CSC310 – Assignment #3

Due: Nov. 14, 2005, 9am at the **START** of class
Worth: 8%
Late assignments not accepted.

1 Mutual Information

Consider a source with an alphabet of four symbols which symbols with equal probability. We encode these four symbols using the codewords 000,011,101,110 and send those codewords over a binary symmetric channel with crossover probability f, receiving three bits at the output.

- Calculate the amount of mutual information (in bits) between the transmitted symbol and the first (leftmost) received bit.
- Calculate the additional amount of mutual information between the transmitted symbol and the second (middle) received bit, assuming you already know the first received bit.
- Calculate the additional amount of mutual information between the transmitted symbol and the third (rightmost) received bit, assuming you already know the first and second received bits.
- What value of f between 0 and 1 maximizes the additional information provided by the third bit?

2 Serial Combination of Channels

A *serial cascade* of channels is a system in which the output of one channel is fed as the input to the next channel in the cascade. Consider a cascade of N identical (but statistically independent) binary symmetric channels, each with crossover probability f.

- Show that the cascade is itself a binary symmetric channel. (Hint: use induction.)
- Find a closed form expression for the crossover probability of the cascade. (Hint: diagonalize the channel transition matrix.)
- Compute the capacity of the cascade, and show that it is strictly positive for any $f < 1/2$ and any N.
- If $f = 0.4$, what is the smallest N for which the capacity is driven below $1/1000$ bits?

3 Parallel Combination of Channels

A *parallel mixture* of two channels with disjoint input alphabets and disjoint output alphabets is constructed by making a master channel whose input alphabet is the union of the input alphabets of the two channels and whose output alphabet is the union of the output alphabets of the two channels. When an input symbol arrives, the master channel passes it to the channel whose input alphabet it belongs to; hence the channels are used "in parallel".

- If the two original channels have input alphabets of size $|A_1|$ and $|A_2|$, and output alphabets of size $|B_1|$ and $|B_2|$, what is the size of the effective input alphabet for the master channel? What is the size of the effective output alphabet for the master channel? How many possible input sequences of length N are there for the master channel? How many possible output sequences of length N?
- If the two original channels have capacities C_1 and C_2, compute the capacity of the parallel mixture.

4 A Simple Error Correcting Code

Consider a code for the N^{th} extension of a channel which has M codewords, each N bits long. We want this code to have the property that it can correct *any single bit corruption* when the N bit long codewords are transmitted across a noisy channel whose properties we do not know.

- What is the smallest permissible Hamming distance between any two codewords? (Recall that the Hamming distance between two codewords is the number of bits on which they differ.)
- If $N = 2^K - 1$, what is the maximum number of codewords we can have in such a code?
- With this maximum number of codewords, what is the rate of the code (again, assuming $N = 2^K - 1$).
- When $N = 7$, construct such a code (ie with the maximum possible rate for that N). List all the codewords. (Hint: try using all possible combinations of the first four bits and something special for the last three bits.)

5 The Data Processing Theorem

The *data processing theorem* states that the mutual information between the input and output of a channel can never be increased by further processing of the output. To prove this theorem, consider a channel with input alphabet A and output alphabet B. Next, consider an arbitrary (instantaneous) data processing step with input alphabet B and output alphabet C. We model the data processor as though it were another channel: for each possible input in the alphabet B it has some probability of outputting each of the symbols in C. (If the data processor is deterministic, these probabilities might all be zero or one.)

- Show that the conditional entropy $H(A|C)$ is greater than or equal to the conditional entropy $H(A|B)$, by showing that their difference is equal to:

$$\sum_{b\in B}\sum_{c\in C}\left[p(b,c)\sum_{a\in A}p(a|b)\left(\log p(a|b)-\log p(a|c)\right)\right]=\sum_{b\in B}\sum_{c\in C}\left(p(b,c)KL[p(a|b)||p(a|c)]\right)$$

  Above, $KL[p||q]$ is the Kullback-Leibler Divergence or cross-entropy between two distributions p and q which is always non-negative. This makes it easy to show that the expression above is non-negative.

- Use the above result to prove the Data Processing Theorem by showing that, for any channel, any input distribution and any data processor the mutual information $I(A;C)$ is always less than or equal to the mutual information $I(A;B)$.