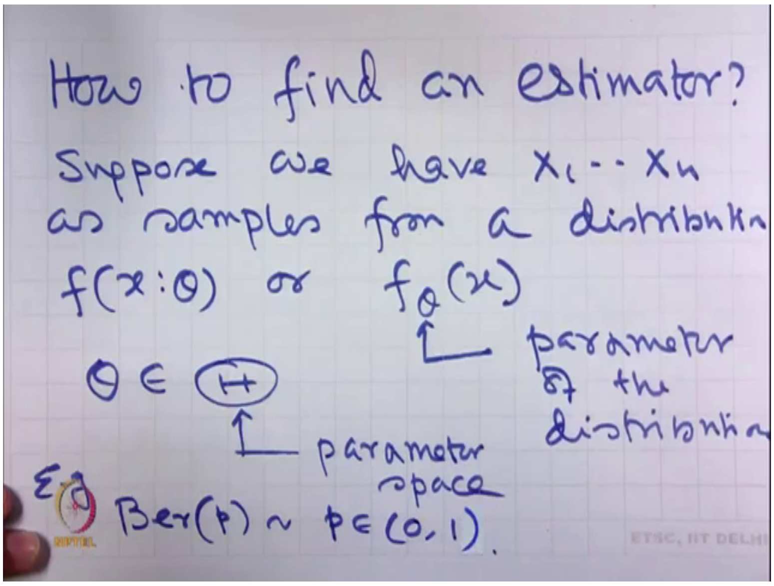
Statistical Inference

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

Lecture - 17
Statistical Inference

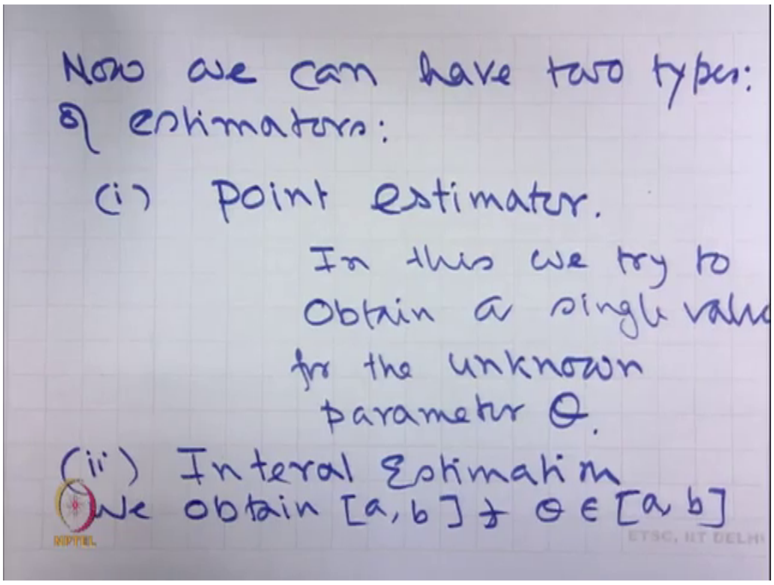
Welcome student to the MOOCs series of lecture on Statistical Inference. This is lecture number 17. Over the last few classes we have been discussing theory of estimation. In particular we have discussed desired properties of an estimator namely unbiasedness, consistency, efficiency, sufficiency which we expect in a good estimator to have.

(Refer Slide Time: 01:00)



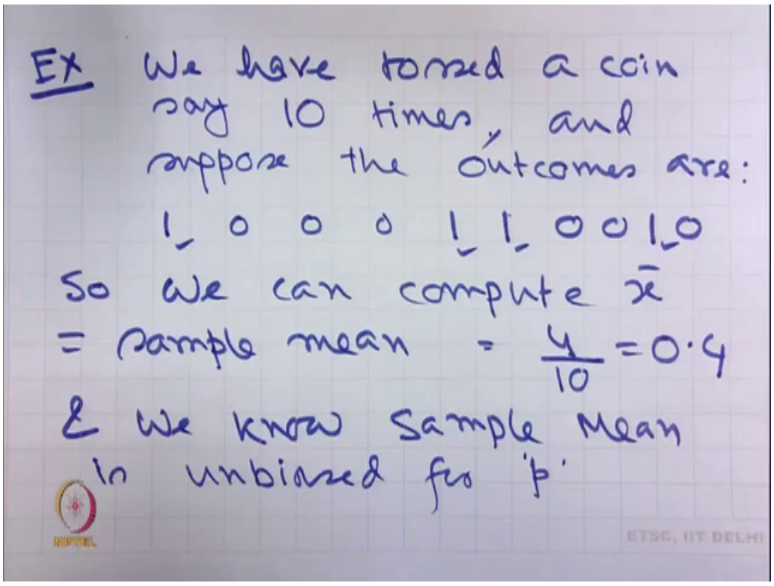
But the main question is how to find an estimator. Suppose we have X 1 X 2 X n as samples from a distribution. If x theta or I may write f theta of x. where theta is the parameter of the distribution. And we know that theta belongs to capital theta the parameters space. For example, Bernoulli p, p belongs to 0, 1.

(Refer Slide Time: 02:32)



Now we can have estimator of two types: one is point estimation or point estimator. In this case, we try to obtain a single value for the unknown parameter theta. And the second one is interval estimation where you try to obtain an interval such that theta belongs to this interval with certain degree of confidence.

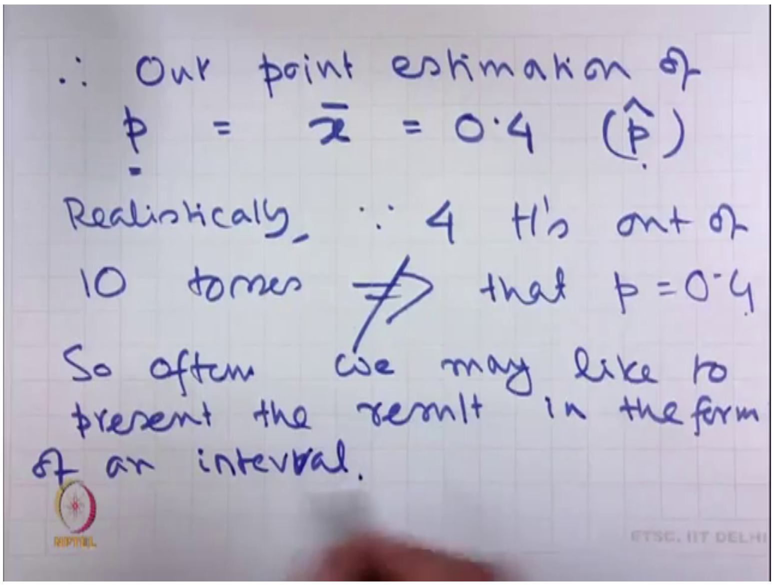
(Refer Slide Time: 04:16)



For example, we have tossed a coin say 10 times. And suppose the outcomes are head tail tail tail head head tail tail head and tail.

So, we can compute x bar is equal to sample mean is equal to 4 divided by 10, 4 comes because there are 4 success is equal to 0.4. And we know that sample mean is an unbiased estimator. Sample mean is unbiased for p.

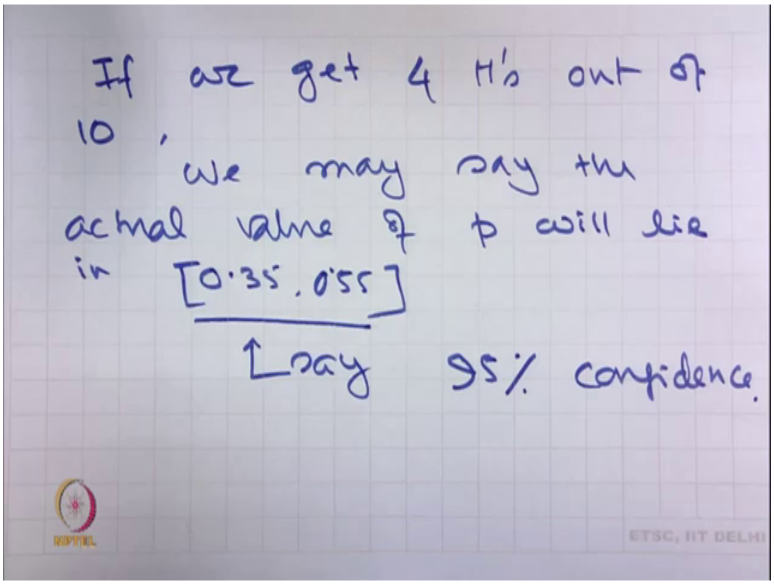
(Refer Slide Time: 06:02)



Therefore, our point estimation of p is equal to x bar is equal to 0.4. And we often denote it as p hat where p hat is an estimated value of the unknown parameter p, but if we think realistically just because we got 4 heads out of 10 tosses does not imply that P is equal to 0.4.

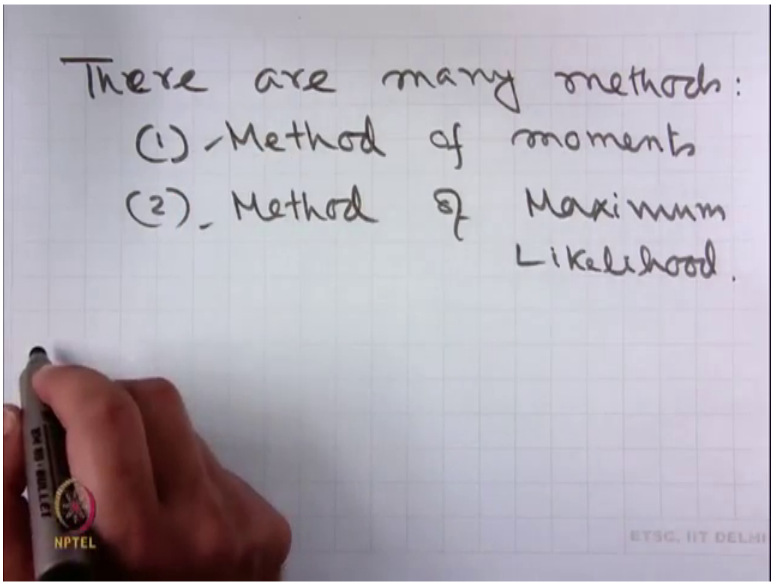
In another set of experiments, you may get with the same coin different values of p hat. So, often we may like to present the result in the form of an interval.

(Refer Slide Time: 07:44)



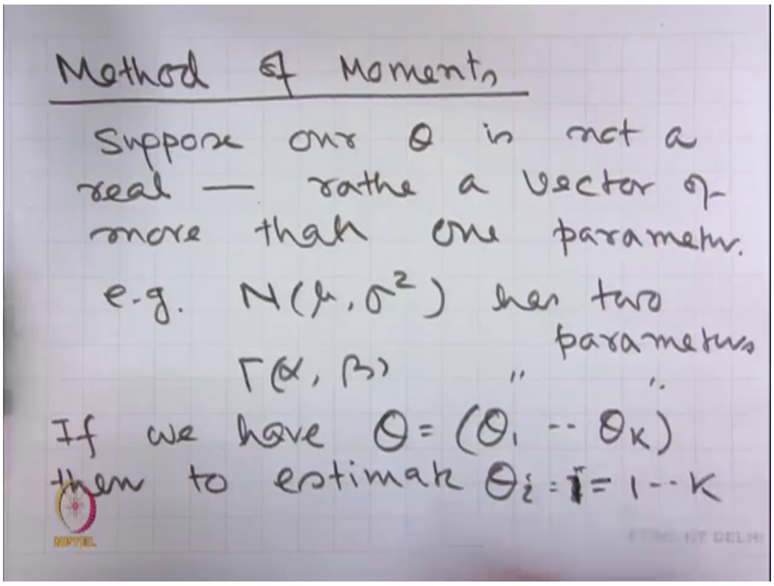
Say for example, if we get 4 heads out of 10. We may say the actual value of p will lie in say 0.35 to say 5. 0.55. So, we are giving an interval and we are saying that the value of the parameter with will lie this interval with certain degree of confidence say 95 percent confidence. When you present the result in this manner, we call it an interval estimation of the parameter. In the present talk I will be discussing point estimation primarily.

(Refer Slide Time: 09:05)



There are many methods, but the two most important to answer method of moments and method of maximum likelihood. So, in this talk, I will be discussing these two methods in detail.

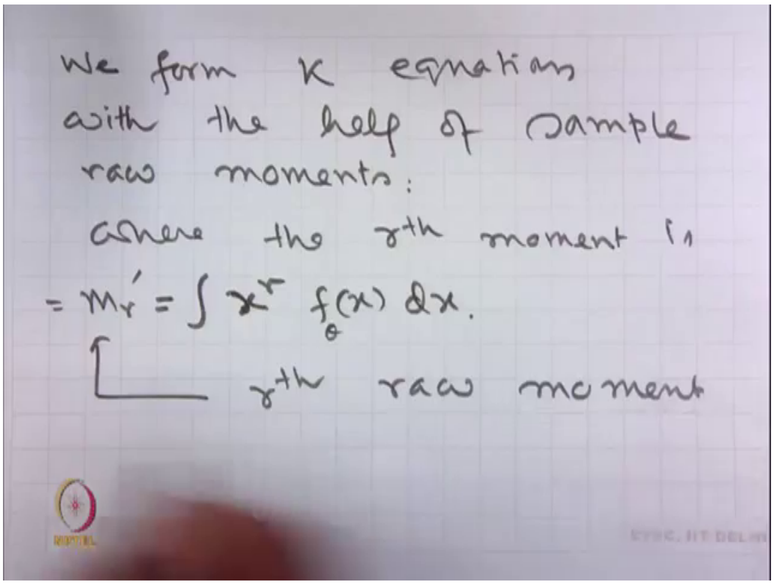
(Refer Slide Time: 09:56)



So, method of moments suppose our theta is not a real rather a vector of more than one parameter. For example, normal mu sigma square has two parameters. Similarly gamma alpha beta has two parameters.

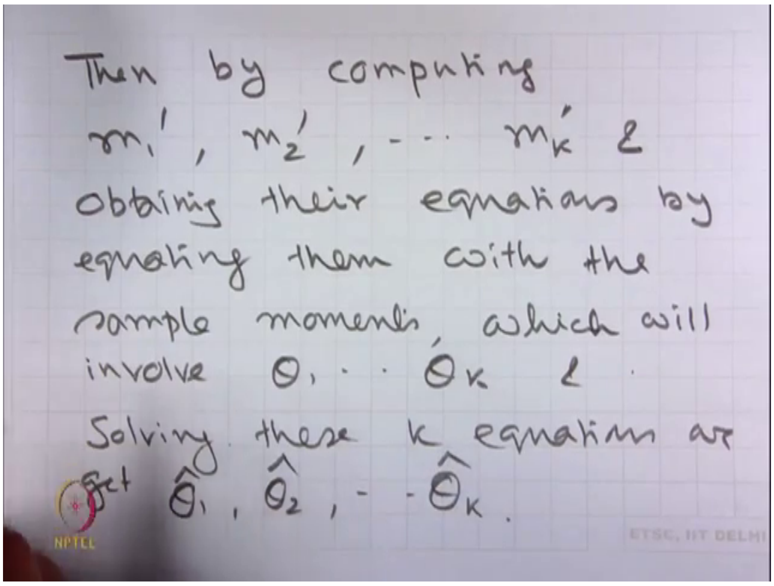
So, if we have theta is equal to a K dimensional vector theta1 theta 2 up to theta K, then to estimate theta versus theta i is equal to i is equal to 1 to K.

(Refer Slide Time: 11:40).



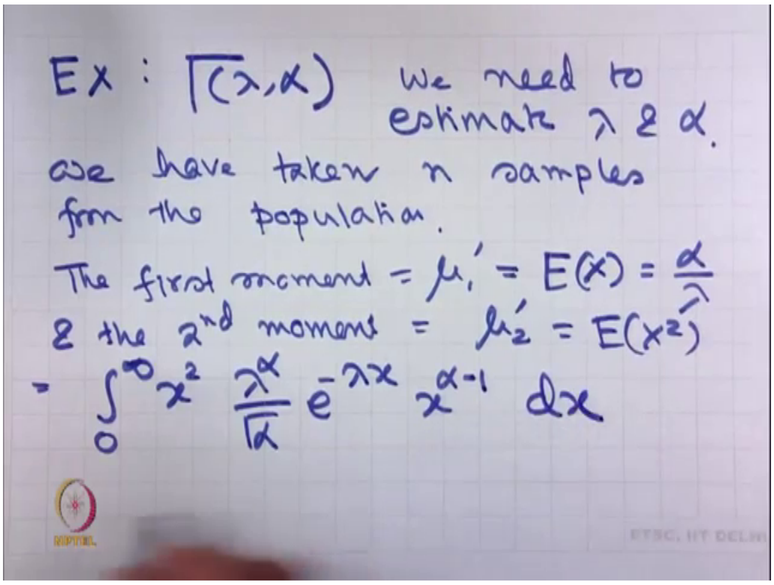
We form K equations with the help of sample raw moments where the rth moment is sigma x to the power r fx dx. This is the rth raw movement.

(Refer Slide Time: 12:50)



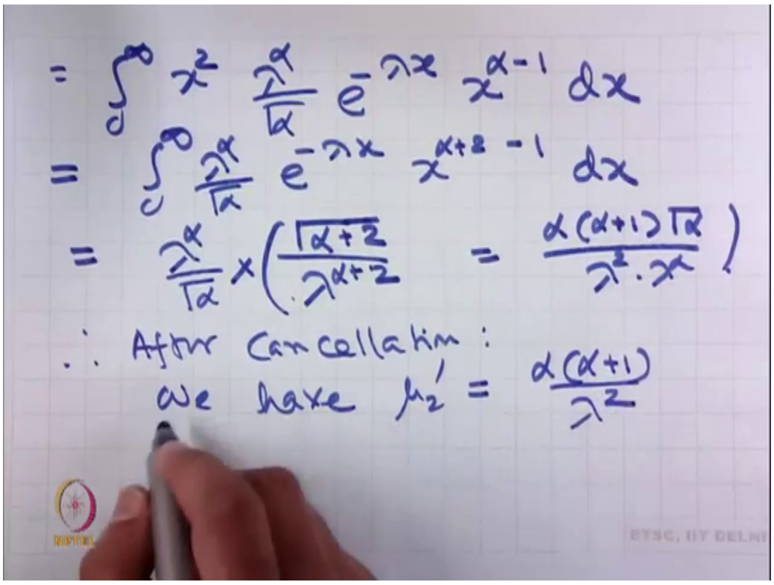
Then by computing m1 prime, m2 prime up to m k prime and obtaining thier equations by equating them with the sample moments, which will involve theta 1, theta 2 up to theta K. And solving this k equations, we get theta 1 hat, theta 2 hat theta K hat.

(Refer Slide Time: 14:22)



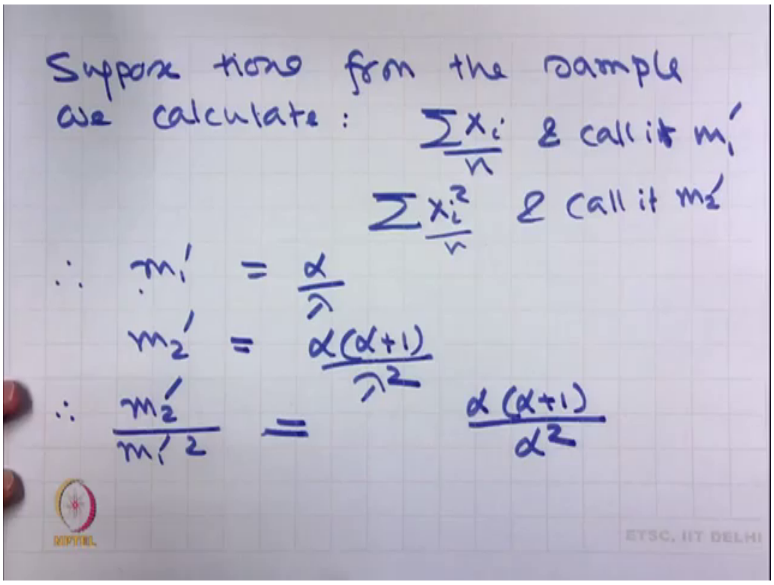
So, let me give you an example: gamma lambda alpha. We need to estimate lambda and alpha. So, what we have done? We have taken say n samples from the population. Then what is the first movement? The first movement is equal to mu 1 prime and we know that for gamma it is expected value of X is equal to alpha upon lambda. And the second movement is equal to mu 2 prime is equal to expected value of X square is equal to 0 to infinity x square lambda power alpha upon gamma alpha e to the power minus lambda x x to the power alpha minus 1 dx is equal to 0 to infinity x square lambda power alpha upon gamma alpha e to the power minus lambda x x to the power alpha minus 1 dx which is is equal to 0 to infinity lambda power alpha upon gamma alpha e to the power minus lambda x x to the power alpha plus 2 minus 1 dx.

(Refer Slide Time: 16:13)



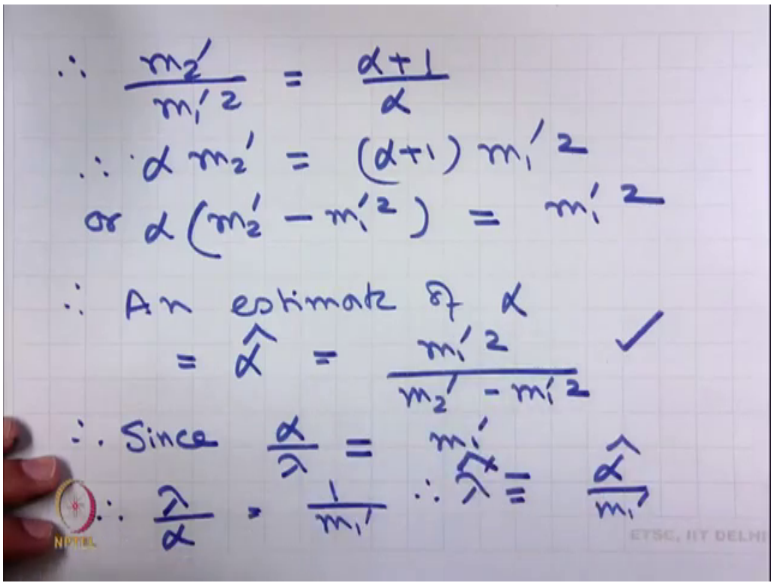
And we know that this is going to be lambda power alpha upon gamma alpha into gamma alpha upon lambda power alpha plus gamma alpha plus 2 upon lambda power alpha plus 2 which is is equal to alpha into alpha plus 1 gamma alpha this part is equal to alpha into alpha plus 1 times gamma alpha into lambda square into lambda power alpha. Therefore, after cancellation we have mu 2 prime is equal to alpha into alpha plus 1 upon lambda square.

(Refer Slide Time: 18:00)



Suppose now, from the sample we calculate sigma Xi by n where n is the sample size and call it m1 prime and sigma Xi square by n and call it m2 prime. Then we can write m 1 prime is equal to alpha over lambda and m2 prime is equal to alpha into alpha plus 1 upon lambda square.Therefore, m2 prime upon m1 prime square is equal to alpha into alpha plus 1 upon alpha square as this lambda square will get cancelled.

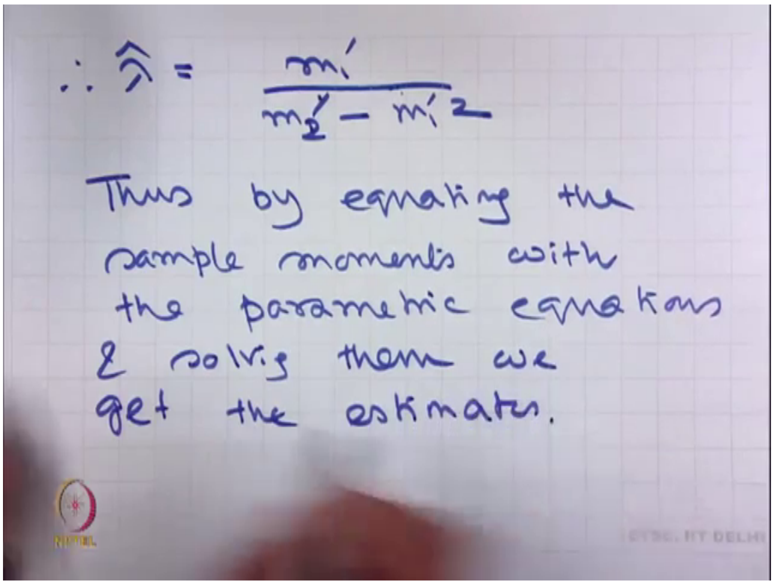
(Refer Slide Time: 19:34)



Therefore, m 2 prime upon m1 prime square is equal to alpha plus 1 upon alpha. Therefore, alpha m 2 prime is equal to alpha plus 1 m1 prime square or alpha times m 2 prime minus 1 prime square is equal to m 1 prime square. Therefore, an estimate of alpha is equal to alpha hat is equal to m 1 prime square upon m 2 prime minus m 1 prime square. Therefore, since we know that alpha upon lambda is equal to m1 prime; therefore, lambda upon alpha is equal to 1 upon m1 prime.

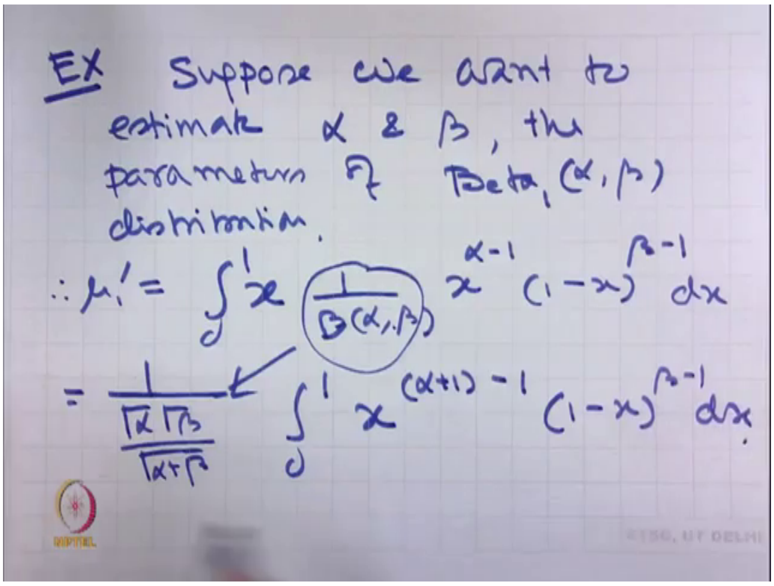
Therefore lambda hat is equal to alpha hat upon m1 prime where alpha alpha hat we are found from here and that cancel swan m1 prime.

(Refer Slide Time: 21:21)



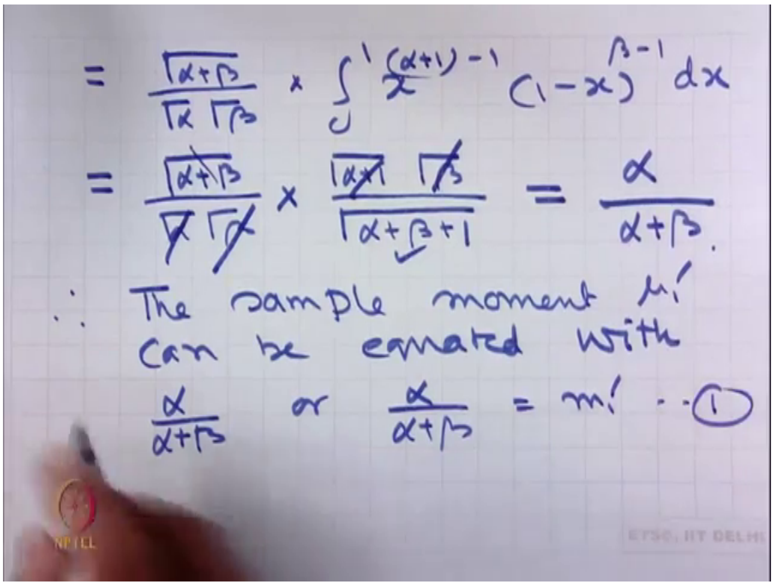
Therefore lambda hat is equal to m1 prime upon m1 m2 prime minus m1 prime square thus by equating the sample moments with the parametric equations and solving them, we get the estimates consider another example.

(Refer Slide Time: 22:24)



Suppose we want to estimate alpha and beta the parameters of beta 1 alpha beta distribution. Again let us considered mu1 prime 0 to 1 x 1 upon beta alpha beta to the power alpha minus 1 1 minus x to the power beta minus 1 dx is equal to gamma alpha gamma beta upon gamma alpha plus beta. This I have taken out multiplied by integration 0 to 1 x to the power alpha plus 1 minus 1 into 1 minus x beta minus 1 dx.

(Refer Slide Time: 24:15)

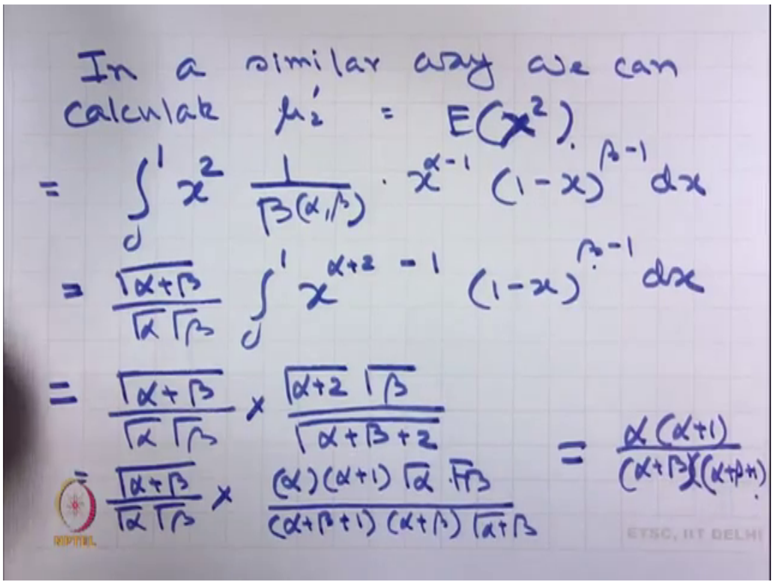


This is is equal to gamma alpha plus beta upon gamma alpha gamma beta into 0 to 1 x to the power alpha plus 1minus 1 1 minus x whole to the power beta minus 1 dx. And this integral gives us beta with alpha plus 1 and beta as there parameters. Therefore, this is is equal to gamma alpha plus beta upon gamma alpha gamma beta multiplied by gamma alpha plus 1 gamma beta gamma alpha plus beta plus 1.

This is is equal to this cancels and we know that gamma alpha plus is 0 equal to alpha times gamma alpha. So, that cancels this gamma alpha with this and we have left with alpha in a similar way this is is equal to alpha plus beta into gamma alpha plus beta. So, this gets cancelled and we are left with alpha plus beta. Therefore, the sample moment can be equated.

Sample first moment can be equated with alpha upon alpha plus beta or alpha upon alpha plus beta is equal to m1 prime. So, this is the first equation.

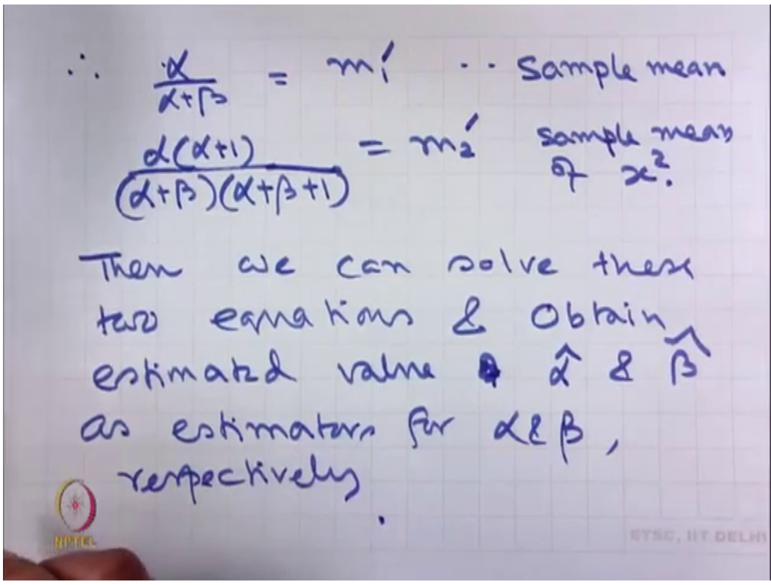
(Refer Slide Time: 26:21)



In a similar way, we can calculate mu 2 prime which is the expected value of X square. So, what it is this is is equal to integration 0 to 1 x square one upon beta alpha beta multiplied by x to the power alpha minus 1 1 minus x to the power beta minus 1 dx.

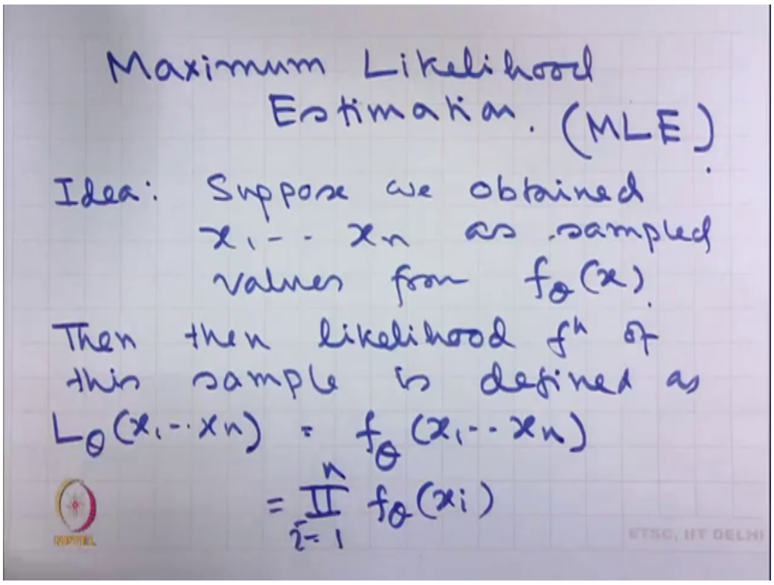
This is equal to as before gamma alpha plus beta upon gamma alpha gamma beta into 0 to 1 x to the power alpha plus 2 minus 1 1 minus x beta minus 1 dx. This is is equal to gamma alpha plus beta upon gamma alpha gamma beta multiplied by now this is giving us beta with alpha plus 2 and this beta. So, we have gamma alpha plus 2 gamma beta upon gamma alpha plus beta plus 2 is equal to gamma alpha plus beta gamma alpha gamma beta into now gamma alpha plus 2 is equal to alpha plus 1 into gamma alpha plus 1e which again is alpha into gamma alpha. So, I can write it as alpha into alpha plus 1 into gamma alpha, then gamma beta and this is alpha plus beta plus 1 into alpha plus beta into gamma alpha plus beta. So, after cancellation, we will have alpha into alpha plus 1 upon alpha plus beta into alpha plus beta plus 1.

(Refer Slide Time: 29:10)



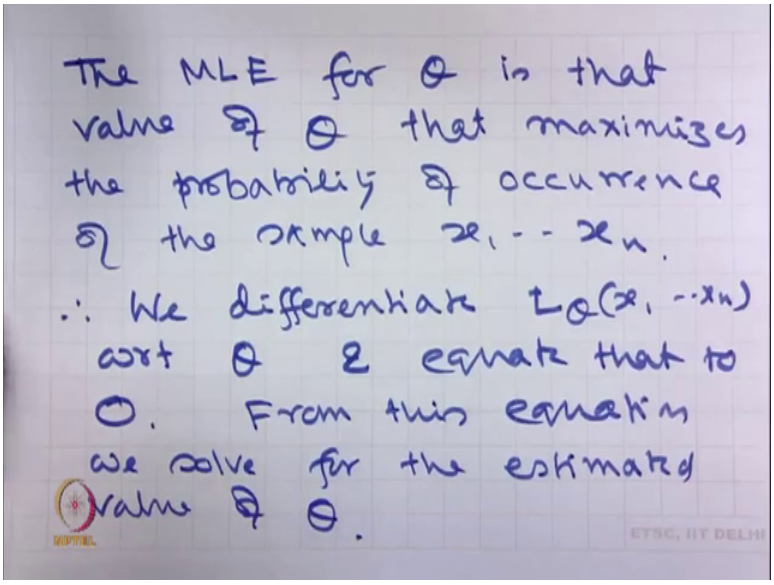
Therefore we have alpha upon alpha plus beta is equal to m1 prime which is the sample mean and alpha, alpha plus 1 upon alpha plus beta into alpha plus beta plus 1 is equal to m2 prime which is sample mean of x square. Then we can solve these two equations and obtain estimated value of alpha hat and beta hat as estimators for alpha and beta respectively. Similar equations can be formed with other distributions and one can obtain the estimates thorough the method of moments in the above way.

(Refer Slide Time: 31:08)



Now, let me discuss maximum likelihood estimators or estimation. In short, we will write it as MLE. The idea is as follows suppose we obtained x 1 x 2 x n as our sample values from f theta x, then the likelihood function of the sample is defined as l theta of x 1 x 2 x n is equal to the joint density of x 1 x 2 x n. And if the samples are independent it is product of f theta of xi i is equal to 1 to n. We have already seen the likelihood function when you are discussing gamma inequality. So, this is not something that is new to us.

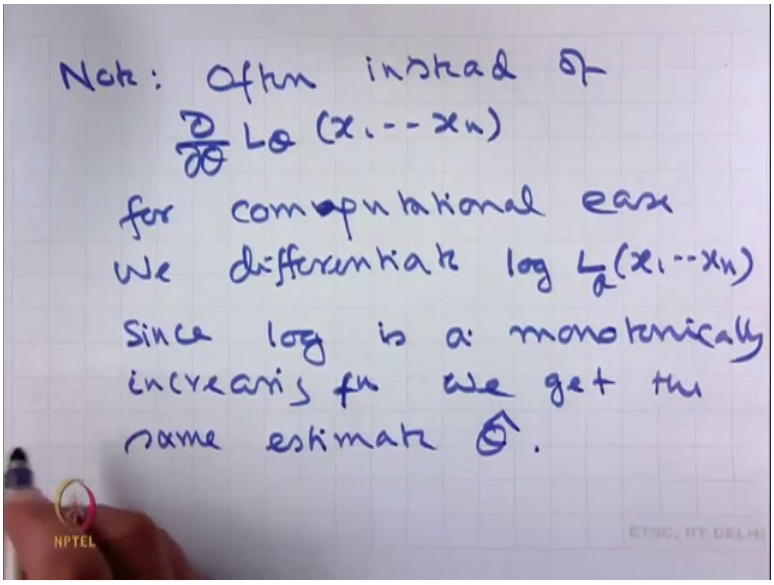
(Refer Slide Time: 33:12).



The idea of likelihood estimation is that we try to obtain that particular value of theta which maximizes the probability of occurrence of the sample x 1 x 2 x n. So, the MLE for theta is that value of theta that maximizes the probability of occurrence of the sample x1 x 2 x n.

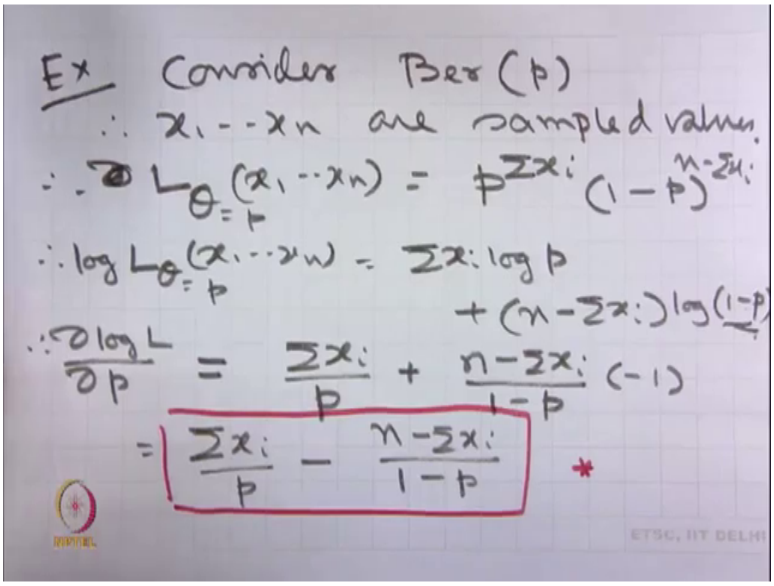
So, how to obtain that? Therefore, we differentiate L theta of x 1 x 2 x n with respect to theta and equate that to 0. From this equation we solve for the estimated value of theta. Of course when we are solving the equation delta L delta theta is equal to 0. We also will have to see that the second order derivative is negative, then only we can ensure that this solution theta hat is actually giving us the maximum probability for the obtained sample x1 x 2 x n.

(Refer Slide Time: 35:50)



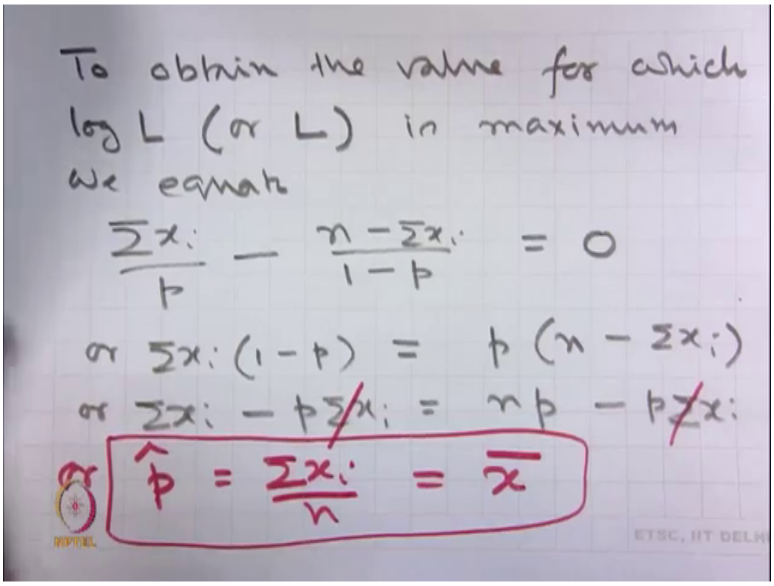
Note often instead of del delta theta of L theta of x 1, x 2 x n. For computational ease, we differentiate log of L theta of x 1, x 2, x n and since log is an increasing function the result is not altered.

(Refer Slide Time: 37:12)



So, let me give you an example. Consider Bernoulli P therefore, x 1, x 2, x n are sampled values. Therefore, log of x 1, x 2, x n is equal to we have already seen this is P to the power sigma x i into 1 minus P whole to the power n minus sigma xi. Therefore log L therefore, log L x 1 x 2 x n is equal to sigma xi log p plus n minus sigma xi log of 1 minus p. Therefore, del log L del p because here the parameter is p. We could write theta is equal to p or theta is equal to p. Therefore, del log L del p is equal to sigma xi upon p plus n minus sigma xi upon log of 1 minus p is equal to 1 upon 1 minus p multiplied by d 1 minus p dp which will give you minus 1 is equal to sigma x i upon p minus n minus sigma x i upon 1 minus p. So, this is star I will need it later when I will be considering the second derivative.

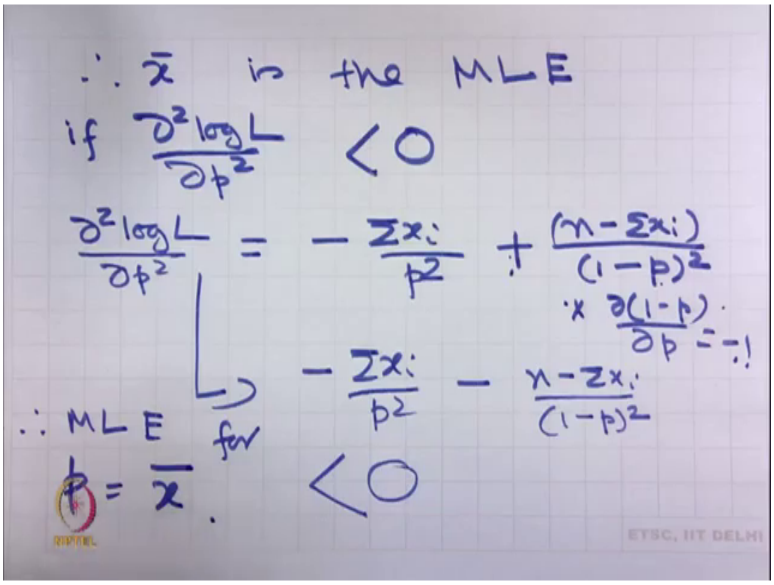
(Refer Slide Time: 39:45)



But to obtain the value for which log L is maximum; that is L is maximum. We equate sigma xi upon p minus n minus sigma x i upon 1 minus p is equal to 0 or sigma x i into 1 minus p is equal to p times n minus sigma x i or sigma x i minus p times sigma x i is equal to n p minus p sigma x i.

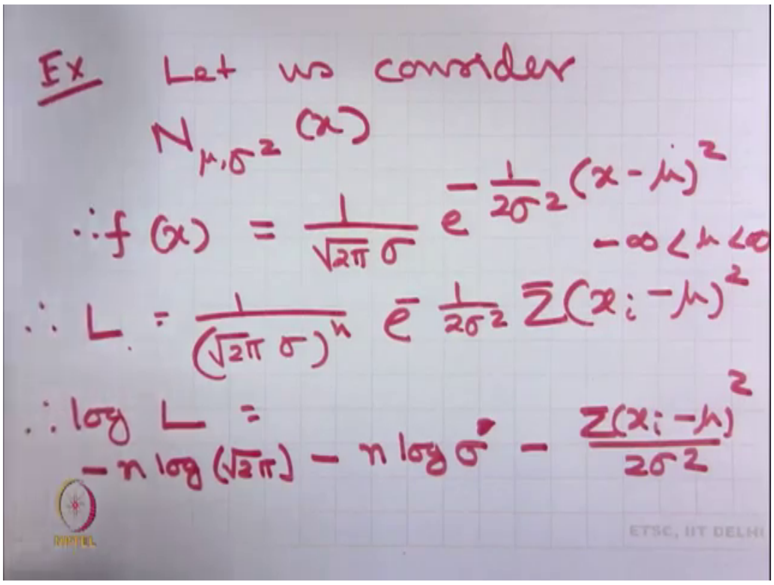
So, this cancels or p hat is equal to sigma x i upon n is equal to sample mean. Therefore, from here also we find that if we consider p to be p hat to be the sample mean then we get delta log L delta p is equal to 0. Therefore, x bar can be a possible solution provided.

(Refer Slide Time: 42:03)



If delta 2 log L delta p square because theta is equal to p is less than 0.We already had delta log L delta p is equal to this. So, now, we are differentiating this with respect to p minus sigma xi upon p square minus n minus sigma x i upon 1 minus p whole square. This will be multiplied by minus 1. So, that will make it plus 1 because 1 minus p whole to the power minus 1 was there this multiplied by d 1 minus p d p which like in earlier case is going to be minus 1. Therefore, delta 2 log L delta p square is equal to minus sigma x i p square minus now this minus again makes it minus n minus sigma x i upon 1 minus p whole square. Since sigma xi can maximum value be n therefore, n minus sigma x i is positive with this negative sign; this becomes negative, this becomes negative, this is positive, this is positive. Therefore, the whole thing is less than 0. Therefore, MLE for p is equal to x bar.

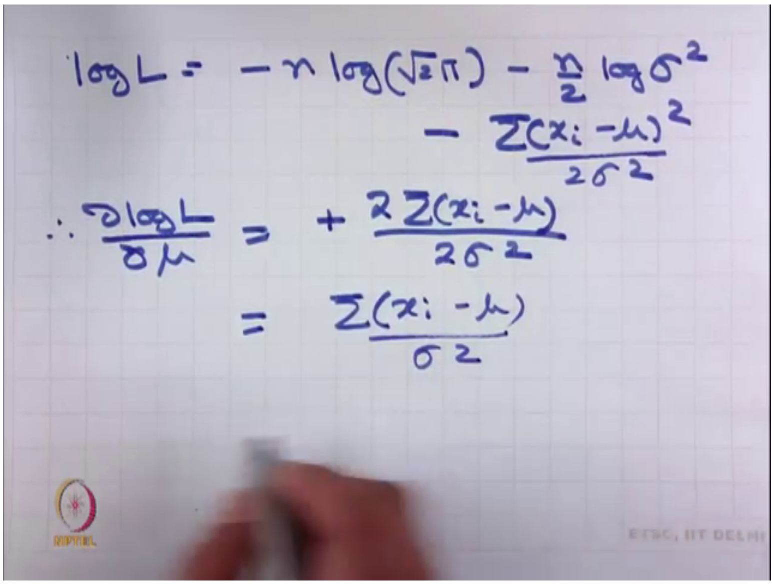
(Refer Slide Time: 44:40)



Let us now consider normal mu sigma square of x. Therefore,

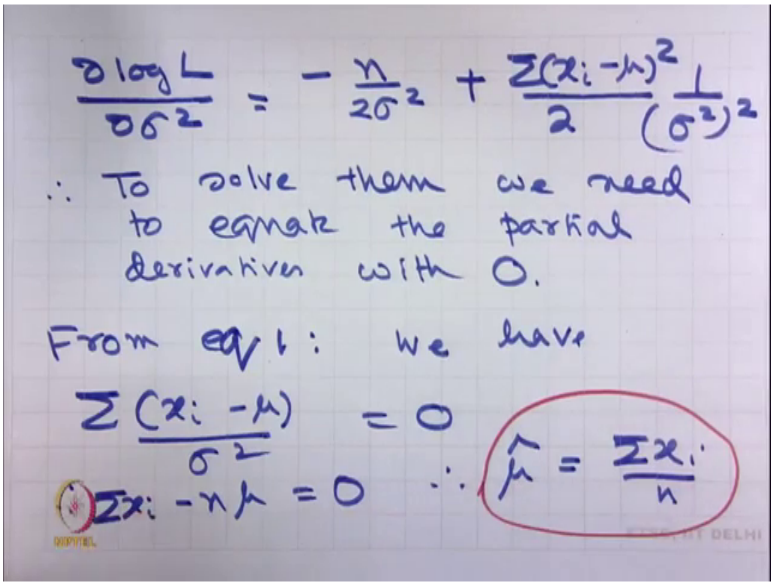
f x is equal to 1 over root over 2 pi sigma e to the power minus 2 sigma square into x minus mu whole square for minus infinity less than mu less than sigma less than infinity. Therefore, L of course, mu sigma square which you are often write as well theta is equal to 1 over root over 2 pi sigma whole to the power n e to the power minus 1 upon 2 sigma square sigma x i minus mu whole square. Therefore log L is equal to minus n log root over 2 pi. This is from here minus n log sigma minus sigma x i minus mu whole square divided by 2 sigma square.

(Refer Slide Time: 47:03)



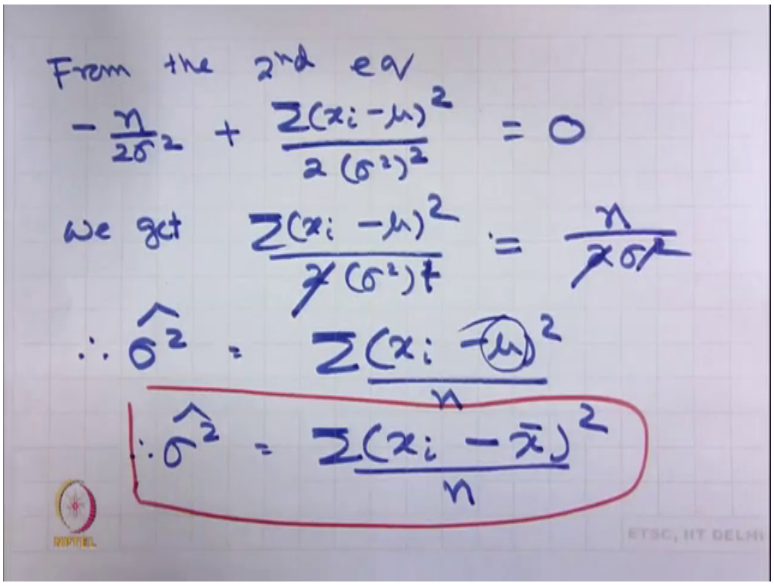
Or if we write it with sigma square as the parameter, if we write sigma square as the parameter, then we can write log L is equal to minus n log root over 2 pi minus n by 2 log sigma square minus sigma x i minus mu whole square upon 2 sigma square. Therefore, del log L del mu is equal to this gets cancelled. Because this gives 0, this gives 0 and what we are left with is minus 2 into sigma x i minus mu. Now this is with minus sign. So, that makes it plus upon 2 sigma square is equal to sigma x i minus mu upon sigma square and del log L.

(Refer Slide Time: 48:48)



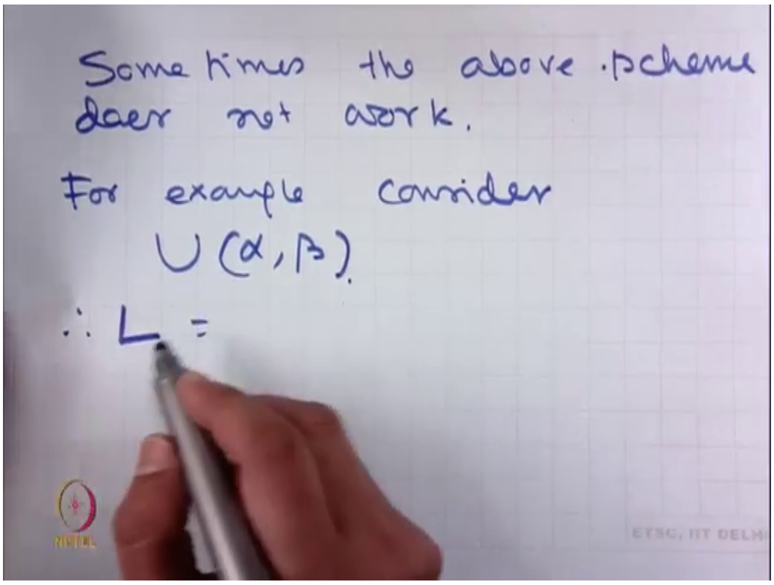
Del sigma square is equal tofrom here we are differentiating it with respect to sigma square. Therefore, log of sigma square derivative is 1 upon sigma square. So, what we are getting is minus n upon 2 sigma square. And from the other one, we are getting plus sigma x i minus mu whole square upon 2 into 1 upon sigma square whole square. Because if you look at this,it is 2 sigma square. So, sigma square to the power minus 1. So, when we are differentiating we are getting that minus will make it plus and 1 upon sigma square whole square. Therefore, to solve them we need to equate the partial derivatives with 0 from equation 1. We have minus x i minus mu upon sigma square is equal to 0. Therefore, sigma x i minus n mu is equal to 0 therefore, mu hat is equal to sigma xi upon n. So, that is the estimate for mu.

(Refer Slide Time: 52:08)



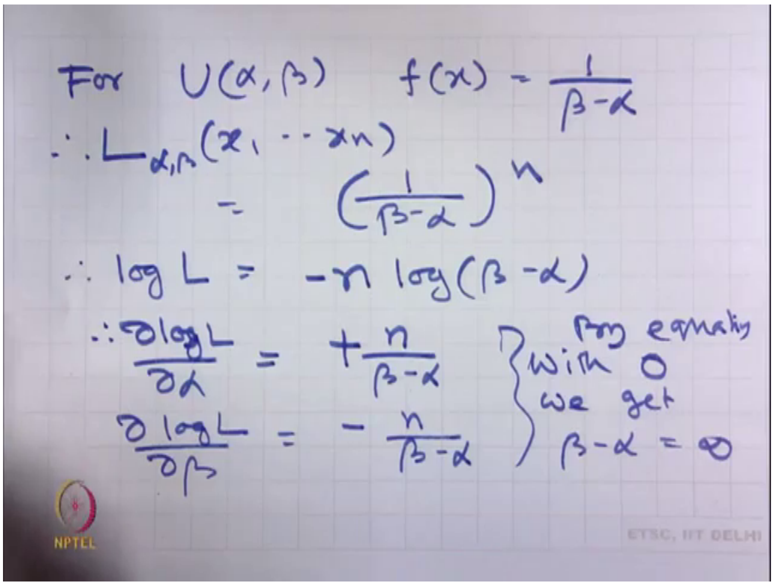
From the second equation which is minus n upon 2 sigma square plus sigma x i minus mu whole square upon 2 sigma square square is equal to 0. We get sigma x i minus mu whole square upon 2 sigma square square is equal to n upon 2 sigma square. So, this cancels one of these cancels. Therefore sigma square hat is equal to sigma x i minus mu whole square upon n. But we do not know mu, we know only mu hat. Therefore, sigma square hat is equal to sigma x i minus x bar whole square up on n. So, that is the maximum likelihood estimate for sigma square. And we already knew that this is not an unbiased estimator because the unbiased estimator for sigma square is sigma x i minus x bar whole square upon n minus 1.When the mu is unknown to us however, MLE gives us these to be the estimator. [FL]

(Refer Slide Time: 54:36)



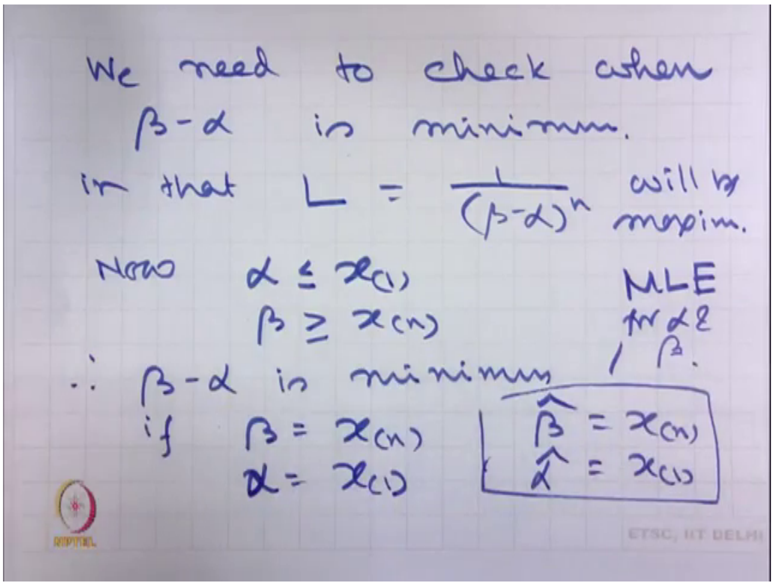
Sometimes, the above trick or above scheme does not work. For example, consider uniform alpha beta.

(Refer Slide Time: 55:20)



Therefore, log is equal to for uniform alpha beta f x is equal to 1 upon beta minus alpha. Therefore, the likelihood function of x 1, x 2, x n is equal to 1 upon beta minus alpha whole to the power n and this does not give us any solution. Therefore, log L is equal to n log beta minus alpha with a minus sign. Therefore, del log L del alpha is equal to minus n upon beta minus alpha and now make it plus and del log L del beta is equal to minus n upon beta minus alpha. By equating 0 by equating with 0, we get beta minus alpha is equal to infinity. Therefore that does not give us a solution. What we can do in the following way?

(Refer Slide Time: 57:08)



We need to check when beta minus alpha is minimum in that case L is equal to 1 upon beta minus alpha whole to the power n will be maximum. Now alpha has to be less than equal to x 1 and beta has to be greater than equal to x n. This is the first order statistics, this is the nth order statistic. Therefore, beta minus alpha is minimum.

If beta is equal to x n and alpha is equal to x 1, therefore x n is the beta hat and alpha hat is equal to x n. These are the MLE for alpha and beta with that I stop here today. In the next lecture, I will give you some properties of maximum likelihood estimator and also I will talk about interval estimation.

Thank you.