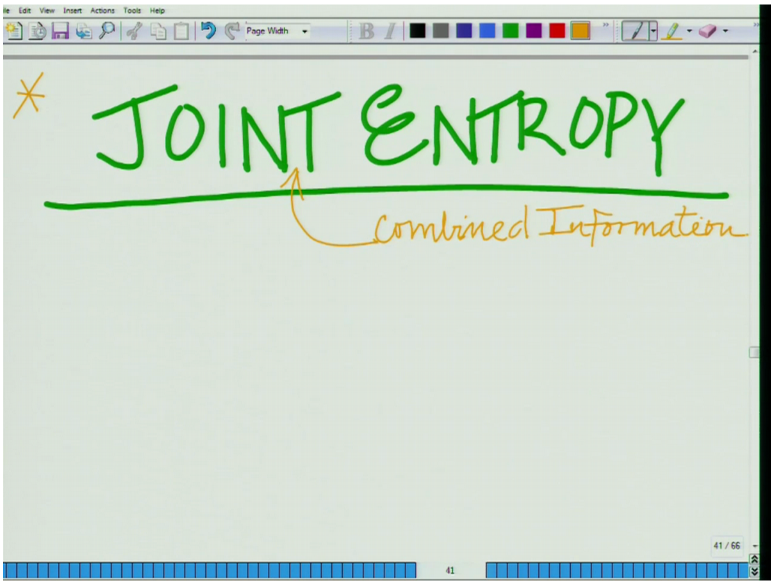
**Principles of Communication Systems - Part II**
**Prof. Aditya K. Jagannatham**
**Department of Electrical Engineering**
**Indian Institute of Technology, Kanpur**

**Lecture - 30**
**Joint Entropy - Definition of Joint Entropy of Two Sources, Simple Example for Joint Entropy Computation**

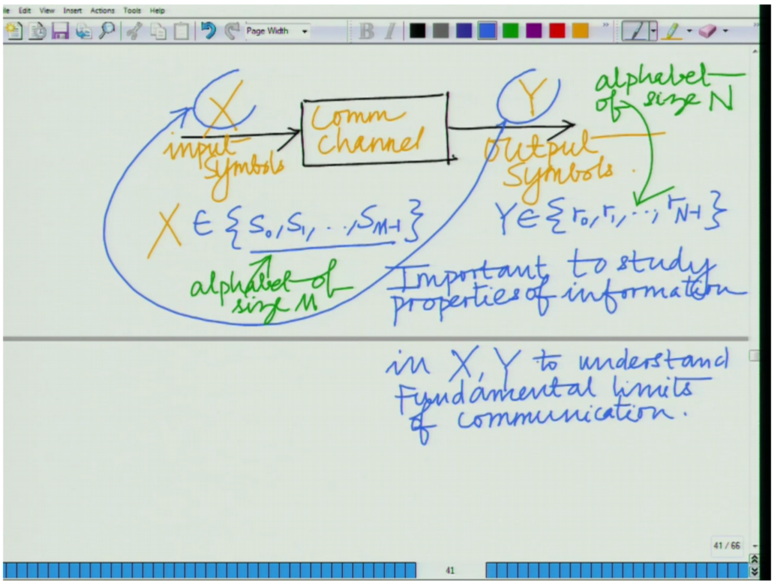
Hello, welcome to another module in this massive open online course. So, in this module let us start looking at another new topic that is entropy. So, far we have looked at different properties of entropy when is entropy maximized let us start looking at another concept that is joint entropy.

(Refer Slide Time: 00:30)




So, we will now start looking at a new concept which is the joint entropy you can think of this has the joint information into the combined information. So, basically you can think of this as the; we will talk more about this later you can think of this as the combined information when you have multiple source employees.

(Refer Slide Time: 01:09)



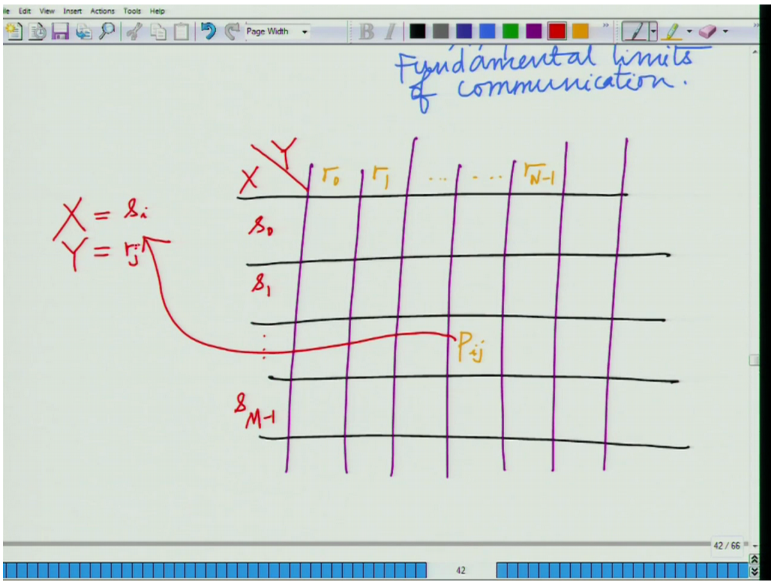

For instance, let us consider what is this relevance to the specific problem that we are interested and which is in the context of communication let us consider a simple communication system. So, we have a channel let us consider a simple example in which we have a channel we have a set of transmitted symbol input symbol. So, this X are your input symbols Y are the output symbols X can belong to the alphabet s naught s 1 s m minus 1 Y can belong to r naught r 1 r m minus 1 . So, now, we want to see what is the. So, now, X is the input Y is the output.

Now, we want to say what are the properties of the joint information or what are the properties of information in X and Y. So, it is now important to study properties of information in X comma Y call it to understand the fundamental limits of communication and the reason being we have X which is being which is X which is the input and Y right which is the output of this communication channel. So, you want to know how much information is communicated across this channel what information does why convey about X how much information does the input X convey about Y what is the total information at X and Y.

So, these are questions which are important to answer in order to characterize fundamental right properties right fundamental aspects pertaining to communication; communication of information on this communication channel and so that is what we are interested in looking at and the framework for which we will start developing in this module first by understanding what is the notion of an joint entropy between these 2 the input X and the output Y.

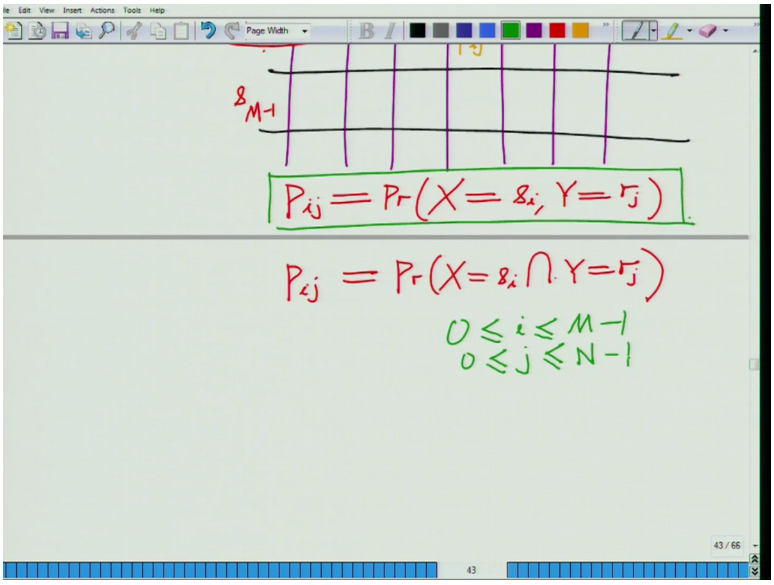
So, let us now look looked at so, of course, we have seen that this is the alphabet of size m alphabet of size m and this is observed that the output and input alphabet size need not be same. So, I am just going to change this to n because the output and input alphabet output alphabet can have different size in comparison to the input alphabet now.

(Refer Slide Time: 04:59)



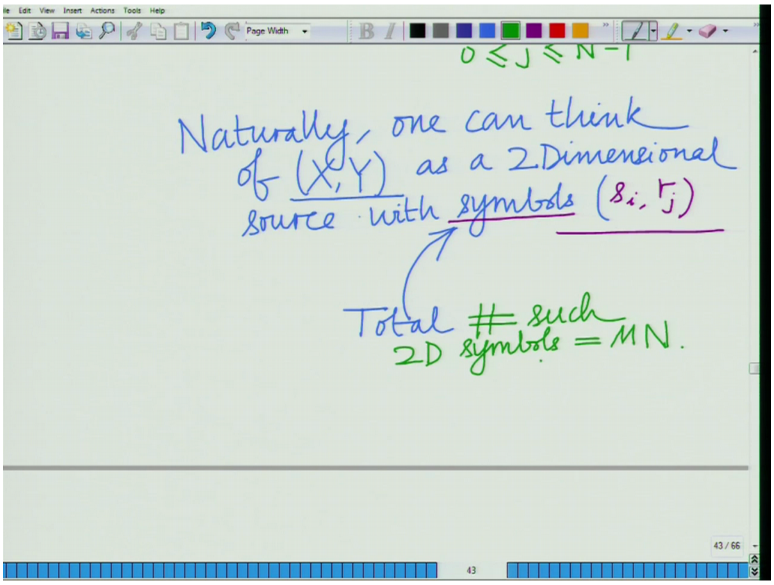

So, let us now understand this. So, let us now make a table for this to describe it better I am going to make a table for this just to describe this better and what I am going to have here are basically your. So, this I am going to divide into X comma y. So, various possible values that as s m minus 1 and various possible alphabet of Y that is r 0 r 1 r n minus 1 and we will look if you look at the i j th entry and call that as P i j this corresponds to P i j basically corresponds to X equal to s i and Y equals the output symbol Y equals r j.

(Refer Slide Time: 06:29)



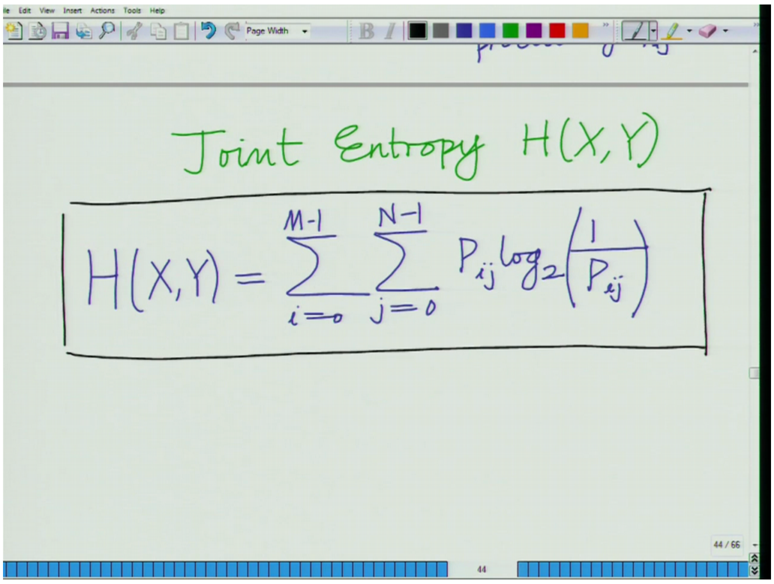

So we are denoting this P i j. So, this 2 dimensional table we are denoting P i j by P i j the probability X equals s i Y equals r j you can also think of this as the probability X equal to s i intersection Y equal to r j that is our P that is our probability we are denoting this by the probability P i j of course, in naturally we must have 0 less than equal to i less than equal to m minus 1. So, 0 less than equal to j less than equal to n minus 1 m is the size of the alphabet X of X from which the input symbols are drawn and n is the size of the alphabet from which the output symbols appear or occurrent the output of this communication channel and now we want to define this notion of joint entropy for this kind of an input output scenario X and Y.

(Refer Slide Time: 07:50)



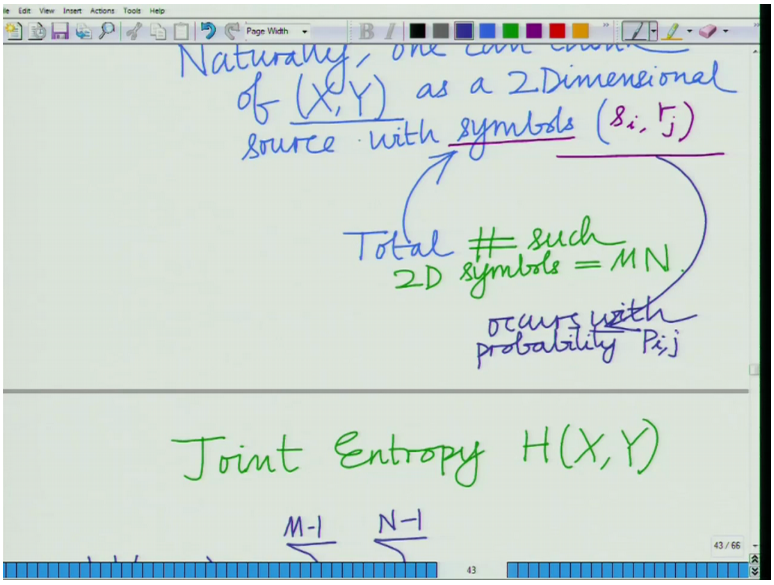

Now, if you look at the total number of. So, so naturally now one way to understand is naturally one can think one can think of X comma Y as a 2 d that is. So, we have looked at. So, 2 d is also implies that you have 2 components X and Y that is you have been ordered pair correct. So, 2 dimensional source with symbols with symbols s i comma r j. So, one can think of this as 2 dimensional symbols and therefore, total number of such 2 d symbols. So, total number of this is total possible combinations of X comma Y m can take X can take m possible m possible symbols n can be choose Y can take from n possible symbol is the total number of 2 d symbols 2 dimensional symbols is m into n.

(Refer Slide Time: 09:35)



And now the entropy the joint entropy is nothing, but the entropy of these 2 dimensional source X comma Y the joint entropies the entropy of this 2 dimensional source X comma Y and that can be defined as follows which is denoted by h of X comma y. So, I will have h of X comma Y we may just write at a little clearly that is basically your h of X comma Y equals well summation i equal to 0 m minus 1 summation j equal to 0 to n minus 1.

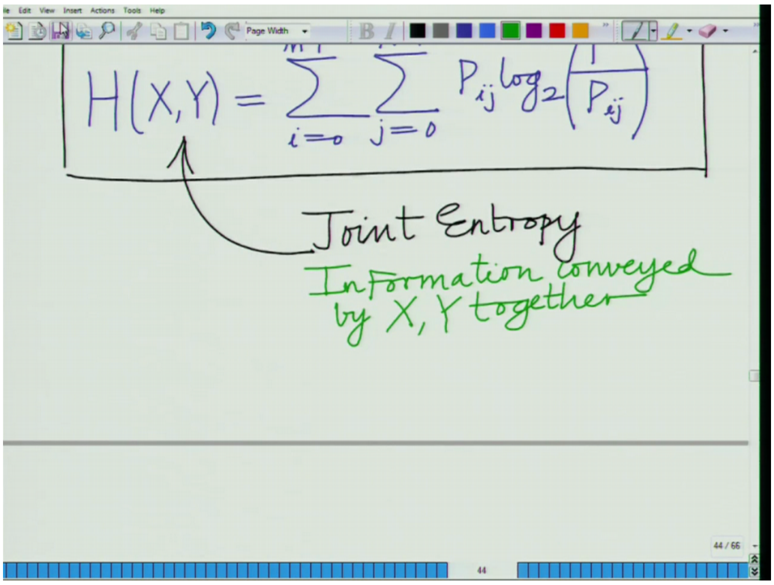
(Refer Slide Time: 10:24)




Remember each such 2 d symbol occurs with probability P i j that is what we have said occurs this s i comma r j into occurs with probability P i j or P i j this let me simply denote it by P i j. So, this is P i j log to the base 2 1 over 1 over P i j. So, this is the joint entropy this is termed as the that is you look at X comma Y z 2 dimensional source look at the probabilities of intersection that is probability that X is equal to s i Y is equal to r j there are total of m n 2 dimensional symbols.

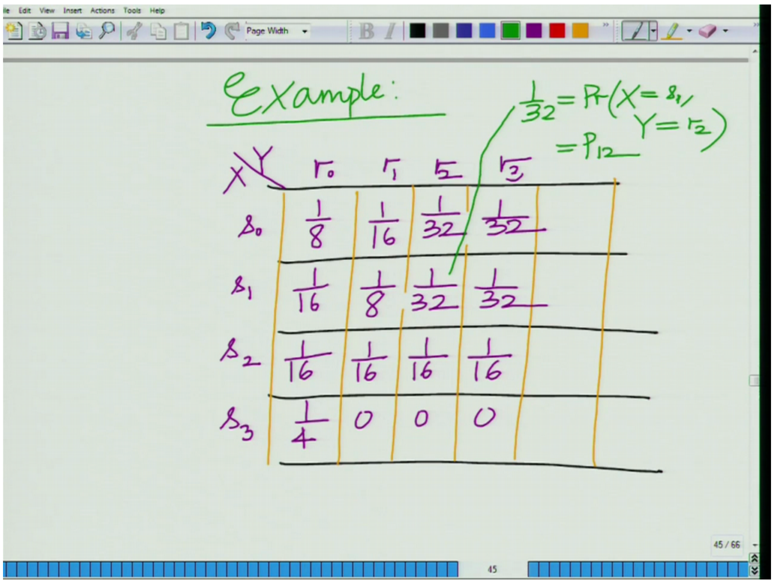
Now, treat this is a 2 dimensional source and compute the entropy of the 2 dimensional source that is submission over all the symbols of this 2 dimensional source probability times log to the base 2 log 1 over probability to the base 2 all right, we have now extending. So, a matter of fact we are now extending the definition the same definition that we had for simple source before to this new that we have calling as that we terming as a 2 dimensional source.

(Refer Slide Time: 11:57)




And this metric which characterizes the joint information the total you can think of this is termed as the joint entropy and you can think of this as the information per symbol in X and Y together that is if you look at X and Y together that is information conveyed by X and Y together that is the meaning of the word joint X comma Y together or basically jointly.

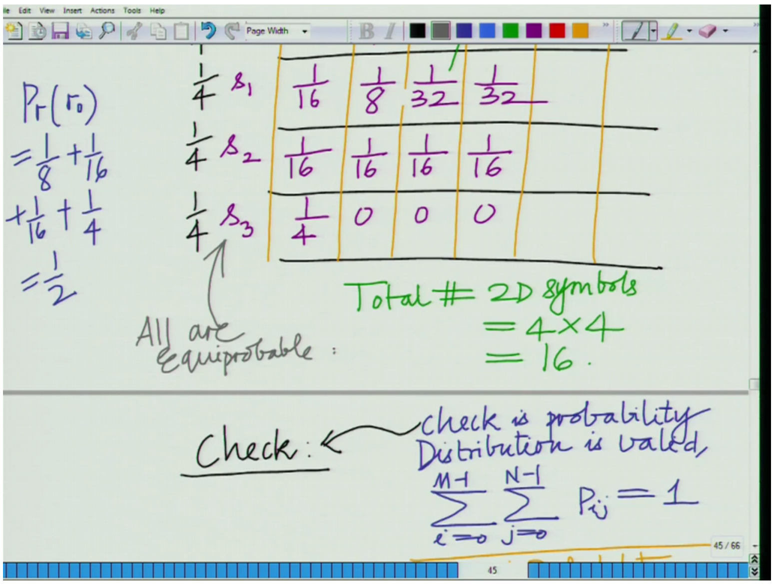
(Refer Slide Time: 12:46)




So, that is basic meaning of this term entropy. So, let us look at an example to understand this better consider the example below let us look at a simple example to understand this concept of joint entropy. So, I have here again X Y r 0 r 1 r 2 r 3 and Y s 0 s 1 s 2 s 3 and the various probabilities are 1 over 8 that is the probability that of X equal to s 0 Y equal to r 0 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 and finally, you have this for instance if you take a look at this 1 by 32 this is equal to the probability that well X is equal to s 1 comma Y is equal to r 2 that is what we have said this is basically your P 1 comma because P i j is the probability that X equal to s i Y is equal to r j. So, this is your P i j.

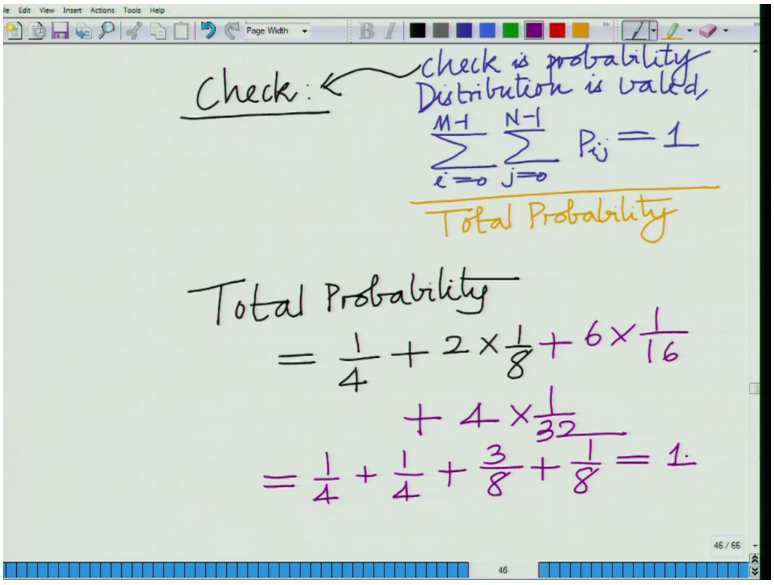
And now so, the total number of 2 d symbols of course, now we can see if because each has each has 4 symbols as 0 s 1 s 2 s 3 and r 0 r 1 r 2 r 3 total number of 2 d symbols this is 4 into 4 equal to 60.

(Refer Slide Time: 15:05)




And I am taking all the possible combinations of the symbols for X and symbols for Y we have total possible 2 d symbols s i comma r j and now just to make a check of course, we know that the total probability must be equal to 1.

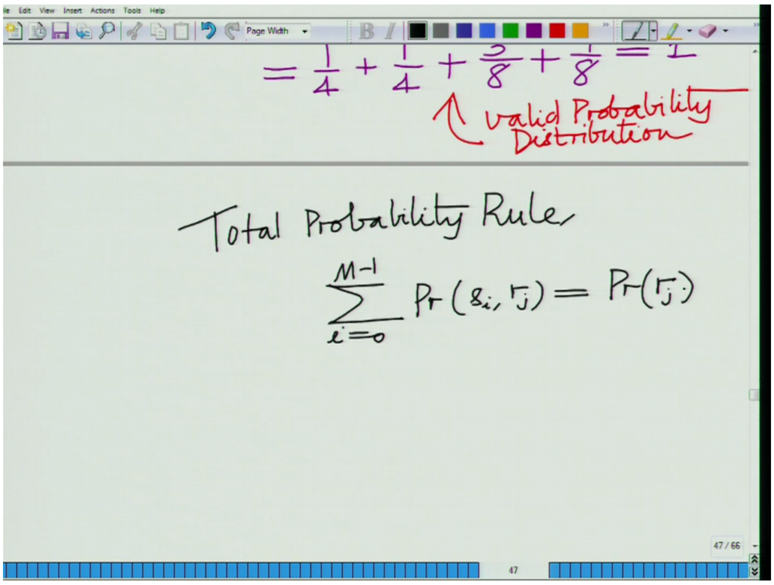
(Refer Slide Time: 15:36)




So, we must have and you can check this you just see if it is a valid distribution to check if the probabilities or probability or the 2 dimensional probability mass function we must have summation i equal to 0 to m minus 1 summation j equal to 0 to n minus 1 equals this must be equal to 1 and we know this from our total probability.

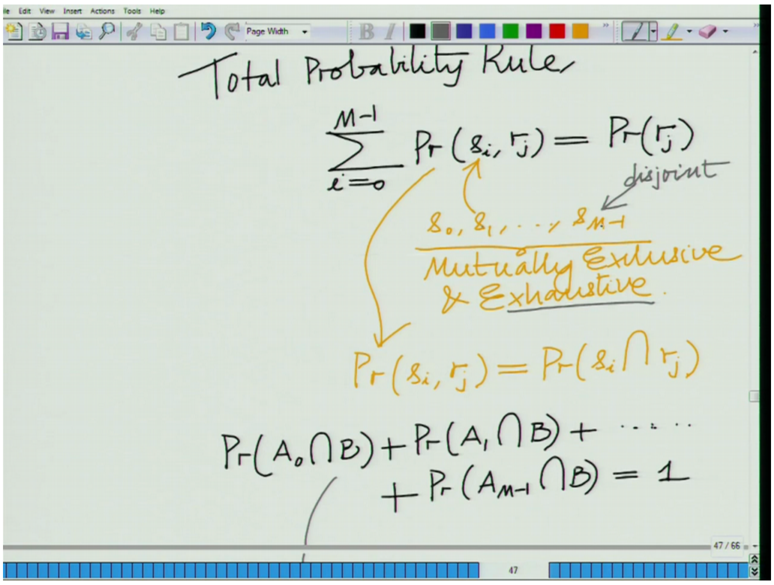
We have some of the probabilities of all the events minus some of the probability. So, this is right this is mutually exclusive and exhaustive. So, the total probability must be add up to 1. So, let us just quickly check it check if the total probability is adding up to 1 all right all though that must be the case, but just helps us traditionally check and confirm it. So, that would be, so we have total probability the total probability where there is one symbol with probability 1 by 4. So, 1 by 4 plus, so, 2 symbols with probability 1 by 8; 1 by 8 plus 6 symbols with probability 1 by 66 into 1 by 16 plus 4 symbols with probability 1 by 32 4 into 1 by 32 that is 1 by 4 plus 1 by 4 plus 3 by 8 plus 1 by 8 and you can check this is equal to 1.

(Refer Slide Time: 17:59)




So, this is basically if a valid probability which implies you have a valid probability you have a valid probability distribution now from the total pro now other in another interesting thing is also from the total probability rule we have summation i equal to 0 to minus 1 probability s i comma r j must be equal to probability of r j that is corresponding to an r j symbol r j if you some over all the s i is probability is corresponding to each s i correct.

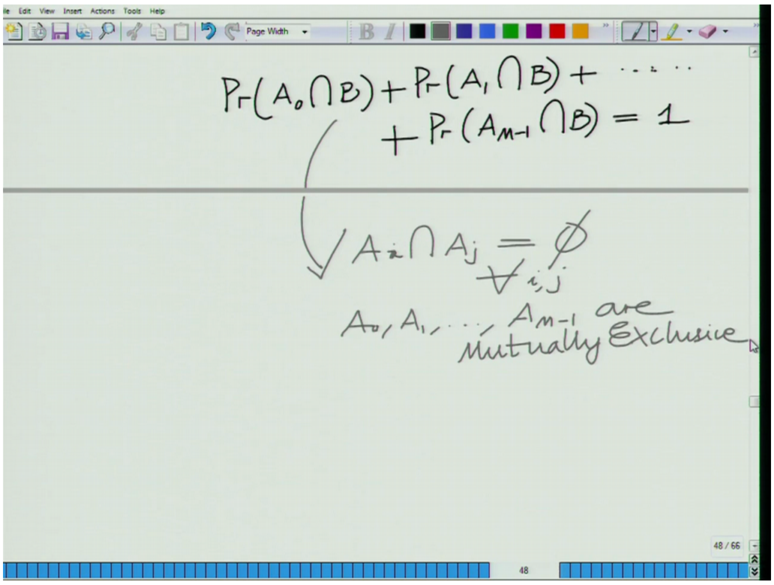
(Refer Slide Time: 19:10)



You will get the net probability corresponding to r j this is from the total probability rule because remember these s i s s 0 s 1 s m minus 1 these are mutually exhaustive these are mutually exclusive and exhaustive that is the intersection of each of these symbols is phi and there union spans the entire set, alright.

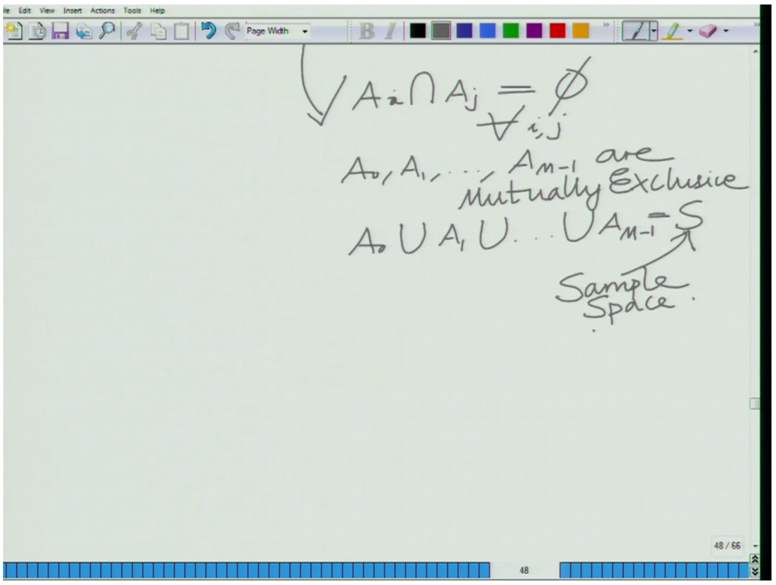
And therefore, if you look at probability s i comma r j this is nothing, but the probability s i intersection with r j and from total probability.

(Refer Slide Time: 20:05)




We have basically probability A naught intersection B plus probability A 1 intersection B plus. So, on up to probability a m minus 1 intersection B is 1 if of course, this you should know from the theory and probability random variables, but anyway just to refresh your memory this if A naught or if A i intersection A j equal to phi or all i comma j that is implies A naught A 1 comma A m minus 1 are mutually exclusive.

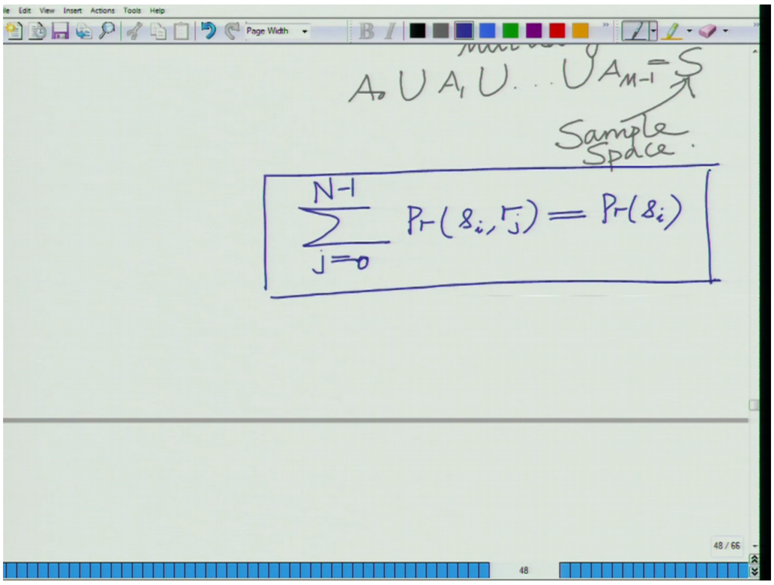
(Refer Slide Time: 21:08)




And they are exhaustive that is A naught A naught union A 1 union A m minus 1 equals s that is the that is the sample space. So, these events are basically a partition of the sample space.

So, there each of these events is disjoint that is their intersection is phi correct and their union spans the entire sample space for instance what we have here in terms of the symbols these symbols s 0 s 1 s m minus 1 these are disjoint in the sense that occurrence of one precludes the occurrence of the other and they are exhaustive in the sense that the symbols must belong to 1 of these s naught s 1 s m minus 1. So, therefore, we must have probability that is therefore, you have some over the probability for each r j if we some over the probability is probability s i comma r j corresponding to each s i comma corresponding to all s i P you get probability r j.

(Refer Slide Time: 22:25)

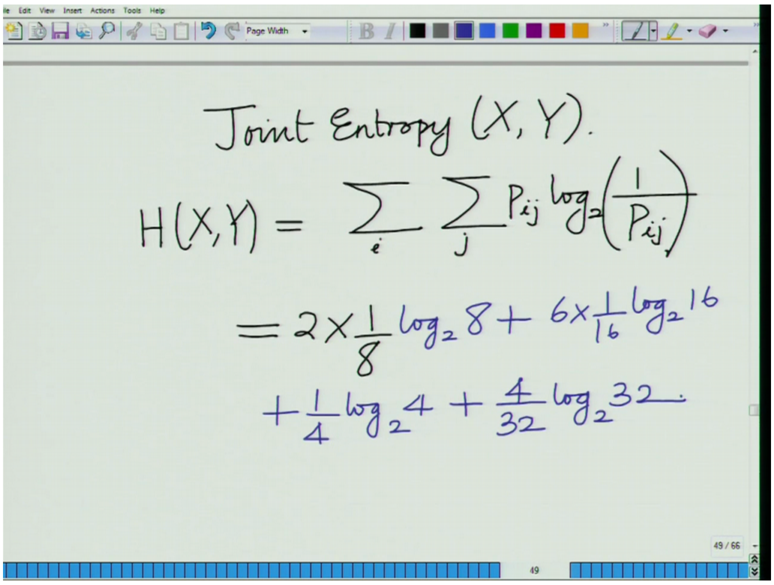



Naturally also if you some over the probabilities of all r j corresponding to corresponding to an s i you get the probability of s i.

This is again according to your total probability therefore, what this means is basically if you some over each column that is r 0 for instance r 0 r 1 r 2 these represent each column. So, if you some over each column you get the probability that is some over the probabilities for each r i corresponding to all the s s you will get the probability correspond total probability corresponding that r i. Similarly if some across the rows that is some over all r j corresponding to each s i then you will get the probability of that s i. So, therefore, we have probability of r 0 equal. So, let me illustrate this thing r 0 equals 1 by 8 plus 1 by 16 plus 1 by 16 plus 1 by 4. So, that will be 1 by plus 1 by 16 plus 1 by 16 plus 1 by 4. So, that will be 1 by 8 1 by out of 4. So, that will be half.

So, probability of r naught equals half probability of r 1 I am just writing it over the column probability r 1 equals 1 by 4 probability of r 2 equals 1 by 8 probability of r 3 is also 1 by 8 and probability of s naught you can check probability of each s naught is 1 by 4 that is all these s s r equiprobable. So, all are equiprobable and therefore, now what you will have is let us proceed to compute the joint entropy the joint entropy.

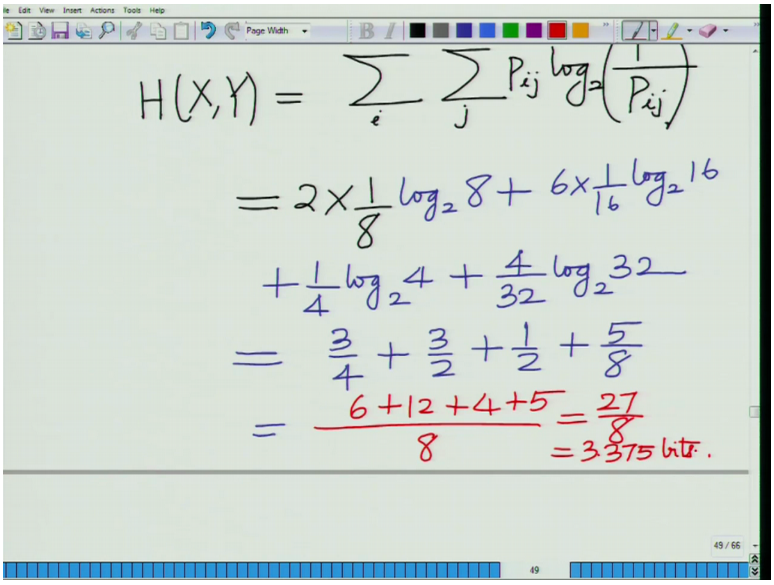
(Refer Slide Time: 24:45)




Now, as we have seen joint entropy X comma Y remember this is h of X comma Y summation over i summation over j P i j log to the base 2 1 over P i j.

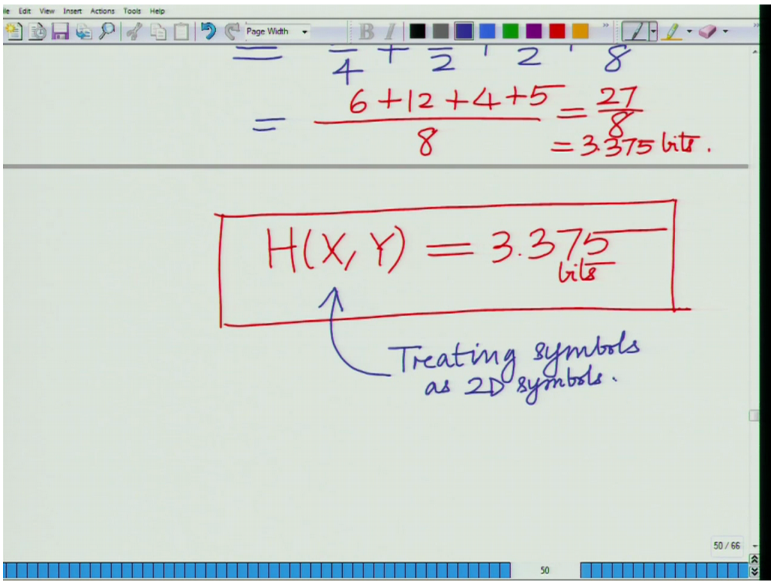
So, this is equal to well there are 2 symbols with probability 1 over 8. So, twice 1 over 8 log to the base 2 8 plus there are 6 symbols with probability 1 over 16 6 into 1 over 16 log to the base 2 1 over 1 over 16 that is 6 1 16 plus that is log 16 to the base 2 1 over 4 log 4 to the base 2 only 1 symbol with probability 1 over 4 and in fact, there are 4 symbols with probability over 32 4 over 32 log 2 to the base 32 and you can see this will be basically you can see this will give you.

(Refer Slide Time: 26:10)




Well this will give you 3 over 4 plus. So, this will be 16 4 divide 1 over 4. So, this will be 3 over 2 plus half plus 5 over 8 that is if you look at this; this will be over 8 you have 6 plus 12 plus 4 plus 5 that is basically you have express for 10 22 plus 5 that is 27 over 8 which is equal to 3.375 bit.

(Refer Slide Time: 27:11)




So, we have the joint information the joint entropy h of X comma Y for this example equals which we have computed treating the symbols as 2 d symbols.

So, basically what we have done now is we have computed the joint entropy of X and Y the (Refer Time: 27:51) is denoted by h of X comma Y. So, in this module basically we will introduce we motivated this problem of joint entropy as considering a simple example of a communication system in which your transmitting symbols X receiving symbols Y. So, we would like to understand the interplay better understand the interplay or the information various aspects of the information between of corresponding to these 2 point it is just a import X and the output Y. We have defined this notion of joint entropy and we have seen a simple example and computed the joint entropy. We have further explored the properties of joint entropy and their relevance in the context of communication in the subsequent modules.

Thank you very much.