Guessing Revisited: A Large Deviations Approach

Manjesh Kumar Hanawal and Rajesh Sundaresan, *Senior Member, IEEE*

***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 briefly discussed. 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.**

***Index Terms*—Guessing, information spectrum, large deviations, length function, source coding.**

I. Introduction

LET $X^n = (X_1, \ldots, X_n)$ 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 and 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*

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

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

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

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

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

Malone and 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 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

$$\lim_{n\to\infty} n^{-1} H_\alpha(P_n) \tag{3}$$

exists for $\alpha = 1/(1+\rho)$. Pfister and 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 and Arikan 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 that if the sequence of distributions of the information spectrum $(1/n) \ln (1/P_n(X^n))$ (see Han [9]) satisfies the large deviation property, then the limiting exponent exists. This is useful because several existing large deviations results can be

[1]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.

Manuscript received December 15, 2008; revised June 24, 2010; accepted July 18, 2010. Date of current version December 27, 2010. 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 by Grant Part (2B) UGC-CAS-(Ph.IV). The material in this paper was presented at the IEEE International Symposium on Information Theory (ISIT 2007), Nice, France, June 2007, and at the IISc Centenary Conference on Managing Complexity in a Distributed World, (MCDES 2008), Bangalore, India, May 2008, and at the National Conference on Communications (NCC 2009), Guwahati, India, January 2009.

M. K. Hanawal is with the INRIA and the University of Avignon, Laboratoire Informatique d'Avignon, Avignon cedex 9 84911, France (e-mail: manjesh.k@gmail.com).

R. Sundaresan is with the ECE Department, Indian Institute of Science, Bangalore 560012, India (e-mail: rajeshs@ece.iisc.ernet.in).

Communicated by H. Yamamoto, Associate Editor for Shannon Theory.

Digital Object Identifier 10.1109/TIT.2010.2090221

readily applied. We then show that all but one previously considered cases in the literature[2] satisfy this sufficient condition. See Examples 1–5 in Section IV.

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

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 [11]. 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 and Shields [12, Th. 2], or the survey article by Shields [13]). One family of sources, that includes most of the sources we consider in this paper 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 and Shields [12, 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 paper 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 rederives prior results by showing that in each case the information spectrum satisfies the LDP. Section V contains proofs and Section VI contains some concluding remarks.

II. 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

$$G : \mathbb{X} \to \{1, 2, \ldots, |\mathbb{X}|\}$$

is a bijection that denotes the order in which the elements of $\mathbb{X}$ are guessed. If $G(x) = g$, then the gth guess is x. 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

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

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

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

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

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

is the length function associated with G.

Observe that

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

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 [14, Th. 1, p. 382].)

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

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

$$L_G(x) - 1 - \log_2 c \leq \log_2 G(x) \leq L_G(x). \tag{9}$$

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

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

[2]These are cases without side information and key-rate constraints. The one exception is an example of Arikan and Merhav [7, Sec. VI-B] for which one can show the existence of Rényi entropy rate rather directly via a subadditivity argument. See our technical report [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)
$$\log_2 G_L(x) \le \log_2 Q_L(x)^{-1} \le L(x). \tag{11}$$

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

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

which proves the left inequality in (11). This inequality was known to Wyner [15].

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

$$\begin{aligned} Q_L(x)^{-1} &= \exp_2\{L(x)\} \cdot \sum_{a\in\mathbb{X}} \exp_2\{-L(a)\} \\ &\le \exp_2\{L(x)\}. \end{aligned}$$

■

Let $\{L(x) \ge B\}$ denote the set $\{x \in \mathbb{X} \mid L(x) \ge B\}$. 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 $B \ge 1$, we have

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

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

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

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

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

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

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

Also

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

Proof: Observe that

$$\begin{aligned} &\mathbb{E}\left[\exp_2\{\rho L(X)\}\right] \\ &\quad \ge \mathbb{E}\left[G_L(X)^\rho\right] && (18) \\ &\quad \ge \mathbb{E}\left[G^*(X)^\rho\right] \\ &\quad \ge \mathbb{E}\left[\exp_2\{\rho L_{G^*}(X)\}\right] \exp_2\{-\rho(1+\log_2 c)\} && (19) \\ &\quad \ge \mathbb{E}\left[\exp_2\{\rho L^*(X)\}\right] \exp_2\{-\rho(1+\log_2 c)\} && (20) \end{aligned}$$

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 $L = L^*$. ■

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.

A. 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

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

If we normalize both sides of (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

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

exists if and only if the limit

$$\lim_{n\to\infty} \inf_{L_n} n^{-1} \ln \mathbb{E}\left[\exp_2\{\rho L_n(X^n)\}\right]$$

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

$$\lim_{n\to\infty} \inf_{L_n} \frac{1}{n} \ln \mathbb{E}\left[\exp\{(\rho \ln 2) L_n(X^n)\}\right] \tag{21}$$

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

B. 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 $\mathbb{E}\left[\exp\left\{(\rho \ln 2) L_n^*(X^n)\right\}\right]$ 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.

$$\lim_{n\to\infty} \frac{1}{n\rho} \ln \mathbb{E}\left[\exp\left\{(\rho \ln 2) L_n(X^n)\right\}\right] = \lim_{n\to\infty} \frac{1}{n\rho} \ln \mathbb{E}\left[\exp\left\{(\rho \ln 2) L_n^*(X^n)\right\}\right]$$

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

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

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 [16]). It follows as a consequence of the above argument that guessing in the increasing order of description lengths is asymptotically optimal. The left-hand side (LHS) 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 [17] characterized this loss as a function of the uncertainty class.

III. 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

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

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 II 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 and Sullivan [4]. The proof is given in Section V-A.

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): [18, 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;
- $\limsup_{n\to\infty} n^{-1} \ln \nu_n\{K\} \leq -\inf_{t\in K} I(t)$ for each closed subset K of $\mathbb{R}$;
- $\liminf_{n\to\infty} n^{-1} \ln \nu_n\{G\} \geq -\inf_{t\in G} I(t)$ 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 ν_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 $(1+\rho) I^*(\rho/(1+\rho))$, and is thus directly related to the large deviations rate function for information spectrum. This is however different from Merhav and 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

$$(1-\alpha)\lim_{n\to\infty} n^{-1}H_\alpha(P_n) = \lim_{n\to\infty} n^{-1}\ln \sum_{x^n\in\mathbb{X}^n} \exp\{-n\alpha F_n(x^n)\}\, U_n(x^n) \quad (22)$$

where

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

and U is the i.i.d. 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., α times) the specific Gibbs free energy of the corresponding statistical mechanical system, if the limit exists. This view point is particularly useful because the i.i.d. process U satisfies a sample path large deviation property. If the information spectrum sequence satisfies the continuity conditions in Varadhan [19, Th. 3.4], then the limiting specific Gibbs free energy exists, and so does the Rényi entropy rate. Our technical report [10] treats an example via this more general approach.

A. Additional Results From Large Deviations Theory

In order to study the examples in Section IV, we state some additional results on LDP of transformed variables. (See [20, Sec. 4.2]), [21, 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 ν_n denote the distribution of ξ_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 $I : \mathcal{X} \to [0,\infty]$. 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 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 $\delta > 0$

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

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

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

where Ω is the underlying sample space.

IV. 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.

A. LDP for Information Spectrum

Example 1 (An i.i.d. Source): This example was first studied by Arikan [3]. Recall that an i.i.d. 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 i.i.d. 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, [18, Th. II.4.1] or [9, eq. (1.9.66-67)])

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

The Legendre-Fenchel dual of the rate function is therefore the cumulant itself ([18, 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[3] with rate function $I^{(2)}_{P_1}(\cdot) = D(\cdot \parallel P_1)$. See for example [18, 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 [18, Th. II.5.1]. For any t in $\mathbb{R}$, let

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

i.e.

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

Then

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

[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.

Using this, we can write

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

thus yielding

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

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 and Sullivan [4]. 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) for this source.

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

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

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 [22]

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

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

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

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 [23, Ch. 1], see also [24, 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 π to the power α. (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 and 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

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

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 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

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

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 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, eq. (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 i.i.d. case and (26) for a Markov source, we get

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

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 and Sullivan [5] considered a class of stationary sources with distribution $P \in \mathcal{M}\left(\mathbb{X}^{\mathbb{N}}\right)$ 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 $\tau : \mathbb{X}^{\mathbb{N}} \to \mathbb{X}^{\mathbb{N}}$ defined by

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

Hypothesis H1 of Pfister and 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).

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

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

where

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

En route to this result, Pfister and 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 $x = (x_1, x_2, \ldots)$, we define $x^n = (x_1, \ldots, x_n)$ as the first n components of x in the usual way. Consider a stationary source P whose letters are $X = (X_1, X_2, \ldots)$. 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: $x, \tau(x), \tau^2(x), \ldots, \tau^{n-1}(x)$. Pfister and 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

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

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

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

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

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

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

defined by $Q \mapsto \int_{\mathbb{X}^{\mathbb{N}}} e_P \, dQ$ is a continuous mapping. Hence by the contraction principle, by setting $\mathcal{X} = \mathcal{M}(\mathbb{X}^{\mathbb{N}})$ 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 and Sullivan's work and use $\overline{D}(Q \parallel P) = \phi(Q) - \overline{H}(Q)$ [5, Prop. 2.1] to write

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

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

Example 5 (Mixed Source): Consider a mixture of two i.i.d. 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 i.i.d. 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, eq. (1.9.41)]). Define

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

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

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

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

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

yielding

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

V. Proofs

We now prove Propositions 6 and 7.

A. 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 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 [25, Prop. 1.4.2] as the following Legendre-Fenchel dual of the Kullback-Leibler divergence, i.e.

$$\begin{aligned}
&\ln\mathbb{E}_{P_n}\left[\exp\{(\rho\ln 2)L_n(X^n)\}\right]\\
&= \sup_{Q_n\in\mathcal{M}(\mathbb{X}^n)}\{(\rho\ln 2)\mathbb{E}_{Q_n}[L_n(X^n)] - D(Q_n\parallel P_n)\}. \quad (31)
\end{aligned}$$

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

$$\begin{aligned}
&\inf_{L_n}\ln\mathbb{E}_{P_n}\left[\exp\{(\rho\ln 2)L_n(X^n)\}\right] \quad (32)\\
&\quad= \inf_{L_n}\sup_{Q_n\in\mathcal{M}(\mathbb{X}^n)}\{\mathbb{E}_{Q_n}[(\rho\ln 2)L_n(X^n)] - D(Q_n\parallel P_n)\}\\
&\quad= \sup_{Q_n\in\mathcal{M}(\mathbb{X}^n)}\inf_{L_n}\{\mathbb{E}_{Q_n}[(\rho\ln 2)L_n(X^n)] - D(Q_n\parallel P_n)\}\\
&\qquad+\Theta(1) \quad (33)\\
&\quad= \sup_{Q_n\in\mathcal{M}(\mathbb{X}^n)}\{\rho H_n(Q_n) - D(Q_n\parallel P_n)\} + \Theta(1) \quad (34)\\
&\quad= \rho H_{\frac{1}{1+\rho}}(P_n) + \Theta(1). \quad (35)
\end{aligned}$$

Equation (33) follows because: (i) the mapping

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

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

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

(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 (34) follows. Last, (35) is a well-known identity which may also be obtained directly by writing the supremum term in (34) as

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

and then applying (31) 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 (32) and (35) by n and let $n\to\infty$ to deduce that (21) exists if and only if the limiting normalized Rényi entropy rate exists. This concludes the proof.

B. Proof of Proposition 7

This is a straightforward application of Varadhan's theorem [19] on asymptotics of integrals. Recall that ν_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 [18, Th. II.7.1.b]) states that if

$$\lim_{M\to\infty}\limsup_{n\to\infty}\frac{1}{n}\ln\int_{t\geq\frac{M}{\beta}}\exp\{n\beta t\}\,d\nu_n(t) = -\infty \quad (36)$$

then the limit

$$\lim_{n\to\infty}\frac{1}{n}\ln\int_{\mathbb{R}}\exp\{n\beta t\}\,\nu_n(dt) = \sup_{t\in\mathbb{R}}\{\beta t - I(t)\} \quad (37)$$

holds. The integral on the LHS in (37) can be simplified by defining the finite cardinality set

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

and by observing that

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

Take logarithms, normalize by n, take limits, and apply (37) to get the desired result. It therefore remains to prove (36).

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 (36) can, therefore, be written as

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

The sequence in n on the LHS of (36) is then

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

a constant sequence. Take the limit as $M \to \infty$ to verify (36). This concludes the proof.

VI. 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. While some of these facts can be gleaned from the works of Pfister & Sullivan [5] and Merhav and 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 paper 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].

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 Int. Symp. Inf. Theory*, Trondheim, Norway, Jun. 1994, pp. 204–204.

[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 *Proc. Conf. Managing Complex. Distributed World, MCDES*, Bangalore, India, May 2008 [Online]. Available: 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*. New York: Springer-Verlag, 2003.

[10] M. K. Hanawal and R. Sundaresan, Guessing Revisited: A Large Deviations Approach 2008, DRDO-IISc Programme in Mathematical Engineering Technical Report No. TR-PME-2008-08.

[11] K. R. Parthasarathy, *Coding Theorems of Classical and Quantum Information Theory*. New Delhi, India: Hindustan Book Agency, 2007.

[12] K. Marton and P. C. Shields, "The positive-divergence and blowing-up properties," *Israel J. Math.*, vol. 86, pp. 331–348, 1994.

[13] P. C. Shields, "The interactions between ergodic theory and information theory," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2079–2093, Oct. 1998.

[14] 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.

[15] A. D. Wyner, "An upper bound on the entropy series," *Inf. Contr.*, vol. 20, no. 2, pp. 176–181, Mar. 1972.

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

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

[18] R. S. Ellis, *Entropy, Large Deviations, and Statistical Mechanics*, ser. Grundlehren der mathematischen Wissenschaften. New York: Springer-Verlag, 1985, vol. 271.

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

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

[21] R. S. Ellis, "The theory of large deviations and applications to statistical mechanics," in *Proc. Lectures for the Int. Seminar on Extreme Events in Complex Dynam.*, Dresden, Germany, Oct. 2006.

[22] 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.

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

[24] F. den Hollander, *Large Deviations*. Providence, RI: Amer. Math. Soc., 2003.

[25] P. Dupuis and R. S. Ellis, *A Weak Convergence Approach to the Theory of Large Deviations*. New York: Wiley, 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 2009, DRDO-IISc Programme in Mathematical Engineering Technical Report No. TR-PME-2009-04.

Manjesh Kumar Hanawal received the B.E. degree in electronics and communication from the National Institute of Technology, Bhopal, and the M.Sc. degree in electrical communication engineering from the Indian Institute of Science, Bangalore, in 2009.

From 2004 to 2007, he was with the Centre for Artificial Intelligence and Robotics (CAIR), Bangalore, as Scientist-B on various cryptographic projects. Currently, he is pursuing the Ph.D. degree at INRIA and the University of Avignon, France.

Rajesh Sundaresan (S'96–M'00–SM'06) received the B.Tech. degree in electronics and communication from the Indian Institute of Technology, Madras, and the M.A. and Ph.D. degrees in electrical engineering from Princeton University, Princeton, NJ, in 1996 and 1999, respectively.

From 1999 to 2005, he was with Qualcomm Inc., Campbell, CA, working on the design of communication algorithms for WCDMA and HSDPA modems. Since 2005, he has been with the Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore. His interests are in the areas of wireless communication networks and information theory.