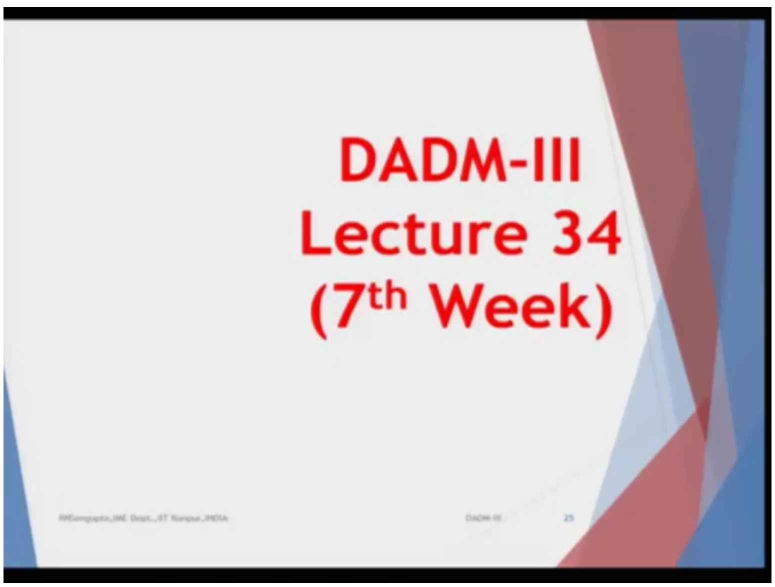
Data Analysis and Decision Making – III
Professor Raghu Nandan Sengupta
Department of Industrial & Management Engineering
Indian Institute of Technology Kanpur
Lecture – 34

Welcome back my dear students. A very good morning. Good afternoon. Good evening to all of you wherever you are in this part of the globe. And as you know this is the D.A.D.M. III which is the Data Analysis and decision making III lecture series under NPTEL MOOC and this total course as it is a third part of the DADM series is for 12 weeks. Contact hours is 30 which when split over the number of lecture is 16 number and each lecture is for half an hour or so. And in a week we have 5 lectures, so after each week we have one examinations, one assignment and after the whole course is over there is one final examination.

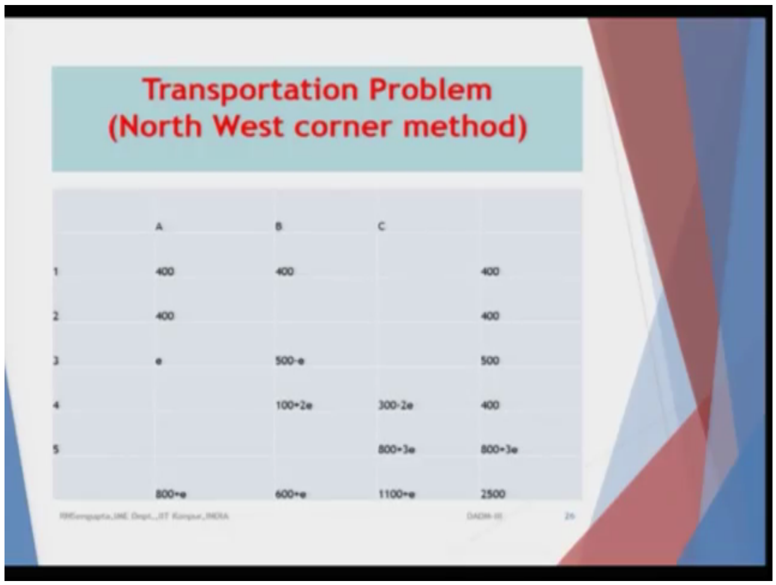
(Refer Slide Time: 1:04)



And as you can see from the slide we are in the 34 lecture. That means we are in the 7th week and with the end of the 34 and then 35th you would basically have the 7th assignment in totality. As you know there are 12 weeks so you will have 12 assignments.

So we were discussing about the concept of this epsilon amount being added for the case or the transportation problem concerning that degeneracy is not to be there and once this degeneracy is removed obviously you will move back to the concept of trying to find out the optimum feasible solution by moving along the corner points what do you do in the linear programming is the same thing like shifting delta amounts you will basically create trying to move from one on the corner points, feasible corner points to the other feasible corner basic concerning that we are starting our basic feasible solution. Till we lead optimum Point which is in this case is the minimization problem.

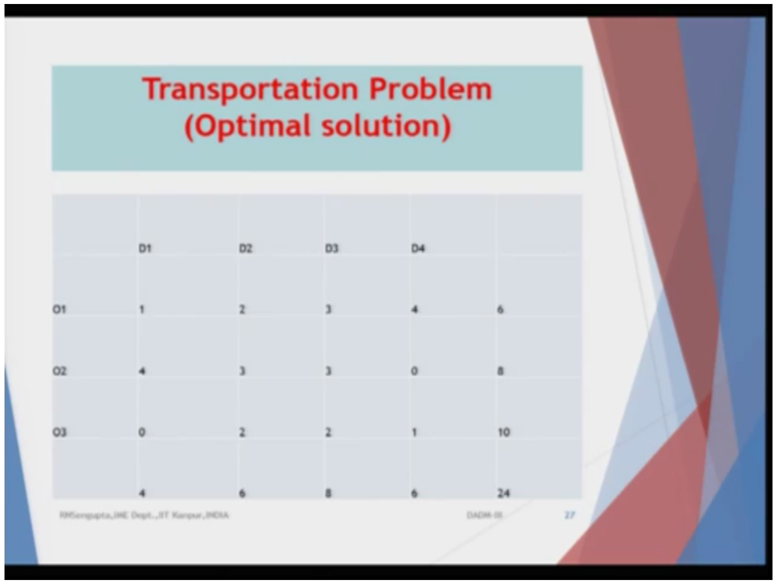
(Refer Slide Time: 2:08)



So here you can see consider in the north west corner method that the degeneracy are if they are there you will basically add and subtract delta amount in each cells accordingly such that you reach the concept of optimality. Now for the transportation problem we will come back to the transportation problem later on considering the sensitive analysis if you remember the sensitive analysis which we considered for the portabitions of the constraints.

Here I will I will discuss later on in this 34th and 35th class some concepts about the DA method and why DA method? Because even though it looks non-linear we will consider linear simple solution methods like as we did in the linear programming for the DA method. This DA method was considered in some detail in DA DA M2, but I will go into the details because if you remember I do remember quite vividly that I did mentioned that the DA method which is data envelope analysis method would be considered in the DADM 3 in details concerning the solution methodologies we have already discussed which is the linear programming.

(Refer Slide Time: 03:24)



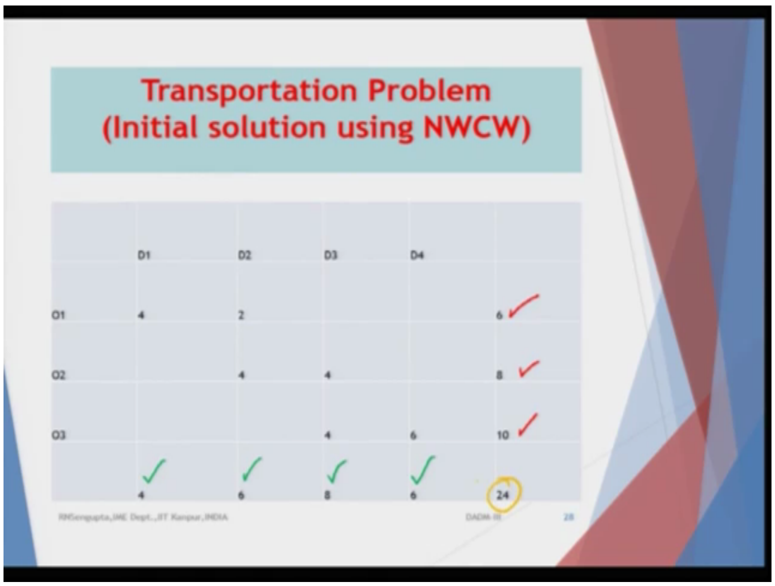
So in the optimal solution problem for the, so I will come to that DADM method later on but in optimal problem solution considering that you have the destinations given as D1 to D4 origins as given as O1 to O3 and you and your total cost structure is given as shown in the matrix so the cost are C11 is 1, C12 is 2, C13 is 3, C14 is 4 in units and the corresponding values for transporting from O2 or getting products as D1, D2, D3, D4 which is basically the values given in the second cell would be second row would be 4, 3, 3, 0.

So remember this 0 cost basically means that whatever transportation you do between O2 and D4 the total cost component per unit is 0, so obviously the total, the total amount which we will transfer we will have a 0 cost. But obviously that it would not be 0 always in the sense the total cost will not be 0 always for the whole structure because you cannot transport more than a requirement or you cannot transport more than what is being supplied from either the origins of the destinations.

And similarly for the third row the values of 0 2 2 1 which would basically again mean the 0 cost is the cost concept which is there between O3 and D1 and the other values of the cost functions are 2 to 1 and if I look at the marginal values the margin values which are basically the corresponding total values of Ai's and Aj's or Bi's depending on how you have been able to nomenclature the values of A's and B's.

So, if you are considering the number of origins where you are transporting the warehouses from you are transporting or the factories from you are producing and then transporting them would be basically given by the Nomenclature I is equal to 1 to M and if you are considering the retailers or the end customer who are there they will be n number and they will J is equal to 1 to n. The values are on the right most columns is 6 8 10 and the total is 24. Similarly, if I consider the overall values which is given for D1 D2 D3 D4 they are 4 6 8 6 so the total is also 24.

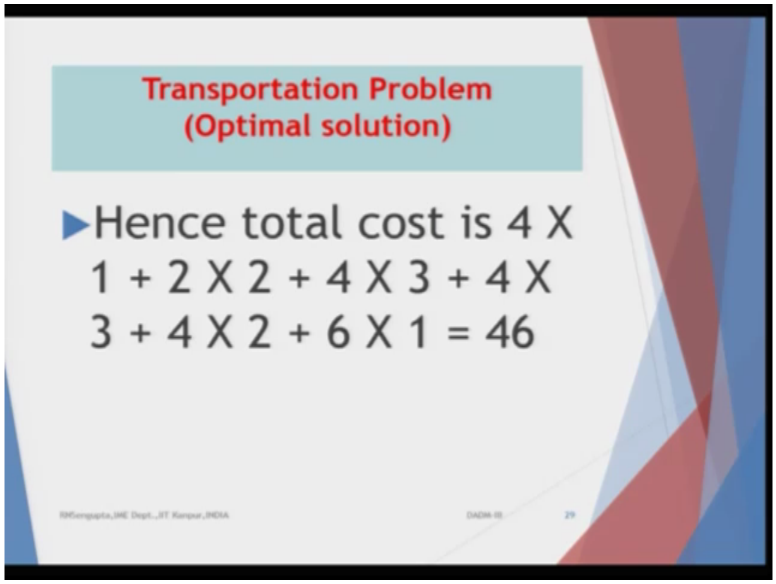
(Refer Slide Time: 05:50)



Now if I consider the simple north west corner method solution, so the values which comes out to be are 4 2 from O1 to D1 and O1 to D2, so the total cost on the right hand side, no total transportation from the right hand side is 6. If I consider O2 so I transport as per the north west corner method I will transport to D24 and D34 so the total is 8. I am basically counting the rows, the values, if I transport products from O3 and D3 if I see the values is 4 and 6. Again the total value is 10, so the right hand side balances.

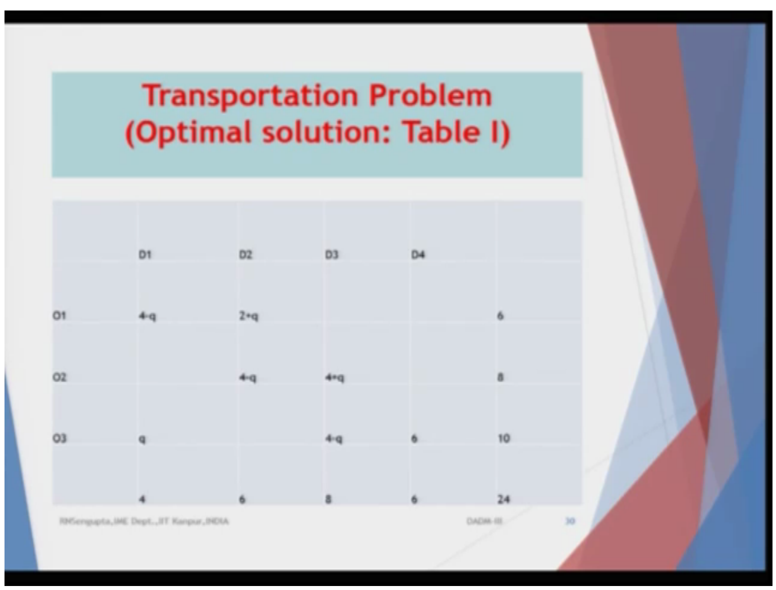
If I consider the column wise the values are 4 along the column so it is 4 here. So I will use a different color concept it is 2 and 4 from O1 to D2 and O2 to D2 is 6. If I consider the next quantum which is O2 to D3 and O3 to D3 it is 8 and the values if I transport from O3 to D4 it is 6. So again the total value comes out to me 24.

(Refer Slide Time: 7:10)



So, hence the total cost would be corresponding to the fact by the total quantum in into the per unit costs, so it would be 4 into 1 plus 2 into 2 plus 4 into 3, so 4 into 1 is 4, 2 into 2 is 4, 4 into 3 is 12, 4 into 3 is 12, 4 into 2 is 8, 6 into 1 is 6 and the total cost is 46.

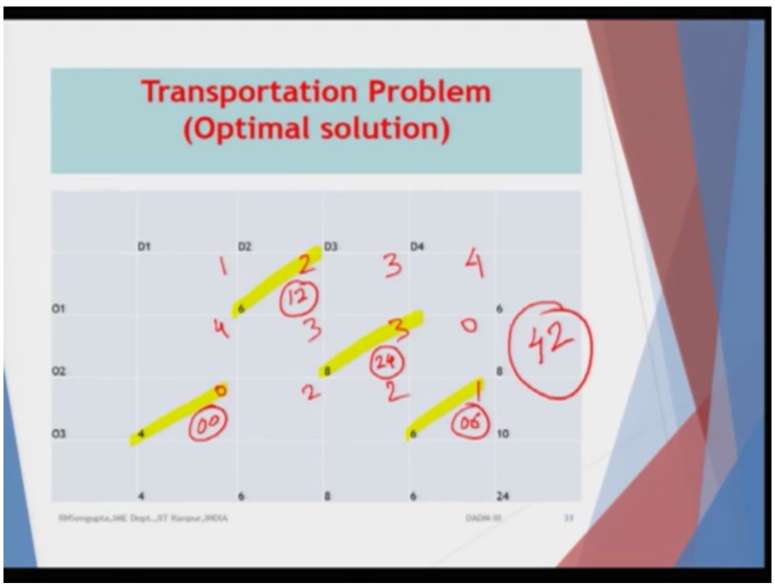
(Refer Slide Time: 7:32)

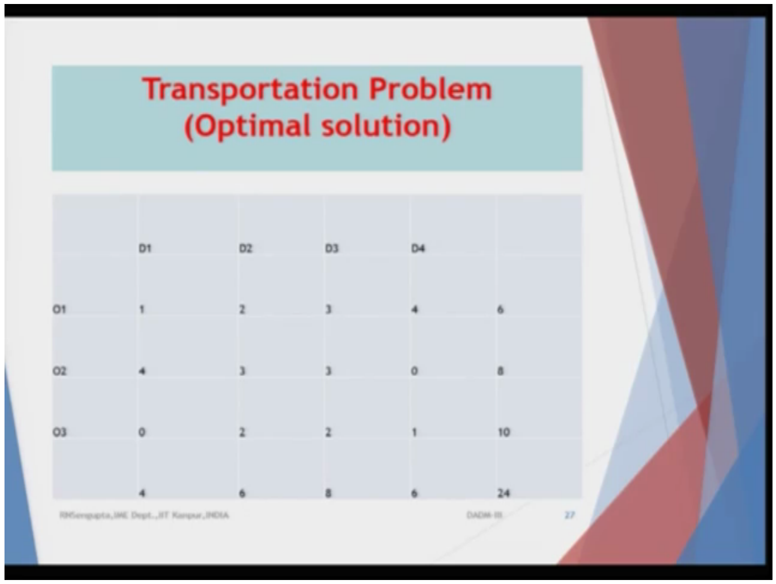


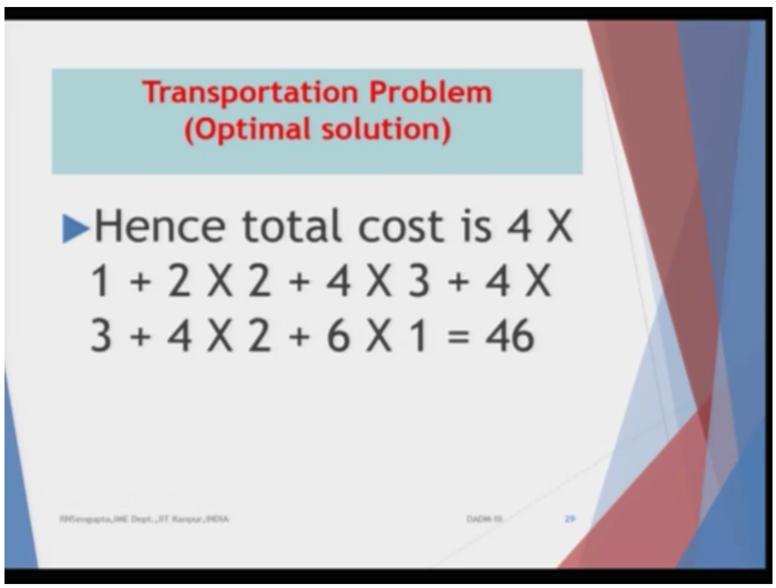
Now, if I consider the values optimum to a solution values if I perturbed the perturbed values will be such that the total cost component in the rows or the columns remains as it is given as 6 8 10 along the right most column and the values being 4 6 8 6 in the bottom most rows. So, perturbations would basically mean that the values of q which we are going to consider we have to find at each step but the values would not be violating the constraints which are there. That means the summation of each row is equal to B values which are there on the right hand side and summation of each values along the columns would be equal to A value.

So A and B are just the nomenclature which I am giving and the total sum of A's and is equal to, should be exactly equal to the total sum of B's. So the total quantum of goods which is being marketed or shipped from all the factories combined should be exactly equal to the total amount or demand which is being demanded and the retailers or the warehouses. So consider I am trying to transfer from the factories to the warehouses.

(Refer Slide Time: 8:39)







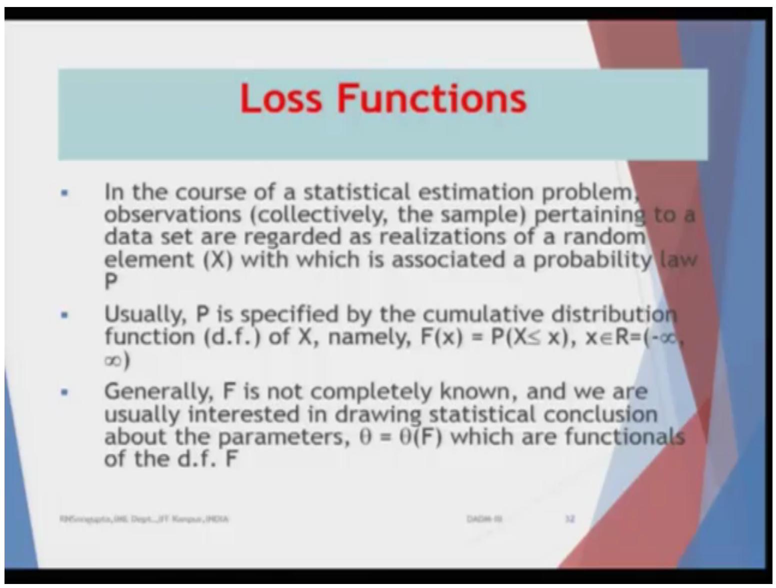
So a perturbation values as we change the total cost structure considering that the epsilon methods which we have, so if we change keep changing the values in a cyclic manner considering that the constraints are met along the rows and along the columns, then the total cost structure now would become, if I consider is 6 between O1 and (D1) D2. So I will write the cost it will be easier.

1 2 3 4 I will write it 1 2 3 4, 4 3 3 0, 1 2 3 4, 1 2 I should use a different color so it will be easier. 1 2 3 4. These are the unit costs 4 3 3 0 and the last row values are 0 2 2 1. So now let us see what are the costs and highlight what are the values are. So I use the highlighting, so 6 to za is

12. Then 8 3 za 24, 12 and 24. So I will keep writing it also. 12 will be the cost and 4 is the cost, 6 is the cost, 0 is the cost.

So I highlight the values so the total cost is 12 plus 24 is 36 that is 6 plus, 6 is 42 and obviously this is 0. So, if I see the values here sorry they were 46, so I have been able to transfer all the amount of goods between origin and destination. Total cost comes out to be if I add them up, so which is 4 unit less. And it can be improved accordingly depending on the problems.

(Refer Slide Time: 11:21)



Now before I come to the data analysis which I did discussed in DADM 2. And what was the significance which I mentioned about the optimization from the simple linear programming concept. First I will discuss something about the loss functions. So loss functions are generally very heavily used and quite important when you are considering the concept of optimization or also in the concept of estimation, we just basically from the point of view of statistics and probability.

Now, what do we mean by loss? Loss means that you have something to predict, something to find out. And if you are not able to predict to the exact value, you have some differences. If there are differences obviously you will (())(12:03) some costs. If they are cost you will want to basically find out what is the overall accumulated cost which you will have based on which you will try to basically minimize your total optimum value whether it is about maximization or minimization, we are, we will be considering in the minimization problems only and it can be done for the maximization problem also, but we will consider the negative side accordingly.

So whenever you are doing any statistical and estimation problem and observations are there you collect the observations which you call a sample. And why do you need a sample? Because there is a population, population has some parameter values which we know in a distribution the parameter values are scale shape and location which we will generally denote by Alpha Beta and Gamma.

These are the symbols and we need to basically find on the scale shape and location parameter because based on that we will be able to find out what is the first moment, what is the second moment and so on and so forth. Because in generally for the case on the normal distribution you have mu and sigma square. In the case of exponential distribution we will have only one parameter if it is basically a exponential distribution which is shifted onto the right, we will basically have two parameter exponential distribution. Similarly, we can have a beta distribution of 3 parameters and so on and so.

Now whenever you find out the sample as I was saying and we our main interest to basically study some characteristics of the population and if the population is not known in all the parameter values we have to take a sample in order to basically find out some characteristics of the population by using the estimates in the sample. And there are different type of studies which can be done.

But our main concern would be when you are taking a sample or the set of observations on the sample. Those should be chosen in such a way and be combined. The combination should be done in such a way to calculate the best estimate of the population parameter such that the error is minimized. So what is the error that would basically come out in different situations.

So whenever you are collecting a sample pertaining to data set you will basically consider that there is the realization of those values and I have already discussed that in DADM 1, so there is a realization of those values, the realization of the values means that there is a random variable when you basically need it and we are trying to understand that the actual value based on which you will try to predict or forecast would be the realized values.

Say for example there is a dice and if I roll it. So before I roll it the X is the random variable based on the fact that what face will come out when you are rolling the dice. So, if it is say for example x and when what it can be either 1 or 2 or 3 or 4 or 5 or 6 and when the actual value is seen by us considering this number 5 we will say the realized value is 5.

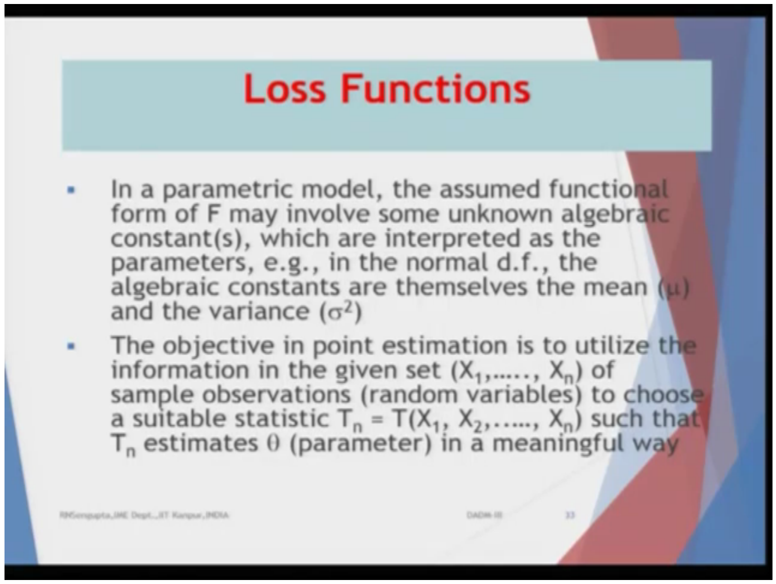
So as I was reading the first point, so you collect a sample pertaining to a data set and they would be regarded as the realization of the random value x with which we will basically have a property law or property distribution on the PDF or PMF as the case may be depending on PDF would be the case when you have basically a continuous distribution, PMF would be the case when you have a discrete distribution.

Usually P the probability law which we want to know is specified with this cumulative distribution as I said which is the cumulative distribution function and it will be basically given by capital F of X and obviously all the properties which are all the 4 properties which are already considered in statistics should hold true when we are commenting about the probability distribution function or the cumulative distribution function capital F of X.

And generally we will consider the x value which is the random variable in the continuous case to be lying between minus infinity to plus infinity. Generally here F is not completely known because if it was known then Alpha, Beta, Gamma which I said as the scale shape and location parameter then obviously it does not suit us to basically collect some set of observation in order to conduct the study where our main concern is to find out this any of this 3 parameter or the combination this parameter which is alpha, beta, gamma.

So generally F is not completely known and we are generally interested in drawing statistical conclusions based on some studies and based on some particular test you have to basically draw some statistical conclusions or more the parameters theta which is basically a function, functional form of capital F and based on the value of Theta we will be able to predict or find all the functional form of capital F as better as possible or as all close as possible to reality.

(Refer Slide Time: 16:46)



When we are doing a paramedic model estimation the assumed functional form of F may involve some on unknown algebraic constants and this the constant we just mentioned is alpha ,beta, gamma for our case. They would be interpreted that the parameters which I told and in generally the normal distribution the algebraic constants or the pattern the parameters are basically given by mu which is the mean and sigma square is basically the square of the standard deviation which is the variance.

The objective in point estimation so generally when we do any inference a test or inference testing on inference study we basically have 3 steps, one is the point estimation, one is the interval estimation and based on the point estimation after the results have been verified using some prior methodologies which are very simple to understand at least for our cause what we are going to deal with, once these are done we basically go into the concept of hypothesis testing.

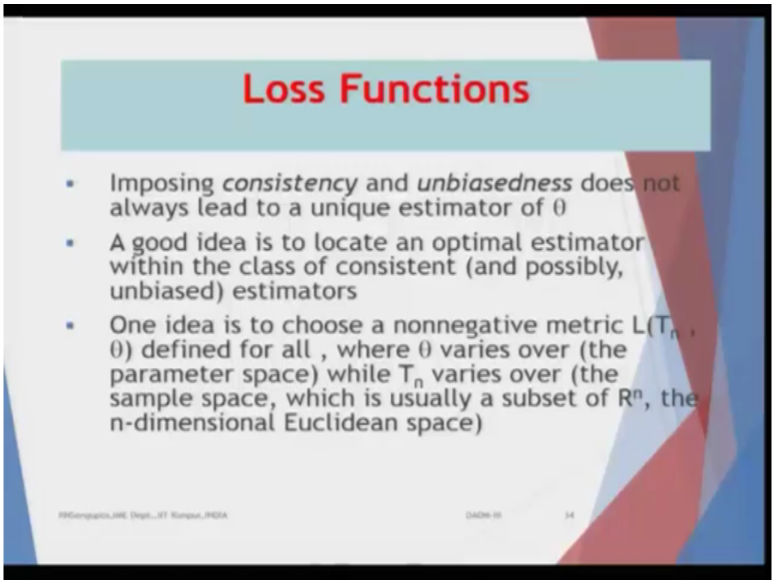
The objective in point estimation is to utilize the information in the given set which is the set is basically I pick up a set of observation which is x1, x2, x3, till xn. These are the set of random observations and we choose a particular statistic. Now statistics is basically the actual functional form based on the sample which we find out that means X1 to Xn what we, whatever set of observation which we have which we call a sample and we combine them in certain form and some functional form we find out that particular statistic that statistics is used as a best estimate for the population parameter for which you want to do some studies. So this is basically the statistic.

So sample observations or random variables are chosen to find out the suitable statistics. So this statistic is a function as I mentioned of X1 to Xn and this n symbol is utilized to denote the number of observations you are going to take. Such that these statistics Tn estimates theta the parameter value in a very meaningful way. And our main concern would be basically to be as close as possible to the estimated, estimate which is Theta such that to minimize the loss or minimize the prediction or minimize the forecasting error whichever we are trying to do.

In general we have seen that in statistics we basically imposed two rules. One is the consistency and unbiasedness and consistency means that when we are trying to basically find out the characteristics of that particular the estimate based on a very large sample size it should be consistent in the sense that in the long run this the overall variance should be as close as possible to the actual one in the sense, the actual differences in the variance from the sample should be as close as possible to 0 such that we are able to predict the overall estimate value or the parameter value as well as accurately as possible that is point 1.

And point 2 is that whenever we are doing the unbiased test our main concern is to basically choose that particular statistic in such a way that in the long run as we keep repeating the observations considering n number set of observations are chosen each time the overall sample statistics in its estimated sense should be as close as possible to the parameter value which is theta, which means the expected value of T in should be exactly equal to theta in the long run. Obviously if there are some differences there would be some biasness which for our case would be only true in the case of in advance statistical studies but for this course as it is DADM 3 which is more to do with operation research we would not be going to that details.

(Refer Slide Time: 20:21)

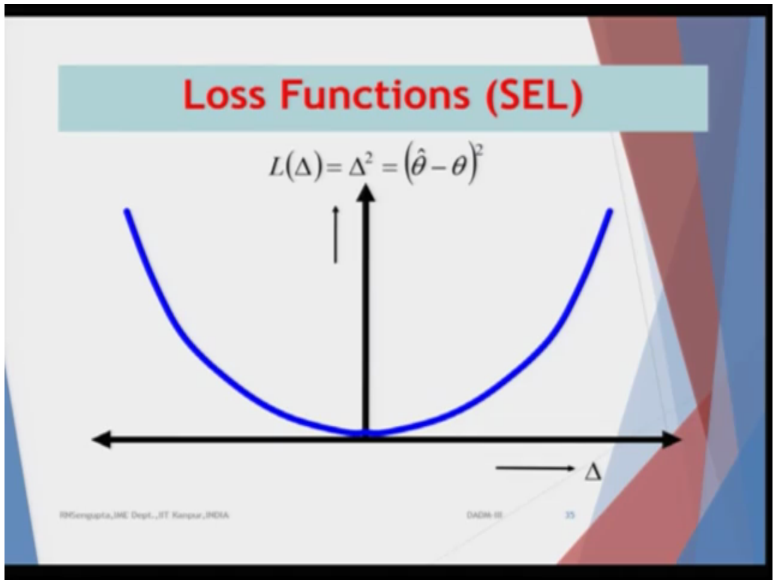


A good idea to choose an optimum missed optimal estimator would within the class of consistent and possible unbiased estimator, estimators would be very important for us because once we find, we are able to find an optimal estimator our job is done and we can utilize that to find out or predict the actual parameter value. In the case we are able to pick up only one set of observation with the sample and we utilize the sample in order to find out the statistic which will should do our work.

One idea when we are trying to find out statistics based on the sample set of observation X1 to Xn would be to find a non-negative metric or measure which would be given by L. The lost function L functional form and it will be divided for all the overall domain space of Tn. Tn is basically the estimate which you are going to find out for theta. So we will try to find on the non-negative metric L which will be defined for all values of Tn where theta values over the parameter space depending on whatever the actual realized values of or actual plausible values of while Tn would vary over the sample space.

So basically have the overall space where you are trying to basically find out that in what range theta varies and as you pick up samples there would be a sample space in which Tn would basically vary. So you will find out that we will try to find out the Tn varies over the sample space which is usually defined as a subset of Rn. Rn is basically the dimension of the sample which you are going to pick up, so n number of observation is very important for us as you increase n obviously the statistic would be in the long run, actually it should be as close as possible to the parameter value. So Tn varies over the sample space which is usually a subset of n and in the case of the Euclidean space it would be of dimension n.

(Refer Slide Time: 22:26)



So let us see this this in graph. So along the x axis we are trying to measure the difference between the statistic and the parameter value. So if the parameter value is basically theta, if the statistic is theta hat then the difference of Theta hat and theta would be denoted by Delta and Delta would we measured along the x axis where any value when I am basically trying to now highlight using this red highlighter any value onto the right would be positive in the sense Theta hat minus Theta positive that means you are overestimating Theta using the value of theta hat.

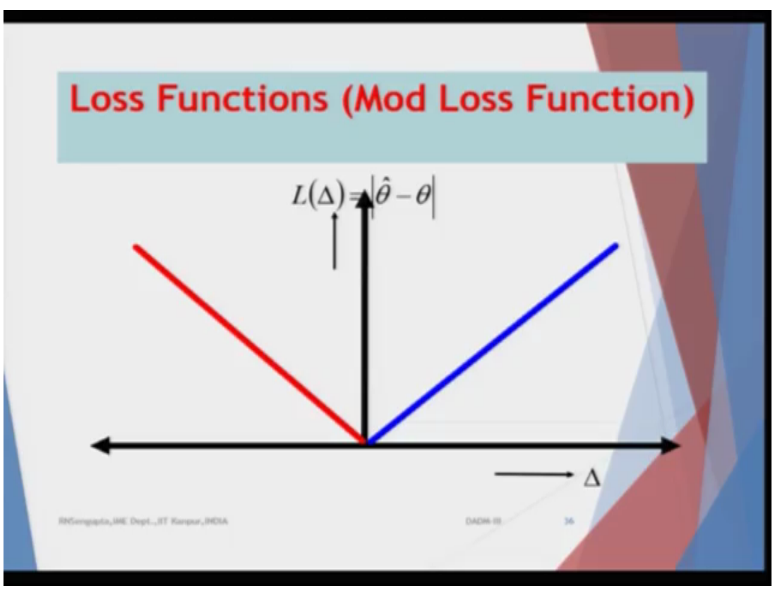
If I go on the left hand side it will be theta hat minus theta would be negative hence we are trying to basically to find out the estimate of Theta using an underestimated value which is theta hat. Now the beauty of the graph is that consider that the fact that if the lost function on the overall positive or negative effect which you trying to find out concerning that you overestimated theta using theta hat or underestimated theta using theta hat the overall shape of the graph would be a quadratic in nature.

Now, why it is quadratic in nature and why we assume it quadratic in nature is that in the long run we will try to basically minimize the loss as far as possible. So trying to basically minimize the quadratic equation would be a kin to the concept that we are trying to basically minimize the variance because what is variance, variance is basically the expected value of as you try to find out and in the case that if there are n observations it would be 1 by n and it would be a summation of Xi minus X Bar, X n bar whole squared provided Mu is not known and we basically utilize X n, X n bar as the best estimate for MU. In the sense that we have already discussed that how you place the best estimate from the sample as the as proxy for the theta value which we do not know.

So if we basically draw the overall lost function as a quadratic one it means that the penalties for both positive and negative values of Delta whether theta hat is greater than theta or theta hat is less than theta it would equally penalized. So if I move two units on to the right or two units on the left and if I consider the quadratic equation is basically theta hat minus theta whole square in that case we will see the overall lost which we will encounter provided you have an overestimation problem or an underestimation problem would always be the same. That is what we of Unit 4 provided that you are not multiplying any constant value with theta minus theta hat whole square.

And as I mentioned that few minutes back trying to minimize the overall estimated value or the expected value of the lost function will be exactly equal to the concept that you are trying to basically minimize the variance.

(Refer Slide Time: 25:18)



In case the lost functions you are using the modulus for lost function or mod loss function in this case also the overestimation and underestimation would be equally penalized, but the interesting fact it would mean that overestimation and underestimation would not be equally penalized considering the overall lost function is quadratic in this case it would just be the mod of the difference of the estimated value on the parameter value. So if I basically find out theta minus theta and I am trying to plot it along the x axis where this positive onto the right of origin which is this is the origin which I consider and it will be, it will be positive on the right of the origin if I consider on the left to the origin it will be considered negative.

So in that case here also the overestimation and underestimation for theta hat minus theta whether it is positive or negative would be of equal quantum, but the interesting fact is that as you as you are in and around the origin there you will find out that concept limb and there differentiability would be a problem because the rate of change of this function at the point 0 happens in such a way that it will be almost difficult to find out that what is the Dy Dx of this curve in case it is needed to solve our problem in more details. So in this case you will basically have mod loss function the mod lost function would be given by the value of mod theta hat minus theta. It looks interesting but trying to basically solve it would be much more tedious then the case when you have basically the quadratic loss function which we just discussed.

In case if the linear loss function, now this would also be true for the case or the quadratic loss function I have not drawn it but I will try to draw it using the graphs which you have to just consider, I will come to that. Considering the loss function is asymmetric linear loss function but is of the modular type in the case that will basically add or multiply some parameter values K1 and K2 for the case when theta hat is basically greater than theta and in the case if theta hat minus theta then basically multiplied by a factor of K2.

So K1 will be multiplied if it is overestimated, (K1 will be multiplied) K2 will we multiplied if it is underestimated. So if the K1 and K2 values are higher and lower then overestimation will be more penalized on the right and underestimation will be less penalized on the left hand side. In case if I just reversed the values that means K2 value is much higher than K1, then the part of the of the slope of the red line which is basically the underestimation should be much more higher on to the left and the the slope of the blue line which is on the right would be much lower.

So with this I will end the 34th class and continue the discussion on the loss function and then come to the estimation problem in the DA concept where you will be trying to place the concept or linear programming in order to understand why DA is important and how it can be modified as a simple linear programming problem. Have a nice day and thank you very much.