CSC 310S
Spring, 2003

February 3, 2003
DUE: March 5 / 7, 2003

Information Theory
Assignment #2

1. Instantaneous Code Construction [20 marks]

Assume that a source produces independent symbols with relative frequencies:

(.41, .16, .15, .14, .08, .03, .02, .01)

(a). Devise *binary* (base 2) codes for this source using each of the three algorithms:

- Shannon Construction algorithm
- Shannon-Fano algorithm
- Huffman's algorithm

Then calculate and compare the average code word lengths and the relative redundancies of the three codes.

(b). For the same set of source symbol probabilities, do the same as in part (a), but this time producing *ternary* (base 3) codes by each of the three algorithms.

[NOTE: Produce the codes prescribed by the algorithms. Don't go on to do "obvious optimizations" to the resulting codes.]

2. Arithmetic Coding [20 marks]

Either by hand or using a program, carry out the tasks below involving *Arithmetic Coding*. (Show your intermediate work by writing down, or printing out from your program, the intermediate values obtained at each stage of the computations.) You may assume that the strings to be encoded and decoded require at most nine decimal places of accuracy in their arithmetic encoded representation (so regular floating point representations can be used; there is no need to employ shifting of bit strings).

Assume a source alphabet of four symbols (A, B, C, and D) with respective *independent* probabilities of occurrence (.4, .3, .2, and .1), associating A with the range [0.0,0.4), D with the range [0.9,1.0), etc.

(a). i. If the string to be encoded starts out with "AAAAA", what is the range of real numbers in which the arithmetic encoding of that string must fall?

ii. And what if the string to be encoded starts out with "DDDDD"?

(b). i. If the *binary represenation* of the number to be decoded starts out with "00000" in the first five places, how many initial characters of the decoded string are known with certainty?

ii. And what if the number to be decoded starts out with "11111"?

(c). Encode the string "ABACADABA" producing a real number in the interval [0,1).

(d). Decode the value .929576450 to produce a string of length nine composed of the four symbols.

3. Calculating Entropy Rates [20 marks]

Consider sources that emit the three symbols, X, Y, and Z. Such sources can be modeled either as *independent*, giving just the relative frequencies of the three symbols (p_X, p_Y, p_Z), or as *Markovian*, giving the first order transition probabilities between each symbol and the next one (p_{IJ} where I and J are any combination of X, Y, and Z).

For each of the four cases below, provide each of the following answers:

(a). What are the equilibrium frequencies with which each of the three symbols occurs, p_X, p_Y, and p_Z ?

(b). What is the *entropy* of the source based on the *independent model* ?

(c). What is the *entropy rate* of the source based on the transition probabilities of the *Markov model* of the source ?

Case (1):

from	to X	to Y	to Z
X	.333..	.333..	.333..
Y	.333..	.333..	.333..
Z	.333..	.333..	.333..

Case (2):

from	to X	to Y	to Z
X	.1	.9	.0
Y	.1	.0	.9
Z	.0	.1	.9

Case (3):

from	to X	to Y	to Z
X	.9	.1	.0
Y	.0	.9	.1
Z	.1	.0	.9

Case (4):

from	to X	to Y	to Z
X	.0	.5	.5
Y	.5	.0	.5
Z	.5	.5	.0