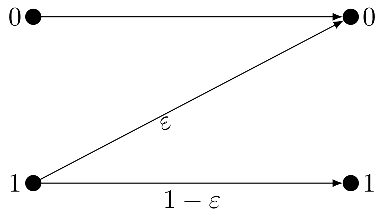
EE5139/EE6139
Solution 9

Semester I, AY2021/2022
Marco Tomamichel

Exercise 9.1 Z-channel (EE5139)

A Z channel is a binary channel with conditional pmf $p(0|0) = 1$, $p(0|1) = \epsilon$.



Suppose $\epsilon = 1/2$, compute the channel mutual information.
Solution: Let $P_X(0) = \alpha$ and $P_X(1) = 1 - \alpha$. The channel mutual information can be calculated as follows

$$\begin{aligned}
I(W) :&= \max_\alpha H(Y) - H(Y|X) \\
&= \max_\alpha H(Y) - \sum_{x=0,1} H(Y|X=x)P_X(x) \\
&= \max_\alpha H_b\Big(\frac{1}{2}(1-\alpha)\Big) - H(Y|X=1)P_X(1) \\
&= \max_\alpha H_b\Big(\frac{1}{2}(1-\alpha)\Big) - (1-\alpha)
\end{aligned}$$

Taking derivative of above expression with respect to α. By calculus, the derivative is 0 when

$$\alpha = \alpha^\star = 1 - \frac{1}{(1/2)(1+2^2)} = \frac{3}{5},$$

and is positive for $\alpha < \alpha^\star$ and negative for $\alpha > \alpha^\star$. Thus,

$$I(W) = H_b(1/5) - 2/5.$$

Exercise 9.2 Type Classes (EE5139)

Let X be a random variable on $\mathcal{X}$ with pmf p_X. The set of sequences of *type* $\lambda \in \mathcal{P}(\mathcal{X})$ is defined as

$$\mathcal{T}^{(n)}(\lambda) := \{\boldsymbol{x} \in \mathcal{X}^n : f_{\boldsymbol{x}} = \lambda\},$$

where $\mathcal{P}(\mathcal{X})$ stands for the set of all distributions over $\mathcal{X}$, and for a given sequence $\boldsymbol{x}$, $f_{\boldsymbol{x}}$ stands for the induced empirical distribution, *i.e.*, $f_{\boldsymbol{x}}(x) := n^{-1} \cdot \sum_{i=1}^n \delta_{x_i,x}$. Let X^n be n i.i.d. copies of X, *i.e.*, $p_{X^n} = p_X^{\otimes n}$. Show that the probability that X^n being any sequence $\boldsymbol{x} \in \mathcal{X}^n$ depends only on its type and p_X, namely

$$p_{X^n}(\boldsymbol{x}) = 2^{-n(H(f_{\boldsymbol{x}}) + D(f_{\boldsymbol{x}} \| p_X))}.$$

Solution: The proof is pretty straightforward.

$$\begin{aligned}
\log p_{X^n}(\boldsymbol{x}) &= \sum_{i=1}^n \log p_X(x_i) \\
&= \sum_{i=1}^n \sum_{x \in \mathcal{X}} \delta_{x_i,x} \cdot \log p_X(x) \\
&= n \cdot \sum_{x \in \mathcal{X}} f_{\boldsymbol{x}}(x) \cdot \log p_X(x) \\
&= -n \cdot \sum_{x \in \mathcal{X}} f_{\boldsymbol{x}}(x) \cdot \left(\log \frac{f_{\boldsymbol{x}}(x)}{p_X(x)} + \log \frac{1}{f_{\boldsymbol{x}}(x)} \right) \\
&= -n(H(f_{\boldsymbol{x}}) + D(f_{\boldsymbol{x}} \| p_X)).
\end{aligned}$$

Exercise 9.3 Channel Coding and List Decoding (EE6139)

In class, we saw that for all rates R below capacity C, there exists a sequence of $(2^{nR}, n)$-codes such that the average error probability tends to zero. Now, suppose we allow the decoder to output a list of 2^{nL} number of messages (instead of one), and decoding is considered successful if and only if the transmitted message is in the list. Show that for all rates $R < C$, there exists a sequence of $(\lceil 2^{n(R+L)} \rceil, n)$-codes[1] such that the average probability of error tends to zero.

Hint: Consider the joint typical set as follows

$$\mathcal{A}_\epsilon^{(n)}(X,Y) = \left\{ (\boldsymbol{x}^n, \boldsymbol{y}^n) \in \mathcal{X}^n \times \mathcal{Y}^n \;\middle|\; \begin{array}{c} \left| \frac{1}{n} \sum_{i=1}^n \log \frac{1}{p(x_i)} - H(X) \right| \le \epsilon, \\ \left| \frac{1}{n} \sum_{i=1}^n \log \frac{1}{p(y_i)} - H(Y) \right| \le \epsilon, \\ \left| \frac{1}{n} \sum_{i=1}^n \log \frac{1}{p(x_i, y_i)} - H(X,Y) \right| \le \epsilon \end{array} \right\}.$$

For any $\epsilon > 0$, the jointly typical sequences satisfy the following properties: if $\tilde{X}^n, \tilde{Y}^n$ are independent, $\tilde{X}^n \sim p^n(x)$, $\tilde{Y}^n \sim p^n(y)$, we have

- $\Pr[(\tilde{X}^n, \tilde{Y}^n) \in \mathcal{A}_\epsilon^{(n)}(X,Y)] \le 2^{-n(I(X;Y)-3\epsilon)}$,
- $\Pr[(\tilde{X}^n, \tilde{Y}^n) \in \mathcal{A}_\epsilon^{(n)}(X,Y)] \ge (1-\epsilon) 2^{-n(I(X;Y)+3\epsilon)}$.

One may consider a random encoder $e : w \mapsto X^n(w) \in \mathcal{X}^n$; and a decoder, upon receiving $\boldsymbol{y} \in \mathcal{Y}^n$, outputs a list of $\tilde{w}$'s such that $(e(\tilde{w}), \boldsymbol{y})$ is jointly typical. (Question: What is/are the error event(s)?)

Solution: Let $p_X^\star = \operatorname{argmax}_{p_X} I(X;Y)$. We generate a random code as follows: For each $w \in \{1, \ldots, 2^{nT}\}$ (T to be determined later), we independently generate a sequence in $\mathcal{X}^n$ in an i.i.d. fashion according to $p_X^\star$ as the encoded sequence of w. Namely, for each w, the encoder output can be represented by a random variable $X^n(w) \in \mathcal{X}^n$ where $X^n(w)$ is distributed according to $(p_X^\star)^{\otimes n}$ and where $\{X^n(w)\}_{w=1,\ldots,2^{nT}}$ are independent. We construct the decoder as follows $d : \cdot \mapsto \mathcal{L}(\cdot)|_{2^L}$, where

$$\mathcal{L}(Y^n) := \left\{ w \in \{1, \ldots, 2^{nT}\} : (X^n(w), Y^n) \in \mathcal{A}_\epsilon^{(n)}(X,Y) \right\},$$

and where the notation "$|_{2^L}$" denote an operation to restrict the size of the set to 2^L by padding (if the set was smaller) or chopping (if the set was larger).

Without loss of generality, assume $w = 1$ was sent, and let Y^n be the random variable describing the corresponding output of the channel. The error events are as follows:

$$\begin{aligned} \mathcal{E}_1 &:= \left\{ (X^n(1), Y^n) \notin \mathcal{A}_\epsilon^{(n)}(X,Y) \right\} \\ \mathcal{E}_2 &:= \left\{ |\mathcal{L}(Y^n)| > 2^{nL} \right\} \end{aligned}$$

By the law of large numbers, we have $\Pr(\mathcal{E}_1) \to 0$ as $n \to \infty$. As for $\mathcal{E}_2$, by Markov's inequality, we have

$$\Pr[\mathcal{E}_2] = \Pr[|\mathcal{L}(Y^n)| > 2^{nL}] \le \frac{\mathbb{E}[|\mathcal{L}(Y^n)|]}{2^{nL}}.$$

It suffices to bound $\mathbb{E}[|\mathcal{L}(Y^n)|]$ now:

$$\begin{aligned} \mathbb{E}[|\mathcal{L}(Y^n)|] &= \mathbb{E}\left[\sum_{w=1}^{2^{nT}} \mathbf{1}\{(X^n(w), Y^n) \in \mathcal{A}_\epsilon^{(n)}(X,Y)\} \right] \\ &\overset{(a)}{\le} 1 + \mathbb{E}\left[\sum_{w=2}^{2^{nT}} \mathbf{1}\{(X^n(w), Y^n) \in \mathcal{A}_\epsilon^{(n)}(X,Y)\} \right] \\ &= 1 + \sum_{w=2}^{2^{nT}} \Pr[(X^n(w), Y^n) \in \mathcal{A}_\epsilon^{(n)}(X,Y)] \\ &\overset{(b)}{\le} 1 + 2^{nT} 2^{-n(I(X;Y)-3\epsilon)}, \end{aligned}$$

[1] In this case, a (M, n)-code is comprised of an encoder $e : \mathcal{M} \to \mathcal{X}^n$ and decoder $d : \mathcal{Y}^n \to \mathcal{P}(\mathcal{M})$ where $|\mathcal{M}| = M$.

where in (a) we upper bounded the term indexed by $w = 1$ by 1, and in (b) we used the fact that $X^n(w)$ and Y^n are independent for $w > 1$ and the fact that

$$\Pr[(X^n, Y^n) \in \mathcal{A}_\epsilon^{(n)}(X, Y)] \leq 2^{-n(I(X;Y)-3\epsilon)},$$

for $X^n \perp Y^n$. As a result,

$$\Pr[\mathcal{E}_2] \leq \frac{1 + 2^{-n(I(X;Y)-T-3\epsilon)}}{2^{nL}}$$

Since we picked p_X to be capacity achieving, it holds that $I(X;Y) = C$, and thus

$$\Pr(\mathcal{E}_2) \leq 2^{-nL} + 2^{-n(C-T+L-3\epsilon)}.$$

By picking $T = C + L - 4\epsilon$, we have

$$\Pr[\mathcal{E}_2] \leq 2^{-nL} + 2^{-n\epsilon}$$

which tends to zero as $n \to \infty$. In other words, any $T < C + L$ shall results in a vanishing probability of error.

Exercise 9.4 Independently generated codebooks (EE6139)

Let $(X, Y) \sim p(x, y)$, and let $p(x)$ and $p(y)$ be their marginals. Consider two randomly and independently generated codebooks $\mathcal{C}_1 = \{X^n(1), \dots, X^n(2^{nR_1})\}$ and $\mathcal{C}_2 = \{Y^n(1), \dots, Y^n(2^{nR_2})\}$. The codewords of $\mathcal{C}_1$ are generated independently each according to $\prod_{i=1}^n p_X(x_i)$, and the codewords for $\mathcal{C}_2$ are generated independently according to $\prod_{i=1}^n p_Y(y_i)$. Define the set

$$\mathcal{C} = \{(x^n, y^n) \in \mathcal{C}_1 \times \mathcal{C}_2 : (x^n, y^n) \in \mathcal{A}_\epsilon^{(n)}(X, Y)\},$$

where $\mathcal{A}_\epsilon^{(n)}$ has been defined in the hint for Exercise 9.3. Show that

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}\left[|\mathcal{C}|\right] = R_1 + R_2 - I(X;Y).$$

Solution: For any $(X^n(i), Y^n(j)) \in \mathcal{C}_1 \times \mathcal{C}_2$, we have

$$(1-\epsilon)2^{-n(I(X:Y)+3\epsilon)} \leq \Pr\{(X^n(i), Y^n(j)) \in \mathcal{A}_\epsilon^{(n)}(X, Y)\} \leq 2^{-n(I(X:Y)-3\epsilon)}$$

for n sufficiently large. Then, for n large enough, we have

$$\begin{aligned}
\mathbb{E}[|C|] &= \mathbb{E}\Bigg[\sum_{i=1}^{2^{nR_1}} \sum_{j=1}^{2^{nR_2}} \mathbf{1}\{(X^n(i), Y^n(j)) \in C\}\Bigg] \\
&= \sum_{i=1}^{2^{nR_1}} \sum_{j=1}^{2^{nR_2}} \Pr\{(X^n(i), Y^n(j)) \in \mathcal{A}_\epsilon^{(n)}(X, Y)\} \\
&\in \left[(1-\epsilon)2^{n(R_1+R_2-I(X:Y)-3\epsilon)}, 2^{n(R_1+R_2-I(X:Y)+3\epsilon)}\right].
\end{aligned}$$

Thus, for any $\epsilon > 0$, we have $N \in \mathbb{N}$ s.t.

$$\frac{1}{n} \log \mathbb{E}\left[|\mathcal{C}|\right] \in \left[R_1 + R_2 - I(X:Y) - 3\epsilon + \frac{1}{n}\log(1-\epsilon), R_1 + R_2 - I(X:Y) + 3\epsilon\right]$$

for $n \geq N$. It follows straightforwardly that $\lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}\left[|\mathcal{C}|\right] = R_1 + R_2 - I(X;Y)$.

Exercise 9.5 Shared Randomness does not increase capacity (EE5139)

Suppose that in the definition of the $(2^{nR}, n)$ code for the DMC $p(y|x)$, we allow the encoder and the decoder to use random mappings. Specifically, let W be an arbitrary random variable independent of the message M and the channel, *i.e.*, $p(y_i|x^i, y^{i-1}, m, w) = p_{Y|X}(y_i|x_i)$ for $i \in [1:n]$. The encoder generates a codeword $x^n(m, W), m \in [1 : 2^{nR}]$, and the decoder generates an estimate $\hat{m}(y^n, W)$. Show that this randomization does not increase the capacity of the DMC.

Solution: Note that

$$nR = H(M) = I(M : Y^n, W) + H(M|Y^n, W).$$

By Fano's inequality, $H(M|Y^n, W) \leq n\epsilon_n$ for some $\epsilon_n \to 0$ as $n \to \infty$. Thus,

$$\begin{aligned}
nR &\leq I(M : Y^n, W) + n\epsilon_n \\
\implies n(R - \epsilon_n) &\leq I(M : W) + I(M : Y^n|W) \\
&= I(M : Y^n|W) \\
&= \sum_{i=1}^{n} I(M : Y_i|Y^{i-1}, W) \\
&\leq \sum_{i=1}^{n} I(M, Y^{i-1}, W : Y_i) \\
&\leq \sum_{i=1}^{n} I(M, Y^{i-1}, W, X^i : Y_i) \\
&= \sum_{i=1}^{n} I(X_i : Y_i),
\end{aligned}$$

where the last equality follows from the fact tat $p(y_i|x^i, y^{i-1}, m, w) = p_{Y|X}(y_i|x_i)$. Finally, we have

$$R - \epsilon_n \leq \frac{1}{n} \sum_{i=1}^{n} I(X_i : Y_i) \leq \max_{P_X} I(X : Y) = C,$$

and the claim is proven.