# COMP2610/COMP6261 - Information Theory

Tutorial 8: Source Coding

Young Lee and Bob Williamson
**Tutors**: Debashish Chakraborty and Zakaria Mhammedi

Week 10 (9th – 13th Oct), Semester 2, 2017

1. *Optimal Coding and Huffman Codes*

   Consider the ensemble $X$ with probabilities $\mathcal{P}_X = \mathbf{p} = \{\frac{1}{2}, \frac{1}{4}, \frac{31}{128}, \frac{1}{128}\}$ and the code $C = \{0, 11, 100, 101\}$.

   (a) What is the entropy $H(X)$?
   (b) What is the expected length $L(C, X)$? Is $C$ an optimal code for $X$?
   (c) What are the code lengths for $X$? Construct a prefix Shannon code $C_S$ for $X$. Compute the expected code length $L(C_S, X)$.
   (d) What are the probabilities $\mathbf{q} = \{q_1, q_2, q_3, q_4\}$ for the code lengths of $C$?
   (e) Compute the relative entropy $D(\mathbf{p}\|\mathbf{q})$. What do you notice about $D(\mathbf{p}\|\mathbf{q})$, $H(X)$, $L(C, X)$, and $L(C_S, X)$?
   (f) Construct a Huffman code $C_H$ for $X$. How does its code lengths compare to $C$ and $C_S$? How do their expected code lengths compare?

2. *Binary Representations*

   Express the following numbers in binary.

   (a) $4.25_{10}$
   (b) $8.1_{10}$

3. *Shannon-Fano-Elias Coding*

   Let $X$ be an ensemble with alphabet $\mathcal{A}_X = \{x_1, x_2, x_3, x_4\}$ and probabilities $p_X = (0.25, 0.5, 0.125, 0.125)$.

   (a) Compute the cumulative distribution function $F(x_i)$ for each symbol $x_i$.
   (b) Compute the symbol intervals $[F(x_{i-1}), F(x_i))$ for each symbol $x_i$. (Recall that we assume for convenience that $F(x_0) = 0$.)
   (c) Compute the modified cumulative distribution function $\bar{F}(x_i)$ for each symbol $x_i$.
   (d) Compute the Shannon-Fano-Elias codewords for each symbol $x_i$.
   (e) Decode the string `10001`.
   (f) Suppose you only use the first $\lceil \log \frac{1}{p(x)} \rceil$ bits to code the above ensemble. Will the result be a prefix code?

4. *Arithmetic Coding*

   (a) For the same ensemble as the previous section, compute an arithmetic code for the sequence `c`. Assume here that $\mathtt{d} = \Box$, i.e. that the symbol `d` denotes the end of stream. Assume also that at each iteration, the probability of the various outcomes is unchanged.
   (b) Compute an arithmetic code for the sequence `ca`, with the same setup as the previous part. Does the codeword for `ca` start with that for `c`?
   (c) Compute an arithmetic code for the sequence `ca` using adaptive probabilities, assuming an initial set of virtual counts $\mathbf{m} = (1, 1, 1)$ and a constant end-of-stream probability $p(\mathtt{d}) = 0.25$.