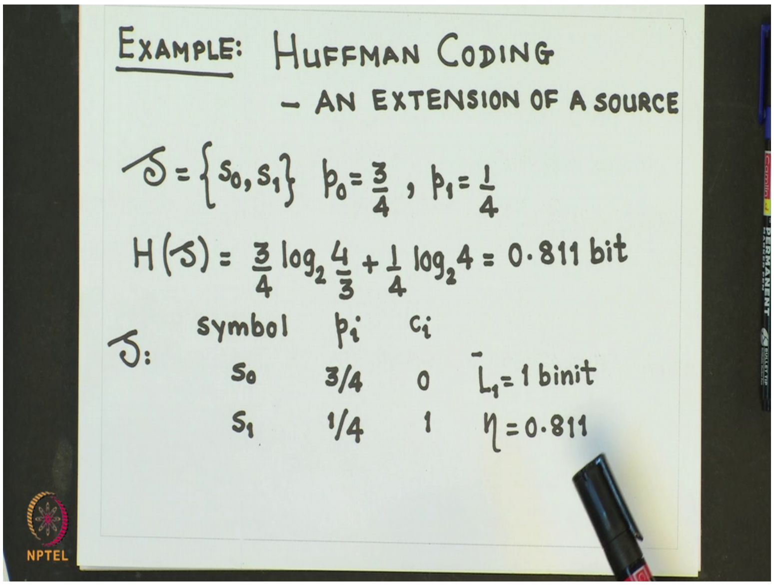
Principles of Digital Communications
Prof. Shabbir N. Merchant
Department of Electrical Engineering
Indian Institute of Technology, Bombay

Lecture – 06
Discrete Memoryless Channels: Mutual Information

In the last class, we studied Huffman coding. We will see the application of Huffman coding to an extension of a source to improve the efficiency of source encoder. So, let us take a example.

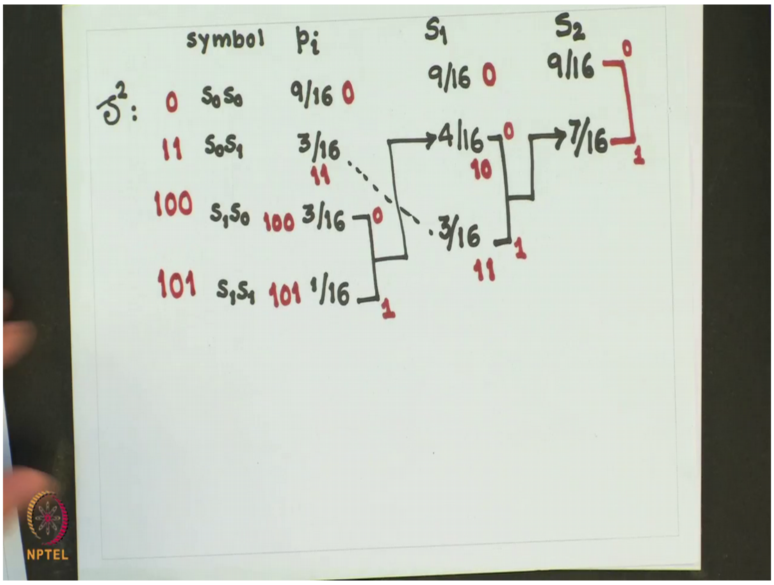
(Refer Slide Time: 00:38)



I have a source with two symbols with the given probabilities as shown here and if we calculate the entropy of this source it turns out to be 0.811 bit. Now, if you were to find a compact code for this source, it is a very trivial, I have two symbols immediately I know that compact code base they consist of two codewords 0 and 1.

So, I require 1 binit and the efficiency of the source encoder in this case would be 0.811. Let us try to improve the efficiency of this source encoder by deploying second extension of the source.

(Refer Slide Time: 01:30)



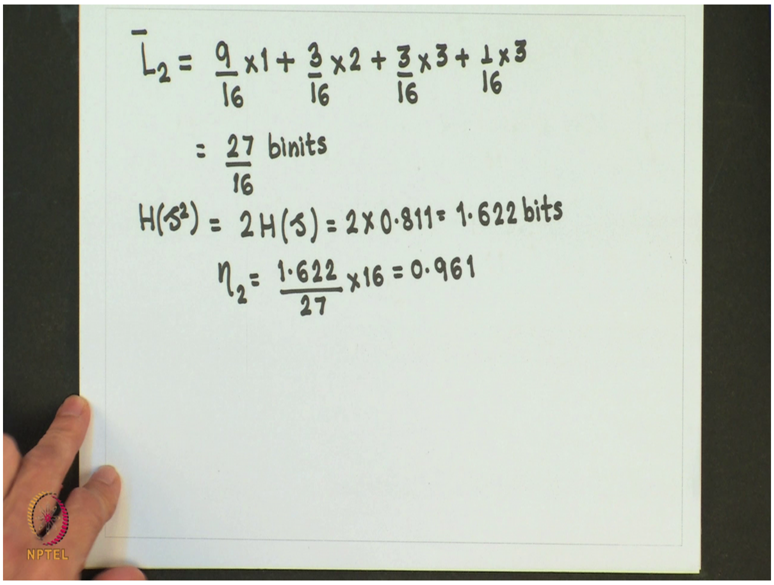
So, if I take the second extension of the source I will have four symbols. Assuming the source to be discrete memoryless source; I can compute the probabilities of the new symbols in the second extension this I have been the state here. Let us deploy Huffman coding for this and get the codeword for this.

So, I will combine this two, this already arrange in a descending order of probabilities I combine this. So, I will get 4 by 16. So, the next reduce source which I call it as S 1 will have three symbols, this will be 4 by 16 and this will be 3 by 16. So, I reduce it once more and then I get S 2 as the reduction, this is 9 by 16 and this will be 7 by 16.

And, now let us assign the labels. So, I start with the last reduction which has only two symbols. So, I give 0, 1 here I will assign 0, 1 arbitrarily here also I will assign 0, 1. So, this symbol in this S 1 will have the same codeword as what is here 9 by 16. So, this will be 0.

This will have you take the codeword 1 and the codeword for this will be 1 0, 1 1 I add 0 and 1 to the codeword 1, then this will go as it is here 1 1 this will also go as it is here 0, then this basically 1, 0. So, I will add 0 to this and 1 to this. So, I get the codeword for this as 0, 1 1, 1 0 0 and 1 0 1, correct.

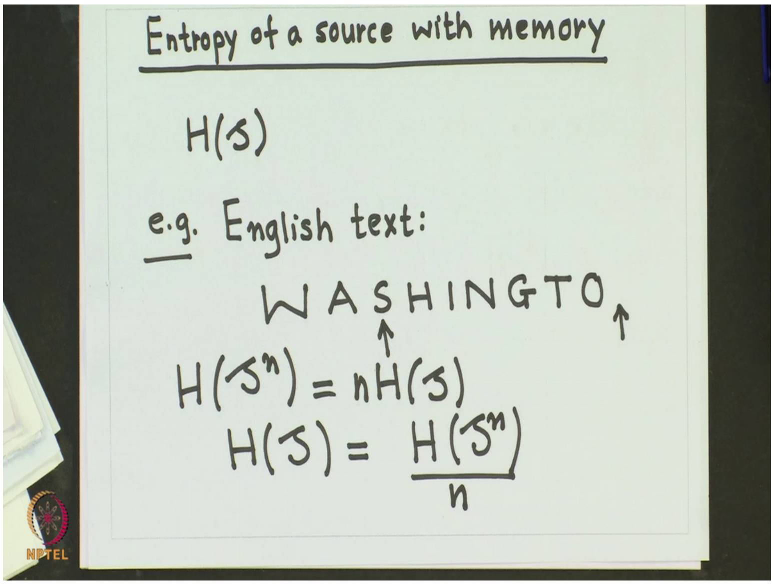
(Refer Slide Time: 04:19)



And, if we compute the length for this I have shown it here you get 27 by 16 binits and if we calculate the entropy of the second extension which is nothing, but twice the entropy of the original source turns out to be 1.622 bits. So, the efficiency turns out to be 0.961 which is an improvement over the earlier figure of 0.811, ok.

So, we can go for higher extensions and try to improve the efficiency of this encoder using the Huffman coding algorithm, ok. Now, so far we have restricted ourselves to information sources which are memoryless. Now, for the sake of completion let me discuss the calculation of entropy for a source with memory.

(Refer Slide Time: 05:30)



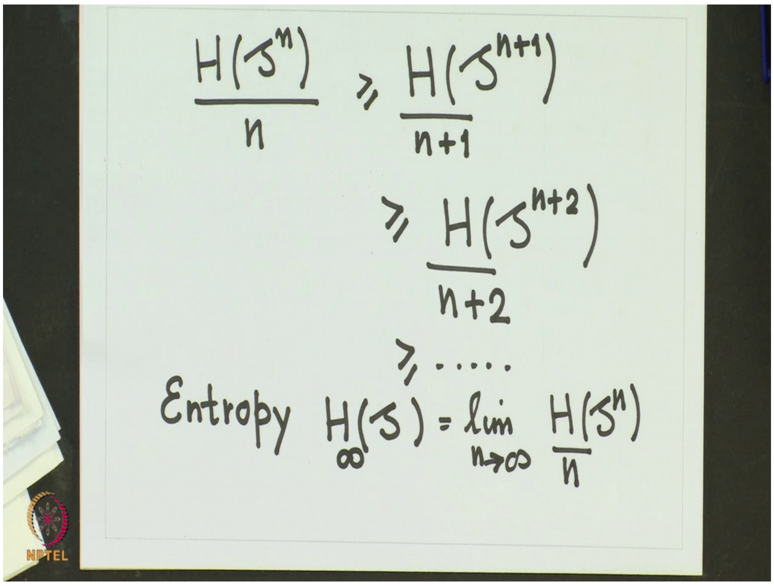
We know that entropy of a discrete memoryless source is average information gain on the occurrence of a symbol from the source and this entropy or the average information is same as the average amount of uncertainty, we have before the occurrence of a symbol.

So, now it is intuitively satisfying to say that this average amount of uncertainty before the occurrence of the symbol will decrease with the number of past observe symbol. Let me take an example of English text and let us assume that the output was this. Now, the amount of uncertainty which I have at this position is; obviously, less than the amount of uncertainty which I have at this position, correct. So, at this position I have observe W and A and at this position I have observed all this x letters. So, I know that most likely here it is going to be N, correct.

So, it is intuitively correct to say that amount of uncertainty decreases with the number of past observe samples. We have also seen that if you take a discrete memoryless source and if you take it n-th extension, then the entropy of the n-th extension is equal to n times the entropy of the original source. So, what does mean that if I were to calculate the entropy of the original discrete memoryless source then I can also obtain it by calculating the entropy of the n-th extension of the source and then dividing it by n. So, this is the entropy per symbol for the n symbol block.

Now, instead of if this source was not memoryless then also I can calculate the entropy of n block of symbols which are being emitted from this source with memory, ok.

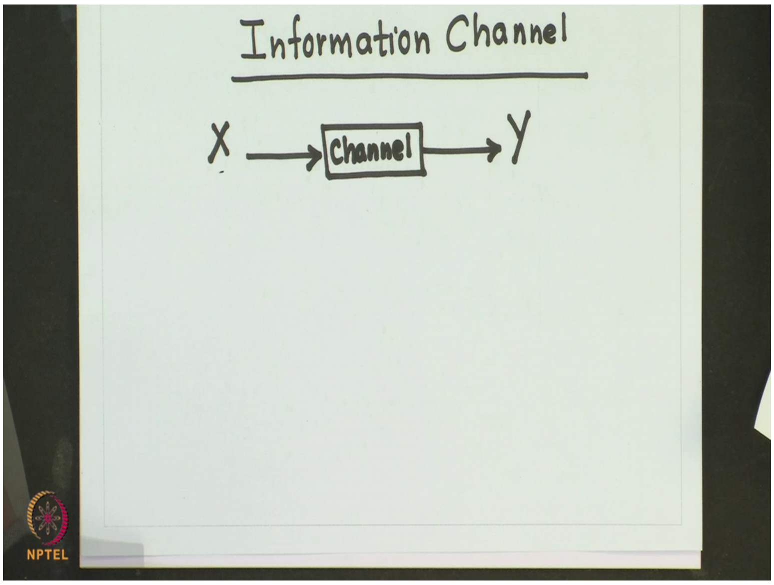
(Refer Slide Time: 08:47)



So, if we calculate the entropy of the nth extension of the source with memory and divided by n, then it is intuitively satisfying and we can also prove it more rigorously mathematically that this quantity should be larger than or equal to H of s n plus 1 divided by n plus 1 and this should be larger than H of s n plus 2 divide by n plus 2, correct.

So, the entropy of a source with memory is defined as. So, we can show that this limit exists. So, far we have discussed the measure of information and also examined the mathematical definition for an information source. The next question which I would like to address is information transmission, ok.

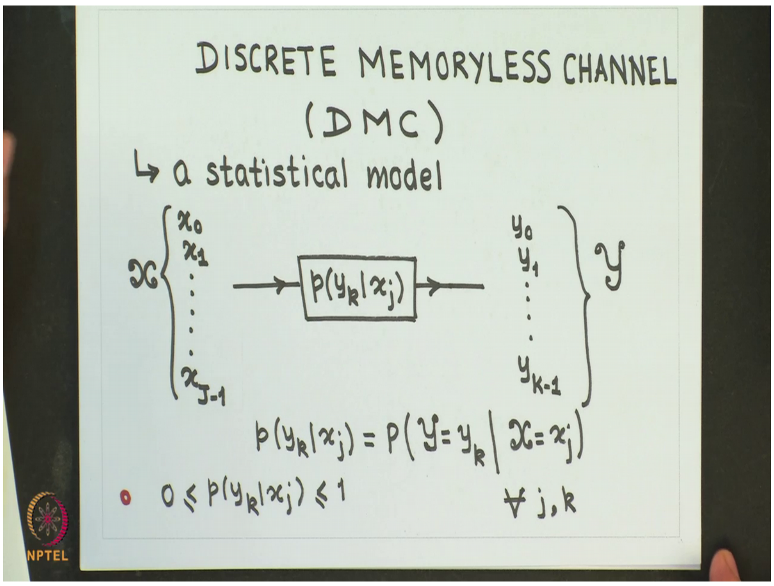
(Refer Slide Time: 10:43)



So, we will start with the definition of information channel. So, the formal way to define would be that information channel or communication channel has a input to it and output to it let me call this input as X and output as Y and this is my communication channel.

So, every unit of time the channel accepts an input selected from an input alphabet and it and in response it emits an output Y which is an output symbol from an output alphabet, correct.

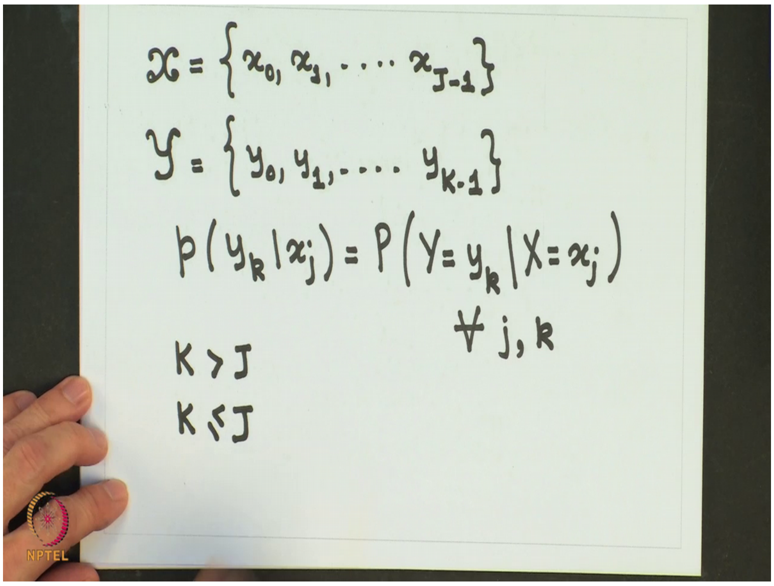
(Refer Slide Time: 12:08)



So, we can formally say that a channel is as shown here and we start with a discrete memoryless channel, the counterpart of discrete memoryless source. So, we define what is known as discrete memoryless channel as a statistical model which has a input which is a symbol from input alphabet x, is a has an output which is a symbol from output alphabet y and there is conditional probabilities associated between the output and the input.

Now, we say the channel is memoryless when the output of the channel at a particular time depends only on the input symbol at that time, correct. So, this is basically the definition for a memoryless channel. So, these are the conditional probabilities which gets specified for a memoryless channel, for all j and k and by the law of probability this condition is always satisfied.

(Refer Slide Time: 13:50)

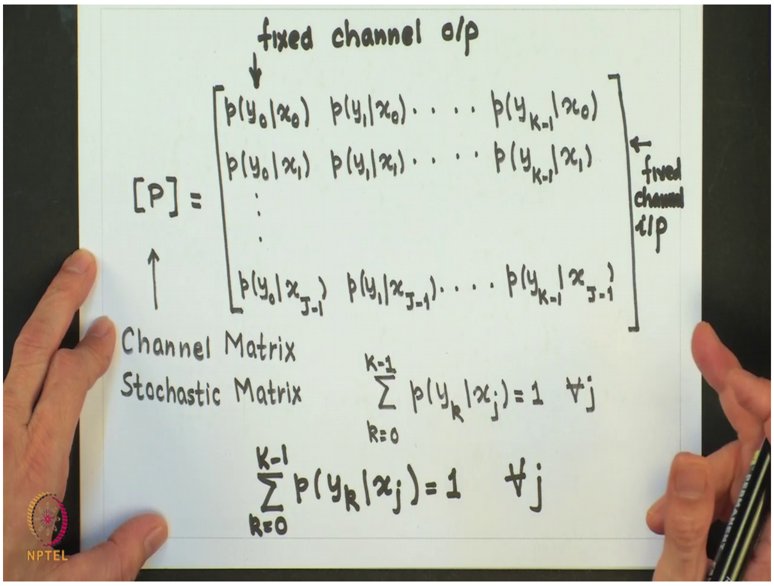


So, we see that basically that we have an input alphabet which I call it as curly x composed of symbols the cardinality of this alphabet is j; that means, there are j symbols and we have the output alphabet which I call it as curly y is composed of y 0, y 1 up to y k minus 1.

So, the cardinality of output alphabet is k and we have the set of transition probabilities y k given x j which is nothing, but probability of output symbol being equal to y k given the input symbol is x j, for all j and k, correct.

Now, it is important to note that j and k are not same, correct. So, it is possible that K is greater than J, for channel coding or K could be less than or equal to J. This condition will happen if two input symbols get mapped to the same output symbols, correct, ok.

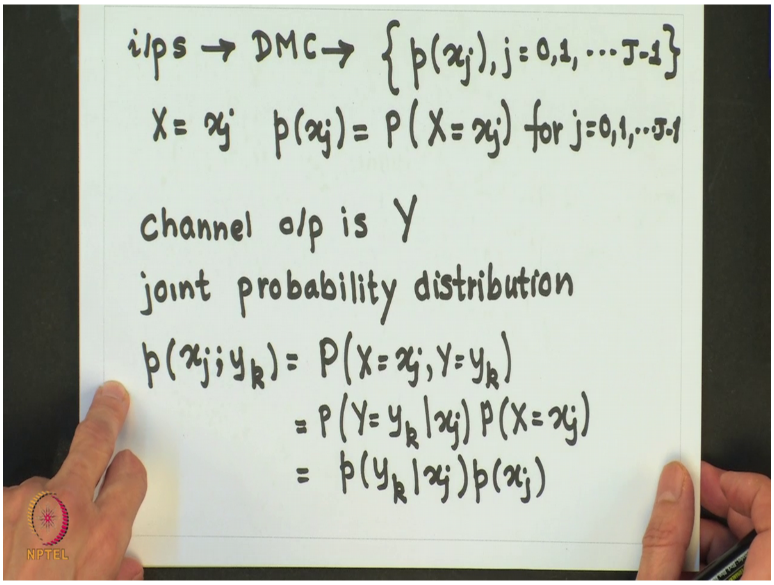
(Refer Slide Time: 15:30)



So, a discrete memoryless channel can be summarized in the form of a matrix. The matrix is shown here. If you look at the elements of the matrix they are given here and this matrix is known has a channel matrix or stochastic matrix. If you look at any of the input row here, then this is a fixed channel input, correct and if you look at any of the column this is the fixed channel output, correct.

And, because this is a fixed channel input it is not difficult to see that this condition will be always valid. Probability of y k given x j for k equal to summation k equal to 0 to capital K minus 1, will be equal to 1, for all j. What it says that, some output will occur for a given input x j, ok.

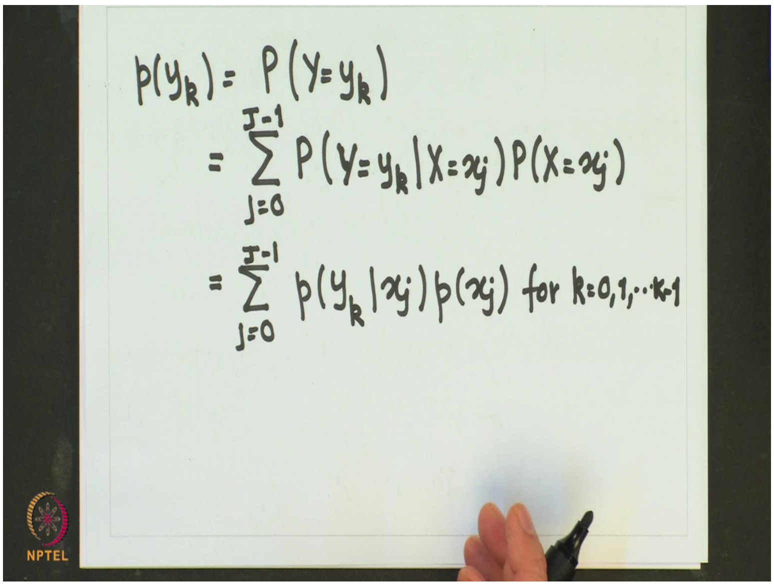
(Refer Slide Time: 17:20)



So, the input to a discrete memoryless channel are selected according to the probabilities p x j; j is equal to 0, 1 up to j minus 1. So, what it means that the channel input X is equal to x j occurs with the probability p x j that is same as probability of the this event occurring. This is known as a priori probability of x j.

Now, channel output is say Y then we can define the joint probability distribution that is probability x j, y k as probability of X is equal to x j, Y is equal to y k and this can be also written by base rule as.

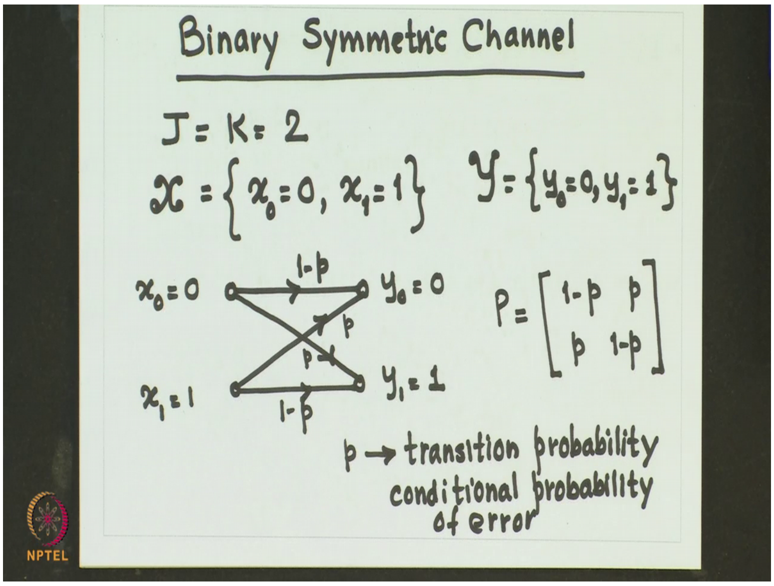
(Refer Slide Time: 19:45)



So, from this we can get the marginally probability distribution of the output as follows j is equal to 0 to J minus 1 K minus 1.

So, let us take a simple case of a binary symmetric channel.

(Refer Slide Time: 21:00)

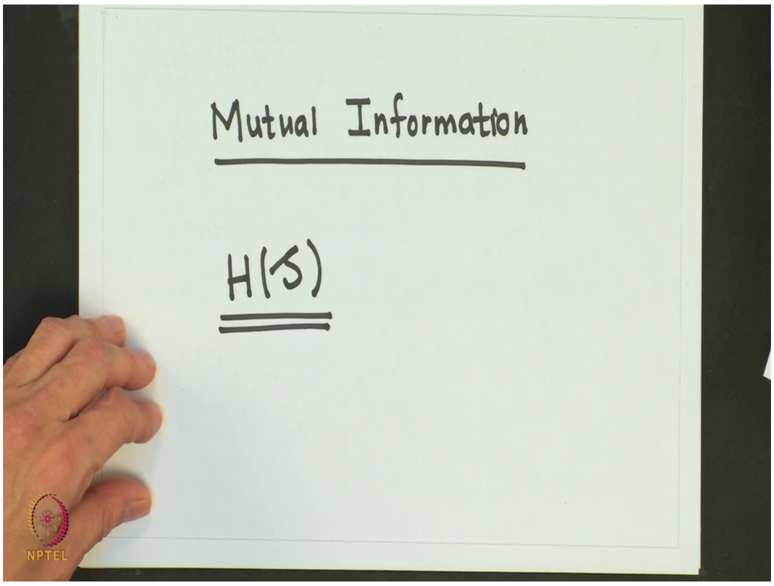


Like we had considered a binary discrete memoryless source, so, we have an equivalent of that. So, we have here J is equal to K is equal to 2, your input alphabet is x 0 is equal to 0 and x 1 is equal to 1 and your output alphabet is equal to y 0 is equal to 0 and y 1 is equal to 1 and this binary symmetric channel can also be shown by this diagram.

So, this tells you basically what is the probability of y 0 equal to 0, given x 0 is equal to 0 is this probability is given by 1 minus p, correct. So, if you look for the channel matrix for this, this will be of the form as shown here. This p is basically is a transition probability or it is also known as conditional probability of error. So, in this diagram both these branches are same p and that is why it is known as a symmetric channel.

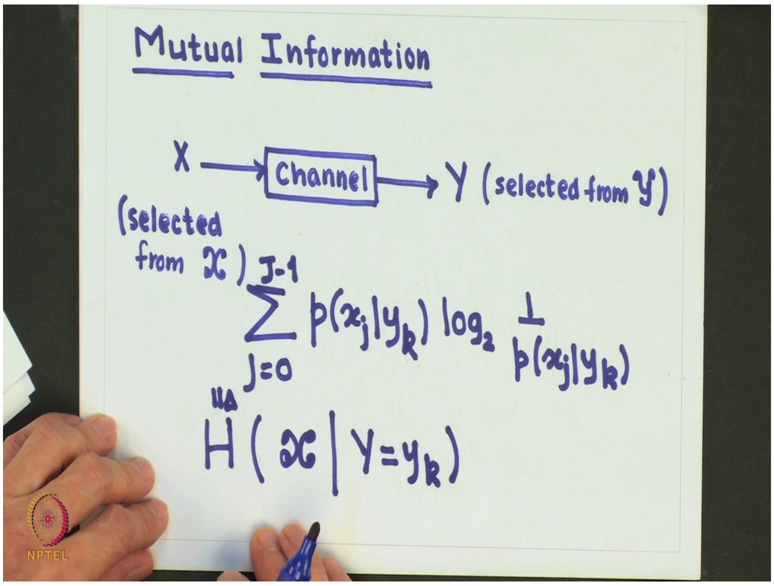
Having define a discrete memoryless channel next question is that the purpose of a communication channel is basically to know the transmitted symbol after observing the output symbol from a channel. So, before you transmitters symbol there is some kind of uncertainty above the transmission. We do not know this symbol is going to be transmitted, that symbol gets transmitted on the channel and then you observe the output of the channel. Now, by looking at the output of the channel ok, I am supposed to know what was the transmission which took place on that channel, ok.

(Refer Slide Time: 24:46)



So, this is the concept of mutual information. Basically, it tells us what is the uncertainty of the input symbol after I have observed the output symbol. So, remember before the transmission starts, the uncertainty about the input symbol is given by this quantity that is the entropy of the source and now, I am interested to calculate what is this uncertainty after I have observed the output.

(Refer Slide Time: 25:36)



So, I have input to the channel X and output as Y. So, based on the reasoning for the definition of information measure I could say that the uncertainty which I have about the input symbol given that I have observed a particular output symbol should be of the form log to the base 2 1 by probability of x j given y k.

So, this is the amount of uncertainty I have about a particular symbol in the input alphabet given I have observed a particular symbol y k. So, now, what would be the average uncertainty about the input symbol given that I have observed the output symbol y k and this can be obtained easily by this summation. So, this will be the average uncertainty I will have about the input symbol given that I observe the output symbol y k. So, this I call it as by definition this is equal to by definition the conditional entropy of x given Y is equal to y k correct.

Now, this quantity itself is a random variable and it will take different values, so, in principle we should able to calculate the average value of this and this will discuss in the next class.

Thank you.