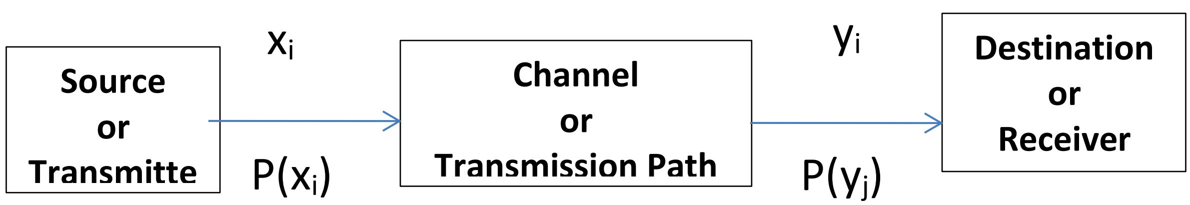
1- Introduction

In previous lecture we perform measure of information and how to calculate self-information, average information or entropy for given source. Also, we find the efficiency of the source according to the type of source discrete or continues.

In this lecture we are going to study the transfer of information from the source to the destination or the receiver via the channel. As described in Lecture#1, the channel or transmission path is the physical connection between the source and destination. The channel usually has negative effect on information transmission so that not all the information (or entropy) is transferred to the receiver, instead a portion of this information is discarded by the channel or the channel adds noise to the transferred information. The model is shown below;

$$p(x_i|y_j)\ or\ p(y_j|x_i)$$



x_i = the transmitted symbol $\qquad$ y_i = the received symbol

The channel is modelled as conditional probability $p(x_i|y_j)\ or\ p(y_j|x_i)$ for all values of x_i and y_j. For the simplest case of binary channel where we have;

$$X = \{ x_1 , x_2 \} \qquad Y = \{ y_1 , y_2 \}$$

We have four conditional probabilities $p(y_j|x_i)$ as follows:

$p(y_1\|x_1)$	Conditional probability of receiving y_1 by the receiver when the source produced x_1 or probability of correct reception of x_1
$p(y_2\|x_2)$	Conditional probability of receiving y_2 by the receiver when the source produced x_2 or probability of correct reception of x_2
$p(y_2\|x_1)$	Conditional probability of receiving y_2 by the receiver when the source produced x_1 or probability of incorrect transition of x_1
$p(y_1\|x_2)$	Conditional probability of receiving y_1 by the receiver when the source produced x_2 or probability of incorrect transition of x_2

E-mail: ak.kadhim@uomus.edu.iq

The above conditional probabilities together with $p(x_i)$, $p(y_j)$, and $p(x_i\,, y_j)$ fully describe the information transmission over the channel.

2- Joint Expectation

The first step towards finding the information transmission over given channel is to average the information when x_i is transmitted and y_j is received. Thus, we have joint experiment and it is required to find the expected values of certain terms (joint expectation);

$$E[h(x,y)] = \sum_{y}\sum_{x} P(x_i, y_j)\,.\,h(x_i, y_j)$$

And for continues joint expectation;

$$E[h(x,y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y).\,h(x,y)\;dxdy$$

3- The Average Mutual Information

The average mutual information (I) is the overall useful information transferred to the receiver from the source. It does not equal to the source information since the effect of the channel usually is negative.

The self-information of the source and the receiver symbols are given by:

$$I_{x_i} = -Log(p(x_i)) \qquad I_{y_j} = -Log(p(y_j))$$

Define the channel conditional information in two different ways according to the conditional probability:

$$I_{xi|yj} = -\,Log\;p\left(x_i|y_j\right) \qquad I_{yj|xi} = -\,Log\;p\left(y_j|x_i\right)$$

The mutual information $I_{x:y}$ is the useful information transferred from the source to the destination can be determined by:

$$I_{x:y} = I_{yj} - I_{yj|xi} = -Log\;p(yj) + Log\;p(yj|xi) = Log\;\frac{p(yj|xi)}{p(yj)}$$

or $$I_{x:y} = I_x - I_{xi|yj} = -Log\, p(xi) + Log\, p(xi|yj) = Log\, \frac{p(xi|yj)}{p(xi)}$$

Taking the average (or expectation) of the mutual information gives the average mutual information (I):

$$I = E[I_{x:y}] = \sum_y \sum_x P(x_i, y_j)\, I_{x:y}$$

$$= \sum_y \sum_x P(x_i, y_j)\, (I_{yj} - I_{yj|xi})$$

$$= \sum_y \sum_x P(x_i, y_j)\, [-Log\, p(y_j) + Log\, p(y_j|x_i)]$$

The final equation can be written as : $$I = \sum_y \sum_x P(x_i, y_j)\, Log\, \frac{p(y_j|x_i)}{p(y_j)} \quad \text{......Eq.1}$$

Using the other mutual information gives: $$I = \sum_y \sum_x P(x_i, y_j)\, Log\, \frac{p(x_i|y_j)}{p(x_i)} \quad \text{....Eq.2}$$

Since I is the average mutual information then its unit is (b-unit of inf./symbol) like the entropy. (I) represents the useful information of the system that is produced by the source and received by the receiver. $0 < I \leq H(x)$, where portion of the information is either lost in the channel or else the channel add noise to the source information (noise is -ve information).

The average mutual information also called:

-Trans-information   -Net information   or   -Information

4- Entropies and Average Mutual Information

We already have the definition of source entropy H(x) which is the average of the self-information of the source. This can be extended to any other probability and its related self-information, like the receiver entropy that can be determined in terms of $p(y_j)$. The followings are the possible entropies of information transmission system:

Meaning	Name	I or Entropy Expression	Eq.No.
Average Mutual Information	I	$\sum_y \sum_x P(x_i, y_j)\ Log \frac{p(y_j \mid x_i)}{p(y_j)}$	1
		$\sum_y \sum_x P(x_i, y_j)\ Log \frac{p(x_i \mid y_j)}{p(x_i)}$	2
Source Entropy	H(x)	$-\sum_x P(x_i)\ Log P(x_i)$	3
Receiver Entropy	H(y)	$-\sum_y P(y_j)\ Log P(y_j)$	4
Conditional or Noise Entropy	H(y|x)	$-\sum_x \sum_y P(x_i, y_j)\ Log P(y_j \mid x_i)$	5
Conditional or Losses Entropy	H(x|y)	$-\sum_x \sum_y P(x_i, y_j)\ Log P(x_j \mid y_i)$	6
System or Joint Entropy	H(x,y)	$-\sum_x \sum_y P(x_i, y_j)\ Log P(x_i, y_j)$	7

Example-1 Show that I = H(x) – H(x|y)

Solution :

Taking the definition of I (L.H.S) using Eq.2 in above table:

$$I = \sum_y \sum_x P(x_i, y_j)\ Log \frac{p(x_i \mid y_j)}{p(x_i)}$$

$$= \sum_y \sum_x P(x_i, y_j)\ [\ Log p(x_i \mid y_j) - Log\ p(x_i)]$$

E-mail: ak.kadhim@uomus.edu.iq

$$= \sum_y \sum_x P(x_i, y_j)\, Log\, p(x_i|y_j) - \sum_y \sum_x P(x_i, y_j)\, Log\, p(x_i)$$

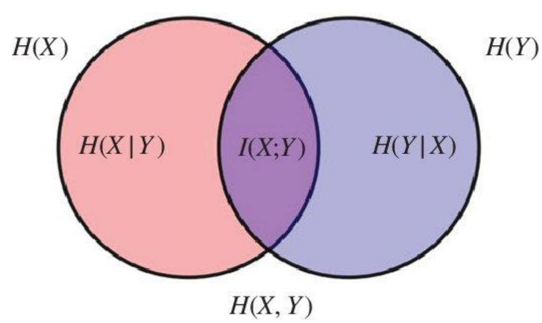
$$= \sum_y \sum_x P(x_i, y_j)\, Log\, p(x_i|y_j) - \sum_x \left(\sum_y P(x_i, y_j)\right) Log\, p(x_i)$$

$$= \sum_y \sum_x P(x_i, y_j)\, Log\, p(x_i|y_j) - \sum_x p(x_i)\, Log\, p(x_i)$$

∴ I = H(x) – H(x|y) using the definitions given by Eq.6 and Eq.3

Q- Similar to solution of above Example, show that:

1- I=H(y)-H(y|x)
2- H(x,y) = H(y) + H(x|y)
3- H(x,y) = H(x) + H(y|x)
4- H(x,y) = H(x) + H(y) – I
5- H(x,y) =I + H(y|x) + H(x|y



Notes on Entropies and I

1- To memorize the above relations, one may use the Venn diagram to represent all the relations described above between H and I.
2- The equation I = H(x) – H(x|y) means that the useful information of the system (or I) is given by the source information H(x) after subtracting the lost information in the channel H(x|y).
3- The equation I = H(y) – H(y|x) means that the useful information of the system (or I) is given by the received information H(y) after subtracting the noise information produced by the channel H(x|y).
4- The perfect system is the case where there is no noise and no losses in the channel (or H(y|x) =0 and H(x|y)=0), thus I=H(x) =H(y).
5- H(x,y) is the system entropy or the whole entropy of the present in the system.

E-mail: ak.kadhim@uomus.edu.iq

5- Discrete Channel Representation

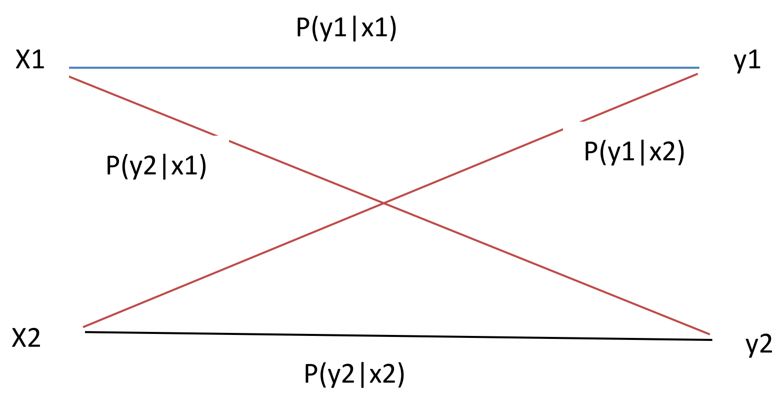
The discrete channel is usually described by the conditional probability $P(y_j|x_i)$. But other probabilities also used since there are different relations to determine this conditional probability.

Since $P(y_j|x_i)$ is two dimensional values then it also can be given in the form of matrix with rows indicating the values of x_i while y_j represents its columns. The summation of each row elements of $P(y_j|x_i)$ matrix is unity (1) as in the next example.

Example-2 Binary Channel

The channel is called binary because its source produces binary values. The source output actually is binary "0" or "1", but here we used x1 and x2 instead. The channel matrix and channel model of an example binary channel is given by:

$$P(y_j|x_i) = \begin{matrix} from/to & y1 & y2 \\ x1 & 0.8 & 0.2 \\ x2 & 0.1 & 0.9 \end{matrix}$$



In the above model the red transitions are the erroneous or incorrect transitions. For example: P(y1|x1) =0.8 means that 80% of x1 are received correct, while P(y2|x1) =0.2 means that 20% of the x1 are received incorrect. Similarly, for x2 90% are received correct, while 10% are received incorrect.

Assuming that p(x1)=0.6 and p(x2)=0.4, then it is possible to calculate all entropies and the average mutual information I as follows;

E-mail: ak.kadhim@uomus.edu.iq

Using the probability theorems to find all probabilities first:

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

The result for all xi and yj is given by: $P(x_i,y_j) =$

from/to	y1	y2
x1	0.48	0.12
x2	0.04	0.36

Using $p(y_j) = \sum_x P(x_i,y_j)$ gives $p(y_j) = [\,0.52 \quad 0.48]$

The above is obtained by the summation of column elements of $P(x_i,y_j)$.

Also, the conditional probability $p(x_i|y_j) = \frac{P(x_i,y_j)}{p(y_j)}$ can be determined:

$P(x_i|y_j) =$

from/to	y1	y2
x1	0.923	0.25
x2	0.077	0.75

Now we can use Eq.1 to Eq.7 given in the Table to find all entropies and I:

For example, to find H(y) one may use Eq.4

$$H(y) = -\sum_y P(y_j)\ LogP(y_j)$$

= - (0.52 Log 0.52 + 0.48 Log 0.48) = 0.9988 Bit/symbol (i.e Log2 is used)

Also H(y|x) can be determined from Eq.5 :

$$H(y|x) = -\sum_x \sum_y P(x_i,y_j)\ LogP(y_j|x_i)$$

= - (0.48 log 0.8 + 0.12 Log 0.2 + 0.04 Log 0.1 + 0.36 Log 0.9)

= 0.6207 Bit/symbol

E-mail: ak.kadhim@uomus.edu.iq

To find (I) we either use Eq.1 or 2 or from Venn diagram I=H(y)- H(y|x) ;

I= 0.9988-0.6207 = 0.3781 Bit/symbol

Q1-Find other entropies either from Eq.s or from Ven diagram if possible (in Bit/Symbol):

1- H(x) (Ans.= 0,971 Bit/symbol)
2- H(x|Y) (Ans.= 0.593 Bit/symbol)
3- H(x,y) (Ans.= 1.564 Bit/symbol)

Q2. For the following binary channel, if p(x1)= 0,6 and p(x2)=0.4, find all entropies and average mutual information (I) by using Eq.1 to 7 and Venn diagram relations.

$$P(y_j|x_i) = \begin{matrix} from/to & y1 & y2 \\ x1 & 0.9 & 0.1 \\ x2 & 0.1 & 0.9 \end{matrix}$$

6- Noiseless and Independent Transmission

Noiseless and Independent Transmission represent the two extreme cases of transferring information over given channel. Considering the general transmission of information defined by source symbols:

$X = \{ x_1 \quad x_2 \quad x_3 \quad x_4 \quad \ldots\ldots.. \; x_N \}$ with their probabilities, and the received symbols: $Y = \{ y_1 \quad y_2 \quad y_3 \quad y_4 \quad \ldots\ldots.. \; y_M \}$ with their probabilities. The channel probabilities are defined for all pairs $(x_i\,, y_j)$ by the conditional channel matrix representing the conditional probabilities $\{ p(y_j|x_i)\}$.

A. Noiseless Channel

Here the number of source symbols must equal the received symbols (N=M). The channel is called noiseless if the channel conditional probabilities are given by:

$$p(y_j|x_i) = 1 \quad for\ i = j$$

and

$$p(y_j|x_i) = 0 \quad for\ i \neq j$$

Using the probability relations, the above also imply that:

$$p(x_i, y_j) = p(x_i).p(y_j|x_i) = p(x_i) \quad for\ i = j$$

and

$$= 0 \qquad for\ i \neq j$$

Using Eq-5 in the Table gives:

$$H(y|x) = -\sum_x \sum_y P(x_i, y_j)\ LogP(y_j|x_i) = 0$$

Similarly, by using other equations in the table revealed that:

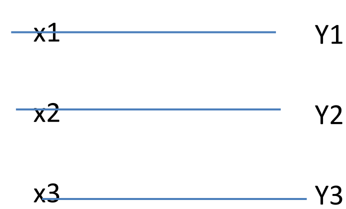
H(x|y)=0, I=H(x)=H(y)=H(x,y)

The Venn-diagram in this case will be represented by single circle representing the above entropies relation. The channel has nothing to do with the reduction of the information and the system is an optimum system in this case.

Example-3 Ternary Noiseless Channel

$$P(y_j|x_i) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

or



B. Independent Transmission

This is the opposite case of the noiseless channel, where the symbols $\{x_i\}$ are independent of the symbols $\{ y_j \}$, for all i and j. Thus,

$$p(x_i, y_j) = p(x_i).p(y_j)$$

Using Eq.1 $I = \sum_y \sum_x P(x_i, y_j)\, Log \frac{p(y_j|x_i)}{p(y_j)}$

$$= \sum_y \sum_x P(x_i, y_j)\, Log \frac{p(x_i,y_j)}{p(x_i).p(y_j)}$$

$$=\sum_y \sum_x P(x_i, y_j)\, Log \frac{p(x_i).p(y_j)}{p(x_i).p(y_j)} = 0$$

Thus, if we have independent transmission then the average mutual information is zero, also (using the relations between entropies and average mutual information):

I=0 , H(y) = H(y|x) , H(x)= H(x|y) , H(x,y) = H(x) + H(y)

The Venn-diagram representing the case of independent is represented by two separate circles.

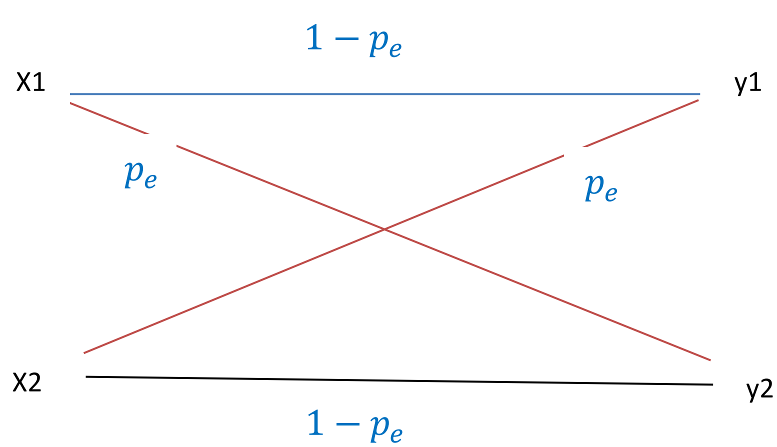
Symmetric Channels

The symmetric channels are those channels that have identical correct transition probabilities for all symbols and similarly for the incorrect transmission probability.

Example-4 Binary Symmetric Channel (BSC)

This is similar to the channel given in Homework#3, its general form is given by:

$$P(y_j|x_i) = \begin{array}{ccc} from/to & y1 & y2 \\ x1 & 1-p_e & p_e \\ x2 & p_e & 1-p_e \end{array}$$



E-mail: ak.kadhim@uomus.edu.iq

Here it is sufficient to describe the whole matrix or the beside model by giving the value of P_e (error probability of the binary symmetric channel). Clearly the above specify that:

$$P(y_2|x_1) = P(y_1|x_2) = p_e$$

and

$$P(y_1|x_1) = P(y_2|x_2) = 1 - p_e$$

It is called symmetric since the correct transition probability of both x_1 and x_2 are the same. Also, the correct transition probability of both x_1 and x_2 are the same

In the question of homework#3 the value of p_e = 0.1, thus:

$$P(y_j|x_i) = \begin{matrix} from/to & y1 & y2 \\ x1 & 0.9 & 0.1 \\ x2 & 0.1 & 0.9 \end{matrix}$$

It is sufficient to look at the matrix rows to decide whether the channel is symmetric or not. If the row components are the same for all rows of the matrix $P(y_j|x_i)$ then the channel is said to be symmetric (this is the definition of symmetry).

Example-5 Ternary Symmetric Channel (TSC)

The following is a ternary symmetric channel:

$$P(y_j|x_i) = \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.1 & 0.7 & 0.2 \\ 0.2 & 0.1 & 0.7 \end{bmatrix}$$

The components of the 1st row (0.7 , 0.2, and 0.1) are the same components of the 2nd and 3rd rows. Also, for every channel the sum of all row components is 1.

E-mail: ak.kadhim@uomus.edu.iq

Q. Specify the type of the following channels (noiseless or/and symmetric) and then draw the model of each channel:

(a) $P(y_j|x_i) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ (b) $P(y_j|x_i) = \begin{bmatrix} 0.8 & 0.2 \\ 0.02 & 0.98 \end{bmatrix}$

(c) $P(y_j|x_i) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ (d) $P(y_j|x_i) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

(e) $P(y_j|x_i) = \begin{bmatrix} 0.8 & 0.2 & 0 \\ 0 & 0.8 & 0.2 \\ 0.2 & 0 & 0.8 \end{bmatrix}$

Notes:

1- In practice all the channels are symmetric and the value of P_e is very important for any information transmission or storage system. It is usually much less than those used in the lecture:

a- In coded speech (or audio) over cellular mobile phone $P_e < 10^{-2}$.
b- In computer data transfer such as database content over Internet $P_e < 10^{-5}$.
c- In compact CD or high-quality music/sound as in multimedia $P_e < 10^{-8}$.

2- The calculation of noise entropy in symmetric channel can be reduced due to symmetry (and rows similarity) from Eq. 5

$$H(y|x) = -\sum_x \sum_y P(x_i, y_j) \; LogP(y_j|x_i)$$

to

$$H(y|x) = -\sum_y P(y_j|x_i) \; LogP(y_j|x_i) \qquad \text{Eq.5B}$$

both relations give the same answer, but the second needs much less calculations.

For BSC with specific P_e, the noise entropy given by Eq.5B is:

$$H(y|x) = -[p_e Log\, p_e + (1-p_e). Log\,(1-p_e)]$$

E-mail: ak.kadhim@uomus.edu.iq