Tony Majestro

Problem Set 7

1. During class, we constructed a sentence by making guesses at the letters, recording the number of guesses for each letter. We went through the entire message; this generated a string of numbers. Given such a sequence of numbers, we can determine their frequencies and then compute the entropy of the resulting sequence of numbers. Ideally, the string and the text should have the same entropies.

Determine the frequencies of this string of guesses – for example, the number 6 appears 3 times among the 101 numbers, and so its frequency is 6/101 – and find the entropy of this string.

The following program determines the frequencies for the string of guesses and calculates the entropy:



This gives an entropy value of 1.432.

1. Construct a Huffman coding of the resulting frequencies and find its weighted average length.

After modifying the above program, it prints out each number and its frequency:



This gives the following output and the following Huffman Tree:

7: 0.010

8: 0.010

11: 0.010

15: 0.010

17: 0.010

4: 0.020

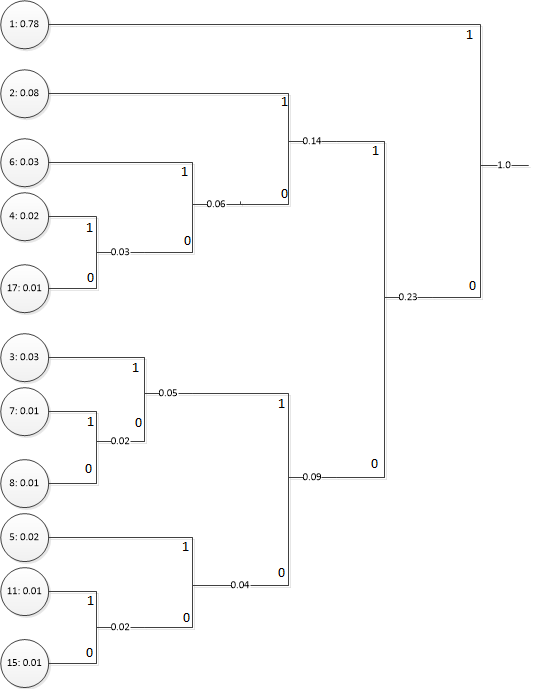
5: 0.020

3: 0.030

6: 0.030

2: 0.079

1: 0.772



This gives the following encodings:

**1: 1**

**2: 011**

**6: 0101**

**4: 01001**

**17: 01000**

**3: 0011**

**7: 00101**

**8: 00100**

**5: 0001**

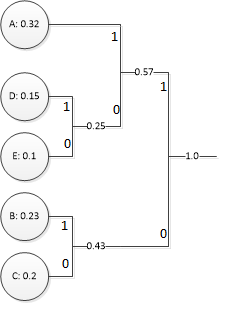
**11: 00001**

**15: 00000**

The weighted average string length is given by multiplying the frequency of each number times the length of each string:

1. Suppose X = {a, b, c, d, e} has the following probability distribution: P(A) = 0.32, P(B) = 0.23, P(C) = 0.2, P(D) = 0.15, and P(E) = 0.1. Use Huffman’s algorithm to find the optimal prefix-free encoding of X. Compare the length of this encoding to H(X).

Using Huffman’s algorithm, we get the following Huffman tree and encoding:

****

A: 11

D: 101

E: 100

B: 01

C: 00

The average length of the encoding is given by the following equation:

The entropy is given by the following equation: