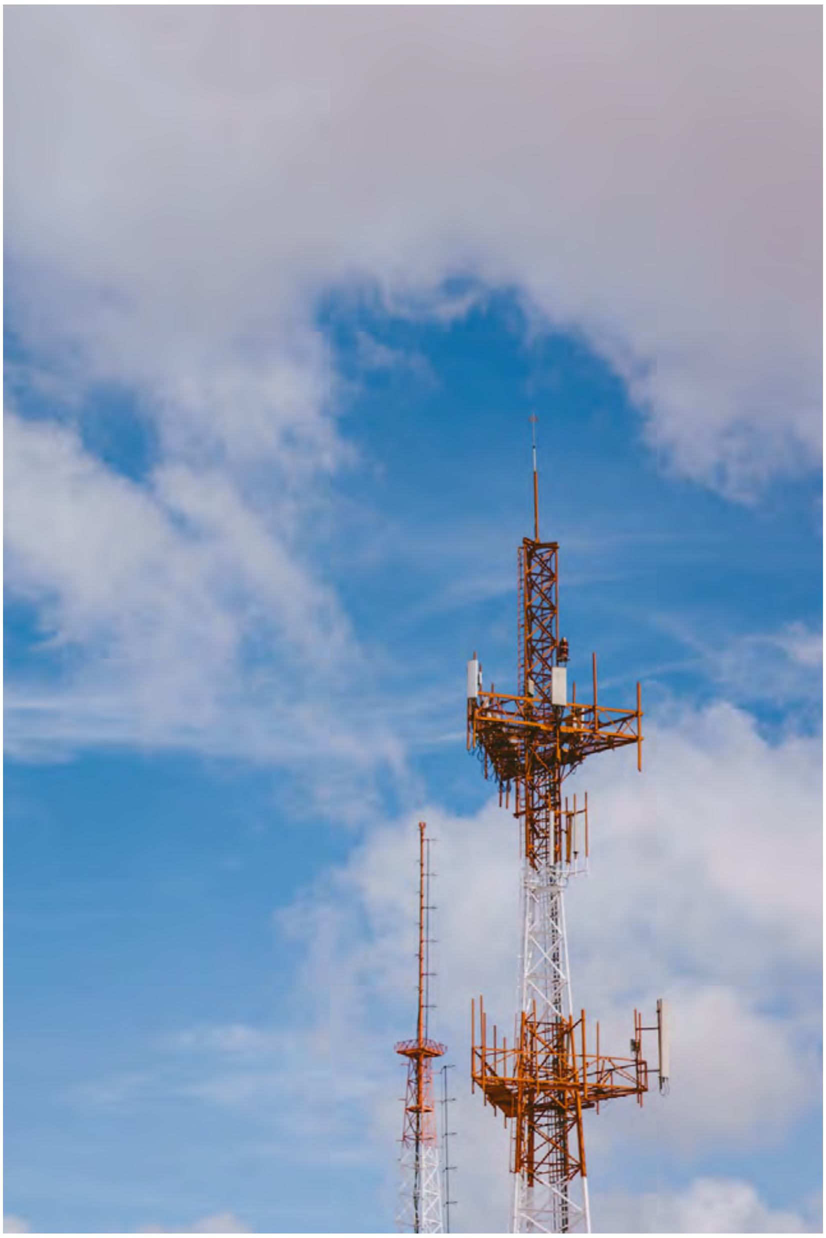


Communication Engineering

INFORMATION THEORY

YOU MAY GET STUDY MATERIAL FROM
AMIESTUDYCIRCLE.COM

INFO@AMIESTUDYCIRCLE.COM

WHATSAPP/CALL: 9412903929

Information Theory

INFORMATION CONTENT OF SYMBOL (I.E. LOGARITHMIC MEASURE OF INFORMATION)

Let us consider a discrete memoryless source (DMS) denoted by X and having alphabet $\{x_1, x_2, \ldots\ldots\ldots x_m\}$. The information content of a symbol x_i denoted by $I(x_i)$ is defined by

$$I(x_i) = \log_b \frac{1}{P(x_i)} = -\log_b P(x_i)$$

Where $P(x_i)$ is the probability of occurrence of symbol x_i.

The unit of $I(x_i)$ is the bit (binary unit) if b = 2, Hartley or decit if b = 10, and nat (natural unit) if b = e. It is standard to use b = 2. Here the unit bit (abbreviated "b") is a measure of information content and is not to be confused with the term `bit' meaning "binary digit". The conversion of these units to other units can be achieved by the following relationships.

$$\log_2 a = \frac{\text{In}a}{\text{In}2} = \frac{\log a}{\log 2}$$

Example

A source produces one of four possible symbols during each interval having probabilities $P(x_1) = \frac{1}{2}$, $P(x_1) = \frac{1}{4}$, *$P(x_3)$ = $P(x_4)$ =* $\frac{1}{8}$. *Obtain the information content of each of these symbols.*

Solution

We know that the information content of each symbol is given as

$$I(x_i) = \log_2 \frac{1}{P(x_i)}$$

Thus, we can write

$$I(x_i) = \log_2 \frac{1}{1/2} = \log_2(2) = 1 \text{ bit}$$

$$I(x_2) = \log_2 \frac{1}{1/4} = \log_2 2^2 = 2 \text{ bits}$$

$$I(x_3) = \log_2 \frac{1}{1/8} = \log_2 2^3 = 3 \text{ bits}$$

$$I(x_4) = \log_2 \frac{1}{1/8} = 3 \text{ bits}$$

Answer

Example

Calculate the amount of information if it is given that $P(x_i) = \frac{1}{4}$.

Solution

We know that amount of information given as

$$I(x_i) = \log_2 \frac{1}{P(x_i)} = \frac{\log_{10} \frac{1}{P(x_i)}}{\log_{10} 2}$$

Substituting given value of $P(x_i)$ in above equation, we get

or $$I(x_i) = \frac{\log_{10} 4}{\log_{10} 2} = 2 \text{ bits}$$ **Answer**

Example

If there are M equally likely and independent symbols, then prove that amount of information carried by each symbol will be,

$$I(x_i) = N \text{ bits}$$

Where $M = 2^N$ and N is an integer.

Solution

Since, it is given that all the M symbols are equally likely and independent, therefore the probability of occurrence of each symbol must be $\frac{1}{M}$.

We know that amount of information is given as,

$$I(x_i) = \log_2 \frac{1}{P(x_i)} \quad \text{(i)}$$

Here, probability of each message is, $P(x_i) = \frac{1}{M}$.

Hence, equation (i) will be

$$I(x_i) = \log_2 M \quad \text{(ii)}$$

Further, we know that $M = 2^N$, hence equation (ii) will be,

$$I(x_i) = \log_2 2^N = N \log_2 2$$

or $$I(x_i) = N \frac{\log_{10} 2}{\log_{10} 2} = N \text{ bits}$$

Example

If $I(x_i)$ is the information carried by symbols x_1 and $I(x_2)$ is the information carried by message x_2, then prove that the amount of information carried compositely due to x_i and x_2 is $I(x_1,x_2) = I(x_1) + I(x_2)$.

Solution

We know that the amount of information is expressed as

$$I(x_i) = \log_2 \frac{1}{P(x_i)}$$

The individual amounts carried by symbols x_1 and x_2 are,

$$I(x_1) = \log_2 \frac{1}{(x_1)} \qquad \text{(i)}$$

and

$$I(x_2) = \log_2 \frac{1}{P(x_2)} \qquad \text{(ii)}$$

Here $P(x_1)$ is probability of symbol x_1 and $P(x_2)$ is probability of symbol x_2. Since message x_1 and x_2 are independent, the probability of composite message is $P(x_1)\ P(x_2)$. Therefore, information carried compositely due to symbols x_1 and x_2 will be,

$$I(x_1, x_2) = \log_2 \frac{1}{P(x_1)P(x_2)} = \log_2 \left[\left(\frac{1}{P(x_1)}\right)\left(\frac{1}{P(x_2)}\right)\right]$$

or

$$I(x_1, x_2) = \log_2 \left(\frac{1}{P(x_1)}\right) + \log_2 \left(\frac{1}{P(x_2)}\right) \qquad \{\text{Since } \log[AB] = \log A + \log B\}$$

Using equations (i) and (ii), we can write RHS of above equation as,

$$I(x_1,x_2) = I(x_1) + I(x_2)$$

Hence Proved

ENTROPY (i.e., AVERAGE INFORMATION)

In a practical communication system, we usually transmit long sequence of symbols from an information sources. Thus, we are more interested in the average information that a source produces than the information content of a single symbol.

In order to get the information content of the symbol, we take notice of the fact that the flow of information in a system can fluctuate widely because of randomness involved into the selection of the symbols. Thus, we require to talk about the average information content of the symbols in a long message.

Thus, for quantitative representation of average information per symbol we make the following assumptions:

(i) The source is stationary so that the probabilities may remain constant with time.

(ii) The successive symbols are statistically independent and come form the source at a average rate of r symbols per second.

The mean value of $I(x_i)$ over the alphabet of source X with m different symbols is given by

$$H(X) = E\left[I\left(x_i\right)\right] = \sum_{i=1}^{m} P\left(x_i\right) I\left(x_i\right)$$

$$= -\sum_{i=1}^{m} P\left(x_i\right) \log_2 P(x_i) \quad \text{b/symbol}$$

The quantity H(X) is called the **entropy** of source X. It is a measure of the *average information content per source symbol*. The source entropy H(X) can be considered as the average amount of uncertainty within source X that is resolved by use of the alphabet.

It may be noted that for a binary source X which generates independent symbols 0 and 1 with equal probability, the source entropy H(X) is

$$H\left(X\right) = -\frac{1}{2}\log_2 - \frac{1}{2}\log_2 = 1 \quad \text{b/symbol}$$

The source entropy H(X) satisfies the following relation:

$$0 \le H\left(X\right) \le \log_2 m$$

Where m is the size (number of symbols) of the alphabet of source X.

Problem

One of the five possible messages Q_1 to Q_5 having probabilities 1/2, 1/4, 1/8, 1/16, 1/16 respectively is transmitted. Calculate the average information

Answer: 1.875 bits/message

INFORMATION RATE

If the time rate at which source X emits symbols, is r (symbol s), the information rate R of the source is given by

$$R = rH(X) \ \text{b/s}$$

Here R is information rate.

H(X) is Entropy or average information

and r is rate at which symbols are generated.

Information rater is represented in average number of bits of information per second. It is calculated as under:

$$R = \left[r \, \text{in} \, \frac{\text{symbols}}{\text{second}}\right] \text{x} \left[H(X) \, \text{in} \, \frac{\text{information bits}}{\text{symbol}}\right]$$

or R = information bits/second

Example

A discrete Memoryless Source (DMS) X has four symbols x_1,x_2,x_3,x_4 with probabilities $P(x_1) = 0.4$, $P(x_2) = 0.3$, $P(x_3) = 0.2$, $P(x_4) = 0.1$.

(i) *Calculate H(X)*

(ii) *Find the amount of information contained in the message $x_1\ x_2\ x_1\ x_3$ and $x_4\ x_3\ x_3\ x_2$ and compare with the H(X) obtained in part (i).*

Solution

We have $H(X) = -\sum_{i=1}^{4} P(x_i)\log_2 P(x_i)$

Simplifying, we get

$$= 0.4 \log_2 0.4 - 0.3 \log_2 0.3 - 0.2 \log 0.2 - 0.1 \log_2 0.1$$

Solving, we get

$$= 1.85 \text{ b/symbol}$$ **Ans.**

(ii) $P(x_1\ x_2\ x_1\ x_3) = (0.4)(0.3)(0.4)(0.2) = 0.0096$

$I(x_1\ x_2\ x_1\ x_3) = -\log_2 0.0096 = 6.70$ b/symbol **Ans.**

Thus, $I(x_1\ x_2\ x_1\ x_2) < 7.4\ [= 4H(X)]$ b/symbol

$P(x_4\ x_3\ x_3\ x_2) = (0.1)(0.2)^2(0.3) = 0.0012$

$I(x_4\ x_3\ x_3\ x_2) = -\log_2 0.0012 = 9.70$ b/symbol

Thus, $I(x_4x_3x_3x_2) > 7.4\ [= 4H(X)]$ b/symbol **Ans.**

Example

Consider a binary memoryless source X with two symbols x_1 and x_2. Prove that H(X) is maximum when both x_1 and x_2 equiproable.

Solution

Let $P(x_1) = \alpha$ so that $P(x_2) = 1 - \alpha$

Thus, $H(X) = -\alpha \log_2 \alpha - (1-\alpha)\log_2(1-\alpha)$ (i)

or $$\frac{dH(X)}{d\alpha} = \frac{d}{d\alpha}[-\alpha\log_2\alpha - (1-\alpha)\log_2(1-\alpha)$$

Using the relation,

$$\frac{d}{dx}\log_b y = \frac{1}{y}\log_b e\frac{dy}{dx}$$

We get $$\frac{dH(X)}{d\alpha} = -\log_2 \alpha + \log_2 (1-\alpha) = \log_2 \frac{1-\alpha}{\alpha}$$

Note that, the maximum value of H(X) requires that

$$\frac{dH(X)}{d\alpha} = 0$$

This means that

$$\frac{1-\alpha}{\alpha} = 1$$

or $$\alpha = \frac{1}{2}$$

Note that H(X) = 0 when α = 0 or 1.

When $P(x_1) = P(x_2) = \frac{1}{2}$, H(X) is maximum and is given by

$$H(x) = \frac{1}{2}\log_2 2 + \frac{1}{2}\log_2 2 = 1 \text{ b/symbol} \qquad \text{(ii)}$$

Hence Proved

Example (AMIE Summer 2006, 10 marks)

Verify the following expression:

$$0 \leq h(X) \leq \log_2 m$$

Where m is the size of the alphabet of X.

Solution

Proof of the lower bound:

Since $0 \leq P(x_i) \leq 1$

$$\frac{1}{P(x_i)} \geq 1 \quad \text{and} \quad \log_2 \frac{1}{P(x_i)} \geq 0$$

The, it follows that

$$P(x_i)\log_2 \frac{1}{P(x_i)} \geq 0$$

Thus, $$H(X) = \sum_{i=1}^{m} P(x_i)\log_2 \frac{1}{P(x_i)} \geq 0 \qquad \text{(i)}$$

Next, we note that

$$P(x_i)\log_2\frac{1}{P(x_i)}=0$$

If and only if $P(x_i) = 0$ or 1. Since

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

When $P(x_i) = 1$ then $P(x_j) = 0$ for $j \neq i$. Thus, only in this case, $H(X) = 0$.

<u>Proof of the upper bound:</u>

Consider two probability distributions $[P(x_i) = P_i]$ and $[Q(x_i) = Q_i]$ on the alphabet $\{x_i\}$, i = 1,2,… m, such that

$$\sum_{i=1}^{m}P_i=1$$

and $$\sum_{i=1}^{m}Q_i=1 \qquad \text{(ii)}$$

We know that

$$\sum_{i=1}^{m}P_i\log_2\frac{Q_i}{P_i}=\frac{1}{\ln 2}\sum_{i=1}^{m}P_i\ln\frac{Q_i}{P_i}$$

Next using the inequality

$$\ln\alpha \le \alpha-1 \quad \alpha\ge 0 \qquad \text{(iii)}$$

and noting that the equality holds obly if $\alpha = 1$, we obtain

$$\sum_{i=1}^{m}P_i\ln\frac{Q_i}{P_i} \le \sum_{i=1}^{m}P_i\left(\frac{Q_i}{P_i}-1\right)=\sum_{i=1}^{m}(Q_i-P_i)=\sum_{i=1}^{m}Q_i-\sum_{i=1}^{m}P_i=0$$

by using Eq. (ii)(iv)

Thus, $$\sum_{i=1}^{m}P_i\log_2\frac{Q_i}{P_i} \le 0 \qquad \text{(v)}$$

Where the equality holds oly if $Q_i = P_i$ for all i.

Setting $$Q_i=\frac{1}{m}, \qquad i = 1,2,\ldots\ldots m \qquad \text{(vi)}$$

We get $$\sum_{i=1}^{m}P_i\log_2\frac{1}{P_i m}=-\sum_{i=1}^{m}P_i\log_2 P_i-\sum_{i=1}^{m}P_i\log_2 m$$

$$=H(X)-\log_2 m-\sum_{i=1}^{m}P_i \qquad \text{(vii)}$$

$$= H(X) - \log_2 m \le 0$$

Hence $H(X) \le \log_2 m$ and also the quality holds only if the symbols in X are equiprobable.

Example

A high-resolution black-and-white TV picture consists of about 2 x 10^6 picture elements and 16 different brightness levels. Pictures and repeated at the rate of 32 per second. All picture elements are assumed to be independent, and all levels have equal likelihood of occurrence. Calculate the average rate of information conveyed by this TV picture source.

Solution

We have

$$H(X) = -\sum_{i=1}^{16} \frac{1}{16} \log_2 \frac{1}{16} = 4 \text{ b/element}$$

$$r = 2\,(10^6)\,(32) = 64\,(10^6) \text{ elements/s}$$

Hence, we have

$$R = r\,H(X) = 64\,(10^0)\,(4) = 256\,(10^0) \text{ b/s} = 256 \text{ Mb/s}$$ **Ans.**

Example

Consider a telegraph source having two symbols, dot and dash. The dot duration is 0.2 s. The dash duration is 3 times the dot duration. The probability of the dot's occurring is twice that of the dash, and the time between symbols is 0.2s. Calculating the information rate of the telegraph source.

Solution

We have

$$P(dot) = 2P\,(dash)$$

$$P(dot + P(dash) = 3P(dash) = 1$$

Thus, $$P(dash) = \frac{1}{3}$$

and $$P(dot) = \frac{2}{3}$$

Further, we know that

$$H(X) = -\,P(dot)\log_2 P(dot) - P(dash)\log_2 P(dash)$$

$$= 0.667\,(0.585) + 0.333\,(1.585) = 0.92 \text{ b/symbol}$$

$$t_{dot} = 0.2s \; t_{dash} = 0.6s \; t_{space} = 0.2\text{ s}$$

Thus, the average time per symbol is

$$T_s = P(dot)\,t_{dot} + P(dash)\,t_{dash} + t_{space} = 0.5333 \text{ s/symbol}$$

and the average symbol rate is

$$r = \frac{1}{T_s} = 1.875 \text{ symbols/s}$$

Thus, the average information rate of the telegraph source will be

$$R = r\,H(X) = 1.875\,(0.92) = 1.725 \text{ b/s}$$ **Ans.**

Example

A discrete source emits one of five symbols once every millisecond with probabilities $1/2, 1/4, 1/8, 1/16$ and $1/16$ *respectively. Determine the source entropy and information rate.*

Solution

We know that the source entropy is given as

$$H(X) = \sum_{i=1}^{m} P(x_i)\log_2 \frac{1}{P(x_i)}$$

or $$H(X) = \sum_{i=1}^{5} P(x_i)\log_2 \frac{1}{P(x_i)} \text{ bits/symbol}$$

or $$H(X) = \frac{1}{2}\log_2(2) + \frac{1}{4}\log_2(4) + \frac{1}{8}\log_2(8) + \frac{1}{16}\log_2(16) + \frac{1}{16}\log_2(16)$$

or $$H(X) = \frac{1}{2} + \frac{1}{2} + \frac{3}{8} + \frac{1}{4} + \frac{1}{4} = \frac{15}{8}$$

or $$H(X) = 1.875 \text{ bits/symbol}$$

The symbol rate $r = \frac{1}{T_b} = \frac{1}{10^{-3}} = 1000$ symbols/sec.

Therefore, the information rate is expressed as

$$R = r\,H(X) = 1000 \times 1.875$$

$$R = 1875 \text{ bits/sec.}$$ **Ans.**

Example (AMIE Summer 2012, 5 marks)

The probabilities of the five possible outcomes of an experiment are given as

$$P(x_1) = \frac{1}{2},\quad P(x_2) = \frac{1}{4},\quad P(x_3) = \frac{1}{8},\quad P(x_4) = P(x_5) = \frac{1}{16}$$

Determine the entropy and information rate if there are 16 outcomes per second.

Solution

The entropy of the system is given as

$$H(X) = \sum_{i=1}^{5} P(x_i)\log_2 \frac{1}{P(x_i)} \quad \text{bits/symbol}$$

or $$H(X) = \frac{1}{2}\log_2(2) + \frac{1}{4}\log_2(4) + \frac{1}{8}\log_2(8) + \frac{1}{16}\log_2(16) + \frac{1}{16}\log_2(16)$$

or $$H(X) = \frac{1}{2} + \frac{2}{4} + \frac{3}{4} + \frac{4}{16} + \frac{4}{16} = \frac{1}{2} + \frac{1}{2} + \frac{3}{8} + \frac{1}{4} + \frac{1}{4}$$

or $$H(X) = \frac{4+4+3+2+2}{8}$$

or $$H(X) = \frac{15}{8} = 1.875 \text{ bits/outcome}$$ **Ans.**

Now, rate of outcomes r = 16 outcomes/sec.

Therefore, the rate of information R will be $R = r\,H(X) = 6 \times \frac{15}{8} = 30$ bits/sec. **Ans.**

Example

An analog signal band limited to 10 kHz is quantized in 8 levels of of a PCM system with probabilities of $1/4, 1/5, 1/5, 1/10, 1/10, 1/20, 1/20$ and $1/20$ *respectively. Find the entropy and the rate of information.*

Solution

We know that according to the sampling theorem, the signal must be sampled at a frequency given as

$$f_s = 10 \times 2 \text{ kHz} = 20 \text{ kHz}$$

Considering each of the eight quantized levels as a message, the entropy of the source may written as

$$H(X) = \frac{1}{4}\log_2 4 + \frac{1}{5}\log_2 5 + \frac{1}{5}\log_2 5 + \frac{1}{10}\log_2 10 + \frac{1}{10}\log_2 10 + \frac{1}{20}\log_2 20 + \frac{1}{20}\log_2 20 + \frac{1}{20}\log_2 20$$

$$H(X) = \frac{1}{4}\log_2 4 + \frac{2}{5}\log_2 5 + \frac{7}{20}\log_2 20$$

H(X) = 2.84 bits/message

As the sampling frequency is 20 kHz then the rate at which message is produced, will be given as

or r = 20000 messages/sec.

Hence the information rate is

$$R = r\,H(X) = 20000 \times 2.84$$

R = 56800 bits/sec. **Ans.**

Problem

An analog signal is band limited to B Hz, sampled at the Nyquist rate and samples are quantized into 4 levels. These four levels are assumed to be purely independent with probabilities of occurrences are

$$p_1 = 1/8, p_2 = 2/8, p_3 = 3/8, p_4 = 2/8$$

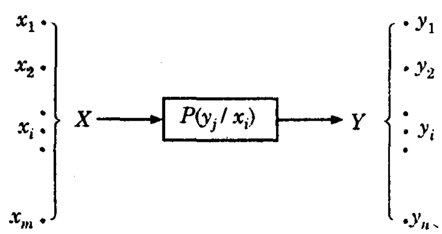
Find the information rate of the source.

Answer: 3.81 B bits/sec

THE DISCRETE MEMORYLESS CHANNELS (DMC)

Channels Representation

A communication channel may be defined as the path or medium through which the symbols flow to the receiver. A discrete memoryless channel (DMC) is a statistical model whit an input X and an output Y as shown in figure given below.



During each unit of the time (signaling interval), the channel accepts an input symbol from X, and in response it generates an output symbol from Y. The channel is "discrete" when the alphabets of X and Y are both finite. It is "memoryless" when the current output depends on only the current input depends on only the current input and not on any of the previous inputs.

A diagram of a DMC with m inputs and n outputs has been illustrated in figure. The input X consists of input symbols $x_1, x_2, \ldots x_m$. The a priori probabilities of these source symbols $P(x_i)$ are assumed to be known. The outputs Y consists of output symbols $y_1, y_2, \ldots y_n$. Each possible input-to-output path is indicated along with a conditional probability $P\left(y_j|x_i\right)$, where $P\left(y_j|x_i\right)$ is the conditional probability of obtaining output y_j given that the input is x_i and is called a **channel transition probability**.

The Channel Matrix

A channel is completely specified by the complete set of transition probabilities. Accordingly, the channel in figure above is often specified by the matrix of transition probabilities $\left[P(Y|X)\right]$, is given by

$$\left[P\left(Y|X\right)\right] = \begin{bmatrix} P\left(y_1|x_1\right) & P\left(y_2|x_1\right) & .... & P\left(y_n|x_1\right) \\ P\left(y_1|x_2\right) & P\left(y_2|x_2\right) & .... & P\left(y_n|x_2\right) \\ ...... & ...... & .... & ...... \\ P\left(y_1|x_m\right) & P\left(y_2|x_m\right) & .... & P\left(y_n|x_m\right) \end{bmatrix} \tag{1}$$

The matrix $[P(Y|X)]$ is called the channel matrix. Since each input to the channel results in some output, each row of the channel matrix must to unity, that is,

$$\sum_{j=1}^{n} P\left(y_1|x_i\right) = 1 \text{ for all i} \tag{2}$$

Now, if the input probabilities P(X) are represented by the row matrix, then we have

$$[P(X)] = [P(x_1)\ P(x_2) \ldots \ldots \ldots \ldots P(x_m)] \tag{3}$$

and the output probabilities P(Y) are represented by the row matrix

$$[P(Y)] = [P(y_1)\ P(y_2) \ldots \ldots \ldots \ldots P(y_n)] \tag{4}$$

$$[P(Y)] = [P(X)]\,[P\left(y_1|x_i\right)] \tag{5}$$

If P(X) is represented as a diagonal matrix, then we have

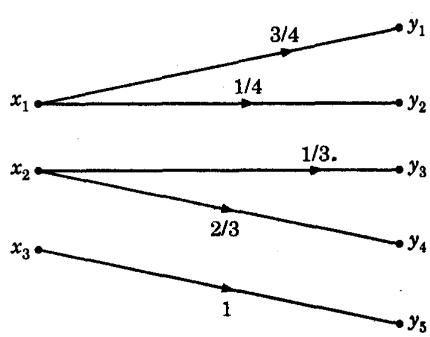
$$[P(X)]_d = \begin{bmatrix} P\left(x_1\right) & 0 & .... & 0 \\ 0 & P\left(x_2\right) & .... & .... \\ .... & .... & .... & .... \\ 0 & 0 & .... & P\left(x_m\right) \end{bmatrix} \tag{6}$$

then $$[P(X,Y)] = [P(X)]\,[P\left(Y|X\right)] \tag{7}$$

If P(X) is represented as a diagonal matrix, then we have

Where the (i,j) element of matrix [P(X,Y)] has the form $P(x_i,y_j)$. The matrix [P(X,Y)] is known as the joint probability matrix, and the element $P(x_i,y_j)$ is the joint probability of transmitting x_i and receiving y_j.

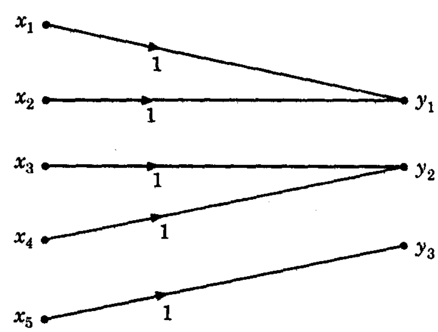
Lossless Channel



A channel describe by a channel matrix with only one non-zero element in each column is called a lossless channel. An example of a lossless channel has been shown in figure and the corresponding channel matrix is given in equations as under:

$$[P(Y|X)] = \begin{bmatrix} \frac{3}{4} & \frac{1}{4} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{3} & \frac{2}{3} & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (8)$$

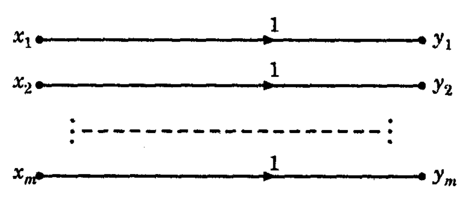
Deterministic Channel



A channel described by a channel matrix with only one non-zero element in each row is called a deterministic channel. An example of a deterministic channel has been shown in figure and the corresponding channel matrix is given by equation as under:

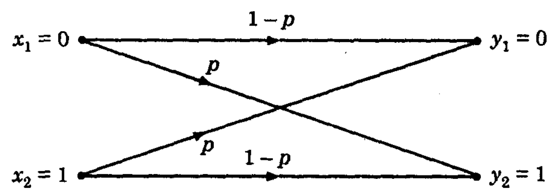
$$[P(Y|X) = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (9)$$

Noiseless Channel



A channel is called noiseless if it is both lossless and deterministic. A noiseless channel has been shown in figure. The channel matrix has only one element in each row and in each column, and this element is unity. Note that the input and output alphabets are of the dame size, that is, m = n for the noiseless channel.

Binary Symmetric Channel (BSC)



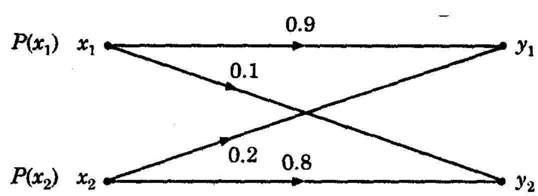
The binary symmetric channel (BSC) is defined by the channel diagram shown in figure and its channel matrix is given by

$$[P(Y|X)] = \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix} \tag{10}$$

The channel has two inputs ($x_1 = 0$, $x_2 = 1$) and two outputs ($y_1 = 0$, $y_2 = 1$). This channel is symmetric because the probability of receiving a 1 if a 0 is sent is the same as the probability of receiving a 0 if a 1 is sent. This common transition probability is denoted by P.

Example

Consider a binary channel shown in figure given below:



(i) *Find the channel matrix of the channel.*

(ii) *Find $P(y_i)$ and $P(y_2)$ when $P(x_1) = P(x_2) = 0.5$.*

(iii) *Find the joint probabilities $P(x_1,y_2)$ and $P(x_2,y_1)$ when $P(x_1) = P(x_2) = 0.5$*

Solution

(i) We known that the channel matrix is given by:

$$[P(Y|X)] = \begin{bmatrix} P(y_1|x_1) & P(y_2|x_1) \\ P(y_1|x_2) & P(y_2|x_2) \end{bmatrix} = \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix}$$

(ii) We know that

$$[P(Y)] = [P(X)]\ [P(Y|X)]$$

$$= [0.5 \quad 0.5] \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix}$$

$$= [0.55 \quad 0.45]$$

$$= [P(y_1) \ \ P(y_2)]$$

Hence, $P(y_1) = 0.55$ and $P(y_2) = 0.45$ Ans.

(iii) Again, we have

$$[P(X,Y)] = [P(X)]_d\,[P(Y|X)]$$

$$= \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix} \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix} = \begin{bmatrix} 0.45 & 0.05 \\ 0.1 & 0.4 \end{bmatrix}$$

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

Hence, $P(x_1,y_2) = 0.05$ and $P(x_2,y_1) = 0.1$ Ans.

THE CONDITIONAL AND JOINT ENTROPIES

Using the input probabilities $P(x_i)$, output probabilities $P(y_i)$, transition probabilities $P(y_j|x_i)$, and joint probabilities $P(x_i,y_j)$, we can define the following various entropy functions for a channel with m inputs and n outputs:

$$H(X) = -\sum_{i=1}^{m} P(xi)\log_2 P(xi) \tag{11}$$

$$H(Y) = -\sum_{j=1}^{n} P(y_j)\log_2 P(y_j) \tag{12}$$

$$H(X|Y) = -\sum_{j=1}^{n}\sum_{i=1}^{m} P(x_i, y_j)\log_2 P(x_i \mid y_j) \tag{13}$$

$$H(Y|X) = -\sum_{j=1}^{n}\sum_{i=1}^{m} P(x_i, y_j)\log_2 P(y_j \mid x_i) \tag{14}$$

$$H(X,Y) = -\sum_{j=1}^{n}\sum_{i=1}^{m} P(x_i, y_j)\log_2 P(x_i, y_j) \tag{15}$$

These entropies can be interpreted as under:

H(X) is the average uncertainty of the channel input, and H(Y) is the average uncertainty of the channel output.

The conditional entropy H(X|Y) is a measure of the average uncertainty remaining about the channel input after the channel output has been observed.

Also, H(Y|X) is sometimes called the *equivocation* of X with respect to Y.

The conditional entropy H(X|Y) is the average uncertainty of the channel output given that X was transmitted. The joint entropy H(X,Y) is the average uncertainty of the communication channel as a whole.

Two useful relationships among the above various entropies are as under:

$$H(X,Y) = H(X|Y) + H(Y) \tag{16}$$

$$H(X,Y) = H(Y|X) + H(X) \tag{17}$$

THE MUTUAL INFORMATION

The mutual information denoted by I(X;Y) of a channel is defined by

$$I(X ; Y) = H(X|Y) - H(Y) \quad \text{b/symbol} \tag{18}$$

Since H(X) represents the uncertainty about the channel input before the channel output is observed and H(X|Y) represents the uncertainty about the channel input after the channel output is observed, the mutual information I(X ; Y) represents the certainty about the channel input that is resolved by observing the channel output.

Properties of (X;Y):

$$I(X;Y) = I(Y;X) \tag{19}$$

$$I(X;Y) \geq 0 \tag{20}$$

$$I(X;Y) = H(Y) - H(Y|X) \tag{21}$$

$$I(X;Y) = H(X) + H(Y) - H(X,Y) \tag{22}$$

Example

Consider a noiseless channel with m input symbols and m output symbols. Prove that

$$H(X) = H(Y) \qquad .(i)$$

and

$$H(Y|X) = 0 \qquad (ii)$$

Solution

For a noiseless channel, the transition probabilities are given by

$$P(y_j | x_i) = \begin{cases} 1 & \text{for } i = j \\ 0 & \text{for } i \neq j \end{cases} \qquad \text{(iii)}$$

Hence,

$$P(x_i, y_j) = P(y_j | x_i) P(x_i)$$

$$= \begin{cases} P(x_i) & \text{for } i = j \\ 0 & \text{for } i \neq j \end{cases} \qquad \text{(iv)}$$

and

$$P(y_j) = \sum_{i=1}^{m} P(x_i, y_j) = P(x_j) \qquad \text{.....(v)}$$

Thus, using equation (iv) and (v) we have

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

Further, we know that

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

$$= -\sum_{i=1}^{m} P(x_i) \sum_{j=1}^{m} \log_2 P(y_j | x_i)$$

$$= -\sum_{i=1}^{m} P(x_i) \log_2 1 = 0$$ **Hence Proved**

Example

Verify the following expression:

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

Solution

We know that

$$P(x_i,y_j) = P(x_i|y_j)\, P(y_j)$$

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

Also we have

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

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

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

$$-\sum_{j=1}^{n} \left[\sum_{i=1}^{m} P(x_i, y_i)\right] \log P(y_j)$$

$$H(X,Y) = H(X|Y) - \sum_{j=1}^{n} P(y_j) \log P(y_i)$$

$$H(X,Y) = H(X|Y) + H(Y)$$ **Hence proved**

Problem

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

Example (AMIE S13, W13, 6 marks)

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

Solution

Mutual information

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

$$= \sum_{j=1}^{n}\sum_{i=1}^{m} r_{ij} \log \frac{r_{ij}}{p_i q_j}$$

where $r_{ij} = P(X = a_i, y = b_i)$

Because of symmetry in I(X,Y) expression

$$I(X,Y) = I(Y,X)$$

i.e. $H(X) - H(X|Y) = H(Y) - H(Y|X)$

Rearranging $H(Y) + H(X|Y) = H(X) + H(Y|X)$

Example (AMIE W94)

Verify the following expression

$$I(X;Y) \geq 0$$

Solution

We have

$$-I(X;Y) = \sum_{i=1}^{m}\sum_{j=1}^{n} P(x_i, y_j) \log_2 \frac{P(x_i)}{P(x_i | y_j)} \quad \text{(i)}$$

Using Bayes' rule, we have

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

We can write equation (i) as under:

$$-I(X;Y) = \frac{1}{\ln 2}\sum_{i=1}^{m}\sum_{j=1}^{n} P(x_i, y_j) \ln \frac{P(x_i)P(y_j)}{P(x_i, y_j)} \quad \text{(ii)}$$

Also, we know that

$$\ln \alpha \leq \alpha - 1$$

Therefore, we have

$$-I(X;Y) \leq \frac{1}{\ln 2}\sum_{i=1}^{m}\sum_{j=1}^{n} P(x_i, y_j)\left[\frac{P(x_i)P(y_j)}{P(x_i, y_j)} - 1\right]$$

or $$-I(X;Y) \leq \frac{1}{\ln 2}\sum_{i=1}^{m}\sum_{j=1}^{n} P(x_i)P(y_j) - \sum_{i=1}^{m}\sum_{j=1}^{n} P(x_i, y_j) \quad \text{(iii)}$$

Since $$\sum_{i=1}^{m}\sum_{j=1}^{n}P(x_i)P(y_j)=\sum_{i=1}^{m}P(x_i)\sum_{j=1}^{n}P(y_j)=(1)(1)=1$$

$$\sum_{i=1}^{m}\sum_{j=1}^{n}P(x_i,y_j)=\sum_{i=1}^{m}\left[\sum_{j=1}^{n}P(x_i,y_j)\right]=\sum_{i=1}^{m}P(x_i)=1$$

Equation (iii) reduces to

$$-I(X;Y) \le 0$$

$$I(X;Y) \ge 0.$$ **Hence proved.**

Example (AMIE Winter 2010, 14 marks)

Consider a BSC with $P(x_1) = \alpha$.

(i) Show that the mutual information I(X;Y) is given by

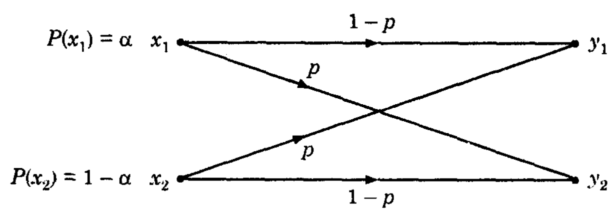
$I(X;Y) = H(Y) + p \log_2 p + (1 - p) \log_2(1 - p)$

(ii) Calculate I(X;Y) for $\alpha = 0.5$ and $p = 0.1$.

(iii) Repeat part (ii) for $\alpha = 0.5$ and $p = 0.5$, and comment on the result.

(iv) Find channel capacity of BSC.

Given figure shows the diagram of the BSC with associated input probabilities.



Solution

(i) We know that

$$[P(X,Y)]\begin{bmatrix}\alpha & 0\\ 0 & 1-\alpha\end{bmatrix}\begin{bmatrix}1-P & P\\ P & 1-P\end{bmatrix}$$

$$[P(X,Y)]=\begin{bmatrix}\alpha(1-p) & \alpha p\\ (1-p)p & (1-\alpha)(1-p)\end{bmatrix}$$

or $$[P(X,Y)]=\begin{bmatrix}P(x_1,y_1) & P(x_1,y_2)\\ P(x_2,y_1) & P(x_2,y_2)\end{bmatrix}$$

Also, we have

$$H(X|Y) = - P(x_1, y_1) \log_2 P(y_1|x_1)$$

$$= - P(x_1, y_2) \log_2 P(y_2|x_2)$$

$- (x_2, y_1) \log_2 P(y_1|x_2) - P(x_2, y_2) \log_2 P(y_2|x_2)$

$= - \alpha(1- \alpha)\ p \log_2 p - (1 - \alpha)(1 - p) \log_2(1 - p)$

$- p \log_2 p - (1 - p) \log_2(1 - p)$

We know that

$I(X;Y) = H(Y) - H(Y|X)$

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

(ii) When $\alpha = 0.5$ and $p = 0.1$, then, we have

$$[P(Y)] = [0.5 \quad 0.5]\begin{bmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{bmatrix} = [0.5 \quad 0.5]$$

Thus, using

$P(y_1) = P(y_2) = 0.5$ we have

$H(Y) = -P(y_1) \log_2 P(y_1) - P(y_2) \log_2 P(y_2)$

$= - 0.5 \log_2 0.5 - 0.5 \log_2 0.5 = 1$

$p \log_2 p + (1 - p) \log_2(1 - p) = 0.1 \log_2 0.1 + 0.9 \log_2 0.9$

$= 0.469$

Thus, $I(X;Y) = 0.469 = 0.531$ **Ans.**

(iii) When $\alpha = 0.5$ and $p = 0.5$, we have

$$[P(Y)] = [0.5 \quad 0.5]\begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix} = [0.5 \quad 0.5]$$

$H(Y) = 1$

$p \log_2 p + (1 - p) \log_2(1 - p) \log_2(1 - p) = 0.5 \log_2 0.5 + 0.5 \log_2 0.5$

$= -1$

Thus, $I(X;Y) = 1 - 1 = 0$

(iv) In part (i), we have proved that

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

which is maximum when H(Y) is maximum. Since, the channel output is binary, H(Y) is maximum when each output has a probability of 0.5 and is achieved for equally likely inputs.

For this case, H(Y) = 1 and the channel capacity is

$$C_s = 1 + p \log_2 p + (1 - p) \log_2 (1 - p)$$

Example (AMIE W96, Summer 2013, 10 marks)

A transmitter has an alphabet of four letters (y_1, y_2, y_3, y_4) and receiver has an alphabet of three letters {z_1, z_2, z_3, z_4}. The joint probability matrix is

$$\begin{array}{cc} & \begin{bmatrix} z_1 & z_2 & z_3 \end{bmatrix} \\ P(y,z) = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} & \begin{bmatrix} 0.30 & 0.05 & 0 \\ 0 & 0.25 & 0 \\ 0 & 0.15 & 0.05 \\ 0 & 0.05 & 0.15 \end{bmatrix} \end{array}$$

Calculate the five entropies H(y), H(z), H(y,z) H(y/z) and H(z/y).

Solution

$$P(y_i) = \sum_j P(y_i, z_j)$$

$$P\left(\frac{y_i}{z_i}\right) = \frac{P(y_i, z_i)}{P(z_j)} \text{ and } z_i = \sum_i P(y_i, z_j)$$

$\therefore$ $P(y_1) = P(y_1, z_1) + P(y_1, z_2) + P(y_1, z_3)$

$= 0.3 + 0.5 + 0 = 0.35$

Similarly

$P(y_2) = 0.25$

$P(y_3) = 0.20$

$P(y_4) = 0.20$

$P(z_1) = 0.3$

$P(z_2) = 0.5$

$P(z_3) = 0.2$

$$P\left(\frac{y_1}{z_1}\right) = \frac{P(y_1 z_1)}{P(z_1)} = \frac{0.3}{0.3} = 1$$

Similarly

$P(y_1/z_2) = 0.1$

$P(y_1/z_3) = 0$

$P(y_2/z_1) = 0$

$P(y_2/z_2) = 0.5$

$P(y_2/z_3) = 0$

$P(y_3/z_1) = 0$

$P(y_3/z_2) = 0.3$

$P(y_3/z_3) = 0.25$

$P(y_4/z_1) = 0$

$P(y_4/z_2) = 0.1$

$P(y_4/z_3) = 0.75$

$$H(y) = +\sum P(y_i)\log\left\{\frac{1}{P(y_i)}\right\} = -\sum P(y_i)\log P(y_i)$$

$= -(0.35 \log 0.35 + 0.25 \log 0.25 + 0.2 \log 0.2 + 0.2 \log 0.2)$

$= 1.958871848$ bits

$H(z) = -(0.3 \log 0.3 + 0.5 \log 0.5 + 0.2 \log 0.2)$

$= 1.485475297$ bits

$$H[y|z_1] = \sum = -\sum P[y_1|z_1]\log P[y|z_1]$$

$$= -\left\{\begin{matrix} P[y_1|z_1]\log P[y_1|z_1] + P[y_2|z_1]\log P[y_2|z_1] \\ +P[y_3|z_1]\log P[y_3|z_1] + P[y_4|z_1]\log P[y_4|z_1]\end{matrix}\right\}$$

$= -(1 \log 1 + 0 \log 0 + 0 \log 0 + 0 \log 0) = 0$

$$H[y|z_2] = -\left\{\begin{matrix} P[y_1|z_2]\log P[y_1|z_2] + P[y_2|z_2]\log P[y_2|z_2] \\ +P[y_3|z_2]\log P[y_3|z_2] + P[y_4|z_2]\log P[y_4|z_2]\end{matrix}\right\}$$

$= -(0.1 \log 0.1 + 0.5 \log 0.5 + 0) = 1.685475296$

Similarly

$$H[y|z_3] = 0.811278124$$

$H[y|z] = \Sigma P(z_j).H(y|z_i)$

$= P(z_1)H(y|z_1) + P(z_2)H(y|z_2) + P(z_3)H(y|z_3)$

$= 1.004993273$ bits

$I(y, z) = H(y) - H(z|y) = 1.48547297 - 0.953878575 = 0.531596722$ bits

$H(z|y) = H(z) - I(y, z) = 1.48547297 - 0.953878575 = 0.531596722$ bits

$I(y, z) = H(y) + H(z) + H(y, z)$

$H(y, z) = 1.958871848 + 1.485475297 - 0.953878575 = 2.49046857$ bits

Problem

Given $[P(x,y)] = \begin{bmatrix} 1/20 & 0 \\ 1/5 & 3/10 \\ 1/20 & 2/5 \end{bmatrix}$

$P(x_1) = 1/20$, $P(x_2) = ½$ and $P(x_3) = 9/20$

Find P(Y/X), H(X) and H(Y).

Answer: $\begin{bmatrix} 1 & 0 \\ 2/5 & 3/5 \\ 1/9 & 8/9 \end{bmatrix}$*; H(X) = 1.234 bits/channel; H(Y) = 0.881 bits/channel*

Channel Capacity per Symbol C_s

The channel capacity per symbol of a DMC is defined as

$$C_s = \max_{\{P(x_i)\}} I(X;Y) \qquad \text{b/symbol} \tag{23}$$

Channel Capacity per Second C

If r symbols are being transmitted per second, then the maximum rate of transmission of information per second is r C_s. This is the channel capacity per second and is denoted by C(b/s), i.e.,

$$C = r\,C_s \quad \text{b/s} \tag{24}$$

Channel Capacity of Lossless Channel

The channel capacity per symbol will be

$$C_s = \max_{\{P(x_i)\}} H(X) = \log_2 m \tag{25}$$

Where m is the number of symbols inX.

Channel capacity of Deterministic Channel

The channel capacity per symbol will be

$$C_s = \max_{\{P(x_i)\}} H(Y) = \log_2 n \tag{26}$$

Where n is the number of symbols in Y.

Channel capacity of noiseless channel

The channel capacity per symbol is

$$C_s = \log_2 m = \log_2 n \tag{27}$$

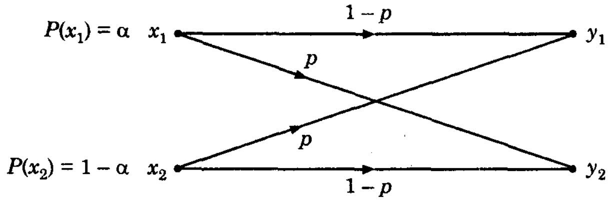
Channel Capacity Of Binary Symmetric Channel (BSC)

The channel capacity per symbol is

$$C_s = 1 + p \log_2 p + (1 - p) \log_2 (1 - p) \tag{28}$$

Example

Find the channel capacity of the binary erasure channel of figure given below



Solution

Let $P(x_1) = \alpha$. Then $P(x_2) = 1 - \alpha$

We have $$[P(Y|X)] = \begin{bmatrix} 1-p & p & 0 \\ 0 & p & 1-p \end{bmatrix} = \begin{bmatrix} P(y_1|x_1) & P(y_2|x_1) & P(y_3|x_1) \\ P(y_1|x_2) & P(y_2|x_2) & P(y_3|x_2) \end{bmatrix}$$

Using eq, (5), we have

$$[P(Y)] = \begin{bmatrix} \alpha & 1 & -\alpha \end{bmatrix} \begin{bmatrix} 1-p & p & 0 \\ 0 & p & 1-p \end{bmatrix}$$

$$= [\alpha\,(1 - p)\;p\;(1 - \alpha)\,(1 - p)]$$

$$= [P(y_1)\;P(y_2)\;P(y_3)]$$

By using eq.(7), we have

$$P(X,Y)] = \begin{bmatrix} \alpha & 0 \\ 0 & 1-\alpha \end{bmatrix} \begin{bmatrix} 1-p & p & 0 \\ 0 & p & 1-p \end{bmatrix}$$

or $$[P(X,Y)] = \begin{bmatrix} \alpha(1-p) & \alpha p & 0 \\ 0 & (1-\alpha)p & (1-\alpha)(1-p) \end{bmatrix}$$

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

In addition , from equation (12) and (14), we can calculate

$$H(Y) = -\sum_{j=1}^{3} P(y_j)\log_2 P(y_j)$$

$$= -\alpha\,(1-p)\log_2 \alpha(1-p) - p\log_2 p - (1-\alpha)(1-p)\log_2[(1-\alpha)(1-p)]$$

$$= (1-p)[-\alpha\log_2 \alpha - (1-\alpha)\log_2(1-\alpha)] - p\log_2 p - (1-p)\log_2\big(1-p)\big)$$

Also, we have $H(Y|X) = -\sum_{j=1}^{3}\sum_{i=1}^{2} P(x_i, y_j)\log_2 P(y_j|x_i)$

$$= -\alpha\,(1 - p)\log_2(1 - p) - \alpha p\log_2 p - (1 - \alpha)\,p\log_2 p - (1 - \alpha)(1 - p)\log_2(1 - p)$$

$$= -p\log_2 p - (1 - p)\log_2(1 - p)$$

Thus, by equations 21, we have

$$I(X;Y) = H(Y) - H(Y|X)$$

$$= (1 - p) [-\alpha \log_2 \alpha - (1 - \alpha) \log_2 (1 - \alpha)] = (1 - p) H(X)$$

And by equations (23), we have

$$C_s = \max_{\{p(X)\}} I(X;Y) = \max_{\{P(x_i)\}} (1-p)H(X)$$

$$= (1-p) \max_{\{P(x_i)\}} H(X) = 1-p \qquad \text{Ans.}$$

Problem

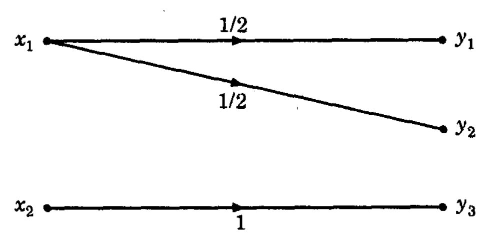
A channel capacity is described by the following channel matrix

$$\begin{bmatrix} 1/2 & 1/2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

(i) *Draw the channel diagram.*

(ii) *Find the channel capacity.*

Answer: (i) See figure. (ii) 1 b/symbol



DIFFERENTIAL ENTROPY

In a continuous channel, an information source produces a continuous signal x(t). The set of possible signals is considered as an ensemble of waveforms generated by some ergodic random process. It is further assumed that x(t) has a finite bandwidth so that x(t) is completely characterized by its periodic sample values. Thus, at any sampling instant, the collection of possible sample values constitutes a continuous random variable X described by it probability density function $f_X(x)$.

The average amount of information per sample value of x(t) is measured by

$$H(X) = -\int_{-\infty}^{\infty} f_X(x) \log_2 f_X(x)\, dx \qquad \text{b/sample} \qquad (29)$$

The entropy H(X) defined by equation (29) is known as the differential entropy of X.

The average mutual information in a continuous channel is defined (by analogy with the discrete case) as

$$I(X;Y) = H(X) - H(X|Y)$$

or $$I(X;Y) = H(Y) - H(Y|X)$$

Where $$H(Y) = -\int_{-\infty}^{\infty} f_Y(y)\log_2 f_Y(y)\,dy \qquad (30)$$

$$H(X\,|\,Y) = -\int_{-\infty}^{\infty} f_{XY}(x,y)\log_2 f_X(x\,|\,y)\,dx\,dy \qquad (31)$$

$$H(Y\,|\,X) = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\log_2 f_Y(y\,|\,x)\,dx\,dy \qquad (32)$$

Example

Find the differential entropy H(X) of the uniformly distributed random variable X with probability density function

$$f_X(x) = \begin{cases} \dfrac{1}{a} & \text{for} \quad 0 \le x \le a \\ 0 & \text{for otherwise} \end{cases}$$

for (i) a = 1, (ii) a = 2, and (iii) a = $\frac{1}{2}$

Solution

Using equation (29), we have

$$H(X) = -\int_{-\infty}^{\infty} f_X(x)\log_2 f_X(x)\,dx$$

$$= -\int_{-a}^{a} \frac{1}{a}\log_2 \frac{1}{a}\,dx = \log_2 a$$

(i) a = 1, H(X) = $\log_2 1 = 0$

(ii) a = 2, H(X) = $\log_2 2 = 1$

(iii) a = $\frac{1}{2}$, H(X) = $\log_2 \frac{1}{2}$ = - $\log_2 2$ = - 1

Note that the differential entropy H(X) is not an absolute measure of information .

Example

With the different entropy of a random variable X defined by equation

$$H(X) = -\int_{-\infty}^{\infty} f_X(x)\log_2 f_X(x)\,dx$$

Find the probability density function $f_X(x)$ for which H(X) is maximum.

Solution

We know that $f_X(x)$ must satisfy the following two conditions:

$$\int_{-\infty}^{\infty} f_X(x)\,dx = 1 \qquad \text{(i)}$$

$$\int_{-\infty}^{\infty} (x-\mu)^2 f_X(x)\,dx = \sigma^2 \qquad \text{(ii)}$$

Where μ is the mean of X and σ^2 it its variance. Since the problem is the maximization of H(X) under constraints of equations (i) and (ii), therefore, we use the method of Lagrange multipliers as under:

$$G\left[f_X(x),\lambda_1,\lambda_2\right] = H(X) + \lambda_1\left[\int_{-\infty}^{\infty} f_X(x)dx - 1\right] + \lambda_2\left[\int_{-\infty}^{\infty}\left(x-\mu\right)^2 f_X(x)dx - \sigma^2\right]$$

$$= \int_{-\infty}^{\infty}\left[-f_X(x)\log_2 f_X(x) + \lambda_1 f_X(x) + \lambda_2\left(x-\mu^2\right)f_X(x)\right]dx - \lambda_1 - \lambda_2\sigma^2 \quad \text{(iii)}$$

Where the parameters λ_1 and λ_2 are the Lagrange multipliers. then the maximization of H(X) requires that

$$\frac{\partial G}{\partial f_X(x)} = -\log_2 f_X(x) - \log_2 e + \lambda_1 + \lambda_2\left(x-\mu\right)^2 = 0 \qquad \text{(iv)}$$

Thus $\log_2 f_X(x) - \log_2 e + \lambda_1 + \lambda_2 (x - \mu)^2$

or $\ln f_X(x) = -1 + \dfrac{\lambda_1}{\log_2 e} + \dfrac{\lambda_2}{\log_2 e}\left(x-\mu\right)^2$

Hence, we obtain

$$f_X(x) = \exp\left[-1 + \frac{\lambda_1}{\log_2 e} + \frac{\lambda_2}{\log_2 e}\left(x-\mu\right)^2\right] \qquad \text{(v)}$$

In view of the constraints of equation (i) and (ii), it is required that $\lambda_2 < 0$. Let

$$\exp\left(-1 + \frac{\lambda_1}{\log_2 e}\right) = a$$

and $\left(\dfrac{\lambda_2}{\log_2 e}\right) = -b^2$

Then, equation (v) can be written as

$$f_X(x) = ae^{-b^2(x-\mu)^2} \qquad \text{(vi)}$$

Substituting equation (vi) into equation (i) and (ii), we get

$$a\int_{-\infty}^{\infty} e^{-b^2(x-\mu)^2}dx = a\frac{\sqrt{n}}{b} = 1 \qquad \text{(vii)}$$

$$a\int_{-\infty}^{\infty}(x-\mu)^2 e^{-b^2(x-\mu)^2}dx = a\frac{\sqrt{\pi}}{2b^3} = \sigma^2 \qquad \text{(viii)}$$

Solving equations (vii) (viii) for a and b^2, we get

$$a = \frac{1}{\sqrt{2\pi\sigma}} \text{ and } b^2 = \frac{1}{2\sigma^2}$$

Substituting these values in equation (vi), we observe that the desired $f_X(x)$ is given by

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-(x-\mu)^2/(2\sigma^2)}$$

Hence proved

Which is the probability density function of Gaussian random variable X of mean μ and variance σ^2.

ADDITIVE WHITE GAUSSIAN NOISE (AWGN) CHANNEL

Shannon-Hartley Law

The capacity C (b/s) of the AWGN channel is given by

$$C = 2BC_s = B\log_2\left(1+\frac{S}{N}\right) \text{ b/s}$$

where B is channel bandwidth which is fixed and S/N is signal to noise ratio.

Above equation is known as *Shannon-Hartley law*.

Proof

Let a, n and y represent the samples of n(t), A(t) and y(t). Assume that the channel is band limited to "B" Hz.

We know that

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

$$H(Y/X) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P(x,y)\log\frac{1}{P(y/x)}dxdy$$

$$= \int_{-\infty}^{\infty} P(x)dx\int_{-\infty}^{\infty} P(y/x)\log\frac{1}{P(y/x)}dy \qquad (1)$$

We know that $y = x + n$

For a given x, y = n + a constant (x)

If $P_N(n)$ represents the probability density function of "n", then $P(y/x) = P_n(y - x)$.

$$\therefore \qquad \int_{-\infty}^{\infty} P(y/x).\log\frac{1}{P(y/x)}dy = \int_{-\infty}^{\infty} P_n(y-x)\log\frac{1}{P_n(y/x)}dy \qquad (2)$$

Let $\quad y - x = z$

$\therefore \qquad dy = -dz$

Hence the above integral reduces to

$$\int_{-\infty}^{\infty} P_n(z).\log\frac{1}{P_n(z)}dz = H(n)$$

H(n) is nothing but the entropy of noise samples.

If the mean squared value of the signal n(t) is S and that of noise n(t) is N, the mean squared value of the output is given by

$$\overline{y^2} = S + N$$

Channel capacity

$$C = \max[I(X;Y)] = \max[H(Y) - H(Y/X)] = \max[H(Y) - H(N)]$$

H(y) is maximum when y is Gaussian and maximum value of H(Y) is $f_m.\log(2\pi e\sigma^2)$

Where f_m is the band limit frequency σ^2 is mean squares value.

$$\therefore \qquad H(y)_{\max} = B\log 2\pi\sigma e(S+N)$$

$$\max[H(n)] = B\log 2\pi e(N)$$

$$\therefore \qquad C = B\log[2\pi e(S+N) - B\log(2\pi eN)$$

$$= B\log\left[\frac{2\pi e(S+N)}{2\pi eN}\right]$$

$$\therefore \qquad C = B\log\left[1+\frac{S}{N}\right] \text{ bits/sec}$$

Example

Consider an AWGN channel with 4 KHz bandwidth and the noise power spectral density $\eta/2 = 10^{-12}$ W/Hz. The signal power required at the receiver is 0.1 mW. Calculate the capacity of this channel.

Solution

Given B = 4000 Hz; S = $0.1(10^{-3})$ W; N = ηB = $2(10^{-12})(4000) = 8(10^{-9})$ W

Thus $$\frac{S}{N} = \frac{0.1(10^{-3})}{8(10^{-9})} = 1.25(10^4)$$

Now $$C = B\log_2\left(1+\frac{S}{N}\right) = 4000\log_2[(1+1.25910^4)] = 54.44(10^3)\text{ b/s}$$

Example

Calculate the capacity of an AWGN channel with a bandwidth of 1 MHz and S/N ratio of 40 dB.

Solution

The capacity of AWGN channel is given by

$$C = \left[B \log_2\left(1+\frac{S}{N}\right)\right] b/s$$

where B is channel bandwidth and S/N is signal to noise ratio.

Given $\quad 10\log_{10}\left(\frac{S}{N}\right) = 40 \Rightarrow \frac{S}{N} = 10^4 \text{ and } B = 1MHz$

Then $\quad C = 1.\log_2(1+10^4) = 13.29\,Mb/\sec$

Example (AMIE Winter 2009, 10 marks)

Find the bandwidth of picture signal in TV if the following data is given where TV picture consists of 3 x 10^5 small picture element. There are 10 distinguishable brightness levels which equally likely to occur. Number of frames transmitted per second = 30 and SNR is equal to 30 dB.

Solution

$$\frac{S}{N} = 30dB$$

$$10\log_{10}\left(\frac{S}{N}\right) = 30$$

or $\quad \log_{10}\left(\frac{S}{N}\right) = 3$

$\therefore \quad \left(\frac{S}{N}\right) = 10^3 = 1000$

Information associated with each picture element

= $\log_2 10$ = (1/0.30103)$\log_{10}(10)$ = 3.32 bits/element

Information associated with each frame = 3.32 x 3 x 10^5

$\therefore$ Information transmitted per second

= 30 x 3.32 x 3 x 10^5

= max. rate of transmission of information = C

$\therefore \quad$ C = 3.32 x 30 x 3 x 10^5 = 29.9 x 10^6 bits/sec = 29.9 Mbps

$$C = B\log_2\left(1+\frac{S}{N}\right) \text{ bits/sec}$$

$$29.9 \times 10^6 = B \times (1/0.30103)\log_{10}(1001)$$

Solving W = 3 mHz

Problem (AMIE Winter 2010, 6 marks)

Calculate the capacity of an standard 4 kHz telephone channel working in the range of 300 to 3400 Hz with a signal to noise ratio of 32 dB.

Answer: 32.953 K bits/sec.

Problem

A Gaussian channel is band limited to 1 Mhz. If the signal power to noise spectral density $S/\eta = 10^5$ Hz, calculate the channel capacity C and the maximum information rate.

Answer: 0.1375 M bits/s, $R_{max} = 1.44(S/\eta) = 0.144$ M bits/s

Hint: $N = \eta B$

CHANNEL CAPACITY OF PCM EQUILISER

With maximum sampling rate of 2W samples per second, the maximum rate of information transmission of the PCM system measured in bits per second is

$$C = 2W\log_2 L\, bits/\text{sec}.$$

With M possible levels, we get $L = M^n$ (M^n distinct code words, each codeword contain n bits).

Now $$C = 2Wn\log_2 M\, bits/\text{sec}.$$

Average transmitted power required to maintain M-ary PCM system operate above error threshold is

$$S = \frac{2}{M}\left[\left(\frac{1}{2}\right)^2 + \left(\frac{3}{2}\right)^2 + ...\left(\frac{M-1}{2}\right)^2\right](\lambda\sigma)^2$$

$$= \lambda^2 N^2\left(\frac{M^2-1}{12}\right)$$

where λ is a constant and $N = N_0B$ = noise variance

Solving for M we get

$$M = \left(1+\frac{12S}{\lambda^2 N_0 B}\right)^{1/2}$$

$\therefore$ Substituting above equation for M in equation for C we get

$$C = Wn\log_2\left(1+\frac{12S}{\lambda^2 N_0 B}\right)$$

BW of pulse duration of 1/2nW is B = λnW

λ value varies between 1 and 2

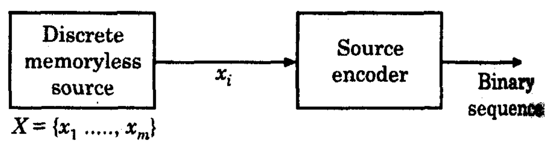
Using minimum value of λ (i.e. 1), we get

$$C = B\log_2\left(1+\frac{12S}{\lambda^2 N_0 B}\right)$$

or
$$C = B\log_2\left(1+\frac{12S}{\lambda^2 N}\right)$$

THE SOURCE CODING

A conversion of the output of a DMS into a sequence of binary symbols (i.e., binary code word) is called **source coding**. The device that performs this conversion s called the source encoder as shown in above figure.



An objective of source coding is to minimize the average bit rate required for representation of the source by reducing the redundancy of the information source.

The Code Length and Code Efficiency

Let X be a DMS with finite entropy H(X) and an alphabet $\{x_1, x_2, \ldots \ldots x_m\}$ with corresponding probabilities of occurrence $P(x_i)$ (i = 1, 2, … … m).

Let the binary code word assigned to symbol x_i by the encoder have length n_i, measured in bits. The length of a code word is the number of binary digits in the code word. The average code word length L, per source symbol is given by

$$L = \sum_{i=1}^{m} P(x_i)n_i$$

The parameter L represents the average number of bits per source symbol used in the source coding process.

Also, the code efficiency η is defined as

$$\eta = \frac{L_{min}}{L}$$

Where L_{min} is the minimum possible value of L. When η approaches unity, the code is said to be efficient.

The code redundancy γ is defined as

$$\gamma = 1 - \eta$$

Kraft Inequality

Let X be a DMS with alphabet $\{x_i\}$ (i = 1, 2, ……m). Assume that the length of the assigned binary code word corresponding to x_i is n_i.

A necessary and sufficient condition for the existence of an instantaneous binary code is

$$K = \sum_{i=1}^{m} 2^{-n_i} \leq 1$$

which is known as *Kraft inequality*.

Example

Consider a DMS X with two symbols x_1 and x_2 nad $P(x_1) = 0.9$, $P(x_2) = 0.1$. Symbols x_1 and x_2 are encoded as follows:

x_1	***$P(x_i)$***	***Code***
x_1	*0.9*	*0*
x_2	*0.1*	*1*

Find the efficiency η and the redundancy γ of this code.

Solution

We know that the average code length L per symbol is

$$L = \sum_{i=1}^{2} P(x_i) n_i = (0.9)(1) + (0.1)(1) = 1 b$$

Also we know

$$H(X) = -\sum_{i=1}^{2} P(x_i) \log_2 P(x_i)$$

$$= -0.9 \log_2 0.9 - 0.1 \log_2 0.1 = 0.469 \text{ b/symbol}$$

Also. the code efficiency η is

$$\eta = \frac{H(X)}{L} = 0.469 = 46.9\ \%$$

And, the code redundancy γ is given by

$$\gamma = 1 - \eta = 0.531 = 53.1\ \%$$ **Ans.**

Example

The second-order extension of the DMS X of question 8.25, denoted by X^2, is formed by taking the source symbols two at a time. The coding of this extension has been shown in given table. Find the efficiency η and the redundancy γ of this extension code.

a_i	$P(a_1)$	Code
$a_1 = x_1x_1$	0.81	0
$a_2 = x_1x_2$	0.09	10
$a_3 = x_2x_1$	0.09	110
$a_4 = x_2x_2$	0.01	111

Solution

We have $L = \sum_{i=1}^{4} P(a_i) n_i$

or $L = 0.81\,(1) + 0.09\,(2) + 0.09\,(3) + 0.01\,(3) = 1.29$ b/symbol

The entropy of the second-order extension of X, $H(X^2)$, is given by

$$H(X^2) = -\sum_{i=1}^{4} P(a_i) \log_2 P(a_i)$$

$$= -\,0.81 \log_2 0.81 - 0.09 \log_2 0.09 - 0.09 \log_2 0.09 - 0.01 \log_2 0.01$$

$H(X^2) = 0.938$ b/symbol

Therefore, the code efficiency η is

$$\eta = \frac{H(X^2)}{L} = \frac{0.938}{1.29} = 0.727 = 72.7\ \%$$

Also, the code redundancy γ will be

$\gamma = 1 - \eta = 0.273 = 27.3\ \%$ **Ans.**

Note that $H(X^2) = 2H(X)$

Example

Verify the following expression:

$$L \geq H(X)$$

Where L is the average code word length per symbol and H(X) is the source entropy.

Also give formula for code efficiency.

Solution

We know that

$$\sum_{i=1}^{m} P_i \log_2 \frac{Q_i}{P_i} \le 0$$

Where the equality holds only if $Q_i = P_i$

Let $$Q_i = \frac{2^{-n_i}}{K} \qquad \text{(i)}$$

Where $$K = \sum_{i=1}^{m} 2^{-n_i} \qquad \text{(ii)}$$

Now, we have

$$\sum_{i=1}^{m} Q_i = \frac{1}{K}\sum_{i=1}^{m} 2^{-n_i} = 1 \qquad \text{(iii)}$$

and $$\sum_{i=1}^{m} P_i \log_2 \frac{2^{-n_i}}{KP_i} = \sum_{i=1}^{m} P_i \left(\log_2 \frac{1}{P_i} - n_i - \log_2 K \right)$$

$$= -\sum_{i=1}^{m} P_i \log_2 P_i - \sum_{i=1}^{m} P_i n_i - \left(\log_2 K\right)\sum_{i=1}^{m} P_i \qquad \text{(iv)}$$

$$= H(X) - L - \log_2 K \le 0$$

From the Kraft inequality, we have

$$\log_2 K \le 0$$

Thus $$H(X) - L \le \log_2 K \le 0$$

or $$L \le H(X)$$

The equality holds when K = 1 and $P_i = Q_i$. **Hence Proved**

L can be made as close to H(X) as desired for some suitably chosen code.

Thus, with L_{min} = H(X), the code efficiency can be rewritten as

$$\eta = \frac{H(X)}{L}$$

ENTROPY CODING

The design of a variable length code such that its average code word length approaches the entropy of DMS is often referred to as entropy coding. In this section we present two examples of entropy coding.

Shannon-Fano Coding

An efficient code can be obtained by the following simple procedure, known as Shannon-Fano algorithm:

1. List the source symbols in order of decreasing probability.

2. Partition the set into two sets that are as close to equiprobables as possible, and assign 0 to the upper set 1 to the lower set.
3. Continue this process, ach time partitioning the sets with as nearly equal probabilities as possible until further partitioning is not possible.
4. An example of Shannon-Fano encoding is shown in Table given below. Note in Shannon-Fano encoding the ambiguity may arise in the choice of approximately equiprobable sets.

Shannon-Fano Encoding

x_i	$P(x_i)$	Step 1	Step 2	Step 3	Step 4	Code 5
x_1	0.30	0	0			00
x_2	0.25	0	1			01
x_3	0.20	1	0			10
x_4	0.12	1	1	0		110
x_4	0.08	1	1	1	0	1110
x_6	0.05	1	1	1	1	1111

H(X) = 2.36 b/symbol L = 2.38 b/symbol

$$\eta = \frac{H(X)}{L} = 0.99$$

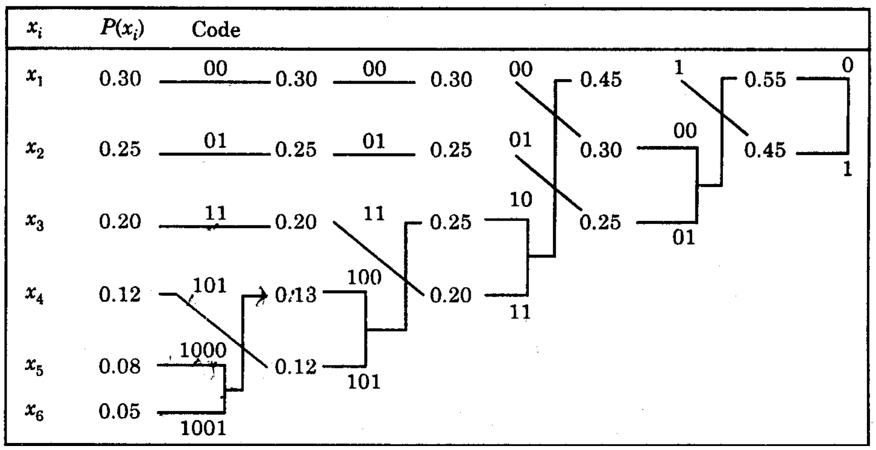
The Huffman Encoding

In general, Huffman encoding results in an optimum code. Thus, it is the code that has the highest efficiency. The Huffman encoding procedure is as follows:

1. List the source symbols in order of decreasing probability.
2. Combine the probabilities of the two symbols having the lowest probabilities, and reorder the resultant probabilities, this step is called reduction 1. The same procedure is repeated until there are two ordered probabilities remaining.
3. Start encoding with the last reduction, which consist of exactly two ordered probabilities. Assign 0 as the first digit in the code words for all the source symbols associated with the first probability; assign 1 to the second probability.
4. Now go back and assign 0 and 1 to the second digit for the two probabilities that were combined in the previous reduction step, retaining all assignments made in Step 3.
5. Keep regression this way until the first column is reached.

An example of Huffman encoding is shown in Table given below:

Huffman Encoding



$$H(X) = 2.36 \quad \text{b/symbol}$$

$$L = 2.38 \quad \text{b/symbol}$$

$$\eta = 0.99$$

Example

A DMS has five equally likely symbols.

(i) Construct a Shannon-Fano code for X, and calculate the efficiency of the code.

(ii) Construct another Shannon-Fano code and compare the results.

(iii) Repeat for the Huffman code and compare the results.

Solution

(i) A Shannon-Fano code (by choosing two approximately equiprobable (0.4 verses 0.6) sets is constructed a flows (see table)

x_i	$P(x_i)$	Step 1	Step 2	Step 3	Code
x_1	0.2	0	0		00
x_2	0.2	0	1		01
x_3	0.2	1	0		10
x_4	0.2	1	1	0	110
x_5	0.2	1	1	1	111

$$H(X) = \sum_{i=1}^{5} P(x_i)\log_2 P(x_i) = 5(-0.2)\log_2 0.2) = 2.32$$

$$L = \sum_{i=1}^{5} P(x_i)n_i = 0.2(2+2+2+3+3) = 2.4$$

The efficiency η is $\quad \eta = \frac{H(X)}{L} = \frac{2.32}{2.4} = 0.967 = 96.7\ \%$ Ans.

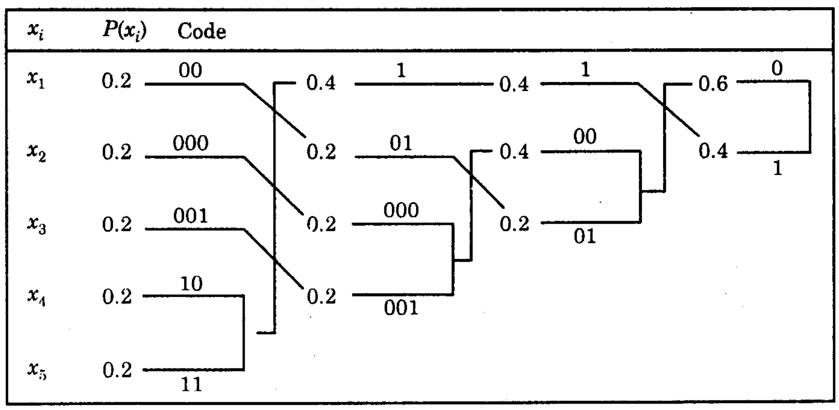
(ii) Another Shannon-Fano code [by choosing another two approximately equiprobable (0.6 verses 0.4) sets] is constructed as follows (see table)

x_i	$P(x_i)$	Step 1	Step 2	Step3	Step 4
x_1	0.2	0	0		00
x_2	0.2	0	1	0	010
x_3	0.2	0	1	1	011
x_4	0.2	1	0		10
x_5	0.2	1	1		11

$$L = \sum_{i=1}^{5} P(x_i)n_i = 0.2(2+3+3+2+2) = 2.4$$

Since the average code word length is the same as that for the code of part (a), the efficiency is the same.

(iii) The Huffman code is constructed as follows (see table)



$$L = \sum_{i=1}^{5} P(x_i)n_i = 0.2(2+3+3+2+2) = 2.4$$

since the average code word length is the same as that for the Shannon-Fano code, the efficiency is also the same.

Example

A DMS X has five symbols x_1,x_2,x_3,x_4, and x_5 with $P(x_1) = 0.4$, $P(x_2) = 0.19$, $P(x_3) = 0.16$, $P(x_4) = 0.15$, and $P(x_5) = 0.1$

(i) Construct a Shannon-Fano code for X, and calculate the efficiency of the code.

(ii) Repeat for the Huffman code and compare the results.

Solution

(i) The Shannon-Fano code is constructed as follows (see table)

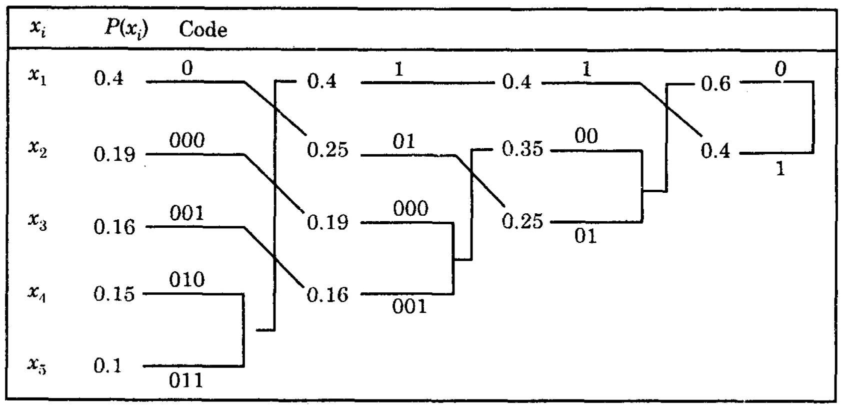
x_i	$P(x_i)$	Step 1	Step 2	Step 3	Code

x_1	0.4	0	0		00
x_2	0.19	0	1		01
x_3	016	1	0		10
x_4	0.15	1	1	0	110
x_5	0.1	1	1	1	111

$$H(X) = \sum_{i=1}^{5} P(x_i) n_i = 0.4(1) + 0.19(2) + 0.16(2) + 0.15(3) + 0.1(3) = 2.25$$

Also $$\eta = \frac{H(X)}{L} = \frac{2.15}{2.25} = 0.956 = 9.56 \ \%$$ Ans.

(ii) The Huffman code is constructed as follows (see table)



$$L = \sum_{i=1}^{5} P(x_i) n_i = 0.4(1) + 0.19 + 0.16 + 0.15 + 0.1(3) = 2.2$$ Ans.

The average code word length of the Huffman code is shorted than that of the Shannon-Fano code, and thus the efficiency is higher than that of the Shannon-Fano code.

Problem

A DMS X has five symbols x_1, x_2,…. x_5 with respective probabilities 0.2, 0.15, 0.05, 0.1 and 0.5.

(i) Construct a Shannon-Fano code for X, and calculate the code efficiency.

(ii) Repeat (i) for the Huffman code.

Answer:

(i)	*Symbols*	x_1	x_2	x_3	x_4	x_5
	Code	*10*	*110*	*1111*	*1110*	*0*

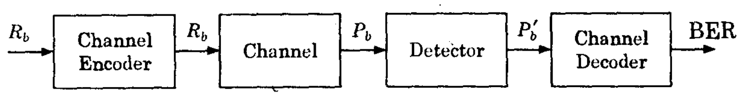
Code efficiency $\eta = 98.6\%$

(ii)	*Symbols*	x_1	x_2	x_3	x_4	x_5
	Code	*11*	*100*	*1011*	*1010*	*0*

Code efficiency $\eta = 98.6\%$

CHANNEL CODING

Channel encoding has the major function of modifying the binary stream in such a way that errors in the received signal can be detected and corrected. These are basically forward error correction (FEC) techniques as compared to Automatic Repeat Request (ARQ) techniques, where error can be corrected at the receiver without the need to request a retransmission. A channel coded digital communication system is shown in figure



Channel Coded Digital Communication System

R_b is the input bit rate, P_b is the output bit error probability and BER is the output bit error rate. A polar NRZ transmission bits matched filter detector would result in a bit error probability given by

$$P_b = \frac{1}{2}\mathrm{erfc}\sqrt{\frac{E_b}{N_0}} \qquad \text{for uncoded transmission}$$

which is the same as BER at the output i.e. if $P_b = 0.00001$. The BER = 1 bit in every 100,000 bits on an average with error control coding, the bit input rate R_b will be changed to new transmission bit rate Rb'. At the receiver, the bit error rate at the output of the channel decoder (BER') would be less than bit error probability Pb' at the input of the decoder.

Error Control Methods

As already stated error control (channel coding) is always associated with adding some redundancy to the source encoded digital signal. As we have seen with repetitive codes, it is possible to only detect errors but not correct them. Also the method adds so much redundancy that it is impractical. There are many advanced error control codes that have been developed over the years. But broadly these are classified as:

- **ARQ** (Automatic Repeat Request) Codes
- **FEC** (Forward Error Control) Codes.

ARQ codes are commonly used in network and computer data communications. They combine an error detecting code (such as cyclic redundancy check - CRC code) with a feedback mechanism that provides retransmission of the erroneously received block of data. ARQ codes use various protocols such as stop and wait protocol, go back 'n' (GBN) protocol, selective repeat (SR) protocol etc. which are mostly popular in computer networks.

In FEC codes, there is no retransmission of the erroneously received data. Most real time systems where a constant throughput and a very small (or no) delay is the requirement. Since

WHATSAPP: +91-9412903929 AMIESTUDYCIRCLE.COM INFO@AMIESTUDYCIRCLE.COM

no feedback is used in FEC codes, these will be designed for the worst channel conditions. The best alternative would be to use a Hybrid scheme which uses both ARQ and FEC techniques depending on the throughput and delay requirement of the system.

As there is no retransmission of messages in FEC codes, these coders have to include sufficient redundancy in the code so that errors can be detected and corrected at the receiver end. FEC codes can be classified as .

- Error concealment codes
- Linear error correcting codes

The *error concealment codes* are normally employed to detect an error (chick noises in magnetic tapes) and then conceal these by some technique as in the case of audio magnetic tapes. The particular sample that has been detected as erroneous will be concealed by replacing the last correctly received sample. This method is called as word repeat method. In another method, known as interpolation, the erroneous sample will be replaced by an interpolated value or an average value of a set of previously correctly received sample values.

The *linear error correcting codes* are very well organised techniques which are broadly classified as

- Algebraic codes or Block codes
- Convolutional codes

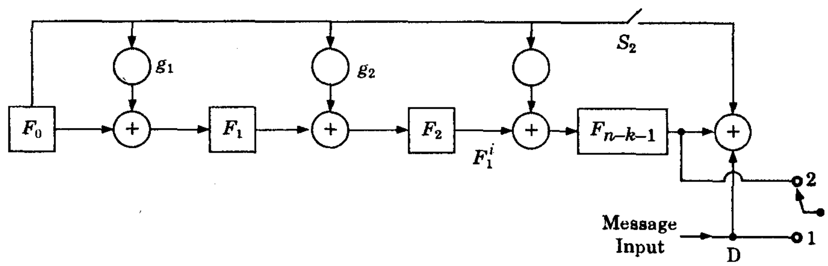
Algebraic Codes. The *algebraic codes or Block Codes* as the name suggests, are those in which a particular message word is encoded as a block of message bits and redundant (also called *parity or check*) bits. The length of the code word is fixed as 'n' bits. If ‘k' is the length of the message word and when the check bits get added, we obtain a code word of length 'n' for each message block of 'k' bits. The conversion from a 'k’ bit message into an ‘n' bit code is generally done by using a generator polynomial. Such a code is called a (n, k) block code. The number of extra/redundant/parity/check bits added is (n - k). There are several types of algebraic block codes. The most commonly used block codes are

- **Parity codes**. By adding a parity/check bit (odd or even), we can detect 1 bit errors but 2 bit errors can not be detected. This is called single parity check code. If any code word (say 1100) is transmitted, we just check the received code word for even parity. If 1100 is received as 1101, we see that the received word has an odd number of 1’s and hence we detect an error. But if there were 2 bits that were received by error, it is not possible even to detect. Correction is rules out. If there are k bits in the message, we have 2^k distinct message words. If (n – k) parity (extra) bits are added to the k message bits to end up with code words of length n bits where n > k is called a (n, k) block code.
- Hamming codes
- Hadamard codes
- Golay codes

- Cyclic codes. Cyclic codes are subclass of linear block codes. In cyclic codes, each code word of the code is the cyclic shift of the other code word of the same code. If (V_0, V_1, V_2, V_3,…… V_{n-1}) is a code word of a code, then (V_{n-1}, V_0, V_1, V_2 …. V_{n-2}) is also a code word of the code. The code polynomial of an (n, k) cyclic code is given as

  $$V(x) = V_0 + V_1x + V_2x^2 + ... + V_{n-1}x^{n-1}.$$

  Here all the coefficients of the above polynomial are 0's and 1's. A cyclic encoder is shown below:

  Its operation is simple. S_1 is kept in position 1 and S_2 is closed. The data bits d_0, d_1, …… d_{k-1} are then applied into the encoding circuit and into the channel with d_{k-1} as first bit through the switch kept in the above position (2). After all the data bits have been encountered into the shift register, the shift register contents are equal to the parity bits. Now S_2 is opened and S_1 is kept in position 2 and the contents of the shift register are shifted into the channel. Then the code word is obtained as

  $$\underbrace{f_0\, f_1 .... f_{n-k-1}}_{\text{parity bits}} \quad \underbrace{d_0 d_1 ..... d_{k-1}}_{\text{data bits}}$$

  During transmission of cyclic codes, errors do occur. Syndrome decoding can be used to correct these errors. Let the received vector be R(x). If E(x) denotes an error vector and S(x) is the syndrome then

  $$S(x) = \text{Rem.}\frac{R(x)}{g(x)}$$

  $$R(x) = V(x) + E(x)$$

  $$\therefore \qquad S(x) = \text{Rem.}\frac{E(x)}{g(x)}$$

  Most efficient code is BCH (Bose Choudhary Hocquenghem) code.

- Non-binary block codes
- Concatenated block codes etc.

Convolution Codes. In convolution codes, the source generates a continuing message sequence of 1's and 0's and the transmitted sequence is generated from this source sequence. The transmitted sequence will not be longer than the message sequence (i.e. no redundant bits are added as in block codes) but error correction capability is achieved by introducing memory into the system. The transmitted sequence is generated by convolving the source sequence with a fixed binary sequence. Hence these are called convolutional codes.

Encoding and Decoding of Codes

1. Information sequence is segmented into message blocks of k successive bits.
2. Each message block is transformed into a larger block of n bits by an encoder as per predetermined set of rules. The (n – k) redundant/additional bits are generated from linear combinations of the message bits.
3. Block codes are generated. Generation of block codes starts with selection of parity check matrix or H – matrix.
4. Block codes are decoded. The generator matrix which is used in encoding operation and in the parity check matrix H, is now used in decoding operation.

ASSIGNMENT

Q.1. (AMIE W05, S09, 4 marks): What do you understand by amount of information? Also define entropy of the source.

Q.2. (AMIE S06, 4 marks): Define uncertainty and entropy.

Q.3. (AMIE W09, 12, S12, 10 marks): Define rate of information, joint entropy, conditional entropy, mutual information and redundancy.

Q.4. (AMIE W08, 12 marks): define entropy of a source. What are the properties of entropy? Define average mutual information over a joint ensemble and explain with an example.

Q.5. (AMIE S11, 4 marks): Define entropy of a source. What is conditional entropy.

Q.6. (AMIE S05, W12, 06): Define channel capacity? Derive an expression for the capacity of binary symmetric channel.

Q.7. (AMIE S08, 6 marks): Derive an expression for the channel capacity of PCM system.

Q.8. (AMIE S08, 6 marks): Deduce the expression

$$C = B\log_2\left(1+\frac{12}{\lambda^2}\frac{S}{N}\right)$$

for channel capacity of a PCM equalizer.

Q.9. (AMIE S13, 10 marks): Derive the Gaussian channel capacity.

Q.10. (AMIE S10, 6 marks): Show that the channel capacity of a channel, disturbed by Gaussian white noise, is given by

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

Q.11. (AMIE W08, 10, 8 marks): State Shannon channel coding theorem and explain the concept of "channel capacity".

Q.12. (AMIE W07, 12 marks): Define channel capacity and channel efficiency for a discrete noisy channel. Describe the properties of binary symmetric channel and binary erase channel. Derive an expression for channel capacity in terms of sampling conditioning.

Q.13. (AMIE S10, 2 marks): What is a binary symmetric channel?

Q.14. (AMIE W10, 8 marks): Define channel capacity. Draw the transition probability diagram of a binary symmetric channel and derive an expression for its capacity in terms of probability.

Q.15. (AMIE W06, 8 marks): Explain the relation between systems capacity and information content of messages.

Q.16. (AMIE W06, 8 marks): Explain the terms: Information, entropy, redundancy, rate of communication.

Q.17. (AMIE W07, 8 marks): What is entropy of a source? How can you find maximum entropy of several alphabets? Define code efficiency and redundancy of codes.

Q.18. (AMIE W06, 4 marks): What do you understand by channel coding?

Q.19. (AMIE S08, 8 marks): State and explain the channel coding theorem. Explain the application of this theorem in binary symmetric channel.

Q.20. (AMIE S07, 6 marks): State and explain the aspect of transmission channel which is defined by Shannon-Hartley law. How does noise affect channel capacity.

Q.21. (AMIE S12, 5 marks): State and explain Shannon-Hartley theorem.

Q.22. (AMIE S07, 6 marks): State and describe the three types of error detection codes and explain how they detect data errors. What are the merits and demerits of channel coding? List them.

Q.23. (AMIE S09, 8 marks): Find the equation for coding efficiency by using source coding theorem.

Q.24. (AMIE S11, 8 marks): What is parity check matrix? How parity check matrix for cyclic codes help in channel coding?

Q.25. (AMIE S06, 10 marks): A source emits one of four symbols s_0, s_1, s_2 and s_3 with probabilities 1/3, 1/6, 1/4 and 1/3 respectively. The successive symbols emitted by the source are statically independent. Calculate the entropy of the source. Derive the formula you use.

Answer: $\frac{1}{3}\log 3 + \frac{1}{6}\log 6 + \frac{1}{4}\log 4$ bits/message

Q.26. (AMIE S07, 8 marks): A source emits one of six symbols α_0, α_1,…….α_5 with probabilities 1/6, 1/8, 1/4, 1/6, 1/6, 1/8 respectively. The symbols emitted by the source are statistically independent. Calculate the entropy of the source and derive the same formula.

Answer: $\frac{1}{6}\log_2(6) + \frac{1}{8}\log_2(8) + \frac{1}{4}\log_2(4) + \frac{1}{6}\log_2(6) + \frac{1}{6}\log_2(6) + \frac{1}{8}\log_2(8)$

Q.27. (AMIE S10, 6 marks): A message source produces five symbols with the probabilities of occurrence as 1/2, 1/6, 1/6, 1/12 and 1/12. Calculate the entropy of the source.

Answer: 1.94 b/symbol

Q.28. (AMIE W13, 5 marks): A source is generating four possible symbols with probabilities of 1/8, 1/8, ¼, ½ respectively. Find entropy and information rate, if the source is generating 1 symbol/ms.

Q.29. (AMIE W05, 10 marks): Consider the two sources S_1 and S_2 emit x_1, x_2, x_3 and y_1, y_2, y_3 with joint probability p(X,Y) as shown in the matrix form. Calculate H(X), H(Y), H(X|Y) and H(Y|X).

$$\begin{array}{cc} & \begin{bmatrix} y_1 & y_2 & y_3 \end{bmatrix} \\ P(x,y) = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} & \begin{bmatrix} 3/40 & 1/40 & 1/40 \\ 1/20 & 3/20 & 1/20 \\ 1/8 & 1/8 & 3/8 \end{bmatrix} \end{array}$$

Answer: 1.298794941 bits, 1.53949107 bits, 1.019901563 bits, 1.260597692 bits

Q.30. (AMIE S08, 6 marks): Five symbols of a discrete memoryless source and their probabilities are given below:

Symbols	Probability	Code Word
s_0	0.4	00
s_1	0.2	10
s_2	0.2	11
s_3	0.1	010
s_4	0.1	011

Determine the average code word length and the entropy of the memoryless source.

Q.31. (AMIE S11, 6 marks): A binary source emits an independent sequence of "0s" and "1s" with probabilities p and 1 - p, respectively. Evaluate the variation of entropy at p = 0.05 and 1.

WHATSAPP: +91-9412903929 AMIESTUDYCIRCLE.COM INFO@AMIESTUDYCIRCLE.COM

Q.32. (AMIE S08, 6 marks): A system has a bandwidth of 4 kHz and a signal to noise ratio of 28 dB at the input of the receiver. If the bandwidth of the channel is doubles while transmitted signal power remains same, determine capacity of the channel.

Q.33. (AMIE S10, 6 marks): A source emits seven symbols with probabilities 1/2, 1/4, 1/8, 1/16, 1/64 and 1/64 respectively. Determine a Huffman code for it.

Q.34. (AMIE S12, 5 marks): For the binary symmetric channel, find the channel capacity for p = 0.9.

Q.35. (AMIE W13, 4 marks): Obtain maximum possible entropy for an 8 symbol source.

Answer: 3 bits/symbols, Hint: $H_{max} = \log_2 M$

WHATSAPP: +91-9412903929 AMIESTUDYCIRCLE.COM INFO@AMIESTUDYCIRCLE.COM