Department of Electrical and Computer Engineering
ECE 776 - Information Theory

Reading

Cover and Thomas, Chapter 5.

Homework 3

1. **Problem 5.3**
The instantaneous code has length $\{l_1, l_2, \ldots, l_m\}$. Let $l_{\max} = \max(l_1, l_2, \ldots, l_m)$. In $\mathcal{D}^*$, there are $D^{l_{\max}}$ sequences with length $l_{\max}$. Of these sequences, because the code is prefix free, there are $D^{l_{\max}-l_i}$ sequences have the codeword $C(x_i)$ as their prefix. Thus, the total number of sequences which has codeword as prefix is
$$\sum_{i=1}^{m} D^{l_{\max}-l_i} = D^{l_{\max}} \sum_{i=1}^{m} D^{-l_i}.$$
Since the instantaneous code strictly satisfies Kraft inequality $\sum_{i=1}^{m} D^{-l_i} < 1$, we have
$$\sum_{i=1}^{m} D^{l_{\max}-l_i} < D^{l_{\max}}.$$
This means that in the $D^{l_{\max}}$ sequences, there is sequence which does not have any code as prefix. Thus, this sequence and its children can not be decoded at all.

2. **Problem 5.8**
We first solve for steady state distribution and entropy rate of this Markov chain. Using the transition probability matrix,
$$\mu^T = \mu^T \begin{bmatrix} 1/2 & 1/4 & 1/4 \\ 1/4 & 1/2 & 1/4 \\ 0 & 1/2 & 1/2 \end{bmatrix},$$
we have $\mu = [2/9, 4/9, 1/3]^T$. For the Markov chain, the entropy rate is
$$\mathcal{H} = H(X_2|X_1) = \sum_{i=1}^{3} \mu_i H(X_2|X_1 = S_i) = 4/3 \text{ bits/symbol}$$
The code can be designed as follows:

	$C(U_n = S_1\| \ \cdot \ )$	$C(U_n = S_2\| \ \cdot \ )$	$C(U_n = S_3\| \ \cdot \ )$	$E[L\|U_{n-1} = S_i]$
$C(\ \cdot \ \|U_{n-1} = S_1)$	0	10	11	1.5 bits/symbol
$C(\ \cdot \ \|U_{n-1} = S_2)$	10	0	11	1.5 bits/symbol
$C(\ \cdot \ \|U_{n-1} = S_3)$	–	0	1	1 bits/symbol

The average code length is $E[L] = \sum_{i=1}^{3} \mu_i E[L|U_{n-1} = S_i] = 4/3$ bits/symbol $= \mathcal{H}$. Thus, the code is optimal.

3. **Problem 5.15**

   (a) Not instantaneous. Since codeword 0 is prefix of codeword 01.

   (b) Uniquely decodable. We can find all 01's first (by find all 1's) and the rest are 0'.

   (c) Non-singular since uniquely decodable.

4. **Problem 5.18**
   The only if part is obvious, since if $C^k(x_1, x_2, \ldots, x_k)$ is not one-to-one, then there exist $(x_1, x_2, \ldots, x_k)$ and $(x'_1, x'_2, \ldots, x'_k)$ such that $C^k(x_1, x_2, \ldots, x_k) = C^k(x'_1, x'_2, \ldots, x'_k)$ (i.e. $C(x)$ is not uniquely decodable).
   Conversely, if $C(x)$ is not uniquely decodable, then there exist $(x_1, x_2, \ldots, x_k)$ and $(x'_1, x'_2, \ldots, x'_n)$ such that $C^k(x_1, x_2, \ldots, x_k) = C^k(x'_1, x'_2, \ldots, x'_n)$. Then, for the sequences $(x_1, x_2, \ldots, x_k, x'_1, x'_2, \ldots, x'_n)$ and $(x'_1, x'_2, \ldots, x'_n, x_1, x_2, \ldots, x_k)$, the $C^{k+n}(\cdot)$ mapping is not one-to-one.

5. **Problem 5.19**
   For any optimal D-ary code, the codeword length must be finite (otherwise, it's not optimal). Let $L_{\max}$ be the length of longest codeword of all codes, then, on the D-ary tree with depth $L_{\max}$, there are only finite sets of $\{l_1, \ldots, l_m\}$ satisfies Kraft inequality (required by prefix condition). Thus,

$$L(C_{\text{opt}}) = \min_{\sum_{i=1}^{m} D^{-l_i} \leq 1} \sum_{i=1}^{m} p_i l_i.$$

   Since $L(C_{\text{opt}})$ is the minimum of finite numbers of continuous function in $\{p_1, \ldots, p_m\}$, it is also continuous in $\{p_1, \ldots, p_m\}$.

6. **Problem 5.21**

   (a) *If all message have equal probability, then there exists an optimal prefix code such that the longest codeword and shortest codeword differ by at most 1 bit.* To see this, we denote $C(x_s)$ and $C(x_l)$ as the shortest and longest codewords respectively. Then, if l_s and l_l differs by more than 1 bit, we can construct a new code such that $C'(x_s) = C(x_s)0$ and $C'(x_l) = C(x_s)1$. Then, the new code $L(C') \leq L(C)$. We can carry on this procedure until all codewords differ by at most 1 bit. Indeed, every optimal source code has this property. Thus, for a source with m messages, $l_s = \lfloor \log m \rfloor$ and $l_l = \lceil \log m \rceil$. We want to have more codewords with length l_s. Thus, we assign $2^{\lfloor \log m \rfloor}$ messages to the nodes at $\lfloor \log m \rfloor$ level. However, there are still $d = m - \lfloor \log m \rfloor$ messages not assigned. To assign these messages, we must take the original assigned message and one non-assigned message and put them as the two children of the original assigned node. Therefore, the code has $2d$ messages with length $\lceil \log m \rceil$ and $m - 2d$ messages with length $\lfloor \log m \rfloor$. The average length is

$$\begin{aligned} L &= \frac{1}{m}(2d\lceil \log m \rceil + (m - 2d)\lfloor \log m \rfloor) \\ &= \frac{1}{m}(2d + m\lfloor \log m \rfloor) = \lfloor \log m \rfloor + \frac{2d}{m} \end{aligned}$$

   (b) The average codeword length equals entropy if and only if $m = 2^k$. It's not hard to see that if $m = 2^k$, $L = k = H(X)$. Conversely,

$$L = \sum_{i=1}^{m} p_i l_i = -\sum_{i=1}^{m} p_i \log 2^{-l_i} = H(X) + D(p||q),$$

   where $p_i = 1/m$ and $q_i = 2^{-l_i}$. $D(p||q) = 0$ if and only if $p_i = q_i = 2^{-l_i} = 1/m$. Thus, $m = 2^k$.

(c) Let $m = 2^k + d$, $0 \le d \le 2^k$. The redundancy is

$$\rho \quad = \quad L - H = \lfloor \log m \rfloor + \frac{2d}{m} - \log m = k + \frac{2d}{2^k + d} - \frac{\ln(2^k + d)}{\ln 2}.$$

Setting

$$\frac{\partial \rho}{\partial d} \quad = \quad 0,$$

we get $d^* = 2^k(2\ln 2 - 1)$. Since $\rho(d)$ is concave in d, the interger d closest to d^* is going to give maximal redundancy and

$$\rho^* = \lim_{k\to\infty} \rho(d) = \rho(d^*) = 0.0861.$$