Problem 1:  **Modulo Channel**

1. Consider the DMC defined as follows: Output $Y = X \oplus_2 Z$ where $X$, taking values in $\{0, 1\}$, is the channel input, $\oplus_2$ is the modulo-2 summation operation, and $Z$ is binary channel noise uniform over $\{0, 1\}$ and independent of $X$. What is the capacity of this channel?

   **Solution:**
   The channel is a BSC with crossover probability $1/2$, so the capacity is zero.

2. Consider the channel of the previous part, but suppose that instead of modulo-2 addition $Y = X \oplus_2 Z$, we perform modulo-3 addition $Y = X \oplus_3 Z$. Now what is the capacity?

   **Solution:**
   Note that now $Y \in \{0, 1, 2\}$. Since $Z$ is Bernoulli($1/2$), the channel output being 1 indicates that $X$ could be 0 or 1 with equal probability. Thus, it gives no information about the input. This simply becomes a Binary Erasure channel with erasure probability $1/2$, and the capacity is $1 - 1/2 = 1/2$.

3. Now suppose the noise $Z$ is no longer independent of the input $X$, but is instead described by the following conditional distribution:

   
   
   $p(Z = z|X = 0) = \begin{cases} 
   1/4 & \text{if } z = 0 \\
   3/4 & \text{if } z = 1,
   \end{cases}$

   and

   $p(Z = z|X = 1) = 1/2 \quad \text{both for } z = 0 \text{ and } z = 1.$

   A random code of size $2^{nR}$ is generated uniformly (that is all codewords are drawn i.i.d. $X \sim \text{Bern}(0.5)$). Find the value $V$ such that if $R < V$ then the average probability of decoding error (average both across the messages and the randomness in the codebook) vanishes with increasing blocklength while if $R > V$ then it does not.

   **Solution:**
   In this problem, several students wrongly interpreted $V$ to be the capacity of the channel. This is incorrect, as the problem gives a random codebook generation according to i.i.d. Bern(0.5) codewords. Recall from the direct and converse theorems in the lecture that the maximum supported rate under a given random codebook is simply the Mutual Information between $X$ and $Y$.

   If $X$ is uniform and the noise $Z$ is distributed as described, one may determine the information between $X$ and $Y$ as follows:

   
   \[ I(X; Y) = H(Y) - H(Y|X) = H(Y) - H(Z|X). \]
For \( X \) chosen Bernoulli(1/2), we have that

\[
H(Z|X) = \frac{1}{2} h_b \left( \frac{1}{4} \right) + \frac{1}{2}.
\]

To compute \( H(Y) \), we first observe that the distribution of \( Y \) when \( X \) is Bernoulli(1/2) is given by \((p_0, p_1, p_2) = (\frac{1}{8}, \frac{5}{8}, \frac{1}{4})\). The entropy of the this distribution may then be calculated.

Performing both computations, we find that \( I(X;Y) = H(Y) - H(Z|X) = 0.393 \). This is the maximum supported rate by the channel under a uniformly generated codebook. Note that this is not the capacity of the channel which is achieved for a different input distribution.
Problem 2: Deletion Channel

20 points

Consider a binary sequence of length \(n\), denoted by \(X^n = (X_1, \ldots, X_n)\). Consider another binary sequence of length \(n\) called deletion pattern, denoted by \(D^n = (D_1, \ldots, D_n)\), which determines how \(X_n\) is to be deleted. Then, the output of the deletion process, denoted by \(y(X^n, D^n)\), is derived from \(X^n\) by deleting the bits at those locations where the deletion pattern is 1.

Consider the following example for \(n = 10\):

\[
X^n = (0, 1, 1, 0, 0, 1, 1, 0)
\]
\[
D^n = (0, 1, 0, 1, 1, 0, 0, 1, 0)
\]
\[
y(X^n, D^n) = (0, 1, 0, 1, 1, 0)
\]

The source sequence \(X^n \in \{0, 1\}^n\) is i.i.d. Bernoulli(1/2), and the deletion pattern \(D^n \in \{0, 1\}^n\) is i.i.d. Bernoulli(\(d\)), independent of \(X^n\).

1. We are interested in computing the mutual information between \(X^n\) and \(y(X^n, D^n)\), 
\(I(X^n; y(X^n, D^n))\). Which of these relations is true?

(a) \(I(X^n; y(X^n, D^n)) = H(y(X^n, D^n)) + H(D^n) - H(y(X^n, D^n)|X^n, D^n)\)

(b) \(I(X^n; y(X^n, D^n)) = H(y(X^n, D^n)) - H(D^n) + H(D^n|X^n, y(X^n, D^n))\)

(c) \(I(X^n; y(X^n, D^n)) = H(y(X^n, D^n)) - H(D^n)\)

(d) \(I(X^n; y(X^n, D^n)) = H(X^n) - H(X^n|y(X^n, D^n)) - H(D^n|X^n, y(X^n, D^n))\)

Solution: (b)

\[
I(y(X^n, D^n); X^n, D^n) = H(y(X^n, D^n)) - H(y(X^n, D^n)|X^n, D^n)
\]
\[
= H(y(X^n, D^n))
\]
\[
= H(X^n, D^n) - H(X^n, D^n|y(X^n, D^n))
\]
\[
= H(X^n) + H(D^n) - H(X^n|y(X^n, D^n)) - H(D^n|X^n, y(X^n, D^n))
\]
\[
= I(X^n; y(X^n, D^n)) + H(D^n) - H(D^n|X^n, y(X^n, D^n))
\]

Rearranging the above terms, we get the desired result

\[
I(X^n; y(X^n, D^n)) = H(y(X^n, D^n)) - H(D^n) + H(D^n|X^n, y(X^n, D^n))
\]

2. Now consider the case where rather than an i.i.d. process, the deletion pattern is a stationary time-homogenous binary Markov chain, independent of \(X^n\). Justify the following steps assuming \(1 \leq m < n\) and the notation \(X^m_m \equiv (X_m, \ldots, X_n)\):


\[
H(X^n|y(X^n, D^n), D_0, D_{n+1})
\]

\[
\geq H(X^n|y(X^n, D^n), y(X^m_{m+1}, D_{m+1}^n), D_0, D_{n+1})
\]

\[
\geq H(X^m|y(X^m, D^m), y(X^m_{m+1}, D^m_{m+1}), D_0, D_{n+1}, D_{m+1})
\]

\[
+ H(X^m_{m+1}|y(X^m, D^m), y(X^m_{m+1}, D^m_{m+1}), D_0, D_{n+1}, D_m)
\]

\[
= H(X^m|y(X^m, D^m), y(X^m_{m+1}, D^m_{m+1}), D_0, D_{n+1}, D_m)
\]

\[
H(X^n|y(X^n, D^n), y(X^m_{m+1}, D^m_{m+1}), D_0, D_{n+1}, D_m - X^m)
\]

\[
H(X^n|y(X^n, D^n), y(X^m_{m+1}, D^m_{m+1}), D_0, D_{n+1})
\]

\[
= H(X^n|y(X^n, D^n), y(X^m_{m+1}, D^m_{m+1}), D_0, D_{n+1})
\]

\[
+ H(X^m_{m+1}|y(X^m, D^m), y(X^m_{m+1}, D^m_{m+1}), D_0, D_{n+1}, D_m)
\]

\[
= H(X^m_{m+1}|y(X^m, D^m), y(X^m_{m+1}, D^m_{m+1}), D_0, D_{n+1}, D_m)
\]

(c) Markov chains

\[
(y(X^m_{m+1}, D^m_{m+1}), D_{n+1}) - D_{m+1} - (X^m, y(X^m, D^m), D_0)
\]

and

\[
(y(X^m, D^m), D_0) - D_m - (X^m_{m+1}, y(X^m_{m+1}, D^m_{m+1}), D_{n+1}).
\]

(d) Stationarity.
Problem 3: Exponential Noise Channel and Exponential Source

Recall that $X \sim \text{Exp}(\lambda)$ is to say that $X$ is a continuous non-negative random variable with density

$$f_X(x) = \begin{cases} 
\lambda e^{-\lambda x} & \text{if } x \geq 0 \\
0 & \text{if } x < 0 
\end{cases}$$

(1)

or, equivalently, that $X$ is a random variable with characteristic function

$$\varphi_X(t) \triangleq E[e^{itX}] = \frac{1}{1 - it/\lambda}.$$  

(2)

Recall also that in this case $EX = 1/\lambda$.

1. Find the differential entropy of $X \sim \text{Exp}(\lambda)$.

Solution:

$$h(X) = -\int_{-\infty}^{\infty} f_X(x) \log(f_X(x)) \, dx$$

(3)

$$= -\int_{0}^{\infty} \lambda e^{-\lambda x} \log(\lambda e^{-\lambda x}) \, dx$$

(4)

$$= -\int_{0}^{\infty} \lambda e^{-\lambda x} \log(\lambda) + \lambda \int_{0}^{\infty} xe^{-\lambda x} \, dx$$

(5)

$$= 1 - \log \lambda.$$  

(6)

2. Prove that $\text{Exp}(\lambda)$ uniquely maximizes the differential entropy among all non-negative random variables confined to $EX \leq 1/\lambda$.

Hint: follow our proof of a similar fact for the Gaussian distribution.

Solution:

Let the probability density of any such non-negative random variable is $f_X$, while $g_X$ is the density of $\text{Exp}(\lambda)$ as in Part (1) above,

$$h(X) = -\int_{0}^{\infty} f_X(x) \log(f_X(x)) \, dx$$

(7)

$$= -\int_{0}^{\infty} f_X(x) \log\left(\frac{f_X(x)}{g_X(x)}\right) \, dx - \int_{0}^{\infty} f_X(x) \log(\lambda e^{-\lambda x}) \, dx$$

(8)

$$= -D(f_X||g_X) - \log(\lambda) \int_{0}^{\infty} f_X(x) \, dx + \lambda \int_{0}^{\infty} xf_X(x) \, dx$$

(9)

$$= 1 - \log \lambda - D(f_X||g_X) - \log(\lambda)$$

(10)

$$\leq 1 - \log \lambda,$$  

(11)

where the last inequality is due to the fact that $D(f_X||g_X) \geq 0$, equality holds if $X = \text{Exp}(\lambda)$.
Fix positive scalars $a$ and $b$. Let $\overline{X}$ be the non-negative random variable of mean $a$ formed by taking $\overline{X} = 0$ with probability $\frac{b}{a+b}$ and, with probability $\frac{a}{a+b}$, drawing from an exponential distribution $Exp(1/(a+b))$. Equivalently stated, $\overline{X}$ is the random variable with characteristic function

$$\varphi_{\overline{X}}(t) = \frac{b}{a+b} + \frac{a}{a+b} \cdot \frac{1}{1-it(a+b)}.$$ (12)

Let $N \sim Exp(1/b)$ and independent of $\overline{X}$.

3. What is the distribution of $\overline{X} + N$?
   
   Tip: simplest would be to compute the characteristic function of $\overline{X} + N$ by recalling the relation $\varphi_{\overline{X}+N}(t) = \varphi_{\overline{X}}(t) \cdot \varphi_{N}(t)$.

   **Solution:**

   $$\varphi_{\overline{X}+N}(t) = \varphi_{\overline{X}}(t) \cdot \varphi_{N}(t)$$ (13)

   $$= \left( \frac{b}{a+b} + \frac{a}{a+b} \cdot \frac{1}{1-it(a+b)} \right) \frac{1}{1-itb}$$ (14)

   $$= \frac{1}{1-it(a+b)} \frac{a+b-itab-itb^2}{(a+b)(1-itb)}$$ (15)

   $$= \frac{1}{1-it(a+b)},$$ (16)

   which is the characteristic function of $Exp(1/a + b)$. Thus $\overline{X} + N$ is distributed as $Exp(1/a + b)$.

4. Find $I(\overline{X}; \overline{X} + N)$.

   **Solution:**

   $$I(\overline{X}; \overline{X} + N) = h(\overline{X} + N) - h(\overline{X} + N|\overline{X})$$ (17)

   $$\overset{N \perp \overline{X}}{=\,} h(\overline{X} + N) - h(N)$$ (18)

   $$= 1 + \log(a + b) - (1 + \log(b))$$ (19)

   $$= \log(1 + \frac{a}{b})$$ (20)

5. Consider the problem of communication over the additive exponential noise channel $Y = X + N$, where $N \sim Exp(1/b)$, independent of the channel input $X$, which is confined to being non-negative and satisfying the moment constraint $EX \leq a$. Find $C(a) = \max I(X; X + N)$, where the maximization is over all non-negative $X$ satisfying $EX \leq a$. What is the capacity-achieving distribution?

   **Hint:** Using findings from previous parts, show that for any non-negative random variable $X$, independent of $N$, with $EX \leq a$, we have $I(X; X + N) \leq I(\overline{X}; \overline{X} + N)$. 

---

Sample Final  Page 6 of 10
Solution:

For any feasible $X$, note that $X + N$ is a non-negative random variable and $E[X + N] = E[X] + E[N] \leq a + b$, thus by the result of Part (2) above, $h(X + N) \leq 1 + \log(a + b)$. Hence,

$$I(X; X + N) = h(X + N) - h(X + N | X) \quad (21)$$

$$N \perp X \Rightarrow h(X + N) - h(N) \quad (22)$$

$$\leq 1 + \log(a + b) - h(N) \quad (23)$$

$$= 1 + \log(a + b) - (1 + \log(b)) \quad (24)$$

$$= \log(1 + \frac{a}{b}) \quad (25)$$

$$= I(X; X + N), \quad (26)$$

Thus $C(a) \leq \log(1 + \frac{a}{b})$. Equality in (*) holds if $X = \overline{X}$ proving $C(a) = \log(1 + \frac{a}{b})$. Maximizing distribution is that of $\overline{X}$.

6. Let now $U \sim \text{Exp}(1/m)$ and consider the rate distortion problem where the expected reconstruction error is not to exceed $D$ and, in addition, the reconstruction is not allowed to exceed the source value. The associated rate distortion function is thus

$$R(D) = \min I(U; V), \quad (27)$$

where the minimum is over joint distributions respecting the given distribution of $U$ and such that $U - V \geq 0$ while $E[U - V] \leq D$. Explain why for any $U, V$ in the feasible set of this minimization, the following equality and inequalities hold:

$$I(U; V) = h(U) - h(U - V | V) \quad (28)$$

$$\geq h(U) - h(U - V) \quad (29)$$

$$\geq \log(m/D). \quad (30)$$

Solution:

$$I(U; V) \overset{(a)}{=} h(U) - h(U - V | V) \quad (31)$$

$$\overset{(b)}{=} h(U) - h(U - V) \quad (32)$$

$$\overset{(c)}{=} \log(m/D), \quad (33)$$

where (a) is due to definition of mutual information and that $h(U | V) = h(U - V | V)$, (b) is due to conditioning reduces entropy and (c) is due the fact as $h(U) = 1 + \log(m)$ by Part (1) of the problem and $h(U - V) \leq 1 + \log(D)$ by Part (2) of the problem.

7. Show that $R(D)$ from the previous part is given by

$$R(D) = \begin{cases} 
\log(m/D) & \text{if } 0 < D \leq m \\
0 & \text{if } D > m.
\end{cases} \quad (34)$$
Hint: Using findings from previous parts, for $0 < D \leq m$ establish existence of $(U, V)$ in the feasible set for which the inequalities in (29) and (30) hold with equality.

Solution:
First consider the case of $D > m$. $R(D) = 0$ is achievable by using just one trivial index for all source sequences, and $V = 0$, when $E[U - V] = 1/m < D$.

Now consider the case when $0 < D \leq m$. From Part (6) above, $R(D) \geq \log(m/D)$. For achievability consider, $Z \sim Exp(1/D)$ and construction random variable $V$ independent of $Z$ which takes value 0 with probability $\frac{D}{m}$ and is a positive exponential random variable $Exp(1/m)$ with probability, $\frac{m-D}{m}$. Let $U = V + Z$ and it is easy to show, on the lines of part (2), $U \sim Exp(1/m)$. Thus $U - V \perp V$ and $I(U; V) = h(U) - h(Z) = \log \frac{m}{D}$. 

Consider the lossless source coding problem in Figure 1. The pair \((X^n, Y^n)\) is generated by i.i.d. drawings of the finite alphabet pair \((X, Y)\), that is \(p(x^n, y^n) = \prod_{i=1}^{n} p_{XY}(x_i, y_i)\). The source sequence \(X^n\) is to be sent near-losslessly when \(Y^n\) is available at both the encoder and the decoder. Thus, a \((2^nR, n)\) code is defined by an encoder \(m(x^n, y^n) \in \{1, 2, \ldots, 2^nR\}\) and a decoder \(\hat{X}^n(m, y^n)\), and the probability of decoding error is defined as \(P_e = P\{\hat{X}^n = X^n\}\), where most explicitly \(\hat{X}^n = X^n(m(X^n, Y^n), Y^n)\). A rate \(R\) is achievable if there exists a sequence of codes with \(P_e \to 0\) as \(n \to \infty\). Let \(R^*\) be the infimum of achievable rates. In this problem you prove that \(R^* = H(X|Y)\).

**Problem 4 : Lossless Source Coding with Side Information**

\[
X^n \xrightarrow{\text{ENCODER}} m(X^n, Y^n) \in \{1 : 2^nR\} \xrightarrow{\text{DECODER}} \hat{X}^n(m, Y^n)
\]

Figure 1: Conditional Lossless Source Coding

1. Prove that any rate \(R > H(X|Y)\) is achievable.
   
   Hint: can use the fact that \(|T^{(n)}_{\epsilon}(X|y^n)| \leq 2^{n(H(X|Y)+\delta(\epsilon))}\) for every \(y^n \in T^{(n)}_{\epsilon}(Y)\).

   **Solution :**
   Fix \(\epsilon > 0\). Given a particular \(y^n\), enumerate all the sequences in \(T^{(n)}_{\epsilon}(X|y^n)\) with \([1 : 2^{n(H(X|Y)+\delta(\epsilon))}]\) indices. Now the encoder and decoder do the following,
   - **Encoding**: If \(y^n \notin T^{(n)}_{\epsilon}(Y)\), index 0 is sent. If \(y^n \in T^{(n)}_{\epsilon}(Y)\), and \(x^n \notin T^{(n)}_{\epsilon}(X|y^n)\) again index 0 is described, else if \(x^n \in T^{(n)}_{\epsilon}(X|y^n)\), the corresponding index of enumeration in \(T^{(n)}_{\epsilon}(X|y^n)\) is described.
   - **Decoding**: If decoder gets 0 it does whatever, else it decipheres \(x^n\) from \(T^{(n)}_{\epsilon}(z|x|y^n)\), as it knows \(y^n\).

   Probability of error,
   \[
P_e = P(Y^n \notin T^{(n)}_{\epsilon}(Y)) + P(Y^n \notin T^{(n)}_{\epsilon}(Y) \cap (X^n, Y^n) \notin T^{(n)}_{\epsilon}(X, Y)) \to 0, \tag{35}
   \]
   as \(n \to \infty\) by law of large numbers. Thus as \(n \to \infty\), the rate achieved is \(R = H(X|Y) + \delta(\epsilon)\), where \(\delta(\epsilon) \to 0\) as \(\epsilon \to 0\).

2. Prove that any rate \(R < H(X|Y)\) is not achievable via the following steps:
(a) For $M = m(X^n, Y^n)$ argue why

$$I(X^n; M|Y^n) \leq nR.$$

**Solution :**

\[
\begin{align*}
nR & \geq H(M) \\ & \overset{(*)}{\geq} H(M|Y^n) \\ & \overset{(**)}{=} H(M|Y^n) - H(M|X^n, Y^n) \\ & = I(X^n; M|Y^n),
\end{align*}
\]

where $(*)$ is due to conditioning reduces entropy and $(**)$ follows from the fact that $M$ is a function of $(X^n, Y^n)$.

(b) Using the previous step, show that

$$H(X^n|M, Y^n) \geq n[H(X|Y) - R].$$

**Solution :** From part (a) above,

\[
\begin{align*}
nR & \geq I(X^n; M|Y^n) \\ & = H(X^n|Y^n) - H(X^n|M, Y^n) \\ & \overset{(*)}{=} nH(X|Y) - H(X^n|M, Y^n),
\end{align*}
\]

where $(*)$ follows from the memoryless nature of the 2-DMS source in the problem. The claim in this question follows after rearranging the last inequality.

(c) Use the previous step and a relation that you know between entropy and probability of error to deduce that if $R < H(X|Y)$ then one cannot get $P_e \to 0$ as $n \to \infty$.

**Solution :**

Invoking Fano’s Inequality (course reader Lemma 6.1, substituting $X = X^n$ and $Y = (M, Y^n)$), we obtain,

$$H(X^n|M, Y^n) \leq 1 + P_e nR,$$

thus combining the above inequality with part (b) above,

$$nR \geq nH(X|Y) - 1 - P_e nR,$$

or

$$P_e \geq \frac{H(X|Y) - R}{R} - \frac{1}{nR},$$

if $R < H(X|Y)$ as $n \to \infty$, $P_e$ will be bounded away from 0.