Guessing and Compression : A Large Deviations Approach

A Thesis

Submitted for the Degree of

Master of Science (Engineering)

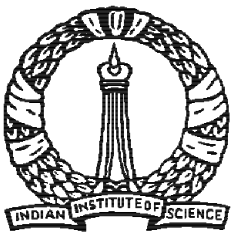
in the **Faculty of Engineering**

by

Manjesh Kumar Hanawal

under the guidance of

Dr. Rajesh Sundaresan



Electrical Communication Engineering

Indian Institute of Science

Bangalore – 560 012 (INDIA)

February 2009

Abstract

The problem of guessing a random string is studied. It arises in the analysis of the strength of secret-key cryptosystems against guessing attacks. Expected number of guesses, or more generally moments of the number of guesses needed to break the cryptosystem grow exponentially with the length of the string. This thesis studies the rate of exponential growth of these moments using the theory of large deviations.

A close relation between guessing and compression is first established. For systems with large key rates, it is shown that if the source's sequence of so-called information spectrum random variables satisfies the large deviation property with a certain rate function, then the limiting guessing exponent exists and is a scalar multiple of the Legendre-Fenchel dual of the rate function. This is then used to rederive several prior results. The large deviations approach brings to light the relevance of information spectrum in determining guessing exponents.

For systems with key-rate constraints, bounds are derived on the limiting guessing exponents for general sources. The obtained bounds are shown to be tight for stationary memoryless, Markov, and unifilar sources, thus recovering some known results. The bounds are obtained by establishing a close relationship between error exponents and correct decoding exponents for fixed rate source compression on the one hand and exponents for guessing moments on the other.

Acknowledgments

I am indebted to my advisor Dr. Rajesh Sundaresan for his valuable guidance and support throughout the duration of my graduation. All the problems considered in thesis are his suggestions. A significant part of results were proved during frequent discussions with him. I thank him for showing lot of patience and giving me enough time to understand the problems. Whenever I made initial progress he always ensured that the final results are quickly obtained and are in the best possible form. He had been always a source of new ideas whenever I had hit road locks. His professional attitude, time management, clarity of thinking, enthusiasm are a great source of inspiration for any one. Apart from regular courses he constantly encouraged me to attend other courses, conferences, workshops and talks which widened my knowledge base. I always cherished every moments with him including those of early morning jogging.

I am fortunate to have been taught by the finest teachers at IISc. I thank all my teachers at IISc especially, from Department of ECE, Department of Mathematics, and Department of CSA. I always admired their tireless efforts to impart knowledge to students. Studies at IISc is so far one of the defining moments in my life.

My father and mother have been the reasons for my success. As ever, they continuously encouraged me to put sincere and hard efforts whenever I expressed unhappiness about my progress. I sincerely thank my parents for giving freedom in pursuing my goals.

I thank my lab mates Arun Padakandla, Darshan, Ajay, Naveen Deshpande, Nidhin, Renu, Krishna Kumar, Ashok for all the technical discussion we had, which always gave me more insight into my research topic. I also thank my hostel mates Srinidi, Bharat, Naveen K P and A-mess dining friends for making my stay at IISc a pleasant one.

This work was supported by the Defence Research and Development Organisation, Ministry of Defence, Government of India, under the DRDO-IISc Programme on Advanced Research in Mathematical Engineering, and by the University Grants Commission under Grant Part (2B) UGC-CAS-(Ph.IV).

Contents

Chapter 1

Introduction

Let X be a random variable taking values on a finite set $\mathbb{X}$. Alice conducts a random experiment and obtains outcome x. Bob is interested in determining the outcome of this experiment, knowing only the alphabet set $\mathbb{X}$ and distribution of the random variable. He submits a sequence of guesses to Alice stepping through the elements in $\mathbb{X}$. Alice replies either "Yes" or "No" to each of Bob's guesses and charges a fixed amount for each guess. Bob should determine the realization with as few guesses as possible to minimise his cost. There are several applications that motivate this problem. Consider cipher systems employed in digital television or DVDs to block unauthorized access to special features. The ciphers used are amenable to such exhaustive guessing attacks and it is of interest to quantify the effort needed by an attacker.

More generally, consider the classical cipher system of Shannon [1] shown in figure 1.1. Let X^n be a message string of length n taking values on $\mathbb{X}^n$. This message should be communicated securely from a transmitter to a receiver, both of which have access to a common secure key U^k of k purely random bits independent of X^n. The transmitter computes the cryptogram $Y = f_n(X^n, U^k)$ and sends it to the receiver over a public channel. The cryptogram may be of variable length, and $R = k/n$ is the key rate of the system. The encryption function f_n is invertible for any fixed U^k. The receiver, knowing Y and U^k, computes $X^n = f_n^{-1}(Y, U^k)$. The functions f_n and f_n^{-1} are published. A wiretapping attacker has access to the cryptogram Y, knows f_n and f_n^{-1}, and attempts

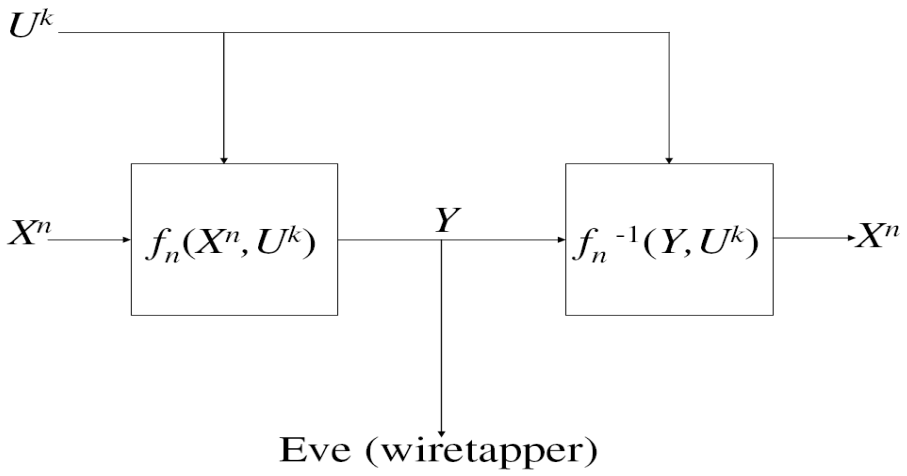


Figure 1.1: **Shannon Cipher System**

to identify X^n without knowledge of U^k. The attacker can use knowledge of the statistics of X^n. The attacker now aims to identify the message using as few guesses as possible, having now tapped the cryptogram (a problem studied by Merhav & Arikan [2]). We assume that the attacker has a test mechanism that tells him whether a guess $\hat{X}^n$ is correct or not. For example, the attacker may wish to attack an encrypted password or personal information to gain access to, say, a computer account, or a bank account via internet, or a classified database [2]. In these situations, successful entry into the system provides the natural test mechanism. We assume that the attacker is allowed an unlimited number of guesses.

In this thesis, we analyse the performance of such guessing attacks using the theory of large deviations.

1.1 Prior work

Let P_n denote the probability mass function (pmf) of the message strings taking values on $\mathbb{X}^n$. Consider the situation where the key rate is sufficiently large (for example, larger than $\log|\mathbb{X}|$) so as to render the cryptogram useless to the attacker. In this case we have *perfect secrecy*. The attacker then simply submits guesses based on the knowledge of P_n.

Massey [3] observed that the expected number of guesses needed to identify the realization is minimized by guessing in the decreasing order of P_n-probabilities. Define the *guessing function* $G_n^* : \mathbb{X}^n \to \{1, 2, \cdots, |\mathbb{X}|^n\}$ to be one such optimal guessing order[1]. $G_n^*(x^n) = g$ implies that x^n is the gth guess. Arikan [4] considered the growth of $\mathbb{E}\left[G_n^*(X^n)^\rho\right]$ as a function of n for an independent and identically distributed (iid) source with marginal pmf P_1 and $\rho > 0$. Moments of guessing are of interest because they describe the tail behavior of $G_n^*(X^n)$; see end of this section. Arikan [4] showed that the growth is exponential in n; the limiting exponent which we will call guessing exponent

$$E(\rho) := \lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}[G_n^*(X^n)^\rho] \tag{1.1}$$

exists and equals $\rho H_\alpha(P_1)$ with $\alpha = 1/(1+\rho)$, where $H_\alpha(P_n)$ is the Rényi entropy of order α for the pmf P_n, given by

$$\frac{1}{1-\alpha} \log \left(\sum_{x^n \in \mathbb{X}^n} P_n(x^n)^\alpha \right), \ \alpha \neq 1. \tag{1.2}$$

Arikan & Merhav remark that their proof in [5] for the limiting guessing exponent is equally applicable to finding the limiting exponent of the moment generating function of compression lengths. Moreover, the two exponents are the same. The latter is a problem studied by Campbell [6].

Malone & Sullivan [7] showed that the limiting exponent $E(\rho)$ of an irreducible Markov chain exists and equals the logarithm of the *Perron-Frobenius eigenvalue* of a matrix formed by raising each element of the transition probability matrix to the power α. From their proof, one obtains the more general result that the limiting exponent exists for any source if the Rényi entropy *rate* of order α,

$$\lim_{n\to\infty} n^{-1} H_\alpha(P_n), \tag{1.3}$$

[1] If there are several strings with the same probability of occurrence, they may be guessed in any order without affecting the expected number of guesses.

exists for $\alpha = 1/(1+\rho)$. Pfister & Sullivan [8] showed the existence of (1.1) for a class of stationary probability measures where the probability of finite-length strings are approximately determined by letter combinations. For such a class, they showed that the guessing exponent has a variational characterization (see (2.23) later). For unifilar sources Sundaresan [9] obtained a simplification of this variational characterization using a direct approach and the method of types.

Merhav & Arikan [2] studied discrete memoryless sources (DMS) for all positive key rates and characterized the best attainable guessing moments required by an attacker. In particular, they showed that for a DMS with governing single letter pmf P_1 on $\mathbb{X}$, the value of the optimal exponent for the ρth moment ($\rho > 0$) is given by

$$E(R,\rho) = \max_{Q} \{\rho \min\{H(Q), R\} - D(Q \parallel P_1)\}. \tag{1.4}$$

The maximization is over all pmfs Q on $\mathbb{X}$, $H(Q)$ is the Shannon entropy of Q, and $D(Q \parallel P)$ is the Kullback-Leibler divergence between Q and P. They also showed that $E(R,\rho)$ increases linearly in R for $R \leq H(P)$, continues to increase in a concave fashion for $R \in [H(P), H']$, where H' is a threshold, and is constant for $R > H'$. Unlike the classical equivocation rate analysis, atypical sequences do affect the behavior of $E(R,\rho)$ for $R \in [H(P), H']$ and perfect secrecy is obtained, i.e., cryptogram is uncorrelated with the message, only for $R > H' > H(P)$. Perfect secrecy is thus clearly obtained when $R > \log|\mathbb{X}|$. Merhav & Arikan also determined the best achievable performance based on the probability of a large deviation in the number of guesses, i.e., the tail behavior of $G_n^*(X^n)$, and showed that the corresponding exponent equals the Legendre-Fenchel transform of $E(R,\rho)$ as a function of ρ. Sundaresan [9] extended the above results to unifilar sources. Hayashi & Yamamoto [10] proved coding theorems for the Shannon cipher system with correlated outputs (X^n, Z^n) where the wiretapper is interested in X^n while the receiver in Z^n.

1.2 Our contribution

We give a large deviations perspective to these results, shed further light on the aforementioned connection between compression and guessing, and unify all prior results on existence of limiting guessing exponents. Specifically, we show that if the sequence of distributions of the *information spectrum* $(1/n)\log(1/P_n(X^n))$ (see Han [11]) satisfies the *large deviation property*, then the limiting exponent exists. This is useful because several existing large deviations results can be readily applied. We then show that all previously considered cases in the literature (without side information and key-rate constraints) satisfy this sufficient condition. The large deviation ideas are already present in the works of Pfister & Sullivan [8] and the method of types approach of Arikan & Merhav [5]. Our work brings out the essential ingredient (the sufficient condition on the information spectrum) and enables us to see all these specific results under one light.

Further, we extend Merhav & Arikan's notion of computational secrecy to general sources. One motivation is that secret messages typically come from the natural languages which can be well-modelled as sources with memory, for e.g., a Markov source of appropriate order. Another motivation is that the study of general sources clearly brings out the connection between guessing and compression, as discussed next.

As with other studies of general sources, *information spectrum* plays a crucial role in this thesis. We show that $E(R,\rho)$ is closely related to

(a) the error exponent of a rate-R source code

(b) the correct decoding exponent of a rate-R source code, when exponentiated probabilities are considered (see Sec.3.3.2).

In particular, the exponents in (a) and (b) appear in the first and second terms when we rewrite $E(R,\rho)$ for a DMS as

$$E(R,\rho) = \max\left\{\rho R - \min_{Q:H(Q)>R} D(Q \parallel P), \min_{Q:H(Q)\leq R}\{\rho H(Q) - D(Q \parallel P)\}\right\}.$$

This brings out the fundamental connection between source coding exponents and key-rate constrained guessing exponents. Further, unlike the case for the probability of a large deviation in the number of guesses [2, Sec. V], both the error exponent and the correct decoding exponent determine $E(R, \rho)$. We extend the above result to general sources by getting upper and lower bounds on $E(R, \rho)$. We then show that these are tight for DMS, Markov, and unifilar sources. The bounds may be of interest even if they are not tight because the upper bound specifies the amount of effort need by an attacker and the lower bound specifies the secrecy strength of the cryptosystem to a designer.

The limiting case as $\rho \downarrow 0$ in (b) yields the classical framework for probability of correct decoding. This special case is related to the work of Han [12] and Iriyama [13] who studied the dual problem of rates required to meet a specified error exponent or a specified correct decoding exponent.

1.3 Organization of the thesis

This thesis is organized as follows. In chapter 2 we focus attention on the system with perfect secrecy. We first study the tight relationship between guessing and compression. We state the relevant large deviations results and the main sufficiency condition. We then re-derive prior results by showing that each case satisfies the sufficient condition. All the proofs are relegated to the end of the chapter. We then conclude the chapter with a concise summary.

In chapter 3 we consider guessing for key-rate constrained systems. We first define the problem precisely and relate it to a modification of Campbell's compression problem [6]. We give bounds on the limits of exponential rate of guessing moments, in terms of information spectrum quantities. We then evaluate these bounds for some specific examples to recover prior results. Yet again, all proofs of claimed results of this chapter are relegated to the end of the chapter. We then conclude the chapter with a summary.

In chapter 4 we make some concluding remarks and discuss some open problems.

Chapter 2

Perfect Secrecy

Let $X^n = (X_1, \cdots, X_n)$ denote n letters of a process where each letter is drawn from a finite set $\mathbb{X}$. The joint probability mass function (pmf) is given by $(P_n(x^n) : x^n \in \mathbb{X}^n)$. Let x^n be a realization and suppose that we wish to guess this realization by asking questions of the form "Is $X^n = x^n$?", stepping through the elements of $\mathbb{X}^n$ until the answer is "Yes". A guessing function

$$G_n : \mathbb{X}^n \to \{1, 2, \cdots, |\mathbb{X}|^n\}$$

is a bijection that denotes the order in which the elements of $\mathbb{X}^n$ are guessed. If $G_n(x^n) = g$, then the gth guess is x^n. We wish to minimize the expected number of guesses, i.e., $\mathbb{E}_{P_n}[G_n(X^n)]$ where the expectation[1] is with respect to P_n. As observed by Massey [3] the expected number of guesses is minimized by guessing in the decreasing order of P_n-probabilities. We denote this optimal guessing strategy by G_n^*.

Our interest in this chapter is to study the exponential growth rate of moments, $\mathbb{E}[G_n^*(X^n)^\rho]$ for a given $\rho > 0$, i.e, to study

$$E(\rho) := \lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}[G_n^*(X^n)^\rho],$$

whenever limit exists[2]. Before we do so we establish equivalence between the problem of

[1] In the subsequent sections we drop subscript P_n and use the notation $\mathbb{E}[G_n(X^n)]$ to mean expectation with respect to P_n when there is no ambiguity.

[2] Results in this chapter can be found in [14] and [15].

guessing and source compression.

2.1 Guessing and Compression

In this section we relate the problem of guessing to one of source compression. An interesting conclusion is that robust source compression strategies lead to robust guessing strategies.

For ease of exposition, let us assume that the message space is simply $\mathbb{X}$. The extension to strings of length n is straightforward. Let $\mathbb{N}$ denote the set of natural numbers. A length function $L : \mathbb{X} \to \mathbb{N}$ is one that satisfies Kraft's inequality

$$\sum_{x \in \mathbb{X}} \exp\{-L(x)\} \leq 1. \tag{2.1}$$

To each guessing function G, we associate a pmf Q_G on $\mathbb{X}$ and a length function L_G as follows.

Definition 2.1.1 *Given a guessing function G, we say Q_G defined by*

$$Q_G(x) = c^{-1} \cdot G(x)^{-1}, \ \forall x \in \mathbb{X}, \tag{2.2}$$

is the pmf on $\mathbb{X}$ associated with G. The quantity c in (2.2) is the normalization constant. We say L_G defined by

$$L_G(x) = \lceil -\log Q_G(x) \rceil \, , \ \forall x \in \mathbb{X}, \tag{2.3}$$

is the length function associated with G.

Observe that

$$c = \sum_{a \in \mathbb{X}} G(a)^{-1} = \sum_{i=1}^{|\mathbb{X}|} \frac{1}{i} \leq 1 + \ln |\mathbb{X}|, \tag{2.4}$$

and therefore the pmf in (2.2) is well-defined. We record the intimate relationship between these associated quantities in the following result. (This is also available in the proof of [16, Th. 1, p.382]).

Proposition 2.1.2 *Given a guessing function G, the associated quantities satisfy*

$$c^{-1} \cdot Q_G(x)^{-1} = G(x) \leq Q_G(x)^{-1}, \tag{2.5}$$

$$L_G(x) - 1 - \log c \leq \log G(x) \leq L_G(x). \tag{2.6}$$

Proof: The first equality in (2.5) follows from the definition in (2.2), and the second inequality from the fact that $c \geq 1$.

The upper bound in (2.6) follows from the upper bound in (2.5) and from (2.3). The lower bound in (2.6) follows from

$$\begin{aligned} \log G(x) &= \log \left(c^{-1} \cdot Q_G(x)^{-1} \right) \\ &= -\log Q_G(x) - \log c \\ &\geq (\lceil -\log Q_G(x) \rceil - 1) - \log c \\ &= L_G(x) - 1 - \log c. \end{aligned}$$

■

We now associate a guessing function G_L to each length function L.

Definition 2.1.3 *Given a length function L, we define the associated guessing function G_L to be the one that guesses in the increasing order of L-lengths. Messages with the same L-length are ordered using an arbitrary fixed rule, say the lexicographical order on $\mathbb{X}$. We also define the associated pmf Q_L on $\mathbb{X}$ to be*

$$Q_L(x) = \frac{\exp\{-L(x)\}}{\sum_{a \in \mathbb{X}} \exp\{-L(a)\}}. \tag{2.7}$$

Proposition 2.1.4 *For a length function L, the associated pmf and the guessing function satisfy the following:*

1. *G_L guesses messages in the decreasing order of Q_L-probabilities;*

2.

$$\log G_L(x) \leq \log Q_L(x)^{-1} \leq L(x). \tag{2.8}$$

Proof: The first statement is clear from the definition of G_L and from (2.7).

Letting $1\{E\}$ denote the indicator function of an event E, we have as a consequence of statement 1) that

$$\begin{aligned} G_L(x) &\leq \sum_{a\in\mathbb{X}} 1\left\{Q_L(a) \geq Q_L(x)\right\} \\ &\leq \sum_{a\in\mathbb{X}} \frac{Q_L(a)}{Q_L(x)} \\ &= Q_L(x)^{-1}, \end{aligned} \tag{2.9}$$

which proves the left inequality in (2.8). This inequality was known to Wyner [17].

The last inequality in (2.8) follows from (2.7) and Kraft's inequality (2.1) as follows:

$$Q_L(x)^{-1} = \exp\{L(x)\} \cdot \sum_{a\in\mathbb{X}} \exp\{-L(a)\} \leq \exp\{L(x)\}.$$

■

Let $\{L(x) \geq B\}$ denote the set $\{x \in \mathbb{X} \mid L(x) \geq B\}$. We then have the following easy to verify corollary to Propositions 2.1.2 and 2.1.4.

Corollary 2.1.5 *For any G, its associated length function L_G, and any $B \geq 1$, we have*

$$\begin{aligned} &\{L_G(x) \geq B + 1 + \log c\} \\ &\quad\subseteq \{G(x) \geq \exp\{B\}\} \\ &\quad\subseteq \{L_G(x) \geq B\}. \end{aligned} \tag{2.10}$$

Analogously, for any L, its associated guessing function G_L, and any $B \geq 1$, we have

$$\{G_L(x) \geq \exp\{B\}\} \subseteq \{L(x) \geq B\}. \tag{2.11}$$

The inequalities between the associates in (2.6) and (2.8) indicate the direct relationship between guessing moments and Campbell's coding problem [6], and that the Rényi entropies are the optimal growth exponents for guessing moments, as highlighted in the

following Proposition.

Proposition 2.1.6 *Let L be any length function on $\mathbb{X}$, G_L the guessing function associated with L, P a pmf on $\mathbb{X}$, $\rho \in (0, \infty)$, L^* the length function that minimizes $\mathbb{E}[\exp\{\rho L^*(X)\}]$, where the expectation is with respect to P, G^* the guessing function that proceeds in the decreasing order of P-probabilities and therefore the one that minimizes $\mathbb{E}[G^*(X)^\rho]$, and c as in (2.4). Then*

$$\frac{\mathbb{E}[G_L(X)^\rho]}{\mathbb{E}[G^*(X)^\rho]} \leq \frac{\mathbb{E}[\exp\{\rho L(X)\}]}{\mathbb{E}[\exp\{\rho L^*(X)\}]} \cdot \exp\{\rho(1 + \log c)\}. \tag{2.12}$$

Analogously, let G be any guessing function, and L_G its associated length function. Then

$$\frac{\mathbb{E}[G(X)^\rho]}{\mathbb{E}[G^*(X)^\rho]} \geq \frac{\mathbb{E}[\exp\{\rho L_G(X)\}]}{\mathbb{E}[\exp\{\rho L^*(X)\}]} \cdot \exp\{-\rho(1 + \log c)\}. \tag{2.13}$$

Also,

$$\left| \frac{1}{\rho} \log \mathbb{E}[G^*(X)^\rho] - \frac{1}{\rho} \log \mathbb{E}[\exp\{\rho L^*(X)\}] \right| \leq 1 + \log c. \tag{2.14}$$

Proof: Observe that

$$\begin{aligned}
&\mathbb{E}[\exp\{\rho L(X)\}] \\
&\quad \geq \mathbb{E}[G_L(X)^\rho] && (2.15)\\
&\quad \geq \mathbb{E}[G^*(X)^\rho] \\
&\quad \geq \mathbb{E}[\exp\{\rho L_{G^*}(X)\}] \exp\{-\rho(1 + \log c)\} && (2.16)\\
&\quad \geq \mathbb{E}[\exp\{\rho L^*(X)\}] \exp\{-\rho(1 + \log c)\}, && (2.17)
\end{aligned}$$

where (2.15) follows from (2.8), and (2.16) from the left inequality in (2.6). The result in (2.12) immediately follows. A similar argument shows (2.13). Finally, (2.14) follows from the inequalities leading to (2.17) by setting $L = L^*$. ■

Thus if we have a length function whose performance is close to optimal, then its associated guessing function is close to guessing optimal. The converse is true as well. Moreover, the optimal guessing exponent is within $1 + \log c$ of the optimal coding exponent

for the length function.

2.1.1 Strings of length n

Let us now consider strings of length n. Let $\mathbb{X}^n$ denote the set of messages and consider $n \to \infty$. Let $\mathcal{M}(\mathbb{X}^n)$ denote the set of pmfs on $\mathbb{X}^n$. By a source, we mean a sequence of pmfs $(P_n : n \in \mathbb{N})$, where $P_n \in \mathcal{M}(\mathbb{X}^n)$. We replace the normalization constant c in (7) by c_n and observe that

$$c_n \leq 1 + n \ln |\mathbb{X}|.$$

If we normalize both sides of equation (2.14) by n, the difference between two quantities as a function of n decays as $O((\log n)/n)$, and vanishes as n tends to infinity. The following theorem follows immediately.

Theorem 2.1.7 *Given $\rho > 0$, the limit*

$$\lim_{n\to\infty} n^{-1} \log \mathbb{E}[G_n^*(X^n)^\rho]$$

exists if and only if the limit

$$\lim_{n\to\infty} \inf_{L_n} n^{-1} \log \mathbb{E}[\exp\{\rho L_n(X^n)\}]$$

exists. Furthermore, the two the limits are equal. □

It is therefore sufficient to restrict our attention to the Campbell's coding problem [6] and study if the limit

$$\lim_{n\to\infty} \inf_{L_n} \frac{1}{n} \log \mathbb{E}[\exp\{\rho L_n(X^n)\}] \tag{2.18}$$

exists, where the infimum is taken over all length functions $L_n : \mathbb{X}^n \to \mathbb{N}$.

2.1.2 Universality

Before we proceed to studying the limit, we make a further remark on the connection between *universal* strategies for guessing and universal strategies for compression.

Let $\mathbb{T}$ denote a class of sources. For each source in the class, let P_n be its restriction to strings of length n and let L_n^* denote an optimal length function that attains the minimum value $\mathbb{E}\left[\exp\{\rho L_n^*(X^n)\}\right]$ among all length functions, the expectation being with respect to P_n. On the other hand, let L_n be a sequence of length functions for the class of sources that does not depend on the actual source within the class. Suppose further that the length sequence L_n is asymptotically optimal, i.e.,

$$\lim_{n\to\infty} \frac{1}{n\rho} \log \mathbb{E}\left[\exp\{\rho L_n(X^n)\}\right] = \lim_{n\to\infty} \frac{1}{n\rho} \log \mathbb{E}\left[\exp\{\rho L_n^*(X^n)\}\right],$$

for every source belonging to the class. L_n is thus "universal" for (i.e., asymptotically optimal for all sources in) the class. An application of (2.12) with c_n in place of c followed by the observation $(1 + \log c_n)/n \to 0$ shows that the sequence of guessing strategies G_{L_n} is asymptotically optimal for the class, i.e.,

$$\lim_{n\to\infty} \frac{1}{n\rho} \log \mathbb{E}\left[G_{L_n}(X^n)^\rho\right] = \lim_{n\to\infty} \frac{1}{n\rho} \log \mathbb{E}\left[G^*(X^n)^\rho\right].$$

Arikan and Merhav [5] provide a universal guessing strategy for the class of discrete memoryless sources (DMS). For the class of unifilar sources with a known number of states, the minimum description length encoding is asymptotically optimal for Campbell's coding length problem (see Merhav [18]). It follows as a consequence of the above argument that guessing in the increasing order of description lengths is asymptotically optimal. The left side of (2.12) is the extra factor in the expected number of guesses (relative to the optimal value) due to lack of knowledge of the specific source in class. Sundaresan [19] characterized this loss as a function of the uncertainty class.

2.2 Large Deviation Results

We begin with some words on notation. Recall that $\mathcal{M}(\mathbb{X}^n)$ denotes the set of pmfs on $\mathbb{X}^n$. The Shannon entropy for a $P_n \in \mathcal{M}(\mathbb{X}^n)$ is

$$H(P_n) = -\sum_{x^n \in \mathbb{X}^n} P_n(x^n) \log P_n(x^n) \tag{2.19}$$

and the Rényi entropy of order $\alpha \neq 1$ is (1.2). The Kullback-Leibler divergence or relative entropy between two pmfs Q_n and P_n is

$$D(Q_n \parallel P_n) = \begin{cases} \displaystyle\sum_{x^n \in \mathbb{X}^n} Q_n(x^n) \log \frac{Q_n(x^n)}{P_n(x^n)}, & \text{if } Q_n \ll P_n, \\ \infty, & \text{otherwise,} \end{cases} \tag{2.20}$$

where $Q_n \ll P_n$ means Q_n is absolutely continuous with respect to P_n. Recall that a source is a sequence of pmfs $(P_n : n \in \mathbb{N})$ where $P_n \in \mathcal{M}(\mathbb{X}^n)$. It is usually obtained via n-length marginals of some probability measure in $\mathcal{M}(\mathbb{X}^{\mathbb{N}})$. Also recall the definitions of limiting guessing exponent in (1.1) and Rényi entropy rate in (1.3) when the limits exist. G_n^* is an optimal guessing function for a pmf $P_n \in \mathcal{M}(\mathbb{X}^n)$. From the results in Section 2.1 on the equivalence between guessing and compression, it is sufficient to focus on the Campbell coding problem (see (2.18)).

Our first contribution is a proof of the following implicit result of Malone & Sullivan [7]. The proof is given in Section 2.4.1.

Proposition 2.2.1 *Let $\rho > 0$. For a source $(P_n : n \in \mathbb{N})$, $E(\rho)$ exists if and only if the Rényi entropy rate (1.3) exists. Furthermore, $E(\rho)/\rho$ equals the Rényi entropy rate.*

The question now boils down to the existence of the limit in the definition of Rényi entropy rate. The theory of large deviations immediately yields a sufficient condition. We begin with a definition.

Definition 2.2.2 (Large deviation property) *[20, Def. II.3.1] A sequence $(\nu_n : n \in$*

$\mathbb{N}$) *of probability measures on* $\mathbb{R}$ *satisfies the* large deviation property (LDP) *with rate function* $I : \mathbb{R} \to [0, \infty]$ *if the following conditions hold:*

- I *is lower semicontinuous on* $\mathbb{R}$*;*
- I *has compact level sets;*
- $\limsup_{n\to\infty} n^{-1} \log \nu_n\{K\} \leq -\inf_{t\in K} I(t)$ *for each closed subset* K *of* $\mathbb{R}$*;*
- $\liminf_{n\to\infty} n^{-1} \log \nu_n\{G\} \geq -\inf_{t\in G} I(t)$ *for each open set* G *of* $\mathbb{R}$*.*

Several commonly encountered sources satisfy the LDP with known and well-studied rate functions. We describe some of these in the examples treated subsequently.

Let ν_n denote the distribution of the information spectrum given by the real-valued random variable $-n^{-1} \log P_n(X^n)$. The following proposition gives a sufficient condition for the existence of the limiting Rényi entropy rate (and therefore the limiting guessing exponent).

Proposition 2.2.3 *Let the sequence of distributions* $(\nu_n : n \in \mathbb{N})$ *of the information spectrum satisfy the LDP with rate function* I*. Then the limiting Rényi entropy rate of order* $1/(1 + \rho)$ *exists for all* $\rho > 0$ *and equals*

$$\beta^{-1} \sup_{t\in\mathbb{R}} \{\beta t - I(t)\},$$

where $\beta = \rho/(1 + \rho)$*. Consequently, the limiting guessing exponent exists and equals*

$$(1 + \rho) \sup_{t\in\mathbb{R}} \{\beta t - I(t)\}.$$

The function $I^*(\beta) := \sup_{t\in\mathbb{R}} \{\beta t - I(t)\}$ is the Legendre-Fenchel dual of the rate function I.

2.2.1 Additional results from Large Deviations Theory

In order to study the examples in Section 2.3, we state some additional results on LDP of transformed variables. (See [21, Sec. 4.2]), [22, Th. 6.12 and 6.14]).

Proposition 2.2.4 (Contraction Principle) *Let $(\xi_n : n \in \mathbb{N})$ denote a sequence of $\mathcal{X}$-valued random variables where $\mathcal{X}$ is a complete separable metric space (Polish space). Let ν_n denote the distribution of ξ_n for $n \in \mathbb{N}$, and the sequence of distributions $(\nu_n : n \in \mathbb{N})$ on $\mathcal{X}$ satisfy the LDP with rate function $I : \mathcal{X} \to [0, \infty]$. Let $\phi : \mathcal{X} \to \mathbb{R}$ be a continuous function. The sequence of distributions of $(\phi(\xi_n) : n \in \mathbb{N})$ on $\mathbb{R}$ also satisfies the LDP with rate function $J : \mathbb{R} \to [0, \infty]$ given by*

$$J(y) = \inf\{I(x) : x \in \mathbb{R}, \phi(x) = y\}.$$

Proposition 2.2.5 (Exponential Approximation) *Let the sequence of distributions of $(\xi_n : n \in \mathbb{N})$ satisfy the LDP with rate function I on $\mathbb{R}$. Assume also that the sequence of random variables $(\zeta_n : n \in \mathbb{N})$ is superexponentially close to $(\xi_n : n \in \mathbb{N})$ in the following sense: for each $\delta > 0$*

$$\limsup_{n\to\infty} \frac{1}{n} \log \Pr\{|\xi_n - \zeta_n| > \delta\} = -\infty. \tag{2.21}$$

Then the sequence of distributions of $(\zeta_n : n \in \mathbb{N})$ also satisfies the LDP on $\mathbb{R}$ with the same rate function I. The condition in (2.21) is satisfied if

$$\lim_{n\to\infty} \sup_{x^n \in \mathbb{X}^n} |\xi_n(x^n) - \zeta_n(x^n)| = 0. \tag{2.22}$$

2.3 Examples

We are now ready to apply Proposition 2.2.3 to various examples. In all the examples that follow, our goal is to show that the sufficient condition for the existence of the limiting guessing exponent holds, i.e., that the sequence of distributions of the information spectrum satisfies the LDP.

Example 2.3.1 (An iid source) This example was first studied by Arikan [4]. Recall that an iid source is one for which $P_n(x^n) = \prod_{i=1}^n P_1(x_i)$, where P_1 is the marginal of X_1. It is then clear that the information spectrum can be written as a sample mean of iid random variables

$$-n^{-1} \log P_n(X^n) = -n^{-1} \sum_{i=1}^n \log P_1(X_i).$$

It is well-known that the sequence $(\nu_n : n \in \mathbb{N})$ of distributions of this sample mean satisfies the LDP with rate function given by the Legendre-Fenchel dual of the cumulant of the random variable $-\log P_1(X_1)$ (see for example [20, Th. II.4.1] or [11, eqn. (1.9.66-67)]):

$$\begin{aligned} \log \mathbb{E}\left[\exp\left\{ \beta(-\log P_1(X_1)) \right\} \right] &= \log \left(\sum_{x \in \mathbb{X}} P_1(x)^\alpha \right) \\ &= (1-\alpha) H_\alpha(P_1). \end{aligned}$$

The Legendre-Fenchel dual of the rate function is therefore the cumulant itself ([20, Th. VI.4.1.e]). An application of Proposition 2.2.3 yields that $(1+\rho)$ times this cumulant, given by $\rho H_\alpha(P_1)$, is the guessing exponent. We thus recover Arikan's result [4].

The rate function I can also be obtained using the *contraction principle* (Proposition 2.2.4) as follows. This method will provide a recipe to obtain the limiting guessing exponent in subsequent examples. Consider a mapping that takes x^n to its empirical pmf in $\mathcal{M}(\mathbb{X})$. Empirical pmf is then a random variable. The distribution of X^n induces a pmf on $\mathcal{M}(\mathbb{X})$. It is well-known that the sequence of distributions of these empirical pmfs, indexed by n, satisfies the *level-2* LDP[3] with rate function $I^{(2)}_{P_1}(\cdot) = D(\cdot \parallel P_1)$. See for example [20, Th II.4.3]. Observe that the mapping from the empirical pmf to the information spectrum random variable is continuous. We can therefore use the contraction principle to get a formula for I in terms of $I^{(2)}_{P_1}(\cdot)$ as follows [20, Th II.5.1]. For any t in

[3]Level-1 refers to sequence of distributions (indexed by n) of sample means, level-2 refers to sample histograms, and level-3 to sample paths.

$\mathbb{R}$, let

$$\theta(t) := \Big\{ Q \in \mathcal{M}(\mathbb{X}) : \sum_{x \in \mathbb{X}} Q(x) \log \frac{1}{P_1(x)} = t \Big\},$$

i.e.,

$$\theta(t) = \Big\{ Q \in \mathcal{M}(\mathbb{X}) : H(Q) + D(Q \parallel P_1) = t \Big\}.$$

Then

$$I(t) = \inf\{ I_{P_1}^{(2)}(Q) : Q \in \theta(t) \}.$$

Using this, we can write

$$\begin{aligned} I^*(\beta) &= \sup_{t \in \mathbb{R}} \Big\{ \beta t - \inf_{Q \in \theta(t)} D(Q \parallel P_1) \Big\} \\ &= \sup_{t \in \mathbb{R}} \sup_{Q \in \theta(t)} \Big\{ \beta t - D(Q \parallel P_1) \Big\} \\ &= \sup_{Q \in \mathcal{M}(\mathbb{X})} \Big\{ \beta (H(Q) + D(Q \parallel P_1)) - D(Q \parallel P_1) \Big\} \\ &= (1+\rho)^{-1} \sup_{Q \in \mathcal{M}(\mathbb{X})} \Big\{ \rho H(Q) - D(Q \parallel P_1) \Big\}, \end{aligned}$$

thus yielding

$$E(\rho) = \sup_{Q \in \mathcal{M}(\mathbb{X})} \Big\{ \rho H(Q) - D(Q \parallel P_1) \Big\}. \tag{2.23}$$

This formula extends to more general sources, as is seen in the next few examples.

Example 2.3.2 (Markov source) This example was studied by Malone & Sullivan [7]. Consider an irreducible Markov chain taking values on $\mathbb{X}$ with transition probability matrix π. Our goal is to verify that the sufficient condition holds and to calculate $E(\rho)$ defined by (1.1) for this source.

Let $\mathcal{M}_s(\mathbb{X}^2)$ denote the set of *stationary* pmfs defined by

$$\mathcal{M}_s\left(\mathbb{X}^2\right) = \Big\{ Q \in \mathcal{M}\left(\mathbb{X}^2\right) : \sum_{x_1 \in \mathbb{X}} Q(x_1, x) = \sum_{x_2 \in \mathbb{X}} Q(x, x_2) \forall x \in \mathbb{X} \Big\}.$$

Denote the common marginal by q and let

$$\eta(\cdot \mid x_1) := \begin{cases} Q(x_1, \cdot)/q(x_1), & \text{if } q(x_1) \neq 0, \\ 1/|\mathbb{X}|, & \text{otherwise.} \end{cases}$$

We may then denote $Q = q \times \eta$, where q is the distribution of X_1 and η the conditional distribution of X_2 given X_1. It is once again well known that the empirical pmf random variable satisfies the level-2 LDP with rate function $I_\pi^{(2)}(Q)$, given by [23]

$$\begin{aligned} I_\pi^{(2)}(Q) &= D(\eta \parallel \pi \mid q) \\ &:= \sum_{x_1 \in \mathbb{X}} q(x_1) D(\eta(\cdot \mid x_1) \parallel \pi(\cdot \mid x_1)). \end{aligned}$$

As in Example 2.3.1, the contraction principle then yields that the sequence of distributions of information spectrum satisfies the LDP with rate function I given by

$$I(t) = \inf\{I_\pi^{(2)}(Q) : Q \in \theta(t)\}.$$

where for t in $\mathbb{R}$, $\theta(t) \subset \mathcal{M}_s(\mathbb{X}^2)$ is defined by

$$\theta(t) = \left\{ Q \in \mathcal{M}_s(\mathbb{X}^2) : \sum_{x_1, x_2} Q(x_1, x_2) \log \frac{1}{\pi(x_2|x_1)} = t \right\}.$$

By Proposition 2.2.1, the limiting guessing exponent exists. Perron-Frobenius theory (Seneta [24, Ch. 1], see also [25, pp.60-61]) yields the cumulant directly as $\log \lambda(\beta)$, where $\lambda(\beta)$ is unique largest eigenvalue (Perron-Frobenius eigenvalue) of a matrix formed by raising each element of π to the power α. (Recall that $\alpha = 1/(1+\rho)$ and $\beta = \rho/(1+\rho)$). Thus $E(\rho) = (1 + \rho) \log \lambda(\beta)$, and we recover the result of Malone & Sullivan [7]. It is useful to note that the steps that led to (2.23) hold in the Markov case (with appropriate changes to entropy and divergence terms) and we may write

$$E(\rho) = \sup_{Q \in \mathcal{M}_s(\mathbb{X}^2)} \Big\{ \rho H(\eta \mid q) - D(\eta \parallel \pi \mid q) \Big\}, \tag{2.24}$$

where $H(\eta \mid q)$ is the conditional entropy of X_2 given X_1 under the joint distribution Q, i.e.,

$$H(\eta \mid q) := -\sum_{x \in \mathbb{X}} q(x) H(\eta(\cdot \mid x)).$$

Example 2.3.3 (Unifilar source) This example was studied by Sundaresan in [9]. A unifilar source is a generalization of the Markov source in Example 2.3.2. Let $\mathbb{X}$ denote the alphabet set as before. In addition, let $\mathbb{S}$ denote a set of finite states. Fix an initial state s_0 and let the joint probability of observing (x^n, s^n) be

$$P_n(x^n, s^n) = \prod_{i=1}^{n} \pi(x_i, s_i \mid s_{i-1})$$

where $\pi(x_i, s_i \mid s_{i-1})$ is the joint probability of (x_i, s_i) given the previous state s_{i-1}. The dependence of P_n on s_0 is understood. Furthermore, assume that $\pi(x_i, s_i \mid s_{i-1})$ is such that $s_i = \phi(s_{i-1}, x_i)$, where ϕ is a deterministic function that is one-to-one for each fixed s_{i-1}. Such a source is called a unifilar source.

$P_{S,X}(s_{i-1}, x_i)$ and ϕ completely specify the process: the initial state S_0 is random with distribution that of marginal of S in $P_{S,X}$, the rest being specified by $P_{X|S}(x_i \mid s_{i-1})$ and ϕ. Example 2.3.2 is a unifilar source with $\mathbb{S} = \mathbb{X}$, $\phi(s_{i-1}, x_i) = x_i$, and $P_{S,X} = q \times \pi$ where q is the stationary distribution of the Markov chain.

Let $\mathcal{M}_s(\mathbb{S} \times \mathbb{X})$ denote the set of joint measures on the indicated space so that the resulting process $(S_n : n \geq 0)$ is a stationary and irreducible Markov chain. Let a $Q \in \mathcal{M}_s(\mathbb{S} \times \mathbb{X})$ be written as $Q = q \times \eta$. For any t in $\mathbb{R}$, let

$$\theta(t) := \left\{ Q \in \mathcal{M}_s(\mathbb{S} \times \mathbb{X}) : \sum_{(s,x)} Q(s,x) \log \frac{1}{\pi(x \mid s)} = t \right\}.$$

Then the sequence of distributions of information spectrum $-n^{-1} \log P_n(X^n)$ satisfies the LDP ([11, eqn. (1.9.30)]) with rate function given (once again via contraction principle) by

$$I(t) = \inf\{D(\eta \parallel \pi \mid q) : Q \in \theta(t)\}.$$

The limiting exponent therefore exists. Following the same procedure that led to (2.23) in the iid case and (2.24) for a Markov source, we get

$$E(\rho) = \sup_{Q \in \mathcal{M}_s(\mathbb{S}\times\mathbb{X})} \Big\{ \rho H(\eta \mid q) - D(\eta \parallel \pi \mid q) \Big\}, \tag{2.25}$$

where $H(\eta \mid q)$ and $D(\eta \parallel \pi \mid q)$ are analogously defined, and the result of Sundaresan [9] is recovered.

Example 2.3.4 (A class of stationary sources) Pfister & Sullivan [8] consider a class of stationary sources with distribution $P \in \mathcal{M}\left(\mathbb{X}^{\mathbb{N}}\right)$ that satisfy two hypotheses H1 and H2 of [8, Sec. II-B]. Hypothesis H1 assumes that for any $\varepsilon > 0$ and any neighborhood of the given source, there is a stationary and ergodic approximation that is absolutely continuous with the given source such that the excess of the given source's Shannon entropy rate over that of the approximation is at most ε. Hypothesis H2 is given by (2.27) below. Under these hypotheses, they prove that $E(\rho)$ exists, and provide a variational characterization analogous to (2.25), i.e.,

$$E(\rho) = \sup_{Q \in \mathcal{M}_s^P} \Big\{ \rho \overline{H}(Q) - \overline{D}(Q \parallel P) \Big\}, \tag{2.26}$$

where $\overline{H}(Q)$ is the Shannon entropy rate, and with P_n and Q_n restrictions of P and Q to n letters

$$\overline{D}(Q \parallel P) = \lim_{n\to\infty} n^{-1} \sum_{x^n} Q_n(x^n) \log \frac{Q_n(x^n)}{P_n(x^n)}.$$

$\mathcal{M}_s^P$ is the set of stationary sources that satisfy $Q_n \ll P_n$ for all n.

En route to this result, Pfister & Sullivan [8] show that the sequence of distributions of the *empirical process* satisfies the *level-3* LDP for sample paths. We first state this precisely, and then use this as the starting point to show the sufficient condition that the information spectrum satisfies the LDP.

For an $x \in \mathbb{X}^{\mathbb{N}}$ given by $x = (x_1, x_2, \cdots)$, we define $x^n = (x_1, \cdots, x_n)$ as the first n components of x in the usual way. In the other direction, given an $x^n \in \mathbb{X}^n$, let $[x^n] \in \mathbb{X}^{\mathbb{N}}$ denote the periodic point in $\mathbb{X}^{\mathbb{N}}$ obtained by repeating $(x_1, \cdots, x_n)$. Let $\tau : \mathbb{X}^{\mathbb{N}} \to \mathbb{X}^{\mathbb{N}}$

denote the shift operator defined by

$$(\tau(x))_i = x_{i+1}, \forall i \in \mathbb{N}.$$

Consider a stationary source P whose letters are $X_1, X_2, \cdots$. Define the empirical process of measures

$$T_n(X^n, \cdot) = n^{-1} \sum_{i=0}^{n-1} \delta_{\tau^i([X^n])}(\cdot).$$

This is a measure on $\mathbb{X}^{\mathbb{N}}$ that puts mass $1/n$ on the following strings:

$$[x^n], \tau([x^n]), \tau^2([x^n]), \cdots, \tau^{n-1}([x^n]).$$

Let $\mathcal{M}_s$ denote set of stationary distributions on $\mathbb{X}^{\mathbb{N}}$. Pfister & Sullivan show that the distributions of the $\mathcal{M}_s$-valued process $T_n(X^n, \cdot)$ satisfies the level-3 LDP with rate function $I_P^{(3)}(\cdot) = \overline{D}(\cdot \parallel P)$ under hypotheses H1 and H2 of their paper. We next use this to show that the sequence of distributions of the information spectrum satisfies the LDP.

Hypothesis H2 of Pfister & Sullivan assumes the existence of a continuous mapping $e_P : \mathbb{X}^{\mathbb{N}} \to \mathbb{R}$ satisfying

$$\lim_{n\to\infty} \sup_{x^n \in \Sigma_n^P} \left| n^{-1} \log P_n(x^n) + \int_{\mathbb{X}^{\mathbb{N}}} e_P \, dT_n([x^n], \cdot) \right| = 0, \tag{2.27}$$

where Σ_n^P denotes the support set of P_n.

By the compactness of $\mathbb{X}^{\mathbb{N}}$, e_P is uniformly continuous. Under the weak topology on the separable metric space $\mathcal{M}_s^P(\mathbb{X}^{\mathbb{N}})$, the mapping

$$\phi : \mathcal{M}_s^P(\mathbb{X}^{\mathbb{N}}) \to \mathbb{R}$$

defined by $Q \mapsto \int_{\mathbb{X}^{\mathbb{N}}} e_P \, dQ$ is a continuous mapping. Hence by the contraction principle, by setting $\mathcal{X} = \mathcal{M}_s$ we get that the sequence of distributions of $(\phi(T_n(X^n, \cdot) : n \in \mathbb{N})$

satisfies the LDP with rate function I given by

$$I(t) = \inf\left\{\overline{D}(Q \parallel P) : Q \in \mathcal{M}_s^P, \phi(Q) = t\right\}.$$

Furthermore, given hypothesis H2 and (2.27), an application of the exponential approximation principle (Proposition 2.2.5) indicates that the sequence of distributions of the information spectrum too satisfies the LDP with the same rate function I, and we have verified that the sufficient condition holds.

What remains is to calculate this rate function. For this, we return to Pfister & Sullivan's work and use $\overline{D}(Q \parallel P) = \phi(Q) - \overline{H}(Q)$ [20, Prop. 2.1] to write

$$I(t) = \inf_{Q \in \mathcal{M}_s^P} \left\{\overline{D}(Q \parallel P) : \overline{H}(Q) + \overline{D}(Q \parallel P) = t\right\}.$$

Finally, the Legendre-Fenchel dual of the rate function is computed as in the steps leading to (2.23)-(2.25), yielding (2.26).

Example 2.3.5 (Mixed source) Consider a mixture of two iid sources with letters from $\mathbb{X}$. We may write

$$P_n(x^n) = \lambda \prod_{i=1}^{n} R(x_i) + (1-\lambda) \prod_{i=1}^{n} S(x_i)$$

where $\lambda \in (0,1)$ with $R, S \in \mathcal{M}(\mathbb{X})$ the two marginal pmfs that define the iid components of the mixture. It is easy to see that the guessing exponent is the maximum of the guessing exponents for the two component sources. We next verify this using Proposition 2.2.3.

The sequence of distributions of the information spectrum satisfies the LDP with rate function given as follows (see Han [11, eqn. (1.9.41)]). Define

$$\begin{aligned}
\theta_1 &= \Big\{Q \in \mathcal{M}(\mathbb{X}) : D(Q \parallel S) - D(Q \parallel R) \geq 0\Big\}, \\
\theta_2 &= \Big\{Q \in \mathcal{M}(\mathbb{X}) : D(Q \parallel S) - D(Q \parallel R) \leq 0\Big\},
\end{aligned}$$

and for $t \in \mathbb{R}$

$$\begin{aligned} A_t &= \theta_1 \cap \Big\{ Q \in \mathcal{M}(\mathbb{X}) : H(Q) + D(Q \parallel R) = t \Big\} \\ B_t &= \theta_2 \cap \Big\{ Q \in \mathcal{M}(\mathbb{X}) : H(Q) + D(Q \parallel S) = t \Big\}. \end{aligned}$$

The rate function (via the contraction principle) is given by

$$I(t) = \min \left\{ \inf_{Q \in A_t} D(Q \parallel R), \inf_{Q \in B_t} D(Q \parallel S) \right\}.$$

From Proposition 2.2.3 we conclude that the limiting guessing exponent exists. $I^*(\beta)$ is then

$$\begin{aligned} &\sup_{t \in \mathbb{R}} \left\{ \beta t - \min \Big\{ \inf_{Q \in A_t} D(Q \parallel R), \inf_{Q \in B_t} D(Q \parallel S) \Big\} \right\} \\ &= \max \Big\{ \sup_{t \in \mathbb{R}} \sup_{Q \in A_t} \Big\{ \beta t - D(Q \parallel R) \Big\}, \\ &\qquad \sup_{t \in \mathbb{R}} \sup_{Q \in B_t} \Big\{ \beta t - D(Q \parallel S) \Big\} \Big\} \\ &= \max \Big\{ \sup_{Q \in \theta_1} \Big\{ \beta H(Q) - (1-\beta) D(Q \parallel R) \Big\}, \\ &\qquad \sup_{Q \in \theta_2} \Big\{ \beta H(Q) - (1-\beta) D(Q \parallel S) \Big\} \Big\} \\ &= (1+\rho)^{-1} \max \Big\{ \sup_{Q} \Big\{ \rho H(Q) - D(Q \parallel R) \Big\}, \\ &\qquad \sup_{Q} \Big\{ \rho H(Q) - D(Q \parallel S) \Big\} \Big\} \\ &= (1+\rho)^{-1} \max \Big\{ \rho H_\alpha(R), \rho H_\alpha(S) \Big\}, \end{aligned}$$

yielding

$$E(\rho) = \max \Big\{ \rho H_\alpha(R), \rho H_\alpha(S) \Big\}.$$

2.4 Proofs

We now prove Propositions 2.2.1 and 2.2.3.

2.4.1 Proof of Proposition 2.2.1

From Theorem 2.1.7 it is sufficient to show that the limit in (2.18) for Campbell's coding problem exists if and only if the Rényi entropy rate exists, with the former ρ times the latter.

Fix n. In the rest of the proof, we use the notation $\mathbb{E}_{P_n}[\cdot]$ for expectation with respect to distribution P_n. The length function can be thought of as a bounded (continuous) function from $\mathbb{X}^n$ to $\mathbb{R}$ and therefore our interest is in the logarithm of its moment generating function of ρ, the cumulant. The cumulant associated with a bounded continuous function (here L_n) has a variational characterization [26, Prop. 1.4.2] as the following Legendre-Fenchel dual of the Kullback-Leibler divergence, i.e.,

$$\log \mathbb{E}_{P_n}\Big[\exp\{\rho L_n(X^n)\}\Big] = \sup_{Q_n \in \mathcal{M}(\mathbb{X}^n)} \Big\{\rho \mathbb{E}_{Q_n}[L_n(X^n)] - D(Q_n \parallel P_n)\Big\}. \tag{2.28}$$

Taking infimum on both sides over all length functions, we arrive at the following chain of inequalities:

$$\begin{aligned}
&\inf_{L_n} \log \mathbb{E}_{P_n}\Big[\exp\{\rho L_n(X^n)\}\Big] && (2.29)\\
&= \inf_{L_n} \sup_{Q_n \in \mathcal{M}(\mathbb{X}^n)} \Big\{\mathbb{E}_{Q_n}[\rho L_n(X^n)] - D(Q_n \parallel P_n)\Big\} \\
&= \sup_{Q_n \in \mathcal{M}(\mathbb{X}^n)} \inf_{L_n} \Big\{\mathbb{E}_{Q_n}[\rho L_n(X^n)] - D(Q_n \parallel P_n)\Big\} + \Theta(1) && (2.30)\\
&= \sup_{Q_n \in \mathcal{M}(\mathbb{X}^n)} \Big\{\rho H_n(Q_n) - D(Q_n \parallel P_n)\Big\} + \Theta(1) && (2.31)\\
&= \rho H_{\frac{1}{1+\rho}}(P_n) + \Theta(1). && (2.32)
\end{aligned}$$

Equation (2.30) follows because (i) the mapping

$$(L_n, Q_n) \mapsto \mathbb{E}_{Q_n}[\rho L_n(X^n)] - D(Q_n \parallel P_n)$$

is a concave function of Q_n, (ii) for fixed Q_n and for any two length functions L_n^1 and L_n^2, for any $\lambda \in [0,1]$, the function $L_n = \lceil \lambda L_n^1 + (1-\lambda) L_n^2 \rceil$ is also a length function and

$$\mathbb{E}_{Q_n}[L_n] = \lambda \mathbb{E}_{Q_n}[L_n^1] + (1-\lambda)\mathbb{E}_{Q_n}[L_n^2] + \Theta(1).$$

(iii) $\mathcal{M}(\mathbb{X}^n)$ is compact and convex, and therefore the infimum and supremum may be interchanged upon an application of a version of Ky Fan's minimax result [27]. This yields a compression problem, the infimum over L_n of expected lengths with respect to a distribution Q_n. The answer is the well-known Shannon entropy $H(Q_n)$ to within 1 bit, and (2.31) follows. Lastly, (2.32) is a well-known identity which may also be obtained directly by writing the supremum term in (2.31) as

$$(1+\rho) \sup_{Q_n \in \mathcal{M}(\mathbb{X}^n)} \left\{ \mathbb{E}_{Q_n}\left[-\left(\frac{\rho}{1+\rho}\right) \log P_n(X^n) \right] - D(Q_n \parallel P_n) \right\}$$

and then applying (2.28) with $-(\rho/(1+\rho)\log P_n(X^n))$ in place of $\rho L_n(X^n)$ to get the scaled Rényi entropy.

Normalize both (2.29) and (2.32) by n and let $n \to \infty$ to deduce that (2.18) exists if and only if the limiting normalized Rényi entropy rate exists. This concludes the proof.

2.4.2 Proof of Proposition 2.2.3

This is a straightforward application of Varadhan's theorem [28] on asymptotics of integrals. Recall that ν_n is the distribution of the information spectrum $n^{-1}\log P_n(X^n)$. Define $F(t) = \beta t$. Since the $(\nu_n : n \in \mathbb{N})$ sequence satisfies the LDP with rate function I, Varadhan's theorem (see Ellis [20, Th. II.7.1.b]) states that if

$$\lim_{M\to\infty} \limsup_{n\to\infty} \frac{1}{n} \log \int_{t \geq \frac{M}{\beta}} \exp\{n\beta t\} \, d\nu_n(t) = -\infty \tag{2.33}$$

then the limit

$$\lim_{n\to\infty} \frac{1}{n} \log \int_{\mathbb{R}} \exp\{n\beta t\} \, \nu_n(dt) = \sup_{t\in\mathbb{R}} \{\beta t - I(t)\} \tag{2.34}$$

holds. The integral on the left side in (2.34) can be simplified by defining the finite cardinality set

$$A_n = \{-n^{-1}\log P_n(x^n) : \forall x^n \in \mathbb{X}^n\} \subset \mathbb{R}$$

and by observing that

$$\begin{aligned}\int_{\mathbb{R}} \exp\{n\beta t\}\ \nu_n(dt) &= \sum_{t\in A_n} \exp\{n\beta t\} \sum_{x^n: P_n(x^n)=\exp\{-nt\}} P_n(x^n) \\ &= \sum_{x^n} P_n(x^n)^{1-\beta} \\ &= \sum_{x^n} P_n(x^n)^{\frac{1}{1+\rho}} = \exp\left\{\beta H_{1/(1+\rho)}(P_n)\right\}.\end{aligned}$$

Take logarithms, normalize by n, take limits, and apply (2.34) to get the desired result. It therefore remains to prove (2.33).

The event $\{t \geq \frac{M}{\beta}\}$ occurs if and only if $\left\{P_n(x^n) \leq \exp\{\frac{-nM}{\beta}\}\right\}$. The integral in (2.33) can therefore be written as

$$\begin{aligned}\sum_{t\in A_n, t\geq \frac{M}{\beta}} \sum_{x^n: P_n(x^n)=\exp\{-nt\}} \exp\{n\beta t\} P_n(x^n) &= \sum_{x^n: P_n(x^n)\leq \exp\{\frac{-nM}{\beta}\}} P_n(x^n)^{\frac{1}{1+\rho}} \\ &\leq |\mathbb{X}|^n \cdot \exp\left\{\frac{-nM}{\beta(1+\rho)}\right\}.\end{aligned}$$

The sequence in n on the left side of (2.33) is then

$$\log|\mathbb{X}| - \frac{M}{\beta(1+\rho)},$$

a constant sequence. Take the limit as $M \to \infty$ to verify (2.33). This concludes the proof.

2.5 Summary

In this chapter we first showed that the problem of finding the limiting guessing exponent is equal to that of finding the limiting compression exponent under exponential costs (Campbell's coding problem). We then saw that the latter limit exists if the sequence of distributions of the information spectrum satisfies the LDP (sufficient condition). The limiting exponent was the Legendre-Fenchel dual of the rate function, scaled by an appropriate constant. It turned out to be the limit of the normalized cumulant of the

information spectrum random variable. While some of these facts can be gleaned from the works of Pfister & Sullivan [8] and Merhav & Arikan [5], our work sheds light on the key role played by the information spectrum and the large deviation property of their distributions. We checked that the sufficient condition held in all previously studied examples (perfect secrecy, no side information).

Existence of limiting cumulants and their differentiability (with respect to ρ) imply the LDP (Gärtner-Ellis theorem [20, Th. II.6.1]). Here however we have proceeded in the reverse direction to conclude existence of the limiting cumulants given the LDP. Our approach enabled us to exploit several well-known results on the LDP and the corresponding rate functions for a wide class of random processes.

Chapter 3

Key Constrained Sources

In this chapter we consider secret transmission of a general source over a Shannon cipher system with an arbitrary positive key rate. The key rate may not be sufficient to achieve perfect secrecy. The best achievable guessing exponent is used as a measure of the cryptosystem's strength against a wiretapper's guessing attacks[1]. We prove upper and lower bounds on the guessing exponents. We also relate them to fixed rate source coding error and correct decoding exponents. We then show that the upper and lower bounds are tight for DMS, Markov, and unifilar sources. Before we begin with the problem statement, we recall the set-up of the Shannon cipher system (see figure 1.1).

Let X^n be a message taking values in $\mathbb{X}^n$. This message should be communicated securely from a transmitter to a receiver, both of which have access to a common secure key U^k of k purely random bits independent of X^n. The transmitter computes the cryptogram $Y = f_n(X^n, U^k)$ and sends it to the receiver over a public channel. The cryptogram may be of variable length, and $R = k/n$ is the key rate of the system. The encryption function f_n is invertible for any fixed U^k. The receiver, knowing Y and U^k, computes $X^n = f_n^{-1}(Y, U^k)$. The functions f_n and f_n^{-1} are published.

[1]Results in this chapter can be found in [29].

3.1 Problem Statement

Let $\mathcal{M}(\mathbb{X}^n)$ be the set of pmfs on $\mathbb{X}^n$. By a source, we mean a sequence of pmfs $(P_n : n \in \mathbb{N})$, where[2] $P_n \in \mathcal{M}(\mathbb{X}^n)$ and $\mathbb{N}$ denotes the set of natural numbers.

For a given cryptogram $Y = y$, define a *guessing strategy*

$$G_n(\cdot \mid y) : \ \mathbb{X}^n \to \{1, 2, \cdots |\mathbb{X}|^n\}$$

as a bijection that denotes the order in which elements of $\mathbb{X}^n$ are guessed. $G_n(x^n \mid y) = g$ indicates that x^n is the gth guess, when the cryptogram is y. With knowledge of P_n, the encryption function f_n, and the cryptogram Y, the attacker can completely calculate all the posterior probabilities of plaintexts $P_{X^n|Y}(\cdot \mid y)$ given the cryptogram. The attacker's optimal guessing strategy is then to guess in the decreasing order of these posterior probabilities $P_{X^n|Y}(\cdot \mid y)$. Let us denote this optimal attack strategy as G_{f_n}. Let $(f_n : n \in \mathbb{N})$ denote the sequence of encryption functions known to the attacker. We assume that attacker employs the optimal guessing strategy.

For a given $\rho > 0$, key rate $R > 0$, define the normalized guessing exponent

$$E_n^g(R, \rho) := \sup_{f_n} \frac{1}{n} \log \mathbb{E}\left[G_{f_n}(X^n \mid Y)^\rho\right].$$

The supremum is taken over all encryption functions. Further define performance limits of guessing moments as in [2]:

$$E_u^g(R, \rho) := \limsup_{n\to\infty} E_n^g(R, \rho) \tag{3.1}$$

$$E_l^g(R, \rho) := \liminf_{n\to\infty} E_n^g(R, \rho). \tag{3.2}$$

Our interest in this chapter is to derive bounds on $E_u^g(R, \rho)$ and $E_l^g(R, \rho)$, and evaluate them for some specific examples.

[2]Sometimes we use P_{X^n} in place of P_n when we refer to the distribution of random vector X^n.

3.2 Guessing with key-rate constraints and source compression

In this section we establish a connection between guessing and source compression subject to a new cost criterion. We next define the related compression quantities. Recall that a length function $L_n : \mathbb{X}^n \to \mathbb{N}$ is a mapping that satisfies Kraft's inequality:

$$\sum_{x^n \in \mathbb{X}^n} \exp\{-L_n(x)\} \leq 1.$$

Every length function yields an attack strategy with a performance characterized as follows.

Proposition 3.2.1 *Let L_n be any length function on $\mathbb{X}^n$. There is a guessing list G_n such that for any encryption function f_n, we have*

$$G_n(x^n \mid y) \leq 2 \exp\left\{\min\left\{L_n(x^n), nR\right\}\right\}.$$

Proof: We use a technique of Merhav and Arikan [2]. Let G_{L_n} denote the guessing function that ignores the cryptogram and proceeds in the increasing order of L_n lengths. Suppose G_{L_n} proceeds in the order $x_1^n, x_2^n, \cdots$. By Proposition 2.1.4, we need at most $\exp\{L_n(x^n)\}$ guesses to identify x^n (This is a simple consequence of the fact that there are at most $\exp\{L_n(x^n)\}$ strings of length less than or equal to $L_n(x^n)$).
As an alternative attack, consider the exhaustive key-search attack defined by the following guessing list:

$$f_n^{-1}\left(y, u_1^k\right), f_n^{-1}\left(y, u_2^k\right), \cdots$$

where $u_1^k, u_2^k, \cdots$ is an arbitrary ordering of the keys. This strategy identifies x^n in at most $\exp\{nR\}$ guesses. Finally, let $G_n(\cdot \mid y)$ be the list that alternates between the two lists, skipping those already guessed, i.e., the one that proceeds in the order

$$x_1^n, f_n^{-1}\left(y, u_1^k\right), x_2^n, f_n^{-1}\left(y, u_2^k\right), \cdots. \tag{3.3}$$

Clearly, for every x^n, we need at most twice the minimum of the two original lists. ■

We now look at a weak converse in the expected sense to the above. Recall from Proposition 2.1.2 that for any guessing function G_n, there exists a length function L_{G_n} that satisfies

$$L_{G_n}(x^n) - 1 - \log c_n \leq \log G_n(x^n) \leq \log L_{G_n}(x^n), \tag{3.4}$$

where $c_n = \sum_{i=1}^{|\mathbb{X}|^n} \frac{1}{i}$.

Proposition 3.2.2 *Fix $n \in \mathbb{N}$, $\rho > 0$. There is an encryption function f_n and a length function L_n such that every guessing strategy G_n (and in particular G_{f_n}) satisfies*

$$\mathbb{E}\left[G(X^n \mid Y)^\rho\right] \geq \frac{1}{(ec_n)^\rho(2+\rho)} \mathbb{E}\left[\exp\left\{\rho \min\left\{L_n\left(X^n\right), nR\right\}\right\}\right],$$

where e denotes $\exp\{1\}$.

Proof: See section 3.5.1. The proof is an extension of Merhav & Arikan's proof of [2, Th.1] to sources with memory. The idea is to identify an encryption mechanism that maps messages of roughly equal probability to each other. Our proof also suggests an asymptotically optimal encryption strategy for sources with memory. ■

Remark 3.2.3 *Note that* $\log c_n \leq \log(1 + n \log |\mathbb{X}|)$, *so that* $(\log c_n)/n = O((\log n)/n)$.

Propositions 3.2.1 and 3.2.2 naturally suggest the following coding problem: identify

$$E_n^s(R, \rho) := \min_{L_n} \frac{1}{n} \log \mathbb{E}\left[\exp\left\{\rho \min\left\{L_n(X^n), nR\right\}\right\}\right]. \tag{3.5}$$

The minimum is taken over all length functions. We may interpret the cost of using length $L_n(x^n)$ as $\exp\left\{\min\{L_n(x^n), nR\}\right\}$, i.e., the cost is exponential in L_n, but saturates at $\exp\{nR\}$ and so all lengths larger than nR enjoy a saturated cost. Then $E_n^s(R, \rho)$ is the minimum normalized exponent of the ρth moment of this new compression cost. In analogy with (3.1) and (3.2) we define

$$E_u^s(R, \rho) = \limsup_{n \to \infty} E_n^s(R, \rho)$$

$$E_l^s(R,\rho) = \liminf_{n\to\infty} E_n^s(R,\rho)$$

The following is a corollary to Propositions 3.2.1 and 3.2.2, and relates $E_n^g(R,\rho)$ and $E_n^s(R,\rho)$.

Corollary 3.2.4 *For a given $R, \rho > 0$, we have*

$$|E_n^s(R,\rho) - E_n^g(R,\rho)| \leq \frac{\log((2ec_n)^\rho(2+\rho))}{n}. \tag{3.6}$$

Proof: Let L_n^* be the length function that achieves $E_n^s(R,\rho)$. Using Proposition 3.2.1, and after taking expectation, we have the guessing strategy G_n that satisfies

$$\begin{aligned}
&\mathbb{E}\left[\exp\left\{\rho \min\left\{L_n^*(X^n), nR\right\}\right\}\right] \\
&\quad\geq \sup_{f_n} \frac{1}{2^\rho}\mathbb{E}\left[G_n(X^n \mid Y)^\rho\right] \\
&\quad\geq \sup_{f_n} \frac{1}{2^\rho}\mathbb{E}\left[G_{f_n}(X^n \mid Y)^\rho\right] \\
&\quad\geq \frac{1}{(2ec_n)^\rho(2+\rho)}\mathbb{E}\left[\exp\left\{\rho \min\left\{L_n(X^n), nR\right\}\right\}\right] \\
&\qquad \text{for some } f_n \text{ and } L_n, \text{ given by Proposition 3.2.2,} \\
&\quad\geq \frac{1}{(2ec_n)^\rho(2+\rho)}\mathbb{E}\left[\exp\left\{\rho \min\left\{L_n^*(X^n), nR\right\}\right\}\right].
\end{aligned}$$

Take logarithms and normalize by n to get (3.6). ■

We now state the equivalence between compression and guessing.

Theorem 3.2.5 (Guessing-Compression Equivalence) *For any $\rho > 0$ and $R > 0$, we have $E_u^s(R,\rho) = E_u^g(R,\rho)$ and $E_l^s(R,\rho) = E_l^g(R,\rho)$.* □

Proof: From Corollary 3.2.4, magnitude of the difference between $E_n^g(R,\rho)$ and $E_n^s(R,\rho)$ decays as $O((\log n)/n)$ and vanishes as $n \to \infty$. ■

Thus, the problem of finding the optimal guessing exponent is the same as that of finding the optimal exponent for the coding problem in (3.5). When $R \geq \log|\mathbb{X}|$, the coding problem in (3.5) reduces to the one considered by Campbell in [6]; this is a case where perfect secrecy is obtained and was studied in chapter 2. Proposition 3.2.1 shows that the optimal length function attaining the minimum in (3.5) yields an asymptotically

optimal attack strategy on the cipher system. Moreover, the encryption strategy in the proof of Proposition 3.2.2 (see section 3.5.1) is asymptotically optimal, from the designer's point of view.

In the rest of the chapter we focus on the equivalent compression problem and find bounds on E_u^s and E_l^s.

3.3 Growth Exponent for the Modified Compression Problem

We begin with some words on notation. Recall that $\mathcal{M}(\mathbb{X}^n)$ denotes the set of pmfs on $\mathbb{X}^n$. The Shannon entropy and the Rényi entropy of order $\alpha \neq 1$ for a $P_n \in \mathcal{M}(\mathbb{X}^n)$ are as defined in (2.19) and (1.2), respectively. The Kullback Leibler divergence or relative entropy between two pmfs Q_n and P_n is given by (2.20).

Let $(X^n : n \in \mathbb{N})$ denote a sequence of random variables on $\mathbb{X}^n$, with corresponding sequence of probability measures denoted by $\mathbf{X} := (P_{X^n} : n \in \mathbb{N})$. Thus $\mathbf{X}$ is a source and X^n its n-letter message output. Abusing notation, we let $\mathcal{M}(\mathbb{X}^{\mathbb{N}})$ denote the set of all sequences $\mathbf{Y} = (P_{Y^n} : n \in \mathbb{N})$ of probability measures, and for each $\mathbf{B} := (B_n \subseteq \mathbb{X}^n : n \in \mathbb{N})$, we define

$$\mathcal{M}(\mathbf{B}) := \left\{ \mathbf{Y} \in \mathcal{M}(\mathbb{X}^{\mathbb{N}}) : \lim_{n\to\infty} P_{Y^n}(B_n) = 1 \right\}.$$

In the rest of this section $\mathbf{X}$ is a fixed source. For any $\mathbf{Y} \in \mathcal{M}(\mathbf{B})$ and $\rho > 0$, define

$$E_u(\mathbf{Y}, \mathbf{X}, \rho) := \limsup_{n\to\infty} \frac{1}{n}\{\rho H(P_{Y^n}) - D(P_{Y^n} \parallel P_{X^n})\}$$

and

$$E_l(\mathbf{Y}, \mathbf{X}, \rho) := \liminf_{n\to\infty} \frac{1}{n}\{\rho H(P_{Y^n}) - D(P_{Y^n} \parallel P_{X^n})\}.$$

We next state a large deviation result that plays a key role in the derivation of bounds on E_u^s and E_l^s.

Proposition 3.3.1 *For all $\rho \geq 0$ and $\mathbf{B} = (B_n \subseteq \mathbb{X}^n : n \in \mathbb{N})$, we have*

$$(1+\rho) \limsup_{n\to\infty} \frac{1}{n} \log \sum_{x^n \in B_n} P_{X^n}^{\frac{1}{1+\rho}}(x^n) = \max_{\mathbf{Y} \in \mathcal{M}(\mathbf{B})} E_u(\mathbf{Y}, \mathbf{X}, \rho) \tag{3.7}$$

$$(1+\rho) \liminf_{n\to\infty} \frac{1}{n} \log \sum_{x^n \in B_n} P_{X^n}^{\frac{1}{1+\rho}}(x^n) = \max_{\mathbf{Y} \in \mathcal{M}(\mathbf{B})} E_l(\mathbf{Y}, \mathbf{X}, \rho) \tag{3.8}$$

The maximum-achieving distribution in (3.7) and (3.8) is $\mathbf{X}^* = (P^*_{X^n} : n \in \mathbb{N})$ where

$$P^*_{X^n}(\cdot) = \frac{P_{X^n}^{\frac{1}{1+\rho}}(\cdot)}{\sum_{y \in B_n} P_{X^n}^{\frac{1}{1+\rho}}(y)}. \tag{3.9}$$

Proof: See section 3.5.2. ■

Remark 3.3.2 *This proposition is a generalization of Iriyama's [13, Prop. 1], which is obtained by setting $\rho = 0$.*

3.3.1 Upper Bound on E_u^s

We first obtain an upper bound on E_u^s. We use $\mathbb{E}_{X^n}[\cdot]$ to denote the expectation with respect to distribution P_{X^n}.

Proposition 3.3.3 (Upper Bound) *Let $R > 0$ and $\rho > 0$. Then*

$$E_u^s(R, \rho) \leq \min_{0 \leq \theta \leq \rho} \left[(\rho - \theta)R + \max_{\mathbf{Y} \in \mathcal{M}(\mathbb{X}^{\mathbb{N}})} E_u(\mathbf{Y}, \mathbf{X}, \theta) \right]$$

Proof: We first recall the useful variational formula [26, Prop. 1.4.2]

$$\log \mathbb{E}_{X^n} \left[\exp\{U(X^n)\} \right] = \sup_{P_{Y^n}} \left\{ \mathbb{E}_{Y^n}[U(Y^n)] - D(P_{Y^n} \parallel P_{X^n}) \right\}. \tag{3.10}$$

for any $U : \mathbb{X}^n \to \mathbb{R}$, where $\mathbb{R}$ denotes set of real numbers. Observe that

$$\begin{aligned}
&\log \mathbb{E}_{X^n}\left[\exp\left\{\rho \min\{L_n(X^n), nR\}\right\}\right] \\
&= \sup_{P_{Y^n}}[\rho \mathbb{E}_{Y^n}\left[\min\{L_n(Y^n), nR\}\right] - D(P_{Y^n} \parallel P_{X^n})] && (3.11)\\
&\leq \sup_{P_{Y^n}}[\rho \min\{\mathbb{E}_{Y^n}\left[L_n(Y^n)\right], nR\} - D(P_{Y^n} \parallel P_{X^n})] && (3.12)\\
&= \sup_{P_{Y^n}}\left\{\min_{0\leq\theta\leq\rho}[(\rho-\theta)nR + \theta\mathbb{E}_{Y^n}\left[L_n(Y^n)\right] - D(P_{Y^n} \parallel P_{X^n})\right\} && (3.13)\\
&= \min_{0\leq\theta\leq\rho}\sup_{P_{Y^n}}\left\{(\rho-\theta)nR + \theta\mathbb{E}_{Y^n}\left[L_n(Y^n)\right] - D(P_{Y^n} \parallel P_{X^n})\right\} && (3.14)\\
&= \min_{0\leq\theta\leq\rho}\left\{(\rho-\theta)nR + \sup_{P_{Y^n}}\left\{\theta\mathbb{E}_{Y^n}\left[L_n(Y^n)\right] - D(P_{Y^n} \parallel P_{X^n})\right\}\right\}.
\end{aligned}$$

In the above sequence of inequalities, (3.11) follows from the variational formula (3.10) with $U(x^n) = \rho \min\{L_n(x^n), nR\}$. Inequality (3.12) follows from Jensen's inequality because $\min\{L_n, nR\}$ is concave in L_n for a fixed nR. Equality (3.13) follows from the identity

$$\rho \min\{a, b\} = \min_{0\leq\theta\leq\rho}\{\theta a + (\rho-\theta)b\}.$$

Equality (3.14) follows because the term within braces is linear in θ, concave in P_{Y^n}, and $\mathcal{M}(\mathbb{X}^n)$ is compact; these permit an interchange of sup and inf by an application of a version of Ky-Fan's minmax theorem [27]. Taking infimum over L_n, and interchanging the infimum over L_n and the min over θ, we get

$$\begin{aligned}
&\inf_{L_n}\log \mathbb{E}_{X^n}\left[\exp\left\{\rho \min\{L_n(Y^n), nR\}\right\}\right] \\
&\leq \min_{0\leq\theta\leq\rho}\left\{(\rho-\theta)nR + \inf_{L_n}\sup_{P_{Y^n}}\left\{\theta\mathbb{E}_{Y^n}\left[L_n(Y^n)\right] - D(P_{Y^n} \parallel P_{X^n})\right\}\right\} \\
&= \min_{0\leq\theta\leq\rho}\left\{(\rho-\theta)nR + \sup_{P_{Y^n}}\left\{\theta\inf_{L_n}\mathbb{E}_{Y^n}\left[L_n(Y^n)\right] - D(P_{Y^n} \parallel P_{X^n})\right\} + O(1)\right\} && (3.15)\\
&= \min_{0\leq\theta\leq\rho}\left\{(\rho-\theta)nR + \sup_{P_{Y^n}}\left\{\theta H(P_{Y^n}) - D(P_{Y^n} \parallel P_{X^n})\right\} + O(1)\right\} && (3.16)\\
&= \min_{0\leq\theta\leq\rho}\left\{(\rho-\theta)nR + \theta H_{\frac{1}{1+\theta}}(P_{X^n}) + O(1)\right\}. && (3.17)
\end{aligned}$$

Equality (3.15) follows because the function inside the inner braces is concave in P_{Y^n}, asymptotically linear in L_n (see proof of Proposition 2.2.1 in Chapter 2), and $\mathcal{M}(\mathbb{X}^n)$ is compact; this allows us to interchange inf and sup. Inequality (3.16) follows because infimum of expected compression lengths over all prefix codes is within 1 bit of entropy. The last equality follows from the well known variational characterization for Rényi entropy,

$$\sup_{P_{Y^n}} \{\theta H(P_{Y^n}) - D(P_{Y^n} \parallel P_{X^n})\} = \theta H_{\frac{1}{1+\theta}}(P_{X^n}), \tag{3.18}$$

a fact that can also be gleaned from the variational formula (3.10). Divide both sides of (3.17) by n and take limit supremum as $n \to \infty$ to get

$$\begin{aligned} E_u^s(R,\rho) &\leq \limsup_{n\to\infty} \min_{0\leq\theta\leq\rho} \left\{ (\rho-\theta)R + \frac{\theta}{n} H_{\frac{1}{1+\theta}}(P_{X^n}) \right\} \\ &\leq \min_{0\leq\theta\leq\rho} \left\{ (\rho-\theta)R + \theta \limsup_{n\to\infty} \frac{1}{n} H_{\frac{1}{1+\theta}}(P_{X^n}) \right\} \qquad (3.19) \\ &= \min_{0\leq\theta\leq\rho} \left\{ (\rho-\theta)R + \max_{\mathbf{Y}\in\mathcal{M}(\mathbb{X}^{\mathbb{N}})} E_u(\mathbf{Y},\mathbf{X},\theta) \right\}, \end{aligned}$$

where inequality in (3.19) follows by noting that

$$\min_{0\leq\theta\leq\rho} \left\{ (\rho-\theta)R + \frac{\theta}{n} H_{\frac{1}{1+\theta}}(P_{X^n}) \right\} \leq \left\{ (\rho-\theta^*)R + \frac{\theta^*}{n} H_{\frac{1}{1+\theta^*}}(P_{X^n}) \right\}$$

for every θ^* satisfying $0 \leq \theta^* \leq \rho$, and the last inequality follows from Proposition 3.3.1. This completes the proof. ■

From the above proof it is clear that the upper bound holds with equality, when Jensen's inequality holds with equality in (3.12), i.e, the random variable

$$(1/n)\min\{L_n(X^n), nR\}$$

tends asymptotically to a constant. This would happen, for example, when normalized encoded lengths concentrate around the entropy rate of the source.

3.3.2 Lower Bound on E_l^s

We now derive a lower bound on E_l^s. For a given distribution P_{Y^n} arrange the elements of set $\mathbb{X}^n$ in the decreasing order of their probabilities as done in Sundaresan [9, Sec. IV]. Enumerate the sequence from 1 to $|\mathbb{X}|^n$. Henceforth refer to a message by its index. Let $T_R(Y^n)$ denote the first $M = \lfloor \exp\{nR\} \rfloor$ elements in the list. We denote the probability of this set by F_{Y^n}, i.e.,

$$F_{Y^n} = \sum_{x^n \in T_R(Y^n)} P_{Y^n}(x^n),$$

and the probability of the complement of this set $T_R^c(Y^n)$ by $F_{Y^n}^c$. Let the restriction of P_{Y^n} to this set $T_R(Y^n)$ be P'_{Y^n}. Let L_n^* denote the length function that attains $E_n^s(R,\rho)$. As the length functions are uniquely decipherable we have $\exp\{L_n^*(i)\} \geq i$.

Proposition 3.3.4 (Lower Bound) *For a given $\rho > 0$ and rate $R > 0$, we have*

$$E_l^s(R,\rho) \geq \max\left\{\rho R + \liminf_{n\to\infty} \frac{1}{n} \log F_{X^n}^c, (1+\rho) \liminf_{n\to\infty} \frac{1}{n} \log \sum_{x^n \in T_R(X^n)} P_{X^n}^{\frac{1}{1+\rho}}(x^n)\right\}. \tag{3.20}$$

Remark 3.3.5 *The first term contains limit infimum of the error exponent for a rate-R source code. The second exponent is the correct decoding exponent for a rate-R code when $\rho \downarrow 0$.*

Proof: The variational formula (3.10) applied to the function $U(x^n) = \rho \min\{L_n(x^n), nR\}$ gives

$$\begin{aligned}
&\min_{L_n} \log \mathbb{E}_{X^n}\left[\exp\left\{\rho \min\left\{L_n(X^n), nR\right\}\right\}\right] \\
&\quad = \min_{L_n} \sup_{P_{Y^n}} \{\rho \mathbb{E}_{Y^n}[\min\{L_n(Y^n), nR\}] - D(P_{Y^n} \parallel P_{X^n})\} \\
&\quad \geq \sup_{P_{Y^n}} \left\{\rho \min_{L_n} \mathbb{E}_{Y^n}\left[\min\{L_n(X^n), nR\}\right] - D(P_{Y^n} \parallel P_{X^n})\right\}
\end{aligned} \tag{3.21}$$

where the interchange of min and sup yields the lower bound in (3.21). Fix a distribution P_{Y^n} and consider the first term in (3.21). Using the enumeration indicated above, we may write

$$\begin{aligned}
&\min_{L_n} \mathbb{E}_{Y^n} \left[\min\{L_n(Y^n), nR\}\right] \\
&= \sum_{i=1}^{|\mathbb{X}|^n} P_{Y^n}(i) \min\{L_n^*(i), nR\} \\
&= \sum_{i=1}^{M} P_{Y^n}(i) \min\{L_n^*(i), nR\} + \sum_{i=M+1}^{|\mathbb{X}|^n} P_{Y^n}(i) nR \\
&\geq \sum_{i=1}^{M} P_{Y^n}(i) \log G_n^*(i) + nRF_{Y^n}^c \qquad (3.22) \\
&\geq F_{Y^n} \sum_{i=1}^{M} \frac{P_{Y^n}(i)}{F_{Y^n}} L_{G_n^*}(i) - \log e(1 + n \log |\mathbb{X}|) + nRF_{Y^n}^c \qquad (3.23) \\
&\geq F_{Y^n} H(P'_{Y^n}) - \log e(1 + n \log |\mathbb{X}|) + nRF_{Y^n}^c. \qquad (3.24)
\end{aligned}$$

Inequality (3.22) follows because $L_n^*(i) \geq \log i = \log G_n^*(i)$ with G_n^* the guessing strategy that guesses in decreasing order of P_{Y^n} probabilities. $L_{G_n^*}$ in (3.23) denotes the length function given by (3.4). Inequality (3.24) follows from the source coding theorem's lower bound. Substitute (3.24) in (3.21), normalize by n, and take limit infimum to get

$$E_l^s(R, \rho) \geq \liminf_{n \to \infty} \frac{1}{n} \sup_{P_{Y^n}} \left\{ F_{Y^n} H(P'_{Y^n}) + F_{Y^n}^c nR - D(P_{Y^n} \parallel P_{X^n}) \right\}.$$

P_{Y^n} may be thought of as a triplet made of P'_{Y^n}, F_{Y^n}, and the restriction of P_{Y^n} to $T_R^c(Y^n)$. We now perform the optimization

$$\sup_{P_{Y^n}} \{F_{Y^n} H(P'_{Y^n}) + F_{Y^n}^c nR - D(P_{Y^n} \parallel P_{X^n})\} \qquad (3.25)$$

in four steps.

Step 1: We first optimize over permutations of strings. It is easy to verify that for the optimization it is enough to restrict attention to those distributions for which the set $T_R(Y^n)$ equals $T_R(X^n)$.

Step 2: We now optimize over restriction of P_{Y^n} to $T_R^c(Y^n)$. Fix a distribution P_{Y^n}, the sum in

$$\sum_{x^n \in T_R^c(X^n)} P_{Y^n}(x^n) \log \frac{P_{Y^n}(x^n)}{P_{X^n}(x^n)}$$

is minimized if for any two elements $x^n, y^n \in T_R(X^n)$, $P_{Y^n}(x^n) \geq P_{Y^n}(y^n) \Leftrightarrow P_{X^n}(x^n) \geq P_{X^n}(y^n)$. Indeed, by the log sum inequality we have

$$\sum_{x^n \in T_R^c(X^n)} P_{Y^n}(x^n) \log \frac{P_{Y^n}(x^n)}{P_{X^n}(x^n)} \geq F_{Y^n}^c \log \frac{F_{Y^n}^c}{F_{X^n}^c},$$

with equality if and only if $P_{Y^n}(x^n) = P_{X^n}(x^n)\frac{F_{Y^n}^c}{F_{X^n}^c}$ for all $x^n \in T_R(P_{X^n})$.

Step 3: To optimize over P'_{Y^n} rewrite (3.25) as

$$\begin{aligned}
&\sup_{P_{Y^n}} \left\{ \rho F_{Y^n} H(P'_{Y^n}) + F_{Y^n}^c \rho n R - \sum_{i=1}^{M} P_{Y^n}(i) \log \frac{P_{Y^n}(i)}{P_{X^n}(i)} - \sum_{M+1}^{\mathbb{X}^n} P_{Y^n}(i) \log \frac{P_{Y^n}(i)}{P_{X^n}(i)} \right\} \\
&= \sup_{P'_{Y^n}, F_{Y^n}} \left\{ \rho F_{Y^n} H(P'_{Y^n}) + F_{Y^n}^c \rho n R - \sum_{i=1}^{M} P_{Y^n}(i) \log \frac{P_{Y^n}(i)}{P_{X^n}(i)} - F_{Y^n}^c \log \frac{F_{Y^n}^c}{F_{X^n}^c} \right\} \quad (3.26) \\
&= \sup_{P'_{Y^n}, F_{Y^n}} \left\{ \rho F_{Y^n} H(P'_{Y^n}) + \rho n R F_{Y^n}^c - F_{Y^n} D(P'_{Y^n} \parallel P'_{X^n}) - D(F_{Y^n} || F_{X^n}) \right\} \\
&= \sup_{F_{Y^n}} \left\{ \rho F_{Y^n} H_{\frac{1}{1+\rho}}(P'_{X^n}) + F_{Y^n}^c \rho n R - D(F_{Y^n} \parallel F_{X^n}) \right\}. \quad (3.27)
\end{aligned}$$

Equality (3.26) is obtained by substituting attained lower bound in Step-2. In (3.27) P'_{Y^n} and P'_{X^n} denote conditional distributions of P_{Y^n} and P_{X^n} given $T_R(Y^n)$ and $T_R(X^n)$ respectively, which we argued were equal in Step-1. $D(F_{Y^n}||F_{X^n})$ denotes the divergence between binary random variables whose probabilities are $\{F_{Y^n}, 1 - F_{Y^n}\}$ and $\{F_{X^n}, 1 - F_{X^n}\}$ respectively. Finally we used variational characterization of Rényi entropy given in (3.18) to arrive at (3.27).

Step 4: We now optimize over $F_{Y^n} \in [0, 1]$. Let Z be a binary random variable defined as

$$Z = \begin{cases} \rho H_{\frac{1}{1+\rho}}(P'_{X^n}) & \text{with probability } F_{Y^n}, \\ \rho n R & \text{with probability } 1 - F_{Y^n} \end{cases}$$

By $\mathbb{E}_{F_{Y^n}}[Z]$ we mean the expectation of Z with respect to the the above distribution. Since Z is a positive random variable, the variational formula yields

$$\sup_{F_{Y^n}} \{\mathbb{E}_{F_{Y^n}}[Z] - D(F_{Y^n} \parallel F_{X^n})\} = \log \mathbb{E}_{F_{X^n}} [\exp\{Z\}] .$$

Continuing with the chain of equalities from (3.27) we get

$$\begin{aligned}
&\sup_{F_{Y^n}} \left\{ F_{Y^n} \rho H_{\frac{1}{1+\rho}}(P'_{X^n}) + F^c_{Y^n} \rho n R - D(F_{Y^n} \parallel F_{X^n}) \right\} \\
&\quad = \log \left\{ F^c_{X^n} \exp\{nR\rho\} + F_{X^n} \left(\sum_{i=1}^{M} {P'_{X^n}}^{\frac{1}{1+\rho}}(i) \right)^{1+\rho} \right\} \\
&\quad = \quad \log \left\{ F^c_{X^n} \exp\{nR\rho\} + \left(\sum_{i=1}^{M} P_{X^n}^{\frac{1}{1+\rho}}(i) \right)^{1+\rho} \right\} . \qquad (3.28)
\end{aligned}$$

Finally normalize both sides of (3.28) by n, take limit infimum, and apply [21, Lemma 1.2.15], which states that the exponential rate of a sum is governed by the maximum of the individual terms' exponential rates, to get the desired result. ■

In the subsequent subsections we further lower bound each of the two terms under max on the right side of (3.20). For an arbitrary source we first recall the source coding error exponent. We also identify the growth rate of sum of exponentiated probabilities of the correct decoding set. We then relate these to the terms in the lower bound obtained in (3.20). We largely follow the approach and notation of Iriyama [13], which we now describe.

For the given $\mathbf{X} = (P_{X^n} : n \in \mathbb{N})$ and a $\mathbf{Y} = (P_{Y^n} : n \in \mathbb{N})$, we define the upper divergence $D_u(\cdot \parallel \cdot)$ and lower divergence $D_l(\cdot \parallel \cdot)$ by

$$D_u(\mathbf{Y} \parallel \mathbf{X}) := \limsup_{n\to\infty} \frac{1}{n} D(P_{Y^n} \parallel P_{X^n})$$

$$D_l(\mathbf{Y} \parallel \mathbf{X}) := \liminf_{n\to\infty} \frac{1}{n} D(P_{Y^n} \parallel P_{X^n}).$$

For a $\mathbf{Y} = (P_{Y^n} : n \in \mathbb{N})$, denote *spectral sup-entropy-rate* [12, Sec. II], [11] as

$$\overline{H}(\mathbf{Y}) := \inf \left\{ \theta : \lim_{n\to\infty} \Pr \left\{ \frac{1}{n} \log \frac{1}{P_{Y^n}(Y^n)} > \theta \right\} = 0 \right\}.$$

Also define, as in [13, Sec. II], the following quantity which determines the performance under mismatched compression:

$$\underline{R}(\mathbf{Y}, \mathbf{X}) := \sup \left\{ \theta : \lim_{n\to\infty} \Pr \left\{ \frac{1}{n} \log \frac{1}{P_{X^n}(Y^n)} < \theta \right\} = 0 \right\}.$$

Decoding Error Exponent

In this subsection we recall the decoding error exponent for fixed-rate encoding of an arbitrary source. We identify the first term in (3.20) as composed of the exponent of minimum probability of decoding error, and obtain a lower bound for it, or alternatively an upper bound on the error exponent. This is made precise in the following definitions.

By an (n, M_n, ϵ_n)-code we mean an encoding mapping

$$\phi_n : \mathbb{X}^n \to \{1, 2, \cdots, M_n\}$$

and a decoding mapping

$$\psi_n : \{1, 2, \cdots M_n\} \to \mathbb{X}^n$$

with probability of error $\epsilon_n := \Pr\{\psi_n(\phi_n(X^n)) \neq X^n\}$. R is r-achievable if for all $\eta > 0$ there exists a sequence of (n, M_n, ϵ_n)-codes such that

$$\limsup_{n\to\infty} \frac{1}{n} \log \frac{1}{\epsilon_n} \geq r \tag{3.29}$$

$$\limsup_{n\to\infty} \frac{1}{n} \log M_n \leq R + \eta. \tag{3.30}$$

The *infimum fixed-length coding rate* for exponent r is

$$\hat{R}(r|\mathbf{X}) = \inf\{R : R \text{ is } r\text{-achievable}\}.$$

On the other hand, the *supremum fixed-length coding exponent* for rate R is

$$\hat{E}(R|\mathbf{X}) = \sup\{r : R \text{ is } r\text{-achievable}\}.$$

Han [13] and Iriyama [11, Sec. 1.9] use a pessimistic definition for fixed rate source coding, i.e., the limit infimum in (3.29), and obtain expressions for the infimum coding rate. For our bounds we need optimistic definitions. Iriyama [13, Eqn. (13)] obtained a lower bound on the infimum coding rate $\hat{R}(r|\mathbf{X})$ under the optimistic definition. We however work with the error exponent, and obtain an upper bound on supremum coding exponent. This suffices to lower bound the first term in (3.20).

Obviously, the best exponent is obtained by encoding only the highest M realizations and hence

$$\hat{E}(R|\mathbf{X}) = \limsup_{n\to\infty} \frac{1}{n} \log \frac{1}{F^c_{X^n}}$$

so that

$$-\hat{E}(R|\mathbf{X}) = \liminf_{n\to\infty} \frac{1}{n} \log F^c_{X^n}.$$

The following Proposition upper bounds the supremum coding exponent.

Proposition 3.3.6 *For any rate $R > 0$,*

$$\hat{E}(R|\mathbf{X}) \leq \inf_{\mathbf{Y}:\underline{R}(\mathbf{Y},\mathbf{X}) - D_u(\mathbf{Y}\|\mathbf{X}) > R} D_u(\mathbf{Y} \parallel \mathbf{X}).$$

Proof: See section 3.5.3. ■

Remark 3.3.7 *Notice that when $R \geq \log|\mathbb{X}|$, we have an infimum over an empty set and hence $\hat{E}(R|\mathbf{X}) = \infty$.*

Correct Decoding Exponent

We now study a generalization of the exponential rate for probability of correct decoding.

For a given (n, M_n, ϵ_n)-code, let

$$A_n := \{x^n \in \mathbb{X}^n : \psi_n(\phi_n(x^n)) = x^n\}$$

denote the set of correctly decoded sequences. For a given $\rho > 0$, R is (r, ρ)-admissible if for every $\eta > 0$ there exists a sequence of (n, M_n, ϵ_n)-codes such that

$$(1+\rho) \liminf_{n\to\infty} \frac{1}{n} \log \sum_{x^n \in A_n} P_{X^n}^{\frac{1}{1+\rho}}(x^n) \geq r \tag{3.31}$$

$$\limsup_{n\to\infty} \frac{1}{n} \log M_n \leq R + \eta. \tag{3.32}$$

The *infimum fixed-length admissible rate* for a given r and $\rho > 0$ is

$$R^*(r, \rho|\mathbf{X}) = \inf\{R : R \text{ is } (r, \rho)\text{-admissible}\}.$$

Clearly, the set $\{R : R \textit{ is } (r, \rho)\text{-admissible}\}$ is closed and so $R^*(r, \rho|\mathbf{X})$ is (r, ρ)-admissible. The *supremum fixed-length coding exponent* for a given R and ρ is

$$E^*(R, \rho|\mathbf{X}) = \sup\{r : R \text{ is } (r, \rho)\text{-admissible}\}.$$

Remark 3.3.8 *The choice of limit infimum in (3.31) makes the definition of admissibility pessimistic. For $\rho \downarrow 0$ the above definitions reduce to the special case of exponential rate for probability of correct decoding (see [11, Sec. 1.10]).*

Clearly, A_n should be $T_R(X^n)$ to maximise the left side of (3.31) and hence

$$E^*(R, \rho|\mathbf{X}) = (1+\rho) \liminf_{n\to\infty} \frac{1}{n} \log \sum_{x^n \in T_R(X^n)} P_{X^n}^{\frac{1}{1+\rho}}(x^n).$$

The following Proposition gives an expression for $E^*(R, \rho|\mathbf{X})$ and generalizes [13, Thm. 4] to any arbitrary $\rho > 0$. En route to its derivation we find the expression for $R^*(r, \rho|\mathbf{X})$.

Proposition 3.3.9 *For any $\rho > 0$ and $r > 0$,*

$$R^*(r, \rho|\mathbf{X}) = \inf_{\mathbf{Y}: E_l(\mathbf{Y}, \mathbf{X}, \rho) \geq r} \overline{H}(\mathbf{Y}) \tag{3.33}$$

$$E^*(R, \rho|\mathbf{X}) = \sup_{\overline{H}(\mathbf{Y}) \leq R} E_l(\mathbf{Y}, \mathbf{X}, \rho). \tag{3.34}$$

Proof: See section 3.5.4 ■

3.3.3 Summary of Bounds on E_u^s and E_l^s

We now combine Propositions 3.3.3, 3.3.4, 3.3.6, and 3.3.9 of the previous subsections to obtain the main result of this chapter.

Theorem 3.3.10 *For a given $\rho > 0$ and $R > 0$,*

$$\begin{aligned} & \max\left\{\rho R - \inf_{\mathbf{Y}:\underline{R}(\mathbf{Y},\mathbf{X})-D_u(\mathbf{Y}\|\mathbf{X})>R} D_u(\mathbf{Y} \parallel \mathbf{X}), \sup_{\overline{H}(\mathbf{Y})\le R} E_l(\mathbf{Y},\mathbf{X},\rho)\right\} \\ & \le \quad E_l^s(R,\rho) \le E_u^s(R,\rho) \\ & \le \quad \min_{0\le\theta\le\rho}\left\{(\rho-\theta)R + \max_{\mathbf{Y}} E_u(\mathbf{Y},\mathbf{X},\theta)\right\} \qquad . \end{aligned} \tag{3.35}$$

□

Proof: The last inequality was proved in Proposition 3.3.3. Proposition 3.3.4 indicates that

$$\begin{aligned} & E_l^s(R,\rho) \\ & \ge \quad \max\left\{\rho R + \liminf_{n\to\infty}\frac{1}{n}\log F_{X^n}^c, (1+\rho)\liminf_{n\to\infty}\frac{1}{n}\log\sum_{x^n\in T_R(X^n)} P_{X^n}^{\frac{1}{1+\rho}}(x^n)\right\} \\ & = \quad \max\left\{\rho R - \hat{E}(R|\mathbf{X}), E^*(R,\rho|\mathbf{X})\right\} \qquad (3.36) \\ & \ge \quad \max\left\{\rho R - \inf_{\mathbf{Y}:\underline{R}(\mathbf{Y},\mathbf{X})-D_u(\mathbf{Y}\|\mathbf{X})>R} D_u(\mathbf{Y} \parallel \mathbf{X}), \sup_{\overline{H}(\mathbf{Y})\le R} E_l(\mathbf{Y},\mathbf{X},\rho)\right\}, \qquad (3.37) \end{aligned}$$

where (3.36) follows from the definitions of $\hat{E}(R|\mathbf{X})$ and $E^*(R,\rho|\mathbf{X})$, and (3.37) from Propositions 3.3.6 and 3.3.9. ■

3.4 Examples

In this section we evaluate the bounds for some examples where they are tight, and recover some known results.

Example 3.4.1 (Perfect Secrecy) First consider the perfect secrecy case, for example, $R \geq \log|\mathbb{X}|$. Because of Remark 3.3.7 and because we may take $\theta = \rho$ in the upper bound in (3.35), the limiting exponential rate of guessing moments simplifies to

$$\sup_{\mathbf{Y}} E_l(\mathbf{Y}, \mathbf{X}, \rho) \leq E_l^s(R, \rho) \leq E_u^s(R, \rho) \leq \max_{\mathbf{Y}} E_u(\mathbf{Y}, \mathbf{X}, \rho).$$

On account of (3.8) in Proposition 3.3.1, sup in the left-most term is achieved. From Proposition 3.3.1, upper and lower bounds are ρ times the liminf and limsup Rényi entropy rates of order $\frac{1}{1+\rho}$. In Proposition 2.2.3 of chapter 2 we showed that whenever the *information spectrum* of the source satisfies the large deviation property with rate function I, the Rényi entropy rate converges and limiting guessing exponent equals the Legendre-Fenchel dual of the scaled rate function $I_1(t) := (1+\rho)I(t)$, i.e.,

$$E_u^s(\rho) = E_l^s(\rho) = \sup_{t \in \mathbb{R}} \{\rho t - I_1(t)\}.$$

In the next examples, we consider the case $R < \log|\mathbb{X}|$.

Example 3.4.2 (An iid source) This example was first studied by Merhav & Arikan [2]. Recall that an iid source is one for which $P_n(x^n) = \prod_{i=1}^n P_1(x_i)$, where P_1 denotes the marginal of X_1. We will now evaluate each term in (3.35).

We first argue that

$$\inf_{\mathbf{Y}: \underline{R}(\mathbf{Y},\mathbf{X}) - D_u(\mathbf{Y}\|\mathbf{X}) > R} D_u(\mathbf{Y} \parallel \mathbf{X}) = \inf_{P_Y : H(P_Y) > R} D(P_Y \parallel P_1). \tag{3.38}$$

To prove that the left side in (3.38) is less than or equal to the right side, let $P_Y \in \mathcal{M}(\mathbb{X})$ be such that $H(P_Y) > R$. Construct an iid source $\hat{\mathbf{Y}}$ $(P_{\hat{Y}^n} : n \in \mathbb{X})$ such that $P_{\hat{Y}_i} = P_Y$ for all $1 \leq i \leq n$. By definition of $D_u(\hat{\mathbf{Y}} \parallel \mathbf{X})$ we get,

$$D_u(\hat{\mathbf{Y}} \parallel \mathbf{X}) = D(P_Y \parallel P_1)$$

and by definition of $\underline{R}(\hat{\mathbf{Y}}, \mathbf{X})$ we have

$$\underline{R}(\hat{\mathbf{Y}}, \mathbf{X}) - D_u(\hat{\mathbf{Y}} \parallel \mathbf{X}) = H(P_Y) > R. \tag{3.39}$$

From (3.39), we conclude that " $\leq$" holds in (3.38).

To prove "$\geq$" in (3.38) we use the following result:

$$\inf_{\mathbf{Y}:\underline{R}(\mathbf{Y},\mathbf{X})-D_u(\mathbf{Y}\|\mathbf{X})>R} D_u(\mathbf{Y} \parallel \mathbf{X}) \geq \inf_{\mathbf{Y}:H_l(\mathbf{Y})>R} D_u(\mathbf{Y} \parallel \mathbf{X}), \tag{3.40}$$

where $H_l(\mathbf{Y}) = \liminf_{n\to\infty} \frac{1}{n} H(P_{Y^n})$. Proof of above inequality follows from a straightforward manipulation of [13, Cor. 1], and is therefore omitted. Because of (3.40) it is sufficient to prove

$$\inf_{\mathbf{Y}:H_l(\mathbf{Y})>R} D_u(\mathbf{Y} \parallel \mathbf{X}) \geq \inf_{P_Y:H(P_Y)>R} D(P_Y \parallel P_1). \tag{3.41}$$

Let $\mathbf{Y}$ be such that $H_l(\mathbf{Y}) > R$. Construct a source $\hat{\mathbf{Y}}$ such that, $P_{\hat{Y}_i} = P_{Y_i}$ for $1 \leq i \leq n$ and $\hat{Y}_1, \hat{Y}_2, \cdots, \hat{Y}_n$ are independent. Let $\mathbf{Z}$ be another source such that, $Z_1, Z_2, \cdots, Z_n$ is an iid sequence with distribution

$$P_{Z_j} = \frac{1}{n} \sum_{i=1}^{n} P_{Y_i}, \quad j = 1, 2, \cdots, n.$$

Now, by convexity of divergence, we have

$$\begin{aligned} D(P_{Y^n} \parallel P_{X^n}) &= D(P_{Y^n} \parallel P_{\hat{Y}^n}) + D(P_{\hat{Y}^n} \parallel P_{X^n}) \\ &\geq D(P_{\hat{Y}^n} \parallel P_{X^n}) \geq D(P_{Z^n} \parallel P_{X^n}) \\ &= nD(P_{Z_1} \parallel P_1) \end{aligned} \tag{3.42}$$

and by concavity of Shannon entropy

$$H(P_{Y^n}) \leq \sum_{i=1}^{n} H(P_{Y_i}) \leq nH(P_{Z_1}). \tag{3.43}$$

Normalize by n take limsup in (3.42) and liminf in (3.43) to get $D_u(\mathbf{Y} \parallel \mathbf{X}) \geq D(P_{Z_1} \parallel P_1)$ and $H(P_{Z_1}) > R$ for a P_{Z_1} that is a limit point of the sequence $(n^{-1}\sum_{i=1}^{n} P_{Y_i}, n \in \mathbb{N})$. From these we conclude that (3.41) holds. This proves (3.38). Following a similar procedure as above, we can bound the other terms in (3.35) for an iid source as

$$\sup_{\mathbf{Y}:\overline{H}(\mathbf{Y})\leq R} E_l(\mathbf{Y}, \mathbf{X}, \rho) \geq \sup_{P_Y : H(P_Y)\leq R} \{\rho H(P_Y) - D(P_Y \parallel P_1)\} \tag{3.44}$$

and

$$\sup_{\mathbf{Y}} E_u(\mathbf{Y}, \mathbf{X}, \theta) = \sup_{P_Y} \{\theta H(P_Y) - D(P_Y \parallel P_1)\}. \tag{3.45}$$

Substitution of (3.38) and (3.44) in the lower bound of (3.35) yields

$$\begin{aligned} E_l^s(R, \rho) &\geq \max\left\{\rho R - \inf_{P_Y : H(P_Y) > R} D(P_Y \parallel P_1), \sup_{P_Y : H(P_Y)\leq R} \{\rho H(P_Y) - D(P_Y \parallel P_1)\right\} \\ &= \sup_{P_Y} \left\{\rho \min\{H(P_Y), R\} - D(P_Y \parallel P_1)\right\}. \end{aligned} \tag{3.46}$$

Similarly substitution of (3.45) in the upper bound of (3.35) yields

$$\begin{aligned} E_u^s(R, \rho) &\leq \min_{0\leq\theta\leq\rho} \left\{(\rho - \theta)R + \sup_{P_Y}\{\theta H(P_Y) - D(P_Y \parallel P_1)\}\right\} \\ &= \sup_{P_Y} \left\{\rho \min_{0\leq\theta\leq\rho} \{(\rho - \theta)R + \theta H(P_Y)\} - D(P_Y \parallel P_1)\right\} \qquad (3.47) \\ &= \sup_{P_Y} \left\{\rho \min\{H(P_Y, R)\} - D(P_Y \parallel P_1)\right\}, \qquad (3.48) \end{aligned}$$

where the interchange of sup and min in (3.47) holds because the function within braces is linear in θ and concave in P_Y. From (3.46) and (3.48), we recover Merhav & Arikan's result (1.4) for an iid source [2, Eqn. (3)].

Example 3.4.3 (Markov source) In this example we focus on an irreducible stationary Markov source taking values on $\mathbb{X}$ and having a transition probability matrix π.

Let $\mathcal{M}_s(\mathbb{X}^2)$ denote the set of *stationary* pmfs defined by

$$\mathcal{M}_s\left(\mathbb{X}^2\right) = \Big\{Q \in \mathcal{M}\left(\mathbb{X}^2\right) : \sum_{x_1 \in \mathbb{X}} Q(x_1, x) = \sum_{x_2 \in \mathbb{X}} Q(x, x_2), \forall x \in \mathbb{X}\Big\}.$$

Denote the common marginal by q and let

$$\eta(\cdot \mid x_1) := \begin{cases} Q(x_1, \cdot)/q(x_1), & \text{if } q(x_1) \neq 0, \\ 1/|\mathbb{X}|, & \text{otherwise.} \end{cases}$$

We may then denote $Q = q \times \eta$, where q is the distribution of X_1 and η the conditional distribution of X_2 given X_1. Following steps similar to the iid case, we have

$$E_u^s = E_l^s = \sup_{Q \in \mathcal{M}_s(\mathbb{X}^2)} \Big\{\rho \min\{H(\eta \mid q), R\} - D(\eta \parallel \pi \mid q)\Big\},$$

where

$$H(\eta \mid q) := \sum_{x \in \mathbb{X}} q(x) H(\eta(\cdot \mid x)).$$

is the conditional one-step entropy, and

$$D(\eta \parallel \pi \mid q) = \sum_{x_1 \in \mathbb{X}} q(x_1) D(\eta(\cdot \mid x_1) \parallel \pi(\cdot \mid x_1)).$$

For a unifilar source the underlying state space forms a Markov chain and the entropy and divergence of the source equals those of the underlying Markov state space source [30, Thm. 6.4.2]. The arguments for the Markov source are now directly applicable to a unifilar source.

3.5 Proofs

3.5.1 Proof of Proposition 3.2.2

Let P_n be any pmf on $\mathbb{X}^n$. Enumerate the elements of $\mathbb{X}^n$ in the decreasing order of their probabilities. For convenience, let $M = \exp\{nR\}$. If M does not divide $|\mathbb{X}|^n$, append a

few dummy messages of zero probability to make the number of messages N a multiple of M. Index the messages from 0 to $N-1$. Henceforth, we identify a message by its index.

Divide the messages into groups of M so that message m belongs to group T_j, where $j = \lfloor m/M \rfloor$, and $\lfloor \cdot \rfloor$ is the floor function. Enumerate the key streams from 0 to $M-1$, so that $0 \leq u \leq M-1$. The function f_n is now defined as follows. For $m = jM + i$ set

$$f_n(jM+i, u) \stackrel{\Delta}{=} jM + (i \oplus u),$$

where $i \oplus u$ is the bit-wise XOR operation. Thus messages in group T_j are encrypted to messages in the same group. The index i identifying the specific message in group T_j, i.e., the last nR bits of m, are encrypted via bit-wise XOR with the key stream. Given u and the cryptogram, decryption is clear – perform bit-wise XOR with u on the last nR bits of y.

Given a cryptogram y, the only information that the attacker gleans is that the message belongs to the group determined by y. Indeed, if $y \in T_j$

$$P_n\{Y = y\} = \frac{1}{M} P_n\{X^n \in T_j\}$$

and therefore

$$P_n\{X^n = m \mid Y = y\} = \begin{cases} \frac{P_n\{X^n = m\}}{P_n\{X^n \in T_j\}}, & \lfloor m/M \rfloor = j, \\ 0, & \text{otherwise}, \end{cases}$$

decreases with m for $m \in T_j$, and is 0 for $m \notin T_j$. The attacker's best strategy $G_{f_n}(\cdot \mid y)$ is therefore to restrict his guesses to T_j and guess in the order $jM, jM+1, \cdots, jM+M-1$. Thus, when $x^n = jM + i$, the optimal attack strategy requires $i+1$ guesses.

We now analyze the performance of this attack strategy as follows.

$$\begin{aligned}
\mathbb{E}\left[G_{f_n}(X^n|Y)^\rho\right] &= \sum_{j=0}^{N/M-1}\sum_{i=0}^{M-1} P_n\{X^n = jM+i\}(i+1)^\rho \\
&\geq \sum_{j=0}^{N/M-1}\sum_{i=0}^{M-1} P_n\{X^n = (j+1)M-1\}(i+1)^\rho \qquad (3.49)\\
&\geq \sum_{j=0}^{N/M-1} P_n\{X^n = (j+1)M-1\}\frac{M^{1+\rho}}{1+\rho} \qquad (3.50)\\
&\geq \frac{1}{1+\rho}\sum_{j=0}^{N/M-1}\sum_{i=0}^{M-1} P_n\{X^n = (j+1)M+i\}M^\rho \\
& \qquad (3.51)\\
&= \frac{1}{1+\rho}\sum_{m=M}^{N-1} P_n\{X^n = m\}M^\rho \qquad (3.52)
\end{aligned}$$

where (3.49) follows because the arrangement in the decreasing order of probabilities implies that

$$P_n\{X^n = jM+i\} \geq P_n\{X^n = (j+1)M-1\}$$

for $i = 0, \cdots, M-1$. Inequality (3.50) follows because

$$\sum_{i=0}^{M-1}(i+1)^\rho = \sum_{i=1}^{M} i^\rho \geq \int_0^M z^\rho\, dz = \frac{M^{1+\rho}}{1+\rho},$$

inequality (3.51) follows because the decreasing probability arrangement implies

$$P_n\{X^n = (j+1)M-1\} \geq \frac{1}{M}\sum_{i=0}^{M-1} P_n\{X^n = (j+1)M+i\}.$$

Thus (3.52) implies that

$$\begin{aligned}
&\sum_{m=0}^{N-1} P_n\{X^n = m\}\,(\min\{m+1, M\})^\rho \\
&= \sum_{m=0}^{M-1} P_n\{X^n = m\}(m+1)^\rho + \sum_{m=M}^{N-1} P_n\{X^n = m\}M^\rho \\
&\leq \mathbb{E}\left[G_{f_n}(X^n|Y)^\rho\right] + (1+\rho)\mathbb{E}\left[G_{f_n}(X^n|Y)^\rho\right] \\
&= (2+\rho)\mathbb{E}\left[G_{f_n}(X^n|Y)^\rho\right]. \qquad (3.53)
\end{aligned}$$

Let G be the guessing function that guesses in the decreasing order of P_n-probabilities without regard to Y, i.e., $G(m) = m+1$. Let L_G be the associated length function, given in (3.4). Now use (3.53), Proposition 2.1.2, and (3.4) to get

$$\begin{aligned}
\mathbb{E}\left[G_{f_n}(X^n|Y)^\rho\right] &\geq \frac{1}{2+\rho}\mathbb{E}\left[(\min\{G(X^n), M\})^\rho\right] \\
&\geq \frac{1}{2+\rho}\mathbb{E}\left[\left(\min\left\{\frac{\exp\{L_G(X^n)\}}{ec_n}, M\right\}\right)^\rho\right] \\
&\geq \frac{1}{(ec_n)^\rho(2+\rho)}\mathbb{E}\left[\exp\left\{\rho\min\left\{L_G(X^n), nR\right\}\right\}\right].
\end{aligned}$$

Since G_{f_n} is the strategy that minimizes $\mathbb{E}\left[G(X^n \mid Y)^\rho\right]$, the proof is complete. ■

3.5.2 Proof of Proposition 3.3.1

We begin with the following Lemma. Recall that $\mathcal{M}(\mathbb{X})$ is the set of all probability measures on $\mathbb{X}$ and $\mathcal{M}(B)$ the subset of $\mathcal{M}(\mathbb{X})$ with support set $B \subseteq \mathbb{X}$:

$$\mathcal{M}(B) = \{\nu \in \mathcal{M}(\mathbb{X}) : \nu(B) = 1\}.$$

Lemma 3.5.1 *For any $\rho > 0, \mu \in \mathcal{M}(\mathbb{X})$ and $B \subseteq \mathbb{X}$*

$$(1+\rho)\log\sum_{x\in B}\mu^{\frac{1}{1+\rho}}(x) = \max_{\nu\in M(B)}\{\rho H(\nu) - D(\nu \parallel \mu)\}.$$

□

Proof: Let $\mu_B(x) = \frac{\mu(x)}{\mu(B)} 1\{x \in B\}$. We then have

$$\begin{aligned}
&(1+\rho)\log \sum_{x\in B} \mu^{\frac{1}{1+\rho}}(x) \\
&= (1+\rho)\log \sum_{x\in B} {\mu_B}^{\frac{1}{1+\rho}}(x) + \log \mu(B) \\
&= (1+\rho) \max_{\nu\in\mathcal{M}(B)} \left\{ \sum_{x\in B} \frac{\rho}{1+\rho}\nu(x)\log\frac{1}{\mu_B(x)} - D(\nu \parallel \mu_B) \right\} + \log\mu(B) \qquad (3.54) \\
&= (1+\rho) \max_{\nu\in\mathcal{M}(B)} \left\{ \frac{\rho}{1+\rho}\left\{H(\nu) + D(\nu \parallel \mu)\right\} - D(\nu \parallel \mu) \right\} \qquad (3.55) \\
&= \max_{\nu\in\mathcal{M}(B)} \left\{\rho H(\nu) - D(\nu \parallel \mu)\right\}. \qquad (3.56)
\end{aligned}$$

where (3.54) follows from the variational formula for Rényi entropy of μ_B. The maximum achieving distribution in (3.56) is $\mu^* \in \mathcal{M}(B)$ given by

$$\mu^*(x) = \frac{\mu^{\frac{1}{1+\rho}}(x)}{\sum_{y\in B}\mu^{\frac{1}{1+\rho}}(y)} 1\{x \in B\}.$$

■

Remark 3.5.2 *[13, Lemma 1] is the special case when $\rho = 0$.*

We now prove (3.8); proof of (3.7) is similar and therefore omitted. We begin by showing "$\leq$" in (3.8). Let $\mathbf{X}^* = (P^*_{X^n} : n \in \mathbb{N}) \in \mathcal{M}(\mathbf{B})$ be as defined in (3.9). It is straightforward to verify by direct substitution that

$$(1+\rho)\log \sum_{x^n\in B_n} P^{\frac{1}{1+\rho}}_{X^n}(x^n) = \rho H(P^*_{X^n}) - D(P^*_{X^n} \parallel P_{X^n}).$$

Normalize by n and take limit infimum, and use the definition of $E_l(\mathbf{X}^*, \mathbf{X}, \rho)$ to get

$$\begin{aligned}
(1+\rho)\liminf_{n\to\infty} \frac{1}{n}\log \sum_{x\in B_n} P^{\frac{1}{1+\rho}}_{X^n}(x) &= E_l(\mathbf{X}^*, \mathbf{X}, \rho) \qquad (3.57) \\
&\leq \max_{\mathbf{Y}\in\mathcal{M}(\mathbf{B})} E_l(\mathbf{Y}, \mathbf{X}, \rho).
\end{aligned}$$

To prove "$\geq$" in (3.8), let $\mathbf{Y} = (P_{Y^n} : n \in \mathbb{N}) \in \mathcal{M}(\mathbf{B})$ be an arbitrary sequence. We may assume that for all sufficiently large n, $P_{Y^n} \ll P_{X^n}$ holds; otherwise $E_l(\mathbf{Y}, \mathbf{X}, \rho) = -\infty$ and the inequality "$\geq$" holds automatically. Define $\mathbf{Y}^* = (P^*_{Y^n} : n \in \mathbb{N}) \in \mathcal{M}(\mathbf{B})$ by

$$P^*_{Y^n}(y) = \frac{P_{Y^n}(y)}{P_{Y^n}(B_n)} 1\{y \in B_n\}.$$

It is clear that $P^*_{Y^n} \in \mathcal{M}(B_n)$ for every n. From Lemma 3.5.1, we have

$$\begin{aligned}(1+\rho)\log \sum_{x^n \in B_n} P_{X^n}^{\frac{1}{1+\rho}}(x^n) &= \max_{P_{Y^n} \in \mathcal{M}(B_n)} \{\rho H(P_{Y^n}) - D(P_{Y^n} \parallel P_{X^n})\} \\ &\geq \rho H(P^*_{Y^n}) - D(P^*_{Y^n} \parallel P_{X^n}). \end{aligned} \tag{3.58}$$

We now study each term on the right side of (3.58). The entropy term is lower bounded as follows.

$$\begin{aligned}&\rho H(P^*_{Y^n}) \\ &= \frac{\rho}{P_{Y^n}(B_n)} \left\{ \sum_{x^n \in B_n} P_{Y^n}(x^n) \log \frac{1}{P_{Y^n}(x^n)} \right\} + \rho \log P_{Y^n}(B_n) \\ &= \frac{\rho}{P_{Y^n}(B_n)} \left\{ H(P_{Y^n}) - \sum_{x^n \in B_n^c} P_{Y^n}(x^n) \log \frac{1}{P_{Y^n}(x^n)} \right\} + \rho \log P_{Y^n}(B_n) \\ &= \frac{\rho}{P_{Y^n}(B_n)} \{H(P_{Y^n}) - P_{Y^n}(B_n^c) H(P_{Y^n}|B_n^c) + P_{Y^n}(B_n^c) \log P_{Y^n}(B_n^c)\} + \rho \log P_{Y^n}(B_n) \\ &\geq \frac{\rho}{P_{Y^n}(B_n)} \{H(P_{Y^n}) - P_{Y^n}(B_n^c) n \log \mathbb{X} + P_{Y^n}(B_n^c) \log P_{Y^n}(B_n^c)\} + \rho \log P_{Y^n}(B_n). \end{aligned} \tag{3.59}$$

The divergence term is upperbounded, as in the proof of Iriyama's [13, Prop. 1], as follows:

$$\begin{aligned}
&D(P^*_{Y^n} \parallel P_{X^n}) \\
&= -\log P_{Y^n}(B_n) + \frac{1}{P_{Y^n}(B_n)} \sum_{x^n \in B_n} P_{Y^n}(x^n) \log \frac{P_{Y^n}(x^n)}{P_{X^n}(x^n)} \\
&= -\log P_{Y^n}(B_n) + \frac{1}{P_{Y^n}(B_n)} D(P_{Y^n} \parallel P_{X^n}) - \frac{1}{P_{Y^n}(B_n)} \sum_{x^n \in B_n^c} P_{Y^n}(x^n) \log \frac{P_{Y^n}(x^n)}{P_{X^n}(x^n)} \\
&\leq -\log P_{Y^n}(B_n) + \frac{1}{P_{Y^n}(B_n)} D(P_{Y^n} \parallel P_{X^n}) - \frac{P_{Y^n}(B_n^c) - P_{X^n}(B_n^c)}{P_{Y^n}(B_n)} \qquad (3.60) \\
&\leq -\log P_{Y^n}(B_n) + \frac{1}{P_{Y^n}(B_n)} D(P_{Y^n} \parallel P_{X^n}) + \frac{1}{P_{Y^n}(B_n)}. \qquad (3.61)
\end{aligned}$$

In inequality (3.60) we used the fact that $\log x \geq 1 - \frac{1}{x}$ for all $x > 0$ and in inequality (3.61) we used the relation $P_{Y^n}(B_n^c) - P_{X^n}(B_n^c) \geq -1$. Substitution of (3.59) and (3.61) in (3.58) and the fact that $\lim_{n\to\infty} P_{Y^n}(B_n) = 1$ yields

$$\begin{aligned}
&(1+\rho) \liminf_{n\to\infty} \frac{1}{n} \log \sum_{x^n \in B_n} P_{X^n}^{\frac{1}{1+\rho}}(x^n) \\
&\geq \liminf_{n\to\infty} \frac{1}{n} \{\rho H(P_{Y^n}) - D(P_{Y^n} \parallel P_{X^n}) - O(1)\} \\
&= E_l(\mathbf{Y}, \mathbf{X}, \rho).
\end{aligned}$$

Since the choice of $\mathbf{Y} = (P_{Y^n} : n \in \mathbb{N}) \in \mathcal{M}(\mathbf{B})$ was arbitrary, we have proved "$\geq$" in (3.8).

From (3.57) and (3.8), it is clear that the maximum is attained by $\mathbf{X}^*$, the distribution defined in (3.9). This completes the proof.

3.5.3 Proof of Proposition 3.3.6

The proof proceeds along lines similar to Iriyama [13, Th. 2] with modifications to account for our interest in $\hat{E}(R|\mathbf{X})$ instead of rate $\hat{R}(r|\mathbf{X})$. We begin with the following Lemma.

Let $(f_n : n \in \mathbb{N})$ denote a sequence of real functions on $(\mathbb{X}^n : n \in \mathbb{N})$. Define $\underline{f}$ by

$$\underline{f}(\mathbf{Y}) := \sup\left\{\theta : \lim_{n\to\infty} \Pr\left\{f_n(Y^n) < \theta\right\} = 0\right\}.$$

Lemma 3.5.3 *For all $a > -\infty$, we have*

$$-\inf_{\mathbf{Y}:\underline{f}(\mathbf{Y})>a} D_u(\mathbf{Y} \parallel \mathbf{X}) \le \liminf_{n\to\infty} \frac{1}{n} \log \Pr\{f_n(Y^n) \ge a\} \le -\inf_{\mathbf{Y}:\underline{f}(\mathbf{Y})\ge a} D_u(\mathbf{Y} \parallel \mathbf{X}).$$

□

Proof: Iriyama's [13, Prop. 2] is the same as above with D_l and limsup. That proof is applicable with obvious changes, and is therefore omitted. ■

Define

$$\tilde{E}(R|\mathbf{X}) = \inf_{\mathbf{Y}:\underline{R}(\mathbf{Y},\mathbf{X})-D_u(\mathbf{Y}\|\mathbf{X})>R} D_u(\mathbf{Y} \parallel \mathbf{X}).$$

We may assume $\tilde{E}(R|\mathbf{X}) < \infty$. Otherwise the Proposition holds trivially. Suppose there exists a sequence of (n, M_n, ϵ_n)-codes such that

$$r = \limsup_{n\to\infty} \frac{1}{n} \log \frac{1}{\epsilon_n} > \tilde{E}(R|\mathbf{X}). \tag{3.62}$$

We will show the contrapositive implication that

$$\limsup_{n\to\infty} \frac{1}{n} \log M_n > R + \eta \ \text{ for some } \ \eta > 0.$$

By definition of $\tilde{E}(R|\mathbf{X})$, since $r > \tilde{E}(R|\mathbf{X})$, there exists $\hat{\mathbf{Y}}$ such that $r > D_u(\hat{\mathbf{Y}} \parallel \mathbf{X})$, and $\underline{R}(\hat{\mathbf{Y}}, \mathbf{X}) - D_u(\hat{\mathbf{Y}} \parallel \mathbf{X}) > R$, i.e., $\underline{R}(\hat{\mathbf{Y}}, \mathbf{X}) - D_u(\hat{\mathbf{Y}} \parallel \mathbf{X}) > R + \eta_1$ for some $\eta_1 > 0$. For arbitrary $\delta > 0$ define

$$T_n := \left\{x^n : \frac{1}{n} \log \frac{1}{P_{X^n}(x^n)} \ge \underline{R}(\hat{\mathbf{Y}}, \mathbf{X}) - \delta\right\}$$

and the set of correctly decoded sequences

$$A_n := \{x^n \in \mathbb{X}^n : \psi_n(\phi_n(x^n)) = x^n\}.$$

Then

$$\epsilon_n = \Pr\{A_n^c\} \geq \Pr\{A_n^c \cap T_n\} = \Pr\{T_n\} - \Pr\{T_n \cap A_n\}$$

and so

$$\Pr\{T_n \cap A_n\} \geq \Pr\{T_n\} - \epsilon_n. \tag{3.63}$$

If we set $f_n(x^n) = \frac{1}{n} \log \frac{1}{P_{X^n}(x^n)}$, we have $\underline{f}(\hat{\mathbf{Y}}) = \underline{R}(\hat{\mathbf{Y}}, \mathbf{X})$. The first inequality of Lemma 3.5.3 then implies that

$$\begin{aligned} \liminf_{n\to\infty} \frac{1}{n} \log \Pr\{T_n\} &\geq -\inf_{\mathbf{Y}:\underline{R}(\mathbf{Y},\mathbf{X}) > \underline{R}(\hat{\mathbf{Y}},\mathbf{X})-\delta} D_u(\mathbf{Y} \parallel \mathbf{X}) \\ &\geq -D_u(\hat{\mathbf{Y}} \parallel \mathbf{X}). \end{aligned}$$

By definition of liminf there exists sufficiently large n_0 such that for all $n \geq n_0$

$$\Pr\{T_n\} \geq \exp\{-n(D_u(\hat{\mathbf{Y}} \parallel \mathbf{X}) + \delta)\}. \tag{3.64}$$

Next, by definition of limsup in (3.62) there exists a subsequence $(n_j : j \in \mathbb{N})$ such that

$$\epsilon_{n_j} \leq \exp\{-n_j(r - \delta)\}, \quad j \in \mathbb{N} \tag{3.65}$$

Also

$$\begin{aligned} \Pr(T_n \cap A_n) &= \sum_{x^n \in T_n \cap A_n} P_{X^n}(x^n) \leq |A_n| \exp\{-n(\underline{R}(\hat{\mathbf{Y}}, \mathbf{X}) - \delta)\} \\ &\leq M_n \exp\{-n(\underline{R}(\hat{\mathbf{Y}}, \mathbf{X}) - \delta)\}. \end{aligned} \tag{3.66}$$

Substitution of (3.64)-(3.66) in (3.63) yields

$$M_{n_j} \exp\left\{-n_j(\underline{R}(\hat{\mathbf{Y}}, \mathbf{X}) - \delta)\right\} \geq \exp\left\{-n_j(D_u(\hat{\mathbf{Y}} \parallel \mathbf{X}) + \delta)\right\} - \exp\left\{-n_j(r - \delta)\right\}$$

which on rearrangement gives

$$M_{n_j} \geq \exp\left\{n_j(\underline{R}(\hat{\mathbf{Y}}, \mathbf{X}) - D_u(\hat{\mathbf{Y}} \parallel \mathbf{X}) - 2\delta)\right\} \cdot \left(1 - \exp\left\{-n_j(r - D_u(\hat{\mathbf{Y}} \parallel \mathbf{X}) - 2\delta)\right\}\right) \tag{3.67}$$

for all sufficiently large j. Choose δ such that $2\delta < \min\left(r - D_u(\hat{\mathbf{Y}} \parallel \mathbf{X}), \eta_1\right)$. Take logarithm, normalize by n, and take limit as $j \to \infty$ in (3.67) to get

$$\begin{aligned} \limsup_{n\to\infty} \frac{1}{n} \log M_n &\geq \underline{R}(\hat{\mathbf{Y}}, \mathbf{X}) - D_u(\hat{\mathbf{Y}} \parallel \mathbf{X}) - 2\delta \\ &> R + \eta_1 - 2\delta = R + \eta \end{aligned}$$

for $\eta = \eta_1 - 2\delta$. This concludes the proof.

3.5.4 Proof of Proposition 3.3.9

We use the following notations in this proof. For each $\mathbf{B} = (B_n : n \in \mathbb{N})$ define

$$|\mathbf{B}| := \limsup_{n\to\infty} \frac{1}{n} \log |B_n|$$

and

$$S(\mathbf{Y}) := \left\{\mathbf{B} : \lim_{n\to\infty} P_{Y^n}(B_n) = 1\right\}.$$

Note that $\mathbf{B} \in S(\mathbf{Y}) \Leftrightarrow \mathbf{Y} \in \mathcal{M}(\mathbf{B})$. We will first prove prove (3.33). Define a set

$$\mathcal{B}(r, \rho|\mathbf{X}) = \left\{\mathbf{B} := (B_n : n \in \mathbb{N}) : (1+\rho) \liminf_{n\to\infty} \frac{1}{n} \log \sum_{x^n \in B_n} P_{X^n}^{\frac{1}{1+\rho}}(x^n) \geq r\right\}. \tag{3.68}$$

Then, by definition,

$$R^*(r, \rho|\mathbf{X}) = \inf\left\{|\mathbf{B}| : \mathbf{B} \in \mathcal{B}(r, \rho|\mathbf{X})\right\}. \tag{3.69}$$

Fix a $\mathbf{B} \in \mathcal{B}(r, \rho|\mathbf{X})$, Proposition 3.3.1 then implies

$$(1+\rho) \liminf_{n\to\infty} \frac{1}{n} \log \sum_{x^n \in B_n} P_{X^n}^{\frac{1}{1+\rho}}(x^n) = \max_{\mathbf{Y}:\mathbf{B}\in S(\mathbf{Y})} E_l(\mathbf{Y}, \mathbf{X}, \rho).$$

We can therefore conclude using (3.68) that the following set equivalence holds

$$\mathcal{B}(r,\rho|\mathbf{X}) = \bigcup_{E_l(\mathbf{Y},\mathbf{X},\rho)\geq r} S(\mathbf{Y}). \tag{3.70}$$

From (3.69) and (3.70) we get

$$\begin{aligned} R^*(r,\rho|\mathbf{X}) &= \inf\left\{|\mathbf{B}| : \mathbf{B} \in \bigcup_{E_l(\mathbf{Y},\mathbf{X},\rho)\geq r} S(\mathbf{Y})\right\} \\ &= \inf_{\mathbf{Y}}\{|\mathbf{B}| : E_l(\mathbf{Y},\mathbf{X},\rho) \geq r, \mathbf{B} \in S(\mathbf{Y})\} \\ &= \inf_{\mathbf{Y}:E_l(\mathbf{Y},\mathbf{X},\rho)\geq r} \overline{H}(\mathbf{Y}), \end{aligned}$$

where last equality follows because

$$\overline{H}(\mathbf{Y}) = \inf\{|\mathbf{B}| : \mathbf{B} \in S(\mathbf{Y})\}$$

as proved by Han & Verdú [31]. This proves (3.33).

We now prove (3.34). We first show that if R is (r,ρ)-admissible then

$$r \leq \sup_{\overline{H}(\mathbf{Y})\leq R} E_l(\mathbf{Y},\mathbf{X},\rho)$$

. Since R is (r,ρ)-admissible, definition of $R^*(r,\rho|\mathbf{X})$ and (3.33) imply

$$R \geq R^*(r,\rho|\mathbf{X}) = \inf_{\mathbf{Y}:E_l(\mathbf{Y},\mathbf{X},\rho)\geq r} \overline{H}(\mathbf{Y}),$$

i.e., for all $\delta > 0$ there exists a $\hat{\mathbf{Y}}$ such that

$$E_l(\hat{\mathbf{Y}},\mathbf{X},\rho) \geq r \quad \text{and} \quad \overline{H}(\hat{\mathbf{Y}}) < R+\delta,$$

which further implies that

$$r \leq \sup_{\overline{H}(\mathbf{Y})<R+\delta} E_l(\mathbf{Y},\mathbf{X},\rho).$$

Since δ was arbitrary, letting $\delta \downarrow 0$ yields

$$r \leq \sup_{\overline{H}(\mathbf{Y}) \leq R} E_l(\mathbf{Y}, \mathbf{X}, \rho),$$

and the converse part is proved.

For the direct part it is sufficient to show that given ρ, any R with

$$r := \sup_{\overline{H}(\mathbf{Y}) \leq R} E_l(\mathbf{Y}, \mathbf{X}, \rho),$$

is (r, ρ)-admissible. By choice of r, for all $\delta > 0$, there exists a $\hat{\mathbf{Y}}$ such that

$$E_l(\hat{\mathbf{Y}}, \mathbf{X}, \rho) > r - \delta \quad \text{and} \quad \overline{H}(\hat{\mathbf{Y}}) \leq R.$$

This implies that

$$\inf_{E_l(\mathbf{Y}, \mathbf{X}, \rho) > r - \delta} \overline{H}(\mathbf{Y}) \leq R.$$

Since δ was arbitrary, let $\delta \downarrow 0$ and use (3.33) to get

$$R \geq \inf_{E_l(\mathbf{Y}, \mathbf{X}, \rho) \geq r} \overline{H}(\mathbf{Y}) = R^*(r, \rho | \mathbf{X}),$$

i.e., is (r, ρ)-admissible. This completes the proof.

3.6 Summary

In this chapter we saw the close connection between the problem of guessing a source realization given a crytogram and the problem of compression with saturated exponential costs. The latter is a modification of a problem posed by Campbell [6]. The exponents for the two problems coincide. We used the information spectrum approach to obtain upper and lower bounds on the guessing exponents. We related the terms in the lower bound to the error exponent and a generalization of correct decoding exponent for fixed length block source codes. We then evaluated these bounds for stationary memoryless, Markov,

unifilar sources, and showed that in these cases the upper and lower bounds are tight.

Chapter 4

Conclusion

In this thesis we analyzed the strength of Shannon cipher systems using guessing exponents as performance metrics. These exponents captured the effort needed by an attacker who employs exhaustive guessing attacks. We first considered the case when key rate was large, i.e., perfect secrecy, and related the problem of finding guessing exponents to one of the compression with exponential costs, a problem introduced and solved by Campbell [6]. We then analyzed this source coding problem using large deviations theory and gave a sufficient condition for the existence of the limiting guessing exponent: the sequence of distributions of the information spectrum should satisfy a large deviation property. We also gave several examples to illustrate the recipe to evaluate the limiting guessing exponent when the sufficiency condition holds.

For the key rate constrained cryptosystem we related the guessing exponent to the exponent of a modified Campbell compression problem with saturated exponentiated costs. We then found upper and lower bounds on the exponents for general sources using the information spectrum approach. These bounds were given in terms of source coding error exponents and correct decoding exponents (with exponentiated probabilities). The bounds were shown to be tight for DMS, Markov, unifilar sources which recovered previously known results.

In both the perfect secrecy and key rate constrained cases, our approach was to relate the problem of guessing to one of compression with exponentiated costs. The information

spectrum played a key role in our study, and tools from large deviations theory readily yielded either the exponents themselves or bounds on these exponents (in the case of the key rate constrained cryptosystem).

We end this thesis with a compilation of some open questions.

- Is there a weaker sufficient condition for the limiting Rényi entropy rate to exist? In other words, can the sufficient condition of chapter 2 that the sequence of distributions of the information spectrum satisfy the LDP be relaxed?

- We showed that the upper and lower bounds on $E(R, \rho)$ in chapter 3 are tight in some special cases (iid, Markov, unifilar sources). But they may not be tight in general. Is there an example where the bounds are not tight? If so, is there a more satisfactory expression for $E(R, \rho)$?

- Suppose that a source is compressed using the Lempel-Ziv compression algorithm [32]. Suppose further that this compressed sequence of bits are transmitted over a constant bit rate channel that has a buffer of finite size. Probability of buffer overflows [18] may be of interest. If the normalized Lempel-Ziv coding lengths satisfy the LDP such probabilities can be evaluated using large deviation theory. What is the most general subset within the set of stationary ergodic sources for which the normalized Lempel-Ziv coding lengths satisfy the LDP?

- How do the results in chapter 3 extend to the case when the receiver is provided with additional side information? Such a problem may be of interest when the attacker has additional information correlated with the source. For example, the time at which the message was sent may yield side information about the message itself. Alternatively, the amount of energy expended by the cryptosystem during its encryption operation may yield critical information about the message. It is interesting to note that Arikan [4] analyzed the perfect secrecy case to lower bound the search effort of sequential decoders given the received signal. The same technique also yields a lower bound to the search effort of a sphere decoder for space-time codes [33]. However, extensions to key rate constrained cryptosystems remain open.

- Recall the work on Hayashi & Yamamoto [10] which considered a cryptosystem encrypting a correlated source (X^n, Z^n) where the receiver was interested in Z^n and the wiretapper in X^n. Can their coding theorems be extended to general sources? How are their answers related to fixed rate source coding exponents.

- If guessing to within a distortion is allowed, can the result of Merhav & Arikan [5] be extended to general sources? Both cases of perfect secrecy and key-rate constrained secrecy remain open. This problem and its connection to compression are of interest in search applications, where a search engine may return all matches that are within a certain proximity of a guess.

Bibliography

[1] C. E. Shannon, "Communication theory of secrecy systems," *Bell Syst. Tech. J.*, vol. 28, no. 3, pp. 565–715, Oct. 1949.

[2] N. Merhav and E. Arikan, "The Shannon cipher system with a guessing wiretapper," *IEEE Trans. Inf. Theory*, vol. 45, no. 6, pp. 1860–1866, Sep. 1999.

[3] J. L. Massey, "Guessing and entropy," in *Proc. 1994 IEEE International Symposium on Information Theory*, Trondheim, Norway, Jun. 1994, p. 204.

[4] E. Arikan, "An inequality on guessing and its application to sequential decoding," *IEEE Trans. Inform. Theory*, vol. IT-42, pp. 99–105, Jan. 1996.

[5] E. Arikan and N. Merhav, "Guessing subject to distortion," *IEEE Trans. Inform. Theory*, vol. IT-44, pp. 1041–1056, May 1998.

[6] L. L. Campbell, "A coding theorem and Rényi's entropy," *Information and Control*, vol. 8, pp. 423–429, 1965.

[7] D. Mallone and W. G. Sullivan, "Guesswork and entropy," *IEEE Trans. Inf. Theory*, vol. 50, no. 4, pp. 525–526, Mar. 2004.

[8] E. Pfister and W. G. Sullivan, "Rényi entropy, guesswork moments, and large deviations," *IEEE Trans. Inf. Theory*, vol. 50, no. 11, pp. 2794–2800, Nov. 2004.

[9] R. Sundaresan, "Guessing based on length functions," in *Proceedings of the Conference on Managing Complexity in a Distributed World, MCDES*, Bangalore, India, May 2008; *also available as* DRDO-IISc Programme in Mathematical Engineering

Technical Report No. TR-PME-2007-02, Feb. 2007.
http://pal.ece.iisc.ernet.in/PAM/tech_rep07/TR-PME-2007-02.pdf.

[10] Y. Hayashi and H. Yamamoto, "Coding theorems for the Shannon cipher system with a guessing wiretapper and correlated source outputs," *IEEE Trans. Inf. Theory*, vol. 54, no. 6, pp. 2808–2817, Jun 2008.

[11] T. S. Han, *Information-Spectrum Methods in Information Theory.* New York: Springer-Verlog, 2003.

[12] ——, "The reliability functions of the general source with fixed-length coding," *IEEE Trans. Inf. Theory*, vol. 46, no. 6, pp. 2117–2132, Sep 2000.

[13] K. Iriyama, "Probability of error for the fixed-length source coding of general sources," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1537–1543, May 2001.

[14] M. K. Hanawal and R. Sundaresan, "Guessing revisited: A large deviations approach," in *Proc. National Conference on Communications*, Guwahati, India, Jan 2009.

[15] ——, "Guessing revisited: A large deviations approach," *DRDO-IISc Programme in Mathematical Engineering Technical Report No. TR-PME-2008-08, Dec.*, 2008, *available at* http://pal.ece.iisc.ernet.in/PAM/tech_rep08/TR-PME-2008-08.pdf.

[16] M. J. Weinberger, J. Ziv, and A. Lempel, "On the optimal asymptotic performance of universal ordering and of discrimination of individual sequences," *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp. 380–385, Mar. 1992.

[17] A.D.Wyner, "An upper bound on the entropy series," *Information and Control*, vol. 20(2), pp. 176–181, Mar. 1972.

[18] N. Merhav, "Universal coding with minimum probability of codeword length overflow," *IEEE Trans. Inf. Theory*, vol. 37, no. 3, pp. 556 – 563, May 1991.

[19] R. Sundaresan, "Guessing under source uncertainty," *IEEE Trans. Inf. Theory*, vol. 53, no. 1, pp. 269–287, Jan. 2007.

[20] R. S. Ellis, *Entropy, Large Deviations, and Statistical Mechanics*. New York: Springer-Verlag, 1985.

[21] A. Dembo and O. Zeitouni, *Large Deviation Techniques and Applications*, 2nd ed. New York: Springer-Verlag, 1998.

[22] R. S. Ellis, "The theory of large deviations and applications to statistical mechanics," Oct. 2006, Lectures for the International Seminar on Extreme Events in Complex Dynamics, Dresden, Germany.

[23] S. Natarajan, "Large deviations, hypotheses testing, and source coding for finite Markov chains," *IEEE Trans. Inf. Theory*, vol. 31, no. 3, pp. 360–365, May 1985.

[24] E. Seneta, *Non-negative Matrices: An Introduction to Theory and Applications*. London: George Allen & Unwin Ltd., 1973.

[25] F. den Hollander, *Large Deviations*. Rhode Island: American Mathematical Society, 2003.

[26] P. Dupuis and R.S.Ellis, *A Weak Convergence Approach to the Theory of Large Deviations*. New York: John Wiley & Sons, 1997.

[27] I. Joó and L. L. Stachó, "A note on Ky Fan's minimax theorem," *Acta Math. Acad. Sci. Hungar.*, vol. 39, pp. 401–407, 1982.

[28] S. R. S. Varadhan, "Asymptotic probabilities and differential equations," *Comm. Pure Appl. Math.*, vol. 19, pp. 261–286, 1966.

[29] M. K. Hanawal and R. Sundaresan, "The Shannon cipher system with a guessing wiretapper: General sources," *DRDO-IISc Programme in Mathematical Engineering Technical Report No. TR-PME-2009-04, Jan.*, 2009, *available at* http://pal.ece.iisc.ernet.in/PAM/tech_rep04/TR-PME-2009-04.pdf.

[30] R. Ash, *Information Theory*. Interscience Publishers, 1965.

[31] T. S. Han and S. Verdú, "Approximation theory of of output statistics," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 752–772, May 1993.

[32] J. Ziv and A. Lempel, "Compression of individual sequences via variable-rate coding," *IEEE Trans. Inf. Theory*, vol. 24, no. 5, pp. 530–536, Sept. 1978.

[33] E. Agrell, T. Eriksson, E. Vardy, and K. Zeger, "Closest point search in lattices," *IEEE Trans. Inf. Theory*, vol. 48, pp. 2201–2214, Aug. 2002.