Essentials of Data Science with R Software – 1
Professor. Shalabh
Department of Mathematics & Statistics
Indian Institute of Technology Kanpur
Lecture No. 61
Method of Maximum Likelihood and Rao Blackwell Theorem

Hello friends welcome to the course Essential of Data Science With R Software - I in which we are trying to understand the basic concepts of probability theory and statistical inference. And I hope now, the lectures are going to be more interesting because we are moving towards the real application that how the things are being obtained.

So, now, you can recall that in the last lecture, we had talked about the methods of estimation, and we had considered the method of moments. Now, exactly on the same lines whatever I had explained you that what is called estimation of parameters same thing is valid for this lecture also. In the last lecture, you had the use of philosophy that was based on the moments. So, you have devised a method of moments.

Now, in this lecture, we are going to devise one more idea one more philosophy to estimate the parameters, and that is based on maximum likelihood and after this, I will try to give you a small result, which is called s Rao Blackwell Theorem that will help you in obtaining some good estimators. So, firstly, let me try to give you an idea that what are these maximum likelihood estimators. So, let me try to give you a very real life simple situation where you are trying to take decisions in our day to day life.

Now, tell me one thing, we have a situation that you have to go from your home to say railway station, and suppose railway station is 5 kilometers away, and this is the time for the office hours so, means you have to go around 5 or 6 o'clock and suppose you have just one hour left for your train to catch.

So, now, suppose your train is at 6 o'clock and at 5 o'clock you are at home and you want to reach the station in time. So, now, you have couple of options by which you can go. One is by walking, second is by using a cycle, third is using some auto rickshaw, fourth here is using a taxi like a car, or fifth here is means, you just try to run.

Which of the method you are going to use? Now, remember one thing this is the traffic time and you have to pass through with a heavily traffic-loaded road. Now, you try to take a pause and try to think about it. But my question is, you can tell me the answer in within fraction of second, but my question is that what are you doing inside your brain because of which you

are trying to take such a decision Try to take a pause and try to think about it. And if you cannot think just try to start the video once again.

So, now what are you trying to do? You have couple of options. Say walking, run, say auto rickshaw, simple rickshaw, cycle, taxi, car etc. And now, your brain starts computing the probability of an event that what is the probability that if you go by this mode of transportation, then you will reach to the station say within one hour.

And you try to compute such probabilities for all the events, and you try to take care of all the possibilities means, there can be traffic, traffic can be high, there can be some traffic jam also near stations, the traffic is usually pretty high, and you are going in the say about 5 o'clock in the evening so the traffic is definitely going to be higher.

Now, you tell me, what you have done. When I started the course, you always felt that that the computation of probability is very difficult, but now ask yourself and tell me you have computed the probability of such a complicated event very easily in a fraction of second. So how can you say that you do not know statistics, and you do not know how to compute the probability.

So, now, let us try to understand what you have done. You have taken, suppose in order to make it more simple, suppose I consider here three options, walking, taking a simple auto rickshaw that is a three wheeler and try to take a car. Now, try to calculate what is the probability that if you walk, you will reach in time? Well, you have to walk five, six kilometers with your luggage so the probability will be say, say 20 percent suppose.

Now in case if you try to take an auto rickshaw, auto rickshaw that they are the 3 wheelers, they are a small, they can go through here and there very easily. And if there is a traffic jam possibly they can take a different route, and they can take some small narrow road and they will make you reach to the station. So, the probability to reach a station is going to be something like 80 percent, 90 percent.

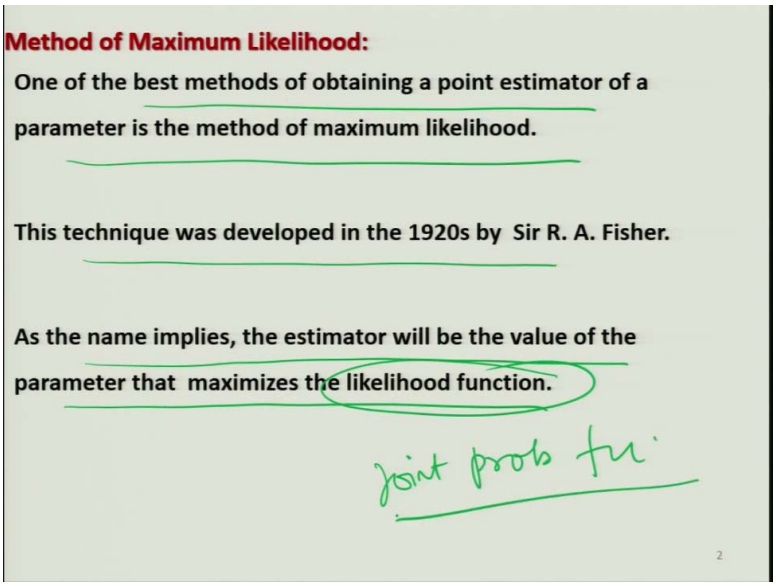
And then the third option is you take your car. Somebody will drop you using your personal car. Now, with the personal car, if there is a traffic jam, you just cannot do anything. If there is a traffic the speed of the car is going to be extremely slow with the risk of traffic jam. So, the probability that you will reach by your own car that is also say 40 percent, 50 percent. Well, these numbers are just hypothetical just to illustrate you, otherwise, you yourself can compute some probability depending on the conditions.

So, now, if you try to see, what are you trying to see, you have here three probabilities 20 percent say 80 percent and say 50 percent. You try to choose the whichever has got the maximum probability and then corresponding to that you try to see what is the mode of transportation by which the probability is maximum that you can reach? That is your auto rickshaw in the given time frame.

So, what you have done, you have obtained a sample of observations and then you have computed the joint probability of all those observations and then you have maximized the probability. And then whatever is the value of the parameter corresponding to the maximum probability that you are trying to say this is the most probable value. What is that? This is maximum likelihood estimator.

So, now, I have given you an example where I should be 100 percent confident that you know the maximum likelihood distribution. The only thing is, you did not knew the name. You knew how to compute it, but you did not know what exactly you are trying to do mathematically. So, that is my objective in this lecture. And let us try to begin this lecture and try to understand what is this method of maximum likelihood to estimate the unknown parameters on the basis of a given sample of data.

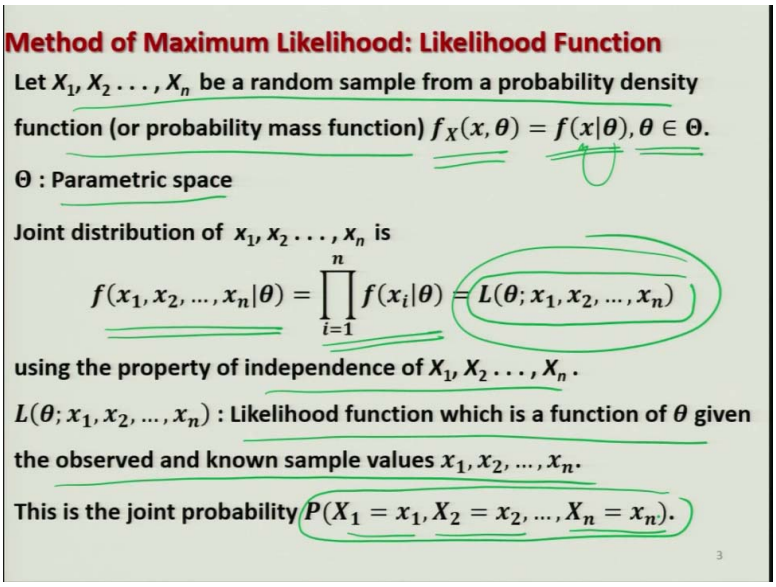
(Refer Slide Time: 07: 15)



So, now, let us come back towards slides and try to begin the lecture. So, one of the best methods of obtaining point estimator of a parameter is the method of maximum likelihood this technique was developed sometimes in 1920s by Sir R.A. Fisher, who is said to be the father of the statistics. As the name implies, the estimator will be the value of the parameter

that maximize the likelihood function. What is the likelihood function? Likelihood function is nothing, but the joint probability function. And you know how to compute the joint probability function either in the case of discrete or continuous random variable as simple as that.

(Refer Slide Time: 07: 54)



So, likelihood function is simply a name or an alternative name for the joint probability function. And when I am trying to say the likelihood function, usually we do not discriminate between the probability mass function or the probability density function. Whatever is that joint function in the case of discrete that is going to be joint probability mass function, and in the case of continuous it is going to be joint probability density function that we try to consider as likelihood function.

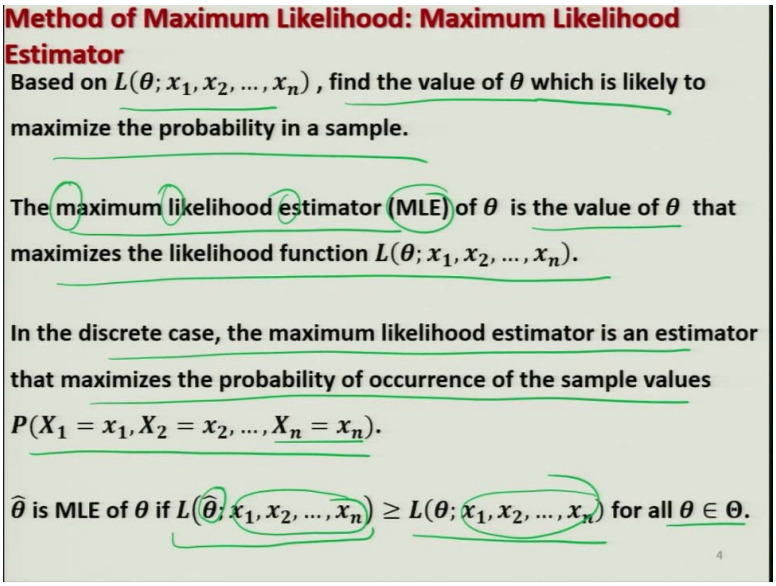
So, first let us try to understand what are the steps that are involved. Well, there can be different steps depending on the problem, but I will try my best to give you a couple of examples, so that I can satisfy you. So, basically, what are you trying to do, you are simply trying to maximize the probability and you are trying to find out the value of the θ corresponding to which given the sample the probability is going to be the maximum.

So, let X_1, X_2,…, X_n be a random sample from sum probability density function or probability mass function say $f_X(x, \theta)$, and we try to write down here as say $f(x|\theta)$, because θ here is a fixed and we are trying to find out the value of θ for a given value of X_1, X_2,…, X_n; the observed sample space and we assume that $\theta \in \Theta$, and Θ is the parameter space.

Now, this is a random sample, so the observations are independent. So, the joint distribution of $X_1, X_2, \ldots, X_n$ can be written here as like this $f(x_1, x_2, \ldots, x_n|\theta)$ this is simply the product of their marginal densities or their product of their individual PDFs. And this is in general indicated by $L(\theta; x_1, x_2, \ldots, x_n)$. So, here we have used the property of independence. And that you know that two random variables are independent if their joint density function f(X, Y) can be expressed as the marginal density function of X and Y that is f(X) into f(Y). So, that is the rule which we have used here.

And now this L in general is going to indicate our joint probability density function, which is now called as likelihood function. So, likelihood function is a function of θ given the observed and known sample values $X_1, X_2, \ldots, X_n$ one which is the joint probability in case of a discrete random variable like, as probability of $X_1 = x_1$ $X_2 = x_2$ and X_n is=a x_n, just like the same concept that you studied earlier.

(Refer Slide Time: 10: 39)

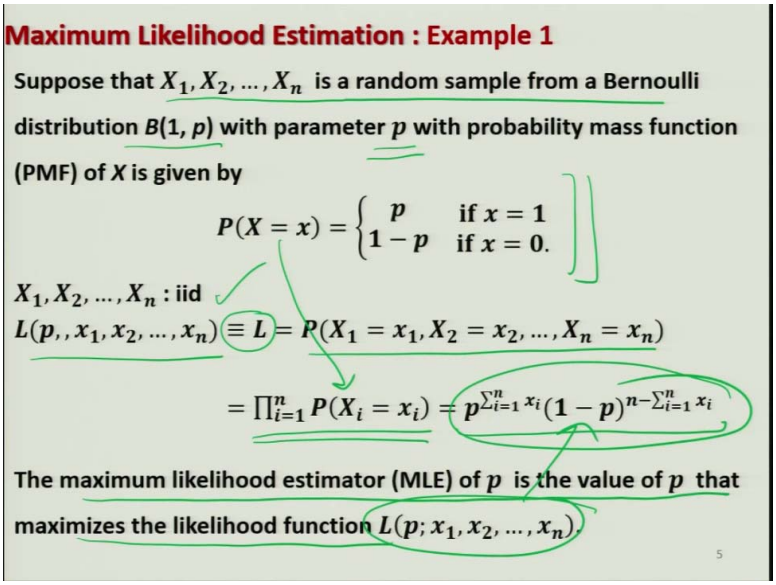


Now, based on this likelihood function, we have to find the value of θ which is likely to maximize the probability in a sample, and the sample is given to us. So, the maximum likelihood estimator, which is abbreviated as m MLE, M for Maximum L from Likelihood E from Estimator, the mle of θ is the value of θ that maximizes the likelihood function, as simple as that. And in case if you are trying to consider the discrete case, the maximum likelihood estimator is an estimator that maximize the probability of occurrence of the sample values $X_1, X_2, \ldots, X_n$.

So, this can be written here as a probability that $X_1=x_1$, $X_2=x_2$ and $X_n= x_n$. The usual approach of writing the joint probability. Now, $X_1, X_2,\ldots, X_n$ are now obtained, they are now given. Now, you simply try to find out the value of θ, say as $\hat{\theta}$. Now the value of the likelihood function at $\theta=\hat{\theta}$ is always going to be greater than the value of the likelihood function at θ for all θ belonging to the parametric space Θ because you are trying to maximize the likelihood function.

You are trying to maximize the probability of observing the sample from a given population whose parameter is θ. So, definitely the likelihood function is going to be the highest at the point of maximum likelihood estimator.

(Refer Slide Time: 12: 14)



And after that, you simply have to solve it, and you will get the maximum likelihood estimator. Well, when we try to estimate it on the basis of given sample of data, then have different types of things are coming into picture and I will try to explain you with some examples.

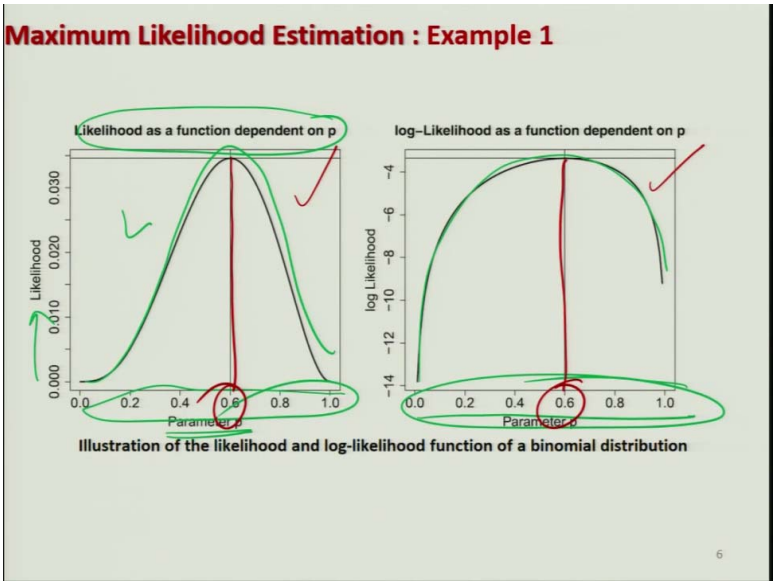
So, the first example I take here where I am considering the sample $X_1, X_2,\ldots, X_n$ from a Bernoulli distribution, this is a random sample, and the parameter here is P whose probability mass function is given by this one. So, your $X_1, X_2,\ldots, X_n$ they are IID so, the likelihood function can be written here as say as the joint probability of $X_1, X_2,\ldots, X_n$ like this one and it is just indicated by here L for the sake of simplicity in writing again and again.

So, now, this probability function can be written as a product of this probability density function for X_i. Now, in case if you try to simply solve it this will come out to be like this $p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i}$. So, the maximum likelihood estimator MLE is the value of p that maximizes the likelihood function, this one, which is now here.

Now, the next question is, how are you going to maximize it? Now, the optimization techniques now come into picture. In case if the likelihood function is complicated, you can use any numerical technique to maximize it. You can use different types of algorithm in case of any complicated function and so, on.

So, now, once you obtain the value of here p from this function, which is maximizing this function, there is a reason for me to be happy that I have got the maximum likelihood estimate of p. Now, how to maximize it? That is not the question which is uncertain in statistics. For that you have to take the help of mathematics and optimization theory. So, in this particular case, we are simply going to use the principle of maxima and minima.

(Refer Slide Time: 14:13)

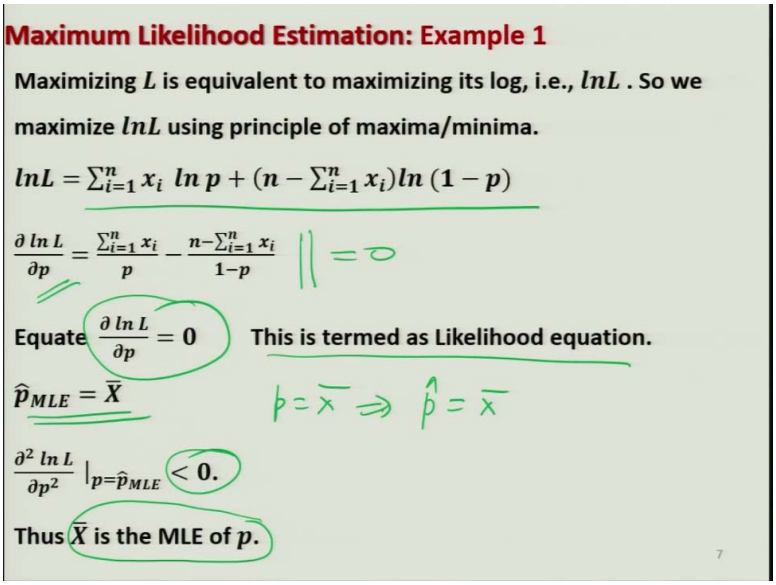


And many times, we will try to transform the likelihood function in some suitable way for example, lock function because both are monotonic functions. So, in case if you try to find out the maximum from the L, then the same maximum can also be obtained by maximizing the log of L. And in case if you try to understand this concept from these two figures. So, you can see here that here we have plotted the likelihood function as a function of p. We have simply plotted here the L for different values of the parameters and their probabilities.

Now, we try to take the log of this likelihood function, and we try to find out their different values and we try to plot it on the same scale. You can see here this scale and this scale, they are the same and the curve will here look like this. But in both cases you can see here that the point at which it is reaching the maximum is here the same that is 0.6 and 0.6. So, either you try to maximize this figure or this figure that will not make much difference or that will actually not make any difference.

Once you are trying to use any numerical technique optimization theory for some complicated function. It may be possible that different algorithms may give you different values, but those values are not going to be much different, and they will be dependent value that is what I was trying to say.

(Refer Slide Time: 15:40)



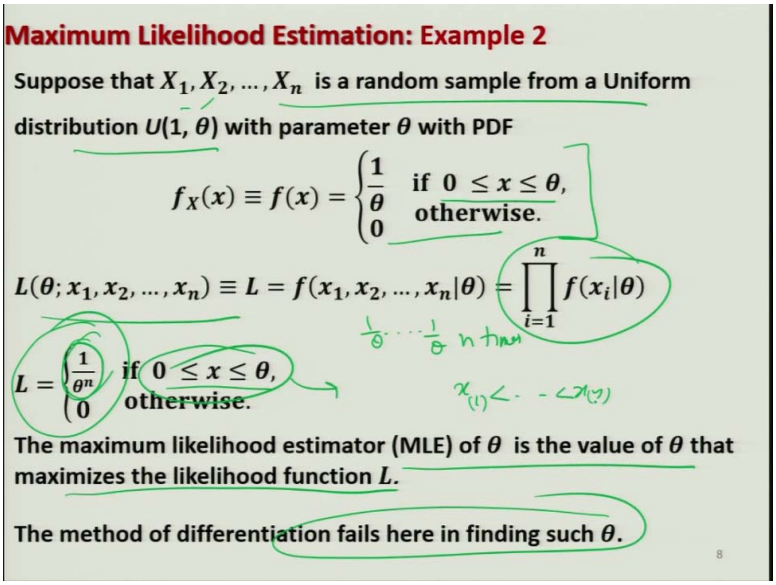
So, now, in this case, we simply try to use the principle of maxima, minima and we try to take here the log and we try to maximize it. So, in case if you try to take here the log it will come out to be here like this $p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i}$. And if you try to simply differentiate it with respect to the parameter it will be here like this and then whatever is the dislocation that is substituted to 0.

So, this equation when we are trying to equate it first derivative or any derivative of the likelihood function or the log likelihood function to zero this is usually termed as likelihood equation. Well, in case if there are more than one parameter say two parameters, then you have to differentiate the likelihood function with respect to each of the parameter then there are going to be two likelihood equations.

So, now, in case if you simply try to differentiate it put it equal to here zero, you obtain here p= $\bar{X}$. And this implies that $\hat{p}$ is simply equal to here the sample mean like this one. Now, the next question is, how to know whether this value is going to give us the maximum or minimum. So, in case if you try to substitute p= $\bar{X}$ in the second derivative of the log likelihood, then this term will come out to be negative, that you can verify, that is a very simple thing.

So, you can see here now, the sample mean is the maximum likelihood estimator of the population proportion in case of Bernoulli distribution. So, that means, in case if you do not know what the value of p, you have obtained a value you have obtained some values in a sample, simply try to find out their automatic mean. And values are going to be in terms of 0 and 1 so the automatic mean will give you the value of or the estimated value of the population proportion.

(Refer Slide Time: 17:29)

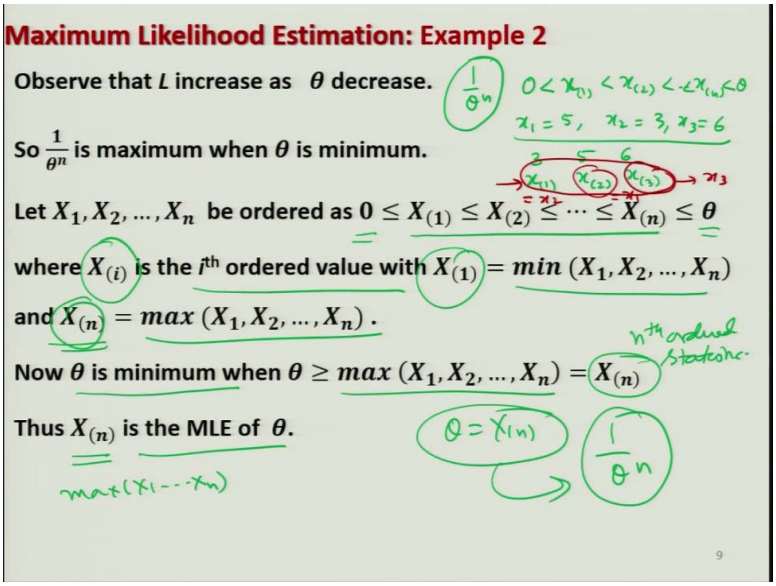


Now, let me try to teach you one more example, where this principle of maxima and minima is not going to work. So, let X_1, X_2,…, X_n be a random sample from a uniform distribution whose parameters are 1 and θ. So, the PDF here is given here like this $f(x) = \begin{cases} \frac{1}{\theta} & \text{if } 0 \le x \le \theta, \\ 0 & \text{otherwise.} \end{cases}$. Now in this case all X_1, X_2,…, X_n they are IIDs. So, the likelihood function can be obtained just by multiplying individual probability density function. So, you simply have to write down here like this and this will become here 1/θ into 1/θ into 1/ θ n times.

So, the likelihood function will come out to here like this $\begin{cases} \frac{1}{\theta^n} & \text{If } 0 \le x \le \theta, \\ 0 & \text{otherwise.} \end{cases}$. Now, as per the rules of the maximum likelihood estimation the maximum likelihood estimator MLE of θ is the value of θ that maximizes the likelihood function this L. Now, in case if you try to employ here the principle of a maximum minimum then it will actually fail this will not give you any result, but this is not the problem of likelihood function, these are the mathematical things that you have to choose an appropriate method to maximize the function.

So, in case if the method of differentiation fails here to finding out θ, it does not mean that the MLE does not exist. So, now, in case if you try to look at this function 1 upon θ and this year range, you can see here you have here the observation X_1, X_2,…, X_n, so what we try to do here that we try to arrange them in some order. So, all these X_1, X_2,…, X_n they are going to lie between zero and θ and that is what I am trying to do here.

(Refer Slide Time: 19:10)



That you have here the function one upon to the power of here and. And now, you are trying to order those observation here as say the X_1 less than X_2 up to here less than X_n. And all of them are going to be lie in the range 0 to θ. What is the meaning of this? Suppose, I have here X_1 equal to suppose 5, X_2 equal to here 3 and X_3 equal to here, suppose, here 6. So now you can see here, whatever is the minimum value out of them this is here 3 three and then here we have here 5 and then we have a 6.

So, we try to denote this minimum value as the X and inside the parentheses we write the 1 in the subscript, and then this here X_2, and the same way for indicating the value of 5 and X_3

that means 3 inside the parentheses in the subscript. So, these are the values of X_1, X_2, X_3 three that we are trying to indicate. So, you can see here that here, this X_1 inside the parenthesis we have written 1, this is same as here, X_2, and this X_2 where I have written the 2 inside the parentheses in the subscript, this is here same as here X_1 and this X_3, this is here, same as here X_3.

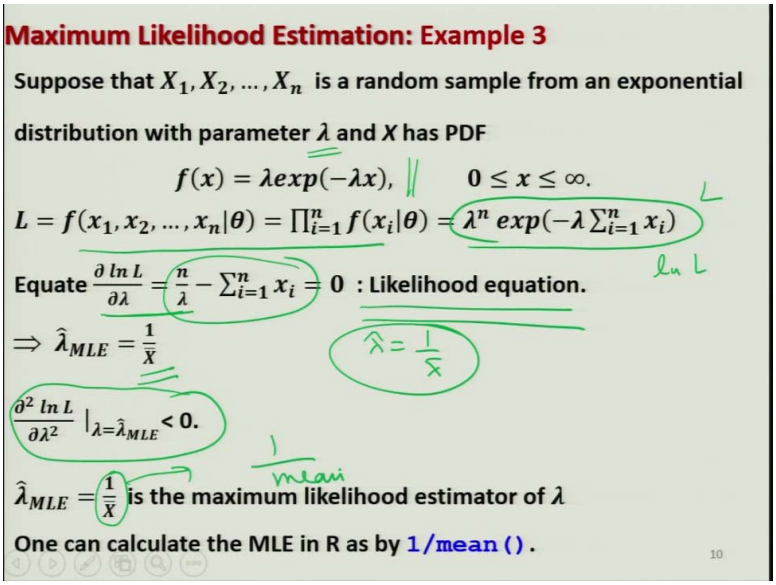
So, you can see here, these X_1, X_2, X_3 in which I have written 1, 2, 3 inside the parentheses, they are the ordered value. That means the X_1, X_2, X_3 are the original observation, they are arranged in some increasing or decreasing order and X_1, X_2, X_3 when I am trying to write down 1, 2, 3 inside the parentheses in that subscript then these are the ordered values. So, we try to arrange the sample values in this order.

So, all the sample values X_1, X_2,…, X_n can be written like this, and they will be lying between 0 and θ because every X_i is lying between 0 and θ. And now, you have to observe under what type of condition this $\frac{1}{\theta^n}$ function is going to take the maximum value. So, one thing you have to just notice that this is actually called as ordered statistics. And means, if I am writing here like this X and then i inside the parentheses in the subscript, this is the ith order value, and this X_1 this will simply be the minimum value out of X_1, X_2,…, X_n and x and here that will be the maximum value of X_1, X_2,…, X_n.

So, now, you can see here your likelihood function is 1 upon θ is power of n. So, now, each of this ordered value is going to lie between 0 and θ. So, I can say that θ is going to take the minimum value when θ is greater than or equal to maximum of X_1, X_2,…, X_n, which is the nth order statistics. So, now, you are going to maximize it. So, θ is minimum when $\theta = X_n$. So, what we want? We want the likelihood function to be maximum, and our likelihood function is in the form of $1/\theta$.

So, obviously, in case if you try to take here $\theta = X_n$ this likelihood function is going to be maximized, and that is why I can say that X_n which is the maximum value of the sample values X_1, X_2,…, X_n is the maximum likelihood estimator of θ. So, you can see here, here you are trying to observe the form of the function of likelihood function, and then you are trying to find out the value of θ which is going to maximizing it.

(Refer Slide Time: 22:58)



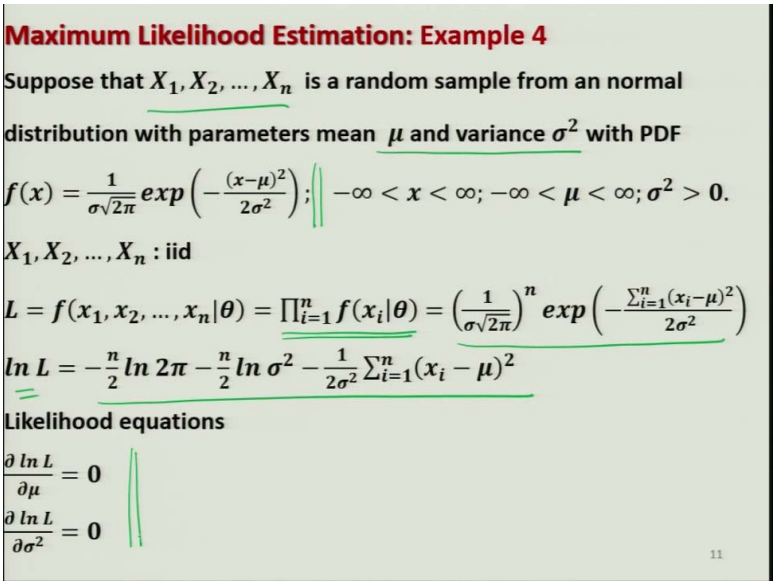
Now, similarly, if I try to take here one more example where X_1, X_2,…, X_n they are coming from an exponential distribution with parameter λ, then X has got this PDF and now, we have to estimate the λ using the method of maximum likelihood. So, we try to obtain here the likelihood function exactly in the same way, we try to multiply the individual PDFs and we obtained this function. So, this is going to be your here L.

Now, you can see here that in case if you try to take the natural log and you try to differentiate it, then it is easier to implement a fistful of maxima and minima. So, we try to differentiate this log of L with respect to the λ and we get here this type of equation $\frac{n}{\lambda} - \sum_{i=1}^{n} x_i$ and try to substitute it to 0. So, this will give us the likelihood equation.

And now, if you try to solve this equation, you will get here say $\lambda = 1/\bar{X}$ and this is going to indicate that $\hat{\lambda} = 1/\bar{X}$, and this $\hat{\lambda}$ is going to be the maximum likelihood estimator of λ. But we need to ensure that this is the value of lambda, which is really going to maximize the likelihood function, and for that, we simply try to check the second order derivative condition of maxima and minima.

And this inform us yes we are correct and $\hat{\lambda} = 1/\bar{X}$ that is 1 upon sample mean is the maximum likelihood estimator of λ for the exponential distribution. Now, in case if you want to compute this value on the basis of R software, this is 1 upon here mean that is all.

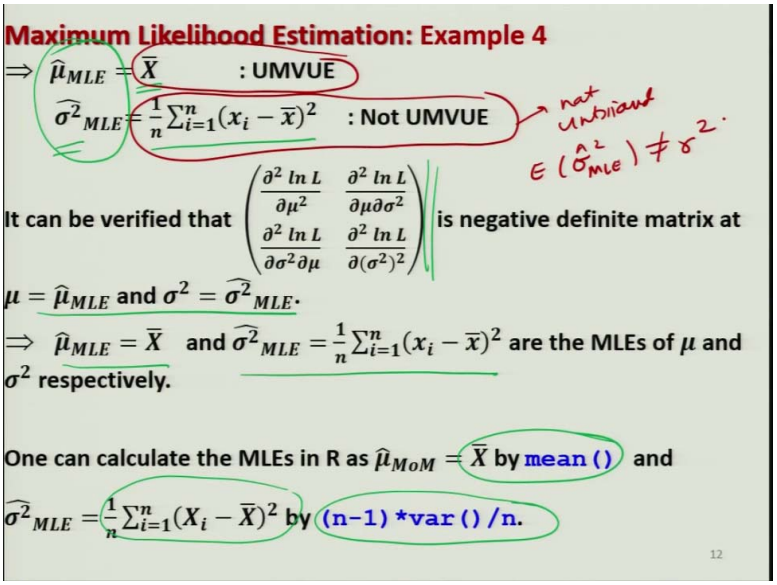
(Refer Slide Time: 24:44)



And similarly, if I try to take here one more example from the normal population with mean, μ and σ^2, so we have got this sample from this population, and the PDF is given here by this $\frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right); \quad -\infty < x < \infty; -\infty < \mu < \infty; \sigma^2 > 0$. Now, I try to write down here the likelihood function just by multiplying all the PDFs for X_1, X_2,…, X_n and this function can be obtained here like this.

Now, you know that by looking at the form of this function it is easier to take the log and then apply the principle of maxima minima. So, I try to take care of a log of this likelihood function which comes out to be here like this and then I try to differentiate this log of L with respect to μ and σ square and we obtain here two likelihood equations.

(Refer Slide Time: 25:37)

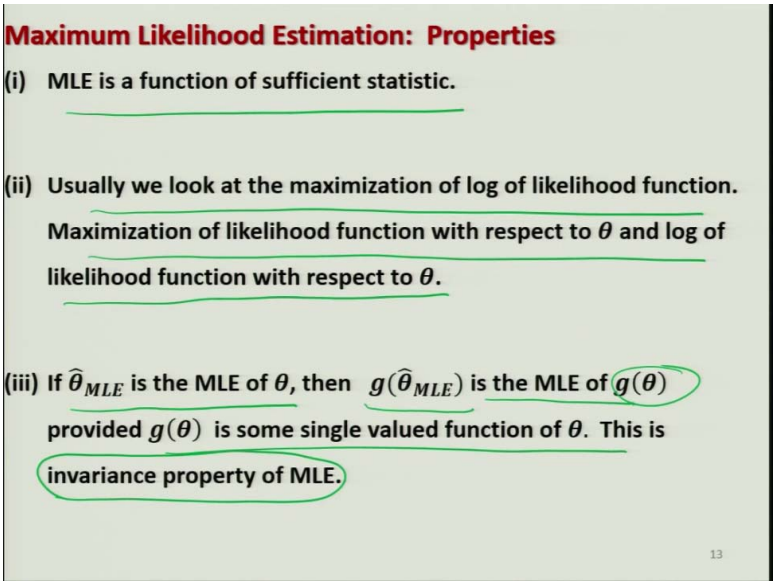


And now, once they are there, it just is solving them we can obtain here that μ hat = $\bar{X}$ and σ square hat= 1/n summation i goes from 1 to n Xi minus $\bar{X}$ whole square. So, these are the maximum likelihood estimates of μ and σ square. And in case if you want to check whether these values are going to provide the maximum of the likelihood function or not, then we have to check whether this matrix is a negative definite matrix or not at μ equal to $\hat{\mu}$ and σ^2 equal to $\hat{\sigma}^2$, and that can be verified that it is coming out to be true.

And after that, we are sure that $\hat{\mu} = \bar{X}$ and $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$ are the MLEs of μ and σ^2 respectively. So, now, in case if you ask me that, how to calculate this thing then I know that this $\bar{X}$ can be computed by mean, and this variance $\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$ can be obtained by this expression n minus 1 upon n into variance of that data vector right.

So, now, that we already have done when we were trying to discuss about the basics of the R software, Now, one thing before I move forward. If you try to see here this $\bar{X}$ this is going to be uniformly minimum variance and bias estimator of μ that we will – that we can show but this estimator of σ^2 is not going to be the uniformly minimum variance and bias estimator and that is clear here at least I know this is not unbiased. Expected value of σ^2 hat MLE is not equal to σ^2. Well, these things are possible, nothing to surprise.

(Refer Slide Time: 27:37)



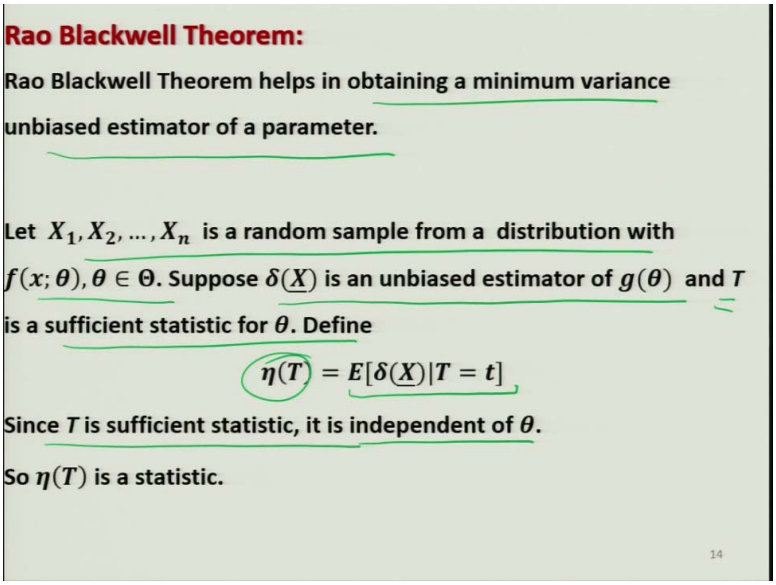
Now, I have taken a couple of examples to explain you. And the moral of the story is that you simply have to write down the likelihood function and then you have to just optimize it. You have to find out the maximum value. So, how to do it that now depends on you. You have a good knowledge of mathematics, now lots of optimization technique, numerical technique, you can use any one of them. And in he R software also, there are couple of commands, but I am not going into that detail, but because my objective here is to give you the basic fundamentals competition is very easy for you.

So, now, some results and some properties of this MLEs. So, first important properties MLE is a function of sufficient statistics. As you can always see, you have seen that say sample mean sample mean, you have already proved that this is a sufficient statistics, and you are obtaining the sample mean to estimate the population mean in the case of normal population.

So, and the second thing is this usually we look at the maximization of the log of likelihood function maximization of likelihood function with respect to θ and log of likelihood function with respect to θ they are going to give you the same value and third property is if $\hat{\theta}$ is the MLE of θ, then any function of $\hat{\theta}$ say g($\hat{\theta}$) is the MLE of g(θ), provided g(θ)is some single valued function of θ.

So, this is called as invariance property of maximum likelihood estimates. And this helps us a lot in finding out different types of estimator for different parameters of the from the same population.

(Refer Slide Time: 29:17)

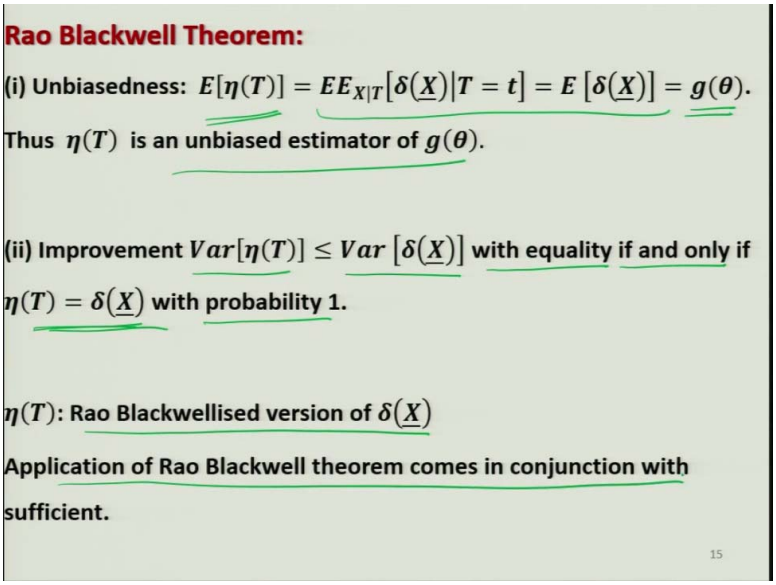


Now, let me try to give you a quick idea about the Rao Blackwell Theorem, because this is going to help us in obtaining a minimum variance unbiased estimator of a parameter. So, that is a very simple thing. One thing you have to keep in mind, the statement of the theorem may not give you a real idea that how are you going to obtain the minimum variance unbiased estimator, but, as soon as I try to give you the working rules, then the life will become very simple.

So, now, let me give you this idea quickly. So, let X_1, X_2,…, X_n be a random sample from some distribution f x θ where θ belongs to θ. And suppose $\delta(\underline{X})$ is an unbiased estimator of some parameter function g(θ). And suppose T is a sufficient statistic for θ. Now, we are trying to define a new statistics $\eta(T)$. And this is defined as expected value of $\delta(\underline{X})$ given T=t.

Now, since you have already assumed that T is a sufficient statistics for θ so by the definition of sufficient statistics, we can say that this $\eta(T)$ is going to be independent of θ. Yes, if you have forgotten it, just go back and try to look into the definition of sufficient statistics. Well, we had done two things, definition and the factorization theorem, but now, I am asking you to look into the definition. So, now, I can see here this that $\eta(T)$ is a statistic.

(Refer Slide Time: 30:53)

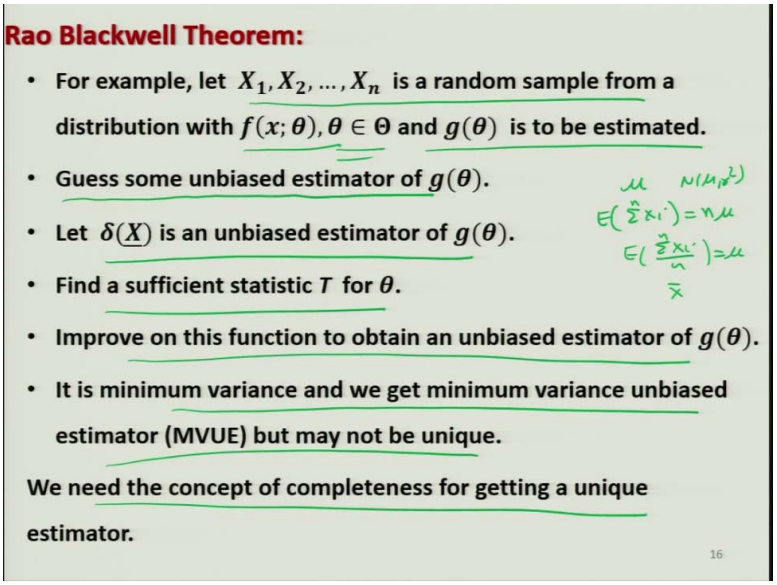


So, now, in case if you try to look at the properties of this $\eta(T)$ so far about unbiasedness you can see here expected value of $\eta(T)$, if you try to find out the expected value here and this will come out to be same as g(θ). So, this $\eta(T)$ is an unbiased estimator of g(θ). And here I have used the rules which you understood that conditional, unconditional expectation of a variable.

And now, in case if you try to find out the variance of $\eta(T)$, under what type of condition it is going to be smaller than the variance of other estimator delta. So, this inequality that variance of $\eta(T)$ is going to be smaller than or equal to variants of $\delta(\underline{X})$ inequality, if and only if $\eta(T)$ = $\delta(\underline{X})$ with probability 1. So that means, if you try to create a statistics with this formulation that will always have a smaller variance than the $\delta(\underline{X})$. And if you try to see what we have done here, you had one estimator $\delta(\underline{X})$, and based on that, you are trying to create another estimator here et using $\delta(\underline{X})$.

So, now, this is going to give you a smaller variance if and only if $\eta(T)$=$\delta(\underline{X})$. This eta-t is called as a Rao Blackwellised version of $\delta(\underline{X})$. So, we are trying to consider an estimator, and then we try to convert it and we try to transform it and this transformation is going to be done by a certain rule, and this will give us a statistics that is called as a Rao Blackwellilsed version of the original estimator. And the application of this Rao Blackwell Theorem comes in conjunction with sufficiency.

(Refer Slide Time: 32:35)



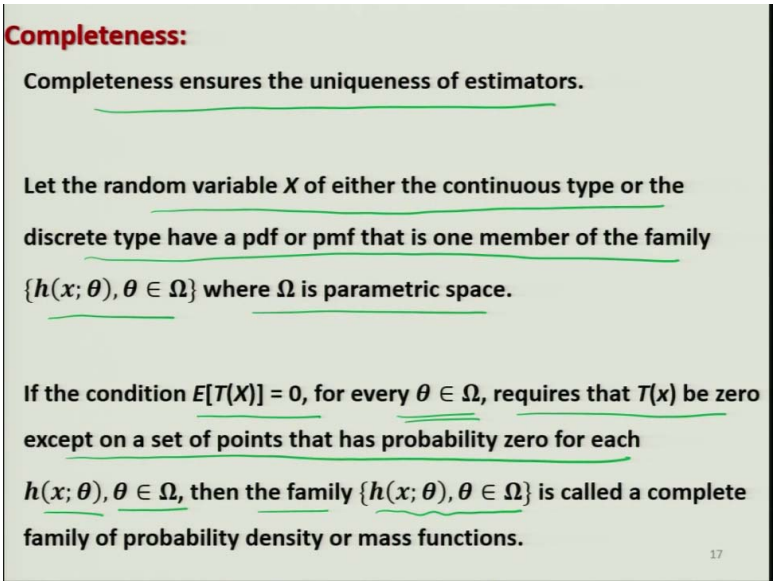
So, now, I try to give you here some rules which are going to make your life very simple that instead of doing all these things, you have to do very simple thing and then you will get here a minimum variance unbiased estimator. So, for example, the first step is suppose the setup is like this that X_1, X_2,…, X_n is a random sample from some distribution f x θ, θ belong to Θ, and g(θ) is the parametric function that we want to estimate. Now, the first step is simply make a guess about some unbiased estimator of g(θ), that is not difficult I am promising you.

And you try to just have it. For example, if you ask me, if I have to estimate the population mean from a $N(\mu, \sigma^2)$, possibly I can say okay, let me start with $\sum_{i=1}^{n} X_i$, I try to find out here the expectation value of summation Xi and it comes out to be here n upon μ. So, now, I want to estimate here μ. So, I can write down here $\sum_{i=1}^{n} X_i/n$ this is equal to here μ. So, now, this X bar becomes an unbiased estimator of μ. So, this type of exercise that we try to do.

So finally, we have here an estimator of g(θ). So, let this $\delta(\underline{X})$ as an unbiased estimator of g(θ). Now, find sufficient statistic T for θ. That you know? You have to simply use that Neyman-Fisher factorization Theorem and now improve on this function to obtain an unbiased estimator of g(θ), simply try to find out sufficient statistics and try to convert it into an unbiased estimator of g(θ).

And it has got minimum variance and we get minimum variance and bias estimator, but this may not be unique. For uniqueness we need here one more concept this is called as completeness. The concept of completeness is we need to get a unique estimator. And now what we have to do, we have to first understand the concept of completeness.

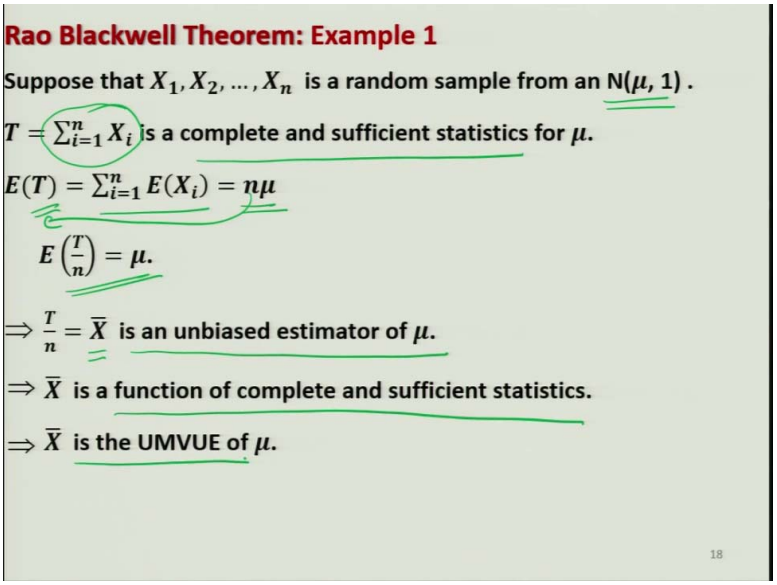
(Refer Slide Time: 34:41)



But definitely I am not going into detail because this will require some more concept which we have not done here like as convergence almost surely, but is still just for the sake of completeness of the lecture I am trying to give you here the definition of the completeness only. So, completeness ensures the uniqueness of estimators. So, now, if I say that, let there would be a random variable X of either continuous or discrete whatever you want having a proper PDF or probability mass function and that is a member of the family say $(x;\theta), \theta \in \Omega$ where Ω is the parameter space.

Now, if the the condition $E[T(X)] = 0$, for every $\theta \in \Omega$, requires that $T(x)$ be zero except on a set of points that has probability zero for each $h(x;\theta), \theta \in \Omega$, then the family $\{h(x;\theta), \theta \in \Omega\}$ is called a complete family of probability density or mass functions or probability mass functions. But anyway, I am not going into these details.

(Refer Slide Time: 35:43)

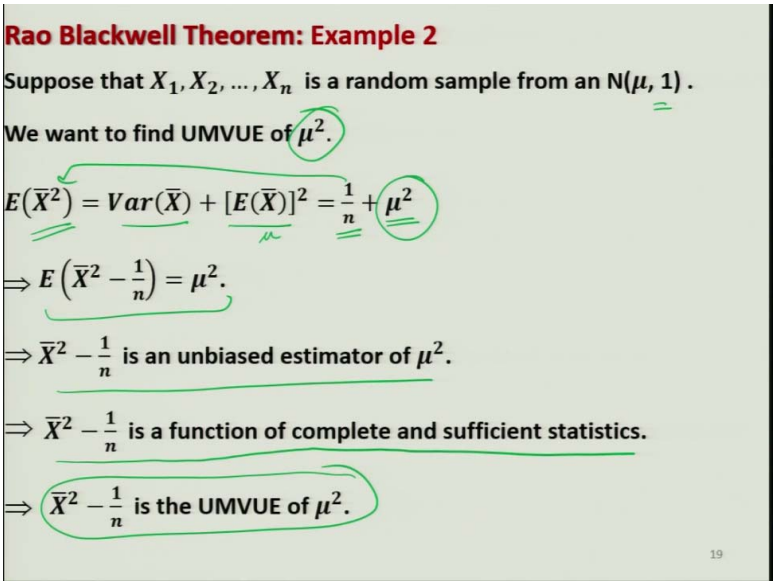


And now, I come to a business that that we try to take here some examples and try to show a show you that how easy it is to find the minimum radius and by estimator. So, the first example is let X_1, X_2,…, X_n be a random sample from a normal population with mean, μ and variance 1 and then you try to find out a sufficient statistics for μ, that we already have found, this was $\sum_{i=1}^{n} X_i$.

Now, you try to ensure that this is also a complete statistics, although, I am not doing it here, but you can believe on me that this $\sum_{i=1}^{n} X_i$ is a complete and sufficient statistics for μ. Now you simply try to take this expectation and try to convert it into an unbiased estimator of μ.

So, if you try to take your expected value of T this will come out to be $\sum_{i=1}^{n} E(X_i) = n\mu$ and if you try to bring this n here, then you can write down here that $E\left(\frac{T}{n}\right) = \mu$. So, this implies that T upon n is actually sample mean $\bar{X}$ this is an unbiased estimator of μ. This is a function of completed sufficient statistics. So, sample mean is the uniformly minimum variance and bias estimator of μ.

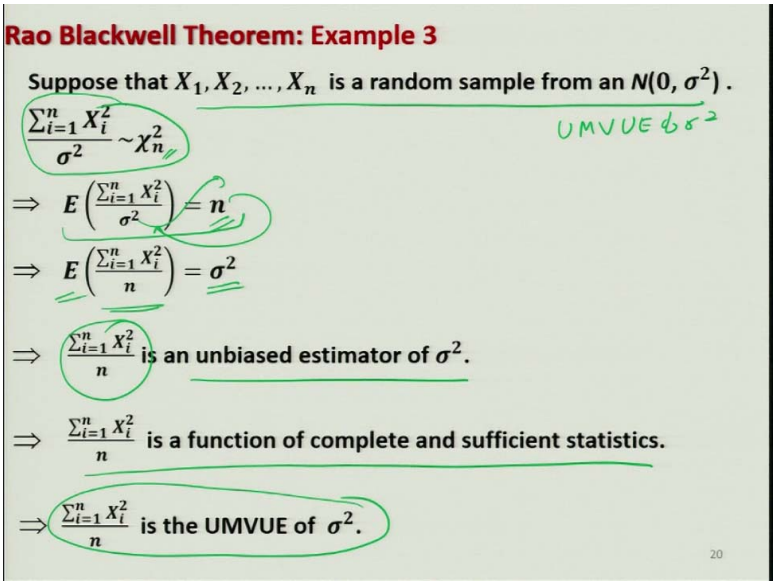
(Refer Slide Time: 36:59)



So, you can see here it was extremely simple to obtain the uniformly minimum variance and bias estimator. Now, in case if you suppose want to find out the uniformly minimum variance and bias estimator of μ squared in the same setup. What we try to do here that we try to find out the $E(\bar{X}^2) = Var(\bar{X}) + [E(\bar{X})]^2$. So, because here the variance is given here as 1 so the variance of X is $\frac{1}{n} + \mu^2$.

And you want an unbiased estimator of this μ^2. So, in case if you try to bring this 1 upon n on the left hand side, I can write down here that $E\left(\bar{X}^2 - \frac{1}{n}\right) = \mu^2$. So, I can now say that $\bar{X}^2 - \frac{1}{n}$ is an unbiased estimator of μ^2 and $\bar{X}^2 - \frac{1}{n}$ minus is a function of complete and sufficient statistics because, you already have proved that $\bar{X}$ is a complete and sufficient statistics for mu. And so, I can see that $\bar{X}^2 - \frac{1}{n}$ is the uniformly minimum variance and bias estimator for μ square. You can see that how it is.

(Refer Slide Time: 38:13)



Now, let me try to take here one more example, that suppose $X_1, X_2, \ldots, X_n$ is a random sample from normal with mean 0 and variance σ^2, σ^2 is unknown and we want to find out

UMVUE of σ^2. So, now, I try to use here a result that we have used many times from the properties of Chi-Square distribution that $\frac{\sum_{i=1}^n X_i^2}{\sigma^2} \sim \chi_n^2$.

And you know that the expected value of a chi-square random variable is the same as the degrees of freedom, and its variance is actually twice of n, but anyway, I try to use here this expectation so expected value of this chi-square variable is equal to its degrees of freedom n so, I can now write down here like this, that σ square goes here and comes here and I can write down here $\frac{\sum_{i=1}^n X_i^2}{n}$ as expected value is now same as σ^2.

So, I can say here that this quantity is an unbiased estimator of σ^2 and $\frac{\sum_{i=1}^n X_i^2}{n}$ is a function of complete and sufficient statistics. So, this is statistics is the uniformly minimum variance unbiased estimator of σ square. So, you can see here it was pretty simple to obtain such estimators.

So, now, we come to an end to this lecture. And I hope, you enjoyed this lecture because you see right from the beginning, our objective was to estimate the unknown parameters. And to reach to this point, we had to understand the basic fundamentals of probability theory and other things. Otherwise, how can you understand how can you find. Means, I have given you

for example, say here μ square in the case of normal μ 1, but if somebody ask me what will be the UMUVE of μ^3 or say μ^4, or something else.

How are you going to find? You are not going to come back to me to ask me that what is the estimator, but you have to depend on your own mind, you have to stand on your own feet, you have to depend on your own hands, that whatever is the function we can find it out. Well, I agree that these are the things where the things are very neat and clear. In case if you are working in the data science sometimes the things becomes complicated but surely, now, you have the basic concepts of all the properties and you can use these tools and data sciences and you can find out a good value of the unknown parameter without much difficulty I would say.

After that it depends on your experience, on your practice that how much practice you have made to find out the estimators of the unknown parameters under different times of conditions. So, I will stop here, but I will request you that you try to take example from the book, from the assignment and try to practice them. And I will see you in the next lecture with the new topic. Till then, good bye.