Essentials of Data Science with R Software – 1
Professor. Shalabh
Department of Mathematics & Statistics
Indian Institute of Technology Kanpur
Lecture No. 58
Cramer Rao Lower Bound and Efficiency of Estimators

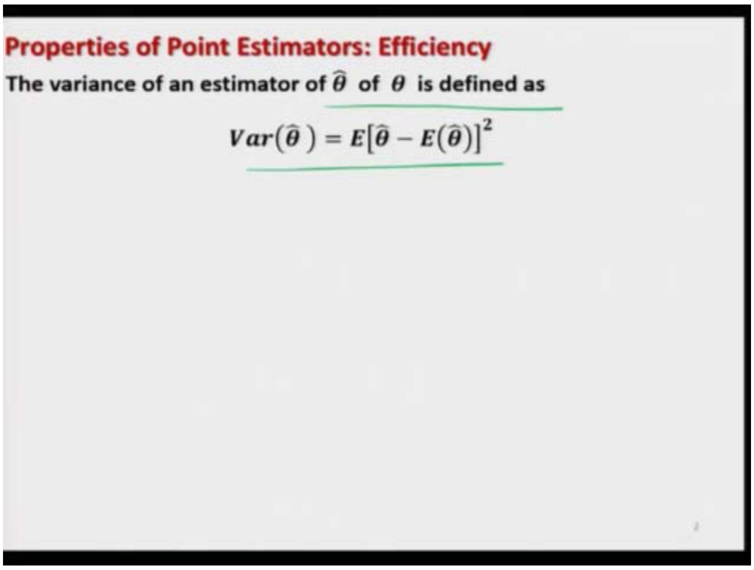
Hello friends welcome to the course essentials of Data Science with R Software – I in which we are trying to understand the basic concepts of probability theory and statistical inference. So, you can recall that in the last lecture, we had discussed about the concept of variability or efficiency of an estimator, and we have seen that this variance or MSE, they are going to help us in finding out an estimator whose variability is as small as possible.

But my next question is, you are trying to search for an estimator which has got the lowest variability, but what is this lowest variability. If you have estimator one, you get some variability in and now you get another estimator say estimator two whose variability is smaller than the first estimator. Now, somebody comes with a terrorist emitter whose variability is smaller than the variability of the first two estimator, so how long you will continue.

So, do not you think that you must also know that what is the lower bound of this variance by which the variability can be reduced. The variability will always be there in the real data set, but if you know that lower bound you know, that the variability of an estimator cannot go beyond this value. So, in very simple words, if you can have an estimator whose variability matches with the lower bound of the variance, then you should be happy that you have got the best estimator.

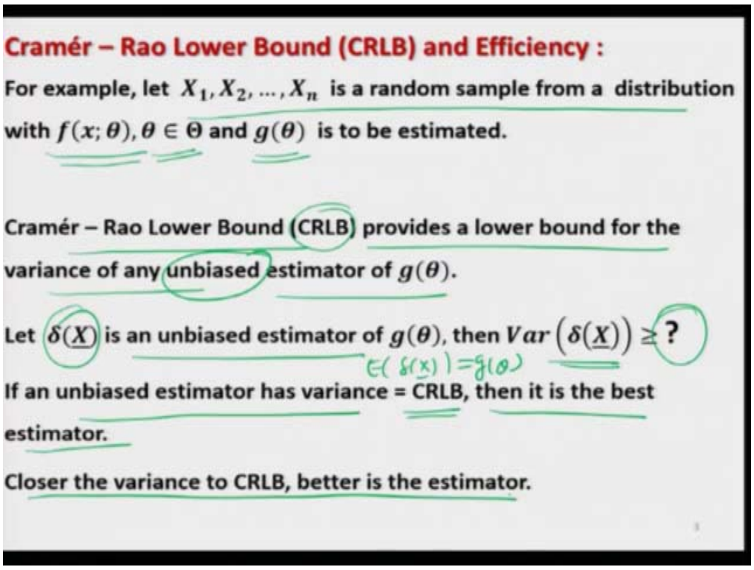
So, that is what we are trying to discuss in this lecture. And this lower bound can be obtained by the Crema Rao Lower Bound. The name of Rao is the name of Professor C.R. Rao, who just celebrated his hundredth birthday in September 2020. And, he is a living father of statistics. So, this was given by him. And, means, Professor Cramer is also there. So, both of them had they given this result, so, that is why this is called as Cramer Rao Inequality or Cramer Rao Lower Bound for the Variance. So, let us try to understand it, that is a very small topic.

(Refer Slide Time: 02:42)



So, let us begin. So, we know that the variance of an estimator $\hat{\theta}$ of θ is defined as say $E[\hat{\theta} - E(\hat{\theta})]^2$, that we know.

(Refer Slide Time: 02:52)

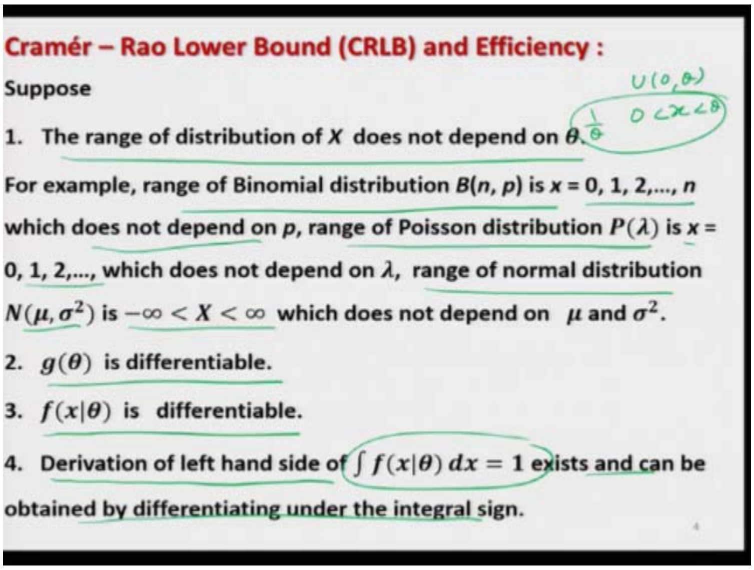


Now, I tried to first state what is the Cramer Rao Lower Bound and how it is related to efficiency and what is this result. So in order to use this Cramer Rao Lower Bound there are some conditions which have to be satisfied. So, now, we try to understand them one by one. So, let X_1, $X_2, \ldots, X_n$ be a random sample from a distribution with $f(x;\theta)$, some probability density function or probability mass function whatever you want $\theta \in \Theta$ and $g(\theta)$ is the parameter that we want to estimate. So, $g(\theta)$ can be θ or that can be a function of the parameters involved in it.

So, that Cramer Rao Lower Bound, which is briefly called a CRLB provides a lower bound for the variance of any unbiased estimator of $g(\theta)$. You had seen that in the earlier lectures, we have considered 3 or 4 possible estimator of the parameter μ from a normal distribution normal μ with $\sigma^2 = 1$, and then we also had found their variances. So, the simple question here is suppose $\delta(\underline{X})$ is an unbiased estimator of $g(\theta)$ that mean expected value of $\delta(\underline{X})$ is equal to $g(\theta)$.

Now, my question is this the variance of $\delta(\underline{X})$ will be greater than or equal to what that I do not know, and that is the answer which is given by this Cramer Rao Lower Bound. And so, obviously, that can be understood very easily that if an unbiased estimator has got a variance, which is equal to the Cramer Rao Lower Bound then it is the best estimator that can estimate the parameters $g(\theta)$. So, obviously now, the simple rule is closer the various to CRLB better is the estimator, and that is why this Cramer Rao Lower Bound is very popular when you are trying to work in real data science.

(Refer Slide Time: 04:48)



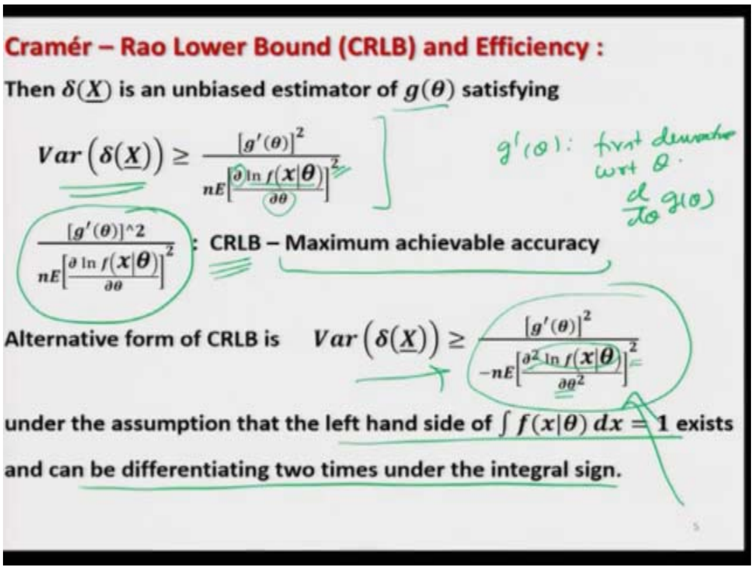
So, there are a couple of assumptions which are needed for the validity and the mathematical derivation of this Cramer Rao Lower Bound, although, I am not going to do here the mathematical derivation, but surely when you want to use it, you have to first make sure that these assumptions are valid. Usually people are not bothered about it. And when they try to use and if any of this assumption is not satisfied, surely that will give you the wrong result.

So, the first assumption is the range of the distribution of X does not depend on θ that is very simple. For example, if you try to see the range of the binomial distribution, B(n, p) is x = 0, 1, 2 up to n, which does not depend on p. Similarly, the range of Poisson distribution is like X goes from 0, 1, 2 … and this does not depend on the λ.

So, and similarly, if you try to see in the case of normal distribution also you have two parameters μ and σ^2, but the range of X is between $-\infty$ and $+\infty$, but this does not mean that all the estimators have got such a thing. An example is the uniform distribution say 0 to θ so in this case X lies between 0 and θ and the form of the PDF is $1/\theta$. So, you can see here in this case, the range is going to depend upon the parameter θ. So, we are not going to consider this type of situation.

So, the second assumption is that this $g(\theta)$ is differentiable and third assumption is this $f(x|\theta)$ is differentiable and fourth assumption is the derivation of left hand side of this quantity that is the $\int f(x|\theta)\, dx$ exists and can be obtained by differentiating under the integral sign. You know, that you cannot take the integral the differentiation sign inside the integral sign just as says there are certain rules which are governing it. So, we as we are not going into that detail, but we are assuming that all these conditions are going to hold true.

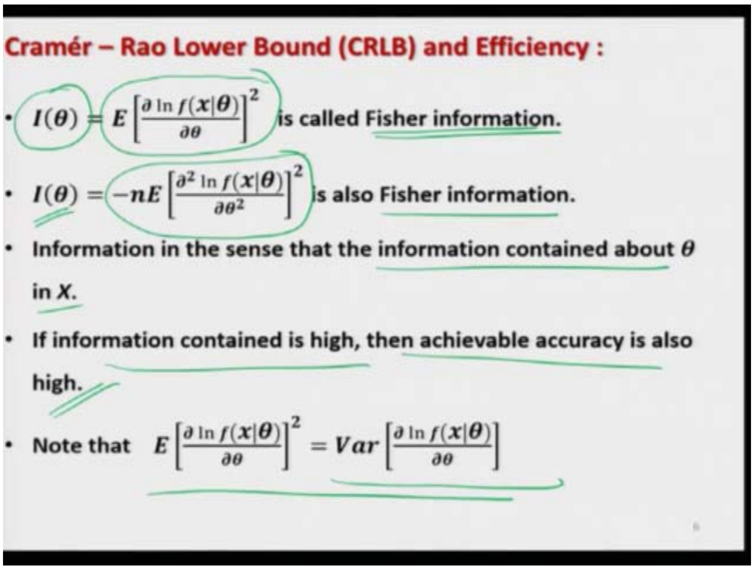
(Refer Slide Time: 06:58)



And believe me, with most of this standard probability density function and mass function, these conditions are generally true. So, now, the Cramer Rao Lower Bound is defined as follows. Then the variance of $\delta(\underline{X})$ which is an unbiased estimator of $g(\theta)$ is given by this quantity that is $\frac{[g'(\theta)]^2}{nE\left[\frac{\partial \ln f(\underline{x}|\theta)}{\partial\theta}\right]^2}$. So, this is $g'(\theta)$ is the first derivative with respect to θ. Something like what you say d by d θ of $g(\theta)$.

So, now, you can compute these quantities and this value on the right hand side which is here this is called as the Cramer Rao Lower Bound. And this indicates the maximum achievable accuracy. So, another form of this variance of $\delta(\underline{X})$ is given by this quantity where this denominator is given $\frac{[g'(\theta)]^2}{-nE\left[\frac{\partial^2 \ln f(\underline{x}|\theta)}{\partial\theta^2}\right]^2}$. And in both cases we are assuming that the left hand side of this $\int f(x|\theta)\, dx = 1$ exist and can be differentiated two times under the integral sign, otherwise, you will not get here the second partial derivative.

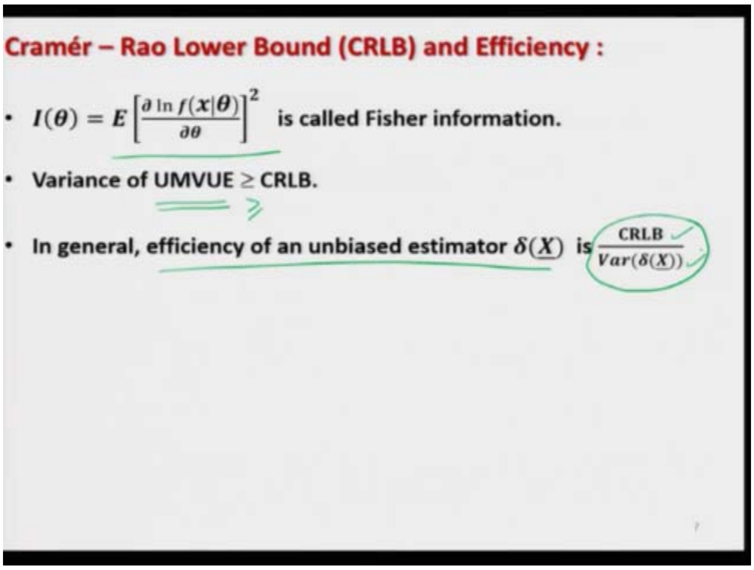
(Refer Slide Time: 08:50)



Now, in this form you can see here this function, $E\left[\frac{\partial \ln f(x|\theta)}{\partial\theta}\right]^2$, this is called as Fisher information and it is usually indicated by the function I(θ). And other form is given by like this, which is the same thing that was coming in the alternative form of the Cramer Rao Lower Bound here in this case.

So, now, this is also called as Fisher information and indicated by I(θ). So, this quantity is called as Fisher information in the sense because it is trying to give us the information that is contained about θ in X. So obviously, if the information contained is high, then the achievable accuracy is also going to be higher, and mathematically, you can see here that $E\left[\frac{\partial \ln f(x|\theta)}{\partial\theta}\right]^2 = Var\left[\frac{\partial \ln f(x|\theta)}{\partial\theta}\right]$ so two that you can see mathematically.

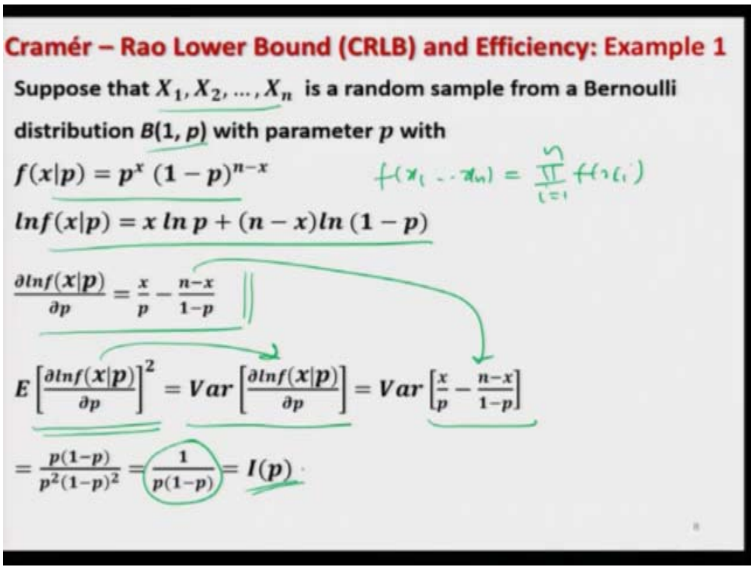
(Refer Slide Time: 10:05)



And so, now you can go find the Fisher information like this. So, now, what you can do here? You already have got the definition of minimum variance unbiased estimators uniformly minimum variance unbiased estimator. So, what will happen? That variance of uniformly minimum variance unbiased estimator will be greater than or equal to CRLB, Cramer Rao Lower Bound. And if this becomes exactly equal to Cramer Rao Lower Bound, then you have the best estimator.

So, in case if you want to compute the efficiency of an unbiased estimator $\delta(\underline{X})$, then it can be obtained using the concept of Cramer Rao Lower Bound as the ratio of Cramer Rao Lower Bound divided by its variance.

(Refer Slide Time: 11:00)

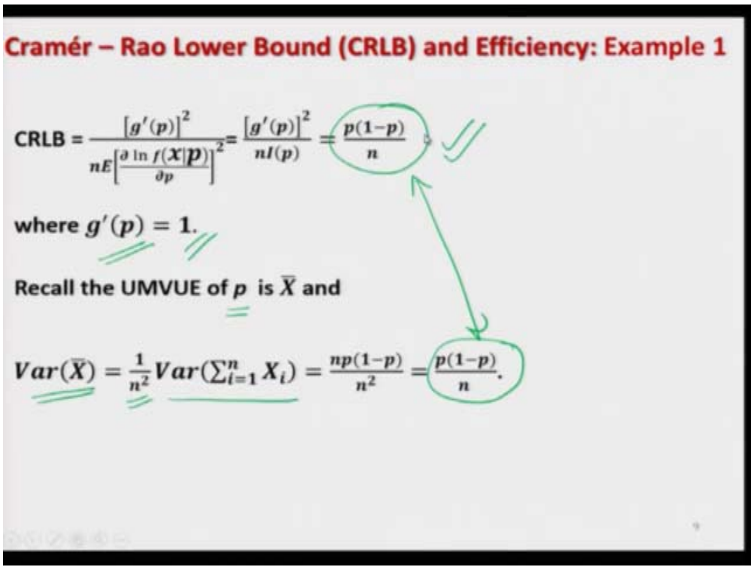


So, let me try to take here an example to show you that how these things can be computed, so, that you are confident that these are not very difficult thing. Suppose, we have what a random sample $X_1, X_2, \ldots, X_n$ from a Bernoulli distribution say binomial 1p and you can see here that if you want to find out the joint consecutive function of $X_1, X_2, \ldots, X_n$, then because they are independent so you can write down f of $X_1, X_2, \ldots, X_n$ as the product of this f (X_i), i goes from 1 to n from there I can write down here.

This joint function f x or a given value of p and now, if you try to take its natural log that is straightforward you try to differentiate it partially, and you will often hear this expression and then you try to square it and then try to take this expectation. This is the same as variance of this quantity inside the parenthesis and this is same as which you have obtained here variance of this quantity $\frac{x}{p} - \frac{n\ x}{1-p}$.

And if you try to compute this variance that can be obtained very easily this will turn out to be $\frac{1}{p(1-p)}$ and because of the information about the parameter p.

(Refer Slide Time: 12:24)

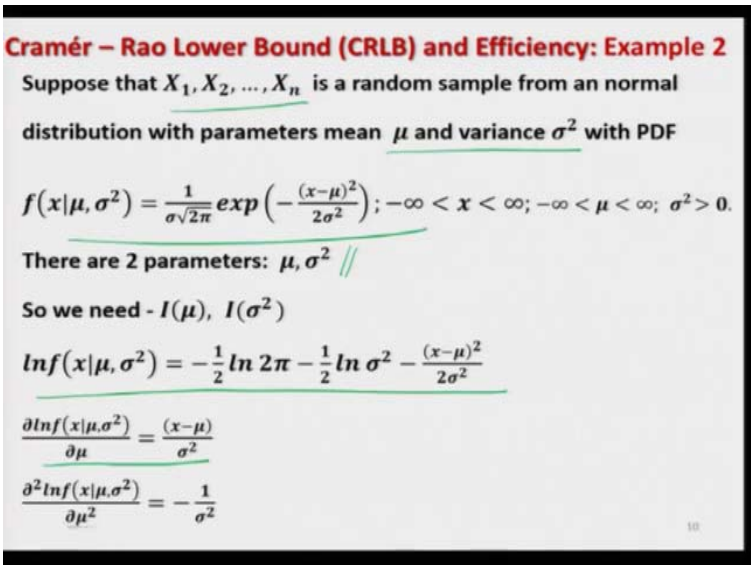


So, now, in case if you try to find out the Cramer Rao Lower Bound, this can be obtained just by substituting all these expressions and it will turn out to be p(1 – p)/n. And you can see here this the $g'(p) = 1$. And now, in case if you try to see here this is the Cramer Rao Lower Bound for this estimator.

And now, in case if you try to recall that we had obtained that small p can be estimated by the sample mean of $X_1, X_2, \ldots, X_n$ in the case of binomial distribution and in this case, if you try to find out the variance of $\bar{X}$, this will come out to be $1/n^2$ variance or summation of Xi and this can be obtained here as p(1 – p)/n and you can see here both the variances are the same.

That means, the Cramer Rao Lower Bound and the variance of $\bar{X}$ they are the same. So, now, you can be confident that you can estimate the population proportion of a binomial distribution by this sample mean without any problem, and that is going to give you a good value.

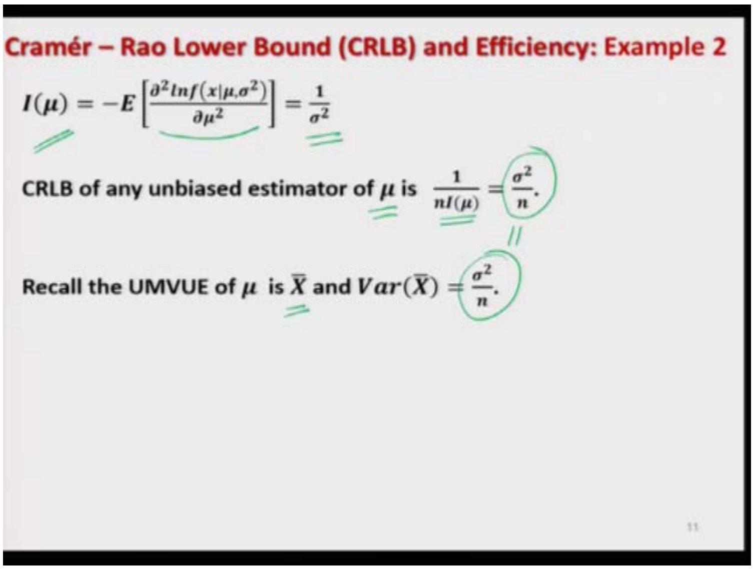
(Refer Slide Time: 13:34)



And similarly, if you try to take one more example, from the continuous random variable, so, let $X_1, X_2, \ldots, X_n$ be a random sample from normal distribution with the mean, μ and variance σ^2 whose PDF is given by like this that you know. So now, we have here two parameters μ and σ^2 so we need to find out the Cramer Rao Lower Bound for μ and σ^2 both.

So, we need to do the same thing. So, if you try to find out their joint density function, and then you try to take its log and then you try to differentiate it first with respect to μ and then with respect to μ square you will get here these quantities. Well, I am not solving it here because they are pretty straightforward thing.

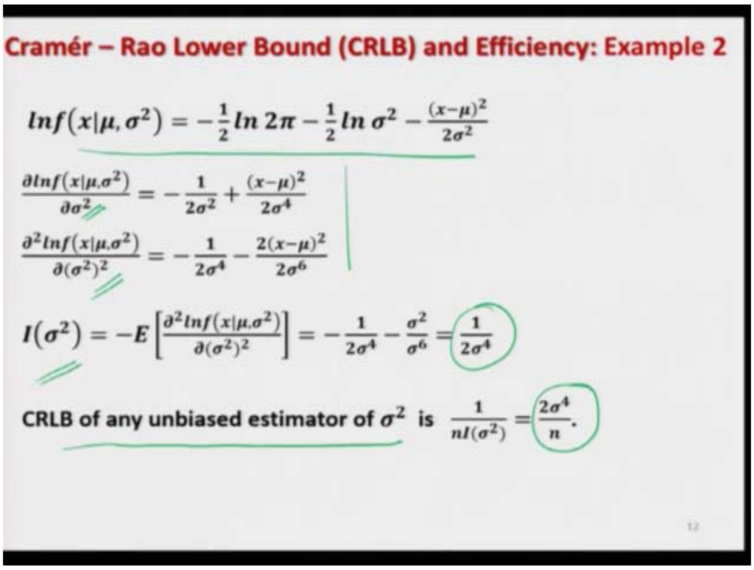
(Refer Slide Time: 14:17)



And then if you try to find out the value of information based on this mu, this will be simply the minus of expectation of this quantity this comes out to be here $1/\sigma^2$. So, the Cramer Rao Lower Bound for any unbiased estimator of μ is coming out to be 1/n times information which is σ^2 /n.

And now, if you try to recall earlier you had estimated the value of μ by $\bar{X}$ and whose variance was coming out to be σ^2/n that means the variance of $\bar{X}$ and Cramer Rao Lower Bound they are matching. So, that means if you want to estimate the parameters μ of the normal population then $\bar{X}$ is the uniformly minimum variance and by estimator of μ.

(Refer Slide Time: 14:58)



And similarly, if you try to do the same exercise for the σ^2, you try to write down the log of this $f(x|\mu,\sigma^2)$ then you try to differentiate it with respect to σ^2 then you try to take the second partial derivative, you will get here these expressions. And based on that if you try to substitute these expressions in the value of $I(\sigma^2)$, you will get here the value here $\frac{1}{2\sigma^4}$. So, the Cramer Rao Lower Bound of any unbiased estimator of σ^2 will come out to be $\frac{2\sigma^4}{n}$.

And now, I come to an end to this lecture, and you can see here this was a very short lecture, but very informative. Now you have got the idea that okay somebody can propose any estimator for estimating there any parameter. And suppose this is an unbiased estimator then in that case, you can compute the Cramer Rao Lower Bound and can judge whether the values which you are trying to obtain for an unknown parameter are good or bad.

So, I will request you try to take some more examples from the book, try to solve them mathematically. These are very simple exercises and I will see you in the next lecture with next topic. Till then, goodbye