Data Analysis and Decision Making - I

Prof. Raghu Nandan Sengupta
Department of Industrial & Management Engineering
Indian Institute of Technology, Kanpur

Lecture - 34
Multivariate extreme value distribution

A warm welcome my dear friends, my dear students. Very good morning, good afternoon, good evening to all of you. This is the DADM 1 course which is Data Analysis and Decision Making under the NPTEL, MOOC series a set of course's. And this course is for 12 weeks which is for 60 lectures total content hours would be 30. And as you know that in each week, we have 5 course classes each class being for half an hour and we are in the 7th week. So, today is the 34th lecture, with another two lectures are each of half an hour. We will we should wrap up the 7th week set of classes.

And hi my name is Raghu Nandan Sengupta from IME department IIT Kanpur. So, if you remember you are discussing about something with a very simple example about copula theory. And based on that we said that or the extreme value distributions and we were discussing that in the extreme value distributions what we are interested is to find out; I will not go into the properties and definitions with full rigor.

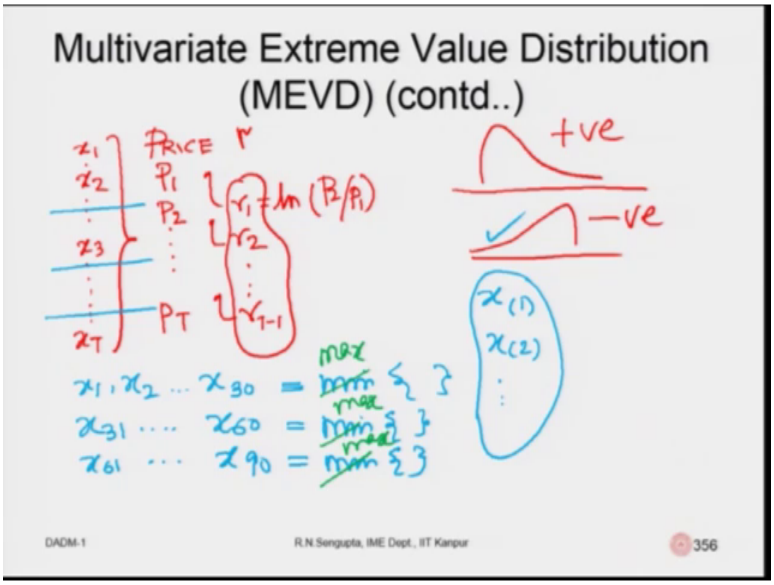
So, given the data said you want to find out the extremes so, the example where they can be used. So, if you remember that I did discuss that copula or which is a linkage function to give two random variables. It is basically utilized to find out the relationship of two random variables like the extremes. Now covariance's are nice linkage functional nice dependents function between two random variables, but their linear in nature. So, in the extremes it may become non-linear say for example, the some financial crisis occurs like see for example, the one in 2008.

So, it had a devastating effect on the banking sector, on overall economy and; obviously, if I want to find out the risk level pertaining to the devastating financial crisis which happened in USA. I want to find out what effect would it happen say for example, all prices or say for example, on the financial sector in India; obviously, in India was shielded because it is own policies it was able to overcome the extent of losses to a greater extent. But say for example, a country like Greece country like other European countries they were affected diversity degree.

So, in that case we will try to basically consider that the relationship between the effect and at the extremes are shocks which happen with huge amount of intensity, but the actual probability of a curry of those shocks are cutting is very rare events consider. Then we basically use extreme value distribution. Extreme value distribution is also used for flows of fluids there we want to understand that and the speed increases or that the velocity increases. How, what effect it will have on the cause oppression front on the force front and so on and so forth.

Now, whenever we are considering extreme value distribution, what we actually do in very simple cases. So, consider the data set is given and as I was demonstrating I will demonstrate again. As the data set is given we will take the minimum value, now the minimum value is basically for a set of data set. So, I will try to basically portray that using in this x this ppt slide only. So, let me add few blank slides and proceed let me add few blank slides. So it will easier I thought I would not, but let me basically go through it in detail.

(Refer Slide Time: 04:13)



Now, consider the let me use the red colour. So, you have a dataset with values of x 1, x 2, x 3 and it goes on. See for example take x T the T can be anything 1000, 2000 whatever. Now I want to find out the extreme value of this in distribution, consider it is extreme value, considers the stock prices for the very simple and case. So, technically we have the so what we are doing is that we have the prices. So, price is given per day so it is PI and for any stock P1, P2 till P T based on that we find out the return. So, return it would not come up here, but it basically would start from the second one, r1 is equal to l n of P2 by P1.

Similarly you will have some r2 continue this way you will have basically the values as given, so there would be one less return. So, when I so this would be teak say for example, coming up to T minus 1; even though the they would technically we should be written one step up. Now, these values the returns which I have are extreme value in the sense I will show it to you there are data points; actually they have been utilized. So, consider if it is a positive distribution so it will be like this, this is a negative distribution it will be like this. Which is basically for the positive this is for the negative this is the losses this is the gains.

Now, what we do we take a block of them, now you use a different colour. So, we take a block it does not mean only x T it can be set of values. So, in the first block you have x 1 x 2 so on and so forth. Then consider your taking till say 30 for the example which is 31 to 60, 61 to 90 and it goes on. Now what I do is that I take the minimum of this, I take the minimum of this I take the minimum this.

So, I take the minimum of this and once I have the minimum. So, consider the minimum I will denote as x this is in a bracket 1. Even though they are today x actually that is the formula for the ordered pairs order different ascending or descending, but I will just use it for simplicity. So, this is the first one this is for the second and so on and so forth. Now once I have these minimums technically it means that I have taken a small sample of whatever the size is, from an actual population which consists of the minimum distribution, which is an extreme value consider either minimum is this one not the maximum.

Now, I do a sort of bootstrapping; bootstrapping is basically consider you have a high boot where there are a lot of shoe eyes, the lace to tie up, you need a lot of tying up to do, those old fashioned boots. So, in bootstrapping it basically means you are tightening up from the lowest level and as you go up you basically do it in such a way, that you are trying to find out the best estimate in the concept of bootstrapping.

So, in bootstrapping technically we take a small sample and increase the size. So, the increase in the files many different methods which we do increase the size in such a way that the actual sample size now explodes lot of actual readings are there. So, the underlying probability of any sample size remains the same. So, say for example, in a sawn small sample you have 1, 2, 3, 4 and 5 these two times. So, the technically of six values 1 occurs 1 by 6, 2 occurs two ways 1 by 6, 3 occurs through 1 by 6, 4 occurs 1 by 6, 5 occurs 2 by 6.

Now, you grow it up in such a way that you basically add up 1, 2, 3, 4 and two number of files many number of times proportionally. So, if you add them proportionally then say for example, you add them 10 number of times. Now the total number of readings would now be basically 60. So, now, in this case 1would occur 10 by 60 which is one 6 which remains as the same relative frequency of the chance of the probability as it was earlier. In the same case number 2 would occur 10 by 60 again which is one-sixth, 3 would occur again 10 by 60 which is one-sixth, 4 would occur again 10 by 60 would see which is one-sixth and 5 will occur 20 by 6 divided by 60 which is two-sixth. So, concatenation in this way or increase in this way basically would give you our replication on the population which you are trying to emulate.

Now, with this data you basically jumbled up, jumbled elements basically shuffle it. Like you have a set of values inside a box and you may say we jumble it up, such that the probability when you pick which number you will pick up is absolutely random. So, when you pick up those numbers and keep repeating it time after time and then the actual sample mean in the long run would be exactly equal to the population mean to give a simple example. What I mean by not by the concept of bootstrapping, but picking up the sample characteristics of the sample averages, time after time would basically replicate the population average.

Consider do you have a this example I think I have repeated in the initial 3 4 lectures of this course. Consider you have a coin and it is an unbiased coin the probabilities are half enough for a head and a tail. But say for example, you toss it 100 number of times for the first case and you get 45 heads and 55 tails. So, obviously the relative frequency of a head and a tail is 45 by 100 which is 0.45 and 55 by 100 which is 0.55 respectively.

Now, keep repeating it do it again 100 times, find out the heads find out the tails note it down. Again do a 100 times key and do it infinite number of times. Now the num the proportions of the heads you will get will; obviously, differ in one case it can be 0.45 in the first case, then it can be 0.46, then in another case it can be 0.60, so (Refer Time: 11:47) any with value is possible. Now if you find out the actual averages of this average, this is total number of heads divided by the total number of tosses in the long run, then obviously that would be equal to 0.5 which is actually the probability.

In the same way what we considering is that, we can basically pick up these observations time after time find out the sample means as the average of the sample mean actual case considering the unbiased and consistency property in the long run should be exactly cooled with the population mean. So obviously, you will use these concepts in order to basically bootstrap and find out that considering extreme value distributions are to be prodded you can do that

Now, this example which I gave for tossing a coin would also be true for the case when here we see a rolling a die, you are playing a game of ludo and considering the it is an unbiased die. And the probabilities are one-sixth, one-sixth, one-sixth for each 1 2 3, then one-sixth 4, one-sixth for 5, one-sixth for 6. Now when you roll it 200 number of times the number of chances of getting either 1 or 2 or 3 or 4 or 5 or 6 actually would definitely be different then than the respective probability which is 1 by 6.

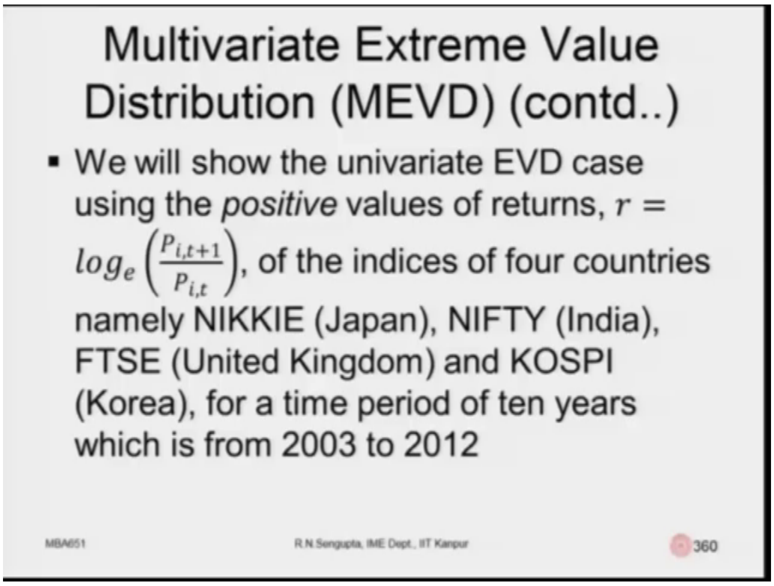
Now, you keep repeating it many number of times infinite number of times and find out the proportions of number of heads. Proportions number of sorry not heads number 1, proportions the number of 2, proportions are number of 3, proportions are number of 4, proportions of number of 5, proportions are number of 6, and in the long run you will find out the probabilities for all of them numbers are painting is 1 by 6. So, you basically use this concept in the very simple case of bootstrapping; obviously, there are many complicated methods, but generally bootstrapping would work accordingly

Now, in the mean the case of the minimum as I have told that you take many of the minimums. And basically concatenate it concatenates means increase the size proportionally jumbled it up and pick up pick up a set of samples or observations and find out you mean and continue doing picking up the samples with replacement remember. So, as that in the long run average of the sample means is equal to the population mean. Similarly you can do the same thing considering the concept of unbiasness and consistency for the variance and so and all these things.

And from them you can find out the parameters of the distribution which I have did tell is the location parameter, shape parameter, and the scale parameter, which we then generally do not buy alpha beta and gamma. Now the question would be; can I do it for the maximum? Yes, you can again do the same thing for the maximum, but maximum for the EVD's, but in that case remember that you will basically take the maximum value for each subsets this 30 which I mean mentioning.

In each case take the maxima so, I will take the maxima take the maxima take the maxima you repeat it, find out the maxima, take it at one go concatenate it do bootstrapping by another means and find out the mean of the overall population. And again you can do for the standard deviation and all these things. These are blank sighs I will just skip. Now this what I have mentioned I will do it I will show you some actual results which I had done for few of the M.B.S and the M.Tech studies so I will just show it to you.

(Refer Slide Time: 15:37)



We will show the in univariate extreme value distribution, I am taking the univariate case because if you remember trying to find out from the multivariate case a ranking system may change. Because we rank the first values and obviously, it does not mean that you are ranking the second way set of random variables or vice versa. So, we will show the univariate extreme value case using the positive values of returns. So, the returns are given as I mentioned log to the base e of the price of today divided by the price of yesterday so that will be the returns of yesterday.

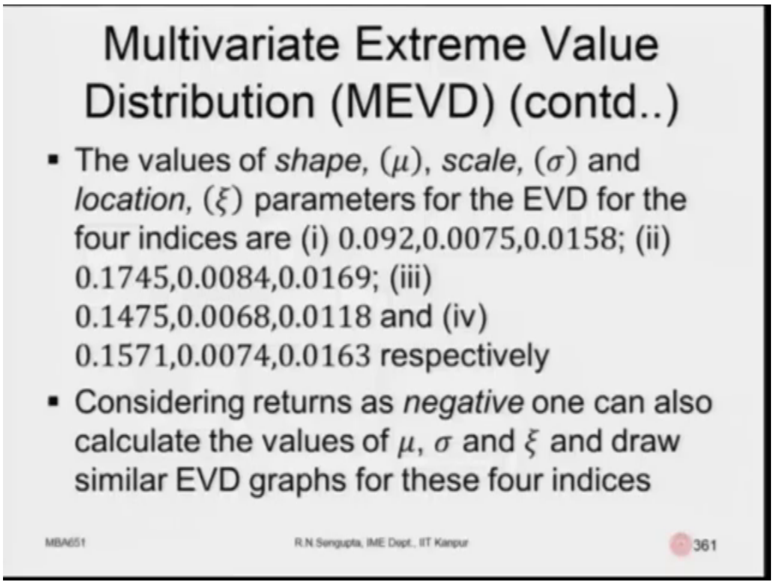
Similarly price for tomorrow divided the price of today, will give me the returns for today and I will proceed accordingly. We do it for the indices of four countries NIKKIE which is for Japan, NIFTY so obviously, in India here of the NAC and the BAC, I am taking the national stock exchange which is for India. FTSE which is financial times index which is for United Kingdom or UK and the KOPSI which is for basically for South Korea. We take the data from 2003 to 2012. So, what we do is that we take the end of the day closing price for this index; index is a conglomeration of stocks depending on whatever number of stocks which are there.

So, once you do that you for you and actually each year on an average you will have about 240 data points I did mention that with considering the Saturdays holidays holiday and the national holidays. So, based on that you have the closing price, closing price on that day and technically you; obviously, people would intuitively think the closing price for today would be the opening phrase by tomorrow technically that should be the case that is not the case that besides a point. So, you have the closing price based on the closing price of 240 parts, but per year. So, you will take basically 2003 to 2012 and those number of some deep days, which you have you will find other corresponding returns by the formula.

So, it will be example log of price on second of January, 2003 you will divide the price of first January 2003 and it continued. Till the last value would belong of the 31st December 2012 divided by the price of 30th December 2012 so, once you do that I have the returns. Now, one thing remember these I am not doing bootstrapping not doing the so (Refer Time: 18:26) I admit did not mention one thing so taking the blocks and doing the bootstrapping is known as block bootstrapping.

So, they can be blocks on the windows which can be overlapping non overlapping in my other concept when I am taking the blocks of x 1 to x 30, x 31 to x 60, x 61 to x 90. So, those blocks are non overlapping you can take overlapping block bootstrap also. But in this case because of the data size is very large you can take the returns as such or you can definitely do the block bootstrapping with overlapping p time periods on without overlapping time periods and then plot the returns from this for different indices.

(Refer Slide Time: 19:15)

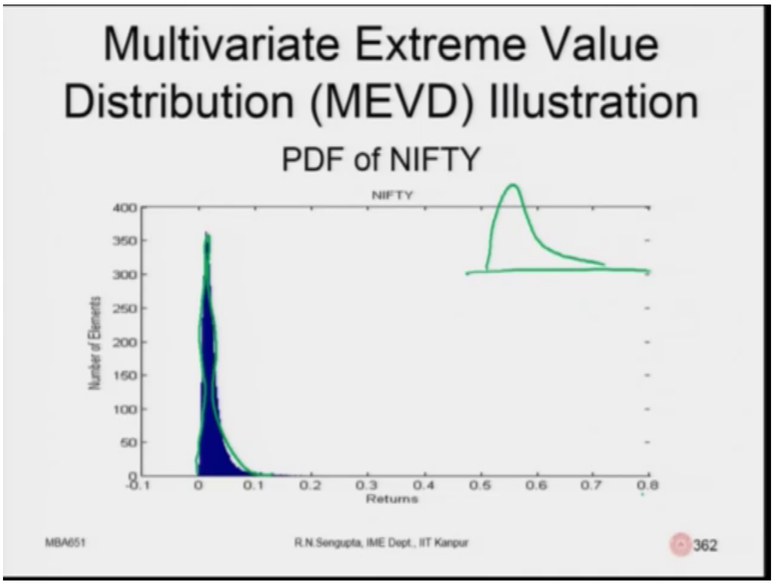


Now, as I mentioned for the these values which you get this expected value after you do either the bootstrapping do not do the bootstrapping from this boots this mean value, variance, skewness kurtosis which are the first with respect to the so first moment respect to the second moment respect to the third moment respect to the fourth moment that this mean variance skewness kurtosis. You can find out the parameters of the distribution. Now, if you remember the parameters concept for a uniform distribution discrete case it is a and b, for a normal distribution is mu and sigma, for exponential it could be 8th lambda or only lambda for poison it can be only lambda so all the n for a minimal it will be n and p. So, all those are the parameters just to recap.

Now technically for any distribution the parameters are in general given as alpha beta gamma. So, these here I consider this mu sigma and eta. So, they are basically the shape parameter, scale parameter and the location parameter and for the EVD's values these for this for the stocks which we take generally for the Indian one, for the FTSI, for the KOSPI and all these things. We find out the parameters of the shapes scale and location as 0.092, 0.0075, 0.0158 that is mu sigma and eta. Similarly for the second case is 0.14745, 0.0084, 0.04169. For the third case which is FTSI it is 0.1475, 0.0068, and 0.0118.

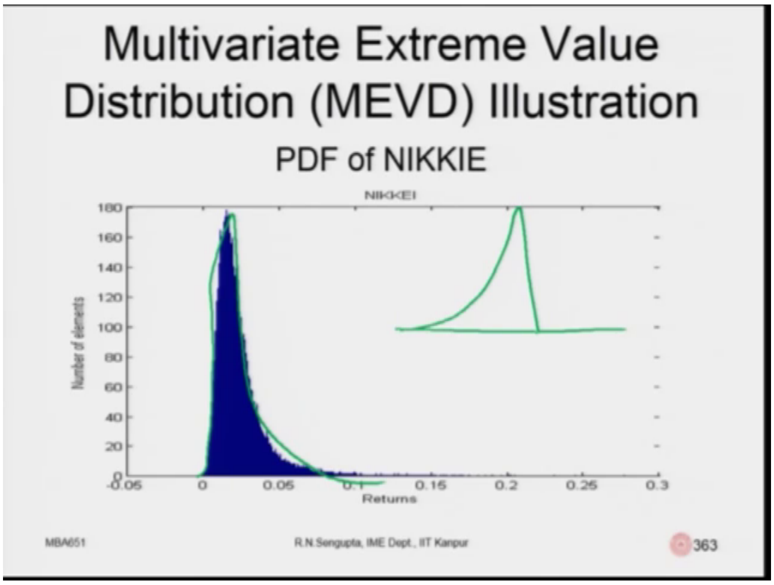
And for the last one which is the KOPSI it is 0.1571, 0.0074, 0.1610, 0.0163 which is basically the values which I have are for the NIKKIE and then the NIFTY then you have the FTSI and if I lastly. Considering the returns negative so we are considering the returns are technically as negative or positive whatever you can consider any moment the maximum. We can also show calculate the values of we can do that find out the values of mu sigma and eta. So, in this minimum negative case will take the minimum and the positive case we will take the maximum. So, we are basically taking the maximum value hence we are taking the positive returns. So, the graph which you are going to we are going to show is not for the minimum on the negative one.

(Refer Slide Time: 22:07)



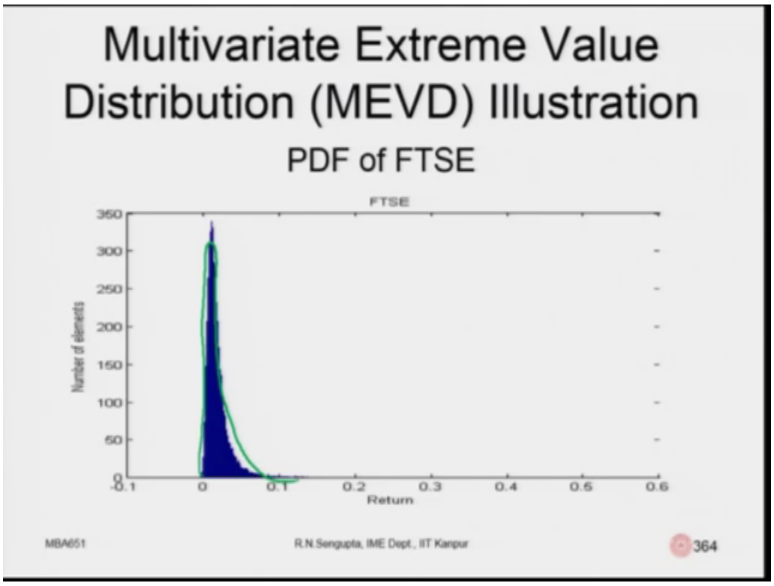
So, this is for the NIFTY so this is the multivariate extreme value distribution example, but only considering the x to the university extreme case. So, if we remember I will just draw it that we had drawn the graph like this so here it is and these are actual values so obviously these are true. This is for the NIFTY and I basically draw the PDF and these values are given the turns are given on the x axis starting from minus 1 to 0 plus 0.8.

(Refer Slide Time: 22:43)



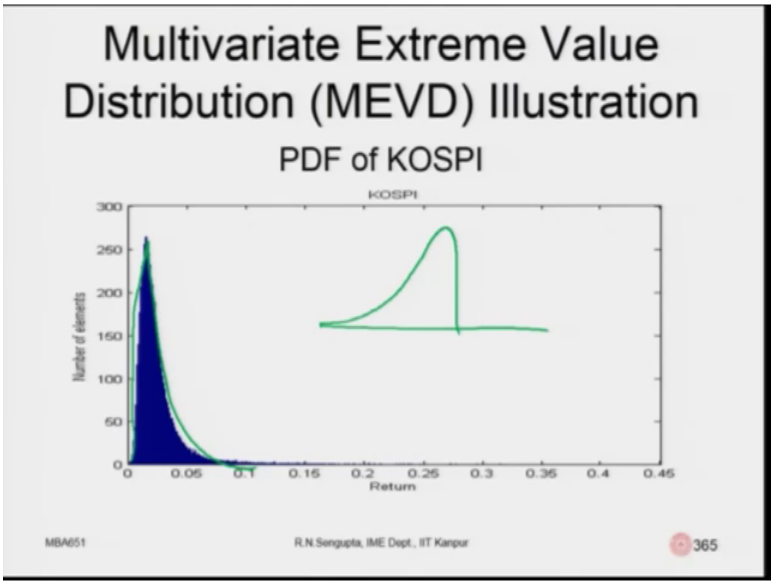
Similarly, if I do it for the NIKKIE this is the PDF of NIKKIE. Then again the value this comes out to be quite nice quite nice as an extreme value distribution for the univariate case for the positive returns. Now, I wished if I take the negative returns just for the information, it will be like that can definitely double check.

(Refer Slide Time: 23:09)



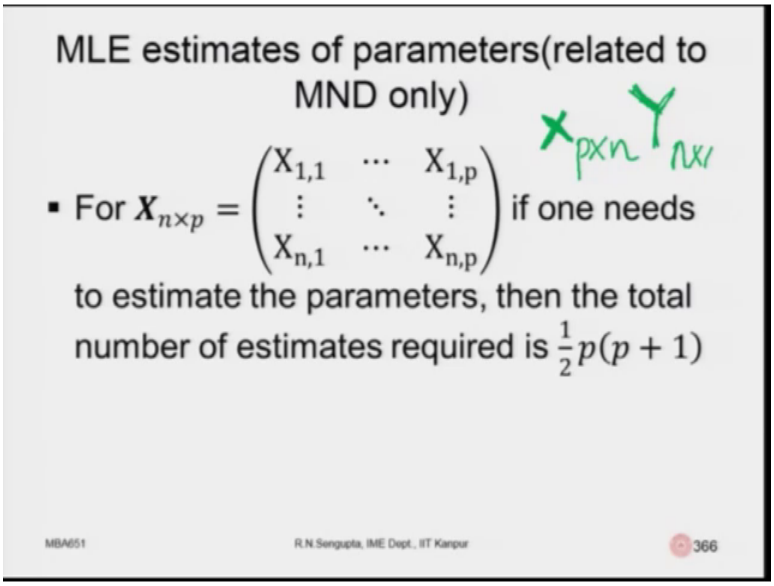
Then I do it for the FTSE for the UK market again returns are like this these are positive returns.

(Refer Slide Time: 23:21)



And finally, when I do it for the KOSPI which is the, for the South Korean market again the values are like this. So, similarly you can do it for the negative one. I am just giving return one to one correspondence so we can understand.

(Refer Slide Time: 23:41)



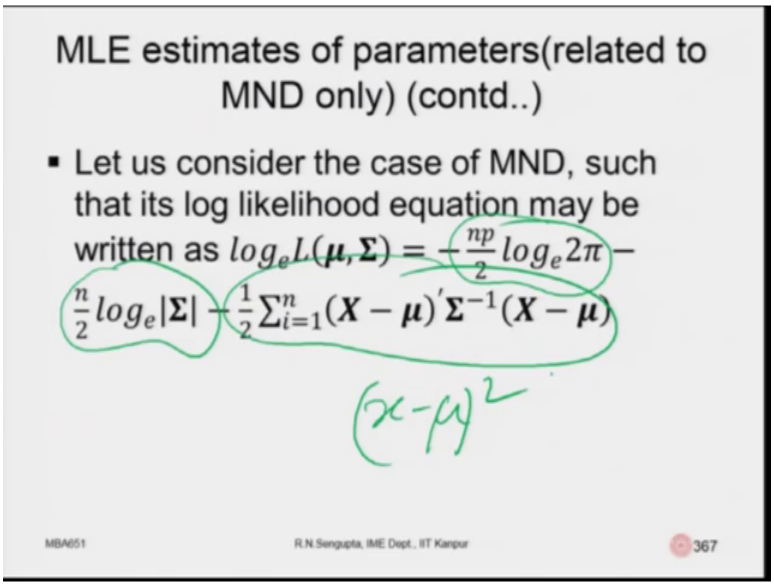
Now we will castle consider the simple methods of trying to basically find out how you do it for the MLE estimate of the parameters related to the multivariate normal distribution only. So, consider that x s are the pam the x is basically a matrix of size n cross p. And I am basically lonely so obviously, this is a nutritional difference here I am noting it down as the first value 1 comma 1 is x 1 1. And I go along the first row till the value which is x 1 comma p. So, these are basically the first values for the p number of random variables.

So, technically you could have considered x p cross and also, but is this just a nomenclature I am following. Second row would be x 2 comma 1 till the last value which is x 2 comma p. So, in that case it will be the values for the second reading for all the random variables 1 to p. And similarly the last row is x n comma 1 till next n comma p which denotes the nth reading for the random variable starting from 1 to p.

So, if you one needs to estimate the parameters then; obviously, the total number of parameters to be estimated would be half into p into p plus 1 depending on the mean values under standard division you are going to calculate. Now I am not digressing just a point so when we go into the multiple linear regression also will consider the same concept time and again that given the ma multivariate normal distribution we will try to basically estimate the betas and proceed accordingly.

So, there obviously, X can be given as p cross n. So, this is a bowl remember. So, that is why I am highlighting it is a matrix. And similarly when you consider Y later on for the multiple linear regression Y would be a vector of size n cross 1. So, those nomenclature of multiplying the rows and columns should be considered accordingly so that there is no dimensionality problem.

(Refer Slide Time: 26:15)



Now, in general if you remember I have not discussed that and did not I did not go into details. So, basically in the multiple linear regression models or multivariate statistical models the method to basically estimate the parameters would be almost the same, the conceptual method would almost be the same as we had done in the univariate case. So, in the univariate case if you remember what we are done we have considered the maximum likelihood estimation problem, in the maximum likelihood estimation problem the meant method of moments.

So, consider the maximum likelihood estimation moment methods, you took that consider the parameters are unknown you pick up a sample and the DLS values are x 1 to x n, only one case univariate. Now, you found out the corresponding probability of x 1 occurring, x 2 occurring, x 3 occurring, to x n occurring. Then I want to basically find out the corresponding probability so I would basically try to find out their IID remember. So, hence I multiply find out the corresponding probability of occurrence of all of them, one at a time and you are basically picking up with replacements only it would not matter.

Whether x 2 is different or x 1 is different from each other. We find out the maximum occurrences of this probability given that alpha and beta are such values which will maximize them. But considering that if the log equation is a monotone increasing function we convert into a logarithmic function and differentiate with respect to alpha beta gamma which ever they are. So obviously, they would be the corresponding parameters which are scaled shape and location. Put it to zero find out the linear set of some decent equation if possible solve it and find out the alpha hat, beta hat, gamma hat.

And; obviously, if they are not then you use different type of iteration methods Runge-Kutta method, the Newton-Raphson method. In many of the cases you have to use the different type of he resting method, genetic algorithm, artificial immune system, you have to use the and colony optimization, all these things could be used. So, we would basically be using those methods on so, but they are from the theoretical point.

So, let us consider the case of the multiple linear multiple normal distribution such that we formulate the log like root functions accordingly. So, the log like dual functions would be exactly the same. So, in the first part you would basically have no parameters with respect to x or the width with respect the parameters. Here is basically something to do with the variance covariance matrix and this is the part which is basically to do the third part with respect to the differences this you remember in the simple case. So, differentiate with respect to alpha beta gamma whichever the parameters put it to 0 and solve it. So, I will continue discussing more than in the 35th class of this NPTEL MOOC series of DADM. So, with this as end this class and have a nice day.

And thank you very much for your attention.