1. Problem 2.4 parts (a), (d), and (e).

(a) Consider

\[ H(X|Z) \leq H(X,Y|Z) = H(Y|Z) + H(X|Y,Z) \leq H(Y|Z) + H(X|Y). \]

(d) Consider

\[
I(X_1, X_2; Y_1, Y_2) = I(X_1, X_2; Y_1) + I(X_1, X_2; Y_2 | Y_1) \\
= I(X_1; Y_1) + I(X_2; Y_1 | X_1) + I(X_2; Y_2 | Y_1) + I(X_1; Y_2 | X_2, Y_1) \\
\overset{(a)}{=} I(X_1; Y_1) + I(X_2; Y_2 | Y_1) \\
\leq I(X_1; Y_1) + I(X_2, Y_1; Y_2) \\
= I(X_1; Y_1) + I(X_2; Y_2) + I(Y_1; Y_2 | X_2) \\
\overset{(b)}{=} I(X_1; Y_1) + I(X_2; Y_2),
\]

where (a) and (b) follow since \( p(y_1, y_2 | x_1, x_2) = p(y_1 | x_1) p(y_2 | x_2) \), which implies that \( Y_1 \rightarrow X_1 \rightarrow X_2 \rightarrow Y_2 \) form a Markov chain.

(e) Consider

\[
I(X_1, X_2; Y_1, Y_2) = I(X_1; Y_1, Y_2) + I(X_2; Y_1, Y_2 | X_1) \\
= I(X_1; Y_1) + I(X_2; Y_2 | Y_1) + I(X_2; Y_2 | X_1) + I(X_2; Y_1 | X_1, Y_2) \\
\geq I(X_1; Y_1) + I(X_2; Y_2 | X_1) \\
\overset{(a)}{=} I(X_1; Y_1) + I(X_2; Y_2, X_1) \\
\geq I(X_1; Y_1) + I(X_2; Y_2),
\]

where (a) follows by the independence of \( X_1 \) and \( X_2 \).


Consider

\[
P\{ (X^n, Y^n) \notin \mathcal{T}_\epsilon^{(n)} | X^n \in \mathcal{T}_\epsilon^{(n)} \} \\
= \sum_{x^n \notin \mathcal{T}_\epsilon^{(n)}} P\{ (X^n, Y^n) \notin \mathcal{T}_\epsilon^{(n)} | X^n \in \mathcal{T}_\epsilon^{(n)}, X^n = x^n \} P\{ X^n = x^n | X^n \in \mathcal{T}_\epsilon^{(n)} \} \\
\leq \max_{x^n \in \mathcal{T}_\epsilon^{(n)}} P\{ (X^n, Y^n) \notin \mathcal{T}_\epsilon^{(n)} | X^n \in \mathcal{T}_\epsilon^{(n)}, X^n = x^n \} \\
= \max_{x^n \in \mathcal{T}_\epsilon^{(n)}} P\{ (X^n, Y^n) \notin \mathcal{T}_\epsilon^{(n)} | x^n \in \mathcal{T}_\epsilon^{(n)} \},
\]

which, by the conditional typicality lemma, tends to zero as \( n \rightarrow \infty \).
3. Problem 2.14 parts (a) and (d).

(a)
\[
P\{(X^n, Y^n, \tilde{Z}^n) \in \mathcal{T}^{(n)}_e(X, Y, Z)\} = \sum_{(x^n, y^n, \tilde{z}^n) \in \mathcal{T}^{(n)}_e(X, Y, Z)} p(x^n, y^n) p(\tilde{z}^n | x^n)
\]
\[
\doteq |\mathcal{T}^{(n)}_e(X, Y, Z)| 2^{-nH(X,Y)} 2^{-nH(Z|X)}
\]
\[
\doteq 2^{nH(X,Y)Z} 2^{-nH(X,Y)1-nH(Z|X)}
\]
\[
= 2^{-nH(Y;Z|X)}.
\]

(d)
\[
P\{(\tilde{X}^n, \tilde{Y}^n, \tilde{Z}^n) \in \mathcal{T}^{(n)}_e(X, Y, Z)\} = \sum_{(\tilde{x}^n, \tilde{y}^n, \tilde{z}^n) \in \mathcal{T}^{(n)}_e(X, Y, Z)} p(\tilde{x}^n) p(\tilde{y}^n | \tilde{z}^n) p(\tilde{z}^n | \tilde{z}^n)
\]
\[
\doteq |\mathcal{T}^{(n)}_e(X, Y, Z)| 2^{-nH(X)} 2^{-nH(Y)} 2^{-nH(Z|X,Y)}
\]
\[
\doteq 2^{nH(X,Y)Z} 2^{-nH(Y)} 2^{-nH(Z|X,Y)}
\]
\[
= 2^{-nI(X;Y)}.
\]


Consider
\[
p(y^n | x^n, m) = \prod_{i=1}^n p(y_i | y^{i-1}, x^i, x_{i+1}^n, m)
\]
\[
= \prod_{i=1}^n p(y_i | y^{i-1}, x^i, x_{i+1}^n, m)
\]
\[
= \prod_{i=1}^n p(y_i | y^{i-1}, x^i, x_{i+1}^n, m)
\]
\[
= \prod_{i=1}^n p(y_i | y^{i-1}, x^i, x_{i+1}^n, m)
\]
\[
= \prod_{i=1}^n p(y_i | y^{i-1}, x^i, m)
\]
\[
= \prod_{i=1}^n p(y_i | y^{i-1}, x^i, m)
\]
\[
= \prod_{i=1}^n p(y_i | y^{i-1}, x^i, m)
\]
\[
= \prod_{i=1}^n p(y_i | y^{i-1}, x^i, m)
\]
\[
= \prod_{i=1}^n p_Y | X (y_i | x_i),
\]

where (a) follows since \(x_{i+1}^n\) is a function of \(m\), which implies \((Y^i, X^i) \to M \to X_{i+1}^n\) form a Markov chain and (b) follows since the channel is memoryless, that is, \((M, X_{i-1}^i, Y_{i-1}^i) \to X_i \to Y_i\) form a Markov chain and \(P\{Y_i = y | X_i = x\} = p_{Y | X}(y | x)\).

5. Problem 3.5.
(a) An index $\hat{m}$ will be chosen by MLD if for every index $m \neq \hat{m}$, \( p(g^n|x^n(\hat{m})) > p(g^n|x^n(m)) \), or equivalently, 
\[
p^{d(x^n(\hat{m}), y^n)}(1 - p)^{n-d(x^n(\hat{m}), y^n)} > p^{d(x^n(m), y^n)}(1 - p)^{n-d(x^n(m), y^n)},
\]
or equivalently, 
\[
d(x^n(\hat{m}), y^n) < d(x^n(m), y^n) \text{ considering the fact that } p < 1/2.
\]

(b) Let \( \mathcal{E} = \{ d(X^n(1), Y^n) > n(p + \epsilon) \} \). Then, the probability of error averaged over codebooks is upper bounded as
\[
P_e^{(n)} = P\{ \hat{M} \neq 1 | M = 1 \}
= P\{ d(X^n(m), Y^n) < d(X^n(1), Y^n) \text{ for some } m \neq 1 | M = 1 \}
= P\{ \mathcal{E}, d(X^n(m), Y^n) < d(X^n(1), Y^n) \text{ for some } m \neq 1 | M = 1 \}
+ P\{ \mathcal{E}^c, d(X^n(m), Y^n) < d(X^n(1), Y^n) \text{ for some } m \neq 1 | M = 1 \}
\leq P\{ \mathcal{E} | M = 1 \}
+ P\{ \mathcal{E}^c \text{ and } d(X^n(m), Y^n) < d(X^n(1), Y^n) \text{ for some } m \neq 1 | M = 1 \}
\leq P\{ d(X^n(1), Y^n) > n(p + \epsilon) | M = 1 \}
+ (2^{nR} - 1) P\{ d(X^n(2), Y^n) \leq n(p + \epsilon) | M = 1 \}.
\]

(c) We know that \( d(X^n(1), Y^n) \) is a binomial random variable with mean \( np \) and variance \( np(1-p) \) given \( \{ M = 1 \} \). Therefore, by the Chebyshev inequality,
\[
P\{ d(X^n(1), Y^n) - np > n\epsilon | M = 1 \} \leq \frac{p(1-p)}{n\epsilon^2},
\]
which tends to zero as \( n \to \infty \). Also, by the Chernoff bound, the second term tends to zero as \( n \to \infty \), provided that \( R < 1 - H(p + \epsilon) \). Therefore, by taking \( \epsilon \to 0 \), any rate \( R < 1 - H(p) = C \) is achievable.

6. Problem 3.7.

(a) We will use the probabilistic method to show that the average probability error over an ensemble of codes tends to zero, from which we can conclude the existence of \( (\phi'_n, \psi'_n) \) with \( P\{ \hat{M}' \neq M' \} \to 0 \) as \( n \to \infty \).

Consider a random permutation \( \sigma \) drawn uniformly from all permutations on \([1 : 2^{nR}]\). (Alternatively, we can consider a random permutation \( \sigma(x) = (x + Z) \mod 2^{nR} \), where \( Z \sim \text{Unif}[1 : 2^{nR}] \).) Then \( \sigma(M') \sim \text{Unif}[1 : 2^{nR}] \). Given \( (\phi_n, \psi_n) \) and \( \sigma \), we consider the encoding function \( \phi'_n(m') = \phi_n(\sigma(m')) \) and the decoding function \( \psi'_n(y^n) = \psi^{-1}(\psi_n(y^n)) \). Now the average probability of error (averaged over the random permutation \( \sigma \)) is
\[
P\{ \hat{M}' \neq M' \} = \sum_m P\{ \sigma(M') = m \} P\{ \hat{M}' \neq M' | \sigma(M') = m \}
= \frac{1}{2^{nR}} \sum_m P\{ \hat{M}' \neq \sigma^{-1}(m) | \sigma(M') = m \}
= \frac{1}{2^{nR}} \sum_m P\{ \sigma(\hat{M}') \neq m | \sigma(M') = m \}
\leq P_{e}^{(n)},
\]

where \((a)\) follows since \(\sigma(M') \sim \text{Unif}[1 : 2^nR]\) and \((b)\) follows by the assumption on the code \((\phi_n, \psi_n)\) for the uniform message. Since the average probability of error is equal to \(P_e^n\), there must exist at least one permutation \(\sigma\) and \((\phi_n', \psi_n')\) with \(P\{M' \neq \hat{M}'\} \leq P_e^n\), which tends to zero as \(n \to \infty\). Alternative proof: By assigning the most likely message to the codeword with the smallest probability of error, the second most likely message to the codeword with the second smallest probability of error, and so on, we can achieve an average probability of error for the non-uniformly distributed message \(M'\) no greater than that for the uniformly distributed message \(M\).

(b) No. Part \((a)\) simply implies that for every distribution on \(M\), there exists a good code (depending on the distribution), while the capacity for the maximal probability of error requires a code that is \(\text{uniformly}\) good for every singleton distribution on \(M\) (and consequently on every distribution).

7. Binary multiplier channel (CT).

(a) Let \(X \sim \text{Bern}(p)\). Then, \(P\{Y = 1\} = \alpha p, H(Y) = H(\alpha p), \) and \(H(Y|X) = pH(Y|X = 1) + (1 - p)H(Y|X = 0) = pH(\alpha).\) Thus,

\[
I(X; Y) = H(\alpha p) - pH(\alpha).
\]

To find \(p\) that maximizes \(I(X; Y)\), take the derivative of \(I(X; Y)\) with respect to \(p\) and find \(p^*\) that makes the derivative equals zero.

\[
\alpha \log(1 - \alpha p^*) - \alpha \log(\alpha p^*) = H(\alpha),
\]

\[
p^* = \frac{1}{\alpha(2^{H(\alpha)/\alpha} + 1)}.
\]

By plugging \(p^*\) into \(I(X; Y)\), we get

\[
C = \log \left(\frac{2^{H(\alpha)/\alpha} + 1}{2^{H(\alpha)/\alpha}}\right).
\]

(b) Note the following.

\[
I(X; Y, Z) = I(X; Z) + I(X; Y|Z)
\]

\[
\stackrel{(a)}{=} I(X; Y|Z)
\]

\[
\stackrel{(b)}{=} H(Y|Z)
\]

Step \((a)\) follows since \(X\) and \(Z\) are independent. Step \((b)\) holds since \(Y = XZ\).

\[
C = \max_{p(x)} H(Y|Z) = \max_{p(x)} (1 - \alpha)H(X) = 1 - \alpha.
\]

Note that \(C\) is the same as the capacity of binary erasure channel with erasure probability \(\alpha\).
8. Cascade of binary symmetric channels (CT).

We only need to consider the crossover probability \( p \leq 0.5 \). If \( p > 0.5 \), the receiver can flip the output, in which case, the crossover probability is \( 1 - p \leq 0.5 \). We use induction to show that the crossover probability is \( (1/2)(1 - (1 - 2p)^n) \). For \( n = 1 \), we can easily check that \( (1/2)(1 - (1 - 2p)^n) = p \). Now assume the crossover probability at \( X_{n-1} \) is \( (1/2)(1 - (1 - 2p)^{n-1}) \). We now show that the crossover probability at \( X_n \) is \( (1/2)(1 - (1 - 2p)^n) \). The cross over probability at \( X_n \) is as follows.

\[
P\{X_n \neq X_0|X_0\} = P\{X_0 \neq X_{n-1}, X_{n-1} = X_n|X_0\} + P\{X_0 = X_{n-1}, X_{n-1} \neq X_n|X_0\}
\]

\[
= (1/2)(1 - (1 - 2p)^{n-1})(1 - p) + (1/2)(1 + (1 - 2p)^{n-1})p
\]

\[
= (1/2)(1 - (1 - 2p)^n).
\]

As \( n \to \infty \), the crossover probability goes to \( 1/2 \), and so \( H(X_0|X_n) \to H(X_0) \).

\[
\lim_{n \to \infty} I(X_0; X_n) = \lim_{n \to \infty} \{H(X_0) - H(X_0|X_n)\} = 0.
\]

Extra problem solutions

1. Entropy of a function of a random variable.

Consider

\[
H(g(X), X) = H(X) + H(g(X)|X) = H(X)
\]

\[
= H(g(X)) + H(X|g(X)) \geq H(g(X)).
\]

Equality holds iff \( H(X|g(X)) = 0 \), i.e., \( g(X) \) is an invertible function.

2. Inequalities (CT).

(a) \( = \), the function \( g(x) = 5x \) is a one-to-one function.

(b) \( \leq \), \( I(X; Y) = I(X, g(X); Y) = I(g(X); Y) + I(X; Y|g(X)) \geq I(g(X); Y) \). Equality holds if \( X \to g(X) \to Y \).

(c) \( = \), can be checked by rearranging the terms.

(d) \( \geq \), Since \( I(X; Z|Y) \geq I(X; Z) \), \( I(X, Y; Z) \geq I(X; Z) + I(Y; Z) \).

3. Conditional mutual information vs. unconditional mutual information (CT).

(a) If \( X \to Y \to Z \), then \( I(X; Y) \geq I(X; Y|Z) \). Equality holds if and only if \( I(X; Z) = 0 \) or \( X \) and \( Z \) are independent. As a simple example of random variable triples satisfying the strict inequality, consider \( X \sim \text{Bern}(1/2) \), \( Y = Z = X \) so that \( I(X; Y) = H(X) - H(X|Y) = H(X) = 1 \), and \( I(X; Y|Z) = H(X|Z) - H(X|Y, Z) = 0 \).

(b) Let \( X \) and \( Y \) be i.i.d. \( \text{Bern}(1/2) \). Let \( Z = X + Y \). Then, \( I(X; Y) = 0 \) and \( I(X; Y|Z) = H(X|Z) = 1/2 \).

4. Entropy of a disjoint mixture (CT).
(a) Since the mixture is disjoint, $S$ can be determined given $X$.

\[ H(X) = H(X, S) = H(S) + p_S(1)H(X|S = 1) + p_S(0)H(X|S = 0) = H(p_1) + p_1H(X_1) + (1 - p_1)H(X_0). \]

(b) We have $dH(X)/dp_1 = \log(1 - p_1) - \log p_1 + H(X_1) - H(X_0)$.

Setting $dH(X)/dp_1 = 0$, the $p_1^*$ that maximizes $H(X)$ is

\[ p_1^* = \frac{2^{H(x_1)}}{2^{H(x_1)} + 2^{H(x_0)}}. \]

Therefore,

\[ 2^{H(x)} \leq \frac{dH(x)|_{p_1 = p_1^*}}{dp_1} = 2^{H(x_1)} + 2^{H(x_0)}. \]

(c) Consider $H(X|S) = H(X, S) - H(S) = p_1H(X_1) + (1 - p_1)H(X_0)$.

5. Data processing (CT).

Consider

\[ I(X_1; X_2, X_3, \ldots, X_n) = I(X_1; X_2) + I(X_1; X_3|X_2) + \ldots + I(X_1; X_n|X_1, \ldots, X_{n-1}) \]

\[ \overset{(b)}{=} I(X_1; X_2). \]

Step (a) holds by chain rule, and (b) by the independence of $X_1$ and $X_i$ given $X_{i-1}$.

6. Typical average lemma and typicality. Proof of the typical average lemma:

Rewrite the average function as follows.

\[ \frac{1}{n} \sum_{i=1}^{n} g(x_i) = \sum_{x \in \mathcal{X}} \pi(x|x^n)g(x). \]

For each $x^n \in \mathcal{T}_c^{(n)}(X)$,

\[ (1 - \epsilon)p(x) \leq \pi(x|x^n) \leq (1 + \epsilon)p(x) \]

for all $x \in \mathcal{X}$. Thus, for any nonnegative function $g(x)$,

\[ \sum_{x \in \mathcal{X}} (1 - \epsilon)p(x)g(x) \leq \sum_{x \in \mathcal{X}} \pi(x|x^n)g(x) \leq \sum_{x \in \mathcal{X}} (1 + \epsilon)p(x)g(x). \]

The typical average lemma follows by noting that $E[g(X)] = \sum_{x \in \mathcal{X}} p(x)g(x)$.

**Proof of Property 1:**

Let $g(x) = -\log p(x)$. Then, $\sum_{i=1}^{n} g(x_i) = -\log p(x^n)$ and $E(g(X)) = H(X)$. By the typical average lemma,

\[ (1 - \epsilon)H(X) \leq \frac{1}{n} \log p(x^n) \leq (1 + \epsilon)H(X), \]
which is equivalent to the following.

\[ 2^{-n(H(X)+\delta(\epsilon))} \leq p(x^n) \leq 2^{-n(H(X)-\delta(\epsilon))} \]

where \( \delta(\epsilon) = \epsilon H(X) \).

**Proof of Property 2:**

Property 2 can be shown as follows.

\[ 1 \geq \sum_{x^n \in X^n} p(x^n) \geq \sum_{x^n \in T_\epsilon^{(n)}(X)} 2^{-n(H(X)+\delta(\epsilon))} \]

\[ = |T_\epsilon^{(n)}(X)|2^{-n(H(X)+\delta(\epsilon))} \]

**Proof of Property 3:**

Let \( 1_{x_i=x} = 1 \) if \( x_i = x \), and 0 otherwise. Note that \( 1_{x_i=x} \) is i.i.d. By weak law of large numbers,

\[ \pi(x | X^n) = \frac{1}{n} \sum_{i=1}^{n} 1_{x_i=x} \rightarrow \mathbb{E}[1_{X_i=x}] = p(x) \]

in probability. That is to say for any \( \epsilon > 0 \),

\[ \lim_{n \rightarrow \infty} P\{|\pi(x | X^n) - p(x)| > \epsilon p(x)\} = 0. \]

We can show \( \lim_{n \rightarrow \infty} P\{X^n \notin T_\epsilon^{(n)}(X)\} = 0 \) using the union bound.

\[ P\{X^n \notin T_\epsilon^{(n)}(X)\} = P\{|\pi(x | X^n) - p(x)| > \epsilon p(x) \text{ for any } x \in X\} \leq \sum_{x \in X} P\{|\pi(x | X^n) - p(x)| > \epsilon p(x)\}. \]

**Proof of Property 4:**

By Property 3 and Property 1, for \( n \) sufficiently large,

\[ (1 - \epsilon) \leq P\{X^n \in T_\epsilon^{(n)}(X)\} \leq |T_\epsilon^{(n)}(X)|2^{-n(H(X)-\delta(\epsilon))}. \]

7. **Proof of Property 1(c):**

By the typical average lemma, for \((x^n, y^n) \in T_\epsilon^{(n)}(X,Y)\) and nonnegative \( g(x, y) \),

\[ (1 - \epsilon)E(g(X,Y)) \leq \frac{1}{n} \sum_{i=1}^{n} g(x_i, y_i) \leq (1 + \epsilon)E(g(X,Y)). \]
Setting $g(x, y) = -\log p(x|y)$, we obtain

$$(1 - \epsilon)H(X|Y) \leq -\frac{1}{n} \sum_{i=1}^{n} \log p_{X|Y}(x_i|y_i) \leq (1 + \epsilon)H(X|Y),$$

or equivalently,

$$2^{-n(H(X|Y)+\delta(\epsilon))} \leq p(x^n|y^n) \leq 2^{-n(H(X|Y)-\delta(\epsilon))},$$

where $\delta(\epsilon) = \epsilon H(X|Y)$.

Alternatively, we can combine the bounds

$$2^{-n(H(X,Y)+\epsilon H(X,Y))} \leq p(x^n,y^n) \leq 2^{-n(H(X,Y)-\epsilon H(X,Y))}$$

and

$$2^{-n(H(Y)+\epsilon H(Y))} \leq p(y^n) \leq 2^{-n(H(Y)-\epsilon H(Y))}$$

and use the fact that $p(x^n|y^n) = p(x^n,y^n)/p(y^n)$ to obtain

$$2^{-n(H(X|Y)+\delta(\epsilon))} \leq p(y^n) \leq 2^{-n(H(Y)-\delta(\epsilon))},$$

where $\delta(\epsilon) = \epsilon(H(X,Y) + H(Y))$, which is looser than what we obtained above. The inequalities for $p(y^n|x^n)$ follows similarly.

**Proof of Property 3:**

Assume without loss of generality that $x^n \in \mathcal{T}^{(n)}_\epsilon(Y|x^n)$; otherwise, $|\mathcal{T}^{(n)}_\epsilon(Y|x^n)| = 0$. Consider

$$1 = \sum_{y^n} p(y^n|x^n)$$

$$\geq \sum_{y^n \in \mathcal{T}^{(n)}_\epsilon(Y|x^n)} p(y^n|x^n)$$

$$\geq |\mathcal{T}^{(n)}_\epsilon(Y|x^n)| 2^{-n(H(X|Y)+\delta(\epsilon))},$$

which implies that

$$|\mathcal{T}^{(n)}_\epsilon(Y|x^n)| \leq 2^{n(H(X|Y)+\delta(\epsilon))},$$

where $\delta(\epsilon) = \epsilon H(X|Y)$.

**Proof of Property 4:**

First assume that $y_i = g(x_i)$ for all $i \in [1:n]$. Then

$$|\{i : (x_i, y_i) = (x, y)\}| = |\{i : (x_i, g(x_i)) = (x, y)\}|$$

$$= \begin{cases} |\{i : x_i = x\}| & \text{if } y = g(x), \\ 0 & \text{otherwise}. \end{cases}$$

Now that $p(x,y) = p(x)$ if $y = g(x)$ and $p(x,y) = 0$ if $y \neq g(x)$, we have for all $(x,y)$ that

$$(1 - \epsilon)p(x,y) \leq \pi(x,y|x^n,y^n) \leq (1 + \epsilon)p(x,y),$$

or equivalently, $(x^n,y^n) \in \mathcal{T}^{(n)}_\epsilon(X,Y)$. Conversely, assume that $(x^n,y^n) \in \mathcal{T}^{(n)}_\epsilon(X,Y)$. Then, $\pi(x,y|x^n,y^n) = 0$ whenever $p(x,y) = 0$. In other words, for every $i \in [1:n]$, $y_i = g(x_i)$. 

8
8. **Problem 3.2.** Let \( X \sim \text{Bern}(\alpha) \). Then \( I(X;Y) = H(Y) - H(Y|X) = H(\alpha/2) - \alpha \).

Using elementary calculus, we have
\[
\frac{d}{d\alpha} I(X;Y) = \frac{1}{2} \log \frac{1 - \alpha/2}{\alpha/2} - 1,
\]
which is equal to zero for \( \alpha = 2/5 \). Hence, the capacity of the Z-channel is \( C = H(1/5) - 2/5 = 0.322 \) bits per transmission.

9. (a) Due to the symmetry of the transition probability matrix for inputs 0 and 1, \( I(X;Y) \) as a function of \( P\{X = 1\} = p \) is symmetric with respect to \( p = 1/2 \). Since \( I(X;Y) \) is concave with respect to \( p \), \( I(X;Y) \) is maximized at \( p = 1/2 \). Thus, capacity is the evaluation of \( I(X;Y) \) with \( p = 1/2 \). With \( X \sim \text{Bern}(1/2) \), \( P\{Y = 0\} = P\{Y = 1\} = (1 - \alpha)/2 \), and \( P\{Y = \epsilon\} = \alpha \). Thus,
\[
C = -(1 - \alpha) \log(\frac{1 - \alpha}{2}) - \alpha \log(\alpha) - H(1 - \alpha - \epsilon, \epsilon, \alpha).
\]

(b) If \( \alpha = 0 \), \( C = 1 - H(\epsilon) \).

(c) If \( \epsilon = 0 \), \( C = 1 - \alpha \).