Homework Set #5

1. **Conditional Entropy**
   Let $(X, Y) \sim p(x, y)$.
   (a) Express $H(X|X+Y)$ in terms of $H(X,Y)$ and $H(X+Y)$.
   (b) Suppose $H(X) > H(Y)$. Is $H(X|X+Y) > H(Y|X+Y)$?

2. **Can side information make a bad situation worse?**
   Suppose we have a horse race with outcome $X \in \{1, 2, ..., m\}$ and side information Y, where $(X, Y) \sim p(x, y) = p(x)p_0(y|x)$. The odds are m for 1.
   (a) Find the growth optimal strategy $b(x)$ and the associated growth rate of wealth $\max_{b(\cdot)} W(b(x), p(x))$ for the gambler.
   (b) Given side information, what is the growth optimal $b(x|y)$ and the associated growth rate? What is the improvement ΔW? Call it ΔW_p.
   Now suppose another gambler believes (incorrectly) that $X \sim q(x)$, and that $(X, Y) \sim q(x)p_0(y|x)$, i.e. he believes the joint distribution is $q(x, y) = q(x)p_0(y|x)$. Note that the conditional distribution $q(y|x) = p_0(y|x)$ is the same as in parts (a) and (b). Thus the noise in the observation of Y given X is the same in each version. Only the estimate of the true distribution of X is different.
   (c) The q gambler now gambles to maximize the growth rate as if $q(x)$ is true, without using side information Y. What is the growth rate W?
   (d) The q gambler is now given side information Y, still believing $q(x, y)$ is the true distribution. Find his optimal $b(x|y)$ and associated growth rate.
   (e) Now calculate ΔW for the q gambler (call it ΔW_q).
   (f) Express the difference $\Delta W_q - \Delta W_p$. Is ΔW_p or ΔW_q larger? This difference has a nice expression. Side information helps more when you are wrong than when you are right.

3. **Bad codes.**
Which of these codes cannot be Huffman codes for any probability assignment?

(a) $\{1, 01, 00\}$.

(b) $\{00, 01, 10, 110\}$.

(c) $\{01, 10\}$.

4. **Huffman coding.**
Consider the random variable

$$X = \begin{pmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 \\ 0.50 & 0.26 & 0.11 & 0.04 & 0.04 & 0.03 & 0.02 \end{pmatrix}$$

(a) Find a binary Huffman code for X.

(b) Find the expected codelength for this encoding.

(c) Find a ternary Huffman code for X.

5. **Codes.**
Let $X_1, X_2, \ldots,$ i.i.d. with

$$X = \begin{cases} 1, & \text{with probability } 1/2 \\ 2, & \text{with probability } 1/4 \\ 3, & \text{with probability } 1/4. \end{cases}$$

Consider the code assignment

$$C(x) = \begin{cases} 0, & \text{if } x = 1 \\ 01, & \text{if } x = 2 \\ 11, & \text{if } x = 3. \end{cases}$$

(a) Is this code nonsingular?

(b) Uniquely decodable?

(c) Instantaneous?

(d) What is the entropy rate of the process

$$Z_1 Z_2 Z_3 \ldots = C(X_1)C(X_2)C(X_3)\ldots?$$

For example, $x^n = 2311\ldots$ gives the encoded process $z^n = 011100\ldots$.

6. **Bad wine.**
One is given 6 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 $(p_1, p_2, \ldots, p_6) = (\frac{7}{26}, \frac{5}{26}, \frac{4}{26}, \frac{4}{26}, \frac{3}{26}, \frac{3}{26})$. Tasting will determine the bad wine.

   Suppose 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 5 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?

7. **Minimum cost codes.**
Words like Run! Help! and Fire! are short, not because they are frequently used, but perhaps because time is precious in the situations in which these words are required. Suppose that $X = i$ with probability $p_i, i = 1, 2, \ldots, m$. Let l_i be the number of binary symbols in the codeword associated with $X = i$, and let c_i denote the cost per letter of the codeword when $X = i$. Thus the average cost C of the description of X is $C = \sum_{i=1}^{m} p_i c_i l_i$.

   (a) Minimize C over all $l_1, l_2, \ldots, l_m$ such that $\sum 2^{-l_i} \leq 1$. Ignore any implied integer constraints on l_i. Exhibit the minimizing $l_1^*, l_2^*, \ldots, l_m^*$ and the associated minimum value C^*.

   (b) How would you use the Huffman code procedure to minimize C over all uniquely decodable codes? Let $C_{Huffman}$ denote this minimum.

   (c) Show that

$$C^* \leq C_{Huffman} \leq C^* + \sum_{i=1}^{m} p_i c_i.$$

8. **Relative entropy is cost of miscoding.**
Let the random variable X have five possible outcomes $\{1, 2, 3, 4, 5\}$. Consider two distributions on this random variable

Symbol	$p(x)$	$q(x)$	$C_1(x)$	$C_2(x)$
1	1/2	1/2	0	0
2	1/4	1/8	10	100
3	1/8	1/8	110	101
4	1/16	1/8	1110	110
5	1/16	1/8	1111	111

(a) Calculate $H(p)$, $H(q)$, $D(p||q)$ and $D(q||p)$.

(b) The last two columns above represent codes for the random variable. Verify that the average length of C_1 under p is equal to the entropy $H(p)$. Thus C_1 is optimal for p. Verify that C_2 is optimal for q.

(c) Now assume that we use code C_2 when the distribution is p. What is the average length of the codewords. By how much does it exceed the entropy $H(p)$?

(d) What is the loss if we use code C_1 when the distribution is q?