Problem 1. Solution:
(a) In this problem, we can calculate the
\[ g(\beta) = \max \{ H(X) : E(X^2) = \beta \} \] (1)
and plot for \( g(\beta) vs \beta \). Then from the \( g(\beta) vs \beta \), we can easily obtain the \( f(\beta) vs \beta \) plot.
First, assume that the input probability density is \( p = (p_1, p_2, ..., p_5) \). Then
\[ E(X^2) = \beta \] (2)
\[ \Rightarrow p_1 (-2)^2 + p_2 (-1)^2 + p_3 (0)^2 + p_4 (1)^2 + p_5 (2)^2 = \beta \] (3)
\[ \Rightarrow 4p_1 + p_2 + p_4 + 4p_5 = \beta \] (4)
From lecture, we know that for a constrained entropy maximization problem,
\[ C(\beta) = \max \{ H(X) : p \cdot c = \beta \} \] (5)
The boltzman distribution gives the maximized entropy. So for this problem, we try to use the result and use analogy to obtain the \( g(\beta) \).
We define the partition function
\[ Z(s) = \sum_{i=1}^{m} e^{-sc_i} \] (6)
where, from Eqn. (4), \( c = (4, 1, 0, 1, 4) \). So
\[ Z(s) = \sum_{i=1}^{m} e^{-sc_i} \] (7)
\[ = 2e^{-4s} + 2e^{-s} + 1 \] (8)
The optimizing input distribution will be the boltzman distribution,
\[ p_{i \text{ max}} = \frac{e^{-sc_i}}{Z(s)} \] (9)
and the corresponding maximized output is
\[ g(\beta) = \log Z(s) - s \frac{Z'(s)}{Z(s)} \] (10)
\[ = \log (2e^{-4s} + 2e^{-s} + 1) - s \frac{-8e^{-4s} - 2e^{-s}}{2e^{-4s} + 2e^{-s} + 1} \] (11)
\[ \beta = - \frac{Z'(s)}{Z(s)} = - \frac{-8e^{-4s} - 2e^{-s}}{2e^{-4s} + 2e^{-s} + 1} \] (12)
Figure 1:
Thus, by using MATLAB, we can plot the parametric functions \( g(\beta) \) vs \( \beta \), as shown in Fig. (1).

To obtain the plot between \( \beta \) and \( f(\beta) \) on \( 0 \leq \beta \leq 8 \), we just need to analyze Fig. (1) a little bit. In the figure, we see that \( g(\beta) \) maximizes at \( \beta_{\text{max}} \approx 2 \), whose exactly value will be gained in part b.

So for \( \beta \leq \beta_{\text{max}} \),

\[
f(\beta) = \max \{ H(X) : E(X^2) \leq \beta \} \quad (13)
= \max \{ H(X) : E(X^2) = \beta \} \quad (14)
= g(\beta) \quad (15)
\]

and for \( \beta_{\text{max}} < \beta \leq 8 \), we have

\[
f(\beta) = \max \{ H(X) : E(X^2) \leq \beta \} \quad (16)
= \max \{ H(X) : E(X^2) = \beta_{\text{max}} \} \quad (17)
= \max \{ g(\beta) \} \quad (18)
\]

Thus, by running MATLAB, we can obtain the accurate plot of \( f(\beta) \) for the range \( 0 \leq \beta \leq 8 \) as shown in Fig. (2).
(b) By running MATLAB, we found that at
\[ \beta_{\text{max}} = 2 \quad (19) \]
the function \( f(\beta) \) achieves its maximum value 1.6094 nats. The corresponding value of \( s \) is
\[ s_{\text{max}} = 0 \quad (20) \]
therefore the input probability density is
\[ p_i = \frac{e^{-sc_i}}{Z(s)} \quad (21) \]
\[ = \frac{e^{-sc_i}}{2e^{-4s} + 2e^{-s} + 1} \quad (22) \]
\[ = \frac{1}{5} \quad (23) \]
so the density is uniformly distributed.

**Problem 2. Solution:**

(a) For an input density \( p = (p_1, p_2, \ldots, p_m) \), the corresponding output density \( r(p) \) satisfies
\[ r_j(p) = \sum_{i=1}^{m} p_i Q_{ij} \quad (24) \]
where \( i \) is the row index and \( j \) is the column index of the \( m \times n \) DMC matrix \( Q \). In this problem, \( m = 3 \).

Then, we have
\[ r(p) = \begin{bmatrix}
\sum_{i=1}^{3} p_1 Q_{1i} & \sum_{i=1}^{3} p_2 Q_{2i} & \sum_{i=1}^{3} p_3 Q_{3i}
\end{bmatrix}^T \quad (25) \]
\[ = \begin{bmatrix}
0.6p_1 + 0.7p_2 + 0.5p_3 \\
0.3p_1 + 0.1p_2 + 0.05p_3 \\
0.1p_1 + 0.2p_2 + 0.45p_3
\end{bmatrix} \quad (26) \]
For initial condition, we set \( p_1 = p_2 = p_3 = 1/3 \), thus
\[ r(p) = \begin{bmatrix}
0.6 & 0.15 & 0.25
\end{bmatrix} \quad (27) \]
We know that
\[ I_i(p) = D(r_i||r(p)) \quad (28) \]
\[ = \sum_{j=1}^{n} r_{ij} \log \frac{r_{ij}}{r_j(p)} \quad (29) \]
hence for the first iteration,
\[ I_1(p) = \sum_{j=1}^{3} r_{1j} \log \frac{r_{1j}}{r_j(p)} \quad (30) \]
\[ = 0.6 \log 0.6 + 0.3 \log 0.15 + 0.1 \log 0.25 \quad (31) \]
\[ = 0.3 \log 2 + 0.1 \log 0.4 \quad (32) \]
\[ = 0.1678 \text{ bits} \quad (33) \]
\[ I_2 (p) = \sum_{j=1}^{3} r_{2j} \log \frac{r_{2j}}{r_j (p)} \]  
\[ = 0.7 \log \frac{0.7}{0.6} + 0.1 \log \frac{0.1}{0.15} + 0.2 \log \frac{0.2}{0.25} \]  
\[ = 0.03279 \text{ bits} \]  
\[ I_3 (p) = \sum_{j=1}^{3} r_{3j} \log \frac{r_{3j}}{r_j (p)} \]  
\[ = 0.5 \log \frac{0.5}{0.6} + 0.05 \log \frac{0.05}{0.15} + 0.45 \log \frac{0.45}{0.25} \]  
\[ = 0.1708 \text{ bits} \]  

Then

\[ I_U (p) = \max_{i=1}^{m} I_i (p) \]  
\[ = \max_{i=1}^{3} I_i (p) \]  
\[ I_L (p) = \log \sum_{i=1}^{m} p_i e^{I_i (p)} \]  
\[ = \log \sum_{i=1}^{3} p_i e^{I_i (p)} \]  

Similarly, by computer programming, we can plot the lower bound \( I_L \) and upper bound \( I_U \) on the capacity (in bits) as a function of the number of iterations for the first 200 iterations as shown in Fig. (3).

(b) The channel capacity will be

\[ C = 0.1616 \text{ bits} \]  
\[ \approx 0.162 \text{ bits} \]  

and the optimizing input probability distribution is

\[ p = (0.5017, 0, 0.4983) \]  

**Problem 3. Solution:**

**Approach one:**
Figure 3:
(a) To obtain the reliable communication, we require that the error probability \( P_e \to 0 \) as the length of the sequence \( n \to \infty \). The size of the decoding net is \( 2^{nH(X|Y)} \), and the size the typical sequence set is \( 2^{nH(X)} \). Thus for any codework to fall in the decoding net, the probability will be

\[
p = \frac{2^{nH(X|Y)}}{2^{nH(X)}} = 2^{-nI(X;Y)} = 2^{-nC}
\]

where \( C \) is the channel capacity.

The new decoding strategy means that when there are \( L \) or more than \( L \) other codewords besides the causal codeword in the net, we claim that decoding fails. (According to Shannon, the causal codeword will always be in the net as \( n \to \infty \)). We regard every \( L \) codewords from the \( 2^{nR} \) codewords being in the decoding net as an event. In other words, when the decoding net has exactly \( L \) other codewords besides the causal codeword, then we say an event happens. Obviously, there are \( N = C_L^{2nR} \) such events, which are labeled as \( E_1, E_2, ..., E_N \). Of course, they are not totally independent of each other. So the error probability is

\[
P_e = \Pr \{ \# \text{ of other codewords in the net } \geq L + 1 \}
\]

\[
= \Pr \{ \text{at least one of } E_i \text{ happens, } i = 1, 2, ..., N \}
\]

\[
= \Pr \{ E_1 \cup E_2 \cup ... \cup E_N \}
\]

\[
\leq \sum_{i=1}^{N} \Pr \{ E_i \} \quad \text{(union bound)}
\]

\[
= \sum_{i=1}^{N} p^L
\]

\[
= \sum_{i=1}^{N} (2^{-nC})^L
\]

\[
= N (2^{-nC})^L
\]

\[
= C_L^{2nR} 2^{-nCL}
\]

\[
= \frac{2^{nR} \left( 2^{nR} - 1 \right) \cdots \left( 2^{nR} - L + 1 \right)}{L!} 2^{-nCL}
\]

\[
= \frac{2^{nR} \left( 2^{nR} - 1 \right) \cdots \left( 2^{nR} - L + 1 \right)}{(2C)^L} \frac{1}{L!}
\]

To obtain the reliable communication, we require that the error probability \( P_e \to 0 \) as the length of the sequence \( n \to \infty \). So

\[
\lim_{n \to \infty} \frac{2^{nR} \left( 2^{nR} - 1 \right) \cdots \left( 2^{nR} - L + 1 \right)}{(2C)^L}
\]

\[
= \lim_{n \to \infty} \frac{2^{nR} \cdot 2^{nR} \cdot \ldots \cdot 2^{nR}}{(2C)^L}
\]

\[
= \lim_{n \to \infty} \frac{(2^{nR})^L}{(2C)^L}
\]

\[
= \lim_{n \to \infty} 2^{-n(C-R)L}
\]

\[
\to 0
\]
Therefore,

\[ R < C \] (63)

(b) When \( L = \log n \), we have

\[ P_e = \frac{2^{nR} (2^{nR} - 1) \ldots (2^{nR} - L + 1)}{(2^{nC})^L} \frac{1}{L!} \] (64)

\[ = \frac{2^{nR} (2^{nR} - 1) \ldots (2^{nR} - \log n + 1)}{(2^{nC})^L} \frac{1}{L!} \] (65)

Because

\[ \lim_{n \to \infty} \log n \frac{2^{nC}}{2^{nR}} = 0 \] (66)

thus

\[ \lim_{n \to \infty} \frac{2^{nR} (2^{nR} - 1) \ldots (2^{nR} - \log n + 1)}{(2^{nC})^L} \frac{1}{L!} \] (67)

\[ = \lim_{n \to \infty} \frac{2^{nR} \cdot 2^{nR} \cdot \ldots \cdot 2^{nR}}{(2^{nC})^L} \] (68)

\[ = \lim_{n \to \infty} \frac{2^{nRL}}{2^{nCL}} \] (69)

\[ = \lim_{n \to \infty} 2^{-n(C-R) \log n} \] (70)

\[ \to 0 \] (71)

thus we still require that

\[ R < C \] (72)

**Approach Two**

(a) To obtain the reliable communication, we require that the error probability \( P_e \to 0 \) as the length of the sequence \( n \to \infty \). The size of the decoding net is \( 2^{nH(X)} \), and the size the typical sequence set is \( 2^{nH(X)} \).

Thus for any codeword to fall in the decoding net, the probability will be

\[ p = \frac{2^{nH(X)}(Y)}{2^{nH(X)}} = 2^{-nI(X;Y)} = 2^{-nC} \] (73)

where \( C \) is the channel capacity.

For the certain communication channel link, the channel capacity keeps unchanged. We are seeking the highest transmission scheme to achieve the channel capacity. In Shannon’s decoding strategy, \( L=1 \), thus the error probability \( p_e \) is the sum of the probabilities of more than 1 codewords falling into the net. As \( n \to \infty \), Shannon states that \( p_e \to 0 \). Now in the new case in which \( L \geq 1 \), the error probability \( p_e \) is the sum of the probabilities of more than \( L \) codewords falling into the net. The two cases are the same if we choose the same typical sequence set and codewords, except for the difference in the decoding strategy. Hence, for
the same $n$, $p_{e(L>1)} < p_{e(L=1)}$. Since Shannon theorem says that we can achieve $R=C$ while $p_{e(L=1)}$ still $\to 0$ as $n \to \infty$, we conclude that for $L>1$, $p_{e(L=1)}$ also $\to 0$ as $n \to \infty$, therefore we can achieve $R\to C$ as well. So

$$R < C$$  \hfill (74)

(b) The difference between this part and part a is that $L$ is constant in part a but is increasing with $n$ in this part. Note that

$$\lim_{n \to \infty} \frac{n}{\log n} = \infty$$  \hfill (75)

and

$$\lim_{n \to \infty} \frac{a^n}{\log n} = \infty$$  \hfill (76)

where $a>1$. Thus compared to the increasing speed of the size of both the decoding net and typical sequence set, the increasing speed of $L$ is very small and is almost 0 in a relative scale. Thus intuitively, we guess that we should get the same result as part a when $L$ is constant.

Actually, by the same analysis as part a, we know that the error probability goes to 0 as $n \to \infty$, because the error probability is smaller than that of $L=1$ case if all the other conditions remain the same. So we know that we can still achieve the channel capacity in rate while we still keep the error probability stay decreasing to 0 as the length of the sequence becomes infinity. Therefore, we remain the reliable communication at the rate such that

$$R < C$$  \hfill (77)