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

**Lecture - 52**
**Loss Function**

Welcome back my dear friends and dear students; a very good morning, good afternoon, good evening to all of you. And this is the DADM which is Data Analysis and Decision Making course on the NPTEL MOOC, this is the DADM - I course under MOOC and we are in this 52nd lecture which we have just started the 2nd lecture for the 11th week and you have another one week to go before we wrap up with the course. As you know, this is a 12 week course total number of hours is 30 total number of lectures is 60 because each lecture being for half an hour and each week we have 5 lectures and this is half an hour each and my name is Raghu Nandan Sengupta from the IME department, IIT Kanpur.

So, we were discussing about loss functions and I mentioned that for loss functions initially they are quadratic. Why quadratic? We consider that they have a one to one similarly with respect to trying to minimize the variance. And if the theta value is an unbiased estimation for T, T is the estimate which you get from the sample. Then finding out the difference squaring up and finding of the expected down value of that would give me the loss which is the variance trying to minimize the variance obviously, it is the best way we can achieve.

Then, we considered that we have the linear loss function where you give if you give equal penalties to T minus theta or t minus theta on the negative side or theta minus T whichever way you denote and if k1 k2 which are the weights are equal which is equal to 1. So, it is a 45 degrees line in the first quadrant and the second quadrant. And in case if k1 is more than k2 or k 2 is more than k1 we will give weightages accordingly; that means, over estimation is more penalized or under estimation is more penalized.

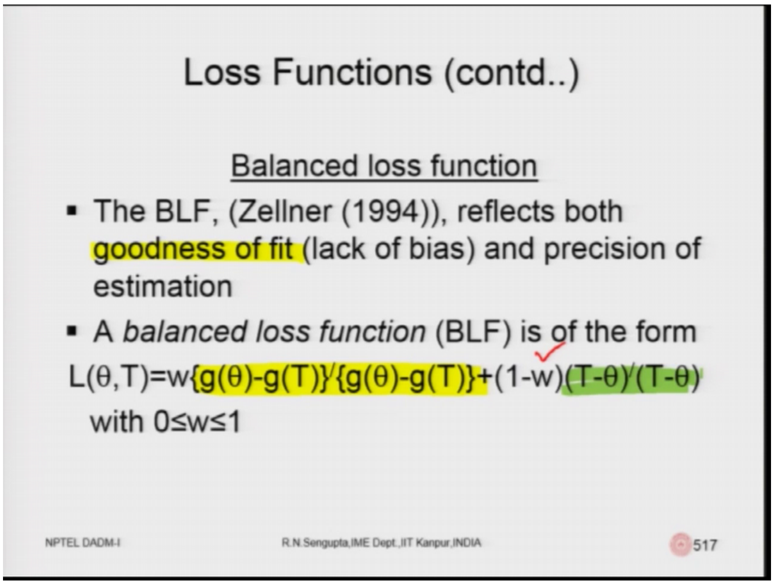
And then later on we also discussed through the diagrams obviously, that even in case if k1 k2 are functional forms of theta also which means for higher values of theta or lower values of theta your overall weightages for the loss function may change rapidly; that means, higher the values are more the weights are; that means, k1 keeps increasing maybe this increase can be linear maybe this increase can be non-linear.

Over and above the loss function, in this concept which we are also discussing and then if underestimation is more penalized or less penalized the values of k would also change accordingly, but consider k I am using just a as a symbol. So, the k by itself is basically a functional form of weights for the thetas or the domain in which we are trying to basically find out the theta.

Then later on, we I in the last few minutes of the 50th lecture I did not mention that if what if the loss function is of the form it is basically the probability or the difference between t minus theta and it has a one to one correspondence with the concept of interval estimation. And this difference being will be loss function will be equal to 1 if T minus theta is greater or less than equal to epsilon. So, that greater than ls can be modelled accordingly and in case it would be otherwise it would be 0.

So, here we are considering the mod means equally distance from the mean value on to the right or the left and they give us very good results corresponding to the fact that we want to find out something to with these loss functions and the interval estimation problems. So, 52nd lecture is about the loss functions.

(Refer Slide Time: 03:56)



Now, Zellner Arnold Zellner in 1984 proposed the balanced loss function. So, balanced loss function is we have more to do with the regression or a multiple linear regression concept. So, when we are doing the multiple linear regression, I always emphasized and I had been talking that you find out the difference of the square of the difference for the which is basically the errors, squared them up. And then differentiate with respect to parameters put those values into 0, find out the parameters in a hat form which is the estimated form and complete the set of other tests which you want to do and but the part is over for the standing guard.

Now, in case of the multiple linear regression what we considered is that these values of the differences which we take for T minus theta or here theta can be say for example, beta is theta and beta hat is t. Or it can be say for example, beta 1, beta 2, beta 3 are the corresponding values of theta 1 theta 2 theta 3 considering is a vector, in that case t1, t2, t3 would be the corresponding estimated values for theta 1, theta 2, theta 3.

Now in Arnold Zellner proposed that what about trying to basically give weightages for the loss function based on two important aspects. One is basically the estimated value which you find out and based on the estimated values of which you found out which is for alpha beta. Then you find out y hat which is basically the best forecasted value of y. What if we give weights to those m the difference of the area errors called corresponding to y hat and y; that means, you are going two step.

First trying to basically find out the parameters values using the estimated values then try to basically give weight some weightages to the loss function corresponding to theta and n and then utilizing some other concept or loss function corresponding to y and y hat. Now, in case if the corresponding loss function with respect to theta and n is basically a quadratic so; obviously, would be quadratic loss function.

And in other case again if it is quadratic in order to the find out the differences or find out the errors of the differences between y which is the actual value and the predicted value is y hat then we will basically have the balanced loss function. This is what we are going to talk discuss about. The balanced loss function reflects both goodness of fit which is the lack of bias and the precision of estimation. So, there are two steps; one is the precision estimation how precise your estimation is and then how good your model basically fits in order to compare the estimated value which is y hat with respect to the actual value which is y.

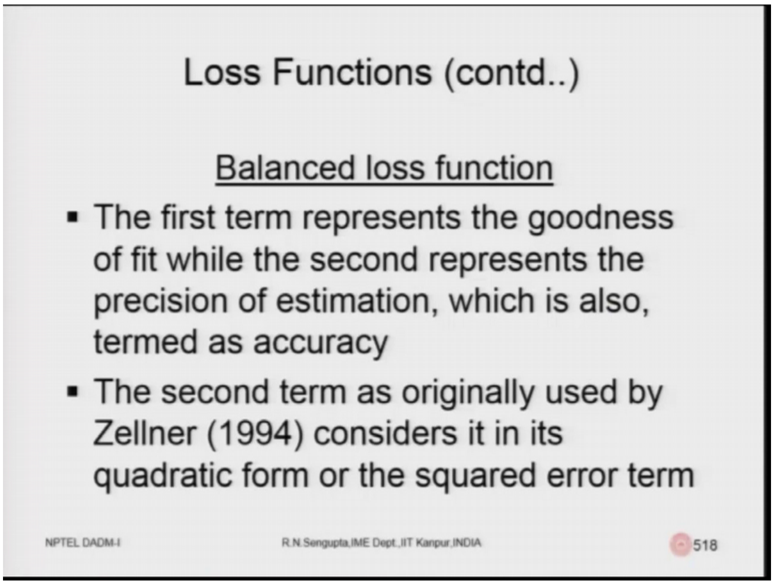
So, y and y hat can be scalar and y and y hat can be vector also depending on that you have different values to predict utilizing the values of x1, x2, x3, xp. A balanced loss function as proposed by Zellner would basically have two parts. So, one is basically the air the weightages which you have which is w for the first part and 1 minus w for the second part. So, what is basically the first part so, if you let me highlight it. So, this if I considered is basically the goodness of fit which I am considering. So, how good or bad the fit is so, here you basically have g theta which is the functional form of theta which can be say for example, y also which is the actual value and g functional form of T would basically the y hat which you want to find out.

So, here what you are trying to find out is the difference between y and y hat. As an example I am giving, squaring them up putting some weights which is w. So, this is basically the goodness of fit let me highlight it with a yellow colour. Now, let me come to the precision estimation this is the estimation which you want to find out let me change in the colour.

So, in this case the precision of estimation was basically again a squared concept which is basically the difference between theta and T or in multiple regression it will be basically difference of alpha and alpha hat or beta and beta hat whatever it is based on the either the univariate case if it is a simple linear regression or the multiple linear regression, whatever it is.

And this is basically the precision of estimation where you give a weightages of 1 minus w. And it should be remembered that Arnold Zellner proposed the balanced loss function as at the weights basically add up to 1 because, you depending on the penalty which you want to put both for precision of estimation as well as for the goodness of fit you will give weights accordingly.

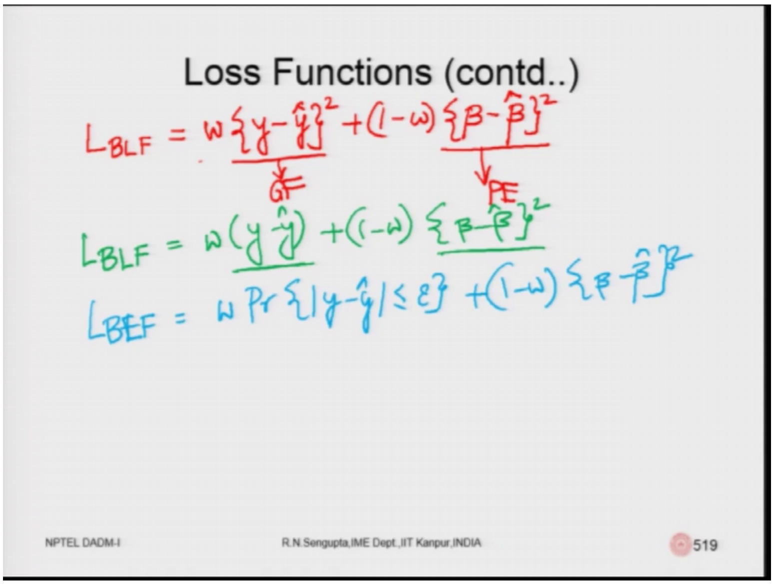
(Refer Slide Time: 08:42)




The first term basically it represents the goodness of fit as I mentioned where the second term represents the precision estimation how good or bad your estimation is so, which is also termed as the accuracy. So, the in case what can be done so, now, I will again create a blank slide and explain few things.

So, first term I am again coming back to my lecture the first term represents the goodness of fit where the second term represents the precision of estimation which is also terms as the accuracy. The second term has originally used by Zellner consist of in it is quadratic form on the squared error sort of form which is there, which means what did Arnold Zellner consider would be like this.

(Refer Slide Time: 09:28)




So, your balanced loss function I will consider is suffix BLF loss function which is balanced loss. I will give weights w and here I am considering that precision of estimation and the goodness of fit. Goodness of fit is basically with respect to y and y hat so, I will put y minus y hat it squared. So, because that was g theta minus gT transpose into g theta minus gT values that that is why it is basically squared plus 1 minus w I am giving the weights to the precision of estimation precision of estimation would be.

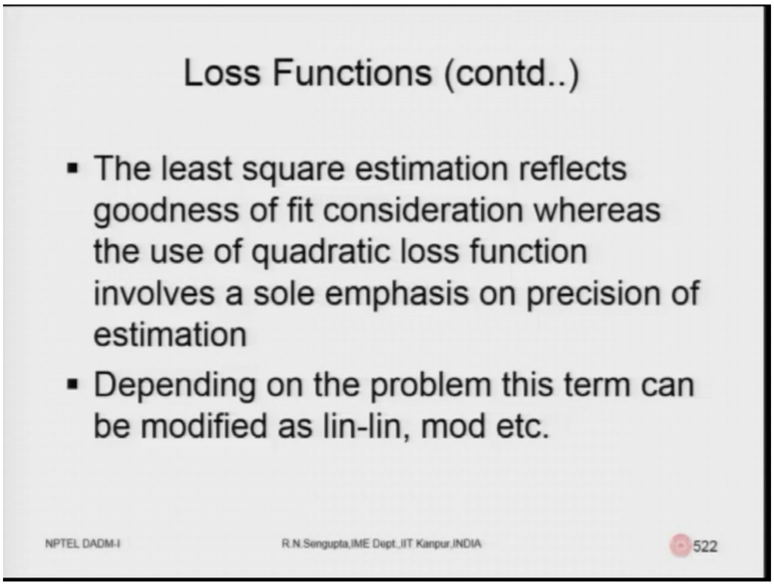
So, I am considering a very simple concept, I am considering there is say for example, only 1 theta you want to find out which is the simple linear regression. So, that case it will basically be B minus B hat. So, this would basically be the precision of estimation. This will basically be the goodness of fit and the loss function balanced loss function can be made accordingly. So, say for example, if you remember I considered what are the very simple loss functions I considered? I considered the linear loss function, the weighted linear loss function the zero one loss functions.

So, the problems can be structured I will use another different colour. Say for example, the balanced loss function can be it say for example, I will consider y minus y hat without the square which is basically linear function plus 1 minus w by beta minus beta hat whole square. So, in this case this is quadratic and this is not quadratic. So, this is basically linear or it can be reversed also; that means, linear part goes through for the second part precision of estimation and the square part comes into the goodness of fit or else it can be like let me use another the balanced loss function can be if you consider the zero one loss function.

So, it will be probability of the difference y minus y hat being less than equal to epsilon. So, if this can be zero and one plus 1 minus w or omega whatever it is. So, this is beta minus beta hat this is square and based on that you precede. So, again here the precision of estimation is quadratic while the goodness of fit is basically the zero one loss function. It can be again changed like the zero one loss function goes to the precision of estimation and the quadratic part comes into the goodness of fit. So, there can be different variance of the loss function, but obviously, there has to be some theoretical (Refer Time: 12:50) in what you have to want to achieve and the practicality of the problem.

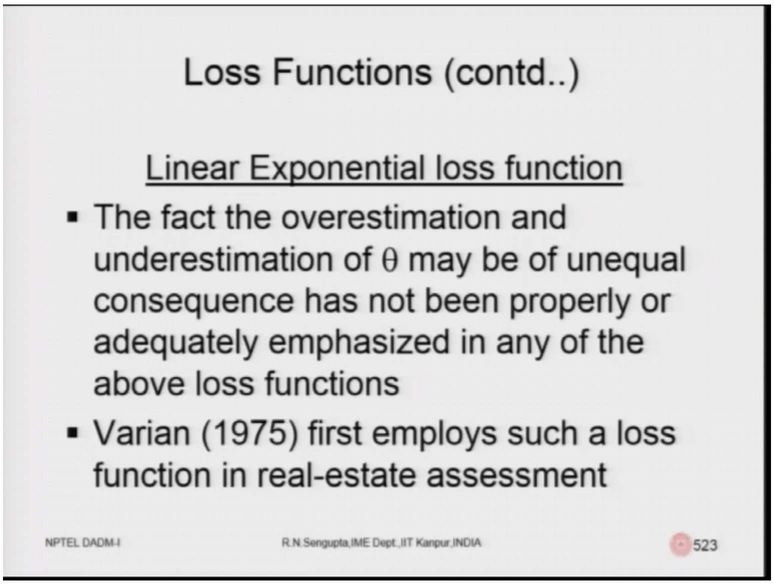
So, this is this was the blank slide I kept it purposefully ok.

(Refer Slide Time: 13:03)




Now, let us continued the discussion so, the least square estimation it reflects the goodness of fit considerations whereas, the use of quadratic loss function involves a sole emphasis on the precision of estimation. So, that will depend on the type of problems which you have. So, depending on the problem these terms can be modified into linear loss function as I mentioned, It can basically be the mod loss function, it can be quadratic loss function, it can be any sorts of loss function in both for the good goodness of it and the precision of estimation.

(Refer Slide Time: 13:38)



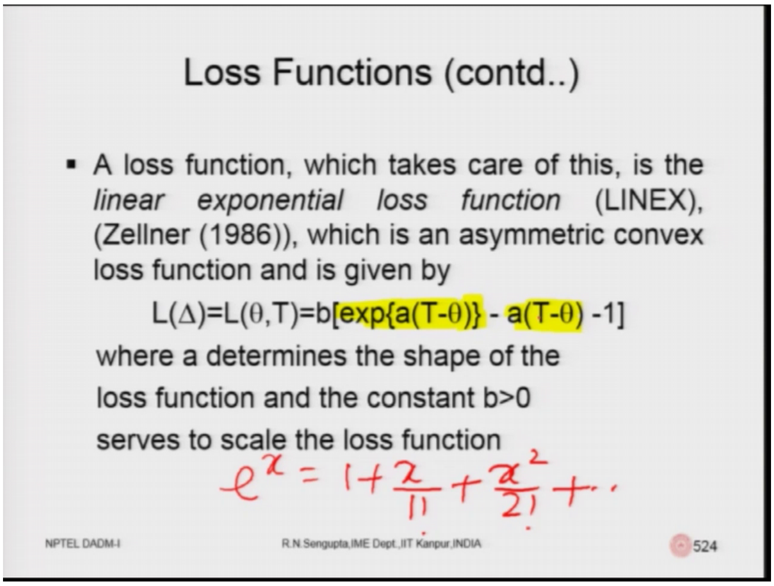

Now, Hal Varian in the year 1975 we use this loss function which is known as the LINEX loss function which is linear exponential loss function. So, LINEX both the word will LINEX which is coming is for the part it is linear or I should use the highlighter my apologies. So, this is LINEX. So, this is the LI NE EX LINEX loss function. So, sorry this double e we not come. While you will use this also is that is the LINEX loss function.

Now, in the LINEX loss function as the words mentions it there are two parts which is the linear part and exponential part and Hal Varian you to first employs employed such a loss function in the real estate as assessment in order to basically find out the prices and the populations and so on and so forth. But later on, the work was left to Hal (Refer Time: 14:56) Arnold Zellner also again who basically found out all the statistical properties of that.

So, now here comes basically the actual description. The fact that over estimation and under estimation of theta which is the parameter of the population may be of unequal consequences, unequal importance, higher means more penalty, lower means less penalty or it may be mean that lower being more penalty higher being less penalty. What I mean by higher low is basically the difference of the estimated value and the actual value, which is Tn or T minus theta.

Theta is basically the parameter value and T or T suffix n is basically the estimated value for theta. The fact that over estimation and under estimation of theta may be of unequal consequences has not been properly or adequately emphasized in any of the above loss functions. So, based on that we will try to basically proceed and learn something about the LINEX loss function.

(Refer Slide Time: 15:55)




So, as I was discussing, after Hal Varian the this work was taken up with Arnold Zellner and he proposed the loss function. A loss function which takes cares of this is the linear exponential loss function on the LINEX loss function which is Zellner 1986 which is given by the asymmetric convex loss function and it is of this form. So, here L delta well delta is basically the difference between T minus theta or theta minus T in whichever you try to basically portray.

So, they would basically be the before I discussed the loss function, they would be two parameters; one is basically a which is the shape parameter and theta a and b is basically the scale parameter, depending on the shape and scale how it can be changed. Now, the loss function basically consists of the linear part which is here and the exponential part which is here.

So, hence it is known as a linear exponential loss function. Now notice very interesting thing if the difference between T and theta is very small; that means, you are trying to basically estimate or find out the loss or corresponding to various precision of estimations very precise. Precision of estimation I am not talking from the viewpoint of balanced loss function is very precise, very nearby. Then look at the exponential term expand it so, if you expand it if we know that the expansion of e to the power x is basically 1 plus x divided by 1 factorial plus x square by 2 factorial and it continues.

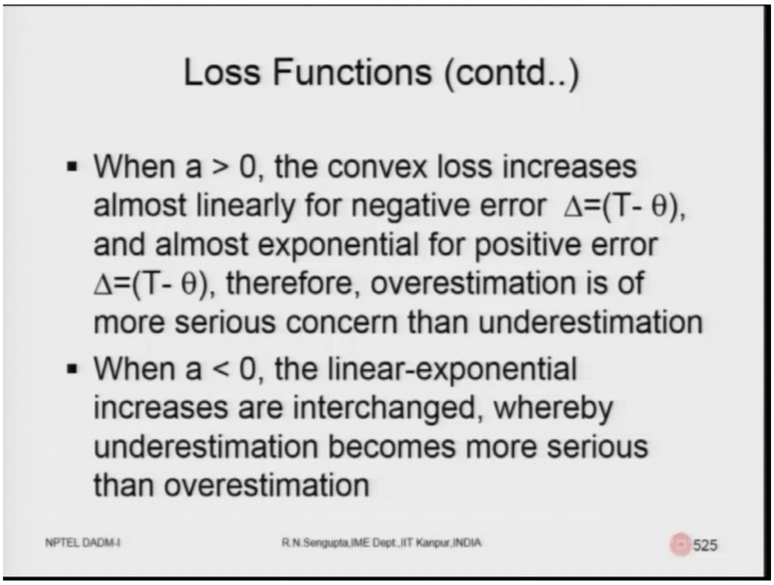
So, if you basically expand it e to the power a T minus theta, the first term would be 1; 1 1 cancels here. Then the second term here is basically plus a T minus theta. So, that term and this second term which is which I am highlighting now also cancels and if you consider the squared error term; that means, the third term which is squared error term is only important and other term becomes very less.

Because, when I mention that the difference between the estimated value and the actual value Tn minus theta is very small. Hence, the cube root 4th part root 5th root can all been ignored so, what we are left is basically the quadratic form. So, hence if you see that the linear exponential loss function for very precise estimation can be converted into a squared error loss depending on the practicality of the problem.

So, here the as per the Zellner, I am just reading it that functional form basically consider b is not there. It does not make would not make much a sense because b being parameter which serves to scale the loss function depending on how you are trying to tackle the problem.

So, again the I will again repeat this is a linear part, there is an exponential part and we saw that the LINEX loss function can be converted into a quadratic loss function depending on the efficiency or the precision of estimation considering that. The 3rd power onwards which means that cubic power and onwards would basically be ignored. When a is greater so obviously, here is where the asymmetric city will come for the loss function and it will become be very clear to all of us. So, if a is positive large number then in that case exponential part would of would be dominating the linear part for the right hand side of the values of the difference between T and theta such that delta is positive.

(Refer Slide Time: 19:17)



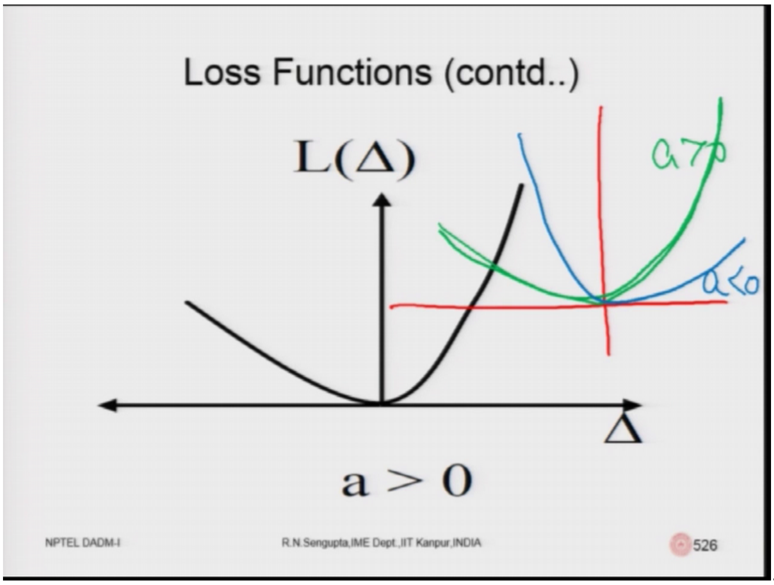

Because, if delta is positive so, obviously and a is also positive then the LINEX the exponential part will dominate the linear part. Hence, here overestimation would be more of a problem - more of a concern and in this case in the case which you are talking about, if a is positive; obviously, in the negative zone where T minus theta is negative obviously, that exponential part would not be dominating the linear part hence it will be more or less linear in nature in the second quadrant. I will come to the graphs later on please wait, but in the other case if we consider say for example, the values of this a is negative.

So, if a is negative and again you are considering Tn minus theta is positive so obviously, the exponential part would not be dominated by the linear part and the linear part and the exponential part when combined together. In that case, so called for our understanding we will say then that the overestimation is less penalized, but if I consider in the underestimation in this case, there we have already have the difference between Tn minus theta as negative my a is also negative. So, negative-negative would make it positive. Hence, you have a exponential part which would be increasing exponentially in the in the second quadrant; that means, for underestimation.

So, hence for a values as positive overestimation is more penalized for a values as negative underestimation is more penalised so obviously, we will be asking that what are the different type of examples which we have I will come and come to that later on.

When a is a greater than 0 the convex loss increases almost linearly for negative errors which is theta delta which is T minus theta and almost exponentially for the positive errors which is T minus theta. Therefore, overestimation is a more serious consequence than underestimation as I just mentioned. When a is negatives less than 0 the linear exponential increase and decrease they are swapped within each other they are interchange. Whereby, the underestimation becomes more serious than overestimation so obviously, you will basically formality of problems accordingly.

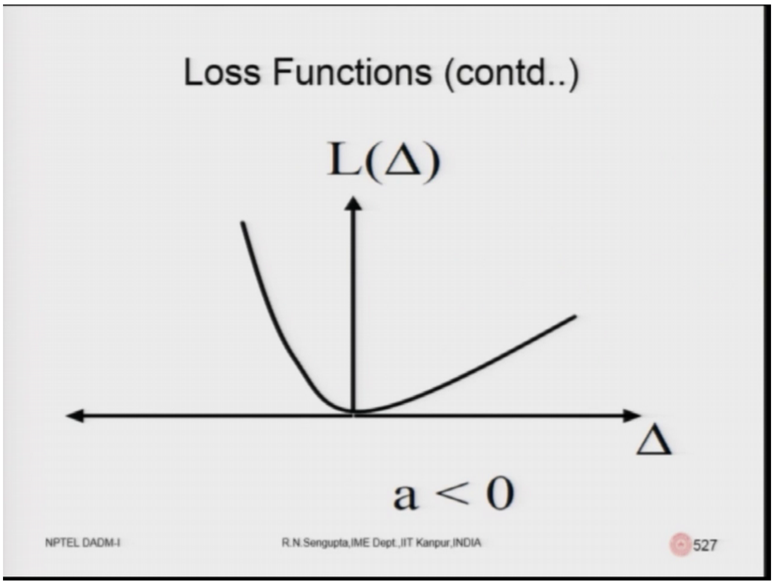
(Refer Slide Time: 21:49)




So, this is the loss function which is basically looks like this, I think I should drawing to it. So, it look like if overestimation is more penalized and use a different colour, overestimation is more penalized then. So, this would be more, this would be less and in case if it is underestimation is more penalized so, this will be the loss function.

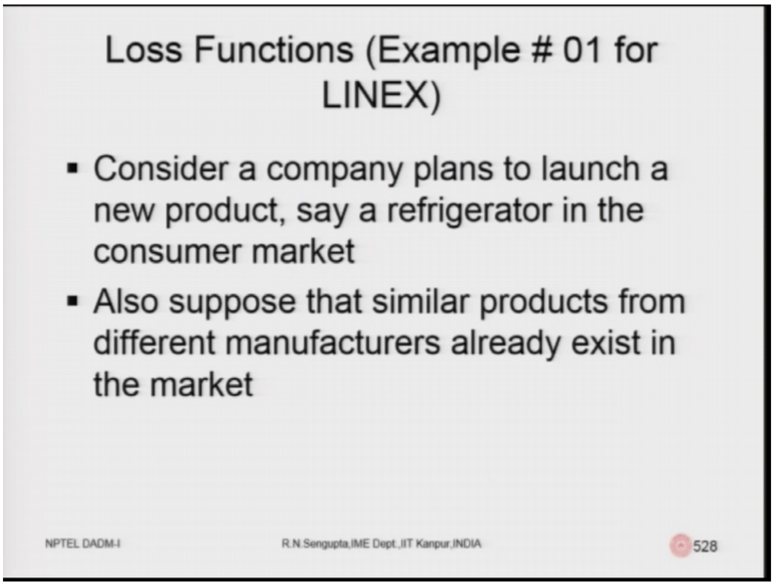
So obviously, in this case a is less than 0, in this case, a is greater than 0, based on that, you will try to do calculation. So, here if you see a greater than 0 over estimation more penalized and underestimation is less penalized.

(Refer Slide Time: 22:50)




Here if you see, for a value is less than negative underestimation is more penalized and over estimation is less penalised. So, an obviously, this one should remember delta value is basically T minus theta and the loss function is also modelled accordingly.

(Refer Slide Time: 23:10)




So, let us consider simple three examples. Consider a company plans to launch a new product say a refrigerator in the market and it is in the consumer market based on this which one to proceed and also suppose the similar products from different manufacturing companies already exists in the market. So obviously, when the company wants to float the problem let me tell you the background.

So obviously, there would be warranty life. So, warranty life means that if the product fails before the warranty life you are liable to change and replace the product as per the norms of the company. And in case if it fails after the warranty life; obviously, the person will basically make the payments and change the product accordingly. Now, the issue is that if it is a new product and there are such other products good competitors in the market who are also new. And you want to basically go for a warranty life which will basically give you the best possible action; that means, maximum the products would not fail during the warranty time within and it will start failing after the warranty time.

So, people would be making the contracts accordingly. Now, consider the warranty life we say for example, 6 months arbitral consider and in case 1 and case 2 in case one you give a warranty life of 8 months and case 2 you give a warranty life of 4 months. Now, if I consider 8 months is 2 plus 2 more than 6 and 4 months warranty is minus 2 with respect to 6 months which is the actual warranty.

Now, if your actual penalty is quadratic, so in that case, 8 minus 6 which is plus 2 whole square and 4 minus 6 which is minus 2 whole square, would give you a penalty which is quadratic which is squared error loss. And obviously, the penalty is both for positive and negative would be of equal amplitude which is fine for our calculations, but we are not taking practicality into consideration.

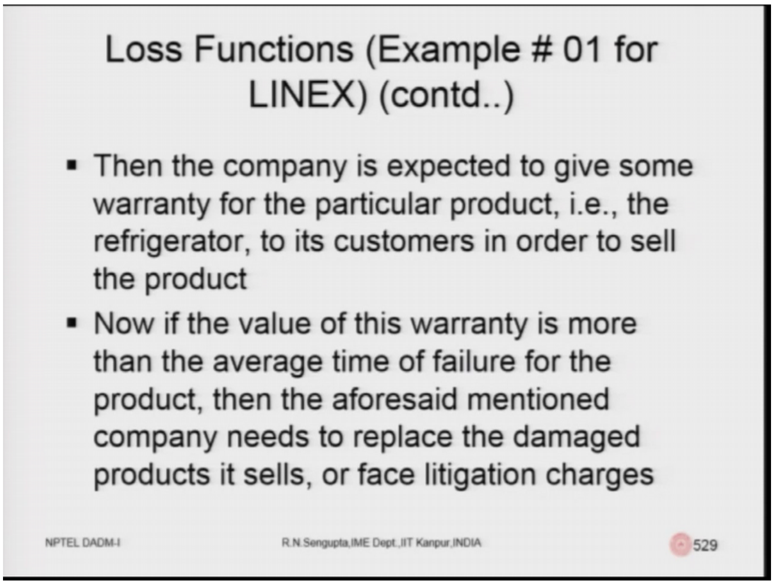
Consider the product basically which you are going to sell. If you give up warranty which is higher, so what will happen is that in the initial period people will be more willing to buy your products because your products warranty life is much higher than the rival products. But, consider in the other hand, maximum on the p of the products, say for example, in case if the warranty life is actually not 8 it is less than 8 and basically products start failing which means that even if initially people brought your product in ;bought your product in large numbers.

But, the failure rate has increased, that means, you have to basically replace them or pay the penalty by either till buying those products again back to them or basically replacing those expensive parts whatever it is.

So, initial profit or revenue has basically been offset by the case that you have not been able to predict the warranty life as it should have been which is basically in and around 6, but you give a estimated value of 8. So, there the losses would be different and they would not be quadratic, they would be unequally penalized. Now, consider the other case where if your warranty life actually should be 6, but you gave 4 which means that people would be more willing to buy your rival products you will lose a huge amount of market in the initial set. But obviously, it may happen that if the product if the products actually are nearer to the actual value which you should have been predicted.

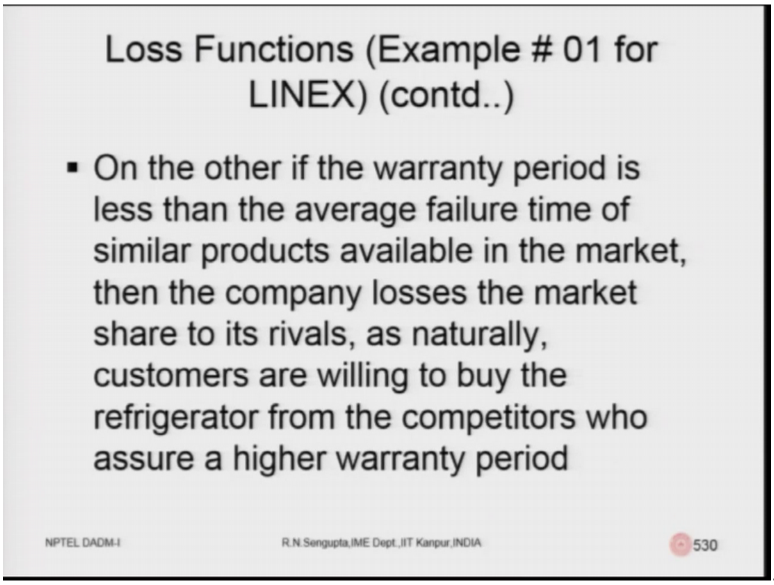
So obviously, later on we will slowly gain some of the market share; that means, again it is unequally penalized. Now, consider this I will read this slide and then again give another example. So, consider a company plans to launch a new product say a is a refrigerator in the consumer market. Also, suppose the similar products exist from different manufactures which are already exist in the market. So, this is the background which I already mentioned I am going to read it again.

(Refer Slide Time: 27:10)




Consider then the company is expected to give such a warranty for the particular product that is the refrigerator to his customers in order to sell the product in the market, which is fine they should give warranty. 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 products it sells or face basically litigation action by the customers.

(Refer Slide Time: 27:36)




On the other hand, if the warranty period is less than the average failure of the similar products in the market then the company losses the market share to it is rivals as naturally the customers are willing to buy the refrigerator from the competitors who assure a higher warranty life. So, as I said that if your warranty life is less than the average time then the companies would lease the initial markets share.

So, all of them will go to the rival party or the other manufacturers, but even if the if they are willing to come back later on and to buy a products, but this amount of offsetting would definitely not be possible in order to take care of your loss. So, what values of a or whether it is positive or negative you would give for this example will depend on what your over estimation and under estimation problems would be with respect to the estimation problems which you basically would have done.

So, this is a refrigeration problem we will see later on. It can be an electrical problem, it can be a civil engineering problem and based on that we will study that how loss functions can be utilized. So, with this I will end this 52nd class and in the 52nd class I just considered the LINEX loss, a simple example of the refrigerator and continued two more examples in the as I said in the electrical engineering, the civil engineering case in the 53rd class and continue more discussion about the multivariate statistical models and other methods.

Thank you very much and have a nice day.