Department of Electrical and Computer Engineering
The Johns Hopkins University
520.137 Introduction to Electrical and Computer Engineering – Fall 2017

Homework Assignment IX

Reading Assignment: Kuc Chapter 7

1. Consider the coding process of the following DNA data set $\{\ A, C, T, G, \text{Blank}\ \}$ whose symbol's frequency of occurrence is $150, 70, 60, 60$ and 50, respectively.

   (a) With the best fixed-length code, how many bits do you need to encode the entire data set?

   (b) Design the Shannon-Fano code. Construct the code table as well as the code tree. What is the average code-word length?

   (c) Design the Huffman code. Construct the code table as well as the code tree. What is the average code-word length?

   (d) What is the total number of bits required to encode the entire data set using Shannon-Fano code? What is the total number of bits required to encode the entire data set using Huffman code? How much saving can we achieved in each case compared to fixed-length code?

   (e) Now, consider the data set $\{\ 5, 2, 0, 1, 3, 7, \text{dummy}\ \}$ whose symbol's frequency of occurrence is $10, 20, 30, 40, 50, 60$, and 0, respectively.
   Design the Huffman code for a ternary number system – a base-3 numeral system using three digits (often called $trits$) $\{0, 1, 2\}$. Show your code table and your ternary Huffman code tree.

2. Consider the coding process with variable-length codes of the student-grade data $\{\ A, B, C, D, F\ \}$ whose source statistics are: $P[A] = 0.5, P[B] = 0.2, P[C] = 0.2, P[D] = 0.05$, and $P[F] = 0.05$.

   (a) What is the entropy of this source? How many bits do we need to represent each symbol using a fixed-length code?

   (b) Design the Shannon-Fano code. Construct the code table as well as the code tree. What is the average code-word length?

   (c) Design the Huffman code. Construct the code table as well as the code tree. What is the average code-word length?

   (d) The source statistics now change to: $P[A] = 0.75, P[B] = 0.08, P[C] = 0.07, P[D] = 0.05$, and $P[F] = 0.05$. Re-design the Shannon-Fano code. Construct the code table as well as the code tree. What is the average code-word length?

   (e) Re-design the Huffman code for the source above. Construct the code table as well as the code tree. What is the average code-word length?

   (f) What happens if we use the old Huffman code to encode the new data set in Part (d) and use the new Huffman code in Part (d) to encode the old data set? What are the resulting two average code-word lengths? Comment on your results.

3. Consider the encoding process with variable-length codes of the message

$$ELECTRICALENGINEERINGISGREAT$$

   (a) What is the estimated entropy of the message? How many bits do we need to represent each symbol using a fixed-length code?

   (b) Design the Shannon-Fano code. Construct the code table as well as the code tree.

   (c) Find the final encoded bit stream with your Shannon-Fano code. How many bits did you spend to encode the message?

   (d) Design the Huffman code. Construct the Huffman table as well as the Huffman tree.

   (e) Find the final encoded bit stream with your Huffman code. How many bits did you spend?

   (f) If there is a single bit error in the Huffman stream, will the error affect one letter in the decoding process or more? Justify your answer by an example.

4. Consider the encoding process with variable-length codes of the following DNA sequence

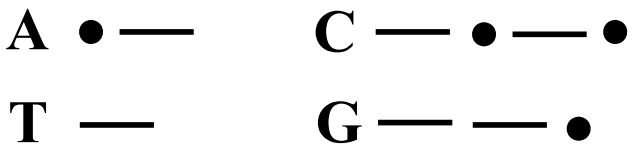
$$AAATCCGTAGCAAACA$$

   (a) What is the entropy of this DNA sequence?

   (b) Design the Shannon-Fano code. Construct the code table as well as the code tree. What is the average code-word length?

   (c) Design the Huffman code. Construct the code table as well as the code tree. What is the average code-word length?

   (d) The Morse code of the 4 symbols $ACTG$ is shown below. What is the average code-word length? Can you come up with the Morse code tree? What problem can you observe from the Morse tree?



   (e) Compare the efficiency of the three different VLC codes: Shannon-Fano, Huffman, and Morse in this case. What is the best code and how does it compare to the entropy?

Due date: **November 29** in class