Chapter Two

1- Channel:

In telecommunications and computer networking, a communication channel or **channel**, refers either to a physical transmission medium such as a wire, or to a logical connection over a multiplexed medium such as a radio channel. A channel is used to convey an information signal, for example a digital bit stream, from one or several *senders* (or transmitters) to one or several *receivers*. A channel has a certain capacity for transmitting information, often measured by its bandwidth in Hz or its data rate in bits per second.

2- Symmetric channel:

The symmetric channel have the following condition:

a- Equal number of symbol in X&Y, i.e. P(Y|X) is a square matrix.

b- Any row in P(Y|X) matrix comes from some permutation of other rows.

For example the following conditional probability of various channel types as shown:

a- $P(Y \mid X) = \begin{bmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{bmatrix}$ is a BSC, because it is square matrix and 1st row is the permutation of 2nd row.

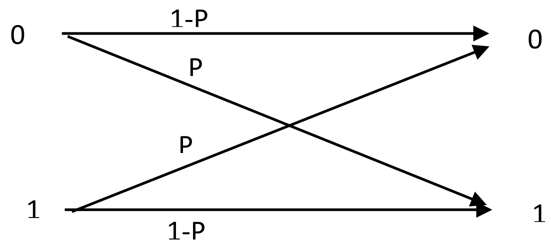
b- $P(Y \mid X) = \begin{bmatrix} 0.9 & 0.05 & 0.05 \\ 0.05 & 0.9 & 0.05 \\ 0.05 & 0.05 & 0.9 \end{bmatrix}$ is TSC, because it is square matrix and each row is a permutation of others.

c- $P(Y \mid X) = \begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0.1 & 0.8 & 0.1 \end{bmatrix}$ is a non-symmetric since since it is not square although each row is permutation of others.

d- $P(Y \mid X) = \begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0.1 & 0.7 & 0.2 \\ 0.1 & 0.1 & 0.8 \end{bmatrix}$ is a non-symmetric although it is square since 2nd row is not permutation of other rows.

2.1- Binary symmetric channel (BSC)

It is a common communications channel model used in coding theory and information theory. In this model, a transmitter wishes to send a bit (a zero or a one), and the receiver receives a bit. It is assumed that the bit is *usually* transmitted correctly, but that it will be "flipped" with a small probability (the "crossover probability").



A **binary symmetric channel with crossover probability *p*** denoted by *BSCp*, is a channel with binary input and binary output and probability of error *p*; that is, if *X* is the transmitted random variable and *Y* the received variable, then the channel is characterized by the conditional probabilities:

$$\Pr(Y = 0 \mid X = 0) = 1 - P$$

$$\Pr(Y = 0 \mid X = 1) = P$$

$$\Pr(Y = 1 \mid X = 0) = P$$

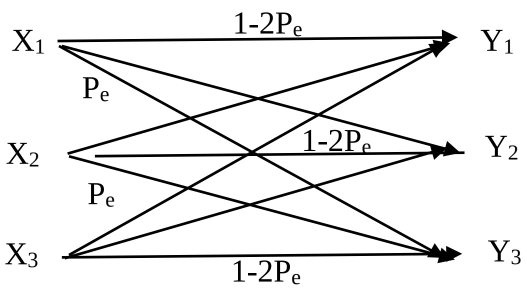
$$\Pr(Y = 1 \mid X = 1) = 1 - P$$

2.2- Ternary symmetric channel (TSC):

The transitional probability of TSC is:

$$P(Y \mid X) = \begin{matrix} \\ x_1 \\ x_2 \\ x_3 \end{matrix} \begin{bmatrix} y_1 & y_2 & y_3 \\ 1-2P_e & P_e & P_e \\ P_e & 1-2P_e & P_e \\ P_e & P_e & 1-2P_e \end{bmatrix}$$

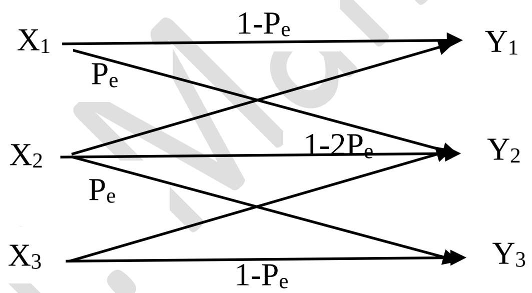
The TSC is symmetric but not very practical since practically x_1 and x_3 are not affected so much as x_2. In fact the interference between x_1 and x_3 is much less than the interference between x_1 and x_2 or x_2 and x_3.



Hence the more practice but nonsymmetric channel has the trans. prob.

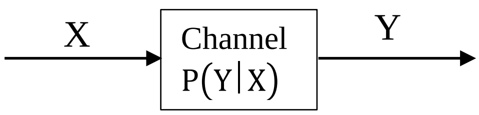
$$P(Y \mid X) = \begin{matrix} \\ x_1 \\ x_2 \\ x_3 \end{matrix} \begin{bmatrix} y_1 & y_2 & y_3 \\ 1-P_e & P_e & 0 \\ P_e & 1-2P_e & P_e \\ 0 & P_e & 1-P_e \end{bmatrix}$$

Where x_1 interfere with x_2 exactly the same as interference between x_2 and x_3, but x_1 and x_3 are not interfere.



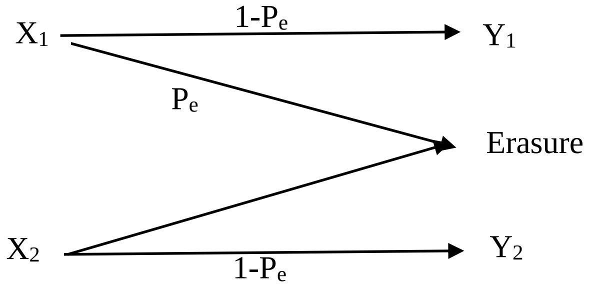
3- Discrete Memoryless Channel:

The Discrete Memoryless Channel (DMC) has an input X and an output Y. At any given time (t), the channel output Y= y only depends on the input X = x at that time (t) and it does not depend on the past history of the input. DMC is represented by the conditional probability of the output Y = y given the input X = x, or $P(Y|X)$.



4- Binary Erasure Channel (BEC):

The Binary Erasure Channel (BEC) model are widely used to represent channels or links that "losses" data. Prime examples of such channels are Internet links and routes. A BEC channel has a binary input X and a ternary output Y.



Note that for the BEC, the probability of "bit error" is zero. In other words, the following conditional probabilities hold for any BEC model:

$$\Pr(Y = \text{"}erasure\text{"} \mid X = 0) = P$$

$$\Pr(Y = \text{"}erasure\text{"} \mid X = 1) = P$$

$$\Pr(Y = 0 \mid X = 0) = 1 - P$$

$$\Pr(Y = 1 \mid X = 1) = 1 - P$$

$$\Pr(Y = 0 \mid X = 1) = 0$$

$$\Pr(Y = 1 \mid X = 0) = 0$$

5- Special Channels:

a- Lossless channel: It has only one nonzero element in each column of the transitional matrix P(Y|X).

$$P(Y \mid X) = \begin{matrix} \\ x_1 \\ x_2 \\ x_3 \end{matrix} \begin{bmatrix} y_1 & y_2 & y_3 & y_4 & y_5 \\ 3/4 & 1/4 & 0 & 0 & 0 \\ 0 & 0 & 1/3 & 2/3 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

This channel has H(X|Y)=0 and I(X, Y)=H(X) with zero losses entropy.

b- Deterministic channel: It has only one nonzero element in each row, the transitional matrix P(Y|X), as an example:

$$P(Y \mid X) = \begin{matrix} x_1 \\ x_2 \\ x_3 \end{matrix} \begin{bmatrix} y_1 & y_2 & y_3 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

This channel has H(Y|X)=0 and I(Y, X)=H(Y) with zero noisy entropy.

c- Ideal channel: It has only one nonzero element in each row and column, the transitional matrix P(Y|X), i.e. it is an identity matrix, as an example:

$$P(Y \mid X) = \begin{matrix} x_1 \\ x_2 \\ x_3 \end{matrix} \begin{bmatrix} y_1 & y_2 & y_3 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

This channel has H(Y|X)= H(X|Y)=0 and I(Y, X)=H(Y)=H(X).

d- Noisy channel: No relation between input and output:

$$\mathrm{H}(\mathrm{X} \mid \mathrm{Y}) = H(Y), \quad \mathrm{H}(\mathrm{Y} \mid \mathrm{X}) = H(X)$$

$$I(X, Y) = 0, \; C = 0$$

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

6- Shannon's theorem:

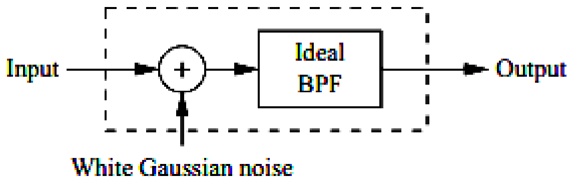
a- A given communication system has a maximum rate of information C known as the channel capacity.

b- If the information rate R is less than C, then one can approach arbitrarily small error probabilities by using intelligent coding techniques.

c- To get lower error probabilities, the encoder has to work on longer blocks of signal data. This entails longer delays and higher computational requirements.

Thus, if R ≤ C then transmission may be accomplished without error in the presence of noise. The negation of this theorem is also true: if R > C, then errors cannot be avoided regardless of the coding technique used.

Consider a bandlimited Gaussian channel operating in the presence of additive Gaussian noise:



The Shannon-Hartley theorem states that the channel capacity is given by:

$$C = Blog_2\left(1 + \frac{S}{N}\right)$$

Where C is the capacity in bits per second, B is the bandwidth of the channel in Hertz, and S/N is the signal-to-noise ratio.

7- Channel Capacity (Discrete channel)

This is defined as the maximum of I(X,Y):

$$C = channel\ capacity = \max[I(X,Y)] \qquad bits/symbol$$

Physically it is the maximum amount of information each symbol can carry to the receiver. Sometimes this capacity is also expressed in bits/sec if related to the rate of producing symbols r:

$$R(X,Y) = r \times I(X,Y) \qquad bits/\sec \quad or\ R(X,Y) = I(X,Y)/\bar{\tau}$$

a- Channel capacity of Symmetric channels:

The channel capacity is defined as $\max[I(X,Y)]$:

$$I(X,Y) = H(Y) - H(Y \mid X)$$

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

But we have

$$P(x_i, y_j) = P(x_i)P(y_j \mid x_i) \qquad \textit{put in above equation yieldes}:$$

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

If the channel is symmetric the quantity:

$$\sum_{j=1}^{m} P(y_j \mid x_i)\log_2 P(y_j \mid x_i) = K$$

Where K is constant and independent of the row number *i* so that the equation becomes:

$$I(X,Y) = H(Y) + K\sum_{i=1}^{n} P(x_i)$$

Hence $\quad I(X,Y) = H(Y) + K \quad$ for symmetric channels

Max of $I(X,Y) = \max[H(Y) + K] = \max[H(Y)] + K$

When Y has equiprobable symbols then $\max[H(Y)] = log_2 m$

Then

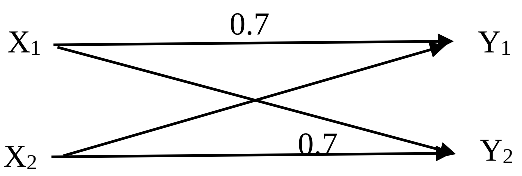
$$I(X,Y) = log_2 m + K$$

Or

$$C = log_2 m + K$$

Example 9:

For the BSC shown:



Find the channel capacity and efficiency if $I(x_1) = 2bits$

Solution:

$$P(Y \mid X) = \begin{bmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{bmatrix}$$

Since the channel is symmetric then

$C = log_2 m + K$ and $n = m$

$where\ n\ and\ m\ are\ number\ row\ and\ column\ repestively$

$$K = 0.7log_2 0.7 + 0.3log_2 0.3 = -0.88129$$

$$C = 1 - 0.88129 = 0.1187 \quad bits/symbol$$

The channel efficiency $\eta = \frac{I(X,Y)}{C}$

$$I(x_1) = -log_2 P(x_1) = 2$$

$$P(x_1) = 2^{-2} = 0.25 \quad then\ P(X) = [0.25 \quad 0.75]^T$$

And we have $P(x_i, y_j) = P(x_i)P(y_j \mid x_i)$ so that

$$P(X,Y) = \begin{bmatrix} 0.7 \times 0.25 & 0.3 \times 0.25 \\ 0.3 \times 0.75 & 0.7 \times 0.75 \end{bmatrix} = \begin{bmatrix} 0.175 & 0.075 \\ 0.225 & 0.525 \end{bmatrix}$$

$$P(Y) = [0.4 \quad 0.6] \quad \rightarrow H(Y) = 0.97095\ bits/symbol$$

$$I(X,Y) = H(Y) + K = 0.97095 - 0.88129 = 0.0896\ bits/symbol$$

Then $\eta = \frac{0.0896}{0.1187} = 75.6\%$

To find the channel redundancy:

$$R = 1 - \eta = 1 - 0.756 = 0.244 \quad or\ 24.4\%$$

1- Channel capacity of nonsymmetric channels:

We can find the channel capacity of nonsymmetric channel by the following steps:

a- Find I(X, Y) as a function of input prob:

$$I(X,Y) = f(P(x_1), P(x_2) \ldots \ldots \ldots, P(x_n))$$

And use the constraint to reduce the number of variable by 1.

b- Partial differentiate I(X, Y) with respect to (n-1) input prob., then equate these partial derivatives to zero.

c- Solve the (n-1) equations simultaneously then find $P(x_1), P(x_2) \ldots \ldots, P(x_n)$ that gives maximum I(X, Y).

d- Put resulted values of input prob. in the function given in step 1 to find $C = \max[I(X,Y)]$.

Example 10:

Find the channel capacity for the channel having the following transition:

$$P(Y \mid X) = \begin{bmatrix} 0.7 & 0.3 \\ 0.1 & 0.9 \end{bmatrix}$$

Assume that $P(x_1) = P \cong 0.862$

Solution: First not that the channel is not symmetric since the 1st row is not permutation of 2nd row.

a- Let $P(x_1) = 0.862 \quad , \quad then\ P(x_2) = 1 - P = 0.138.$

$$P(X,Y) = P(X) \times P(Y \mid X)$$

$$\therefore \quad P(X,Y) = \begin{bmatrix} 0.7 \times 0.862 & 0.3 \times 0.862 \\ 0.1(1 - 0.862) & 0.9(1 - 0.862) \end{bmatrix}$$

$$= \begin{bmatrix} 0.6034 & 0.2586 \\ 0.0138 & 0.1242 \end{bmatrix}$$

From above results $P(Y) = [0.6172 \quad 0.3828]$

$$H(Y) = -\sum_{j=1}^{m} P(y_j) \log_2 P(y_j)$$

$$\therefore H(Y) = 0.960 \ \ bits/symbol$$

We have

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

$$H(Y \mid X) = -[0.6034\ ln0.7 + 0.2586 ln0.3 + 0.0138 ln0.1 + 0.1242 ln0.9]/ln$$

$$H(Y \mid X) = 0.8244 \quad \text{bits/symbol}$$

$$C = \max[I(X,Y)] = H(Y) - H(Y \mid X)$$

$$= 0.96021 - 0.8244 = 0.1358\ bits/symbol.$$

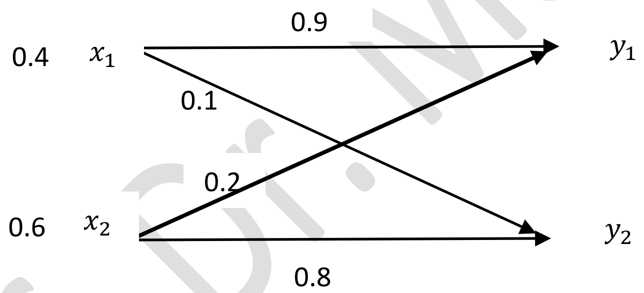
Review questions:

A binary source sending x_1 with a probability of 0.4 and x_2 with 0.6 probability through a channel with a probabilities of errors of 0.1 for x_1 and 0.2 for x_2.Determine:

1- Source entropy.
2- Marginal entropy.
3- Joint entropy.
4- Conditional entropy $H(Y \mid X)$.
5- Losses entropy $H(X \mid Y)$.
6- Transinformation.

Solution:

1- The channel diagram:



Or $P(Y \mid X) = \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix}$

$$H(X) = -\sum_{i=1}^{n} p(x_i)\, log_2 p(x_i)$$

$$H(X) = -\frac{[0.4\ln(0.4) + 0.6\ln(0.6)]}{ln2} = 0.971\frac{bits}{symbol}$$

2- $P(X,Y) = P(Y \mid X) \times P(X)$

$$\therefore P(X,Y) = \begin{bmatrix} 0.9 \times 0.4 & 0.1 \times 0.4 \\ 0.2 \times 0.6 & 0.8 \times 0.6 \end{bmatrix} = \begin{bmatrix} 0.36 & 0.04 \\ 0.12 & 0.48 \end{bmatrix}$$

$$\therefore P(Y) = [0.48 \quad 0.52]$$

$$H(Y) = -\sum_{j=1}^{m} p(y_j)\, log_2 p(y_j)$$

$$H(Y) = -\frac{[0.48\ln(0.48) + 0.52\ln(0.52)]}{\ln(2)} = 0.999\ bits/symbol$$

3- $H(X,Y)$

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

$$H(X,Y) = -\frac{[0.36\ln(0.36) + 0.04\ln(0.04) + 0.12\ln(0.12) + 0.48\ln(0.48)]}{\ln(2)}$$

$$= 1.592\ bits/symbol$$

4- $H(Y|\mathrm{X})$

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

$$H(Y|\mathrm{X}) = -\frac{[0.36\ln(0.9) + 0.12\ln(0.2) + 0.04\ln(0.1) + 0.48\ln(0.8)]}{\ln(2)}$$

$$= 0.621\frac{bits}{symbol}$$

Or $\quad H(Y|\mathrm{X}) = H(X,Y) - H(X) = 1.592 - 0.971 = 0.621\frac{bits}{symbol}$

5- $H(X|\mathrm{Y}) = H(X,Y) - H(Y) = 1.592 - 0.999 = 0.593\ bits/symbol$

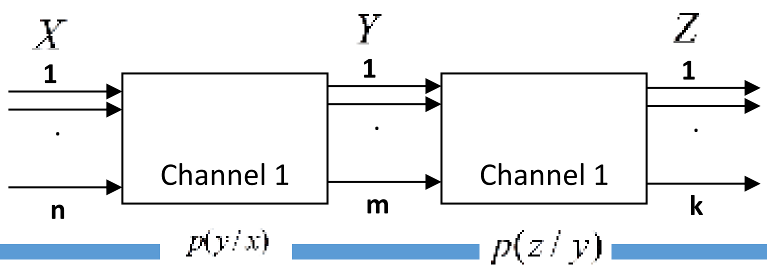
6- $I(X,Y) = H(X) - H(X|\mathrm{Y}) = 0.971 - 0.593 = 0.378$ bits/symbol

2- Cascading of Channels

If two channels are cascaded, then the overall transition matrix is the product of the two transition matrices.

$$p(z/x) = p(y/x).p(z/y)$$

$(n\times k)$ *matrix* $\quad (n\times m)$ *matrix* $\quad (m\times k)$ *matrix*

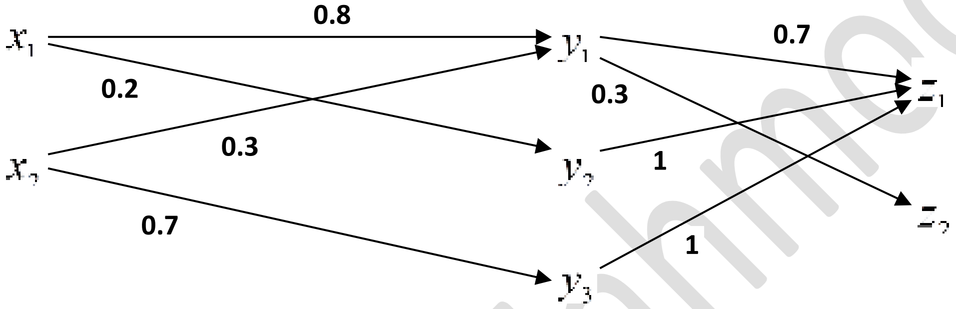


For the series information channel, the overall channel capacity is not exceed any of each channel individually.

$$I(X,Z) \le I(X,Y) \quad \& \quad I(X,Z) \le I(Y,Z)$$

Example:

Find the transition matrix $p(z/x)$ for the cascaded channel shown.



$$p(y/x) = \begin{bmatrix} 0.8 & 0.2 & 0 \\ 0.3 & 0 & 0.7 \end{bmatrix}, \quad p(z/y) = \begin{bmatrix} 0.7 & 0.3 \\ 1 & 0 \\ 1 & 0 \end{bmatrix}$$

$$p(z/x) = \begin{bmatrix} 0.8 & 0.2 & 0 \\ 0.3 & 0 & 0.7 \end{bmatrix} \begin{bmatrix} 0.7 & 0.3 \\ 1 & 0 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 0.76 & 0.24 \\ 0.91 & 0.09 \end{bmatrix}$$