Sample problems for the 7th lecture (April 19)

1. Define the joint typical set as:

$$\mathcal{F}_n(\delta) := \Big\{ (x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n \ : |h_1(x^n) - E[h_1(X^n)]| < \delta,$$
$$|h_2(y^n) - E[h_2(Y^n)]| < \delta, \text{ and } |h_3(x^n, y^n) - E[h_3(X^n, Y^n)]| < \delta \Big\}.$$

Suppose $\mathcal{X} = \mathcal{Y} = \{0, 1, 2\}$ and $n = 3$. Assume $\{(X_i, Y_i)\}_{i=1}^n$ are i.i.d. and

$$P_{Y|X}(y|x) = \begin{cases} \frac{1}{2}, & x = y; \\ \frac{1}{4}, & x \neq y \end{cases} \text{ and } P_X(x) = \begin{cases} \frac{1}{2}, & x \neq 1 \\ 0, & x = 1 \end{cases}$$

Answer the following questions.

(a) Let
$$h_1(x^n) = -\frac{1}{n} \log_2 P_{X^n}(x^n).$$
Find $h_1(000)$ and $h_1(111)$.

(b) Further, let
$$h_2(y^n) = -\frac{1}{n} \log_2 P_{Y^n}(y^n)$$
and
$$h_3(x^n, y^n) = -\frac{1}{n} \log_2 P_{X^n, Y^n}(x^n, y^n).$$
Find the typical set if $\delta = 0.2$.

(c) Suppose we choose the codebook $\mathcal{C} = \{000, 222\}$. Find one output y^3 that is jointly typical with none of the codewords. Also, find one output y^3 that is jointly typical with more than one codewords.

(d) Continue from (c). Find the error rate of the typical set decoding.

Solution.

(a) $P_{X^3}(000) = P_{X_1}(0)P_{X_2}(0)P_{X_3}(0) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$. Hence, $h_1(000) = -\frac{1}{3} \log_2 \left(\frac{1}{8}\right) = 1$. Similarly, we compute

$$P_{X^3}(111) = P_{X_1}(1)P_{X_2}(1)P_{X_3}(1) = 0$$

and hence $h_1(111) = \infty$.

(b) First, the set of x^3 that satisfies

$$|h_1(x^n) - E[h_1(X^n)]| = |h_1(x^n) - 1| < \delta$$

include

$$\mathcal{G}_1 \triangleq \{000, 002, 020, 022, 200, 202, 220, 222\}.$$

Second, we can obtain from

$$P_Y(y) = \sum_{x=0}^{2} P_X(x) P_{Y|X}(y|x) = \begin{cases} \frac{3}{8}, & y = 0, 2 \\ \frac{1}{4}, & y = 1 \end{cases}$$

which implies $H(Y) = \frac{11 - 3\log_2(3)}{4} \approx 1.561$. We then obtain

#0+#2	#1	$P_{Y^3}(y^3)$	$\lvert h_2(y^n) - E[h_2(Y^n)]\rvert$
3	0	$\left(\frac{3}{8}\right)^3$	$\left\lvert \log_2\left(\frac{8}{3}\right) - \frac{11-3\log_2(3)}{4} \right\rvert \approx 0.146$
2	1	$\left(\frac{3}{8}\right)^2 \left(\frac{1}{4}\right)$	$\left\lvert \frac{2\log_2(8/3)+2}{3} - \frac{11-3\log_2(3)}{4} \right\rvert \approx 0.049$
1	2	$\left(\frac{3}{8}\right) \left(\frac{1}{4}\right)^2$	$\left\lvert \frac{\log_2(8/3)+4}{3} - \frac{11-3\log_2(3)}{4} \right\rvert \approx 0.244$
0	3	$\left(\frac{1}{4}\right)^3$	$\left\lvert 2 - \frac{11-3\log_2(3)}{4} \right\rvert \approx 0.439$

where #0, #1 and #2 represent the number of 0, 1 and 2 occurrences in y^3, respectively. Accordingly, the set of y^3 that satisfies

$$|h_2(y^n) - E[h_2(Y^n)]| < \delta$$

include

$$\begin{aligned} \mathcal{G}_2 \bigcup \tilde{\mathcal{G}}_2 &= \underbrace{\{100, 010, 001, 102, 012, 021, 120, 210, 201, 122, 212, 221\}}_{\mathcal{G}_2} \\ &\quad \bigcup \underbrace{\{000, 002, 020, 022, 200, 202, 220, 222\}}_{\tilde{\mathcal{G}}_2}. \end{aligned}$$

Finally,

$$H(X, Y) = H(X) + H(Y|X) = 1 + H(\tfrac{1}{2}, \tfrac{1}{4}, \tfrac{1}{4}) = \frac{5}{2}.$$

Thus, (x^3, y^3) is in the typical set only when

$$\begin{aligned} P_{X^3,Y^3}(x^3, y^3) &= P_{X^3}(x^3) \cdot P_{Y^3|X^3}(y^3|x^3) \\ &= \tfrac{1}{8} \cdot (\tfrac{1}{2} \cdot \tfrac{1}{4} \cdot \tfrac{1}{4}) \text{ or } \tfrac{1}{8} \cdot (\tfrac{1}{2} \cdot \tfrac{1}{2} \cdot \tfrac{1}{4}) \end{aligned}$$

Checking the pairs from $\mathcal{G}_1$ and $\mathcal{G}_2 \bigcup \tilde{\mathcal{G}}_2$ yields that

$$\begin{aligned}
\mathcal{F}_3(\delta) \;=\; & \{(000,100),(000,010),(000,001),(000,102),\\
& (000,012),(000,021),(000,120),(000,210),(000,201),\\
& (000,002),(000,020),(000,022),(000,200),(000,202),(000,220),\\
& (002,100),(002,010),(002,001),(002,102),\\
& (002,012),(002,021),(002,201),(002,122),(002,212),\\
& (002,000),(002,020),(002,022),(002,200),(002,202),(002,222),\\
& (020,100),(020,010),(020,001),(020,012),\\
& (020,021),(020,120),(020,210),(020,122),(020,221),\\
& (020,000),(020,002),(020,022),(020,200),(020,220),(020,222)\\
& (022,010),(022,001),(022,102),(022,012),\\
& (022,021),(022,120),(022,122),(022,212),(022,221),\\
& (022,000),(022,002),(022,020),(022,202),(022,220),(022,222),\\
& (200,100),(200,010),(200,001),(200,102),\\
& (200,120),(200,210),(200,201),(200,212),(200,221),\\
& (200,000),(200,002),(200,020),(200,202),(200,220),(200,222),\\
& (202,100),(202,001),(202,102),(202,012),\\
& (202,210),(202,201),(202,122),(202,212),(202,221),\\
& (202,000),(202,002),(202,022),(202,200),(202,220),(202,222),\\
& (220,100),(220,010),(220,021),(220,120),\\
& (220,210),(220,201),(220,122),(220,212),(220,221),\\
& (220,000),(220,020),(220,022),(220,200),(220,202),(220,222),\\
& (222,102),(222,012),(222,021),(222,120),\\
& (222,210),(222,201),(222,122),(222,212),(222,221),\\
& (222,002),(222,020),(222,022),(222,200),(222,202),(222,220)\}.
\end{aligned}$$

(c) Any y^3 outside $\mathcal{G}_2 \bigcup (\tilde{\mathcal{G}}_2 \setminus \{000, 222\})$ is not jointly typical with any codeword. In fact, all y^3 in $\mathcal{G}_2 \bigcup (\tilde{\mathcal{G}}_2 \setminus \{000, 222\})$ are jointly typical with at least one codeword, among which $y^3 = 102$, 012, 021, 120, 210, 201, 002, 020, 022, 200, 202 and 220, are jointly typical with both codewords.

(d) The input-output pairs that will be decoded correctly include

$$\{(000,100),(000,010),(000,001),(222,122),(222,212),(222,221)\},$$

which occurs with probability $\frac{1}{2} \cdot \frac{1}{16} \cdot 6 = \frac{3}{16}$. Note that adding the term of $\frac{1}{2}$ is because the prior probability for sending 000 and 222 is respectively $\frac{1}{2}$.

The input-output pairs that will be decoded incorrectly include

$$\{(000,122),(000,212),(000,221),(222,100),(222,010),(222,001)\},$$

which occurs with probability $\frac{1}{2}\cdot\frac{1}{4^3}\cdot 6=\frac{3}{64}$.

For all remaining input-output pairs, the decoder simply makes a random guess. Hence, the error rate is

$$\frac{1}{2}\left(1-\frac{3}{16}-\frac{3}{64}\right)+\frac{3}{64}=\frac{55}{128}\approx 0.43.$$

Note: The typical set decoder performs well only when n is very, very large. When n is as small as 3, such a decoding approach indeed performs poor.

2. Fixe a set $\mathcal{F}_n(\delta)$ that satisfies

$$|\mathcal{F}_n(\delta)|\le 2^{n(H(X,Y)+\delta)} \text{ and } P_{X^n,Y^n}(\mathcal{F}_n^c(\delta))<\delta$$

and for $(x^n,y^n)\in\mathcal{F}_n(\delta)$,

$$P_{X^n}(x^n)\le 2^{-n(H(X)-\delta)} \text{ and } P_{Y^n}(y^n)\le 2^{-n(H(Y)-\delta)}.$$

(For your information, Shannon's channel coding theorem only requires these two conditions.) For a given codebook

$$\mathcal{C}_n{:=}\{\boldsymbol{c}_1,\boldsymbol{c}_2,\ldots,\boldsymbol{c}_{M_n}\},$$

we define the encoder $f_n(\cdot)$ and decoder $g_n(\cdot)$, respectively, as follows:

$$f_n(m)=\boldsymbol{c}_m \quad \text{for } 1\le m\le M_n,$$

and

$$g_n(y^n)=\begin{cases} m, & \text{if } \boldsymbol{c}_m \text{ is the only codeword in } \mathcal{C}_n \\ & \text{satisfying } (\boldsymbol{c}_m,y^n)\in\mathcal{F}_n(\delta); \\ \text{any one in } \{1,2,\ldots,M_n\}, & \text{otherwise.}\end{cases}$$

(a) Express the conditional probability of error λ_m, given that $\boldsymbol{c}_m$ is transmitted, in terms of $\mathcal{D}_m$, where

$$\mathcal{D}_m:=\mathcal{F}_n(\delta|\boldsymbol{c}_m)\setminus\bigcup_{m'=1,m'\ne m}^{M_n}\mathcal{F}_n(\delta|\boldsymbol{c}_{m'})$$

and

$$\mathcal{F}_n(\delta|x^n) := \{y^n \in \mathcal{Y}^n \ : \ (x^n, y^n) \in \mathcal{F}_n(\delta)\}.$$

Hint: $\mathcal{D}_m$ is the set of y^n that will be definitely decoded to $\boldsymbol{c}_m$ without randomization.

(b) Show that

$$\lambda_m \leq P_{Y^n|X^n}\left(\mathcal{D}_m^c \middle| \boldsymbol{c}_m\right) = 1 - P_{Y^n|X^n}\left(\mathcal{D}_m \middle| \boldsymbol{c}_m\right).$$

Hint: By definition, $\{\mathcal{D}_m\}_{m=1}^{M_n}$ are disjoint.

(c) Show that

$$P_{Y^n|X^n}\left(\mathcal{D}_m^c \middle| \boldsymbol{c}_m\right) \leq P_{Y^n|X^n}\left(\mathcal{F}_n^c(\delta|\boldsymbol{c}_m) \middle| \boldsymbol{c}_m\right) + \sum_{m'=1, m'\neq m}^{M_n} P_{Y^n|X^n}\left(\mathcal{F}_n(\delta|\boldsymbol{c}_{m'}) \middle| \boldsymbol{c}_m\right).$$

Hint: $\mathcal{D}_m \subset \mathcal{F}_n(\delta|\boldsymbol{c}_m)$.

(d) The upper bound in (c) shows that the error bound for typical set decoding is characterized by

$$P_{Y^n|X^n}\left(\mathcal{F}_n^c(\delta|\boldsymbol{c}_m) \middle| \boldsymbol{c}_m\right) \text{ and } P_{Y^n|X^n}\left(\mathcal{F}_n(\delta|\boldsymbol{c}_{m'}) \middle| \boldsymbol{c}_m\right).$$

The expected value of $P_{Y^n|X^n}\left(\mathcal{F}_n^c(\delta|\boldsymbol{c}_m) \middle| \boldsymbol{c}_m\right)$ with respect to random $\boldsymbol{c}_m$ with distribution P_{X^n} is

$$\begin{aligned} & E\left[P_{Y^n|X^n}\left(\mathcal{F}_n^c(\delta|\boldsymbol{c}_m) \middle| \boldsymbol{c}_m\right)\right] \\ &= \sum_{\boldsymbol{c}_m \in \mathcal{X}^n} P_{X^n}(\boldsymbol{c}_m) P_{Y^n|X^n}\left(\mathcal{F}_n^c(\delta|\boldsymbol{c}_m) \middle| \boldsymbol{c}_m\right) \\ &= P_{X^n,Y^n}\left(\mathcal{F}_n^c(\delta)\right) < \delta. \end{aligned}$$

Thus if the second term in (c) is expectedly small, then the expected decoding error will be small.

Show that $2^{-n(I(X;Y)-3\delta)}$ is the upper bound of the expected value of $P_{Y^n|X^n}\left(\mathcal{F}_n(\delta|\boldsymbol{c}_{m'}) \middle| \boldsymbol{c}_m\right)$ with respect to independent $\boldsymbol{c}_m$ and $\boldsymbol{c}_{m'}$ with common distribution P_{X^n}.

(e) Show that the expected value of the second term in (c) with respect to i.i.d. $\{\boldsymbol{c}_m\}_{m=1}^{M_n}$ can be made smaller than $2^{-n\delta}$, i.e.,

$$E\left[\sum_{m'=1,m'\neq m}^{M_n} P_{Y^n|X^n}\left(\mathcal{F}_n(\delta|\boldsymbol{c}_{m'})\middle|\boldsymbol{c}_m\right)\right] \leq 2^{-n\delta}$$

if

$$\frac{1}{n}\log_2(M_n) \leq I(X;Y) - 4\delta.$$

Hint: Use (d).

Solution.

(a) (See the example in Problem 1(d).) Given that $\boldsymbol{c}_m$ is transmitted, the decoder decodes correctly when

$$y^n \in \mathcal{D}_m.$$

The decoder decodes incorrectly when

$$y^n \in \bigcup_{m'=1,m'\neq m}^{M_n} \mathcal{D}_{m'}.$$

For the remaining y^n, a random guess is performed. Hence,

$$\begin{aligned}
\lambda_m &= \frac{M_n-1}{M_n}\left(1 - P_{Y^n|X^n}(\mathcal{D}_m|\boldsymbol{c}_m) - P_{Y^n|X^n}\left(\bigcup_{m'=1,m'\neq m}^{M_n}\mathcal{D}_{m'}\middle|\boldsymbol{c}_m\right)\right)\\
&\quad + P_{Y^n|X^n}\left(\bigcup_{m'=1,m'\neq m}^{M_n}\mathcal{D}_{m'}\middle|\boldsymbol{c}_m\right)\\
&= \frac{M_n-1}{M_n}\left(1 - P_{Y^n|X^n}\left(\bigcup_{m'=1}^{M_n}\mathcal{D}_{m'}\middle|\boldsymbol{c}_m\right)\right) + P_{Y^n|X^n}\left(\bigcup_{m'=1,m'\neq m}^{M_n}\mathcal{D}_{m'}\middle|\boldsymbol{c}_m\right).
\end{aligned}$$

(b)

$$\begin{aligned}
\lambda_m &= \frac{M_n-1}{M_n}\left(1 - P_{Y^n|X^n}\left(\bigcup_{m'=1}^{M_n}\mathcal{D}_{m'}\middle|\boldsymbol{c}_m\right)\right) + P_{Y^n|X^n}\left(\bigcup_{m'=1,m'\neq m}^{M_n}\mathcal{D}_{m'}\middle|\boldsymbol{c}_m\right)\\
&\leq 1 - \left(P_{Y^n|X^n}\left(\bigcup_{m'=1}^{M_n}\mathcal{D}_{m'}\middle|\boldsymbol{c}_m\right) - P_{Y^n|X^n}\left(\bigcup_{m'=1,m'\neq m}^{M_n}\mathcal{D}_{m'}\middle|\boldsymbol{c}_m\right)\right)\\
&= 1 - P_{Y^n|X^n}\left(\mathcal{D}_m\middle|\,\boldsymbol{c}_m\right).
\end{aligned}$$

Note: This shows that the proof of Shannon's channel coding theorem simply regards the random guess as a definite error.

(c) With $\mathcal{D}_m \subset \mathcal{F}_n(\delta|\boldsymbol{c}_m)$, we derive

$$\begin{aligned}
P_{Y^n|X^n}\left(\mathcal{D}_m^c \middle| \boldsymbol{c}_m\right) &= P_{Y^n|X^n}\left(\mathcal{F}_n^c(\delta|\boldsymbol{c}_m)\middle|\boldsymbol{c}_m\right) + P_{Y^n|X^n}\left(\mathcal{D}_m^c \setminus \mathcal{F}_n^c(\delta|\boldsymbol{c}_m)\middle|\boldsymbol{c}_m\right) \\
&= P_{Y^n|X^n}\left(\mathcal{F}_n^c(\delta|\boldsymbol{c}_m)\middle|\boldsymbol{c}_m\right) + P_{Y^n|X^n}\left(\mathcal{F}_n(\delta|\boldsymbol{c}_m) \setminus \mathcal{D}_m\middle|\boldsymbol{c}_m\right) \\
&\leq P_{Y^n|X^n}\left(\mathcal{F}_n^c(\delta|\boldsymbol{c}_m)\middle|\boldsymbol{c}_m\right) + P_{Y^n|X^n}\left(\bigcup_{m'=1, m'\neq m}^{M_n} \mathcal{F}_n(\delta|\boldsymbol{c}_{m'})\middle|\boldsymbol{c}_m\right) \\
&\leq P_{Y^n|X^n}\left(\mathcal{F}_n^c(\delta|\boldsymbol{c}_m)\middle|\boldsymbol{c}_m\right) + \sum_{m'=1, m'\neq m}^{M_n} P_{Y^n|X^n}\left(\mathcal{F}_n(\delta|\boldsymbol{c}_{m'})\middle|\boldsymbol{c}_m\right)
\end{aligned}$$

(d)

$$\begin{aligned}
&E\left[P_{Y^n|X^n}\left(\mathcal{F}_n(\delta|\boldsymbol{c}_{m'})\middle|\boldsymbol{c}_m\right)\right] \\
&= \sum_{\boldsymbol{c}_m\in\mathcal{X}^n}\sum_{\boldsymbol{c}_{m'}\in\mathcal{X}^n} P_{X^n}(\boldsymbol{c}_m)P_{X^n}(\boldsymbol{c}_{m'})P_{Y^n|X^n}\left(\mathcal{F}_n(\delta|\boldsymbol{c}_{m'})\middle|\boldsymbol{c}_m\right) \\
&= \sum_{\boldsymbol{c}_m\in\mathcal{X}^n}\sum_{\boldsymbol{c}_{m'}\in\mathcal{X}^n} P_{X^n}(\boldsymbol{c}_m)P_{X^n}(\boldsymbol{c}_{m'}) \sum_{y^n\in\mathcal{F}_n(\delta|\boldsymbol{c}_{m'})} P_{Y^n|X^n}\left(y^n\middle|\boldsymbol{c}_m\right) \\
&= \sum_{\boldsymbol{c}_{m'}\in\mathcal{X}^n}\sum_{y^n\in\mathcal{F}_n(\delta|\boldsymbol{c}_{m'})} P_{X^n}(\boldsymbol{c}_{m'}) \sum_{\boldsymbol{c}_m\in\mathcal{X}^n} P_{X^n}(\boldsymbol{c}_m)P_{Y^n|X^n}\left(y^n\middle|\boldsymbol{c}_m\right) \\
&= \sum_{\boldsymbol{c}_{m'}\in\mathcal{X}^n}\sum_{y^n\in\mathcal{F}_n(\delta|\boldsymbol{c}_{m'})} P_{X^n}(\boldsymbol{c}_{m'})P_{Y^n}(y^n) \\
&= \sum_{(\boldsymbol{c}_{m'}, y^n)\in\mathcal{F}_n(\delta)} P_{X^n}(\boldsymbol{c}_{m'})P_{Y^n}(y^n) \\
&\leq \sum_{(\boldsymbol{c}_{m'}, y^n)\in\mathcal{F}_n(\delta)} 2^{-n(H(X)-\delta)}2^{-n(H(Y)-\delta)} \\
&= |\mathcal{F}_n(\delta)|2^{-n(H(X)-\delta)}2^{-n(H(Y)-\delta)} \\
&\leq 2^{n(H(X,Y)+\delta)}2^{-n(H(X)-\delta)}2^{-n(H(Y)-\delta)} \\
&= 2^{-n(H(X)+H(Y)-H(X,Y)-3\delta)} \\
&= 2^{-n(I(X;Y)-3\delta)}.
\end{aligned}$$

(e)

$$
\begin{aligned}
& E\Bigg[\sum_{m'=1, m'\neq m}^{M_n} P_{Y^n|X^n}\Big(\mathcal{F}_n(\delta|\boldsymbol{c}_{m'})\Big|\boldsymbol{c}_m\Big)\Bigg] \\
&= \sum_{m'=1, m'\neq m}^{M_n} E\Big[P_{Y^n|X^n}\Big(\mathcal{F}_n(\delta|\boldsymbol{c}_{m'})\Big|\boldsymbol{c}_m\Big)\Big] \\
&\leq \sum_{m'=1, m'\neq m}^{M_n} 2^{-n(I(X;Y)-3\delta)} \\
&= (M_n-1)2^{-n(I(X;Y)-3\delta)} \\
&\leq M_n 2^{-n(I(X;Y)-3\delta)}.
\end{aligned}
$$

Thus, if $M_n \leq 2^{n(I(X;Y)-4\delta)}$, then

$$M_n 2^{-n(I(X;Y)-3\delta)} \leq 2^{-n\delta}.$$

Note: This implies that the maximum code size to have a vanishing error is $I(X;Y) - 4\delta$ for arbitrarily small δ. As a result, maximizing $I(X;Y)$ over all input P_X becomes a natural next step.

3. The communication system can be simplified as

$$W \to X^n \to Y^n \to \hat{W},$$

where W is a uniformly distributed message over M_n possibilities, and hence $H(W) = \log_2(M_n)$. Let P_e denote the error of estimating W based on Y^n.

(a) Show that

$$\log_2 M_n \quad \leq \quad 1 + P_e \cdot \log_2 M_n + I(X^n; Y^n).$$

Hint: Fano's inequality and data processing inequality, i.e.,

$$H(W|Y^n) \leq h_b(P_e) + P_e \log_2(M_n - 1) \text{ and } I(W;Y^n) \leq I(X^n;Y^n).$$

(b) For DMS $P_{X^n}(x^n) = \prod_{i=1}^n P_X(x_i)$ and DMC $P_{Y^n|X^n} = \prod_{i=1}^n P_{Y|X}(y_i|x_i)$, we have

$$\frac{1}{n} I(X^n;Y^n) = I(X;Y).$$

Show that if

$$\frac{1}{n}\log_2(M_n) > 2I(X;Y),$$

, then

$$P_e > \frac{1}{2} - \frac{1}{n \cdot I(X;Y)}.$$

Hint: Use (a).

Solution.

(a)

$$\begin{aligned}
\log_2 M_n &= H(W) \\
&= H(W|Y^n) + I(W;Y^n) \\
&\leq H(W|Y^n) + I(X^n;Y^n) \\
&\leq 1 + P_e(\mathcal{C}_n) \cdot \log_2 M_n + I(X^n;Y^n).
\end{aligned}$$

(b) The result in (a) implies

$$\begin{aligned}
P_e &\geq 1 - \frac{1 + I(X^n;Y^n)}{\log_2 M_n} \\
&= 1 - \frac{\frac{1}{n} + \frac{1}{n}I(X^n;Y^n)}{\frac{1}{n}\log_2 M_n} \\
&= 1 - \frac{\frac{1}{n} + I(X;Y)}{\frac{1}{n}\log_2 M_n} \\
&> 1 - \frac{\frac{1}{n} + I(X;Y)}{2I(X;Y)} \\
&= \frac{1}{2} - \frac{1}{n \cdot I(X;Y)}.
\end{aligned}$$

Note: Hence, P_e is bounded away from zero as n large.