1)(a) \( E = \{01, 00, 101, 100, 110, 111\} \) is prefix-free and has lower rate \( \Rightarrow \) code is not optimal
\( \Rightarrow \) code is not a Huffman code.

(b) \( H(x) \leq 2 \leq H(x) + \frac{1}{2} \)
\( H(x) \leq 1.6 \leq H(x) + \frac{1}{3} \)

\( \Rightarrow \)
\( 2 \geq H(x) \geq 1.5 \)
\( 1.6 \geq H(x) \geq 1.6 - \frac{1}{3} \)

\( \Rightarrow \)
\( 1.5 \leq H(x) \leq 1.6 \)

We know \( 0 \leq H(x) \leq \log_2 Q \Rightarrow \log_2 Q \geq 1.5 \)
\( \Rightarrow Q \geq 2^{1.5} = 2.83 \)
\( \Rightarrow Q \geq 3. \)

(Actually, you can show \( Q > 4 \) as follows: Suppose \( Q = 3 \). Take blocks \( N = 3 \) at a time and employ a code with codeword lengths \( \{3, 3, 3, 3, 3, 3, 4, 4\} \). The maximal rate of this code is
\( \frac{1}{2} \cdot \frac{1}{9} (3 \cdot 7 + 2 \cdot 4) = 1.61 < 2 \)

This implies \( R_{\text{HUFF}} \leq 1.61 \), which violates our assumption. Thus, \( Q > 4 \).)
[c] No. For any \( m \) and \( n \) source, suppose I build a \((\text{suboptimal})\) lossless code on blocks of length \( mn \) by concatenating \( m \) Huffman codes of length \( n \).

Then

\[ R_{\text{subopt}}^{(mn)} = R^{(n)} \]

Since Huffman coding is optimal,

\[ R^{(mn)} \leq R_{\text{subopt}}^{(mn)} \leq R^{(n)} \]
\[ R = \frac{1}{1000} \left( 1 \cdot 512 + 3 \cdot 128 \cdot 3 + 5 \cdot 32 \cdot 3 + 5 \cdot 8 \cdot 1 \right) \]
\[ = 0.184 \ \text{output} \ \text{bits per input bit} \]
\[ = 0.728 \ \text{output bits per input bit} \]

\[ H(x) = \text{output bits} = 0.72193 \]

\[ \text{known} \quad H(x) \leq R_{\text{max}} \leq H(x) + 1/N \]
\[ 0.72193 \leq 0.728 \leq 1.05 \]
Form blocks

<table>
<thead>
<tr>
<th>Block</th>
<th>Probability</th>
<th>( p(i) \times 10^0 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>000</td>
<td>512</td>
<td>1</td>
</tr>
<tr>
<td>001</td>
<td>128</td>
<td>3</td>
</tr>
<tr>
<td>010</td>
<td>128</td>
<td>3</td>
</tr>
<tr>
<td>100</td>
<td>128</td>
<td>3</td>
</tr>
<tr>
<td>011</td>
<td>32</td>
<td>5</td>
</tr>
<tr>
<td>110</td>
<td>32</td>
<td>5</td>
</tr>
<tr>
<td>101</td>
<td>32</td>
<td>5</td>
</tr>
<tr>
<td>111</td>
<td>8</td>
<td>7</td>
</tr>
</tbody>
</table>

\[ R_{5-p} = \frac{1}{1000} \left( 1.512 + 3.128.3 + 5.32.3 + 7.31 \right) \]

\[ = 0.2 \text{ bits/block} \]

\[ = 0.733 \text{ output bits/input bit} \]

Know \( R_{\text{ Huffman}} \leq R_{5-p} \leq H(x) + 1 \) \( /N \)

\[ 0.728 \leq 0.733 \leq 1.05 \]
\( H((x_1)) = \frac{1}{n-1} \sum_{i=1}^{n} H(x_i, x_1, x_2, \ldots, x_n) \)

\[= \frac{1}{n-1} \sum_{i=1}^{n-1} H(x_1, x_i) \quad (\text{if } x_i \neq x_1) \]

\[= \frac{1}{n-1} \sum_{i=1}^{n-1} H(x_2, x_i) \quad (\text{if } x_1 = x_2) \]

\[= \frac{1}{n-1} \sum_{i=1}^{n-1} H(x_2, x_i) \quad (\text{if } x_1 = x_2) \]

\[= \frac{1}{n-1} \sum_{i=1}^{n-1} E_{x_1} [H(x_2|x)] \]

\[= N_2(0.9) \cdot 0.8 + N_2(0.6) \cdot 0.2 \]

\[= 0.419 \cdot 0.8 + 0.97045 \cdot 0.2 \]

\[= 0.569 \]

\( R_{\text{MUR}} = \frac{1}{n-1} \sum_{i=1}^{n-1} H((x_i)) = 0.569 \)

\( 0.569 < 0.72193 \)

Correlated sources are easier to code.
3) (a) $I(x; y | z) = 0$

False

Suppose $H(x) > 0$, $y = x$, and $z = 0$ (always)

Then $I(x; y | z) = I(x; y) = H(x) > 0$.

(b) $H(x | z) = H(x)$

False

Let $z = y = x$ and $H(x) > 0$

Then $H(x | z) = 0 \neq H(x)$

(c) $H(x | y, z) = H(x | y)$

True

$0 = I(x; z | y) = H(x | y) - H(x | y, z)$

(d) $I(x; z) = H(y)$

True

$\sqrt{\text{Data Proc. Theorem}}$

$\sqrt{\text{Data Proc. Theorem}}$

$I(x; z) \leq I(x; y) = H(y) - H(y | x) \leq H(y)$
(e) \( H(Y|Z) \leq H(X|Z) \)

False

Suppose \( Y \) is independent of \( X \) and \( Z \) is independent of \( Y \) and \( H(Y) > H(X) \).

Then \( H(Y|Z) = H(Y) > H(X) = H(X|Z) \)

(f) \( I(X; (Y, Z)) = I(X; Y) \)

True

\[
I(X; (Y, Z)) = H(X) - H(X|Y, Z) \\
= H(X) - H(X|Y) \quad \text{(from (e))} \\
= I(X; Y)
\]
4) Rough Idea:

Getting the rough idea is pretty easy:

since \( A_n \) contains essentially all of the probability
as \( N \to \infty \).

\[
A_n^{(x)} \subseteq B_n \quad \text{(roughly)}
\]

and thus

\[
|A_n^{(x)}| \leq |B_n| \leq 2^{6n+10^w}
\]

\[
\Rightarrow \quad 2^{n(H(x)-c)} \leq 2^{n(6x+10)}
\]

Since \( c \) is arbitrary, \( H(x) \leq 10 \).

Rigorous:

Suppose \( H(x) > 10 \). Now consider sequences \( B_n \) that are also in \( A_n^{(x)} \) for any \( x > 0 \).

\[
P(B_n \cap A_n^{(x)}) \leq 2^{n(6x+10)} \cdot 2^{-n(H(x)-c)}
\]

\[
= 2^{n(7x+10-H(x))}
\]

\[
< \delta_1, \quad N \text{ large} \quad (because \ 10-H(x) \to 0)
\]

But \( P(A_n^{(x)}) \geq 1-\delta_2 \) for any \( x > 0 \), \( N \text{ large} \)

and thus

\[
P(B_n) = P(B_n \cap A_n^{(x)}) + P(B_n \cap A_n^{(x)})
\]

\[
< \delta_1 + \delta_2
\]

Since \( \delta_1, \delta_2 \) are arbitrary, \( P(B_n) \neq 1 \). Contradiction.
5) (a) Use construction from class. Pick any 
\( R > H(x) \) and \( S > 0 \). Pick \( \delta = \frac{R - H(x)}{2} \).

**Method**

1. Encode sequences in \( A_c^{(n)} \) with less than \( \lceil N(H(x) + \delta) \rceil \) bits. (Possible by AEP \#2 part 3)

2. Encode sequences in \( A_c^{(n)} \) with \( (0, 0, \ldots, 0) \).

Thus, code length \( \leq N(H(x) + \delta) + 1 \)

\[
\text{rate} \leq H(X) + \delta + \frac{1}{N} < R \quad \text{for} \quad N \geq N_1 = \frac{2}{R - H(X)}
\]

and, by the AEP \#3 \( N_2 \) s.t. \( \forall N \geq N_2 \),

\[
P_e \leq \mathbb{P}(A_c^{(n)}) < \delta.
\]

Thus, choose \( N = \max(N_1, N_2) \) and we have the code.

(b) Suppose \( R < H(x) \) and pick any \( \delta > 0 \).

Lower bound: consider only errors in the typical set, which has probability

\[
\mathbb{P}(A_c^{(n)}) > 1 - \delta \quad \text{for} \quad N \text{ large enough.}
\]
We can only code $2^{n^2}$ typical sequences each of maximal probability $2^{-n(H(x)-\varepsilon)}$.

Thus,

$$p_e > 1 - \varepsilon - 2^{n^2} 2^{-n(H(x)-\varepsilon)}$$

$$= 1 - \varepsilon - 2^{n(R-H(x)+\varepsilon)}$$

(\text{Choose } \varepsilon = \frac{H(x)-R}{2}. \text{ For any } s > 0, \exists N_1, s.t.

$$\forall N \geq N_1,$$

$$p_e > 1 - \frac{s}{2} - 2^{n(R-H(x)+\varepsilon)}$$

and \exists N_2, s.t.

$$2^{n(R-H(x)+\varepsilon)} < \frac{s}{2}$$

\implies \quad p_e > 1 - s \quad \text{for } \forall s > 0 \text{ and } N \text{ large enough}

\implies \lim_{p \to 0} p_e = 1.
(a) From class, we know this is simply
given by
\[ H(x) = -\sum_{x \in X} p_x(x) \log_2 p_x(x) \]
\[ = 1.75 \]

(b) We need to encode the entire typical set of size
\[ 2^{N H(x)} = 2^{1.75N} \]

How many sequences of length \( k \) have \( \leq k/4 \) 1's? Assume we have an i.i.d. binary source \( (Y_i) \) with \( P(Y_i = 0) = \frac{3}{4}, P(Y_i = 1) = \frac{1}{4} \).
Let \( A^{(k)}_1 \) : set of sequences with \( \leq k/4 \) 1's.
\[ 1 \geq P(A^{(k)}_1) = \sum_{x \in A^{(k)}_1} p_x(x) \]
\[ \geq |A^{(k)}_1| \cdot 2^{-k \cdot H_2(1/4)} \]
(since all sequences in \( A^{(k)}_1 \) have higher probability than those with approximately \( k/4 \) 1's).
\[ \Rightarrow |A^{(k)}_1| \leq 2^{k H_2(1/4)} \]

But \( |A^{(k)}_1| \geq 2^{k H_2(1/4)} \) (it includes the typical set).
\[ \Rightarrow |A^{(k)}_1| \geq 2^{k H_2(1/4)} \]

We must have \( 2^{k H_2(1/4)} \geq 2^{1.75N} \)
\[ \Rightarrow k/ \cdot \frac{1.75}{H_2(1/4)} = 1.75/0.81 \geq 2.16. \]
Let $C =$ capacity of the cascade.

For any $p_x(x)$,

$$I(x; \tilde{y}) \geq I(x; y) \quad \text{(data processing inequality)}$$

$$\Rightarrow \max_{p_x(x)} I(x; \tilde{y}) \geq \max_{p_y(y)} I(x; y)$$

$$\Rightarrow C \leq C_1$$

Suppose $C > C_2$, then there exists $p_x(x)$ (with associated $p_y(y)$) such that

$$C = I_p(x; \tilde{y}) > C_2 = \max_{p_y(y)} I(y; \tilde{y})$$

But, by the data processing inequality,

$$I_p(y; \tilde{y}) \geq I_p(x; \tilde{y})$$

$$\Rightarrow I_p(y; \tilde{y}) > \max_{p_y(y)} I(y; \tilde{y})$$

and thus our supposition is false $\Rightarrow C \leq C_2$.

$$\Rightarrow C \leq \min(C_1, C_2)$$
8) The key here is rigor. Since the channel is a one, I know:

\[ C = \max \ I(x; y) \]

\[ I(x; y) = H(y) - H(y|x) \]

\[ \leq 1 - E_x \left[ H(y|x = x) \right] \quad \text{(since } y \text{ is binary)} \]

\[ = 1 - \left[ p_x(0) \mathcal{H}_2(p) + p_x(1) + p_x(2) \mathcal{H}_2(p) \right] \]

\[ \leq 1 - \mathcal{H}_2(p) \quad \text{(since } \mathcal{H}_2(p) \leq 1 \text{)} \]

\[ \quad \text{and } p_x(0) + p_x(1) + p_x(2) = 1 \]

which can be achieved with

\[ p_x(0) = \frac{1}{2} \]

\[ p_x(1) = 0 \]

\[ p_x(2) = \frac{1}{2} \]

Thus,

\[ C = 1 - \mathcal{H}_2(p) \]
9) (a) \( H(\gamma | X = x) = \mathcal{N}_2(\rho), \ \forall x \)

Thus, 
\( I(X; \gamma) = H(\gamma) - \mathcal{N}_2(\rho) \leq 1 - \mathcal{N}_2(\rho) \)

achieved by \( p(x_k = 0) = \frac{1}{2}, \ p(x_k = 1) = 0, \ p(x_k = 2) = \frac{1}{2} \)

\[ \Rightarrow C = 1 - \mathcal{N}_2(\rho) \]

(b) \( H(\gamma | X = x) = \mathcal{N}_2(\rho), \ \forall x \)

\[ p(\gamma = 0) = \rho \quad (\text{very unlikely}) \]
\[ p(\gamma = 1) = (1 - \rho) \]

\[ \Rightarrow C = \mathcal{N}_2(\rho) - \mathcal{N}_2(\rho) = 0 \]

(c) \( H(\gamma | X = x) = \log_2 3, \ \forall x \)

\[ \Rightarrow I(X; \gamma) = H(\gamma) - \log_2 3 \]

\[ \leq 2 - \log_2 3 \]

which is achieved when \( p_x(\gamma) = 1/2, \ \forall x \)

\[ \Rightarrow C = 2 - \log_2 3 \]

\[ = 0.415 \]
Choose suboptimal decision rule:

$$x(y) = \text{argmax}_x \rho_x(x)$$

$$p_e = P(x \neq \text{argmax}_x \rho_x(x))$$

$$= 1 - \max_x \rho_x(x)$$

$$\leq 1 - \frac{1}{|x|}$$

$$= \frac{|x| - 1}{|x|}$$

The optimal decision rule must be at least as good.

(b)

$$\mathcal{H}_{1|x_1}(\rho) = \mathcal{H}_2(\rho) + \rho \log_2(|x| - 1)$$

$$\frac{d}{d\rho} \mathcal{H}_{1|x_1}(\rho) = -\log_2 \frac{\rho}{1-\rho} + \log_2 (|x| - 1) = 0$$

$$\Rightarrow \frac{\rho}{1-\rho} = |x| - 1$$

$$\Rightarrow \rho = \frac{|x| - 1}{|x|}$$

(second derivative is indeed negative also)
(c) I will give the rule,
\[ I(X, Y) = H(x) \]
\[ \Rightarrow H(x | y) = 0 \]
\[ \Rightarrow H(x | y = \gamma) = 0 \quad \forall \gamma \]
\[ \Rightarrow \exists \bar{x}(\gamma) \text{ s.t. } p_{x | y}(\bar{x}(\gamma) | y) = 1 \quad \forall \gamma \]
\[ \Rightarrow \text{ choose } \hat{x}(\gamma) = \bar{x}(\gamma) \]
\[ \Rightarrow p_e = 0. \]