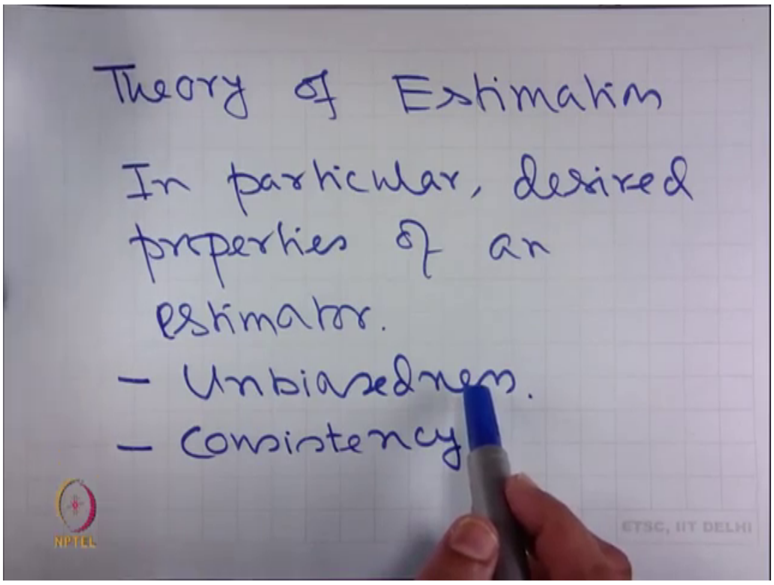
Statistical Inference
Prof. Niladari Chatterjee
Department of Mathematics
Indian Institute of Technology, Delhi

Lecture – 13
Statistical Inference

Welcome students to the MOOCS lecture series, on Statistical Inference. This is lecture number 13 and you know that we are currently discussing Theory of Estimation.

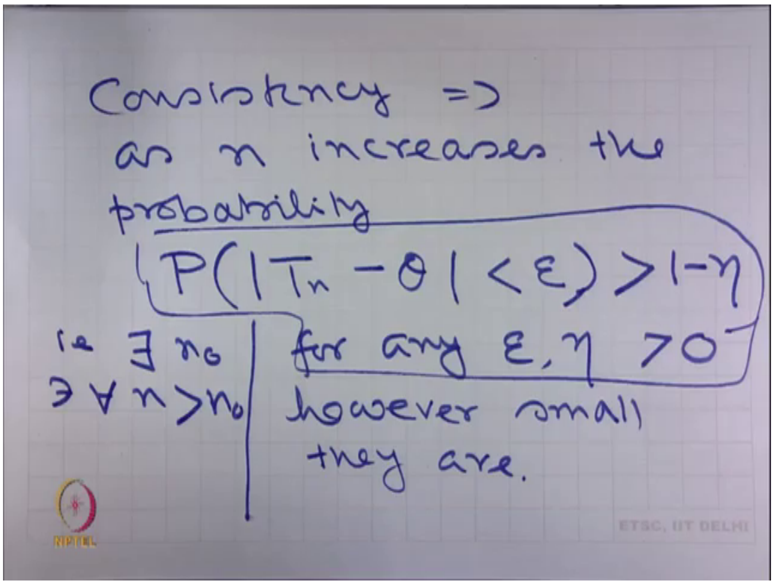
(Refer Slide Time: 00:36)



In particular we are looking at properties of or desired properties of an estimator, in this respect, we have discussed two properties namely unbiasedness and consistency and estimator T x 1 x 2 xn is said to be unbiased to estimate parameter theta, if the expected value of T is theta.

And we have seen that there are many unbiased estimators, but all of them are not consistent. Why consistency is important?

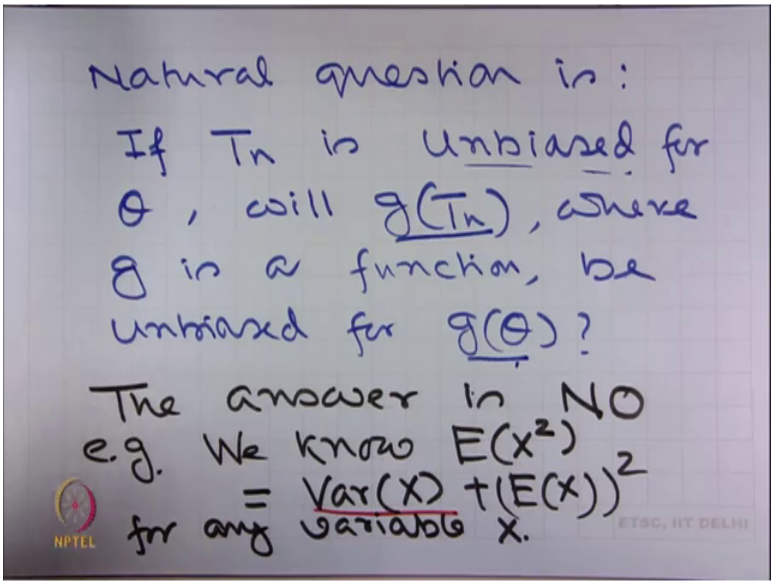
(Refer Slide Time: 02:03)



Consistency implies as n increases the probability that is P of modulus of Tn minus theta less than epsilon becomes greater than 1 minus eta for any epsilon and eta greater than 0; however, small they are or in other words there exist N naught such that for all N greater than N naught this property holes; that means, after certain number, if we choose samples of size beyond that, then we know that the probability that the estimator Tn will come arbitrarily close to the parameter theta that it is estimating and that probability is going to be as close to one as possible.

And we have seen sufficient conditions for consistency that if the expected value of Tn converges to theta and the variance of T n goes to 0 as N goes to infinity, then we know that, that particular Tn is going to be consistent for theta.

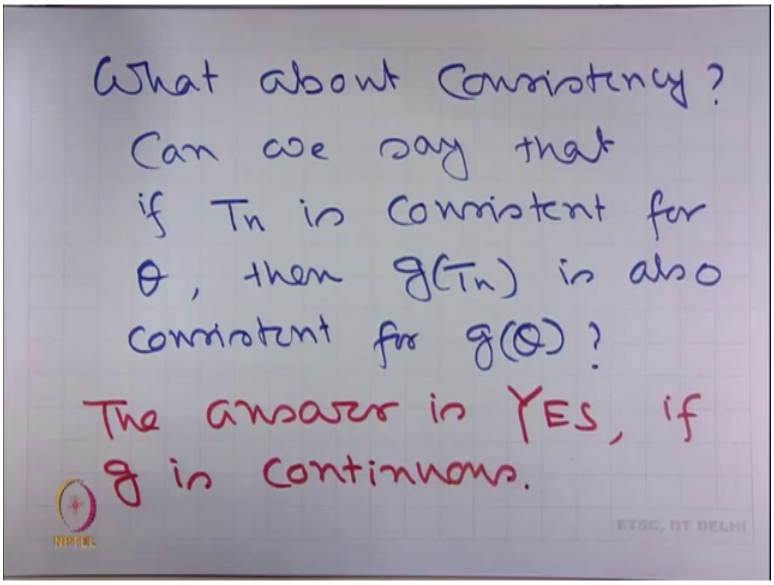
(Refer Slide Time: 04:21)



Natural question is, if Tn is unbiased for theta will g of T n, where g is a function be unbiased for g theta. It is a very natural question, because many a time we try to estimate not exactly the theta, but some function of theta. Can we use the unbiasedness property to get an unbiased estimator for g theta?

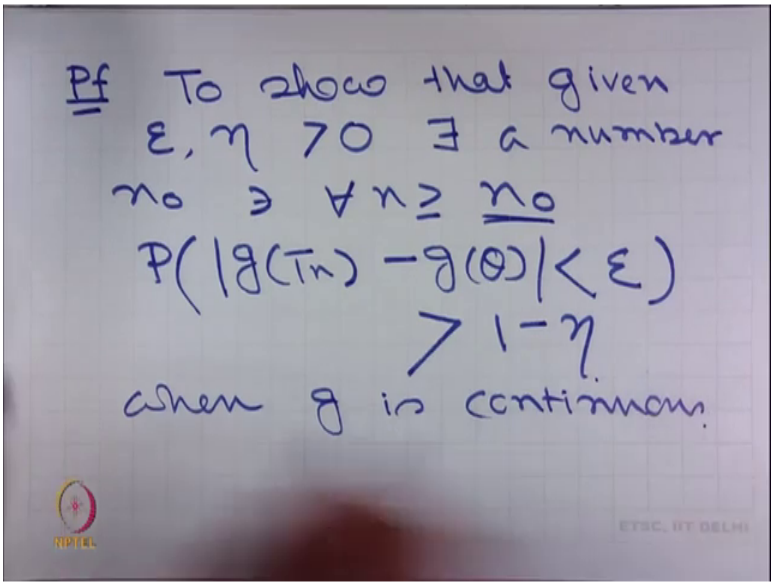
Unfortunately the answer is no. For example, we know expected value of X square is equal to variance of X plus expected value of X whole square for any variable X. Therefore, if we focus on Tn, an expected value of Tn is theta, we see that expected value of T n square is going to be variance of T n plus theta square and therefore, we can see that X square is not unbiased for theta square, because there is the bias in the form of variance of X. What about consistency?

(Refer Slide Time: 06:55)



Can we say that if Tn is consistent for theta then g of Tn is also consistent. For g theta and the answer is yes, if g is continuous proof to show that given epsilon and eta greater than 0, there exist a number n naught, such that for all n greater than equal to n naught probability.

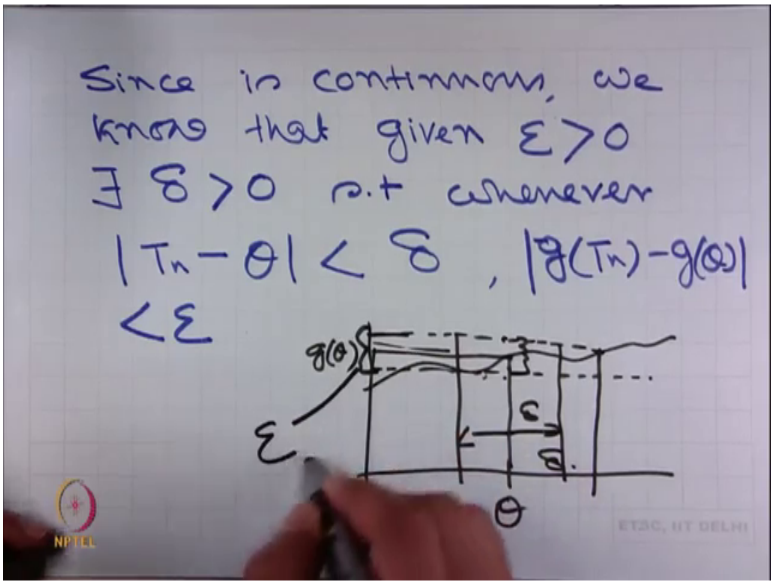
(Refer Slide Time: 08:08)



Modulus of g Tn minus g theta less than epsilon is greater than 1 minus eta.

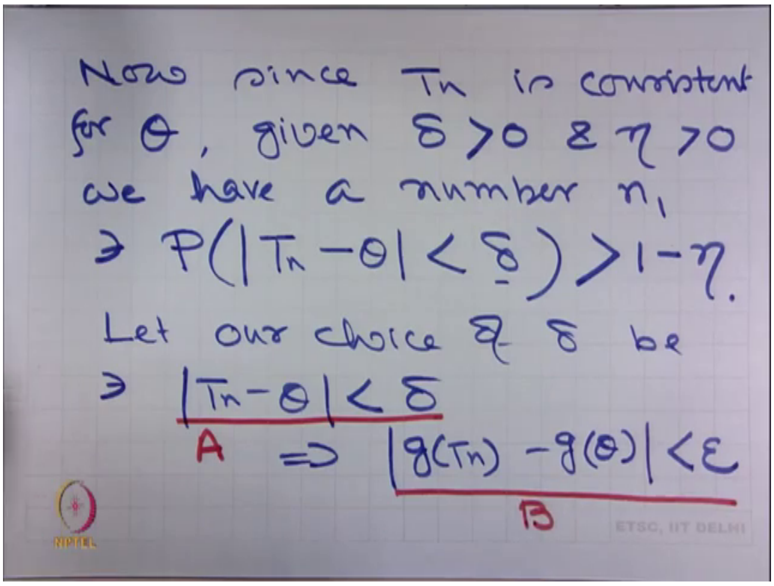
So, we have to identify such a n naught when g is continuous, since g is continuous by definition of continuity.

(Refer Slide Time: 09:19)



We know that given epsilon greater than 0 there exist delta greater than 0, such that whenever modulus of Tn minus theta is less than delta modulus of g Tn minus g theta is less than epsilon or suppose, this is theta and g is a continuous function and this is g theta and suppose, we create an epsilon neighborhood of around the g theta then we see that for all n, in this neighborhood sorry, for all values of the variable say X in the neighborhood. Let us call it delta, which may be like this. So, let us call it delta we can see that modulus of g theta minus g Tn is within the limit epsilon. So, this is our epsilon.

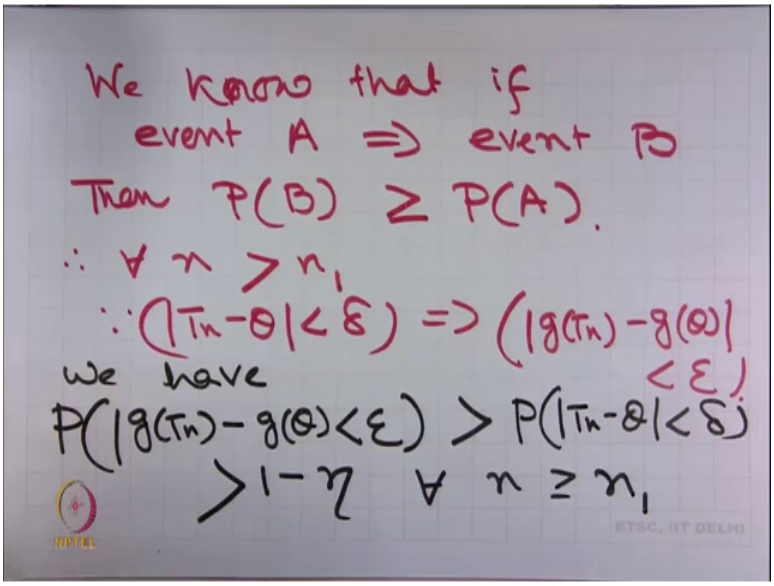
(Refer Slide Time: 11:48)



Now, since Tn is consistent for theta given delta greater than 0 and eta greater than 0, we have a number in one such that probability modulus of Tn minus theta, less than delta is greater than 1 minus eta. Let our choice of delta be such that Tn minus theta less than delta implies g Tn minus g theta less than epsilon.

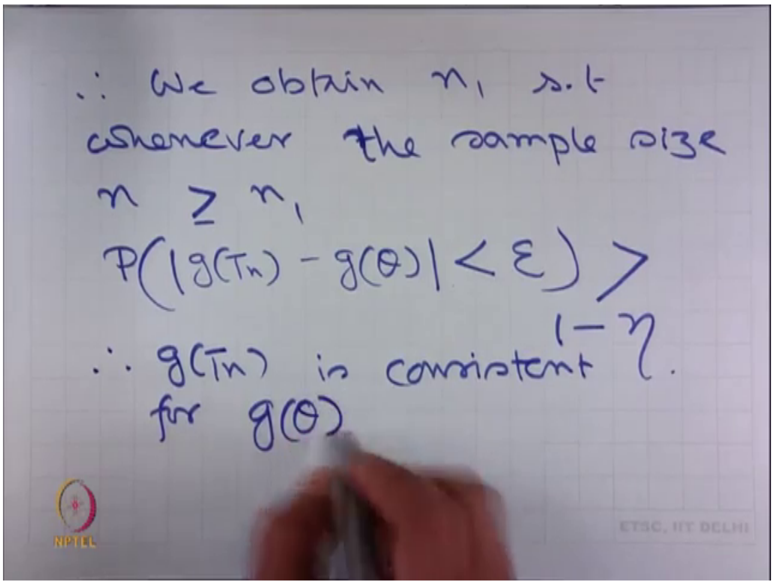
So, let us call this event A and let us call this event B.

(Refer Slide Time: 13:31)



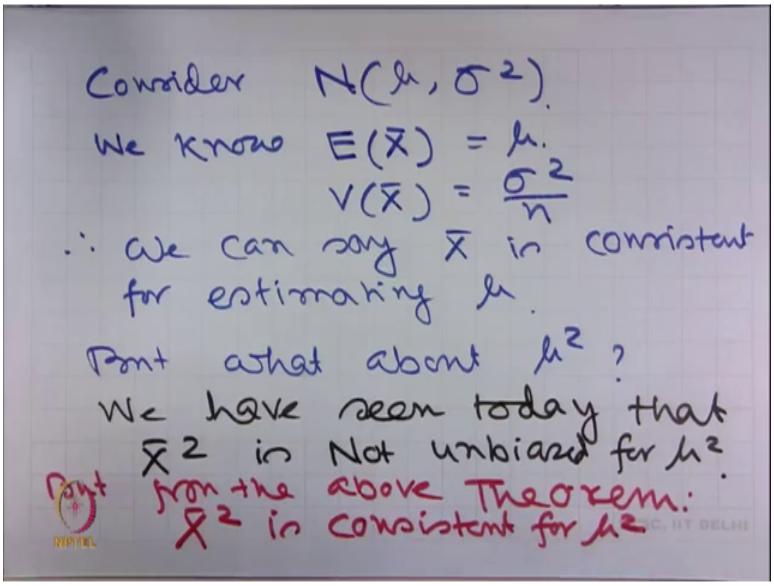
We know that, if event A implies event B then probability of B is greater than equal to probability of A. Therefore, for all n greater than n 1 since Tn minus theta less than delta implies g Tn minus g theta less than epsilon, we have probability g Tn minus g theta, less than epsilon is greater than probability Tn minus theta less than delta, which is greater than 1 minus eta for all n greater than equal to n 1.

(Refer Slide Time: 15:07)



Therefore, we obtained n 1, such that whenever this sample size n is greater than equal to n 1 the probability g Tn minus g theta less than epsilon is greater than 1 minus eta therefore, g Tn is consistent for g theta. What is the advantage consider?

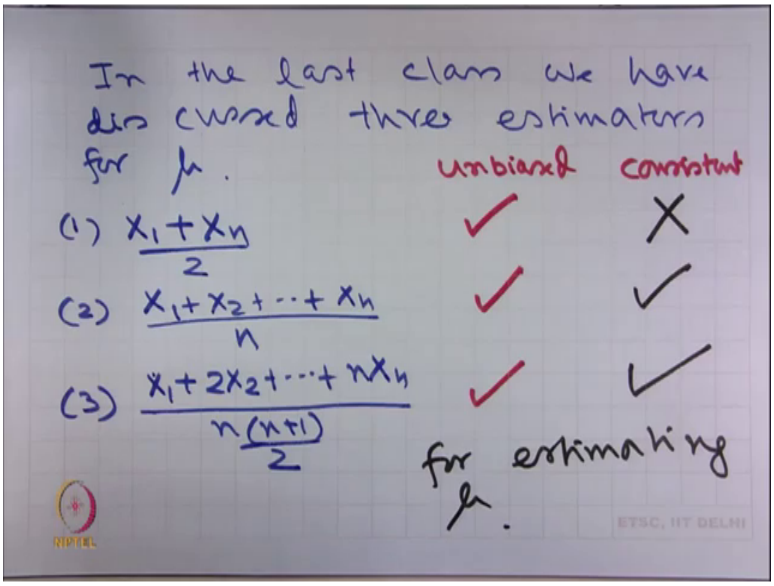
(Refer Slide Time: 16:05)



Normal mu sigma square, we know that the expected value of X bar is equal to mu variance of X bar is equal to sigma square by N. Therefore, we can say X bar is consistent for estimating mu, this we get from the theorem that approved in the previous class, but what about mu square?

We have seen today that X bar square is not unbiased for mu square, but from the above theorem what we get that X bar square is consistent for mu square.

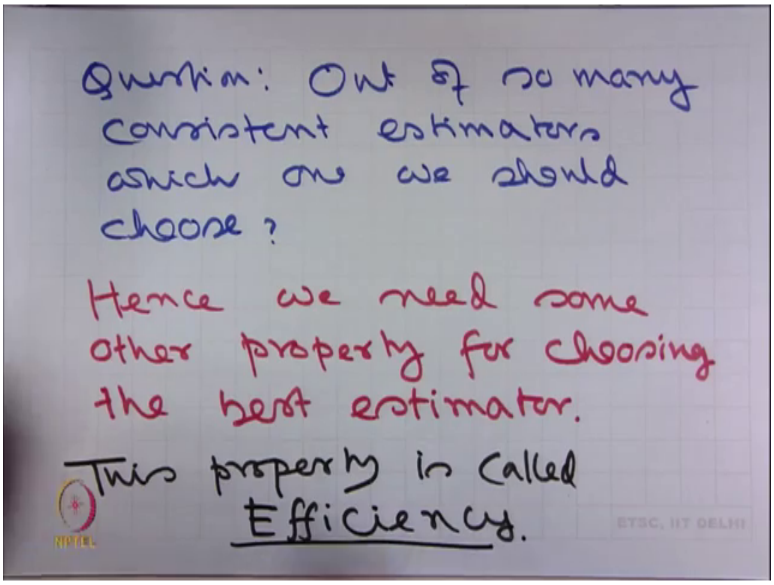
(Refer Slide Time: 18:15)



In the last class, we have discussed three estimators for mu one is, X 1 plus say Xn by 2, second is X 1 plus X 2 plus Xn by n and third one is X 1 plus 2 X 2 plus up to n Xn by n into n plus 1 by 2.

If we look at these three estimators, we found that this is unbiased, this is unbiased and this is also unbiased in estimating mu, but this is not consistent, but this is consistent and this is consistent for estimating mu, which you know is the population mean. Therefore, there is a question; we have many unbiased estimators, among them, some of them are consistent and that can be a large number, because we can choose estimators in such a way. So, that they become consistent, that is they are various reduces to 0 as n goes to infinity. Therefore, the question is out of so, many consistent estimators which one we should choose?

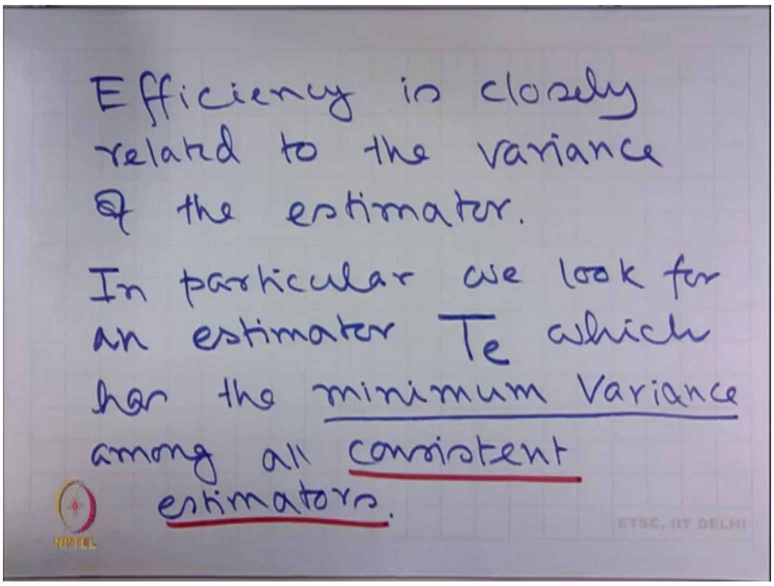
(Refer Slide Time: 20:37)



This is a very pertinent question, because we have been given several estimators, we need to choose the best one among them and we have already observed that, just unbiasedness and consistency, they do not give us a single choice of estimator to estimate a particular parameter theta. Just that in the previous case, we have used theta is equal to mu the population mean.

Therefore, we need some other property for choosing the best estimator, this property is called efficiency.

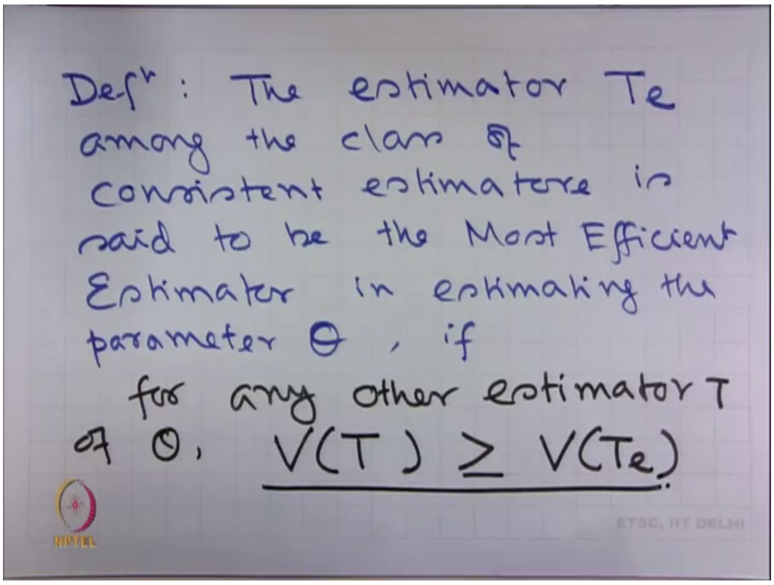
(Refer Slide Time: 22:44)



Now, efficiency is closely related to the variance of the estimator, if the variance is less; that means, if the square of the deviation from the expected value is less then the advantage is that we are more sure that we are closer to the estimate and therefore, the more efficient an estimator is the less is it is variance.

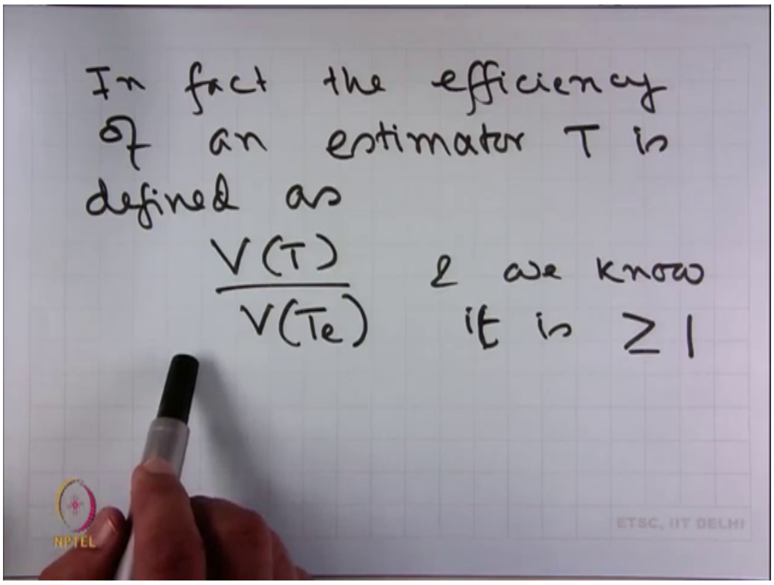
In particular, we look for an estimator Te, which has the minimum variance among all consistent estimators and it makes sense that we are only searching minimum variance among the consistent estimators, because if an estimator is not consistent, then we cannot say that it is coming close and close to the parameter theta as n goes to infinity, because this variance is not decreasing as n increases.

(Refer Slide Time: 25:17)



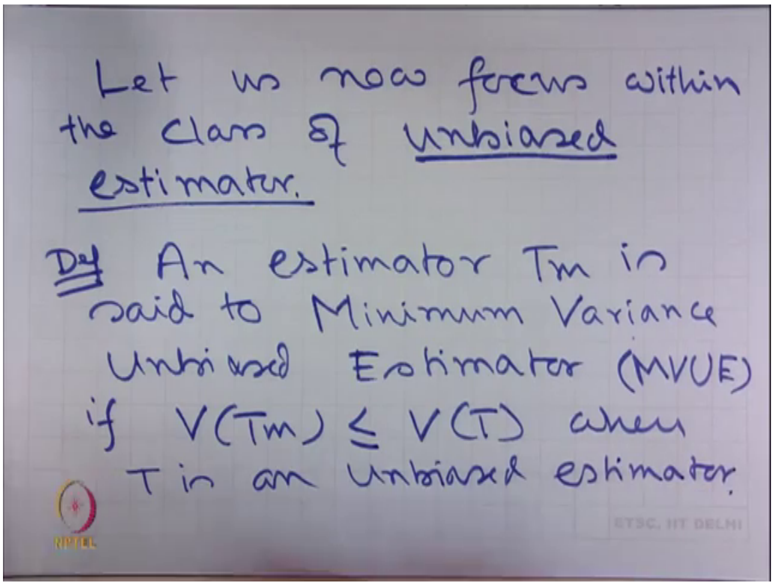
The estimator Te among the class of consistent, estimators is said to be the most efficient estimator in estimating the parameter theta. If for any other estimator T of theta, variance of T is greater than; equal to variance of T e that is the estimator T, whose variance is the minimum among all other consistent estimators that is called the minimum variance estimator or most efficient estimator.

(Refer Slide Time: 27:17)



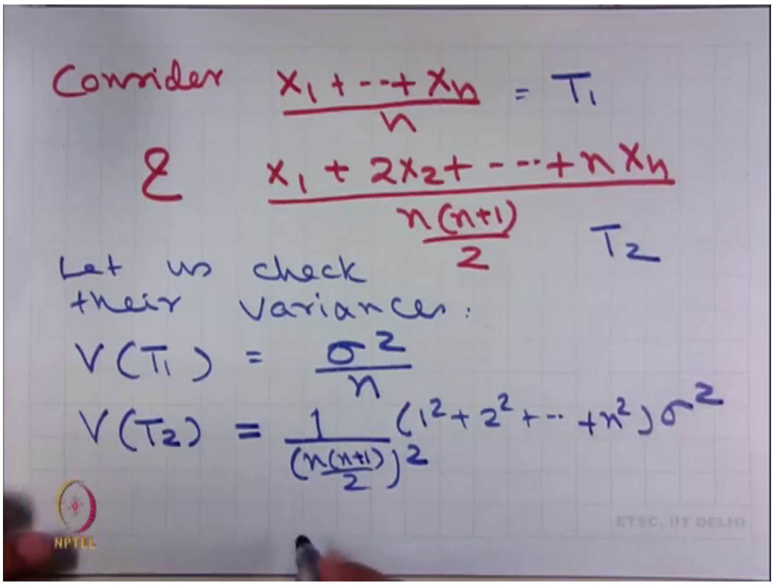
In fact, the efficiency of an estimated T is defined as variance of T divided by variance of Te and we know it is greater than equal to 1, it will be 1, if variance of T is same as variance of Te otherwise, as Te as the minimum variance, variance of T has to be greater than variance of Te and therefore, its efficiency has to be greater than equal to 1 for this class.

(Refer Slide Time: 28:33)



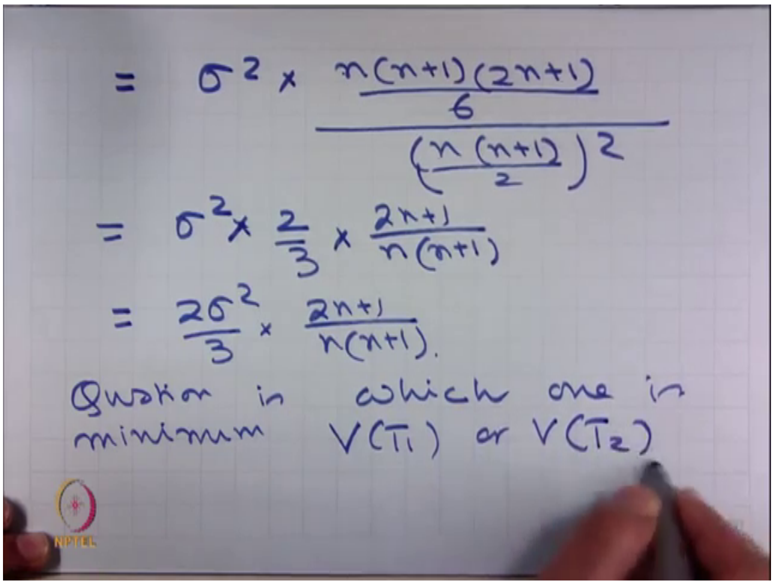
Let us now focus within the class of unbiased estimator and estimator. Tm is said to the minimum variance unbiased estimator that is MVUE, if variance of T m is less than equal to variance of T, where T is an unbiased estimator that is in the class of unbiased estimator, we are looking for an estimated Tm, whose variance is minimum.

(Refer Slide Time: 30:26)



Consider X 1 plus X n by n and X 1 plus 2 X 2 plus n X n by n into n plus 1 by 2. Let us call them T 1 and T 2 respectively. We know both are unbiased for mu, we know both are consistent. So, let us check they are variances, variance of T 1 as we know is sigma square by n and variance of T 2 is equal to, we have calculated in other day, but let me do it again. This is 1 upon n into n plus 2 by 2 whole square into 1 square plus 2 square up to n square sigma square, which is equal to sigma square into n into n plus 1 into 2 n plus 1 by 6 divided by n into n plus 1 by 2 whole square, which is equal to sigma square into 2 by 3.

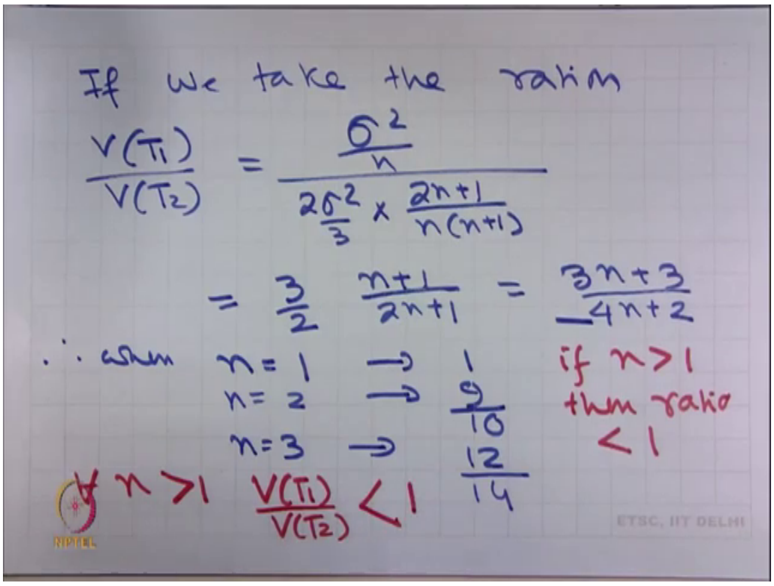
(Refer Slide Time: 32:10)



This is 4, which goes to the numerator 6 is there into 2 n plus 1 upon n into n plus 1 is equal to 2 sigma square by 3 into 2 n plus 1 upon n into n plus 1.

Question is, which one is minimum variance of T 1 or variance of T 2?

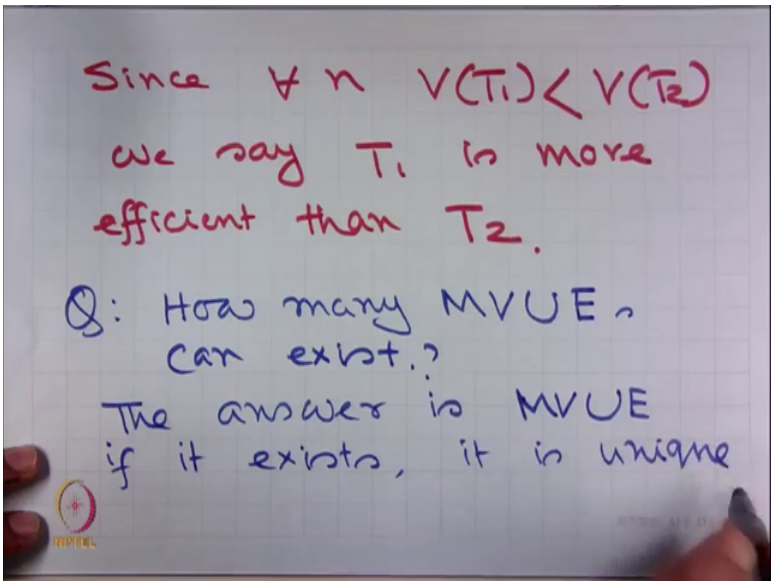
(Refer Slide Time: 33:37)



If we take the ratio variance of T 1 upon variance of T 2, this is equal to sigma square by n upon 2 sigma square by 3 into 2 n plus 1 upon n into n plus 1 is equal to 3 by 2 n plus 1 upon 2 n plus 1 is equal to 3 n plus 3 upon 4 n plus 2.

Therefore when n is equal to 1, the ratio is 6 by 6 is equal to 1, n is equal to 2. The ratio is 9 by 10, n is equal to 3. The ratio is 12 by 14 and we see that, if n is greater than 1, then ratio is less than 1 and as n increases. We can, we can see that the numerator is increased by 3, but denominator is being increased by 4 therefore, for all n greater than 1 variance of T 1 upon variance of T 2 is less than 1 or in other words.

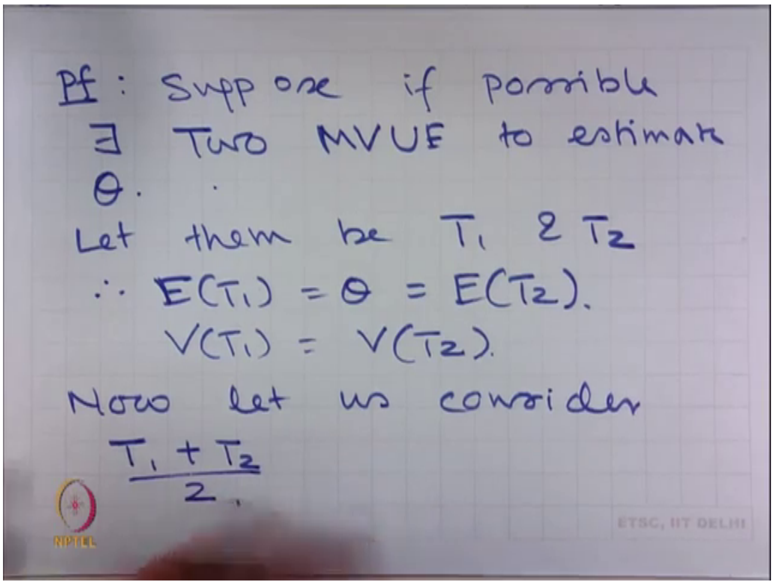
(Refer Slide Time: 35:59)



Since, for all n variance of T 1 is less than variance of T 2. We say T 1 is more efficient than T 2.

Question, the obvious question is how many minimum variance unbiased estimators can exist? Can there be more than one minimum variance unbiased estimator to estimate a parameter theta? The answer is MVUE; if it exist it is unique proof.

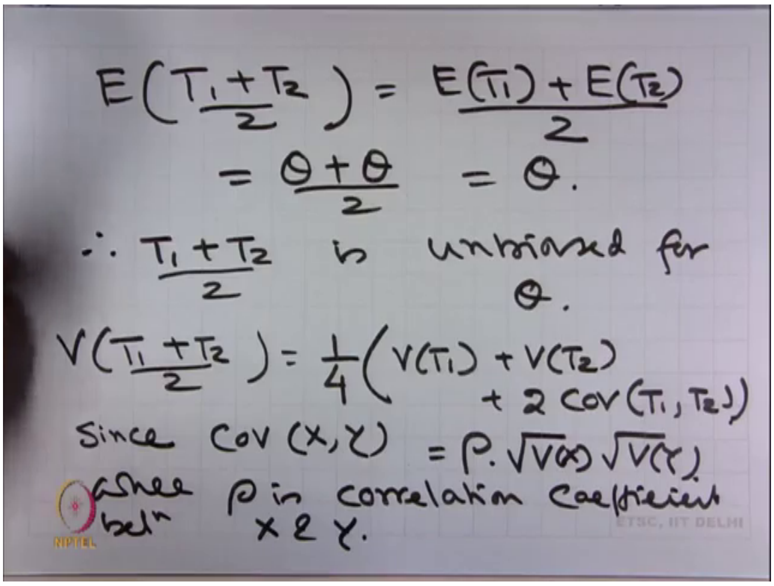
(Refer Slide Time: 37:24)



Suppose, if possible there exist 2 MVUV to estimate theta. Let them be T 1 and T 2 therefore, expected value of T 1 is equal to theta is equal to expected value of T 2 and variance of T 1 is equal to variance of T 2.

Now, let us consider T 1 plus T 2 by 2, expected value of T 1 plus T 2 by 2 is equal to expected value of T 1 plus.

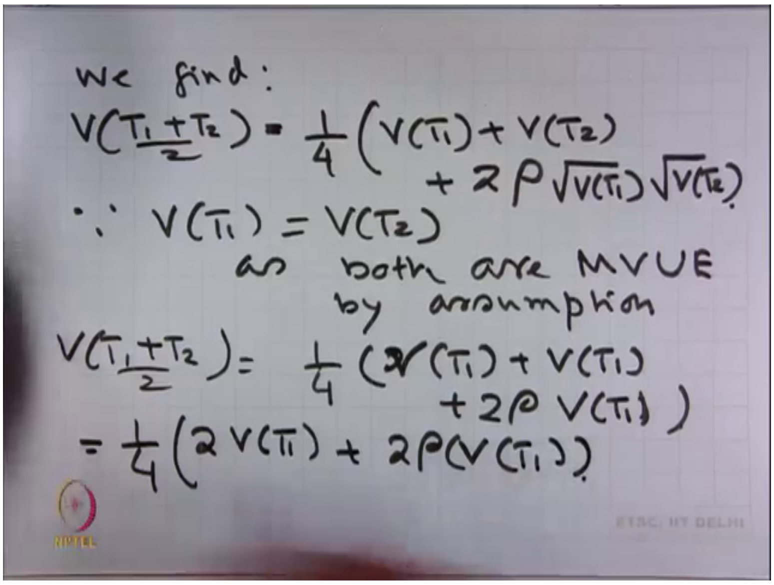
(Refer Slide Time: 38:35)



Expected value of T 2 upon 2 by linearity of expectation is equal to theta plus theta by 2 is equal to theta. Therefore, T 1 plus T 2 by 2 is unbiased for theta and variance of T 1 plus T 2 by 2 is equal to 1 by 4 times variance of T 1 plus variance of T 2 plus 2 times covariance of T 1 T 2.

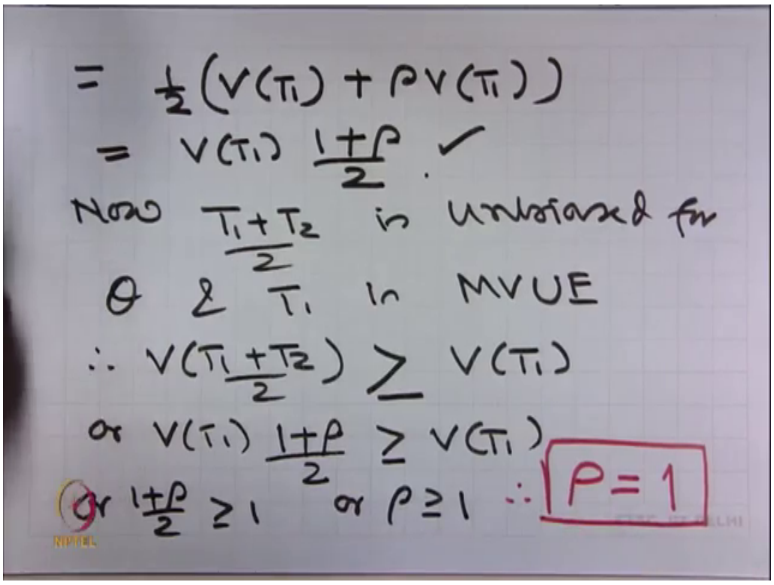
Since, covariance of X comma Y is equal to rho times root over variance of X root over variance of Y, where rho is correlation coefficient between X and Y.

(Refer Slide Time: 40:21)



We find variance of T 1 plus T 2 by 2 is equal to 1 by 4 variance of T 1 plus variance of T 2 plus 2 times rho root over variance of T 1 root over variance of T 2. Since, variance of T 1 is equal to variance of T 2 as both are MVUE by assumption variance of T 1 plus T 2 by 2 is equal to 1 by 4 into variance of T 1 plus variance of T 1 replacing variance of T 2 by variance of T 1 plus 2 rho. This is root over variance of T 1; this is root over variance of T 2. So, I can write it as variance of T 1 is equal to 1 by 4 times two variance of T 1 plus 2 rho into variance of T 1.

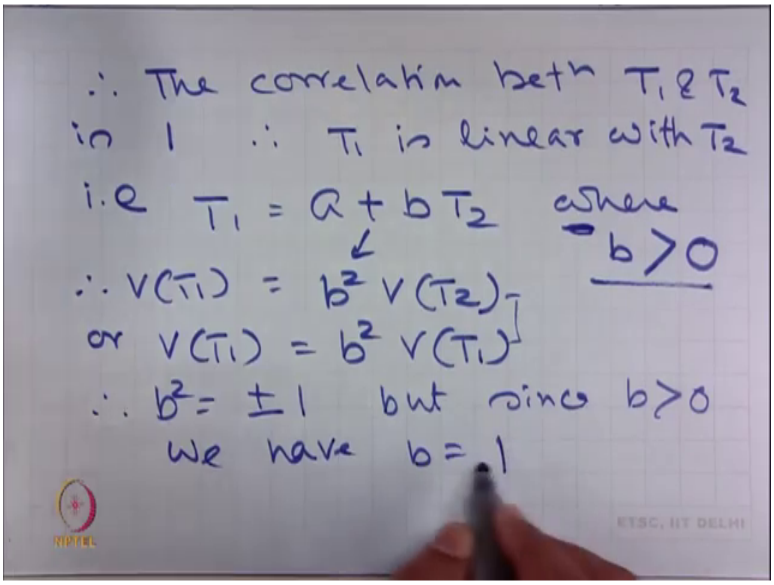
(Refer Slide Time: 42:01)



This is equal to half into variance of T 1 plus rho variance of T 1 is equal to variance of T 1 into 1 plus rho by 2.

Now, T 1 plus T 2 by 2 is unbiased for theta and T 1 is minimum variance unbiased estimator. Therefore, variance of T 1 plus T 2 by 2 is greater than variance of T 1 or variance of T 1 into 1 plus rho by 2 from here is greater than equal to variance of T 1 or 1 plus rho by 2 is greater than equal to 1 or rho greater than equal to 1 and since, rho is a correlation coefficient and therefore, the rho lies only between minus 1 to 1. We can say that rho is equal to 1. What does it mean?

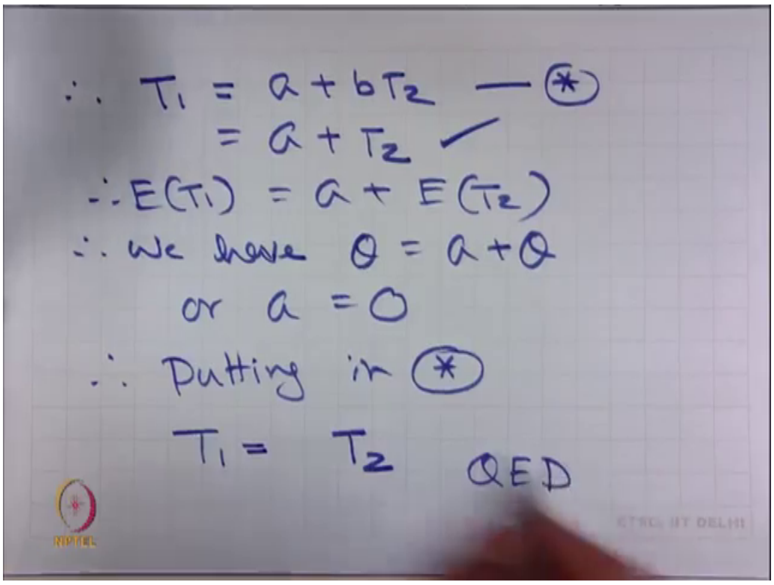
(Refer Slide Time: 43:50)



It means that the correlation between T 1 and T 2 is 1 and therefore, T 1 is linear with T 2 that is T 1 is equal to a plus b T 2, where b is greater than 0.

Therefore, variance of T 1 is equal to from here, b square times variance of T 2 or variance of T 1 is equal to b squared times variance of T 1 has both of them are MVUE therefore, they are equal therefore, b squared is equal to plus minus 1, but since b greater than 0. We have b is equal to 1.

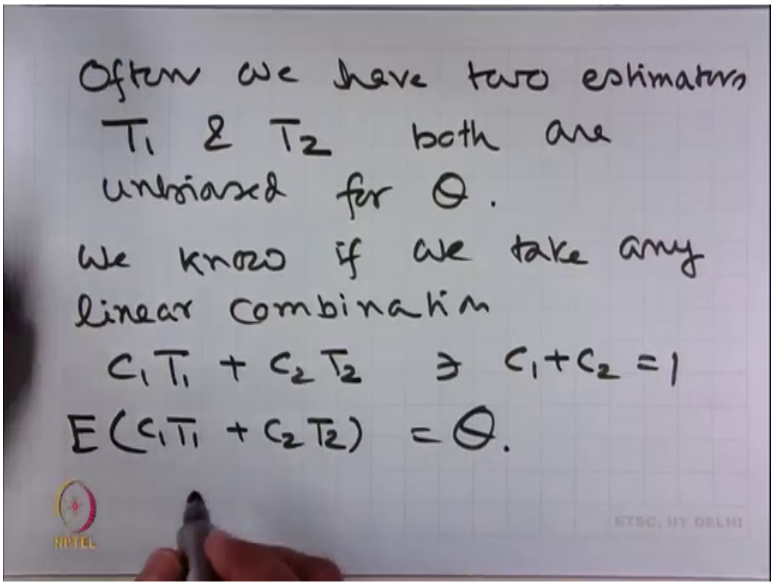
(Refer Slide Time: 45:17)



Therefore, T 1, which is equal to a plus b T 2 is actually a plus T 2 therefore, expected value of T 1 is equal to a plus expected value of T 2 and since, both are unbiased for theta therefore, we have theta is equal to a plus theta or a is equal to 0.

Therefore, putting in star T 1 is equal to T 2 Kvd; that means, there can be only one minimum variance unbiased estimator now often.

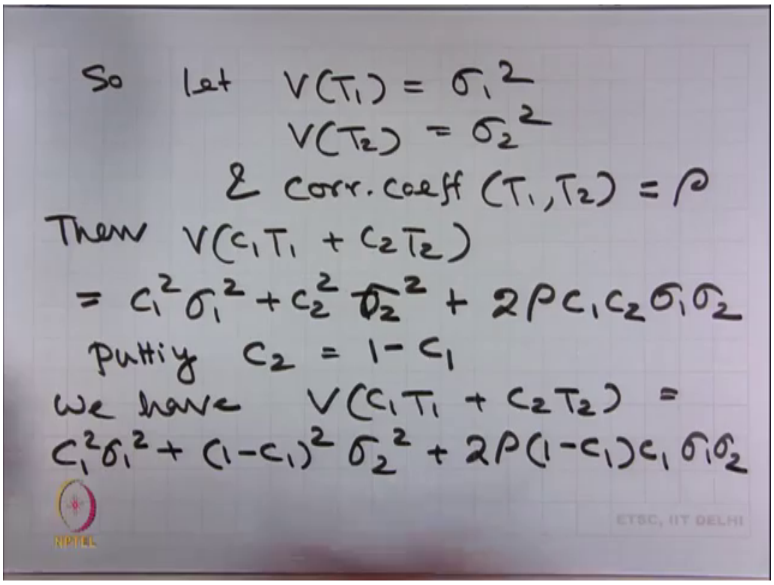
(Refer Slide Time: 46:25)



We have two estimators T 1 and T 2, both are unbiased for theta, we know if we take any linear combination C 1 T 1 plus C 2 T 2 such that C 1 plus C 2 is equal to 1 then expected value of C 1 T 1 plus C 2 T 2 is equal to theta therefore, given any two unbiased estimators. We can generate any number of unbiased estimators for theta by taking a linear combination C 1 plus C 1 T 1 plus C 2 T 2 and that is going to be unbiased.

The question is should we try all of them to find out which one among them is most efficient or in other words among the all, the linear combinations C 1 T 1 plus C 2 T 2, such that C 1 plus C 2 is equal to 1, which one should we take as the most efficient of them.

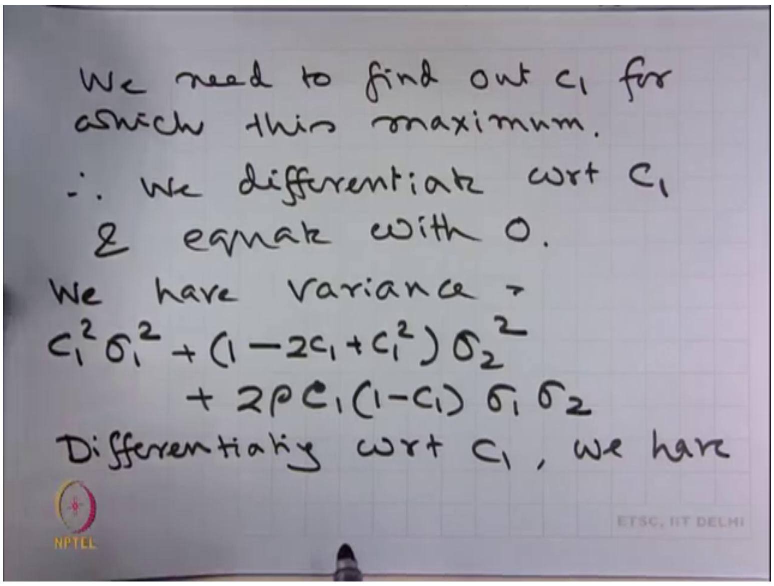
(Refer Slide Time: 48:23)



So, let variance of T 1 is equal to sigma 1 square variance of T 2 is equal to sigma 2 square and correlation coefficient of T 1 and T 2 is equal to rho then variance of C 1 T 1 plus C 2 T 2 is equal to C 1 square sigma 1 square plus C 2 square sigma 2 square plus 2 times rho C 1 C 2 sigma 1 sigma 2.

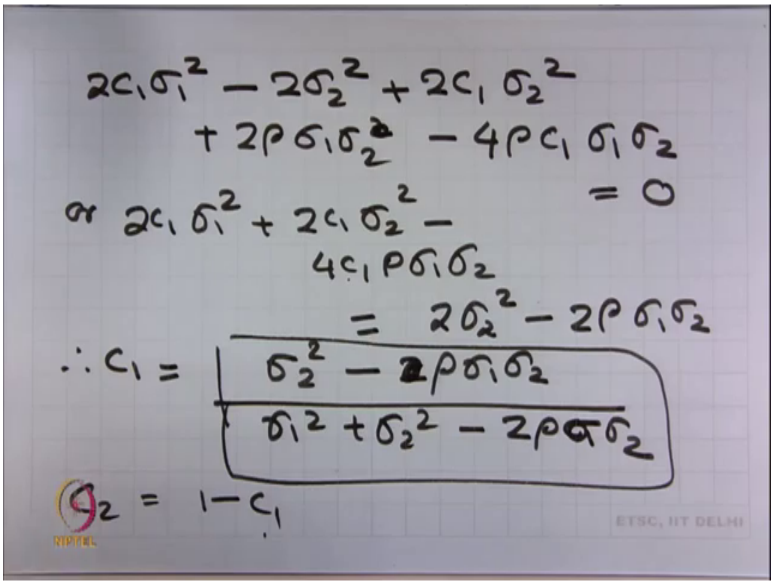
Putting C 2 is equal to 1 minus C 1, we have variance of C 1 T 1 plus C 2 T 2 is equal to C 1 square sigma 1 square plus 1 minus C 1 whole square sigma 2 square plus 2 rho 1 minus C 1 into C 1 sigma 1 sigma 2.

(Refer Slide Time: 50:13)



So, we need to find out C 1 for which this is maximum. Therefore, we differentiate with respect to C 1 and equate with 0, we have variance is equal to C 1 square sigma 1 square plus 1 minus 2 C 1 plus sigma 1 square plus C 1 square into sigma 2 square plus 2 rho C 1 into 1 minus C 1 into sigma 1 sigma 2.

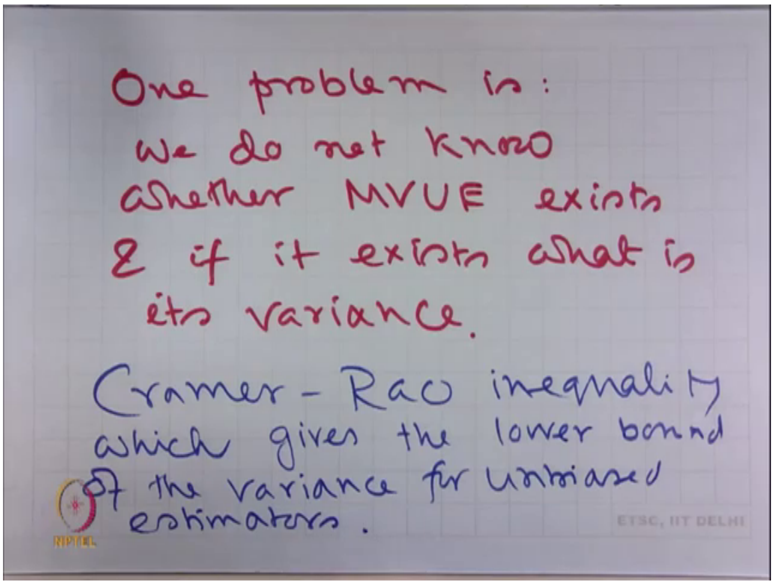
(Refer Slide Time: 51:54)



By differentiating with respect to C 1 we have 2 C 1 sigma 1 square minus 2 sigma 2 square plus 2 C 1 sigma 2 square plus 2 rho sigma 1 sigma 2 minus 4 rho C 1 sigma 1 sigma 2 is equal to 0 or 2 C 1 sigma 1 square plus 2 C 1 sigma 2 square minus 4 C 1 rho sigma 1 sigma 2 is equal to 2 sigma 2 square minus 2 rho sigma 1 sigma 2. Therefore, C 1 is equal to sigma 2 square minus 2 rho sigma 1 sigma 2 divided by sigma 1 square plus sigma 2 square minus 2 rho sigma 1 sigma 2 or this is the coefficient for T 1 and 1 minus C 1 should give you C 2, which is the coefficient for T 2 such that C 1 T 1 plus C 2 T 2 will have the minimum variance.

Therefore, when we have two such estimators, we can always choose in which linear combination we should take them so, that the variance of the summation is minimum.

(Refer Slide Time: 54:27)



One problem is, we do not know whether MVUE exist and if it is exist, what is it? Is variance to find the answer? To this we have something called Crammar Rao inequality, which gives the lower bound of the variance for unbiased estimators. This is not for all estimators, there is a class of estimators, which observe certain properties, which we call regularity or we say regularity assumption under those assumptions. Crammar Rao inequality gives a lower bound for the variance of the unbiased estimators, what is the advantage?

The advantage is that if we find a particular estimator, whose variance is equal to that bound, then we are sure that this is the minimum variance unbiased estimator, because there cannot be any other unbiased estimator, whose variance is lower than that of course, if we focus on the set of estimators, which observe the regularity assumptions. In the next class I shall deal with these and solve certain problems students, see you in the next class.

Thank you so, much.