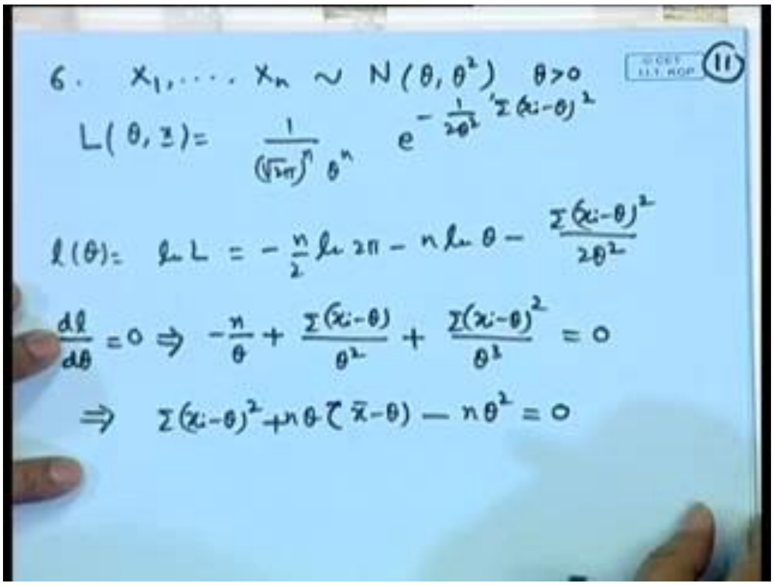
Probability and Statistics
Prof. Somesh Kumar
Department of Mathematics
Indian Institute of Technology, Kharagpur

Lecture - 58
Examples on MLE - II, MSE

Now, here we will discuss also the situations where the form of the maximum likelihood estimator may not be determined explicitly it may not exist or in case of certain situations we may have non unique maximum likelihood estimator.

(Refer Slide Time: 00:46)

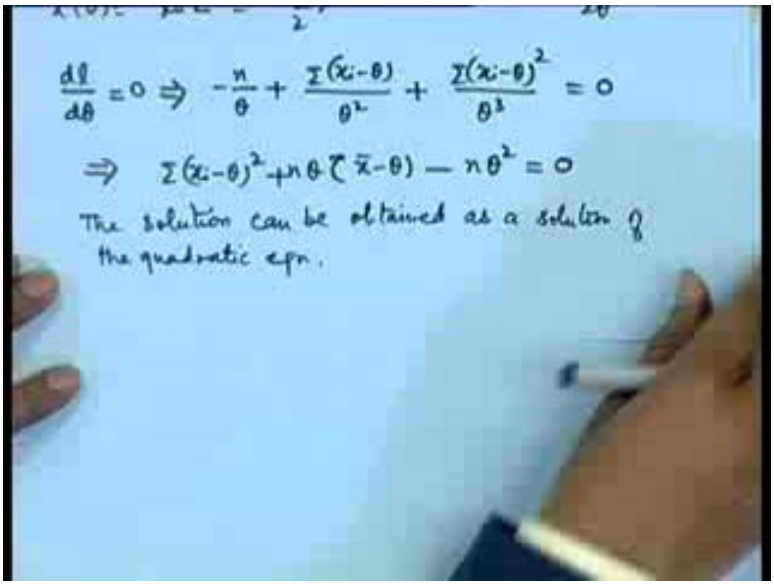


Let us take say X 1, X 2, X n follows n normal theta, theta is square; that means, I am considering the situation where the mean and the standard deviation are the same. So, naturally theta has to be positive here. Now here the likelihood function if you write then following the earlier set up it becomes 1 by root 2 pi to the power n, theta to the power n, e to the power minus 1 by twice theta is square, sigma x i minus theta whole square.

So, the log of the likelihood function that is minus n by tool log of 2 pi, minus n log theta, minus sigma x i minus theta whole square by twice theta square. So, what is the likelihood equation here? d l by d theta equal to 0 that gives minus n by theta plus, now if I consider here this term this is consisting of theta in the numerator as well as in the denominator, so the derivative will come in 2 terms sigma; x i minus theta by theta is square and plus sigma x i minus theta square by theta cube is equal to 0. So, if theta is taken to be positive I can strike of 1 theta and we can write the equation as. So, I multiply by theta cube in the full equation and we get it as sigma x i minus theta is square, plus theta into n x bar minus theta minus n theta is square equal to 0.

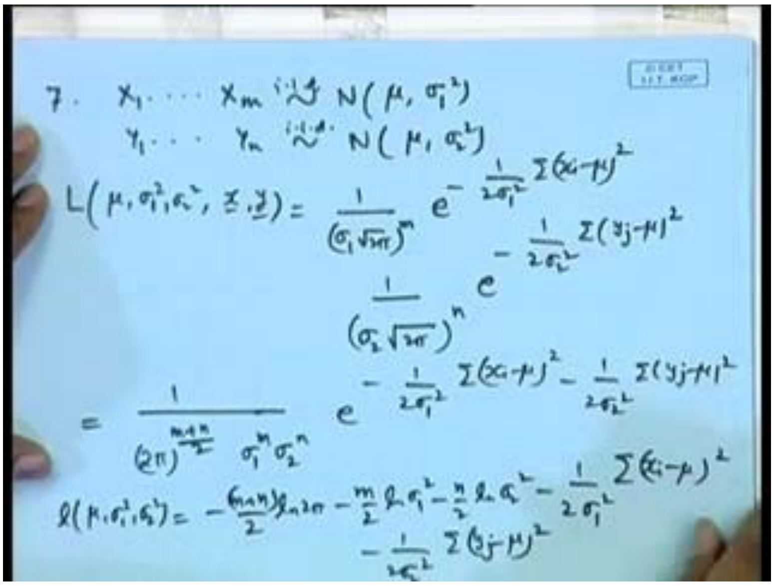
Here you can see that the solution of this equation can be obtained in the terms of solution of a quadratic equation.

(Refer Slide Time: 03:11)



The solution can be obtained.

(Refer Slide Time: 03:40)

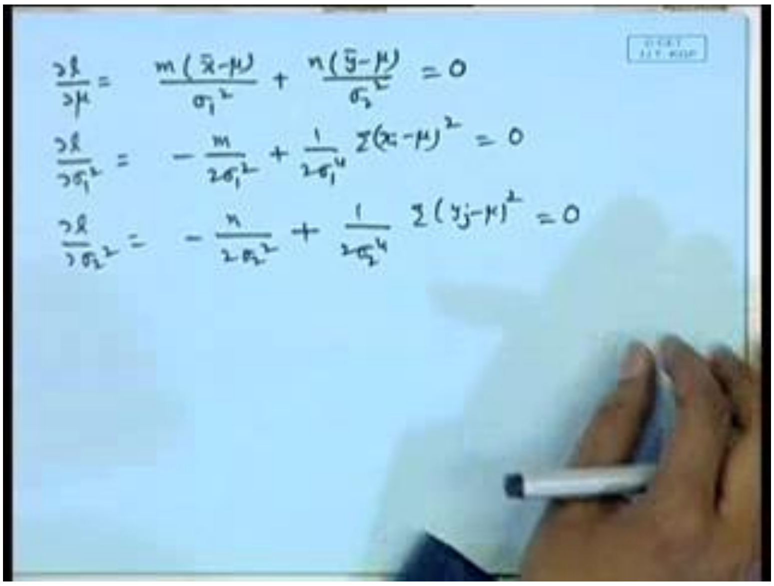


Let me take another example of similar nature, there the form may be even more difficult. Now this is popularly called the problem of common mean in the statistical inference. So, we have a random sample from a normal population with mean mu and variance sigma 1 is square, and another random sample Y 1, Y 2 Y n from a normal population with mean mu and variance sigma 2 square. So, in particular sigma 1 square sigma 2 square may be different, but the mean is common. So, here you write down the likelihood function here 3 parameters are there: mu, sigma 1 square, sigma 2 square and 2 samples x and y are there.

So, the likelihood function will involve 1 by sigma 1 root 2 pi to the power n, e to the power minus 1 by 2 sigma 1 square, sigma x i minus mu square and 1 by sigma 2 root 2 pi to the power n, e to the power minus 1 by 2 sigma 2 square sigma y j minus mu whole square. So, the terms can be simplified 1 by 2 pi to the power m plus n by 2, sigma 1 to the power n, sigma 2 to the power n, e to the power minus 1 by 2 sigma 1 square, sigma x i minus mu square, minus 1 by 2 sigma 2 square, sigma y j minus mu square.

So, the log likelihood function is equal to minus m plus n by 2, l n 2 pi minus n by 2 l n, minus m by 2 l n sigma 1 is square, minus n by 2 l n sigma 2 square, minus 1 by 2 sigma 1 is square, sigma x i minus mu square minus 1 by 2 sigma 2 square, sigma y j minus mu square.

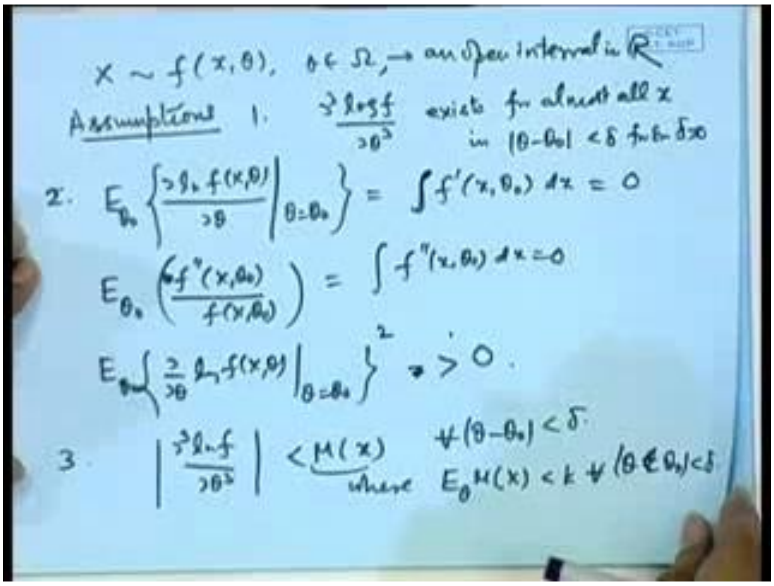
(Refer Slide Time: 05:54)



So, if we consider the likelihood equations; the equations are del l by del mu is equal to 0 that gives us m x bar minus mu by sigma 1 square plus n, y bar minus mu by sigma 2 square is equal to 0. If I consider del l by del sigma 1 square that gives us minus m by 2 sigma 1 square, plus 1 by 2 sigma 1 to the power 4 sigma x I, minus mu square del l by del sigma 2 square is equal to minus n by 2 sigma 2 square, plus 1 by 2 sigma to the power 4, sigma y j minus mu whole square.

Obviously you can see that if I obtain the value of mu here from here, it involves sigma 1 square and sigma 2 square. So, substituting here we get highly non-linear equation since sigma 1 square and sigma 2 square and the solutions for them cannot be obtained in the explicit form. So, numerical methods can be use to obtain the solutions therefore, the question arises that what are the situations where the maximum likelihood is estimator will exist or it will not exist. So, we have certain regularity conditions under which the maximum likelihood estimator always exists; let me briefly mention about this here the likelihood equations. So, we state in the following form let us have the following assumptions.

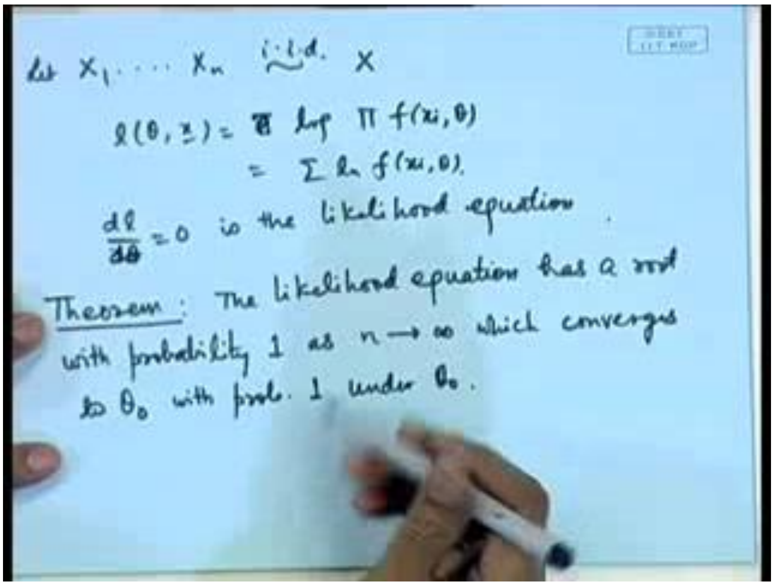
(Refer Slide Time: 07:50)



So, we consider X has a distribution f x theta, where theta belongs to omega and this omega is an open interval in the real line so; that means, I am considering one dimensional case, the assumptions are the third order derivative respective theta exists for almost all x in for some delta greater than 0. So, around some neighborhood of a point theta naught; second assumption is that at a point theta naught this expectation becomes 0; basically it means that the density can be integrated differentiated under the integral sin, now this integral is a generate notation this could be summation also in case we are dealing with the discrete distributions. So, in particular we assume up to higher order. That means, if we consider f double prime X theta naught by f x theta naught where this derivative is respect to theta then this should also be 0 and this square is greater than 0.

And the third order derivative is bounded in a neighborhood of where M x is also integrable function.

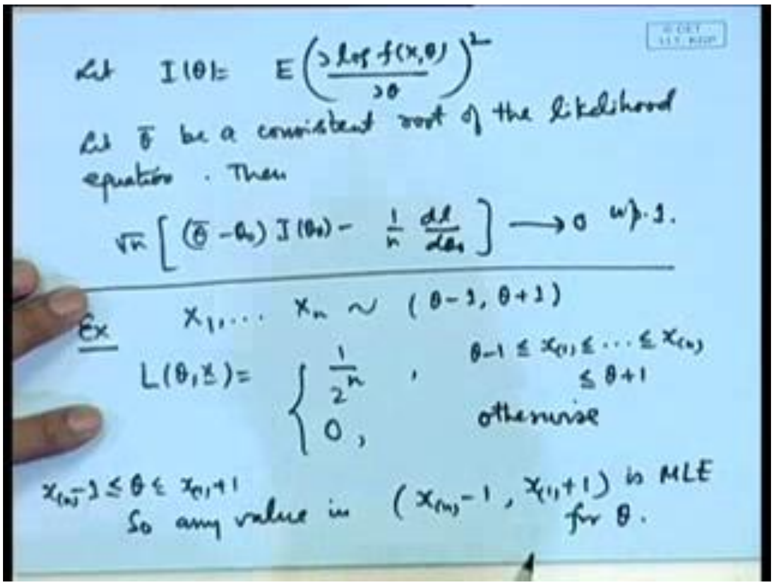
(Refer Slide Time: 10:25)



So under these assumptions on the distribution, now let us consider X 1, X 2, X n to be i. i. d as X and we define the likelihood equation as the log of product f x i theta, that is actually sigma, then d l by d theta is equal to 0 is the likelihood equation. So, we have the following result - The likelihood equation has a root with probability 1 as n tends to infinity, which converges to theta naught with probability 1 under theta naught.

So, this is an important result that is under certain regularity conditions the maximum likelihood estimator can always be found and it also converts this to the parameter with probability one that means it is also consistent. The second thing is that the asymptotic distribution is also normal.

(Refer Slide Time: 12:20)

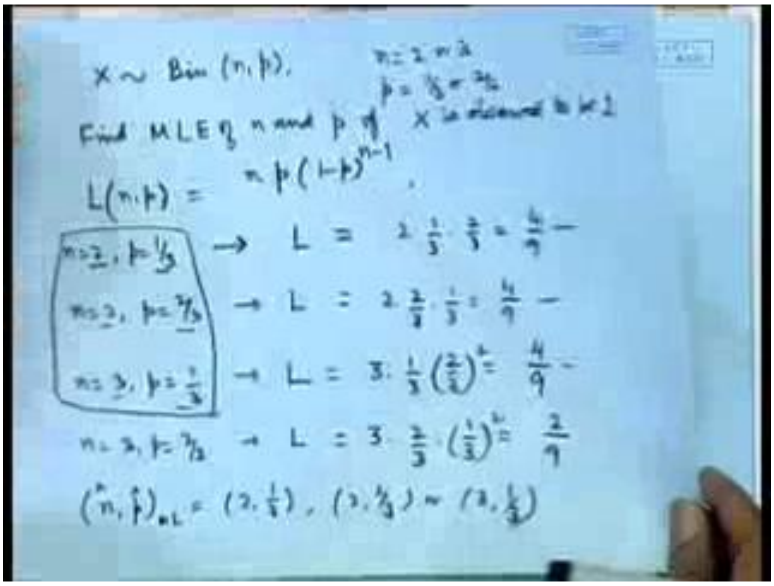


So, if I define say I theta as expectation del log f x theta by del theta whole square, which is actually called the information; let theta bar be a consistent root of the likelihood equation then so we are continuing with those assumptions, the asymptotic distribution of is 0 with probability 1. We further consider the case when the maximum likelihood estimators may not be unique.

Let us takes X 1, X 2, X n follow a uniform distribution on the interval say theta minus 1 to theta plus 1. So, the likelihood function in this particular case will be equal to 1 by 2 to the power n, for theta minus 1 less than or equal to X 1, and so on less than or equal to x n less than or equal to theta plus 1. So, you can see here that theta is less than or equal to x 1 plus 1 and it is also greater than or equal to x n minus 1. So, any value of theta between these 2 limits will be maximum likelihood estimator. So, any value in the interval x n minus 1 to X 1, plus 1 is maximum likelihood estimate for theta. So, we have a situation here where the maximum likelihood estimator is not unique.

However for convenience one may take the average of the end points that is X 1, plus X n by 2 as the maximum likelihood estimator in this case. Now we consider the case where the techniques of direct differentiation or even considering like that the behavior of the likelihood function as an increasing function or decreasing function may not be appropriate. I am talking about the case where we may have to take each value one by one and then check which one will give the maximum likelihood estimator; let me explain through an example.

(Refer Slide Time: 15:38)



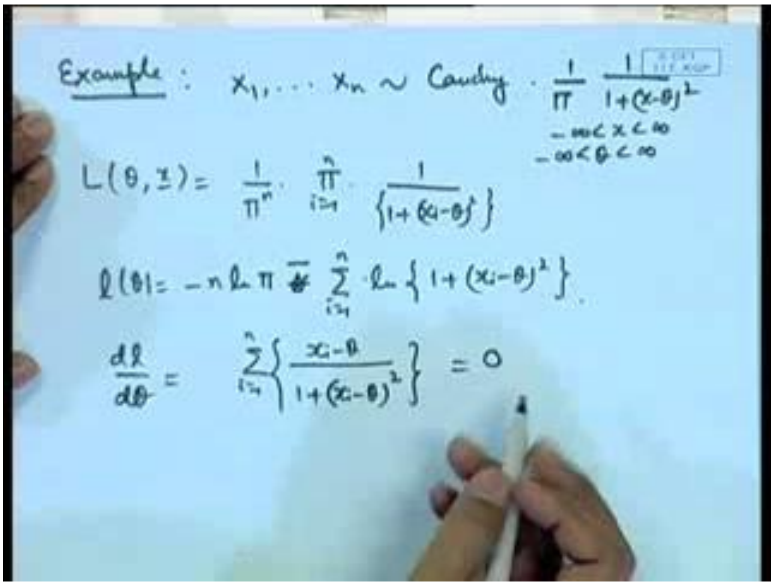
Suppose I consider binomial n, p and n is either 2 or 3 and p is either 1 by 3 are 2 by 3.

Now, we want to find out the maximum likelihood estimator of n and p if x is observed to be 1. Now this is the situation where we actually write down the values of the likelihood function at each of this parameter values that is n is equal to 2 p is equal to 1 by 3 n is equal to 2 p is equal to 2 by 3 etcetera and then see which one is the largest. So, let us write down the likelihood function here, the likelihood function here is n c x. So, since X is 1 so it is n, p to the power x means p, 1 minus p to the power n minus 1. So, since x is equal to 1 is already observed we are having exactly these values. So, we have the 4 values here when n is equal to 2, p is equal to 1 by 3 corresponded to this the likelihood function is equal to twice 1 by 3 into 2 by 3 that is equal to 4 by 9.

Then n is equal to 2 and p is equal to 2 by 3 the likelihood function value turns out to be twice 2 by 3, 1 by 3 that is again 4 by 9; when n is equal to 3 and p is equal to 1 by 3 the likelihood function value is equal to 3, 1 by 3, 2 by 3 is square that is equal to 4 by 9 again and n is equal to 3, p is equal to 2 by 3; here the likelihood function value turns out to be 3, 2 by 3, 1 by 3 is square which is equal to 2 by 9. If we are looking at the maximization of the likelihood function then you can observe here that this value this value and this value they are all same and they are the maximum this value is actually a smaller.

That means any of the configurations 2, 1 by 3, 2, 2 by 3 3, 1 by 3 they are as likely to give the maximum value as a new other value therefore, the maximum likelihood estimator for n and p can be considered to be pair 2, 1 by 3, 2, 2 by 3, or 3, 1 by 3 now this is another case because here four possible configurations are their out of that 3 or corresponding to the maximum likelihood value.

(Refer Slide Time: 18:58)



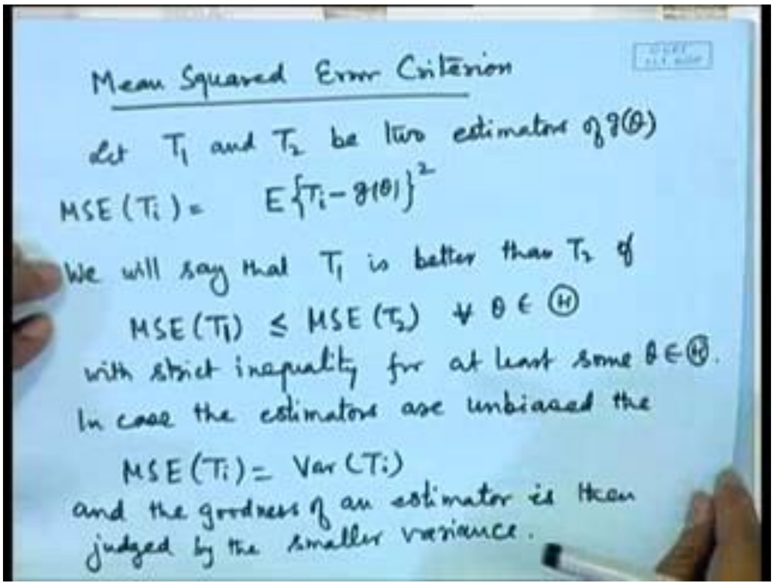
Let us take another example where the argument may follow a different path consider say X 1, X 2, X n following say Cauchy distribution with the density function 1 by pi, 1 by 1 plus x minus theta is square; where x is any real number theta is any real number.

Let us look at the likelihood function that is equal to 1 by pi to the power n and product i is equal to 1 to n, 1 by 1 plus x i minus theta is square. So, log likelihood function is then equal to minus n plus sigma i is equal to 1 to n. So, log of this term we can write it as a minus 1 plus x i minus theta is square. So, d l by d theta is equal to 0 that is the likelihood equation will be equal to i is equal to 1 to n, x i minus theta divided by 1 plus x i minus theta is square this is equal to 0; you can easily see that it is a non-linear equation and the explicit solution does not exist. So, one has to use numerical methods for finding out the solution of this equation.

So, this is another example where we may not have the solution of the likelihood equation in the direct form; however, if the assumptions that we mentioned earlier or satisfied, then we can prove that the solution will exist with probability 1 and it will be consistent estimator. Now next we come to the concept of judging the goodness of the estimators. So, for judging the goodness of the estimator's one may consider the variability aspect for example, unbiasedness is one judgment because whether the estimator is biased or unbiased. So, if estimator is unbiased it will be considered to be better than the biased estimator. If an estimator is consistent another estimator is inconsistent then again we may consider the consistent estimator to be better than the inconsistent estimator.

However if we are having several consistent estimators or several unbiased estimators, then how to compare among them? So, one of the popular criteria is to look at the variability of the estimator.

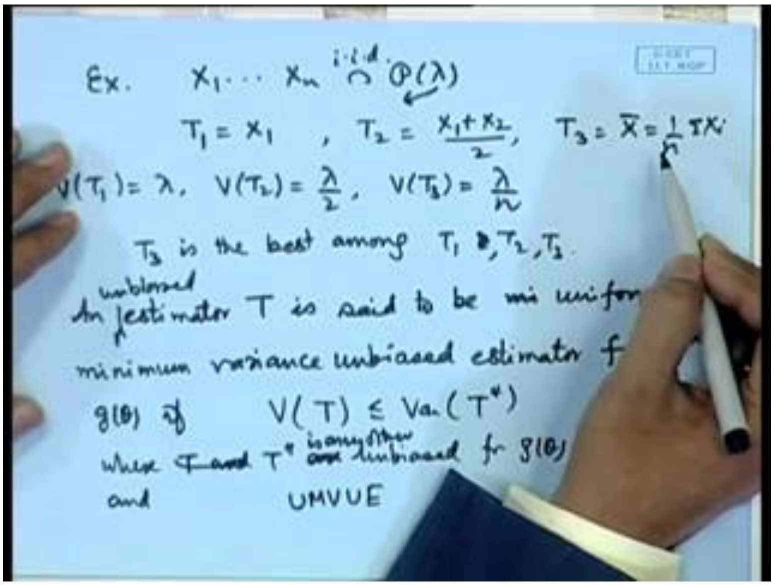
(Refer Slide Time: 21:49)



So, we have the so called mean is squared error criterion. Let T 1 and T 2 be two estimators of theta say g theta then the mean is squared error of T i is defined as expectation of T i minus g theta is squared. So, this is giving a measure of variability of the estimator and we will say that T 1 is better than T 2 if mean squared error of T i is T 1 is less than or equal to the mean squared error of T 2 for all parameters with a strict inequality for at least some theta.

Now, if the estimators are unbiased suppose T i is unbiased for g theta then this is reducing to the variance of T i and in that case the criteria can be written as in case the estimators are unbiased the mean squared error of T i simply becomes variance of T I, and the goodness of an estimator is than judged by the a smaller variance that is a smaller the variance the estimator is better.

(Refer Slide Time: 24:17)



As a simple example let us consider the exercise considered yesterday; suppose I am having X 1, X 2 x n following say poisson lambda distribution and I am considering estimation of lambda. So, let me write estimator T 1 as X 1, this is unbiased, let me write estimator T 2 has X 1, plus X 2 by 2 let me write estimator say T 3 has x bar which is actually the mean of all the observations.

Now, let us look at variance of T 1 that is lambda; if I look at variance of T 2 that is lambda by 2; if I look at variance of T 3 that is lambda by n. So, clearly here T 3 is the best among T 1 and T 2 among T 1, T 2 and T 3. So, this gives a procedure for checking that which estimators are better. Now among the unbiased estimators the one which has a smallest variance we call it minimum variance unbiased estimator. So, an estimator T is said to be uniformly minimum variance unbiased estimator for g theta, if variance of T is less than or equal to variance of say T star where T and T star are unbiased for g theta.

So, an unbiased estimator T said to be uniformly when memory is unbiased inter for g theta if variance of T is less than or equal to variance T star where T star is any other. So, for any other estimator if variance is a smaller for T, then definitely it is having the smallest. So, we use a terminology UMVUE for example, in this case of poisson distribution X bar will be uniformly minimum variance unbiased estimator.

Now the question arises that how to check that it is uniformly minimum variance unbiased estimator or how to find a uniformly minimum variance unbiased estimator because here we have already got it and then we can check, but then the total number of estimators are infinite and therefore, how to find that. So, we have to develop a method for finding out the UMVUE.

So, in the next lecture we will be considering these methods there is some additional terminology called sufficiency and completeness which is quite useful; also there are methods of obtaining lower bounds which can be found. So, in the next class we will be considering that.