Practice Problems Midterm Examination

Instructions:	Do all problems.	Points are as indicated.	Open book, open notes.

Put answers on answer sheets provided.

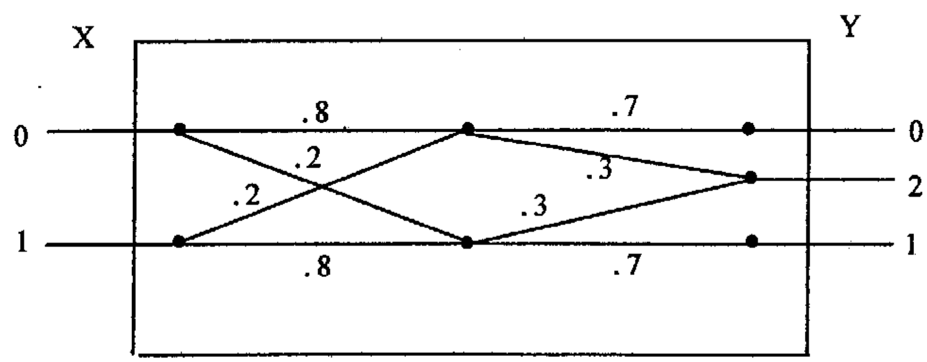
1. An i.i.d. source produces sequences with symbols from the set $\{0, 1, 2, 3\}$ with probabilities $\{.6, .2, .1, .1\}$, respectively.
   (a) Calculate the entropy of the source (base 3) *(5 points)*
   (b) Assume that pairs of symbols from the source are encoded using a Huffman code with code symbols $\{0, 1, 2\}$. Calculate the average length of the Huffman code words. *(15 points)*
   (c) Compare the average number of source symbols per code symbol with the entropy. *(5 points)*

2. For the source described in problem 1, assume that the source output is first parsed into 10 variable length phases using the Tunstall procedure and then these phases are encoded using a Huffman code with symbols from the alphabet $\{0, 1, 2\}$:
   (a) Calculate the average length of the Tunstall phases *(10 points)*
   (b) Calculate the average length of the Huffman code words *(10 points)*
   (c) Calculate the average number of source symbols per code symbol and compare your results with the entropy. *(5 points)*

3. Let a biased die be tossed until the number 6 occurs for the first time. On each toss, let the probability of the outcomes $\{1, 2, 3, 4, 5, 6\}$ be $\left\{\frac{1}{10}, \frac{2}{10}, \frac{3}{10}, \frac{1}{10}, \frac{1}{10}, \frac{2}{10}\right\}$, respectively. Let the tosses be statistically independent. Let Z be a r.v. indicating the possible outcomes of the die tosses and let N be a r.v. indicating the number of throws. Calculate (base 2)

E [N]	H (Z | N)	H (N)	H (Z)	I (Z ; N)	
(5 points)	*(15 points)*	*(5 points)*	*(3 points)*	*(2 points)*	*(total of 30 points)*

4. Consider the communications channel shown below where the channel is made up of the cascade of a binary symmetric channel and a binary erasure channel.



   (a) Prove that a uniform input distribution maximizes I(X ; Y) (you may use results proved in class). *(5 points)*
   (b) Calculate (base 2) C = Max I(X ; Y) *(10 points)*
   (c) Compare the capacity found in (b) with the capacities of the individual channels: i.e., the binary symmetric channel and the binary erasure channel. *(5 points)*

* * *

ECE 154C -- Spring 2009
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

8. The experiment is to throw a pair of fair dice, one at a time. Call the first value X and the second value Y. (X and Y are statistically independent and each take on the values {1, 2, ..., 6} with probabilities {1/6. 1/6, ..., 1/6}. Let S represent the real sum of X and Y, that is, let S=X+Y. Let Z represent the modulo two summation of X and Y, that is, let Z = (X + Y) modulo 2.

Compute the following entropies:

(a) H(X), H(Y), H(S), H(S), H(Z)

(b) H(X,Y), H(X,S), H(X,Z), H(S,Z)

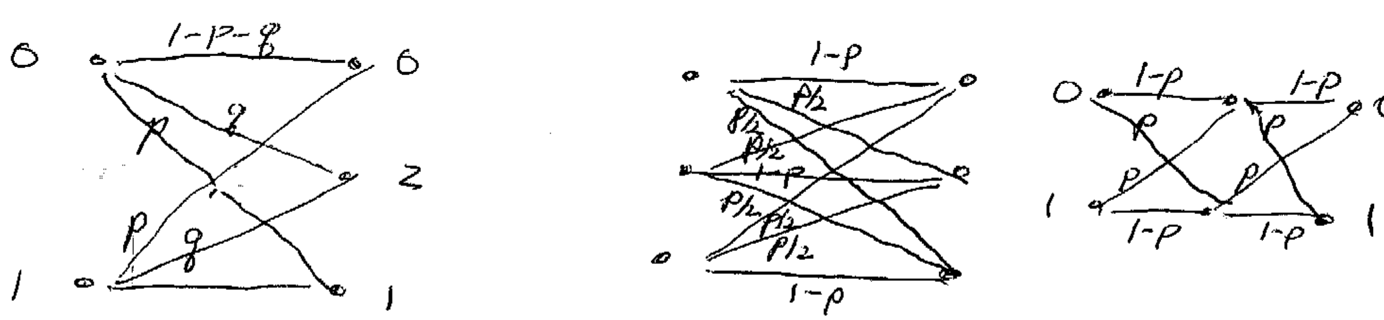
(c) H(X,Y,S), H(X,Y,Z), H(X,S,Z)

(d) H(X,Y,S,Z)

(e) H(X|S), H(X|Z), H(X,Y|S), H(X,Y|Z), H(S|Z)

---

9. Find the capacities of the following discrete memoryless channels



(a) (b) (c)