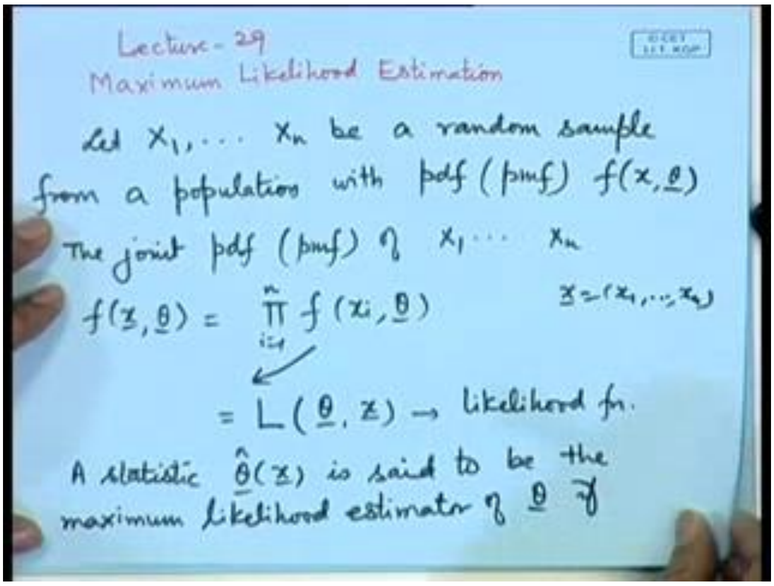
Probability and Statistics
Prof. Somesh Kumar
Department of Mathematics
Indian Institute of Technology, Kharagpur

Lecture - 57
Examples on MLE – I

In the last lecture I introduced method of maximum likelihood estimation, besides that we had also discussed the method of least square estimation and the method of moments. So, now, I continue the discussion on the method of maximum likelihood estimation which is actually the last among the three mention methods that I have given. So, let me repeat the definition of the maximum likelihood definition that how it is obtained, so firstly we define likelihood function.

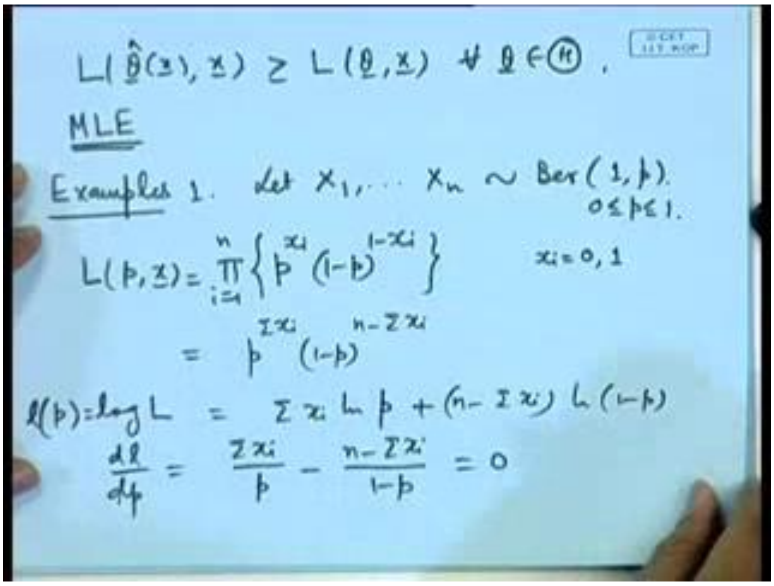
(Refer Slide Time: 00:53)



So, if I have X 1, X 2, X n let X 1, X 2, X n be a random sample from a population with either pdf or pmf say f x theta. So, what we write the joint pdf or pmf of X 1, X 2 X n. So, we write it as say f x theta equal to product of f x i theta I is equal to 1 to n; we call this now see when X 1, X 2, X n is observed to be some value small x 1, small x 2 small x n then we call this as the likelihood function of theta and we may put x here just to denote its dependence on x also this is called the likelihood function.

So, what we consider a statistic theta hat x is said to be the maximum likelihood estimator of theta if L theta hat X is greater than or equal to L theta X for all theta belonging to the parametric space.

(Refer Slide Time: 02:55)

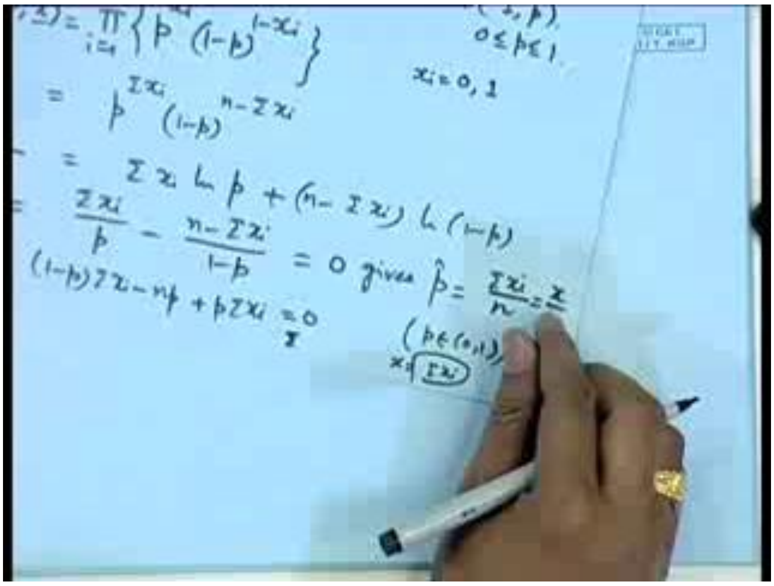


So, in short we use the terminology MLE for maximum likelihood estimator; a popular technique to obtain the optimizing value of theta is to take logarithm of L and then consider the derivative of L with respect to the parametric functions and then equate to 0 and solve those equations.

So, let me explain the method through some examples let X 1, X 2, X n follows a Bernoulli distribution. So, here p is the unknown parameter, we want to find out the maximum likelihood estimator for p. So, we write down the likelihood function here that will be equal to p to the power x i, 1 minus p to the power 1 minus x i product i is equal to 1 to n. If you look at the inside quantity this is the probability mass function of x i. So, here each of the x i is can take value 0 or 1.

So, this can be simplified as p to the power sigma x i, 1 minus p to the power n minus sigma x i. So, we have to differentiate this with respect to p and solve it. So, what we can do we can consider log of l, which I call as a small l of p that is equal to sigma x i log p plus n minus sigma x i log of 1 minus p. So, d l by d p is equal to sigma x i by p minus n minus sigma x i by 1 minus p that is equal to 0 now then this gives.

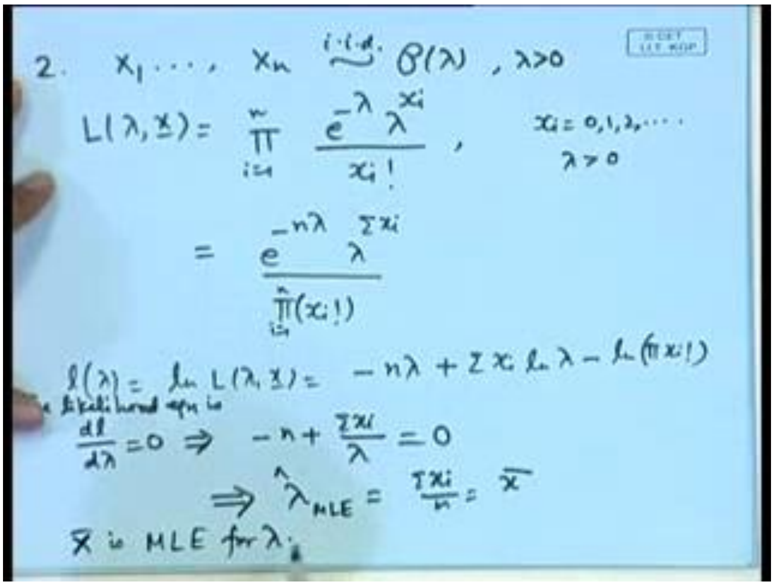
(Refer Slide Time: 05:25)



So, after some simplification we will get p hat is equal to so in fact you can solve it let us combine the term, so it will give me 1 minus p sigma x i minus n p plus p sigma x i divided by. So, this we can put p is belonging to the interval 0 to 1.

We can take the case p is equal to 0 and p is equal to 1 separately. So, this becomes sigma x i minus n p is equal to 0; that means, p hat is equal to sigma x i by n, which I can call say x by n; where x is the sigma of x i. If you compare it with the method of moment estimator actually it is the same thing here; that means, the maximum likelihood estimator of the probability of success is actually the number of successes in n trails divided by the number of trails, that is the proportion of the number of successes which is a very logical estimator and it is coming through the method of maximum likelihood estimation.

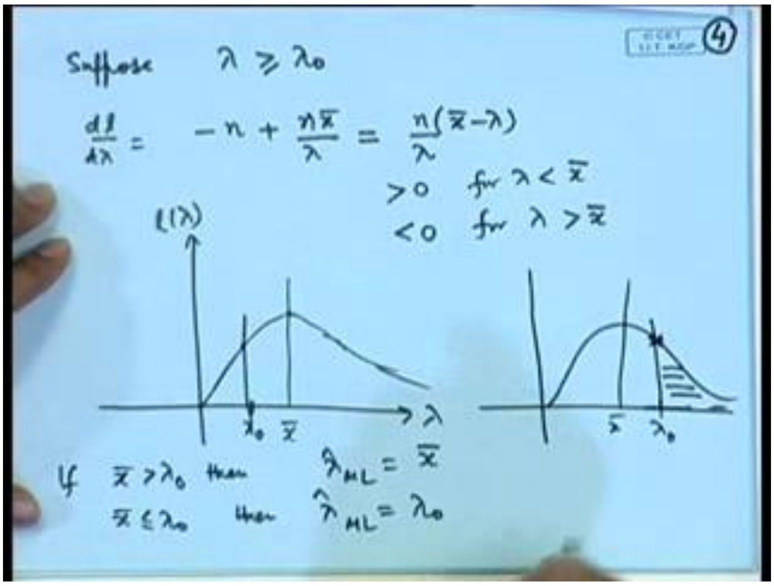
(Refer Slide Time: 06:42)



Let us take say X 1, X 2, X n to be a random sample form poisson distribution with parameter lambda; let us write down the likelihood function that is equal to e t o the power minus lambda, lambda to the power x i, divided by x i factorial product i is equal to 1 to n. Here each of the x i is can take values 0, 1, 2 and so on and can dies positive. So, this is equal to e to the power minus n lambda, lambda to the power sigma x i divided by product x i factorial i is equal to 1 to n.

So, we can write the log likelihood function x minus n lambda plus sigma x i log of lambda minus log of product x i factorial. So, the likelihood equation d l by d lambda is equal to 0, that is minus n plus sigma x i by lambda is equal to 0. So, the solution of this gives lambda hat MLE is equal to sigma x i by n that is x bar. So, x bar is the maximum likelihood estimator for lambda. Note here that in these 2 cases the maximum likelihood estimator and the method of moments estimators are same; however, as we will see it is not a rule, in many cases they will not be the same let me take an example of that kind.

(Refer Slide Time: 08:46)

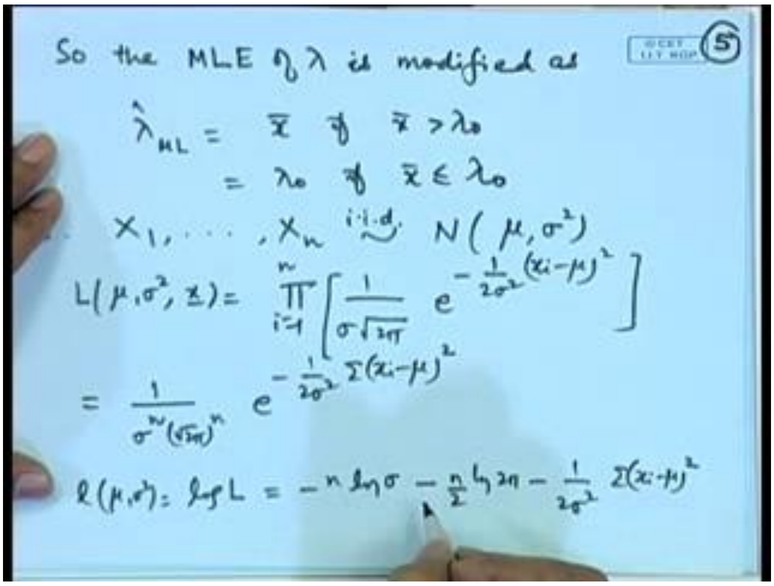


Another thing that we can observe here the maximization is done over the parameter space. So, in case there is a modification for example in the previous exercise, suppose we know that lambda is say greater than or equal to lambda naught; that means, we know that the rate of arrival in a Poisson process is bigger than a prescribed quantity. In that case if we look at x bar this is actually the maximization over the full parameter space and we may have a situation where x bar is actually less than lambda naught, in that case this does not satisfy the property that likelihood function is maximized over the parameter space.

So, what we do? Then consider the behavior of the likelihood function, what we are getting is d l by d lambda is equal to. So, we write it in minus n plus n x bar by lambda that we can write as x bar minus lambda, n by lambda that is n x bar by lambda minus lambda minus n. So, you can see here that if lambda is less than x bar then this is positive; it is less than 0 for lambda greater than x bar. So, we look at the behavior of the function it is increasing up to x bar and then decreasing after x bar.

So, let us plot it as a function of lambda the likelihood function; that means, on this side I have log likelihood, it is increasing and then it is decreasing this is the point x bar. Now suppose lambda naught is here, if lambda naught is here then you consider the maximum value x bar is satisfying this condition and therefore, x bar remains the MLE; that is if x bar is greater than lambda naught then lambda hat ML is x bar. However, we may have a situation where this is x bar and lambda naught is say here, in that case you see the parameter space is this; that means, the maximum value that is occurring is at actually lambda naught; that means, if x bar is less than or equal to lambda naught then lambda hat ML is equal to lambda naught.

(Refer Slide Time: 11:38)

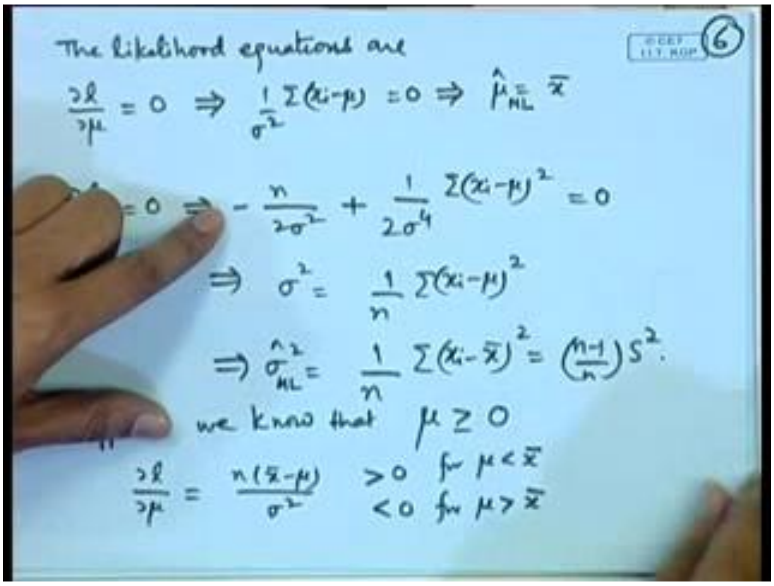


So, our estimators is then modified, so the maximum likelihood estimator of lambda is modified as let me write it as lambda hat ML this is equal to x bar; if x bar is greater than lambda naught it is equal to lambda naught, if x bar is less than or equal to lambda naught. Now this brings into focus another important property or you can say another important aspect of the maximum likelihood estimator, which would have been mixed by the method of moments; because in the method of moments the restriction on the parameter space does not play any role their we simply look at the moment which is not affected by the restriction on the parameter space and so the method of moment estimator remains as x bar. Whereas here you see the effect of the restriction on the parameter space is getting effected on the maximum likelihood estimator which is actually a reasonable thing, because if x bar is less than or equal to lambda naught which is going outside the parameter space, we should not take x bar as an estimate for lambda.

Let us take another popular example X 1, X 2, X n follows normal mu sigma square. If we consider the likelihood function here then it is a function of 2 parameters mu and sigma square. So, L mu sigma square x bar, X that is equal to product of 1 by sigma root 2 pi, e to the power minus 1 by 2 sigma square x i minus mu square. Now this term we will simplify this can be written as 1 by sigma to the power n root 2 pi to the power n, e to the power minus 1 by 2 sigma square sigma x i minus mu square.

So, the log likelihood function minus n log sigma, minus n log n by 2 log 2 pi, minus 1 by 2 sigma square, sigma x i minus mu square. Now note here that I have written it as minus n log sigma; now one may ask that if parameter is sigma square then whether we should write it as sigma. Now answer is that both are because maximum likelihood estimation is invariant under the transformation of the parameters, suppose I obtain the MLE of sigma square in place of sigma and then I want for sigma, then I should simply take the square root of that. So, I may write it like this also it does not make any difference.

(Refer Slide Time: 14:50)



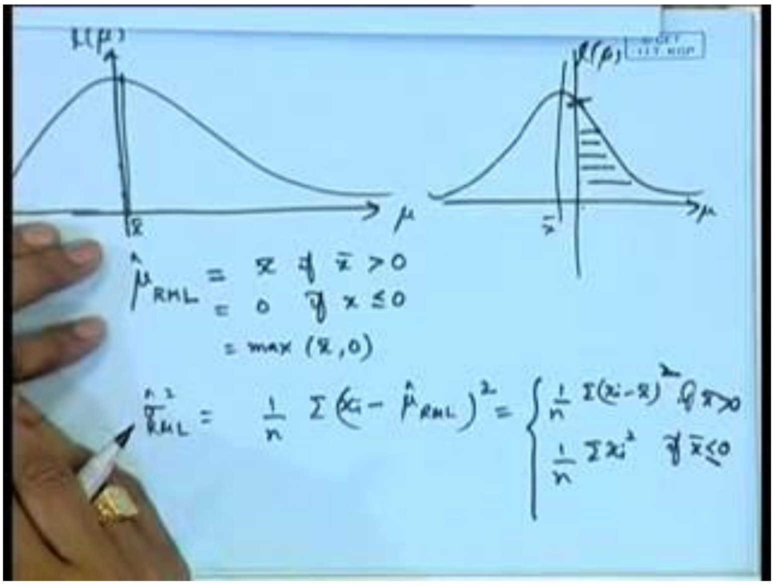
So, the likelihood equations in this case will be del l by del mu is equal to 0; that means, sigma x i minus mu 1 by sigma square is equal to 0, which will give mu hat is equal to x bar and del l by del sigma square is equal to 0 that gives minus n by 2 sigma square, plus 1 by 2 sigma to the power 4, sigma x i minus mu square, this gives sigma square is equal to 1 by n sigma x i minus mu square; that means, a equation for sigma square involves mu. So, since we have already got the solution for mu we can substitute; so sigma hat square that is a maximum likelihood estimator becomes that is n minus 1 by n S square.

You can note here that it is similar to the method of moment's estimator in this case. So, in many situations the method of moment's estimators and the maximum likelihood estimates concede, but there are other cases where they may not concede. We already have seen the example that if the parameter space gets modified then the maximum likelihood estimator gets modified. For example, in this particular situation suppose I consider suppose we know that mu is greater than or equal to 0 now this type of situation occurs when through some experience we already know that the mean is actually non negative although, the variable may be normally distributed, but because of the way the experiment has been framed or any other reason the parameter space is restricted; that means, we know that the mean is greater than or equal to 0.

Now, you see here we have the maximum likelihood estimator for mu as x bar and for sigma square it is n minus 1 by n s square. Now if we see that x bar by observation gives us a negative value then it will become unreasonable estimators for mu, which is taken to be greater than or equal to 0. So, we can analyze it in a proper way by looking at the behavior of the log likelihood function which we have actually maximized. So, if we look at with respect to mu we have del l by del mu as equal to. So, del l by del mu that function we got it as n times x bar minus mu by sigma square.

So, you can see that it is greater than 0 for mu less than x bar and it is less than 0 for mu greater than x bar; that means, the function is increasing up to x bar and decreasing beyond x bar.

(Refer Slide Time: 18:06)

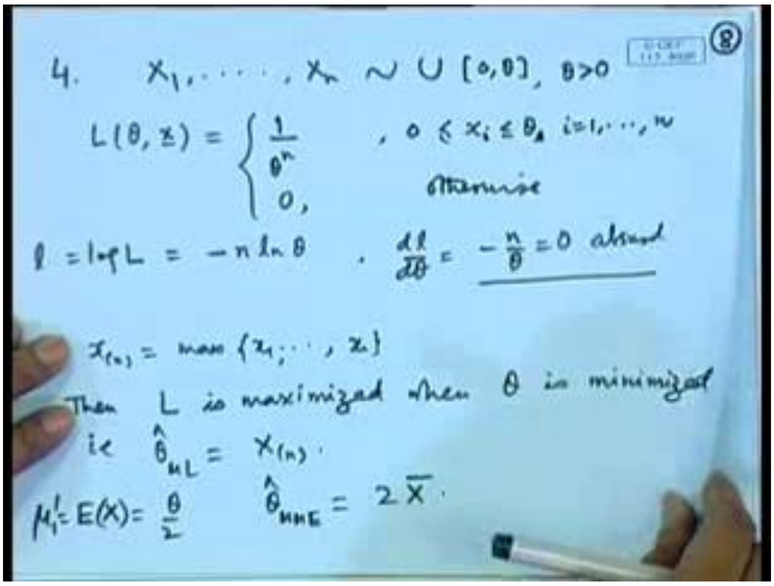


That means the shape of the likelihood function as a function of mu is something like; so this could be the maximizing choice. Now if x bar is bigger than 0 then we may take x bar, but we may have a situation where x bar may be less than 0. So, if the parameter space is this then the maximum is occurring at 0 itself.

So, the modified or you can say the restricted maximum likelihood estimator becomes x bar if x bar is greater than 0 it is equal to 0, if x bar is less than or equal to 0; which we can actually call as maximum of x bar and 0. Now is there any effect on the estimator for sigma square, the answer is yes because the maximum likelihood estimator for sigma square was obtained by substituting the estimator for mu in this second likelihood equation, so we will get sigma hat square ML as 1 by n sigma x i minus mu hat are ML square.

So, this we can write as when x bar is positive then this is the old 1 that is 1 by n sigma x i minus x bar whole square if x bar is positive; and it is equal to 1 by n sigma x i square if x bar is negative or less than or equal to 0. We will show later on that when the parameter space is restricted, this restricted maximum likelihood estimator has a greater performance as compared to the usual maximum likelihood estimator that we obtained before; we will define the criteria of better a little later.

(Refer Slide Time: 20:35)



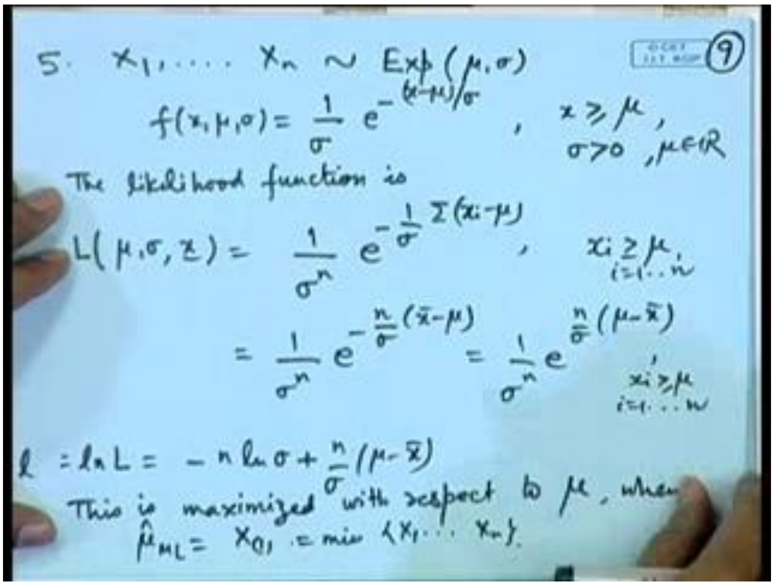
Let us take another example where this process of argument does not seem to work; let us consider say a random sample from uniform distribution on the interval say 0 to theta where theta is a positive real number. So, here the likelihood function is simply 1 by theta to the power n for 0 less than or equal to x i less than or equal to theta, for i is equal to 1 to n it is 0 otherwise. Now you see if you follow the previous procedure of taking logarithm and differentiating what I will get? If I take log of L then that will give me minus n l n theta and if I differentiate let me call it is small l so d l by d theta that will give me minus n by theta. So, if I put this equal to 0 this is actually observed. So, why this is happening?

This is happening because we are not taking care of the region, when we are differentiating and putting equal to 0 basically we are trying to find out the minimum and maximum over a range of parameter. Here the range of the parameter value is dependent upon x i so that is not been taken care by this kind of process. So, let us look at a direct argument, our aim is to maximize the likelihood function 1 by theta to the power n with respect to theta; since theta is in the denominator it corresponds to the minimization with respect to theta, what is a minimum value of theta?

The minimum value of theta will be actually since E theta is greater than each of the x i the minimum value that theta can take will be the maximum of x 1, x 2, x n. So, if I call x n as the maximum of x 1, x 2, x n that is largest order statistic then L is maximized when theta is minimized that is theta hat ML is equal to X n. So, this is the maximum likelihood estimator for theta. Suppose I want to find out the method of moments estimator here, what is that? Consider the mean of this distribution that is the first moment that is theta by 2.

So, theta hat method of moments estimator that will be twice and in place of mu 1 prime will put X bar. So, you can see the situation here the method of moment estimator and the maximum likelihood estimator are totally different. In fact, they are here it is the maximum of the observation and here it is average 2 times the average of the observations; further this example illustrates that the method of taking logarithm and differentiating does not always work.

(Refer Slide Time: 24:00)

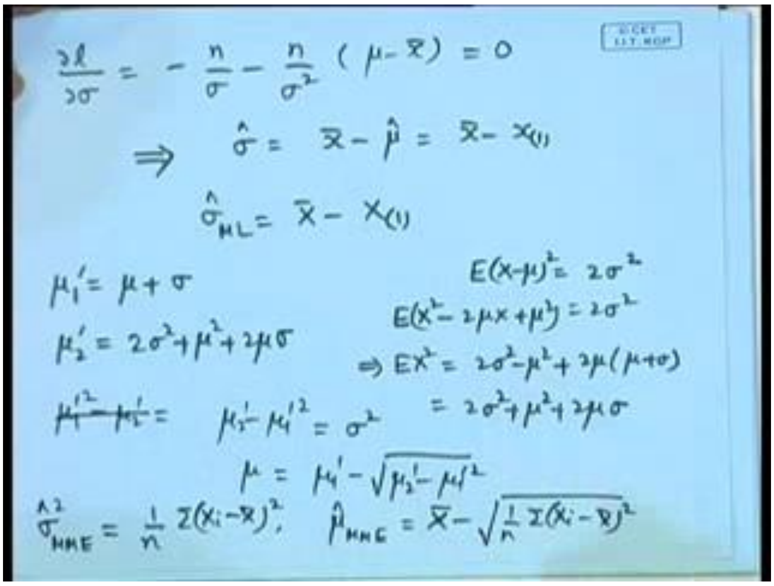


Let us take another examples say X 1, X 2, X n follow exponential distribution let me consider say 2 parameter exponential distribution mu sigma; that means, the density function is 1 by sigma e to the power x minus mu by sigma that is the density of; here x is greater than or equal to mu sigma is positive and mu is any real number. The likelihood function will be 1 by sigma to the power n, e to the power minus 1 by sigma, sigma x i minus mu, each x i is greater than or equal to mu which we can write as 1 by sigma to the power n, e to the power minus n by sigma x bar minus mu or 1 by sigma to the power n, e to the power n by sigma mu minus x bar.

Now, if we want to maximize this function with respect to mu, the likelihood equation and then derivative will not give a result like in the uniform case because the log likelihood function minus n l n sigma, plus n by sigma mu minus x bar; you can see here if I differentiate with respect to mu and put equal to 0, I get an absurd result so; however, with respect to sigma we can do that, but for finding out the maximum likelihood estimator with respect to mu we can use a direct argument, this is a increasing function of mu. So, the maximum value will be attained when mu attends its maximum value.

Since mu is always less than or equal to x i, the maximization will occur when mu is the minimum of the x i. So, this is maximized with respect to mu when mu hat ml is equal to X 1 that is the minimum of the observations. So, we can substitute this here and get the estimator for sigma also.

(Refer Slide Time: 27:00)



Consider the derivative of this with respect to sigma that will give minus n by sigma, minus n by sigma square, mu minus x bar this is equal to 0. So, this gives sigma hat is equal to x bar minus mu hat that is equal to x bar minus x 1.

So, sigma hat ML is equal to X bar minus X 1, let us see what is the method of moments estimator in this case; to obtain the method of moments estimator we have to consider the moments of this distribution, now here it is a 2 parameter distribution I will have to find out first 2 moments. So, mu 1 prime is equal to mu plus sigma, mu 2 prime now for that we can consider certain transformation expectation of X minus mu square that is equal to twice sigma square, so expectation of X square minus 2 mu X plus mu square is equal to twice sigma square.

So, expectation of X square is equal to 2 sigma square, minus mu square plus twice mu expectation of X that is mu plus sigma that is equal to twice sigma square plus mu square plus 2 mu sigma. So, the second moment is twice sigma square plus mu square plus 2 mu sigma. Now if we consider mu 1 prime square minus mu 2 prime rather I consider mu 2 prime minus mu 1 prime square then that gives me sigma square and therefore, mu becomes mu 1 prime minus square root of mu 2 prime minus mu 1 prime square.

So, the method of moments estimators for mu and sigma square will be obtained as sigma hat square MME that is equal to 1 by n sigma x i minus x bar whole square and mu hat MME will be equal to X bar minus a square root 1 by n sigma x i minus x bar whole square. Note here that the maximum likelihood estimators and the method of moment estimators are quite different from each other and again therefore, the question arises that which of this is better.