Solutions Manual
Tutorial 4: Introduction to Information Theory

University of Windsor
Department of Electrical and Computer Engineering
ELEC 4190 - Digital Communications

Note: You should try the problems before and during the tutorial session. Your solutions will not be collected or graded.

1. The Nyquist rate of the sampling is $2 \times 6000 = 12000$ samples/sec and the sampling frequency is $f_s = 12000 + 2000 = 14000$ samples/sec.

   The entropy is

$$\begin{aligned} H(x) &= -\sum_i p_i \log_2 p_i \\ &= -(0.2 \log_2 0.2 + 0.1 \log_2 0.1 + 0.15 \log_2 0.15 \\ &\quad + 0.05 \log_2 0.05 + 0.3 \log_2 0.3 + 0.2 \log_2 0.2) \\ &= 2.4087 \text{ bits/source output} \end{aligned}$$

   The information rate in bits per second is $f_s \times H(X) = 14000 \times 2.4087 = 33721.8$ bps.

2.

   (a) The entropy is

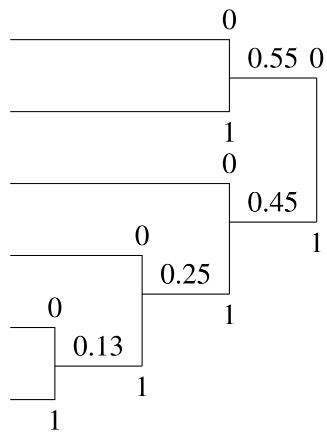
$$\begin{aligned} H(x) &= -\sum_i p_i \log_2 p_i \\ &= -(0.3 \log_2 0.3 + 0.25 \log_2 0.25 + 0.2 \log_2 0.2 \\ &\quad + 0.12 \log_2 0.12 + 0.08 \log_2 0.08 + 0.05 \log_2 0.05) \\ &= 2.36 \text{ bits/source output} \end{aligned}$$

(b) The average codeword length is lower bounded by the entropy of the source for error free reconstruction. So, the minimum possible average codeword length is $H(X) = 2.36$.

(c) Uniformly distributed source means $p_i = 1/6$ for all outputs. In this case, the entropy is $H(X) = \log_2 N = \log_2 6 = 2.585$ bits/source output which is higher.

(d) To design the Huffman code, we start be sorting source outputs in decreasing order of their probabilities and follow the procedure below.

x_i	Code	$P(x_i)$
x_1	00	0.30
x_2	01	0.25
x_3	10	0.20
x_4	110	0.12
x_5	1110	0.08
x_6	1111	0.05



The average codeword length is

$$\begin{aligned}\bar{L} &= \sum_{i=1}^{N} p_i l_i \\ &= 2 \times 0.3 + 2 \times 0.25 + 2 \times 0.2 + 3 \times 0.12 + 4 \times 0.08 + 4 \times 0.05 \\ &= 2.38 \text{ bits/source output}\end{aligned}$$

It is clear that the average codeword length is very close the source entropy.

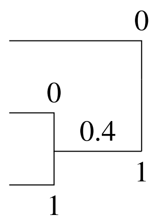
3. A discrete memoryless source has an alphabet $\{a_1, a_2, a_3\}$ with corresponding probabilities $\{0.1, 0.3, 0.6\}$.

(a) The entropy is

$$\begin{aligned}H(x) &= -\sum_{i} p_i \log_2 p_i \\ &= -\left(0.1 \log_2 0.1 + 0.3 \log_2 0.3 + 0.6 \log_2 0.6\right) \\ &= 1.2955 \text{ bits/source output}\end{aligned}$$

(b) To design the Huffman code, we start be sorting source outputs in decreasing order of their probabilities and follow the procedure below.

x_i	Code	$P(x_i)$
a_3	0	0.6
a_2	10	0.3
a_1	11	0.1



The average codeword length is

$$
\begin{aligned}
\bar{L} &= \sum_{i=1}^{N} p_i l_i \\
&= 1 \times 0.6 + 2 \times 0.3 + 2 \times 0.1 \\
&= 1.4 \text{ bits/source output}
\end{aligned}
$$

(c) For the second extension (i.e., $n = 2$) of the source, the new alphabet is

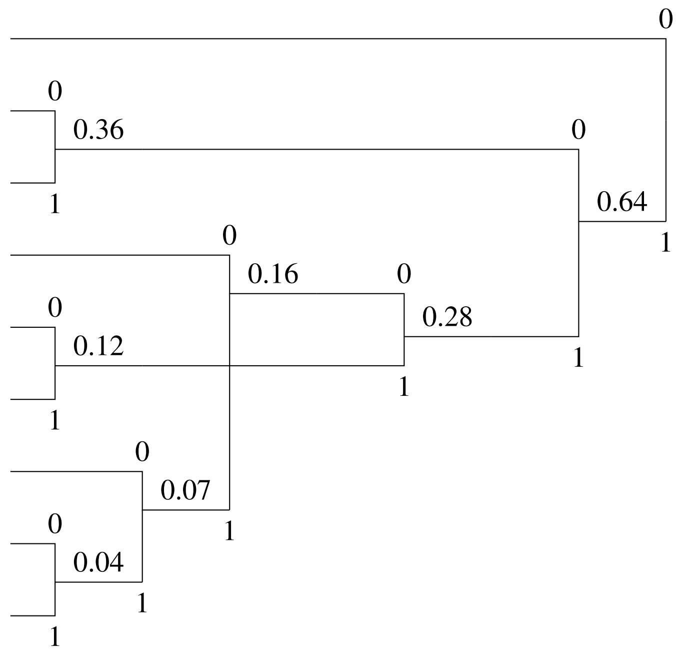
$$\{a_1a_1, a_1a_2, a_1a_3, a_2a_1, a_2a_2, a_2a_3, a_3a_1, a_3a_2, a_3a_3\}$$

Note that the size is $3^n = 9$ source outputs in this case.
The corresponding probabilities are

$$\{0.01, 0.03, 0.06, 0.03, 0.09, 0.18, 0.06, 0.18, 0.36\}$$

Note that each value is obtained by multiplying the corresponding two probabilities. For example, the first value is 0.1×0.1 which corresponds to the first output a_1a_1, the second value is 0.1×0.3 which corresponds to the second output a_1a_2, etc.
To design the Huffman code, we start be sorting source outputs in decreasing order of their probabilities and follow the same procedure.

x_i	Code	$P(x_i)$
a_3a_3	0	0.36
a_2a_3	100	0.18
a_3a_2	101	0.18
a_2a_2	1100	0.09
a_1a_3	1110	0.06
a_3a_1	1111	0.06
a_1a_2	11110	0.03
a_2a_1	111110	0.03
a_1a_1	111111	0.01



The average codeword length is

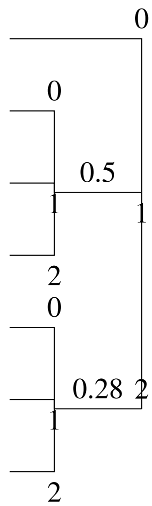
$$\bar{L}_2 = \sum_{i=1}^{N} p_i l_i$$
$$= 2.61 \text{ bits/source output pair}$$

This means that the average number of required binary letters per each source output letter $\bar{L} = \bar{L}_2/2 = 1.3$ bits/source output.

(d) Huffman coding of the original source requires 1.4 bits per source output letter (i.e., efficiency is $1.2955/1.4 = 92.54\%$). Huffman coding of the second extension of the source requires 1.3 bits per source output letter (i.e., efficiency is $1.2955/1.3 = 99.65\%$) and thus it is more efficient.

4. To design the Huffman code, we start be sorting source outputs in decreasing order of their probabilities and follow the same procedure.

x_i	Code	$P(x_i)$
x_1	0	0.22
x_2	10	0.18
x_3	11	0.17
x_4	12	0.15
x_5	20	0.13
x_6	21	0.1
x_7	22	0.05



The average codeword length is

$$\bar{L} = \sum_{i=1}^{N} p_i l_i$$
$$= 1.78 \text{ ternary symbol/source output}$$

For a fair comparison of the average codeword length with the entropy of the source, we should compute the entropy of ternary code with logarithms in base 3. So,

$$H(X) = -\sum_i p_i \log_3 p_i$$
$$= 1.7047 \text{ ternary symbol/source output}$$

As expected, $\bar{L} \geq H(X)$.

5.

Location	Content	Codeword
1	B	0B
2	A	0A
3	BB	1B
4	BBB	3B
5	AB	2B
6	BBA	3A
7	ABB	5B
8	ABA	5A
9	AA	2A
10	BBBA	4A
11	BA	1A
12	ABAA	8A
13	ABAB	8B
14	BAB	11B
15	AAA	9A
16	ABAAB	12B
17	BAA	11A

6. Given $P(X=0,Y=0)=P(X=0,Y=1)=P(X=1,Y=1)=1/3$ and $\sum_i \sum_j p(x_i,y_j)=1$, we get that $P(X=1,Y=0)=0$.

   The marginal probabilities are

$$P(X=0)=P(X=0,Y=0)+P(X=0,Y=1)=2/3$$
$$P(X=1)=P(X=1,Y=0)+P(X=1,Y=1)=1/3$$
$$P(Y=0)=P(X=0,Y=0)+P(X=1,Y=0)=1/3$$
$$P(Y=1)=P(X=0,Y=1)+P(X=1,Y=1)=2/3$$

   Then,

$$H(X)=-\sum_i p(x_i)\log_2 p(x_i)=0.9183 \text{ bits/source output}$$
$$H(Y)=-\sum_j p(y_j)\log_2 p(y_j)=0.9183 \text{ bits/source output}$$

   The joint entropy is

$$H(X,Y)=-\sum_i \sum_j p(x_i,y_j)\log_2 p(x_i,y_j)=1.585 \text{ bits/source output}$$

The conditional entropy is

$$H(X|Y) = H(X,Y) - H(Y) = 0.6667 \text{ bits/source output}$$
$$H(Y|X) = H(X,Y) - H(X) = 0.6667 \text{ bits/source output}$$

The mutual information is

$$I(X;Y) = H(X) - H(X|Y) = 0.2516 \text{ bits/source output}$$

7. The minimum required channel capacity is equal to the entropy of the source. Hence,

$$R = H(X) = -\sum_i p(x_i) \log_2 p(x_i) = 2.8155 \text{ bits/source output}$$

For a BSC, the maximum capacity is at most 1, so this source cannot be transmitted via a BSC.

8. Using Shannon's formula, we get

$$C = W \log_2(1 + \text{SNR}) = 3200 \log_2(1 + 10^{\frac{35}{10}}) = 37.2 \text{ kbps}$$

9. Noiseless channel means that the crossover probability $p = 0$. In this case, $H(X|Y) = H(Y|X) = 0$, hence, $I(X;Y) = H(X) = H(Y)$. The channel capacity is the maximum of $H(X)$ (or $H(Y)$) which is $\log_2 N$ bits/symbol where N is the number of source outputs.

10. First, we need to find the mutual information. Assume the marginal probability of the inputs are $P(X = A) = a$, $P(X = B) = b$, and $P(X = C) = c$ such that $a + b + c = 1$.

The marginal probabilities are

$$\begin{aligned}
P(Y = 1) &= P(Y = 1|X = A)P(X = A) + P(Y = 1|X = B)P(X = B) + P(Y = 1|X = C)P(X = C) \\
&= a + 0.5b + 0.5c = 0.5(1 + a) \\
P(Y = 2) &= P(Y = 2|X = A)P(X = A) + P(Y = 2|X = B)P(X = B) + P(Y = 2|X = C)P(X = C) \\
&= 0 + 0.5b + 0.5c = 0.5b + 0.5c = 0.5(1 - a)
\end{aligned}$$

The joint probabilities are

$$\begin{aligned}
P(X = A, Y = 1) &= P(Y = 1|X = A)P(X = A) = a \\
P(X = A, Y = 2) &= P(Y = 2|X = A)P(X = A) = 0 \\
P(X = B, Y = 1) &= P(Y = 1|X = B)P(X = B) = 0.5b \\
P(X = B, Y = 2) &= P(Y = 2|X = B)P(X = B) = 0.5b \\
P(X = C, Y = 1) &= P(Y = 1|X = C)P(X = C) = 0.5c
\end{aligned}$$

$$P(X=C,Y=2)=P(Y=2|X=C)P(X=C)=0.5c$$

Then,

$$\begin{aligned}H(Y)&=-\sum_j p(y_j)\log_2 p(y_j)\\&=-0.5(1+a)\log_2 0.5(1+a)-0.5(1-a)\log_2 0.5(1-a)\end{aligned}$$

and

$$\begin{aligned}H(Y|X)&=-\sum_i\sum_j p(x_i,y_j)\log_2 p(y_j|x_i)\\&=-(a\log_2 1+0+0.5b\log_2 0.5+0.5b\log_2 0.5+0.5c\log_2 0.5+0.5c\log_2 0.5)\\&=b+c=1-a\end{aligned}$$

The mutual information is

$$\begin{aligned}I(X;Y)&=H(Y)-H(Y|X)\\&=-0.5(1+a)\log_2 0.5(1+a)-0.5(1-a)\log_2 0.5(1-a)-(1-a)\end{aligned}$$

The channel capacity is the maximum of the mutual information. So, we need to find the input probability distribution that maximizes the mutual information by taking the derivative.

$$\begin{aligned}\frac{\partial}{\partial a}I(X;Y)=0&=-0.5\log_2 0.5(1+a)-0.5(1+a)\frac{1}{(1+a)\ln 2}\\&\quad+0.5\log_2 0.5(1-a)-0.5(1-a)\frac{-1}{(1-a)\ln 2}+1\\&=-0.5\log_2\frac{1-a}{1+a}+1\end{aligned}$$

Using the channel symmetry for B and C, we know that $b=c$. Since, $a+b+c=1$, we get $b=c=\frac{1}{5}$.

The capacity is

$$C=\max I(X;Y)=0.7219-0.4=0.3219 \text{ bits/transmission}$$