ECE 534: Elements of Information Theory, Fall 2009
Homework 4
Solutions

5.4 (Luke Vercimak)

1.

Codeword Length	Codeword	X	Probability						
1	1	x_1	0.49	0.49	0.49	0.49	0.49	0.51	1
2	00	x_2	0.26	0.26	0.26	0.26	0.26	0.49	
3	011	x_3	0.12	0.12	0.12	0.13	0.25		
5	01000	x_4	0.04	0.05	0.08	0.12			
5	01001	x_5	0.04	0.04	0.05				
5	01010	x_6	0.03	0.04					
5	01011	x_7	0.02						

2. Find the expected code length for this encoding.

$$L(C) = \sum_x p(x)l(x) \tag{1}$$
$$= (0.49)(1) + (0.26)(2) + (0.12)(3) + (0.04)(5)(2) + (0.03)(5) + (0.02)(5) \tag{2}$$
$$= 2.013 \text{ bits} \tag{3}$$

3.

Codeword Length	Codeword	X	Probability			
1	0	x_1	0.49	0.49	0.49	1
1	1	x_2	0.26	0.26	0.26	
2	20	x_3	0.12	0.12	0.35	
2	22	x_4	0.04	0.09	0.25	
3	210	x_5	0.04	0.04		
3	211	x_6	0.03			
3	212	x_7	0.02			

5.20 (Jonathan Waxman) Suppose that $X = i$ with probability $p_i, i = 1, 2, ..., m$. Let l_i be the number of binary symbols in the codeword associated with $X = i$, and let c_i denote the cost per letter of the codeword when $X = i$.

(a) The average cost C of the description of X is $C = \sum_{i=1}^m p_i c_i l_i$. Given the constraint $\sum_{i=1}^m 2^{-l_i} \leq 1$, the optimal cost over l_i is found by minimizing the Lagrange function

$$J = \sum_{i=1}^m p_i c_i l_i + \lambda \left(\sum_{i=1}^m 2^{-l_i} \right).$$

Differentiate with respect to l_i:

$$\frac{\partial J}{\partial l_i} = p_i c_i - \lambda 2^{-l_i} \ln 2.$$

Setting the result equal to zero yields

$$2^{-l_i} = \frac{p_i c_i}{\lambda \ln 2}.$$

Substituting this into the constraint gives

$$\sum_{i=1}^{m} \frac{p_i c_i}{\lambda \ln 2} = 1.$$

$$\lambda = \frac{1}{\ln 2} \sum_{i=1}^{m} p_i c_i.$$

Hence,

$$\begin{aligned}
p_i c_i &= \frac{1}{\ln 2} \left(\sum_{i=1}^{m} p_i c_i \right) 2^{-l_i} \ln 2 \\
&= \left(\sum_{i=1}^{m} p_i c_i \right) 2^{-l_i}. \\
l_i^* &= -\log_2 \frac{p_i c_i}{\sum_{i=1}^{m} p_i c_i}. \\
C^* &= -\sum_{i=1}^{m} p_i c_i \log_2 \frac{p_i c_i}{\sum_{i=1}^{m} p_i c_i}.
\end{aligned}$$

(b) To use the Huffman code procedure, one would simply use the probability distribution $\frac{p_i c_i}{\sum_{i=1}^{m} p_i c_i}$.

(c) Consider the quantity $\left\lceil -\log_2 \frac{p_i c_i}{\sum_{i=1}^{m} p_i c_i} \right\rceil$. Then

$$
\begin{array}{lll}
-\log_2 \frac{p_i c_i}{\sum_{i=1}^m p_i c_i} & \leq \left\lceil -\log_2 \frac{p_i c_i}{\sum_{i=1}^m p_i c_i} \right\rceil & \leq -\log_2 \frac{p_i c_i}{\sum_{i=1}^m p_i c_i} + 1 \\
-p_i c_i \log_2 \frac{p_i c_i}{\sum_{i=1}^m p_i c_i} & \leq p_i c_i \left\lceil -\log_2 \frac{p_i c_i}{\sum_{i=1}^m p_i c_i} \right\rceil & \leq -p_i c_i \log_2 \frac{p_i c_i}{\sum_{i=1}^m p_i c_i} + p_i c_i \\
-\sum_{i=1}^m p_i c_i \log_2 \frac{p_i c_i}{\sum_{i=1}^m p_i c_i} & \leq \sum_{i=1}^m p_i c_i \left\lceil -\log_2 \frac{p_i c_i}{\sum_{i=1}^m p_i c_i} \right\rceil & \leq -\sum_{i=1}^m p_i c_i \log_2 \frac{p_i c_i}{\sum_{i=1}^m p_i c_i} + p_i c_i \\
C^* & \leq C_{Huffman} & \leq C^* + \sum_{i=1}^m p_i c_i
\end{array}
$$

5.30 (Venkatakumar Srinivasan)

5.30a $H(p), H(q), D(p||q)$ and $D(q||p)$

$$
\begin{aligned}
H(p) &= -\sum_x p(x) \log p(x) \\
&= \frac{1}{2}\log 2 + \frac{1}{4}\log 4 + \frac{1}{8}\log 8 + \frac{2}{16}\log 16 \\
&= \frac{1}{2} + \frac{1}{2} + \frac{3}{8} + \frac{1}{2} \\
&= \frac{3}{2} + \frac{3}{8} \\
&= \frac{15}{8} \\
&= 1.875
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
H(q) &= -\sum_x q(x) \log q(x) \\
&= \frac{1}{2}\log 2 + 4 \times \frac{1}{8}\log 8 \\
&= \frac{1}{2} + \frac{3}{2} \\
&= 2
\end{aligned}
$$

$D(p||q) = \sum p(x) \log \frac{p(x)}{q(x)}$

$$
\begin{aligned}
&= \frac{1}{2}\log\frac{1/2}{1/2} + \frac{1}{4}\log\frac{1/4}{1/8} + \frac{1}{8}\log\frac{1/8}{1/8} + 2 \times \frac{1}{16}\log\frac{1/16}{1/8} \\
&= \frac{1}{4}\log 2 + \frac{1}{8}\log 1/2 \\
&= \frac{1}{4}\log 2 - \frac{1}{8}\log 2 \\
&= \frac{1}{8}
\end{aligned}
$$

$D(q||p) = \sum q(x) \log \frac{q(x)}{p(x)}$

$$
\begin{aligned}
&= \frac{1}{2}\log\frac{1/2}{1/2} + \frac{1}{8}\log\frac{1/8}{1/4} + \frac{1}{8}\log\frac{1/8}{1/8} + 2\times\frac{1}{8}\log\frac{1/8}{1/16} \\
&= \frac{1}{8}\log 1/2 + \frac{1}{4}\log 2 \\
&= \frac{1}{4}\log 2 - \frac{1}{8}\log 2 \\
&= \frac{1}{8}
\end{aligned}
$$

5.30b *Average codeword lengths of* C_1 *and* C_2

The average codeword length for C_1 for distribution p,

$$
\begin{aligned}
&= 1\cdot\frac{1}{2} + 2\cdot\frac{1}{4} + 3\cdot\frac{1}{8} + 4\cdot\frac{1}{16} + 4\cdot\frac{1}{16} \\
&= \frac{1}{2} + \frac{1}{2} + \frac{3}{8} + \frac{1}{2} \\
&= \frac{3}{2} + \frac{3}{8} \\
&= \frac{15}{8} = H(p)
\end{aligned}
$$

The average codeword length for C_2 for distribution q,

$$
\begin{aligned}
&= 1\cdot\frac{1}{2} + 3\cdot\frac{1}{8} + 3\cdot\frac{1}{8} + 3\cdot\frac{1}{8} + 3\cdot\frac{1}{8} \\
&= \frac{3}{2} + \frac{1}{2} \\
&= 2 = H(q)
\end{aligned}
$$

5.30c *Average codeword lengths of* C_2 *when distribution is* p

$$
\begin{aligned}
E_p L_2 &= 1\cdot\frac{1}{2} + 3\cdot\frac{1}{4} + 3\cdot\frac{1}{8} + 3\cdot\frac{1}{16} + 3\cdot\frac{1}{16} \\
&= \frac{1}{2} + \frac{3}{4} + \frac{3}{8} + \frac{3}{8} = 2
\end{aligned}
$$

This exceeds H(p) by $\frac{1}{8} = D(p||q)$.

5.30d *Average codeword lengths of* C_1 *when distribution is* q

$$
\begin{aligned}
E_q L_1 &= 1\cdot\frac{1}{2} + 2\cdot\frac{1}{8} + 3\cdot\frac{1}{8} + 4\cdot\frac{1}{8} + 4\cdot\frac{1}{8} \\
&= \frac{1}{2} + \frac{1}{4} + \frac{3}{8} + 1 = \frac{17}{8}
\end{aligned}
$$

This exceeds H(q) by $\frac{1}{8} = D(q||p)$.

5.32 (Yiwei Song)

(a) The expected number of tastings required is

$$\frac{8}{23}*1+\frac{6}{23}*2+\frac{4}{23}*3+\frac{2}{23}*4+(\frac{2}{23}+\frac{1}{23})*5=\frac{55}{23}$$

(b)The one with highest probability (the first one in this case) should be tasted first.
(c)The Huffman code of this probability distribution is

$$C_i=\{00,01,11,101,1000,1001\}$$

The minimum expected number of tastings required equals expected length of Huffman code, which is

$$E(L)=\sum_{i=1}^{6}p_il_i=\frac{54}{23}$$

(d) Since it does not matter whether we determine the first bit or second bit firstly. The first tasted mixture could be
mixture of 1,2 or
mixture of 3,4,5,6 or
mixture of 2,3 or
mixture of 1,4,5,6
And according to the fact that exchanging the codewords of the same length gives another optimal code, the first tasted mixture could also be
mixture of 1,3 or
mixture of 2,4,5,6

5.33 (Nathan Schneider)

a)

The binary Huffman codewords are given by $\begin{pmatrix} 0 & 0.6 \\ 10 & 0.3 \\ 11 & 0.1 \end{pmatrix}$
i.e. the Huffman code has codewords of length 1, 2, and 2.

The Shannon codeword lengths are given by $l(x)=\left\lceil \log(\frac{1}{p(x)})\right\rceil$: $\begin{pmatrix} 1 & 0.6 \\ 2 & 0.3 \\ 4 & 0.1 \end{pmatrix}$

b)
For $D \geq 3$, the Huffman code will produce all codewords of length one. Thus, all Shannon codewords must be of length one for the expected Shannon codeword length to be equal to the expected Huffman codeword length. This requires the smallest integer D such that $\lceil \log_D(10)\rceil = 1$. The D that satisfies this condition is $D = 10$.