Chapter one

Review of Probability

1- Definition: A probabilistic model is a mathematical description of an uncertain situation.

A probability of an event A: If an experiment has A_1, A_2,……. A_n, outcomes, then:

$$Prob(A_i) = P(A_i) = \lim_{N\to\infty} \frac{n(A_i)}{N}$$

Where $n(A_i)$= no. of times event (outcomes) (A_i) occurs

N= total number of trails.

2- Not that

$$1 \geq P(A_i) \geq 0, \quad \text{and}$$

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

If $P(A_i) = 1$ then A_i is certain event

When the sample space **Ω** has a finite number of equally likely outcomes, so that the discrete uniform probability law applies. Then, the probability of any event A is given by

$$P(A) = \frac{Number\ of\ elements\ of\ A}{Number\ of\ elements\ of\ \Omega}$$

3- Conditional Probability:

Conditional probability provides us with a way to reason about the outcome of an experiment, based on partial information. For any event A, gives us the conditional probability of A given B, denoted by P(A | B). For example, suppose that all six possible outcomes of a fair die roll are equally likely. If we are told that the outcome is even, we are left with only three possible outcomes, namely, 2, 4, and 6.

P(the outcome is 6 |the outcome is even) = 1/3

The definition of conditional probability when all outcomes are equally likely, is given by:

$$P(A \mid B) = \frac{number\ of\ elements\ of\ A\ \cap\ B}{number\ of\ elements\ of\ B}$$

Or

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

Where $P(B) > 0$

Example: We toss a fair coin three successive times. We wish to find the conditional probability P(A | B) when A and B are the events

A = {more heads than tails come up}, B = {1st toss is a head}.

The sample space consists of eight sequences,

Ω = {HHH, HHT, HTH, HTT, THH, THT, TTH, TTT},

$$P(B) = \frac{4}{8}$$

$$P(A \cap B) = \frac{3}{8}$$

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{3}{8}}{\frac{4}{8}} = 3/4$$

4- Total Probability Theorem

Let $A_1,...,A_n$ be disjoint events that form a partition of the sample space (each possible outcome is included in one and only one of the events A_1 ,...,A_n) and assume that $P(A_i) > 0$, for all i = 1,...,n. Then, for any event B, we have

$$P(B) = P(A_1 \cap B) + \cdots + P(A_n \cap B)$$

$$= P(A_1)P(B \mid A_1) + \cdots + P(A_n)P(B \mid A_n).$$

$$P(A_i) = \frac{\text{number of elements of } A_i}{\text{total number of possible outcomes}}$$

Bayes' Rule: Let A_1, A_2,...,A_n be disjoint events that form a partition of the sample space, and assume that P(Ai) > 0, for all i. Then, for any event B such that P(B) > 0, we have

$$P(A_i \mid B) = \frac{P(A_i\ )P(B|A_i)}{P(B)}$$

$$= \frac{P(A_i\ )P(B|A_i)}{P(A_1\ )P(B|A_1) + .. + P(A_n\ )P(B|n)}$$

5- Independence:

We say that the events A_1, A_2,...,A_n are independent if

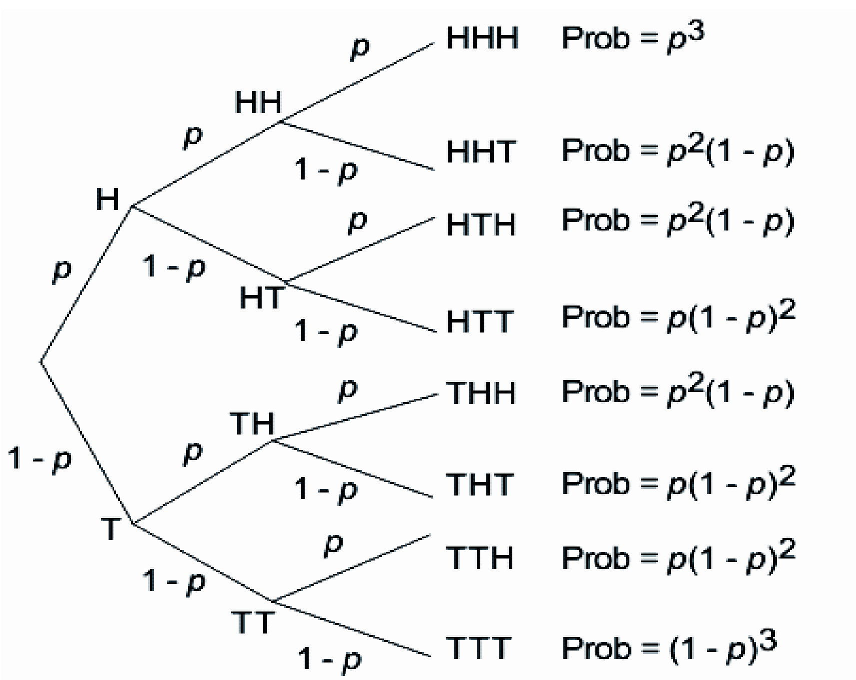
$$\mathbf{P}\left(\bigcap_{i \in S} A_i\right) = \prod_{i \in S} \mathbf{P}(A_i), \qquad \text{for every subset } S \text{ of } \{1, 2, \ldots, n\}.$$

6- Independent Trials

If an experiment involves a sequence of independent but identical stages, we say that we have a sequence of independent trials. In the special case where there are only two possible results at each stage, we say that we have a sequence of independent Bernoulli trials. Consider an experiment that consists of *n* independent tosses of a biased coin, in which the probability of "heads" is *p*, where *p* is some number between 0 and 1.

Let us now consider the probability

p(k)=P(k heads come up in an n-toss sequence),



The probability of any given sequence that contains k heads is $P^k(1-P)^{n-k}$, so we have

$$P(k) = \binom{n}{k} P^k (1-P)^{n-k}$$

The numbers $\binom{n}{k}$ (called "n choose k") are known as the binomial coefficients, while the probabilities p(k) are known as the binomial probabilities.

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!} \qquad\qquad k = 1, 2, \ldots\ldots..n$$

Note that the binomial probabilities p(k)must add to 1, thus showing the binomial formula:

$$\sum_{k=0}^{n} \binom{n}{k} P^k (1-P)^{n-k} = 1$$

7- Random variable

Mathematically, a random variable is a real-valued function of the experimental outcome. For example, in an experiment involving the transmission of a message, the time needed to transmit the message, the number of symbols received in error, and the delay with which the message is received are all random variables.

A random variable is called discrete if its range (the set of values that it can take) is finite or at most countably infinite.

On the other hand, the random variable that associates with a the numerical value

$$sign(a) = \begin{cases} 1 & if\ a > 0 \\ 0 & if\ a = 0 \\ -1 & if\ a < 0 \end{cases}$$

Is discrete.

8- Probability Mass Function (PMF)

The most important way to characterize a random variable is through the probabilities of the values that it can take. If x is any possible value of X, the probability mass of x, denoted $p_X(x)$, is the probability of the event$\{X = x\}$ consisting of all outcomes that give rise to a value of X equal to x:

$$p_X(x) = P(\{X = x\})$$

For example, let the experiment consist of two independent tosses of a fair coin, and let X be the number of heads obtained. Then the PMF of X is

$$p_X(x) = \begin{cases} \frac{1}{4} & if\ x = 0 \quad or\ x = 2 \\ 1/2 & if\ x = 1 \\ 0 = & otherwise \end{cases}$$

Note that

$$\sum_x p_X(x) = 1$$

The Bernoulli Random Variable: Consider the toss of a biased coin, which comes up a head with probability p, and a tail with probability $1-p$. The Bernoulli random variable takes the two values 1 and 0, depending on whether the outcome is a head or a tail:

$$X = \begin{cases} 1 & if\ a\ head \\ 0 & if\ a\ tail \end{cases}$$

Its PMF is

$$p_X(x) = \begin{cases} P & if\ x = 1 \\ 1 - P & if\ x = 0 \end{cases}$$

9- Probability Density Function (PDF)

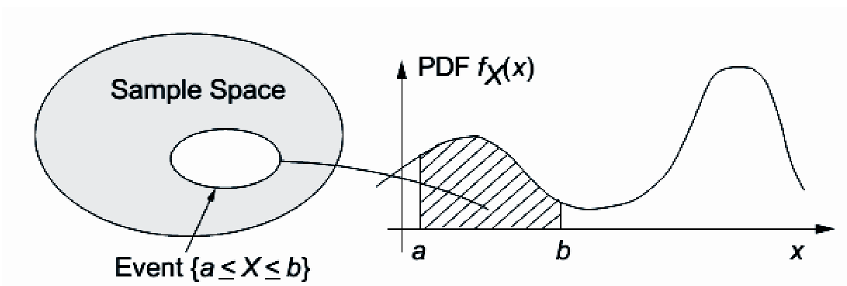
A random variable X is called continuous if its probability law can be described in terms of a nonnegative function X, called the probability density function of X, or PDF for short, which satisfies

$$P(X \in B) = \int_B f_X(x)dx$$

for every subset B of the real line. In particular, the probability that the value of X falls within an interval is

$$P(a \leq X \leq b) = \int_a^b f_X(x)dx$$

and can be interpreted as the area under the graph of the PDF as shown

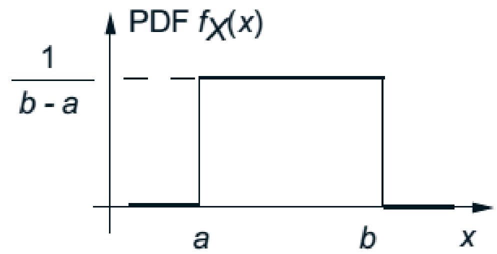


More generally, we can consider a random variable X that takes values in an interval [a, b], and again assume that all subintervals of the same length are equally likely. We refer to this type of random variable as uniform or uniformly distributed. Its PDF has the form

$$f_X(x) = \begin{cases} c & if\ a \le x \le b \\ 0 & therwise \end{cases}$$

Where c is constant

$$1 = \int_a^b c dx \qquad 1 = 1 = c\int_a^b dx = c(b-a)$$



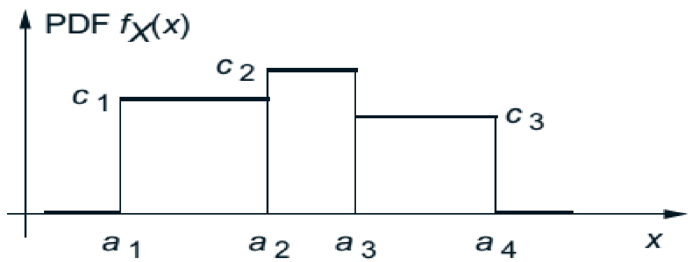
$$c = \frac{1}{b-a}$$

Generalizing, consider a random variable X whose PDF has the piecewise constant form

$$f_X(x) = \begin{cases} c_i & if\ a_i \le x \le a_{i+1} \quad i = 1, 2, \dots\dots, n-1 \\ & therwise \end{cases}$$

Where a_1, a_2,..., a_n are some scalars with a_i <a_i+1for all i, and c_1,c_2,...,c_n are some nonnegative constants

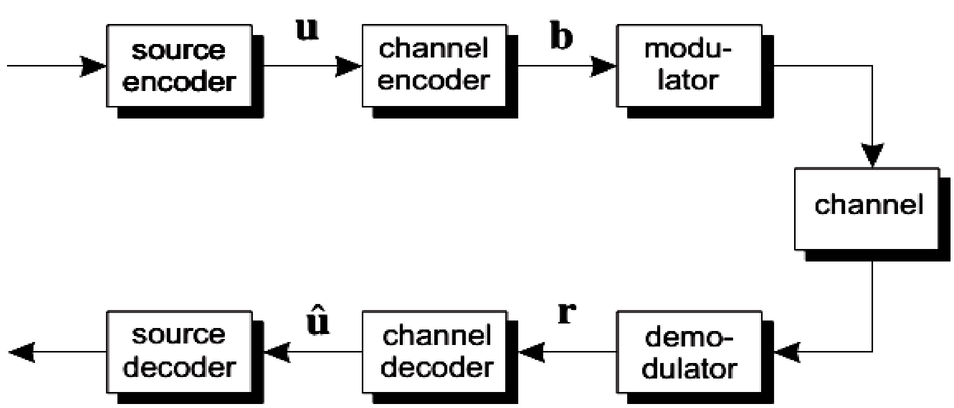
$$1 = \int_{a_1}^{a_n} f_X dx \quad = \sum_{i=1}^{n-1} \int_{a_1}^{a_{i+1}} c_i dx = \sum_{i=1}^{n-1} c_i(a_{i+1} - a_i)$$



Chapter Two

1- Digital Communication system

The reliable transmission of information over noisy channels is one of the basic requirements of digital information and communication systems. Because of this requirement, modern communication systems rely heavily on powerful channel coding methodologies. For practical applications these coding schemes do not only need to have good coding characteristics with respect to the capability of detecting or correcting errors introduced on the channel. They also have to be efficiently implementable, e.g. in digital hardware within integrated circuits. In Figure 1 the basic structure of a digital communication system is shown which represents the architecture of the communication systems in use today.



An important result of information theory is the finding that error-free transmission across a noisy channel is theoretically possible – as long as the information rate does not exceed the so-called channel capacity. In order to quantify this result, we need to measure information. Within Shannon's information theory this is done by considering the statistics of symbols emitted by information sources.

2- Entropy

In information theory, entropy is the average amount of information contained in each message received. Here, message stands for an event, sample or character drawn from a distribution or data stream. Entropy thus characterizes our uncertainty about our source of information.

Self- information:

In information theory, self-information is a measure of the information content associated with the outcome of a random variable. It is expressed in a unit of information, for example bits, nats, or Hartley, depending on the base of the logarithm used in its calculation.

A bit is the basic unit of information in computing and digital communications. A bit can have only one of two values, and may therefore be physically implemented with a two-state device. These values are most commonly represented as 0 and 1.

A nat is the natural unit of information, sometimes also nit or nepit, is a unit of information or entropy, based on natural logarithms and powers of e, rather than the powers of 2 and base 2 logarithms which define the bit. This unit is also known by its unit symbol, the nat.

The hartley (symbol Hart) is a unit of information defined by International Standard IEC 80000-13 of the International Electrotechnical Commission. One hartley is the information content of an event if the probability of that event occurring is 1/10. It is therefore equal to the information contained in one decimal digit (or dit).

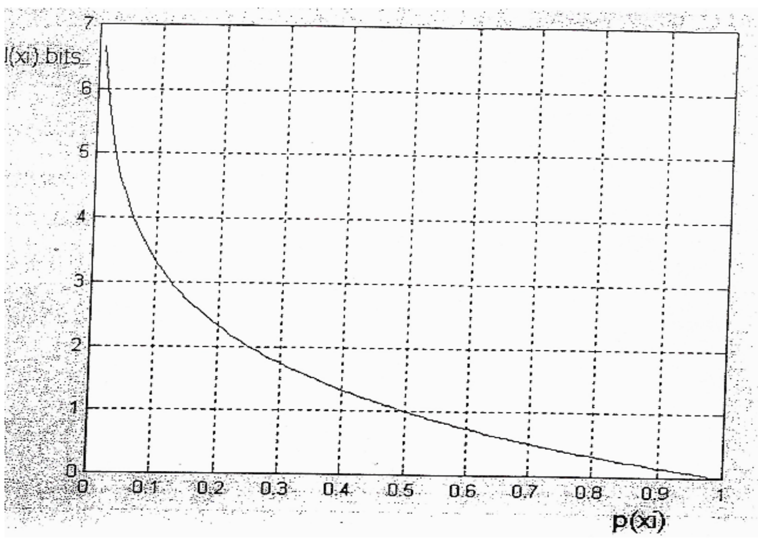
1 Hart ≈ 3.322 Sh ≈ 2.303 nat.

The amount of self-information contained in a probabilistic event depends only on the probability of that event: the smaller its probability, the larger the self-information associated with receiving the information that the event indeed occurred.

Suppose that the source of information produces finite set of message $x_1, x_2, \dots \dots . x_n$with prob. $p(x_1), p(x_2), \dots \dots \dots . P(x_n)$ and such that

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

1- Information is zero if $P(x_i) = 1$ (certain event)
2- Information increase as $P(x_i)$ decrease to zero
3- Information is a +ve quantity



The log function satisfies all previous three points hence:

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

Where $I(x_i)$ is self information of (x_i) and if:

i- If “a” =2 , then $I(x_i)$ has the unit of bits

ii- If “a”= e = 2.71828, then $I(x_i)$ has the unit of nats iii- If “a”= 10, then $I(x_i)$ has the unit of hartly

Recall that $\log_a x = \frac{lnx}{lna}$

Example 1:

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

Solution:

$$P(1) = P(2) = \cdots \ldots \ldots . = P(6) = \frac{1}{6}$$

$$I(4) = -\log_2\left(\frac{1}{6}\right) = \frac{\ln(\frac{1}{6})}{ln2} = 2.5849 \ \ bits$$

Example 2:

A biased coin has P(Head)=0.3. Find the amount of information gained if you are told that a tail will appear.

Solution:

$$P(tail) = 1 - P(Head) = 1 - 0.3 = 0.7$$

$$I(tail) = -\log_2(0.7) = -\frac{ln0.7}{ln2} = 0.5145 \ \ \ bits$$

4- Source Entropy:

If the source produces not equiprobable messages then $I(x_i), i = 1, 2, \ldots \ldots \ldots, n$ are different. Then the statistical average of $I(x_i)$ over i will give the average amount of uncertainty associated with source X. This average is called source entropy and denoted by $H(X)$, given by:

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

$$\therefore \quad H(X) = -\sum_{i=1}^{n} P(x_i) \log_a P(x_i)$$

Example 3:

Find the entropy of the source producing the following messages:

$$Px_1 = 0.25, \qquad Px_2 = 0.1, \;\; Px_3 = 0.15, \;\; and\; Px_4 = 0.5$$

Solution:

$$H(X) = -\sum_{i=1}^{n} P(x_i) \log_a P(x_i)$$

$$= -\frac{[0.25ln0.25 + 0.1ln0.1 + 0.15ln0.15 + 0.5ln0.5]}{ln2}$$

$$H(X) = 1.7427\ \frac{bits}{symbol}$$

Example 4:

Find and plot the entropy of binary source.

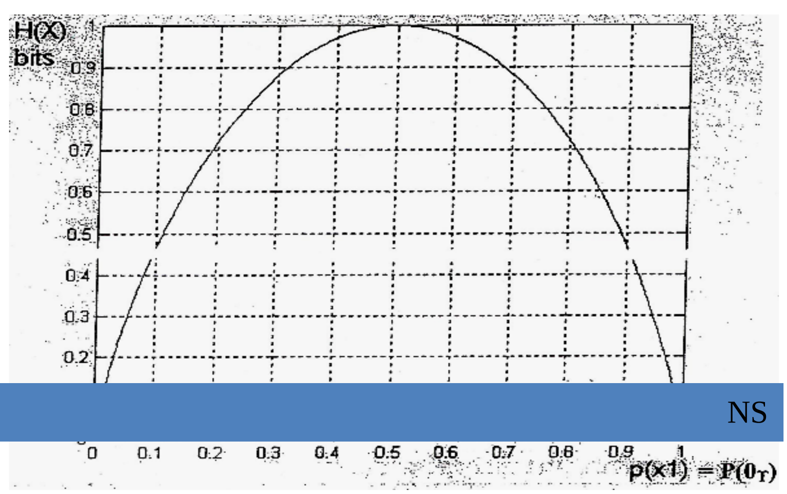
$$P(0_T) + P(1_T) = 1$$

$$H(X) = -[P(0_T)\log_2 P(0_T) + (1 - P(0_T))\ \log_2(1 - P(0_T))]\ bits/symbol$$

If $P(0_T) = 0.2, then\; P(1_T) = 1 - 0.2 = 0.8, and\; put\; in\; above\; equation,$

$$H(X) = -[0.2\log_2(0.2) + 0.8\log_2(0.8)] = 0.7$$

Not that H(X) is maximum equal to 1(bit) if: $P(0_T) = P(1_T) = 0.5$ as shown in figure.



If all messages are equiprobable, then $P(x_i) = 1/n$ so hat:

$$H(X) = H(X)_{max} = -[\frac{1}{n}\log_a\left(\frac{1}{n}\right)] \times n = -\log_a\left(\frac{1}{n}\right) = \log_a n \;\; bits/symbol$$

And $H(X) = 0$ if one of the message has the prob of a certain event.

5- Source Entropy Rate:

It is the average rate of amount of information produced per second.

$$R(X) = H(X) \times rate\ of\ producing\ the\ symbols \quad = \frac{bits}{sec} = bps$$

The unit of H(X) is bits/symbol and the rate of producing the symbols is symbol/sec, so that the unit of R(X) is bits/sec.

Sometimes $R(X) = \frac{H(X)}{\bar{\tau}}$,

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

$\bar{\tau}$ is the average time duration of symbols, τ_i is the time duration of the symbol x_i.

Example 6:

A source produces dots '.' And dashes '-' with P(dot)=0.65. If the time duration of dot is 200ms and that for a dash is 800ms. Find the average source entropy rate.

Solution:

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

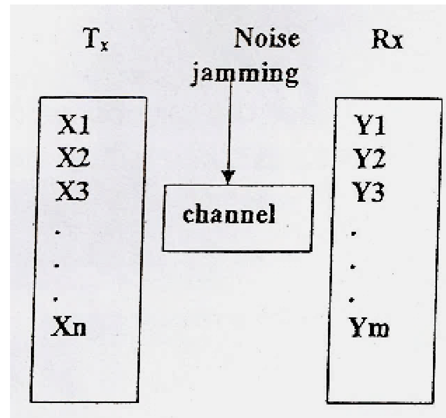
$H(X) = -[0.65\log_2(0.65) + 0.35\log_2(0.35)] = 0.934\ bits/symbol$

$\bar{\tau} = 0.2 \times 0.65 + 0.8 \times 0.35 = 0.41\ sec$

$$R(X) = \frac{H(X)}{\bar{\tau}} = \frac{0.34}{0.41} = 2.278\ bps$$

6- Mutual Information:

Consider the set of symbols $x_1, x_2, \ldots., x_n$, the transmitter T_x my produce. The receiver R_x may receive $y_1, y_2 \ldots\ldots\ldots. y_m$. Theoretically, if the noise and jamming, then the set X=setY. However and due to noise and jamming, there will be a conditional prob $P(y_j \mid x_i)$:



1- $P(x_i)$ to be what is so called the apriori prob of the symbol x_i, which is the prob of selecting x_i for transmission.

2- $P(y_j \mid x_i)$ to be what is called the aposteriori prob of the symbol x_i after the reception of y_j.

The amount of information that y_j provides about x_i is called the mutual information between x_i and y_i. This is given by:

$$I(x_i, y_j) = \log_2\left(\frac{aposterori\ prob}{apriori\ prob}\right) = \log_2\left(\frac{P\left(y_j \mid x_i\right)}{P(x_i)}\right)$$

Properties of $\boldsymbol{I(x_i, y_j)}$:

1- It is symmetric, $I\left(x_i, y_j\right) = I(y_j, x_i)$.

2- $I\left(x_i, y_j\right) > 0$ if aposteriori prob> apriori prob, y_j provides +ve information about x_i.

3- $I\left(x_i, y_j\right) = 0$ if aposteriori prob= apriori prob, which is the case of statistical independence when y_j provides no information about x_i.

4- $I\left(x_i, y_j\right) < 0$ if aposteriori prob< apriori prob, y_j provides -ve information about x_i, or y_j adds ambiguity.

Example 7:

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 can't decide anything about x_i as if we transmit a deterministic signal x_i but the receiver receives noise like signal y_j that is completely has no correlation with x_i. Then x_i and y_j are statistically independent so that $P\left(x_i \mid y_j \right) = P(x_i) and\ P\left(y_j \mid x_i \right) = P(x_i)\ for\ all\ i\ and\ j, then$:

$$I\left(x_i, y_j\right) = \log_2 1 = 0\ \ for\ all\ i\ \&\ j, then\ I(X, Y) = 0$$

7- Transinformation (average mutual information):

It is the statistical average of all pair $I(x_i, y_j)\ , i = 1, 2, \ldots\ldots, n,\ \ j = 1, 2, \ldots\ldots, m$.

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

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

$$I(X, Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} P\left(x_i, y_j\right) \log_2 \left(\frac{P\left(y_j \mid x_i \right)}{P\left(y_j\right)} \right) \frac{bits}{symbol}$$

or

$$I(X, Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} P(x_i, y_j) \log_2 \left(\frac{P\left(x_i \mid y_j \right)}{P(x_i)} \right)\ bits/symbol$$

Expand above equation:

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

And we have

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

And by substituting:

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

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

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

8- Marginal Entropies:

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

$$H(Y) = -\sum_{j=1}^{m} P\left(y_j\right)\log_2 P\left(y_j\right) \qquad \frac{bits}{symbol}$$

9- Joint entropy and conditional entropy:

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

$$H(X,Y) = H(XY) = -\sum_{j=1}^{m}\sum_{i=1}^{n} P\left(x_i, y_j\right)\log_2 P\left(x_i, y_j\right) \qquad bits/symbol$$

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

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

Return to first equation, we have: $P(x_i, y_j) = P(x_i)P(y_j \mid x_i)$, put inside log term

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

But

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

Put it in above equation yields:

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

So that $H(X,Y) = H(X) + H(Y \mid X)$

Example 8:

The joint probability of a system is given by:

$$P(X,Y) = \begin{array}{c} x_1 \\ x_2 \\ x_3 \end{array}\begin{bmatrix} 0.5 & 0.25 \\ 0 & 0.125 \\ 0.0625 & 0.0625 \end{bmatrix}$$

Find:

1- Marginal entropies.

2- Joint entropy

3- Conditional entropies.

4- The mutual information between x_1 and y_2.

5- The transinformation.

6- Draw the channel model.

1- $P(X) = \begin{bmatrix} x_1 & x_2 & x_3 \\ 0.75 & 0.125 & 0.125 \end{bmatrix} \quad P(Y) = \begin{bmatrix} y_1 & y_2 \\ 0.5625 & 0.4375 \end{bmatrix}$

$$H(X) = -[0.75\ln(0.75) + 2 \times 0.125\ln(0.125)]/ln2$$

$$= 1.06127\ bits/symbol$$

$$H(Y) = -[0.5625\ln(0.5625) + 0.4375\ln(0.4375)]/ln2$$

$$= 0.9887\ bits/symbol$$

2-

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

$$H(X,Y) = -\frac{[0.5\ln(0.5) + 0.25\ln(0.25) + 0.125\ln(0.125) + 2 \times 0.0625\ln(0.0625)]}{ln2}$$

$$= 1.875 \qquad bits/symbol$$

3- $H(Y \mid X) = H(X,Y) - H(X) = 1.875 - 1.06127 = 0.813\ \frac{bits}{symbol}$

$$H(X \mid Y) = H(X,Y) - H(Y) = 1.875 - 0.9887 = 0.886\ bits/symbol$$

4- $I(x_1, y_2) = \log_2\left(\frac{P(x_1|y_2)}{P(x_1)}\right)$, $but\ P(x_1 \mid y_2) = P(x_1, y_2)/P(y_2)$

$$I(x_1, y_2) = \log_2\left(\frac{P(x_1,y_2)}{P(x_1)P(y_2)}\right) = \log_2\frac{0.25}{0.75 \times 0.4375} = -0.3923 \quad bits$$

That means y_2 gives ambiguity about x_1

5- $I(X,Y) = H(X) - H(X \mid Y) = 1.06127 - 0.8863 = 0.17497\ bits/symbol.$

6- To draw the channel model, must find P(Y|X) matrix from P(X, Y) matrix by dividing its rows by the corresponding P(x_i):

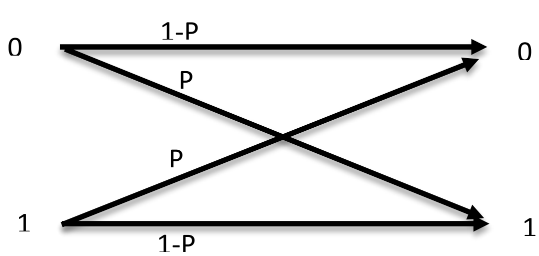
$$P(X \mid Y) = \begin{array}{c} x_1 \\ x_2 \\ x_3 \end{array}\begin{bmatrix} 0.5/0.75 & 0.25/0.75 \\ 0/0.125 & 0.125/0.125 \\ 0.0625/0.125 & 0.0625/0.125 \end{bmatrix} = \begin{array}{c} x_1 \\ x_2 \\ x_3 \end{array}\begin{bmatrix} 2/3 & 1/3 \\ 0 & 1 \\ 0.5 & 0.5 \end{bmatrix}$$

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.

B**inary 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 abit (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:

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

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

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

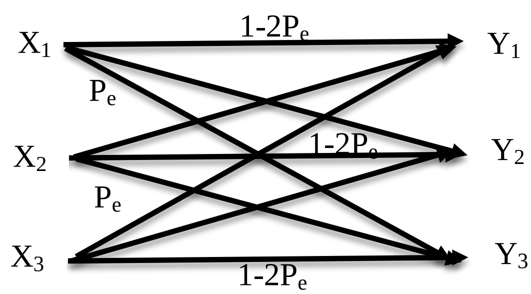
$$\Pr(Y = 1 \mid X = 1) = 1 - P$$

Ternary symmetric channel (TSC):

The transitional probability of TSC is:

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

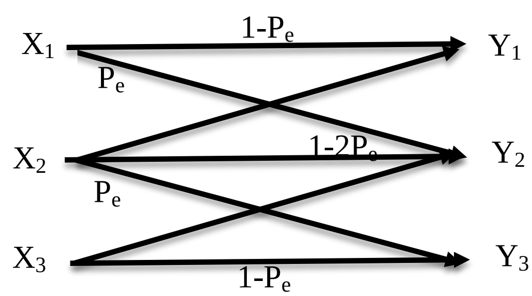
The TSC is symmetric but not very practical since practically x_1 and x_3 are not affected so much as x_2. In fact the interference between x_1 and x_3 is much less than the interference between x_1 and x_2 or x_2 and x_3.



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

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

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.



Special Channels:

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

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

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

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

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

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

3- Noiseless 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:

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

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

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:

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

1- Channel capacity of Symmetric channels:

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.

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

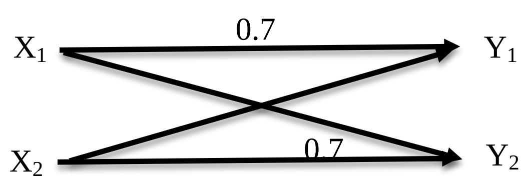
$$I(X,Y) = log_2 m + K$$

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

$C = log_2 m + K \quad$ 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 $\eta = \frac{0.0896}{0.1187} = 75.6\%$

2- 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:

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

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

a- Let $P(x_1) = P, \quad then\ P(x_2) = 1 - P$, hence instead of having two variable, we will have only one variable P.

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

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

From above results $P(Y) = [0.1 + 0.6P \quad 0.9 - 0.6P]$

We have

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

$$H(Y \mid X) = -[0.7Pln0.7 + 0.3Pln0.3 + 0.1(1-P)ln0.1 + 0.9(1-P)ln0.9]/ln2$$

$$= -[0.7Pln0.7 + 0.3Pln0.3 + 0.1ln0.1 - 0.1Pln0.1 + 0.9ln0.9 - 0.9Pln0.9]ln2$$

b- $\frac{\partial H(Y|X)}{\partial P} = \frac{dH(Y|X)}{dP} = -[0.7ln0.7 + 0.3ln0.3 - 0.1ln0.1 - 0.9ln0.9]/ln2$

$$\frac{dH(Y \mid X)}{dP} = -[-0.2858781]/ln2$$

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

Then $H(Y) = -[(0.1 + 0.6P)ln(0.1 + 0.6P) + (0.9 - 0.6P)ln(0.9 - 0.6P)]/ln2$

$$\frac{dH(Y)}{dP} = -[0.6ln(0.1+0.6P) + 0.6 - 0.6ln(0.9-0.6P) - 0.6]/ln2$$

$$\therefore \ \frac{dH(Y)}{dP} = -\frac{[0.6\{ln(0.1+0.6P) - ln(0.9-0.6P)\}]}{ln2} = -\frac{\left[0.6\ln\left(\frac{0.1+0.6P}{0.9-0.6P}\right)\right]}{ln2}$$

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

So that $\frac{I(Y,X)}{dP} = \frac{dH(Y)}{dP} - \frac{dH(Y|X)}{dP} = 0$

$$-\frac{\left[0.6\ln\left(\frac{0.1+0.6P}{0.9-0.6P}\right)\right]}{ln2} + \frac{0.285781}{ln2} = 0$$

$$-\left[0.6\ln\left(\frac{0.1+0.6P}{0.9-0.6P}\right)\right] = -0.285781$$

$$\ln\left(\frac{0.1+0.6P}{0.9-0.6P}\right) = 0.47630167$$

$$\left(\frac{0.1+0.6P}{0.9-0.6P}\right) = e^{0.47630167} = 1.6101087$$

$$0.1 + 0.6P = 1.6101087(0.9 - 0.6P)$$

$$\therefore \quad P \cong 0.862 \qquad put\ in\ H(Y) equation\ to\ get$$

$$H(Y) = -[(0.1 + 0.6 \times 0.862)ln(0.1 + 0.6 \times 0.862) + (0.9 - 0.6 \times 0.862)ln(0.9 - 0.6 \times 0.862)]/ln2$$

$$\therefore H(Y) = 0.96021 \quad bits/symbol$$

Similarly we can substitute in H(Y| X) equation to get

$$\mathrm{H}(\mathrm{Y} \mid \mathrm{X}) = 0.66354 \quad \mathrm{bits/symbol}$$

d-

$$C = \max[I(X,Y)] = 0.96021 - 0.66354 = 0.29666\ bits/symbol.$$

Shannon's theorem:

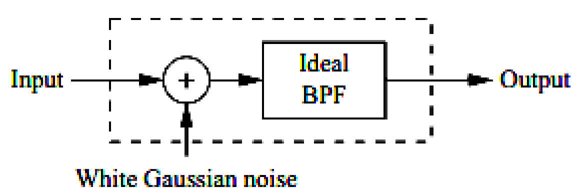
1- A given communication system has a maximum rate of information C known as the channel capacity.

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 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.

Nyquist Rate:

When a continuous function, x(t), is sampled at a constant rate, fs (samples/second), there is always an unlimited number of other continuous functions that fit the same set of samples. But only one of them is band-limited to ½ fs (hertz), which means that its Fourier transform, X(f), is 0 for all |f| ≥ ½ fs, which is called the Nyquist criterion. In terms of a function's own bandwidth (B), the Nyquist criterion is often stated as fs > 2B. And 2B is called the Nyquist rate for functions with bandwidth B. When the Nyquist criterion is not met (B > ½ fs), a condition called aliasing occurs, which results in some inevitable differences between x(t) and a reconstructed function that has less bandwidth.

Hartley's law

If the amplitude of the transmitted signal is restricted to the range of [−*A* ... +*A*] volts, and the precision of the receiver is ±Δ*V* volts, then the maximum number of distinct pulses *M* is given by

$$M = 1 + \frac{A}{\Delta V}.$$

Hartley constructed a measure of the line rate *R* as:

$$R = f_p \log_2(M),$$

Where f_p is the pulse rate, also known as the symbol rate, in symbols/second