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

Lecture – 13
Balanced Loss Function

Welcome back my dear friends very good morning, good afternoon and good evening to all of you. Do you know this is a DADM-II which is Data Analysis and Decision Making two course under NPTEL MOOC series. And this course is for 30 hours which is for 12 weeks and each week we have 5 lectures each being for half an hour and we are in the 3rd week as you can see from the slide which is the 13th lecture. So, we have already completed 11 12 and we and the 3rd class for the 3rd week. And as you know that after each I do repeat that and please bear with me for after each week we have one assignment.

So, in totality there will be 12 assignments and after the end of the course there will be an end sem or final examination based on whatever coverage has been done. And my good name is Raghu Nandan Sengupta from the IME department IIT Kanpur. So, if you remember we are discussing about loss functions. So, to give up again a recap even though it may be a repetition again I will say that I am going slow because the general feedback which I am which I got from DADM-I and TQM too which courses which I have taught. So, they I should I will should pace it a little bit because considering we have 60 lectures to be conducted.

So, whenever you are trying to estimate your main aim was basically to find out the estimate of the parameters based on the fact the estimation estimated value or the function should have two properties from the statistical point of view which have already discussed in DADM-I was basically unbiasness another one was basically consistency.

Now, whenever you are doing that it may not be possible. So, that is why you try to utilize the concept of the loss function. And in loss function I did mention that its very nice theoretically very proper prim and proper we get good results that if we consider the unruly square equation for the estimation of the loss functions. Similarly so, ordinary least square would be quadratic one so, which in which case both over estimation and under estimation equally penalized.

Now, it could also be turned into a fact that if you want to have equal estimation or equal errors in estimation both over estimation under estimation we can use the lin lin or linear linears loss functions which are 45 degrees line. These diagrams I have already discussed and then you can basically have a lin lin function corresponding to the fact that over estimation will be more penalized then in then underestimation in that case the line in the first quadrant would be greater than 45 and line in the second quadrant would be less than 45. And in case if underestimation more penalized will basically and then over estimation will have the line in the second quadrant to be more than 45; that means, we are looking from the left hand side and the line in the first quadrant would be less than 45.

But it was soon found that Hal Varian in the 1960’s from the economics point of view gave this loss function which is LINEX loss function linear exponential part. So, the first part is one part is Linnaeus second part is exponential and Zellner in the 1980’s basically came out with the fine statistical properties about this loss function.

Now, if to give an example, if a which is the parameter value is greater than 0 and you have overestimation. So, in that case over estimation will be more penalized than under estimation and in cases a is negative so, in that case underestimation is more penalize than over estimation. So, in the former case where over estimation more penalized because a being positive exponential function will dominate the linear function in the first quadrant.

So, obviously, it will be exponentially increasing and for the case when a is positive in the same case if you go to the second quadrant in that case linear function will slowly start dominating as delta increases on to the left I am looking at the real line number line in front of me. So, a linear function will dominate the exponential function so, it will be underestimation will be less penalized.

So, if you switch the concepts; that means, a is non negative in that case overestimation would be less penalized because the linear part will dominate the exponential part as deltas tends to increase goes more on to the right. And in case when in the same sequence if you consider underestimation will be more penalized than overestimation because in that case exponential part will slowly dominate the linear part.

Now, to give an example I did see in DADM-I definitely something about the examples. So, I will consider that and then I will read those example which are stated. So, consider this loss functions in which the LINEX loss function. So, if you are trying to build a dam the examples I have already given. So, building the dam would be and if the height is 120 feet and in two cases it is 122 and another second case is 118.

So, if you use the quadratic loss function it is basically 122 minus 120 which is plus 2 whole square is this 4 and in the second case is 118 minus 120 is minus 2 whole square is 4. So, in both in cases where overestimation underestimation equally penalized this plus 2 and minus 2 would have the same consequences in the practical concept of the theoretical concept.

But, I did also mention that that is practically not actually possible because in the first case when you build a plus 2 meter height the initial cost is man hours lost, extra material cost and there is an overshooting of the cost. But the positive side is that if a flood comes then, the propensity or the probability of the flood breaching the dam which is already of 122 feet height would be much much less.

So, the catastrophic loss would not be there. So, an initial cost would be higher which is not as catastrophic as phenomenal as in the natural calamity loss, but in the case when the height is 118 there the cost initially for man hours material is much less, but when the flood comes the propensity of the flood breaching the dam is much higher. So, the whole consequence would be a devastating natural calamity loss, loss of manpower, loss of cattle agricultural everything will be flooded and inundated.

So, in this case we will consider practically underestimation to be more penalized then over estimation. Now, if we consider the second example that was first one from the civil engineering second which we consider say for example, from the electrical engineering you have a machine, the machine has fuse fuses or trip switches vacuum circuit breakers and the actual overall warranty of those vacuum circuit breaker is 6 months.

So, consider initially two cases; in case one you overestimate with 6 by 8 and another case you basically over esteem underestimate 6 by 4. So, in the first case the difference is 8 minus 6 is 2 whole square of that is 4, in the second case the difference is 6 minus 8 is minus 2 whole square of that is 4. So, if you use the quadratic loss function in both the cases it is equally penalized, but actually the situation is not that consider case one where you overestimate.

So, in that case what will happen that you will basically be tempted to stop the machine after the 6 months warranty life check the machine and change the vacuum circuit breakers well. In that case we are doing the extra production month over and above the 6 months would; obviously, be beneficial for you on the production front profit front, but the problem is that if there is a catastrophic voltage fluctuation then they would may be huge amount of manpower loss, accident and the whole machine may be destroyed and people may be hurt and so on and so forth.

Now, if you consider the case two when you underestimate so; obviously, the production would be affected you will stop the machine much before hand then 6 months when you want to change the circuit breakers, but the probability of any catastrophic loss for the failing of the warranty life of that vacuum circuit breakers would not be there because you are replacing those vacuum circuit breakers much before so; obviously, they may be some one or two rare cases, but in general it would not be there. So, the overall loss in the second case would be initially only man powers loss, but catastrophic loss would not be there man power loss because you have stopped the production.

So, if you consider this actual practical situation you will see that you would rather consider overestimation to be more penalize than underestimation and solve the problems accordingly. Consider third example where we are not sure that what the values of a would be in the first case values of a was negative for the dam case, in the second case value was a was positive for the electricals machine part and for this case say for example, marketing or it consider the marketing problem.

So, you have a product in the market and you consider the warranty life of the machine to be 1 years or 12 months and you consider two situations; in one case it is overestimated to 15 months another case is underestimated by 9 so, 12 is basically 9. So, in the first case 15 minus 3 minus 12 is plus 3 plus 3 whole square is 9 in the secondary is 9 minus 12 is minus 3 minus 3 whole square is 9. So, if you consider that the squared error loss in both the cases it is equally penalized as 9 and 9.

Now, see what are the actual situation the which may be different because considering that we have to decide on the sign of a. So, in the first case if you basically give a warranty life higher than the actually it should be which is 12 months so; obviously, people will be more tempted to buy a product you basically capture a good market share. But the probability of these machines or the products failing so, products can be say for example, a refrigerator fridge, AC, coolers whatever it is failing would be much higher.

So, people and in case it is a warranty time it means you have to basically replace those machine replace those products. So, if the probability of failure of those machines is much higher because you have given a warranty life of 15 months and when in actuality and in practical sense it should be 12 months. So, many permissions would be failing and you have to basically replace them so, they would be a business loss or your overall goodwill in the market would be lost.

So, you make a huge loss later on and your competitors gain again the market. But in the second case, when you basically predict the warranty life to be 9 months. So, in this initial case what will happen then; obviously, products which are 12 months being been proposed by your competitors people would definitely be more willing to buy this product. So, initially you lose the market share, but it may so, happen that in the long run that as the you basically depress the product and you slowly pick up. So, you will be tempted to give a little bit more higher warranty life.

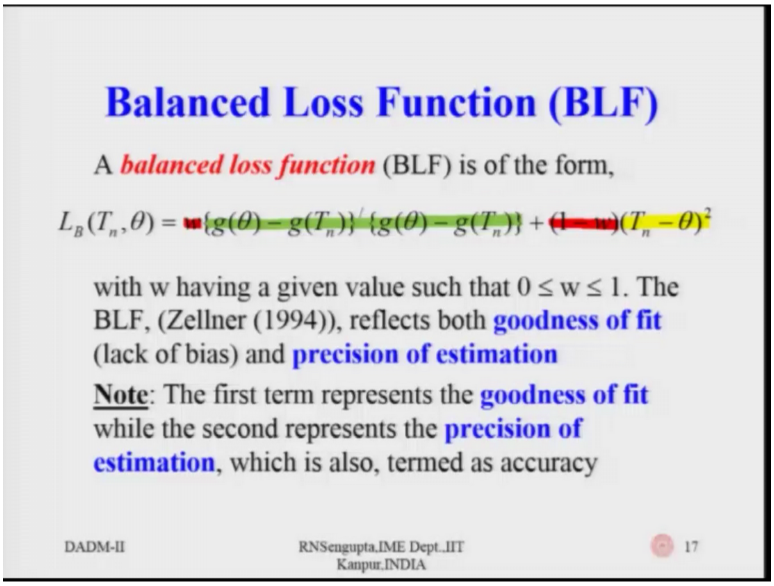
So, people may be tempted to basically come under your marketing scheme and buy that product. So, initial loss is basically compensated by a increase in the market share later on. So, the value of a which you will decide for this case on the marketing the a is plus or minus would depend on what you think is more important for you. Is it say for example, more of market share loss or is it may basically gaining of market share which is positive or is basically losing the customer satisfaction and litigation case is being filed by the customers because the products fail much before then what you have basically said that should be the warranty life.

So, you have to make a decision that what the value of a should be. Now I also discussed that the as you expand the value on the value of the LINEX loss with respect to e that means, take the expansion of e the first two terms which is 1 and a delta by 1 cancels the only term left is basically the second power a square delta square by 2 factorial and then higher power. So, if we ignore the higher powers basically the LINEX loss in an around the value of 0 becomes a quadratic loss function. Now in regression model what we generally consider is that in regression model we consider and I have already discussed that in DADM-I so, I will again repeated in regression and then come to the actual use in DADM-II.

Why I am discussing that in regression model we consider that you have the errors. Errors is basically the difference in the actual value of y and the estimated value of the forecasted value of y which is y hat which is basically the error you squared the error, sum the errors for all the values of n which you are going to take. Differentiate this square sum of the square of the errors with respect to the parameters, parameters would be for the multiple linear regression would be the first parameter would be alpha, second one is beta 1, then third beta 2 so on and so forth. Till beta k you have basically k plus 1 equations considering alpha partially differentiate these sum of the square of the errors with respect to all these parameters put them to 0.

So, all these k plus 1 equations and get the k plus 1 parameters which would now be alpha hat beta 1 hat, beta 2 hat, till beta k hat. But now general Zellner again in later years proposed a loss function which will now discuss.

(Refer Slide Time: 14:28).



So, it is basically a balanced loss function balance means you are balancing the loss with respect to some criteria which is there. So, I will basically discuss this loss function as it is and then I will come to the slide for first let me give you the background.

Now, whenever you are estimating in the multiple linear regression case what you are trying to do is basically first step is to find out the alpha hats, beta hats, that is alpha hat, beta 1 hat, beta 2 hat so on and so forth and then use those hat values which is the estimated values in order to basically estimate the y hat value. So, they basically you are doing a two step process. Now Zellner basically further extended that and basically had a loss function where you basically penalize these two loss with the weight age of the delta and 1 minus delta where the sum; obviously of delta and 1 minus delta should be 1.

So, the balanced loss function looks like this and I am going to consider both the losses for the estimation part and the forecasting part estimation being for the parameters and forecasting being for the value of y as both being quadratic. So, that mean basically highlight. So, delta is basically w here. So, the first part is what I will do is highlight with the yellow one is the estimation point; that means, you have theta as the actual parameter and you are estimating that using a sample size of n. So, which is T n is the parameter estimate.

So, T n minus theta whole square would give you the squared error loss with respect to the estimation problem and you basically multiply that with 1 minus delta 1 is the w which is the weight. That is point 1 and we will basically use that term as precision of estimation I am not highlighting and I am just basically pointing out with the pointer this is the precision of estimation because your main task was initially to estimate the parameter values.

Next once precision estimation is done will basically try to forecast the error term or find out the error term. So, that basically comes for the goodness of it again I am not highlighting it I am just moving my pointer there which is the goodness of fit or the lack of bias. So, the goodness of fit would basically be the now there you have basically a functional value of theta. So, it will basically be g theta and when you basically estimate that using the T n value which is the estimated value of theta then it will be a function of g T n.

So, again I find out the errors corresponding to the forecasting errors. So, this is basically is again a quadratic loss function because you take the transpose multiplied by the same matrix or the vectors and the error term which is there is again is w. So, you should take the sum of 1 minus w and w it becomes 1. So, a balanced loss function now I will read will have w being the values giving to the weights w is between 0 and 1 the balanced loss function proposed by Zellner in 1994 reflects both goodness of fit and the lack of bias at the precision of estimation. The first term note presents the goodness of fit while the second basically represents the accuracy of the initial estimation value. I would not discuss, but I will basically highlight it using the qualitative concept.

So, generally technically what you will find out is that when you consider the balanced loss function and you want to basically or say for example, let us go one step backward. So, you basically have a exponential distribution or you basically have um gamma distribution or you basically have a normal distribution so; obviously, when the estimated values are found out considering the squared error loss you have basically squared error estimate.

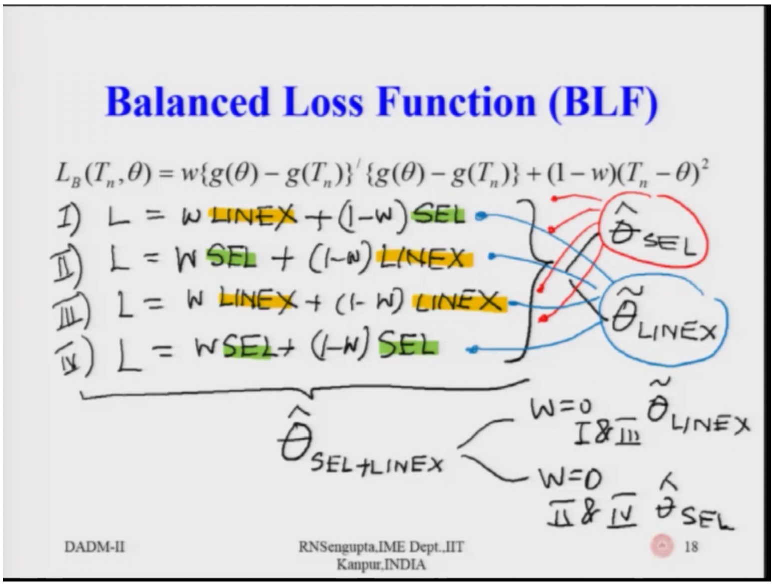
So, in this case when you have the squared error estimate you will basically have from the normal distribution the best estimate for the sample mean is x n bar which for the population parameter which is the mean value which is mu the best estimate is basically the sample estimated which is x n bar. Now similarly when you change it to the LINEX loss again we see that we have a different estimate and you can find out what is the estimate and it has been found out by Zellner.

So, you can and you will basically have a nuisance parameter based on which you will basically say that this is the best estimate of the mean value which is mu under the condition when you have the LINEX loss functions. Similarly, if you have gamma distribution or this extreme value distribution or for the exponential distribution all the parameters would basically have one type of estimate found under the ordinary least square and another type of estimate found under the LINEX loss.

Now, it has been proved that if you are able to find a convex combination of this loss function that is not in the balanced loss function is basically for the case for the multiple linear regression. When you consider for the estimation of say for example, the parameters of any distribution if you basically take a convex combination of this loss function which is LINEX loss and squared error loss then the general estimate which you find out using these two combination of the loss function actually becomes the actual estimate either under the case of LINEX or under the case of squared error loss depending on what is the value of lambda or w you are trying to choose.

So, at the boundary conditions the values of this estimate which you find out would basically equal to exactly equal to under squared error loss and another case it will be exactly equal to the LINEX loss. So, which is a beauty of this the new estimate which can find out it can be proved. Now similarly you if you basically venture on to the task that if you want to find out say for example, balanced loss function considering 4 different combinations. So, what are the combinations let me go one by one.

(Refer Slide Time: 20:32).



Now, I will basically draw it. So, this is the balance this is the second slide. So, I have just made it in order to explain. So, there are IV cases, case I. So, L I am only using l it will basically be the balanced loss function first would be LINEX w into LINEX plus 1 minus w into squared error loss, IInd case is w into squared error loss plus 1 minus LINEX, IIIrd case is LINEX loss balanced loss function sorry LINEX plus 1 minus w LINEX and the final case which you have already done, but I will just mention it again is w squared error loss plus 1 minus w squared error loss.

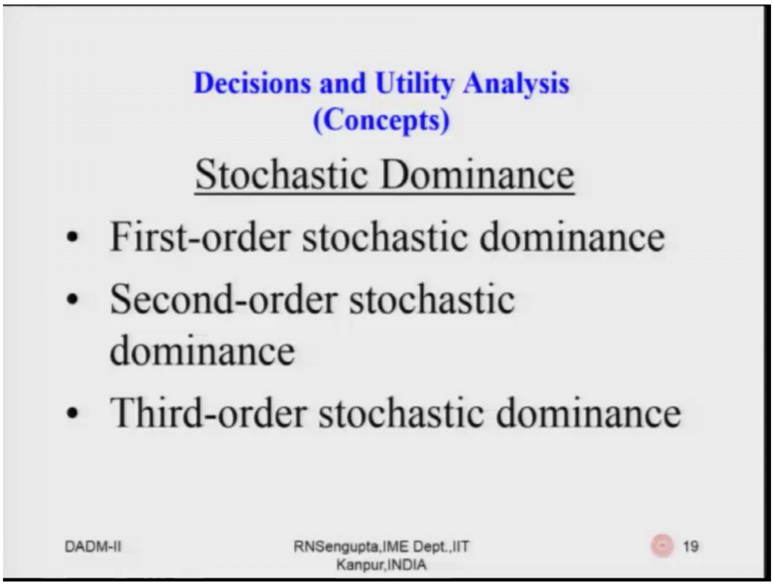
Now, let me highlight them. So, I am using LINEX here, I am using LINEX here, I am using LINEX here, I am using LINEX here, I am using LINEX here and I am basically mark the squared error, squared error here, squared error here, squared error here, squared error here . Now and now actually the task is what? The task is for all these combinations I have two different estimates to utilize number one I will try to utilize theta hat which will be the estimated value of under squared error loss and another denote by theta tilde a which is the estimated value under LINEX.

So, basically we will try to combine point one actually combined is squared error loss with all these 4 combinations that is point one. And in the other case we will consider the LINEX loss under these 4 conditions and then for try to find out depending on the situations which you have they can be different type of loss function which will be basically giving us estimates under the loss function which will given us different answers point one. Point number two what will be interested is that well is it possible to find out a general estimate for all these 4 cases, 4 cases means here a general form of theta say for example, hat I am using the hat which is now basically a combination of SEL plus LINEX depending on the value of w and at the boundary conditions whether when w is 0 and w is 1 for all these cases we get. So, w is 0 for this case for say for example, for the Ist and the IIIrd case we should basically get theta tilde LINEX for the Ist and the IIIrd and for the case again for the Ist for say for example, for the IInd and the IVth if this is 0 so, this is for the Ist and IIIrd for the IInd actually it should be theta hat squared error.

So, if you are able to prove it. So, it will be a good step where actually in the case of distributions we have been. So, generally we will need to find out that whether you can do it for the case for the well balanced loss function and obviously, we will consider some assumptions.

Though so, this is just a precursor based on which we will basically try to come back and consider it later. So, again remember we will have basically 4 combination squared error LINEX, LINEX squared error, then LINEX LINEX and squared error squared error for the combinations for the second part is basically first precision of estimation then biasness or goodness of it. And for all this case we take the cases of LINEX loss estimate and the squared error estimate which are already noted in literature find a general form of the estimate such that our boundary conditions w is equal to 0 or w is equal to 1 we get actually what it should be depending on the case where we get it is the estimate under the squared error loss and estimated under the LINEX loss.

(Refer Slide Time: 26:19).



We will consider the simple case of stochastic dominance and I will basically go a little bit slow here also with examples by the way we did discuss many things about the balanced loss function this was more of precursor problems anything else would be carried on later on because this concept of LINEX loss will be coming up later on when we consider different type of other nonparametric decision making.

So, first consider the first order stochastic dominance, third second is the second order stochastic dominance and we will also consider the third order stochastic dominance. So, we will basically go one by one for the first order, second order, third order and consider them with examples. So, with this before continuing with the first order one I would like to close it here and continue with the discussions for the for the stochastic dominance and come back to the concept of utilizing the LINEX loss later on in the multiple decision making problems have a nice day and thank you very much.