Datta Meghe College of Engineering, Airoli, Navi Mumbai

Information Theory and Source Coding

Digital Communication
(ELC 504)

- S N Vaidya

What is information?

- Consider three hypothetical statements
  1. It is going to rain today
  2. There will be thunderstorm and flooding today
  3. Their will be snow fall.

- The information content of any message is closely related to the past knowledge of the occurrence of event and the level of uncertainty it contains with respect to the recipient of the message

- The amount of information received from the knowledge of occurrence of an event is related to the probability or the likelihood of occurrence of the event.
- The message related to an event least likely to occur contains more information.

- Let us consider a communication system which contains message
- $m_1, m_2, \ldots \ldots \ldots \ldots \ldots \ldots ..$ With probabilities of occurrence $p_1, p_2, \ldots \ldots \ldots \ldots \ldots \ldots ..$

$$p_1 + p_2 + \cdots \ldots \ldots \ldots \ldots = 1$$

- Let the transmitter, select message m_k of probability p_k
- Then, by way of definition of the term information that the system has conveyed an amount of information I_k is given by

$$I_k = \log_2 \frac{1}{p_k}$$

- If p is the probability of occurrence of a message and I is the information gained from the message, it is evident from the discussion that

$$P \rightarrow 1, I \rightarrow 0 \text{ and} P \rightarrow 0, I \rightarrow \infty$$

Engineering measure of information

Problem on Information

- A binary symbol occurs with a probability of 0.75. Determine the information associated with the symbol in bits, nats and Hartley.

Average Information , Entropy

- Suppose we have M different and independent messages $m_1, m_2, \dots \dots \dots \dots \dots \dots ..$ with probabilities if occurrence $p_1, p_2, \dots \dots \dots \dots \dots \dots ..$ Suppose further that during a long period of transmission a sequence of L messages has been generated.
- Then, if L is very large, we may express that in the L message sequence we transmitted $p_1 L$ message of m_1, $p_2 L$ message of m_2, etc.
- The total information in such a sequence will be
- $I_{total} = p_1 L \log_2 \frac{1}{p_1} + p_2 L \log_2 \frac{1}{p_2} + \dots\dots\dots\dots\dots\dots\dots\dots\dots$

- The average information per message interval represented by the symbol H will then be

$$H \equiv \frac{I_{total}}{L} = p_1 \log_2 \frac{1}{p_1} + p_2 \log_2 \frac{1}{p_2} + \cdots \ldots \ldots \ldots \ldots \ldots \ldots \ldots ..$$

$$H = \sum_{k=1}^{M} p_k \log_2 \frac{1}{p_k}$$

- This average information is also referred to by the term entropy
- We have seen that when there is only a single possible message (p_k =1) , the receipt of that message conveys no information.
- At the other extreme, as $p_k \to 0, I_k \to \infty$.
- However, since

$$\lim_{p \to 0} p \log \frac{1}{p} = 0$$

- The average information associated with an extremely unlikely message , as well as an extremely likely message is zero

- Consider the case of just two message with probabilities p and $(1-p)$
- The average information per message is

$$H = p\log_2\frac{1}{p} + (1-\mathrm{p})\log_2\frac{1}{1-p}$$

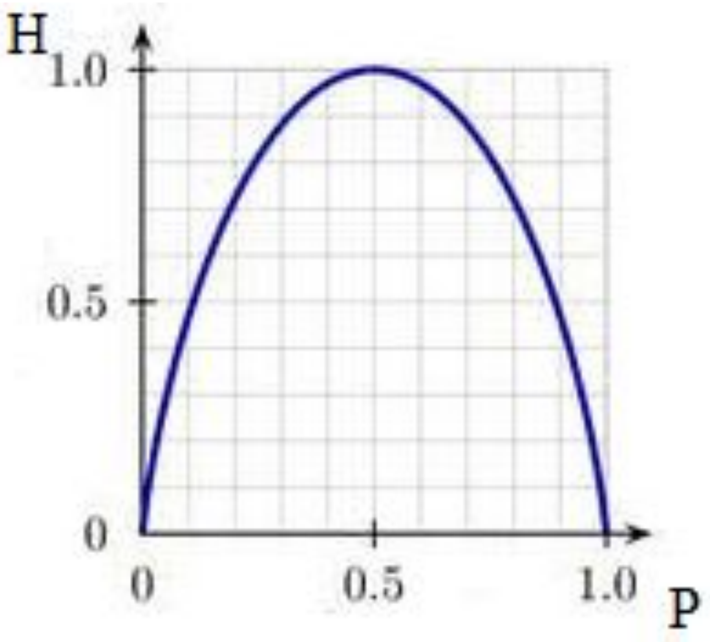
- A plot of H as a function of p is shown below
- The maximum occurs at $p = {}^{1}/_{2}$ i.e. when the

two messages are equally likely i.e.

$$H_{max} = \frac{1}{2}\log_2 2 + \frac{1}{2}\log_2 2$$

$$= \log_2 2$$

$$H_{max} = 1 \text{ bit/message}$$



Problem on Entropy

- A discrete memoryless source is capable of transmitting three distinct symbols, m_0, m_1 and m_2. Their probabilities are ½, ¼ and ¼ respectively. Calculate the source entropy.

Information Rate

- Average rate at which information is transferred is called information rate

$$R = rH \frac{bits}{sec}$$

Where

r= symbol rate

H= Entropy

Problem on Information rate

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

Source Encoding

- The process of efficient representation of data generated by a discrete source is known as source encoding
- The source encoder convert the symbol sequence into the binary sequence by assigning codeword to the symbol in the input sequence.
- The codeword assigned may either be fixed length or variable length
- In a fixed length codeword, fixed length binary codeword is assigned to each symbol whereas, in a variable length codeword more bits are assigned to rarely occurring symbols and less bits are assigned to frequently occurring symbols

Problem

- A and B are two cities with weather conditions Sunny, Cloudy , Rainy and Foggy

A			B		
Sunny	¼	00	Sunny	½	0
Cloudy	¼	01	Cloudy	1/8	1110
Rainy	¼	10	Rainy	1/8	110
Foggy	1/4	11	Foggy	1/4	10

- Average Info for A=$2 \times \frac{1}{4} + 2 \times \frac{1}{4} + 2 \times \frac{1}{4} + 2 \times \frac{1}{4} = 2$

$$B = 1 \times \frac{1}{2} + 4 \times \frac{1}{8} + 3 \times \frac{1}{8} + 2 \times \frac{1}{4} = 1.875$$

Source Coding

- The objective in designing a source encoder is to find out unique binary code word Ci of the length n_i bits for the message m_i for $i = 1,2,3 \dots \dots \dots q$ such that the average number of bits per symbol used in the coding scheme is as close to the average information per symbol.
- Goal is to find an efficient description of information sources – Reduce required bandwidth – Reduce memory to store
- Memoryless –If symbols from source are independent, one symbol does not depend on next
- Memory – elements of a sequence depend on one another, e.g. UNIVERSIT_?

Shannon-Fano Coding

- Steps
  1. Arrange the message m_i in order of decreasing probability
  2. Divide the messages into two halves such that the sum of probabilities are as close as possible.
  3. Repeat the above step for each half until individual elements are left
  4. Append 1 to one half and 0 to the other half
  5. Collect the bits
  6. That is the codeword

Problem

- Consider 8 messages m_1 *through* m_8 with probabilities ½,1/8,1/8,1/16,1/16,1/16,1/32 and 1/32

Message	Probability	I	II	III	IV	V	No.of bits/message
M_1	½	0					0 (1)
M_2	1/8	1	0	0			100 (3)
M_3	1/8	1	0	1			101 (3)
M_4	1/16	1	1	0	0		1100 (4)
M_5	1/16	1	1	0	1		1101 (4)
M_6	1/16	1	1	1	0		1110 (4)
M_7	1/32	1	1	1	1	0	11110 (5)
M_8	1/32	1	1	1	1	1	11111 (5)

I. Entropy

$$H = \sum_{k=1}^{8} p_k \log_2 \frac{1}{p_k}$$

$$= \frac{1}{2}\log_2 \frac{1}{2} + 2 \times \frac{1}{8}\log_2 \frac{1}{8} + 3 \times \frac{1}{16}\log_2 \frac{1}{16} + 2 \times \frac{1}{32}\log_2 \frac{1}{32}$$

$$H = 2\frac{5}{16}$$

II. Average codeword length

$$L = \sum_{K=1}^{8} Probability \times (length\ of\ bits)$$

$$= \frac{1}{2} \times 1 + \frac{1}{8} \times 3 + \frac{1}{8} \times 3 + \frac{1}{16} \times 4 + \frac{1}{16} \times 4 + \frac{1}{16} \times 4 + \frac{1}{32} \times 5 + \frac{1}{32} \times 5$$

$$L = 2\frac{5}{16}$$

III Code Efficiency

$$\eta = \frac{H}{L}$$

$\eta = 1$

IV Code Redundancy

$$r = 1 - \eta$$

$r = 0$

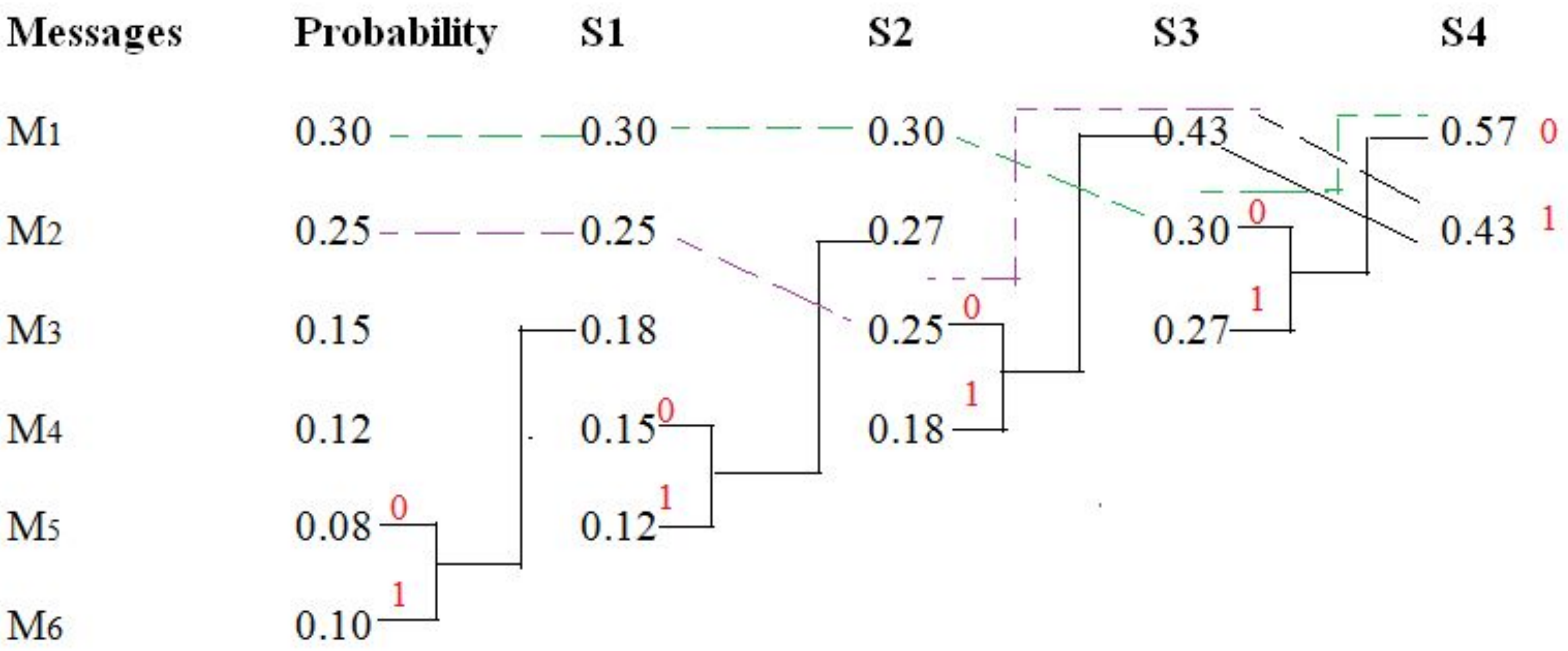
Huffman Coding

- Steps

1. Arrange the message m_i in order of decreasing probability
2. Combine two messages of lowest probability and write their combined value.
3. Keep the combination at the top
4. Repeat step 3 till only two remains
5. Next encode by starting from the final reduction and assigning bit '0' to the top probability and '1'to the bottom one.
6. Repeat step 5 for each combination
7. The encoded message then is represented by a bit it goes through starting from the final reduction.

Problem

- Consider six messages M_1 through M_6 with probabilities 0.3, 0.25,0.15,0.12,0.08,0.10 respectively. Calculate using Huffman Code:- Average length, Entropy, Code efficiency and Redundancy.



Messages	Probability	Code word
M_1	0.30	00
M_2	0.25	10
M_3	0.15	010
M_4	0.12	011
M_5	0.08	110
M_6	0.10	111

Problem

Q. Consider alphabets a, b, c, d, e and f with probability 0.4, 0.2, 0.1, 0.1, 0.1 and 0.1. Construct Huffman code and find

a. Entropy

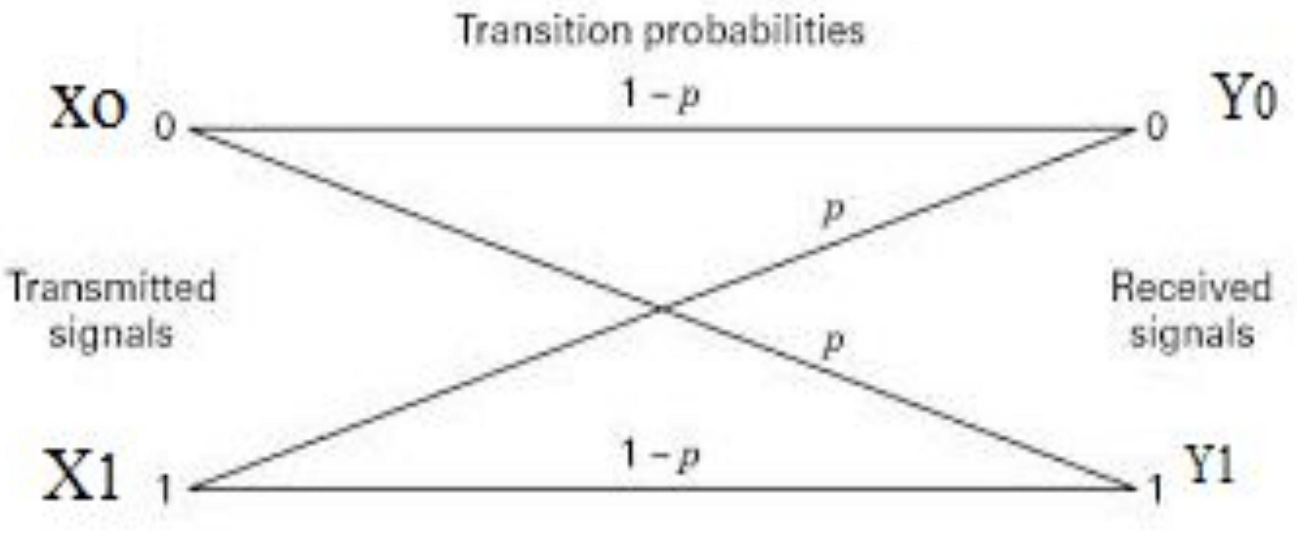
b. Average code word length

c. Code Efficiency

d. Code Redundancy

Types of Channel

- 1. Discrete memoryless Channel
- 2. Binary Symmetric Channel



Mutual Information

- The Mutual Information between two random variables measures non-linear relations between them. Besides, it indicates **how much information can be obtained from a random variable** by observing another random variable.
- It is closely linked to the concept of **entropy**. This is because it can also be known as the reduction of **uncertainty** of a random variable if another is known. Therefore, a high mutual information value indicates a large reduction of uncertainty whereas a low value indicates a small reduction. If the mutual information is zero, that means that the two random variables are **independent**.

How mutual information is calclauted

- Let $P(y_k/x_j)$ be the transition probabilities.
- Mutual Information is given as
- $I(X;Y) = \sum_{J=0}^{J-1} \sum_{K=0}^{K-1} P(x_j/y_k) \log_2 \left[\frac{P(y_k/x_j)}{P(y_k)}\right]$

CHANNEL CAPACITY

Shannon's Theorem, Channel Capacity

Theorem Given a source of M equally likely messages, with $M \gg 1$, which is generating information at a rate R. Given a channel with channel capacity C. Then, if

$$R \leq C$$

there exists a *coding* technique such that the output of the source may be transmitted over the channel with a probability of error in the received message which may be made arbitrarily small.

Capacity of Gaussian Channel/ Shannon-Hartley Theorem

Theorem The channel capacity of a white, bandlimited gaussian channel

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

where B is the channel bandwidth, S the signal power, and N is the total noise within the channel

Bandwidth- S/N Tradeoff

The Shannon-Hartley theorem, indicates that a noiseless gaussian channel ($S/N = \infty$) has an infinite capacity. On the other hand, while the channel capacity does increase, it does not become infinite as the bandwidth becomes infinite because, with an increase in bandwidth, the noise power also increases. Thus for a fixed signal power and in the presence of white gaussian noise the channel capacity approaches an upper limit with increasing bandwidth. We now calculate that limit. Using $N = \eta B$

we have $$C = B \log_2\left(1 + \frac{S}{\eta B}\right) = \frac{S}{\eta}\frac{\eta B}{S}\log_2\left(1 + \frac{S}{\eta B}\right)$$

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

We recall that $\lim_{x\to 0}(1 + x)^{1/x} = e$ (the naperian base), and identifying x as $x = S/\eta B$, we find that

$$C_\infty = \lim_{B\to\infty} C = \frac{S}{\eta}\log_2 e = 1.44\frac{S}{\eta}$$