Module 1

Introduction to Digital Communications and Information Theory

Lesson 3

Information Theoretic Approach to Digital Communications

After reading this lesson, you will learn about

- ***Scope of Information Theory***
- ***Measures of self and mutual information***
- ***Entropy and average mutual information***

The classical theory of information plays an important role in defining the principles as well as design of digital communication systems. Broadly, Information Theory provides us answer to questions such as:

a) What is information and how to quantify and measure it?
b) What kind of processing of information may be possible to facilitate transmission of information efficiently through a transmission medium?
c) What if at all, are the fundamental limits for such processing of information and how to design efficient schemes to facilitate transmission (or storage) of information?
d) How should devices be designed to approach these limits?

More specifically, knowledge of information theory helps us in efficient design of digital transmission schemes and also in appreciating scope and limits of the extent of improvement that may be possible.

In this Lesson, we introduce the concepts of 'information' the way they are used in Digital Communications. It is customary and convenient to define information through the basic concepts of a statistical experiment.

Let us consider a statistical experiment, having a number of outcomes, rather than a single outcome. Examples of several such experiments with multiple outcomes can be drawn from the theory of Digital Communications and a few are mentioned below.

Examples of random experiments with multiple outcomes

1. A signal source with finite and discrete number of possible output letters may be viewed as a statistical experiment wherein an outcome of interest may be 'sequences of source letters'.

2. Another example, frequently referred to, is that of a discrete input-discrete output communication channel where the input and output letters (or sequences) may be the outcomes of interest. We discuss more about such a channel later.

Joint Experiment

Let us consider a two-outcome experiment. Let, 'x' and 'y' denote the two outcomes where, 'x' denotes a selection from the set of alternatives

$$X = \{a_1, a_2, ...a_K\} \qquad 1.3.1$$

K is the number of elements in the set X. Similarly, let 'y ' denote a selection from the set of alternatives

$Y = \{b_1, b_2, ...b_J\}$, with 'J' being the number of elements in set Y.

Now, the set of pairs

$\{a_k, b_j\}$ $\quad 1 \le k \le K$ and $1 \le j \le J$, forms a joint sample space.

Let, the corresponding joint probability of $P(x = a_k \ \& \ y = b_j)$ over the joint sample space be denoted as,

$P_{XY}(a_k, b_j)$ where, $1 \le k \le K$ and $\quad 1 \le j \le J$ $\qquad$ 1.3.2

We are now in a position to define a *joint ensemble XY* that consists of the joint sample space and the probability measures of its elements $\{a_k, b_j\}$.

An *event* over the joint ensemble is defined as a subset of elements in the joint sample space.

Example: In XY joint ensemble, the event that x= a_k corresponds to the subset of pairs $\{(a_k, b_1); (a_k, b_2); .....(a_k, b_J)\}$. So, we can write an expression of $P_X(a_k)$ as:

$$P_X(a_k) = \sum_{j=1}^{J} P_{XY}(a_k, b_j), \text{ or in short, } P(x) = \sum_{y} P(x, y) \qquad 1.3.3$$

Conditional Probability

The conditional probability that the outcome 'y' is 'b_j' given that the outcome x is a_k is defined as

$$P_{Y|X}(b_j | a_k) = \frac{P_{XY}(a_k, b_j)}{P_X(a_k)}, \text{ assuming that, } P(a_k) > 0 \qquad 1.3.4$$

or in short, $\quad P(y|x) = \dfrac{p(x, y)}{p(x)}$

Two events x = a_k and y = b_j are said to be statistically independent if,

$$P_{XY}(a_k, b_j) = P_X(a_k).P_Y(b_j)$$

So in this case,

$$P_{Y|X}(b_j | a_k) = \frac{P_X(a_k).P_Y(b_j)}{P_X(a_k)} = P_Y(b_j) \qquad 1.3.5$$

Two ensembles X and Y will be statistically independent if and only if the above condition holds for each element in the joint sample space, i.e., for all j-s and k-s. So, two ensembles X and Y are statistically independent if all the pairs (a_k, b_j) are statistically independent in the joint sample space.

The above concept of joint ensemble can be easily extended to define more than two (many) outcomes.

We are now in a position to introduce the concept of 'information'. For the sake if generality, we first introduce the definition of 'mutual information'.

Mutual Information

Let, $X = \{a_1, a_2, ..., a_K\}$
$Y = \{b_1, b_2, ..., b_J\}$ and

in a joint ensemble with joint probability assignments $P_{XY}(a_k, b_j)$.

Now, the mutual information between the events $x=a_k$ and $y=b_j$ is defined as,

$$I_{X;Y}(a_k;b_j) \triangleq \log_2 \frac{P_{X|Y}(a_k|b_j)}{P_X(a_k)}, \text{ bits}$$

$$= \log_2 \frac{\text{a posteriori probability of 'x'}}{\text{a priori probability of 'x'}}, \text{ bits}$$

i.e, Mutual Information $= \log_2 \dfrac{\text{ratio of conditional probability of } a_k \text{ given by } b_j}{\text{probability of } a_k}$

1.3.6

This is the information provided about the event $x = a_k$ by the occurrence of the event $y = b_j$. Similarly it is easy to note that,

$$I_{Y;X}(b_j;a_k) = \log_2 \frac{P_{Y|X}(b_j|a_k)}{P_Y(b_j)} \qquad 1.3.7$$

Further, it is interesting to note that,

$$I_{Y;X}(b_j;a_k) = \log_2 \frac{P_{Y|X}(b_j|a_k)}{P_Y(b_j)} = \log_2 \frac{P_{XY}(a_k,b_j)}{P_X(a_k)} \cdot \frac{1}{P_Y(b_j)}$$

$$= \log_2 \frac{P_{X|Y}(a_k|b_j)}{P_X(a_k)} = I_{X;Y}(a_k;b_j)$$

$$\therefore I_{X;Y}(a_k;b_j) = I_{Y;X}(b_j;a_k) \qquad 1.3.8$$

i.e, the information obtained about a_k given b_j is same as the information that may be obtained about bj given that $x = a_k$.

Hence this parameter is known as 'mutual information'. Note that mutual information can be negative.

Self-Information

Consider once again the expression of mutual information over a joint ensemble:

$$I_{X;Y}(a_k;b_j) = \log_2 \frac{P_{X|Y}(a_k|b_j)}{P_X(a_k)} \text{ bits,} \qquad 1.3.9$$

Now, if $P_{X|Y}(a_k|b_j) = 1$ i.e., if the occurrence of the event $y=b_j$ certainly specifies the outcome of $x = a_k$ we see that,

$$I_{X;Y}(a_k;b_j) = I_X(a_k) = \log_2 \frac{1}{P_X(a_k)} \text{ bit} \qquad 1.3.10$$

This is defined as the self-information of the event $x = a_k$. It is always non-negative. So, we say that self-information is a special case of mutual information over a joint ensemble.

Entropy (average self-information)

The *entropy* of an ensemble is the average value of all self information over the ensemble and is given by

$$H(X) = \sum_{k=1}^{K} P_X(a_k) \log_2 \frac{1}{P_X(a_k)}$$

$$= -\sum_x p(x) \log p(x), \text{ bit} \qquad 1.3.11$$

To summarize, self-information of an event $x = a_k$ is

$$I_X(a_k) = \log_2 \frac{1}{P_X(a_K)} \text{ bit}$$

while the *entropy* of the ensemble X is

$$H(X) = \sum_{k=1}^{K} P_X(a_k) \log_2 \frac{1}{P_X(a_k)} \text{ bit.}$$

Average Mutual Information

The concept of averaging information over an ensemble is also applicable over a joint ensemble and the average information, thus obtained, is known as *Average Mutual Information:*

$$I(X;Y) \triangleq \sum_{k=1}^{K} \sum_{j=1}^{J} P_{XY}(a_k;b_j) \log_2 \frac{P_{X|Y}(a_k|b_j)}{P_X(a_k)} \text{ bit} \qquad 1.3.12$$

It is straightforward to verify that,

$$I(X;Y) = \sum_{k=1}^{K}\sum_{j=1}^{J} P_{XY}(a_k, b_j)\log_2 \frac{P_{X|Y}(a_k|b_j)}{p_X(a_k)}$$

$$= \sum_{k=1}^{K}\sum_{j=1}^{J} P_{XY}(a_k, b_j)\log_2 \frac{P_{Y|X}(b_j|a_k)}{p_Y(b_j)} = \text{I(Y;X)} \qquad 1.3.13$$

The unit of *Average Mutual Information* is bit. Average mutual information is not dependent on specific joint events and is a property of the joint ensemble. Let us consider an example elaborately to appreciate the significance of self and mutual information.

Example of a Binary Symmetric Channel (BSC)

Let us consider a channel input alphabet X = {a_1, a_2} and a channel output alphabet Y = { b_1, b_2}. Further, let $P(b_1|a_1) = P(b_2|a_2) = 1-\varepsilon$ and $P(b_1|a_2) = P(b_2|a_1) = \varepsilon$. This is an example of a binary channel as both the input and output alphabets have two elements each. Further, the channel is symmetric and unbiased in its behavior to the two possible input letters $a_1$and a_2. Note that the output alphabet Y need not necessarily be the same as the input alphabet in general.

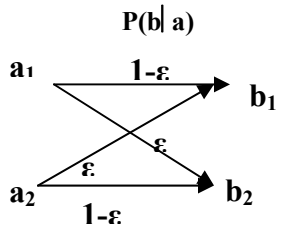

Fig.P.1.3.1 Representation of a binary symmetric channel; ε indicates the probability of an error during transmission.

Usually, if a_1 is transmitted through the channel, b_1 will be received at the output provided the channel has not caused any error. So, ε in our description represents the probability of the channel causing an error on an average to the transmitted letters.

Let us assume that the probabilities of occurrence of a_1 and a_2 are the same, i.e. $P_X(a_1) = P_X(a_2) = 0.5$. A source presenting finite varieties of letters with equal probability is known as a *discrete memory less source (DMS)*. Such a source is unbiased to any letter or symbol.

Now, observe that,

$$P_{XY}(a_1, b_1) = P_{XY}(a_2, b_2) = \frac{1-\in}{2}$$

And $P_{XY}(a_1, b_2) = P_{XY}(a_2, b_1) = \frac{\in}{2}$

Note that the output letters are also equally probable, i.e.

$$\sum_{j=1}^{2} P(y_j|x_i)p(x_i) = \frac{1}{2} \text{ for}$$

So, $P_{X|Y}(a_1|b_1) = P_{X|Y}(a_2|b_2) = 1-\epsilon$ and $P_{X|Y}(a_1|b_2) = P_{X|Y}(a_2|b_1) = \epsilon$

Therefore, possible mutual information are:

$I_{X;Y}(a_1;b_1) = \log_2 2(1-\epsilon) = I_{X;Y}(a_2;b_2)$ and $I_{X;Y}(a_1;b_2) = \log_2 2\epsilon = I_{X;Y}(a_2;b_1)$

Next, we determine an expression for the average mutual information of the joint ensemble XY:

$$I(X;Y) = (1-\epsilon)\log_2 2(1-\epsilon) + \epsilon \log_2 2\epsilon \text{ bit}$$

Some observations

(a) Case#1: Let, $\epsilon \rightarrow 0$. In this case,

$I_{X;Y}(a_1;b_1) = I_{X;Y}(a_2;b_2) = \log_2 2(1-\epsilon)$ bit = 1.0 bit (approx.)

$I_{X;Y}(a_1;b_2) = I_{X;Y}(a_2;b_1) = \log_2 2\epsilon$ bit $\rightarrow$ a large negative quantity

This is an almost noiseless channel. When a_1 is transmitted we receive b_1 at the output almost certainly.

(b) Case#2: Let us put $\epsilon = 0.5$. In this case,

$I_{X;Y}(a_1;b_1) = I_{X;Y}(a_1;b_2) = 0$ bit

It represents a very noisy channel where no information is received. The input and output are statistically independent, i.e. the received symbols do not help in assessing which symbols were transmitted.

(c) Case#3 : Fortunately, for most practical scenario, $0 < \epsilon << 0.5$ and usually,

$$I_{X;Y}(a_1;b_1) = I_{X;Y}(a_2;b_2) = \log_2 2(1-\epsilon) \text{ bit} = 1.0 \text{ bit (approx.)}$$

So, reception of the letter $y=b_1$ implies that it is highly likely that a_1 was transmitted. Typically, in a practical digital transmission system, $\epsilon \cong 10^{-3}$ or less.

The following figure, **Fig 1.3.1** shows the variation of I(X;Y) vs. ϵ.

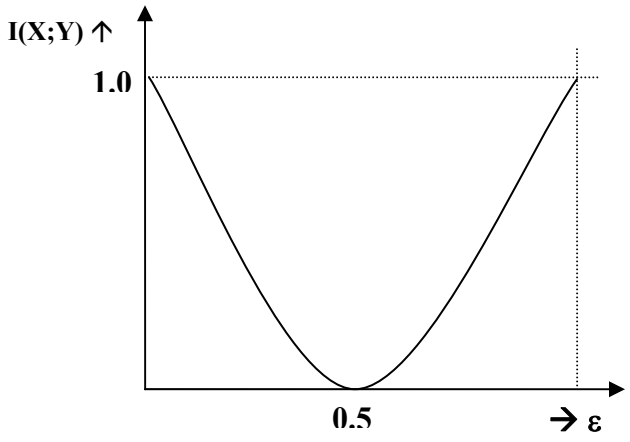

Fig. 1.3.1 *Variation of average mutual information vs. ε for a binary symmetric channel*

Problems

Q1.3.1) Why the theory of information is relevant for understanding the principles of digital communication systems?

Q1.3.2) Consider a random sequence of 16 binary digits where the probability of occurrence is 0.5. How much information is contained in this sequence?

Q1.3.3) A discrete memory less source has an alphabet {1,2,…,9}. If the probability of generating digit is inversely proportional to its value. Determine the entropy of source.

Q1.3.4) Describe a situation drawn from your experience where concept of mutual information may be useful.