EE276: Homework #4 Solutions

Due on Friday Feb 9, 5pm

1. **Channel capacity.**
   Find the capacity of the following channels with given probability transition matrices, where the element p_{ij} of the matrix represents $p(y = j|x = i)$:

   (a) $\mathcal{X} = \mathcal{Y} = \{0, 1, 2\}$

   $$p(y|x) = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{bmatrix}$$

   (b) $\mathcal{X} = \mathcal{Y} = \{0, 1, 2\}$

   $$p(y|x) = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \end{bmatrix}$$

   (c) $\mathcal{X} = \mathcal{Y} = \{0, 1\}$ (the Z-channel)

   $$p(y|x) = \begin{bmatrix} 1 & 0 \\ 1/2 & 1/2 \end{bmatrix}$$

   Solution:

   (a) $\mathcal{X} = \mathcal{Y} = \{0, 1, 2\}$

   $$p(y|x) = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{bmatrix}$$

   This is a symmetric channel and by the known result for symmetric channel (please read section 7.2 in Cover & Thomas to understand this properly. The basic idea is similar to that used for the BSC):

   $$C = \log |\mathcal{Y}| - H(\mathbf{r}) = \log 3 - \log 3 = 0. \tag{1}$$

   In this case, the output is independent of the input.

   (b) $\mathcal{X} = \mathcal{Y} = \{0, 1, 2\}$

   $$p(y|x) = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \end{bmatrix}$$

   Again the channel is symmetric:

   $$C = \log |\mathcal{Y}| - H(\mathbf{r}) = \log 3 - \log 2 = 0.58 \text{ bits}$$

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

$$\begin{aligned} H(Y|X) &= \Pr(X=0)\cdot 0 + \Pr(X=1)\cdot 1 = \alpha \\ H(Y) &= h_2(\Pr(Y=1)) = h_2(\alpha/2) \\ I(X;Y) &= H(Y) - H(Y|X) = h_2(\alpha/2) - \alpha \end{aligned}$$

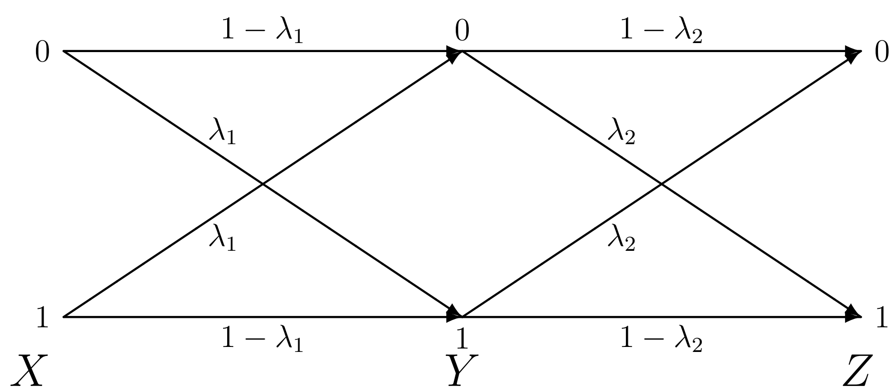
Since $I(X;Y)$ is strictly concave on α (why?) and $I(X;Y)=0$ when $\alpha=0$ and $\alpha=1$, the maximum mutual information is obtained for some value of α such that $0<\alpha<1$.

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 $h_2(1/5) - 2/5 = 0.722 - 0.4 = 0.322$.

2. **Cascaded BSCs.**
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 are the capacities C_1 and C_2 of $p_1(y|x)$ and $p_2(z|y)$, respectively?

(b) We now cascade these channels as shown in the figure. Thus $p_3(z|x) = \sum_y p_1(y|x)p_2(z|y)$. What is the capacity C_3 of $p_3(z|x)$?

(c) What is the capacity of the cascade in part b) if the receiver can view *both* Y and Z, i.e., the channel is defined by the transition $p((y,z)|x) = p_1(y|x)p_2(z|y)$?

(d) Now let us actively intervene between channels 1 and 2, rather than passively transmit 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 $W \longrightarrow x^n(W) \longrightarrow y^n \longrightarrow \tilde{y}^n(y^n) \longrightarrow z^n \longrightarrow \hat{W}$ where W represents the information bits.)
Note: In part (d), think of capacity as the maximum achievable rate for reliable

communication, not as the maximum mutual information, and argue about both the achievability and converse.

Solution:

(a) C_1 is just a capacity of a $BSC(\lambda_1)$. Thus, $C_1 = 1 - h_2(\lambda_1)$. Similarly, $C_2 = 1 - h_2(\lambda_2)$.

(b) First observe that the cascaded channel is also a BSC. Since the new BSC has a crossover probability of $p_3 = \lambda_1(1-\lambda_2) + (1-\lambda_1)\lambda_2 = \lambda_1 + \lambda_2 - 2\lambda_1\lambda_2$,

$$C_3 = 1 - h_2(\lambda_1 + \lambda_2 - 2\lambda_1\lambda_2).$$

As an aside, note that the new channel is noisier than the original two since by concavity of $h_2(p)$,

$$h_2((1-\lambda_1)\lambda_2 + \lambda_1(1-\lambda_2)) \geq \lambda_2 h_2(1-\lambda_1) + (1-\lambda_2)h_2(\lambda_1) = h_2(\lambda_1)$$

Similarly for $h_2(\lambda_2)$. Thus, $C_3 \leq \min\{C_1, C_2\}$.

(c) Note that Z becomes irrelevant once we observe Y. Thus, the capacity of this channel is just $C_1 = 1 - h_2(\lambda_1)$.
More formally, note that $I(X; Z|Y) = 0$ by the conditional independence of X and Z given Y. Hence, we have $I(X; Y, Z) = I(X; Y)$. Hence, the capacity is $\max_{p_X} I(X; Y, Z) = \max_{p_X} I(X; Y) = C_1$.

(d) Since we are allowed to decode the intermediate outputs and reencode them prior to the second transmission, any rate less than both C_1 and C_2 can be achievable and at the same time any rate greater than either C_1 or C_2 will cause $P_\epsilon^{(n)} \to 1$ exponentially (also by data processing, $I(X^n; Z^n) \leq \min\{I(X^n; Y^n), I(Y^n; Z^n)\}$). Hence, the overall capacity is the minimum of two capacities,

$$\min(C_1, C_2) = \min(1 - h_2(\lambda_1), 1 - h_2(\lambda_2)).$$

3. **Choice of channels.**
Let $\mathcal{C}_1 \equiv (\mathcal{X}_1, p_1(y_1|x_1), \mathcal{Y}_1)$ and $\mathcal{C}_2 \equiv (\mathcal{X}_2, p_2(y_2|x_2), \mathcal{Y}_2)$ be two channels with capacities C_1, C_2 respectively. Assume the input and output alphabets for the two channels are disjoint, i.e., $\mathcal{X}_1 \cap \mathcal{X}_2 = \emptyset$ and $\mathcal{Y}_1 \cap \mathcal{Y}_2 = \emptyset$. Consider a channel $\mathcal{C}$, which is a union of the two channels $\mathcal{C}_1$, $\mathcal{C}_2$, where at each time, one can send a symbol over $\mathcal{C}_1$ or over $\mathcal{C}_2$ but not both. In this problem, we calculate the capacity of $\mathcal{C}$ in terms of the C_1 and C_2.

Define θ as an indicator random variable denoting whether channel $\mathcal{C}_1$ or $\mathcal{C}_2$ is used in a particular transmission. Let X and Y denote the input and output for the channel $\mathcal{C}$. Note that X follows a mixture distribution over the (disjoint) alphabets $\mathcal{X}_1$ and $\mathcal{X}_2$:

$$X = \begin{cases} X_1 & \text{with probability } \alpha \\ X_2 & \text{with probability } 1-\alpha \end{cases}$$

where X_1 and X_2 are random variables taking values in $\mathcal{X}_1$ and $\mathcal{X}_2$, respectively.

(a) Argue that $H(\theta|X) = H(\theta|Y) = 0$.

(b) Show that
$$I(X;Y) = h_2(\alpha) + \alpha I(X_1;Y_1) + (1-\alpha) I(X_2;Y_2)$$
where Y_1 and Y_2 are the channel outputs when X_1 and X_2 are transmitted through $\mathcal{C}_1$ and $\mathcal{C}_2$, respectively.
Hint: Start with $I(Y;X,\theta) = I(Y;\theta) + I(Y;X|\theta) = I(Y;X) + I(Y;\theta|X)$.

(c) Let C be the capacity of the channel $\mathcal{C}$. Maximize the expression in (b) over α, P_{X_1}, P_{X_2} to show that $2^C = 2^{C_1} + 2^{C_2}$.

(d) Let $C_1 = C_2$. Then show that $C = C_1 + 1$ and give an intuitive explanation.

Solution:

(a) Since the output alphabets $\mathcal{Y}_1$ and $\mathcal{Y}_2$ are disjoint, θ is a function of Y and so $H(\theta|Y) = 0$. Similarly for $H(\theta|X) = 0$

(b) Consider,

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

Now $I(Y;\theta|X) = H(\theta|X) - H(\theta|X,Y) = 0$ because given X, θ is deterministic. Therefore,

$$\begin{aligned} I(X;Y) &= I(Y;\theta) + I(X;Y|\theta) \\ &= H(\theta) - H(\theta|Y) + \alpha I(X_1;Y_1) + (1-\alpha) I(X_2;Y_2) \\ &= h_2(\alpha) + \alpha I(X_1;Y_1) + (1-\alpha) I(X_2;Y_2) \end{aligned}$$

(c) It follows from (b) that

$$C = \sup_{\alpha} \left\{ h_2(\alpha) + \alpha C_1 + (1-\alpha) C_2 \right\}.$$

Maximizing over α one gets the desired result. The maximum occurs for $h_2'(\alpha) + C_1 - C_2 = 0$, with an optimal $\alpha^* = 2^{C_1}/(2^{C_1} + 2^{C_2})$.
Evaluating $I(X;Y)$ at α^* yields the channel capacity:

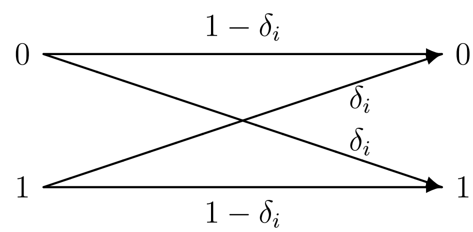
$$C = \frac{C_2 + \log_2\left(1 + 2^{C_1 - C_2}\right) + C_1 + \log_2\left(1 + 2^{-C_1 + C_2}\right)}{\log_2\left(2^{C_1} + 2^{C_2}\right)}$$

Thus,

$$2^C = 2^{C_1} + 2^{C_2}$$

(d) The result follows from (c). Intuitively, if we have two identical channels, then in each transmission, we can transmit one extra bit through our choice of the channel. As an extreme example, suppose both channels are BSC(0.5) channels. Then $C_1 = C_2 = 0$, but $C = 1$. This is because we have two possible channels and we can communicate 1 bit/transmission by sending through channel 1 if the input bit is 0 and sending through channel 2 if the input bit is 1.

4. **Time-varying channels.**
Consider a time-varying discrete *memoryless* channel. Let $Y_1, Y_2, \ldots, Y_n$ be conditionally independent given $X_1, X_2, \ldots, X_n$, with conditional distribution given by $p(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^{n} p_i(y_i|x_i)$ (where $p_i(y_i|x_i)$ is a $BSC(\delta_i)$ as shown in figure). Let $\mathbf{X} = (X_1, X_2, \ldots, X_n)$, $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)$.



In this problem, we show that

$$\max_{P_{\mathbf{X}}} I(\mathbf{X};\mathbf{Y}) = \sum_{i=1}^{n} (1 - h(\delta_i))$$

(a) Show that $I(\mathbf{X};\mathbf{Y}) \leq \sum_{i=1}^{n}(1 - h(\delta_i))$ for any $P_{\mathbf{X}}$ (It may be helpful to wait until Tuesday's lecture, and to use a chain of inequalities similar to the channel coding converse proof).

(b) Find a distribution over $\mathbf{X}$ for which $I(\mathbf{X};\mathbf{Y}) = \sum_{i=1}^{n}(1 - h(\delta_i))$.

Solution:

(a) We can use a similar chain of inequalities as in the proof of the converse to the channel coding theorem. Hence

$$\begin{aligned} I(X^n;Y^n) &= H(Y^n) - H(Y^n|X^n) \\ &= H(Y^n) - \sum_{i=1}^{n} H(Y_i|Y^{i-1}, X^n) \\ &= H(Y^n) - \sum_{i=1}^{n} H(Y_i|X_i) \end{aligned}$$

since by the definition of the channel, Y_i depends only on X_i and is conditionally independent of everything else. Continuing the series of inequalities, we have

$$\begin{aligned} I(X^n;Y^n) &= H(Y^n) - \sum_{i=1}^{n} H(Y_i|X_i) \\ &\leq \sum_{i=1}^{n} H(Y_i) - \sum_{i=1}^{n} H(Y_i|X_i) \\ &\leq \sum_{i=1}^{n}(1-h(p_i)) \end{aligned}$$

The first inequality follows by chain rule + conditioning reduces entropy (equality when Y_i's are independent). The second inequality holds because Y_i is binary (equality when Y_i's are uniform in $\{0,1\}$).

(b) We have equality when $X_1, X_2, ..., X_n$ is chosen i.i.d. $\sim Bern(1/2)$ (and so Y_i's are also i.i.d. and $Bern(1/2)$). Hence

$$\max_{p(\mathbf{x})} I(\mathbf{X};\mathbf{Y}) = \sum_{i=1}^{n}(1-h(p_i))$$

5. **Maximum Differential Entropy** (It may also be helpful to wait until Tuesday's lecture to solve this problem, even though the necessary concepts have been defined last Thursday.)

   (a) Show that among all distributions supported in an interval [a,b], the uniform distribution maximizes differential entropy.

   (b) Let X be a continuous random variable with $\mathcal{E}[X^4] \leq \sigma^4$ and Y be a continuous random variable with a probability density function $g(y) = c\exp\left(-\frac{y^4}{4\sigma^4}\right)$ where $c = \frac{1}{\int_{-\infty}^{\infty}\exp\left(-\frac{y^4}{4\sigma^4}\right)dy}$. Show that

$$h(X) \leq h(Y)$$

   with equality if and only if X is distributed as Y.
   [Hint: you can use the fact that $\mathcal{E}[Y^4] = \sigma^4$.]

Solution:

(a) Denote by $u(x)$ the uniform distribution, with $x \in [a,b]$, such that $u(x) = \frac{1}{b-a}$ if $x \in [a,b]$, and 0 otherwise. Let $g(x)$ be any distribution supported in the interval $[a,b]$. Then, we have

$$
\begin{aligned}
0 &\leq D(g||u) && (2)\\
&= \int g(x)\log\frac{g(x)}{u(x)} && (3)\\
&= \int g(x)\log\left((b-a)g(x)\right) && (4)\\
&= \log(b-a) + \int g(x)\log g(x) && (5)\\
&= \log(b-a) - H(X), && (6)
\end{aligned}
$$

which implies $H(X) \leq \log(b-a)$.

On the other hand, note that if x is uniformly distributed in the interval $[a, b]$, we have

$$
\begin{aligned}
H(X) &= \int u(x)\log\frac{1}{u(x)}\\
&= \int u(x)\log(b-a)\\
&= \log(b-a),
\end{aligned}
$$

which finishes the proof.

(b) Since $\mathcal{E}[X^4] \leq \sigma^4 = \mathcal{E}[Y^4]$, we have

$$
\begin{aligned}
D(f_X||g) =&\mathcal{E}\left[\log\frac{f_X(X)}{g(X)}\right]\\
=&-h(X) + \mathcal{E}\left[-\log g(X)\right]\\
=&-h(X) + \mathcal{E}\left[-\log c + \frac{X^4}{4\sigma^4}\log e\right]\\
\leq&-h(X) + \mathcal{E}\left[-\log c + \frac{Y^4}{4\sigma^4}\log e\right]\\
=&-h(X) + \mathcal{E}\left[-\log g(Y)\right]\\
=&-h(X) + h(Y).
\end{aligned}
$$

Therefore, $h(Y) \geq h(X) + D(f_X||g) \geq h(X)$.