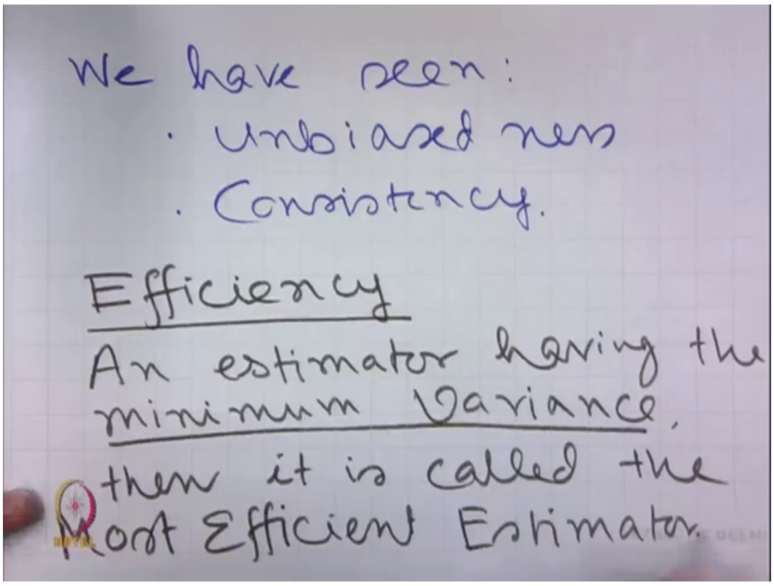
Statistical Inference
Prof. Niladri Chatterjee
Department of Mathematics
Indian Institute of Technology, Delhi

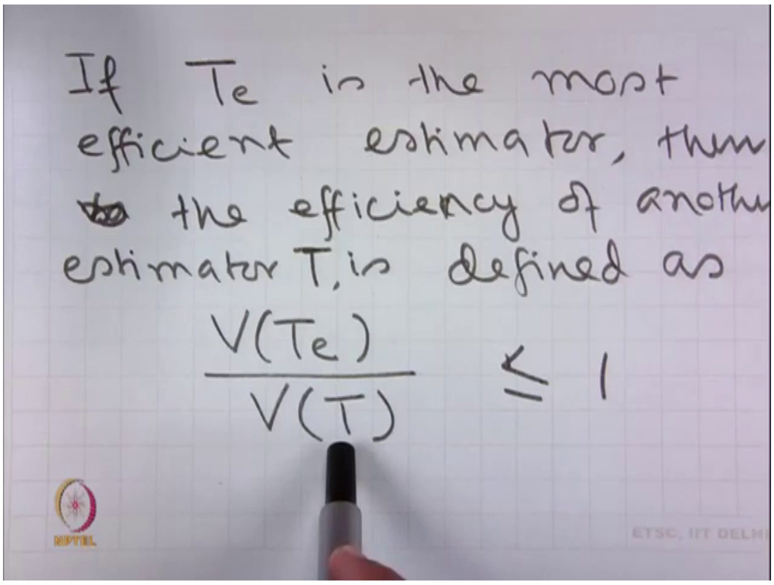
Lecture – 14
Statistical Inference

(Refer Slide Time: 00:42)



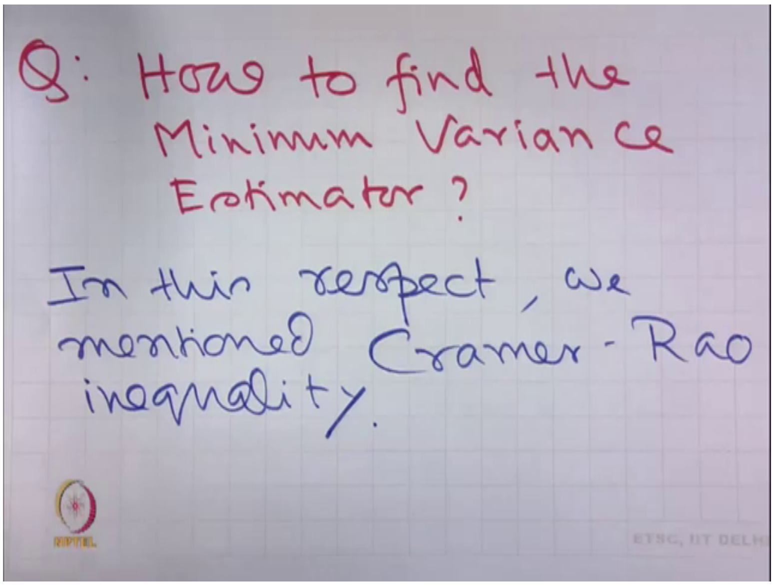
Welcome students to the MOOCs series of lectures on Statistical Inference. This is lecture number 14. In the last two lectures, we have been discussing the properties of an estimators. In particular we have seen two properties, unbiasedness and of course consistency. Towards the end of the last class, I was discussing what is called efficiency. An estimator having the minimum variance; that is, among all possible estimators. If the variance is minimum, then it is called the most efficient estimator.

(Refer Slide Time: 02:10)



In fact, if T e is the most efficient estimator, then variance then the efficiency of another estimator is defined as variance of T e divided by variance of T. So, let me call it T. In other words, if T is any estimator, then its efficiency is measured by comparing its variance with the minimum variance estimator. Obviously, this is less than equal to 1, it is one. If we are looking at some estimator T, whose variance is equal to T, variance of whose variance is equal to the variance of T e. Otherwise, it is more than the variance of T, and therefore this value has to be less than 1.

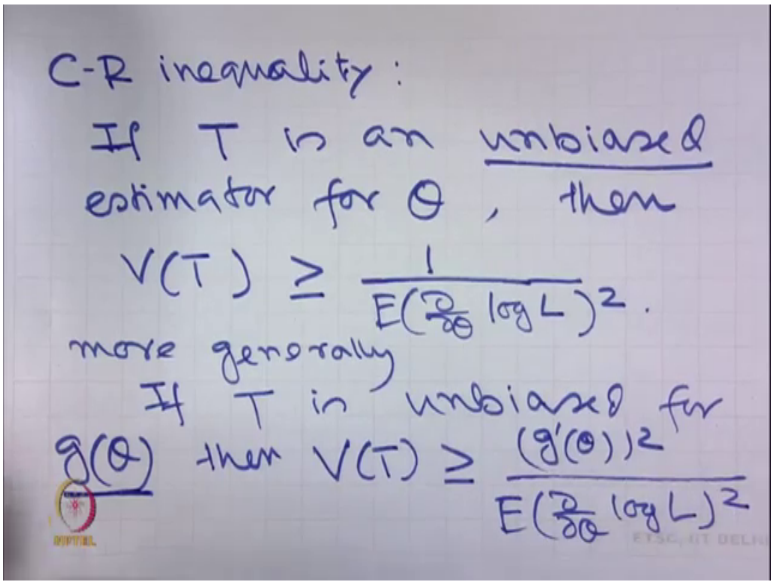
(Refer Slide Time: 03:57)



The question is how to find the minimum variance estimator, this is very important. As there is no (Refer Time: 04:30) of estimators right. We have already seen that, there can be any number of estimator for estimating a particular parameter theta. But, the problem is how do we know~~,~~? ~~this~~ This is going to be the minimum variance.

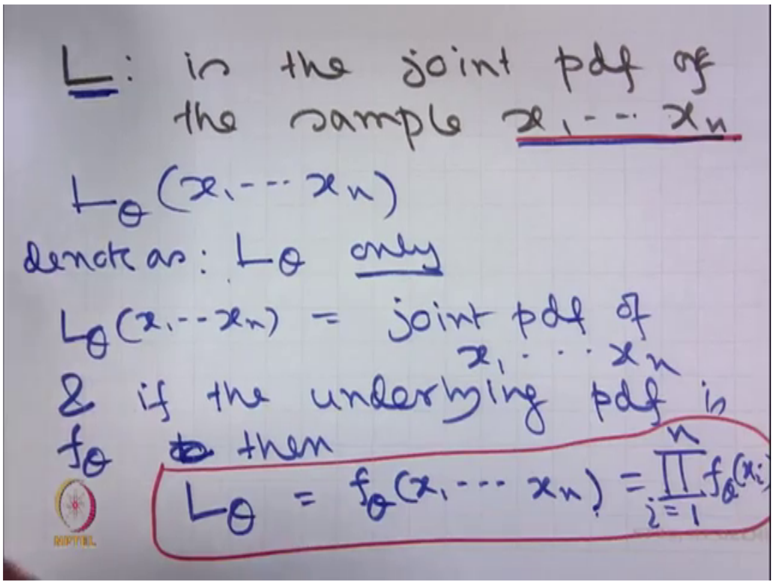
The advantage there by is that, if we know that the minimum variance has to be a some particular value say v, and if we get an estimator T whose variance is same as ~~v~~V, then we know that it is the minimum variance estimator. In this respect, we mentioned Cramer-Rao inequality. This is not for all the classes of estimators. This is applicable to a restricted class of estimator, but still it is very very useful for statistical inference.

(Refer Slide Time: 05:46)



The C-R inequality or Cramer-Rao inequality, it states that if T is an unbiased estimator, so this is very important. We are looking at within the class of unbiased estimators for theta, then variance of T is greater than or equal to 1 upon expected value of del del theta log of L whole square. More generally, if T is unbiased for g theta, which is a function of the parameter theta, and we want to estimate g theta, then variance of T is greater than or equal to g prime theta square upon the same quantity, which is expected value of del del theta log L whole square.

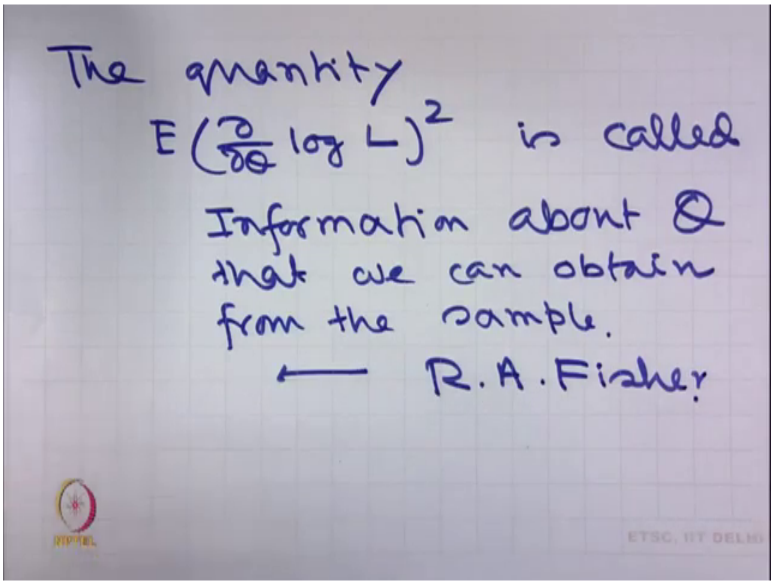
(Refer Slide Time: 07:33)



In this respect, what is L, L is the joint pdf of the sample x 1, x 2, x n or in other words we want to estimate theta or some function of it. We have taken a sample x 1, x 2, x n, we are looking at that joint density function of x 1, x 2, x n, which we call L. Of course, L depends upon theta. So, in some books you may ~~find,~~find L theta x 1, x 2, x n. In this case, sometimes I will be using L theta only, because it is understood that, it is based upon the sample x 1, x 2, x n.

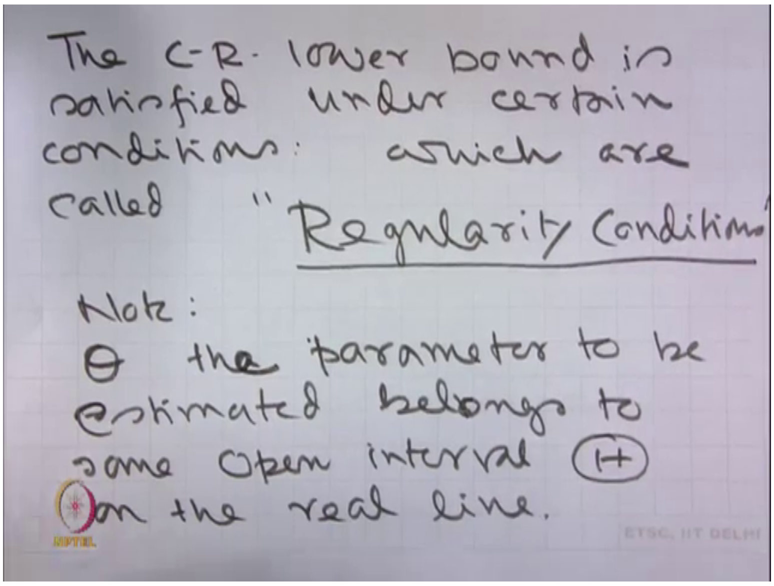
Now, what is L theta or L theta of x 1, x 2, x n, it is joint pdf of x 1, x 2, x n. And if the underlying pdf is f theta, because this is the pdf that involves theta, then L theta is equal to f theta of x 1, x 2, x n. And if they are independent, then we can write it as product of f theta of xi, i is equal to 1 to n. So, do not get confused with this L theta, it is something that we know very well. Since, the samples are independent most of them will be assuming, this form that L theta is equal to the product of the individual density of the sample x 1, x 2, and x n.

(Refer Slide Time: 10:11)



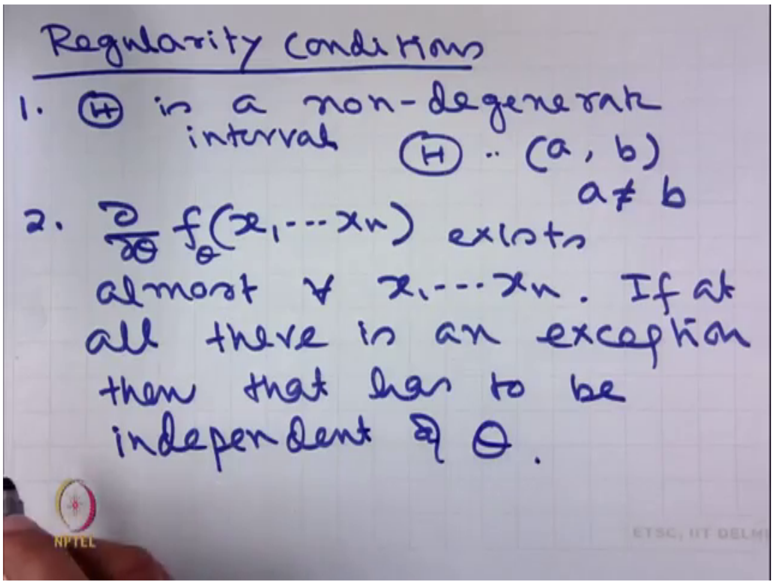
The quantity expected value of del del theta of log L whole square is called the information about theta, that we can obtain from the sample. And this name is being given by R. A. Fisher ok.

(Refer Slide Time: 11:12)



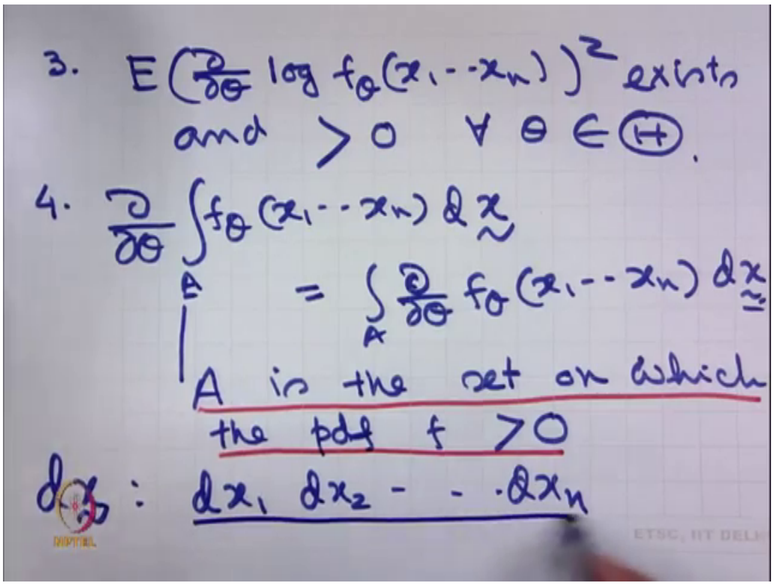
Now, as I mentioned earlier that the Cramer-Rao lower bound is satisfied under certain conditions which are called regularity conditions. Note that ~~theta,~~theta; the parameter to be estimated belongs to some open interval theta on the real line.

(Refer Slide Time: 12:47)



Theta is a non-degenerate interval that means, theta has to be of the form a to b, a not equal to b, so that differentiation with respect to theta make sense. The derivative of the likelihood function exists almost for all x 1, x 2, x n. If at all there is an exception that means, if it does not exist on some interval, if there exists such an interval, then that has to be independent of theta.

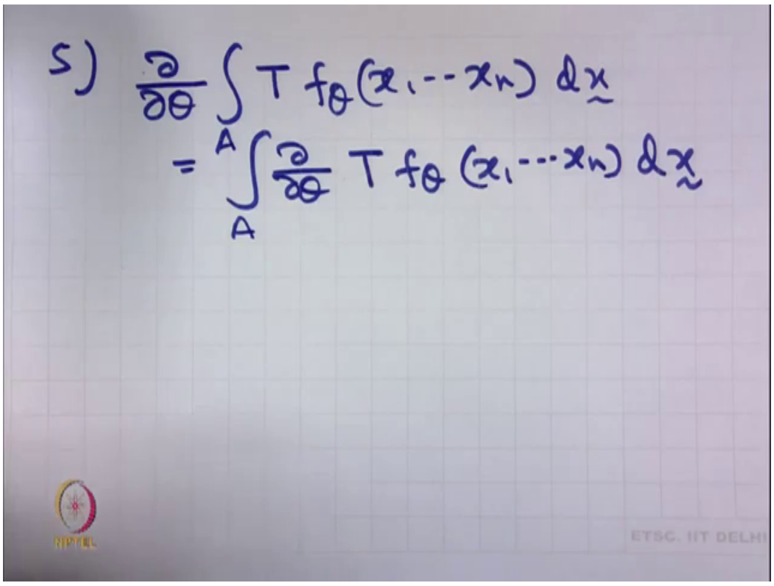
(Refer Slide Time: 14:55)



3, expected value of del del theta log of f theta of x 1, x 2, x n whole square exists and greater than 0 for all theta belonging to theta, because we know in the Cramer- Rao bound this comes in the denominator. So, if it is not defined or if it is equal to 0, then that term is not valid, therefore this has to be one of the constraints. Del del theta of f theta of x 1, x 2, x n, d x is equal to del del theta of f theta x 1, x 2, x n, d x.

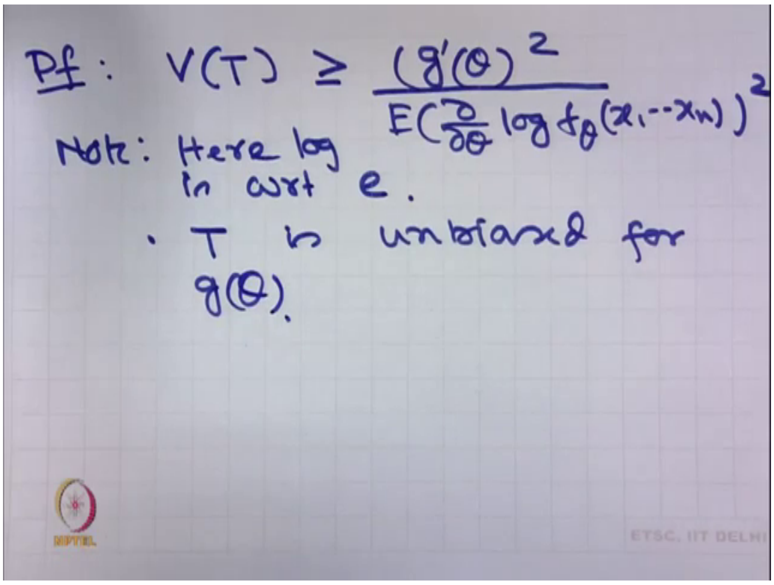
Now, you may ask what is this notation, and what is A. So, A is the set on which the pdf f is greater than 0 or in other words we are looking at only the region on which the probability density function is greater than 0. And what is d x, it is basically d x 1, d x 2, d x n. Say for example, here I am integrating f theta x 1, x 2, x n and to be mathematically precise, I have to integrate it with respect to x 1, then with respect to x 2 and with respect to x n. So, it is basically invariable integration. To make to keep it notational is ~~simpler,~~simpler; I am using that notation d x with a tilde, which actually means this quantity.

(Refer Slide Time: 17:53)



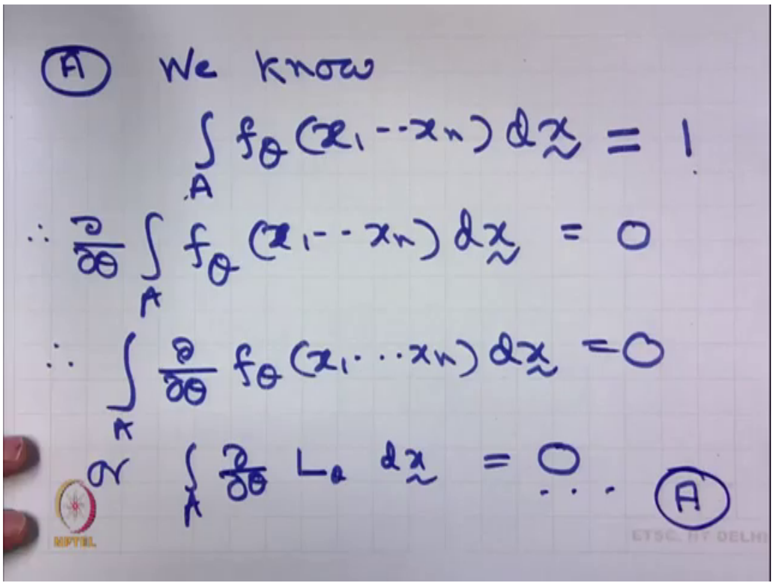
And number 5 is del del theta of the estimator T f theta x 1, x 2, x n, d x is equal to integration of so if you look at 4 and 5, you can see that basically we are allowing the differentiation to moved into the integral sign. And this is possible, if this limit of ~~integration,~~integration does not depend upon theta ok.

(Refer Slide Time: 19:05)



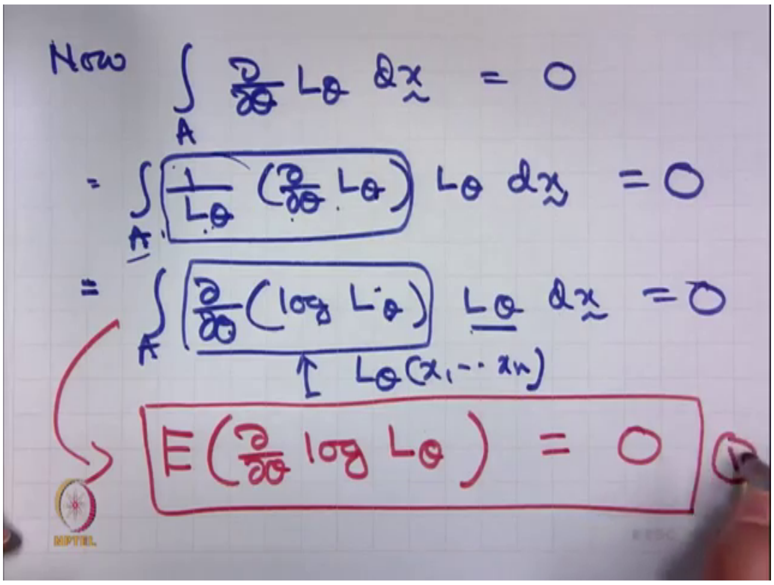
With this assumption, now let us prove the Cramer-Rao bound that is proof that variance of T is greater than equal to g prime theta whole square upon expected value of del del theta log f theta of x 1, x 2, x n whole square. Note, here logarithm is with respect to e and also note T is unbiased for g theta ok. So, we begin the proof in the following way.

(Refer Slide Time: 20:05)



We know, integration over a of f theta x 1, x 2, x n, d x that means, I am looking at the joint density function. And I am integrating it over, the entire possible range, and that has to be equal to 1, because the total probability is 1. Therefore, del del theta over A of f theta x 1, x 2, x n, d x is equal to now, I am differentiating this with respect to theta, and this is going to be 0. Now, by the regularity condition that differentiation under integration, we can write it as follows or we can write it as so this is the first finding A. I will come back to it later.

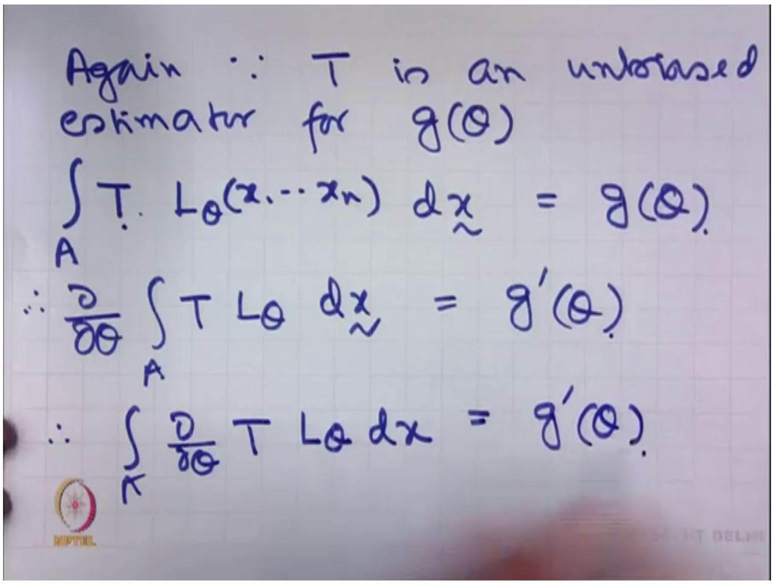
(Refer Slide Time: 21:49)



Now, integration of del del theta of L theta d x is equal to 0. So, let me write it as integration of 1 by L theta del del theta L theta into L theta d x is equal to 0. I am dividing and multiplying by L theta, and since I am considering the range, where L theta is greater than 0, this makes sense.

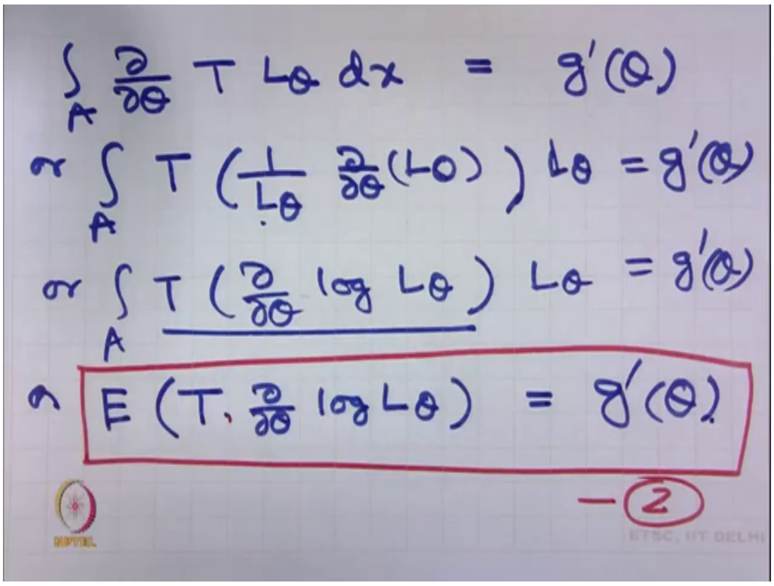
Because, if we differentiate log of L theta with respect to theta and taking that partial derivative, then what we are getting, this is 1 upon L theta into del del theta of log L. So, this entire quantity, I can write it as this. And that we are multiplying by L theta of x 1, x 2, x n and integrating. So, what is this, it is a function of x 1, x 2, x n, because this L theta is L theta of x 1, x 2, x n. Therefore, this whole quantity is nothing but expected value of del del theta of log of L theta. And just now, we observe that this is is equal to 0. So, this is our finding 1.

(Refer Slide Time: 24:06)



Again since T is an unbiased estimator for g theta, integration over a T of L theta, let me write it with x 1, x 2, x n, but as we have seen, I am often leaving out this part. If the notation is understood, d x is equal to g theta, because that is the expectation of T. Now, differentiating it with respect to theta we get, I am leaving out this.

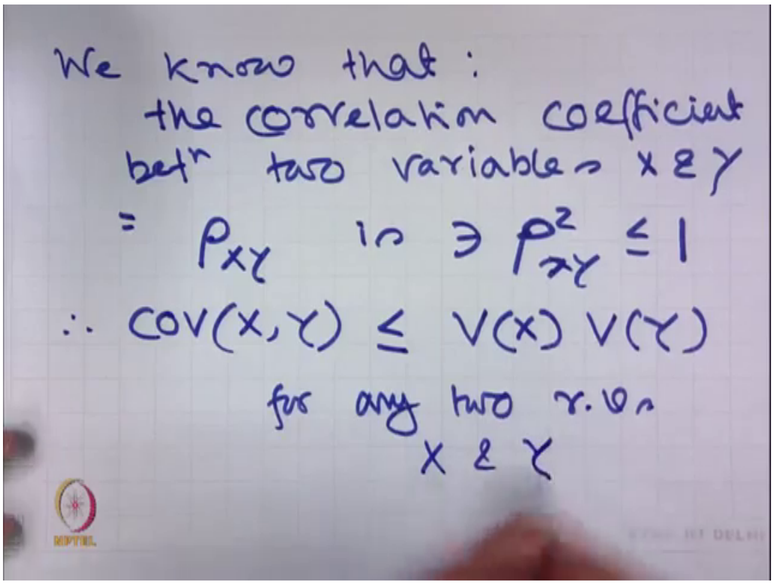
(Refer Slide Time: 25:37).



Now, by regularity condition, we can push it inside or as before, we divide and multiply by L theta. So, I have multiplied and divided by L theta or integration over A del del theta of log of L theta into L theta is equal to g prime theta or this is expected value of T del

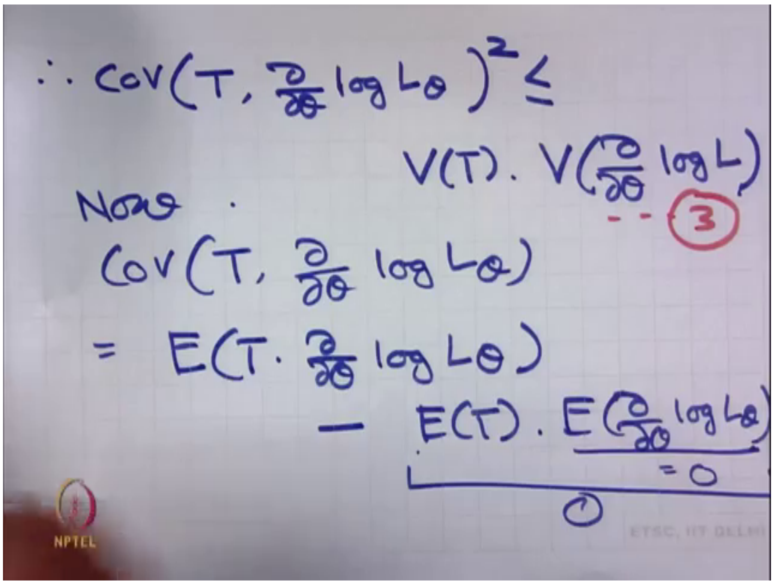
del theta log of L theta, this is is equal to g prime theta. So, if you look at in 1, we have got expected value of del del theta log theta is equal to 0. And here, we have got expected value of T into del del theta log theta is equal to g prime theta. Let me call it equation 2.

(Refer Slide Time: 37:32)



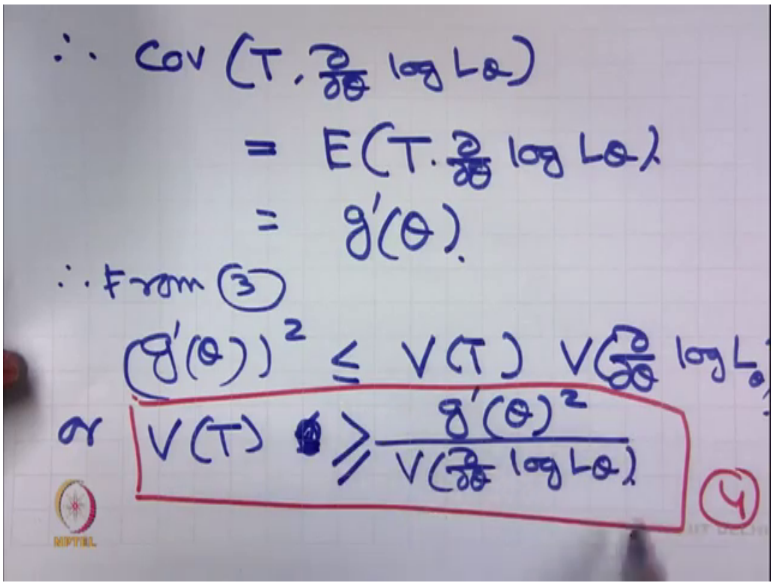
Now, we know that covariance the, we know that the correlation coefficient between two variables X and Y, which we called rho XY is such that rho square XY is less than equal to 1 that is, because the mod value of rho XY has to be less than equal to 1. Now, what is rho? rho is equal to covariance of XY upon root over variance of X and variance of Y. Therefore, covariance of X, Y square is less than equal to variance of X into variance of Y, and this is true for any two random variables X and Y.

(Refer Slide Time: 29:15)



Therefore, we can write that covariance between T and del del theta of log L theta is less than equal to variance of T into variance of del del theta of log L square. Now, covariance between T and del del theta log L theta can be written as expected value of T times del del theta log L theta minus expected value of T into expected value of del del theta log L theta. This as we have seen is equal to 0, therefore this quantity is 0.

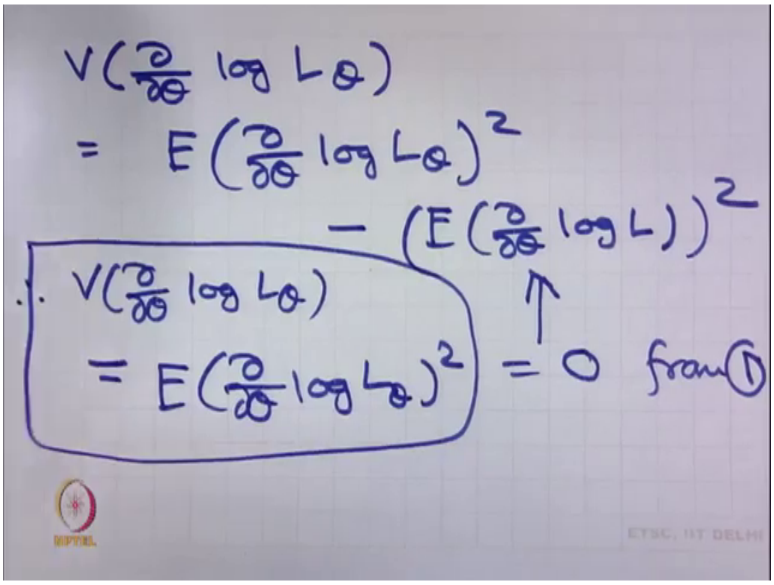
(Refer Slide Time: 30:47)



Therefore, we are left with covariance between T and del del theta log L theta is equal to expected value of T del del theta log L theta, and which we have found that expected value of T del del theta log log L theta is equal to g prime theta. Therefore, this is is equal to g prime theta.

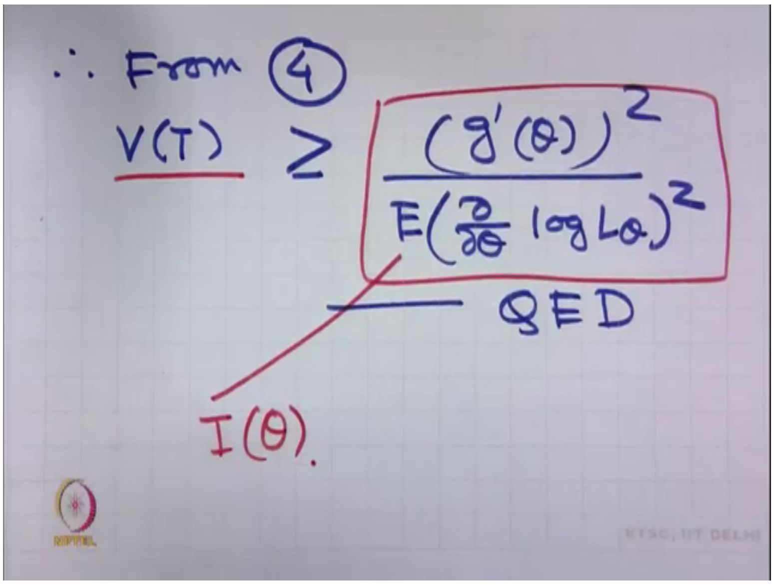
Now, let us look at this, let me call it 3. Therefore, by putting the value g prime theta here, we get g prime theta square is less than equal to variance of T into variance of del del theta log L theta or variance of T is less than equal to g prime theta whole square upon variance of del del theta log L theta. Let us call it 4 sorry I made a mistake there variance of T has to be greater than equal to this quantity.

(Refer Slide Time: 33:10)



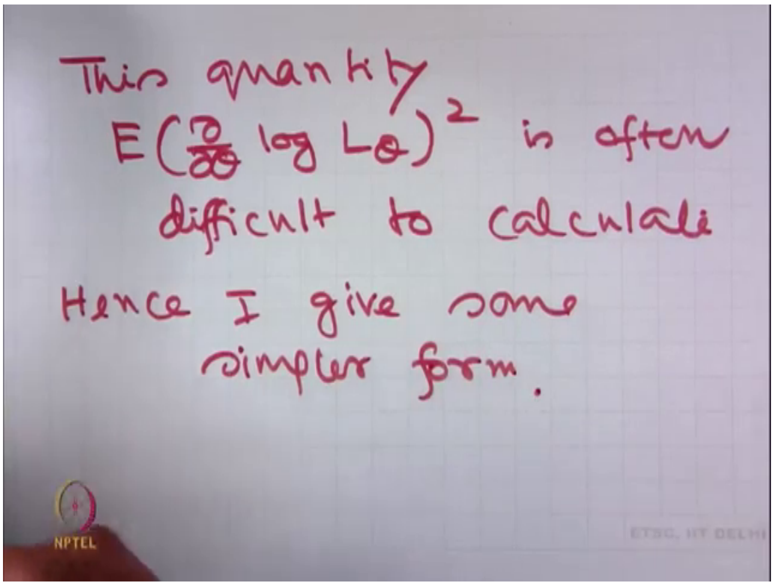
Now, in 4 let us look at what is variance of del del theta log L theta is equal to expected value of del del theta log L theta whole square minus expected value of del del theta log L square. And this is equal to 0 from 1. Therefore, variance of del del theta log L theta is equal to expected value of del del theta log L theta whole square, putting this in 4.

(Refer Slide Time: 34:11)



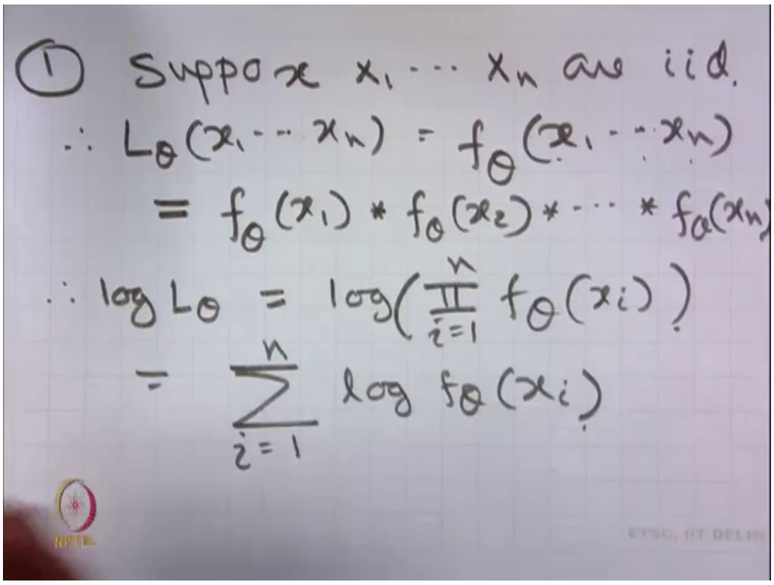
From 4, which is variance of T is greater than equal to g prime theta whole square upon variance of del del theta log L theta, we can write it as variance of T is greater than equal to g prime theta whole square upon expected value of del del theta log of L theta whole square. So, this is the result, that we were trying to prove that variance of T has to be greater than equal to this, where this is called the I theta.

(Refer Slide Time: 35:14).



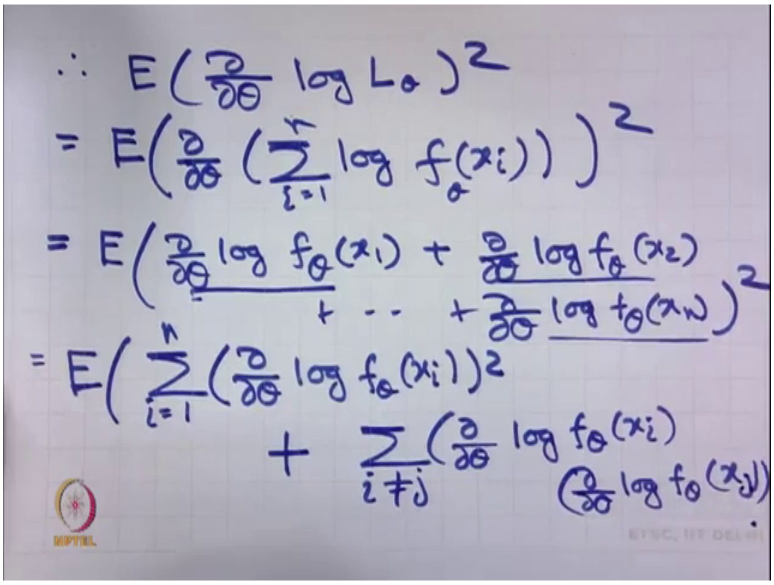
Now, this quantity expected value of del del theta log L theta whole square is often difficult to calculate. Hence, I give some simpler form.

(Refer Slide Time: 35:57)



Suppose x 1, x 2, x n are i i d that means, they are independent and identically distributed. Therefore, what is L theta of x 1, x 2, x n this is is equal to f theta of x 1, x 2, x n, which is the joint pdf of the sample values. Since, these are independent, we can write this as f theta of x 1 into f theta of x 2 into f theta of x n. Therefore, log of L theta is equal to log of the product. And since log of product is equal to some of the logs, this is same as sigma, i is equal to 1 to n log of f theta x i.

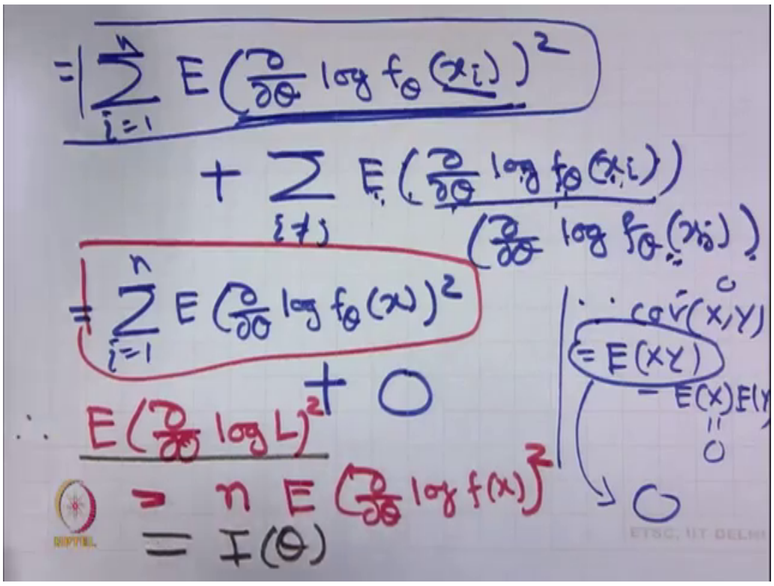
(Refer Slide Time: 37:32)



Therefore, expected value of del del theta log L theta whole square is equal to expected value of del del theta of sigma log of f x i, i is equal to 1 to n whole square is equal to expected value of del del theta of log of f theta x 1 plus del del theta of log of f theta of x 2 plus del del theta of log of f theta of x n whole square. So, now you will understand the difficulty, it becomes the sum of n terms whole square.

So, this we are writing as expected value of sigma i is equal to 1 to n del del theta of log of f theta of xi whole square, because I am collecting the individual square terms plus sigma over i not equal to j del del theta of log of f theta x i into del del theta of log of f theta x j. And I am looking at expectation of that one, bringing the expectation because of the linearity of expectation.

(Refer Slide Time: 39:44).



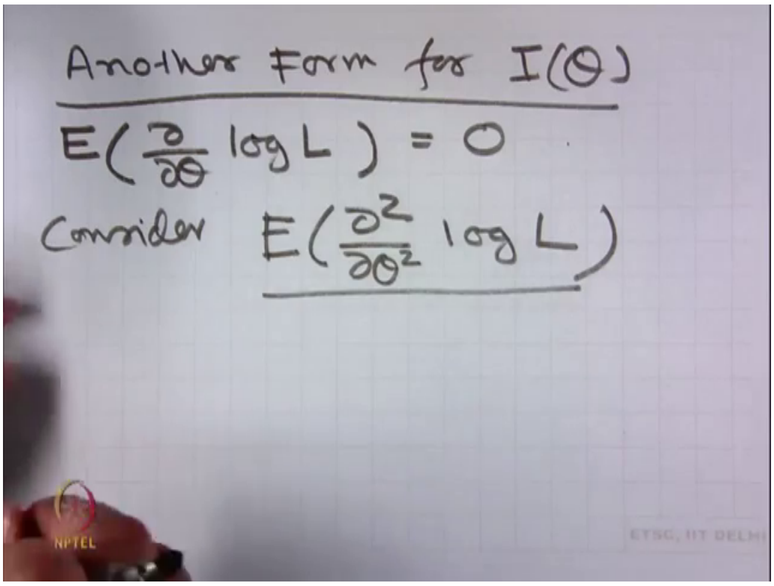
This is is equal to sigma i is equal to 1 to n expected value of del del theta of log of f theta x i whole square plus sigma i not equal to j expected value of del del theta log of f theta of x i into del del theta of log of f theta of x j, i not equal to j. Now, note that all x i's are identically distributed. Therefore, this term is same for all i. Therefore, this particular term boils down to sigma i is equal to 1 to n expected value of del del theta of log of f theta of x square plus let us consider this part it is the expected value of two terms, the product of two terms, one is based on x i, other is based on x j. And x i and x j are independent.

Therefore, the covariance between them is has to be 0, and also we have seen that expected value of this is 0, because even for that greater thing the L we have found that expected value of del del theta log L is equal to 0 by the same technique. We can find out that expectation of del del theta log of f theta x i is equal to 0. Therefore, what we find that this is the product of two terms, which are independent.

Therefore, their covariance is ~~θ,~~0; moreover their individual expectation is 0. And therefore, since covariance of X, Y is equal to expected value of X Y minus expected value of X into expected value of Y. In this case, we find this is 0, this is 0, therefore expected value of X Y is also going to be 0 or in other words expected value of del del theta log f of X Y multiplied by del del theta log f of x j. This is going to be 0, for each pair i j, i not equal to j. Therefore, I make it 0. I hope that, you understood the logic.

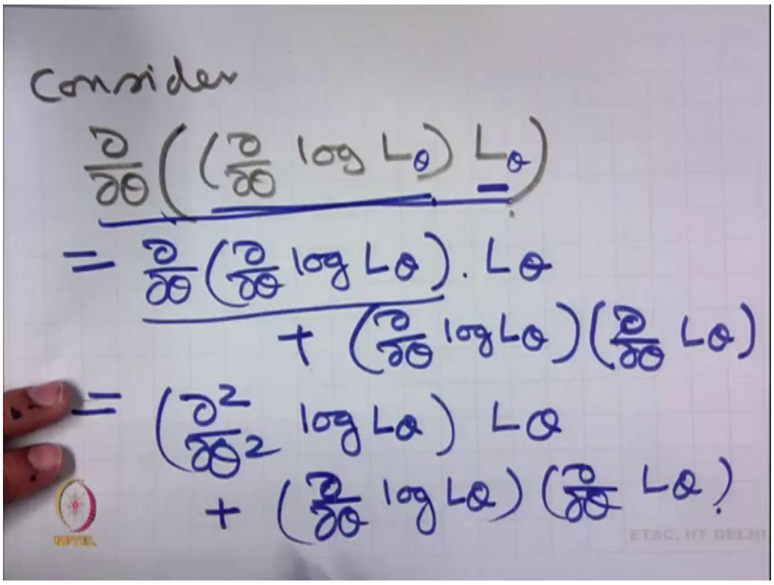
Therefore, what we find that expected value of del del theta of log L whole square is actually this term, which is nothing but n times expected value of del del theta log of f x square. So, this is another interesting form of I theta, in solving problems instead of using this we may often use. This as the denominator of the right hand side of the lower bound or right hand side of the Cramer- Rao inequality.

(Refer Slide Time: 44:00).



Now, let me discuss another form for I theta. We know that the expected value of del del theta log L is equal to 0, we have already seen that. Now, consider expected value of let us try to compute this. What is this?

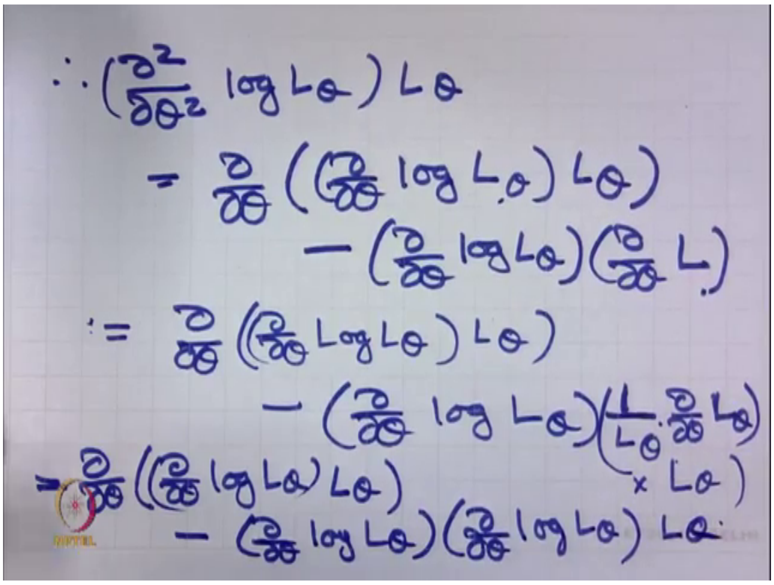
(Refer Slide Time: 44:58).



Let us consider, the product del del theta log L and L. And let us consider, it is partial derivative with respect to theta, we know that these are also dependent on theta. By by rule of multiplication for derivatives, we can write it as del del theta of del del theta log L theta times L theta plus del del theta log L theta into all right.

This is very straightforward, because it is first function derivative of 1st function into 2nd function plus 1st function into derivative of 2nd function. This we can write as del 2 del square del theta square of this we can write as del square del theta square of log L theta times L theta plus del del theta of log L theta times del del theta of L theta.

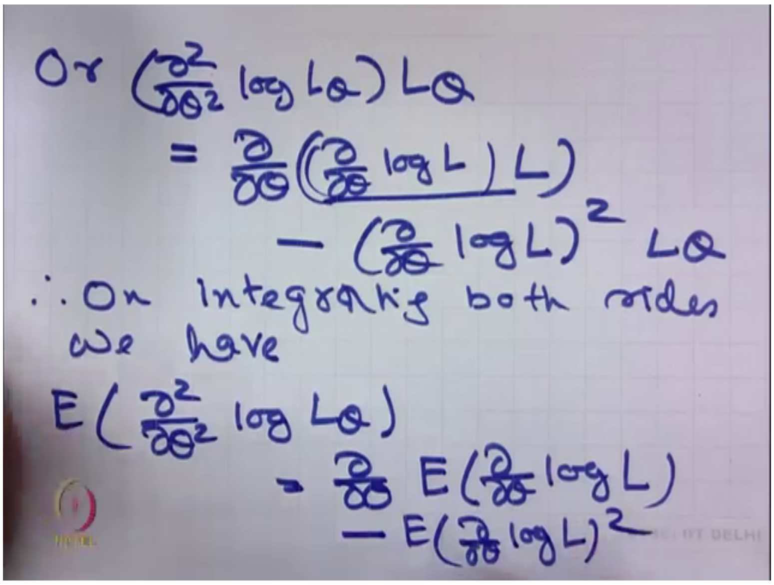
(Refer Slide Time: 47:13)



Therefore, del squared del theta square log of L theta into L theta is equal to del del theta of del del theta of log L into L theta minus del del theta of log L theta into del del theta of L. We started with this, and we have obtained this.

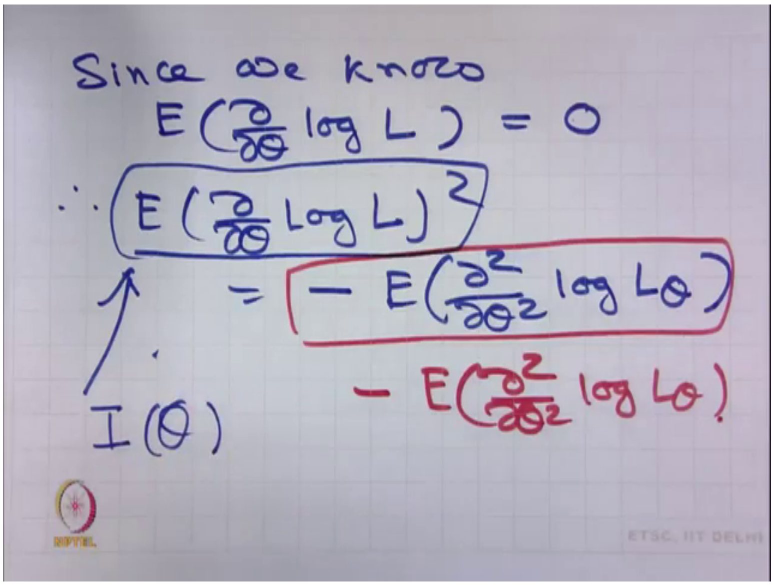
And therefore, what we are getting is del square del theta square log L theta L theta is equal to del del theta of del del theta log L theta multiplied by L theta minus del del theta log L theta into del del theta of L is equal to del del theta of del del theta log L theta into L theta minus del del theta of log L theta into 1 by L theta into del del theta log of L into L theta is equal to del del theta of del del theta log L theta into L theta minus del del theta of log L theta. And this, we can write is that we have already seen that it is del del theta of log of L theta into L theta.

(Refer Slide Time: 49:40).



Or del square del theta square log of L theta into L theta, can be written as del del theta log L whole square into L theta. This is coming, because this is the same term, so it is square of del del theta log L theta square. Therefore, on integrating both sides, we have the expected value of del square del theta square log of L theta is equal to del del theta of expected value of del del theta of log L, I am integrating with respect to the pdf. Therefore, I get expected value of that one minus expected value of del del theta log L whole square.

(Refer Slide Time: 51:26)



Since, we know the expected value of del del theta of log L is equal to 0, this we have found, before this part is going to be 0. Therefore, we can see that the expected value of del del theta log L whole square is equal to minus of this thing. Expected value of del del theta log L square is equal to minus of expected value of del 2 del theta square log of L theta. And we know that this is the I theta.

Hence, we get another expression for I theta, which is this minus of expected value of del theta del theta square log of L theta. So, we have got three different forms. One from the actual proof, but from there we have derived two different forms, one is this one, and the other one is n into expected value of del del theta log f whole square.

So, in problem solving, we shall be using one of these forms. And we will be able to solve certain problems. So, in the next class, I will solve a few problems using Cramer-Rao inequality. And show that, how we obtain the lower bound in certain cases when we are dealing with the unbiased estimators for a function g theta of the parameter theta.

Ok students, thank you, seeing you in the next class.

Thanks.