Statistical Inference

Prof. Somesh Kumar

Department of Mathematics

Indian Institute of Technology, Kanpur

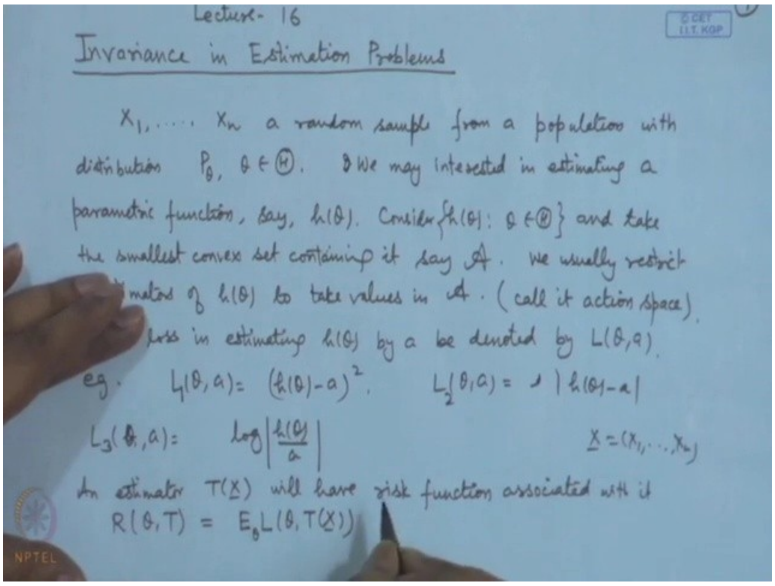
Module No. # 01

Lecture No. # 16

Invariance – I

Today I will introduce the concept of invariance in estimation problems, why do we need invariance? And what is invariance? Earlier we have seen that see we have a large class of estimators now, if we insist on certain criteria such as unbiasedness, consistency etcetera. Then we are restricting the class of available estimators for example, if we apply the criteria of unbiasedness, then, we are considering only those estimators which are unbiased.

(Refer Slide Time: 01:43)



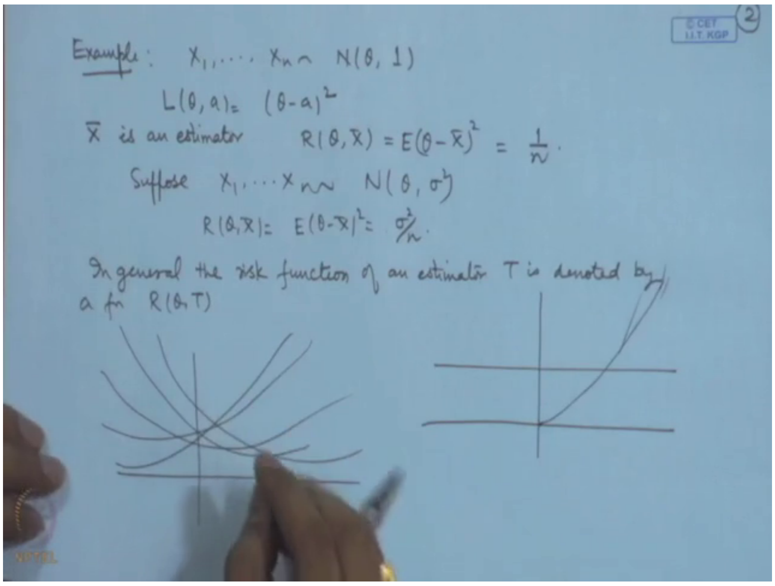
And then it is possible or it may be possible to choose the best among them according to another criteria such as minimum variance unbiased estimator. Similarly, among the consistent estimators we may choose the ones, which has asymptotically normal distribution so we call it kian estimators. In a similar way, this invariance also attempts, it is an attempt to reduce the class of available estimators by applying an additional criteria and then it may be possible to choose among them the best.

So, we will call it best equivalent estimator. So, let me introduce the concept first so, as before we are considering, we have a sample x 1 x 2 x n a random sample from a population with distribution p theta, theta belonging to theta. In general we will be interested in; so, we may be interested in estimating a parametric function say, h theta. So, what we consider? See usually, the space of values of theta so, according to that x theta will also vary. So, let us consider say the space h theta, theta belonging to theta.

And consider consider this and take the smallest covex set containing it say, let me give a notation script A. So, we usually restrict the estimators of h theta to take values in A. So, we call it say actions space. Next what we do? We consider a certain criteria I have already discussed for example, mean squared error, but in place of mean squared error we can consider a general function, we call it loss function, let the loss in estimating h theta by a be denoted by say L theta a.

So for example, we may have L theta a is equal to say h theta minus a square L theta a let me put a L 1 L 2 could be for example, modulus of h theta minus a, we may take say L 3 as say log of h theta by a say, modulus of this and like that we can define various such things so, these are called loss functions. Now, an estimator T x that is where x, I am denoting by x 1, x 2, x n then will have risk function associated with it let me call it R theta T that is equal to expectation of L theta T x so, this is called a risk function.

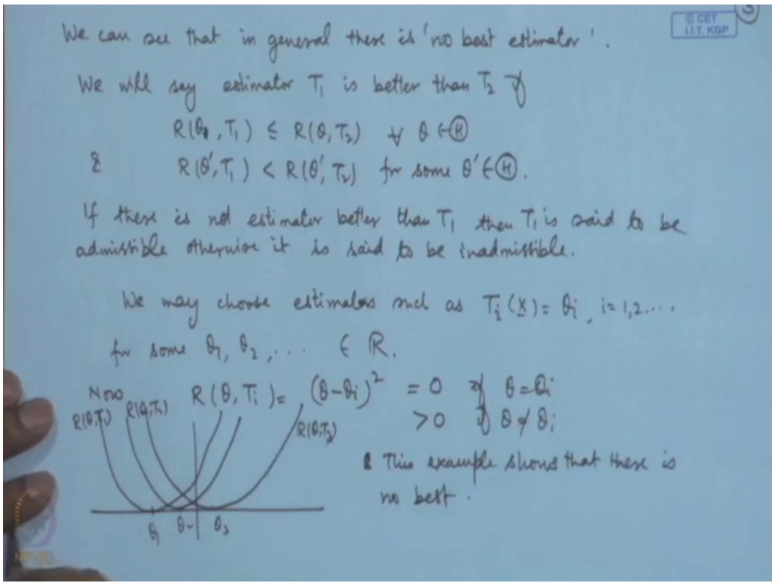
(Refer Slide Time: 05:57)



Let me illustrate this thing, let us take this example, say x 1, x 2, x n follows normal theta 1 we are considering say loss function is equal to theta minus a square, x bar is an estimator so, risk of this square is equal to theta minus x bar square expectation, which is equal to 1 by n this is a constant value. Suppose, in place of normal theta 1, we had theta sigma square and we had considered the same loss function. In that case R theta x bar will be equal to expectation of theta minus x bar square is equal to sigma square by n so, it becomes a function of the parameter.

So, in general the risk function of an estimator T is denoted by a function R theta T. So, now you may have situation like this, that for a given estimator the like here, it is 1 by N so, it is something like this, but if you are considering sigma square by N and sigma square may change. In that case depending upon the value of sigma square you may have curve, it may be like for sigma square equal to 0. And if we plot it as a function of sigma square, then it goes up and up, as sigma square tends to infinity it goes to infinity, at sigma square equal to 0 this is equal to 0. In general the risk function of various estimators for a certain parametric function will be depicted by certain graphs. Easily you can see that with respect to this, criteria there is no best.

(Refer Slide Time: 08:20)



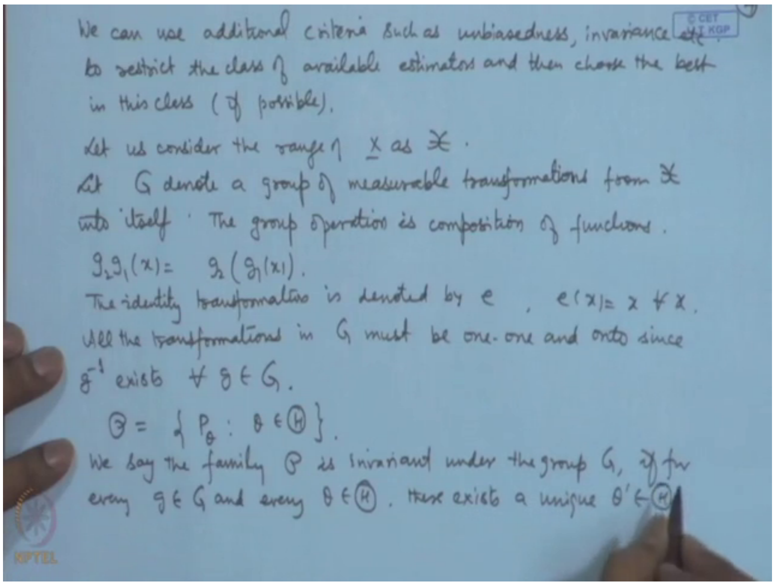
We can see that, in general there is no best estimator. What is the meaning of this statement? Because we will say that estimator say T 1 is better than T 2, if R theta 1; R theta T 1 is less than or equal to R theta T 2 for all theta. And R theta prime T 1 is strictly less than R theta prime T 2 for some theta prime belonging to theta. So, if there is no estimator better than T 1, then T 1 is said to be admissible, otherwise it is said to be inadmissible. So, if you consider say the squared error loss function, we may choose estimators such as say T i x is equal to theta i for i is equal to 1, 2 and so on for some theta 1 theta 2 etcetera, belonging to the real line.

Now, if you consider R theta T i then that is equal to theta minus theta i square. That is obviously equal to 0, if theta is equal to theta i and it is greater than 0 if theta is not equal to theta i. So, if you consider the plot of each of these things suppose, this is value theta 1 then R theta, theta 1 or R theta T 1 that will be something like this. If you consider theta 2 here, then the risk function of so, this is say R theta T 1 this is R theta T 2 suppose, theta 3 is here then its risk function will be like this R theta T 3.

So you can easily see that there is no best this example shows that there is no best. The problem of point estimation can be stated as the problem of finding out the best estimator, the one which has the minimum risk throughout, but this example shows that it is not possible to have best estimator. Therefore, what are the other practical options? There are; we can actually say that the class of all the estimators is not ordered, it cannot be completely ordered.

So, what we can do? We can introduce some additional criteria such as unbiasedness, invariance etcetera and therefore, we have a smaller class and within that class we can try to find out the best choice so, invariance is one such thing.

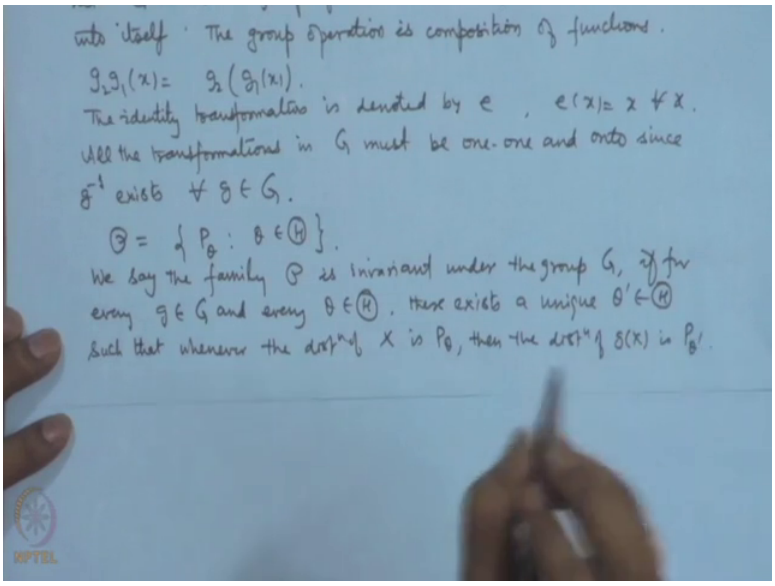
(Refer Slide Time: 12:31)



We can use additional criteria such as unbiasedness, invariance etcetera to restrict the class of available estimators and then choose the best in this class if possible. Now, let me introduce the concept of invariance. Let us consider the range of x as script x. So, we introduce let G denote a group of measurable transformations from x into itself. The group operation is composition of functions (No Audio From 14:11 to 14:22) that is we define g 2 g 1 of x as g 2 of g 1 of x and there is an identity if it is a group. The identity transformation is the identity function this is denoted by e, that is e of x is equal to x for all x.

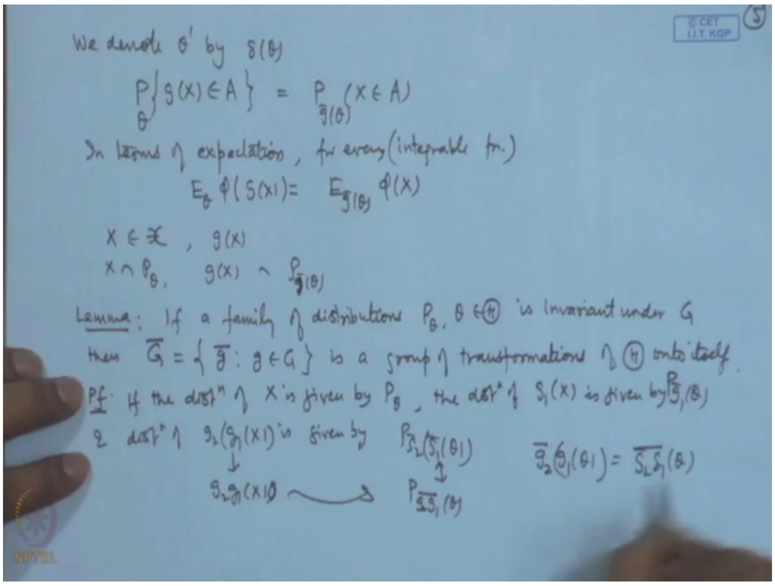
Now, all the transformations in G must be one-one and onto since g inverse exists for all G and measurable T is required because if I say x is a random variable, then g x must also be a random variable. So, now let us consider the family of distributions so, p is the family of distributions. So, we say that the family p is invariant under the group G if for every g belonging to G and every theta belonging to theta, there exists a unique theta prime belonging to theta.

(Refer Slide Time: 16:21)



Such that whenever the distribution of x is p theta, then the distribution of g x is p theta prime.

(Refer Slide Time: 16:43)

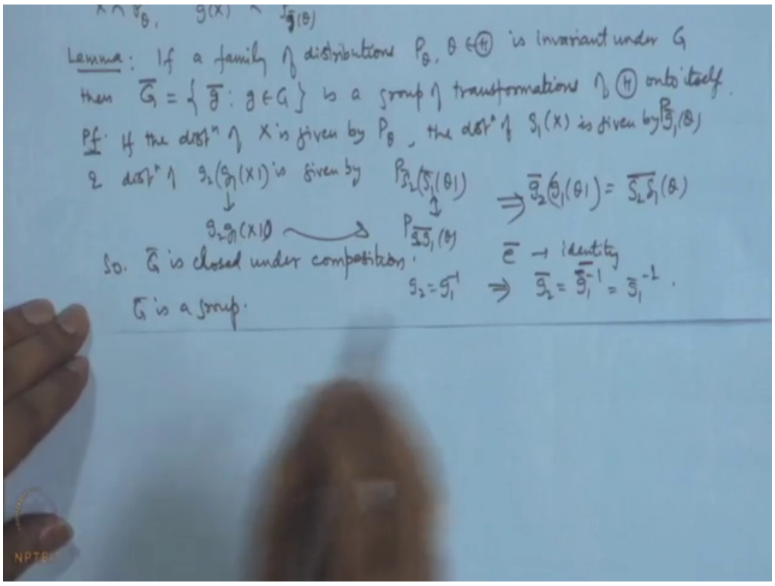


And we denote this theta prime by g of theta so, what we are saying is essentially? That probability, that g x belonging to A when theta is a true parameter value is same as probability of x belonging to A, then the true parameter value is g bar theta. In terms of expectation, this condition is saying that for every integrable function, integrable in the sense of expectation.

Expectation of phi g x is equal to expectation of (No Audio From: 17:38 to 17:46) so, what we are saying is that for x belonging to x we are introducing g of x and if x is having distribution p theta then g x is having distribution p g bar theta so, there is an association here.

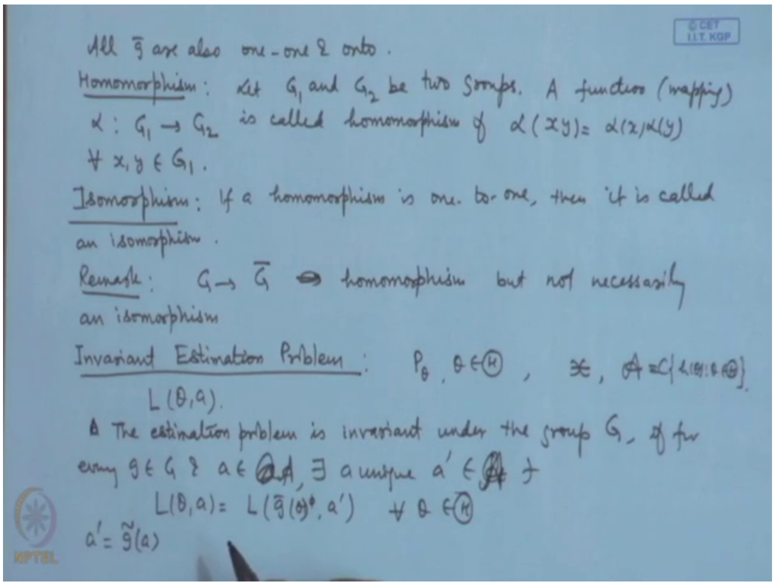
And then we have the following lemma, if a family of distributions p theta is invariant under G, then the corresponding group G bar which is obtained by the collection of g bar corresponding to every G, we have a g bar so, this group is a group of transformations of theta onto itself. So, if the distribution of x is given by p theta, the distribution of g 1 x is given by g 1 bar theta; sorry p of g 1 bar theta and the distribution of say g 2 of g 1 x is given by p g 2 bar g 1 bar theta. But this is equal to g 2 of g 1 of x and this distribution is given by p of g 2 g 1 bar theta so, these two should be same, because of the uniqueness you are getting that g 2 bar g 1 bar theta is equal to g 2 bar g 1; g 2 g 1 bar theta.

(Refer Slide Time: 20:14)



So, closer property is satisfied, is closed under composition. Now, if we consider e bar that is the identity element, if we consider say, if we choose here g 2 is equal to g 1 inverse then, what you will get? g 2 bar is equal to g 1 inverse bar that is equal to g 1 bar inverse so, this implies that G bar is a group.

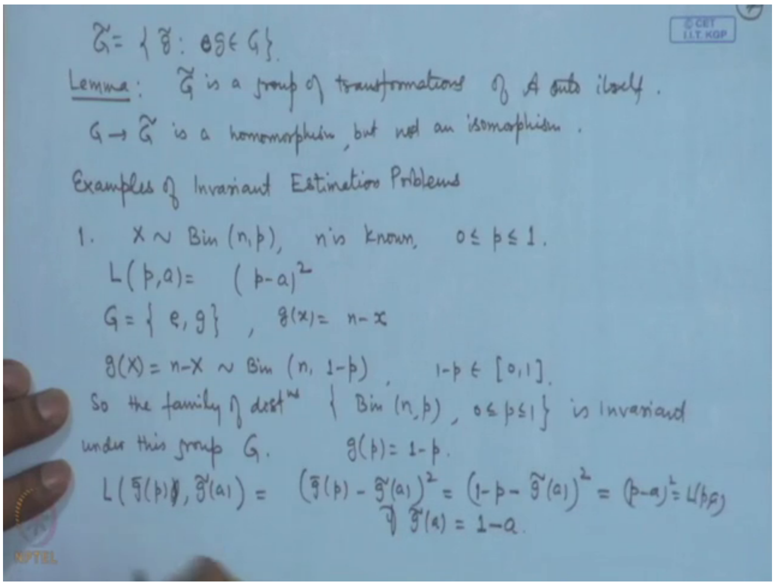
(Refer Slide Time: 21:08)



(No Audio From: 21:00 to 21:07) So, all g bar are also one-one and onto. Now, we observe an interesting property let me define, what is a homomorphism? Let G and G; let G 1 and G 2 be two groups. A function or mapping say, alpha from G 1 to G 2 is called homomorphism if alpha of x y is equal to alpha x into alpha y for all x y belonging to G 1. And isomorphism, if a homomorphism is one to one, then it is called an isomorphism, so, we make a remark here; that G to G bar this is homomorphism, but not necessarily an isomorphism.

(No Audio From: 23:16 to 23:27) Let us define, what is an invariant estimation problem? So, we have a family of distributions, we are considering a certain loss function that is L theta a, and we have a group of transformations. So, we have family of distributions, we have the space of the values of the random variable, we have the action space that is the convex closer of the h theta values and we have a loss function. So, we say that the estimation problem is invariant under the group say G, if for every g and a, there exists a unique say a prime such that a is the space of the estimators, such that L theta a is equal to L g bar theta a prime for all theta so, we denote this a prime as g tilde a. So, we have introduced another group now.

(Refer Slide Time: 25:44)

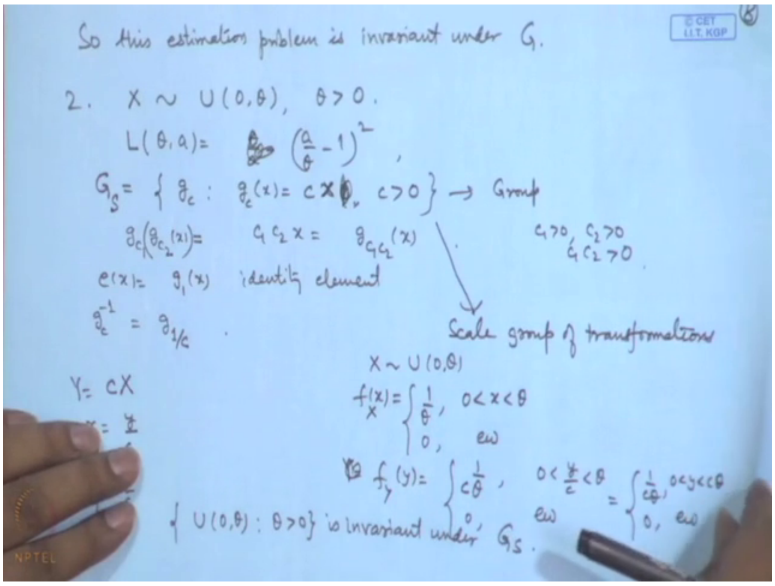


G tilde is the group, corresponding to the group G as before we have that G tilde is a group of transformations of A into; onto itself. And once again the mapping from G to G tilde is a homomorphism, but not an isomorphism. Let us consider examples of invariant estimation problems. (No Audio From: 27:46 to 27:02) Let us consider say X follows binomial n p, n is known, and p is any value between 0 and 1. Let us consider the problem of estimating p under the loss function say p minus a square so, this is a squared error loss function. Let us consider the group consisting of two elements where e is the identity element and g is an element, which takes x to n minus x here.

Now, under this transformation first of all, let us see whether the family of distributions is invariant so, if you look at the distribution of g X that is n minus X. If X follows binomial n p, then the distribution of n minus X is binomial n, 1 minus p, because n minus X denotes the number of failures the probability of a failure is 1 minus p there are n trials. So, the distribution of n minus x is binomial n, 1 minus p so, if p lies between 0 to 1 then 1 minus p also lies between 0 to 1, 1 minus p also lies between 0 to 1. So the family of distributions that is binomial n p distributions here, where n is known, but p varies between 0 to 1 this is invariant under this group G.

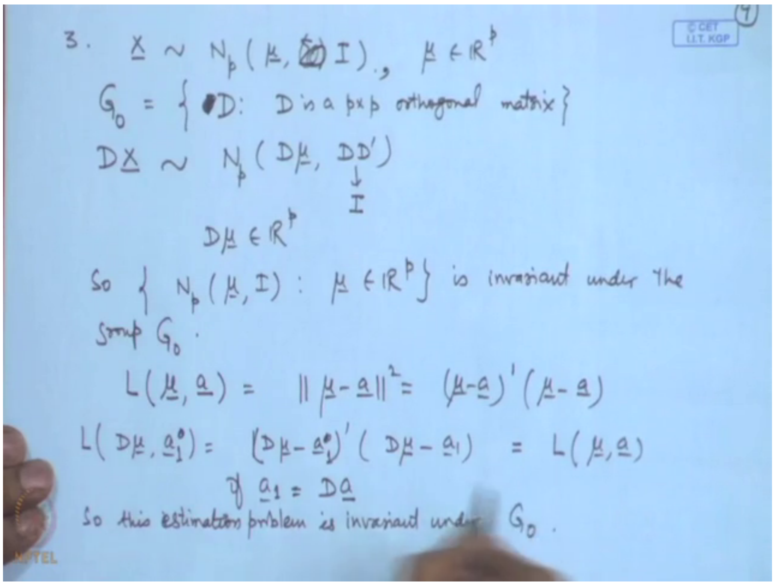
Let us look at the whether the estimation problem is invariant or not. Now, under this transformation what is g of p? That is equal to 1 minus p. So, if I consider say L of g bar p and g tilde a that is equal to g bar p minus g tilde a square, that is equal to 1 minus p minus g tilde a square now, that will be equal to p minus a square that is L p a, if g tilde a is equal to 1 minus a.

(Refer Slide Time: 29:50)



Therefore, so this estimation problem is invariant under the group G. Let us take another problem say x follows uniform distribution on the interval 0 to theta, where theta is positive.

(Refer Slide Time: 34:09)



Let us consider the loss function in estimating theta as say theta by a, a by theta minus 1 square that is a minus theta whole square divided by theta square so, in place of the squared error, we have considered a quadratic loss function. Let us consider the group of transformations g c, where g c x is equal to c x and c is positive first of all we can see whether it is a group of transformations. If you consider say composition say g of c 1 g of c 2 then that is equal to c 1 of c 2 of x that is g of c 1 c 2 of x and if c 1 is positive c 2 is positive then c 1 c 2 is also positive so, it is closed under the composition.

The identity element is given by g of 1 that is corresponding to c is equal to 1, this is the identity element. And g c inverse is actually equal to g of 1 by c because if you take g c and g of 1 by c on that then you will get the identity element. So, this is a group. We actually call it a scale group of transformations we can use the notation G s for the scale. Let us see, whether this estimation problem is invariant under the scale group. So let us consider the distribution of c X, if X follows uniform 0 theta that means the density function of f x is equal to 1 by theta between the point 0 to theta it is 0 elsewhere.

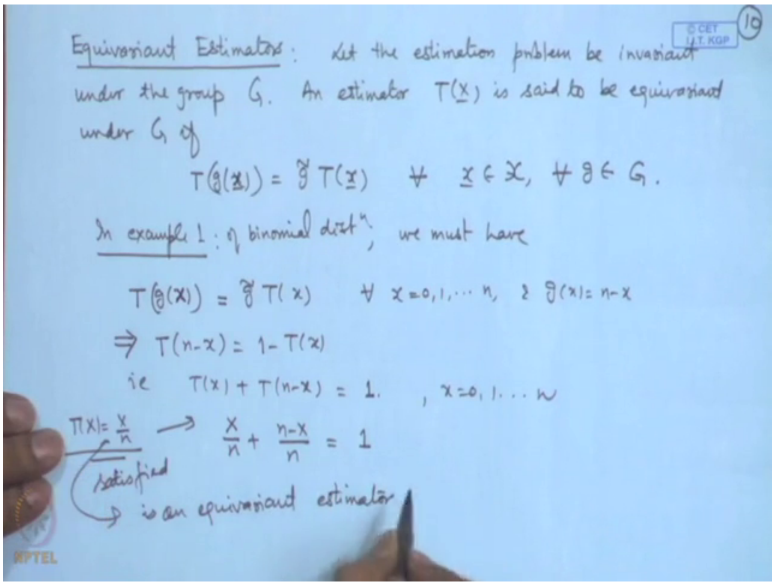
Let us consider y is equal to c X that is x is equal to y by c so, the density of y is then equal to 1 by theta, 0 less than y by c less than theta and you have d x by d y is equal to 1 by c so, 1 by c will come here that is 0 elsewhere. So, this we can write as 1 by c theta 0 less than y less than c theta 0 elsewhere. Now, notice here theta is a positive number c is positive so, c theta is also a positive number, we can replace c theta by theta prime. So, we can conclude that the family uniform 0 theta distributions, this family is invariant under the scale group of transformations G s. Let us consider X following this is the multivariate normal distribution, multivariate normal mu and say sigma or here let me put identity.

I consider the group of transformations, this is a p dimensional vector I consider the group of p by p dimensional let me use some other notation this is p here, let me use the notation say D where D is a p by p orthogonal matrix. Here mu is a p dimensional vector in the p dimensional euclidean space and I consider the group of transformations as the group of all orthogonal matrices. So, if you consider distribution of D X then that will be N p D mu, D D transpose.

But if it is orthogonal matrix then D D transpose will be equal to I so, this becomes I here. So, D mu is again a vector in the p dimensional euclidean space so, this family of distributions N p mu I, where mu belongs to R p is invariant under the group of orthogonal transformations let me put G o. Let us further introduce estimation here by taking a loss function, let us introduce a loss function as the norm of mu minus a square that is mu minus a prime mu minus a.

Let us consider say L of D mu a prime, then that is equal to D mu minus a prime in place of a prime let me write a 1 here because prime is used for transpose here D mu minus a 1. Now, this will be equal to L mu a, if a 1 is equal to D of a, because of the orthogonality D prime D will become identity. So, this estimation problem is invariant under the orthogonal group g.

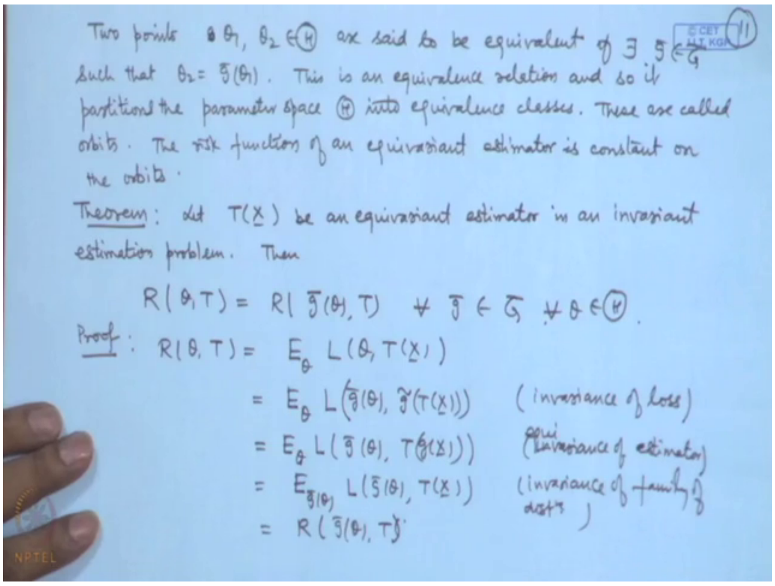
(Refer Slide Time: 37:50)



Now, we define equivariant estimators. Let the estimation problem be invariant under the group G. Then an estimator T x is said to be equivariant under the group G, if T of g x is equal to g tilde of T x for all x for all G. Let us take the example of binomial distribution that I have discussed just now. So, in the example one of binomial distribution let us consider the form of an equivariant estimator here. We should have T of g x is equal to g tilde of T x for all x is equal to 0 1 to N and g x is n minus x.

So, this condition will give us T of n minus x is equal to 1 minus T x. That is T x plus T n minus x is equal to 1. See for example, if I take T x is equal to x by n then this is satisfying this condition x by n plus n minus x by n that is equal to 1 so, this is satisfied. So, this is an equivariant estimator.

(Refer Slide Time: 41:29)

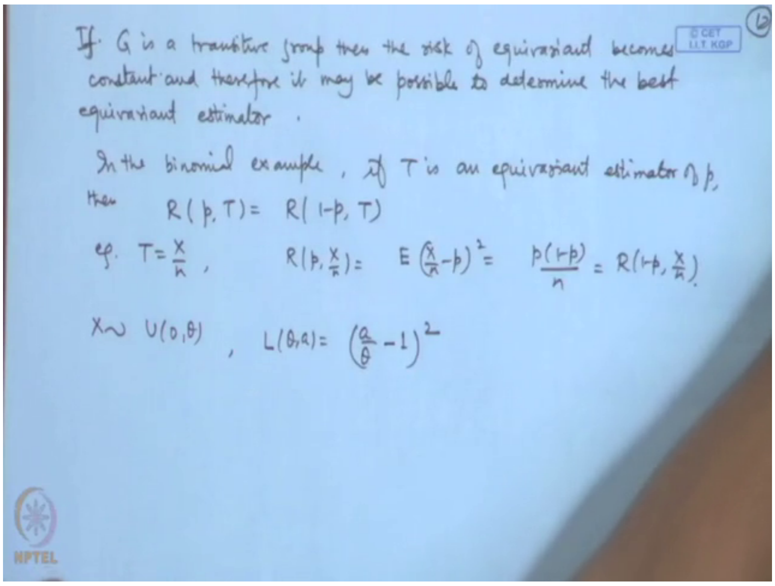


Now, one very important property about equivariant estimator is that, when we consider the risk function of the equivariant estimator, then it is constant for all parametric values where those parametric values can be reached from a given parameter point by means of g. So, let us define what is called an orbit? So, two points say theta 1 and theta 2 in the parameter space are said to be equivalent, if there exists g bar belonging to G bar such that, theta 2 is equal to g bar of theta bar. Then this is an equivalent relation and so, it partitions the parameter space into equivalence classes, these are called orbits. The risk function of an equivariant estimator is constant on the orbits, we have the following theorem.

Let T x be an equivariant estimator in an invariant estimation problem (No Audio From: 43:35 to 43:46) then the risk function of T is equal to risk function of T and g bar theta for all G bar and for all theta. Let us look at the proof of this, the risk function of T is equal to expectation of L theta T x. Now, loss function is invariant therefore, we can express it as L of g bar theta and g tilde of T x invariance of loss. Now, this we can write as expectation of L g bar theta T of g of x, because g tilde T of x is equal to T of g of x, because the invariance of estimator or rather equivariance of the estimator.

Now, if x has distribution theta then g x has a distribution g bar theta so, we can express it as g bar theta L g bar theta T of x, that is invariance of the family of distributions. But this is nothing but the risk of T at the point g bar theta.

(Refer Slide Time: 45:53)



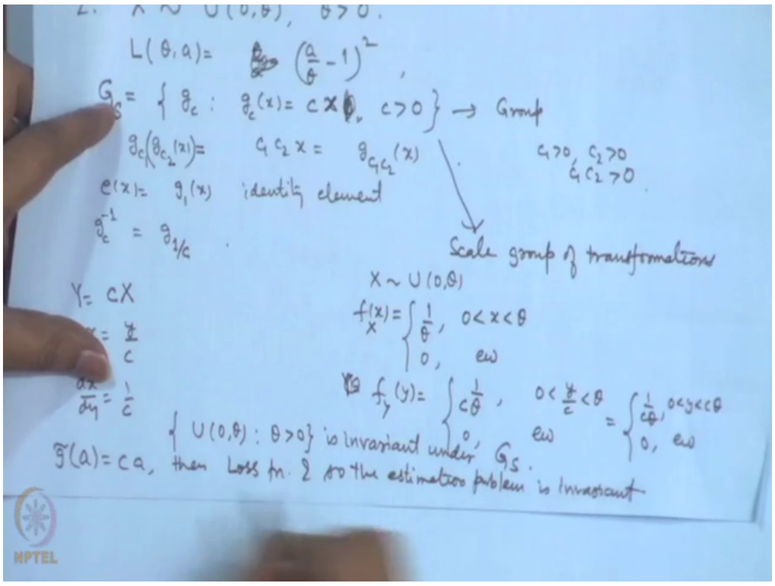
So if you have a, if g is a transitive group then the risk of equivariant estimator becomes constant. If it becomes constant you can think of minimizing it and therefore, it may be possible to determine the best equivariant estimator.

(No Audio From: 46:43 to 47:05)

Now, let us take examples here in the case of binomial distribution here, the value p is going to 1 minus p therefore, this group is not a transitive group, a transitive group means that from any point of time we can reach any other point. So, in the case of binomial distribution the risk function will be a function of in the binomial example, if T is an equivariant estimator of p then, risk of p is equal to risk of T at 1 minus p. For example, if I take T is equal to X by n then, what is the risk of X by n? That is expectation of X by n minus p square that is nothing but p into 1 minus p by n, which is same as R 1 minus p X by n.

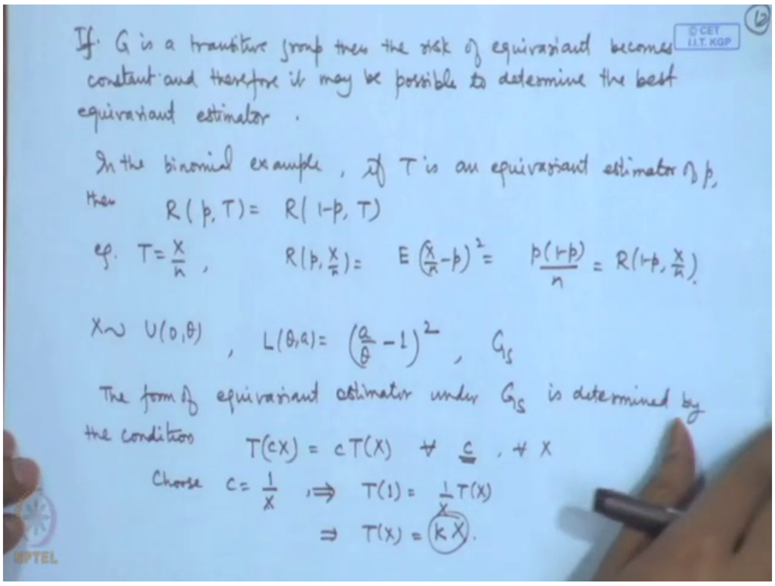
But this is not free from the parameter however, there can be situations let us consider X following say uniform 0 theta distribution and we had taken the loss function as say a by theta minus 1 square. Here our estimator will shift to here we are considering that the family of distributions is invariant under this.

(Refer Slide Time: 49:11)



But what about the problem of estimation here? If we consider a g tilde a is equal to c a, then loss function and so, the estimation problem is invariant.

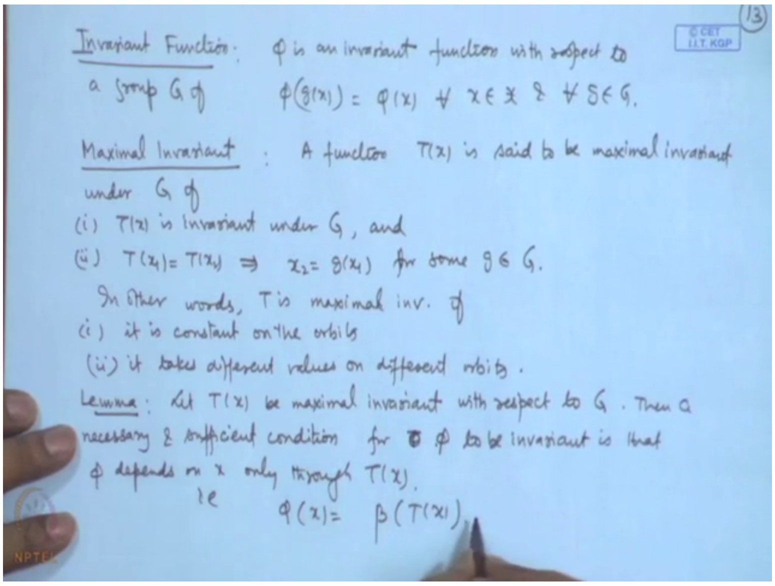
(Refer Slide Time: 49:37)



So, what will be the form of an invariant estimator then? Let us consider the form of equivariant estimator under the group G s, the scale group is determined by the condition T of c X is equal to c of T X for all c and of course, for all X. Now, here c is any positive value, choose c is equal to 1 by X. So, what we will get here? This will give us T of 1 is equal to 1 by X T of x this implies T X is of the form now, this T 1 is a constant so this is a constant times X.

So, this is the form of an equivariant estimator here that is a multiple of X. Now, this raises another question. In the previous problem of binomial distribution I substituted; I considered n minus X there in this one sense it is for all c, I have substituted c is equal to 1 by X and the form of an equivariant estimator is turning out to be k X. What is this term k x actually? Why we are substituting c is equal to 1 by X? In fact this is actually leading us to a maximal invariant. So, what is a maximal invariant function? Let me give you a definition of that here.

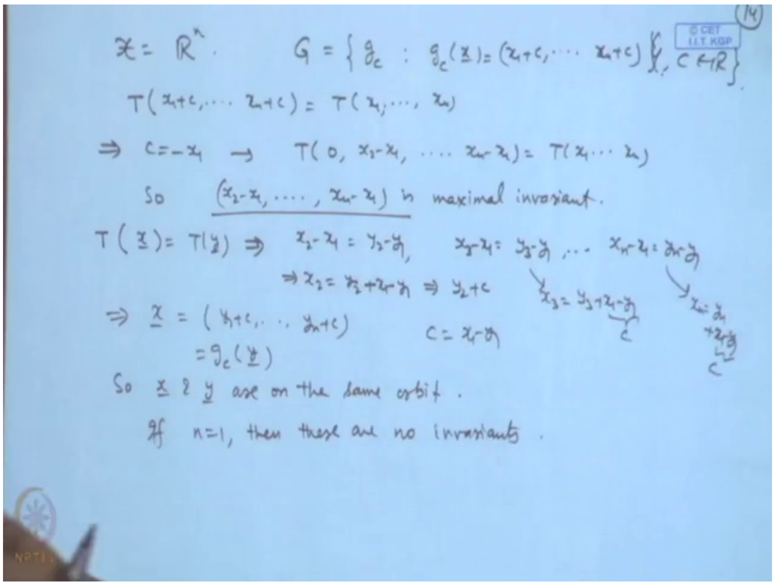
(Refer Slide Time: 51:43)



(No Audio From: 51:32 to 51:41) So, what is an invariant function firstly? Invariant function, phi is an invariant function with respect to a group G, if say phi of g x is equal to phi of x for all x and for all G. So, then we define maximal invariant. (No Audio From: 52:26 to 52:35) A function T x is said to be maximal invariant under G, if T x is invariant under G and T x 1 is equal to T x 2 implies x 2 is equal to g of x 1 for some G. That means on the say if it is on the same orbit then the value will be the same, on the different orbits that value will be different.

We can say in other words T is maximal invariant, if it is constant on the orbits and secondly it takes different values on different orbits. We have the following lemma, let T x be maximal invariant with respect to the group of transformations G, then a necessary and sufficient condition for T or say for another function phi to be invariant is that, phi depends on x only through T x. That is phi is a function of T x.

(Refer Slide Time: 55:24)



Let me explain through some example let us consider say N dimensional euclidean space and G is the group of translations. So, g c of x is equal to x 1 plus c x 2 plus c x n plus c where c is any real number. Then if we consider say T of x 1 plus c and so on, x N plus c is equal to T of x 1 x 2 x n. Then this I can choose c is equal to say minus of x 1 then that will give me T of 0 x 2 minus x 1 and so on x n minus x 1 is equal to T of x 1 x 2 x n. So, x 2 minus x 1 and so on x n minus x 1 is maximal invariant, we can prove this actually.

Let us take say T of say x is equal to T of y that means, say x 2 minus x 1 is equal to y 2 minus y 1 x 3 minus x 1 is equal to say, y 3 minus y 1 and so on x n minus x 1 is equal to say, y n minus y 1. So, each of this I can write as see this x 2 minus x 1 then I can write as x 2 is equal to say, y 2 plus x 1 minus y 1. That I can write as say, y 2 plus c similarly, here you say x 3 is equal to y 3 plus x 1 minus y 1 similarly, I can say x n is equal to y plus x 1 minus y 1.

So, this I can write as c x 1 minus y 1 so, x that means x is equal to y 1 plus 3 and so on y n plus c that is equal to g c of y. So, x and y are on the same orbit so, this is maximal invariant. If I take n is equal to 1, then there are no invariants the one which I derived just now, in the previous case when I took k x here so, there is no invariant here. Whereas, if I had taken n observations here, I would have got x 2 by x 1 x 3 by x 1 etcetera, I will explain it in the following lecture.