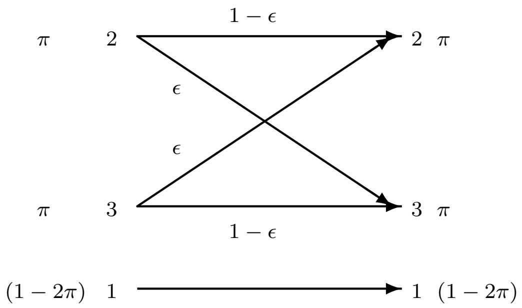
Solutions: Homework Set # 8

Problem 1



(a) Kuhn-Tucker condition:

$$\sum_y P(y \mid x) \log \frac{P(y \mid x)}{P(x)} = C \quad \forall \; P(x) > 0$$

For $x = 1$, this will be $\log \frac{1}{1-2\pi}$ and for $x = 2$ and $x = 33$ it would be $(1-\epsilon)\log\frac{1-\epsilon}{\pi} + \epsilon \log \frac{\epsilon}{\pi}$ Thus,

$$\begin{aligned}
&\log \frac{1}{1-2\pi} = (1-\epsilon)\log\frac{1-\epsilon}{\pi} + \epsilon\log\frac{\epsilon}{\pi} \\
&\Rightarrow \frac{1}{1-2\pi} = \left(\frac{1-\epsilon}{\pi}\right)^{1-\epsilon}\left(\frac{\epsilon}{\pi}\right)^{\epsilon} \\
&\Rightarrow \frac{\pi}{1-2\pi} = (1-\epsilon)^{1-\epsilon}\epsilon^{\epsilon} = \delta \\
&\Rightarrow \pi = \delta - 2\pi\delta && (1) \\
&\Rightarrow (1+2\delta)\pi = \delta && (2) \\
&\Rightarrow \pi = \frac{\delta}{1+2\delta} \\
&C = \log\frac{1}{1-2\pi} = \log\frac{1}{1-\frac{2\delta}{1+2\delta}} = \log(1+2\delta) = \log\left(1+2(1-\epsilon)^{1-\epsilon}\epsilon^{\epsilon}\right)
\end{aligned}$$

(b) If $\epsilon = 0 \to \lim_{\epsilon\to 0}\epsilon^{\epsilon} = 1$. Then,

$$\begin{aligned}
C &= \log 1 + 2 = \log 3 \quad \text{bits} \\
&= \log \mid \mathcal{Y} \mid
\end{aligned}$$

If $\epsilon = \frac{1}{2}$, then

$$C = \log\left(1 + 2\left(\frac{1}{2}\right)^{\frac{1}{2}}\left(\frac{1}{2}\right)^{\frac{1}{2}}\right) = \log 2 = 1 \quad \text{bits}$$

If $\epsilon = 0$, we have a perfect noiseless channel with $|\mathcal{X}| = |\mathcal{Y}| = 3$.

$$I(X;Y) = H(Y) - H(Y \mid X) = H(Y) \quad \text{since} \quad H(Y \mid X) = 0$$
$$\rightarrow \max H(Y) = C = \log 3$$

If $\epsilon = \frac{1}{2}$, $X = 2$ and $X = 3$ will be received completely random. So we can say that only one bit is transmitted, Hence, $C = 1$ bit.

Problem 2

(a) We know for each of two channels there is an input distribution $p_1^*(x)$ and $p_2^*(x)$ such that they achieve capacitites C_1 and C_2 respectively. Assume sender choses the alphabet $\mathcal{X}_1$ with probability p and $\mathcal{X}_2$ with probability $(1-p)$. Therefore,

$$Pr\{x : x \in \mathcal{X}_1 \cup \mathcal{X}_2\} = \begin{cases} p\ \ p_1^*(x) & x \in \mathcal{X}_1 \\ (1-p)\ \ p_2^*(x) & x \in \mathcal{X}_2 \end{cases}$$

$$\begin{aligned} Pr\{Y = y\} &= \sum_{x \in \mathcal{X}_1 \cup \mathcal{X}_2} Pr\{X = x\} Pr\{Y = y \mid X = x\} \\ &= p \sum_{x \in \mathcal{X}_1} p_1^*(x) Pr\{Y = y \mid X = x\} \\ &+ (1-p) \sum_{x \in \mathcal{X}_2} p_2^*(x) Pr\{Y = y \mid X = x\} \end{aligned}$$

$$Pr\{Y = y \mid X = x\} = \begin{cases} p_1(y \mid x) & x \in \mathcal{X}_1 \text{ and } y \in \mathcal{Y}_1 \\ p_2(y \mid x) & x \in \mathcal{X}_2 \text{ and } y \in \mathcal{Y}_2 \\ 0 & \text{otherwise} \end{cases}$$

$$\Rightarrow Pr\{Y = y\} = \begin{cases} p\ \ P_1^*(y) & y \in \mathcal{Y}_1 \\ (1-p)\ \ P_2^*(y) & y \in \mathcal{Y}_2 \end{cases}$$

while $\begin{array}{l} P_1^*(y) = \sum p_1^*(x) p_1(y \mid x) \\ P_2^*(y) = \sum p_2^*(x) p_2(y \mid x) \end{array}$

We write Kuhn-Tucker conditions. If $x_i \in \mathcal{X}_1$:

$$\begin{aligned} \sum p(y \mid x_i) \log \frac{p(y \mid x_i)}{p(y)} &= \sum_{y \in \mathcal{Y}_1} p_1(y \mid x_i) \frac{p_1(y \mid x_i)}{p P_1^*(y)} \\ &= \sum_{y \in \mathcal{Y}_1} p_1(y \mid x_i) \frac{p_1(y \mid x_i)}{P_1^*(y)} - \log p \sum_{y \in \mathcal{Y}_1} p_1(y \mid x_i) \\ &= C_1 - \log p \end{aligned}$$

If $x_i \in \mathcal{X}_2$, similarly we get

$$\sum p(y \mid x_i) \log \frac{p(y \mid x_i)}{p(y)} = C_2 - \log (1-p)$$

Now with proper choice of p, we can satisfy K-T conditions for the new union channel, thus:

$$C_1 - \log p = C_2 - \log (1-p) \quad \Rightarrow \quad \frac{p}{1-p} = 2^{C_1 - C_2} \Rightarrow p = \frac{1}{1 + 2^{C_1 - C_2}}$$

Thus,

$$\begin{aligned} &C = C_1 + \log (1 + 2^{C_1 - C_2}) \\ &2^C = 2^{C_1}(1 + 2^{C_1 - C_2}) \\ \Rightarrow &2^C = 2^{C_1} + 2^{C_2} \end{aligned}$$

(b)

$$\begin{aligned} I(X_1, X_2; Y_1, Y_2) &= H(Y_1, Y_2) - H(Y_1, Y_2 \mid X_1, X_2) \\ P(x_1, x_2, y_1, y_2) &= P(x_1, x_2)P(y_1 \mid x_1, x_2)P(y_2 \mid y_1, x_1, x_2) \\ &= P(x_1, x_2)P(y_1 \mid x_1)P(y_2 \mid x_2) \\ \rightarrow H(Y_1, Y_2 \mid X_1, X_2) &= H(Y_1, \mid X_1) + H(Y_2 \mid X_2) \end{aligned}$$

Since $Y_1 \rightarrow X_1 \rightarrow X_2 \rightarrow Y_2$ forms a Markov chain.

$$\begin{aligned} I(X_1, X_2; Y_1, Y_2) &= H(Y_1, Y_2) - H(Y_1, \mid X_1) - H(Y_2 \mid X_2) \\ &\geq H(Y_1) - H(Y_1, \mid X_1) + H(Y_2) - H(Y_2 \mid X_2) \end{aligned}$$

with equality iff Y_1 and Y_2 are independent, i.e. X_1 and X_2 are independent.

$$\rightarrow \max I(X_1, X_2; Y_1, Y_2) = C$$

(c) Using part (a), we have $C = \log (2^{C_1} + \log 2^{C_2})$, $C_1 = 1 - H_2(P)$ and $C_2 = 0$

$$\rightarrow C = \log (2^{1 - H_2(P)} + 2^0) = \log (1 + 2^{1 - H_2(p)})$$

Problem 3

(a) We note that channel S_1 is symmetric. Hence, $\Pi = 1 - \Pi = \frac{1}{2}$.

$$\begin{aligned} &P(y = +1) = \frac{1}{2}(1 - \epsilon) \\ &P(y = -1) = \frac{1}{2}(1 - \epsilon) \\ &P(y = \epsilon) = \frac{1}{2}\epsilon + \frac{1}{2}\epsilon = \epsilon \\ &C_1 = \sum_y P(y \mid x_i) \log \frac{P(y \mid y_i)}{P(y)} = (1 - \epsilon) \log \frac{1 - \epsilon}{0.5(1 - \epsilon)} + \epsilon \log \frac{\epsilon}{\epsilon} = 1 - \epsilon \end{aligned}$$

In class we saw that the capacity of the Z-channel with $P_{error} = \gamma$ is

$$C_2 = \log\left(1 + \delta(1-\gamma)\right)$$

where $\delta = \gamma^{\frac{\gamma}{1-\gamma}}$.

For $\gamma = 0.5$, $C = \log\left(1 + 0.50.5\right) = \log\frac{5}{4}$

(b)

$$\begin{bmatrix} \pi_1 & \pi_2 \end{bmatrix} \begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix} = \begin{bmatrix} \pi_1 & \pi_2 \end{bmatrix}$$

$$\pi_1(1-a) + (1-\pi_1)b = \pi_1 \rightarrow \pi_1 - \pi_1 a + b - \pi_1 b = \pi_1$$

$$\rightarrow \pi_1 = \frac{b}{a+b} \qquad \pi_2 = \frac{a}{a+b}$$

(c) In fact, we only derive a transmission rate and show that this rate is achievable. Showing the optimality of this rate is more technical.

We claim that $R = \pi_1 C_1 + \pi_2 C_2$ is achievable. We propose the following simple scheme to transmit at rate R. The transmitter designs two capacity achieving codes C_1 and C_2 for transmission over the two erasure and Z channels, respectively. Depending on the state of the channel, it sends the first remaining bit (the next bit in the sequence) of the corresponding codeword. Since the receiver also knows the state of the channel, it can collect all the bits and then re-arrange them to obtain channel output as if the same codeword was used for transmission on a single binary symmetric channel. For large duration of time, the channel would be in the bad state for π_1 fraction of time, and in the good state for the remaining π_2 fraction. Since we transmit at capacity rate for each fraction, the total rate would be $R = \pi_1 C_1 + \pi_2 C_2$.

(d)

$$
\begin{aligned}
C_{NSI} &= \frac{1}{n} I(X^n; Y^n) \\
&\leq \frac{1}{n} I(X^n; Y^n, S^n) \\
&= \frac{1}{n} I(X^n; S^n) + \frac{1}{n} I(X^n; Y^n|S^n) \\
&= \frac{1}{n} \left(H(Y^n|S^n) - H(Y^n|X^n, S^n)\right) \\
&\quad (I(X^n; S^n) = 0 \text{ since the encoder does not have access to the knowledge of channel state}) \\
&= \frac{1}{n} \sum_i \left(H(Y_i|S^n Y_1^{i-1}) - H(Y_i|X^n, S^n, Y_1^{i-1})\right) \\
&\leq \frac{1}{n} \sum_i \left(H(Y_i|S_i) - H(Y_i|X^n, S^n, Y_1^{i-1})\right) \quad \text{Conditioning reduces entropy} \\
&\leq \frac{1}{n} \sum_i \left(H(Y_i|S_i) - H(Y_i|X_i, S_i)\right) \quad X^n, S^n, Y_1^{i-1} \longrightarrow S_i, X_i \longrightarrow Y_i \text{ forms a Markov chain} \\
&= \frac{1}{n} \sum_i I(X_i; Y_i|S_i) \\
&= \frac{1}{n} \sum_i I(X_i; Y_i, S_i) \\
&\quad (I(X_i; S_i) = 0 \text{ since the encoder does not have access to the knowledge of channel state})
\end{aligned}
$$

Note that X_i is NOT only dependent on Y_i and S_i and thus $H(X_i|Y^n, S^n, X_1^{i-1}) \neq H(X_i|Y_i, S_i)$! So the hint was not the right path to prove $C_{NSI} \leq \frac{1}{n} \sum_i I(X_i; Y_i|S_i)$.

Problem 4 (**Channel with Memory**)

In this problem we consider the memory-less channel $Y_i = Z_i X_i$ with input alphabet $X_i \in \{-1, 1\}$.

(a) In this part we assume that $\{Z_i\}$ are i.i.d. with

$$
Z_i = \begin{cases} 1, & p = 0.5, \\ -1, & p = 0.5. \end{cases}
$$

Let us send a sequence of capacity achieving codewords of length n over this channel.

Then for the capacity of such a channel we can write

$$
\begin{aligned}
C &= \frac{1}{n} I(X^n; Y^n) \\
&= \frac{1}{n} \left[H(Y^n) - H(Y^n|X^n) \right] \\
&= \frac{1}{n} \left[H(Y^n) - H(Z^n) \right] \\
&\overset{(1)}{=} \frac{1}{n} \left[H(Y^n) - \sum_{i=1}^{n} H(Z_i) \right] \\
&\overset{(2)}{=} \frac{1}{n} \left[H(Y^n) - n \right] \\
&\leq 0,
\end{aligned}
$$

where (1) follows because Z_i are i.i.d. and (2) follows because $H(Y^n)$ is at most n. So for the capacity in this case we have $C = 0$.

(b) In this part of problem we assume that Z is randomly chosen at the beginning of transmission and remains fixed during the transmission so $Z_i = Z$ where we have

$$
Z = \begin{cases} 1, & p = 0.5, \\ -1, & p = 0.5. \end{cases}
$$

Again for the capacity we can write

$$
\begin{aligned}
C &= \frac{1}{n} I(X^n; Y^n) \\
&= \frac{1}{n} \left[H(Y^n) - H(Y^n|X^n) \right] \\
&= \frac{1}{n} \left[H(Y^n) - H(Z^n) \right] \\
&\overset{(1)}{=} \frac{1}{n} \left[H(Y^n) - \sum_{i=1}^{n} H(Z_i|Z^{i-1}) \right] \\
&\overset{(2)}{=} \frac{1}{n} \left[H(Y^n) - 1 \right] \\
&\overset{(3)}{=} \frac{1}{n} \left[n - 1 \right] \xrightarrow{n \to \infty} 1 \quad \text{bits},
\end{aligned}
$$

where (1) follows from the chain rule, (2) follows because $H(Z_i|Z^{i-1}) = 0$ for $i \geq 2$ and $H(Z_i|Z^{i-1}) = H(Z) = 1$ for $i = 1$, and (3) follows because the uniform input distribution makes the output distribution uniform which maximizes $H(Y^n)$.

Problem 5

In this problem we consider a BSC with crossover probability $0 < \epsilon < 1$ represented by $X_i = Y_i + Z_i \mod 2$, where X_i, Y_i, and Z_i are, respectively, the input, the output, and the noise variable at time i. Then

$$
\mathbb{P}[Z_i = 0] = 1 - \epsilon \quad \text{and} \quad \mathbb{P}[Z_i = 1] = \epsilon
$$

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) We can write

$$
\begin{aligned}
I(X^n; Y^n) &= H(Y^n) - H(Y^n|X^n) \\
&= H(Y^n) - H(Z^n) \\
&\overset{(1)}{\leq} n - H(Z^n) \\
&= n\left[1 - \frac{1}{n}H(Z^n)\right] \\
&\overset{(2)}{\leq} n\left[1 - \mathcal{H}(Z)\right]
\end{aligned}
$$

where $\mathcal{H}(Z)$ is the entropy rate of the random process $\{Z_i\}$. Note that
(1) follows from the fact that $H(Y^n) \leq n$, and
(2) follows because we have $\mathcal{H}(Z) \leq \frac{1}{n}H(Z^n)$. To show this last inequality we have to show that the sequence $f_n = \frac{1}{n}H(Z^n)$ is a non-increasing sequence because we know that $\mathcal{H}(Z) = \lim_{n\to\infty} f_n$.

To this end we write

$$
\begin{aligned}
f_{n+1} - f_n &= \frac{1}{n+1}H(Z^{n+1}) - \frac{1}{n}H(Z^n) \\
&= \frac{1}{n+1}\sum_{i=1}^{n+1} H(Z_i|Z^{i-1}) - \frac{1}{n}\sum_{i=1}^{n} H(Z_i|Z^{i-1}) \\
&= \frac{1}{n+1}H(Z_{n+1}|Z^n) + \sum_{i=1}^{n}\left[\frac{1}{n+1}H(Z_i|Z^{i-1}) - \frac{1}{n}H(Z_i|Z^{i-1})\right] \\
&= \frac{1}{n+1}H(Z_{n+1}|Z^n) - \frac{1}{n(n+1)}\sum_{i=1}^{n} H(Z_i|Z^{i-1}) \\
&= \frac{1}{n+1}\left[H(Z_{n+1}|Z^n) - \frac{1}{n}\sum_{i=1}^{n} H(Z_i|Z^{i-1})\right].
\end{aligned}
$$

From the lecture we know that for the stationary process $\{Z_i\}$ the sequence $H(Z_i|Z^{i-1})$ is a non-increasing sequence so we have

$$
H(Z_{n+1}|Z^n) \leq H(Z_i|Z^{i-1}) \quad \text{For } i = 1, \ldots, n,
$$

so we have

$$
\begin{aligned}
f_{n+1} - f_n &= \frac{1}{n+1}\left[H(Z_{n+1}|Z^n) - \frac{1}{n}\sum_{i=1}^{n} H(Z_i|Z^{i-1})\right] \\
&\leq 0,
\end{aligned}
$$

which shows that $\mathcal{H}(Z) \leq \frac{1}{n}H(Z^n)$.

(b) From part (a) we know that we can choose the input distribution such that makes the output distribution uniform so make the inequality (1) tight. Then we can observe that by increasing the block length n we have $\frac{1}{n}H(Z^n) \to \mathcal{H}(Z)$ so we have shown that by increasing the block length and choosing the input distribution properly we can achieve the upper bound in part (a).

(c) In part (b) we have shown that we can achieve

$$I(X^n; Y^n) = n(1 - \mathcal{H}(Z)).$$

From part (a) we know that

$$\mathcal{H}(Z) \leq f_n \leq f_{n-1} \leq \cdots \leq f_1 = H(Z_1) = h_2(\epsilon),$$

so we have that

$$I(X^n; Y^n) = n(1 - \mathcal{H}(Z)) \geq n(1 - h_2(\epsilon)) = n \cdot C,$$

where C is the capacity of the BSC is it is memory-less.