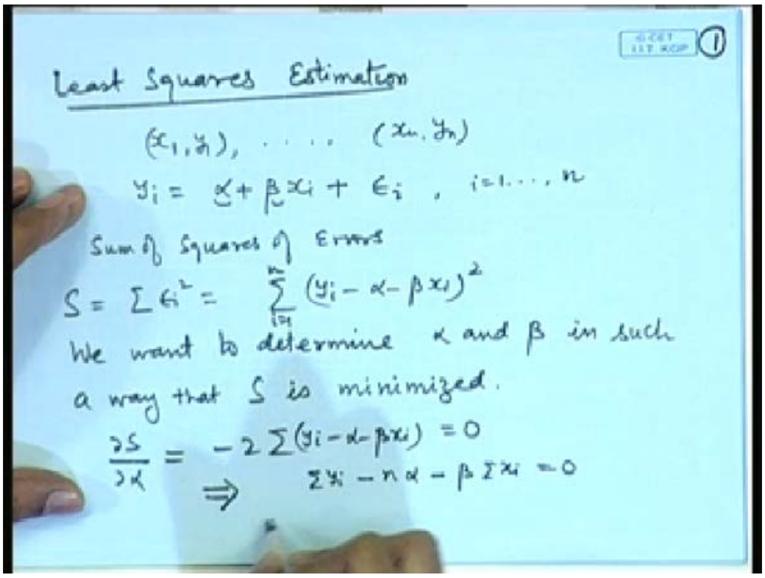
Probability and Statistics
Prof. Dr. Somesh Kumar
Department of Mathematics
Indian Institute of Technology, Kharagpur

Module No. # 01
Lecture No. # 28
Estimation-II

Our 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 were proposed were like methods of least square, the method of moments the maximum likelihood estimation etcetera. So, let me introduce these.

(Refer Slide Time: 01:05)



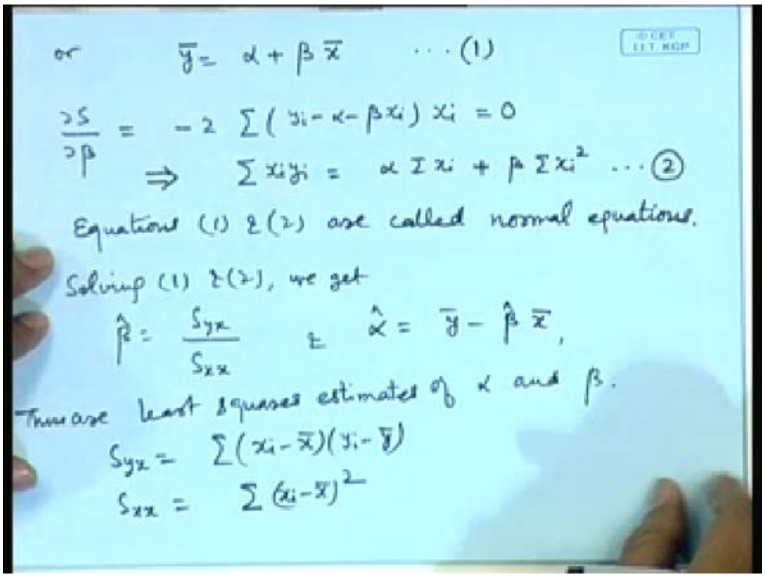
So, in the least square methods least squares. Here we assume that the data is obtained in the form of variables like x 1, y 1, x 2, y 2, x n, y n. For example, this may be related to certain relationship like certain variables which are related in the sense that y i's could be the weights of the persons 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 off springs etcetera. So, there could be various kind 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 errors 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 minimum excuse me s is minimized.

Now, by looking at the nature of this function it is easy to see that the minimizing choices of an 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 squared quadratic function in both alpha and beta. That means, it is a bowl 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.

(Refer Slide Time: 04:17)

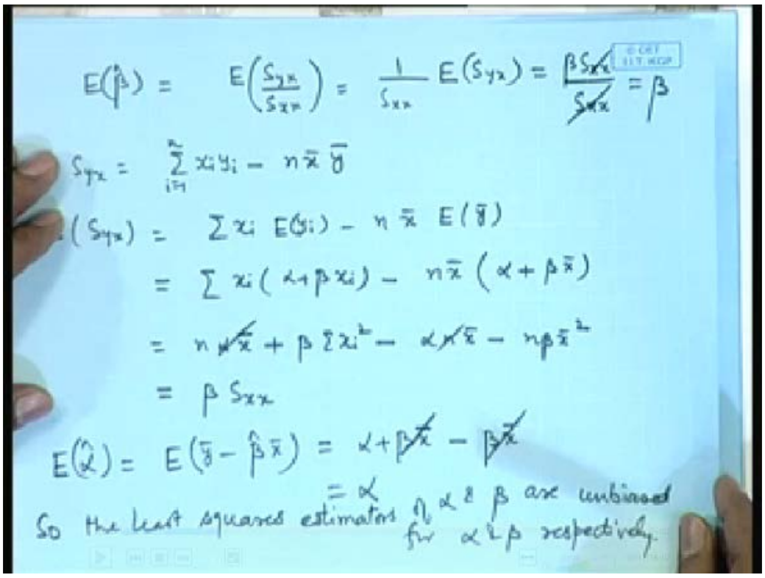


Or, we can write it as y bar is equal to alpha plus beta x bar that is the first equation. Second equation you obtain by del s by del beta equal to 0; that means, minus twice sigma y i minus alpha minus beta x i into x i is equal to 0. After simplification this is resulting in sigma x i y i is equal to alpha sigma x i minus 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 head is equal to s y x by s x x and alpha head is equal to y bar minus beta head 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 head and beta head 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 introduced a random error here epsilon i. So, if we make epsilon i is a I i d and if I put normal 0 sigma square then it can be easily shown that expectation of beta head.

(Refer Slide Time: 07:00)



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 is a normal 0 sigma square then y i follows normal alpha plus beta x i and sigma square. So, 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 to 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.

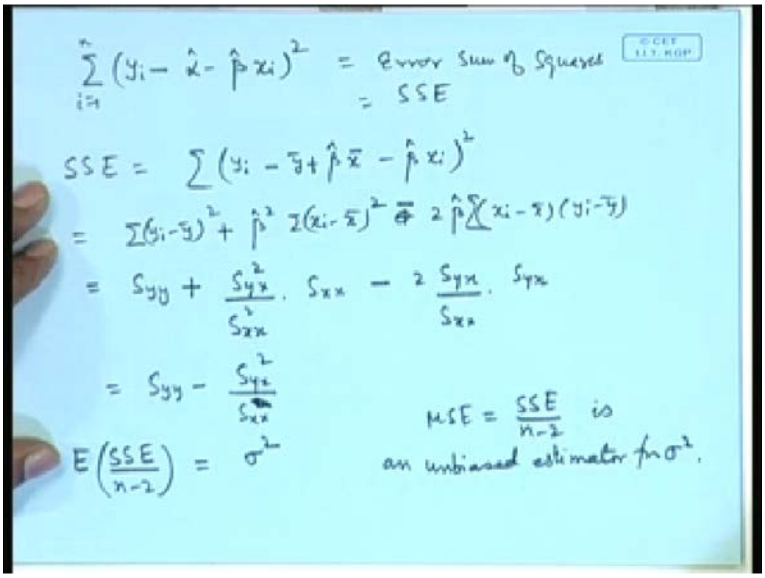
So, if we substitute this value here I will get beta s x x by s x x which cancels out that is equal to beta. So, this beta had least square estimate of beta is an unbiased estimate for beta

Similarly, if I look at expectation of alpha head that is equal to expectation of y bar minus beta head x bar

Now, expectation of y bar is alpha plus beta x bar and expectation of beta head we have proved to be beta. So, this cancels out. So, the least squares estimates 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: 10:17)



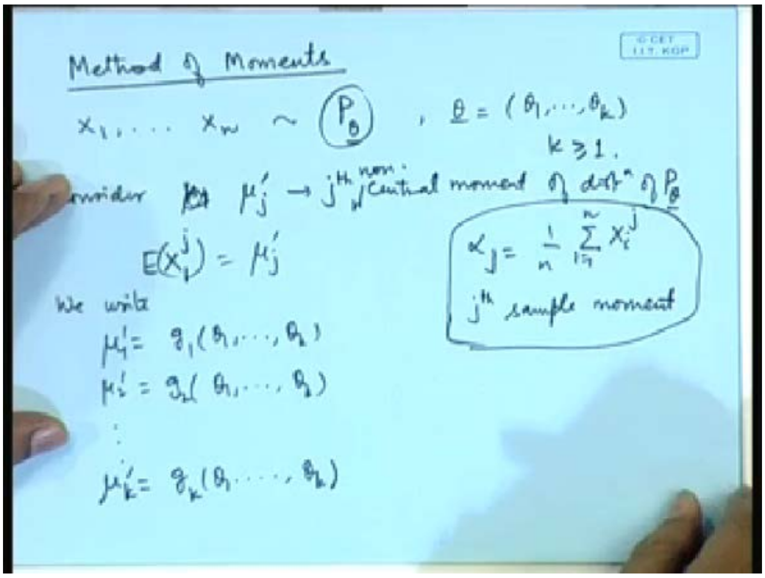
We may put y i minus alpha head minus beta head x i whole square what is this term our initial error sum of squares was y i minus alpha minus beta x i once I have estimated alpha and beta and I substitute then this is my estimated value of y i this small y i is the actual value which has been observed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

So, this is actually the error sum of squares which we call s s e. So, this is the value of this now if I substitute this y i minus alpha head. So, alpha head is equal to y bar minus beta head x bar minus beta head 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 head square sigma x i minus x bar whole square plus twice beta head x i minus x bar y i minus y bar summation with a minus sign here. I have taken the cross product term here. So, this is equal to s y y plus now beta head 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 s s e divided by n minus 2; that will be equal to sigma square. So, this we call m s e mean sum of squares due to error this is an unbiased estimator.

In modern statistic this particular analysis is coming under the topic of regulation analysis where we study various kind of relationships between given variables.

So, suppose we are given variables x i(s) and y i(s) or x 1 x 2 x k and y where y is a response variable 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 used for all types of linear model including those which are used in the analysis of variance. So, we will not pursue too much about this in this particular discussion.

(Refer Slide Time: 14:09)

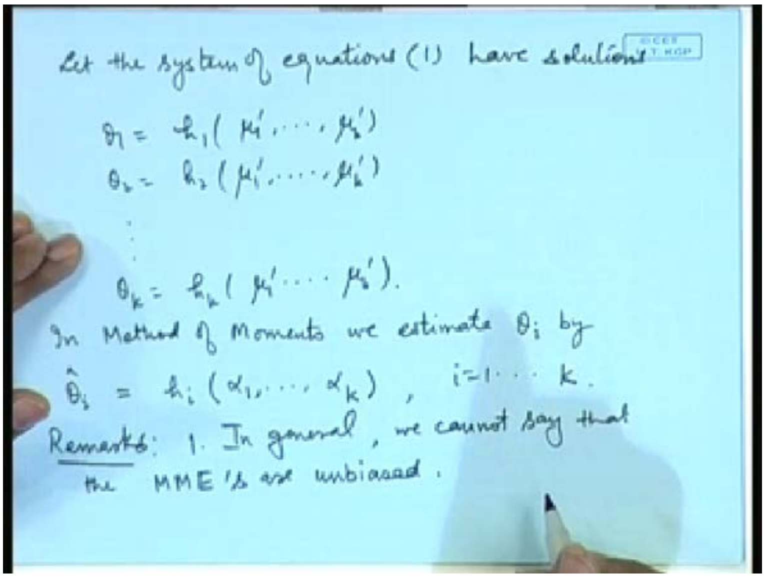


Next, we consider the method of moments; this is attributed to Carl Pearson. The model is as follows: if we have a random sample say x 1 x to x n from a population with parameter theta, so, theta may be in general 1 dimensional 2 dimensional or k dimensional. So, here k is greater than or equal to 1.

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 from here that at mean expectation of x 1 to the power j that is actually mu j prime.

So, in general these moments will be certain parametric functions. For example, I may have say, mu 1 prime is equal to g 1 of theta 1 theta to 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 k th row 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 j th sample moment. If we consider this let us write this system of equations as 1 ok

(Refer Slide Time: 16:55)



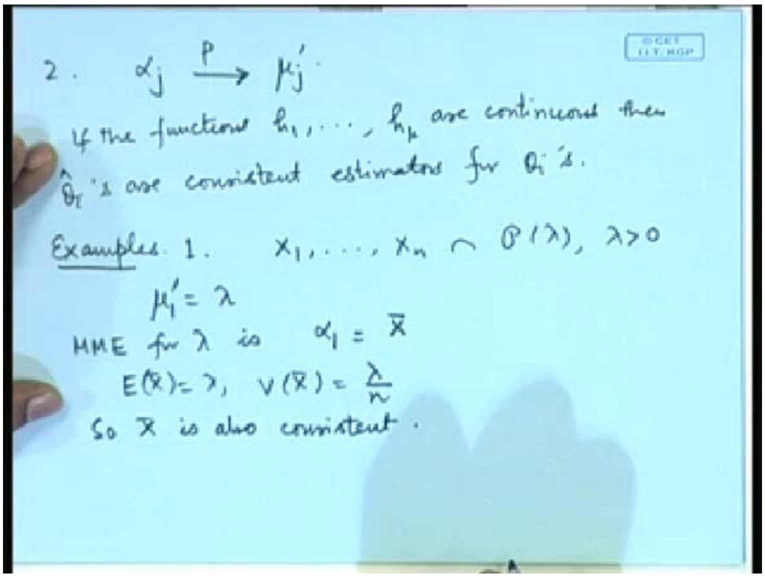
Let the system of equations 1 have solutions say theta 1 is sum function of h 1 of mu 1 prime mu 2 prime mu k prime theta 2 is sum function h 2 of mu 1 prime mu 2 prime mu k prime theta k is equal to sum function of mu 1 prime mu 2 prime mu k prime.

In method of moments, we estimate theta i by theta i head. 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 moments estimators of the parameters. So, basically, what is the philosophy behind the methods of moments? I am estimating the j th population moment by the j th central moment; that is mu j prime is estimated by alpha j.

So, whatever parametric function is coming toward 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. So, 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 moments estimators. So, I use a word m m e‘s

In general, they may not be unbiased. So, sometimes they may be unbiased sometimes they may be may not be unbiased; however, consistency may be true in particular.

(Refer Slide Time: 19:38)

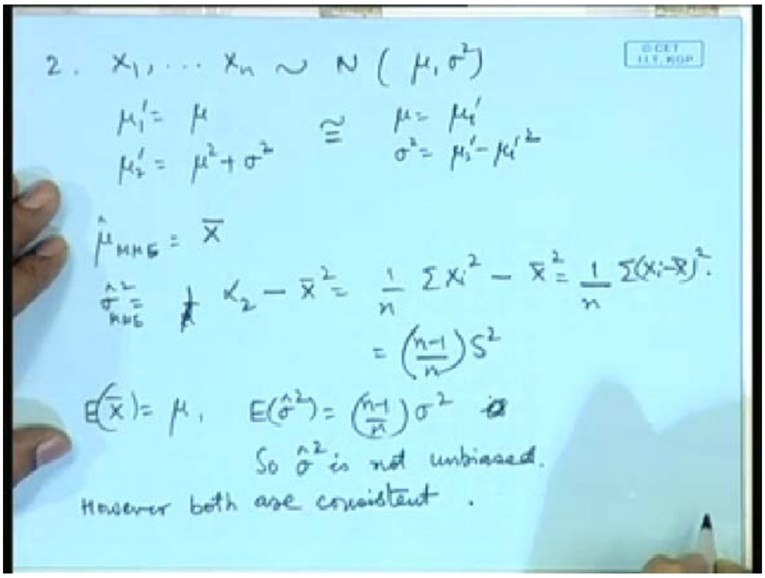


We have say alpha j; this is consistent for mu j prime by the last numbers - this convergence is valid. So, if the functions h 1 h 2 h k are continuous then theta i heads are consistent estimators for theta i's. So, consistency may be true and in most of the practical cases this may actually happen; however, unbiasness is not guaranteed

Let me explain this method by solving certain example. You may see that many times 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 parson distributions is same as the mean. So, variance will become lambda by n. So, x bar is also consistent.

So, you can see actually in many times this method of moment estimator may be extremely simple; let me take another example.

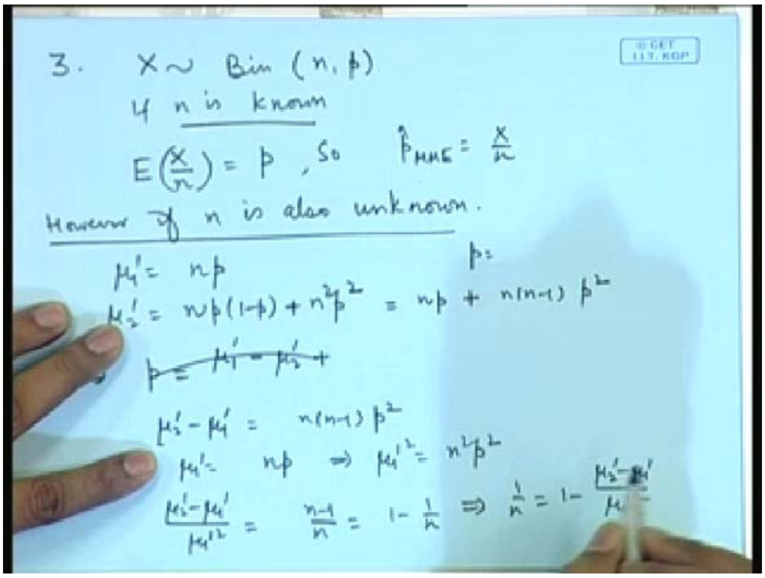
(Refer Slide Time: 21:50)



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 head 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 x square.

So, actually it is not x square. In fact, you can see expectation of x bar is mu, but expectation of sigma head square is n minus 1 by n sigma square. So, this is sigma head square is not unbiased already I have proved that x bar as well as sigma head square; they are consistent.

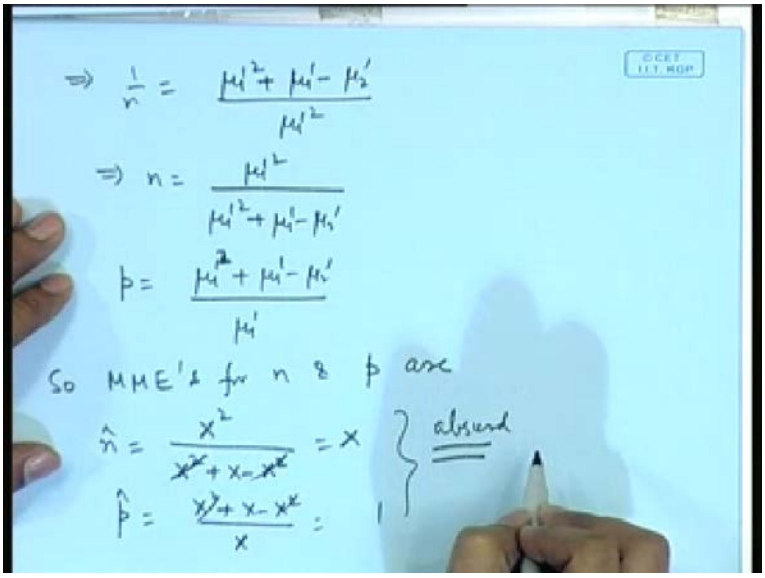
(Refer Slide Time: 23:47)



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 head m m e 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 trails 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 1 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 cumbersome we get here p is equal to mu 1 prime minus mu 2 prime plus. So, we 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 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)



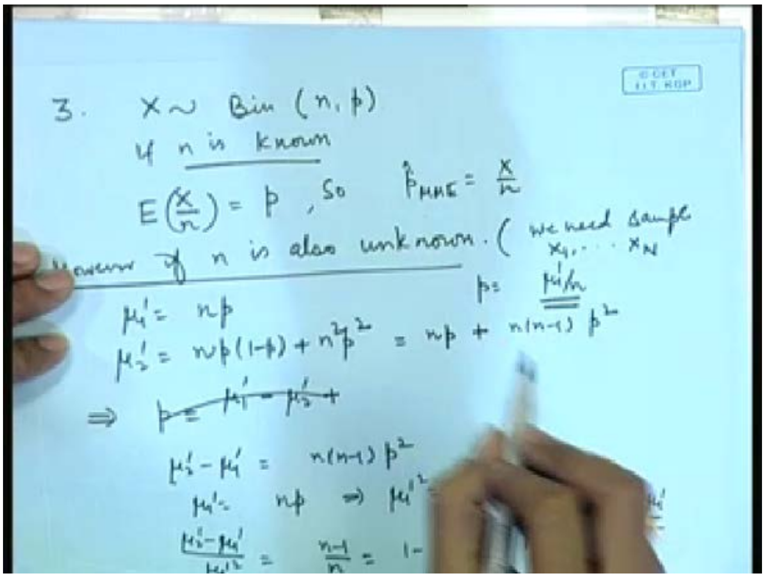
This we can write as 1 by n is equal to mu 1 prime square plus mu 1 prime minus mu 2 prime by mu 1 prime square. So, n is equal to mu 1 prime square by mu 1 prime square plus mu 1 prime minus mu 2 prime.

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; so, we just substitute there. So, mu 1 prime divided by this (Refer Slide Time: 27:16). So, that will give me mu 1 prime square plus mu 1 prime minus mu 2 prime by mu 1 prime.

So, method of moments 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.

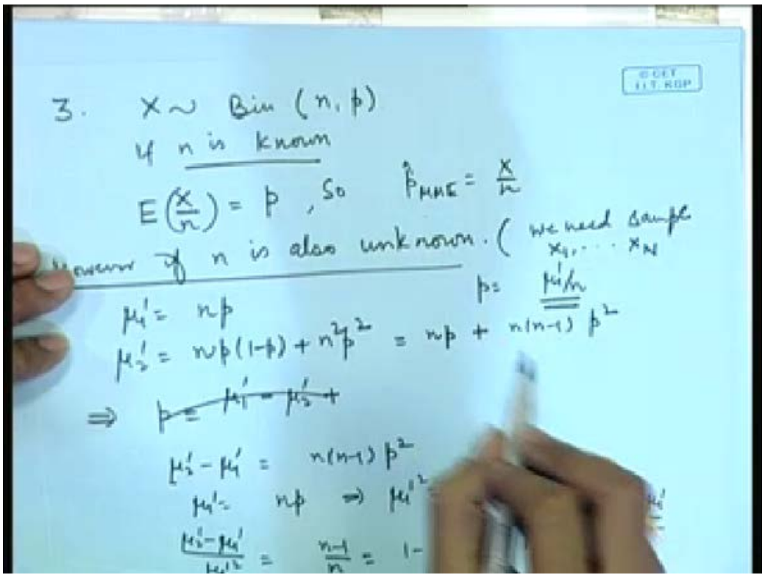
If you put p then you will get x square plus x minus x square divided by x which is cancelling out and you get only 1. This is leading to absurd situation. Now, why this is coming? Since I have here 2 observations, 2 parameters n and p, it is not possible to estimate both of them with 1 observation; that means, I need to take a sample here.

(Refer Slide Time: 28:42)



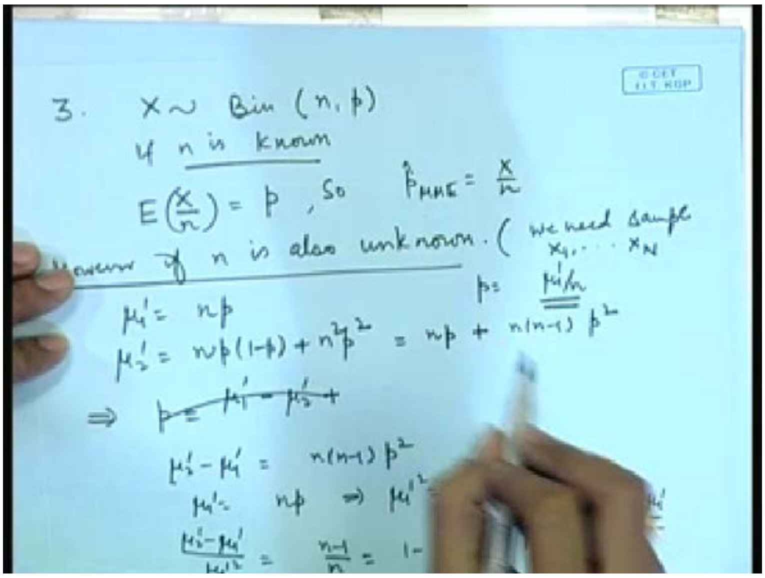
So, when n is known it is alright that I use x by n, but if n is unknown we need sample. So, let me say sample is x 1 x 2 x capital N.

(Refer Slide Time: 29:00)



So, in that case this situation can be resolved.

(Refer Slide Time: 29:07)



So, here the m m e‘s will be: N head m m e 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 divided by x bar minus 1 by N sigma x i minus x bar whole square and p head m m e 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 unbiasness, etcetera, is ruled out because, we cannot actually evaluate expectations of ratios of this type of functions. Consistency can still be considered because, X bar will be consistent for p and for n p. that means, if I consider x bar by capital n then it will be consistent for p, etcetera. So, the consistency may hold, but the unbiasness is totally ruled out. In fact, it cannot be even checked.

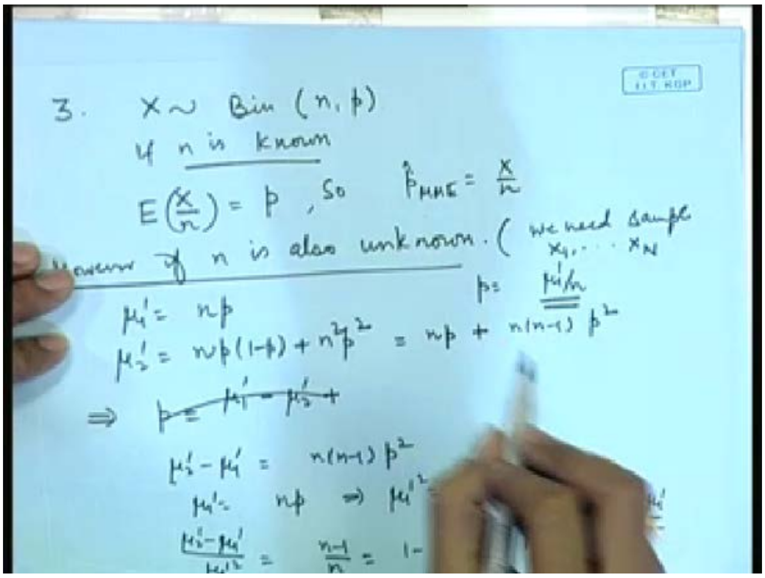
Let us take another example. Suppose, we consider a 2 parameter uniform distribution in a 2 parameter uniform distribution we have a and b as the parameters.

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. Now, 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 you 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 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. So, the method of moments estimators are obtained by substituting alpha 1 and alpha 2 for mu 1 prime and mu 2 prime. So, I will get x bar minus square root 3 by n sigma x i minus x bar whole square and b head 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 moments estimators may not be very convenient to handle. In fact, again I ask here to check the unbiasness expectation of x bar may be a plus b by 0. But, calculation of the expectation of this quantity is not that simple and therefore, in general the method of moment’s estimator does not seem to give very nice looking estimates. In some of the situations of course, like in the Poisson distribution case, or normal distribution case, we got nice solutions. But, in many of the 2 parameters or more number of parameter situations, that method of moments estimators may not be always very nice. Let me take one more example here.

(Refer Slide Time: 33:36)

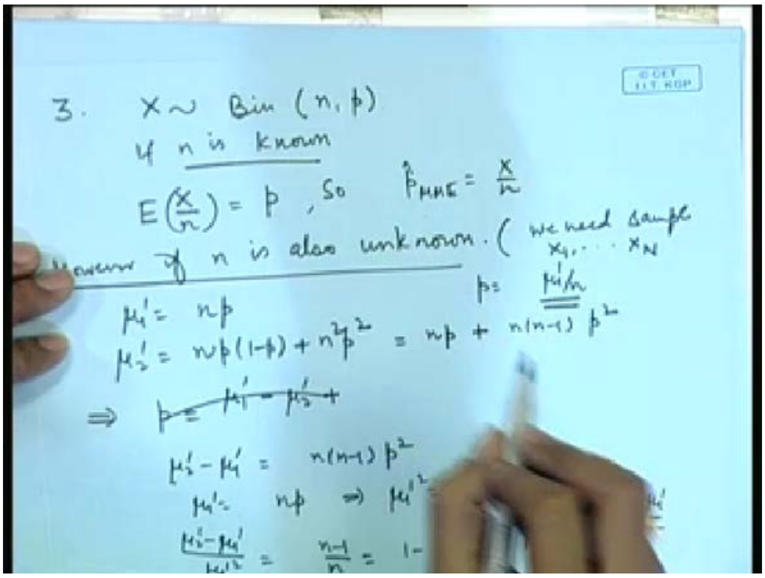


Let x 1 x 2 x n follow gamma distribution and I will consider 2 parameter situations. So, a 2 parameter situation gamma p alpha where p is the shape parameter and 1 by alpha is the scale parameter. So, here mu 1 prime is equal to p by alpha and mu 2 prime is equal to p into p plus 1 by alpha square.

So, once again if we want to solve it, so, we will get it as p square by alpha square plus p by alpha square. So, mu 2 prime minus mu 1 prime square is equal to p by alpha square let us take the ratio here. So, alpha is equal to mu 1 prime by mu 2 prime minus mu 1 prime square and therefore P is equal to mu 1 prime square by mu 2 prime minus mu 1 prime square. So, alpha head m m e is equal to x bar divided by 1 by n sigma x i minus x bar whole square and p head m m e is equal to x bar square divided by 1 by n sigma x i minus x bar square.

Notice here that if I had taken p to be known here then the problem will be extremely simple because, in that case I can simply substitute in the first equation itself x bar and the estimate of alpha will become p by x bar. However, if 2 parameter situation is there the solutions are not straight forward to obtain.

(Refer Slide Time: 35:32)



Let us take say, beta distribution x 1 x to x n follows a beta distribution with parameter say alpha and beta here mu 1 prime is equal to alpha by alpha plus beta and mu 2 prime is equal to alpha into alpha plus 1 divided by alpha plus beta into alpha plus beta plus 1.

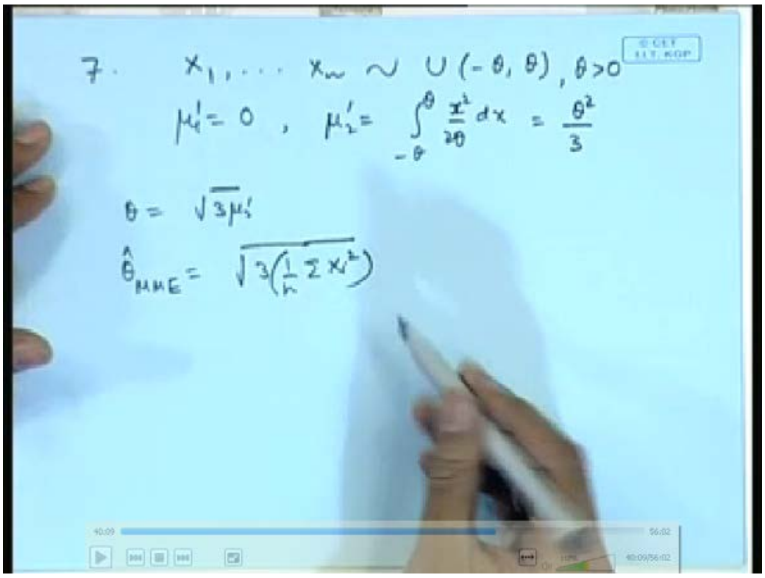
Now, if we solve these equations, this is quite cumbersome and the solutions turn out to be alpha is equal to mu 1 prime into mu 1 prime minus mu 2 prime divided by mu 2 prime minus mu 1 prime square and beta is equal to 1 minus mu 1 prime into mu 1 prime minus mu 2 prime divided by mu 2 prime minus mu 1 prime square.

So, substituting for mu 1 prime as x x bar and for mu 2 prime 1 by n sigma x i square, we get the method of moments estimators as x bar into. So, if I take this 1 x bar minus 1 by n sigma x i square divided by 1 by n sigma x i minus x bar whole square similarly beta head m m e is equal to 1 minus x bar into x bar minus 1 by n sigma x i square divided by 1 by n sigma x i minus x bar whole square.

In all these cases, you can see that checking of the unbiasness is extremely difficult. In fact, there will not be unbiased. Of course, consistency may still hold most of the situations. If the denominator is non vanishing for example, here, the x bar will be consistent for mu 1 prime and 1 by n sigma x i square will be consistent for mu 2 prime and therefore, there will be consistent, but not unbiased. Same thing is true in the gamma case also; that these 2 will be consistent as x bar will be consistent for mu 1 prime and sigma x i square by n will be consistent for mu 2 prime, but again they will be not be unbiased.

So, this leads to… and another thing that we are observing is that the calculation of the estimates is not in a very convenient looking form although the distribution are not too complicated because, these are some of the basic distribution like gamma distribution, uniform distribution or beta distribution. So, the estimates are not looking very nice. Another thing could be that here we have seen that we have to write the number of equations equal to the number of parameters.

(Refer Slide Time: 38:50)

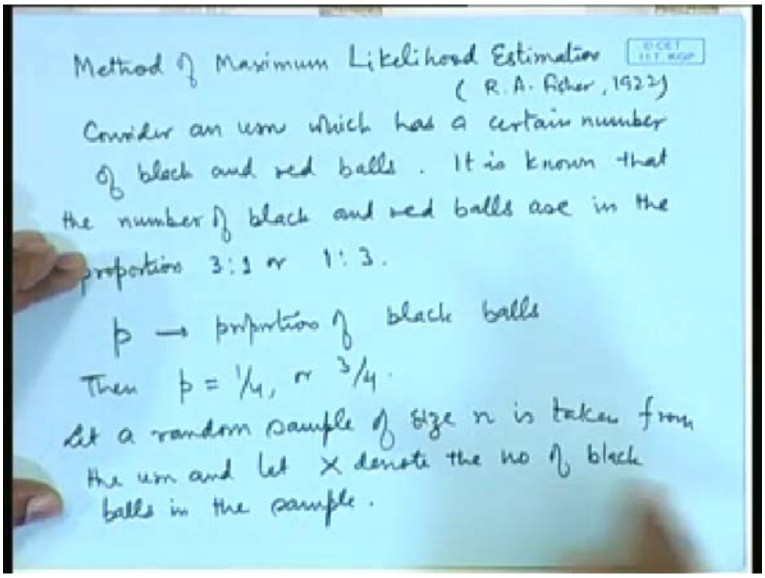


But, there may be some situations where this kinds of situations may arise. Like, if I consider say, uniform minus theta to theta; that means, it is a uniform distribution on a symmetric interval, but we do not know the range of the interval. Therefore, theta is unknown. So, here the problem is to estimate theta if I consider the first moment actually it is 0. So, this does not gives us any information about how to estimate the theta.

So, now if we go by strictly by the rule of the method of moments estimation then this is not an estimable function then however, some practical solutions can be given. We may look at mu 2 prime; that means, a second moment that is 1 by 2 theta and we look at x square d x from minus theta to theta then it is equal to theta square by 3. So, if we make use of this then theta is equal to root 3 mu 2 prime and we may use the method of moments estimator as 3 and the 1 by n sigma x i square. So, this is not strictly in accordance with the method of moments; but, anyway this can be considered as a practical solution.

Similar kind of situation may occur in various multi parameter cases where one or the other of the parameter may not be directly coming in the form of the expectation. There, another method of estimation that was introduced in the first of the 28th century is the method of maximum likelihood.

(Refer Slide Time: 40:43)



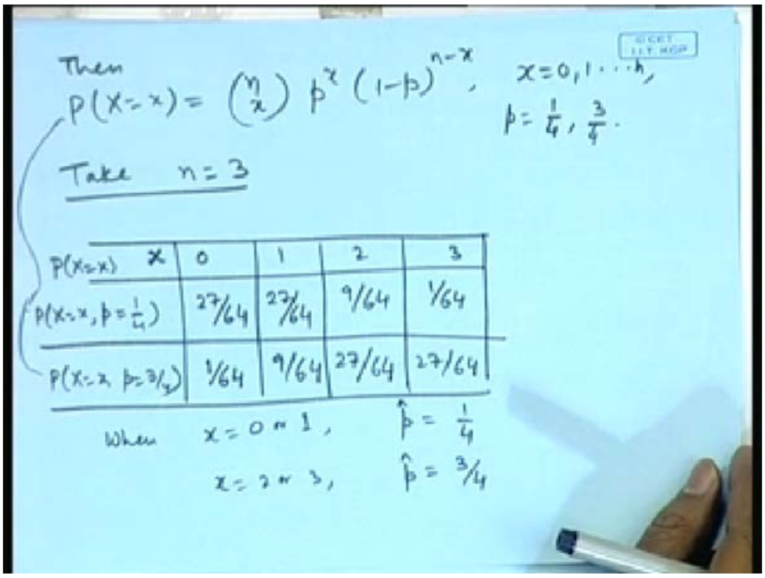
estimation this was introduced by r a fisher in 1922.

The earlier methods like the least square method - it is based on the data. That means, if I have x i y i and we write a relationship and then we estimate the parameter in the method of moments we look at the form of the distribution and we basically look at the moments the moment structure are used however, fisher said that we should may use of the full probability model because the moments structure does not make use of the full full use of the distribution because 2 different distributions may have the equal moments say first moment or second moment may be equal for 2 distributions and still you may get the different estimates. So, the method of maximum likelihood said that we should look at the form of the distribution itself and make full use of that.

So, let us consider a simple example. Consider an urn which has a certain number of say black and red balls. So, it is known that the number of black and red balls are in the proportion; it is either 3 is to 1 or 1 is to 3. That means, we do not know which one or more numerous. So, if we say that there are n number of balls and the x is the numbers of the black balls then, we do the proportion x by n may be either 1 by 4 or 3 by 4. So, now,. what is the problem? Now, the problem is that we want to estimate this proportion to know that whether blacks are more or the reds are more.

So, let me denote p is the proportion of say black balls then p is either 1 by 4 or 3 by 4. Now, let us see how to frame a statistical inference problem and then an estimation problem from here what we can do we can take a random sample from the urn so. Let a random sample of size say, n is taken from the urn and let us say, x denote the number of black balls in the sample then, you can write actually this distribution of x.

(Refer Slide Time: 44:32)



Probability that x equal to x that will be actually n c x p to the power x 1 minus p to the power n minus x where x can take values 0, 1, 2, n and p can take value 1 by 4; 3 by 4. So, the parameter space is restricted to 2 values 1 by 4 and 3 by 4 and if you want to estimate that then, we have to conclude whether p is equal to 1 by 4 or p is equal to 3 by 4. So, let us work it out for a particular value of n.

So, let us take say n is equal to 3 if I take n is equal to 3 then what are the various probabilities for this on this side I will write down what is the probability of x equal to x and on this side I will consider the values of p p is equal to…so, what is the probability of x is equal to x, when p is equal to say 1 by 4? What is the probability of x equal to x when p is equal to 3 by 4? And, on this side let us take the various values of x x can take value 0, 1, 2 and 3. So, since n is equal to 3 what is the probability that x equal to 0 that will be equal to 1 minus p cube. So, if p is equal to 1 by 4, it is simply 3 by 4 cube. That is, 27 by 64 when p is equal to 3 by 4 1 minus p cube becomes 1 by 4 cube that is 1 by 64.

What is the probability of x equal to 1? That is, 3 c 1 that is 3 p into 1 minus p square. So, when p is equal to 1 by 4, this 3 by 4 square becomes 9 by 64; 9 by 16 and then you have 1 by 4.

So, it becomes again 27 by 64; whereas, this value will become when p is equal to 3 by 4 then this is 3 by 4 and this should become 1 by 4 square. So, this will become 9 by 64 when x equal to 2 you will get 3 p square into 1 minus p. So, for p is equal to 1 by 4 this becomes 9 by 64 and for p is equal to 3 by 4, this becomes 27 by 64 for x equal to 3 this probability simply becomes p cube. So, it is 1 by 64 and 27 by 64.

So, as a layman what will be our observation? Our observation is that when we have taken a random sample of size 3 from the urn and we observe that what this sample contains. So, we note down the number of black balls and the red balls.

So, if we observe that there are no black balls; that means, all the balls are red the probability of that is higher corresponding to p is equal to 1 by 4. That means, as a heuristic thinking, it will mean that there are actually less number of black balls; that means, p is equal to 1 by 4 is the likelihood; likely value or you can say most likely value.

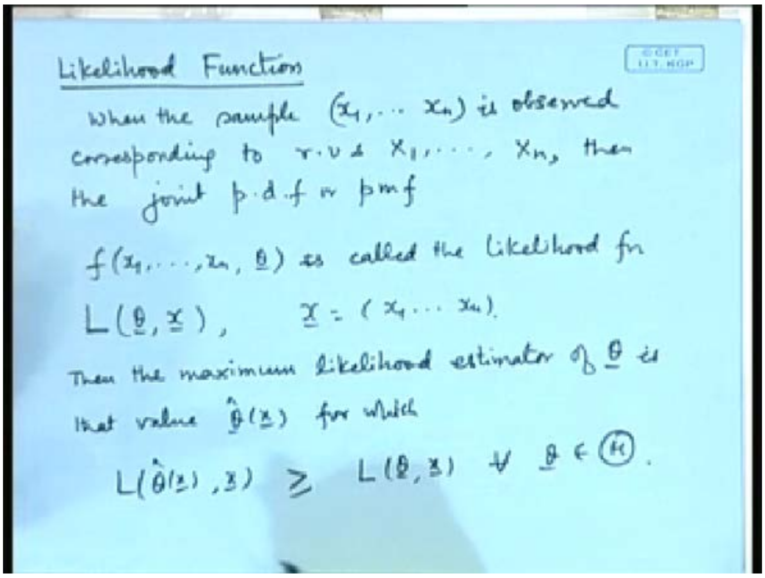
Similarly, when x equal to 1 then, corresponding to p is equal to 1 by 4, you have a higher probability. That means, it shows that it is more likely that the number of black balls is less in the urn because, 3 times you have drawn and you are getting only either none or 1 black ball. Whereas, if you get all the 3 times a red ball a black ball then the probability is higher for p is equal to 3 by 4. That means, it is more likely that the number of black balls is more than the number of red balls and the same observation is true when x is equal to 2 is observed.

So, what we are doing? We are looking at the actual sample observed. So, what is the probability or what is the parameter value for which the likelihood of that sample being observed is highest? That means, we are looking at the maximum likelihood function. So, what is n c x p to the power x 1 minus p to the power n minus x actually this is the probability mass function of the random variable x; here we consider p to be unknown parameter and x is the random variable.

So, x small x is the values of that random variable. Now, I am giving a different interpretation to this. I am calling it the likelihood function of this for different values of p and I am looking at that value of p for which this is maximized. So, you can see that when x equal to 0 or 1 then I am taking p head to be 1 by 4 and when x equal to 2 or 3 then I am taking p is equal to 3 by 4. Why I am taking these values because corresponding to these value the likelihood function that is this value is actually higher.

Now, this is the case which I had explained through 2 points in the parameter space. If I have infinite number of points are more than 2 points then, we simply look at the maximum value over the entire parameter space. That means, the problem is reducing to maximize this function which I call the likelihood function. So, we introduce the term called the likelihood function of a sample.

(Refer Slide Time: 50:46)



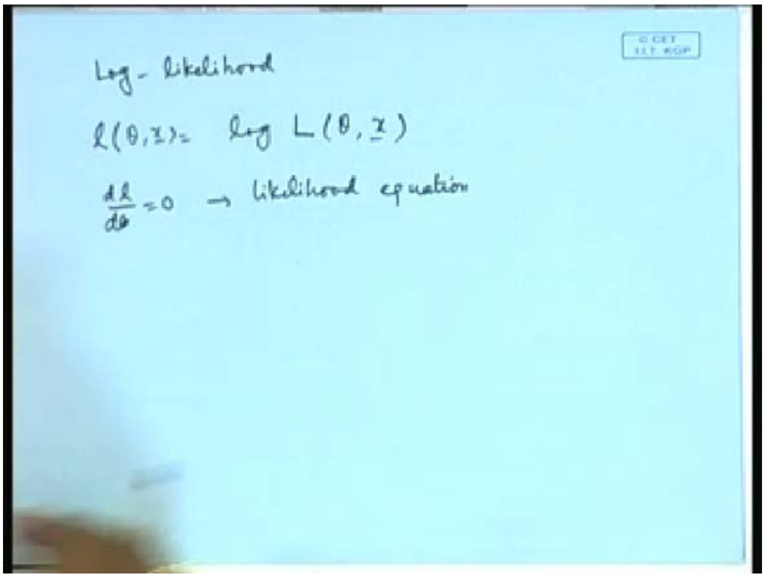
So, we call likelihood function. So, when the sample x 1 x 2 x n is observed corresponding to random variables x 1 x 2 x n then the joint probability density function or probability mass function that is f x 1 x 2 x n and the parameter that we call is called the likelihood function that is l theta x; where x is of course, x 1 x 2 x n then the maximum likelihood estimator of theta is that value theta head x for which l theta head x is greater than or equal to l theta for all theta belonging to theta.

So, basically it becomes a optimization problem; that means, we have to find the points of the likelihood function. Now, for that we may use the methods of calculus or any other analytical methods. In general, we have seen that the parameter space will be an interval and therefore, we have to use the usual methods of calculus such as differentiation putting equal to 0, checking the second derivative or any other analytical method. Sometimes we may be able to tell straight away from the likelihood function or sometimes we have to use this kind of analysis.

Now, if you look at the form of the distribution such as binomial distribution poisson distribution normal distribution they are falling in the form of distributions in the exponential family. So, when the distribution are in the exponential family a major term is coming actually as an exponent e to the power something or something to the power something.

So, if I take log of that then that terms that which are coming as x i terms they are coming in the linear form therefore, in most of the cases it is more convenient to handle log of the likelihood function.

(Refer Slide Time: 54:12)



So, we consider log likelihood that we denote by small l theta x that is equal to log of l theta x. In the beginning we will consider the case when theta could be a scalar and then we also consider the case n theta is a vector.

For example, in the case of normal distribution, etcetera and we will see that what are the solutions and the another point is that sometimes this method of taking the log, etcetera may not be very convenient. In fact, it may not lead to a solution in that case we may have to use direct analytical methods for finding out the solutions of the likelihood function.

A general method is like we differentiate. So, we write d l by d theta is equal to 0; this is called likelihood equation. So, there may be situation where we will get a solution and sometimes a solution may not be there. So, we will see a result where we say that the likelihood function always has a solution and we will work out these things. So, in the next class we will continue the discussion on the maximum likelihood estimators.

We will see that, what are the properties that this maximum likelihood estimator satisfy; then, how do they compare with the other methods like the methods of moment estimators, etcetera.