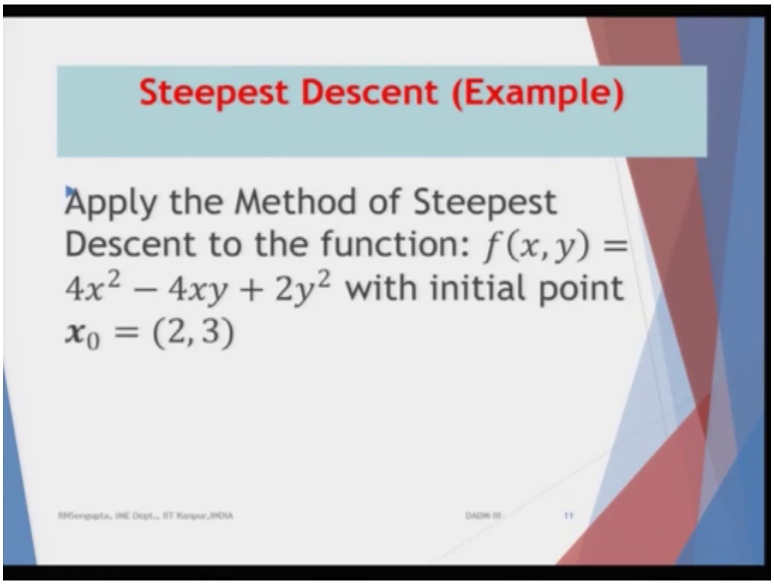
Data Analysis and Decision Making 3
Professor. Raghu Nandan Sengupta
Department of Industrial and Management Engineering
Indian Institute of Technology Kanpur India
Lecture 52

Welcome back my dear friends. A very good morning, good afternoon, good evening to all of you, wherever you are in this part of the globe. And this is the DADM 3, which is data analysis and decision making 3 course under the NPTEL MOOC series. And as you know, this total course duration is for 12 weeks, which is 30 contact hours and total number of lectures is 60. And as you can see, we are in the 11^{th} week, and which is the 52^{nd} lecture, the second lecture in 11^{th} week. (And we were discuss) And my good name is Raghu Nandan Sengupta, I am from IME department, IIT Kanpur. And as you know, we have already taken 10 lectures, 10-week lectures, which is 5 lectures each week.

And you have already appeared 10 assignments, after the 11^{th} week you will appear for the 11^{th} assignment. And after you complete the total course, which is 12 weeks, as I already mentioned, you will have already taken 12 assignments, you will appear for the final examination. So, you were discussing about the steepest, in the quadratic programming, steepest descent method and we will consider a problem accordingly to solve it and then proceed into the concept of elaborating and robust optimisation later on, then again come back in the solution methodology for the simple problems.

(Refer Slide Time: 01:35)



Steepest Descent (Example)

- We first compute the steepest descent direction from: $\nabla f(x, y) = (8x - 4y, 4y - 4x)$ to obtain $\nabla f(\boldsymbol{x_0}) = \nabla f(2, 3) = (4, 4)$
- Then minimize the function $\varphi(t) = f\big((2, 3) - t(4, 4)\big) = f(2 - 4t, 3 - 4t)$ by computing $\varphi^{'}(t) = -\nabla \text{f}(2 - 4\text{t}, 3 - 4\text{t}).(4, 4) = -\big(8(2 - 4\text{t}) - 4(3 - 4\text{t}), 4(3 - 4\text{t}) - 4(2 - 4\text{t})\big).(4, 4) = -(-16\text{t} + 4, 4).(4, 4) = 64 - 32$
- This strictly convex function has a strict global minimum when $\varphi^{'}(t) = 64t - 32$, or $t = \frac{1}{2}$, as can be seen by nothing that $\varphi^{''}(t) = 64 > 0$

Steepest Descent (Example)

- We therefore set $x_1 = x_0 - \frac{1}{2}\nabla f(x_0) = (2, 3) - \frac{1}{2}(4, 4) = (0, 1)$
- Continuing the process, we have $\nabla f(x_1) = \nabla f(0, 1) = (-4, 4)$, and by defining $\varphi(t) = f\big((0, 1) - t(-4, 4)\big) = f(4t, 1 - 4t)$, we obtain $\varphi^{'}(t) = -\big(8(4t) - 4(1 - 4t), 4(1 - 4t) - 4(4t)\big).(-4, 4) = 320t - 32$
- We have $\varphi^{'}(t) = 0$ when $t = \frac{1}{10}$, and because $\varphi^{''}(t) = 320$, this critical point is a strict global minimizer. We therefore set : $x_2 = x_1 - \frac{1}{10}\nabla f(x_1) = (0, 1) - \frac{1}{10}(-4, 4) = \left(\frac{2}{5}, \frac{3}{5}\right)$

Steepest Descent (Example)

- Repeating this process yields $x_3 = \left(0, \frac{2}{10}\right)$
- We can see that the Method of Steepest Descent produces a sequence of iterates x_k that is converging to the strict global minimizer of $f(x, y)$ at $x^* = (0, 0)$

So, we will apply the method of steepest descent to the function in, where f of x has function form of 4x square minus 4xy plus 2ay square and where x-naught the initial point is 2, 3. Now, look at the problem, which is very simple, I am sure all of you know how to solve it. If it was only 4 x square or 2xy, obviously we would have differentiated it with respect to x or y. Found out the second differentiate and found out whether it was grater than 0 or less than 0 or equal to 0 and commented whether it was a maximization problem or the minimization problem.

And we would have closed the discussion and found out the rate of change of the function. Now, here one of the function forms is x into y, which meana that is a function when both the variables x and y are jointly affecting the function form. And we are going to start at 2 and 3. So, this x naught is a variable which I am taking. So, remember one thing this x which is shown bold it only gives you the co-ordinate, this have nothing to do with the x variable in that direction, in that dimension. So, we could have basically denoted like z naught also, but x-naught is basically 2 or 3, where x value is 2 and y value is 3.

We first compute this steepest descent direction, form the rate of change of the function. So, the rate of the change of the function is, let us go one by one. So, the function form is 4x square minus 4xy plus 2y square. So, let us differentiate with respect to x, so it is partial differentiation what I had told earlier. So, it will be 4x square when differentiated with x, will be 8x. 4xy when differentiated with x would be 4y and 2y square will be 0. So, this is what we have, it is 8x minus 4y.

Similarly, when we take the differentiation of this function with respect to y, 4x square would basically be 0, because we are taking that partial differentiation and remember in partial differentiation, I am sorry to repeat it, you will all know that differentiation being taken from y would basically mean x being considered as 0. The second term would be minus 4x and the third term would be 4y. So let us see, so minus 4x is here plus 4y. And once we put the value of x-naught, so in plug it there it will be 8 into 2, which is 16 minus 4 into 12, 4 into 3 is 12, 16 minus 12 is 4, the first value is 4. And the second value is 4y, 4 into 3 is 12, minus 4 into 2 is 8. 12 minus 8 is 4, which is right.

Then we want to basically minimise the function form based on the rate of change of function which is there, it is 4, 4. Del of f 2, 3 is 4, 4. Now, see here what it means is at 2 and 3 the rate of change of function, which is happening in x direction and in the y direction, considering y is constant for x and x is constant for y, is at the rate of 4 and 4. So, obviously if you replace with, not with 2 and 3, with some other value like 0, 0 or 5, 6 what the rate of change function could definitely be different. So, you want to basically find it out accordingly.

So, where you start off, obviously would matter that what is the rate of change of the function. So, then the second point says that, then minimise the function 5t, such that you would basically find out that the functional form, which is 2, 3 when you started of and the rate of change of the function at depending at is 4 and 4 at 2, 3 what is the value. So, once we find out, so it would be 2 minus, these 4, 4 and 2, 3 is probably 2 minus 40, the x value, and 3 minus 40 which is y value. So, computing the rate of change of function again we have 2 minus 40 and 3 minus 40 multiplied by 4 and 4 which is rate of change of the function, gives you the value of 64 minus 32, which comes out to be 32 itself.

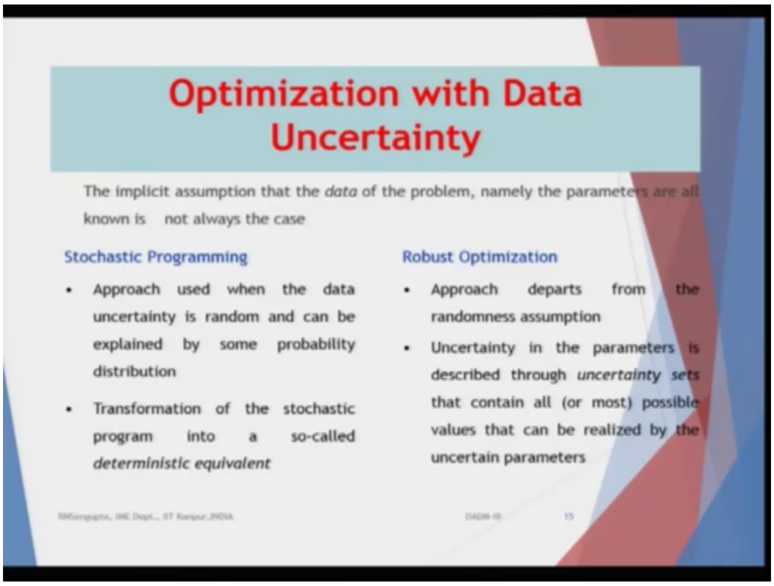
So, this value gives 32 means this is strictly convex because this is greater than 0, not greater than or equal to 0 but greater than 0. So, this is strictly convex function and as the strictly global minimum where the (5), rate of change of the function 5t would be given 64t minus 32 or t is equal to half, as can be seen that by noting down that 5 double dash, which is rate of change of second derivative would be greater than 0, hence it is obviously basically a strictly convex function.

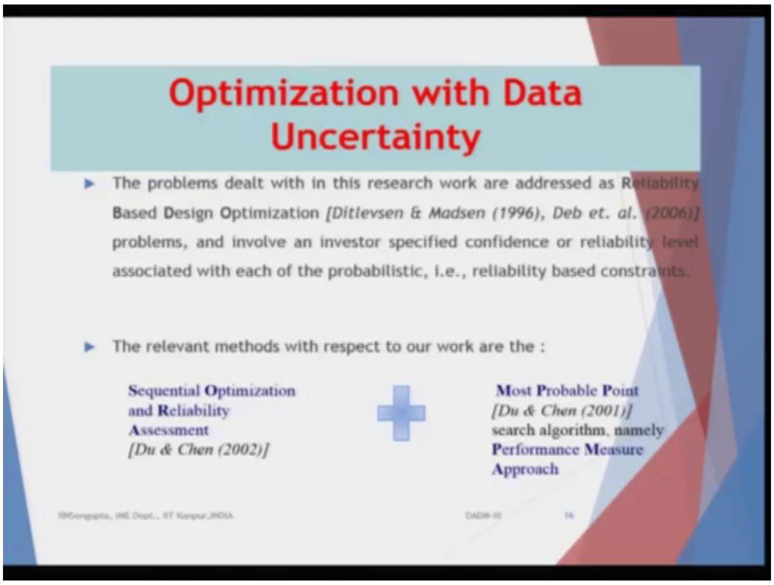
In case it was greater than, it was 0, obviously in that case it would have been a convex function only. So, we therefore set now in the second stage, we therefore set the x-naught value depending on the rate of change of the function which is already there, so 4 and 4 would the rate of change of the function, this half, this is the t value, what is the quantum of t you are going to move, so, as that you reach that position where you can find out x, so x1 comes out to be 0, 1. Again you continue the process, find out the del of x1. So, del of x1 value should be given as minus 4 and 4.

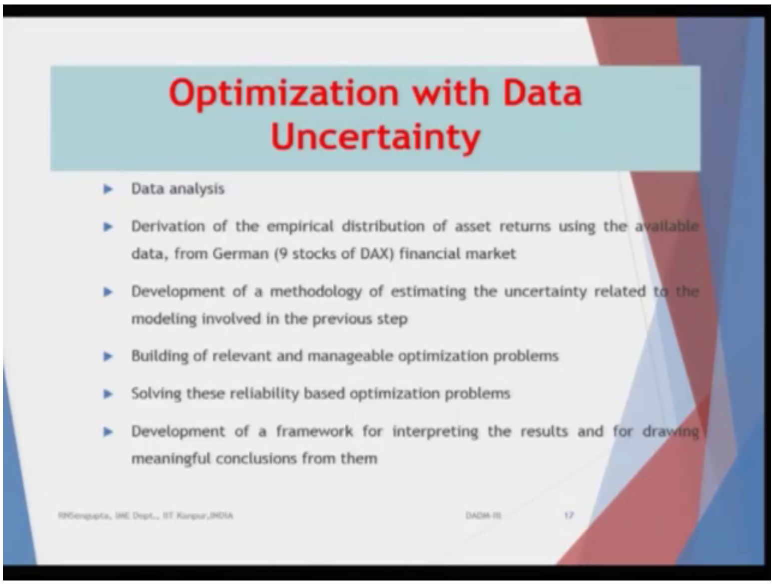
So, this is the rate of change of the function that means x is now increasing at a negative rate which is minus 4, but y keeps increasing on the positive rate of plus 4. We define that functions such that again we find out the value of t, which now comes out to be 1 by 10. And because 5 of double differentiation with respect to t is 320. So, basically again a strictly convex function and once we find out that functional form, we find out x2 which comes out to be 2 by 5 and 3 by 5. Again, you continue doing it till you basically reach the maximum position. So, repeating the process would again yield x3 is 0 and 2 by 10.

We can see the method of steepest descent produces sequence of iterative process that converges tot the strictly global minimum point, x star which will be 0, 0. So, you are basically producing to the strictly global point and you will continue to finding out x4, x5, x6 and so on and so forth. So, what means it is important that were you start would be giving you some information that how fast you can basically proceed to the global minimum or the global maximum considering you are trying to form a problem in the diverse manner.

(Refer Slide Time: 08:42)







Now, we will consider something to do with optimisation with data and uncertainty or stochasticity. And before that I will basically build up the story accordingly. So, there are different type of dates which are available, so data can be for finance, data can be for speed of a car, data can be for the rainfall, data can be for the prices of grains or some products, data can be for, say for example, tensile strength of material, data can be flow of fluids in channels, anything can be considered as data.

Data can be, say for example, the height of students whose height you want to measure to find out that what is the body mass index considering the weight and height. It can be different thing, it can be speed of a car, it can be the rainfall, humidity whatever it is. So, we will basically consider and till now by the way, till now if you remember, in all the problems we have been considering we always consider the data was fixed, that is deterministic.

So, all the variables we have taken, the parameters we have take for the simplest method, for the concept of the duality of the problem, the vougeles approximation method, the north-west corner method, the concept of branch and bound concept. Everywhere we considered or say for example the gomory cut methods, all the processes, we considered the data or the variables or the information based on which we were trying to basically formulate the problem or all deterministic nature. That means there is no variability once the data is fixed it is fixed or the information is given is fixed it is fixed.

The implicit assumption is that the data of the problem based on which you are trying to solve the problem, namely the parameters, we will consider them as the parameters are all known and that may not be true in that case. So, considering that all known has been the fact when we are considering the problem in the deterministic sense but in the probabilistic sense they may not be known. So, that means with some certainty we can say that there is some variance in the data.

So, generally we can either utilise the stochastic programming or you can use the robust optimisation and we will also see the concept of trying to utilise the reliability optimisation. In stochastic programming we will use the approach when the data uncertainty is random and can be explained by some probability distribution, mark these words very important. So, we will consider in the case of reliability optimisation or stochastic programming that there is some under lying distribution based on which you will try to basically find out that what would be the mean value of the data based on which you will try to optimise or what is the variance, which is the second moment, what is the (()) (11:28), all the cortices which is the third moment or forth moment based on which you may be tempted to solve the problem using some methodology.

In robust optimisation the approach departs from the randomness assumption. So, we will consider it is nondeterministic. That means the values are not known. But we would not consider underlying there is any distribution per say which will give us some information that how to model the data and how to utilise the mean value based on which we can proceed. So, that means a non-parametric concept, or some perturbation set would be considered, where the perturbations will give you and the nominal value, will give you some information based on which we can model the robust optimisation methodology. The transformation in the under stochastic programming, transformation in the stochastic programming into so called determinist equivalent would be important we will see that.

While uncertainty when we consider the concept of robust optimisation, uncertainty in the parameters is described through uncertainty sets which I mentioned, that contain or all or most possible values that can be realised by the uncertain parameters based on the fact that there is some perturbation. Perturbation is considering the, there is an atom and it is vibrating, so if it is vibrating obviously it will be partum from its nominal values or the mean values.

So, we will consider that, that perturbation is given by the perturbation set there are different sets which we will consider in the robust optimisation. While in stochastic programming and reliability method we will consider there is a underlying distribution based on which we will proceed. The problems which we will consider basically has been addressed in one of the good papers is which is known as reliability, but there are other papers also, I am just mentioning few of them, which is the reliability based design optimisation method by Ditlevsen and Madsen and (()) (13:28) in respect to 1996 or 2006 in this problems.

And involved, an investor or the decision makers some specified confidence level or reliability level which is associated which each and every probabilistic or reliability-based constraints which are been there in the problem. Now, here we should pause and think whenever you have an optimisation problem, what you have in a very simple linear programming you have an objective function or linear one and there are constraints. So, each and every constraint which area there, there are decision variables and they are right hand side equalities and inequalities. In the simple case, all of them are deterministic and we solve the problem.

But in the case, we are going to consider the, either the stochastic programming or the concept of reliability-based optimisation or the concept of robust optimisation we will intensively consider that both it can be in the actual sense. That both the decision variables as well as the right hand side of the factors which you have or the parameters based on which we are trying to optimise or the objective functions, all of them are stochastic or all of them are have a reliability, distribution or all of them are robust.

Which means the level of confidence which you want to have for each, and every decision variable will depend on the decision maker is going to model it. Now, there can be different, defiantly different kind of reliability levels for each and every constraints we will consider that proceed accordingly. But in order to solve these problems we will consider first the simulation models in the reliability sense and then go into the robust optimisation, give the results accordingly. Some very simple proofs, I will only state the final results and then give the results in the simulation-based concept also.

The relevant methods with respect to our work or all other discussion which we will have would be 2, we are trying to basically take two of the methods which is the sequential optimisation and reliability assessment method and another is the most probable methods such algorithm, namely the performance measured approach which is basically used. So, here what you are trying to do is that you are trying to basically combine two of the methods in the very simplistic sense considering there are two spaces, in space one or x space you have the actual optimisation probable.

You will solve it and when you are solving it you will use either the syntax methods, the vowels approximation method, the branch and bound method, the Gomory method whatever, the quadratic programming, steepest descent whatever it is. And once you get the first iteration stage, that actually x-naught value which you have, the first iterative value based on which you will proceed to the next step if your area only going to solve the deterministic one. You will take that x-naught transform it using the (()) (16:32) transformation, this is one of the transformations I will come to that later on, into the u space.

Now, u space is where you have transformed them into the most probable point concept where you want to find out the probable point depending on the level of reliability which is there. Optimise that x-naught point, which now has been converted into u-naught or u1. Once u-naught is basically or u1 has been optimised to u2, you again back transform or do the inverse transformation into the x- space, get the point of x2, again optimise x2 to x3 star, star I basically mean the each and every optimisation stage what is the result you are going to get.

Once x3 star is found out again you transform into u-space, optimise the u1, get say for example u4 star, again retransform or basically do the reverse function or inverse mapping into the x-space and continue doing this at each and every state. So, as that you at each and very state you basically find you that how you are performance of the algorithm is performing considering the optimum value which you want to reach is basically what is the difference between the nth state of iteration with resect to the nth plus 1 iteration state.

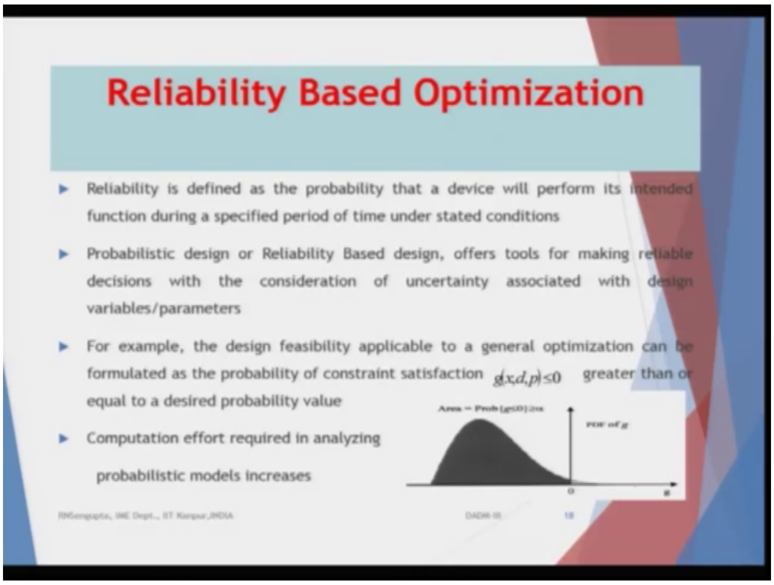
So, as the difference between the iterative process decreases, you basically are assured that you are basically trying to reach optimality case, as the case should be. So, the relevant methods with respect to the work, we will be using the SORA method or the MPP point, which is based on the performance measure approach. So, our main concern would be the MPP point, how you find out. So, what we will do generally we will see that lat. But first let us basically give the data's very simple analysis then we will come to methodologies. We will have our data analysis part, try to basically find out how the data's can be collected, how the in sample and out sampling can be done.

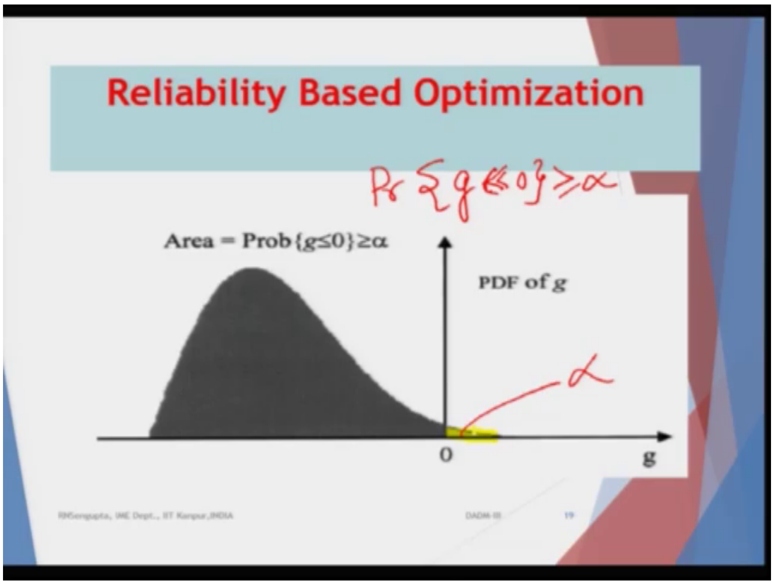
So, the derivations of the imperial distribution or the asset returns using the available data from the German stock exchange and the 9 stocks of the DAX of the financial markets would be considered. You will develop or development of the methodologies would be considered of estimating the uncertainty related to the modelling involved in the previous step. So, if you remember I said the decision maker has a level of reliability. So, the level of reliability which will be proposed by, in order to solve the problem will depend on both the market conditions.

Market conditions, considering from the point of view of solving the financial optimisation point as well as the case when you were considering both the (deci) as well as the decision maker's reliability of the level of confidence also. We will be building the relevant manageable optimization problems based on this fact. Then you will solve this reliability-based optimization problem as I said. So, you will basically optimize in the x- space, whatever the optimization method is, as I said it can be anything, any deterministic one. Once in the first step the x-naught is found out you, you basically transform into the u-space, optimize and find out the MPP point or most probable point and then again do the reverse mapping and take that point x-space.

Continue doing this process of back and forth trying to optimize in the x-space and trying to find the MPP point in the u-space till you are able, you are certain that you have reached the optimality conditions as per the norm which has been set by you. So, as I said you will solve these reliability-based optimization problems. You will develop a framework for interpreting the results and for drawing meaningful conclusion based on which you can say that how good or bad these results are, considering they are not no more deterministic, but they are probabilistic.

(Refer Slide Time: 20:35)





So, what is reliability. So, whenever if you consider the concept of very simple statistics, we say that we are confident. So, we are saying that we are confident and 90 percent of time, it means that if we conduct the experiment 100 number of times, 90 number of times the results basically agrees with the results which is biased or 90 of the times we agree with the results and 10 percent of the times it is not true.

So, what we want to do is that, first we want to basically have the concept of reliability also mathematically modelled in the same way as we basically say about the concept of level of confidence and then see that the concept of level of confidence which we see in (()) (21:18), how it can be in the general sense of (()) (21:22), how it can be brought down into the picture and basically utilised accordingly. So, reliability is the defined as the probability that a device will perform in the mathematical sense, will perform as intended function during a specified period of time under stated conditions.

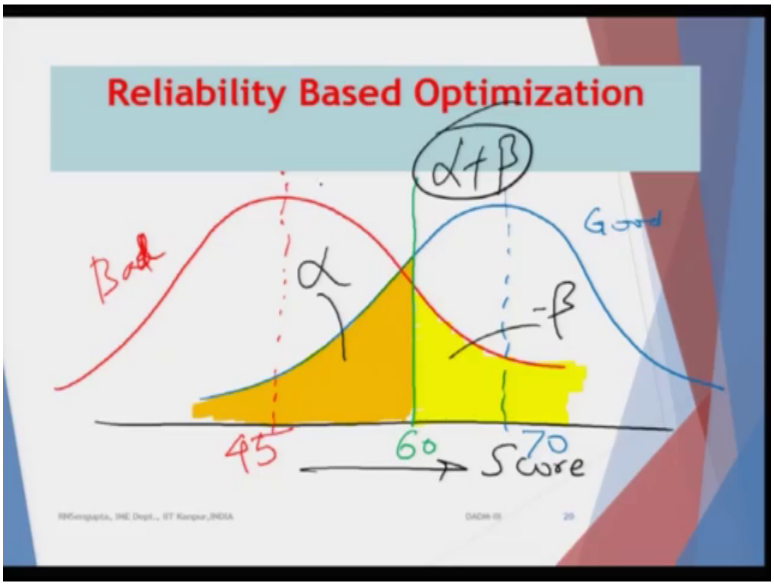
So, as I said 90 percent- 95 percent whatever it is. Probabilistic design or reliability-based design tools, they are they offer and help us in making reliable decisions with the consideration of uncertainty associated with the design variables of the parameters which are there. So, obviously whenever there are parameters, whenever there are variables, they are non-deterministic many of the cases or maximum of the cases, so how do we incorporate that to find out the level of reliability based on which we are confident that we can basically state that it does not with a certain probability. For example, the design feasibility applicable to a general optimisation problem can be formulated, as the probability of the constraints satisfactions.

So, the constraint satisfaction is you have a functional form of gxdp, where g is basically a function and x is the decision variables you want to take, d are the deterministic parameters which can be scalar or vector and p is the set of probabilistic parameters which are there and it is less than or equal to 0. So, it could have been greater than or equal to 0 also, and obviously equality can also be there. We want to basically find out that it is greater than or equal to desired probability level.

So, this value that gxdp, this is the constraints is greater than or equal to or less than or equal to some value of reliability which is there. So, what we do is that, actually we want to basically have this function, probability of g is less than or equal to 0 and probability of that being true is alpha. So, what we do is that you basically plot that function, considering it is 1 dimensional 1 or 2 dimensional 1 whatever it is. Now, once you plotted, if it is less than 0 obviously the 0 value is here, so I will basically try to draw it in this, so this I will expand.

So, here it is. So, in the 2 dimensional 1 I have been able to plot consider, I want to assume probability of g less than equal to 0, so these are function, is greater than alpha. So, the function form is here and less than equal to, which means there is alpha, and this is the area. So, it could have been on the right-hand side also, depending. Now, in the two in the higher dimension it can be done accordingly. Now, what actually it means is this, what we will do and before that I will try to basically give you a simple example from the hypothetical case. I have already discussed this problem in DADM 1, but I will basically build up the background accordingly.

(Refer Slide Time: 24:50)



So, the problem is like this. I will first draw the diagrams and then come into it. So, consider the problem is like this, as you can see in the slide. So, there are two normal distributions and on the x axis I have the score, so the background is like this. You are the manager of a bank and you are approached by different people to basically tell or ask them that they ask for a loan. And obviously before you give a loan or you take the decision whether a loan should be given, you basically mark some, give some points to the background of the person.

So, the background is an whether he is working, whether he has a government job, private job, whether he or she has a business, whether he or she has own house or whether the person has his own vehicle, whether the person has an account and what is the overall credit worthiness of the key person with respect to other loans he or she has taken, so based on that you basically assign different points. So, once you assign a point you think that for the set of people whose points are 60 and above, you will basically give the loan and any person less than 60 would deny that loan.

And consider that the set of persons who would be given the loan, which is blue distribution, normal distribution and their average scores is 70 and the set of people who would be denied the loan, which is the bad one, is the red in the normal distribution and their score is 45. So, what is interesting is this, I will basically have two areas, this yellow one would be the set of people who should have been denied the loan, but they are getting the loan.

And this orange one is the set of people who should be have been given the loan or the first person would be the people who had been denied the loan, but they are getting the loan. And the second set of people which are now, highlighting, is the set of people who should be given the loan, but they are not given. So, these are errors and these errors I will basically use as the concept of, so this I will consider as alpha and this is considered as beta, so look at the interesting fact. I tried to basically shift the 60 on to the right or the left. So, shifting on to the right basically increases alpha, so alpha is what, it is a sort of a business loss and what is beta, beta is basically bad dept which is lost.

So, and in shifting that 60 point to the right will obviously increase alpha but decrease beta. On the other hand, trying to shift 60, the average personal loan, if you shift it to the left it will decrease alpha but increase beta, so what is to be done. The main fact would remain, that trying to basically find out, that to minimize alpha and beta at same time is not possible. So, the best we would basically to be, to minimise the sum of alpha and beta.

But in probable, in (()) (29:11) testing what we do is that we basically keep beta fixed at certain level and try to basically find out the level of confidence which is alpha and based on that fact we basically pass on the decision, whether h-not is to be accepted or h-not is not to be accepted or rejected. So, with this I will close the 52nd lecture and continue more discussion about the reliability point of view and then go into the concept of robust optimisation later on. Thank you very much and have a nice day.