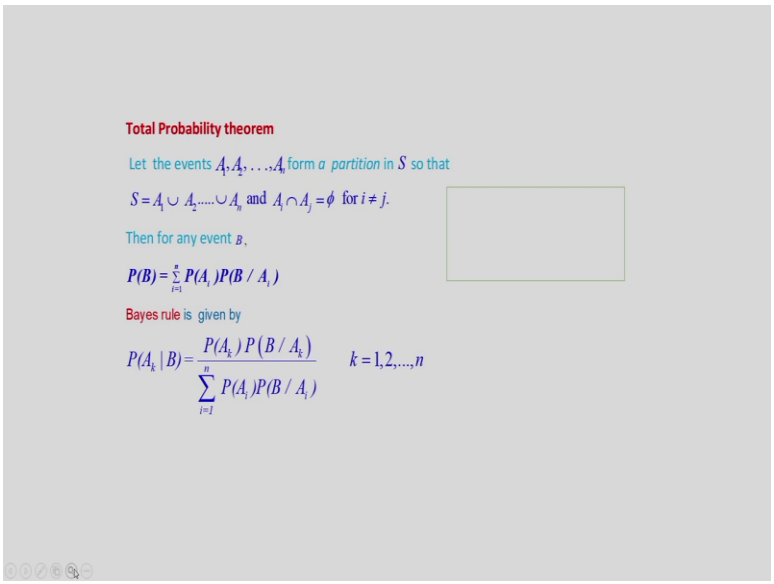
Statistical Signal Processing
Prof. Prabin Kumar Bora
Department of Electronics and Electrical Engineering
Indian Institute of Technology – Guwahati

Lecture - 38
Review 1

(Refer Slide Time: 00:41)

Hello students we are approaching the end of the course in this lecture. We will review some of the important concepts our course involved signal modeling by random processes, estimating the model parameters, signal estimation by optimal linear filters. We will review the first two topics first.

(Refer Slide Time: 01:02)

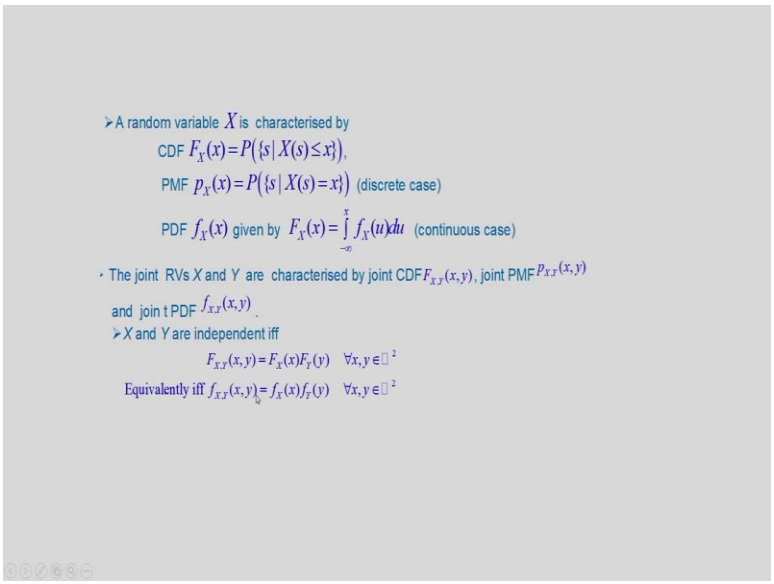


We will start with some basic probability concepts first one is total probability theorem suppose a 1 a 2 up to a n form a partition on the sample space S. So, a 1 a 2 may be up to a n this form a partition on the sample space as that means union of all these events is the sample space S and they do not have any intersection among them. Then any event being suppose being an event the probability of event B we establish that this is equal to summation P of Ai into P of B given Ai, i going from 1 to n.

Now using the total probability theorem we can derive the Bayes rule that is probability of A k given B suppose some event B has occurred what is the probability that a particular event A k occurs so that is the probability of A k given B and this is given why this result probability of A k into probability of B given A k divided by this probability of B is obtained to the total probability theorem. So, that way Bayes theorem is important for us here this probability of all those events with form a partitions these probabilities are known as the apriori probability or prior probability.

And this is the conditional probability then using this apriori probability and conditional probability or the be given A k we can find the Bayes rule. So, Bayes rule will determine the posterior probability, probability of A k given B and this we can determine using this formula.

(Refer Slide Time: 03:09)

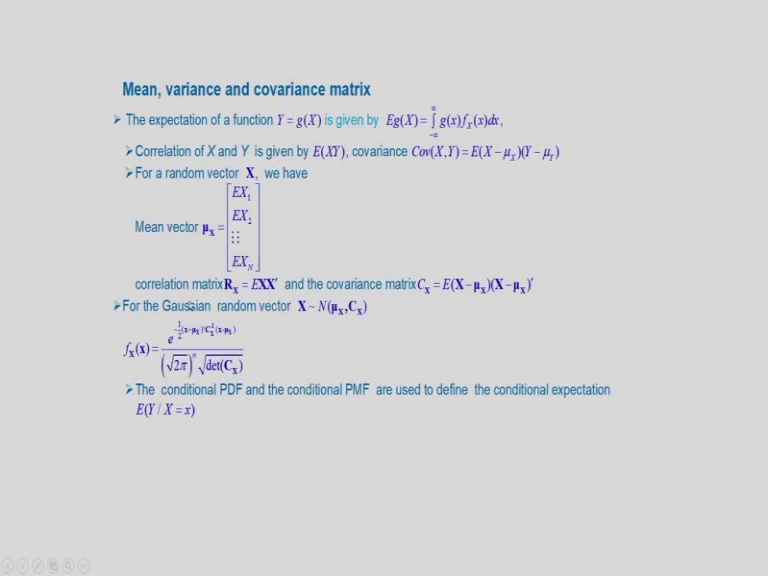


Next we define a random variable you know that random variable is characterized by a CDF F X of x later on we simplified the notation we simply write this is equal to F of x. This is the probability of the event s such that X s is less than equal to small x. Similarly if X is a discrete random variable we define the probability mass function small px x is equal to probability of those s for which X s is equal to small x, probability of s such that X s is equal to X.

And in the case of discrete case we define the PDF. So, PDF small fx and here the CDF is related to the PDF through this integration. So, PDF at X CDF at X is the integration of F x u du u going from - infinity to x, so this was the definition of PDF. Joint random variable x and y are characterized by the joint CDF that is capital F xy at point xy or joint PMF small pxy at point xy or joint PDF small F xy at point xy.

Later on we simplified these notations so we do not simply Fxy here similarly this is p xy and this one is small f xy and the concept of independence is important 2 events 2 random variables X and Y are independent if the joint CDF is product of the marginal CDF. So, F xy is equal to Fx into Fy for all xy belonging to R 2, so this was the definition similarly in terms of joint PDF also we can define it. So, joint PDF f point xy is the product of the marginal PDF. So that way we define the independence.

(Refer Slide Time: 06:11)



So next we saw how we can characterize random variables in terms of its mean variance moments etc. The expectation of a function Y is equal to gX is given by this formula E gX is equal to integration gx fx dx x going from - infinity to + infinity. So, this is the definition of expectation of any function of gX. So, we can use this definition to find out mu is equal to E of X so we will put gX is equal to X here.

E of x square then variance Sigma x square is equal to E of x - mu whole square so we put gX is equal to X – mu. So, that way we can find out various expectations given the PDF or PMF and for two random variables x and y we have the joint expectation E xy. So, it is also called correlation and similarly covariance of xy is E of x - mu x into y - mu y this is the function covariance. And if we have two or more random variables we can represent the random variable by a vector.

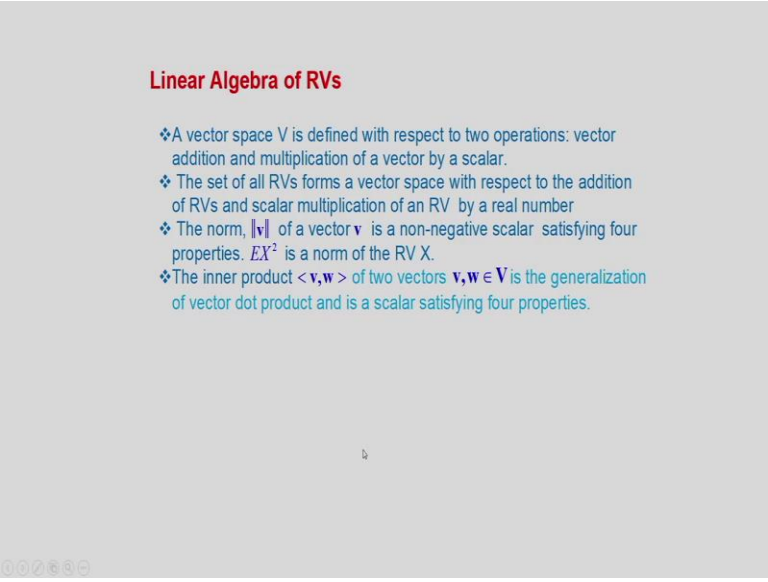
For example this x may be represented as a vector that is there are N random variables like this suppose then this can be represented as a vector. And now given the random vector we can find out the mean vector mu x, mu is a vector its components are E of X 1 E of X 2 up to E of X N. Similarly we can define correlation metrics to measure the correlation among the random variables we define the correlation matrix it is given by E of X given X transpose that means E of this one is the defined as E of that it X is this X 1 X 2 up to X N.

Then transpose will be a row vector X 1 X 2 up to X N so this will give us a matrix and then we will take the expectation of individual elements. Then covariance matrix similarly is defined C x is equal to E of X - mu x into X - mu x transpose. And one important distribution is the Gaussian distribution for multiple random variables we have the Gaussian random vector and this is denoted by normal me did the one by the parameter mu x vector and the covariance C x matrix.

And its distribution is given by fx is equal to e to the power - half of x - mu x transpose into C x inverse into x - mu x divided by root over 2 pi to the power n into square root of determinant of C x, so this is the expression for the multivariate Gaussian or vector Gaussian random vector. And we also defined the conditional PDF for example conditional PDF we define f of x given y is equal to f xy divided by f y provided f y is not equal to 0.

And once we define the conditional PDF or conditional PMF similarly we can define we can find out the conditional expectation E of Y given X is equal to X so this is equal to this is equal to and it is y f of y given x dy integration from - infinity to infinity. So, this is the definition of conditional expectation.

(Refer Slide Time: 11:10)



Then we discussed about linear algebra random variables. We discussed that a vector space V is defined with respect to two operations vector addition and multiplication of a vector by a scalar. The set of all random variables form a vector space with respect to the addition of random

variables and scalar multiplication when random variable of a random variable by a real number this scalar in case of real vector space either is a real number. So, therefore a vector space is defined with respect to two operations here one is the summation of two random variable that is addition and the other one is multiplication of a random variable by a scalar.

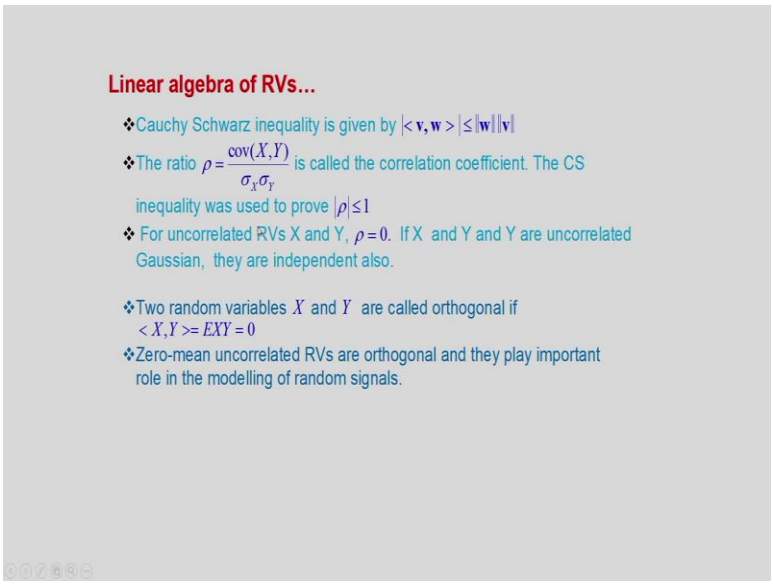
Then we define a normal way vector V this is the notation is a non-negative scalars satisfying four properties. So, what are those properties that is norm of X is always greater than equal to 0 number two norm of X is equal to 0 if X is equal to 0, number two if and only if, number three suppose r is a scalar then norm of rx is equal to absolute value of r into norm of x. So, this is for r belonging to real line.

Similarly 4 is norm of x + y suppose x and y are two vectors then norm of x + y is less than equal to norm of x + norm of y. So, norm of X + y is less than equal to norm of x + norm of y this is known as a triangular inequality. So, we also saw that E of X square is the norm of the random variable X we saw that your X square satisfies these four properties so it is a norm. Similarly we define the inner product of two vectors this is the symbol inner product of v w.

So, this also generalization of dot product of ordinary vectors and it is a scalar satisfying 4 properties here also we saw 4 properties are there and those are like; so number one is that is xy inner product of XY is equal to inner product of Y into inner product of Y X this is for all X Y. Then number 2 inner product of X, X is equal to norm of X square, number 3 is if you consider r X, X that will be equal to r times inner product of XX.

Then that number 4 is if I considered a suppose inner product of X, Y + Z that will be equal to inner product of XY + inner product of X Z, so that way we can define the inner product. So, in the case of random variables E of XY is an inner product operation E of XY is an inner product operation that we have solved.

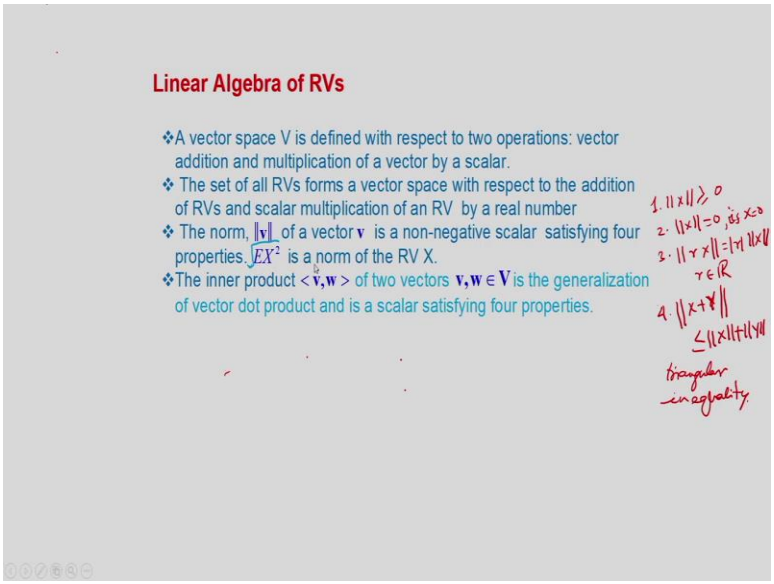
(Refer Slide Time: 15:48)



Then we established one important property that is Cauchy-Schwarz inequality the tip which is given by no magnitude of inner product of v and w is less than equal to norm of w into norm of v this is the Cauchy Schwarz inequality and because of this E of XY suppose if we take E of XY maginitute is there is less than square root of EX square into square root of EY square, so norm of X is EX square root of E of X square norm of Y is square root of EY square.

So, that way magnitude of E of XY; is less than equal to square root of EX square into square root of EY square that is known as the Cauchy Schwarz inequality in the case of vector space.

(Refer Slide Time: 16:52)



Square root of E of X square is a norm of the random variable X. So, once we define the norm we can show that this quantity square root of E of X square is a norm which satisfy these four properties. Similarly we define the inner product of two vectors. So, it also satisfy 4 properties number one is inner product of XY is in equal to inner product of YX. Number 2 is inner product of X, X is equal to it is a norm square, number 3 inner product of any scalar multiplier r X, Y that is equal to r times norm of r into inner product of XY.

Then number 4 is inner product of X + Y, W is equal to inner product of XW plus inner product of YW. So, that way these 4 properties if it is satisfied by a scalar of portion like this then it is an inner product operation. Then we establish the Cauchy Schwarz inequality which is given by mod of inner product of v and w is less than equal to norm of w into norm of v. So, this we can directly apply to establish this relationship that is magnitude of E XY is less than equal to magnitude of E of X square root of E of X square into square root of E of Y square.

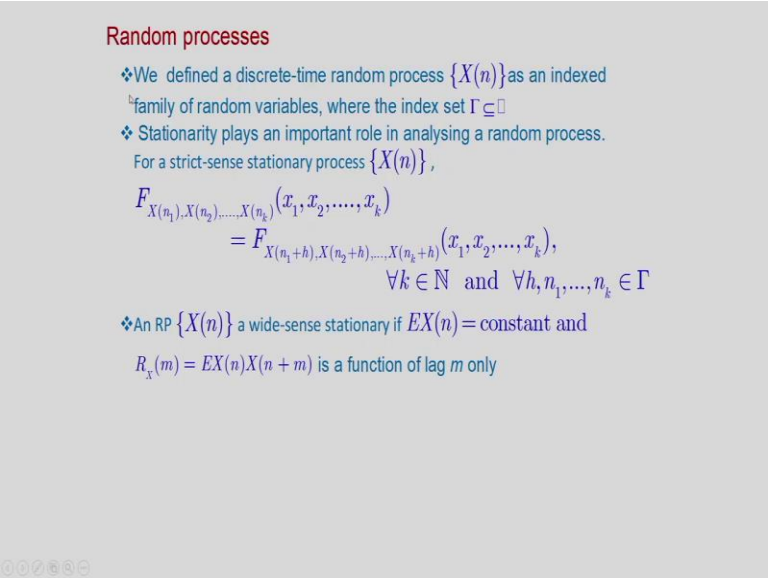
So this result directly follows from Cauchy Schwarz inequality and that also resulted in that covariance of magnitude of covariance of XY is less than square root of Sigma square into Sigma Y square so that way we get this ratio rho is equal to covariance of XY divided by Sigma X into Sigma Y and it is called the correlation coefficient and we can use the Cauchy Schwarz inequality to prove that mod of Rho that is the correlation coefficient is less than equal to 1.

For uncorrelated random variables X and Y, rho is equal to 0 so X and Y are uncorrelated, uncorrelated if Rho is equal to 0, so and this will imply that E of X Y is equal to EX into EY. If X and Y are uncorrelated Gaussian they are independent also usually uncorrelatedness does not imply independence but if X and Y are uncorrelated Gaussian they are independent also.

Now using the concept of in our product we can define orthogonal random variables two random variables X and Y are called orthogonal if inner product of XY that is equal to E of XY is equal to 0 so this is the definition of orthogonal random variable 0 mean uncorrelated random variables are orthogonal because in that case uncorrelated random variable E of XY is equal to EX into EY and 0 mean if any of the mean is 0 then this will become 0 therefore it will become orthogonal.

So, zero mean uncorrelated random variables are orthogonal and they play important role in modeling of random signals.

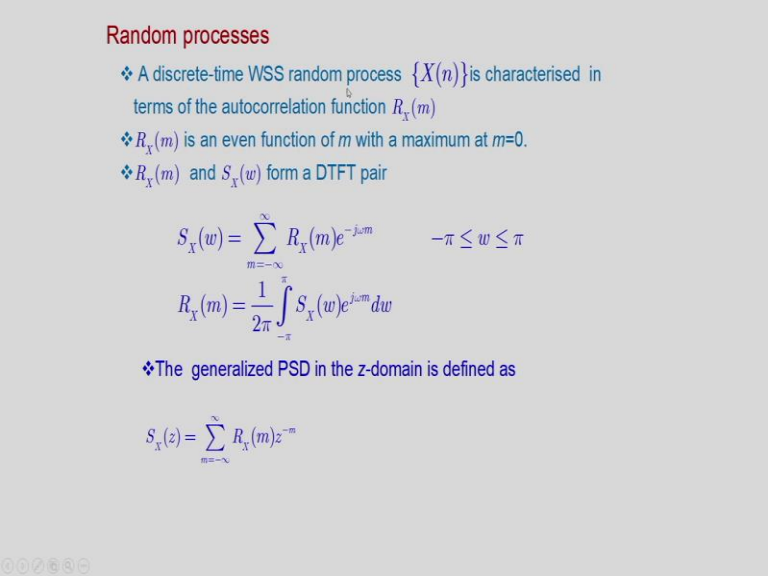
(Refer Slide Time: 21:28)



Next we discuss the concept of random process we defined a discrete-time random process Xn and then index family of random variables where the index set tau is a subset of set of integer and characterizing a random process by probability is difficult very complex therefore we define some easier concept that is stationarity and wide-sense stationarity. For a strict sense stationary random process, for a strict sense stationary process Xn the joint distribution at instant n 1 and 2 up to n k is same as the joint distribution at instant n 1 + h n 2 + h up to nk + h.

So joint distribution is invariant under the reciliations of the time axis so this was the definition of the a strict sense stationarity. And then we consider stationarity in a very narrower sense a random process xn is wide sense stationary if you have Xn is constant and autocorrelation function, now autocorrelation function is defined as Y over X n into X of n + M joint expectation of Xn and Xn + m is the function of lag m only so it will not depend on this parameter n but it will depend only the difference parameter like parameter m only that way we define the WSS process.

(Refer Slide Time: 23:32)

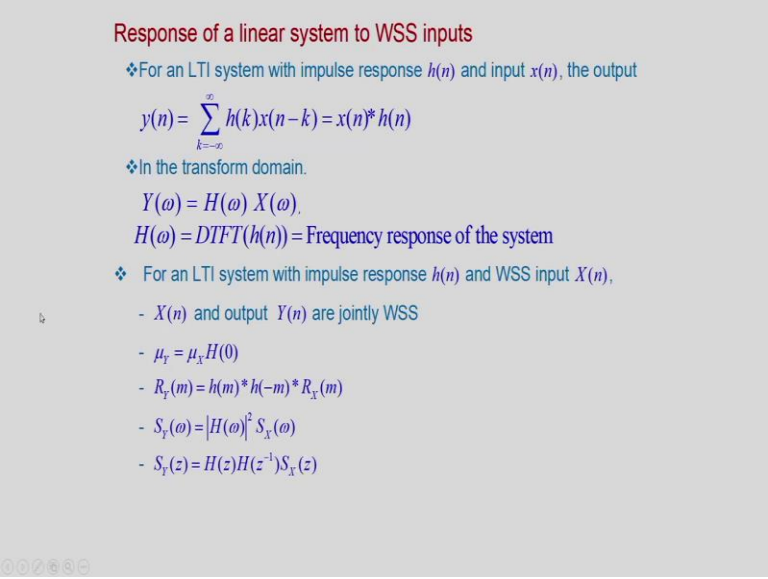


A discrete term WSS random variable a discrete-time WSS random process Xn is characterized in terms of the autocorrelation function Rx of m and we observe that Rx of m is an even function of m with max with a maximum at m is equal to 0. Then we define the power spectral density that in the case of WSS random process that is power spectral density SX Omega this is the average power part frequency this is defined by that is S x Omega is equal to summation R x of m into e to the power - j omega m, m going from minus infinity to plus infinity.

So this is the discrete time Fourier transform of the autocorrelation sequence. So, a S x of Omega is equal to summation Rx of m into e to the power - j Omega m, m going from minus infinity to plus infinity. So, this is this discrete time Fourier transform of the autocorrelation sequence and here Omega is defined uniquely from - PI to PI because this function is a periodic function just like any other DTFT function therefore it is uniquely defined for Omega lying between minus PI and PI.

And autocorrelation is related with power spectral density by this relationship Rx of m is equal to 1 by 2 pi integration Sx Omega e to the power j Omega m d Omega Omega going from minus PI to PI. So, this way we define power spectral density and how Auto correlation function is related to power spectral density and for our analysis generalized power spectral density is important this is the Z transform of the autocorrelation sequence and given by this formula S x that is equal to summation Rx of m Z to the power – m, m going from minus infinity to infinity.

(Refer Slide Time: 25:54)



Next we discussed about the response of a linear system to WSS inputs. So, we know that for a linear time-invariant system the input-output relationship is this yn is equal to summation hk into x of n - k k going from minus infinity to infinity this is the convolution relationship and we denote it by xn start hn and in the transform domain we take the discrete-time Fourier transform of both then Y Omega is equal to H Omega into X Omega where each Omega is the DTFT of HN that is impulse response sequence and it is called the frequency response of the system.

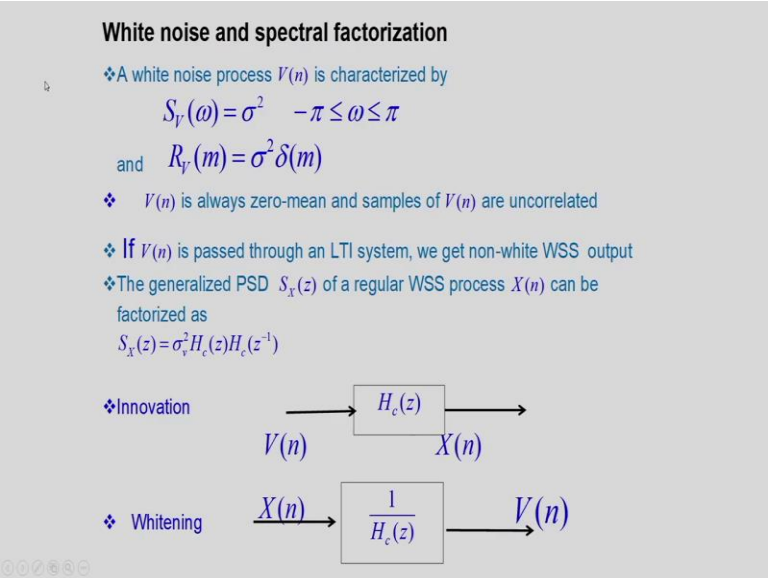
Now in the case of LTI system with WSS input we know deep following that input Xn and the output Yn are jointly WSS that we established and expected value of the output Mu y that is equal to MU x times H of 0 what is H 0? H 0 is the frequency response at 0 frequency that is equal to summation of hn, n going from minus infinity to plus infinity n is equal to h of 0. Then we establish that Ry m that is a top correlation of the output is hm is convolved with hm -m then convolved witht Rx of m.

So, Ry of m is hm convolved hm - m convolved with Rx of m. And in the frequency domain power spectral density Sy Omega is equal to mod of H Omega square into Sx Omega. So, these we get from here Fourier transform of this will be a Sy Omega this one will be H Omega this is a star Omega because it is a hm -m together a it will be norm of H Omega or magnitude of H

Omega square into power spectral density of X that is Sx of Omega so that way as Sy Omega is equal to mod of H Omega square into Sx Omega.

And in this z transform domain we have Sy z is equal to Hz into Hz inverse into Sx z so this is the relationship in the z transform domain.

(Refer Slide Time: 29:14)



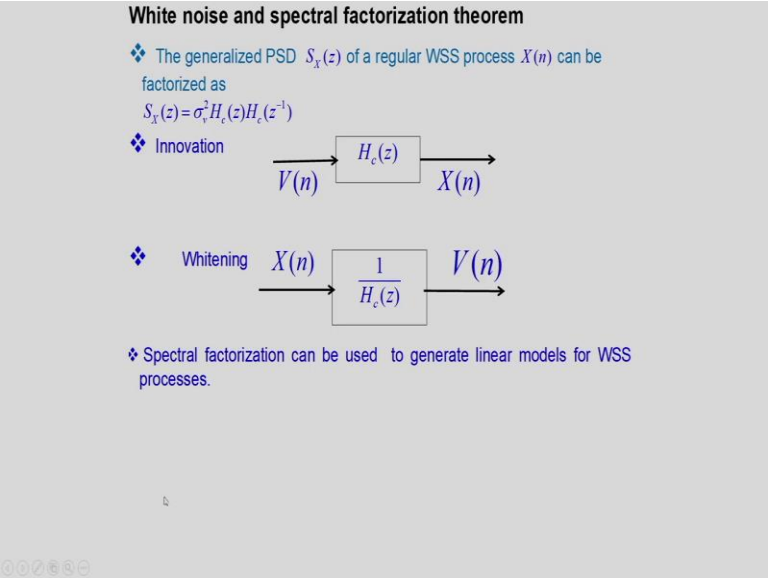
Then we defined white noise, white noise process Vn is characterized by constant power spectral density SP Omega is equal to Sigma square Omega lying between minus PI and PI. So, power spectral density is constant it will imply that autocorrelation function will be Sigma square into Delta m where Delta m is the impulse Delta m is equal to 1, 1 for m is equal to 0 is equal to 0 otherwise. And what noise is already remain and samples of VN are uncorrelated so this implies that samples of Vn are uncorrelated.

Now one important result if Vn is passed through an LTI system we get a non-white WSS, non-white, white sense stationary process. So, taking clue from this idea we establish the spectral factorization theorem that is a Sx z, the power spectral density of any random process or WSS random process under certain condition known as the regularity condition for a regular random process the generalized PSD Sx z from this result we establish that the generalized PSD Sx z that of a regular WSS process Xn can be factorized as Sx z is equal to sigma V square into Hz into Hc z inverse where is Hc z is a minimum phase transfer function.

And similarly it is Hc z will be a maximum phase and because L transfer function exceeded the minimum phase causal transformation and sigma V square is a constant which can be interpreted as the variance of a white noise. And from this result we have the innovation representation of the signal that is Xn is the output of a linear time-invariant minimum phase system with transfer function Hz which input Vn so this is the result.

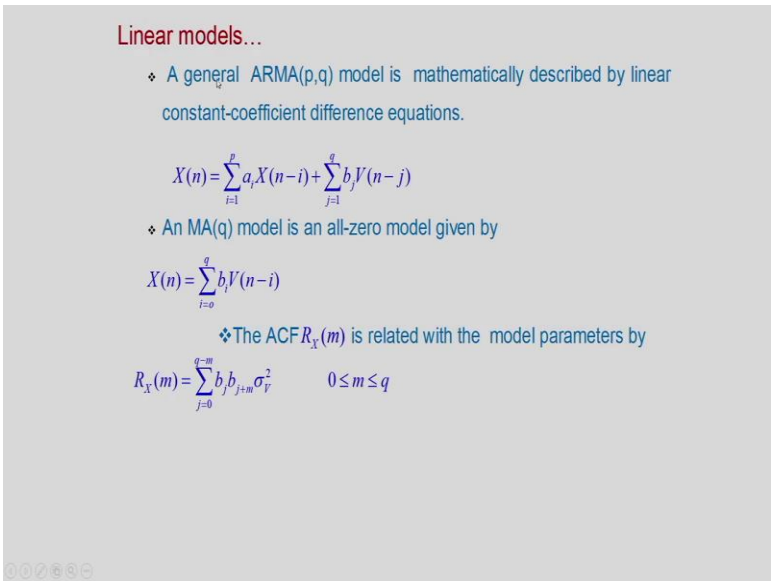
And similarly since this Hc z is invertible minimum phase we can pass Xn through one by a filter up transfer function 1 by Hc z and we will get back Vn this is known as whitening a widening a WSS process.

(Refer Slide Time: 32:14)



Now this is the result important result now spectral factorization can be used to generate linear models for WSS process. So, we can pass Vn through a filter to get Xn this is the WSS process therefore WSS process can be expressed in terms of this filter and the input Vn that way we get the linear models for WSS.

(Refer Slide Time: 32:50)

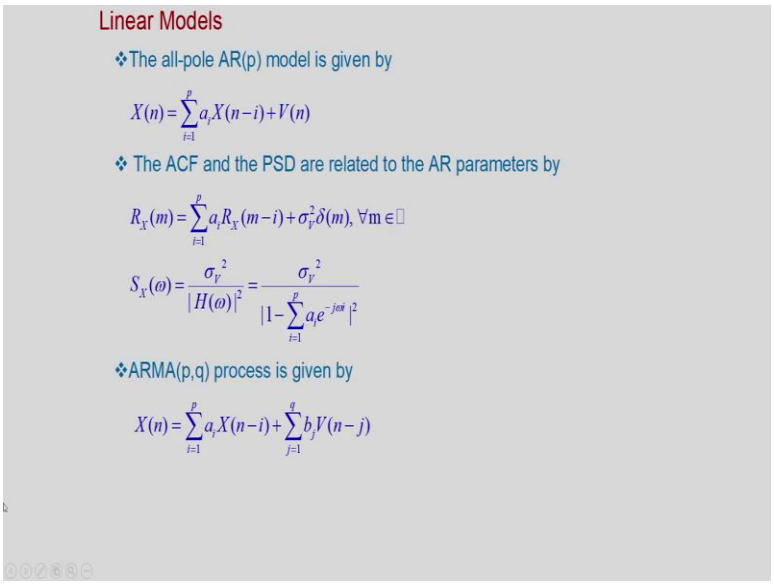


We discussed about linear models is general ARMA pq model is mathematically described by a linear constant-coefficient difference equation that is Xn is equal to summation ai into X of n - i i going from 1 to p + summation B j into V of n – j j going from 1 to q so this is the model ARMA p q this is the p term this is the q term, this is known as the auto regression and this is this part is known as the moving average.

And for particular case and MA q model is in all the other model given by Xn is equal to summation bi into V of n - i only V components are there white noise components are there so it is a linear combination of white noise, it is a linear company or combination of white noise samples. Now this is the MA q model and it is autocorrelation function Rx m is related to the model parameters by this formula this we can establish easily that Rx of m is equal to bj into bj + m sigma v square for j going from 0 to q – m.

And here m lies between 0 and q and Rx of - m we can find out to be Rx of m - m is also equal to R x of m therefore here we can write instead of m mod of m. So, this is the autocorrelation function of a moving average process.

(Refer Slide Time: 34:48)

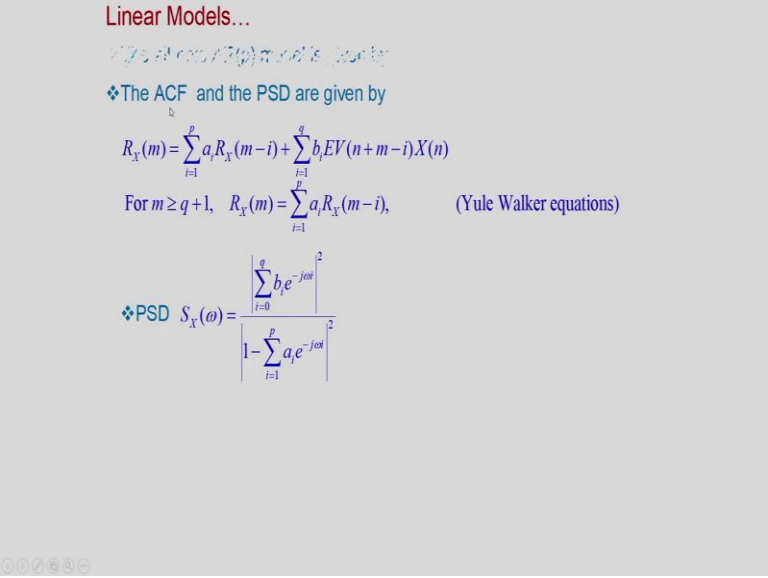


Similarly we define the all-pole model AR p model this is given by Xn is equal to summation ai into X of n - I + Vn i doing from 1 to p, so this is the AR p model. The ACF and the PSD are related to the AR parameters by this relationship autocorrelation function is satisfied these relationship Rx of m is equal to summation ai into Rx of m – i i going from 1 to p + sigma v square delta m where m is their integer m belong to z.

So this is the autocorrelation function related with the model parameters. Similarly we can find out the power spectral density also this is Sxy Omega this is given by sigma V square by mod of H Omega Square and this is this we can write Sigma V square divided by mod of 1 - summation ai e to the power - j Omega i are going from 1 to p whole square. So, this is the power spectral density of an AR p model.

So, if Xn is a AR p power spectral density is given by this and for an ARMA pq process this is the model there is a p parameter there is a q parameter so that way Xn is equal to summation ai x n - i i going from 1 to p plus summation bj V n – j j going from 1 to q so this is the ARMA pq model.

(Refer Slide Time: 36:35)

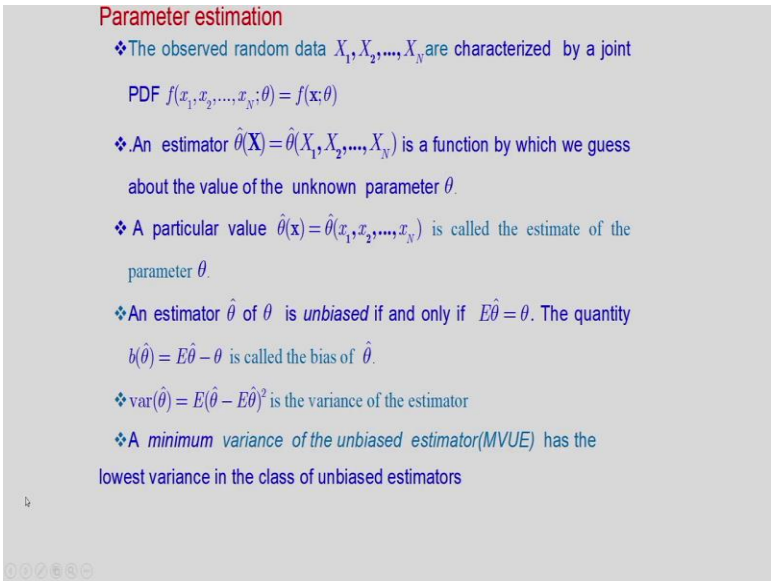


And the autocorrelation function and the power spectral density of the ARMA pq model is given by this relationship Rx of m is equal to summation ai into Rxof m - i i going from 1 to p plus the contribution from the moving average part this is given by this summation bi into EV of n + m - i into Xn i going from 1 to q and we observe that for m greater than equal to q + 1 this Rx of m will be simply given by this part summation ai into Rx of m – i i going from 1 to p this is the Yule walker equation for ARMA p, q model.

And power spectral density since it has two part will be given by Sx Omega into numerator is the moving average part summation bi into e to the power - j Omega i i going from 0 to q whole square summation bi into e to the power - j Omega i i going from 0 to q mod whole square divided by so PSD Sx Omega is given by mod of the numerator polynomial square that is summation bi into a to the power j Omega i i going from 0 to Q and then mod square.

So this is the numerator part similarly denominator part is contribution from the moving average part and this is given by the norm of 1 - summation bi into e to the power - j Omega i i going from 1 to p whole square so this is the power spectral density for a ARMA pq model. So, that we saw the modeling of random process in terms of linear model needs to discuss parameter estimation.

(Refer Slide Time: 38:54)



The parameter estimation problem is like this given the observed random data X1 X2 up to Xn and the probability model that is f x1 x2 up to xn as a function of theta this is the probability model and random data is modeled by this probability. In parameter estimation the observed random data x1 x2 up to xn are characterized by a joint PDF F x1 x2 up to xn as a function of theta that is f H theta.

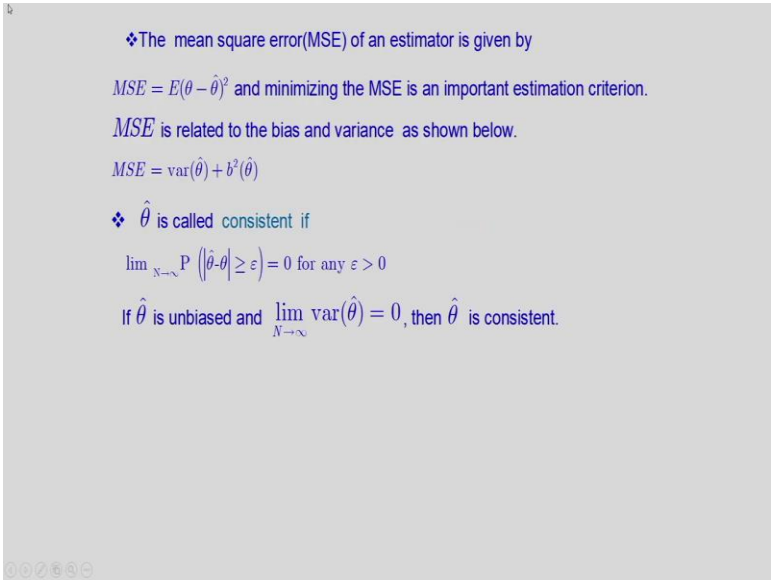
And estimate our theta hat X that is theta it is a function theta hat X 1 X 2 up to Xn is a function by which we guess about the value of the unknown parameter theta. So, using this functional relationship we will find out some acceptable value for the unknown parameter theta this we denote it by theta hat X. A particular value of the estimator that is theta hat small x that is equal to theta hat small x 1 small x2a small xn is called an estimate of the parameter theta.

So theta hat x vector x that is the estimation rule and here it is an estimator and if we consider particular values of the random vectors that is x 1 x 2 up to x N then we have an estimate of the parameter. So, theta hat small x estimate theta hat we got is the estimator. And estimator theta hat of theta is unbiased if and only if E of theta hat is equal to theta hat that unbiased estimator we define. And if it is a biased estimator we have the bias term E of theta hat is equal to E of theta hat - theta is called bias of the theta hat.

We also define the variance of theta hat it is given by E of theta hat – E of theta hat whole square that is expected value of theta hat - E of theta hat whole squared this is the variance of theta hat. So, one very desirable properties minimum variance unbiased estimator that is the estimator is unbiased and the variance is lowest among the class of unbiased estimator. So, a minimum variance unbiased estimator MVUE has the lowest variance in the class of unbiased estimator.

So, we define one important property which is most desirable that is minimum variance unbiased estimator MVUE. So, if the estimator is unbiased and it has the variance lowest among the variances in the class of unrest estimator then it is known as the minimum variance unbiased estimator.

(Refer Slide Time: 42:24)

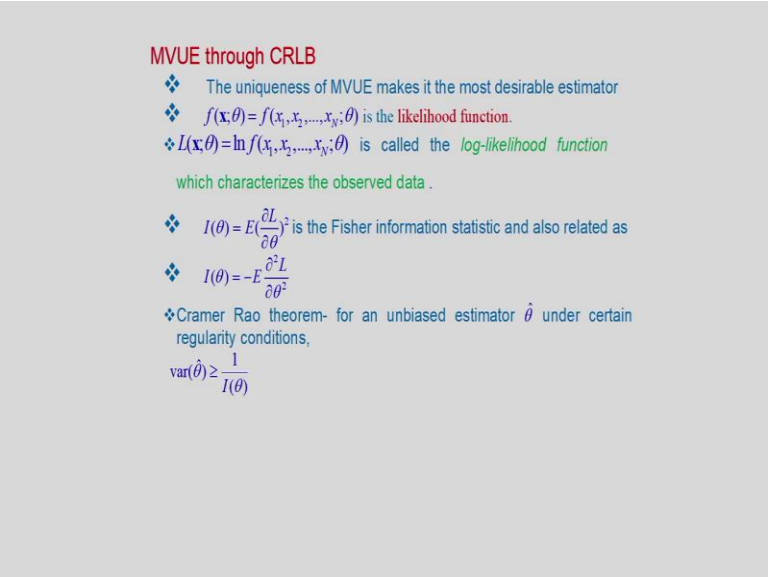


We also define the mean square error MSE of an estimator given by E of theta - theta hat whole square that is the MSE and it is related to variance and bias by this relationship MSE is equal to variance of theta hat + biased theta square and some asymptotic property we define theta hat is called consistent if limit probability that theta hat minus theta is greater than any Epsilon this probability goes down to 0 as N tends to infinity that means the probability that theta hat will be deviated from theta by any arbitrary amount as N tends to infinity is equal to 0.

This probability will converge to 0 that way then we will say that theta hat is a consistent estimator. In that case that means theta hat will be very close to the original parameter. Now

particularly if theta hat is unbiased this definition of consistency can be simplified if theta hat is unbiased then and limit of variance theta hat as N tends to infinity is equal to 0 then theta hat is consistent. So, in the case of unbiased estimator this is the test for consistency variance of theta hat is equal to 0 as N tends to infinity. Limit of variance of theta hat is equal to 0 as N tends to infinity.

(Refer Slide Time: 44:14)



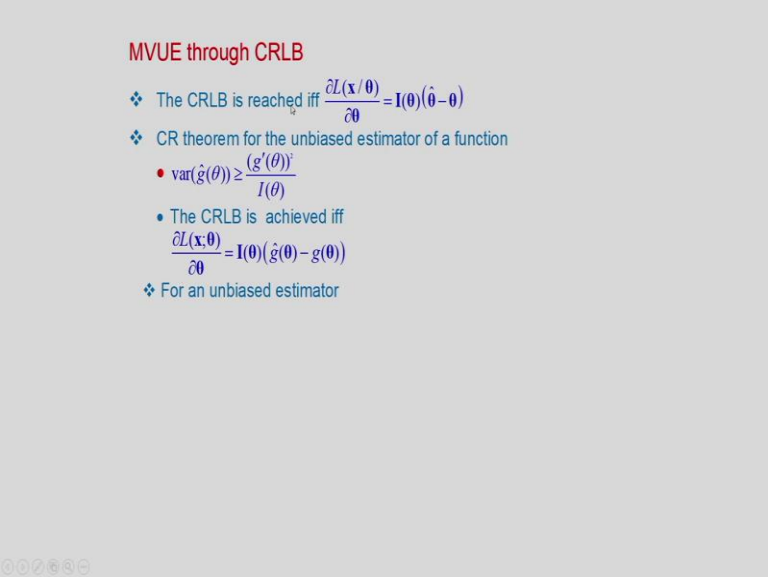
And we thought we established that a very desirable property for estimation is MUE minimum variance unbiased estimator property and we search for this MVUE we first test was true Cramer Rao lower bound. The uniqueness of the MVUE the uniqueness of MVUE makes it the most desirable estimator minimum variance unbiased estimator is the most desirable estimator because it is unique. Now to find the MVUE through CRLB we need the likelihood function that is f of x theta given by this, this is design density as a function of theta and this is called the likelihood function.

And then we defined a log likelihood function log of likelihood function as a function of theta this is called a log likelihood function likelihood function and log likelihood function. they characterize the observed data. Then we define one parameter known as the Fisher information statistic I theta is equal to E of that is log likelihood function E of del del theta whole Square. This is the Fisher information statistic and we showed that this is equal to - E of del L del theta

square either we can take del Del theta square and then take the expected value or the del L Del theta square and then take the expected value with a negative sign.

So these are the definition of Fisher information statistics and Cramer Rao theorem state that for an unbiased estimator theta hat under certain regularity conditions variance of theta hat is always greater than equal to 1 by I theta where I theta is the Fisher information statistic so that way it gives a bound on the variance how much smaller variance we can obtain.

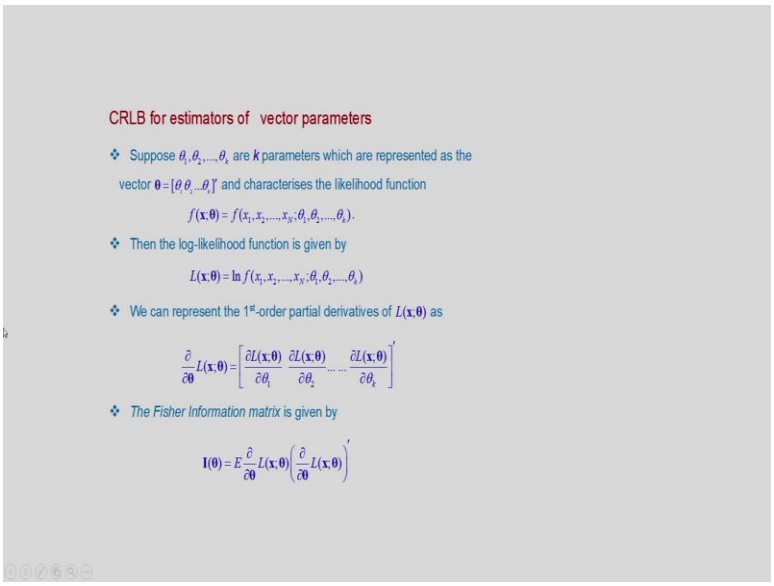
(Refer Slide Time: 47:00)



Then this here will be is reached if and only if this partial derivative of log likelihood function can be written as a product of two terms I theta into theta hat - theta so this is the information statistic into theta hat - theta if the derivative partial derivative of the log likelihood function can be factorized like this then CLRB will be reached that means variance of theta, theta hat will be equal to 1 by I theta that is the under this condition.

And also we can define the CRLB in the case of a function so variance of g hat theta where he X theta is an estimator for g theta is always bound by the g dash theta whole square divided by I theta this is the case when instead of a parameter we have a function. And in this case also CRLB is achieved if del del theta is equal to I theta into g hat theta - g theta so this is the factorization relation and if it is satisfied then this CRLB is achieved and in that case the corresponding estimator will be MVUB because it is variance is minimum.

(Refer Slide Time: 48:49)

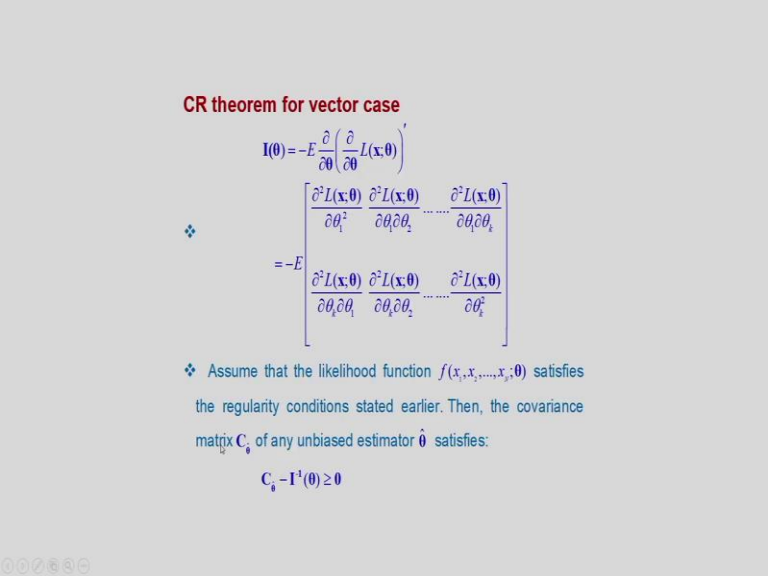


Then we discussed this CRLB for multiple parameter or vector parameter case. So, suppose theta 1 theta 2 up to theta k are k parameters which are represented as this vector and which characterizes the likelihood function. So, likelihood function is characterized by this set of parameters. Then log likelihood function will be so this is the likelihood function they this is the joint PDF, we are considering PDF only.

This is the joint PDF as a function of theta 1 theta 2 up to theta k then if we take the log likelihood function then Lx theta is equal to log of f that is the log likelihood function. And now we can take the first order partial derivative which respect to all parameters theta 1 theta 2 upto theta k and we present it as a column vector. So, that is the partial derivative of the log likelihood function with respect to theta vector that is a column vector.

And the Fisher information matrix is now given by I theta this is a matrix now that is equal to E of del L del theta that is a column vector multiplied by del del theta del L del theta transpose that is a row vector. So, if we multiply this column vector by this row vector we get a matrix and then we take the expected value of individual element to get this in Fisher information matrix.

(Refer Slide Time: 50:40)

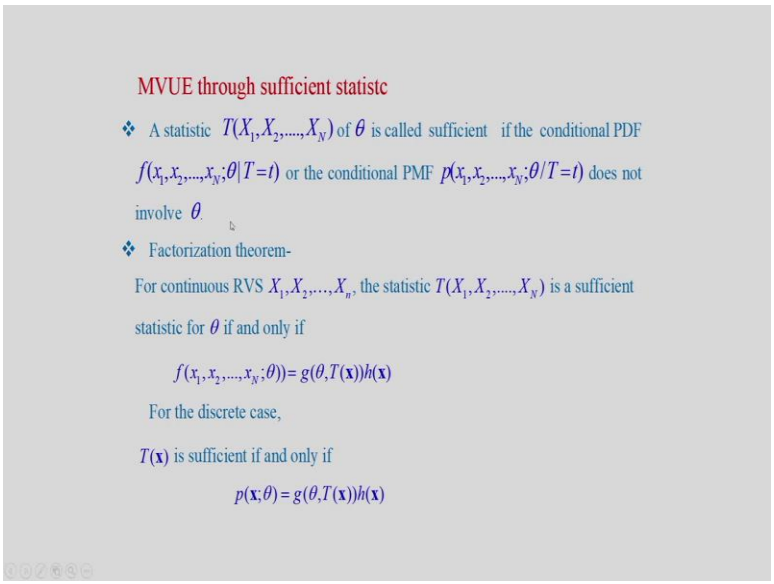


Now CR theorem for vector cases I theta we will find out the I theta metrics that is information matrix Fisher information matrix by this relationship this is the I theta matrix and this can be also shown to be equal to - E of del del theta del del theta L this can be also shown to be equal to I theta is equal to I theta matrix is equal to - of E of del del theta vector of del L del theta vector. So, that way this will get as a matrix like this its component is this second order partial derivative.

So this is the second order partial derivative matrix and then we take the expectation of the individual elements with a negative sign that is the Fisher information matrix. Now assume that the likelihood function that is f x theta satisfy the regularity condition certain regularity conditions that we have stated in the lecture. Then the covariance matrix E theta hat or when the unbiased estimator theta hat vector satisfy this relationship that is C theta hat - I inverse theta hat matrix C theta hat – inverse of matrix is always greater than equal to 0 C theta hat matrix – inverse of theta matrices always greater than equal to 0 matrix.

What does it mean? Is that did C theta hat - inverse of I theta that must be a positive semi definite matrix this is a positive semi definite matrix. This is the semi definite matrix. So, this way we defined the CRLB for a vector parameter case.

(Refer Slide Time: 53:07)

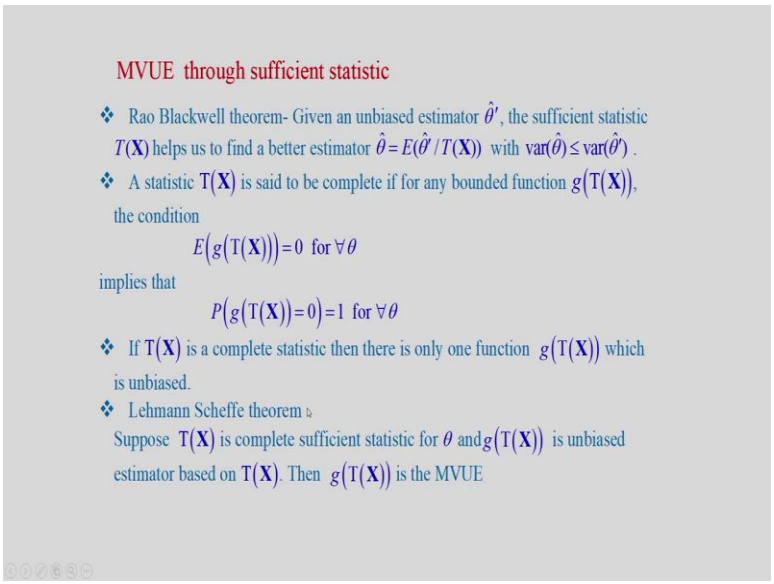


Then we derived MVUE through sufficient statistic so if statistics T X 1 X 2 up to X N of theta is called a sufficient statistic if the conditional PDF that is the PDF of x 1 x 2 up to x N as a function of theta given T is equal to T is equal to small t this is the value of this statistic or the conditional PMF p of x 1 x 2 up to x N as a function of theta given T is equal to small t does not depend on theta does not involve any theta term then we say that this statistic is sufficient statistic that with information contain indeed statistic is sufficient to estimate the unknown parameter theta.

Then we established one important theorem what is known as a factorization theorem for a continuous random variable case for continuous random variables X 1 X 2 up to XN this statistic is a sufficient this statistics T X 1 X 2 up to XN is a sufficient statistic for theta if and only if this likelihood function is a product of two terms one is g theta T x into hx product of two terms one is g theta Tx which is a function of theta and Tx other term is hx which is a function of X only it does not involve theta or Tx.

So for the discrete case similarly we have Tx is sufficient if P of X theta that is the probability mass function is a function of this quantity g of theta Tx into hx where hx is independent of any theta term or Tx term. So, that way we defined the we derived factorization theorem to test the sufficient statistic.

(Refer Slide Time: 55:21)

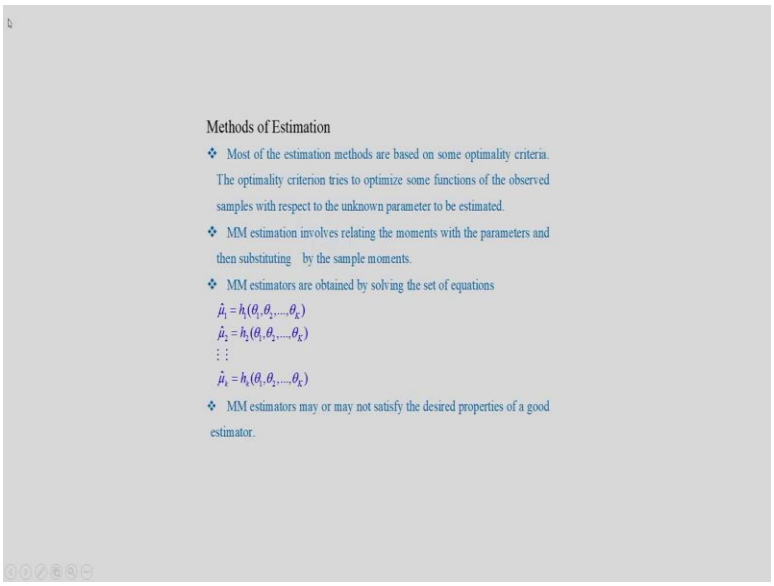


Then we discussed the law Blackwell theorem given an unbiased estimate theta hat des this sufficient statistic TX helps us to find a better estimate theta hat what is that that is equal to E of theta des TX its variance will be a less than equal to the variance of theta hat des so that is the important property we will have we can get an estimator theta hat which is given by this relationship and which is unbiased this theta hat X will be unbiased and its variance will be lower than the variance of the other unbiased estimator theta hat des.

So this theorem shows us how to obtain an unbiased estimator for theta given a sufficient statistic. Then we define the complete statistic a statistic the Tx is said to be complete if the for any bounded function g TX the condition E of g TX is equal to 0 for all theta implies that probability that g Tx is equal to 0 is equal to 1 for all theta. So, this is the condition if this is 0 expected value is 0 then corresponding probability must be equal to 1.

So this is the complete statistic if Tx is a complete statistic then there is only one a function of g Tx which is unbiased that is also important result and then we got the Lehmann Scheffe theorem. Suppose the TX is a complete sufficient statistic for theta and g Tx is unbiased estimator based on Tx then g Tx is an MVUE so this is important result and that means if it is a function of the of a complete sufficient statistic if the unbiased estimator is a function of a complete sufficient statistic then it will be MVUE. So that way we see how we can get MVUE through a complete sufficient statistic.

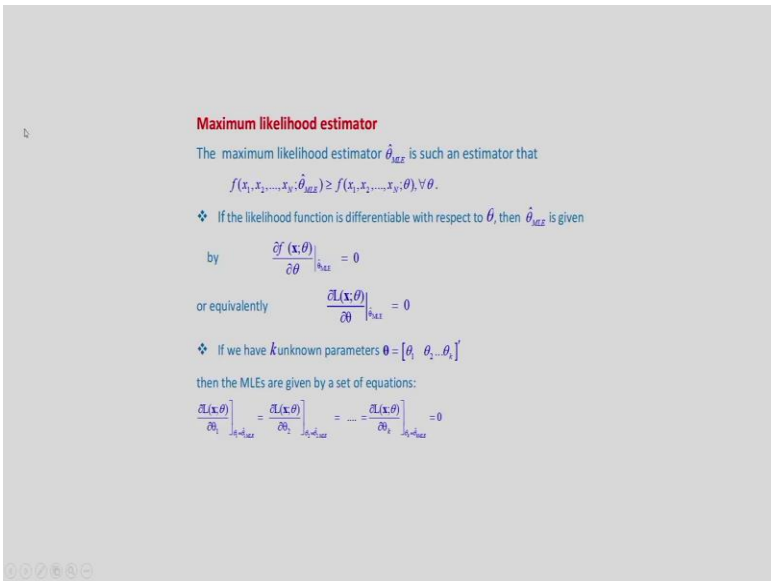
(Refer Slide Time: 57:50)



Then we discuss a few methods of estimation that means method of moments and method of maximum likelihood estimation and Bayesian estimation. So, in MM estimator method of moment estimation involves relating the moments with the parameters and then substituting by the sample moments. So, what we do we relate the moments of the distribution with the parameters and then in those relationship we substitute the moments by the corresponding sample moment and then we estimate the parameter.

So that way suppose mu 1 hat is a function of h 1 theta 1 theta 2 upto theta k, mu 2 hat is a function of h2 theta 1 theta 2 upto theta k like that we can determine mu k hat and they are related by this relationship therefore we can find out theta 1 hat beta 2 hat upto theta k hat by solving this set of equation. MM estimators may or may not satisfy the desired properties of good estimators. So they; but it is a very simple estimator.

(Refer Slide Time: 59:20)

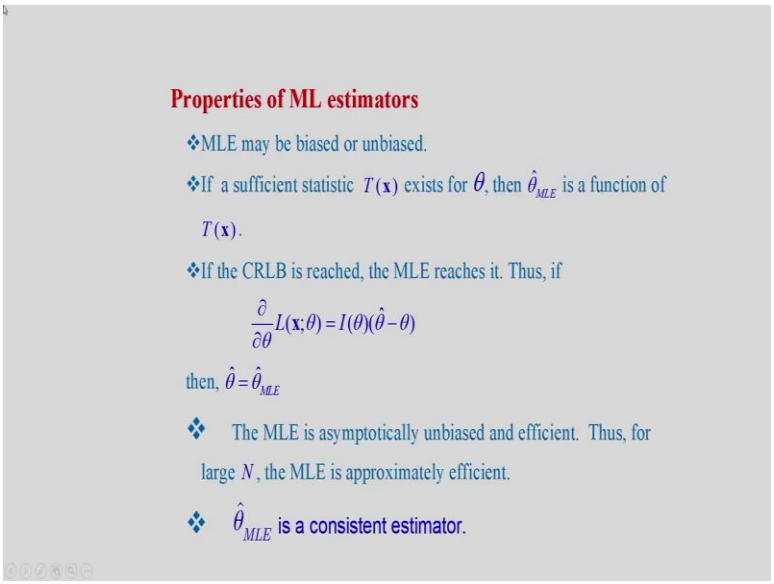


Then we discuss the maximum likelihood estimator, the maximum likelihood estimator theta hat MLE is such an estimator that the likelihood function as a function of theta hat MLE is greater than equal to the likelihood function as a function of theta for any theta. So, for all theta will condition consider this relationship that the likelihood function of theta hat MLE is always greater then this estimate is the maximum likelihood estimator.

And therefore this likelihood function is maximum for theta hat MLE this condition can be written in terms of this derivative condition, partial derivative of del f del theta at theta MLE is equal to 0. The likelihood function if the likelihood function is differentiable with respect to theta then theta hat MLE is given by this relationship del f del theta f theta hat MLE is equal to 0 or in terms of log likelihood function we can write del L del theta f theta at MLE is equal to 0.

And similarly if we have k unknown parameters theta 1 theta 1 theta vector that is equal to theta 1 theta 2 up to theta k transpose then the MLE's are obtained by this set of equation del del L del theta 1 up to del theta MLE is equal to 0 del L del theta 2 f theta 2 health MLE is equal to 0 del L del theta 1 at theta 1 hat MLE is equal to 0 del L del theta 2 theta 2 hat MLE is equal to 0 like that where we can get a set of equations.

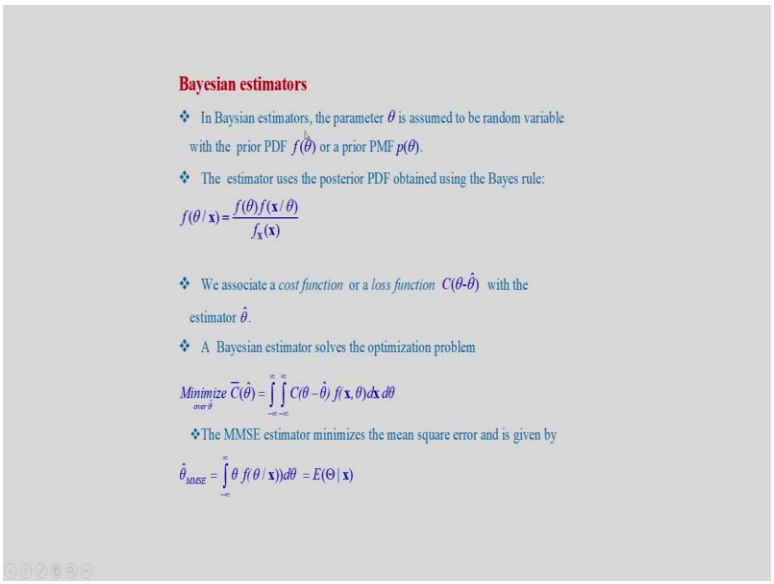
(Refer Slide Time: 1:01:25)



And we established the properties of ML estimate or MLS estimator maybe biased or unbiased if a sufficient statistic Tx exists for theta then theta hat MLE is a function of that statistic that is one important property. Again if CRLB is reached if the condition partial shall be equality CRLB equality is satisfied then MLE will reached that condition. So, if CRLB is reached than MLE recessive. So, that means in the case where Cramer Rao bounded satisfied with equality in the case maximum likelihood estimator will have the Cramer Rao lower bound for the variance.

So that therefore if del L del theta is equal to I theta into theta hat - theta then theta hat must be that MLE this is the condition. The MLE is asymptotically unaised and efficient thus for large N MLE is asymptotically unbiased for large N it will satisfy the Cramer lower bound and it is asymptotically unbiased also. Then in all cases theta hat MLE is an consistent estimator.

(Refer Slide Time: 1:03:11)



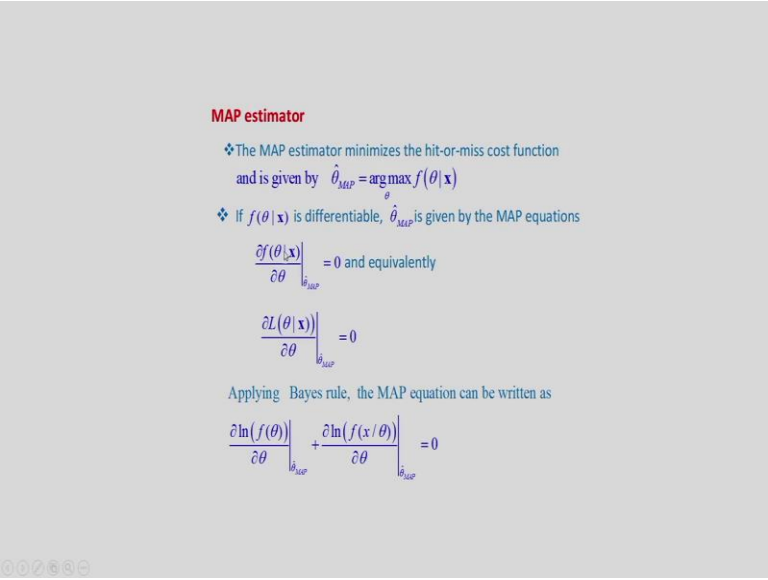
Then we discussed about Bayesian estimators in Bayesian estimators the parameter theta is assumed to be random variable so in MLE or method of moments we consider theta to be a unknown deterministic quantity but in Bayesian approach it is a random quantity. So, therefore prior PDF f theta or by prior PMF P theta must be satisfied. The estimator uses the posterior PDF obtained by this relation f of theta given X this is the posterior PDF it is given by f theta into f of x given theta divided by f x.

And Bayesian estimator we associate a cost function or a loss function C of theta - theta hat with the estimator theta hat. So, the average value of this cost function is known as the Bayesian risk we want to minimize z. A Bayesian estimate or solve the optimization problem minimize average value of theta hat that is the risk over theta hat and this is given by this relationship. So, because it is a function of X and theta therefore it is integration with respect to X and theta.

So that way this function we have to minimize there are different cost function we saw that means square error cost function. So, if we consider the cost function at the squared error cost function we will get the minimum mean square error estimate or minimum mean square error estimator minimizes the mean square error and it is given by this mean square error is E of theta hat - theta whole square.

This is the mean square error and this is minimized by the MMSE estimator and this is given by this relationship theta hat MMSE is conditional expectation theta hat MMSE is the conditional expectation of theta given X. So, that way the MMSE estimator is a Bayesian estimator which minimizes this cost function E of theta - theta whole square and it is given by this result E of theta given X.

(Refer Slide Time: 1:05:40)



Similarly we derived the map estimator maximum a posteriori probability estimator. This minimizes another cost function what is known as the hit or miss was discontent we considered. So, this is suppose the error, if error is within - Delta by 2 and + Delta by 2 then this cost is zero otherwise cost is one that is the hit or miss cost function and the map estimate or minimizes the hit or miss cost function and is given by theta hat MAP is R max theta hat MAP of theta given X this is the posterior PDF and for what values of theta this posterior PDF is maximized that is the theta hat map. So that the decorative mode of the posterior PDF.

If f of theta x is differentiable then theta hat map is given by the map equation what is the map equation and that is partial derivative of the posterity PDF theta at a map a theta hat map is equal to 0 or this we can write in terms of log likelihood function also del L del theta del L theta given X del theta at theta hat map is equal to 0. And we can apply the Bayes rule to find out f theta given X then we can show that this map equations will result into this condition partial derivative

of likelihood function at that map plus partial derivative of the posterior density at theta hat map must be equal to 0.

So this is the bay map equations in terms of the likelihood function and posterior density function, thank you.