ECE 534: Elements of Information Theory

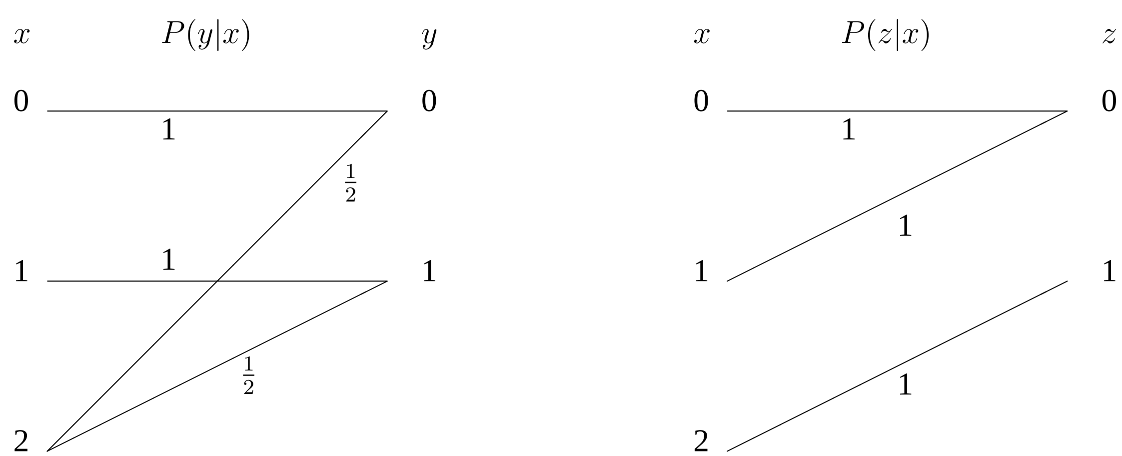
Solutions to Midterm Exam (Spring 2006)

Problem 1 [20 pts.]

A discrete memoryless source has an alphabet of three letters, $x_i, i = 1, 2, 3$, with probabilities 0.4, 0.4, and 0.2, respectively.
(a) Find the binary Huffman code for this source and determine the average number of bits needed for each source letter.
(b) Suppose two letters at a time are encoded into a binary sequence. Find the Huffman code and the average number of bits needed per source letter.
Solution:
(a) One possible Huffman code is: $C(1) = 0$, $C(2) = 10$, and $C(3) = 11$. The average number of bits per source letter is $0.4 + (0.4 + 0.2) \times 2 = 1.6$.
(b) One possible code construction is: $C(11) = 000$, $C(12) = 001$, $C(13) = 100$, $C(21) = 010$, $C(22) = 011$, $C(23) = 110$, $C(31) = 111$, $C(32) = 1010$, and $C(33) = 1011$. The average number of bits per source letter is $[3 + (0.08 + 0.04)]/2 = 1.56$.

Problem 2 [20 pts.]

A source X produces letters from a three-symbol alphabet with the probability assignment $P_X(0) = 1/4$, $P_X(1) = 1/4$, and $P_X(2) = 1/2$. Each source letter x is transmitted through two channels simultaneously with outputs y and z and the transition probabilities indicated below:



Calculate $H(X)$, $H(Y)$, $H(Z)$, $H(Y, Z)$, $I(X;Y)$, $I(X;Z)$, $I(X;Y|Z)$, and $I(X;Y,Z)$.
Solution:

$$H(X) = \frac{1}{4}\log_2 4 + \frac{1}{4}\log_2 4 + \frac{1}{2}\log_2 2 = 1.5$$

The probability distribution of Y is $P_Y(0) = \frac{1}{4} + \frac{1}{2}\frac{1}{2} = \frac{1}{2}$. Therefore,

$$H(Y) = 1$$

Similarly, $P_Z(0) = P_Z(1) = \frac{1}{2}$ and

$$H(Z) = 1$$

If $Z = 1$, then $X = 2$, and $Y = 0, 1$ with equal probability; if $Z = 0$, then $X = 0, 1$ with equal probability and as a consequence, $Y = 0, 1$ with equal probability. Therefore,

$$H(Y|Z) = P_Z(0)H(Y|Z = 0) + P_Z(1)H(Y|Z = 1) = 1$$

and

$$H(Y, Z) = H(Z) + H(Y|Z) = 1 + 1 = 2$$

Since $H(Y|X = 0) = H(Y|X = 1) = 0$ and $H(Y|X = 2) = 1$, we have

$$I(X; Y) = H(Y) - H(Y|X) = 1 - \frac{1}{2} = 0.5$$

Since $H(Z|X) = 0$, we have

$$I(X; Z) = H(Z) - H(Z|X) = H(Z) = 1$$

Since Z is completely determined by X, $H(Y|X, Z) = H(Y|X) = 0.5$. We have

$$I(X; Y|Z) = H(Y|Z) - H(Y|X, Z) = 1 - 0.5 = 0.5$$

and

$$I(X; Y, Z) = I(X; Z) + I(X; Y|Z) = 1 + 0.5 = 1.5$$

Problem 3 [30 pts.]
a) Prove that the number of elements in the δ-typical set $A(n, \delta)$ satisfies $|A(n, \delta)| \leq 2^{nH(X)+n\delta}$.
b) Prove that $|A(n, \delta)| \geq (1 - \delta)2^{nH(X)-n\delta}$ for sufficiently large n.
c) Prove that the expected length L of a D-ary prefix code for a random variable X satisfies $L \log D \geq H(X)$. (Hint: use Kraft's Inequality $\sum_x D^{-l_x} \leq 1$.)
Solution:
a) $A(n, \delta)$ contains the set of sequences $(x_1, \ldots, x_n)$, such that $|\frac{1}{n}\sum_{i=1}^{n} \log \frac{1}{P_X(x_i)} - H(X)| \leq \delta$. Therefore, every sequence in $A(n, \delta)$ satisfies:

$$2^{-nH(X)-n\delta} \leq P(x_1, \ldots, x_n) \leq 2^{-nH(X)+n\delta} \tag{1}$$

Using the lower bound in (1), we have

$$1 \geq P[A(n, \delta)] = \sum_{(x_1,\ldots,x_n)\in A(n,\delta)} P(x_1, \ldots, x_n) \geq |A(n, \delta)|2^{-nH(X)-n\delta}$$

and hence

$$|A(n, \delta)| \leq 2^{nH(X)+n\delta}$$

b) We have from the law of large numbers that $P[A(n,\delta)] \geq 1-\delta$ for all sufficiently large n. Using the upper bound in (1), we have for all sufficiently large n,

$$1-\delta \leq P[A(n,\delta)] = \sum_{(x_1,\ldots,x_n)\in A(n,\delta)} P(x_1,\ldots,x_n) \leq |A(n,\delta)| 2^{-nH(X)+n\delta}$$

and hence

$$|A(n,\delta)| \geq (1-\delta)2^{nH(X)-n\delta}$$

c) Denoting the length of the codeword for $a \in \mathcal{A}$ as l_a, we have

$$\begin{aligned} L\log D - H(X) &= \sum_{a\in\mathcal{A}} P_X(a) l_a \log D - \sum_{a\in\mathcal{A}} P_X(a)\log\frac{1}{P_X(a)} \\ &= -\sum_{a\in\mathcal{A}} P_X(a)\log\frac{D^{-l_a}}{P_X(a)} \\ &\geq -(\log e)\sum_{a\in\mathcal{A}} P_X(a)\left(\frac{D^{-l_a}}{P_X(a)} - 1\right) \\ &= -(\log e)\left(\sum_a D^{-l_a} - 1\right) \\ &\geq 0 \end{aligned}$$

where the first inequality comes from $\ln x \leq x - 1$ and the last inequality comes from Kraft's inequality.

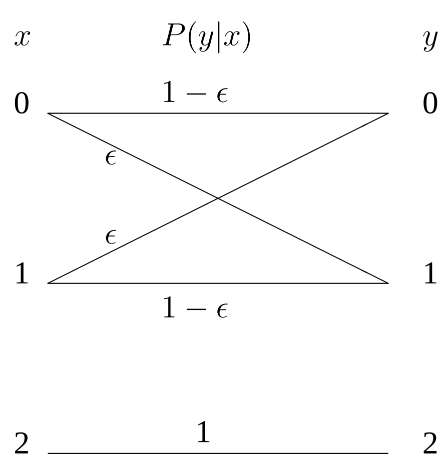
Problem 4 [30 pts.]

Consider n discrete memoryless channels with capacities $C_1, C_2, \ldots, C_n$, respectively. Both the input and the output alphabet sets of different channels are disjoint. We define the sum channel of these n channels as a channel that has all n channels available for use but only one channel may be used at any given time.

(a) Prove that the capacity of the sum channel is given by

$$C = \log_2 \sum_{i=1}^{n} 2^{C_i}$$

(b) Use the above result to find the capacity of the following channel:



Solution:

(a) Denote the input and output of the ith channel as X_i and Y_i. Denote the input and output alphabets of the ith channel as $\mathcal{A}_i$ and $\mathcal{B}_i$, respectively. Denote the input and output of the sum channel as X and Y. Suppose the ith channel is used with probability p_i, and the input distribution for the ith channel is P_{X_i}. The input distribution of the sum channel is:

$$P_X(x) = \begin{cases} p_i P_{X_i}(x) & x \in \mathcal{A}_i, \quad i = 1, \dots, n \\ 0 & \text{otherwise} \end{cases}.$$

The output distribution is therefore

$$P_Y(y) = \begin{cases} p_i \sum_{x \in \mathcal{A}_i} P_{X_i}(x) P_{Y_i|X_i}(y|x) & y \in \mathcal{B}_i, \quad i = 1, \dots, n \\ 0 & \text{otherwise} \end{cases}.$$

Since $P_{Y_i}(y) = \sum_{x \in \mathcal{A}_i} P_{X_i}(x) P_{Y_i|X_i}(y|x)$, we have

$$\begin{aligned} H(Y) &= \sum_y P_Y(y) \log \frac{1}{P_Y(y)} \\ &= \sum_i \sum_{y \in \mathcal{B}_i} p_i P_{Y_i}(y) \log \frac{1}{p_i P_{Y_i}(y)} \\ &= \sum_i p_i \log \frac{1}{p_i} + \sum_i p_i H(Y_i). \end{aligned}$$

The conditional entropy is

$$\begin{aligned} H(Y|X) &= \sum_x P_X(x) H(Y|X=x) \\ &= \sum_i \sum_{x \in \mathcal{A}_i} p_i P_{X_i}(x) H(Y|X=x) \\ &= \sum_i p_i H(Y_i|X_i). \end{aligned}$$

The mutual information can then be obtained as

$$\begin{aligned} I(X;Y) &= H(Y) - H(Y|X) \\ &= \sum_i p_i \log \frac{1}{p_i} + \sum_i p_i I(X_i;Y_i). \end{aligned}$$

To maximize the mutual information, we need to choose P_{X_i} such that the capacity of the ith channel is achieved, leading to

$$\begin{aligned} C &= \sup_{P_X} I(X;Y) \\ &= \sup_{(p_1,\dots,p_n)} \left[\sum_i p_i \log \frac{1}{p_i} + \sum_i p_i C_i \right] \end{aligned}$$

The Lagrangian of this optimization problem is

$$L(p_1, \dots, p_n, \lambda) = \sum_i p_i \log \frac{1}{p_i} + \sum_i p_i C_i + \lambda(\sum_i p_i - 1)$$

Solving the following equations:

$$\frac{\partial L}{\partial p_i} = \log \frac{1}{p_i} - \log e + C_i + \lambda = 0, \qquad i = 1, \ldots, n$$
$$\frac{\partial L}{\partial \lambda} = \sum_i p_i - 1 = 0,$$

we obtain the optimal values of $(p_1, \ldots, p_n)$ as

$$p_i = \frac{2^{C_i}}{\sum_i 2^{C_i}}, \quad i = 1, \ldots, n.$$

The channel capacity is thus

$$C = \log \sum_i 2^{C_i}$$

(b) This channel can be decomposed into the sum of a binary symmetric channel with capacity $1 - h(\epsilon)$ and a channel with zero capacity. The capacity of the sum channel is thus

$$C = \log\left(1 + 2^{1-h(\epsilon)}\right)$$