1. Find the channel capacity of the following discrete memoryless channel:

\[ Z \]

\[ X \rightarrow Y \]

where \( \Pr\{Z = 0\} = \Pr\{Z = a\} = \frac{1}{2} \). The alphabet for \( x \) is \( X = \{0, 1\} \). Assume that \( Z \) is independent of \( X \). Observe that the channel capacity depends on the value of \( a \).

**Solution :**

\[
Y = X + Z \quad X \in \{0, 1\}, \quad Z \in \{0, a\}
\]

We have to distinguish various cases depending on the values of \( a \).

- \( a = 0 \). In this case, \( Y = X \), and \( \max I(X; Y) = \max H(X) = 1 \). Hence the capacity is 1 bit per transmission.
- \( a \neq 0, \pm 1 \). In this case, \( Y \) has four possible values 0, 1, \( a \), and 1 + \( a \). Knowing \( Y \), we know the \( X \) which was sent, and hence \( H(X|Y) = 0 \). Hence \( \max I(X; Y) = \max H(X) = 1 \), achieved for an uniform distribution on the input \( X \).
- \( a = 1 \). In this case, \( Y \) has three possible output values, 0, 1, and 2. The channel is identical to the binary erasure channel with \( a = 1/2 \). The capacity of this channel is 1/2 bit per transmission.
- \( a = -1 \). This is similar to the case when \( a = 1 \) and the capacity here is also 1/2 bit per transmission.

2. Consider a 26-key typewriter.

(a) If pushing a key results in printing the associated letter, what is the capacity \( C \) in bits?

(b) Now suppose that pushing a key results in printing that letter or the next (with equal probability). Thus \( A \rightarrow A \) or \( B \), \( \cdots \), \( Z \rightarrow Z \) or \( A \). What is the capacity?

(c) What is the highest rate code with block length one that you can find that achieves zero probability of error for the channel in part (b).

**Solution :**

(a) If the typewriter prints out whatever key is struck, then the output \( Y \), is the same as the input \( X \), and

\[
C = \max I(X; Y) = \max H(X) = \log 26,
\]

attained by a uniform distribution over the letters.
(b) In this case, the output is either equal to the input (with probability $\frac{1}{2}$) or equal to the next letter (with probability $\frac{1}{2}$). Hence $H(Y|X) = \log 2$ independent of the distribution of $X$, and hence

$$C = \max I(X;Y) = \max H(Y) - \log 2 = \log 26 - \log 2 = \log 13$$

attained for a uniform distribution over the output, which in turn is attained by a uniform distribution on the input.

(c) A simple zero error block length one code is the one that uses every alternate letter, say $A, C, E, \cdots, W, Y$. In this case, none of the codewords will be confused, since $A$ will produce either $A$ or $B$, $C$ will produce $C$ or $D$, etc. The rate of this code,

$$R = \frac{\log(\# \text{ codewords})}{\text{Block length}} = \frac{\log 13}{1} = \log 13$$

In this case, we can achieve capacity with a simple code with zero error.

3. Consider a binary symmetric channel with $Y_i = X_i \oplus Z_i$, where $\oplus$ is mod 2 addition, and $X_i, Y_i \in \{0, 1\}$. Suppose that $\{Z_i\}$ has constant marginal probabilities $p(Z_i = 1) = p = 1 - p(Z_i = 0)$, but that $Z_1, Z_2, \cdots, Z_n$ are not necessarily independent. Let $C = 1 - H(p)$. Show that

$$\max_{p(x_1, x_2, \cdots, x_n)} I(X_1, X_2, \cdots, X_n; Y_1, Y_2, \cdots, Y_n) \geq nC$$

Comment on the implications.

**Solution:**

When $X_1, X_2, \cdots, X_n$ are chosen i.i.d. $\sim$ Bern($\frac{1}{2}$),

$$I(X_1, \cdots, X_n; Y_1, \cdots, Y_n) = H(X_1, \cdots, X_n) - H(X_1, \cdots, X_n|Y_1, \cdots, Y_n)$$

$$= H(X_1, \cdots, X_n) - H(Z_1, \cdots, Z_n|Y_1, \cdots, Y_n)$$

$$\geq H(X_1, \cdots, X_n) - H(Z_1, \cdots, Z_n)$$

$$\geq H(X_1, \cdots, X_n) - \sum H(Z_i)$$

$$= n - nH(p)$$

Hence, the capacity of the channel with memory over $n$ uses of the channel is

$$nC^{(n)} = \max_{p(x_1, \cdots, x_n)} I(X_1, \cdots, X_n; Y_1, \cdots, Y_n)$$

$$\geq I(X_1, \cdots, X_n; Y_1, \cdots, Y_n)_{p(x_1, \cdots, x_n) = \text{Bern}(\frac{1}{2})}$$

$$\geq n(1 - H(p))$$

$$= nC$$

Hence, channels with memory have higher capacity. The intuitive explanation for this result is that the correlation between the noise decreases the effective noise; one could use the information from the past samples of the noise to combat the present noise.

4. Consider the channel $Y = X + Z$ (mod 13), where

$$Z = \begin{cases} 
1, & \text{with probability } \frac{1}{3} \\
2, & \text{with probability } \frac{1}{3} \\
3, & \text{with probability } \frac{1}{3}
\end{cases}$$

and $X \in \{0, 1, \cdots, 12\}$. 


(a) Find the capacity.
(b) What is the maximizing \( p^*(x) \)?

**Solution:**

(a)

\[
C = \max_{p(x)} I(X; Y) \\
= \max_{p(x)} H(Y) - H(Y|X) \\
= \max_{p(x)} H(Y) - \log 3 \\
= \log 13 - \log 3 \\
= \log \frac{13}{3}
\]

which is attained when \( Y \) has a uniform distribution, which occurs (by symmetry) when \( X \) has a uniform distribution.

(b) The capacity is achieved by a uniform distribution on the inputs, that is,

\[
p(X = i) = \frac{1}{13} \quad \text{for } i = 0, 1, \cdots, 12.
\]

5. Using two channels

(a) Consider two discrete memoryless channels \((\mathcal{X}_1, p(y_1|x_1), \mathcal{Y}_1)\) and \((\mathcal{X}_2, p(y_2|x_2), \mathcal{Y}_2)\) with capacities \( C_1 \) and \( C_2 \) respectively. A new channel \((\mathcal{X}_1 \times \mathcal{X}_2, p(y_1|x_1) \times p(y_2|x_2), \mathcal{Y}_1 \times \mathcal{Y}_2)\) is formed in which \( x_1 \in \mathcal{X}_1 \) and \( x_2 \in \mathcal{X}_2 \), are *simultaneously* sent, resulting in \( y_1, y_2 \). Find the capacity of this channel.

(b) Find the capacity \( C \) of the union 2 channels \((\mathcal{X}_1, p(y_1|x_1), \mathcal{Y}_1)\) and \((\mathcal{X}_2, p(y_2|x_2), \mathcal{Y}_2)\) where, at each time, one can send a symbol over channel 1 or channel 2 but not both. Assume the output alphabets are distinct and do not intersect. Show \( 2^C = 2^{C_1} + 2^{C_2} \).

**Solution:**

(a) To find the capacity of the product channel \((\mathcal{X}_1 \times \mathcal{X}_2, p(y_1, y_2|x_1, x_2), \mathcal{Y}_1 \times \mathcal{Y}_2)\), we have to 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 transition probabilities are given as \( p(y_1, y_2|x_1, x_2) = p(y_1|x_1)p(y_2|x_2) \),

\[
p(x_1, x_2, y_1, y_2) = p(x_1, x_2)p(y_1, y_2|x_1, x_2) \\
= p(x_1, x_2)p(y_1|x_1)p(y_2|x_2)
\]

Therefore, \( Y_1 \rightarrow X_1 \rightarrow X_2 \rightarrow Y_2 \) forms a Markov chain and

\[
I(X_1, X_2; Y_1, Y_2) = H(Y_1, Y_2) - H(Y_1|X_1, X_2) \\
= H(Y_1, Y_2) - H(Y_1|X_1, X_2) - H(Y_2|X_1, X_2) \\
= H(Y_1, Y_2) - H(Y_1|X_1) - H(Y_2|X_2) \\
\leq H(Y_1) + H(Y_2) - H(Y_1|X_1) - H(Y_2|X_2) \\
= I(X_1; Y_1) + I(X_2; Y_2)
\]
where (1) and (2) follow from Markovity and (3) is met with equality of $X_1$ and $X_2$ are independent and hence $Y_1$ and $Y_2$ are independent. Therefore

$$C = \max_{p(x_1, x_2)} I(X_1, X_2; Y_1, Y_2)$$

$$\leq \max_{p(x_1, x_2)} I(X_1; Y_1) + \max_{p(x_1, x_2)} I(X_2; Y_2)$$

$$= \max_{p(x_1)} I(X_1; Y_1) + \max_{p(x_2)} I(X_2; Y_2)$$

$$= C_1 + C_2$$

with equality iff $p(x_1, x_2) = p^*(x_1)p^*(x_2)$ and $p^*(x_1)$ and $p^*(x_2)$ are the distributions that maximize $C_1$ and $C_2$ respectively.

(b) Let

$$\begin{cases} 
1, & \text{if the signal is sent over the channel 1} \\
2, & \text{if the signal is sent over the channel 2} 
\end{cases}$$

Consider the following communication scheme: The sender chooses between two channels according to Bern($\alpha$) coin flip. Then the channel input is $X = (\theta, X_\theta)$.
Since the output alphabets $Y_1$ and $Y_2$ are disjoint, $\theta$ is a function of $Y$, i.e. $X \rightarrow Y \rightarrow \theta$.
Therefore,

$$I(X; Y) = I(X; Y, \theta)$$

$$= I(X_\theta, \theta; Y, \theta)$$

$$= I(\theta; Y, \theta) + I(X_\theta; Y|\theta)$$

$$= I(\theta; Y, \theta) + I(X_\theta; Y|\theta)$$

$$= H(\theta) + \alpha I(X_\theta; Y|\theta = 1) + (1 - \alpha) I(X_\theta; Y|\theta = 2)$$

$$= H(\alpha) + \alpha I(X_1; Y_1) + (1 - \alpha) I(X_2; Y_2)$$

Thus, it follows that

$$C = \sup_{\alpha} \{H(\alpha) + \alpha C_1 + (1 - \alpha)C_2\}$$

which is a strictly concave function on $\alpha$. Hence, the maximum exists and by elementary calculus, one can easily show $C = \log_2(2^{C_1} + 2^{C_2})$, which is attained with $\alpha = 2^{C_1}/(2^{C_1} + 2^{C_2})$.

6. Consider a time-varying discrete memoryless binary symmetric channel. Let $Y_1, Y_2, \ldots, Y_n$ be conditionally independent given $X_1, X_2, \ldots, X_n$, with conditional distribution given by $p(y^n|x^n) = \prod_{i=1}^n p_i(y_i|x_i)$, as shown below.

(a) Find $\max_{p(x)} I(X^n; Y^n)$.
(b) We now ask for the capacity for the time invariant version of this problem. Replace each $p_i, 1 \leq i \leq n$, by the average value $\bar{p} = \frac{1}{n} \sum_{j=1}^{n} p_j$, and compare the capacity to part (a).

Solution:

(a)

$$I(X^n;Y^n) = H(Y^n) - \sum_{i=1}^{n} H(Y_i|X_i)$$

$$\leq \sum_{i=1}^{n} H(Y_i) - \sum_{i=1}^{n} H(Y_i|X_i)$$

$$\leq \sum_{i=1}^{n} (1 - H(p_i))$$

with equality if $X_1, \cdots, X_n$ are chosen i.i.d. $\sim$ Bern($\frac{1}{2}$). Hence

$$\max_{p(x)} I(X_1, \cdots, X_n; Y_1, \cdots, Y_n) = \sum_{i=1}^{n} (1 - H(p_i))$$

(b) Since $H(p)$ is concave on $p$, by Jensen’s inequality,

$$\frac{1}{n} \sum_{i=1}^{n} H(p_i) \leq H\left(\frac{1}{n} \sum_{i=1}^{n} p_i\right) = H(\bar{p})$$

i.e.,

$$\sum_{i=1}^{n} H(p_i) \leq nH(\bar{p})$$

Hence,

$$C_{\text{time-varying}} = \sum_{i=1}^{n} (1 - H(p_i))$$

$$= n - \sum_{i=1}^{n} H(p_i)$$

$$\geq n - nH(\bar{p})$$

$$= \sum_{i=1}^{n} (1 - H(\bar{p}))$$

$$= C_{\text{time invariant}}$$

7. Suppose a binary symmetric channel of capacity $C_1$ is immediately followed by a binary erasure channel of capacity $C_2$. Find capacity $C$ of the resulting channel.
Now consider an arbitrary discrete memoryless channel \((X, p(y|x), Y)\) followed by a binary erasure channel, resulting in an output
\[
\tilde{Y} = \begin{cases} Y, & \text{with probability } 1 - \alpha \\ e, & \text{with probability } \alpha \end{cases}
\]
where \(e\) denotes erasure. Thus the output \(Y\) is erased with probability \(\alpha\). What is the capacity of this channel?

Solution:

(a) Let \(C_1 = 1 - H(p)\) be the capacity of the BSC with parameter \(p\), and \(C_2 = 1 - \alpha\) be the capacity of the BEC with parameter \(\alpha\). Let \(\tilde{Y}\) denote the output of the cascaded channel, and \(Y\) the output of the BSC. Then, the transition rule for the cascaded channel is simply
\[
p(\tilde{y}|x) = \sum_{y=0,1} p(\tilde{y}|y)p(y|x)
\]
for each \((x, \tilde{y})\) pair.

Let \(X \sim \text{Bern}(\pi)\) denote the input to the channel. Then,
\[
H(\tilde{Y}) = H(((1 - \alpha)\pi(1 - p) + p(1 - \pi)), \alpha, (1 - \alpha)(p\pi + (1 - p)(1 - \pi)))
\]
and also
\[
H(\tilde{Y}|X = 0) = H((1 - \alpha)(1 - p), \alpha, (1 - \alpha)p)
\]
\[
H(\tilde{Y}|X = 1) = H((1 - \alpha)p, \alpha, (1 - \alpha)(1 - p)) = H(\tilde{Y}|X = 0)
\]

Therefore,
\[
C = \max_{p(x)} I(X; \tilde{Y})
\]
\[
= \max_{p(x)} \{H(\tilde{Y}) - H(\tilde{Y}|X)\}
\]
\[
= \max_{p(x)} \{H(\tilde{Y})\} - H(\tilde{Y}|X)
\]
\[
= \max_{p(x)} \{H(((1 - \alpha)\pi(1 - p) + p(1 - \pi)), \alpha, (1 - \alpha)(p\pi + (1 - p)(1 - \pi)))\}
\]
\[
- H((1 - \alpha)(1 - p), \alpha, (1 - \alpha)p)
\]
Note that the maximum value of $H(\tilde{Y})$ occurs when $\pi = 1/2$ by the concavity and symmetry of $H(\cdot)$. (We can check this also by differentiating (4) with respect to $\pi$.) Substituting the value $\pi = 1/2$ in the expression for the capacity yields

$$C = H((1 - \alpha)/2, \alpha, (1 - \alpha)/2) - H((1 - p)(1 - \alpha), \alpha, p(1 - \alpha))$$

$$= (1 - \alpha)(1 + p \log p + (1 - p) \log(1 - p))$$

$= C_1 C_2$

(b) For the cascade of an arbitrary discrete memoryless channel (with capacity $C$) with the erasure channel (with the erasure probability $\alpha$), we will show that

$$I(X; \tilde{Y}) = (1 - \alpha)I(X; Y)$$

(5)

Then, by taking suprema of both sides over all input distributions $p(x)$, we can conclude the capacity of the cascaded channel is $(1 - \alpha)C$.

Proof of (5):

Let

$$E = \begin{cases} 1, & \tilde{Y} = e \\ 0, & \tilde{Y} = Y \end{cases}$$

Then, since $E$ is a function of $Y$,

$$H(\tilde{Y}) = H(\tilde{Y}, E)$$

$$= H(E) + H(\tilde{Y}|E)$$

$$= H(\alpha) + \alpha H(\tilde{Y}|E = 1) + (1 - \alpha)H(\tilde{Y}|E = 0)$$

$$= H(\alpha) + (1 - \alpha)H(Y),$$

where the last equality comes directly from the construction of $E$. Similarly,

$$H(\tilde{Y}|X) = H(\tilde{Y}, E|X)$$

$$= H(E|X) + H(\tilde{Y}|X, E)$$

$$= H(E) + \alpha H(\tilde{Y}|X, E = 1) + (1 - \alpha)H(\tilde{Y}|X, E = 0)$$

$$= H(\alpha) + (1 - \alpha)H(Y|X),$$

whence

$$I(X; \tilde{Y}) = H(\tilde{Y}) - H(\tilde{Y}|X) = (1 - \alpha)I(X; Y)$$

8. We wish to encode a Bernoulli($\alpha$) process $V_1, V_2, \cdots$ for transmission over a binary symmetric channel with error probability $p$.

Find conditions on $\alpha$ and $p$ so that the probability of error $p(\hat{V}^n \neq V^n)$ can be made to go to zero as $n \to \infty$.

Solution :

Suppose we want to send a binary i.i.d. Bern($\alpha$) source over a binary symmetric channel with error
probability $p$. By the source-channel separation theorem, in order to achieve the probability of error that vanishes asymptotically, i.e. $P(V^n \neq V^n) \to 0$, we need the entropy of the source to be less than the capacity of the channel. Hence,

$$H(\alpha) + H(p) < 1,$$

or, equivalently,

$$\alpha^\alpha (1 - \alpha)^{1 - \alpha} p^p (1 - p)^{1-p} < \frac{1}{2}.$$

9. Let $(X_i, Y_i, Z_i)$ be i.i.d. according to $p(x, y, z)$. We will say that $(x^n, y^n, z^n)$ is jointly typical [written $(x^n, y^n, z^n) \in A_\epsilon^{(n)}$] if

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

Now suppose that $(\hat{X}^n, \hat{Y}^n, \hat{Z}^n)$ is drawn according to $p(x^n)p(y^n)p(z^n)$. Thus, $\hat{X}^n$, $\hat{Y}^n$, $\hat{Z}^n$ have the same marginals as $p(x^n, y^n, z^n)$ but are independent. Find (bounds on) Pr{$(\hat{X}^n, \hat{Y}^n, \hat{Z}^n) \in A_\epsilon^{(n)}$} in terms of the entropies $H(X)$, $H(Y)$, $H(Z)$, $H(X,Y)$, $H(X,Z)$, $H(Y,Z)$ and $H(X,Y,Z)$.

Solution:

$$\Pr\{(\hat{X}^n, \hat{Y}^n, \hat{Z}^n) \in A_\epsilon^{(n)}\} = \sum_{(x^n, y^n, z^n) \in A_\epsilon^{(n)}} p(x^n)p(y^n)p(z^n)$$

$$\leq \sum_{(x^n, y^n, z^n) \in A_\epsilon^{(n)}} 2^{-n(H(X)+H(Y)+H(Z)-3\epsilon)}$$

$$\leq |A_\epsilon^{(n)}| 2^{-n(H(X)+H(Y)+H(Z)-3\epsilon)}$$

$$\leq 2^{2n(H(X,Y,Z)+\epsilon)} 2^{-n(H(X)+H(Y)+H(Z)-3\epsilon)}$$

$$\leq 2^{n(H(X,Y,Z)-H(X)-H(Y)-H(Z)+4\epsilon)}$$

Also,

$$\Pr\{(\hat{X}^n, \hat{Y}^n, \hat{Z}^n) \in A_\epsilon^{(n)}\} = \sum_{(x^n, y^n, z^n) \in A_\epsilon^{(n)}} p(x^n)p(y^n)p(z^n)$$

$$\geq \sum_{(x^n, y^n, z^n) \in A_\epsilon^{(n)}} 2^{-n(H(X)+H(Y)+H(Z)+3\epsilon)}$$

$$\geq |A_\epsilon^{(n)}| 2^{-n(H(X)+H(Y)+H(Z)+3\epsilon)}$$

$$\geq (1 - \epsilon) 2^{n(H(X,Y,Z)-\epsilon)} 2^{-n(H(X)+H(Y)+H(Z)-3\epsilon)}$$

$$\geq (1 - \epsilon) 2^{n(H(X,Y,Z)-H(X)-H(Y)-H(Z)-4\epsilon)}$$

Note that the upper bound is true for all $n$, but the lower bound only hold for $n$ large.
10. Twenty questions.

(a) Player A chooses some object in the universe, and player B attempts to identify the object with a series of yes-no questions. Suppose that player B is clever enough to use the code achieving the minimal expected length with respect to player A’s distribution. We observe that player B requires an average 38.5 questions to determine the object. Find a rough lower bound to the number of objects in the universe.

(b) Let $X$ be uniformly distributed over $\{1, 2, \cdots, m\}$. Assume that $m = 2^n$. We ask random questions: Is $X \in S_1$? Is $X \in S_2$? \cdots until only one integer remains. All $2^m$ subsets $S$ of $\{1, 2, \cdots m\}$ are equally likely.

i. How many deterministic questions are needed to determine $X$?

ii. Without loss of generality, suppose that $X = 1$ is the random object. What is the probability that object 2 yields the same answers as object 1 for $k$ questions?

iii. What is the expected number of objects in $\{2, 3, \cdots, m\}$ that have the same answers to the questions as those of the correct object 1?

Solution:

(a) 

$$37.5 = L^* - 1 < H(X) \leq \log |X|$$

and hence number of objects in the universe $> 2^{37.5} = 1.94 \times 10^{11}$.

(b) i. Obviously, Huffman codewords for $X$ are all of length $n$. Hence, with $n$ deterministic questions, we can identify an object out of $2^n$ candidates.

ii. Observe that the total number of subsets which include both object 1 and object 2 or neither of them is $2^{m-1}$. Hence, the probability that object 2 yields the same answers for $k$ questions as object 1 is $(2^{m-1}/2^m)^k = 2^{-k}$.

iii. Let

$$1_j = \begin{cases} 1, & \text{object } j \text{ yields the same answers for } k \text{ questions as object 1} \\ 0, & \text{otherwise.} \end{cases}$$

for $j = 2, \cdots, m$

Then

$$E[N] = E\left[ \sum_{j=2}^{m} 1_j \right]$$

$$= \sum_{j=2}^{m} E[1_j]$$

$$= \sum_{j=2}^{m} 2^{-k}$$

$$= (m - 1)2^{-k}$$

$$= (2^n - 1)2^{-k}$$

where $N$ is the number of objects in $\{2, 3, \cdots, m\}$ with the same answers.