# Lecture three

**Husam Abduldaem Mohammed**

## Conditional and joint Entropy

The intersymbol influence reduces uncertainty and thereby reduces the amount of information produced. We account for this effect by using conditional probability and conditional entropy.

A source producing dependent symbols is said to be redundant, meaning that symbols are generated which are not absolutely essential to convey the information.

For a very long passages of printed English, the conditional entropy may be as low as 0.5 or less bit/symbol because of intersymbol influence. Thus with a suitable coding, printed English theoretically could be transmitted in binary form with an average of the binary digits per symbol.

We just define entropy as the average information per symbol for a certain event $[x_i]$.

Let us consider two ensemble spaces $[x_i]$ and $[y_j]$. The event of having certain outcome of $[x_i]$ and $[y_j]$ simultaneously is a joint event and have a specified **joint probability $p(x_i, y_j)$.** Such event has a certain amount of **self information** which is related to **$p(x_i, y_j)$.** The average information over all the spaces of $[x_i]$ and $[y_j]$ is called **joint entropy.**

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

Another important event is the event of having a certain $x_i$ when a certain $y_j$ occurred. This event has a certain probability which the **conditional probability $p(x_i / y_j)$.** The average information per symbol of such event is called **conditional entropy**

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

similarly

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

As shown above H(X / Y) is related to the uncertainty of having $x_i$ when $y_j$ is occurred.

For a communication system with input [$x_i$] and output [$y_j$].

Noise affect the values of conditional probability such as to increase the uncertainty ( or decrease $p(x_i / y_j)$) of determining the input from the output.

## Example:

a simple language consists of only two symbols A and B produced in a long continuous sequence. Find the single, joint and conditional probabilities of A and B, assuming that the values found from the limited sequence below are typical of a very long sequence ( assumed that the 2 1st letter is A in order to have 20 pairs of pairs of symbols ). Evaluate the conditional entropy for the sequence, and hence deduce the redundancy of the language.

AABBBAAAABBAAABBBAAA

## Solution:

p(A) = 12/20     p(B)=8/20, by counting As and Bs

p(AA)= 9/20     p(BB) 5/20

p(AB)= 3/20     p(BA) 3/20,   by counting pairs

p(A/B)=3/8,     p(B/B)=5/8,  by counting As or Bs after B

p(A/A) =9/12     p(B/A) = 3/12,  by counting As or Bs after A

There are four pairs of symbols AA, BB, AB and BA. Therefore

H(j/I) = -[p(AA) log p(A/A) + p(BB) log (p(B/B)) + p(AB) log p(A/B) + p(BA) log (B/A)]

= 9/20 log 9/12 + 5/20 log 5/8 + 3/20 log 3/12 + 3/20 log 3/8

= 0.868 bit/symbol

if no intersymbol influence had been present the information would have been given by

$$H(i) = -\sum p(i) \log p(i) = -[p(A)\log(pA) + p(B)\, ogp(B)]$$

= -(0.6 log 0.6 + 0.4 log 0.4)

= 0.971 bits/symbol

the redundancy is given by

R = 1 - actual entropy / maximum entropy

= 1- 0.868/ 1= 13 percent.

**Example:(shaum)**

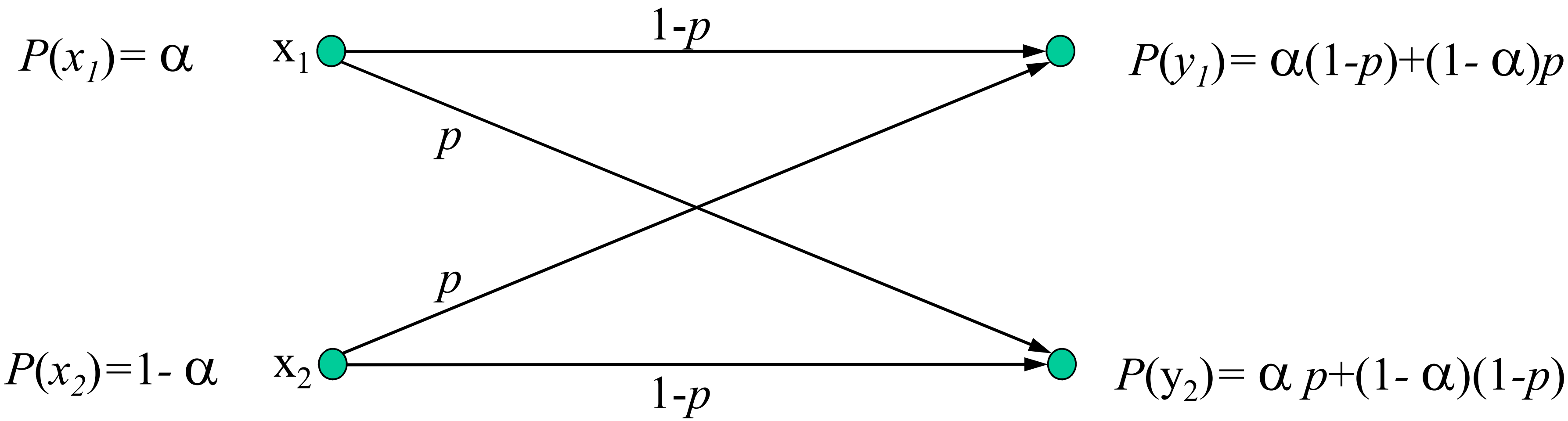
Consider BSC with $p(x_1)= \alpha$.

1- show that the mutual information I(X;Y) is given by

$$I(X;Y)=H(Y)+p\log_2(p)+(1-p)\log_2(1-p)$$

2- calculate I(X;Y) for $\alpha = 0.5$ and p=0.1.

3- repeat (2) for $\alpha = 0.5$ and p=0.5.



**Solution:**

**1-**

## Channel capacity

**Entropy rate (R)** measures the amount of information prodced by a source in a given time

**channel capacity (C):** is a measure of the amount of information a channel can transfer per unit time. Its unit is bits/s.

## Shannon- Hartley theorem

the capacity of a channel with bandwidth **B** and additive Gaussian band limited white noise is

$$C = B \log_2(1 + S/N) \text{ bits/s}$$

where S and N are the average signal power and noise power, respectively, at the output of the channel.

**Fundamental theorem:**

given a channel of capacity C and a source having entropy rate R, then if $R \leq C$, there exists a coding technique such that the output of the source can be transmitted over the channel with an arbitrary small frequency of errors, inspite of the presence of noise. If R > C, it is not possible to transmit without errors.

**Example:**

calculate the capacity of a low pass channel with a usable bandwidth of 3000 Hz and S/N = $10^3$ at the channel output, assume that the channel noise to be Gaussian and white.

**Solution:**

$$C = B \ \log_2(1 + S/N)$$

$$= 3000 \log_2 (1+1000)$$

$$= 30000 \text{ bit/s.}$$

the parameters used in this example are typical of standard voice grade telephone line. The

**Example:** A CRT terminal is used to enter alphanumeric data into a computer. The CRT is connected to the computer through a voice grade telephone line having a usable bandwidth of 3000 Hz and an output S/N of 10 dB. Assume that the terminal has 128 character and that the data sent from the terminal consist of independent sequence of equiprobable characters.

1) find the capacity of the channel.

2) find the maximum (theoretical) rate at which data can be transmitted from the terminal to the computer without error.

**Solution:**

1) the capacity is $C = B \log_2(1+ S/N)$

$$= (3000) \log_2(1+10) = 10{,}378 \text{ bits/s.}$$

2) average information content/character:

$$H = \log_2(128) = 7 \text{ bits/character}$$

and the average information rate of the source $R = r_s H$. For errorless transmission. We need $R = r_s H < C$ or

$$7r_s < 10378$$

$$r_s < 1482$$

hence the maximum rate at which data can be transmitted without error is 1482 character / s.

$C = B \log_2 (1 + S/N)$

$N = \eta B$ where $\eta / 2$ is the noise power spectral density.

$C = B \log_2 (1 + S / \eta B)$

$= (S/\eta)(\eta B / S)(\log_2 (1 + S / \eta B))$

$= (S/\eta) \log_2 (1 + S / \eta B)^{\eta B / S}$

but

$$\lim_{x \to \infty} (1 + x)^{1/x} = e$$

And letting $x = S/\eta B$

$$\lim_{B \to \infty} C = \frac{S}{\eta} \log_2 e$$

$$= 1.44\ S/\eta$$

**Channel representation:-**

Memory less discrete channels are completely sepcified by the set of conditional the relate probability of each output state to the input probability

Each possible input -to-output path is indicted a long with a conditional probability $P_{ij}$ with is concise nation for $p(y_j/x_i)$ thus $p_{ij}$ is the conditonal probability of obtaining output $y_j$ given that the input $x_i$ and is called a channel transition probability

A channel is often specified by the matrix of transition probability [p(y/x)] where for the channel in above

$$[p(y/x)] = \begin{bmatrix} p(y_1/x_1) & p(y_2/x_2) & & p(y_3/x_3) \\ \hline p(y_1/x_2) & p(y_2/x_2) & \cdots & p(y_3/x_2) \\ \hline \\ \hline \\ \end{bmatrix}$$

Since each input to the channel results in some output each new row of the channel matrix must sut to the unity

The channel matrix is useful in driving the output probability give the input probabilities

for example :-

If the input probability p(x) are represented by te row matrix

$$[p(x)]=[p(x_1) \quad p(x_2)]$$

then

$$[p(y)]=[p(y_1) \quad p(y_2) \quad p(y_3)]$$

which is computed by

$$[p(y)]=[p(x)][p(y/x)]$$

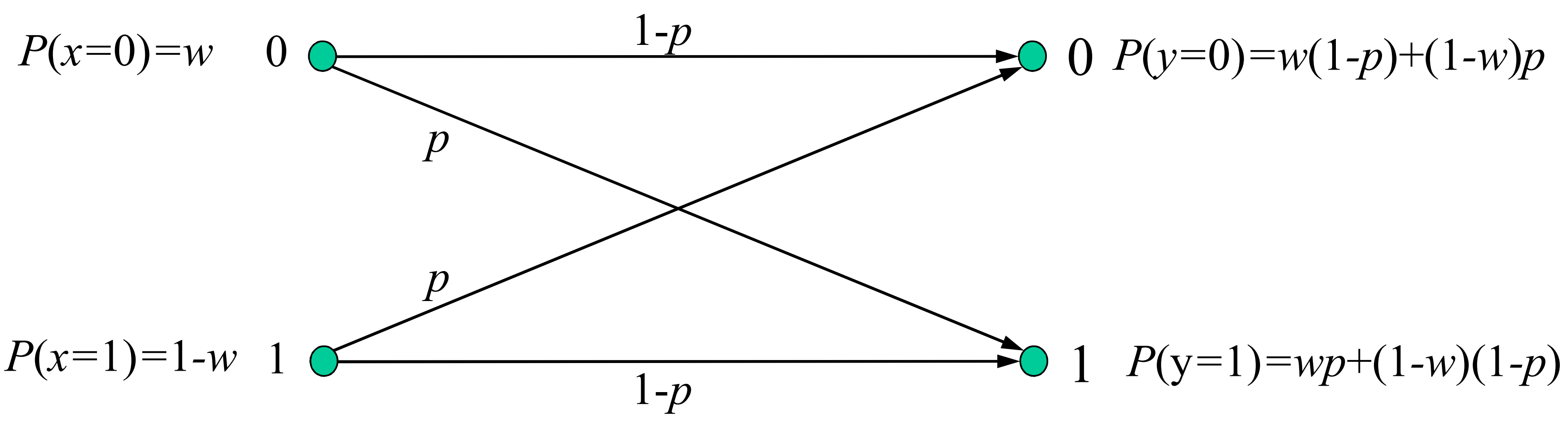
**Joint probability matrix :-** [p(x)]is diagonal ⟶ [p($x_1$y)]

each element p($x_i$) p($y_j$/$x_i$) or p($x_j$/$y_j$)

P($x_i y_j$) is the joint probability of transmitting $x_i$ and receiving $y_j$

## Binary Symmetric Channel

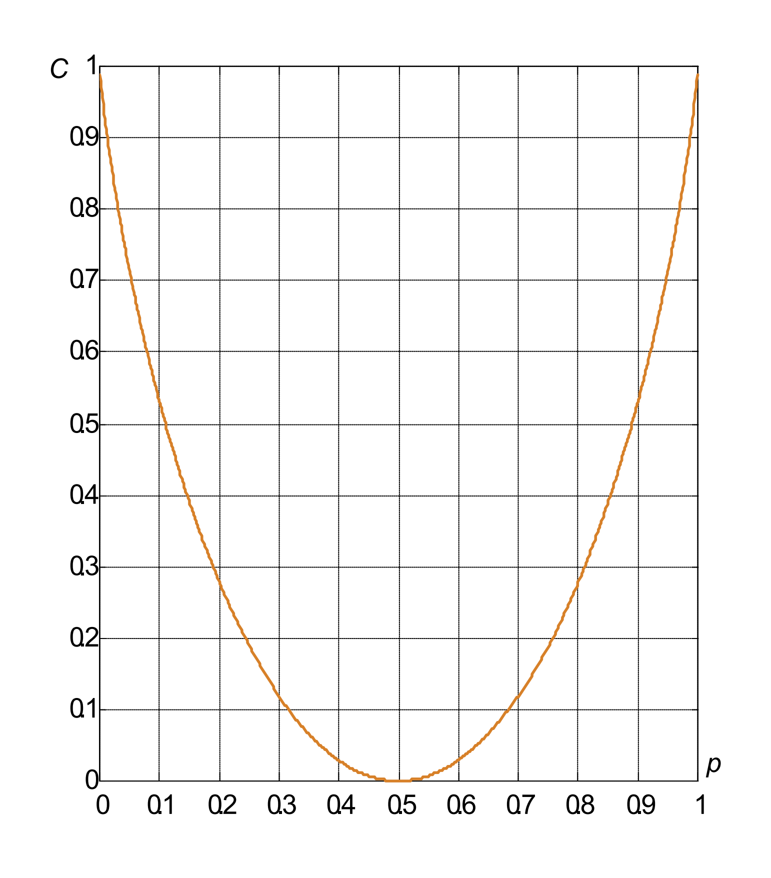
Denote probability of BSC input *x=0* as *w.* Then probabilities of output symbols will be as shown on the graph below.



Because with any input there are exactly two possible output symbol values, one of them having probability $p$, $H(Y|X)=h(p)$ (see $h(\cdot)$ here), and $I(X;Y)=H(Y)-H(Y|X)\leq 1-h(p)$, capacity of BSC

$$C = 1 - h(p)].$$

BSC capacity C versus crossover probability $p$




It is seen that when $p=0$ and channel becomes error-free capacity equals 1bit/symb which is of no wonder because at the input we can have at most one bit of information per symbol. The same we have when $p=1$ because in this case channel is again deterministic and receiving end knows that it just changes any input symbol into an opposite one. When $p=1/2$ output symbols 0 and 1 are equiprobable whatever symbol is applied to input. Therefore, no information is conveyed from the input to the output of BSC (connection between input and output is completely cut off). Certainly, it is always natural to accept $0<p<1/2$.