948.
[16] T. S. Han and O. Uchida, “Universal fixed-to-variable source coding in the finite blocklength regime,” in Proc. IEEE Int. Symp. Inf. Theory, Jul. 2013, pp. 649–653.
[17] O. Kosut and L. Sankar, “Asymptotics and non-asymptotics for universal fixed-to-variable source coding,” IEEE Trans. Inf. Theory, vol. 63, no. 6, pp. 3757–3772, Jun. 2017.
[18] T. S. Han, Information-Spectrum Methods in Information Theory. New York, NY, USA: Springer, 2003.
[19] S. A. Savari, S. M. H. T. Yazdi, N. Abedini, and S. P. Khatri, “On optimal and achievable fix-free codes,” IEEE Trans. Inf. Theory, vol. 58, no. 8, pp. 5112–5129, Aug. 2012.
[20] S. M. H. T. Yazdi and S. A. Savari, “On the relationships among optimal symmetric fix-free codes,” IEEE Trans. Inf. Theory, vol. 60, no. 8, pp. 4567–4583, Aug. 2014.
[21] K. A. S. Immink, “A practical method for approaching the channel capacity of constrained channels,” IEEE Trans. Inf. Theory, vol. 43, no. 8, pp. 1389–1399, Sep. 1997.
[22] Ryo Nomura

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.