PRINCIPLES OF COMMUNICATIONS

UNIT-4

LECTURE-2

Uncertainity, Information, and Entropy

Probabilistic experiment involves the observation of the output emitted by a discrete source during every unit of time. The source output is modeled as a discrete random variable, S,which takes on symbols form a fixed finite alphabet.

$$\mathcal{S} = s_0, s_1, s_2, \cdots, s_{k-1}$$

with probabilities

$$P(S = s_k) = p_k, k = 0, 1, \cdots, K-1$$

We assume that the symbols emitted by teh source during successive signaling intervals are statistically independent. A source having the jproperties jist described is called *discrete memoryless source*, memoryless in the sense that the symbol emitted at any time is independent of previous choices.

We can define the amount of information contained in each symbols.

$$I(s_k) = \log\left(\frac{1}{p_k}\right)$$

Here, generally use $\log_2$ since in digital communications we will be talking about bits. The above expression also tells us that when there is more uncertainity(less probability) of the symbol being occured then it conveys more information. Some properties of information are summarized here:

1. for certain event i.e, $p_k = 1$ the information it conveys is zero, $I(s_k) = 0$.
2. for the events $0 \le p_k \le 1$ the information is always $I(s_k) \ge 0$.
3. If for two events $p_k > p_i$, the information content is always $I(s_k) < I(s_i)$.
4. $I(s_k s_i) = I(s_k) + I(s_i)$ if s_k and s_i are statistically independent.

The amount of information $I(s_k)$ produced by the source during an arbitrary signalling interval depends on the symbol s_k emitted by the source at that time. Indeed, $I(s_k)$ is a discrete random variable that takes on the values $I(s_0), I(s_1), \cdots, I(s_{K-1})$ with probabilites $p_0, p_1, \cdots, p_{K-1}$ respectively. The mean of $I(s_k)$ over teh source alphabet S is given by

$$
\begin{aligned}
H(S) &= E[I(s_k)] \\
&= \sum_{k=0} p_k I(s_k) \\
&= \sum_{k=0} p_k \log_2\left(\frac{1}{p_k}\right)
\end{aligned}
$$

This important quantity is called entropy of a discrete memoryless source with source alphabet S. It is a measure of average information content per source symbol.

Some properties of Entropy

The entropy $H(\mathcal{S})$ of a discrete memoryless source is bounded as follows:

$$0 \leq H(\mathcal{S}) \leq log_2(K)$$

where K is the radix of the alphabet $\mathcal{S}$ of the source. Furthermore, we may make two statements:

1. $H(\mathcal{S}) = 0$, if and only if the probability $p_k = 1$ for some k, and the remaining probabilities in teh set are all zero; this lower bound on entropy corresponds to no uncertainity.
2. $H(\mathcal{S}) = log_2(K)$, if and only if $p_k = \frac{1}{K}$ for all k; this upper bound on entropy corresponds to maximum uncertainity.

Shannon Source Coding Theorem

An important problem in communication is the efficient representation of data generated by a discrete source. The process by which this representation is accomplished is called source encoding.

Our primary interest is in the development of an efficient source encoder that satisfies two functional requirements:

1. The code words produced by the encoder are in binary form.
2. The source code is uniquely decodable, so that the original source sequence can be reconstructed perfectly from the encoded binary sequence.

We define the average code word length, $\bar{L}$, of the source encoder as