Homework Set 2 **EECS 651** **Due: Wed., Jan. 24, 2007**

Homework submission policy:

Homework is due in class on the listed date. However, as an extended deadline, homework may also be turned in to my office by 9 AM the following morning. If the door is closed, just push it under the door. Do not put it in my mailbox. No homework will be accepted after 9 AM. If you slide it under my door, write the date and time at the top of the paper.

Write your solutions neatly and submit them in the order of the problems.

Remember the collaboration policy discussed in the course information handout.

1. Do there exist prefix codes with the following sets of codeword lengths?

   (a) {2,2,3,3,3,5,6,6,6,6,7} .

   (b) {2,3,3,3,4,4,4,4,4}

   (c) {2,2,2,3,3}

   (d) For any set for which there does exist a code, draw the binary tree of a code with these lengths.

2. Find an example of a set of probabilites for which $R^* \geq H + .9$. Hint: a binary source will suffice. This shows that R^* can be very close to $H + 1$.

3. Show by example that a prefix code with lengths $l_i = \lceil - \log_2 P_i \rceil$ does not necessarily have minimum average length.

4. Show that if each probability in the set $\{P_1,...,P_M\}$ is a negative power of 2, then a Shannon code is an optimal prefix code.

5. Consider an IID source with the following set of probabilities:

   {.25,.2,.1,.1,.1,.1,.05,.05,.05}.

   (a) Find the entropy of the source.

   (b) Find two different prefix codes (first-order) with minimum rate. The codes should have different sets of lengths.

   (c) Compare the entropy and the rate of the codes found in (b). Do they differ by a "reasonable" amount?

   (d) Find a Shannon code for this set of probabilities.

   (e) Find the rate of the Shannon code and compare to the entropy and the rate of the minimum rate codes.

6. Suppose someone designs a prefix code (not necessarily optimal) for a set of probabilities $\{p_1,...,p_N\}$ and it has average length 3. What can be said about the entropy H of this set of probabilities? (Hint: Find the best upper and/or lower bounds to H that can be obtained with the given information.)

   (b) Repeat the problem, only now assume the prefix code is optimal.

7. Let X be a random variable with probability distribution $\{P_1,...,P_M\}$ and let Y be an independent random variable with probability distribution $\{Q_1,...,Q_M\}$. Let $W = (X,Y)$ be the random vector with NM outcomes. Show that $H(W) = H(X) + H(Y)$. Show this directly from first principles without using any other entropy relationships or formulas.

8. Simplified code for a large alphabet source: In this problem we study a source for which one can design a lossless code that is simpler than a conventional first-order variable-length code, and nearly as good. Consider a source with alphabet $A = \{1,2,\ldots,2^{16}\}$ and probabilities $\{P_1, P_2, \ldots, P_{2^{16}}\}$ such that for $i = 1,\ldots,4096$, $P_{16(i-1)+1} \cong P_{16(i-1)+2} \cong \ldots \cong P_{16(i-1)+16}$. One commonly used approach is the following: Each $x \in A$ can be uniquely represented as $x = 16(i-1) + j$, where $1 \le i \le 4096$ and $1 \le j \le 16$. The integer i is called the "class" of x and j is called its "index" within the class. The Huffman algorithm is used to design a code for the class i and a fixed-length code is used for the index j. This is one of the ideas used when designing the lossless coding part of JPEG.

   (a) Quantify the way in which this code is simpler than a conventional first-order variable-length code.

   (b) Show that this approach achieves a rate close to the entropy of the source.