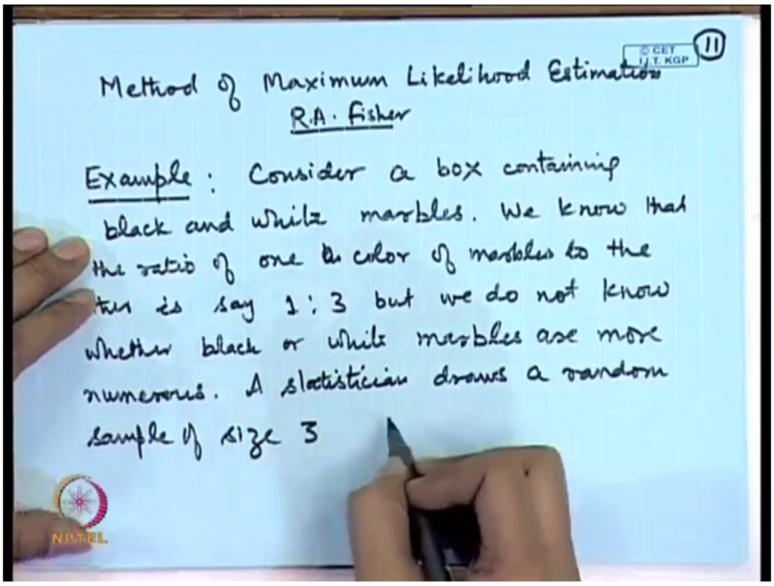
Statistical Inference
Prof. Somesh Kumar
Department of Mathematics
Indian Institute of Technology, Kharagpur

Lecture - 08
Finding Estimators -II

(Refer Slide Time: 00:48)



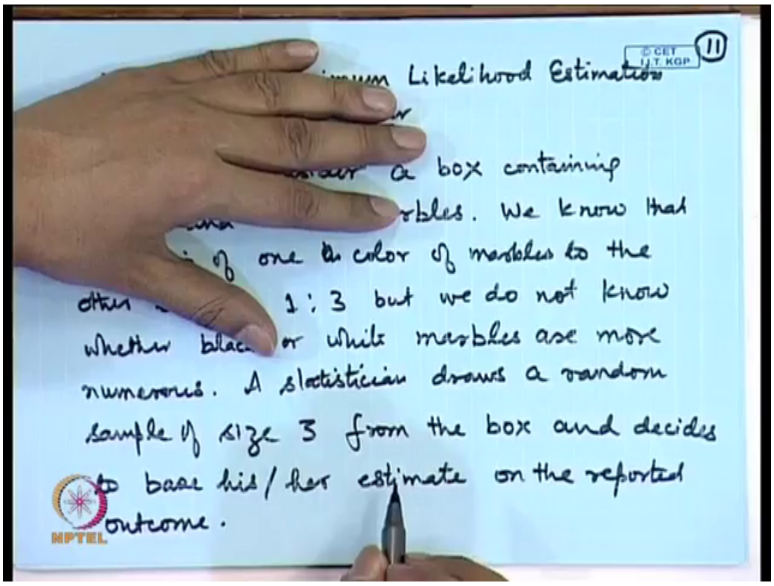
After this method of moments in the historical order, the method of maximum likelihood estimation is considered to be the most popular and the currently most commonly used method of estimation. This was developed by R. A. Fisher the method of maximum likelihood estimation. So, this method was developed by R. A. Fisher. Here he considers that all the problems in the inference should be based on the likelihood function. So, what is the 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 our 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 one 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 ones or white ones are 3 times the black ones that is not known. So, we only know we have a partial information that one of them is three times the other the number of marbles.

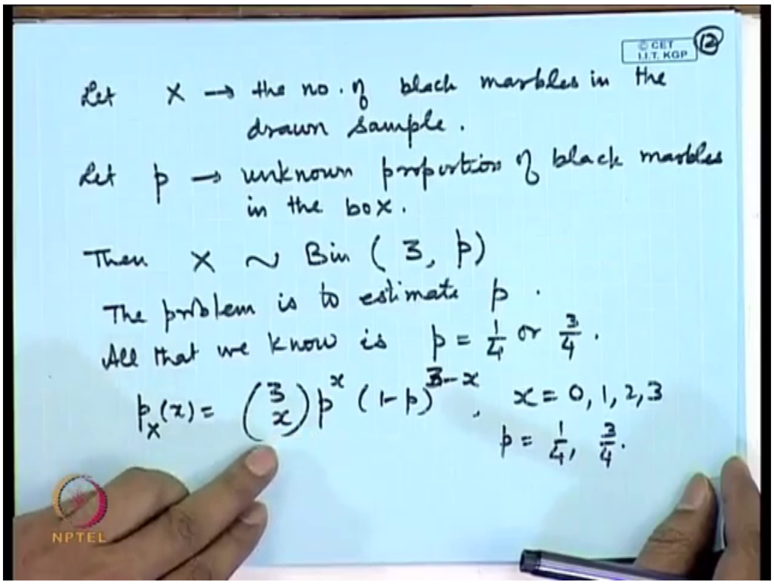
So, a statistician decides to take a random sample from the box, and based on that random sample he will give a inference, that means, he will give an estimator for the ratio. So, we can consider this problem in this following fashion. A statistician draws a random sample of size 3 from the box and decides to base his estimate on the reported outcome.

(Refer Slide Time: 04:24)



So, if you are drawing three marbles, now these three marbles may have some black or white marbles among themselves.

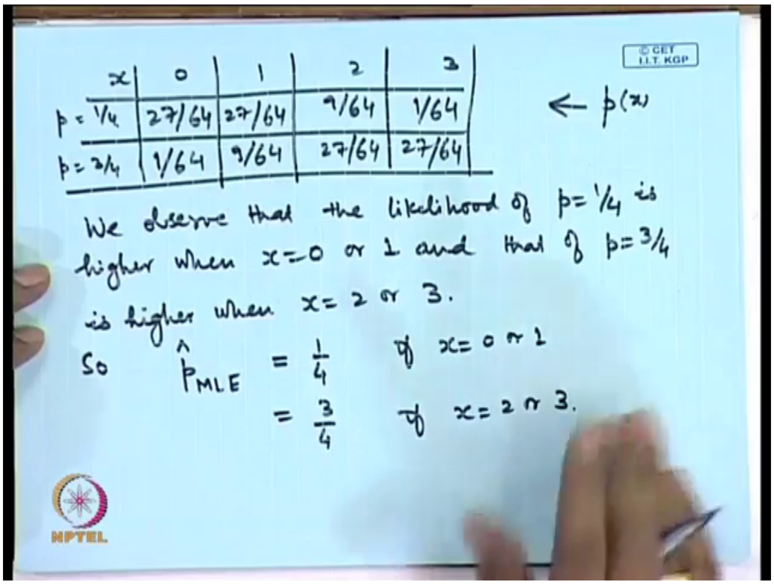
(Refer Slide Time: 05:05)



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 described by a probability model here. For example, X is a now at each draw, you will have a black marble or a white marble ok. So, you can consider it as a binomial trial, where if you draw a black marble, you will consider it as a success and if you draw white marble, you will consider it as a failure. The probability of success is p. So, in the total three 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 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: 07:48)



So, let me write it in the form of a table x can take value is 0 1, 2 and 3l 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 one, if we put p equal to 0, I get 1 minus p cube. So, 1 minus p cube if p is equal to 1 by 4 it becomes 3 by 4 cube that is 27 by 64; then p is equal to 3 by 4, this becomes 1 by 64.

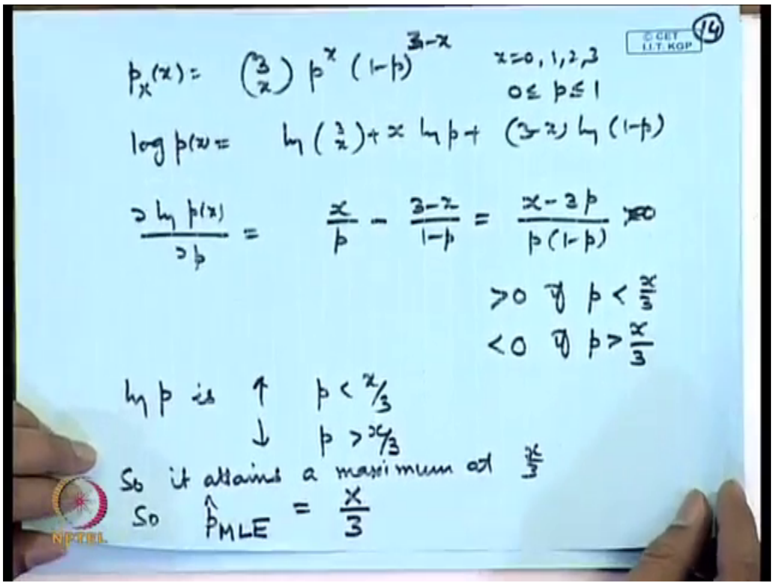
Similarly, let us consider x is equal to 1 if I take x is equal to 1, this is 3 p into 1 minus p square. So, again for p is equal to 1 by 4 this value turns out to be 27 by 64; whereas, that for 3 by 4, it turns out to be 9 by 64. In a similar way, I can consider other values for p equal to 1 by 4 and 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, if x equal to 2 or 3, p is equal to 3 by 4 has 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 observe 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 it. 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 exists 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. 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.

(Refer Slide Time: 11:55)



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 uncountable 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. Now, this is 3 minus x here.

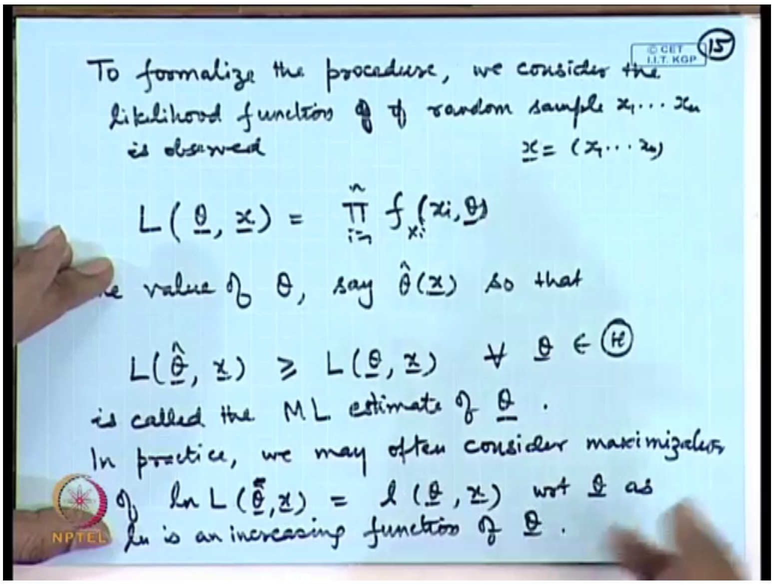
So, we may consider if you use the simple method of calculus, you have to look at the derivative or the behavior of the function as it is increasing or decreasing. So, for example, we may consider derivatives. Now, a simple method could be to take log of p that becomes log of 3 x x log p 3 minus x log 1 minus p. Now, this maximization of this with respect to p is same as maximization of this, because log is an 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 peak 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 the number of successes in 3 trials, then x by 3 is the sample proportion. And I am using the 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 add n here, then the only change 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.

(Refer Slide Time: 15:53)

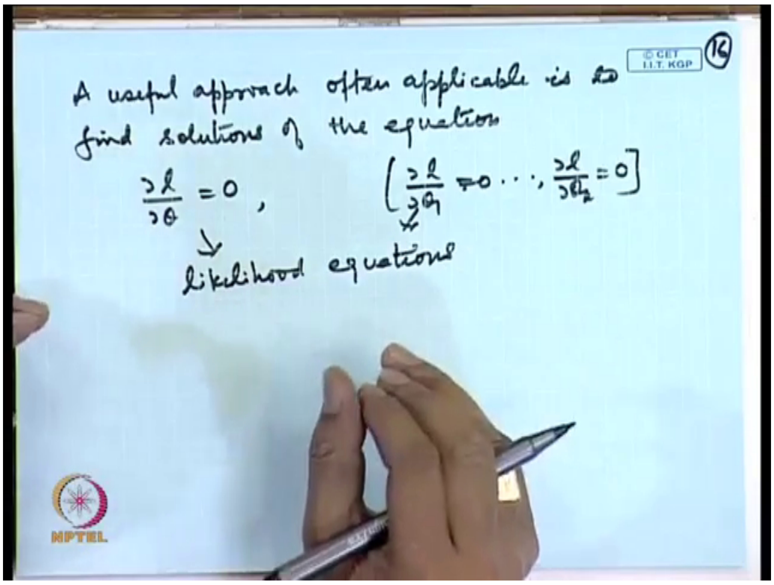


So, to formalize this method of maximum likelihood now we can say so to formalize the procedure, we consider the likelihood function of a random sample. So, likelihood function basically it is a I am considering it as a function of the parameter. So, if random sample say x 1, x 2, x n is observed then 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 s evaluated at the points x 1, x 2, x n which are actually the observed values.

So, I am considering it as actually a function of theta. So, the value of theta say theta hat x, so that L theta hat 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 1st case, we did not take log. We actually wrote the values here because only two possibilities were there

Whereas, in the 2nd 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 and in many practical problems it is applicable. So, in practice we may often consider maximization of log of let me call it a small l of theta, x with respect to theta as ln is a an increasing function of theta.

(Refer Slide Time: 19:02)



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 forthcoming class. So, today’s class I end up at this point.