SC222: Tutorial Sheet 12

Problems based on Information Theory

(1) *Huffman coding.* Consider the random variable

$$X = \begin{pmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 \\ 0.49 & 0.26 & 0.12 & 0.04 & 0.04 & 0.03 & 0.02 \end{pmatrix}$$

(a) Find a binary Huffman code for X.

(b) Find the expected code length for this encoding.

(c) Find a ternary Huffman code for X.

(2) *Bad codes.* Which of these codes cannot be Huffman codes for any probability assignment?

(a) {0, 10, 11}

(b) {00, 01, 10, 110}

(c) {01, 10}

(3) *Shannon codes and Huffman codes.* Consider a random variable X that takes on four values with probabilities $(\frac{1}{3}, \frac{1}{3}, \frac{1}{4}, \frac{1}{12})$.

(a) Construct a Huffman code for this random variable.

(b) Show that there exist two different sets of optimal lengths for the codewords. That is, show that codeword length assignments $(1, 2, 3, 3)$ and $(2, 2, 2, 2)$ are both optimal.

(c) Conclude that there are optimal codes with codeword lengths for some symbols that exceed the Shannon code length $\left\lceil \log \frac{1}{p(x)} \right\rceil$.

(4) *Huffman code.* Find the **(a)** *binary* and **(b)** *ternary* Huffman codes for the random variable X with probabilities

$$p = \left(\frac{1}{21}, \frac{2}{21}, \frac{3}{21}, \frac{4}{21}, \frac{5}{21}, \frac{6}{21}\right).$$

(c) Calculate $L = \sum p_i l_i$ in each case.

(5) *Bad wine.* One is given six bottles of wine. It is known that precisely one bottle has gone bad (tastes terrible). From inspection of the bottles it is determined that the probability p_i that the i^{th} bottle is bad is given by $(p_1, p_2, \ldots, p_6) = \left(\frac{8}{23}, \frac{6}{23}, \frac{4}{23}, \frac{2}{23}, \frac{2}{23}, \frac{1}{23}\right)$. Tasting will determine the bad wine. Suppose that you taste the wines one at a time. Choose the order of tasting to minimize the expected number of tastings required to determine the bad bottle. Remember, if the first five wines pass the test, you don't have to taste the last.

(a) What is the expected number of tastings required?

(b) Which bottle should be tasted first?

Now you get smart. For the first sample, you mix some of the wines in a fresh glass and sample the mixture. You proceed, mixing and tasting, stopping when the bad bottle has been determined.

(c) What is the minimum expected number of tastings required to determine the bad wine?

(d) What mixture should be tasted first?

(6) *Generating random variables.* One wishes to generate a random variable X

$$X = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1 - p\,. \end{cases}$$

You are given fair coin flips $Z_1, Z_2, \ldots$. Let N be the (random) number of flips needed to generate X. Find a good way to use $Z_1, Z_2, \ldots$ to generate X. Show that $E[N] \leq 2$.

(7) Channel capacity. Consider the discrete memoryless channel $Y = X + Z \pmod{11}$, where

$$Z = \begin{pmatrix} 1, & 2, & 3 \\ \frac{1}{3}, & \frac{1}{3}, & \frac{1}{3} \end{pmatrix}$$

and $X \in \{0, 1, \ldots, 10\}$. Assume that Z is independent of X.

(a) Find the capacity.

(b) What is the maximizing $p^*(x)$?

(8) *Z-channel.* The Z-channel has binary input and output alphabets and transition probabilities $p(y|x)$ given by the following matrix:

$$Q = \begin{bmatrix} 1 & 0 \\ 1/2 & 1/2 \end{bmatrix} \qquad x, y \in \{0, 1\}$$

Find the capacity of the Z-channel and the maximizing input probability distribution.

(9) *Unused symbols.* Show that the capacity of the channel with probability transition matrix

$$P_{y|x} = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & \frac{1}{3} & \frac{2}{3} \end{bmatrix}$$

is achieved by a distribution that places zero probability on one of input symbols. What is the capacity of this channel? Give an intuitive reason why that letter is not used.