# 1 Channel model

The discrete communication channels can be modeled by representing them by a set of input alphabets $X = \{x_1, x_2, \cdots, x_i\}$ consisting of '$i$' symbols, a set of output alphabets $Y = \{y_1, y_2, \cdots, y_j\}$ consisting of '$j$' symbols and a set of conditional probabilities $P(y_j/x_i)$. In other words, a *Discrete Memoryless Channel* ($DMC$) is completly described by the set of transition probabilities $P(y_j/x_i)$, where $x_i$ denotes input symbol, $y_j$ denotes output symbol and $P(y_j/x_i)$ denotes the probabilities of receiving symbol $y_j$ given that symbol $x_i$ was sent Fig. 1.

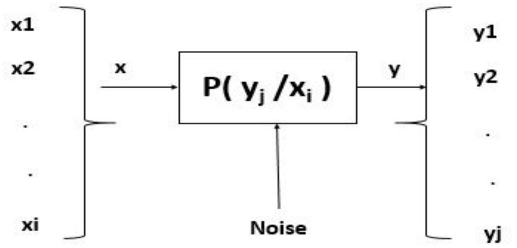


Figure 1: A discrete communication channel

## 1.1 Channel Matrix(Probability Transition Matrix)

The conditional probabilities $P(y_j/x_i)$ are due to channel impairments or channel noise that modifies the input symbols $x_i$ into $y_j$, resulting in error at the output of the channel $e$. This discrete-input, discrete-output channel is characterized by the set $X = \{x_1, x_2, \cdots, x_i\}$ of the possible inputs, the set $Y = \{y_1, y_2, \cdots, y_j\}$ of the possible outputs and a set of conditional probabilities that relate the possible outputs to the possible inputs. The conditional probability $P(y_j/x_i)$ is defined as the channel transition probability,denoted by $P_{ji}$. These conditional probabilities can be represented in the form of matrix with all input symbols represented row-wise and output symbols column-wise; such a matrix is known as *Probability Transition Matrix* ($PTM$) or simply *channel matrix* (or *noise matrix*), i.e.,

$$[P(\tfrac{y_j}{x_i})] = \begin{array}{c} \\ x_1 \\ x_2 \\ \vdots \\ x_n \end{array} \begin{array}{c} \begin{array}{cccc} y_1 & y_2 & \cdots & y_m \end{array} \\ \begin{pmatrix} P(\frac{y_1}{x_1}) & P(\frac{y_2}{x_1}) & \dots & P(\frac{y_m}{x_1}) \\ P(\frac{y_1}{x_2}) & P(\frac{y_2}{x_2}) & \dots & P(\frac{y_m}{x_2}) \\ \vdots & \vdots & \dots & \vdots \\ P(\frac{y_1}{x_n}) & P(\frac{y_2}{x_n}) & \dots & P(\frac{y_m}{x_n}) \end{pmatrix} \end{array}$$

Let us consider a *Discrete Memoryless Source* ($DMS$) having an input alphabet $X = \{x_1, x_2, x_3\}$ and an output alphabet $Y = \{y_1, y_2, y_3, y_4\}$. The various transition probabilities are marked on a channel diagram as show in Fig. 2.

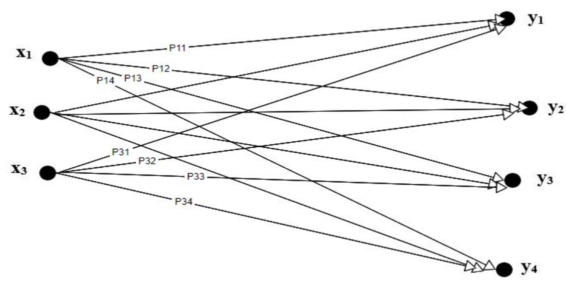


Figure 2: Channel diagram

In Fig. 2, $P_{11}$ is the probability that $y_1$ will be received when $x_1$ was transmitted (because of channel noise), that is,

$$P_{11} = P(\frac{y_1}{x_1}); P_{12} = P(\frac{y_2}{x_1}); P_{34} = P(\frac{y_4}{x_3})$$

Therefore,

$$[P(\frac{Y}{X})] = \begin{array}{c|cccc} & y_1 & y_2 & y_3 & y_4 \\ x_1 & P_{11} & P_{12} & P_{13} & P_{14} \\ x_2 & P_{21} & P_{22} & P_{23} & P_{24} \\ x_3 & P_{31} & P_{32} & P_{33} & P_{34} \end{array}$$

Since each input to the channel results in some output,each row of the channel matrix must be unity, i.e.,

$$\sum_{j=1}^{n} P(\frac{y_j}{x_i}) = 1 \qquad \forall i \tag{1}$$

Further, the input probabilities $[P(x)]$ are represented by the row matrix

$$[P(x_i)] = [P(x_1), P(x_2), \cdots, P(x_n)], i = 1, 2, 3, \cdots, n$$

and the output probabilities $[P(y)]$ are represented by the row matrix

$$[P(y_j)] = [P(y_1), P(y_2), \cdots, P(y_m)], j = 1, 2, 3, \cdots, m$$

From probability theory, we know that

$$P(AB) = P(\frac{B}{A}).P(A)$$

Here, $P(AB)$ is the joint probability of $A$ and $B$. In this case, if we let $A = x_i$ and $B = y_j$, then

$$P(x_i, y_j) = P(\frac{y_j}{x_i}).P(x_i) \tag{2}$$

Here, $P(x_i, y_j)$ is the joint probability of $x_i$ and $y_j$. If we add all the joint probabilities for a fixed $y_j$, then we get $P(y_j)$, that is,

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

This gives the probability of getting the symbol $y_j$, as follows:

$$P(y_j) = \sum_{i=1}^{n} P(\frac{y_j}{x_i}).P(x_i), j = 1, 2, ...., m \tag{3}$$

Suppose that $j^{th}$ symbol is received when $i^{th}$ symbol is transmitted; an error will result in this case. Hence, probability of error (or error probability) $P_e$ can be written as

$$P_e = \sum_{j=1, j\neq 1}^{m} P(y_j) \tag{4}$$

that is, all the probabilities will contribute to the error except $i = j$ (as in this case correct symbol is received):

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

## 1.2 Joint Probability Matrix

We already know that if $x_i$ are the inputs and $y_j$ are the outputs, then the joint probability is given by

$$P(x_i, y_j) = P(\frac{y_j}{x_i}).P(x_i) = P(\frac{x_i}{y_j}).P(y_j) \tag{5}$$

and also we defined the $PTM$ as

$$[P(\frac{y_j}{x_i})] = \begin{bmatrix} P(\frac{y_1}{x_1}) & P(\frac{y_2}{x_1}) & \ldots & P(\frac{y_m}{x_1}) \\ P(\frac{y_1}{x_2}) & P(\frac{y_2}{x_2}) & \ldots & P(\frac{y_m}{x_2}) \\ \vdots & \vdots & \ldots & \vdots \\ P(\frac{y_1}{x_n}) & P(\frac{y_2}{x_n}) & \ldots & P(\frac{y_m}{x_n}) \end{bmatrix}$$

Multiplying first row of $P(\frac{y_j}{x_i})$ by $P(x_1)$, second row by $P(x_2)$ and so on, we get

$$[P(\frac{y_j}{x_i})].P(x_i) = \begin{bmatrix} P(\frac{y_1}{x_1}).P(x_1) & P(\frac{y_2}{x_1}).P(x_1) & \dots & P(\frac{y_m}{x_1}).P(x_1) \\ P(\frac{y_1}{x_2}).P(x_2) & P(\frac{y_2}{x_2}).P(x_2) & \dots & P(\frac{y_m}{x_2}).P(x_2) \\ \vdots & \vdots & \dots & \vdots \\ P(\frac{y_1}{x_n}).P(x_n) & P(\frac{y_2}{x_n}).P(x_n) & \dots & P(\frac{y_m}{x_n}).P(x_n) \end{bmatrix}$$

But we know that $P(\frac{y_j}{x_i}).P(x_i) = P(x_i, y_j)$. Thus, we get

$$[P(x_i, y_j)] = \begin{array}{c} \\ i=1 \\ i=2 \\ \vdots \\ i=n \end{array} \begin{array}{c} \begin{array}{cccc} j=1 & j=2 & \cdots & j=m \end{array} \\ \begin{pmatrix} P(x_1, y_1) & P(x_1, y_2) & \dots & P(x_1, y_m) \\ P(x_2, y_1) & P(x_2, y_2) & \dots & P(x_2, y_m) \\ \vdots & \vdots & \dots & \vdots \\ P(x_n, y_1) & P(x_n, y_2) & \dots & P(x_n, y_m) \end{pmatrix} \end{array} \tag{6}$$

This is the *Joint Probability Matrix* ($JPM$) which has the following properties:

1. $\sum_{i=1}^{n} P(x_i, y_j) = P(y_j)$, that is, the sum of all the elements of the $j^{th}$ column of $JPM$ gives the probability of the $j^{th}$ output-probability of all output symbols.

2. $P(x_i) = \sum_{j=1}^{m} P(\frac{x_i}{y_j}).P(y_j)$, $i = 1, 2, 3, \cdots, n$, that is,

$$P(x_i) = \sum_{j=1}^{m} P(x_i, y_j), \quad i = 1, 2, 3, \cdots, n$$

since, $P(\frac{x_i}{y_j}).P(y_j) = P(x_i, y_j)$ (i.e., probability of the $i^{th}$ input can be obtained by adding all the elements in the ith row of $JPM$-probability of input symbols).

3. $\sum_{i=1}^{n} P(x_i)$.
Hence,

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

i.e., the sum of all the elements of $JPM$ is equal to unity.

### 1.2.1 Binary Symmetric Channel

Let us consider a binary information channel, meaning a discrete channel where there are two input symbols $x_0$ and $x_1$ and two output symbols $y_0$ and $y_1$ as shown in Fig. 3 along with associated probabilities. The output

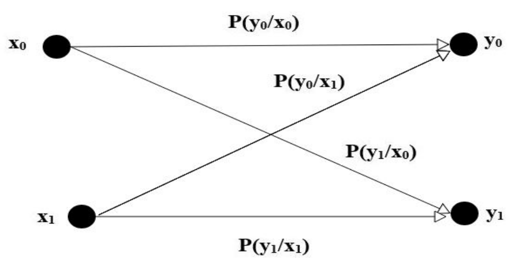


Figure 3: Binary information channel

symbol probabilities are given by

$$P(y_0) = P(\frac{y_0}{x_0}).P(x_0) + P(\frac{y_0}{x_1}).P(x_1)$$

$$P(y_1) = P(\frac{y_1}{x_1}).P(x_1) + P(\frac{y_1}{x_0}).P(x_0) \tag{7}$$

In matrix form, these give the PTM for a binary symmetric channel(BSC):

$$\begin{bmatrix} P(y_0) \\ P(y_1) \end{bmatrix} = [P(x_0)P(x_1)] \begin{bmatrix} P(\frac{y_0}{x_0}) & P(\frac{y_1}{x_0}) \\ P(\frac{y_0}{x_1}) & P(\frac{y_1}{x_1}) \end{bmatrix}$$

The binary information channel, shown in Fig. 3, is said to be symmetric if

$$P(\frac{y_0}{x_0}) = P(\frac{y_1}{x_1}) = P$$

and

$$P(\frac{y_0}{x_1}) = P(\frac{y_1}{x_0}) = 1 - P$$

Then the channel diagram for $BSC$ can be shown in Fig. 4. We get

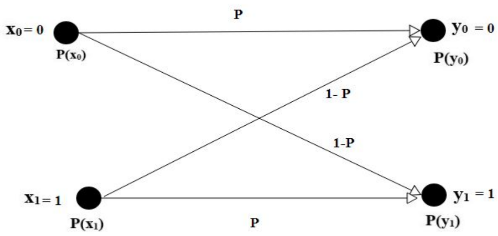


Figure 4: Binary Symmetric channel

$$\begin{bmatrix} P(y_0) \\ P(y_1) \end{bmatrix} = [P(x_0)P(x_1)] \begin{bmatrix} P & 1-P \\ 1-P & P \end{bmatrix}$$

The $BSC$ has two inputs ($x_0 = 0, x_1 = 1$) and two outputs ($y_0 = 0, y_1 = 1$). This is a symmetric channel as the probability of receiving a '0' when a '0' is sent is the same as the probability of receiving a '1' if a '1' is sent, that is $P$, which is the common transition probability.
Now $(1 - P)$ is naturally the error probability because a '1' is received when a '0' is transmitted and similarly a '0' is received when '1' is transmitted. If $P = \frac{1}{2}$, $BSC$ channel diagram will be as shown in Fig. 5.
We calculate the output probabilities, joint probabilities and conditional probabilities for the $BSC$.

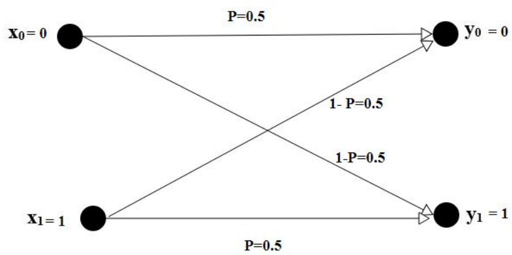


Figure 5: Binary Symmetric channel with P=0.5

## 2 System Entropies

The input $X$ and output $Y$ of a channel are sources with their own entropies, $H(X)$ and $H(Y)$, as shown in Fig. 6.

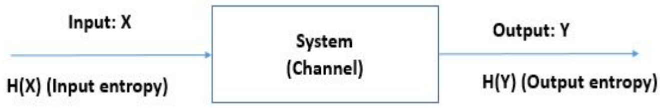


Figure 6: System entropies

The input entropy $H(X)$ is given by

$$H(X) = \sum_{i=1}^{n} P(x_i).log_2 \frac{1}{P(x_i)} = -\sum_{i=1}^{n} P(x_i).log_2(P(x_i)) bits/symbol \tag{8}$$

The output entropy $H(Y)$ is given by

$$H(Y) = -\sum_{j=1}^{m} P(y_j).log_2(P(y_j))bits/symbol \tag{9}$$

These source entropies $H(X)$ and $H(Y)$ represent average amounts of information going into and coming out of system (channel) per symbol. Equivalently, these are the average uncertainty about the input and output. The entropy of the input symbols before transmission is generally known as 'priory entropy'. The entropy of the input symbols after transmission and reception of a particular symbol, say, $Y_j$, is defined as the posteriori or conditional entropy denoted by $H(X/Y_j)$, and is given by

$$H(X/Y_j) = \sum_i P(\frac{x_i}{y_j})log(\frac{1}{P(x_i/y_j)}) \tag{10}$$

This represents the uncertainty about $X$ given that $Y_j$ is received or equivalently how much more information one would gain by knowing $X$. Averaging over all $Y_j$,we get the equivocation of $X$ with respect to $Y$ denoted by $H(X/Y)$, that is,

$$H(\frac{X}{Y}) = E[H(\frac{X}{Y_j})] = \sum_{j=1}^{m} P(y_j).H(\frac{X}{Y_j}) = \sum_{j=1}^{m} P(y_j)\sum_{i=1}^{n} P(\frac{x_i}{y_j}).log(\frac{1}{P(x_i/y_j)})$$

That is,

$$H(\frac{X}{Y}) = \sum_j\sum_i P(y_j).P(\frac{x_i}{y_j}).log_2\frac{1}{P(x_i/y_j)}$$

$$\implies H(X/Y) = \sum_j\sum_i P(x_i, y_j)log_2\frac{1}{P(x_i/y_j)}bits/symbol \tag{11}$$

$H(X/Y)$ that is, equivocation,specifies the receivers average uncertainty about $X$ when receiving $Y$. That is to say, it specifies the average uncertainty about the input of the channel being channel output is known. It represents the amount of information lost due to noise (Inter Symbol Interference ($ISI$)) with respect to any of the output symbols.
Similarly, if $x_i$ is sent, then the uncertainity about $Y$ is the conditional entropy. In the above equation for $H(X/Y)$, if $x$ and $y$ are interchanged, we get

$$H(\frac{Y}{X}) = \sum_j\sum_i P(y_j, x_i).log_2\frac{1}{P(y_j/x_i)}$$

But

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

Therefore,

$$H(\frac{Y}{X}) = \sum_j\sum_i P(x_i, y_j).log_2\frac{1}{P(y_j/x_i)} \tag{12}$$

$H(Y/X)$ is the equivocation of $Y$ with respect to $X$ representing the senders average uncertainty about $Y$ when $X$ is known or equivalently how much extra information is gained by also knowing $Y$.
An observer trying to guess both the input of the channel $X$ and its output $Y$ of the channel will have an average uncertainty given by the joint entropy:

$$H(\frac{X}{Y}) = \sum_i\sum_j P(x_i, y_j).log_2\frac{1}{P(x_i, y_j)} = -\sum_i\sum_j P(x_i, y_j).log_2 P(x_i, y_j)$$

We know that

$$P(A, B) = P(\frac{A}{B}).P(B)$$

Therefore,

$$P(x_i, y_j) = P(\frac{x_i}{y_j}).P(y_j)$$

Hence,

$$H(X, Y) = -\sum_i\sum_j P(x_i, y_j).log_2[P(\frac{x_i}{y_j}).P(y_j)]$$

Now

$$log_2[P(\frac{x_i}{y_j}).P(y_j)] = log_2P(\frac{x_i}{y_j}) + log_2P(y_j)$$

Therefore,

$$\begin{aligned}
H(X,Y) &= -\sum_i\sum_j P(x_i,y_j).log_2[P(\frac{x_i}{y_j})] - \sum_i\sum_j P(x_i,y_j).log_2P(y_j) \\
&= \sum_i\sum_j P(x_i,y_j).\frac{1}{log_2P(x_i/y_j)} - \sum_j\{\sum_i P(x_i,y_j)\}log_2P(y_j) \\
&= H(\frac{X}{Y}) - \sum_i P(y_j).log_2P(y_i)
\end{aligned}$$

Hence, First term $= H(\frac{X}{Y})$ and $\sum_i P(x_i,y_j) = P(y_j)$ We get

$$H(X,Y) = H(\frac{X}{Y}) + H(Y) \tag{13}$$

Similarly,it can be shown that

$$H(X,Y) = H(\frac{Y}{X}) + H(X)$$

We know that

$$P(x_i,y_j) = P(\frac{y_j}{x_i})P(x_i)$$

Therefore,

$$\begin{aligned}
H(X,Y) &= -\sum_i\sum_j P(x_i,y_j).log_2[P(\frac{y_j}{x_i}).P(x_i)] \\
&= -\sum_i\sum_j P(x_i,y_j).log_2[P(\frac{y_j}{x_i})] - \sum_i\sum_j P(x_i,y_j).log_2P(x_i) \\
&= \sum_i\sum_j P(x_i,y_j).\frac{1}{log_2P(y_j/x_i)} - \sum_i\{\sum_j P(x_i,y_j)\}log_2P(x_i) \\
&= H(\frac{Y}{X}) - \sum_i P(x_i).log_2P(x_i)
\end{aligned}$$

Hence, First term $= H(\frac{Y}{X})$ and $\sum_j P(x_i,y_j) = P(x_i)$. We get,
$H(X,Y) = H(\frac{Y}{X}) + \sum_i P(x_i)log_2(\frac{1}{P(x_i)})$
Therefore,

$$H(X,Y) = H(\frac{Y}{X}) + H(X) \tag{14}$$

If the channel is such that $X$ and $Y$ are statistically independent, then

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

But in real life,one would expect $X$ and $Y$ to be related rather than independent, in which case

$$H(X,Y) = H(\frac{Y}{X}) + H(X)$$

This relation that we have already proved confirms the interpretation of $H(Y/X)$ as the extra information conveyed by $Y$ when $X$ is already known. It is analogous to the rule $\mid X \cup Y \mid = \mid X \mid + \mid Y/X \mid$ for finite sets $X$ and $Y$. Similarly,

$$H(\frac{X}{Y}) = H(Y) + H(\frac{X}{Y})$$

which corresponds to $\mid X \cup Y \mid = \mid Y \mid + \mid X/Y \mid$. We call $H(X)$, $H(Y)$, $H(X/Y)$, $H(Y/X)$ and $H(X,Y)$ the system entropies. They depend on both the channel characteristics and the input $X$, which between them will determine $Y$.
Summarizing, we have :

1. Measure of the average uncertainty of the channel input :

$$H(X) = -\sum_i P(x_i)log_2P(x_i)$$

2. Measure of the average uncertainty at the output of the channel :

$$H(Y) = -\sum_{j} P(y_j) log_2 P(y_j)$$

3. Measure of the average uncertainty about the input of the channel being the output of the channel is known- equivocation of $Y$ with respect to $X$ :

$$H(\frac{X}{Y}) = -\sum_{i}\sum_{j} P(x_i, y_j).log_2 P(\frac{x_i}{y_j})$$

4. Measure of the average uncertainty at the output of the channel given that $X$ was transmitted :

$$H(\frac{Y}{X}) = -\sum_{i}\sum_{j} P(x_i, y_j).log_2 P(\frac{y_j}{x_i})$$

5. Measure of the average amount of uncertainty of the communication channel as a whole :

$$H(X,Y) = H(X) + H(\frac{Y}{X}) = H(Y) + H(\frac{X}{Y})$$