Unit 1
PART B
Note that the answer for the questions is indicated via the page number given alongside. Match the page numbers to the scanned pages given below to get the answer.

1. What is Entropy? Explain the properties and types of Entropy? (1-7 to 1-11)
2. Explain about the relationship between Joint and Conditional Entropy. (1-7)
3. The Joint Probability Matrix is given as

$$\begin{bmatrix} 0.3 & 0.05 & 0 \\ 0 & 0.25 & 0 \\ 0 & 0.15 & 0.05 \\ 0 & 0.05 & 0.15 \end{bmatrix}$$

Find all the Entropies and Mutual Information.

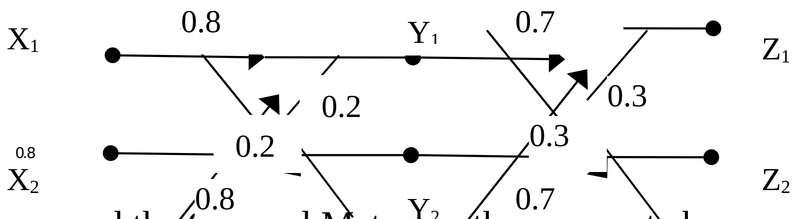
4 Prove that the Upper bound on Entropy is given as $H_{max} \leq \log_2 M$. Here 'M' is the number of messages emitted by the source. (1-11)

5. Prove that $H(X,Y)=H(X/Y)+H(Y)$ (1-72)
$=H(Y/X)+H(X)$

6. (i) A channel has the following Channel matrix.

$$P(Y/X) = \begin{bmatrix} 1-P & P & 0 \\ 0 & P & 1-P \end{bmatrix}$$ (1-70)

(a)Draw the Channel diagram.
(b)If the source has equally likely outputs, Compute the probabilities associated with the channel outputs for P=0.2 (6 marks)
(ii)Two BSC's are connected in cascade as shown in the figure.

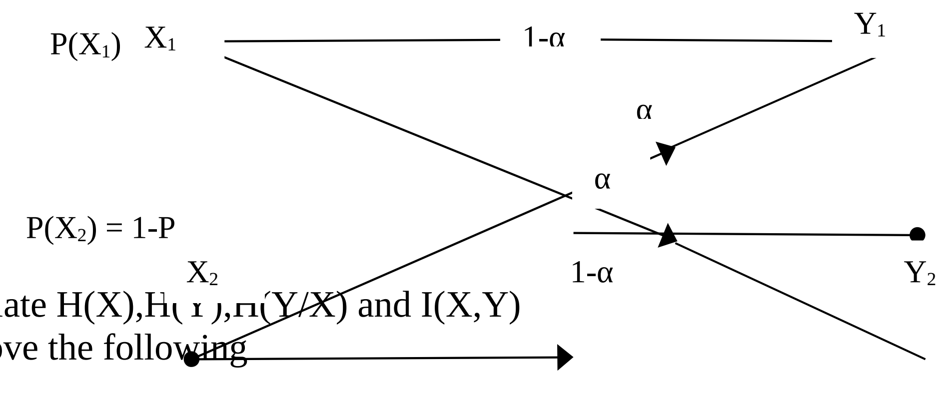


(a)Find the Channel Matrix of the resultant channel. (1-74)
(b)Find P (Z_1) and P (Z_2), if $P(X_1)=0.6$, $P(X_2)=0.4$

7. (i)Prove that the Mutual information of the channel is Symmetric.
$I(X, Y)=I(Y, X)$ (6 marks) (1-87)
(ii)Prove that the mutual information is always positive
$I(X, Y) \geq 0$ (6marks) (1-89)

8.Prove the following relationships:
a) $I(X,Y)=H(X)-H(X/Y)$ (1-87)
b) $I(X,Y)=H(Y)-H(Y/X)$ (1-89)

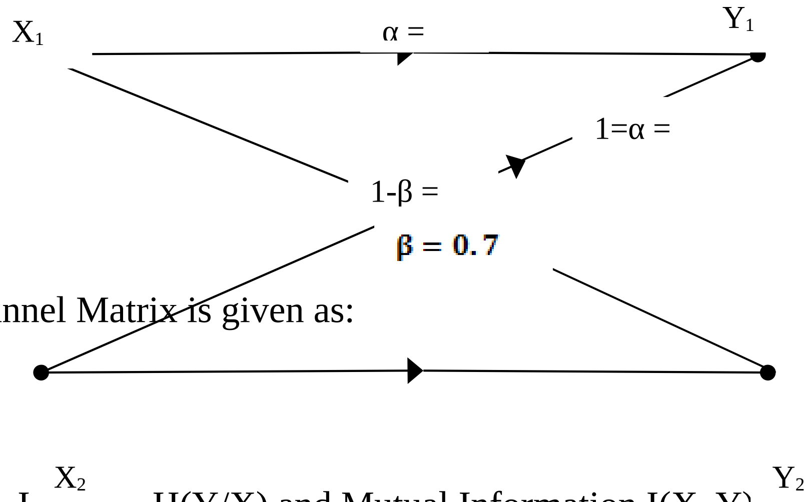
9.(i) Consider the Binary Symmetric Channel shown in the figure. (1-91)



Calculate H(X),H(Y),H(Y/X) and I(X,Y)
(ii)Prove the following (1-90)

$I(X,Y)=H(X)+H(Y)-H(X,Y)$

10. (a)Find the Mutual Information and Channel capacity for the channel shown in the figure. Given that $P(X_1) = 0.6$ and $P(X_2) = 0.4$ (6 Marks) (1-94)



(b)A Binary Channel Matrix is given as:

$$\begin{bmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{10} & \frac{9}{10} \end{bmatrix}$$

Determine H(X), H(X/Y), H(Y/X) and Mutual Information I(X, Y) (1-98)

Review Questions

1. *Explain the concept of amount of information. Also explain what is infinite information and zero information.*
2. *With the help of an example give physical interpretation of amount of information.*

Unsolved Example

1. *A source emits four symbols with probabilities,* $p_0 = 0.4$, $p_1 = 0.3$, $p_2 = 0.2$ and $p_3 = 0.1$. *Find out the amount of information obtained due to these four symbols.* **[Ans. : 8.703 bits]**

1.4 Entropy (Average Information)

Consider that we have M-different messages. Let these messages be $m_1, m_2, m_3, \ldots m_M$ and they have probabilities of occurrence as $p_1, p_2, p_3, \ldots p_m$. Suppose that a sequence of L messages is transmitted. Then if L is very very large, then we may say that,

$p_1\, L$ messages of m_1 are transmitted,

$p_2\, L$ messages of m_2 are transmitted,

$p_3\, L$ messages of m_3 are transmitted,

$\vdots \qquad \vdots$

$p_M\, L$ messages of m_M are transmitted.

Hence the information due to message m_1 will be,

$$I_1 = \log_2\left(\frac{1}{p_1}\right)$$

Since there are $p_1\, L$ number of messages of m_1, the total information due to all message of m_1 will be,

$$I_{1(total)} = p_1\, L \log_2\left(\frac{1}{p_1}\right)$$

Similarly the total information due to all messages m_2 will be,

$$I_{2(total)} = p_2\, L \log_2\left(\frac{1}{p_2}\right) \qquad \text{and so on.}$$

Thus the total information carried due to the sequence of L messages will be,

$$I_{(total)} = I_{1(total)} + I_{2(total)} + \ldots\ldots + I_{M(total)} \qquad \ldots (1.4.1)$$

$$\therefore\ I_{(total)} = p_1\, L \log_2\left(\frac{1}{p_1}\right) + p_2\, L \log_2\left(\frac{1}{p_2}\right) + \ldots + p_M\, L \log_2\left(\frac{1}{p_M}\right) \qquad \ldots (1.4.2)$$

The average information per message will be,

$$\text{Average information} = \frac{\text{Total information}}{\text{Number of messages}}$$

$$= \frac{I_{(total)}}{L} \quad \text{... (1.4.3)}$$

Average information is represented by *Entropy*. It is represented by H. Thus,

$$\text{Entropy (H)} = \frac{I_{(total)}}{L} \quad \text{... (1.4.4)}$$

From equation 1.4.2 we can write above equation as,

$$\text{Entropy (H)} = p_1 \log_2\left(\frac{1}{p_1}\right) + p_2 \log_2\left(\frac{1}{p_2}\right) + \ldots + p_M \log_2\left(\frac{1}{p_M}\right) \quad \text{... (1.4.5)}$$

We can write above equation using $\sum$ sign as follows :

$$\text{Entropy : } \quad H = \sum_{k=1}^{M} p_k \log_2\left(\frac{1}{p_k}\right) \quad \text{... (1.4.6)}$$

1.4.1 Properties of Entropy

1. Entropy is zero if the event is sure or it is impossible. i.e.,

$$H = 0 \quad \text{if} \quad p_k = 0 \text{ or } 1.$$

2. When $p_k = 1/M$ for all the 'M' symbols, then the symbols are equally likely. For such source entropy is given as $H = \log_2 M$.

3. Upper bound on entropy is given as,

$$H_{max} = \log_2 M$$

Proofs of these properties is given in next examples.

Example 1.4.1 : *Calculate entropy when* $p_k = 0$ *and when* $p_k = 1$.

Solution : Consider equation 1.4.6,

$$H = \sum_{k=1}^{M} p_k \log_2\left(\frac{1}{p_k}\right)$$

Since $p_k = 1$, the above equation will be,

$$H = \sum_{k=1}^{M} \log_2\left(\frac{1}{1}\right) = \sum_{k=1}^{M} \frac{\log_{10}(1)}{\log_{10}(2)}$$

$$= 0 \qquad \text{Since } \log_{10} 1 = 0$$

Now consider the second case when $p_k = 0$. Instead of putting $p_k = 0$ directly let us consider the limiting case i.e.,

$$H = \sum_{k=1}^{M} p_k \log_2\left(\frac{1}{p_k}\right) \quad \text{By equation 1.4.6}$$

With p_k tending to '0' above equation will be,

$$H = \sum_{k=1}^{M} \lim_{p_k \to 0} p_k \log_2\left(\frac{1}{p_k}\right)$$

The RHS of above equation will be zero when $p_k \to 0$. Hence entropy will be zero. i.e.,

$$H = 0$$

Thus entropy is zero for both certain and most rare message.

Example 1.4.2 : *A source transmits two independent messages with probabilities of p and $(1-p)$ respectively. Prove that the entropy is maximum when both the messages are equally likely. Plot the variation of entropy (H) as a function of probability 'p' of the messages.*

Solution : We know that entropy is given as,

$$H = \sum_{k=1}^{M} p_k \log_2\left(\frac{1}{p_k}\right)$$

For two messages above equation will be,

$$H = \sum_{k=1}^{2} p_k \log_2\left(\frac{1}{p_k}\right) = p_1 \log_2\left(\frac{1}{p_1}\right) + p_2 \log_2\left(\frac{1}{p_2}\right)$$

Here we have two messages with probabilities $p_1 = p$ and $p_2 = 1-p$. Then above equation becomes,

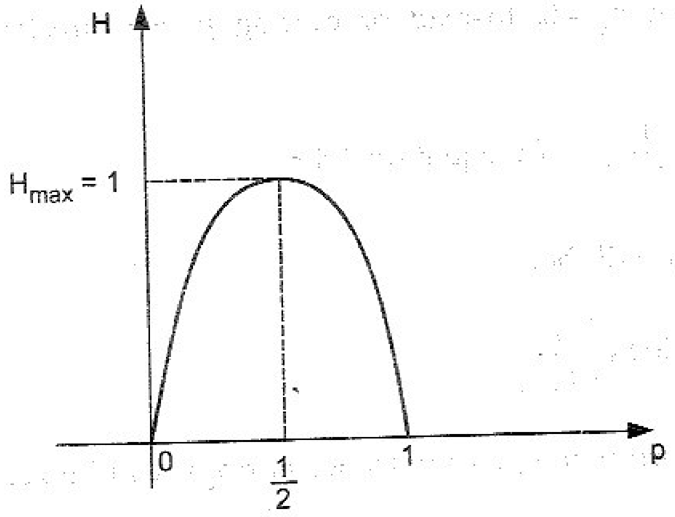
$$H = p \log_2\left(\frac{1}{p}\right) + (1-p) \log_2\left(\frac{1}{1-p}\right) \quad \ldots (1.4.7)$$

A plot of H as a function of p is plotted in Fig. 1.4.1.

As shown in Fig. 1.4.1, Entropy is maximum at $p = \frac{1}{2}$. Putting $p = \frac{1}{2}$ in equation 1.4.7 we get,

$$H_{max} = \frac{1}{2} \log_2 (2) + \frac{1}{2} \log_2 (2)$$

$$= \log_2 (2) = \frac{\log_{10} (2)}{\log_{10} (2)} = 1 \text{ bit/message.}$$



p	H
0	0
0.2	0.7215
0.4	0.9709
0.5	1
0.6	0.9709
0.8	0.7215
1	0

Fig. 1.4.1 Plot of entropy 'H' with probability 'p' for two messages. The maximum of H occurs at $p = \frac{1}{2}$, i.e. when messages are equally likely

This shows that H_{max} occurs when both the messages have same probability, i.e. when they are equally likely.

Example 1.4.3 : *Show that if there are 'M' number of equally likely messages, then entropy of the source is* $\log_2 M$.

Solution : We know that for 'M' number of equally likely messages, probability is,

$$p = \frac{1}{M}$$

This probability is same for all 'M' messages. i.e.,

$$p_1 = p_2 = p_3 = p_4 = \ldots p_M = \frac{1}{M} \quad \ldots (1.4.8)$$

Entropy is given by equation 1.4.6,

$$H = \sum_{k=1}^{M} p_k \log_2 \left(\frac{1}{p_k}\right)$$

$$= p_1 \log_2 \left(\frac{1}{p_1}\right) + p_2 \log_2 \left(\frac{1}{p_2}\right) + \ldots + p_M \log_2 \left(\frac{1}{p_M}\right)$$

Putting for probabilities from equation 1.4.8 in above equation we get,

$$H = \frac{1}{M} \log_2 (M) + \frac{1}{M} \log_2 (M) + \ldots + \frac{1}{M} \log_2 (M)$$

(Add 'M' number of terms)

In the above equation there are 'M' number of terms in summation. Hence after adding these terms above equation becomes,

$$H = \log_2 (M) \quad \text{... (1.4.9)}$$

Example 1.4.4 : *Prove that the upper bound on entropy is given as $H_{max} \leq \log_2 M$. Here 'M' is the number of messages emitted by the source.*

Solution : To prove the above property, we will use the following property of natural logarithm :

$$\ln x \leq x-1 \quad \text{for} \quad x \geq 0 \quad \text{... (1.4.10)}$$

Let us consider any two probability distributions $\{p_1, p_2, \ldots\ldots p_M\}$ and $\{q_1, q_2, \ldots q_M\}$ on the alphabet $X = \{x_1, x_2 \ldots. x_M\}$ of the discrete memoryless source. Then let us consider the term $\sum_{k=1}^{M} p_k \log_2 \left(\frac{q_k}{p_k}\right)$. This term can be written as,

$$\sum_{k=1}^{M} p_k \log_2 \left(\frac{q_k}{p_k}\right) = \sum_{k=1}^{M} p_k \frac{\log_{10}\left(\frac{q_k}{p_k}\right)}{\log_{10} 2}$$

Multiply the RHS by $\log_{10} e$ and rearrange terms as follows :

$$\sum_{k=1}^{M} p_k \log_2 \left(\frac{q_k}{p_k}\right) = \sum_{k=1}^{M} p_k \frac{\log_{10} e}{\log_{10} 2} \cdot \frac{\log_{10}\left(\frac{q_k}{p_k}\right)}{\log_{10} e}$$

$$= \sum_{k=1}^{M} p_k \log_2 e \cdot \log_e \left(\frac{q_k}{p_k}\right)$$

Here $\log_e \left(\frac{q_k}{p_k}\right) = \ln \left(\frac{q_k}{p_k}\right)$. Hence above equation becomes,

$$\sum_{k=1}^{M} p_k \log_2 \left(\frac{q_k}{p_k}\right) = \log_2 e \sum_{k=1}^{M} p_k \ln \left(\frac{q_k}{p_k}\right)$$

From the equation 1.4.10 we can write

$\ln \left(\frac{q_k}{p_k}\right) \leq \left(\frac{q_k}{p_k} - 1\right)$ Hence above equation becomes,

$$\sum_{k=1}^{M} p_k \log_2 \left(\frac{q_k}{p_k}\right) \leq \log_2 e \sum_{k=1}^{M} p_k \left(\frac{q_k}{p_k} - 1\right)$$

The conditional entropy $H(X / Y)$ is an average measure of uncertainty in X after Y is received. In other words $H(X / Y)$ represents the information lost in the noisy channel.

Example 1.9.2 : *Prove that*

$$H(X, Y) = H(X/Y) + H(Y)$$
$$= H(Y/X) + H(X)$$

Solution : Consider equation 1.9.18,

$$H(X,Y) = \sum_{i=1}^{M} \sum_{j=1}^{M} P(x_i, y_j) \log_2 \frac{1}{P(x_i, y_j)}$$

$$= -\sum_{i=1}^{M} \sum_{j=1}^{M} P(x_i, y_j) \log_2 P(x_i, y_j) \quad \text{... (1.9.20)}$$

From probability theory we know that,

$$P(AB) = P(A / B)\, P(B)$$

$$\therefore \quad P(x_i, y_j) = P(x_i / y_j)\, P(y_j)$$

Putting this result in the $\log_2$ term of equation (1.9.20) we get,

$$H(X,Y) = -\sum_{i=1}^{M} \sum_{j=1}^{M} P(x_i, y_j) \log_2 [P(x_i / y_j)\, P(y_j)]$$

We know that $\log_2 \left[P\left(x_i / y_j\right) P\left(y_j\right)\right] = \log_2 P\left(x_i / y_j\right) + \log_2 P\left(y_j\right)$.

Hence above equation becomes,

$$H(X,Y) = -\sum_{i=1}^{M} \sum_{j=1}^{M} P(x_i, y_j) \log_2 P(x_i / y_j)$$
$$-\sum_{i=1}^{M} \sum_{j=1}^{M} P(x_i, y_j) \log_2 P(y_j)$$

$$= \sum_{i=1}^{M} \sum_{j=1}^{M} P(x_i, y_j) \log_2 \frac{1}{P(x_i / y_j)}$$
$$-\sum_{j=1}^{M} \left\{ \sum_{i=1}^{M} P(x_i, y_j) \right\} \log_2 P(y_j)$$

The first term in above equation is $H(X/Y)$ as per equation 1.9.5. From the standard probability theory,

$$\sum_{i=1}^{M} P(x_i, y_j) = P(y_j)$$

Hence $H(X, Y)$ will be written as,

$$H(X,Y) = H(X/Y) - \sum_{j=1}^{M} P(y_j) \log_2 P(y_j)$$

$$= H(X/Y) + \sum_{j=1}^{M} P(y_j) \log_2 \frac{1}{P(y_j)}$$

As per the definition of entropy, the second term in the above equation is $H(Y)$. Hence,

$$H(X,Y) = H(X/Y) + H(Y) \qquad \text{... (1.9.21)}$$

Thus the first given equation is proved. From the probability theory we know that

$$P(AB) = P(B/A)\, P(A)$$

$$\therefore \qquad P(x_i, y_j) = P(y_j / x_i)\, P(x_i)$$

Putting this result in the $\log_2$ term of equation (1.9.20) we get,

$$H(X,Y) = -\sum_{i=1}^{M} \sum_{j=1}^{M} P(x_i, y_j) \log_2([P(y_j / x_i)\, P(x_i)]$$

$$= -\sum_{i=1}^{M} \sum_{j=1}^{M} P(x_i, y_j) \log_2 P(y_j / x_i)$$

$$-\sum_{i=1}^{M} \sum_{j=1}^{M} P(x_i, y_j) \log_2 P(x_i)$$

$$= \sum_{i=1}^{M} \sum_{j=1}^{M} P(x_i, y_j) \log_2 \frac{1}{P(y_j / x_i)}$$

$$-\sum_{i=1}^{M} \left\{ \sum_{j=1}^{M} P(x_i, y_j) \right\} \log_2 P(x_i)$$

As per equation 1.9.11, the first term of above equation is $H(Y/X)$. And from standard probability theory,

$$\sum_{j=1}^{M} P(x_i, y_j) = P(x_i)$$

Hence $H(X,Y)$ will be written as,

$$H(X,Y) = H(Y/X) - \sum_{i=1}^{M} P(x_i)\log_2 P(x_i)$$

$$= H(Y/X) + \sum_{i=1}^{M} P(x_i)\log_2 \frac{1}{P(x_i)}$$

As per the definition of entropy, the second term in the above equation is $H(X)$. Hence

$$H(X,Y) = H(Y/X) + H(X) \quad \text{... (1.9.22)}$$

Thus the second part of the given equation is proved.

Example 1.9.3 : *Two BSC's are connected in cascade as shown in Fig. 1.9.4.*

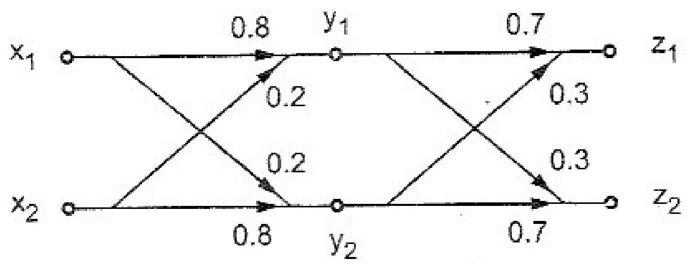


Fig. 1.9.4 BSC of Example 1.9.3

i) Find the channel matrix of resultant channel.

ii) Find $P(z_1)$ and $P(z_2)$ if $P(x_1) = 0.6$ and $P(x_2) = 0.4$

Solution : i) To obtain channel matrix using equation (1.9.1)

Here we can write two matrices of the channel as follows :

$$P(Y/X) = \begin{bmatrix} P(y_1/x_1) & P(y_2/x_1) \\ P(y_1/x_2) & P(y_2/x_2) \end{bmatrix} = \begin{bmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{bmatrix}$$

$$\text{and} \quad P(Z/Y) = \begin{bmatrix} P(z_1/y_1) & P(z_2/y_1) \\ P(z_1/y_2) & P(z_2/y_2) \end{bmatrix} = \begin{bmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{bmatrix}$$

Hence resultant channel matrix is given as,

$$P(Z/X) = P(Y/X) \cdot P(Z/Y)$$

1.9.2 Binary Symmetric Channel (BSC)

The binary communication channel of Fig. 1.9.1 is said to be symmetric if $P(y_0 / x_0) = P(y_1 / x_1) = p$. Such channel is shown in Fig. 1.9.2.

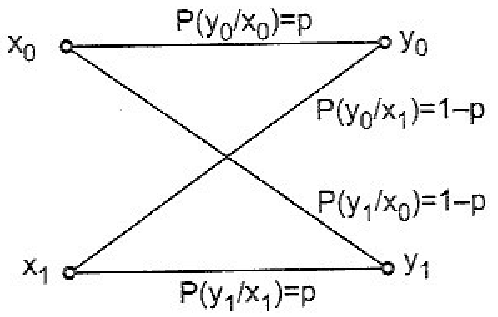


Fig. 1.9.2 Binary symmetric channel

For the above channel, we can write equation (1.9.15) as,

$$\begin{bmatrix} P(y_0) \\ P(y_1) \end{bmatrix} = \begin{bmatrix} P(x_0) & P(x_1) \end{bmatrix} \begin{bmatrix} p & 1-p \\ 1-p & p \end{bmatrix} \quad \text{... (1.9.16)}$$

1.9.3 Binary Erasure Channel (BEC)

The channel diagram of binary erasure channel is shown in Fig. 1.9.3. Note that there are two input symbols and three output symbols. The middle symbol y_2 represents error (e). This error is detected and *erased*. Hence this channel is also called Binary Erasure Channel (BEC).

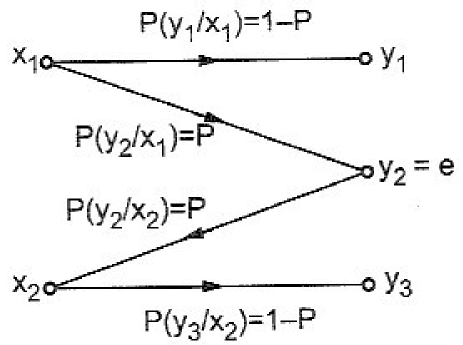


Fig. 1.9.3 Binary erasure channel

From above figure we can write the channel matrix for BEC as follows :

$$P(Y/X) = \begin{bmatrix} P(y_1/x_1) & P(y_2/x_1) & P(y_3/x_1) \\ P(y_1/x_2) & P(y_2/x_2) & P(y_3/x_2) \end{bmatrix}$$

$$= \begin{bmatrix} 1-p & p & 0 \\ 0 & p & 1-p \end{bmatrix}$$

Example 1.9.1 : *For the BEC of Fig. 1.9.3 calculate the probabilities associated with the channel outputs for p = 0.2, if the source outputs are equally likely.*

Solution : It is given that source emits x_1 and x_2 with equal probabilities. Hence,

$$P(x_1) = \frac{1}{2} \quad \text{and} \quad P(x_2) = \frac{1}{2}$$

The output probabilities are given as,

$$\begin{bmatrix} P(y_1) \\ P(y_2) \\ P(y_3) \end{bmatrix} = \begin{bmatrix} P(x_1) & P(x_2) \end{bmatrix} \begin{bmatrix} 1-p & p & 0 \\ 0 & p & 1-p \end{bmatrix}$$

It is given that p = 0.2. Putting values in above equation,

$$\begin{bmatrix} P(y_1) \\ P(y_2) \\ P(y_3) \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 0.8 & 0.2 & 0 \\ 0 & 0.2 & 0.8 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{2} \times 0.8 + 0 \\ \frac{1}{2} \times 0.2 + \frac{1}{2} \times 0.2 \\ \frac{1}{2} \times 0 + \frac{1}{2} \times 0.8 \end{bmatrix} = \begin{bmatrix} 0.4 \\ 0.2 \\ 0.4 \end{bmatrix}$$

Thus $P(y_1) = 0.4$, $P(y_2) = 0.2$ and $P(y_3) = 0.4$.

1.9.4 Equivocation (Conditional Entropy)

The conditional entropy $H(X / Y)$ is called equivocation. It is defined as,

$$H(X / Y) = \sum_{i=1}^{M} \sum_{j=1}^{M} P(x_i, y_j) \log_2 \frac{1}{P(x_i / y_j)} \quad \text{... (1.9.17)}$$

And the joint entropy $H(X, Y)$ is given as,

$$H(X, Y) = \sum_{i=1}^{M} \sum_{j=1}^{M} P(x_i, y_j) \log_2 \frac{1}{P(x_i \; y_j)} \quad \text{... (1.9.18)}$$

The conditional entropy $H(X / Y)$ represents uncertainty of X, on average, when Y is known. Similarly the conditional entropy $H(Y / X)$ represents uncertainty of Y, on average, when X is transmitted. $H(Y / X)$ can be given as,

$$H(Y / X) = \sum_{i=1}^{M} \sum_{j=1}^{M} P(x_i, y_j) \log_2 \frac{1}{P(y_j / x_i)} \quad \text{... (1.9.19)}$$

$$\sum_{j=1}^{M} P(x_i, y_j) = P(x_i)$$

Hence $H(X, Y)$ will be written as,

$$H(X,Y) = H(Y/X) - \sum_{i=1}^{M} P(x_i) \log_2 P(x_i)$$

$$= H(Y/X) + \sum_{i=1}^{M} P(x_i) \log_2 \frac{1}{P(x_i)}$$

As per the definition of entropy, the second term in the above equation is $H(X)$. Hence

$$H(X,Y) = H(Y/X) + H(X) \quad \text{... (1.9.22)}$$

Thus the second part of the given equation is proved.

Example 1.9.3 : *Two BSC's are connected in cascade as shown in Fig. 1.9.4.*

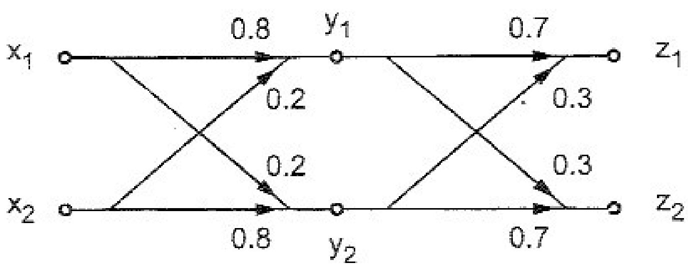


Fig. 1.9.4 BSC of Example 1.9.3

i) Find the channel matrix of resultant channel.

ii) Find $P(z_1)$ and $P(z_2)$ if $P(x_1) = 0.6$ and $P(x_2) = 0.4$

Solution : i) To obtain channel matrix using equation (1.9.1)

Here we can write two matrices of the channel as follows :

$$P(Y/X) = \begin{bmatrix} P(y_1/x_1) & P(y_2/x_1) \\ P(y_1/x_2) & P(y_2/x_2) \end{bmatrix} = \begin{bmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{bmatrix}$$

$$\text{and} \quad P(Z/Y) = \begin{bmatrix} P(z_1/y_1) & P(z_2/y_1) \\ P(z_1/y_2) & P(z_2/y_2) \end{bmatrix} = \begin{bmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{bmatrix}$$

Hence resultant channel matrix is given as,

$$P(Z/X) = P(Y/X) \cdot P(Z/Y)$$

$$= \begin{bmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{bmatrix} \begin{bmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{bmatrix}$$

$$= \begin{bmatrix} 0.62 & 0.38 \\ 0.38 & 0.62 \end{bmatrix}$$

ii) To obtain $P(z_1)$ and $P(z_2)$:

The probabilities of Z_1 and Z_2 are given as,

$$P(Z) = P(X)\ P(Z / X)$$

$$= [P(x_1)\ P(x_2)] \begin{bmatrix} 0.62 & 0.38 \\ 0.38 & 0.62 \end{bmatrix}$$

$$= [0.6\ \ 0.4] \begin{bmatrix} 0.62 & 0.38 \\ 0.38 & 0.62 \end{bmatrix}$$

$$= [0.524\ \ 0.476]$$

Thus $P(Z_1) = 0.524$ and

$P(Z_2) = 0.476$

1.9.5 Rate of Information Transmission over a Discrete Channel

The entropy of the symbol gives average amount of information going into the channel. i.e.,

$$H(X) = \sum_{i=1}^{M} p_i \log_2\left(\frac{1}{p_i}\right) \quad \text{... (1.9.23)}$$

Let the symbols be generated at the rate of *'r'* symbols per second. Then the average rate of information going into the channel is given as,

$$D_{in} = rH(X) \text{ bits /sec} \quad \text{... (1.9.24)}$$

Errors are introduced in the data during the transmission. Because of these errors, some information is lost in the channel. The conditional entropy $H(X / Y)$ is the measure of information lost in the channel. Hence the information transmitted over the channel will be,

$$\text{Transmitted information} = H(X) - H(X / Y) \quad \text{... (1.9.25)}$$

Hence the average rate of information transmission D_t across the channel will be,

$$D_t = [H(X) - H(X / Y)]r \text{ bits/sec} \quad \text{... (1.9.26)}$$

When the noise becomes very large, then X and Y become statistically independent. Then $H(X / Y) = H(X)$ and hence no information is transmitted over the channel. In case of

It is given as,

$$I(X;Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} P(x_i, y_j)\, I(x_i, y_j)$$

Thus $I(x_i, y_j)$ is weighted by joint probabilities $P(x_i, y_j)$ over all the possible joint events. Putting for $I(x_i, y_j)$ from equation 1.10.1 we get,

$$I(X;Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} P(x_i, y_j) \log_2 \frac{P(x_i / y_j)}{P(x_i)} \quad \text{... (1.10.2)}$$

1.10.1 Properties of Mutual Information

The mutual information has following properties :

i) The mutual information of the channel is symmetric. i.e.,

$$I(X;Y) = I(Y;X)$$

ii) The mutual information can be expressed in terms of entropies of channel input or output and conditional entropies. i.e,.

$$I(X;Y) = H(X) - H(X/Y)$$

$$= H(Y) - H(Y/X)$$

Here $H(X/Y)$ and $H(Y/X)$ are conditional entropies.

iii) The mutual information is always positive, i.e.,

$$I(X;Y) \geq 0$$

iv) The mutual information is related to the joint entropy $H(X, Y)$ by following relation :

$$I(X;Y) = H(X) + H(Y) - H(X,Y)$$

Example 1.10.1 : *Prove that the mutual information of the channel is symmetric i.e.,*

$$I(X;Y) = I(Y;X)$$

Solution : Let us consider some standard relationships from probability theory. These are as follows.

$$P(x_i\, y_j) = P(x_i / y_j)\, P(y_j) \quad \text{... (1.10.3)}$$

and

$$P(x_i\, y_j) = P(y_j / x_i)\, P(x_i) \quad \text{... (1.10.4)}$$

Here $P(x_i\, y_j)$ is the joint probability that x_i is transmitted and y_j is received.

$P(x_i / y_j)$ is the conditional probability of that x_i is transmitted and y_j is received.

$P(y_j / x_i)$ is the conditional probability that y_j is received and x_i is transmitted.

$P(x_i)$ is the probability of symbol x_i for transmission.

$P(y_j)$ is the probability that symbol y_j is received.

From equation 1.10.3 and equation 1.10.4 we can write,

$$P(x_i / y_j)\, P(y_j) = P(y_j / x_i)\, P(x_i)$$

$$\therefore \quad \frac{P(x_i / y_j)}{P(x_i)} = \frac{P(y_j / x_i)}{P(y_j)} \qquad \text{... (1.10.5)}$$

The average mutual information is given by equation 1.10.2 as,

$$I(X; Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} P(x_i, y_j) \log_2 \frac{P(x_i / y_j)}{P(x_i)} \qquad \text{... (1.10.6)}$$

Hence we can write $I(Y; X)$ as follows.

$$I(Y; X) = \sum_{i=1}^{n} \sum_{j=1}^{m} P(x_i, y_j) \log_2 \frac{P(y_j / x_i)}{P(y_j)} \qquad \text{... (1.10.7)}$$

From equation 1.10.5 the above equation can be written as,

$$I(Y; X) = \sum_{i=1}^{n} \sum_{j=1}^{m} P(x_i, y_j) \log_2 \frac{P(x_i / y_j)}{P(x_i)}$$

$$= I(X; Y) \text{ from equation 1.10.6}$$

Thus

$$\boxed{I(X; Y) = I(Y; X)} \qquad \text{... (1.10.8)}$$

Thus the mutual information of the discrete memoryless channel is symmetric.

Example 1.10.2 : *Prove the following relationships.*

$$I(X; Y) = H(X) - H(X / Y)$$

$$I(X; Y) = H(Y) - H(Y / X)$$

Solution : Here $H(X / Y)$ is the conditional entropy and it is given as,

$$H(X / Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} P(x_i, y_j) \log_2 \frac{1}{P(x_i / y_j)} \qquad \text{... (1.10.9)}$$

$H(X / Y)$ is the information or uncertainty in X after Y is received. In other words $H(X / Y)$ is the information lost in the noisy channel. It is the average conditional self information.

$$= \sum_{j=1}^{m} \sum_{i=1}^{n} P(x_i, y_j) \log_2 \frac{1}{P(y_j)}$$

$$- \sum_{i=1}^{n} \sum_{j=1}^{m} P(x_i, y_j) \log_2 \frac{1}{P(y_j / x_i)} \quad \text{... (1.10.14)}$$

The conditional entropy $H(Y / X)$ is given as,

$$H(Y / X) = \sum_{i=1}^{n} \sum_{j=1}^{m} P(x_i, y_j) \log_2 \frac{1}{P(y_j / x_i)} \quad \text{... (1.10.15)}$$

Here $H(Y / X)$ is the uncertainty in Y when X was transmitted. With this result, equation 1.10.14 becomes,

$$I(Y; X) = \sum_{j=1}^{m} \sum_{i=1}^{n} P(x_i, y_j) \log_2 \frac{1}{P(y_j)} - H(Y / X) \quad \text{... (1.10.16)}$$

Here let us use the standard probability equation,

$$\sum_{i=1}^{n} P(x_i, y_j) = P(y_j) \quad \text{... (1.10.17)}$$

Hence equation 1.10.16 becomes,

$$I(Y; X) = \sum_{j=1}^{n} P(y_j) \log_2 \frac{1}{P(y_j)} - H(Y / X)$$

The entropy is given by equation 1.4.6. Hence first term of above equation represents $H(Y)$. Hence above equation becomes,

$$\boxed{I(Y; X) = H(Y) - H(Y / X)} \quad \text{... (1.10.18)}$$

Note that the above result is similar to that of equation 1.10.13.

Example 1.10.3 : *Prove that the mutual information is always positive i.e.,*

$$I(X;Y) \geq 0$$

Solution : Mutual information is given by equation 1.10.2 as,

$$I(X;Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} P(x_i, y_j) \log_2 \frac{P(x_i / y_j)}{P(x_i)} \quad \text{... (1.10.19)}$$

From equation 1.10.3, $P(x_i / y_j) = \dfrac{P(x_i, y_j)}{P(y_j)}$

Putting above value of $P(x_i / y_j)$ in equation 1.10.19,

$$I(X;Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} P(x_i, y_j) \log_2 \frac{P(x_i, y_j)}{P(x_i)\, P(y_j)}$$

We know that $\log_2 \frac{x}{y}$ can be written as, $-\log_2 \frac{y}{x}$. Hence above equation becomes,

$$I(X;Y) = -\sum_{i=1}^{n} \sum_{j=1}^{m} P(x_i, y_j) \log_2 \frac{P(x_i)\, P(y_j)}{P(x_i, y_j)}$$

This equation can be written as,

$$-I(X;Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} P(x_i, y_j) \log_2 \frac{P(x_i)\, P(y_j)}{P(x_i, y_j)} \qquad \text{... (1.10.20)}$$

Earlier we have derived one result given by equation 1.4.11. It states that,

$$\sum_{k=1}^{m} p_k \log_2 \left(\frac{q_k}{p_k} \right) \leq 0$$

This result can be applied to equation 1.10.20. We can consider p_k be $P(x_i\ y_i)$ and q_k be $P(x_i)\ P(y_j)$. Both p_k and q_k are two probability distributions on same alphabet. Then equation 1.10.20 becomes,

$$-I(X;Y) \leq 0$$

i.e.,

$$I(X;Y) \geq 0 \qquad \text{... (1.10.21)}$$

The above equation is the required proof. It says that mutual information is always non negative.

Example 1.10.4 : *Prove the following,*

$$I(X\,;\,Y) = H(X) + H(Y) - H(X, Y)$$

Solution : In example 1.10.1 we have derived following relation :

$$H(X, Y) = H(X / Y) + H(Y)$$

$$\therefore \quad H(X / Y) = H(X, Y) - H(Y) \qquad \text{... (1.10.22)}$$

Mutual information is given by equation 1.10.13 also i.e.,

$$I(X\,;\,Y) = H(X) - H(X / Y)$$

Putting for $H(X / Y)$ in above equation from equation 1.10.22,

$$I(X\,;\,Y) = H(X) + H(Y) - H(X, Y) \qquad \text{... (1.10.23)}$$

Thus the required relation is proved.

$P(y_j / x_i)$ is the conditional probability that y_j is received and x_i is transmitted.

$P(x_i)$ is the probability of symbol x_i for transmission.

$P(y_j)$ is the probability that symbol y_j is received.

From equation 1.10.3 and equation 1.10.4 we can write,

$$P(x_i / y_j)\, P(y_j) = P(y_j / x_i)\, P(x_i)$$

$$\therefore \quad \frac{P(x_i / y_j)}{P(x_i)} = \frac{P(y_j / x_i)}{P(y_j)} \qquad \text{... (1.10.5)}$$

The average mutual information is given by equation 1.10.2 as,

$$I(X;Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} P(x_i, y_j) \log_2 \frac{P(x_i / y_j)}{P(x_i)} \qquad \text{... (1.10.6)}$$

Hence we can write $I(Y; X)$ as follows.

$$I(Y;X) = \sum_{i=1}^{n} \sum_{j=1}^{m} P(x_i, y_j) \log_2 \frac{P(y_j / x_i)}{P(y_j)} \qquad \text{... (1.10.7)}$$

From equation 1.10.5 the above equation can be written as,

$$I(Y;X) = \sum_{i=1}^{n} \sum_{j=1}^{m} P(x_i, y_j) \log_2 \frac{P(x_i / y_j)}{P(x_i)}$$

$$= I(X;Y) \text{ from equation 1.10.6}$$

Thus

$$\boxed{I(X;Y) = I(Y;X)} \qquad \text{... (1.10.8)}$$

Thus the mutual information of the discrete memoryless channel is symmetric.

Example 1.10.2 : *Prove the following relationships.*

$$I(X;Y) = H(X) - H(X/Y)$$

$$I(X;Y) = H(Y) - H(Y/X)$$

Solution : Here $H(X/Y)$ is the conditional entropy and it is given as,

$$H(X/Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} P(x_i, y_j) \log_2 \frac{1}{P(x_i / y_j)} \qquad \text{... (1.10.9)}$$

$H(X/Y)$ is the information or uncertainty in X after Y is received. In other words $H(X/Y)$ is the information lost in the noisy channel. It is the average conditional self information.

Consider the equation 1.10.2,

$$I(X;Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} P(x_i, y_j) \log_2 \frac{P(x_i / y_j)}{P(x_i)}$$

Let us write the above equation as,

$$I(X;Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} P(x_i, y_j) \log_2 \frac{1}{P(x_i)}$$

$$- \sum_{i=1}^{n} \sum_{j=1}^{m} P(x_i, y_j) \log_2 \frac{1}{P(x_i / y_j)}$$

From equation 1.10.9, above equation can be written as,

$$I(X;Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} P(x_i, y_j) \log_2 \frac{1}{P(x_i)} - H(X/Y) \quad \text{... (1.10.10)}$$

Here let us use the standard probability relation which is given as follows :

$$\sum_{j=1}^{m} P(x_i, y_j) = P(x_i)$$

Hence equation 1.10.10 will be,

$$I(X;Y) = \sum_{i=1}^{n} P(x_i) \log_2 \frac{1}{P(x_i)} - H(X/Y) \quad \text{... (1.10.11)}$$

First term of the above equation represents entropy. i.e.,

$$H(X) = \sum_{i=1}^{n} P(x_i) \log_2 \frac{1}{P(x_i)} \quad \text{... (1.10.12)}$$

Hence equation 1.10.11 becomes,

$$\boxed{I(X;Y) = H(X) - H(X/Y)} \quad \text{... (1.10.13)}$$

Here note that $I(X;Y)$ is the average information transferred per symbol across the channel. It is equal to source entropy minus information lost in the noisy channel is given by above equation.

Similarly consider the average mutual information given by equation 1.10.8,

$$I(Y;X) = \sum_{i=1}^{n} \sum_{j=1}^{m} P(x_i, y_j) \log_2 \frac{P(y_j / x_i)}{P(y_j)}$$

$$= \sum_{j=1}^{m} \sum_{i=1}^{n} P(x_i, y_j) \log_2 \frac{1}{P(y_j)}$$

$$-\sum_{i=1}^{n} \sum_{j=1}^{m} P(x_i, y_j) \log_2 \frac{1}{P(y_j / x_i)} \quad \text{... (1.10.14)}$$

The conditional entropy $H(Y / X)$ is given as,

$$H(Y / X) = \sum_{i=1}^{n} \sum_{j=1}^{m} P(x_i, y_j) \log_2 \frac{1}{P(y_j / x_i)} \quad \text{... (1.10.15)}$$

Here $H(Y / X)$ is the uncertainty in Y when X was transmitted. With this result, equation 1.10.14 becomes,

$$I(Y; X) = \sum_{j=1}^{m} \sum_{i=1}^{n} P(x_i, y_j) \log_2 \frac{1}{P(y_j)} - H(Y / X) \quad \text{... (1.10.16)}$$

Here let us use the standard probability equation,

$$\sum_{i=1}^{n} P(x_i, y_j) = P(y_j) \quad \text{... (1.10.17)}$$

Hence equation 1.10.16 becomes,

$$I(Y; X) = \sum_{j=1}^{n} P(y_j) \log_2 \frac{1}{P(y_j)} - H(Y / X)$$

The entropy is given by equation 1.4.6. Hence first term of above equation represents $H(Y)$. Hence above equation becomes,

$$\boxed{I(Y; X) = H(Y) - H(Y / X)} \quad \text{... (1.10.18)}$$

Note that the above result is similar to that of equation 1.10.13.

Example 1.10.3 : *Prove that the mutual information is always positive i.e.,*

$$I(X;Y) \geq 0$$

Solution : Mutual information is given by equation 1.10.2 as,

$$I(X;Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} P(x_i, y_j) \log_2 \frac{P(x_i / y_j)}{P(x_i)} \quad \text{... (1.10.19)}$$

From equation 1.10.3, $P(x_i / y_j) = \frac{P(x_i, y_j)}{P(y_j)}$

Example 1.10.5 : *Consider the binary symmetric channel shown in Fig. 1.10.1. Calculate H(X), H(Y), H(Y/X) and I(X; Y).*

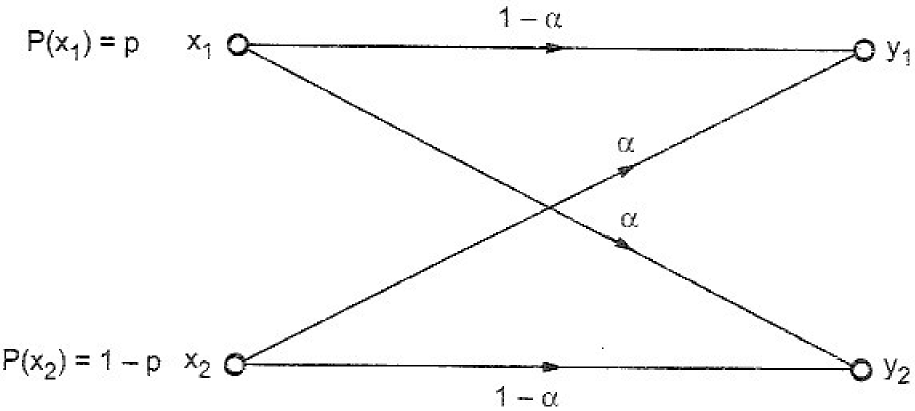


Fig. 1.10.1 Binary symmetric channel

Solution : Here for BSC, the various probabilities are as follows :

$$P(x_1) = p \quad \text{and} \quad P(x_2) = 1-p$$

$$P(y_1 / x_2) = P(y_2 / x_1) = \alpha$$

$$P(y_1 / x_1) = P(y_2 / x_2) = 1-\alpha$$

(i) To obtain H(X)

We know that X has two symbols x_1 and x_2. Their probabilities are p and $1-p$. Equation 1.4.7 gives the entropy for such source i.e.,

$$H(X) = p\log_2\left(\frac{1}{p}\right)+(1-p)\log_2\left(\frac{1}{1-p}\right) \qquad \text{... (1.10.24)}$$

This is also called as *Horseshoe* function and it is denoted by Ω ( ). i.e.

$$\Omega(p) = p\log_2\left(\frac{1}{p}\right)+(1-p)\log_2\left(\frac{1}{1-p}\right) \qquad \text{... (1.10.25)}$$

Thus equation 1.10.24 gives entropy of X.

(ii) To obtain H(Y)

The standard probability relation states that

$$P(y_j) = \sum_{i=1}^{n} P(x_i\ y_j) \qquad \text{... (1.10.26)}$$

Putting above value of $P(x_i / y_j)$ in equation 1.10.19,

$$I(X;Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} P(x_i, y_j) \log_2 \frac{P(x_i, y_j)}{P(x_i)\, P(y_j)}$$

We know that $\log_2 \frac{x}{y}$ can be written as, $-\log_2 \frac{y}{x}$. Hence above equation becomes,

$$I(X;Y) = -\sum_{i=1}^{n} \sum_{j=1}^{m} P(x_i, y_j) \log_2 \frac{P(x_i)\, P(y_j)}{P(x_i, y_j)}$$

This equation can be written as,

$$-I(X;Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} P(x_i, y_j) \log_2 \frac{P(x_i)\, P(y_j)}{P(x_i, y_j)} \quad \text{... (1.10.20)}$$

Earlier we have derived one result given by equation 1.4.11. It states that,

$$\sum_{k=1}^{m} p_k \log_2 \left(\frac{q_k}{p_k} \right) \leq 0$$

This result can be applied to equation 1.10.20. We can consider p_k be $P(x_i\ y_i)$ and q_k be $P(x_i)\ P(y_j)$. Both p_k and q_k are two probability distributions on same alphabet. Then equation 1.10.20 becomes,

$$-I(X;Y) \leq 0$$

i.e.,

$$I(X;Y) \geq 0 \quad \text{... (1.10.21)}$$

The above equation is the required proof. It says that mutual information is always non negative.

Example 1.10.4 : *Prove the following,*

$$I(X\,;Y) = H(X) + H(Y) - H(X, Y)$$

Solution : In example 1.10.1 we have derived following relation :

$$H(X, Y) = H(X / Y) + H(Y)$$

$$\therefore \quad H(X / Y) = H(X, Y) - H(Y) \quad \text{... (1.10.22)}$$

Mutual information is given by equation 1.10.13 also i.e.,

$$I(X\,;Y) = H(X) - H(X / Y)$$

Putting for $H(X / Y)$ in above equation from equation 1.10.22,

$$I(X\,;Y) = H(X) + H(Y) - H(X, Y) \quad \text{... (1.10.23)}$$

Thus the required relation is proved.

Hence equation 1.10.34 becomes,

$$I(X;Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} P(y_j / x_i)\ P(x_i)\ \log_2 \frac{P(y_j / x_i)}{\sum_{i=1}^{n} P(y_j / x_i)\ P(x_i)} \quad \text{... (1.10.35)}$$

In the above equation observe that, the mutual information an be obtained from transition probabilities $P(y_j / x_i)$ and $P(x_i)$. Transition probabilities $P(y_j / x_i)$ are the characteristic of the channel but $P(x_i)$ are independent of the channel.

The channel capacity of the discrete memoryless channel is given as maximum average mutual information. The maximization is taken with respect to input probabilities $P(x_i)$. i.e.,

$$C = \max_{P(x_i)} I(X;Y) \quad \text{... (1.10.36)}$$

Example 1.10.6 : *Find the mutual information and channel capacity for the channel shown in Fig. 1.10.2. Given that* $P(x_1) = 0.6$ *and* $P(x_2) = 0.4$.

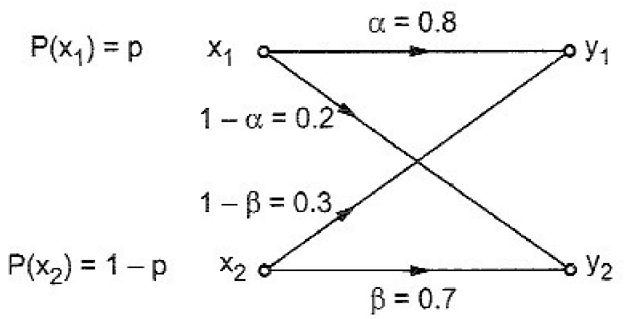


Fig. 1.10.2 Communication channel (Nonsymmetric binary channel)

Solution : Here note that the given channel is binary but not symmetric. Hence we have to derive equations for this channel using basic principles.

(i) To obtain source entropy $H(X)$:

Let $P(x_1) = p$. Then $P(x_2) = 1-p$. Here p = 0.6 given.

Entropy of the source is given as,

$$H(X) = p \log_2 \frac{1}{p} + (1-p) \log_2 \frac{1}{(1-p)}$$

$$= 0.6 \log_2 \frac{1}{0.6} + 0.4 \log_2 \frac{1}{0.4}$$

$$= 0.4422 + 0.5287 = 0.9709 \text{ bits/symbol}$$

(ii) To obtain $H(Y)$:

From equation 1.19.12 we can obtain probabilities of output symbols as,

$$\begin{bmatrix} P(y_1) \\ P(y_2) \end{bmatrix} = \begin{bmatrix} P(x_1) & P(x_2) \end{bmatrix} \begin{bmatrix} P(y_1/x_1) & P(y_2/x_1) \\ P(y_1/x_2) & P(y_2/x_2) \end{bmatrix}$$

Here $P(y_1/x_1) = \alpha, \quad P(y_2/x_1) = 1-\alpha$

$$P(y_1/x_2) = 1-\beta, \quad P(y_2/x_2) = \beta$$

$$\therefore \quad \begin{bmatrix} P(y_1) \\ P(y_2) \end{bmatrix} = \begin{bmatrix} p & (1-p) \end{bmatrix} \begin{bmatrix} \alpha & 1-\alpha \\ 1-\beta & \beta \end{bmatrix} = \begin{bmatrix} p\alpha + (1-p)(1-\beta) \\ p(1-\alpha) + (1-p)\beta \end{bmatrix}$$

Thus the two probabilities are,

$$P(y_1) = p\alpha + (1-p)(1-\beta) = (0.6)\ (0.8) + (0.4)\ (0.3) = 0.6$$

and $$P(y_2) = p(1-\alpha) + \beta\,(1-p) = (0.6)\ (0.2) + (0.4)\ (0.7) = 0.4$$

Now entropy of output symbols can be obtained as,

$$H(Y) = P(y_1)\log_2 \frac{1}{P(y_1)} + P(y_2)\log_2 \frac{1}{P(y_2)}$$

$$= 0.6\log_2 \frac{1}{0.6} + 0.4\log_2 \frac{1}{0.4} = 0.9709 \text{ bits/symbol}$$

Thus $$H(X) = H(Y)$$

(iii) To obtain $H(Y/X)$:

We know that $H(Y/X)$ is given as,

$$H(Y/X) = \sum_{i=1}^{n} \sum_{j=1}^{m} P(x_i\, y_j)\ \log_2 \frac{1}{P(y_j/x_i)}$$

We know that $P(x_i\, y_j) = P(y_j/x_i)\ P(x_i)$

$$P(x_1\, y_1) = P(y_1/x_1)\ P(x_1) = \alpha p$$

$$P(x_1\, y_2) = P(y_2/x_1)\ P(x_1) = (1-\alpha)\ p$$

$$P(x_2\, y_1) = P(y_1/x_2)\ P(x_2) = (1-\beta)\ (1-p)$$

$$P(x_2\, y_2) = P(y_2/x_2)\ P(x_2) = \beta(1-p)$$

(ii) To obtain $H(Y)$:

From equation 1.19.12 we can obtain probabilities of output symbols as,

$$\begin{bmatrix} P(y_1) \\ P(y_2) \end{bmatrix} = \begin{bmatrix} P(x_1) & P(x_2) \end{bmatrix} \begin{bmatrix} P(y_1/x_1) & P(y_2/x_1) \\ P(y_1/x_2) & P(y_2/x_2) \end{bmatrix}$$

Here $P(y_1/x_1) = \alpha, \quad P(y_2/x_1) = 1-\alpha$

$$P(y_1/x_2) = 1-\beta, \quad P(y_2/x_2) = \beta$$

$$\therefore \quad \begin{bmatrix} P(y_1) \\ P(y_2) \end{bmatrix} = \begin{bmatrix} p & (1-p) \end{bmatrix} \begin{bmatrix} \alpha & 1-\alpha \\ 1-\beta & \beta \end{bmatrix} = \begin{bmatrix} p\alpha+(1-p)(1-\beta) \\ p(1-\alpha)+(1-p)\beta \end{bmatrix}$$

Thus the two probabilities are,

$$P(y_1) = p\alpha+(1-p)(1-\beta) = (0.6)\ (0.8) + (0.4)\ (0.3) = 0.6$$

and $$P(y_2) = p(1-\alpha)+\beta\,(1-p) = (0.6)\ (0.2) + (0.4)\ (0.7) = 0.4$$

Now entropy of output symbols can be obtained as,

$$H(Y) = P(y_1)\log_2 \frac{1}{P(y_1)} + P(y_2)\log_2 \frac{1}{P(y_2)}$$

$$= 0.6\log_2 \frac{1}{0.6} + 0.4\log_2 \frac{1}{0.4} = 0.9709 \text{ bits/symbol}$$

Thus $$H(X) = H(Y)$$

(iii) To obtain $H(Y/X)$:

We know that $H(Y/X)$ is given as,

$$H(Y/X) = \sum_{i=1}^{n} \sum_{j=1}^{m} P(x_i y_j) \log_2 \frac{1}{P(y_j/x_i)}$$

We know that $P(x_i y_j) = P(y_j/x_i)\ P(x_i)$

$$P(x_1 y_1) = P(y_1/x_1)\ P(x_1) = \alpha p$$

$$P(x_1 y_2) = P(y_2/x_1)\ P(x_1) = (1-\alpha)\ p$$

$$P(x_2 y_1) = P(y_1/x_2)\ P(x_2) = (1-\beta)\ (1-p)$$

$$P(x_2 y_2) = P(y_2/x_2)\ P(x_2) = \beta(1-p)$$

Hence $H(Y/X)$ can be written as,

$$H(Y/X) = P(x_1 y_1) \log_2 \frac{1}{P(y_1/x_1)} + P(x_1 y_2) \log_2 \frac{1}{P(y_2/x_1)}$$

$$+ P(x_2 y_1) \log_2 \frac{1}{P(y_1/x_2)} + P(x_2 y_2) \log_2 \frac{1}{p(y_2/x_2)}$$

$$= \alpha p \log_2 \frac{1}{\alpha} + (1-\alpha)\, p \log_2 \frac{1}{1-\alpha} + (1-\beta)(1-p)$$

$$\log_2 \frac{1}{1-\beta} + \beta(1-p) \log_2 \frac{1}{\beta} \quad \text{... (1.10.37)}$$

Putting values in above equation,

$$H(Y/X) = (0.8)(0.6) \log_2 \frac{1}{0.8} + (0.2)(0.6) \log_2 \frac{1}{0.2} + (0.3)(0.4) \log_2 \frac{1}{0.3}$$

$$+ (0.7)(0.4) \log_2 \frac{1}{0.7}$$

$$= 0.1545 + 0.2786 + 0.2084 + 0.1441 = 0.7855 \text{ bits/symbol}$$

(iv) To obtain mutual information :

Mutual information is given by equation 1.10.18 as,

$$I(Y;X) = H(Y) - H(Y/X) = 0.9709 - 0.7855 = 0.1853 \text{ bits/symbol.}$$

(v) To obtain channel capacity :

Let us write generalized equations for $H(Y)$ and $H(Y/X)$. In part (i) we have obtained $P(y_1)$ and $P(y_2)$ as,

$$P(y_1) = p\alpha + (1-p)(1-\beta)$$

$$P(y_2) = p(1-\alpha) + \beta(1-p)$$

$$\therefore \quad H(Y) = P(y_1) \log_2 \frac{1}{P(y_1)} + P(y_2) \log_2 \frac{1}{P(y_2)}$$

$$= [p\alpha + (1-p)(1-\beta)] \log_2 \frac{1}{[p\alpha + (1-p)(1-\beta)]}$$

$$+ [p(1-\alpha) + \beta(1-p)] \log_2 \frac{1}{[p(1-\alpha) + \beta(1-p)]}$$

Hence mutual information is given as,

$$I(Y;X) = H(Y) - H(Y/X)$$

Putting expressions of $H(Y)$ and $H(Y/X)$ in above equation,

$$\begin{aligned} I(Y;X) = & \left[p\alpha+(1-p)(1-\beta)\right]\log_2 \frac{1}{\left[p\alpha+(1-p)(1-\beta)\right]} \\ & +\left[p(1-\alpha)+\beta(1-p)\right]\log_2 \frac{1}{\left[p(1-\alpha)+\beta(1-p)\right]} \\ & -\alpha p \log_2 \frac{1}{\alpha}-(1-\alpha)p \log_2 \frac{1}{1-\alpha} \\ & -(1-\beta)(1-p)\log_2 \frac{1}{1-\beta}-\beta(1-p)\log_2 \frac{1}{\beta} \end{aligned} \quad \text{... (1.10.38)}$$

For channel capacity, maximum value of $I(Y;X)$ is obtained for p. We have to obtain the value of 'p', which maximizes $I(Y;X)$. Hence above equation is differentiated with respect to 'p' and equated to zero. i.e.,

$$\frac{dI(Y;X)}{dp} = 0$$

Solving above equation we get following value of 'p'. i.e.,

$$p = \frac{\beta\, 2^{\left(\frac{d}{\alpha+\beta-1}\right)}+\beta-1}{\left(1+2^{\frac{d}{\alpha+\beta-1}}\right)(\alpha+\beta-1)} \quad \text{... (1.10.39)}$$

where, $$d = -\alpha\log_2 \frac{1}{\alpha}-(1-\alpha)\log_2 \frac{1}{1-\alpha}+(1-\beta)\log_2 \frac{1}{1-\beta}+\beta \log_2 \frac{1}{\beta}$$

Here it is given that $\alpha = 0.8$ and $\beta = 0.7$, then above equation becomes,

$$\begin{aligned} d &= -(0.8)\log_2 \frac{1}{0.8}-(0.2)\log_2 \frac{1}{0.2}+(0.3)\log_2 \frac{1}{0.3} + 0.7 \log_2 \frac{1}{0.7} \\ &= -0.2575 - 0.4643 + 0.521 + 0.3602 \\ &= 0.1594 \end{aligned}$$

Hence value of 'p' becomes,

$$p = \frac{(0.7)2^{\left(\frac{0.1594}{0.8+0.7-1}\right)}+0.7-1}{\left(1+2^{\left(\frac{0.1594}{0.8+0.7-1}\right)}\right)(0.8+0.7-1)}$$

$$= \frac{0.5731}{1.12364} = 0.51$$

Thus the value of p = 0.51 will maximize the channel capacity. i.e.,

$$C = I(Y;X)\Big|_{p=0.6367}$$

Putting values in equation 1.10.38,

$$C = [(0.51)(0.8)+(0.49)(0.3)] \log_2 \frac{1}{(0.51)(0.8)+(0.49)(0.3)}$$

$$+ [(0.51)(0.2)+(0.7)(0.49)] \log_2 \frac{1}{(0.51)(0.2)+(0.7)(0.49)}$$

$$-(0.8)(0.51)\log_2 \frac{1}{0.8} - (0.2)(0.51)\log_2 \frac{1}{0.2}$$

$$-(0.3)(0.49)\log_2 \frac{1}{0.3} - (0.7)(0.49)\log_2 \frac{1}{0.7}$$

= 0.4714 + 0.5198 – 0.1313 – 0.2368 – 0.2553 – 0.1764

= 0.1913

Example 1.10.7 : *A binary channel matrix is given as,*

$$\begin{array}{ccc} & y_1 & y_2 & \leftarrow \textit{outputs} \\ \textit{Inputs} \begin{cases} x_1 \\ x_2 \end{cases} & \multicolumn{2}{c}{\begin{bmatrix} 2/3 & 1/3 \\ 1/10 & 9/10 \end{bmatrix}} \end{array}$$

Determine $H(X)$, $H(X/Y)$, $H(Y/X)$ *and mutual information* $I(X;Y)$.

Solution : (i) Given data :

Here source probabilities are not given. Hence let us assume probabilities of x_1 and x_2 as,

$$P(x_1) = \frac{1}{3} \quad \text{and} \quad P(x_2) = \frac{2}{3}$$

The channel matrix can be compared with the standard matrix. i.e.,

$$P = \begin{bmatrix} P(y_1/x_1) & P(y_2/x_1) \\ P(y_1/x_2) & P(y_2/x_2) \end{bmatrix} = \begin{bmatrix} 2/3 & 1/3 \\ 1/10 & 9/10 \end{bmatrix}$$

From above matrix, it is clear that $P(y_1/x_1) \neq P(y_2/x_2)$, hence this is nonsymmetric binary channel.

(ii) To obtain input symbol entropy $H(X)$:

Entropy is given as,

$$H(X) = P(x_1)\log_2\frac{1}{P(x_1)} + P(x_2)\log_2\frac{1}{P(x_2)}$$

$$= \frac{1}{3}\log_2 3 + \frac{2}{3}\log_2\frac{3}{2} = 0.9182 \text{ bits/symbol}$$

(iii) To obtain output symbol entropy $H(Y)$:

From equation 1.9.12 we can obtain probabilities of output symbols as,

$$\begin{bmatrix} P(y_1) \\ P(y_2) \end{bmatrix} = \begin{bmatrix} P(x_1) & P(x_2) \end{bmatrix} \begin{bmatrix} P(y_1/x_1) & P(y_2/x_1) \\ P(y_1/x_2) & P(y_2/x_2) \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{3} & \frac{2}{3} \end{bmatrix} \begin{bmatrix} 2/3 & 1/3 \\ 1/10 & 9/10 \end{bmatrix}$$

$$= \begin{bmatrix} 0.2889 \\ 0.7111 \end{bmatrix}$$

Thus $P(y_1) = 0.2889$ and $P(y_2) = 0.7111$

Entropy of Y can be calculated as,

$$H(Y) = P(y_1)\log_2\frac{1}{P(y_1)} + P(y_2)\log_2\frac{1}{P(y_2)}$$

$$= 0.2889\log_2\frac{1}{0.2889} + 0.7111\log_2\frac{1}{0.7111}$$

$$= 0.8672 \text{ bits/symbol}$$

(iv) To calculate joint entropy $H(X,Y)$:

For joint entropy we require joint probabilities $P(x_i, y_j)$. We know that,

$$P(x_i y_j) = P(y_j/x_i)\,P(x_i)$$

$$\therefore \quad P(x_1 y_1) = P(y_1/x_1)P(x_1) = \frac{2}{3}\times\frac{1}{3} = \frac{2}{9}$$

$$P(x_1 y_2) = P(y_2/x_1)\,P(x_1) = \frac{1}{3}\times\frac{1}{3} = \frac{1}{9}$$

$$P(x_2 y_1) = P(y_1/x_2)\,P(x_2) = \frac{1}{10}\times\frac{2}{3} = \frac{2}{30}$$

$$P(x_2 y_2) = P(y_2 / x_2)\ P(x_2) = \frac{9}{19} \times \frac{2}{3} = \frac{18}{30}$$

Joint entropy is given as,

$$H(X, Y) = \sum_{i=1}^{M} \sum_{j=1}^{M} P(x_i y_j) \log_2 \frac{1}{P(x_i y_j)}$$

$$= P(x_1 y_1) \log_2 \frac{1}{P(x_1 y_1)} + P(x_1 y_2) \log_2 \frac{1}{P(x_1 y_2)}$$

$$+ P(x_2 y_1) \log_2 \frac{1}{P(x_2 y_1)} + P(x_2 y_2) \log_2 \frac{1}{P(x_2 y_2)}$$

$$= \frac{2}{9} \log_2 \frac{9}{2} + \frac{1}{9} \log_2 9 + \frac{2}{30} \log_2 \frac{30}{2} + \frac{18}{30} \log_2 \frac{30}{18}$$

$$= 1.5365 \ \text{bits/symbol}$$

(v) To obtain conditional entropies $H(X/Y)$ and $H(Y/X)$:

$H(Y/X)$ is given as,

$$H(Y/X) = H(X,Y) - H(X) = 1.5365 - 0.9182$$

$$= 0.6183 \ \text{bits/symbol}$$

and $H(X/Y)$ is given as,

$$H(X/Y) = H(X,Y) - H(Y) = 1.5365 - 0.8672$$

$$= 0.6692 \ \text{bits/symbol}$$

(vi) To obtain mutual information $I(X;Y)$:

Mutual information is given as,

$$I(X;Y) = H(X) - H(X/Y) = 0.9182 - 0.6692$$

$$= 0.249 \ \text{bits/symbol}$$

Also $$I(X;Y) = H(Y) - H(Y/X) = 0.8672 - 0.6183$$

$$= 0.249 \ \text{bits/symbol}$$

Thus both the equations have same result.