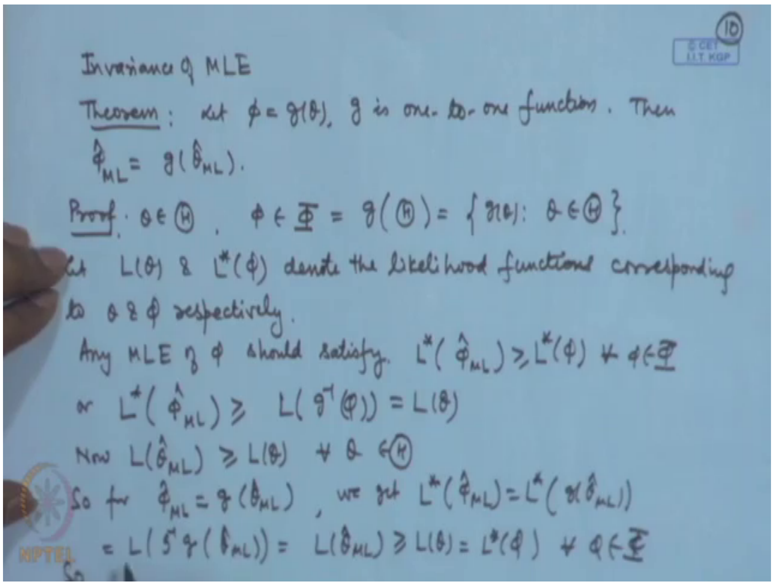
Statistical Inference
Prof. Somesh Kumar
Department of Mathematics
Indian Institute of Technology, Kharagpur

Lecture – 14
Properties of MLES - II

Form now, there are certain other properties of the maximum like fluid estimators like invariance which make it very attractive. What is the meaning of invariance? Suppose, we are able to obtain as a natural parameters theta, theta 1, theta 2 etcetera say suppose we have a one parameter problem and we have theta. So, we obtain the MLE of theta; however, suppose in the given problem it may be required that theta square is the quantity of interest 1 by theta is a quantity of interest, log of theta is a quantity of interest, in that case we can substitute the maximum likelihood estimator in that function.

In general, if we are considering function g theta, then g theta hat ML will be the actual MLE of g theta.

(Refer Slide Time: 01:13)



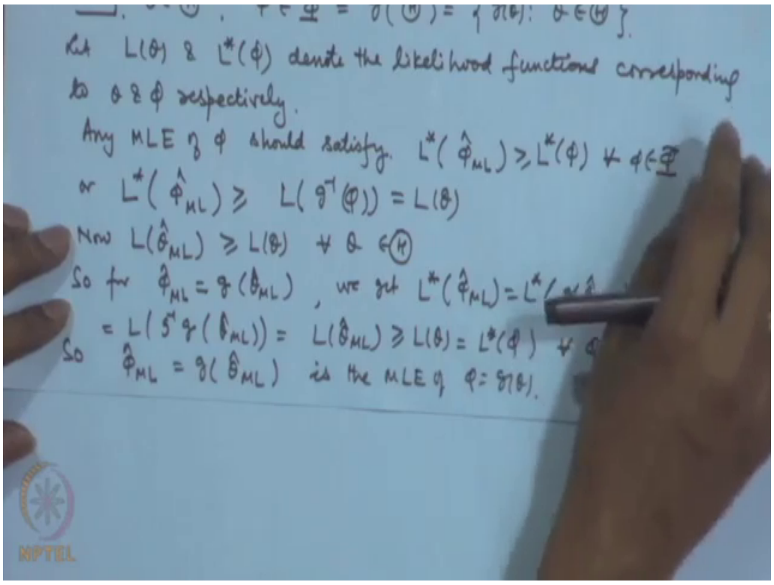
Now, this property I will be proving in two forms; invariance of MLE. Firstly, I will prove it for the one-one functions and then, actually I will give the general proof which is true for any function.

Let, phi be g theta where g is a one to one function, then phi hat ML is equal to g of theta hat ML. Let us look at the proof of this. So, suppose my parameter spaced for theta is capital script theta and phi we wrote as capital phi and this is actually the g theta space, that is the set of all g theta values as theta varies over a script theta. Let L theta and L star phi denote the likelihood functions corresponding to theta and phi respectively. Essentially, they are the same function because l theta is obtained as the joint distribution written at the point theta.

Now, in that you substitute because g theta is equal to phi. So, if we substitute in terms of phi here in that function, then that will be a function of phi and we denoted by L star. So, they are actually same functions, but written in as functions of different variables. Now, any maximum likelihood estimator of phi should satisfy L star phi hat ML greater than or equal to L star phi for all phi belonging to a script phi or L star phi hat ML greater than or equal to L g inverse phi that is equal to L theta. Now, L theta had ML it is greater than or equal to L theta for all theta belonging to theta a script theta.

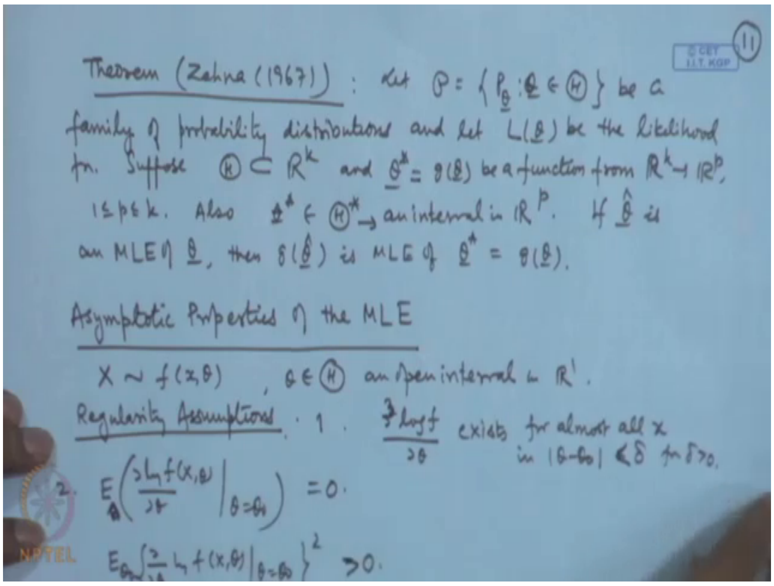
So, for phi hat ML equal to g of theta hat ML, we get l star phi hat ML is equal to L star g of theta hat ML that is equal to L g inverse g of theta hat ML that is equal to L of theta hat ML which is greater than or equal to L theta that is equal to L star phi for all phi.

(Refer Slide Time: 04:51)



So, you can say that phi hat ML is equal to g of theta hat ML is the MLE of phi is equal to g theta. Now, naturally this result is true we have proved for g being a one-one function. However, even if we have any function the same invariance property can be used as a justification for this was provided by Zahna 1967, I will state the result without proof here.

(Refer Slide Time: 05:26)



Let p theta belonging to script theta be a family of probability distributions and let L theta be the likelihood function. Suppose, theta is a subset of k dimensional Euclidean space and theta star be a function from R k to say R p where p will be less than or equal to k. Also, we assume that the range of theta star, this is an interval in R p. So, if theta hat is an MLE of theta, then g of theta hat is MLE of. The proof of this requires that for every value of a g theta, if there are several values which will lead to the same value because now we the function can be a an even function, then the likelihood function for theta star is defined as the maximum of all those values.

In the case of one to one transformation, what we have done is L theta is L of g inverse phi which we call L star phi whereas, in the case of an even function there can be several values corresponding to one value of phi, in this case theta star there can be several values. So, what we will do that for all those values we take the maximum of the likelihood function and then we maximize that.

So, when we associate maximum for each inverse image, what will happen is that, we are actually creating a one to one function and therefore, this theorem is once again applicable. However, I am just keeping the proof here for the details we can look at the paper by Zahna in 1967. I will now give some more asymptotic properties of the maximum likelihood estimators. So, let me call it large sample properties or asymptotic properties of the maximum likelihood estimator.

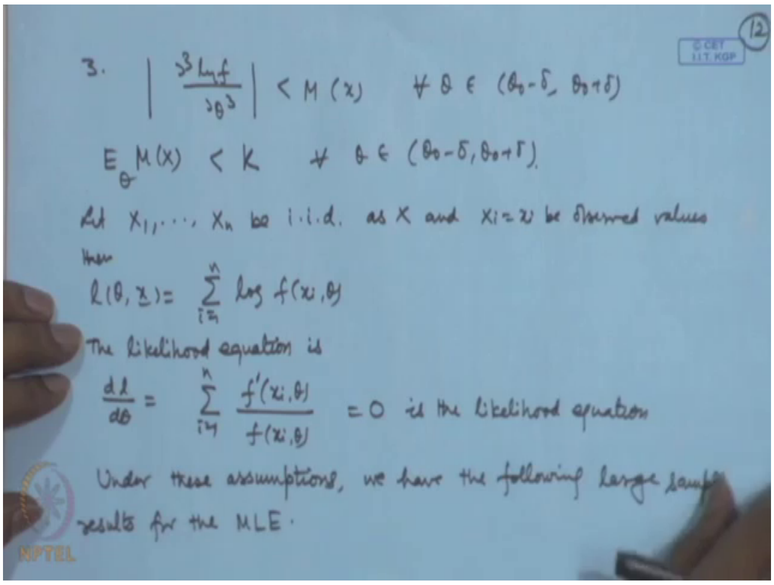
Now, these properties are true under certain conditions which we usually call regularity conditions. Now, these conditions were initially given by Cramer and these are usually called Cramer Rao or, Fisher Cramer Rao regularity conditions. So, I will just call it regularity conditions.

So, in general we are considering a class of probability distributions. Now, they may have probability densities or probability mass functions. So, let me write that the density function or the mass function, this is a general notation I am using theta belongs to a script theta, this is an open interval in real line, then we have the following regularity assumptions. The assumptions are as follows.

That the up to the third order derivative exists and this should be for all theta; however, it is enough if we assume it in a neighborhood of the solution. Suppose, we know that the solution exists around theta naught, then if we assume this derivative existing and in interval or in a neighborhood of theta naught, then it is enough; less than delta for some delta positive.

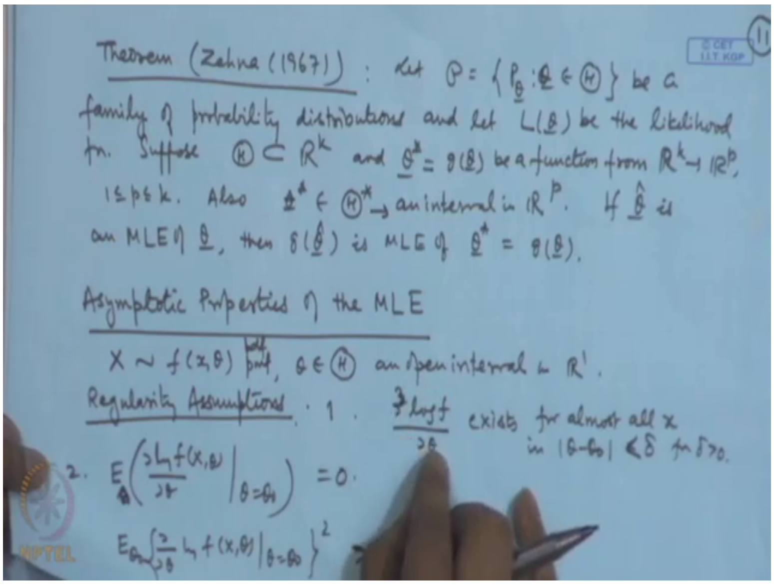
The second condition is that expectation of del log f by del theta at theta is equal to theta naught that is equal to 0, expectation of theta naught del by del theta log of f x theta at theta is equal to theta naught whole square that is positive.

(Refer Slide Time: 11:11)



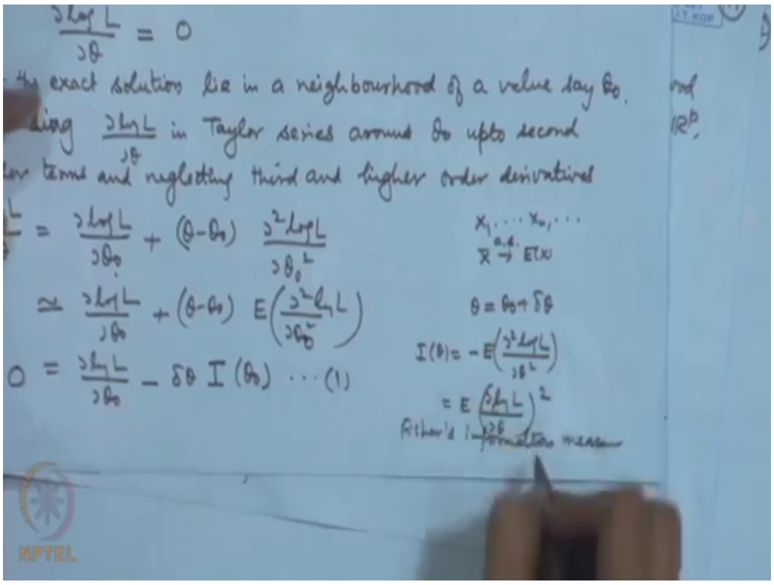
The third condition is that the third derivative of the density or the mass function is bounded in a neighborhood of theta naught and this function itself is having finite expectation a bounded expectation at any point in the interval theta naught minus delta to theta naught plus delta. So, if we are considering X 1, X 2, X n independent and identically distributed as x and observed values are then the likelihood equation is the log likelihood is and the likelihood equation is d l by d theta is equal to sigma f prime x i theta. Now, here prime means derivative with respect to theta, this is the likelihood equation.

(Refer Slide Time: 13:02)



Let us look at the conditions once again whatever assumptions we have made here f can be pdf or pmf theta belongs to the parameter space which is an open interval in the real line, we are assuming the derivative up to the third order exist. And, the assumption is at least for an interval in the neighborhood of the solution and then expectation of the first order derivative at theta is equal to theta naught was to vanish the expectation of first derivative square that should be positive. In fact, we have defined earlier this as the information function.

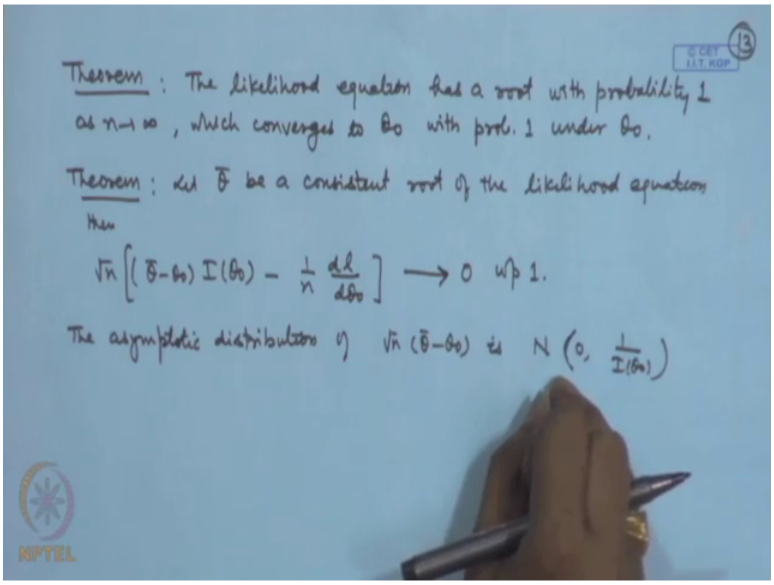
(Refer Slide Time: 13:48)



If we look at this one expectation of del log L by del theta whole square, this is the called fishers information major Fisher's information measure. We will talk more about it somewhat later third assumption is that the third area third order derivative is bounded by an integrable function. Now, under these conditions, we have under these assumptions we have the following large sample results for the maximum likelihood estimator.

I will stated it in the form of theorems without proof.

(Refer Slide Time: 14:47)



The first result says that the likelihood equation has a root with probability 1 as n tends to infinity and the root converges to theta naught with probability 1 under theta naught. So, this says that the likelihood equation has a root with probability 1 and the root converges to theta not with probability 1. So, this is a very important result and the second one says that the let theta bar be a consistent root of the likelihood equation, then square root n theta bar minus theta naught into I theta naught the information measure we have defined minus 1 by n d l by d theta naught, this converges to 0 with probability 1.

Now, what does it mean that square root n theta bar minus theta naught I theta naught minus 1 by n d l by d theta naught. This result is something like that the root of the likelihood equation that is the maximum likelihood estimator is asymptotically efficient because what term we are getting theta bar minus theta naught, you see if you take it to this side 1 by n d l by d theta naught divided by I theta naught. And, the asymptotic distribution of root n theta bar minus theta naught is normal 0 1 by I theta naught that is as n becomes large, the distribution of root and theta bar minus theta naught converges to that of a normal distribution with mean 0 and variance 1 by I theta naught, where I theta naught is the Fisher's information measure.

The proofs of these results are not very difficult actually they use the laws of large numbers and the central limit theorem at various points. I am skipping the proof here and what is more important is that this is true under fairly general conditions for example, the assumptions that I have stated now these assumptions are true for say binomial distribution, say for Poisson distribution, say for normal distribution, say for gamma distribution; that means, there is a large class of distributions particularly the distributions in the exponential family which satisfy these conditions Gaussian distribution is not in the exponential family, but even that also satisfies this condition. So, there is a fairly large class of distributions and densities which will actually satisfy this property.

So, under fairly general conditions we can say that the likelihood equation has a solution, the solution is consistent with probability 1; that means, it converges to the true value with probability one moreover the asymptotic distribution is normal and it is also second order efficients in the sense of Rao. So, that makes the use of likelihood maximum likelihood estimators a fairly important practice in the statistical theory.

See this property which I have written at the end that is a square root n theta bar minus theta naught or say square root and theta bar minus theta naught is asymptotically normal. This is also known as consistent as asymptotic normal property or kind estimator. So, if an estimator is consistent as well as it is asymptotic distribution is normal. So, naturally these are having some desirable properties, we also say best asymptotically normal estimator that is ban estimator so, can ban. So, under certain conditions such estimator have exist and maximum likelihood estimators are more likely to satisfy these properties. I will be completing this discussion now.