Statistical Inference

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

Lecture - 09
Finding Estimators- III

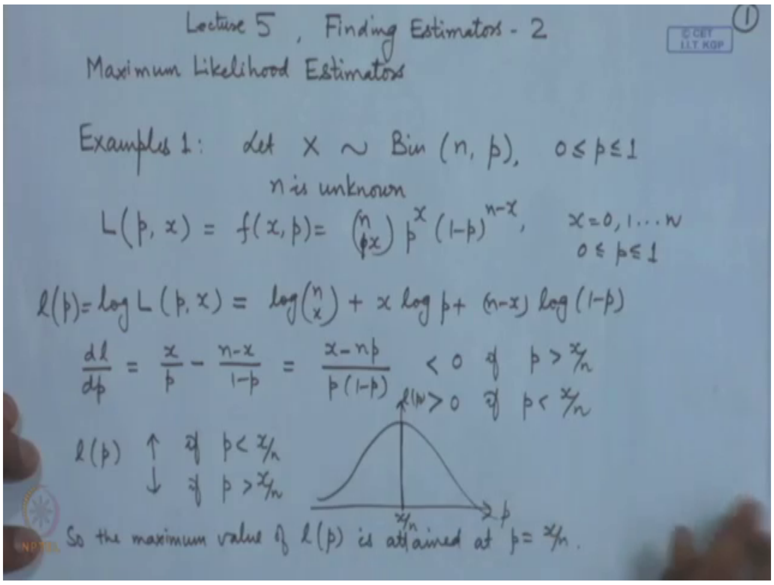
So, friends, in my earlier lecture, I have told the some criteria for judging the goodness of estimators. For example, unbiasness is one criteria, consistency is one criteria, that means, if an estimator is unbiased it is in general preferable to an estimator which is not unbiased. Similarly, an estimator which is consistent is preferable to an estimator which is inconsistent.

So, after there are many other criteria's which we will be discussing in the further discussion, then we dwelt upon on how to find out the new estimators or how to propose the estimators. We have mainly discussed two methods, one is the method of moments. The method of moments concentrated on equating the sample moments with the population moments and thereby obtaining the estimates of the parameters.

The method is quite simple. And in generate quite good estimators, but then there are certain criteria which does not satisfy. For example, in many cases, we saw that the method of moments estimator were not unbiased, although in many cases they were consistent. Another popular method which was introduced in 1920s by R.A. Fisher is the well known method of maximum likelihood estimation. Here the idea is that whenever a sample is observed we look at the probability of that sample being observed, and what is the parameter value for which this probabilities or likelihood is maximized.

So, we define what is known as a likelihood function. In the previous class, I have given an example illustrating that and the general form of a likelihood function. Today, we start with various applications that is in many probability models, what are the method of maximum likelihood estimators. So, we call in general MLEs that is the Maximum Likelihood Estimators.

(Refer Slide Time: 02:38)



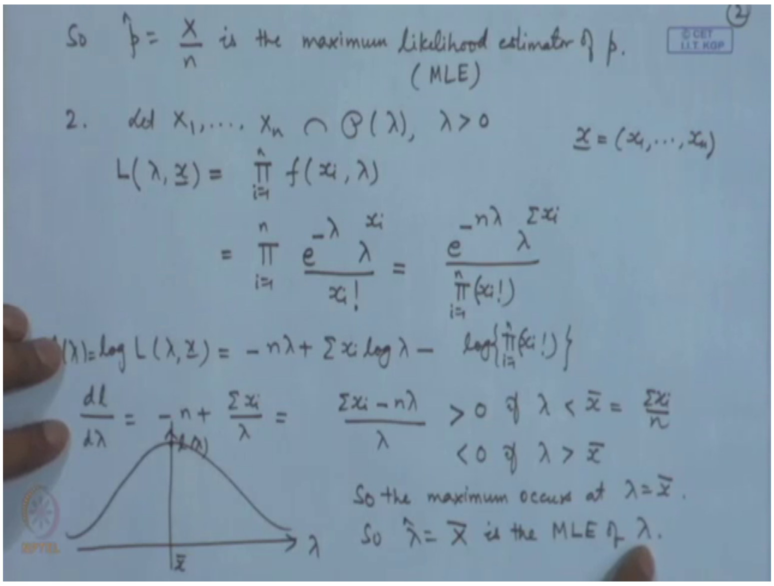
So, let me discuss some applications which are applicable to popular distributional models; maximum likelihood estimators. So, let me start with some familiar examples. Let X follow a binomial n p distribution. Now, if we say that n is known, then p is the parameter let us consider the likelihood function.

So, the likelihood function is written as a function of the parameter which is actually the density function, and in this particular case it is n c x p to the power x 1 minus p to the power n minus x. Here x takes values 0, 1 to n, and p is a number between 0 to 1. Our objective is to maximize this likelihood function with respect to p. A usual practice is to take the log of likelihood function which we call log likelihood, and we use an another notation a small l for this. So, a small l p is equal to log of likelihood that is equal to log of n c x plus x log p plus n minus x log of 1 minus p.

Now, if you look at this function, we can apply the usual method of the calculus for finding out the maximum with respect to p. So, we can consider for example, derivative of this with respect to p. So, this vanishes you get x by p minus n minus x by 1 minus p, which we can write as x minus n p divided by p into 1 minus p.

Now, if you notice this thing, this is less than 0, if p is greater than x by n and it is greater than 0, if p is less than x by n. So, we can see from here that l p this will be increasing if p is less than x by n; and it is decreasing if p is greater than x by n. Therefore, the shape of the likelihood function is something like this. If you are plotting l p, then it is attaining the maximum at the point x by n. So, the maximum value of l p is attained at p is equal to x by n.

(Refer Slide Time: 05:59)



So, we say that p hat is equal to X by n is the maximum likelihood estimator of p. Now, you notice this thing X by n is actually a sample proportion. So, we are getting that the sample proportion is the maximum likelihood estimator of the population proportion p. So, this is a natural estimator and from the method of maximum likelihood estimator, we are actually getting that as an estimator.

Let me take some more examples for the popular distributional models suppose I have a random sample X 1, X 2, X n from a Poisson distribution with parameter say lambda. Our interest is to find out the maximum likelihood estimator for the parameter lambda. As you recall the parameter lambda in the Poisson distribution represents the average arrival rate or the mean of the process in which that Poisson distribution is generated.

So, if you write down the likelihood function, L lambda and let me use the notation x for the sampled observations x 1, x 2, x n, this is nothing but the joint probability mass function of x 1, x 2, x n written at the points x 1, x 2, x n. So, this is nothing but product i is equal to 1 to n. Now, this is for one particular x i, if we write it is e to the power minus lambda, lambda to the power x i divided by x i factorial. So, this can be further simplified e to the power minus n lambda, lambda to the power sigma x i divided by product of x i factorial.

Now, as you notice we have to maximize this function with respect to lambda and this function here lambda is occurring in the exponent as well as lambda has an exponent. Therefore, it will be convenient if once again in place of the likelihood function we consider log likelihood function. So, we take log of this we call it log likelihood that is equal to minus n lambda plus sigma x i log of lambda minus log of product x i factorial.

Once again if you observe, this is a non-linear function of lambda we can apply the usual method of analysis for finding out the maximum with respect to lambda. So, let us consider the simple derivation of this with respect to lambda, so that is equal to minus n plus sigma x i by lambda that is equal to sigma x i minus n lambda divided by lambda. Easily you consider it is greater than 0 if lambda is less than x bar where x bar is actually sigma x i by n. And it is less than 0, if lambda is greater than x bar.

So, naturally if you plot the behavior of the L function, so suppose this is my x axis represents lambda, on the y axis represent L of lambda. Then for lambda less than x bar the value is positive of the derivative. Therefore, the L lambda function will be increasing. And for lambda greater than x bar this d l by d lambda is negative, therefore, this l lambda will be a decreasing function. Therefore, the maximum occurs at lambda is equal to x bar. So, the maximum occurs at lambda is equal to x bar.

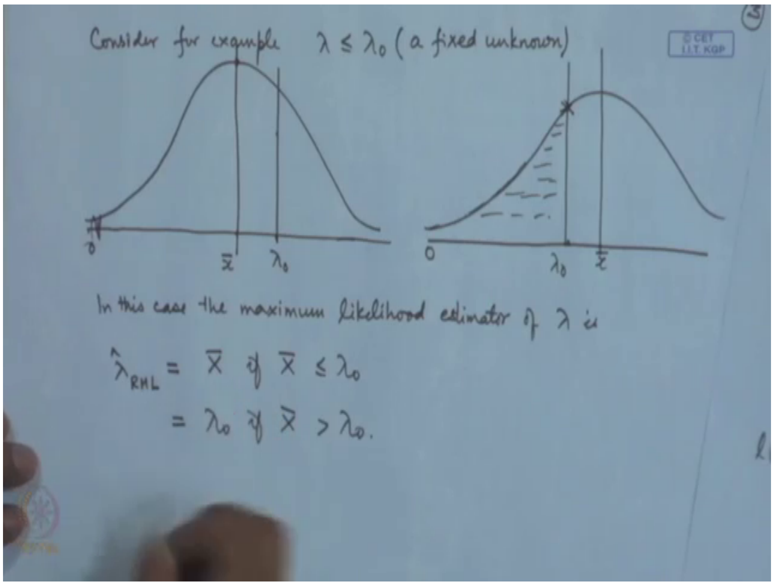
So, we say that lambda hat is equal to X bar is the maximum likelihood estimator of lambda. Once again you observe here, this is a sample mean. And in this particular case, it turns out that the sample mean is the maximum likelihood estimator of lambda. In the method of moments also, we would have got the same estimator because expectation of X bar would have been equal to because the first moment is lambda, and first sample moment is X bar. So, this would have also been the method of moment’s estimator for lambda in the case of Poisson distribution.

However, in the case of maximum likelihood estimator we have a restriction. Restriction means that whatever be the required parameter space, so maximization is over only that region; that thing is not necessarily satisfied suppose we are considering the method of moments. Because there we simply quit the sample moments with the population movements, we do not bother about what is the region of the parameter that means, the region where the parameter can vary.

Similarly, when we apply the concept of unbiasness or consistency, we do not look at the parameter is space. In that sense, the maximum likelihood estimation is more powerful and all in compassing procedure, because it takes into account what is the sampled observations as well as what is the required parameter space where you are actually considering the estimation. In that sense this has more applicability and acceptance for the user point of view.

To give an example in this case I have taken lambda to be greater than 0, that means, the arrival rate is positive which is true in general for a poison process. But suppose your physical constraints restrict the parameter space. For example, it could be a service queue, where if the number of required percent exceeds a certain number, then the service that means, and no more persons are allowed ok.

(Refer Slide Time: 13:02)

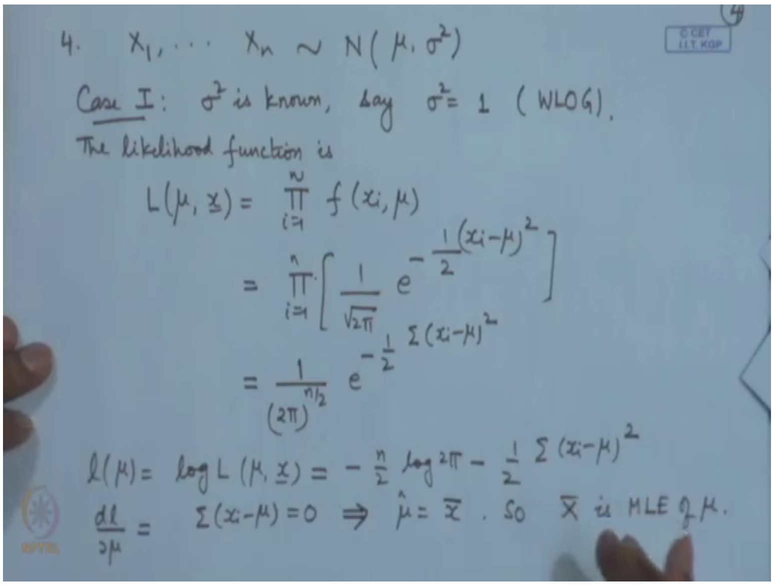


Then you may have a situation of this nature, consider for example, lambda is less than or equal to lambda naught, where lambda naught is a fixed unknown. Now, in this case, if you see we have here looked at the maximum value lambda is equal to X bar. Now, you may have two cases let me make the plot here. So, see this is X bar. Now, there maybe two cases, it could happen that lambda naught value is here. If lambda naught is here, then the maximum of the likelihood function in the region 0 to suppose this is starting point is your 0. So, 0 to lambda naught the maximum value is still occurring at lambda naught, it is still occurring at x bar.

Whereas, you may have another situation where your lambda naught maybe on this site your X bar is here. Now, if you look at the likelihood function, we are concerned only for this portion. And therefore, if you see the maximum value that is occurring here that is at lambda naught. So, we cannot say here that the maximum likelihood estimator is X bar, it is actually lambda naught.

So, in this case the maximum likelihood estimator of lambda is let me call it lambda hat RML restricted ML. So, this is equal to X bar if X bar is less than or equal to lambda naught, it is equal to lambda naught if X bar is greater than lambda naught. So, you note here that this estimator is certainly different from the method of moments estimator for this problem, because the method of moments estimator does not take care of this fact that lambda is bounded by lambda naught. So, the answer would have been still X bar for the method of moment estimator.

(Refer Slide Time: 15:46)



Let me explain the situation with some other examples also. Let us for example, take X 1, X 2, X n following normal mu sigma square distribution. Now, I consider different cases because when we are dealing with the two parameter problem, then there may be some information regarding one parameter or there may be information regarding both the parameters. I will consider all these cases.

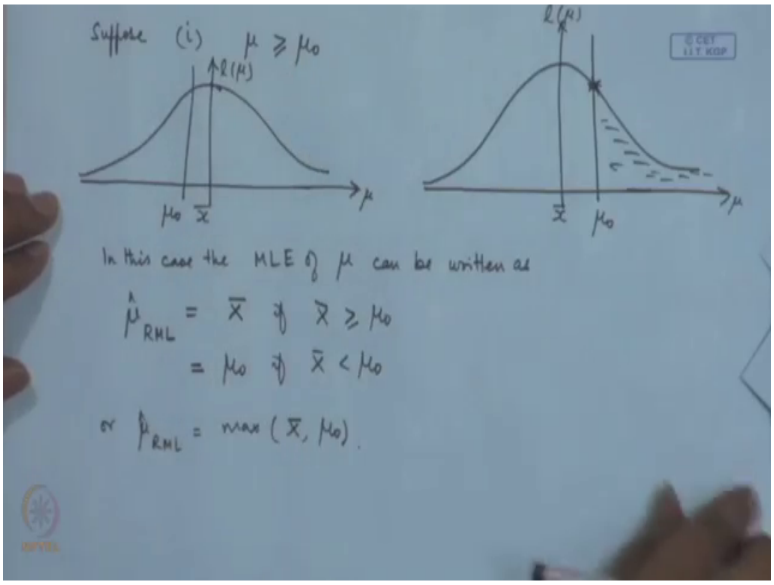
Let us take say case one say sigma square is known. So, in that case without loss of generality we can take sigma square to be one without loss of generality. So, if we write down the likelihood function, the likelihood function is LL mu, x, because when sigma square is known only one parameter is occurring here. So, it is the joint density function of X 1, X 2, X n add the observed values a small x 1 a, small x 2 a, small x n that is equal to product i is equal to 1 to n 1 by root 2 pi e to the power minus 1 by 2 x i minus mu square.

Now, we try to write it in a slightly compact fashion. So, you get 2 pi to the power n by 2 e to the power minus 1 by 2 sigma x i minus mu square. So, as before you can see here mu is occurring in the exponent, therefore it is beneficial if we consider the log likelihood. So, we consider log likelihood function as minus n by 2 log 2 pi minus half sigma x i minus mu square. In order to maximize this with respect to mu, we consider simple derivative with respect to mu which gives us sigma x i minus mu is equal to 0, which are extremely simple solution mu hat is equal to x bar. So, X bar is the maximum likelihood estimator of mu.

Now, if you look at here the parameter space for mu is minus infinity to infinity for sigma square it is 0 to infinity. So, when we took sigma square is equal to 1, the parameter is space is simply 0 to minus infinity to infinity. And if you look at the X bar, X bar is likely to be any value because in the normal distribution case the variable lies on the real line, and therefore, the average value will also lie on the real line.

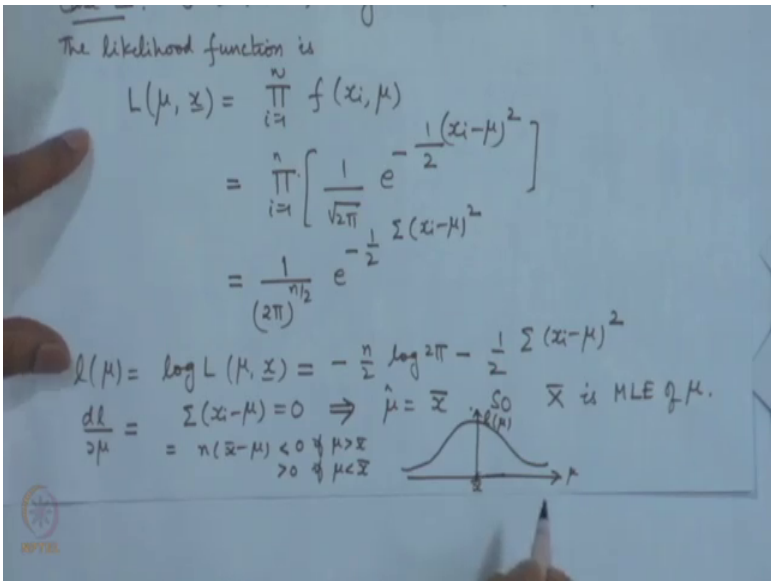
Now, if we had considered the method of moments estimator in this problem, then for mu the method of moments estimator also would have been X bar. However, let us consider say a slightly different situation.

(Refer Slide Time: 19:24)



In the same case suppose we know from the physical considerations that mean mu is either greater than or equal to mu naught less than or equal to mu naught, or it lies in an interval say mu 1 to mu 2. So, let me take one case say mu is greater than or equal to mu naught. Now, you look at the behavior of the likelihood function. So, we have observed here d l by d d mu is equal to sigma x i minus mu.

(Refer Slide Time: 19:58)



Now, this you can write as n times x bar minus mu. Now, once again you notice this, this is less than 0 if mu is greater than x bar; it is greater than 0 if mu is less than x bar. So, the nature of the likelihood function would have been of this nature that if this is mu on the x axis, on the y axis we plot L mu then for mu less than x bar the likelihood function is the log likelihood function is increasing and it is decreasing thereafter. Therefore a maximum is occurring at x bar.

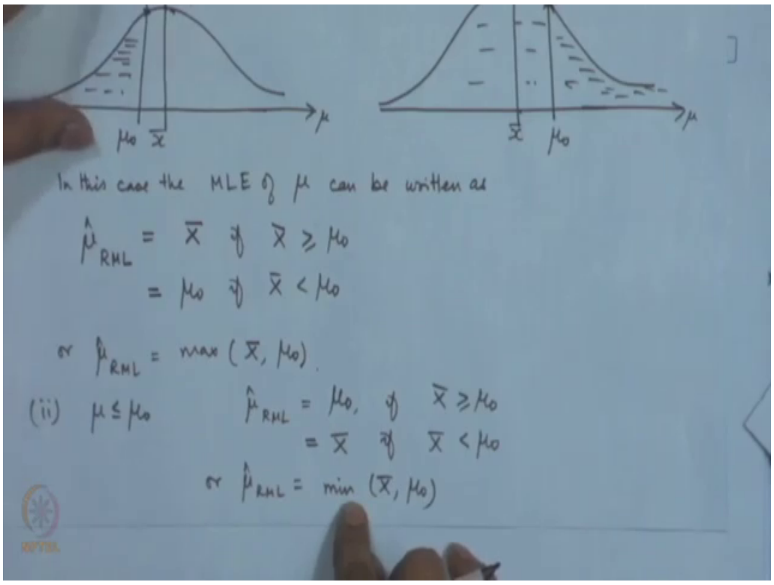
Now, if I use the restriction mu is greater than or equal to mu naught, then there are two cases. Let us make the plot of the likelihood function. On this side we show mu; and on this side we show L mu. So, we may have a situation that say mu naught is here. Now, our parameter spaces mu greater than or equal to mu naught. So, if you see it carefully our region of consideration is on the right side of this x equal to mu naught, this mu is equal to mu naught, now the maximum value that mu is equal to x bar that is occurring within this region. So, the maximum likelihood estimator for mu is still remains x bar.

Let us look at the other case suppose mu naught is on the right side here. Now, there is a problem mu is greater than or equal to mu naught. So, our region of maximization is only this; now in this region if you see the likelihood function is decreasing the maximum value is attained at mu naught. Therefore, your formal maximum likelihood estimator has got modify.

So, we in this case the maximum likelihood estimator of mu can be written as mu hat let me put RML just to denote a restriction that is equal to X bar, if X bar is greater than or equal to mu naught; and it is equal to mu naught if x bar is less than mu naught. Or, we can also expressed in this version that mu hat RML is equal to maximum of X bar and mu naught.

So, immediately you can notice that it has got changed from the original maximum likelihood estimator. And therefore, it is certainly different from the method of moments estimator also. Because, this procedure takes care of the exact parameter space where the maximization problem is solved which is not true in the method of moments estimator. I will consider other type of restriction for this problem.

(Refer Slide Time: 23:40)



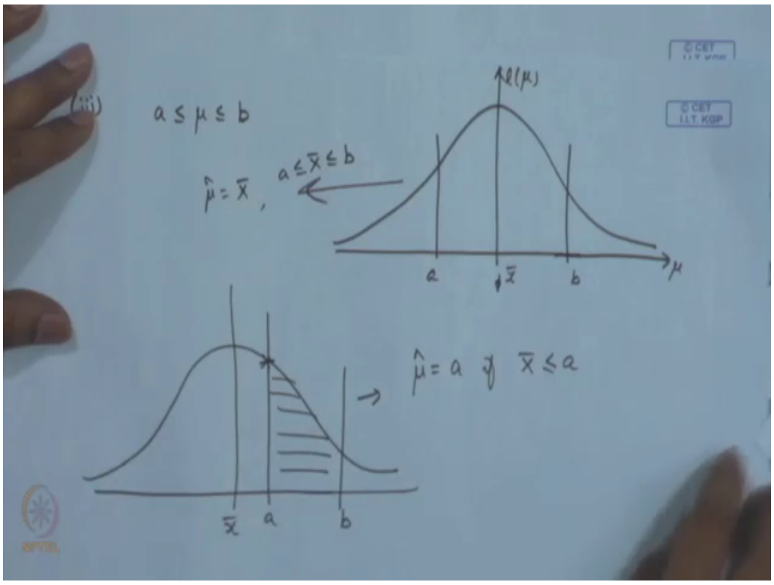
So, let us takes a mu less than or equal to mu naught. Now, if you take mu less than or mu naught, we can go back to the same graph and see this. If mu is less than or equal to mu naught and mu naught is in this position then our region of maximization is here. Therefore, the maximum value is occurring at mu naught; that means, I will say that mu hat RML, it is equal to mu naught if X bar is greater than or equal to mu naught and it is equal to.

Now, in this case, if you see if mu naught is on this side then our region of maximization is this full thing and here the maximum is occurring at x bar. So, it is equal to X bar if X bar is less than mu naught. So, this you can also say in other words as mu hat RML is equal to minimum of X bar and mu naught.

So, notice here if we have the full region we get X bar as the maximum likelihood estimator for mu in the case of estimating the mean of a normal distribution when the variance is known. When there are certain restrictions like a lower bound placed or an upper bound placed for the parameter mu, then accordingly the maximum likelihood estimator gets modified. In this case, it is becoming maximum of X bar and mu naught; and in this case, it is becoming minimum of X bar and mu naught.

Let me take another kind of restriction. In many of the practical problems it may happen that the mean mu lies between two values. For example, you look at the average income levels, you look at the average rainfall, you look at the average weight, average height, so there are various parameters which occurring the practical situations which are actually bounded in nature. They are not unbounded that means, we cannot say that they take values from minus infinity to infinity. So, when that information is available to us in that case we should utilize that, and our estimator should reflect that.

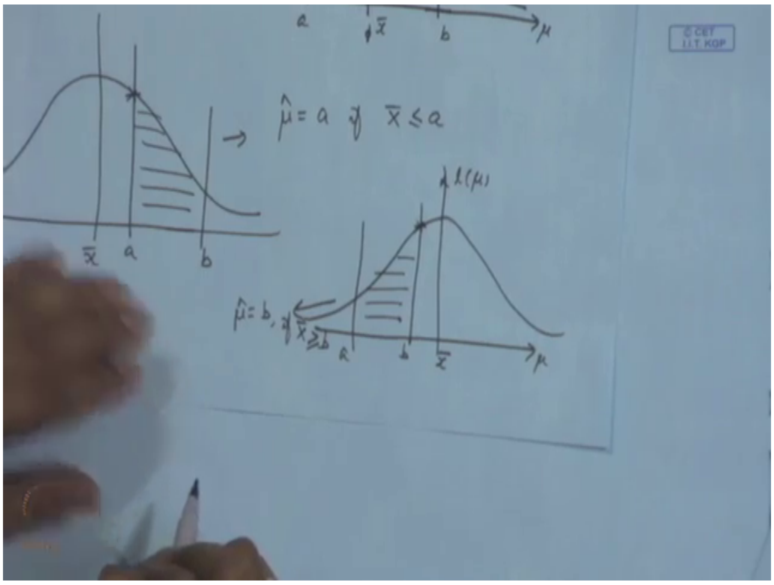
(Refer Slide Time: 26:05)



That means, let me take the third restriction of this nature that say a is less than or equal to mu is less than or equal to b. Now, this is even more interesting we look at the likelihood function as we have plotted in this particular case. So, if your a and b is for example, containing x bar that is x bar lies between a to b then the maximum occurs as usual at x bar.

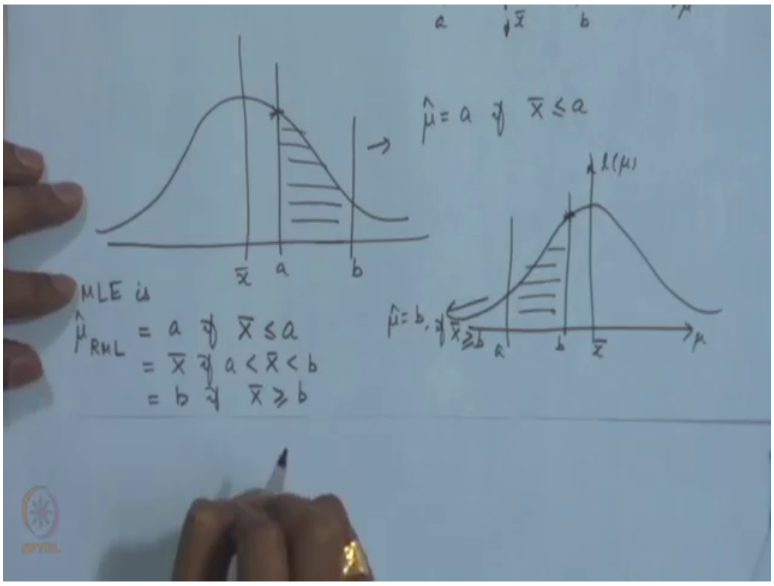
However, you could have had other kind of situations. So, in this case in this case mu hat is equal to X bar that means, when X bar is lying between a to b. You consider another situation for example, a and b are here. If a and b are here, then we have to look at the maximum of the likelihood function within this region alone and obviously the maximum occurs at a.

(Refer Slide Time: 27:46)



So, in this particular case then the maximum likelihood estimator is becoming a if X bar is less than a. And a similar situation would occur if we consider say a and b are to the left of x bar, in this case our maximization problem is restricted to this region. And if you see, the maximum is occurring at b. So, in this particular case then mu hat will become equal to b, that means, if X bar is greater than or equal to b.

(Refer Slide Time: 28:27)



Therefore our solution for the full problem of new lying between a to b is that mu hat RML it is equal to a if X bar is less than or equal to a. It is equal to X bar if a is less than

X bar less than b. And it is equal to b if X bar is greater than or equal to b. So, if there is any prior information about the parameter, the method of maximum likelihood estimation takes care of that.

So, thank you today.