Data Analysis and Decision Making - II
Prof. Raghu Nandan Sengupta
Department of Industrial & Management Engineering
Indian Institute of Technology, Kanpur

Lecture – 12
Loss Function

Welcome mat my dear friends, very good morning, good afternoon and good evening to all of you. This is the DADM-2 which is Data Analysis and Decision Making-2 course under the NPTEL MOOC lecture series. And as you know this course is for 30 hours which means that its spread over 12 weeks and we are in the third week as you can see from the slide and this is the 12th lecture which is the second lecture in the third week. And as you know each week we have five lectures each for half an hour and after each week we have one assignment.

So, by the time you are seeing this 13th third week 12th lecture you would definitely have completed two assignments and then after this twelve assignments they would be the final answer based on or like question paper based on all the concepts and topics which have been covered. And my name is Raghu Nandan Sengupta from IME department IIT Kanpur. So, if you remember we are discussing about loss functions; that means, get to give a very brief background or recap if we can say so.

So, we had something to estimate, estimate means find out and that is basically the parameter of the distribution or some value and when we take a sample they would definitely be an error like when you have trying to find out the probability of getting a half of an unbiased coin if you toss that coin thousand number of times, the exact number of heads need not be 0.5 100 by 1000; that means, 0.5 it may be either more or less and; obviously, as you keep doing it that is besides the one if it keep doing it that actual ratio for more and more number of tosses in the infinite case becomes exact equal to 0.5 tending towards 0.5.

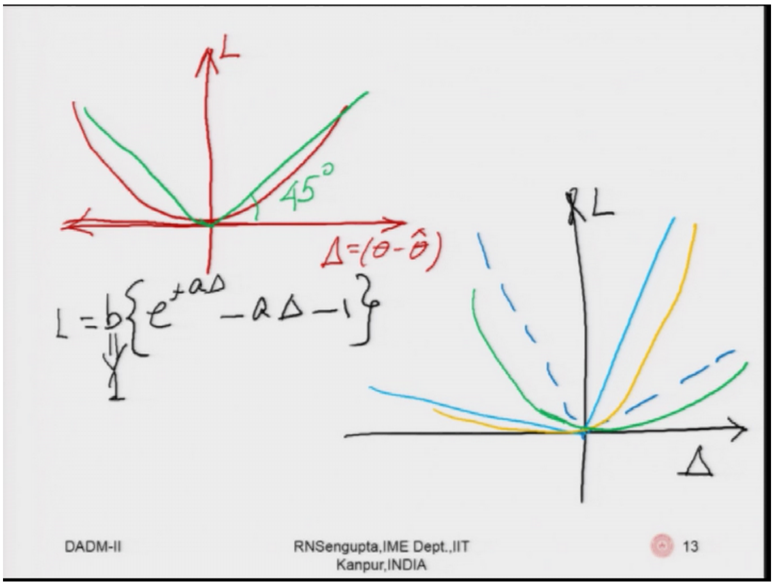
Now, whenever you are you are estimating or trying to find out. So, as I said that the parameter or the value which you want to want to find out is known gives given by the symbol theta and the estimate which you find out from the sample is given by the theta hat. The hat is basically the way we denote. I will I will try to use the slide also in trying to draw because if I remember correctly for the last two lectures which was basically in the 10th and the tenth 11th I was talking a lot of things, but technically I should have also written the same similar things on the slides we would have made it much easier I will do it definitely today.

Because that I found out later on when I was going through the lectures for editing part they were definitely these questions which came to my mind. So, any coming back to this 12th lecture in the DADM-2, so, whenever you are trying to basically estimate you will basically have the concepts of an error and this error is basically the difference between the actual value and the estimated value.

So, the errors technically we consider in order to make our life simple because the calculations are very nice very compact statistically easy to prove many things. We consider the errors to be equally penalized quadratic in nature and; obviously, I also remember if I am pretty sure about that. I did not mention that when you are considering the quadratic loss function or quadratic utility function you will consider the distribution of the returns based on the utility to be normal and vice versa. So, this is an if and only if conditions quadric loss function results in a normal distribution for the returns and normal distribution the returns I mean gives us a quadratic loss function for the decision making.

So, whenever you are considering the quadratic loss, you basically consider equally per equal penalization and this equal penalization can also be true if you have the lin lin loss function which is linear linears loss function. So, it would basically look I will draw it here or in a new fresh on what let me. So, it is easier for us. So, you have the lin lin loss function. So, when I am talking about the quadratic loss function, these slides are not a part of the discussion I will just remove it just for the explanation.

(Refer Slide Time: 04:53).



So, the for the quadratic loss this is y axis is the loss function which I will denote by L and this x axis I denote by delta, delta is basically the difference between the estimate in the error for the quadratic loss you have this. So, this should be touching this x axis. Anyway consider this is now if it is a lin lin loss function let me use another color say for example, green and they are equally penalized. So, this is 45 degrees line and this is also 45 degrees line.

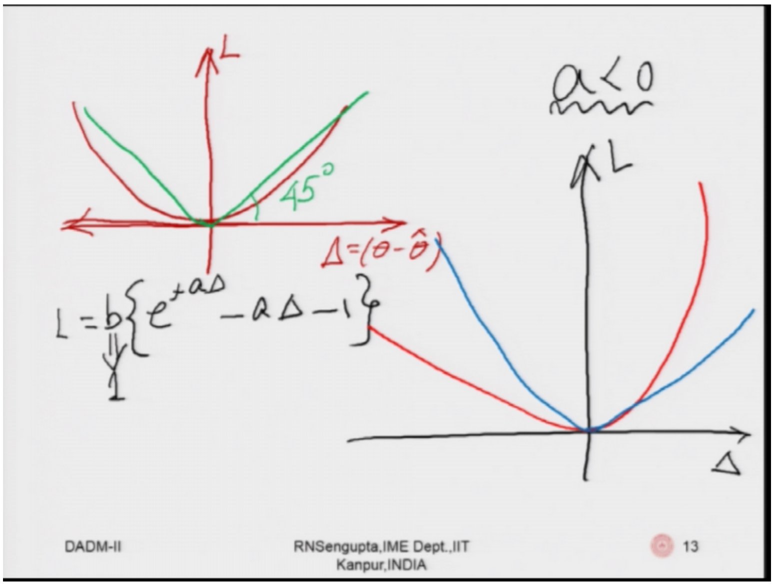
So, technically they are equally penalized and obviously, you can tilt this quadratic loss function on more on to the second quadrant on more on to the first quadrant similarly for the lin lin loss function. So, if it happens it will be like this. So, let me draw the axis. So, for the linear linear loss function maybe it is like which I have already drawn, but I am again still doing it. So, this is L the loss function this is delta.

In case you have the quadratic one tilted let me use the color cream because orange it would be like a semicircle just tilted accordingly quadratic consider this is a quadric one here you use the green something like that. So, in unequally penalized.

Now, later on I also mentioned about the linear exponential loss function on the linex loss function and to give a background it was basically proposed by Hal Varian in the 1960’s and Zellner from proved many very nice properties of this loss function from the statistical point of view. And if you consider the loss function it is given by the form which and again write it. So, the loss there is a scale parameter b which we can consider for the simple case as 1 it would not affect.

So, e to the power minus e to e to the power plus a lambda minus a lambda minus 1. So, this lambda is the difference between the parameter value in the estimated value. Now, we have also discussed I am repeating it because the examples would become very clear when you consider that and please bear with me. So, when we consider the theta value as positive. So, let me erase it.

(Refer Slide Time: 08:21)



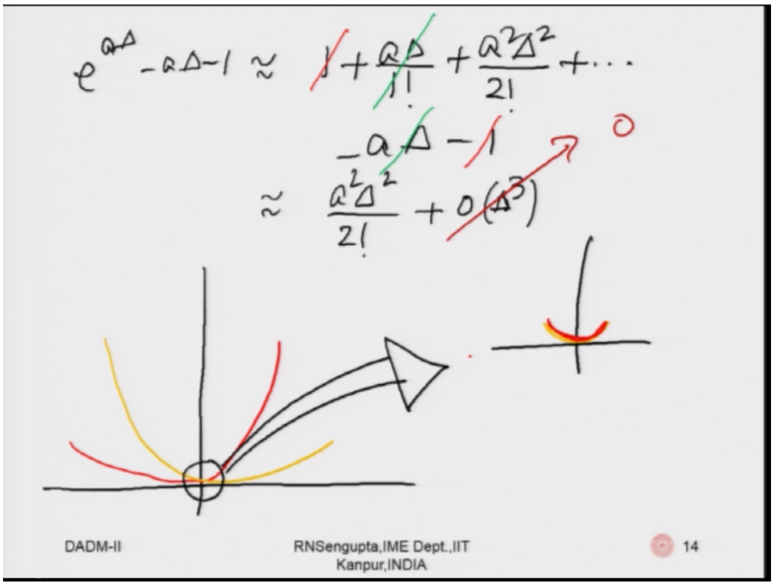
So, this is the axis and if I have and this, I am delta L and if I am considering here a as positive remember this is one case scenario.

So, if I have this I will use that color. So, as you expand it as delta increases; obviously, e to the power plus a delta will keep increasing exponentially while the negative part a minus a delta and minus 1 they would increase, but they would increase linearly. So, the exponential part would dominate the linear part. So, it will be exponential is the exponential something like that, I have not drawn it accurately let me try something like this and if you see delta as negative on the left hand side first second quadrant.

So, in that case e to the power minus value will decrease exponentially. So, it will be dominated by the linear part. So, it will be like this something like this. So, in this case over estimation is more penalized than under estimation for a is positive. Now consider a as negative I use another color. So, consider this dark blue. So, in this case when a is negative then as delta increases in the in the positive sense so, it is e to the power minus. So, this will basically decrease exponentially because 1 by e to the power that value and the negative minus of minus a delta will become positive. So, it will be linear, but linear would slowly dominate as delta increases. So, it will be linear here and if you consider the second quadrant. So, in that case you will basically have minus of minus a deltas becomes positive. So, then in the left hand side which is the second quadrant the exponential term will dominate the linear part.

So, you will basically have underestimation being more penalized than overestimation. Another thing which I did not mention again without writing that is why I thought I will spend some time a little bit slow I know that I am going a little bit slow, but still. So, if you basically have very small values of delta.

(Refer Slide Time: 11:15)

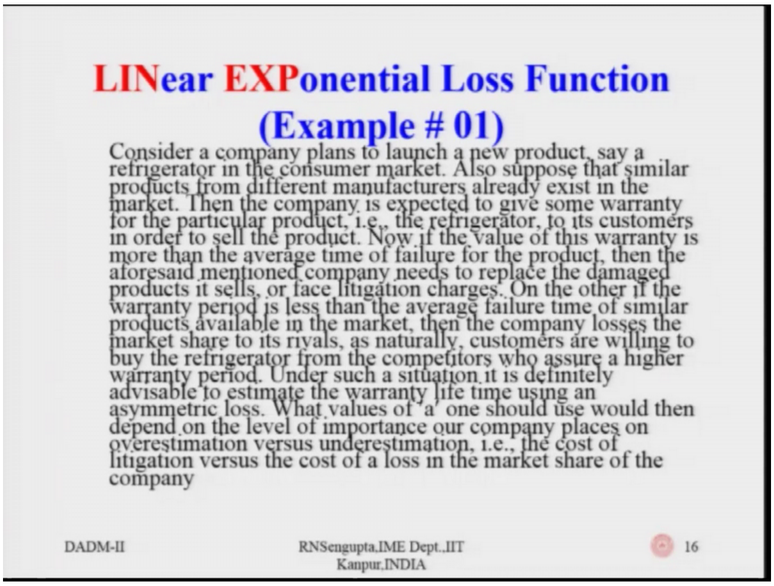


So, what you do is you expand it. So, it is e to the power a delta minus a delta minus 1 expand it 1 plus a delta 1 factorial plus a square delta square by 2 factorial plus increases minus a delta minus 1.

So, in these case if you have this 1 1 cancels, this value of a delta by factorial 1 a delta cancels. So, in the long run the value if you ignore the higher terms third order and higher. So, this will become a square delta square by 2 plus of order this order value I am saying is basically cube in the sense as the increase cube higher power they can be considered as almost tending to 0. So, if you consider a square delta square by factorial 2 you see this is basically quadratic loss function or the squared error loss function which means they are the linex loss function for very small values of the delta value.

So, which means that if you have and the both positive and negative values of a. So, if you zoom in; that means, you are zooming in at this portion. So, if we zoom in. So, you basically have the values of they would be quadratic almost exactly very small values. So, this is the concept which I wanted to share.

(Refer Slide Time: 13:30)



So, let us consider few examples. So, consider a company plans to launch in which product. So, consider is in a refrigerator or an AC or say for example, a TV, fridge whatever it is and for each product which you are selling you as the main owner of the company or the CEO of the company you want to give some warranty life. Consider the warranty life is 6 months which means that way the product fails before the 6 months your company would replace that product or change the parts as it is in the warranty and if it fails after the 6 months; obviously, the person would be paying the actual price until unless he or she has some yearly contract for the product which he or she has bought from you from your company.

Now, consider that there are competitors in the market and the 6 months which you are given may not be accurate. So, the consider the values which predicted in by that competitor actually it is 4 months and you basically predict 6 months. So, in this scenario what happens? If you give 6 months competitor gives 4 months the initially people will be more interested to buy your product because you are giving a higher warranty life extra 2 months.

So, you basically increase yourselves increase the market share, but one of the flip side negative side what happens? Actually the product will start failing those type of products will start failing in an around 4 months. So, if they start failing on an average around 4 months so; obviously, you have to replace this products they would be a business loss, they would be a loss of the product, people would definitely not be talking in positive terms of other products. So, you basically face maybe you may face a litigation loss or whatever it is.

So, initially the market side which you have one or which you had garnered us or gathered would slowly be frittered away considering that you lose the trust of the customers. Consider the second scenario; you have predicted 6 months actually guarantee time warranty time for that product is 8 months.

Now, in both sorry in both the cases case 1when it was 6 and 4 and in this second case when it is 6 and 8 if you consider the difference of the of the actual value and the predicted value is basically in the first case it would be would be 4 minus 6 whole square is 4, 4 minus 6 sorry it is 2 whole square is 4 and in the second case it is 8 minus 6 plus 2 whole square is 4. So, in these case if you use the quadratic loss function it will be equally penalizing both the cases plus 4 plus 4.

Now, continuing to the second case, if it is 8 and you are 6. So, initially you lose the market people would not be willing to buy a product, but it may be possible that they that they see that the product actual life is 6 months people would be slowly tempted to buy your product because you they will think the competitor basically not giving the good deal. So, initial low sales would be compensated by higher sales later on.

So, which one is good for you depending on how you want to formulate the problem would definitely dictate the value of a based on which you are going to formulate the idea of the loss function. So, with this let me read it. Consider the company plans to launch a new product say a refrigerator in the consumer market also suppose that similar products from different manufacturers already exist in the market. Then the company is expected to give some warranty for the particular product that is the refrigerator to its customers in order to sell the product.

Now, if the value of this warranty is more than the average time of failure for the product, then the aforesaid mentioned company needs to replace the damaged product it sells or face litigation charges by the customers. On the other hand, if the warranty period is less than the average failure time of similar products available in the market, then the company loses the market share to its rival as naturally customers are willing to buy the refrigerator from the competitors who assure a higher warranty period or higher warranty life.

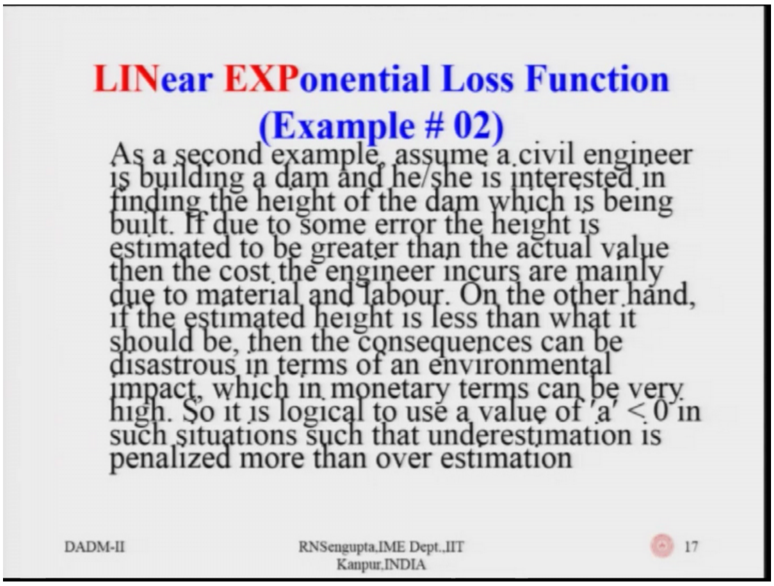
Under such a situation it is definitely advisable to estimate the warranty time life time using an asymmetric loss function because both over estimation and under distribution as I said 6 and 4 and another case 6 and 8 would have different consequence on the financial front considering both the loss of your business customers, conference being lost your market share being lost or gained can it may be the other way around. So, they could definitely be or unequal proportions.

So, what values of a one should use would then depend on the level of importance our company or your company wants to place on the overestimation versus underestimation case that is the cost of litigation or the negative loss which the company faces versus the cost of loss in the market share of the company. So, one is litigation in case if you give a warranty line more than what I have what is actually is.

So, people have bought your product initially, but now this see the at the product is failing so; obviously, they would file cases and another case would be loss of market share because you have given a warranty life less so obviously, people would be more interested to buy the product from the competitors because they have given a warranty life which is higher.

So, in this case a has to be decided accordingly.

(Refer Slide Time: 19:13)



Consider the second example; I will read it for first let me give the background. So, consider you are a civil engineer or an engineer or an executive project manager and you think that you want to build a dam and consider the dam height is 120 feet. So, consider 120 feet is actually what it should be and consider there are two scenarios. Scenario 1 you build the height to 122 feet and in another case you basically build the height to 118 feet.

Now in the first case when you wheel it to the height of 122 feet, then the difference is 122 minus 120 which is plus 2 whole square of that is 4. In the second case when you build it to 118 feet, 118 minus 120 is minus 2 minus 2 whole square is 4. So, in the case when you are doing an symmetric loss which quadratic loss; obviously, it would be 4 4 in both the cases as I just mentioned.

But now let us consider the cases separately, in the first case if you build the height of 122 feet. So, initial cost which you had to incur which was the extra cost would be the man hours lost, material cost, then the salary of these people and so on and so forth. So, basically you would have spend extra amount of money to build that extra 2 feet height of the dam. Now consider what are the consequence of that. So, later on say for example, the flood comes. So, on an average you will it is definitely true what I am going to say that the height of the dam is 120 is the minimum is on an average which should basically be able to raise the water pressure.

So, any height of 122 feet which is more than 120 feet; the overall loss to the environment and the catastrophic loss would definitely be not there and it will be minimized. So, in case even if there is a water height for the dam exceeding 120, but less than 1 122 we are able to sustain that catastrophic environment loss.

So, initial cost was high, but we are able to save catastrophic huge amount of loss. In social in social framework villages being washed, environment, the forest trees and the farmland being washed floods happening in a nearby town so and so forth.

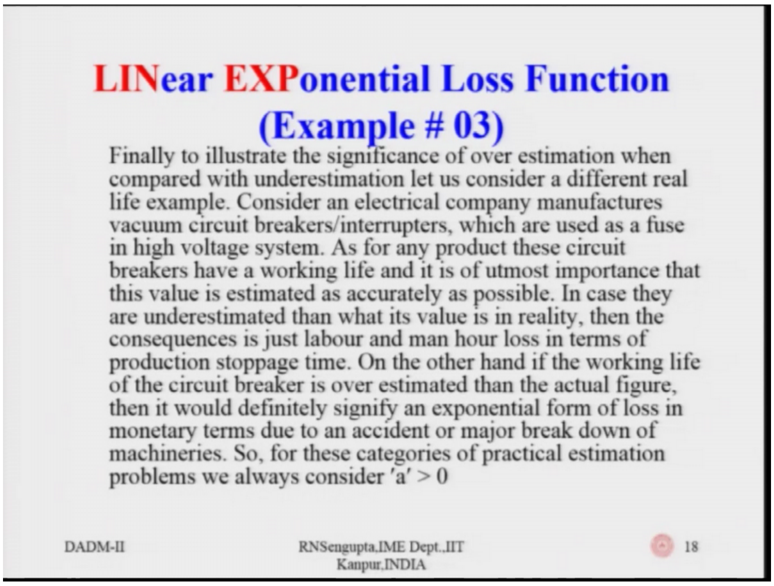
Now, consider the picture when it was it is 118 feet. So, initial cost for building the dam is much lower because you have not spend that man hours, the material cost, the salary and all this thing for that extras 2 feet extra means 118 to 120, but consider the what happens later on. The flood comes and the propensity of the flood or the probability of the flood to breach the dam for a height which is less than 118 would be much much higher so; obviously, they would be catastrophic loss in the environment such that the initial cost which you are able to save some profit you are able to make is absolute washed away by the huge amount of flood, where it will inundate the the plains or the overall area in around the dam.

So, let me read the problem before that so; obviously, it would mean that over estimation would be less penalized than under estimation in this case. So, we will take a value of a as negative. So, let me read the problem. As a second example, assume a civil engine is building a dam and he or she is interested in finding the height of the dam which is being built.

If due to some error the height is estimated is to be greater than the actual value then the cost the engineer incurs are mainly due to material and labor as I mentioned; that means, the extra amount of cost on the other hand if the estimated height is less than what it should be then the consequence can be disastrous in terms on an environmental impact which is monetary whose monetary value can be very high.

So, it is logical that in this example we will consider value of a is equal to negative less than 0 in such situations such that the underestimation is more penalized then over estimation.

(Refer Slide Time: 23:12)



Let us consider the third example, consider you are your big electrical some sophisticated equipment manufacturer that and in this machine which basically is used by the production line of a big company, that machine which you are going to supply has vacuum circuit breakers. So, they are trip switches or something sort of fuse.

Now, the average life or the warranty life of the and where when they need to be replaced because considering the fluctuation the voltages, you think the actual life of thus those vacuum circuit breakers should be 6 months. So that means, in an around 6 months you will the user or to whom you have supplied the machine will stop the machine, change this vacuum circuit breakers, do a thorough maintenance test and analysis and then only start the production.

Now, consider as I said that the warranty life is 6 months. Now consider two cases; in one case you have given a warranty life actually it should be 6, but you have given a warranty life 4 months and in another case you have given a warrant life of 8 months. So, in both the cases 4 minus 6 is minus 2 whole square of that is 4, 8 minus 6 is plus 2 whole square of this is 4. So, if you consider the quadratic loss function on the squared error loss function in both the cases this is equally penalized, but let us see the picture in practical sense.

In the case when you are basically under estimating it; that means, 6 is 4 then; obviously, the person to whom you have supplied the machine will stop the manufacturing process in an around 4 months so; obviously, maintains would be done. So, the initial loss which he or she will face would basically be the stoppage of work and man hours and production loss, but; obviously, as the machine needs as the strip switches or the vacuum circuit predators needs to be changed in an around 6 months if he or she is doing it at the fourth month depending on the information which has been supplied to him or her then the probability of any vacuums are stripping or causing some huge amount of like a loss in the machinery parts and man material and loss of life is basically minimal. So, it is not there.

On the other hand considered in place of 6 you have given a warranty life of 8 months. So, what will happen the person would basically continue using that machine and in an around that 6 months 8 months he or she will basically replace those vacuum circuit breakers. So obviously, those extra 2 months that is 6 to 8 at least 6 where he or she should have stopped as the machine to change that vacuum circuit breakers so; obviously, they would be huge amount of extra production, but the loss would be that the probability the propensity of the machine tripping and either a fire breaking out or basically human lives being loss would be much higher. So, in that case the losses would be catastrophic.

So, in this case under in the case when we are considering that overestimation be more problematic, while in the civil engineering case it was under estimation was more problematic. So, in this electrical case over estimation more problematics hence, you will consider a value of a is greater than 0. So, let me read the problem.

Finally, to illustrate the significance of overestimation when compared with the underestimation let us consider a different real life example, consider an electrical company manufactures of vacuum circuit breakers or interrupters, which are used as fuse in high voltage system as for any product these circuit breakers that is vacuum circuit breakers have a working life and it is of utmost important that this value is estimated as accurately as possible.

In case they are underestimated then what is its value in reality that is to be basically found out then the consequence is that, in case they are underestimated so, as I said then the consequences is just labor and man hour loss in terms of production stoppage time on the other hand if the working time of the circuit breaker is over estimated; that means, in place of 6 it is 8 months then the actual figure then the actual figure which is 6 months then it would definitely signify an exponential form loss in monetary terms due to an accident or major breakdown of machineries.

So, for this categories we will consider practical estimation of the problem where the value of a would be greater than 0. So, with this I will end the 12th class and continue discussing more about the stochastic dominance concepts and later on pick up the concepts of utility function how they can be utilized more in the practical sense.

Thank you very much and have a nice day.