جامعة الفراهيدي
الكلية التقنية الهندسية
هندسة تقنيات الاتصالات

Al-Farahidi University

المرحلة الثانية

نظرية معلومات

Lec:2

Sura saad

الكورس الاول

2023/10/15

السعر: 2000
التسلسل:8

الاسم:

Statistics:
Bayes' Theorem

UCD maths support centre

Bayes' Theorem (or Bayes' Rule) is a very famous theorem in statistics. It was originally stated by the Reverend Thomas Bayes.

If we have two events A and B, and we are given the conditional probability of A given B, denoted P(A|B), we can use Bayes' Theorem to find P(B|A), the conditional probability of B given A.

Bayes' Theorem: $$P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B')P(B')}$$

where P(B′) is the probability of B not occurring.

Example:

Q: In a factory there are two machines manufacturing bolts. The first machine manufactures **75%** of the bolts and the second machine manufactures the remaining **25%**. From the first machine **5%** of the bolts are defective and from the second machine **8%** of the bolts are defective. A bolt is selected at random, what is the probability the bolt came from the first machine, given that it is defective?

A:
Let A be the event that a bolt is defective and let B be the event that a bolt came from Machine 1.

Check that you can see where these probabilites come from!
$P(B) = 0.75 \quad P(B') = 0.25 \quad P(A|B) = 0.05 \quad P(A|B') = 0.08$

Now, use Bayes' Theorem to find the required probability:

$$P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B')P(B')}$$

$$= \frac{0.05 \times 0.75}{0.05 \times 0.75 + 0.08 \times 0.25}$$

$$= 0.3846$$

Try this:

Exercise: Among a group of male pensioners, 10% are smokers and 90% are nonsmokers. The probability of a smoker dying in the next year is 0.05 while the probability for a nonsmoker is 0.005. Given one of these pensioners dies in the next year, what is the probability that he is a smoker?

Information Theory

<u>Information theory</u> present us the performance characteristics of an ideal or optimum communication system.

- *The performance of an ideal system provides a meaningful basis for comparing the performance of realizable systems ;*
- *It illustrates the gain in performance that can be obtained by implementing more complicated transmission and detection schemes.*
- *Motivation for the study of information theory is provided by <u>Shannon`s coding theorem,</u> Sometimes referred to as <u>Shannon`s 2nd theorem</u> : If a source has an information rate less than the channel capacity ,there exists an encoding procedure such that the source output can be transmitted over the channel with an arbitrarily small probability of error .*

- This is truly surprising statement, which led to that transmission, and reception can be a accomplished with negligible error , even with the presence of noise.

An understanding of this process called <u>coding.</u>

<u>Definitions</u>

1. *Information: any information about an object .*
2. *Message: defined information.*
3. *Signal: physical form of the massage.*

Information theory is a subject that deals with information and data transmission from one point to another.

The block-diagram of any digital communication is

Source of information → Encoder → Channel → Decoder → Listner , Receptor

Audio	*Modulation*	*Speaker*
Video	*Ciphering*	*TV*
Telex	*Error Correction*	*Computer*
Computer	*Data Compression*	*or Any Other Equipment*

Note:

The concept of information related to the probability .Any signal that convoys information is unpredictable (random) , but not vise versa ,i.e not any random signal convoys information(noise is a random signal convoying no information).

Self Information

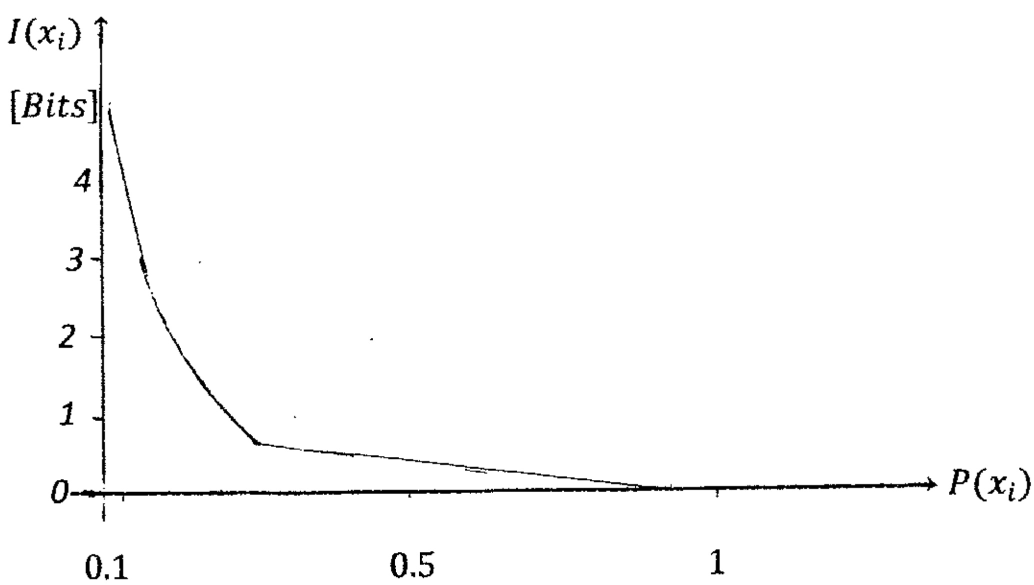
Suppose that the source of information produces finite set of messages $x_1, x_2, \ldots\ldots, x_n$ *with probabilities* $P(x_1), P(x_2), \ldots, P(x_n)$ *and such that*

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

Self Information is

$$I(x_i) = Log_a \frac{1}{P(x_i)} = -Log\ _aP(x_i)$$

- *If* $a = 2$, *the unit of information is the binary unit or bit.*
- *if* $a = e = 2.71828$ *then* $I(x_i)$ *has the unit on nats.*
- *If* $a = 10$, *then* $I(x_i)$ *has the units of Hartley.*



- *Information is zero if* $P(x_i) = 1$ *(Certain event).*
- *Information increases as* $P(x_i)$ *decreases to zero.*
- *Information is + value quantity .*
- *The function that relates* $P(x_i)$ *with information of* x_i *denoted by* $I(x_i)$ *is called* *self information* *of* x_i. *The log function satisfies all previous above points.*

Example

Find the amount of information contained in a black & white (B/w) TV picture, if we assume that each picture has 2.10^5 *dots (pixels and picture elements) and each pixel has* 8 *equiprobable and distinguishable level of brightness ?*

Solution

$$P(each\ level) = \frac{1}{8}\ since\ equiprobale = -log_2 P(x_i)$$

Information / pixel = $-log_2(1/8) = 3\ Bits$

Information / Picture = *Information / pixel * No. of pixels =*$3 * 2.10^5 = 600\ Kbits$

Note $log_a x = \frac{lnx}{\ln a}$

HW

Repeat previous example for color TV wit 16 equiprobable colors and eight equiprobable levels of brightness?

Answer 7 bits.

Entropy

The word entropy finds its roots in Greek entropia which means a turning toward or transformation .The word was used to describe the measure of disorder .

Source Entropy

If the source produces *not equeprobable* *messages ,then* $I(x_i), i = 1,2, \dots .n$ *are different.*

Then the statistical average of $I(x_i)$*over* i *will give the average amount of uncertainty associated with the source* X. *This average is called* *source entropy,* *and denoted by* $H(X)$.

$$H(X) = \sum P(x_i) I(x_i)$$

Or

$$H(x) = -\sum_{i=1}^{n} P(x_i) log_2 P(x_i) \qquad Bits/Symbol$$

Example

Find the entropy of the source producing the following messages:

$$P(X) = \begin{matrix} x_1 & x_2 & x_3 & x_4 \\ 0.25 & 0.1 & 0.15 & 0.5 \end{matrix}$$

Solution

$$H(x) = -\frac{[0.25 \, \text{Ln}\, 0.25 + \; 0.1 Ln \; 0.1 + 0.15 Ln 0.15 + 0.5 \; Ln \; 0.5\,]}{Ln2} = 1.7427 \; {}^{bits}/_{symbol}$$

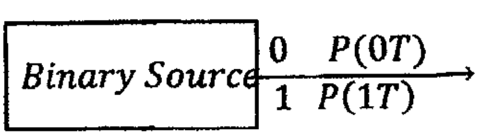
Note

Usually and according to our previous study in logic and digital electronics we are not familiar with fractions of bits. Here in communication these occur due to averaging, i.e. for the previous example, the 1.7427 is the average , the if the source produces say 10^5 *messages , then the amount of information is 174270 bits .*

Example

Find and plot the entropy Of binary source.

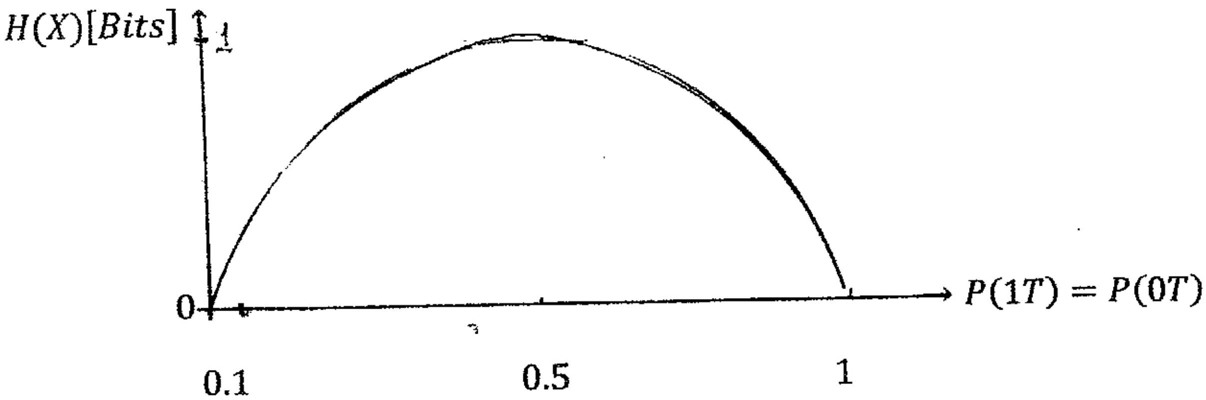
$$P(0T) + P(1T) = 1$$



$$H(X) = -\sum_{i=1}^{n} P(x_i) Log_2 P(x_i) = [P(0T) Log_2 P(0T) + \; (1 - P(0T)) Log_2 P(1 - P(0T))]$$

Note that $H(X)$ is maximum bits/ symbol equals to 1 bits if $P(0T) = P(1T) = 0.5$

$$H(X) = -2[0.5\,Log_2\,0.5\,] = -2\left[0.5\frac{Ln0.5}{Ln2}\right] = -2[0.5 - 1] = 1 \;\; Bits\,/symbol$$



<u>Notes :</u>

1. *In general $H(X) = H(x)_{max} = Log_2 n\, Bits/symbol$ if all messages are equiprobable ,i.e. $P(x_i) = \frac{1}{n}$,then $H(X) = -\sum_{i=1}^{n} P(x_i) I(x_i) = -\sum_{i=1}^{n} P(x_i) Log_2 P(x_i)$.*
2. $H(X) = H(X)_{max} = -\left[\frac{1}{n} Log_2\left(\frac{1}{n}\right) * n\right] =$

   $= Log_2 n \;\; Bits/Symbol$
3. *$H(X) = 0$, if one of the messages has the probability of a certain event.*

<u>Source Entropy Rate</u>

This the average rate of amount of information produced per second, it is denoted by $R(X)$ and is given by : $R(X) = H(X)^{*}$ Rate of producing the symbols .

Where the units of $R(X)$ is $Bits\,/Sec$ (bps), if $H(X)$ is in $bits\,/symbol$,and the rate of producing symbols is $Symbol/Sec$.

Sometimes $R(X)$ is also given as:

$R(X) = \frac{H(X)}{\bar{\tau}}$ Where $\bar{\tau} = \sum_{i=1}^{n} \tau_i P(x_i)$ - Average time duration of symbols, τ_i is the time duration of the symbol x_i.

Example1

A source produces dot "." and dashes "_" with $P(dots) = 0.65$. If the time duration of a dot is $200mS$ and for a dash is $800mS$. Find the average source entropy?

$P(dot) = 0.65$, then $P(dash) = 1 - P(dot) = 1 - 0.65 = 0.35$

$$H(X) = -\sum_{i=1}^{n} P(x_i)Log_2P(x_i) = [(0.65Log_2 0.65) + 0.35Log_2 0.35] = 0.934\ Bits/Symbol$$

$\tau_{dot} = 0.2\ Sec$, $\tau_{dash} = 0.8\ sec$, $\bar{\tau} = 0.2 * 0.65 + 0.8 * 0.35 = 0.41\ Sec$

Then $R(X) = \frac{H(X)}{\bar{\tau}} = \frac{0.934}{0.41} = 2.278bps.$

Example 2

In a telex link, information's arranged in block of 8 characters. The 1st position (character) in each block is always kept the same for the synchronization purposes. The remaining places are filled randomly from the English alphabets with equal probability .IF the system produces 400 blocks / sec , find the average source entropy rate?

Synchronization Character							

← Block →

Each of the 7 positions behaves as a source that may that may produce randomly one of the English alphabets with probability of 1/26, hence :

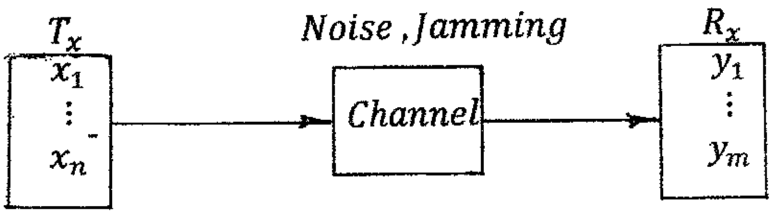
Information / position $= Log_2\left(\frac{1}{26}\right) = 4.7\ Bits$.

Information / Block $= 7 * 4.7 = 32.9\ Bits$

We exclude the 1st character since it has no information having the probability of certain event (contains synchronization character only). Then

$$R(X) = Information/Block * rate\ of\ producing\ Block/sec = 32.9 * 400 =$$
$$= 13160\ Bits/sec$$

Mutual Information



$x_1, x_2, \dots \dots . x_n$ *-set of symbols produced by the transmitter.*

$y_1, y_2, \dots, y_m$ *- symbols ,that may receive by the receiver* R_x .

Theoretically, if the noise and jamming is zero then the set $X = set\ Y$ *and* $= m$.

However, due to the noise and jamming there will be a conditional prob. $P(y_j/x_i)$.

Definitions

1. $P(x_i)$ *to be what is called the apriori of probability of the symbol* x_i *, which the probability of selecting* x_i *for transmission .*
2. $P(x_i/y_j)$ *to be what is called the aposteriori probability of the symbol* x_i *after the reception* y_j.

The amount of information that y_j *provides about* x_i *is called mutual information between* x_i *and* y_j .

This is given by :

$$I(x_i, y_j) = Log_2 \frac{aposteriori\ probability}{aperori\ probability} = Log_2 \frac{P(x_i/y_j)}{P(x_i)}$$

Note that also and since $P(x_i).P(y_j/x_i) = P(y_j).P(x_i/y_j) = P(x_i; y_j)$ *then*

$$I(x_i, y_j) = Log_2 \frac{P(x_i/y_j)}{P(x_i)} = Log_2 \frac{P(y_j/x_i)}{P(y_j)} = I(y_j, x_i)$$

Note

$P(x_i/y_j) \neq P(y_j/x_i)$ *in general .In fact* $P(y_j/x_i)$ *gives the probability of* y_j *given that* x_i *is transmitted as we are at the* T_x *and we transmit* x_i *and we ask about the probability of receiving* y_j *instead .*

The probability $P(x_i/y_j)$ *is the probability of* x_i *given we receive* y_j *as if we are at the* R_x *and we receive* y_j *and we ask about it was coming from* x_i.

Properties of Mutual information

1. *It is symmetric i.e.* $I(x_i, y_j) = I(y_j, x_i)$.
2. $I(x_i, y_j) > 0$ *,if aposteriori probability > aperiori probability ,* y_j *provides positive value information about* x_i .
3. $I(x_i, y_j) = 0$*, if aposteriori probability =aperiori probability, which is in the case of statistical independence, when* y_j *provides no information about* x_i .
4. $I(x_i, y_j) < 0$ *,if aposteriori probability <aperiori probability,* y_j *provides negative information about* x_i *, i.e.* y_j ***adds ambiguity*** .Type equation here.

Transformation (Average Mutual) Information

This is statistical averaging of all the pair $(x_i, y_j), i = 1,2, \dots, n, j = 1,2, \dots, m$.

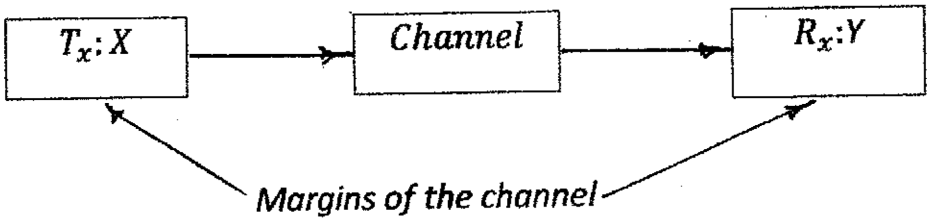
This is denoted by $I(X, Y)$ *and is given by :*

$$I(X,Y) = \sum_{i=1}^{n}\sum_{j=1}^{m} I(x_i . y_j) . P(x_i, y_j = \sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i, y_j) \, Log_2 \frac{P(x_i/y_j)}{P(x_i)} \; Bits/symbol$$

Or

$$I(X,Y) = \sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i, y_j) \, Log_2 \frac{P(y_j/x_i)}{P(y_j)} \; Bits/symbol$$

Mutual Entropies :



Marginal entropies are a term usually used to denote both source entropy $H(X)$ defined as before and the receiver entropy $H(Y)$,given by :

$$H(Y) = -\sum_{j=1}^{m} P(y_j) Log_2 P(y_j) \; Bits/Symbol$$

Joint and Conditional Entropy

The average amount of information, associated with the pair (x_i, y_j) called joint or system entropy $H(X,Y)$:

$$H(X,Y) = H(Y,X) = -\sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i, y_j) Log_2 P(x_i, y_j) \; Bits/Symbol$$

The average amount of information associated with the pairs (y_j/x_i) and (x_i/y_j) are called conditional entropies $H(Y/X)$ and $H(X/Y)$, they are given by :

$$H(Y/X) = -\sum_{j=1}^{m}\sum_{i=1}^{n} P(x_i, y_j) \, Log_2 P(y_j/x_i) \; Bits/Symbol$$

$$H(X/Y) = -\sum_{j=1}^{m}\sum_{i=1}^{n} P(x_i, y_j) \, Log_2 P(x_i/y_j) \; Bits/Symbol$$

Example1

Show that $H(X,Y) = H(X) + H(Y/X)$

This is a very useful identity to ease calculations in problem solving.

To prove it, then we know that:

$$H(X,Y) = -\sum_{j=1}^{m}\sum_{i=1}^{n} P(x_i, y_j) Log_2 P(x_i, y_j)$$

But $P(x_i,y_j) = P(x_i).P(y_j/x_i)$, putting this inside the Log term only ,then

$$H(X,Y) = -\sum_{j=1}^{m}\sum_{i=1}^{n} P(x_i/y_j)log_2P(x_i) - \sum_{j=1}^{m}\sum_{i=1}^{n} P(x_i/y_j)Log_2P(y_j/x_i)$$

After reversing the order of summation of the 2nd term:

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

$H(X,Y) = -\sum_{i=1}^{n} P(x_i)Log_2P(x_i) - \sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i,y_j)log_2P(y_j/x_i)$, In this equation, the 1st term in fact $H(X)$ and the 2nd term with – sign is $H(Y/X)$

,then

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

<u>Home work</u>

Show that $H(X,Y) = H(Y) + H(X/Y)$

<u>Example 2</u>

Show that $I(X,Y) = H(X) - H(X/Y)$

$$I(X,Y) = P(x_i,y_j)I(x_i,y_j) = \sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i,y_j)Log_2\frac{P(x_i/y_j)}{P(x_i)} =$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i,y_j)Log_2P(x_i/y_j) - \sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i,y_j)Log_2P(x_i)$$

As before ,we reverse the order of summation of the 2nd term ,then $\sum_{j=1}^{m} P(x_i,y_j) = P(x_i)$,and

$$I(X,Y) = \sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i,y_j)Log_2P(x_i/y_j) - \sum_{i=1}^{n} P(x_i)Log_2P(x_i) = H(X) - H(X/Y)$$

Example3

Show that $I(X,Y)$ is zero for extremely noisy channel?

Solution

For extremely noisy channel, then y_j gives no information about x_i (the receiver cannot decide anything about x_i , as if we transmit a deterministic signal , but the receiver receives noise like.

Signal y_j that is completely has no correlation withx_i). Then x_i and y_j are statistically independent and $P(x_i/y_j) = P(x_i)$ and $P(y_j/x_i) = P(y_j)$ for all i and j.

Then

$I(x_i, y_j) = Log_2 1 = 0$ *for all i and j.*

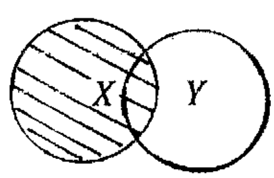
$$I(x_i, y_j) = Log_2 \frac{P(x_i/y_j)}{P(x_i)} = Log_2 \frac{P(x_i)}{P(x_i)} = Log_2 1 = 0$$

Thus $I(X,Y) = 0$

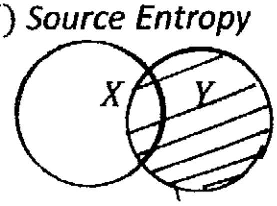
Venn Diagram Representations of Entropies

$X = Transmitter\ T_x$

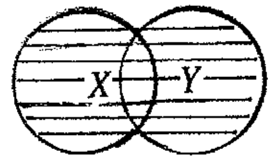
$Y = Reciever\ R_x$



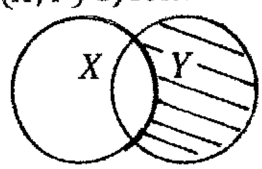
$H(X)$ Source Entropy



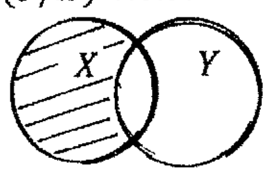
$H(Y)$ Receiver Entropy



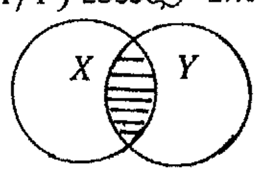
$H(X,Y)$ *System Entropy*



$H(Y/X)$ *Noise Entropy*



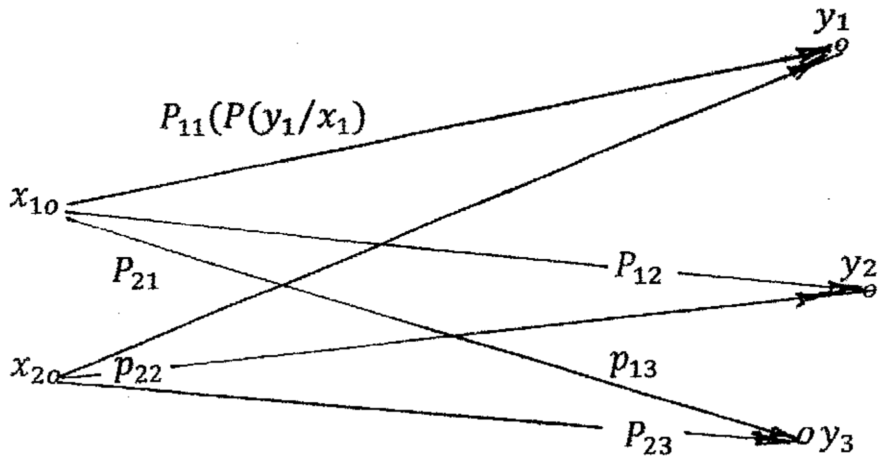
$H(X/Y)$ *Losses Entropy*



$I(X,Y)$ *Transinfomation*

Channel Representation

1. *The channel will be assumed to be memory less.*
2. *For each channels output (o/p) at given time is a function of the channel input (i/p) at that time ,and is not a function of previous channel inputs.*
3. *Memory less discrete channels are completely specified by the set of conditional probabilities that relate the probability of each o/p state to the i/p probabilities .*
4. *An example illustrates the technique. A diagram of a channel with two i/p's and 3 o/p's is illustrated in figure:* (12 -28)



Each possible i/p –to-o/p path is indicated along with conditional probability P_{ij} which is concise notation $P(y_j/x_i)$ .Thus P_{ij} is the conditional probability of obtaining o/p $y_j$that the i/p $x_i$and is called a <u>channel transitions probabilities .</u>

We can see from the figure that the channel is completely specified by the complete set of transition probabilities $P(Y/X)$.where the channel in the figure.

$$[P(Y/X)] = \begin{matrix} & y_1 & y_2 & y_3 \\ x_1 & P(y_1/x_1) & P(y_2/x_1) & P(y_3/x_1) \\ x_2 & P(y_1/x_2) & P(y_2/x_2) & P(y_3/x_2) \end{matrix}$$

<u>Notes</u>

1)Since each i/p to the channel results in some o/p ,each row of the Channel matrix must <u>sum to unity .</u>

2) The matrix is useful in deriving the o/p probabilities given the i/p probabilities.

For example ,if the i/p Probabilities are represented by <u>row matrices :</u>

$$[P(X)] = [P(x_1) \quad P(x_2)]$$

$$[P(Y)] = [P(y_1) \quad P(y_2) \quad P(y_3)]$$

Which is computed by :

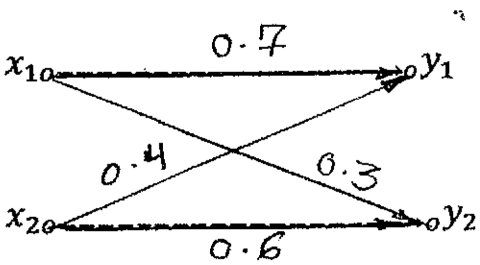
$[P(X,Y)] = [P(X)][P(Y/X]$,This equation yields a matrix$[P(X,Y)]$.Each element in the matrix has the form $P(x_i)P(y_j/x_i)$ or $P(x_i,y_j)$.

This matrix is as the <u>joint probability matrix</u>, and the term $P(x_i, y_j$ is the joint probability of transmitting x_i and receiving y_j .

<u>Example</u>

Consider the binary i/p – o/p channel shown in the figure. The matrix of <u>transition probabilities</u> is :

$$[P(Y/X)] = \begin{matrix} & y_1 & y_2 \\ x_1 & 0.7 & 0.3 \\ x_2 & 0.4 & 0.6 \end{matrix} \quad ,$$



If the i/p probabilities are $P(x_1) = 0.5$ and $P(x_2)$, find the o/p $P(Y)$,and the joint probability matrix ?

1)The joint probability matrix

$$[P(x,y)] = [P(X)][P(Y/X)] = [0.5 \quad 0.5]\begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix} = \begin{bmatrix} (0.5)(0.7) & (0.5)(0.3) \\ (0.5)(0.4) & (0.5)(0.6) \end{bmatrix}$$

$$= \begin{bmatrix} 0.35 & 0.15 \\ 0.2 & 0.3 \end{bmatrix} = \begin{bmatrix} P(x_1,y_1) & P(x_1,y_2) \\ P(x_2,y_1) & P(x_2,y_2) \end{bmatrix}$$

2) Matrix $[P(Y)]$

$P(y_1) = P(x_1,y_1) + P(x_1,y_1) = 0.35 + 0.2 = 0.55$

$P(y_2) = P(x_2,y_1) + P(x_2,y_2) = 0.15 + 0.3 = 0.45$

$[P(Y)] = [o.55 \quad 0.45]$

Note: The sum of all elements in the joint matrix =1.

Example

The Joint probability of a system is given by:

$$P(X,Y)=\begin{matrix} \\ x_1 \\ x_2 \\ x_3 \end{matrix}\begin{matrix} y_1 & y_2 \\ \left[\begin{matrix} 0.5 \\ 0 \\ 0.0625 \end{matrix}\right. & \left.\begin{matrix} 0.25 \\ 0.125 \\ 0.0625 \end{matrix}\right] \end{matrix}$$

Find :

1. *Marginal entropies;*
2. *Joint entropy ;*
3. *Conditional entropies ;*
4. *The mutual information between x_1 and y_2 ;*
5. *The transinformation, and draw the channel model?*

Solution

1. *First we find $P(X)$ and $P(Y)$ from $P(X,Y)$ by summing the rows and columns.*

$$[P(X)] = \begin{matrix} x_1 & x_2 & x_3 \\ [0.75 & 0.125 & 0.125] \end{matrix}$$

$$[P(Y)] = \begin{matrix} y_1 & y_2 \\ [0.5625 & 0.4375] \end{matrix}$$ *, then*

$$H(X) = -\sum_{i=1}^{n} P(x_i)Log_2P(x_i) = -\sum_{i=1}^{3} P(x_i)Log_2P(x_i)\ /Ln2$$
$$= -[0.75\ Ln0.75 + 2 * (0.125Ln0.125]/Ln2 = 1.06127\ Bits/symbol$$
$$H(Y) = -\frac{[0.5625Ln0.5625+0.4375Ln0.4375]}{Ln2} = 0.9887\ Bits/symbol$$

$$2.\ H(X,Y) = -\sum_{j=1}^{m}\sum_{i=1}^{n} P(x_i,y_j)Log_2P(x_i,y_j)$$

From Joint matrix by rows

$$H(X,Y) = -\ [0.5\ Log_2\,0.5 + 0.25Log_2 0.25 + 0.125Log_2 0.125 + 2 * (0.0625Log_2 0.0625]$$

$$H(X,Y) = 1.875\ Bits/Symbol \qquad (15\text{-}28)$$

3. $H(Y/X) = H(X,Y) - H(X) = 1.875 - 1.06127 = 0.813\, Bits/symbol$

$H(X/Y) = H(X,Y) - H(Y) = 1.875 - 0.9887 = 0.886\ Bits/Symbol$

4. $I(x_1, y_2) = Log_2 \frac{P(x_1/y_2)}{P(x_1)}$,but $P(x_1/y_2) = \frac{P(x_1,y_2)}{P(y_2)}$, *Thus* $P(x_1, y_2) = P(y_2) * P(x_1/y_2)$

$$I(x_1, y_2) = Log_2 \frac{P(x_1, y_2)}{P(x_1)P(y_2)} = Log_2 \frac{0.25}{(0.75).(0.4375)} = -0.3923 Bits$$

That means y_2 *gives ambiguity about* x_1 .

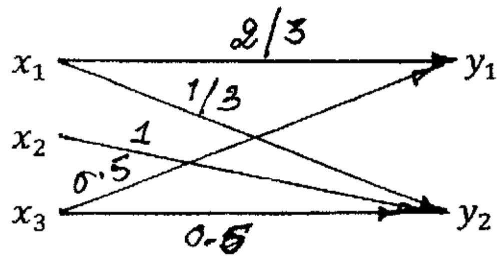
5. $I(X.Y) = H(X) - H(X/Y) = 1.06127 - 0.8863 = 017497\ Bits/symbol$

6.To draw the channel model we find $[P(Y/X]$ *matrix from the joint matrix* $[P(X,Y)]$ *by dividing its rows by the corresponding* $P(x)$*:*

$$[P(Y/X)] = \begin{matrix} & y_1 & y_2 \\ x_1 & 0.5/0.75 & 0.25/0.75 \\ x_2 & 0/0.125 & 0.125/0.125 \\ x_3 & 0.0625/0.125 & 0.0625/0.125 \end{matrix} = \begin{matrix} & y_1 & y_2 & \\ x_1 & 2/3 & 1/3 & = 1 \\ x_2 & 0 & 1 & = 1 \\ x_3 & 0.5 & 0.5 & = 1 \end{matrix}$$

Note that the sum of each row in $[P(Y/X)]$ *matrix is unity.*

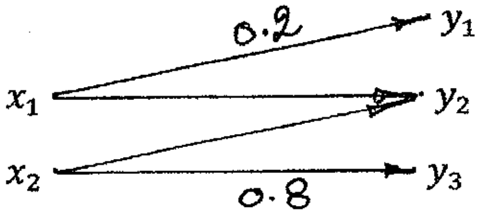
7.The channel model



The drawing of channel model depends on the conditional matrix.

H.W

For the channel model shown



Find:

-Source Entropy rate ,if $\tau_{x1} = 1mS,\ \tau_{x2} = 2mS$ *and* $I(x_1) = 2$ *Bits ?*

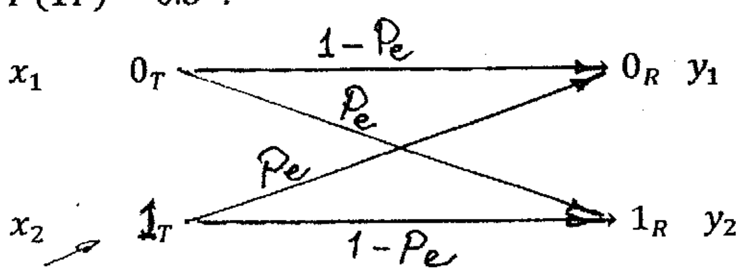
-The transinformation ?

Answer $R(X) = 463.58$ Bits/Sec.

$I(X,Y) = 0.70797$ Bits/Sec.

Example

Find and plot the transinformation for a binary symmetric channel (BSC) ,shown if $P(0T) = P(1T) = 0.5$ *?*



Solution

This BSC is very known channel with practical values of $P_e \ll 1$*.If we denote* $0_T = x_1, 1_T = x_2, 0_R = y_1, 1_R = y_2$ *,then*

$$P(Y/X) = \begin{matrix} & y_1 & y_2 \\ x_1 & 1-P_e & P_e \\ x_2 & P_e & 1-P_e \end{matrix}$$

$$P(Y/X) = \frac{P(X,Y)}{P(X)}, P(X,Y) = P(Y/X).P(X)$$

$$P(X,Y) = \begin{matrix} & y_1 & y_2 \\ x_1 & 0.5(1-P_e) & 0.5P_e \\ x_2 & 0.5\,P_e & 0.5(1-P_e) \end{matrix}$$

$P(y_1) = 0.5(1-P_e) + 0.5P_e = 0.5\,,\ P(y_2) = 0.5$

$Then[P(Y)] = \begin{bmatrix} y_1 & y_2 \\ 0.5 & 0.5 \end{bmatrix}$ *and*

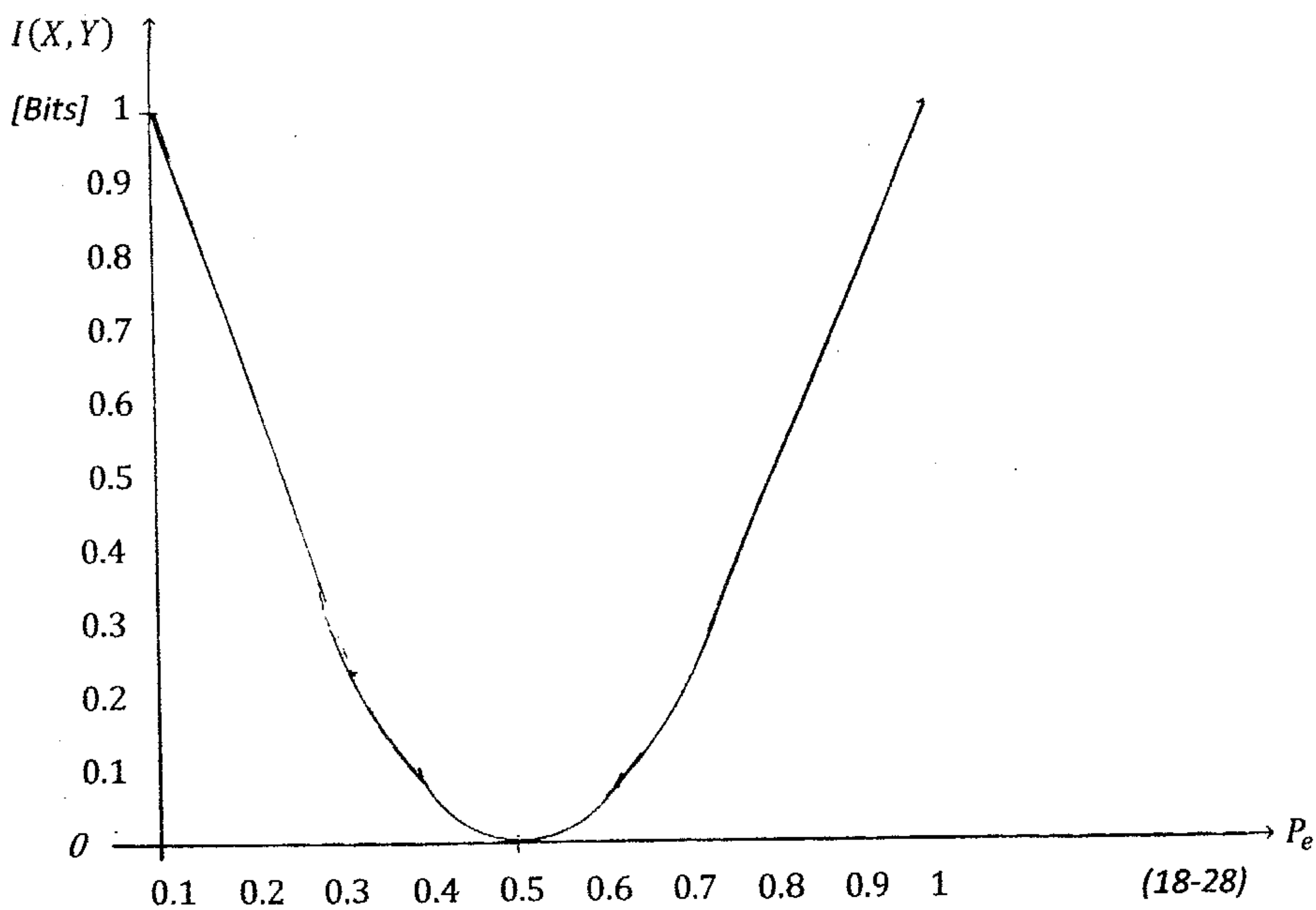
$H(Y) = H(Y)_{max} = 1Bit\ \left(Log_2 \frac{1}{0.5}\right) = Log_2 2 = 1$

Next we find $H(Y/X) = -\sum_{j=1}^{m}\sum_{i=1}^{n} P(x_i, y_j) Log_2(y_j/x_i)$

$H(Y/X) = -\{[0.5(1-P_e)log_2(1-P_e)] * 2 + [0.5P_e Log_2 P_e] * 2\} =$

$H(Y/X) = -[(1-P_e)log_2(1-P_e) + P_e Log_2 P_e]$

$$I(X,Y) = H(Y) - H(Y/X) = 1 + (1-P_e)log_2(1-P_e) + P_e Log_2 P_e$$



(18-28)

Homework

A BSC has $P(Y/X) = \begin{bmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{bmatrix}$, *if* $I(0_T) = 3$ *bits, find system and losses entropies ?*

Answer

Other Special Channels

1. *Lossless Channel*

   This Channel has nonzero element in each column of the transitional matrix $[P(Y/X)]$ *as an example:*

$$[P(Y/X)] = \begin{array}{c} \\ \end{array}\begin{matrix} y_1 & y_2 & y_3 & y_4 & y_5 \end{matrix}$$
$$[P(Y/X)] = \begin{bmatrix} 3/4 & 1/4 & 0 & 0 & 0 \\ 0 & 0 & 1/3 & 2/3 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

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

   H. W: Draw the channel model?

2. *Deterministic Channel*

This channel has only nonzero element in each row of the transition matrix $[P(Y/X)]$ *,as an example:*

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

This channel has $H(Y/X) = 0$ *and* $I(X,Y) = H(Y)$ *with zero noise entropy.*

H.W

Draw the channel model?

$C = r * Max[I(X,Y)]\ Bits/Sec$ (2)

Where r is the number of symbols produced per second.

C also expressed as

$C = Max[R(X,Y)]$ (3)

Where :

$R(X,Y) = r * I(X,Y)\ Bits/sec$ *Or*

$R(X,Y) = \frac{I(X,Y)}{\bar{\tau}}$ (4)

Where $\bar{\tau} = \sum_{i=1}^{n} \tau_i P(x_i)$ *Average time duration of symbols,* τ_i *is the time duration of the symbol* x_i.

The maximization of $I(x,Y)$ *is done with respect to the to the i/p probability* $P(X)$ *,or the o/p probability for a constant channel conditions .i.e. with* $P(Y/X)$ *being a constant.*

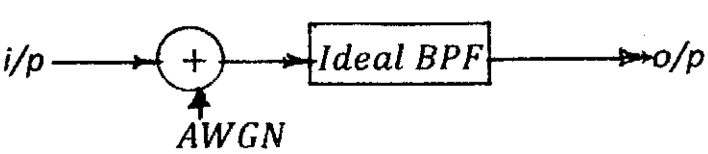
<u>Shannon's Theorem</u>

1. *A given communication system has the maximum rate of information "C" known as the channel capacity.*
2. *If the information rate R is less than C ,then one can approach arbitrarily small error probabilities by using intelligent coding techniques .*
3. *To get lower error probabilities, the encoder has to work on longer blocks of signal data. This entails longer delays and high computational requirements.*

Thus if $R \leq C$*, 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 avoided regardless of the coding technique used.*

Consider a band limited Gaussian channel operation, in the presence of additive noise.



,BPF – Band pass filter

AWGF -Additive White Gaussian Noise

Noiseless Channel

This channel has only one nonzero element in each row and column of the transitional matrix $[P(Y/X)]$ i.e.it is an identity matrix .As an example :

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

This has $H(X/Y) = H(Y/X) = 0$ and $I(X,Y) = I(X) = H(Y)$.

H.W : Draw the channel model?

Source Efficiency

$$\eta = \frac{H(X)}{H(X)_{max}} = H(X)/Log_2 n$$

Source Redundancy

$1 - \eta = 1 - [H(X)/Log_2 n]$ *, this equation can be given as percentage.*

Channel Capacity

Discrete Channel

Is a channel whose i/p is discrete.

Channel Capacity

Defined as the maximum of mutual information $I(X,Y)$.

$$C = Max[I(X,Y)] \;\; Bita/symbol \qquad (1)$$

Where C is the channel capacity.

Physically it is 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" .

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

$$C = BLog_2(1 + (S/N))$$

Where:

C *– The channel capacity in Bits/ Sec;*

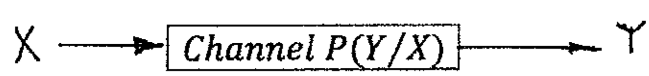
B *– The bandwidth of the Channel in HZ;*

S/N *– Signal – to noise ratio.*

Discrete Memory less Channel (DMC)

DMC has an i/p X and o/p Y . At any given time (t) , the channel o/p $Y = y$ *only depends on the i/p* $X = x$ *at time t and does not depend on the past history of the i/p .*

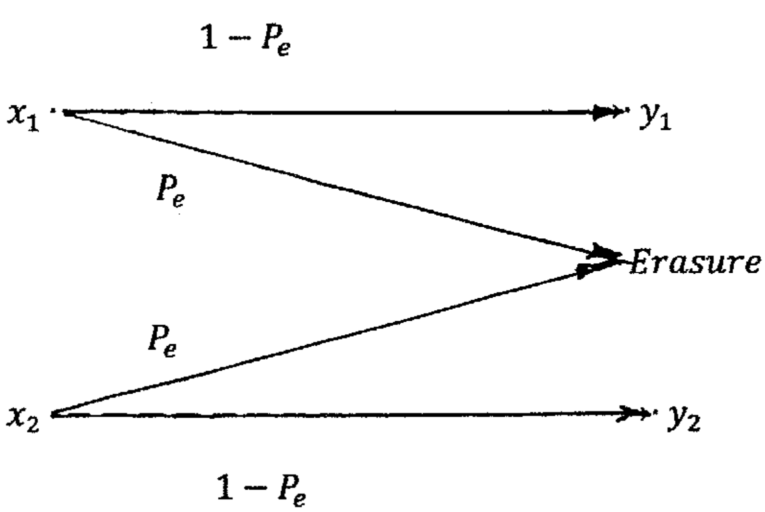
DMC is represented by the conditional probability $P(Y/X)$.



Binary Erasure Channel(BEC)

BEC model are widely used to represent channels or links that losses data .

Prime examples of such channels are the internet links and routes ABCE channels ,has a binary i/p X and ternary o/p Y .



Where P_e *-Erase probability, e-erasure.*

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

$P_r(Y = erasure\,/X = 0) = P$

$P_r(Y = erasure/X = 1) = P$

$P_r(Y = 0/X = 0) = 1 - P_e$ *Right*

$P_r(Y = 1/X = 1) = 1 - P_e$ *Right*

$P_r(Y = 0/X = 1) = 0$ *Error*

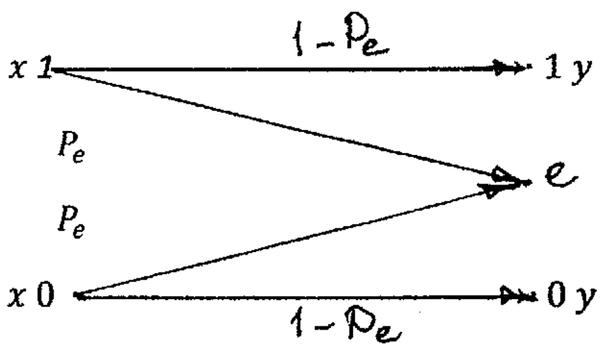
$P_r(Y = 1/X = 0) = 0$ *Error*

The channel model is:

$$\begin{array}{c} \\ x_1 \\ x_2 \end{array}\begin{array}{c} \begin{array}{ccc} y_1 & E & y_2 \end{array} \\ \begin{bmatrix} 1-P_e & P_e & 0 \\ 0 & P_e & 1-P_e \end{bmatrix} \end{array}$$

The BEC, in this model a T_x sends a bit (0 or 1) and the receiver either receives that bit or it receives a message the not received (erased).

This channel is used frequently in information theory because it is one of the simplest channel to analyze. The BEC was introduced by Peter Eilas in 1954 in MTI.



This channel is error free .

P_e *–erase probability. Binary input , ternary output .*$X(0,1)$ $Y(0,1,e)$

Channel Capacity of Discrete Symmetric Channels

The general definition of Symmetric Channels is that channel, where

1. $n = m$ *equal numbers of symbols in X and Y, i.e.* $[P(Y/X)]$ *is a square matrix.*

2. *Any row in* $[P(Y/X)]$ *matrix comes from some permutation of other rows.*

Example :

$[P(Y/X] = \begin{bmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{bmatrix}$ *is a binary symmetric channel(BSC) where n=m=2, and 1st row is the permutation of the 2nd row.*

Channel Capacity of such symmetric channel is easy to find using the following derivation:

To find $max[I(X,Y]$, *then :*

$$I(X,Y) = H(Y) - H(Y/X) = H(Y) + \sum_{j=1}^{m}\sum_{i=1}^{n} P(x_i)\,P(y_j/x_i)\,Log_2P(y_j/x_i)$$

If the channel is symmetric, then the quantity $\sum_{j=1}^{m} P(y_j/x_i)\,Log_2P(y_j/x_i)$ *is a constant independent of the row number, so if this comes out of the double summation, the remaining*

Is $\sum_{i=1}^{n} P(x_i)$ *which equal unity, hence is:*

$$I(X,Y) = H(Y) + \sum_{J=1}^{m} P(y_j/x_i)\,Log_2P(y_j/x_i) = H(Y) + K$$

Where $K = \sum_{J=1}^{m} P(y_j/x_i)\,Log_2P(y_j/x_i)$

Hence:

$I(X,Y) = H(Y) + K$ *For symmetric channels only.*

Now to find max of $I(X,Y) = \text{Max}[H(Y) + K] = Max[H(Y)\,] + K$

And since $Max[H(Y)\,] = Log_2m$ *when Y has equiprobable symbols, then the channel capacity*

$C = Log_2m + K\ \ Bits/Symbols\ \ For\ symmetric\ channels\ only.$

Channel Efficiency and Redundancy

Channel Efficiency

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

And the channel redundancy $R = 1 - \eta = 1 - \frac{I(X,Y)}{C}$

Notes

1. $I(X,Y)$*, becomes maximum equals to "C" only if the condition for the maximization is satisfied, i.e. only if* Y *has equiprobable symbols .*

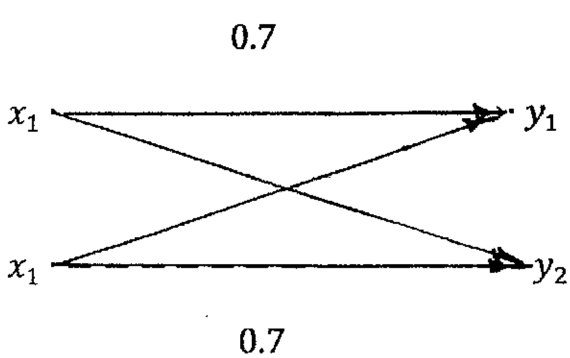
This condition yields that X *has also equiprobable symbols ,since if the output of symmetric channel is equiprobable , then its in* X *is also equiprobable .*

2.For symmetric channel only , and to ease calculations , the following formulae can be used

$$\boxed{I(X,Y) = H(Y) + K}$$

Example

For the BSC shown:



Find the channel capacity and efficiency if $(x_1) = 2\ Bits$ *?*

Solution

First we write $[P(Y/X)]$ *as*

$$[P(Y/X)] = \begin{matrix} & y_1 & y_2 \\ x_1 & 0.7 & 0.3 \\ x_2 & 0.3 & 0.7 \end{matrix}$$ *and since the channel is symmetric ,the* $C = Log_2 m + K$

$n = m = 2$

And

$$K = \sum_{j=1}^{m} P(y_j/x_i) Log_2(y_j/x_i) = 0.7 Log_2 0.7 + 0.3\, Log_2 0.3 = -0.88129\, Bits\ /Symbols$$

Then

$C = Log_2 m + K = Log_2 2 + K = 1 - 0.8812 = 0.1187\, Bits/Symbol$

To find the channel efficiency , we must find $I(X,Y)$ *.First we find* $P(x_1)$ *from* $I(x_1) = -Log_2 P(x_1) = 2$, *giving* $P(x_1) = 2^{-2} = 0.25$ *,then*

$$[P(X)] = \begin{matrix} x_1 & x_2 = 1 - x_1 \\ [0.25 & 0.75] \end{matrix}$$ *, multiplying with*$[P(Y/X]$ *,get*

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

Then summing the columns, to obtain

$[P(Y)] = [0.4 \quad 0.6]$ *from which*

$H(Y) = 0.97095\ \ Bits/symbol$

$I(X,Y) = H(Y) + K = 0.97095 - 0.8812 = 0.0896\, Bits/Symbol$

Then $\eta = \frac{I(X,Y)}{C} = \frac{0.0896}{0.1187} = 75.6\%$

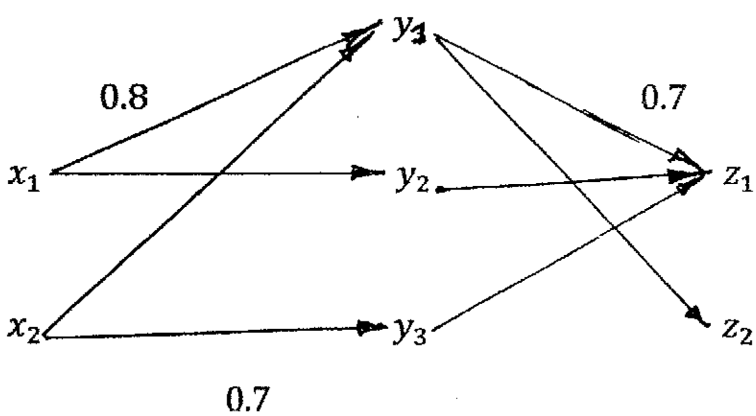
<u>Cascading of Channels</u>

If two channels are cascaded as shown, then the overall transition matrix of the equivalent channel is the matrix multiplication of the transitional probabilities of two cascaded channels.

$[P(Z/X)] = [P(Y/X)] * [P(Z/Y)]$

<u>*Example*</u>

Find the transition matrix $[P(Z/X)]$ *for the cascaded channels shown?*



<u>*Solution:*</u>

$$[P(Y/X)] = \begin{matrix} & y_1 & y_2 & y_3 \\ x_1 & 0.8 & 0.2 & 0 \\ x_2 & 0.3 & 0 & 0.7 \end{matrix}$$

$$[P(Z/Y)] = \begin{matrix} & z_1 & z_2 \\ y_1 & 0.7 & 0.3 \\ y_2 & 1 & 0 \\ y_3 & 1 & 0 \end{matrix}$$

$$[P(Z/X)] = [P(Y/X)] * [P(Z/Y)]$$

$$[P(Z/X)] = \begin{bmatrix} 0.8 & 0.2 & 0 \\ 0.3 & 0 & 0.7 \end{bmatrix} * \begin{bmatrix} 0.7 & 0.3 \\ 1 & 0 \\ 1 & 0 \end{bmatrix} = \begin{matrix} & z_1 & z_2 \\ x_1 & 0.76 & 0.24 \\ x_2 & 0.91 & 0.09 \end{matrix}$$

0.8 * 0.7 + 0.2 * 1 + 0 * 1 =
0.56 + 0.2 = 0.76

0.8 * 0.3 + 0.2 × 0 + 0 × 0
= 0.24

0.3 * 0.7 + 0 * 1 + 0.7 * 1 =
= 0.21 + 0.7 = 0.91

0.3 * 0.3 + 0 × 0 + 0 * 0
= 0.09

(27-28)

Example :

$$P(x_1) \quad P(x_2)$$

For the previous example, find $[P(Y)]$ *and* $[P(Z)]$ *, if* $[P(X)] = [0.7 \quad 0.3]$ *?*

$$[P(X,Y)] = [P(Y/X] * [P(X)]$$

$$\begin{array}{cc} & \begin{array}{ccc} y_1 & y_2 & y_3 \end{array} \\ [P(X,Y)] = \begin{array}{c} x_1 \\ x_2 \end{array} & \begin{bmatrix} 0.56 & 0.14 & 0 \\ 0.09 & 0 & 0.21 \end{bmatrix} \begin{array}{c} P(x_1) \\ P(x_2) \end{array} \end{array}$$

$$\begin{array}{ccc} y_1 & y_2 & y_3 \end{array}$$

$$[P(Y)] = [o.65 \quad 0.14 \quad 0.2]$$

$$[P(Y,Z)] = [P(Y/Z)] * [P(Y)]$$

$$\begin{array}{cc} & \begin{array}{cc} z_1 & z_2 \end{array} \\ [P(Y,Z)] = \begin{array}{c} y_1 \\ y_2 \\ y_3 \end{array} & \begin{bmatrix} 0.455 & 0.195 \\ 0.14 & 0 \\ 0.21 & 0 \end{bmatrix} \begin{array}{c} P(y_1) \\ P(y_2) \\ P(y_3) \end{array} \end{array}$$

$$\begin{array}{cc} z_1 & z_2 \end{array}$$

$$[P(Z)] = [0.805 \quad 0.195]$$