Statistical Inference

Prof. Somesh Kumar

Department of Mathematics

Indian Institute of Technology, Kharagpur

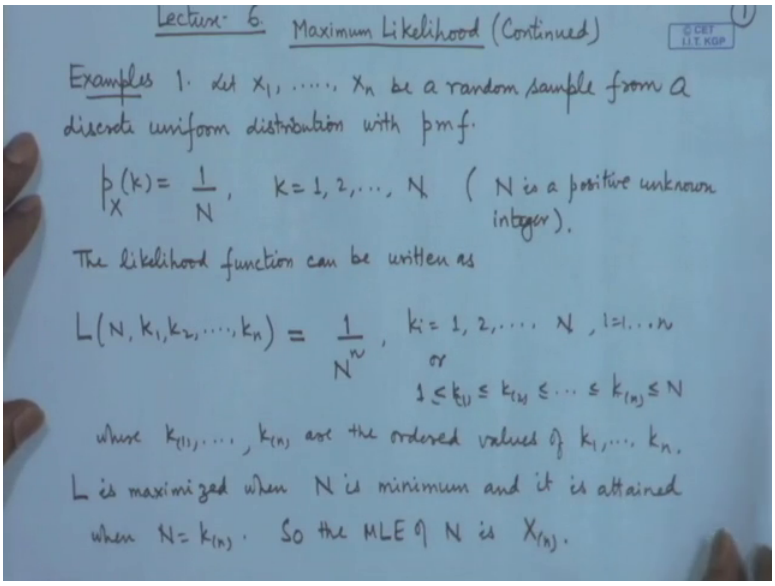
Module No. # 01

Lecture No. # 06

Finding Estimators – III

Yesterday we have discussed in detail various probability models, and how to find out themaximum likelihood estimators for that. We have seen here that the effect of changing the parameter space, or effect of the prior information on the parameter space plays an important role in the maximum likelihood estimation, which makes it different from the other methods, such as unbiased estimation or the method of moment's estimation. So, today I will explain this method with the help of several other examples, and we will discuss certain importantlarge sample properties of the maximum likelihood estimators.

(Refer Slide Time: 01:04)

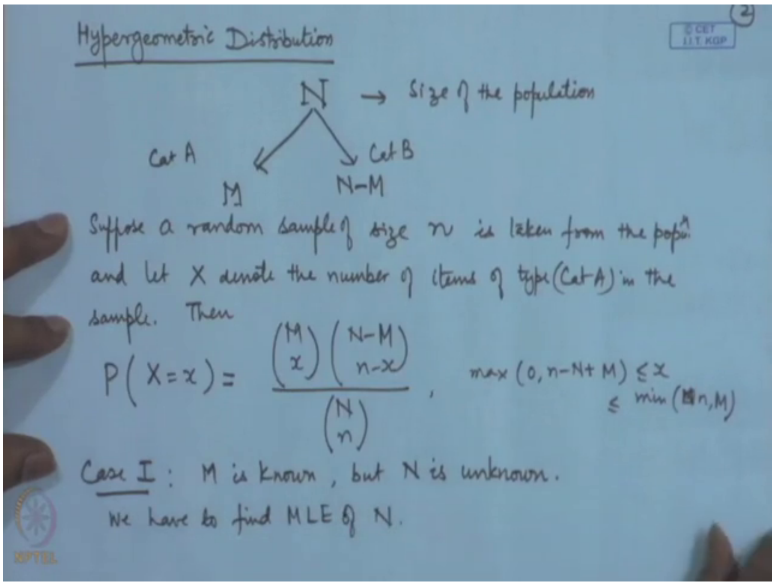


Let, me start with the, a couple of examples on discreet distributions. So, let us consider, say a discrete uniform distribution.Let x 1 x 2 x n be a random sample, from a discreet uniform distribution. So, a discreet uniform distribution is usually concentrated on n points, and normally we take the points from 1 to n, and each one will be equal probability. So, we can consider the probability mass function as follows, with probabilitymass function given by. So, we write p x k is equal to 1 by n,where k can take values 1 2N. Now, in this case there may not be any inference problem, if we know on how many points the distribution is concentrated. The inference problem arises if we do not know how many points are there. So, this type of situation may arise, where we know that each possibilities with equal probability, but how many possibilities are there that may not be known. So in that case, we may be interested in estimating that number. So, we are assuming here that n is a positive unknown integer. So, we proceed asbefore, we write down the likelihood function, which is the joint distribution of x 1 x 2 x n.

So, we consider points x 1 is equal to k 1, x 2 is equal to k 2, x n is equal to k n. So, we can write it in the following fashion. The likelihood function can be written as L N, and as I mentioned, we are considering the points k 1 k 2 k n, which are the observed values of the random variables x 1 x 2 x n respectively. So, that is equal to 1 by N to the power n, where each of the k i's can take values 1 2 N or i is equal to 1 to N. Now, the problem here is to, maximize this function with respect to n. As n is appearing at the denominator, it will be the minimum value of n. So, this will be maximized when n is taking the minimum value. Now, what is the minimum value of n that is possible here. So, this reason we can write it in a more appropriate fashion; that k 1, suppose i order then, k 2 up to k n, then the reason can be written as; one less than or equal to k 1, less than or equal to k 2, less than or equal to k n, less than or equal to N.

So, from here it is clear that the minimum value of n, that is possible is the maximum value of k 1 k 2 k n, wherek 1 k 2 k n are the ordered values of k 1 k 2 k n. So, L is maximized, when N is minimum and it is attained, when N is equal to k n. Now k n corresponds to the largest order statistics here. So, we conclude that the maximum likelihood estimator of N is x n. You can notice the analogy with the continuous uniform distribution, which we discussed in the previous class. In the continuous uniformed distribution on the interval zero to theta, the maximum likelihood estimator for theta was also the largest order statistics, that is x n. So, in the discreet uniform case also the same thing is happening. The only difference here is that, here x i's are taking positive integral values here.

(Refer Slide Time: 06:35)



Let us take another important discreet distribution; that is hyper geometric distribution. Now a hyper geometric distribution is usually considered in the following fashion, that there is a large populationof size n. This is the size of the population. Now this population is divided into two parts. Let us say category A and category B. The entire population,for example we may divide, aemployees of an organization by two categories that is those who are in the supervisory position and those who are in theworkingconditions, and that is the,they are the lower level employees and the higher level employees.

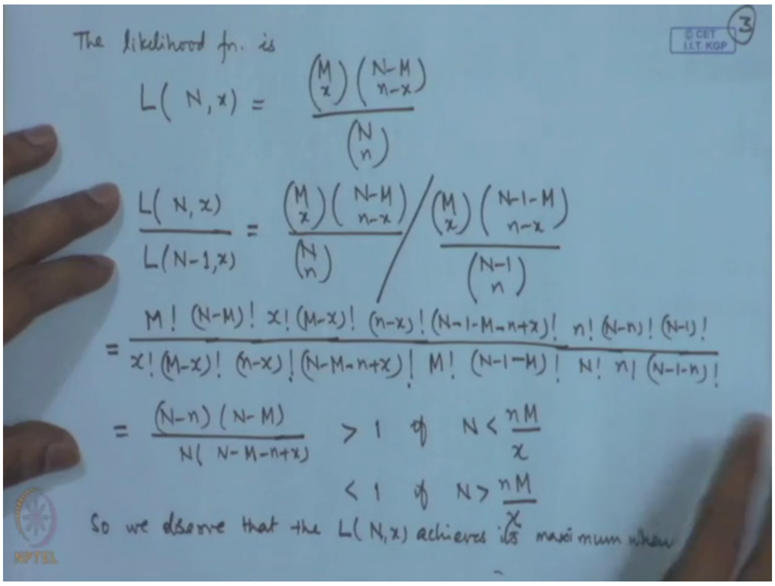
We may divide the patients into two groups; say those who are havingcommunicable diseases, those who do not have communicable diseases. We may dividesection of a student in to the students, who are following engineering discipline and the others who are studying saymedical discipline. So, we have a large population, and the population size of one category is M, and therefore the other category population has N minus M numbers. Suppose, we take a random sample a random sample of size small n, is taken from the population. And let x denote the number of items,items means it could be persons or anything, of type category A in the sample. Then the probability distribution of x is given by M c x,N minus M c n minus x divided by N c n.

Now obviously, this random variable x, it can take values from 0 1 to N, because in a random sample of size n, you may have none of this category and all of the other category, 1 of 1 category, n minus 1 of another category and so on. However, this is also subject to the restrictions of the total elements of each type, and therefore we may write the restrictions in a more strict sense as; that is x is a integer between maximum of 0 and n minus N plus M to minimum of n M. Now, when we look at this probability model, there can be two different cases; one case could be, that the total population size is unknown. Now this type of situation arises for example, in estimatingsay, we have a lake and a company which is involved in the fishing. It may like to estimate that how much of fish amount will be available in the lake, if they start the fishing operations. Now; obviously, one cannot take out the water from the lake, and count the how many fish will be there.

So, we assume that the size of the population that is capital N is unknown. Now one may conduct the following experiment, which is known as capture recapture technique. We take a random sample of size capital N from the lake. The fish that are taken out they are tagged; that means, they are marked with something, then they are shifted back to the lake. So, that they get mixed up with the entire population of the fish. Later on we consider a random sample of size n, from thefish once again; from the lake we again take a random sample of size n. Now out of that you look at how many of them are tagged, and how many of them are untagged. So, now, this capital M is known to you andcapital N is not known to us, and the problem will come how to estimate capital N.

Similarly, there can be another problem, where the total population size is known, we may like to estimate how many people are suffering from a certain diseases or a certain virus.For example, how many people are infected with H I Vvirus. In that case, we again take a sample of size n. And in that sample, x will denote the number of people who are actually infected with the virus, and then on the basis of that we estimate N. So, in this case capital N may be known, but capital M is unknown. So, when we consider this hyper geometric model, there are two cases. So, case one is that M is known, but N is unknown. So in this case, we have to find the maximum likelihood estimator of N. In order to do that, we write the likelihood function. Now in this case the observation is, thesample of size n has been taken, and x is the number of items of type category A. So, this is the recorded item. So, this function itself denotes the likelihood function in this particular case, because this is the probability mass function of the observation here.

(Refer Slide Time: 13:50)

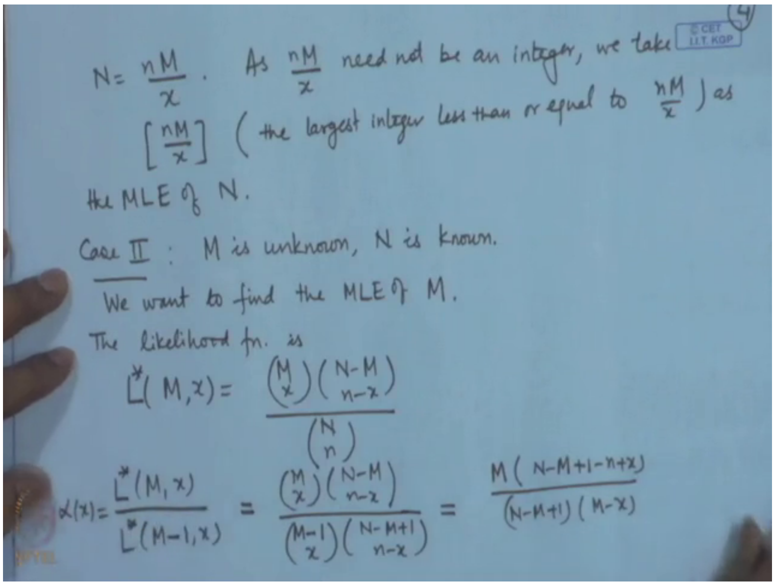


So, the likelihood function is, let me call it LN here. So, that is equal to M c x,N minus M c n minus x,N c n. And we need to maximize this with respect to capital N. Now the methods that I mentioned in the previous examples, cannot be directly implemented here. The main reason is that here n is an integer, so we cannot apply differentiation procedure taking log etcetera. So, we carry out a different analysis. Let us write down, we try to see the increasing or decreasing nature of this function in a straightforward fashion. Let us consider for example, the value of the likelihood function at N, and the value of the likelihood function at N minus 1. So, this is M c x, c,N minus M c n minusx divided by N c n. And then this whole thing is divided by M c x,N minus 1 minus M c n minus x divided by N minus 1 c n. We may expand the factorial here, so we will get M factorial divided by x factorial.

So, this entire thingturns out to be like M factorial divided by x factorial,M minus x factorial, then we have N minus M factorial, divided by n minus x factorial, then N minus M minus n plus x factorial. This whole thing is then divided by these terms, so M factorial, x factorial into M minus xfactorial, then we have N minus 1 minus M factorial, n minus x factorial, and then N minus 1 minus M minus n plus x factorial, then further we have n minus, and then we had this N c n and N minus 1 c n. So, we write that also,N factorial, n factorial N minus n factorial. And in the similar way this will be N minus 1 factorial, n factorial,N minus 1 minus n factorial. So, it is easy that one can simplify these terms, and we get it as N minus n, into N minus M divided by N into N minus Mminusn plus x. Now you notice that, this is greater than 1 if N is less than n M by x, and it is less than 1 if N is greater than n M by x.

Now obviously, you can see N is taking integer values from 1 2 and so on.Now this ratio; that is LN x, divided by LN minus 1 x. So, what we are observing here is that, if i increase N. From N minus 1 to n if I go, then this ratio is greater; that means,it is an increasing function of N, when N is less than n M by x. Andwhen N is bigger than n M by x, then this value starts decreasing.Therefore, you can say that this function increases till this and then decreases,therefore the maximum of L N function is achieved when N is equal to n M by x. Now, naturally n M by x need not be an integer, although x n and M are integers, but this expression need not be an integer. So, we may take the integral portion of n M by x as the, maximum likelihood estimator for N.

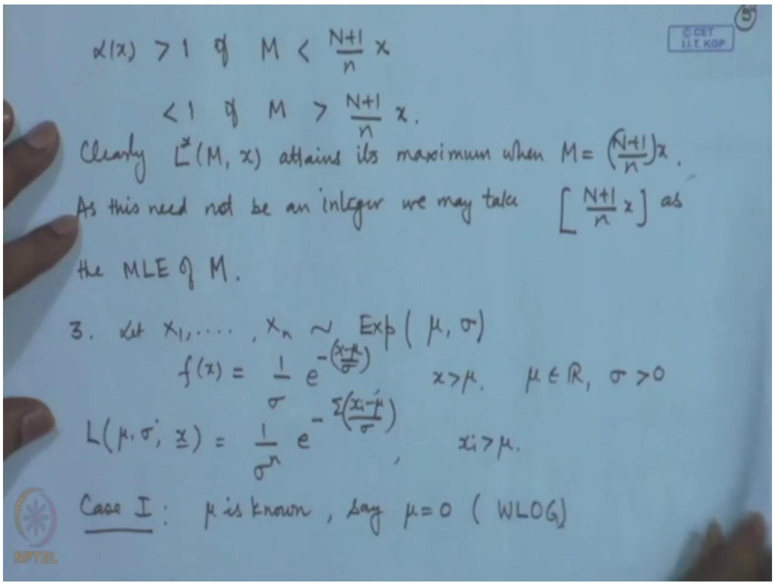
(Refer Slide Time: 19:08)



So, we observe that, the L function achieves its maximum, when N is equal to n M by x. As n M by x need not be an integer, we take n M by x integral portion; that is the largest integer, less than or equal to n M by x, as the maximum likelihood estimator of N. Now, let us take up the other case, when M is unknown,M is unknown and N is known. So, here we want to find out, the maximum likelihood estimator of M.Now, once again, if you consider this likelihood function here. I wrote here it as a function of N, because this is coming from the probability mass function of x. Here M and N both are involved. Now, if N is known and M is unknown, I will consider the likelihood function as a function of N. So, the likelihood function will become. Although, it will be the same expression, it will be written as L M x. Let me call it L star.

So, this is M c x,N minus M c n minus x. Now as before, we have to consider the maximization of this, with respect to M. now M is an integer and the factorials are involved here,therefore 1 cannot apply the usual methods of analysis; such as differentiation etcetera, rather we try to see the behavior of this in a straightforward fashion. So, once again we write L star M x divided by L star M minus 1 x. Now that is equal to M c x,N minus M c n minus x, when we write this ratio N c n will be same, so that will cancel out, and we will get M minus 1 c x,N minus M plus 1 c n minus x. Now as before, we can simplify this, and the term turns out to be M into N minus M plus 1 minus n plus x, divided by N minus M plus 1 into M minus x. Now, once again we observed that this ratio, let me call itsay alpha x.

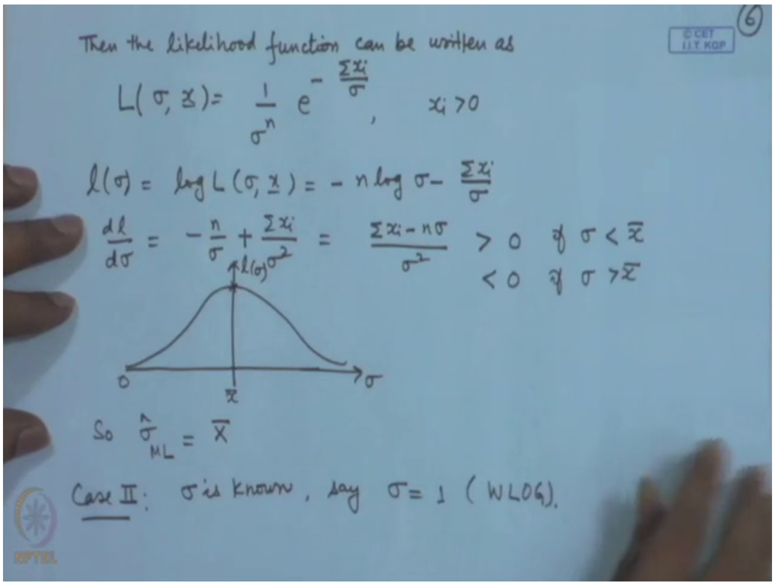
(Refer Slide Time: 22:45)



So, if we observe this ratio, alpha x is greater than 1 if M is less than N plus 1 by n x, and it is less than 1 if M is greater than N plus 1 by n x. So, we can easily see that, the L star function, it is increasing for M less than N plus 1 by n x, and it will start decreasing for M greater than this. Therefore, the maximum will be attained at N plus 1 by n x, and therefore we can consider the integral part of this, as the maximum likelihood estimator for N. Clearly L star M attains its maximum, when M is equal to N plus 1 by n x. As this need not be aninteger, we may take the integral portion of this as the MLE of M. So, here we have seen that, in the discreet case the method of obtaining the maximum likelihood estimators differs little bit. We have not considered another important distribution which arises quite often instatistical modeling; that is a exponential distribution. Now, the exponential distribution once again has two parameters. It may have a scale parameter; it may have a location parameter. So, I will consider a general model, and then we look at thesolution here. Let x 1 x 2 x n followexponential mu sigma distribution, when I say this we are writing down the density function as, 1 by sigma e to the power minus x minus mu by sigma, where x is greater than mu.

Here mu can be any real number and sigma is positive. In the usual study which are related to reliability and life testing, there mu is considered as the minimum guarantee time and there mu will be positive, but in many other applications it need not beso. So I am taking the general case where mu can take any real value, and sigma of course, is associated with the average,therefore sigma is greater than 0. So, we consider the likelihood function here, 1 by sigma to the power n e to the power minus sigma x i minus mu by. Now when we are dealing with the two parameters situation; one may have different cases. It may happen that the minimum guarantee time is fixed, and therefore we may take it to be 0. It may happen that sigma is fixed, and therefore we may take it to be one. So, we consider these cases. So, case 1; let us consider say mu is known, so we may take without loss of generality, this to be 0. If that is so, then we may write the likelihood function. If we substitute mu is equal to 0, the form of this function becomes much simple.

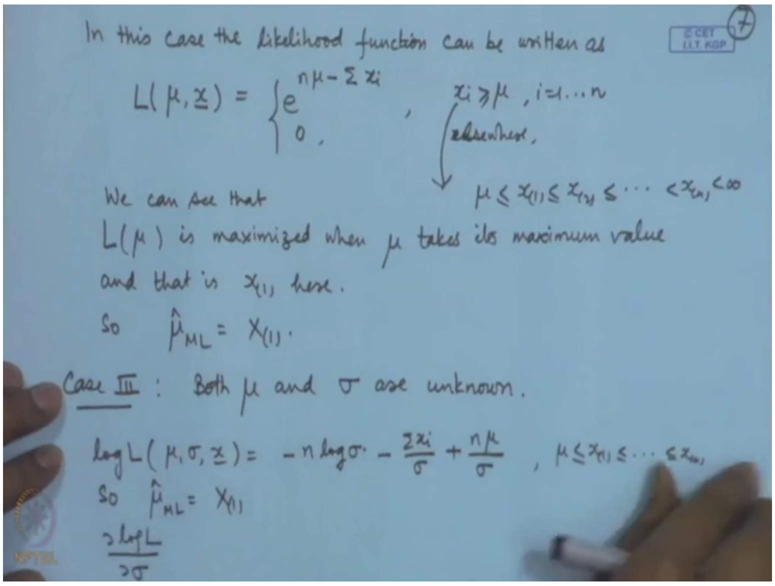
(Refer Slide Time: 27:07)



And we get it as, then the likelihood function can be written as,L sigma x, as 1 by sigma to the power n, e to the power minus sigma x i by sigma, where each x i will be greater than zero. So, we write down the log likelihood function that is equal to minus n log of sigma, minus sigma x I by sigma. So, now, this is a straightforward function for sigma, we can consider the derivative with respect to sigma, and we get minus n by sigma minus sigma x i. So, this will become plus sigma x i by sigma square, which gives us sigma x i minus n sigma by sigma square; obviously, you can study its behavior. It will be greater than 0 if sigma is less than x bar. It will be less than 0 if sigma is greater than x bar. So, if we consider the, plotting of the curve as a function ofsigma. If we plot L sigma, now sigma is of course positive, so this is starting from zero.

So, this is increasing till x bar and thereafter it is decreasing, because our derivative is positive, for sigma less thanx bar and it is less than 0, for sigma greater than x bar.Therefore, easily you can see that the maximum occurs at x bar. So, the maximum likelihood estimator of sigma turns out to be the mean of the distribution. Now, here as beforelike we have considered in the normal distribution, one may have additional information about sigma.For example, sigma may be having an upper bound; such as sigma less than or equal to sigma naught, or sigma greater than or equal to sigma naught or sigma may lie in an interval. In that case, the solutions will, for the maximum likelihood estimator will get modified accordingly, as we have discussed in the case of normal distribution. So, I will be skipping thosedescriptions here. Let us take up the second case, whensigma is known when sigma is known we cantake it to be one without loss of generality. Now in this case the likelihood function can be written as.

(Refer Slide Time: 30:17)

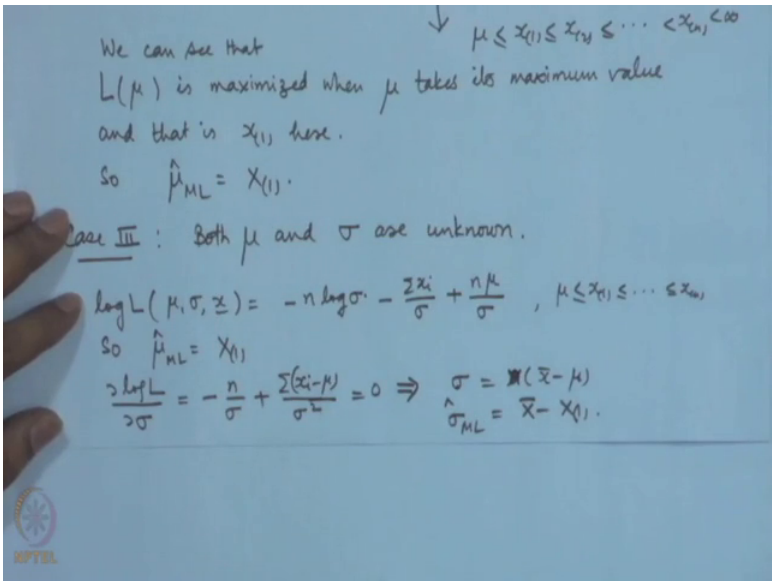


So, this is now a function of mu, because sigma is known. So, if you look at the form that I have discussed here, 1 by sigma to the power n e to the power minus sigma x i minus mu by sigma. So, here if I put sigma is equal to 1 this term vanishes, and you are leftwith only the exponent term, which I can simply write as e to the power n mu, minus sigma x i. So, e to the power n mu minus sigma x i, and of course each x i is greater than mu. And obviously, this is 0 if. Let me say elsewhere, each of x i has to be greater than mu in this particular case. Now, if you look at this function, we have to maximize this with respect to mu here.And this n mu is occurring in the exponent without any multiplication or any other involvement of any other term. So, naturally you can easily see that, the maximization will occur for the maximum value of mu.

Now, what is the maximum possible value of mu. Now, mu is less than each of the x i’s,therefore this reason can be written as mu less than x 1 less than x 2 and so on. Therefore, the maximum value of mu can be only x 1. So, the maximum likelihood estimator of mu is x 1 in this case. We can see that L mu is maximized, when mu takes its maximum value and that is x 1 here. Once again here this x 1 x 2 x n denotes the order statistics of theoriginal observations. So, mu head M L is equal to the minimum of the observations. So, you have seen in the uniformed distribution, we got the maximum of the observations. And in this particular case we are getting the minimum of the observations. Now, let us take the more important case, when both the parameters mu and sigma are unknown.

Now let us go back to theoriginal 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, forconvenience 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 attend 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 M L will remain to be x 1.

(Refer Slide Time: 35:16)

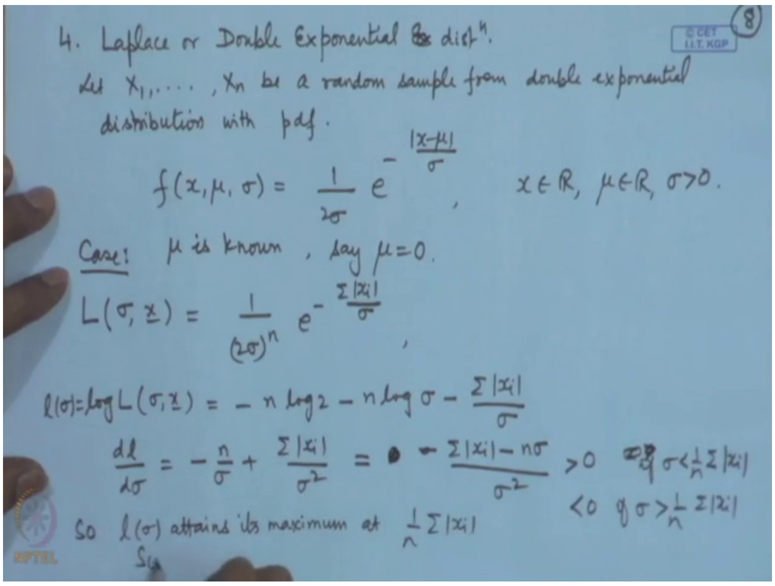


However, for maximization with respect to sigma, we can apply the usual calculus here. So, you can consider derivative with respect to sigma; that will be equal to minus n by sigma, plus sigma x I by. So, we may actually put it together, because this was this term. Now this is equal to 0. If you put 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 m l. So, sigma head M L 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, when sigma was known or sigma is unknown, the estimation of 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, I have additional prior informationabout mu is there, in the form say mu less than or equal to zero. Basically, it means that the minimum guarantee time is upper bounded, by some number say mu naught, which we have brought down to zero. 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 is starts from minus infinity; that means, it is 0, and then at 0 it will be 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.So, in this case mu head M L, which I will call restricted M L.This will become minimum of x 1 and 0.

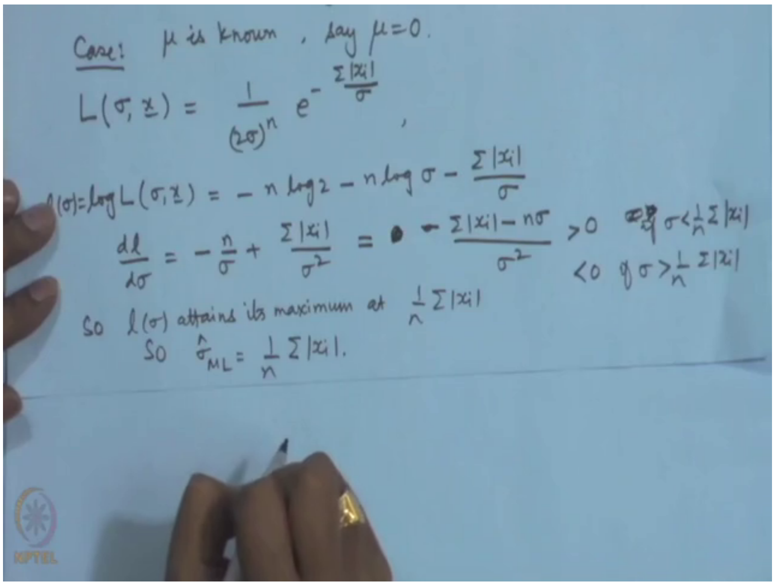
So, the role of prior information is important here. You consider the second situation, suppose in place of mu less than or equal to 0 we had mu greater than 0,or greater than or equal to 0 in that case there will be no change, because x 1 is greater than or equal to mu, which will remain greater than 0. So, the maximum occurrence is 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, andthis is you can say beauty of the maximum likelihood procedure, that it takes care of each situation individually. So, this is totallybased on the likelihood function.

(Refer Slide Time: 39:16)



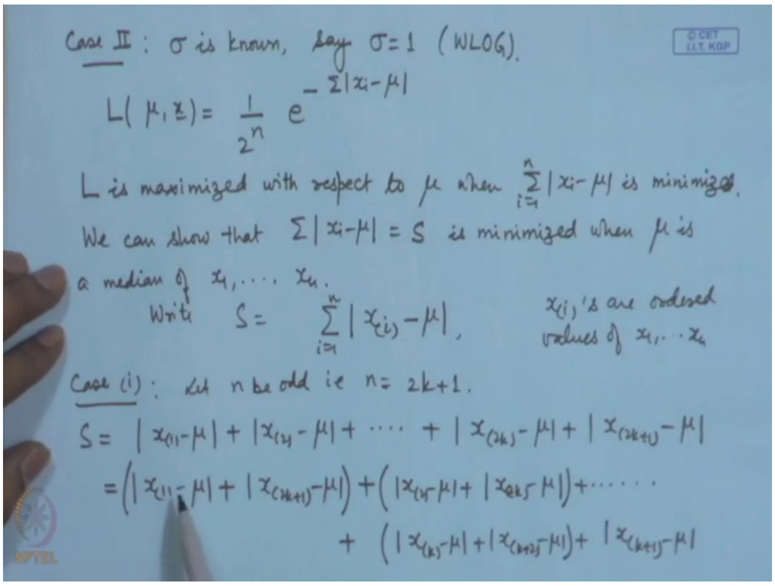
Now, let us consider another important estimation, which is known as Laplace or Double Exponential Distribution.Laplace or Double Exponential Distribution; So, let x 1 x 2 x n be a random samplefrom double exponential distribution, with the probability density function. Here x is any real number, the 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 d l by d sigma that is equal to minus n by sigma, plus sigma modulus x I by sigma square. So, 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. 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.

(Refer Slide Time: 39:16)



So, the maximum occurs at 1 by n sigma modulus x i. So, l sigma attains its maximum at 1 by n sigma modulus x i. So, the maximum likelihood estimator of equal to 1 by n sigma modulus x i.

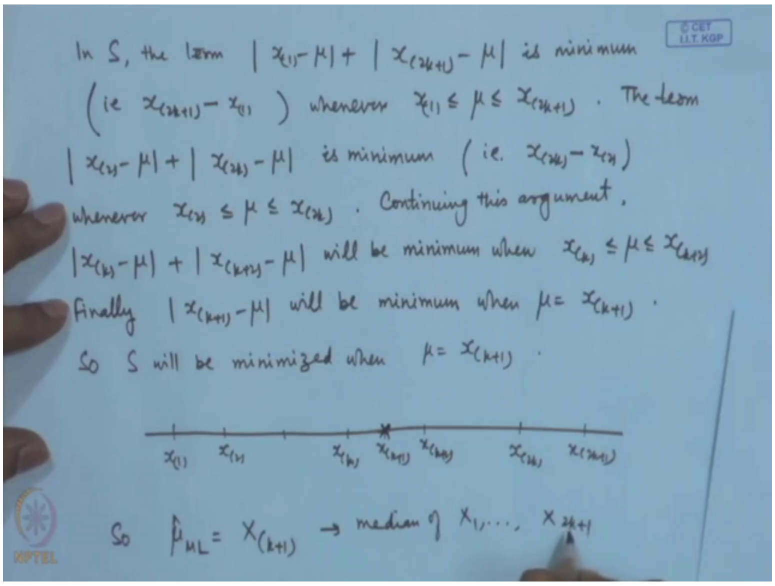
(Refer Slide Time: 43:00)



Let us take the second case, when sigma is knownand once again, since sigma is a scale parameter we may take it to be 1, without loss 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 of x i minus mu is minimized. Now, one can show that, this is minimized when mu is the median of the observations, because thismodulus 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 can considered the ordered x i's, ordered values of x 1 x 2 x n; that means x 1 is the minimum x 2 is the second minimum and so on as before. Now we give argument in two cases. Let us take n; that is n is equal to something like 2 k plus 1. Now, this some S we express like this, x 1 minus mu plus x 2 minus mu plus and so on.x2 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 plus 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 I have clubbed together. So, if you look at these 2. Here it is the x 1 and this is x 2 k plus 1. If I consider mu to be any value between these 2, then this will turn out to be x 2 k plus 1 minus x 1 that will be the minimum value.So, let us write the complete argument here.

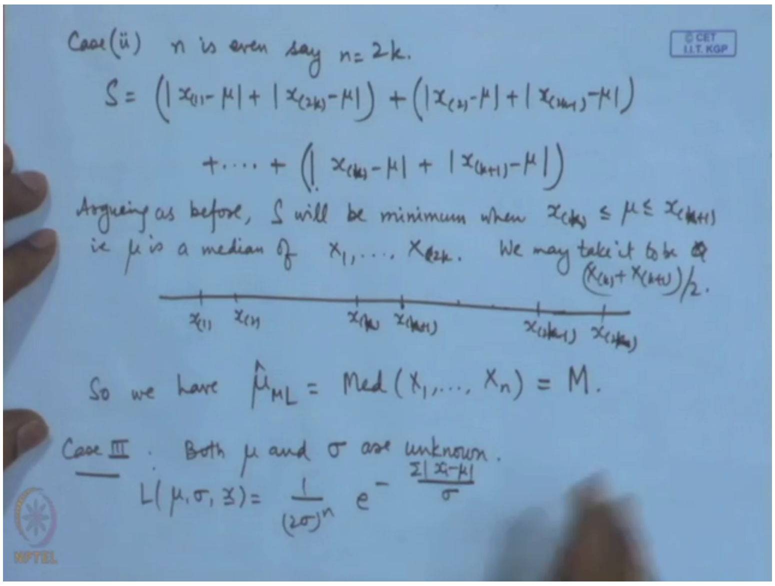
(Refer Slide Time: 47:25)



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 this, this and thistogether then this together and so on. We have derived the condition further, minimization of each of these. 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 first one, mu should be a new value between these two. From the second one mu should be a new value between these two, from the third 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 the 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 M L 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 middlevalue will be median here.

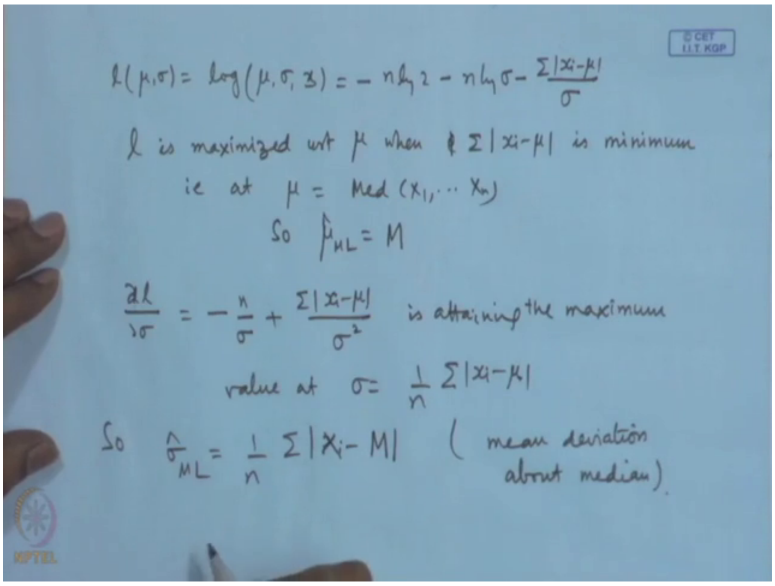
(Refer Slide Time: 52:14)



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 m minus mu plus 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 m x m plus 1, x 2 m minus 1 x 2 m, this will bek here. So, here if you see, the first term here will be minimum when mu lies between the largest intervals. 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 values, that is x k plus x k plus 1 by 2. So, this we conclude that mu is a median ofx 1 x 2, x2 k. So, where we may take it to be x k, plus x k plus 1 by 2. So, we have mu head M L 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: 56:20)



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, whensigma 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 M L 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 mu 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, you have already obtained the solution formu, if you 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 M. So, sigma head M L is equal to 1 by n sigma modulus x i minus M which is nothing, but the mean deviation about median. So, todayfriends we have discussedvarious probability models, and we have discussed the maximum likelihood estimators for those models. I have tried to covervarious cases here. And another thing is that we will take up some different cases, where either the maximumlikelihood 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.