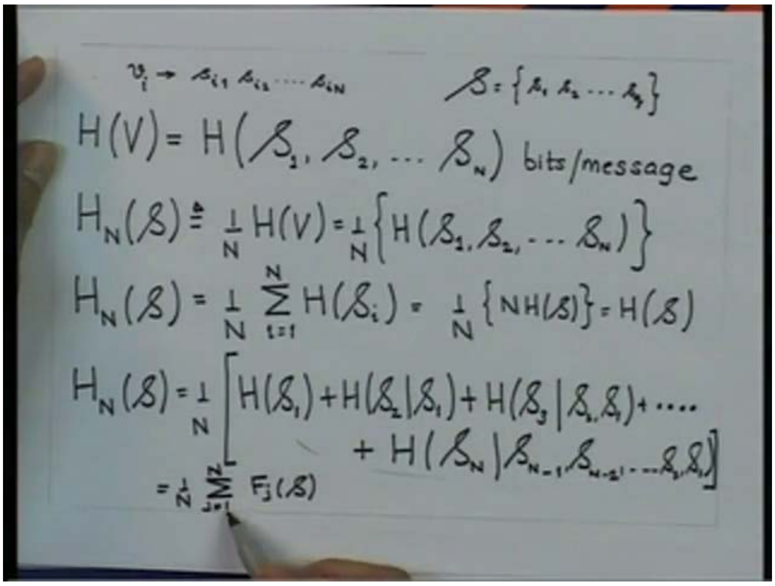
Information Theory and Coding
Prof. S. N. Merchant
Department of Electrical Engineering
Indian Institute of Technology, Bombay

Lecture - 6
Asymptotic Properties of Entropy and Problem Solving in Entropy

In the last class, we looked at the procedure to evaluate information of a Markov source with arbitrary memory based on conditional information measures. Today, we will look at the evaluation of information of a Markov source with the arbitrary memory, based on joint information. In order to do that, let me take a simple example. Let us assume that I have a message V i form of symbols of length n.

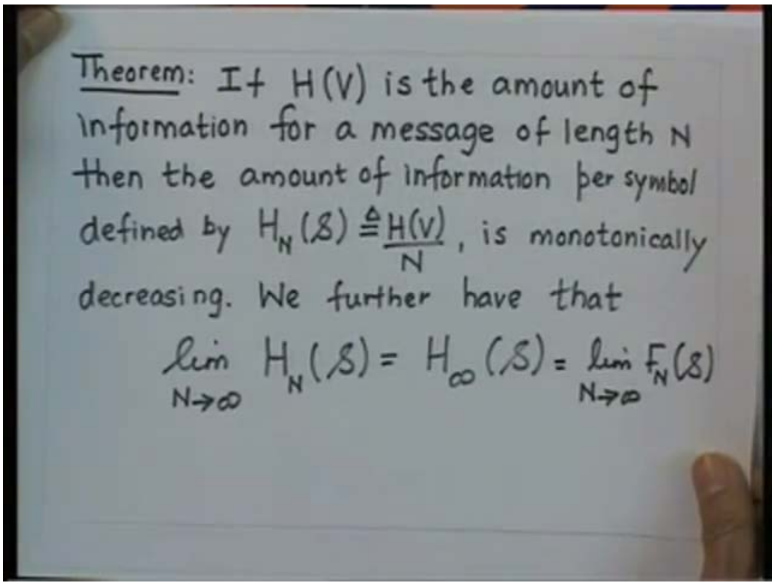
(Refer Slide Time: 01:18)



Previously, we had studied messages of length one only. So, let us consider messages of length N; this s i 1, s i 2, s i N each of this can come from this Markov source from this arbitrary memory, whose alphabet is from s 1 to s q. Now, I am interested in evaluating the information, which is there in this message given by V i. So, on an average the information in this message of length N can be evaluated by this expression out here, this is will be bits per message. Now, if I am interested in finding out the information per symbol than H N s is defined as 1 by N of H V. So, we can say this is by definition and this is nothing but equal to 1 by N joint information.

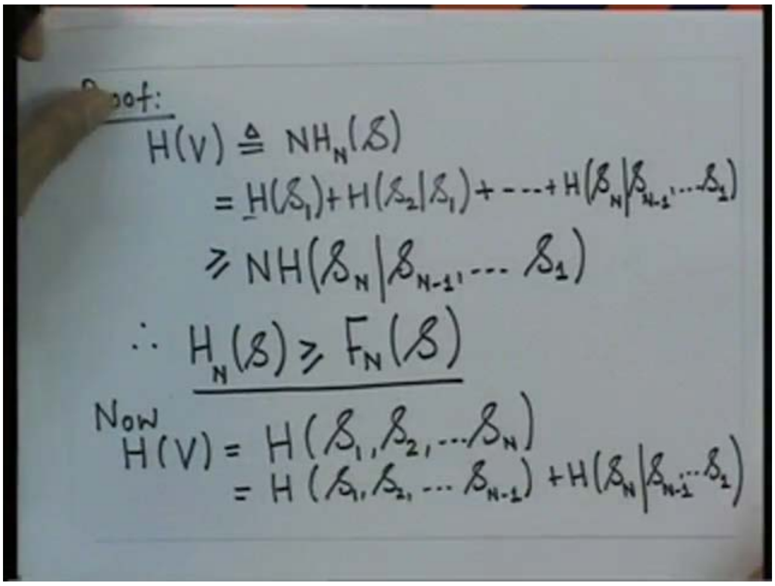
Now, we are also seeing last time that I can write H N s is equal to this quantity. If I assume that all the symbols, which consider a message of length N are independent. In that case H N s turns out to be same as H s, but if the symbols in these messages are not independent, then I can write a very generic expression for H N s, which is nothing but an average information per symbol. In that message will be given by this expression out here. That is 1 by N j summation j equal to 1 to N, F j s, this we had seen last time. We are also seeing that F N s is a monotony decreasing function of N. Today what we will try to prove first is that H N s also is a monotonically decreasing function of N. In order to do that let me state the theorem first.

(Refer Slide Time: 03:51)



The theorem says if H of V is the amount of formation for a message of length N. Then the amount of information per symbol defined by H N s is equal to s V by N is monotonically decreasing. This is the first result, which will try to prove. We further have that limit of N tending to infinity of H N s will be equal to H infinity s H infinity was the entropy, which we defined for a Markov source. This is nothing but equal to limit of N tending to infinity of F N s, so will show that both the limits are equal to H infinity s.

(Refer Slide Time: 06:34)

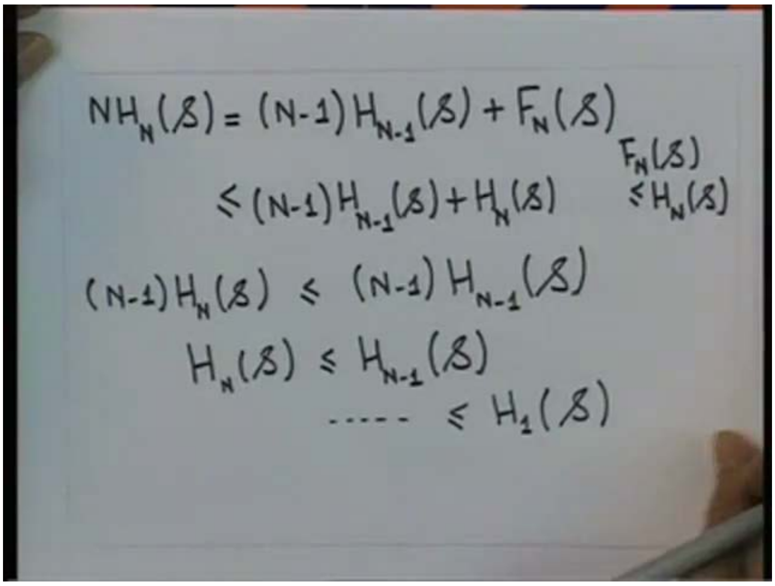


So, let us try to prove this theorem H V by definition is equal to N times H N s. So, this is equal to H of s 1 plus H of s 2 given s 1 plus H of s N given s N minus 1 to s 1. Now, we have seen last time that each of this quantity out here is related by this relationship. The additional information, which I get at is less than additional information, which I get at N minus 1. This way I can continue up to the time instant one.

So, based on this relationship, which we derived last time, which is equivalent to F N s less that equal to F n minus s, which we have proved, that this is monotony decreasing function of N and limit of this is equal to H infinity s. That was the final on top of the Markov source based on this relationship out here. I can write that this expression out here is greater than or equal to N times H of s N s N minus 1 up to s 1. Therefore, H N s, which is nothing but H V divide by N is greater than or equal to H of s N given s N minus 1 s 1. That is nothing but by definition F n s, this we had seen this definition of F N s is equal to this we had seen last time.

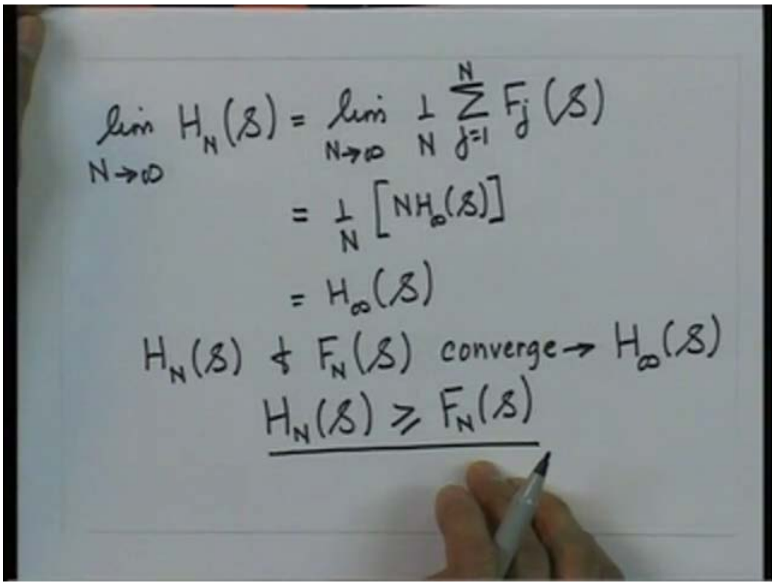
Now, H of V is equal to H of s 1 s 2 up to s N this entropy joint entropy can be broken up as H of s 1 s tow N minus 1 joint information, which I have from N minus 1 symbols plus the additional information, which I get from N symbol. So, this relationship can be written like this. Now, from this relationship and this relationship I can write N of H N s is equal to N minus 1 H of N minus 1 s plus F N s.

(Refer Slide Time: 10:00)



Now, we are just seeing that F N s is always less than F N s is always less than equal to H N s. So, from this I can write N minus 1 H N minus 1 s plus H N s. Now, just trying to simplify this expression will get N minus 1 H N s is less than equal to N minus 1 H of N minus 1 s, which implies H N s is less than equal to H my H N minus 1 this. So, I can proceed like this and show that this is less than equal to H 1 s. Now, since H N s is always greater than equal to 0 and H N s monotonic decreases with N. This follows that this quantity should converge to a limit as N tends to infinity. Let us try to find out what is the value of H N s as N tends to infinity. So, limit of N tending to infinity of H N s would be limit of N tending to infinity of 1 by N summation F j s, j is equal to 1 to N.

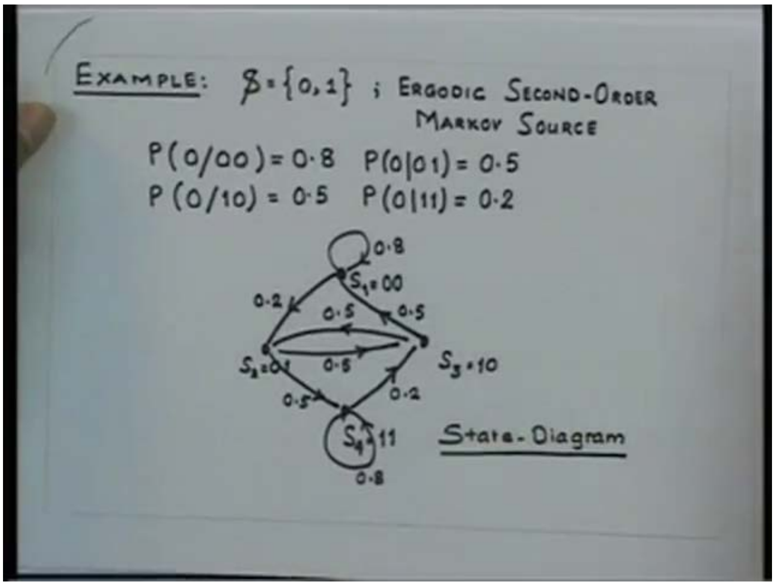
(Refer Slide Time: 12:04)



Now, we know that as N tends to infinity F j s because in monotonically decreasing function of N converges to H infinity s. Therefore, this is nothing but 1 by N n times H infinity s and this is equal to H infinity s. So, what we have proved that both H N s and F N s converge to the same limit, which is given by H infinity s 1. This is nothing but the entropy of a Markov source with arbitrary memory, but this condition is also valid, which we proved H N s is always less than greater than equal to F N s.

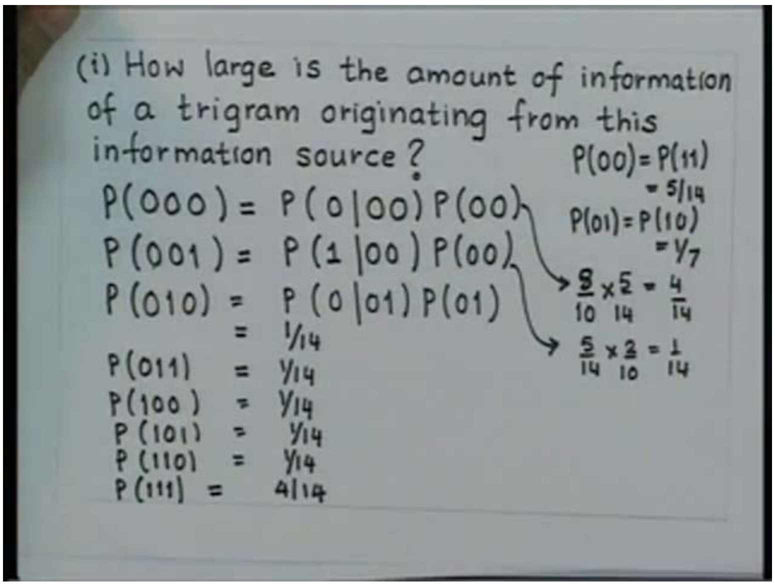
What is implies that H N s is a worst approximation to the actual amount of information H infinity s, but the advantage of H N s is its simplicity in evaluation. So, with this result let us proceed ahead and solve an example, which will help us to understand, appreciate the material, which we have covered so far. So, I will try to explain whatever we have done so far, in relation to a Markov process with the help of an example. Let us assume I have an example given here, I have a binary source 0 and 1.

(Refer Slide Time: 14:35)



This binary source is second order Markov source. This conditional probability is have been specified probability of 0 given given 0 is 0.8 another transitional properties have been specified. We had a look at this example earlier too. Now, the first thing is to draw a straight diagram for this Ergodic second-order Markov source. To that the straight diagram for that is given as follows, we are also had a look at this said diagram earlier in the course of our lecture.

(Refer Slide Time: 15:26)



What I mean by this statement is, that if I look at the messages, which are being form this source as messages of length 3. Then what is the amount of information, which is contained in that message. So, let us try to evaluate this in order to evaluate this what I have to is basically to calculate the amount of information per trigram, the probabilities of a trigrams are determined first.

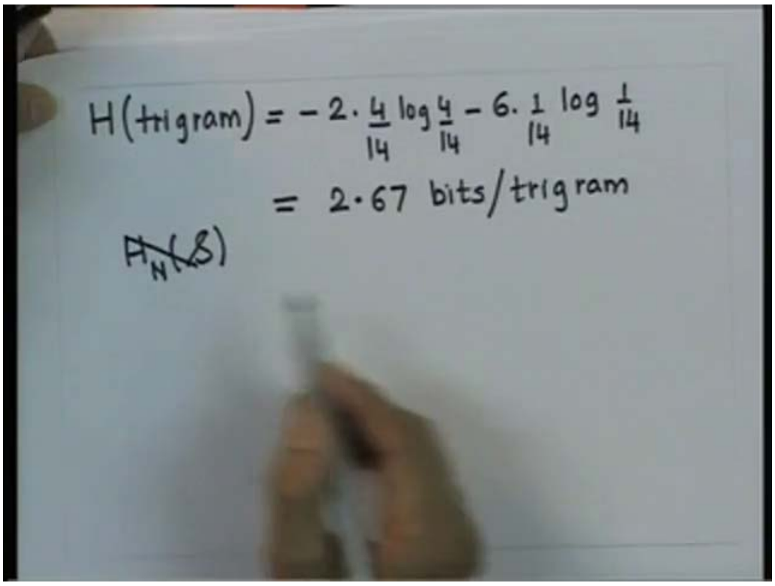
Now, in order to this what I have to find out is probability of 0 0 0. This is nothing but probability of emitting 0 given i was in the state 0 0. That is the probability of being in the state 0 0. Now, we had earlier calculated the probabilities of the states as probability of 0 0 equal to probability of 0 1 is equal to 5 by 14. Probability of 0 1 equal to probability of 1 0 is equal to 1 by 7. We have had a look at the procedure to calculate. This probability is assuming that this probability having calculate earlier. Let us try to calculate the probability of the messages of length 3.

So, the different combinations, which I can have of the messages of length 3 would be like this. So, first one is probability of 0 0 would be given by this next would be probability of 0 0 1. This would be given by what is the probability of emission of one given i was in the state 0 0 and probability of 0 0, when I am writing this probability of 0 zero 0 zero one I assume that this is the latest symbol. This is the previous symbol.

So, this is a further away from this symbol. So, to find out a probability of 0 zero 1 is that I have to find out what is the transition probability 1 given 0 zero and of being in the state 0 zero. So, I have to evaluate all this probabilities. Now, probability of 0 1 0 is again probability of zero, given i was on stage 0 one multiplied by probability of 0 1. Now, this is the quantity probability of 0 given 0 zero is 8 by 10 because 0.8 has been given to us. This is the probability, which has been given the probability of 0 zero

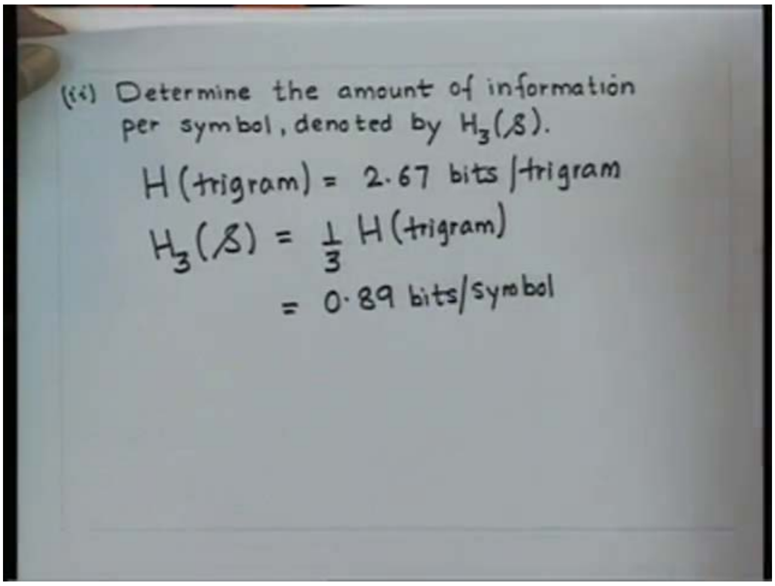
We have calculated that is nothing but a 5 by 14. So, this turns out to be 4 by 14. Similarly, this quantity turns out to be 5 by 14 multiplied by 2 by 10 is equal to 1 by 14 this quantity turns out to be 1 by 14. So, based on this idea I can calculate the probability is of 0 1 1 that turns out to be 1 by 14. Probability of 1 0 zero is again 1 by 14 probability of 1 0 1 is equal to 1 by 14 probability of 1 1 0 is 1 by 14. Finally, property of 1 1 1 is equal to 4 by 14. Once I have these probabilities, I can use the expression for the calculation of the entropy.

(Refer Slide Time: 20:07)



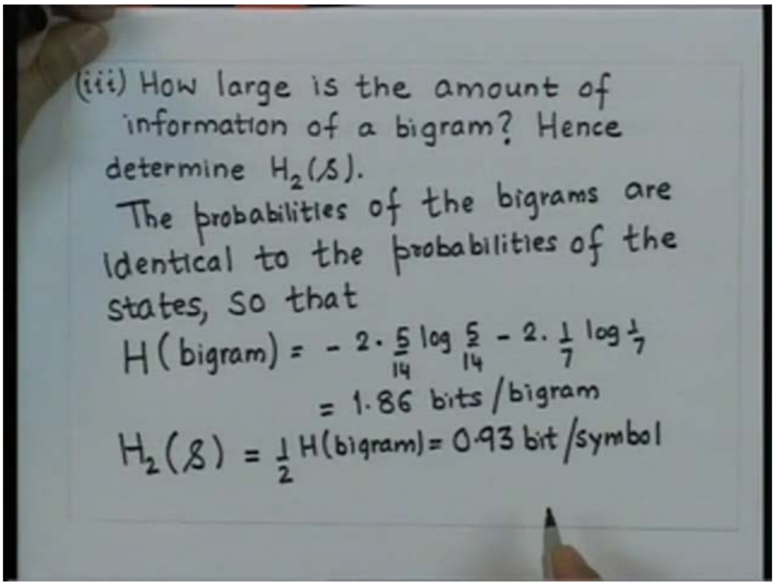
Then calculate entropy or the information in the trigram as equal to minus twice 4 into 14 log of 4 by 14 minus 6 times 1 by 14, log of 1 by 14 and this turns out to be 2.67 bits per trigram. So, for the example under consideration we have for the messages of length 3. The information in those messages is 2.67 bits per message is nothing but a trigram that is symbols of length 3. Now, the next question is once I know this is basically what we are calculating is H of N s. In our case this N was nothing but now the next question is that. Once I know the message in a trigram than determine amount of information per symbol, which is denoted by H 3 s. Now, if I want to calculate F 3 s.

(Refer Slide Time: 22:03)



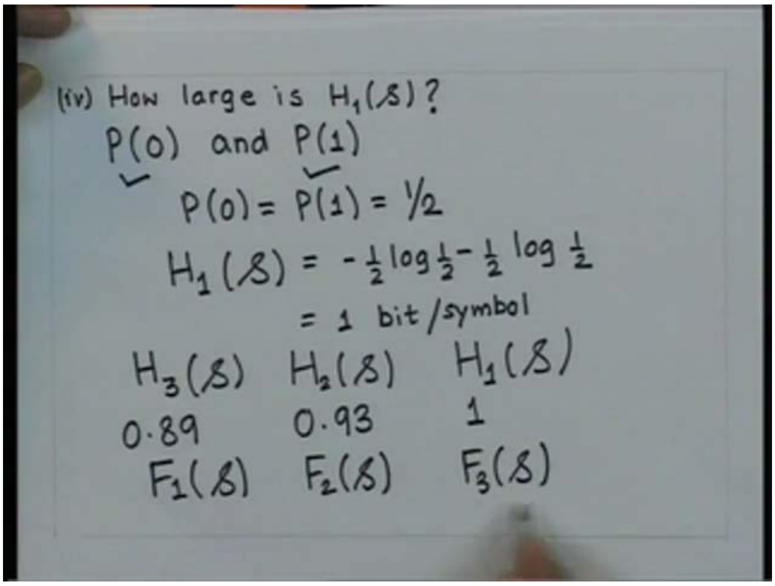
It is very simple, since I know my H of trigram was equal to 2.67bits per trigram. This is nothing but H V and H 3 s 1 by 3 H of trigram. This is nothing but 0.89 bits per symbol. Now, the next question is I have calculated, if I were to reduce this messages of from the length 3 to length 2 then what happens to the amount of information. To answer that question let us solve this. So, next what I am interested is how large is the amount of information of a bigram. Hence, determine H 2 s.

(Refer Slide Time: 23:06)



So, the probabilities of the bigrams are identical to the probabilities of the states and since we have calculated the probability of the states earlier, so we can say. So, that information from the bigram would be minus twice 5 by 14 log of five by 14 minus twice 1 by 7 log of 1 by seven. This turns out to be 1.86 bits per bigram this is your again H V here are the messages of length 2. Hence, H 2 s which is nothing but half H of bigram is equal to 0.93 bit per symbol. Now, let us look at the messages of length 1. So, if I look at the messages of length one than how large is which H 1 s to calculate.

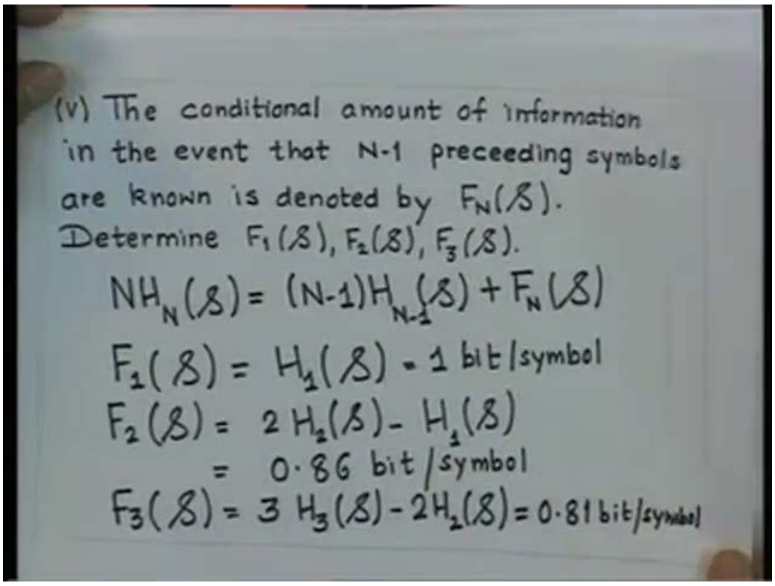
(Refer Slide Time: 25:12)



This I should know the probabilities of P 0 and P 1. Now, once I know the probability of 0 and probability of 1, I can calculate H 1 s we have seen basically how to calculate the probability of 0 and probability of 1 for this example earlier. So, we had seen that probability of 0 is equal to probability of one and that turns out to be equal half. So, from this I get H 1 s is equal to minus half log minus half minus log half and that is equal to 1 bit per symbol.

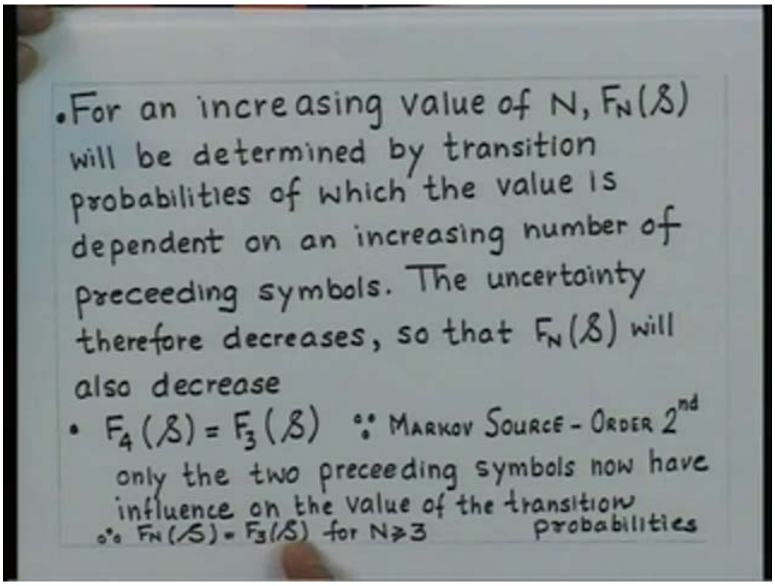
Now, we have calculated H 3 s H 2 s and H 1 s the value of H 3 s we got was 0.89 bits per symbol value of H 2 s. We got was 0.93 bits per symbol and value of H 1 was 1 bit per second 1 bit per symbol. So, what this means that as I keep on increasing N the value monotonically decreases from 1 I have reached up to 2.89. So, the final limit of H 3 of H N s would be H infinity s. Now, if I were interested in to calculate F 1 s F 2 s and F 3 s, let us try evaluate this. So, the condition amount of information in the event that N minus 1 preceding symbols are known is denoted by F N s.

(Refer Slide Time: 27:27)



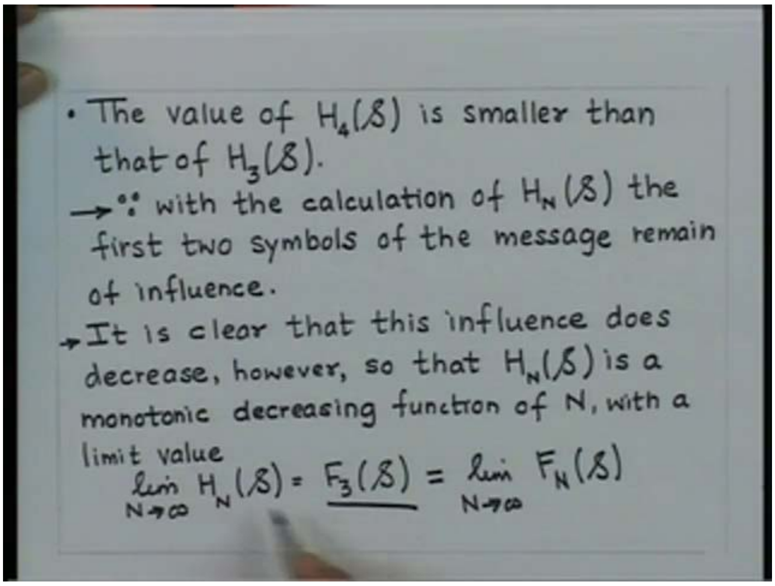
So, what I am interested is to determine F 1 s F 2 s and F 3 s we have derived that N H N sis equal to nminus one H N minus 1 s plus F N s. Now, if I want to calculate F 1 s, since I know H 1 from F ones is nothing but H 1 s. That is nothing but one bit per symbol F 2 s from this relationship I will get as twice H 2 s minus H 1 s. Now, H 2 s and H 1 s I have calculate earlier. So, this turns out to be 0.86 bit per symbol. Finally, my F 3 s is equal to thrice H 3 s minus twice H 2 s. This comes out to be 0.81 bit per symbol.

(Refer Slide Time: 29:45)



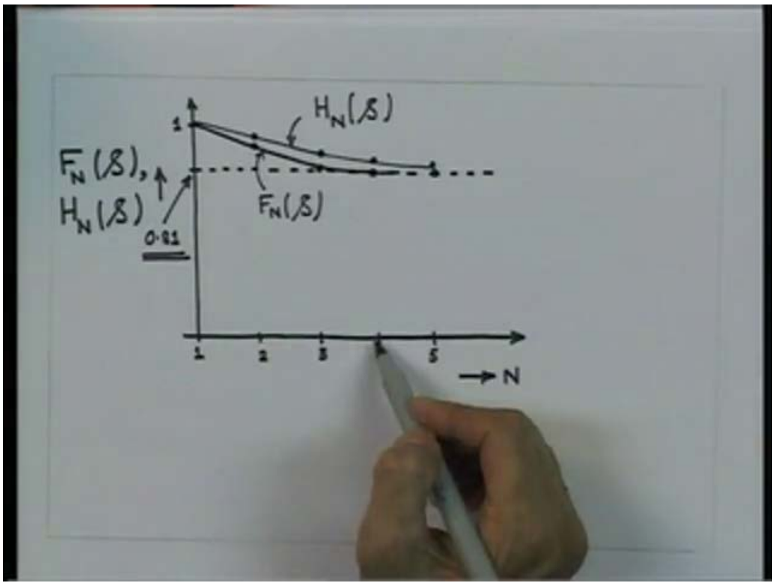
Now, what we have observed that for an increasing value of N F N s will be determined by the transition probabilities, of which the value is dependent on an increasing number of preceding symbols. So, the uncertainty therefore decreases, therefore F N s will also decrease. Now, in our case if you try to calculate F 4 s F 4 s will turn out to be same as F 3 s because Markov source is order second. So, the only two only the two preceding symbols, now have an influence on the value of the transition probabilities. Therefore, F N s will be equal to F 3 s for N greater than equal to 3. Now, another interesting result is F N s does not change beyond F 3 s, but what happens to H 4 s? if you look at H 4 s calculation.

(Refer Slide Time: 31:04)



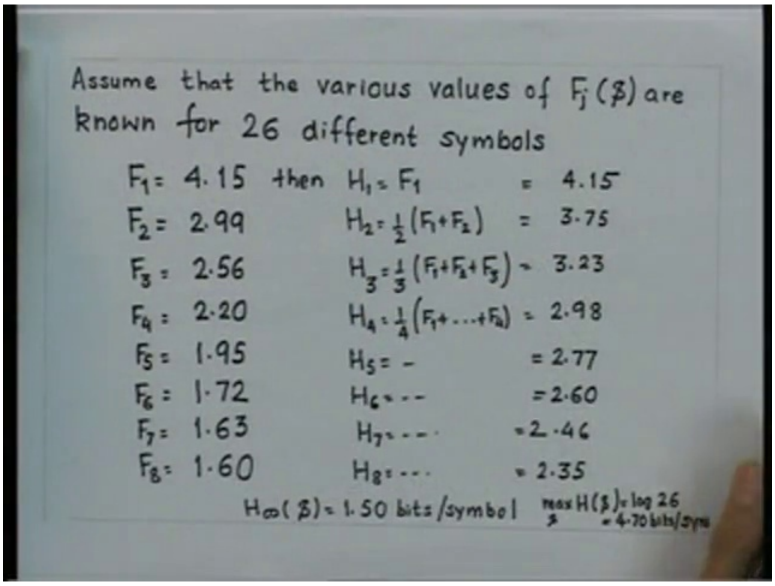
Then you will find that the value of H 4 s is smaller than that of H 3 s, because with the calculation of H N s the first two symbols of the message remain of influence. It is clear that this influence does decrease however. So, that H N s is a monotonic decreasing function of N with a limit value. So, if you take the limit of H N s N tending to infinity you will get as same as F 3 s. Because, F 3 s is nothing but limit of N tending to infinity of F N s, because our source is the second order Markov source F N s is for aim greater than or equal to 3 is same as F 3 s. Now, if you plot the values of H N s and F N s as function of N, what you will get is this graph.

(Refer Slide Time: 32:19)



It is interesting to examine this graph on the x-axis I have increasing value of N on the y-axis I have plotted F N s or H N s. If you look at the plot of F N s for N equal to 1 we found out to value to be of fn F 1 s was 1 for F N s for N equal 2 was lower than that. For F 3 was still lower, but F 1 s f is all stabilised the value of F threes, but this is not true of H N s. H N s goes monotonically decreasing function, but as I keep ongoing beyond N greater than 3. You will find that H N s keeps on decreasing. Finally, it approaches the value of F 3 s and that is equal to 0.81. So, what I would say that entropy of this source, which we had considered is 0.81 bits per symbol. Let us take another example to understand the concepts in a much better way.

(Refer Slide Time: 33:34)



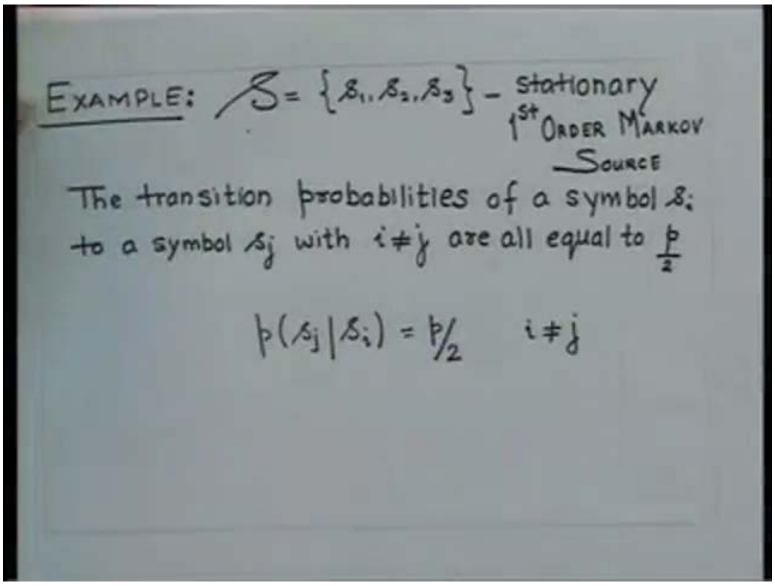
Assume that the various values of F j s are known for 26 different symbols. This could be an example of English literature. So, let us assume that I know F 1, so that will be equal to 4.15 bits. If I can calculate F 2, which I am calculation of all this F 1 F 2 F 3 is based on the conditional information measure. So, F 2 if I can calculate if i know it is 2.99. I can calculate H 2 o, which is equal to 1 by 2 F 1 plus F 2 and that turns out to be 3.75. So, I can keep on calculating this F 1 F 2 F 3 F four up to F 8.

Similarly, I can calculate the values for H 8 from F, if you look at these values of F n and H N there is a interesting result. You will find that always H N s is larger than f n that is the one-point, which you have to notice another is f n is always a monotonically decreasing function of N. Similarly, you will find that H N is monotonically decreasing function of N. Now, if you look at the way these f n decreases f n decreases much faster than H N. When I reach F 8 the value for entropy turns out to be 1.60 bits per symbol, whereas 2.25 bits per symbol.

If you continue this process than you will find H infinity s stabilises or converges to a value of 1.50 bits per symbol. H N s will also converge to the same value one-point bits per symbol, but the convergence rate will be slower than that for f n. If I were to find out what is the maximum value, which I can get of an information from this source, which consist of 26 different symbol the value will be obviously equal to log of, 26 that will happen. I assume the symbols are all independent and that turns out to be 4.7 bits per symbol.

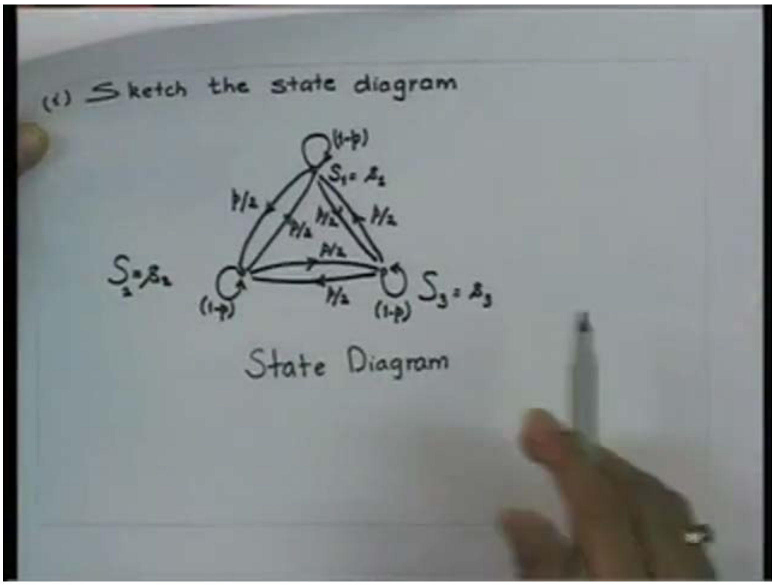
What is to what I want to convey from this example is that if I take a source, which consist of 26 different symbols. Then the entropy of that source based on the assumption that all the symbols are I will get that entropy to be equal to 4.7 bits per symbol, but if i look at that source with a memory. Then the entropy turns out to be much lower than 4.70 bits. It is equal to 1.5 bits per symbol. So, there is a lot of discrepancy between these two entropy. Now, to get a better feel will take one more example before we conclude our discussions on the modelling of these sources, from the information point of view. Let us take another example.

(Refer Slide Time: 37:35)



Let me assume that I have a source, which emits three symbols s 1 s 2 s 3. This source is a stationary and I also assume that this source is a first-order Markov source. It has been given to me that the transition probabilities of a symbol s i to a symbol s j with i naught equal to j are all equal too p by 2, p is a variable. So, what it means is that probability of s j given s i is equal to p by 2 for i naught equal to j. Now, let us try to calculate the different values for information measures, which we have studied so far. The first thing that is to be done before we do the calculation is to draw the state diagram for this source. The state diagram for the source will give us all the information, which is essential for us to calculate all the transition probabilities. So, if I take the state diagram for this would be something like this

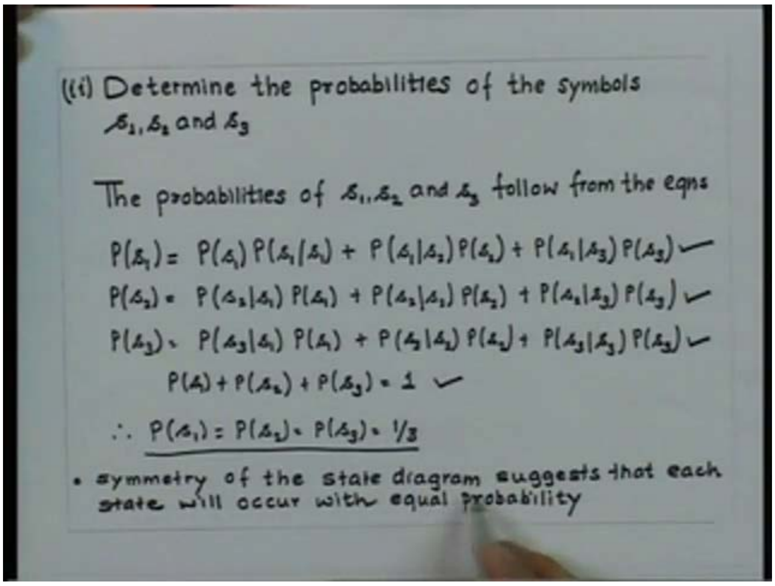
(Refer Slide Time: 39:12)



So, the state diagram would be given by this, this is the state diagram for source under discussion. Since, we have this source, which is a Markov source of order one. There are three symbols we have three states this three states are identical to the three symbols. So, s 1 is equal to s small s 1 capital S 2, which denotes the state's is same as the symbol s 2. The third state S 3 is identical to the symbol s 3. Now, it has been given to us that probability of s j given to s i is equal to p by 2 for i naught equal to j. So, if i take an example for this state than the probably going from this to this is p by 2 and from here to here is p by 2.

So, the probability of being in the same state would be calculated by addition of this p by 2 plus p by 2, that comes out to p 1 minus p would be the probability of remaining in the state. So, similarly I can calculate for state s 2 and s 3 and this would be the state diagram, which I get for source under discussion. Next, is basically what I am interested is to calculate the probabilities of the symbols s 1 s 2 and s 3. Now, since the state's and the symbols in this example are the same calculation of the probabilities of the symbol is equivalent to calculation of the probabilities of the state's s 1 s 2 and s 3. We had seen how to do that let us try to calculate these probabilities.

(Refer Slide Time: 41:22)

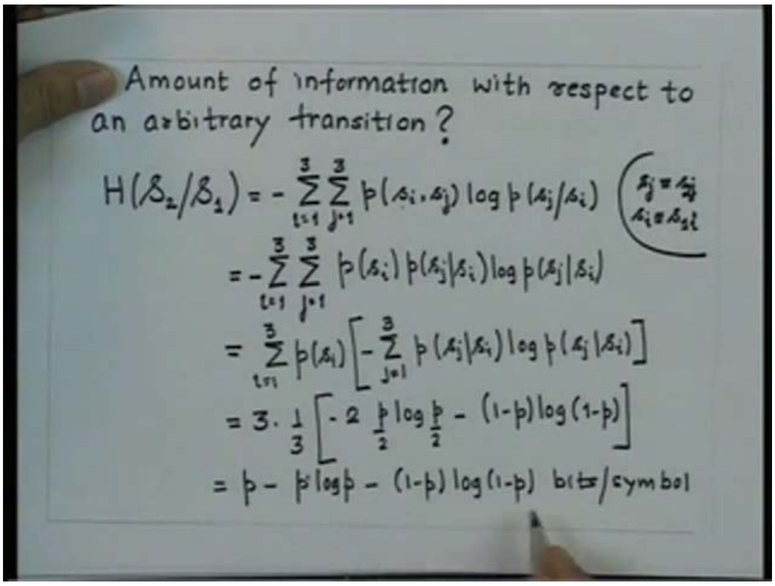


So, the probabilities of symbols s 1 s 2 and s 3 or in other words the probabilities of the state capitals one capital s 2 and s 3 follow from this equation probability of s 1 would be, what is the probability of emission of s 1, when i state s 1 and multiplied by the probability of the state s 1 itself. Plus, what is the probability of emission of s 1 when I am state s 2 multiplied by probability of s 2 itself.

So, this would be the expression, which I get for calculation of the probability of the symbol s 1. In this case it is equivalent to calculation of the probability of the state s 1 that is capital s 1. Similarly, this equation follow for symbol s 2 and symbol s 3. I know that probability of s 1 plus probability of s 2 plus probability of s 3 is equal to 1 Therefore, if I substitute this values if I solve all this equation, what I will get is this result probability of s 1 is equal to probability of s 2 is equal to probability of s 3 is equal to 1.

Now, because of the symmetry of the state diagram this result was expected. So, the symmetry suggests that each state will occur with equal probability. Now, during the course of a study of Markov source. We also talk about the information associated with the arbitrary transition. In order to understand that concept in a better way, let us try to find out the amount of information with respect to an arbitrary transition for this example under discussion.

(Refer Slide Time: 43:40)

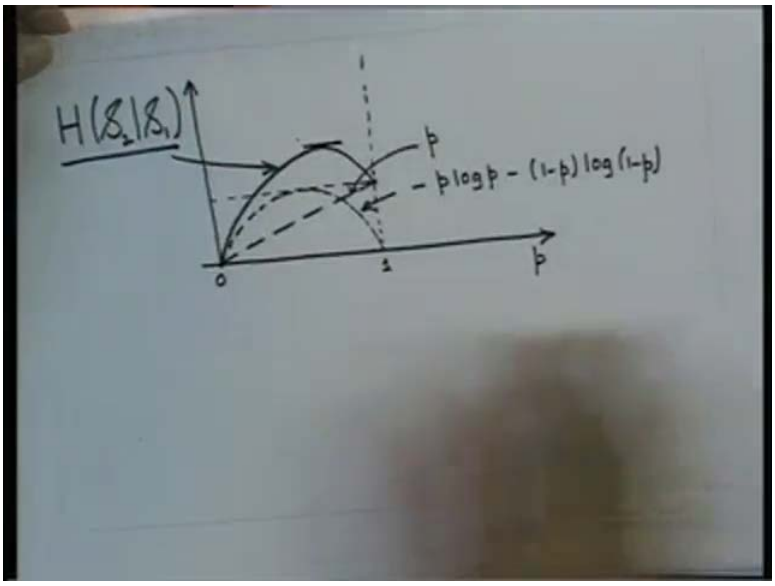


So, I am interested in the calculation of information with respect to an arbitrary transition. What I mean by that I am interested in calculating average information of s 2 given s 1 in the light of our earlier discussion this s j is equivalent to s 2 j and s i is equivalent to s 1 i. So, by definition H of s 2 given H 1 would be probability of s i s j joint probability multiplied by what is the information, which I get when there is a transition from s i to s j. The amount of information, which I get from one transition from one particular transition s i to s j is minus log of p s j given si and the average value would be this.

So, if I try to simplify this value I will get as this is equal to probability of s i probability of s j given s i log of probability of s j given s i. Now, this I can simplify as I can remove probability of s i outside the bracket log of probability s j given s i this can be simplified. Because, probability of s i’s are all equally probable three times 1 by third minus 2 p by 2 log of p by 2 minus 1 minus p log of 1 minus p. This is a value, which I get when I am in a particular state. Since, all the probabilities of the states are probable will just multiply by 3.

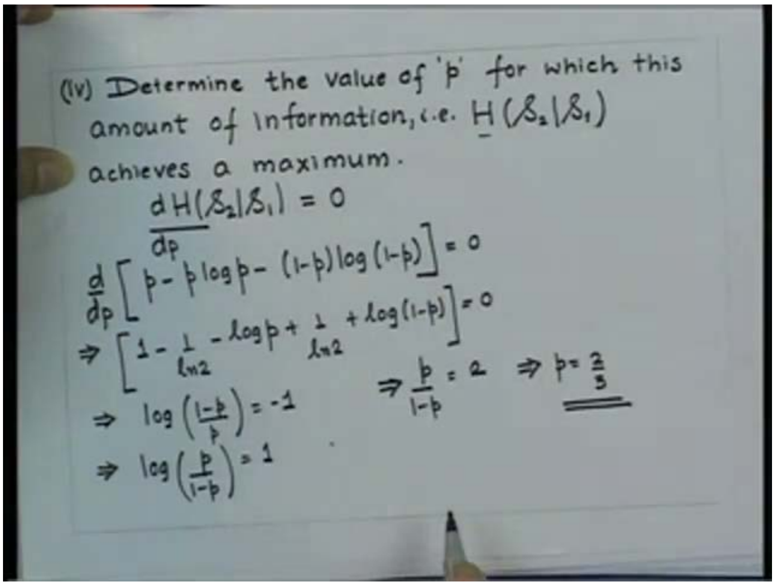
If I simplify this expression it turns out to be p minus p log p minus 1 minus p log 1 minus p bits per symbol. If you plot this the plot of this expression as a function of p can be, this is familiar expression, which we get, this is the entropy function we have. So, this also entropy of a binary source, this is this I can plot as a straight-line. So, entropy the amount of information with respect an arbitrary transition is sum of two functions. This is entropy function another is a linear function.

(Refer Slide Time: 47:20)



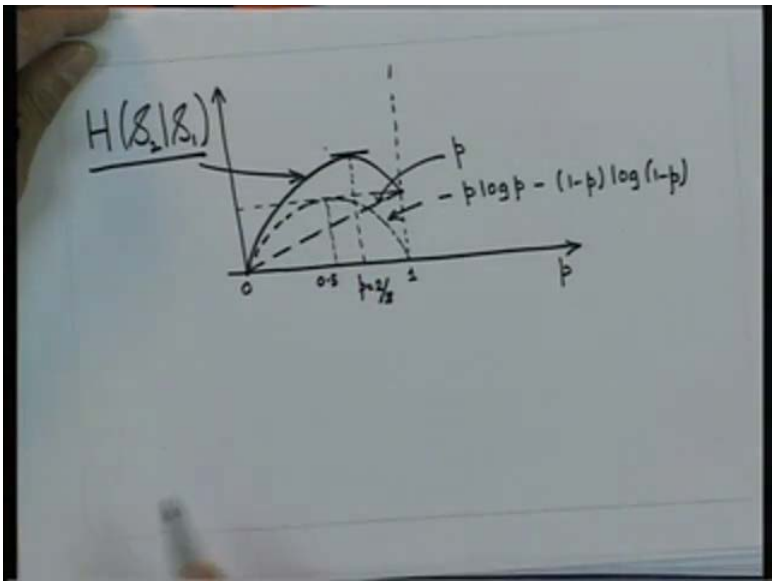
So, we can plot this as this would be the result for my entropy function. That is nothing but minus p log p minus 1 minus p log 1 minus p. This value will range from 0 to 1 and this will be my p. So, my H of s 2 given s 1 would be the sum of this and that will turn out to be something like this. So, if you look at this additional information, this is the graph of this. Now, this graph shows that there is some kind there is a peak out here let us try to evaluate what is the value of this maximum? What is a maximum value which this achieves? For what value of p it achieves? So to determine the value of p for which this amount of information, that is s 2 H of s 2 given s 1 achieves a maximum it is not very difficult.

(Refer Slide Time: 49:23)



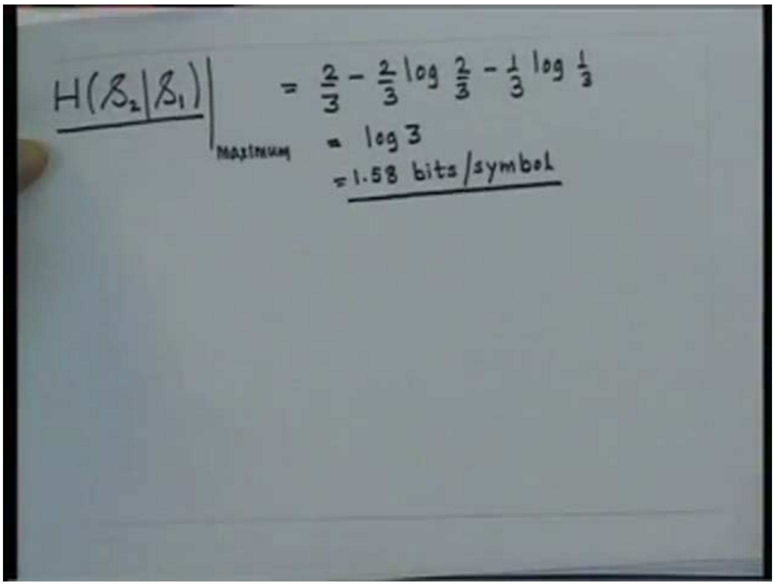
To determine the value of p for which this amount of information achieves a maximum is given by derivative of H of s 2 given H 1 s 1 is equal to 0. Now, let us take the derivative of it p minus p log p minus 1 minus p log of 1 minus p is equal to 0. This implies derivative of this is one derivative of p log p will be 1 by log 2 minus log p plus derivative of this quantity plus log of 1 minus p. This implies that log of 1 minus p upon p is equal to minus 1, which implies log of p over 1 minus p is equal to 1. Therefore, this implies that p over 1 minus p is equal to and this implies that p is equal to 2 by 3.So, the maximum value of this a value at which the maximum will occur.

(Refer Slide Time: 51:20)



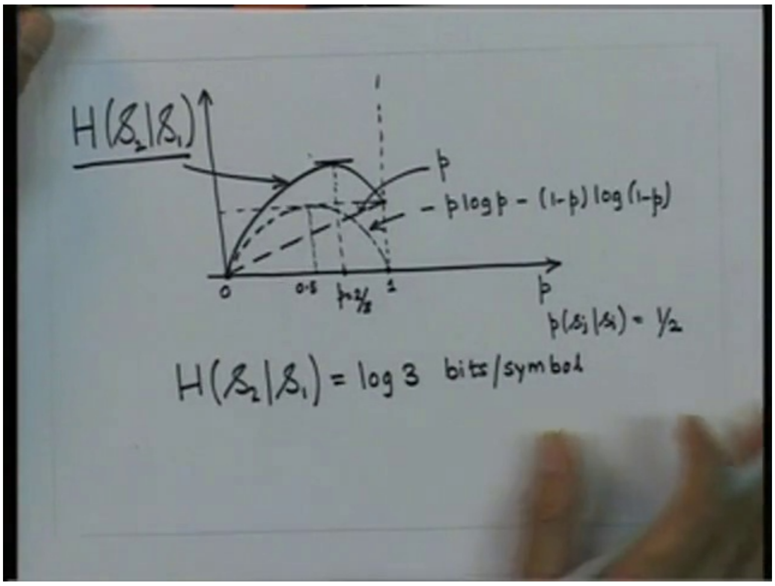
So, in this graph this maximum value will occur for p equal to 2 by 3, this occurs at 0.5. Now, what is the maximum value? If we calculate that, maximum value I have to evaluate that expression for p equal to two third.

(Refer Slide Time: 51:42)



If I substitute that I get two third log and this turns out to be log 3. This is equal to 1.58 bits per symbol. So, this is the maximum value, which I get for this quantity. This is the arbitrary transitions, now let us look at this graph. If you look at this graph H 2 s H of s 2 given s 1 for p equal to 0 I get this value to be 0.

(Refer Slide Time: 52:59)



Now, this is expected, because at this point p equal to 0 the source remains in the same state. Therefore, there is no uncertainty and therefore, the amount of information is 0. When p is equal to 1 the source has an equal chance of coming into one of both other states from any given state, because probability of s j given s i is equal to half in that case. In that case the information the transition information will be 1 bit.

Now, when p is equal two third every transition has the same probability of occurring. Therefore, the source behaves as if there are three independent symbols, what it means? If a previous symbol is given the uncertainty is not reduced. Therefore, H of s 2 s 1 turns out to be same as then copy of the source, when the source symbols are independent and that is log 3 bits per symbol.

So, with this we complete our discussions about the modelling of the sources. From information point of view, we have looked at different types of sources, sources with which are memory less. That is zero memory and the sources, which have memory. In that category we have considered a Markov source of arbitrary order. We have also looked at the calculations of the information for this source. Now, with this base we will go ahead and look at coding aspect for these sources.