2nd Semester 2009

Solutions to Set #3
Data Compression, Huffman code, Shannon Code

1. **Huffman coding.**
   Consider the random variable

$$X = \begin{pmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 \\ 0.50 & 0.26 & 0.11 & 0.04 & 0.04 & 0.03 & 0.02 \end{pmatrix}$$

   (a) Find a binary Huffman code for X.
   (b) Find the expected codelength for this encoding.
   (c) Extend the Binary Huffman method to Ternarry (Alphabet of 3) and apply it for X.

   Solution: Huffman coding.

   (a) The Huffman tree for this distribution is

Codeword								
1	x_1	0.50	0.50	0.50	0.50	0.50	0.50	1
01	x_2	0.26	0.26	0.26	0.26	0.26	0.50	
001	x_3	0.11	0.11	0.11	0.11	0.24		
00011	x_4	0.04	0.04	0.08	0.13			
00010	x_5	0.04	0.04	0.05				
00001	x_6	0.03	0.05					
00000	x_7	0.02						

   (b) The expected length of the codewords for the binary Huffman code is 2 bits. ($H(X) = 1.99$ bits)

   (c) The ternary Huffman tree is

Codeword					
0	x_1	0.50	0.50	0.50	1.0
1	x_2	0.26	0.26	0.26	
20	x_3	0.11	0.11	0.24	
21	x_4	0.04	0.04		
222	x_5	0.04	0.09		
221	x_6	0.03			
220	x_7	0.02			

This code has an expected length 1.33 ternary symbols. ($H_3(X) =$ 1.25 ternary symbols).

2. **Codes.**
Let $X_1, X_2, \ldots,$ i.i.d. with

$$X = \begin{cases} 1, & \text{with probability } 1/2 \\ 2, & \text{with probability } 1/4 \\ 3, & \text{with probability } 1/4. \end{cases}$$

Consider the code assignment

$$C(x) = \begin{cases} 0, & \text{if } x = 1 \\ 01, & \text{if } x = 2 \\ 11, & \text{if } x = 3. \end{cases}$$

(a) Is this code nonsingular?

(b) Uniquely decodable?

(c) Instantaneous?

(d) Entropy Rate is defined as

$$H(\mathcal{X}) \triangleq \lim_{n\to\infty} \frac{H(X^n)}{n}. \tag{1}$$

What is the entropy rate of the process

$$Z_1 Z_2 Z_3 \ldots = C(X_1) C(X_2) C(X_3) \ldots?$$

Solution: Codes.

(a) **Yes**, this code is nonsingular because $C(x)$ is different for every x.

(b) **Yes**, this code is uniquely decodable. Reversing the codewords

$$C'(x) = \begin{cases} 0, & \text{if } x = 1 \\ 10, & \text{if } x = 2 \\ 11, & \text{if } x = 3 \end{cases}$$

gives an instantaneous code, and thus a uniquely decodable code. Therefore the reversed extension is uniquely decodable, and so the extension itself is also uniquely decodable.

(c) **No**, this code is not instantaneous because $C(1)$ is a prefix of $C(2)$.

(d) The expected codeword length is

$$L(C(x)) = 0.5 \times 1 + 0.25 \times 2 + 0.25 \times 2 = \frac{3}{2}.$$

Further, the entropy rate of the i.i.d. X^n is

$$H(\mathcal{X}) = H(X) = H(.5, .25, .25) = \frac{3}{2}.$$

So the code is a uniquely decodable code with $L = H(\mathcal{X})$, and therefore the sequence is maximally compressed with $H(\mathcal{Z}) = 1$ bit. If $H(\mathcal{Z})$ were less than its maximum of 1 bit then the Z^n sequence could be further compressed to its entropy rate, and X^m could also be compressed further by blockcoding. However, this would result in $L_m < H(\mathcal{X})$ which contradicts theorem 5.4.2 of the text. So $H(\mathcal{Z}) = 1$ bit.

Note that the Z^n sequence is not i.i.d. $\sim \mathrm{Br}(\frac{1}{2})$, even though $H(\mathcal{Z}) = 1$ bit. For example, $P\{Z_1 = 1\} = \frac{1}{4}$, and a sequence starting $10\ldots$ is not allowed. However, once $Z_i = 0$ for some i then Z_k is Bernoulli($\frac{1}{2}$) for $k > i$, so Z^n is asymptotically Bernoulli($\frac{1}{2}$) and gives the entropy rate of 1 bit.

3. **Compression**

(a) Give a Huffman encoding into an alphabet of size D = 2 of the following probability mass function:

$$\left(\frac{1}{2}, \frac{1}{8}, \frac{1}{8}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}\right)$$

(b) Assume you have a file of size 1,000 symbols where the symbols are distributed i.i.d. according to the pmf above. After applying the Huffman code, what would be the pmf of the compressed binary file and what would be the expected length?

4. **Solution:**

The code is presented in Fig 1.

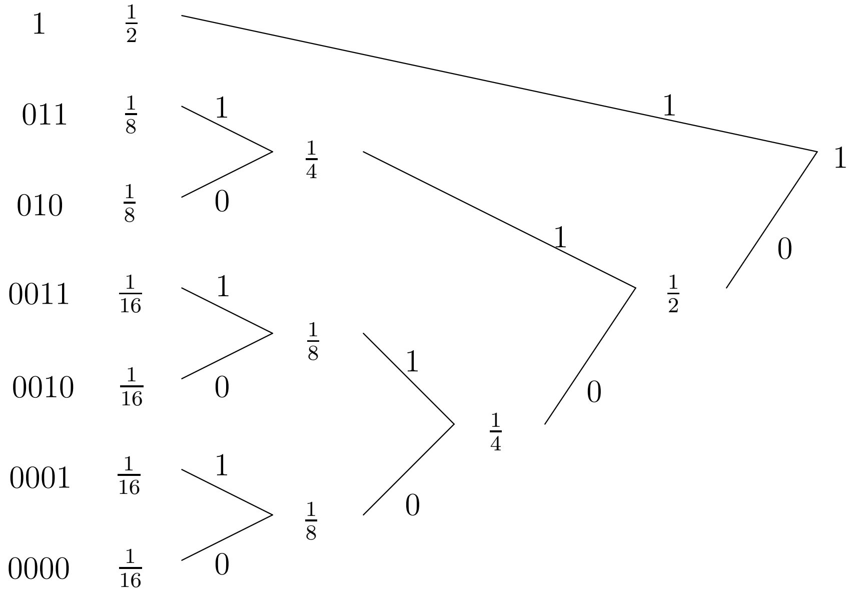

Figure 1: Huffman

(a) Huffman code is optimal code and achieves the entropy for dyadic distribution. If the distribution of the digits is not $Bernoulli(\frac{1}{2})$ you can compress it further. The binary digits of the data would be equally distributed after applying the Huffman code and therefore $p_0 = p_1 = \frac{1}{2}$.
The expected length would be:

$$E[l] = \frac{1}{2}\cdot 1 + \frac{1}{8}\cdot 3 + \frac{1}{8}\cdot 3 + \frac{1}{16}\cdot 4 + \frac{1}{16}\cdot 4 + \frac{1}{16}\cdot 4 + \frac{1}{16}\cdot 4 = 2.25$$

Therefore, the expected length of 1000 symbols would be 2250 bits.

5. **Entropy and source coding of a source with infinite alphabet** (15 points)

Let X be an i.i.d. random variable with an infinite alphabet, $\mathcal{X} = \{1, 2, 3, ...\}$. In addition let $P(X = i) = 2^{-i}$.

(a) What is the entropy of the random variable?

(b) Find an optimal variable length code, and show that it is indeed optimal.

Solution

(a)

$$\begin{aligned} H(X) &= -\sum_{x\in\mathcal{X}} p(x)\log p(x) \\ &= -\sum_{i=1}^{\infty} 2^{-i}\log_2(2^{-i}) \\ &= -\sum_{i=1}^{\infty} \frac{-i}{2^i} = 2 \end{aligned}$$

(b) Coding Scheme:

1	0
2	10
3	110
4	1110
5	11110
.	.
.	.
.	.

Average Length:

$$L^* = \sum_{i=1}^{\infty} p(x=i)L(i) = \sum_{i=1}^{\infty} \frac{i}{2^i} = 2 = H(X)$$

Hence it is the Optimal Code.

6. **Bad wine.**

One is given 6 bottles of wine. It is known that precisely one bottle has gone bad (tastes terrible). From inspection of the bottles it is determined that the probability p_i that the i^{th} bottle is bad is given by $(p_1, p_2, \ldots, p_6) = (\frac{7}{26}, \frac{5}{26}, \frac{4}{26}, \frac{4}{26}, \frac{3}{26}, \frac{3}{26})$. Tasting will determine the bad wine.

Suppose you taste the wines one at a time. Choose the order of tasting to minimize the expected number of tastings required to determine the bad bottle. Remember, if the first 5 wines pass the test you don't have to taste the last.

(a) What is the expected number of tastings required?

(b) Which bottle should be tasted first?

Now you get smart. For the first sample, you mix some of the wines in a fresh glass and sample the mixture. You proceed, mixing and tasting, stopping when the bad bottle has been determined.

(c) What is the minimum expected number of tastings required to determine the bad wine?

(d) What mixture should be tasted first?

Solution: Bad wine.

(a) If we taste one bottle at a time, the corresponding number of tastings are $\{1, 2, 3, 4, 5, 5\}$ with some order. By the same argument as in Lemma 5.8.1, to minimize the expected length $\sum p_i l_k$ we should have $l_j \leq l_k$ if $p_j > p_k$. Hence, the best order of tasting should be from the most likely wine to be bad to the least.

The expected number of tastings required is

$$\begin{aligned}\sum_{i=1}^{6} p_i l_i &= 1 \times \frac{7}{26} + 2 \times \frac{5}{26} + 3 \times \frac{4}{26} + 4 \times \frac{4}{26} + 5 \times \frac{3}{26} + 5 \times \frac{3}{26} \\ &= \frac{75}{26} \\ &= 2.88\end{aligned}$$

(b) The first bottle to be tasted should be the one with probability $\frac{7}{26}$.

(c) The idea is to use Huffman coding.

(01)	7	7	8	11	15	26
(11)	5	6	7	8	11	
(000)	4	5	6	7		
(001)	4	4	5			
(100)	3	4				
(101)	3					

The expected number of tastings required is

$$\begin{aligned}\sum_{i=1}^{6} p_i l_i &= 2 \times \frac{7}{26} + 2 \times \frac{5}{26} + 3 \times \frac{4}{26} + 3 \times \frac{4}{26} + 3 \times \frac{3}{26} + 3 \times \frac{3}{26} \\ &= \frac{66}{26} \\ &= 2.54\end{aligned}$$

Note that $H(p) = 2.52$ bits.

(d) The mixture of the first, third, and forth bottles should be tasted first, (or equivalently the mixture of the second, fifth and sixth).

7. **Relative entropy is cost of miscoding.**
Let the random variable X have five possible outcomes $\{1, 2, 3, 4, 5\}$. Consider two distributions on this random variable

Symbol	$p(x)$	$q(x)$	$C_1(x)$	$C_2(x)$
1	1/2	1/2	0	0
2	1/4	1/8	10	100
3	1/8	1/8	110	101
4	1/16	1/8	1110	110
5	1/16	1/8	1111	111

(a) Calculate $H(p)$, $H(q)$, $D(p||q)$ and $D(q||p)$.

(b) The last two columns above represent codes for the random variable. Verify that the average length of C_1 under p is equal to the entropy $H(p)$. Thus C_1 is optimal for p. Verify that C_2 is optimal for q.

(c) Now assume that we use code C_2 when the distribution is p. What is the average length of the codewords. By how much does it exceed the entropy $H(p)$?

(d) What is the loss if we use code C_1 when the distribution is q?

Solution: Relative entropy is cost of miscoding.

(a)

$$\begin{aligned} H(p) &= \sum_i -p_i \log p_i \\ &= -\frac{1}{2}\log\frac{1}{2} - \frac{1}{4}\log\frac{1}{4} - \frac{1}{8}\log\frac{1}{8} - 2\cdot\frac{1}{16}\log\frac{1}{16} \\ &= \frac{15}{8}. \end{aligned}$$

Similarly, $H(q) = 2$.

$$\begin{aligned} D(p||q) &= \sum_i p_i \log\frac{p_i}{q_i} \\ &= \frac{1}{2}\log\frac{1/2}{1/2} + \frac{1}{4}\log\frac{1/4}{1/8} + \frac{1}{8}\log\frac{1/8}{1/8} + 2\cdot\frac{1}{16}\log\frac{1/16}{1/8} \\ &= \frac{1}{8}. \end{aligned}$$

Similarly, $D(q||p) = \frac{1}{8}$.

(b) The average codeword length for C_1 is

$$El_1 = \frac{1}{2}\cdot 1 + \frac{1}{4}\cdot 2 + \frac{1}{8}\cdot 3 + 2\cdot\frac{1}{16}\cdot 4 = \frac{15}{8}.$$

Similarly, the average codeword length for C_2 is 2.

(c)

$$E_p l_2 = \frac{1}{2}\cdot 1 + \frac{1}{4}\cdot 3 + \frac{1}{8}\cdot 3 + 2\cdot\frac{1}{16}\cdot 3 = 2,$$

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

(d) Similarly, $E_q l_1 = \frac{17}{8}$, which exceeds $H(q)$ by $D(q||p) = \frac{1}{8}$.

8. **Shannon code.** Consider the following method for generating a code for a random variable X which takes on m values $\{1, 2, \ldots, m\}$ with probabilities $p_1, p_2, \ldots, p_m$. Assume that the probabilities are ordered so that $p_1 \geq p_2 \geq \cdots \geq p_m$. Define

$$F_i = \sum_{k=1}^{i-1} p_i, \tag{2}$$

the sum of the probabilities of all symbols less than i. Then the codeword for i is the number $F_i \in [0, 1]$ rounded off to l_i bits, where $l_i = \lceil \log \frac{1}{p_i} \rceil$.

(a) Show that the code constructed by this process is prefix-free and the average length satisfies

$$H(X) \leq L < H(X) + 1. \tag{3}$$

(b) Construct the code for the probability distribution $(0.5, 0.25, 0.125, 0.125)$.

Solution to Shannon code.

(a) Since $l_i = \lceil \log \frac{1}{p_i} \rceil$, we have

$$\log \frac{1}{p_i} \leq l_i < \log \frac{1}{p_i} + 1 \tag{4}$$

which implies that

$$H(X) \leq L = \sum p_i l_i < H(X) + 1. \tag{5}$$

The difficult part is to prove that the code is a prefix code. By the choice of l_i, we have

$$2^{-l_i} \leq p_i < 2^{-(l_i - 1)}. \tag{6}$$

Thus F_j, $j > i$ differs from F_i by at least 2^{-l_i}, and will therefore differ from F_i is at least one place in the first l_i bits of the binary expansion of F_i. Thus the codeword for F_j, $j > i$, which has length $l_j \geq l_i$, differs from the codeword for F_i at least once in the first l_i places. Thus no codeword is a prefix of any other codeword.

(b) We build the following table

Symbol	Probability	F_i in decimal	F_i in binary	l_i	Codeword
1	0.5	0.0	0.0	1	0
2	0.25	0.5	0.10	2	10
3	0.125	0.75	0.110	3	110
4	0.125	0.875	0.111	3	111

The Shannon code in this case achieves the entropy bound (1.75 bits) and is optimal.