Statistical Inference

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

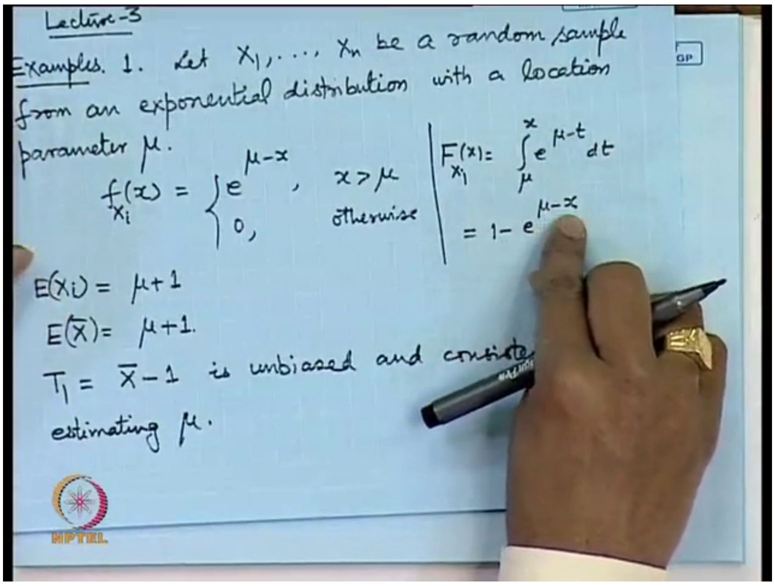
Module No. # 01
Lecture No. # 03

Basic Concepts of Point Estimations – II

In last lecture, I introduced two basic concepts of point estimation namely unbiased estimation and consistency.One of the properties that is unbiasedness it is related to the estimator being equal to the true value on the average, that means if we have many samples then the average of that will be equal to the true value.Whereas, consistency is a large sample property that means if we take a sample to be large enough then the probability that it is close enough to the true value of the parameter is almost equal to one.And so, the two properties have somewhat different applications and as well as implications and many times we try to combine various properties of the estimators.

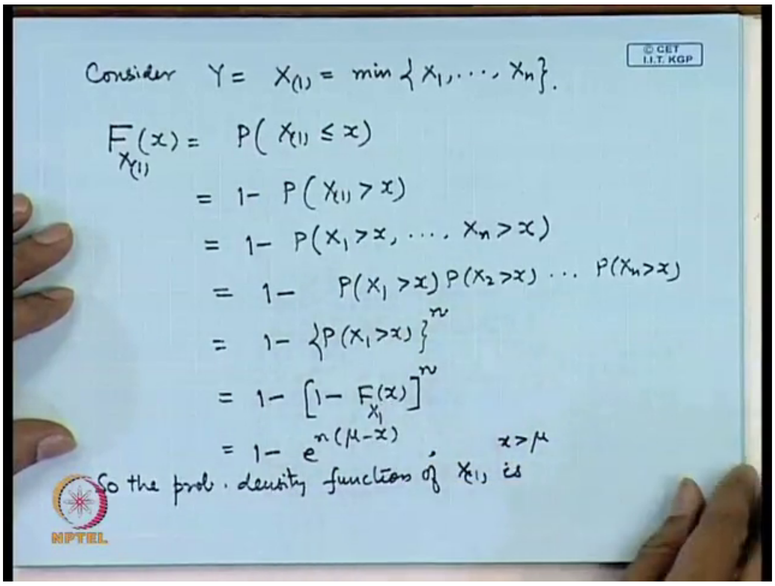
So, I had shown in the last lecture some sort of invariance of the consistent estimators.For example, if t is a consistent estimator for theta then g of t n, where g is a continuous function will be consistent for g theta.Similarly, if I have t n to be consistent for theta and I have sequences of numbers a n and b n, such is that a n converges to one and b n converges to zero then a n t n plus b n also is a consistent estimator for theta.

(Refer Slide Time: 02:05)



So, let me give a few examples, let us consider say let X 1 X 2 X n be a random sample from an exponential distribution with a location parameter mu.That means, I am considering the density function of say X i to be e to the power mu minus x where x is greater than mu and 0 otherwise. So, this is actually the well known shifted exponential distribution here mu denotes the minimum guarantee time of the component are the life.So, here if you see expectation of X i is equal to mu plus 1. So, mu plus 1 is the first moment and therefore, if I consider expectation of x bar that is also going to be mu plus 1. So, by weak law of large numbers we get x bar as a consistent estimator for mu plus 1 and if I take 1 to the left hand side then we get let me call it T 1 so T 1 is equal to X bar minus 1 is unbiased and consistent for estimating mu that is a minimum guarantee time.

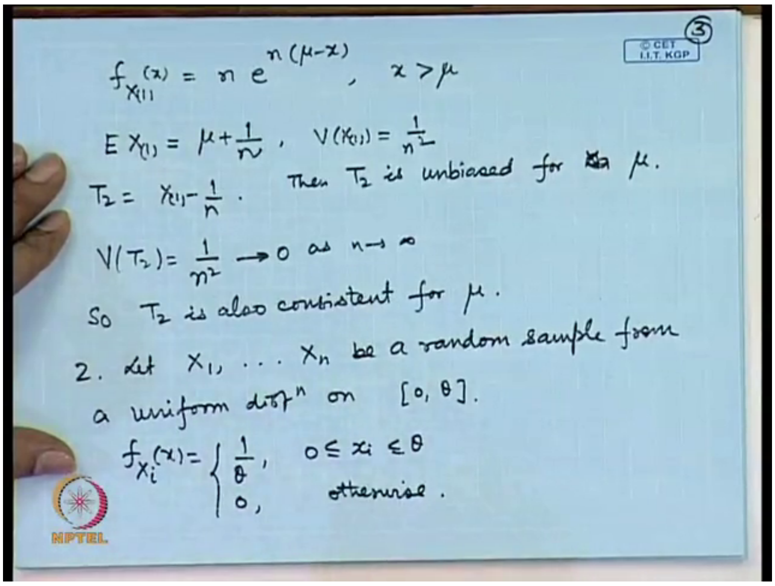
(Refer Slide Time: 04:20)



Now, in this problem let me introduce another estimator,let us consider say Y is equal to X 1 here this X 1 denotes the minimum of the observations X 1 X 2 X n 1 can derive the distribution of X 1.In fact in general if I want to find out the distribution of this I can find it in the following way I can either say c d f of this that is probability of X 1 less than or equal to x.This can be written as 1 minus probability of X 1 greater than X that I can write as 1 minus probability that now, if the minimum is greater than X this is equivalent to saying each of the X i is are greater than X.Now, here X 1 X 2 X n a is a random sample therefore, X 1 X 2 X n are independently and identically distributed random variables.

So, this can be actually written as 1 minus probability of X 1 greater than x into probability of X 2 greater than x and so on probability of X n greater than x that is 1 minus probability of X 1 greater than x to the power n that is equal to 1 minus now, this is again 1 minus c d f of X 1 itself.Now, if I have this as the probability density function I can write down the corresponding c d f here, that is integral from mu to x of e to the power mu minus t d t that is equal to 1 minus e to the power mu minus x.So, if I substitute 1 minus e to the power mu minus x here, I will get 1 minus e to the power n times mu minus x.So, the probability density function of X 1 is now this can be obtained by considering derivative of this of course, this value I have written for X greater than mu if x is less than mu then this is going to be zero.

(Refer Slide Time: 07:20)

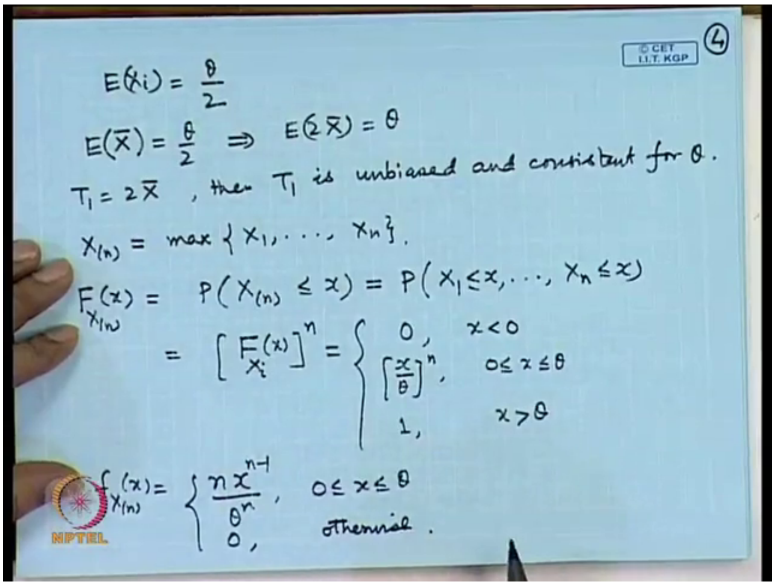


So, if we consider derivative of this I will get the density function of X 1 as n time's e to the power n mu minus x where x is greater than mu. So, this is the probability density function of the minimum of the observations if I have considered a random sample from, an exponential distribution with a location parameter mu.Now, this is the usual 2 parameter exponential distribution here the scale parameter is 1 by n and location parameter is mu. So, if I consider the expectation of X 1 that is equal to mu plus 1 by n.So, if I take T 2 as X 1 minus 1 by n then T 2 is also unbiased for X 1 for mu so, I have got another unbiased estimator.

Now, in this one I can consider variance also what is variance of X 1 for example, y variance of X 1 here is 1 by n square. So, variance of T 2 is also 1 by n square, because it is variance of X 1 itself the variance of a function does not change if I make a change of origin.Now, consider the result that if expectation is equal to the parameter and the variance goes to 0 T 2 is unbiased and it is variance goes to 0 as n tends to infinity so T 2 is also consistent for mu. So, in this problem we have considered two estimators,one is based on the sample mean this is unbiased and consistent and at the same time we have considered T 2, which is based on the minimum of the observations this is also unbiased and consistent.

So, that brings us to the question that if I have more than one estimator satisfying certain given desirable properties then, which one you should use. So, in this direction I will give you 1 more example, let us consider say X 1 X 2 X n be a random sample from a uniformed distribution on the interval 0 to theta, that means I am considering the density of x i as 1 by theta 0 less than or equal to x i less than or equal to theta it is equal to 0 otherwise.

(Refer Slide Time: 11:08)

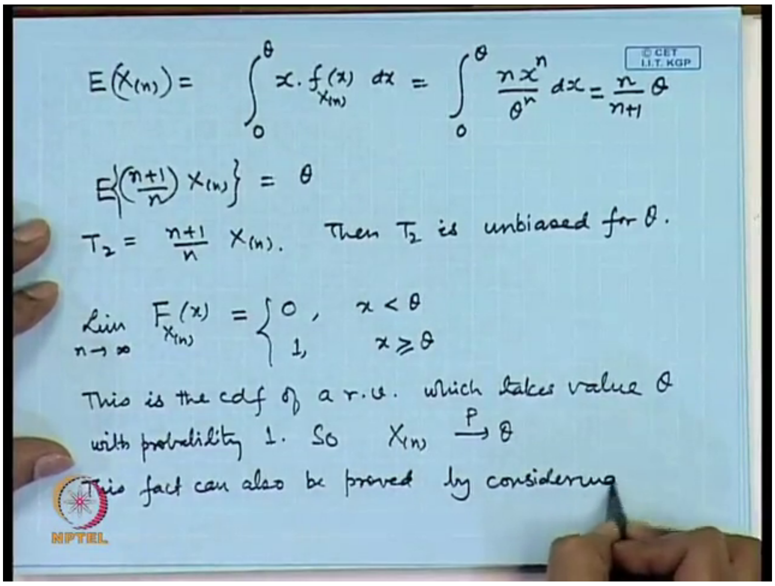


Now, in the uniformed distributionwe know expectation of x i is the middle point of the interval that is theta by 2. So, immediately we conclude that expectation of x bar is theta by 2 this implies that expectation of 2 x bar is equal to theta so, if I call T 1 is equal to 2 x bar then T 1 is unbiased and consistent for theta.Now, in this problem let me consider another one let us consider say X n now X n I am calling to be the maximum of the observations.As in the previous case we can derive the distribution of X n let us consider the c d f of this so, this is equal to probability of X n less than or equal to x.

Now, this statement that the maximum is less than or equal to x is equivalent to that each of the observations is less than or equal to x.And once again using the fact that x i's are independently and identically distributed this is equivalent to same each of the x i's c d f at x so, this is simply this to the power n.Now, for the uniformed distribution the c d f is it is equal to 0 if x i is less than 0 it is equal to x by theta, if 0 is less than or equal to x is less than or equal to theta it is equal to 1 if x is greater than theta.So, if we use this c d f here this becomes 0 if x is less than 0 it is equal to x by theta to the power n, if 0 less than or equal to x is less than or equal to theta it is equal to 1 if x is greater than theta.

One may find out the probability density function from here, by considering the derivative because this is a continuous distribution so, you get the density function as n x to the power n minus 1 by theta to the power n, if x lies between 0 to theta and it is 0 otherwise.

(Refer Slide Time: 13:56)

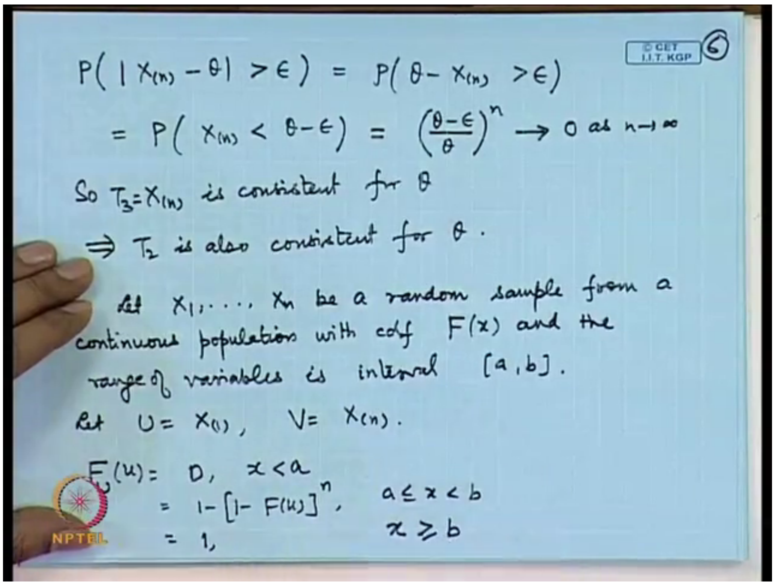


Let us consider say expectation of X n now so, expectation of X n is equal to integral x into the density function of X n from 0 to theta, that is equal to integral 0 to theta n x to the power n by theta to the power n d x.So, as we can see easily the integral of x to the power n will be x to the power n plus 1 by n plus 1 and if we substitute the limits from 0 to theta I will get theta to the power n plus 1.And in the denominator I have theta to the power n so that will cancel and therefore, this value will be equal to n by n plus 1 theta.If I adjust this n by n plus 1 on the left hand side I get this is equal to theta. So, if I use the notation T 2 as n plus 1 by n X n then T 2 is unbiased we had obtained estimator T 1 as 2 x bar which is unbiased and now I have obtained T 2.

Let us check say probability of X n minus modulus theta greater than epsilon whether it tends to 1 or not, that we can check from here also.If we take the limit of this cumulative distribution function now, here x is less than or equal to theta. So, this value will tend to 0 if x is less than theta and whenever, x is greater than or equal to theta it is becoming 1.So, if I take the limit of if I consider say limit of f X n as n tends to infinity then this is equal to 0 for x less than theta and it is equal to 1 for x greater than or equal to theta.

Now, this denotes the distribution of a random variable which takes value only theta this is the c d f of a random variable, which takes value theta with probability 1. So, basically we approved that X n converges to theta in distribution, but theta is a constant therefore, it is equivalent to saying X n converges to theta in probability.In fact this fact can also be proved in a different way I will consider directly the definition of convergence in probability.

(Refer Slide Time: 17:32)



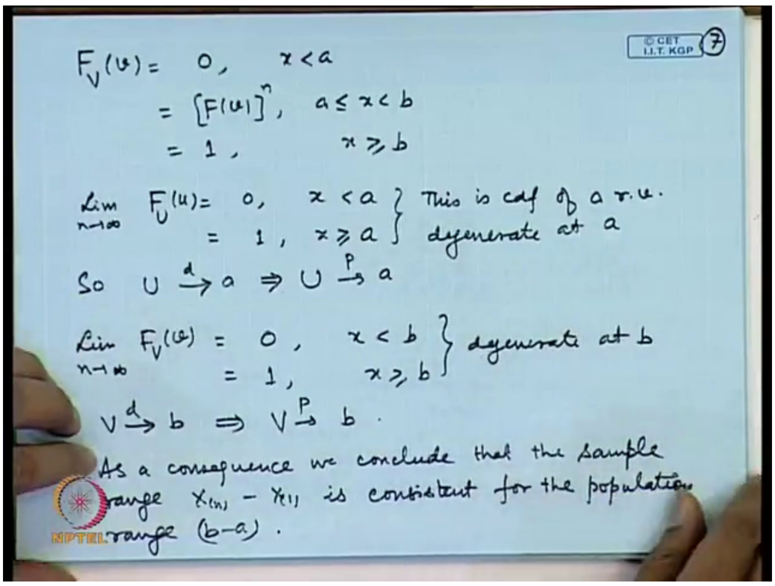
Let me take say probability of modulus X n minus theta say greater than epsilon. Now, the distribution of X n is in the interval 0 to theta that means, X n is always below theta so, if we consider this modulus of X n minus theta this is same as theta minus X n.So, this is equivalent to probability of theta minus X n greater than epsilon which I can write as probability of X n less than theta minus epsilon.Now, this is nothing but the distribution function of X n at the point theta minus epsilon since X n is having a continuous distribution whether, we put less than or less than or equal to it does not make a difference therefore, this value is equal to theta minus epsilon by theta to the power n as we have derived just in the previous sheet here.

So, now you can see epsilon is a positive number so, theta minus epsilon is less than theta therefore, if I take the limit as n tends to infinity this will go to 0 so, X n is consistent for theta let me call it t 3.If we look at the coefficient n plus 1 by n this goes to one as n tends to infinity so, we have the result that if t n is consistent for theta and a n goes to one, then a n t n is also consistent therefore, if we use this fact T 2 is also consistent for theta.So, T 2 is unbiased and consistent T 1 is also unbiased and consistent so, I have given you two different distributions where for estimating of one parameter I m getting two different unbiased and consistent estimators.

So, that shows that actually, we need additional criteria to distinguish between or to choose between where yes competing estimators.From the previous two exercises, we can also find something more important if we look at the form of the distribution function of the maximum and the minimum there is some specific structure here.For example, when I took the limit here I got only 0 1 here similarly, in the distribution of the minimum, we had that X 1 is converging to mu so, minimum was converging to the lower limit and here the maximum is converging to the upper limit theta.

In fact if we have any continuous distributions then this is a general fact so I will state it in the following results so, let me give it as exercise.Let X 1 X 2 X n be a random sample from a continuous population with c d f say capital F x and the range of variables is interval a to b of course, this interval may be open or close that does not make any difference if we are handling a continuous distribution.Let us define say u is equal to the minimum and v is equal to X n, then the claim is that u is a consistent estimator for a and V is a consistent estimator for b.So, the proof will use the steps which we have derived just now that is the c d f of u that is 1 minus so actually, it is equal to 0 for x less than a it is equal to 1 minus 1 minus f u to the power n for a less than or equal to x less than b it is equal to 1 for x greater than or equal to b.

(Refer Slide Time: 22:44)



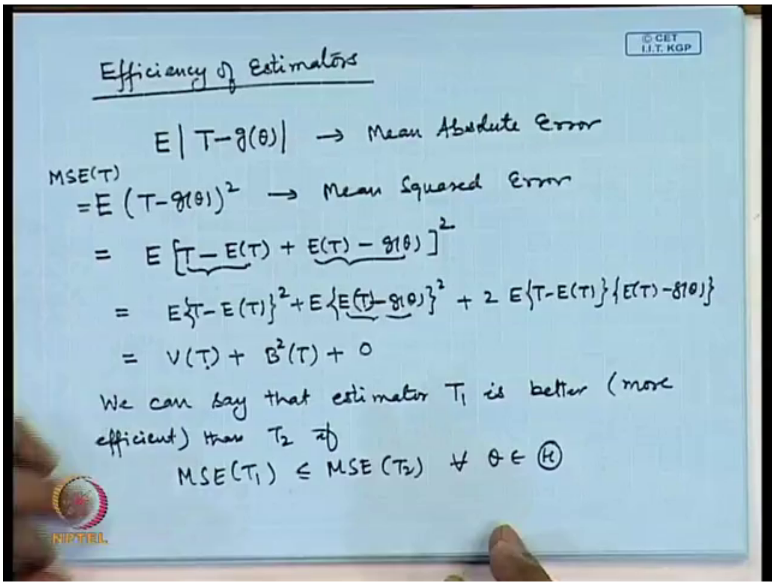
Similarly, if I consider say F v then it is equal to x 0 for x less than a it is equal to F v to the power n for a less than or equal to x less than b it is equal to 1 for x greater than or equal to b.Notice here that this equality or inequality does not make any difference here, because it is continuous distribution.So, if I take the limits here now F is a number between zero to one so if I take the limits here this number will go to zero. So, this is going to 1 so if I take the limit as n tends to infinity of F u F am getting 0 for x less than a and it is equal to 1 for x greater than or equal to a.

So, this is c d f of a random variable which is simply degenerate at a so, we can conclude that u converges to a in distribution and therefore, u converges to a in probability, because convergence and distribution and probability are equivalent if the right hand side is a constant.Similarly, if I consider limit of F V as n tends to infinity then this is also 0 for x less than b and it is equal to 1 for x greater than or equal to b.So, once again V is converging to b this random variable is degenerated at b so V tends to b in distribution or V tends to b in probability.So, if a continuous distribution is having a range a to b then the smallest order statistics converges to the lowest value or the lowest value in the range.And the largest order statistic converges to so, these can be treated as the consistence estimators of these respective parameters.

So, this actually gives some easy applications basically for example, we want to find out a consistent estimator further range.For example, here range may be b minus a then easily you can say that v minus u that is the maximum minus the minimum sample range is the consistent estimator for the population range.As a consequence we conclude that the sample range that is X n minus X 1 is consistent for the population range b minus a.We have some further special cases here for example, lower limit could be minus infinity or the upper limit could be plus infinity in that case for example, if the lower limit is minus infinity then X 1 does not converge in probability.Similarly, if the highest value is unbounded that means, b is infinity then X n does not converge in probability. (No audio from 26:31 to 26:39)

Next we introduce the concept of efficiency as we have seen that there can be situations where we have more than one consistent estimator we may have more than one estimator, which is unbiased as well as consistent. So, in that case we introduced the concept of efficiency of estimators. (No audio from 27:05 to 27:15)

(Refer Slide Time: 27:05)



Thefor judging the efficiency of the estimators we consider something called as expected error.We have seen unbiasedness so, in unbiasedness we had expectation of T is equal to the given parametric functions say g theta. So, if it is not unbiased expectation of T minus g theta is a biased or you can say expected error, but in the there is a danger in using biased as a simple you know criteria for a goodness of an estimator.Because sometimes, the negative bias and the negative errors and the positive errors may cancel out each other so, on the average the estimator may become unbiased. But actually it is not a good estimator.

We have seen the examples for example, in the estimation of e to the power minus three lambda we had an estimator minus 2 to the power x in Poisson distribution, which was stating values always away from the range, but the errors where positive and negative both very large errors and they were cancelling out each other. So, simply using expectation of x minus g theta that is biased as a measure is a dangerous thing.So, one may look at other measures for example, why not consider absolute error and then take expectation. So, one may consider expectation of say T minus g theta absolute value so this is called the mean absolute error or one may consider expectation of T minus g theta whole square, which is called the mean squared error.

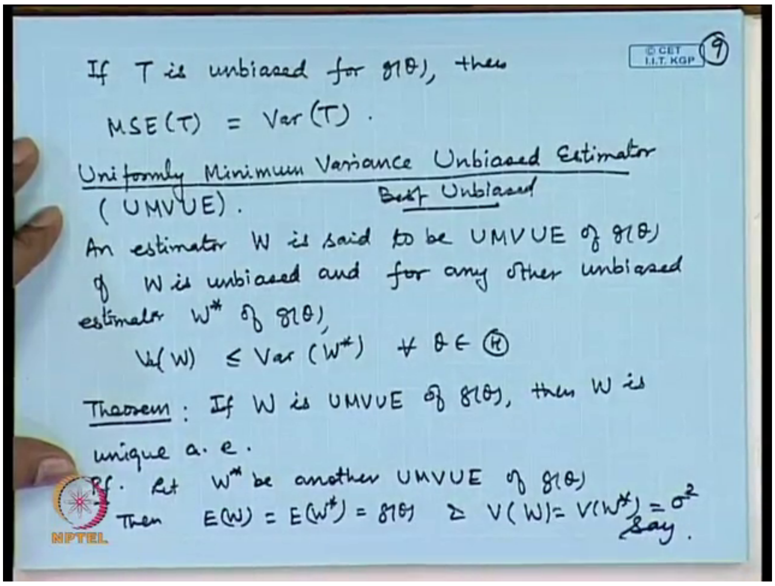
So, I will pay some attention to this in the definitions in the first case we are simply looking at the amount or you can say magnitude of the error that we have committed in estimating g theta by T and then we take the average of that.In the second one we are considering the squares so, if you think as a layman then probably you will feel that the first one is an appropriate measure for the error or you can say average error.However, in practice the evaluation of expectation of modulus T minus g theta is quite complex.Second point is that if we look at mathematically this function is not easy to handle, the main problem is that modulus function is not a smooth function, because it is having a corner that is at T is equal to g theta it is not smooth.

Whereas if we look at the mean squared error it is easy to evaluate and it has a simple interpretation which is quite so, what I do I add and subtract expectation t here. So, let us consider this as one term and this as one term so this becomes expectation of t minus expectation of T square plus expectation of T minus g theta expectation of that square plus twice expectation T minus expectation T into expectation T minus g theta.So, let us look at these terms the first term is simply the variance of T.The second term is fixed term so expectation will be the same value, because we already taking an expectation here this term is nothing but the bias of the estimator T.

And if you look at the cross product term here then this term is a constant so expectation applies to this and this becomes 0.So, we have that mean squared error let me call it MSE of T that is equal to variance plus the biased.Now, this is quite significant interpretation, if I have two estimator say T 1 and T 2 and we only say that variance of T 1 is less than variance of T 2, then we are controlling only one quantity. However, it may turn out that there is another estimator say T three which is which may be actually biased, but it is variance is much less so that the overall mean squared error is smaller. So, the average squared error will be less so, one can use mean squared error is as a good criteria for judging the goodness of an estimator.

So, we will say that we can say that estimator say T 1 is better, which is actually a terminology for more efficient than T 2 if mean squared error of T 1 is less than or equal to mean squared error of T 2 for all theta. So, if the two mean squared errors are equal then they will be same.Now, in the context of unbiased estimation this concept of mean squared error being smaller is equivalent to variance being smaller.

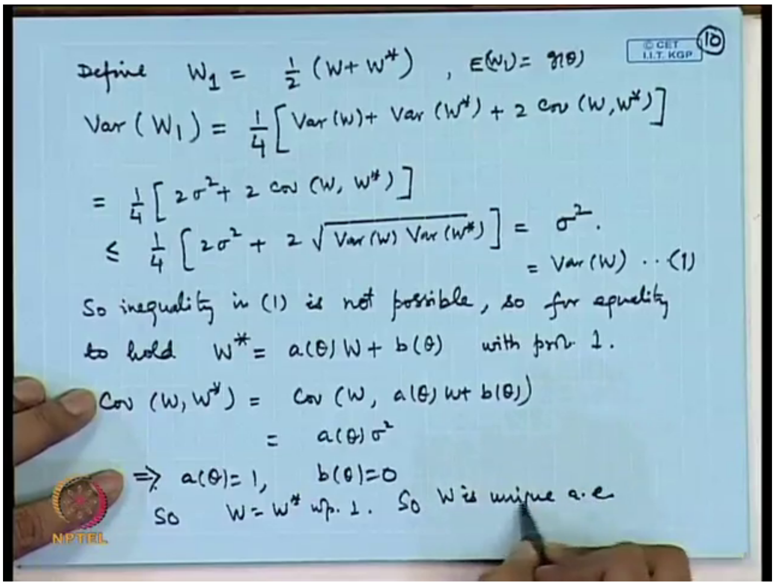
(Refer Slide Time: 33:41)



For example, if the estimator T is unbiased then bias will be 0 and this means squared error will be equal to the variance. If T is unbiased for g theta then mean squared error of T is called to be variance of T.Now, we define uniformly minimum variance unbiased estimators that is UMVUE.So, an estimator w is said to be UMVUE of a g theta if w is unbiased and for any other unbiased estimator say W star of g theta variance of W will be less than or equal to variance of W star that means, it will have the minimum variance throughout the parameter a space.

The first result in this direction is about the uniqueness of the UMVUE, if so we also use the terminology best unbiased estimator etcetera. So, if W is UMVUE of say g theta then W is unique almost everywhere. So, let W star b another UMVUE then by definition expectation of W expectation of W star both are same as g theta and variance of W and variance of W star are also same let us call it say sigma square fine.

(Refer Slide Time: 36:41)



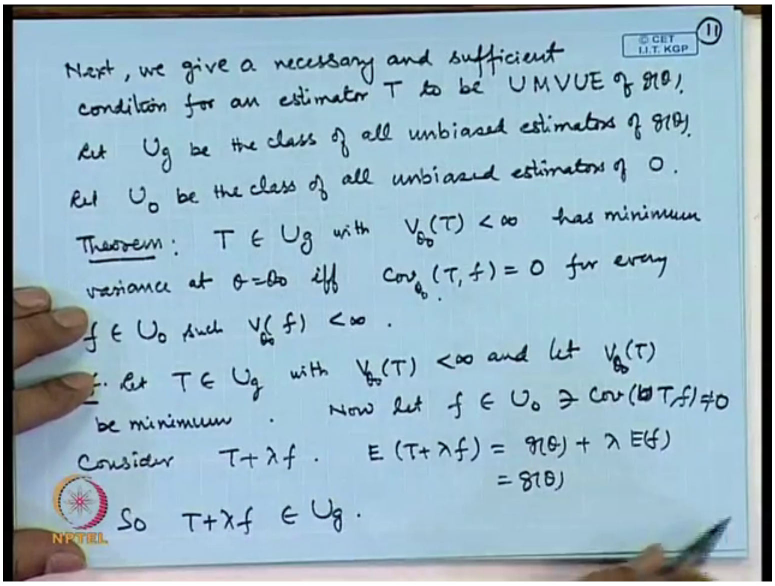
Now, let me define say W 1 as half W plus W star then what is the variance of W 1?We can apply the formula for a linear combination of variables so, variance of a constant times that is that constant square times variance of W plus W star, which is becoming variance of W plus variance of W star plus twice covariance between W and W star. Now, we are assuming variance of W and variance of W star to be sigma square so it becomes 1 by 42 sigma square plus twice covariance W W star.Now, covariance square is less than or equal to the product of the variances the well known Cauchy-Schwarz inequality so, this becomes 1 by 42 sigma square plus twice square root of variance W into variance of W star, but these are both sigma squares so this is simply becoming sigma square so 2 sigma square plus 2 sigma square so it becomes sigma square, which is the variance of W or W star.

So, what we are proving?If W is UMVUE W star is another UMVUE then I am able to get another estimator W 1 which is also unbiased, because if I take expectation of W 1 here that is again g theta as both W and W star are unbiased and it is variance is less than or equal to the variance of W.So, let me call this equation number 1 inequality in 1 is not possible, because our original claim is that W and W star are UMVUE so another unbiased estimator cannot have variance less than them. So, at the most it can have equal so that means we should have equality.

Now, how this inequality came inequality came from this condition of the correlation between w and w star being less than 1 so that means, correlation must be 1 that is covariance is equal to the square root of the variances that means, W and W star are linearly related with probability one. So, for equality to hold W star must be linearly related to W with probability one.Now, once again you have unbiasedness so, if you are saying unbiasedness then what should be the condition here and also, if I look at say covariance here between W and W star, then that is equal to covariance between W and a theta W plus b theta so that is equal to a theta into sigma square.

So, that means because this covariance between W W star is equal to variance W so, a theta is 1 and b theta will be 0 because unbiasedness is there, because expectation W star must be a theta so that is simply becoming g theta plus b theta so b theta must be 0.So, what we are concluding here that W is equal to W star with probability 1 that means, W is unique almost everywhere. So, you cannot have two different unbiased UMVU E if they are two different then they are equal almost everywhere.

(Refer Slide Time: 41:29)



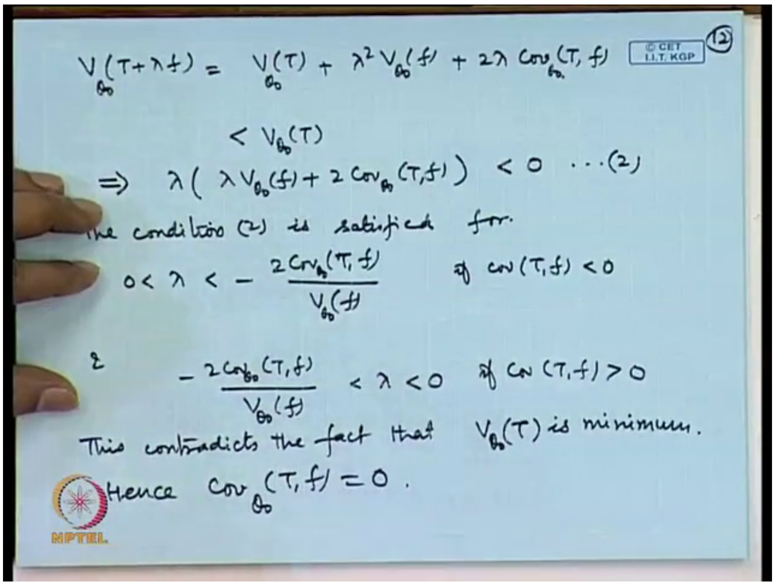
Now, next I give a necessary and sufficient condition for an estimator to be UMVUE.

(No audio 41:24 to 41:57)

So, let us consider let U g be the class of all unbiased estimators of g theta let U 0 be the class of all unbiased estimators of 0. So, we have the following necessary and sufficient condition so,T belongs to U g with variance of t to be finite so this has minimum variance at theta is equal to theta naught, if and only if covariance of t with say f is 0 for every f belonging to U 0 for, which variance of f is finite.That means, if an estimator is having covariance 0 that means, it is uncorrelated with every unbiased estimator of 0 then this will be UMVUE of a function g let me prove this here. So, let T the unbiased estimator of g and it is variance be finite and let variance theta naught T be minimum.

Now, let us consider f belonging to U 0 such that covariance between T and F is not 0. So, I am assuming contrary to what we have to prove so we will arrive at a contradiction. So, let us consider say T plus lambda F now if I take expectation of T plus lambda F then it is equal to expectation T that is g theta plus lambda times expectation F that is 0 so it is equal to g theta. So, this new function which I have created T plus lambda F is also unbiased.

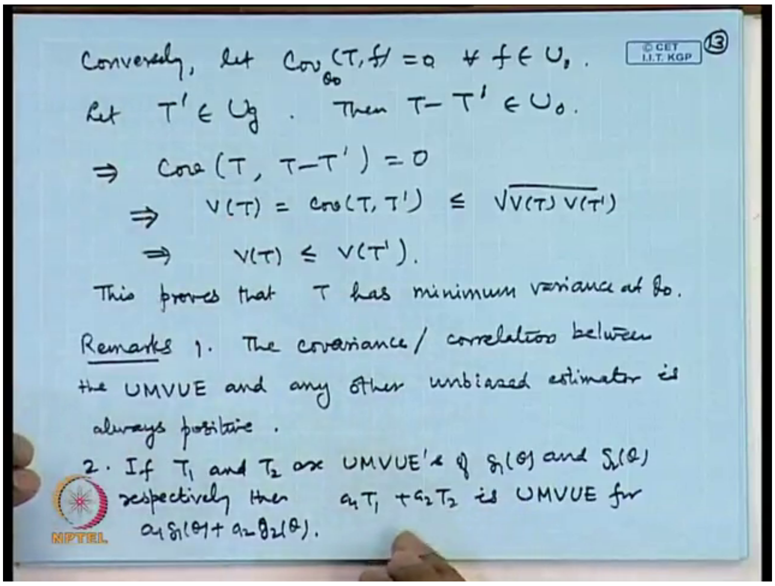
(Refer Slide Time: 45:03)



Now, let us take variance of T plus lambda f so that is equal to variance of T plus lambda square times variance of f plus twice lambda covariance between T and f.Now, if I put a condition here that this is less than variance of theta naught T, then this thing cancels out and it is reducing to a quadratic being less than zero.That means, this condition is equivalent to lambda into lambda V theta naught f plus twice covariance theta naught T f less than 0.So, this condition obviously, can be satisfiedthe condition 2 is satisfied for 0 less than lambda less than minus twice covariance T f by variance f of course, all these evaluations are at the point theta naught.

If covariance of T and f is negative and for minus twice covariance theta naught T f by variance theta naught f less than lambda less than 0 if this is positive.That means whatever, be the value of covariance between T and f whether it is positive or negative I am able to obtain a range of lambda values, such that the variance of T plus lambda f is less than variance of T this is a contradiction to the fact that I assumed that variance of t is minimum at theta naught. So, where is the mistake?The mistake is that I am assuming that covariance between T and f is not 0 so this is wrong.So, this contradicts the fact that variance theta naught T is minimum hence, we must have covariance between T and f equal to 0.

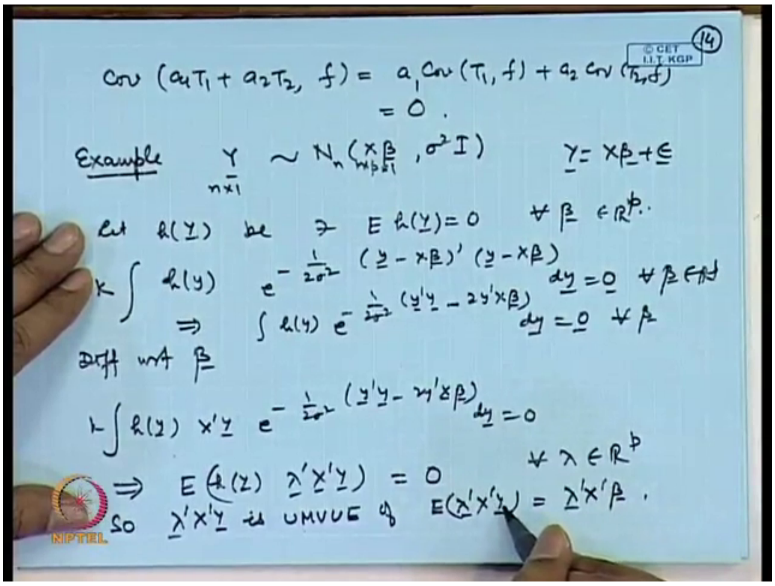
(Refer Slide Time: 48:05)



Now, let us take the converse of this conversely let cos theta between T and into variance of T prime.So, obviously, this is equivalent to saying that variance of T is less than or equal to variance of T prime. So, if I am taking covariance of T to be 0 with every unbiased estimator of 0 and I am taking another unbiased estimator t prime of g theta then I m getting that, the unbiased the variance of T is less than or equal to variance of T prime, this proves that T has minimum variance at theta.Another thing which you can conclude from here I have proved that if t is UMVUE then covariance between T and T prime that is equal to variance of T that means this is always positive.

So, we are also concluding from here that the covariance or you can say correlation between the UMVUE and any other unbiased estimator is always positive.And other interesting property about the UMVUE is that, if T 1 and T 2 are U M V U Es of g 1 theta and g 2 theta respectively then a 1 T 1 plus a 2 T 2 is UMVU E for a 1 g 1 theta plus a 2 g 2 theta.That means some sort of linearity property is also true for the UMVUE, although it is true for the unbiased estimation, but it is not clear that it will be true for UMVUE, but that is true here in fact 1 can look at a very simple proof of this.

(Refer Slide Time: 51:05)



If I consider say covariance of a 1 T 1 plus a 2 T 2 with an unbiased estimator of 0 then it is equal to a times covariance between T 1 and f plus a 2 times covariance between T 2 and f.Now, if T 1 and T 2 are UMVUE these are 0 so this is simply 0 so by the previous theorem this result follows.As an application of this theorem let us consider linear model and try to obtain UMVUE.So, let us consider the gauss Markov linear model so y is an n by 1 vector with mean x beta and variance covariance matrix as sigma square y.So, actually it is the part of the gauss Markov linear model where we write it as plus epsilon and epsilon follows normal 0 sigma square I.

So, let us consider say h y be a real valued function such that expectation of h y is say 0 for all beta this may be say n by p this may be p by 1 etcetera. So, if you write this statement expectation h y is 0 it is equivalent to h y into the density function of y this is the multivariate normal distribution so, it is e to the power minus 1 by 2 sigma square y minus x beta prime y minus x beta.And some coefficient will come which I am writing as a constant this is equal to 0 for all beta belonging to r p this is a multivariate integral here.

Now, you differentiate both the sides with respect to beta then I will get h y then derivative of this will give this term into the derivative of this with respect to y that gives me x prime y e to the power minus 1 by 2 sigma square.In fact here I can simplify beforehand I can write the term which is not involved in why I cannot separate out and take to the other side.So, this is reducing to so if you differentiate this you will get x prime y here and the same term here. So, this is equivalent to saying expectation of h y into some coefficient lambda prime x prime y is equal to 0 for all lambda belonging to r p.

So, what is this one this is a linear function so, by the previous theorem what we are saying is that lambda prime x prime y is UMVUE of expectation of lambda prime x prime y that is lambda prime x prime beta.In the gauss Markov theory of linear models, we had proved that lambda prime x prime y is the best linear unbiased estimator of lambda prime x prime beta.Here we are proving that it is not only best linear unbiased it is actually, best unbiased that is it is the UMVUE for this.Although I have made a small mistake here it is lambda prime x prime x beta so for this it is becoming best unbiased estimator.

In the forthcoming classes we will consider methods for finding out estimators just now in the previous two classes; we have considered the properties of the estimator some desirable criteria. However, there must be some methods by which we can derive these estimators. So, we will do some well known methods and as well as we will explore further how to find out the best unbiased estimators.