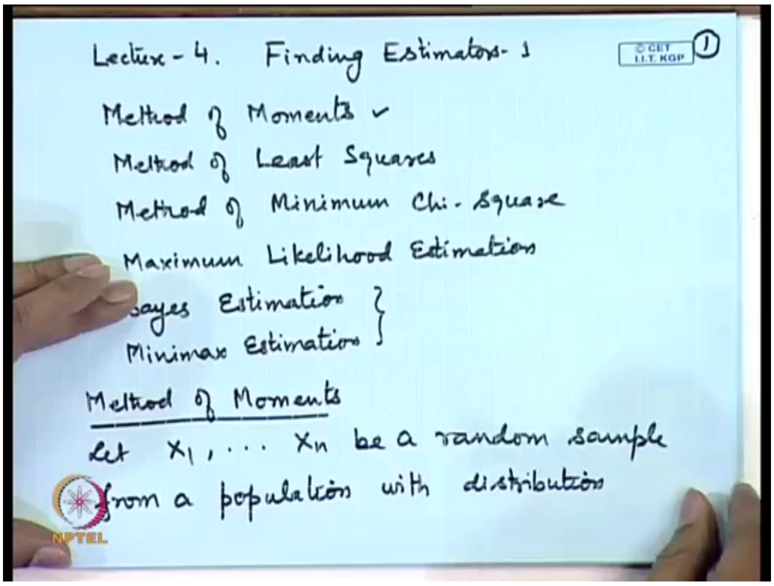
Statistical Inference
Prof. Somesh Kumar
Department of Mathematics
Indian Institute of Technology, Kharagpur

Lecture - 07
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 moreover as we have already seen such as uniform distribution or an exponential distribution that we may have several estimators, maybe 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: 01:19)

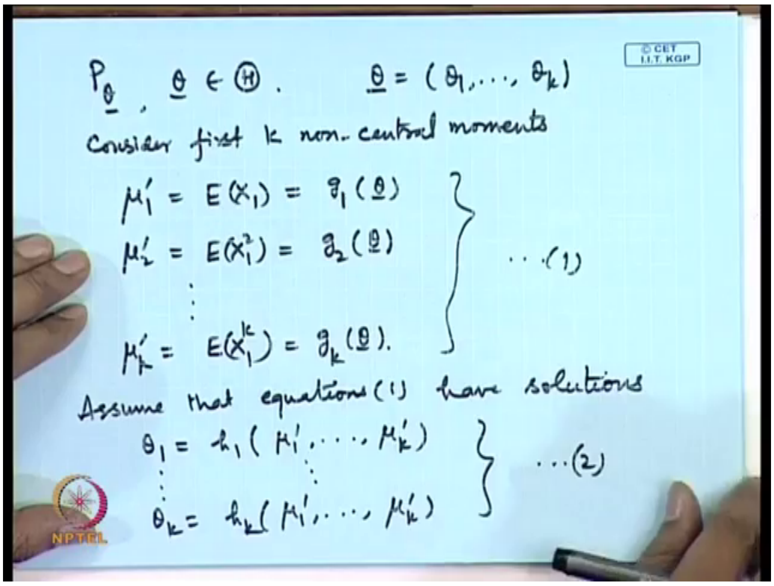


So, some of the well known methods which are used are the method of moments, the method of least squares, the method of minimum chi square, then maximum likelihood estimation and then there are certain new procedure such as Bayes estimation, minimax estimation. The last two procedures which I have mentioned, 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 identified as say P theta, theta belongs to theta.

(Refer Slide Time: 03:13)



So, here in general I am considering theta to be a vector parameter that means, theta may have component say theta 1, theta 2, theta k. As we have already talked about for example, if you consider a normal distribution, usually it is characterized by two parameters mu and sigma square. So, in that case theta is mu sigma square.

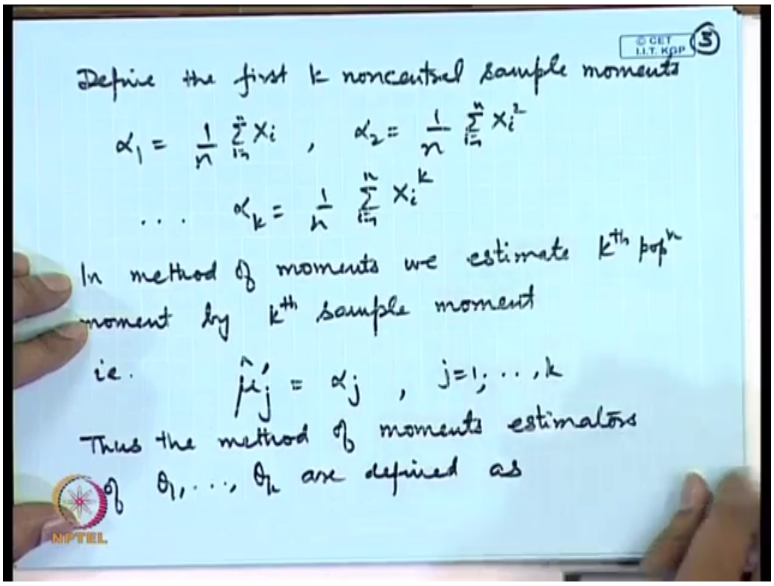
Similarly, if we consider a Poisson distribution, it is characterized by a single parameter say lambda. We may have a Weibull distribution, we may have a gamma distribution. So, these are variously a described by 2, 3 or 5 parameters, etcetera. So, in general if we have k dimension parameter, we consider 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 these moments, they are going to be functions of the parameter. So, let us call this function as a g 1 of theta. Similarly, mu 2 prime that is the second moment of the 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 this k equations. So, each of this is a function of theta 1, theta 2, theta k. So, we assume that this equations 1, 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-center moments of the population, we substitute these by the corresponding sample moments.

(Refer Slide Time: 06:14)

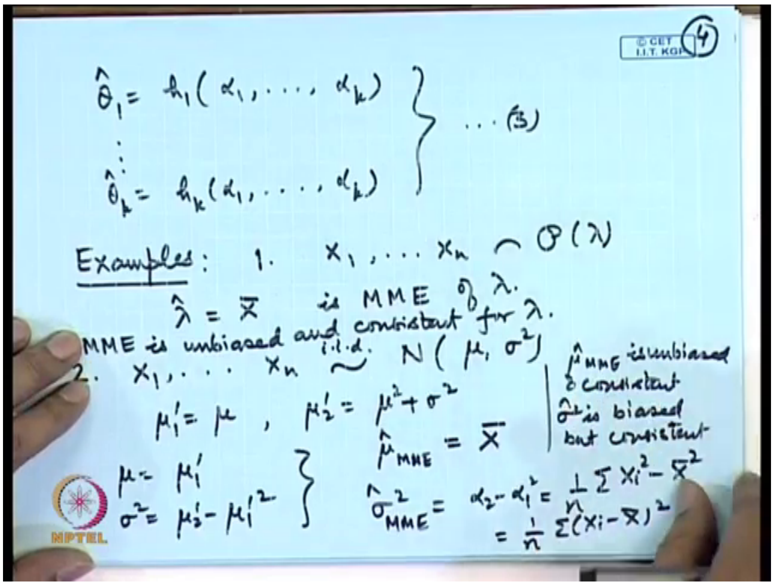


So, let us defines say sample moments as define the first k non-central sample moments that means, let me define say alpha 1 is equal to 1 by n sigma X i, alpha 2 is say 1 by n sigma X i square, i is equal to 1 to n. In general, so alpha k is equal to 1 by n sigma X i to the power k, i is equal to 1 to n.

In method of moments, we estimate kth population moment by kth sample moment that is I am writing that mu j hat mu j prime hat 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 hat is equal to h 1 of alpha 1, alpha 2, alpha k and so on. Theta k hat is equal to h k of alpha 1, alpha 2, alpha k.

(Refer Slide Time: 08:08)



Now, the question may come that if we are having the solutions to this equations; if the solutions to this equations are obtainable in the explicit form, then only we can write down the solution for the method of moments. There may be some cases, where you may have say two parameters or three parameters, but two or three equations may not lead to the solutions in that case we may take extra moments here.

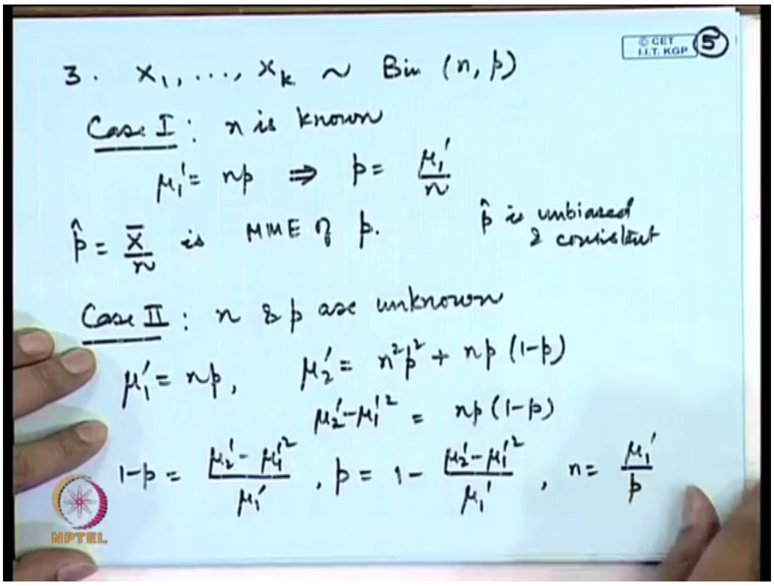
So, let me start with certain examples here, the simplest one for example I may consider say X 1, X 2, X n follow a Poisson lambda distribution. Now, this is the one parameter case, so I need to take up only the first moment. Now, we know that the first moment of the Poisson distribution is lambda and the first sample moment is X bar. So, lambda hat as 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 are parameters here, unknown parameters. Let us take here, mu 1 prime in normal distribution the means is mu; mu 2 prime is equal to the second moment is mu square plus sigma square.

So, if you solve this we get mu is equal to mu 1 prime and sigma square is equal to mu 2 prime minus mu 1 prime a square. So, this is the system which is equivalent to this system that theta i 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 estimators for mu hat mu let me call it MME, 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 a 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 minus X bar square. So, there is a clear cut case of comparison between the method of moment's estimator and an unbiased estimator, in this particular problem.

(Refer Slide Time: 11:47)



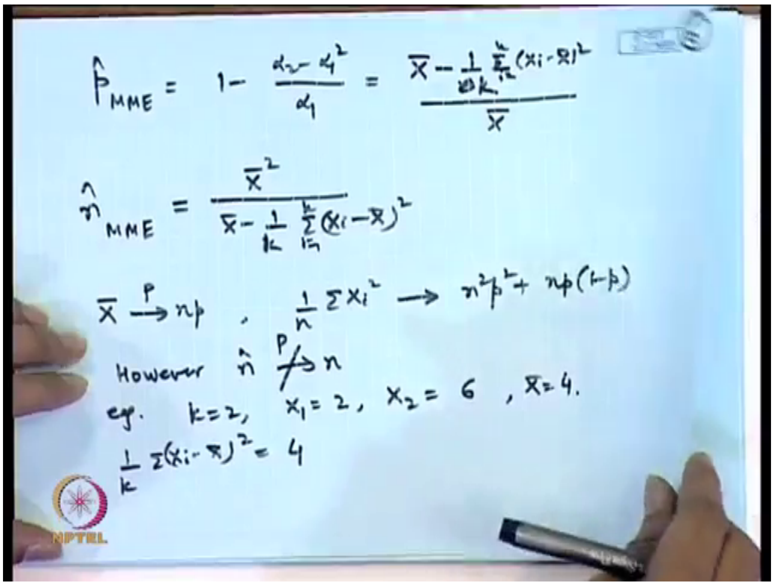
Let us take, say X 1, X 2, X k following a binomial distribution with parameters n and p. Quite difficult situations in binomial distribution deal with the situations, where n is known. So, if we have n is known, then parameter is p here. And if I considered the first moment here, first moment of the 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 you want write down the solution p is equal to mu 1 prime by n. So, we get here p hat is equal to X bar by n, so this is method of moments estimator of p. Since here, only one parameter was there be considered only one 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 two 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 as 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 moment's estimator for n and p.

(Refer Slide Time: 14:37)



So, let us look at this value here. We get p hat MME 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 a 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 square.

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 change quite drastically here.

In the context of these excises, let us also see some other properties which we had earlier for example, unbiasness. Now, you see in the Poisson distribution case expectation of X bar is equal to lambda. So, the method of moment's estimator is actually unbiased. It will also be consistent if we apply the weak log 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 MME is unbiased and consistent for lambda. Let us take up the second one; normal distribution is example, here if we are looking at X bar, when X bar is unbiased for mu and also it is consistent. However, if you look at the estimator for sigma i square, you can notice here that it is not unbiased; however, it will remain consistent, because it is actually n minus 1 by n X square. So, since s square was consistent and n minus 1 by n converges to 1, this also converges this to 1.

Therefore, here you are having that mu hat MME is unbiased and consistent, however sigma hat square is biased, but consistent. So, this brings us to important property that the method of moments estimators need not always be unbiased. Now, in these two excises they are consistent.

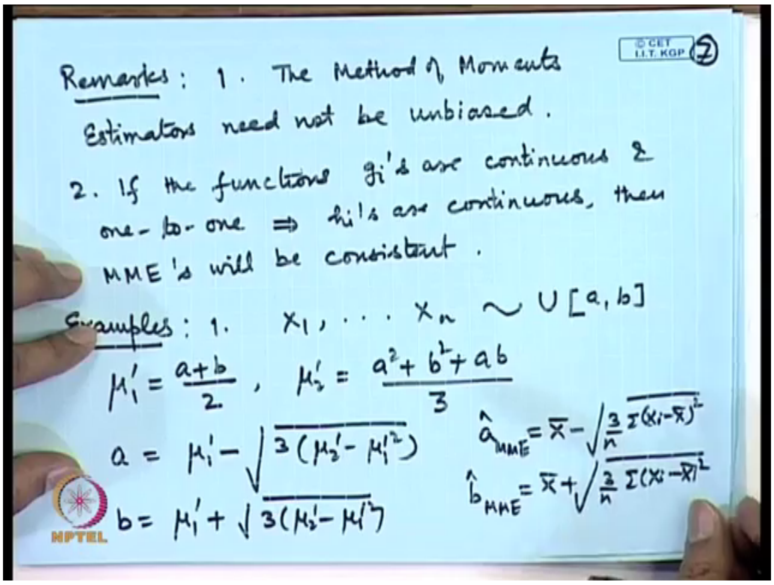
So, again the question arises whether they will be consistent always, let us take up the next case. Here, X bar by n this is unbiased as well as consistent so, p hat is unbiased and consistent. Let us take a second case when both the parameters, where unknown. Here if you see, since X bar was unbiased for p; so this cannot be unbiased, because this is quite different.

If you 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 going to n p and this going to the variance term that is n p into 1 minus p here.

So, this does not converge actually, because for convergence in probability we are established one property that is the invariance property, but the invariance properties only for the continues functions. Here this not a continues functions, because the denominator may become 0 however, n hat does not converge to n in probability. We may take one illustration here, you may take say k equal to 2; let me take observation say X 1 is equal 2, X 2 is equal say 6. So, X bar is equal to 4 and if I calculate one 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 the method. Sometimes the properties of unbiasedness and consistency hold, sometimes they do not hold.

(Refer Slide Time: 20:23)



So, let me give it comments here. The method of moments estimators need not be unbiased. If the functions say g i's are continues and one-to-one; in that case inverse functions will exist and they will be continues h is are continues, then MME's will be consistent that means, you are not always consistent, but under certain conditions they will be consistent.

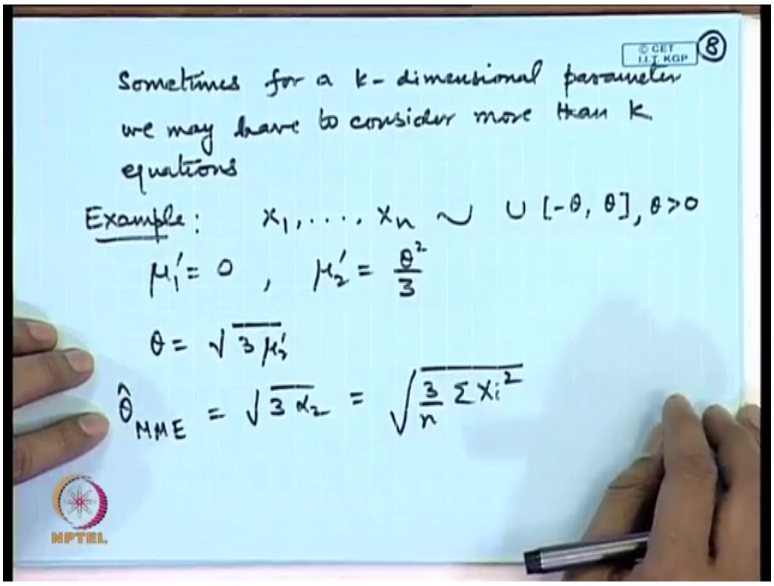
Let us take up another case let us take say X 1, X 2, X n a random sample form a uniform distribution say I am the interval a to b. Again you may different conditions for example, it maybe one parameter situation; that means, a may be known, or b may be known, or both maybe unknown. So, I will consider the case when both a and b are unknown, so that means, we have two 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 you 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 the basically equations in two unknowns and they are non-linear equations. However, one can solve it by making use of certain elementary relation such as a minus b is equal to a 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 we substitute the alpha 1 and alpha 2 here, then the method of moments estimator turn out to be X bar minus a square root 3 by n sigma X i minus X bar square. And b hat MME as X bar plus a 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.

(Refer Slide Time: 24:52)

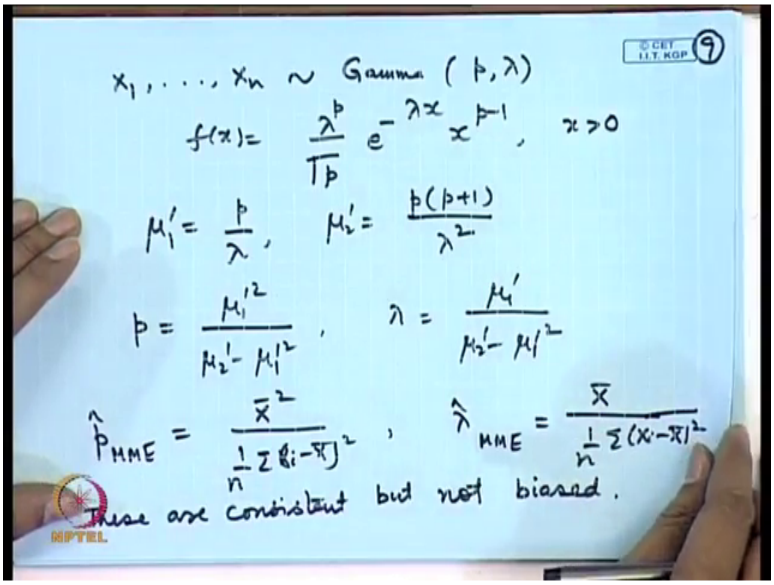


I mentioned that if I am having a k-dimensional parameter then we may usually considered k equations. So, why usually, because sometimes the k equations may not give us the desirable result. Let us take a very simple example sometimes for a k-dimensional parameter we may have to consider more than k 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. So, if I the substitute 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 a 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. I end up the section with two more examples; one for a two parameter gamma distribution, and one for a two parameter beta distribution.

(Refer Slide Time: 27:14)

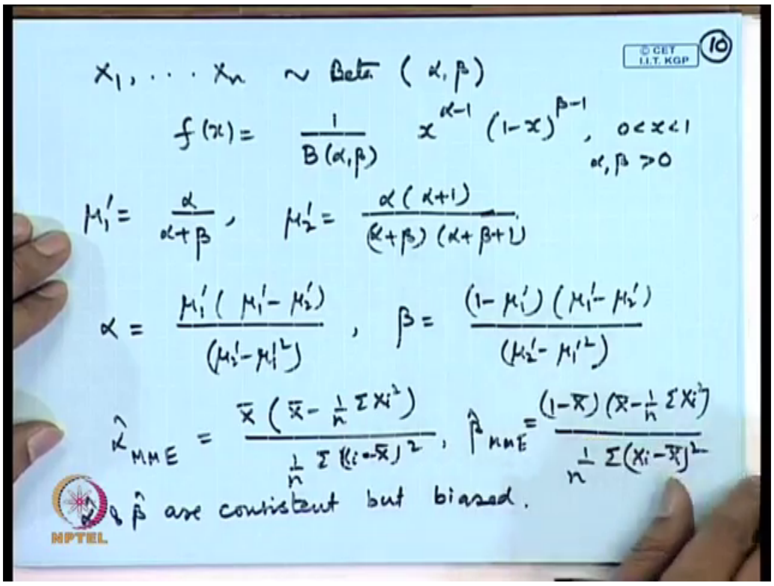


Let us take say X 1, X 2, X n following a gamma distribution with parameter say p and lambda. So, here lambda is corresponding to the rate of the corresponding Poisson process, that means, I am taking the density function is equal to lambda to the power p by gamma p e to the power minus lambda x x to 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 hat MME 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 moment 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: 29:28)



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 two 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 a 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 you substitute mu 1 prime as alpha 1 and mu 2 prime as alpha 2, we get the method of moments estimator 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 hat MME 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 hat and beta hat MME's they are consistent, but biased. 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 this function, so from here because of the basic weak log large numbers X bar are alpha 1 converges to mu 1 prime in probability and alpha 2 prime converges 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 the actually corresponding unbiased estimators. So, today's class I end up at this point.