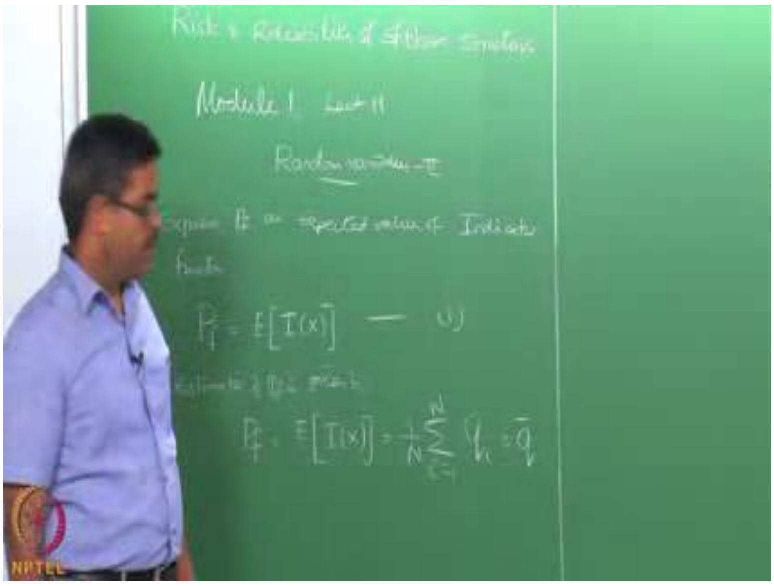
Risk and Reliability of Offshore Structures
Prof. Srinivasan Chandrasekaran
Department of Ocean Engineering
Indian Institute of Technology, Madras

Module – 01
Lecture – 11
Random Variables – II

Welcome to the 11th lecture on module-1 on the online course title Risk and Reliability of Offshore Structures.

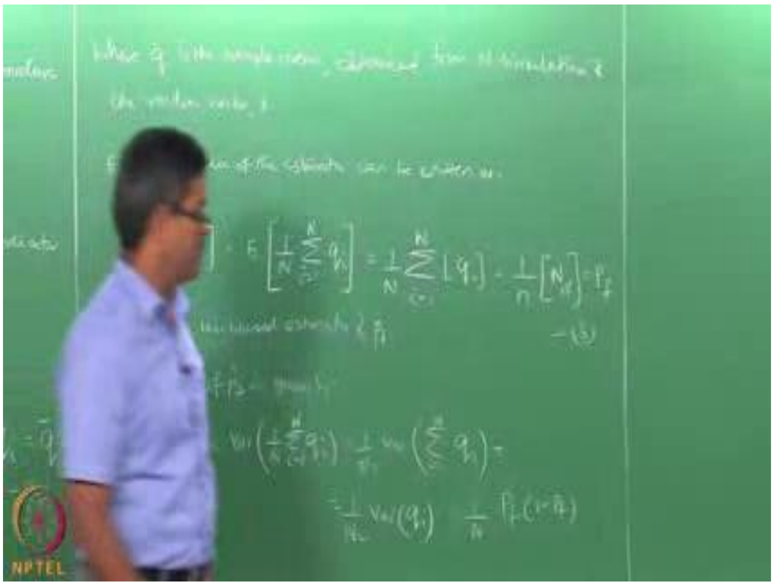
(Refer Slide Time: 00:24)



Today, we will talk about the 11th lecture. I will continue to discuss more issues on random variables. In the last lecture, we already said, what is actually the physical meaning of a random variable, and how a random variable can be generated under different conditions, including (Refer Time: 01:10) type variables. We also said how Monte Carlo simulation method can be used to generate random variable of a conditional distribution.

Let us slightly rewind back, and start from Monte Carlo simulation method. One can always express probability of failure as an expected value of an indicator function. So, let us say, I want to express probability of failure as an expected value of indicator function, which I can write as, probability of failure can be given by an expected value of indicator function. Let this be equation 1. Now, estimated or probability of failure is then given by, which can be 1 by N of sum of N equals 1 to N; i equals 1 to N, because we know this (Refer Time: 02:47); q i, which we know, this is q bar, where q bar is the sample mean obtained, is the sample mean obtained from N simulations of the random vector X.

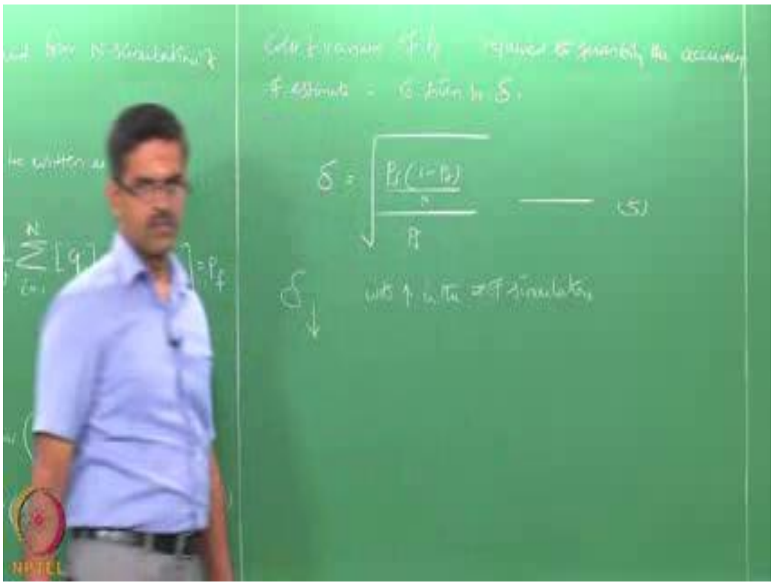
(Refer Slide Time: 03:02)



It is interesting and important for all of us to know that, probability of failure is now also a random variable in the present discussion, because, the argument is a random variable here.

Therefore, the expected value of this estimator can be given by, or can be written as expected value of, let us say, P hat failure is expected value of 1 by N of summation of i equals 1 to N of q i, which can be written as 1 by N of summation of i equals 1 to N of q i, which is nothing, but, 1 by N of N of probability of failure, which is nothing, but, probability of failure. And, we call this as equation number 3. If you look at this equation closely, one can say that, the P hat is an unbiased estimator of probability of failure. Now, the variance of this value is given by variance of probability of failure hat, is variance of 1 by N of summation q i, which can be said as 1 by N 2 of variation of summation of i equals 1 to N of q i, which can be said as, 1 by N 2 of variance of q i, which is nothing, but, 1 by N probability of failure 1 minus probability of failure; call this equation number 4.

(Refer Slide Time: 06:31)



Now, the coefficient of variation of P hat of f which is used to quantify the accuracy of the estimate, because this is required to quantify the accuracy of the estimate, which is given by, let us say, a function delta, where delta can be said as root of probability of failure, 1 minus probability of failure, by N, divided by probability of failure. Interestingly, if you look at the coefficient of variation expression, one can see that, delta decreases, with increase in the number of simulations.

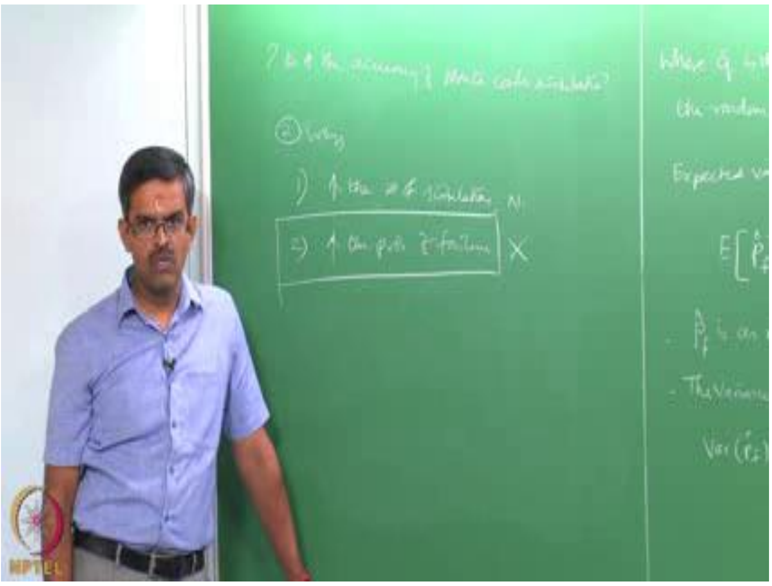
What does it mean? This implies the statement that, the estimate of probability of failure improves, as the analysis proceeds; because, the number of simulations are more and more; as you proceed with the analysis, the estimate of probability of failure will be kept on improving. The coefficient of variance can be used to decide as and when the simulation should be stopped. So now, the question is, what is the tolerable, or a target value? Now, the tolerable, or target value for delta need to be fixed, if you really want to stop the simulation.

(Refer Slide Time: 08:32)

So, we call this as delta target. Let this be delta target and this should be specified, and therefore, if it reaches this value, simulation can be terminated. Usually, delta target is kept between the braces of 0.01 to 0.05, for a typical simulation. So, if we look at an example, let us say, let delta target is set as 0.05; then, approximately, you require 39600 simulations. So, that is the order of simulation. Now, in the whole discussion, we actually do not know the probability of failure. So, the probability of failure is an unknown value, because, we are interested in only estimating this. Therefore, this is unknown a prior to the analysis; only after the analysis, I know this.

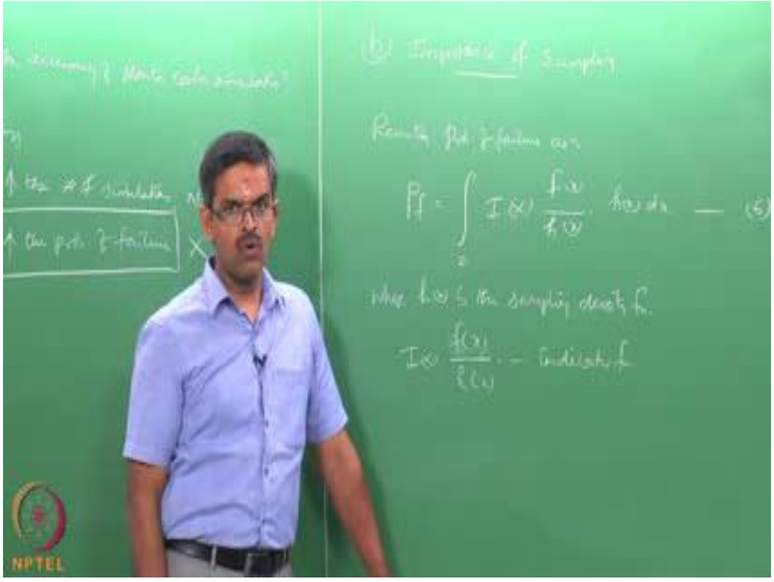
Interestingly, this implies a very interesting statement that, the number of simulations required to compute probability of failure is also not known until the analysis is complete. So, they are interrelated. What target value you fix, that will govern the number of simulation you require. Since you do not know, at what number of simulation you will be able to fix up the delta target, which can give you a probability failure with a closer accuracy, the number of simulations required to achieve the probability of failure, which is said to be an accurate value, is also not known. Therefore, there is a question which is coming in mind, how to improve this accuracy. Another question asked is, how to improve the accuracy of the simulation.

(Refer Slide Time: 11:37)



How to improve the accuracy of the simulation? There are two ways by which we are going to do this. Point number one, by increasing the number of simulations. In my exercise, I am taking this as N. The second could be, by increasing the probability of failure. Obviously, as a risk analyst, or as a reliability engineer, you would not prefer the second option. You want to make the estimate of probability of failure as accurate as possible. So, based upon the better engineering judgment, one will not prefer to increase the probability of failure; rather, one will be interested to address the problem by increasing the number of simulations. So, the moment we agree that, it is important to increase the number of simulation, then, the question comes, what is the importance of sampling?

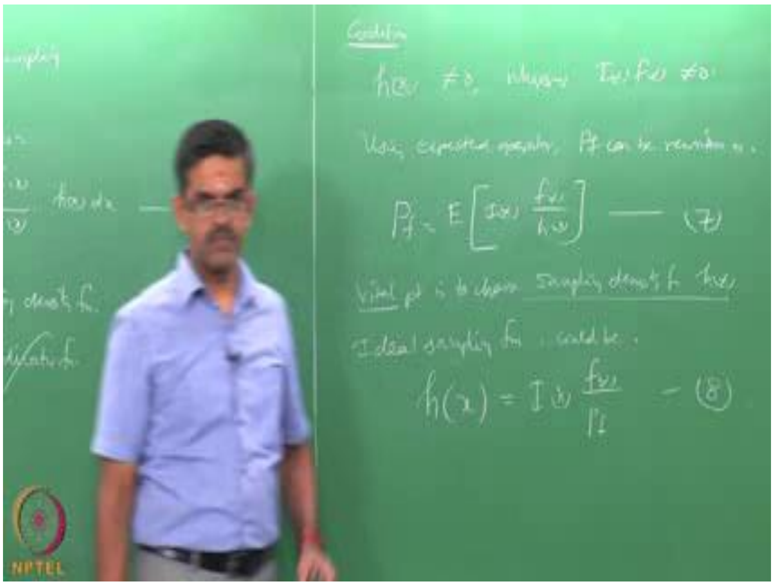
(Refer Slide Time: 13:03)



So, now, what is the importance of sampling? Sampling is important, ladies and gentleman, since reliability estimates are probabilistic based. One of the methods of reducing the variables in Monte Carlo simulation, or estimate of the probability of failure of a component system, is by understanding the importance of sampling. Let us try to rewrite the probability of failure slightly, in a different way.

Let us say, call probability of failure as I x f x h x h x d x, where h of x is the sampling density function, and i of x, f of x, h of x, now becomes the indicator. It is therefore important to note that, the sampling density function h of x is chosen to remain as non0, wherever i of h of x is 0.

(Refer Slide Time: 15:03)



So, the condition is, the sampling density function h of x should remain non–0, wherever i x f of x is non–0. Now, the question asked is, why this condition is essential. This is important due to the fact that, no regions of failure domain are excluded from the analysis. So, wherever it is non-0, we should be able to capture those points, where we do not exclude them in the no regions of failure domain. So, using the definition of the expected operator, one can rewrite the probability of failure as. So, using the expected operator, probability of failure can be rewritten as. So, the probability of failure is expected value of i of x, f of x, h of x.

Earlier, it was only i of x; now, the indicator function is modified. So, let me call this as equation number 7. The vital point in the whole discussion is to choose the sampling density function. So, the vital point is to choose the sampling density function, which is h of x, such that, sampling is done more frequently from the failure domain, not from the safe domain.

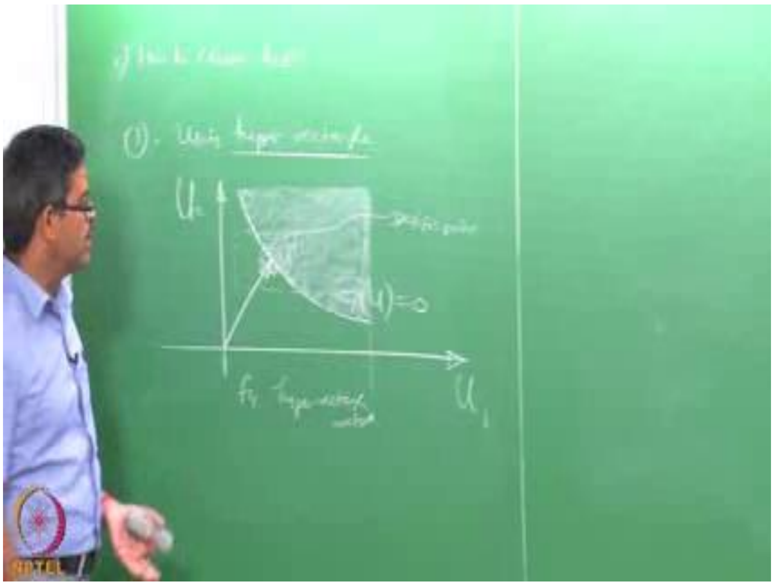
So, therefore, ideal sampling function could be h of x, i of x, f of x, probability of failure; because, I want to choose the sampling density function such that, sampling is done more from the failure domain; that is very important. So, the vital point is, we choose sampling density function such that, sampling is done more in the failure domain. So, if that is the case, then this condition is valid. So, that sampling density function ideally should be given by this expression. For this choice of equation 8 to be valid, mean of the estimate identical to property of failure with a 0 variance for any N.

(Refer Slide Time: 18:14)

So, for equation 8 to be valid, mean of the estimate should be identical to probability of failure, with a non 0 variance, for any N; that is very important condition. But interestingly, this ideal situation is not practical. Why? This situation or this condition is not practical. Why? In simple, it is not practical, because, the probability of failure is what is computed from the whole exercise; you cannot only say the non-0 variance. Therefore, selecting a sampling density function is a critical step.

So, selecting h of x is a critical step. In fact, a very poor choice of sampling density function can increase the variance of probability of failure, thereby making the Monte Carlo simulation very crude. So, several sampling functions are published in the literature; please look at the reference given in the NPTEL website; there are many papers addressing this. We will pick up one such method, and explain this in detail. So, now, the question is focused on how to choose the sampling density function, because, it is very important.

(Refer Slide Time: 20:14)



If the sampling density function is not properly chosen, it can lead to a very crude Monte Carlo simulation. One of the methods, proposed in the literature, one such method proposed in the literature is with hyper rectangle, using hyper rectangle. So, let us try to plot this. So, these are my u 1 and u 2; that is my, this is my domain of failure; that is my hyper rectangle. I shade this; this is my hyper rectangle; this area. Within this, I can pick up any particular value, and choose a design point, which is indicated as u dash. So, the domain here is given as g of u is 0. So, this is one of the method by which the sampling density function h of x can be selected. This method is called hyper rectangle method, given by Shinozuka in 1983.

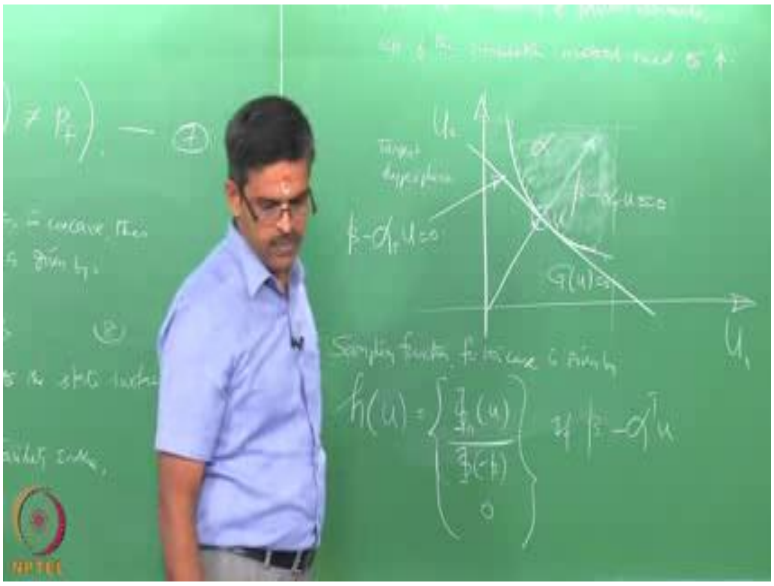
In this method, the whole scenario is about the design point. This point, where the system is tangent, and this becomes normal, is what we call as the design point. The whole difficulty is to choose this design point, because, the whole method is centered about the design point. Unfortunately, the estimate of probability of failure obtained using this approach, is biased, as the procedure does not assign any sampling density to regions of failure domain. The sampling density is not completely assigned to the regions of failure domain. Therefore, the result obtained from this method is slightly biased; mathematically, will not be equal to the true probability of failure.

(Refer Slide Time: 23:24)

So, if it is known that the safe set in the standard normal space is concave, if the safe set in the standard normal space is concave, as shown in this figure here, then, the sampling over the half space can be defined by a different equation. If the space in the safe domain is concave, then, sampling over the half domain is given by alpha transpose u should be greater than or equal to beta, where alpha is a unit normal to the surface, to the straight surface at the design point, and beta is called the first order reliability index, which has been tested upon by Hasofer-Lind later, which we will discuss in the second module; Hasofer-Lind method which focuses again, on how to obtain the design point, and therefore, from that, how to get the reliability index.

Now, to improve the accuracy of this estimate of the failure, to improve the accuracy of failure estimate, one should improve the efficiency of the simulation method, the efficiency of the simulation method need to be improved.

(Refer Slide Time: 25:56)

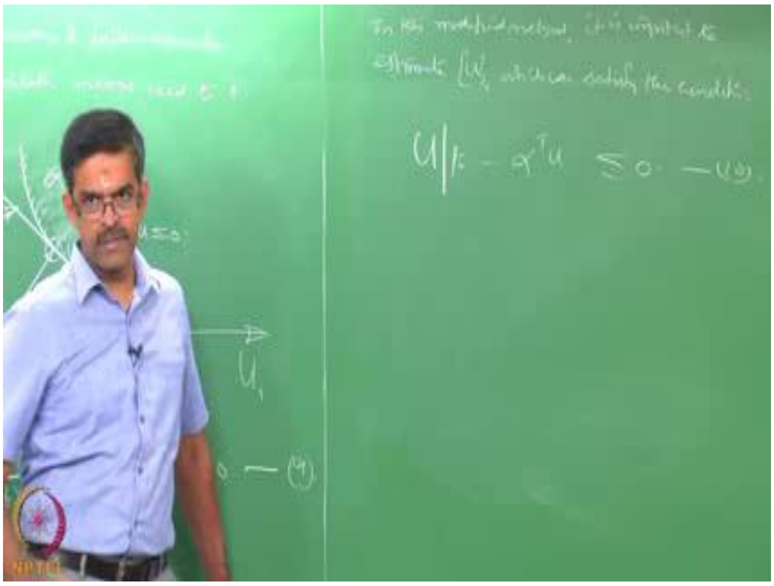


Let us alternately see another method. Graphically, this is u 1; that is a u 2. I have a failure domain again plotted. Let us say, I draw a tangent hyper plane. Let me draw a tangent hyper plane; wherever this plane is intersecting a point, identify that as the design point; and, I can now say, this plane can be beta minus alpha T u is 0. So, obviously, this is going to be, g of u is 0, and this is going to be alpha, and this is now going to be beta minus alpha T u, is less than or equal to 0.

And, of course, we call this point as u dash. So, graphically, the hyper rectangle method is improved by drawing a tangent hyper plane, and then, identifying a design point, based on which, now, I can define the sampling density function h of x accordingly, so that, the failure probability estimate can be improved. Now, the sampling function for this case is given by, let us say, h of u, is phi n of u, which is the normal function, by phi of minus beta, where beta is the reliability index, or 0, if beta minus alpha transpose u should be less than or equal to 0; this is equation number 9.

In this approach, it is important to estimate the value of u, such that, a specific condition can be satisfied.

(Refer Slide Time: 30:11)



So, in this method, in this, let us say, modified method, it is important to estimate u, because, the function is about u, which can satisfy the condition u, given beta minus alpha transpose u, should be less than or equal to 0. There are different steps involved in generating this sample. We will look into the steps in the next lecture in detail. So, now, we are working on random variables, how to generate random variables, what are different methods available, and how Monte Carlo simulation can be improved, what is the importance of sampling, in the whole exercise and discussion, how to choose an appropriate sampling density function h of x, which can improve, or decline the accuracy of probability of failure, because, you must choose the sampling density function h of x in such a manner that, most of the points of h of x, or h of u, are taken from the failure domain. We have also seen in this lecture, what is the necessity to obtain delta target so that, one can truncate, or one can stop the simulation of Monte Carlo method.

Approximately, you require about 40000 simulations, if you say, my delta target is about 0.05, or, let us say, 5 percent. If we do not choose an appropriate sampling density function h of x, it may lead to a very crude way of simulating the sample, which can result in very high inaccuracy of estimating probability of failure. Of course, there are some parallel references available, which have been given in the literature. Please read them, and try to correlate the equations written, and the methods explained in the lecture, so that, if you have any difficulties, please do write to me; we will try to explain it in detail.

Thank you.