S-72.340 INFORMATION THEORY

Exercise 6, October 25, 2000 Solutions

1. The symmetry of the channel can be utilized here. At first, we can write:

$$I(X;Y) = I(Y;X) = H(Y) - H(Y|X)\,.$$

The elements of the transition probability matrix $\mathbf{P}(y|x)$ are given by

$$\mathbf{P}(y|x)_{i,j} = P(Y = j|X = i) = \begin{cases} 1-p & \text{when } i = j \\ \frac{p}{M-1} & \text{when } i \neq j\,. \end{cases}$$

It is easy to see that the rows of the matrix, which represent conditional distributions of Y with different values of X, are permutations of each other and therefore have the same entropy $H(\mathbf{r})$.

$$\begin{aligned} H(Y|X) &= -\sum_{x,y} P(x,y)\log P(y|x) \\ &= -\sum_{x} P(x)\Big[\sum_{y} P(y|x)\log P(y|x)\Big] \\ &= H(\mathbf{r})\cdot\sum_{x} P(x) = H(\mathbf{r}) \\ &= -(M-1)\cdot\frac{p}{M-1}\log\frac{p}{M-1} - (1-p)\log(1-p) \\ &= -p\log\frac{p}{M-1} - (1-p)\log(1-p)\,. \end{aligned}$$

Further, $H(Y) \leq \log M$, with equality if and only if all the output symbols are equally probable. This is achieved if the probability of each input symbol is $1/M$, since in that case, for arbitrary $y \in \{0,1,\ldots,M-1\}$,

$$\begin{aligned} P(y) &= \sum_{x} P(x)P(y|x) \\ &= \frac{1}{M}\cdot(1-p) + (M-1)\cdot\frac{1}{M}\cdot\frac{p}{M-1} \\ &= \frac{1}{M}\,. \end{aligned}$$

So, we get the capacity of the channel:

$$\begin{aligned} C &= \log M - H(\mathbf{r}) \\ &= \log M + p \log \frac{p}{M-1} + (1-p)\log(1-p)\,, \end{aligned}$$

which is the same result as was obtained for $I(X;Y)$ in Problem 4 of Exercise 5.

2. The distribution $\mathbf{p} = \{p_0, p_1, p_2\}$ of the input symbols must be determined so that $I(X;Y)$ is maximized.

   By the Bayes theorem,

$$\begin{aligned} P_{X|Y}(0|0) &= \frac{P_{XY}(0,0)}{P_Y(0)} = \frac{P_X(0)P_{Y|X}(0|0)}{P_X(0)P_{Y|X}(0|0) + P_X(2)P_{Y|X}(0|2)} \\ &= \frac{p_0 \cdot \frac{1}{3}}{p_0 \cdot \frac{1}{3} + p_2 \cdot \frac{1}{3}} = \frac{p_0}{p_0 + p_2}\,. \end{aligned}$$

   In the same way, we get the other conditional probabilities $P(x|y)$:

$$\begin{aligned} &P(0|1) = \frac{p_0}{p_0+p_1} &&P(0|3) = p_0 &&P(1|1) = \frac{p_1}{p_0+p_1} \\ &P(1|2) = \frac{p_1}{p_1+p_2} &&P(1|3) = p_1 &&P(2|0) = \frac{p_2}{p_0+p_2} \\ &P(2|2) = \frac{p_2}{p_1+p_2} &&P(2|3) = p_2\,. \end{aligned}$$

   The average mutual information is given by

$$\begin{aligned} I(\mathbf{p}) = I(X;Y) &= \sum_{x,y} P(x,y) \log \frac{P(x|y)}{P(x)} \\ &= -\frac{p_0}{3}\Big[\log(p_0+p_2) + \log(p_0+p_1)\Big] \\ &\quad - \frac{p_1}{3}\Big[\log(p_0+p_1) + \log(p_1+p_2)\Big] \\ &\quad - \frac{p_2}{3}\Big[\log(p_0+p_2) + \log(p_1+p_2)\Big]\,. \end{aligned}$$

   Next, we define the Lagrangian function $L(\mathbf{p}, \lambda)$ as

$$L(\mathbf{p}, \lambda) = -3I(\mathbf{p}) + \lambda \cdot (p_0 + p_1 + p_2 - 1)\,.$$

During differentiation of L, we assume that the logarithms are natural.

$$\begin{aligned}
\frac{\partial L}{\partial p_0} &= \log(p_0+p_2) + \log(p_0+p_1) + \frac{p_0}{p_0+p_2} + \frac{p_0}{p_0+p_1} \\
&\quad + \frac{p_1}{p_0+p_1} + \frac{p_2}{p_0+p_2} + \lambda = 0 \\
&\quad \Rightarrow\ \lambda = -2 - \log(p_0+p_1) - \log(p_0+p_2)\,. \\
\frac{\partial L}{\partial p_1} &= 2 + \log(p_0+p_1) + \log(p_1+p_2) + \lambda = 0 \\
&\quad \Rightarrow\ \log(p_0+p_2) = \log(p_1+p_2)\ \Rightarrow\ p_0 = p_1 \\
\frac{\partial L}{\partial p_2} &= 2 + \log(p_0+p_2) + \log(p_1+p_2) + \lambda = 0 \\
&\quad \Rightarrow\ \cdots \Rightarrow\ p_0 = p_2 \\
&\quad \Rightarrow\ p_0 = p_1 = p_2 = \frac{1}{3}\,.
\end{aligned}$$

So, in this case too, the capacity of the channel turns out to be the same as the average mutual information obtained in Exercise 5:

$$C = \frac{2}{3}\log\frac{3}{2}\,.$$