Large Deviations of Probability Rank

Erdal Arikan
Electrical Eng. Dept.
Bilkent University
06533 Ankara, Turkey
arikan@ee.bilkent.edu.tr

Abstract — **Consider a pair of random variables (X, Y) with distribution P. The probability rank function is defined so that $G(x|y) = 1$ for the most probable outcome x conditional on $Y = y$, $G(x|y) = 2$ for the second most probable outcome, and so on, resolving ties between elements with equal probabilities arbitrarily. The function G was considered in [1] in the context of finding the unknown outcome of a random experiment by asking questions of the form 'Is the outcome equal to x?' sequentially until the actual outcome is determined. The primary focus in [1], and the subsequent works [2], [3], was to find tight bounds on the moments $\mathbf{E}[G(X|Y)^\theta]$. The present work is closely related to these works but focuses more directly on the large deviations properties of the probability rank function.**

I. Results

The aim of this work is to determine the large deviation exponent of $\ln G$,

$$\lim_{n\to\infty} n^{-1} \ln P[\ln G(X^n|Y^n) > nL], \tag{1}$$

for a sequence of pairs of r.v.'s (X^n, Y^n) under various assumptions regarding their distribution. Special instances of this problem correspond to finding the error exponent in source and channel coding problems of information theory. E.g., if we regard X^n as an input of length n to a noisy channel and Y^n as the channel output, $P[\ln G(X^n|Y^n) > nL]$ is the probability of decoding error for a list decoder with list size e^{nL}. We begin by noting that the mean of $\ln G$ is closely related to the Shannon entropy.

Proposition 1 *For (X, Y) a pair of jointly distributed random variables,*

$$-\ln(1+\ln M) + H(X|Y) \le \mathbf{E}[\ln G(X|Y)] \le H(X|Y) \tag{2}$$

where M is the maximum over all y of the range of X conditioned on $Y = y$.

We study large-deviations of $\ln G(X^n|Y^n)$ under the assumption that the sequence of functions

$$\varphi_n(\theta) \triangleq \frac{1}{n} \ln \mathbf{E}[G(X^n|Y^n)^\theta] \tag{3}$$

converges to a limit $\varphi(\theta)$. We let $R_{\varphi'}$ denote the range of φ'. Now, the Gärtner-Ellis theorem [4, p.15] gives

Proposition 2 *For any $L \in R_{\varphi'}$,*

$$\lim_{n\to\infty} n_{-1} \ln P[\ln G(X^n|Y^n) > nL] = \varphi(\theta_L) - \theta_L \varphi'(\theta_L) \tag{4}$$

where $\theta_L = \inf\{\theta : \varphi'(\theta) = L\}$.

For the special case where (X^n, Y^n) is a pair of random vectors with i.i.d components, we recall from [1] that for any $\theta \ge 0$

$$\lim_{n\to\infty} \varphi_n(\theta) = \varphi(\theta) = \ln \sum_y \left[\sum_x P(x,y)^{1/(1+\theta)}\right]^{1+\theta} \tag{5}$$

This yields the source coding error exponent (with side information Y^n). The well-known source coding error exponent [5, p.37] is obtained by omitting the side information term.

Another special case of interest is when X^n represents a codeword from a block code with block length n and rate R. Then, $P(x^n) = e^{-nR}$ if x^n is a codeword and 0 otherwise. This distribution is called the code's empirical distribution and denoted Q_n below. The r.v. Y^n represents the channel output when X^n is transmitted. We recall from [1] that for $\theta \ge 0$,

$$\varphi_n(\theta) = \theta R - n^{-1} E_0(\theta, Q_n) + o(n) \tag{6}$$

where E_0 is is Gallager's function [6, p. 138] and $o(n)$ is a quantity that goes to zero as n goes to infinity. Proposition 2 now yields the well-known sphere-packing bound for list-decoding.

In the case of $L = 0$, which corresponds to ordinary ML decoding, Proposition 2 may not apply since 0 may not belong in $R_{\varphi'}$. In this case, Gärtner-Ellis theorem yields only a lower-bound.

Proposition 3 *Let $\{(X^n, Y^n)\}$ be a sequence of input-output pairs for a noisy channel such that $\{\varphi_n\}$ converges to a limit φ. Then,*

$$\liminf_{n\to\infty} n^{-1} \ln P[\ln G(X^n|Y^n) > 0] \ge -\theta_0 \varphi'_+(\theta_0) \tag{7}$$

where $\theta_0 = \inf\{\theta : \varphi(\theta) > 0\}$ and φ'_+ denotes right-derivative.

It can be shown that this bound is equivalent to the familiar sphere-packing lower bound [6, p. 157], except it is formulated in terms of code empirical distributions.

References

[1] E. Arikan, "An inequality on guessing and its application to sequential decoding," *IEEE Trans. Inform. Theory,* vol. IT-42, no. 1, pp. 99-105, January 1996.

[2] E. Arikan and N. Merhav, "Guessing subject to distortion," *IEEE Trans. Inform. Theory,* vol. IT-44, no. 3, pp. 1041–1056, May 1998.

[3] E. Arikan and N. Merhav, "Joint source-channel coding and guessing with application to sequential decoding," *IEEE Trans. Inform. Theory,* vol. IT-44, no. 5, pp. 1756–1769, September 1998.

[4] J. A. Bucklew, *Large Deviation Techniques in Decision, Simulation, and Estimation.* New York: Wiley, 1990.

[5] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems.* New York: Academic, 1981.

[6] R. G. Gallager, *Information Theory and Reliable Communication.* New York: Wiley, 1968.