Information Theory: Principles and Applications
Homework 4 - Due: April 30, 2010

1. Fano's inequality gives a lower bound on the probability of error in terms of the conditional entropy. Now, we will derive an *upper* bound on the error probability of an *maximum a posteriori probability* (MAP) decoder in terms of the conditional entropy.

   Let X and Y be random variables with joint distribution $p_{XY}(x, y)$. Suppose X is the transmitted message and Y is the received symbol. Let $\widehat{X}$ be the MAP estimate of X obtained from Y.

   Given that $Y = y$ is received, the MAP decoder will decide on a $\widehat{x}$ for which

   $$P(X = \widehat{x}|Y = y) \geq P(X = x|Y = y) \quad \text{for all } x$$

   (a) Show that when $Y = y$ the MAP decoder makes an error with probability

   $$P(X \neq \widehat{X}|Y = y) = 1 - \max_x p_{X|Y=y}(x)$$

   (b) Show that

   $$P(X \neq \widehat{X}) = \sum_{y \in \mathcal{Y}} [1 - \max_x p_{X|Y=y}(x)] p_Y(y)$$

   (c) Show that for any random variable U

   $$\begin{aligned} H(U) &\geq \sum_{u \in \mathcal{U}} p_U(u)[1 - p_U(u)](\log e) \\ &\geq [1 - \max_u p_U(u)](\log e) \end{aligned}$$

   (d) Show that

   $$H(X|Y = y) \geq [1 - \max_x p_{X|Y=y}(x)](\log e)$$

   (e) Show that

   $$P(\widehat{X} \neq X)(\log e) \leq H(X|Y)$$

2. Decoding rules

   (a) Suppose that the integers from 1 to M are encoded into channel inputs x_1 to x_M. Let y be the channel output and let the transition probabilities $P(y|x = x_m)$ for all $m = 1, \ldots, M$ be given. If the cost of decoding a transmitted message m as m' is given by $C(m, m')$ and a probability distribution $p_X(x)$ is given on the (input) integers from 1 to M, derive the decoding rule that yields minimum average cost.

   (b) Show that using $C(m, m') = 1$ if $m \neq m'$ and $C(m, m') = 0$ if $m = m'$, minimizing the average cost is the same as minimizing the probability of error and that scheme corresponds to the MAP decoder.

3. A discrete memoryless channel is characterized by the matrix

$$P(Y|X) = \begin{bmatrix} 1/2 & 1/3 & 1/6 \\ 1/6 & 1/2 & 1/3 \\ 1/3 & 1/6 & 1/2 \end{bmatrix}$$

   where the (i, j) entry gives the probability of $Y = y_j$ given that $X = x_i$. If $P(X = x_1) = 1/2$ and $P(X = x_2) = P(X = x_3) = 1/4$, find the decision scheme that minimizes the probability of error and calculate the corresponding probability of error for that scheme.

4. *Hamming distance is indeed a distance.* Prove the following properties from the Hamming distance

   (a) $d_H(\mathbf{x}, \mathbf{y}) \geq 0$, with equality if and only if $\mathbf{x} = \mathbf{y}$.

   (b) $d_H(\mathbf{x}, \mathbf{y}) = d_H(\mathbf{y}, \mathbf{x})$.

   (c) $d_H(\mathbf{x}, \mathbf{z}) \leq d_H(\mathbf{x}, \mathbf{y}) + d_H(\mathbf{y}, \mathbf{z})$. (triangular inequality)

5. The generator matrix of a linear binary block code is

$$\mathbf{G} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}$$

   (a) Find the corresponding parity check matrix for this code.

   (b) Find the codeword corresponding to the message $\mathbf{m} = (1\ 0\ 1\ 0)$?

   (c) Construct the syndrome table for this code.

(d) If the received word is $\mathbf{r} = (1\ 1\ 1\ 0\ 0\ 0\ 1)$, find via syndrome decoding the codeword selected by the decoder and the corresponding message at the decoder output.

6. Consider a binary block code with M codewords and blocklength n. That is, each codeword is sequence of n bits. Suppose this code can correct up to (and including) t errors.

   (a) For a codeword $\mathbf{c}$ consider the set $\mathcal{S}$ of all binary sequences for which the decoder decides $\mathbf{c}$. Show that

   $$|\mathcal{S}| \geq \sum_{j=0}^{t} \binom{n}{j}$$

   (b) *Sphere Packing Bound:* Show that the number of codewords M satisfies

   $$M \leq \frac{2^n}{\sum_{j=0}^{t} \binom{n}{j}}$$

   (c) Hamming codes can correct up to 1 error, and for blocklength $2^m - 1$ they contain $2^{2^m - m - 1}$ codewords. Show that Hamming codes satisfy the sphere packing bound with equality. Thus, for these blocklengths they are the highest rate single error correcting codes.

   (d) We want binary codes with length $n = 255$ and capable of correcting $t = 1, 2, 3$ errors. According to the sphere packing bound, what is the mininum number of parity bits needed for each value of t?

7. Consider the following method of constructing a binary block code of minimum distance d and block length n:

   - Start with the list of all 2^n binary sequences as potential codewords, and an empty list of chosen codewords.
   - While the list of potential codewords is not empty, pick any of its members and add it to the list of chosen codewords. Remove it and all sequences that are distance less than or equal to $d - 1$ from it from the list of potential codewords.
   - The constructed code is the set of chosen codewords.

(a) Argue that this procedure will yield a code of minimum distance at least d.

(b) *Gilbert-Varshamov Bound:* Show that at each iteration of the second step listed above, one codeword is added to the list of chosen codewords, and at most $\sum_{i=0}^{d-1} \binom{n}{i}$ sequences are deleted from the list of potential codewords. Conclude that the number of chosen codewords M satisfies

$$M \geq \frac{2^n}{\sum_{j=0}^{d-1} \binom{n}{j}}$$