Solutions to Homework # 5

ECE 6605 Information Theory 10/27/03
Prof. Steven W McLaughlin

1. a) Following are the tables, the numbers labeling the tables refer to how many codewords have been encoded when this table is obtained:

Table 0

X	$C(X)$
α	1
β	2

Table 1

X	$C(X)$
α	1
β	2
$\alpha\alpha$	3

Table 2

X	$C(X)$
α	1
β	2
$\alpha\alpha$	3
$\alpha\alpha\alpha$	4

Table 3

X	$C(X)$
α	1
β	2
$\alpha\alpha$	3
$\alpha\alpha\alpha$	4
$\alpha\beta$	5

Table 4

X	$C(X)$
α	1
β	2
$\alpha\alpha$	3
$\alpha\alpha\alpha$	4
$\alpha\beta$	5
$\beta\beta$	6

Table 5

X	$C(X)$
α	1
β	2
$\alpha\alpha$	3
$\alpha\alpha\alpha$	4
$\alpha\beta$	5
$\beta\beta$	6
$\beta\alpha$	7

Table 6

X	$C(X)$
α	1
β	2
$\alpha\alpha$	3
$\alpha\alpha\alpha$	4
$\alpha\beta$	5
$\beta\beta$	6
$\beta\alpha$	7
$\alpha\beta\beta$	8

Table 7

X	$C(X)$
α	1
β	2
$\alpha\alpha$	3
$\alpha\alpha\alpha$	4
$\alpha\beta$	5
$\beta\beta$	6
$\beta\alpha$	7
$\alpha\beta\beta$	8
$\beta\beta\alpha$	9

Table 8

X	$C(X)$
α	1
β	2
$\alpha\alpha$	3
$\alpha\alpha\alpha$	4
$\alpha\beta$	5
$\beta\beta$	6
$\beta\alpha$	7
$\alpha\beta\beta$	8
$\beta\beta\alpha$	9
$\alpha\alpha\alpha\beta$	10

Table 9

X	$C(X)$
α	1
β	2
$\alpha\alpha$	3
$\alpha\alpha\alpha$	4
$\alpha\beta$	5
$\beta\beta$	6
$\beta\alpha$	7
$\alpha\beta\beta$	8
$\beta\beta\alpha$	9
$\alpha\alpha\alpha\beta$	10
$\beta\alpha\alpha$	11

The encoding then is as follows

$$\underbrace{\alpha}_{1}\underbrace{\alpha\alpha}_{3}\underbrace{\alpha}_{1}\underbrace{\beta}_{2}\underbrace{\beta}_{2}\underbrace{\alpha\beta}_{5}\underbrace{\beta\beta}_{6}\underbrace{\alpha\alpha\alpha}_{4}\underbrace{\beta\alpha}_{7}\underbrace{\alpha\alpha\alpha}_{4}$$

b) Once again we will give the tables as they are built and then go into the decoded word and the correct encoding.

Table 0

X	$C(X)$
c	1
d	2

Table 1

X	$C(X)$
c	1
d	2
$\text{c}\underline{\text{d}}$	3

Table 2

X	$C(X)$
c	1
d	2
$\text{c}\underline{\text{d}}$	3
$\text{d}\underline{\text{c}}$	4

Table 3

X	$C(X)$
c	1
d	2
$\text{c}\underline{\text{d}}$	3
$\text{d}\underline{\text{c}}$	4
$\text{c}\underline{\text{c}}$	5

Table 4

X	$C(X)$
c	1
d	2
$\text{c}\underline{\text{d}}$	3
$\text{d}\underline{\text{c}}$	4
$\text{c}\underline{\text{c}}$	5
$\text{cd}\underline{\text{d}}$	6

Table 5		Table 6		Table 7		Table 7 (corr)		Table 8 (corr)	
X	$C(X)$	X	$C(X)$	X	$C(X)$	X	$C(X)$	X	$C(X)$
c	1	c	1	c	1	c	1	c	1
d	2	d	2	d	2	d	2	d	2
$\text{c}\underline{\text{d}}$	3	$\text{c}\underline{\text{d}}$	3	$\text{c}\underline{\text{d}}$	3	$\text{c}\underline{\text{d}}$	3	$\text{c}\underline{\text{d}}$	3
$\text{d}\underline{\text{c}}$	4	$\text{d}\underline{\text{c}}$	4	$\text{d}\underline{\text{c}}$	4	$\text{d}\underline{\text{c}}$	4	$\text{d}\underline{\text{c}}$	4
$\text{c}\underline{\text{c}}$	5	$\text{c}\underline{\text{c}}$	5	$\text{c}\underline{\text{c}}$	5	$\text{c}\underline{\text{c}}$	5	$\text{c}\underline{\text{c}}$	5
$\text{cd}\underline{\text{d}}$	6	$\text{cd}\underline{\text{d}}$	6	$\text{cd}\underline{\text{d}}$	6	$\text{cd}\underline{\text{d}}$	6	$\text{cd}\underline{\text{d}}$	6
$\text{dc}\underline{\text{c}}$	7	$\text{dc}\underline{\text{c}}$	7	$\text{dc}\underline{\text{c}}$	7	$\text{dc}\underline{\text{c}}$	7	$\text{dc}\underline{\text{c}}$	7
		$\text{cc}\underline{\text{c}}$	8	$\text{cc}\underline{\text{c}}$	8	$\text{cc}\underline{\text{c}}$	8	$\text{cc}\underline{\text{c}}$	8
				$\text{dcc}\underline{\text{c}}$	9	$\text{cd}\underline{\text{c}}$	9	$\text{cd}\underline{\text{c}}$	9
								$\text{ccc}\underline{\text{d}}$	10

Therefore the decoded message will be

$$\underbrace{c}_{1}\underbrace{d}_{2}\underbrace{c}_{1}\underbrace{cd}_{3}\underbrace{dc}_{4}\underbrace{cc}_{5}\underbrace{c}_{1}\underbrace{dcc}_{7}\underbrace{cd}_{3}$$

However the correct encoding (obtained using table 7 & 8 in their corrected versions) should have been

$$\underbrace{c}_{1}\underbrace{d}_{2}\underbrace{c}_{1}\underbrace{cd}_{3}\underbrace{dc}_{4}\underbrace{cc}_{5}\underbrace{cd}_{3}\underbrace{ccc}_{8}\underbrace{d}_{2}$$

2. To obtain the codeword for the source output, we start with the interval $[0,1]$ and divide it, select the interval of the first source output, expand it and then divide it again as shown in figure 1.

   We then use the midpoint of the last segment, which is 0.012471805, find its binary representation and truncate it to $\lceil -\log_2 p' \rceil + 1$ where $p' = 0.01250461 - 0.012439$ bits, this gives us 15 bits and therefore the codeword is 000000110011000

   The entropy of a symbol is $H(0.1) = 0.469$ bits/symbol, therefore the entropy of the encoded sequence is $8(0.469) = 3.752$ bits, however we are using 15 bits.

   Using $L_n = \frac{1}{n}H(X_1,\ldots,X_n) + \frac{2}{n}$ we get that the average codeword length for eight symbols is $L_n = 0.469 + \frac{2}{8} = 0.719$ bits. For this particular codeword we have $\frac{15}{8} = 1.875$ bits/symbol

3. We need to find $C = \max_{p(x)} I(X;Y) = H(Y) - H(Y|X)$. We first calculate $H(Y|X)$ using the data from the transition matrix, let $Pr(X=0) = \pi$ and $Pr(X=1) = 1-\pi$, we have:

$$\begin{aligned}
H(Y|X) &= \sum_{X\in 0,1} p(x) \sum_{Y\in 0,1} p(y|x)\log_2 1/p(y|x) \\
&= \sum_{X\in 0,1} p(x)H(Y|x) \\
&= \pi[1\log_2 1 + 0\log_2 0] + (1-\pi)2\frac{1}{2}\log_2\frac{1}{2} \\
&= 1-\pi
\end{aligned}$$

   Now we also know that
   $Pr(Y=0) = Pr(Y=0|X=0)Pr(X=0) + Pr(Y=0|X=1)Pr(X=1) = \pi + \frac{1}{2}(1-\pi) = \frac{1+\pi}{2}$ and

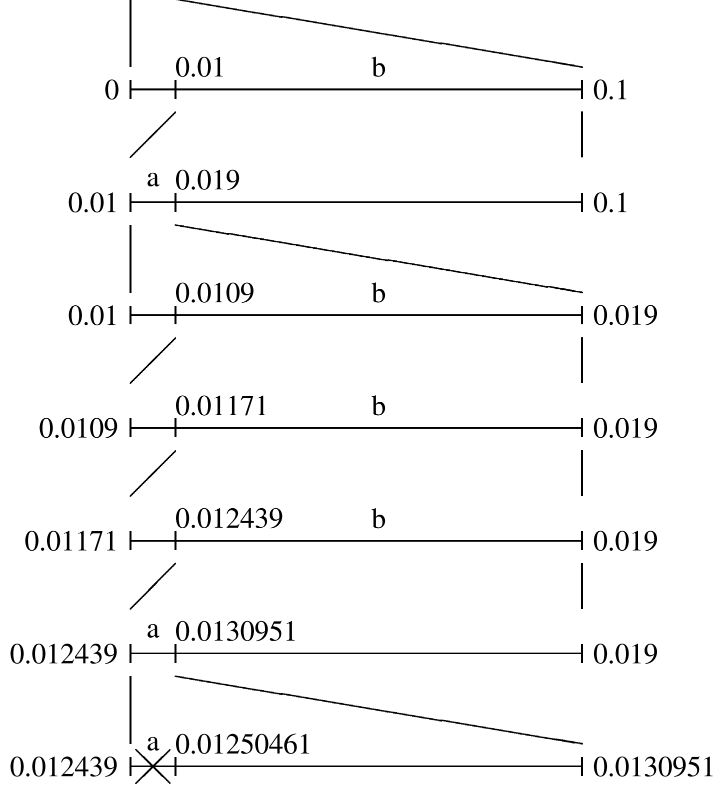

Figure 1: Arithmetic Coding procedure

$Pr(Y=1) = Pr(Y=1|X=0)Pr(X=0) + Pr(Y=1|X=1)Pr(X=1) = 0 + \frac{1-\pi}{2} = \frac{1-\pi}{2}$. Therefore we get that $H(Y) = H(\frac{1-\pi}{2})$. Then if we label $p = 1-\pi$ we get that $I(X;Y) = H(\frac{p}{2}) - p$. Then to find C we need to maximize over $p(x)$, we then take the derivative and set it equal to zero we get:

$$\frac{\partial I(X;Y)}{\partial p} = \frac{1}{2}\log_2 \frac{1-\frac{p}{2}}{\frac{p}{2}} - 1$$

Then to maximize we set $\frac{\partial I(X;Y)}{\partial p} = 0$ and we get

$$\begin{aligned}
\frac{1}{2}\log_2 \frac{1-\frac{p}{2}}{\frac{p}{2}} - 1 &= 0 \\
\log_2 \frac{1-\frac{p}{2}}{\frac{p}{2}} &= 2 \\
\frac{1-\frac{p}{2}}{\frac{p}{2}} &= 4 \\
1-\frac{p}{2} &= 2p \\
1 &= \frac{5p}{2} \\
\frac{2}{5} &= p
\end{aligned}$$

This is reasonable, since $Pr(X=1)=p$ is the probability of the noisy input to the channel (remember, 0 is received perfectly). Then we can calculate capacity in bits as $H(1/5)-2/5=0.722-0.4=0.322$.

4. In this problem we have to consider special cases of the value a.

$a=0$ In this case the channel reduces to a BSC with crossover probability $p=0$, therefore $C=1$.

$a=\pm 1$ In both of these cases the channel has 3 outputs, either $\{0,1,2\}$ if $a=1$ or $\{-1,0,1\}$ if $a=-1$. In both cases the channel becomes a Binary Erasure Channel, which was discussed in class, and $C=1-\alpha$. In our particular case $\alpha=1/2$ and therefore $C=1/2$.

$a\in\mathscr{R}-\{0,\pm 1\}$ In this case if we know Y we know which X was sent and therefore $H(X|Y)=0$. This means that $\max_{p(x)} I(X;Y)$ becomes $\max_{p(x)} H(X)$ which is clearly equal to one and therefore $C=1$

5. Assume that the X_i are chosen i.i.d. with $p=1/2$ then

$$\begin{aligned}
\frac{1}{n}I(X_1,X_2,\ldots,X_n;Y_1,Y_2,\ldots,Y_n) &= \frac{1}{n}[H(X_1,X_2,\ldots,X_n)-H(X_1,X_2,\ldots,X_n|Y_1,Y_2,\ldots,Y_n)] \\
&= \frac{1}{n}[H(X_1,X_2,\ldots,X_n)-H(Z_1,Z_2,\ldots,Z_n|Y_1,Y_2,\ldots,Y_n)] \\
&\geq \frac{1}{n}[H((X_1,X_2,\ldots,X_n)-H(Z_1,Z_2,\ldots,Z_n)] \\
&\geq \frac{1}{n}[H(X_1)\sum_{i=1}^{n}1-\sum_{i=1}^{n}H(Z_i)] \\
&\geq 1-H(p)
\end{aligned}$$

where we used for the second equality the equation $Y_i=X_i\oplus Z_i$ which tells us those entropies are the same and for the first inequality the fact that conditioning reduces entropy. For the second inequality we used $H(X_1,X_2,\ldots,X_n)\geq\sum_{i=1}^{n}H(X_i)$ with equality iff the X_i are independent.

Now we know that

$$\max_{p(X_1,X_2,\ldots,X_n)} I(X_1,X_2,\ldots,X_n;Y_1,Y_2,\ldots,Y_n)\geq I(X_1,X_2,\ldots,X_n;Y_1,Y_2,\ldots,Y_n)_{p(x_1,x_2,\ldots,x_n)=\text{Bern}(\frac{1}{2})}$$

and since $C=\frac{1}{n}\max_{p(X_1,X_2,\ldots,X_n)} I(X_1,X_2,\ldots,X_n;Y_1,Y_2,\ldots,Y_n)$ we have $C\geq 1-H(p)$ thus channels with memory have higher capacity. Intuitively we can explain it by observing that the memory implies that there will be correlation between the noise samples, we could then use a detector that extracts the information from the past samples to combat the present noise, thereby increasing capacity.