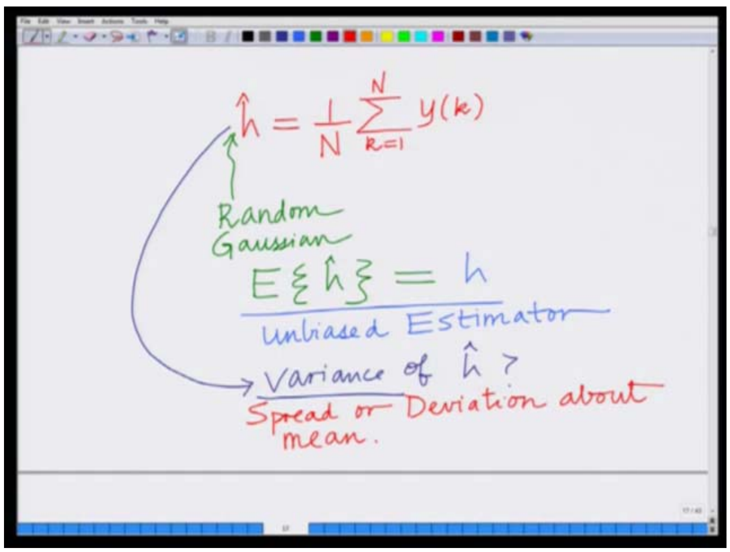
Estimation for Wireless Communication –MIMO/OFDM on Cellular and Sensor Networks
Professor A K Jagannatham
Department of Electrical Engineering
Indian Institute of Technology, Kanpur
Lecture Number- 04
Properties of Maximum Likelihood (ML) Estimate-Variance and Spread Around Mean

Hello, welcome to another module in this massive open online course on Estimation for Wireless Communication. So we are looking at the properties of maximum likelihood estimate and we have said that in our simple scenario of noisy observation in our wireless sensor network, we have the maximum likelihood estimate h hat, which is given as h hat equals 1 over n submission k equals 1 to n y of k.

(Refer Slide Time: 00:43)



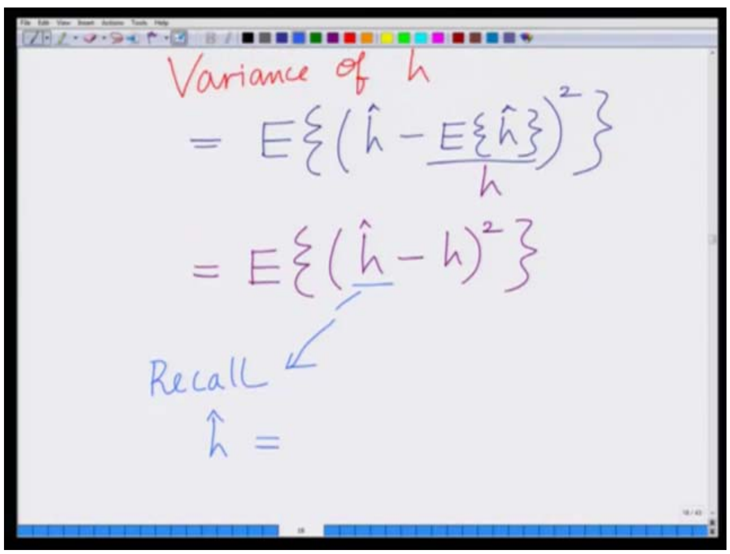
We have said that this estimate h hat is random Gaussian in nature. And we had in fact characterised the mean, that is the expected value of this estimate h hat, we had said is equal to the true value of the underlying parameter h and this is known as, to such an estimator so therefore such an estimate, this estimator the sample mean is basically an unbiased estimator.

That is the average value of the estimate is equal to the true value of the underlying parameter. That characterises the mean of the estimate, that is the expected value of h hat. Let us now look at the other aspect, that is the variance of the estimate. So let us now look at what is the variance of h hat what is the variance of h hat, remember the variance is always also a measure of the spread or the deviation about the mean.

The spread or the deviation about the mean remember, although on an average the estimator is the true underlying parameter if the spread about the mean is too much, then the estimator has a very poor performance. He would like the estimator that is the average to be equal to the true value of the underlying parameter; at the same time we would also like the spread that is the deviation of around the true parameter to be as low as possible.

So characterising this variance about the true underlying parameter of the spread of the estimate around the true underlying parameter h is also important, yeah. And this variance is basically nothing but a measure of this spread, the spread about the mean. So what is the variance, the variance of h hat if you might remember for any random variable, the variance of h hat for a random variable h hat is the expected value of the squared deviation.

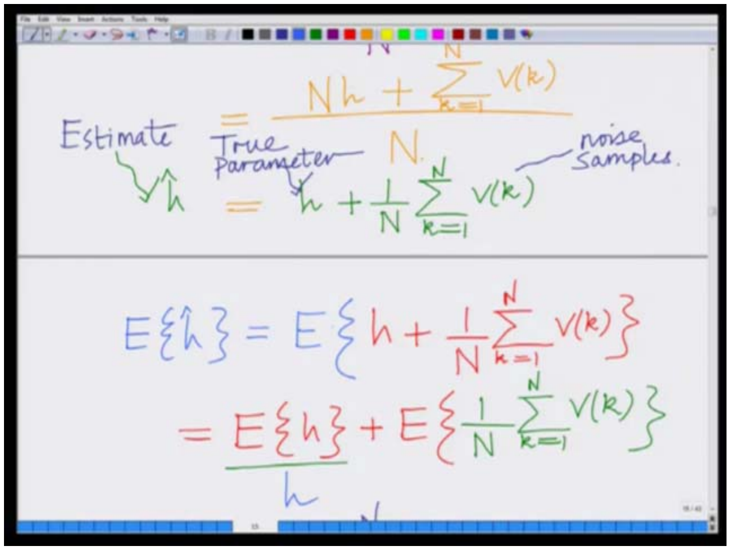
(Refer Slide Time: 03:13)



That is h hat squared deviation about the mean that is h hat minus expected value of h hat whole whole square, this is the expression for the variance. That is expected value of h hat minus the expected value of h hat minus expected value of h hat whole square. But we know that this expected value of h hat this is equal to h that is the true parameter itself. So this is simple, so this is nothing but expected value of h hat minus h whole square, yeah.

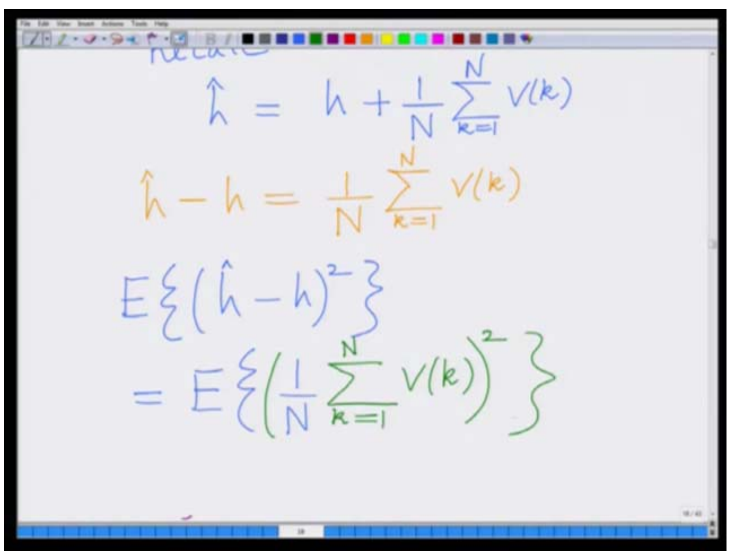
So this is the expected value of h hat minus h whole square. But recall that we have shown previously that this h hat, now recall that we had shown previously that is we had simplified.If you just go a little bit above.

(Refer Slide Time: 04:10)



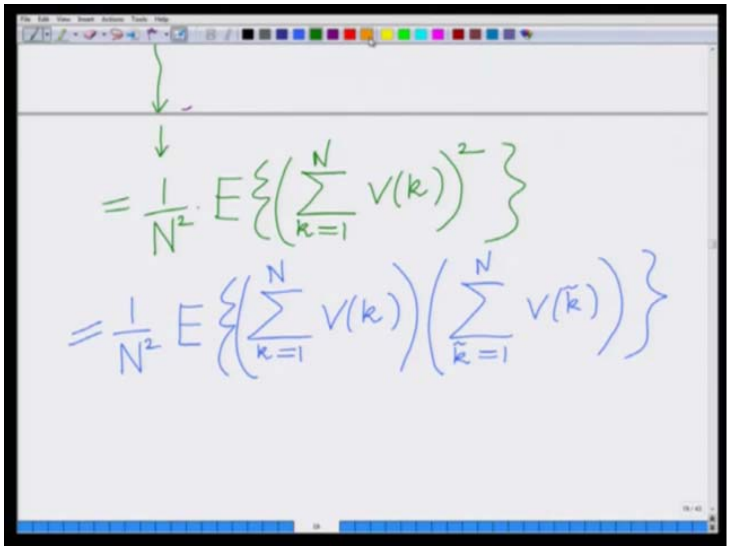
We had simplified h hat equals, that is your h hat equals h + h + 1 over n submission k equals 1 to V k. We will use that again, so I have h hat equals h + 1 over n submission k equals 1 to V k.

(Refer Slide Time: 04:30)



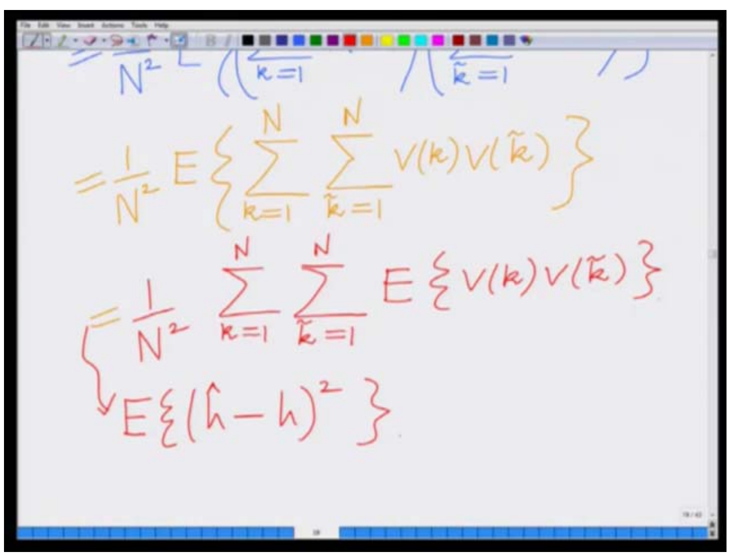
Which basically implies that h hat minus equals 1 over n submission k equals 1 to n V k. Therefore now I can compute this expected value of h hat minus h whole square is nothing but the expected value of 1 over n submission k equal to 1 to n V k whole square because h hat minus h is nothing but submission 1 over n V k. And this is basically equal to 1 over n square, the 1 over n is a constant, so 1over n square can come out.

(Refer Slide Time: 05:48)



Expected value of submission k equal to 1 to n V k whole square. Now the square of the term, square of something, I can write it as its product with itself. Therefore, I can change, I can as submission k equal to 1 to n V k with submission k tilde equals 1 to n, that is the product of 2 terms. Just changing the index and rewriting the terms, submission k equal to 1 to n V k times submission k tilde k equal to 1 to n V k.

(Refer Slide Time: 06:44)

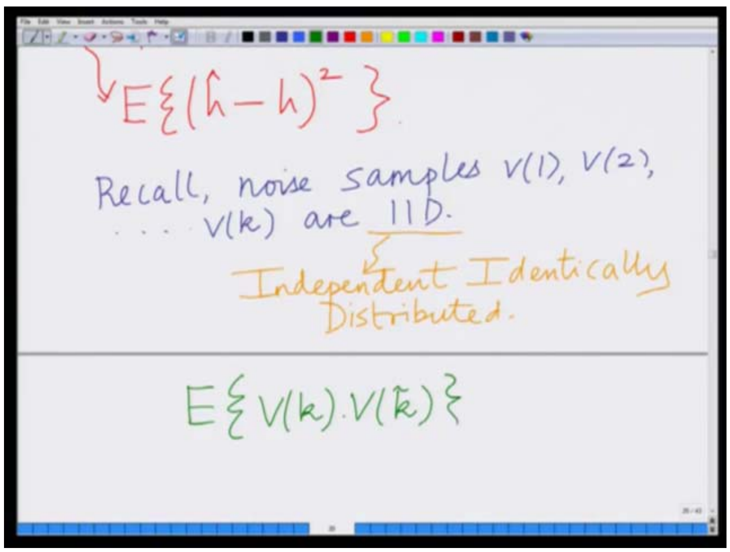


Which I can now write as basically your 1 over n square, expected value of submission k equals 1 to n submission k tilde equals 1 to n V k V k tilde. And now taking the expectation operator inside, I can rewrite this as 1 over n square submission submission k equal to 1 to n submission k tilde equals 1 to n expected value of V k V k tilde, yeah. So this is what is this, this is your simplified expression for expected value of h hat minus h whole square.

That is the expected value of h hat minus h whole square equals 1 over n square submission k equal to 1 to n submission k tilde equals 1 to n expected value of the product terms all terms of the form expected value of V k V k tilde yeah. Now we are going to use an important property of the Gaussian noise samples that we are we had assumed that is the Gaussian noise samples are independent and identically distributed.

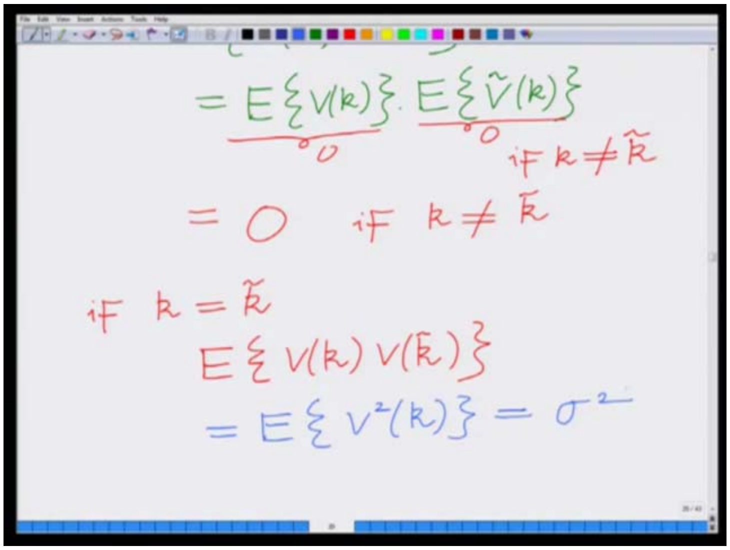
What does it mean to say that the Gaussian nice samples are independent, so recall that recall that these are IIDs.

(Refer Slide Time: 09:05)



That is what are these, these are independent distributed, which means if I take any 2 noise samples, expected V k into and look at the product expected value of the product V k V k tilde, this is equal to the expected value of V k times expected value of V tilde k.

(Refer Slide Time:10:00)

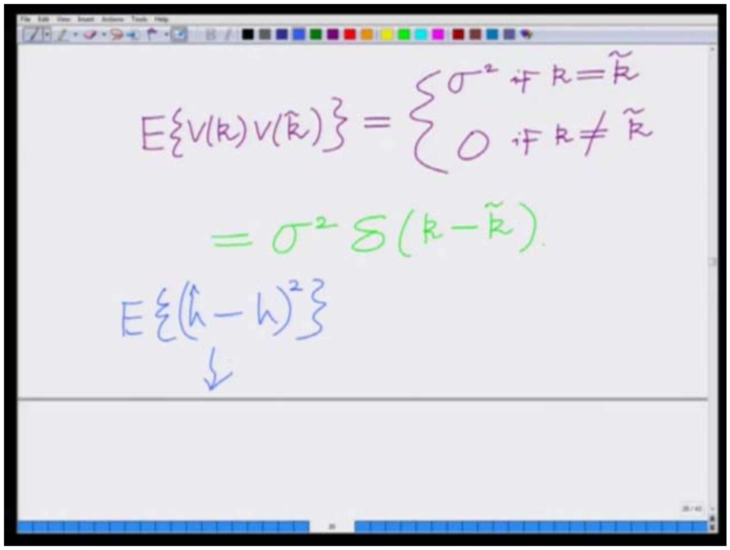


If when is if k is not equal to, that is expected value of the product V k V k tilde is expected value of V k times expected value of V k tilde. However, we know that the noise samples are 0 mean, therefore expected value of V k and expected value of V k tilde are both equal to 0. Therefore we have this is equal to 0, each expected value of V k tilde equal to 0 therefore the expected value of the product is 0 if k is not equal to k tilde.

On the other hand, if k equal to k tilde we have expected value of V k times V k tilde equals basically your expected since k equal to k tilde, this is basically expected value of V square k and expected value of V square k is nothing but the variance or basically the noise of the power sigma square, that is what we had also noted earlier.

Therefore, we have this beautiful property because the noises the noise samples are independent and we can identically distributed, we have expected value of V k V k tilde equals sigma square if k equals k tilde and 0 if k is not equal to k tilde, therefore to summarise this we have expected value of V k, the product V k V k tilde equals sigma square if k equal to k tilde, 0 if k is not equals k tilde.

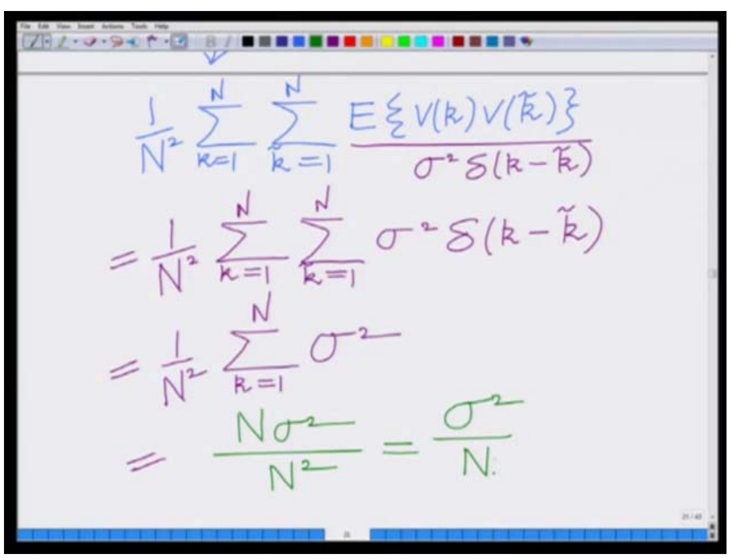
(Refer Slide Time: 11:45)



Which means, this can be basically summarised as being equal to sigma square Delta function of k minus sigma square delta function of k minus k tilde. Where delta of k minus k tilde that is [eq] that is equals 1 if k equals k tilde and 0 otherwise this is sigma square Delta k minus k tilde. Therefore now my variance which is remember expected value of h minus h hat minus h square.

Recall, expected value of, I can use this property to simplify the variance of the estimate as.

(Refer Slide Time: 12:55)

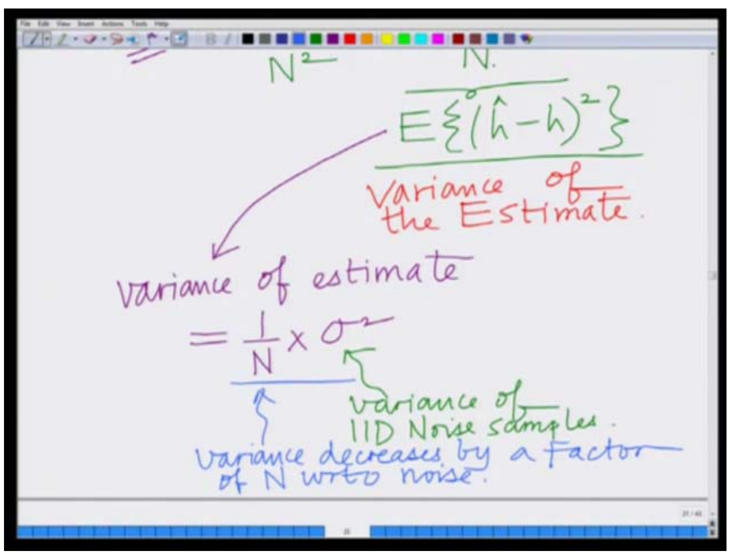


This is equal to, remember to this variance is equal to 1 over n square submission k equal to 1 to n submission k tilde equals 1 to n expected value of V k times V k tilde. And remember and this we have simplified as sigma square Delta k minus k tilde.

Therefore this is equal to 1 over n square submission k equals 1 to n submission k tilde equals 1 to n sigma square Delta k minus k tilde and this term will only survive if k tilde equals K, therefore one of the submissions in this will go away, because only terms for each K, I will have only one term which survives where k tilde equals K. Therefore, for each k I have k equals 1 to n sigma square.

Which is basically your n times sigma square divided by n square, which is basically sigma square divided by n and therefore what is this quantity sigma square divided by n, this is the variance of the estimate.

(Refer Slide Time: 14:22)



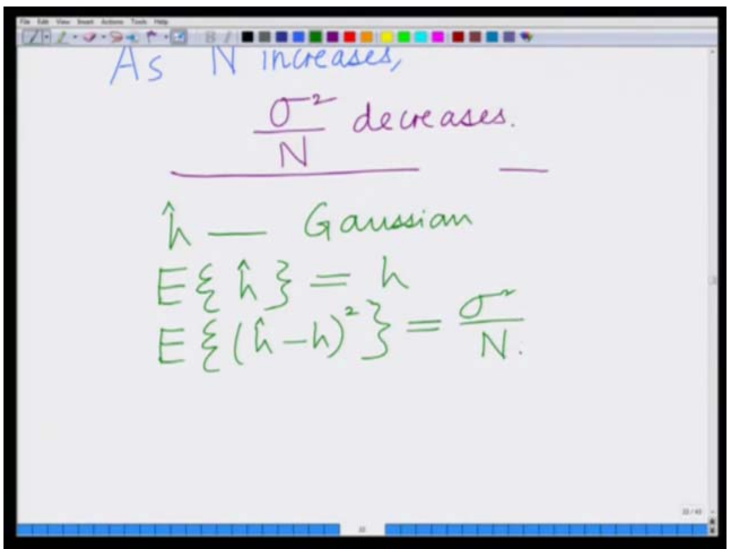
That is expected value of h hat minus h whole square or in other words basically also the variance of, the variance of the estimate. So we have derived this elegant expression for variance of the estimate.

We have shown that the variance of the estimate is sigma square divided by n where what is sigma square; remember sigma square is the variance or the noise power of each of the individual noise samples V 1, V 2 up to V n. So we are saying that if the noise samples V 1, V 2, V n R IIDs with mean 0 variance sigma square, the maximum likelihood estimator and estimate which is unbiased.

That is the mean of the estimator is equal to the true parameter h and the variance of the estimator is sigma square divided by n, that is 1 over n times the variance of the individual noise samples. So this is the variance of the estimate and therefore we have variance of estimate equals 1 over n times sigma square where sigma square equals your variance of the IID.

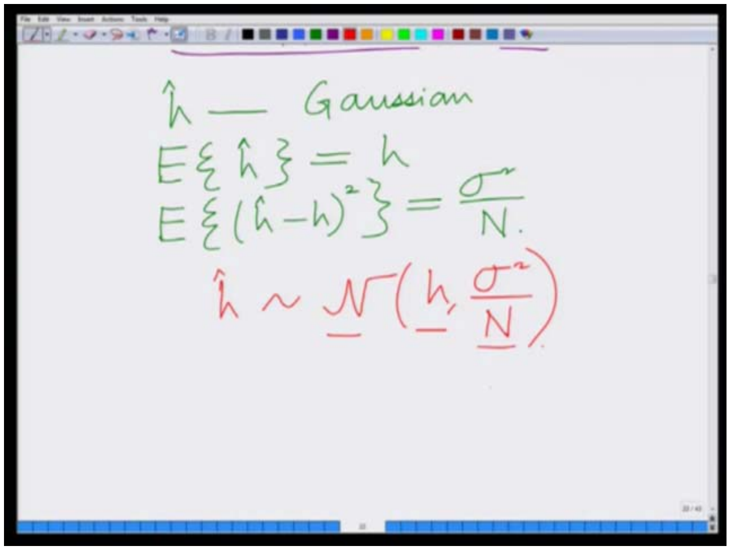
So what is this; so we are observing variance decreases by factor of n with respect to the noise. So what this tells us that by a factor of the variance of the estimate is 1 over n times the variance of the noise. And as n increases, and the important thing to observe is the following thing as n increases, the variance progressively keeps decreasing, that is the important thing. As n increases, sigma square over n decreases and this is an important.

(Refer Slide Time:16:54)



And lets us see what this let us see what this implies, therefore to summarise what we have is h hat is Gaussian, that is the estimate is random Gaussian, the mean of the estimate expected value of h hat is equal to h and the variance, that is expected value of h hat minus h whole square is equal to Sigma Square divided by n. And I can also denote this as basically h hat is your estimate h hat is Gaussian with mean h.

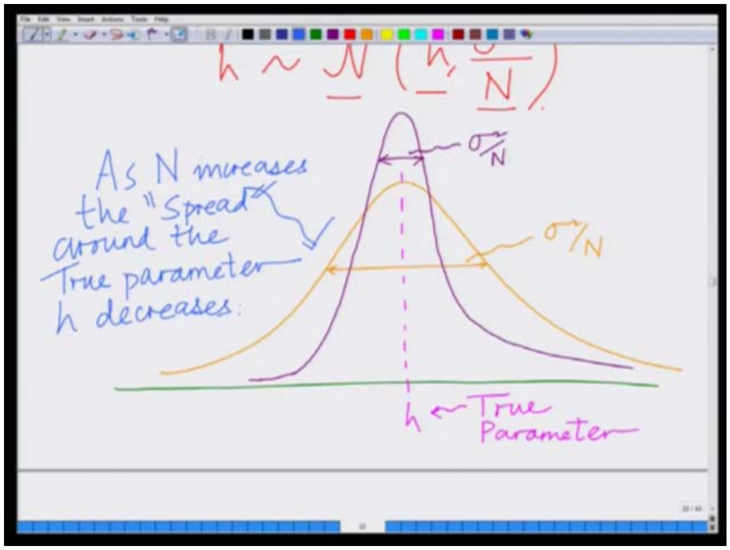
(Refer Slide Time:17:46)



Remember, n is the symbol for the normal distribution or the Gaussian distribution that is Gaussian with mean h and variance sigma square divided by n. That is this compact notation shows that h hat is normally distributed or Gaussian distributed with mean given by the true parameter h and variance given by sigma square over n where sigma square is the variance of the of each of the IID noise samples.

And why is this important? This is important for the following reason, because now let us plot the PDF of h. This h is random in nature, but you will observe something interesting when you will plot the PDF of h. So let us say, the PDF of h for a certain n is given by this Gaussian with spread sigma square by n. Remember the mean, it is centred around the true parameter h.

(Refer Slide Time:18:45)



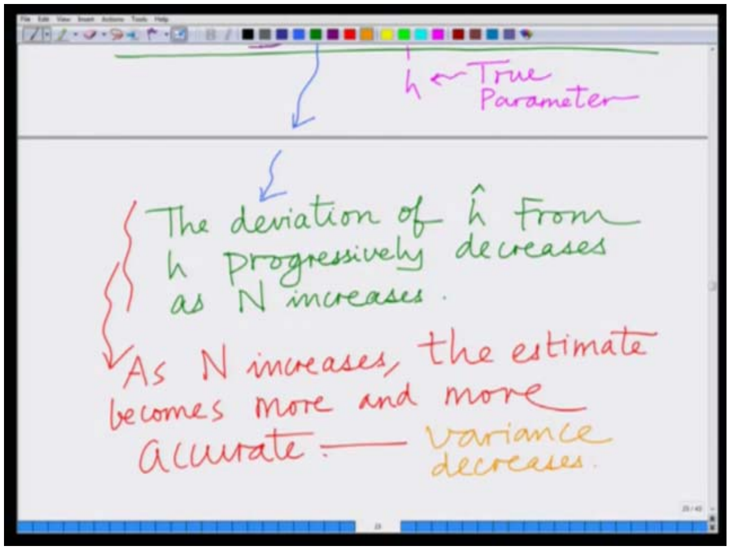
That is this what is h this is the true parameter which is the mean, the true parameter which is the, now as you increase n what happens, the centre remains the same because remember the mean does not depend on n, but what happens to the spread? Remember the spread which is sigma square divided by n, this decreases.

So you can observe as n increases, the spread that is the deviation or the variance around the true parameter h around the true parameter h decreases and this is a very important property. That even though the estimate is random in nature, on an average it is equal to the mean h and the variance of the estimator h hat which is sigma square over n progressively decreases as n increases.

Which means that Gaussian corresponding to h hat shrinks that is it becomes more and more picky around the mean, that is h. Which means the deviation from the true parameter that is, all though h hat remember we said h hat is not necessarily equal to h. Although h hat is not necessarily equal to h, the deviation from h, right the deviation from the true parameter h progressively decreases as n increases and that is an important point.

That is the deviation of h hat from h, the variance or basically your deviation of h hat or the spread of h hat from h progressively decreases, remember.

(Refer Slide Time:21:06)

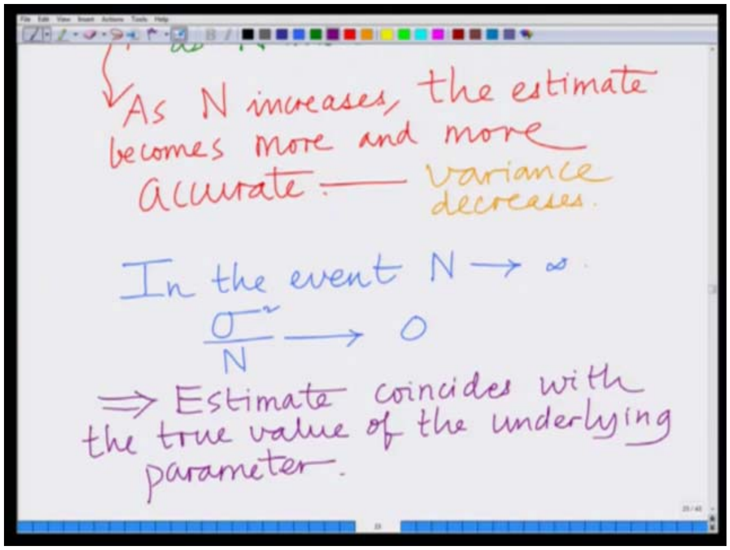


Progressively decreases and increases, therefore as n increases the estimate becomes more and more accurate in the sense that the variance decreases.

Therefore this is an important property. What we are saying is, because the estimate is random, we cannot exactly say that it is equal to the true it is never going to be equal to the true parameter h. However, what we can do something very important that basically because the variance is decreasing as n is increasing therefore progressively it is getting closer.

That is, it is a Gaussian distribution but it is getting more and more picky around the true parameter h. Therefore, in some in a certain sense it is becoming closer and closer to the true parameter h. And in fact if an n tends assume (()) (22:36) when n tends to infinity, in fact in the event n tends to infinity, we have Sigma Square over n tends to 0, that is variance tends to 0 which basically implies that your estimate coincides with the of the of the underlying parameter h.

(Refer Slide Time:23:03)



So what we are saying is, that is in when n becomes very large that is n tends to infinity, then sigma square by n tends to 0, that is the variance is 0 which means the estimate h hat coincides with the true parameter h. And this is the behaviour of the maximum likelihood estimate which is very interesting behaviour. That is first what we have demonstrated is this maximum likelihood estimate on average is the true parameter.

Not only that, the spread or the variance around the true parameter h is basically sigma square divided by n which means, as the variance as n the number of samples or the number of observations increases, the variance become smaller and smaller, that is the deviation becomes deviation around the true parameter become smaller and smaller.

That is the error that is remember the error or this deviation decreases progressively and as in (())(24:26) as n becomes infinity, the variance becomes 0 and therefore the estimate coincides with the dynamiter and this is the interesting behaviour exhibited by the ML estimate or the maximum likelihood estimate.

Alright, so what we have done so far in this mooch, we are at a point where we have derived them maximum likelihood estimate based on the likelihood function and not only that, but we have characterised the behaviour and the properties, interesting behaviour exhibited by this maximum likelihood estimate. We will explore other expects of this estimate in the future module. Thank you very much.