CSC 310, Spring 2002 — Solutions to Assignment #3

Question 1: 15 marks.

First, we need to find the output probabilities, by considering all ways of getting a particular output:

$$\begin{aligned} q_{-1} &= p_{-1}P_{-1,-1} + p_0P_{0,-1} + p_{+1}P_{+1,-1} \;=\; p_{-1} + p_0/3 \\ q_0 &= p_{-1}P_{-1,0} + p_0P_{0,0} + p_{+1}P_{+1,0} \;=\; p_0/3 \\ q_{+1} &= p_{-1}P_{-1,+1} + p_0P_{0,+1} + p_{+1}P_{+1,+1} \;=\; p_0/3 + p_{+1} \end{aligned}$$

We now find the backward probabilities using Bayes' Rule: $Q_{ij} = p_iP_{ij}/q_j$:

$$Q_{-1,-1} = \frac{p_{-1}}{p_{-1}+p_0/3},\quad Q_{0,-1} = \frac{p_0/3}{p_{-1}+p_0/3},\quad Q_{+1,-1} = 0$$

$$Q_{-1,0} = 0,\quad Q_{0,0} = 1,\quad Q_{+1,0} = 0$$

$$Q_{-1,+1} = 0,\quad Q_{0,-1} = \frac{p_0/3}{p_{+1}+p_0/3},\quad Q_{+1,-1} = \frac{p_{+1}}{p_{+1}+p_0/3}$$

Question 2: 15 marks.

We get $H(\mathcal{B}\,|\,\mathcal{A})$ by averaging the entropy of the output given a specific input with respect to the distribution of inputs:

$$H(\mathcal{B}\,|\,\mathcal{A}) = p_{-1}H(\mathcal{B}\,|\,a{=}-1) + p_0H(\mathcal{B}\,|\,a{=}0) + p_{+1}H(\mathcal{B}\,|\,a{=}+1)$$

Since the output is always equal to the input when the input is ±1, $H(\mathcal{B}\,|\,a=-1) = H(\mathcal{B}\,|\,a=+1) = 0$. The three output symbols are equally likely when the input is 0, so $H(\mathcal{B}\,|\,a{=}0) = \log(3)$.

The answer is therefore $H(\mathcal{B}\,|\,\mathcal{A}) = p_0\log(3)$. The base of the log could be whatever you choose, but I'll use base 2 logs below.

Question 3: 15 marks.

$H(\mathcal{B}) = -q_{-1}\log(q_{-1}) - q_0\log(q_0) - q_{+1}\log(q_{+1})$. Substituting the values for q_{-1}, q_0, and q_{+1} found in the answer to Question 1, we find that

$$H(\mathcal{B}) = -\,(p_{-1}+p_0/3)\log(p_{-1}+p_0/3) - (p_0/3)\log(p_0/3) - (p_{+1}+p_0/3)\log(p_{+1}+p_0/3)$$

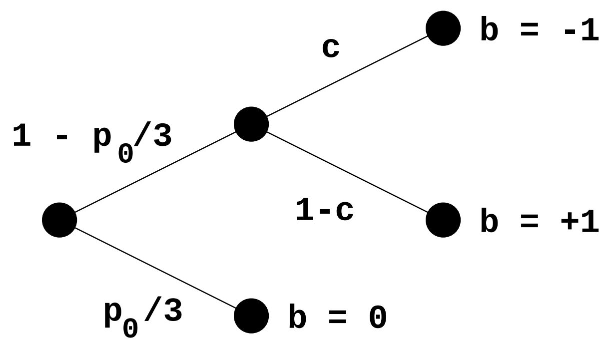
Question 4: 20 marks.

We can use the formula $I(\mathcal{A},\mathcal{B}) = H(\mathcal{B}) - H(\mathcal{B}\,|\,\mathcal{A})$. We now consider p_0 to be fixed, and ask what values of p_{-1} and p_{+1} maximize $I(\mathcal{A},\mathcal{B})$. Note that $p_{-1} = 1 - p_0 - p_{+1}$.

With p_0 fixed, $H(\mathcal{B} \mid \mathcal{A})$ is fixed, since according to Question 2, it is equal to $p_0 \log(3)$. So maximizing $I(\mathcal{A}, \mathcal{B})$ is the same as maximizing $H(\mathcal{B})$, which we found in Question 3.

The brute-force approach to maximizing $H(\mathcal{B})$ is to replace p_{-1} by $1 - p_0 - p_{+1}$ in the formula from Question 3, differentiate with respect to p_{+1}, and solve for the derivative being zero, as it should be at the maximum. (One would also need to check that the maximum doesn't instead occur at one of the extreme points where p_{+1} or p_{-1} is zero.)

An easier way is to note that the output distribution given by q_{-1}, q_0, and q_{+1} can be viewed in terms of a two step decision process, described by the following tree:



where $c = (p_{-1} + p_0/3) \,/\, (1 - p_0/3)$. The entropy of this distribution is the entropy of the first decision, which is $H(p_0/3)$, plus $1 - p_0/3$ times the entropy of the second decision, which is $H(c)$. The only part of this that depends on p_{-1} and p_{+1} is $H(c)$, so the values for p_{-1} and p_{+1} that maximize $H(\mathcal{B})$ are the values that maximize $H(c)$. When $p_{-1} = p_{+1} = (1 - p_0)/2$, we get $c = 1/2$. Since $H(1/2) = 1$ is the maximum possible value for $H(c)$, we see that $H(c)$ and hence $H(\mathcal{B})$ are maximized when $p_{-1} = p_{+1}$.

Question 5: 35 marks.

The channel capacity is the maximum value of $I(\mathcal{A}, \mathcal{B})$ for any input distribution. We know from Question 4 that the input distribution that produces the maximum value will have $p_{-1} = p_{+1} = (1 - p_0)/2$. Substituting for these will leave p_0 as the only variable we need to maximize with respect to.

Using the results from Questions 2 and 3, we get that

$$\begin{aligned}
I(\mathcal{A}, \mathcal{B}) &= H(\mathcal{B}) - H(\mathcal{B} \mid \mathcal{A}) \\
&= -\,(p_{-1} + p_0/3) \log(p_{-1} + p_0/3) \;-\; (p_0/3) \log(p_0/3) \;-\; (p_{+1} + p_0/3) \log(p_{+1} + p_0/3) \\
&\quad -\, p_0 \log(3) \\
&= -\,(1/2 - p_0/6) \log(1/2 - p_0/6) \;-\; (p_0/3) \log(p_0/3) \;-\; (1/2 - p_0/6) \log(1/2 - p_0/6) \\
&\quad -\, p_0 \log(3)
\end{aligned}$$

Taking the derivative (noting that the derivative of $x \log x$ is $1 + \log x$), we get

$$\begin{aligned}
dI(\mathcal{A}, \mathcal{B})/dp_0 &= (1 + \log(1/2 - p_0/6))/6 - (1 + \log(p_0/3))/3 + (1 + \log(1/2 - p_0/6))/6 - \log(3) \\
&= (1 + \log(1/2 - p_0/6))/3 - (1 + \log(p_0/3))/3 - \log(3)
\end{aligned}$$

Equating this to zero in order to find the maximum, we get that

$$\begin{aligned}
(1+\log(1/2-p_0/6))/3 &= (1+\log(p_0/3))/3+\log(3) \\
\log(1/2-p_0/6)) &= \log(p_0/3))+3\log(3) \\
1/2-p_0/6 &= 3^3 p_0/3 \\
1/2 &= (9+1/6)p_0 \\
p_0 &= 3/55 \;=\; 0.0545
\end{aligned}$$

From this, we find that $p_{-1} = p_{+1} = (1-p_0)/2 = 0.4727$.

With these values for p_0, p_{-1}, and p_{+1}, we get that $I(\mathcal{A},\mathcal{B}) = 1.0265$ bits (using base two logs). This is slightly greater than when the extreme value of $p_0 = 0$ is used (one bit), and much greater than when $p_0 = 1$ is used (zero bits), so this is indeed the maximum of $I(\mathcal{A},\mathcal{B})$, and therefore the capacity of the channel.

Note that $p_0 = 0$ corresponds to just using the inputs -1 and $+1$, which are transmitted without error. By using the input of 0, we can increase the capacity, even though this input is transmitted very unreliably, though the gain in this case is quite small.