Introduction to Information and Coding Theory August 3, 2016
(Prof. Permuter Haim, Mr. Oron Sabag and Mr. Iddo Gattengo)

Final Exam - Moed A
Total time for the exam: 3 hours!

Important: For **True / False** questions, copy the statement to your notebook and write clearly true or false. You should prove the statement if true, and provide counterexample otherwise.

1) **Cascaded BSCs (21 Points)**: Given is a cascade of k identical and independent binary symmetric channels, each with crossover probability α.
   a) In the case where no encoding or decoding is allowed at the intermediate terminals, what is the capacity of this cascaded channel as a function of k, α.
   b) Now, assume that encoding and decoding is allowed at the intermediate points, what is the capacity as a function of k, α.
   c) What is the capacity of each of the above settings in the case where the number of cascaded channels, k, goes to infinity?

   Solution:
   a) Cascaded BSCs result a new BSC with a new parameter, β. Therefore, the capacity is $C_a = 1 - H_2(\beta)$ and β can be found as $\beta = 0.5(1-(1-2\alpha)^k)$. Another expression that was written is $\beta = \sum_{\{i\leq k: i \text{ is odd}\}} \binom{k}{i}\alpha^i(1-\alpha)^{k-i}$.
   b) We have seen in HW that in the case of encoding and decoding the capacity of the cascaded channel equals $C_b = \min\{C_i\}$. Since all channels are identical, we have that $C_i = 1 - H_2(\alpha)$.
   c) In (a), $\beta \to 0.5$ as $k \to \infty$ so $C_a \to 0$. For (b), the number of cascaded channels does not change the capacity which remains $C_b = 1 - H_2(\alpha)$.
2) **True or False on conditional independence probabilities (10 Points)**:
   Given are three discrete random variables X, Y, Z.
   a) **True**/**False**: If $X \perp\!\!\!\perp Y$ then $X \perp\!\!\!\perp Y|Z$.
   b) **True**/**False**: If $X \perp\!\!\!\perp Y|Z$ then $X \perp\!\!\!\perp Y$.

   Solution:
   a) False. For example, take two independent random variables $X \sim Bern(0.5)$ and $Y \sim Bern(\alpha)$ and let $Z = X \bigoplus Y$. It is clear that $I(X;Y) = 0$, while $I(X;Y|Z) = H(X|Z) = H_2(\alpha) > 0$.
   b) False. For example, $X = Y = Z$ for some discrete random variable with $H(X) > 0$. In this case, $I(X;X|X) = 0$, while $I(X;X) = H(X) > 0$.
3) **Disjoint sets on discrete random variable (28 Points)**:
   Let X_0 and X_1 be discrete random on the alphabets $\mathcal{X}_0 = \{1, ..., m\}$ and $\mathcal{X}_1 = \{m+1, ..., n\}$, respectively. Let θ be a binary random variable with $P(\theta = 1) = p$, for some p. Let

$$X = \begin{cases} X_0 & \text{if } \theta = 0 \\ X_1 & \text{if } \theta = 1. \end{cases}$$

   a) Find $H(X)$ in terms of $H(X_0)$, $H(X_1)$, and p.
   b) Prove the inequality: $2^{H(X)} \leq 2^{H(X_0)} + 2^{H(X_1)}$.
   c) Find a sufficient and necessary condition for equality to hold in (b). The condition should be stated using $H(X_0)$, $H(X_1)$, and p only.
   d) Using (b), prove that $H(X) \leq \log|\mathcal{X}|$.

   Solution:
   a) Consider the following equalities:

$$\begin{aligned} H(X) &\overset{(a)}{=} H(X,\theta) \\ &= H(\theta) + H(X|\theta) \\ &= H_2(p) + (1-p)H(X_1) + pH(X_0), \end{aligned}$$

   where (a) follows from the fact that $H(\theta|X) = 0$.
   b) Note that $H(X)$ is a concave function in p. Therefore, we can take maximum on the parameter p in order to show the inequality. The first derivative is:

$$\frac{d}{dp}H(X) = \log\left(\frac{1-p}{p}\right) - H(X_1) + H(X_0).$$

   The maximizer can be found as $p^* = \frac{2^{H(X_1)}}{2^{H(X_0)}+2^{H(X_1)}}$ and substituting it back into $H(X)$ gives the desired inequality.
   c) Since $H(X)$ is a concave function in p, we know from the previous question that $p = p^*$ is a sufficient and necessary condition.
   d) For any random variable X on $\mathcal{X} = \{1, \ldots, n\}$ with $p(x)$, we should show that there exists X_0 and X_1 as given in the problem so we can use inequality (b). Define an a random variable, θ, which equals $\theta = 0$ if $X \in \{1, \ldots, n-1\}$ and

$\theta = 1$ if $X = n$. Now, define a random variable X_0 that has distribution $p_{X_0}(x) = p_{X|\theta}(x|\theta = 0)$ and X_1 in a similar way. Since $H(X_1) = 0$, we have that $2^{H(X)} \leq 2^{H(X_0)} + 1$ and we can repeat the same procedure for X_0 until we have $2^{H(X)} \leq n$ which gives the desired inequality.

4) **True or False on the concatenation order (10 Points)**: Given are channel A and channel B both have binary inputs and binary outputs. The channels are concatenated so the output of the channel A is the input to channel B and the capacity of this channel is denoted by $C_{A\to B}$. The definition of $C_{B\to A}$ is similar but channel B comes first.

   a) **True/False**: If channels A and B are binary symmetric channels, then $C_{A\to B} = C_{B\to A}$.
   b) **True/False**: For arbitrary binary channels, the order of the concatenation has no effect on the capacity.

   Solution:

   a) True. For two BSCs, we can write the output as $Y = (X \bigoplus Z_1) \bigoplus Z_2$. Since the XOR operator is commutative and associative we can also write this channel as $Y = (X \bigoplus Z_2) \bigoplus Z_1$ which is exactly the channel in the different order. Therefore, they have equal capacities.
   b) False. Let us denote the capacity of a general binary channel as $C(p_{0,0}, p_{1,1})$ where $p_{i,j}$ is $p(y = j|x = i)$. We take Z-channel with parameter α and S-channel with parameter β to construct our example. If the Z channel comes first then the capacity is $C(1-\beta, 1-\alpha\bar{\beta})$ and for the case where the S channel comes first is $C(1-\bar{\alpha}\beta, 1-\alpha)$ which is also equals $C(\bar{\alpha}\beta, \alpha)$ (why??). By taking $\beta \geq 0.5$ and α that satisfy $1-\beta = \bar{\alpha}\beta$ we have that

$$C_{A\to B} = C(1-\beta, 1-\alpha\bar{\beta})$$
$$C_{B\to A} = C(1-\beta, \alpha).$$

   All left is to calculate $C(\alpha, \beta) = \max_p H_2(\bar{\alpha}p + \beta\bar{p}) - pH_2(\alpha) - \bar{p}H_2(\beta)$. By taking the first derivative, we have $(\bar{\alpha} - \beta)\log(\frac{1-z}{z}) - H_2(\alpha) + H_2(\beta) = 0$, where $z = \bar{\alpha}p + \beta\bar{p}$. Arranging the equation gives $z^* = 1/(1 + 2^{\frac{H(\beta)-H(\alpha)}{\bar{\alpha}-\beta}})$. Now, any choice of $\alpha\beta$, for example, $\alpha = \frac{1}{2}$, $\beta = \frac{2}{3}$ gives $C_{A\to B} \neq C_{B\to A}$.

5) **Huffman Code (31 Points)** : Let X^n be a an i.i.d. source that is distributed according to p_X:

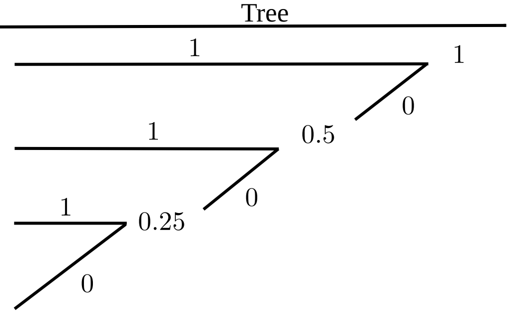
x	0	1	2	3
$p_X(x)$	0.5	0.25	0.125	0.125

   a) What is the optimal lossless compression rate R^* for the source sequence? (4 points)
   Solutions:
   The optimal lossless compression rate is given by the entropy of X

$$\begin{aligned} R^* =& H(X) \\ =& -0.5 \cdot \log 0.5 - 0.25 \cdot \log 0.25 - 2 \cdot 0.125 \cdot \log 0.125 \\ =& 1.75 \end{aligned}$$

   b) Build a binary Huffman code for the source X. (4 points)
   Solutions:

$c(x)$	x	$p_X(x)$	Tree
1	0	0.5	
01	1	0.25	
001	2	0.125	
000	3	0.125	



   c) What is the expected length of the resulting compressed sequence. (4 points)
   Solutions:
   Denote the length of a codeword by $L(c(x_i))$. Then

$$\begin{aligned} L(c^n(x^n)) =& \sum_{i=1}^{n} L(c(x_i)) \\ =& \sum_{i=1}^{n} [p_X(0) \cdot L(c(0)) + p_X(1) \cdot L(c(1)) + p_X(2) \cdot L(c(2)) + p_X(3) \cdot L(c(3))] \\ =& n\,(0.5 \cdot 1 + 0.25 \cdot 2 + 0.125 \cdot 3 + 0.125 \cdot 3) \\ =& 1.75n \end{aligned}$$

The expected length of the sequence is $nR^* = 1.75n$.
Note that the distribution on X is dyadic, and therefore the Huffman code is optimal.
Therefore, $nR = nH(X)$.

d) What is the expected number of zeros in the resulting compressed sequence. (5 points)
Solutions:
Let $N(0|c)$ denote the number of zeros in a codeword c, and $c^n(x^n) = [c(x_1), c(x_2), \ldots, c(x_n)]$

$$\begin{aligned}
\mathbb{E}\left[N(0|c^n(X^n))\right] =& \mathbb{E}\left[\sum_{i=1}^{n} N(0|c(X_i))\right] \\
=& \sum_{i=1}^{n} \mathbb{E}\left[N(0|c(X_i))\right] \\
=& \sum_{i=1}^{n} \left[p_X(0)\cdot N(0|c(0)) + p_X(1)\cdot N(0|c(1)) + p_X(2)\cdot N(0|c(2)) + p_X(3)\cdot N(0|c(3))\right] \\
=& \sum_{i=1}^{n} \left[0.5\cdot 0 + 0.25\cdot 1 + 0.125\cdot 2 + 0.125\cdot 3\right] \\
=& 0.875n
\end{aligned}$$

Since the code is optimal, the number of zeros is half of the expected length (see the following sub-question).

e) Let $\tilde{X}^n$ be an another source distributed i.i.d. according to $p_{\tilde{X}}$.

$\tilde{x}$	0	1	2	3
$p_{\tilde{X}}(\tilde{x})$	0.3	0.4	0.1	0.2

What is the expected length of compressing the source $\tilde{X}$ using the code constructed in (b). (4 points)
Solutions:
Denote the length of a codeword by $L(c(x_i))$. Then

$$\begin{aligned}
L(c^n(x^n)) =& \sum_{i=1}^{n} L(c(x_i)) \\
=& \sum_{i=1}^{n} \left[p_X(0)\cdot L(c(0)) + p_X(1)\cdot L(c(1)) + p_X(2)\cdot L(c(2)) + p_X(3)\cdot L(c(3))\right] \\
=& n\left(0.3\cdot 1 + 0.4\cdot 2 + 0.1\cdot 3 + 0.2\cdot 3\right) \\
=& 2n
\end{aligned}$$

f) Answer (d) for the code constructed in (b) and the source $\tilde{X}^n$. (5 points)
Solutions:

$$\begin{aligned}
\mathbb{E}\left[N(0|c^n(\tilde{X}^n))\right] =& \mathbb{E}\left[\sum_{i=1}^{n} N(0|c(\tilde{X}_i))\right] \\
=& \sum_{i=1}^{n} \mathbb{E}\left[N(0|c(\tilde{X}_i))\right] \\
=& \sum_{i=1}^{n} \left[p_{\tilde{X}}(0)\cdot N(0|c(0)) + p_{\tilde{X}}(1)\cdot N(0|c(1)) + p_{\tilde{X}}(2)\cdot N(0|c(2)) + p_{\tilde{X}}(3)\cdot N(0|c(3))\right] \\
=& \sum_{i=1}^{n} \left[0.3\cdot 0 + 0.4\cdot 1 + 0.1\cdot 2 + 0.2\cdot 3\right] \\
=& 1.2n
\end{aligned}$$

Note that the expected number of zeros is not half of the expected length. It implies that the code is not optimal.

$$R^* = H(\tilde{X}) = 1.846$$

g) Is the relative portion of zeros (the quantity in (d) divided by the quantity in (c)) after compressing the source X^n and the source $\tilde{X}^n$ different? For both sources, explain why there is or there is not a difference. (5 points)
Solutions:
For X^n we used optimal code with varying length. Therefore, the expected number of zeros is half of the compressed sequence. However, ee used a code that is not optimal for $\tilde{X}^n$. Henceforth, the compression rate is not optimal, and the expected number of zeros is not necessarily half of the expected length. Note that the expected length is not optimal too,

since $H(\tilde{X}) \approxeq 1.8464$, which is not equal to $\frac{\mathbb{E}\left[L(c(\tilde{X^n}))\right]}{n}$.

Good Luck!