Random codes in communication 10 July 2011

Final Exam

1) **True or False** (30 points)
Copy each relation to your notebook and write **true** or **false**. Then, if it's true, prove it. If it is false give a counterexample or prove that the opposite is true.
   a) Let X, Y be two random variables. Then $H(X - Y) \leq H(X|Y)$. [4 points]

      Solution: **False**.

$$H(X-Y) \geq H(X-Y|Y) \\ = H(X|Y).$$

   b) For any finite alphabet random variables $H(X, Y, Z) - H(X, Y) \geq H(X, Z) - H(X)$. [4 points]

      Solution: **False**.

$$H(X,Y,Z) - H(X,Y) = H(Z|X,Y) \\ \leq H(Z|X) \\ = H(X,Z) - H(Z).$$

   c) Which of the following sequence of code-lengths are a valid binary huffman codes(can be more than one answer)? [4 points]
      - 1,2,3,3 V
      - 1,2,2,3 X
      - 1,3,3,3 X
      - 2,2,2,2 V

   d) Let $\{X_i\}_{i\geq 1}$ be an i.i.d. source distributed according to P_X. In addition, let $\{Y_i\}_{i\geq 1}$ and $\{Z_i\}_{i\geq 1}$ be two i.i.d. side information sequences that may be available at the encoder and decoder of a lossless source coding setting. If $I(X;Y) > I(X;Z)$, then the minimum rate that is needed to compress $\{X_i\}_{i\geq 1}$ losslessly with side information $\{Y_i\}_{i\geq 1}$ is smaller than the minimum rate that is needed to compress $\{X_i\}_{i\geq 1}$ losslessly with with side information $\{Z_i\}_{i\geq 1}$. Assume the side information is known both to the encoder and decoder. [4 points]

      Solution: **True**. $I(X;Y) > I(X;Z)$, hence $H(X|Z) > H(X|Y)$ and the minimum rate that is needed to compress $\{X_i\}_{i\geq 1}$ losslessly is with side information $\{Y_i\}_{i\geq 1}$.

   e) Let X be a continues alphabet random variable. For which of the following channels with power constraint $\frac{1}{n}\sum_{i=1}^{n} x_i^2(m) \leq P$, the capacity is not finite (there can be more than one answer): [10 points]
      i) $Y = X$ V
      ii) $Y = X + Z$, where $Z \sim N(0, 1)$ X
      iii) $Y = X + Z$, where $Z \sim N(0, 1)$ with probability 0.9, and $Z = 0$ with probability 0.1. V

      Provide a scheme that achieves the infinite capacity, if exists.

Solution: Since X has continues alphabet, (a) and (c) - which has no noise - has infinite capacity. The scheme is for (c), but also apply for (a):

- Take the interval $[0, \sqrt{P}]$ and divide it into 2^{nR} points- $j\frac{\sqrt{P}}{2^{nR}}$, $j = 1, 2, ..., 2^{nR}$.
- For every message j, send $x_i(m) = j\frac{\sqrt{P}}{2^{nR}}$ for all $i = 1, 2, ..., n$.
- The decoder declares the message at the first time that $y = j\frac{\sqrt{P}}{2^{nR}}$ for some j.

An error accurse if $Z \neq 0$ for all i. Hence, $P(er) = 0.9^n$, which goes to zero with increasing n.

f) Assume a memoryless channel given by $p(y|x)$, and the capacity is given by $C = \max_{p(x)} I(X;Y)$. The capacity can be strictly increased by forming the output to be $Y_1 = f(Y)$. [4 points]

Solution: **False**. Since $H(X|Y) = H(X|Y, f(Y)) \geq H(X|f(Y))$. It can also be explained by the the fact that $X - Y - g(Y)$ is a Markov chain.

2) **Joint Entropy** (15 points) Consider n different discrete random variables, named $X_1, X_2, ..., X_n$. Each random variable separately has an entropy $H(X_i)$, for $1 \leq i \leq n$.

a) What is the upper bound on the joint entropy $H(X_1, X_2, ..., X_n)$ of all these random variables $X_1, X_2, ..., X_n$ given that $H(X_i)$, for $1 \leq i \leq n$ are fixed? [3 points]
b) Under what conditions will this upper bound be reached?
c) What is the lower bound on the joint entropy $H(X_1, X_2, ..., X_n)$ of all these random variables? [4 points]
d) Under what condition will this upper bound be reached? [4 points]
e) Assume the vector $[X_1, X_2, ..., X_n]$ is observed many times and one would like to compress it. Denote at time i the observed random vector as $[X_{1,i}, X_{2,i}, ..., X_{n,i}]$. The distribution of $[X_{1,i}, X_{2,i}, ..., X_{n,i}]$ is according to $[X_1, X_2, ..., X_n]$ for all i and for $i \neq j$ $[X_{1,i}, X_{2,i}, ..., X_{n,i}]$ is independent of $[X_{1,j}, X_{2,j}, ..., X_{n,j}]$.
Given that you have the possibility to optimally compress without any loss a sequence of random variable distributed i.i.d. but not a sequence of random vector. Provide a coding schemes for an optimal lossless compression for the random vector $[X_1, X_2, ..., X_n]$ given that they are distributed according to the distributions you found in subexercise 2b and subexercise 2d.

Solution:

a) The upper bound is $\sum_{i=1}^{n} H(X_i)$.
b) It can be achieved if all X_is are independent.
c) The lower bound is $H(X_i)$, where X_i has the largest entropy.
d) It can be achieved if for all $j \neq i$: $X_j = f_j(X_i)$ for some function f_j.
e) As in subexercise 2b, all X_is are independent, thus we must encode each of them apart. For subexercise 2d, we only encode the X_i with the largest entropy, and use the fact that any other X_j is a function of it to decode.

3) **Blahut-Arimoto's algorithm** (35 points) Recall, that the capacity of a memoryless channel is given by

$$C = \max_{p(x)} I(X;Y).$$

Solving this optimization problem is a difficult task for the general channel. In this question we develop an iterative algorithm for finding the solution for a fixed channel $p(y|x)$.

a) First, prove that the mutual information as a function of $p(x)$ and $p(x|y)$ as below. [3 points]

$$I(X;Y) = \sum_{x,y} p(x)p(y|x) \log \frac{p(x|y)}{p(x)}.$$

b) Show that $I(X;Y)$ as written above is convex in both $p(x),\ p(x|y)$ (Hint. You may use the Log-Sum-inequality). [8 points]
c) Find an expression for $p(x)$ that maximizes $I(X;Y)$ when $p(x|y)$ is fixed (Hint. You may use the Lagrange multipliers method with the constraint $\sum_x p(x) = 1$). [10 points]
d) Find an expression for $p(x|y)$ that maximizes $I(X;Y)$ when $p(x)$ is fixed (Hint. You may use the Lagrange multipliers method with constraints $\sum_x p(x|y) = 1$ for all y). [9 points]
e) Using (d), conclude that $C = \max_{p(x),p(x|y)} I(X;Y)$. [5 points]

The algorithm is performed by maximizing in each iteration over another variable; first over $p(x)$ when $p(x|y)$ is fixed, then over $p(x|y)$ when $p(x)$ is fixed, and so on. This iterative algorithm converges, and hence one can find the capacity of any DMC $p(y|x)$ with reasonable alphabet size.

Solution:

a) Since $I(X;Y) = H(X) - H(X|Y)$, the answer is obvious.
b) Recall, that the Log-Sum inequality is

$$\sum_{i=1}^{n} a_i \log \frac{a_i}{b_i} \geq \left(\sum_{i=1}^{n} a_i \right) \log \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} b_i}.$$

Hence

$$\begin{aligned}(\lambda p_1(x) + & (1-\lambda)p_2(x)) \log \tfrac{\lambda p_1(x)+(1-\lambda)p_2(x)}{\lambda p_1(x|y)+(1-\lambda)p_2(x|y)} \\ & \leq \lambda p_1(x) \log \tfrac{p_1(x)}{p_1(x|y)} + (1-\lambda)p_2(x) \log \tfrac{p_2(x)}{p_2(x|y)}.\end{aligned}$$

Taking the reciprocal of the logarithms yields

$$\begin{aligned}(\lambda p_1(x) + & (1-\lambda)p_2(x)) \log \tfrac{\lambda p_1(x|y)+(1-\lambda)p_2(x|y)}{\lambda p_1(x)+(1-\lambda)p_2(x)} \\ & \geq \lambda p_1(x) \log \tfrac{p_1(x|y)}{p_1(x)} + (1-\lambda)p_2(x) \log \tfrac{p_2(x|y)}{p_2(x)}.\end{aligned}$$

Multiplying by $p(y|x)$ and summing over all $x,\ y$, and letting $\mathcal{I}(p(x), p(x|y))$ be the mutual information as in (a), we obtain

$$\begin{aligned}\mathcal{I}(\lambda p_1(x) + & (1-\lambda)p_2(x), \lambda p_1(x|y) + (1-\lambda)p_2(x|y) \\ & \geq \lambda \mathcal{I}(p_1(x), p_1(x|y)) + (1-\lambda)\mathcal{I}(p_2(x), p_2(x|y)).\end{aligned}$$

c) Define the lagrangian

$$L = \sum_{x,y} p(x)p(y|x) \log \frac{p(x|y)}{p(x)} + \mu(\sum_x p(x) - 1),$$

and differentiate over $p(x)$. Solving $\frac{\partial L}{\partial p(x)} = 0$ provides us with

$$p(x) = \frac{\prod_y p(x|y)^{p(y|x)}}{\sum_x \prod_y p(x|y)^{p(y|x)}}.$$

d) Define the lagrangian

$$J = \sum_{x,y} p(x)p(y|x) \log \frac{p(x|y)}{p(x)} + \mu(y)(\sum_x p(x|y) - 1),$$

and differentiate over $p(x|y)$. Solving $\frac{\partial J}{\partial p(x|y)} = 0$ provides us with

$$p(x|y) = \frac{p(x)p(y|x)}{\sum_x p(x)p(y|x)}.$$

e) The expression for $p(x|y)$ is the one that corresponds to $p(x)$, and hence maximizing over $p(x),\ p(x|y)$ is the same as over $p(x)$ alone.

4) **Source-channel coding problem**(20 points) Consider the source-channel coding problem given in Fig. 1, where V, X, Y, W have a Binary alphabet. The source V is i.i.d. Bernoulli (p), and the channel is in Fig. 2.

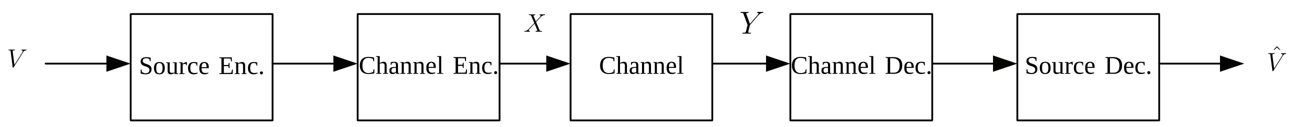

Fig. 1. A source-channel coding problem

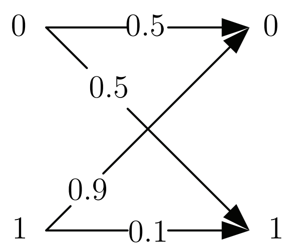

Fig. 2. The channel.

a) What is the capacity of the channel given in Fig. 2? [8 points]
b) Assume that error-free bits can be transmitted through the channel. What is the minimum rate in which the source V can be encoded such that the source decoder can reconstruct the source V losslessy? [4 points]
c) For what values of p can the source V be reconstructed losslessly using the scheme in Fig. 1 (you may use the inverse of H, i.e., $H^{-1}(q)$)? [4 points]
d) Would the answer to 4c changes if a joint source-channel coding and decoding is allowed? [4 points]

Solution:

a) We first write the mutual information function: $I(X;Y) = H(Y) - H(Y|X)$. Note that $H(Y|X) = pH(0.5) + (1-p)H(0.1)$, where $p = p(x = 0)$. As for $H(Y)$, note that Y is distributed$\sim B(0.9 - 0.4p)$. Hence,

$$I(X;Y) = -pH(0.5) - (1-p)H(0.1) - (0.9-0.4p)\log(0.9-0.4p) - (0.1+0.4p)\log(0.1+0.4p)$$

Solving $\frac{\partial I(X;Y)}{\partial p} = 0$ leaves us with

$$0.4 \log \frac{0.9 - 0.4p}{0.1 + 0.4p} = 1 - H(0.1),$$

or $p = 0.4623$. Thus, $C = 0.1476$.

b) The minimum rate, of course, is $R = H(p)$ since $V \sim B(p)$.

c) We require that $R \leq C$. Recall that $H^{-1}(C)$ has to values, given by $a, 1-a$. Thus, $p \leq a,\ p \geq 1 - a$ is the answer.

d) No, it would not change. This is due to the fact that for DMCs, it doesn't matter if you do joint source-channel decoding (source-channel separation theorem).

Good Luck!