CSEP 521 - Spring 2005
Assignment 5

Due 5/5/05

1. We have seen that a Huffman code is a fixed length to variable length code and the Golomb code is a variable length to variable length code. There are some distinct advantages to variable length to fixed length codes. For example, if a bit is flipped in a a variable length to fixed length code then the decoder's error is limited. Tunstall coding is a popular method of creating a variable to fixed length code. As an example consider the code for a three letter alphabet $\{a, b, c\}$, with probabilities .8, .1, and .1, respectively.

aaa	000
aab	001
aac	010
ab	011
ac	100
b	101
c	110

   Notice that there are no codes for the strings a and aa that may terminate a string. These two special cases are handled by the escape code 111. The code 111 indicates the input string has terminated with either an a or an aa. In this case one more bit is needed to indicate which of the two it is. Examples are: $aaaabbc$ codes to 000 011 101 110 and $abaabaaaba$ codes to 011 001 000 101 111 0.

   (a) Compute the average bit rate for the Tunstall code above. You can assume a very long string whose length goes to infinity to ignore the escape code.

   (b) Compute the optimal Huffman code using two symbol batches. That is, use the "symbols" $\{aa, ab, ac, ba, bb, bc, ca, cb, cc\}$. Compute the average bit rate for this Huffman code ignoring input strings of odd length.

   (c) Compare your results from the Tunstall code and the Huffman code to entropy.

2. An alternative run length coder is called the γ-code. Recall that a run length coder really codes sequences of integers which are the the number of zeros between the ones in a binary string. In the γ-code, the integer $n \geq 0$ is coded by first writing $n+1$ in binary, then preceding it with m 0's where $m+1$ is the number of bits that were just written. The γ-codes are started in the table:

number	code
0	1
1	010
2	011
3	00100
4	00101
5	00110
6	00111
7	0001000
8	0001001
$\vdots$	$\vdots$

(a) Explain why the γ-code is uniquely decodable.

(b) Encode the binary string $0^5 10^8 10^{10} 1$ using the γ-code. This string is first transformed to three integers, then coded using the γ code. (Note 0^n is the string of n 0's.)

(c) Give an expression for the length of the γ-code of n as a function of n.

3. 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.