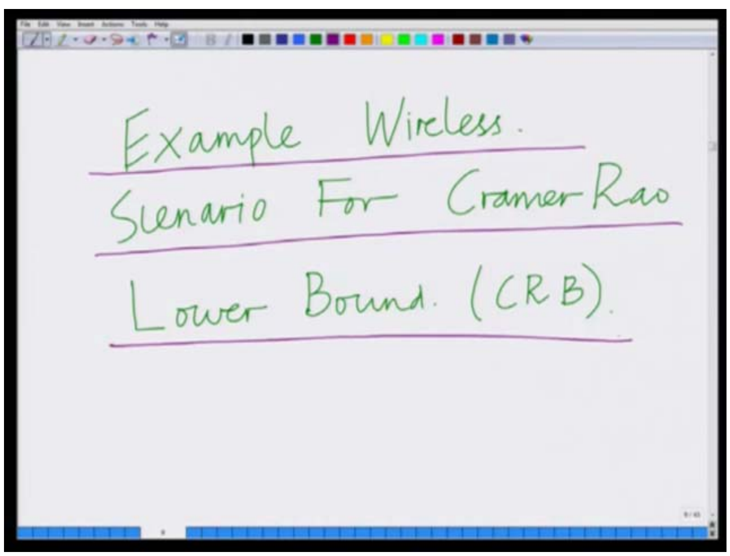
Estimation for Wireless Communications-MIMO/OFDM Cellular and Sensor Networks.
Professor Aditya K Jagannatham.
Department of Electrical Engineering.
Indian Institute of Technology Kanpur.
Lecture -12.
Cramer Rao Bound (CRB) Example - Wireless Sensor Network.

Hello, welcome to another module in this massive open online course on estimation for wireless communication systems. In the previous module we have looked at the Cramer Rao bound which provides a fundamental lower bound on the variance of an unbiased estimator. Right, we have derived the expression for the Cramer Rao lower bound which characterises the lowest possible variance that can be achieved by an estimator. Alright.

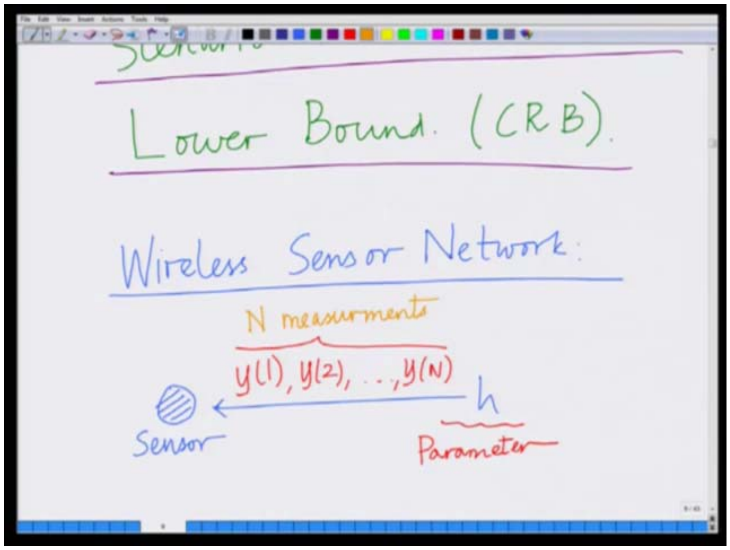
So, now in this module let us try to look at an example to understand this computation of the Cramer Rao lower bound better. So, let us look at an example in the context of a wireless communication system to compute the Cramer Rao lower bound.

(Refer Slide Time: 0:49)



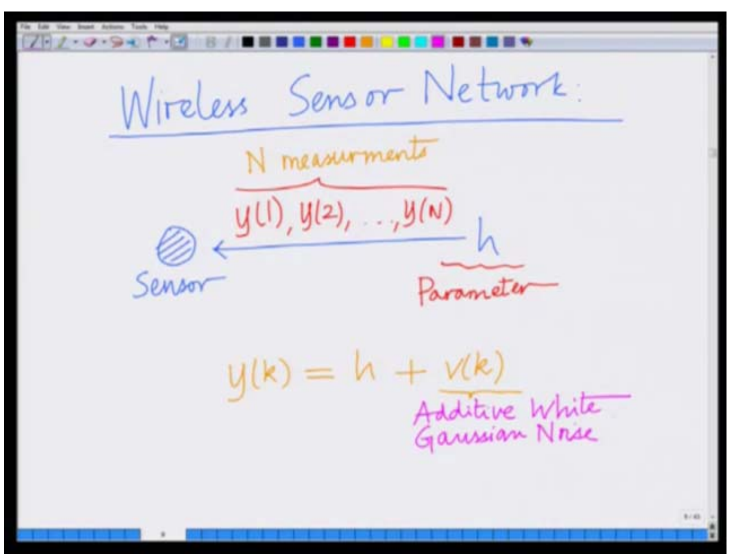
So, in this module, let us look at an example wireless scenario to compute the Cramer Rao lower bound which we also simply call as the Cramer Rao bound are basically the CRB. So, example wireless scenario to compute the Cramer Rao bound, ya.

(Refer Slide Time: 1:38)



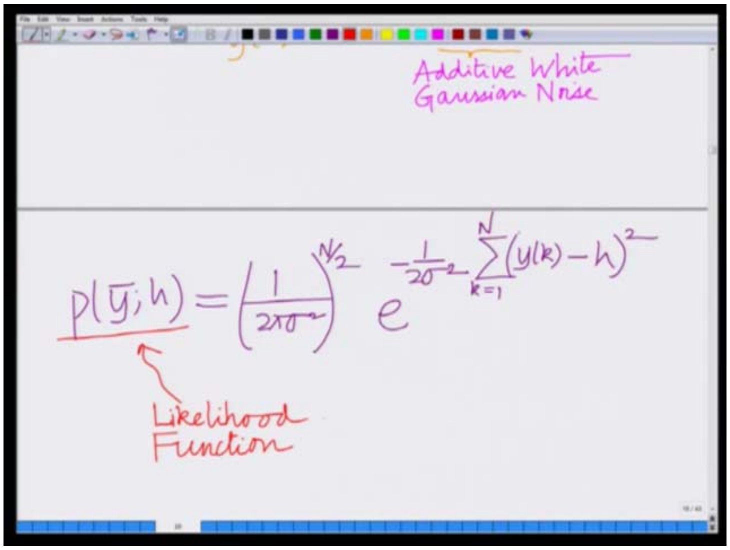
And let us go back to our wireless sensor scenario, let me remind you, we have already considered our wireless sensor network scenario in which we have a sensor node, so this is our sensor and which is trying to estimate a parameter H, this is the unknown parameter to be estimated and we have made N measurements Y1, Y2,… Up to YN. These we have called our N observations or N measurements. So, we are considering our wireless sensor network scenario where there is a sensor node which has made N observations of Y1, Y2,… Up to YN of the parameter H to be estimated and each observation YK is given as the parameter H plus the noise VK. That is the parameter being observed in additive white Gaussian noise.

(Refer Slide Time: 2:57)



So, we have YK equals H plus VK where VK is additive white Gaussian noise. In fact we have also derived the likelihood function for this, it is nothing but the joint density of the observations Y1, Y2,… Up to YN that is denoted by the observation vector Y bar.

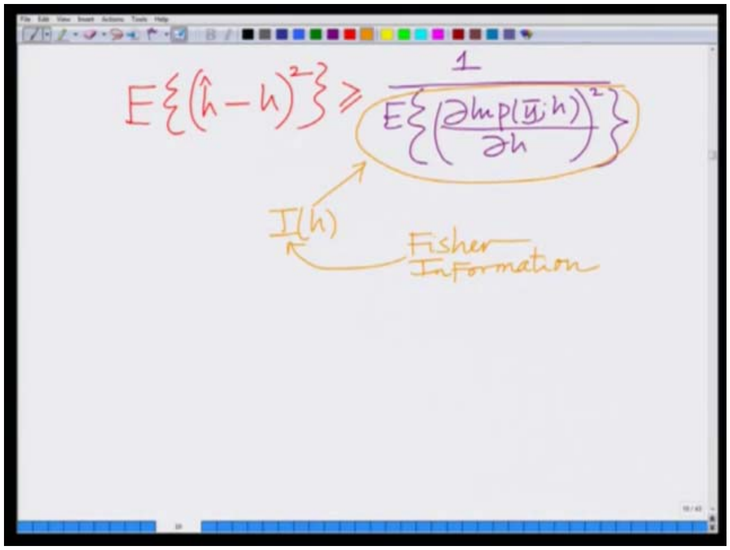
(Refer Slide Time: 3:34)



So we have the likelihood function and you can also recall that this likelihood function which has the joint density of the observations viewed as a function of unknown parameter that is 1 over 2 pie Sigma square, so let me write it a little bit clearly, that is the likelihood function P of Y bar characterised by the parameter H is 1 over 2 pie Sigma square raised to the power of N by 2 times E raised to -1 over 2 Sigma square summation K equals 1 to N YK minus H whole square. What is this, this is your likelihood function of the, this is the likelihood function corresponding to the parameter H. Which we said we have derived from the probability density of the observations Y1, Y2,… Up to YN which is denoted by the vector Y bar, viewed, that is the probability density function when we view it as a function of the unknown parameter H, this is the likelihood function, all right.

So, now to compute the Cramer Rao bound or the Cramer Rao lower bound, I have to 1st start by basically computing the Fisher information, that the I of H. That is what we said about the Cramer Rao lower bound, remember, the Cramer Rao lower bound corresponding to parameter H is given as

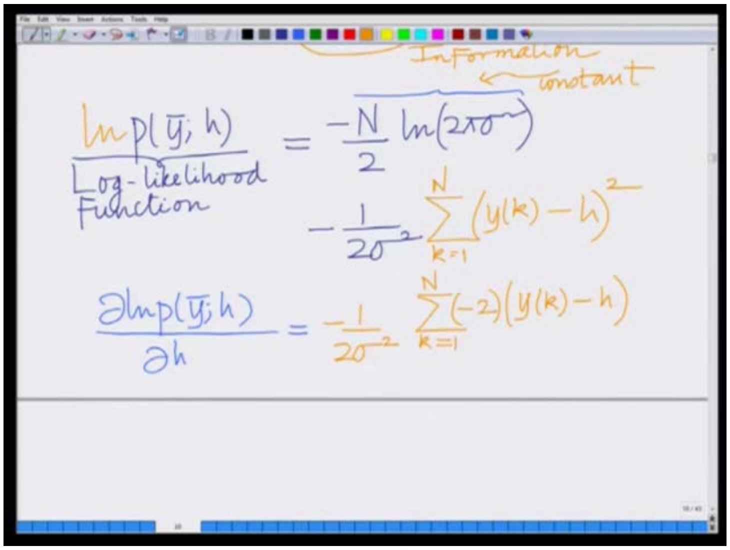
(Refer Slide Time: 5:16)



expected H hat minus 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, ya. And this quantity here, that is the expected value of the square of the derivative of the log likelihood functions, this is nothing but your future information. This is basically the Fisher information corresponding to the parameter H and we have also said the higher the Fisher information, the larger the Fisher information, the larger the Fisher information, the more information the likelihood function conveys about the parameter H, hence the lower is the mean square error.

Naturally if the information is more, the error is going to be lower and that is an interesting property of that, that is an interesting intuition from the Cramer Rao bound. Alright, so to do that, to compute the Fisher information, we start by computing 1st the log likelihood function from the likelihood function and then the derivative of the log like you would function. So, we have the expression for the like you would function, that is what we have already seen.

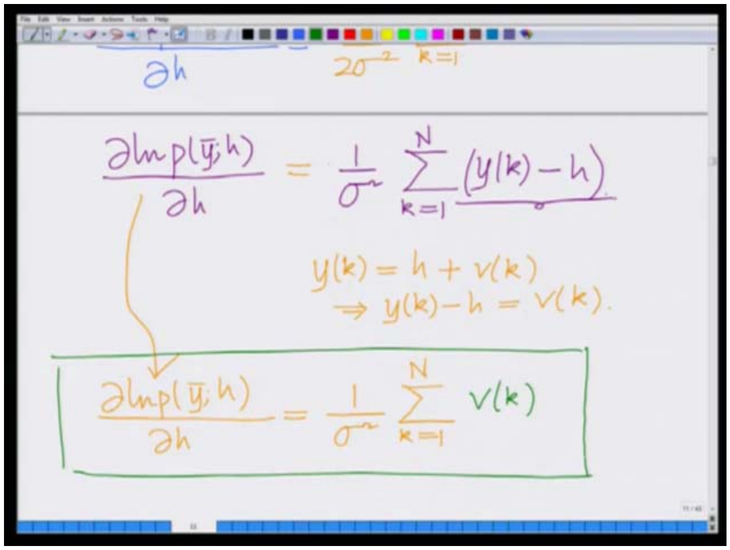
(Refer Slide Time: 6:37)



Now let us compute the log likelihood function, the natural logarithm of the likelihood function, this is your log likelihood function and this is, we have also derived the log likelihood function N by 2 log 2 pie Sigma square, in fact minus N by 2 log 2 pie Sigma square -1 over 2 Sigma square summation K equals 1 to N YK minus H whole square.

This is your log likelihood function. Now if I compute the derivative of the log likelihood function, you can see that derivative of the log likelihood function with respect to the parameter H, you can see this part is a constant, this does not depend, that is this minus N by 2 log of 2 pie Sigma square, this is not depend on H, so that the derivative of that with respect to a is 0, so what I am left with is -1 over 2 Sigma square summation K equals to 1 to N, the derivative of basically your YK minus H whole square which is -2 times YK minus H. So, the negative sign because there is a negative in front of the H,

(Refer Slide Time: 8:16)

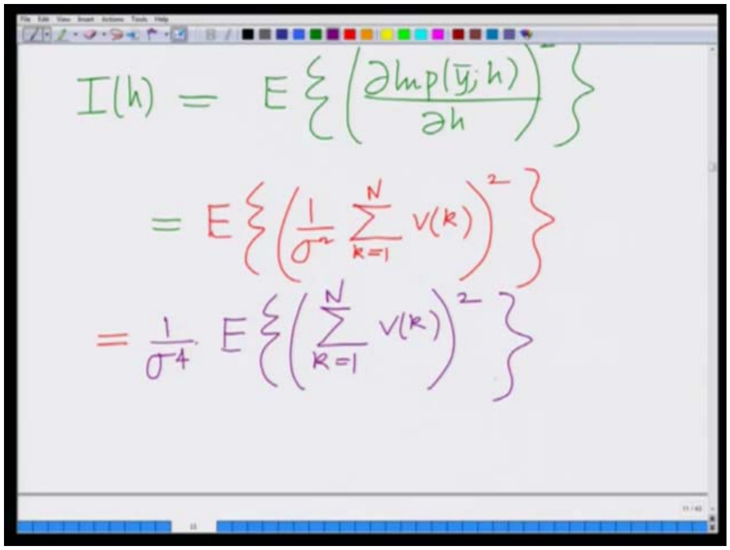


and therefore this can be simplified as the derivative of the log likelihood function is basically 1 over, the 2's cancel, so this is 1 over Sigma square K equals 1 to N YK minus H.

So, we have computed the log likelihood function from that we have computed the derivative of the log likelihood function and we can see that the derivative of the log likelihood function is 1 over Sigma square summation K equals 1 to N YK minus H. And we already know that YK equals H plus VK where VK is the noise, so YK minus H is nothing but VK, that is the noise. So, we know, basically now if you realise and that is also very clear, let me repeat that YK equals H plus VK implies YK minus H is nothing but basically VK which is the noise, this implies the derivative of the log likelihood function, your Y bar H with respect to H equals 1 over Sigma square instead of YK minus H, simply replace that by your quantity VK and this is basically the derivative of your log likelihood function.

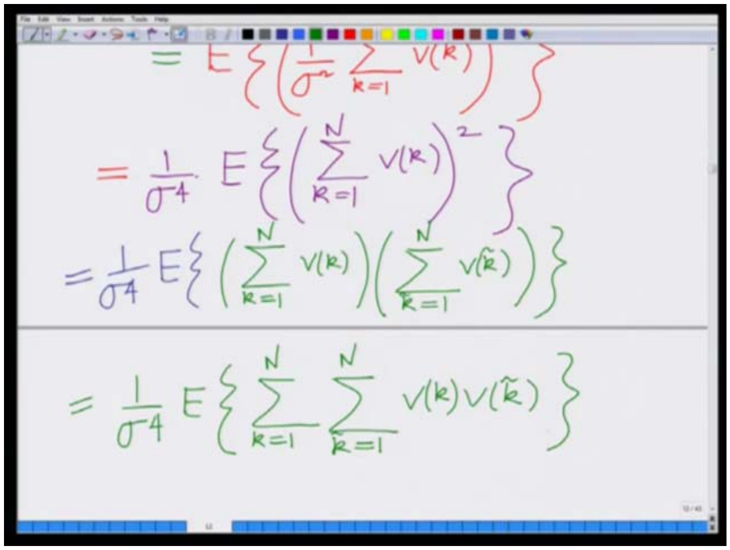
That is 1 over Sigma square summation K equals 1 to N summation VK, that is the derivative of the log likelihood function. Now the Fisher information is the expected value of the square of the derivative of the log likelihood function.

(Refer Slide Time: 10:10)



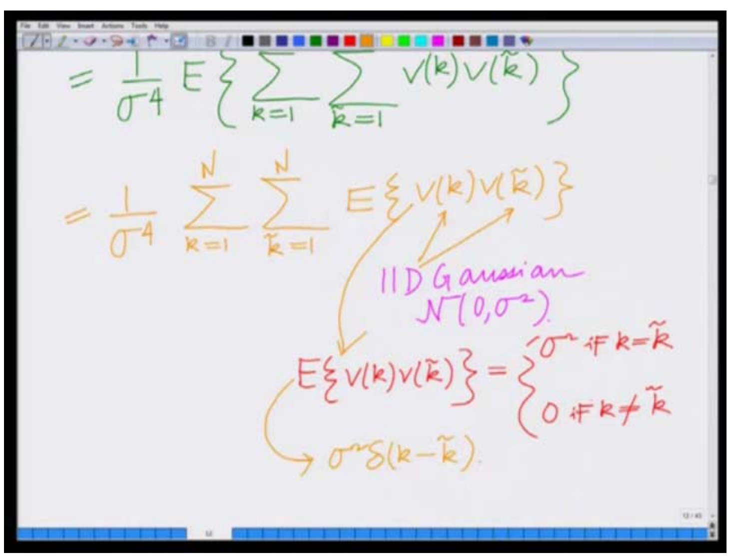
So, let us now compute the Fisher information which is the next step to evaluate the Cramer Rao lower bound, the Fisher information is the expected value of the square of the derivative of the log likelihood function that is Y bar H, the square of the derivative of the log likelihood function which is the expected value of basically we computed the derivative of the log likelihood function, that is 1 over Sigma square summation K equals 1 to N VK whole square. Now one over Sigma square is a constant, so that will come out of the expectation, so that is basically 1 over Sigma square square, that is one over Sigma 4, 1 over Sigma raised to the power 4, expected value of summation K equals 1 to N VK whole square. Yah, so that is one over Sigma raised to the power 4 expected value of summation K equals 1 to N VK whole square.

(Refer Slide Time: 11:25)



Now this I can also write as basically your 1 over Sigma to the power of 4 expected value of summation VK VK whole square , I can write it as summation K equals 1 to N VK Times itself but I am going to change the index and this is the trick we have seen several times before, K tilde equals 1 to N VK tilde and now if I expand this product of 2 summations, I have expected value of summation K equals 1 to N summation K equals 1 to N VK, VK tilde and now since expectation operator is linear,

(Refer Slide Time: 12:21)



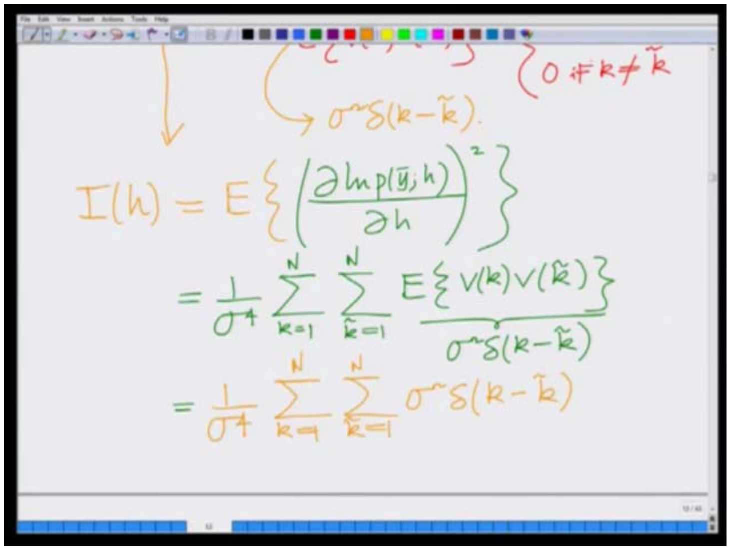
I move the expectation operator inside and what I have is basically 1 over Sigma raised to the power of 4 summation K equals 1 to N, summation K tilde equals 1 to N expected value of VK, VK tilde.

And this is something that we have seen several times before, that is expected value of VK VK tilde, remember we have said that the noise is white Gaussian, additive white Gaussian. And also we have said basically that means that the noise samples or other way to say that is basically the noise samples are independent identically distributed Gaussian random variables. Right, so these noise samples VK, VK tilde, these are IID Gaussian. Alright, and we have explored the properties of these noise samples in detail many times earlier.

They are IID Gaussian, which means they are identical Gaussian which means each of them has mean 0, that is Gaussian with mean 0 and variance Sigma square denoted by N 0, Sigma square and they are independent, which means basically the expected value of VK into VK tilde is basically equal to 0 is K is not equal to K tilde and is equal to Sigma square is K equals to K tilde. So, let me just summarise that once again so that you can just quickly recollect this property.

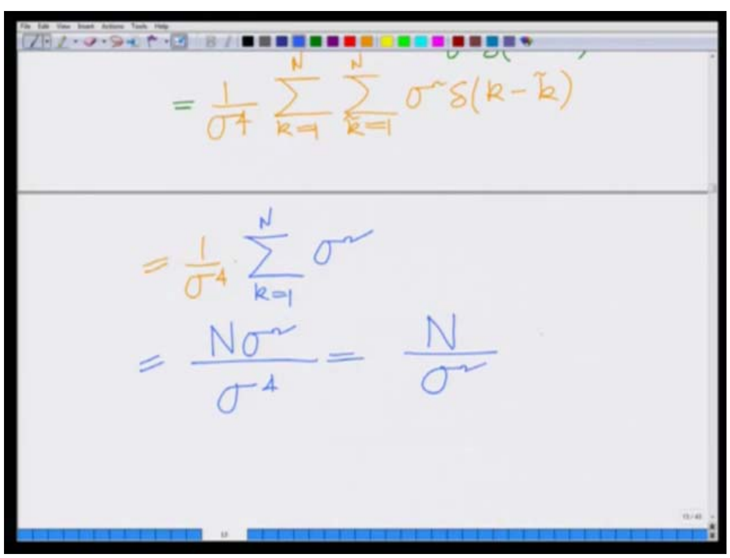
Since they are dependent, we have expected value of VK, VK tilde is basically equal to Sigma square is K equals K tilde and 0 if K is not equal to K tilde, this basically means that expected value VK, VK tilde is basically Sigma square delta, that is the discrete Delta K minus K tilde. Right, Sigma square Delta K minus K tilde is basically Sigma square K equals K tilde and 0 is K is not equal to K tilde. I am going to now substitute this expression for the correlation that is respected value of VK, VK tilde in basically the expression for I of H, that is the Fisher information of the parameter H that we have derived above.

(Refer Slide Time: 14:48)



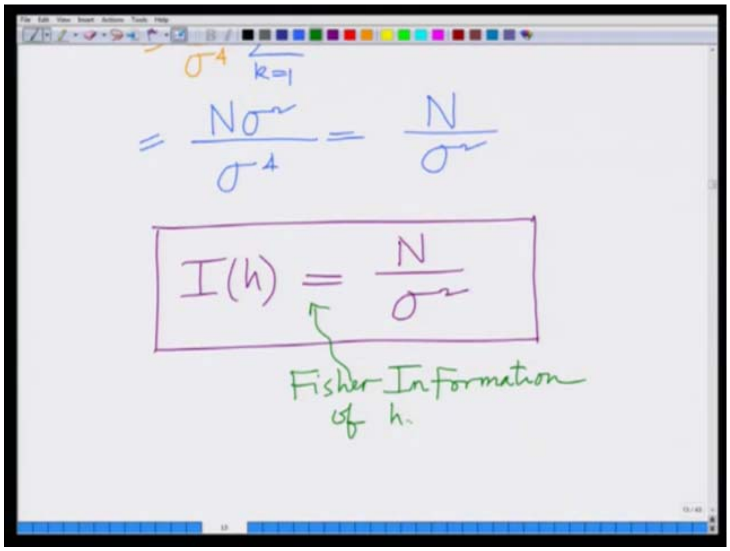
So, therefore we have I of H equals basically your expected value of the square of the derivative of the log likelihood function which is basically 1 over Sigma 4 summation K equals 1 to N, summation K tilde equals 1 to N, expected value of VK VK tilde and we have just seen that this is basically Sigma square Delta of K minus K tilde which means this is equal to 1 over Sigma raised to the power of 4, K equals 1 to N, K tilde equals 1 to N, K tilde equals 1 to N, Sigma square Delta K minus K tilde which means that the summation K tilde equals 1 to N, only the term corresponding to each K tilde equals K will survive.

(Refer Slide Time: 16:03)



So, this is basically 1 over Sigma to the power of 4, K equals 1 to N, Sigma square for each K tilde equals K which is basically equals N times Sigma square. Summation K equals 1 to N Sigma square is N times Sigma square divided by your Sigma raised to the power of 4, so this is equal to N, this is equal to, this is nothing but N divided by Sigma square and this is a very interesting result.

(Refer Slide Time: 16:43)

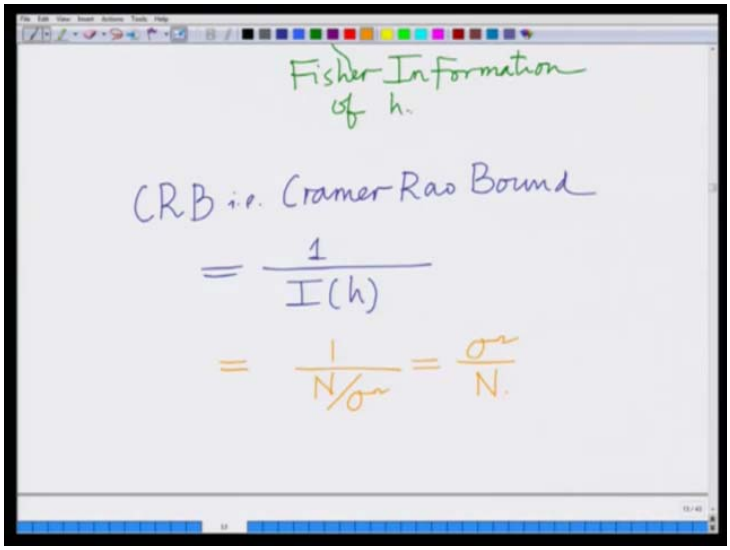


What we have just established is that the Fisher information for this sensor network scenario is N divided by Sigma square. And what is this, this is basically your Fisher information.

Fisher information of the parameter H. So, for this wireless sensor network example that we have seen double time before, in fact one of the 1st example that we have seen in the context of maximum likelihood estimation, so we have not demonstrated or derived an expression for the Fisher information of the parameter H and we have demonstrated this Fisher information denoted by I of H is equal to N divided by Sigma square by N is the number of observations, Sigma square is the variance, is the variance of each of the Gaussian noise samples and also this Fisher information is basically nothing but the expected value of the average value of the square of the derivative of the log likelihood function of the parameter H.

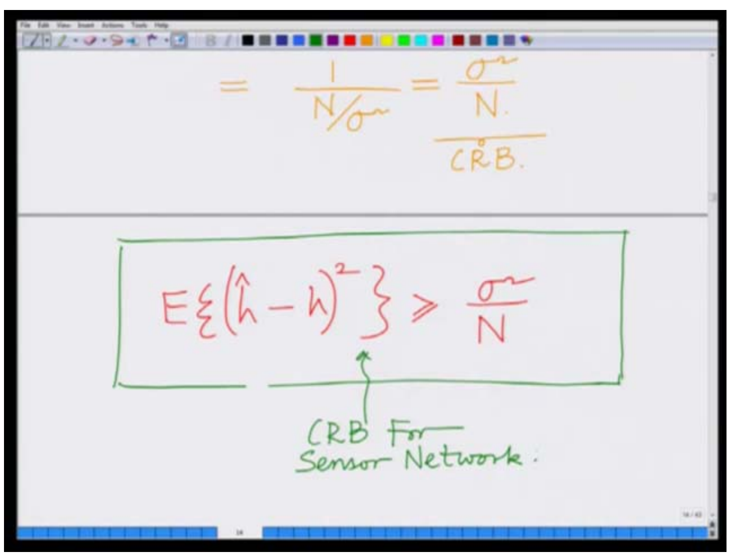
And now computing the Cramer Rao bound is simple, the Cramer Rao bound is simply the inverse of the Fisher information, that is 1 over I of H,

(Refer Slide Time: 18:06)



therefore now the Cramer Rao bound, that is your Cramer Rao bound is basically equal to 1 over the Fisher information which is basically which is basically 1 over N divided by Sigma square which is basically equal to Sigma square divided by N, this is the Cramer Rao bound. This is your CRB, which means this is a lowest variance achievable by any unbiased estimator

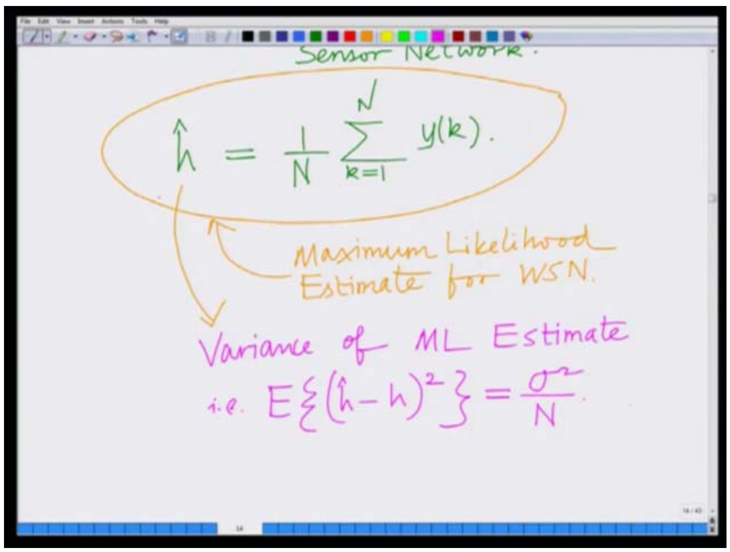
(Refer Slide Time: 18:49)



and therefore we have the Cramer Rao bound which is basically expressed as expected value of H hat minus H whole square is greater than or equal to Sigma square divided by N. And this is basically the Cramer Rao bound or the sensor network scenario.

This is your CRB this is the Cramer Rao bound for the sensor network or the wireless sensor network. So, we are saying we have derived the Cramer Rao bound which is the minimum variance achievable by in any unbiased estimator and that we have demonstrated is equal to Sigma square divided by N. So, the variance of any estimator which yields an unbiased estimate of the parameter H has to be necessarily greater than or equal to this quantity Sigma square by N which is the Cramer Rao bound. And you will notice something very interesting that we already have an estimator which uses Cramer Rao bound and that is nothing but the maximum likelihood estimate.

(Refer Slide Time: 20:03)

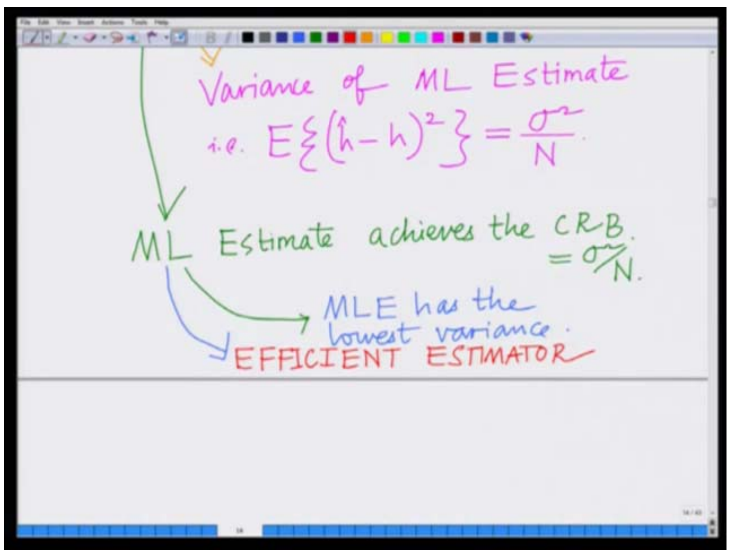


If you recall, for our maximum likelihood estimate a hat equals 1 over N summation K equals 1 to N summation K equals 1 to N of Y of K, this is basically what is this, this is basically your maximum likelihood estimate for the wireless sensor network.

This is basically the maximum likelihood estimate for the wireless sensor network. And the variance of the maximum likelihood estimate, variance of the maximum likelihood estimate is in fact, you will be happy to recall variance of your maximum likelihood estimate that is expected value of H hat minus H whole square is in fact Sigma square divided by N. Yah. So, what is interesting that is not only are we demonstrated through this Cramer Rao bound approach that the lowest variance achievable is Sigma square divided by N. We already have an estimator in fact which achieve this Cramer Rao bound, that is in fact the maximum likelihood estimate is also the sample mean of the observation that is 1 over N summation K equals to 1 to N Y of K, this is a sample mean of the observations Y1, Y2,… Up to YN, we said this is the maximum likelihood estimate which is derived by maximising the likelihood of, maximising the likelihood of, maximising the likelihood function with respect to the parameter H and this estimate, the maximum likelihood estimate already achieves this lowest variance, that is Sigma square divided by N. So, the maximum likelihood estimate achieves the Cramer Rao bound and therefore not only and, therefore there cannot be any other unbiased estimator estimator which has a lower variance that the maximum likelihood estimate.

So, the maximum likelihood estimator indeed turns out to be, turns out to be the best suited estimator for this scenario.

(Refer Slide Time: 22:27)



So, the maximum likelihood, so to summarise basically our maximum likelihood estimator, our ML estimator, the ML estimator achieves the CRB which is equal to Sigma square by N and therefore ML estimate has the ML estimate which is also abbreviated as MLE has the lowest MLE or basically the ML ML estimator has the lowest variance for any unbiased estimator. And such an estimator is said to be basically an efficient.

Any estimator which achieves the Cramer Rao lower bound is said to be a efficient, it is an efficient estimator, implies that it achieves the Cramer Rao lower bound, that is it gives the lowest possible variance, in this case, that is also the mean squared error. Since this is also an unbiased estimator, the variance is also basically your mean square error. So, basically what we have demonstrated in this in this module, as we have taken a concrete example considering a sensor, considering a wireless sensor network scenario with N observations, we have derived the facial information, the like you, the log likelihood and from that basically the Fisher information, from the inverse of the Fisher information, we have derived the Cramer Rao lower bound for parameter estimation in this wireless sensor network. We have demonstrated that this Cramer Rao bound is basically given by Sigma square divided by N which is the lowest variance achievable by any unbiased estimator of the parameter H.

And not only that, interestingly the maximum likelihood estimate, the simplistic seeming maximum likelihood estimate but yet which is very powerful that we have derived earlier also achieves this Cramer Rao bound which means it has the lowest variance amongst the class of all unbiased estimators, so therefore there cannot be any other unbiased estimator which has a lower variance. Which means that the ML estimator is the best estimator for this estimation scenario and also therefore this is also known as an efficient estimator since it achieves the Cramer Rao lower bound. We will stop this module here and explore other aspects in the subsequent modules. Thank you very much.