Data Analysis and Decision Making-III
Professor Raghu Nandan Sengupta
Department of Industrial & Management Engineering
Indian Institute of Technology Kanpur
Lecture 24

Welcome back my dear friends. A very good morning, good afternoon, good evening to to all of you, wherever you are in this part of the globe. And as you know, this is a DADM three, which is data analysis and decision making three course under NPTEL MOOC series. And you are all aware, so please excuse me if I am repeating. This course total duration is 12 weeks, which total contact hours is 30 hours. And which when converted into number of lectures is 16 number. Means each lectures is for half an hour. And considering the fact, that each week we have five lectures. Each being for half an hour.

After that we have an assignment. So if you see the slide, we are in the fifth week, 24th lecture- that means the 24th and 25th will basically end your 5th week. And all of you have already taken four assignments. After this you will take the fifth assignment, after the 25th is over. And after the end of this course, you will basically have the final examination to take for the evaluation. And my good name is Raghu Nandan Sengupta from the IME department at IIT Kanpur.

Now in the last ~~(1:28)~~ (())(1:28) ~~of~~ end of the 23rd lecture, I was repeating one of the slide, which I have already discussed. The reason being that it was a very generic nature of how you write the problem. You basically have ~~a sets,~~ a set of objective function. And we initially, for all these number of days we considered, that there was only one, one objective function with some assumptions- like linearity, or the prices, proportionality~~.~~ ~~T~~they can be broken down into number of infinite small decision variables, all these things.

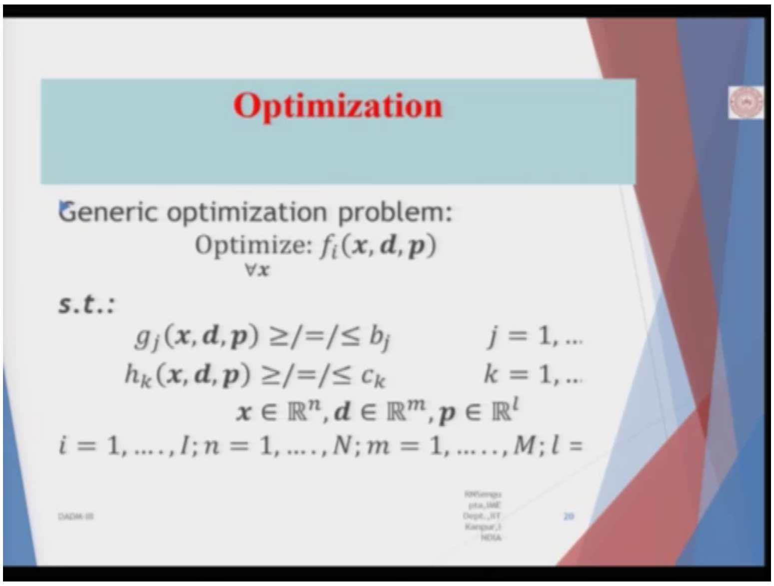
And the constraints were also, not probabilistic. They were all deterministic. If you remember, I did repeat it time and again. The constraints were specific to either the greater than type, less than type, equal to. And equality basically came up later, as we added the slack and the surplus, depending on how you formulate the problem. We also considered that the axes were, as greater than zero as per the ~~options~~assumptions.

Now, generally whenever you are trying to solve the problem, in one of the problem we saw that the slack and surplus, technically which should have been zero are not zero. In the sense that some amount of slack or surplus would be left, in order to basically attain the maximum, which is the basic feasible solution which was optimal. So we will try basically dwell on that, about sensitive analysis and all this concept as we go ahead.

So again this slide comes, where we are considering in an optimization framework of maximization or minimization, whatever it is, of only one, I is equal to 1, you will basically have a single objective. And if it is a linear, it will be a linear problem. And the constraints which was considering, like you consider the constraints, which was there for factory 1, 2, 3, which is plant 1, 2, 3 or machine 1, 2, depending on how your formulated the problem. The constraints were basically deterministic in nature. And as it says, like G Js or H Ks, are basically the constraints depending on the set of variables, which is X.

Here all we are taking are X or D. D is basically the deterministic constraints and deterministic parameters. And P is the probabilistic parameters. They are all vectors. And number of vectors would denote the variability in the problems. So X basically be of dimension N. It means there are N number of variables which ~~leads~~ need to be solved to arrive at the problem, at the solution.

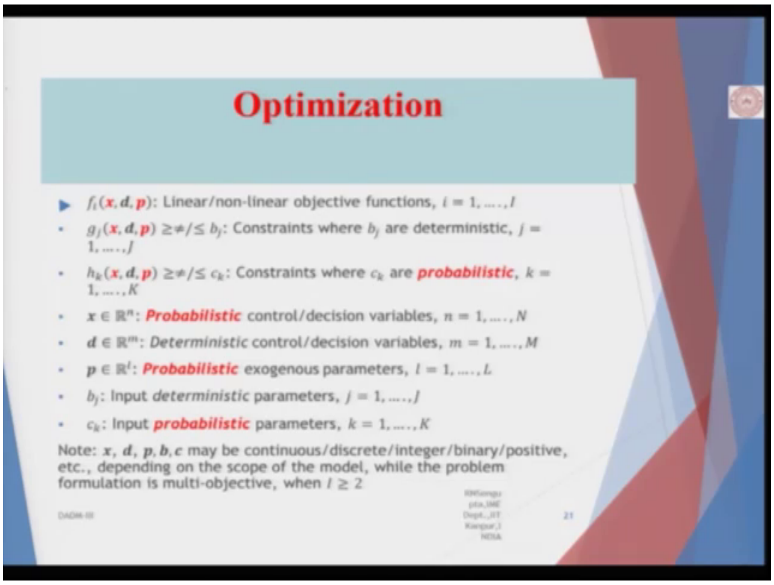
(Refer Slide Time: 04:25)



Now when you are considering G Js or greater than equal to less than. So it obviously means that it would be, there would be some slacks or surplus, depending on how the problem is formulated. But the issue of basically trying to have another set of constraints K in number, would K is equal to 1 to capital K. And J is equal to 1 to capital J. Basically we segregating the sets of constraints into some which are probabilistic and some which are deterministic.

Now this will come later on, when we, when we give concept to the diagrams, which I will come within few minutes.

(Refer Slide Time: 05:09)



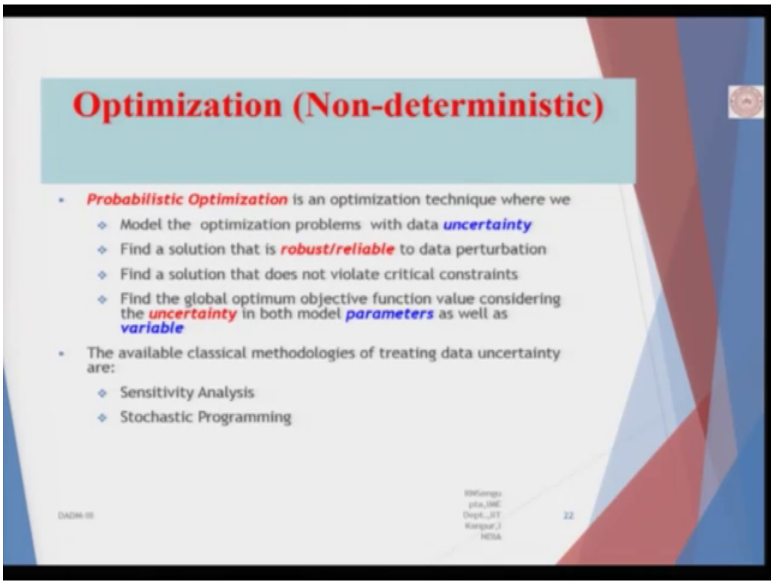
Now we have the linear, nonlinear objective functions- I is equal to 1 to capital I. So if it is linear and I is 1, we had whatever we have been discussing till now. The linear of the objective function optimization problem, simplex method we have been utilizing. And in case, if there is no probabilistic constraint. So if you see the third bullet point that HK X,D,P. And we are in; here in this problem we are considering the red color to signify the probabilistic nature of either the decision variables or the parameters, exogenous parameters or the input variables also.

If X is not a probabilistic or P is not probabilistic or C is not probabilistic, obviously all of them would be deterministic. And you basically solve the problem as given. In the simplistic sense whether the linear programming or the nonlinear programming. And combined, you have the constraint as greater than, less than, equal to. Where on the right side, you have the BJs. J is equal 1 to capital J. So B concept I am taking is exactly as it was there in the operation, optimization problem we had solved - AX is equal to B. Now if D, P and C are all deterministic, then the set of constraint basically becomes capital G plus capital K. And that is exactly equal to the number of rows and the number of columns, which we have for the problem, when you have basically solved using the Gauss–Jordan concept.

Now why I have considered a little bit before hand the probabilistic one, is to give an intuitive feel that what we will be doing later on, in a very simplistic way, was . First to consider that the parameters which are not under our control and some of the decision variables which you want to take are I have some distribution. Like say for example, you are designing a car. And you know that the weight of the door or the weight of the chassis sash, whatever you want to design or a weight of the roof of the car, is under your control. So which is a parameter, which is given by the designer. And your main decision problem is basically to find out the cost, reduction in the cost, minimum cost. So, these would be set of deterministic parameters.

While the parameters which are probabilistic, which are not under your control, which you know has to be considered or they have to be considered in the problem, would we say for example- the speed of the wind or say for example, the road condition. All this things would be parameters which are non-deterministic, probabilistic. Here X,D,P, BC maybe continuous, discrete integer, binary or positive or whatever the framework of the domain of this X, D, P, BC would be depending on the scope of the problem. And as I mentioned, which I am mentioning again, that I value would be 1 for the single objective function. Would be more than 1 for a bi-objective, if it is two, it is a bi-objective problem or higher it will be more than 2, which is multi-objective problem.

(Refer Slide Time: 08:40)



Now where the actual interesting part comes up is, we will discuss in the diagram. So in the simplistic case of a probabilistic optimization, this would be an optimization technique, where the model or when we solve the model constraints that there is some data uncertainty. Data uncertainty as I mentioned in the car design problem. It can be say for example, for design the dam. It can be for the designing of a portfolio. It can be designing of different heat chambered or hook cooling towers in mechanical design or civil engineering design. It can be designing of say for example, the amount of money ~~I am~~ you are going to spend for investment in government projects, where the returns are flexible, which are not under your control depending on the market conditions. So there can be different examples based on which we can understand the problem.

So you will model the optimization problem with data uncertainty, which I have mentioned. And, when you are trying to solve the problem, when it is not deterministic, you will basically take some robust and reliable solution. Robust and the reliable, the word reliable is what we will be coming back in the deterministic sense also. We will consider some robust and reliable solutions. Reliable means to what degree they are true. Or with what certainty we can say that these answers would be coming out or to what level of confidence can I say, that yes these values would be true for the problem I am going to solve.

And robust in the sense, that any perturbations in the variables, in whether the deterministic one or probabilistic one, would keep our answer as close as possible to the deterministic value, which you want to find out. Such that, we are certain with a high amount of confidence that the answers would be almost the same, even if the external parameter values are changed. So that would obviously would be the answer.

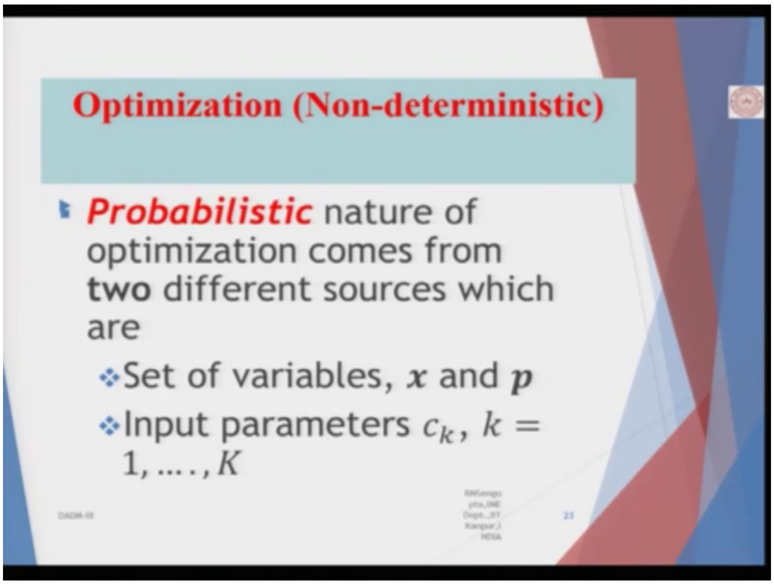
Because in case, say for example, we are deigning a car. And the cost factor is coming out to be depending on whatever different type of materials we use. And if the, if the information set from outside, like the wind speed will vary largely or the road conditions would vary largely. Even if that factor is there, and if the cost are as low as possible. And they are in and around the value, which you have found. Then we are certain that our answer or optimization performance has been very robust, in the sense that we are able to give good results as, as decided.

We would definitely try to find a solution. And that may not be very applicable for the, the deterministic sense. We would definitely like to find out solutions, which do not violate the basic, ~~the basic,~~ the feasible region which we have. That means, the problem which you want to find out, the solution which you want to find out, would be one of the solutions along the corner spaces. And if it is not, obviously we will try to check whether they are violating the constraint, whether they are out of the feasible region.

So that should not happen. Like yes, the external factors are there. But that should not affect the answer or the feasible point to be infeasible, now become infeasible. So that has also to be looked into. We would basically find an~~, an~~ objective function, optimum objective function, considering the uncertainty in the initial case. And then, slowly relax the uncertainty later on. And check how the answers in the uncertain case, considering probabilistic variables, probabilistic P values, probabilistic constant, how they affect the deterministic answer.

–And we will consider both at the same time, such that the both the variables- decision variables as well as the parameters are uncertain. So obviously there would be a joined distribution. How we will model that, we will see later part in this course. Now the available classical methodology, based on which we treat data uncertainty would be- sensitive analysis, which we will consider in few slides later on. Either in the twenty fourth lecture or in the twenty fifth lecture. Or else, we use stochastic programming, robust optimization, reliable optimization based on~~,~~ how we have been able to model that problem, ~~concerning~~ considering with the distribution is one of the factors we are going to consider. Or it is basically distribution would be there. But we would not be considering parametric form for that distribution as such.

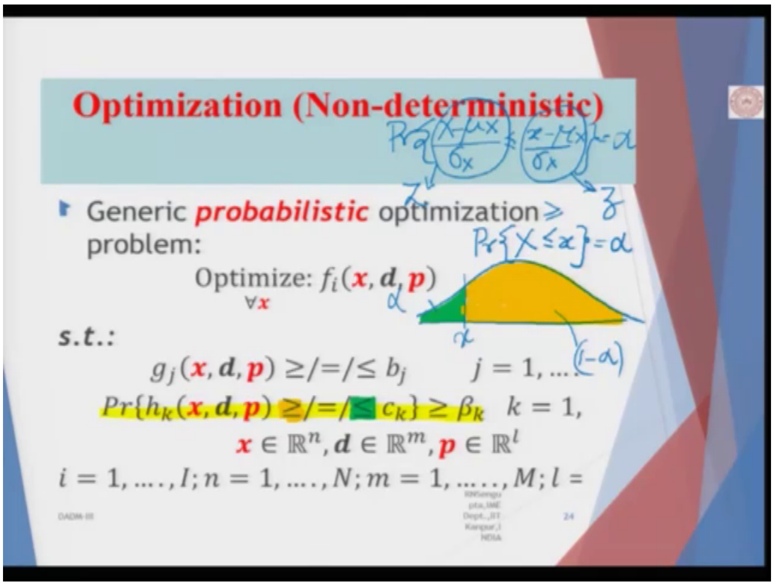
(Refer Slide Time: 13:25)



Now whenever you are doing this probabilistic one. And will come later on the sensitive one. Sensitive one is basically how sensitive the answer is with respect to the perturbation of the constraints or perturbations of the variable or perturbations of the parameter values. So probabilistic nature of optimization comes from two different sources. Set of the variables, which is X and P and how they are changing and how the parameter values, external parameters are changing. And it will also be depend on the input parameters.

If you remember, CKs were also non, non-deterministic. So how they are changing and what effect they have on the constraint would also dictate because if the right hand side CKs are changing, obviously it would mean that the constraint space or the feasible region would also be changing, which means that you will have different sets of interior point based on which you will do a search, in the very simplistic sense what I am saying.

(Refer Slide Time: 14:22)



Now when we come to the optimization problem, so there will be some philosophical change of a modulate. So in the general, the probabilistic; so the problem or the model which I have discussed, in a very general sense was the deterministic one. Now, if you bring probabilistic sense in the optimization problem, then you will basically have an optimization, where X and P are non deterministic. Hence, they are shown in red color. That means the decision variables also is not known with certainty. And the parameter values which are affecting our decision model is also definite not under our control.

And that is, there is a distribution, if we know it is well and good. If you do not know, then you have to basically assume something and proceed accordingly. Now we divide the total constraint into set. So initially it was deterministic. You will only consider the deterministic part. So there would, the total number of variables would be as I mentioned would be capital J plus capital K. Capital J is the number of constraint which has, which are being considered as per this slide here, a deterministic. And the capital K number of constraint, which are now probabilistic.

Here is why they are, the interesting part is. If they are probabilistic as per the concepts. So obviously, the right hand side K, CKs are unknown. So they are changing, that means they are perturbing. So, if there perturbations are there, the movements are there, they are shifting either outside or inside, based on the, the initial surface which you have. So that it is a straight line or a plane or a hyper plane, depending on how the problem has been formulated in, in one dimension; in two dimensions, three dimensions or higher dimension.

But now, a probability will come or a chance or the level of reliability which you want to assign. Now this concept of reliability would be formulated into fronts. Two fronts is that, if we consider the distribution per se to be normal, which we will consider to make our life simple. Then we will basically consider the perturbations in such a way that we will consider, we will include normality. And solve the problem, which I am already told many times but I still repeat. So if you (con), if you see the set of capital K number of constraints here, which I am highlighting. Then you have a reliability beta. So how reliable you are? It is beta1, beta2 beta3, which is 90 percent, 95 percent, 97 percent, 99 percent. Higher and lower the values of beta will give you, how reliable you are, how less reliable you are.

Now think from the normal distribution. So if you consider the normal distribution, and I am from my side, if I am considering the values are increasing. So X has basically from minus infinity to plus infinity. I am not considering standard normal. And the probability distribution for X is given. The middle value is the mean median mode is mew suffix X, which is the mean value. And obviously, there is a standard deviation.

Now if I consider this, this highlighted part and I write like this. So I am using the color blue. So generally, we know probability PR of capital X less than equal to small X, if this is equal to alpha. So what we do is that we have the distribution, is a univariate normal distribution. It can be converted into multivariate normal distribution also; the problem would not be that difficult.

So you can consider some X values. And you want the X is the realise ~~X~~ value, what we can understand. And we want that, so I will use color highlighter green. So I would basically consider capital X is onto the left hand side or right hand side depending on the problem has been formulated. So this less than type would come for the case when you have this less than type. So in case, it is greater than type; let me use another color. So this is greater than type, this is the greater than type. And in that case, it will be assume this, where the problem formulation would be of the greater than type.

So when you are considering the normal distribution, life is very simple. So what you do is, I will just write it here only. So you will consider the probability- X minus ~~mew~~ mu X by sigma suffix X. So this sigma suffice X or ~~mew~~ mu X what I am considering is in the univariate case. Corresponding to the fact, if the multivariate case, you will; you can convert using the multivariate normal distribution or whatever the dimension is. And less than equal to this small X ~~mew~~ mu X divided by sigma X is equal to Alpha. So you are basically transforming from the X distribution to the Z distribution.

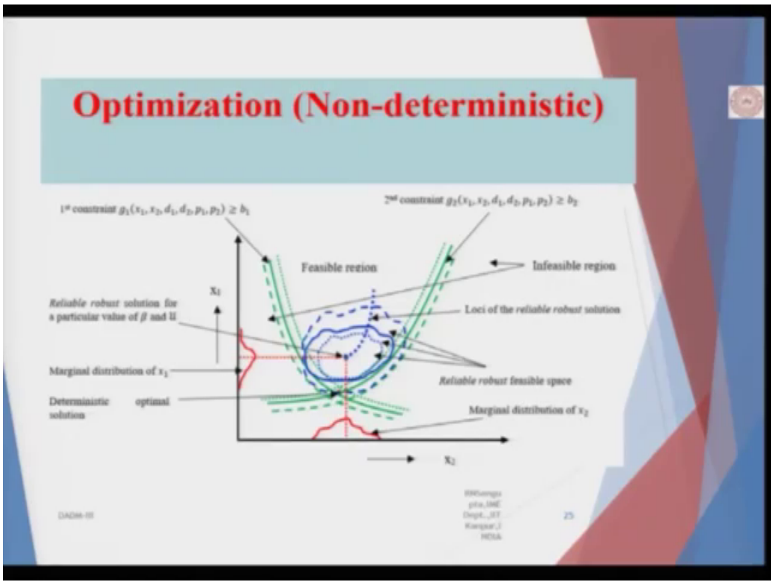
And when you transform, what you are doing is that you are converting this X minus ~~mew~~ mu X divided by sigma into capital Z, which is the standard normal deviate. And this one- X minus ~~mew~~ mu X, small X minus ~~mew~~ mu X by sigma as small Z, the real Z value. So this would mean that I would basically have the distribution centre normal, find it out, plug in and finish my answer and solve the problem.

Now the question is, I am coming to the sensitive analysis. But question is that what happens, if the distribution pulse is not normal. It can be anything. How do I find out? So either you do simulation, that means you simulate it in such a way, that you find out how many number times out of thousand simulation, million, 1 million simulation, 2 million simulation ~~that~~ if we do, how many times this is valid. Based on that fact, you will find out the alpha value. Because alpha value is not changing. The overall area, which I have here, is basically alpha and the area on the right hand side, which is the saffron color, is 1 minus alpha. So that, because this total sum is always 1. So you will always ensure that.

Now in the in the simple univariate normal, normal distribution, you can either simulate or use simple transformation to convert into a normal distribution. And then solve the problem. But in case, if it is a multivariate non-normal distribution, you will basically have to simulate and basically find out the distribution on an orthogonal plane for each and every of the, of the axis, which are non-normal. Try to find out their orthogonal transformation in the normal, non-normal surface. And then find it out, to solve the problem, which I am going to come to that later on.

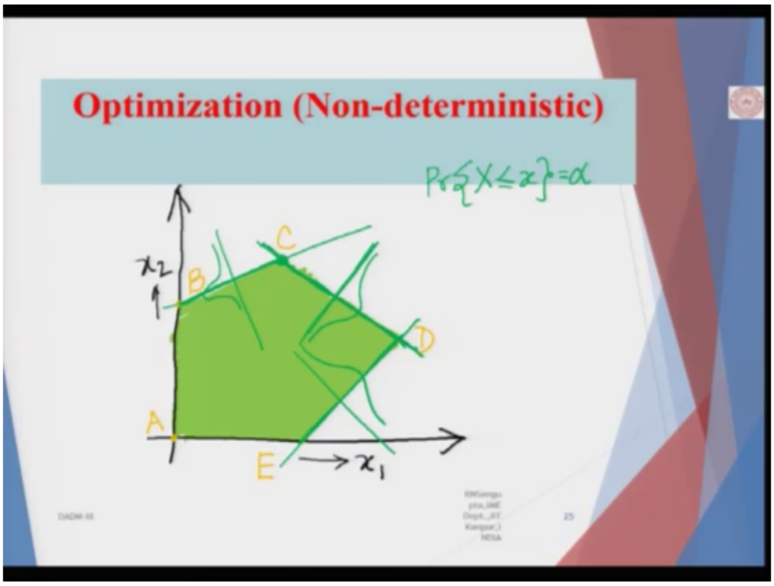
But how would it look, when you are trying to solve the deterministic part, then the normal distribution part with constraints sensitive analysis and perturbation. Then the case, where you have the non-deterministic part of the perturbations. So this would be discussed in within few moments, using simple diagram, conceptually. And then, I will basically come to the problem solution.

(Refer Slide Time: 23:12)



Now I will spend some time in this diagram. It maybe a little bit more involved and as I go step by step. But I will try to basically go into the details. So first consider. It is only two dimension problem. And you have along the Y axis one of the variable which is X2, and along the X axis you have one of the variable which is XX; along the Y axis you have X1, sorry, along the X axis you have X2. So these are the two variables only. And let me draw it before I come to this diagram, let me draw it.

(Refer Slide Time: 23:55)



So this one, 383 so I create 3 blank slide and then proceed. So I basically draw. First let me start with the (lini) optimation, which I have already done. So this is a one dimension X1. Next, actually let me plot X2. The way you draw X1 X2 does not change the concept as such. So your space, which we have been already doing- this was one constraint, the second one, the third one. So the points based on which you will do your movement was A, B, C, D and E. And your feasible region; so this was the feasible region. I will mark this with green color and then basically go into trying to basic, give this perturbation in such a way, so it is easy for all of us to understand. So this is the feasible region which we want.

Now what happens? So considered this green line, so this is the point. Consider this is the one of the constraint. So that was one of the sets, which was in GJ and HK, if you remember. Consider that is, either GJ or HK and both are deterministic. Now assume, that it is the probabilistic, due to some reason- X is variable, varying, B is varying whatever. So if I look at it from the point of view, if it is a normal distribution then the distribution, the perturbation in the univariate case, would look like this. This, it is perturbing.

So if I have a perturbation or movement for this constraint ED, it is basically normal. Or if I have a perturbation, say for example, for BC constraint; here normalities have been drawn in such a way, in order to basically to differentiate there is variance difference. Now if they, if normality is there only for one, I solve the problem using probability- X less than equal to small X is equal to alpha or greater than alpha, less than alpha. And I, I am able to proceed. But what happens if they are combined?

So in the case, they are combined. The, there are two scenarios. I will come back to this diagram once again. There are two scenarios. Scenario 1, where the variances are same for the two dimension one. And scenario 2, the variance are different. And another concept would be if they are not orthogonal to each other, if they basically merge. So I will basically try to draw it using the diagrams, explain it and then proceed to solve the problems later on using the concept of very simple sensitive analysis for the deterministic case~~. A~~and then go into the probabilistic case later on. One week, once we come to the robust and ~~innovative~~ reliability problem. With this I will close this class. And have a nice day and thank you very much.