Digital Communication

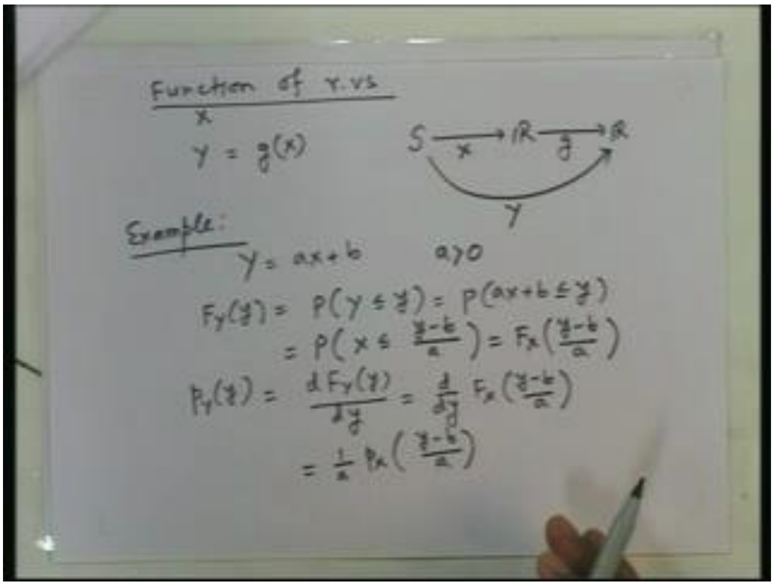
Prof. Bikash Kumar Dey
Department of Electrical Engineering
Indian Institute of Technology, Bombay

Lecture - 05
Probability and Random Processes
(Part-2)

Hello everyone. In this class we will continue our discussion on probability and random processes. In the last class, we defined sample space then event independent random variables, first random variables, independent random variables. Then, cumulative distribution function, density function of a continuous random variable then joint probability distribution and density functions. And, we also defined independent random variables. Now, in this class we will start with functions of random variables. If we have a random variable and if we take a function of the random variable with a real value, that is the function takes the value of the random variable and transforms it to another real number.

So, if it has a function the function of the random variable is also of course random. Because the random variable is has some randomness and the function will also have some randomness, the value of the function will also be random. So, the function value will be also a random variable and what about its cumulative distribution function and density function. We would like to characterize that in terms of the function and the original random variable. So, let us do that.

(Refer Slide Time: 02:28)



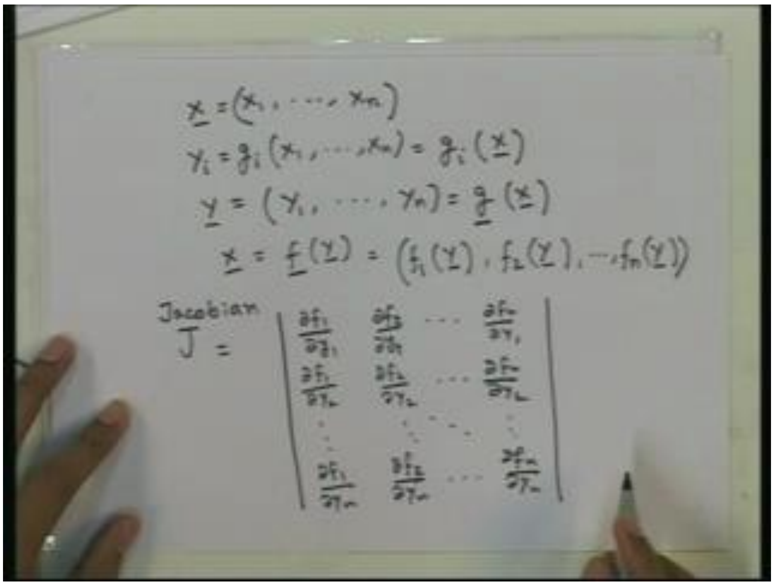
If you have a random variable X, so we are discussing function of random variables X and we consider another random variable which is a function of X. So, basically we have the sample space then the original random variable X is a function of S into R. The value is a real number then we take another function g, which converts it to another real number. So, then this function together something here goes to here and then ultimately goes to here. So, this mapping from S to R itself, this mapping which is g and X together is also a mapping of S into R and that is our random variable Y.

So, this random variable Y also will have some probability distribution and if it is a continuous random variable it will have density function. So, let us consider an example. Y equal to a X plus b and we take a greater than zero. If a is less than 0 the results are similar, we will not consider that here. So, what is the cumulative distribution function of Y in terms of the cumulative distribution function of X? That is what we would like to find out. So, F Y of y we want to find. So, this is by definition P probability that Y is less than equal to small y and; that means, probability that a X plus b is less than equal to small y.

Now, what does this mean? This means that capital X is less than equal to by minus b by a. So, this is a probability that capital X is less than equal to y minus b by a. But what is this, this is nothing but the cumulative distribution function of X at y minus b by a. So, we have expressed the cumulative distribution function of the new random variable Y in terms of the cumulative distribution function of the original random variable X, what we wanted. Now, we also want to see the relation between their density functions, if the density function of the both the random variables exists. So, let us find out.

So, P Y y density function that is what we want to find out. This is d F Y y by d y by definition is the definition. Now this is nothing but d y of this is equal to this. So, F X y minus b by a. Now, what is this is nothing but 1 by a, then the derivative of this with respect to this whole thing. So, that is nothing, but P x Y minus b by a. So, this is the relation between that density functions. Now, if we have some multiple random variables and we their some multiple functions of those multiple random variables. Then let us see what happens to their joint density functions and joint distribution functions.

(Refer Slide Time: 06:25)



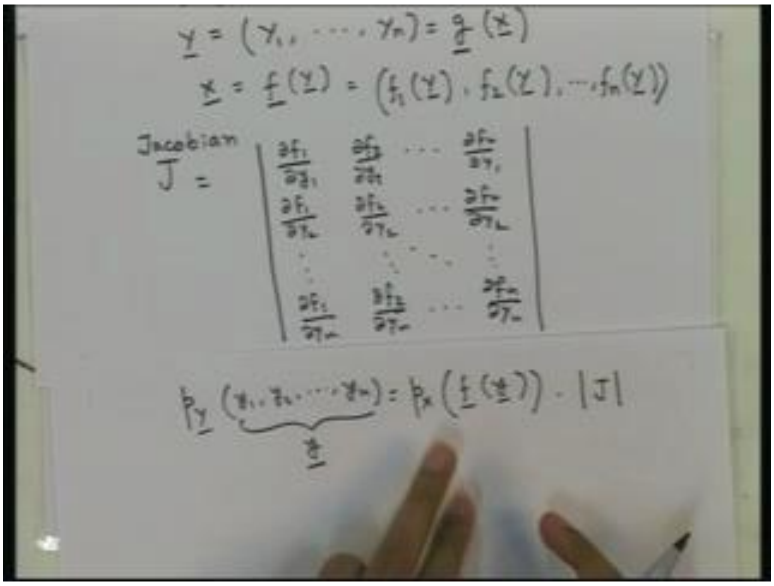
So, we have a set of random variables X 1 to X n and we will write then in short just for notational simplicity we will write them as a vector X. So, this is the vector of these random variables. And we take n number of functions of these vectors. So, the function take g i the i th function takes all these values and then maps it to one real number. So, it is a function of all these random variables. So, g i of X 1 to X n is the function of all these random variables that is it is a function of this random vector. And, this function is also a random variable because it is a function of some random variables.

So, this function is called the i th random variable Y i. Then, this is this can be also written as g i of the vector X. And, this there are i n number of such functions g 1 to g n. So, we get Y 1 to Y n and we denote those random variables new random variables by a vector Y. So, this is our new random vector which is a function of the old original random vector. And, we assume that the set of functions is invertible meaning by if we know Y the value of Y you can get back what is the value of X. So, in that case if that happens. So, this is basically we denote these functions together as a function vector g of X.

So, this random vector is a function of the original random vector. So, that each component here actually is obtained by one function of that random vector. And this g is invertible and we assume that the inverse is called f. So, X is f of Y, so is an inverse function. Now, so this is basically, so every component the first component of X will be X 1 will be obtained by the first component of the function, if there is one function f 1, which gives us X 1, f 2 gives us X 2 and f n gives us X n. So, these are the component functions. Now, given any such transformation any such transformation of vector, we can compute what is called Jacobean of the transformation.

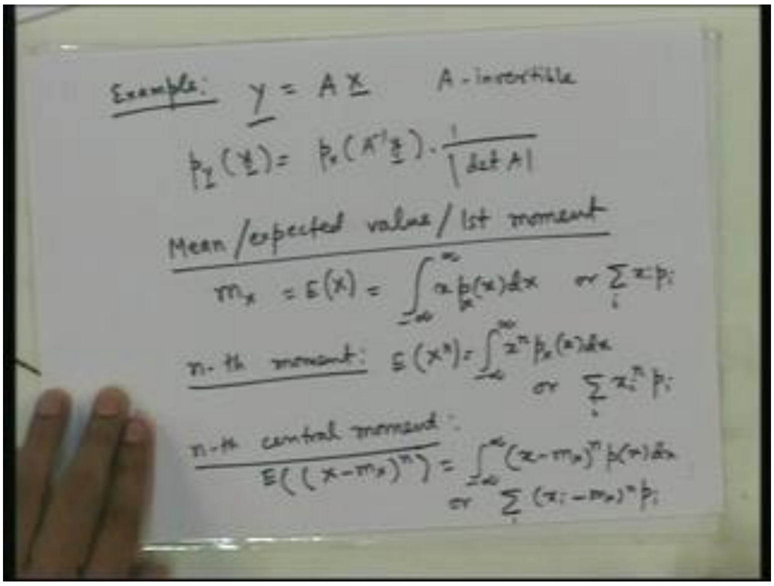
What is the Jacobean? Jacobean is the determinant of the following matrix. Jacobean J is the determinant of Del f 1 Del Y 1. Remember that this f 1 is a function of this whole vector, this f 1 of Y 1 Y 2 and till Y n. We are not writing those arguments it is implied
So, Del f 2 by y 2 Del Y 1 Del fn by Del y 1, Del f 1 by Del y 2 Del f 2 by Del y 2 Del fn by Del y 2, Del f 1 by Del y n Del f 2 by Del y n Del fn by Del y n. This is the determinant of this matrix is the it is called the Jacobean of this transformation f. Then we, one can show that if we know the density function of X, we can find the density function of the new vector random vector Y in terms of it in the following way.

(Refer Slide Time: 11:06)



p y the density function of y at y 1 y 2 y n is p X, so this is the vector y. Take the inverse, but you have to multiply by the Jacobean. So, this Jacobean is taken Jacobean of the inverse transform. So, this is the original transform is g. We take inverse transformation and we take the Jacobean of that. So, this density evaluator of the function inverse and then that has to be multiplied by the Jacobean of the inverse transform. And that quite confirms with what we did for the simple function Y equal to a X plus b. Because here we actually did that the computed the inverse X in terms of Y this and there we multiply b Y 1 by a. What is 1 by a? It is the Jacobean of the inverse transformation inverse transformation is X equal to y minus b by a y minus b by a, the derivative of that is 1 by a. So, we have multiplied b Y 1 by a.

(Refer Slide Time: 12:42)

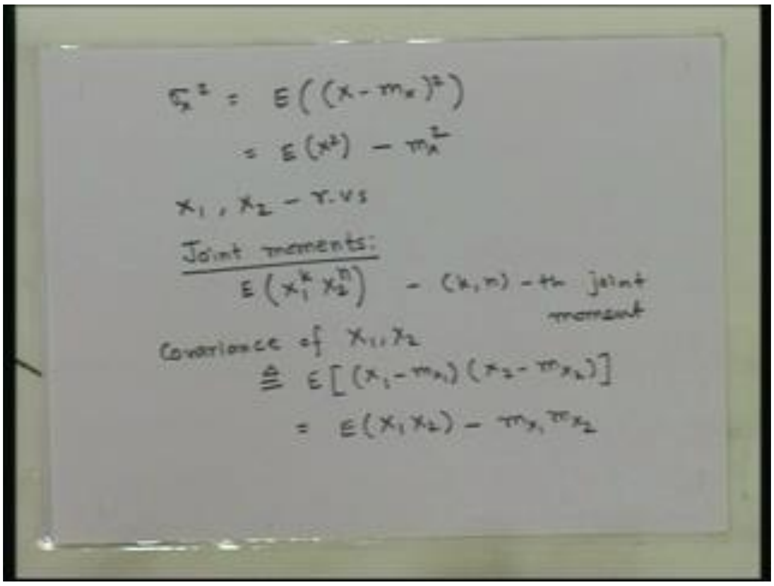


So example; A special case is this simple mapping, via this vector is a matrix invertible matrix times X. A has to be invertible because we need that the inverse transformation is the inverse transformation exist. So, this A is inverted. Then p Y y is p x A inverse y times 1 by determinant A. It is a very simple nice formula and one can check that this is what happens because what is the inverse transformation X equal to A inverse Y. And what is Jacobean of that transformation it will be just determinant of A inverse. So, determinant of A inverse is nothing but 1 by determinant of A. So, we do that.

So, now there also important quantity is defined for some for random variable some average quantities. So, we will define those quantities. So, mean or expected value is called mean or expected value or first moment of a random variable. m x the mean of X is also denoted by E X. It is basically minus infinity to infinity x p x d x if it is a continuous random variable then density function exists. So, we can compute the average value in this way while taking the average we cannot simply integrate, but we have to multiply by the density function. And if it is discrete random variable we can take the summation for all the values and their probabilities, product of the value and probability. Similarly, n th moment is defined as simply expected value of X power n.

So, this will be simply X power n instead of X here. Then same as before this or if it is discrete random variable we will have x i power n p i. Now, this is n th moment, but we can compute we can define what is called n th central moment. In X there is some bias, this is the bias the it is actually it has some mean value may not be zero the average value may not be zero. So, we actually bring that average value to zero by subtracting the mean value and then take the moment that is called the central moment. So, it is defined as X minus m x. We subtract m x from X and this becomes a new random variable which has mean m x. And, then this power n you take the expectation. So, this will be x minus mx power n p x d x or if it is a discrete random variable you take xi minus mx power n p i summation we step into equation.

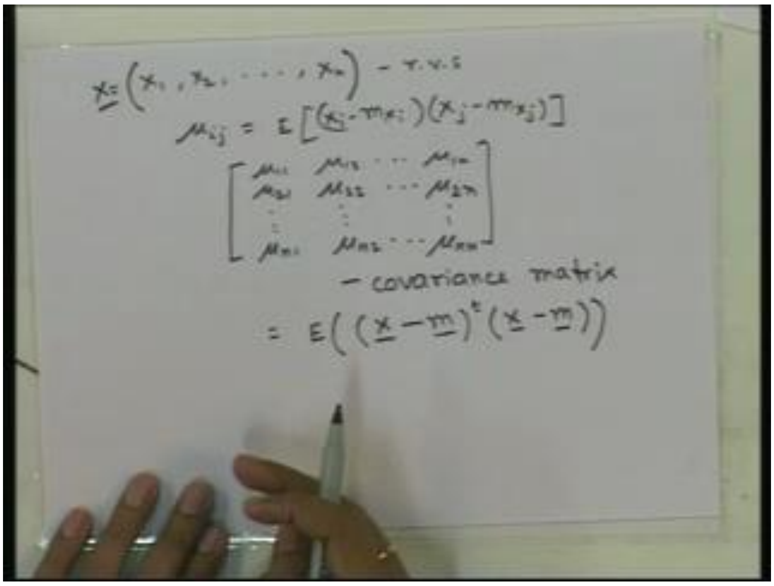
(Refer Slide Time: 16:54)



Variance is denoted by sigma X square. It is defined as X expectation of X minus m x square. It is the second central moment second central moment is called the variance and one can do with. I can show after some manipulation that, this is same as expectation of X square the central moment the second moment minus the square of the first moment. Now, if we have two random variables X 1 and X 2, then the joint moments can be defined as X 1 power k X 2 power n the product the expectation of the product. Now, this expectation we will be taken; obviously, with respect to the joint distribution of the two random variables.

The either if there are continuous the joint density function of if they are discrete take the joint probability distribution probability mass function. So, this is the k n th joint moment. Previously, we had one random variable we had one index we called it n th moment. Now, we have two because there are two random variables and call it k n th joint moment. Now, similarly, we can compute the central moments we can subtract the mean values and then get the central moments. And, covariance is of special importance of X 1 and X 2 it is defined as it is defined as expectation of X 1 minus m x 1 times X 2 minus m x 2. And one can show that this will also be the same as expectation of X 1 X 2 minus m x 1 m x 2. So, this is the covariance of X 1 and X 2 is two random variables. And, if we now have instead of two random variables say n number of random variables.

(Refer Slide Time: 19:52)

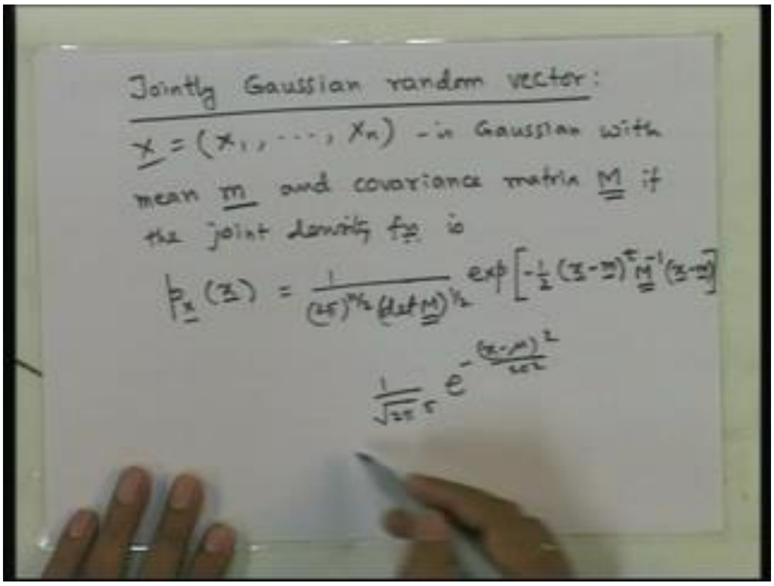


So, X 1, X 2, X n the n number of random variables. Then, we can define n square number of covariance between any pair of these random variables. So, for any i and j in this range 1 to n, we chose any two random variables. They may be same also; we can compute the covariance between those two random variables. If the random variable is same you chose the same random variable two times, you will get this as the variance of that random variable. So, Xi minus m x i times X j minus m x j and then we can take the matrix with this components and that is called the covariance matrix.

So, we take this matrix mu 11 mu 1 2 mu 1 n, mu 2 1 mu 2 2 mu 2 n. So, on till mu n 1 mu n 2 mu n n. Then this is the covariance matrix. Basically, all the covariances are written as components of a matrix and that this matrix is very useful. And, one can also write this matrix as in a vector notation very nicely as expectation of the vector X. So, this is the vector, this is the X vector and then this mean vector is the basically a vector with m x 1 m x 2 and m x n there is a mean vector. Take the transpose that is now you get a column instead of row vector, you make it column then you multiply that by this same thing, but this is column, this is row. So, you get a matrix at row multiplication.

Then every component of the matrix is a product of this type and then you take the expectation of all the components you get this matrix. So, this is the covariance matrix. Now, one particular type of random vector is very important in this course and we will just write that down that is Jointly Gaussian random vector.

(Refer Slide Time: 22:40)



So, join so this is a set of vectors a set of random variables. So, X is X 1 to X n, n number of random variables and this is called a Gaussian random vector this is Gaussian with mean. So, this since this is a vector of random variables the mean will also be a vector mean of X 1 mean of X 2 and. So, on till mean of X n. So, that together they will form a vector. So, mean m is a vector and the covariance matrix M. So, this is a Gaussian with mean m and covariance matrix M. If the density function, if the joint density function is given by p X at the vector x is 1 by 2 pi power n by 2 determinant M power half. So, this is the covariance matrix e power exponential of minus half x minus m transpose matrix M inverse and then x minus m.

One can check that the single random variable that we defined comes as a special case of this. So, if you take one random variable in this that is n equal to 1 what happens? This mean vector becomes the mean of the random variable this covariance matrix becomes the covariance is the variance of the random variable sigma square. Then what is the determinant of that matrix determinant of the scalar sigma square? It is sigma square itself. Then when you take power half you get sigma here and n equal to 1. So, root over

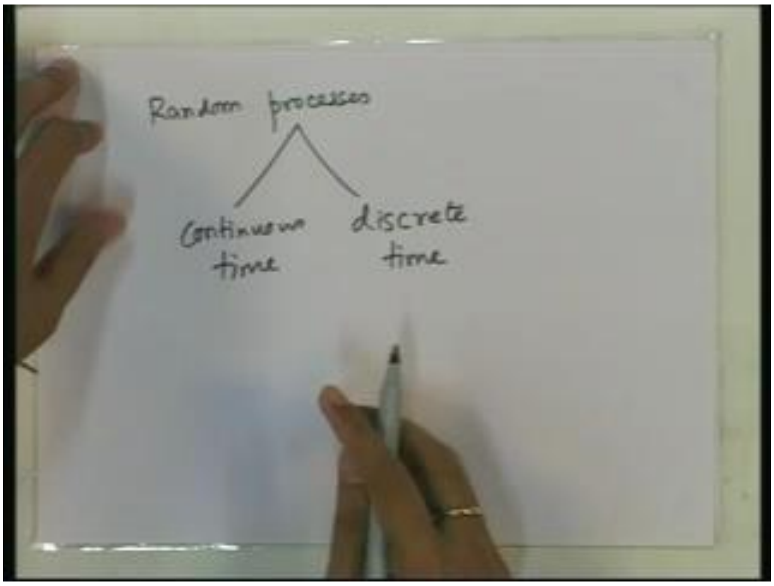
2 pi times sigma and what you would get here it matches with the density function that we considered before.

And, so we had Gaussian single variable Gaussian random variable density function was 1 by root over 2 pi sigma e to the power minus x minus mu whole square by 2 sigma square. So, mu is our m now, it is a single component. So, here it matched now here half x minus mu transpose is same as itself. Because, it is a single component vector is a scalar and M inverse is nothing, but sigma square inverse 1 by sigma square and half comes here, minus here and the square this and this together gives you square. So, it matches with this for a single random variable. So, this is a special case of this. Now, we define what is called the stochastic process.

So far we are discussing random variables. Now, random variable is a single random quantity where as in practice in the in nature many signals that we get are random. So, many random functions are obtained or random functions or random signals are obtained in nature. So, they are not they do not fall in this category. Because they are not simple one value, but the function has many values at different time. So, for example, the noise generated of the register as I said before that if you do not now consider only one time instance. The voltage measure at one time instance that voltage might be changing with time. So, it is a function of time actually. So, that is a random signal. So, that is such a random quantity random function is called random process.

So, another example may be the output of a source generating a signal random signal. For example, if there is a microphone the output of the microphone there is some voltage generated depending on the content of the speech. So, that voltage is a random signal and those are modeled as random processes. So, those are called the random processes. And, the time in the random process may be a may be discrete or continuous. It may be continuous, it is mostly most of the times the natural phenomenon that we observe they generate continuous time signals. So, continuous time random process, but we can also sample then it becomes another random process with discrete time. So, the random process may be discrete time or continuous time. So, we have random processes continuous time or discrete time.

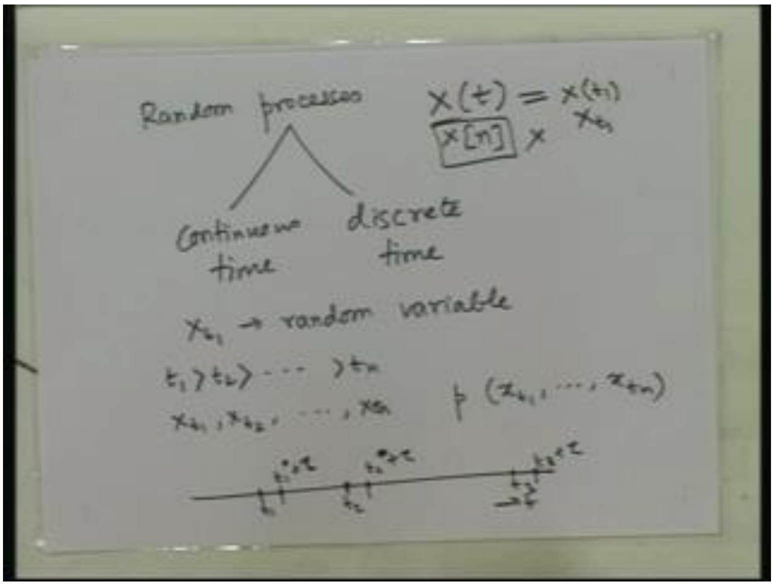
(Refer Slide Time: 28:55)



And, we observe if we measure the function for a particular resistor. If we see what function we get the how the voltage is changing and if you plot it then we get one realization of the random process. It does not tell you all that is there about the random process. We are only seeing one particular instance of the one particular realization of the random process or stochastic process is also called stochastic process. So, such a function is also called a sample function. It is like a particular outcome of an experiment. we are not observing all the outcomes we are observing one particular outcome such as in this case of random process we called that a single realize a sample function. If the observation is a function. So, now, if we now take the value of a random process at a particular time at a fixed time t 1.

So, the random process will be denoted by X t or X n if it is a discrete time random process. So, if we now for the rest of the class we will consider mostly a continuous time random process. So, this notation will not be used. And, now if you fix the time t to be t equal to t 1, let us say then which denote that by X t 1 or sometimes by X subscript t 1 this will mean same thing as this. So, now if you take this is also a random variable because the functions are random. So, the value at t equal to t 1 is also random. So, at any time the value is random. So, it is a random variable it is a real value it takes real value. So, if you take any time the random process value of the random process at that time is a random variable. Value of the random process at any other time is another random variable. So, if you take n number of times you get n number of random variables. So, X t 1 for any t 1 is a random process a random variable.

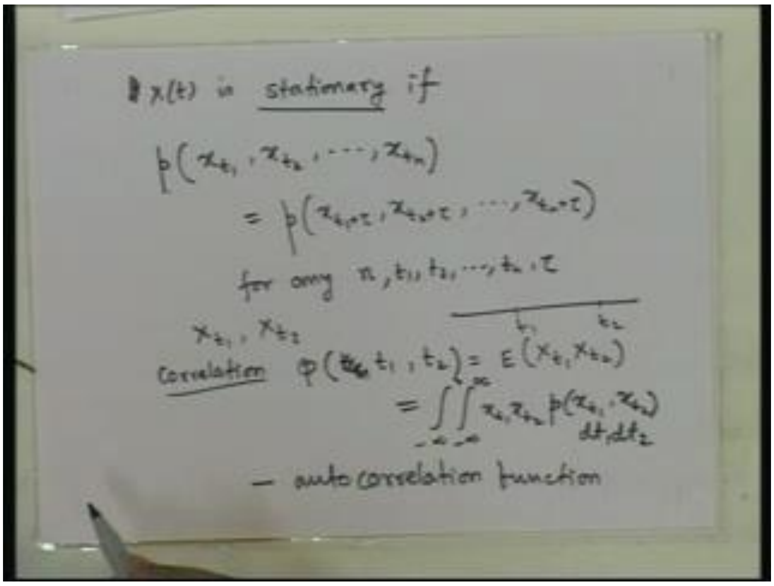
(Refer Slide Time: 31:39)



And if you take some times t 1 t 2 t n, n number of times in decreasing order then we can take X t 1, X t 2, X t n they are random variables. And, how do we characterize such n number of random variables; obviously, by the joint p d f if there if we are considering a continuous time random process continuous value random processes. The value is continuous then we have a probability density function or as which is abbreviated as p d f. So, we have a joint p d f which is denoted by p X t 1 to X t n. So, now if we, so this density functions if it does not change if we shift all the time instances by the it is same amount.

So, if we have if we chose a. So, this is time if we chose this and this time t 1 t 2 t 3. Now, if we shift all the times instances by say one second shift to here; shift to here shift to here. We take t 1 prime t 2 prime, that is t 1 plus say tau, t 2 plus tau and t 3 plus tau. Then the density function of these random variables together these 3 random variables is same as the density function of these 3 random variables. If that happens then we call that we say that the random process is stationary. That means the behavior of the random process does not change with time the statistical properties of the random process does not change with time. You take the joint distribution of some random variables split at different times if we shift all the time instances by the same a amount then the density function does not change the joint density function does not change.

(Refer Slide Time: 34:23)



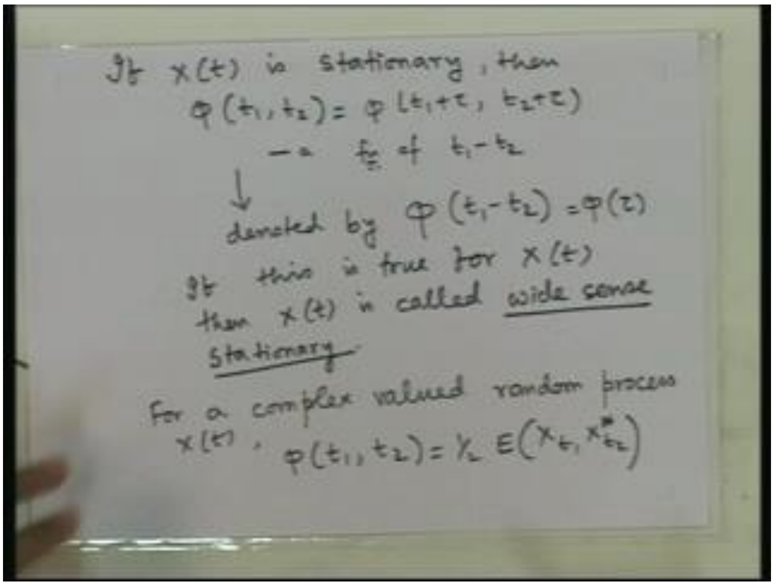
So, in mathematical notation it means that P. So, X t is called stationary. If so X t is called stationary if p x t 1 x t 2 x t n. Actually, we should write x t 1 capital X t 1 capital X t n here because this is the density function of joint density of those random variables evaluated at these values. But for to avoid cumbersome notations we are not writing them here. So, this should be if this is same as P x t 1 plus tau x t 2 plus tau x t n plus tau. If these 2 are the same density functions for any number of points n any times chosen that is t 1 t 2 to t n and any shift chosen any shift tau. So, then this random process is called stationary.

And, this is called strict stationary. In fact, this is called a strictly stationary if this condition is satisfied. There is a weaker definition that we will come across later that is called wide sense stationary process. Often that is a sufficient condition as in many applications. We do not need stationary process, but we need wide sense stationary process most of the times. Uh Okay. So, we have defined what a stationary process is. Now, correlation between two random variables x t 1 and x t 2 the correlation between these two random variables is defined called as phi t 1, t 2 and this is nothing but expectation of X t 1 X t 2 is the definition.

Now, this is; obviously, if you write it in terms of integration it is nothing, but minus infinity to infinity, minus infinity to infinity X t 1 X t 2 P X t 1 X t 2 d t 1 d t 2 the expectation of the product. And, this is a function of t 1 and t 2 and we have chosen 2 random variables at 2 different times. But if you change the 2 times the correlation changes, so this correlation changes. So, this is a function of two variables t 1 and t 2 and this is called the autocorrelation function, because we have taken the same random process and taking 2 different time instances and compute the computing correlation between the random variables at those 2 different time instances.

So, this is called the autocorrelation function of the random process that we are considering. So, this is called autocorrelation function. Now, if the process random process X t is stationary then this will depend only on, this will not change if I shift both the times by the same quantity. So, if the process is stationary then what will happen is.

(Refer Slide Time: 38:41)



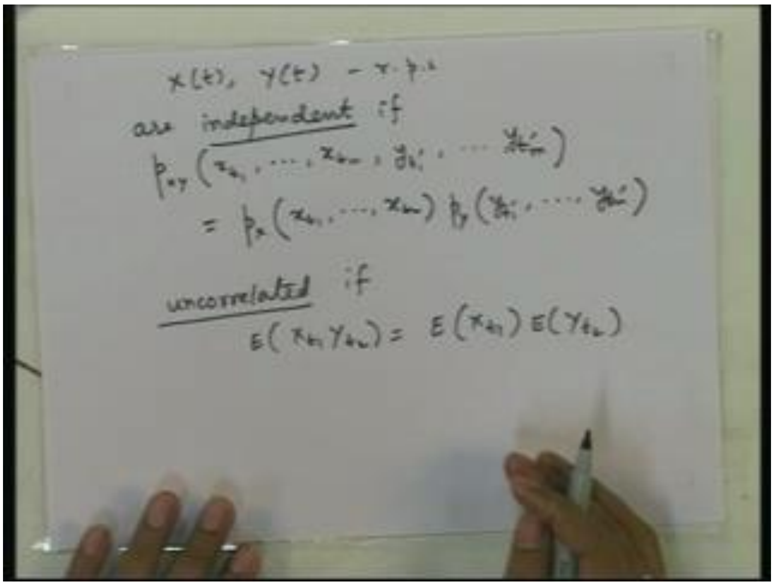
If the process is stationary then phi t 1 t 2 will be same as phi t 1 plus tau t 2 plus tau for any tau. So, one can check that a very easily from the definition because this is invariant on the shift of the times. So, this is actually a function of how much gap is there between t 1 and t 2. Where, the gap is same if you take any 2 of the times instances. So, that the gap is same then we can say that this these 2 time instances is the shift is the shift of these 2 time instances. So, it is the this is a function of the gap between the times t 1 and t 2. So, this is so this is a function of t 1 minus t 2.

So, it is denoted by phi t 1 minus t 2. So, we can call it a single variable it is a function of only the gap and if that gap is called tau then it is nothing, but phi tau. So, it is a function of the gap between the random variables. And, So for stationary process the autocorrelation function is a function of only the gap between the time instances. Now, this condition may be satisfied even by a process which is not stationary. So, that kind of process is called wide sense stationary. So, if this is true for X t even if it is not stationary, but it is true for X t then X t is called wide sense stationary. So, it is a wide sense stationary process.

Now, we have been discussing from the beginning only the real random variable. That is the mapping is into the real number set of real numbers. But we can also consider complex value random variables and it is everything is same except for some. Say, minor changes in the definitions like the definition of covariance and correlation. So, there whenever there are two random variables coming in the expectation it one will be conjugate that is that is all the difference. So, for a complex valued random process X t then X t has a real part and imaginary X t is say at plus b t I b t.

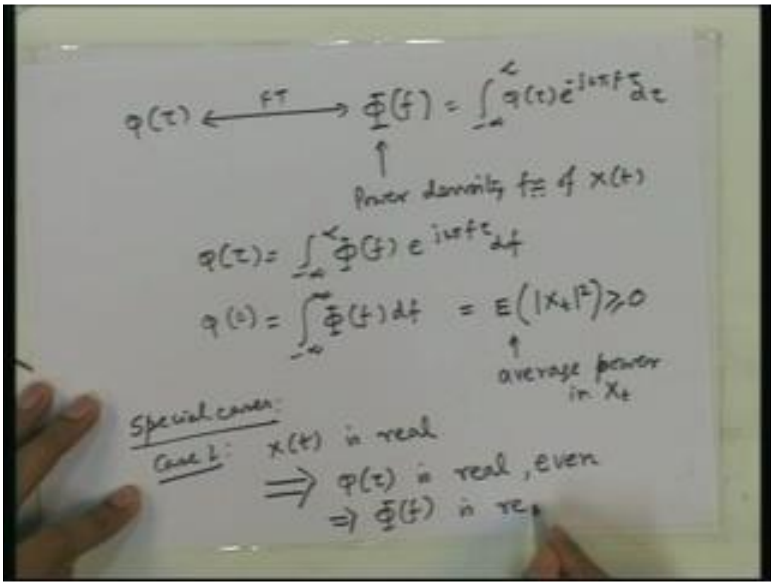
So, then the then the autocorrelation function will be defined as, the correlation itself will be defined as expectation of X t 1 times not X t 2, but X t 2 conjugate. So, this half is not really very important this is there are 2 parts real and imaginary. So, this is scaled by half. So, for a random process.

(Refer Slide Time: 42:51)



If there are now two random processes X t and Y t. Then we say that those two random processes are this is these are random processes and we say that they are statistically independent. These are independent if we take the random variable X at say x t 1 t 2 till t n and Y at t 1 prime t 2 prime till tm prime. And, we take the corresponding random variables joint p d f. Then that will factor as the product of these 2 separately x t 1 to x t n times p Y y t 1 prime to Y t n prime. And, it will be it is called uncorrelated if expectation of X t 1 Y t 2 is expectation of X t 1 times expectation of Y t 2. These are all usual definitions for this is all t 1 and t 2. Now, if we take the Fourier transform of the autocorrelation function for a stationary wide sense stationary process and that is called the power density spectrum. So, if we have for a stationary process we have the autocorrelation function is a function of a single the gap between the times.

(Refer Slide Time: 44:58)

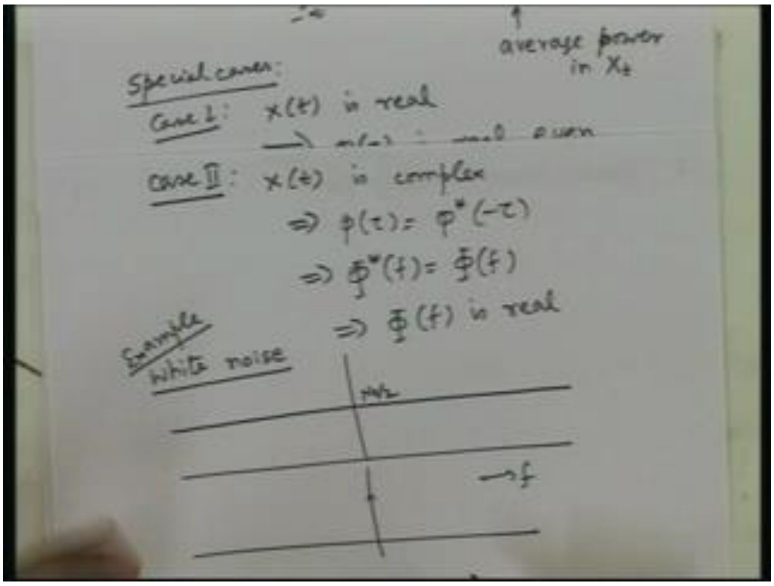


We can take the Fourier transform and let us call it phi of f is defined to be minus infinity to infinity phi tau e to the power minus j 2 pi f tau d tau. This is called the power density spectrum of X t. It is a very important function; this is the Fourier transform of the autocorrelation function. And, we can compute the inverse Fourier transform and get back autocorrelation function from the power density spectrum as, phi tau equal to minus infinity to infinity phi f e to the power j 2 pi f tau d f inverse Fourier transform. Now, from here we can see that phi 0 if you compute the autocorrelation function at 0 at 0 that is nothing, but integral minus infinity to infinity phi f d f.

This is the area under called power the that that function area under the power spectral density function. But what is this is also by definition this is nothing, but expectation of mod X t square if it is real then it mod is not necessary if it is complex then X t X t conjugate. So, it will give you mod X t square. And, this is nothing but the energy of the signal at t average energy average power in X t. So, this is always of course, greater than equal to zero and so, this is average power in X t and what is this is the area under the curve phi f. So, area under this power spectral density function is the average power in X t.

So, this actually gives you this function gives you how much power is there in what frequency. How much average power of X t is there in what frequency that is why this is the power spectral density? So, this is the power spectral density power density spectrum Now, we can take some we can see some special cases. Common special cases are when X t is real and; that means, that will mean that phi tau 1 can check that phi tau is real and even function. So, that phi f the power density spectrum will also be real and even.

(Refer Slide Time: 48:30)

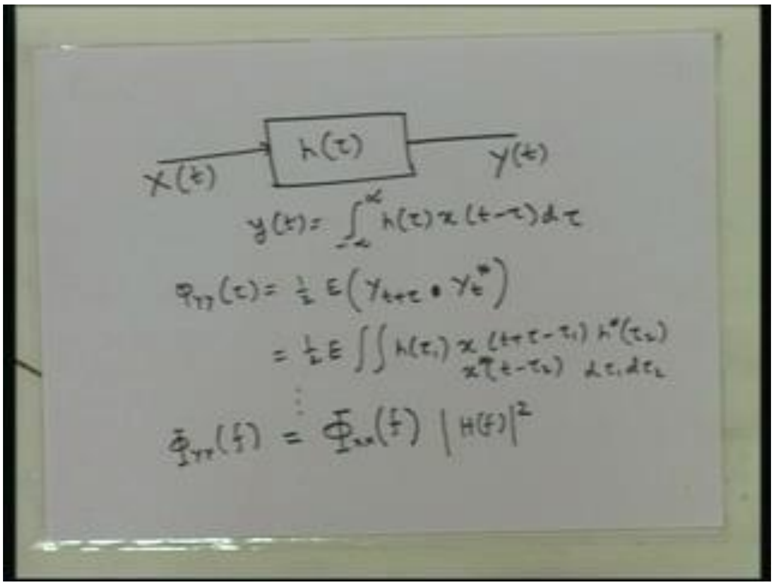


And then the next special case is when X t is complex then what will happen is phi tau will be phi conjugate minus tau. It is because the because of the way of phi tau is defined phi tau is defined to the expectation of X t X t plus tau conjugate. So, if we take X t plus tau X t conjugate and that is the whole product is a conjugate of the previous thing. So, as a result at minus tau phi minus tau will be phi tau conjugate or phi tau is phi minus tau conjugate. So, this will imply phi star f is same as phi f; that means, the conjugate is same as itself this. So, what does it mean phi f is real.

So, we have discussed about power spectral and power density spectrum and one very important type of noise that we come across is what is called white noise. Where the power spectral density is a flat function. So, example white noise where there is a flat power spectral density. In the two side, if you draw it in two side then the density is usually denoted by n naught by 2, where n naught is the one sided a density function. Where if you add the density at minus frequency and plus frequency you get n naught. So, this is n naught by 2, two sided density function the amplitude is n naught by 2.

So, what is the autocorrelation function of the white noise then? It is a delta function; it is an inverse Fourier transform of this. So, it is delta function; that means, that the autocorrelation at non zero gaps is zero; that means, if you take two time instances and consider the random variables of a white noise at those two time instances then the correlation between them is zero. So, there is no correlation between the values of the noise at 2 different time instances. So, uncorrelated noise, so white noise is basically a uncorrelated in time. Now, it is also in this context is also very important to see what happens when you pass a stationary process through a linear time invariant filter.

(Refer Slide Time: 51:44)



So, if you have a filter with impulse response h tau and if you pass an X t a random process through it. Say, we have a microphone and a voltage is generated and we pass it through the through a system through a filter with impulse response h tau what will be spectrum power spectral density of the output process? So, Y t is minus infinity to infinity h tau x t minus tau is the convolution of input and output input on the filtered impulse response. And, to compute the power spectral density of the output process we need to first find out what is the autocorrelation function of this process.

So, this is also stationary because this is stationary. So, phi y tau is half we considered in general a complex process. for real also it will be it will be the it will be similar derivation. Expectation of a Y t plus tau Y t conjugate it is the product. So, once we write this expectation in terms of their this can be written as this integration each one can be written in terms of this integration and then we get a expectation of double integration h tau 1 x t plus tau minus tau 1. We can call it alpha and beta. We can change the variables later and then for this we can take h conjugate tau 2 x t minus tau 2 conjugate then d tau 1 d tau 2.

So, we can take this and we can find out what is the Fourier transform we have to take another integral and find out the Fourier transform. And, with some manipulation one can show that this will be with some manipulation. One can show that this will be the power spectral sorry not this. But the Fourier transform of this that is phi Y Y f this will be the power spectral density of the input process times H f mode square. So, H f is the frequency response of this system. That is this is the impulse H f is impulse the is the Fourier transform of h tau. Then this is the basically power response and this autocorrelation of this power spectral density will be multiplied by this frequency response to get the output power density spectrum.

It this quite satisfactory, because we note that if you transmit if we pass a deterministic signal also through such a filter, the Fourier transform of this is multiplied by the Fourier transform of this to get the Fourier transform of the output. The output Fourier transform becomes the product for the Fourier transform of the impulse response and the input. So, here for our random process also a similar thing happened only this is now we have to consider the power spectral density. Power spectral density of the output is the power spectral density of the input times the frequency power frequency response of the filter.

So, in this class we have first considered somewhat happens to the how we can get the density function probability, density function of a function of a random variable from the density function of the original random variable. We have also considered functions of random vector and seen how to express the probability density function of the new vector in terms of the original vector's density function. Then, we have defined random process and we have defined correlation function of a random process. We have defined, what is the stationary random process?

Then, wide sense stationary random process and we have defined autocorrelation function and the Fourier transform of the autocorrelation function is called the power spectral density or power density spectrum. And, I have seen a particular example of a power density spectrum there is a simple example, but it is a very important example because it come we will come across this in this course very often it is a white noise. The white noise is a particular process for which the autocorrelation function is a delta function that is the power density spectrum is flat. And, we have then discussed what happens to a to the power density spectrum of a random process if we pass it through a linear time invariant filter.

The output of the filter will also be the random process, but that will have the power density spectrum which is the product of the input process power density spectrum and the frequency response of the filter. So for example, if we pass white noise through a low pass filter white noise has this flat power spectral density power density spectrum. So, when we pass it through a low pass filter the output will have a low pass power density spectrum. There is that kind of results will follow from what we discussed just now.

Thank you.