# 5 Discrete Memoryless Channels

## 5.1 Homework 7

- Chapter 7, Problems 3,7,8,9,10,11
- Supplementary Problems:
  1. Let $X$ and $Y$ be the input and output of a BSC. Show that if $\epsilon = 0.5$, then $X$ and $Y$ are independent.
  2. Show in Example 7.8 that $H(Y \mid E) = (1-\gamma) h_b(a)$

**3. Memory increases capacity.** Consider a BSC with crossover probability $0 < \epsilon < 1$ represented by $X_i = Y_i + Z_i \bmod 2$, where $X_i, Y_i$, and $Z_i$ are respectively, the input, the output, and the noise variable at time $i$. Then

$$\operatorname{Pr}\{Z_i = 0\} = 1-\epsilon \quad \text{and} \quad \operatorname{Pr}\{Z_i = 1\} = \epsilon$$

for all $i$. We assume that $\{X_i\}$ and $\{Z_i\}$ are independent, but we make no assumption that $Z_i$ are i.i.d. so that the channel may have memory.
a) Prove that

$$I(\mathbf{X}; \mathbf{Y}) \leq n - h_b(\epsilon)$$

b) Show that the upper bound in (a) can be achieved by letting $X_i$ be i.i.d. bits taking the values 0 and 1 with equal probability and $Z_1 = Z_2 = \cdots = Z_n$
c) Show that with the assumptions in (b), $I(\mathbf{X}; \mathbf{Y}) > nC$, where $C = 1 - h_b(\epsilon)$ is the capacity of the BSC if it is memoryless.
Typo: $Y_i = X_i + Z_i, \qquad X_i + Y_i = (X_i + X_i) + Z_i = 0 + Z_i = Z_i)$

**Answer** (a)

$$\begin{aligned}
I(\mathbf{X}; \mathbf{Y}) &= H(\mathbf{Y}) - H(\mathbf{Y} \mid \mathbf{X}) \\
&\leq \sum_{i=1}^{n} H(Y_i) - \sum_{i=1}^{n} H\left(Y_i \mid \mathbf{Y}^{i-1}, \mathbf{X}\right) \\
&\leq n \cdot 1 - H(Y_1 \mid \mathbf{X}) \\
&= n - h_b(\epsilon)
\end{aligned}$$

(b) In order to achieve this upper bound, we have to 1) make $H(\mathbf{Y}) = \sum_{i=1}^{n} H(Y_i)$ and $H(Y_i) = 1$, i.e., the output distribution of the BSC is uniform. This can be done by letting $p(X_i)$ be the uniform distribution on $\{0, 1\}$. 2) $H\left(Y_i \mid \mathbf{Y}^{i-1}, \mathbf{X}\right) = 0, i \geq 2$, that is the random variable $Z_i, i \geq 2$ are fixed (same as $Z_1$).
(c) With assumption in b holds

$$I(\mathbf{X}; \mathbf{Y}) = n - h_b(\epsilon) \geq n - n h_b(\epsilon) = n(1 - h_b(\epsilon)) = nC$$

7 . Let

$$P(\epsilon) = \left[\begin{array}{cc} 1-\epsilon & \epsilon \\ \epsilon & 1-\epsilon \end{array}\right]$$

be the transition matrix for a BSC with crossover probability $\epsilon$. Define $a * b = (1-a)b + a(1-b)$ for $0 \leq a, b \leq 1$
a) Prove that a DMC with transition matrix $P(\epsilon_1) P(\epsilon_2)$ is equivalent to a BSC with crossover probability $\epsilon_1 * \epsilon_2$. Such a channel is the cascade of two BSCs with crossover probabilities $\epsilon_1$ and $\epsilon_2$ respectively.
b) Repeat (a) for a DMC with transition matrix $P(\epsilon_2) P(\epsilon_1)$.
c) Prove that

$$1 - h_b(\epsilon_1 * \epsilon_2) \leq \min(1 - h_b(\epsilon_1), 1 - h_b(\epsilon_2))$$

This means that the capacity of the cascade of two BSCs is upper bounded by the capacity of either of the two BSCs.
d) Prove that a DMC with transition matrix $P(\epsilon)^n$ is equivalent to a BSC with crossover probabilities $\frac{1}{2}(1 - (1 - 2\epsilon)^n)$

**Answer** (a) The composite transition probability is

$$p(Y|X) = \sum_{Z=0,1} p(Y|Z)p(Z|X)$$

Given $p(Y|Z) = P(\epsilon_2), p(Z|X) = P(\epsilon_1)$ the above could be write down in matrix format, that is

$$\begin{aligned} p(\epsilon) = P(\epsilon_2)P(\epsilon_1) &= \begin{bmatrix} 1-\epsilon_2 & \epsilon_2 \\ \epsilon_2 & 1-\epsilon_2 \end{bmatrix} \begin{bmatrix} 1-\epsilon_1 & \epsilon_1 \\ \epsilon_1 & 1-\epsilon_1 \end{bmatrix} \\ &= \begin{bmatrix} (1-\epsilon_2)(1-\epsilon_1)+\epsilon_1 * \epsilon_2 & (1-\epsilon_2)\epsilon_1+(1-\epsilon_1)\epsilon_2 \\ (1-\epsilon_2)\epsilon_1+(1-\epsilon_1)\epsilon_2 & (1-\epsilon_2)(1-\epsilon_1)+\epsilon_1 * \epsilon_2 \end{bmatrix} \\ &= \begin{bmatrix} (1-\epsilon_2)(1-\epsilon_1)+\epsilon_1 * \epsilon_2 & (1-\epsilon_2)\epsilon_1+(1-\epsilon_1)\epsilon_2 \\ (1-\epsilon_2)\epsilon_1+(1-\epsilon_1)\epsilon_2 & (1-\epsilon_2)(1-\epsilon_1)+\epsilon_1 * \epsilon_2 \end{bmatrix} \end{aligned}$$

Note that, denoting $\epsilon^* = (1-\epsilon_2)\epsilon_1 + (1-\epsilon_1)\epsilon_2$, and

$$1-\epsilon^* = 1-((1-\epsilon_2)\epsilon_1+(1-\epsilon_1)\epsilon_2) = (1-\epsilon_2)(1-\epsilon_1)+\epsilon_1 * \epsilon_2$$

We obtain

$$p(\epsilon^*) = \begin{bmatrix} (1-\epsilon_2)(1-\epsilon_1)+\epsilon_1 * \epsilon_2 & (1-\epsilon_2)\epsilon_1+(1-\epsilon_1)\epsilon_2 \\ (1-\epsilon_2)\epsilon_1+(1-\epsilon_1)\epsilon_2 & (1-\epsilon_2)(1-\epsilon_1)+\epsilon_1 * \epsilon_2 \end{bmatrix} = \begin{bmatrix} 1-\epsilon^* & \epsilon^* \\ \epsilon^* & 1-\epsilon^* \end{bmatrix}$$

therefore equivalent to a BSC with crossover probability $\epsilon^* = \epsilon_1 * \epsilon_2$
(b) As seen from the above matrix multiplication, the formula is symmetric about $\epsilon_1$ and $\epsilon_2$. Therefore the conclusion holds for $p(\epsilon_1)p(\epsilon_2)$
(c)

$$C_1 = 1-H_b(\epsilon_1) \qquad C_2 = 1-H_b(\epsilon_2) \qquad C_3 = 1-H_b(\epsilon_1 * \epsilon_2)$$

From Markov Chain $X-Z-Y$ we can see that

$$C_3 = I(X;Y) \leq I(X;Z,Y) = I(X;Z) = C_1$$

$$C_3 = I(X;Y) \leq I(X,Z;Y) = I(Z;Y) = C_2$$

Thus

$$C_3 = \min \leq \{C_1, C_2\}$$

which is exactly same as

$$1-h_b(\epsilon_1 * \epsilon_2) \leq \min(1-h_b(\epsilon_1), 1-h_b(\epsilon_2))$$

(d) For $n=1$, the $\frac{1}{2}(1-(1-2\epsilon)^n) = \epsilon$ trivially holds. Suppose it holds for $n = k-1, k \geq 2$, then for $n = k$,

$$\begin{aligned} \Pr(Y_k = 1 \mid X = 0) &= \Pr(Y_k = 1 \mid Y_{k-1} = 0, X = 0)\Pr(Y_{k-1} = 0 \mid X = 0) + \\ &\quad \Pr(Y_k = 1 \mid Y_{k-1} = 1, X = 0)\Pr(Y_{k-1} = 1 \mid X = 0) \\ &= \Pr(Y_k = 1 \mid Y_{k-1} = 0)\Pr(Y_{k-1} = 0 \mid X = 0) + \Pr(Y_k = 1 \mid Y_{k-1} = 1)\Pr(Y_{k-1} = 1 \mid X = 0) \\ &= \epsilon\left(1 - \frac{1}{2}\left(1-(1-2\epsilon)^{k-1}\right)\right) + (1-\epsilon)\frac{1}{2}\left(1-(1-2\epsilon)^{k-1}\right) \\ &= \frac{1}{2}\left(1-(1-2\epsilon)^k\right) \end{aligned}$$

We proved by mathematical induction.

**8. Symmetric channel.** A DMC is symmetric if the rows of the transition matrix $p(y \mid x)$ are permutations of each other and so are the columns. Determine the capacity of such a channel.
See Section 4.5 in Gallager [129] for a more general discussion.

**Answer** Note that

$$\begin{aligned} H(Y \mid X) &= -\sum_x \sum_y P(y \mid x) P(x) \log P(y \mid x) \\ &= -\sum_x P(x) \left( \sum_y P(y \mid x) \log P(y \mid x) \right) \end{aligned}$$

Note that by symmetric assumption, $H(\Pi) = -\sum_y P(y \mid x) \log P(y \mid x)$ is independent of $x$. Thus,

$$\begin{aligned} H(Y \mid X) &= -\sum_x P(x) \left( \sum_y P(y \mid x) \log P(y \mid x) \right) \\ &= \sum_x P(x) H(\Pi) = H(\Pi) \end{aligned}$$

Since the capacity is

$$\begin{aligned} C = \max_x I(X;Y) &= \max_x H(Y) - H(Y|X) \\ &= \max_x H(Y) - H(\Pi) \end{aligned}$$

Recall that In Theorem 2.43, we have proved that for any random variable

$$H(X) \leq \log |\mathcal{X}|$$

Therefore

$$C \leq \log |\mathcal{Y}| - H(\Pi)$$

where $H(\Pi) = -\sum_y P(y \mid x) \log P(y \mid x)$ and the equality is attained when $Y$ is uniform.

9. Let $C_1$ and $C_2$ be the capacities of two DMCs with transition matrices $P_1$ and $P_2$, respectively, and let $C$ be the capacity of the DMC with transition matrix $P_1 P_2$. Prove that $C \leq \min(C_1, C_2)$

**Answer** Think of the constructed Markov Chain $X - Z - Y$, where $Z$ denoted the DMC in the middle phase

$$C_3 = I(X;Y) \leq I(X;Z,Y) = I(X;Z) = C_1$$

$$C_3 = I(X;Y) \leq I(X,Z;Y) = I(Z;Y) = C_2$$

Similar to proof in Problem 7.

$$C_3 = \min \leq \{C_1, C_2\}$$

**10. Two parallel channels.** Let $C_1$ and $C_2$ be the capacities of two DMCs $p_1(y_1 \mid x_1)$ and $p_2(y_2 \mid x_2)$, respectively. Determine the capacity of the DMC

$$p(y_1, y_2 \mid x_1, x_2) = p_1(y_1 \mid x_1) p_2(y_2 \mid x_2)$$

Hint: Prove that

$$I(X_1, X_2; Y_1, Y_2) \leq I(X_1; Y_1) + I(X_2; Y_2)$$

if $p(y_1, y_2 | x_1, x_2) = p_1(y_1 | x_1) p_2(y_2 | x_2)$

**Answer**

$$
\begin{aligned}
I(X_1, X_2; Y_1, Y_2) &= \sum_{x_1,x_2,y_1,y_2} p(x_1, x_2, y_1, y_2) \log \frac{p(x_1, x_2, y_1, y_2)}{p(x_1, x_2)p(y_1, y_2)} \\
&= \sum_{x_1,x_2,y_1,y_2} p(x_1, x_2, y_1, y_2) \log \frac{(y_1, y_2|x_1, x_2)\, p(x_1, x_2)}{p(x_1, x_2)p(y_1, y_2)} \\
&= \sum_{x_1,x_2,y_1,y_2} p(x_1, x_2, y_1, y_2) \log \frac{p_1\,(y_1|x_1)\, p_2\,(y_2|x_2)}{p(y_1, y_2)} \\
&= \sum_{x_1,x_2,y_1,y_2} p(x_1, x_2, y_1, y_2) \log p_1\,(y_1|x_1) + \sum_{x_1,x_2,y_1,y_2} p(x_1, x_2, y_1, y_2) \log p_2\,(y_2|x_2) - \\
&\quad \sum_{x_1,x_2,y_1,y_2} p(x_1, x_2, y_1, y_2) \log p(y_1, y_2) \\
&= \sum_{x_1,y_1} p(x_1, y_1) \log p_1\,(y_1|x_1) + \sum_{x_2,y_2} p(x_2, y_2) \log p_2\,(y_2|x_2) - \sum_{y_1,y_2} p(y_1, y_2) \log p_1\,(y_1, y_2) \\
&= I\,(X_1; Y_1) + I\,(X_2; Y_2) + H(Y_1, Y_2) - H(Y_1) - H(Y_2) \\
&\leq I\,(X_1; Y_1) + I\,(X_2; Y_2)
\end{aligned}
$$

Therefore we know that the new $C \leq C_1 + C_2$

11. In the system below, there are two channels with transition matrices $p_1\,(y_1 \mid x)$ and $p_2\,(y_2 \mid x)$. These two channels have a common input alphabet $\mathcal{X}$ and output alphabets $\mathcal{Y}_1$ and $\mathcal{Y}_2$, respectively, where $\mathcal{Y}_1$ and $\mathcal{Y}_2$ are disjoint. The position of the switch is determined by a random variable $Z$ which is independent of $X$, where $\Pr\{Z = 1\} = \lambda$
a) Show that

$$I(X; Y) = \lambda I\,(X; Y_1) + (1 - \lambda) I\,(X; Y_2)$$

b) The capacity of the system is given by $C = \max_{p(x)} I(X; Y)$. Show that $C \leq \lambda C_1 + (1 - \lambda) C_2$, where $C_i = \max_{p(x)} I\,(X; Y_i)$ is the capacity of the channel with transition matrix $p_i\,(y_i \mid x)\,, i = 1, 2$
c) If both $C_1$ and $C_2$ can be achieved by a common input distribution, show that $C = \lambda C_1 + (1 - \lambda) C_2$

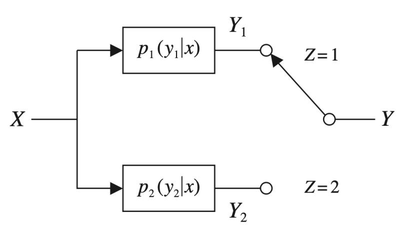


Figure 16: Problem 11.

**Answer** (a)

$$E(Y) = \lambda E(Y_1) + (1 - \lambda) E(Y_2)$$

$$
\begin{aligned}
I(X; Y) &= E \log \frac{p(X, Y)}{p(X)p(Y)} \\
&= E \log \frac{p(X, (Y_1, Y_2, Z))}{p(X)p(Y_1, Y_2, Z)} \\
&= E\left(\mathbb{1}_{Z=1} \log \frac{p(X, Y_1)}{p(X)p(Y_1)}\right) + E\left(\mathbb{1}_{Z=2} \log \frac{p(X, Y_2)}{p(X)p(Y_2)}\right) \\
&= \lambda I\,(X; Y_1) + (1 - \lambda) I\,(X; Y_2)
\end{aligned}
$$

(b)

$$
\begin{aligned}
\max_{p(x)} I(X;Y) &= \max_{p(x)} \lambda I\left(X;Y_1\right) + (1-\lambda) I\left(X;Y_2\right) \\
&\leq \max_{p(x)} \lambda I\left(X;Y_1\right) + \max_{p(x)} \lambda I\left(X;Y_2\right) \\
&= \lambda C_1 + (1-\lambda) C_2
\end{aligned}
$$

(c) The equality achieved when $\operatorname{argmax} I\left(X;Y_1\right) = \operatorname{argmax} I\left(X;Y_2\right)$, otherwise the two maximum values cannot be achieved simultaneously.

> Let $X$ and $Y$ be the input and output of a BSC. Show that if $\epsilon = 0.5$, then $X$ and $Y$ are independent.

**Answer**

$$
I(X;Y) = H(Y) - H(Y|X) = H(Y) - h_b(0.5) = 1 - 1 = 0
$$

Therefore, $X, Y$ independent

> Show in Example 7.8 that $H(Y \mid E) = (1-\gamma) h_b(a)$, where
>
> $$E = \begin{cases} 0 \text{ if } Y \neq e \\ 1 \text{ if } Y = e \end{cases}$$

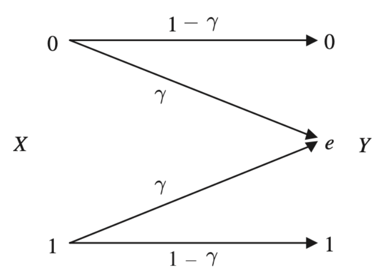


Figure 17: Example 7.8.

$$
\begin{aligned}
H(Y \mid E) = &-\sum_{e,y} p(e,y) \log \frac{p(e,y)}{p(e)} \\
&= -p(e=1, y=e) \log \frac{p(e=1, y=e)}{p(e=1)} - p(e=0, y=1) \log \frac{p(e=0, y=1)}{p(e=0)} - p(e=0, y=0) \log \frac{p(e=0, y=0)}{p(e=0)} \\
&= -0 - (1-a)(1-\gamma) \log(1-a) - a(1-\gamma) \log(a) \\
&= (1-\gamma) h_b(a)
\end{aligned}
$$