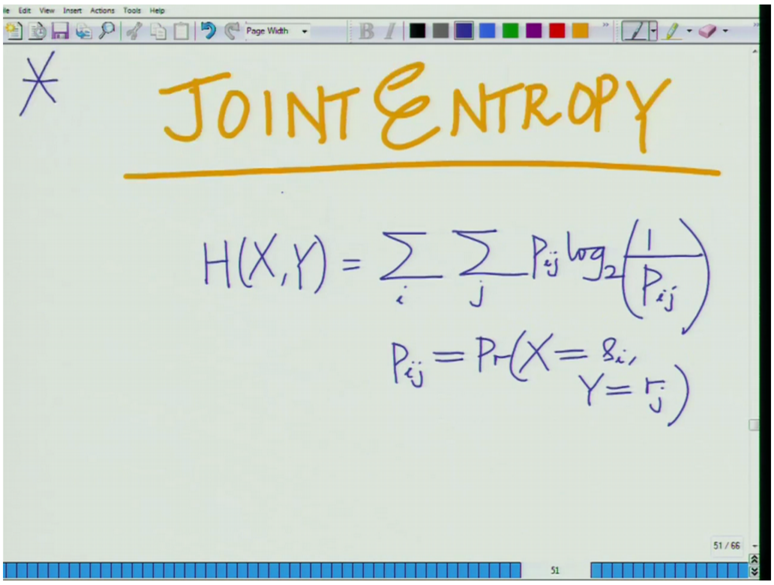
Principles of Communication Systems - Part II
Prof. Aditya K. Jagannatham
Department of Electrical Engineering
Indian Institute of Technology, Kanpur

Lecture – 31
Properties of Joint Entropy, Relation between Joint Entropy and Marginal Entropies

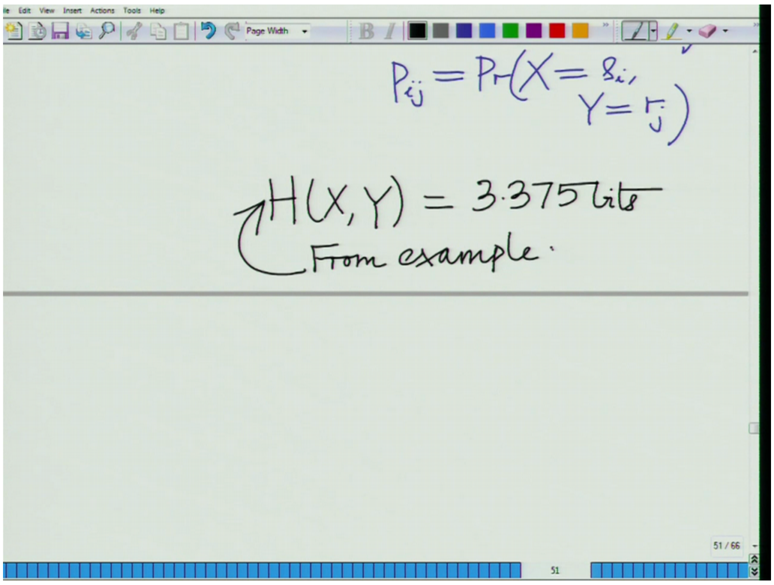
Hello, welcome to another module in this massive open online course. So, let us continue our discussion on joint entropy that is we are looking at the joint entropy of 2 sources that is H of X comma Y.

(Refer Slide Time: 00:27)



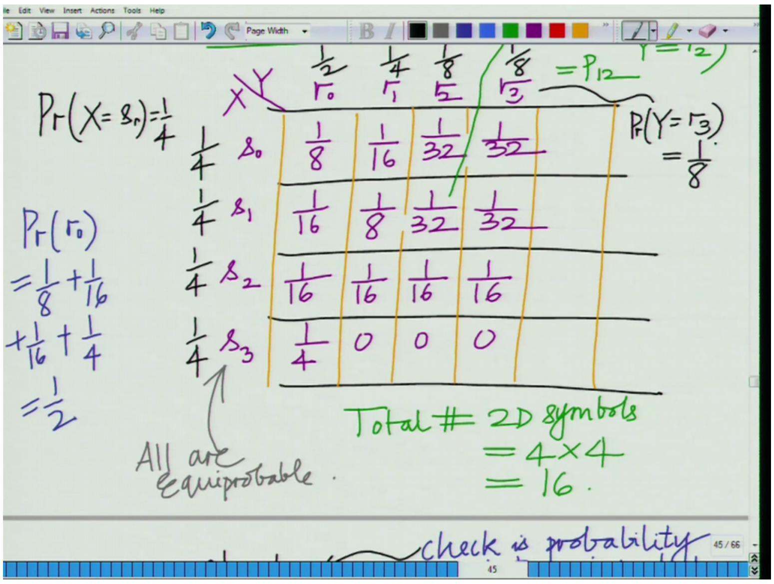
So, we are looking at the notion of joint entropy and we have seen that the joint entropy can be defined as H of X comma Y equals that is treating this as 2 dimensional symbols summation overall i comma j summation over all i j log to the base 2 1 over P of i j where P of i j equals the probability X equal to s i comma Y equals r j we have seen this joint entropy. We also calculated from our example that h X Y equals well that is equals 3.37 bits they also we have seen from example.

(Refer Slide Time: 01:23)



Now, let us make an interesting observation we have also seen the probabilities if you remember for the example we have also computed the probabilities of the symbols.

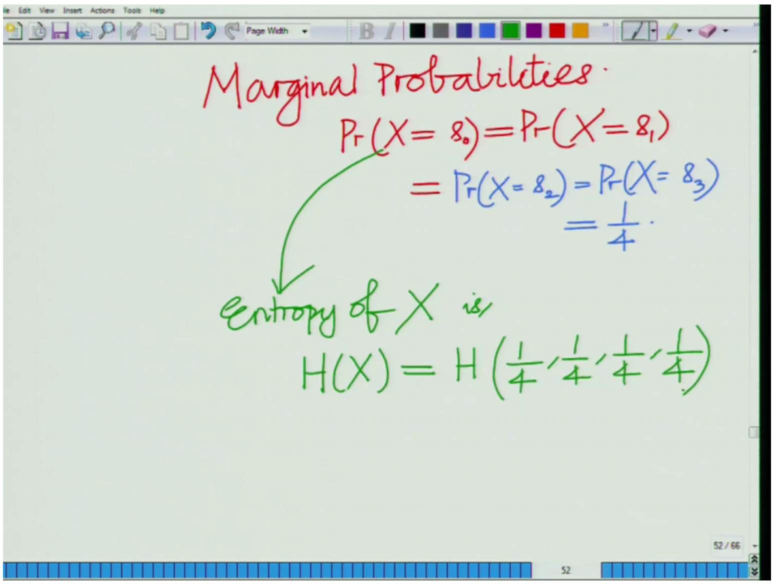
(Refer Slide Time: 01:54)



For instance probability of s you seen that s naught s 1 s 2 s 3 are equiprobable. So, the probability that X takes any of these symbols X is any of the symbols is 1 by 4 for instance probability X equals s naught equals 1 by 4 further we have also for instance the probability that Y equals r 3 equals 1 by 8.

So, if you look at X and Y individually the; these are known as the marginal probability. So, we are compute also computed the marginal probabilities for X and Y correct. So, we have know we know the joint probabilities P i comma j that is what we have called probability X equal s i.

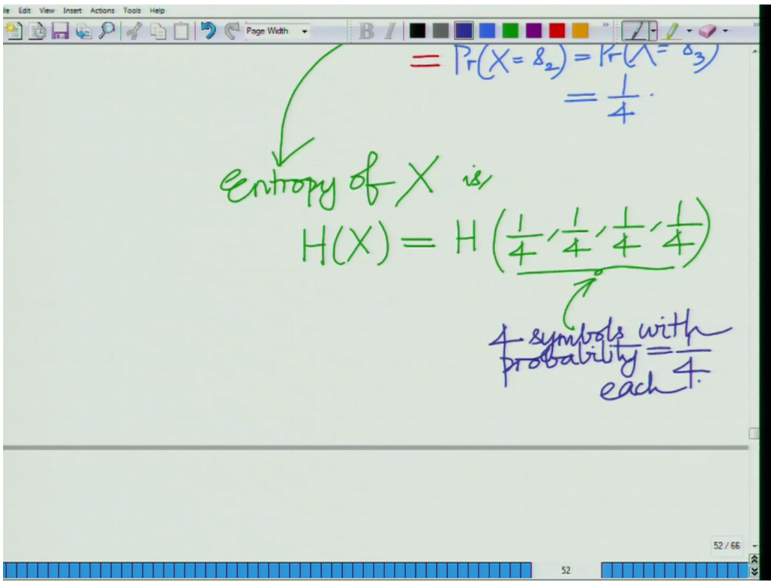
(Refer Slide Time: 02:42)



So, we have also known computed the marginal probabilities for instance probability X equals s 0 equals probability X equals s 1 equals probability X equals s 2 equals probability X equals s 3 equals 1 by 4.

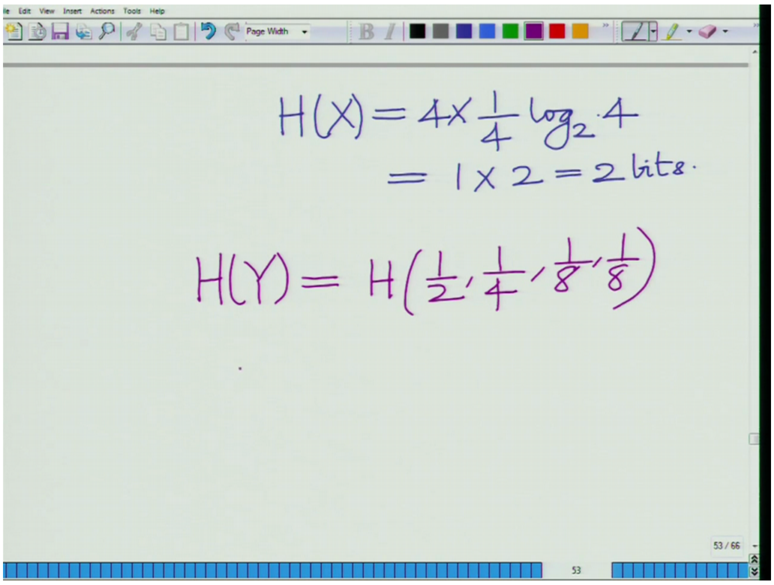
Hence the entropy of this source is H of X corresponds to simply the entropy of a source with symbols with probabilities 1 by 4 each correct I use this notation to denote that.

(Refer Slide Time: 03:53)



There are basically 1 by 4 each source with 4 symbols with which are equiprobable probability equal to 1 by 4 each which means we have H of X equals 4 times 1 by 4 log 2 to the base log to the base to 1 over 1 by 4 is log 4.

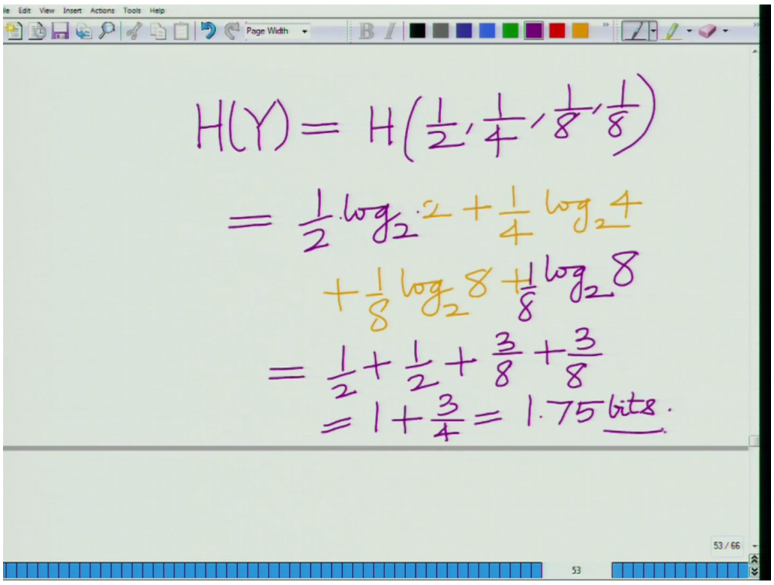
(Refer Slide Time: 04:14)



That is basically log 4 to the base 2. So, that is equal to 1 times 2 equals 2 bits.

Similarly, we have computed entropy of Y corresponds to the entropy of a source with the marginal probabilities we have calculate the marginal probabilities probability of r 0 is half r 1 is 1 by 4 r 2 and r 3 r 1 by 80. So, this entropy we have already. So, this entropy is half.

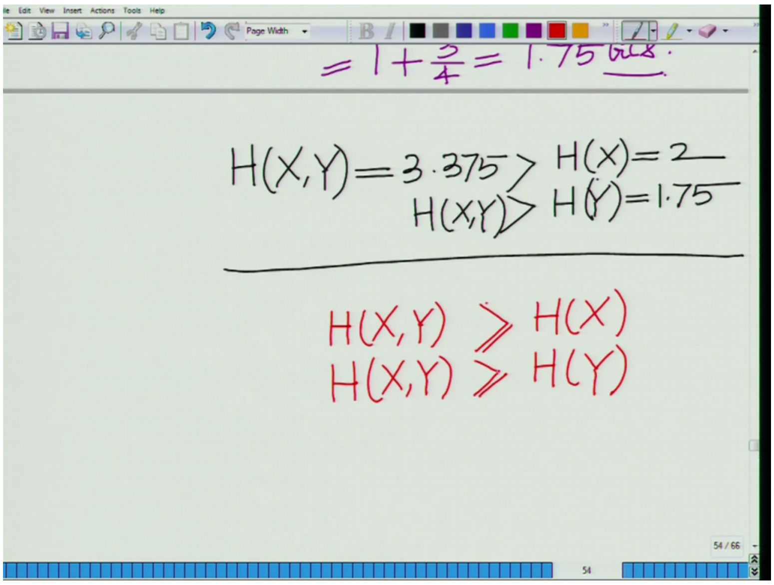
(Refer Slide Time: 05:04)



Let us just computed once more log to the base 2 1 over half that is basically log to the base 2 of 2 plus 1 by 4 plus 1 by 8 plus log 1 by 8 log to the base 2 8. So, this is basically half plus half plus 3 by 8 plus 3 by 8 equals basically 1 plus 3 by 4. So, that would be 7 by 4 which is equal to basically 1.75 bits.

So, we have calculated H of X and H of Y from the marginal probability there marginal probability mass functions we have seen that H of X that is the entropy of source X as 2 bits entropy of Y is 1.75.

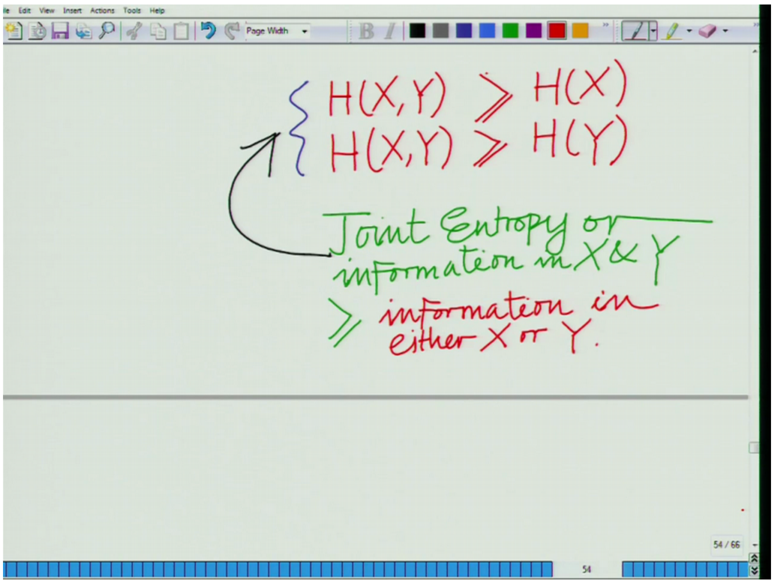
(Refer Slide Time: 06:41)



Now, you will observe something very interesting will realize 2 interesting properties one you can see H of X Y equals well 3.375 and this is greater than both H of X which is equal to 2 and it is also greater than H of X Y is also greater than H of Y which is equal to 1.75.

So, we have basically H of X comma Y is greater than H of X H of X comma Y is greater than or greater than equal to rather greater than equal to h or Y and this property can be shown and this is an interesting property.

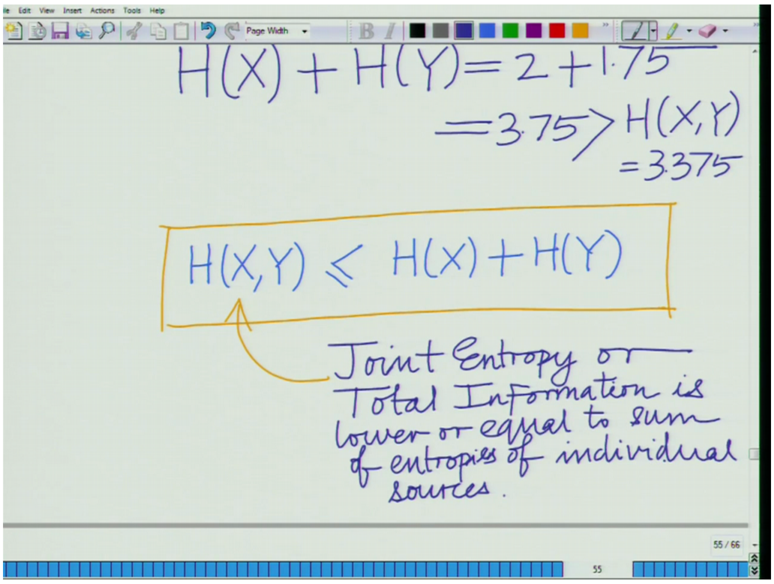
(Refer Slide Time: 07:48)



What this says is the joint entropy that is joint entropy of 2 sources is greater than or equal to the individual right the individual entropy of either source correct. So, if I look at the joint entropy of 2 sources the basically the; we said there entropy is nothing, but a measure of information right it is a means to quantify the information.

So, that the joint entropy of 2 sources cannot be smaller than the information in either source, so, when you looking at 2 sources together the amount of information they convey together cannot be smaller than the amount of information each either source either source conveys in isolation. So, that is an important point that is very intuitive, but it shows that this property follows through the definition of joint entropy that is joint entropy or information in X and Y is greater than equal to information in X information in either information in either X or Y, alright. So, the information the joint information cannot be lower than the information in the individual sources that is also intuitive and this property can be shown.

(Refer Slide Time: 09:32)

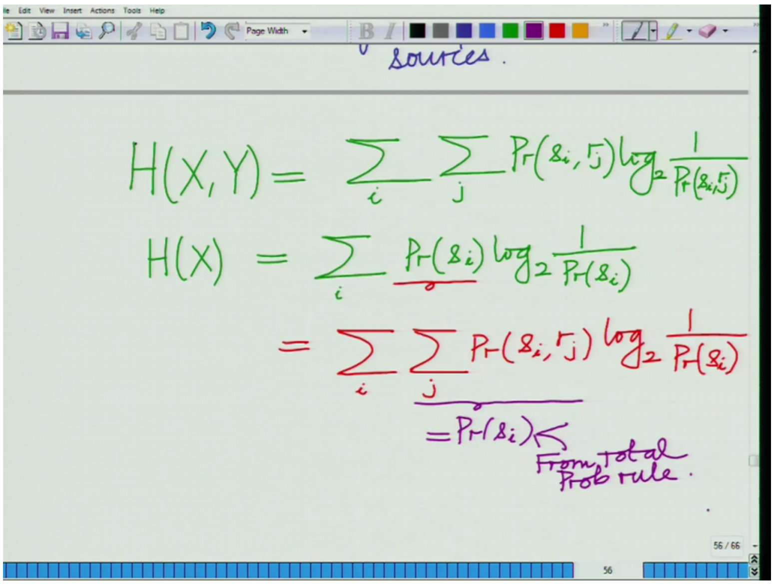


Now, something even more interesting is the fact that if you look at H of X plus H of Y this is equal to well this is equal to 2 plus 1.75 equals 3 point 7 five which is greater than H of X comma which is equal to simply 3.375 correct. So, what you are observing is H of X comma Y the property at your observing is at H of X comma Y is less than or equal to H of X plus H of Y and this is the very interesting property.

What this says is the joint information in X comma Y is less than the sum of the information in the individual sources. So, what you seeing previously is in the information is greater than the information and that of the joint information joint entropy is greater than that of the entropy in either source, but what this result is also telling us the next result is at the joint information is less than the sum of the component information that is its less than the sum of the entropies of the constituent sources. So, the total information joint entropy or total information is lower or equal to the sum of or rather the sum of entropies the sum of the entropies of the individual sources.

So, this is an interesting process of while the joint entropy is greater than the entropy of either source the joint entropy is less than or lower less than or equal to the sum of the entropies of the individual sources. Let us try establishing this result this is an, this is an important result.

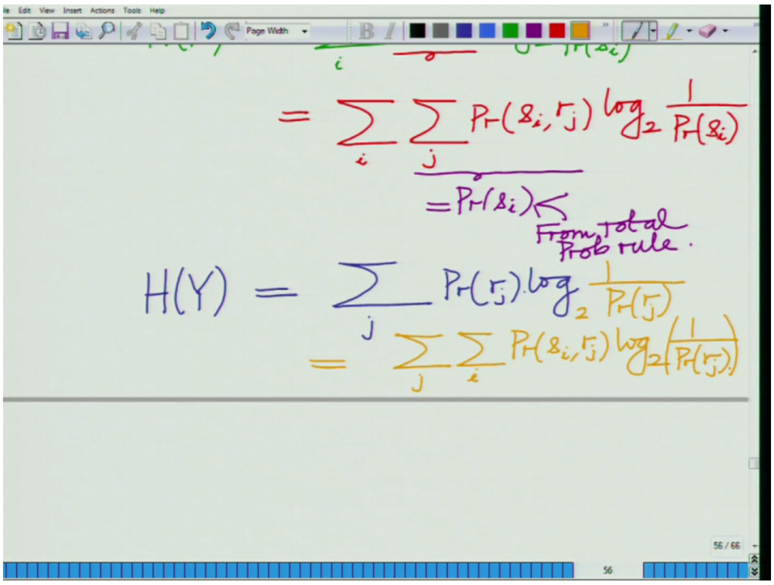
(Refer Slide Time: 12:16)



So, let us try establish in this we have well from our definition we have these are information theory contains several such interesting inequalities. So, we have H of X Y is summation over i summation over j probability of s i that is X equal to s i Y equal to r j i am simply writing a probability of s i comma r j 1 over log to the base 2 1 over probability of s i comma r j H of X equals summation over i probability of s i log to the base 2 1 over probability of s i now i can write this using total probability rule correct. I can write this using the total probability rule remember we have already seen that I can write this as probability of s i is nothing, but a summation probability over j. So, each s i I can also. So, each s i, I can also sum over j.

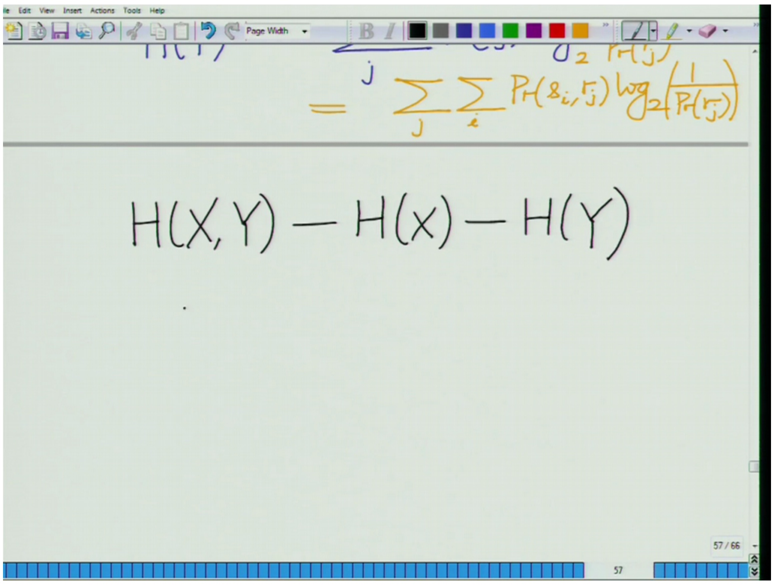
Sum over j probability of s i comma r j because times log to the base 2 1 over probability of s i because look at this; this quantity here is nothing, but this is equal to probability of s i this follows from total probability rule because we have seen that probability of s i each s i is nothing, but I can sum it I can take the joint probability s i comma r j and summit sum the probabilities corresponding to all possible r j. So, all possible values of the index j that is the follows from that is equal to the probability of s i and that follows from the total probability rule.

(Refer Slide Time: 14:51)

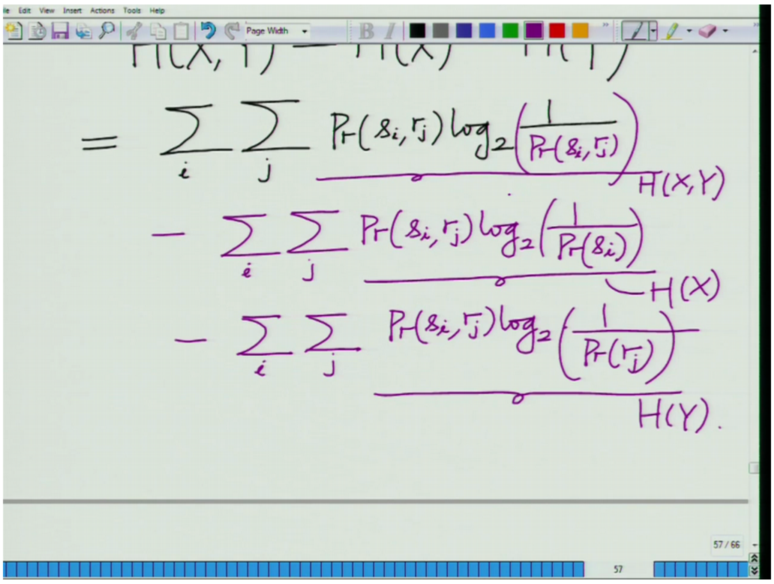


Similarly, we have H of Y is summation j probability of r j log to the base 2 1 over probability of r j now I can sum; I can write the summation over i using the total probability rule this is probability s i comma r j log to the base 2 1 over probability of r j.

(Refer Slide Time: 15:40)

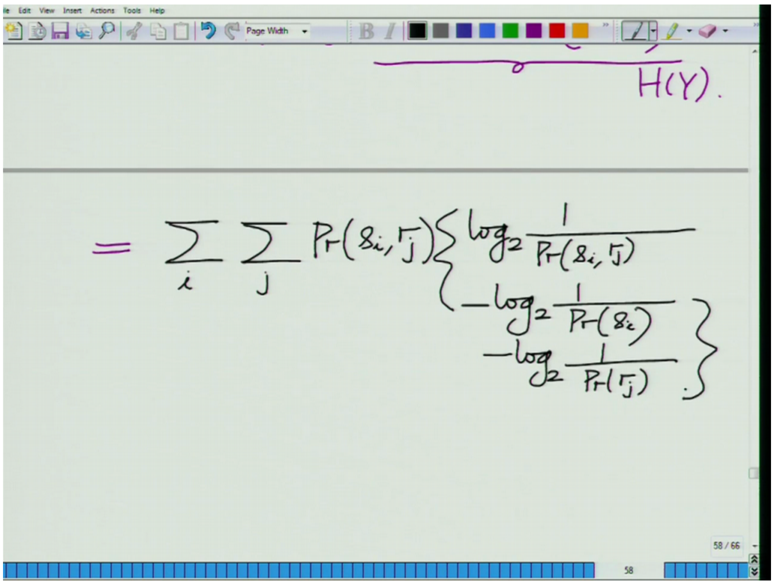


(Refer Slide Time: 15:55)



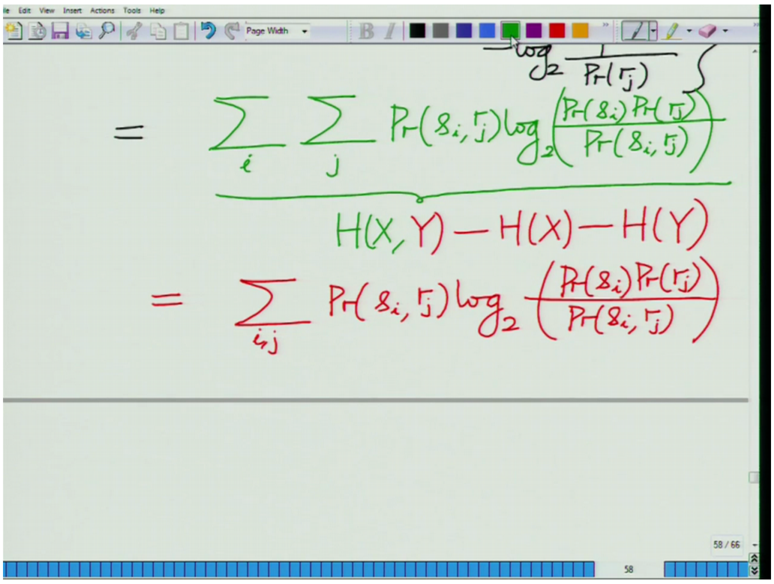
And now if you look at h X Y minus h X minus h Y you will see this is nothing, but well summation over i comma j probability s i comma r j log to the base 2 1 over probability s i comma r j minus. So, this is your h X comma Y minus summation over i comma j probability of s i comma r j log to the base 2 1 over probability of s i this is your h X minus sum over i comma j probability s i comma r j log to the base 2 1 over probability of r j and this is your h Y and now if can see if you can take this probability of s i r j common.

(Refer Slide Time: 17:28)



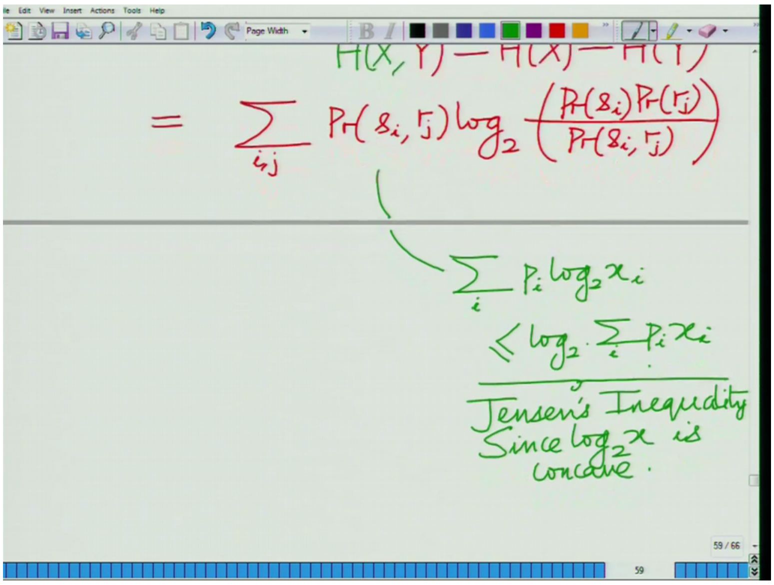
You will see this is nothing, but summation of over i comma j probability of s i comma r j into log to the base 2 1 over probability s i comma r j minus log to the base 2 1 over probability s i minus log to the base 2 1 over probability which is equal to well summation over equal to summation or i comma j probability of s i r j log 2 to the base probability of s i probability of r j divided by probability of s i r j.

(Refer Slide Time: 18:22)



so this is the final expression we have for H of X i am sorry H of X comma Y minus H of X minus H of Y and now you can see this is nothing, but basically correct and I can write this is the double summation is confusing i can simply write it as a single summation; summation over all possible values of i comma j probability of s i comma r j log to the base 2 probability of s i into probability of r j divided by probability of s i comma r j and now you can see this is of the form summation over i P i log to the base 2 X i.

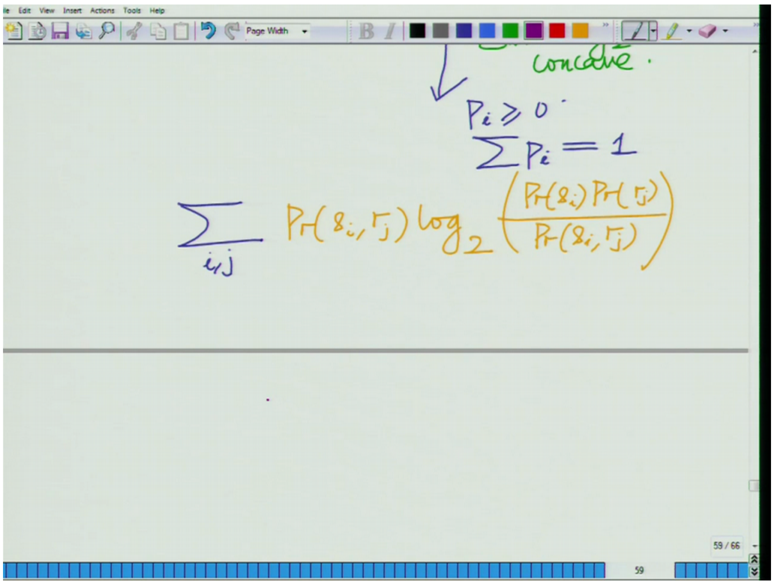
(Refer Slide Time: 20:00)



And we have seen by Jensen's inequality this is well a P i f of X i is less than or equal to f of summation over i because log X concave correct because the log is a concave this is basically Jensen's inequality we are again using or rather going to use Jensen's inequality since the function log 2 to the power since log to the power X is concave and therefore, what we were going to have. So, what we have seen is Jensen's inequality we have summation P i f of X i that is f if f is a concave function.

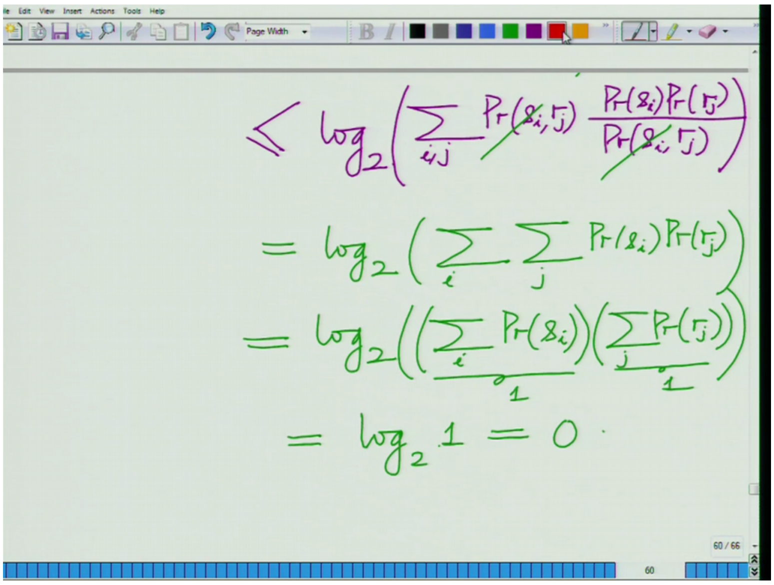
We have summation P i f of X i is less than or equal to less than or equal to f of the function of summation P i X i.

(Refer Slide Time: 21:23)



Where of course, the P i is are probability that is there is their positive non negative quantities and they sum to 1 that is of course, it goes without saying that the P i is are such that e P i is greater than equal to 0 and summation of P i equals 1 and therefore, what we have is if you can look at this you have i comma j i comma j probability s i comma r j log to the base 2 probability of s i into probability of r j divided by probability of s i comma r j by Jensen's inequality.

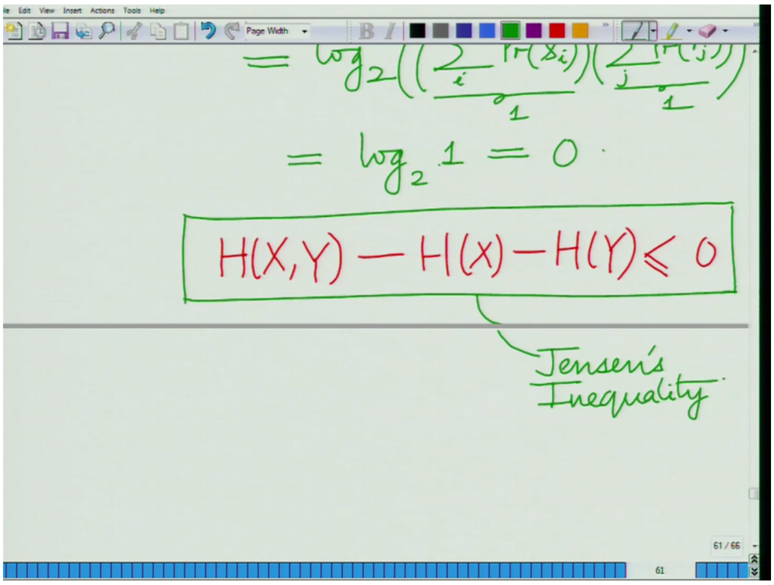
(Refer Slide Time: 22:18)



This is less than or equal to log to the base 2 summation i comma j probability of s i comma r j times probability of s i probability of r j divided by probability of s i into r j and now you can see this probability of s i comma r j this cancels this is equal to log to the base 2 summation i comma j.

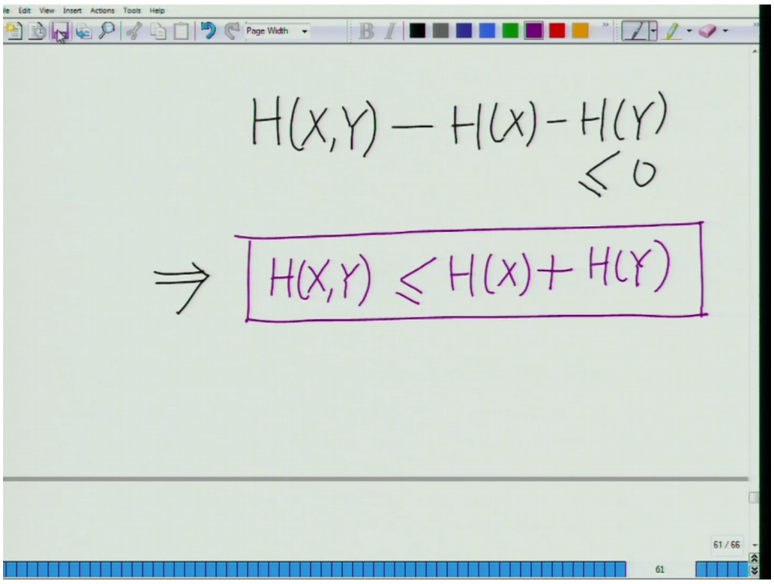
Now, you can write it as summation i and summation j spitted into 2 summations probability of s i and product of probability of s i probability of r j which is nothing, but if you look at this; this is log to the base 2 summation over i probability of s i into summation over j probability of r j now you can see the summation of probabilities by total probability rule this is 1, this is 1. So, this is log to the base 2 of 1 and which is 0. So, you can see this is equal to log to the base 2 of 1 and this is 0 and therefore, the net result the net result that you have remember this is all simplification of H of X Y minus h X minus h y. So, the net result that you have is H of X Y minus h X minus h Y is less than or equal to 0 where we have employed Jensen's inequality to derive this result, alright.

(Refer Slide Time: 24:15)



So, we have H of X comma Y minus h X minus h Y less than or equal to 0 and this is simply in application of sum simplification using total probability rule and also basically by using Jensen's inequality which is very important as we will see that is an information theory this inequality is used extensively to derive various interesting property. Such as we also seen this in the context or deriving the maximum entropy that is the source that is the probability density function of the probability mass function for the symbols of the source which maximizes the entropy of the source and we are seeing that the maximum entropy occurs when all the symbols of the source are equiprobable.

(Refer Slide Time: 25:15)



And now this is helpers derive this result that H of X comma Y minus H of X minus H of Y is less than or equal to 0 which basically implies that well H of X comma Y less than or equal to H of X plus H of Y.

So, that is the interesting part about this. So, joint information joint entropy is less than or equal to the sum of the entropies of the individual sources alright. So, what we have seen in this module is we have further seen the refine this motion of entropy all right we have seen the properties of this joint entropy earlier. So, we are earlier define joint entropy alright we are calculated joint entropy from a simple example we have also seen that the joint entropy is greater than or equal to the entropies of either source right entropies of either or either of the individual sources further the joint entropy is less than or equal to the sum of the entropies of the individual sources and this fact we are formally verified by proving it and proving it with the aid of Jensen's inequality alright. So, let us, we will stop this module here and look at other aspects in the subsequent modules.

Thank you very much.