Quiz 1 COMP 423 Feb. 1, 2008

Answer all questions in the exam book. You may keep this exam sheet.

Calculators or other electronic devices are NOT permitted.

There are a total of 15 points.

1. **(2 points)**

   Consider an alphabet with four symbols and probabilities:

   $$p(A_1) = .35, \quad p(A_2) = .1, \quad p(A_3) = .15, \quad p(A_4) = .4$$

   (a) Construct a Huffman code such that, at each merge step, the child labeled 0 has probability less than or equal to the child labeled 1. Be sure to write the code words explicitly. A binary tree representation is not enough.

   (b) Give an expression of $\overline{\lambda} - H$ for this code, where $\overline{\lambda}$ is the average code length and H is the entropy.

2. **(3 points)**

   Consider an ASCII file with n bytes, which is parsed into strings from the set $\{S_1, S_2, \ldots, S_M\}$ and then encoded.

   (a) Assuming the strings S_i and their codewords $C(S_i)$ are known to both the encoder and decoder, give expressions for:

   - the number n of bytes of the original file,
   - the total number of bits of the encoded file

   which are based on the number n_i of occurences of each string S_i, and any other relevant parameters.

   (b) When initially choosing the codewords $C(S_i)$, should one consider the length of the strings S_i? Briefly explain.

3. **(3 points)**

In class, we defined Golomb codes by a constant b which was a power of 2.

(a) Give an expression for the code word lengths for $b = 2$, and draw the binary tree representation of the code for $i = 1, 2, ..10$.

(b) For what probability $p(i)$ would this code be an optimal prefix code?

(c) Suppose we were to define a Golomb code for $b = 3$. Noting b is not a power of 2, how could one design a Golomb code for this case? Draw a binary tree representation of your suggested code (for $i = 1, ...10$), and be sure to label the leaves.

Hint: You should assume that $p(i)$ is a decreasing function of i.

4. **(3 points)**

Consider the set of integers $\{1, 2, \ldots, N\}$ where N is finite, and a probability function

$$p(i) = \frac{a}{i}$$

on these integers, where a is a proportionality constant.

(a) Give an expression for a. A summation is sufficient.

(b) Suppose we wanted to choose a (near) optimal prefix code for this set of N integers. What should the codeword lengths λ_i be?

(c) Sketch out the binary tree representation of such a code. Assume N is large.

5. **(2 points)**

Use Jensen's inequality to derive an upper bound on:

$$\sum_{i=1}^{N}(i + \log(i + 1))$$

6. **(2 points)**

Use move-to-front to encode:

`Matt_ate_the_mat`

Give the sequence of codewords, assuming the move-to-front list starts empty.