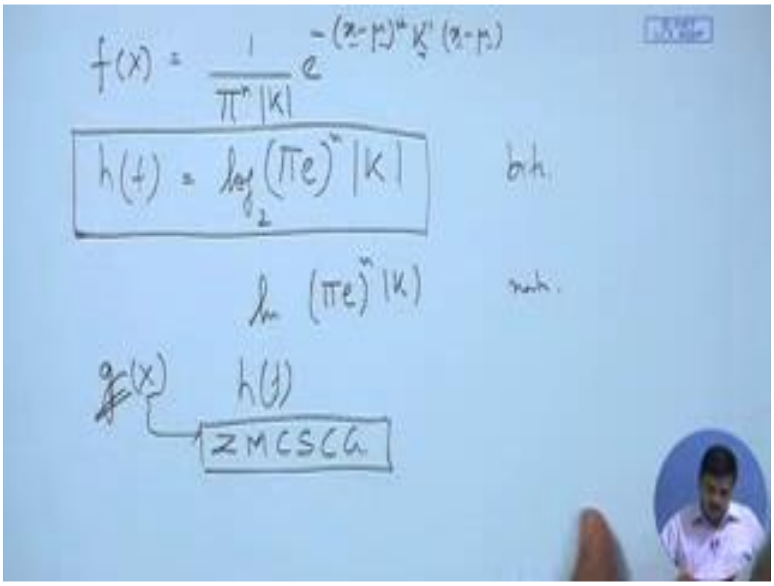
Fundamentals of MIMO Wireless Communication
Prof. Suvra Sekhar Das
Department of Electronics and Communication Engineering
Indian Institute of Technology, Kharagpur

Lecture - 34
Fundamentals of Information Theory – 4

We continue in this lecture with the differential entropy in the previous lecture. We have seen that Gaussian distribution or multi varied Gaussian distribution maximizes the entropy of another of any random variable with the same co variance matrix. We have given the results for real random vectors see if you have complex random vector then the result is very, very important and important for our case. So, we extend the result what we have achieved before to that of the complex random vectors. At this point I would like to remind you or tell you that these results are very fundamental and you can find the reference in a paper by Nazeer and of 1993 where they have shown that the 0 mean circular symmetric complex Gaussian random vector maximizes the entropy.

So, we will be just showing you the result which is almost straight forward given whatever we have described all though the derivation of this result is not. So, straight forward, but for our case very interest in the results mainly with whatever buildup we have achieved in the previous lecture in this 1 we are going to extend that and show the result for the complex Gaussian case.

(Refer Slide Time: 01:30)

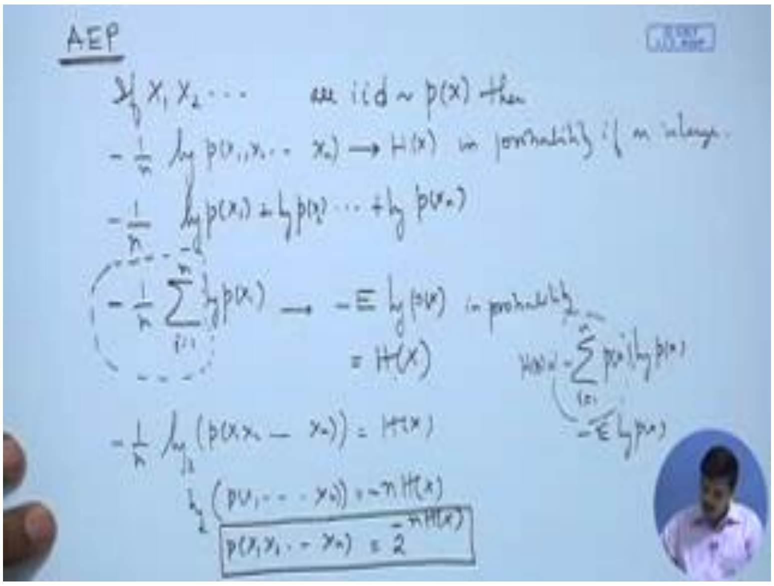


So, in the case of a complex Gaussian the f of x is given by 1 upon pi raise to the power of n times determinant of k e to the power of minus x minus mu hermitian times k inverse times x minus mu and remember all of these are vectors this is matrix. So, this is the expression of f x. So, if we have to calculate h of f what will be getting is and what is important h of f will be getting as log 2 of pi e raise to the power of n times determinant of k in bits or if we have l n pi e to the power of n of a k in nabs right. So, this is the expression of h of f, when x is a complex Gaussian distributed with n components and again if we try to see if g is random if g is a distribution of a set up random variables like g of x like we have done over here like we have done in the previous lecture.

In this case again we could show that h of g is upper bounded by h of f which is given by this expression. So, basically this is the maximum entropy provided x is 0 mean circular symmetric complex Gaussian and we have explained this definition and also there is definition that g is proper. So, if you do this these terms have to be proper and this would be a very, very important expression in our results on capacity. So, with this we move on further and we would like to give you explanation or an intuitive feeling of a how do we get the expression of capacity.

So, we are not going to derive the exact expression of capacity because that is not our n goal because we will be using the expression it is important that we have an intuitive understanding of what we are talking about. So, for this we will be using again all the things that we have established and what we need at this point is partition of energy or partition property not as partition of energy as we learned in typical physics.

(Refer Slide Time: 03:52)



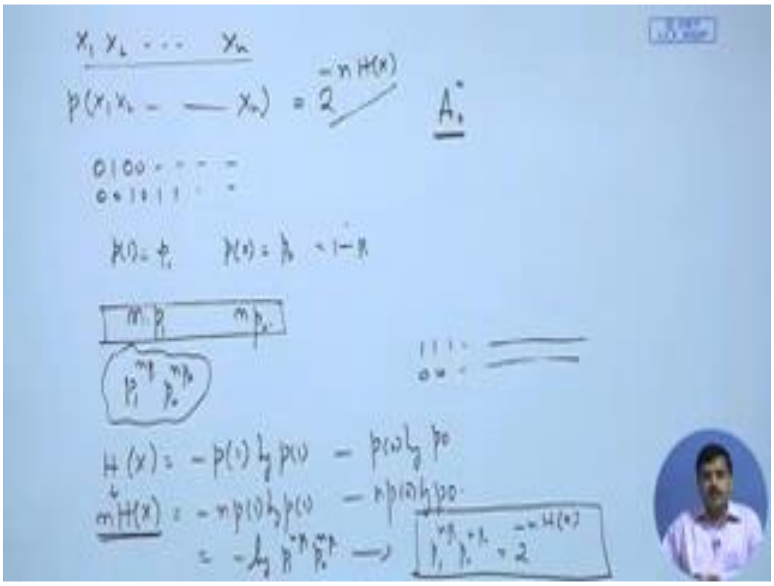
So, basically we will be talking about a e p and a e p states that if x 1 x 2 and so on are independent identically distributed following distribution of p of x then minus 1 upon n log p x 1 x 2 up to x n tends to h x in probability. If n is large right. So, this statement is very, very in indicated on this gives us a very important results on it is quite nice to understand what it tells us. So, if we take a look at this expression on the left hand side that we have here it is 1 minus 1 upon n log of p x 1 x 2 up to x n. So, this we could also write as minus 1 upon n log p x 1 plus log p x 2 and. So, on plus log p x n which is basically minus n sum of i equals 1 to n log p x i.

If we look at this it is almost sample mean of this particular things this we could write it as almost expectation of log p x right this is what we could write because this if n is very, very large then by weak log large numbers, when n is very, very large we have very large value of n by weak log large numbers we could write this as the expectation of these variables and this would be also. So, there is minus expectation of sorry this is a minus sign minus expectation.

So, minus expectation of log p of x if we would remember then we could write this as h of x because h of x was defined as a sum over i equals to 1 to n p x log p x with a minus sign. So, this was basically the leading to the expectation operation, but when n is very, very large well weak log of large numbers we said we could write this as expectation and from this this we have written it as minus expectation of log p of x. So, this was defined as h of x in in integral form instead of the summation there was integral sign.

Basically, now we are using this definition and write saying that this could be written in probability in probability sense as h of x now what are the things that lead that that this leads to. So, if we move down further again starting from this point what we have is minus 1 upon n log of 2 p x 1 x 2 x n is equal to h of x. So, clearly we have minus n we have n h x we have a minus. So, this will log 2 p x 1 to x n is basically this right? So, we could write p of x 1 x 2 up to x n is equal to 2 to the power of minus h x minus n h x. So, what is this p x 1 x 2 x n being equal to 2 to the power of minus n h x? So, if we look at this expression in details we get some intuitive understanding of what is the meaning of this.

(Refer Slide Time: 07:55)



So, x 1 x 2 up to x n is basically a sequence of observations that we get sequence of random variables right and we have said earlier that it is an i i d you said here it is an i i d with individual distribution p of x. So, x 1 to x n is basically a sequence and p of x 1 x 2 x n is a probability of this particular sequence to occur and without choosing a particular sequence we said that this appears to be 2 to the power of minus n h x. So, this in turn means that if i take a random sequence x 1 to x n n is very large the probability of that sequence is 2 to the power of n h x and as I take as many sequences all of those sequences have these probability.

For instance, I could take 0 1 0 0 up to a certain number i could take 0 0 1 0 and. So, on and. So, forth up to certain number x n the probability of an arbitrary sequence is 2 to the power of n h x this is somewhat difficult to get initially, but if we look at what we are doing here is i am drawing an outcome

I am drawing a particular sequence x 1 followed by x 2 x 3 and x 4 and I say the probability of x 1 is same that of probability of x 2 and so on and so forth, they are all independent and they are all identically distributed. So, if we take let say 0s and ones coming out and then we say the probability of 1 is equal to let say some p 1 and probability of a 0 is let say p 0 or which is equal to 1 minus p 1 and we have sub sequence what we are going to get is the number of ones in such sequence is basically n times p 1 when n is very, very large number of 0s in such sequence is basically again n in to p 0.

So, if I ask what is the probability of this particular sequence it is basically the probability of 1 raise to the power of p 1 times probability of 0 raise to the power of n p 0 because all of them are i i d right. So, so this is basically the probability of that particular sequence and all we are saying is that that that value is equal to 2 to the power of minus n h x and there could be many such sequences that that comes up i mean it is not all of the sequences, but what it is essentially saying that if i pick an output if I take n x 1 from that from the source and what is the probability of that being 1 it is basically p 1 if I take another or most that being 1 is basically p 1. So, if I keep doing this experiment for a very long time and generate a very long sequence on an average I am going to get n times p 1 number of ones and n times p 0 number of 0s.

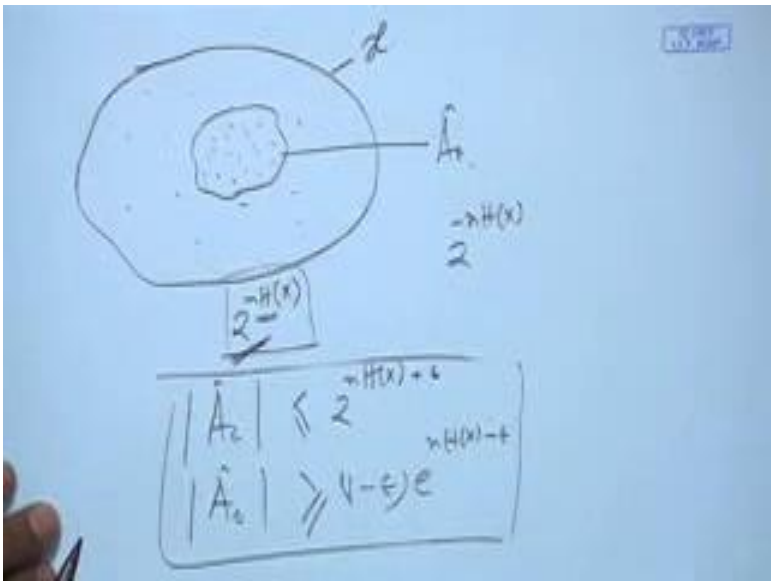
So, just try to imagine that you are doing this experiment over and over again for very large sequence x n it would turn out that each 1 of them would have the probability 2 to the power of minus n h x this is in probability. So, if is there could be other sequences there could be sequences which are like all ones there could be sequences like all 0s, but if n is very large we could generally say that just a typical sequence if I take out a sequence from that a typical sequence would have n times p 1 number of ones and n times p 0 number of 0s in that.

So, the over all probability of the sequence is p 1 to the power of n p 1 and p 0 to the power of n p 0 if you would take the now if I say that what is h x, if I say what is h x, h x is basically minus p 1 that is probability of getting 1 times log of probability of getting a 1 minus p 0 times log of p probability of getting a 0 right t 0 and if there are n such symbols i would say we have n h x is equal to minus n p 1 log of p 1 minus n p 0 log of p 0 and basically if you take a look at this it is basically log of minus log p 1 to the power of n p 1 times p 0 to the power of n p 0 and again this particular sequence p 1 n p 1 p 0 to the power of n p 0 is equal to 2 to the power of minus n h x.

So, started with h x, if I take n such sequences, n such symbols what I get is n h x n times h x and the right hand side is straight forward. So, what we get is the probability that sequence is indeed 2 to the power of minus n h x. So, one is to spend little bit of trying in trying to capture what it is talking about. So, we usually define a set known as the typical the set of typical sequences the set of typical sequences all we say is that it contains sequences which occurred with very high probability and that probability is 2 to the power of minus n h x and all the sequences in that set are equi-probable.

So that means, I am drawing a very large sequence a very, very large sequence we can clearly say that number of ones in them is n p 1 number of 0s is n p 0 and there could be many such sequences and each of the sequences occur with equal probability provided the occurrence ones in 0s are independent and they are identically distributed in any of the locations. So, those sequences which are likely to occur or high likely to occur are known as typical sequences and their probability of those sequences given by 2 to the power of minus n h x and they are all equi probable.

(Refer Slide Time: 14:17)

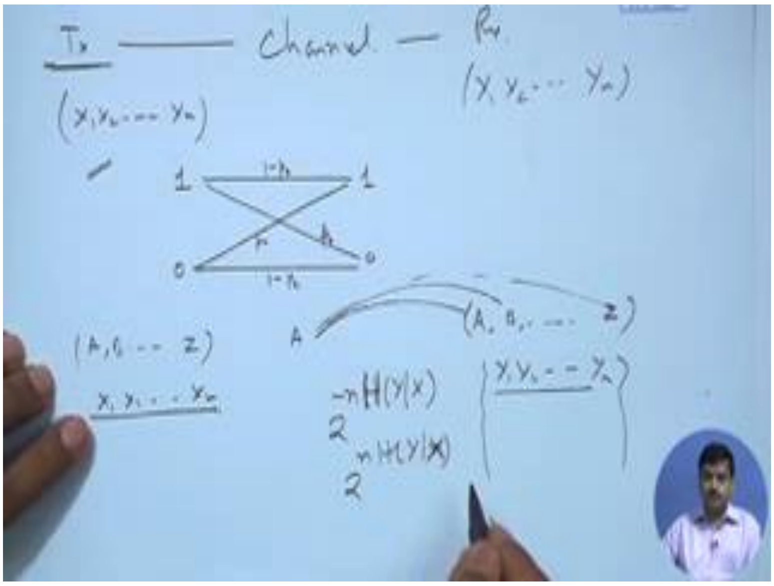


So, if we consider a big set of sequences, if we consider a big set of sequences we will find that there is a subset of sequences this is the full set this is the set of typical sequences usually indicated in in this manner which has all these all these sequences.

So, and if they are all equi probable all we can say is that the number of sequences in this set is basically 2 to the power of nearly equal to 2 to the power of n h x. So, this this is also an outcome of this particular thing. So, number of sequences in that is basically 2 to the power of n h x in in this typical sequence or in other words what is usually said is that the number of element in n is less than or equal to 2 to the power of n h x plus epsilon and number of sets in the typical sequence is greater than or equal to 1 minus epsilon 2 to the power of minus n h x minus epsilon for large n.

So, basically this bounds. So, what we are trying to say, we are not going to the proof of this although it is not difficult to prove it because again we are interest in the result the number of such typical sequences is 2 to the power of n h x. So, we have it is also can be shown that it can be easily quickly told in terms of 2 to the power of minus n h x is the probability of a sequence all the sequence are equal equi probable. So, the number of sequence in that case is 2 to the power of n h x you can also justify in that sense. So, once we have said this we move on to try to understand that how do we take advantage of this particular thing. So, again detail description of this can we found in any book on information theory, now suppose we have typical communication channel and there is a transmitter there is a channel there is a receiver.

(Refer Slide Time: 16:19)



So, at the transmitter what we take is sequences right. So, we take sequences all we are saying now is the transmitter generates a sequence and the sequence is the sequence of symbols and if it is, if the symbols are equi probable then we can form a typical sequence and each such sequence occurs with probability of 2 to the power of minus n h x where n is the number of symbols in that sequence and what we get over here is y's y 1 y 2 up to y n generated by x 1 to x n and the channel lies in between.

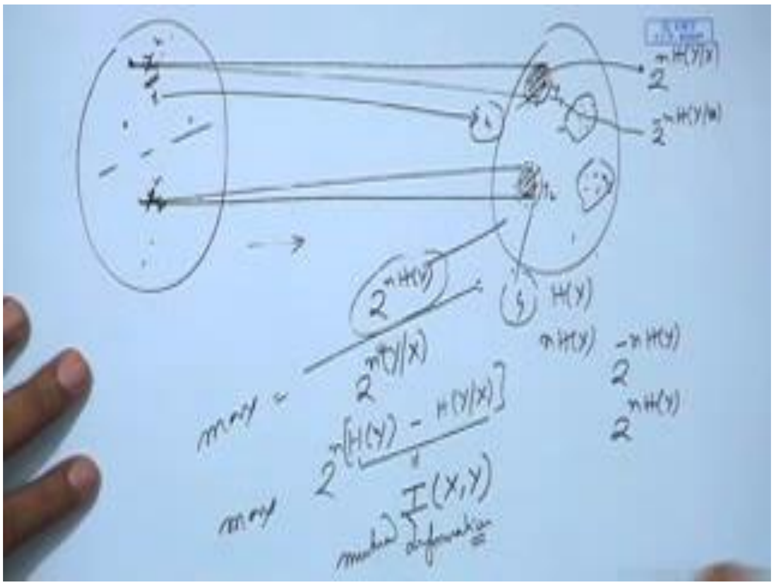
So, if we take a very simple case where there is a 1 and a 0 that is been generated by the source what the channel could do is 1 could remain a 1 one could become a 0 0 could remain a 0 or 0 could become a one. So, this is 1 of the simplest channel that we can think of where 1 becoming a 0 is identified the probability of error and 1 minus p e is proved it as correct in the symmetric channel case we are going to have this as p e and this as 1 minus p e as well.

So, basically every symbol that goes out has a probability of becoming some symbols. So, basically if I take the example of a p up to z, let us have this 26 characters when I sent a there is sudden probability that could remain a there is sudden probability could become b and so on with the certain probability of a becoming z. So, if I take any particular sequence from this I take x 1, x 2 up to x n i would find a corresponding sequence of it becoming certain y 1, y 2 up to y n where y 1, y 2, y n can take values from this.

Now, since we have talk about the sequence been generated at the source we could also say that at the receiver because of a particular sequence there is sequence that is get generated now clearly because of 1 in this example 1 could become 1 or 1 could become a 0 and in this example a when transmitted could become a could become a b or could become a z with certain probability this particular sequence could produce many sequences right and the typical sequence has a probability of 2 to the power of minus of n h y conditional x y given x

So, what it is trying to say is basically when a particular sequence is taken and if I look at all the sequences that get generated because of this typical sequence will have a probability this and the number of such sequences would be 2 to the power of minus n h 2 to the power of n h y given x.

(Refer Slide Time: 19:26)



So, we can draw broader picture and say that, suppose I have a source which selects from a certain points or certain sequence now this point indicates sequences x 1, x 2 and so on, when it goes to the channel it produces y 1 consider this is the sequence y 2 y 3 and so on. So, this x 1 sequence this is x 1 n indicating a particular sequence produces a certain set of outputs. So, this generates certain set of outputs this generates a certain set of outputs and the number of such sequences in these in this particular 1 is 2 to the power of n h y even x is that is what is being meant and each outcome with a probability of 2 to the power of minus n h y given x.

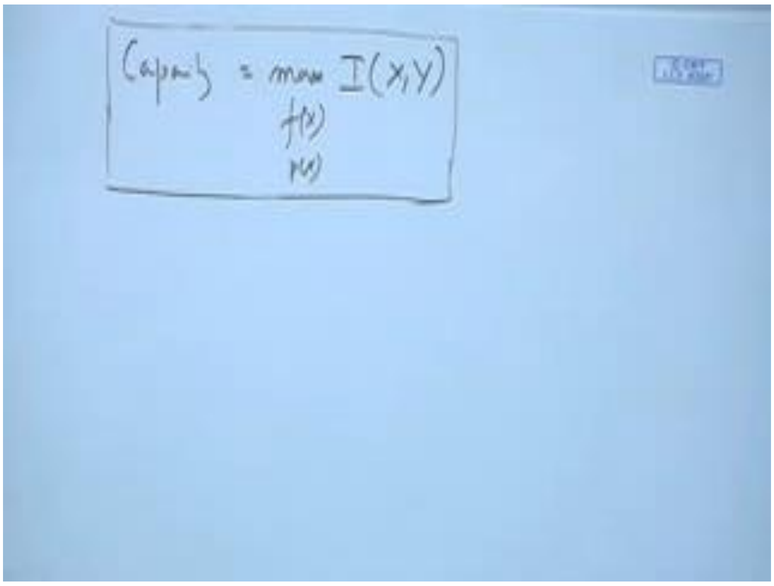
So, what is the number of typical sequences in y, if h of y indicates the entropy of this output? So, basically n h y is the entropy of the output when n symbols are taken together. So, the typical sequence would have a probability of 2 to the power minus n h y and the number of such typical sequences would be 2 to the power of n h y. So, there are 2 to the power of n h y number of typical sequences 1 particular sequence produces 2 to the power of n h y even x.

So, what we are saying is out of all these sequences 1 sequence produces. So, many sequences this particular sequence should produce so many sequences. So, if we have to achieve low error probability or if we have to say that when x 1 sequence is transmitted x 1 n or when x 2 n sequence is transmitted all of these sequences are received right with certain probability all of them are is equal with the probability and these sequences are received with equal probability when x 2 sequences is sent the job of the receiver is to identify what sequences is been transmitted. So, if we have to ensure that when x 1 is transmitted the receiver does not decode it as x 2 it decodes only as x 1 then we have to ensure sufficient separation of the sequences.

In other words, if this contain let say 4 sequences, for example, for this contains another four sequences we have to ensure that when a particular sequence generates any of these four sequences it would the receiver would say the particular sequence 1 was generated where as when these four sequences are generated the receiver would say that x 2 was generated. So, have to ensure there is separation between these sequences. So, in total if there are 2 to the power of n h y typical sequences and if 1 particular sequence at the input produces 2 to the power of n h y given x a number of such sequences, we can have 2 to the power of n h y upon 2 to the power of n h y given x number of different sequences.

So, we could have so many such different sequences in this there would be so many such sequences. So, this number, this count 1, 2, 3, 4 this number would be given by this right. So, this is equal to 2 to the power of n h y minus h of y given x. So, what would like to do is to increase the transmission rate that means send as many sequences as possible. So, we want to maximize the number of separable groups at the receiver. So, that maximum number of symbols at the transmitter could be send without confusing without any ambiguity. So, basically we are trying to maximize this ratio. So, at the job would be to maximize this ratio. So, that maximization of that ratio would be maximization of this term we should mean maximization of this term. So, maximization of this term now you could easily recognize this as i x y and which is the mutual information.

(Refer Slide Time: 23:45)



So, what we can get an indication from this particular explanation is that why do we have that channel capacity denoted as max of i x y over all possible distributions of the source symbols either f x or p x in case of probability mass function. So, what we have indicated over here is that this is the definition of capacity that 1 would typically follow and we have given a intuitive explanation of y such a description comes. So, using if we look at this description we have i x y which we have defined for both this discrete and random and discrete and continuous variables we have defined h y and h of y given x and what we now need to do is to is to find the distribution of source symbols or the distribution of the source for which this explanation is maximized.

Again just giving you hints we have seen that the distribution which maximizes this capacity expression or this mutual information which is determined by h y is again we have derived for continuous random variables it is the Gaussian distribution. So, we will continue this particular lecture and in the next one we will see that the Gaussian distribution maximizes the mutual information between the source and distribution for a Gaussian channel and that we will again finally, lead to the expression of capacity.

Thank you.