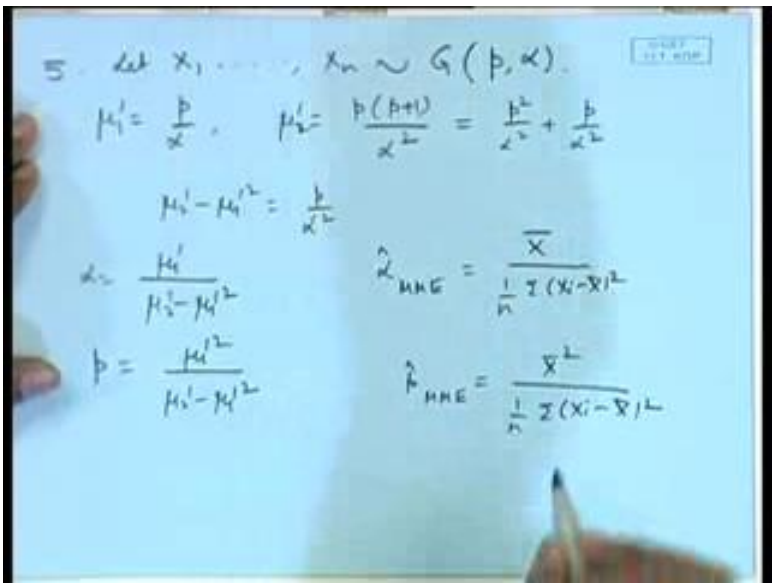
Probability and Statistics
Prof. Somesh Kumar
Department of Mathematics
Indian Institute of Technology, Kharagpur

Lecture – 56
Examples on MME, MLE

(Refer Slide Time: 00:23)

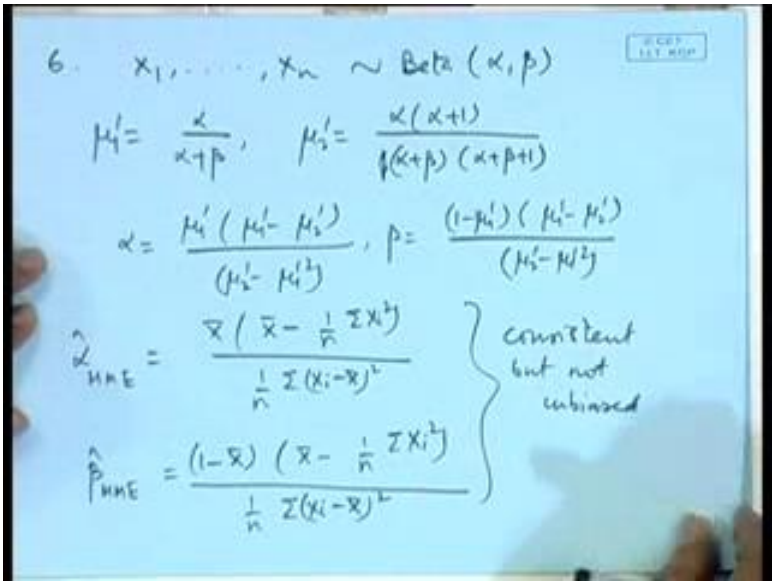


Let me take one more example here; let X 1, X 2, X n follow gamma distribution, and I will consider 2 parameter situations. So, a 2 parameter situation gamma p, alpha where p is the shape parameter and 1 by alpha is the scale parameter. So, here mu 1 prime is equal to p by alpha and mu 2 prime is equal 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 hat MME is equal to X bar divided by 1 by n, sigma X i minus X bar whole square and p hat MME is equal to X bar square divided by 1 by n sigma X i minus X bar whole 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 a straight forward to obtain.

(Refer Slide Time: 02:19)



Let us takes a beta distribution X 1, X 2, X n follow 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.

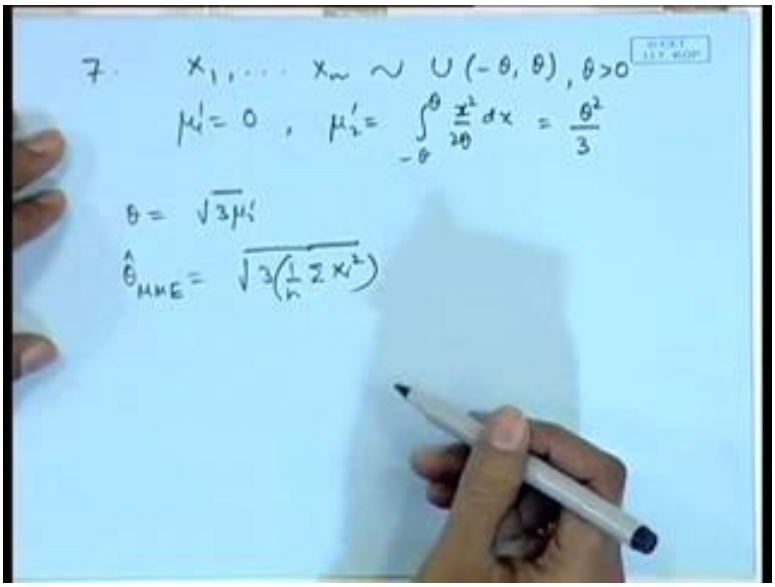
Now, if we solve these equations this is quite cumbersome, and the solutions turn out to be alpha is equal to mu 1 prime into mu 1 prime minus mu 2 prime divided by mu 2 prime minus mu 1 prime square, and beta is equal to 1 minus mu 1 prime into mu 1 prime minus mu 2 prime, divided by mu 2 prime minus mu 1 prime square. So, substituting for mu 1 prime as X 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 one 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 hat MME, 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 unbiasedness is extremely difficult, in fact they will not be unbiased. Of course consistency may still hold in 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, they 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 not 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 distributions 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: 05:37)

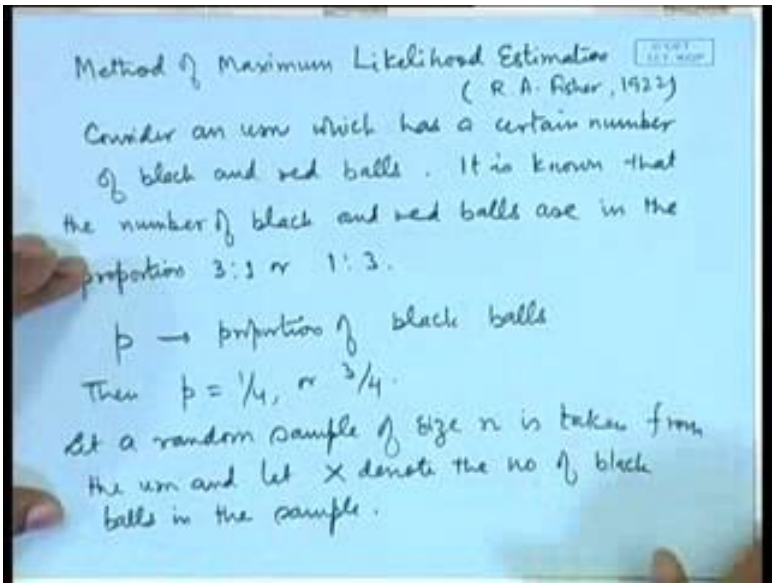


But there may be some situations where this kind of situation may arise, like if I consider say uniform minus theta to theta; that means, it is a uniform distribution on a symmetric interval, but we do not know the range of the interval therefore, theta is unknown.

So, here the problem is to estimated theta; if I consider the first moment actually it is 0. So, this does not give us any information about how to estimate the theta. So, now, if we go by a strictly by the rule of the method of moments estimation, then this is not an estimable function then; however, some particle solutions can be given we may look at mu 2 prime; that means, the second moment that is 1 by 2 theta and we look at x square dx 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 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 particle 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 half of the 20th century is the method of maximum likelihood estimation; this was introduced by R.A Fisher in 1922.

(Refer Slide Time: 07:29)



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 structures are used. However, fisher said that we should make use of the full probability model, because the moments structured does not make use of the 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.

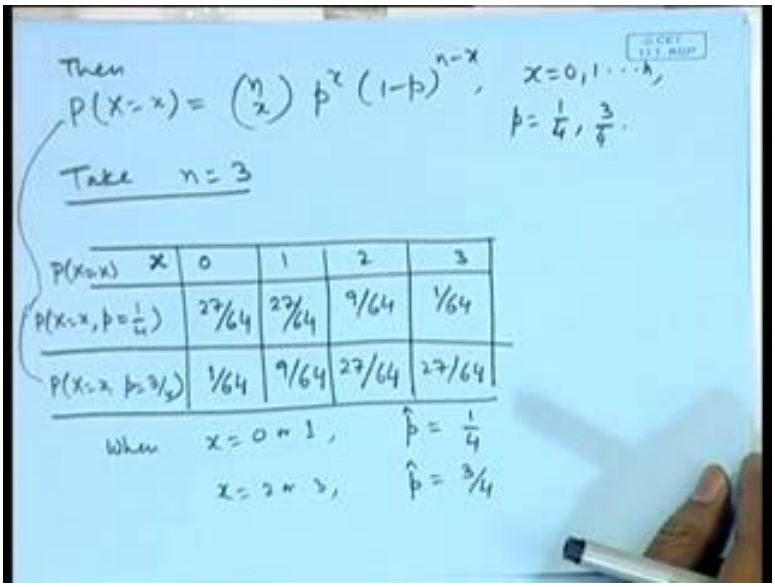
So, the method of maximum likelihood said that we should look at the form of the distribution itself and make full use of that. So, let us consider a simple example, consider an urn which has a certain number of say black and red balls. So, it is known

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 are more numerous.

So, if we say that there is n number of balls and x is the number 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 the either 1 by 4 or 3 by 4. Now let us see how 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 n is taken from the urn and let say X denote the number of black balls in the sample then you can write actually the distribution of x.

(Refer Slide Time: 11:20)



Probability that X is 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 to n and p can be 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 is 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 is 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 is 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 is 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 will 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 is equal to 3 this probability simply becomes p cube so it is 1 by 64 and 27 by 64.

So, as a layman what we will 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 1 by 4 is the likely values or you can say most likely value.

Similarly, when x is 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, 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.

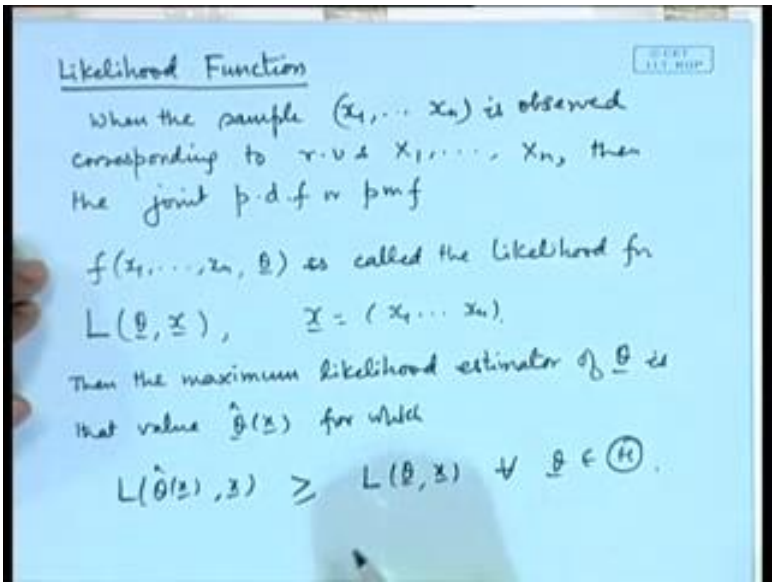
So, what we are doing? We are looking at the actual sample of observed. So, what is the probability or what is the parameter value for which the likelihood of that is sample

being observed is highest? That means, we are looking at the maximum likelihood function. So, what is the 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 is 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 is equal to 0 or 1, then I am taking p hat to be 1 by 4 and when x is 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 values likelihood function that is this value is actually higher.

Now, this is the case which I explained through 2 points in the parameter space, if I have in finite 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: 17:33)



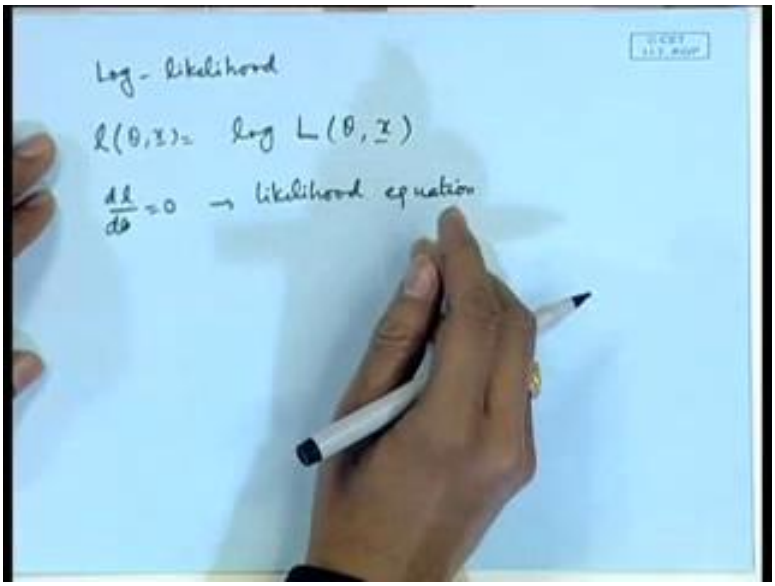
So, we called likelihood function. So, when the sample x 1, x 2, x n is observed corresponding to random variables x 1, x 2, x n then the joint probability density function or probability mass function that is f x 1, x 2 x n and the parameter that we call is called

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 hat x for which L theta hat x is greater than or equal to L theta for all theta belonging to theta.

So, basically it becomes optimization problem; that means, you have to find the extremism 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 methods. Sometimes we may be able to tell a straight away from the likelihood function or sometimes we have to use this kind of analysis.

So, now if you look at the form of the distribution such as binominal 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 are 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 case it is more convenient to handle log of the likelihood function.

(Refer Slide Time: 21:00)



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

also consider the case when theta is a vector for example, in the case of normal distribution etcetera and we will see that what are the solutions and another point is at sometimes this method of taking a 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 the solution may not be there. So, we will see a result where we say that the likelihood function always as 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 at these maximum likelihood estimator satisfy and then how do they compare with the other methods like the method of moments estimator etcetera.