ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
School of Computer and Communication Sciences

Handout 5 Information Theory and Coding
Homework 3 October 2, 2012

PROBLEM 1. Consider two discrete memoryless sources. Source 1 has an alphabet of 6 symbols with the probabilities, 0.3, 0.2, 0.15, 0.15, 0.1, 0.1. Source 2 has an alphabet of 7 letters with probabilities 0.3, 0.25, 0.15, 0.1, 0.1, 0.05, 0.05. Construct a binary ($D = 2$) Huffman code and a ternary ($D = 3$) Huffman code for each source. Find the average number of code letters per source symbol in each case. Hint: observe that a ternary tree has an odd number of leaves. A fictitious symbol with probability 0 might therefore be needed for the code construction.

PROBLEM 2.

(a) A source has an alphabet of 4 letters, a_1, a_2, a_3, a_4, and we have the condition $P(a_1) > P(a_2) = P(a_3) = P(a_4)$. Find the smallest number q such that $P(a_1) > q$ implies that $n_1 = 1$ where n_1 throughout this problem is the length of the codeword for a_1 in a Huffman code.

(b) Show by example that if $P(a_1) = q$ (your answer in part (a)), then a Huffman code exists with $n_1 > 1$.

(c) Now assume the more general condition, $P(a_1) > P(a_2) \geq P(a_3) \geq P(a_4)$. Does $P(a_1) > q$ still imply that $n_1 = 1$? Why or why not?

(d) Now assume that the source has an arbitrary number K of letters with $P(a_1) > P(a_2) \geq \cdots \geq P(a_K)$. Does $P(a_1) > q$ now imply $n_1 = 1$?

(e) Assume $P(a_1) \geq P(a_2) \geq \cdots \geq P(a_K)$. Find the largest number q' such that $P(a_1) < q'$ implies that $n_1 > 1$.

PROBLEM 3. The following problem concerns a technique known as run length coding. Along with being a useful technique, it should make you look carefully into the sense in which Huffman coding is optimal. A source produces a sequence of independent binary digits with probabilities $P(\mathbf{0}) = 0.9$ and $P(\mathbf{1}) = 0.1$. We shall encode this sequence in two stages, first counting the number of 0's between successive 1's in the source output, and then encoding these counts into binary code words. The first stage of encoding maps source sequences into intermediate digits by the following rule:

Source Sequence	Intermediate Digits (# of zeros)
1	0
01	1
001	2
0001	3
⋮	⋮
00000001	7
00000000	8

Thus the following sequence is encoded as follows:

1	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**1**	**0**	**0**	**0**	**0**	**1**
0,			2,								8,			2,	0,					4

The final stage of encoding assigns a code word of length 1 to the intermediate digit 8 and codewords of length 4 to the other intermediate digits.

(a) Justify, in whatever detail you find convincing to yourself that the overall code is uniquely decodable.

(b) Find the average number $\bar{N}$ of source digits per intermediate digit.

(c) Find the average number $\bar{M}$ of encoded binary digits per intermediate digit.

(d) Show, by appeal to the law of large numbers, that for a very long source sequence of source digits, the ratio of the number of encoded binary digits to the number of source digits will with high probability be close to $\bar{M}/\bar{N}$. Compare this ratio to the average number number of code letters per source letter for a Huffman code encoding four source digits at a time.

Problem 4. For a stationary process $X_1, X_2, \ldots$, show that

(a) $\frac{1}{n}H(X_1, \ldots, X_n) \leq \frac{1}{n-1}H(X_1, \ldots, X_{n-1})$.

(b) $\frac{1}{n}H(X_1, \ldots, X_n) \geq H(X_n|X_{n-1}, \ldots, X_1)$.

Problem 5. Show, for a Markov chain, that

$$H(X_0|X_n) \geq H(X_0|X_{n-1}), \quad n \geq 1.$$

Thus, initial state X_0 becomes more difficult to recover as time goes by.

Problem 6. Let $X_1, X_2, \ldots$ be i.i.d., each with probability distribution $p(x)$. Show that with probability one

$$\lim_{n\to\infty} p(X_1, \ldots, X_n)^{1/n}$$

exists, and find its value. Hint: use the AEP.

Compare this to

$$\lim_{n\to\infty} E[p(X_1, \ldots, X_n)^{1/n}].$$

Problem 7. Let X, Y, Z be discrete random variables. Prove the validity of the following inequalities and find the conditions for equality:

(a) $I(X, Y; Z) \geq I(X; Z)$.

(b) $H(XY|Z) \geq H(X|Z)$.

(c) $I(X; Z|Y) \geq I(Z; Y|X) - I(Z; Y) + I(X; Z)$.

(d) $H(X, Y, Z) - H(X, Y) \leq H(X, Z) - H(X)$.