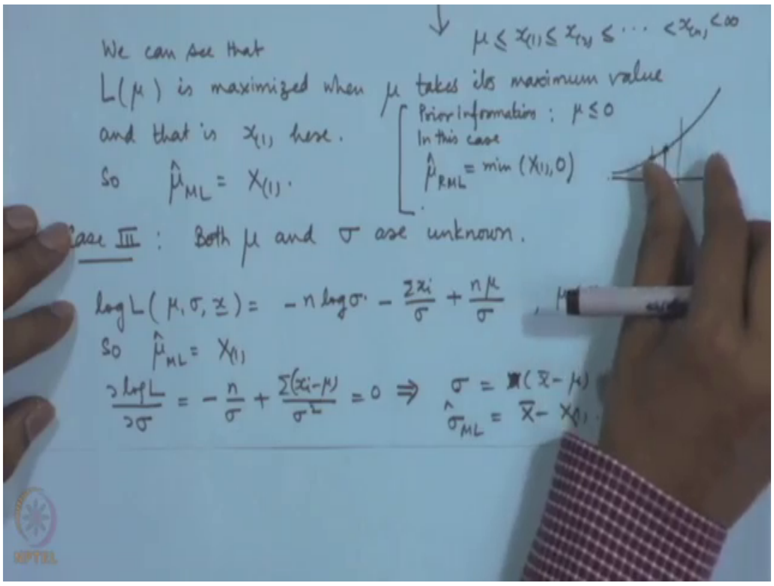
Statistical Inference
Prof. Somesh Kumar
Department of Mathematics
Indian Institute of Technology, Kharagpur

Lecture – 12
Finding Estimators - VI

(Refer Slide Time: 00:21)



Now, let us take the more important case when both the parameters mu and sigma r unknown. Now, let us go back to the original likelihood function, it was 1 by sigma to the power n e to the power minus sigma x i minus mu by sigma. So, we consider the now this is having two parts; one part is involving only mu and other part is involving sigma also. So, for come convenience we take the log of this, so log of likelihood function that is equal to minus n log sigma minus sigma x i by sigma plus n mu by sigma.

Now, you can see here the role of mu is quite different. And when we consider the maximization with respect to mu, it will be attained at the maximum value of mu. So, we can easily then see that as before the maximum value that it can take is so mu head ML will remain to be X 1. However, for maximization with respect to sigma, we can apply the usual calculus here. So, you can consider derivative with respect to sigma that will be equal to minus n by sigma plus sigma x i by so, we may actually put it together because this was this term. Now, this is equal to 0, if you put this, you get sigma is equal to n times x bar divided by n. So, this n gets cancelled out. So, the maximum likelihood estimator for sigma will be obtained by simply replacing mu by mu head ML.

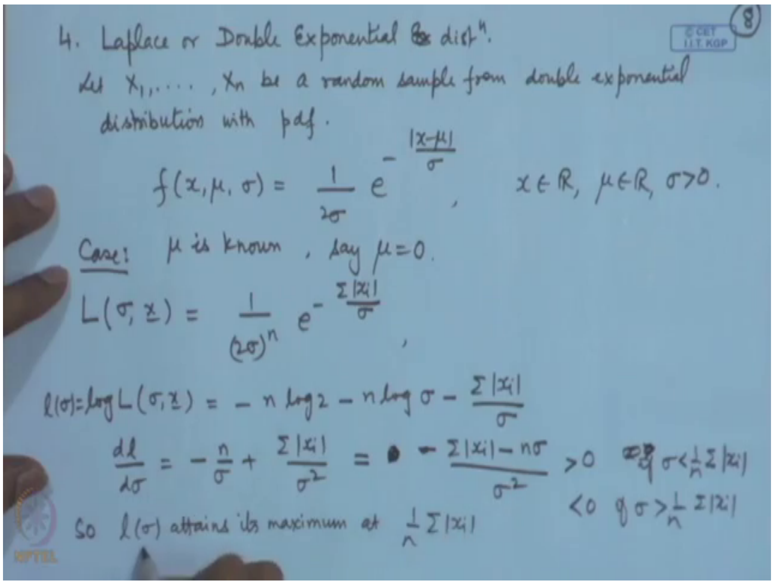
So, sigma head ML is equal to X bar minus X 1. Now, you can see here the effect of partial information and the effect of no information. When the partial information about the parameters was there, then in the case of the estimator of sigma we got X bar, but now you see it is changed to X bar minus X 1. Whereas, the effect on the estimation of mu is not there for when sigma was known or sigma is unknown, the estimation is mu is still the same.

Now, in this case I will also consider some special cases. Here let us consider when sigma was known, suppose have additional prior information, suppose additional prior information about mu is there in the form say mu less than or equal to 0. Basically it means that the minimum guarantee time is upper bounded by some number say mu naught which we have brought down to 0. Now, in this case what will happen? If we look at the form of the likelihood function, this function is an increasing function; this function is an increasing function of mu it starts from minus infinity that means, it is 0 and then at 0, it will e to the power something and then thereafter.

Now, if you see if x 1 is here, then the maximum is occurring at this point. Whereas, if x 1 is here with respect to 0, then the maximum is occurring here. Then in this case mu head ML which I will call restricted ML, this will become minimum of X 1 and 0. So, the role of prior information is important here. You consider a second situation suppose in place of mu less than or equal to 0, we had mu greater than 0 or greater than or equal to 0, in that case there will be no change because x 1 is greater than or equal to mu it will remain greater than 0. So, the maximum occurrence at x 1 which is within the zone. So, there will not be any change in the maximum likelihood estimator when I am considering the prior information mu greater than or equal to 0.

So, you can actually see that the role of the prior information is different in different situations and this is the you can say beauty of the maximum likelihood procedure that it takes care of each situation individually. So, this is totally based on the likelihood function.

(Refer Slide Time: 06:20)

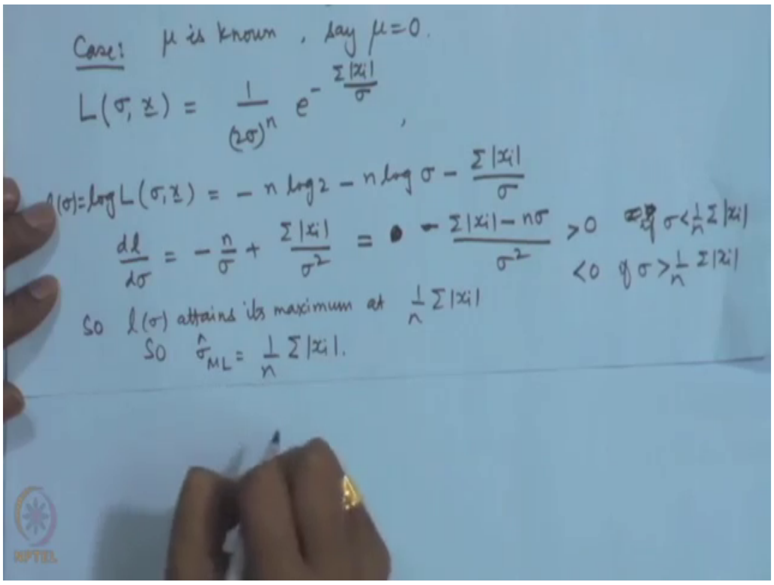


Now, let us consider another important distribution which is known as Laplace or double exponential distribution. So, let X 1, X 2, X n be a random sample from double exponential distribution with the probability density function. Here x is any real number; the parameter mu is any real number and sigma is a positive parameter. As before we may have different situations like mu may be known, so we may put it to be 0; when sigma may be known and we may put it to be 1 etcetera.

So, let us consider the case when say mu is known say mu is equal to 0. So, in this case the likelihood function is 1 by 2 sigma to the power n e to the power minus sigma modulus x i by sigma. So, the log likelihood function that is equal to minus n log 2 minus n log sigma minus sigma modulus x i by sigma. So, if we consider dl by d sigma that is equal to minus n by sigma plus sigma modulus x i by sigma square.

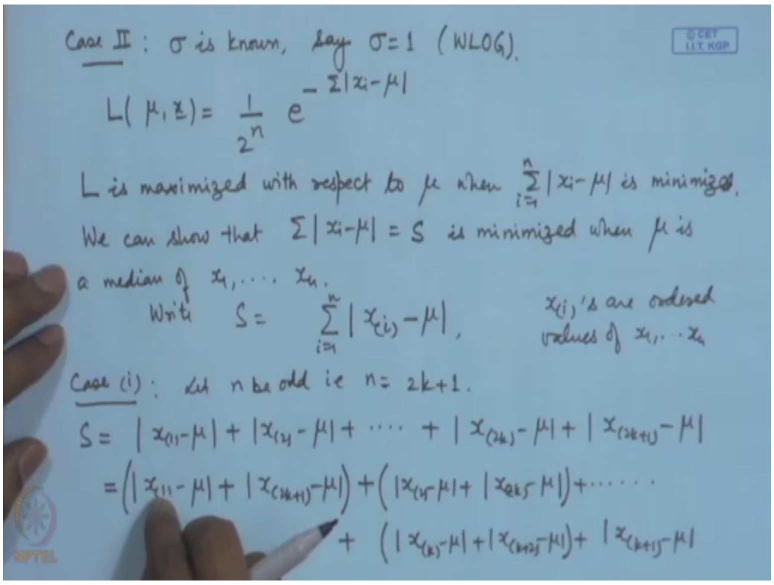
If you put this equal to 0, of course, you can adjust the term this is equal to sigma modulus x i minus n sigma by sigma square as you can easily see that it is greater than 0. If sigma is greater than if sigma is less than 1 by n sigma modulus x i and it is less than 0 if sigma is greater than 1 by n sigma modulus x i. So, the maximum occurs at 1 by n sigma modulus x i. So, l sigma attains it is maximum at 1 by n sigma modulus x i.

(Refer Slide Time: 09:44)



So, the maximum likelihood estimator of equal to 1 by n sigma modulus x i.

(Refer Slide Time: 09:58)



Let us take the second case when sigma is known. And once again since sigma is a scale parameter, we may take it to be 1 without laws of generality. In this case the likelihood function is equal to 1 by 2 to the power n e to the power minus sigma modulus x i minus mu. Now, you see here, this will be maximized with respect to mu if sigma of modulus x i minus mu is minimized. L is maximized with respect to mu, when sigma of modulus x i minus mu is minimized.

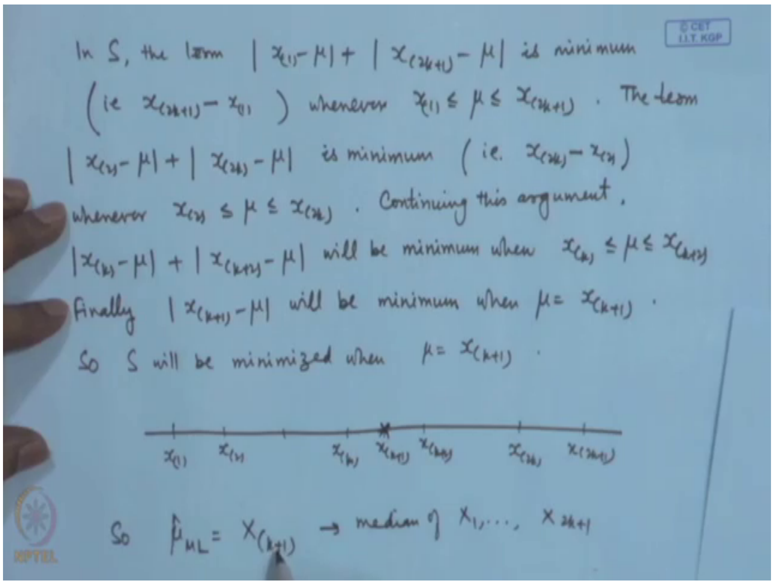
Now, one can show that this is minimized when mu is the median of the observations, because, this modulus term is coming, therefore, you cannot use the usual differentiation procedure here. However, we can give a direct argument, we can show here that sigma modulus of x i minus mu let me call it S is minimized when mu is a median of x 1, x 2, x n.

Let me consider two cases. So, we write S as a sigma and in place of the x i's we consider the ordered x i s ordered value of x 1, x 2, x n, that means, x 1 is the minimum x 2 is a second minimum and so on as before. Now, we give argument in two cases let us take say n that is say n is equal to something like 2 k plus 1. Now, this some S, we place like this x 1 minus mu plus x 2 minus mu plus and so on x 2 k minus mu plus x 2 k plus 1 minus mu.

This we express as say x 1 minus mu plus x 2 k plus 1 minus mu that means, I have taken the first term and the last term. Then I take the second term and the second last term x 2 minus mu and x 2 k minus mu and so on that is finally, we will have x k minus mu minus x k plus 2 minus mu. And the last term then will be remaining that is x k plus 1 minus mu.

What we do, we look at the minimization of each of these terms which are clubbed together. So, if you look at these two, here it is the x 1 and this is x 2 k plus 1. If I consider mu to be any value between these two, then this will turn out to be x 2 k plus 1 minus x 1 that will be the minimum value.

(Refer Slide Time: 14:24)



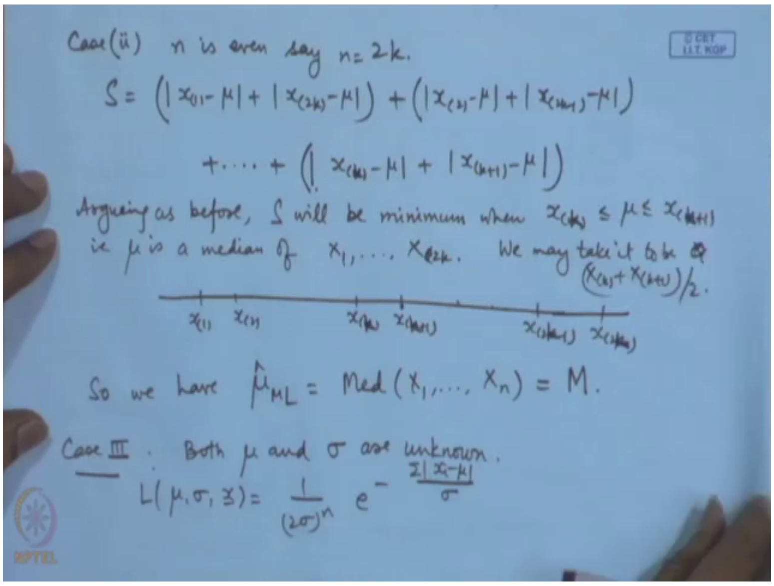
So, let us write the complete argument here. In S, the term x 1 minus mu plus x 2 k plus 1 minus mu is minimum that is the value will be x 2 k plus 1 minus x 1, whenever I choose mu to be a number between x 1 and x 2 k plus 1. Similarly, the term x 2 minus mu plus x 2 k minus mu; this is minimum and of course, the minimum value will be x 2 k minus x 2, whenever x 2 is less than or equal to mu less than or equal to x 2 k. So, in that way, if you look at all the sums, they will be minimum, whenever mu lies between the two values which are involved in those two terms.

So, if we continue this argument, the term x k minus mu plus x k plus 2 minus mu will be minimum, when x k is less than or equal to mu less than or equal to x k plus 2. Finally, x k plus 1 minus mu will be minimum, when mu is equal to x k plus 1. We have considered the term by term minimization of this S. So, we have taken these this and this together, then these two together and so on. We have derived the condition for the minimization of each of this.

Now, therefore, the overall minimum will be attained if all the conditions are simultaneously satisfied. Now, if you see all the conditions to be simultaneously satisfied, what will be the condition. This is the widest interval because this is from minimum to the maximum. This interval is the second and so on. So, if I look at this scale here x 1, x 2, x 2 k, x 2 k plus 1 somewhere you have x k x k plus 1 and x k plus 2. So, from the 1st one, mu should be any value between these two; from the 2nd one, mu should be any value between these two; from the 3rd one and so on and finally you are getting the value that is x k plus 1.

So, if mu is x k plus 1, each of these terms that I have clubbed together, they will be minimum. Therefore, overall S will be minimized. So, S will be minimized when mu is equal to x k plus 1, because this will satisfy all the conditions. So, we conclude that mu head ML is equal to X k plus 1 that is actually the median of X 1, X 2, X 2 k plus 1 because when the number of observations is odd, the middle value will be the median here.

(Refer Slide Time: 19:12)



Now, let us consider the case when n may be even; n is even say n is equal to 2 k. Now, in this case once again we may consider the clubbing in the similar fashion. However, this last term will not be there. Therefore, we will write the clubbing in this fashion x 1 minus mu plus x 2 k minus mu plus x 2 minus mu plus x 2 k minus 1 minus mu and so on. In the final, it will be x n minus mu plus x sorry x k minus mu and x k plus 1 minus mu.

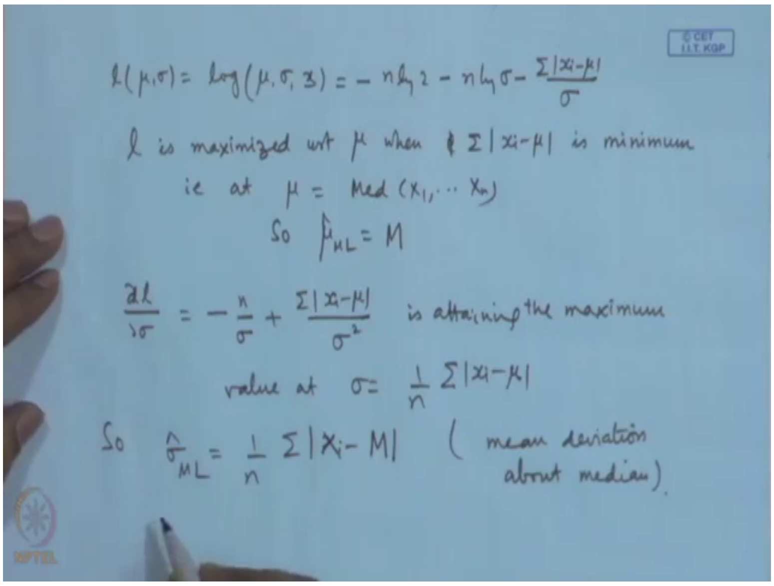
So, if we give the argument as before; arguing as before, S will be minimum when x m is less than or equal to mu less than or equal to x m plus 1. Because, now on a scale x 1, x 2, x n, x n plus 1, x 2 m minus 1 x 2 m, so here this first sorry this will be k here. So, here if you see the first term will be minimum when mu lies between the largest interval;

the second one will be minimum when the mu lies between x 2 to x 2 k minus 1 and so on. The last sum will be minimized when mu lies between x k to x k plus 1.

Now, if mu lies between x k to x k plus 1 when we have even number of observations that is x 1, x 2, x 2 k any number between x k to x k plus 1 is called a median. For convenience many times we take the average of these two value that is x k plus x k plus 1 by 2. So, this we conclude that mu is a median of X 1, X 2, X 2 k, so where we may take it to be X k plus X k plus 1 by 2. So, we have mu head ML equal to the median of X 1, X 2, X n. In both the cases, we are getting median; we denote it by say M.

So, now let us consider the important case. When both the parameters may be unknown, so both mu and sigma are unknown, in this case the likelihood function is equal to 1 by 2 sigma to the power n e to the power minus sigma modulus x i minus mu by sigma.

(Refer Slide Time: 23:17)



So, we take the log here that is equal to minus n log 2 minus n log sigma minus sigma x i minus mu by sigma. So, as before the maximization with respect to mu will occur when sigma of modulus x i minus mu is minimum. And we have already shown that this is occurring when mu is a median. So, l is maximized with respect to mu, when sigma of modulus x i minus mu is minimized that is at mu equal to median of X 1, X 2, X n. So, mu head ML is equal to the median which we are calling M.

Now, you look at the solution for sigma. If we consider the derivative of l with respect to sigma, we get minus n by sigma plus sigma modulus x i minus my by sigma square. And as before if we argue this is attaining the maximum value at sigma is equal to 1 by n sigma modulus x i minus mu. Now, we have already obtained the solution for mu. If we substitute it here you get the maximum value of the maximized value of likelihood function for sigma equal to 1 by n sigma modulus of x i minus n. So, sigma head ML is equal to 1 by n sigma modulus X i minus M which is nothing but the mean deviation about median mean deviation about median.

So, today friends we have discussed various probability models and we have discussed the maximum likelihood estimators for those models, I have tried to cover various cases here. And another thing is that we will take up some different cases where either the maximum likelihood estimator is not unique, it may not exist. And then we will consider the large sample properties of the maximum likelihood estimators in the next class.