Quiz 1 COMP 423 Feb. 8, 2006

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

Calculators or other electronic devices are NOT permitted.

The Quiz is marked out of 20.

1. **(4 points)**

   Consider an alphabet with four symbols and probabilities:

   $$p(A_1) = .4, \quad p(A_2) = .35, \quad p(A_3) = .2, \quad p(A_4) = .05$$

   (a) Give an expression of the entropy defined by the $p(A_i)$'s.

   (b) 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.

   (c) Give an expression of the average code length of this Huffman code.

   (d) Decode the bit string 100101111010

2. **(3 points)**

   (a) Give an example of a prefix code on $\{A_1, A_2, \ldots, A_6\}$ with codeword lengths 1,3,3,3,4,4 respectively.

   (b) For your code in (a), what probabilities $p(A_i)$ would make the average codeword length equal to the entropy?

   (c) Give an example of probabilities $p(A_i)$ for which your code is an optimal prefix code, and the average codeword length is greater than the entropy, *and* the probabilities *of each symbol* are different from (b).

3. **(5 points)**

   Draw the binary tree representations of:

   (a) a unary code

   (b) a Golomb code ($b = 4$),

   (c) an Elias code

   for the integers $\{1, 2, \ldots, 15\}$.

4. **(2 points)**

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

(a)

$$\sum_{i=N_1}^{N_2} \log \log i, \qquad \text{where } N_1 < N_2 \ .$$

<u>Hint:</u> use fact that

$$\sum_{i=N_1}^{N_2} i \;=\; \frac{(N_2 - N_1 + 1)(N_2 + N_1)}{2} \ .$$

(b)

$$\sum_{i=1}^{N} \sqrt{i}$$

<u>Hint:</u> Jensen's inequality holds for *any* convex function.

5. **(3 points)**

Consider Bernoulli trials with probability p_0 for the occurrence of a 0 and probability $1 - p_0$ for the occurrence of a 1.

(a) What is the entropy for a single Bernoulli trial?

(b) What is the entropy for run lengths?

<u>Hint:</u>

- A run length i is defined as in class, namely $i - 1$ 0's followed by a 1.
- Use the same definition of entropy as seen in class (but for $N = \infty$).
- Use the identities:

$$\sum_{k=0}^{\infty} x^k \;=\; \frac{1}{1-x} \qquad\qquad \sum_{k=1}^{\infty} k\ x^k \;=\; \frac{x}{(1-x)^2}$$

6. **(3 points)**

Consider a probability function

$$p(i) \;=\; i\ 2^{-(i+1)}$$

on the positive integers $i \in \{1, 2, 3, \ldots\}$.

Define a code $C(i)$ such that the average codeword length is as close to the the entropy as possible (given the time constraints of this exam). Be sure to specify $\lambda(i)$ for this code.