# CSE 490 GZ
# Assignment 4
Practice for the Midterm on February 8, 2002

1. Consider the probability distribution $a : 1/4, b : 1/2, c : 1/4$.

   (a) Use arithmetic coding with scaling to code the string *bbbbba*. Show the steps in the process and the value of $C$ which keeps track of the number of complementary bits to be output after a 0 or 1 is output. I chose this example because the scaled interval are very easy to calculate.

   (b) Use arithmetic decoding with scaling to decode 00000000001 (10 zeros followed by a 1) assuming the string decoded is of length 6.

2. Consider the string abracadabraabracadabra.

   (a) Run the Sequitur algorithm on this string to produce a context-free grammar. Show the steps along the way.

   (b) Encode this grammar in the basic encoding in the alphabet $\{a, b, c, d, r\}$.

3. In this example we explore Sequitur on some pathological strings.

   (a) Use Sequitur to find the grammar for $a^4$, $a^8$ $a^{16}$.

   (b) Generalize to give a grammar for $a^n$ for $n$ a power of 2.

   (c) Assuming a two letter alphabet compute the compression ratio for Sequitur, as a function of $n$, for strings of the form $a^n$ where $n$ is a power of 2.

4. In this example we explore LZ77 on pathological strings. For this problem we assume LZ77 (solution B), that is, the search buffer has size the length of the string and the look-ahead buffer has size 0.

   (a) Use LZ77 (solution B) to find the sequence of tuples for $a^4$, $a^8$ $a^{16}$.

   (b) Generalize to give the sequence of tuples for $a^n$ for $n$ a power of 2.

   (c) Assuming a two letter alphabet compute the compression ratio for LZ77 (solution B), as a function of $n$, for strings of the form $a^n$ where $n$ is a power of 2. Use the simple fixed length code for this.