Name: *Student ID:*

1/20	
2/20	
3/20	
4/20	
5/20	
T/100	

CSC 310 — Mid-term Test — 2004-02-27

For all questions, show enough of your work to indicate how you obtained your answer. No books or notes are allowed. You may use a calculator.
This test is 50 minutes in length.
The five questions are worth equal amounts.

1. You would like to encode a sequence of symbols that come from an alphabet with $d + 3$ symbols. You want to encode symbols a_1, a_2, and a_3 using codewords that are three bits long. You want to encode symbols $a_4, a_5, \ldots, a_{d+3}$ using codewords that are eight bits long. What is the maximum value of d for which this will be possible, if the code must be uniquely decodable?

2. Let C be a uniquely decodable binary code for an alphabet with I symbols. Assume that the probabilities of these symbols are all non-zero. Prove that if C is optimal (ie, has minimum expected codeword length), all codewords of C are at most I bits long.

3. [20 marks] Find a binary Huffman code for the source alphabet $\{a_1, a_2, a_3, a_4, a_5\}$, with symbol probabilities $p_1 = 0.11$, $p_2 = 0.05$, $p_3 = 0.10$, $p_4 = 0.70$, and $p_5 = 0.04$. Show this code as a table giving the codeword for every source symbol. Show your work.

4. This question concerns arithmetic coding for a binary source alphabet in which the symbol probabilities are $p_0 = 1/3$ and $p_1 = 2/3$. Assume the arithmetic encoder operates as described in the lectures, with the coding interval (initially $[0, 1)$) shrinking as each symbol is encoded, bits then being output if they are determined, and the coding interval being expanded as bits are output, or when it is entirely in the middle of the coding range. Fill in the table below to show how this procedure operates as the symbols 1, 1, and 0 are encoded. Assume that symbol 0 is allocated the low part of the coding region (nearer 0).

Input symbol	Interval after shrinking for this symbol	Bits output at this point (if any)	Interval after expansion
1			
1			
0			

5. Suppose we are compressing a source of symbols from a binary alphabet, $\{0, 1\}$, using arithmetic coding. We model the very first symbol by saying that 0 and 1 are equally likely. For the second and later symbols, we will use probabilities from a first-order Markov model.

   For the two questions below, consider the following sequence of symbols to be compressed:

   $$1 \ 1 \ 1 \ 1 \ 0$$

   a) Suppose that both the encoder and the decoder fix the probabilities of transitions in the Markov chain as follows:

   $$\begin{aligned} P(X_n = 1 \,|\, X_{n-1} = 0) &= 1/4 \\ P(X_n = 1 \,|\, X_{n-1} = 1) &= 7/8 \\ P(X_n = 0 \,|\, X_{n-1} = 0) &= 3/4 \\ P(X_n = 0 \,|\, X_{n-1} = 1) &= 1/8 \end{aligned}$$

   Approximately how many bits will the arithmetic coder output when compressing this sequence? You may write an arithmetic expression for this rather than an actual number.

   b) Suppose instead that the encoder and decoder don't know the probabilities for transitions, but estimate them adaptively using the Laplace scheme (adding one to the observed counts).

   Approximately how many bits will the arithmetic coder output when compressing this sequence? You may write an arithmetic expression for this rather than an actual number.