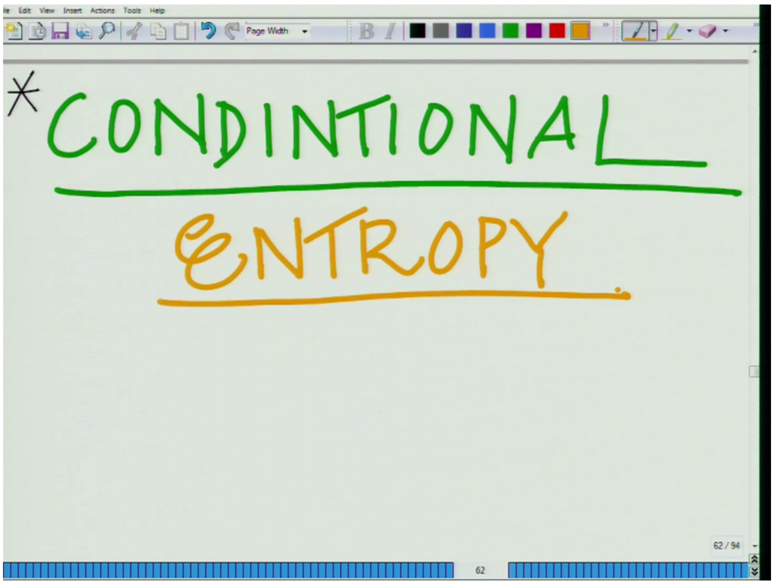
Principles of Communication Systems - Part II
Prof. Aditya K. Jagannatham
Department of Electrical Engineering
Indian Institute of Technology, Kanpur

Lecture – 32
Conditional Entropy, Example of Conditional Entropy, Properties of Conditional Entropy

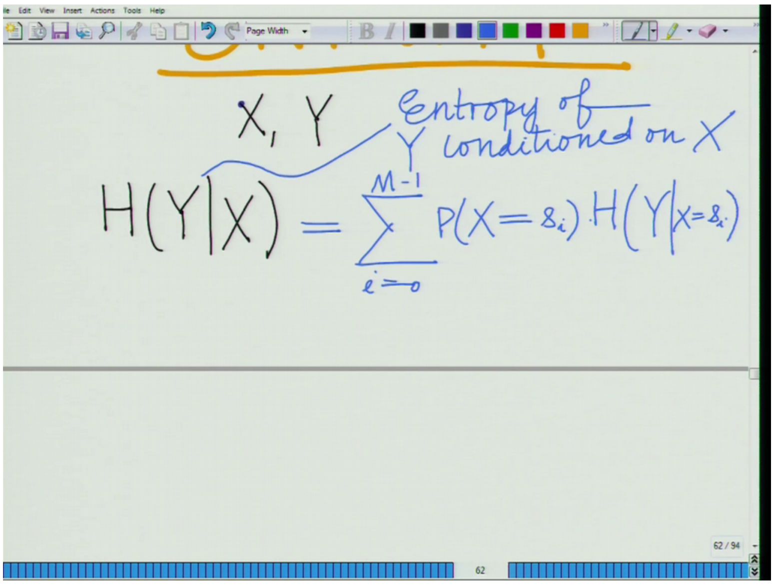
Hello, welcome to another module in this massive open online course. So far we have seen several concepts in information theory; we have seen the definition of entropy, joint entropy.

(Refer Slide Time: 00:30)



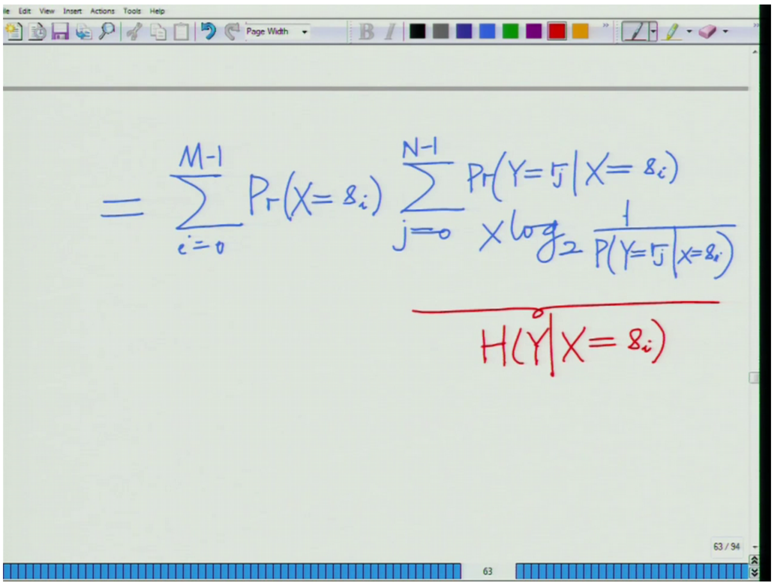
Let us look at yet another related quantity which is the conditional entropy. So, we will start looking at another important quantity in information theory that is the conditional entropy. Entropy as you have already said characterizes it is a measure of information, correct? It is a conditional entropy. So, let us say we have again 2 sources X and Y, we have X and Y then H of the conditional entropy remember, similar H of Y conditioned on X.

(Refer Slide Time: 01:12)



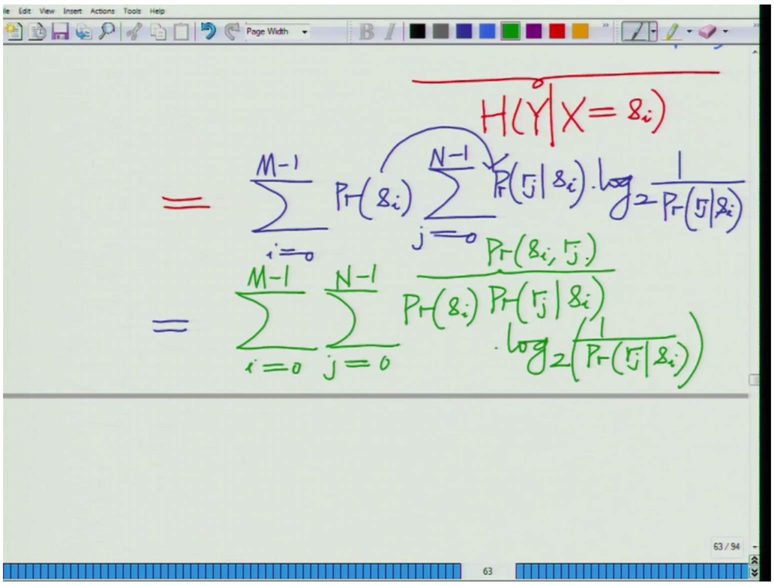
That is entropy of Y conditioned on X equals well summation i equal to 0, i equal to 0 to M minus 1. Probability of X equal to s i times the entropy of Y conditioned on X equal to s i which is equal to well.

(Refer Slide Time: 02:09)



Let us look at how these quantities are defined, which is equal to i equal to 0 to M minus 1 probability X equal to s i time summation j equal to 0, correct? J equal to 0 to M minus 1 probability of Y equal to r j conditioned on X equal to s i times log to the base 2 1 over probability, that is this is the entropy of Y given X equal to s i. So, we are using the conditional probabilities. So, this is the entropy of Y conditioned on X equal to s i this is the probability of X equal to s i. So, this is basically the entropy of Y condition on the source of X or entropy of Y given X.

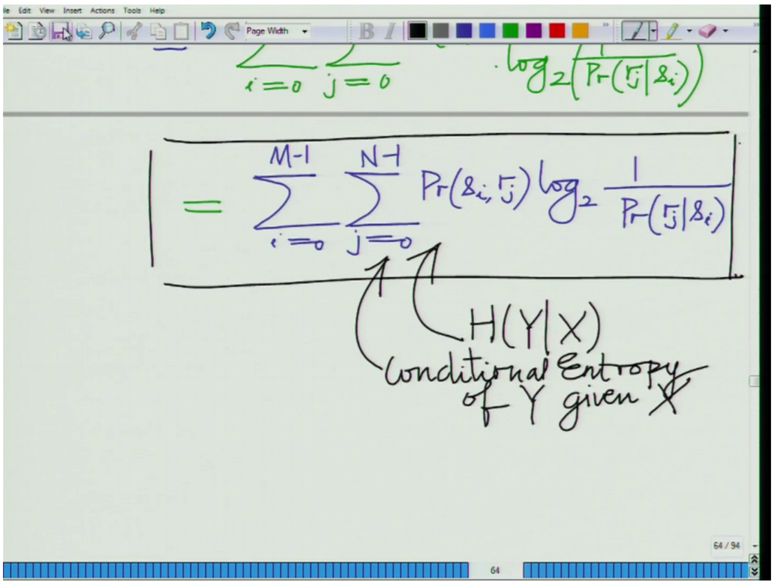
(Refer Slide Time: 03:35)



Now this can also be written as follows. So, this can let me just simplify the notation little bit this is equal to i equal to 0 M minus 1 probability of s i time summation j equal to 0 to N minus 1 probability of r j given s i times log to the base 2 1 over probability of r j given s i.

Now, if you take probability of s i inside and multiply this take probability of s i into probability of r j given s i that is Y equal to r j given X equal to s i that is simply probability of s i comma r j. So, this will be summation i equal to 0 summation j equal to 0 to N minus 1 probability of s i into probability Y equal to r j given X equal to s i times log 1 by probability r j given s i, and of this quantity is basically probability of s i comma r j. That is X equal to s i and Y equal to r j.

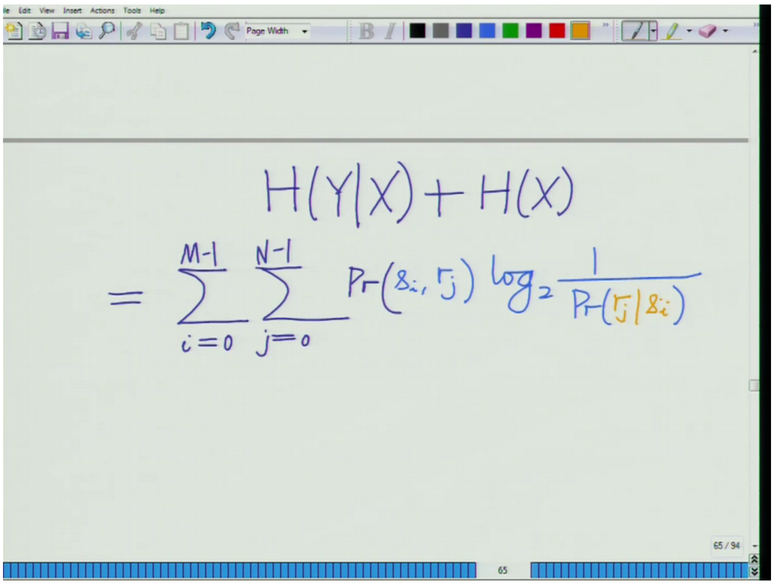
(Refer Slide Time: 05:14)



So, this is equal to sum i equal to 0 to M minus 1 sum j equal to 0 to N minus 1 probability s i comma r j log to the base 2 1 over probability r j given s i. This is the quantity conditional entropy, entropy of Y given X. This is the conditional entropy of, conditional entropy of this is the entropy of Y given X.

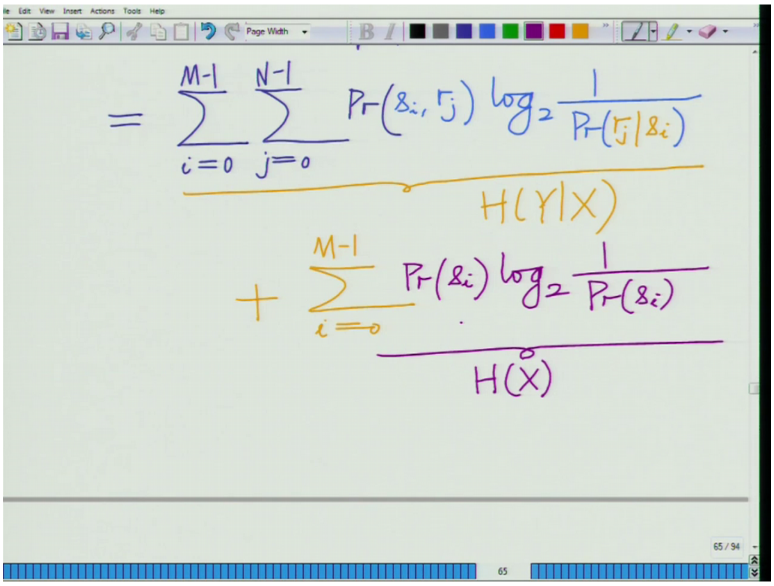
So, this is the conditional entropy of so, this is basically the uncertainty in Y, correct? The uncertainty in Y given X having observed X, or having given the source X, what is the uncertainty in Y, what is the uncertainty in Y. So, this is the in definition of conditional entropy.

(Refer Slide Time: 06:49)



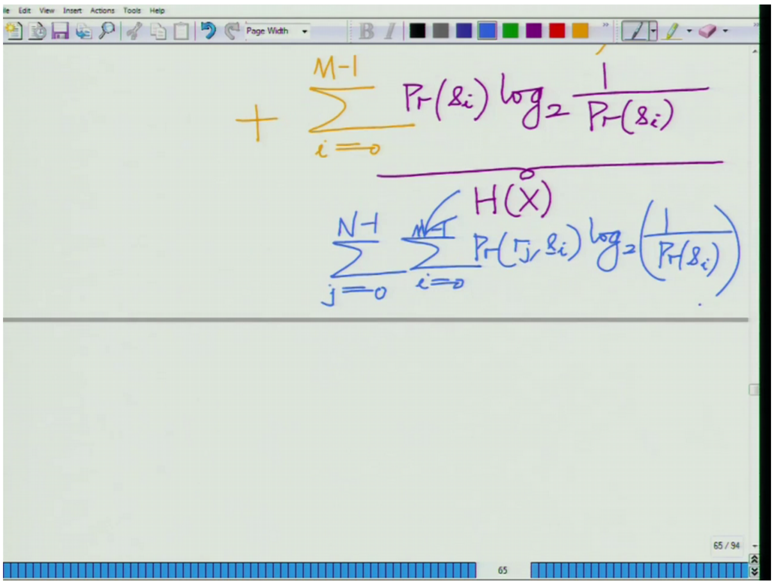
Now let us look at an interesting property of this quantity conditional entropy. H of Y given X plus H of X I would like to look at what this quantity is this quantity is summation first time I am going to write H of Y given X this is probability of s i comma r j log 2 to the base 1 by probability, I am sorry r j given s i.

(Refer Slide Time: 07:40)



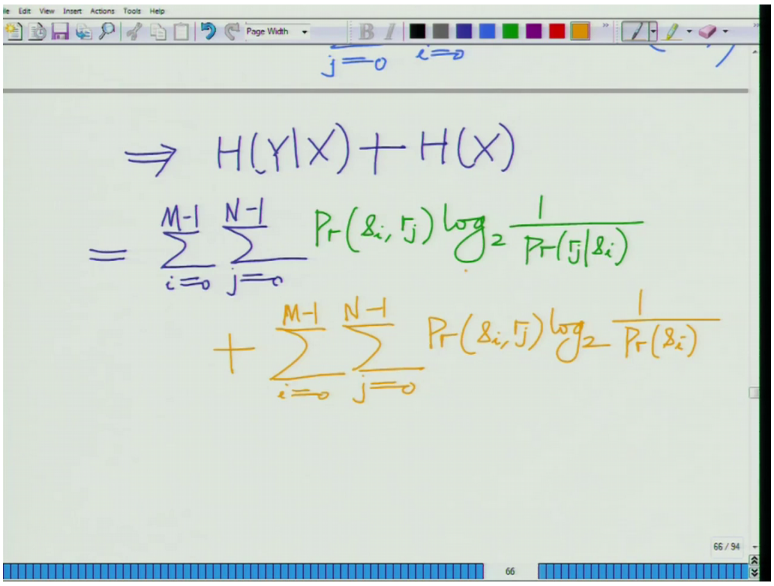
This is basically your H of Y given X plus summation i equal to 0 to M minus 1 probability entropy of X. This we have seen several times before probability of s i times log 2 to the base probability of s i summation over all i.

(Refer Slide Time: 08:24)



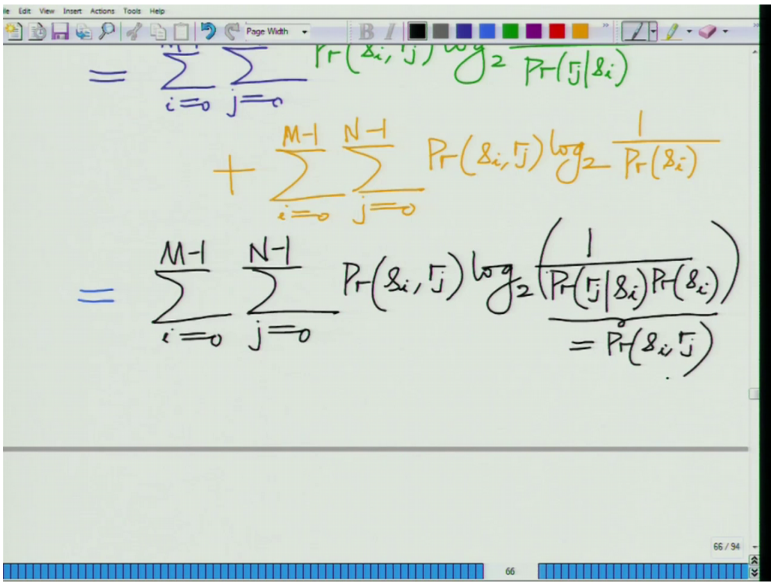
So, H Y given X plus H of X. Now look at this, this entropy of X I can also write as, H of X I can also write this as remember probability of s i is nothing but probability of s i r j summed over all j. So, I can also write this as well, j equal to 0 N minus 1. I can also introduce a summation over all j i equal to 0 to M minus 1 probability of r j s i 1 over probability of s i. And therefore, H of Y given X plus H of X you can see ah,

(Refer Slide Time: 09:03)



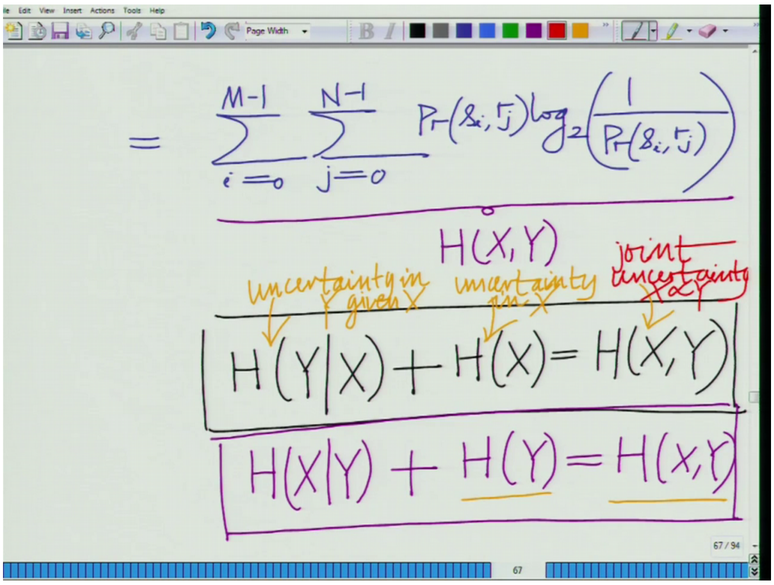
This was this is basically summation i equal to 0 to M minus 1 summation j equal to 0 to N minus 1, probability s i comma r j log 2 to the log to the base 2 1 over probability r j given s i plus summation i equal to 0 to M minus 1, summation j equal to 0 to N minus 1 probability s i comma r j log to the 2 to the base 1 over probability s i.

(Refer Slide Time: 10:18)



And now if you add these 2, now you can see I can take probability of s i r j common. So, that with the summation i equal to 0 to M minus 1 summation j equal to 0 to N minus 1 probability s i comma r j and when you add the 2 logarithm the arguments multiply. So, this is log 2 1 over probability if r j given s i into probability of s i. And now observe that probability r j given s i into probability of s i this is equal to probability of nothing but s i comma r j. That is intersection of s i comma r j and therefore, this is going to be.

(Refer Slide Time: 11:06)

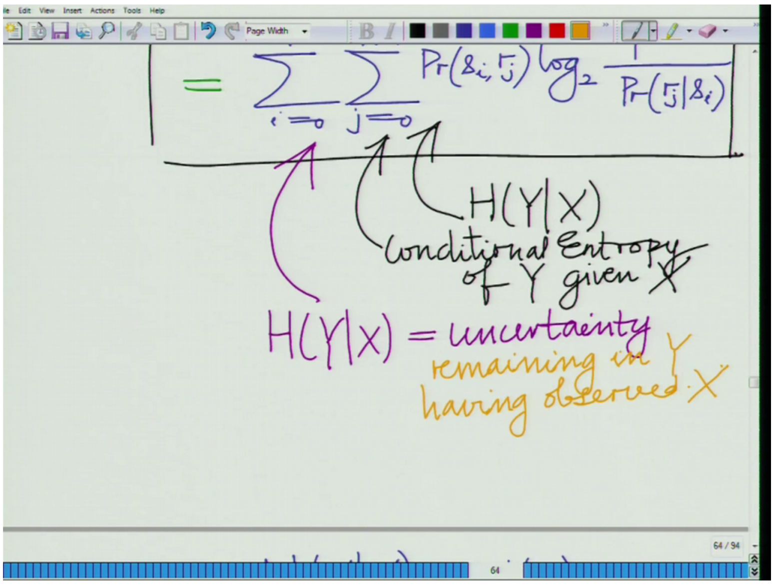


Now, you can observe this is summation i equal to 0 to M minus 1 summation j equal to 0 to N minus 1 r j equal to 0 to N minus 1 probability s i comma r j log 2 to the base 1 over probability s i comma r j, and if you can recollect this is nothing but the joint entropy of X comma Y.

So, the very interesting relation that we have here is that H Y conditional entropy of Y conditional entropy X H, Y given X plus H X is nothing but the joint entropy. This is the conditional entropy H Y given X plus H X equals H X. Similarly you can also show naturally, is nothing special about X and Y. So, if I reverse the rules of X and Y H X given Y plus H X of Y is also equal to H, this is another interesting property of this is basically another way to express again. So, I have H Y given X plus H X is equal to the joint entropy H X comma Y, which is the same as H X given Y plus H of Y and. So, we have also seen H Y given X is the conditional entropy you can also see this is basically nothing but the uncertainty remember entropy is the uncertainty, entropy is the uncertainty.

So, conditional entropy is the uncertainty remaining in Y. One can say so, you have conditional entropy is nothing but the uncertainty remember is the information conveyed by a source or is a uncertainty in a source. So, we can say the conditional entropy is the uncertainty remaining in Y having observed X.

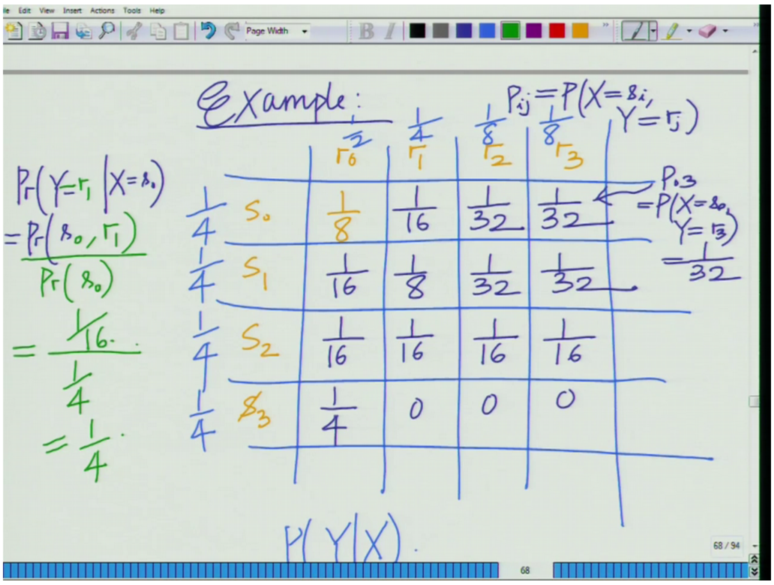
(Refer Slide Time: 13:15)



So, this is H Y given X equals uncertainty remaining, the uncertainty remaining in Y having observed X. And H X given Y and of course, this is a joint information in X and Y this information. So, uncertainty so, let us look at it this way let us write it over here. Uncertainty remaining in Y given X, H of X is uncertainty in X this is joint information or joint uncertainty in X and Y, joint uncertainty in X and Y. So, uncertainty in X plus the uncertainty remaining in Y given X, that is a joint information of X and Y joint entropy of X comma Y. So, we have this interesting property.

Let us look at a simple example to understand this better all right. So, let us look at a simple example, let us go back to our previous example that we had where we had this table of probability X.

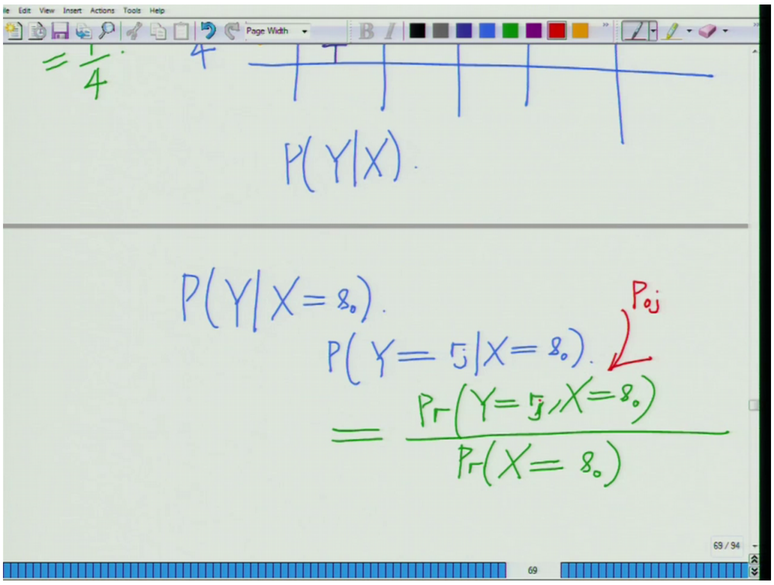
(Refer Slide Time: 15:16)



So, this is what you have, this is your s 0, s 1, s 2, s 3. This is your r 0, r 1, r 2 at 3 the joint probabilities are 1 by 8, 1 by 16, 1 by 32, 1 by 32 1 by 16, 1 by 8, 1 by 32, 1 by 32, 1 by 16, 1 by 16, 1 by 16, 1 by 16, 1 by 4. Now each entry p of i j remember equals p of X equal to s i Y equals r j. For instance this denotes p of 0 comma 3 equals p of X equals s 0 comma Y equals r 3 equals 1 by 32. And from this we have computed the marginal probabilities of each s 0 probability of X equal to s 0 is simply the sum of the probabilities in the corresponding row. That is in sum the probabilities of s 0 corresponding to each r j, that is p of s 0 r j is summation over j gives the probability of s 0 that is, we know we know that by the total probability route.

So, that is we have computed this before. So, probability of each s 0 all s zeros are equiprobable, probabilities 1 by 4 is probability of r naught is half r 1, 1 by 4 star 1 by 8 1 by 8.

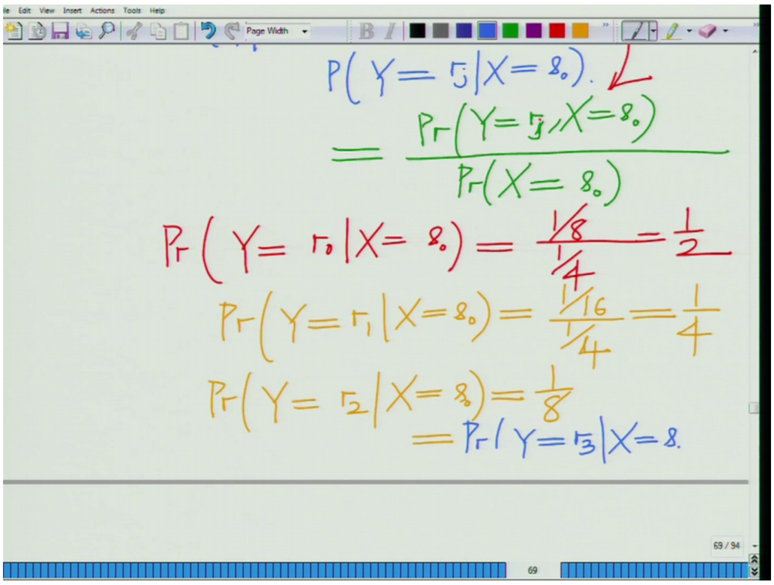
(Refer Slide Time: 17:37)



Now what we have to do is we have to compute, let us say the joint probability. So, we want to compute the probability of well Y given X. Let us first start by computing probability of Y given X equal to s 0. For that we will need the probability of Y equals r j given X equal to s 0. Now if you look at this for instance probability of Y equal to s 0, given X equal to s 0 equals probability given X Y equal to, sorry r, r 1. Let us say Y equal to r 1 Y equal to r 1 equals probability of s 0 comma r 1 divided by probability of s 0 comma r s 0 comma r 1, for instance that is 1 by 16 divided by, 1 by 16 divided by probability of s 0 probability of s 0 is 1 by 4 equals 1 by 4.

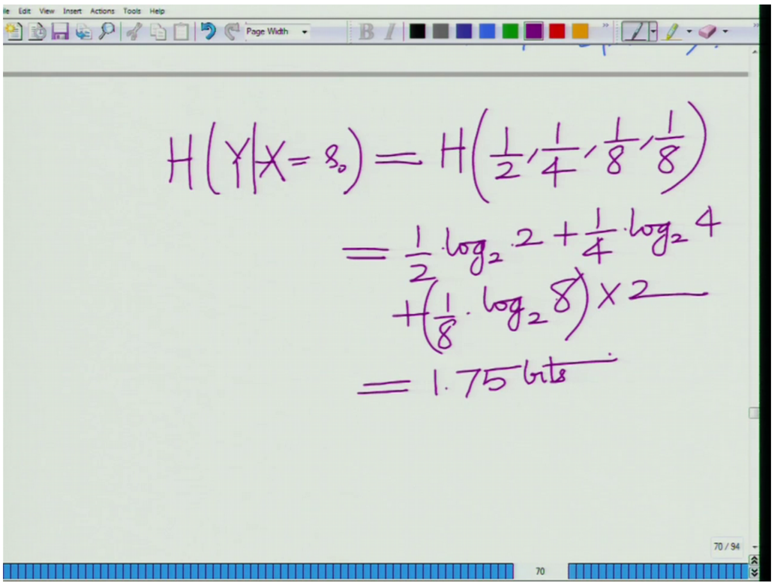
So, probability of remember we are using we are using the simple principle that probability of a given b is probability of a intersection b divided by probability of b. We already know the for various joint probabilities, for instance if we want to compute probability of s 0 r j given X Y equal to r j given X equal to s 0 that is simply equals probability of Y equals r j comma X equal to s 0, given the probability that X equals s 0. We already know these quantities, this is nothing but p 0 comma j that is probability of X equal to r 0 comma Y equal to r j. And probability of X 0 we already know this probability these are the marginal probabilities.

(Refer Slide Time: 19:59)



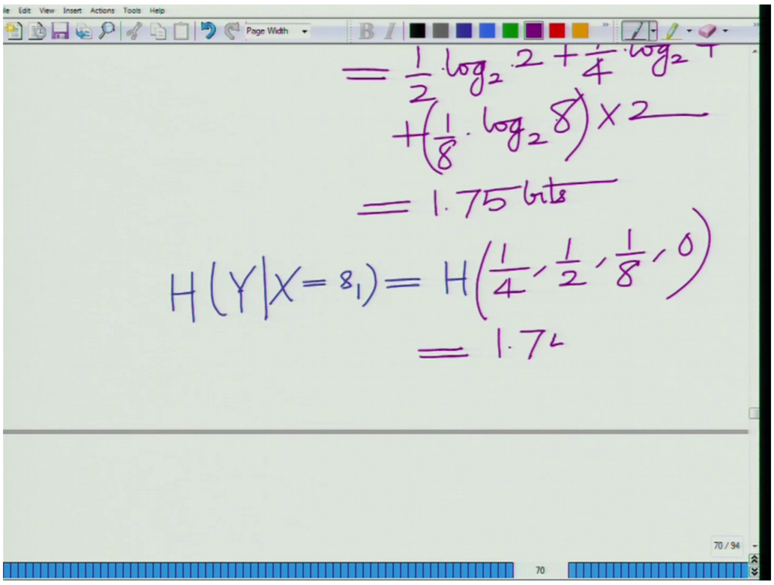
So, we see for instance probability of Y equal to r 0 given X equal to r 0 that is the probability of r 0, s 0 that is 1 by 8 divide the probability X equal to s 0 that is 1 by 4. So, this is 1 by 8 divided by 1 by 4 equals well half similarly probability Y equals r 1 given X equals s 0. We have already calculated that that is basically 1 by 16 divided by 1 by 4 equals, 1 by 4 probability Y equals r 2 given X equal to s 0 equals, well 1 by 32 divided by 1 by 4 1 by 8 which is also equal to the probability that Y equal to r 3 given X equal to s 0.

(Refer Slide Time: 21:10)



So, the entropy given, entropy of Y given X equal to s 0. Corresponds to entropy of a source with the 4 probabilities, probability Y equal to r 0 given X equal to s 0 that is half Y equal to r 1 given X equal to s 0 that is 1 by 4 and the remaining probabilities that is 1 by 8 1 by 8, which is basically again, we confirm the entropy of the source this is basically half log to the base 2 1 over half that is 2 1 by 4 log to the base 2 4 plus 1 by 8 log to the base 2 8 and find there are 2 symbols of probability 1 by 8. So, into multiplied by 2. So, this is 1.75 bits. Similarly one can calculate H of Y.

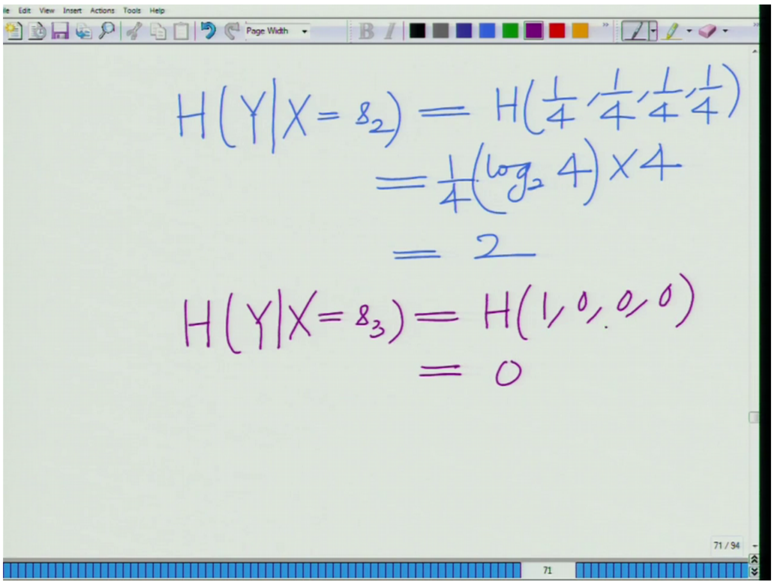
(Refer Slide Time: 22:09)



So, we have so, using the conditional probabilities of the various symbols Y equal to r 0, r 1, r 2, r 3. Conditioned on the fact that X equal to 0 we are able to calculate the conditional entropy of Y given X equal to s 0.

Next the conditional entropy of Y given X equal to s 1. This is equal to the entropy corresponding to the probabilities well 1 by 4 again the same 1 by 4, 1 by 2, 1 by 8 comma 0 which is equal to 1.75, H of Y given X equal to s 2.

(Refer Slide Time: 22:50)

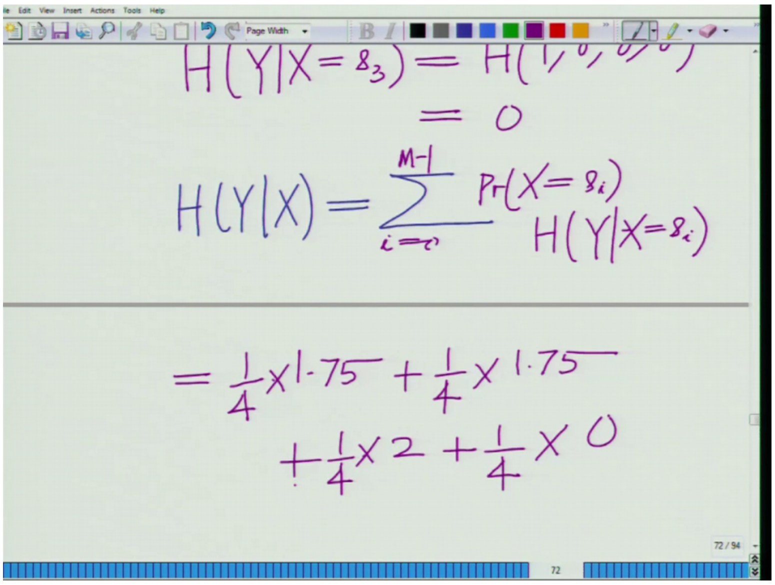


This is equal to H of, this probabilities will be equal 1 by 4, 1 by 4, 1 by 4, 1 by 4 because, if you look at given X equal to s 2 all the symbols r j the joint probability of all symbols is 1 by 16. So, each probability that is Y equal to that Y equal to r j given X equal to s 2 will be 1 by 16 divided by 1 by 4 that is 1 by 4.

So, the conditional probabilities will all be 1 by 4 in this case and the joint the entropy will be well, 1 by 4 log to the base 2 4 into there are 4 symbols with probability 1 by 4 So, this is basically 2. And finally, H of Y given X equal to s 3 this will be equal to H of 1, 0, 0, 0 because you can see in this case the joint probability of all symbols other than the symbol r 0, r 0. So, there conditional probability of the 0, conditional probability of X equal to Y equal to r 1, r 0 given X equal to X 3 will be 1 by 4 divided by 1 by 4 that is equal to 1 ok.

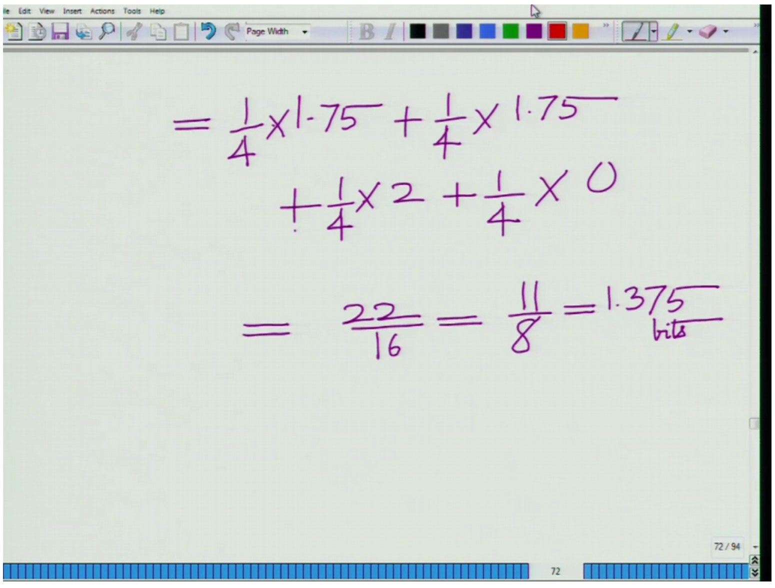
So, that is an entropy of a source with probabilities 1, 0, 0 and this is basically certain which means given X equal to s 3 the only symbol Y can take will be r, r 0, because the rest all have 0 probabilities. So, there is no uncertainty which means this conditional entropy given X equal to s 3 will naturally be 0. And that you can get also from this calculation corresponding to probabilities 1, 0, 0, 0.

(Refer Slide Time: 24:42).



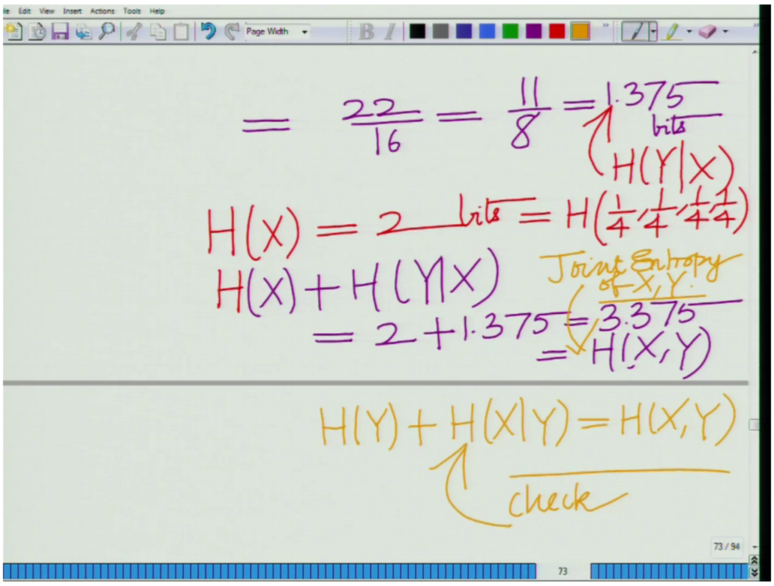
And therefore, finally, we have H of Y given X equals summation well i equal to 0 to M minus 1 M equal to 4 in this case probability X equal to s i into entropy H of Y given X equal to s i which is equal to well, that is equal to probability of X equal to s 0 is 1 by 4 into entropy X given Y given X equal to s 0 that is 1.75. Plus probability of X equal to s 1 is 1 by 4 into entropy Y given X equal to s 1 that is 1.75, plus 1 by 4 probability of X equal to s 2 entropy Y given X equal to s 2 which is equal to 2 plus 1 by 4 it is probability X equal to s 3 into entropy Y given X equal to s 3 is 0.

(Refer Slide Time: 25:48)



And if you calculate this you will get the answer this is equal to 22 by 16 equals 11 by 8 which is equal to well, this is equal 22 by I think this equal to 1.375 because this is equal to 1.375.

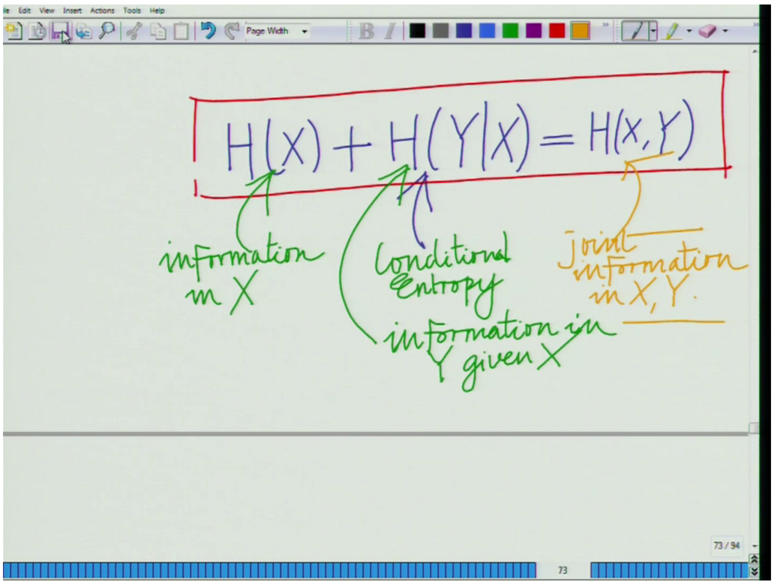
(Refer Slide Time: 26:14)



And now if you remember, H of X equal to 2 we have derived this is basically this quantity is H of Y given X, all the source symbols have all the symbols of X of uniform probability 1 over 4. So, H of X equal to 2 we have calculated this H of X equal to 2 this is equal to H of because all the marginal probabilities are equal to 1 by 4, and you can check that H of X plus H of Y given X is equal to 2 plus 1.375 equals 3.375 which is indeed equal to H of X given Y. We had also computed H of X given Y which is now the joint entropy of X comma Y.

Similarly, you can also check that H of well, Y plus H of X given Y should also be equal to H of X comma Y, you can also check this. So, you can check you can check this. So, basically that is what we have we have seen the definition of conditional entropy. So, in this module what we have seen is we have explored yet another concept that is the conditional entropy. The conditional entropy of Y given X and X given Y which is the conditional entropy of Y given X is the uncertainty remaining in Y having observed X.

(Refer Slide Time: 27:59)



So, we have H of X plus H of, the important property that we have seen H of H of Y given X equals H of Y. This is conditional entropy or information in Y uncertainty in Y given X. This is the information in X. This is the joint information in X comma Y; this is the joint information in X comma Y all right. So, we have seen conditional entropy, it is properties and an example to illustrate how to compute the conditional entropy all right.

So, we will stop here and continue with other aspects in the subsequent modules.

Thank you very much.