S-72.340 INFORMATION THEORY

Exercise 7, October 31, 2000 Solutions

1. Consider the function $f(t) = -t \log t$. By differentiating twice, we get

$$f''(t) = -\frac{1}{t \log e} < 0 ,$$

that is, $f(t)$ is a strictly concave function. Thus, for all t_1, t_2 and for all $\lambda \in [0,1]$,

$$f(\lambda t_1 + (1-\lambda)\, t_2) \geq \lambda f(t_1) + (1-\lambda) f(t_2) ,$$

with equality only if $\lambda = 0$ or $\lambda = 1$. From this, it follows that, for all x and for all $\lambda \in [0,1]$,

$$\begin{aligned}
& - \Big[\lambda Q_1(x) + (1-\lambda) Q_2(x)\Big] \log \Big[\lambda Q_1(x) + (1-\lambda) Q_2(x)\Big] \\
& \geq \lambda \Big[-Q_1(x) \log Q_1(x)\Big] + (1-\lambda) \Big[-Q_2(x) \log Q_2(x)\Big] \\
\Rightarrow & -\sum_x Q(x) \log Q(x) \geq \lambda \Big[-\sum_x Q_1(x) \log Q_1(x)\Big] + (1-\lambda) \Big[-\sum_x Q_2(x) \log Q_2(x)\Big] \\
\Rightarrow & \, H(X) \geq \lambda H_1(X) + (1-\lambda) H_2(X) ,
\end{aligned}$$

with equality only if $\lambda = 0$ or $\lambda = 1$.

2. Denote $g_1(t) = t(2-t)$ and $g_2(t) = -(t+1)^2$. These functions are both concave, since

$$g_1''(t) = g_2''(t) = -2 < 0\,.$$

Now consider two points, $\boldsymbol{\alpha}_1 = (\alpha_{1,1}, \alpha_{1,2})$ and $\boldsymbol{\alpha}_2 = (\alpha_{2,1}, \alpha_{2,2})$, and let $\lambda \in [0,1]$.

$$\begin{aligned}
&\lambda f(\boldsymbol{\alpha}_1) + (1-\lambda) f(\boldsymbol{\alpha}_2)\\
=&\lambda\left[\alpha_{1,1}(2-\alpha_{1,1}) - (\alpha_{1,2}+1)^2\right] + (1-\lambda)\left[\alpha_{2,1}(2-\alpha_{2,1}) - (\alpha_{2,2}+1)^2\right]\\
=&\lambda\alpha_{1,1}(2-\alpha_{1,1}) + (1-\lambda)\,\alpha_{2,1}(2-\alpha_{2,1})\\
-&\lambda(\alpha_{1,2}+1)^2 - (1-\lambda)(\alpha_{2,2}+1)^2\\
=&\lambda g_1(\alpha_{1,1}) + (1-\lambda) g_1(\alpha_{2,1}) + \lambda g_2(\alpha_{1,2}) + (1-\lambda) g_2(\alpha_{2,2})\\
\le& g_1\left[\lambda\alpha_{1,1} + (1-\lambda)\,\alpha_{2,1}\right] + g_2\left[\lambda\alpha_{1,2} + (1-\lambda)\,\alpha_{2,2}\right]\\
=& f\left[\lambda\boldsymbol{\alpha}_1 + (1-\lambda)\boldsymbol{\alpha}_2\right],
\end{aligned}$$

and thus f is a concave function.

3. At first, we try to find a point where partial derivatives with respect to both variables are equal to 0.

$$\begin{aligned}
\frac{\partial f}{\alpha_1} &= 2 - 2\alpha_1 = 0 \Rightarrow \alpha_1 = 1\,.\\
\frac{\partial f}{\alpha_2} &= -2(\alpha_2 + 1) = 0 \Rightarrow \alpha_2 = -1 < 0\,.
\end{aligned}$$

Since there is no such point in the region, the maximum of the function in this region must occur on the boundary. The boundary consists of the non-negative α_2-axis, where $\alpha_1 = 0$, and the non-negative α_1-axis, where $\alpha_2 = 0$. The maximum occurs at the point where the partial derivative with respect to the non-zero variable is equal to 0 and the partial derivative with respect to the other variable is negative. In this case, the point where the maximum occurs is $(1,0)$, and the maximum value of f is

$$f[(1,0)] = 1 \cdot (2-1) - (0+1)^2 = 0\,.$$

4. The probability vector $\mathbf{p} = (p_0, p_1)$ must be determined so that the average mutual information $I(X;Y)$ is maximized. According to Theorem 4.4.1 of the lectures, assuming that $p_0\,,\, p_1 \neq 0$, the capacity C can be found by solving

$$I(x=0;Y) = I(x=1;Y) = C\,,$$

with the additional condition: $p_0 + p_1 = 1$. The binary entropy function

$$H(p) = -p\log p - (1-p)\log(1-p)$$

will be used.

$$\begin{aligned}
I(x=0;Y) &= \sum_y P(y|0)\log\frac{P(y|0)}{P(y)} \\
&= (1-\epsilon)\log\frac{1-\epsilon}{p_0(1-\epsilon)+p_1\delta} + \epsilon\log\frac{\epsilon}{p_0\epsilon + p_1(1-\delta)}\,. \\
I(x=1;Y) &= \delta\log\frac{\delta}{p_0(1-\epsilon)+p_1\delta} + (1-\delta)\log\frac{1-\delta}{p_0\epsilon+p_1(1-\delta)}\,. \\
I(x=0;Y) &= I(x=1;Y)
\end{aligned}$$

$$\begin{aligned}
&\Rightarrow (1-\epsilon)\log(1-\epsilon) - (1-\epsilon)\log\left[p_0(1-\epsilon)+p_1\delta\right] + \epsilon\log\epsilon \\
&\quad -\epsilon\log\left[p_0\epsilon + p_1(1-\delta)\right] - \delta\log\delta + \delta\log\left[p_0(1-\epsilon)+p_1\delta\right] \\
&\quad -(1-\delta)\log(1-\delta) + (1-\delta)\log\left[p_0\epsilon+p_1(1-\delta)\right] = 0 \\
&\Rightarrow (\delta+\epsilon-1)\log\left[p_0(1-\epsilon)+p_1\delta\right] + (1-\delta-\epsilon)\log\left[p_0\epsilon+p_1(1-\delta)\right] = H(\epsilon)-H(\delta) \\
&\Rightarrow \log\frac{p_0(1-\epsilon)+p_1\delta}{p_0\epsilon+p_1(1-\delta)} = \frac{H(\epsilon)-H(\delta)}{\delta+\epsilon-1} \\
&\Rightarrow \frac{p_0(1-\epsilon)+p_1\delta}{p_0\epsilon+p_1(1-\delta)} = 2^{\frac{H(\epsilon)-H(\delta)}{\delta+\epsilon-1}} \\
&\Rightarrow \frac{(1-\delta-\epsilon)p_0+\delta}{(\delta+\epsilon-1)p_0+1-\delta} = 2^{\frac{H(\epsilon)-H(\delta)}{\delta+\epsilon-1}} \\
&\Rightarrow (1-\delta-\epsilon)p_0 + \delta = (\delta+\epsilon-1)2^{\frac{H(\epsilon)-H(\delta)}{\delta+\epsilon-1}}p_0 + (1-\delta)2^{\frac{H(\epsilon)-H(\delta)}{\delta+\epsilon-1}} \\
&\Rightarrow p_0 = \frac{(1-\delta)2^{\frac{H(\epsilon)-H(\delta)}{\delta+\epsilon-1}} - \delta}{(1-\delta-\epsilon)\left(1+2^{\frac{H(\epsilon)-H(\delta)}{\delta+\epsilon-1}}\right)} \\
&\quad\; p_1 = 1-p_0 = \frac{1-\epsilon\left(1+2^{\frac{H(\epsilon)-H(\delta)}{\delta+\epsilon-1}}\right)}{(1-\delta-\epsilon)\left(1+2^{\frac{H(\epsilon)-H(\delta)}{\delta+\epsilon-1}}\right)}\,.
\end{aligned}$$

The capacity can now be obtained in a straightforward way by using the expression for $I(x=0;Y)$ or $I(x=1;Y)$. For example,

$$\begin{aligned} C &= I(x=0;Y) \\ &= (1-\epsilon)\log(1-\epsilon) + \epsilon\log\epsilon \\ &- (1-\epsilon)\log\left[p_0(1-\epsilon)+p_1\delta\right] - \epsilon\log\left[p_0\epsilon + p_1(1-\delta)\right]. \end{aligned}$$

After some algebraic manipulations, we get

$$C = \log\left(1 + 2^{\frac{H(\epsilon)-H(\delta)}{\delta+\epsilon-1}}\right) + \frac{(1-\epsilon)H(\delta) - \delta H(\epsilon)}{\delta+\epsilon-1}.$$

If $\epsilon = \delta$, we have a binary symmetric channel (BSC), and we get the familiar result

$$C = 1 - H(\epsilon) = 1 - H(\delta).$$