Homework Set #6

Due 26 November 2009, 6 pm, in INR036

Problem 1 (Cascade network)

Consider the channel shown in Fig. 1 which is formed by cascading two binary symmetric channels. The source S sends $X \in \{0,1\}$ through the first channel and the relay node R receives $Y \in \{0,1\}$ where $\Pr(Y \neq X) = p \leq \frac{1}{2}$. Then the relay R produces $U \in \{0,1\}$ and feeds it to the second channel and the destination D receives $V \in \{0,1\}$ where $\Pr(U \neq V) = q \leq \frac{1}{2}$. We denote by $C(p) = 1 - H_2(p)$ and $C(q) = 1 - H_2(q)$ the capacities of the first and the second channel, respectively, where $H_2(\cdot)$ is the binary entropy function. We use C to denote the capacity of the network from X to V.

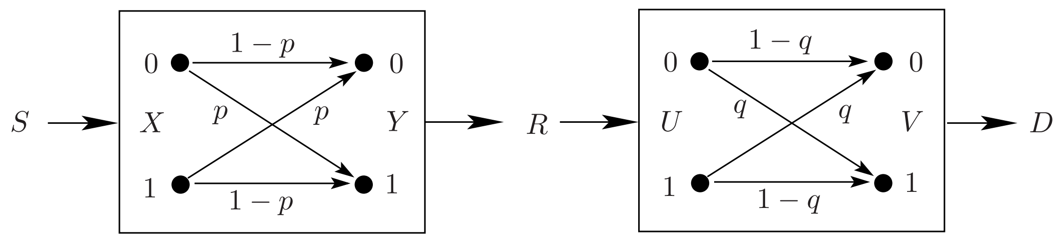

Figure 1: A cascade of two binary symmetric channels

(a) Show that $C \leq C(p)$ and $C \leq C(q)$.

(b) Assume that there is no processing allowed at the relay, *i.e.,* the relay node just forwards its received bit to the destination. Find the capacity of the channel $C' = \max_{p(x)} I(X;V)$. *Hint:* Do the two channels create a single binary symmetric channel in this scenario?

(c) Now assume that the relay node can wait to receive an arbitrary number of bits, process them, and produce a sequence $U_1, \ldots, U_n$ which is a function of its observed signal $Y_1, \ldots, Y_{n-1}$ to transmit to the destination node. Show that $C = \min\{C(p), C(q)\}$ is achievable, by demonstrating a scheme that achieves it.

(d) Compare C and C'. Which one is larger? Explain why.

Problem 2 (Binary multiplier channel)

Consider the channel $Y = XZ$, where X and Z are independent binary random variables that take on values 0 and 1. The random variable Z is Bernoulli(α) [*i.e.*, $\mathbb{P}[Z = 1] = \alpha$].

(a) Suppose that the receiver can observe Z as well as Y. What is the capacity of this channel in this case?
Hint: Have you seen this channel before?

(b) Now assume that the receiver has only access to Y. Find the capacity of the channel in this case and find the maximizing distribution on X. Assuming $\mathbb{P}[X=1]=p$ you should find the value of p such that maximizes the mutual information $I(X;Y)$.
Hint:

1: *Expand the mutual information as* $I(X;Y) = H(Y) - H(Y|X)$.

2: *Find the probability* $\mathbb{P}[Y=1]$ *and write the entropy of* Y.

3: *Then you have to find an expression for* $H(Y|X)$. *To this end, you may use the following expansion*

$$H(Y|X) = H(Y|X=0)\mathbb{P}[X=0] + H(Y|X=1)\mathbb{P}[X=1].$$

4: *Now you have an expression for the mutual information with respect to the parameters* α *and* p. *You have to maximize this expression with respect to* p.

Problem 3 (Jointly typical sequences)

We consider a binary erasure channel (BEC) with erasure probability α.

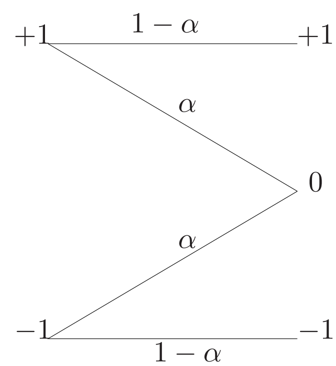

Figure 2: Binary erasure channel with erasure probability α

The input distribution that achieves capacity is the uniform distribution (i.e. $p(x) = (\frac{1}{2}, \frac{1}{2})$).

(a) The jointly typical set $A_\epsilon^{(n)}(X,Y)$ is defined as the set of sequences $x^n \in \{+1,-1\}^n, y^n \in \{-1,0,+1\}^n$ that satisfy equations

$$\left| -\frac{1}{n} \log p(x^n) - H(X) \right| < \epsilon \tag{1}$$

$$\left| -\frac{1}{n} \log p(y^n) - H(Y) \right| < \epsilon \tag{2}$$

$$\left| -\frac{1}{n} \log p(x^n, y^n) - H(X,Y) \right| < \epsilon. \tag{3}$$

The first two equations correspond to the conditions that x^n and y^n are in $A_\epsilon^{(n)}(X)$ and $A_\epsilon^{(n)}(Y)$, respectively. Consider the last condition, which can be rewritten as

$$-\frac{1}{n} \log p(x^n, y^n) \in (H(X,Y) - \epsilon, H(X,Y) + \epsilon).$$

To compute $p(x^n, y^n)$ one approach is to look at the binary erasure channel as a binary multiplier channel $Y = XZ$, where Z is a binary random variable that is equal to 0 with probability α, and is independent of X. In that case,

$$\begin{aligned} p(x^n, y^n) =& p(x^n)p(y^n|x^n) \\ =& p(x^n)p(z^n|x^n) \\ =& p(x^n)p(z^n) \\ =& \left(\frac{1}{2}\right)^n (1-\alpha)^{n-k}\alpha^k. \end{aligned}$$

where k is the number of places that we have erasure(0) in the sequence y^n.
Show that the condition that (x^n, y^n) are jointly typical is equivalent to the condition that x^n is typical and $z^n = x^n y^n$ is typical. (If $x^n = (x_1, \cdots, x_n)$, $y^n = (y_1, \cdots, y_n)$, then by $z^n = x^n y^n$ we mean that $\forall 1 \leq i \leq n$ $z_i = x_i y_i$ and $z^n = (z_1, \cdots, z_n)$)
Hint: Find $H(X)$, $H(Y)$, $H(X, Y)$, $H(Z)$.

(b) Now consider random coding for the multiplier channel. As in the proof of the channel coding theorem, assume that 2^{nR} codewords $X^n(1), X^n(2), \ldots, X^n(2^{nR})$ are chosen uniformly over the 2^n possible binary sequences of length n. One of these codewords is chosen and sent over the channel. The receiver looks at the received sequence and tries to find a codeword in the code that is jointly typical with the received sequence. This corresponds to finding a codeword $X^n(i)$ such that $X^n(i)Y^n \in A_\epsilon^{(n)}(Z)$. For a fixed codeword $x^n(i)$, what is the probability that the received sequence Y^n is such that $(x^n(i), Y^n)$ is jointly typical?

(c) Now consider a particular received sequence y^n. Assume that we choose a sequence X^n at random, uniformly distributed among all the 2^n possible binary $n-$sequences. What is the probability that the chosen sequence is jointly typical with this y^n?

(d) Given that a particular codeword was sent, the probability of error (averaged over the probability distribution of the channel and over the random choice of codewords) can be rewritten as

$$Pr(\text{error}|x^n(1)\text{ sent}) = \sum_{y^n : y^n \text{ causes error}} p(y^n|x^n(1)).$$

There are two kinds of error: the first occurs if the received sequence y^n is not jointly typical with the transmitted codeword, and the second occurs if there is another codeword jointly typical with the received sequence. Using the results of the preceding parts, calculate this probability of error. By the symmetry of the random coding argument, this does not depend on which codeword was sent.

Problem 4 (Deterministic Channel)

In this problem, we consider channels in which the output is a deterministic function of the input, i.e., if we denote the input of the channel by X and the output by Y, then $Y = f(X)$, where f is a deterministic function.

(a) Let $X \in \{1, 2, \cdots, n\}$ and $Y \in \{1, 2, \cdots, m\}$ and $Y = f(X)$, where f is a surjective function. What is the capacity of the channel? How does the answer change if f was not surjective?
Definition: A function $f : A \to B$ is surjective if and only if for every b in the target set B there is at least one a in the domain A such that $f(a) = b$.

(b) Consider now the following linear deterministic channel. The channel is defined by an $m \times n$ matrix A with entries in $\mathbb{F}_2$. The input to the channel is a vector $X = [x_1, \cdots, x_n]^t$, with entries in $\mathbb{F}_2$. The output of the channel Y is then obtained by $Y = AX$. All the operations are done in $\mathbb{F}_2$; i.e., they are all modulo 2. This means that the output of the channel would be a vector of length m with entries in $\mathbb{F}_2$. What is the capacity of this channel?

Definition: The finite field $\mathbb{F}_2$ is a set with two elements, $\{0, 1\}$, which is occupied with two operations; the addition and the multiplication modulo 2.

(c) Consider the scenario where we cascade two linear deterministic channels. Use your answer to part (a) of exercise one to prove the following fact about two matrices P and T,

$$\text{rank}(PT) \leq \min\{\text{rank}(P), \text{rank}(T)\}.$$