Homework Set #7

Due 3 December 2009, 6 pm, in INR036

Problem 1 (BSC with feedback)

Suppose that feedback is used on a binary symmetric channel with parameter p. Each time a Y is received, it becomes the next transmission. Thus, X_1 is Bern($\frac{1}{2}$), $X_2 = Y_1$, $X_3 = Y_2$, ..., $X_n = Y_{n-1}$.

(a) Find $\lim_{n\to\infty} \frac{1}{n} I(X^n; Y^n)$.

(b) Show that for some values of p, this can be higher than capacity.

(c) Using this feedback transmission scheme, $X^n(W; Y^n) = (X_1(W), Y_1, Y_2, \ldots, Y_{n-1})$, what is the asymptotic communication rate achieved; that is, what is $\lim_{n\to\infty} \frac{1}{n} I(W; Y^n)$?

Note: Note that $\lim_{n\to\infty} \frac{1}{n} I(X^n; Y^n)$ is not the rate at which this scheme can convey information from source to destination. To find the communication rate for this scheme we should find the mutual information between transmitted message W and received sequence of bits Y^n namely $\lim_{n\to\infty} \frac{1}{n} I(W; Y^n)$. This is in fact the quantity that we calculate in part (c).

Problem 2 (Channel with dependence between the letters)

Consider the following channel over a binary alphabet that takes in 2−bit symbols and produces a 2−bit output, as determined by the following mapping: $00 \to 01$, $01 \to 10$, $10 \to 11$, and $11 \to 00$. Thus, if the 2−bit sequence 01 is the input to the channel, the output is 10 with probability 1. Let X_1, and X_2 denote the two input symbols and Y_1, Y_2 denote the corresponding output symbols.

(a) Calculate the mutual information $I(X_1, X_2; Y_1, Y_2)$ as a function of the input distribution on the four possible pairs of inputs.

(b) Show that the capacity of a pair of transmissions on this channel is 2 bits.

(c) Show that under the maximizing input distribution, $I(X_1; Y_1) = 0$. Thus, the distribution on the input sequence that achieves capacity does not necessarily maximize the mutual information between individual symbols and their corresponding outputs.

Problem 3

For arbitrary sequence of $\mathcal{X}$−valued RV's $X^k \triangleq X_1 \cdots X_k, Y^k \triangleq Y_1 \cdots Y_k$ show that

$$H(X^k|Y^k) \leq \mathbb{E}d_H(X^k, Y^k)\log(|\mathcal{X}|) + kH\left(\frac{1}{k}\mathbb{E}d_H(X^k, Y^k)\right)$$

Hint: Use chain rule, and Fano's inequality.

Problem 4 (Capacity per unit cost)

Consider a DMC with cross-over probability $W_{Y|X}(\cdot|\cdot)$ with input $x \in \mathcal{X}$ and output $y \in \mathcal{Y}$. It costs $c(x)$ to send symbol $x \in \mathcal{X}$ over the channel. Assume that $c(x) > 0$, $\forall x \in \mathcal{X}$.

For some $x \in \mathcal{X}$, let us define the information divergence between $W_{Y|X=x}$ (which we write as $W_{Y|X}$ for shorthand) and P_Y as

$$D(W_{Y|X} \parallel P_Y) \triangleq \sum_{y \in \mathcal{Y}} W_{Y|X}(y|x) \log \frac{W_{Y|X}(y|x)}{P_Y(y)}.$$

Note that $D(W_{Y|X} \parallel P_Y)$ is still a function of $x \in \mathcal{X}$.

(a) Given any distribution P_Y, show that for any choice of input distribution $\tilde{P}_X(x)$,

$$\frac{\sum_{x \in \mathcal{X}} \tilde{P}_X(x) D(W_{Y|X} \parallel P_Y)}{\sum_{x \in \mathcal{X}} \tilde{P}_X(x) c(x)} \leq \max_{x \in \mathcal{X}} \frac{D(W_{Y|X} \parallel P_Y)}{c(x)}$$

Hint: You may use the following fact. For $a, b, c, d > 0$ with $\frac{a}{b} \leq \frac{c}{d}$, we always have $\frac{a}{b} \leq \frac{a+c}{b+d} \leq \frac{c}{d}$.

(b) Let $\tilde{P}_X$ and P_X be arbitrary input distributions and let $\tilde{P}_Y(y) = \sum_{x \in \mathcal{X}} \tilde{P}_X(x) W_{Y|X}(y|x)$ and $P_Y(y) = \sum_{x \in \mathcal{X}} P_X(x) W_{Y|X}(y|x)$ be the resulting output distributions. Show that

$$\sum_{x \in \mathcal{X}} \tilde{P}_X(x) D(W_{Y|X} \parallel P_Y) - \sum_{x \in \mathcal{X}} \tilde{P}_X(x) D(W_{Y|X} \parallel \tilde{P}_Y) \geq 0$$

and conclude that

$$\frac{\sum_{x \in \mathcal{X}} \tilde{P}_X(x) D(W_{Y|X} \parallel \tilde{P}_Y)}{\sum_{x \in \mathcal{X}} \tilde{P}_X(x) c(x)} \leq \max_{x \in \mathcal{X}} \frac{D(W_{Y|X} \parallel P_Y)}{c(x)}$$

Hint: Use properties of information divergence (Kullback-Leibler distance) or Jensen's inequality.

(c) Suppose we are given an input distribution $P_X^*(x)$ such that

$$\frac{D(W_{Y|X} \parallel P_Y^*)}{c(x)} \leq \lambda, \ \forall x \in \mathcal{X}$$

and

$$\frac{D(W_{Y|X} \parallel P_Y^*)}{c(x)} = \lambda, \ \forall x : P_X^*(x) > 0$$

where $P_Y^*(y) = \sum_x P_X^*(x) W_{Y|X}(y|x)$.

Using part (b) show that for any $\tilde{P}_X(x)$, and $\tilde{P}_Y(y) = \sum_{x \in \mathcal{X}} \tilde{P}_X(x) W_{Y|X}(y|x)$,

$$\frac{\sum_{x \in \mathcal{X}} \tilde{P}_X(x) D(W_{Y|X} \parallel \tilde{P}_Y)}{\sum_{x \in \mathcal{X}} \tilde{P}_X(x) c(x)} \leq \lambda$$

with equality if and only if $\tilde{P}_Y(y) = P_Y^*(y)$.

(d) Using the result found in **(c)**, what can you conclude about the optimizing input distribution for finding the *capacity per unit cost* C_{cost}, defined as,

$$C_{\text{cost}} = \max_{P_X(x)} \frac{I(X;Y)}{\mathbb{E}[c(x)]}.$$