**UNIT IV**

**Information Theory**

Contents

Information Theory:

* Discrete messages
* Concept of amount of information and its properties
* Average information
* Entropy and its properties
* Information rate
* Mutual information and its properties
* Illustrative Problems

Source Coding:

* Introduction
* Advantages
* Hartley Shannon’s theorem
* Bandwidth –S/N trade off
* Shanon- Fano coding,
* Huffman coding
* Illustrative Problems

Information Theory

Information theory deals with representation and the transfer of information.

There are two fundamentally different ways to transmit messages: via discrete signals and via continuous signals. ... For example, the letters of the English alphabet are commonly thought of as discrete signals.

Information sources

Definition:

The set of source symbols is called the source alphabet, and the elements of the set are called the symbols or letters.

The number of possible answers ‘ r ’ should be linked to “information.” “Information” should be additive in some sense. We define the following measure of information:



Where ‘ r ’ is the number of all possible outcome so far an do m message U.

Using this definition we can confirm that it has the wanted property of additivity:



The basis ‘b’ of the logarithm b is only a change of units without actually changing the amount of information it describes.

Classification of information sources

1. Discrete memory less.
2. Memory.

Discrete memory less source (DMS) can be characterized by “the list of the symbols, the probability assignment to these symbols, and the specification of the rate of generating these symbols by the source”.

1. Information should be proportion to the uncertainty of an outcome.
2. Information contained in independent outcome should add.

Scope of Information Theory

1. Determine the irreducible limit below which a signal cannot be compressed.
2. Deduce the ultimate transmission rate for reliable communication over a noisy channel.
3. Define Channel Capacity - the intrinsic ability of a channel to convey information.

The basic setup in Information Theory has:

– a source,

– a channel and

– destination.

The output from source is conveyed through the channel and received at the destination. The source is a random variable S

which takes symbols from a finite alphabet i.e.,

S = {s0, s1, s2, ・ ・ ・ , sk−1}

With probabilities

P(S = sk) = pk where k = 0, 1, 2, ・ ・ ・ , k − 1 and

k−1,Xk=0 ,pk = 1

The following assumptions are made about the source

1. Source generates symbols that are statistically independent.
2. Source is memory less i.e., the choice of present symbol does not depend on the previous choices.

Properties of Information

1. Information conveyed by a deterministic event is nothing
2. Information is always positive.
3. Information is never lost.
4. More information is conveyed by a less probable event than a more probable event

Entropy:

The Entropy (H(s)) of a source is defined as the average information generated by a discrete memory less source.

Information content of a symbol:

Let us consider a discrete memory less source (DMS) denoted by X and having the alphabet {U1, U2, U3, ……Um}. The information content of the symbol xi, denoted by I(xi) is

defined as



I (U) = log b

= - log b P(U)

Where P (U) is the probability of occurrence of symbol U

Units of I(xi):

For two important and one unimportant special cases of b it has been agreed to use the following names for these units:

b =2(log2): bit,

b = e (ln): nat (natural logarithm),

b =10(log10): Hartley.

The conversation of these units to other units is given as



log2a=

Uncertainty or Entropy (i.e Average information)

Definition:

In order to get the information content of the symbol, the flow information on the symbol can fluctuate widely because of randomness involved into the section of symbols.

The uncertainty or entropy of a discrete random variable (RV) ‘U’ is defined as

H(U)= E[I(u)]= 



Where PU (·) denotes the probability mass function (PMF) 2 of the RV U, and where the support of P U is defined as



We will usually neglect to mention “support” when we sum over PU (u) · logb PU (u), i.e., we implicitly assume that we exclude all u

With zero probability PU (u) =0.

Entropy for binary source

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



H (u) = - log2 - log2 = 1 b/symbol

Bounds on H (U)

If U has r possible values, then 0 ≤ H(U) ≤ log r,

Where

H(U)=0 if, and only if, PU(u)=1 for some u,

H(U)=log r if, and only if, PU(u)= 1/r ∀ u.

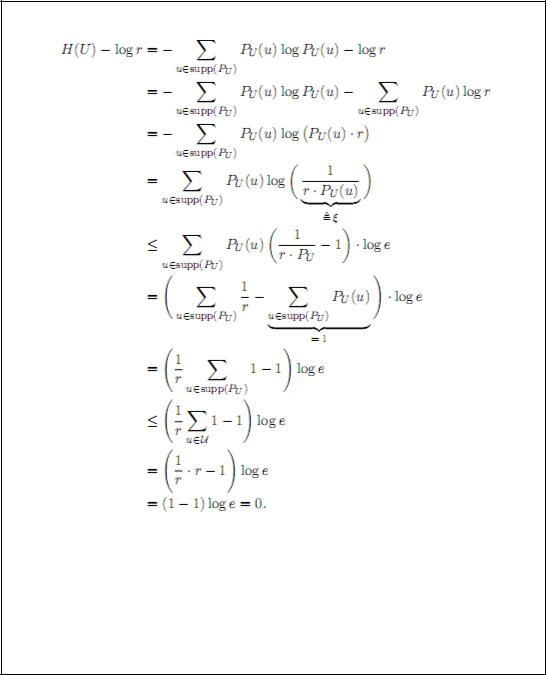
Hence, H(U) ≥ 0.Equalitycanonlybeachievedif −PU(u)log2 PU(u)=0



For all u ∈ supp (PU), i.e., PU (u) =1forall u ∈ supp (PU).

To derive the upper bound we use at rick that is quite common in.

Formation theory: We take the deference and try to show that it must be non positive.



Equality can only be achieved if

* 1. In the IT Inequality ξ =1,i.e.,if 1r·PU(u)=1=⇒ PU(u)= 1r ,for all u;

1. |supp (PU)| = r.

Note that if Condition1 is satisfied, Condition 2 is also satisfied.

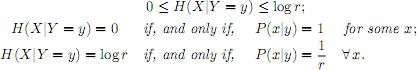
Conditional Entropy

Similar to probability of random vectors, there is nothing really new about conditional probabilities given that a particular event Y = y has occurred.

The conditional entropy or conditional uncertainty of the RV X given the event Y = y is defined as



Note that the definition is identical to before apart from that everything is conditioned on the event Y = y



Note that the conditional entropy given the event Y = y is a function of y. Since Y is also a RV, we can now average over all possible events Y = y according to the probabilities of each event. This will lead to the averaged.

Mutual Information

Although conditional entropy can tell us when two variables are completely independent, it is not an adequate measure of dependence. A small value for H(Y| X) may implies that X tells us a great deal about Y or that H(Y) is small to begin with. Thus, we measure dependence using *mutual information*:

I(X,Y) =H(Y)–H(Y|X)



Mutual information is a measure of the reduction of randomness of a variable given knowledge of another variable. Using properties of logarithms, we can derive several equiva-lent definitions

I(X,Y)=H(X)–H(X| Y)

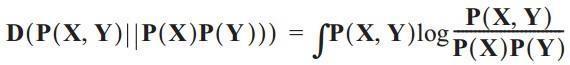
I(X,Y) = H(X)+H(Y)–H(X,Y) = I(Y,X)

In addition to the definitions above, it is useful to realize that mutual information is a particular case of the Kullback-Leibler divergence. The KL divergence is defined as:



KL divergence measures the difference between two distributions. It is sometimes called the relative entropy. It is always non-negative and zero only when p=q; however, it is not a distance because it is not symmetric.

In terms of KL divergence, mutual information is:



In other words, mutual information is a measure of the difference between the joint probability and product of the individual probabilities. These two distributions are equivalent only when X and Y are independent, and diverge as X and Y become more dependent.

**UNIT IV**

**Information Theory**

Contents

Information Theory:

* Discrete messages
* Concept of amount of information and its properties
* Average information
* Entropy and its properties
* Information rate
* Mutual information and its properties
* Illustrative Problems

Source Coding:

* Introduction
* Advantages
* Hartley Shannon’s theorem
* Bandwidth –S/N trade off
* Shanon- Fano coding,
* Huffman coding
* Illustrative Problems

Information Theory

Information theory deals with representation and the transfer of information.

There are two fundamentally different ways to transmit messages: via discrete signals and via continuous signals. ... For example, the letters of the English alphabet are commonly thought of as discrete signals.

Information sources

Definition:

The set of source symbols is called the source alphabet, and the elements of the set are called the symbols or letters.

The number of possible answers ‘ r ’ should be linked to “information.” “Information” should be additive in some sense. We define the following measure of information:



Where ‘ r ’ is the number of all possible outcome so far an do m message U.

Using this definition we can confirm that it has the wanted property of additivity:



The basis ‘b’ of the logarithm b is only a change of units without actually changing the amount of information it describes.

Classification of information sources

1. Discrete memory less.
2. Memory.

Discrete memory less source (DMS) can be characterized by “the list of the symbols, the probability assignment to these symbols, and the specification of the rate of generating these symbols by the source”.

1. Information should be proportion to the uncertainty of an outcome.
2. Information contained in independent outcome should add.

Scope of Information Theory

1. Determine the irreducible limit below which a signal cannot be compressed.
2. Deduce the ultimate transmission rate for reliable communication over a noisy channel.
3. Define Channel Capacity - the intrinsic ability of a channel to convey information.

The basic setup in Information Theory has:

– a source,

– a channel and

– destination.

The output from source is conveyed through the channel and received at the destination. The source is a random variable S

which takes symbols from a finite alphabet i.e.,

S = {s0, s1, s2, ・ ・ ・ , sk−1}

With probabilities

P(S = sk) = pk where k = 0, 1, 2, ・ ・ ・ , k − 1 and

k−1,Xk=0 ,pk = 1

The following assumptions are made about the source

1. Source generates symbols that are statistically independent.
2. Source is memory less i.e., the choice of present symbol does not depend on the previous choices.

Properties of Information

1. Information conveyed by a deterministic event is nothing
2. Information is always positive.
3. Information is never lost.
4. More information is conveyed by a less probable event than a more probable event

Entropy:

The Entropy (H(s)) of a source is defined as the average information generated by a discrete memory less source.

Information content of a symbol:

Let us consider a discrete memory less source (DMS) denoted by X and having the alphabet {U1, U2, U3, ……Um}. The information content of the symbol xi, denoted by I(xi) is

defined as



I (U) = log b

= - log b P(U)

Where P (U) is the probability of occurrence of symbol U

Units of I(xi):

For two important and one unimportant special cases of b it has been agreed to use the following names for these units:

b =2(log2): bit,

b = e (ln): nat (natural logarithm),

b =10(log10): Hartley.

The conversation of these units to other units is given as



log2a=

Uncertainty or Entropy (i.e Average information)

Definition:

In order to get the information content of the symbol, the flow information on the symbol can fluctuate widely because of randomness involved into the section of symbols.

The uncertainty or entropy of a discrete random variable (RV) ‘U’ is defined as

H(U)= E[I(u)]= 



Where PU (·) denotes the probability mass function (PMF) 2 of the RV U, and where the support of P U is defined as



We will usually neglect to mention “support” when we sum over PU (u) · logb PU (u), i.e., we implicitly assume that we exclude all u

With zero probability PU (u) =0.

Entropy for binary source

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



H (u) = - log2 - log2 = 1 b/symbol

Bounds on H (U)

If U has r possible values, then 0 ≤ H(U) ≤ log r,

Where

H(U)=0 if, and only if, PU(u)=1 for some u,

H(U)=log r if, and only if, PU(u)= 1/r ∀ u.

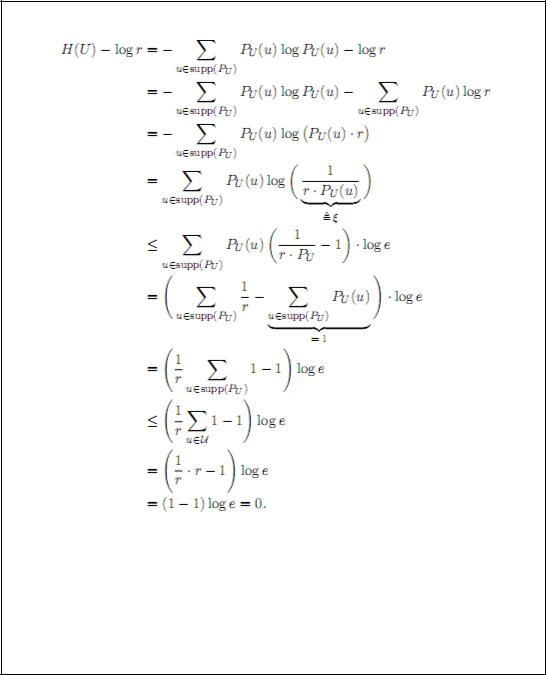
Hence, H(U) ≥ 0.Equalitycanonlybeachievedif −PU(u)log2 PU(u)=0



For all u ∈ supp (PU), i.e., PU (u) =1forall u ∈ supp (PU).

To derive the upper bound we use at rick that is quite common in.

Formation theory: We take the deference and try to show that it must be non positive.



Equality can only be achieved if

* 1. In the IT Inequality ξ =1,i.e.,if 1r·PU(u)=1=⇒ PU(u)= 1r ,for all u;

1. |supp (PU)| = r.

Note that if Condition1 is satisfied, Condition 2 is also satisfied.

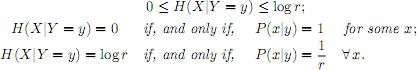
Conditional Entropy

Similar to probability of random vectors, there is nothing really new about conditional probabilities given that a particular event Y = y has occurred.

The conditional entropy or conditional uncertainty of the RV X given the event Y = y is defined as



Note that the definition is identical to before apart from that everything is conditioned on the event Y = y



Note that the conditional entropy given the event Y = y is a function of y. Since Y is also a RV, we can now average over all possible events Y = y according to the probabilities of each event. This will lead to the averaged.

Mutual Information

Although conditional entropy can tell us when two variables are completely independent, it is not an adequate measure of dependence. A small value for H(Y| X) may implies that X tells us a great deal about Y or that H(Y) is small to begin with. Thus, we measure dependence using *mutual information*:

I(X,Y) =H(Y)–H(Y|X)



Mutual information is a measure of the reduction of randomness of a variable given knowledge of another variable. Using properties of logarithms, we can derive several equiva-lent definitions

I(X,Y)=H(X)–H(X| Y)

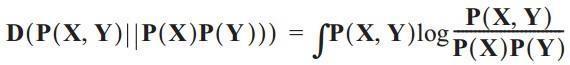
I(X,Y) = H(X)+H(Y)–H(X,Y) = I(Y,X)

In addition to the definitions above, it is useful to realize that mutual information is a particular case of the Kullback-Leibler divergence. The KL divergence is defined as:



KL divergence measures the difference between two distributions. It is sometimes called the relative entropy. It is always non-negative and zero only when p=q; however, it is not a distance because it is not symmetric.

In terms of KL divergence, mutual information is:



In other words, mutual information is a measure of the difference between the joint probability and product of the individual probabilities. These two distributions are equivalent only when X and Y are independent, and diverge as X and Y become more dependent.