

DRDO–IISc Programme on  
Advanced Research in Mathematical  
Engineering

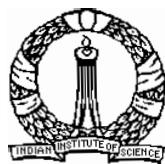
Guessing Revisited: A Large Deviations Approach

(TR-PME-2008-08)  
by

Manjesh Kumar Hanawal and Rajesh Sundaresan

Department of Electrical Communication Engineering, Indian Institute of Science,  
Bangalore

15 December 2008  
revised 24 June 2010



Indian Institute of Science  
Bangalore 560 012

# Guessing Revisited: A Large Deviations Approach

Manjesh Kumar Hanawal and Rajesh Sundaresan

15 December 2008

revised 24 June 2010

## ABSTRACT

The problem of guessing a random string is revisited. A close relation between guessing and compression is first established. Then it is shown that if the sequence of distributions of the information spectrum 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. Other sufficient conditions related to certain continuity properties of the information spectrum are also studied. This approach highlights the importance of the information spectrum in determining the limiting guessing exponent. All known prior results are then re-derived as example applications of our unifying approach. Going beyond the known examples, the existence of the limiting guessing exponent for a source that models a general ferromagnet with summable interactions is also established.

**Keywords:** guessing, length function, source coding, information spectrum, large deviations.

## 1 Introduction

Let  $X^n = (X_1, \dots, X_n)$  denote  $n$  letters of a process where each letter is drawn from a finite set  $\mathbb{X}$  with joint probability mass function (pmf)  $(P_n(x^n) : x^n \in \mathbb{X}^n)$ . 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”. We wish to do this using the minimum expected number of guesses. 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 (Merhav & Arikan [1]).

Massey [2] observed that the expected number of guesses is minimized by guessing in the decreasing order of  $P_n$ -probabilities. Define the *guessing*

function

$$G_n^* : \mathbb{X}^n \rightarrow \{1, 2, \dots, |\mathbb{X}|^n\}$$

to be one such optimal guessing order<sup>1</sup>.  $G_n^*(x^n) = g$  implies that  $x^n$  is the  $g$ th guess. Arikán [3] considered the growth of  $\mathbb{E}[G_n^*(X^n)^\rho]$  as a function of  $n$  for an independent and identically distributed (iid) source with marginal pmf  $P_1$  and  $\rho > 0$ . He showed that the growth is exponential in  $n$ ; the limiting exponent

$$E(\rho) := \lim_{n \rightarrow \infty} \frac{1}{n} \ln \mathbb{E}[G_n^*(X^n)^\rho] \quad (1)$$

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

$$\frac{1}{1 - \alpha} \ln \left( \sum_{x^n \in \mathbb{X}^n} P_n(x^n)^\alpha \right), \quad \alpha \neq 1. \quad (2)$$

Malone & Sullivan [4] 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  $\alpha$ . From their proof, one obtains the more general result that the limiting exponent exists for any source if the Rényi entropy *rate* of order  $\alpha$ ,

$$\lim_{n \rightarrow \infty} n^{-1} H_\alpha(P_n), \quad (3)$$

exists for  $\alpha = 1/(1 + \rho)$ . Pfister & Sullivan [5] showed the existence of (1) for a class of stationary probability measures, beyond Markov measures, that are supported on proper subshifts of  $\mathbb{X}^\mathbb{N}$  [5]. A particular example is that of shifts generated by finite-state machines. For such a class, they showed that the guessing exponent has a variational characterization (see (25) later). For unifilar sources Sundaresan [6] obtained a simplification of this variational characterization using a direct approach and the method of types.

Merhav & Arikán remark that their proof in [7] 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 [8].

Our contribution is to 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 the following.

- If the sequence of distributions of the *information spectrum*

$$(1/n) \ln(1/P_n(X^n))$$

---

<sup>1</sup>If there are several sequences with the same probability of occurrence, they may be guessed in any order without affecting the expected number of guesses.

(see Han [9]) 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 but one previously considered cases in the literature (without side information and key-rate constraints) satisfy this sufficient condition. See Examples 1-5 in section 4. The one exception is an example of Arikan & Merhav [7, Sec. VI-B] which is addressed next.

- If the information spectrum sequence satisfies some continuity conditions, subadditivity arguments or an application of Laplace's method as generalized by Varadhan [10, Th. 3.4] enable us to show that the Rényi entropy rate (3) exists and therefore so does the limiting guessing exponent. The exception indicated above is a source with a certain mixing property (Arikan & Merhav [7, Sec. VI-B]). While this source's information spectrum sequence need not satisfy the large deviation property, it does have a limiting exponent (for guessing and moment generating function of compression lengths). Our argument to show this is more direct (in comparison to [7, Sec. VI-B]) and exploits subadditivity. See Example 6 of section 4. This exception also serves to shed light on the additional properties one obtains from having the large deviation property for the information spectrum.
- We then demonstrate that the Rényi entropy rate, and therefore the limiting exponent, exists for a family of sources represented by any *summable ferromagnetic interaction* on  $\mathbb{Z}$  [11, Ch. IV]. See Example 7 of section 4.

The large deviation theoretic ideas are already present in the works of Pfister & Sullivan [5] and the method of types approach of Arikan & Merhav [7]. Our work however brings out the essential ingredient (the sufficient conditions on the information spectrum), enables us to see the previously obtained specific results under one light, and apply them to a new class of sources representing summable ferromagnetic interactions.

The quest for a general sufficient condition under which the information spectrum satisfies a large deviation property is a natural line of inquiry, and one of independent interest, in view of the Shannon-McMillan-Breiman theorem which asserts that the information spectrum of a stationary and ergodic source converges to the Shannon entropy almost surely and in  $L_q$ , for all  $q \geq 1$ ; see for example [12]. In particular, the large deviation property implies exponentially fast convergence to entropy. In the several specific examples we consider, the information spectrum does satisfy the large deviation property. One sufficient condition for the weaker property of exponentially fast convergence to entropy is the so-called *blowing up property*. (See Marton & Shields [13, Th. 2], or the survey article by Shields [14]). One family of sources, that includes most of the sources we consider in this re-

port and goes beyond, is that of *finitary encodings* of memoryless processes, also called finitary processes. These are known to have the blowing-up property, and therefore exponentially fast convergence to entropy (see Marton & Shields [13, Th. 3]). It is an interesting open question to see if finitary processes, or what other sources with the blowing up property, satisfy the large deviation property.

The rest of the report is organized as follows. Section II studies the tight relationship between guessing and compression. Section III states the relevant large deviations results and the main sufficiency results. Section IV re-derives prior results by showing that in each case the information spectrum satisfies the LDP. The example of Merhav & Arikán [7, Sec. VI-B] and the source modeling ferromagnetic interactions are also addressed. Section V contains proofs and section VI contains some concluding remarks.

## 2 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 and will be returned to shortly. A guessing function

$$G : \mathbb{X} \rightarrow \{1, 2, \dots, |\mathbb{X}|\}$$

is a bijection that denotes the order in which the elements of  $\mathbb{X}$  are guessed. If  $G(x) = g$ , then the  $g$ th guess is  $x$ . Let  $\mathbb{N}$  denote the set of natural numbers. A length function

$$L : \mathbb{X} \rightarrow \mathbb{N}$$

is one that satisfies Kraft's inequality

$$\sum_{x \in \mathbb{X}} \exp_2\{-L(x)\} \leq 1, \quad (4)$$

where we have used the notation  $\exp_2\{-L(x)\} = 2^{-L(x)}$ . To each guessing function  $G$ , we associate a PMF  $Q_G$  on  $\mathbb{X}$  and a length function  $L_G$  as follows.

**Definition 1** *Given a guessing function  $G$ , we say  $Q_G$  defined by*

$$Q_G(x) = c^{-1} \cdot G(x)^{-1}, \quad \forall x \in \mathbb{X}, \quad (5)$$

*is the PMF on  $\mathbb{X}$  associated with  $G$ . The quantity  $c$  in (5) is the normalization constant. We say  $L_G$  defined by*

$$L_G(x) = \lceil -\log_2 Q_G(x) \rceil, \quad \forall x \in \mathbb{X}, \quad (6)$$

*is the length function associated with  $G$ .*

Observe that

$$c = \sum_{a \in \mathbb{X}} G(a)^{-1} = \sum_{i=1}^{|\mathbb{X}|} \frac{1}{i} \leq 1 + \ln |\mathbb{X}|, \quad (7)$$

and therefore the PMF in (5) is well-defined. We record the intimate relationship between these associated quantities in the following result. (This is also available in the proof of [15, Th. 1, p.382]).

**Proposition 1** *Given a guessing function  $G$ , the associated quantities satisfy*

$$c^{-1} \cdot Q_G(x)^{-1} = G(x) \leq Q_G(x)^{-1}, \quad (8)$$

$$L_G(x) - 1 - \log_2 c \leq \log_2 G(x) \leq L_G(x). \quad (9)$$

*Proof:* The first equality in (8) follows from the definition in (5), and the second inequality from the fact that  $c \geq 1$ .

The upper bound in (9) follows from the upper bound in (8) and from (6). The lower bound in (9) follows from

$$\begin{aligned} \log_2 G(x) &= \log_2 (c^{-1} \cdot Q_G(x)^{-1}) \\ &= -\log_2 Q_G(x) - \log_2 c \\ &\geq (\lceil -\log_2 Q_G(x) \rceil - 1) - \log_2 c \\ &= L_G(x) - 1 - \log_2 c. \end{aligned}$$

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

**Definition 2** *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*

$$Q_L(x) = \frac{\exp_2\{-L(x)\}}{\sum_{a \in \mathbb{X}} \exp_2\{-L(a)\}}. \quad (10)$$

**Proposition 2** *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.

$$\log_2 G_L(x) \leq \log_2 Q_L(x)^{-1} \leq L(x). \quad (11)$$

*Proof:* The first statement is clear from the definition of  $G_L$  and from (10).

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

$$\begin{aligned} G_L(x) &\leq \sum_{a \in \mathbb{X}} 1\{Q_L(a) \geq Q_L(x)\} \\ &\leq \sum_{a \in \mathbb{X}} \frac{Q_L(a)}{Q_L(x)} \\ &= Q_L(x)^{-1}, \end{aligned} \tag{12}$$

which proves the left inequality in (11). This inequality was known to Wyner [16].

The last inequality in (11) follows from (10) and Kraft's inequality (4) as follows:

$$Q_L(x)^{-1} = \exp_2\{L(x)\} \cdot \sum_{a \in \mathbb{X}} \exp_2\{-L(a)\} \leq \exp_2\{L(x)\}.$$

■

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 1 and 2.

**Corollary 3** *For a given  $G$ , its associated length function  $L_G$ , and any  $B \geq 1$ , we have*

$$\begin{aligned} &\{L_G(x) \geq B + 1 + \log_2 c\} \\ &\subseteq \{G(x) \geq \exp_2\{B\}\} \\ &\subseteq \{L_G(x) \geq B\}. \end{aligned} \tag{13}$$

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

$$\{G_L(x) \geq \exp_2\{B\}\} \subseteq \{L(x) \geq B\}. \tag{14}$$

The inequalities between the associates in (9) and (11) indicate the direct relationship between guessing moments and Campbell's coding problem [8], and that the Rényi entropies are the optimal growth exponents for guessing moments, as highlighted in the following Proposition.

**Proposition 4** *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  $\mathbb{E}[\exp_2\{\rho L^*(X)\}]$ , 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 (7). Then*

$$\frac{\mathbb{E}[G_L(X)^\rho]}{\mathbb{E}[G^*(X)^\rho]} \leq \frac{\mathbb{E}[\exp_2\{\rho L(X)\}]}{\mathbb{E}[\exp_2\{\rho L^*(X)\}]} \cdot \exp_2\{\rho(1 + \log_2 c)\}. \tag{15}$$

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

$$\frac{\mathbb{E}[G(X)^\rho]}{\mathbb{E}[G^*(X)^\rho]} \geq \frac{\mathbb{E}[\exp_2\{\rho L_G(X)\}]}{\mathbb{E}[\exp_2\{\rho L^*(X)\}]} \cdot \exp_2\{-\rho(1 + \log_2 c)\}. \quad (16)$$

Also,

$$\left| \frac{1}{\rho} \log_2 \mathbb{E}[G^*(X)^\rho] - \frac{1}{\rho} \log_2 \mathbb{E}[\exp_2\{\rho L^*(X)\}] \right| \leq 1 + \log_2 c. \quad (17)$$

*Proof:* Observe that

$$\begin{aligned} & \mathbb{E}[\exp_2\{\rho L(X)\}] \\ & \geq \mathbb{E}[G_L(X)^\rho] \end{aligned} \quad (18)$$

$$\begin{aligned} & \geq \mathbb{E}[G^*(X)^\rho] \\ & \geq \mathbb{E}[\exp_2\{\rho L_{G^*}(X)\}] \exp_2\{-\rho(1 + \log_2 c)\} \end{aligned} \quad (19)$$

$$\geq \mathbb{E}[\exp_2\{\rho L^*(X)\}] \exp_2\{-\rho(1 + \log_2 c)\}, \quad (20)$$

where (18) follows from (11), and (19) from the left inequality in (9). The result in (15) immediately follows. A similar argument shows (16). Finally, (17) follows from the inequalities leading to (20) by setting  $L = L^*$ .  $\blacksquare$

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  $1 + \log_2 c$  of the optimal coding exponent for the length function.

## 2.1 Strings of length $n$

Let us now consider strings of length  $n$ . Let  $\mathbb{X}^n$  denote the set of messages and consider  $n \rightarrow \infty$ . 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 (17) by  $n$ , the difference between two quantities as a function of  $n$  decays as  $O((\log_2 n)/n)$ , and vanishes as  $n$  tends to infinity. The following theorem follows immediately, with a change of base to natural logarithms.

**Theorem 5** *Given  $\rho > 0$ , the limit*

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

*exists if and only if the limit*

$$\liminf_{n \rightarrow \infty} \liminf_{L_n} n^{-1} \ln \mathbb{E}[\exp_2\{\rho L_n(X^n)\}]$$

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

It is therefore sufficient to restrict our attention to the Campbell's coding problem [8] and study if the limit

$$\liminf_{n \rightarrow \infty} \frac{1}{L_n} \ln \mathbb{E}[\exp\{(\rho \ln 2)L_n(X^n)\}] \quad (21)$$

exists, where the infimum is taken over all length functions  $L_n : \mathbb{X}^n \rightarrow \mathbb{N}$  and exponentiation is with respect to the base of the natural logarithm.

## 2.2 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  $\mathbb{E}[\exp\{(\rho \ln 2)L_n^*(X^n)\}]$  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.,

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{1}{n\rho} \ln \mathbb{E}[\exp\{(\rho \ln 2)L_n(X^n)\}] \\ &= \lim_{n \rightarrow \infty} \frac{1}{n\rho} \ln \mathbb{E}[\exp\{(\rho \ln 2)L_n^*(X^n)\}], \end{aligned}$$

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 (15) with  $c_n$  in place of  $c$  followed by the observation  $(1 + \log_2 c_n)/n \rightarrow 0$  shows that the sequence of guessing strategies  $G_{L_n}$  is asymptotically optimal for the class, i.e.,

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{1}{n\rho} \ln \mathbb{E}[G_{L_n}(X^n)^\rho] \\ &= \lim_{n \rightarrow \infty} \frac{1}{n\rho} \ln \mathbb{E}[G^*(X^n)^\rho]. \end{aligned}$$

Arikan and Merhav [7] 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 [17]). 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 (15) 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 [18] characterized this loss as a function of the uncertainty class.

### 3 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

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

and the Rényi entropy of order  $\alpha \neq 1$  is (2). The Kullback-Leibler divergence or relative entropy between two pmfs  $Q_n$  and  $P_n$  is

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

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) and Rényi entropy rate in (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 on the equivalence between guessing and compression, it is sufficient to focus on the Campbell coding problem.

Our first contribution is a proof of the following implicit result of Malone & Sullivan [4]. The proof is given in Section 5.1.

**Proposition 6** *Let  $\rho > 0$ . For a source  $(P_n : n \in \mathbb{N})$ ,  $E(\rho)$  exists if and only if the Rényi entropy rate (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 3 (Large deviation property)** [11, 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  $I : \mathbb{R} \rightarrow [0, \infty]$  if the following conditions hold:*

- *I is lower semicontinuous on  $\mathbb{R}$ ;*
- *I has compact level sets;*
- $\limsup_{n \rightarrow \infty} n^{-1} \ln \nu_n\{K\} \leq -\inf_{t \in K} I(t)$  for each closed subset  $K$  of  $\mathbb{R}$ ;

- $\liminf_{n \rightarrow \infty} n^{-1} \ln \nu_n\{G\} \geq -\inf_{t \in G} I(t)$  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  $\nu_n$  denote the distribution of the information spectrum given by the real-valued random variable  $-n^{-1} \ln P_n(X^n)$ . The following proposition gives a sufficient condition for the existence of the limiting Rényi entropy rate (and therefore the limiting guessing exponent).

**Proposition 7** *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  $I$ . Proposition 7 says that, under the sufficient condition, the limiting guessing exponent equals  $(1 + \rho)I^*(\rho/(1 + \rho))$ , a direct relation to the large deviations rate function for information spectrum. This is however different from Merhav & Arikan's [7, Th. 2] for memoryless sources which states that the limiting guessing exponent is the Legendre-Fenchel dual of the source coding *error exponent* function. We refer the reader to Merhav and Arikan [7, Sec. IV] for further interesting connections between source coding error exponent, guessing exponent, and two other exponents related to lossy source coding.

Let us briefly discuss another approach to verify the existence of Rényi entropy rate (see Proposition 6). With  $\alpha = 1/(1 + \rho)$ , we can rewrite  $1 - \alpha$  times the Rényi entropy rate in (3) as

$$\begin{aligned} & (1 - \alpha) \lim_{n \rightarrow \infty} n^{-1} H_\alpha(P_n) \\ &= \lim_{n \rightarrow \infty} n^{-1} \ln \sum_{x^n \in \mathbb{X}^n} \exp \{-n\alpha F_n(x^n)\} U_n(x^n), \end{aligned} \tag{22}$$

where

$$F_n(x^n) := (-n^{-1} \ln P_n(x^n) - (\ln |\mathbb{X}|)/\alpha),$$

and  $U$  is the iid process on  $X^\mathbb{N}$  with uniform marginal on  $\mathbb{X}$ . One can then view  $\alpha \in (0, 1)$  as the inverse temperature (when  $\rho > 0$ ) of a statistical

mechanical system,  $F_n(x^n)$  as the energy of the configuration  $x^n$ , and the right side of (22) as a scaled version of (i.e.,  $\alpha$  times) the specific Gibbs free energy of the corresponding statistical mechanical system, if the limit exists. This view point is particularly useful because the iid process  $U$  satisfies a sample path large deviation property. If the information spectrum sequence satisfies the continuity conditions in Varadhan [10, Th. 3.4], then the limiting specific Gibbs free energy exists, and so does the Rényi entropy rate. We illustrate a similar approach in Example 7 in section 4.

### 3.1 Additional results from Large Deviations Theory

In order to study the examples in Section 4, we state some additional results on LDP of transformed variables. (See [19, Sec. 4.2]), [20, Th. 6.12 and 6.14]).

**Proposition 8 (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  $\nu_n$  denote the distribution of  $\xi_n$  for  $n \in \mathbb{N}$ , and let the sequence of distributions  $(\nu_n : n \in \mathbb{N})$  on  $\mathcal{X}$  satisfy the LDP with rate function  $I : \mathcal{X} \rightarrow [0, \infty]$ . Let  $\phi : \mathcal{X} \rightarrow \mathbb{R}$  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  $J : \mathbb{R} \rightarrow [0, \infty]$  given by*

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

**Proposition 9 (Exponential Approximation)** *Suppose that the sequence of distributions of  $(\xi_n : n \in \mathbb{N})$  satisfies 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  $\delta > 0$*

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \ln \Pr\{|\xi_n - \zeta_n| > \delta\} = -\infty. \quad (23)$$

*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 (23) is satisfied if*

$$\lim_{n \rightarrow \infty} \sup_{\omega \in \Omega} |\xi_n(\omega) - \zeta_n(\omega)| = 0, \quad (24)$$

*where  $\Omega$  is the underlying sample space.*

## 4 Examples

We are now ready to apply Proposition 7 and related techniques to various examples. In first five 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. In the last two examples, we show that the limiting Rényi entropy rate exists using more general approaches.

#### 4.1 LDP for information spectrum

**Example 1 (An iid source)** *This example was first studied by Arikan [3]. 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*

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

*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  $-\ln P_1(X_1)$  (see for example [11, Th. II.4.1] or [9, eqn. (1.9.66-67)]):*

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

*The Legendre-Fenchel dual of the rate function is therefore the cumulant itself ([11, Th. VI.4.1.e]). An application of Proposition 7 yields that  $(1 + \rho)$  times this cumulant, given by  $\rho H_\alpha(P_1)$ , is the guessing exponent. We thus recover Arikan's result [3].*

The rate function  $I$  can also be obtained using the contraction principle (Proposition 8) 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<sup>2</sup> with rate function  $I_{P_1}^{(2)}(\cdot) = D(\cdot \parallel P_1)$ . See for example [11, 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  $I_{P_1}^{(2)}(\cdot)$  as follows [11, Th II.5.1]. For any  $t$  in  $\mathbb{R}$ , let

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

---

<sup>2</sup>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.

i.e.,

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

Then

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

Using this, we can write

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

thus yielding

$$E(\rho) = \sup_{Q \in \mathcal{M}(\mathbb{X})} \left\{ \rho H(Q) - D(Q \parallel P_1) \right\}. \quad (25)$$

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

**Example 2 (Markov source)** This example was studied by Malone & Sullivan [4]. Consider an irreducible Markov chain taking values on  $\mathbb{X}$  with transition probability matrix  $\pi$ . Our goal is to verify that the sufficient condition holds and to calculate  $E(\rho)$  defined by (1) for this source.

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

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

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  $\eta$  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 [21]

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

As in Example 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

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

By Proposition 6, the limiting guessing exponent exists. Perron-Frobenius theory (Seneta [22, Ch. 1], see also [23, pp.60-61]) yields the cumulant directly as  $\ln \lambda(\beta)$ , where  $\lambda(\beta)$  is unique largest eigenvalue (Perron-Frobenius eigenvalue) of a matrix formed by raising each element of  $\pi$  to the power  $\alpha$ . (Recall that  $\alpha = 1/(1+\rho)$  and  $\beta = \rho/(1+\rho)$ ). Thus  $E(\rho) = (1+\rho) \ln \lambda(\beta)$ , and we recover the result of Malone & Sullivan [4]. It is useful to note that the steps that led to (25) hold in the Markov case (with appropriate changes to entropy and divergence terms) and we may write

$$E(\rho) = \sup_{Q \in \mathcal{M}_s(\mathbb{X}^2)} \left\{ \rho H(\eta | q) - D(\eta \| \pi | q) \right\}, \quad (26)$$

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

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

**Example 3 (Unifilar source)** This example was studied by Sundaresan in [6]. A unifilar source is a generalization of the Markov source in Example 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

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

where  $\pi(x_i, s_i | s_{i-1})$  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  $\pi(x_i, s_i | s_{i-1})$  is such that  $s_i = \phi(s_{i-1}, x_i)$ , where  $\phi$  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  $\phi$  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  $P_{X|S}(x_i | s_{i-1})$  and  $\phi$ . Example 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

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

Then the sequence of distributions of information spectrum  $-n^{-1} \ln P_n(X^n)$  satisfies the LDP ([9, 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 (25) in the iid case and (26) for a Markov source, we get

$$E(\rho) = \sup_{Q \in \mathcal{M}_s(\mathbb{S} \times \mathbb{X})} \left\{ \rho H(\eta \mid q) - D(\eta \parallel \pi \mid q) \right\}, \quad (27)$$

where  $H(\eta \mid q)$  and  $D(\eta \parallel \pi \mid q)$  are analogously defined, and the result of Sundaresan [6] is recovered.

**Example 4 (A class of stationary sources)** Pfister & Sullivan [5] considered a class of stationary sources with distribution  $P \in \mathcal{M}(\mathbb{X}^{\mathbb{N}})$  that satisfies two hypotheses H1 and H2 of [5, Sec. II-B], which we will now describe.

Let  $\mathcal{M}^P(\mathbb{X}^{\mathbb{N}})$  denote the set of sources that satisfy  $Q_n \ll P_n$  for all  $n \in \mathbb{N}$ , where  $P_n$  and  $Q_n$  are restrictions of  $P$  and  $Q$  to  $n$  letters. Note that it may be possible that a  $Q \in \mathcal{M}^P(\mathbb{X}^{\mathbb{N}})$  is not absolutely continuous with respect to  $P$ . Also, let  $\mathcal{M}_s^P(\mathbb{X}^{\mathbb{N}}) \subset \mathcal{M}^P(\mathbb{X}^{\mathbb{N}})$  denote the subset of stationary sources with respect to the shift operator  $\tau : \mathbb{X}^{\mathbb{N}} \rightarrow \mathbb{X}^{\mathbb{N}}$  defined by

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

Hypothesis H1 of Pfister & Sullivan [5] assumes that for any neighborhood of a stationary source  $Q \in \mathcal{M}_s^P(\mathbb{X}^{\mathbb{N}})$  and any  $\varepsilon > 0$ , there exists an ergodic  $Q' \in \mathcal{M}_s^P(\mathbb{X}^{\mathbb{N}})$  in that neighborhood such that  $\overline{H}(Q') \geq \overline{H}(Q) - \varepsilon$ , where  $\overline{H}(Q)$  is the Shannon entropy rate of source  $Q$ . Their hypothesis H2 is given by (30) below.

Under these hypotheses, Pfister & Sullivan [5] proved that  $E(\rho)$  exists, and provided a variational characterization analogous to (27), i.e.,

$$E(\rho) = \sup_{Q \in \mathcal{M}_s^P(\mathbb{X}^{\mathbb{N}})} \left\{ \rho \overline{H}(Q) - \overline{D}(Q \parallel P) \right\}, \quad (28)$$

where

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

En route to this result, Pfister & Sullivan [5] showed 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  $x = (x_1, x_2, \dots)$ , we define  $x^n = (x_1, \dots, x_n)$  as the first  $n$  components of  $x$  in the usual way. Consider a stationary source  $P$  whose letters are  $X = (X_1, X_2, \dots)$ . Define the empirical process of measures

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

This is a measure on  $\mathbb{X}^{\mathbb{N}}$  that puts mass  $1/n$  on the following strings:  $x, \tau(x), \tau^2(x), \dots, \tau^{n-1}(x)$ . Pfister & Sullivan showed that the distributions of the  $\mathcal{M}(\mathbb{X}^{\mathbb{N}})$ -valued process  $T_n(X, \cdot)$  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 ([5, Prop. 2.2-2.3]). Furthermore,

$$\overline{D}(Q \parallel P) = +\infty, \quad Q \notin \mathcal{M}_s^P(\mathbb{X}^{\mathbb{N}}), \quad (29)$$

so that we may restrict  $\overline{D}(\cdot \parallel P)$  to  $\mathcal{M}_s^P(\mathbb{X}^{\mathbb{N}})$ . We next use this to show that the information spectrum satisfies the LDP.

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

$$\lim_{n \rightarrow \infty} \sup_{x \in \Sigma_n^P} \left| n^{-1} \ln P_n(x^n) + \int_{\mathbb{X}^{\mathbb{N}}} e_P \, dT_n(x, \cdot) \right| = 0, \quad (30)$$

where  $\Sigma_n^P = \{x \in \mathbb{X}^{\mathbb{N}} : P_n(x^n) > 0\}$ .

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

$$\phi : \mathcal{M}(\mathbb{X}^{\mathbb{N}}) \rightarrow \mathbb{R}$$

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

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

where the restriction of the infimum to  $\mathcal{M}_s^P(\mathbb{X}^{\mathbb{N}})$  follows from (29). Furthermore, given hypothesis H2 and (30), an application of the exponential approximation principle (Proposition 9) 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  $\overline{D}(Q \parallel P) = \phi(Q) - \overline{H}(Q)$  [5, Prop. 2.1] to write

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

Finally, the Legendre-Fenchel dual of the rate function is computed as in the steps leading to (25)-(27), yielding (28).

**Example 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 7.

The sequence of distributions of the information spectrum satisfies the LDP with rate function given as follows (see Han [9, eqn. (1.9.41)]). Define

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

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

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

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 7 we conclude that the limiting guessing exponent exists.

$I^*(\beta)$  is then

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

yielding

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

## 4.2 Other approaches

Thus far, we considered examples where we could verify the sufficient condition that the information spectrum satisfies the LDP. In the next two examples, we directly address the existence of the Rényi entropy rate.

**Example 6 (Sources with a certain mixing property)** *Arikan & Merhav [7] considered a stationary source with memory satisfying the following property: there is a finite and positive  $B$  such that for all  $m, n$ ,  $u^m \in \mathbb{X}^m$  and  $v^n \in \mathbb{X}^n$ , the following condition holds:*

$$\left| \ln \frac{P_{m+n}((u^m, v^n))}{P_m(u^m)} - \ln P_n(v^n) \right| \leq B,$$

where  $(u^m, v^n)$  is a concatenation of the two strings. This condition is equivalent to

$$-B \leq \ln P_{m+n}((u^m, v^n)) - \ln P_m(u^m)P_n(v^n) \leq B. \quad (31)$$

They considered a more general setting of guessing subject to distortion. Under the assumption in (31), and when specialized to the guessing without distortion setting, they show that the limiting exponent exists and equals

$$\lim_n n^{-1} \max_{Q_n \in \mathcal{M}(\mathbb{X}^n)} [\rho H(Q_n) - D(Q_n \parallel P_n)].$$

We re-derive this result using subadditivity, and further explore the relationship between the sufficient condition (31) and the rather stringent sufficient condition of an LDP for information spectrum.

It is a simple matter to verify, on account of only the upper bound inequality in (31) and stationarity, that

$$\begin{aligned} & \ln \left( \sum_{x^{m+n}} P_{m+n}(x^{m+n})^\alpha \right) \\ & \leq \ln \left( \sum_{x^m} P_m(x^m)^\alpha \right) + \ln \left( \sum_{x^n} P_n(x^n)^\alpha \right) + \alpha B \end{aligned}$$

when  $\alpha \in (0, 1)$ , so that the Rényi entropy of order  $\alpha$  (see (2)) plus a constant is subadditive for  $\alpha \in (0, 1)$ , i.e.,

$$\begin{aligned} & H_\alpha(P_{m+n}) + \frac{\alpha B}{1-\alpha} \\ & \leq \left( H_\alpha(P_m) + \frac{\alpha B}{1-\alpha} \right) + \left( H_\alpha(P_n) + \frac{\alpha B}{1-\alpha} \right). \end{aligned}$$

This subadditivity, the fact that  $H_\alpha(P_n) \leq \ln |\mathbb{X}|^n < \infty$ , and the fact that

$$\lim_n n^{-1} \frac{\alpha B}{1-\alpha} = 0$$

imply that the Rényi entropy rate of order  $\alpha$  exists [19, Lem. 6.1.11], and

$$\begin{aligned} & \lim_{n \rightarrow \infty} n^{-1} H_\alpha(P_n) \\ & = \inf_{n \geq 1} n^{-1} H_\alpha(P_n) \\ & = \inf_{n \geq 1} n^{-1} \sup_{Q_n} \{ \rho H(Q_n) - D(Q_n \parallel P_n) \}, \end{aligned}$$

where the last equality comes from the well-known variational characterization of Rényi entropy with  $\alpha = 1/(1+\rho)$  (see also (44) that comes later). As we have used only the upper bound inequality in (31), this is a modest sharpening of Arikan & Merhav's result.

To understand the more demanding sufficient condition that an LDP for information spectrum exists, we begin with normalized cumulant of the information spectrum:

$$\begin{aligned} c_n(\beta) & := n^{-1} \ln \mathbb{E} [\exp \{-\beta \ln P_n(X^n)\}] \\ & = n^{-1} \ln \left( \sum_{x^n \in \text{Support}(P_n)} P_n(x^n)^{1-\beta} \right). \end{aligned}$$

Clearly,  $c_n(\beta)$  is a finite sum over the support set of  $P_n$  and is therefore finite for all  $\beta \in \mathbb{R}$ . Moreover, on account of both the upper bound and the

lower bound inequalities in (31),  $nc_n(\beta) + |1 - \beta|B$  is subadditive for each  $\beta \in \mathbb{R}$ , and therefore the following limit exists:

$$c(\beta) := \lim_{n \rightarrow \infty} c_n(\beta) = \inf_{n \geq 1} c_n(\beta) < \infty.$$

The natural candidate for the rate function for information spectrum is then the Legendre-Fenchel transform

$$I(t) := \sup_{\beta \in \mathbb{R}} \{\beta t - c(\beta)\}.$$

An application of [11, Th. II.6.1] shows that  $I$  is convex, lower semicontinuous on  $\mathbb{R}$ , nonnegative, has compact level sets, and satisfies  $\inf_{t \in \mathbb{R}} I(t) = 0$ . It also satisfies the upper large deviation bound (see Definition 3). However, for the lower large deviation bound to hold, and therefore for the information spectrum to have an LDP, a sufficient condition is that  $c(\beta)$  be differentiable for all  $\beta \in \mathbb{R}$ , which may not hold.

Even the considerably weaker condition that  $I(t)$  attains its infimum at the unique point  $H$ , the Shannon entropy rate, may not hold. This weaker condition is equivalent to  $c(\beta)$  being differentiable at  $\beta = 0$  with derivative  $H$ . While  $\lim_{n \rightarrow \infty} c'_n(0)$  equals  $H$  for the stationary source, a sufficient condition for the validity of the interchange of differentiation at  $\beta = 0$  and the limit over  $n$  is uniform convergence of the sequence of functions

$$c'_n(\beta) = -n^{-1} \frac{\sum_{x^n \in \text{Support}(P_n)} P_n(x^n)^{1-\beta} \ln P_n(x^n)}{\sum_{a^n \in \text{Support}(P_n)} P_n(a^n)^{1-\beta}}$$

over a closed interval of  $\beta$  that contains 0; even this may not hold. Under the assumptions of this section, differentiability of  $c(\beta)$  at  $\beta = 0$  is also equivalent to exponentially fast convergence of  $n^{-1} \ln P_n(X^n)$  to  $H$ , which is further equivalent to strict convexity of  $c(\beta)$ ; see [11, Th. II.6.3 and Th. VII.2.1]. While we do not yet have explicit counterexamples, we anticipate that these need not hold in general.

**Example 7 (Summable ferromagnetic interaction on  $\mathbb{Z}$ )** With  $\mathbb{X} = \{-1, 1\}$ , we now consider two-sided stationary sources on  $(\mathbb{X}^{\mathbb{Z}}, \mathcal{B}(\mathbb{X}^{\mathbb{Z}}))^3$  that are models for ferromagnetic interactions. The value 1 corresponds to spin in a particular direction and -1 to spin in the opposite direction. As before, we let  $\mathcal{M}(\mathbb{X}^{\mathbb{Z}})$  and  $\mathcal{M}_s(\mathbb{X}^{\mathbb{Z}})$  represent the set of probability measures and the subset of stationary probability measures, respectively, on  $\mathcal{B}(\mathbb{X}^{\mathbb{Z}})$ . Let the interaction strengths be specified by a symmetric function  $J : \mathbb{Z} \rightarrow \mathbb{R}_+$  which

---

<sup>3</sup>The  $\sigma$ -algebra  $\mathcal{B}(\mathbb{X}^{\mathbb{Z}})$  is the one generated by cylinder sets of  $\mathbb{X}^{\mathbb{Z}}$ .

satisfies  $J(i - j) = J(j - i)$  for each  $i$  and  $j$  in  $\mathbb{Z}$ . We shall assume that  $J$  is summable:

$$\sum_{k \in \mathbb{Z}} J(k) < \infty.$$

Let

$$\Lambda_n = \{-n, -n+1, \dots, -1, 0, 1, \dots, n-1, n\}.$$

Fix a boundary configuration  $\tilde{x} \in \mathbb{X}^{\Lambda_n^c}$  outside  $\Lambda_n$ . Two special boundary conditions denoted  $\tilde{0}$  and  $\tilde{1}$  shall be made up of all zeros or all ones, respectively, on  $\mathbb{X}^{\Lambda_n^c}$ . Let us define the Hamiltonian or interaction energy of a configuration  $x \in \Lambda_n$  with boundary condition  $\tilde{x}$  as

$$\begin{aligned} H_{n, \tilde{x}}(x) = & -\frac{1}{2} \sum_{i, j \in \Lambda_n} J(i - j) x_i x_j \\ & - \sum_{i \in \Lambda_n} \left( \sum_{j \in \Lambda_n^c} J(i - j) \tilde{x}_j \right) x_i. \end{aligned}$$

Let us also define the interaction energy ignoring the boundary effect by

$$\overline{H}_n(x) = -\frac{1}{2} \sum_{i, j \in \Lambda_n} J(i - j) x_i x_j.$$

Fix a temperature  $T > 0$ . Our focus shall be on infinite volume Gibbs states, i.e., stationary probability measures  $P^T$  on  $(\mathbb{X}^{\mathbb{Z}}, \mathcal{B}(\mathbb{X}^{\mathbb{Z}}))$  at temperature  $T$ . These are such that the conditional probability of  $x \in \mathbb{X}^{\Lambda_n}$  given a boundary condition  $\tilde{x} \in \mathbb{X}^{\Lambda_n^c}$  satisfies

$$P^T \left\{ X = x \mid \tilde{X} = \tilde{x} \text{ on } \Lambda_n^c \right\} = \frac{\exp\{-(1/T)H_{n, \tilde{x}}(x)\}}{Z(n, T, \tilde{x})} \quad (32)$$

where  $Z(n, T, \tilde{x})$  is the normalization constant or partition function:

$$Z(n, T, \tilde{x}) = \sum_{x \in \mathbb{X}^{\Lambda_n}} \exp\{-(1/T)H_{n, \tilde{x}}(x)\}. \quad (33)$$

For the validity of (32) for a given  $P^T$ , we refer the reader to Ellis [11, Ch. IV] or Liggett [24, Ch. IV]. There may be several different  $P^T$  having the same conditional distribution given a boundary condition, i.e., the right-hand side of (32). This phenomenon of multiplicity (first-order phase transition) will not concern us because we shall work with a given  $P^T$ . Let  $P_{\Lambda_n}^T$  be the restriction of  $P^T$  to the set  $\Lambda_n$ ; this plays the role of  $P_n$  in earlier examples. We shall use the notation

$$P_{\Lambda_n}^T(x \mid \tilde{x}) = P^T \left\{ X = x \mid \tilde{X} = \tilde{x} \text{ on } \Lambda_n^c \right\}$$

for the conditional measures given the boundary condition. Clearly,  $P_{\Lambda_n}^T(\cdot)$  is a convex combination of the conditional measures

$$\left\{ P^T \left\{ X = \cdot \mid \tilde{X} = \tilde{x} \text{ on } \Lambda_n^c \right\} : \tilde{x} \in \mathbb{X}^{\Lambda_n^c} \right\}.$$

We will also have occasion to use the probability measure on the finite set  $\Lambda_n$  defined by

$$\overline{P}_{\Lambda_n}^T(x) = \frac{\exp\{-(1/T)\overline{H}_n(x)\}}{\overline{Z}(n, T)},$$

where  $\overline{Z}(n, T)$  is the corresponding normalization constant or partition function.

By stationarity, existence of the Rényi entropy rate for  $P^T$  is equivalent to existence of the limit

$$\lim_{n \rightarrow \infty} |\Lambda_n|^{-1} \ln \sum_{x \in \mathbb{X}^{\Lambda_n}} (P_{\Lambda_n}^T(x))^\alpha. \quad (34)$$

Liggett [24, Cor. 1.30] shows that there is a certain stochastic ordering among  $P_{\Lambda_n}^T(\cdot \mid \tilde{0})$ ,  $P_{\Lambda_n}^T(\cdot)$ , and  $P_{\Lambda_n}^T(\cdot \mid \tilde{1})$ . In particular, we have

$$P_{\Lambda_n}^T(x \mid \tilde{0}) \leq P_{\Lambda_n}^T(x) \leq P_{\Lambda_n}^T(x \mid \tilde{1}) \quad \forall n \geq 0, \quad \forall x \in \mathbb{X}^{\Lambda_n},$$

and therefore the existence of the limit in (34) follows from the claim that

$$\lim_{n \rightarrow \infty} |\Lambda_n|^{-1} \ln \sum_{x \in \mathbb{X}^{\Lambda_n}} (P_{\Lambda_n}^T(x \mid \tilde{x}))^\alpha \quad (35)$$

exists and is independent of the choice of  $\tilde{x}$ .

To establish this last claim and provide a variational characterization of the value in (35), observe that

$$\begin{aligned} & \lim_{n \rightarrow \infty} |\Lambda_n|^{-1} \ln \sum_{x \in \mathbb{X}^{\Lambda_n}} (P_{\Lambda_n}^T(x \mid \tilde{x}))^\alpha \\ &= \lim_{n \rightarrow \infty} |\Lambda_n|^{-1} \ln \sum_{x \in \mathbb{X}^{\Lambda_n}} \frac{\exp\{-(\alpha/T)H_{n,\tilde{x}}\}}{Z(n, T, \tilde{x})^\alpha} \end{aligned} \quad (36)$$

$$= \lim_{n \rightarrow \infty} |\Lambda_n|^{-1} [\ln Z(n, T/\alpha, \tilde{x}) - \alpha \ln Z(n, T, \tilde{x})] \quad (37)$$

$$= \lim_{n \rightarrow \infty} |\Lambda_n|^{-1} [\ln \overline{Z}(n, T/\alpha) - \alpha \ln \overline{Z}(n, T)] \quad (38)$$

$$\begin{aligned} &= \sup_{Q \in \mathcal{M}_s(\mathbb{X}^{\mathbb{Z}})} \left\{ \frac{\alpha}{2T} \sum_k J(k) \mathbb{E}_Q[X_0 X_k] - \overline{D}(Q \parallel U) \right\} \\ &\quad - \alpha \sup_{Q \in \mathcal{M}_s(\mathbb{X}^{\mathbb{Z}})} \left\{ \frac{1}{2T} \sum_k J(k) \mathbb{E}_Q[X_0 X_k] - \overline{D}(Q \parallel U) \right\}, \end{aligned} \quad (39)$$

where (36) follows after substitution of (32) in (35), (37) follows from the definition in (33), (38) follows from an application of [11, Lem. IV.6.2] to each term in the expression thus yielding a limit that does not depend on  $\tilde{x}$ , and finally, (39) follows from the variational characterizations [11, Th. IV.7.3(a)] of limiting specific Gibbs free energy at two temperatures  $T/\alpha$  and  $T$ . The quantity  $\overline{D}(\cdot \parallel U)$  is the level-3 large deviation rate function for the iid process  $U$ . Both suprema in (39) are attained at stationary Gibbs states of the corresponding temperatures. The limiting guessing exponent is then  $(1+\rho)$  times the right-hand side of (39). This proves the claim that the limit in (35) exists, and thus the Rényi entropy rate exists.

To contrast the approach of this last example with those of earlier examples, we remark that we did not take the route of showing that the information spectrum satisfies a large deviation property. But instead, we exploited the fact that the iid process  $U$  satisfies a level-3 large deviation property and that the information spectrum satisfies certain continuity conditions, a fact that is exploited in the proof of [11, Th. IV.7.3(a)]. Varadhan's [10, Th. 3.4] provides a more general sufficient condition that can be applied to establish existence of the Rényi entropy rate for other sources.

## 5 Proofs

We now prove Propositions 6 and 7.

### 5.1 Proof of Proposition 6

From Theorem 5 it is sufficient to show that the limit in (21) for Campbell's coding problem exists if and only if the Rényi entropy rate exists, with the former  $\rho$  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  $\rho$ , the cumulant. The cumulant associated with a bounded continuous function (here  $L_n$ ) has a variational characterization [25, Prop. 1.4.2] as the following Legendre-Fenchel dual of the Kullback-Leibler divergence, i.e.,

$$\begin{aligned} & \ln \mathbb{E}_{P_n} \left[ \exp\{(\rho \ln 2)L_n(X^n)\} \right] \\ &= \sup_{Q_n \in \mathcal{M}(\mathbb{X}^n)} \left\{ (\rho \ln 2)\mathbb{E}_{Q_n}[L_n(X^n)] - D(Q_n \parallel P_n) \right\}. \end{aligned} \quad (40)$$

Taking infimum on both sides over all length functions, we arrive at the

following chain of inequalities:

$$\inf_{L_n} \ln \mathbb{E}_{P_n} \left[ \exp\{(\rho \ln 2)L_n(X^n)\} \right] \quad (41)$$

$$= \inf_{L_n} \sup_{Q_n \in \mathcal{M}(\mathbb{X}^n)} \left\{ \mathbb{E}_{Q_n}[(\rho \ln 2)L_n(X^n)] - D(Q_n \parallel P_n) \right\}$$

$$= \sup_{Q_n \in \mathcal{M}(\mathbb{X}^n)} \inf_{L_n} \left\{ \mathbb{E}_{Q_n}[(\rho \ln 2)L_n(X^n)] - D(Q_n \parallel P_n) \right\}$$

$$+ \Theta(1) \quad (42)$$

$$= \sup_{Q_n \in \mathcal{M}(\mathbb{X}^n)} \left\{ \rho H_n(Q_n) - D(Q_n \parallel P_n) \right\} + \Theta(1) \quad (43)$$

$$= \rho H_{\frac{1}{1+\rho}}(P_n) + \Theta(1). \quad (44)$$

Equation (42) follows because (i) the mapping

$$(L_n, Q_n) \mapsto \mathbb{E}_{Q_n}[(\rho \ln 2)L_n(X^n)] - D(Q_n \parallel P_n)$$

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

$$L_n = \left\lceil \lambda L_n^{(1)} + (1 - \lambda) L_n^{(2)} \right\rceil$$

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 [26]. 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  $\ln 2$  nats, and (43) follows. Lastly, (44) is a well-known identity which may also be obtained directly by writing the supremum term in (43) as

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

and then applying (40) with  $-(\rho/(1+\rho) \ln P_n(X^n))$  in place of  $(\rho \ln 2)L_n(X^n)$  to get the scaled Rényi entropy.

Normalize both (41) and (44) by  $n$  and let  $n \rightarrow \infty$  to deduce that (21) exists if and only if the limiting normalized Rényi entropy rate exists. This concludes the proof.

## 5.2 Proof of Proposition 7

This is a straightforward application of Varadhan's theorem [10] on asymptotics of integrals. Recall that  $\nu_n$  is the distribution of the information spectrum  $n^{-1} \ln P_n(X^n)$ . 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 [11, Th. II.7.1.b]) states that if

$$\lim_{M \rightarrow \infty} \limsup_{n \rightarrow \infty} \frac{1}{n} \ln \int_{t \geq \frac{M}{\beta}} \exp\{n\beta t\} d\nu_n(t) = -\infty \quad (45)$$

then the limit

$$\lim_{n \rightarrow \infty} \frac{1}{n} \ln \int_{\mathbb{R}} \exp\{n\beta t\} \nu_n(dt) = \sup_{t \in \mathbb{R}} \{\beta t - I(t)\} \quad (46)$$

holds. The integral on the left side in (46) can be simplified by defining the finite cardinality set

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

and by observing that

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

Take logarithms, normalize by  $n$ , take limits, and apply (46) to get the desired result. It therefore remains to prove (45).

The event  $\{t \geq \frac{M}{\beta}\}$  occurs if and only if

$$\left\{ P_n(x^n) \leq \exp\left\{\frac{-nM}{\beta}\right\} \right\}.$$

The integral in (45) can therefore be written as

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

The sequence in  $n$  on the left side of (45) is then

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

a constant sequence. Take the limit as  $M \rightarrow \infty$  to verify (45). This concludes the proof.

## 6 Conclusion

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. We also looked at other conditions on the information spectrum to establish existence of the limiting exponent and applied it to a class of sources modeling a general ferromagnet with summable interactions. While some of these facts can be gleaned from the works of Pfister & Sullivan [5] and Merhav & Arikan [7], our work sheds light on the key role played by the information spectrum. It will be of interest to find a rich class of sources beyond those listed in this report for which the information spectrum satisfies the LDP.

Results on guessing with key-rate constraints for a general source are provided using the above information spectrum approach in [27].

## Acknowledgements

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).

## References

- [1] 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.
- [2] J. L. Massey, “Guessing and entropy,” in *Proc. 1994 IEEE International Symposium on Information Theory*, Trondheim, Norway, Jun. 1994, p. 204.
- [3] E. Arikan, “An inequality on guessing and its application to sequential decoding,” *IEEE Trans. Inf. Theory*, vol. 42, pp. 99–105, Jan. 1996.
- [4] D. Malone and W. G. Sullivan, “Guesswork and entropy,” *IEEE Trans. Inf. Theory*, vol. 50, no. 4, pp. 525–526, Mar. 2004.

- [5] 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.
- [6] 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](http://pal.ece.iisc.ernet.in/PAM/tech_rep07/TR-PME-2007-02.pdf).
- [7] E. Arikan and N. Merhav, “Guessing subject to distortion,” *IEEE Trans. Inf. Theory*, vol. 44, pp. 1041–1056, May 1998.
- [8] L. L. Campbell, “A coding theorem and Rényi’s entropy,” *Information and Control*, vol. 8, pp. 423–429, 1965.
- [9] T. S. Han, *Information-Spectrum Methods in Information Theory*. Springer-Verlag, 2003.
- [10] S. R. S. Varadhan, “Asymptotic probabilities and differential equations,” *Comm. Pure Appl. Math.*, vol. 19, pp. 261–286, 1966.
- [11] R. S. Ellis, *Entropy, Large Deviations, and Statistical Mechanics*, ser. Grundlehren der mathematischen Wissenschaften. New York: Springer-Verlag, 1985, vol. 271.
- [12] K. R. Parthasarathy, *Coding Theorems of Classical and Quantum Information Theory*. TRIM no. 45, Hindustan Book Agency, 2007.
- [13] K. Marton and P. C. Shields, “The positive-divergence and blowing-up properties,” *Israel J. Math.*, vol. 86, pp. 331–348, 1994.
- [14] P. C. Shields, “The interactions between ergodic theory and information theory,” *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2079–2093, Oct. 1998.
- [15] 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.
- [16] A.D.Wyner, “An upper bound on the entropy series,” *Information and Control*, vol. 20(2), pp. 176–181, Mar. 1972.
- [17] N. Merhav, “Universal coding with minimum probability of codeword length overflow,” *IEEE Trans. Inf. Theory*, vol. 37, no. 3, pp. 556 – 563, May 1991.
- [18] R. Sundaresan, “Guessing under source uncertainty,” *IEEE Trans. Inf. Theory*, vol. 53, no. 1, pp. 269–287, Jan. 2007.
- [19] A. Dembo and O. Zeitouni, *Large Deviation Techniques and Applications*, 2nd ed. New York: Springer-Verlag, 1998.
- [20] 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.
- [21] 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.
- [22] E. Seneta, *Non-negative Matrices: An Introduction to Theory and Applications*. London: George Allen & Unwin Ltd., 1973.
- [23] F. den Hollander, *Large Deviations*. Rhode Island: American Mathematical Society, 2003.
- [24] T. M. Liggett, *Interacting Particle Systems*, ser. Grundlehren der mathematischen Wissenschaften. New York: Springer-Verlag, 1985, vol. 276.
- [25] P. Dupuis and R.S.Ellis, *A Weak Convergence Approach to the Theory of Large Deviations*. New York: John Wiley & Sons, 1997.
- [26] 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.
- [27] 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\\_rep09/TR-PME-2009-04.pdf](http://pal.ece.iisc.ernet.in/PAM/tech_rep09/TR-PME-2009-04.pdf).