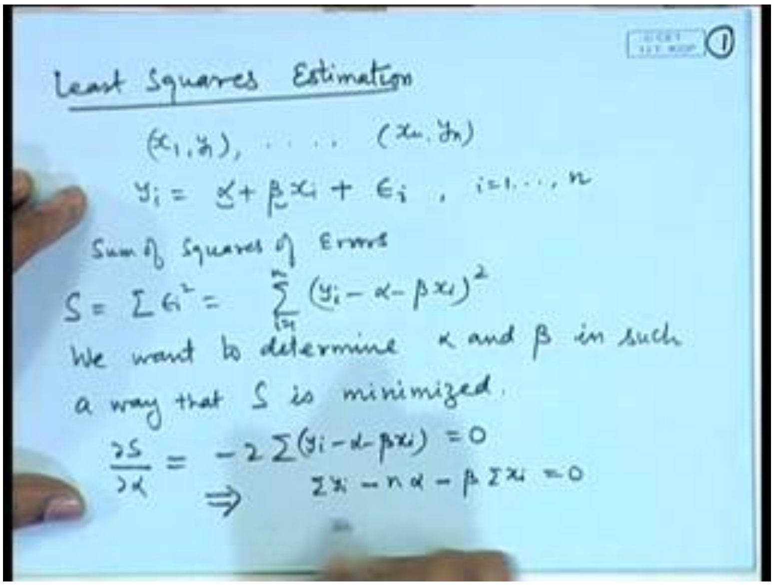
Probability and Statistics
Prof. Somesh Kumar
Department of Mathematics
Indian Institute of Technology, Kharagpur

Lecture - 55
LSE, MME

Previous lecture I introduced certain criteria for estimation; that means, if we are taking an estimator what properties it must satisfy. There are various other properties and we will be discussing them in detail. But now let me introduce the methods for finding out estimators because it is alright to say that this estimator is unbiased or this is consistent, but how do we get them. So, I mentioned in the brief introduction to the history that the initial methods that where proposed we are like method of least squares, the method of moments, the maximum likelihood estimation etcetera. Let me introduce these.

(Refer Slide Time: 01:04)

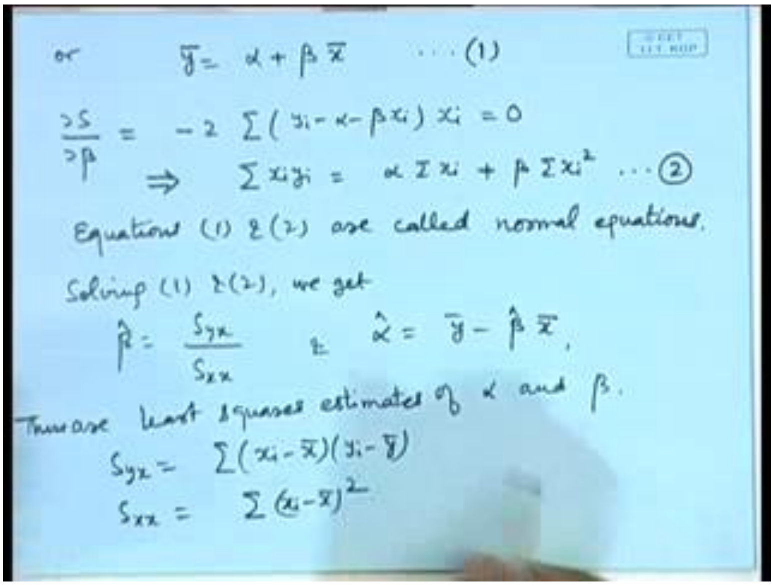


In the least square methods; so here we assume that the data is obtained in the form of variables like x 1 y 1, x 2 y 2, x n y n. For example, this may be related to certain relationship like certain variables which are related in the sense that y i's could be the weights of the persons and x i's could be the heights, x i's could be the heights of the parents and y i's could be the heights of the office springs etcetera. So, there could be various kinds of things where they are related.

However, the relationship is to be determined in the form of a linear relationship such as y i is equal to say alpha plus beta x i. So, we assume that the actual observations will introduce certain error say epsilon i's. So, the purpose is that we should estimate the parameters of the model that is alpha and beta in such a way that the sum of squares of errors that is sigma y i minus alpha minus beta x i square; let me call it S. So, in the least square methods, we want to find out alpha and beta such that this S is a minimu; excuse me, S is minimized.

Now, by looking at the nature of this function, it is easy to see that the minimizing choices of alpha and beta will be obtained when the first order derivative of S with respect to alpha and beta are equal to 0 because here it is a squared quadratic function and both alpha and beta; that means, its bowled shaped function and therefore, the minimization will be occurred when the first derivative is 0. So, del S by del alpha that is equal to minus twice sigma y i minus alpha minus beta x i is equal to 0, which we can write as sigma y i minus n alpha minus beta sigma x i is equal to 0 or we can write it as y bar is equal to alpha plus beta x bar that is the first equation.

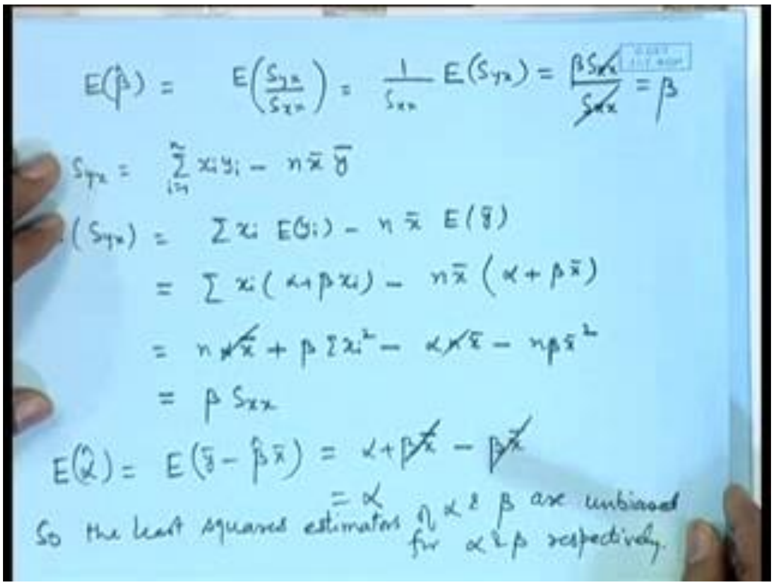
(Refer Slide Time: 04:21)



Second equation is obtained by del S by del beta equal to 0; that means, minus twice sigma y i minus alpha minus beta x i into x i is equal to 0. After simplification this is resulting in sigma x i y i is equal to alpha sigma x i plus beta sigma x i square. Let me call it equation number 2. So, equations 1 and 2 are called normal equations. So, if we solve equations 1 and 2 then we get the least square estimates of alpha and beta. So, for example, if we solve it we will get this solving 1 and 2, we get beta hat is equal to S y x by S x x and alpha hat is equal to y bar minus beta hat x bar. So, these are least squares estimates of alpha and beta.

Here S y x is sigma x i minus x bar into y i minus y bar and S x x is equal to sigma x i minus x bar whole square. It can be shown that this alpha hat and beta hat are actually unbiased. Now for that we have to make certain assumptions on the model, we have assumed that y i and x i are related through this relationship and we have introduce a random error here epsilon i. So, if we make epsilon i's are i.i.d and if i put normal 0 sigma square.

(Refer Slide Time: 07:03)

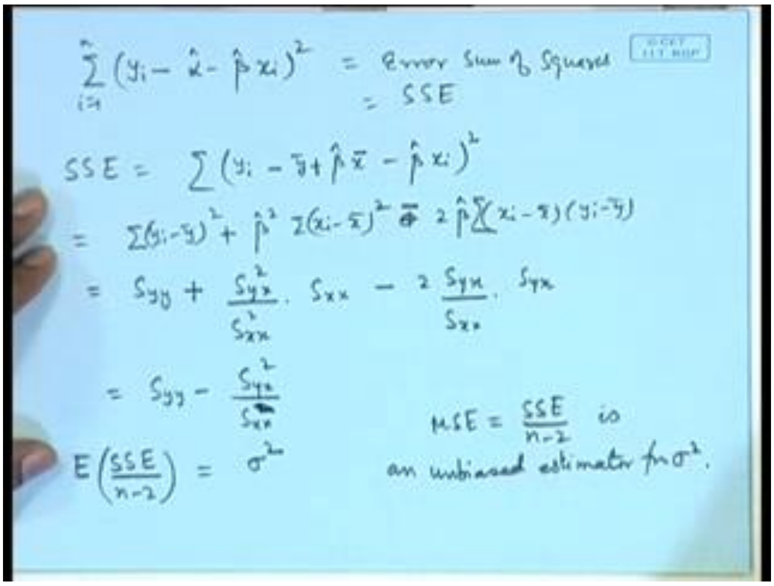


Then it can be easily shown that expectation of beta hat, so that will be expectation of S y x by S x x that is equal to; now in this model S x x I can keep out and when we are assuming here that epsilon i's are normal 0 sigma square then y i follows normal alpha plus beta x i and sigma square.

If we make use of this it is related to expectation of S y x. Now let me calculate this S y x term. So, S y x is equal 2 sigma x i y i minus n x bar y bar. So, if I take expectation of this, this is equal to sigma x i expectation of y i minus n x bar expectation of y bar. So, that is equal to sigma x i alpha plus beta x i minus n x bar. Now if I know the expectation of y i if I substitute for each of them here I will get expectation of y bar as alpha plus beta x bar. So, this term after simplification becomes n alpha x bar plus beta sigma x i square minus alpha n x bar minus n beta x bar square. So, this is becoming beta S x x. So, if we substitute this value here, I will get beta S x x by S x x which cancels out that is equal to beta. This beta at least square estimate of beta is an unbiased estimate for beta.

Similarly if I look at expectation of alpha hat that is equal to expectation of y bar minus beta hat x bar; now expectation of y bar is alpha plus beta x bar and expectation of beta hat we have prove to be beta so this cancels out. So, the least square estimators of alpha and beta are unbiased for alpha and beta respectively. We may also consider after substitution an estimate for sigma square in this model.

(Refer Slide Time: 10:18)



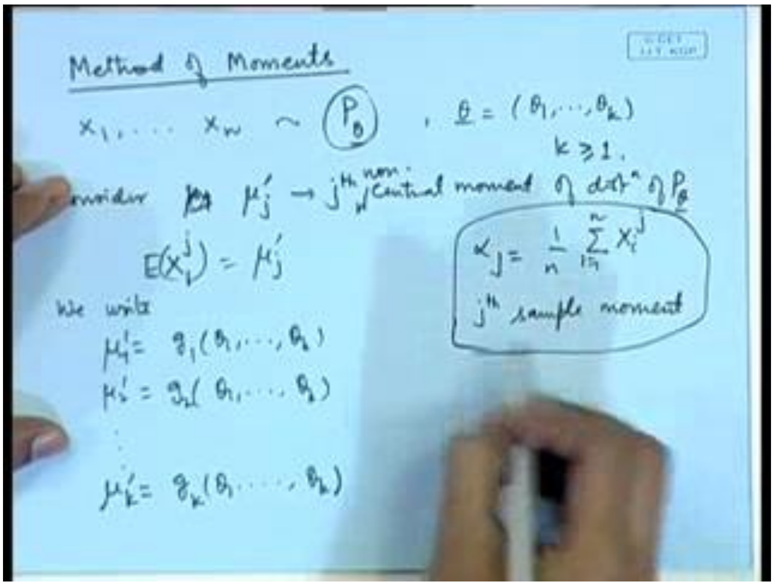
We may put y i minus alpha hat minus beta hat x i whole square what is this term our initial error some of a squares was y i minus alpha minus beta x i. Once I have estimated alpha and beta and I substitute then this is my estimated value of y I; this small y i is the actual value which has been observed, whereas from the model I can estimate it to be alpha hat plus beta hat x i. So, this is actually the error sum of squares which we call SSE. So, this is the value of this.

Now if I substitute this y i minus alpha hat. So, alpha hat is equal to y bar minus beta hat x bar minus beta hat x I, so this becomes plus here whole square. Let us simplify this term that is equal to sigma y i minus y bar whole square plus beta hat square sigma x i minus x bar whole square plus twice beta hat x i minus x bar y i minus y bar summation with a minus sin here, I have taken the cross product term here.

This is equal to S y y plus; now beta hat is equal to S y x by S x x. So, this is S y x square by S x x square into S x x minus twice S y x by S x x into S y x. So, that is equal to S y y minus S y x square by S x x square, sorry S x x. Then it can be shown that this will have expectation of SSE divided by n minus 2 that will be equal to sigma square. So, this we call MSE mean some of a squares due to error. This is an unbiased estimator. In modern statistic this particular analysis is coming under the topic of regression analysis where we study various kinds of relationships between given variables. So, suppose we are given variables x i is and y i is or x 1, x 2, x k and y there y is the response your able and x 1, x 2, x k are the explanatory variables then we fit various kind of relationships between y and x 1, x 2, x k.

And we derive the estimates of the parameters through least squares method. This methodology is also use for all types of linear models including those which are used in the analysis of variance. So, we will not perceive too much about this in this particular discussion.

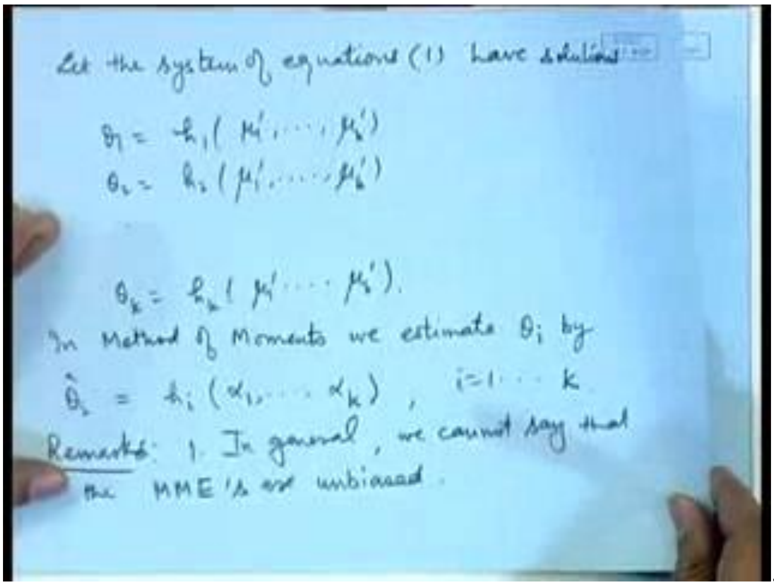
(Refer Slide Time: 14:12)



Next we consider the method of moments; this is attributed to Karl Pearson. The model is as follows, if we have a random sample say x 1, x 2, x n from a population with parameter theta. So, theta may be in general 1 dimensional 2 dimensional or k dimensional. So, here k is greater than or equal to 1, what we do? We consider mu i or say mu j prime that is the j th central moment sorry non central moment of distribution of p theta. That means, if i say that x 1, x 2, x n is a random sample form here that at means expectation of x 1 to the power j that is actually mu j prime.

In general theses moments will be certain parametric functions; for example, I may have say mu 1 prime is equal to g 1 of theta 1, theta 2, theta k mu 2 prime may be some function say g 2 of theta 1, theta 2, theta k. If I have k parametric function k parameters then I write up to kth. Now estimate let us write here I define alpha j. So, alpha j, I define to be 1 by n sigma x i to the power j i is equal to 1 to n; that means, the jth sample moment. If we consider this let us write this system of equations as 1.

(Refer Slide Time: 16:57)

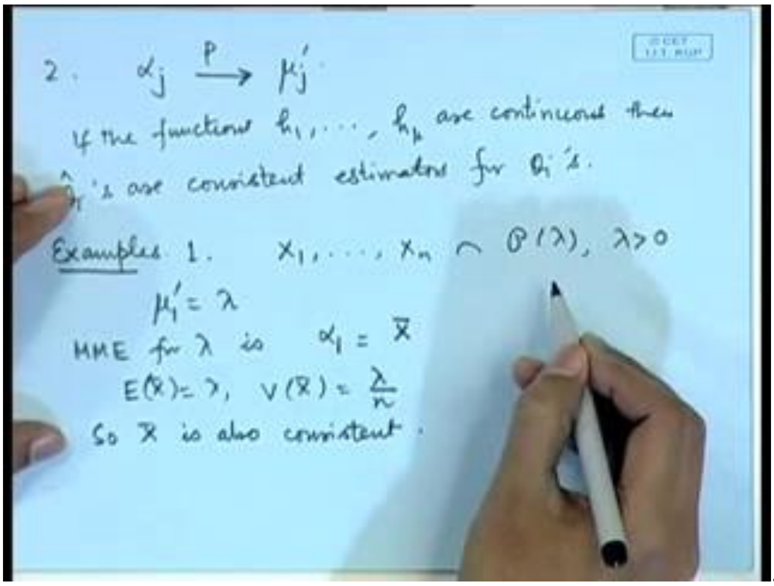


Let the system of equations 1 have solutions- say theta 1 is some function of h 1 of mu 1 prime, mu 2 prime, mu k prime; theta 2 is some function h 2 of mu 1 prime, mu 2 prime, mu k prime; theta k is equal to some function of mu 1 prime, mu 2 prime, mu k prime.

In method of moments we estimate theta i by theta i hat, let me call it h i and in place of mu 1 prime, mu 2 prime, mu k prime substitute alpha 1, alpha 2, alpha k for i is equal to 1 to k. So, these are called method of moment's estimators of the parameters. So, basically what is the philosophy behind the method of moments? I am estimating the jth population moment by the jth central moment that is mu j prime is estimated by alpha j. So, whatever parametric function is coming to us for estimation, we substitute the corresponding because the parametric functions will be sum functions of the moments and then whatever moment term is coming there we simply substitute the corresponding sample moment there.

In short this is the method of moments. So, like we had seen that the least square estimates are unbiased. But in general we cannot say that the method of moment's estimators, so I will use the word MME is in general they may not be unbiased. So, sometimes they may be unbiased, sometimes they may not be unbiased. However, consistency may be true.

(Refer Slide Time: 19:40)

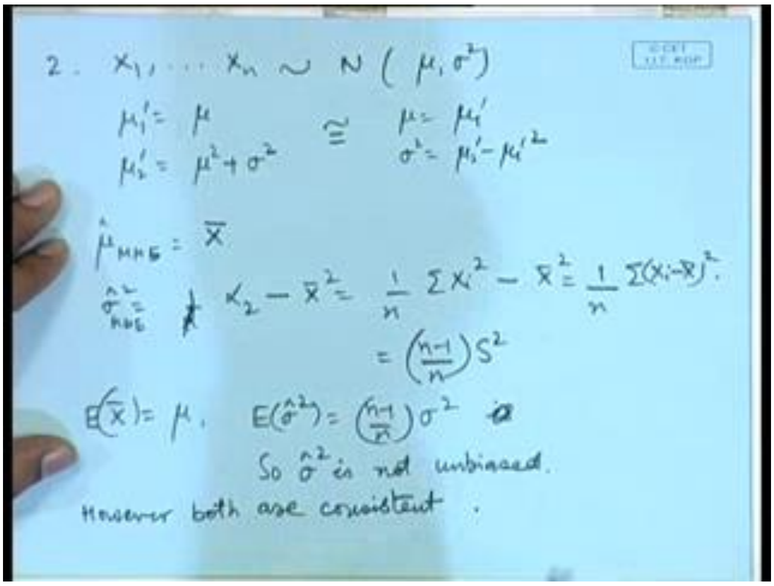


In particular we have say alpha j; this is consistent for mu j prime by the weak law of large numbers this convergence is valid. So, if the functions h 1, h 2, h k are continuous then theta i hats are consistent estimators for theta i is. So, consistency may be true and in most of the practical cases this may actually happen. However, unbiasedness is not guaranteed.

Let me explain this method by solving certain example, you may see that many time this is extremely simple, suppose I say x 1, x 2, x n follow Poisson lambda distribution. So, here you can see only 1 parameter is coming. So, we look at only the first moment mu 1 prime is lambda; that means, method of moments estimator for lambda is simply alpha 1 that is x bar. So, we had actually seen that x bar is here unbiased. In fact, it will be consistent also because variance of x bar will be expectation of x bar is lambda and variance of x bar will be actually lambda by n because the variance of a Poisson distribution is same as the beam. So, variance will become lambda by n.

So, x bar is also consistent. So, you can see actually many times this method of moments estimator may be extremely simple.

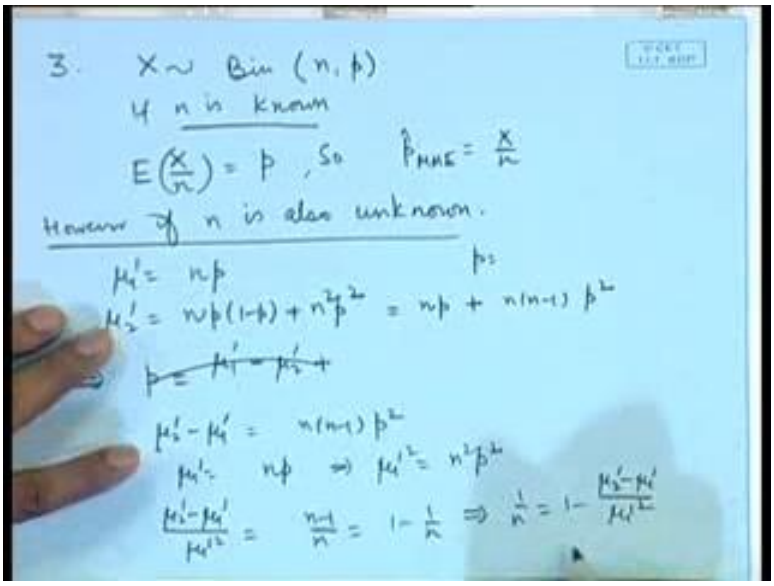
(Refer Slide Time: 21:50)



Let me take another example say x 1, x 2, x n follow normal mu sigma square, now this is a 2 parameter situation. So, we will write 2 moments. So, mu 1 prime is equal to mu, what is mu 2 prime? Mu 2 prime is mu square plus sigma square that is a second moment. So, from here mu is equal to mu 1 prime the solution and sigma square is equal to mu 2 prime minus mu 1 prime square. So, method of moments estimators will be for mu, it will be simply x bar and for sigma hat square this will be 1 by n that is your alpha 2 minus x bar square that is 1 by n sigma x i square minus x bar square which we can write as 1 by n sigma x i minus x bar whole square.

You can also write it as n minus 1 by n S square. So, which is not S square, in fact, you can see expectation of x bar is mu, but expectation of sigma hat square is n minus 1 by n sigma square. So, this is sigma hat square is not unbiased already I have proved that x bar as well as sigma hat square they are consistent.

(Refer Slide Time: 23:53)

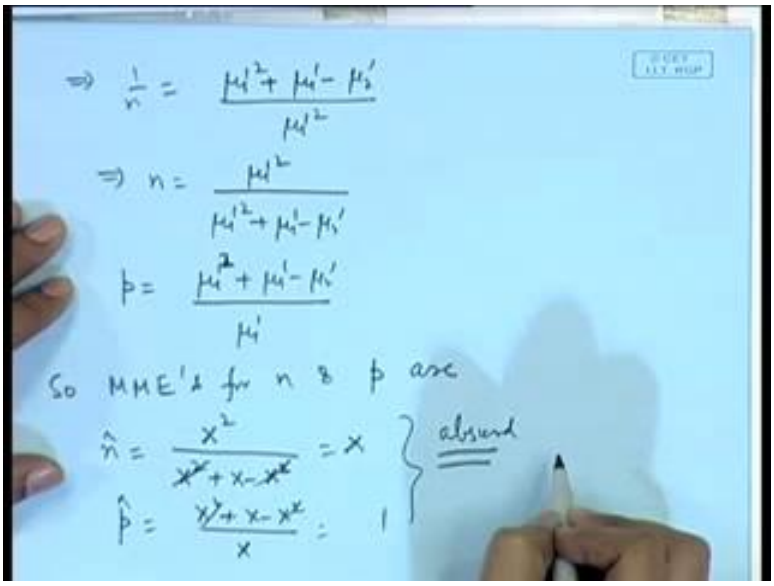


Let us consider say x following binomial n p, now if n is known then you have expectation of x by n is equal to p. So, p hat MME is simply x by n; however, if n is also unknown there may be a situation where we do not know how many number of trials have been conducted in the binomial distribution and then I have to estimate both n and p, in that case then you may write the 2 moments you write mu one prime is equal to n p and mu 2 prime is equal to n p into 1 minus p plus n square p square.

From here, we solve for n and p. So, you can look at the solutions the values will turn out to be slightly combustion we get here p is equal to mu 1 prime minus mu 2 prime plus. So, you may just write down the values here p is actually n p that is n p plus n square minus n that is n into n minus 1 p square. So, mu 2 prime minus mu 1 prime is divided by is equal to n into n minus 1 p square and mu 1 prime is equal to n p. So, this implies mu 1 prime square is equal to n square p square. So, we divide; if we divide, we will get mu 2 prime minus mu 1 prime divided by mu 1 prime square is equal to n minus 1 by n that is 1 minus 1 by n.

This implies 1 by n is equal to 1 minus mu 2 prime minus mu 1 prime divided by mu 1 prime square.

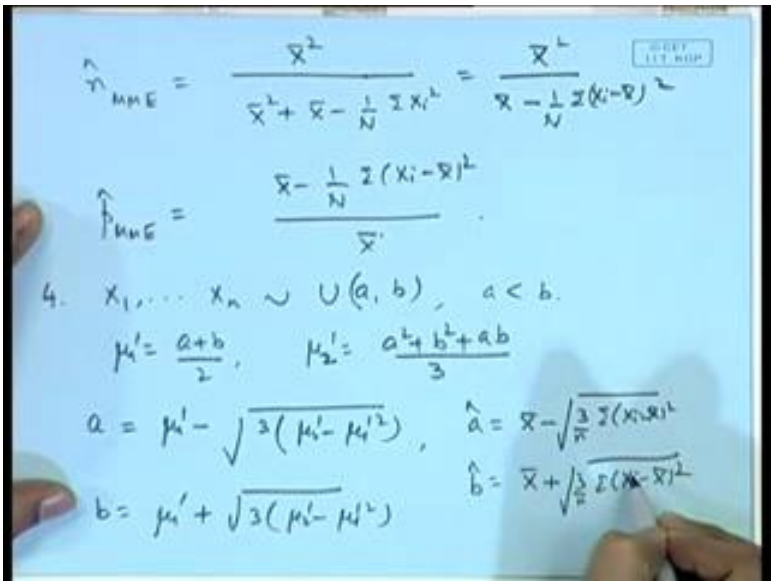
(Refer Slide Time: 26:53)



This we can write as 1 by n is equal to mu 1 prime square plus mu 1 prime minus mu 2 prime by mu 1 prime square. So, n is equal to mu 1 prime square by mu 1 prime square plus mu 1 prime minus mu 2 prime similarly p is equal to because from the first 1 p is equal to mu 1 prime by n. So, if n is already determined we just substitute there, so, mu 1 prime divided by this so that will give me mu 1 prime square plus mu 1 prime minus mu 2 prime by mu 1 prime. So, method of moment's estimators for n and p are so for n, it will become now here I have taken only 1 observation x. So, we simply substitute x square divided by now this will lead to some peculiar problem which you can see x square plus x minus x square.

This cancels out, you get only x, if you put p then you will get x square plus x minus x square divided by x which is canceling out and you get only one this is leading to absurd situation. Now why this is coming? Since I have here 2 observations, 2 parameters n and p it is not possible to estimate both of them with one observation; that means, I need to take a sample here. So, when n is known it is alright that is use x by n, but if n is unknown we need sample. So, let me say sample is x 1 x 2 x capital N. So, in that case this situation can be resolved.

(Refer Slide Time: 29:07)



So, here the MME is will be N hat MME is equal to now x bar square divided by x bar square plus x bar minus 1 by n sigma x i square which we can also write as x bar square divide by x bar minus 1 by n sigma x i minus x bar whole square.

And p hat MME will be equal to x bar minus 1 by n sigma x i minus x bar whole square divided by x bar. So, here you can see the form is quite complicated and the question of checking unbiasedness etcetera is ruled out because we cannot actually evaluate the expectations of ratios of this type of functions consistency, you can still be considered because x bar will be consistent for p and for n p; that means, if I considered x bar by capital N then that will be consistent for p etcetera. So, the consistency may hold, but the unbiasedness is totally ruled out. In fact, it cannot be even checked let us take another example, suppose we consider a 2 parameter uniform distribution, in a 2 parameter uniform distribution we have a and b as the parameters.

Now, let us consider say first moment here the first moment is a plus b by 2 and the second moment is a square plus b square plus a b by 3, How many times you will see that when we have multi parameter situation the solutions of the equation may not be trivial because the equations need not be necessarily linear in general they may be non-linear equations as we have seen in the binomial case and same thing is true in the uniform distribution case also. So, if you solve these things, you will get a as mu 1 prime minus a square root 3 into mu 2 prime minus mu 1 prime square and b is equal to mu 1 prime plus square root 3 mu 2 prime minus mu 1 prime square.

The method of moments estimators are obtained by substituting alpha 1 and alpha 2 forming 1 prime and mu 2 prime. So, I will get x bar minus a square root 3 by n sigma x i minus x bar whole square and b hat is equal to x bar plus root 3 by n sigma x i minus x bar whole square. So in fact, you can see that many times the form of the method of moment's estimators may not be very convenient to handle. In fact, again if I ask here to check the unbiasedness expectation of x bar may be a plus b by 2, but calculation of the expectation of this quantity is not that simple and therefore, in general the method of moments estimator does not seen to give very nice looking estimates.

In some of the situations of course, like in the Poisson distribution case are normal distribution case; we got nice solutions, but in many of the 2 parameter are more number of parameter situations that the method of modes estimators may not be always very nice.