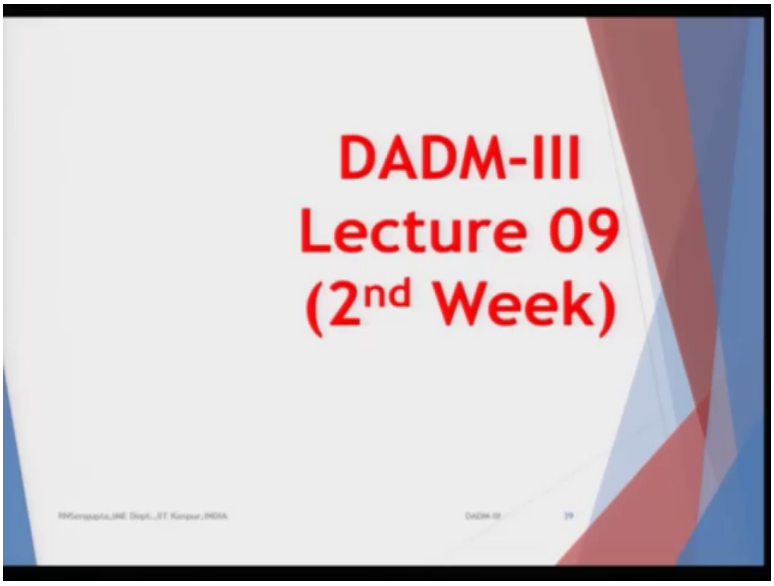
Data Analysis and Decision Making - 3
Professor Raghu Nandan Sengupta
Department of Industrial and Management Engineering
Indian Institute of Technology Kanpur
Lecture 09

Welcome back my dear friends, a very good morning, good afternoon, good evening to all of you wherever you are in this part of this globe and as you know this the DADM-3 which is Data Analysis and Decision Making-3 course on the NPTEL mock series. As you know this course total duration is 12 weeks which when converted into number of contact hours is 30 hours and number of lectures would be 60 that means each lecture is for half an hour and each week we will have 5 lectures, so after each week we will have assignment.

(Refer Slide Time: 00:55)



So considering that you are or we are in lecture number 9 which is the second week you have already taken assignment number 1 and after the 10th one we are all set to take up assignment number 2.

And my good name is Raghu Nandan Sengupta from the IME department IIT Kanpur. So if you remember we are discussing about utility theory and consequences and how they can be utilized for optimization. So I will try to basically further proceed in the concept of utility theory try to

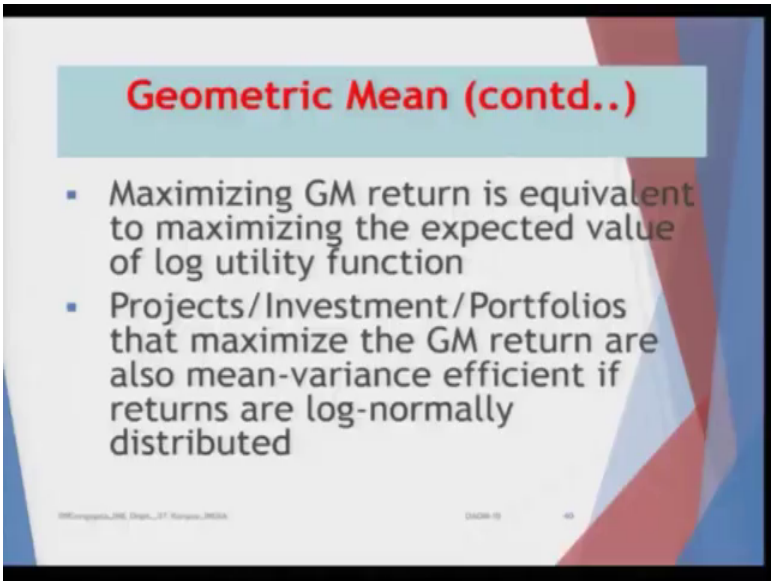
wrap it up by 10th lecture and then from the 11th one try to start of with the concept of optimization and the realistic concept and later on see that how utility theory could also be utilized for the optimization problem.

Now, when we are doing utility theory to give a very brief review obviously all of you are very well acquainted what we have considered, we have basically considered the concept of utility, what is expected value and obviously based on the fact we can also find out the concept of the variances, then we considered the different (con) the ideas of non-satiation and how risk lover, risk indifferent, and risk haters can be clubbed according to second derivation of UW being greater than 0, equal to 0 or less than 0.

Well that will give you the trend how the first derivative is behaving and now will see the first derivative will always considered to be positive because we are aware that human beings are non-satiated that more you give to him or her more he or she wants. Then, we consider the concept of absolute risk aversion property, relative risk aversion property and what are the significances of A prime and R prime were discussed then we went into the 4 different types of utility functions and how the properties of A, A prime, R, R prime can be utilized to find out the characteristics of the utility functions were also mentioned.

Later on, then we went into the concept of certainty values, certainty equivalent value what it means in the actual sense is, was also mentioned. And then we proceeded that how we can draw the utility function considering the 45 degree line where UW is equal to W and so on and so forth. So we were discussing about the concept of geometric mean maximization, so geometric mean maximization if you remember I did mentioned that you want to maximize the overall return based on geometric mean and obviously we know there are 3 types of means, one is the arithmetic mean, one is the harmonic mean and one is the geometric mean and based on that will try to expand and try to utilize this concept in order to arrive that how geometric mean can be utilized for the maximization process.

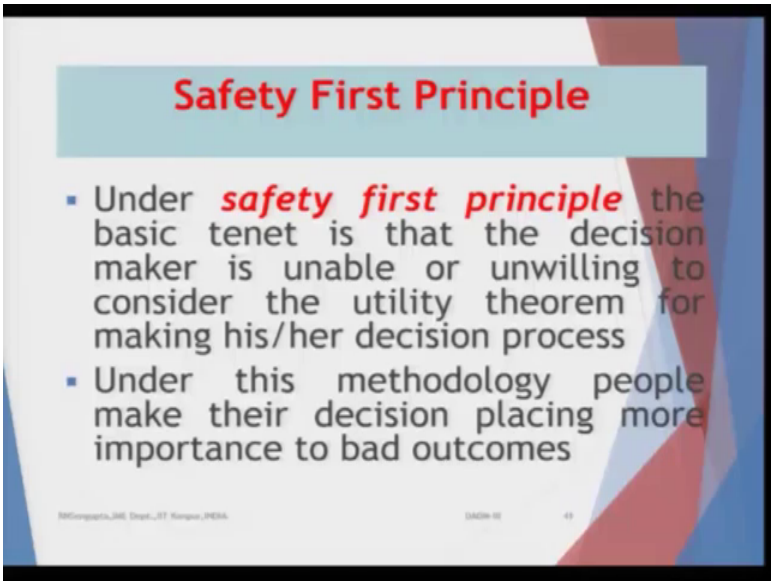
(Refer Slide Time: 03:34)



So maximizing geometric mean returns is equivalent to maximizing the expected value of log utility function. So log utility functions we have already considered is log of EW. So if you are trying to basically maximize the geometric mean it would also mean that you are trying to maximize the log utility function, so there would be one to one correspondence between them. Now for any projects, investments, portfolios or decision making which basically maximizes the geometric mean also are also mean variance efficient if returns are log normally distributed.

So if the returns of the decisions are log normally distributed then you be rest assured that the mean variance efficient concept would give you the portfolio which would also be obtained when you are considering the maximization of the geometric mean concept for that portfolio, set up portfolios. Now we will basically go through safety first principle that means how safe you are, you want to be considering the basic tenants that is decision maker is unable to or unwilling to consider the utility theorem for making his or her decision process, so you want to be basically play it safe and here I will try to draw that there are obviously diagram so I will try to draw the diagrams in order to make things much clear

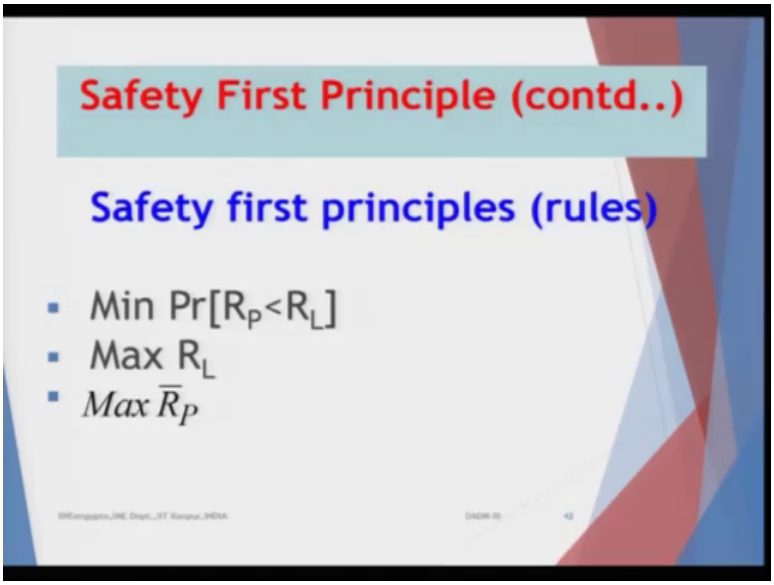
(Refer Slide Time: 04:52)



So under this methodology that means safety first principle people make their decision placing more importance to bad outcomes that means I am more what you say more sensitivity to risk being on the negative side than being risk on the positive side. That means or let me put our decision I loose in one case 100 rupees and in another case in the same so called gamble I win 100 rupees. Now my liking or my good fielding for a getting an extra 100 rupees would be there.

So let me quantify it say for example on a level of 1 to 10, let me say that I am on a positive level of plus 5 when I win 100 rupees. Now consider on the other sense that if I lose 100 rupees, so technically that is equally penalized than my loss or my level of dissatisfaction should also be minus 5 on the other side. But you will here will consider that loosing would make us more unhappy that gaining 500 or 100 whatever I am saying, so in that case the level of dissatisfaction when I lose that money would basically be more than 5, in more than 5 in the quantum sense, in the negative sense it would may be minus 6, minus 7, minus 8 whatever it is that means I am putting more emphasis on the losses in the quantum sense.

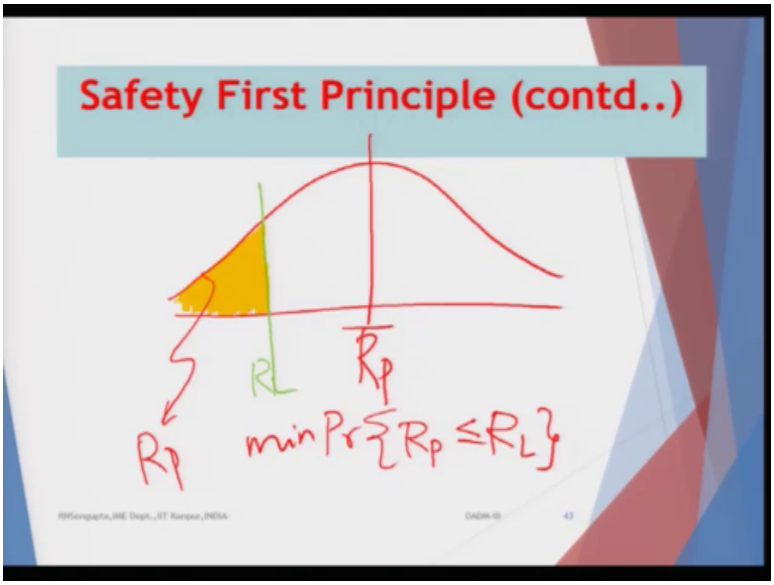
(Refer Slide Time: 06:21)



So safety first principle can be basically be analyzed by this 3 formulas and if you look at them they have the essence of optimization in a very simple way. I will try to come on that later on but it would be repetition now it may not make any significance but it is basically to inculcate the interest and later on we will see that they would make sense. So let me draw the diagram, so go one by one. So I will define what they mean, so let we draw it. So the first one is minimization of probability of RP being less than RL, so what is RP and RL? So (P) word P, I am using is for a portfolio and R capital are is the return, so R suffix P is basically the portfolio returns of the collective decisions returns.

Now what is RL? RL is a fixed value which you have set for myself, so what I set for myself may not be true for you or for the third person because the reason is that my level of appetite of risk would be different from person to person. So hence when I am considering RL value that would be higher or lower depending on the level of appetite one has for risk.

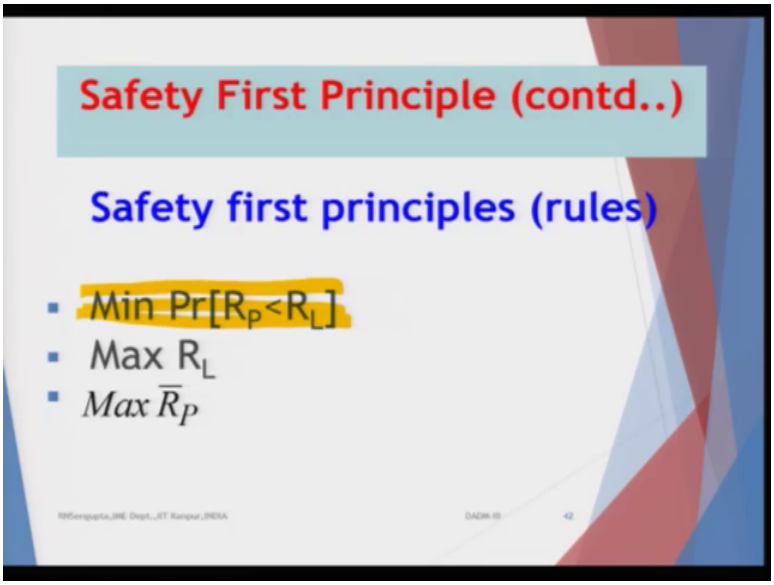
(Refer Slide Time: 07:37)



So let me draw it considering the normal distribution to be true. So this is the RP bar which is the average return of the portfolio, this is the distribution of RP. Now what I mean is probability of RP is less than RL and we will try to minimize it, so there is reason also for why I am trying to basically put the minimization sign. Now consider RL, so this is RL on the left hand side, so I want to basically minimize the probability that RL would basically be lying RP would be lying less than RL. So this hashed area is basically the area for which (RL) RP is less than RL and I want to minimize it which means that technically in a sense I am trying to push the overall distribution which is normal more on to the right that means this red part the distribution which you see would be moving more on to the right from my side.

Technically it means that if it is a normal distribution then the average of RP is increasing that means RP bar is increasing and going more to the right which means that if I am more happy with the higher values of RP bar on an average it also means that the overall area on the probabilistic sense for some values of RP to lie on to the left of RL would slowly start decreasing that means the area will start shrinking.

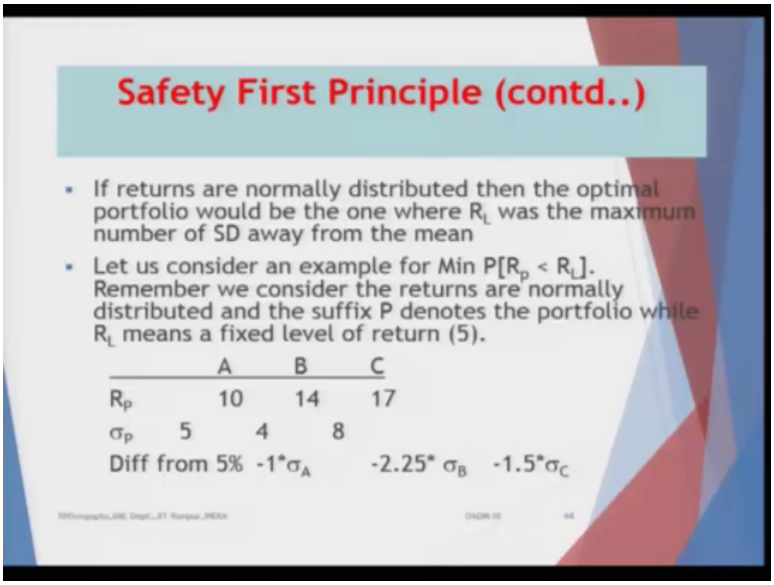
(Refer Slide Time: 10:00)



So this would mean that RP is less than RL, so RP on the left hand side should shrink. Now considering maximizing RL, so consider RL is fixed at some value and my distribution is also given so what it says that I am trying to maximize RL but it does not mean that I am going to shift RL as it is on to the right without changing the distribution because the overall area on to the left or the value over which I would like the investment to investment to be which is RL as it starts increasing obviously my mine the number of assets which I have on the decision variables based on which I am trying to make a decision is there will also start moving on to the right.

So in general the overall distribution will start moving on to the right which would mean the value of RP bar would start increasing accordingly. So the first value also give us the concept the RP bar is increasing, next one also give us the information that RL maximization will also lead to the fact there RP bar is increasing and the last one is maximizing average value of RP bar which technically means the distribution by itself will be moving on to the right.

(Refer Slide Time: 11:38)



So if returns are normally distributed, so in this continuation then the optimum portfolio would be the one where RL was the maximum number of standard deviations away from the mean, that means how far it is from the left or the right that will (cover) and the concept of left or right would be relative considering that you are trying to draw the loss distribution or the profit distribution I will come to that later on. Let us also consider an example of minimization, probability of RP less than RL, So remember we consider the returns are normally distribution as we considered and the suffix P denotes the portfolio while RL means the fixed level of return which for our case is given as 5 percentage or whatever it is.

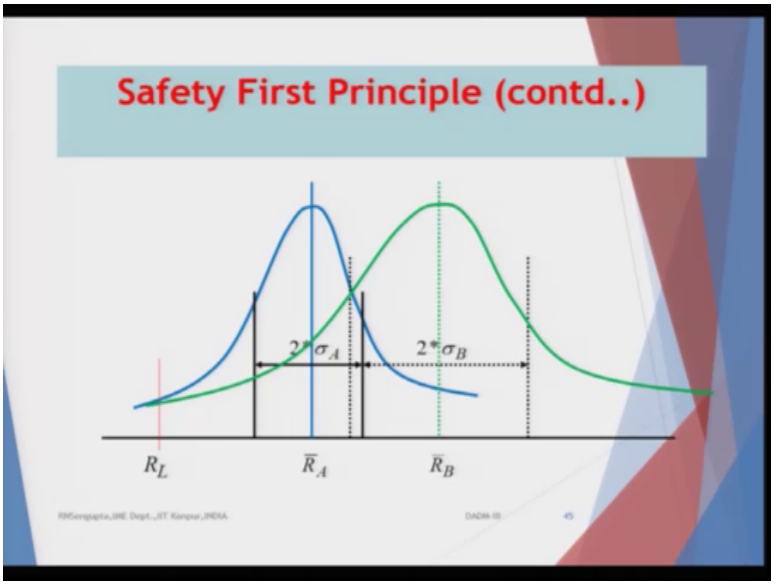
Now if that is the case, the values of RL for A, B, C RP for the 3 cases 10, 14 and 17 so sigma values so called A, B, C investments for the corresponding standard deviation of the portfolio values would be 5, 4, 8 so it would mean that once you find out the differences 5 percent or whatever percentage is there. So you need to out that 5 percent of that would be how many times further away from 1 sigma because 1 sigma would basically cover about 67 percentage of the overall area from minus 1 sigma to plus 1 sigma.

So in this case when we put these values it comes out to be minus 2.25 sigma B ok. Now when I am considering the values of B, A and C as the investments I want to basically find out that percentage difference from the 5 percent which I have already stipulated as given in the fixed level of return mean fixed level of return. So it will mean that the value would be to, so the first

case would minus 1 to multiplied by the sigma A value that means sigma A value has to be basically shifted on to the left by 1 unit in order that it is able to meet the criteria as it is given in the problem.

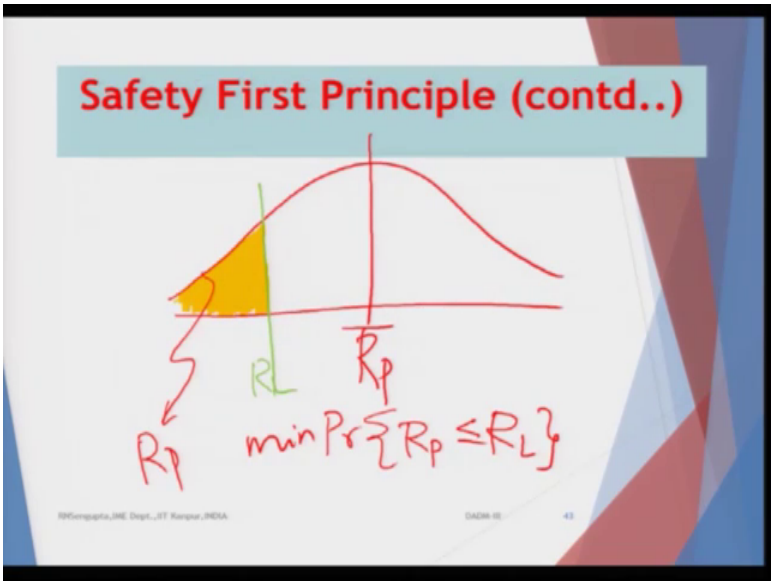
So now when I consider sigma B, the value is 2.25 more on to the left and when I use sigma C it is minus 1.5e to the left. Now what is important to know is that, this values of minus 1, minus 2.25 and minus 1.5 have been build up considering the standard deviation on the mean value for the investment. So that would become clear and this is very true when you are considering the normal distribution.

(Refer Slide Time: 14:32)



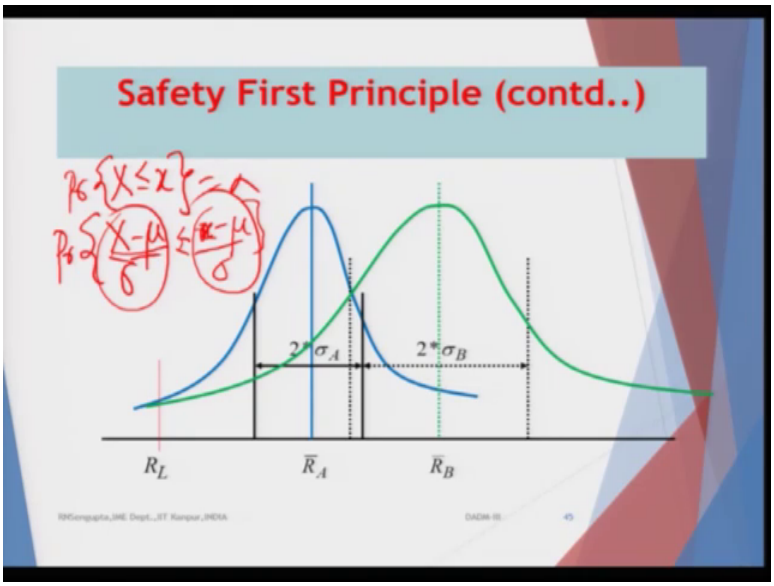
So here is the normal distribution. So if you remember if the diagram which I have drawn I will just quickly go there.

(Refer Slide Time: 14:40)



So here the diagrams are all normal in nature that means distributionally they are normal in nature but in general practical senses we would not consider them to be normal.

(Refer Slide Time: 14:54)

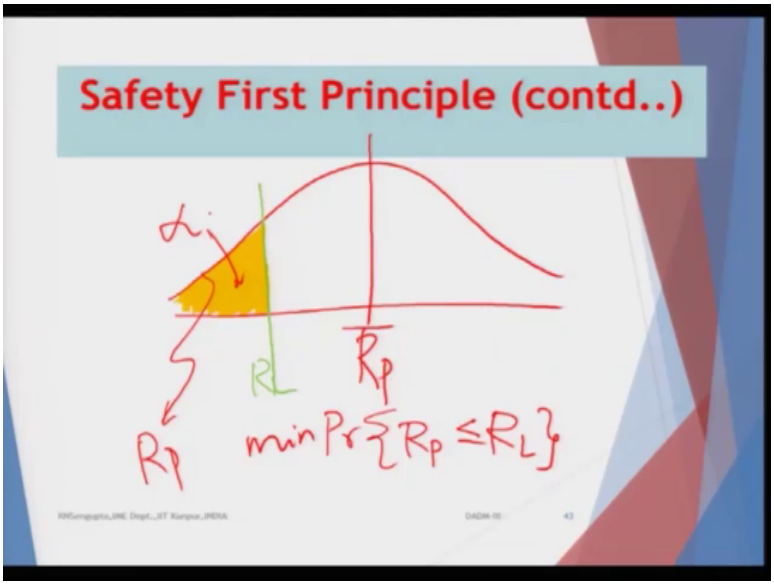


So this is the safety first principles two diagrams, the green one is basically for that decision B hence the mean value which you see, so mean value if you see for RB is the straight line green one and the corresponding values of decision A which is also the mean value is given by R bar A and the mean value has been drawn here in order to meet one understand that what is dispersion or deviation on to the right or the left for the decision A.

Now remember one thing, the more it shrinks the distribution more it shrinks even if the mean value is the lower or higher on the sum its lower, then the effect in trying to find for how many standard deviation would be further more not complicated would basically adapt more value. So, what we are doing is we are trying to find out this one so it means probability of X minus Mu value by sigma, so this Mu and sigma are basically the Mu and the sigma for the X random variable is x minus Mu by sigma.

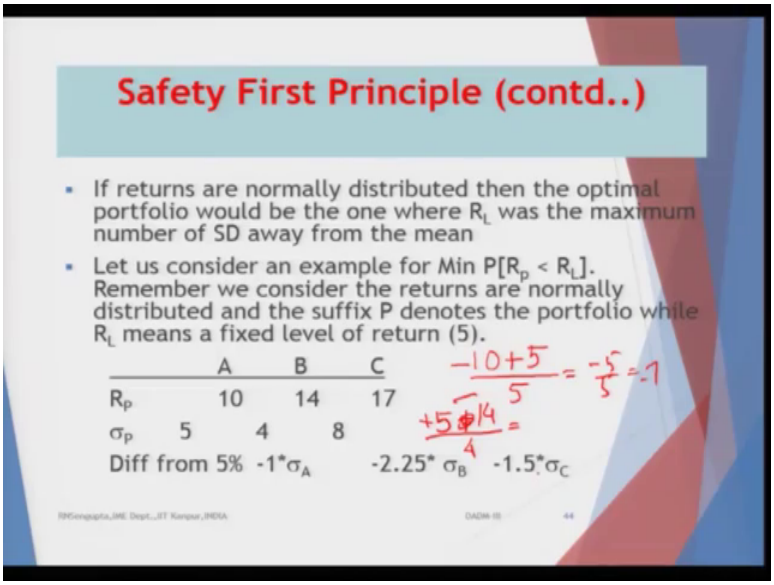
So as you are using it the value of alpha does not change, now what does this become? This becomes capital Z which is standard normal deviate random variable and this value would basically be a some fixed value why it is because you know Mu you know sigma then plugging it plug it in along the value of X would basically give you the ratio and ratio means ratio of X minus Mu on the numerator and sigma on denominator, so you can basically decide where your overall, what you call, so it will help you to find out the values of alpha, alpha means basically how many, what percentage you are there on to the left or right.

(Refer Slide Time: 17:23)



So if I go back to this diagram, so this would basically be given, the value of alpha, similarly you can do on the right hand side, so this value can be increased or decreased corresponding to the probabilities which can be found out on this problem.

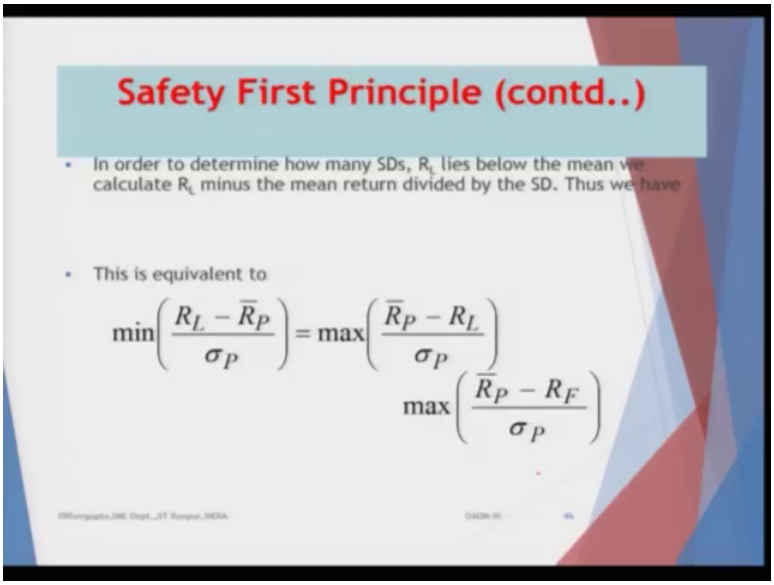
(Refer Slide Time: 17:40)



So now if I look here, in this problem why it minus 1? Is very simple, I told you it is, so it would be 10 minus this 5 value which is coming from the mean, so RL value, so its minus 1 because, this minus, this is plus it will be 5 minus 10 by 5, so this becomes minus 5 by 5 is equal to minus 1. So similarly if I want to find out this, it would be minus 5 plus 14 divided by 8, so this would basically become minus 9 by no sorry-sorry my mistake, my mistake. So this is going with, so this will be minus 5, 14, this will be 4, so let us find out what this value, this would be minus, so this would be plus, this is minus, so this will be minus 9 by 4, 4 into 2, 8 and then obviously the value comes out to 2.25, so when I am talking about minus 1.25, so it will basically be, from 5 minus 17 divided by 8.

So that will give you, what is the standard deviation on to the left or the right I should go. So,if it is very wide, dispersed, so obviously in that case the denominator nearer to the value would dictate that how the strong or weak the dispersion is how spread out they are.

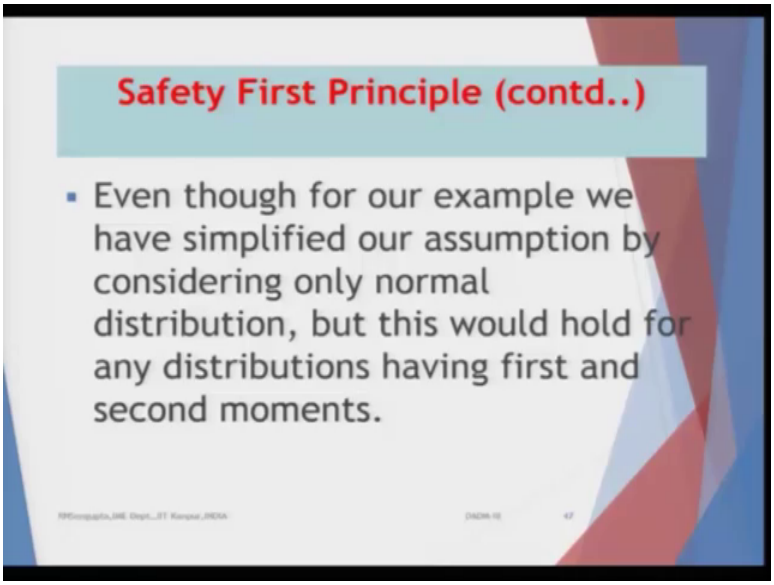
(Refer Slide Time: 19:42)



Now in order to find out how many standard deviations RL lies below the mean, we calculate the value of RL minus the mean return divided by the standard deviation that is what exactly what I did. So this is basically equivalent to you to check this, ok, by this value. This is RP bar which is the average value and technically that should be coming out in the discussions later on.

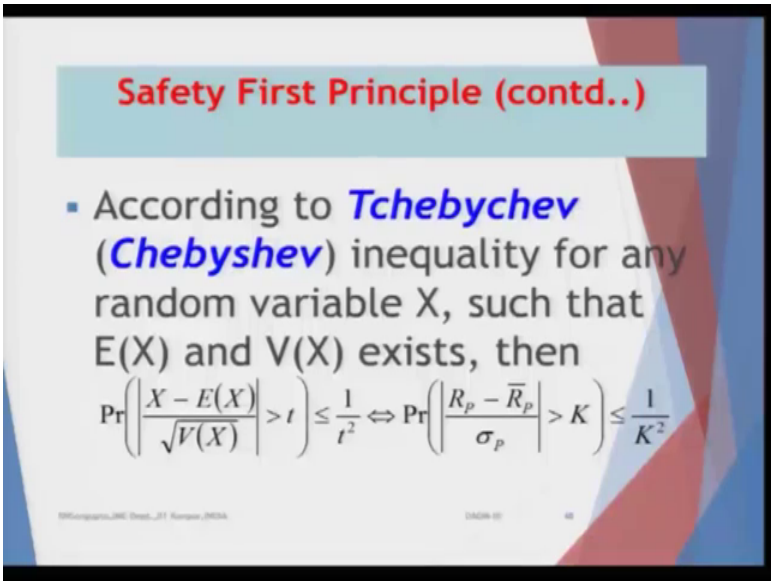
So minimization of that will be maximization on negative of that, which means I basically change the signs in the numerator and then basically what you will find out is that, this RL value can be replaced by the so called risk free interest, that means the data based on which you are trying to do the investment for any financial institutions whatever it is or the bank and the value which you will get from here would tell you how far or how close it is to the what I should say, how far or how close it is to the values of overall loss which you are going to sustain.

(Refer Slide Time: 20:52)



So even though for an example, we have simplified our assumption by considering the normal, this is very important point for you to understand, so even though we have considered the normal distribution, so but this concept would hold true for any distribution provided the first and the second moment exists. Why the first moment exists? Because you are in that case what you are taking? You are taking RL or RF minus the mean value, so that means the first moment would give you the information on the mean value and similarly when you divide by sigma, which means the second moment should also exist.

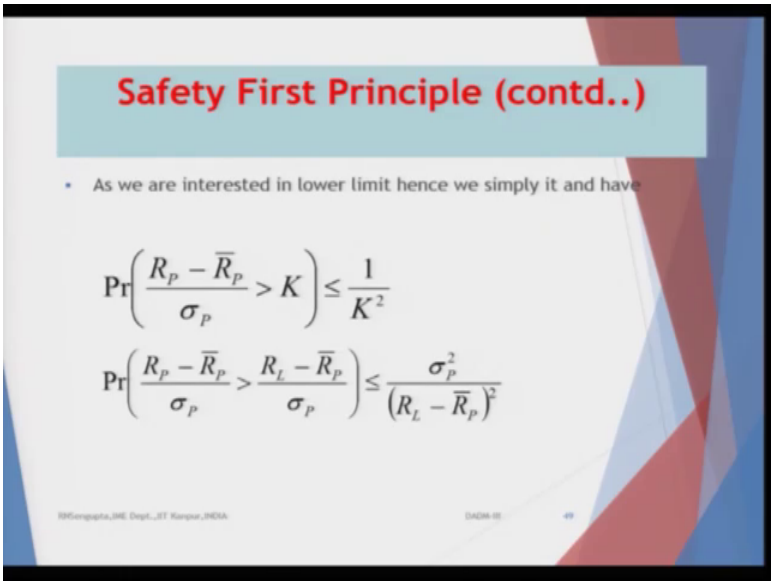
(Refer Slide Time: 21:32)



Now according to Tchebychev's (Chebyshev) Inequality for any random variable X, such that expected variable and variance exists that means the first and the second moment exists. Then the mod of the difference between so now we are basically going into a bandwidth, so shrink it larger such that the constraint side basically changes drastically, hence your space based on means you will do your search would also be minimized.

So according to Tchebychev's (Chebyshev) Inequality for any random variable X, such that expected value and variance increase, which would mean the probability of the ratio of the mod of the difference of X minus expected value of X divided by variance should basically be greater than t and this t (you) based on a simple assumptions you have to find it out. So basically this becomes I replace X by Rp and obviously of X by Rp bar and the variance now becomes, variance is basically sigma square, so when once you tick out of the square root, this value is greater than K, K value which you will be basically find out as it is. So, it is less than equal to 1 by t square, so t is basically the value of the, of so called Tchebychev's (Chebyshev) Inequality which we will take in order to solve on problems accordingly. It is basically a type of bound.

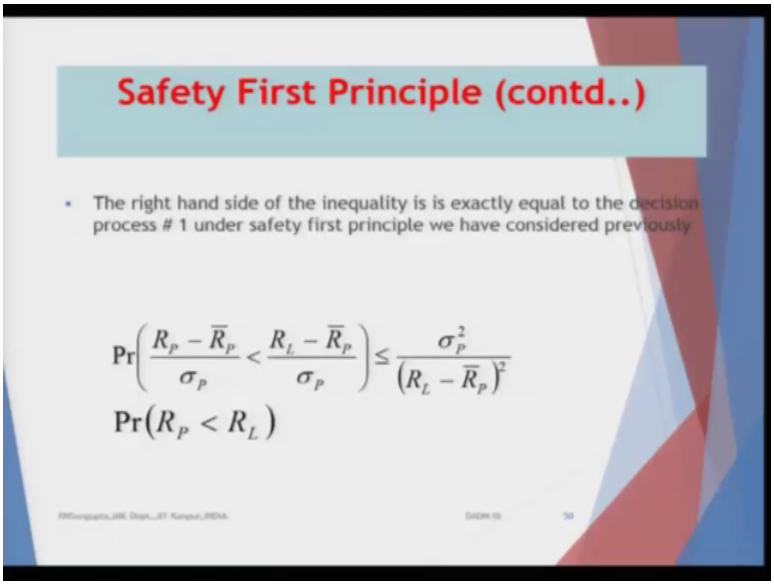
(Refer Slide Time: 22:54)



So as we are interested in lower limit, hence we simpler way simplify it and write it down, so once I basically write it down, it comes out to be true that, ok. Now remember the probability of the excess return over the average value divided by standard deviation is greater than some K value, so replace the K value with a quantity which you know and then basically corresponding to that, the K square value also changes, so what you are trying to do is that, using Tchebychev's Inequality you are trying to basically find out the maximum bound.

So in optimization problem when you find it, if you are able to solve the problem, then you know that the bounds would give you the maximum limit till which you can, are able to find the solutions.

(Refer Slide Time: 23:43)

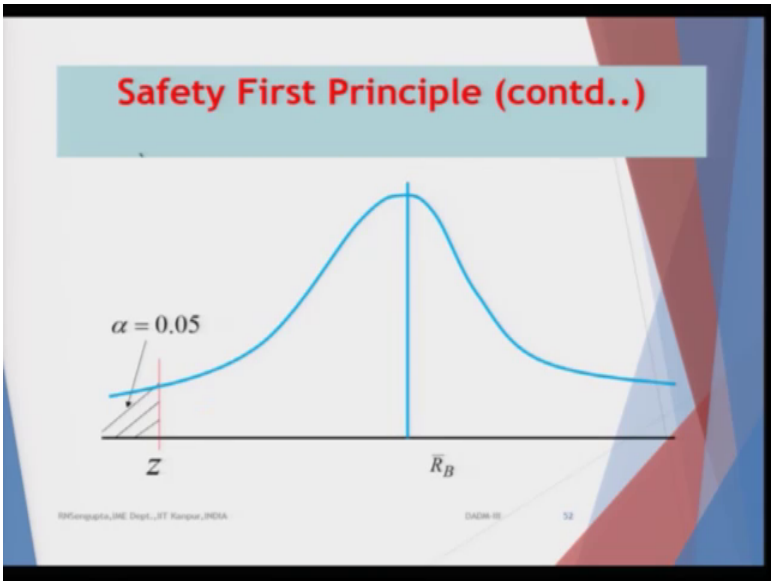


The right hand side of the inequalities (in) is exactly equal to the decision process number 1 under safety first principle, we have considered, so once we consider probability Rp minus Rp bar divided by sigma is less than RL minus Rp bar, so that divide by sigma, so that will give me the value of the K or the bound maximum to which level this value can change.

Now coming to the next criteria, so the next criteria basically means that we want to basically maximize Rp, so pushing Rp would mean the distribution by itself is also going to the right, which means the mean value Rp bar is also increasing. But as I am increasing the change of the values of the of pushing RL to the right, hence the distribution also changes, would basically mean that the probability of the value which is onto the left hand side from my side based on the area which is left hand side from RL should also be minimized or should also be basically given a certain value, such that putting that in the constraint would be able to solve you and give you some answer realistically such that you are able to understand and solve it accordingly.

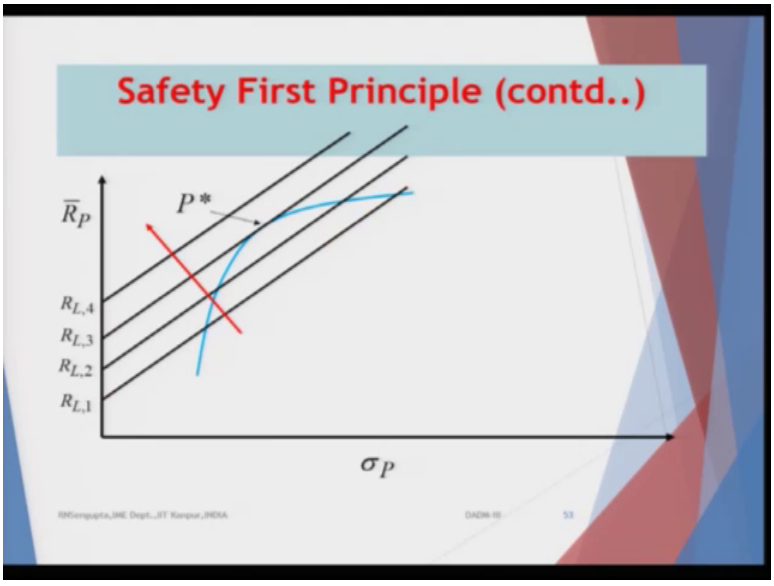
So if I put alpha as 0.05 then we should have the values as shown here, Rp minus Rp bar by sigma p is less than equal to RL and Rp divided by sigma p, so that will give you the bound based on which you can do the shifting of RL.

(Refer Slide Time: 25:25)



So in this graph what we are looking at, we are finding out the value of alpha, so alpha can change, it can go from 0.05 to 0.10,0.20 and so on and so forth. But obviously it will be in that as alpha increases hence the Z value which is RL or Rp value will also start moving hence the distribution by itself will be moving.

(Refer Slide Time: 25:54)



Now in the safety first principle, if you continue, so basically it would mean that for the case when you are considering moving of the efficient frontier it would mean that the values would increase such that the combine value of your decision variables such that you are able to

maximize Rp, Rp bar because Rp maximization would also mean the average value also shifts and to the right from my side and sigma p would happen in such a way that we are able to find the maximum bound. So with this, I will end this 9th lecture and continue the 10th lecture with some basic fundamentals of utility-3 in the last stage and then basically continue with more definitions of optimization or in the part of DADM-3 as required. Have a nice day and thank you very much.