On The Most Significant Bit w.r.t. Side Information

Gowtham Ramani Kumar, ISL, Stanford

Introduction:

Given a single i.i.d. source $X^n \sim \prod_{i=1}^{n} p(x_i)$, one can find efficient schemes to compress it. However, one may not always be interested in X^n. One may instead be interested in a correlated sequence Y^n. For our setting, we simply assume $X^n, Y^n \sim \prod p(x_i, y_i)$. We also restrict ourselves to the simple case where X_i, Y_i are i.i.d. Bern(1/2) and are related via a BSC with crossover probability α.

We are interested in the following question: If I am allowed to say only 1 bit of information about the X^n sequence, and my goal is to convey the maximum possible amount of information about the Y^n sequence, what is the 1 bit I must specify?

Problem statement:

Given: $X^n, Y^n \sim \prod p(x_i, y_i)$

where $p(x,y) = \begin{pmatrix} \frac{1-\alpha}{2} & \frac{\alpha}{2} \\ \frac{\alpha}{2} & \frac{1-\alpha}{2} \end{pmatrix}, x, y \in \{-1,1\}$

We are interested in a function $b: X^n \to \{-1,1\}$ that maximizes

$$I(b; Y^n)$$

Related results:

- It is impossible to find $b(X^n), \tilde{b}(Y^n)$ so that $b = \tilde{b}$ with high probability unless $\alpha = 0$. [1]
- If the requirement is to compress X^n at a rate R while maximizing $\frac{1}{\mathrm{n}} I(M; Y^n)$, then the initial efficiency
$$\lim_{R \to 0} \frac{\frac{1}{\mathrm{n}} I(M; Y^n)}{R} = \rho^2$$
where $\rho = EXY = (1 - 2\alpha)$ is the Renyi correlation between X and Y. [2]
- In [3], the authors discuss the information-bottleneck method: a generalization of rate-distortion function with distortion $d(x, \tilde{x})$ depending on the joint-statistics $p(x,y)$.

Motivation:

- Goal changed from minimizing distortion to describing a correlated sequence.
  - Example 1: X^n is a sound file and Y^n is the set of words in that sound file. [3]
  - Example 2: X^n is an image of people in a bar and Y^n is a list of their names. [3]
  - Example 3: X^n is side-information and Y^n is a horse-race. [2]
- Random coding fails for this problem!
  - A random bit $b(X^n)$ is independent of Y^n.
  - In fact we can generate $\mathrm{n}(H(X|Y) - \epsilon)$ random bits and guarantee independence.

Inner bounds:

- The trivial inner bound: $b(X^n) = X_1$ achieves $I(X_1; Y^n) = 1 - H(\alpha)$.
- One can attempt to construct a more sophisticated inner bound: $b(X^n) = 1(X^n$ has more 1's than 0's).

To compute this inner bound:

Let $\bar{X} = \frac{\sum X_i}{\sqrt{n}}, \bar{Y} = \frac{\sum Y_i}{\sqrt{n}}$. Then by CLT, $\bar{X}, \bar{Y}$ are jointly Gaussian with unit variance and covariance $\rho = EXY$.

It turns out this inner-bound is worse than the trivial inner bound.

Outer bound:

It was shown in [2] that if $U - X - Y$, then

$$\frac{I(U;Y)}{I(U;X)} \leq \rho^2$$

Here, $b - X^n - Y^n$ and $I(b; X^n) \leq 1$. Hence,

$$I(b; Y^n) \leq \rho\text{^}2$$

Here, $\rho = 1 - 2\alpha$.

A hypothesis:

It appears plausible that the trivial inner bound is optimal, i.e., for any bit $b(X^n)$,

$$I(b; Y^n) \leq 1 - H(\alpha)$$

Proof ideas?

Comments? Questions? Suggestions?

References:

[1] H.S. Witsenhausen, "On sequences of pairs of dependent random variables", SIAM J. Appl. Maths, Vol. 28, Jan. 1975.

[2] Elza Erkip, Member, IEEE, and Thomas M. Cover, Fellow, IEEE, " The Efficiency of Investment Information", IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. 44, NO. 3, MAY 1998.

[3] Naftali Tishby, Fernando C. Pereira, William Bialek, " The Information Bottleneck Method", The 37th annual Allerton Conference on Communication, Control, and Computing, Sep 1999: pp. 368–377.