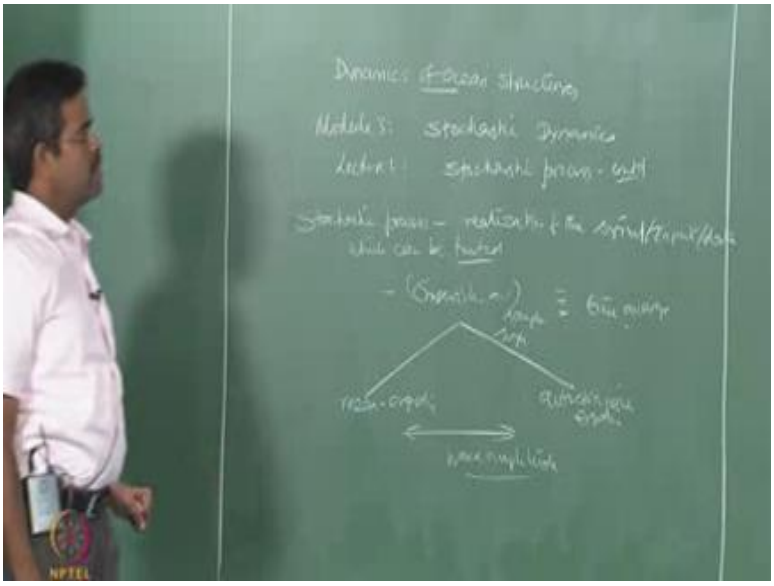
Dynamics of Ocean Structures
Prof. Srinivasan Chandrasekaran
Department of Ocean Engineering
Indian Institute of Technology, Madras

Lecture – 44
Stochastic Process – contd

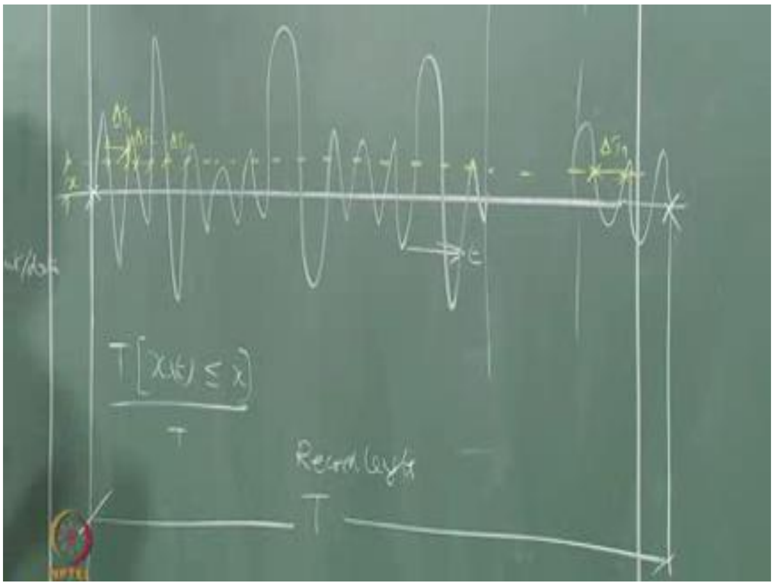
In the last lecture we discussed about the stochastic process introduction. We said that the basic difference between stochastic and random is that more or less they are same except of the stochastic process gives you the realization of the signal or the input or the data which can be trusted.

(Refer Slide Time: 00:25)



Why I am using the word trusted because the ensemble average of the sample of a single signal is as same as that of the time average. So, one can define this by two ways one make it mean ergodic, make it auto covariance ergodic. A classical example of the system which is a stochastic process can be the wave amplitude. Now, one can really understand a very interesting physical meaning of (Refer Time: 01:41) process to be ergodic for the stochastic process.

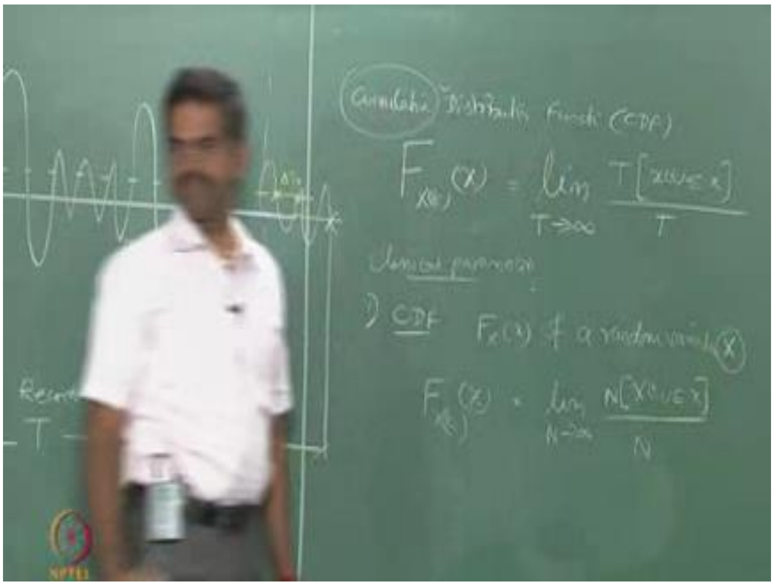
(Refer Slide Time: 01:45)



Let us take this specific signal. Let us have signal like this which is along the time scale where the signal is we will be after specific window let us say the signal continues. Now I have a mean value or any specific value of exceedance let the value be drawn here, let us call this value as some x. So, what I am interested is to know what is that sample which is less than and equal to x over a time and then average it for the record of t. The T is actually the record length; let us say I start my record from this point to let us say this point. So, this becomes my T what we call as record length.

Now, looking for the time at which the sample data is lower or equal to some specific threshold data or some control value or some representative value within the signal. So, if I read this signal then I can easily point out these are the values which are lower than the x. So, I can call for example this as delta T 1 similarly I can call this value as delta T 2 similarly delta T 3 delta T 3 are nothing but the time values. So, as I proceed I can call this value as delta T n where there is some n value. So, explicitly this can be written mathematically like this which we call cumulative distribution function which is CDF expressed as capital F x of T of x, which is now going to be mathematically the limit T times infinity T of x of T less than x of T.

(Refer Slide Time: 03:57)

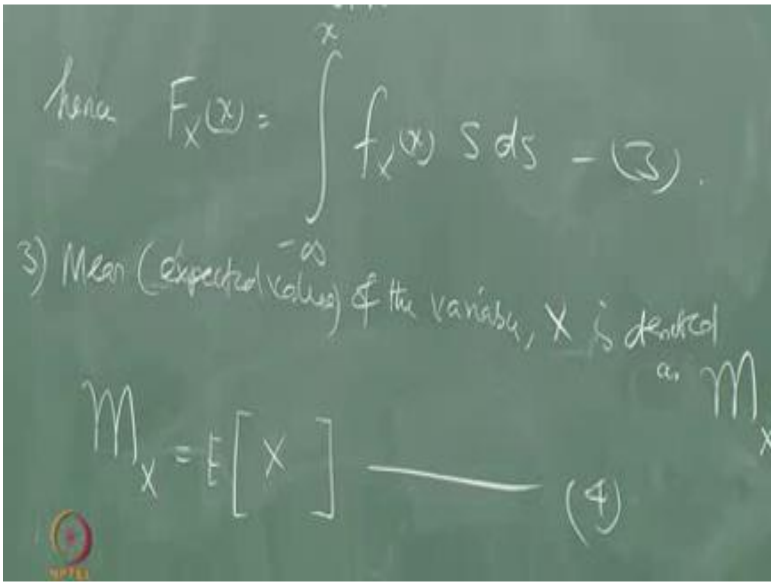


So, the explanation what we have given here is mathematically nothing but the cumulative distribution function.

Why cumulative? Because I am looking at the record from 0 to T, so it is cumulative why is distribution because I am looking at the distribution of the entire signal for the whole record length capital T. Now, I can interestingly see this is going to be actually time average these values will T is applicable only when the processes remains stochastic which is now mean ergodic as well as auto covariance ergodic. So, therefore, to make or to apply a stochastic process or stochastic dynamics in a given process the process should qualify to remain as ergodic, otherwise you cannot do this. So, let us quickly see some of the classical let us say parameters which are required in stochastic analysis - one must what we call the cumulative distribution function, which we already said there, but let us formally write it down for any random variable.

Let us say the cumulative distribution function F of x of x of a random variable capital X. Suppose I wanted to find out the cumulative distribution function of a specific value of x T 1 of x for over a record length limit n tends to infinity. So, average of N, N of x T one less than X. So, in this case I am using capital X because is my variable.

(Refer Slide Time: 07:03)

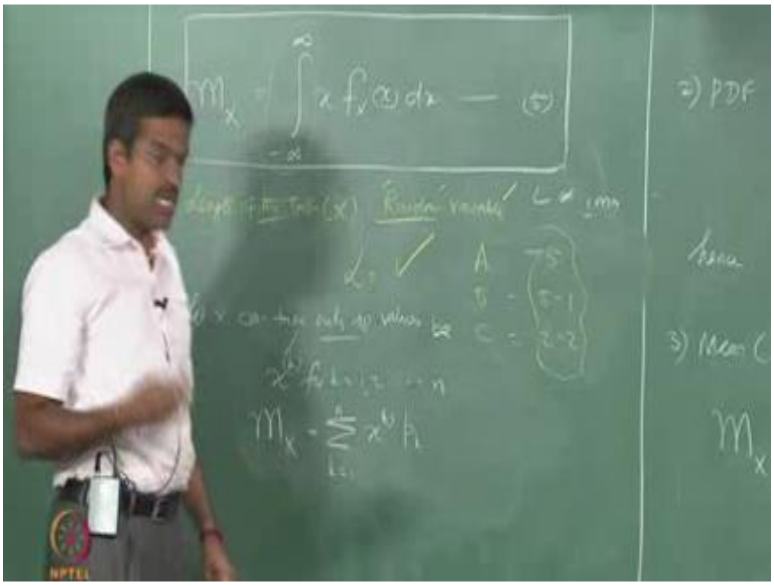


Then the next important parameter which people look at is which is indicated as small f of capital X of x where x is the random variable which is a differential of CDF. So, I will call equation number 1, equation number 2. So, by this definition hence the cumulative distribution function which is F of X of x can be an integral of the probabilistic function which is F of X of x sds, whereas the integration is only from minus infinity till x. The reason being the cumulative distribution function is checking the period city or the validity only with respect to x that is the reason, because here this is nothing but expanded as any variable or any variable less than or equal to control value.

Therefore the integration of the function only will be from the negative value of infinity to that of the positive value alone that is the reason why we are saying this way. So, this can be a reverse function like this which integration of the PDF which can be obtained, once we know these two one can find the mean value. Mean is otherwise called as expected value of the variable x is given by is denoted as small m of capital X as x is a variable. I am deliberately using capital X here taking x as a variable just to make it clear that where we are substituting capitalize, so m of X.

M X can be given by a simple expression which is expected value of the variable - this is equation 5.

(Refer Slide Time: 09:49)



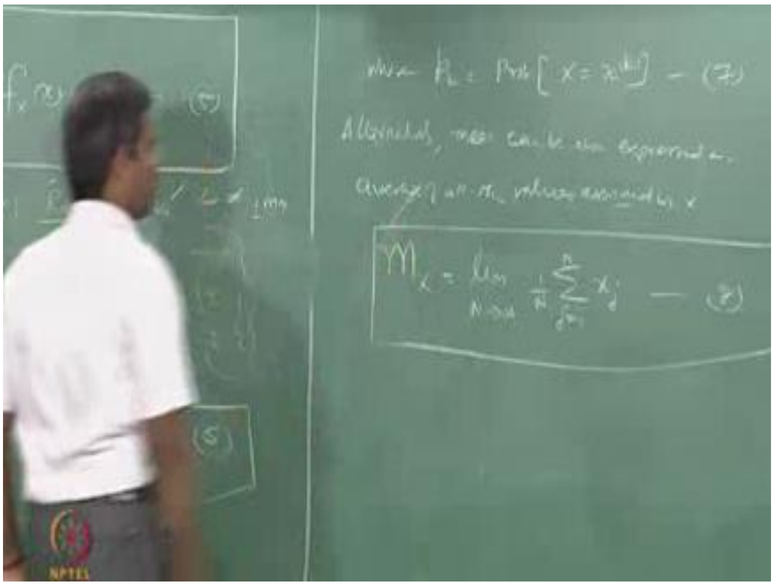
Now remember interestingly, the mean actually considers the value from all the bounds that is represented actually here, whereas the p d f is only consider the bound from the value of infinity to d x because x actually threshold value for fatigue CDF. So, m X can be given by this. Let us retain this equation because I need this equation later for comparing for another case, so will have this equation ready with us.

Now let us take another parallel example and derive m X in the different understanding let us say for example, you are looking for the length of the table. Let us take length of the table is a variable, why it is a variable? Because random variable I already said yesterday random variable is that number or the term random is the variable can take any value in a given set, but the variable already has a fixed assigned number we do not know the value therefore we are selecting it randomly, why randomly? A person will say the length of table is 5 meters other person will say the length of table is 5.1 meters, it is say 2.2 meter all these are random numbers. But the length of the table is neither of them or all of them, but how over the length of the table L is a fixed value, so does not change. Though the meaning is a variable the variability comes only on choice, but the value is fixed actually.

So, let us say I want to check the length of the table which is a variable X, which is a random variable and; obviously, length of the table will be normally in the whole number. We do not have length of the table is 5.15 unnecessarily you are converting feet into a meter or vice versa generally it is a whole number. So, we do not want fractional numbers in this.

On the other hand, x can take a specific value. Let x can take only specific values for example, you may say how specific value length of the table cannot be 1 millimetre I mean. So, it cannot take a random value of highly random in nature see this got a specific value only, some decent assumption. So, this got a specific value. Let that specific value be x of k for k equals 1 2 up to n. Here one is interested to know the mean of this value, one is interested to know mean of this because it now taking a random variables I wants to know mean of this because I am talking about the variable as capital X, m X here which can be simply sum of all possible values varying from k equal to 1 to n that is a variable here of that of x of k that is the value what it can take multiplied by sum p of k. So, what is p of k?

(Refer Slide Time: 13:18)

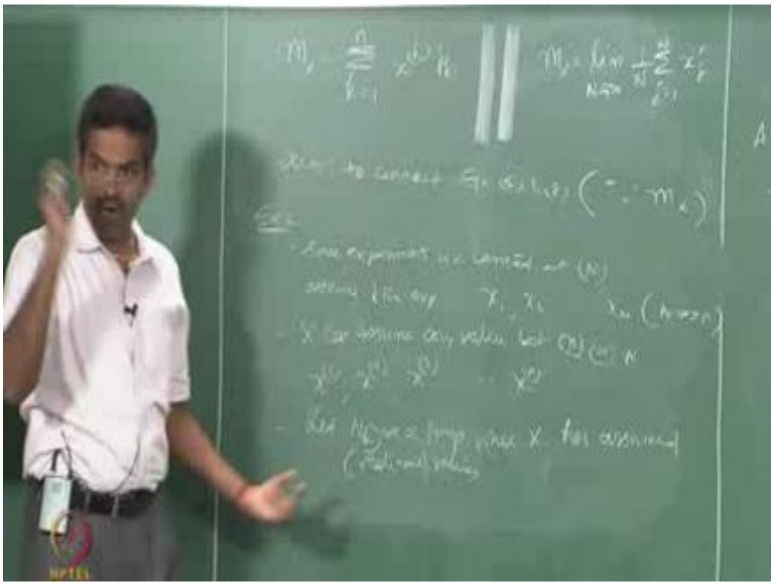


Where I call this equation as equation 6, where p of k here is the probability x will be exactly equal to x of T. Here is x of T is a whole number, k is the count, k is the count, and it is k actually I do not know whether you are able to see it, this is k - k is the count there is no T here.

Let see equation 7, alternatively, mean can be also expressed as I can also express mean the other way. The average of all the values assumed by k, so m X can be also limit n tends to infinity 1 by m mode, 1 to n x of j - equation 8. Then interestingly, if you look at the equation of 6 and 8 or equation 6 and 8 - 6 and 8 both of are different form actually because in 6 we do not have an average over the sample data and we have a probability of the chosen value of x associated with the whole number what it can assume, whereas in 8 we do not have the condition. So, simple average, but both are giving mean. So, I must be able to connect them statistically both are actually mean, this is also mean equation for finding out the mean of the variable this is also.

So, I have to connect these two statistically, let see how to bridge this. I will retain this equation as (Refer Time: 15:38). So, now, I have two equations.

(Refer Slide Time: 15:41)

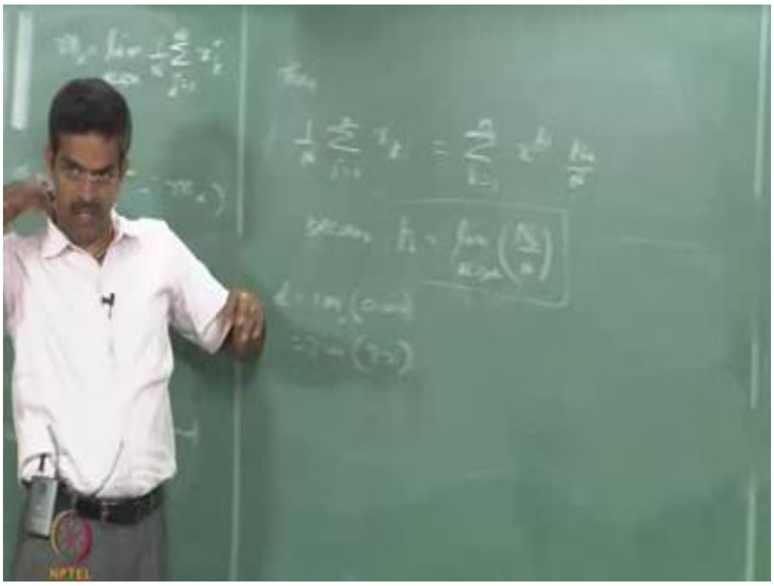


One is m X is summation of k equals 1 to n x k of p k, the other is m X equal to limit n tends to infinity 1 by n of sum of j equals 1 to n x of j. So, I have to bridge this, they are both are mean. Now to bridge this let us discuss one more example. Now the aim is to connect equation 6 and 8 why both are actually m of X, I want to connect this. Now let us take an example, the example is we are conducting some experiments some experiments are carried out let say n set of values - we got some n set of values. The outcomes of the experiment are x 1, x 2, x N, where N is sufficiently larger than n, sufficiently larger than n, N is sufficiently larger than n.

Now, it means that x can assume any value between either n or n because both are mean that is it can take x 1, x 2, x 3, up to x n, but amongst these values of what x is assuming all values are not relevant, why? Because we already said x cannot take certain values beyond the bound of this for example, length of the table cannot be one millimetre the one millimetre are one is also a random number, but it cannot take. It means, it is not that bound of n it can take though N can be very higher than n, but my mean should be bounded only on certain chosen values which are relevant to the data this is called physical realization is stochastic process, this is where we are different from random process. We have to physically realize this value to that of the example or the problem. So, it means X can take any value because both are giving me the mean, but there are amongst any value the only relevant values of x 1, x 2, x 3, till x small n.

Now, x is assuming any value, but only these are relevant it means there are groups. Let N k be a group where X has assumed the possible realized values, why N k? Because k is the count for n, k is the count for n. So, what I am doing is amongst the value what capital x can take and grouping them in such a manner that some are relevant values, some are irrelevant, some are imaginary and forming groups, let one group of N k represent those k values based on which the mean was calculated.

(Refer Slide Time: 19:40)



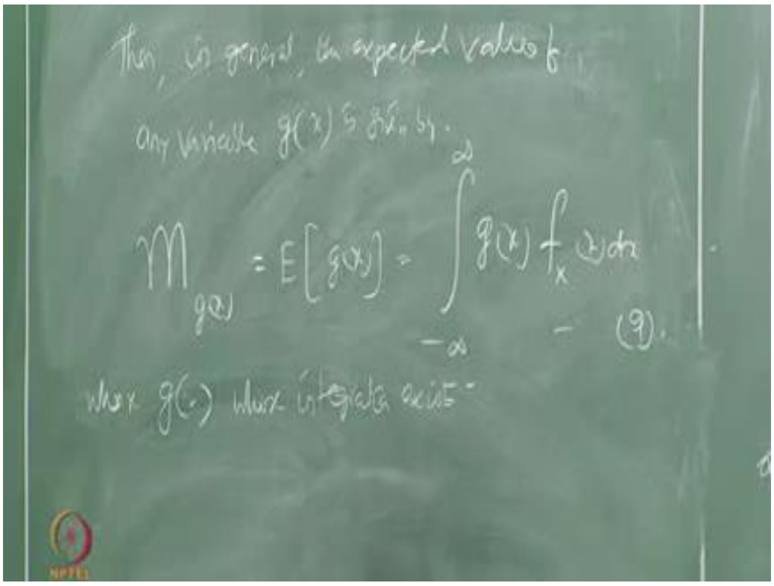
1 by N of summation of j equals 1 to capital N of x j, should be made equal to summation of k are one to small n x of k instead of p k we said this N k by n where why? Because the probability of k which is suiting the value what is the probability you understand this; for example, some in the classroom some person one person says the value of length of the table is 1 millimetre, 1 meter; I associate a factor or a multiply to

that value say that this is irrelevant to the probability of accepting this value as my table length is may be 0.001 percent.

So, it will automatically deplete a contribution of this number on the mean. So, the p k is nothing, but the probability value of the realization on the original sample, you understand. If the real table length is 3 meters somebody says 3.1 meter than the probability will be (Refer Time: 21:01) 95 percent or 90 percent. Then multiplier will take care of the relevance or realization of this value in the original sample. Now this realization or the multiplier realization in terms of probability is now converted to this because p k is nothing but limit n tends to infinity N k by N, is it OK. For example, you have only 6 numbers. So, N k will be 6 and N is also 6. So, all your values will have a probability of 100 percent. So, the mean will be exactly equal to the average of those 6 values which will be same as this.

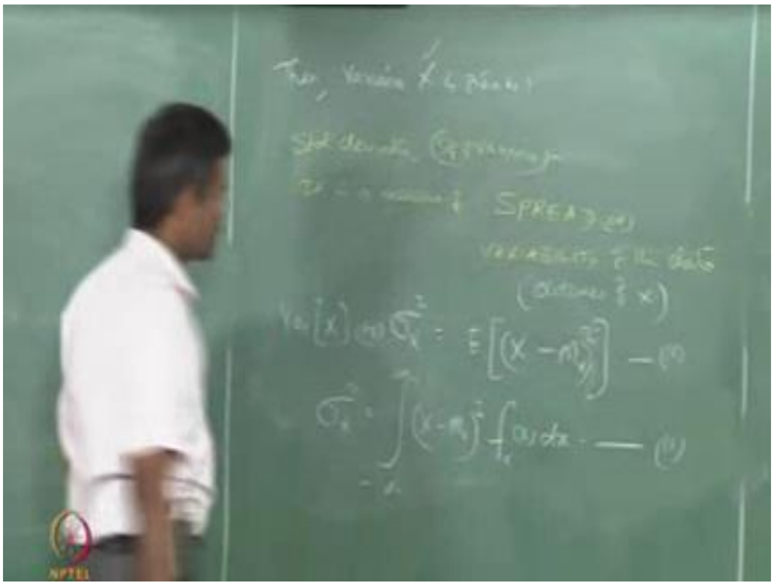
So, if my p k is giving this validity, these two means will converge with the same value - one is for the random process one is for the stochastic process.

(Refer Slide Time: 22:23)



Having said this, then in general the expected value of any variable g of x is then given by I want to generalize this - which is m, because initially when the random variable of x I wrote this as m x. Now the random variable g of x, so I should say this as m g of x, when the random variable was x I wrote this as exported value of x. Now I should write exported value of g of x, when my random variable was x I should integrate this to form my PDF as a multiplier of x PDF of x. In this case g of x f of x d x, where this is equation number 9; where g of the function, may be any function E x g y etcetera whose integration exists that goes without saying otherwise you cannot integrate this. Parallely then let us work out the variance.

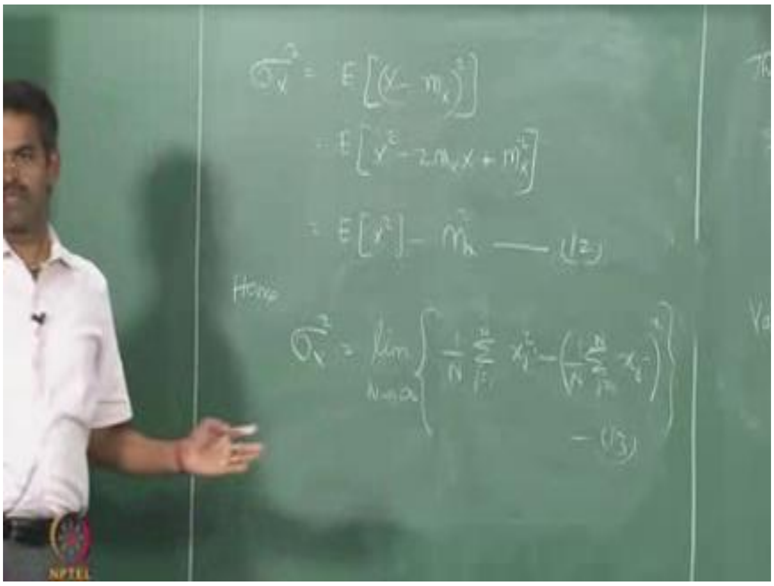
(Refer Slide Time: 24:20)



Variance X is given by; now, one can people ask the question that what is importance of variance. Let us say standard deviation which is the square root of variance is important because it is a measure of spread or variability of the data or to be very specific outcomes of the variable. Will tell you the spread of the outcome of the variable, so it is very important therefore, one should know how to calculate the variance for the given variable X. So, variance of X is generally is indicated can be also called as sigma X square because it is sigma X is a standard deviation which is square root of the variance which is nothing but the expected value of the variable minus the mean square; let us write it properly this is square expected value of this function.

Now, I want to such to substitute this back here, because mean already we know is given by this function. So, therefore, sigma x square or variance can be given by an integral of minus infinity to plus infinity, g of x is X minus m of X and integral of f sorry PDF of f of capital X small x d x - equation number 10, 11. Having said this, this one is square; square.

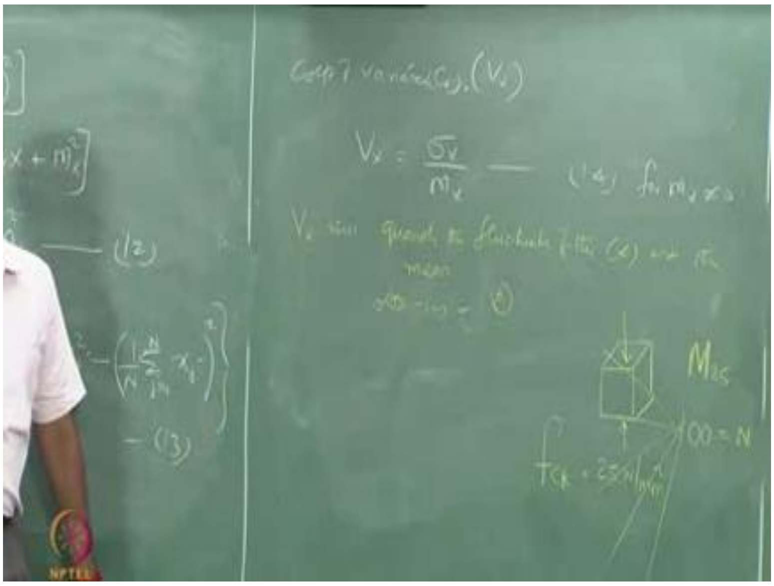
(Refer Slide Time: 27:10)



So, one can expand this expected value of X minus m X square which can be said as expanded value of X square minus 2 m X X plus m X square, which can be mathematically said as the expected value of X square minus m square X - equation number 12.

Hence, sigma X square can be now written as limit N tends to infinity; this is one, which is going to be 1 by N of summation of x j, j equals 1 to n of square that is what it means; let me use a different bracket here minus. I am now talking about this which is 1 by n of summation of j equals 1 to n of x j the whole square - equation number 13. Now, one can ask the question what is the interest of finding out the mean and standard deviation which are first (Refer Time: 29:03) of a given random variable.

(Refer Slide Time: 29:08)



If you know both the means and standard deviation I can find out the coefficient of variance C v or V x let us this is symbol is V x. Vx is given by the ratio of sigma x (Refer Time: 29:27) equation 14. Of course, for mean for non zero mean process for m x minus not equal zero because there are some process which is zero mean process.

Now, one can ask a physical question what is the importance of finding about the C v or V x. Indication of V x will qualify or let us say quantify the fluctuations of the variable with respect to the mean. Let us physically see this question, I have a cube let us say the cube is designed to be M 25, it means the characteristic comprises strength of the cube should be 25 Newton per millimetre square after a cube (Refer Time: 30:38) 28 days. So, how do you do it? I tested in a testing machine feel a cube or crush the cube and find out the crushing strength of the cube. I have crossed 100 samples, found out all the data.

Now, one can easily find the mean and standard deviation of the given data from this equation mean is nothing but sum of all divided by N, and standard deviation is nothing, but the equation what we already have we can find. If you know the value of V x you will exactly know amongst 100 sample which has come from, let us say 10 different sources you can easily find out if the variance is very high then it is away from the mean then the sample is either too strong or too weak. So, it will give me the fluctuation, not give me good or bad, it gives the fluctuation with respect to the mean of the given problem.

So, it is a good index for people to know what is that sample which is either too bad or too good, from the given lot of n which is 100. So, this will help you to group the sample in terms of a category. You can say sample which is having closer to 98 to 100 can be group A. So, this will help you to do that. So, that is a requirement therefore, in a given stochastic process (Refer Time: 32:05) if the variable is following in a ergodic process which has got mean and auto covariance ergodicity defined in the given sample single sample from the given ensemble is qualified to represent the whole ensemble as an input data for example, wave load. For that sample one can find the mean and standard deviation, and take it forward for finding out the; defining the load in terms of statistical properties which will be used in dynamic analysis directly.