MODULE – 3 INFORMATION CHANNELS

STRUCTURE

1. Objectives
2. Introduction
3. Communication channel
4. Channel model and channel capacity
5. Mutual information
6. Review questions.
7. Outcomes.

OBJECTIVES

After completion of this module the student will be able

1. To learn about different Communication channel in communication systems.
2. To find channel capacity of different channels in communication system.
3. To develop channel matrix and to find out mutual information in channel.

3.1 COMMUNICATION CHANNELS:

Observe that the matrix is necessarily a square matrix. The principal diagonal entries are the self-impedances of the respective ports. The off diagonal entries correspond to the transfer or mutual impedances. For a passive network the impedance matrix is always symmetric i.e. $Z^T = Z$, where the superscript indicates transposition.

Similarly, a communication network may be uniquely described by specifying the joint probabilities (JPM). Let us consider a simple communication network comprising of a transmitter (source or input) and a receiver (sink or output) with the interlinking medium-the channel as shown in Fig 4.1.

Fig 4.1 A Simple Communication System

This simple system may be uniquely characterized by the ' Joint probability matrix' (JPM), $P(X, Y)$ of the probabilities existent between the input and output ports.

$$P(X,Y) = \begin{bmatrix} p(x_1, y_1) & p(x_1, y_2) & P(x_1, y_3) & \ldots & P(x_1, y_n) \\ P(x_2, y_1) & P(x_2, y_2) & P(x_2, y_3) & \ldots & P(x_2, y_n) \\ P(x_3, y_1) & P(x_3, y_2) & P(x_3, y_3) & \ldots & P(x_3, y_n) \\ \vdots & \vdots & \vdots & & \vdots \\ P(x_m, y_1) & P(x_m, y_2) & P(x_m, y_3) & \ldots & P(x_m, y_n) \end{bmatrix} \quad \ldots\ldots\ldots\ldots (4.1)$$

For jointly continuous random variables, the joint density function satisfies the following:

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x, y)\,dx\,dy = 1 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (4.2)$$

$$\int_{-\infty}^{+\infty} f(x, y)\,dy = f_X(x) \quad \ldots\ldots\ldots\ldots (4.3)$$

$$\int_{-\infty}^{+\infty} f(x, y)\,dx = f_Y(y) \quad \ldots\ldots\ldots \ldots\ldots\ldots\ldots\ldots (4.4)$$

We shall make use of their discrete counterpart as below:

\ $\sum_{\forall k}\sum_{\forall j} p(x_k, y_j) = 1$,Sum of all entries of JPM .... (4.5)

$\sum_{\forall j} p(x_k, y_j) = p(x_k)$, Sum of all entries of JPM in the k^{th} r ....(4.6)

$\sum_{\forall k} p(x_k, y_j) = p(y_j)$, Sum of all entries of JPM in the j^{th} column .... (4.7)

And also

Thus the joint probabilities, as also the conditional probabilities (as we shall see shortly) form complete finite schemes. Therefore for this simple communication network there are five probability schemes of interest viz: $P(X)$, $P(Y)$, $P(X, Y)$, $P(X|Y)$ and $P(Y|X)$. Accordingly there are five entropy functions that can be described on these probabilities:

H(X): Average information per character or symbol transmitted by the source or the entropy of the source.

H(Y): Average information received per character at the receiver or the entropy of the receiver.

H(X, Y): Average information per pair of transmitted and received characters or the average uncertainty of the communication system as a whole.

H (X|Y): A specific character yj being received. This may be the result of the transmission of one of the xk with a given probability. The average value of the Entropy associated with this scheme when yj covers all the received symbols i.e., E {H (X|yj)} is the entropy H (X|Y), called the 'Equivocation', a measure of information about the source when it is known that Y is received.

H (Y|X) : Similar to H (X|Y), this is a measure of information about the receiver.

The marginal Entropies $H(X)$ and $H(Y)$ give indications of the probabilistic nature of the transmitter and receiver respectively. $H(Y|X)$ indicates a measure of the 'noise' or 'error' in t he channel and the equivocation $H(X|Y)$ tells about the ability of recovery or reconstruction of the transmitted symbols from the observed output symbols.

The above idea can be generalized to an ***n***- port communication system, problem being similar to the study of random vectors in a product space (***n***-dimensional random variables Theory). In each product space there are finite numbers of probability assignments (joint, marginal and conditional) of different orders, with which we may associate entropies and arrive at suitable physical interpretation. However, concepts developed for a two-dimensional scheme will be sufficient to understand and generalize the results for a higher order communication system.

3.2 JOINT AND CONDITIONAL ENTROPIES:

In view of Eq (4.2) to Eq (4.5), it is clear that all the probabilities encountered in a two dimensional communication system could be derived from the JPM. While we can compare the JPM, therefore, to the impedance or admittance matrices of an ***n***-port electric network in giving a unique description of the system under consideration, notice that the JPM in general, need not necessarily be a square matrix and even if it is so, it need not be symmetric.

We define the following entropies, which can be directly computed from the JPM.

$$H(X, Y) = p(x_1, y_1) \log \frac{1}{p(x_1, y_1)} + p(x_1, y_2) \log \frac{1}{p(x_1, y_2)} + \ldots + p(x_1, y_n) \log \frac{1}{p(x_1, y_n)}$$

$$+ p(x_2, y_1) \log \frac{1}{p(x_2, y_1)} + p(x_2, y_2) \log \frac{1}{p(x_2, y_2)} + \ldots + p(x_2, y_n) \log \frac{1}{p(x_1, y_1)}$$

$$+\ldots p(x_m, y_1)\log\frac{1}{p(x_m, y_1)} + p(x_m,y_2)\log\frac{1}{p(x_m, y_2)} +\ldots p(x_m,y_n)\log\frac{1}{p(x_m, y_n)} \text{ or}$$

$$H(X, Y) = \sum_{k=1}^{m}\sum_{j=1}^{n} p(x_k, y_j)\log\frac{1}{p(x_k, y_j)} \quad \ldots\ldots\ldots(4.9)$$

$$H(X) = \sum_{k=1}^{m} p(x_k)\log\frac{1}{p(x_k)}$$

Using Eq (4.6) <u>only for the multiplication term</u>, this equation can be re-written as:

$$H(X) = \sum_{k=1}^{m}\sum_{j=1}^{n} p(x_k, y_j)\log\frac{1}{p(x_k)} \quad \ldots\ldots \quad (4.10)$$

Similarly, $H(Y) = \sum_{j=1}^{n}\sum_{k=1}^{m} p(x_k, y_j)\log\frac{1}{p(y_j)}$ ………………. (4.11)

Next, from the definition of the conditional probability we have:

$$P\{X = x_k \mid Y = y_j\} = \frac{P\{X = x_k, Y = y_j\}}{P\{Y = y_j\}}$$

i.e., $p(x_k \mid y_j) = p(x_k, y_j) / p(y_j)$

Then $\sum_{k=1}^{m} p(x_k \mid y_j) = \frac{1}{p(y_j)}\sum_{k=1}^{m} p(x_k, y_j) = \frac{1}{p(y_j)}.p(y_j) = 1$ ………. (4.12)

Thus, the set $[X \mid y_j] = \{x_1 \mid y_j,\ x_2 \mid y_j \cdots\ x_m \mid y_j\}$; $P[X \mid y_j] = \{p(x_1 \mid y_j),\ p(x_2 \mid y_j)\cdots\ p(x_m \mid y_j)\}$, forms a complete finite scheme and an entropy function may therefore be defined for this scheme as below:

$$H(X \mid y_j) = \sum_{k=1}^{m} p(x_k \mid y_j)\log\frac{1}{p(x_k \mid y_j)}.$$

Taking the average of the above entropy function for all admissible characters received, we have the average " ***conditional Entropy" or "Equivocation":***

$$H(X \mid Y) = E\{H(X \mid y_j)\}_j$$
$$= \sum_{j=1}^{n} p(y_j)\, H(X \mid y_j)$$
$$= \sum_{j=1}^{n} p(y_j)\sum_{k=1}^{m} p(x_k \mid y_j)\log\frac{1}{p(x_k \mid y_j)}$$

Or $H(X \mid Y) = \sum_{j=1}^{n}\sum_{k=1}^{m} p(x_k, y_j)\log\frac{1}{p(x_k \mid y_j)}$ ………… (4.13)

Eq (4.13) specifies the " ***Equivocation*** ". It specifies the average amount of information n eeded to specify an input character provided we are allowed to make an observation of the output produced by that input. Similarly one can define the conditional entropy $H(Y \mid X)$ by:

$$H(Y \mid X) = \sum_{k=1}^{m} \sum_{j=1}^{n} p(x_k, y_j) \log \frac{1}{p(y_j \mid x_k)} \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (4.14)$$

Observe that the manipulations, made in deriving Eq 4.10, Eq 4.11, Eq 4.13 and Eq 4.14, are intentional. ' ***The entropy you want is simply the double summation of joint probability multiplied by logarithm of the reciprocal of the probability of interest'*** . For example, if you want joint entropy, then the probability of interest will be joint probability. If you want source entropy, probability of interest will be the source probability. If you want the equivocation or conditional entropy, $H(X|Y)$ then probability of

interest will be the conditional probability $p(x_K|y_j)$ and so on.

All the five entropies so defined are all inter-related. For example, consider Eq (4.14). We have:

$$H(Y \mid X) = \sum_k \sum_j p(x_k, y_j) \log \frac{1}{p(y_j \mid x_k)}$$

Since, $\dfrac{1}{p(y_j \mid x_k)} = \dfrac{p(x_k)}{p(x_k, y_j)}$

We can straight away write:

$$H(Y|X) = \sum_k \sum_j p(x_k, y_j) \log \frac{1}{p(y_j \mid x_k)} - \sum_k \sum_j p(x_k, y_j) \log \frac{1}{p(x_k)}$$

Or $\quad H(Y \mid X) = H(X, Y) - H(X)$

That is: $\quad H(X, Y) = H(X) + H(Y \mid X) \quad \ldots\ldots\ldots \quad (4.15)$

Similarly, you can show: $H(X, Y) = H(Y) + H(X \mid Y) \quad \ldots\ldots\ldots \quad (4.16)$

Consider $H(X) - H(X|Y)$. We have:

$$H(X) - H(X|Y) = \sum_k \sum_j p(x_k, y_j) \left(\log \frac{1}{p(x_k)} - \log \frac{1}{p(x_k \mid y_j)} \right)$$

$$= \sum_k \sum_j p(x_k, y_j) \log \frac{p(x_k, y_j)}{p(x_k) \cdot p(y_j)} \quad (4.17)$$

Using the logarithm inequality derived earlier, you can write the above equation as:

$$H(X) - H(X|Y) = \log e \sum_k \sum_j p(x_k, y_j) \ln \frac{p(x_k, y_j)}{p(x_k) \cdot p(y_j)}$$

$$\geq \log e \sum_k \sum_j p(x_k, y_j) \left(1 - \frac{p(x_k) . p(y_j)}{p(x_k, y_j)} \right)$$

$$\geq \log e \left(\sum_k \sum_j p(x_k, y_j) - \sum_k \sum_j p(x_k).p(y_j) \right)$$

$$\geq \log e \left(\sum_k \sum_j p(x_k, y_j) - \sum_k p(x_k). \sum_j p(y_j) \right) \geq 0$$

Because $\sum_k \sum_j p(x_k, y_j) = \sum_k p(x_k) = \sum_j p(y_j) = 1$. Thus it follows that:

$$H(X) \geq H(X|Y) \quad \ldots\ldots\ldots \quad (4.18$$

Similarly, $H(Y) \geq H(Y|X)$ ………….. (4.19)

Equality in Eq (4.18) & Eq (4.19) holds iffy $P(x_k, y_j) = p(x_k).p(y_j)$; *i*.e., if and only if input symbols and output symbols are statistically independent of each other.

NOTE : Whenever you write the conditional probability matrices you should bear in mind the property described in Eq.(4.12), i.e. For the CPM (conditional probability matrix) $P(X|Y)$, if you add all the elements in any column the sum shall be equal to unity. Similarly, if you add all elements along any row of the CPM, $P(Y|X)$ the sum shall be unity

Example 4.1

Determine different entropies for the JPM given below and verify their relationships.

$$P(X,Y) = \begin{bmatrix} 0.2 & 0 & 0.2 & 0 \\ 0.1 & 0.01 & 0.01 & 0.01 \\ 0 & 0.02 & 0.02 & 0 \\ 0.04 & 0.04 & 0.01 & 0.06 \\ 0 & 0.06 & 0.02 & 0.2 \end{bmatrix}$$

Using $p(x_k) = \sum_{j=1}^{n} p(x_k, y_j)$, we have, by adding entries of $P(X, Y)$ row-wise we get:

$P(X) = [0.4, 0.1, 0.04, 0.15, 0.28]$

Similarly adding the entries column-wise we get:

$P(Y) = [0.34, 0.13, 0.26, 0.27]$

Hence we have:

$$H(X,Y) = 3 \times 0.2 \log\frac{1}{0.2} + 0.1 \times \log\frac{1}{0.1} + 4 \times 0.01 \log\frac{1}{0.01} + 3 \times 0.02 \log\frac{1}{0.02} + 2 \times 0.04 \log\frac{1}{0.04} + 2 \times 0.06 \log\frac{1}{0.06}$$

$$= 3.188311023 \text{ bits /sym}$$

$$H(X) = 0.4 \log\frac{1}{0.4} + 0.13 \log\frac{1}{0.13} + 0.04 \log\frac{1}{0.04} + 0.15 \log\frac{1}{0.15} + 0.28 \log\frac{1}{0.28}$$

$$= 2.021934821 \text{ bits / sym}$$

$$H(Y) = 0.34 \log\frac{1}{0.34} + 0.13 \log\frac{1}{0.13} + 0.26 \log\frac{1}{0.26} + 0.27 \log\frac{1}{0.27}$$

$$= 1.927127708 \text{ bits / sym}$$

Since $p(x_k | y_j) = \dfrac{P(x_k, y_j)}{P(y_j)}$ we have:

(Divide the entries in the j^{th} column of the JPM of $p(y_j)$

$$P(X|Y) = \begin{bmatrix} \frac{0.2}{0.34} & 0 & \frac{0.2}{0.26} & 0 \\ \frac{0.1}{0.34} & \frac{0.01}{0.13} & \frac{0.01}{0.26} & \frac{0.01}{0.27} \\ 0 & \frac{0.02}{0.13} & \frac{0.02}{0.26} & 0 \\ \frac{0.04}{0.34} & \frac{0.04}{0.13} & \frac{0.01}{0.26} & \frac{0.06}{0.27} \\ 0 & \frac{0.06}{0.13} & \frac{0.02}{0.26} & \frac{0.20}{0.27} \end{bmatrix}$$

$$\therefore H(X|Y) = 0.2\log\frac{0.34}{0.2} + 0.2\log\frac{0.26}{0.2} + 0.1\log\frac{0.34}{0.1}$$
$$+ 0.01\log\frac{0.13}{0.01} + 0.01\log\frac{0.26}{0.01} + 0.01\log\frac{0.27}{0.01}$$
$$+ 0.02\log\frac{0.13}{0.02} + 0.02\log\frac{0.26}{0.02} + 0.04\log\frac{0.34}{0.04}$$
$$+ 0.04\log\frac{0.13}{0.04} + 0.01\log\frac{0.26}{0.01} + 0.06\log\frac{0.27}{0.06}$$
$$+ 0.06\log\frac{0.13}{0.06} + 0.02\log\frac{0.26}{0.02} + 0.2\log\frac{0.27}{0.2}$$

$=1.261183315$ ***bits / symbol***

Similarly, dividing the entries in the k^{th} row of JPM by $p(x_k)$, we obtain the CPM $P(Y|X)$. Then we have:

$$P(Y|X) = \begin{bmatrix} \frac{0.2}{0.4} & 0 & \frac{0.2}{0.4} & 0 \\ \frac{0.1}{0.13} & \frac{0.01}{0.13} & \frac{0.01}{0.13} & \frac{0.01}{0.13} \\ 0 & \frac{0.02}{0.04} & \frac{0.02}{0.04} & 0 \\ \frac{0.04}{0.15} & \frac{0.04}{0.15} & \frac{0.01}{0.15} & \frac{0.06}{0.15} \\ 0 & \frac{0.06}{0.28} & \frac{0.02}{0.28} & \frac{0.20}{0.28} \end{bmatrix}$$

Thus by actual computation we have

$H(X, Y) = 3.188311023$ ***bits/Sym*** $H(X) = 2.02193482$ ***bit/Sym*** $H(Y) = 1.927127708$ ***bits/Sym***

$H(X|Y) = 1.261183315$ ***bits/Sym*** $H(Y|X) = 1.166376202$ ***bits/Sym***

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

$H(X) > H(X|Y)$ and $H(Y) > H(Y|X)$

$$\text{And } H(Y|X) = 2 \times 0.2\log\frac{0.4}{0.2} + 0.1\log\frac{0.13}{0.1} + 3 \times 0.01\log\frac{0.13}{0.01} + 2 \times 0.02\log\frac{0.04}{0.02}$$

$$+ 2 \times 0.04\log\frac{0.05}{} + 0.01\log\frac{0.15}{} + 0.06\log\frac{0.15}{} + 0.06\log\frac{0.28}{}$$

$$\qquad 0.04 \qquad 0.01 \qquad 0.06 \qquad 0.06$$

$$+ 2 \times 0.02 \log \frac{0.28}{0.02} = 1.166376202 \text{ bits / sym} .$$

3.3 Mutual information:

On an average we require $H(X)$ bits of information to specify one input symbol. However, if we are allowed to observe the output symbol produced by that input, we require, then, only $H(X|Y)$ bits of information to specify the input symbol. Accordingly, we come to the conclusion, that on an average, observation of a single output provides with $[H(X) - H(X|Y)]$ bits of information. This difference is called '***Mutual Information***' or '***Transinformation***' of the channel, denoted by $I(X, Y)$. Thus:

$$I(X, Y) \triangleq H(X) - H(X|Y) \qquad \ldots\ldots\ldots\ldots (4.20)$$

Notice that in spite of the variations in the source probabilities, $p(x_k)$ (may be due to noise in the channel), certain probabilistic information regarding the state of the input is available, once the conditional probability $p(x_k | y_j)$ is computed at the receiver end. The difference between the initial uncertainty of the source symbol x_k, i.e. $\log 1/p(x_k)$ and the final uncertainty about the same source symbol x_k, after receiving y_j, i.e. $\log 1/p(x_k | y_j)$ is the information gained through the channel. This difference we call as the mutual information between the symbols x_k and y_j. Thus

$$I(x_k, y_j) = \log \frac{1}{p(x_k)} - \log \frac{1}{p(x_k | y_j)}$$

$$= \log \frac{p(x_k | y_j)}{p(x_k)} \qquad \ldots\ldots\ldots\ldots (4.21\ a)$$

Or
$$I(x_k, y_j) = \log \frac{p(x_k . y_j)}{p(x_k).p(y_j)} \qquad \ldots\ldots\ldots\ldots (4.21\ b)$$

Notice from Eq. (4.21*a*) that

$$I(x_k) = I(x_k, x_k) = \log \frac{p(x_k | x_k)}{p(x_k)} = \log \frac{1}{p(x_k)}$$

This is the definition with which we started our discussion on information theory! Accordingly $I(x_k)$ is also referred to as 'Self Information

It is clear from Eq (3.21b) that, as $\dfrac{p(x_k, y_j)}{p(x_k)} = p(y_j \mid x_k)$,

$$I(x_k, y_j) = \log \frac{p(y_j \mid x_k)}{p(y_j)} = \log \frac{1}{p(y_j)} - \log \frac{1}{p(y_j \mid x_k)}$$

Or $$I(x_k, y_j) = I(y_j) - I(y_j \mid x_k) \quad \text{.......} \quad (4.22)$$

Eq (4.22) simply means that "the Mutual information ' is symmetrical with respect to its arguments.i.e.

$$I(x_k, y_j) = I(y_j, x_k) \quad \text{..............} \ (4.23)$$

Averaging Eq. (4.21b) over all admissible characters x_k and y_j, we obtain the average information gain of the receiver:

$$I(X, Y) = E\{I(x_k, y_j)\}$$
$$= \sum_k \sum_j I(x_k, y_j). p(x_k, y_j)$$
$$= \sum_k \sum_j p(x_k, y_j).\log \frac{p(x_k, y_j)}{p(x_k)p(y_j)} \quad \text{............} (4.24)$$

From Eq (4.24) we have:

1) $$I(X, Y) = \sum_k \sum_j p(x_k, y_j) \log \frac{1}{p(x_k)} . - \log \frac{1}{p(x_k \mid y_j)} = H(X) - H(X \mid Y) \quad \text{......} (4.25)$$

2) $$I(X, Y) = \sum_k \sum_j p(x_k, y_j) [\log \frac{1}{p(y_j)} . - \log \frac{1}{p(y_j \mid x_k)}$$
$$= H(Y) - H(Y \mid X) \quad \text{......} \quad (4.26)$$

3) $$I(X,Y) = \sum_k \sum_j p(x_k, y_j) \log \frac{1}{p(x_k)} + . \sum_{k} \sum_{j} p(x_k, y_j) \log \frac{1}{p(y_j)} - \sum_K \sum_J p(x_k, y_j) \log \frac{1}{p(x_k y_j)}$$

Or $$I(X, Y) = H(X) + H(Y) - H(X, Y) \quad \text{......} \quad (4.27)$$

Further, in view of <u>Eq.(4.18) & Eq.(4.19)</u> we conclude that, " ***even though for a particular received symbol, yj, H(X) – H(X | Yj) may be negative, when all the admissible ou tput symbols are covered, the average mutual information is always non- negative***". That is to say, we cannot loose information on an average by observing the output of a channel. An easy method, of remembering the various relationships, is given in Fig 4.2.Althogh the diagram resembles a Venn-diagram, it is not, and the diagram is only a tool to remember the relationships. That is all. You cannot use this diagram for proving any result.

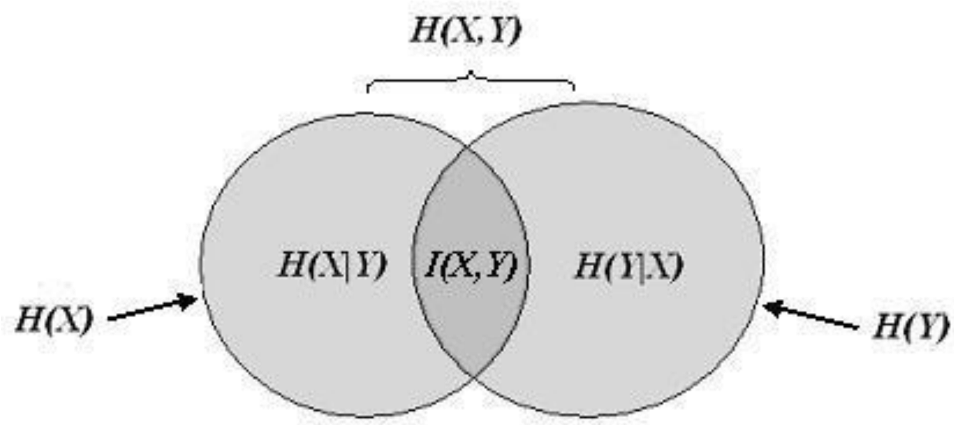


Fig 4.2 Entropy Relations

The entropy of ***X*** is represented by the circle on the left and that of ***Y*** by the circle on the right. The overlap between the two circles (dark gray) is the mutual information so that the remaining (light gray) portions of $H(X)$ and $H(Y)$ represent respective equivocations. Thus we have

$$H(X \mid Y) = H(X) - I(X, Y) \text{ and } H(Y \mid X) = H(Y) - I(X, Y)$$

The joint entropy $H(X, Y)$ is the sum of $H(X)$ and $H(Y)$ except for the fact that the overlap is added twice so that

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

Also observe $H(X, Y) = H(X) + H(Y|X)$

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

For the JPM given in Example 4.1, $I(X, Y) = 0.760751505$ bits / sym

Shannon Theorem: Channel Capacity:

Clearly, the mutual information ***I (X, Y)*** depends on the source probabilities apart from the channel probabilities. For a general information channel we can always make ***I(X, Y) = 0*** by choosing any one of the input symbols with a probability one or by choosing a channel with independent input and output. Since ***I(X, Y)*** is always nonnegative, we thus know the minimum value of the Transinformation. However, the question of max ***I(X, Y)*** for a general channel is not easily answered.

Our intention is to introduce a suitable measure for the efficiency of the channel by making a comparison between the actual rate and the upper bound on the rate of transmission of information. Shannon's contribution in this respect is most significant. Without botheration about the proof, let us see what this contribution is.

Shannon's theorem: on channel capacity ("coding Theo rem")

It is possible, in principle, to device a means where by a communication system will transmit information with an arbitrary small probability of error, provided that the information rate ***R***(=***r***×***I*** (***X***,***Y***),where ***r*** is the symbol rate) is less than or equal to a rate ' ***C***' called "channel capacity".

The technique used to achieve this objective is called coding. To put the matter more formally, the theorem is split into two parts and we have the following statements.

Positive statement:

" *Given a source of **M** equally likely messages, with **M>>1**, which is generating information at a rate **R**, and a channel with a capacity **C.** If **R ≤ C**, then there exists a coding technique such that the output of the source may be transmitted with a probability of error of receiving the message that can be made arbitrarily small*".

This theorem indicates that for ***R≤ C*** transmission may be accomplished without error even in the presence of noise. The situation is analogous to an electric circuit that comprises of only pure capacitors and pure inductors. In such a circuit there is no loss of energy at all as the reactors have the property of storing energy rather than dissipating.

Negative statement:

" *Given the source of **M** equally likely messages with **M>>1**, which is generating information at a rate **R** and a channel with capacity **C**. Then, if **R>C**, then the probability of error of receiving the message is close to unity for every set of **M** transmitted symbols*".

This theorem shows that if the information rate ***R*** exceeds a specified value ***C***, the error probability will increase towards unity as ***M*** increases. Also, in general, increase in the complexity of the coding results in an increase in the probability of error. Notice that the situation is analogous to an electric network that is made up of pure resistors. In such a circuit, whatever energy is supplied, it will be dissipated in the form of heat and thus is a "lossy network".

You can interpret in this way: Information is poured in to your communication channel. You should receive this without any loss. Situation is similar to pouring water into a tumbler. Once the tumbler is full, further pouring results in an over flow. You cannot pour water more than your tumbler can hold. Over flow is the loss.

Shannon defines " ***C***" the channel capacity of a communication channel a s the maximum value of Transinformation, $I(X, Y)$:

$$C \triangleq Max\ I(X, Y) = Max\ [H(X) - H(Y|X)] \qquad \ldots\ldots\ldots\ldots (4.28)$$

The maximization in Eq (4.28) is with respect to all possible sets of probabilities that could be assigned to the input symbols. Recall the maximum power transfer theorem: 'In any network, maximum power will be delivered to the load only when the load and the source are properly matched'. The device used for this matching purpose, we shall call a "transducer ". For example, in a radio receiver, for optimum response, the impedance of the loud speaker will be matched to the impedance of the output power amplifier, through an output transformer.

This theorem is also known as "The Channel Coding Theorem" (Noisy Coding Theorem). It may be stated in a different form as below:

$$R \le C \text{ or } r_S\, H(S) \le r_C\, I(X,Y)_{Max} \text{ or} \{ H(S)/T_S\} \le \{ I(X,Y)_{Max}/T_C\}$$

"If a discrete memoryless source with an alphabet 'S' has an entropy H(S) and produces symbols every 'T_S' seconds; and a discrete memoryless channel has a capacity $I(X,Y)_{Max}$ and is used once every T_C seconds; then if

There exists a coding scheme for which the source output can be transmitted over the channel and be reconstructed with an arbitrarily small probability of error. The parameter C/T_C is called the critical rate. When this condition is satisfied with the equality sign, the system is said to be signaling at the critical rate.

Conversely, if $\frac{H(S)}{T_S} > \frac{I(X,Y)_{Max}}{T_C}$***, it is not possible to transmit information over the channel and reconstruct it with an arbitrarily small probability of error***

A communication channel, is more frequently, described by specifying the source probabilities ***P(X)*** & the conditional probabilities ***P (Y|X)*** rather than specifying the JPM. The CPM, ***P (Y|X),*** is usually refereed to as the ' ***noise characteristic*** ' of the channel. Therefore unless otherwise specified, we shall understand that the description of the channel, by a matrix or by a 'Channel diagram' always refers to CPM, ***P (Y|X)***. Thus, in a discrete communication channel with pre-specified noise characteristics (i.e. with a given transition probability matrix, ***P (Y|X))*** the rate of information transmission depends on the source that drives the channel. Then, the maximum rate corresponds to a proper matching of the source and the channel. This ideal characterization of the source depends in turn on the transition probability characteristics of the given channel.

Redundancy and Efficiency:

A redundant source is one that produces 'dependent' symbols. (Example: The Markov source). Such a source generates symbols that are not absolutely essential to convey information. As an illustration, let us consider the English language. It is really unnecessary to write "U" following the letter "Q". The redundancy in English text is e stimated to be 50%(refer J Das etal, Sham Shanmugam, Reza, Abramson, Hancock for detailed discussion.) This implies that, in the long run, half the symbols are unnecessary! For example, consider the following sentence.

" ***Y.u sh..ld b. abl. t. re.d t.is ev.n tho… sev.r.l l.t..rs .r. m.s..ng*** "

However, we want redundancy. Without this redundancy abbreviations would be impossible and any two dimensional array of letters would form a crossword puzzle! We want redundancy even in communications to facilitate error detection and error correction. Then how to measure redundancy? Recall that for a Markov source, $H(S) < H(\bar{S})$, where $\bar{S}$ is an ad- joint, zero memory source. That is, when dependence creeps in, the entropy of the source will be reduced and this can be used as a measure indeed!

" ***The redundancy of a sequence of symbols is measured by noting the amount by which the entropy has been reduced***".

When there is no inter symbol influence the entropy at the receiver would be ***H(X)*** for any given set of messages $\{X\}$ and that when inter symbol influence occurs the entropy would be ***H (Y|X)***. The difference [***H(X) –H (Y|X)***] is the net reduction in entropy and is called " ***Absolute Redundancy***". Generally it is measured relative to the maximum entropy and thus we have for the " ***Relative Redundancy***" or simply, ' ***redundancy***' ***, E***

E = (Absolute Redundancy) ÷ H(X)

Or $$E = 1 - \frac{H(Y\,|\,X)}{H(X)}$$ .... (4.29)

Careful observation of the statements made above leads to the following alternative definition for redundancy,

$$E = 1 - \frac{R}{C}$$ .... (4.30)

Where R is the actual rate of Transinformation (mutual information) and C is the channel capacity. From the above discussions, a definition for the efficiency, η for the channel immediately follows:

$$\eta = \frac{\text{Actual rate of mutual information}}{\text{maximum possible rate}}$$

That is. $$\eta = \frac{R}{C}$$ .... (4.31)

and $$\eta = 1 - E$$ .... (4.32)

3.4 Capacity of Channels:

While commenting on the definition of 'Channel capacity', Eq. (4.28), we have said that maximization should be with respect to all possible sets of input symbol probabilities. Accordingly, to arrive at the maximum value it is necessary to use some Calculus of Variation techniques and the problem, in general, is quite involved.

Example 3.2: Consider a Binary channel specified by the following noise characteristic (channel matrix):

$$P(Y\,|\,X) = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & \frac{3}{4} \end{bmatrix}$$

The source probabilities are: $p(x_1) = p,\ p(x_2) = q = 1-p$

Clearly, $H(X) = -p \log p - (1 - p) \log (1 - p)$

We shall first find JPM and proceed as below:

$$P(X,Y) = \begin{bmatrix} p(x_1).\,p(y_1\,|\,x_1) & p(x_1).\,p(y_2\,|\,x_1) \\ p(x_2).\,p(y_1\,|\,x_2) & p(x_2).\,p(y_2\,|\,x_2) \end{bmatrix} = \begin{bmatrix} \frac{p}{2} & \frac{p}{2} \\ \frac{1-p}{4} & \frac{3(1-p)}{4} \end{bmatrix}$$

Adding column-wise, we get:

$$p(y_1) = \frac{p}{2} + \frac{1-p}{4} = \frac{1+p}{4} \quad \text{and} \quad p(y_2) = \frac{p}{2} + \frac{3(1-p)}{4} = \frac{3-p}{4}$$

Hence $H(Y) = \frac{1+p}{4}\log\frac{4}{1+p} + \frac{3-p}{4}\log\frac{4}{3-p}$

And $H(Y|X) = \frac{p}{2}\log 2 + \frac{p}{2}\log 2 + \frac{1-p}{4}\log 4 + \frac{3(1-p)}{4}\log\frac{4}{3}$

$$I(X, Y) = H(Y) - H(Y|X) = 1 - \frac{3\log 3}{4} p + \frac{3\log 3}{4} - \frac{1+p}{4}\log(1+p) - \frac{3(1-p)}{4}\log(3-p)$$

Writing $\log x = \log e \times \ln x$ and setting $\frac{dI}{dp} = 0$ yields straight away:

$p = \frac{3a-1}{1+a} = 0.488372093$, Where $a = 2^{(4-3\log 3)} = 0.592592593$

With this value of p, we find $I(X, Y)_{Max} = 0.048821$ **bits /sym**

For other values of p it is seen that $I(X, Y)$ is less than $I(X, Y)_{max}$

Although, we have solved the problem in a straight forward way, it will not be the case

p	. 0.2	. 0.4	. 0.5	. 0.6	. 0.8
I(X,Y) Bits / sym	. 0.32268399	. 0.04730118	. 0.04879494	. 0.046439344	. 0.030518829

When the dimension of the channel matrix is more than two. We have thus shown that the channel capacity of a given channel indeed depends on the source probabilities. The computation of the channel capacity would become simpler for certain class of channels called the 'symmetric 'or 'uniform' channels.

Muroga's Theorem **:**

The channel capacity of a channel whose noise characteristic, $P(Y|X)$, is square and non-singular, the channel capacity is given by the equation:

$$C = \log \sum_{i=1}^{i=n} 2^{-Q_i} \qquad \text{......................(4.33)}$$

Where Q_i are the solutions of the matrix equation $P(Y|X).Q = [h]$, where $h = [h_1, h_2, h_3, h_4 \ldots h_n]^t$ are the row entropies of $P(Y|X)$.

$$\begin{bmatrix} p_{11} & p_{12} & p_{13} & \ldots & p_{1n} \\ p_{21} & p_{22} & p_{23} & \ldots & p_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ p_{n1} & p_{n2} & p_{n3} & \ldots & p_{nn} \end{bmatrix} \begin{bmatrix} Q_1 \\ Q_2 \\ \vdots \\ Q_n \end{bmatrix} = \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_n \end{bmatrix}$$

From this we can solve for the source probabilities (i.e. Input symbol probabilities):

$$[p_1, p_2, p_3 \ldots p_n] = [p_1', p_2', p_3' \ldots p_n']\, P^{-1}[Y|X]$$, provided the inverse exists.

However, although the method provides us with the correct answer for Channel capacity, this value of C may not necessarily lead to physically realizable values of probabilities and if $P^{-1}[Y|X]$ does not exist ,we will not have a solution for Q_i`s as well. One reason is that we are not able to incorporate the inequality constraints $0 \leq p_i \leq 1$ .Still, within certain limits; the method is indeed very useful.

Example 3.2: Consider a Binary channel specified by the following noise characteristic (channel matrix):

$$P(Y|X) = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & \frac{3}{4} \end{bmatrix}$$

The row entropies are:

$$h_1 = \frac{1}{2}\log 2 + \frac{1}{2}\log 2 = 1 \text{ bit / symbol.}$$

$$h_2 = \frac{1}{4}\log 4 + \frac{3}{4}\log\frac{4}{3} = 0.8112781 \text{ bits / symbol.}$$

$$P^{-1}[Y|X] = \begin{bmatrix} 3 & -1 \\ -1 & 2 \end{bmatrix}$$

$$\begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix} = P^{-1}[Y|X] \cdot \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} = \begin{bmatrix} 1.3774438 \\ 0.6225562 \end{bmatrix}$$

$$C = \log\left[2^{-Q_1} + 2^{-Q_2}\right] = 0.048821 \text{ bits / symbol, as before.}$$

$$\text{Further, } p_1' = 2^{-Q_1 - C} = 0.372093 \text{ and } p_2' = 2^{-Q_2 - C} = 0.627907.$$

$$[p_1 \;\; p_2] = [p_1' \;\; p_2'] \cdot P^{-1}[Y|X] = [0.488372 \;\; 0.511628]$$

Giving us $p = 0.488372$

Example 3.3:

Consider a 3×3 channel matrix as below:

$$P[Y|X] = \begin{bmatrix} 0.4 & 0.6 & 0 \\ 0.5 & 0 & 0.5 \\ 0 & 0.6 & 0.4 \end{bmatrix}$$

The row entropies are:

$$h_1 = h_3 = 0.4 \log (1/0.4) + 0.6 \log (1/0.6) = 0.9709505 \text{ bits / symbol.}$$

$$h_2 = 2 \times 0.5 \log (1/0.5) = 1 \text{ bit / symbol.}$$

$$P^{-1}[Y|X] = \begin{bmatrix} 1.25 & 1 & -1.25 \\ 5/6 & -2/3 & 5/6 \\ -1.25 & 1 & 1.25 \end{bmatrix}$$

$$\begin{bmatrix} Q_1 \\ Q_2 \\ Q_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1.0193633 \\ 1 \end{bmatrix}$$

$$C = \log \{2^{-1} + 2^{-1.0193633} + 2^{-1}\} = 0.5785369 \text{ bits / symbol.}$$

$$p'_1 = 2^{-Q_1 - C} = 0.3348213 = p'_3,\ p'_2 = 2^{-Q_2 - C} = 0.3303574.$$

Therefore, $p_1 = p_3 = 0.2752978$ and $p_2 = 0.4494043$.

Suppose we change the channel matrix to:

$$P[Y|X] = \begin{bmatrix} 0.8 & 0.2 & 0 \\ 0.5 & 0 & 0.5 \\ 0 & 0.2 & 0.8 \end{bmatrix} \quad P^{-1}[Y|X] = \begin{bmatrix} 0.625 & 1 & -0.625 \\ 2.5 & -4 & 2.5 \\ -0.625 & 1 & 0.625 \end{bmatrix}$$

We have:

$h_1 = h_3 = 0.721928$ ***bits / symbol***, and $h_2 = 1$ ***bit / symbol.***

This results in:

$$Q_1 = Q_3 = 1;\ Q_2 = -0.39036.$$

$$C = \log \{2 \times 2^{-1} + 2^{+0.39036}\} = 1.2083427 \text{ bits / symbol.}$$

$$p'_1 = 2^{-Q_1 - C} = 0.2163827 = p'_3,\ p'_2 = 2^{-Q_2 - C} = 0.5672345$$

Giving: $p_1 = p_3 = 1.4180863$ ***and*** $p_2 =$ ***Negative***!

Thus we see that, although we get the answer for ***C*** the input symbol probabilities computed are not physically realizable. However, in the derivation of the equations, as already pointed out, had we included the conditions on both input and output probabilities we might have got an excellent result! But such a derivation becomes very formidable as you cannot arrive at a numerical solution! You will have to resolve your problem by graphical methods only which will also be a tough proposition! The formula can be used, however, with restrictions on the channel transition probabilities. For example, in the previous problem, for a physically realizable p_1, p_{11} should be less than or equal to ***0.64***. (Problems 4.16 and 4.18 of Sam Shanmugam to be solved using this method)

Symmetric Channels:

The Muroga's approach is useful only when the noise characteristic ***P [X|Y]*** is a square and invertible matrix. For channels with $m \neq n$, we can determine the Channel capacity by simple inspection when the channel is "***Symmetric" or "Uniform".***

Consider a channel defined by the noise characteristic:

$$P[Y|X] = \begin{bmatrix} p_{11} & p_{12} & p_{13} & \cdots & p_{1n} \\ p_{21} & p_{22} & p_{23} & \cdots & p_{2n} \\ p_{31} & p_{32} & p_{33} & \cdots & p_{3n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ p_{m1} & p_{m2} & p_{m3} & \cdots & p_{mn} \end{bmatrix} \quad \ldots\ldots\ldots\ldots\ldots (4.34)$$

This channel is said to be Symmetric or Uniform if the second and subsequent rows of the channel matrix are certain permutations of the first row. That is the elements of the second and subsequent rows are exactly the same as those of the first row except for their locations. This is illustrated by the following matrix:

$$P[Y|X] = \begin{bmatrix} p_1 & p_2 & p_3 & \cdots & p_n \\ p_{n-1} & p_2 & p_n & \cdots & p_4 \\ p_3 & p_2 & p_1 & \cdots & p_5 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ p_n & p_{n-1} & p_{n-2} & \cdots & p_1 \end{bmatrix} \quad \ldots\ldots\ldots\ldots\ldots (4.35)$$

Remembering the important property of the conditional probability matrix, ***P*** [***Y***|***X***], that the sum of all elements in any row should add to unity; we have:

$$\sum_{j=1}^{n} p_j = 1 \quad \ldots\ldots\ldots\ldots\ldots\ldots (4.36)$$

The conditional entropy ***H*** (***Y***|***X***) for this channel can be computed from:

$$H(Y|X) = \sum_{k=1}^{m}\sum_{j=1}^{n} p(x_k, y_j)\log\frac{1}{p(x_k, y_j)}$$

$$= \sum_{k=1}^{m} p(x_k).\sum_{j=1}^{n} p(y_j|x_k)\log\frac{1}{p(y_j|x_k)}$$

However, for the channel under consideration observe that:

$$\sum_{k=1}^{m} p(x_k).\sum_{j=1}^{n} p(y_j \mid x_k)\log\frac{1}{p(y_j \mid x_k)} = \sum_{j=1}^{n} p_j \log\frac{1}{p_j} = h \quad \text{......... (4.37)}$$

is a constant, as the entropy function is symmetric with respect to its arguments and depends only on the probabilities but not on their relative locations. Accordingly, the entropy becomes:

$$H(Y \mid X) = \sum_{k=1}^{m} p(x_k).h = h \quad \text{................(4.38)}$$

as the source probabilities all add up to unity.

Thus the conditional entropy for such type of channels can be computed from the elements of any row of the channel matrix. Accordingly, we have for the mutual information:

$$I(X, Y) = H(Y) - H(Y|X) = H(Y) - h$$

Hence, C = Max $I(X, Y)$ =Max $\{H(Y) - h\}$ = Max $H(Y) - h$

Since, $H(Y)$ will be maximum if and only if all the received symbols are equally probable and as there are n – symbols at the output, we have:

$$H(Y)_{Max} = \log n$$

Thus we have for the symmetric channel:

$$C = \log n - h \quad \text{.............. (4.39)}$$

The channel matrix of a channel may not have the form described in Eq (3.35) but still it can be a symmetric channel. This will become clear if you interchange the roles of input and output. That is, investigate the conditional probability matrix $P(X|Y)$.

We define the channel to be symmetric if the CPM, $P(X|Y)$ has the form:

$$P(X \mid Y) = \begin{matrix} p_1 & p_m \; p_{m-1} & p_2 & \ldots & p_m \; p_{m-1} \\ p_2 & & p_6 & \ldots & \\ p_3 & p_4 & p_m & \ldots & p_{m-2} \\ M & M & M & M & M \\ p_m & p_1 & p_{m-3} & \ldots & p_1 \end{matrix} \quad \text{..................(4.40)}$$

That is, the second and subsequent columns of the CPM are certain permutations of the first column. In other words entries in the second and subsequent columns are exactly the same as in the first column but for different locations. In this case we have:

$$H(X \mid Y) = \sum_{j=1}^{n}\sum_{k=1}^{m} p(x_k, y_j)\log\frac{1}{p(x_k \mid y_j)} = \sum_{j=1}^{n} p(y_j)\sum_{k=1}^{m} p(x_k \mid y_j)\log\frac{1}{p(x_k \mid y_j)}$$

Since $\sum_{j=1}^{n} p(y_j) = 1$ and $\sum_{k=1}^{m} p(x_k \mid y_j)\log\frac{1}{p(x_k \mid y_j)} = \sum_{k=1}^{m} p_k \log\frac{1}{p_k} = h'$ is a constant, because all entries in any column are exactly the same except for their locations, it then follows that:

$$H(X|Y) = h' = \sum_{k=1}^{m} p_k \log \frac{1}{p_k} \qquad \text{............ (4.41)}$$

*Remember that the sum of all entries in any column of Eq (3.40) should be unity.

As a consequence, for the symmetry described we have:

$$C = \text{Max } [H(X) - H(X|Y)] = \text{Max } H(X) - h'$$

Or $$C = \log m - h' \qquad \text{............(4.42)}$$

Thus the channel capacity for a symmetric channel may be computed in a very simple and straightforward manner. Usually the channel will be specified by its noise characteristics and the source probabilities [i.e. $P(Y|X)$ and $P(X)$]. Hence it will be a matter of simple inspection to identify the first form of symmetry described. To identify the second form of symmetry you have to first compute $P(X|Y)$ – tedious!

Example 3.4:

Consider the channel represented by the channel diagram shown in Fig 3.3:

The channel matrix can be read off from the channel diagram as:

$$P(Y|X) = \begin{matrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{3} & \frac{1}{3} \end{matrix}$$

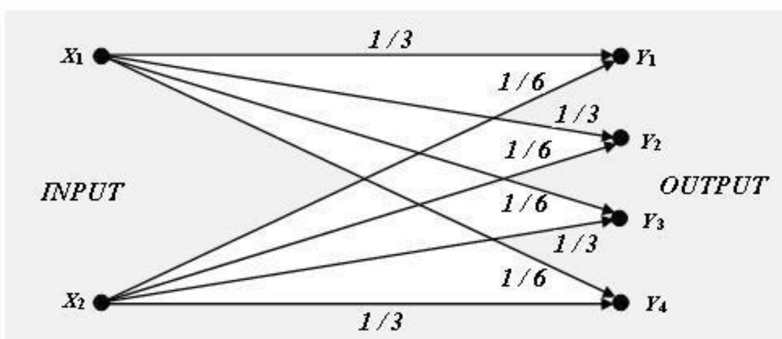


Fig 4.3 A Symmetric Channel

Clearly, the second row is a permutation of the first row (written in the reverse order) and hence the channel given is a symmetric channel. Accordingly we have, for the noise entropy, h (from either of the rows):

$$H(Y|X) = h = 2 \times \frac{1}{3} \log 3 + 2 \times \frac{1}{6} \log 6 = 1.918295834 \text{ bits / symbol.}$$

$$C = \log n - h = \log 4 - h = 0.081704166 \text{ bits / symbol.}$$

Example 4.5:

A binary channel has the following noise characteristic:

$$\begin{array}{c|cc} & 0 & 1 \\ \hline 0 & \frac{2}{3} & \frac{1}{3} \\ 1 & \frac{1}{3} & \frac{2}{3} \end{array}$$

(a) If the input symbols are transmitted with probabilities ***3 / 4*** and ***1 / 4*** respectively, find $H(X)$, $H(Y)$, $H(X, Y)$, $H(Y|X)$ and $I(X, Y)$.
(b) Find the channel capacity, efficiency and redundancy of the channel.
(c) What are the source probabilities that correspond to the channel capacity?

To avoid confusion, let us identify the input symbols as x_1 and x_2 and the output symbols by y_1 and y_2. Then we have:

$P(x_1) = 3/4$ and $p(x_2) = 1/4$

$$P(X|Y) = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} \end{bmatrix}$$

$$H(Y|X) = h = \frac{2}{3}\log\frac{3}{2} + \frac{1}{3}\log 3 = \log 3 - \frac{2}{3} = 0.918295833 \text{ bits / symbol}.$$

$$H(X) = \frac{3}{4}\log\frac{4}{3} + \frac{1}{4}\log 4 = \log 4 - \frac{3}{4}\log 3 = 2 - \frac{3}{4}\log 3 = 0.811278125 \text{ bits / symbol}.$$

Multiplying first row of $P(Y|X)$ by $p(x_1)$ and second row by $p(x_2)$ we get:

$$P(X,Y) = \begin{bmatrix} \frac{2}{3}\times\frac{3}{4} & \frac{1}{3}\times\frac{3}{4} \\ \frac{1}{3}\times\frac{1}{4} & \frac{2}{3}\times\frac{1}{4} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} \\ \frac{1}{12} & \frac{1}{6} \end{bmatrix}$$

Adding the elements of this matrix columnwise, we get: $p(y_1) = 7/12$, $p(y_2) = 5/12$.

Dividing the first column entries of $P(X, Y)$ by $p(y_1)$ and those of second column by $p(y_2)$, we get:

$$P(X|Y) = \begin{bmatrix} \frac{6}{7} & \frac{3}{5} \\ \frac{1}{7} & \frac{2}{5} \end{bmatrix}$$

From these values we have:

$$H(Y) = \frac{7}{12}\log\frac{12}{7} + \frac{5}{12}\log\frac{12}{5} = 0.979868756 \text{ bits / symbol}.$$

$$H(X,Y) = \frac{1}{2}\log 2 + \frac{1}{4}\log 4 + \frac{1}{12}\log 12 + \frac{1}{6}\log 6 = 1.729573958 \text{ bits / symbol}.$$

$$H(X|Y) = \frac{1}{2}\log\frac{7}{6} + \frac{1}{4}\log\frac{5}{3} + \frac{1}{12}\log 7 + \frac{1}{6}\log\frac{5}{2} = 0.74970520 \text{ bits / symbol}$$

$$H(Y|X) = \frac{1}{2}\log\frac{3}{2} + \frac{1}{4}\log 3 + \frac{1}{12}\log 3 + \frac{1}{6}\log\frac{3}{2} = \log 3 - \frac{2}{3} = h \quad (as before).$$

$$I(X,Y) = H(X) - H(X|Y) = 0.061572924 \text{ bits / symbol}.$$

$$= H(Y) - h = 0.061572923 \text{ bits / symbol}.$$

$$C = \log n - h = \log 2 - h - 1 - h = 0.081704167 \text{ bits / symbol}.$$

$$Efficiency, \eta = \frac{I(X,Y)}{C} = 0.753608123 \text{ or } 75.3608123\%$$

$$Re\,dundancy, E = 1 - \eta = 0.246391876 \text{ or } 24.6391876\%$$

To find the source probabilities, let $p(x_1) = p$ and $p(x_2) = q = 1 - p$. Then the ***JPM*** becomes:

$$P(X,Y) = \begin{matrix} \frac{2}{3}p, & \frac{1}{3}p \\ \frac{1}{3}(1-p), & \frac{2}{3} \end{matrix}$$

Adding columnwise we get: $p(y_1) = \frac{1}{3}(1+p)$ ***and*** $p(y_2) = \frac{1}{3}(2-p)$

For $H(Y) = H(Y)_{max}$, we want $p(y_1) = p(y_2)$ and hence $1+p = 2-p$ or $p = \frac{1}{2}$

Therefore the source probabilities corresponding to the channel capacity are: $p(x_1) = 1/2 = p(x_2)$.

Binary Symmetric Channels (BSC): (Problem 2.6.2 – S imon Haykin)

The channel considered in Example 3.6 is called a 'Binary Symmetric Channel' or (***BSC***). It is one of the most common and widely used channels. The channel diagram of a ***BSC*** is shown in Fig 3.4. Here ' ***p***' is called the error probability.

For this channel we have:

$$H(Y|X) = p\log\frac{1}{p} + q\log\frac{1}{q} = H(p) \qquad (4.43)$$

$$H(Y) = [p - \alpha(p-q)]\log\frac{1}{[p-\alpha(p-q)]} + [q + \alpha(p-q)]\log\frac{1}{[q+\alpha(p-q)]} \qquad \ldots(4.44)$$

$I(X, Y) = H(Y) - H(Y|X)$ and the channel capacity is:

$$C = 1 + p\log p + q\log q \qquad \ldots\ldots\ldots\ldots(4.45)$$

This occurs when $\alpha = 0.5$ *i.e.* $P(X=0) = P(X=1) = 0.5$

In this case it is interesting to note that the equivocation, $H(X|Y) = H(Y|X)$.

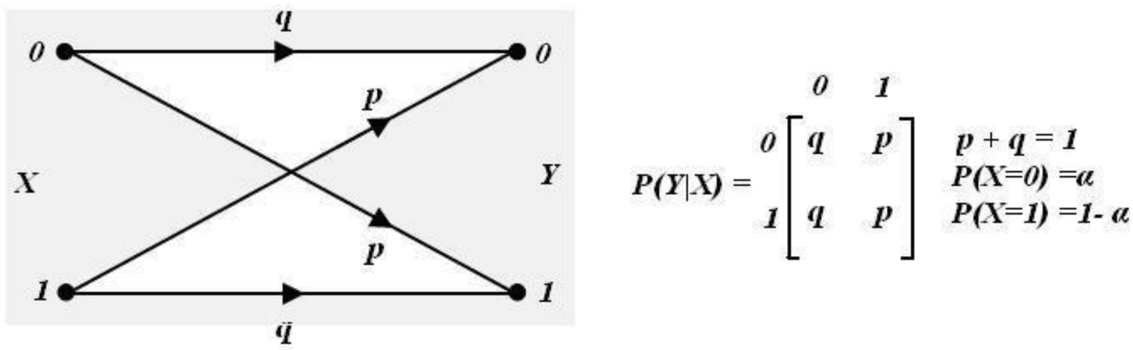


Fig 4.4 *Channel diagram of a BSC and its Channel Matrix.*

An interesting interpretation of the equivocation may be given if consider an idealized communication system with the above symmetric channel as shown in Fig 4.5.

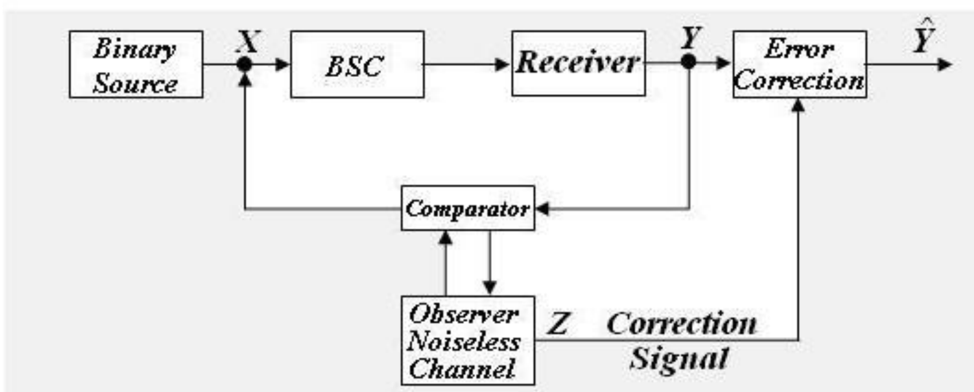


Fig 4.5 *An Idealized Binary Communication System.*

The observer is a noiseless channel that compares the transmitted and the received symbols. Whenever there is an error a ' ***1*** ' is sent to the receiver as a correction signal and appropriate correction is effected. When there is no error the observer transmits a ' ***0*** ' indicating no change. Thus the observer supplies additional information to the receiver, thus compensating for the noise in the channel. Let us compute this additional information .With ***P (X=0) = P (X=1) = 0.5***, we have:

Probability of sending a '1' = Probability of error in the channel .

Probability of error = P (Y=1|X=0).P(X=0) + P (Y=0|X=1).P(X=1)

$$= p \times 0.5 + p \times 0.5 = p$$

$$\therefore \text{Probability of no error} = 1 - p = q$$

Thus we have P ***(Z = 1) = p*** and ***P (Z = 0) =q***

Accordingly, additional amount of information supplied is:

$$= p \log \frac{1}{p} + q \log \frac{1}{q} = H\,(X \mid Y) = H\,(Y \mid X) \qquad \text{........ (4.46)}$$

Thus the additional information supplied by the observer is exactly equal to the equivocation of the source. Observe that if ' ***p*** ' and ' ***q*** ' are interchanged in the channel matrix, the trans -information of the channel remains unaltered. The variation of the mutual information with the probability of error is shown in Fig 3.6(*a*) for ***P (X=0) = P (X=1) = 0.5***. In Fig 4.6(*b*) is shown the dependence of the mutual information on the source probabilities.

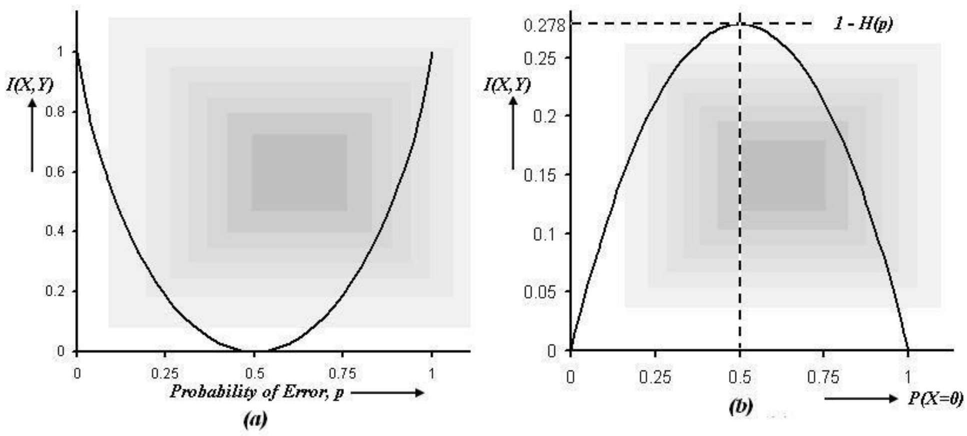


Fig 4.6 *Mutual Information of a BSC*

Binary Erasure Channels (*BEC*):

The channel diagram and the channel matrix of a ***BEC*** are shown in Fig 3.7.

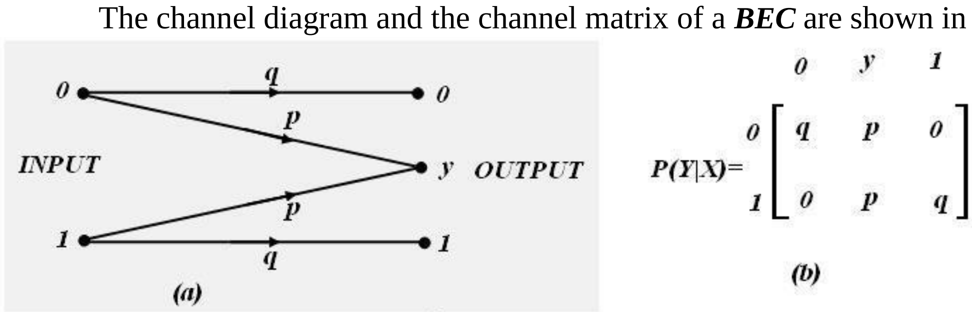


Fig 4.7 *Binary Erasure Channel.*

BEC is one of the important types of channels used in digital communications. Observe that whenever an error occurs, the symbol will be received as ' ***y***' and no decision will be made about the information but an immediate request will be made for retransmission, rejecting what have been received (***ARQ*** techniques), thus ensuring ***100%*** correct data recovery. Notice that this channel also is a symmetric channel and we have with ***P(X = 0) =*** α, ***P(X = 1) = 1 -*** α.

$$H(Y|X) = p\log\frac{1}{p} + q\log\frac{1}{q} \qquad \ldots\ldots\ldots\ldots\ldots\ldots \qquad (4.47)$$

$$H(X) = \alpha\log\frac{1}{\alpha} + (1-\alpha)\log\frac{1}{(1-\alpha)} \qquad \ldots\ldots\ldots\ldots\ldots\ldots \qquad (4.48)$$

The ***JPM*** is obtained by multiplying first row of ***P (Y|X)*** by α and second row by ***(1–*** α***).***
We get:

$$P(X,Y) = \begin{bmatrix} q\alpha & p\alpha & 0 \\ 0 & p(1-\alpha) & q(1-\alpha) \end{bmatrix} \qquad \ldots\ldots\ldots\ldots\ldots\ldots \qquad (4.49)$$

Adding column wise we get: ***P (Y)*** = $[q\alpha,\ p,\ q(1-\alpha)]$ .... (4.50)
From which the CPM ***P (X|Y)*** is computed as:

$$P(X|Y) = \begin{matrix} 1 & \alpha & 0 \\ 0 & (1-\alpha) & 1 \end{matrix} \qquad \ldots\ldots\ldots\ldots\ldots\ldots \qquad (4.51)$$

$$\therefore H(X|Y) = \alpha q \log 1 + \alpha p \log \frac{1}{\alpha} + (1-\alpha) p \log \frac{1}{(1-\alpha)} + (1-\alpha) q \log 1$$

$$= pH(X)$$

$$\therefore I(X, Y) = H(X) - H(X|Y) = (1-p)\, H(X) = q\, H(X) \qquad \ldots\ldots \qquad (4.52)$$

$$\therefore C = Max\ I(X, Y) = q \text{ bits / symbol.} \qquad \ldots\ldots \qquad (4.53)$$

In this particular case, use of the equation ***I(X, Y) = H(Y) – H(Y | X)*** will not be correct, as ***H(Y)*** involves ‘ ***y***’ and the information given by ‘ ***y***’ is rejected at the receiver.

Deterministic and Noiseless Channels: (Additional Information)

Suppose in the channel matrix of Eq (3.34) we make the following modifications.

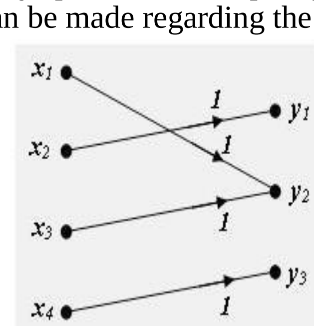
a) Each row of the channel matrix contains one and only one nonzero entry, which necessarily should be a ‘ ***1***’. That is, the channel matrix is symmetric and has the property, for a given k and j, $P(y_j|x_k) = 1$ and all other entries are ‘ 0’. Hence given x_k, probability of receiving it as y_j is ***one***. For such a channel, clearly

$$H(Y|X) = 0 \text{ and } I(X, Y) = H(Y) \qquad \ldots\ldots\ldots\ldots\ldots \qquad (4.54)$$

Notice that it is not necessary that ***H(X) = H(Y)*** in this case. The channel with such a property will be called a ‘ ***Deterministic Channel***’.

Example 4.6:

Consider the channel depicted in Fig 3.8. Observe from the channel diagram shown that the input symbol x_k uniquely specifies the output symbol y_j with a probability one. By observing the output, no decisions can be made regarding the transmitted symbol!!



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

Fig 4.8

b) Each column of the channel matrix contains one and only one nonzero entry. In this case, since each column has only one entry, it immediately follows that the matrix $P(X|Y)$ has also one and only one non zero entry in each of its columns and this entry, necessarily be a ‘ ***1***’ because:

If $p(y_j|x_k) = \alpha,\ p(y_j \mid x_r) = 0,\ r \neq k,\ r = 1, 2, 3 \cdots m.$

Then $p(x_k, y_j) = p(x_k) \times p(y_j|x_k) = \alpha \times p(x_k)$,

$$p\ (x_r,\ y_j)\ =\ 0,\ r \neq k,\quad r = 1, 2, 3\cdots\ m.$$

$$\therefore\ p\ (y_j)\ =\ \sum_{r=1}^{m} p(\ x_r,\ y_j\ )\ =\ p\ (x_k,\ y_j)\ =\ \alpha\, p\ (x_k)$$

$$\therefore\ p(\,x_k \mid y_j\,) = \frac{p(x_k, y_j)}{p(\,y_j\,)} = 1, \text{ and } p(\,x_r \mid y_j\,) = 0\,,\quad \forall r \neq k\,, r = 1, 2\,, 3, \ldots m\,.$$

It then follows that $H\ (X|Y)\ =\ 0$ and $I\ (X,\ Y)\ =\ H\,(X)$ ……... (4.55)

Notice again that it is not necessary to have $H(Y)\ =\ H(X)$. However in this case, converse of (***a***) holds. That is one output symbol uniquely specifies the transmitted symbol, whereas for a given input symbol we cannot make any decisions about the received symbol. The situation is exactly the complement or mirror image of (***a***) and we call this channel also a deterministic channel (some people call the channel pertaining to case (***b***) as 'Noiseless Channel', a classification can be found in the next paragraph). Notice that for the case (***b***), the channel is symmetric with respect to the matrix $P\ (X|Y)$.

Example 3.7:

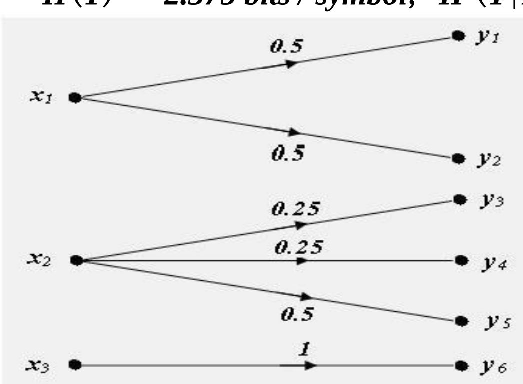
Consider the channel diagram, the associated channel matrix, $P\ (Y|X)$ and the conditional probability matrix $P\ (X|Y)$ shown in Fig 3.9. For this channel, let

$p\ (x_1)=0.5,\ \ p(x_2)\ =\ p(x_3)\ =\ 0.25.$

Then $p\ (y_1)\ =\ p\ (y_2)\ =\ p(y_6)\ =0.25,\ \ p(y_3)\ =\ p(y_4)\ =0.0625$ and $p(y_5)\ =\ 0.125.$

It then follows $I(X,\ \ Y)\ =\ H(X)\ =1.5$ ***bits / symbol***,

$H(Y)\ =\ 2.375$ ***bits / symbol***, $H\ (Y|X)\ =\ 0.875$ ***bits / symbol*** and $H\ (X|Y)\ =\ 0.$



$$P(Y|X) = \begin{bmatrix} 0.5 & 0.5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.25 & 0.25 & 0.5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\therefore\ P(X|Y) = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Fig 4.9 *Channel diagram for Example 4.7*

(Irrespective of P(X), the reader should verify this)

c) Now let us consider a special case: The channel matrix in Eq (3.34) is a square matrix and all entries except the one on the principal diagonal are zero. That is:

$$p\ (y_k|x_k) = 1 \text{ and } p(y_j|x_k)=0\ \forall k \neq j$$

Or in general, $p\ (y_j|x_k) = \delta_{jk}$, where δ_{jk}, is the '***Kronecker delta***', i.e. $\delta_{jk} = 1$ if $j = k$
$= 0$ if $j \neq k$.

That is, $P\ (Y|X)$ is an Identity matrix of order '***n***' and that $P\ (X|Y)\ =\ P\ (Y|X)$ and $p(x_k,\ \ y_j)\ =\ p(x_k)\ =\ p(y_j)$ can be easily verified.

For such a channel it follows:

$$H(X|Y) = H(Y|X) = 0 \text{ and } I(X, Y) = H(X) = H(Y) = H(X, Y) \qquad (4.56)$$
……

We call such a channel as " ***Noiseless Channel***". Notice that for the channel to be noiseless, it is necessary that there shall be a one-one correspondence between input and output symbols. No information will be lost in such channels and if all the symbols occur with equal probabilities, it follows then:

$$C = I(X, Y)_{Max} = H(X)_{Max} = H(Y)_{Max} = \log n \text{ bits / symbol.}$$

Thus a noiseless channel is symmetric and deterministic with respect to both descriptions ***P (Y|X)*** and ***P (X|Y)***.

Finally, observe the major concept in our classification. In case (***a***) for a given transmitted symbol, we can make a unique decision about the received symbol from the source end. In case (***b***), for a given received symbol, we can make a decision about the transmitted symbol from the receiver end. Whereas for case (***c***), a unique decision can be made with regard to the transmitted as well as the received symbols from either ends. This uniqueness property is vital in calling the channel as a 'Noiseless Channel'.

d) To conclude, we shall consider yet another channel described by the following ***JPM***:

$$P(X,Y) = \begin{bmatrix} p_1 & p_1 & p_1 & \ldots & p_1 \\ p_2 & p_2 & p_2 & \ldots & p_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ p_m & p_m & p_m & \ldots & p_m \end{bmatrix}$$

$$\text{with } \sum_{k=1}^{m} p_k = \frac{1}{n} \quad i.e.\ p(y_j) = \frac{1}{n},\ \forall\, j = 1, 2, 3, \ldots n.$$

This means that there is no correlation between x_k and y_j and an input x_k may be received as any one of the y_j's with equal probability. In other words, the input-output statistics are independent!!

This can be verified, as we have $p(x_k, y_j) = p_k$

$$= np_k . \sum_{k=1}^{m} p_k = p(x_k).p(y_j)$$

$\therefore\ p(x_k|y_j) = np_k$ and $p(y_j|x_k) = 1/n$

Thus we have:

$$H(X,Y) = n.\sum_{k=1}^{m} p_k \log \frac{1}{p_k}, \quad H(X) = \sum_{k=1}^{m} np_k \log \frac{1}{np_k} = n\sum_{k=1}^{m} p_k \log \frac{1}{p_k} + \log \frac{1}{n}$$

$$H(Y) = \sum_{j=1}^{n} p(y_j) \log \frac{1}{p(y_j)} = \log n,$$

$$H(X|Y) = H(X),\ H(Y|X) = H(Y) \text{ and } I(X,Y) =$$

∴Such a channel conveys no information whatsoever. Thus a channel with independent input-output structure is similar to a network with largest internal loss (purely resistive network), in contrast to a noiseless channel which resembles a lossless network.

Some observations:

For a deterministic channel the noise characteristics contains only one nonzero entry, which is a ‘ ***1***’, in each row or only one nonzero entry in each of its columns. In either case there exists a linear dependence of either the rows or the columns. For a noiseless channel the rows as well as the columns of the noise characteristics are linearly independent and further there is only one nonzero entry in each row as well as each column, which is a ‘ ***1***’ that appears only on the principal diagonal (or it may be on the skew diagonal). For a channel with independent input-output structure, each row and column are made up of all nonzero entries, which are all equal and equal to ***1/n***. Consequently both the rows and the columns are always linearly dependent!!

Franklin.M.Ingels makes the following observations:

1) If the channel matrix has only one nonzero entry in each column then the channel is termed as “ ***loss-less channel***”. True, because in this case $H\ (X|Y) = 0$ and $I(X, Y) = H(X)$, i.e. the mutual information equals the source entropy.

2) If the channel matrix has only one nonzero entry in each row (which necessarily should be a ‘ ***1’***), then the channel is called “ ***deterministic channel***”. In this case there is no ambiguity about how the transmitted symbol is going to be received although no decision can be made from the receiver end. In this case $H\ (Y|X) = 0$, and $I(X, Y) = H(Y)$.

3) An “ ***Ideal channel***” is one whose channel matrix has only one nonzero element in each row and each column, i.e. a diagonal matrix. An ideal channel is obviously both loss-less and deterministic. Lay man’s knowledge requires equal number of inputs and outputs-you cannot transmit 25 symbols and receive either 30 symbols or 20 symbols, there shall be no difference between the numbers of transmitted and received symbols. In this case

$$I(X,Y) = H(X) = H(Y); \text{ and } H(X|Y) = H(Y|X) = 0$$

4) A “ ***uniform channel***” is one whose channel matrix has identical rows ex cept for permutations OR identical columns except for permutations. If the channel matrix is square, then every row and every column are simply permutations of the first row.

Observe that it is possible to use the concepts of “ *sufficient reductions*” and make the channel described in (1) a deterministic one. For the case (4) observe that the rows and columns of the matrix (Irreducible) are linearly independent.

Additional Illustrations:

Example 3.9

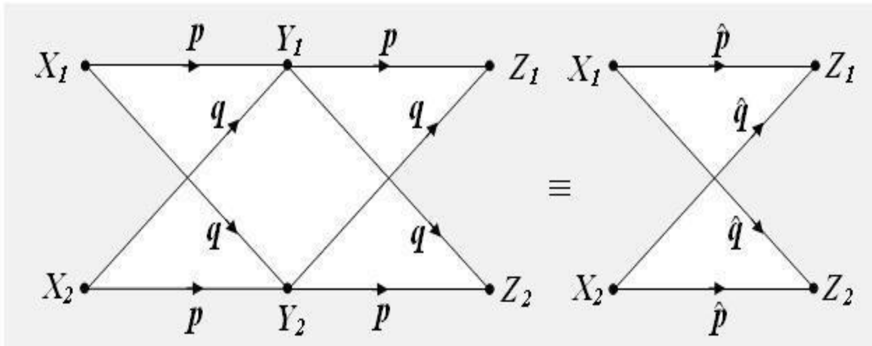


Fig 4.10 *Two BSC's in Cascade*

Consider two identical ***BSC's*** cascaded as shown in Fig 4.10. Tracing along the transitions indicated we find:

$$p(z_1|x_1) = p^2 + q^2 = (p+q)^2 - 2pq = (1-2pq) = p(z_2|x_2) \text{ and } p(z_1|x_2) = 2pq = p(z_2|x_1)$$

Labeling $\hat{p} = 1 - 2pq$, $\hat{q} = 2pq$ it then follows that:

$$I(X, Y) = 1 - H(q) = 1 + p \log p + q \log q$$
$$I(X, Z) = 1 - H(2pq) = 1 + 2pq \log 2pq + (1-2pq) \log (1-2pq).$$

If one more identical ***BSC*** is cascaded giving the output (u_1, u_2) we have:

$$I(X, U) = 1 - H(3pq^2 + p^3)$$

The reader can easily verify that $I(X, Y) \geq I(X, Z) \geq I(X, U)$

Example 4.9:

Let us consider the cascade of two noisy channels with channel matrices:

$$P(Y|X) = \begin{bmatrix} \frac{1}{6} & \frac{1}{6} & \frac{2}{3} \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \end{bmatrix} \quad P(Z|Y) = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{3} & \frac{2}{3} & 0 \\ 0 & \frac{1}{3} & \frac{2}{3} \end{bmatrix}, \text{ with } p(x_1) = p(x_2) = 0.5$$

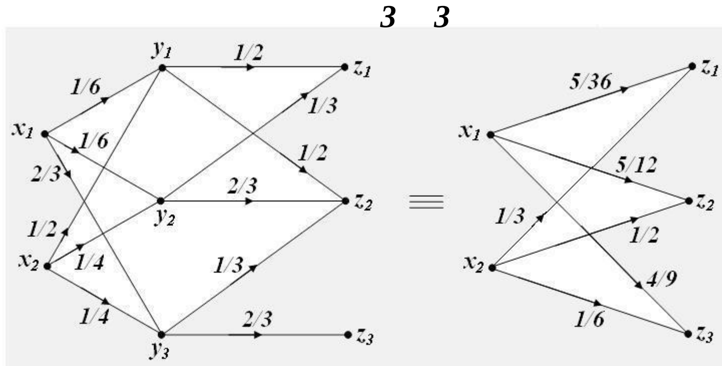


Fig 4.11 *Cascade of two noisy channels.*

The above cascade can be seen to be equivalent to a single channel with channel matrix:

$$P(Z|X) = \begin{bmatrix} \frac{5}{36} & \frac{5}{12} & \frac{4}{9} \\ \frac{1}{3} & \frac{1}{2} & \frac{1}{6} \end{bmatrix}$$

The reader can verify that: ***I(X, Y) = 0.139840072 bits / symbol.***

I(X, Z) = 0.079744508 bits / symbol.

Clearly ***I(X, Y) > I(X, Z).***

Example 3.10: Let us consider yet another cascade of noisy channels described by:

$$P(Y|X) = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix} \qquad P(Z|Y) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{2}{3} & \frac{1}{3} \\ 0 & \frac{1}{3} & \frac{2}{3} \end{bmatrix}$$

The channel diagram for this cascade is shown in Fig 4.12. The reader can easily verify in this case that the cascade is equivalent to a channel described by:

$$P(Z|X) = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix} = P(Y|X) \; ;$$

Inspite of the fact, that neither channel is noiseless, here we have ***I(X, Y) = I(X, Z).***

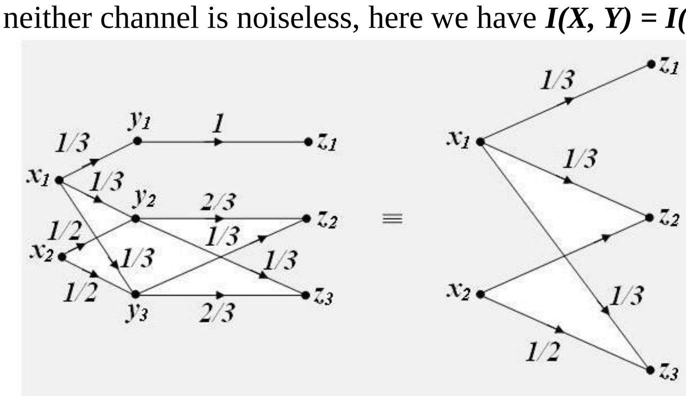


Fig 4.12 *Cascade of noisy channels of Example* **4.10**

Review Questions:

1. What are important properties of the codes?

2. what are the disadvantages of variable length coding?

3. Explain with examples:
4. Uniquely decodable codes, Instantaneous codes

5. Explain the Shannon-Fano coding procedure for the construction of an optimum code

6. Explain clearly the procedure for the construction of compact Huffman code.

7. A discrete source transmits six messages symbols with probabilities of 0.3, 0.2, 0.2, 0.15, 0.1, 0.5. Device suitable Fano and Huffmann codes for the messages and determine the average length and efficiency of each code.

8. Consider the messages given by the probabilities 1/16, 1/16, 1/8, ¼, ½. Calculate H. Use the Shannon-Fano algorithm to develop a efficient code and for that code, calculate the average number of bits/message compared with H.

9. Consider a source with 8 alphabets and respective probabilities as shown:

   A B C D E F G H
   0.20 0.18 0.15 0.10 0.08 0.05 0.02 0.01

   Construct the binary Huffman code for this. Construct the quaternary Huffman and code and show that the efficiency of this code is worse than that of binary code

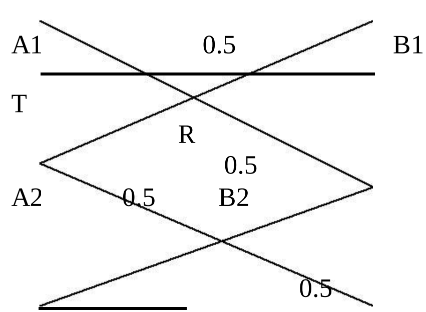
10. Define Noiseless channel and deterministic channel.

11. A source produces symbols X, Y,Z with equal probabilities at a rate of 100/sec. Owing to noise on the channel, the probabilities of correct reception of the various symbols are as shown:

P (j/i)	X	Y	z
X	¾	¼	0
y	¼	½	¼
z	0	¼	¾

Determine the rate at which information is being received.

12. Determine the rate of transmission l(x,y) through a channel whose noise characteristics is shown in fig. P(A1)=0.6, P(A2)=0.3, P(A3)=0.1



OUTCOME:

1. Able to understand different Communication channel in communication systems.
2. Capable of finding channel capacity of different channels in communication system.
3. Able to develop channel matrix and mutual information in channel.