Information Theory: Principles and Applications
Homework 2 - Due: April 16, 2010

1. Determine if the following binary code is uniquely decodable. If the code is not uniquely decodable, construct an ambiguous sequence.

x_i	$\mathcal{C}(x_i)$
x_1	010
x_2	0001
x_3	0110
x_4	1100
x_5	00011
x_6	00110
x_7	11110
x_8	101011

2. A source with an alphabet of size 4 has symbol probabilities 1/3, 1/3, 2/9, 1/9.

   (a) Use the Huffman algorithm to find an optimal binary prefix-free code for this source.

   (b) Use the Huffman algorithm to find another optimal binary prefix-free code with a different set of lengths.

   (c) Find another prefix-free code that is optimal but cannot result from using the Huffman algorithm.

3. Consider a source with an alphabet of 9 symbols with probability distribution given by $(0.49; 0.26; 0.12; 0.03; 0.03; 0.03; 0.02; 0.01; 0.01)$.

   (a) Find a binary Huffman code for this source.

   (b) Find a quaternary Huffman code for this source.

   (c) What is the average length of the codewords for both coders?

   (d) Which one of the coders is more efficient?

4. Which of these codes cannot be Huffman codes for any probability assignment. Justify your answer.

   (a) $\{0, 10, 11\}$

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

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

5. A source has an alphabet of 4 letters, x_1, x_2, x_3, x_4 with probabilities p_1, p_2, p_3, p_4.

   (a) Suppose that $p_1 > p_2 = p_3 = p_4$. Find the smallest number q such that $p_1 > q$ implies that the length of the codeword for x_1, that is $l(x_1) = 1$, in a Huffman code.

   (b) Show by example that if $p_1 = q$, then a Huffman code exists with $l(x_1) > 1$.

   (c) Now assume the more general condition $p_1 > p_2 \geq p_3 \geq p_4$. Does $p_1 > q$ still imply that $l(x_1) = 1$? Justify your answer.

   (d) Now assume that the source has an arbitrary number K of letters with $p_1 > p_2 \geq \ldots \geq p_K$. Does $p_1 > q$ still imply that $l(x_1) = 1$?

   (e) Assume that $p_1 \geq p_2 \geq \ldots p_K$. Find the largest number q' such that $p_1 < q'$ implies that $l(x_1) > 1$?

6. Consider a source with M equiprobable symbols

   (a) Let $k = \lceil \log M \rceil$. Show that, for a binary Huffman code, the only possible codeword lengths are k and $k - 1$.

   (b) As a function of M, find how many codewords have length $k = \lceil \log M \rceil$. What is the expected codeword length, $\overline{L}$, in bits per source symbol?

   (c) Define $y = M/2^k$. Express $\overline{L} - \log M$ as a function of y. Find the maximum value of this function over $0.5 < y \leq 1$. This illustrates that the entropy bound $\overline{L} < H(X) + 1$ is rather loose in this equiprobable case.

7. Consider the following method for generating a code for a random variable X which takes on M values $\{1, \ldots, M\}$ with probabilities $p_1, \ldots, p_M$, respectively. Assume that the probabilities are ordered so that $p_1 \geq p_2 \geq \ldots \geq p_M$. Let:

$$F_i = P(X \leq i - 1) = \sum_{i=1}^{i-1} p_i \qquad F_1 = 0;$$

The codeword for i is the binary expansion of F_i rounded off to $\lceil \log 1/p_i \rceil$ bits. Example: $1/2 = .1000$, $1/4 = .0100$, $5/8 = .1010$.

(a) Show that the code constructed by this process is prefix-free and the average length satisfies $H(X) \leq \overline{L} \leq H(X) + 1$.

(b) Construct the code for the following probability distribution: $(0.5; 0.25; 0.125; 0.125)$ and determine the average lengtho of the codewords?

8. Inequalities and Weak Law of Large Numbers

(a) (Markov's inequality) For any nonnegative random variable X and any $a > 0$ show that

$$P(X \geq a) \leq \frac{E[X]}{a}$$

Find a random variable that achieves this inequality with equality.

(b) (Chebyshev's inequality) Let Y be a random variable with mean m_Y and variance σ_Y^2. Show that for any $\epsilon > 0$

$$P(|Y - m_Y| \geq \epsilon) \leq \frac{\sigma_Y^2}{\epsilon^2}$$

(c) Let $Z_1, \ldots, Z_n$ be a sequence of independent and identically distributed random variables with mean μ and variance σ^2. Let $\overline{Z} = \frac{1}{n}\sum_{i=1}^{n} Z_i$ be the sample mean. Show that:

$$P(|\overline{Z} - \mu| \geq \epsilon) \leq \frac{\sigma^2}{n\epsilon^2}$$

Therefore $P(|\overline{Z} - \mu| \geq \epsilon) \to 0$ as $n \to \infty$. This is known as the weak law of large numbers.

9. A discrete memoryless source emits a sequence of statistically independent binary digits with probabilities $P(X = 1) = 0.005$ and $P(X = 0) = 0.995$. The digits are taken 100 at a time and a binary codeword is provided for every sequence of 100 digits containing three or fewer ones.

(a) Assuming that all codewords are the same length, find the minimum length required to provide codewords for all sequences with three or fewer ones.

(b) Calculate the probability of observing a source sequence for which no codeword has been assigned.

(c) Use Chebyshev's inequality to bound the probability of observing a source sequence for which no codeword has been assigned. Compare this bound with the actual probability computed in part (b).

10. Finding the typical set: Suppose a discrete memoryless source emits 0 and 1 with probability 0.6 and 0.4 respectively. For $\epsilon = 0.1$, what is $A_\epsilon^{(25)}$? What is the probability of $A_\epsilon^{(25)}$? How many sequences belong to $A_\epsilon^{(25)}$? Compare this result with the upper and lower bounds.

    Obs.: Use a computer to build a table of probabilities for sequences with k zeros, $0 \leq k \leq 25$, and to find those sequences that are in the typical set.

11. (Extra) The typical set $A_\epsilon^{(n)}$ defined last class is often called a weakly typical set, in contrast to another kind of typical set called strongly typical set. Assume a discrete memoryless source and let $N_i(\mathbf{x}^n)$ be the number of symbols in a string $\mathbf{x}^n$ of length n taking on the value i. The probability for symbol i is p_i. Then the strongly typical set $A_\epsilon^{*(n)}$ is defined as follows:

$$A_\epsilon^{*(n)} = \left\{ \mathbf{x}^n : p_i(1-\epsilon) < \frac{N_i(\mathbf{x}^n)}{n} < p_i(1+\epsilon); \quad \text{for all } i \in \mathcal{X} \right\}$$

    (a) Show that $p_{\mathbf{X}^n}(\mathbf{x}^n) = \prod_i p_i^{N_i(\mathbf{x}^n)}$

    (b) Show that every $\mathbf{x}^n \in A_\epsilon^{*(n)}$ has the following property

$$H(X)(1-\epsilon) < \frac{-\log p_{\mathbf{X}^n}(\mathbf{x}^n)}{n} < H(X)(1+\epsilon)$$

    (c) Show that if $\mathbf{x}^n \in A_\epsilon^{*(n)}$, then $\mathbf{x}^n \in A_{\tilde{\epsilon}}^{(n)}$ with $\tilde{\epsilon} = H(X)\epsilon$, that is, $A_\epsilon^{*(n)} \subset A_{\tilde{\epsilon}}^{(n)}$.

    (d) Show that for any $\delta > 0$ and all sufficiently large n,

$$P(\mathbf{X}^n \in A_\epsilon^{*(n)}) \leq \delta$$

    Hint: Taking each letter i separately, $1 \leq i \leq |\mathcal{X}|$, show that for all sufficiently large n, $P(|N_i/n - p_i| \geq \epsilon) \leq \delta/|\mathcal{X}|$.

    (e) Show that for all $\delta > 0$ and all sufficiently large n.

$$(1-\delta)2^{n(H(X)-\epsilon)} < |A_\epsilon^{*(n)}| < 2^{n(H(X)+\epsilon)} \tag{1}$$

    Note that parts (d) and (e) constitute the same theorem for the strong typical set as the results showed in the class establishes for the weakly typical set. Typically the n required for (1) to hold (with the correspondence in part (c) between ϵ and $\tilde{\epsilon}$ is considerably larger than the one required for the weakly typical set result to hold. Strong typical sets are used to prove results in rate-distortion theory.