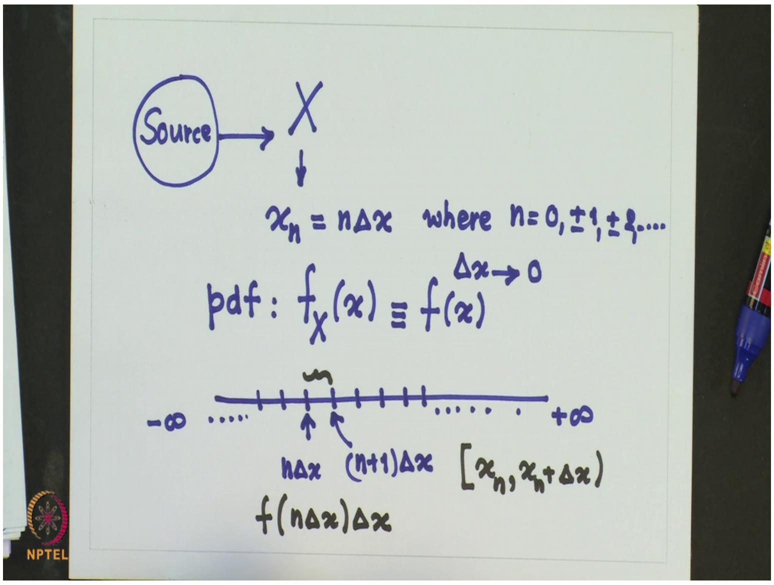
Principles of Digital Communications
Prof. Shabbir N. Merchant
Department of Electrical Engineering
Indian Institute of Technology, Bombay

Lecture – 10
Differential Entropy – I

Hello, welcome back. So, far in our study we have limited our discussion to discrete sources and discrete channel. Now we will address the issue of defining information measure for sources which generates continuous data or analog data. The motivation for this is basically to arrive at the Shannon's important theorem which relates the channel capacity to the engineering parameters like bandwidth and signal to noise ratio.

So, let us look at a source which generates continuous data and we will model it as follows.

(Refer Slide Time: 01:21)



The output of the source would be treated as a continuous random variable X correct. So, this continuous random variable can be viewed as a limiting form of a discrete random variable that assumes the value x n equal to n times delta x; where n is equal to 0, plus minus 1, plus minus 2 up to plus minus infinity and delta x approaches 0 correct.

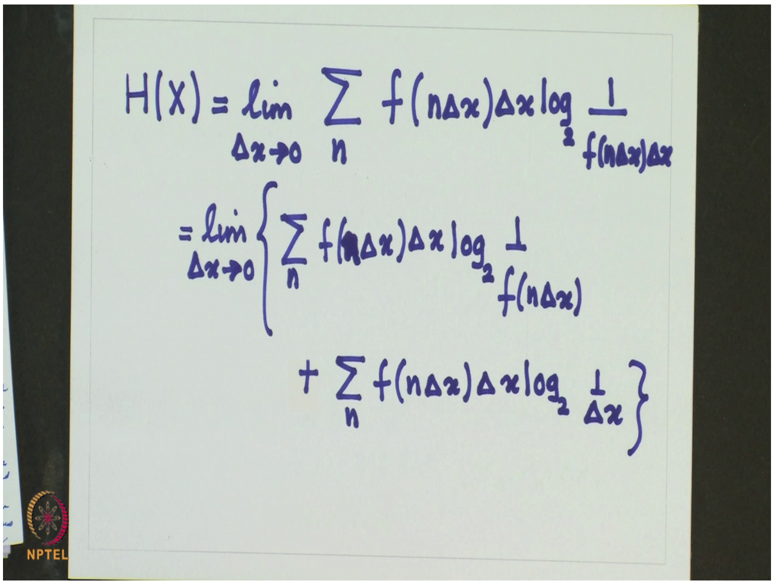
And with this continuous random variable, we have what is probability density function associated with it; we call it in short PDF. So, this PDF would be given by this function shown here, for sake of simplicity many a times we will write this as just f x we will strip it off the random variable x from here just to make the writing the expression little easier.

Now, without loss of generality let us assume that this random variable domain is from minus infinity to plus infinity and since we are considering it as a limiting form of a discrete random variable, what it means basically it is divided into small intervals of size delta x correct ok. So, if you take any particular interval out here correct this is n delta x and let me call this is n plus 1 delta x correct.

Now, by definition the random variable x assumes a value in this interval which is your x n, x n plus delta x. So, the probability of random variable x lying in this interval would be given by f times n delta x multiplied by delta x approximately. Yes delta x tends to 0 this approximation becomes better ok.

So, now let us try to extend the definition of entropy for a discrete messages or discrete random variable to this continuous random variable, but we will again view it as a limiting form of a discrete random variable.

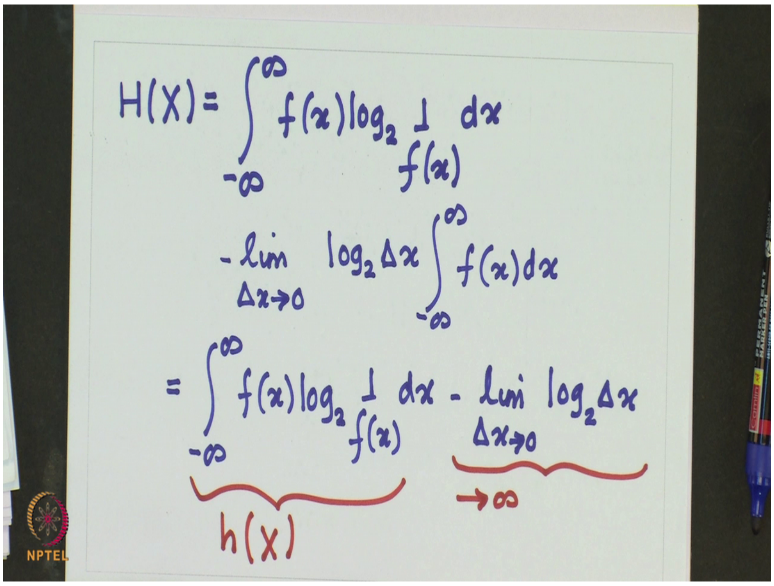
(Refer Slide Time: 05:13)



In that case I will be able to write my entropy as; this is the probability correct of that interval and then log to the base 2 of 1 by correct.

So, permitting delta x tends to 0 the ordinary entropy of the continuous random variable X takes this limiting form ok. So, this we can simplify as n time this is sorry this is n times delta x delta x log to the base 2. So, what I have done here basically this expression I have broken up into two parts ok. So, this is what I get from here ok.

(Refer Slide Time: 07:20)



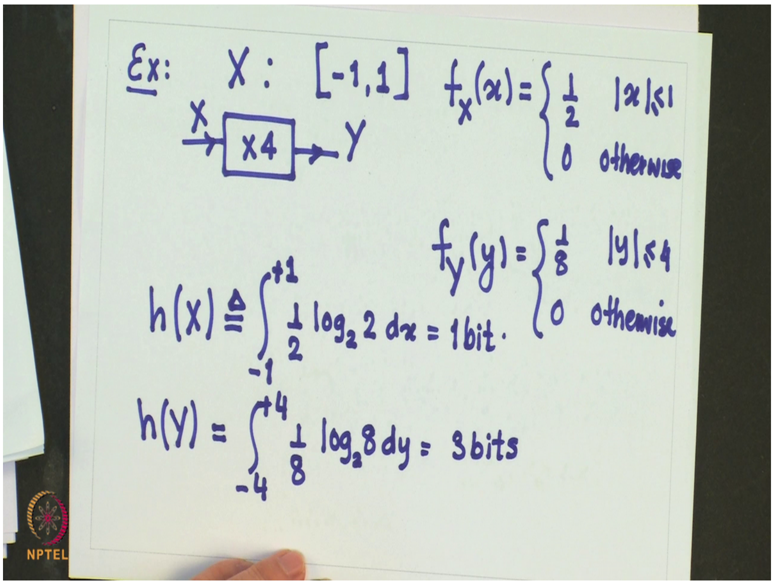
So, now if you take the limiting case when delta x tends to 0 what I would get is H X is equal to so this comes from the first term here; this is basically from the first term and the second term is rewritten as follows. So, this I can again rewrite as because f x is the PDF. So, the integral here is equal to 1; so I get this quantity.

Now, look at this quantity if you look at this quantity including the minus sign; this will tend to infinity as delta x tends to 0. Now intuitively this is expected. So, what will happen basically when delta x tends to 0, this tends to infinity. So, this right side basically will tend to infinity and the reason basically is obviously, because there are infinite number of values between say minus infinity, plus infinity or between any particular range. So, the uncertainty is very high of getting a particular value correct; the probability of getting a particular value is almost 0 correct. So, if you look from that point of view and to the uncertainty becomes infinity and the entropy should be equal to infinite ok.

So, now this is infinite I cannot define this to avoid this problem associated with the term log 2 delta x; we basically define this quantity as what is known as differential entropy and this term basically serves as a reference correct. This is actually acceptable because remember that for information transmission, we are actually interested in the difference between the entropy terms that have a common reference correct.

So, the if they have a common reference the difference of that information of the difference between the corresponding differential entropy terms will be still finite correct, if the reference is same. Therefore, it is perfectly justified for me to define this term as information measure; it is a differential relative measure for a continuous source. Absolute entropy obviously, is infinite for a continuous source intuitively it is fine.

(Refer Slide Time: 11:00)



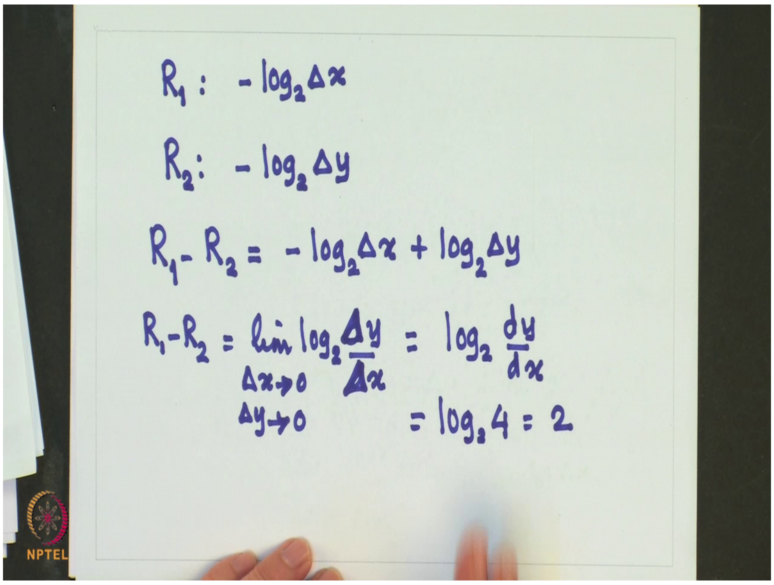
So, let me take one example to help you appreciate this. So, let us assume that we have a random variable X which is uniformly distributed between the range minus 1 to say plus 1 correct and this signal is passed through an amplifier of gain say 4; an output I get basically another random variable which is Y.

Now obviously, from here we can find out the PDF of X; that would be equal to what? It is a uniform random variable. So, this will be equal to half for x less than equal to 1 or is equal to 0 otherwise; correct and the PDF for random variable Y would be given as; so now the range will be minus 4 to plus 4; again it will be uniform distribution correct; uniform PDF. So, and this would be equal to 0 otherwise correct.

So, now let us try to calculate the differential entropy for this. So, by definition this would be equal to plus minus 1 your f x is half; so log base 2 1 by f x so this will be equal to 2 dx and this will turn out to be 1 bit correct. And if you take the differential entropy of the random variable Y, this would be equal to this expression correct and this would be equal to 3 bits correct.

So, you get a difference of 2 bits correct, but it is important to realize that in this case what is the reference for random variable X? For random variable X the reference is minus log to the base 2 delta x correct and for Y; the reference is minus log to the base 2 delta y.

(Refer Slide Time: 13:26)

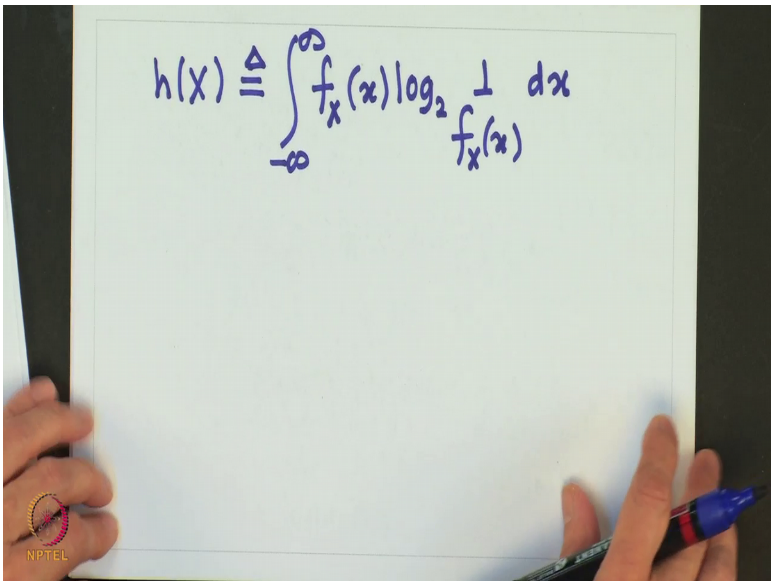


Now if you take the difference between these two references I get it as minus log and if we take this in the limiting case, where both delta x tends to 0 and delta y tends to 0 what I will get it is equal to dy by dx correct; so log of this correct.

So, this basically it is; so limiting case of this would be equal to this. So, actually this is limiting case I should have written this delta y, delta x this is equal to log to the base 2 of dy by dx correct and this is equal to log 2; this is 4 times so this is 4; so this is equal to 2. So, what I get basically that this x and y have obviously, because there is no gain in the information it is just amplified the signal is amplified. So, I do not gain any information as such.

So, from the absolute entropy for both x and y should be the same correct and therefore, you see that this; the there is a difference between the differential relative entropies because of the difference in the references ok. So, so but if the references are same then when you take the difference it will be nullified fine. Now having defined the differential entropy the next question is that this definition; let us define it again.

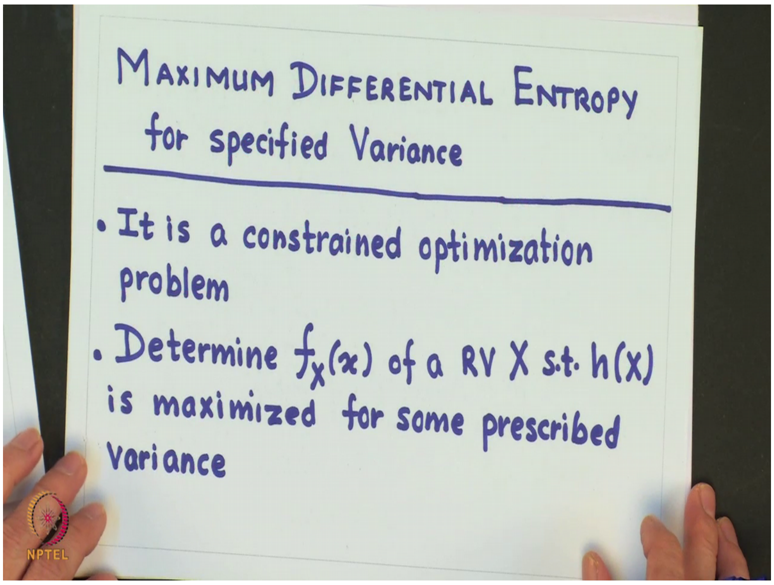
(Refer Slide Time: 15:54)



So, so taking this as your definition for differential entropy now we have seen for a discrete sources that the maximum entropy turns out to be when your probability distribution function for the output is uniform correct. So, is it possible for us to find out a distribution, a PDF for a continuous case such that this becomes maximum correct.

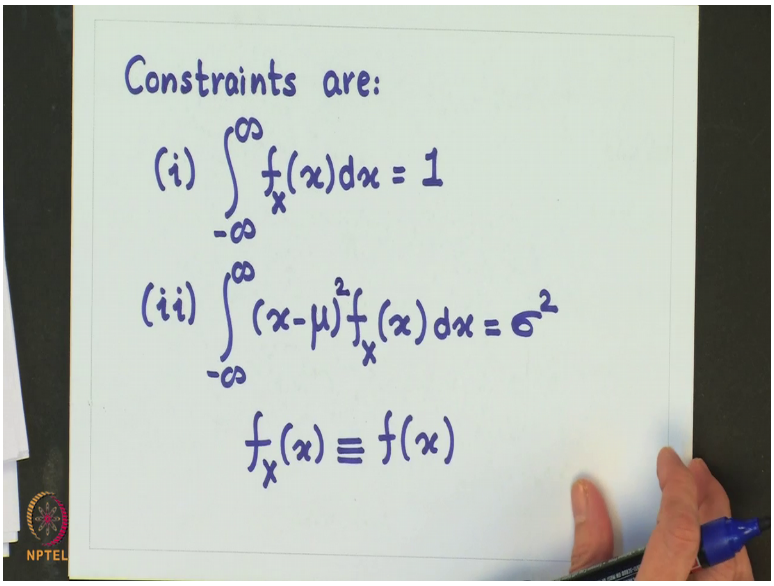
So, now my problem is basically I want to find out the maximum differential entropy, but I will have to put some constraint on x correct. So, I could say ok fine given maximum value of x what would be the maximum differential entropy? Correct, I could solve that problem, but for this class what will consider is basically what is of interest to me is; basically if I give you the variance of this random variable to be a constant and take that as a constraint, then for that constraint can you find out the PDF which will maximize this h X ok.

(Refer Slide Time: 17:29)



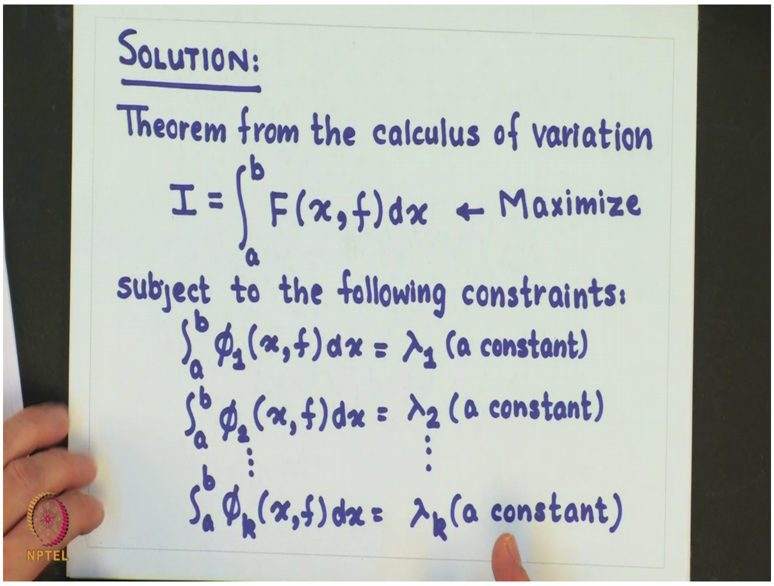
So, my problem is as follows I want to maximize h X correct, but since f x is a PDF there will be some constraint on that and I also specify the constraint that the variance of my input signal is fixed ok. So, this is a constrained optimization problem; my problem is to determine PDF of a random variable X; such that differential entropy is maximized for some prescribed variance.

(Refer Slide Time: 18:18)



So, I specify this variance. So, let us try to solve this; so this is my problem and the constraints are given like this. The first constraint is on the PDF property, it should be 1 and the this constraint has been specified by me that the variance; so this should be the variance of my PDF is a constant, mu is the mean value of the random variable x and for convenience basically we will try to represent this f X x by just f x ok.

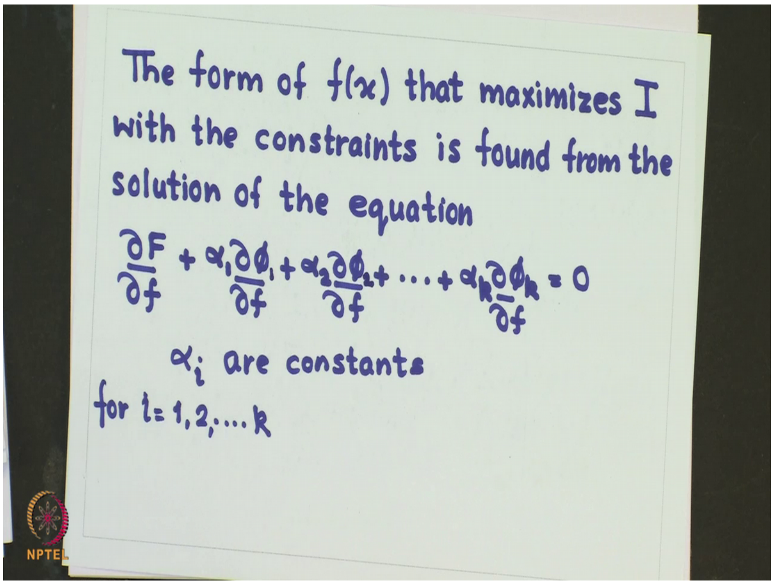
(Refer Slide Time: 19:05)



So, the solution to this basically is provided from the calculus of variation and what does the theorem from the calculus of variation say is this; given this integral I of this form, this is a function of some variable x and it is also function of another function f which could be function of x correct. I want to maximize this subject to the following constraint; so there are different functions are being given up to phi 1, phi 2, phi k and all those functions basically satisfy this constraint where your lambdas are basically constant correct.

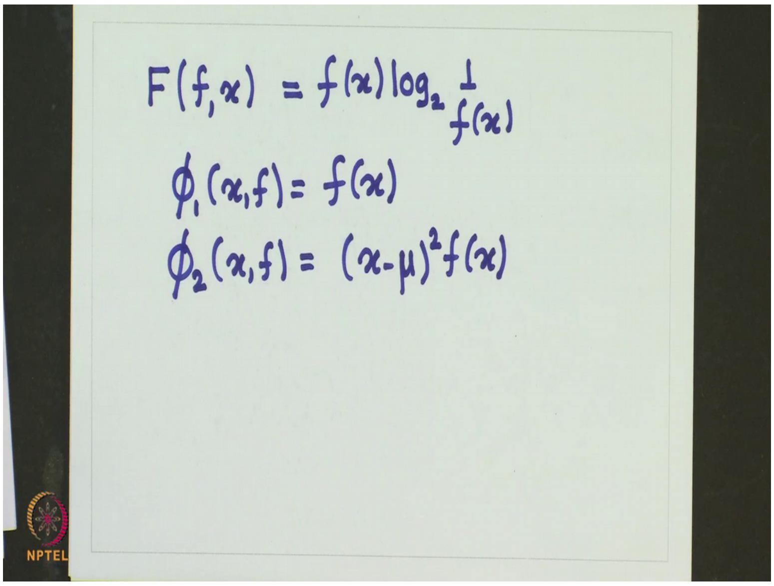
Now, without going into the proof for this; the solution to this problem optimization problem with the constraint is provided here.

(Refer Slide Time: 20:04)



So, the form of f x that maximizes this I is found from the solution of this equation correct. So, you take the partial derivative of capital F with respect to small f and then phi s; these are phi 1, phi 2 up to phi k and your alpha i s are constant for i equal to 1 to k.

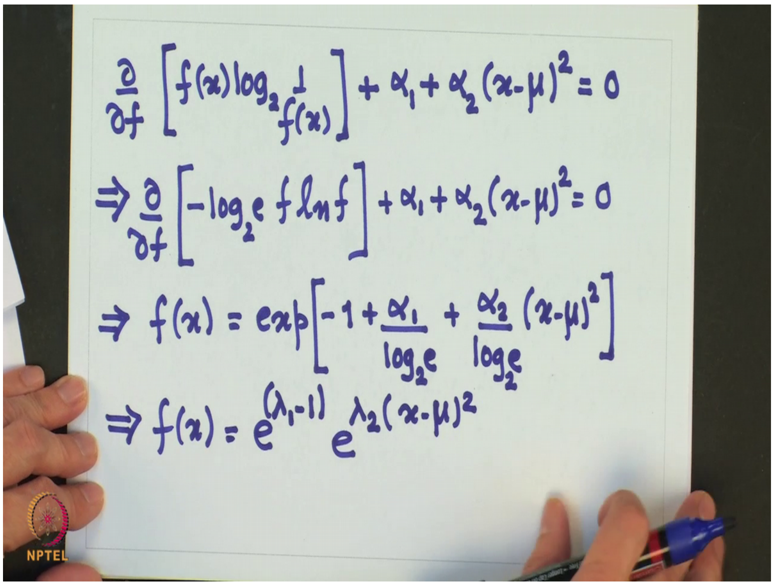
(Refer Slide Time: 20:43)



So, in our case we want to optimize the differential entropy; so our function capital F f x this is equal to f x log to the base 2 1 by f x and your phi 1 x f is equal to f x and your phi 2 x f is equal to x minus mu square f x fine.

So, we have mapped our given function to this capital F and then we will use the procedure specified here to find the solution for this ok. So, let us do it quickly.

(Refer Slide Time: 21:41)

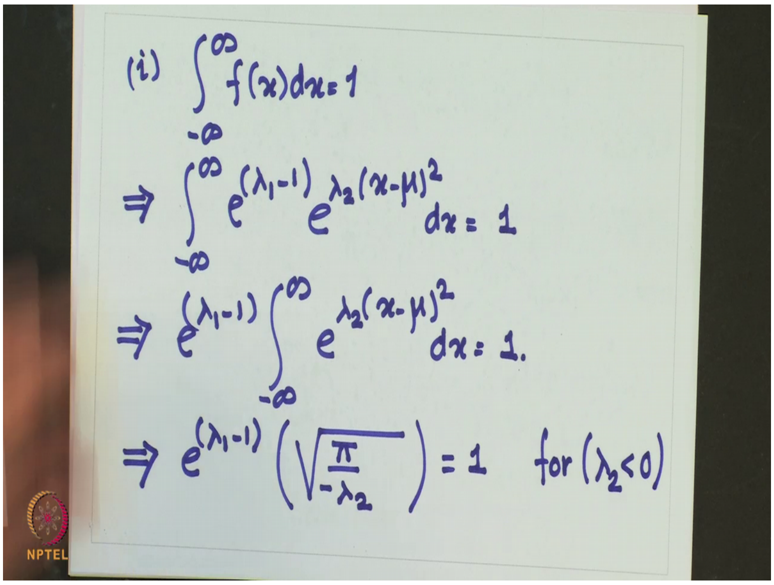


So first thing is take a take the derivative of this and when we take the derivative of this with f x and then you have to take the derivative of this also with f x and after taking a derivative you will have to multiply by the constants alpha 1 and alpha 2. So, it is simple to see that if I do this basically what I am going to get here is alpha 1 plus alpha 2 x minus mu square; x minus mu bracket squared sorry this is equal to 0.

So, now this will imply as derivative of; we will just change this to the log form natural log. So, for the sake of simplicity I am just writing instead of writing f x I am just writing it f here, it will simplify our expressions ok. So, this will imply that f x is equal to exponential simple; I will take the derivative you will get this expression. So, this implies that my f x is equal to e raised to lambda 1 minus 1 e raised to lambda 2 x minus mu square.

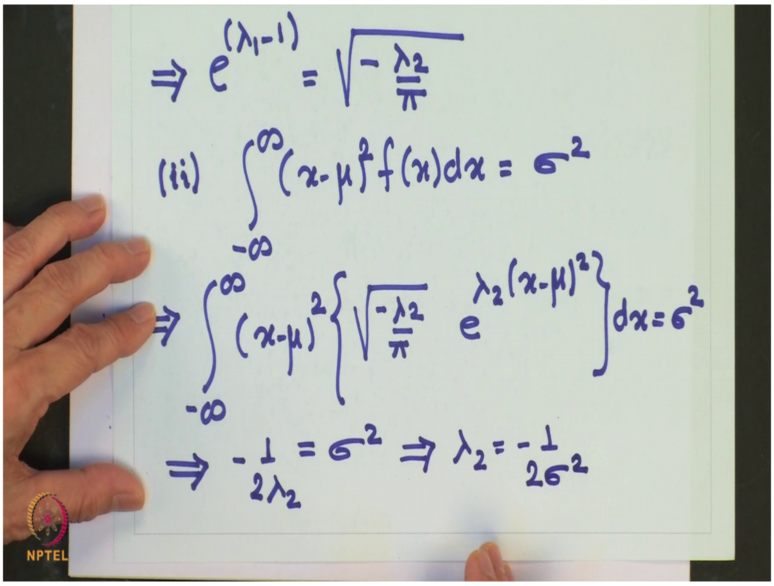
Now, alpha 1 log 2 divided by this quantity I am going to change it as lambda 1 correct and this quantity will become lambda 2; these are again a different constants correct related to alpha 1 and alpha 2 respectively ok. And now we will use this is the expression; I get for my f x correct and then this expression has to satisfy the constraints correct. So, the first constraint we have is basically is this constraint correct.

(Refer Slide Time: 24:21)



So, if you plug in that f x here this will imply that this should be equal to 1. So, this implies that e raised to lambda 1 minus 1 is equal to 1. Now this is equal imply that e raised to lambda 1 minus 1 this integral is standard and you will get it equal to pi by minus lambda 2, this is equal to 1 for lambda 2 less than 0 correct; this has to be satisfied ok.

(Refer Slide Time: 25:42)



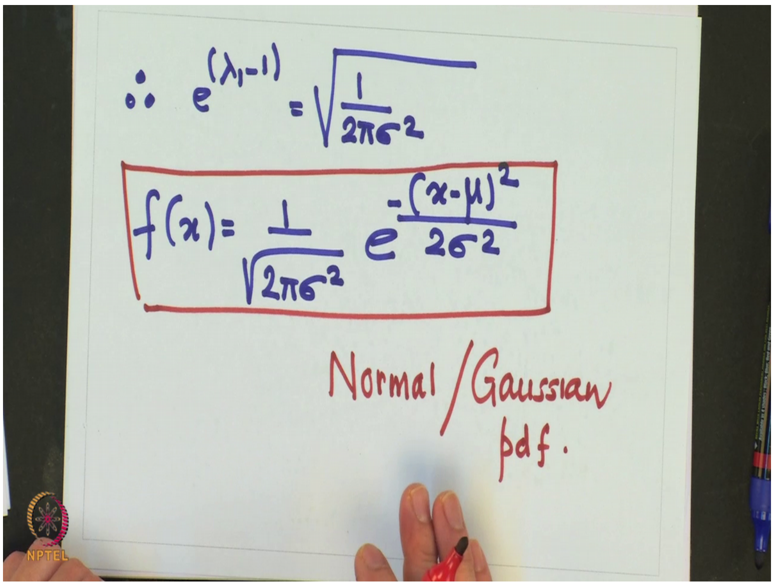
So, straightforward basically and from here I say that this implies that e raised to lambda 1 minus 1 is equal to minus lambda 2 by pi ok and now let us use the constraint second constraint which says that PDF should be equal to sigma square.

So this will imply; substitute the f x which we got just now. So, this will be equal to minus this is for e raised to lambda minus 1 from this here correct, then I write it e raised to lambda 2 x minus mu square d x is equal to sigma square ok.

Now, what you could do is basically here we could change the variables here x minus mu equal to z and then solve it, this is a standard again. So, if you basically if you look at this the solution to this is very simple. So, this implies this and you can even this basically corresponds to something like a normal distribution where you have a mean equal to mu and if you choose lambda 2 which satisfies this and that the above equation will be also true.

So, it is very simple to see that this should be satisfied from here correct; this condition should be satisfied. So, this implies that lambda 2 is equal to 1 is equal to minus 1 by 2 sigma square correct.

(Refer Slide Time: 27:48)



So, from this we get therefore, e raised to lambda 1 minus 1 is equal to square root of 1 by 2 pi sigma square.

So, what I get is my f x is equal to 1 by root 2 pi sigma squared e raised to minus x minus mu squared upon 2 sigma square. So, this is the final relationship which I get and this is not difficult for you to identify as normal or Gaussian PDF correct. So, this PDF will maximize my differential entropy h X given that the variance of my random variable is fixed to sigma squared ok.

So, we will continue the this discussion in the next class, we will try to find out what is the differential entropy for this normal or Gaussian PDF.

Thank you.