EL606: Homework 6 – Solutions

Due date: Nov. 9, 2004

1. *Problem 3, Ch. 8, Cover&Thomas*
   The capacity of the channel

$$Y = X + Z \qquad X \in \{0, 1\}, \quad Z \in \{0, a\}$$

   depends on the value of a:

   (a) *Case* $a = 0$
   In this case, the channel is noiseless. Since its input is binary, the capacity is 1 bit per transmission.

   (b) *Case* $a \neq 0, \pm 1$
   In this case, when $X = 0$, Y is either 0 or a and when $X = 1$, Y is either 1 or $1 + a$. Although the channel is noisy, knowing Y, uniquely determines the X which was sent, and hence $H(X|Y) = 0$. Consequently, $\max I(X;Y) = \max H(X) = 1$, achieved for an uniform distribution on the input X.

   (c) *Case* $a = 1$
   In this case, when $X = 0$, Y is either 0 or 1 and when $X = 1$, Y is either 1 or 2. Consequently, the channel is identical to the binary erasure channel discussed in class, with the erasure probability $\epsilon = 1/2$. The capacity of this channel is $1 - \epsilon = 1/2$ bit per transmission.

   (d) *Case* $a = -1$
   We again have a binary erasure channel with the erasure probability $\epsilon = 1/2$.

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

$$Y_i = X_i \oplus Z_i,$$

   where

$$Z_i = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1 - p \end{cases}$$

   and Z_i are not independent.

$$\begin{aligned} I(X_1, X_2, \ldots, X_n; Y_1, Y_2, \ldots, Y_n) =& H(X_1, X_2, \ldots, X_n) - H(X_1, X_2, \ldots, X_n | Y_1, Y_2, \ldots, Y_n) \\ =& H(X_1, X_2, \ldots, X_n) - H(Z_1, Z_2, \ldots, Z_n | Y_1, Y_2, \ldots, Y_n) \\ \geq& H(X_1, X_2, \ldots, X_n) - H(Z_1, Z_2, \ldots, Z_n) \\ \geq& H(X_1, X_2, \ldots, X_n) - \sum H(Z_i) \\ =& H(X_1, X_2, \ldots, X_n) - nH(p) \\ =& n - nH(p), \end{aligned}$$

   with equality if $X_1, X_2, \ldots, X_n$ are chosen i.i.d. $\sim$ Bern($\frac{1}{2}$). The capacity of the channel

with memory over n uses of the channel is

$$\begin{aligned} nC^{(n)} =& \max_{p(x_1,x_2,\ldots,x_n)} I(X_1, X_2, \ldots, X_n; Y_1, Y_2, \ldots, Y_n) \\ \geq& I(X_1, X_2, \ldots, X_n; Y_1, Y_2, \ldots, Y_n)_{p(x_1,x_2,\ldots,x_n)=\text{Bern}(\frac{1}{2})} \\ \geq& n(1 - H(p)) \\ =& nC. \end{aligned}$$

Hence channels with memory have higher capacity. The intuitive explanation for this result is that the correlation between the noise decreases the effective noise; one could use the information from the past samples of the noise to combat the present noise.

3. *Problem 7, Ch. 8, Cover&Thomas*

   (a) If the typewriter prints out whatever key is struck, then the output, Y, is the same as the input, X, and
   $$C = \max I(X; Y) = \max H(X) = \log 26,$$
   attained by a uniform distribution over the letters.

   (b) In this case, the output is either equal to the input (with probability $\frac{1}{2}$) or equal to the next letter (with probability $\frac{1}{2}$). Hence $H(Y|X) = \log 2$ independent of the distribution of X, and hence
   $$C = \max I(X; Y) = \max H(Y) - \log 2 = \log 26 - \log 2 = \log 13,$$
   attained for a uniform distribution over the output, which in turn is attained by a uniform distribution on the input.

   (c) A simple zero error block length one code is the one that uses every alternate letter, say A,C,E,. . . ,W,Y. In this case, none of the codewords will be confused, since A will produce either A or B, C will produce C or D, etc. The rate of this code,
   $$R = \frac{\log(\#\text{ codewords})}{\text{Block length}} = \frac{\log 13}{1} = \log 13.$$
   In this case, we can achieve capacity with a simple code with zero error.

4. *Problem 12, Ch. 8, Cover&Thomas*
   We can use the same chain of inequalities as in the proof of the converse to the channel coding theorem. Hence

$$\begin{aligned} I(X^n; Y^n) =& H(Y^n) - H(Y^n|X^n) \\ =& H(Y^n) - \sum_{i=1}^{n} H(Y_i|Y_1, \ldots, Y_{i-1}, X^n) \\ =& H(Y^n) - \sum_{i=1}^{n} H(Y_i|X_i), \end{aligned}$$

since by the definition of the channel, Y_i depends only on X_i and is conditionally independent of everything else. Continuing the series of inequalities, we have

$$\begin{aligned} I(X^n; Y^n) =& H(Y^n) - \sum_{i=1}^{n} H(Y_i|X_i) \\ \leq& \sum_{i=1}^{n} H(Y_i) - \sum_{i=1}^{n} H(Y_i|X_i) \\ \leq& \sum_{i=1}^{n} (1 - h(p_i)), \end{aligned}$$

with equality if $X_1, X_2, \ldots, X_n$ is chosen i.i.d. $\sim$ Bern(1/2). Hence

$$\max_{p(\mathbf{x})} I(X_1, X_2, \ldots, X_n; Y_1, Y_2, \ldots, Y_n) = \sum_{i=1}^{n} (1 - h(p_i)).$$