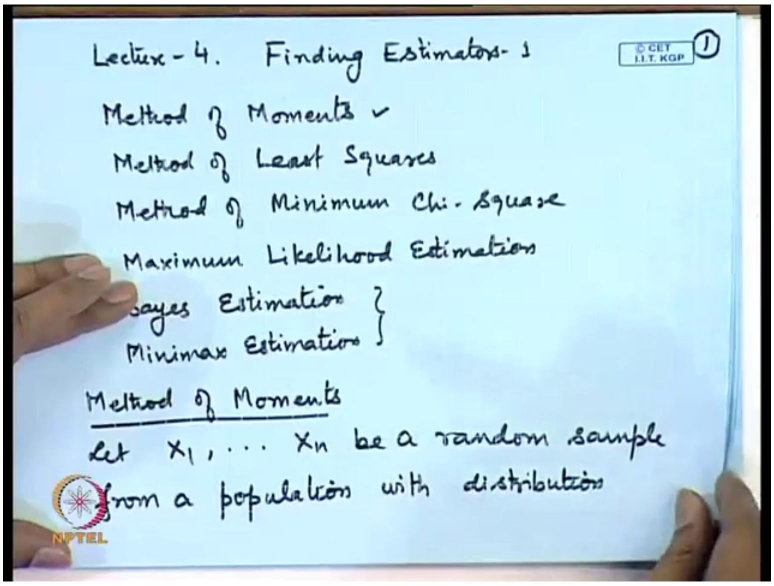
Statistical Inference
Prof. Somesh Kumar
Department of Mathematics
Indian Institute of Technology, Kharagpur

Lecture No. # 04
Finding Estimators – I

In the previous two lectures, I have discussed certain desirable properties for estimators. However, still we are not clear how to derive estimators for various kind of parameters. It may be one thing to say that we can estimate a population mean by a sample mean, a population variance by sample variance or a population range by sample range, but many a times we are having more complicated situations, and more over as we have already seen such as uniform distribution or an exponential distribution that we may have several estimators; may be one is based on the mean, another is based on say order statistics etcetera. So, there must be some procedures or methodology by which we should be able to derive the estimators.

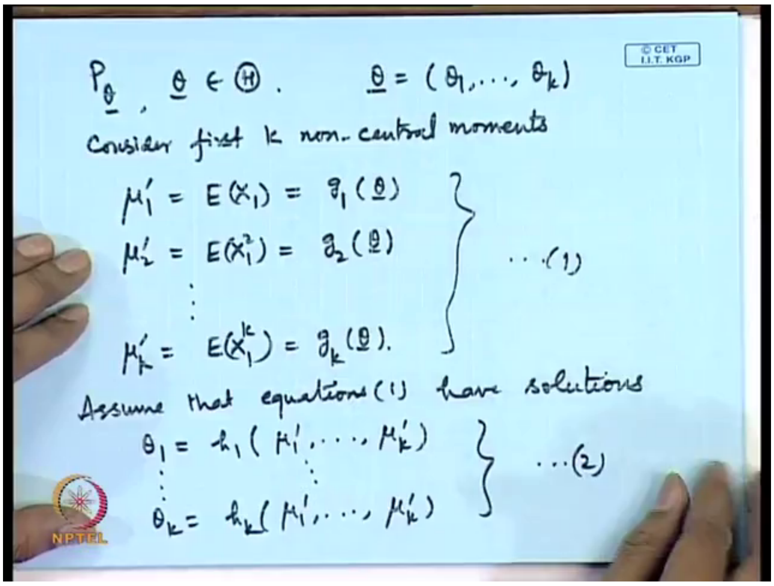
(Refer Slide Time: 01:17)



So, some of the well known methods which are used are the method of moments, the method of least squares, the method of minimum chi-square, then maximum likelihood estimation and then there are certain new procedures such as bayes estimation, minimax estimation. The last two procedures which I have mentioned they are based on decision theoretic concepts and we may not be able to cover much of this in this particular course.

Historically, the method of moments seems to be the oldest one introduced by Karl Pearson. So, let me start from here. The method of moments. Let us consider that we have a random sample X 1, X 2, X n be a random sample from a population with say distribution which is indentified as say p theta, theta belongs to theta.

(Refer Slide Time: 03:11)

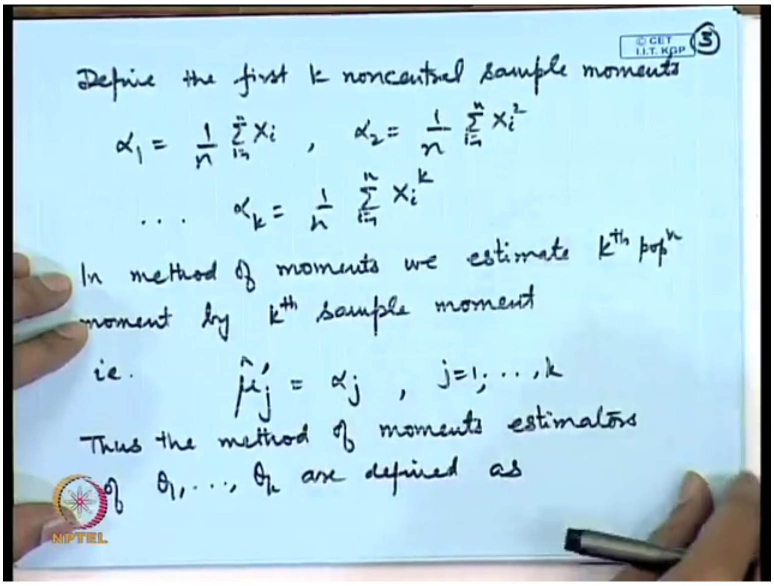


So, here in general I am considering theta to be a vector parameter; that means, theta may have components say theta 1, theta 2, theta k. As we have already talked about for example, if we consider a normal distribution usually, it is characterized by two parameters mu and sigma square. So, in that case theta is mu and sigma square. Similarly, if we consider a poisson distribution it is characterized by a single parameter say lambda. We may have a weibull distribution, we may have a gamma distribution. So, these are variously described by 2, 3 or 5 parameters etcetera. So, in general if we have a k dimensional parameter we consider k moments.

So, let us consider first k non-central moments. That means, we calculate say mu 1 prime which is expectation of say X 1 that is, now all of this moments they are going to be functions of the parameter. So, let us call these functions as say g 1 of theta. Similarly, mu 2 prime that is a second moment of this distribution this will be another function of theta let us call It g 2 and so on. Let us write say mu k prime is equal to expectation of X 1 to the power k that is g k of theta. Now, we assume that these k equations. So, each of this is a function of theta 1 theta, 2 theta, theta k.

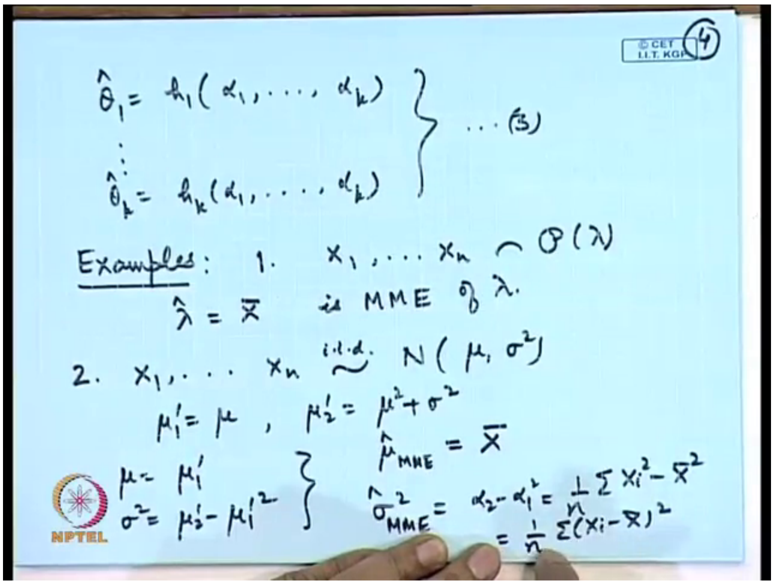
So, we assume that these equations one they have solutions. Assume that the system of equations 1 have solutions. Now, the solutions will be in the form; that means, I am saying theta 1 is h 1 of say mu 1 prime, mu 2 prime, mu k prime and so on. Theta k is h k of mu 1 prime, mu 2 prime, mu k prime. Let us call It 2. In method of moments what we do in place of this mu 1 prime, mu 2 prime, mu k prime which are the first k non-central moments of the population we substitute these by the corresponding sample moments.

(Refer Slide Time: 06:12)



So, let us define say sample moments as define the first k non-central sample moments; that means, let me define say alpha 1 is equal to 1 by n sigma X I, alpha 2 is say 1 by n sigma X i square, i is equal to 1 to n in general. So, alpha k is equal to 1 by n sigma X i to the power k, i is equal to 1 to n. In method of moments we estimate k-eth population moment by k-eth sample moment, that is I am writing that mu j head, mu j prime head is equal to alpha j, for j is equal to 1 to k. So, these values we substitute here thus the method of moments estimators of theta 1, theta 2, theta k are defined as theta 1 head is equal to h 1 of alpha 1, alpha 2, alpha k and so on.

(Refer Slide Time: 08:06)



Theta k head is equal to h k of alpha 1, alpha 2, alpha k. Now, the question may come that If we are having the solutions to these equations, If the solutions to these equations are obtainable in the explicit form then only we can write down this solution for the method of moments. There may be some cases where you may have say 2 parameters or 3 parameters, but 2 or 3 equations may not lead to the solutions. In that case we may take extra moments here. So, let me start with certain examples here the simplest 1 for example, I may consider say X 1, X 2, X n follow a poisson lambda distribution.

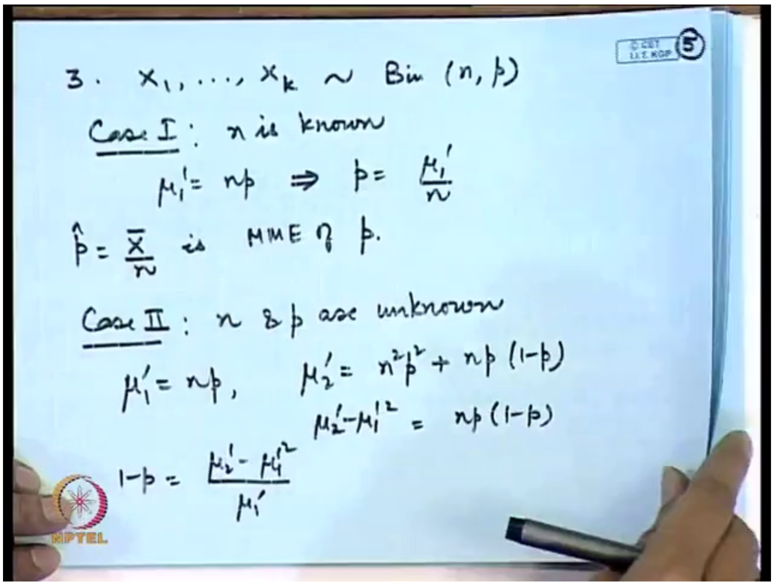
Now, this is a 1 parameter case. So, I need to take up only the first moment. Now, we know that the first of the poisson distribution is lambda and the first sample moment is X bar. So, lambda head is equal to X bar. So, this is the method of moment estimator of lambda. Let us take say X 1, X 2, X n following normal mu sigma square distribution where both mu and sigma square are parameters here, unknown parameters. Let us take here mu 1 prime in normal distribution the mean is mu. mu 2 prime is equal to the second moment is mu square plus sigma square.

So, If we solve this we get mu is equal to mu 1 prime and sigma square is equal to mu 2 prime minus mu 1 prime square. So, this is the system which is equivalent to this system that theta is are written in terms of the mu i primes. So, now, we substitute alpha 1 for mu 1 prime and alpha 2 for mu 2 prime. So, the method of moments estimator for mu head mu let me call

It M M E that is denoting the method of moments estimator of mu. It is simply X bar that is alpha 1 and for sigma square it is equal to alpha 2 minus alpha 1 square.

Let us see what is the value of this. It is 1 by n sigma X i square minus X bar square which I can write as 1 by n sigma X i minus X bar whole square. Notice, here in the previous classes when I was discussing unbiased estimation I derived the unbiased estimator of sigma square as s square that was 1 by n minus 1 sigma X i minus X bar square. So, there is a clear cut case of comparison between the method of moments estimator and an unbiased estimator in this particular problem.

(Refer Slide Time: 11:45)

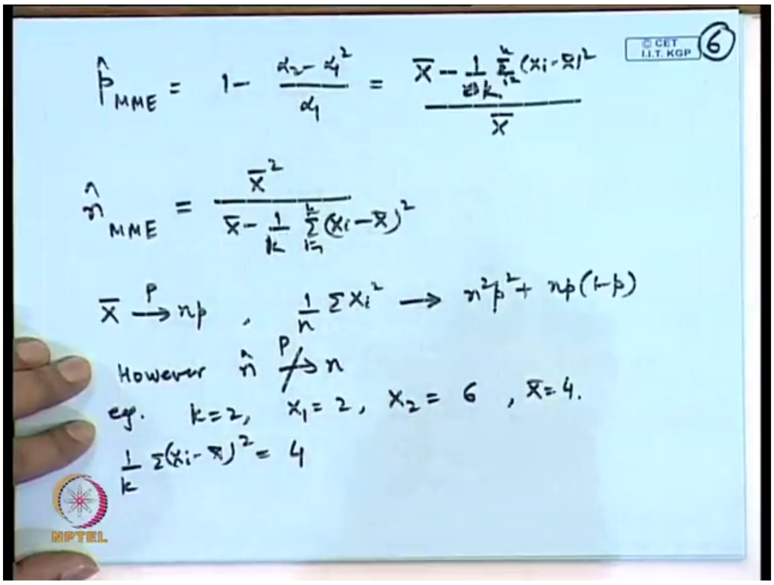


Let us take say X 1, X 2, X k following a binomial distribution with parameters n and p. Quite typical situations in binomial distribution deal with the situations where n is known. So, If we have n is known then parameter is p here and If I consider the first moment here, first moment of the binomial distribution is n p. So, this is to be estimated by alpha 1; that means, X bar is an estimate of n p. So, If we want to write down the solution p is equal to mu 1 prime by n. So, we get here p head is equal to X bar by n.

So, this is method of moments estimator of p. Since, here only 1 parameter was there we considered only 1 equation. Now, let us take the more general case where n and p both are unknown. When both are unknown then we will have to take up the first 2 moments. So, mu 1 prime is equal to n p and mu 2 prime is equal to n square, p square plus n p into 1 minus p. In the binomial distribution the second moment is equal to this value here. Now, we can solve this equation actually if we take up say mu 2 prime minus mu 1 prime square, I get n p into 1 minus p. So, If I divide this equation by this, I get 1 minus p is equal to mu 2 prime minus mu 1 prime square by mu 1 prime.

So, the solution for p has come and If I substitute that value of p here I get the value of n. So, I get p is equal to 1 minus mu 2 prime minus mu 1 prime square divided by mu 1 prime and n is equal to mu 1 prime by p. So, now, by substituting alpha 1 and alpha 2 for mu 1 prime and mu 2 prime I get the method of moments estimator for n and p.

(Refer Slide Time: 14:35)



So, let us look at this value here. We get p head M M E as 1 minus alpha 2 minus alpha 1 square by alpha 1, here alpha 1 is X bar and alpha 1 square is 1 by n sigma X i square. So, If you substitute those values this turns out to be X bar minus 1 by n. In this case It is 1 by k sigma X i minus X bar square i is equal to 1 to k divided by X bar and n is estimated by X bar square divided by X bar minus 1 by k sigma X i minus X bar square i is equal to 1 to k. Notice here, when n was known then the estimate for p was simply X bar by n whereas, now you can see it has changed quite drastically here.

In the context of these exercises let us also see some other properties which we had earlier for example, unbiasedness now you see in the poisson distribution case expectation of X bar is equal to lambda. So, the method of moments estimator is actually unbiased. It will also be consistent if we apply the weak law of large numbers, as we have already seen that if the first moment exist the sample mean is always a consistent estimator for the population mean. So, in this case M M E is unbiased and consistent for lambda. Let us take up the second 1. Normal distribution example. Here, If we are looking at X bar then X bar is unbiased for mu and also it is consistent.

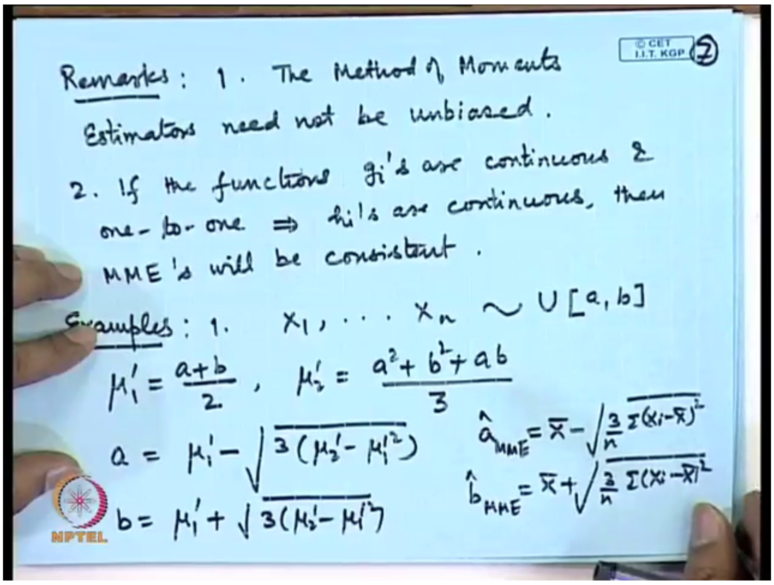
However, If we look at the estimator for sigma square you can notice here that It is not unbiased; however, It will remain consistent because it is actually n minus 1 by n s square. So, since s square was consistent and n minus 1 by n converges to 1 this also converges to 1 therefore, here you are having that mu head M M E is unbiased and consistent; however, sigma head square is biased, but consistent. So, this brings us to important property that the unbiased, the method of moments estimators need not always be unbiased.

Now, in these 2 exercises they are consistent. So, again the question arises whether they will be consistent always? Let us take up the next case. Here X bar by n this is unbiased as well as consistent. So, p head is unbiased and consistent. Let us take the second case when both the parameters were unknown. Here if you see since X bar was unbiased for p. So, this cannot be unbiased because this is quite different. If we take up the limits here. So, we are having X bar converges to n p in probability. We are having 1 by n sigma X i square that is alpha 2, this converges to the second moment that is n square p square plus n p into 1 minus p.

So, both of these are convergent in probability. Now, let us look at this quantity in the denominator you are having X bar minus 1 by k this quantity. So, If you look at the limit here this is going to n p and this is going to the variance term that is n p minus into 1 minus p here. So, this does not converge actually because for convergence in probability we have established 1 property that is the invariance property, but the invariance property is only for the continuous functions. Here this is not a continuous function because the denominator may become 0. However, n head does not converge to n in probability.

You may take 1 illustration here. You may take say k equal to 2. Let me take observation say X 1 is equal to 2, X 2 is equal to say 6. So, X bar is equal to 4 and If I calculate 1 by k sigma X i minus X bar square that is also equal to 4. So, this denominator actually becomes 0 and the probability of this is positive because I am taking it to be actual observations here. Therefore, we conclude here that the method of moments estimator need not be consistent also. So, this is a method sometimes the properties of unbiasedness consistency holds, sometimes they do not hold.

(Refer Slide Time: 20:22)



So, let me give it comments here. The method of moments estimators need not be unbiased. If the functions say g i's are continuous and one-to-one. In that case inverse functions will exist and they will also be continuous, h i's are continuous, then M M E's will be consistent. That means, they are not always consistent, but under certain conditions they will be consistent. Let us take up another case. Let us take say X 1, X 2, X n a random sample from a uniform distribution say on the interval a to b.

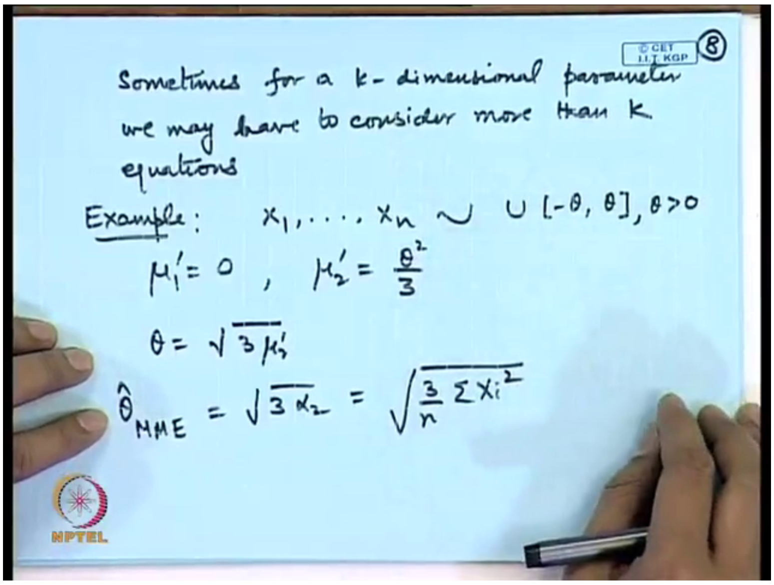
Again you may have different conditions for example, It may be one parameter situation; that means, a may be known or b may be known or both may be unknown.

So, I will consider the case when both a and b are unknown. So that means, we have 2 parameters. So, we write down the first two moments the mean is a plus b by 2 and the second moment is a square plus b square plus a b by 3. So, we need to solve this. If we solve this, we get a is equal to mu 1 prime minus square root 3 mu 2 prime minus mu 1 prime square and b as mu 1 prime plus square root 3 times mu 2 prime minus mu 1 prime square. So, these are basically 2 equations in 2 unknowns and they are non-linear equations. However, 1 can solve it by making use of certain elementary relations such as a minus b is equal to square root of a plus b whole square minus 4 a b.

So, I am assuming since the interval is from a to b. So, I am taking a to be less than b. So, I am taking minus value here and plus value here. So, If you substitute the alpha 1 and alpha 2 here then the method of moments estimators turn out to be X bar minus square root 3 by n sigma X i minus X bar square and b head M M E as X bar plus square root 3 by n sigma X i minus X bar square. In this particular case we may see that these estimators may be consistent. Now the reason for that is that alpha 1 is consistent for mu 1 prime and alpha 2 is consistent for this. And this is continuous function here and that is the inverse functions that we have considered they are continuous therefore, this will be consistent. However, they are not unbiased.

I mentioned that If I am having a k-dimensional parameter then we may usually consider k equations. So, why usually because sometimes the k equations may not give us the desirable result.

(Refer Slide Time: 24:50)

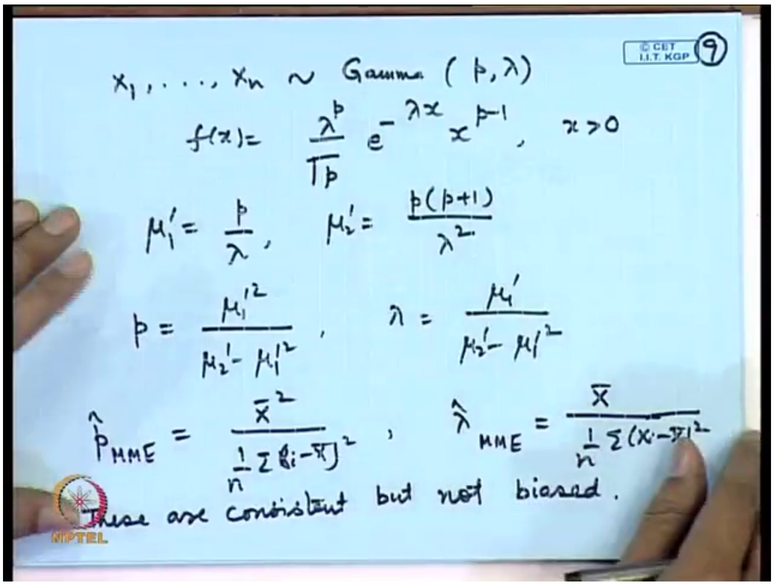


Let us take a very simple example. Sometimes for a k-dimensional parameter we may have to consider more than k equations. So, let us take an example for this situation say X 1, X 2, X n follow a uniform distribution on the interval say minus theta to theta where theta is a positive number. Now, in this case let us see the first moment mu 1 prime is actually 0. So, this does not give any information about theta and therefore, how to estimate. So, a natural thing is to consider the second moment here. The second moment here turns out to be, if we substitute in the previous formula of this you will get theta square by 3 because a is minus theta and b is plus theta.

See, If I get substitute it here I will get theta square plus theta square minus theta square by 3. So, that is theta square by 3. So, a solution to this is equal to square root of 3 mu 2 prime.

So, I may take the method of moments estimator as square root 3, alpha 2 that is square root 3 by n sigma X i square. So, here since the first moment did not give us any solution for theta I am using second moment.

(Refer Slide Time: 26:58)

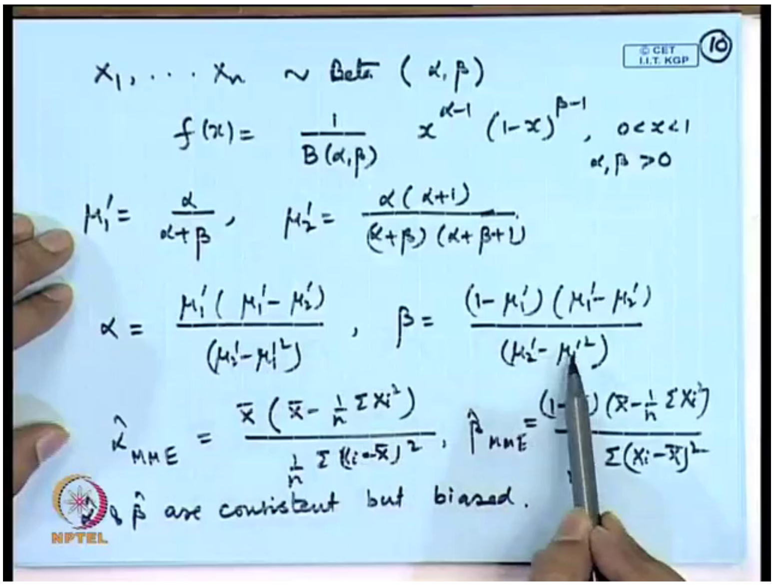


I end up this section with the two more examples, one for a two parameter gamma distribution and one for a two parameter beta distribution. Let us take say X 1, X 2, X n following a gamma distribution with parameter say p and lambda. So, here lambda is corresponding to the rate of the corresponding poisson process. That means, I am taking the density function is equal to lambda to the power p by gamma p, e to the power minus lambda x, x to the power p minus 1, X is greater than 0. Now, in this distribution the first moment is p by lambda and the second moment is equal to p into p plus 1 by lambda square.

So, quite easily we can solve this. The solution is in the form p is equal to mu 1 prime square by mu 2 prime minus mu 1 prime square and lambda is equal to mu 1 prime by mu 2 prime minus mu 1 prime square. So, the method of moments estimators are easily obtained as X bar square divided by 1 by n sigma X i minus X bar square and lambda head M M E is equal to X bar divided by 1 by n sigma X i minus X bar square. One can easily check that these are consistent, but not unbiased.

So, generally the method of moments estimators will be consistent, but usually they will not be unbiased. In fact, the typical situations where they will be unbiased is only when you are having the first moment only. So, in that case the sample mean is unbiased for the population mean and therefore, unbiasedness will be satisfied.

(Refer Slide Time: 29:26)



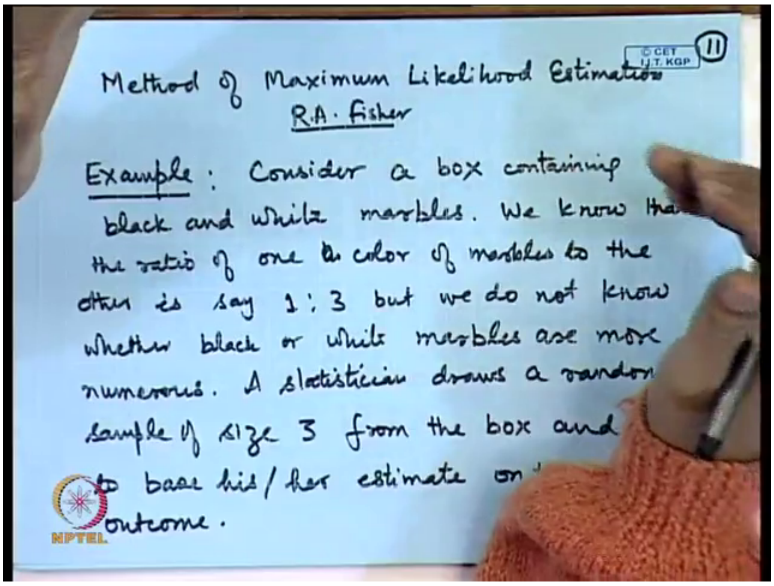
Similarly, let us take up say beta distribution, say with parameters alpha and beta. That means, I am considering the density function as equal to 1 by beta alpha beta, X to the power alpha minus 1, 1 minus X to the power beta minus 1 where x is between 0 and 1 and alpha and beta both are unknown positive parameters. The first 2 moments of a beta distribution are alpha by alpha plus beta and alpha into alpha plus 1 divided by alpha plus beta into alpha plus beta plus 1. So, we can solve these equations by firstly dividing and then subtracting by 1 etcetera.

So, the form of the solution is that 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, mu 1 prime minus mu 2 prime divided by mu 2 prime minus mu 1 prime square. So, If we substitute mu 1 prime as alpha 1 and mu 2 prime as alpha 2 we get the method of moments estimators as X bar into well this is X bar minus 1 by n sigma X i square divided by 1 by n sigma X i minus X bar square and similarly beta head M M E is equal to 1 minus X bar into the same term here that is X bar minus 1 by n sigma X i square divided by 1 by n sigma X i minus X bar square. In this case also If we look at this thing these estimators are consistent, but not unbiased.

So, this alpha head and beta head M M E's they are consistent, but bi[ased]- consistency is obvious because these things have turned out to be a to be continuous functions. In fact, the denominator is always positive because mu 2 prime minus mu 1 prime square is actually the population variance and if you look at these functions. So, from here because of the basic weak law of large numbers X bar or alpha 1 converges to mu 1 prime in probability and alpha 2 prime converge to mu 2 prime in probability. So, If you substitute these things here these things also remain consistent.

However, they are not unbiased. In fact, later on when we discuss the theory of finding out unbiased estimators we will see what will be actually the corresponding unbiased estimators. After this method of moments in the historical order the method of maximum likelihood estimation is considered to be the most popular and the currently most commonly used method of estimation this was developed by R A Fisher

(Refer Slide Time: 33:20)

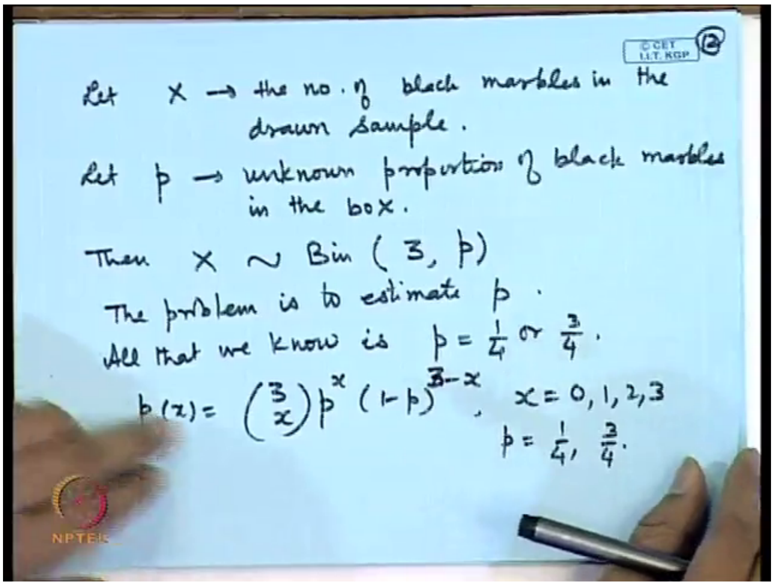


The method of maximum likelihood estimation. So, this method was developed by RA Fisher. Here, he considers that all the problems in the inference should be based on the likelihood function. So, what is a likelihood function? He considered that when we are having a an experiment and we are formulating a random variable then the probability distribution of the random variable is dependent upon the unknown feature of the parameter. So, usually we consider f x, but actually f x theta, that means the density function or the probability mass function is a function of variable as well as the parameter.

So, the idea that the Fisher gave that we should consider those values of theta for which this is maximum. So, that is why it is called the method of maximum likelihood. That means, the values which are most probable. So, I will introduce an example here. Consider. So, we have consider a box containing say black and white marbles. So, some number is there. We know that the ratio of 1 color of marble to the other is say 1 is to 3, but we do not know whether black or white marbles are more numerous. That means, the incidence of black marbles is more. That means, they are 3 times the white one's or white is are 3 times the black one's, that is not known.

So, we only know we have a partial information that 1 of them is 3 times the other. The number of marbles. So, a statistician decides to take a random sample from the box and based on that random sample he will give a inference. That means, he will give an estimator for the ratio. So, we can consider this problem in this following fashion. A statistician draws a random sample of size 3 from the box and decides to base his estimate on the reported outcome. So, If you are drawing 3 marbles. Now, these 3 marbles may have some black or white marbles among themselves.

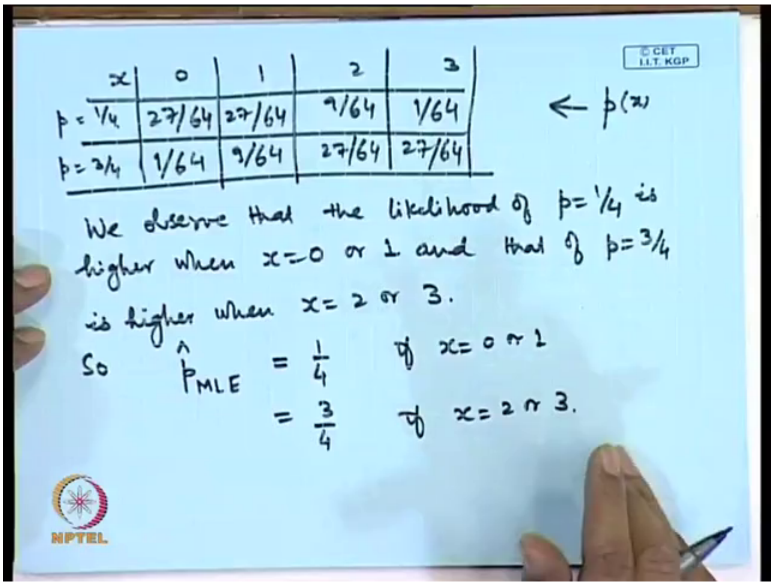
(Refer Slide Time: 37:38)



So, let us denote. Let x denote the number of black marbles in the drawn sample. Let us consider say p as the unknown proportion of black marbles in the box. Then we can easily describe by a probability model here. For example, x is a, now at each draw you will have a black marble or a white marble. So, you can consider it as a Bernoullian trial where If you draw a black marble you will consider it as a success and If you draw a white marble you will consider it as a failure, the probability of success is p.

So, in the total 3 trials x is the number of successes here. So, this can be easily described by a Binomial model; that means, I am saying x follows a Binomial distribution with parameters 3 and p. And the problem is, the problem is to estimate p. All that we know is that p is 1 by 4 or 3 by 4. So, partial information is available, but we do not know exactly. So, on the basis of my reported x, that means, on the basis of my sample we should take a decision whether p should be 1 by 4 or 3 by 4. So, we may write down the probability mass function of x as 3 c x, p to the power x, 1 minus p to the power n mi[nus]-, n is 3, 3 minus x, x is equal to 0, 1, 2, 3 and p is either 1 by 4 or 3 by 4. Let us write down the probabilities of various possibilities based on this probability mass function.

(Refer Slide Time: 40:21)



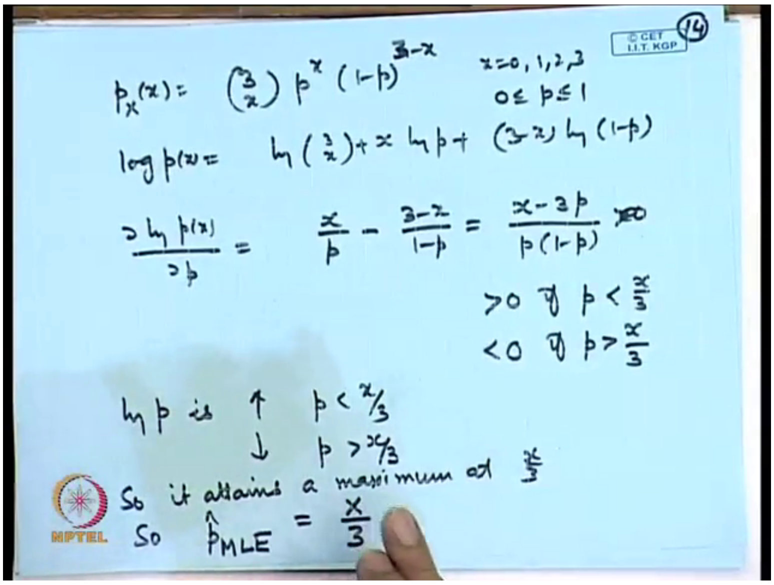
So, let me write it in the form of a table. x can take values 0, 1, 2 and 3 and p can take value 1 by 4 and 3 by 4. On the basis of this I tabulate in this table the value of p x. What is the probability that x is equal to 0 when p is equal to 1 by 4? So, in this 1 If we put p equal to 0 I get 1 minus p q. So, 1 minus p q, If p is equal to 1 by 4 it becomes 3 by 4 cube that is 27 by 64. When p is equal to 3 by 4 this becomes 1 by 64. Similarly, let us consider x is equal to 1. If I take x is equal to 1 this is 3 p into 1 minus p square. So, again for p is equal to 1 by 4 this value turns out to be 27 by 64

Whereas at for 3 by 4 it turns out to be 9 by 64. In a similar way I can consider other values for p equal to 1 by 4 and x is equal to 2 the value is 9 by 64 here and for p is equal to 3 by 4 this value is 27 by 64. The last values are 1 by 64 and 27 by 64. Now, you see If x is equal to 0 then p is equal to 1 by 4 gives a higher likelihood. Similarly, If x equal to 1 p is equal to 1 by 4 gives a higher likelihood. Whereas, at If x equal 2 or 3 p is equal to 3 by 4 have a higher likelihood.

So, we based on this discussion we may write our estimators as we observe that the. So, this p x now I am calling as likelihood. We obse rve that the likelihood of p is equal to 1 by 4 is higher when x is equal to 0 or 1 and that of p is equal to 3 by 4 is higher when x equal to 2 or 3. So, the maximum likelihood estimator for p can be written as 1 by 4 if x is equal to 0 or 1 and it is equal to 3 by 4 if x is equal to 2 or 3. Now, this is the mathematical part of i t.

Let us look at physical interpretation. If in a random sample of size 3, we observe that there are no black balls then we have a feeling that there are less number of black balls. Similarly, If we observe all the 3 are black then we should have a feeling that there are more black balls and therefore, 3 by 4 should be the estimate and similarly interpretation exist for 1 and 2 also. So, therefore, this method of maximum likelihood estimator is actually a you can say an intuitively appealing procedure because what It says that now based on my sample I have already drawn the sample and let me use that information for coming to a conclusion about the value of the parameter.

(Refer Slide Time: 44:28)



That means in general If we did not have this information that p is equal to 1 by 4 or 3 by 4 then the problem will transform like this. That means, I am having say my function p x as say 3 c x p to the power x 1 minus p to the power n minus x and now here p is any number between 0 to 1. So, naturally now the problem is more complicated. I cannot make a table of this nature for all values of p because there are uncountably many values of p in the interval 0 to 1. However, I can use the usual methods of analysis or calculus to find out the value of p which will maximize this function, this is 3 minus x here.

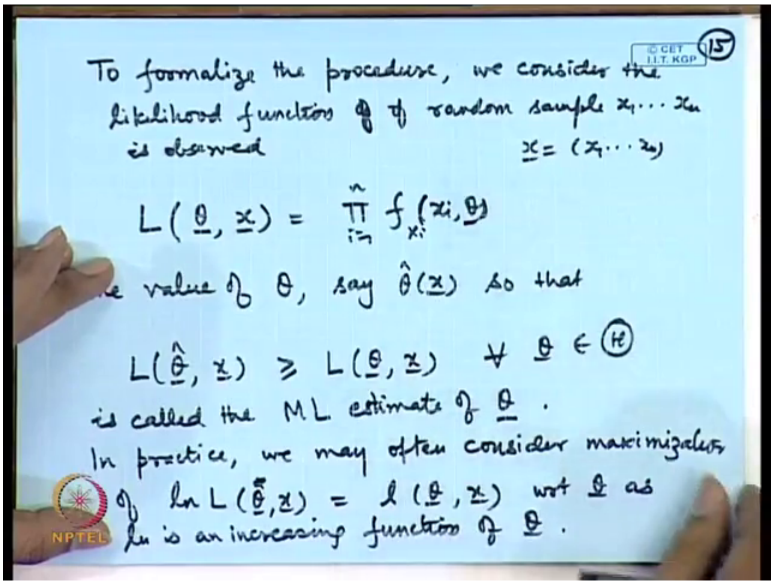
So, we may consider if you use the simple method of calculus you have to look at the derivative or the behavior of the function as It is increasing or decreasing. So, for example, we may consider derivatives. Now, a simple method could be to take log of p that becomes log of 3 x, x log p, 3 minus x log 1 minus p. Now, this maximization of this with respect to p same as maximization of this because log is in increasing function a one-to-one increasing function.

So, I can consider derivative of this with respect to p which gives us x by p minus 3 minus x by 1 minus p which after adjustment I can see it as x minus 3 p by p into 1 minus p. Now, you look at this. This function is positive if p is less than x by 3. It is negative if p is greater than x by 3. So, If we look at this derivative with respect to p as a function of p then I am having the behavior as positive or negative in certain region. As a consequence I know the behavior of log p; that means, log p is increasing up to p less than x by 3 and It is decreasing for p greater than x by 3.

So, there is a p attained at x by 3. So, It attains a maximum at x by 3. So, we can consider the maximum likelihood estimator of p as x by 3 which is actually the sample proportion, because if I have conducted the experiment 3 times and x is a number of successes in 3 trials then x by 3 is a sample proportion and I am using this sample proportion as an unbiased estimate as a maximum likelihood estimate of the population proportion p. So, nothing you can say unreasonable here and in fact, this seems to be a quite general method now.

So, in place of 3 if I had n here then the only changes would have become here that I would have got x by n, which actually matches with your method of moments estimator. It was also unbiased and consistent estimator. So, to formalize this method of maximum likelihood now we can say.

(Refer Slide Time: 48:26)

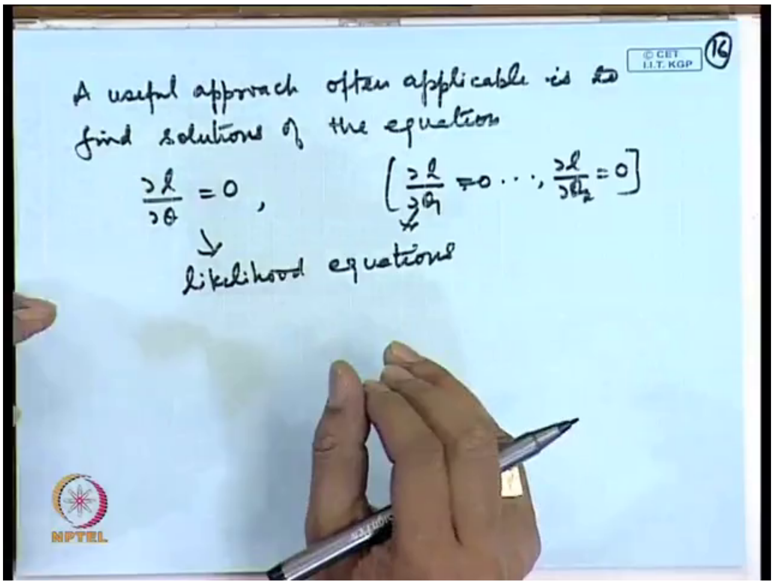


So, to formalize the procedure we consider the likelihood function of a random sample. So, likelihood sample basically it is, I am considering it is a function of the parameter. So, If random sample say X 1, X 2, X n is observed then the likelihood function which is actually a function of theta and of course, x also where x is actually X 1, X 2, X n It is actually the joint density function of X 1, X 2, X n actually.

So, it is evaluated at the points X 1, X 2, X n which are actually the observed values. So, I am considering it as a actually function of theta. So, the value of theta say theta head x. So, that L theta head x is greater than or equal to L theta x for all theta is called the maximum likelihood estimate of theta. Now, in practice one may take log depending upon what type of function you are having. So, for example, in the first case we did not take log, we actually wrote the values here because only 2 possibilities were there.

Whereas in the second case it is a continuous function and differentiable function. So, we took log and then differentiated. So, there is no hard and fast rule for this. However, this procedure of taking logarithm is also considered quite standard in many practical problems it is applicable. So, in practice we may often consider maximization of log of let me call it small L of theta x with respect to theta as l n is a an increasing function of theta.

(Refer Slide Time: 51:35)



A useful approach often applicable is to find solutions of the equation del l by del theta is equal to 0 If It is a scalar parameter or If I have a vector parameter then I may have to consider several equations. So, these are called likelihood equations. Again the question arises whether we can always solve it, whether we can always solve it explicitly or sometimes implicit solutions will be there or sometimes solutions will not exist. So, these and the other properties of the maximum likelihood estimators we will take up in the forth coming class. So, today's class I end up at this point.