Tutorial 4: Introduction to Information Theory

University of Windsor
Department of Electrical and Computer Engineering
ELEC 4190 - Digital Communications

Note: You should try the problems before and during the tutorial session. Your solutions will not be collected or graded.

1. An information source can be modeled as a bandlimited process with a bandwidth of 6000 Hz. This process is sampled at a rate higher than the Nyquist rate to provide a guard band of 2000 Hz. We observe that the resulting samples take values in the set: $\mathcal{A} = \{-4, -3, -1, 2, 4, 7\}$ with corresponding probabilities $\{0.2, 0.1, 0.15, 0.05, 0.3, 0.2\}$. What is the entropy of the discrete time source in bits per output (sample)? What is the information generated by this source in bits per second?

2. A discrete memoryless source has an alphabet $\{x_1, x_2, x_3, x_4, x_5, x_6\}$ with corresponding probabilities $\{0.3, 0.25, 0.2, 0.12, 0.08, 0.05\}$.

   (a) Find the entropy of this source.

   (b) What is the minimum required average codeword length to represent this source for error-free reconstruction?

   (c) Compare the entropy with the entropy of a uniformly distributed source with the same alphabet.

   (d) Design a Huffman code for the source and compare the average length of the Huffman code with the entropy of the source.

3. A discrete memoryless source has an alphabet $\{a_1, a_2, a_3\}$ with corresponding probabilities $\{0.1, 0.3, 0.6\}$.

   (a) Find the entropy of this source.

   (b) Design a Huffman code for the source. What is the average codeword length?

(c) Design a Huffman code for the second extension of the source (take two letters at a time). What is the average codeword length? What is the average number of required binary letters per each source output letter?

(d) Which is a more efficient coding scheme: the Huffman coding of the original source or the Huffman coding of the second extension of the source?

4. Design a ternary Huffman code, using 0, 1, and 2 as letters, for a source with output alphabet probabilities given by $\{0.05, 0.1, 0.15, 0.17, 0.18, 0.22, 0.13\}$. What is the resulting average codeword length? Compare the average codeword length with the entropy of the source. Note that, for a fair comparison of the average codeword length with the entropy of the source, you should compute the entropy of ternary code with logarithms in base 3.

5. Find the Lempel-Ziv source code for the source sequence below and recover the original sequence from the Lempel-Ziv source code. Use decimal numbers to represent the dictionary location.

```
BABBBBBABBBAABBABAAABBBABAABAAABABBABAAAABAABBAA
```

6. Two binary random variables X and Y are distributed according to the joint distributions $P(X = 0, Y = 0) = P(X = 0, Y = 1) = P(X = 1, Y = 1) = 1/3$. Compute $H(X)$, $H(Y)$, $H(X, Y)$, $H(X|Y)$, $H(Y|X)$, and $I(X; Y)$.

7. A discrete memoryless source produces eight outputs with probabilities 0.05, 0.07, 0.08, 0.1, 0.1, 0.15, 0.2, and 0.25. What is the minimum channel capacity required to transmit this source reliably? Can this source be reliably transmitted via a binary symmetric channel?

8. Determine the capacity of a telephone line with a bandwidth of 3.2 kHz. Assume AWGN with a signal-to-noise ratio of 35 dB.

9. Show that for a noiseless channel, $I(X; Y) = H(X) = H(Y)$, and determine the channel capacity per symbol.

10. For the channel shown below, find the channel capacity and the input distribution that achieves capacity.