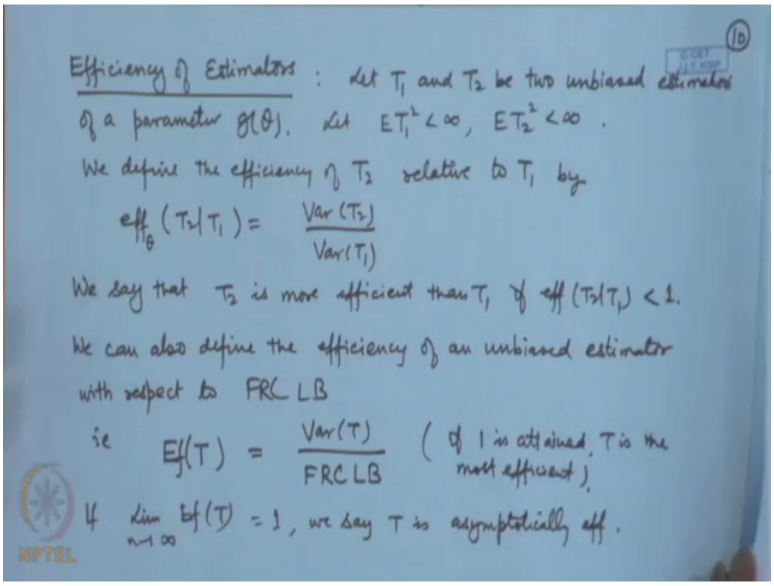
Statistical Inference
Prof. Somesh Kumar
Department of Mathematics
Indian Institutes of Technology, Kharagpur

Lecture – 18
Lower Bounds for Variance – IV

We have discussed the concept of minimum variance; that means, I am among unbiased estimators the estimator which has the minimum variance is some considered to be the best. In general, we can always compare two unbiased estimators by comparing their variances. That means, the one which has the smaller variance will be considered to be more stable or better. So, there is a classical concept of efficiency of estimators based on this. Let me discuss that here efficiency of estimators.

(Refer Slide Time: 00:47)



So, let T 1 and T 2 be two unbiased estimators of a parameter say g theta. And let us assume that they have the finite second moment, this condition is required because the variance must exist. So, we define the efficiency of T 2 relative to T 1 by so, we use a notation e f T 2 given T 1 it is equal to variance of T 2 divided by variance of T 1.

Naturally, if the variances are equal then the efficiency will be equal to 1. If the efficiency is less than 1; that means, variance of T 2 is less than variance of T 1; that means, T 2 is more efficient than T 1. Conversely if the efficiency is more than 1 then variance of T 2 will become bigger than variance of T 1; that means, T 1 is better than T

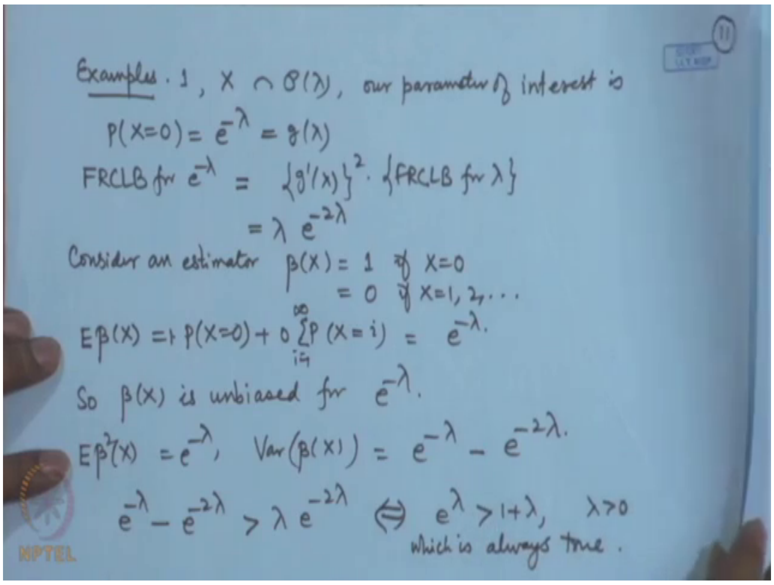
2. So, we say that T 2 is more efficient than T 1 if efficiency function is less than 1. Now, this is regarding any two estimators. Now, in general give an any unbiased estimator we can consider its efficiency with respect to the Rao Cramer lower bound.

So, for example, we may consider an estimator which attains the FRC lower bound if that is so, then that is a benchmark or you can say the best thing. So, anything which is bigger than that its efficiency will be considered with respect to that; that means, its efficiency will be bigger than 1. So, we can also define the efficiency of an unbiased estimator with respect to the FRC lower bound that is, we may say let me give another notation we will call it E notation.

So, efficiency of an unbiased estimator E f efficiency of an estimator T we define as variance of T divided by FRC lower bound for the variance of an unbiased estimator for that parameter. Suddenly, we know that sometimes this may be attained sometimes this may not be attained. So, these definitions are not full-proof another thing is that in certain cases we may not consider unbiased estimators. Because, if we consider only mean squared error as a criteria it may turn out that the mean squared error is less than the variance by combining certain terms.

We can also consider that although this may not be attained, but asymptotically it may be attained. So, we can give a definition that if limit of this is equal to 1, then we say that T is asymptotically efficient. So, here if 1 is attained, T is the most efficient. Let us look at some examples here.

(Refer Slide Time: 05:25)



Let us go back to the Poisson example and for convenience let me restrict attention to 1 observation suppose, X follows Poisson lambda. And here, our parameter of interest is say probability X is equal to 0 that is e power minus lambda. Of course, we may ask the question that why we are considering this function.

Now, usually a Poisson distribution is the distribution of the number of arrivals, number of occurrences, during a given time interval or during a given area or during a given space etcetera. Now, what happens for example, if you are considering a q service q then how many people are arriving that will denote the number X. Then certainly it is of interest to know that if X is equal to 0; that means, there is a slack period. Because, if in a service q it may happen that we have to imply service personnel; that means, the person who will be giving the service.

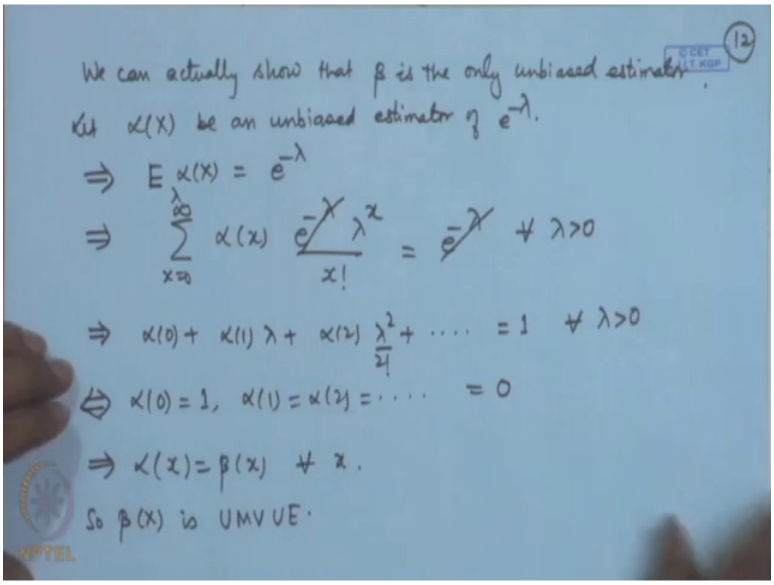
For example, it is a railway ticket counter, it is a ticket counter at a cinema hall or it is a it is a service counter at a popular say cafe. So, therefore, persons are required there are person all are required. In the when there are no person; that means, when X is equal to 0 we need not deploy the people or we may deploy less number of people. So, certainly in such cases it is of interest to know or estimate the probability of 0 occurrence.

So, this gives us this parametric function e to the power minus lambda suddenly it is a non-linear function of lambda therefore, the variance of an unbiased estimator of e to the power minus lambda can never attain the lower bound. Let us look at this what will be the lower bound? FRC lower bound for e to the power minus lambda that is equal to g prime lambda square into the FRC for lambda. For lambda it is lambda by n and if n is equal to 1 this is simply lambda.

The derivative of g lambda is e to the power minus lambda with a minus sign when we squared it we get e to the power minus 2 lambda. So, this is lambda this is a lower bound. Now, let us consider an estimator say beta X is equal to 1 if X is equal to 0 it is equal to 0 if X is equal to 1 2 and so on. Then, if you look at expectation of beta X that is equal to 1 into probability X equal to 0 plus 0 into probability is equal to X is equal to say i i is equal to 1 to infinity. So, this becomes 0. So, this is e to the power minus lambda.

So, beta X is unbiased for e to the power minus lambda. However, if you look at expectation beta square, now this will again be same and therefore, variance of beta that is also e to the power minus lambda minus e to the power minus 2 lambda. Now, if you compare this with the lower bound e to the power minus lambda minus e to the power minus 2 lambda greater than lambda e to the power minus 2 lambda. Because, this is equivalent to e to the power lambda greater than 1 plus lambda for lambda positive which is always true. So, you can see that this lower bound is not attained.

(Refer Slide Time: 09:57)

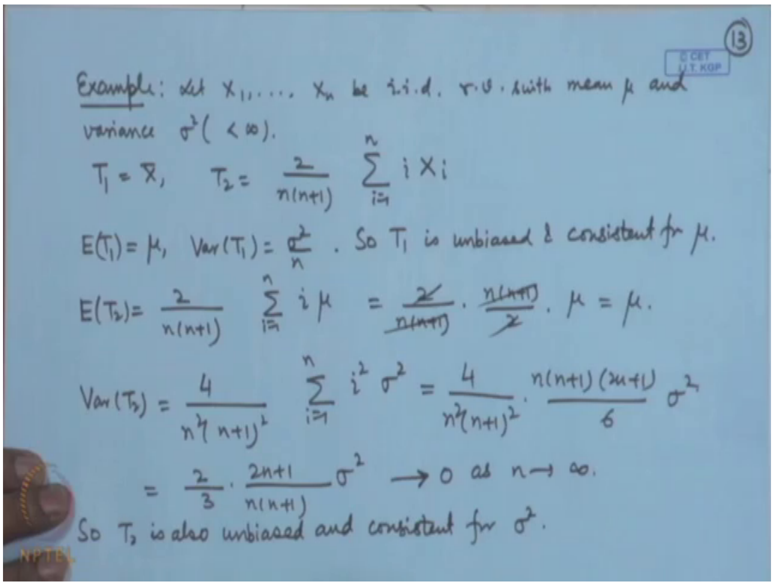


However, we can use another argument to actually prove that beta X is we can actually show that beta is the only unbiased estimator. We can proceed by the basic principles let us consider say alpha X let alpha X be an unbiased estimator of e to the power minus lambda. Then we should have expectation of alpha X equal to e to the power minus lambda.

Now, let us write down this relation alpha x e to the power minus lambda lambda to the power x by x factorial is equal to e to the power minus lambda for all lambda. Now, this e to the power minus lambda you can remove from both the sides because this is a positive term. So, this is reducing to then alpha 0 plus alpha 1 into lambda plus alpha 2 into lambda square by 2 factorial and so on is equal to 1. So, left hand side is a power series in lambda and right hand side is simply a constant.

So, this is true if and only if the coefficients match; that means, alpha zero must be 1 and alpha 1 alpha 2 and so on all of them must be 0 which is the same as the function beta because beta 1 beta 0 was 1 and beta 1 beta 2 and so on all of them were 0. So, this alpha function and beta functions are the same. So, beta must be UMVUE. So, although here the lower bound is not attained, but actually beta will be the most efficient estimator here. Let me give an example of comparing two unbiased estimators with respect to their variances.

(Refer Slide Time: 12:11)



The estimators may both may be unbiased, both may be consistent etcetera. So, let us take another example. I am not taking any distributional form let us consider say X 1 X 2

X n be independent and identically distributed random variables with say mean mu and variance sigma square; obviously, we are assuming that variance is finite here.

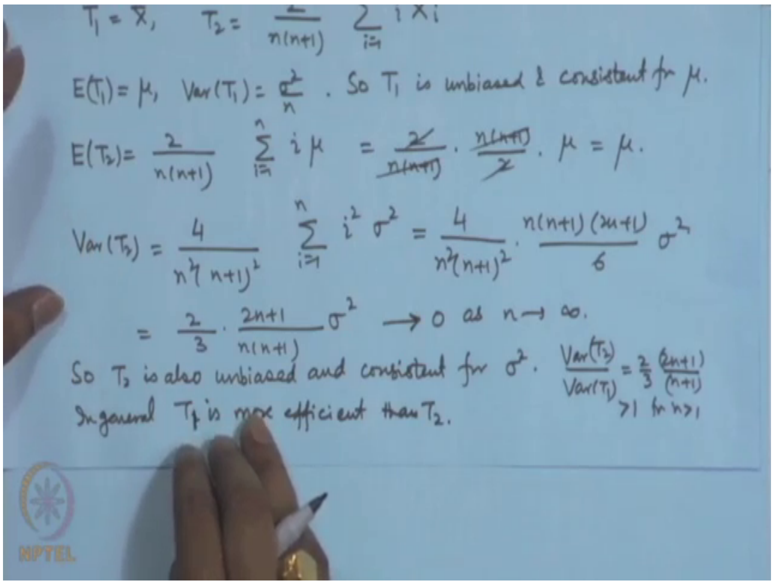
Now, you consider 2 unbiased estimators here let me take say T 1 is equal to X bar and T 2 is equal to 2 by n into n plus 1 sigma i X i i is equal to 1 to n. Now obviously, if you look at expectation of T 1 this we have seen that sample mean is unbiased for the population mean, the variance of this is equal to sigma square by n. So, if estimator is unbiased and its variance converges to 0 then we also know that it will be consistent. So, what we are seeing that T 1 is unbiased and consistent for estimating mu.

Now, if you look at T 2. So, that is equal to expectation of T 2 is equal to 2 by n into n plus 1 sigma i is equal to 1 to n i expectation of X i that is again mu. So, sigma i is n into n plus 1 by 2. So, you get 2 by n into n plus 1 into n into n plus 1 by 2 into mu. So, these terms cancel out you get only mu. So, T 2 is also unbiased let us look at variance of T 2.

Now, variance of T 2 if you take this is constant. So, it will become square 4 n square into n plus 1 square sigma i is equal to 1 to n i square into variance of X i, variance of X i is sigma square. Since, we have assumed independence of the observations the correlation or covariance term will not come here you get this. Now, sigma is square we have the formula so you get 4 n square n plus 1 whole square n into n plus 1 into 2 n plus 1 by 6 sigma square. So, after simplification you get it as 2 by 3; 2 n plus 1 divided by n into n plus 1 sigma square.

So, as n tends to infinity this goes to 0. So, T 2 is also unbiased and consistent for sigma square. However, let us compare the variances what is variance of T 2 by variance of T 1.

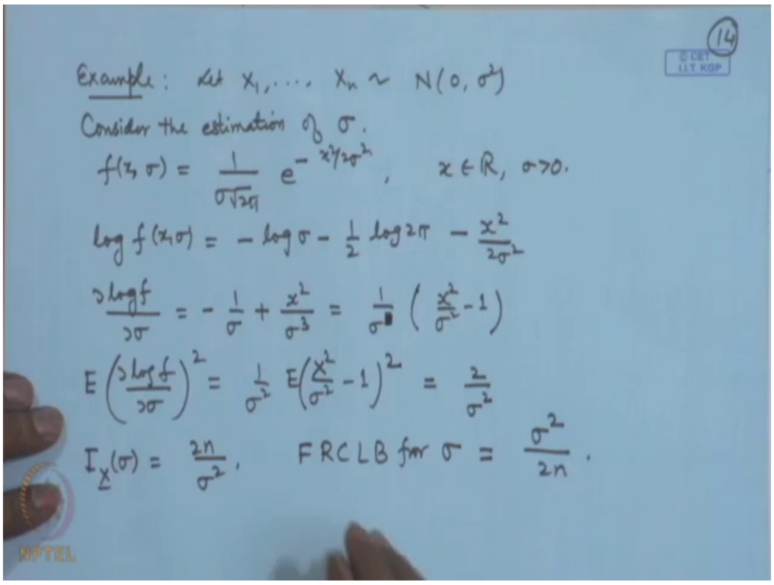
(Refer Slide Time: 15:41)



Variance of T 2 divided by variance of T 1. So, sigma square is coming here sigma square is appearing here by n by n so that will cancel out. So, you get the term as 2 by 3 2 n plus 1 divided by n plus 1.

Obviously, this is always greater than 1 for n greater than 1. If n is equal to 1 of course, this will be equal to 1 and if n is equal to 1 actually T 1 and T 2 are both equal to X 1 so that case is of not any interest. So, in general T 2 T 1 is more efficient than T 2. So, here you have seen we have 2 estimators both of which are unbiased as well as consistent for the sample mean, but one of them can be preferred over the other if we are applying the criteria of a smaller variance.

(Refer Slide Time: 16:59)



Now, let me also take another distribution say suppose we consider a random sample from a normal distribution, where mean I was assume to be 0 and variance is sigma square. We have already discussed this example in the context of estimation of sigma square when mu are some fixed volume mu naught. Now, whenever mu is some fixed value mu naught you can always shift the observations so that the mean can be made to be 0.

Now, suppose my interest is not to consider estimation of sigma square, but the estimation of sigma. So, consider the estimation of sigma say. Now let us look at the lower bound the density function is of the form 1 by sigma root 2 pi e to the power minus x square by 2 sigma square, where x is of course, any real value. Log is equal to minus log sigma minus 1 by 2 log 2 pi minus x square by 2 sigma square; so, derivative of this with respect to sigma that is minus 1 by sigma minus. Now derivative of this will become 0 then derivative of 1 by sigma square is minus 2 by sigma cube. So, it will become x square by sigma cube that is equal to 1 by sigma cube we can write it as 1 by sigma x square by sigma square minus 1.

So, expectation of del log f by del sigma is equal to 1 by sigma square expectation of X square by sigma square minus 1 whole square. Now, if X follows normal 0 sigma square then X by sigma follows normal 0 1 X square by sigma square will follow chi square on

1 degree of freedom. So, therefore, this will have expectation 1 and therefore, expectation of the variable minus its mean square that is going to be the variance.

Now, variance of a chi square is twice its degrees of freedom. So, this term will become equal to 2 so, this is simply equal to 2 by sigma square. Say if we consider the information that will be equal to 2 n by sigma square. So, the FRC lower bound for estimation of sigma that will be equal to sigma square by 2 n. In the following class I will consider two estimators for this see whether they any of them attain the lower bound and also compare them so, that I will be doing in the following lecture.