Information Theory

1.1 Review of probability

Self Probability: If an experiment has A_1, A_2, A_3 ….. A_n outcome then:

Prob. (A) = P(A) = $\lim\limits_{N \to \infty} = \dfrac{n(A_i)}{N}$

$n(A_i)$: number of times outcome A_i occurs.
N : number of repeated times.

Note that:

- ❖ $1 \geq P(A) \geq 0$ and $\sum_{i=1}^{n} P(A_i) = 1$
- ❖ If $P(A_i)=1$ A_i is a certain event.

Joint Probability : If we have two experiments A, B. exp. A has A_1, A_2, A_3 ….. A_n outcome and exp. B has B_1, B_2, B_3 ….. B_m outcome, then $P(A_i,B_j)$=joint prob. of A occurs from exp. A_i and B occurs from exp. B.

$$\sum_{j=1}^{m}\sum_{i=1}^{n} P(A_i, B_j) = 1$$

$P(A_i,B_j)$ can be written in matrix for $P(A_i,B_j)$:

A_i in the source
B_j in the receiver

$$\begin{array}{cccccc} & B_1 & B_2 & . & . & B_m \\ A_1 & . & . & . & . & . \\ A_2 & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ A_n & . & . & . & . & . \end{array}$$

Note that :

- ❖ $\sum_{i=1}^{n} P(A_i, B_j) = P(B_j)$ sum of the i_{th} column.
- ❖ $\sum_{j=1}^{m} P(A_i, B_j) = P(A_i)$ sum of the j_{th} row.

Conditional probability: two experiments A & B with their outcome affect on each other.

$P(A_i / B_j)$ = conditional prob. Of A_i given that B_j is already occurred in exp. B.

$P(B_j / A_i)$ = conditional prob. Of B_j given that A_i is already occurred in exp. A.

$P(A_i / B_j)$ or $P(B_j / A_i)$ can be written in matrix form.

$$\begin{array}{c|ccccc} & B_1 & B_2 & . & . & B_m \\ \hline A_1 & . & . & . & . & . \\ A_2 & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ A_n & . & . & . & . & . \end{array}$$

Note that:

- ❖ $P(A_i,B_j) = P(A_i).\ P(B_j / A_i) = P(B_j).\ P(A_i / B_j)$
- ❖ $\sum_{i=1}^{n} P(A_i / B_j) = 1$ and $\sum_{j=1}^{m} P(Bj / A_i) = 1$

Statical Independent : if A_i has no effect on the prob. of B_j then its called independent.

$$P(A_i / B_j) = P(A_i)$$
$$P(B_j / A_i) = P(B_j)$$
$$P(A_i,B_j) = P(A_i).\ P(B_j)$$

Ex: Two experiments A & B has the joint prob. matrix

$$P(A_i,B_j) = \begin{matrix} 0.1 & 0.25 \\ 0 & 0.2 \\ 0.25 & 0.2 \end{matrix}$$ ***. Find P(A),P(B), P(A/B) & P(B/A)***

Sol:

$$P(A_1)= \sum_{j=1}^{2} P(A_1,B_j) = 0.1+0.25 = 0.35$$

$$P(A_2)= \sum_{j=1}^{2} P(A_2,B_j) = 0+0.2 = 0.2$$

$$P(A_3)= \sum_{j=1}^{2} P(A_3,B_j) = 0.25+0.2 = 0.45$$

$$P(B_1)= \sum_{i=1}^{3} P(A_i,B_1) = 0.1+0+0.25 = 0.35$$

$$P(B_2)= \sum_{i=1}^{3} P(A_i,B_2) = 0.25+0.2+0.2 = 0.65$$

$$P(A_i,B_j) = P(A_i).\ P(B_j / A_i)$$

$$P(Bj/Ai) = \frac{P(Ai,Bj)}{P(Ai)} = \begin{matrix} \frac{0.1}{0.35} & \frac{0.25}{0.35} \\ 0 & \frac{0.2}{0.2} \\ \frac{0.25}{0.45} & \frac{0.2}{0.45} \end{matrix} = \begin{matrix} \frac{2}{7} & \frac{5}{7} \\ 0 & \frac{1}{1} \\ \frac{5}{9} & \frac{4}{9} \end{matrix}$$

$$P(Ai/Bj) = \frac{P(Ai,Bj)}{P(B_j)} = \begin{matrix} \frac{0.1}{0.35} & \frac{0.25}{0.65} \\ 0 & \frac{0.2}{0.65} \\ \frac{0.25}{0.35} & \frac{0.2}{0.65} \end{matrix} = \begin{matrix} \frac{2}{7} & \frac{5}{13} \\ 0 & \frac{4}{13} \\ \frac{5}{7} & \frac{4}{13} \end{matrix}$$

Random Variable (R.V):

a. **Discrete R.V.:** Recall the case of dice, each face is numbered as 1,2,…,6. if the dice is fair, then:-

P(1) = P(2) = P(3) = ….P(6) = 1/6

Also $\sum_{1}^{6} P(X_i) = 1$

$\overline{X} =$ mean of R.V= $\sum X_i . P(X_i)$

$\overline{X^2} =$ mean square of R.V= $\sum X_i^2 . P(X_i)$

$\sigma^2 =$ variance of R.V. = $\overline{X^2}$ - $\left(\overline{X}\right)^2$

H.W1: Find $\overline{X}$, $\overline{X^2}$ and σ^2 for previous example.

b. **Continuous R.V.:** Here X can be all real values not discrete then we call P(X)=PDF=Prob. Density function that gives the prob. That X lies between any two points X_1 & X_2.

$$P(X_2 > X > X_1) = \int_{X_1}^{X_2} P(X)dx$$

note that :

- $\int_{-\infty}^{\infty} P(X)dx = 1$

$$\overline{X} = \int_{-\infty}^{\infty} X.P(X)dx$$

$$\overline{X^2} = \int_{-\infty}^{\infty} X^2.P(X)dx$$

$$\sigma^2 = \overline{X^2} - \left(\overline{X}\right)^2$$

- If X is a random voltage signal then $\overline{X} =$ D.C value of X, $\overline{X^2} =$ total power (normalized on X) & $\sigma^2 =$ A.C. power of X.

Ex: If X I a continuous R.V. having the following PDF. Find:

- ***a. Constant k*** ***c.*** $\overline{X}$ ***,*** $\overline{X^2}$ ***,*** σ^2
- ***b. P(X>1)***

Sol:

a. $\int_{-\infty}^{\infty} P(X)dx = 1$ $\int_{-2}^{2} P(X)dx = \int_{-2}^{2} \frac{1}{2}.K.dx = \frac{1}{2}.K.(4)$

$K = \frac{1}{2}$

b. $P(X>1) = \int_{1}^{2} P(X)dx = \int_{1}^{2} \frac{1}{2} - \frac{X}{4} = \frac{X}{2} - \frac{X^2}{8}\Big|_{1}^{2} = \frac{1}{8}$

c. $\overline{X} = \int_{-2}^{0} X.(\frac{1}{2} + \frac{X}{4})dx + \int_{0}^{2} X.(\frac{1}{2} - \frac{X}{4})dx = 0$

$\overline{X^2} = 2.\int_{0}^{2} X^2.(\frac{1}{2} - \frac{X}{4})dx = \frac{2}{3}$

$\sigma^2 = \overline{X^2} - \left(\overline{X}\right)^2 = \frac{2}{3} - 0 = \frac{2}{3}$

H.W2: two dice are thrown, the sum of points appearing on the two dice is a random variable (X). Find the value of the R.V. taking by X & corresponding probabilities.

H.W3: If P(X) = $\frac{a}{2}e^{-a|x|}$ ***, find*** $\overline{X}$ ***,*** $\overline{X^2}$ ***and*** σ^2

1.2 Introduction to information theory

The purpose of a communication system is to carry information-bearing baseband signal generated by an information source from one point to another over a communication channel, with high efficiency and reliability. Figure 1 illustrates the functional diagram and the basic elements of a digital communication system.

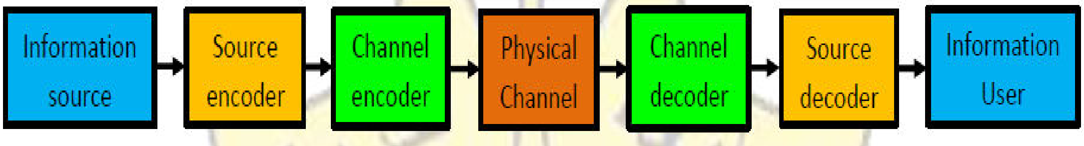


Figure 1. Communication system

- The information source may be either an analogue signal, such as an audio or video signal or a digital signal, such as the output of the computer that is discrete in time and has a finite number of the computer characters knows as information sequence.
- source encoding is a process of efficiently converting the output of either an analogue or digital source into a sequence of binary digits. It is also called data compression.

- Information theory provides a quantitative measure of the information contained in message signal and allows us to determine the capacity of a communication system to transfer this information from source to destination. Through the use of coding, redundancy can be reduced from message signal so that channels can be used with improved efficiency.

Self Information:

Suppose that te source of information produces finite set of messages X_1, X_2, …. X_n with prob. $P(X_1)$, $P(X_2)$,…$P(X_n)$, such that $\sum_{i=1}^{n} P(A_i) = 1$.

The amount of information gained from knowing that the source produces the messages X_i as follows:

1. Information is zero if $P(X_i)=1$.
2. Information increases as $P(X_i)$ decreases.
3. information is a positive quantity.

The function that relates $P(X_i)$ with information of X_i called:
$I(X_i)$ = self information of X_i.

$$I(X_i) = -\log_a P(X_i)$$

- the unit of $I(X_i)$ depends on a:
  1. If a=2, $I(X_i)$ has the unit of bits.
  2. if a=e=2.718, $I(X_i)$ has the unit of nat.
  3. if a=10, $I(X_i)$ has the unit of hartly.

Note that:

- $\text{Log}_a\, p = \dfrac{Ln(P)}{Ln(a)}$

Ex: A fair dice is thrown, find the amount of information gained if you are told that 4 will appear.

Sol :

Fair dice = $P(1)=P(2)=P(3)=P(4)=P(5)=P(6)=\frac{1}{6}$

$I(4)=-\text{Log}_a P(4) = -\text{Log}_2 \frac{1}{6} = \text{Log}_2 6$

$I(4)= \dfrac{Ln(6)}{Ln(2)} = 2.5844$ bit

Ex: Find the amount of information containing in a black and white TV picture if each picture has $2*10^5$ dots (pixels) and each pixel has 8 equal prob. Level of brightness.

Sol:

Information / pixel = $-\log_2 P(\text{level}) = -\log_2 \frac{1}{8} = 3$ bits

Information / picture = $3*2*10^5$ = 600 K bits

Source Entropy :
If $I(X_i)$, i=1,2,…n are different for a source producing un equal probability symbols, then the statical average of $I(X_i)$ will give the average amount of uncertainty associated with the source X, this average is called source entropy and denoted by H(X) and measured by bit per symbol.

$$H(X)=\sum_{i=1}^{n} P(X_i) I(X_i)$$

$$H(X)= -\sum_{i=1}^{n} P(X_i).\log_2 P(X_i)$$

Ex: Find the entropy of source producing the symbols.
P(X)=[0.25 0.1 0.15 0.5]

Sol:

$$H(X)= -\sum_{i=1}^{n} P(X_i).\log_2 P(X_i) = -\frac{1}{Ln(2)}[0.25 \text{ Ln } (0.25) + 0.1 \text{ Ln } (0.1) + 0.15 \text{ Ln } (0.15) + 0.5 \text{ Ln } (0.5)]$$

H(X)= 1.7427 bits/symbols

Ex: Find and plot the entropy of a binary source.

Sol:
P(0) + P(1) = 1
P(1) = 1 – P(0)

$$H(X)=\sum_{i=1}^{2} P(X_i).\log_2 P(X_i)$$
$$= -[P(0).\log_2 P(0) + (1-P(0). \log_2 (1-p(0))]$$

Note that:
- ❖ $H(X)=\log_2 n$, if the n symbols $X_1, X_2, \ldots X_n$ are equal probability
- ❖ H(X)=0, if one of the symbols has prob. = 1

Source Entropy Rate R(X) :

This is average rate amount of information produced per second.
R(X)=H(X)* rate of producing symbols

Rate of producing Symbols = $\frac{1}{\tau'}$

$$\tau' = \sum_{i=1}^{n} \tau_i . P(X_i)$$

τ' = average time duration of symbols
τ_i= time duration of X_i

$$R(X) = \frac{H(X)}{\tau'}$$

Ex: A source produces dots "•" & dashes "-" with probability P(dot) = 0.65, if time duration of a dot is 200 ms and that for a dash is 800 ms. Find the average source entropy R(X).

Sol:

P(dot) = 0.65 P(dash)=1-P(dot) =1-0.65 =0.35

τ dot = 200 ms , τ dash = 800 ms

$\tau' = \sum_{i=1}^{2} \tau_i . P(X_i)$ = [200*0.65 + 800*0.35]= 410 ms

H(X)= - [0.65 $\log_2$ (0.65) + 0.35 $\log_2$ (0.35)] = 0.934 bit/symbol

R(X) = $\frac{H(X)}{\tau'}$ = $\frac{0.934}{410}$ = 2.278 bit/sec

Mutual Information :

Consider the set of symbols X_1, $X_{2......}$ X_n can be produced. The receiver may receive Y_1, $Y_{2......}$ Y_m. if the noise and jamming are zero the set X = set Y and (n=m), however, due to noise and jamming, there will be conditional probability P(Y/X).

Definition:
$P(X_i)$ is called a priori prob. Of the symbol X_i which is the prob. Of selecting X_i for transmission.
$P(X_i/Y_i)$ is known a posteriori prob. Of X_i after the reception of Y_i.

The amount of information that Y_i provides about X_i is called "Mutual Information" between X_i & Y_i . This is given by :

$$I(X_i, Y_i) = \log_2 \text{(a posteriori prob.)/(a priori prob.)} = \log_2 \frac{P(X_i/Y_i)}{P(X_i)}$$

Note that :

- $I(X_i, Y_i) = I(Y_i, X_i) = \log_2 \frac{P(Y_i/X_i)}{P(Y_i)}$

Properties of $I(X_i, Y_i)$:

1. $I(X_i, Y_i)$ is symmetric i.e. $I(X_i, Y_i) = I(Y_i, X_i)$
2. $I(Y_i, X_i) > 0$, if a posterior prob. > priori prob.
   Then Yi provides +ve information about X_i.
3. $I(Y_i, X_i) = 0$, if a posterior prob. = priori prob.
   Then Yi provides no information about X_i.
4. $I(Y_i, X_i) < 0$, if a posterior prob. = priori prob.
   Then Yi provides or adds ambiguity (fuzzy) to X_i.

Marginal Entropy:
A term usually used to denote both source entropy H(X) & receiver entropy (Y(X).

$$H(Y) = -\sum_{j=1}^{m} P(Y_j).\log_2 P(Y_j) \quad \text{bit/symbol}$$

Joint & Conditional Entropies:

The average amount of information associated with the pair (X_i, Y_i) is called joint (system) entropy.

$$H(X,Y)= H(XY)= - \sum_{j=1}^{m}\sum_{i=1}^{n} P(X_i,Y_j).\log_2 P(X_i,Y_j)$$

The average amount of information associated with the pair (X_i / Y_i) & (Y_i / X_i) are called conditional entropy.

$$H(Y/X)= - \sum_{j=1}^{m}\sum_{i=1}^{n} P(X_i,Y_j).\log_2 P(Y_j / X_i) \quad \text{Noise Entropy.}$$

$$H(X/Y)= - \sum_{j=1}^{m}\sum_{i=1}^{n} P(X_i,Y_j).\log_2 P(X_i / Y_j) \quad \text{Losses Entropy.}$$

TransInformation:

Average mutual information, this is statical average of all pairs I(X_i, Y_i)

$$I(X,Y)= \sum_{j=1}^{m}\sum_{i=1}^{n} P(X_i,Y_j).I(X_i,Y_j)$$

$$=\sum_{j=1}^{m}\sum_{i=1}^{n} P(X_i,Y_j).\log_2 \frac{P(X_i / Y_j)}{P(X_i)}$$

$$=\sum_{j=1}^{m}\sum_{i=1}^{n} P(X_i,Y_j).\log_2 \frac{P(Y_j / X_i)}{P(Y_J)}$$

It is measured by bits/symbol.

Ex: Show that H(X,Y) = H(X) + H(Y/X)

Sol:

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

$$= -\sum_{j=1}^{m}\sum_{i=1}^{n} P(X_i,Y_j).\log_2 P(X_i)\ P(Y_j/X_i)$$

$$= -\sum_{j=1}^{m}\sum_{i=1}^{n} P(X_i,Y_j).\log_2 P(X_i) - \sum_{j=1}^{m}\sum_{i=1}^{n} P(X_i,Y_j).\log_2 P(Y_j/X_i)$$

$$= -\sum_{i=1}^{n} P(X_i).\log_2 P(X_i) - \sum_{j=1}^{m}\sum_{i=1}^{n} P(X_i,Y_j).\log_2 P(Y_j/X_i)$$

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

H.W. 4: show that H(X,Y)=H(Y)+ H(X/Y)

H.W. 5: show that I(X,Y)=H(X)-H(X/Y)

H.W. 6: show that I(X,Y)=H(Y)-H(Y/X)

- ❖ The prove above shows that the transinformation I(X,Y) is the Net average information gained at R_x which is the difference between the source information produced by the source H(X) and the information lost in the channel H(X/Y) or (H(Y/X) due to noise and jamming.

Ex: The joint prob. is given by

$$P(X_i,Y_j)=\begin{matrix} 0.5 & 0.25 \\ 0 & 0.125 \\ 0.0625 & 0.0625 \end{matrix}$$

Find : ***1. Marginal entropies***
2. System Entropies
3. Noise and losses entropies
4. Mutual information between X_1 and Y_2
5. Transinformation
6. Draw the channel model

Sol:

1. $P(X_i) = \sum_{j=1}^{2} P(X_i,Y_j) = [0.75 \quad 0.125 \quad 0.125]$

$P(Y_i) = \sum_{i=1}^{3} P(X_i,Y_j) = [0.5625 \; 0.4375]$

$$H(X) = -\sum_{i=1}^{3} P(X_i).\log_2 P(X_i) = -\frac{1}{\ln(2)}[0.75\ln(0.75) + 2*0.125\ln(0.125)]$$
$$= 1.06127 \text{ bits/symbol}$$
$$H(Y) = -\sum_{j=1}^{2} P(Y_j).\log_2 P(Y_j) = -\frac{1}{\ln(2)}[0.5625\ln(0.5625) + 0.4375\ln(0.4375)]$$
$$= 0.9887 \text{ bits/symbol}$$

2. $H(X,Y)=)= -\sum_{j=1}^{2}\sum_{i=1}^{3} P(X_i,Y_j).\log_2 P(X_i,Y_j) = -\frac{1}{\ln(2)}[0.5\ln(0.5)+$
$0.25\ln(0.25)+0.125\ln(0.125)+2*0.0625\ln(0.0625)]$
$=1.875$ bits/symbols

3. $H(Y/X)= H(X,Y) - H(X) = 1.875 – 1.06127 = 0.81373$ bit/symbol.

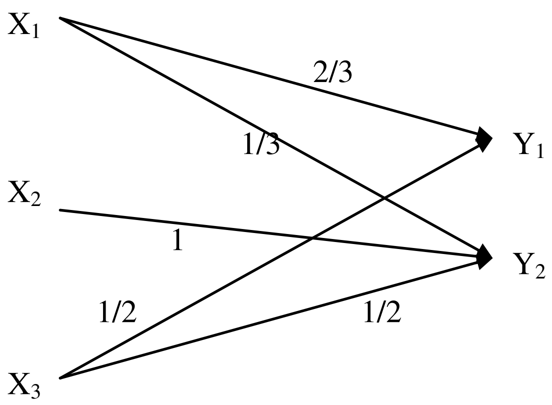
$H(X/Y)= H(X,Y) - H(Y) = 1.875 – 0.9887 = 0.8863$ bit/symbol

4. $I(X_1, Y_2) = \log_2 \frac{P(X_1/Y_2)}{P(X_1)}$ since $P(X_1/Y_2) = \frac{P(X_1,Y_2)}{P(Y_2)}$

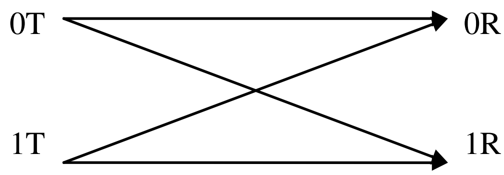
$$= \log_2 \frac{P(X_1,Y_2)}{P(X_1)P(Y_2)}$$
$$= \log_2 \frac{0.25}{0.75*0.4375} = -0.3923 \text{ bits}$$

5. $I(X,Y)=H(X)-H(X/Y) = 0.17497$ bits/symbol
6. To draw a channel, we find $P(Y_j/X_i) = \frac{P(X_i, Y_j)}{P(X_i)}$

$$P(Y_j/X_i) = \begin{matrix} \frac{0.5}{0.75} & \frac{0.25}{0.75} \\ \frac{0}{0.125} & \frac{0.125}{0.125} \\ \frac{0.0625}{0.125} & \frac{0.0625}{0.125} \end{matrix} = \begin{matrix} \frac{2}{3} & \frac{1}{3} \\ \frac{0}{1} & \frac{1}{1} \\ \frac{1}{2} & \frac{1}{2} \end{matrix}$$



Ex: Find & plot the transinformation for binary symmetric channel (BSC) shown below if P(0T)=P(1T)=0.5



Sol:

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

Let $0T = X_1$ & $1T = X_2$

$0R = Y_1$ & $1R = Y_2$

$$P(Y/X)= \begin{matrix} 1-Pe & Pe \\ Pe & 1-Pe \end{matrix}$$

$$P(X_i,Y_j)=P(X_i).P(Y_j/X_i)= \begin{matrix} \frac{1-Pe}{2} & \frac{Pe}{2} \\ \frac{Pe}{2} & \frac{1-Pe}{2} \end{matrix}$$

$P(Y_j)= [0.5 \;\; 0.5]$

$$H(Y/X)= -\sum_{j=1}^{m}\sum_{i=1}^{n} P(X_i,Y_j).\log_2 P(Y_j/X_i)$$

$$= -\frac{1}{\ln 2}[2*\frac{1-Pe}{2}\ln(1-Pe)+2*\frac{Pe}{2}\ln(Pe)]$$

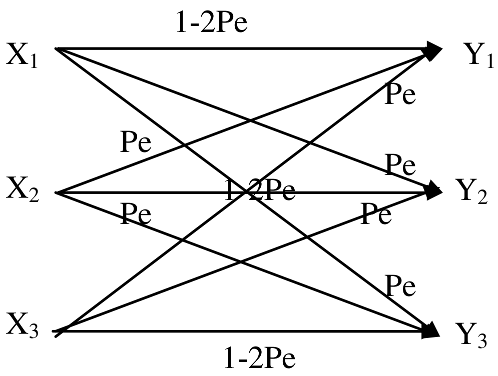
$$= -\frac{1}{\ln 2}[(1-Pe)\ln(1-Pe)+(Pe)\ln(Pe)]$$

$$I(X,Y)=1+\frac{1}{\ln 2}[(1-Pe)\ln(1-Pe)+(Pe)\ln(Pe)]$$

Pe	I(X,Y)
0	1
0.5	0
1	1

Ternary Symmetric Channel (TSC)

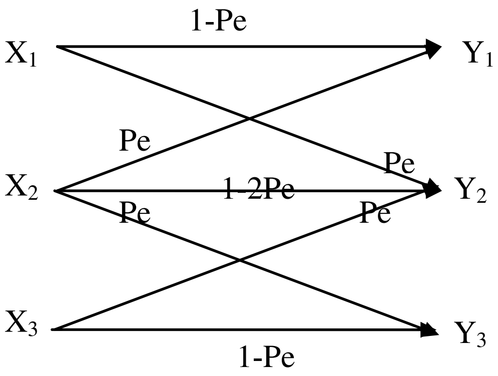
$$P(\mathrm{Y/X})=\begin{bmatrix} 1-2Pe & Pe & Pe \\ Pe & 1-2Pe & Pe \\ Pe & Pe & 1-2Pe \end{bmatrix}$$



- This TSC is symmetric but not practical since (doesn't effect that much on X_3).
- There is no chance that X_1 is received as Y_3 or X_3 as X_1 . Hence, a non symmetrical channel but more practical is shown :-

Transision channel matrix

$$P(\mathrm{Y/X})=\begin{bmatrix} 1-Pe & Pe & 0 \\ Pe & 1-2Pe & Pe \\ 0 & Pe & 1-Pe \end{bmatrix}$$



Other special channels:

1. ***Lossless channel***: This channel has only one non-zero elements in each column of P(Y/X) . This channel has H(X/Y)=0 and I(X,Y)=H(X)

$$P(\mathrm{Y/X})=\begin{bmatrix} \frac{3}{4} & \frac{1}{4} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{3} & \frac{2}{3} & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

2. ***Deterministic channel*** : This channel has only one non-zero element in each row of P(Y/X). This has H(Y/X)=0 and I(X,Y)=H(Y)

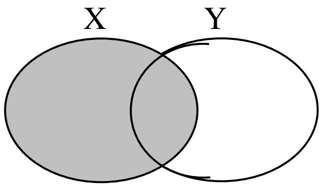
$$P(Y/X)=\begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

3. ***Noiseless Channel:*** This channel has n=m and P(Y/X) is an identity matrix. I(X,Y)=H(X/Y)=H(Y/X) & H(Y/X)=H(X/Y).

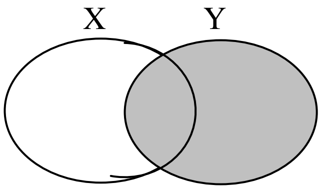
$$P(Y/X)=\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

note that this channel is both deterministic and lossless.

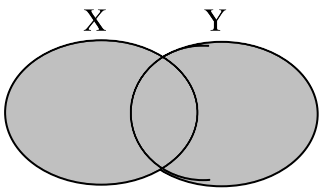
Veen Diagram of Representation of the channel.



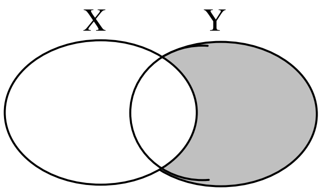
H(X)



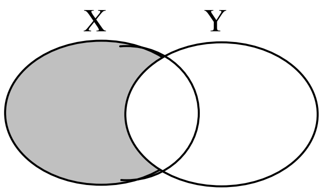
H(Y)



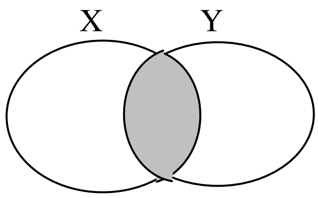
H(X,Y)



H(Y/X)



H(X/Y)



I(X,Y)

Channel Capacity C:

The limiting rate of information transmission through a channel is called channel capacity. Channel capacity is the maximum rate at which reliable transmission of information over the channel is possible.
At data rate < C reliable transmission over the channel is possible.
At data rate >C reliable transmission is not possible.

The channel capacity is also defined as the maximum of I(X,Y) measured in bit / symbol.

1. **Channel capacity for symmetric channel**
   A general definition of symmetric channel is that n=m and each channel is a permutation of the other rows.

   To find $I(X,Y)_{max}$ for such symmetric channel then:

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

$$= H(Y) + \sum_{j=1}^{m}\sum_{i=1}^{n} P(X_i, Y_j).\log_2 P(Y_j / X_i)$$

$$= H(Y) + \sum_{j=1}^{m}\sum_{i=1}^{n} P(X_i)P(Y_j / X_i).\log_2 P(Y_j / X_i)$$

$$= H(Y) + \sum_{j=1}^{m} P(Y_j/X_i).\log_2 P(Y_j/X_i)$$

$$\sum_{j=1}^{m} P(Y_j/X_i).\log_2 P(Y_j/X_i) = K = \text{constant}$$

$$= H(Y) + K$$

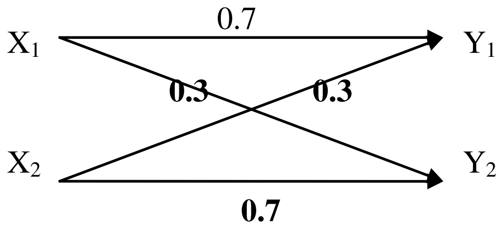
$$C = I(X,Y)_{max} = [\, H(Y) + K \,]_{max} = \log_2 m + K$$

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

$$\text{Channel Redundancy} = R = 1 - \frac{I(X,Y)}{C}$$

Ex: Find the channel capacity for BSC shown then find channel redundancy, if $I(X_1)$=2 bits

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



Sol:

$$C = I(X,Y)_{max} = [\ H(Y) + K\ ]_{max} = \log_2 m + K = 1+K$$

$$\sum_{j=1}^{2} P(Y_j/X_i).\log_2 P(Y_j/X_i) = K$$

$$0.7\log_2 0.7+0.3\log_2 0.3=-0.61086$$

$$C=1-0.61086 = 0.389 \text{ bit / symbol}$$

To find channel redundancy we must find I(X,Y)

If $I(X_1) = 2= - \log_2 P(X_1)$

$P(X_1) = 1/4$

$P(X_2)=3/4$

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

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

$$H(Y)= \frac{-1}{\ln(2)}[(0.4).\ln(0.4)+(0.6)\ln(0.6)] = 0.97095 bit/symbol$$

$$I(X,Y)=0.97075-0.61086=0.36 \text{ bit/symbol}$$

$$R=1-0.36/0.38914=7.46\%$$

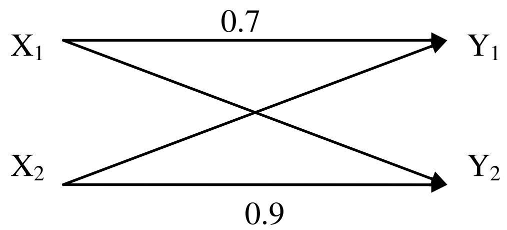
2. Channel capacity for non-symmetric channel

Procedure:

- Find I(X,Y) as a function of input prob.
  $I(X,Y)=f[P(X_1), P(X_2),\ldots P(X_n)]$
  $\sum_{i=1}^{n} P(X_i)=1$ & use this constrain to reduce the number of variables by 1.
- To maximize I(X,Y), differentiate I(X,Y) with respect to $P(X_1), P(X_2),\ldots P(X_n)$ and then equate to zero.
- Find the input prob. $P(X_1), P(X_2),\ldots P(X_n)$ that make I(X,Y) max.

Ex: find the channel capacity for the channel shown:

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



Sol:

Let $P(X_1)=p$ then $P(X_2)=1-p$

I(X,Y)= H(Y) - H(Y/X)-f(p)

$$P(X,Y)=P(X).P(Y/X)=\begin{bmatrix}p & 1-p\end{bmatrix}\begin{bmatrix}0.7 & 0.3\\ 0.1 & 0.9\end{bmatrix}=\begin{bmatrix}0.7p & 0.3p\\ 0.1(1-p) & 0.9(1-p)\end{bmatrix}$$

P(Y)= [0.1+0.6p 0.9-0.6p]

$$H(Y)=\frac{-1}{\ln(2)}[(0.1+0.6p).\ln(0.1+0.6p)+(0.9-0.6p)\ln(0.9-0.6p)]$$

$$H(Y/X) = -\sum_{j=1}^{m}\sum_{i=1}^{n} P(X_i,Y_j).\log_2 P(Y_j/X_i)$$

$$= \frac{-1}{\ln(2)}[0.7p.\ln 0.7+0.3p\ln 0.3+0.1(1-p)\ln 0.1+0.9(1-p)\ln 0.9]$$

$$\frac{\partial H(Y/X)}{\partial p}=\frac{-1}{\ln(2)}[0.7.\ln 0.7+0.3\ln 0.3-0.1\ln 0.1-0.9\ln 0.9]$$

$$= \frac{-1}{\ln(2)}[-0.285781]$$

$$\frac{\partial H(Y)}{\partial p} = \frac{-1}{\ln(2)}[0.6 + 0.6.\ln(0.1+0.6p) - 0.6 - 0.6.\ln(0.9-0.6p)]$$

$$= \frac{-1}{\ln(2)}[0.6\ln\frac{0.1+0.6p}{0.9-0.6p}]$$

$$\frac{\partial I(X,Y)}{\partial p} = \frac{\partial H(Y)}{\partial p} - \frac{\partial H(Y/X)}{\partial p} = Zero$$

$$0.6\ln\frac{0.1+0.6p}{0.9-0.6p} + 0.285781 = 0$$

$$\left[\ln\frac{0.1+0.6p}{0.9-0.6p} = -\frac{0.285781}{0.6}\right].\ln^{-1}$$

$$\frac{0.1+0.6p}{0.9-0.6p} = e^{\frac{0.285781}{0.6}} \Rightarrow p = 0.47187$$

Substitute p in H(Y) and H(Y/X) to find $I(X,Y)_{max}$

H(Y) = -0.96021
H(Y?X)= 0.66354

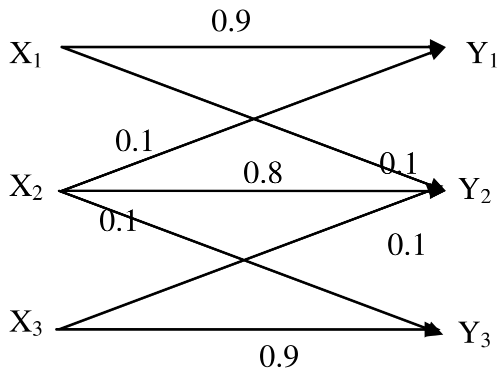
I(X,Y)= H(Y) - H(Y/X) = 0.2966 bits/symbol

NOTE:

Sometimes to ease calculation, we are asked to find channel capacity when channel is non-symmetric but there are some similarities between some symbols (not all). In such case we can satisfy that by assuming theses symbols are equal probability and proceed as in previous example.

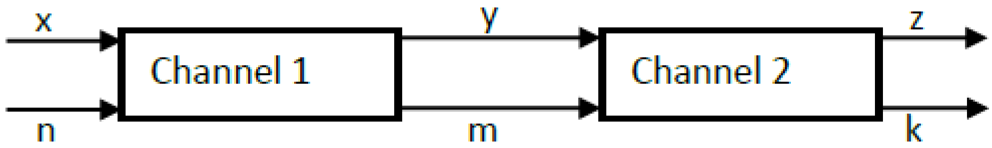
Ex:

We can assume
$P(X_1) = P(X_3) = p$
Then
$P(X_2) = 1-2p$



Cascading of channels

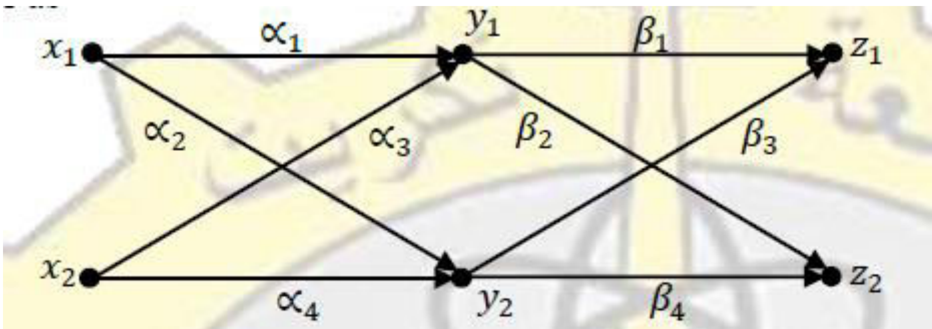
If two channels are cascaded then the over all transition matrix is the product of the two transition matrices.



$$[p(z|x)] = [p(y|x)][p(z|y)]$$

$$n \times k \quad = \quad (n \times m)\ (m \times k)$$

For example, a binary satellite communication system can often be represented by the cascade combination of two binary channels. The first one represents the uplink and the second represents the downlink. These channels can be combined together as



By determining all possible paths X_i to Z_j , it is clear that the following probabilities can be used to find the overall channel.

$$p_{11} = \propto_1 \beta_1 + \propto_2 \beta_3$$

$$p_{12} = \propto_1 \beta_2 + \propto_2 \beta_4$$

$$p_{21} = \propto_3 \beta_1 + \propto_4 \beta_3$$

$$p_{22} = \propto_3 \beta_2 + \propto_4 \beta_4$$

Thus, the overall channel matrix can be expressed as

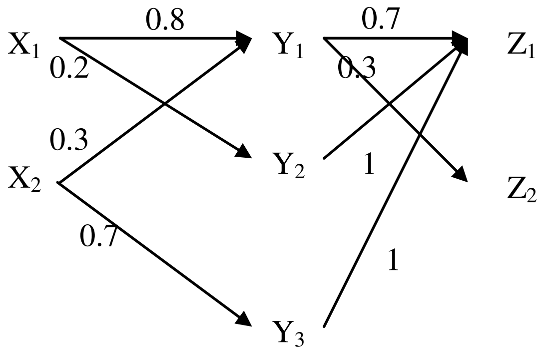
$$[p(z|x)] = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} = \begin{bmatrix} \propto_1 & \propto_2 \\ \propto_3 & \propto_4 \end{bmatrix} \begin{bmatrix} \beta_1 & \beta_2 \\ \beta_3 & \beta_4 \end{bmatrix}$$

Ex: find the transition matrix P(Z/X) for the cascaded channels shown P(X) = [0.7 0.3]

Sol:

$$P(Y/X) = \begin{bmatrix} 0.8 & 0.2 & 0 \\ 0.3 & 0 & 0.7 \end{bmatrix}$$

$$P(Z/Y) = \begin{bmatrix} 0.7 & 0.3 \\ 1 & 0 \\ 1 & 0 \end{bmatrix}$$



$$P(Z/X)=P(Y/X).P(Z/X)= \begin{bmatrix} 0.76 & 0.24 \\ 0.91 & 0.09 \end{bmatrix}$$

H.W 7 : find the joint prob. and then find P(Y) & P(Z) of the example.

P(X,Y)=P(X)P(Y/X)
P(X,Z)=P(X)P(Z/X)