UC Irvine
EECS 141B
Spring 2014

HOMEWORK 3
(Due 5/22/2014)

1. The following are a set of classes for noiseless source codes.

    1. *Fixed-Length Codes:* A code whose codeword length is fixed.
    2. *Variable-Length Codes:* A code whose codeword length is not fixed.
    3. *Distinct Codes:* A code for which each codeword is distinguishable from all other codewords.
    4. *Prefix Codes:* A code in which no codeword is a prefix of another.
    5. *Uniquely Decodable Codes:* A code for which the original source sequence can be reconstructed perfectly from the encoded binary sequence. All prefix codes are uniquely decodable but not all uniquely decodable codes are prefix codes.

    Consider the following table where a source of size 4 has been encoded in binary codes with symbol 0 and 1.

s_i	Code 1	Code 2	Code 3	Code 4	Code 5	Code 6
s_1	00	00	0	0	0	1
s_2	01	01	1	10	01	01
s_3	00	10	00	110	011	001
s_4	11	11	11	111	0111	0001

    Fill in the following table by placing a check mark in the appropriate box if a code belongs to a particular class.

Class	Code 1	Code 2	Code 3	Code 4	Code 5	Code 6
Fixed-Length						
Variable-Length						
Distinct						
Prefix						
Uniquely Decodable						

2. Define the entropy of a discrete source $\mathcal{S}$ with probabilities p_k, $k = 0, 1, \ldots, K-1$, as

$$H(\mathcal{S}) = \sum_{k=0}^{K-1} p_k \log_2\left(\frac{1}{p_k}\right) = -\sum_{k=0}^{K-1} p_k \log_2(p_k).$$

(a) Show that

$$H(\mathcal{S}) - \log_2 K = \log_2(e) \sum_{k=0}^{K-1} p_k \ln\left(\frac{1}{Kp_k}\right).$$

(b) Let $f(x) = \ln(x) - (x-1)$. By using calculus on $f(x)$ show that

$$\ln(x) \leq (x-1), \qquad x > 0$$

with equality at $x = 1$.

(c) Using the result in part (b) above in part (a), show that

$$H(\mathcal{S}) \leq \log_2(K)$$

with equality if and only if $p_k = 1/K$, $k = 0, 1, \ldots .K-1$.

3. Consider a discrete memoryless source with alphabet $\mathcal{S} = \{s_0, s_1, \ldots, s_{K-1}\}$ with probabilities $\{p_0, p_1, \ldots, p_{K-1}\}$, respectively. The nth extension of this source is another discrete memoryless source with source alphabet $\mathcal{S}^n = \{\sigma_0, \sigma_1, \ldots, \sigma_{M-1}\}$ where $M = K^n$. Each σ_j corresponds to a unique, ordered concatenation of symbols $s_{j_1}, s_{j_2}, \ldots, s_{j_n}$ from $\mathcal{S}$, $j = 0, 1, \ldots, M-1$, $0 \leq j_1, j_2, \ldots, j_n \leq K-1$. Note

$$P(\sigma_j) = P(s_{j_1})P(s_{j_2}) \cdots P(s_{j_n})$$

and

$$\sum_{j=0}^{M-1} P(\sigma_j) = \sum_{j_1=0}^{K-1}\sum_{j_2=0}^{K-1} \cdots \sum_{j_n=0}^{K-1} P(s_{j_1})P(s_{j_2}) \cdots P(s_{j_n}).$$

(a) Show that

$$\sum_{j=0}^{M-1} P(\sigma_j) = 1.$$

(b) Show that

$$\sum_{j=0}^{M-1} P(\sigma_j) \log_2\left(\frac{1}{p_{j_k}}\right) = H(\mathcal{S}), \qquad k = 1, 2, \ldots, n.$$

(c) Hence, show that

$$H(\mathcal{S}^n) = \sum_{j=0}^{M-1} P(\sigma_j) \log_2\left(\frac{1}{P(\sigma_j)}\right) = nH(\mathcal{S}).$$

(d) Calculate the maximum entropy of an ASCII symbol.

(e) Calculate the maximum entropy of an extended ASCII symbol.

(f) Do you think the entropy of an English text approaches that of the maximum entropy of an ASCII symbol? Why or why not?

4. Let a source alphabet $\mathcal{S}$ be specified in terms of its symbols and probabilities as in the following table.

s_k	s_0	s_1	s_2	s_3	s_4	s_5
p_k	0.30	0.24	0.21	0.14	0.07	0.04

(a) Calculate the entropy $H(\mathcal{S})$.

(b) Construct the Huffman code for $\mathcal{S}$. Tabulate the codeword corresponding to each symbol s_k.

(c) Calculate the average code length $\bar{L}$.

(d) Calculate the efficiency of the code, $\eta = H(\mathcal{S})/\bar{L}$.