Data Analysis and Decision Making - II

Prof. Raghu Nandan Sengupta
Department of Industrial & Management Engineering
Indian Institute of Technology, Kanpur

Lecture – 45
Reliability Based Optimization

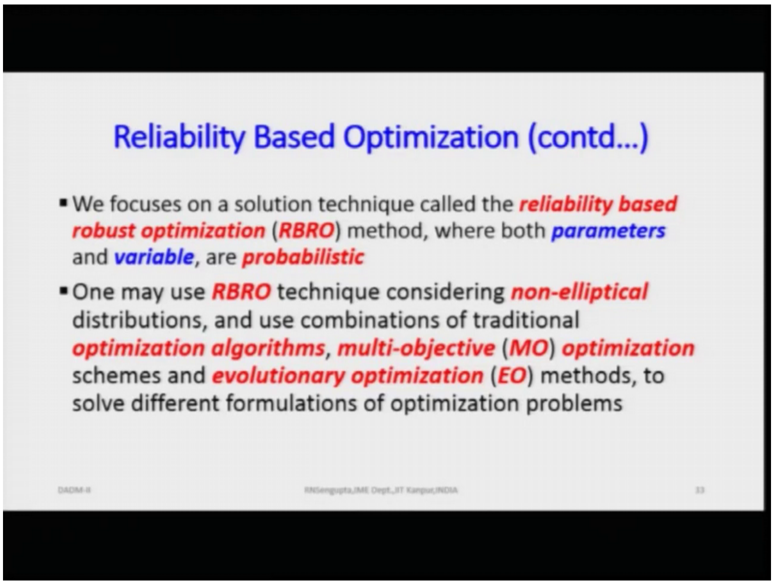
Welcome back, my dear friends a very good morning, good afternoon, good evening to all of you wherever you are in this part of this globe whether in India or Abroad. And as you know this is the DA DM 2 which is Data Analysis and Decision Making II course under the NPTEL MOOC series. And this total course is duration is spread over 12 weeks and the total number of contact hour is 30 hours and the total number of lecture 60. And as you know, this each week, we have 5 lectures, each lecture being for a half an hour and today is the 45th lecture which is the end of the 9th week. So, after this lecture ends you will basically be taking the assignment number 9 and my good name is Raghu Nandan Sengupta from the IME department at IIT, Kanpur.

So, if you remember in the last class which was in the 44th lecture, we will discussing that in the optimization problem and I will come to that re reason why I considered the optimization problem specifically for the multi criteria decision making. We will come to this optimization problem solution techniques later on in DA DM 3 also, but I thought I will give you an basic idea.

So, the parameters which are internal to the system values are unknown. So, you can either estimate, use some computational techniques or at the worst case scenario they are unknown. Other sets of variables which are outside like the example which I gave for the car, the wind speed, external temperature humidity, the tiltation of the road all these things are external. So, these variables are also unknown and they may also affect the overall solution based on which you are trying to find on the best optimal solution of the objective problem.

Now, here we will consider a very simple technique, conceptually I will give you the idea which is known as the reliability based robust optimization techniques where both the parameters and the variables are probabilistic; that means, they are both unknown so, one may use.

(Refer Slide Time: 02:19)



So, if you want to use the RBRO which is robust reliability based robust optimization technique, we will we can consider the distribution based on which you are going to consider the constraints are non elliptical or they are not normal.

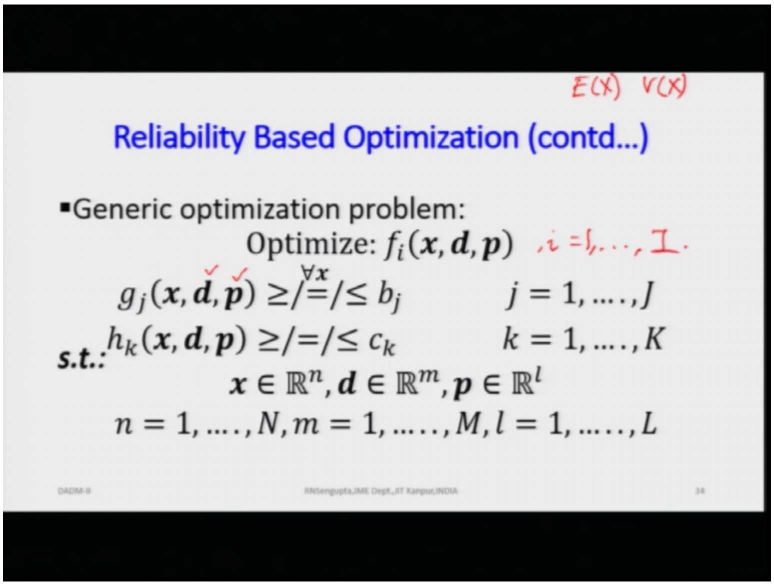
Now, this is very important, I will tell you why and this will make sense, because all of you have, you have done DA DM 1 decision making you must remain, you must have read when I taught was the concept of central limit theorem, the use of normal distribution, use of standard normal table and the formulas for that. Those would be utilized here it in order to make you understand that why non- elliptical and non- normal distributions, if they are it will make things difficult for us to solve in a very simple sense.

So, one may use the RBRO technique considering non- elliptical distribution. And you can use different combinations or different traditional optimization techniques like simplex method, linear programming, branch and bound quadratic programming. Then integer programming concept can be utilized to solve this problem, along with the fact that you would be utilizing different type of multi objective formulation goal programming so on and so forth. You can use different about evolutionary algorithm or heuristic methods you want to solve them.

So, what you are trying to do? You will try to combine this traditional method along with the concept of reliability or stochastic concepts I am not considering the stochastic programming as such I am trying to basically discuss the concept of reliability based robust optimization combining them with the traditional techniques. And give you a feel that how it the problem can be solved conceptually and; obviously, there would be solution techniques for that.

So, this I am a idea based on which we will consider is this robot reliability based robust optimization I will only consider the initial part in the sense that how the reliability part can be formulated in a very simple pictorial sense such that you can get an idea that how it can we can proceed, proceed to solve that. While this traditional optimization techniques would be taking up later on in DADM 3. Now do not be too much bothered about the formulation, I am just giving on the idea basic idea.

(Refer Slide Time: 04:46)



So, generic optimization problem is in a very simple sense like this you want to optimize some f is. So, this f i would be basically going from 1 to N. So, there are basically n number object functions and you want to optimize them. So, this is a basically a multiple optimization problem. So, if you want to optimize the problem which you, I just saw if you want to optimize it can be that you want to maximize the return minimize the risk or it can be say for example, for the again the financial problem.

Consider that you want to maximize the odd moments, the even moments and say ma. Now, let me check how to do it, the variance. So, you will basically try to maximize the odd moments which is the expected value the 3rd moment, the 5th moment so on and so.

And you will try to basically minimize the even moments which is the variance, then the 4th moment so. If I consider the moments, it will be the mean the variance the skewness and the kurtosis. So, if I take the mean and the skewness I will try bit basically try to maximize them. If I try consider the variance of the kurtosis, I will try to minimize them.

So, they can be more than such moments for the problem for finance, they can be different type of optimal objective problems say for example, for the car design it can be you want to minimize the cost can be one you want to basically maximize the strength, you want to basically maximize the speed or you want to basically maximize the crash proofness or you want to minimize the fuel efficiency. So, they can be different type of objective functions. So, those objective functions will be considered under the umbrellas of I is equal to 1 2 3 4 till capital N.

Now, what and you will basically try to optimize this function based on the fact that these x are the decision variables. So, x if there are decision variables, it basically would go in to the case where the values of x, the x value variables would be, vector. So, it can be say for example, I am trying to optimize and I am trying to utilize the problems way where as I mentioned for the car example, it can be the speed, it can be say the strength of the material, it can be say for example, the crash worthiness, it can be fuel efficiency, it can be the boot space, it can be the safety features, other safety features, how fast or how slow the break can be applied or the airbag quality can it can be the boot space, it can be the style and all these things are considered.

So, when I consider this the constraints are as follows. I have the constraints considering that there are say for example, two types of constraints one set of constraints I will consider at the deterministic one, on the other sets of constraints, I will consider the probabilistic one.

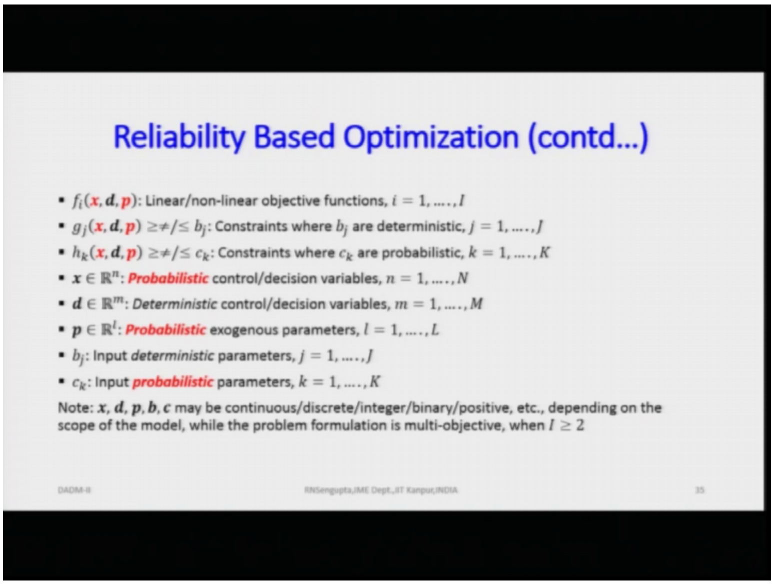
So, the deterministic ones I will consider as g j j basically changes from 1 to capital J. So, there are 1 2 3 4 till capital J number of such constraints which can be clubbed as deterministic. And this x is the decision variable which I have already mentioned. While d and p have their own concept or own notion where p is the set of parameter or the variables. I will not use the word parameters, these are the variable values which have the external influences like if you remember for the car example I had said, it can be say for example, there is wind speed the, the tiltation of the road, the quality of the road, humidity, temperature, snowfall, direction or say for example, the if any hurricane is coming, what is the maximum speed of the of the wind which will affect the stability of the car.

And we will consider the d set of, of PAM vector to be some set of deterministic variables or deterministic parameters which are external or internal to system which we know beforehand. So, these deterministic p sets of parameters are variables are are non-deterministic or stochastic for which we may know the distribution or we may not know the distribution. Distribution can be either the marginal case of the joint ka distribution case and x is the set of variables based on which you want to optimize.

The next set of variables which are capital K in number, small k is equal to 1 to capital K are the set of probabilistic constraints. Again will have x d and p given the same notion while c k I forgot to mention one thing, the b j's which I have for the first set of constraints which are deterministic and c k which I have for the second set of constraints which are non- deterministic are the constraint values based on which you will try to optimize. And we are purposefully taking the sets of constraints as deterministic and on probabilistic in order to make you understand that if they were normal how easily the problem could have been solved.

I will come to that later on within few minutes if you remember I had mentioned that accordingly. The dimension of x is as I said is small n which can go from 1 n is equal to 1 to capital N, the dimension of d is m, m can go from 1 to capital M and dimension of p is small l, small l can go from 1 to capital L and remember this value of i is equal to 1 to capital I.

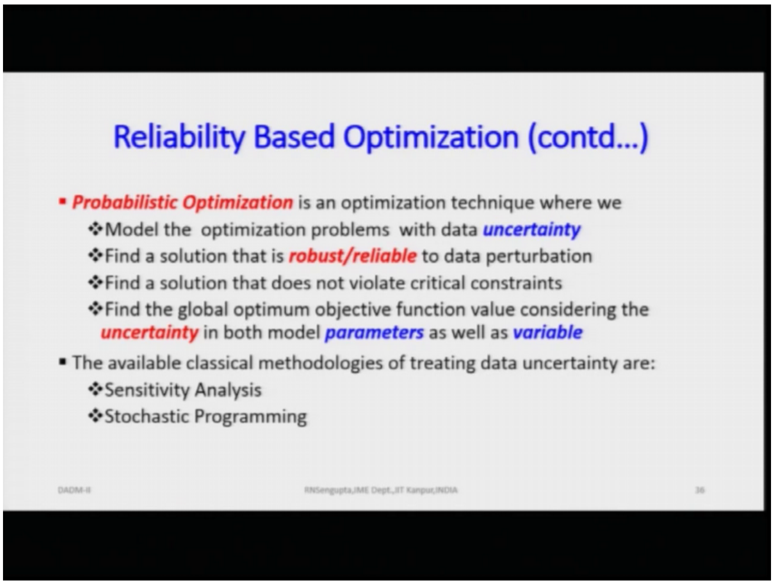
(Refer Slide Time: 11:04)



Now, here it is which have already mentioned I am just mentioning it again the f i's are the linear non-linear objective functions, i is changing from 1 to capital I. And remember the red colour which I mentioned here for x and p, where x is the decision variable which is probabilistic p is the parameter and the variable values which are also probabilistic. While d is the determinist sets, deterministic sets of variables or values for which the values are known. There can be both internal or external to the system, g j's are the constraints where b j's b suffix j's are deterministic nature j is equal to 1 to capital J.

While the constraint h k whatever the h k is constraints are, are probabilistic in nature. Based on the fact that c k is probabilistic, x is probabilistic control decision variables d is the deterministic control decision variables, p is the probabilistic exogenous decision variables and as I said b j's and c k's j's and k's are in the suffix are the input deterministic parameter values and input probabilistic parameter values. So, based on that the idea, we want to basically optimize and find on the best optimum solution constraining is a multi criteria decision making and multi objective decision making.

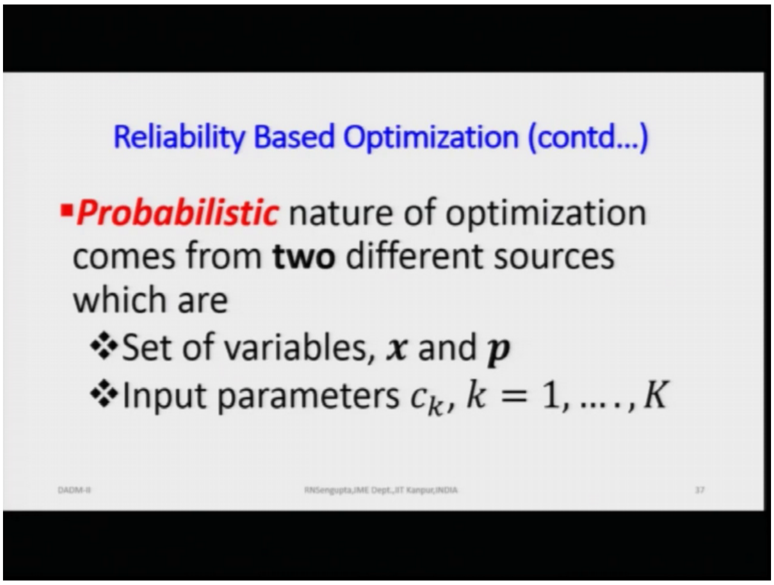
(Refer Slide Time: 12:32)



Now, where the problem comes? So, probabilistic optimization is important in optimization sense, because it models the optimization problems with data and is uncertainty. Data uncertainty can be that there is some reading errors, the actual distribution is unknown to me or unknown to us to the decision maker or they can be underlay distribution which should consider into taken, taken into consideration to formulate the problem. We need to find a solution that is robust and reliable based on the, data perturbation sets or beta data perturbation values which are going to take.

You want to find a solution that does not violate the, the critical constraints of the ka set of constraint based on which we are trying to formulate the problem. So, the set of constraints was basically g j's an h k' s based on which we are trying to optimize. We want to also formulate find the global optimum objective function values considering the uncertainties, certainties which are already there for the parameters and the variables. The available class of methodologies for treating data uncertainty can be through sensitivity analysis and through stochastic programming, but we will consider here the very simple idea that how it can be done using the reliability based robust optimization technique.

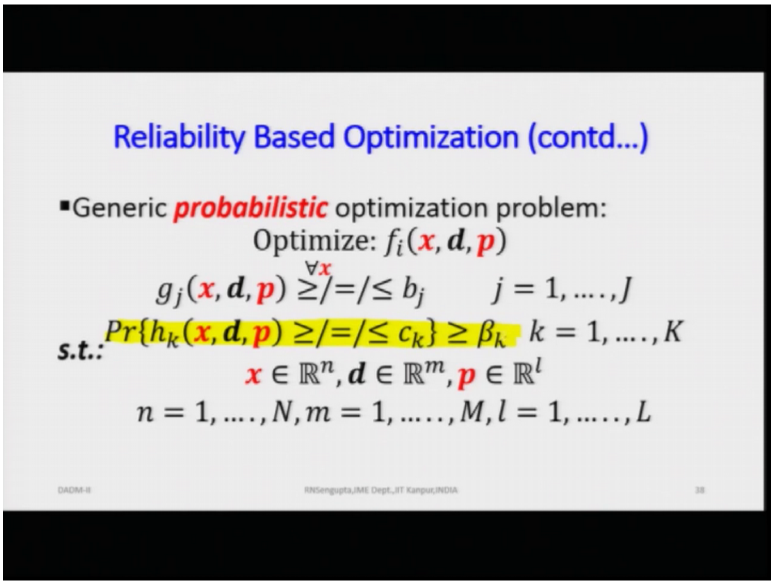
(Refer Slide Time: 13:50)



The probabilistic nature of the optimization comes from two values which I have already mentioned, it is from x which is the decision variables and the p is the parameters which are external to the system point 1. And it also comes from the input parameters which is c k which is the set of the right hand side inequality which you have for the list of non-deterministic constraints which are already there which is c k, k is equal to 1 to capital K.

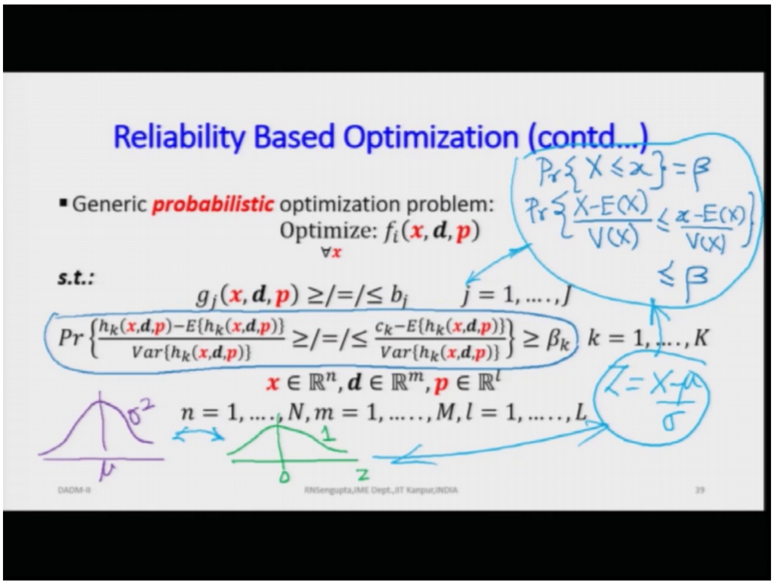
Now, once you are able to see that there is probabilistic values in the or probabilistic constraints in the problem what you will do is that you will write the same problem in a very simple sense. Now I will give you two cases case 1, I will consider the distributions are symmetric and the class is normal distribution. Then I will explain that if the classes are not normal, how do you solve it? The first would be the normal case, which will be easy for us to understand. So, let me first explain the problem and then come to the, the probabilistic constraint considering normal distribution.

(Refer Slide Time: 15:07)



So, the optimisation problem remains the same you optimize f i’s, i is equal to 1 to capital I for a function which is a function of x d and p x is the decision variable p is the parameter, d is the deterministic the set of parameters. The first set of object of the constraint g capital J in number are all mm deterministic and again the function values are x d and p. While the probabilistic part I am considering in such a sense that I would consider the probability that the constraints would be violated or not violated by level of reliability which I will mention as beta k. So, beta k, k values will change from 1 to capital K. Now, let me highlight this in order to make you understand how this can be done.

(Refer Slide Time: 16:17)



This part I will only concentrate on the probabilistic part. So, it will be easier for me. So, let me consider it is normal. So, what I do? So, I will just go slowly and try to formulate it. So, it is easier for enough to so it will be the expected value and the expected value for would be for this function.

So, what I will have? So, let me write it down so it is easy now. So, this is the expected value and this I have, it will be the variance. So, because what I am trying to use is the concept of standard deviate one. See if I do this here so, I have to do in the right hand side. So, this comes here. So, let me select in such a way so; obviously, decrease the font size I think this would be true and let me go through it. So, what I have done, if you check if this distribution is normal which I said for the first case we will consider 2. So, what we are doing is this I will use the blue colour here.

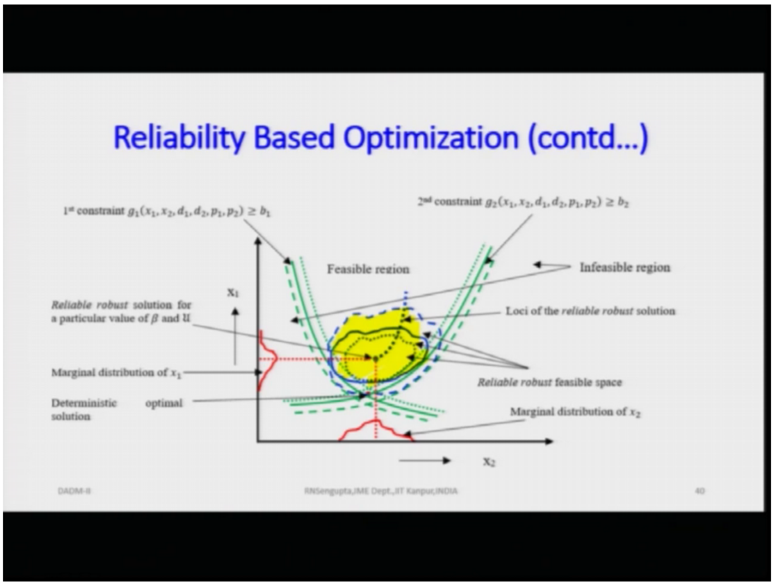
So, actually this whole part, actually it was this probability of capital X either less than equal to small x or greater than equal to small x does not matter is equal to beta which means probability of X minus E x which is the expected value divided by variance of X less than equal to x minus E of X divided by V of X this is less than beta this would always be true, because actually you are transforming let me use, this is true.

So, you are transforming the x distribution is being transformed to the Z distribution this is a normal. So, I have not been able to draw it properly. So, this would be Z distribution with 0 mean and 1 a standard deviation and this had mu and sigma square as the respective mean value in the standard deviation. So, I have been able to transform it and that transformation, we already know Z is equal to X minus mu by sigma which I have used for this case and for this diagram.

So, I have been able to utilize it this in such a way that I will incorporate the concept of normal distribution in order to convert this and solve this problem very simply, but life is not that simple. In the sense the normal distribution assumption which we considered may not hold true in a maximum of the cases. So, if it does not how do we solve it, but before that what problem do we face.

So, this slide I just made in order to make you understand that how the concept of normality would play a role in order to give you a good picture about the concept of reliability or robustness in the constraint problem.

(Refer Slide Time: 21:55)



Now let us spend some time with the diagram it is I have tried to make it as less cluttered as possible, but the diagrams have been such that I will go slowly in order to explain. So, consider this, consider the dark green bold line. So, before that I am considering a two-dimensional problem where along the y axis I am considering the decision variable x 1 along the x axis, I am considering the decision variable x 2. Now when I want to solve x 1 and x 2 using the simple linear programming, I would have a feasible region and I want to find out that point in the feasible region which will give me the optimum solution.

Let me consider the feasible region as this one, the one of the constraints which is given as g 1, I am not going to talk about all the parameter values. Consider g 1 as the constraint. So, this green bold one is the constraint bond with for the first and g 2 this is this green bold line which is the constraint boundary for the second. Consider for the simplicity of the problem that when we solve the problem the boundary of these two constraints in the determinate sense, in the deterministic optimal solution which is the red dot, if you can see it from there.

So, this is the deterministic solution. So, once we solve it we are happy and utilize this for solving our problem, but now consider what problem it occurs if both the constraints are non-deterministic point 1 point number 2, if they are non-deterministic in the simplistic sense if they are normal. And in the third case, we will consider that it is non-deterministic and they are non-normal also. So, let us consider the normal case.

So, first if they are normal case then the perturbation or the movement of the non-deterministic, the probabilistic boundaries would be such that if I consider a normal distribution movement of the boundary space or the boundary means the green bold lines would happen say for example, it happens with a level of probability of say for example, 99 percent. Both of them consider for both of them then; obviously, you will draw a normal distribution and find out what is the total area, area coverage on to the right which is the infeasible set on to the left with the feasible set and solve the problems accordingly.

So, in that case when you consider that both the variances of constraint 1 and constraints 2 as are same then the overall space which you will have between which you will find out the best optimal solution and the centre of gravity of that point would be the optimum solution would be a circle here the circle has not been drawn, because we are considering the non- normal distribution to be true. If the non- dimensional distribution is true then if you want to find out the marginal distributions of say for example, constraint 1 and constraint 2 then the marginal distributions I will use the red colour only.

So, in order to let me use the wait blue what colour do I have to make things, let me use the violet one. So, in this case if the marginal distributions of constraint 1, constraint 2 is normal then the distribution is normal for this case. And if the distribution is normal for this case it is normal, but now as they are non-normal hence the red pdf’s comes. So, let me erase it, they come in a sense that the overall distribution values are such that the proportions of which of if you consider 99 percent coverage, then the overall coverage on to the right or the left of the mean value would be different that is point 1.

Point number 2, if I consider that overall joint distribution of these marginals which are non-normal then rather than me as rather than be a circle, it will be a very weird overall coverage area where finding out the centre of gravity of these areas would be difficult. So, if I consider this small dotted one as it expands, expands in the sense have the reliability increases or decreases this whole area will increase and decrease in such a way the loci of the centre of gravity would move such that it will be important for us to find out the low the centre of gravity based on which you can find out the, the reliable solution.

Now, in the case if it is normal in the higher dimension case in a two dimensional case it is a circle in the higher dimensional case, in the third dimension it will be a sphere. And as you go up, it will be high sphere. So, hence finding out the overall centre of gravity of a sphere or an higher sphere when you are trying to find out the multivariate distribution which is normal becomes easy, but in the case of the reliability based optimization when the distributions are non-normal, we will try to utilize some other method.

In a very simplistic sense I will give you the outline such that we are able to understand how they can be utilized in order to solve the problem. With this I will try to close this 45th lecture which is the end of the 9 week and try to wrap up this concept of reliability based operation in the next class which will be the first class of the 10th week and try to start the new concepts a new topics accordingly.

Thank you very much and have a nice day.