Statistical Inference

Prof. Somesh Kumar

Department of Mathematics

Indian Institute of Technology, Kharagpur

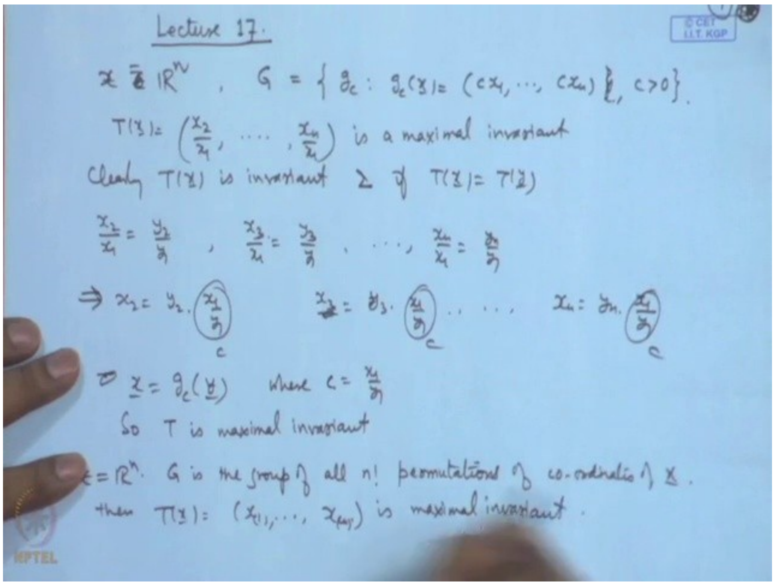
Module No. # 01

Lecture No. # 17

Invariance – II

So, let us continue the discussion on the equivariant estimators and maximal invariance, I had considered one example of the finding out maximal invariant let me take one or two more examples also.

(Refer Slide Time: 00:35)



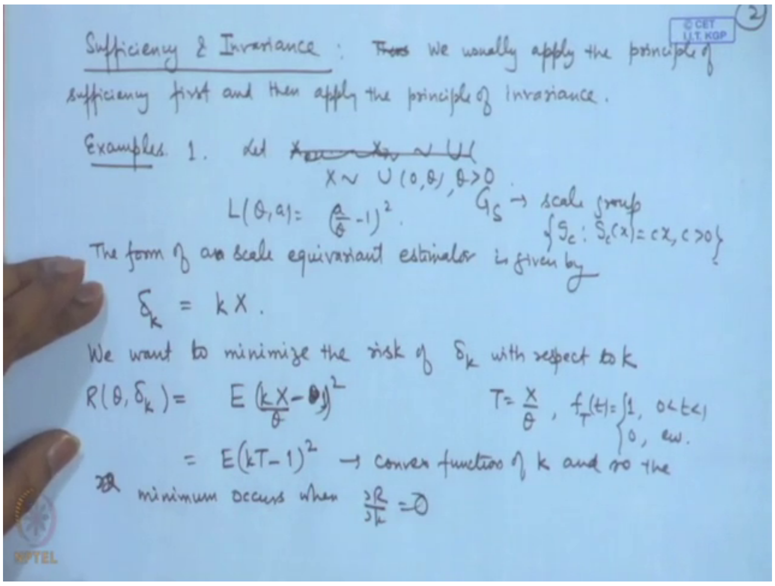
So, let us consider say again x as n dimensional Euclidean space and G is the group of scale transformations. So, I am considering for n points then g c of x is equal to c x 1 c x 2 c x n, where c is any positive real number. In this case T x is equal to x 2 by x 1 and so on x n by x 1 this is a maximal invariant.Once again you can see that as x 2 goes to c x 2, x 1 goes to c x 1 so, this ratio becomes same as x 2 by x 1 x n goes to c x n, x 1 goes to c x n so, the ratio goes to x n by x 1 etcetera. So, these are all this is an invariant function.So, clearly T x is invariant and also if I take two points T x is equal to T y, then what I get x 2 by x 1 is equal to y 2 by y 1, x 3 by x 1 is equal to say y 3 by y 1 and soon x n by x 1 is equal to say y n by y n.

Then this implies x 2 is equal to y 2 x 1 by y 1 x 3 by x 3 is equal to say y 3 into x 1 by y 1 and so on x n is equal to y n into x 1 by y 1 so this is c. So, this implies that x is equal to g c of y where we have chosen c to be x 1 by y 1 so, T is maximal invariant.Let us take another example say x is equal to say R n and G is the group of all n factorial permutations of coordinates of x. Then T x is equal to order statistics x 1 x 2 x n this is maximal invariant. (No audio from: 03:30 to 03:39) Now, let us look at the importance of invariance in determining the best as I mentioned earlier that we want to use this concept to reduce the class of available estimators. And in that reduced class if it is possible to find the best one then we are having some sort of optimal estimator in that class.

Now, when we start the reduction then we have to find out the equivariant estimator we have seen one example here in the binomial case, the condition that we are getting here is that we should have T of n minus x is equal to 1 minus T of x in the case of uniformed distribution we got the form of the equivariant estimator as a multiple of x.Now, if you take multiple of x then you realize here for example, 2 x is unbiased it is also the method of moments estimator for this problem. But suppose I have n observations x 1 x 2 x n in that case if I straightforwardly apply the concept of invariance then the estimator will turn out to be a function of it will turn to be x 1 into a function of x 2 by x 1 x 3 by x 1 and soon.

On the other hand we had seen that the maximum likelihood estimator is x n, the complete sufficient statistics is x n. Now, in that case why not we restrict attention firstly to the sufficient statistics and then we apply the concept of invariance.

(Refer Slide Time: 05:36)



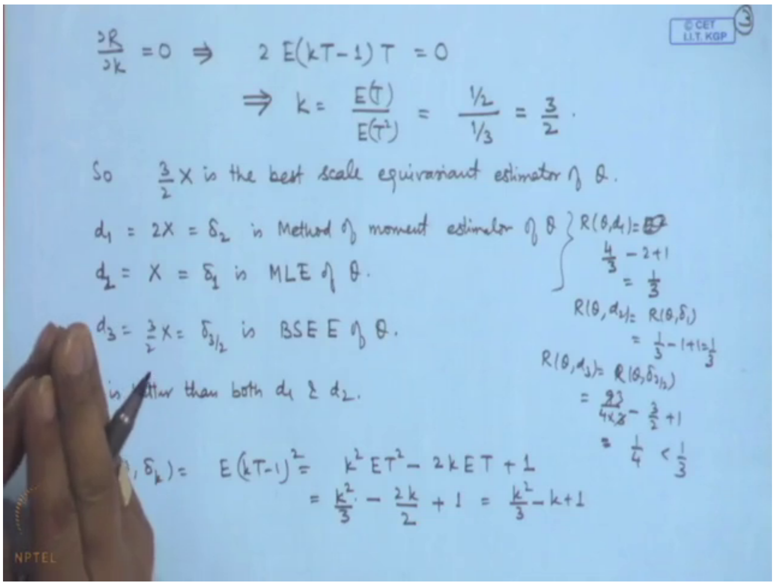
So, let me justify this thing sufficiency and invariance. So, there is a natural question that when the two criteria of sufficiency and invariance are there then which one should be applied first, can the applications of these in any order give the same answer or that will lead to the same solution etcetera? The general answer to these questions have been attempted by many researchers and under certain conditions certain results have been obtained. But we will follow a practical approach here, in most of the estimation problems we usually deal with convex loss functions.If the loss function is convex, the class of estimator which is based on the sufficient statistics supersites or you can say given any estimator, which is not based on the sufficient statistics.

We can find an estimator which is better than that using the rao black well theorem.Therefore, we can restrict attention to the class of estimators which are based on sufficient statistics. Now, if we apply the invariance on this class of estimators then we are considering much smaller class so, we will follow this approach here.So, we usually apply the principle of sufficiency first and then apply the principle of so, let us see in certain problems when the group of transformations is transitive we may actually end up with getting the best equivariant estimator.

So, let us start with the uniformed distribution problem. (No audio from: 07:58 to 08:11) Let me firstly consider this x following uniform 0 theta problem which I introduced earlier. We got the form of the group was a scale group that is g c g c x is equal to c x where c is positive here theta is positive.The form of a scale equivariant estimator is given by let me use the term d delta k that is equal to k times X.We want to minimize the risk of delta k with respect to k so, let us consider the risk of delta k that is equal to expectation of k X minus theta square sorry k X by theta minus 1 square. Let us substitute say T is equal to X by theta, then what is the distribution of T that is uniform 0 1.

So, we can write it as expectation of k T minus 1 square that means the risk function of the best of the equivarent estimator is independent of the parameter.Now, this is true because this is scale group of transformations is a transitive group. Because if I consider any two points theta 1 and theta 2 on the positive real line then, they can be reached from the other one.For example, I take theta is equal to 2 and theta 1 is equal to 2 and theta 2 is equal to 3 then if I choose c is equal to 3 by 2 then 3 by 2 times 2 is equal to 3 that means there exist a transformation so that I can reach theta 2 from theta 1.Therefore, the risk function will be constant and therefore, we can find out the best choice here.

(Refer Slide Time: 11:27)



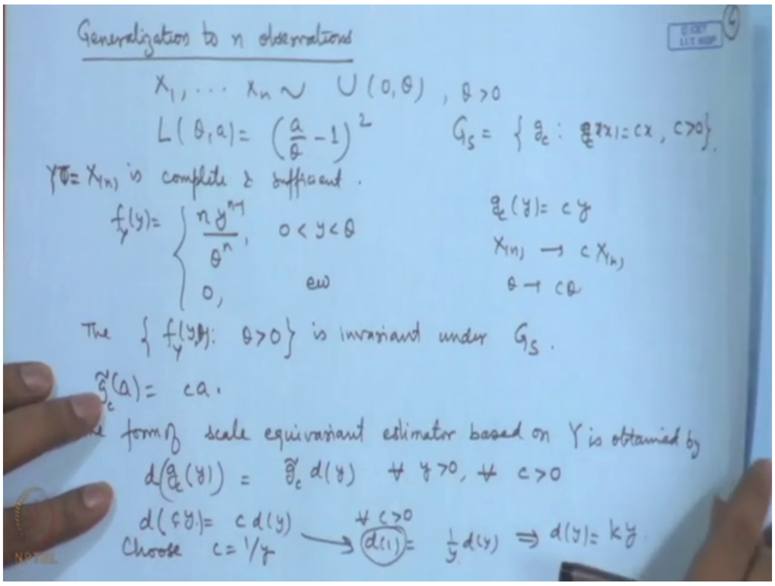
So,the this is now naturally a convex function of k and so, the minimum occurs when del R by del k is equal to 0.Now, this del R by del k is equal to 0 you can calculate del R by del k that will give us twice expectation k T minus 1 into T is equal to 0 this means k is equal to expectation of T by expectation of T square. Now, in the case of uniformed distribution the mean is half and expectation T square is 1 by 3 that is equal to 3 by 2.So, 3 by 2 X is the best scale equivariant estimator of theta.Now, let us just have a comparison between various estimators for this problem. See let me give the notations here say d 1 that is 2 X which is actually equal to delta 2 under this notation because delta k is k X so, this is method of moments estimator here. (No audio from: 12:42 to 12:49)

And d 1 sorry d 2 that is equal to X that is equal to delta 1 is the maximum likelihood estimator of theta.And this is d 3 that is equal to 3 by 2 X that is actually delta 3 by 2 this is the best scale equivariant estimator of theta.Naturally d 3 is better than both d 1 and d 2 also let us compare d 1 and d 2 what is the risk of d 1 that is equal to expectation of 2.So, actually we have the general form here we can calculate the risk function of delta k that is expectation of k T minus 1 square that is equal to k square expectation of T square minus twice k expectation of T plus 1 that is equal to k square expectation of T square is 1 by 3 minus twice k expectation of T is half plus 1 that is equal to k square by 3 minus k plus 1.

So, from here the risk function of the method of moments estimator that will be equal to 4 by 3 minus 2 plus 1 that is equal to 1 by 3.The risk function of d 2 that is the maximum likelihood estimator that is the risk of delta 1 that is equal to 1 by 3 minus 1 plus 1 that is equal to 1 by 3.And the risk of d 3 that is the best scale equivariant estimator that is obtained by putting k equal to 3 by 2 that is equal to 9 by 4 into 3 minus 3 by 2 plus 1 so, that is equal to 1 by 4 which is less than 1 by 3.So, in this particular case the method of moments estimator and the maximum likelihood estimators they have the same risk and the best scale equivariant estimator is better than both of them.

So, here you can see the concept of invariance helps us in reducing the mean squared error because here, the risk criteria is actually the mean squared criteria mean squared error criteria.

(Refer Slide Time: 16:12)

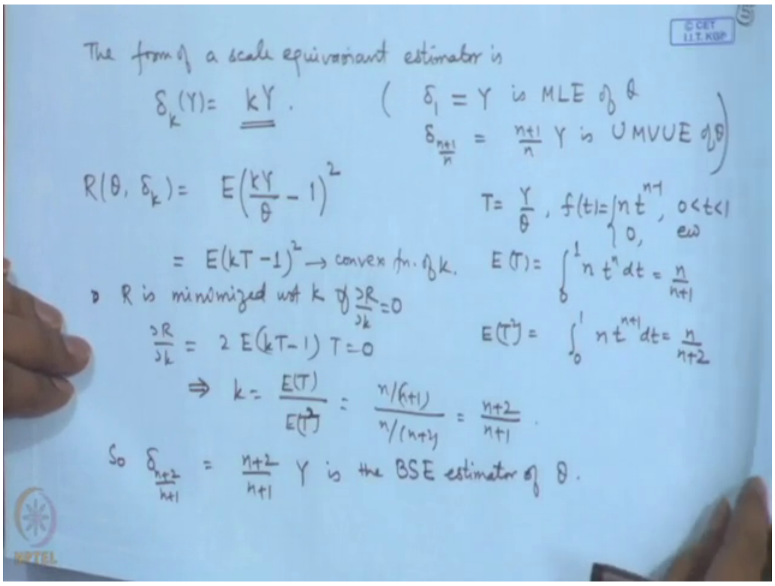


Now,I will generalize this problem here I have considered only one observation from the uniformed distribution.Now, in place of one observation suppose I have n observation let us consider generalization to n observations. If I have generalization to n observations that is x 1 x 2 x n follows uniform 0 theta, the last in estimating theta is once again the same I am considering to keep the problem invariant the group of transformations is the scale group of transformations. (No voice from: 16:42 to 16:52) Now, here x n is let me call it say T is complete and sufficient rather let me call it y. And we know the distribution of y the distribution of y is n y to the power n minus 1 by theta to the power n.

Then the distributions f y f y theta where theta greater than 0, they will remain invariant because if I consider the distribution of see if I take g c of y then that will be equal to c of y. Because if each of the observation is shifted by c x then x n that is the maximum will also be shifted by c.So, x n goes to c x n and therefore, theta will go to c theta for this density also therefore, this density family of distribution is invariant under the scale group of transformations. And same thing will happen to the a also that is g c tilde a this will become equal to c a theta goes to c theta and a goes to c a.

So, we can consider the form of scale equivariant estimator based on y because this is the complete sufficient statistics. So, rather than starting from x 1 x 2 x n we will initially itself restrict attention to x 1 that is the x n that is the sufficient statistic. So, this is obtained by considering the condition that is d of g c y equal to g c tilde of d y for all y and for all c.Now, this condition gives d of c y is equal to c of d y so, you can choose c is equal to 1 by y. So, there is no invariant here actually no maximal invariant here you will get d of 1 is equal to 1 by y d of y this implies d of y is of the form this d 1 is a constant a constant times y.

(Refer Slide Time: 20:14)

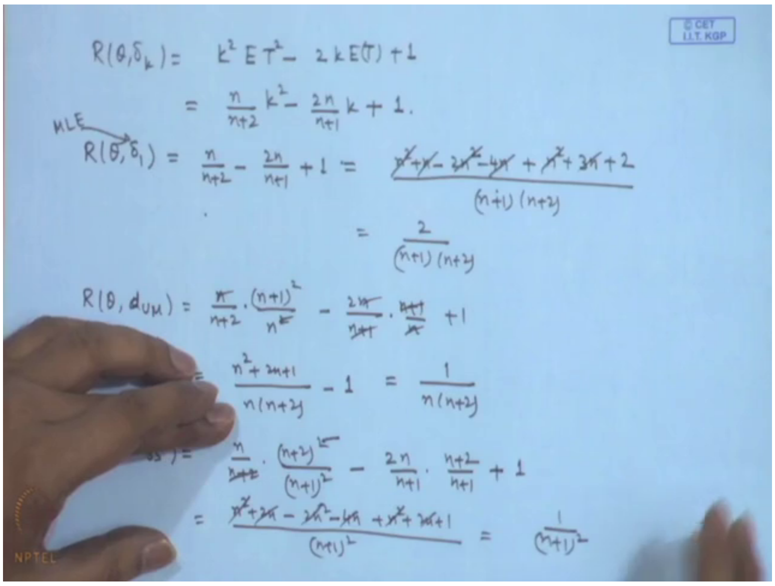


So, we can write the form of a scale equivariant estimator is then let us call it delta k of y is equal to k times of y. In fact if I consider delta 1 that is equal to Y this is the maximum likelihood estimator and if I take delta is equal to n plus 1 by n that is n plus 1 by n Y this is the minimum variance unbiased estimator of theta so these two things are known to us.Now, let us try to see whether I get something else by considering the minimization of the risk function with respect to k. So, let us consider the risk function of delta k that is equal to expectation of k Y by theta minus 1 square.See once again we can look at this density if I define say T is equal to Y by theta then the density of T is nothing but n t to the power n minus 1 0 less than T less than 1 0 otherwise because this is the density of y.

So, if I consider the density of y by theta then 1 by theta d y will be equal to d T so, this density reduces to n t to the power n minus 1.So, if I look at expectation of T that is equal to integral n t to the power n d t 0 to 1 that is equal to n by n plus 1.And expectation of T square turns out to be 0 to 1 n t to the power n plus 1 d t that is equal to n by n plus 2 so, this is equal to expectation of k T minus 1 square.If we consider the minimization with respect to k R is minimized with respect to k if del R by del k equal to 0 why because this is nothing but a convex function of k this is convex function of k.

So, this del R by del k this gives twice expectation k T minus 1 into T equal to 0 that means k is equal to expectation of T by expectation of T square and that is equal to n by n plus 1 divided by n by n plus 2 that is equal to n plus 2 by n plus 1 so, delta of n plus 2 by n plus 1 that is n plus 2 by n plus 1 y is the best scale equivariant estimator of theta.Once again we can look at the relative risk comparison or risk improvement here naturally here you have the coefficient 1 here you have coefficient n plus 1 by n and here you have n plus 2 by n plus 1 so, this is the best. That is the risk of this will be the smaller than both of this now let us look at the overall comparison of the risk values here.

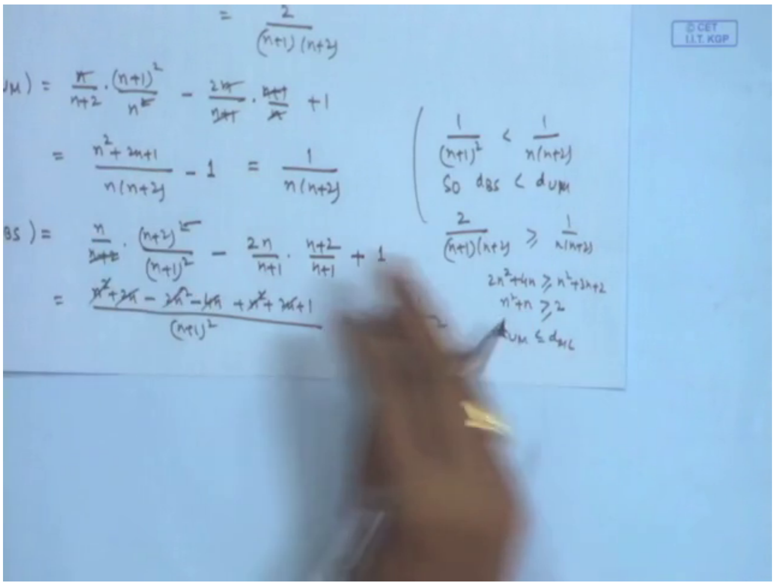
(Refer Slide Time: 24:10)



So, what is the value of R theta delta k that is k square expectation of T square minus twice k expectation of T plus 1 that is equal to n by n plus 2 k square minus twice n by n plus 1 k plus1.So, if I look at the risk of say the maximum likelihood estimator here k equal to 1 so,I get n by n plus 2 minus twice n by n plus 1 plus 1 so, you can simplify this that is equal to n square plus n minus twice n square minus 4 n plus n square plus 3 n plus 2 divided by n plus 1 into n plus 2 so this get simplified that is equal to 2 divided by n plus 1 by n plus 2.If we look at the risk of the unbiased estimator minimum variance unbiased estimator then that is equal to it is obtained by the value k equal to n plus 1 by n here.

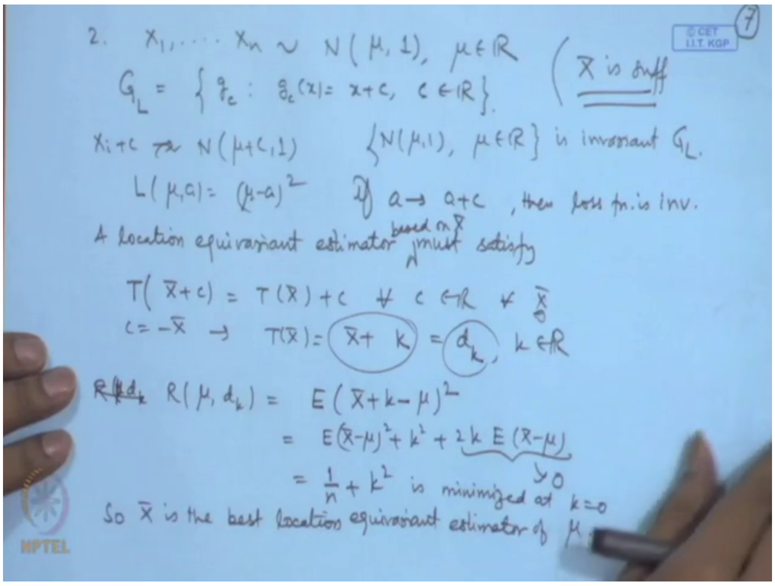
So, here if we substitute that we get n by n plus 2 n plus 1 square by n square minus twice n by n plus 1 into n plus 1 by n plus 1 so, here these terms cancelled out and we get here n square plus 2 n plus 1 divided by n into n plus 2 this is minus 1 because this is minus 2 plus 1 that is equal to once again 1 by n into n plus 2.So, naturally you can see that this is greater than let us also look at the risk of the best scale equivariant estimator that is equal to n by n plus 2 into n plus 2 square by n plus 1 square minus twice n by n plus 1 into n plus 2 by n plus 1 plus 1 so this can be simplified that is equal to n square plus 2 n minus twice n square minus 4 n plus n square plus 2 n plus 1 divided by n plus 1 square 1 by n plus 1 square.

(Refer Slide Time: 27:27)



So, if we compare now this 1 by n plus 1 square less than 1 by n into n plus 2 because this is n square plus 2 n and this is n square plus 2 n plus 1 so d b S is better than d u m.Similarly, if I compare the MLE and the UMVUE then 2 by n plus 1 into n plus 2 and 1 by n into n plus 2 let us look at the comparison here.So, this term if I multiply I get twice n square plus 4 n and here I get n square plus 3 n plus 2 so, if I put greater than this is reducing to n square plus n greater than 2 so, this condition is true that means d U M is better than d M L.So, in this particular case the best scale equivariant estimator turns out to be the best among the three given estimators.

(Refer Slide Time: 28:56)

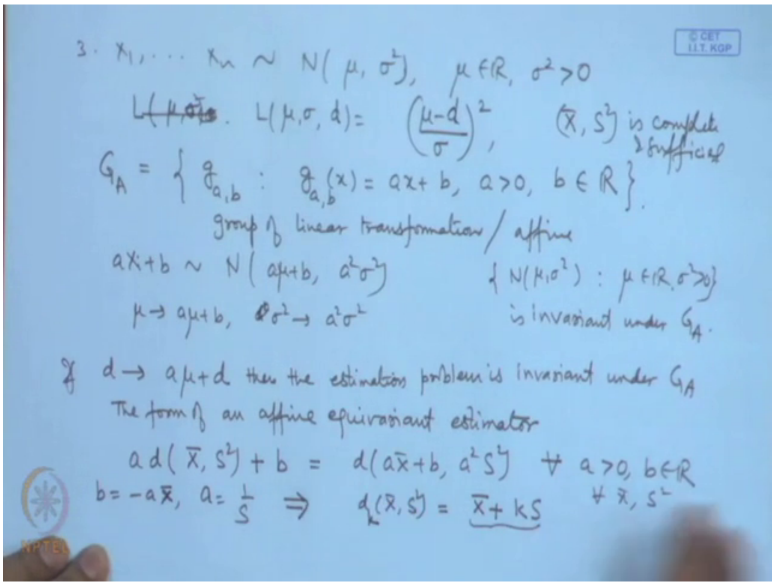


Let us take the normal distribution case x 1 x 2 x n follows normal mu 1 where mu is a real number we consider the group of translations or the location transformations g c of x is equal to x plus c where c is any real number.Now, you observe here that X i plus c that will follow normal mu plus c 1 so, if mu is a real number then mu plus c is also a real number so, the family is invariant this family is invariant under the location group. Let us take the loss function here as mu minus a square now if mu goes to mu plus c, a should go to a plus c so if I take a going to a plus c then the loss function remains invariant and therefore, the estimation problem is invariant.

So, now let us find out the form of a location equivariant estimator this must satisfy now in this problem X bar is sufficient.So, we will restrict attention to the distribution of X bar based on X bar. So, we will have T of X bar plus c is equal to T of X bar plus c for all c and for all XX bar so you choose c is equal to minus X bar so, you will get T X bar is equal to X bar plus a constant so, we call it say d k where k is any real number. So, this is the form of a location equivariant estimator that it is X bar plus a constant.

So, if I consider the risk function of this, that is equal to expectation of X bar plus k minus mu whole square that is equal to expectation of X bar minus mu square plus k square plus twice k expectation of X bar minus mu this is 0 so, you are left with 1 by n plus k square this is minimized at k equal to 0.So,X bar is the best location equivariant estimator of mu.

(Refer Slide Time: 32:30)

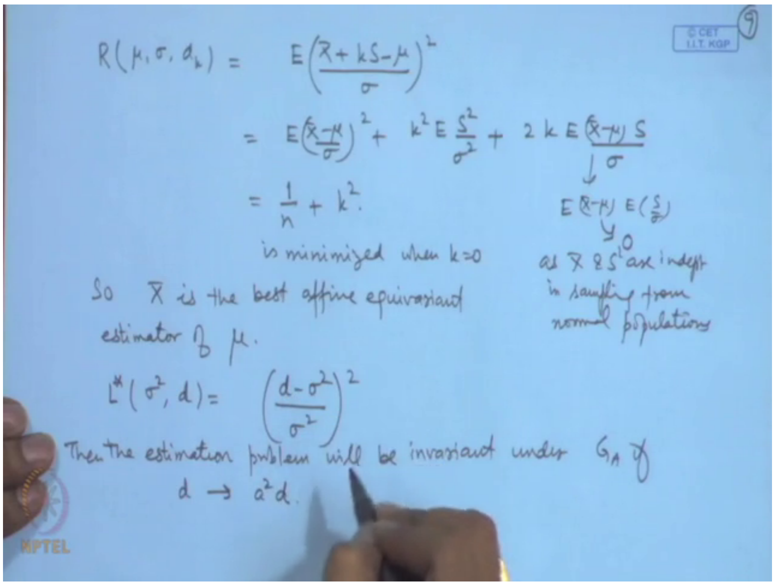


Let us generalize this problem;I consider here the variance also to be unknown x 1 x 2 x n following say normal mu sigma square. Now, we maintain the same loss function no sorry we change the loss function as mu minus a by sigma whole square.Let us consider the group of transformations which are linear or affine (No audio from: 33:20 to 33:27)a x plus b where a is a positive number and b is any real number this is called group of linear transformations or affine transformations so, we will use the notation g a.Now, under this the distribution of a X i plus b is normal a mu plus b a square sigma square so, a mu plus b is another real numbers a square sigma square is positive therefore, the family is invariant.

The family of normal mu sigma square distributions this is invariant under the group g a.Naturally you are seeing that a mu goes to a mu plus b a square sigma square goes to a sigma square goes to a square sigma square therefore, if I want to consider the let me put here d then d should go to a mu plus d. Then the estimation problem is invariant, (No audio from: 34:53 to 35:04) then the estimation problem is invariant under the group g.Now, to derive the form of a affine equivarentestimator (No audio from: 35:13 to 35:22) let us consider the form of an affine equivarent estimator then you will get d of X bar S square so, here in this problem X bar and S square is complete and sufficient.

So, a d X bar S square plus b is equal to d of aX bar plus b a square S square for all a and for all b and for all X bar and S square.So, you choose b is equal to minus aX bar and a is equal to 1 by S, then this will give us d of X bar S square is equal to X bar plus some constant times S.So, this is the form of an affine equivarent estimator let us look at the risk function here.

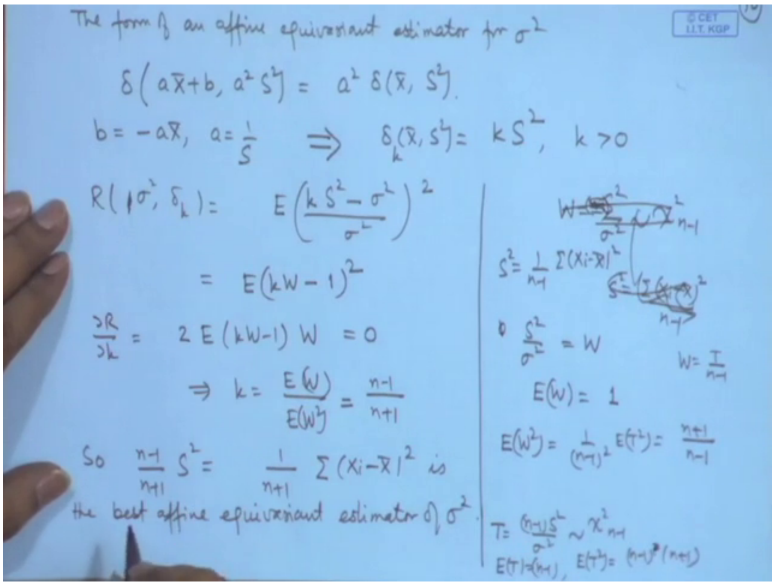
(Refer Slide Time: 36:55)



Let us consider the risk function of d k that is equal to expectation of X bar plus k S minus mu by sigma whole square that is equal to expectation of X bar minus mu by sigma whole square plus k square expectation of S square by sigma square plus twice k expectation of X bar minus mu S by sigma.Now, in the sampling from normal distribution we know that X bar and S square they are independently distributed. So, this term will become 0 because this will become expectation of X bar minus mu into expectation of S by sigma and this is 0 as X bar and S square are independent in sampling from normal populations.

So, this term simply becomes 1 by n plus k square expectation of S square is sigma square so, this is simply one.Once again this is minimized when k equal to 0 so,X bar is the best affine equivariant estimator of mu.Suppose I take the loss function here that means the problem is of estimating sigma square. If I consider this then the estimation problem remains invariant under the affine group if d goes to a square d, because sigma square goes to a square sigma square so, this will get this will remain invariant if d goes to a square d.

(Refer Slide Time: 39:47)



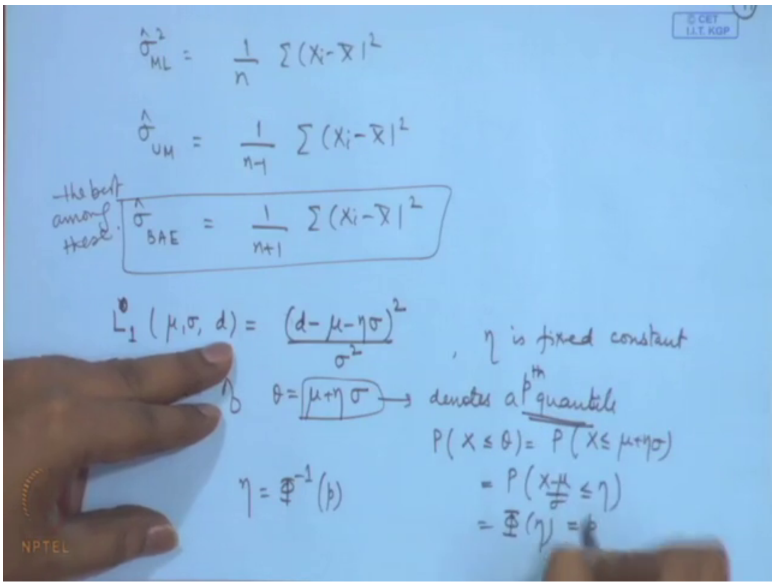
So, the form of an affine equivariant estimator for sigma square this should satisfy delta of a x bar plus b a square S square is equal to a square delta of x bar S square. So, you choose b is equal to minus a X bar and a is equal to 1 by S so, then this will imply that delta of X bar S square is nothing but a constant times S square because this will become 0 this will become 1 so where k is a positive real.Let us look at the minimization of the risk of delta k that is equal to expectation of k S square minus sigma square by sigma square whole square let us use the notation.

SW is equal to say S square by sigma square that follow chi square n minus 1 actually it will depend upon what notation for S square I am using if I am using S square is equal to sigma X i minus XS bar whole square by n minus 1 then we should have n minus 1 S square by sigma square follows chi square on n minus 1 degrees of freedom.So, we can use some modified notation here because I substituted here expectation of S square is equal to sigma square so, actually I am choosing the definition of S square as 1 by n minus 1 sigma X i minus X bar whole square.

So, let us use the notation(No audio from: 41:59 to 42:09) so, S square is equal to 1 by n minus 1 sigma X i minus X bar whole square and let us write this S square by sigma square as say W then this is reducing to expectation of k times W minus 1 whole square.Now, we can find out the minimization del R by del k that will give twice expectation k W minus 1 into W this will give k time is equal to expectation of W minus expectation of W square.Now, let us look at these terms here expectation of W is equal to 1 because expectation of S square is equal to sigma square. What is expectation of W square?That will be equal to so if I calculate say T is equal to n minus 1 S square by sigma square that is chi square on n minus 1 degrees of freedom.

So, expectation of T is n minus 1 and expectation of T square is equal to n minus 1 square into n plus 1 n minus 1 into n plus 1 so, expectation of W is equal to because here W is equal to T by n minus 1.So, expectation of W square is 1 by n minus 1 square into expectation of T square so that gives us n plus 1 by n minus 1.So, this is equal to then expectation of W is equal to n minus 1 by n plus 1 so, n minus 1 by n plus 1 S square that is equal to 1 by n plus 1 sigma X i minus X bar whole square is the best affine equivariant estimator of sigma square.

(Refer Slide Time: 45:07)

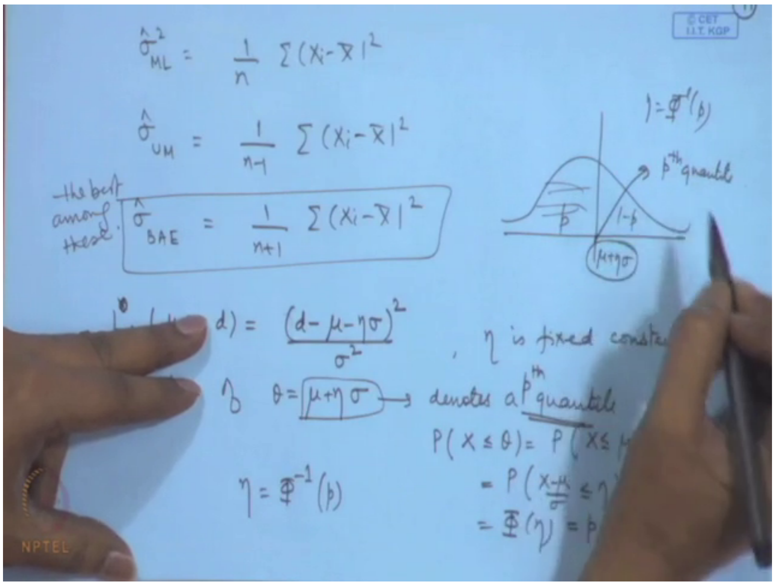


So, let us take the whole thing in the perspective here for normal mu sigma square distribution we had sigma square M L as 1 by n sigma X i minus X bar whole square. We had the minimum variance unbiased estimator as 1 by n minus 1 sigma X i minus X bar whole square whereas, the best affine equivariant estimator is now1 by n plus 1 sigma X i minus X bar whole square. Since all of these are only multiples of sigma X i minus X bar whole square this is the best among these two these three.So, you can see here that the principle of invariance allows us to improve upon the given estimators.

In the estimators which we have obtained using the method of maximum likelihood estimator method of moments or by the method of minimum variance unbiased estimation etcetera.Of course in this problem it can be further shown that even this best affine equivariant estimator can be further improved, but that is by another approach. So, we will not be doing that approach here now for this problem I want to give one more application here let us consider another loss function.Say L star in place of star let us put L 1 mu sigma d is equal to d minus mu minus eta sigma whole square divided by sigma square.

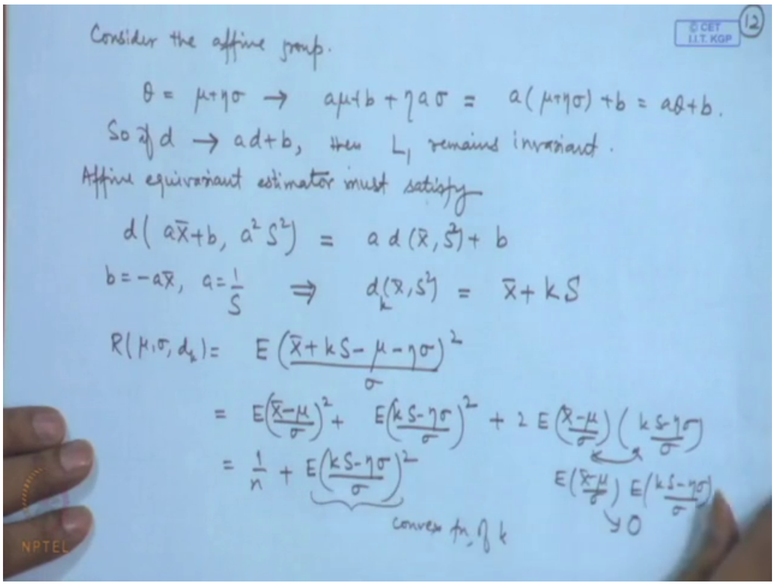
Now, what is this problem this can be considered as estimation of theta is equal to mu plus eta sigma where eta is a fixed constant.See in the case of normal distribution mu plus eta sigma denotes a quantile because here if I consider probability of X less than or equal to theta that is probability of X less than or equal to mu plus eta sigma that is equal to probability of X minus mu by sigma less than or equal to eta that is equal to phi of eta where phi is the standard normal c d f.

(Refer Slide Time: 47:47)



So if I put this is equal to p that is eta is equal to phi inverse p then this the p th quantile.Quantile of order p that means if I am considering the distribution here then this probability is p this probability is 1 minus p. So, this point is mu plus eta sigma where eta is given by phi inverse p so, this the p th quantile.So, like mean median etcetera the quantiles also of interest to be estimated to denote or to find out various locations on the distribution let us consider the estimation of this here.

(Refer Slide Time: 48:30)

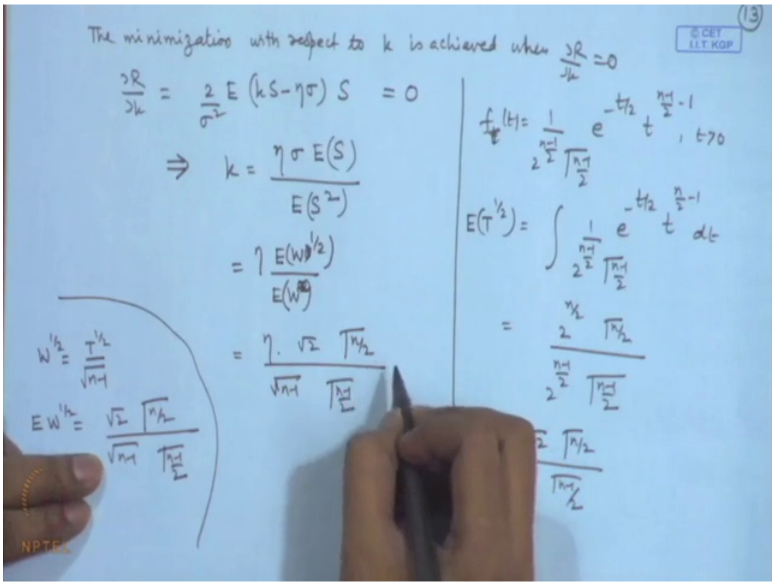


And so, consider the affine group now we have already seen here that mu goes to a mu plus b sigma will go to a sigma. So, what will happen to theta?Theta is equal to mu plus eta sigma so, this will go to a mu plus b plus eta a sigma that is equal to a mu plus eta sigma plus b that is equal to a theta plus b.Therefore, d must go to a d plus b so, if that is happening then L 1 remains invariant. So, affine equivariant estimator if I consider affine equivariant estimator that must satisfy d of aX bar plus b a square S square is equal to a d X bar S square plus b.

So, if you choose b is equal to minus a X bar and a is equal to 1 by S then we get the form of d X bar S square as X bar plus because this will become a constant and we get minus a X bar and then we put 1 by aso, S is multiplied by this constant plus this x bar.That is the form which is the same which we obtained for the estimation of mu.Now, if you consider the risk function of this estimator that is expectation of X bar plus k S minus mu minus eta sigma by sigma whole square then we can write it as expectation of X bar minus mu by sigma whole square plus expectation of k S minus eta sigma by sigma whole square plus two times expectation of X bar minus mu by sigma into k S minus eta sigma by sigma.

Now, here these terms are independent because in the sampling from normal distribution X bar and S square are independent and this term becomes 0.So, the problem has reduced to minimization of this term expectation of k S minus eta sigma by sigma whole square.Now, this is a convex function of k so, the minimization will occur when we differentiate this with respect to k and put equal to 0.

(Refer Slide Time: 52:17)

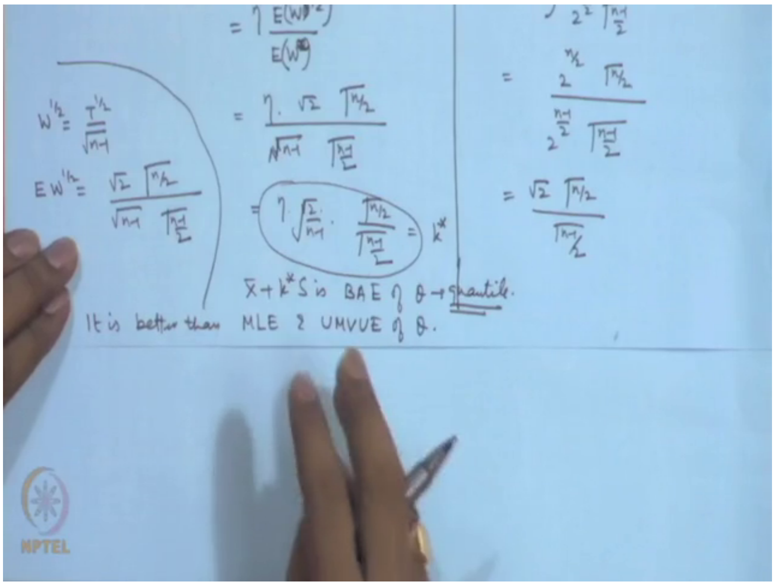


So, the minimization with respect to k is achieved when del R by del k equal to 0 and del R by del k is actually equal to twice expectation k S minus eta sigma into S equal to 0 that will give us k is equal to eta sigma expectation of S divided by expectation of S square.Now, if we use the no terminology used in the derivation for estimation of sigma square we had defined T as n minus 1 S square by sigma square then this is following chi square distribution on n minus 1 degrees of freedom.

Now, expectation of S is coming in the terms of T to the power half so, let us calculate expectation of T to the power half. We have the density function of t as 1 by 2 to the power n minus 1 by 2 gamma n minus 1 by 2 e to the power minus t by 2 t to the power n minus 1 by 2 minus 1 where t is positive.So, expectation of T to the power half that will be 1 by 2 to the power n minus 1 by 2 gamma n minus 1 by 2 e to the power minus t by 2 t to the power n by 2 minus 1 d t that is equal to gamma n by 2 2 to the power n by 2 divided by 2 to the power n minus 1 by 2 gamma n minus 1 by 2 that is equal to root 2 gamma n by 2 divided by gamma n minus 1 by 2.

Now, S here we can consider S by sigma and S square by sigma square so, this is reducing to expectation of W divided by expectation of W square sorry expectation of W to the power half because here I get S by sigma so, this is W to the power half and expectation of W.Now, what is W to the power half that is equal to T to the power half by square root n minus 1 so, expectation of W to the power half is equal to root 2 gamma n by 2 by root n minus 1 gamma n minus 1 by 2.So, this terms we substitute here eta root 2 gamma n by 2 divided by square root n minus 1 gamma n minus 1 by 2 and then expectation of W is equal to 1.

(Refer Slide Time: 56:02)



So, this is the term that we get of course, we can write it as eta root 2 by n minus 1 gamma n by 2 by gamma n minus 1 by 2 so this is the best choice of k and we are getting X bar plus let me call this as the this choice as say k star k star S is the best affine equivariant estimator of theta that is the quantile. Once again you can see this is different from the maximum likelihood estimator so, it is better than MLE and UMVUE of theta.So, today we have discussed the concept of invariance in detail and we have seen that finding out the best invariant estimator under a certain group of transformations many times leads us to much better estimators than the conventional methods.

That we have discussed till now like the maximum likelihood estimation or the minimum variance unbiased estimation.Another point which is to be noted here we have not discussed it here till now is that there are procedures which can also lead to improvement over the equivarient estimators. But that is part of some advanced discussion if we find time we will be able to cover it somewhat later. So,I will also introduce some decision theory concepts such as the Bayes estimation and the mini max estimation in the next lectures.