EL606: Homework 5 – Solutions

Due date: Nov. 2, 2004

1. *Problem 12, Ch. 12, Cover&Thomas*
   We first parse the string, looking for strings that we have not seen before. Thus, the parsing yields 0,00,000,1,10,101,0000,01,1010,1. There are 10 phrases, and therefore we need 4 bits to represent the pointer to the prefix. Thus, using the scheme described in the text, we encode the string as (0000,0),(0001,0), (0010,0), (0000,1), (0100,0), (0101,1), (0011,0), (0001,1), (0110,0),(0000,1). (The last phrase, though it is not really a new phrase, is handled like a new phrase).

2. *Problem 1, Ch. 8, Cover&Thomas*

   (a) The statistician calculates $\tilde{Y} = g(Y)$. Since $X \to Y \to \tilde{Y}$ forms a Markov chain, we can apply the data processing inequality. Hence for every distribution on x,

   $$I(X;Y) \geq I(X;\tilde{Y}).$$

   Let $\tilde{p}(x)$ be the distribution on x that maximizes $I(X;\tilde{Y})$. Then

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

   Thus, the statistician is wrong and processing the output does not increase capacity.

   (b) We have equality (no decrease in capacity) in the above sequence of inequalities only if we have equality in data processing inequality, i.e., for the distribution that maximizes $I(X;\tilde{Y})$, we have $X \to \tilde{Y} \to Y$ forming a Markov chain.

3. *Problem 2, Ch. 8, Cover&Thomas*

   (a) By Bayes rule, for any events A and B,

   $$\Pr(A|B) = \frac{\Pr(A)\Pr(B|A)}{\Pr(B)}.$$

   In this case, we wish to calculate the conditional probability of a_1 given the channel output. Thus we take the event A to the event that the source produced a_1, and B to be the event corresponding to one of the 8 possible output sequences. Thus $\Pr(A) = 1/2$, and $\Pr(B|A) = \epsilon^i(1-\epsilon)^{3-i}$, where ϵ is the number of ones in the received sequence. $\Pr(B)$ can then be calculated as $\Pr(B) = \Pr(a_1)\Pr(B|a_1) + \Pr(B|a_2)$. Thus we can calculate

   $$\Pr(a_1|000) = \frac{\frac{1}{2}(1-\epsilon)^3}{\frac{1}{2}(1-\epsilon)^3 + \frac{1}{2}\epsilon^3}$$

   $$\Pr(a_1|100) = \Pr(a_1|010) = \Pr(a_1|001) = \frac{\frac{1}{2}(1-\epsilon)^2\epsilon}{\frac{1}{2}(1-\epsilon)^2\epsilon + \frac{1}{2}\epsilon^2(1-\epsilon)}$$

   $$\Pr(a_1|110) = \Pr(a_1|011) = \Pr(a_1|101) = \frac{\frac{1}{2}(1-\epsilon)\epsilon^2}{\frac{1}{2}(1-\epsilon)\epsilon^2 + \frac{1}{2}\epsilon(1-\epsilon)^2}$$

   $$\Pr(a_1|111) = \frac{\frac{1}{2}\epsilon^3}{\frac{1}{2}\epsilon^3 + \frac{1}{2}(1-\epsilon)^3}$$

(b) If $\epsilon < 1/2$, then the probability of a_1 given 000,001,010 or 100 is greater than 1/2, and the probability of a_2 given 110,011,101 or 111 is greater than 1/2. Therefore, the decoding rule above chooses the source symbol that has maximum probability given the observed output. This is the *maximum a posteriori* decoding rule, and is optimal in that it minimizes the probability of error. To see that this is true, let the input source symbol be X, let the output of the channel be denoted by Y and the decoded symbol be $\hat{X}(Y)$. Then

$$\begin{aligned}
\Pr(E) =& \Pr(X \neq \hat{X}) \\
=& \sum_y \Pr(Y = y) \Pr(X \neq \hat{X} | Y = y) \\
=& \sum_y \Pr(Y = y) \sum_{x \neq \hat{x}(y)} \Pr(x | Y = y) \\
=& \sum_y \Pr(Y = y) \left(1 - \Pr(\hat{x}(y) | Y = y)\right) \\
=& \sum_y \Pr(Y = y) - \sum_y Pr(Y = y) \Pr(\hat{x}(y) | Y = y) \\
=& 1 - \sum_y Pr(Y = y) \Pr(\hat{x}(y) | Y = y)
\end{aligned}$$

and thus to minimize the probability of error, we have to maximize the second term, which is maximized by choosing $\hat{x}(y)$ to the the symbol that maximizes the conditional probability of the source symbol given the output.

(c) The probability of error can also be expanded

$$\begin{aligned}
\Pr(E) =& \Pr(X \neq \hat{X}) \\
=& \sum_x \Pr(x) \Pr(\hat{X} \neq x) \\
=& \Pr(a_1) \Pr(Y = 011, 110, 101, \text{ or } 111) \\
& \quad + \Pr(a_2) \Pr(Y = 000, 001, 010 \text{ or } 100) \\
=& \frac{1}{2} \left(3\epsilon^2(1-\epsilon) + \epsilon^3\right) + \frac{1}{2} \left(3\epsilon^2(1-\epsilon) + \epsilon^3\right) \\
=& 3\epsilon^2(1-\epsilon) + \epsilon^3.
\end{aligned}$$

(d) By extending the same arguments, it is easy to see that the decoding rule that minimizes the probability of error is the maximum a posteriori decoding rule, which in this case is the same as the maximum likelihood decoding rule (since the two input symbols are equally likely). So we choose the source symbol that is most likely to have produced the given output. This corresponds to choosing a_1 if the number of 1's in the received sequence is n or less, and choosing a_2 otherwise. The probability of error is then equal to (by symmetry) the probability of error given that a_1 was sent, which is the probability that $n+1$ or more 0's have been changed to 1's by the channel. This probability is

$$\Pr(E) = \sum_{i=n+1}^{2n+1} \binom{2n+1}{i} \epsilon^i (1-\epsilon)^{2n+1-i}$$

This probability goes to 0 as $n \to \infty$, since this is the probability that the number of 1's is $n+1$ or more, and since the expected proportion of 1's is $n\epsilon < n+1$, by the weak law of large numbers the above probability goes to 0 as $n \to \infty$.

4. *Problem 5, Ch. 8, Cover&Thomas*

$$Y = X + Z (\text{mod } 11)$$

where

$$Z = \begin{cases} 1 & \text{with probability } 1/3 \\ 2 & \text{with probability } 1/3 \\ 3 & \text{with probability } 1/3 \end{cases}$$

In this case,

$$H(Y|X) = H(Z|X) = H(Z) = \log 3,$$

independent of the distribution of X, and hence the capacity of the channel is

$$\begin{aligned} C &= \max_{p(x)} I(X;Y) \\ &= \max_{p(x)} H(Y) - H(Y|X) \\ &= \max_{p(x)} H(Y) - \log 3 \\ &= \log 11 - \log 3, \end{aligned}$$

which is attained when Y has an uniform distribution, which occurs (by symmetry) when X has an uniform distribution.

(a) The capacity of the channel is $\log(11/3)$ bits/transmission.

(b) The capacity is achieved by an uniform distribution on the inputs. $\Pr(X = i) = 1/11$ for $i = 0, 1, \ldots, 10$.

5. *Problem 9, Ch. 8, Cover&Thomas*
First we express $I(X;Y)$, the mutual information between the input and output of the Z-channel, as a function of $p = \Pr(X = 1)$:

$$H(Y|X) = \Pr(X=0) \cdot \underbrace{H(Y|X=0)}_{=0} + \Pr(X=1) \cdot \underbrace{H(Y|X=1)}_{=1} = p$$

Note that $\Pr(Y = 1) = p/2$, and thus

$$\begin{aligned} H(Y) &= h(\Pr(Y=1)) = h(p/2) \\ I(X;Y) &= H(Y) - H(Y|X) = h(p/2) - p \end{aligned}$$

The maximum mutual information is obtained for some value of p such that $0 < p < 1$.

Using elementary calculus, we determine that

$$\frac{d}{dp} I(X;Y) = \frac{1}{2} \log_2 \frac{1 - p/2}{p/2} - 1\,,$$

which is equal to zero for $p = 2/5$. (It is reasonable that $\Pr(X = 1) < 1/2$ because $X = 1$ is the noisy input to the channel.) So the capacity of the Z-channel in bits is $H(1/5) - 2/5 = 0.722 - 0.4 = 0.322$.