Practice Problems for Midterm

1. For an i.i.d. source with probabilities {0.3, 0.25, 0.15, 0.1, 0.1, 0.05 and 0.05} find a binary Huffman code and calculate its average length. Evaluate the efficiency of the code. Repeat this problem for a ternary Huffman code.

2. If one were to encode the third extension of the source described in Problem 1 using a Huffman code with quaternary code symbols, how many dummy symbols would be used in constructing the Huffman code?

3. Assume the same source as described in Problem 1 but now we want to encode the source using a Tunstall code where each code source phrase is encoded into 4 binary digits. Find such a code and evaluate the average number of code symbols per source symbol. What is the efficiency of such a code?

4. Assume that we use the source phrases found in Problem 3 but we now encode them using a ternary Huffman code. What is the efficiency of the resultant Tunstall-Huffman coding scheme?

5. Consider an i.i.d binary source, S_0, where the probability of a 0 is equal to p. If p is close to 1, the source will emit long strings of 0's, separated by 1's. A typical output sequence of this source would be: 0 0 1 0 0 0 0 1 0 1 0 1 1 0 0 0 0 1 0 1 ....

   Consider a new source S with symbols s_0, s_1, s_2, s_3, s_4, ... which emits the symbol s_i when the source S_0 emits a run of i 0's followed by a 1. That is for the above binary output of the source S_0, the source S emits the sequence: s_2 s_4 s_1 s_0 s_4 s_1 ...
   (a) Find the entropy (base 2) of the source S. Do not leave your answer in terms of an infinite series but rather evaluate the infinite sum. The answer should be a function of p.
   (b) Find $H(S)/H(S_0)$ where $H(S_0)$ (base 2) is the entropy of source S_0.
   (c) Calculate the average number of source symbols from source S_0 in a symbol from source S.
   Interpret your answer in terms of what you found in part (b).

6. Consider the following variable to fixed length coding scheme for the source S_0 described in the previous problem:

Source Output	Code Word
1	000
01	001
001	010
0001	011
00001	100
000001	101
0000001	110
0000000	111

   Calculate the average number of binary code digits per source digit. Compare your answer to H (S_0) (base 2) for p=0.9.

7. Consider a Lempel Ziv compression with a window size of 16. Assume that the encoder has already compressed the text

PETER PIPER PICKED A PECK OF PICKLED PEPPERS HOW MANY P

so that the last 16 letters are still in the window.

Assume that an ASCII symbol requires 7 binary digits and that we used the code introduced in class to encode the length of a match.. How many binary digits will result when compressing the remainder of the text which is

ECKS OF PICKLED PEPPERS DID PETER PIPER PICK

More ...

Practice Problems for Midterm

1. Consider an i.i.d ternary source, where the probability of a 0 is equal to 0.6, the probability of a 1 is equal to 0.3 and the probability of 2 is equal to 0.1.
   - **(a)** Give a quaternary Huffman code (that is a code with symbols from the alphabet {0,1,2,3}) for each pair of source symbols. Put the code words in the table on the answer sheet.
   - **(b)** Calculate the average length of the Huffman code words.
   - **(c)** Find the efficiency of the Huffman code. That is, compute the ratio of the appropriate entropy to the average number of Huffman code symbols per source symbol.

2. Again consider the source S_0 described in problem 1:
   - **(a)** Find a Tunstall code for this source with 9 phrases, where each phrase is encoded into two ternary code symbols. Put the phrases on the answer sheet.
   - **(b)** For this code, compute the average number of ternary code symbols per ternary source symbol. (Hint. First find the average length of the source phrases.)
   - **(c)** Compute the efficiency of this code. That is, compute the ratio of the appropriate entropy to the average number of code symbols per source symbol.

3. Answer the following short questions. .

(a) **TRUE OR FALSE. EXPLAIN YOUR ANSWER.**

For any two discrete random variable X and Y, $H(X,Y) > H(X)$.

(b) **TRUE OR FALSE. EXPLAIN YOUR ANSWER.**

For any discrete random variable X, $H(X) \geq 0$.

(c) **TRUE OR FALSE. EXPLAIN YOUR ANSWER.**

Let X and Y be two binary random variables and let $Z = X \oplus Y$, where $\oplus$ means modulo 2 addition. Then $H(X.Z) = H(X) + H(Y)$.

(d) Assume that you are using the Lempel Ziv algorithm with a window size of 16. Assume that you have already have encoded the phrase "MARY HAD A LITTLE" and you are about to encode the remaining phrase " LAMB" where the first symbol to be encoded is a space. What do you transmit?

(e) Assume that you are using the Lempel Ziv algorithm with a window size of N. You find that it is better <u>not</u> to encode matches of a single letter but rather to encode it as a flag bit 0 (meaning "no match") followed by the 8 bit ASCII symbol for that letter.. What is the smallest value of N for which this is true? **EXPLAIN YOUR ANSWER.**