MODULE 1 INFORMATION THEORY

STRUCTURE

1. Objectives
2. Introduction
3. Measure of information
4. Average information Content(entropy) in long independent sequences
5. Mark-off statistical model for information sources
6. Review questions.
7. Outcomes.

OBJECTIVES

After completion of this module the student will be able

1. To learn about the probability theory, linear algebra, random processes and communication systems.
2. To learn the basic concept of information theory and coding.
3. To find the average information content.

1.1INTRODUCTION

The block diagram of an information system can be drawn as shown in figure. the meaning of the word "information "in information theory is "message" or "intelligence". This message may be an electrical message such as voltage, current or power or speech message • picture message such as fascimile or television or music message. A source which produces these messages is called "information source".

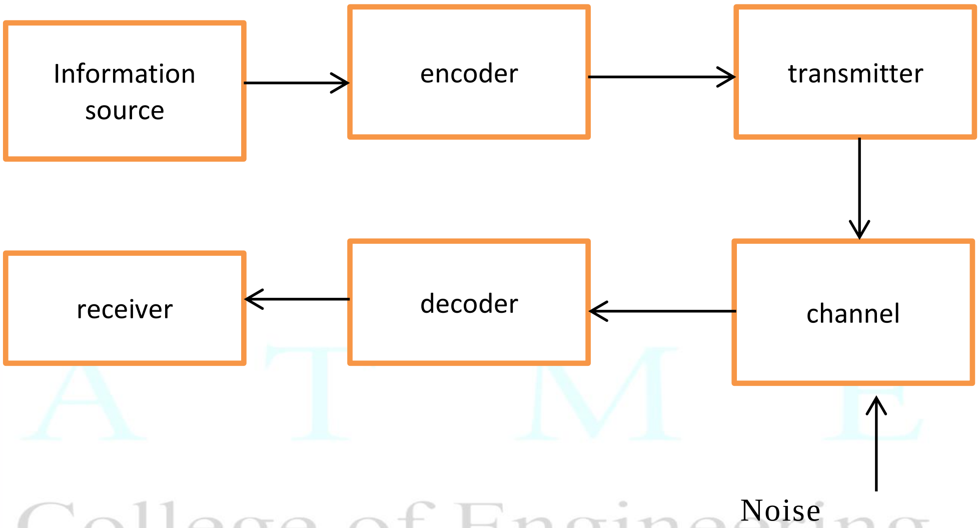


Figure: Block diagram of an Information source

Information sources can be classified into two categories: analog information sources id discrete information sources. Analog information sources, such as a microphone actuated (speech, or a TV camera scanning a scene, emit one or more continuous amplitude electrical finals with respect to time. The output of discrete information sources such as a teletype or the numerical output of a computer consists of a sequence of discrete symbols or letters. An analog information source can be transformed into a discrete information source through the ocess of sampling and quantizing.

Discrete information sources are characterized by (a) source alphabet (b) symbol rate) source alphabet probabilities and (d) probabilistic dependance of symbols in a sequence.

Example of source alphabet (discrete information source) a teletype having 26 letters of the English Alphabet plus several special characters such as full stop, comma etc. along with numerals.

The Symbol rate refers to the rate at which the teletype produces characters. Ex: If the teletype operates at the speed of 10 characters/sec, then the symbol rate is said to be symbols/sec.

If the teletype is producing messages in English language, then some letters appear more frequently than others. For example, the letter F appears more often than the letter Z and if a word starts with Q, the next letter will be U and so on. These structural properties of symbol sequences can be characterized by probability of occurrence of the individual symbols and by the conditional properties of occurrence of symbols (i.e., probabilistic dependence).

SOURCE ENCODER: Let the input to the source encoder he a string Of source yymbti from the source Alphabet S = {s1, s,, ..., sn }occurring at a rate of "r," symbols/sec.

The source encoder converts the symbol sequence into it binary sequence of U's and v by assigning code-words to the symbols in the input sequence. Binary coding is preferit, because of its high efficiency of transmission and also the case with which they can ii transmitted over the channel [other types of coding such as ternary, quarternary coding etct discussed in unit 31. The simplest way of coding is to assign a fixed length binary code-we, to each symbol in the input sequence. But, fixed-length coding of individual symbols in, source output is efficient only if the symbols occur with equal probabilities in a statistics independent sequence. In most practical situations, the symbols occur with uricqt4 probabilities. The source encoder, then assigns variable length code-words to these symb6 The important parameters of a source encoder namely block size, length of code-words, average data rate and the encoder efficiency.

TRANSMITTER: The transmitter couples the input message signal to the channel. While it may sometimes be possible to couple the input transducer directly to the channel. It is often necessary to process and modify the input signal for efficient transmission over the channel. Signal processing operations performed by the transmitter include arnplification, filtering and modulation. The most important of these operations is modulation - a process designed to match the properties of the transmitted signal to the channel through the use of carrier wave.

CHANNEL: A communication channel provides the electrical connection between the source and the destination. The channel may be a pair of wires (2-line transmission system or a telephone cable or free space over which the information bearing signal is radiated. Due to physical limitations, communication channels have only finite bandwidth and the information bearing signal suffers amplitude and phase distortion as it travels over the channel. In add° to the distortion, the signal power also decreases due to attenuation of the channel.

DECODER AND RECEIVER: The source decoder converts the binary output of the channel decoder into a symbol sequence. The decoder for fixed-length code-words is quite simple, but the decoder for a system using variable-length code-words will be very complex. Therefore, the function of the decoder is to convert the corrupted signals into a symbol sequence and the function of the receiver is to identify the symbol sequence and match it with the correct sequence.

In 1948, C.E. SHANNON, known as "Father of Information Theory", published a treatise on the mathematical theory of communication in which he established basic theoretical bounds for the performances of communication systems. Shannon's theory is based on probabilistic models for information sources and communication channels. In the forthcoming sections, we present some of the important aspects of Shannon's work.

1.2 MEASURE OF INFORMATION

In order to know and compare the "information content" of various messages produced by an information source, a measure is necessary to quantitatively know that information content. For this, let us consider an information source producing independent sequence of symbols from source alphabet S = {s1, s2,… sq} with probabilities P = {p1, p2, …pq} respectively.

Let S_K be a symbol chosen for transmission at any instant of time with a probability equal to pK. Then the "Amount of Information" or "Self-Information" of message S_K (provided it is correctly identified by the receiver) is given by

$$1_K = \log(1/P_K)$$

If the base of the logarithm is 2, then the units are called "BITS", which is the short form of "Binary Units". If the base is "10", the units are "HARTLEYS" or "DECITS". If the base is "e", the units are "NATS" and if the base, in general, is "r", the units are called "r-ary units".

The most widely used unit of information is "BITS" where the base of the logarithm is 2. Throughout this hook, log to the base 2 is simply written as log and the units can be assumed to the bits, unless or otherwise specified.

Example 1.1 : The binary symbols '0' and '1' are transmitted with probabilities ¼ and ¾ respectively. Find the corresponding self-informations.

Solution

Self-information in a '0' = $I_0 = \log(1/P_0) = \log 4 = 2$ bits.

Self-information in a '1' = $I_1 = \log(1/P_1) = \log 4/3$

I_1=0.415 bits.

Thus, it can be observed that more information is carried by a less likely message. Logarithmic expression is chosen for measuring information because of the follow reasons:

1. The information content or self-information of any message cannot be negative. Each message must contain certain amount of information.

2. The lowest possible self-information is "zero" which occurs for a sure event since P (sure event) = 1.

1.2.1 Zero-Memory Source:- It represents a model of a discrete information source emitting sequence of symbols from a fixed finite source alphabet S = {s_1, s2,….sq} Successive symbols are selected according to some fixed probability law and are statistically independent of one another. This means that there is no connection between any two symbols and that the urce has no memory. Such type of sources are called **"memoryless"** or **"zero-memory"** sources.

1.3 AVERAGE INFORMATION CONTENT (ENTROPY) OF SYMBOLS IN LONG INDEPENDENT SEQUENCES

Let us consider a zero-memory source producing independent sequences of symbols. while the receiver of these sequences may interpret the entire message as a single unit, communication systems often have to deal with individual symbols. Let us consider the source alphabet S = {s1, s2, sq} with probabilities P = (p1, p2, pq} respectively.

Let us consider a long independent sequence of length L symbols. This long sequence en contains

P1L number of messages of type s1

P2L number of messages of type s2,

and PqL number of messages of type sq.

Average self-information is also called entropy is given by

$$\mathbf{H(S)=\sum_{i=1}^{q} pilog(\frac{1}{pi})\ bits/message\ symbol}$$

Note that the definition of MS) given in equation (1.4) is based on time-averaging-11: definition is valid for ensemble averages provided the source is ergodic The source entropy given by equation (1.4) is similar to the expression for entropy;.. statistical mechanics. The source entropy can be interpreted as follows. On the average.., can expect to get H(S) bits of information per symbol in long messages from the information source eventhough we cannot say in advance what symbol sequences will occur in these messages. Thus H(S) represents the "average uncertainty" or the 'average amount of the source.

Illustration: Let us consider a binary source with source alphabet S={s1,s2} with probabilities P={1/256, 255/256}

Solution:

$H(S)=\sum_{i=1}^{q} pilog(\frac{1}{pi})$ bits/message symbol

The entropy H(s)=0.037 bits/message symbol

INFORMATION RATE : Let us suppose that the symbols are emitted by the source at fixed time rate "rs" symbols/sec. The "average source information rate R," in bits/sec is fined as the product of the average information content per symbol and the message symbol rate rs

R = rsH(S) bits/sec or BPS

1.4 PROPERTIES OF ENTROPY

The entropy function is given by equation (1.4) for a source alphabet S = {s1, s2,..sq} with P = {p1, p2, pq} where q = number of source symbols, as

$H(S)=\sum_{i=1}^{q} pilog(\frac{1}{pi})$ bits/message symbol

Many interesting properties can be observed as listed below:

The entropy function is continuous for every independent variable PK in the interval (0,1). i.e., if Pi(varies continuously from 0 to 1, so does the entropy function.[Note: Entropy function vanishes at both pK = 0 and pK = 1].

2. The entropy function is a symmetrical function of its arguments. i.e., H [pK, (1 — pK)] = H [(1 — pK), pK] for all K = 1,2, …q i.e., the value of H(S) remains the same irrespective of the locations of the probabilities. i.e., as long as the probabilities are same, it does not matter in which order they are arranged. Thus the sources SA, SB and SC with probabilities.

PA ={P1,P2,P3}

PB ={P2,P3,P1}

PA ={P3,P1,P2}

Such that $\sum_{i=1}^{3} Pi = 1$ will all have the same entropy i.e.,H(SA)= H(SB)= H(SC)

1.4.1 Extremal Property

Let us consider the same source S with q symbols S={s1, s2,…….. sq} with probabilities P={P1,P2,…..Pq}. the entropy of S is given by equation

$H(S)=\sum_{i=1}^{q} pi\log(\frac{1}{pi})$

The entropy has an upper bound is logq-H(S).

The lower bound for H(S) is zero.

1.4.2 Property of Additivity

Suppose that we split the symbol sq into 'n' subsymbols such that sq= sq1, sq2,…… sqn . occurring with probabilities Pq1, Pq2,……… Pqn such that

$\sum_{j=1}^{q} Pqj$

Source Efficiency=H(S)/H(S)max

Redundancy= 1-source efficiency

1.5MARK OFF MODEL FOR INFORMATION SOURCES ASSUMPTION

A source puts out symbols belonging to a finite alphabet according to certain probabilities depending on preceding symbols as well as the particular symbol in question.

1.5.1. Define a random process

A statistical model of a system that produces a sequence of symbols stated above is and which is governed by a set of probs. is known as a random process.

Therefore, we may consider a discrete source as a random process

And the converse is also true.

i.e. A random process that produces a discrete sequence of symbols chosen from a finite set may be considered as a discrete source.

1.5.2. Discrete stationary Mark off process

Provides a statistical model for the symbol sequences emitted by a discrete source.

General description of the model can be given as below:

1. At the beginning of each symbol interval, the source will be in the one of "n" possible states 1, 2, ….. n

Where "n" is defined as

$n \leq (M)^m$

M = no of symbol / letters in the alphabet of a discrete stationery source,

m = source is emitting a symbol sequence with a residual influence lasting

„m" symbols.

i.e. m: represents the order of the source.

m = 2 means a 2nd order source

m = 1 means a first order source.

The source changes state once during each symbol interval from say i to j. The probabilityy of this transition is P_{ij}. P_{ij} depends only on the initial state i and the final state j but does not depend on the states during any of the preceeding symbol intervals.

2. When the source changes state from I to j it emits a symbol. Symbol emitted depends on the initial state i and the transition ij.

3. Let $s_1, s_2, \ldots s_M$ be the symbols of the alphabet, and let $x_1, x_2, x_3, \ldots\ldots x_k, \ldots\ldots$ be a sequence of random variables, where x_k represents the k^{th} symbol in a sequence emitted by the source. Then, the probability that the k^{th} symbol emitted is s_q will depend on the previous symbols $x_1, x_2, x_3, \ldots\ldots\ldots, x_{k-1}$ emitted by the source.
   i.e., $P(X_k = s_q / x_1, x_2, \ldots\ldots, x_{k-1})$

4. The residual influence of $x_1, x_2, \ldots\ldots, x_{k-1}$ on x_k is represented by the state of the system at the beginning of the k^{th} symbol interval.
   i.e. $P(x_k = s_q / x_1, x_2, \ldots\ldots, x_{k-1}) = P(x_k = s_q / S_k)$
   When S_k in a discrete random variable representing the state of the system at the beginning of the k^{th} interval.Term states" is used to remember past history or residual influence in the same context as the use of state variables in system theory / states in sequential logic circuits.

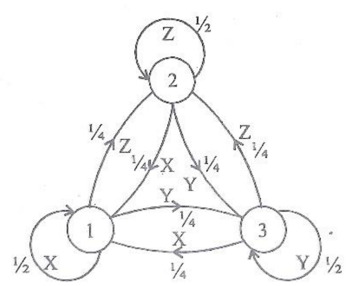
1.5.3 System Analysis with regard to Markoff sources

Representation of Discrete Stationary Markoff sources:

1. Are represented in a graph form with the nodes in the graph to represent states and the transition between states by a directed line from the initial to the final state.

2. Transition probs. and the symbols emitted corresponding to the transition will be shown marked along the lines of the graph.

A typical example for such a source is given below.



It is an example of a source emitting one of three symbols X, Y, and Z

- The probability of occurrence of a symbol depends on the particular symbol in question and the symbol immediately proceeding it.
- Residual or past influence lasts only for a duration of one symbol.

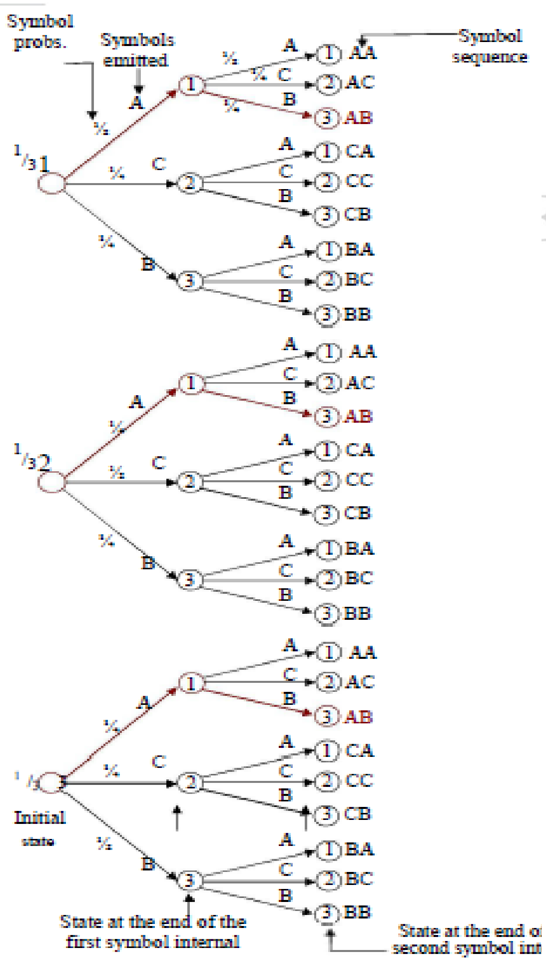
Last symbol emitted by this source

The last symbol emitted by the source can be A or B or C. Hence past history can be represented by three states- one for each of the three symbols of the alphabet.

State transition and symbol generation can also be illustrated using a tree diagram.

1.6 Tree diagram

Tree diagram is a planar graph where the nodes correspond to states and branches correspond to transitions. Transitions between states occur once every T_s seconds. Along the branches of the tree, the transition probabilities and symbols emitted will be indicated. Tree diagram for the source considered



Use of the tree diagram

Tree diagram can be used to obtain the probabilities of generating various symbol sequences.

Generation a symbol sequence say AB

This can be generated by any one of the following transitions:

$$1 \longrightarrow 2 \longrightarrow 3$$

OR

$$2 \longrightarrow 1 \longrightarrow 3$$

OR

$$3 \longrightarrow 1 \longrightarrow 3$$

Therefore probabilities of the source emitting the two – symbol sequence AB is given by

$$P(AB) = P(S_1 = 1, S_2 = 1, S_3 = 3)$$

Or

$$P(S_1 = 2, S2 = 1, S3 = 3) \quad \text{----- (1)}$$

Or

$$P(S_1 = 3, S2 = 1, S3 = 3)$$

Note that the three transition paths are disjoint.

Therefore $P(AB) = P(S1 = 1, S2 = 1, S3 = 3) + P(S1 = 2, S2 = 1, S3 = 3)$
$+ P(S1 = 2, S2 = 1, S3 = 3)$ ----- (2)

The first term on the RHS of the equation (2) can be written as

$P(S1 = 2, S2 = 1, S3 = 3)$

$= P(S1 = 1)\, P(S2 = 1 / S1 = 1)\, P(S3 = 3 / S1 = 1, S2 = 1)$

$= P(S1 = 1)\, P(S2 = 1 / S1 = 1)\, P(S3 = 3 / S2 = 1)$

Review questions:

1. Explain the terms (i) Self information (ii) Average information (iii) Mutual Information (iv) Efficiency (v) Redundancy.

2. Discuss the reason for using logarithmic measure for measuring the amount of information.

3. Explain the concept of amount of information associated with message. Also explain what infinite information is and zero information.

4. A binary source emitting an independent sequence of 0‟s and 1‟s with pro babilities p and (1-p) respectively. Plot the entropy of the source.

5. Explain the concept of information, average information, information rate and redundancy as referred to information transmission.

6. Suppose that a large field is divided into 64 squares. In the dark night, a cow has entered in this field and it is equally likely to be in any of the squares. This cow is located by a member of searching party who sends back information giving the location of the cow as 43 rd square. Calculate the amount of information obtained in the reception of the message.

Outcome

Able to understand the concept of probability theory, linear algebra, random processes and develop applications for communication systems.
Able to apply the knowledge to find the self-information, entropy and mutual information.
Able to apply the knowledge to analyze the Mark off statistical models.

Resources

- https://en.wikipedia.org/wiki
- www.inference.phy.cam.ac.uk/mackay/itprnn/1997/l1/node7.html
- web.ntpu.edu.tw/~yshan/intro_lin_**code**.pdf
- users.ece.cmu.edu/~koopman/des_s99/**coding**/
- elearning.vtu.ac.in/P4/EC63/S11.pdf