Estimation for Wireless Communications-MIMO/OFDM Cellular and Sensor Networks.

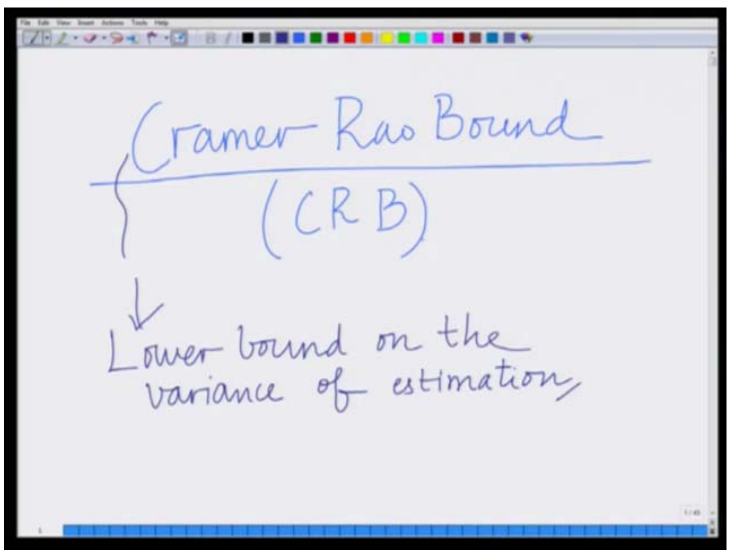
Professor Aditya K Jagannatham.
Department of Electrical Engineering.
Indian Institute of Technology Kanpur.

Lecture -11.

Cramer Rao Bound (CRB) For Parameter Estimation.

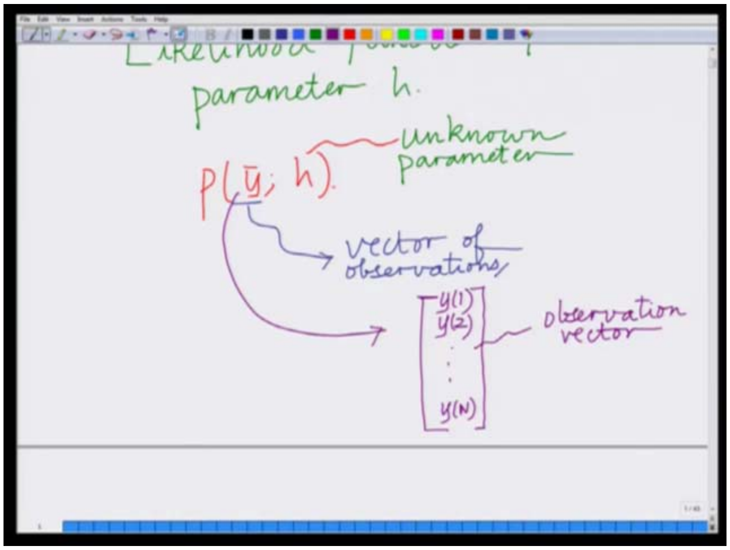
Hello, welcome to another module in this massive open online course on estimation for wireless communications. So, so far we have looked at the maximum likelihood estimation, both in the context of sensor network and also in the context of a wireless communication system. Where we looked at channel, maximum likelihood channel estimation for a wireless communication system. Let us now look at something analytical and a bit more fundamental, that is we are going to talk about the Cramer Rao lower Bound.

(Refer Slide Time: 0:44)



So, what we are going to talk about today is known as the Cramer Rao bound and its also abbreviated as CRB. And what is the Cramer Rao bound, the Cramer Rao bound, represents a fundamental lower bound on the variance of an estimate. So, the CRB gives a convenient way to characterise the performance of an estimator. What it gives is the best achievable performance of an estimator, that is the lowest possible variance that can be achieved by an estimator. So, it is a lower bound on the variance achievable by an estimator. So, it yields, so the Cramer Rao bound is a lower bound on the variance of estimation. Or in other words that the variance of any estimator has to be greater than this, the Cramer Rao lower bound.

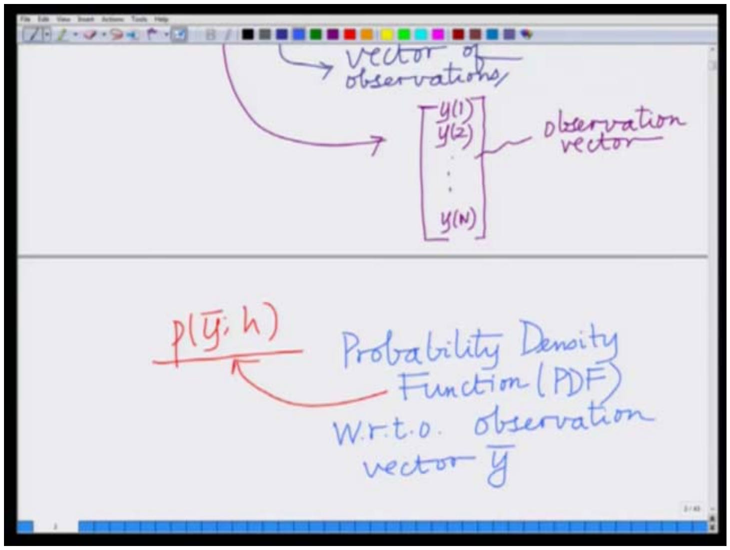
(Refer Slide Time: 2:00)



So, to derive the Cramer Rao bound, let us start with the likelihood function corresponding to the parameter H. Remember, we have already looked at the likelihood function for the unknown parameter H and this likelihood function is denoted by P of Y bar parameterised by H where H is your, H is the unknown parameter and Y bar is your observation vector or vector of observations. So, to derive this Cramer Rao bound, which is the lower bound on the variance of any estimator, we start as usual with something that is very fundamental to the context of estimation, that is the likelihood function. The likelihood function as we are all very familiar with by now, it is denoted by P Y bar; H where H denotes the unknown parameter and Y bar denotes the vector of observations, remember Y bar is the N dimensional vector, we have N observations Y1, Y2… YN and Y bar is this vector Y1, Y2… YN.

So, Y bar is this vector, so again just to refresh your memory, Y bar is the observation vector Y1, Y2… Up to YN, this is the, this is the observation vector Y bar which is Y1, Y2… Up to YN.

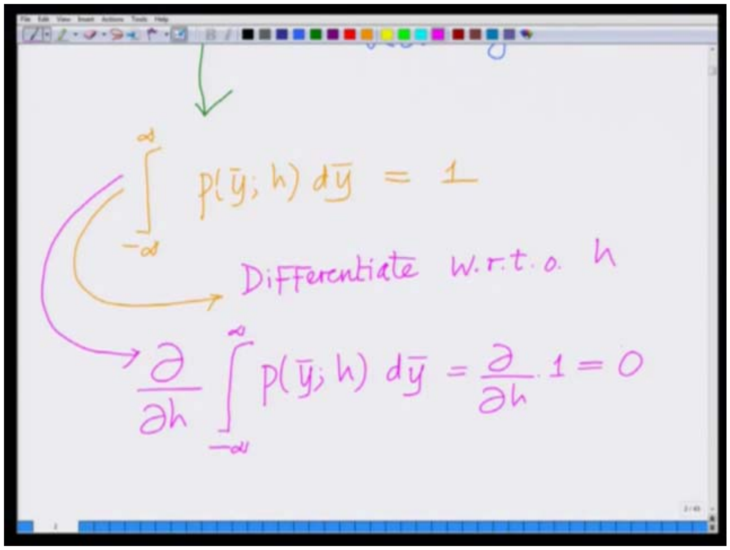
(Refer Slide Time: 4:01)



Now also recall that this is observed likelihood function P Y bar H has a dual role. Remember the way derived it, this is a likelihood function with respect, this is the probability density function of the observations Y1. So this is the probability and this is an important point, this is the probability density function that is the PDF with respect to with respect to the observations or the observation vector with respect to the observation vector Y bar. So, recall that this likelihood function is nothing but the probability density function of the observations Y1, Y2… Up to YN, that is the joint probably the density function of the function of the observations Y1, Y2… Up to YN parameterised by the unknown parameter H.

And when we view it as a function of the unknown parameter H, this is a likelihood function. Therefore, the likelihood function is also probability density function, hence naturally, since the integral of the probability density function is 1, this probability density function must integrate to 1.

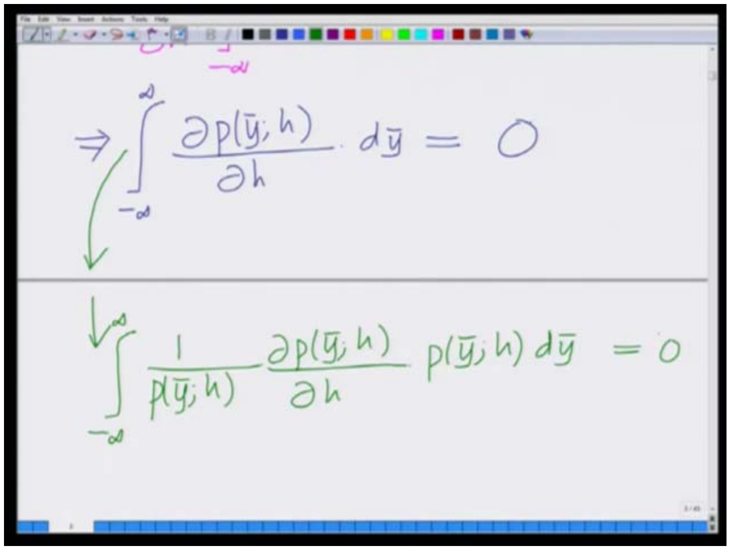
(Refer Slide Time: 5:32)



Therefore, since this likelihood function is also a probability density function with respect to the observations, we must have integral - infinity to infinity, it naturally follows that integral - infinity to infinity P Y bar H dY equal to 1, dY bar is equal to 1. That is integral of this probability density function is equal to 1.

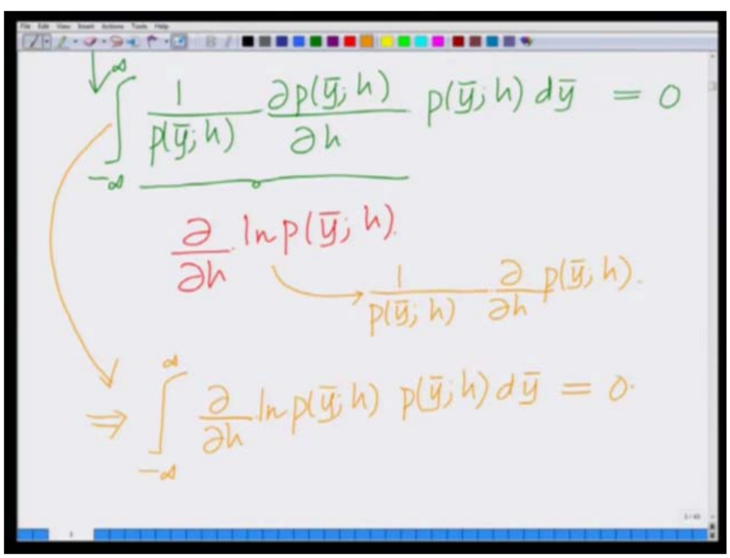
Now differentiating this with respect to H, what we do now if we differentiate this with respect to H, differentiate with respect to H and therefore what we have a dow by dow H of integral - infinity to infinity PY bar parameterised by H dY bar equals dow by dow H, derivative of the right inside is a derivative of 1 with respect to H but the derivative of this constant 1 with respect to H is 0, which means the partial derivatives of the quantity on the left with respect to H is 0.

(Refer Slide Time: 6:46)



Now we move this derivative operation inside, this implies basically we are moving this inside, what we have a is integral - infinity to infinity dow the probability density function, the partial derivative of the probability density function with respect to an unknown parameter H times dY bar is equal to 0. That is what we have. And now observe something, I can multiply and divide by this probability density function PY bar parameterised by H, so I am multiplying and dividing and now multiplying by the probability density function of Y bar parameterised by H which is equal to 0.

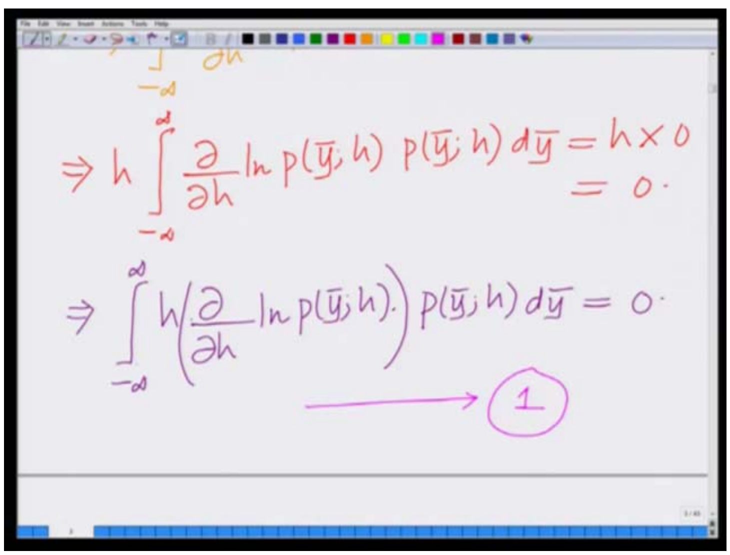
(Refer Slide Time: 7:54)



Now if you observe this quantity, this quantity is nothing but the partial derivative of the log likelihood function. This is dow by dow H log of, the log, the natural logarithm of the probability density function of Y bar parameterised by H because the probability density function of the log likelihood, that is the probability, that is the derivative, partial derivative of the logarithm of of PY bar parameterised by H is 1 over PY bar parameterised by H times the derivative of PY bar parameterised by H. Alright. So, this is nothing but again just to rehash the same thing, this is nothing but basically 1 over PY bar weight of the log likelihood is 1 over P bar PY bar parameterised by H times the derivative of P Y bar parameterised by H with respect to H and that is what I have.

Therefore now I can write, this implies that - infinity to infinity, the derivative of the log likelihood PY by parameterised by H times PY bar parameterised by H dY bar equals equals 0.

(Refer Slide Time: 9:30)

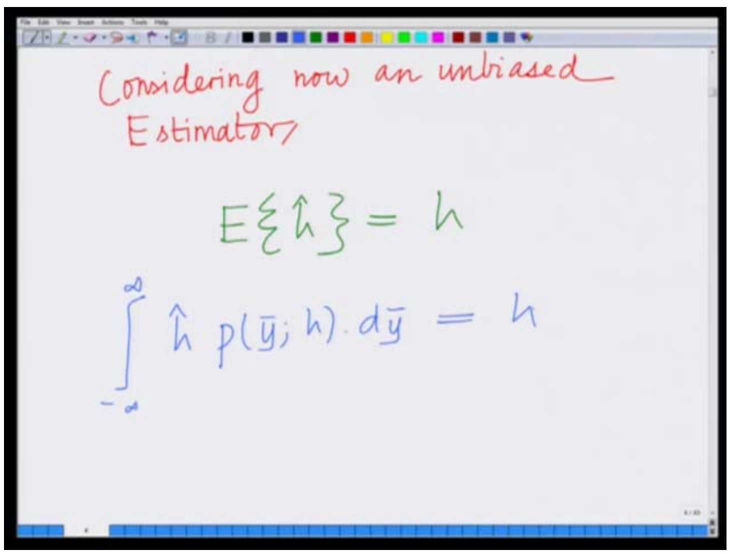


And now I can multiply by H on both sides, so multiplying by the unknown parameter H on both sides I have - infinity to infinity, integral - infinity to infinity the derivative of the log likelihood PY bar parameterised by H times PY bar parameterised by H dY bar is H times 0 which is again 0. And now finally moving H inside, I remember I can move the integral H inside the integral because the integral is with respect to Y bar, so I can move H inside the integral dow by dow H, the logarithm of PY bar H times PY bar H dY bar equal to 0.

And therefore let us call this result as your result number 1. So, I am calling this as basically the result number 1, what I have is integral - infinity to infinity H times that the derivative of the log likelihood function log PY bar parameterised by H times PY bar parameterised by H dY bar is basically equal to 0. This integral on the left is equal to 0. And we have derived this property which we are subsequently going to employer in deriving the Cramer Rao rule or the Cramer Rao bound for the variance of estimation of the parameter.

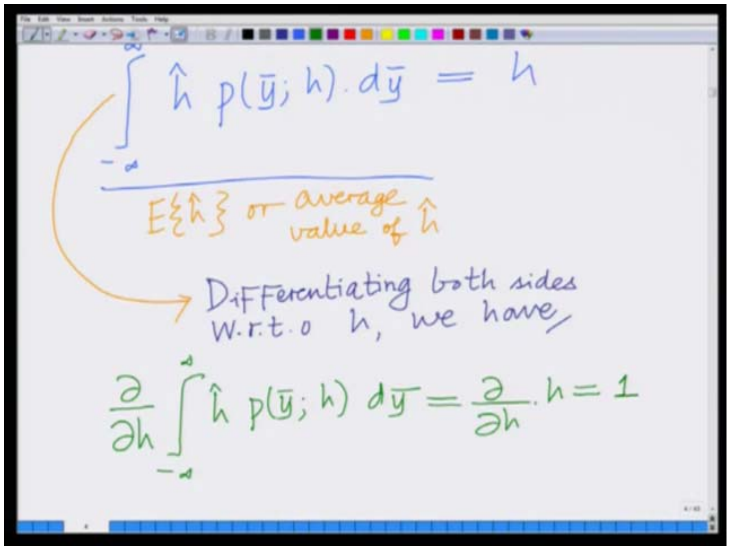
Alright, now let us also consider an unbiased estimator of the unknown parameter H which means the expected value of the estimator is expected value of the estimate H hat is always H, that is what we have seen in the previous modules. That is the unbiased estimator is one such that the **est** expected value of the average value of the estimate is equal to the true value of the unknown parameter H.

(Refer Slide Time: 11:40)



So, considering now an unbiased estimator, considering now an unbiased estimator, what I have is that the expected value of H hat, that is the average value of the unknown of the estimate is is always equal to the true value of the underlying unknown parameter H. Which means writing this in mathematical terms, this means the expected value is nothing but the expected value of the quantity is nothing but the quantity times the probability density function times dY bar which we are saying is equal to H.

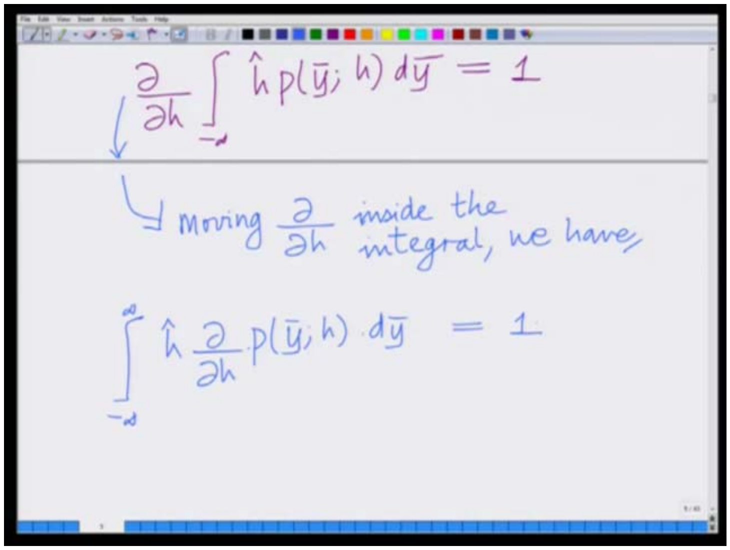
(Refer Slide Time: 12:41)



What is this, this quantity here is basically the expected value of H hat multiplied by the probability density function and integrated between - infinity to infinity, therefore this is nothing but expected value of H hat or basically your average value of H hat. The average value of the average value of the average value of the unknown average value of the estimate H hat.

Now again differentiating this with respect to H, so what we are going to do again, we are considering an unbiased estimator, that is the expert could value of H hat is equal to H which means they integral - infinity to infinity H hat times the probability density function PY bar parameterised by H dY bar is equal to H. Now we differentiate on both sides with respect to H. Of course when we differentiate on the right with respect to H, that derivative is one because that derivative of H with respect to H is simply one. So, basically let us differentiate now both sides with respect to H, differentiating both sides with respect to H, what we have is basically dow by dow H, the partial with respect to H, - infinity to infinity H hat PY bar H dY bar and differentiating now on the right what we have a is dow by dow H of H but the derivative of H with respect to H equals 1.

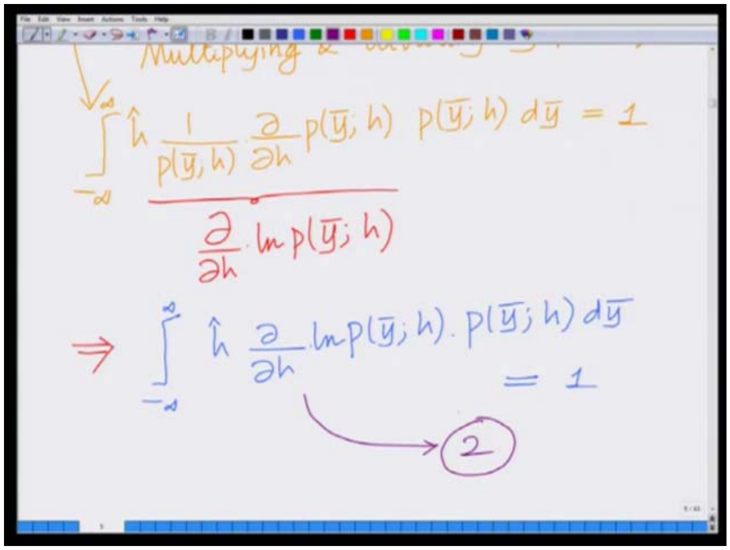
(Refer Slide Time: 14:43)



Therefore to put it simply, what we have is dow by dow H or is dow by dow H integral - infinity to infinity H hat PY bar parameterised by H dY bar equals 1. Now moving the derivative inside the integral, so basically now moving dow by dow H inside the integral, what we have is basically, look at this, this is H hat times dow by dow H PY bar parameterised by H dY bar equals 1. Where when we move the derivative inside the integral, we do not need to consider the derivative of H hat with respect to H because H hat is the estimator and estimator depends only on the observations Y bar. So, it does not depend on the unknown parameter H.

So, we are moving the derivative dow by dow H directly to P that is the probability density function PY bar parameterised by H and now you can see I can once again multiply and divide by PY bar.

(Refer Slide Time: 16:13)

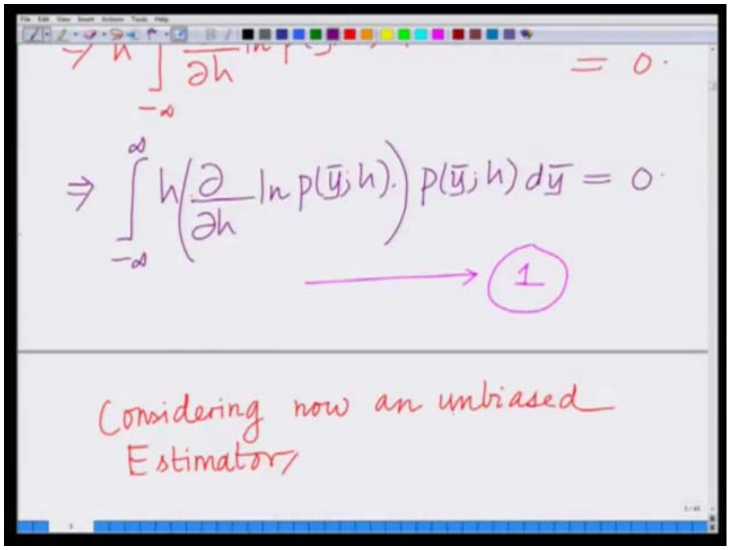


So now multiplying and dividing, multiplying and dividing by PY bar parameterised by H, I have integral - infinity to infinity H hat 1 over 1st I am dividing with respect to the with respect to the probability density function dow by dow H PY bar H, now I multiplying the probability density function PY bar parameterised by H dY bar equals 1.

And now if you look at once again, similar to what we have done before, if you look at this quantity 1 over PY bar parameterised by H times dow by dow H of PY bar parameterised by H, this is nothing but the partial derivative of basically the log likelihood function, partial derivative of the log of the probability density function of the observations Y bar parameterised by H, therefore this implies basically that - infinity to infinity H hat dow by dow H log P of Y bar parameterised by H times P of Y bar parameterised by H dY bar is equal to 1 and let us call this as a result number 2.

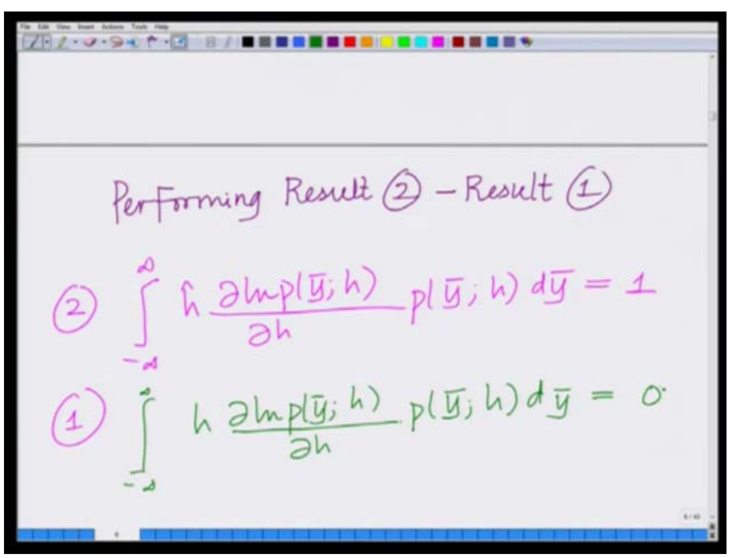
So, what I am denoting this, so I have derived 2 results basically, this is your result number 2.

(Refer Slide Time: 18:17)



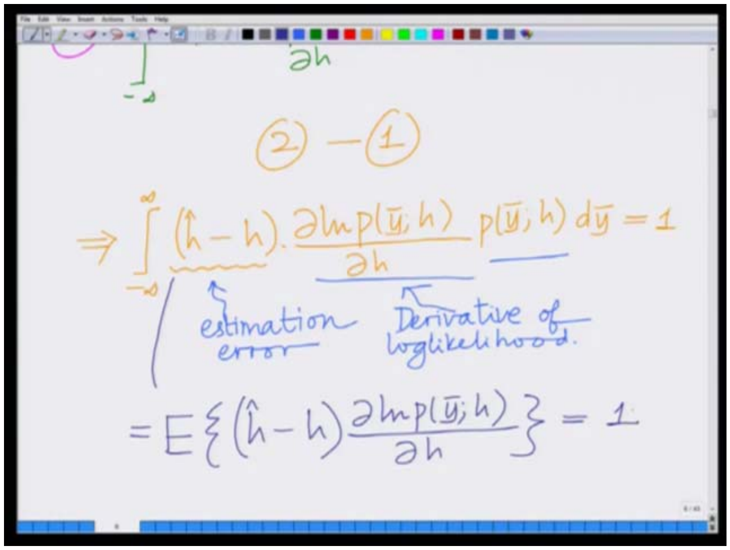
So, previously we have derived previously 1 and now we have derived the result number 2 and now let us subtract 1 from 2. So let us look at these 2 results, we have result number 1 over here and we have result number 2 over here and now performing result number 2 - result 1, what we have, remember, what we have, let us write again for the sake of convenience, I think it would be better if we write both the results, result number 2 is basically integral - infinity to infinity H hat dow ln PY bar H by dow H PY bar H dY bar equals 1 and result 1 is basically that integral - infinity to infinity H times dow ln PY bar H divided by dow H PY bar H dY bar is equal to 0. In fact is equal to 0.

(Refer Slide Time: 18:35)



Now, performing 2-1, now you can clearly see when I make result 2 - result 1 that implies,

(Refer Slide Time: 19:56)

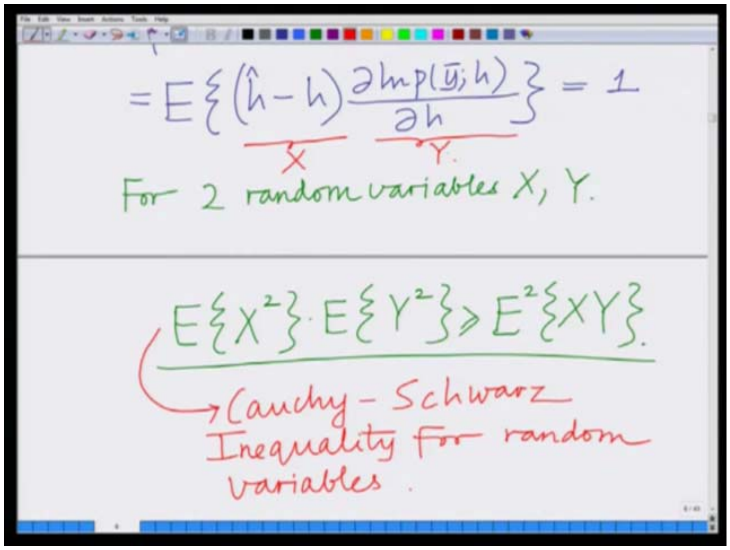


clearly implies, what does it imply, this implies now you can clearly see this implies H hat - H times the derivative of the log likelihood function with respect to H times the probability density function of Y bar parameterised by H or basically the likelihood function with respect to H dY bar is equal to 1. All right and this is the central result, and if you look at this quantity, H hat - H, this is nothing but your estimation error, right.

So, what we have here is basically H hat - H which is basically the estimation error, so we have derived interesting result for the estimation error and now if you look at this, this is basically the estimation error, this is the derivative of the log likelihood, what is this, this is the derivative of your, this is the derivative of your log likelihood, now remember we are multiplying this by the probability density of the probability density function of the observation vector Y bar which means this is basically nothing but the expected value. Look at this, this is basically equal to the expected value of the error times the derivative of the log likelihood function with respect where H. And we are saying this expected value of this product is equal to 1.

So, take a look at this again and this requires some understanding and some clear thinking, all right. So, please over this derivation again, what we have derived is we have derived interesting result which states that they, the expected value of the product of this estimation error H hat - H times the derivative of the log likelihood function of the parameter H with respect to H is equal to 1, that is excited value of this product is equal to 1.

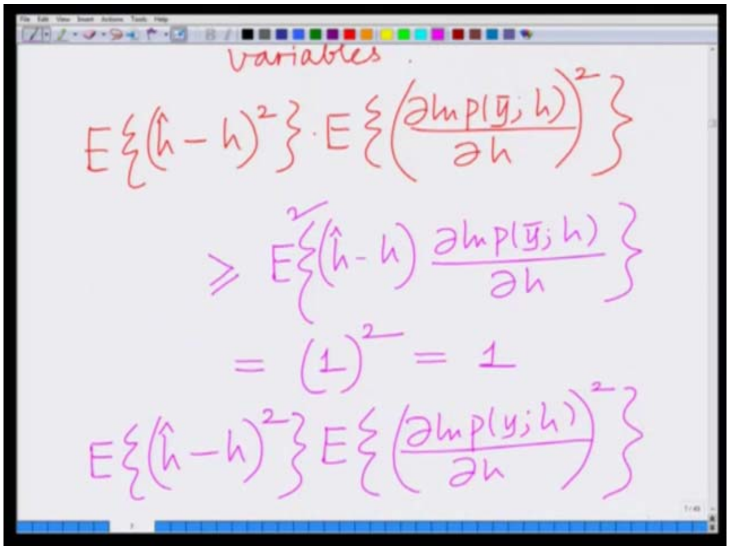
(Refer Slide Time: 22: 31)



And now therefore we have, we can now use the Cauchy Schwarz inequality which basically states that the expected value of XY is greater than or equal to the expected value of or the expected value of which states that if you have 2 random variables, for any 2 random variables, for 2 random variables XY, X, Y, it must be that the expected value, the square of the expected value of the product XY is greater than or basically they the product expected value of X square times expected value of Y square must be greater than the square of the expected value of the product XY and therefore now what. Now we can use this interesting result, this is basically your Cauchy Schwarz inequality for random variables.

This is your, this is the Cauchy Schwarz inequality, this is the Cauchy Schwarz inequality for random variables and now therefore we can use this Cauchy Schwarz inequality on this product by treating this H hat - H by treating this random variable H hat - H as a random variable X by treating this derivative of the log likelihood function as a random variable Y

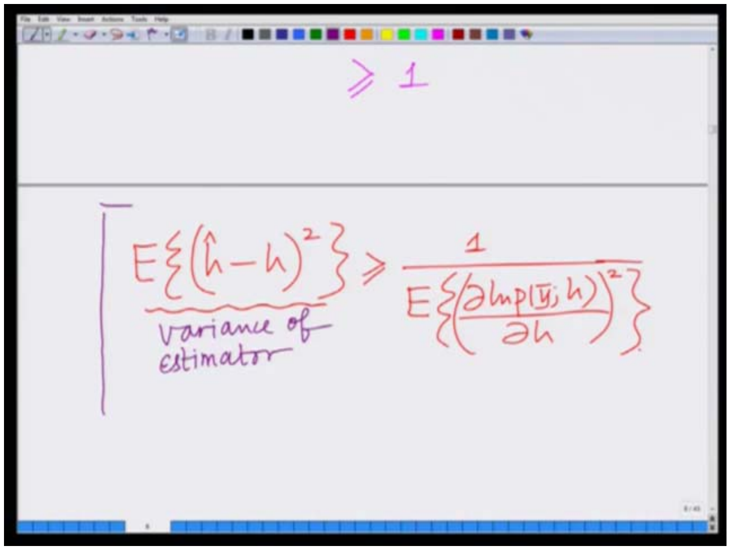
(Refer Slide Time: 24:29)



and therefore we have expected H hat - H whole square times expected derivative of the log likelihood function with respect to H whole square is greater than or equal to, if greater than or equal to the expected value of H hat - H square of the expected value of H hat - H times the derivative of the log likelihood function which is basically equal to the product. The expected value of the product is 1, so this is basically greater than or equal to 1 square which is equal to 1.

Therefore now we can write the expected value of H hat - H whole square times the derivative of the log likelihood function expected value of H hat - H whole square times expected value of dow derivative of the log likelihood function whole square, expected value of the derivative of the log likelihood function whole square is greater than or equal to 1.

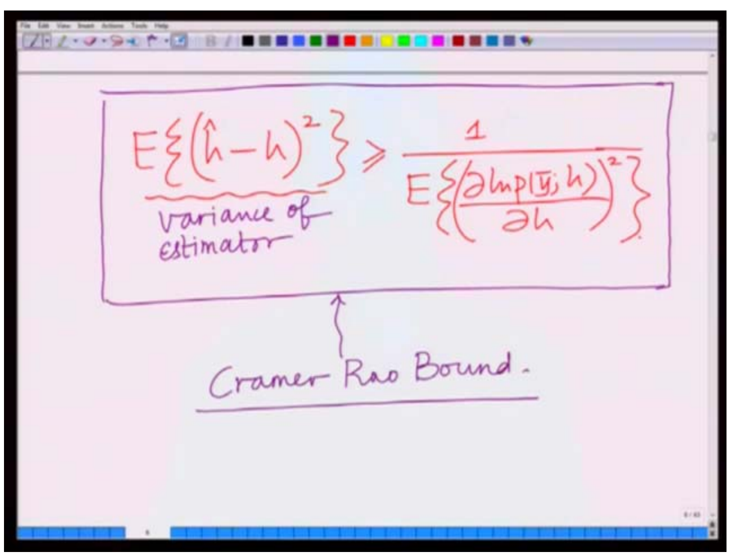
(Refer Slide Time: 26:15)



Therefore now moving this quantity to the right, we have expected value of H hat - H whole square is greater than or equal to 1 over the expected value of the square of the derivative of the log likelihood function and this is basically… Now you can see this on the left is basically your variance of the estimator that is the square of the deviation and this therefore, it gives a fundamental bound on the variance, this is therefore your Cramer Rao bound.

So, it says that the variance of this estimator…

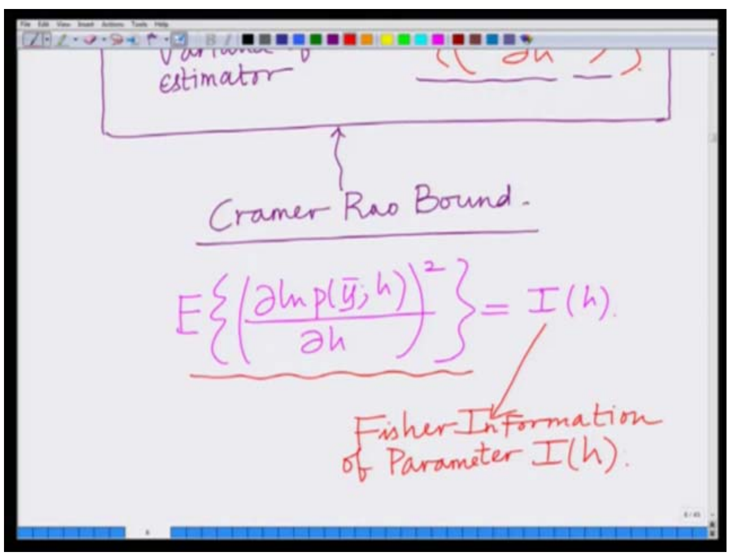
(Refer Slide Time: 27:12)



So this is basically nothing but, this is basically your Cramer Rao, **…** So, what we have derived, if we have derived this result which states that the expected value of H hat - H square, that is basically the variance of this estimator and that is for any particular estimator as long as it is unbiased, that is the variance of any particular estimate H hat, alright, not necessarily the maximum likelihood estimate but any particular estimate H hat as long as it is unbiased, the variance is always greater than or equal to 1 over the expected value of the square of the derivative of the log likelihood function of the parameter H with respect to H.

All right and this is the fundamental bound on the variance of any estimator and this is known as the, fundamental bound and the variance of any unbiased estimator for that matter, speaking more precisely and exactly, and this fundamental bound on the variance of any unbiased estimator is known as the Cramer Rao bound or more explicitly the Cramer Rao lower bound. Alright, it is automatically understood that it is the lower bound for the variance of estimation therefore this is also known as the Cramer Rao lower bound or the Cramer Rao bound. And as we have already said, this is abbreviated as CRB. This is the Cramer Rao bound.

(Refer Slide Time: 28:49)

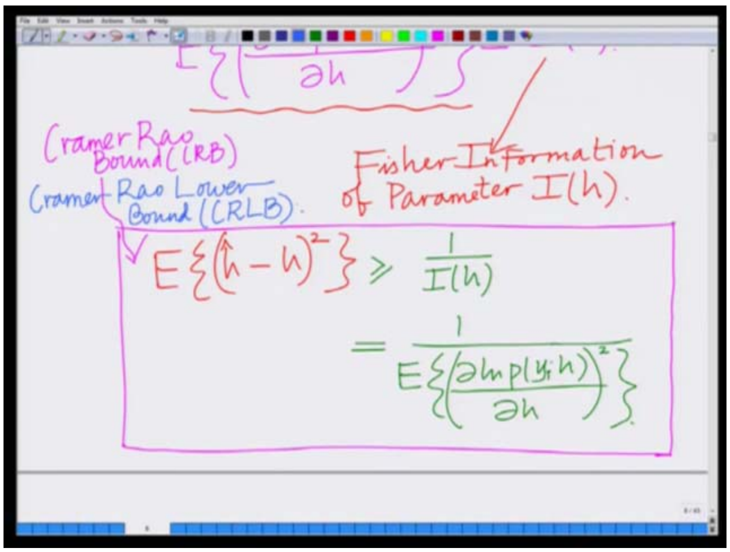


And look at this, this quantity, if you look at this quantity, it is interesting, if you look at this quantity in the denominator, this quantity, the expected value of the square of the derivative of the log likelihood function, this is known as the Fisher information I of H. This is known as the Fisher information I of the parameter H, so this basically this quantity is denoted by the Fisher information.

The Fisher information of the parameter, the Fisher information of the parameter H, in some sense this basically quantifies the information that the log likelihood function provides about the unknown parameter H. And therefore the larger the Fisher information, naturally the lower is the estimation variance, that is basically the larger the amount of information that your log likelihood function provide with respect to H, the larger is the Fisher information and therefore the estimation variance because the information provided is higher, we expect the variance, the variance of estimation to be lower and that is indeed what is reflected in this Cramer Rao lower bound.

That is the variance, the minimum variance, the lower bound on the variance of estimation is basically given by the inverse of the Fisher information.

(Refer Slide Time: 30:34)



So, we can also write this as basically expected value of H hat - H whole square is greater than or equal to the inverse of the Fisher information that is 1 over E the derivative of the square, average value of the square of the derivative of the log likelihood function. And this is basically your Fisher information. This is basically the result for the Cramer Rao bound. This is basically your, let me again write this explicitly, this is again the Cramer Rao bound for parameter estimation or also the Cramer Rao, the Cramer Rao, the Cramer Rao bound or basically the Cramer Rao lower bound for the estimation of the parameter H.

So, what we have derived today is something very fundamental, we have considered basically an unbiased estimator, H hat for any unknown parameter H and we have derived a fundamental or bound on the variance of estimation for this parameter H and we demonstrate and this we have said, this fundamental bound on the variance of any unbiased estimator is given by the Cramer Rao bound or the Cramer Rao lower bound, we have derived an expression for this Cramer Rao lower bound and we have demonstrated that the estimation variance, the average value, that is the expected or average value of H hat - H whole square is greater than or equal to 1 over the expected value of the square of 1 over the expected value average value of the square of that derivative of the log likelihood function of H with respect to H and this basically, this expected value the square of the derivative of the log likelihood function is basically also the, is basically also the Fisher information of the unknown parameter H.

Alright, so the Cramer Rao lower bound represents yields a fundamental bound on the, fundamental lower bound on the variance of any unbiased estimator H hat of any parameter H. So, will stop this module here and we will explore other aspects of this and in fact other applications and examples of the Cramer Rao lower bound in the subsequent modules. Thank you very much.