Risk and Reliability of Offshore Structures
Prof. Srinivasan Chandrasekaran
Department of Ocean Engineering
Indian Institute of Technology, Madras

Module - 01
Lecture - 10
Random Variables

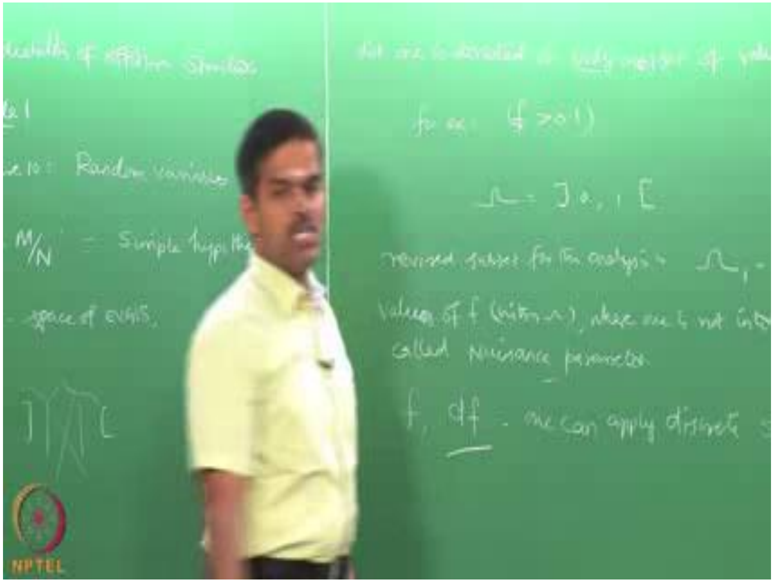
(Refer Slide Time: 00:24)

Welcome to the online course on risk and reliability of offshore structures. We are talking about lectures on module-1. In module 1, we are essentially focusing probability and possible reasoning. In this lecture, which is the 10 lecture, we will introduce you some characteristics of random variables. What are random variables, how they can be simulated, how they can model or how they can be in ensemble for a given analysis?

Before you move into the discussions of random variables, let us recollect what we discuss in the last lecture. We said that hypothesis is very important for sampling statistics in probability in possible reasoning. Let us talk about simple and compound hypothesis. The hypothesis what we discussed so far in the last lecture referring to generally a single parameter where single parameter f can be simply M by N such hypothesis are otherwise called as simple hypothesis. Because in the space let us say of events sigma of all parameters that are possible for an analysis f is represented only be one single point that is why its call simple hypothesis.

However, in many cases, one will be interested to test the hypothesis for the parameters lying within some sub set of the space value. We are not interested for the entire space of events, we are interested only in a range of space of events while particular value of failure f one may not be interested for or one may be very much interested. So, on the other hand, physically if a space of event is very large, you are not interested in finding out the analysis or for the entire set of range of space of events, you are interested only in certain range of this or you may not be interested in certain range of this. So, in that case how would we actually handle these problems?

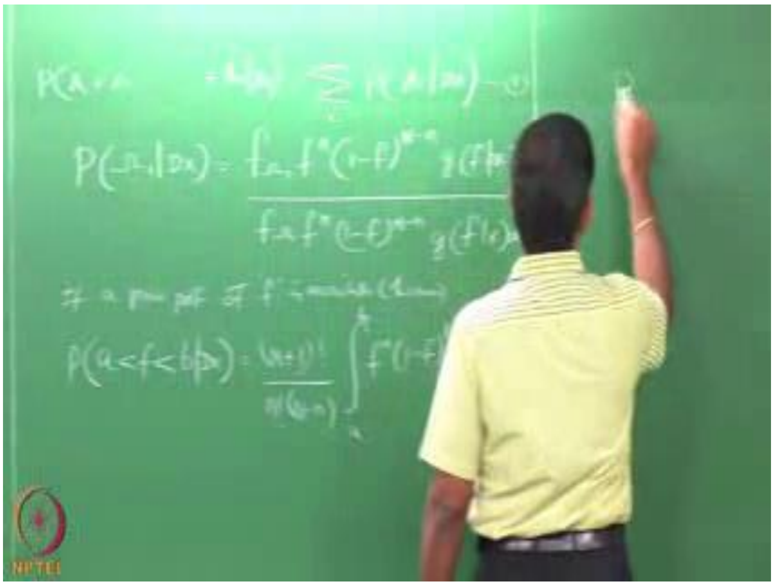
(Refer Slide Time: 03:04)



Now, let us say one is interested in only one set of values of failure. Let us say let one is interested in only one set of values of probability of failure; let us call this as f. Let us say for example, f greater than 0.1, one will be interested; but of course, in the space of events 0 and 1. So, in the space of event 0 and 1, one is interested in knowing or analyze in situation only for f exceeding 0.1. Therefore, now the revise subset for the analysis is let say I call this as sigma 1, where I say now my subset is 0.1 to 1, because I am interested in values exceeding 0.1.

Finally, see that one is not interested in certain actual values of f; on certain actual values of f, one may not be interested. For example, if the value is less than 0.1, it does not actually simulate any failure such values are called nuisance parameters. The values of f of course, within the subset where one is not interested are called nuisance parameters. If the chosen problem has no other parameter other than the variable f and their difference interval is d f which are mutually exclusive then a discrete sum rule can be directly applied what we discuss so far. So, you so got only f values and the interval is d f which you are interested in then one can apply the discrete some rule why I am saying discrete sum rule because I am interested only on a specific interval of the subset sigma 1.

(Refer Slide Time: 05:49)

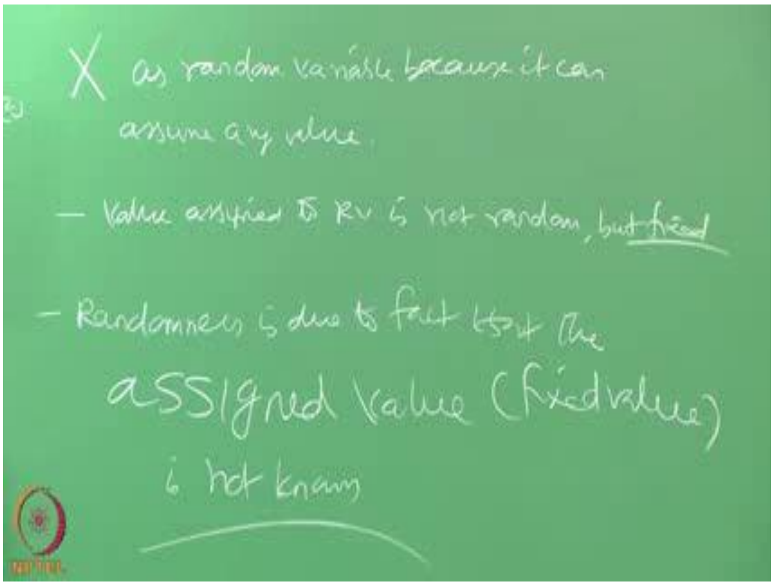


Then I can write probability of A 1 plus A 2 and so on plus A n given d x, we already know d is results of the hypothesis or the test conducted experiments investigated. Given d x can be simply said as summation of i of probability of A i given d x. Let me call this equation number 1. Now the above equation which is one can be also generalized into an integral form as variable A i becomes more numerous; in that case one can easily get rid of the nuisance parameter by integrating it. Let us see how; so probability of given d x where I am looking only at the subset is given as f sigma 1, f n 1 minus f n minus n where capital n is a total set of values what we observed and n is the value just failed in that g of f f x dx divided by f of the original subset sigma f n 1 minus f n minus n g of f given x. Let me call this as equation number 2.

In case a uniform prior PDF is available to you for a single parameter f, if prior PDF that is probability it is the function of f is available or I could say it is known to me then the integral given in equation two reduces to an incomplete beta function let us say that given by probability of a less than f less than b given Dx the number a and b represents specific events as be discuss in the last lecture.

Which can be now said as n plus 1 factorial by n factorial n minus n integrating from the limits of interest a to b f n 1 minus f of n minus n are equation number 3. A set of discrete hypothesis can always be identified by one or more numerical indicates or indices which can identify the in such cases hypothesis testing transforms into a problem of estimation that is the importance of hypothesis testing where we can handle it a simple and compound hypothesis as we just now saw in few minutes.

(Refer Slide Time: 09:46)



Let us now talk about a different kind of set of variables available which are actually useful in doing reliability analysis of offshore structures what we call as random variables. Now, let us quickly see physically to understand what do you mean by random variable, why variable we are not can take any value in a given analysis, but why the randomness is it assigned to the variable because a variable can take any value or is it assigned because we do not know the value what the variable takes. Random variable is not so called random variable because you can assume any value, it is not called as random variable, because it can assign or it can assume any value, it is not because of this reason we call this random variable. The value assigned to the random variable is not random; the value assigned to the random variable is not random, but it is fixed.

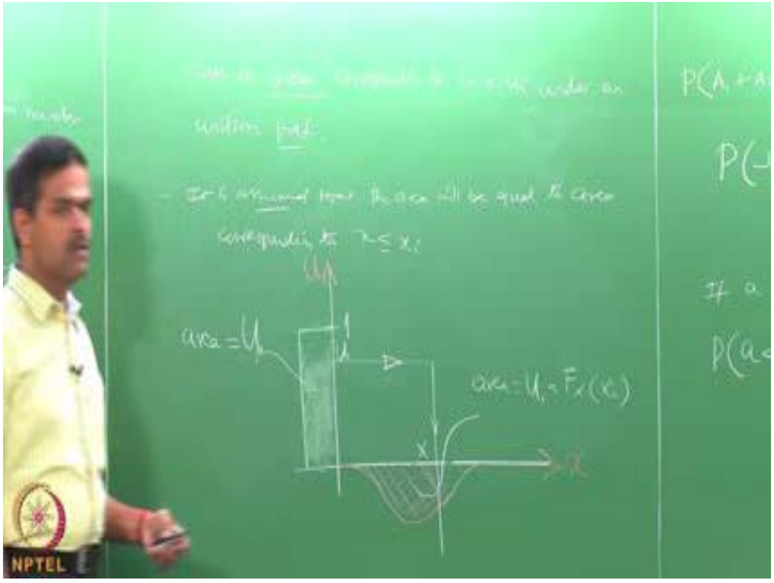
Then why we call this random, randomness is essentially due to the fact that the assigned value to this variable is not known. So, randomness is due to the fact that the assigned value which is actually a fixed value is not known. So, since the value is not known you will be guessing this value in a wide range that is why we say it is randomness being guessed by range of possible values, these variables are termed as random variables is that clear.

(Refer Slide Time: 12:23)

Now, random number is therefore, a representation of the event that has randomness associated with itself. It is only a convention to deal with the random event mathematically. Let us quickly see how do we generate these random numbers; there are many ways of doing it. Let say I have specified distribution; I want to generate random number to fit into a specified distribution.

Let us say sequence of random numbers x i let us say let x i be a sequence of random numbers so x i is actually x i, where i can be 0, 1, 2 and so on from a non uniform probability defined by the cumulative density function which can be generated as below. So, I have an existing a predefined CDF. I want to generate a random number to fit to this CDF. So, what we do first generate a sequence of uniform random numbers. Let us say from 0 to 1, where U i and i can take any value from 0 and so on. Then determine x i as f minus x of U i and i varies from 0, 1, 2 etcetera, this procedure operates by setting an area corresponding to U less than U i under uniform PDF.

(Refer Slide Time: 14:23)



So, this of course, sets an area corresponding to U less than U i which is under an uniform PDF that is very important, I am trying to fit a random number for existing distribution. Now in this case there is a basic assumption it is assumed that the area will be equal to this area will be equal to the area corresponding to x less than x i or equal to x i under the PDF of interest. I can draw this graphically as you see here. Let us say I have a curve say this is my x, this is my u, and I have a uniform curve here.

So, this value of course corresponds to 1; let say this area corresponds to U i. And this area let say not the complete area, but the corresponding area how do we get this let say I have a corresponding area which is only up to here, so I have a corresponding area. I mark a corresponding area here, so this is U 1 and of course this is my x. And this area now this area will be equal to U 1, which is an assumption that is going to be F of x x i.

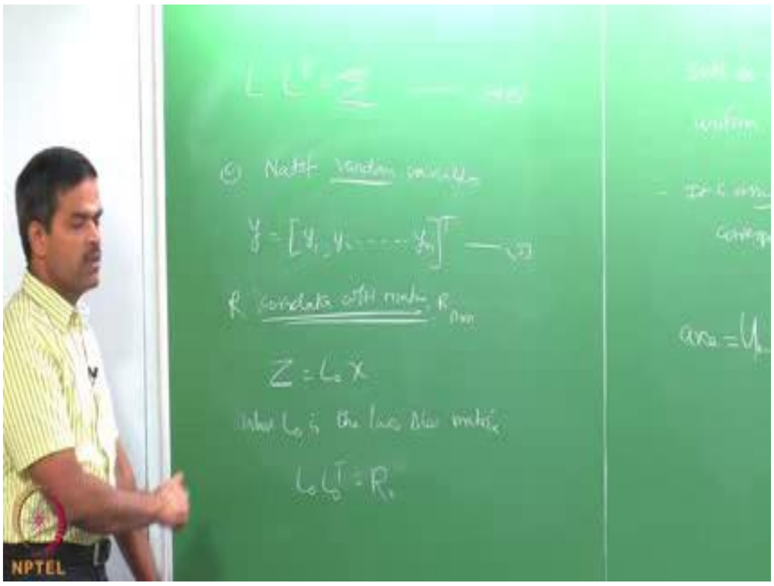
(Refer Slide Time: 17:42)

So for x to follow a standard normal distribution, one should generate pairs of random variables u 1 and u 2 uniformly. For x to follow a standard normal distribution, one should generate a pairs of random variables, let say u 1 and u 2 uniformly, and this should be distributed in the range 0 to 1. So, how to generate this we can use the following equations let us say x 1 and x 2 square root of 2 natural logarithm of u 1 sin 2 pi u 2 square root of minus 2 logarithm u 1 cos 2 pi u 2 - call the equation number 2. If we look at this equation please see that if the uniform random variables are statistically independent, then the standard normal random variables generated using the above equation two will also be statistically independent.

(Refer Slide Time: 19:51)

Let us now talk about multiple random variables. A set of random variable y which will be - y 1, y 2, y 3, till y n which is actually a joint random variable which means that M - the vector M and the covariance sigma. One need to generate the sequence of random numbers or random variables that are statistically independent one can use the following relationship for this. Let us say x can be x 1, x 2, x n transpose. And now if you want to generate y, simply say it is L x plus M where M is the mean vector and sigma is a covariance matrix - call equation number 4. In this case, L is the lower triangular matrix this can be found out from Cholesky decomposition which can be determined using Cholesky decomposition of the covariance matrix.

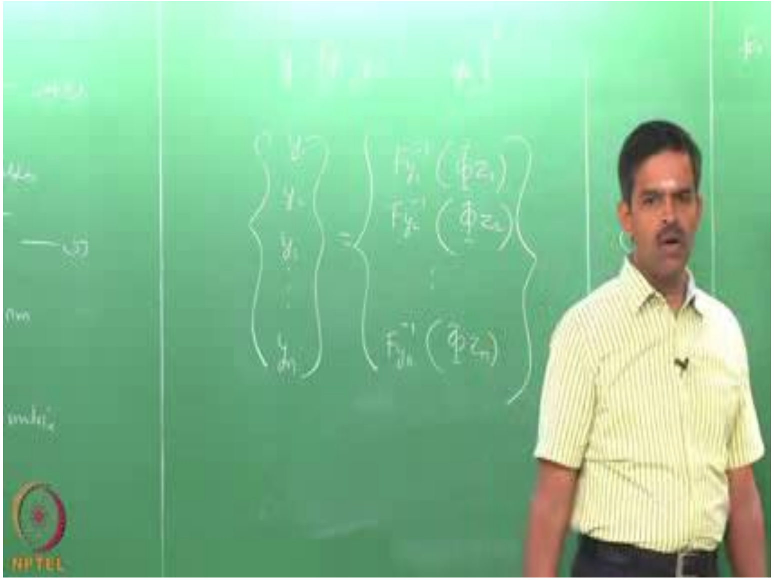
(Refer Slide Time: 22:14)



That is one has to fulfill the requirement that L L transpose should be the covariance matrix it calls the equation number 4(a) just in continuation with the set of equations written there. Now, you have to repeat this procedure n number of times for how many variables you want to generate as multiple random variables. There is another way by which you can generate random variable that is what we call Natof random variables. Now, to generate Natof type random variable of y equals y 1, y 2, y n transpose which is now a vector of Natof distributed random variable with marginal cumulative distribution functions with marginal CDFs. And the correlation coefficient matrix R is of size the correlation coefficient matrix R will have size n by n, because n is the number of variables of generating.

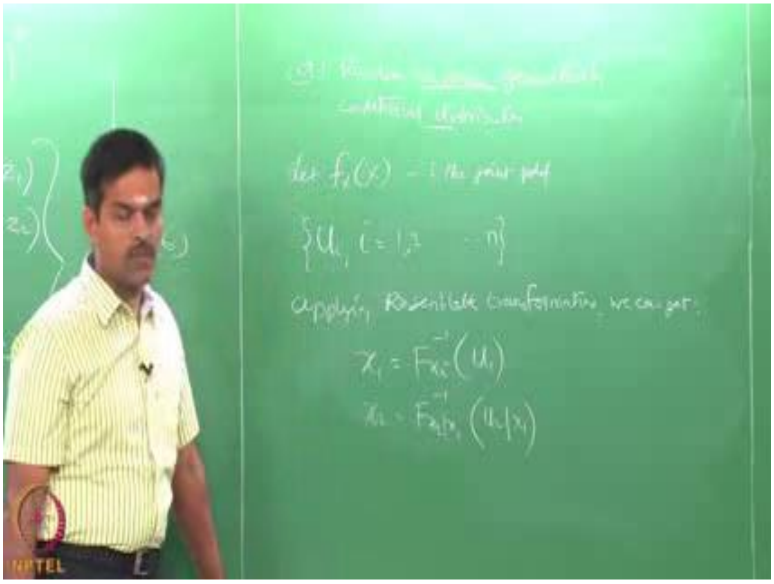
One can generate now a sequence of n random variables which are statistically independent one can use the following relation. Let us say Z is L naught x, where L naught is a lower triangular matrix, that satisfies L 0, L 0 transpose is my R 0, where R is the correlation coefficient matrix. Now R 0 of course, is the modified correlation coefficient matrix for the Natof random variables.

(Refer Slide Time: 24:38)



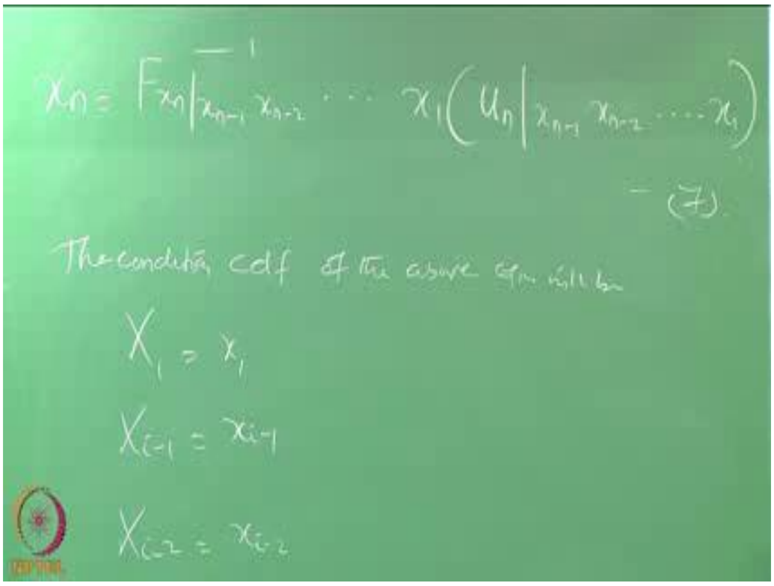
Now, the outcome of this generation will be determined as follows. We know y is y 1, y 2, y n transpose. So, I want to establish and expand this series of y 1, y 2, y 3 y n is now equal to F y 1 inverse phi of z 1; similarly F y 2 inverse phi of z 2 and so on F y n inverse phi of z n, I call this equation number 6. One can repeat this above procedure n number of times as many y or variables you require. So far, in the previous cases we saw that I have a predefined prior understood distribution whose PDF and CDF are known to me, I am trying to fit the random variable for an existing prior define distribution.

(Refer Slide Time: 26:13)



Now, I want to now generate random variable by conditional distribution. Let say I want my random variable should obey type distribution. So, random variable generated by conditional distribution. To generate an outcome of a vector, which is a random variable whose joint PDF, let f of x of x is the joint PDF defined by the conditional distribution; one is interested to generate a sequence of n statistically independent uniformly distributed random variable, which is given by this. So, there should be statistically independent and uniform distributed random variables. Then one can apply Rosenblatt transformations as I am now writing here. So, applying Rosenblatt transformation one can get x 1 is F x I inverse of u 1, x 2 F x 2 x 1 inverse of u 2 x 1 and so on.

(Refer Slide Time: 28:26)

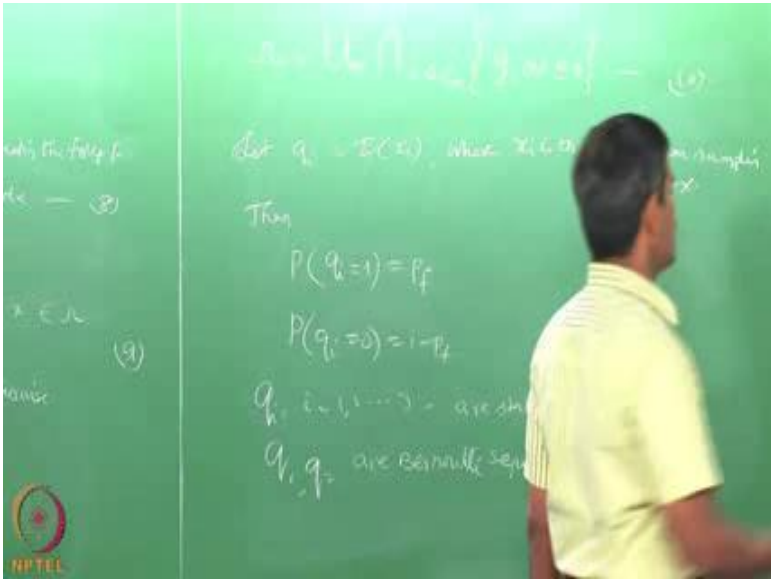


So, in such case, I can write a general expression x n F x n x n minus 1 x n minus 2 of inverse x 1 of u 1 given x n minus 1 x n minus 2 till x 1, call the equation number 7. The above set of equations only generate will have conditional CDFs of x i. The conditional CDFs of the above equations will be let us say x 1 will be x 1 x I minus 1 will be x i minus 1 similarly x i minus 2 will be x i minus 2. One can repeat the above procedure as many outcomes as you want for x which will be the random set of variable which in statistically independent and they will obey to a conditional distribution.

(Refer Slide Time: 30:08)

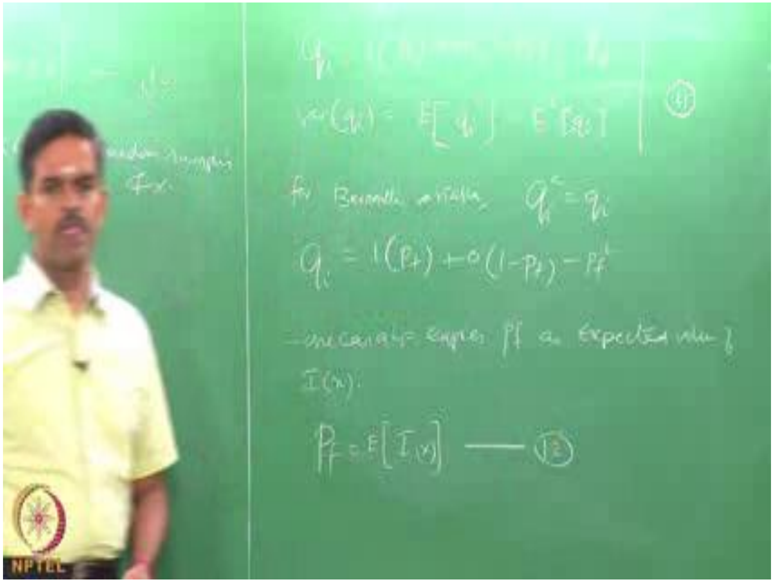
Alternately, one can also use Monte Carlo simulation in this simulation of random variables; the main problem actually lies with the evaluation of specific function which is this function. So, the difficulty is with evaluating the following function integral for x I of x f of x t x, where I of x is an indicator function which is given by 1 and 0 if is an element of sigma or otherwise equation 9.

(Refer Slide Time: 31:36)



Where, sigma is 2 m the intersection I element of C m of the function g i x based on 0. Now, let us say q i is I x i, where x i is i th random sampling of x. Then t of q i equals 1 will be the probability of failure and p of q i equal 0 will be 1 minus probability of failure which is reliability. So, q i, I can v from 1, 2, 3 etcetera which are statistically independent.

(Refer Slide Time: 33:31)



Of course, q 1, q 2 are Bernoulli sequence which can be given by the following equation q i can be given by 1 of p a plus 0 of 1 minus p f. which actually p f. And variance of q i can be said as expected value of q i square minus expected square value of q i. We all know for Bernoulli's variable for let me call this equation number 11, for Bernoulli variable q i square is same as q i. Therefore, the equation can be rewritten as variance q I can be simply or 1 of p f plus 0 1 minus p f, I am substituting back minus p f square that is what I will get as q i. Alternately, one can also express probability of failure as expected value of indicator function. One can also express probability failure as expected value of indicator function that is probability of failure is expected value of I f x.

(Refer Slide Time: 35:39)

Estimator of probability of failure is then given by probability hat failure is expected value of the indicator function which is nothing but 1 by n of summation of I equals 1 to n q i just q bar. Where q bar is the sample mean obtain from n simulation of random vector x. It is important dear friends to see that the probability failure is also a random variable in the present simulation. Therefore, the expected value of this estimator is given by so expected value of p hat I am talking about p hat I will then connect it a probability of failure which can be given by the expected value of 1 by n of summation of q i i 1 to n. Which can also said as 1 by n N probability of failure which can be simply probability of failure of this equation number 13.

(Refer Slide Time: 38:00)

Now, we will get probability of failure hat as an unbiased value hence p hat failure will be an unbiased estimator of probability of failure. And the variance of p hat failure can be given by variance of p hat failure is variance of 1 by N of summation of i equals 1 to n of q i which can be 1 by N 2 of variance of summation of i equals 1 to n of q i which nothing but 1 by N 2 of variance of q i which can be said as 1 by N of probability of failure 1 minus probability of failure equation number 14. Now, the coefficient of variation and further discussion of random variables using Monte-Carlo simulation discussed in the next lecture.

In this lecture, we are discuss about some specific characteristics of random variable how one can generate a random variable using a conditional distribution, how one can fit the generated random variable for a predefined a prior known PDF and CDF for a given function. I hope you will flow this lecture, try to look into the support material from references given in the NPTEL website should you have any difficulty or a questions please post to me, will try to explain as clear as possible.

Thank you very much.