Homework 7 (2023) — SOLUTIONS

JAIST — School of Information Science — I232 Information Theory

1. Consider a commander of an army besieged in a fort for whom the only means of communication to his allies is a set of carrier pigeons. Assume that each carrier pigeon can carry 8 bits of information (i.e. one byte), that pigeons are released once every 5 minutes and that each pigeon takes 3 minutes to reach its destination.
   (a) Assume that all pigeons reach his allies safely. What is the capacity of this link in bits/hour?
   (b) Now assume that the enemy tries to shoot down the pigeon, and they manage to hit a fraction α of them. Since the pigeons are sent at a constant rate, the allies know when the pigeons are missing. What is the capacity of this link in bits/hour?
   (c) Now assume that every time the enemy shoots down a pigeon, they send out a dummy pigeon with a uniform random 8-bit sequence. What is the capacity of this link, in bits/hour?

Solution:

(a) 12 pigeons reach the destination per hour, so the capacity is:

$$\frac{60}{5} \cdot 8 = 96 \text{ bits/hour}$$

(b) The capacity of the binary erasure channel is $1 - \alpha$ bits per channel use. According to (a), the capacity is:

$$96(1 - \alpha) \text{ bits/hour}$$

(c) We know that the enemy shoots down a pigeon with probability α. Given that each bird is carrying 8 bits of information, we have 256 possible messages for each pigeon sent. When the enemy replaces the pigeon by a dummy one, there is a small probability 1/256 that the dummy pigeon will have the same message as the original message. Thus, the chance of receiving the original information is $1 - \alpha + \alpha/256$. If the replaced pigeon does not contain the same message as the shot-down pigeon, the probability of getting any of the the 255 wrong messages is uniform $\frac{\alpha}{256}$. As a result, we can get a symmetric channel matrix. Then we can compute the capacity as follows:

$$\begin{aligned}
C &= \max_{p(y)} I(\mathsf{X};\mathsf{Y}) \\
&= H(\mathsf{Y}) - H(\mathsf{Y}|\mathsf{X}) \\
&= \log_2(256) - H(\mathbf{r}) \\
&= 12(8 - H\left(1 - \alpha + \alpha/256, \frac{\alpha}{256}, \frac{\alpha}{256}, \frac{\alpha}{256}, \ldots, \frac{\alpha}{256}\right)) \\
&= 96 - 12h(\frac{255}{256}\alpha) - 12 \cdot \frac{255}{256}\alpha \log 255
\end{aligned}$$

2. *Errors and Erasures Channel.* Consider a 2-input, 3-output DMC, with $\mathcal{X} = \{0, 1\}$ and $\mathcal{Y} = \{0, ?, 1\}$. Let the probability of error be α and let the probability of erasure be ϵ, so the channel conditional probabilities are:

$$p_{\mathsf{Y}|\mathsf{X}}(y|x) = \begin{bmatrix} 1 - \alpha - \epsilon & \epsilon & \alpha \\ \alpha & \epsilon & 1 - \alpha - \epsilon \end{bmatrix}$$

Assume that the capacity achieving input distribution is $p_{\mathsf{X}}(0) = p_{\mathsf{X}}(1) = \frac{1}{2}$.

   (a) Find the capacity of the errors and erasures channel.

Solution: The capacity-achieving input distribution is $p_{\mathsf{X}}(0) = p_{\mathsf{X}}(1) = \frac{1}{2}$.

$$
\begin{aligned}
I(\mathsf{X};\mathsf{Y}) &= H(\mathsf{Y}) - H(\mathsf{Y}|\mathsf{X}). \\
\text{Preparation:} \quad p_{\mathsf{Y}}(0) = p_{\mathsf{Y}}(1) &= \frac{1}{2}(1-\epsilon-\alpha) + \frac{1}{2}\alpha = \frac{1}{2}(1-\epsilon) \\
p_{\mathsf{Y}}(?) &= \frac{1}{2}\epsilon + \frac{1}{2}\epsilon = \epsilon \\
\text{So,} \quad H(\mathsf{Y}) &= -\sum_{y\in\mathcal{Y}} p_{\mathsf{Y}}(y)\log p_{\mathsf{Y}}(y) \\
&= -[2\Big(\frac{1}{2}(1-\epsilon)\log\frac{1}{2}(1-\epsilon)\Big) + \epsilon\log\epsilon] \\
&= -(1-\epsilon)\log(1-\epsilon) + (1-\epsilon) - \epsilon\log\epsilon \\
&\text{Alternate answer:} (1-\epsilon) + h(\epsilon)
\end{aligned}
$$

$$
\begin{aligned}
H(\mathsf{Y}|\mathsf{X}) &= p_{\mathsf{X}}(0)H(\mathsf{Y}|\mathsf{X}=0) + p_{\mathsf{X}}(1)H(\mathsf{Y}|\mathsf{X}=1) \\
&= \frac{1}{2}h(1-\epsilon-\alpha,\epsilon,\alpha) + \frac{1}{2}h(\alpha,\epsilon,1-\epsilon-\alpha) = h(\alpha,\epsilon,1-\epsilon-\alpha) \\
&= -\alpha\log\alpha - \epsilon\log\epsilon - (1-\epsilon-\alpha)\log(1-\epsilon-\alpha) \\
&\text{Alternate answer:} (1-\epsilon)h(\frac{\alpha}{1-\epsilon}) + h(\epsilon) \\
\text{So,} \quad C &= H(\mathsf{Y}) - H(\mathsf{Y}|\mathsf{X}) \\
&= -(1-\epsilon)\log(1-\epsilon) + (1-\epsilon) - \epsilon\log\epsilon \\
&\quad - \Big(-\alpha\log\alpha - \epsilon\log\epsilon - (1-\epsilon-\alpha)\log(1-\epsilon-\alpha)\Big) \\
&= -(1-\epsilon)\log(1-\epsilon) + (1-\epsilon) + \alpha\log\alpha + (1-\epsilon-\alpha)\log(1-\epsilon-\alpha) \\
&\text{Alternate answer: } (1-\epsilon)(1-h(\frac{\alpha}{1-\epsilon}))
\end{aligned}
$$

(b) Verify that if $\epsilon = 0$, the capacity of the BSC is obtained.

Solution: Evaluating the above expression with $\epsilon = 0$, we have:

$$
\begin{aligned}
C &= 1 + \alpha\log\alpha + (1-\alpha)\log(1-\alpha) \\
&= 1 - h(\alpha)
\end{aligned}
$$

which is the capacity of the BSC with error probability α.

(c) Verify that if $\alpha = 0$, the capacity of the BEC is obtained.

Solution: With $\alpha = 0$ and $0\log 0 = 0$, we have:

$$C = 1 - \epsilon$$

which is the capacity of the BEC.

3. Find the capacity and the capacity-achieving input distribution of the following channels.

(a)

$$
p_{\mathsf{Y}|\mathsf{X}}(y|x) = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}
$$

(b)

$$p_{\mathsf{Y}|\mathsf{X}}(y|x) = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix}$$

(c)

$$p_{\mathsf{Y}|\mathsf{X}}(y|x) = \begin{bmatrix} p & 1-p & 0 & 0 \\ 1-p & p & 0 & 0 \\ 0 & 0 & q & 1-q \\ 0 & 0 & 1-q & q \end{bmatrix}$$

Solution:

(a)

$$\begin{aligned} I(\mathsf{X};\mathsf{Y}) &= H(\mathsf{Y}) - H(\mathsf{Y}|\mathsf{X}) \\ &= H(\mathsf{Y}) - \sum_{x\in\mathcal{X}} H(\mathsf{Y}|\mathsf{X}=x)p_{\mathsf{X}}(x) \end{aligned}$$

It should be clear that $H(\mathsf{Y}) = \log 3$, independent of the input distribution. And, $H(\mathsf{Y}|\mathsf{X} = x) = \log 3$ for all x. Then:

$$\begin{aligned} I(\mathsf{X};\mathsf{Y}) &= \log 3 - \log 3 \sum_{x\in\mathcal{X}} p_{\mathsf{X}}(x) \\ C &= 0 \end{aligned}$$

(b) $H(\mathsf{Y}|\mathsf{X} = x)$ is independent of x, and is equal to $h_3(0.5, 0.5, 0)$. Then:

$$\begin{aligned} I(\mathsf{X};\mathsf{Y}) &= H(\mathsf{Y}) - H(\mathsf{Y}|\mathsf{X}) \\ &\le \log 3 - h_3(0.5, 0.5, 0) \end{aligned}$$

Note that the *columns* of $p_{\mathsf{Y}|\mathsf{X}}(y|x)$ sum to a constant 1. If X is uniform $p_{\mathsf{X}}(x) = \frac{1}{3}$, then

$$p_{\mathsf{Y}}(y) = \sum_{x\in\mathcal{X}} p_{\mathsf{Y}|\mathsf{X}}(y|x)p_{\mathsf{X}}(x) = \frac{1}{3}\sum_{x\in\mathcal{X}} p_{\mathsf{Y}|\mathsf{X}}(y|x) = \frac{1}{3}$$

That is, if X is uniformly distributed, then Y is uniformly distributed.

To maximize mutual information $I(\mathsf{X};\mathsf{Y}) \le \log 3 - h_3(0.5, 0.5, 0)$, choose Y to be uniformly distributed, so:

$$\begin{aligned} C &= \log_3 - h_3(0.5, 0.5, 0) \\ &= 1.585 - 1 \\ &= 0.585 \text{ bits} \end{aligned}$$

(c) This is two parallel BSCs, the p channel and the q channel At any time, one of the two channels is used. Create a new random variable Z:

$$p_{\mathsf{Z}}(z) = \begin{cases} \alpha & z = \text{“use p channel”} \\ 1-\alpha & z = \text{“use q channel”} \end{cases}$$

Because z can be determined from the channel output y, $\mathsf{X} \to \mathsf{Y} \to \mathsf{Z}$. Write mutual information $I(\mathsf{X};\mathsf{Y},\mathsf{Z})$ two ways:

$$\begin{aligned} I(\mathsf{X};\mathsf{Y},\mathsf{Z}) &= I(\mathsf{X};\mathsf{Z}) + I(\mathsf{X};\mathsf{Y}|\mathsf{Z}) \\ I(\mathsf{X};\mathsf{Y},\mathsf{Z}) &= I(\mathsf{X};\mathsf{Y}) + I(\mathsf{X};\mathsf{Z}|\mathsf{Y}) \end{aligned}$$

Since $\mathsf{X} \to \mathsf{Y} \to \mathsf{Z}$, $I(\mathsf{X};\mathsf{Z}|\mathsf{Y}) = 0$, so:

$$\begin{aligned}
I(\mathsf{X};\mathsf{Y}) &= I(\mathsf{X};\mathsf{Z}) + I(\mathsf{X};\mathsf{Y}|\mathsf{Z}) \\
&= H(\mathsf{Z}) - H(\mathsf{Z}|\mathsf{X}) + \alpha I(\mathsf{X}_{\mathrm{p}};\mathsf{Y}_{\mathrm{p}}) + (1-\alpha) I(\mathsf{X}_{\mathrm{q}};\mathsf{Y}_{\mathrm{q}}) \\
&= h(\alpha) + \alpha I(\mathsf{X}_{\mathrm{p}};\mathsf{Y}_{\mathrm{p}}) + (1-\alpha) I(\mathsf{X}_{\mathrm{q}};\mathsf{Y}_{\mathrm{q}}) \\
C &= \max_{\alpha} h(\alpha) + \alpha C_{\mathrm{p}} + (1-\alpha) C_{\mathrm{q}},
\end{aligned}$$

where $C_{\mathrm{p}} = 1 - h(p)$ and $C_{\mathrm{q}} = 1 - h(q)$ are the capacities of the two BSCs. The solution, found by taking the derivative with respect to α, is:

$$\alpha^* = \frac{2^{C_{\mathrm{p}}}}{2^{C_{\mathrm{p}}} + 2^{C_{\mathrm{q}}}}.$$

Then, the capacity is:

$$C = \log\left(2^{1-h(p)} + 2^{1-h(q)}\right).$$