Solutions to Homework Set #4
Channel and Source coding

1. **Lossless source coding with side information.**
   Consider the lossless source coding with side information that is available at the encoder and decoder, where the source X and the side information Y are i.i.d. $\sim P_{X,Y}(x, y)$.

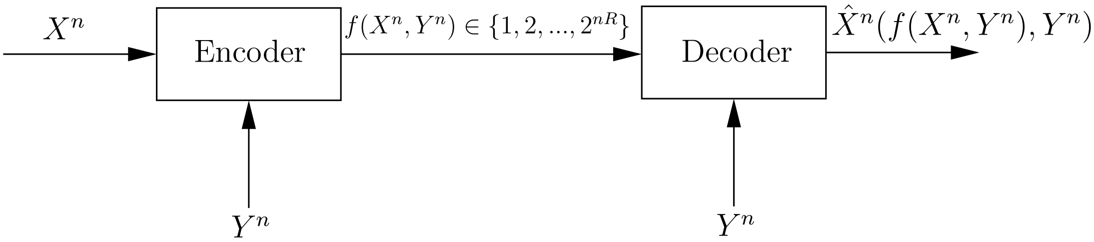


Figure 1: Lossless source coding with side information at the encoder and decoder.

Show that a code with rate $R < H(X|Y)$ can not be achievable, and interpret the result.

Hint: Let $T \triangleq f(X^n, Y^n)$. Consider

$$\begin{aligned} nR &\geq H(T) \\ &\geq H(T|Y^n), \end{aligned} \tag{1}$$

and use similar steps, including Fano's inequality, as we used in the class to prove the converse where side information was not available.

Solution Sketch of the solution (please fill in the explanation for each step):

$$\begin{aligned} nR &\geq H(T) \\ &\geq H(T|Y^n), \\ &\geq I(X^n; T|Y^n) \\ &= H(X^n|Y^n) - H(X^n|T, Y^n) \\ &= nH(X|Y) - \epsilon_n, \end{aligned}$$

where $\epsilon_n \to 0$.

2. **Preprocessing the output.**
   One is given a communication channel with transition probabilities $p(y \mid x)$ and channel capacity $C = \max_{p(x)} I(X;Y)$. A helpful statistician preprocesses the output by forming $\tilde{Y} = g(Y)$, yielding a channel $p(\tilde{y}|x)$. He claims that this will strictly improve the capacity.

   (a) Show that he is wrong.

   (b) Under what conditions does he not strictly decrease the capacity?

   Solution: Preprocessing the output.

   (a) The statistician calculates $\tilde{Y} = g(Y)$. Since $X \to Y \to \tilde{Y}$ forms a Markov chain, we can apply the data processing inequality. Hence for every distribution on x,

   $$I(X;Y) \geq I(X;\tilde{Y}).$$

   Let $\tilde{p}(x)$ be the distribution on x that maximizes $I(X;\tilde{Y})$. Then

   $$C = \max_{p(x)} I(X;Y) \geq I(X;Y)_{p(x)=\tilde{p}(x)} \geq I(X;\tilde{Y})_{p(x)=\tilde{p}(x)} = \max_{p(x)} I(X;\tilde{Y}) = \tilde{C}.$$

   Thus, the helpful suggestion is wrong and processing the output does not increase capacity.

   (b) We have equality (no decrease in capacity) in the above sequence of inequalities only if we have equality in data processing inequality, i.e., for the distribution that maximizes $I(X;\tilde{Y})$, we have $X \to \tilde{Y} \to Y$ forming a Markov chain. Thus, $\tilde{Y}$ should be a sufficient statistic.

3. **The Z channel.**
   The Z-channel has binary input and output alphabets and transition probabilities $p(y|x)$ given by the following matrix:

   $$Q = \begin{bmatrix} 1 & 0 \\ 1/2 & 1/2 \end{bmatrix} \qquad x, y \in \{0,1\}$$

   Find the capacity of the Z-channel and the maximizing input probability distribution.

Solution: The Z channel.
First we express $I(X;Y)$, the mutual information between the input and output of the Z-channel, as a function of $\alpha = \Pr(X = 1)$:

$$\begin{aligned} H(Y|X) &= \Pr(X=0)\cdot 0 + \Pr(X=1)\cdot 1 = \alpha \\ H(Y) &= \mathbf{H}(\Pr(Y=1)) = \mathbf{H}(\alpha/2) \\ I(X;Y) &= H(Y) - H(Y|X) = \mathbf{H}(\alpha/2) - \alpha \end{aligned}$$

Since $I(X;Y)$ is strictly concave on α (why?) and $I(X;Y) = 0$ when $\alpha = 0$ and $\alpha = 1$, the maximum mutual information is obtained for some value of α such that $0 < \alpha < 1$.

Using elementary calculus, we determine that

$$\frac{d}{d\alpha} I(X;Y) = \frac{1}{2}\log_2 \frac{1-\alpha/2}{\alpha/2} - 1\,,$$

which is equal to zero for $\alpha = 2/5$. (It is reasonable that $\Pr(X = 1) < 1/2$ since $X = 1$ is the noisy input to the channel.) So the capacity of the Z-channel in bits is $H(1/5) - 2/5 = 0.722 - 0.4 = 0.322$.

4. **Using two channels at once.**
Consider two discrete memoryless channels $(\mathcal{X}_1, p(y_1 \mid x_1), \mathcal{Y}_1)$ and $(\mathcal{X}_2, p(y_2 \mid x_2), \mathcal{Y}_2)$ with capacities C_1 and C_2 respectively. A new channel $(\mathcal{X}_1 \times \mathcal{X}_2, p(y_1 \mid x_1) \times p(y_2 \mid x_2), \mathcal{Y}_1 \times \mathcal{Y}_2)$ is formed in which $x_1 \in \mathcal{X}_1$ and $x_2 \in \mathcal{X}_2$, are *simultaneously* sent, resulting in y_1, y_2. Find the capacity of this channel.

Solution: Using two channels at once.
To find the capacity of the product channel $(\mathcal{X}_1 \times \mathcal{X}_2, p(y_1, y_2|x_1, x_2), \mathcal{Y}_1 \times \mathcal{Y}_2)$, we have to find the distribution $p(x_1, x_2)$ on the input alphabet $\mathcal{X}_1 \times \mathcal{X}_2$ that maximizes $I(X_1, X_2; Y_1, Y_2)$. Since the transition probabilities are given as $p(y_1, y_2|x_1, x_2) = p(y_1|x_1)p(y_2|x_2)$,

$$\begin{aligned} p(x_1, x_2, y_1, y_2) &= p(x_1, x_2)p(y_1, y_2|x_1, x_2) \\ &= p(x_1, x_2)p(y_1|x_1)p(y_2|x_2), \end{aligned}$$

Therefore, $Y_1 \to X_1 \to X_2 \to Y_2$ forms a Markov chain and

$$\begin{aligned}
I(X_1, X_2; Y_1, Y_2) &= H(Y_1, Y_2) - H(Y_1, Y_2|X_1, X_2) \\
&= H(Y_1, Y_2) - H(Y_1|X_1, X_2) - H(Y_2|X_1, X_2) \qquad (2) \\
&= H(Y_1, Y_2) - H(Y_1|X_1) - H(Y_2|X_2) \qquad (3) \\
&\le H(Y_1) + H(Y_2) - H(Y_1|X_1) - H(Y_2|X_2) \qquad (4) \\
&= I(X_1; Y_1) + I(X_2; Y_2),
\end{aligned}$$

where Eqs. (2) and (3) follow from Markovity, and Eq. (4) is met with equality if X_1 and X_2 are independent and hence Y_1 and Y_2 are independent. Therefore

$$\begin{aligned}
C &= \max_{p(x_1,x_2)} I(X_1, X_2; Y_1, Y_2) \\
&\le \max_{p(x_1,x_2)} I(X_1; Y_1) + \max_{p(x_1,x_2)} I(X_2; Y_2) \\
&= \max_{p(x_1)} I(X_1; Y_1) + \max_{p(x_2)} I(X_2; Y_2) \\
&= C_1 + C_2.
\end{aligned}$$

with equality iff $p(x_1, x_2) = p^*(x_1)p^*(x_2)$ and $p^*(x_1)$ and $p^*(x_2)$ are the distributions that maximize C_1 and C_2 respectively.

5. **A channel with two independent looks at Y.**
   Let Y_1 and Y_2 be conditionally independent and conditionally identically distributed given X. Thus $p(y_1, y_2|x) = p(y_1|x)p(y_2|x)$.

   (a) Show $I(X; Y_1, Y_2) = 2I(X; Y_1) - I(Y_1; Y_2)$.

   (b) Conclude that the capacity of the channel

   $$X \longrightarrow \boxed{} \longrightarrow (Y_1, Y_2)$$

   is less than twice the capacity of the channel

   $$X \longrightarrow \boxed{} \longrightarrow Y_1$$

Solution: A channel with two independent looks at Y.

(a)

$$
\begin{aligned}
I(X;Y_1,Y_2) &= H(Y_1,Y_2) - H(Y_1,Y_2|X) && (5)\\
&= H(Y_1) + H(Y_2) - I(Y_1;Y_2) - H(Y_1|X) - H(Y_2|X) && (6)\\
&\qquad \text{(since } Y_1 \text{ and } Y_2 \text{ are conditionally independent given } X\text{)} && (7)\\
&= I(X;Y_1) + I(X;Y_2) - I(Y_1;Y_2) && (8)\\
&= 2I(X;Y_1) - I(Y_1;Y_2) \quad \text{(since } Y_1 \text{ and } Y_2 \text{ are conditionally identically distributed)} && (9)
\end{aligned}
$$

(b) The capacity of the single look channel $X \to Y_1$ is

$$C_1 = \max_{p(x)} I(X;Y_1). \tag{10}$$

The capacity of the channel $X \to (Y_1, Y_2)$ is

$$
\begin{aligned}
C_2 &= \max_{p(x)} I(X;Y_1,Y_2) && (11)\\
&= \max_{p(x)} 2I(X;Y_1) - I(Y_1;Y_2) && (12)\\
&\leq \max_{p(x)} 2I(X;Y_1) && (13)\\
&= 2C_1. && (14)
\end{aligned}
$$

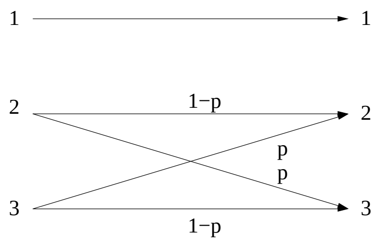
Hence, two independent looks cannot be more than twice as good as one look.

6. **Choice of channels.**
Find the capacity C of the union of 2 channels $(\mathcal{X}_1, p_1(y_1|x_1), \mathcal{Y}_1)$ and $(\mathcal{X}_2, p_2(y_2|x_2), \mathcal{Y}_2)$ where, at each time, one can send a symbol over channel 1 or over channel 2 but not both. Assume the output alphabets are distinct and do not intersect.

(a) Show $2^C = 2^{C_1} + 2^{C_2}$.

(b) What is the capacity of this Channel?

Solution: Choice of channels.



(a) Let

$$\theta = \begin{cases} 1, & \text{if the signal is sent over the channel 1} \\ 2, & \text{if the signal is sent over the channel 2} \end{cases}.$$

Consider the following communication scheme: The sender chooses between two channels according to Bern(α) coin flip. Then the channel input is $X = (\theta, X_\theta)$.
Since the output alphabets $\mathcal{Y}_1$ and $\mathcal{Y}_2$ are disjoint, θ is a function of Y, i.e. $X \to Y \to \theta$.
Therefore,

$$\begin{aligned} I(X;Y) &= I(X;Y,\theta) \\ &= I(X_\theta,\theta;Y,\theta) \\ &= I(\theta;Y,\theta) + I(X_\theta;Y,\theta|\theta) \\ &= I(\theta;Y,\theta) + I(X_\theta;Y|\theta) \\ &= H(\theta) + \alpha I(X_\theta;Y|\theta=1) + (1-\alpha)I(X_\theta;Y|\theta=2) \\ &= H(\alpha) + \alpha I(X_1;Y_1) + (1-\alpha)I(X_2;Y_2). \end{aligned}$$

Thus, it follows that

$$C = \sup_\alpha \left\{ H(\alpha) + \alpha C_1 + (1-\alpha)C_2 \right\},$$

which is a strictly concave function on α. Hence, the maximum exists and by elementary calculus, one can easily show $C = \log_2(2^{C_1} + 2^{C_2})$, which is attained with $\alpha = 2^{C_1}/(2^{C_1} + 2^{C_2})$.
If one interprets $M = 2^C$ as the effective number of noise free symbols, then the above result follows in a rather intuitive manner: we have $M_1 = 2^{C_1}$ noise free symbols from channel 1, and $M_2 =$

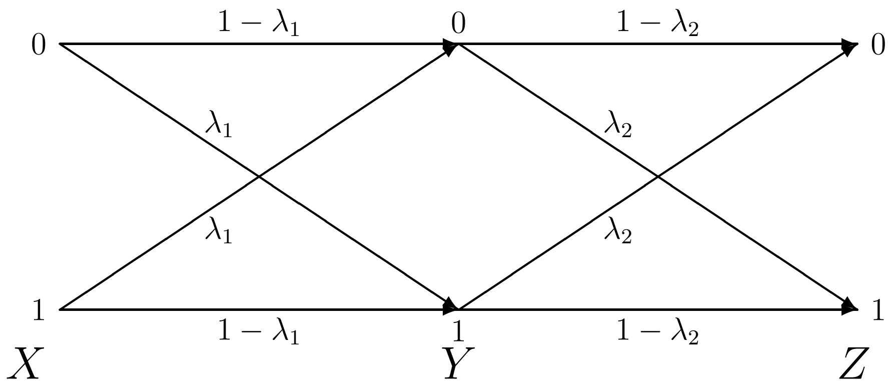
2^{C_2} noise free symbols from channel 2. Since at each step we get to choose which channel to use, we essentially have $M_1 + M_2 = 2^{C_1} + 2^{C_2}$ noise free symbols for the new channel. Therefore, the capacity of this channel is $C = \log_2(2^{C_1} + 2^{C_2})$.

This argument is very similiar to the effective alphabet argument given in Problem 19, Chapter 2 of the text.

(b) From part (b) we get capacity is

$$\log(2^{1-H(p)} + 2^0).$$

7. **Cascaded BSCs.**
Consider the two discrete memoryless channels $(\mathcal{X}, p_1(y|x), \mathcal{Y})$ and $(\mathcal{Y}, p_2(z|y), \mathcal{Z})$.
Let $p_1(y|x)$ and $p_2(z|y)$ be binary symmetric channels with crossover probabilities λ_1 and λ_2 respectively.



(a) What is the capacity C_1 of $p_1(y|x)$?

(b) What is the capacity C_2 of $p_2(z|y)$?

(c) We now cascade these channels. Thus $p_3(z|x) = \sum_y p_1(y|x)p_2(z|y)$. What is the capacity C_3 of $p_3(z|x)$? Show $C_3 \le \min\{C_1, C_2\}$.

(d) Now let us actively intervene between channels 1 and 2, rather than passively transmitting y^n. What is the capacity of channel 1 followed by channel 2 if you are allowed to decode the output y^n of channel 1 and then reencode it as $\tilde{y}^n$ for transmission over channel 2? (Think $W \longrightarrow x^n(W) \longrightarrow y^n \longrightarrow \tilde{y}^n(y^n) \longrightarrow z^n \longrightarrow \hat{W}$.)

(e) What is the capacity of the cascade in part c) if the receiver can view *both* Y and Z?

Solution: Cascaded channels.

(a) *Brute force method:* Let $C_1 = 1 - H(p)$ be the capacity of the BSC with parameter p, and $C_2 = 1 - \alpha$ be the capacity of the BEC with parameter α. Let $\tilde{Y}$ denote the output of the cascaded channel, and Y the output of the BSC. Then, the transition rule for the cascaded channel is simply

$$p(\tilde{y}|x) = \sum_{y=0,1} p(\tilde{y}|y)p(y|x)$$

for each $(x, \tilde{y})$ pair.
Let $X \sim$Bern(π) denote the input to the channel. Then,

$$H(\tilde{Y}) = H((1-\alpha)(\pi(1-p)+p(1-\pi)),\, \alpha,\, (1-\alpha)(p\pi+(1-p)(1-\pi)))$$

and also

$$\begin{aligned}
H(\tilde{Y}|X=0) &= H((1-\alpha)(1-p),\, \alpha,\, (1-\alpha)p) \\
H(\tilde{Y}|X=1) &= H((1-\alpha)p,\, \alpha,\, (1-\alpha)(1-p)) \;=\; H(\tilde{Y}|X=0).
\end{aligned}$$

Therefore,

$$\begin{aligned}
C &= \max_{p(x)} I(X;\tilde{Y}) \\
&= \max_{p(x)} [H(\tilde{Y}) - H(\tilde{Y}|X)] \\
&= \max_{p(x)} [H(\tilde{Y})] - H(\tilde{Y}|X) \\
&= \max_{\pi} [H((1-\alpha)(\pi(1-p)+p(1-\pi)),\, \alpha,\, (1-\alpha)(p\pi+(1-p)(1-\pi)))] \\
&\quad - H((1-\alpha)(1-p),\, \alpha,\, (1-\alpha)p). \qquad (15)
\end{aligned}$$

Note that the maximum value of $H(\tilde{Y})$ occurs when $\pi = 1/2$ by the concavity and symmetry of $H(\cdot)$. (We can check this also by differentiating Eq. (15) with respect to π.)

Substituting the value $\pi = 1/2$ in the expression for the capacity yields

$$\begin{aligned} C &= H((1-\alpha)/2,\, \alpha,\, (1-\alpha)/2) - H((1-p)(1-\alpha),\, \alpha,\, p(1-\alpha)) \\ &= (1-\alpha)(1 + (1-p)\log(1-p) + p\log p) \\ &= C_1 C_2. \end{aligned}$$

(b) *Elegant method:*
For the cascade of an arbitrary discrete memoryless channel (with capacity C) with the erasure channel (with the erasure probability α), we will show that

$$I(X;\tilde{Y}) = (1-\alpha)I(X;Y). \tag{16}$$

Then, by taking suprema of both sides over all input distributions $p(x)$, we can conclude the capacity of the cascaded channel is $(1-\alpha)C$.
Proof of Eq. (16):
Let

$$E = \begin{cases} 1, & \tilde{Y} = e \\ 0, & \tilde{Y} = Y \end{cases}.$$

Then, since E is a function of Y,

$$\begin{aligned} H(\tilde{Y}) &= H(\tilde{Y}, E) \\ &= H(E) + H(\tilde{Y}|E) \\ &= H(\alpha) + \alpha H(\tilde{Y}|E=1) + (1-\alpha)H(\tilde{Y}|E=0) \\ &= H(\alpha) + (1-\alpha)H(Y), \end{aligned}$$

where the last equality comes directly from the construction of E. Similarly,

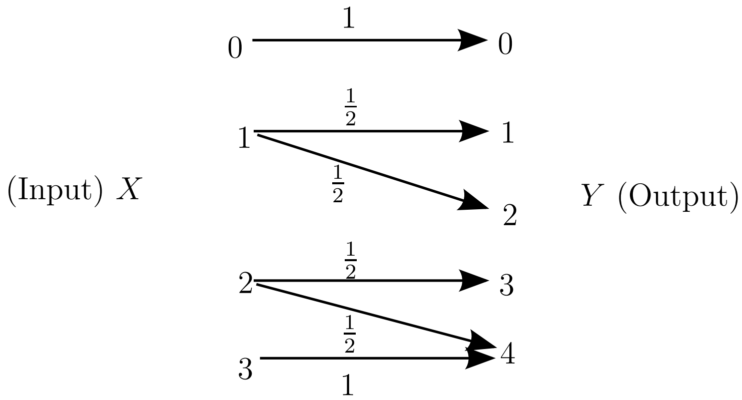
$$\begin{aligned} H(\tilde{Y}|X) &= H(\tilde{Y}, E|X) \\ &= H(E|X) + H(\tilde{Y}|X, E) \\ &= H(E) + \alpha H(\tilde{Y}|X, E=1) + (1-\alpha)H(\tilde{Y}|X, E=0) \\ &= H(\alpha) + (1-\alpha)H(Y|X), \end{aligned}$$

whence

$$I(X;\tilde{Y}) = H(\tilde{Y}) - H(\tilde{Y}|X) = (1-\alpha)I(X;Y).$$

8. **Channel capacity**

(a) What is the capacity of the following channel



(b) Provide a simple scheme that can transmit at rate $R = \log_2 3$ bits through this channel.

Solution for Channel capacity

(a) We can use the solution of previous home question:

$$C = \log\left(2^{C_1} + 2^{C_2} + 2^{C_3}\right)$$

Now we need to calculate the capacity of each channel:

$$C_1 = \max_{p(x)} I(X;Y) = H(Y) - H(Y|X) = 0 - 0 = 0$$

$$C_2 = \max_{p(x)} I(X;Y) = H(Y) - H(Y|X) = 1 - 1 = 0$$

$$\begin{aligned} C_3 = \max_{p(x)} I(X;Y) &= \max_{p(x)} \{H(Y) - H(Y|X)\} \\ &= \max_{p(x)} \left[-\frac{1}{2}p_2 \log\left(\frac{1}{2}p_2\right) - \left(\frac{1}{2}p_2 + p_3\right)\log\left(\frac{1}{2}p_2 + p_3\right)\right] - p_2 \end{aligned}$$

Assigning $p_3 = 1 - p_2$ and derive against p_2:

$$\frac{dI(X;Y)}{dp_2} = -\frac{p_2}{2} \cdot \frac{1}{2} \cdot \frac{1}{\frac{p_2}{2}} - \frac{1}{2}\log\left(\frac{p_2}{2}\right) + \frac{2-p_2}{2} \cdot \frac{1}{2} \cdot \frac{1}{\frac{2-p_2}{2}} + \frac{1}{2}\log\left(\frac{2-p_2}{2}\right) - 1 = 0$$

And as result $p_2 = \frac{2}{5}$:

$$C_3 \approx 0.322$$

And, finally:

$$C = \log(2^0 + 2^0 + 2^{0.322}) \approx 1.7$$

(b) Here is a simple code that achieves capacity.
Encoding: You just use ternary representation of the message and send using 0,1,2 but no 3 (or 0,1,3 but no 2) of the input channel.
Decoding: map the ternary output into the message.