# Non-Asymptotic and Asymptotic Analyses of Source Coding: An Approach from the Viewpoint of the Overflow Probability

# 情報源符号化の有限長解析と漸近解析
## —オーバーフロー確率に着目したアプローチ—

February 2018

Shota SAITO

齋藤 翔太

# Non-Asymptotic and Asymptotic Analyses of Source Coding: An Approach from the Viewpoint of the Overflow Probability

# 情報源符号化の有限長解析と漸近解析 —オーバーフロー確率に着目したアプローチ—

February 2018

Waseda University

Graduate School of Fundamental Science and Engineering

Department of Pure and Applied Mathematics, Research on Information Theory

Shota SAITO

齋藤 翔太

# Acknowledgements

I would like to thank my adviser Professor Toshiyasu Matsushima at Waseda University for his comments, suggestions, support, and encouragement during my time as a undergraduate and graduate student at Waseda University. I was greatly influenced by his viewpoint and enthusiasm for research. Furthermore, his guidance enabled me to explore many interesting topics in information theory and related research areas such as probability theory, statistics, error correcting codes, and machine learning.

I am grateful to my co-advisers Professors Shin'ichi Oishi, Masahide Kashiwagi, Yasutaka Shimizu, and Shigeichi Hirasawa at Waseda University for their advice during the revisions of the dissertation. In particular, Dr. Shigeichi Hirasawa gave me continuous encouragement during my time as a graduate student at Waseda University.

I wish to acknowledge valuable discussions with Dr. Ryo Nomura at Senshu University, Dr. Hideki Yagi at the University of Electro-Communications, Dr. Nozomi Miya at Aoyama Gakuin University, Dr. Shunsuke Horii at Waseda University, Dr. Tota Suko at Waseda University, Professor Manabu Kobayashi at Shonan Institute of Technology, Dr. Goki Yasuda at Toshiba Corporation, Dr. Tomohiko Saito at Shonan Institute of Technology, Professor Yoshihumi Ukita at Yokohama College of Commerce, Professor Yasunari Maeda at Kitami Institute of Technology, Dr. Takahiro Yoshida at Yokohama College of Commerce, Dr. Daiki Koizumi at Otaru University of Commerce, Dr. Takashi Suenaga at NTT DATA Corporation, Dr. Shuhei Kuwata at NTT DOCOMO, INC, Dr. Kazuhiko Minematsu at NEC Corporation, Professor Masayuki Goto at Waseda University. In particular, I am indebted to Dr. Ryo Nomura and Dr. Hideki Yagi for their helpful comments and suggestions.

I would like to thank all the members of Matsushima laboratory for their help and discussions. Especially, I appreciate interesting discussions with Mr. Yuta Nakahara and Mr. Jun Yoshizawa.

I am grateful to anonymous reviewers and associate editors for their helpful comments when I submitted the article to the journal. Their comments improved the contents and presentation of my research.

Finally, I really appreciate the warm support of my family: my parents, sister, and grandparents.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

## 1.1 Background

Information theory is a basic mathematical theory that deals with a digital communication. It started from the work by C. E. Shannon in 1948 [39]. Today, more and more digital data are exchanged on a network. Further, a large amount of data can be obtained in recent years. To accumulate, transmit, and analyze such digital data, information theory plays a fundamental role.

A *source coding* is one of the fundamental and important research topics in information theory. To explain the mathematical model of source coding, we take *variable-length lossless source coding* as an example. A stochastic process $\{X_i\}_{i=1}^{\infty}$ is said to be a *source*, where $X_1, X_2, \ldots,$ are discrete random variables. Various assumptions on $\{X_i\}_{i=1}^{\infty}$ define various sources. For example, $\{X_i\}_{i=1}^{\infty}$ is said to be a *stationary memoryless source* if $X_1, X_2, \ldots,$ are i.i.d. random variables; a source $\{X_i\}_{i=1}^{\infty}$ is said to be a *stationary ergodic source* if $\{X_i\}_{i=1}^{\infty}$ satisfies the stationarity and ergodicity. Let $x_1, x_2, \ldots, x_n$ be a realization of the random variables $X_1, X_2, \ldots, X_n$. For a shorthand notation, we denote $X_1, X_2, \ldots, X_n$ as $X^n$ and $x_1, x_2, \ldots, x_n$ as $x^n$. A sequence $x^n$ are mapped to another sequence by an injective function (called an *encoder*)

$$f_n : \mathcal{X}^n \to \mathcal{U}^\star, \tag{1.1}$$

where $\mathcal{U} := \{0, 1, \cdots, K-1\}$ ($K$ is a positive integer greater than 2) and $\mathcal{U}^\star$ is a set of all sequences composed of elements of $\mathcal{U}$ and the empty string $\Lambda$. For example, if $K = 2$, then $\mathcal{U} = \{0, 1\}$ and

$$\mathcal{U}^\star = \{\Lambda, 0, 1, 00, 01, 10, 11, 000, 001, \cdots\}. \tag{1.2}$$

An element of $\mathcal{U}^\star$ is said to be a *codeword*. Length of a codeword is said to be a *codeword length* and denoted by $\ell(f_n(x^n))$ for a source sequence $x^n$. After the codeword is sent to a receiver, the receiver recovers the original sequence $x^n$ by a function (called a *decoder*)

$$g_n : \mathcal{U}^\star \to \mathcal{X}^n \tag{1.3}$$

such that

$$\mathbb{P}[X^n \neq g_n(f_n(X^n))] = 0 \tag{1.4}$$

for all $n = 1, 2, \ldots$. A pair of the encoder and the decoder $(f_n, g_n)$ is said to be a *code*.

Table 1.1: Categorization of a source coding problem

| | derivation of a theoretical fundamental limit | construction and evaluation of an optimal code |
|---|---|---|
| non-asymptotic analysis | | |
| asymptotic analysis | | |

As shown in Table 1.1, one way to categorize various studies of a source coding problem is to divide them into the following viewpoints:

- the derivation of a theoretical fundamental limit or the construction and evaluation of an optimal code,
- the non-asymptotic analysis or the asymptotic analysis.

The detail of each item is described in the following.

### 1.1.1 The derivation of a theoretical fundamental limit or the construction and evaluation of an optimal code

The objective of study for a source coding problem mainly consists of two parts: the derivation of a theoretical fundamental limit, and the construction and evaluation of an optimal code.

The first objective of study is to derive a theoretical limitation under a certain performance criterion. By deriving the theoretical limitation, we can clarify whether there is room for improvement to the current technology. The standard approach to show the theoretical fundamental limit is showing an *achievability result* and a *converse result*. The achievability result shows that an operation is possible to a quantity. The converse result shows that the operation is not possible over the quantity. For example, in the variable-length lossless source coding, one of the major criteria on codeword length is the *mean codeword length* per source symbol. It is defined by the expectation of a codeword length per source symbol, i.e.,

$$\frac{1}{n} \sum_{x^n \in \mathcal{X}^n} p_{X^n}(x^n) \ell(f_n(x^n)), \tag{1.5}$$

where $p_{X^n}(x^n)$ denotes a probability mass function of $X^n$. For a stationary memoryless source, it is shown in [39] that, as $n \to \infty$, the mean codeword length per source symbol is greater than or equal to the *Shannon entropy*

$$H(X) = -\sum_{x \in \mathcal{X}} p_X(x) \log_K p_X(x). \tag{1.6}$$

This is the converse result. Further, it is also shown in [39] that, as $n \to \infty$, there exists a sequence of a code $\{(f_n, g_n)\}_{n=1}^{\infty}$ for which the mean codeword length per source symbol is asymptotically smaller than or equal to the entropy. This is the achievability result. Combining the converse result and the achievability result, it is clarified that

$$\lim_{n \to \infty} \frac{1}{n} \sum_{x^n \in \mathcal{X}^n} p_{X^n}(x^n) \ell(f_n(x^n)) = H(X). \tag{1.7}$$

As we have explained in the previous paragraph, one of the research topics in source coding is to characterize the fundamental limit of a certain criterion (e.g., the mean codeword length per source symbol) by using a quantity defined by a probability distribution of a source (e.g., the Shannon entropy). Not only the mean codeword length but also other criteria have been proposed, and various studies have been done to clarify the theoretical limitation of these criteria. For example, in a certain problem setting, some studies have characterized the fundamental limit by using the *Rényi entropy* [37]

$$H_\alpha(X^n) = \frac{1}{1-\alpha} \log_K \left( \sum_{x^n \in \mathcal{X}^n} (p_{X^n}(x^n))^\alpha \right) \quad \text{for} \quad \alpha \in (0,1) \cup (1, \infty), \tag{1.8}$$

and other studies have characterized the fundamental limit by using the *smooth Rényi entropy* [34], [35]

$$H_\alpha^\gamma(X^n) = \frac{1}{1-\alpha} \log_K \left( \inf_{q \in \mathcal{B}^\gamma(p_{X^n})} \sum_{x^n \in \mathcal{X}^n} (q(x^n))^\alpha \right) \quad \text{for} \quad \gamma \geq 0, \alpha \in (0,1) \cup (1, \infty), \tag{1.9}$$

where $\mathcal{B}^\gamma(p_{X^n})$ is a set of functions $q : \mathcal{X}^n \to [0,1]$ such that $q(x^n) \leq p_{X^n}(x^n)$ for all $x^n \in \mathcal{X}^n$ and $\sum_{x^n \in \mathcal{X}^n} q(x^n) \geq 1 - \gamma$.

The existence of a code achieving the fundamental limit is guaranteed by the achievability result. However, how to construct such a code is not clear. Therefore, the second objective of study is to construct an optimal code to achieve the theoretical fundamental limit and to analyze the performance of the code. For example, the *Shannon code* [39], the *Huffman code* [14], and the *arithmetic code* (e.g., [23], [38]) have been proposed. The mean codeword length per source symbol of these codes approaches the Shannon entropy for a stationary memoryless source as the blocklength

$n \to \infty$. These codes are designed under the assumption that a probability distribution of a source is known. On the other hand, even if a probability distribution of a source is unknown, several studies have proposed the code whose mean codeword length per source symbol approaches the Shannon entropy. Such codes are said to be *universal codes*.

Among universal codes, the Bayes code (e.g., [2], [5], [24]), which is elaborated in Chapters 7 and 8 in this dissertation, is one of the codes whose mean codeword length per source symbol approaches the Shannon entropy at the fastest speed. Roughly speaking, the Bayes code works as follows. Suppose that the probability mass function of a source sequence $X^n$ is represented by $p_{\theta_*^k}(x^n)$, where $\theta_*^k \in \Theta^k \subset \mathbb{R}^k$ is an unknown parameter. When a class of a probability mass function

$$\{p_{\theta^k} : \theta^k \in \Theta^k \subset \mathbb{R}^k\} \tag{1.10}$$

and a prior probability density function of the parameter $w(\theta^k)$ are known, the Bayes code estimates the probability of $x^n$ as

$$\int_{\Theta^k} w(\theta^k) p_{\theta^k}(x^n) d\theta^k, \tag{1.11}$$

and utilizes it as a coding probability of the arithmetic coding. The mean codeword length of the Bayes code has been evaluated up to constant terms for some major sources (see, e.g., [2], [8]).

### 1.1.2 The non-asymptotic analysis or the asymptotic analysis

As we have shown in Section 1.1.1, Shannon [39] has derived the fundamental limit under the setting that the blocklength $n$ goes to infinity, i.e., $n \to \infty$. Such analysis is said to be the *asymptotic analysis*. From the early days of information theory, various studies on the asymptotic analysis have been done.

However, the blocklength $n$ is finite in the actual use of digital devices. Thus, the *non-asymptotic analysis*, which deals with the case where the blocklength $n$ is finite, has attracted attention recently. This is also said to be the *finite blocklength analysis*.

## 1.2 Purpose of the dissertation

This dissertation deals with the *overflow probability* as a performance criterion. It is defined as the probability of a codeword length per source symbol exceeding a threshold $R \geq 0$, i.e., the overflow probability is defined by

$$\mathbb{P}\left[\frac{1}{n}\ell(f_n(X^n)) > R\right]. \tag{1.12}$$

The main purposes of this dissertation are summarized as the following (P1) and (P2) (see also Tables 1.2 and 1.3):

**(P1)** The first research purpose is to derive the non-asymptotic theoretical fundamental limits on the overflow probability for several source coding problems. As by-products, we derive the asymptotic theoretical fundamental limits based on the non-asymptotic results.

**(P2)** The second research purpose is to evaluate the asymptotic performance of the Bayes code from the viewpoint of the overflow probability.

Table 1.2: The first research purpose in the dissertation

| | derivation of a theoretical fundamental limit | construction and evaluation of an optimal code |
|---|---|---|
| non-asymptotic analysis | (P1) | |
| asymptotic analysis | (P1) | |

Table 1.3: The second research purpose in the dissertation

| | derivation of a theoretical fundamental limit | construction and evaluation of an optimal code |
|---|---|---|
| non-asymptotic analysis | | |
| asymptotic analysis | | (P2) |

The motivation of these research purposes are as follows.

### Research motivation for (P1)

For a non-asymptotic analysis, it is important to focus on the distribution of a codeword length. This is because codeword lengths vary from the mean codeword length in the finite blocklength setting. The overflow probability represents the tail probability of the distribution of a codeword length. Thus, the overflow probability is one way to capture the distribution of the codeword length.

The overflow probability has close relationship to the *fixed-length source coding* in which a codeword length is fixed. For example, studies such as [18] and [29] have stated the equivalence between the error probability in fixed-length source coding and the overflow probability in variable-length lossless source coding. Roughly speaking, this is explained as follows. To construct the optimal fixed-length code (fixed-length code that has the smallest error probability), we set the encoder as follows:

- For most likely source symbols (for example, $2^k - 1$), the encoder assigns a unique codeword of length $k$ bits, and

- for the rest of the source symbols, the encoder assigns the remaining codeword of length $k$ bits. (In this case, a decoding error occurs.)

On the other hand, to construct the optimal variable-length lossless code (variable-length lossless code that has the smallest overflow probability), we set the encoder as follows:

- For most likely source symbols (for example, $2^k - 1$), the encoder assigns a unique codeword whose length does not exceed $k$ bits,

and

- for the rest of the source symbols, the encoder assigns a codeword whose length satisfies the Kraft inequality $\sum_{x^n \in \mathcal{X}^n} 2^{-\ell(f_n(x^n))} \leq 1$. (In this case, an overflow occurs.)

From the above construction, it can be shown that the error probability of the optimal fixed-length code and the overflow probability of the optimal variable-length lossless code are equivalent.

The preceding paragraph has explained the relationship between fixed-length source coding and variable-length lossless source coding. These two regimes are essentially the same. The difference is that fixed-length source coding treats least likely source symbols as *decoding error* while variable-length lossless source coding treats them as *overflow*. Thus, some questions to ask are as follows.

- In variable-length source coding *allowing errors* (variable-length source coding that allows the decoding error probability), what result can we obtain? Are there any relationship between the overflow probability and the error probability of variable-length source coding?

- The variable-length *lossy* source coding under the *excess distortion probability* can be viewed as the generalization of variable-length source coding allowing errors. In this problem setting, what result can we obtain? Are there any relationship between the overflow probability and the excess distortion probability?

- Can we characterize the fundamental limit of the preceding problems (variable-length lossless source coding, variable-length source coding allowing errors, and variable-length lossy source coding) in some unified manner?

Motivated by these questions, the research purpose (P1) aims to give some insights to these questions.

**Research motivation for (P2)**

The major approach to the asymptotic analysis of the mean codeword length is based on the law of large numbers because the mean codeword length is the mean of the distribution of a codeword length. On the other hand, the major approach to the asymptotic analysis of the overflow probability is based on the central limit theorem (CLT) or the Berry-Esséen inequality because the overflow probability is the tail probability of the distribution of a codeword length. Thus, the method to analyze the mean codeword length and the method to analyze the overflow probability differ. This fact indicates the following: if some code has been evaluated based on only the mean codeword length, a novel insight on the code can be obtained from the evaluation based on the overflow probability.

As we have stated in Section 1.1.1, the mean codeword length of the Bayes code has been analyzed by various previous studies. However, the performance of this code on the overflow probability is not known. Thus, the research purpose (P2) aims to obtain a new insight on the Bayes code from the analysis based on the overflow probability.

## 1.3 Organization of the dissertation

This dissertation consists of nine chapters.

As preliminaries, Chapter 2 introduces basic notations used throughout this dissertation. Then, some typical sources and mathematical models of a source coding problem are described. Finally, the major performance criteria on codeword length are presented.

Table 1.4: Chapters 3, 4, 5, and 6 in the dissertation

| | derivation of a theoretical fundamental limit | construction and evaluation of an optimal code |
|---|---|---|
| non-asymptotic analysis | Chapters 3, 4, 5, 6 | |
| asymptotic analysis | Chapters 3, 4, 5, 6 | |

As shown in Table 1.4, Chapters 3, 4, 5, and 6 correspond to the research purpose (P1). These chapters derive the fundamental theoretical limits of

- variable-length lossless source coding (in Chapter 3),
- variable-length source coding allowing errors (in Chapter 4),
- variable-length lossy source coding (in Chapter 5),
- fixed-length Slepian-Wolf source coding (in Chapter 6).

Although some previous studies have derived the fundamental limits on the overflow probability for these source coding problems, they are characterized by various methods and various quantities in each problem. In contrast, Chapters 3–6 of this dissertation characterize all the fundamental limits by using the *smooth max entropy* or related quantity. This is one of the features of this dissertation. The smooth max entropy $H^{\gamma}(X^n)$ is defined by the limit $\alpha \downarrow 0$ of the smooth Rényi entropy $H_{\alpha}^{\gamma}(X^n)$. It is shown in [43] that the smooth max entropy can be written as

$$H^{\gamma}(X^n) = \min_{\substack{\mathcal{Z}^n \subset \mathcal{X}^n: \\ \mathbb{P}[X^n \in \mathcal{Z}^n] \geq 1-\gamma}} \log_K |\mathcal{Z}^n|. \tag{1.13}$$

Chapter 3 considers the problem of variable-length lossless source coding. We derive the non-asymptotic coding theorems by using the smooth max entropy. We use the explicit code construction technique, which is used throughout Chapters 3–5, to show the achievability results. Further, we show the asymptotic coding theorems, which are easily derived from the non-asymptotic results. This chapter contains fundamental and basic ideas used in Chapters 4 and 5 and is positioned as a starting point of Chapters 4 and 5.

Chapter 4 deals with the problem of variable-length source coding allowing errors. Based on the smooth max entropy, the non-asymptotic coding theorems are derived for both stochastic codes and deterministic codes. The main results clarify the difference between the stochastic codes and the deterministic codes. Further, they also clarify the relationship between the overflow probability and the error probability. Moreover, the asymptotic coding theorems are obtained based on the non-asymptotic fundamental limits.

Chapter 5 treats the problem of variable-length lossy source coding. We first define the smooth max entropy-based quantity. Then, using this quantity, novel non-asymptotic coding theorems are obtained for both stochastic codes and deterministic codes. The main results clarify the difference between the stochastic codes and the deterministic codes. Further, they also show the relationship between the overflow probability and the excess distortion probability. Finally, we show asymptotic coding theorems based on the non-asymptotic results.

As stated in Section 1.2, variable-length source coding under the criterion of the overflow probability has close relationship to fixed-length source coding. In Chapter 6, we consider the fixed-length Slepian-Wolf coding problem, which is one of the major problems in fixed-length source coding. This problem deals with the case where one decoder jointly decodes two codewords encoded by separate encoders for two correlated sources. For this problem, we give another characterization of the second-order achievable rate region by using the quantity related to the smooth max entropy and the conditional smooth max entropy.

As shown in Table 1.5, Chapters 7 and 8 correspond to the research purpose (P2). These chapters evaluate the performance of the Bayes code under the overflow probability.

Table 1.5: Chapters 7 and 8 in the dissertation

| | derivation of a theoretical fundamental limit | construction and evaluation of an optimal code |
|---|---|---|
| non-asymptotic analysis | | |
| asymptotic analysis | | Chapters 7, 8 |

Chapter 7 deals with the case where a positive overflow probability is allowed. This chapter analyzes the minimum threshold of the overflow probability of the Bayes code. The upper and lower bounds on it are obtained. To prove the main results, we use the asymptotic evaluation of the codeword length of the Bayes code and the Berry-Esséen bound. This result clarifies one of the advantages of the Bayes codes under the overflow probability.

Chapter 8 considers the case where the overflow probability of the Bayes code vanishes asymptotically. First, this chapter derives the necessary and sufficient condition of the overflow probability of the Bayes code vanishing asymptotically. To show this result, the asymptotic normality of the codeword length of the Bayes code plays a crucial role. Next, this chapter analyzes the behavior of the overflow probability of the Bayes code for the *moderate deviation regime* in which the overflow probability approaches zero and the threshold approaches the theoretical limit at the same time. The obtained result also clarifies one of the advantages of the Bayes codes under the overflow probability.

Chapter 9 is devoted to the conclusion of the dissertation. In this chapter, concluding remarks and future works are presented.

# Chapter 2

# Preliminaries

## 2.1 Introduction

This chapter defines basic notations, several sources, mathematical models of a source coding problem, and major criteria of a source coding problem. Section 2. 2 introduces basic notation. Section 2. 3 defines some sources that are treated in this dissertation. Section 2. 4 presents mathematical models of

- variable-length lossless source coding,
- variable-length source coding allowing errors,
- variable-length lossy source coding,
- fixed-length source coding.

Section 2. 5 describes major performance criteria of the above source coding problems.

## 2.2 Notation

This section presents basic notations that are used throughout this dissertation.

A *source* is an indexed sequence of random variable[1] $\{X_i\}_{i=1}^{\infty}$, where $X_i$ is a random variable taking a value in a set $\mathcal{X}$. The set $\mathcal{X}$ is said to be a *source alphabet*, and we assume that $\mathcal{X}$ is a finite set unless otherwise noted. A random variable is denoted by an upper-case letter and a realization of a random variable is denoted by a lower-case letter. For example, $X_1$ is a random variable taking a value on $\mathcal{X}$ and $x_1 \in \mathcal{X}$ is a realization of $X_1$. Let $\mathcal{X}^n$ be the $n$-th Cartesian product of $\mathcal{X}$. A length $n$ sequence $(X_1, \ldots, X_n) \in \mathcal{X}^n$ is denoted by $X^n$ and a realization of $X^n$ is denoted by $x^n$.

[1]A random variable is defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where $\Omega$ is a sample space, $\mathcal{F}$ is a $\sigma$-field on $\Omega$, and $\mathbb{P}$ is a probability measure on $(\Omega, \mathcal{F})$. In this dissertation, to simplify the presentation, we do not explicitly state the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ on which a random variable is defined.

Let a probability distribution of $X$ be $P_X$. A probability mass function is denoted by $p_X(x)$. That is, $p_X(x) := P_X(\{x\})$ for $x \in \mathcal{X}$. Further, a joint probability distribution of $X^n$ is denoted by $P_{X^n}$ and a joint probability mass function is denoted by $p_{X^n}(x^n)$.

Throughout the dissertation, log denotes a logarithm of base 2. Further, ln denotes a logarithm of base $e$, i.e., ln is a natural logarithm.

## 2.3 Definition of sources

### 2.3.1 Stationary ergodic source

A *stationary ergodic source* is often assumed in information theory. Before describing the definition of a stationary ergodic source, we first define a *stationary source*.

**Definition 2.3.1** *A source $\{X_i\}_{i=1}^{\infty}$ is said to be stationary source if*

$$\begin{aligned} &\mathbb{P}[X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n] \\ &= \mathbb{P}[X_{l+1} = x_1, X_{l+2} = x_2, \ldots, X_{l+n} = x_n] \end{aligned} \tag{2.1}$$

*for any $n, l \in \mathbb{N}$ and $x_1, x_2, \ldots, x_n \in \mathcal{X}$, where $\mathbb{N}$ denotes a set of natural numbers.*

A stationary ergodic source is defined as follows.

**Definition 2.3.2** *A source $\{X_i\}_{i=1}^{\infty}$ is said to be stationary ergodic source if $\{X_i\}_{i=1}^{\infty}$ is a stationary source and it satisfies, with probability one,*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(X_{i+1}, X_{i+2}, \ldots, X_{i+k}) = E[f(X_1, X_2, \ldots, X_k)] \tag{2.2}$$

*for any $k \in \mathbb{N}$ and any integrable function $f$ on $\mathcal{X}^k$, where the expecctation in the right hand side is taken with respect to $P_{X^k}$.*

### 2.3.2 Stationary ergodic Markov source

A source defined by a sequence of random variables with dependence is the fundamental and important source in information theory. The typical example is a *stationary ergodic Markov source*, which is defined in this subsection.

First, a $k$-th order Markov chain is defined as follows.

**Definition 2.3.3** *Let $k \in \mathbb{N}$. A sequence of random variable $\{X_i\}_{i=1}^{\infty}$ is said to be a $k$-th order Markov chain if it satisfies*

$$\begin{aligned} &\mathbb{P}[X_n = x_n \mid X_{n-1} = x_{n-1}, X_{n-2} = x_{n-2}, \ldots, X_1 = x_1] \\ &= \mathbb{P}[X_n = x_n \mid X_{n-1} = x_{n-1}, X_{n-2} = x_{n-2}, \ldots, X_{n-k} = x_{n-k}] \end{aligned} \tag{2.3}$$

*for any $n > k$. Regarding a $k$-th order Markov chain, $\mathcal{X}^k$ is said to be a state space and an element of a state space $\mathcal{X}^k$ is said to be a state.*

Next, we define a stationary ergodic $k$-th order Markov source.

**Definition 2.3.4** *A source $\{X_i\}_{i=1}^{\infty}$ is said to be a stationary ergodic $k$-th order Markov source if $\{X_i\}_{i=1}^{\infty}$ is a stationary ergodic and $k$-th order Markov chain. If $k$ is finite, we call a stationary ergodic $k$-th order Markov source as a stationary ergodic finite order Markov source.*

### 2.3.3 Stationary memoryless source

This section introduces a source described by a single probability distribution $P_X$. It is said to be a *stationary memoryless source.*

**Definition 2.3.5** *A source $\{X_i\}_{i=1}^{\infty}$ is said to be a stationary memoryless source if*

$$p_{X^n}(x^n) = \prod_{i=1}^{n} p_X(x_i) \tag{2.4}$$

*for any $x^n = x_1x_2\dots x_n \in \mathcal{X}^n$.*

### 2.3.4 General source

So far, we have defined several sources by specifying the probabilistic structure. However, information theory also deals with a source allowing *arbitrary* probabilistic structure. Such a source is said to be a *general source.* Thus, a general source need not to satisfy a consistency condition:

$$p_{X^n}(x^n) = \sum_{x\in\mathcal{X}} p_{X^{n+1}}(x^n x). \tag{2.5}$$

The general source is very general, and it includes foregoing stationary ergodic source, stationary ergodic Markov source, and stationary memoryless source. A general source is denoted as $\mathbf{X} = \{X^n\}_{n=1}^{\infty}$ in this dissertation.

**Example 2.3.1** *We show one example for a general source, which does not satisfy consistency condition (2.5). Let a source alphabet be $\mathcal{X} = \{0, 1\}$. Furthermore, let $\theta \in (0, 1/2) \cup (1/2, 1)$ and define $t_n$ by $t_n = x_1 + x_2 + \ldots + x_n$ for a sequence $x^n = x_1x_2\dots x_n$. Then, we consider the source $\{X_i\}_{i=1}^{\infty}$ that has a probability mass function defined by*

$$p_{X^n}(x^n) = \begin{cases} 1/2^n & (n \text{ is odd}), \\ \theta^{t_n}(1-\theta)^{n-t_n} & (n \text{ is even}). \end{cases} \tag{2.6}$$

*That is, $\{X_i\}_{i=1}^{\infty}$ is distributed according to the uniform distribution if $n$ is odd and the Bernoulli distribution if $n$ is even.*

## 2.4 Mathematical models of a source coding problem

### 2.4.1 Fixed-length source coding

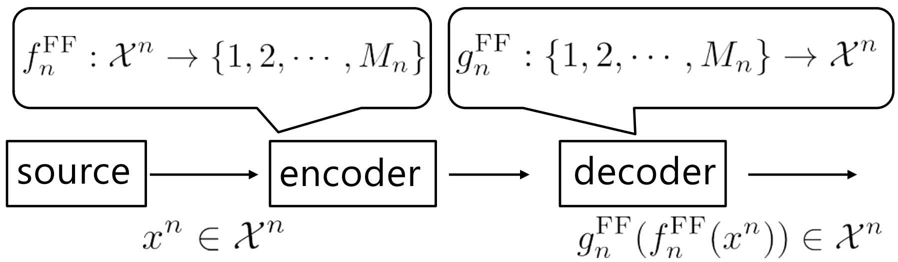


Figure 2.1: Fixed-length source coding

This section explains *fixed-length source coding*[2] depicted in Fig. 2.1.

A *fixed-length source code* is a pair of $(f_n^{\mathrm{FF}}, g_n^{\mathrm{FF}})$ defined as follows. An *encoder* $f_n^{\mathrm{FF}}$ is defined by

$$f_n^{\mathrm{FF}} : \mathcal{X}^n \to \{1, 2, \cdots, M_n\}, \tag{2.7}$$

where $M_n$ is a positive integer. A *decoder* is defied by

$$g_n^{\mathrm{FF}} : \{1, 2, \cdots, M_n\} \to \mathcal{X}^n. \tag{2.8}$$

A fixed-length source coding allows a decoding error probability. An analysis is usually carried out under the condition that the decoding error probability does not exceed $\epsilon \in [0, 1)$ for all $n = 1, 2, \ldots$, i.e.,

$$\mathbb{P}[X^n \neq g_n^{\mathrm{FF}}(f_n^{\mathrm{FF}}(X^n))] \leq \epsilon \tag{2.9}$$

or under the condition that the decoding error probability does not exceed $\epsilon \in [0, 1)$ asymptotically, i.e.,

$$\limsup_{n\to\infty} \mathbb{P}[X^n \neq g_n^{\mathrm{FF}}(f_n^{\mathrm{FF}}(X^n))] \leq \epsilon. \tag{2.10}$$

**Remark 2.4.1** *Since a codeword $i \in \{1, 2, \cdots, M_n\}$ is an integer, it is transformed into a sequence when it is transmitted to a receiver. For example, to transform $i \in \{1, 2, \cdots, M_n\}$ into a binary sequence, we transform it into a binary sequence with length $\lceil \log M_n \rceil$, where $\lceil a \rceil$ denotes the smallest integer that is larger than or equal to $a$. Thus, the codeword length is fixed. This is the reason why the above setup is called the fixed-length source coding.*

[2] It is also said to be almost-lossless fixed-to-fixed compression.

### 2.4.2 Variable-length lossless source coding

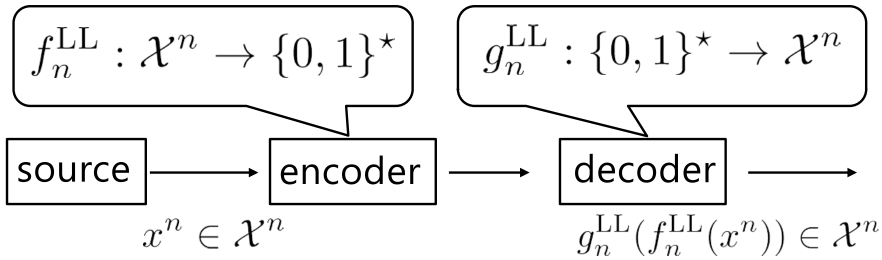


Figure 2.2: Variable-length lossless source coding

In Section 2.4.1, we have described fixed-length source coding, where the codeword length is fixed. However, we can consider the setup for which the codeword length is variable. In Sections 2.4.2, 2.4.3, and 2.4.4, we describe such a setup. First, this section explains *variable-length lossless source coding*[3] depicted in Fig. 2.2.

A *variable-length lossless source code* is a pair of $(f_n^{\mathrm{LL}}, g_n^{\mathrm{LL}})$ defined as follows. An *encoder* $f_n^{\mathrm{LL}} : \mathcal{X}^n \to \{0,1\}^\star$ is an injective function, where $\{0,1\}^\star$ denotes the set of all binary strings and the empty string $\Lambda$, i.e.,

$$\{0,1\}^\star := \{\Lambda, 0, 1, 00, 01, 10, 11, 000, 001, \cdots\}. \tag{2.11}$$

Further, $f_n^{\mathrm{LL}}(x^n) \in \{0,1\}^\star$ is said to be a *codeword* and its length is said to be a *codeword length*. A *decoder* $g_n^{\mathrm{LL}} : \{0,1\}^\star \to \mathcal{X}^n$ is a function such that

$$g_n^{\mathrm{LL}}(f_n^{\mathrm{LL}}(x^n)) = x^n \tag{2.12}$$

for all $x^n \in \mathcal{X}^n$.

The above definition ensures that

$$\mathbb{P}[X^n \neq g_n^{\mathrm{LL}}(f_n^{\mathrm{LL}}(X^n))] = 0 \tag{2.13}$$

for any $n = 1, 2, \ldots$.

### 2.4.3 Variable-length source coding allowing errors

In Section 2.4.2, we have defined variable-length lossless source coding in which a decoding error probability is zero, i.e., (2.13) holds. In this section, we describe a source coding problem in which a decoding error probability is allowed.

As depicted in Fig. 2.3, a *variable-length source code allowing errors* is a pair of $(f_n^{\mathrm{AE}}, g_n^{\mathrm{AE}})$ defined as follows. An *encoder* $f_n^{\mathrm{AE}}$ is defined by

$$f_n^{\mathrm{AE}} : \mathcal{X}^n \to \{0,1\}^\star, \tag{2.14}$$

[3] It is also said to be lossless fixed-to-variable compression.

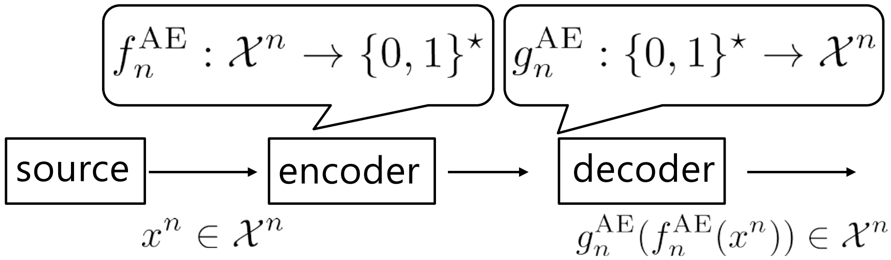


Figure 2.3: Variable-length source coding allowing errors

where the encoder $f_n^{\mathrm{AE}}$ needs not to be an injective mapping. That is, the following situation is allowed:

$$x^n \neq (x^n)' \Rightarrow f_n^{\mathrm{AE}}(x^n) = f_n^{\mathrm{AE}}((x^n)'). \tag{2.15}$$

A *decoder* is defied by

$$g_n^{\mathrm{AE}} : \{0,1\}^\star \to \mathcal{X}^n. \tag{2.16}$$

Since we do not assume that an encoder $f_n^{\mathrm{AE}}$ is injective, a decoding error can be occurred. Therefore, in some cases, analysis is carried out under the condition that a decoding error probability does not exceed $\epsilon \in [0,1)$ for all $n = 1, 2, \ldots$, i.e.,

$$\mathbb{P}[X^n \neq g_n^{\mathrm{AE}}(f_n^{\mathrm{AE}}(X^n))] \leq \epsilon, \tag{2.17}$$

or under the condition that a decoding error probability does not exceed $\epsilon \in [0,1)$ asymptotically, i.e.,

$$\limsup_{n\to\infty} \mathbb{P}[X^n \neq g_n^{\mathrm{AE}}(f_n^{\mathrm{AE}}(X^n))] \leq \epsilon. \tag{2.18}$$

### 2.4.4 Variable-length lossy source coding

In Section 2.4.3, we have stated variable-length source code allowing errors. In this case, we consider the event that a source sequence equals a decoded sequence or not. One way to generalize this setting is to introduce a function that measures the difference between a source sequence and a decoded sequence. This kind of source coding problem is said to be a *variable-length lossy source coding* problem. In this section, we describe a variable-length lossy source coding depicted in Fig. 2.4.

A *variable-length lossy source code* is a pair of $(f_n^{\mathrm{LS}}, g_n^{\mathrm{LS}})$ defined as follows. An *encoder* $f_n^{\mathrm{LS}}$ is defined by

$$f_n^{\mathrm{LS}} : \mathcal{X}^n \to \{0,1\}^\star, \tag{2.19}$$

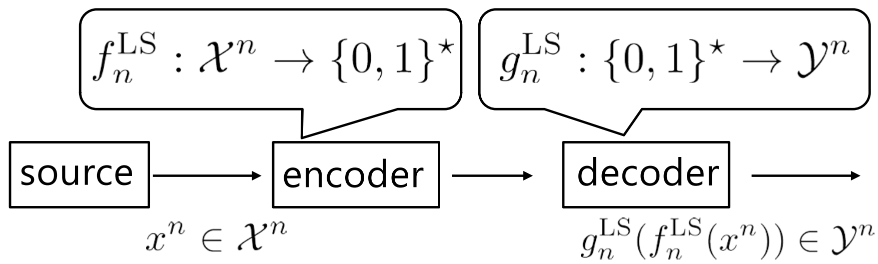


Figure 2.4: Variable-length lossy source coding

where the encoder $f_n^{\mathrm{LS}}$ needs not to be an injective mapping. A *decoder* is defied by

$$g_n^{\mathrm{LS}} : \{0,1\}^\star \to \mathcal{Y}^n, \tag{2.20}$$

where a set $\mathcal{Y}$ is said to be a *reproduction alphabet*. This dissertation assumes that $\mathcal{X}$ and $\mathcal{Y}$ are both finite sets unless otherwise noted.

A function $d_n : \mathcal{X}^n \times \mathcal{Y}^n \to [0, +\infty)$ measures the difference between an original source sequence $x^n \in \mathcal{X}^n$ and the decoded sequence $y^n \in \mathcal{Y}^n$. This function is called a *distortion* measure.

There are two major criteria on distortion measure. The one is the *average distortion* and the other is the *excess distortion probability*.

The average distortion is defined as

$$E\left[d_n(X^n, g_n^{\mathrm{LS}}(f_n^{\mathrm{LS}}(X^n)))\right] \tag{2.21}$$

and in some cases, analysis is carried out under the condition that the average distortion does not exceed $\epsilon \in [0,1)$ for all $n = 1, 2, \ldots$, i.e.,

$$E\left[d_n(X^n, g_n^{\mathrm{LS}}(f_n^{\mathrm{LS}}(X^n)))\right] \leq \epsilon \tag{2.22}$$

or under the condition that the average distortion does not exceed $\epsilon \in [0,1)$ asymptotically, i.e.,

$$\limsup_{n\to\infty} E\left[d_n(X^n, g_n^{\mathrm{LS}}(f_n^{\mathrm{LS}}(X^n)))\right] \leq \epsilon. \tag{2.23}$$

On the other hand, the excess distortion probability is defined as

$$\mathbb{P}\left[d_n(X^n, g_n^{\mathrm{LS}}(f_n^{\mathrm{LS}}(X^n))) > D\right] \tag{2.24}$$

for $D \geq 0$, and in some cases, analysis is carried out under the condition that the excess distortion probability does not exceed $\epsilon \in [0,1)$ for all $n = 1, 2, \ldots$, i.e.,

$$\mathbb{P}\left[d_n(X^n, g_n^{\mathrm{LS}}(f_n^{\mathrm{LS}}(X^n))) > D\right] \leq \epsilon \tag{2.25}$$

or under the condition that the excess distortion probability does not exceed $\epsilon \in [0,1)$ asymptotically, i.e.,

$$\limsup_{n\to\infty} \mathbb{P}\left[d_n(X^n, g_n^{\mathrm{LS}}(f_n^{\mathrm{LS}}(X^n))) > D\right] \leq \epsilon. \tag{2.26}$$

**Remark 2.4.2** *The error probability, which is considered in Section 2.4.3, can be seen as a special case of the excess distortion probability. To see this, we consider the excess distortion probability*

$$\mathbb{P}[d_n(X^n, Y^n) > D]. \tag{2.27}$$

*In (2.27), let* $X^n$ *and* $Y^n$ *be random variables taking values in* $\mathcal{X}^n$*, set* $D = 0$*, and let the distortion measure be*

$$d_n(x^n, y^n) = I\{x^n \neq y^n\}, \tag{2.28}$$

*for* $x^n, y^n \in \mathcal{X}^n$*, where* $I\{\cdot\}$ *denotes an indicator function. Then, it is easily verified that (2.27) reduces to the error probability*

$$\mathbb{P}[X^n \neq Y^n]. \tag{2.29}$$

## 2.5 Performance criteria on codeword length

The typical criterion on codeword length in fixed-length source coding defined in Section 2.4.1 is the *coding rate* defined by

$$\frac{\log M_n}{n}. \tag{2.30}$$

On the other hand, there are two major performance criteria on codeword length for variable-length coding defined in Sections 2.4.2, 2.4.3, and 2.4.4. They are the *mean codeword length* and the *overflow probability*. For a variable-length code $(f_n, g_n)$, the mean codeword length is defined by

$$E\left[\frac{1}{n}\ell(f_n(X^n))\right], \tag{2.31}$$

where $\ell(f_n(x^n))$ denotes a codeword length of $x^n \in \mathcal{X}^n$; the overflow probability is defined by

$$\mathbb{P}\left[\frac{1}{n}\ell(f_n(X^n)) > R\right] \tag{2.32}$$

for $R \geq 0$. The constant $R$ is said to be a *threshold* of the overflow probability or a *rate* of the overflow probability.

# Chapter 3

# Non-asymptotic and asymptotic analyses of variable-length lossless source coding

## 3.1 Introduction

This chapter deals with the problem of variable-length lossless source coding for a general source. For this problem, studies such as [18] and [30] have derived the fundamental limit on the minimum threshold of the overflow probability. They have characterized the fundamental limit by focusing on the distribution of the self-information $\frac{1}{n}\log\frac{1}{p_{X^n}(X^n)}$.

This chapter considers the derivation of the fundamental limit on the minimum threshold of the overflow probability as in [18] and [30]. However, this study focuses on the distribution of source symbol $X$ ordered in decreasing probability. As we will discuss in Section 3.8, this viewpoint is compatible with the definition of the smooth max entropy, and we derive the non-asymptotic characterizations on the minimum threshold of the overflow probability by using the smooth max entropy. To show the achievability results, we use the explicit code construction instead of using the random coding argument. This technique is an important and basic one that is used throughout Chapters 3–5. Moreover, the proof of our achievability result clarifies

- the similarity of variable-length lossless code under the overflow probability and fixed-length code,

and

- the difference between a prefix code and a non-prefix code.

Further, using the results obtained in the non-asymptotic regime, we establish the asymptotic coding theorems. It is worth noticing that this chapter is positioned as a starting point of Chapters 4 and 5.

The organization of this chapter is as follows. Section 3.2 sets up the problem formulation. Section 3.3 introduces the smooth max entropy, which plays a significant role in the main results. Section 3.4 describes the prior works. Non-asymptotic coding theorems for prefix codes and non-prefix codes are derived in Section 3.5 and Section 3.6, respectively. Using these theorems, we show asymptotic coding theorem in Section 3.7. Finally, Section 3.8 discusses the main results and concludes this chapter.

## 3.2 Problem formulation

Let $\mathcal{X}$ be a source alphabet, which is a finite set. Let $X$ be a random variable taking a value in $\mathcal{X}$ and $x$ be a realization of $X$. The probability distribution of $X$ is denoted as $P_X$ and the probability mass function of $X$ is denoted as $p_X(x)$.

This chapter analyzes the variable-length lossless prefix codes and non-prefix codes defined as follows. First, an encoder of a prefix code $f^{\mathrm{p}} : \mathcal{X} \to \{0,1\}^\star$ is defined as an injective function satisfying Kraft's inequality

$$\sum_{x \in \mathcal{X}} 2^{-\ell(f^{\mathrm{p}}(x))} \leq 1, \tag{3.1}$$

where $\ell(f^{\mathrm{p}}(x))$ denotes the codeword length of codeword $f^{\mathrm{p}}(x)$ for $x \in \mathcal{X}$. A decoder $g^{\mathrm{p}} : \{0,1\}^\star \to \mathcal{X}$ is a function such that

$$g^{\mathrm{p}}(f^{\mathrm{p}}(x)) = x \tag{3.2}$$

for all $x \in \mathcal{X}$. Next, an encoder of a non-prefix code $f$ is defined as an injective function such that $f : \mathcal{X} \to \{0,1\}^\star$. A decoder $g : \{0,1\}^\star \to \mathcal{X}$ is a function such that

$$g(f(x)) = x \tag{3.3}$$

for all $x \in \mathcal{X}$. For example, in [18] and [41], removing the prefix condition for variable-length lossless source coding is discussed.

Using the overflow probability, we define an $(R,\delta)_{\mathrm{p}}$ code and $(R,\delta)$ code.

**Definition 3.2.1** *Given $R \geq 0$ and $\delta \in [0,1)$, an $(R,\delta)_{\mathrm{p}}$ code is a prefix code satisfying*

$$\mathbb{P}[\ell(f^{\mathrm{p}}(X)) > R] \leq \delta. \tag{3.4}$$

*Further, an $(R,\delta)$ code is a non-prefix code satisfying*

$$\mathbb{P}[\ell(f(X)) > R] \leq \delta. \tag{3.5}$$

The purpose of this chapter is to analyze the infimum of the threshold of the overflow probability:

$$R^*_{\mathrm{p}}(\delta) := \inf\{R : \ \exists \text{ an } (R,\delta)_{\mathrm{p}} \text{ code}\}, \tag{3.6}$$

$$R^*(\delta) := \inf\{R : \ \exists \text{ an } (R,\delta) \text{ code}\}. \tag{3.7}$$

We consider the following problem formulation in the asymptotic analysis. Let $\mathcal{X}^n$ be the $n$-th Cartesian product of $\mathcal{X}$. Let $X^n$ be a random variable taking a value in $\mathcal{X}^n$ and $x^n$ be a realization of $X^n$. Furthermore, let $\mathbf{X} = \{X^n\}_{n=1}^{\infty}$ denote a general source. The joint probability distribution of $X^n$ is denoted as $P_{X^n}$ and the joint probability mass function of $X^n$ is denoted as $p_{X^n}(x^n)$. An encoder of a prefix code $f_n^{\mathrm{p}} : \mathcal{X}^n \to \{0,1\}^\star$ is defined as an injective function satisfying Kraft's inequality

$$\sum_{x^n \in \mathcal{X}^n} 2^{-\ell(f^{\mathrm{p}}(x^n))} \leq 1. \tag{3.8}$$

A decoder $g_n^{\mathrm{p}} : \{0,1\}^\star \to \mathcal{X}^n$ is a function such that

$$g_n^{\mathrm{p}}(f_n^{\mathrm{p}}(x^n)) = x^n \tag{3.9}$$

for all $x^n \in \mathcal{X}^n$. Next, an encoder of a non-prefix code $f_n$ is defined as an injective function such that $f_n : \mathcal{X}^n \to \{0,1\}^\star$. A decoder $g_n : \{0,1\}^\star \to \mathcal{X}^n$ is a function such that

$$g_n(f_n(x^n)) = x^n \tag{3.10}$$

for all $x^n \in \mathcal{X}^n$.

We define an $(n, R, \delta)_{\mathrm{p}}$ code and an $(n, R, \delta)$ code as follows.

**Definition 3.2.2** *Given $R \geq 0$ and $\delta \in [0,1)$, a prefix code $(f_n^{\mathrm{p}}, g_n^{\mathrm{p}})$ satisfying*

$$\mathbb{P}\left[\frac{1}{n}\ell(f_n^{\mathrm{p}}(X^n)) > R\right] \leq \delta \tag{3.11}$$

*is said to be an $(n, R, \delta)_{\mathrm{p}}$ code. Moreover, a non-prefix code $(f_n, g_n)$ satisfying*

$$\mathbb{P}\left[\frac{1}{n}\ell(f_n(X^n)) > R\right] \leq \delta \tag{3.12}$$

*is said to be an $(n, R, \delta)$ code.*

The asymptotic fundamental limit is the following minimum threshold.

**Definition 3.2.3** *Given $\delta \in [0,1)$, $R_{\mathrm{p}}(\delta|\mathbf{X})$ is the infimum of the threshold $R$ such that there exists an $(n, R, \delta)_{\mathrm{p}}$ code for all $n \geq n_0$ with some $n_0 > 0$. Further, $R(\delta|\mathbf{X})$ is the infimum of the threshold $R$ such that there exists an $(n, R, \delta)$ code for all $n \geq n_0$ with some $n_0 > 0$.*

Moreover, another asymptotic fundamental limit is defined as follows.

**Definition 3.2.4** *Given $\delta \in [0,1)$, $\hat{R}_{\mathrm{p}}(\delta|\mathbf{X})$ is the infimum of the threshold $R$ such that there exists a code $(f_n^{\mathrm{p}}, g_n^{\mathrm{p}})$ satisfying*

$$\limsup_{n\to\infty} \mathbb{P}\left[\frac{1}{n}\ell(f_n^{\mathrm{p}}(X^n)) > R\right] \leq \delta. \tag{3.13}$$

*Further, $\hat{R}(\delta|\mathbf{X})$ is the infimum of the threshold $R$ such that there exists a code $(f_n, g_n)$ satisfying*

$$\limsup_{n\to\infty} \mathbb{P}\left[\frac{1}{n}\ell(f_n(X^n)) > R\right] \leq \delta. \tag{3.14}$$

The main results of this chapter are to analyze the fundamental theoretical limitations by using the smooth max entropy. In the next section, we describe the definition of the smooth max entropy.

## 3.3 Smooth max entropy

The smooth max entropy is defined as follows.

**Definition 3.3.1** *Given $\gamma \in [0,1)$, the smooth max entropy $H^\gamma(X)$ is defined as*

$$H^\gamma(X) = \min_{\substack{\mathcal{Z}\subset\mathcal{X}: \\ \mathbb{P}[X\in\mathcal{Z}]\geq 1-\gamma}} \log|\mathcal{Z}|, \tag{3.15}$$

*where $|\cdot|$ represents the cardinality of the set.*

From the definition, it is easy to see that the smooth max entropy $H^\gamma(X)$ is a monotonically non-increasing function of $\gamma$.

**Remark 3.3.1** *For $\gamma \geq 0$ and $\alpha \in (0,1)\cup(1,\infty)$, the smooth Rényi entropy of order $\alpha$[1] has first introduced by Renner and Wolf [34] and has redefined in [35] as follows:*

$$H_\alpha^\gamma(X^n) = \frac{1}{1-\alpha}\log\left(\inf_{q\in\mathcal{B}^\gamma(P_{X^n})}\sum_{x^n\in\mathcal{X}^n}(q(x^n))^\alpha\right), \tag{3.16}$$

*where $\mathcal{B}^\gamma(P_{X^n})$ is a set of functions $q : \mathcal{X}^n \to [0,1]$ such that*

$$q(x^n) \leq p_{X^n}(x^n) \tag{3.17}$$

*for all $x^n \in \mathcal{X}^n$ and*

$$\sum_{x^n\in\mathcal{X}^n} q(x^n) \geq 1-\gamma. \tag{3.18}$$

[1]Strictly speaking, this should be called $\gamma$-smooth Rényi entropy of order $\alpha$.

*It is known that $H_\alpha^\gamma(X^n)$ is a monotonically non-increasing function of $\alpha \in (0,1) \cup (1,\infty)$. Therefore, $H_\alpha^\gamma(X^n)$ takes the maximum value when $\alpha \downarrow 0$. Based on this fact, the smooth Rényi entropy of order zero, i.e., the quantitiy $\lim_{\alpha\downarrow 0} H_\alpha^\gamma(X^n)$, is said to be the smooth max entropy in [13]. Uyematsu [43] has shown that the smooth max entropy can be defined as the form in Definition 3.3.1.*

*The smooth Rényi entropy is a generalization of the Shannon entropy [39] and Rényi entropy [37]. This quantity is used in cryptography (see, e.g., [34], [35], [36]), source coding (see, e.g., [34], [43], [46], [47], [48]), and random number generation (see, e.g., [44], [45]).*

## 3.4 Related previous works

Kontoyiannis and Verdú [18] have derived the non-asymptotic fundamental limit $R^*(\delta)$ as in the following theorem.

**Theorem 3.4.1 ([18])** *For any $a \geq 0$, define $\delta$ by*

$$\delta = \mathbb{P}\left[\log \frac{1}{p_X(X)} \geq a\right]. \tag{3.19}$$

*Then, $R^*(\delta)$ is given by*

$$R^*(\delta) = \lceil \log(1 + M(2^a)) \rceil - 1, \tag{3.20}$$

*where $M(\beta)$ is defined by*

$$M(\beta) = \mathbb{P}\left[\log \frac{1}{p_X(X)} < \log \beta\right]. \tag{3.21}$$

**Remark 3.4.1** *Theorem 3.4.1 treats the restricted $\delta$ such that (3.19). On the other hand, our study deals with any $\delta \in [0,1)$.*

Nomura et al. [30] have derived the next result on $\hat{R}_{\mathrm{p}}(\delta|\mathbf{X})$.

**Theorem 3.4.2 ([30])** *For any $\delta \in [0,1)$, $\hat{R}_{\mathrm{p}}(\delta|\mathbf{X})$ is given by*

$$\hat{R}_{\mathrm{p}}(\delta|\mathbf{X}) = \inf\left\{R : \limsup_{n\to\infty} \mathbb{P}\left[\frac{1}{n} \log \frac{1}{p_{X^n}(X^n)} \geq R\right] \leq \delta\right\}. \tag{3.22}$$

**Remark 3.4.2** *The previous study [30] has only considered the asymptotic setting. On the other hand, our study considers the non-asymptotic setting as well as the asymptotic setting.*

## 3.5 Non-asymptotic coding theorem for prefix codes

The next lemma shows the achievability result.

**Lemma 3.5.1** *For any $\delta \in [0, 1)$, there exists an $(R, \delta)_{\mathrm{p}}$ code such that*

$$R = \lfloor H^{\delta}(X) + 1 \rfloor. \tag{3.23}$$

(**Proof**) Let $x_i$ be the element of $\mathcal{X}$ that has the $i$-th largest probability. That is, it holds that

$$p_X(x_1) \geq p_X(x_2) \geq p_X(x_3) \geq \cdots. \tag{3.24}$$

Next, let $k^* \geq 1$ be the integer satisfying

$$\sum_{i=1}^{k^*-1} p_X(x_i) < 1 - \delta, \tag{3.25}$$

$$\sum_{i=1}^{k^*} p_X(x_i) \geq 1 - \delta. \tag{3.26}$$

Using $k^*$, we define a set $\mathcal{A} \subset \mathcal{X}$ as

$$\mathcal{A} = \{x_1, x_2, \ldots, x_{k^*}\}. \tag{3.27}$$

From the definition of $\mathcal{A}$, we have

$$\mathbb{P}[X \in \mathcal{A}] \geq 1 - \delta. \tag{3.28}$$

Further, the definition of the smooth max entropy and $\mathcal{A}$ establish the relationship

$$H^{\delta}(X) = \log |\mathcal{A}|, \tag{3.29}$$

which yields

$$|\mathcal{A}| \leq 2^{\lfloor H^{\delta}(X)+1 \rfloor} - 1. \tag{3.30}$$

Now, the encoder and decoder are constructed as follows.

**[Encoder]**

- For the most probable $2^{\lfloor H^{\delta}(X)+1 \rfloor} - 1$ source symbols (i.e., $x_1, \ldots, x_{2^{\lfloor H^{\delta}(X)+1 \rfloor}-1}$), the encoder $f^{\mathrm{p}}$ assigns codeword whose codeword length is $\lfloor H^{\epsilon}(X) + 1 \rfloor$ bits,

and

- for the rest of the source symbols, the encoder $f^{\mathrm{p}}$ assigns codeword whose codeword length satisfies Kraft's inequality (3.1).

**[Decoder]** The decoder is the inverse function of the encoder $f^{\mathrm{p}}$.

Then, from (3.28), (3.30), and the prefix code defined above, it holds that

$$\mathbb{P}[\ell(f^{\mathrm{p}}(X)) > \lfloor H^{\delta}(X) + 1 \rfloor] \leq \mathbb{P}[X \notin \mathcal{A}] \leq \delta. \tag{3.31}$$

This completes the proof. □

The next lemma shows the converse result.

**Lemma 3.5.2** *For any $\delta \in [0, 1)$, any $(R, \delta)_{\mathrm{p}}$ code satisfies*

$$R \geq H^{\delta}(X). \tag{3.32}$$

**(Proof)** Fix an $(R, \delta)_{\mathrm{p}}$ code arbitrarily. Then, it holds that

$$\mathbb{P}[\ell(f^{\mathrm{p}}(X)) > R] \leq \delta \tag{3.33}$$

from the definition of an $(R, \delta)_{\mathrm{p}}$ code. Next, let $\mathcal{S}^{\mathrm{p}}(R)$ be defined as

$$\mathcal{S}^{\mathrm{p}}(R) = \{x \in \mathcal{X} : \ell(f^{\mathrm{p}}(x)) \leq R\}. \tag{3.34}$$

Then, (3.33) is rewritten as

$$\mathbb{P}[X \in \mathcal{S}^{\mathrm{p}}(R)] \geq 1 - \delta. \tag{3.35}$$

Therefore, the definition of the smooth max-entropy and (3.35) establish

$$H^{\delta}(X) \leq \log |\mathcal{S}^{\mathrm{p}}(R)|. \tag{3.36}$$

On the other hand, we have

$$1 \overset{(a)}{\geq} \sum_{x \in \mathcal{X}} 2^{-\ell(f^{\mathrm{p}}(x))} \geq \sum_{x \in \mathcal{S}^{\mathrm{p}}(R)} 2^{-\ell(f^{\mathrm{p}}(x))} \overset{(b)}{\geq} |\mathcal{S}^{\mathrm{p}}(R)| 2^{-R}. \tag{3.37}$$

where $(a)$ follows from (3.1); $(b)$ follows from (3.34). Hence, (3.37) gives

$$\log |\mathcal{S}^{\mathrm{p}}(R)| \leq R. \tag{3.38}$$

Therefore, (3.32) is obtained from (3.36) and (3.38). □

By Lemmas 3.5.1 and 3.5.2, the next result is obtained.

**Theorem 3.5.1** *For any $\delta \in [0, 1)$, it holds that*

$$H^{\delta}(X) \leq R_{\mathrm{p}}^{*}(\delta) \leq \lfloor H^{\delta}(X) + 1 \rfloor. \tag{3.39}$$

**Remark 3.5.1** *The upper and lower bounds in Theorem 3.5.1 are tight in the following sense: (i) There exists a source for which the lower bound holds in equality. (ii) For any $\gamma > 0$, there exists a source for which $R_{\rm p}^*(\delta)$ is larger than $\lfloor H^\epsilon(X) + 1 \rfloor - \gamma$.*

*To verify (i), one example, for which the lower bound holds in equality, is given. Given $\delta \in [0, 1)$ and $R \in \mathbb{N}$ satisfying $\delta < 1/2^R$, suppose that*

$$\mathcal{X} = \{x_1, x_2, \ldots, x_{2^R}\} \tag{3.40}$$

*and*

$$p_X(x_i) = \frac{1}{2^R} \quad (i = 1, 2, \ldots, 2^R). \tag{3.41}$$

*For this source, we shall show that the lower bound in Theorem 3.5.1 holds in equality. For any $(R, \delta)_{\rm p}$ code, Lemma 3.5.2 guarantees that*

$$R \geq H^\delta(X). \tag{3.42}$$

*Since (3.41) and $\delta < 1/2^R$, the definition of the smooth max entropy yields*

$$H^\delta(X) = \log |\mathcal{X}| = R. \tag{3.43}$$

*Therefore, it is proved that the equality in (3.42) holds.*

*To verify (ii), let $\mathcal{A}$ be the set defined as in (3.27), and suppose that the source satisfies $|\mathcal{A}| = 2^K$ for some $K \in \mathbb{N}$ (in this case, $H^\delta(X) = K$ because of (3.29)). For any $\gamma > 0$, we calculate the overflow probability for the threshold $\lfloor H^\delta(X) + 1 \rfloor - \gamma$. For the threshold $\lfloor H^\delta(X) \rfloor + 1 - \gamma$, the code that has the minimum overflow probability is constructed as follows:*

- *for the most probable $2^{H^\delta(X)} - 1$ source symbols, the encoder $f^{\rm p}$ assigns codeword whose codeword length is $H^\delta(X)$ bits,*

*and*

- *for the rest of the source symbols, the encoder $f^{\rm p}$ assigns codeword whose codeword length satisfies Kraft's inequality (3.1).*

*Then, from the construction of the code and the relationship*

$$2^{H^\epsilon(X)} - 1 = 2^{\log |\mathcal{A}|} - 1 = |\mathcal{A}| - 1, \tag{3.44}$$

*it is verified that one element in $\mathcal{A}$ (define this element as $x^*$) has codeword length that is larger than $H^\delta(X)$ bits. Thus, it holds that*

$$\mathbb{P}[\ell(f^{\rm p}(X)) > \lfloor H^\delta(X) + 1 \rfloor - \gamma] = \mathbb{P}[X \notin \mathcal{A}] + \mathbb{P}[X = x^*] > \delta, \tag{3.45}$$

*where the last inequality is due to the definition of $\mathcal{A}$. Because (3.45) holds for the prefix code with the minimum overflow probability for the threshold $\lfloor H^\delta(X) + 1 \rfloor - \gamma$, (3.45) also holds for any prefix code for the threshold $\lfloor H^\delta(X) + 1 \rfloor - \gamma$. Hence, (3.45) and the definition of $R_{\rm p}^*(\delta)$ establish*

$$R_{\rm p}^*(\delta) > \lfloor H^\delta(X) + 1 \rfloor - \gamma. \tag{3.46}$$

*This concludes that there exists a source for which $R_{\rm p}^*(\delta)$ is larger than $\lfloor H^\delta(X) + 1 \rfloor - \gamma$ for any $\gamma > 0$.*

## 3.6 Non-asymptotic coding theorem for non-prefix codes

This section analyzes the fundamental limit on the overflow probability for non-prefix codes. We emphasize that this section contains basic ideas used throughout Chapters 4 and 5 (for details, see the discussion in Section 3.8).

The the next lemma shows the achievability result.

**Lemma 3.6.1** *For any $\delta \in [0,1)$, there exists an $(R,\delta)$ code such that*

$$R = \lfloor H^{\delta}(X) \rfloor. \tag{3.47}$$

**(Proof)**

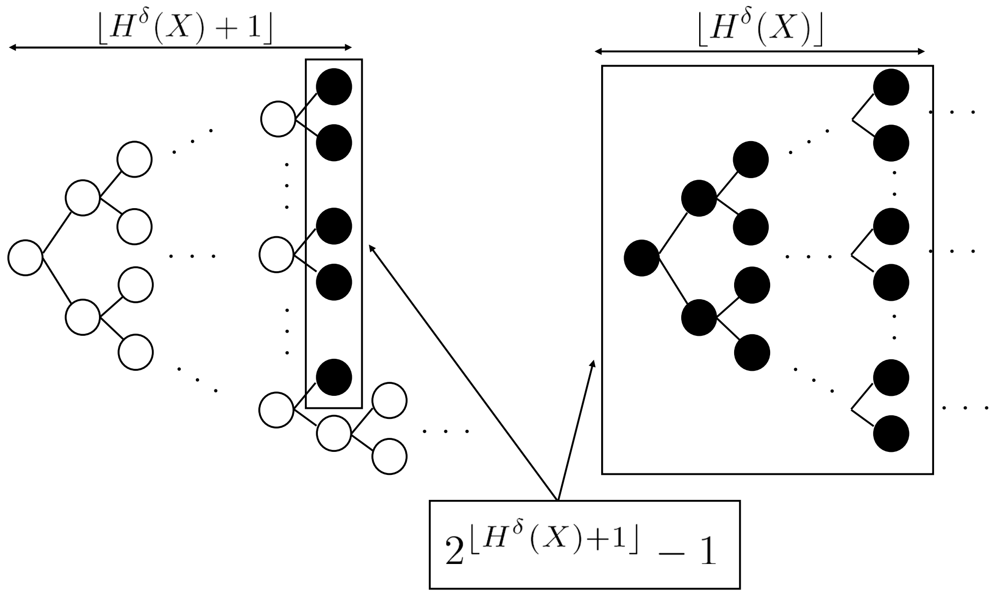


Figure 3.1: Relationship between the interior nodes and leaf nodes of a code tree

First, as shown in Fig. 3.1, the following fact is easily verified:
(♠) *In a code tree, the following quantities are equivalent:*

- *the number of leaf nodes whose depth is $\lfloor H^{\delta}(X)+1 \rfloor$,*
- *the number of interior nodes within the depth of $\lfloor H^{\delta}(X) \rfloor$.*

Next, the encoder $f$ is defined. Let $x_i$ be the element of $\mathcal{X}$ which has the $i$-th largest probability. That is, it holds that $p_X(x_1) \geq p_X(x_2) \geq p_X(x_3) \geq \cdots$. Then, the encoder $f$ maps a source symbol $x$ in the order of decreasing probability to

$$\{0,1\}^{\star} = \{\Lambda, 0, 1, 00, 01, 10, 11, 000, 001, \cdots\}$$

in the lexicographic order, i.e.,

$$f(x_1) = \Lambda, \tag{3.48}$$
$$f(x_2) = 0, \tag{3.49}$$
$$f(x_3) = 1, \tag{3.50}$$
$$f(x_4) = 00, \tag{3.51}$$
$$f(x_5) = 01, \tag{3.52}$$
$$f(x_6) = 10, \tag{3.53}$$
$$f(x_7) = 11, \tag{3.54}$$
$$f(x_8) = 000, \ldots. \tag{3.55}$$

Further, the decoder is defined as the inverse function of the encoder.

Then, from the non-prefix code defined above, the fact (♠), and (3.30), it holds that

$$\mathbb{P}[\ell(f(X)) > \lfloor H^{\delta}(X) \rfloor] \leq \mathbb{P}[X \notin \mathcal{A}] \leq \delta, \tag{3.56}$$

where $\mathcal{A}$ is the set defined as in (3.27). □

The next lemma shows the converse bound.

**Lemma 3.6.2** *For any $\delta \in [0, 1)$, any $(R, \delta)$ code satisfies*

$$R > H^{\delta}(X) - 1. \tag{3.57}$$

(**Proof**) Fix an $(R, \delta)$ code arbitrarily. Then, we have

$$\mathbb{P}[\ell(f(X)) > R] \leq \delta \tag{3.58}$$

from the definition of an $(R, \delta)$ code. Next, let $\mathcal{S}(R)$ be defined as

$$\mathcal{S}(R) = \{x \in \mathcal{X} : \ell(f(x)) \leq R\}. \tag{3.59}$$

Then, (3.58) leads to

$$\mathbb{P}[X \in \mathcal{S}(R)] \geq 1 - \delta. \tag{3.60}$$

Hence, the definition of the smooth max entropy and (3.60) yields

$$H^{\delta}(X) \leq \log |\mathcal{S}(R)|. \tag{3.61}$$

On the other hand, it follows that

$$|\mathcal{S}(R)| \overset{(a)}{\leq} 1 + 2 + 2^2 + \cdots + 2^{\lfloor R \rfloor} = 2^{\lfloor R \rfloor + 1} - 1 < 2^{R+1} \tag{3.62}$$

where $(a)$ is due to the definition of $f$ and $\mathcal{S}(R)$. Thus, (3.62) gives

$$\log |\mathcal{S}(R)| < R + 1 \tag{3.63}$$

Therefore, (3.57) is obtained from (3.61) and (3.63). □

Combination of Lemmas 3.6.1 and 3.6.2 establishes the next result.

**Theorem 3.6.1** *For any $\delta \in [0,1)$, it holds that*

$$H^{\delta}(X) - 1 < R^{*}(\delta) \leq \lfloor H^{\delta}(X) \rfloor. \tag{3.64}$$

**Remark 3.6.1** *The upper bound in Theorem 3.6.1 is tight in the sense that for any $\gamma > 0$, there exists a source for which $R^{*}(\delta)$ is larger than $H^{\delta}(X) - \gamma$. To verify the tightness of the upper bound in Theorem 3.6.1, let $A$ be the set defined as in (3.27), and suppose that the source satisfies $|A| = 2^{K}$ for some $K \in \mathbb{N}$. For this source, the tightness of the upper bound is proved by using the same argument as in Remark 3.5.1.*

*On the other hand, since the lower bound in Theorem 3.6.1 does not hold in equality, it is not possible to discuss the tightness on the lower bound as in Theorem 3.5.1.*

## 3.7 Asymptotic coding theorem

The next theorem characterizes $R_{\mathrm{p}}(\delta|\mathbf{X})$ by the smooth max entropy.

**Theorem 3.7.1** *For any $\delta \in [0,1)$, it holds that*

$$R_{\mathrm{p}}(\delta|\mathbf{X}) = \limsup_{n\to\infty} \frac{1}{n} H^{\delta}(X^{n}). \tag{3.65}$$

(**Proof**) First,

$$R_{\mathrm{p}}(\delta|\mathbf{X}) \leq \limsup_{n\to\infty} \frac{1}{n} H^{\delta}(X^{n}) \tag{3.66}$$

is proved. From Lemma 3.5.1, there exists a code $(f_n^{\mathrm{p}}, g_n^{\mathrm{p}})$ satisfying

$$\mathbb{P}\left[\frac{1}{n}\ell(f_n^{\mathrm{p}}(X^{n})) > \frac{1}{n} H^{\delta}(X^{n}) + \frac{1}{n}\right] \leq \delta. \tag{3.67}$$

Fix $\gamma > 0$ arbitrarily. Then, it holds that

$$\frac{1}{n} H^{\delta}(X^{n}) \leq \limsup_{n\to\infty} \frac{1}{n} H^{\delta}(X^{n}) + \gamma \tag{3.68}$$

and

$$\frac{1}{n} \leq \gamma \tag{3.69}$$

for all $n \geq n_0$ with some $n_0 > 0$. Then, from (3.67), we have

$$\mathbb{P}\left[\frac{1}{n}\ell(f_n^{\mathrm{p}}(X^{n})) > \limsup_{n\to\infty} \frac{1}{n} H^{\delta}(X^{n}) + 2\gamma\right] \leq \delta \tag{3.70}$$

for all $n \geq n_0$. Thus, from (3.70), $(f_n^{\mathrm{p}}, g_n^{\mathrm{p}})$ is an $(n, R, \delta)_{\mathrm{p}}$ code with

$$R = \limsup_{n \to \infty} \frac{1}{n} H^{\delta}(X^n) + 2\gamma \tag{3.71}$$

for all $n \geq n_0$. Since $\gamma > 0$ is arbitrary, this indicates the desired inequality (3.66).

Next,

$$R_{\mathrm{p}}(\delta|\mathbf{X}) \geq \limsup_{n \to \infty} \frac{1}{n} H^{\delta}(X^n) \tag{3.72}$$

is shown. For any $(n, R, \delta)_{\mathrm{p}}$ code, Lemma 3.5.2 gives

$$nR > H^{\delta}(X^n). \tag{3.73}$$

Therefore, it holds that

$$R \geq \limsup_{n \to \infty} \frac{1}{n} H^{\delta}(X^n) \tag{3.74}$$

for any $(n, R, \delta)_{\mathrm{p}}$ code. Hence, we have (3.72). □

The same discussion as that in Theorem 3.7.1 yields the next result on $R(\delta|\mathbf{X})$.

**Theorem 3.7.2** *For any $\delta \in [0, 1)$, it holds that*

$$R(\delta|\mathbf{X}) = \limsup_{n \to \infty} \frac{1}{n} H^{\delta}(X^n). \tag{3.75}$$

Furthermore, by almost the same proof of Theorems 3.7.1 and 3.7.2, we have the next result on $\hat{R}_{\mathrm{p}}(\delta|\mathbf{X})$ and $\hat{R}(\delta|\mathbf{X})$.

**Corollary 3.7.1** *For any $\delta \in [0, 1)$, it holds that*

$$\hat{R}_{\mathrm{p}}(\delta|\mathbf{X}) = \lim_{\tau \downarrow 0} \limsup_{n \to \infty} \frac{1}{n} H^{\delta+\tau}(X^n) \tag{3.76}$$

$$\hat{R}(\delta|\mathbf{X}) = \lim_{\tau \downarrow 0} \limsup_{n \to \infty} \frac{1}{n} H^{\delta+\tau}(X^n). \tag{3.77}$$

**Remark 3.7.1** *The difference between (3.65) and (3.76) comes from the difference of the asymptotic constraint, i.e., (3.65) is the result under the constraint of (3.11) for all $n \geq n_0$ with some $n_0$, while (3.76) is the result under the constraint of (3.13). Similarly, the same remark holds between (3.75) and (3.77).*

**Remark 3.7.2** *The result (3.76) can be proved by combining the results in [29] and [43]. This is because the study [29] has established the "equivalence theorem" between the overflow probability in variable-length lossless source coding and the error probability in fixed-length source coding. However, this chapter has shown that (3.76) can be verified by different approach from the combination of the results in [29] and [43].*

## 3.8 Discussion and conclusion of this chapter

Some discussions on the main results are stated as follows.

1) *Variable-length lossless source coding under the overflow probability vs. fixed-length source coding:* Suppose that there exists an $(R, \delta)_{\mathrm{p}}$ code. Further, let $x_i$ denote the element of $\mathcal{X}$ that has the $i$-th largest probability. That is, we have

$$p_X(x_1) \geq p_X(x_2) \geq p_X(x_3) \geq \cdots . \tag{3.78}$$

Now, we consider the following fixed-length encoder $\varphi$:

- For the most probable $2^R - 1$ source symbols (i.e., $x_1, \ldots, x_{2^R-1}$), the encoder $\varphi$ assigns a unique codeword whose codeword length is $R$ bits,

and

- for the rest of the source symbols, the encoder $\varphi$ assigns the remaining codeword whose codeword length is $R$ bits.

The decoder $\psi$ is defined as the inverse function of the encoder $\varphi$. Then, the above construction and the existence of an $(R, \delta)_{\mathrm{p}}$ code guarantee that the coding rate of $(\varphi, \psi)$ is $R$ and the error probability of $(\varphi, \psi)$ is less than or equal to $\delta$. Therefore, $R^*_{\mathrm{p}}(\delta)$ (the minimum threshold of the overflow probability) represents the minimum coding rate in fixed-length source coding. Further, the above argument shows that optimal variable-length lossless source coding under the overflow probability and the optimal fixed-length source coding are essentially the same. The only difference is the treatment of least likely source symbols: fixed-length source coding treats least likely source symbols as decoding error while variable-length lossless source coding treats them as overflow.

2) *Non-prefix code vs. prefix code:* Comparing Theorem 3.5.1 and Theorem 3.6.1, it is observed that the results for the optimal non-prefix code and the optimal prefix code differ 1 bit. This is because we have the relationship between leaf nodes and interior nodes of a code tree which is shown in Fig. 3.1. In view of this relationship, results for a non-prefix code can be easily converted to results for a prefix code and vice versa. Since descriptions for a non-prefix code are simpler than that of a prefix code, we only consider a non-prefix code in the following Chapters 4 and 5.

3) *Asymptotic distribution of the self-information vs. non-asymptotic distribution of source symbols:* As shown in Section 3.4, Nomura et al. [30] have derived the fundamental limit on the overflow probability by focusing on the tail probability of the asymptotic distribution of the self-information $\frac{1}{n} \log \frac{1}{p_{X^n}(X^n)}$ (see the left picture of Fig. 3.2). On the other hand, this study derives the fundamental limit on the overflow probability by focusing on the tail probability of the distribution

of $X$ ordered in decreasing probability such that $p_X(x_1) \geq p_X(x_2) \geq p_X(x_3) \geq \cdots$ (see the right picture of Fig. 3.2). Therefore, both Nomura et al. [30] and this study focus on least likely source symbols. However, it is worth noticing that our viewpoint of the tail probability (i.e., the tail probability of the distribution of $X$ ordered in decreasing probability) is compatible with the definition of the smooth max entropy. This is one reason why we succeeded in the characterization of the fundamental limit on the overflow probability by using the smooth max entropy.

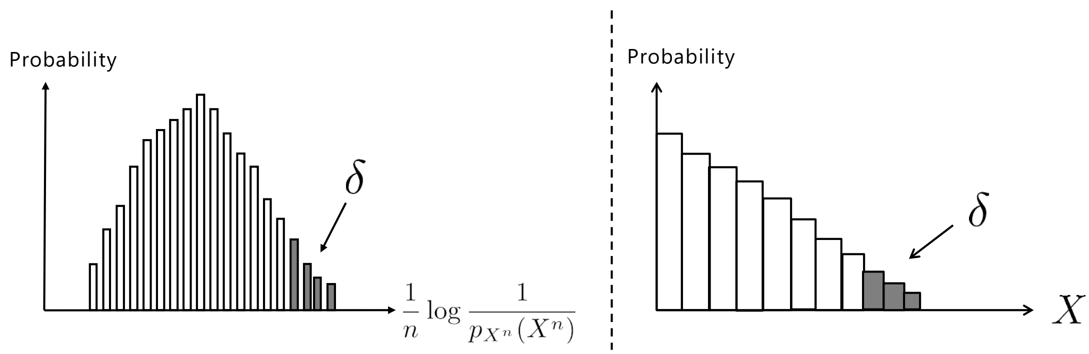


Figure 3.2: The distribution of the self-information (left) and the distribution of source symbols ordered in decreasing probability (right)

To summarize this chapter, we considered the problem of variable-length lossless source coding for a general source. The non-asymptotic coding theorems were obtained for both prefix codes and non-prefix codes, and the difference between these codes were shown. The smooth max entropy was an important quantity to characterize the fundamental limits. To show the achievability results, we used the explicit code construction instead of using the random coding argument. Further, this chapter showed the asymptotic coding theorems, which were easily derived from the non-asymptotic results. Note that this chapter contains fundamental and basic ideas used throughout Chapters 4 and 5, and is the gateway to Chapters 4 and 5.

# Chapter 4

# Non-asymptotic and asymptotic analyses of variable-length source coding allowing errors

## 4.1 Introduction

In Chapter 3, we have considered variable-length lossless source coding, i.e., variable-length source coding in which the error probability is zero. On the other hand, this chapter allows a positive error probability and deals with the problem of variable-length source coding allowing errors for a general source.

For this problem, Nomura and Yagi [33] have treated the case where both the overflow probability and the error probability may be positive. They have focused on the asymptotic distribution of the self-information $\frac{1}{n}\log\frac{1}{p_{X^n}(X^n)}$ and derived the asymptotic characterization on the minimum threshold of the overflow probability.

On the other hand, as we have explained in Chapter 3, we consider the distribution of $X$ ordered in decreasing probability, and evaluate the minimum threshold of the overflow probability by using the smooth max entropy. As shown in Chapter 3, we use the explicit code construction, instead of using the random coding argument, to show the achievability results. It is clarified that the proof of the achievability result in this chapter is a generalization of that in Chapter 3. Our proof of achievability results clarifies the difference between the deterministic encoder and the stochastic encoder. Further, our achievability results show that the overflow probability and the error probability are trade-off. Moreover, the obtained results clarify the benefit of allowing a positive error probability instead of zero error case. Using the results obtained in the non-asymptotic regime, we establish the asymptotic coding theorems.

This chapter is organized as follows. Section 4.2 describes the problem formulation. Section 4.3 shows the related prior work. Section 4.4 derives the non-asymptotic coding theorem for stochastic codes. Section 4.5 shows the non-asymptotic coding theorem for deterministic codes. Based on the results obtained in the non-asymptotic setting, Section 4.6 derives the asymptotic coding theorem. Finally, Section 4.7 discusses the main results and concludes this chapter.

## 4.2 Problem formulation

Let $\mathcal{X}$ be a source alphabet, which is a finite set. Let $X$ be a random variable taking a value in $\mathcal{X}$ and $x$ be a realization of $X$. The probability distribution of $X$ is denoted as $P_X$ and the probability mass function of $X$ is denoted as $p_X(x)$.

As we have seen in Chapter 3, results between a prefix code and a non-prefix code differ at most one bit. Thus, we discuss only non-prefix codes in the following. The pair of an encoder and a decoder $(f, g)$ is defined as follows. An encoder $f$ is defined as $f : \mathcal{X} \to \{0, 1\}^\star$. An encoder $f$ can be a stochastic code and produces a non-prefix code. Further, we allow $f(x) = f(x')$ for $x \neq x'$. For $x \in \mathcal{X}$, the codeword length of $f(x)$ is denoted as $\ell(f(x))$. A deterministic decoder $g$ is defined as $g : \{0, 1\}^\star \to \mathcal{X}$.

Using the error probability and the overflow probability, we define an $(R, \epsilon, \delta)$ code.

**Definition 4.2.1** *Given $R \geq 0$ and $\epsilon, \delta \in [0, 1)$, a code $(f, g)$ satisfying*

$$\mathbb{P}[X \neq g(f(X))] \leq \epsilon, \tag{4.1}$$

$$\mathbb{P}[\ell(f(X)) > R] \leq \delta \tag{4.2}$$

*is called an $(R, \epsilon, \delta)$ code.*

The fundamental limits are the minimum thresholds $R^*(D, \epsilon, \delta)$ and $\tilde{R}(D, \epsilon, \delta)$ for given $\epsilon$ and $\delta$.

**Definition 4.2.2** *Given $\epsilon, \delta \in [0, 1)$, we define*

$$R^*(\epsilon, \delta) := \inf\{R : \exists \text{ an } (R, \epsilon, \delta) \text{ code}\}, \tag{4.3}$$

$$\tilde{R}(\epsilon, \delta) := \inf\{R : \exists \text{ a deterministic } (R, \epsilon, \delta) \text{ code}\}. \tag{4.4}$$

We consider the following problem formulation in the asymptotic analysis. Let $\mathcal{X}^n$ be the $n$-th Cartesian product of $\mathcal{X}$. Let $X^n$ be a random variable taking a value in $\mathcal{X}^n$ and $x^n$ be a realization of $X^n$. Moreover, let $\mathbf{X} = \{X^n\}_{n=1}^\infty$ denote a general source. The joint probability distribution of $X^n$ is denoted as $P_{X^n}$ and the joint probability mass function of $X^n$ is denoted as $p_{X^n}(x^n)$. An encoder $f_n : \mathcal{X}^n \to \{0, 1\}^\star$ is a possibly stochastic code and produces a non-prefix code. A decoder $g_n : \{0, 1\}^\star \to \mathcal{X}^n$ is a deterministic code.

We define an $(n, R, \epsilon, \delta)$ code as follows.

**Definition 4.2.3** *Given $R \geq 0$ and $\epsilon, \delta \in [0, 1)$, a code $(f_n, g_n)$ satisfying*

$$\mathbb{P}[X^n \neq g_n(f_n(X^n))] \leq \epsilon, \tag{4.5}$$

$$\mathbb{P}\left[\frac{1}{n}\ell(f_n(X^n)) > R\right] \leq \delta \tag{4.6}$$

*is called an $(n, R, \epsilon, \delta)$ code.*

The asymptotic fundamental limit is the following minimum threshold.

**Definition 4.2.4** *Given $\epsilon, \delta \in [0,1)$, $R(\epsilon, \delta|\mathbf{X})$ is the infimum of the threshold $R$ such that there exists an $(n, R, \epsilon, \delta)$ code for all $n \geq n_0$ with some $n_0 > 0$.*

Furthermore, another asymptotic fundamental limit is defined as follows.

**Definition 4.2.5** *Given $\epsilon, \delta \in [0,1)$, $\hat{R}(\epsilon, \delta|\mathbf{X})$ is the infimum of the threshold $R$ such that there exists a code $(f_n, g_n)$ satisfying*

$$\limsup_{n\to\infty} \mathbb{P}[X^n \neq g_n(f_n(X^n))] \leq \epsilon, \tag{4.7}$$

$$\limsup_{n\to\infty} \mathbb{P}\left[\frac{1}{n}\ell(f_n(X^n)) > R\right] \leq \delta. \tag{4.8}$$

## 4.3 Related work

Nomura and Yagi [33] have derived the next result on $\hat{R}(\epsilon, \delta|\mathbf{X})$.

**Theorem 4.3.1 ([33])** *For any $\epsilon, \delta \in [0,1)$ satisfying $\epsilon + \delta < 1$, it holds that*

$$\hat{R}(\epsilon, \delta|\mathbf{X}) = G_{\epsilon,\delta}(\mathbf{X}), \tag{4.9}$$

*where $G_{\epsilon,\delta}(\mathbf{X})$ is defined by*

$$G_{\epsilon,\delta}(\mathbf{X}) = \inf\left\{R : \lim_{\nu\downarrow 0}\limsup_{n\to\infty} \inf_{\mathbb{P}[X^n\in\mathcal{A}_n]\geq 1-\epsilon-\nu} \mathbb{P}\left[-\frac{1}{n}\log\frac{p_{X^n}(X^n)}{\mathbb{P}[X^n\in\mathcal{A}_n]} \geq R, X^n \in \mathcal{A}_n\right]\right\}. \tag{4.10}$$

Furthermore, the study [33] has shown the next result, which is another characterization of $\hat{R}(\epsilon, \delta|\mathbf{X})$.

**Theorem 4.3.2 ([33])** *For any $\epsilon, \delta \in [0,1)$ satisfying $\epsilon + \delta < 1$, it holds that*

$$G_{\epsilon,\delta}(\mathbf{X}) = \tilde{H}_{\epsilon,\delta}(\mathbf{X}) = \overline{H}_{\epsilon+\delta}(\mathbf{X}), \tag{4.11}$$

*where $\tilde{H}_{\epsilon,\delta}(\mathbf{X})$ is defined by*

$$\tilde{H}_{\epsilon,\delta}(\mathbf{X}) = \inf\left\{R : \lim_{\nu\downarrow 0} F(\epsilon+\nu, R) \leq \delta\right\}, \tag{4.12}$$

$$F(\tau, R) = \limsup_{n\to\infty} \inf_{\mathbb{P}[X^n\in A_n]\geq 1-\tau} \mathbb{P}\left[\frac{1}{n}\log\frac{1}{p_{X^n}(X^n)} \geq R, X^n \in A_n\right], \tag{4.13}$$

*and $\overline{H}_{\epsilon+\delta}(\mathbf{X})$ is defined by*

$$\overline{H}_{\epsilon+\delta}(\mathbf{X}) = \inf\left\{R : \limsup_{n\to\infty} \mathbb{P}\left[\frac{1}{n}\log\frac{1}{p_{X^n}(X^n)} > R\right] \leq \epsilon+\delta\right\}. \tag{4.14}$$

Note that the study by Nomura and Yagi [33] has considered the asymptotic analysis, whereas this study deals with the non-asymptotic analysis as well as the asymptotic analysis.

## 4.4 Non-asymptotic coding theorem for stochastic codes

The next lemma shows the achievability result on $R$ of an $(R, \epsilon, \delta)$ code.

**Lemma 4.4.1** *For any* $\epsilon, \delta \in [0, 1)$ *satisfying* $\epsilon + \delta < 1$*, there exists an* $(R, \epsilon, \delta)$ *code such that*

$$R = \lfloor H^{\epsilon+\delta}(X) \rfloor. \tag{4.15}$$

(**Proof**) Before the construction of the encoder and decoder is described, some notations are introduced.

- Let $x_i$ be the element of $\mathcal{X}$ which has the $i$-th largest probability. That is, it holds that $p_X(x_1) \geq p_X(x_2) \geq p_X(x_3) \geq \cdots$.
- Let $i^* \geq 1$ be the integer satisfying

$$\sum_{i=1}^{i^*-1} p_X(x_i) < 1 - \epsilon - \delta, \tag{4.16}$$

$$\sum_{i=1}^{i^*} p_X(x_i) \geq 1 - \epsilon - \delta. \tag{4.17}$$

- Let $k^* \geq 1$ be the integer satisfying

$$\sum_{i=1}^{k^*-1} p_X(x_i) < 1 - \epsilon, \tag{4.18}$$

$$\sum_{i=1}^{k^*} p_X(x_i) \geq 1 - \epsilon. \tag{4.19}$$

  This definition yields $k^* \geq i^*$.

- Let $\alpha$ and $\beta$ be defined as

$$\alpha := \sum_{i=1}^{k^*-1} p_X(x_i), \tag{4.20}$$

$$\beta := 1 - \epsilon - \alpha. \tag{4.21}$$

- Let $w_i$ be the $i$-th binary string in $\{0, 1\}^\star$ in the increasing order of the length and ties are arbitrarily broken. For example, $w_1 = \Lambda, w_2 = 0, w_3 = 1, w_4 = 00, w_5 = 01$, etc.

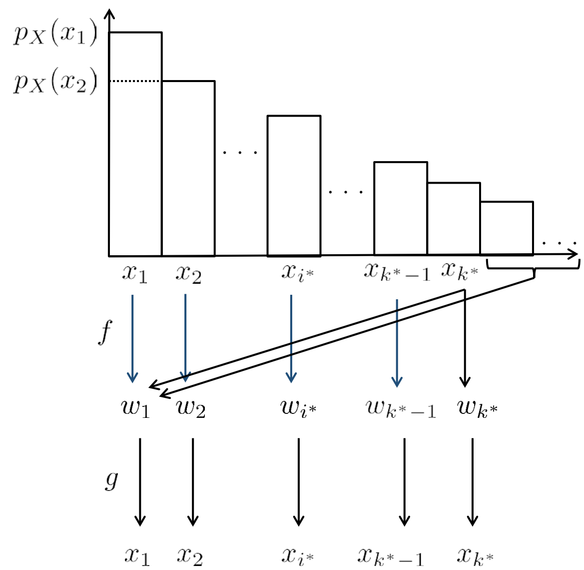


Figure 4.1: Construction of the encoder and the decoder in variable-length compression allowing errors

The encoder $f : \mathcal{X} \to \{0,1\}^\star$ and decoder $g : \{0,1\}^\star \to \mathcal{X}$ are constructed as follows (see Fig. 4.1).

**[Encoder]**

1) For $i = 1, \ldots, k^* - 1$, set $f(x_i) = w_i$.

2) For $x_{k^*}$, set[1]

$$f(x_{k^*}) = \begin{cases} w_{k^*} & \text{with prob. } \frac{\beta}{p_X(x_{k^*})}, \\ w_1 & \text{with prob. } 1 - \frac{\beta}{p_X(x_{k^*})}. \end{cases} \tag{4.22}$$

3) For $x \in \{x_{k^*+1}, x_{k^*+2}, \ldots\}$, set $f(x) = w_1$.

**[Decoder]** Set $g(w_i) = x_i$ $(i = 1, \ldots, k^*)$.

Now, the error probability is evaluated. The construction of the encoder and the decoder yields

$$g(f(x_i)) = x_i \tag{4.23}$$

[1]Note that we have $p_X(x_{k^*}) \geq \beta$ from (4.19).

for $i = 1, \ldots, k^* - 1$. Furthermore, $g(f(x_{k^*})) = x_{k^*}$ holds with probability $\beta / p_X(x_{k^*})$ for $x_{k^*}$. Thus, it holds that

$$\mathbb{P}[g(f(X)) = X] = \sum_{i=1}^{k^*-1} p_X(x_i) + \mathbb{P}[f(X) = w_{k^*}, X = x_{k^*}] \tag{4.24}$$

$$= \alpha + \beta \tag{4.25}$$

$$= 1 - \epsilon. \tag{4.26}$$

This leads to

$$\mathbb{P}[g(f(X)) \neq X] = \epsilon. \tag{4.27}$$

Next, the overflow probability is evaluated. The construction of the encoder verifies that $\ell(w_i) = \lfloor \log i \rfloor$ $(i = 1, \ldots, k^*)$. Moreover, the definition of the smooth max entropy gives $H^{\epsilon+\delta}(X) = \log i^*$. Hence, setting $R = \lfloor \log i^* \rfloor = \lfloor H^{\epsilon+\delta}(X) \rfloor$, we have

$$\mathbb{P}[\ell(f(X)) > R] \leq \sum_{i=i^*+1}^{k^*-1} \mathbb{P}[X = x_i] + \mathbb{P}[f(X) = w_{k^*}, X = x_{k^*}] \tag{4.28}$$

$$= \sum_{i=1}^{k^*-1} \mathbb{P}[X = x_i] - \sum_{i=1}^{i^*} \mathbb{P}[X = x_i] + \beta \tag{4.29}$$

$$\leq \alpha - (1 - \epsilon - \delta) + \beta \tag{4.30}$$

$$= \delta, \tag{4.31}$$

where the last inequality follows from the definition of $\alpha$ and (4.17); the last equality is due to the definition of $\beta$. □

The next lemma shows the converse bound on $R$ of an $(R, \epsilon, \delta)$ code.

**Lemma 4.4.2** *For any $\epsilon, \delta \in [0, 1)$ satisfying $\epsilon + \delta < 1$, any $(R, \epsilon, \delta)$ code satisfies*

$$R > H^{\epsilon+\delta}(X) - 1. \tag{4.32}$$

(**Proof**) Fix $(R, \epsilon, \delta)$ code $(f, g)$ arbitrarily. The definition of an $(R, \epsilon, \delta)$ code gives

$$\mathbb{P}[X \neq g(f(X))] \leq \epsilon, \tag{4.33}$$

$$\mathbb{P}[\ell(f(X)) > R] \leq \delta. \tag{4.34}$$

Let $\mathcal{S}$ and $\mathcal{T}$ be defined as

$$\mathcal{S} := \{x \in \mathcal{X} : x \neq g(f(x))\}, \tag{4.35}$$

$$\mathcal{T} := \{x \in \mathcal{X} : \ell(f(x)) > R\}. \tag{4.36}$$

Furthermore, let

$$\mathcal{V} := \mathcal{S}^c \cap \mathcal{T}^c. \tag{4.37}$$

Then, it holds that

$$\begin{aligned}\mathbb{P}[X \in \mathcal{V}] &= \mathbb{P}[X \in \mathcal{S}^c \cap \mathcal{T}^c] && (4.38)\\ &\overset{(a)}{=} \mathbb{P}[X \in (\mathcal{S} \cup \mathcal{T})^c] && (4.39)\\ &= 1 - \mathbb{P}[X \in \mathcal{S} \cup \mathcal{T}]\\ &\overset{(b)}{\geq} 1 - (\mathbb{P}[X \in \mathcal{S}] + \mathbb{P}[X \in \mathcal{T}]) && (4.40)\\ &\overset{(c)}{\geq} 1 - (\epsilon + \delta), && (4.41)\end{aligned}$$

where

- $(a)$ follows from De Morgan's laws,
- $(b)$ is due to union bound,
- $(c)$ follows from (4.33)–(4.36).

Hence, the definition of the smooth max entropy and (4.41) yield

$$H^{\epsilon+\delta}(X) \leq \log |\mathcal{V}|. \tag{4.42}$$

On the other hand, it holds that

$$|\mathcal{V}| \overset{(a)}{\leq} 1 + 2 + 2^2 + \cdots + 2^{\lfloor R \rfloor} = 2^{\lfloor R \rfloor + 1} - 1 < 2^{R+1}, \tag{4.43}$$

where $(a)$ follows from (4.37), i.e., we have $\ell(f(x)) \leq R$ and $x = g(f(x))$ for any $x \in \mathcal{V}$ and this yields $(a)$. Finally, (4.32) is obtained from (4.42) and (4.43). □

Combination of Lemmas 4.4.1 and 4.4.2 gives the following result on $R^*(\epsilon, \delta)$.

**Theorem 4.4.1** *For any $\epsilon, \delta \in [0, 1)$ satisfying $\epsilon + \delta < 1$, it holds that*

$$H^{\epsilon+\delta}(X) - 1 < R^*(\epsilon, \delta) \leq \lfloor H^{\epsilon+\delta}(X) \rfloor. \tag{4.44}$$

Theorem 4.4.1 shows that $R^*(\epsilon, \delta)$ can be specified within one bit in the interval not exceeding $H^{\epsilon+\delta}(X)$ regardless of the values of $\epsilon$ and $\delta$.

**Remark 4.4.1** *In Theorem 4.4.1, consider the special case of $\epsilon = 0$. In this case, (4.44) coinsides with (3.64). Furthermore, Theorem 4.4.1 shows the merit of allowing the positive error probability. Because the smooth max entropy $H^{\gamma}(X)$ is a monotonically non-increasing function of $\gamma$, $H^{\epsilon+\delta}(X) \leq H^{\delta}(X)$ holds. Therefore, comparing (3.64) with (4.44), it is observed that the upper and lower bounds of the minimum threshold of the overflow probability can be smaller than the zero error (lossless) case by allowing the non-zero error probability.*

## 4.5 Non-asymptotic coding theorem for deterministic codes

The next lemma shows the achievability result on $R$ of a deterministic $(D, R, \epsilon, \delta)$ code.

**Lemma 4.5.1** *For any $\epsilon, \delta \in [0, 1)$ satisfying $\epsilon + \delta < 1$, there exists a deterministic $(R, \epsilon, \delta)$ code such that*

$$R = \left\lfloor H^{\epsilon+\delta}(X) + \frac{2 \log e}{2^{H^{\epsilon+\delta}(X)}} \right\rfloor. \tag{4.45}$$

(**Proof**) First, some notations are defined.

- Let $k^* \geq 1$ be the integer satisfying (4.18) and (4.19).
- Define $\gamma$ as

  $$\gamma = 1 - \sum_{i=1}^{k^*} p_X(x_i). \tag{4.46}$$

  Then, it holds that $\gamma \leq \epsilon$.
- Let $j^* \geq 1$ be the integer satisfying

  $$\sum_{i=1}^{j^*-1} p_X(x_i) < 1 - \gamma - \delta, \tag{4.47}$$

  $$\sum_{i=1}^{j^*} p_X(x_i) \geq 1 - \gamma - \delta. \tag{4.48}$$

The deterministic encoder $f : \mathcal{X} \to \{0, 1\}^\star$ and decoder $g : \{0, 1\}^\star \to \mathcal{X}$ are constructed as follows.

**[Encoder]**

1) For $i = 1, \ldots, k^*$, set $f(x_i) = w_i$.
2) For $x \in \{x_{k^*+1}, x_{k^*+2}, \ldots\}$, set $f(x) = w_1$.

**[Decoder]** Set $g(w_i) = x_i$ $(i = 1, \ldots, k^*)$.

Now, we evaluate the error probability. From the definition of the encoder and decoder, we have

$$\mathbb{P}[X = g(f(X))] = \sum_{i=1}^{k^*} p_X(x_i) \tag{4.49}$$

$$\geq 1 - \epsilon. \tag{4.50}$$

Thus, it holds that

$$\mathbb{P}[X \neq g(f(X))] \leq \epsilon. \tag{4.51}$$

Next, the overflow probability is evaluated. The definition of the encoder establishes

$$\mathbb{P}[f(X) = w_1] = p_X(x_1) + \gamma, \tag{4.52}$$
$$\mathbb{P}[f(X) = w_i] = p_X(x_i) \quad (i = 2, \ldots, k^*). \tag{4.53}$$

Setting $R = \lfloor \log \min(j^*, k^*) \rfloor$, it holds that[2]

$$\begin{aligned}
\mathbb{P}\left[\ell(f(X)) > R\right] &\leq 1 - \sum_{i=1}^{j^*} \mathbb{P}[f(X) = w_i] && (4.54)\\
&= 1 - \left( \sum_{i=1}^{j^*} p_X(x_i) + \gamma \right) && (4.55)\\
&\leq 1 - ((1 - \gamma - \delta) + \gamma) && (4.56)\\
&= \delta, && (4.57)
\end{aligned}$$

where the last inequality follows from (4.48).

Therefore, the code $(f, g)$ is a deterministic $(R, \epsilon, \delta)$ code with $R = \lfloor \log \min(j^*, k^*) \rfloor$.

Let $i^*$ be the integer satisfying (4.16) and (4.17). Then, the proof of Lemma 4.4.1 gives

$$\log i^* = H^{\epsilon+\delta}(X). \tag{4.58}$$

The fact that $\gamma \leq \epsilon$ yields the inequality $i^* \leq j^*$ and $i^* \leq k^*$. This means that $i^* \leq \min(j^*, k^*)$. If $i^* = \min(j^*, k^*)$, $\min(j^*, k^*) \leq i^* + 2$ obviously holds. Then, assuming that $i^* < \min(j^*, k^*)$, if

$$\min(j^*, k^*) \leq i^* + 2 \tag{4.59}$$

holds, then

$$\log \min(j^*, k^*) \leq \log(i^* + 2) \leq \log i^* + \frac{2 \log e}{i^*} \tag{4.60}$$

is derived[3], and we obtain (4.45).

[2] If $R = \lfloor \log k^* \rfloor$, then $\mathbb{P}[\ell(f(X)) > R] = 0$.

[3] The rightmost inequality follows from Taylor's expansion.

In the following, the inequality $\min(j^*, k^*) \leq i^* + 2$ is proved. The first step of the proof is the following inequality:

$$\sum_{i=1}^{j^*} p_X(x_i) - \sum_{i=1}^{i^*} p_X(x_i) \overset{(a)}{\leq} 1 - \gamma - \delta + p_X(x_{j^*}) - (1 - \epsilon - \delta) \tag{4.61}$$

$$= p_X(x_{j^*}) + \epsilon - \gamma \tag{4.62}$$

$$\overset{(b)}{\leq} p_X(x_{j^*}) + p_X(x_{i^*+1}), \tag{4.63}$$

where $(a)$ follows from (4.17) and (4.47); $(b)$ follows from

$$\epsilon - \gamma \leq \left(1 - \sum_{i=1}^{k^*-1} p_X(x_i)\right) - \left(1 - \sum_{i=1}^{k^*} p_X(x_i)\right) \tag{4.64}$$

$$= p_X(x_{k^*}) \tag{4.65}$$

$$\leq p_X(x_{i^*+1}). \tag{4.66}$$

Inequality (4.63) is equivalent to

$$\sum_{i=1}^{j^*-1} p_X(x_i) \leq \sum_{i=1}^{i^*+1} p_X(x_i). \tag{4.67}$$

Thus, $j^* - 1 \leq i^* + 1$ is obtained, which implies that $\min(j^*, k^*) \leq i^* + 2$. □

Combination of Lemma 4.5.1 and the fact that $R^*(\epsilon, \delta) \leq \tilde{R}(\epsilon, \delta)$ gives the following result on $\tilde{R}(\epsilon, \delta)$.

**Theorem 4.5.1** *For any $\epsilon, \delta \in [0, 1)$ satisfying $\epsilon + \delta < 1$, it holds that*

$$H^{\epsilon+\delta}(X) - 1 < \tilde{R}(\epsilon, \delta) \leq \left\lfloor H^{\epsilon+\delta}(X) + \frac{2 \log e}{2^{H^{\epsilon+\delta}(X)}} \right\rfloor. \tag{4.68}$$

By Theorem 4.5.1, $\tilde{R}(\epsilon, \delta)$ can be specified in the interval within four bits. Thus, it is observed that this result is slightly weaker than that for stochastic codes (Theorem 4.4.1).

## 4.6 Asymptotic coding theorem

The next theorem characterizes $R(\epsilon, \delta|\mathbf{X})$ by the smooth max entropy.

**Theorem 4.6.1** *For any $\epsilon, \delta \in [0, 1)$,*

$$R(\epsilon, \delta|\mathbf{X}) = \limsup_{n \to \infty} \frac{1}{n} H^{\epsilon+\delta}(X^n). \tag{4.69}$$

(**Proof**) First,

$$R(\epsilon, \delta|\mathbf{X}) \leq \limsup_{n\to\infty} \frac{1}{n} H^{\epsilon+\delta}(X^n) \tag{4.70}$$

is proved. From Lemma 4.5.1, there exists a code $(f_n, g_n)$ satisfying

$$\mathbb{P}[X^n \neq g_n(f_n(X^n))] \leq \epsilon, \tag{4.71}$$

$$\mathbb{P}\left[\frac{1}{n}\ell(f_n(X^n)) > \frac{1}{n}H^{\epsilon+\delta}(X^n) + \frac{2\log e}{n2^{H^{\epsilon+\delta}(X^n)}}\right] \leq \delta. \tag{4.72}$$

Fix $\gamma > 0$ arbitrarily. Then, it holds that

$$\frac{1}{n}H^{\epsilon+\delta}(X^n) \leq \limsup_{n\to\infty} \frac{1}{n}H^{\epsilon+\delta}(X^n) + \gamma \tag{4.73}$$

and

$$\frac{2\log e}{n2^{H^{\epsilon+\delta}(X^n)}} \leq \gamma \tag{4.74}$$

for all $n \geq n_0$ with some $n_0 > 0$. Hence, from (4.72), we have

$$\mathbb{P}\left[\frac{1}{n}\ell(f_n(X^n)) > \limsup_{n\to\infty} \frac{1}{n}H^{\epsilon+\delta}(X^n) + 2\gamma\right] \leq \delta \tag{4.75}$$

for all $n \geq n_0$. Thus, from (4.71) and (4.75), $(f_n, g_n)$ is an $(n, R, \epsilon, \delta)$ code with

$$R = \limsup_{n\to\infty} \frac{1}{n}H^{\epsilon+\delta}(X^n) + 2\gamma \tag{4.76}$$

for all $n \geq n_0$. Since $\gamma > 0$ is arbitrary, this indicates the desired inequality (4.70).

Next,

$$R(\epsilon, \delta|\mathbf{X}) \geq \limsup_{n\to\infty} \frac{1}{n} H^{\epsilon+\delta}(X^n) \tag{4.77}$$

is shown. For any $(n, R, \epsilon, \delta)$ code, Lemma 4.4.2 gives

$$nR > H^{\epsilon+\delta}(X^n) - 1. \tag{4.78}$$

Therefore, it holds that

$$R \geq \limsup_{n\to\infty} \frac{1}{n} H^{\epsilon+\delta}(X^n) \tag{4.79}$$

for any $(n, R, \epsilon, \delta)$ code. Hence, (4.77) is obtained. □

**Remark 4.6.1** *Theorems 4.4.1 and 4.5.1 show that, in the non-asymptotic regime, the results on the minimum threshold are different for stochastic and deterministic encoders. However, Theorem 4.6.1 indicates that, in the asymptotic regime, the restriction to deterministic encoders does not affect the minimum threshold.*

The next result on $\hat{R}(\epsilon, \delta|\mathbf{X})$ is derived by almost the same proof of Theorem 4.6.1.

**Corollary 4.6.1** *For any $\epsilon, \delta \in [0, 1)$, it holds that*

$$\hat{R}(\epsilon, \delta|\mathbf{X}) = \lim_{\tau \downarrow 0} \limsup_{n \to \infty} \frac{1}{n} H^{\epsilon+\delta+\tau}(X^n). \tag{4.80}$$

**Remark 4.6.2** *Similar to Remark 3.7.1, the difference between (4.69) and (4.80) is due to the difference of the asymptotic constraint, i.e., (4.69) is the result under the constraint of (4.5) and (4.6) for all $n \geq n_0$ with some $n_0$, while (4.80) is the result under the constraint of (4.7) and (4.8).*

## 4.7 Discussion and conclusion of this chapter

We state some discussions regarding the main results in this chapter.

1) *Generalization of the proof of the achievability result:* The proof of Lemma 4.4.1 can be viewed as a generalization of the proof of Lemma 3.6.1. Indeed, setting $\epsilon = 0$ in the proof of Lemma 4.4.1, the construction of the encoder in the proof of Lemma 4.4.1 coincides with that in the proof of Lemma 3.6.1.

2) *Trade-off relationship between the error probability and the overflow probability:* In Lemma 4.4.1, Eq. (4.15) indicates that the threshold $R$ is determined by the sum of $\epsilon$ (the tolerated error probability) and $\delta$ (the tolerated overflow probability). Suppose that $\epsilon + \delta = C$ for some constant $C$. Then, $\epsilon$ and $\delta$ has a trade-off relationship, i.e., if $\epsilon$ is high, then $\delta$ is low and vice versa. Therefore, we see that the overflow probability and the error probability has a trade-off relationship in the case that the sum of $\epsilon$ and $\delta$ is a constant.

3) *Stochastic code vs. deterministic code:* Comparing the stochastic code in the proof of Lemma 4.4.1 and the deterministic code in the proof of Lemma 4.5.1, it is observed that the only difference between these codes is the treatment of the source symbol $x_{k^*}$. That is, for the stochastic code in the proof of Lemma 4.4.1, randomization is implemented at the source symbol $x_{k^*}$.

To summarize this chapter, we considered the problem of variable-length source coding allowing errors for a general source. Based on the smooth max entropy, the non-asymptotic coding theorems were derived for both stochastic codes and deterministic codes. To prove the achievability results, the explicit code construction was used. The asymptotic coding theorems were obtained based on the non-asymptotic fundamental limits. Note that basic ideas of the proofs in this chapter are based on those in Chapter 3.

# Chapter 5

# Non-asymptotic and asymptotic analyses of variable-length lossy source coding

## 5.1 Introduction

In Chapter 4, we have considered variable-length source coding allowing errors. This chapter generalizes this setting and deals with the problem of variable-length lossy source coding under the excess distortion probability.

For this problem, Yagi and Nomura [49] have considered the case where either the overflow probability or the excess distortion probability does not exceed a positive constant asymptotically. Nomura and Yagi [32] have treated the case where the probability of union of events that the excess distortion occurs and the overflow occurs does not exceed a positive constant asymptotically. Both of these results have focused on the asymptotic distribution of the self-mutual information $\frac{1}{n}\log\frac{p_{Y^n|X^n}(Y^n|X^n)}{p_{Y^n}(Y^n)}$ and derived the asymptotic characterization on the minimum threshold of the overflow probability.

We evaluate the fundamental limit on the overflow probability as in [32] and [49]. The superficial differences between these previous works and this study are

1. this study considers the case where both the excess distortion probability and the overflow probability may be positive,
2. this study investigates both non-asymptotic and asymptotic cases.

However, we emphasize that the essential difference is that we focus on the distribution of $X$ ordered in decreasing probability. This viewpoint enables us to evaluate the minimum threshold of the overflow probability by using a new smooth max entropy-based quantity. To show the achievability results, we use the explicit code construction, which can be viewed as a generalization of that in Chapter 4. Similar to Chapter 4, our proof of achievability results clarifies the difference between the deterministic encoder and the stochastic encoder. Further, our achievability results show that the overflow probability and the excess distortion probability are trade-off. Using the non-asymptotic results, asymptotic coding theorem is established.

This chapter is organized as follows. Section 5.2 sets up the problem formulation. Section 5.3 describes the related previous work. Section 5.4 introduces the notion of majorization and Schur concavity. Then, this section defines the smooth max entropy-based quantity, which plays a significant role in the main results. Section 5.5 derives the non-asymptotic coding theorems for stochastic codes. The non-asymptotic coding theorems for deterministic codes are shown in Section 5.6. Section 5.7 derives the asymptotic coding theorem. Section 5.8 discusses and concludes this chapter. Finally, Section 5.9 summarizes some properties that the achievability part and the converse part have in common throughout Chapters 3–5.

## 5.2 Problem formulation

Let $\mathcal{X}$ be a source alphabet and $\mathcal{Y}$ be a reproduction alphabet, where both are finite sets. Let $X$ be a random variable taking a value in $\mathcal{X}$ and $x$ be a realization of $X$. The probability distribution of $X$ is denoted as $P_X$ and the probability mass function of $X$ is denoted as $p_X(x)$. A distortion measure $d$ is defined as $d : \mathcal{X} \times \mathcal{Y} \to [0, +\infty)$.

As we have seen in Chapter 3, results between a prefix code and a non-prefix code differ at most one bit. Thus, we discuss only non-prefix code in the following. The pair of an encoder and a decoder $(f, g)$ is defined as follows. An encoder $f$ is defined as $f : \mathcal{X} \to \{0, 1\}^\star$. An encoder $f$ may be a stochastic code and produces a non-prefix code. For $x \in \mathcal{X}$, the codeword length of $f(x)$ is denoted as $\ell(f(x))$. A deterministic decoder $g$ is defined as $g : \{0, 1\}^\star \to \mathcal{Y}$.

Using the excess distortion and the overflow probabilities, we define a $(D, R, \epsilon, \delta)$ code.

**Definition 5.2.1** *Given $D, R \geq 0$ and $\epsilon, \delta \in [0, 1)$, a code $(f, g)$ satisfying*

$$\mathbb{P}[d(X, g(f(X))) > D] \leq \epsilon, \tag{5.1}$$

$$\mathbb{P}[\ell(f(X)) > R] \leq \delta \tag{5.2}$$

*is called a $(D, R, \epsilon, \delta)$ code.*

The fundamental limits that are analyzed in this chapter are the minimum thresholds $R^*(D, \epsilon, \delta)$ and $\tilde{R}(D, \epsilon, \delta)$ for given $D$, $\epsilon$, and $\delta$.

**Definition 5.2.2** *Given $D \geq 0$ and $\epsilon, \delta \in [0, 1)$, we define*

$$R^*(D, \epsilon, \delta) := \inf\{R : \exists \text{ a } (D, R, \epsilon, \delta) \text{ code}\}, \tag{5.3}$$

$$\tilde{R}(D, \epsilon, \delta) := \inf\{R : \exists \text{ a deterministic } (D, R, \epsilon, \delta) \text{ code}\}. \tag{5.4}$$

**Remark 5.2.1** *Consider the special case $\delta = 0$. A $(D, R, \epsilon, 0)$ code gives a fixed-length code achieving rate $R$ and the excess distortion probability $\leq \epsilon$. Hence, $R^*(D, \epsilon, 0)$ or $\tilde{R}(D, \epsilon, 0)$ represents the fundamental limit in fixed-length lossy source coding: the minimum coding rate of a fixed-length code under the excess distortion criterion.*

We consider the following problem setting in the asymptotic analysis. Let $\mathcal{X}^n$ and $\mathcal{Y}^n$ be the $n$-th Cartesian product of $\mathcal{X}$ and $\mathcal{Y}$, respectively. Let $X^n$ be a random variable taking a value in $\mathcal{X}^n$ and $x^n$ be a realization of $X^n$. Also, let $\mathbf{X} = \{X^n\}_{n=1}^{\infty}$ denote a general source. The joint probability distribution of $X^n$ is denoted as $P_{X^n}$ and the joint probability mass function of $X^n$ is denoted as $p_{X^n}(x^n)$. A distortion measure $d_n$ is defined as $d_n : \mathcal{X}^n \times \mathcal{Y}^n \to [0, +\infty)$. An encoder $f_n : \mathcal{X}^n \to \{0,1\}^\star$ is produces a non-prefix code and possibly a stochastic code. A decoder $g_n : \{0,1\}^\star \to \mathcal{Y}^n$ is a deterministic code.

An $(n, D, R, \epsilon, \delta)$ code is defined as follows.

**Definition 5.2.3** *Given $D, R \geq 0$ and $\epsilon, \delta \in [0, 1)$, a code $(f_n, g_n)$ satisfying*

$$\mathbb{P}\left[\frac{1}{n} d_n(X^n, g_n(f_n(X^n))) > D\right] \leq \epsilon, \tag{5.5}$$

$$\mathbb{P}\left[\frac{1}{n} \ell(f_n(X^n)) > R\right] \leq \delta \tag{5.6}$$

*is called an $(n, D, R, \epsilon, \delta)$ code.*

The asymptotic fundamental limit is the following minimum threshold.

**Definition 5.2.4** *Given $D \geq 0$ and $\epsilon, \delta \in [0, 1)$, $R(D, \epsilon, \delta|\mathbf{X})$ is the infimum of the threshold $R$ such that there exists an $(n, D, R, \epsilon, \delta)$ code for all $n \geq n_0$ with some $n_0 > 0$.*

Furthermore, another asymptotic fundamental limit is defined as follows.

**Definition 5.2.5** *Given $D \geq 0$ and $\epsilon, \delta \in [0, 1)$, $\hat{R}(D, \epsilon, \delta|\mathbf{X})$ is the infimum of the threshold $R$ such that there exists a code $(f_n, g_n)$ satisfying*

$$\limsup_{n\to\infty} \mathbb{P}\left[\frac{1}{n} d_n(X^n, g_n(f_n(X^n))) > D\right] \leq \epsilon, \tag{5.7}$$

$$\limsup_{n\to\infty} \mathbb{P}\left[\frac{1}{n} \ell(f_n(X^n)) > R\right] \leq \delta. \tag{5.8}$$

## 5.3 Related previous work

Yagi and Nomura [49] have derived the results about $\hat{R}(D, 0, \delta|\mathbf{X})$ and $\hat{R}(D, \epsilon, 0|\mathbf{X})$. Before showing their results, we define some notations and the quantities $\overline{I}_\gamma(\mathbf{X}; \mathbf{Y})$ and $\overline{D}_\gamma(\mathbf{X}; \mathbf{Y})$.

First, let a general source $\mathbf{X} = \{X^n\}_{n=1}^{\infty}$. Let another general source be $\mathbf{Y} = \{Y^n\}_{n=1}^{\infty}$ taking values in $\{\mathcal{Y}^n\}_{n=1}^{\infty}$. A joint probability mass function of $Y^n$ is denoted as $p_{Y^n}(x^n)$. Further, $p_{Y^n|X^n}(y^n|x^n)$ denotes a conditional probability mass function of $Y^n$ given $X^n$. Then, the quantities $\overline{I}_{\gamma}(\mathbf{X};\mathbf{Y})$ and $\overline{D}_{\gamma}(\mathbf{X};\mathbf{Y})$ are defined as follows.

**Definition 5.3.1** *For any* $\gamma \in [0,1)$*,*

$$\overline{I}_{\gamma}(\mathbf{X};\mathbf{Y}) := \inf\{a : F(a \mid \mathbf{X},\mathbf{Y}) \le \gamma\}, \tag{5.9}$$

*where*

$$F(a \mid \mathbf{X},\mathbf{Y}) := \limsup_{n\to\infty} \mathbb{P}\left[\frac{1}{n}\log\frac{p_{Y^n|X^n}(Y^n|X^n)}{p_{Y^n}(Y^n)} > a\right]. \tag{5.10}$$

*Moreover, for any* $\gamma \in [0,1)$*,*

$$\overline{D}_{\gamma}(\mathbf{X};\mathbf{Y}) := \inf\{a : J(a \mid \mathbf{X},\mathbf{Y}) \le \gamma\}, \tag{5.11}$$

*where*

$$J(a \mid \mathbf{X},\mathbf{Y}) := \limsup_{n\to\infty} \mathbb{P}\left[\frac{1}{n}d_n(X^n,Y^n) > a\right]. \tag{5.12}$$

*Further,*

$$\overline{I}(\mathbf{X};\mathbf{Y}) := \overline{I}_0(\mathbf{X};\mathbf{Y}) \tag{5.13}$$
$$\overline{D}(\mathbf{X};\mathbf{Y}) := \overline{D}_0(\mathbf{X};\mathbf{Y}). \tag{5.14}$$

The result in [49] is stated in the next theorem.

**Theorem 5.3.1 ([49])** *For any* $D \ge 0$ *and* $\delta \in [0,1)$*,* $\hat{R}(D,0,\delta|\mathbf{X})$ *is given by*

$$\hat{R}(D,0,\delta|\mathbf{X}) = \inf_{\mathbf{Y}:\overline{D}(\mathbf{X},\mathbf{Y})\le D} \overline{I}_{\delta}(\mathbf{X};\mathbf{Y}). \tag{5.15}$$

*Further, for any* $D \ge 0$ *and* $\epsilon \in [0,1)$*,* $\hat{R}(D,\epsilon,0|\mathbf{X})$ *is given by*

$$\hat{R}(D,\epsilon,0|\mathbf{X}) = \inf_{\mathbf{Y}:\overline{D}_{\epsilon}(\mathbf{X},\mathbf{Y})\le D} \overline{I}(\mathbf{X};\mathbf{Y}). \tag{5.16}$$

## 5.4 Smooth max entropy-based quantity

One of the useful properties of the smooth max entropy[1] is *Schur concavity*. This property is used in the proof of the achievability result in the main theorem. The notion of *majorization* is first reviewed before the definition of a Schur concave function.

[1]The smooth max entropy is defined as in (3.15).

**Definition 5.4.1** *Let $\mathbb{R}_+$ be the set of non-negative real numbers. Further, let $m$ be a positive integer and $\mathbb{R}_+^m$ be the $m$-th Cartesian product of $\mathbb{R}_+$. Suppose that $\mathbf{x} = (x_1, \ldots, x_m) \in \mathbb{R}_+^m$ and $\mathbf{y} = (y_1, \ldots, y_m) \in \mathbb{R}_+^m$ satisfy*

$$x_i \geq x_{i+1}, \quad y_i \geq y_{i+1} \quad (i = 1, 2, \ldots, m-1). \tag{5.17}$$

*For $k = 1, \ldots, m-1$, if $\mathbf{x} \in \mathbb{R}_+^m$ and $\mathbf{y} \in \mathbb{R}_+^m$ satisfy*

$$\sum_{i=1}^{k} x_i \leq \sum_{i=1}^{k} y_i \quad \text{and} \quad \sum_{i=1}^{m} x_i = \sum_{i=1}^{m} y_i, \tag{5.18}$$

*then we say that $\mathbf{y}$ majorizes $\mathbf{x}$ (it is denoted as $\mathbf{x} \prec \mathbf{y}$ in this dissertation).*

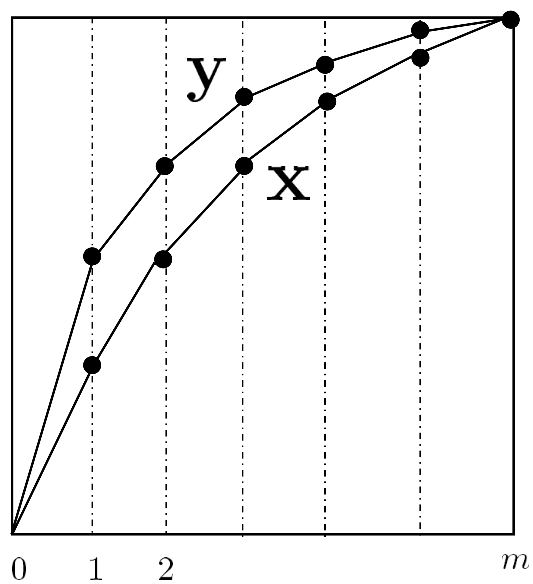


Figure 5.1: The illustration of $\mathbf{x} \prec \mathbf{y}$

Setting $1, \ldots, m$ as the horizontal axis and $\sum_{i=1}^{k} x_i, \sum_{i=1}^{k} y_i$ as the vertical axis, we can draw $\mathbf{x} \prec \mathbf{y}$ as in Fig. 5.1.

Schur concave functions are defined as the following definition.

**Definition 5.4.2** *A function $h(\cdot) : \mathbb{R}_+^m \to \mathbb{R}$ is said to be a Schur concave function if*

$$h(\mathbf{y}) \leq h(\mathbf{x}) \tag{5.19}$$

*for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}_+^m$ satisfying $\mathbf{x} \prec \mathbf{y}$.*

The definitions of the smooth max entropy and Schur concave functions indicate that the smooth max entropy is a Schur concave function[2].

Next, a new quantity is introduced based on the smooth max entropy. This quantity plays a significant role in producing the main results.

**Definition 5.4.3** *Given $D \geq 0$ and $\epsilon, \delta \in [0, 1)$, $G^{\delta}_{D,\epsilon}(X)$ is defined as*

$$G^{\delta}_{D,\epsilon}(X) := \min_{\substack{P_{Y|X}: \\ \mathbb{P}[d(X,Y)>D]\leq\epsilon}} H^{\delta}(Y) \tag{5.20}$$

$$= \min_{\substack{P_{Y|X}: \\ \mathbb{P}[d(X,Y)>D]\leq\epsilon}} \min_{\substack{\mathcal{W}\subset\mathcal{Y}: \\ \mathbb{P}[Y\in\mathcal{W}]\geq 1-\delta}} \log|\mathcal{W}| \tag{5.21}$$

*where $P_{Y|X}$ denotes a conditional probability distribution of $Y$ given $X$.*

**Remark 5.4.1** *For a given $D \geq 0$ and $\epsilon \in [0, 1)$, suppose that the following inequality holds:*

$$\mathbb{P}\left[\inf_{y\in\mathcal{Y}} d(X, y) > D\right] > \epsilon \tag{5.22}$$

*In this case, there are no codes whose excess distortion probability is less than or equal to $\epsilon$. Conversely, (5.22) holds if such codes do not exist for given $D$ and $\epsilon$. In such a case, $R^*(D, \epsilon, \delta)$ and $\tilde{R}(D, \epsilon, \delta)$ are defined as $R^*(D, \epsilon, \delta) = +\infty$ and $\tilde{R}(D, \epsilon, \delta) = +\infty$. Further, if (5.22) holds, $G^{\delta}_{D,\epsilon}(X)$ is defined as $G^{\delta}_{D,\epsilon}(X) = +\infty$. This is because there is no conditional probability distribution $P_{Y|X}$ on $\mathcal{Y}$ satisfying $\mathbb{P}[d(X, Y) > D] \leq \epsilon$.*

## 5.5 Non-asymptotic coding theorem for stochastic codes

The next lemma shows the achievability result on $R$ of a $(D, R, \epsilon, \delta)$ code. Note that the proof of this lemma is parallel with that of Lemma 4.4.1 (see the discussion in Section 5.8 for details).

**Lemma 5.5.1** *Assume that $G^{\delta}_{D,\epsilon}(X) < +\infty$. For any $D \geq 0$ and $\epsilon, \delta \in [0, 1)$, there exists a $(D, R, \epsilon, \delta)$ code such that*

$$R = \lfloor G^{\delta}_{D,\epsilon}(X) \rfloor. \tag{5.23}$$

(**Proof**) First, some notations are defined.

- For $y \in \mathcal{Y}$ and $D \geq 0$, the $D$-ball $\mathcal{B}_D(y)$ is defined as

$$\mathcal{B}_D(y) := \{x \in \mathcal{X} : d(x, y) \leq D\}. \tag{5.24}$$

The illustration of this $D$-ball $\mathcal{B}_D(y)$ is in Fig. 5.2.

[2]By using the notion of majorization, the previous study [16] has shown that the smooth Rényi entropy of order $\alpha$ is a Schur concave function for $0 \leq \alpha < 1$ and a Schur convex function for $\alpha > 1$.

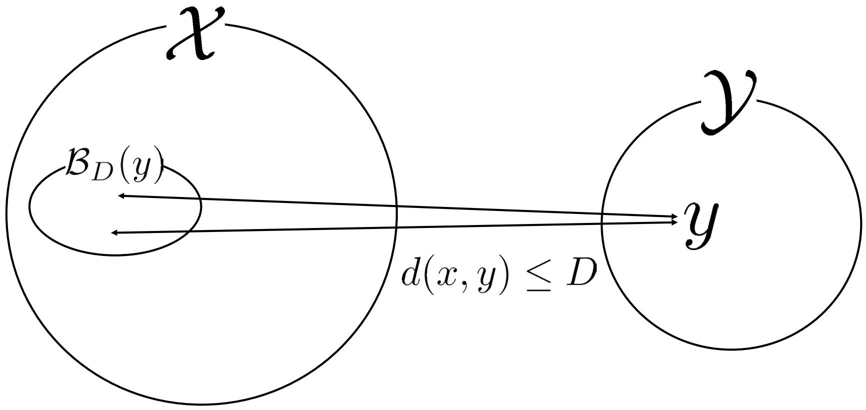


Figure 5.2: The $D$-ball $\mathcal{B}_D(y)$

- The following procedure[3] defines $y_i$ $(i = 1, 2, \cdots)$. Let $y_1$ be defined as

$$y_1 := \arg\max_{y \in \mathcal{Y}} \mathbb{P}[X \in \mathcal{B}_D(y)], \tag{5.25}$$

and for $i = 2, 3, \cdots$, let $y_i$ be defined as

$$y_i := \arg\max_{y \in \mathcal{Y}} \mathbb{P}\left[X \in \mathcal{B}_D(y) \setminus \bigcup_{j=1}^{i-1} \mathcal{B}_D(y_j)\right]. \tag{5.26}$$

- For $i = 1, 2, \ldots$,

$$\mathcal{A}_D(y_i) := \mathcal{B}_D(y_i) \setminus \bigcup_{j=1}^{i-1} \mathcal{B}_D(y_j). \tag{5.27}$$

From the definition, it holds that

$$\bigcup_{j=1}^{i} \mathcal{A}_D(y_j) = \bigcup_{j=1}^{i} \mathcal{B}_D(y_j) \quad (i \geq 1), \tag{5.28}$$

$$\mathcal{A}_D(y_i) \cap \mathcal{A}_D(y_j) = \emptyset \quad (\forall i \neq j), \tag{5.29}$$

$$\mathbb{P}[X \in \mathcal{A}_D(y_1)] \geq \mathbb{P}[X \in \mathcal{A}_D(y_2)] \geq \cdots . \tag{5.30}$$

[3] In this paper, we assume that $\mathcal{X}$ and $\mathcal{Y}$ are finite sets. However, we can assume countably infinite $\mathcal{X}$ and $\mathcal{Y}$ if this operation is admitted for countably infinite $\mathcal{X}$ and $\mathcal{Y}$.

- If $\epsilon + \delta < 1$, let $i^* \geq 1$ be the integer satisfying

$$\sum_{i=1}^{i^*-1} \mathbb{P}[X \in \mathcal{A}_D(y_i)] < 1 - \epsilon - \delta, \tag{5.31}$$

$$\sum_{i=1}^{i^*} \mathbb{P}[X \in \mathcal{A}_D(y_i)] \geq 1 - \epsilon - \delta. \tag{5.32}$$

  If $\epsilon + \delta \geq 1$, then $i^* := 1$.

- Let $k^* \geq 1$ be the integer satisfying

$$\sum_{i=1}^{k^*-1} \mathbb{P}[X \in \mathcal{A}_D(y_i)] < 1 - \epsilon, \tag{5.33}$$

$$\sum_{i=1}^{k^*} \mathbb{P}[X \in \mathcal{A}_D(y_i)] \geq 1 - \epsilon. \tag{5.34}$$

  From this definition, it holds that $k^* \geq i^*$.

- Let $\alpha$ and $\beta$ be defined as

$$\alpha := \sum_{i=1}^{k^*-1} \mathbb{P}[X \in \mathcal{A}_D(y_i)], \tag{5.35}$$

$$\beta := 1 - \epsilon - \alpha. \tag{5.36}$$

- Let $w_i$ be the $i$-th binary string in $\{0,1\}^\star$ in the increasing order of the length and ties are arbitrarily broken. For example, $w_1 = \lambda, w_2 = 0, w_3 = 1, w_4 = 00, w_5 = 01$, etc.

The encoder $f : \mathcal{X} \to \{0,1\}^\star$ and the decoder $g : \{0,1\}^\star \to \mathcal{Y}$ are constructed as follows (see Fig. 5.3).

**[Encoder]**

1) For $x \in \mathcal{A}_D(y_i)$ $(i = 1, \ldots, k^* - 1)$, set $f(x) = w_i$.

2) For $x \in \mathcal{A}_D(y_{k^*})$, set[4]

$$f(x) = \begin{cases} w_{k^*} & \text{with prob. } \frac{\beta}{\mathbb{P}[X \in \mathcal{A}_D(y_{k^*})]}, \\ w_1 & \text{with prob. } 1 - \frac{\beta}{\mathbb{P}[X \in \mathcal{A}_D(y_{k^*})]}. \end{cases} \tag{5.37}$$

3) For $x \notin \bigcup_{i=1}^{k^*} \mathcal{A}_D(y_i)$, set $f(x) = w_1$.

[4]Note that we have $\Pr\{X \in \mathcal{A}_D(y_{k^*})\} \geq \beta$ from (5.34).

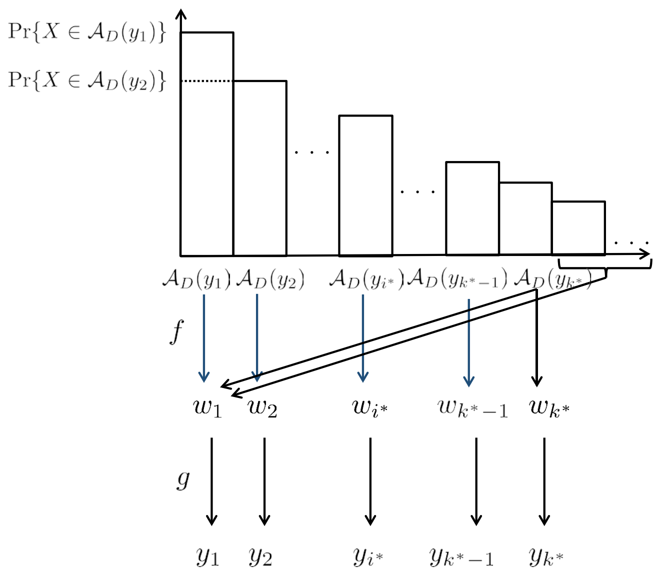


Figure 5.3: Construction of the encoder and the decoder in lossy compression

**[Decoder]** Set $g(w_i) = y_i$ $(i = 1, \ldots, k^*)$.

Now, the excess distortion probability is evaluated. It holds that $d(x, g(f(x))) \leq D$ for $x \in \mathcal{A}_D(y_i)$ $(i = 1, \ldots, k^* - 1)$ since $g(f(x)) = y_i$. Furthermore, $d(x, g(f(x))) \leq D$ holds with probability $\beta / \mathbb{P}[X \in \mathcal{A}_D(y_{k^*})]$ for $x \in \mathcal{A}_D(y_{k^*})$. Thus,

$$\mathbb{P}[d(X, g(f(X))) \leq D] = \sum_{i=1}^{k^*-1} \mathbb{P}[X \in \mathcal{A}_D(y_i)] + \mathbb{P}[f(X) = w_{k^*}, X \in \mathcal{A}_D(y_{k^*})] \quad (5.38)$$

$$= \alpha + \beta = 1 - \epsilon. \quad (5.39)$$

Therefore, we have

$$\mathbb{P}[d(X, g(f(X))) > D] = \epsilon. \quad (5.40)$$

Next, the overflow probability is evaluated. The construction of the encoder verifies that $\ell(w_i) = \lfloor \log i \rfloor$ $(i = 1, \ldots, k^*)$. Hence, setting $R = \lfloor \log i^* \rfloor$, we have

$$\mathbb{P}[\ell(f(X)) > R] \le \sum_{i=i^*+1}^{k^*} \mathbb{P}[f(X) = w_i] \tag{5.41}$$

$$= \sum_{i=i^*+1}^{k^*-1} \mathbb{P}[X \in \mathcal{A}_D(y_i)] + \mathbb{P}[f(X) = w_{k^*}, X \in \mathcal{A}_D(y_{k^*})] \tag{5.42}$$

$$= \sum_{i=1}^{k^*-1} \mathbb{P}[X \in \mathcal{A}_D(y_i)] - \sum_{i=1}^{i^*} \mathbb{P}[X \in \mathcal{A}_D(y_i)] + \beta \tag{5.43}$$

$$\le \alpha - (1 - \epsilon - \delta) + \beta \tag{5.44}$$

$$= \delta, \tag{5.45}$$

where the last inequality follows from the definition of $\alpha$ and (5.32); the last equality is due to the definition of $\beta$.

The foregoing result shows that the code $(f, g)$ is a $(D, R, \epsilon, \delta)$ code with $R = \lfloor \log i^* \rfloor$. Thus, if

$$\log i^* = G^{\delta}_{D,\epsilon}(X) \tag{5.46}$$

is shown, the proof of the theorem is completed. First, $Y := g(f(X))$. Notice that

$$p_Y(y_1) \overset{(a)}{=} \mathbb{P}[X \in \mathcal{A}_D(y_1)] + \mathbb{P}\left[X \in \bigcup_{i \ge k^*+1} \mathcal{A}_D(y_i)\right] + \mathbb{P}[f(X) = w_1, X \in \mathcal{A}_D(y_{k^*})] \tag{5.47}$$

$$\overset{(b)}{=} \mathbb{P}[X \in \mathcal{A}_D(y_1)] + \mathbb{P}[d(X, g(f(X))) > D] \tag{5.48}$$

$$\overset{(c)}{=} \mathbb{P}[X \in \mathcal{A}_D(y_1)] + \epsilon, \tag{5.49}$$

$$p_Y(y_i) = \mathbb{P}[X \in \mathcal{A}_D(y_i)] \quad (i = 2, \ldots, k^* - 1). \tag{5.50}$$

where $(a)$ and $(b)$ follow from the definition of the encoder and decoder; $(c)$ is due to (5.40). Then[5],

$$\sum_{i=1}^{i^*-1} p_Y(y_i) = \sum_{i=1}^{i^*-1} \mathbb{P}[X \in \mathcal{A}_D(y_i)] + \epsilon < 1 - \delta, \tag{5.51}$$

$$\sum_{i=1}^{i^*} p_Y(y_i) = \sum_{i=1}^{i^*} \mathbb{P}[X \in \mathcal{A}_D(y_i)] + \epsilon \ge 1 - \delta, \tag{5.52}$$

$$p_Y(y_1) \ge p_Y(y_2) \ge \cdots \ge p_Y(y_{k^*}). \tag{5.53}$$

[5]If $i^* = k^*$, the equality in (5.52) does not hold. However, $\sum_{i=1}^{i^*} p_Y(y_i) \ge 1 - \delta$ is true because $\sum_{i=1}^{i^*} p_Y(y_i) = 1$.

This implies that $\log i^* = H^\delta(Y)$. Hence, if

$$H^\delta(Y) = G^\delta_{D,\epsilon}(X) \tag{5.54}$$

is shown, the desired equation $\log i^* = G^\delta_{D,\epsilon}(X)$ is obtained.

The following lemma is useful to show (5.54).

**Lemma 5.5.2** *If $P_{Y^*}$ which is induced by $P_{Y^*|X}$ satisfying $\mathbb{P}[d(X,Y^*) > D] \le \epsilon$ majorizes any $P_{\tilde{Y}}$ which is induced by $P_{\tilde{Y}|X}$ satisfying $\mathbb{P}[d(X,\tilde{Y}) > D] \le \epsilon$, then $H^\delta(Y^*) = G^\delta_{D,\epsilon}(X)$ holds.*

(**Proof**) This lemma is obtained by combining the definition of $G^\delta_{D,\epsilon}(X)$ and the fact that the smooth max entropy is a Schur concave function. □

In view of the above lemma, we prove that $P_Y$ majorizes any $P_{\tilde{Y}}$ induced by $P_{\tilde{Y}|X}$ satisfying $\mathbb{P}[d(X,\tilde{Y}) > D] \le \epsilon$. To show this fact, we assume the following condition (♠) and show a contradiction.

(♠) *There is a $P_{\tilde{Y}}$ satisfying $\mathbb{P}[d(X,\tilde{Y}) > D] \le \epsilon$ but not being majorized by $P_Y$.*

Let $y_{\pi(1)}$ give the largest $p_{\tilde{Y}}(y)$ in $\mathcal{Y}$, $y_{\pi(2)}$ give the largest $p_{\tilde{Y}}(y)$ in $\mathcal{Y} \setminus \{y_{\pi(1)}\}$, $y_{\pi(3)}$ give the largest $p_{\tilde{Y}}(y)$ in $\mathcal{Y} \setminus \{y_{\pi(1)}, y_{\pi(2)}\}$, etc. That is, $p_{\tilde{Y}}(y_{\pi(1)}) \ge p_{\tilde{Y}}(y_{\pi(2)}) \ge \cdots \ge p_{\tilde{Y}}(y_{\pi(k^*)})$ and $p_{\tilde{Y}}(y_{\pi(k^*)}) \ge p_{\tilde{Y}}(y_{\pi(i)})$ for all $i = k^*+1, k^*+2, \ldots$. The fact that the support of $p_Y$ is $\{1, 2, \ldots, k^*\}$ and the assumption (♠) show that there exists a $1 \le j_0 \le k^* - 1$ satisfying

$$\sum_{i=1}^{j_0} (p_{\tilde{Y}}(y_{\pi(i)}) - p_Y(y_i)) > 0. \tag{5.55}$$

On the other hand, the excess distortion probability under $P_X P_{\tilde{Y}|X}$ is evaluated as

$$\mathbb{P}[d(X,\tilde{Y}) > D]$$

$$\ge \sum_{x\in\mathcal{X}} \sum_{i=1}^{j_0} p_X(x) p_{\tilde{Y}|X}(y_{\pi(i)}|x) I\{d(x, y_{\pi(i)}) > D\} \tag{5.56}$$

$$= \sum_{x\in\mathcal{X}} \sum_{i=1}^{j_0} p_X(x) p_{\tilde{Y}|X}(y_{\pi(i)}|x) - \sum_{x\in\mathcal{X}} \sum_{i=1}^{j_0} p_X(x) p_{\tilde{Y}|X}(y_{\pi(i)}|x) I\{x \in \mathcal{B}_D(y_{\pi(i)})\} \tag{5.57}$$

$$= \sum_{i=1}^{j_0} p_{\tilde{Y}}(y_{\pi(i)}) - \sum_{x\in\mathcal{X}} p_X(x) \sum_{i=1}^{j_0} p_{\tilde{Y}|X}(y_{\pi(i)}|x) I\{x \in \mathcal{B}_D(y_{\pi(i)})\} \tag{5.58}$$

$$\ge \sum_{i=1}^{j_0} p_{\tilde{Y}}(y_{\pi(i)}) - \mathbb{P}\left[X \in \bigcup_{i=1}^{j_0} \mathcal{B}_D(y_{\pi(i)})\right] \tag{5.59}$$

where $I\{\cdot\}$ denotes the indicator function and the last inequality is due to

$$\sum_{i=1}^{j_0} p_{\tilde{Y}|X}(y_{\pi(i)}|x) I\{x \in \mathcal{B}_D(y_{\pi(i)})\} \leq I\left\{x \in \bigcup_{i=1}^{j_0} \mathcal{B}_D(y_{\pi(i)})\right\} \tag{5.60}$$

for all $x \in \mathcal{X}$. For the second term in (5.59), it holds that

$$\mathbb{P}\left[X \in \bigcup_{i=1}^{j_0} \mathcal{B}_D(y_{\pi(i)})\right] \overset{(a)}{\leq} \mathbb{P}\left[X \in \bigcup_{i=1}^{j_0} \mathcal{B}_D(y_i)\right] \tag{5.61}$$

$$\overset{(b)}{=} \sum_{i=1}^{j_0} \mathbb{P}[X \in \mathcal{A}_D(y_i)] \tag{5.62}$$

$$\overset{(c)}{=} \sum_{i=1}^{j_0} p_Y(y_i) - \epsilon. \tag{5.63}$$

where

- $(a)$ follows from the definition of $y_i$,
- $(b)$ follows from (5.28) and (5.29),
- $(c)$ follows from (5.49) and (5.50).

Plugging (5.63) into (5.59) establishes

$$\mathbb{P}[d(X, \tilde{Y}) > D] \geq \sum_{i=1}^{j_0} (p_{\tilde{Y}}(y_{\pi(i)}) - p_Y(y_i)) + \epsilon > \epsilon, \tag{5.64}$$

where the last inequality is due to (5.55). This is a contradiction to the fact that $\mathbb{P}[d(X, \tilde{Y}) > D] \leq \epsilon$. □

**Remark 5.5.1** *The random coding argument is not used to prove the achievability result. Instead, an explicit code construction is given. This is similar to Feinstein's cookie-cutting argument [6]. The constructed code satisfies the properties*[6] *of the optimal code discussed in [20].*

The next lemma shows the converse bound on $R$ of a $(D, R, \epsilon, \delta)$ code.

**Lemma 5.5.3** *For any $D \geq 0$ and $\epsilon, \delta \in [0, 1)$, any $(D, R, \epsilon, \delta)$ code satisfies*

$$R > G_{D,\epsilon}^{\delta}(X) - 1. \tag{5.65}$$

[6] Kostina et al. [20] have studied the optimal variable-length code that achieves the minimum mean codeword length under the constraint of the excess distortion probability. In [20], several properties of the optimal code are shown.

(**Proof**) Fix arbitrary $(D, R, \epsilon, \delta)$ code $(f, g)$, and set $Y := g(f(X))$. The definition of a $(D, R, \epsilon, \delta)$ code gives

$$\mathbb{P}[\ell(f(X)) > R] \leq \delta, \tag{5.66}$$

$$\mathbb{P}[d(X, Y) > D] \leq \epsilon. \tag{5.67}$$

Let $\mathcal{T}$ be defined as

$$\mathcal{T} := \{g(f(x)) \in \mathcal{Y} : x \text{ satisfies } \ell(f(x)) > R\}. \tag{5.68}$$

Then, (5.66) is rewritten as

$$\mathbb{P}[Y \in \mathcal{T}] \leq \delta \tag{5.69}$$

Hence, it holds that

$$\mathbb{P}[Y \in \mathcal{T}^c] \geq 1 - \delta, \tag{5.70}$$

where the superscript "$c$" represents the complement. Inequality (5.70) and the definition of the smooth max entropy establish

$$H^\delta(Y) \leq \log |\mathcal{T}^c|. \tag{5.71}$$

On the other hand, since $\ell(g^{-1}(y)) \leq \lfloor R \rfloor$ for $y \in \mathcal{T}^c$,

$$|\mathcal{T}^c| \leq 1 + 2 + \cdots + 2^{\lfloor R \rfloor} = 2^{\lfloor R \rfloor + 1} - 1 < 2^{R+1}. \tag{5.72}$$

Combining (5.71) and (5.72) yields

$$H^\delta(Y) < R + 1. \tag{5.73}$$

Thus, the second inequality in (5.67) gives

$$G^\delta_{D,\epsilon}(X) < R + 1. \tag{5.74}$$

This completes the proof. □

Combination of Lemmas 5.5.1 and 5.5.3 immediately yields the following result on $R^*(D, \epsilon, \delta)$.

**Theorem 5.5.1** *For any $D \geq 0$ and $\epsilon, \delta \in [0, 1)$, it holds that*

$$G^\delta_{D,\epsilon}(X) - 1 < R^*(D, \epsilon, \delta) \leq \lfloor G^\delta_{D,\epsilon}(X) \rfloor. \tag{5.75}$$

Theorem 5.5.1 shows that the minimum threshold $R^*(D, \epsilon, \delta)$ can be specified within one bit in the interval not greater than $G^\delta_{D,\epsilon}(X)$, regardless of the values $D$, $\epsilon$, and $\delta$. Because we use an explicit construction of good codes rather than the random coding argument in the proof of Lemma 5.5.1, such a result is obtained.

## 5.6 Non-asymptotic coding theorem for deterministic codes

The next lemma shows the achievability result on $R$ of a deterministic $(D, R, \epsilon, \delta)$ code. Note that the proof of this lemma is parallel with that of Lemma 4.5.1 (see the discussion in Section 5.8 for details).

**Lemma 5.6.1** *Assume that $G^{\delta}_{D,\epsilon}(X) < +\infty$. For any $D \geq 0$ and $\epsilon, \delta \in [0, 1)$, there exists a deterministic $(D, R, \epsilon, \delta)$ code such that*

$$R = \left\lfloor G^{\delta}_{D,\epsilon}(X) + \frac{2 \log e}{2^{G^{\delta}_{D,\epsilon}(X)}} \right\rfloor. \tag{5.76}$$

**(Proof)**

First, some notations are defined.

- Let $k^* \geq 1$ be the integer satisfying (5.33) and (5.34).
- Define $\gamma$ as

$$\gamma = 1 - \sum_{i=1}^{k^*} \mathbb{P}[X \in \mathcal{A}_D(y_i)]. \tag{5.77}$$

  Then, it holds that $\gamma \leq \epsilon$.

- Let $j^* \geq 1$ be the integer satisfying

$$\sum_{i=1}^{j^*-1} \mathbb{P}[X \in \mathcal{A}_D(y_i)] < 1 - \gamma - \delta, \tag{5.78}$$

$$\sum_{i=1}^{j^*} \mathbb{P}[X \in \mathcal{A}_D(y_i)] \geq 1 - \gamma - \delta. \tag{5.79}$$

The *deterministic* encoder $f : \mathcal{X} \to \{0, 1\}^\star$ and decoder $g : \{0, 1\}^\star \to \mathcal{Y}$ are constructed as follows.

**[Encoder]**

1) For $x \in \mathcal{A}_D(y_i)$ $(i = 1, \ldots, k^*)$, set $f(x) = w_i$.
2) For $x \notin \bigcup_{i=1}^{k^*} \mathcal{A}_D(y_i)$, set $f(x) = w_1$.

**[Decoder]** Set $g(w_i) = y_i$ $(i = 1, \ldots, k^*)$.

Now, the excess distortion probability is evaluated. The definition of the encoder and decoder gives

$$\mathbb{P}[d(X, g(f(X))) \leq D] = \sum_{i=1}^{k^*} \mathbb{P}[X \in \mathcal{A}_D(y_i)] \tag{5.80}$$

$$\geq 1 - \epsilon. \tag{5.81}$$

Therefore, it holds that

$$\mathbb{P}[d(X, g(f(X))) > D] \leq \epsilon. \tag{5.82}$$

Next, the overflow probability is evaluated. The definition of the encoder yields

$$\mathbb{P}[f(X) = w_1] = \mathbb{P}[X \in \mathcal{A}_D(y_1)] + \gamma, \tag{5.83}$$

$$\mathbb{P}[f(X) = w_i] = \mathbb{P}[X \in \mathcal{A}_D(y_i)] \quad (i = 2, \dots, k^*). \tag{5.84}$$

Setting $R = \lfloor \log \min(j^*, k^*) \rfloor$, it holds that[7]

$$\mathbb{P}\left[\ell(f(X)) > R\right] \leq 1 - \sum_{i=1}^{j^*} \mathbb{P}[f(X) = w_i] \tag{5.85}$$

$$= 1 - \left( \sum_{i=1}^{j^*} \mathbb{P}[X \in \mathcal{A}_D(y_i)] + \gamma \right) \tag{5.86}$$

$$\leq 1 - ((1 - \gamma - \delta) + \gamma) \tag{5.87}$$

$$= \delta, \tag{5.88}$$

where the last inequality is due to (5.79).

Thus, the code $(f, g)$ is a deterministic $(D, R, \epsilon, \delta)$ code with $R = \lfloor \log \min(j^*, k^*) \rfloor$.

Let $i^*$ be the integer satisfying (5.31) and (5.32). Then, the proof of Lemma 5.5.1 gives

$$\log i^* = G^{\delta}_{D,\epsilon}(X). \tag{5.89}$$

Since $\gamma \leq \epsilon$, it is easily verified that $i^* \leq j^*$ and $i^* \leq k^*$. This means that $i^* \leq \min(j^*, k^*)$. If $i^* = \min(j^*, k^*)$, it is obvious that $\min(j^*, k^*) \leq i^* + 2$. Thus, assuming that $i^* < \min(j^*, k^*)$, we shall show that

$$\min(j^*, k^*) \leq i^* + 2. \tag{5.90}$$

This inequality leads to

$$\log \min(j^*, k^*) \leq \log(i^* + 2) \leq \log i^* + \frac{2 \log e}{i^*}, \tag{5.91}$$

[7] If $R = \lfloor \log k^* \rfloor$, then $\mathbb{P}[\ell(f(X)) > R] = 0$.

where the rightmost inequality is due to Taylor's expansion. This yields the desired result (5.76).

In the following, $\min(j^*, k^*) \le i^* + 2$ is shown. The first step to show this result is the following inequality:

$$
\begin{aligned}
&\sum_{i=1}^{j^*} \mathbb{P}[X \in \mathcal{A}_D(y_i)] - \sum_{i=1}^{i^*} \mathbb{P}[X \in \mathcal{A}_D(y_i)] && (5.92)\\
&\quad \overset{(a)}{\le} 1 - \gamma - \delta + \mathbb{P}[X \in \mathcal{A}_D(y_{j^*})] - (1 - \epsilon - \delta) && (5.93)\\
&\quad = \mathbb{P}[X \in \mathcal{A}_D(y_{j^*})] + \epsilon - \gamma && (5.94)\\
&\quad \overset{(b)}{\le} \mathbb{P}[X \in \mathcal{A}_D(y_{j^*})] + \mathbb{P}[X \in \mathcal{A}_D(y_{i^*+1})]. && (5.95)
\end{aligned}
$$

where $(a)$ follows from (5.32) and (5.78); $(b)$ follows from

$$
\begin{aligned}
\epsilon - \gamma &\le \left(1 - \sum_{i=1}^{k^*-1} \mathbb{P}[X \in \mathcal{A}_D(y_i)]\right) - \left(1 - \sum_{i=1}^{k^*} \mathbb{P}[X \in \mathcal{A}_D(y_i)]\right) && (5.96)\\
&= \mathbb{P}[X \in \mathcal{A}_D(y_{k^*})] && (5.97)\\
&\le \mathbb{P}[X \in \mathcal{A}_D(y_{i^*+1})]. && (5.98)
\end{aligned}
$$

Inequality (5.95) is equivalent to

$$
\sum_{i=1}^{j^*-1} \mathbb{P}[X \in \mathcal{A}_D(y_i)] \le \sum_{i=1}^{i^*+1} \mathbb{P}[X \in \mathcal{A}_D(y_i)]. \tag{5.99}
$$

Thus, $j^* - 1 \le i^* + 1$ is obtained. This implies that $\min(j^*, k^*) \le i^* + 2$. □

From Lemma 5.6.1 and the fact that $R^*(D, \epsilon, \delta) \le \tilde{R}(D, \epsilon, \delta)$, the following result on $\tilde{R}(D, \epsilon, \delta)$ is obtained.

**Theorem 5.6.1** *For any $D \ge 0$ and $\epsilon, \delta \in [0, 1)$, it holds that*

$$
G_{D,\epsilon}^{\delta}(X) - 1 < \tilde{R}(D, \epsilon, \delta) \le \left\lfloor G_{D,\epsilon}^{\delta}(X) + \frac{2 \log e}{2^{G_{D,\epsilon}^{\delta}(X)}} \right\rfloor. \tag{5.100}
$$

Theorem 5.6.1 indicates that $\tilde{R}(D, \epsilon, \delta)$ can be specified in the interval within four bits. This is the slightly weaker result than that for stochastic codes (Theorem 5.5.1).

# 5.7 Asymptotic coding theorem

The next theorem gives the characterization of $R(D, \epsilon, \delta | \mathbf{X})$ by the smooth max entropy-based quantity $G_{D,\epsilon}^{\delta}(X^n)$.

**Theorem 5.7.1** *For any $D \geq 0$ and $\epsilon, \delta \in [0,1)$, it holds that*

$$R(D,\epsilon,\delta|\mathbf{X}) = \limsup_{n\to\infty} \frac{1}{n} G^{\delta}_{D,\epsilon}(X^n), \tag{5.101}$$

*where $G^{\delta}_{D,\epsilon}(X^n)$ is defined as*

$$G^{\delta}_{D,\epsilon}(X^n) := \min_{\substack{P_{Y^n|X^n}: \\ \mathbb{P}[d_n(X^n,Y^n)>nD]\leq\epsilon}} H^{\delta}(Y^n). \tag{5.102}$$

(**Proof**) First,

$$R(D,\epsilon,\delta|\mathbf{X}) \leq \limsup_{n\to\infty} \frac{1}{n} G^{\delta}_{D,\epsilon}(X^n) \tag{5.103}$$

is shown. From Lemma 5.6.1, there exists a code $(f_n, g_n)$ satisfying

$$\mathbb{P}\left[\frac{1}{n} d_n(X^n, g_n(f_n(X^n))) > D\right] \leq \epsilon, \tag{5.104}$$

$$\mathbb{P}\left[\frac{1}{n}\ell(f_n(X^n)) > \frac{1}{n}G^{\delta}_{D,\epsilon}(X^n) + \frac{2\log e}{n2^{G^{\delta}_{D,\epsilon}(X^n)}}\right] \leq \delta. \tag{5.105}$$

Fix $\gamma > 0$ arbitrarily. Then, it holds that

$$\frac{1}{n}G^{\delta}_{D,\epsilon}(X^n) \leq \limsup_{n\to\infty}\frac{1}{n}G^{\delta}_{D,\epsilon}(X^n) + \gamma \tag{5.106}$$

and

$$\frac{2\log e}{n2^{G^{\delta}_{D,\epsilon}(X^n)}} \leq \gamma \tag{5.107}$$

for all $n \geq n_0$ with some $n_0 > 0$. Then, from (5.105), we have

$$\mathbb{P}\left[\frac{1}{n}\ell(f_n(X^n)) > \limsup_{n\to\infty}\frac{1}{n}G^{\delta}_{D,\epsilon}(X^n) + 2\gamma\right] \leq \delta \tag{5.108}$$

for all $n \geq n_0$. Thus, from (5.104) and (5.108), $(f_n, g_n)$ is an $(n, D, R, \epsilon, \delta)$ code with

$$R = \limsup_{n\to\infty}\frac{1}{n}G^{\delta}_{D,\epsilon}(X^n) + 2\gamma \tag{5.109}$$

for all $n \geq n_0$. Since $\gamma > 0$ is arbitrary, this indicates the desired inequality (5.103).

Next,

$$R(D,\epsilon,\delta|\mathbf{X}) \geq \limsup_{n\to\infty} \frac{1}{n} G^{\delta}_{D,\epsilon}(X^n) \tag{5.110}$$

is shown. For any $(n, D, R, \epsilon, \delta)$ code, Lemma 5.5.3 gives

$$nR > G^{\delta}_{D,\epsilon}(X^n) - 1. \tag{5.111}$$

Therefore, it holds that

$$R \geq \limsup_{n\to\infty} \frac{1}{n} G^{\delta}_{D,\epsilon}(X^n) \tag{5.112}$$

for any $(n, D, R, \epsilon, \delta)$ code. Hence, we have (5.110). □

Theorem 5.7.1 characterizes the minimum threshold on the overflow probability by the quantity related to the *entropy*. On the other hand, previous works such as [32] and [49] have characterized it by the quantity related to the *mutual information*.

**Remark 5.7.1** *Theorem 5.7.1 shows that, in the asymptotic regime, the restriction to deterministic encoders does not affect the minimum threshold of the overflow probability. On the other hand, Theorems 5.5.1 and 5.6.1 show that, in the non-asymptotic regime, the results on the minimum threshold are different for stochastic and deterministic encoders.*

The next result on $\hat{R}(D, \epsilon, \delta|\mathbf{X})$ is obtained by almost the same proof of Theorem 5.7.1.

**Corollary 5.7.1** *For any $D \geq 0$ and $\epsilon, \delta \in [0, 1)$, it holds that*

$$\hat{R}(D, \epsilon, \delta|\mathbf{X}) = \lim_{\tau\downarrow 0} \limsup_{n\to\infty} \frac{1}{n} G^{\delta+\tau}_{D,\epsilon+\tau}(X^n). \tag{5.113}$$

**Remark 5.7.2** *Similar to Remark 3.7.1, the difference between (5.101) and (5.113) is due to the difference of the asymptotic constraint, i.e., (5.101) is the result under the constraint of (5.5) and (5.6) for all $n \geq n_0$ with some $n_0$, while (5.113) is the result under the constraint of (5.7) and (5.8).*

## 5.8 Discussion and conclusion of this chapter

Some discussions on the main results are described as follows.

1) *Generalization of the proof of the achievability result:* The proof of Lemma 5.5.1 can be viewed as a generalization of the proof of Lemma 4.4.1. To see this, we set $D = 0$, $\mathcal{X} = \mathcal{Y}$, and

$$d(x, y) = I\{x \neq y\} \tag{5.114}$$

for any $x, y \in \mathcal{X}$, where $I\{\cdot\}$ denotes an indicator function. Then, the construction of the encoder in the proof of Lemma 5.5.1 coincides with that in the proof of Lemma 4.4.1. Similarly, the proof of Lemma 5.6.1 can be viewed as a generalization of the proof of Lemma 4.5.1.

2) *Trade-off relationship between the excess distortion probability and the overflow probability:* In Lemma 5.5.1, Eq. (5.23) and (5.46) show that the threshold $R$ is determined by the sum of $\epsilon$ (the tolerated excess distortion probability) and $\delta$ (the tolerated overflow probability). Suppose that $\epsilon + \delta = C$ for some constant $C$. Then, $\epsilon$ and $\delta$ has a trade-off relationship, i.e., if $\epsilon$ is high, then $\delta$ is low and vice versa. Therefore, we see that the overflow probability and the excess distortion probability has a trade-off relationship in the case that the sum of $\epsilon$ and $\delta$ is a constant.

3) *Stochastic code vs. deterministic code:* Comparing the stochastic code in the proof of Lemma 5.5.1 and the deterministic code in the proof of Lemma 5.6.1, it is observed that the only difference between these codes is the treatment of the source symbols in $\mathcal{A}_D(y_{k^*})$. That is, for the stochastic code in the proof of Lemma 5.5.1, randomization is implemented at the source symbols $x \in \mathcal{A}_D(y_{k^*})$.

To summarize this chapter, we considered the problem of variable-length lossy source coding for a general source. First, the smooth max entropy-based quantity was defined. Then, using this quantity, novel non-asymptotic coding theorems were obtained for both stochastic codes and deterministic codes. The explicit code construction was used to prove the achievability results. Finally, asymptotic coding theorems were shown based on the non-asymptotic results. Notice that basic ideas of the proofs in this chapter are based on those in Chapters 3 and 4.

## 5.9 Discussion of Chapters 3–5

To summarize Chapters 3–5, we discuss some properties that the achievability part and the converse part have in common throughout Chapters 3–5.

### 5.9.1 Discussion on the achievability part in Chapters 3–5

In the following, we discuss some properties of the codes in the proofs of the achievability results in Chapters 3–5.

1) *Basic strategy on the construction of the encoder:* The basic strategy to construct the encoder is as follows:

   - For most likely source symbols, we encode these symbols so as *not* to suffer overflow or error (excess distortion).
   - For least likely source symbols, we allow these symbols to suffer overflow or error (excess distortion).

2) *Difference between stochastic codes and deterministic codes:* In Chapters 4 and 5, we considered the stochastic codes and deterministic codes. As pointed out

in Section 4.7 and Section 5.8, the only difference between stochastic and deterministic codes is the treatment of the source symbol $x_{k^*}$ (in Chapter 4) and source symbols in $\mathcal{A}_D(y_{k^*})$ (in Chapter 5). If we implement the randomization for $x_{k^*}$ (or elements in $\mathcal{A}_D(y_{k^*})$),

- we can make the error probability (or the excess distortion probability) exactly $\epsilon$,
- we can make the overflow probability exceeding the threshold $\lfloor H^{\epsilon+\delta}(X) \rfloor$ (or $\lfloor G^{\delta}_{D,\epsilon}(X) \rfloor$) less than or equal to $\delta$.

On the other hand, if the randomization is not used for $x_{k^*}$ (or elements in $\mathcal{A}_D(y_{k^*})$), the error probability (or the excess distortion probability) is less than or equal to $\epsilon$, not exactly $\epsilon$. In this case, to make the overflow probability smaller than $\delta$, we must set the threshold

$$\left\lfloor H^{\epsilon+\delta}(X) + \frac{2 \log e}{2^{H^{\epsilon+\delta}(X)}} \right\rfloor \tag{5.115}$$

(or $\left\lfloor G^{\delta}_{D,\epsilon}(X) + \frac{2 \log e}{2^{G^{\delta}_{D,\epsilon}(X)}} \right\rfloor$), which is slightly greater than $\lfloor H^{\epsilon+\delta}(X) \rfloor$ (or $\lfloor G^{\delta}_{D,\epsilon}(X) \rfloor$).

3) *Relationship between the overflow probability and the error probability (excess distortion probability) and another construction of the encoder:* In Chapters 4 and 5, the fundamental limit on the overflow probability is determined by the sum of $\epsilon$ (the tolerated error or excess distortion probability) and $\delta$ (the tolerated overflow probability). In the proof of the achievability results in Chapters 4 and 5, the encoder was constructed as shown in Fig. 5.4 (see also Fig. 4.1 and Fig. 5.3). In this construction,

   - the least likely source symbols (which correspond the part painted in gray in Fig. 5.4) are coded so as to have error (excess distortion),
   - the source symbols whose indices are greater than $i^*$ and less than $k^*$ (which correspond the part painted in black in Fig. 5.4) are coded so as to overflow.

   However, the fact that the fundamental limit is determined by the sum of $\epsilon$ and $\delta$ indicates that we can construct another encoder as shown in Fig. 5.5. That is, we define $\tilde{k}$ appropriately, and we can construct the encoder whose error probability (or excess distortion probability) is less than or equal to $\epsilon$ and overflow probability is less than or equal to $\delta$ as follows:

   - the source symbols whose indices are greater than $i^*$ and less than $\tilde{k}$ (which correspond the part painted in gray in Fig. 5.5) are coded so as to have error (excess distortion),
   - the least likely source symbols (which correspond the part painted in black in Fig. 5.5) are coded so as to overflow.

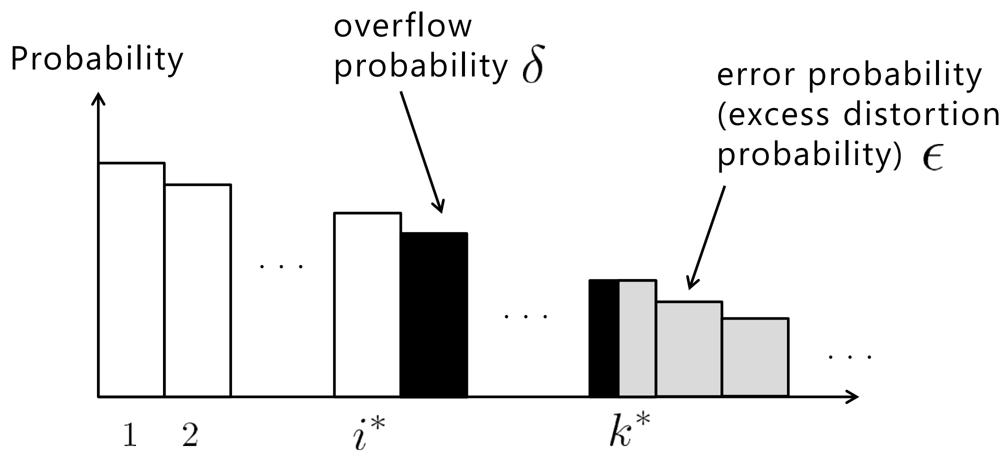


Figure 5.4: Construction of the encoder in Chapters 4 and 5 with error probability (excess distortion probability) $\epsilon$ and overflow probability $\delta$

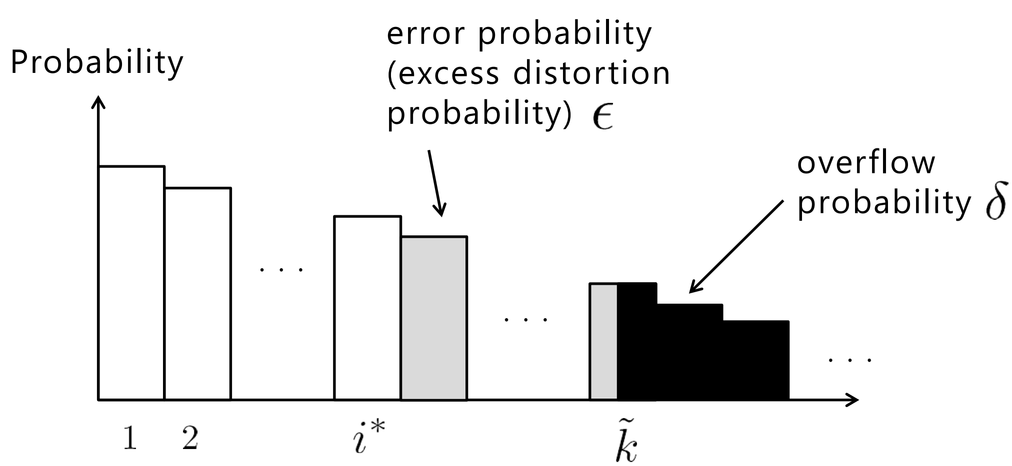


Figure 5.5: Another construction of the encoder with error probability (excess distortion probability) $\epsilon$ and overflow probability $\delta$

### 5.9.2 Discussion on the converse part in Chapters 3–5

In the proofs of the converse results in Chapters 3–5, we evaluated the cardinality of *"good" sets* by using the smooth max entropy (or the quantity defined by the smooth max entropy). That is,

- in Chapter 3, we evaluated the cardinality of *the set of source symbols whose codeword length does not overflow* by using the smooth max entropy (see (3.36) and (3.61));
- in Chapter 4, we analyzed the cardinality of *the set of source symbols which neither overflow nor have error* by using the smooth max entropy (see (4.42));
- in Chapter 5, we examined the cardinality of *the set of source symbols which neither overflow nor have excess distortion* by using the quantity defined by the smooth max entropy (see (5.74)).

# Chapter 6

# Second-order achievable rate region of Slepian-Wolf coding problem

## 6.1 Introduction

Through Chapters 3–5, we have considered variable-length source coding problems and characterized the minimum threshold by using the smooth max entropy. As stated in Section 1.2, the variable-length source coding problem with the overflow probability and the fixed-length source coding problem are closely related. Thus, this fact raises the following question: can we characterize the fundamental limit in fixed-length source coding based on the smooth max entropy? Chapter 6 gives one of the answers for this question.

This chapter deals with the fixed-length Slepian-Wolf coding problem [40]: for two correlated sources, one decoder jointly decodes two codewords encoded by separate encoders. Slepian and Wolf [40] have first clarified the achievable rate region for this problem. After this work, studies have been done to clarify the achievable rate region in various problem settings.

This chapter focuses the achievable rate region of Slepian-Wolf coding problem for a general source. For this problem, based on the *information spectrum methods* [10], Miyake and Kanaya [25] have derived the achievable rate region under the condition that the error probability vanishes. Further, Han [10] has treated the case where a positive error probability is allowed and derived the achievable rate region. These achievable rate regions are said to be the *first-order* achievable rate region. Recently, the *second-order* achievable rate region, a finer evaluation of the achievable rate region than the first-order case, has been analyzed by Nomura and Han [31] based on the information spectrum methods.

Instead of using the information spectrum methods, Uyematsu and Matsuta [46] have analyzed the *first-order* achievable rate region based on the quantity related to the *smooth max entropy* [34] and the *conditional smooth max entropy* [35]. However, this work has not derived the *second-order* achievable rate region.

This chapter extends the results of [46] to the *second-order* case. Further, this chapter shows the relationship between the derived achievable rate region and the rate region defined by the smooth max entropy and the conditional smooth max entropy. Moreover, the relationship is clarified between the two functions: the function which characterizes the second-order achievable rate region in [31] and that in our study.

A map of this chapter and previous studies are shown in Table 6.1.

Table 6.1: The map of Chapter 6 and previous studies

| | information spectrum | smooth max entropy |
|---|---|---|
| first-order achievable rate region | Miyake and Kanaya [25]<br>Han [10] | Uyematsu and Matsuta [46] |
| second-order achievable rate region | Nomura and Han [31] | Chapter 6 |

The organization of this chapter is as follows. Section 6.2 explains the problem formulation. First, an $(n, M_n^{(1)}, M_n^{(2)}, \epsilon_n)$ code is introduced. Second, first-order achievable rate region is defined. Then, second-order achievable rate region is defined. Section 6.3 describes the prior works. Section 6.4 shows the non-asymptotic key lemmas, which play a fundamental role in producing the main results. Section 6.5 derives the main results in this chapter. The proofs of the main results are shown in Section 6.6. Finally, Section 6.7 concludes this chapter.

## 6.2 Problem formulation

### 6.2.1 $(n, M_n^{(1)}, M_n^{(2)}, \epsilon_n)$ code

Let $\mathcal{X}_1, \mathcal{X}_2$ be source alphabets of two correlated sources, where $\mathcal{X}_1, \mathcal{X}_2$ are finite or countably infinite sets. Let $(\mathbf{X}_1, \mathbf{X}_2) = \{(X_1^n, X_2^n)\}_{n=1}^{\infty}$ denote a general correlated source [10], where $(X_1^n, X_2^n)$ is a random variable taking a value in $\mathcal{X}_1^n \times \mathcal{X}_2^n$. Let $(\mathbf{x}_1, \mathbf{x}_2)$ be a realization of a random variable $(X_1^n, X_2^n)$. The probability distribution of $(X_1^n, X_2^n)$ is denoted as $P_{X_1^n X_2^n}$ and the probability mass function of $(X_1^n, X_2^n)$ is denoted as $p_{X_1^n X_2^n}(\mathbf{x}_1, \mathbf{x}_2)$. Further, the conditional probability distribution of $X_i^n$ given $X_j^n$ is denoted as $P_{X_i^n|X_j^n}$ for $(i,j) = (1,2)$ and $(2,1)$. The conditional probability mass function of $X_i^n$ given $X_j^n$ is denoted as $p_{X_i^n|X_j^n}(\mathbf{x}_i|\mathbf{x}_j)$ for $(i,j) = (1,2)$ and $(2,1)$.

Two encoders and a decoder of the Slepian-Wolf coding problem, which are shown in Fig. 6.1, are defined as follows. Encoders $\phi_n^{(1)}$ and $\phi_n^{(2)}$ are defined as

$$\phi_n^{(1)} : \mathcal{X}_1^n \to \mathcal{M}_n^{(1)} := \{1, 2, \ldots, M_n^{(1)}\}, \tag{6.1}$$

$$\phi_n^{(2)} : \mathcal{X}_2^n \to \mathcal{M}_n^{(2)} := \{1, 2, \ldots, M_n^{(2)}\}, \tag{6.2}$$

where $M_n^{(1)}$ and $M_n^{(2)}$ are positive integers. A decoder $\psi_n$ is defined as

$$\psi_n : \mathcal{M}_n^{(1)} \times \mathcal{M}_n^{(2)} \to \mathcal{X}_1^n \times \mathcal{X}_2^n. \tag{6.3}$$

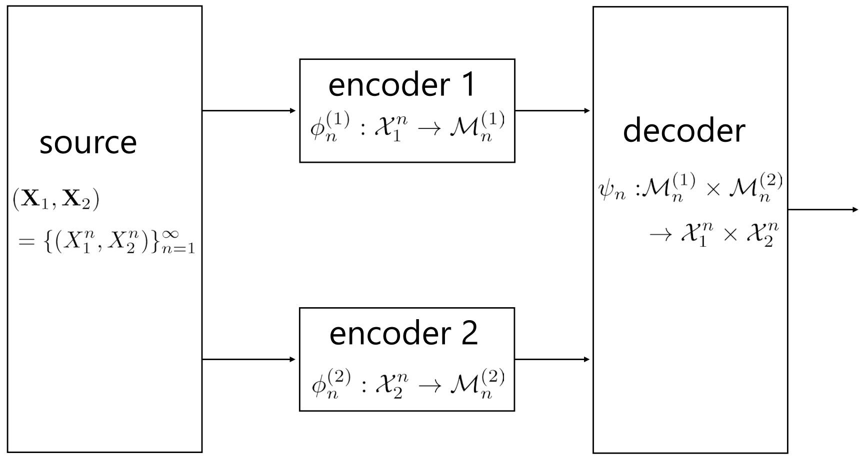


Figure 6.1: Slepian-Wolf source coding

Furthermore, the probability of error $\epsilon_n$ is defined as

$$\epsilon_n = \mathbb{P}[(X_1^n, X_2^n) \neq \psi_n(\phi_n^{(1)}(X_1^n), \phi_n^{(2)}(X_2^n))]. \tag{6.4}$$

Then, a pair of encoders and a decoder $(\phi_n^{(1)}, \phi_n^{(2)}, \psi_n)$ with $\mathcal{M}_n^{(1)}, \mathcal{M}_n^{(2)}$ and $\epsilon_n$ is called an $(n, M_n^{(1)}, M_n^{(2)}, \epsilon_n)$ code.

### 6.2.2 First-order achievable rate region

Let $R_1, R_2$ be non-negative real numbers. The notion of $\epsilon$-achievability is defined as follows.

**Definition 6.2.1 ([10])** *Let $\epsilon \in [0, 1)$. A rate pair $(R_1, R_2)$ is $\epsilon$-achievable if there exists a sequence of $(n, M_n^{(1)}, M_n^{(2)}, \epsilon_n)$ codes satisfying*

$$\limsup_{n\to\infty} \frac{1}{n} \log M_n^{(1)} \leq R_1, \tag{6.5}$$

$$\limsup_{n\to\infty} \frac{1}{n} \log M_n^{(2)} \leq R_2, \tag{6.6}$$

$$\limsup_{n\to\infty} \epsilon_n \leq \epsilon. \tag{6.7}$$

Then, the $\epsilon$-achievable rate region is defined as follows.

**Definition 6.2.2 ([10])** *The $\epsilon$-achievable rate region is defined as*

$$\mathcal{R}(\epsilon|\mathbf{X}_1, \mathbf{X}_2) = \{(R_1, R_2) : (R_1, R_2) \text{ is } \epsilon\text{-achievable}\}. \tag{6.8}$$

The $\epsilon$-achievable rate region is also said to be the *first-order* achievable rate region.

### 6.2.3 Second-order achievable rate region

Let $L_1, L_2$ be non-negative real numbers. The notion of $(a_1, a_2, \epsilon)$-achievability is defined as follows.

**Definition 6.2.3 ([31])** *Let $(a_1, a_2) \in \mathbb{R}^2$ and $\epsilon \in [0, 1)$. A rate pair $(L_1, L_2)$ is $(a_1, a_2, \epsilon)$-achievable if there exists a sequence of $(n, M_n^{(1)}, M_n^{(2)}, \epsilon_n)$ codes satisfying*

$$\limsup_{n\to\infty} \frac{1}{\sqrt{n}} \log \frac{M_n^{(1)}}{e^{na_1}} \le L_1, \tag{6.9}$$

$$\limsup_{n\to\infty} \frac{1}{\sqrt{n}} \log \frac{M_n^{(2)}}{e^{na_2}} \le L_2, \tag{6.10}$$

$$\limsup_{n\to\infty} \epsilon_n \le \epsilon. \tag{6.11}$$

Then, the $(a_1, a_2, \epsilon)$-achievable rate region is defined as follows.

**Definition 6.2.4 ([31])** *The $(a_1, a_2, \epsilon)$-achievable rate region is defined as*

$$\mathcal{L}(a_1, a_2, \epsilon|\mathbf{X}_1, \mathbf{X}_2) = \{(L_1, L_2) : (L_1, L_2) \text{ is } (a_1, a_2, \epsilon)\text{-achievable}\}. \tag{6.12}$$

The $(a_1, a_2, \epsilon)$-achievable rate region is also said to be the *second-order* achievable rate region.

## 6.3 Previous works

### 6.3.1 First-order achievable rate region based on the information spectrum methods

The functions $F_n(R_1, R_2)$ and $F(R_1, R_2)$ are defined as follows [10].

$$\begin{aligned} F_n(R_1, R_2) := \mathbb{P}\Bigg[ &\frac{1}{n} \log \frac{1}{p_{X_1^n|X_2^n}(X_1^n|X_2^n)} \ge R_1 \\ &\text{or} \quad \frac{1}{n} \log \frac{1}{p_{X_2^n|X_1^n}(X_2^n|X_1^n)} \ge R_2 \\ &\text{or} \quad \frac{1}{n} \log \frac{1}{p_{X_1^n X_2^n}(X_1^n, X_2^n)} \ge R_1 + R_2 \Bigg], \end{aligned} \tag{6.13}$$

$$F(R_1, R_2) := \limsup_{n\to\infty} F_n(R_1, R_2). \tag{6.14}$$

The $\epsilon$-achievable rate region is characterized by $F(R_1, R_2)$.

**Theorem 6.3.1 ([10])** *For any $\epsilon \in [0, 1)$, the $\epsilon$-achievable rate region is given by*

$$\mathcal{R}(\epsilon|\mathbf{X}_1, \mathbf{X}_2) = \mathrm{Cl}\left(\{(R_1, R_2) : F(R_1, R_2) \le \epsilon\}\right), \tag{6.15}$$

*where $\mathrm{Cl}(\cdot)$ denotes the closure operation.*

### 6.3.2 Second-order achievable rate region based on the information spectrum methods

For $(a_1, a_2) \in \mathbb{R}^2$, the functions $F_n(L_1, L_2|a_1, a_2)$ and $F(L_1, L_2|a_1, a_2)$ are defined as follows [31].

$$
\begin{aligned}
&F_n(L_1, L_2|a_1, a_2) \\
&= \mathbb{P}\Bigg[\frac{-\log p_{X_1^n|X_2^n}(X_1^n|X_2^n) - na_1}{\sqrt{n}} \geq L_1 \\
&\qquad \text{or} \quad \frac{-\log p_{X_2^n|X_1^n}(X_2^n|X_1^n) - na_2}{\sqrt{n}} \geq L_2 \\
&\qquad \text{or} \quad \frac{-\log p_{X_1^n X_2^n}(X_1^n, X_2^n) - n(a_1 + a_2)}{\sqrt{n}} \geq L_1 + L_2\Bigg], \qquad (6.16) \\
&F(L_1, L_2|a_1, a_2) := \limsup_{n\to\infty} F_n(L_1, L_2|a_1, a_2). \qquad (6.17)
\end{aligned}
$$

The $(a_1, a_2, \epsilon)$-achievable rate region is characterized by $F(L_1, L_2|a_1, a_2)$.

**Theorem 6.3.2 ([31])** *For any $(a_1, a_2) \in \mathbb{R}^2$ and $\epsilon \in [0, 1)$, the $(a_1, a_2, \epsilon)$-achievable rate region is given by*

$$
\mathcal{L}(a_1, a_2, \epsilon|\mathbf{X}_1, \mathbf{X}_2) = \mathrm{Cl}\left(\{(L_1, L_2) : F(L_1, L_2|a_1, a_2) \leq \epsilon\}\right).
$$

### 6.3.3 First-order achievable rate region based on the smooth max entropy

First, some definitions in [46] are introduced. Let $\mathcal{S}(M_n^{(1)}, M_n^{(2)})$ be defined as

$$
\begin{aligned}
\mathcal{S}(M_n^{(1)}, M_n^{(2)}) = \{T_n \subset \mathcal{X}_1^n \times \mathcal{X}_2^n : |T_n| \leq M_n^{(1)} M_n^{(2)} \\
\text{and} \quad \max_{\mathbf{x}_2 \in \mathcal{X}_2^n} |\{\mathbf{x}_1 \in \mathcal{X}_1^n : (\mathbf{x}_1, \mathbf{x}_2) \in T_n\}| \leq M_n^{(1)} \\
\text{and} \quad \max_{\mathbf{x}_1 \in \mathcal{X}_1^n} |\{\mathbf{x}_2 \in \mathcal{X}_2^n : (\mathbf{x}_1, \mathbf{x}_2) \in T_n\}| \leq M_n^{(2)}\}, \qquad (6.18)
\end{aligned}
$$

where $|\cdot|$ represents the cardinality of a set. By using $\mathcal{S}(\cdot, \cdot)$, $G_n(R_1, R_2)$ and $G(R_1, R_2)$ are defined as

$$
G_n(R_1, R_2) := 1 - \max_{T_n \in \mathcal{S}(e^{nR_1}, e^{nR_2})} \mathbb{P}[(X_1^n, X_2^n) \in T_n], \qquad (6.19)
$$

$$
G(R_1, R_2) := \limsup_{n\to\infty} G_n(R_1, R_2). \qquad (6.20)
$$

The previous study [46] has shown that the $\epsilon$-achievable rate region is characterized by $G(R_1, R_2)$.

**Theorem 6.3.3 ([46])** *For any $\epsilon \in [0,1)$, the $\epsilon$-achievable rate region is given by*

$$\mathcal{R}(\epsilon|\mathbf{X}_1, \mathbf{X}_2) = \mathrm{Cl}(\{(R_1, R_2) : G(R_1, R_2) \le \epsilon\}). \tag{6.21}$$

The previous study [46] has also investigated the relationship between the $\epsilon$-achievable rate region $\mathcal{R}(\epsilon|\mathbf{X}_1, \mathbf{X}_2)$ and the rate region defined by the smooth max entropy and the conditional smooth max entropy. Before stating this result, we review the definitions of the smooth max entropy and the conditional smooth max entropy.

The smooth max entropy has first introduced by Renner and Wolf [34] and the conditional smooth max entropy by Renner and Wolf [35]. Later, Uyematsu [43] has shown that the smooth max entropy can be defined as the following definition.

**Definition 6.3.1 ([34], [35], [43])** *For $\epsilon \in [0,1)$, the smooth max entropy $H^\epsilon(X_1^n, X_2^n)$ is defined as*

$$H^\epsilon(X_1^n, X_2^n) = \min_{\substack{\mathcal{A}_n \subset \mathcal{X}_1^n \times \mathcal{X}_2^n: \\ \mathbb{P}[(X_1^n, X_2^n) \in \mathcal{A}_n] \ge 1-\epsilon}} \log |\mathcal{A}_n|. \tag{6.22}$$

*The conditional smooth max entropy $H^\epsilon(X_i^n|X_j^n)$ for $(i,j)$ = (1,2) and (2,1) is defined as*

$$H^\epsilon(X_i^n|X_j^n) = \min_{\substack{\mathcal{A}_n \subset \mathcal{X}_1^n \times \mathcal{X}_2^n: \\ \mathbb{P}[(X_1^n, X_2^n) \in \mathcal{A}_n] \ge 1-\epsilon}} \log \max_{\mathbf{x}_j \in \mathcal{X}_j^n} |\{\mathbf{x}_i \in \mathcal{X}_i^n : (\mathbf{x}_1, \mathbf{x}_2) \in \mathcal{A}_n\}|. \tag{6.23}$$

Then, for any $\epsilon \in [0,1)$, the rate region defined by the smooth max entropy and the conditional smooth max entropy is given by

$$\begin{aligned}\tilde{\mathcal{R}}(\epsilon|\mathbf{X}_1, \mathbf{X}_2) = \Bigg\{ &(R_1, R_2) : R_1 \ge \limsup_{n\to\infty} \frac{1}{n} H^\epsilon(X_1^n|X_2^n) \\ &\text{and} \quad R_2 \ge \limsup_{n\to\infty} \frac{1}{n} H^\epsilon(X_2^n|X_1^n) \\ &\text{and} \quad R_1 + R_2 \ge \limsup_{n\to\infty} \frac{1}{n} H^\epsilon(X_1^n, X_2^n) \Bigg\}. \end{aligned} \tag{6.24}$$

The relationship between $\tilde{\mathcal{R}}(\epsilon|\mathbf{X}_1, \mathbf{X}_2)$ and $\mathcal{R}(\epsilon|\mathbf{X}_1, \mathbf{X}_2)$ is given as follows.

**Theorem 6.3.4 ([46])** *For any $\epsilon \in (0,1)$, it holds that*

$$\tilde{\mathcal{R}}(\epsilon/3|\mathbf{X}_1, \mathbf{X}_2) \subset \mathcal{R}(\epsilon|\mathbf{X}_1, \mathbf{X}_2) \subset \lim_{\delta \downarrow 0} \tilde{\mathcal{R}}(\epsilon + \delta|\mathbf{X}_1, \mathbf{X}_2). \tag{6.25}$$

The study [10] has characterized the $\epsilon$-achievable rate region by the function $F_n(R_1, R_2)$. On the other hand, the study [46] has characterized the $\epsilon$-achievable rate region by the function $G_n(R_1, R_2)$. The relationship between $F_n(R_1, R_2)$ and $G_n(R_1, R_2)$ is given as follows.

**Theorem 6.3.5 ([46])** *For any $\gamma > 0$ and $n = 1, 2, \ldots$, we have*

$$F_n(R_1 + \gamma, R_2 + \gamma) - 3e^{-n\gamma} \le G_n(R_1, R_2) \le F_n(R_1, R_2). \tag{6.26}$$

## 6.4 Non-asymptotic key lemmas

The following lemmas, which are valid for finite blocklength $n$, play a significant role to prove main theorems in this chapter.

**Lemma 6.4.1** *Let $M_n^{(1)}, M_n^{(2)}$ be arbitrary positive integers. Then, for any $\gamma > 0$ and $n = 1, 2, \ldots$, there exists an $(n, M_n^{(1)}, M_n^{(2)}, \epsilon_n)$ code such that*

$$\epsilon_n \leq 1 - \max_{T_n \in \mathcal{S}(M_n^{(1)} e^{-\sqrt{n}\gamma}, M_n^{(2)} e^{-\sqrt{n}\gamma})} \mathbb{P}[(X_1^n, X_2^n) \in T_n] + 3e^{-\sqrt{n}\gamma}. \tag{6.27}$$

(**Proof**) The proof of this lemma is similar to that of [46, Lemma 3]. First, let $\mathcal{T}_n$ be any element in $\mathcal{S}(M_n^{(1)} e^{-\sqrt{n}\gamma}, M_n^{(2)} e^{-\sqrt{n}\gamma})$. Then, the random coding argument is used.

1. **Random coding**: For $\mathbf{x}_1 \in \mathcal{X}_1^n$, $i \in \mathcal{M}_n^{(1)}$ is generated according to the uniform distribution over $\mathcal{M}_n^{(1)}$ and set $\phi_n^{(1)}(\mathbf{x}_1) = i$. For $\mathbf{x}_2 \in \mathcal{X}_2^n$, $j \in \mathcal{M}_n^{(2)}$ is generated according to the uniform distribution over $\mathcal{M}_n^{(2)}$ and set $\phi_n^{(2)}(\mathbf{x}_2) = j$.

2. **Construction of the decoder**: A decoder receives $(i, j) \in \mathcal{M}_n^{(1)} \times \mathcal{M}_n^{(2)}$. Then, we construct the decoder as follows. If there exists a unique $(\mathbf{x}_1, \mathbf{x}_2) \in \mathcal{T}_n$ satisfying
$$\phi_n^{(1)}(\mathbf{x}_1) = i \tag{6.28}$$
and
$$\phi_n^{(2)}(\mathbf{x}_2) = j, \tag{6.29}$$
set $\psi_n$ as
$$\psi_n(i, j) = (\mathbf{x}_1, \mathbf{x}_2). \tag{6.30}$$
If there exists no such pair or if there exist more than one such pair, set $\psi_n$ as an any specified element in $\mathcal{X}_1^n \times \mathcal{X}_2^n$.

3. **Evaluation of the error probability**: The analysis of the error probability is the same as that of [10], and it holds that
$$E[\epsilon_n] \leq 1 - \mathbb{P}[(X_1^n, X_2^n) \in \mathcal{T}_n] + 3e^{-\sqrt{n}\gamma}, \tag{6.31}$$
where $E[\cdot]$ is the average over the above random coding. This means that there exists at least one $(n, M_n^{(1)}, M_n^{(2)}, \epsilon_n)$ code satisfying
$$\epsilon_n \leq 1 - \mathbb{P}[(X_1^n, X_2^n) \in \mathcal{T}_n] + 3e^{-\sqrt{n}\gamma}. \tag{6.32}$$
Therefore, taking the maximum of $\mathbb{P}[(X_1^n, X_2^n) \in \mathcal{T}_n]$ over $\mathcal{T}_n$, we conclude that there exists an $(n, M_n^{(1)}, M_n^{(2)}, \epsilon_n)$ code satisfying
$$\epsilon_n \leq 1 - \max_{\mathcal{T}_n \in \mathcal{S}(M_n^{(1)} e^{-\sqrt{n}\gamma}, M_n^{(2)} e^{-\sqrt{n}\gamma})} \mathbb{P}[(X_1^n, X_2^n) \in \mathcal{T}_n] + 3e^{-\sqrt{n}\gamma}. \tag{6.33}$$

Hence, we complete the proof. $\square$

The next lemma is Lemma 2 in [46].

**Lemma 6.4.2 ([46])** *For any* $(n, M_n^{(1)}, M_n^{(2)}, \epsilon_n)$ *code, we have*

$$\epsilon_n \geq 1 - \max_{T_n \in \mathcal{S}(M_n^{(1)}, M_n^{(2)})} \mathbb{P}[(X_1^n, X_2^n) \in \mathcal{T}_n]. \tag{6.34}$$

## 6.5 Main results

First, the functions $G_n(L_1, L_2|a_1, a_2)$ and $G(L_1, L_2|a_1, a_2)$ are defined. To characterize the second-order achievable rate region, these functions play a fundamental role.

**Definition 6.5.1** *For* $(a_1, a_2) \in \mathbb{R}^2$*, the functions* $G_n(L_1, L_2|a_1, a_2)$ *and* $G(L_1, L_2|a_1, a_2)$ *are defined as*

$$G_n(L_1, L_2|a_1, a_2) := 1 - \max_{T_n \in \mathcal{S}(e^{na_1 + \sqrt{n}L_1}, e^{na_2 + \sqrt{n}L_2})} \mathbb{P}[(X_1^n, X_2^n) \in \mathcal{T}_n], \tag{6.35}$$

$$G(L_1, L_2|a_1, a_2) := \limsup_{n \to \infty} G_n(L_1, L_2|a_1, a_2). \tag{6.36}$$

The following characterization of the $(a_1, a_2, \epsilon)$-achievable rate region is one of the main results.

**Theorem 6.5.1** *For any* $\epsilon \in [0, 1)$ *and* $(a_1, a_2) \in \mathbb{R}^2$*, the* $(a_1, a_2, \epsilon)$*-achievable rate region is given by*

$$\mathcal{L}(a_1, a_2, \epsilon|\mathbf{X}_1, \mathbf{X}_2) = \mathrm{Cl}(\{(L_1, L_2) : G(L_1, L_2|a_1, a_2) \leq \epsilon\}). \tag{6.37}$$

The rate region by the smooth max entropy and the conditional smooth max entropy is defined as follows.

**Definition 6.5.2** *For any* $\epsilon \in [0, 1)$ *and* $(a_1, a_2) \in \mathbb{R}^2$*, let* $\tilde{\mathcal{L}}(a_1, a_2, \epsilon|\mathbf{X}_1, \mathbf{X}_2)$ *be*

$$\begin{aligned} \tilde{\mathcal{L}}(a_1, a_2, \epsilon|\mathbf{X}_1, \mathbf{X}_2) = \Bigg\{ (L_1, L_2) : L_1 &\geq \limsup_{n \to \infty} \frac{H^\epsilon(X_1^n|X_2^n) - na_1}{\sqrt{n}} \\ \text{and} \quad L_2 &\geq \limsup_{n \to \infty} \frac{H^\epsilon(X_2^n|X_1^n) - na_2}{\sqrt{n}} \\ \text{and} \quad L_1 + L_2 &\geq \limsup_{n \to \infty} \frac{H^\epsilon(X_1^n, X_2^n) - n(a_1 + a_2)}{\sqrt{n}} \Bigg\}. \end{aligned} \tag{6.38}$$

Then, we have the following relationship between the second-order achievable rate region $\mathcal{L}(a_1, a_2, \epsilon|\mathbf{X}_1, \mathbf{X}_2)$ and $\tilde{\mathcal{L}}(a_1, a_2, \epsilon|\mathbf{X}_1, \mathbf{X}_2)$. This result is similar to Theorem 6.3.4.

**Theorem 6.5.2** *For any $\epsilon \in (0,1)$ and $(a_1, a_2) \in \mathbb{R}^2$, we have*

$$\tilde{\mathcal{L}}(a_1, a_2, \epsilon/3|\mathbf{X}_1, \mathbf{X}_2) \subset \mathcal{L}(a_1, a_2, \epsilon|\mathbf{X}_1, \mathbf{X}_2) \subset \lim_{\delta \downarrow 0} \tilde{\mathcal{L}}(a_1, a_2, \epsilon + \delta|\mathbf{X}_1, \mathbf{X}_2). \quad (6.39)$$

Theorem 6.5.2 does not cover the case of $\epsilon = 0$. The next theorem treats the case of $\epsilon = 0$. This result clarifies that the outer bound of Theorem 6.5.2 holds with equality in the case of $\epsilon = 0$.

**Theorem 6.5.3** *Any pair $(L_1, L_2) \in \lim_{\delta \downarrow 0} \tilde{\mathcal{L}}(a_1, a_2, \delta|\mathbf{X}_1, \mathbf{X}_2)$ is $(a_1, a_2, 0)$-achievable. That is, it holds that*

$$\mathcal{L}(a_1, a_2, 0|\mathbf{X}_1, \mathbf{X}_2) = \lim_{\delta \downarrow 0} \tilde{\mathcal{L}}(a_1, a_2, \delta|\mathbf{X}_1, \mathbf{X}_2). \quad (6.40)$$

The prior work [31] has characterized the second-order achievable rate region by $F_n(L_1, L_2|a_1, a_2)$. On the other hand, this study characterizes the second-order achievable rate region by $G_n(L_1, L_2|a_1, a_2)$. If $(R_1, R_2)$ is chosen so that $R_i = a_i + (L_i/\sqrt{n})$ for $i = 1, 2$ in Theorem 6.3.5, the following relationship is immediately obtained.

**Theorem 6.5.4** *For any $\gamma > 0$, $(a_1, a_2) \in \mathbb{R}^2$, and $n = 1, 2, \ldots$, it holds that*

$$F_n(L_1 + \gamma, L_2 + \gamma|a_1, a_2) - 3e^{-\sqrt{n}\gamma} \leq G_n(L_1, L_2|a_1, a_2) \leq F_n(L_1, L_2|a_1, a_2). \quad (6.41)$$

# 6.6 Proofs of main results

## 6.6.1 Proof of Theorem 6.5.1

**Direct part:**

For any $(L_1, L_2) \in \mathrm{Cl}(\{(L_1, L_2) : G(L_1, L_2|a_1, a_2) \leq \epsilon\})$ and any $\gamma > 0$, let

$$M_n^{(1)} = e^{na_1 + \sqrt{n}(L_1 + 2\gamma)}, \quad (6.42)$$
$$M_n^{(2)} = e^{na_2 + \sqrt{n}(L_2 + 2\gamma)}. \quad (6.43)$$

Then, from Lemma 6.4.1, there exists an $(n, M_n^{(1)}, M_n^{(2)}, \epsilon_n)$ code such that

$$\epsilon_n \leq 1 - \max_{\mathcal{T}_n \in \mathcal{S}(e^{na_1 + \sqrt{n}(L_1+\gamma)}, e^{na_2 + \sqrt{n}(L_2+\gamma)})} \mathbb{P}[(X_1^n, X_2^n) \in \mathcal{T}_n] + 3e^{-\sqrt{n}\gamma} \quad (6.44)$$
$$= G_n(L_1 + \gamma, L_2 + \gamma|a_1, a_2) + 3e^{-\sqrt{n}\gamma}. \quad (6.45)$$

From (6.45) and the assumption on $(L_1, L_2)$, it holds that

$$\limsup_{n\to\infty} \epsilon_n \leq \limsup_{n\to\infty} G_n(L_1 + \gamma, L_2 + \gamma|a_1, a_2) \quad (6.46)$$
$$= G(L_1 + \gamma, L_2 + \gamma|a_1, a_2) \quad (6.47)$$
$$\leq \epsilon. \quad (6.48)$$

Since $\gamma > 0$ is arbitrary, (6.48) indicates that any

$$(L_1, L_2) \in \mathrm{Cl}\left(\{(L_1, L_2) : G(L_1, L_2|a_1, a_2) \leq \epsilon\}\right)$$

is $(a_1, a_2, \epsilon)$-achievable.

**Converse part:**

Let $(L_1, L_2)$ is $(a_1, a_2, \epsilon)$-achievable. Then, there exists a sequence of $(n, M_n^{(1)}, M_n^{(2)}, \epsilon_n)$ codes satisfying

$$\limsup_{n\to\infty} \frac{1}{\sqrt{n}} \log \frac{M_n^{(1)}}{e^{na_1}} \le L_1, \tag{6.49}$$

$$\limsup_{n\to\infty} \frac{1}{\sqrt{n}} \log \frac{M_n^{(2)}}{e^{na_2}} \le L_2, \tag{6.50}$$

$$\limsup_{n\to\infty} \epsilon_n \le \epsilon. \tag{6.51}$$

From (6.49) and (6.50), it holds that

$$M_n^{(1)} \le e^{na_1+\sqrt{n}(L_1+\gamma)}, \tag{6.52}$$

$$M_n^{(2)} \le e^{na_2+\sqrt{n}(L_2+\gamma)}, \tag{6.53}$$

for any $\gamma > 0$ and sufficiently large $n$.

Then, (6.52), (6.53), and Lemma 6.4.2 establish the following inequality for sufficiently large $n$.

$$\begin{aligned}\epsilon_n &\ge 1 - \max_{T_n\in\mathcal{S}(e^{na_1+\sqrt{n}(L_1+\gamma)}, e^{na_2+\sqrt{n}(L_2+\gamma)})} \mathbb{P}[(X_1^n, X_2^n) \in \mathcal{T}_n] \\ &= G_n(L_1+\gamma, L_2+\gamma|a_1, a_2).\end{aligned} \tag{6.54}$$

From (6.51) and (6.54), it follows that

$$\epsilon \ge \limsup_{n\to\infty} \epsilon_n \ge \limsup_{n\to\infty} G_n(L_1+\gamma, L_2+\gamma|a_1, a_2) \tag{6.55}$$

$$= G(L_1+\gamma, L_2+\gamma|a_1, a_2). \tag{6.56}$$

Since $\gamma > 0$ is arbitrary,

$$(L_1, L_2) \in \mathrm{Cl}\left(\{(L_1, L_2) : G(L_1, L_2|a_1, a_2) \le \epsilon\}\right) \tag{6.57}$$

is concluded.

### 6.6.2 Proof of Theorem 6.5.2

**Proof of the inner bound:**

For any $\gamma > 0$ and any $\epsilon > 0$, let $M_n^{(1)}$ and $M_n^{(2)}$ be any positive integers satisfying

$$H^\epsilon(X_1^n|X_2^n) + \sqrt{n}\gamma \le \log M_n^{(1)} \le H^\epsilon(X_1^n|X_2^n) + 2\sqrt{n}\gamma, \tag{6.58}$$

$$H^\epsilon(X_2^n|X_1^n) + \sqrt{n}\gamma \le \log M_n^{(2)} \le H^\epsilon(X_2^n|X_1^n) + 2\sqrt{n}\gamma, \tag{6.59}$$

$$H^\epsilon(X_1^n, X_2^n) + 2\sqrt{n}\gamma \le \log M_n^{(1)} M_n^{(2)} \le H^\epsilon(X_1^n, X_2^n) + 4\sqrt{n}\gamma. \tag{6.60}$$

For the $M_n^{(1)}, M_n^{(2)}$, Lemma 6.4.1 guarantees that there exists an $(n, M_n^{(1)}, M_n^{(2)}, \epsilon_n)$ code satisfying

$$\epsilon_n \leq 1 - \max_{T_n \in \mathcal{S}(M_n^{(1)} e^{-\sqrt{n}\gamma}, M_n^{(2)} e^{-\sqrt{n}\gamma})} \mathbb{P}[(X_1^n, X_2^n) \in \mathcal{T}_n] + 3e^{-\sqrt{n}\gamma}. \tag{6.61}$$

On the other hand, let $\mathcal{T}_n^{(12)}, \mathcal{T}_n^{(1)}$, and $\mathcal{T}_n^{(2)}$ be subsets of $\mathcal{X}_1^n \times \mathcal{X}_2^n$ satisfying

$$\begin{cases} \log |\mathcal{T}_n^{(12)}| = H^\epsilon(X_1^n, X_2^n), & (6.62) \\ \mathbb{P}[(X_1^n, X_2^n) \in \mathcal{T}_n^{(12)}] \geq 1 - \epsilon, & (6.63) \end{cases}$$

$$\begin{cases} \log \max_{\mathbf{x}_2 \in \mathcal{X}_2^n} |\{\mathbf{x}_1 \in \mathcal{X}_1^n : (\mathbf{x}_1, \mathbf{x}_2) \in \mathcal{T}_n^{(1)}\}| = H^\epsilon(X_1^n | X_2^n), & (6.64) \\ \mathbb{P}[(X_1^n, X_2^n) \in \mathcal{T}_n^{(1)}] \geq 1 - \epsilon, & (6.65) \end{cases}$$

$$\begin{cases} \log \max_{\mathbf{x}_1 \in \mathcal{X}_1^n} |\{\mathbf{x}_2 \in \mathcal{X}_2^n : (\mathbf{x}_1, \mathbf{x}_2) \in \mathcal{T}_n^{(2)}\}| = H^\epsilon(X_2^n | X_1^n), & (6.66) \\ \mathbb{P}[(X_1^n, X_2^n) \in \mathcal{T}_n^{(2)}] \geq 1 - \epsilon, & (6.67) \end{cases}$$

and let $\mathcal{T}_n^*$ be $\mathcal{T}_n^* := \mathcal{T}_n^{(12)} \cap \mathcal{T}_n^{(1)} \cap \mathcal{T}_n^{(2)}$. Then, (6.63), (6.65), (6.67), and the union bound yield

$$1 - \mathbb{P}[(X_1^n, X_2^n) \in \mathcal{T}_n^*] \leq 3\epsilon. \tag{6.68}$$

Further, from (6.62), (6.64), (6.66), and the left inequalities of (6.58), (6.59), (6.60), it holds that

$$\mathcal{T}_n^* \in \mathcal{S}(M_n^{(1)} e^{-\sqrt{n}\gamma}, M_n^{(2)} e^{-\sqrt{n}\gamma}). \tag{6.69}$$

Thus, (6.61), (6.68), and (6.69) establish

$$\epsilon_n \leq 1 - \mathbb{P}[(X_1^n, X_2^n) \in \mathcal{T}_n^*] + 3e^{-\sqrt{n}\gamma} \tag{6.70}$$

$$\leq 3\epsilon + 3e^{-\sqrt{n}\gamma}. \tag{6.71}$$

From the right inequalities of (6.58), (6.59), (6.60) and the inequality (6.71), it is possible to construct a sequence of $(n, M_n^{(1)}, M_n^{(2)}, \epsilon_n)$ codes satisfying

$$\frac{1}{\sqrt{n}} \log \frac{M_n^{(1)}}{e^{na_1}} \leq \frac{H^\epsilon(X_1^n | X_2^n) - na_1}{\sqrt{n}} + 2\gamma, \tag{6.72}$$

$$\frac{1}{\sqrt{n}} \log \frac{M_n^{(2)}}{e^{na_2}} \leq \frac{H^\epsilon(X_2^n | X_1^n) - na_2}{\sqrt{n}} + 2\gamma, \tag{6.73}$$

$$\frac{1}{\sqrt{n}} \log \frac{M_n^{(1)}}{e^{na_1}} + \frac{1}{\sqrt{n}} \log \frac{M_n^{(2)}}{e^{na_2}} \leq \frac{H^\epsilon(X_1^n, X_2^n) - n(a_1 + a_2)}{\sqrt{n}} + 4\gamma, \tag{6.74}$$

$$\epsilon_n \leq 3\epsilon + 3e^{-\sqrt{n}\gamma}. \tag{6.75}$$

Taking $\limsup_{n\to\infty}$ of the above inequalities, we have

$$\limsup_{n\to\infty} \frac{1}{\sqrt{n}} \log \frac{M_n^{(1)}}{e^{na_1}} \leq \limsup_{n\to\infty} \frac{H^\epsilon(X_1^n|X_2^n) - na_1}{\sqrt{n}} + 2\gamma, \tag{6.76}$$

$$\limsup_{n\to\infty} \frac{1}{\sqrt{n}} \log \frac{M_n^{(2)}}{e^{na_2}} \leq \limsup_{n\to\infty} \frac{H^\epsilon(X_2^n|X_1^n) - na_2}{\sqrt{n}} + 2\gamma, \tag{6.77}$$

$$\limsup_{n\to\infty} \left( \frac{1}{\sqrt{n}} \log \frac{M_n^{(1)}}{e^{na_1}} + \frac{1}{\sqrt{n}} \log \frac{M_n^{(2)}}{e^{na_2}} \right) \leq \limsup_{n\to\infty} \frac{H^\epsilon(X_1^n, X_2^n) - n(a_1 + a_2)}{\sqrt{n}} + 4\gamma, \tag{6.78}$$

$$\limsup_{n\to\infty} \epsilon_n \leq 3\epsilon. \tag{6.79}$$

Since $\gamma > 0$ is arbitrary, $\tilde{\mathcal{L}}(a_1, a_2, \epsilon/3|\mathbf{X}_1, \mathbf{X}_2) \subset \mathcal{L}(a_1, a_2, \epsilon|\mathbf{X}_1, \mathbf{X}_2)$ is obtained.

**Proof of the outer bound:**

For any sequence of codes that satisfies $\limsup_{n\to\infty} \epsilon_n \leq \epsilon$, we have $\epsilon_n \leq \epsilon + \delta$ for any $\delta > 0$ and sufficiently large $n$. Thus, Lemma 6.4.2 yields

$$1 - \max_{T_n \in \mathcal{S}(M_n^{(1)}, M_n^{(2)})} \mathbb{P}[(X_1^n, X_2^n) \in \mathcal{T}_n] \leq \epsilon_n \leq \epsilon + \delta \tag{6.80}$$

for any $\delta > 0$ and sufficiently large $n$.

Next, let $\mathcal{T}_n^* \in \mathcal{S}(M_n^{(1)}, M_n^{(2)})$ be defined as the set that maximizes $\mathbb{P}[(X_1^n, X_2^n) \in \mathcal{T}_n]$ in (6.80). Then, the definitions of the smooth max entropy and the conditional smooth max entropy establish

$$H^{\epsilon+\delta}(X_1^n, X_2^n) \leq \log |\mathcal{T}_n^*| \leq \log M_n^{(1)} M_n^{(2)}, \tag{6.81}$$

$$H^{\epsilon+\delta}(X_1^n|X_2^n) \leq \log \max_{\mathbf{x}_2 \in \mathcal{X}_2^n} |\{\mathbf{x}_1 : (\mathbf{x}_1, \mathbf{x}_2) \in \mathcal{T}_n^*\}| \leq \log M_n^{(1)}, \tag{6.82}$$

$$H^{\epsilon+\delta}(X_2^n|X_1^n) \leq \log \max_{\mathbf{x}_1 \in \mathcal{X}_2^n} |\{\mathbf{x}_2 : (\mathbf{x}_1, \mathbf{x}_2) \in \mathcal{T}_n^*\}| \leq \log M_n^{(2)}. \tag{6.83}$$

Therefore, for any $(a_1, a_2, \epsilon)$-achievable rate pair $(L_1, L_2)$,

$$\limsup_{n\to\infty} \frac{H^{\epsilon+\delta}(X_1^n, X_2^n) - n(a_1 + a_2)}{\sqrt{n}} \leq L_1 + L_2, \tag{6.84}$$

$$\limsup_{n\to\infty} \frac{H^{\epsilon+\delta}(X_1^n|X_2^n) - na_1}{\sqrt{n}} \leq L_1, \tag{6.85}$$

$$\limsup_{n\to\infty} \frac{H^{\epsilon+\delta}(X_2^n|X_1^n) - na_2}{\sqrt{n}} \leq L_2. \tag{6.86}$$

Since $\delta > 0$ is arbitrary, $\mathcal{L}(a_1, a_2, \epsilon|\mathbf{X}_1, \mathbf{X}_2) \subset \lim_{\delta\downarrow 0} \tilde{\mathcal{L}}(a_1, a_2, \epsilon + \delta|\mathbf{X}_1, \mathbf{X}_2)$ is obtained.

### 6.6.3 Proof of Theorem 6.5.3

First, fix arbitrary sequence $\{\delta_i\}$ satisfying $1 > \delta_1 > \delta_2 > \cdots \to 0$. For each $i = 1, 2, \ldots$, by letting $\epsilon = \delta_i$ in the proof of the inner bound of Theorem 6.5.2, we can prove that any $(L_1, L_2) \in \tilde{\mathcal{L}}(a_1, a_2, \delta_i|\mathbf{X}_1, \mathbf{X}_2)$ is $(a_1, a_2, 3\delta_i)$-achievable. That is, for each $i = 1, 2, \ldots$, there exists a sequence of $(n, M_n^{(1)}, M_n^{(2)}, \epsilon_n)$ codes satisfying

$$\limsup_{n\to\infty} \frac{1}{\sqrt{n}} \log \frac{M_n^{(1)}}{e^{na_1}} \leq L_1, \tag{6.87}$$

$$\limsup_{n\to\infty} \frac{1}{\sqrt{n}} \log \frac{M_n^{(2)}}{e^{na_2}} \leq L_2, \tag{6.88}$$

$$\limsup_{n\to\infty} \epsilon_n \leq 3\delta_i. \tag{6.89}$$

From (6.89), it is observed that for an arbitrary $\gamma > 0$, there exists a sequence of positive integers $\{n_i\}$ satisfying

$$\epsilon_n \leq 3\delta_i + \gamma \quad (\forall i = 1, 2, \ldots; \forall n \geq n_i) \tag{6.90}$$

and $n_1 < n_2 < \cdots \to +\infty$. For each $n = 1, 2, \ldots$, let $i_n$ be the integer $i$ satisfying $n_i \leq n < n_{i+1}$. Then, it holds that

$$\epsilon_n \leq 3\delta_{i_n} + \gamma \quad (n = 1, 2, \ldots). \tag{6.91}$$

Therefore, by noticing that $\delta_{i_1} \geq \delta_{i_2} \geq \delta_{i_3} \geq \cdots \to 0$ $(n \to \infty)$ and $\gamma > 0$ is arbitrary, it holds that

$$\lim_{n\to\infty} \epsilon_n = 0. \tag{6.92}$$

This result shows that any pair $(L_1, L_2)$ in

$$\lim_{i\to\infty} \tilde{\mathcal{L}}(a_1, a_2, \delta_i|\mathbf{X}_1, \mathbf{X}_2) = \lim_{\delta\downarrow 0} \tilde{\mathcal{L}}(a_1, a_2, \delta|\mathbf{X}_1, \mathbf{X}_2)$$

is $(a_1, a_2, 0)$-achievable.

## 6.7 Conclusion of this chapter

This chapter considered the fixed-length Slepian-Wolf coding problem for a general source. The second-order $(a_1, a_2, \epsilon)$-achievable rate region $\mathcal{L}(a_1, a_2, \epsilon|\mathbf{X}_1, \mathbf{X}_2)$ was characterized by the function $G(L_1, L_2|a_1, a_2)$. The rate region $\tilde{\mathcal{L}}(a_1, a_2, \epsilon|\mathbf{X}_1, \mathbf{X}_2)$ was defined based on the smooth max entropy and the conditional smooth max entropy. The relationship between $\mathcal{L}(a_1, a_2, \epsilon|\mathbf{X}_1, \mathbf{X}_2)$ and $\tilde{\mathcal{L}}(a_1, a_2, \epsilon|\mathbf{X}_1, \mathbf{X}_2)$ was shown. Further, the relationship between $G_n(L_1, L_2|a_1, a_2)$ and $F_n(L_1, L_2|a_1, a_2)$ was clarified.

# Chapter 7

# Asymptotic analysis of the Bayes code allowing positive overflow probability

## 7.1 Introduction

This chapter allows a positive overflow probability up to $\epsilon$ and evaluates the $\epsilon$-coding rate[1] of the Bayes code for a stationary ergodic finite order Markov source. The results show the advantage of the Bayes code and give the new insight of the Bayes code under the overflow probability. Related previous studies are stated as follows.

In the non-universal setting (i.e., a probability distribution of a source is known), Nomura et al. [28] have analyzed the $\epsilon$-coding rate of the Shannon code for a stationary memoryless source. Further, Kontoyiannis and Verdú [18] have evaluated the $\epsilon$-coding rate of the optimal non-prefix code defined in [18] for a stationary memoryless source and a stationary ergodic finite order Markov source.

On the other hand, in the universal setting (i.e., a probability distribution of a source is unknown), Kosut and Sankar [21] have shown the $\epsilon$-coding rate of the Type Size code for a stationary memoryless source. In [22], they also have derived the $\epsilon$-coding rate of the code based on the two-stage description for a stationary memoryless source. For a stationary ergodic first order Markov source, Iri and Kosut [15] have derived the $\epsilon$-coding rate of the Type Size code.

A map of this chapter and previous studies are shown in Table 7.1.

This chapter is organized as follows. Section 7.2 shows the problem formulation. Section 7.3 explains the related previous works. Section 7.4 describes the Bayes code. Section 7.5 derives the main results. Finally, Section 7.6 discusses the main results and concludes this chapter.

[1]The $\epsilon$-coding rate is the same as the minimum threshold of the overflow probability discussed in Chapters 3–5. However, following the related previous works such as [15], [21], and [22], we use this terminology in this chapter.

Table 7.1: The map of Chapter 7 and previous studies

| | stationary memoryless source | Markov source |
|---|---|---|
| non-universal setting | Nomura et al. [28]<br>Kontoyiannis and Verdú [18] | Kontoyiannis and Verdú [18] |
| universal setting | Kosut and Sankar [21]<br>Kosut and Sankar [22] | Iri and Kosut [15]<br>Chapter 7 |

## 7.2 Problem formulation

Let $\mathcal{X} = \{0, 1, \ldots, K\}$ be a finite source alphabet and $X$ be a random variable taking a value in $\mathcal{X}$. In this chapter, we treat a stationary ergodic finite order Markov source. Let $\mathcal{S}$ be a set of states (state space) of a stationary ergodic finite order Markov source and let $\theta_{i,s_j}$ be a probability of occurrence of the symbol $i \in \mathcal{X}$ under the state $s_j \in \mathcal{S}$ $(j = 0, 1, \ldots, |\mathcal{S}| - 1)$. Then, define $\theta_{s_j}^K$ by

$$\theta_{s_j}^K = (\theta_{0,s_j}, \theta_{1,s_j}, \ldots, \theta_{K-1,s_j})^T, \tag{7.1}$$

where we assume that $\theta_{i,s_j} > 0$ for $i \in \mathcal{X}$, $j = 0, 1, \ldots, |\mathcal{S}| - 1$ and $\sum_{i=0}^{K} \theta_{i,s_j} = 1$ for $j = 0, 1, \ldots, |\mathcal{S}| - 1$. Further, let $\theta^k$ be

$$\theta^k = (\theta_{s_0}^K, \theta_{s_1}^K, \ldots, \theta_{s_{|\mathcal{S}|-1}}^K)^T, \tag{7.2}$$

where $k = K|\mathcal{S}|$. Moreover, $p_{\theta^k}$ denotes a probability mass function of a stationary ergodic finite order Markov source defined by $\theta^k$. The parameter space of $\theta^k$ is denoted as $\Theta^k = (0, 1)^k$.

Let $w(\theta^k)$ be the prior probability density function of $\theta^k$. We assume that $w(\theta^k) > 0$ for any $\theta^k \in \Theta^k$ and $w(\theta^k)$ is three times continuously differentiable. Further, we assume that $w(\theta^k)$ and a class of parameterized distribution of a source $\{p_{\theta^k} : \theta^k \in \Theta^k = (0, 1)^k\}$ are known, but the true parameter $\theta_*^k \in \Theta^k$ is unknown.

The $k \times k$ Fisher information matrix $I(\theta^k)$ is defined as

$$I(\theta^k) = \lim_{n \to \infty} \frac{1}{n} E_{p_{\theta_*^k}} \left[ -\frac{\partial^2 \ln p_{\theta^k}(X^n)}{\partial \theta^k (\partial \theta^k)^T} \right], \tag{7.3}$$

where $E_{p_{\theta_*^k}}[\cdot]$ denotes the expectation by $p_{\theta_*^k}$. As shown in Goto et al. [9, Example 5], the determinant of the Fisher information matrix, $\det I(\theta^k)$, is continuous with respect to $\theta^k$ under a stationary ergodic finite order Markov source. Further, Goto et al. [9, Example 5] have shown that for a stationary ergodic finite order Markov source, it holds that[2]

$$-\frac{1}{n}\frac{\partial^2 \ln p_{\theta^k}(X^n)}{\partial\theta^k(\partial\theta^k)^T} \to I(\theta^k) \quad \text{a.s.} \tag{7.4}$$

Let $\hat{\theta}^k$ denote the maximum likelihood estimator. Regarding $\hat{\theta}^k$, it holds that

$$\hat{\theta}^k = \theta^k_* + O\left(\left(\frac{\ln\ln n}{n}\right)^{1/2}\right) \quad \text{a.s.} \tag{7.5}$$

for a stationary ergodic finite order Markov source [8].

The following quantities play a crucial role in the main results of this chapter.

**Definition 7.2.1** *We define $H_{\theta^k}(X^n)$, $H_{\theta^k}(\mathbf{X})$, $\sigma^2_{\theta^k}(X^n)$, $\sigma^2_{\theta^k}(\mathbf{X})$, $\sigma_{\theta^k}(X^n)$, and $\sigma_{\theta^k}(\mathbf{X})$ as follows.*

$$H_{\theta^k}(X^n) := E_{p_{\theta^k_*}}\left[\ln\frac{1}{p_{\theta^k}(X^n)}\right], \tag{7.6}$$

$$H_{\theta^k}(\mathbf{X}) := \lim_{n\to\infty}\frac{1}{n}H_{\theta^k}(X^n), \tag{7.7}$$

$$\sigma^2_{\theta^k}(X^n) := V_{p_{\theta^k_*}}\left[\ln\frac{1}{p_{\theta^k}(X^n)}\right], \tag{7.8}$$

$$\sigma^2_{\theta^k}(\mathbf{X}) := \lim_{n\to\infty}\frac{1}{n}\sigma^2_{\theta^k}(X^n), \tag{7.9}$$

$$\sigma_{\theta^k}(X^n) := \sqrt{\sigma^2_{\theta^k}(X^n)}, \tag{7.10}$$

$$\sigma_{\theta^k}(\mathbf{X}) := \sqrt{\sigma^2_{\theta^k}(\mathbf{X})}, \tag{7.11}$$

*where $V_{p_{\theta^k_*}}[\cdot]$ denotes the variance by $p_{\theta^k_*}$. Further, $H_{\theta^k_*}(X^n)$ is defined as*

$$H_{\theta^k_*}(X^n) := E_{p_{\theta^k_*}}\left[\ln\frac{1}{p_{\theta^k_*}(X^n)}\right]. \tag{7.12}$$

*We define $H_{\theta^k_*}(\mathbf{X})$, $\sigma^2_{\theta^k_*}(X^n)$, $\sigma^2_{\theta^k_*}(\mathbf{X})$, $\sigma_{\theta^k_*}(X^n)$, and $\sigma_{\theta^k_*}(\mathbf{X})$ in the same way.*

The quantity $H_{\theta^k_*}(X^n)$ is said to be *entropy* and $H_{\theta^k_*}(\mathbf{X})$ is said to be *entropy rate*. Also, $\sigma^2_{\theta^k_*}(X^n)$ is said to be *varentropy* and $\sigma^2_{\theta^k_*}(\mathbf{X})$ is said to be *varentropy rate*[3]. Since we assume a stationary ergodic finite order Markov source, the varentropy rate exists [18] as well as the entropy rate does. Further, $\sigma^2_{\theta^k_*}(\mathbf{X}) < \infty$ holds for a stationary ergodic finite order Markov source [18]. In this dissertation, we assume $\sigma^2_{\theta^k_*}(\mathbf{X}) > 0$.

This chapter analyzes the $\epsilon$-coding rate defined as follows.

[2]In this dissertation, "$\mathbb{P}_{\theta^k_*}$-almost surely" is abbreviated as "a.s." Thus, (7.4) is equivalent to

$$\mathbb{P}_{\theta^k_*}\left[-\frac{1}{n}\frac{\partial^2 \ln p_{\theta^k}(X^n)}{\partial\theta^k(\partial\theta^k)^T} \to I(\theta^k)\right] = 1.$$

[3]The quantity $\sigma^2_{\theta^k_*}(\mathbf{X})$ is also said to be the *minimal coding variance* in [17].

**Definition 7.2.2** *Let $\ell(X^n)$ denote a codeword length of a code for a source sequence $X^n$. Then, for $\epsilon \in [0,1)$, the $\epsilon$-coding rate for the code is defined as*

$$R_\ell^*(n,\epsilon,\theta_*^k) = \inf\left\{R : \mathbb{P}_{\theta_*^k}[\ell(X^n) > nR] \le \epsilon\right\}. \tag{7.13}$$

# 7.3 Related previous works

## 7.3.1 Previous works: distribution of a source is known

This subsection describes previous studies under the setup that the probability distribution of a source is known.

Let $\ell_S(\cdot)$ denote a codeword length of the Shannon code, i.e.,

$$\ell_S(x^n) = \left\lceil \ln \frac{1}{p_{\theta_*^k}(x^n)} \right\rceil \tag{7.14}$$

for $x^n \in \mathcal{X}^n$. Then, the $\epsilon$-coding rate of the Shannon code, $R_{\ell_S}^*(n,\epsilon,\theta_*^k)$, is calculated as follows from the result of [29] for a stationary memoryless source:

$$R_{\ell_S}^*(n,\epsilon,\theta_*^k) = H_{\theta_*^k}(X) + \frac{\sigma_{\theta_*^k}(X)}{\sqrt{n}} Q^{-1}(\epsilon) + o\left(\frac{1}{n}\right), \tag{7.15}$$

where $Q(z)$ is defined as

$$Q(z) = \int_z^\infty \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt \tag{7.16}$$

and $Q^{-1}(z)$ denotes its inverse function for $z \in \mathbb{R}$.

Let $\ell_O(\cdot)$ denote a codeword length of the optimal variable-length code defined in [18], where "optimal" means that the overflow probability of the code is smallest among all variable-length code without prefix constraint. The upper and lower bounds of the $\epsilon$-coding rate of this code are evaluated as follows in [18] for a stationary ergodic finite order Markov source:

$$R_{\ell_O}^*(n,\epsilon,\theta_*^k) \le H_{\theta_*^k}(\mathbf{X}) + \frac{\sigma_{\theta_*^k}(\mathbf{X})}{\sqrt{n}} Q^{-1}(\epsilon) + \frac{C_1}{n}, \tag{7.17}$$

$$R_{\ell_O}^*(n,\epsilon,\theta_*^k) \ge H_{\theta_*^k}(\mathbf{X}) + \frac{\sigma_{\theta_*^k}(\mathbf{X})}{\sqrt{n}} Q^{-1}(\epsilon) - \frac{1}{2n}\ln n - \frac{C_2}{n}, \tag{7.18}$$

where $\epsilon \in (0, 1/2)$ and $C_1$ and $C_2$ are positive constants.

## 7.3.2 Previous works: distribution of a source is unknown

This subsection describes previous studies under the setup that the probability distribution of a source is unknown.

Let $\ell_{\mathrm{TS}}(\cdot)$ denote a codeword length of the prefix code based on the two-stage description[4]. The $\epsilon$-coding rate of this code is given as follows in [22] for a stationary memoryless source:

$$R^*_{\ell_{\mathrm{TS}}}(n,\epsilon,\theta^k_*) = H_{\theta^k_*}(X) + \frac{\sigma_{\theta^k_*}(X)}{\sqrt{n}}Q^{-1}(\epsilon) + \frac{k}{2n}\ln n + O\left(\frac{1}{n}\right). \tag{7.19}$$

The non-prefix code based on the type class size, which is said to be the Type Size code, has been proposed in [21]. Let $\ell_{\mathrm{Type}}(\cdot)$ denote a codeword length of the Type Size code. The upper bound on the $\epsilon$-coding rate of this code is evaluated as follows in [21] for a stationary memoryless source:

$$R^*_{\ell_{\mathrm{Type}}}(n,\epsilon,\theta^k_*) \leq H_{\theta^k_*}(X) + \frac{\sigma_{\theta^k_*}(X)}{\sqrt{n}}Q^{-1}(\epsilon) + \frac{k-2}{2n}\ln n + O\left(\frac{1}{n}\right). \tag{7.20}$$

Let $\ell_{\mathrm{np}}(\cdot)$ denote a codeword length of a non-prefix code. It is proved in [22] that the $\epsilon$-coding rate for any non-prefix code satisfies the following inequality for a stationary memoryless source:

$$\sup_{\theta^k\in\Theta^k}\left[R^*_{\ell_{\mathrm{np}}}(n,\epsilon,\theta^k) - H_{\theta^k}(X) - \frac{\sigma_{\theta^k}(X)}{\sqrt{n}}Q^{-1}(\epsilon)\right] \geq \frac{k-2}{2n}\ln n + O\left(\frac{1}{n}\right). \tag{7.21}$$

Lastly, it is shown in [15] that the Type Size code satisfies the following inequality for a stationary ergodic first-order Markov source:

$$\sup_{\theta^k\in\Theta^k}\left[R^*_{\ell_{\mathrm{Type}}}(n,\epsilon,\theta^k) - H_{\theta^k}(\mathbf{X}) - \frac{\sigma_{\theta^k}(\mathbf{X})}{\sqrt{n}}Q^{-1}(\epsilon)\right] = \frac{k-2}{2n}\ln n + O\left(\frac{1}{n}\right). \tag{7.22}$$

## 7.4 Bayes code

The Bayes code is one of universal variable-length prefix codes (see, e.g., [5], [2], [24]). For a source sequence $x^n$, the Bayes code utilizes the arithmetic coding probability $P_B(x^n)$, where $P_B(\cdot)$ is the probability that minimizes the Bayes risk function defined as

$$\int_{\Theta^k} w(\theta^k)\left\{E_{p_{\theta^k}}\left[\ln\frac{1}{P_B(X^n)}\right] - H_{\theta^k}(X^n)\right\}d\theta^k. \tag{7.23}$$

In other words, the Bayes code is the optimal code in the sense that it minimizes the mean codeword length averaged with the prior probability density function $w(\theta^k)$. The arithmetic coding probability of the Bayes code $P_B(x^n)$ is given by $\int_{\Theta^k} w(\theta^k)p_{\theta^k}(x^n)d\theta^k$ (see, e.g., [24]), and the codeword length of the Bayes code $\ell_B(x^n)$ is

$$\ell_B(x^n) = -\ln\int_{\Theta^k} w(\theta^k)p_{\theta^k}(x^n)d\theta^k. \tag{7.24}$$

[4]This means that the first stage encodes the type of a sequence and the second stage encodes the index of the sequence within the type class (see, e.g., [3, Chap. 13])

For a stationary memoryless source and a stationary ergodic finite order Markov source, the mean codeword length of the Bayes code has been analyzed up to constant terms (see, e.g., [2], [8]). Furthermore, the codeword length of the Bayes code for a source sequence $X^n$ has been evaluated as follows for a stationary ergodic finite order Markov source [8]:

$$\ell_B(X^n) = \ln \frac{1}{p_{\hat{\theta}^k}(X^n)} + \frac{k}{2} \ln \frac{n}{2\pi} + \ln \frac{\sqrt{\det I(\hat{\theta}^k)}}{w(\hat{\theta}^k)} + o(1) \quad \text{a.s.} \tag{7.25}$$

## 7.5 Main results

Before showing the main results in this chapter, we show some lemmas.

**Lemma 7.5.1** *For a stationary ergodic finite order Markov source, the codeword length of the Bayes code $\ell_B(X^n)$ for a source sequence $X^n$ is given by*

$$\ell_B(X^n) = \ln \frac{1}{p_{\theta_*^k}(X^n)} + \frac{k}{2} \ln \frac{n}{2\pi} + O(\ln \ln n) \quad \text{a.s.} \tag{7.26}$$

(**Proof**) Taylor's expansion of $\ln(1/p_{\theta_*^k}(X^n))$ around $\hat{\theta}^k$ yields

$$\begin{aligned} \ln \frac{1}{p_{\theta_*^k}(X^n)} =& \ln \frac{1}{p_{\hat{\theta}^k}(X^n)} + \frac{1}{2}(\theta_*^k - \hat{\theta}^k)^T \frac{\partial^2 \ln p_{\theta^k}(X^n)}{\partial \theta^k (\partial \theta^k)^T} \bigg|_{\theta^k = \hat{\theta}^k} (\theta_*^k - \hat{\theta}^k) \\ &+ o\left( (\theta_*^k - \hat{\theta}^k)^T \frac{\partial^2 \ln p(X^n|\theta^k)}{\partial \theta^k (\partial \theta^k)^T} \bigg|_{\theta^k = \hat{\theta}^k} (\theta_*^k - \hat{\theta}^k) \right). \end{aligned} \tag{7.27}$$

Therefore, it holds that

$$\begin{aligned} \ln \frac{1}{p_{\hat{\theta}^k}(X^n)} =& \ln \frac{1}{p_{\theta_*^k}(X^n)} - \frac{1}{2}(\theta_*^k - \hat{\theta}^k)^T \frac{\partial^2 \ln p(X^n|\theta^k)}{\partial \theta^k (\partial \theta^k)^T} \bigg|_{\theta^k = \hat{\theta}^k} (\theta_*^k - \hat{\theta}^k) \\ &+ o\left( (\theta_*^k - \hat{\theta}^k)^T \frac{\partial^2 \ln p_{\theta^k}(X^n)}{\partial \theta^k (\partial \theta^k)^T} \bigg|_{\theta^k = \hat{\theta}^k} (\theta_*^k - \hat{\theta}^k) \right). \end{aligned} \tag{7.28}$$

Combination of (7.4) and (7.5) gives

$$\begin{aligned} &(\theta_*^k - \hat{\theta}^k)^T \frac{\partial^2 \ln p_{\theta^k}(X^n)}{\partial \theta^k (\partial \theta^k)^T} \bigg|_{\theta^k = \hat{\theta}^k} (\theta_*^k - \hat{\theta}^k) \\ &= -n(\theta_*^k - \hat{\theta}^k)^T \left( -\frac{1}{n} \frac{\partial^2 \ln p_{\theta^k}(X^n)}{\partial \theta^k (\partial \theta^k)^T} \bigg|_{\theta^k = \hat{\theta}^k} \right) (\theta_*^k - \hat{\theta}^k) \\ &= O(\ln \ln n) \quad \text{a.s.} \end{aligned} \tag{7.29}$$

Thus, from (7.25), (7.28), and (7.29), we obtain (7.26). □

**Lemma 7.5.2** *Let $\phi(z)$ be the probability density function of the standard normal distribution for $z \in \mathbb{R}$. Then, we have the following inequality for $a, b \geq 0$:*

$$Q(a) \geq Q(b) - \phi(b)(a-b). \tag{7.30}$$

(**Proof**) Taylor's expansion leads to

$$Q(a) = Q(b) + Q^{'}(b)(a-b) + \frac{Q^{''}(\xi)}{2}(a-b)^2, \tag{7.31}$$

where $\xi$ lies between $a$ and $b$. Because $Q^{''}(z) \geq 0$ for $z \geq 0$ and $Q^{'}(z) = -\phi(z)$, the inequality (7.30) holds. □

Using the result in [27], Kontoyiannis and Verdú [18] have derived the next inequality.

**Lemma 7.5.3 ([18])** *For a stationary ergodic finite order Markov source, there exists a finite positive constant $A$ such that for all $n \geq 1$, it holds that*

$$\sup_{z \in \mathbb{R}} \left| \mathbb{P}_{\theta_*^k}\left[ \frac{\ln \frac{1}{p_{\theta_*^k}(X^n)} - nH_{\theta_*^k}(\mathbf{X})}{\sqrt{n}\sigma_{\theta_*^k}(\mathbf{X})} > z \right] - Q(z) \right| \leq \frac{A}{\sqrt{n}}. \tag{7.32}$$

The inequality (7.32) is said to be the Berry-Esséen bound. To bound the overflow probability in the proof of the main results, inequalities (7.30) and (7.32) are used.

The $\epsilon$-coding rate of the Bayes code, $R^*_{\ell_B}(n, \epsilon, \theta_*^k)$, is upper and lower bounded as shown in the following theorems.

**Theorem 7.5.1** *For $\epsilon \in (0, 1/2)$ and all $n$ large enough, the $\epsilon$-coding rate of the Bayes code for a stationary ergodic finite order Markov source is bounded as*

$$R^*_{\ell_B}(n, \epsilon, \theta_*^k) \leq H_{\theta_*^k}(\mathbf{X}) + \frac{\sigma_{\theta_*^k}(\mathbf{X})}{\sqrt{n}} Q^{-1}(\epsilon) + \frac{k}{2n}\ln n + O\left(\frac{1}{n}\right), \tag{7.33}$$

$$R^*_{\ell_B}(n, \epsilon, \theta_*^k) \geq H_{\theta_*^k}(\mathbf{X}) + \frac{\sigma_{\theta_*^k}(\mathbf{X})}{\sqrt{n}} Q^{-1}(\epsilon) + \frac{k}{2n}\ln n + \frac{C_l(n)}{n} + O\left(\frac{1}{n}\right), \tag{7.34}$$

*where $C_l(n)$ is a negative term such that $o(\ln n)$ and satisfies the following condition:*

$$\frac{1}{\sqrt{n}}\{C_l(n) + O(\ln \ln n)\} \to -0. \tag{7.35}$$

(**Proof**) First, (7.33) is shown. For simplicity, we abbreviate $H = H_{\theta_*^k}(\mathbf{X})$ and $\sigma = \sigma_{\theta_*^k}(\mathbf{X})$. Let $U(n)$ be defined as

$$U(n) = nH + \sqrt{n}\sigma Q^{-1}\left(\epsilon - \frac{A}{\sqrt{n}}\right) + \frac{k}{2}\ln\frac{n}{2\pi} + \ln\frac{\sqrt{\det I(\theta_*^k)}}{w(\theta_*^k)} + C, \tag{7.36}$$

where $C$ is a positive constant and $A$ is the constant in (7.32). Then, it holds that

$$
\begin{aligned}
&\mathbb{P}_{\theta_*^k}[\ell_B(X^n) > U(n)] \\
&\overset{(a)}{=} \mathbb{P}_{\theta_*^k}\left[\frac{\ln \frac{1}{p_{\hat{\theta}^k}(X^n)} - nH}{\sqrt{n}\sigma} > Q^{-1}\left(\epsilon - \frac{A}{\sqrt{n}}\right) + \frac{C + o(1)}{\sqrt{n}\sigma}\right] && (7.37)\\
&\overset{(b)}{\leq} \mathbb{P}_{\theta_*^k}\left[\frac{\ln \frac{1}{p_{\theta_*^k}(X^n)} - nH}{\sqrt{n}\sigma} > Q^{-1}\left(\epsilon - \frac{A}{\sqrt{n}}\right) + \frac{C + o(1)}{\sqrt{n}\sigma}\right] && (7.38)\\
&\overset{(c)}{\leq} Q\left(Q^{-1}\left(\epsilon - \frac{A}{\sqrt{n}}\right) + \frac{C + o(1)}{\sqrt{n}\sigma}\right) + \frac{A}{\sqrt{n}} && (7.39)\\
&\overset{(d)}{\leq} Q\left(Q^{-1}\left(\epsilon - \frac{A}{\sqrt{n}}\right)\right) - \frac{C + o(1)}{\sqrt{n}\sigma}\phi\left(Q^{-1}\left(\epsilon - \frac{A}{\sqrt{n}}\right) + \frac{C + o(1)}{\sqrt{n}\sigma}\right) + \frac{A}{\sqrt{n}} && (7.40)\\
&= \epsilon - \frac{C + o(1)}{\sqrt{n}\sigma}\phi\left(Q^{-1}\left(\epsilon - \frac{A}{\sqrt{n}}\right) + \frac{C + o(1)}{\sqrt{n}\sigma}\right) && (7.41)\\
&\leq \epsilon, && (7.42)
\end{aligned}
$$

where

- $(a)$ follows from (7.25), (7.36), and the fact that

$$\ln \frac{w(\hat{\theta}^k)\sqrt{\det I(\theta_*^k)}}{w(\theta_*^k)\sqrt{\det I(\hat{\theta}^k)}} = o(1) \quad \text{a.s.} \tag{7.43}$$

  holds from (7.5) and the continuity of $w(\theta^k)$ and $\det I(\theta^k)$,

- $(b)$ follows from the fact that $\hat{\theta}^k$ is the maximum likelihood estimator,
- $(c)$ follows from the Berry-Esséen bound (7.32),
- $(d)$ follows from the following reason: because $Q^{-1}(\epsilon - A/\sqrt{n}) > 0$ and $Q^{-1}(\epsilon - A/\sqrt{n}) + (C + o(1))/\sqrt{n}\sigma > 0$ for all $n$ large enough and $\epsilon \in (0, 1/2)$, we substitute $Q^{-1}(\epsilon - A/\sqrt{n})$ and $Q^{-1}(\epsilon - A/\sqrt{n}) + (C + o(1))/\sqrt{n}\sigma$ for $a$ and $b$ respectively in (7.30).

Thus, (7.42) and the definition of $R^*_{\ell_B}(n, \epsilon, \theta_*^k)$ yield

$$nR^*_{\ell_B}(n, \epsilon, \theta_*^k) \leq U(n). \tag{7.44}$$

Hence, (7.44) and some calculation establish (7.33).

Next, (7.34) is shown. Let $L(n)$ be defined as

$$L(n) = nH + \sqrt{n}\sigma Q^{-1}\left(\epsilon + \frac{A}{\sqrt{n}}\right) + \frac{k}{2}\ln\frac{n}{2\pi} + C_l(n). \tag{7.45}$$

Further, let $o^-(1)$ be a term such that $\lim_{n\to\infty} o^-(1) = -0$. Then, it holds that

$$\begin{aligned}
&\mathbb{P}_{\theta_*^k}[\ell_B(X^n) > L(n)] \\
&\overset{(a)}{=} \mathbb{P}_{\theta_*^k}\left[\frac{\ln \frac{1}{p_{\theta_*^k}(X^n)} - nH}{\sqrt{n}\sigma} > Q^{-1}\left(\epsilon + \frac{A}{\sqrt{n}}\right) + o^-(1)\right] &(7.46)\\
&\overset{(b)}{\geq} Q\left(Q^{-1}\left(\epsilon + \frac{A}{\sqrt{n}}\right) + o^-(1)\right) - \frac{A}{\sqrt{n}} &(7.47)\\
&\overset{(c)}{\geq} Q\left(Q^{-1}\left(\epsilon + \frac{A}{\sqrt{n}}\right)\right) - \phi\left(Q^{-1}\left(\epsilon + \frac{A}{\sqrt{n}}\right)\right) o^-(1) - \frac{A}{\sqrt{n}} &(7.48)\\
&= \epsilon - \phi\left(Q^{-1}\left(\epsilon + \frac{A}{\sqrt{n}}\right)\right) o^-(1) &(7.49)\\
&\geq \epsilon, &(7.50)
\end{aligned}$$

where

- $(a)$ follows from (7.26), (7.35), and (7.45),
- $(b)$ follows from the Berry-Esséen bound (7.32), and
- $(c)$ follows from the following reason: because $Q^{-1}(\epsilon + A/\sqrt{n}) + o^-(1) > 0$ for all $n$ large enough and $\epsilon \in (0, 1/2)$, we substitute $Q^{-1}(\epsilon + A/\sqrt{n}) + o^-(1)$ and $Q^{-1}(\epsilon + A/\sqrt{n})$ for $a$ and $b$ respectively in (7.30).

Thus, (7.50) and the definition of $R^*_{\ell_B}(n, \epsilon, \theta_*^k)$ yield

$$nR^*_{\ell_B}(n, \epsilon, \theta_*^k) \geq L(n). \tag{7.51}$$

Therefore, (7.51) and some calculation give the desired result (7.34). □

**Remark 7.5.1** *The concrete expression of the term $C_l(n)$ in (7.34) has not been obtained yet. The more precise evaluation of this term is one of the future works.*

## 7.6 Discussion and conclusion of this chapter

Comparing (7.33), (7.34) with (7.17), (7.18), a new insight of the Bayes code is obtained. The upper and lower bounds (7.17), (7.18) are the results for non-universal optimal code, where optimal means that the code has the smallest overflow probability. On the other hand, the upper and lower bounds (7.33), (7.34) are the results for the Bayes code. As explained in Section 7.4, the Bayes code is designed to minimize the *mean codeword length* averaged with the prior probability density function $w(\theta^k)$. Thus, it is not designed to minimize the *overflow probability*. However, the Bayes code behaves similarly to the optimal code; the first and second terms of the upper bound

(7.33) and the lower bound (7.34) are the same as those of (7.17) and (7.18); the first term is $H_{\theta^k_*}(\mathbf{X})$, and the second term is $\frac{\sigma_{\theta^k_*}(\mathbf{X})}{\sqrt{n}}Q^{-1}(\epsilon)$.

To summarize this chapter, we analyzed the $\epsilon$-coding rate of the Bayes code for a stationary ergodic finite order Markov source. The upper and lower bounds on the $\epsilon$-coding rate were obtained. The main ingredients of the proof of the main result were the asymptotic evaluation of the codeword length of the Bayes code and the Berry-Esséen bound.

# Chapter 8

# Asymptotic analysis of the Bayes code under vanishing overflow probability

## 8.1 Introduction

Chapter 7 has treated the case where the overflow probability is allowed up to $\epsilon$. In contrast to this setting, this chapter considers the case where the overflow probability vanishes asymptotically. This problem setting is closely related to the *moderate deviation regime* in information theory. In the following, we describe the moderate deviation regime.

In information theory, asymptotic studies with the criterion of the overflow probability can be divided into three regimes[1]:

1) The *large deviation regime* in which the overflow probability goes to zero asymptotically and behaves like $\exp\{-nr\}$ for some $r > 0$ (e.g., [42]).[2]

2) The *central limit theorem regime* (it is also said to be the *normal approximation regime* or *second order regime*) in which a positive overflow probability is allowed (this is the setting considered in Chapter 7).

3) The *moderate deviation regime* in which the overflow probability goes to zero asymptotically and behaves like $\exp\{-n^t r\}$ for some $r > 0$ and $t \in (0, 1)$ (e.g., [1]).

The overflow probability and its rate[3] have the trade-off relationship. That is, if the rate is large, the overflow probability is small and vice versa. The relationship between

[1]Regarding this classification, see, e.g., [11].

[2]The notation $n$ denotes a length of a source sequence.

[3]As mentioned in Section 2.5, a rate of the overflow probability is the same as a threshold of the overflow probability.

the overflow probability and its rate in the above regimes 1) – 3) is summarized as follows:

i) In the large deviation regime, the overflow probability goes to zero, *but* the rate is away from the ideal asymptotic limit (i.e., the entropy).

ii) In the central limit theorem regime, the rate goes to the entropy with the speed of $O(1/\sqrt{n})$, *but* the overflow probability is away from zero[4].

iii) In the moderate deviation regime, the overflow probability goes to zero *and* the rate goes to the ideal asymptotic limit (entropy) with the speed slower than $O(1/\sqrt{n})$, that is, slower than in the central limit theorem regime.

As the previous study [1] has pointed out, the moderate deviation regime combines the desired features of the central limit theorem regime and the large deviation regime. That is to say, the moderate deviation regime deals with the case where the rate goes to the ideal asymptotic limit (entropy) *and* the overflow probability goes to zero, whereas the overflow probability is away from zero in the central limit theorem regime and the rate is away from the ideal asymptotic limit (entropy) in the large deviation regime. In the viewpoint of practical applications, it is desirable that the rate is close to the ideal asymptotic limit (entropy) *and* the overflow probability is close to zero.

For a stationary memoryless source, Altuğ et al. [1] have evaluated the behavior of the optimal code; here, optimal means that the overflow probability is smallest among all variable-length codes. However, the study [1] has only treated the *non-universal* setting. Hence, the evaluation for a *universal* code in the moderate deviation regime is a problem to work with.

In the framework of the moderate deviation regime, this chapter analyzes the behavior of the Bayes code for a stationary memoryless source. Although the Bayes code is not designed to minimize the overflow probability, our result shows that the behavior of the overflow probability of the Bayes code is similar to that of the optimal non-universal code investigated in [1]. Furthermore, this chapter derives a necessary and sufficient condition of the overflow probability of the Bayes code vanishing asymptotically.

This chapter is organized as follows. Section 8.2 derives the necessary and sufficient condition of the overflow probability of the Bayes code vanishing asymptotically. Section 8.3 analyzes the overflow probability of the Bayes code in the moderate deviation regime. First, a problem formulation is described. Then, the previous study is shown. Next, the main result is derived. Section 8.4 discusses the main results and concludes this chapter.

[4] For example, see the previous results and the main results stated in Chapter 7 in this thesis.

## 8.2 Necessary and sufficient condition of the overflow probability of the Bayes code vanishing asymptotically

In this section, we treat a stationary ergodic finite order Markov source and use the same notations as in Chapter 7. First, the prior works are described. Then, we derive the necessary and sufficient condition that the overflow probability of the Bayes code approaches zero asymptotically.

### 8.2.1 Previous works

For a source sequence $X^n$, let the codeword length of the Shannon code be $\ell_S(X^n)$. Further, let the codeword length of the optimal code defined in [18] be $\ell^\star(X^n)$.

Kontoyiannis and Verdú [18] have shown that $\ell_S(X^n)$ and $\ell^\star(X^n)$ satisfy the asymptotic normality for a stationary ergodic finite order Markov source:

$$\frac{\ell_S(X^n) - H_{\theta_*^k}(X^n)}{\sqrt{\sigma^2_{\theta_*^k}(X^n)}} \xrightarrow{d} \mathcal{N}(0,1), \tag{8.1}$$

$$\frac{\ell^\star(X^n) - H_{\theta_*^k}(X^n)}{\sqrt{\sigma^2_{\theta_*^k}(X^n)}} \xrightarrow{d} \mathcal{N}(0,1). \tag{8.2}$$

where "$\xrightarrow{d}$" denotes the convergence in distribution.

Combining this result with the result in [28], the next theorem is obtained. This theorem gives the necessary and sufficient condition that the overflow probability of the Shannon code and the optimal code approach zero asymptotically.

**Theorem 8.2.1** *Let $\{\eta_n\}_{n=1}^\infty$ be a sequence such that $\eta_n > 0$ for $n = 1, 2, \ldots$. Then, for a stationary ergodic finite order Markov source,*

$$\lim_{n\to\infty} \mathbb{P}_{\theta_*^k}[\ell_S(X^n) > \eta_n] = 0 \tag{8.3}$$

*and*

$$\lim_{n\to\infty} \mathbb{P}_{\theta_*^k}[\ell^\star(X^n) > \eta_n] = 0 \tag{8.4}$$

*hold if and only if the sequence $\{\eta_n\}_{n=1}^\infty$ satisfies, for all $T \in (1, \infty)$,*

$$\lim_{n\to\infty} \left\{\eta_n - \left(H_{\theta_*^k}(X^n) + T\sqrt{\sigma^2_{\theta_*^k}(X^n)}\right)\right\} = \infty. \tag{8.5}$$

### 8.2.2 Results of this study

The next theorem shows that the codeword length of the Bayes code satisfies the asymptotic normality for a stationary ergodic finite order Markov source.

**Theorem 8.2.2** *For a stationary ergodic finite order Markov source, it holds that*

$$\frac{\ell_B(X^n) - H_{\theta_*^k}(X^n)}{\sqrt{\sigma^2_{\theta_*^k}(X^n)}} \xrightarrow{d} \mathcal{N}(0,1). \tag{8.6}$$

(**Proof**) Since $\sigma^2_{\theta_*^k}(X^n) = O(n)$ holds for a stationary ergodic finite order Markov source (see, e.g., [18]), (7.26) yields

$$\frac{\ell_B(X^n) - \ln \frac{1}{p_{\theta_*^k}(X^n)}}{\sqrt{\sigma^2_{\theta_*^k}(X^n)}} \longrightarrow 0 \quad \text{a.s.} \tag{8.7}$$

This gives

$$\frac{\ell_B(X^n) - \ln \frac{1}{p_{\theta_*^k}(X^n)}}{\sqrt{\sigma^2_{\theta_*^k}(X^n)}} \to 0 \quad \text{in probability,} \tag{8.8}$$

where "in probability" denotes the convergence in probability.

Further, it holds that

$$\frac{\ln \frac{1}{p_{\theta_*^k}(X^n)} - H_{\theta_*^k}(X^n)}{\sqrt{\sigma^2_{\theta_*^k}(X^n)}} \xrightarrow{d} \mathcal{N}(0,1) \tag{8.9}$$

for a stationary ergodic finite order Markov source [18].

Therefore, combination of (8.8), (8.9), and Slutsky's theorem (e.g., [12]) yield

$$\frac{\ell_B(X^n) - H_{\theta_*^k}(X^n)}{\sqrt{\sigma^2_{\theta_*^k}(X^n)}} = \frac{\ell_B(X^n) - \ln \frac{1}{p_{\theta_*^k}(X^n)}}{\sqrt{\sigma^2_{\theta_*^k}(X^n)}} + \frac{\ln \frac{1}{p_{\theta_*^k}(X^n)} - H_{\theta_*^k}(X^n)}{\sqrt{\sigma^2_{\theta_*^k}(X^n)}} \tag{8.10}$$

$$\xrightarrow{d} \mathcal{N}(0,1). \tag{8.11}$$

This is the desired result. □

Prior works such as (8.1) and (8.2) are results for non-universal codes. On the other hand, Theorem 8.2.2 is the result for the Bayes code (universal code). Theorem 8.2.2 and some calculations as in [28] establish the necessary and sufficient condition that the overflow probability of the Bayes code approaches zero asymptotically.

**Theorem 8.2.3** *Let $\{\eta_n\}_{n=1}^{\infty}$ be a sequence such that $\eta_n > 0$ for $n = 1, 2, \ldots$. Then, for a stationary ergodic finite order Markov source,*

$$\lim_{n\to\infty} \mathbb{P}_{\theta_*^k}[\ell_B(X^n) > \eta_n] = 0 \tag{8.12}$$

*holds if and only if the sequence $\{\eta_n\}_{n=1}^{\infty}$ satisfies, for all $T \in (1, \infty)$,*

$$\lim_{n\to\infty} \left\{ \eta_n - \left( H_{\theta_*^k}(X^n) + T\sqrt{\sigma^2_{\theta_*^k}(X^n)} \right) \right\} = \infty. \tag{8.13}$$

## 8.3 Overflow probability of the Bayes code in the moderate deviation regime

### 8.3.1 Problem formulation

Let $\mathcal{X}$ be a finite source alphabet and $X$ be a random variable taking a value in $\mathcal{X}$. A realization of $X$ is denoted as $x$. Let $X_1, X_2, \cdots$ be i.i.d. random variables with the probability distribution $P_{\theta_*^k}$, where $\theta_*^k \in \Theta^k = (0,1)^k$ is the $k$-dimensional parameter. Further, let $p_{\theta_*^k}$ be a probability mass function corresponding to $P_{\theta_*^k}$, i.e., $p_{\theta_*^k}(x) := P_{\theta_*^k}(\{x\})$ for $x \in \mathcal{X}$. A sequence of random variables $X_1, \ldots, X_n$ is denoted as $X^n$ and the realization of $X^n$ is denoted as $x^n$.

The entropy $H_{\theta_*^k}(X)$ and the varentropy $\sigma^2_{\theta_*^k}(X)$ are defined as

$$H_{\theta_*^k}(X) = E_{p_{\theta_*^k}}\left[\ln \frac{1}{p_{\theta_*^k}(X)}\right] \tag{8.14}$$

and

$$\sigma^2_{\theta_*^k}(X) = V_{p_{\theta_*^k}}\left[\ln \frac{1}{p_{\theta_*^k}(X)}\right], \tag{8.15}$$

where $E_{p_{\theta_*^k}}[\cdot]$ and $V_{p_{\theta_*^k}}[\cdot]$ denote the expectation and variance with respect to $p_{\theta_*^k}$, respectively. We assume $0 < \sigma^2_{\theta_*^k}(X) < \infty$.

A variable-length lossless source code is a pair of encoder and decoder $(f_n, g_n)$ defined as follows. An encoder $f_n : \mathcal{X}^n \to \{0,1\}^\star$ is an injective function. A decoder $g_n : \{0,1\}^\star \to \mathcal{X}^n$ is a function such that $g_n(f_n(x^n)) = x^n$ for all $x^n \in \mathcal{X}^n$.

### 8.3.2 Related previous work

In this subsection, we assume that the true parameter $\theta_*^k$ is known. That is, the probability distribution of a source is known. Let $\mathrm{P_e}(\mathbb{P}_{\theta_*^k}, n, R)$ be defined as

$$\mathrm{P_e}(\mathbb{P}_{\theta_*^k}, n, R) = \min \mathbb{P}_{\theta_*^k}[\ell(f_n(X^n)) > nR], \tag{8.16}$$

where the minimum is taken over all variable-length lossless source codes. In other words, $\mathrm{P_e}(\mathbb{P}_{\theta_*^k}, n, R)$ is the overflow probability of the optimal code in the sense that the overflow probability is smallest among all variable-length codes. Regarding $\mathrm{P_e}(\mathbb{P}_{\theta_*^k}, n, R)$, the previous study [1] has proved the next result.

**Theorem 8.3.1 ([1])** *Let $\{R_n\}_{n=1}^{\infty}$ be a sequence of rates such that*

$$R_n = \frac{1}{n} H_{\theta_*^k}(X^n) + \tau_n, \tag{8.17}$$

*where $\{\tau_n\}_{n=1}^{\infty}$ is a sequence satisfying*

$$\lim_{n\to\infty} \tau_n = +0, \tag{8.18}$$

$$\lim_{n\to\infty} \sqrt{n}\tau_n = \infty. \tag{8.19}$$

*Then, it holds that*

$$\lim_{n\to\infty} \frac{\ln \mathrm{P_e}(\mathbb{P}_{\theta_*^k}, n, R_n)}{n\tau_n^2} = -\frac{1}{2\sigma_{\theta_*^k}^2(X)}. \tag{8.20}$$

**Remark 8.3.1** *Theorem 8.3.1 clarifies the following fact: for a sequence of rates $\{R_n\}_{n=1}^{\infty}$ that goes to the entropy with the speed slower than $O(1/\sqrt{n})$,*

$$\mathrm{P_e}(\mathbb{P}_{\theta_*^k}, n, R_n) \approx \exp\left\{-\frac{1}{2\sigma_{\theta_*^k}^2(X)} n\tau_n^2\right\} \tag{8.21}$$

*for sufficiently large $n$. Equation (8.21) shows that*

- *the overflow probability of the optimal code approaches zero*
- *the overflow probability of the optimal code behaves like* $\exp\{-n^t r\}$ *with $r = 1/2\sigma_{\theta_*^k}^2(X) > 0$ and some $t \in (0, 1)$.*

*This indicates that the study in [1] is divided into the moderate deviation regime (see the case 3) in Section 8.1). Moreover, (8.21) shows that the slower $\tau_n$ goes to zero, the faster the overflow probability $\mathrm{P_e}(\mathbb{P}_{\theta_*^k}, n, R_n)$ goes to zero. That is to say, $\tau_n$ is the parameter that controls the trade-off between the speed of the overflow probability going to zero and the speed of the rate approaching the entropy.*

### 8.3.3 Result of this study

The next theorem shows the behavior of the Bayes code in the moderate deviation regime.

**Theorem 8.3.2** *Let $\{R_n\}_{n=1}^{\infty}$ be a sequence of rates such that*

$$R_n = \frac{1}{n} H_{\theta_*^k}(X^n) + \tau_n, \tag{8.22}$$

*where $\{\tau_n\}_{n=1}^{\infty}$ is a sequence satisfying*

$$\lim_{n\to\infty} \tau_n = +0, \tag{8.23}$$

$$\lim_{n\to\infty} \sqrt{n}\tau_n = \infty. \tag{8.24}$$

*Then, it holds that*

$$\lim_{n\to\infty} \frac{\ln \mathbb{P}_{\theta_*^k}[\ell_B(X^n) > nR_n]}{n\tau_n^2} = -\frac{1}{2\sigma_{\theta_*^k}^2(X)}. \tag{8.25}$$

(**Proof**) Equation (8.25) is proved by showing

$$\limsup_{n\to\infty} \frac{\ln \mathbb{P}_{\theta_*^k}[\ell_B(X^n) > nR_n]}{n\tau_n^2} \leq -\frac{1}{2\sigma_{\theta_*^k}^2(X)}, \tag{8.26}$$

$$\liminf_{n\to\infty} \frac{\ln \mathbb{P}_{\theta_*^k}[\ell_B(X^n) > nR_n]}{n\tau_n^2} \geq -\frac{1}{2\sigma_{\theta_*^k}^2(X)}. \tag{8.27}$$

From Theorem 8.3.1, it is easy to prove (8.27). Indeed, it holds that

$$\mathbb{P}_{\theta_*^k}[\ell_B(X^n) > nR_n] \geq \mathrm{P_e}(P_{\theta_*^k}, n, R) \tag{8.28}$$

from the definition of $\mathrm{P_e}(P_{\theta_*^k}, n, R_n)$, and (8.27) is obtained by combining (8.28) with (8.20).

Next, (8.26) is shown. Let $\{t_n\}_{n=1}^{\infty}$ be a sequence such that $t_n \geq 0$ for $n = 1, 2, \ldots$. Then, we have

$$\begin{aligned}
&\mathbb{P}_{\theta_*^k}[\ell_B(X^n) > nR_n] \\
&\overset{(a)}{=} \mathbb{P}_{\theta_*^k}\left[\ln \frac{1}{p_{\theta_*^k}(X^n)} + \frac{k}{2}\ln\frac{n}{2\pi} + O(\ln\ln n) > nR_n\right] &(8.29)\\
&\overset{(b)}{=} \mathbb{P}_{\theta_*^k}\left[\ln \frac{1}{p_{\theta_*^k}(X^n)} > H_{\theta_*^k}(X^n) + n\tau_n - \frac{k}{2}\ln\frac{n}{2\pi} + O(\ln\ln n)\right] &(8.30)\\
&\overset{(c)}{\leq} \exp\left\{-t_n\left(H_{\theta_*^k}(X^n) + n\tau_n - \frac{k}{2}\ln\frac{n}{2\pi} + O(\ln\ln n)\right)\right\}\\
&\quad \cdot E_{p_{\theta_*^k}}\left[\exp\left\{t_n \ln\frac{1}{p_{\theta_*^k}(X^n)}\right\}\right], &(8.31)
\end{aligned}$$

where

- $(a)$ follows from (7.26),

- $(b)$ follows from (8.22),
- $(c)$ is due to the Chernoff bound.

On the other hand, the Taylor expansion around $t_n = 0$ yields

$$\ln E_{p_{\theta_*^k}}\left[\exp\left\{t_n \ln \frac{1}{p_{\theta_*^k}(X^n)}\right\}\right] = H_{\theta_*^k}(X^n)t_n + \frac{1}{2}\sigma^2_{\theta_*^k}(X^n)t_n^2 + \sum_{m=3}^{\infty} \frac{\zeta^m_{\theta_*^k}(X^n)}{m!} t_n^m, \quad (8.32)$$

where $\zeta^m_{\theta_*^k}(X^n)$ denotes the $m$-th cumulant of $\ln(1/p_{\theta_*^k}(X^n))$.

Taking the logarithm of both sides of (8.31) and plugging (8.32) establish

$$\begin{aligned}
&\ln \mathbb{P}_{\theta_*^k}[\ell_B(X^n) > nR_n] \\
&\le -t_n \left(H_{\theta_*^k}(X^n) + n\tau_n - \frac{k}{2}\ln\frac{n}{2\pi} + O(\ln\ln n)\right) \\
&\quad + H_{\theta_*^k}(X^n)t_n + \frac{1}{2}\sigma^2_{\theta_*^k}(X^n)t_n^2 + \sum_{m=3} \frac{\zeta^m_{\theta_*^k}(X^n)}{m!} t_n^m \qquad (8.33) \\
&= -n\tau_n t_n + \frac{1}{2}\sigma^2_{\theta_*^k}(X^n)t_n^2 + \frac{k}{2}\ln\left(\frac{n}{2\pi}\right)t_n + O(\ln\ln n)t_n + \sum_{m=3} \frac{\zeta^m_{\theta_*^k}(X^n)}{m!} t_n^m. \qquad (8.34)
\end{aligned}$$

Dividing both sides of (8.34) by $n\tau_n^2$, we have

$$\begin{aligned}
\frac{\ln \mathbb{P}_{\theta_*^k}[\ell_B(X^n) > nR_n]}{n\tau_n^2} &\le -\frac{t_n}{\tau_n} + \frac{1}{2n\tau_n^2}\sigma^2_{\theta_*^k}(X^n)t_n^2 + \frac{k}{2n\tau_n^2}\ln\left(\frac{n}{2\pi}\right)t_n \qquad (8.35) \\
&\quad + \frac{O(\ln\ln n)t_n}{n\tau_n^2} + \frac{1}{n\tau_n^2}\sum_{m=3} \frac{\zeta^m_{\theta_*^k}(X^n)}{m!} t_n^m. \qquad (8.36)
\end{aligned}$$

In the following, we analyze the terms in the right-hand side of (8.36) by setting $t_n = \tau_n/\sigma^2_{\theta_*^k}(X)$ and using the assumption of $\tau_n$.

The first and second terms in the right-hand side of (8.36) are calculated as

$$-\frac{t_n}{\tau_n} + \frac{1}{2n\tau_n^2}\sigma^2_{\theta_*^k}(X^n)t_n^2 = -\frac{1}{\sigma^2_{\theta_*^k}(X)} + \frac{1}{2\sigma^2_{\theta_*^k}(X)} = -\frac{1}{2\sigma^2_{\theta_*^k}(X)}, \qquad (8.37)$$

where we use the fact that $\sigma^2_{\theta_*^k}(X^n) = n\sigma^2_{\theta_*^k}(X)$ because a stationary memoryless source is assumed (see, e.g., [4]).

Next, the third term in the right-hand side of (8.36) is evaluated as

$$\begin{aligned}
\frac{k}{2n\tau_n^2}\ln\left(\frac{n}{2\pi}\right)t_n &= \frac{k}{2n\tau_n\sigma^2_{\theta_*^k}(X)}\ln\left(\frac{n}{2\pi}\right) \qquad (8.38) \\
&\to 0 \quad (n \to \infty). \qquad (8.39)
\end{aligned}$$

Then, the fourth term in the right-hand side of (8.36) is evaluated as

$$\frac{O(\ln\ln n)t_n}{n\tau_n^2} = \frac{O(\ln\ln n)}{n\tau_n\sigma^2_{\theta^k_*}(X)} \tag{8.40}$$

$$\to 0 \quad (n\to\infty). \tag{8.41}$$

Finally, the fifth term in the right-hand side of (8.36) is calculated as

$$\frac{1}{n\tau_n^2}\sum_{m=3}^{\infty}\frac{\zeta^m_{\theta^k_*}(X^n)}{m!}t_n^m = \frac{1}{n}\sum_{m=3}^{\infty}\frac{\zeta^m_{\theta^k_*}(X^n)}{m!}\frac{\tau_n^{m-2}}{\left(\sigma^2_{\theta^k_*}(X)\right)^m} \tag{8.42}$$

$$\to 0 \quad (n\to\infty), \tag{8.43}$$

where we use the fact that $\zeta^m_{\theta^k_*}(X^n) = O(n)$ (see, e.g., [4]).

Thus, from (8.36)–(8.43), we obtain (8.26). □

## 8.4 Discussion and conclusion of this chapter

We compare the previous result (Theorem 8.3.1) and the main result (Theorem 8.3.2). The Bayes code is designed to minimize the *mean codeword length* averaged with the prior probability density function $w(\theta^k)$. Therefore, it is not designed to minimize the *overflow probability*. However, a comparison of (8.25) with (8.20) shows that the behavior of the overflow probability of the Bayes code is similar to that of the optimal (optimal means that the *overflow probability* is smallest) non-universal code.

When the sequence $\{\eta_n\}_{n=1}^{\infty}$ satisfies the condition (8.13), Theorem 8.2.3 does not give the speed of convergence of the overflow probability. Hence, one of the research questions is how is the speed of convergence when the sequence $\{\eta_n\}_{n=1}^{\infty}$ meets the condition (8.13). Theorem 8.3.2 gives one of the answers for this question. That is, if $\eta_n$ equals $nR_n = H_{\theta^k_*}(X^n) + n\tau_n$, then[5] the speed that the overflow probability of the Bayes code goes to zero is about $\exp\{-n\tau_n^2/2\sigma^2_{\theta^k_*}(X)\}$.

The summary of this chapter is as follows. First, this chapter analyzed the necessary and sufficient condition of the overflow probability of the Bayes code approaching zero asymptotically. To derive the result, the asymptotic normality of the Bayes code played an important role. Next, this chapter evaluated the overflow probability of the Bayes code in the moderate deviation regime. The result showed that the behavior of the overflow probability of the Bayes code is similar to that of the optimal non-universal code.

[5]Due to the condition of $\tau_n$ and the fact that $\sigma^2_{\theta^k_*}(X^n) = n\sigma^2_{\theta^k_*}(X)$ because a stationary memoryless source is assumed (see, e.g., [4]), it is easy to see that $\eta_n = H_{\theta^k_*}(X^n) + n\tau_n$ satisfies (8.13).

# Chapter 9

# Concluding remarks and future works

## 9.1 Concluding remarks

This dissertation mainly consists of two parts. The first part of this dissertation discussed the non-asymptotic analysis as well as the asymptotic analysis of the theoretical fundamental limits on the overflow probability for several source coding problems. Chapters 3, 4, 5, and 6 correspond to the first part.

In Chapter 3, we considered the variable-length lossless source coding problem for a general source. The non-asymptotic and asymptotic fundamental limits were characterized by using the smooth max entropy for both prefix codes and non-prefix codes. To show the achievability results, the explicit code construction was used. This technique was utilized throughout Chapters 3–5.

In Chapter 4, we considered the variable-length source coding allowing errors for a general source. The non-asymptotic and asymptotic fundamental limits were characterized by using the smooth max entropy. The obtained results indicated that the overflow probability and the error probability are trade-off. Further, by comparing the results in Chapter 3 with the results in this chapter, the benefit of allowing the positive overflow probability was clarified.

In Chapter 5, we considered the variable-length lossy source coding problem for a general source. The non-asymptotic and asymptotic fundamental limits were characterized by using the smooth max entropy-based quantity. The results clarified that the overflow probability and the excess distortion probability are trade-off.

In Chapter 6, we considered the fixed-length Slepian-Wolf coding problem for a general source. The second-order achievable rate region was characterized by using the function related to the smooth max entropy and the conditional smooth max entropy.

The second part of this dissertation discussed the asymptotic analysis of the overflow probability of the Bayes code. Chapters 7 and 8 correspond to the second part.

Chapter 7 considered the case where a positive overflow probability is allowed.

In this chapter, we evaluated the upper and lower bounds on the $\epsilon$-coding rate of the Bayes code for a stationary ergodic finite order Markov source. The main result showed that the Bayes code behaves similarly to the optimal code; the first and second terms of the upper bound and the lower bound are the same as those of the optimal code; the first term is $H_{\theta^k_*}(\mathbf{X})$, and the second term is $\frac{\sigma_{\theta^k_*}(\mathbf{X})}{\sqrt{n}}Q^{-1}(\epsilon)$.

Chapter 8 treated the case where the overflow probability of the Bayes codes approaches zero asymptotically. In this chapter, we analyzed the necessary and sufficient condition of the overflow probability of the Bayes code approaching zero asymptotically. Further, we evaluated the overflow probability of the Bayes code in the moderate deviation regime. Our result clarified that the behavior of the overflow probability of the Bayes code is similar to that of the optimal non-universal code.

## 9.2 Future works

In Chapters 3–6, we have derived various general formulas. However, from such general formulas, coding theorems for a source with a specific probabilistic structure have not obtained yet. For example, previous works such as [19] and [20] have obtained the fundamental limits for a stationary memoryless source after deriving the general formulas. Therefore, one of the future works is the derivation of a theoretical limit for a source with specific probabilistic structure (e.g., a stationary memoryless source and a stationary ergodic Markov source).

Regarding the studies in Chapter 7, one research direction is analyzing the $\epsilon$-coding rate of the Bayes code for various sources. For example, previous study such as [26] has analyzed the $\epsilon$-coding rate of the Bayes code for a piecewise stationary memoryless source. Another future work is the evaluation of the Bayes code in the moderate deviation regime for various sources.

# Bibliography


[1] Y. Altuğ, A. B. Wagner, and I. Kontoyiannis "Lossless compression with moderate error probability," Proc. IEEE Int. Symp. Inf. Theory (ISIT), Istanbul, Turkey, pp. 1744–1748, July 2013.

[2] B. S. Clarke and A. R. Barron, "Information-theoretic asymptotics of Bayes methods," IEEE Trans. Inf. Theory, vol.36, no.3, pp.453–471, May 1990.

[3] T. M. Cover and J. A. Thomas, Elements of Information Theory, 2nd ed., John Wiley & Sons, 2006.

[4] T. Courtade and S. Verdú, "Cumulant generating function of codeword lengths in optimal lossless compression," Proc. IEEE Int. Symp. Inf. Theory (ISIT), Honolulu, HI, USA, pp.2494–2498, June-July 2014.

[5] L. D. Davisson, "Universal noiseless coding," IEEE Trans. Inf. Theory, vol. 19, no. 6, pp. 783–795, Nov. 1973.

[6] A. Feinstein, "A new basic theorem of information theory," IRE Trans. Inf. Theory, vol.4, pp. 2–22, Sept. 1954.

[7] W. Feller, An Introduction to Probability and Its Applications, vol.1 and 2, John Wiley & Sons, 1966.

[8] M. Gotoh, T. Matsushima, and S. Hirasawa, "A generalization of B. S. Clarke and A. R. Barron's asymptotics of Bayes codes for FSMX sources," IEICE Trans. Fundamentals, vol.E81-A, no.10, pp.2123–2132, Oct. 1998.

[9] M. Goto, T. Matsushima, and S. Hirasawa, "An analysis of the difference of code lengths between two-step codes based on MDL principle and Bayes codes," IEEE Trans. Inf. Theory, vol.47, no.3, pp.927–944, March 2001.

[10] T. S. Han, Information-Spectrum Methods in Information Theory, Springer, 2003.

[11] M. Hayashi and S. Watanabe, "Non-asymptotic and asymptotic analyses on Markov chains in several problems," Proc. Information Theory and Applications Workshop (ITA), San Diego, CA, USA, pp. 1–10, Feb. 2014.

[12] R. V. Hogg, J. W. McKean, and A. T. Craig, Introduction to Mathematical Statistics, 6th ed., Pearson Prentice Hall, 2005.

[13] T. Holenstein and R. Renner, "On the randomness of independent experiments," IEEE Trans. Inf. Theory, vol. 57, no. 4, pp. 1865-1871, Apr. 2011.

[14] D. A. Huffman, "A method for the construction of minimum redundancy codes," Proc. IRE, vol. 40, no. 9, pp. 1098–1101, Sept. 1952.

[15] N. Iri and O. Kosut, "Third-order coding rate for universal compression of Markov sources," Proc. IEEE Int. Symp. Inf. Theory (ISIT), Hong Kong, China, pp. 1996–2000, June 2015.

[16] H. Koga, "Characterization of the smooth Rényi entropy using majorization," Proc. Information Theory Workshop (ITW), Seville, Spain, pp. 1–5, Sept. 2013.

[17] I. Kontoyiannis, "Second-order noiseless source coding theorems," IEEE Trans. Inf. Theory, vol.43 no.4, pp.1339-1341, July 1997.

[18] 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.

[19] V. Kostina and S. Verdú, "Fixed-length lossy compression in the finite blocklength regime," IEEE Trans. Inf. Theory, vol. 58, no. 6, pp. 3309–3338, June 2012.

[20] 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.

[21] O. Kosut and L. Sankar, "Universal fixed-to-variable source coding in the finite blocklength regime," Proc. IEEE Int. Symp. Inf. Theory (ISIT), Istanbul, Turkey, pp.649–653, July 2013.

[22] O. Kosut and L. Sankar, "New results on third-order coding rate for universal fixed-to-variable source coding," Proc. IEEE Int. Symp. Inf. Theory (ISIT), Honolulu, HI, USA, pp.2689-2693, June–July 2014.

[23] G. G. Langdon and J. J. Rissanen, "A simple general binary source code," IEEE Trans. Inf. Theory, vol. 28, no. 5, pp. 800–803, Sept. 1982.

[24] T. Matsushima, H. Inazumi, and S. Hirasawa, "A class of distortionless codes designed by Bayes decision theory," IEEE Trans. Inf. Theory, vol.37, no.5, pp.1288–1293, Sept. 1991.

[25] S. Miyake and F. Kanaya,"Coding theorems on correlated general sources," IEICE Trans. Fundamentals, vol. E78-A, no. 9, pp. 1063-1070, Sept. 1995.

[26] T. Moriya, S. Saito, N. Miya, and T. Matsushima, "Evaluation of the minimum overflow threshold of Bayes codes for non-stationary sources," Proc. the 37th Symposium on Information Theory and its Applications (SITA), Toyama, Japan, pp.22–27, Dec. 2014 (in Japanese).

[27] S. V. Nagaev, "More exact statements of limit theorems for homogeneous Markov chains," Theory Probab. Appl., vol. 6, no. 1, pp.62-81, 1961.

[28] R. Nomura, T. Matsushima, and S. Hirasawa, "On the probability of length overflow of lossless codes," IEICE Trans. Fundamentals, vol.J90-A, no.4, pp.292–304, Apr. 2007 (in Japanese).

[29] R. Nomura, T. Matsushima, and S. Hirasawa, "A note on the $\epsilon$-overflow probability of lossless codes," IEICE Trans. Fundamentals, vol.E90-A, no.12, pp.2965-2970, Dec. 2007.

[30] R. Nomura, T. Matsushima, and S. Hirasawa, "Information spectrum approach to overflow probability of variable-length codes with conditional cost function," Proc. IEEE Int. Symp. Inf. Theory (ISIT), Cambridge, MA, USA, pp. 1097–1101, July 2012.

[31] R. Nomura and T. S. Han, "Second-order Slepian-Wolf coding theorems for non-mixed and mixed sources," IEEE Trans. Inf. Theory, vol. 60, no. 9, pp. 5553–5572, Sept. 2014.

[32] R. Nomura and H. Yagi, "Variable-length lossy source coding allowing some probability of union of overflow and excess distortion," Proc. IEEE Int. Symp. Inf. Theory (ISIT), Barcelona, Spain, pp. 2958–2962, July 2016.

[33] R. Nomura and H. Yagi, "Overflow probability of variable-length codes allowing non-vanishing error probability," Proc. IEEE Information Theory Workshop (ITW), Kaohsiung, Taiwan, Nov. 2017.

[34] R. Renner and S. Wolf, "Smooth Rényi entropy and applications," Proc. IEEE Int. Symp. Inf. Theory (ISIT), Chicago, USA, page 232, June-July 2004.

[35] R. Renner and S. Wolf, "Simple and tight bounds for information reconciliation and privacy amplification," Proc. ASIACRYPT, LNCS 3788, Chennai (Madras), India, pp. 199-216, Dec. 2005.

[36] R. Renner, S. Wolf, and J. Wullschleger, "Trade-offs in information-theoretic multi-party one-way key agreement," International Conference on Information Theoretic Security (ICITS), LNCS 4883, Madrid, Spain, pp. 65– 75, May 2007.

[37] A. Rényi, "On measures of entropy and information," 4th Berkley Symposium on Mathematics, Statistics and Probability, pp. 547–561, 1960.

[38] J. J. Rissanen, "Generalized Kraft inequality and arithmetic coding," IBM Journal of Research and Development, vol. 20, no. 3, pp. 198–203, May 1976.

[39] C. E. Shannon, "A mathematical theory of communication," Bell System Technical Journal, vol. 27, pp. 379–423, pp. 623–656, 1948.

[40] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," IEEE Trans. Inf. Theory, vol. 19, no. 9, pp. 471–480, July 1973.

[41] W. Spankowski and S. Verdú, "Minimum expected length of fixed-to-variable lossless compression without prefix constraints," IEEE Trans. Inf. Theory, vol. 57, no. 7, pp. 4017–4025, July 2011.

[42] O. Uchida and T. S. Han, "The optimal overflow and underflow probabilities of variable-length coding for the general source," IEICE Trans. Fundamentals, vol. E84-A, no. 10, pp. 2457–2465, Oct. 2001.

[43] T. Uyematsu, "A new unified method for fixed-length source coding problems of general sources," IEICE Trans. Fundamentals, vol. E93-A, no. 11, pp. 1868–1877, Nov. 2010.

[44] T. Uyematsu, "Relating source coding and resolvability: a direct approach," Proc. IEEE Int. Symp. Inf. Theory (ISIT), Austin, USA, pp. 1350–1354, June 2010.

[45] T. Uyematsu and S. Kunimatsu, "A new unified method for intrinsic randomness problems of general sources," Proc. IEEE Information Theory Workshop (ITW), Sevilla, Spain, pp. 1–5, Sept. 2013.

[46] T. Uyematsu and T. Matsuta, "Revisiting the Slepian-Wolf coding problem for general sources: a direct approach," Proc. IEEE Int. Symp. Inf. Theory (ISIT), Honolulu, HI, USA, pp. 1336–1340, June–July 2014.

[47] T. Uyematsu and T. Matsuta, "Source coding with side information at the decoder revisited," Proc. IEEE Int. Symp. Inf. Theory (ISIT), Hong Kong, China, pp. 1570–1574, June 2015.

[48] N. A. Warsi, "One-shot source coding with coded side information available at the decoder," Proc. IEEE Int. Symp. Inf. Theory (ISIT), Istanbul, Turkey, pp. 3070–3074, July 2013.

[49] H. Yagi and R. Nomura, "Variable-length coding with epsilon-fidelity criteria for general sources," Proc. IEEE Int. Symp. Inf. Theory (ISIT), Hong Kong, China, pp. 2181–2185, June 2015.

# List of Publications

| | Title, Transaction or Proceeding, Publication Date, Authors |
|---|---|
| | (**Journals**) |
| 1. ○ | Evaluation of Overflow Probability of Bayes Code in Moderate Deviation Regime, IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, vol.E100-A, no.12, pp.2728–2731, Dec. 2017, Shota Saito, Toshiyasu Matsushima |
| 2. | Spatially "Mt. Fuji" Coupled LDPC Codes, IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, vol.E100-A, no.12, pp.2594–2606, Dec. 2017, Yuta Nakahara, Shota Saito, Toshiyasu Matsushima |
| 3. ○ | Threshold of Overflow Probability Using Smooth Max-Entropy in Lossless Fixed-to-Variable Length Source Coding for General Sources, IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, vol.E99-A, no.12, pp.2286–2290, Dec. 2016 , Shota Saito, Toshiyasu Matsushima |
| 4. ○ | Second-Order Achievable Rate Region of Slepian-Wolf Coding Problem in Terms of Smooth Max-Entropy for General Sources, IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, vol.E99-A, no.12, pp. 2275–2280, Dec. 2016, Shota Saito, Toshiyasu Matsushima |
| 5. ○ | Evaluation of the Bayes Code from Viewpoints of the Distribution of Its Codeword Lengths, IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, vol.E98-A, no.12, pp.2407–2414, Dec. 2015, Shota Saito, Nozomi Miya, Toshiyasu Matsushima |

| | Title, Transaction or Proceeding, Publication Date, Authors |
|---|---|
| | (**International Conferences (with review)**) |
| 6. ○ | Variable-Length Lossy Compression Allowing Positive Overflow and Excess Distortion Probabilities, 2017 IEEE International Symposium on Information Theory (ISIT2017), Aachen, Germany, June 2017, Shota Saito, Hideki Yagi, Toshiyasu Matsushima |
| 7. ○ | Threshold of Overflow Probability in Terms of Smooth Max-Entropy for Variable-Length Compression Allowing Errors, 2016 International Symposium on Information Theory and Its Applications (ISITA2016), Monterey, California, USA, Oct.-Nov. 2016, Shota Saito, Toshiyasu Matsushima |
| 8. ○ | Evaluation of Overflow Probability of Bayes Code in Moderate Deviation Regime, 2016 International Symposium on Information Theory and Its Applications (ISITA2016), Monterey, California, USA, Oct.-Nov. 2016, Shota Saito, Toshiyasu Matsushima |
| 9. | Spatially "Mt. Fuji" Coupled LDPC Codes, 2016 International Symposium on Information Theory and Its Applications (ISITA2016), Monterey, California, USA, Oct.-Nov. 2016 , Yuta Nakahara, Shota Saito, Toshiyasu Matsushima |
| 10. ○ | Fundamental limit and pointwise asymptotics of the Bayes code for Markov sources, 2015 IEEE International Symposium on Information Theory (ISIT2015), Hong Kong, June 2015, Shota Saito, Nozomi Miya, Toshiyasu Matsushima |
| 11. ○ | Evaluation of the minimum overflow threshold of Bayes codes for a Markov source, 2014 International Symposium on Information Theory and Its Applications (ISITA2014), Melbourne, Oct. 2014, Shota Saito, Nozomi Miya, Toshiyasu Matsushima |

| | Title, Transaction or Proceeding, Publication Date, Authors |
|---|---|
| | (**Domestic Conferences**) |
| 12. | Variable-Length Intrinsic Randomness Problem Allowing Non-Vanishing Underflow Probability (in Japanese), IEICE Technical Report (IT), Chiba, Japan, July 2017, Jun Yoshizawa, Shota Saito, Toshiyasu Matsushima |
| 13. | Variable-Length Lossy Compression Allowing Positive Overflow and Excess Distortion Probabilities, Proceedings of 39th Symposium on Information Theory and its Applications (SITA2016), Gifu, Japan, Dec. 2016, Shota Saito, Hideki Yagi, Toshiyasu Matsushima |
| 14. | A Study on Message Passing Algorithm for Counting Short Cycles in Sparse Bipartite Graphs (in Japanese), IEICE Technical Report (IT), Osaka, Japan, Jan. 2016, Yuta Nakahara, Shota Saito, Toshiyasu Matsushima |
| 15. | A Note on the Optimal Data Prediction under Bayes Criterion for Multidimensional Linear Regression Models Assuming Latent Variables (in Japanese), IEICE Technical Report (PRMU), Osaka, Japan, Jan. 2016, Mikio Ushioda, Shota Saito, Toshiyasu Matsushima |
| 16. | A Note on the Computational Complexity Reduction Method of the Optimal Prediction under Bayes Criterion in Semi-Supervised Learning (in Japanese), IEICE Technical Report (PRMU), Osaka, Japan, Jan. 2016, Yuto Nakano, Shota Saito, Toshiyasu Matsushima |
| 17. | On the Achievable Overflow Threshold of Variable-Length Coding Using Smooth Max-Entropy for General Sources (in Japanese), Proceedings of 38th Symposium on Information Theory and its Applications (SITA2015), Okayama, Japan, Nov. 2015, Shota Saito, Toshiyasu Matsushima |

| | Title, Transaction or Proceeding, Publication Date, Authors |
|---|---|
| 18. | Variations of the Strong Converse Theorem on the Intrinsic Randomness Problem for General Sources (in Japanese), IEICE Technical Report (IT), Tokyo, Japan, July 2015, Shota Saito, Toshiyasu Matsushima |
| 19. | Another Representation on the Second-Order Achievable Rate Region of Slepian-Wolf Coding Problem for General Sources (in Japanese), IEICE Technical Report (IT), Fukuoka, Japan, Mar. 2015, Shota Saito, Nozomi Miya, Toshiyasu Matsushima |
| 20. | Evaluation of the Minimum Overflow Threshold of Bayes Codes for Non-Stationary Sources (in Japanese), Proceedings of 37th Symposium on Information Theory and its Applications (SITA2014), Toyama, Japan, Dec. 2014, Takashi Moriya, Shota Saito, Nozomi Miya, Toshiyasu Matsushima |
| 21. | Asymptotic Normality and LIL of the Codeword Length of Bayes Codes for Stationary Ergodic Markov Sources (in Japanese), Proceedings of 37th Symposium on Information Theory and its Applications (SITA2014), Toyama, Japan, Dec. 2014, Shota Saito, Nozomi Miya, Toshiyasu Matsushima |
| 22. | Universality of Spatially Coupled Punctured LDPC Codes for Decode-and-Forward in Erasure Relay Channels (in Japanese), Proceedings of 37th Symposium on Information Theory and its Applications (SITA2014), Toyama, Japan, Dec. 2014, Yuta Nakahara, Shota Saito, Akira Kamatsuka, Toshiyasu Matsushima |
| 23. | A Note on Attack Against Nonlinear Combiner Generator Using Sum - Product Algorithm (in Japanese), IEICE Technical Report (IBISML), Aichi, Japan, Nov. 2014, Kota Kubo, Shota Saito, Akira Kamatsuka, Toshiyasu Matsushima |
| 24. | Evaluation of Minimum Overflow Threshold for Bayes Codes (in Japanese), Proceedings of 36th Symposium on Information Theory and its Applications (SITA2013), Shizuoka, Japan, Nov. 2013, Shota Saito, Nozomi Miya, Ryo Nomura, Toshiyasu Matsushima |