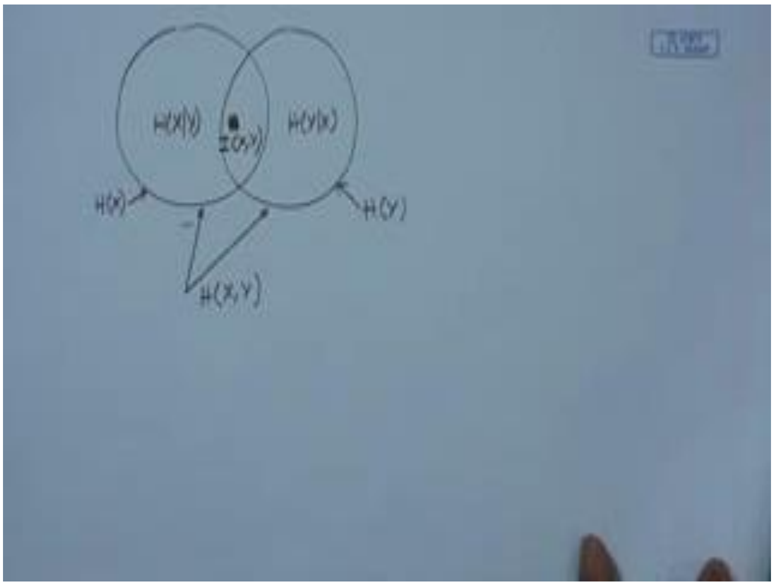
Fundamentals of MIMO Wireless Communication
Prof. Suvra Sekhar Das
Department of Electronics and Communication Engineering
Indian Institute of Technology, Kharagpur

Lecture - 32
Fundamentals of Information Theory – II

Welcome to the lectures in fundamentals of MIMO Wireless Communications. Currently we are looking at fundamentals of information theory, where we are doing a revision of the important definition and terms which are useful. Finally, in the expression of capacity from MIMO systems the last expression that we have derived in the previous lecture is that of mutual information.

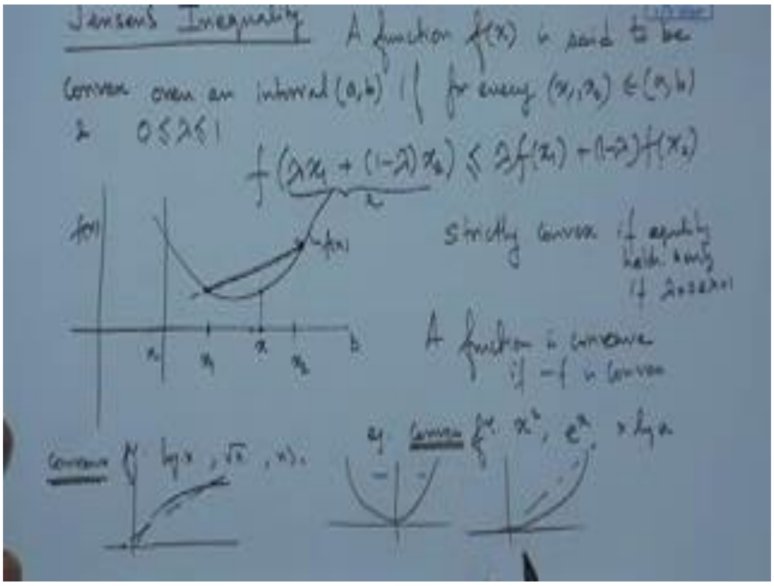
(Refer Slide Time: 00:41)



So, there is also corresponding a diagrammatic representation of mutual information through when diagrams or set diagram. If this we would be called as h of x if this is h of x and this is h of y and this this particular part in in that case basically this is I of x comma y. So, if you look at I f x comma y I f x comma y is it is h x plus h y minus h x y.

So, basically this together is h of x y right, and it could also match with this with this result that what we have is it is h of y minus of h of y given x. So, h of y this particular part is h of y t comma x and this is h of x given y. So, in that case all those things match. So, I of x y is equal to h of y minus h of y given x I of x y is basically h of x minus h of x given y this I am taking out and the finally, the last one is this h of x plus h of y minus this one, what we will be left with this gets added two time if, I do h of x plus h of y and then I take away the union of them union in which this appears only one in h x it gets added one in h y this section gets added once. When I take out the union I take out this I take of this part I take of once from this one is left. So, that is also given by this description.

(Refer Slide Time: 02:51)



So, this is also well described through this particular diagram diagrammatic representation. So, with this we move forward to define an important inequality known as Ensens or Jensens Inequality this is a very, very important inequality which we will use in some of the proofs, where we show the positivity of the some of the important constraints and from which also derive some of the important results. So, with Jensens inequality we start by saving that function f of x is said to be convex over and interval a b if for every x1 comma x2 which is element of a and b and or certain lambda which lies between 0 and 1 if f this function f of lambda times x1 plus 1 minus lambda times x2 is less than or equal to lambda times f of x1 plus 1 minus lambda times f of x2.

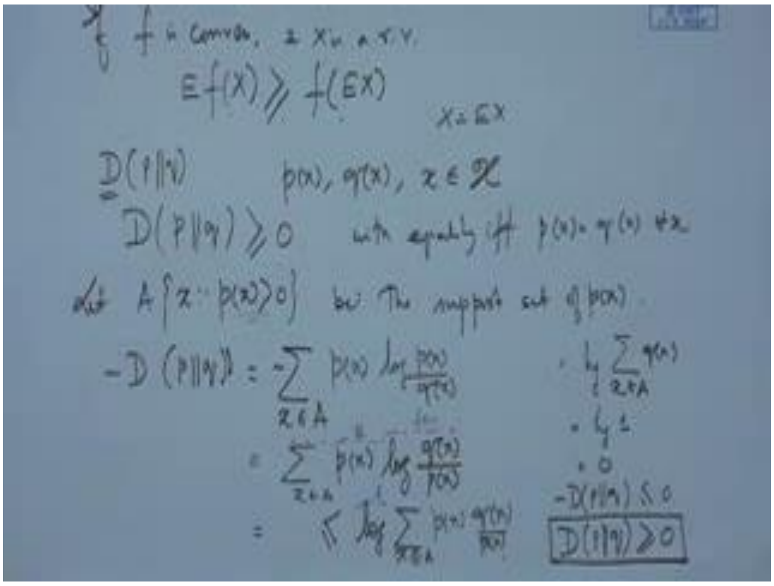
So, what this is same effectively is that if I have an interval between a and b I take 2 points x1 and x2 in this. Lambda lying between 0 and 1 and lambda times x1 plus 1 minus lambda times x2 is something in between this. Because if you take lambda equals to 0 you are getting x2 if getting lambda equals to one this term becomes 0. It is basically x1. Any other value would be somewhere in between. So, f of a value which is somewhere in between is less than or equal to the value at that points. If this is f of x sorry this is the point a, that I have even I would draw this as the axis this is x f of x. So, whatever is the value the f x here and let say a f x has a value there, I take these values and again if you see lambda and 1 minus lambda is basically taking the mean of it.

So, if let say f x goes like this, f of x lambda times x1 plus one minus lambda times x2 is somewhere here and that is this value is x this is f x. So, this is x, f of this x goes here and f of x1 is here f of x2 is here. So, any lambda times f x1 when lambda is 0 this becomes a value x f two. When lambda is 0 at x2 it is a f x2 when, lambda is one this is x1 this is f x1 because these term goes to 0 as. So, lambda anything between 0 and 1 would be any value between these 2 points.

So, basically it is same that the function value of the function between a and b lies below the cord connecting these 2 points. So, that functions is a convex function. So, that is what is described by this convex function and the function is set to be strictly convex. If equality holds and only if lambda is equal to 0 or lambda is equal to 1; that means, equality holds only at these 2 points in that case it would be said that the function is strictly convex otherwise it is just convex function again this will be used in some of the important proofs and in this domain of MIMO wireless communications. This is very important understanding an important function in inequality that will be used often.

So, a function is concave we could write function is concave if minus f is convex we say reverse of it. So, this is just description of it a, some of convex functions examples of convex functions b x squared in to the part of x, x log x. So, if we look at e to the power of x it goes like this it is a cord if we look at x squared it goes like this. So, there is utilizing below the cord and so on and so forth. Concave function one of them is log x into y x for x greater than 0. So, here we would see that deformation of log x typically would be going like this. So, the function lies above the co ordinate case. So, these are concave functions.

(Refer Slide Time: 08:17)



So, these are convex functions right. So, this they taken be used and the result of this definition we could write that if f is convex and x is a random variable we could write expected look at this is the average value. So, that e of f x is greater than or equal to f of e of x.

So, we could write it in this formula concise location and if it is strictly convex we could write that the equality holds and in that case x equals to e of x; that means, a for the case when x is constant this will be definitely curve discretely convex, we move on the forward and we would like to look at this, if this d of p given q would like look at a this particular thing given the jensens in equality. So, let us say that p of x and q of x due to distributions when x is element of some set right. So, this will be 2 probabilities the mass functions we would to do prove that d of p from q is greater than or equal to 0 and with equality if p is equal to q. So, with equality if and only if p of x is equal to q of x for all x right this is what we would like to see. So, let for this we define let a be defined as x for which p of x is greater than 0 and is also defined as be the support set of p of x right in that case we would begin by writing minus d p from q could be written as minus x element of (Refer Time: 11:00) for the proof we would started with the minus you could have done it with the plus also things would have been similar, but here would followed this way p of x log p x by q x this as per definition.

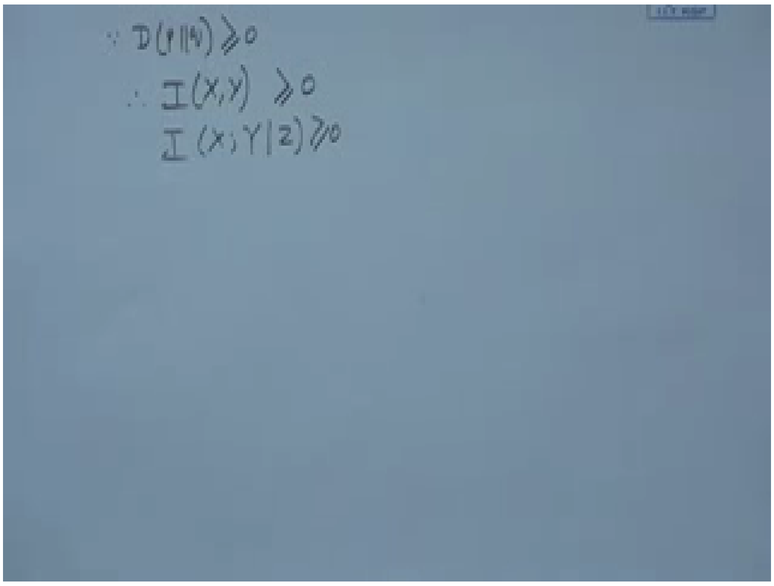
So, getting over the support what the support means we are taking the case for which p x is greater than 0; that means, we are not taking the case p x is equal to 0. So, this is what we have written. So, we could write this as x element of a p of x is now since we move the minus side in side we could write this as log q of x by p of x right. So, this is just minus I moving in that is this gets inward because this is raised to the power of minus 1. Now we would use this Jensens inequality at this point yeah we will use the Jensens already we have it here we already have it here. So, this particular function we could say because log is a concave function is less than or equal to log of some of x element of this support set of a p x times q x by p x look at this what we have done, now for this for convex functions we have e of f x is greater than or equal to f of e x.

Now, what we have is e of f x this is the e this is the whole thing is basically the expectation of pressure of f x right. So, what we is we have this for convex function is greater than, but for concave function it is reversed it is less than. So, that is what we have over here. So, log of e of x log, log is a function comes over here. So, this is the function and expectation moves inside. So, here now it is straight forward p x p x can cancels. So, we could write this as equal to log of sum over x element of this support set of a times q of x. So, over this support set of x this term that is sum over the probability mass function is equal to 1 then only there is a probability mass function and this as only none 0 values. So, basically this is equal to log of one which is equal to 0.

So, now what we have minus d p from q is less than or equal to 0. So, basically we have minus d p from q is less than or equal to 0 or d of p that even could be is greater than or equal to 0 this will again be used in some of the further definitions. So, one of the things which we can easily see from this particular thing is that d is the related entropy mutual information is relative entropy. So, sorry this is also defined in terms of relative entropy of the joint distribution with respect to the independent distributions the marginal distribution product of marginal distributions.

So, if this are relative entropy is not negative in that case mutual information is also not negative we have seen entropy is also non-negative. So, if we would go by the definition of mutual information that is a h x minus h of y given x. So, that is what we have h x minus h of x given y. So, in that case it again some important interesting results would come out because of a particular result that we have been able to derive in this particular case.

(Refer Slide Time: 14:17)

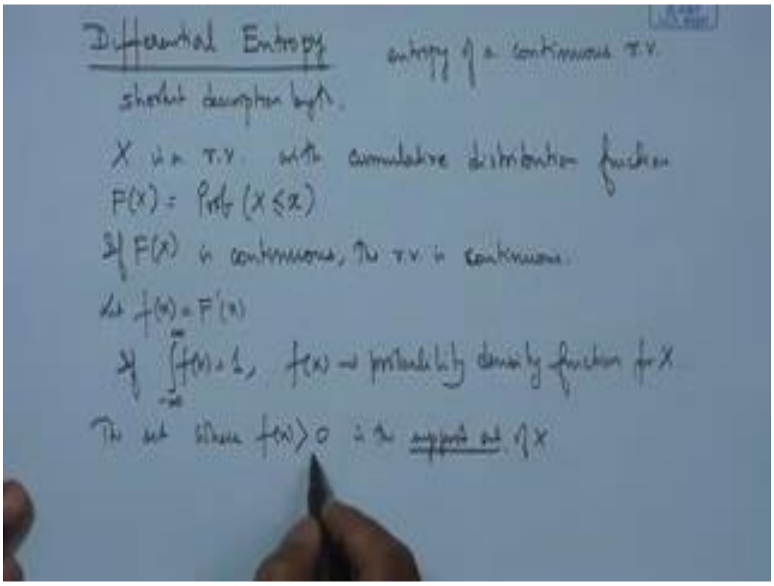


So, we look forward we would write them down here first in foremost since because d p from q is greater than or equal to 0 therefore, we could say that I of x y is also greater than or equal to 0 right and we could say that I of x and y which is conditional z it should also be greater than or equal to 0. So, these are some of the important results that we have from the previous definition; that means, this particular definition that we have got.

So, with this a description whatever we have described till now, is basically for discrete random variables. So, as we have said that we have these discrete sources and we have this discrete channel like this binary symmetric channel and in other cases we would also have the wave from channels. So, wave from channels are continues. So, when we talk about continuous channels we cannot what with these discrete sources with these discrete random variables.

So, we would also need to define this entropy and relative entropy and everything for continuous random variables. So, we will move on to a definition for the appropriate entropy term when we have continuous random variables, once we are done with these things then we will move on to look at the expression of capacity for a discrete sources and discrete channel and then, we will move on to the wave form channel we are all we will be using the results as has been described in the previous lecture as well as in this particular lecture.

(Refer Slide Time: 16:02)



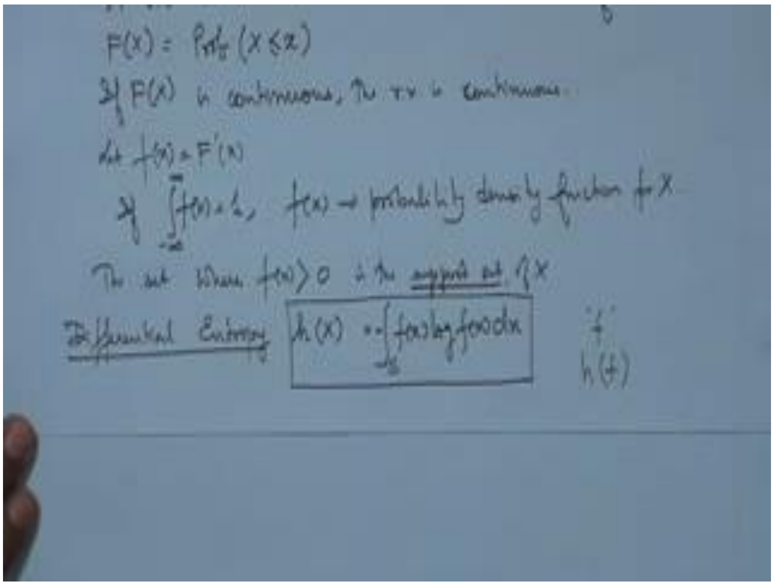
So, we move further and we would like to define differential entropy right. So, differential entropy is basically the entropy of a continuous random variable now you will see the difference when, you look at the continuous random variable and that is why we need to give a different term and we will not use the term as entropy be differential entropy we would not go and explain the details of why it could be possibly called the deferential entropy, but rather we would see the implications of a continuous random variable and why it should be handled carefully and why at least different name should be given.

So, we look for go ahead. So, we are looking at the entropy of a continuous random variable. So, this is also a related to differential entropy is also related to the shortest length of random variables this is also similar to the entropy, but there are certain is which are different which does not make the entropy of a continuous random variable to be treated exactly as a same way as that of discrete random variable. So, move forward with this these are also gives us the shortest description length.

So, this is this also a gives us and now we would say, that suppose x is a random variable with cumulative distribution function f x defined as the probability that x is less than a particular x. If f x is continuous then we would say the random variable is continuous and we would say let f of x is equal to f prime of x f prime of x is indicating a the derivative right and when, this exists and if a integral minus infinity to infinity f of x is equal to one we would call f of x as the probability density function right probability density function for x. So, these are some of the things which are necessary and we would call the set where f of x is greater than 0 is the support set of x.

Now, this is important because we are not taking the case where a f of x is equal to 0 because that is not the supported set.

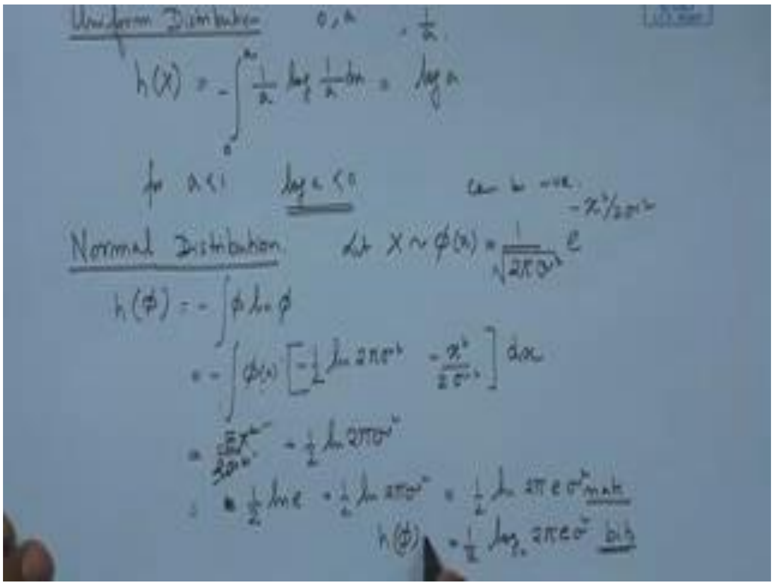
(Refer Slide Time: 19:36)



So, supported set would mean the range of x, where there is a non 0 probability of x occurring otherwise we would have almost infinite range of values of x that we have to deal with. So, with this we move forward and define the differential entropy h of x. So, now, we will use the term small h instead of capital H which we would define earlier of the continuous random variable x as integrate over the support set s indicates the support sets we are not integrating over the entire range of f of x log of f x d x with a minus. So, this defines the differential entropy. So, for continuous random variable instead of using the term entropy, we would be using the term differential entropy and we will be almost deriving the expressions as we have done for the discrete random variable and here. Also if you further see that this description is depended on f.

So, since this is depended on f it is sometimes also represented as h of f. So, because this you will be defined over the support set of s. So, that that is very, very important now, just to see one of the important implications that when we take this as a continuous random variables.

(Refer Slide Time: 21:00)



So, suppose i-i take uniform distribution, suppose I take uniform distribution and for which we say that the random variable is distributed between 0 to a and so, it is density is basically one by a. So, if I have to calculate h of x h of x would be minus mean by this description it ranges from 0 to a it is defined in entire range one upon a log of 1 by a d x.

So, this integral would lead to log of a. So, whenever for a less than 1 what do we get we get log of a less than 0 so; that means, what you are getting is h of x which is the differential entropy can be negative. Whereas what we have seen for entropy of discrete random variables it is non-negative it is a 0 or greater than 0, but here what we have seen is that it can be a negative value. So, therefore, we cannot give with the same as entropy because we have a different variable to deal with that is a continuous random variable and therefore, we would we would usually call it differential entropy right. Moving forward again we have an important result which is also used further. So, if we would take the normal distribution.

So, when we take the normal distribution we let x distributed as phi of x this is the notation that x distributed is phi of x and for normal distribution it is well known it is given as one by 2, 2 pi sigma squared where sigma is the standard deviation of x e to the power of minus x squared by 2 sigma squared. That means we are taking 0 mean when you say normal distribution we are taking a 0 mean and the description that we are giving is given by this particular expression. So, if we have to calculate h of. So, as we had said we could use h of f.

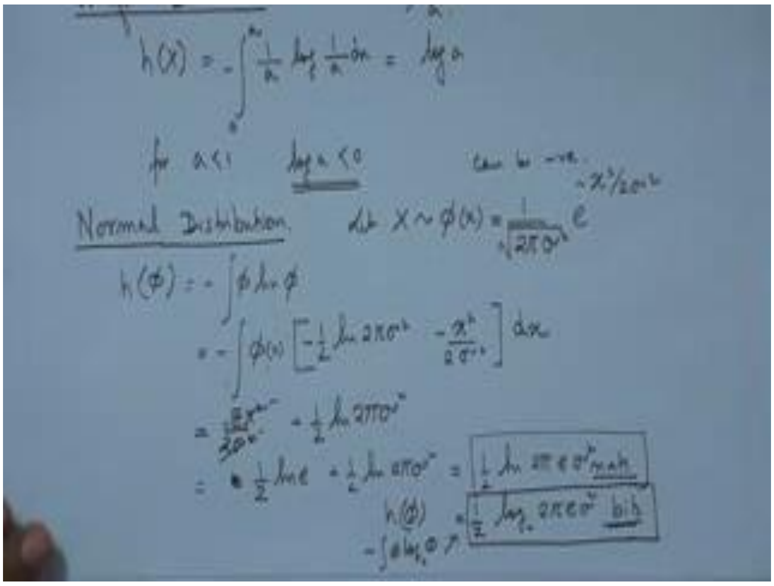
So, since phi is denoting the normal distribution in this case h of phi is basically minus 5, l n phi. So, we are taking I n for sake of inside this particular case we could do it for log I am just giving a log two result at the end of it. So, this is equal to integral 5 of x times. So, if we take the natural logarithm of this what you end up with is this term. So, log of this plus log of this log of this is basically minus half l n 2 pi sigma squared right and a the other part what we are left with is a when, you take a log of it this goes put.

So, basically minus x squared upon 2 sigma squared, this is what you are left with in this case. So, minus sign a minus in a, a minus sign is all minus sign go away. So, you are left with a plus sign. If you are taking this particular if you are taking. So, here we have d of x right. So, if we are taking this particular term let say this particular term, we have 1 by 2 sigma squared integral phi x x squared is basically e of x squared.

So, in this term is multiplied this term is integrated this particular term with this particular term 2 sigma squared is bellow 5 x x squared d x is basically e of x squared, because this is the probability of density function and what you have is the rest of the terms plus sign or half it is a constant l n 2 pi sigma squared integral phi x d x is basically 1. So, we have to only this term this we could write it as e of x squared is basically sigma squared. So, this and this goes out. So, we have half from half we are going to write instead of 1 I could write l n of e so; that means, it is the same thing basically this goes out with this two remains a plus half l n two pi sigma squared.

So, what you have there that leads to half l n 2 pi e sigma squared in n ats, because we have taken natural logarithm over here and if you take in bits. So, it would turn out to be h of phi in bits would turn out to be half log base 2, 2 pi e sigma squared in bits. So, for arriving at this instead of taking l n over here 1 as to take log base 2 and then you would arrive at this particular expression.

(Refer Slide Time: 26:43)



So, you can try this as exercise instead of taking. So, basically h phi when it is defined as minus then it would be defined as minus integral phi log base 2 of phi. So, through this you are going get this particular expression there. So, this is again a very important result which we will use later on right.

In this particular lecture, we would stop at this point where we have started a journey into the differential entropy descriptions of continuous random variables we have to do something more in the future lectures. When we finally, come up to the expression of capacity of which will be using in defining capacity of MIMO channels.

Thank you.