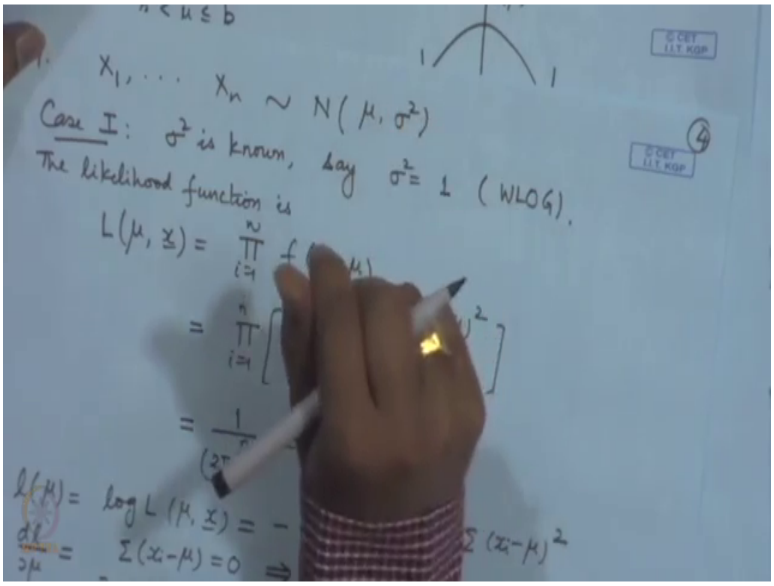
Statistical Inference
Prof. Somesh Kumar
Department of Mathematics
Indian Institute of Technology, Kharagpur

Lecture - 10
Finding Estimators- IV

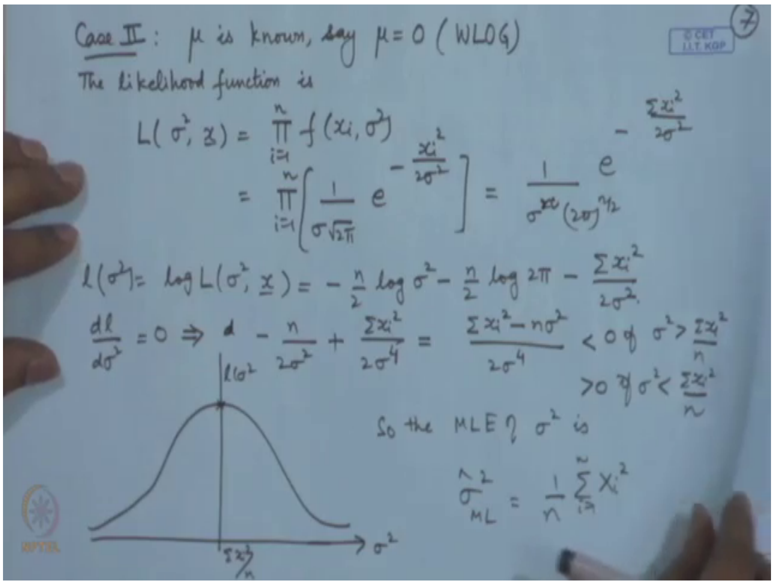
Now, let me take additional cases in the case of normal distribution.

(Refer Slide Time: 00:24)



See here we have taken the case for estimating mu, because sigma square was known. Now, you may have another identical situation where mu may be known and we may be interested in the estimation of sigma square.

(Refer Slide Time: 00:45)



So, let us look at this situation then say mu is known. If mu is known, then without loss of generality we my put mu is equal to 0 because we can always shift all the observations by mu naught. For example, if I say that mu is known equal to mu naught then we may put it is equal to 0.

So, now you look at the likelihood function, notice here the problem gets modified in the maximum likelihood estimation as soon as the information about the parameters is changed. So, the likelihood function is the product of the density functions of x 1, x 2, x n that is equal to 1 by sigma root 2 pi e to the power minus x i square by 2 sigma square i is equal to 1 to n.

So, we can write it in a more compact fashion. It becomes equal to 1 by sigma square sigma to the power n 2 pi to the power n by 2 e to the power minus sigma x i square by 2 sigma square. Notice here that sigma is occurring in the denominator as well as it is occurring in the denominator of the exponent. Therefore it is beneficial to considered the log likelihood function that is equal to minus n by 2 log of sigma square minus n by 2 log of 2 pi minus sigma x i square by 2 sigma square.

So, we considered the likelihood equation that is d l by d sigma square is equal to 0. So, so when you differentiate this you get minus n by 2 sigma square plus sigma x i square by twice sigma to the power 4. Notice here that I am considering sigma square as a parameter, one may be miss lead by considering sigma as a parameter, and then you may be getting a slightly different derivative here. So, later on we will show that the two procedures will lead to the same answer, the identical answer. That means, whether you are considering estimation of sigma are you are considering estimation of sigma square, it should not lead to contradictory a statements.

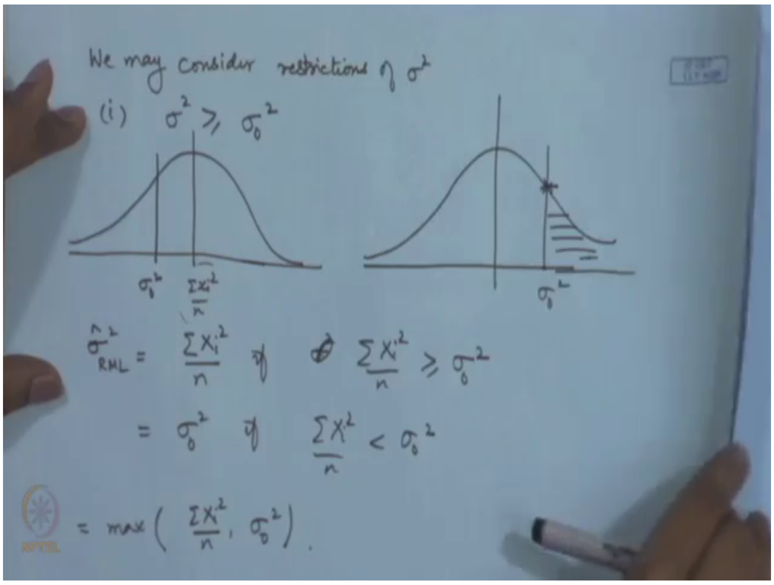
Now, we write it in a slightly modified fashion sigma x i square minus n sigma square by twice sigma to the power four. So, notice here this will be less than 0 if sigma square is greater than sigma x i square by n; and it is greater than 0 if sigma square is less than sigma x i square by n.

So, if we look at the plot of the likelihood function, then naturally the likelihood function is increasing up to sigma x i square by n, because the derivative is positive for sigma square less than sigma x i square by n. So, it is increasing up to this, and there after it is decreasing. So, the maximum occurs at sigma x i square by n. So, the maximum likelihood estimator of sigma square is we will write m l just to denote that it is the maximum likelihood estimator that is turning out to be 1 by n sigma x i square.

Now, you can look at the variation in place of mu is equal to 0 if we had put mu is equal to mu naught, then what would have been the modification. Here, we would have got x i minus mu naught whole square. Therefore when we considered the derivative here we would have got then increasing and decreasing nature for sigma x i minus mu naught whole square by n. Thereby, the answer would have been 1 by n sigma x i minus mu naught whole square.

So, now once again let me show you the effect of the prior information in this. Suppose on sigma square we have certain information because as you know sigma square is the variance. Now, the variances are the reciprocal of that is known as the precision. So, the variability may be known in advance or it may have certain restrictions.

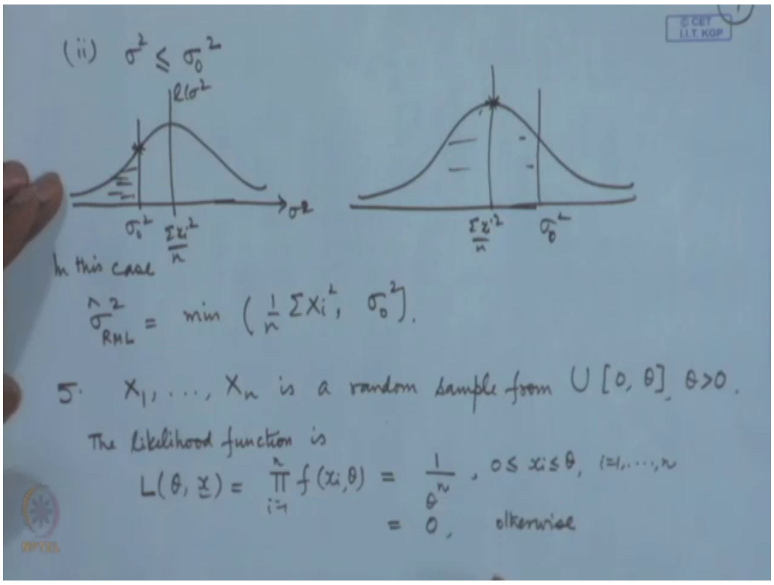
(Refer Slide Time: 06:02)



For example, we may consider say restrictions on sigma square. Say for example, sigma square may be greater than or equal to sigma naught square. Now, if you consider sigma square greater than or equal to sigma naught square, then in this case there will be two cases, because sigma naught square may occur here are sigma naught square may occur here.

So, let us see. This is sigma x i square by n, and it may happen that sigma naught a square is here. So, in this case, the maximum occurs at this point. Whereas, if sigma naught a square occurs here, in that case our region of maximization is here because sigma square is greater than or equal to sigma naught a square, in that case the maximum will occur at sigma naught square. So, we conclude that sigma hat square restricted m l is equal to sigma x i square by n if sigma square if sigma x i square by n is greater than or equal to sigma naught square. It is equal to sigma naught square if sigma x i square by n is less than sigma naught a square. That means, we can write it as maximum of sigma x i square by n and sigma naught square. In a similar way, one may considered the case of an upper bound on sigma square.

(Refer Slide Time: 08:03)



Let me take sigma square less than or equal to sigma naught square. So, once again if we look at the plot of the likelihood function, in that case if sigma naught square is occurring here, now this is our region of maximization. So, the maximum will occur at sigma naught square, whereas if sigma naught square occurs here, then this is our region of maximization, and we get the maximum here.

So, in this case the maximum likelihood estimator of sigma square will be minimum of 1 by n sigma x i square sigma naught square. So, the effect of the information or the prior information about the parameter plays a role in the maximum likelihood estimation. And that is one important feature which distinguishes the method of maximum likelihood estimation from various other methods.

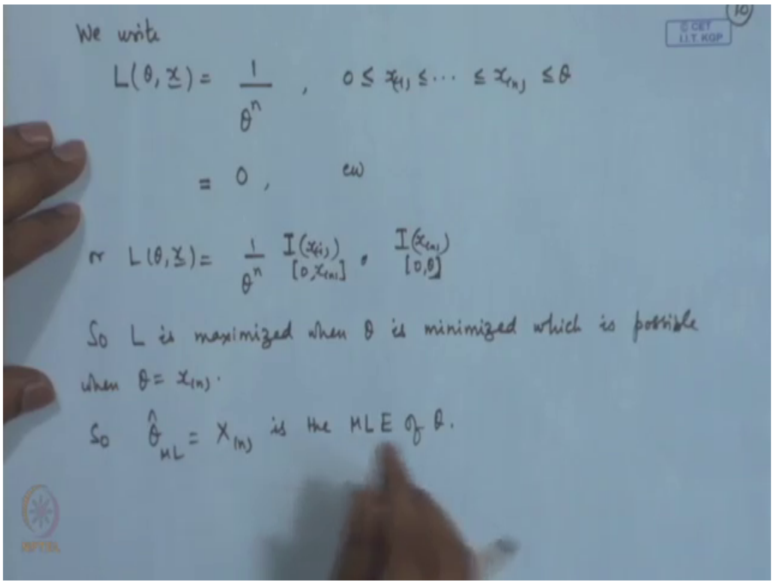
The example that I have discussed take into account that the likelihood function or the log of the likelihood function is a nice are you can say smooth function, because we are able to differentiate and carry out the usual arguments of the analysis. Now, in certain situations that may not be possible.

Let me take up another case say x 1, x 2, x n is a random sample from uniform zero theta distribution, where theta is the unknown parameter which is certainly positive. We are interested in the maximum likelihood estimation for theta. If you recollect the method of moments estimator for theta was 2 x bar, because the mean of the uniform distribution is theta by 2. So, the first sample movement that is x bar would be the method of moments estimator for theta by 2, that means, 2 x bar will be the method of moments estimator for theta. Let us look at the maximum likelihood estimator here.

So, the likelihood function is L theta x that is equal to product of f x i theta i is equal to 1 to n. Now, this we write as 1 by theta to the power n because the density function of the uniform distribution on the interval 0 to theta it is 1 by theta. So, it will become 1 by theta to the power n. But at the same time let us not forget that each of the x i's lies between 0, and theta this is for i is equal to 1 to n. Now, a we should also write that it is 0 at other places.

Now, a common thing which we have been applying a real that you take the log of this and differentiate with respect to theta and put equal to 0. Now, in this case what it would lead to you will get minus n log theta. And if you differentiate will get minus n by theta which you put equal to 0 will give you and absurd answer. The reason for these absurdities that we have not taken care of the full likelihood function; the full likelihood function takes into account this portion also.

(Refer Slide Time: 12:19)



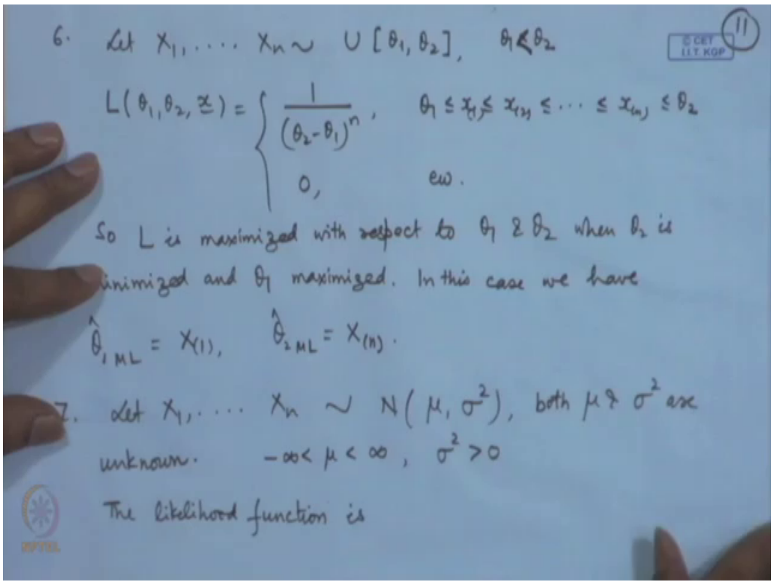
So, we write it in a slightly more compact fashion as follows. We may right the likelihood function as 1 by theta to the power n 0 less than or x 1 less than or equal to x n less than or equal to theta. Or we can also write it as 1 by theta to the power n i here we can say that all the x i s are from 0 to x n and multiplied by x n itself lies between 0 to theta.

Now, if you look at the maximization of this with respect to theta, now the theta is occurring in the denominator, so that means, what is the minimum value of theta the minimum possible value of theta is x n theta cannot be below x n because of the observations each of the observations lies between 0 to theta. So, L is maximized when theta is minimized which is possible when theta is equal to x n.

So, theta hat ML is equal to x n is the maximum likelihood estimator of theta what is the maximum of the observations. So, you can see here the result is quite different from the method of moments estimation here, because in the method of moments we would have got 2 x bar. So, this is certainly different and later on we will study the criteria that which one should be preferred here; that means, weather MME is better here, or ML is better here which one should prefer. So, we will discuss about those criteria later on.

This example shows that one should not blindly use the differentiation and put equal to 0, because this will not give the answer in this particular situation. Similar thing would occur for example, if I consider two parameter you uniform distribution.

(Refer Slide Time: 15:01)



Suppose, I take a random sample from uniform theta 1 to theta 2, where theta 1 is certainly less than or equal to theta 2. So, in this particular case, we have two unknown parameters here. And we considered the maximum likelihood estimation. So, as before we considered the likelihood function, and this will be it is equal to 0 elsewhere.

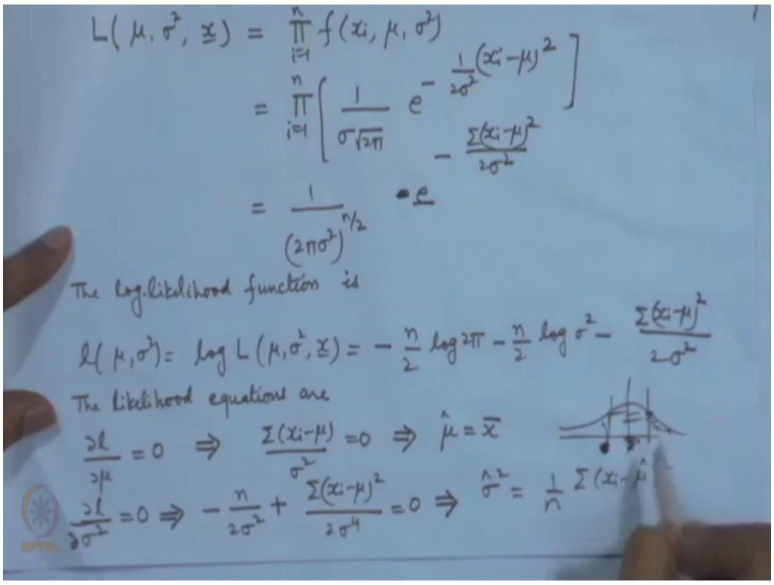
Now, you notice the likelihood function here. The likelihood function has theta 2 minus theta 1 in the denominator which is the positive quantity. And we are looking at the maximization; that means, theta 2 minus theta 1 should be minimum. That means, theta 2 should be minimum and theta 1 should be maximum. Now, if you look at the nature of the observations, all the observations lie between theta 1 to theta 2. Therefore, the minimum of the observations is certainly greater than or equal to theta 1. And the maximum of the observations is certainly less than or equal to theta 2. So, L is maximized with respect to theta 1 and theta 2, when theta 2 is minimized, and theta 1 maximized.

So, in this case, we have theta 1 hat maximum likelihood estimator is equal to the minimum of the observations. And theta 2 hat ML is equal to the maximum of the observations. And now this is an example where have considered two parameter problem. So, the method of maximum likelihood estimator can be used for the maximization of the likelihood function, when there can be more than one parameter. And in that case the maximization should be consider with respect to all the parameters. So, in this case, you can see the simultaneous maximum is occurring.

Now, let us go back to the case of normal distribution that I discussed earlier here I had taken special cases. If you see carefully, if we consider normal mu sigma square here, I have taken sigma squared to be known. So, in effect I have reduced a two one parameter problem. Similarly, if you look at mu is known, then once again the parameter has been reduced to sigma square alone. So, in effect this problem also reduced to one parameter problem. However, in general both the parameters in a normal distribution maybe unknown and in that case let us look at the solution.

So, let me discuss in detail. So, we have x 1 x 2 x n a random sample from normal mu sigma square as before. However, both mu and sigma square are unknown. So, in general you remember that in the normal distribution the mean parameter may vary from minus infinity to plus infinity and the variance parameter will be from 0 to infinity. Now, in this case when we want to find out the maximum likelihood estimator we will like to find out for both mu and sigma square.

(Refer Slide Time: 19:48)



So, let us write down the likelihood function. So, the likelihood function is L mu sigma square x. Notice here that this has become function of both mu and sigma square now. So, this is a joint density function as before in the earlier cases I had substituted special values of mu were sigma square as the case was. In this case, we will have to write down the full form of the density function of a normal distribution that is 1 by sigma root 2 pi e to the power minus 1 by 2 sigma square x i minus mu whole square.

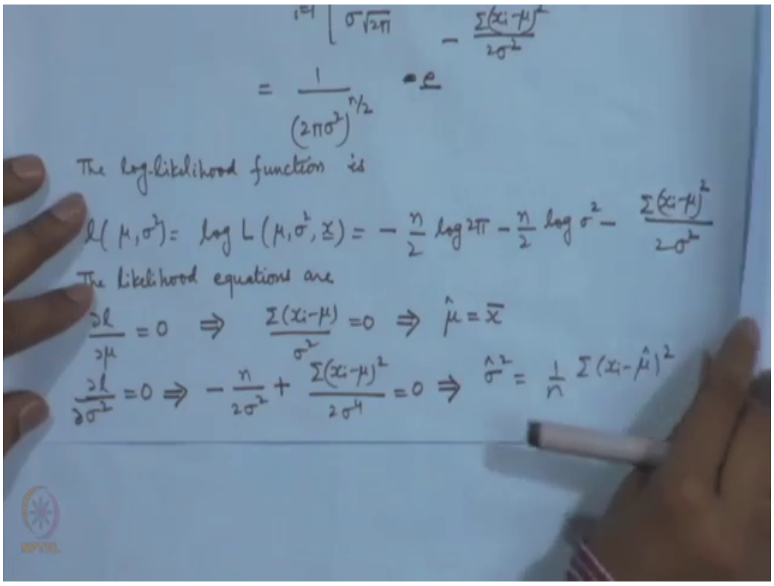
So, we write it in a slightly more compact fashion. This becomes 1 by 2 pi sigma square to the power n by 2 e to the power that when you take. So, it will become e to the power minus sigma x i minus mu square by twice sigma square. Again you observe the parameters for which we need the estimators they are occurring in the exponent as well as they are occurring in the main form here.

So, it will be beneficial if we considered the log likelihood as before. So, the log likelihood L mu sigma square log of L mu sigma square x that is equal to minus n by 2 log of 2 pi minus n by 2 log of sigma square minus sigma x i minus mu square divided by twice sigma square. This equation this function involves mu and sigma square two variables. We need to maximize this with respect to both mu and sigma square.

So, since this function is still very nice a smooth function, so we can still use the direct calculus method for example by taking the first order derivatives putting them equal to 0. They are giving us the likelihood equation. The solutions of that will be the points of a minimum or maximum which we can check separately that they would be actually leading to the maximization points, they will not be the points of minimum.

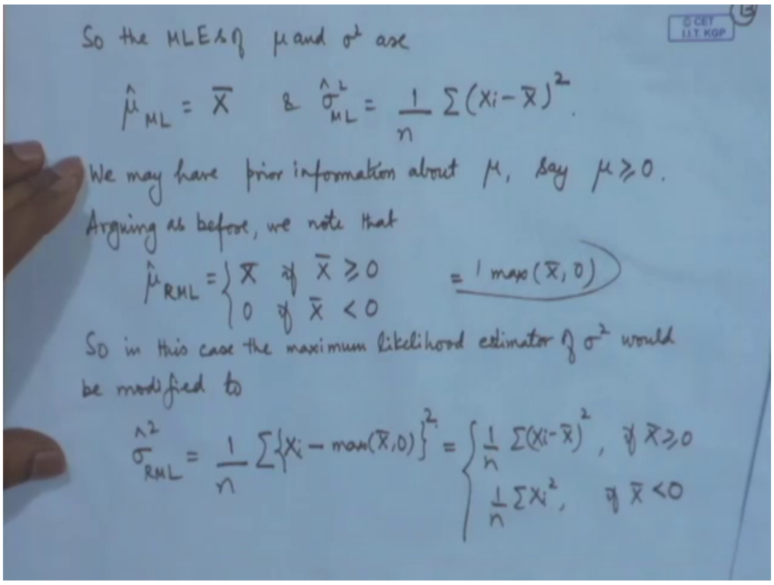
So, in this case for example, we write down the likelihood equations. The likelihood equations are del l by del mu is equal to 0 that is sigma x i minus mu by sigma square is equal to 0 which we can further write because this can be easily simplified sigma square is in the denominator that would give me mu hat is equal to x bar.

(Refer Slide Time: 23:18)



The other equation is del l by del sigma square is equal to 0 which will give me minus n by twice sigma square minus plus sigma x i minus mu square by twice sigma to the power 4 equal to 0. Which will give me sigma square is equal to 1 by n sigma x i minus mu hat square. Actually the equation is sigma square is equal to 1 by n sigma x i minus mu square, we substitute the value of mu from the first equation and substitute here.

(Refer Slide Time: 24:07)



So, the maximum likelihood estimator then turn out to be, so the maximum likelihood estimators of mu and sigma square are mu hat ML is equal to x bar and sigma hat square ML is equal to 1 by n sigma x i minus x bar whole square. In this case, you may notice at these are the same as the method of moment estimators for this particular problem. But once again as I mentioned earlier a method of maximum likelihood can take care of many other possibilities also.

For example, we may have say prior information about mu say mu is greater than or equal to 0. In that case once again we look at the likelihood function here we are getting n x bar minus mu. So, if you plot the behavior with respect to mu, then the maximum is occurring at x bar.

But if x bar is greater than 0, I will consider 0 here. And this region is coming. So, the maximum likelihood estimator will be x bar. However, if 0 occurs on this side and then we have this portion then the maximum will occur at 0. So, are going as before we know that mu hat restricted ML will be equal to x bar if x bar is greater than or equal to 0. It will be equal to 0 if x bar is less than 0, which we can actually write as maximum of x bar and 0.

Now, if we use this in that case the second equation the solution will get modified, because for sigma square the estimator was 1 by n sigma x i minus the estimator of mu.

And if the estimator for mu gets modified, immediately the estimator for sigma square I will also get modified.

So, in this case, the maximum likelihood estimator of sigma square would be modified to sigma hat square RML is equal to 1 by n sigma x i minus maximum of x bar 0, which we can write as 1 by n sigma x I minus x bar whole square if x bar is greater than or equal to 0. And it will become 1 by n sigma x I square if x bar is less than 0. So, the placing of additional information about the parameter changes the maximum likelihood estimators.

I will consider a few more examples in the next class and also then we will see there are certain desirable properties which are basically called the large sample properties that the maximum likelihood estimator satisfy, and because of this the method as wide applicability among a statisticians. So, in the tomorrow's class we will consider various properties of the maximum likelihood estimators. And then, we will proceed to determining the criteria for judging the goodness of the estimators.

Thank you today.