ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
School of Computer and Communication Sciences

Handout 4 Information Theory and Coding
Homework 2 Sep. 26, 2017

PROBLEM 1. A source has an alphabet of 4 letters. The probabilities of the letters and two possible sets of binary code words for the source are given below:

Letter	Prob.	Code I	Code II
a_1	0.4	1	1
a_2	0.3	01	10
a_3	0.2	001	100
a_4	0.1	000	1000

For *each* code, answer the following questions (no proofs or numerical answers are required).

(a) Is the code prefix-free?

N.B. A "prefix-free" code is also known as an *instantaneous* code.

(b) Is the code uniquely decodable?

(c) Give an heuristic description of the purpose of the first letter in the code words of code II.

PROBLEM 2. Let $\bar{M} = \sum_i p_i l_i^{2015}$ be the 2015th moment (i.e., the expected value of the 2015th power) of the code word lengths l_i associated with an encoding of a random variable X with distribution p. Let $\bar{M}_1 = \min \bar{M}$ over all prefix-free codes for X; and let $\bar{M}_2 = \min \bar{M}$ over all uniquely decodable codes for X. What relationship exists between $\bar{M}_1$ and $\bar{M}_2$?

PROBLEM 3. Consider the following method for constructing binary code words for a random variable U which takes values $\{a_1, \dots, a_m\}$ with probabilities $P(a_1), \dots, P(a_m)$. Assume that $P(a_1) \geq P(a_2) \geq \cdots \geq P(a_m)$. Define

$$Q_1 = 0 \quad \text{and} \quad Q_i = \sum_{k=1}^{i-1} P(a_k) \quad \text{for } i = 2, 3, \dots$$

The code word assigned to the letter a_i is formed by finding the binary expansion of $Q_i < 1$ (i.e, $1/2 = .100\dots$, $1/4 = .0100\dots$, $5/8 = .1010\dots$) and letting the codeword be the first l_i bits of this expansion where $l_i = \lceil -\log_2 P(a_i) \rceil$.

(a) Construct binary code words for the probability distribution $\{1/4, 1/4, 1/8, 1/8, 1/16, 1/16, 1/16, 1/16\}$.

(b) Prove that the method described above yields a prefix-free code and the average codeword length $\bar{L}$ satisfies

$$H(X) \leq \bar{L} < H(X) + 1.$$

PROBLEM 4. A random variable takes values on an alphabet of K letters, and each letter has the same probability. These letters are encoded into binary words using the Huffman procedure so as to minimize the average code word length. Let j and x be chosen such that $K = x2^j$, where j is an integer and $1 \leq x < 2$.

(a) Do any code words have lengths not equal to j or $j + 1$? Why?

(b) In terms of j and x, how many code words have length j?

(c) What is the average code word length?

PROBLEM 5. 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 and a ternary Huffman code for each source. Find the average number of code letters per source symbol in each case.

Hint: A ternary code is a mapping of source symbols to $\{0, 1, 2\}^*$. 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 6.

(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 $n_1 > 1$.