Solutions: Homework Set # 7

Problem 1

(a) We need to prove that

$$(x^n, y^n) \in A_\epsilon^{(n)}(X,Y) \Leftrightarrow x^n \in A_\epsilon^{(n)}(X), z^n \in A_\epsilon^{(n)}(Z),$$

where $y^n = x^n \oplus z^n$. First we calculate some entropies:

$$\begin{aligned} H(X) =& 1 \\ H(Z) =& H(Y|X) = -0.1\log(0.1) - 0.9\log(0.9) = H_2(0.1) \\ H(X,Y) =& H(X) + H(Y|X) = 1 + H(Y|X) = 1 + H_2(0.1). \end{aligned}$$

Notice that $p(x^n) = (\frac{1}{2})^n$, $\forall x^n$, hence $-\frac{1}{n}\log p(x^n) = 1$ which implies $x^n \in A_\epsilon^{(n)}(X)$ for any x^n. On the other hand, $(x^n, y^n) \in A_\epsilon^{(n)}(X,Y)$ implies that $-\frac{1}{n}\log p(x^n, y^n) \in (H(X,Y) - \epsilon, H(X,Y) - \epsilon)$. Since $p(x^n, y^n) = (\frac{1}{2})^n(1-p)^{n-k}p^k = p(x^n)p(z^n)$, where k is the number of places x and y differ, it follows that

$$\begin{aligned} -\frac{1}{n}\log p(x^n, y^n) =& -\frac{1}{n}\log(\frac{1}{2})^n(1-p)^{n-k}p^k \\ =& -\frac{1}{n}\log p(x^n)p(z^n) \\ =& -\frac{1}{n}\log p(x^n) - \frac{1}{n}\log p(z^n) \\ =& 1 - \frac{1}{n}\log p(z^n) \in (H(X,Y) - \epsilon, H(X,Y) - \epsilon). \end{aligned}$$

So, $-\frac{1}{n}\log p(z^n) \in (H(Z) - \epsilon, H(Z) - \epsilon)$, i.e. z^n is typical. Using the same equations in reverse direction, we see that assuming typical z^n implies that $-\frac{1}{n}\log p(x^n, y^n) \in (H(X,Y) - \epsilon, H(X,Y) - \epsilon)$.

It remains to show that y^n is typical. Notice that $y_i = x_i \oplus z_i$. Then

$$\begin{aligned} p(Y_i = 0) =& p(X_i = 0, Z_i = 0) + p(X_i = 1, Z_i = 1) \\ =& p(X_i = 0)p(Z_i = 0) + p(X_i = 1)p(Z_i = 1) \\ =& \frac{1}{2}(1-p) + \frac{1}{2}p = \frac{1}{2}. \end{aligned}$$

Hence, $-\frac{1}{n}\log p(y^n) = 1$, $\forall y^n$ i.e. every y^n is typical.

(b) For a fixed codeword $x^n(i)$, the event that $(x^n(i), Y^n)$ is jointly typical is equivalent to the event that $z^n \in A_\epsilon^{(n)}(Z)$. From the AEP, for large enough n,

$$Pr\left(z^n \in A_\epsilon^{(n)}(Z)\right) > 1 - \epsilon$$

(c) Let A denote the event that there exists a codeword $x^n(i),\ i \neq 1$ which is jointly typical with Y_n. Using the union bound and the fact that $x^n(i)$'s are $i.i.d$, we see that

$$\begin{aligned} Pr\left(A\right) =& Pr\left(E_2 \cup \ldots \cup E_{2^{nR}} \middle| \, x^n(1) \text{ was sent}\right) \\ \leq& \sum_{i=1}^{2^{nR}} Pr(E_i | x^n(1)) \\ =& \sum_{i=1}^{2^{nR}} Pr(E_2 | x^n(1)) \\ \leq& 2^{nR} Pr\left((x^n, Y^n) \in A_\epsilon^{(n)}(X,Y)\right). \end{aligned}$$

By joint AEP We know that the probability that y^n is jointly typical with an independent $x^n(i)$ is $2^{-n(I(X;Y)-R)}$. So,

$$Pr\left(A\right) \leq 2^{n\epsilon} 2^{-n(I(X;Y)-R)}.$$

(d)

$$\begin{aligned} Pr(\text{error} | x^n(1) \text{ was sent}) \leq& Pr((x^n, y^n) \notin A_\epsilon^{(n)}(X,Y) + Pr(A) \\ <& \epsilon + 2^{n\epsilon} 2^{-n(I(X;Y)-R)}, \end{aligned}$$

which is smaller than 2ϵ when $R < I(X;Y)$ and n is sufficiently large. The probability of error is then

$$Pr(\text{error}) = \sum_{i=1}^{2nR} Pr(\text{error} | x^n(i) \text{ was sent}) Pr(x^n(i) \text{ was sent}) = Pr(\text{error} | x^n(1) \text{ was sent}) < 2\epsilon.$$

Problem 2

(a) Each of the channels are binary symmetric, and their capacity can be found as

$$C_G = 1 - H_2(P_G)$$

$$C_B = 1 - H_2(P_B)$$

As we have seen before, the optimal input distribution for a binary symmetric channel is the uniform distribution, regardless of the cross-over probability of the channel:

$$\Pr(X = 0) = \Pr(X = 1) = \frac{1}{2}.$$

(b) Having the transition matrix $P = \begin{bmatrix} 1-g & g \\ b & 1-b \end{bmatrix}$, the stationary distribution of the Markov process is $\pi = [\pi_B \ \pi_G]$ such that $\pi = \pi P$.

$$\begin{aligned} \pi_B = \pi_B(1-g) + \pi_G b = \pi_B(1-g) + (1-\pi_B)b \\ \Rightarrow \begin{cases} \pi_B = \frac{b}{b+g} \\ \pi_G = \frac{g}{b+g}. \end{cases} \end{aligned}$$

(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 $R = \pi_B C_B + \pi_G C_G = 1 - \frac{bH_2(P_B)+gH_2(P_G)}{b+g}$ is achievable. We propose the following simple scheme to transmit at rate R. The transmitter designs two capacity achieving codes $\mathcal{C}_B$ and $\mathcal{C}_G$ for transmission over the bad and good channels, respectively. Depends 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_B fraction of time, and in the good state for the remaining π_G fraction. Since we transmit at capacity rate for each fraction, the total rate would be $R = \pi_B C_B + \pi_G C_G$.

(d) Since the channel alphabets for the states are two disjoint sets ($\mathcal{Y}_B = \{0, 1\}$ and $\mathcal{Y}_G = \{2, 3\}$), the receiver can determine the channel state based on the received symbol.

(e)

$$\begin{aligned}
C_{NSI} &\overset{(a)}{=} \max_{p(x)} \frac{1}{n} I(X^n; Y^n, S^n) \\
&\overset{(b)}{=} \max_{p(x)} \frac{1}{n} I(X^n; Y^n | S^n) \\
&\overset{(c)}{=} \max_{p(x)} \frac{1}{n} \left(H(Y^n|S^n) - H(Y^n|X^n, S^n) \right) \\
&\overset{(d)}{\leq} \max_{p(x)} \frac{1}{n} \left(\sum_{i=1}^{n} H(Y_i|S^n) - \sum_{i=1}^{n} H(Y_i|X_i, S_i) \right) \\
&\overset{(e)}{\leq} \max_{p(x)} \frac{1}{n} \left(\sum_{i=1}^{n} H(Y_i|S_i) - \sum_{i=1}^{n} H(Y_i|X_i, S_i) \right) \\
&\overset{(f)}{=} \max_{p(x)} \frac{1}{n} \sum_{i=1}^{n} I(X_i; Y_i|S_i)
\end{aligned}$$

- (a) is just the definition of the capacity, given the fact that the receiver can determine the state of the channel from the output;
- (b) holds since the transmitter has no side information about the channel state, and so X^n is independent of S^n:

$$I(X^n; Y^n, S^n) = \underbrace{I(X^n; S^n)}_{0} + I(X^n; Y^n|S^n).$$

- (c) is just expansion of the mutual information.
- in (d), we have used the chain rule, $H(Y^n|S^n) = \sum_{i=1}^{n} H(Y_i|Y^{i-1}, S^n)$ and increased each term by removing Y^{i-1} from conditioning. Note that conditioning reduces the entropy. The same argument holds because the first summation is the expansion of $H(Y^n|S^n)$ using the chain rule. The second term is replaced by a larger term, since removing conditioning reduces the entropy. Note that the output of the channel only depends on its input and state, and is independent of anything else given X and S. Therefore, this inequality is tight.

- (e) is again by the fact that conditioning reduces entropy. This is also tight, since there is no feedback in the channel, and therefore previous states of the channel do not affect the input nor the output of the channel.
- (f) is by definition of mutual information. Note that is the capacity where the current state of the channel is known.

Problem 3

(a) To find the capacity of the product channel, we must 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 joint distribution $p(x_1, x_2, y_1, y_2) = p(x_1, x_2)p(y_1|x_1)p(y_2|x_2)$,

$$Y_1 \to X_1 \to X_2 \to Y_2$$

forms a Markov chain and thus

$$\begin{aligned}
I(X_1, X_2; Y_1, Y_2) &= H(Y_1, Y_2) - H(Y_1, Y_2|X_1, X_2) && (1)\\
&= H(Y_1, Y_2) - H(Y_1|X_1, X_2) - H(Y_2|X_1, X_2) && (2)\\
&= H(Y_1, Y_2) - H(Y_1|X_1) - H(Y_2|X_2) && (3)\\
&\leq H(Y_1) + H(Y_2) - H(Y_1|X_1) - H(Y_2|X_2) && (4)\\
&= I(X_1; Y_1) + I(X_2; Y_2) && (5)
\end{aligned}$$

where the second and third equalities follow from the stated Markovity. Furthermore, the inequality 4 holds with equality for independent X_1 and X_2. Thus

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

(b) 1.

$$\begin{aligned}
I(X; Y_1, Y_2) &= H(Y_1, Y_2) - H(Y_1, Y_2|X)\\
&= H(Y_1, Y_2) - H(Y_1|X) - H(Y_2|X)\\
&= H(Y_1) + H(Y_2|Y_1) - 2H(Y_1|X)\\
&= H(Y_1) + H(Y_2) - 2H(Y_1|X) - I(Y_1; Y_2)\\
&= 2H(Y_1) - 2H(Y_1|X) - I(Y_1; Y_2)\\
&= 2I(X; Y_1) - I(Y_1; Y_2),
\end{aligned}$$

and the fifth equality follows from

$$p(y_1) = \sum_x p(y_1, x) = \sum_x p(x)p(y_1|x) = \sum_x p(x)p(y_2|x) = \sum_x p(x, y_2) = p(y_2).$$

2.

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

Problem 4

As this is a symmetric channel the capacity is achieved by using an uniform input distribution $p(i) = \frac{1}{|\mathcal{X}|} = \frac{1}{5}$.

(a)

$$\begin{aligned} C =& I(X_{\text{uniform}}; Y) \\ =& H(X) - H(Y|X) \\ =& \sum_{i=1}^{5} \frac{1}{5} \log 5 - \sum_{i=1}^{5} \frac{1}{5} H(Y|X = i) \\ =& \log 5 - 1. \end{aligned}$$

(b) Consider the following code:

$$\begin{aligned} f(m_0) =& 00 \\ f(m_1) =& 12 \\ f(m_2) =& 24 \\ f(m_3) =& 31 \\ f(m_4) =& 43. \end{aligned}$$

One can verify that the possible outputs for any two of these codewords are distinct. Hence, given the output, we can never confuse different inputs.

If the messages $m_0 \ldots m_4$ are equiprobable, we are essentially transmitting $\log 5$ bits of information while using the channel only twice. The rate of this zero-error code is therefore $\frac{\log 5}{2}$. Notice that $1 < \frac{\log 5}{2} < \log 5 - 1$. In fact, the zero-error capacity for this channel is equal to the Shannon capacity. For more details, see "On the Shannon capacity of a graph" by Laszlo Lovasz, IEEE Transactions on Information Theory, vol. IT-25, pp 1-7, Jan 1979.