Problem Set 3

These problems have been selected from the Course Textbook by Cover and Thomas.

5.3 *Slackness in the Kraft Inequality.* A particular prefix code over a D-ary alphabet $\mathcal{D}$ has word lengths $\ell_1, \ell_2, \ldots, \ell_m$ which satisfy the strict inequality

$$\sum_{i=1}^{M} D^{-\ell_i} < 1.$$

Show that there exist arbitrarily long sequences in $\mathcal{D}^*$ which cannot be parsed into a concatenation of codewords.

5.5 *More Huffman codes.* Find the binary Huffman code for the source with probabilities $\left(\frac{1}{3}, \frac{1}{5}, \frac{1}{5}, \frac{2}{15}, \frac{2}{15}\right)$. Argue that this code is also optimal for the source with probabilities $\left(\frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}\right)$.

5.9 *Optimal code lengths that require one bit above entropy.* The source coding theorem shows that the optimal code for a random variable X has an expected length less than $H(X) + 1$. Give an example of a random variable for which the expected length of the optimal code is close to $H(X) + 1$ [i.e., for any $\epsilon > 0$, construct a distribution for which the optimal code has $L \geq H(X) + 1 - \epsilon$].

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

(a) Construct Huffman codes for this random variable and show that there exist two different sets of optimal codeword lengths.

(b) Conclude that there are optimal codes with codeword lengths for some symbols that *exceed* the Shannon codeword length $\lceil -\log p(x) \rceil$.

5.24 *Optimal codes for uniform distributions.* Consider a random variable with m equiprobable outcomes. The entropy H of this source is $\log_2 m$ bits.

(a) Describe the optimal binary prefix code for this source and compute the average codeword length L_m.

(b) For which values of m does $L_m = H$?

(c) We know that $L < H + 1$ for any probability distribution. The *redundancy* of a variable-length code is defined as $\rho = L - H$. For which value(s) of m, where $2^k \leq m < 2^{k+1}$ for some integer k, is the redundancy of the optimal code maximized? What is the limiting value of this worst-case redundancy as $k \to \infty$?

5.32 *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 ith bottle is bad is given by

$$(p_1, p_2, \dots, p_6) = \frac{1}{23}(8, 6, 4, 2, 2, 1).$$

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) Which bottle should be tasted first?

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

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 mixture should be tasted first?

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

4.1 *Doubly stochastic matrices.* An $n \times n$ matrix $P = [P_{ij}]$ is said to be *doubly stochastic* if $P_{ij} \geq 0$ for all i and j and $\sum_j P_{ij} = 1$ for all i and $\sum_i P_{ij} = 1$ for all j. (Thus all rows and columns are probability vectors.) A *permutation matrix* is a doubly stochastic matrix with precisely one nonzero entry (which must be equal to one) in each row (and hence in each column). It can be shown that every doubly stochastic matrix can be written as a convex combination of permutation matrices.

(a) Let $\mathbf{a} = (a_1, a_2, \dots, a_n)$, with $a_i \geq 0$ for all i, and $\sum_i a_i = 1$, be a probability vector. Let $\mathbf{b} = \mathbf{a}P$, where P is doubly stochastic. Show that $\mathbf{b}$ is a probability vector and that $H(b_1, \dots, b_n) \geq H(a_1, \dots, a_n)$. Thus, stochastic mixing increases entropy. *Hint:* apply the log sum inequality: for non-negative numbers $a_1, \dots, a_n$ and $b_1 \dots, b_n$, we have

$$\sum_{i=1}^{n} a_i \log \frac{a_i}{b_i} \geq \left(\sum_{i=1}^{n} a_i \right) \log \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} b_i}$$

with equality if and only if $\frac{a_i}{b_i} = c$ for every i, where c is a constant.

(b) Show that a stationary distribution μ for a doubly stochastic matrix P is the uniform distribution.

(c) Conversely, prove that if the uniform distribution is a stationary distribution for a Markov transition matrix P, then P is doubly stochastic.

4.7 *Entropy rates of Markov chains.*

(a) Find the entropy rate of the two-state Markov chain with transition matrix

$$P = \begin{bmatrix} 1 - p_{01} & p_{01} \\ p_{10} & 1 - p_{10} \end{bmatrix}$$

(b) What values of p_{01}, p_{10} maximize the entropy rate?

(c) Find the entropy rate of the two-state Markov chain with transition matrix

$$P = \begin{bmatrix} 1 - p & p \\ 1 & 0 \end{bmatrix}$$

(d) Find the maximum value of the entropy rate of the Markov chain of part (c). We expect that the maximizing value of p should be less than $\frac{1}{2}$, since the 0 state permits more information to be generated than the 1 state.

(e) Let $N(t)$ be the number of allowable state sequences of length t for the Markov chain of part (c). Find $N(t)$ and calculate

$$H_0 = lim_{t \to \infty} \frac{1}{t} \log N(t).$$

[*Hint*: Find a linear recurrence that expresses $N(t)$ in terms of $N(t-1)$ and $N(t-2)$. Why is H_0 an upper bound on the entropy rate of the Markov chain? Compare H_0 with the maximum entropy found in part (d).]

4.12 *Entropy rate of a dog looking for a bone.* A dog walks on the integers, possibly reversing direction at each step with probability $p = 1/10$. Let $X_0 = 0$. The first step is equally likely to be positive or negative. A typical walk might look like this:

$$(X_0, X_1, \ldots) = (0, -1, -2, -3, -4, -3, -2, -1, 0, 1, \ldots)$$

(a) Find $H(X_0, X_1, X_2, \ldots, X_n)$.

(b) Find the entropy rate of the dog.

(c) What is the expected number of steps that the dog takes before reversing direction?

4.33 *Chain inequality.* Let $X_1 \to X_2 \to X_3 \to X_4$ form a Markov chain. show that

$$I(X_1; X_3) + I(X_2; X_4) \le I(X_1; X_4) + I(X_2; X_3).$$