ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

School of Computer and Communication Sciences

Handout 14 — Introduction to Communication Systems

Solutions to Graded Homework 2 — November 6, 2008

PROBLEM 1.

1. Let $p_2 = p_3 = p_4 = x$ and $p_1 = y$. Clearly, $3x + y = 1$. Also for symbol a_1 to get the smallest length, 1, it should be picked last in the Huffman procedure. This implies that $y > 2x$. Thus we have $1 - 3x > 2x$ which implies that $x < \frac{1}{5}$. As a result $y > 1 - \frac{3}{5} = \frac{2}{5}$. Thus $q = \frac{2}{5}$.

2. If $p_1 = \frac{2}{5}$ then at the second step of the Huffman procedure we can chose either symbol a_1 as one of the two symbols with smallest probabilities or not which leads to either $n_1 = 2$ or $n_1 = 1$.

3. For the general case, we will prove that the sum of the two smallest probabilities p_3+p_4 is less than or equal to $\frac{2}{5}$. If we can prove this, then again as argued previously we would have $n_1 = 1$ since $p_1 > \frac{2}{5}$ and $p_1 > p_2$. To prove the above claim, assume the contrary. Thus assume that $p_3 + p_4 > \frac{2}{5}$. This implies that at least one of p_3 or p_4 is strictly greater than $\frac{1}{5}$. Now since $p_2 \geq p_3 \geq p_4$, this implies that $p_2 \geq \frac{1}{5}$. As a result $p_2 + p_3 + p_4 > \frac{3}{5}$ which would mean that $p_1 < \frac{2}{5}$, a contradiction (because we are given that $p_1 > \frac{2}{5}$).

PROBLEM 2.

1. (i) The Huffman tree is a complete binary tree of depth 3. This is show in Figure 1.

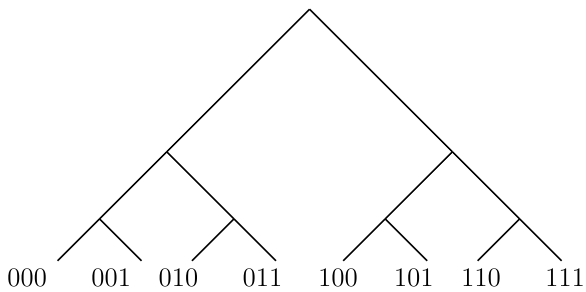

Figure 1: Huffman tree for Problem 2.1. It is a complete binary tree of depth 3.

(ii) The Huffman tree is the complete binary tree of depth n.

(iii) The entropy is equal to n. Just knowing that the entropy is n and there are 2^n symbols allows me to consider a code which has all the codewords of length n and enumerate all possible $n-$tuples. Clearly this code has average length equal to n. Also it is easy to check that the code is prefix-free, which makes the code an optimal one. In fact this is the only possible optimal code for such a source.

2. The length of the codewords are $1, 2, \ldots, n-2, n-1, n-1$. The Huffman tree is shown in Figure 2.

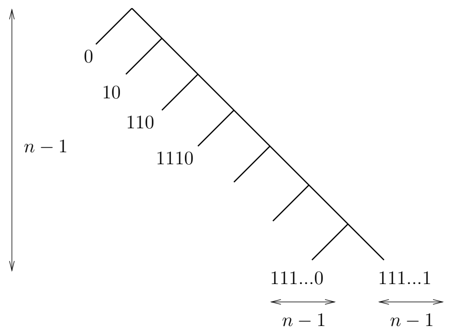

Figure 2: Huffman tree for Problem 2.2

Problem 3. 1. The average length is given by

$$\sum_i p_i \log_2 \frac{1}{q_i} = \sum_{i=1}^{n-1} i p_i + (n-1) p_n$$

Note that here when we ask "average length of your code", we mean that since we think the source has distribution given by q_i and since the probabilities are di-adic (meaning inverse power of 2), one possible optimal code would have lengths given by $\log_2 \frac{1}{q_i}$, and this is what we use.

2.

$$\begin{aligned}
-D(p||q) &= -\sum_i p_i \log(\frac{p_i}{q_i}) \\
&= \sum_i p_i \log(\frac{q_i}{p_i}) \\
&\leq \sum_i p_i (\frac{q_i}{p_i} - 1) \\
&= 1 - 1 = 0
\end{aligned}$$

where we used the fact that $\log(x) \leq x - 1$ for all $x \geq 0$. Note that $\log(x) < x - 1$ for all x except at $x = 1$ where there is equality, thus $D(p||q)$ is zero if and only if $p_i = q_i$ for all i.

3.

$$\begin{aligned}
\sum_i p_i \log(\frac{1}{q_i}) &= \sum_i p_i \log(\frac{p_i}{q_i}) + \sum_i p_i \log \frac{1}{p_i} \\
&= H(S) + D(p||q)
\end{aligned}$$

Problem 4. 1.

$$H(S_2) = \sum_{i \neq 21} p_i \log \frac{1}{p_i} + p'_{21} \log \frac{1}{p'_{21}} + p''_{21} \log \frac{1}{p''_{21}}$$

where $p'_{21} + p''_{21} = p_{21}$. We have

$$p'_{21} \log \frac{1}{p'_{21}} + p''_{21} \log \frac{1}{p''_{21}} > p_{21} \log \frac{1}{p_{21}} = p'_{21} \log \frac{1}{p_{21}} + p''_{21} \log \frac{1}{p_{21}}$$

which is true since $\log \frac{p_{21}}{p'_{21}} \geq 0$ and $\log \frac{p_{21}}{p''_{21}} \geq 0$. Thus $H(S_2) > H(S_1)$.

2. Let C_2 be an optimal code for S_2. We can create a code for S_1 by taking the same codewords as C_2 for all the alphabets except u and for u we take the codeword which has the smallest length amongst u and ü in C_2. Clearly the new code for S_1 is still uniquely decodable and its average length is smaller than L_2 by construction. Thus clearly any optimal code for S_1 will be better than one constructed above, hence $L_1 \leq L_2$.

Solutions for second part of Problem 4:

1. Clearly the code is still prefix-free since adding the tail bits do not make the codewords prefix of any other code.

2.

$$\begin{aligned} L_1' - L_1 &= \sum_{i \neq 21} p_i l_i + p_{21}'(l_{21} + 1) + p_{21}''(l_{21} + 1) - \sum_{i \neq 21} p_i l_i - p_{21} l_{21} \\ &= p_{21} \end{aligned}$$

3. Since L_2 is the length of the optimal code for S_2 we have $L_2 \leq L_1'$. Thus we have

$$\begin{aligned} L_2 - L_1 \leq L_1' - L_1 \\ = p_{21} \end{aligned}$$