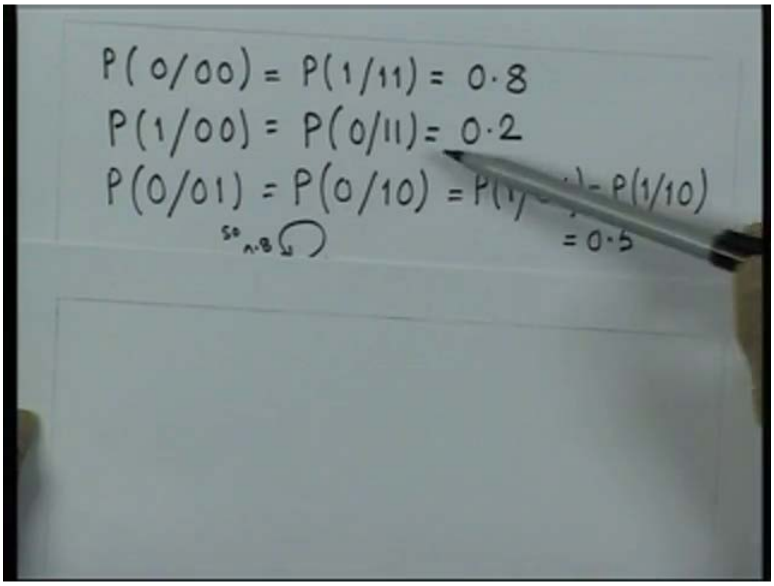
Information of Theory and Coding
Prof. S. N. Merchant
Department of Electrical Engineering
Indian Institute of Technology, Bombay

Lecture - 4
Adjoint of an Information Source, Joint and Conditional Information Measures

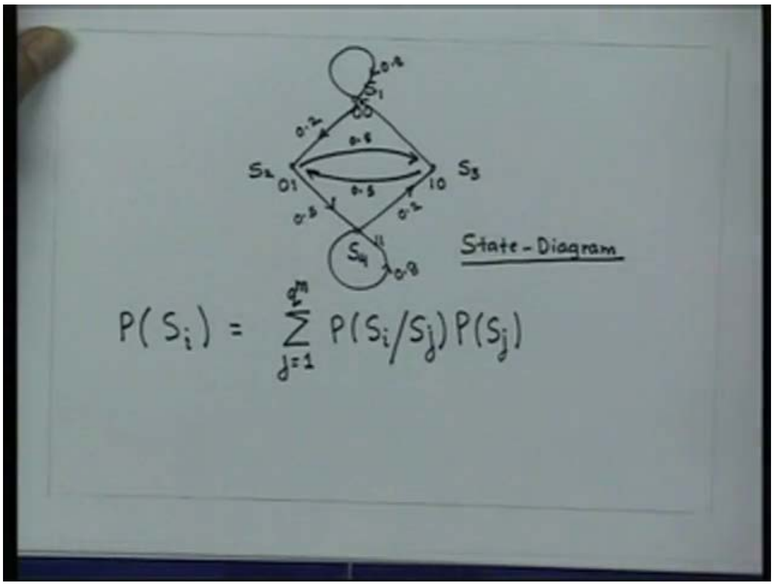
We have looked at the definition of an M th order Markov source. M th order Markov source was defined as the source, which emits symbol wherein the occurrence of each symbol is dependent up on M preceding symbols. We also looked at the definition of an ergodic Markov source, where we said that the probability distribution, over the set of states for an ergodic Markov source, remains constant, it does not change with time, what we are interested is in calculation of entropy for an M th order Markov source. In order to do this, the first thing that we have to calculate is the probability distribution over the set of states. Let us take an example and find out the procedure to do this.

(Refer Slide Time: 01:54)



Let me go back to the previous example, which we had looked earlier, where, I have a second order Markov source and the conditional transition probabilities have been given, as shown here. So, this source consists of two binary symbols 0 and 1, and it is in a second order source. Therefore, you have four states, and the transition probabilities have been given, as shown on, as shown here.

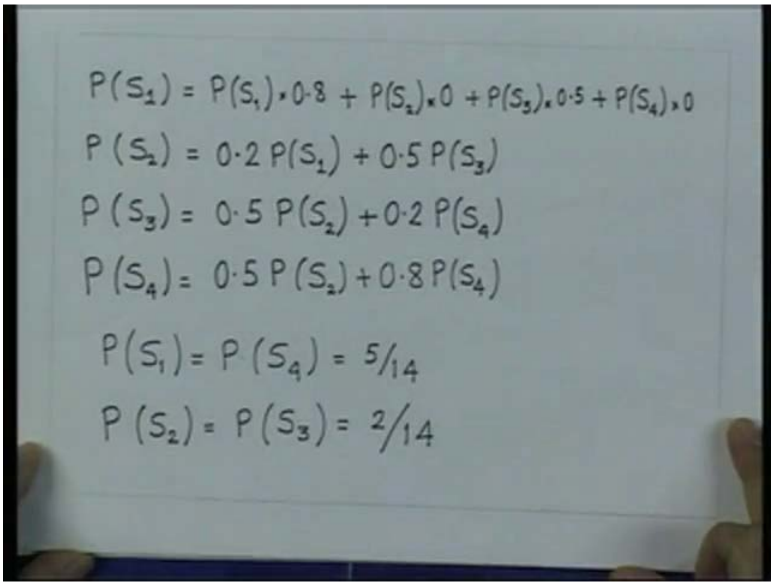
(Refer Slide Time: 02:40)



For this, we can draw the state diagram very easily as I have four states S1, S2, S3, S4 and the conditional, the state diagram for this is as follows. We had drawn the same diagram earlier too, so this is my state diagram for our example, under discussion. Now, if it is ergodic Markov source then I can find out what is probability of these four states, based on the conditional symbol probabilities. Let us look at that now, for any particular state if I want to calculate what is the probability of state, that would be given by this following expression. Probability of being in a particular state Si, given i was in a state as j.

And what is the probability of the state is j that is given by PSj and this is sum over, all the values from j equal to 1. In our case there are four states, so you will have four in general it would be q raised to m, where q is a size of the source alphabet. And m is the order of the Markov process, so using this expression I can find out what is my probability of states.

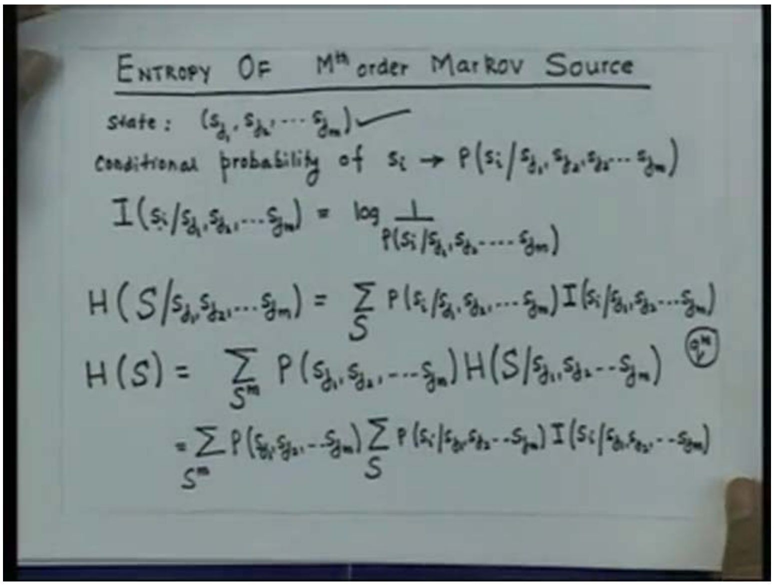
(Refer Slide Time: 05:02)



Now, for our specific example this can be reduced to writing something like this. I can find out what is my probability of S1, obviously my probability of S1 is, probability of S1 given, i was in S1. That is 0.8 and what is the probability of state S1 that is what I am supposed to calculate. I cannot come to S1 state from S2 nor I can come to state S1, from S4. So, all this are 0, whereas I can come to state S1from S3.

So I get probability of S1 given S3 is 0.5 and probability of S3 itself is indicated here. So, I have four simultaneous equations and if I solve these four simultaneous equations, what I get are these values. So, these are the probability distribution of an ergodic second-order Markov source. Now, once we have this quantity, we can find out what is the entropy of the same source, which is under discussion. Let us look at, the calculation of the entropy for such a source. Now, if I want to do that, let us try to develop a general expression for an M th order Markov source.

(Refer Slide Time: 06:48)



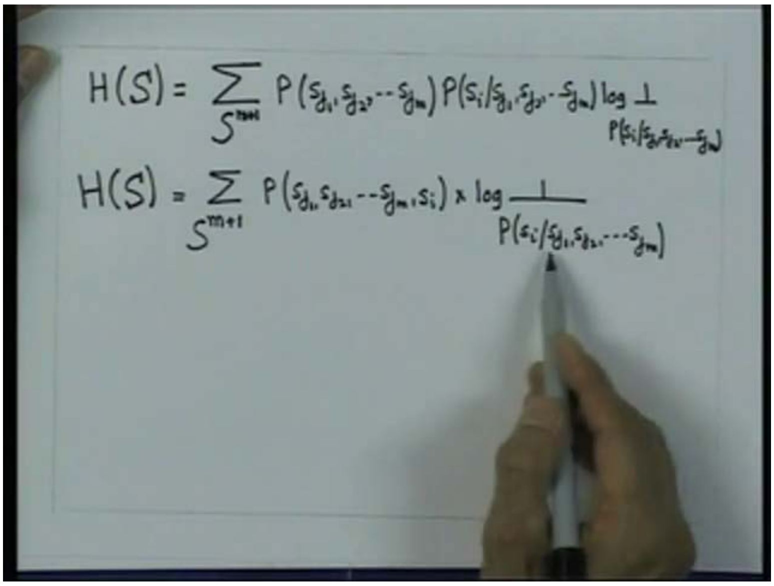
So, we will calculate now entropy of MTh order Markov source. So, let us assume that, the source was in the state given by Sj1, Sj2 up to Sjn. If the source is in the state given by this then there are m conditional probabilities, there are q conditional probabilities, associated with this state q because the length of the source alphabet is q. So, let us find out what is the conditional probability, conditional probability of the receiving symbol Si. Given I was in the state will be, nothing but probability of Si given Sj1, Sj2, Sj3 up to Sjm.

So, when I receive a particular symbol Si given, I was in this state. Then the amount of uncertainty resolved or the amount of information which I gain on the occurrence of Si, in this state. Will be given by, I of this, denotes the self-information, which I get on the occurrence of Si given, I was in state Sj1 Sj2 up to Sjm is nothing but by definition as seen earlier log of 1 by probability of Si given Sj1, Sj2 up to Sjm. So, this is the amount of information which I get, on occurrence of a particular symbol Si. Now, if I average the information which I get, when the source in the state given by this. Then that would be, H of S, Sj1, Sj2 up to Sjm is equal to, probability of Si given Sj1, Sj2 up to Sjm.

Information which I get is this, this is sum over the source alphabet S. So, this is the average amount of information per symbol while, we are in state this. Now, this state itself can change so if I average this quantity over all the states then what I will get is the entropy of that MTh order Markov source. So, to get the entropy from this quantity, I had to take the average so the entropy of the source would be given by, probability of being in a particular state. That is given by this, multiply by the average information that I get when, I am in that state.

This quantity, is obviously averaged over all the states so there are a number of states, which we can have is given by q raised to m. So, I should sum this quantity for all the states qm states. Now, to denote that summation I will just use an abbreviation, I will denote this as sm. So, when I write sm it means that, I am summing this over different q raised to m states. So, let us try to simplify this quantity what I will get is sm, this quantity comes from here. So, I can write this now, I can take this summation outside.

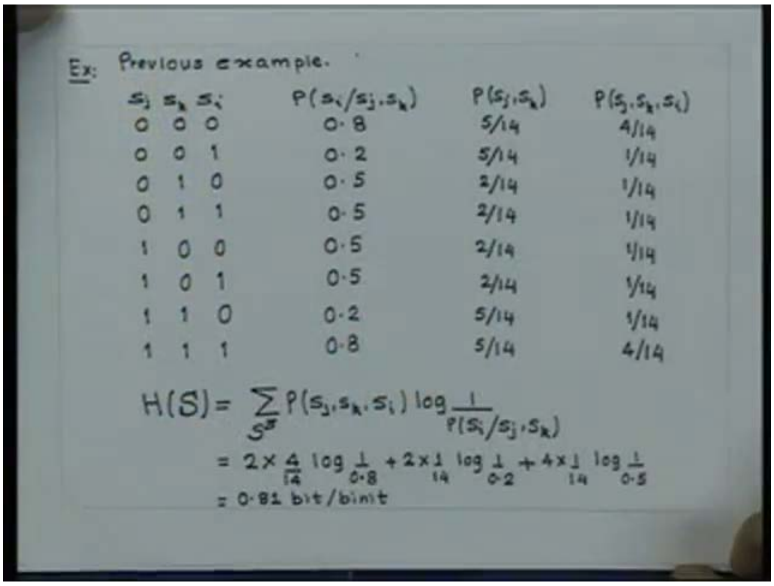
(Refer Slide Time: 13:46)



And if I do that, what I get is H of S, is equal to probability of Sj1, Sj2 up to Sjm log of. Now, this summation is over this state S and also over the source S. So, to denote that in abbreviated form I will write it as S plus m 1. So, this can be simplified as, this quantity can be simplified as multiplied by log of probability of Si, Sj1, Sj2 up to Sjm. So, this is the final expression which we get for the calculation of entropy of an MTh order Markov source. Let us take an example and try to evaluate this expression to get, the feel of what we are doing.

So, let us go back to our previous example, which we had considered second-order Markov source, for which we just calculated the probability of states. So, if I take that example and if I am interested in calculation for that example, this entropy then in our case m is equal to 2. So, what will happen this summation will go for over S raised to 3 and this quantity out here, will be two elements here, there will be one element here and then the conditional probability given there.

(Refer Slide Time: 16:51)



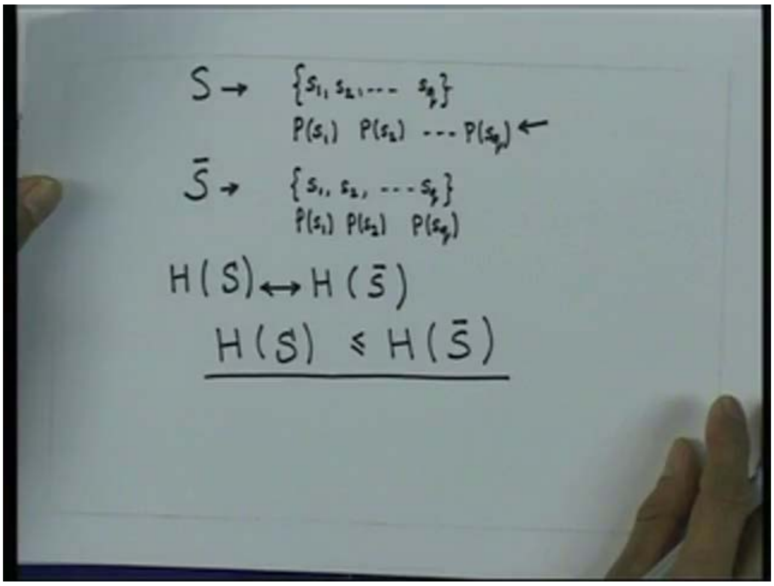
So, that can be written as so this is the previous example which I have. Let us try to understand this, what I have written here. So, Sj, Sk denotes the particular state I have four states so 0 0 0 1 0 1 1 0 1 1, these are the four states which I have. And in each state the source can emit either 0 or 0, so when it is 0 0 I can emit 0 or it can emit 1. Similarly, when it is in the state 0 1, it could either emit 0 or 1 so Sj, Sk, Si so these are the various combinations which I can have in this expression. This is what I am writing, I am supposed to get this expression evaluated to do this, I will require the states. So, I have written down all the various combination of Sj, Sk, Sr.

I will also require the conditional symbol probabilities so conditional symbol probabilities equivalent of this, for this problem is given directly from the specifications of the problem. This was given to us as a specification of the problem, probability of Sj, Sk is nothing but the probability of a particular state. So, we have 0 0 0 1 1 0 1 1 so this we had just seen, how to calculate and the values we have calculated are given here. Once I have this, I can find out what is probability of Sj, Sk, Si by simply multiplication of this two, what I will get this values.

And then you just take this values, I have this values for each of the, this, I have got eight values here, i have got eight values here, just plug them into this. And what I get is 0.81 bit per minute so if I have a second-order Markov source then the entropy of that source is given by 0.81 bit per minute, if the source was not a second-order Markov source.

And it was 0 memory source then we have shown that when the probability of occurrence of the symbol 0 and 1 are equi probable then the entropy of that 0 memory source turned out to be 1 bit per minute. So, we find that when, there is a dependency upon the symbols, even if the source is a binary source then the entropy of that source is less than a 0 memory source. Now, let us look, let us take some clue from this example and try to discuss it in a more generic fashion.

(Refer Slide Time: 20:21)



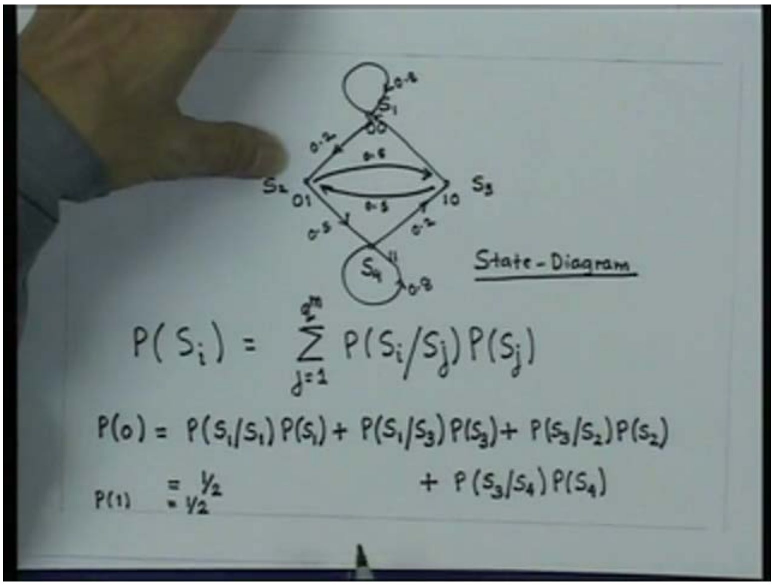
Suppose, if I have a source S which is a Markov source could be of any order. And this source S will obviously have, some source alphabet let us assume, they are given by S1, S2 and Sq. Now, for this source S1, S2, Sq given by the source alphabet, I can always calculate the probability distribution of the states. Now, given the probability distribution of the states, it is also possible for me to calculate the first order symbol probabilities. That mean unconditional probabilities I can find out, what I mean by that I am interested in finding out what is the probability of P of S1, P of S2 and P of Sq. These are known as unconditional symbol probabilities, now let us assume that I can calculate for a moment.

Let me assume that I can do this, for any Markov source now, if I have another source, let me call another source is S bar, this source S bar also has, the same source alphabet that is S1, S2, Sq, but this source S bar is a 0 memory source. So, if it is a 0 memory source then it has got only unconditional probabilities and let me assume that, unconditional probabilities for this source is identical to the one, which we have for the source S, So, the unconditional probabilities are also identical and they are given by PS2, PSq. Now, the question is what is the relationship between HS and HS bar?

Is there some kind of a relationship existing between these two sources, what it means basically that in all respect, the source S and S bar are same. Source alphabets are same and the unconditional symbol probabilities are same, but the source S emit symbol wherein, the occurrence of each symbol is dependent upon the preceding symbol. Whereas, S bar is a 0 memory source so is there any kind of relationship between these two. We will very shortly see that I can show that, entropy of HS is always less than equal to entropy of S bar, a very important relationship.

Now, to prove this relationship in general for an M th order Markov source is little bit involved so what I will do is, try to simplify this source S, for our derivation will assume that this source S is first order Markov process. So, for the first order Markov process, will try to find out another source S bar, given by this specification. Now, if you look at the previous example, that is a second-order Markov source where we had calculated, the probability of the states then it is also possible for me to go, little beyond and calculate the unconditional symbol probabilities.

(Refer Slide Time: 25:02)

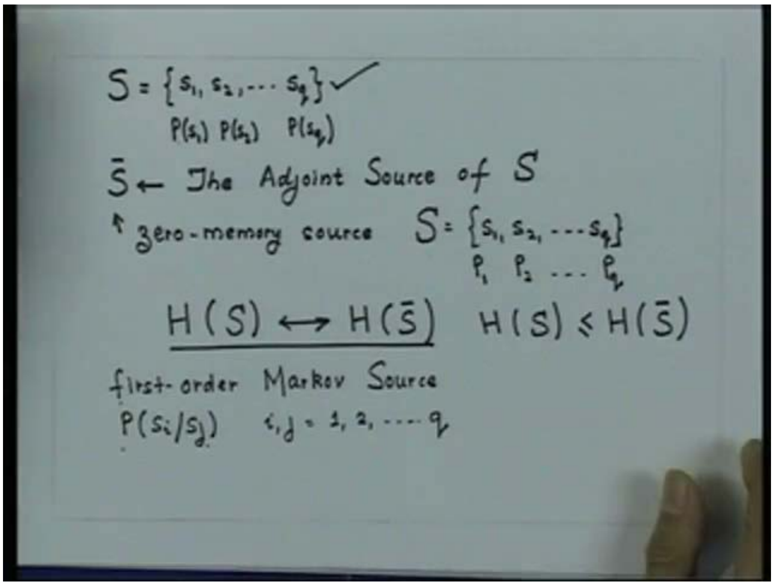


Now, it is not very difficult to show that, going back to this example which I have out here, I can, if I were interested in finding out, what are my unconditional symbol probabilities, for this ergodic second-order Markov source. Then for symbol zero I can write down as probability of occurrence of 0 would be when, probability of occurrence of 0 would be when I am in state S1 and I go back in state S1, because if I go from state S1 to state S1 it means that, 0 has occurred. And what is the probability of occurrence of being in the state S1, so if I use this kind of information I can calculate, and what is the unconditional symbol probabilities of 0?

So, let us do this write down this expression for the 0 symbol. So, probability of 0 would be probability of, going to state S1 when I am in state S1 and what is the probability of the state S1? So, that is 1 then I can come to this if I am in state S3 than if 0 occurs I go to S1 then what is the probability of coming to state S1 from S3 multiplied by probability of S3 itself. Plus, I have probability from S3 to S2 so if 0 occurs from S2 to S3 so if I, what is the probability of going in state S3, when I am state S2 multiplied by probability of state S2 itself. And finally, I have probability of I can come from, when I go from state s4 to s3 that can occur only when 0 occurs.

So, probability of S3 given, I was in S4, probability of S4 itself. Now, if I, I know all this value this is 0.8, this is given out here in the state diagrams. And we have already calculated probability of PS1, PS2, PS3 and PS4 if you plug on all those value out here, you will get probability of 0, to be equal to half. Similarly, you can calculate the probability of 1 is equal to half so this we have done, I have shown you how to do it for a simple case of a second-order Markov process, but the similar procedure can be adopted, for a general MTh order Markov process. So, what I want to say is that given a Markov source MTh order Markov source, I can always calculate it the unconditional symbol probabilities. So, let us try to define.

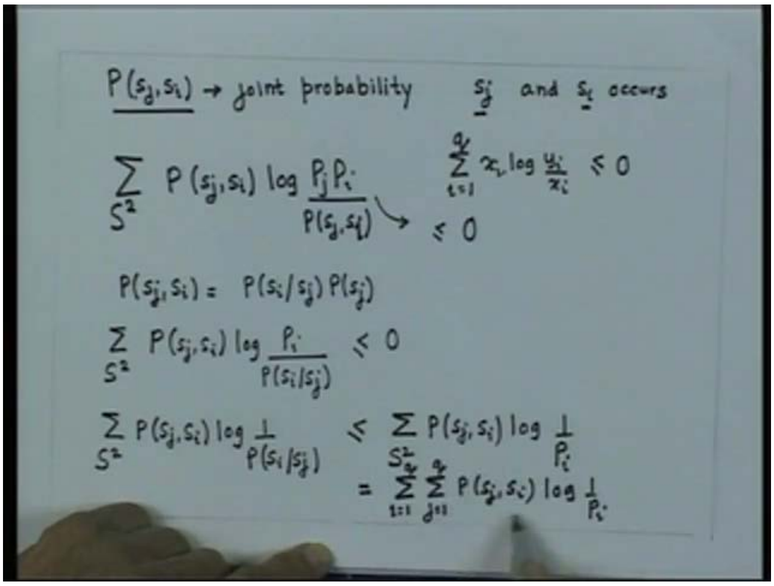
(Refer Slide Time: 29:17)



So, if I have a Markov source S, I will consider just now it is first order. So, let me define MTh order Markov source consisting of source symbols S1, S2 and Sq, this is MTh order Markov process. And with each of these symbols, I have associated unconditional probabilities, these are unconditional probability of the symbol. I can say the first order of symbol probabilities now, let me define another source S bar and this S bar is known as the adjoint source, of S. This source S bar, which is adjoint source of S. S is a 0 memory source with the identical source alphabet S given as S1, S2, Sq and identical first-order symbol probabilities PS1, PS2, PSq and denote this in abbreviated form P1, P2 and Pq.

Now, we are interested in calculation of relationship between HS and HS bar. Now, to calculate this relationship which we said, that would be something of the form HS, is always less than equal to HS bar. I will assume that, my source S is a first-order Markov source. So, for a first-order Markov source, my source alphabet, let me assume is this, the conditional probabilities, which will be involved would be Si given, Sj for i j equal to 1 2 up to q. So, I had the conditional probabilities, I had the source alphabet and now, let us try to calculate what is the entropy for this source?

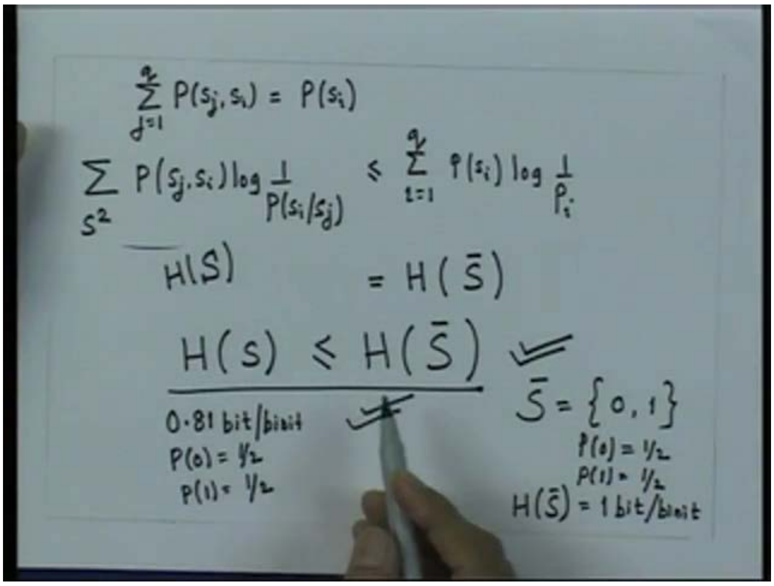
(Refer Slide Time: 33:17)



Let us define probability of Sj, Si as the joint probability, this joint probability that the source, is in the state specified by Sj and Si, occurs. Signing the state Sj and then Si occurs, when such a thing happens this is the probability of Si, occurring when I am in a state Sj. And this is denoted as joint probability now, let us try to examine the double summation of the form. This domain double summation is over Sj and Si so listen right. Now, we have seen earlier that there is a relationship of the form xi log of yi over xi, i equal to 1 to q is always less than equal to 0. Where xi and yi are the set of probabilities now similarly, here this joint probabilities P Sj Si if I sum it over j and I, the value is 1.

Similarly, Pj multiplied by Pi if I sum it over j and I, it is always equal to 1, what it means from this relationship that, this quantity out here is always less than equal to 0. Now, we also know from the base rule that probability of Sj Si is equal to probability of Si given, Sj multiplied by probability of, if you plug-in this expression into this then I can write as Pi over, probability of Si given Sj is less than equal to 0, which I can further simplify as, probability of Sj Si log of probability of Si given, Sj will be less than equal to, probability of Si Sj log of 1 by Pi. If you look on the right-hand side and concentrate on this term, this is nothing but equal to double summation of probability of Sj Si log of 1 by Pi, the summation is over I is equal to 1 to q, j is equal to 1 to q.

(Refer Slide Time: 37:46)



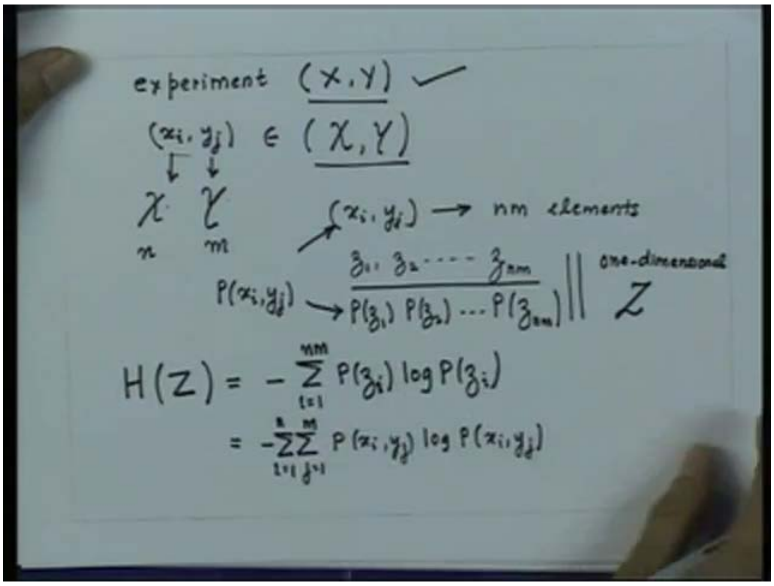
And this summation, that is probability of Sj Si, j equal to 1 to q, is nothing by probability of Si. So finally, what I get is, i equal to 1 to q probability of PSi log of 1 by Pi. This quantity is nothing by our definition as, entropy of HS bar, and this quantity out here, it is very easy to identify in the light of what we have done, earlier today. This expression which we derive for entropy of an MTh order Markov's source, in our case a source is first-order so this become S2 and this expression simplifies, to this expression. So, this expression is nothing but entropy of Markov source so from this relationship we get as H of S is nothing but less than equal to H of S bar.

This we have shown it for a first-order Markov source, but this is also valid in general, for any order Markov source. Going back to a previous example, HS got as 0.81 bit per minute and for the same example, we calculate the first-order symbol probabilities were P0 half and P1 again half. So, my adjoint source for the example which we consider earlier, would be S bar would be nothing but I have a source alphabet consisting of 0 and 1. And my probability of 0 would be the same as my unconditional symbol probabilities, for the source S and that is, nothing but equal to half. And probability of 1 again is equal to same identical to this so it is a half and to calculate entropy for this is simple H of S bar is 1 bit per minute.

So, what it shows that this expression is valid, this we have shown for our example discussed earlier. Now, to go little further and find out little more in-depth about MTh order Markov source, I need to define some few more definitions. With that intention, let us consider a following example suppose, I carry out an experiment, this experiment consist of tossing of a die and a coin simultaneously.

So, when I toss a die and coin simultaneously and if I am interested, in the outcome of this experiment then I will have two outcomes for this experiment. One is basically, the number which I can get on the die and the other is basically, either its head or tail. So, I can formulate in many cases, experiments which have two outcomes.

(Refer Slide Time: 42:40)



So, I will say that experiment having two outcomes denoted by let us say x and y. So, x and y this way which the way I have given here denotes, an experiment. Now, the outcome of these experiments are the values xi and yj, they will be occurring jointly. xi comes from a sample space, let us denote it as, this way and let us denote yj coming from a sample space x y like this.

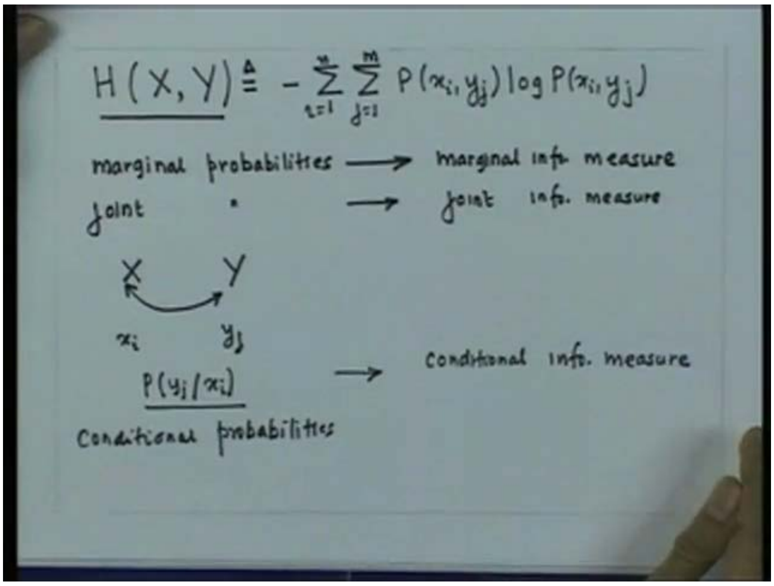
So, for this experiment x y, I have the elements xi yj belonging to a sample space, given by this. In the way I would like to define some kind of information measure to, the outcome of this experiment where the outcomes are not a single events. But they are two events so how do I define information measure for such an experiment.

It is very simple to do, if you follow the following procedure now, the number of events which can occur given by xi yj will depend on the size of the sample space x. And the sample space y, let me assume that the size of the sample space x is, n elements and the size of the sample space y is m elements. So, totally I will have nm elements, let me indicate these nm elements given by the joint occurrence of this two events, as z1, z2 up to znm. Now, with each of this event xi yj, I will have probability of occurrence given by probability of xi given yj. Now, in the light of our definition, of in terms of z this is nothing but equivalent to probability of z1, probability of z2 and probability of znm1.

So, an experiment which had two outcomes has been reduced to, an experiment which consists of single elements, given by z1 to znm1 with its probability of occurrence. Now, this has been reduced to a one-dimensional sample space, let us call this one-dimensional sample space as Z. Now, if I were to find out what is the information, associated with this experiment.

Then, going back to our definition I can define the information measure for this experiment is nothing but given by this simple expression, where my I goes from 1 to nm. So, this is the information measure associated with this experiment Z now, this is going back to our original notation. I can write this as nothing but p of xi yz log of P of xi yz and my summation will be double summation, I equal to 1 so I have been able to calculate the information measure for this experiment, my original experiment which was given by this sample space.

(Refer Slide Time: 48:16)

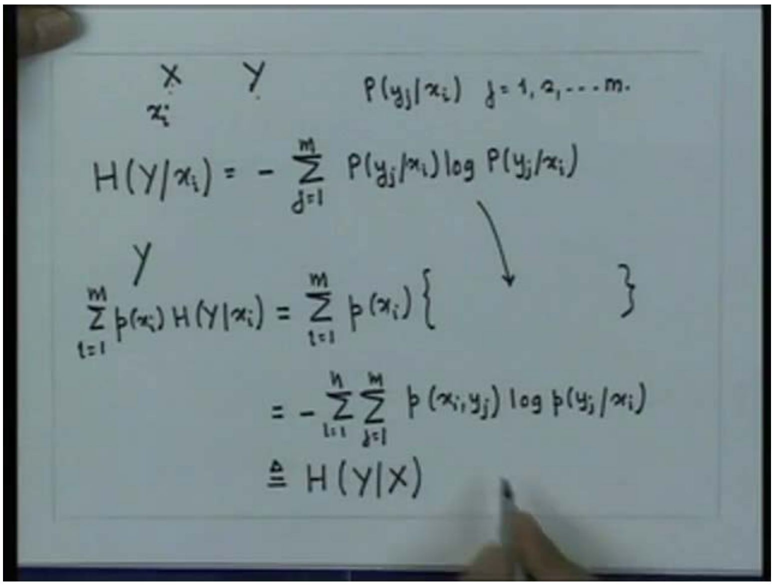


So, I can use this argument to define, my information measure for joint events as x y equal to, I is equal to 1 to n, j is equal to 1 to n. So,. This is the information measure which I will define for my experiment consisting of two events. So, we have seen that if I have marginal probabilities or just simple first order probabilities or unconditional probabilities. Then with marginal probabilities, we had associated an information measure. And that was marginal information measure, that is equivalent to the self-information, which we had defined earlier. Now, we have seen that if I have joined probabilities, in the form of P of xi yj, I have joined probabilities then I can define joint information measure.

This is joint information measure so question that arises is that, supposed to have two different experiments x and y and possibly the outcome of this two different experiments are related in some manner. And the question arises, is it possible for me to define something, what is known as conditional information measures because I can always associate conditional probabilities, with those two experiments x and y. For, let us assume that the elements of this experiment x is xi and elements of the experiment y are yj.

Then, we have probability of yj given xi, if that the relationship existing between yj and xi then I can define my what is probability of yj given, xi. If these are given to me then is it possible for me to calculate it. This is nothing but conditional probabilities so these are conditional probabilities. Then based on this conditional probabilities, I can also define what is known as conditional information measure. Let us look into definition of conditional information measure.

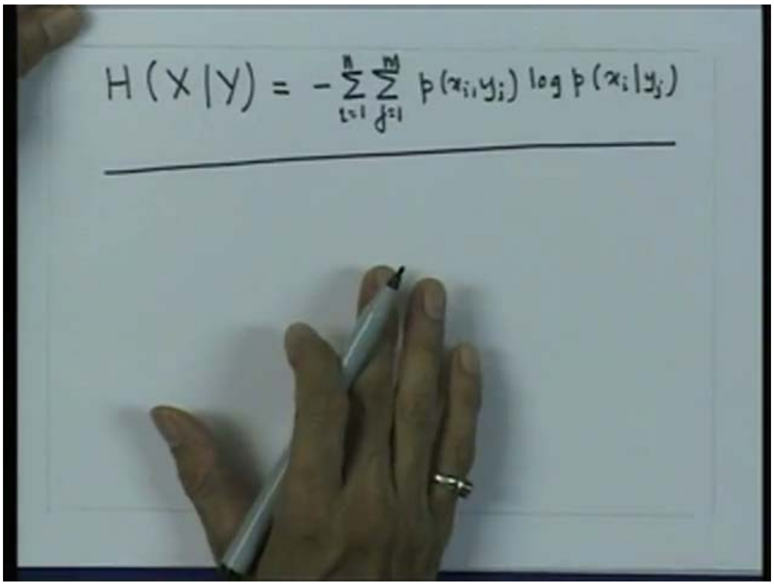
(Refer Slide Time: 52:08)



To understand this, let us assume that I have two probabilistic experiments x and y. And I am interested in amount of information with regard to y, under the condition that outcome xi has already occurred. And then since we have been given conditional probabilities and also. So, let us calculate by definition the average information, which I will get about the experiment y. When xi has occurred would be, by definition given as y of x i is nothing but probability of yj given xi log of probability of P given, xi j is equal to 1 to n. Now, by averaging this quantity over all xi then what I get is the average amount of information of y given, for knowledge of x, that is very simple.

What I have do is take the average of this quantity so average of this quantity is nothing but P of xi H of y of xi, i is equal to 1 to n. And if we just write this expression, is equal to 1 to m and substitute this expression out here, we can simplify this very easily to, i is equal to 1 to m j is equal to 1, probability of xi yj log of probability of y j given xi. I have changed my notation from capital P to small p they are same.

(Refer Slide Time: 55:23)



So, this is what I will get so this is the amount of average information of y given foreknowledge of x and this by definition, I will call it as H of y given x. Based on this argument, I can also define what is my H of X given, Y and that would be simply nothing but P of xi yj log of P of xi given, yj, i is equal 1 to n, j is equal to 1 to n. This is the average amount of information of x given, foreknowledge of y. Now, based on all this definitions, will try to probe a little further into emit order Markov source.