CSE 490 G

Assignment 2

Due Friday, January 20, 2006

1. Consider the alphabet $\{a, b, c, d, e, f\}$ with fixed prefix code

a	00
b	01
c	100
d	101
e	110
f	111

   Use the adaptive Huffman code to decode the following binary string.

   $$11010111010100000001101$$

   Show the decoder's tree after each symbol is decoded.

2. Consider a four symbol alphabet $\{a, b, c, d\}$ with probabilities $P(a) = .1, P(b) = .3, P(c) = .5, P(d) = .1$.

   (a) Use the greedy algorithm to construct a Tunstall code whose output codes have length 4. Be sure to reserve at least one output code as an "escape" code for inputs not coded by the Tunstall code.

   (b) Design a fixed code for the strings that require it.

   (c) Compute the average bit rate (in bits per symbol) for your code.

   (d) Compute the first-order entropy for this model. What is the percentage difference of the Tunstall code from entropy.

3. Suppose we are in the case where we we know the probability of 0 is much more than 1/2, but we don't know it precisely. The question is how to design an adaptive Golomb code that will work well. One way to do this is by doubling the order until the first 1 is found, then use the optimal Golomb code from that point on. This is called the *doubling algorithm*. For example, suppose the input is $0^{12}10^{10}10^{13}1$. The following table show how the coding proceeds:

order	input	output	calculation for next order
1	0	1	$2 = 2 \times 1$
2	00	1	$4 = 2 \times 2$
4	0000	1	$8 = 2 \times 4$
8	000001	0101	$9 = \lceil -1/\log_2(12/13) \rceil$
9	0^9	1	$15 = \lceil -1/\log_2(21/22) \rceil$
15	01	00010	$8 = \lceil -1/\log_2(22/24) \rceil$
8	0^8	1	$11 = \lceil -1/\log_2(30/32) \rceil$
11	000001	01010	$9 = \lceil -1/\log_2(35/38) \rceil$

Note that after coding $0^{12}1$ we have seen exactly one 1 out of 13 input symbols. This is why we switch to order 9. Similarly, after coding $0^{12}10^9$, we have seen exactly one 1 out of 22 total symbols so we switch to order 15, and so on.

(a) Code the string $0^8 10^{15} 10^8 1$ using the doubling algorithm. What is the compression ratio?

(b) Decode the string 111100110000101111 using the doubling algorithm. What is the compression ratio?

(c) Compute the γ-code for $0^8 10^{15} 10^8 1$ (the same as the γ-code for the sequence 8 15 8) and compare with Golomb code with doubling for the same string.

4. Consider the model with three symbols $\{a, b, c\}$ with probabilities $P(a) = 1/2$, $P(b) = 1/4$, and $P(c) = 1/4$. Assume an arithmetic coder with the partition of $[0, 1)$ with a first, b second, and c third.

   (a) Using the arithmetic coding algorithm to find the interval for the string *babc*. Compute the tag, short code and prefix code for this string.

   (b) Using arithmetic coding decode the string 010111 which encodes a string of length 4.