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

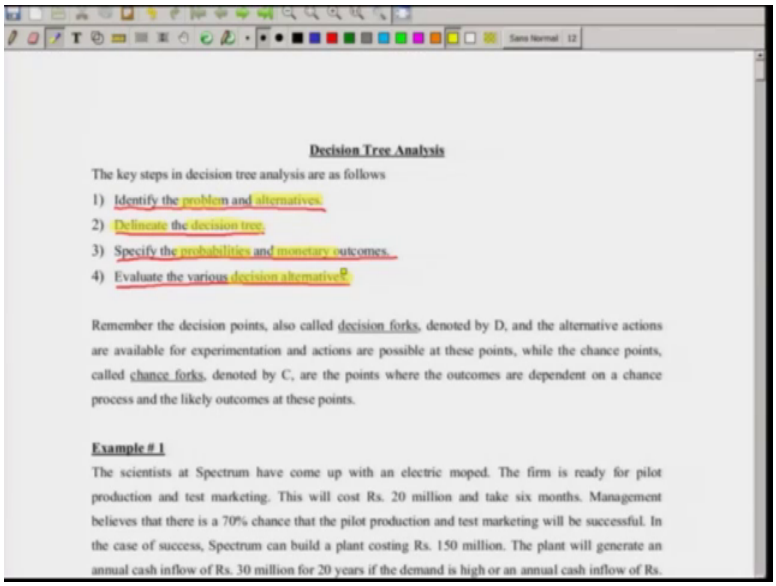
Lecture – 18
Decision Trees

Very good morning, good afternoon, good evening to all my dear friends and students, welcome back to this DADM II lecture under the NPTEL MOOC series, which is the Data Analysis and Decision Making II. And as you know we are discussing the non-parametric methods like d a h p, topsis, electra, magnet all these methods. And this total course duration is for 12 weeks, which is for 30 hours each week we have 5 lectures, each being for half an hour. And after each week we have assignments. So, we are in the 4th week, and we have already completed two classes in the 4th week. And we are going to the 18th lecture for this DADM 2. And my name is Raghunandan Sengupta from the IME department IIT Kanpur in India.

So, if you remember last class, which was the 17th lecture we discussed about the formulation simple formulation of DA, we consider the output oriented model, we consider the input oriented model. I did not discuss the radial model, I will come to that later on when we discuss how the solution (Refer Time: 01:17) should be done. I have only given the formulation if you remember in the first stage.

Now, we will discuss a very simple technique, where in my utility analysis is can be used in, in a very big way to find out the decision you are going to take. Now, this is basically known as the decision tree analysis. So, in the decision tree analysis, because all the concepts we have covered the theoretical background is already done. So, I will immediately go in to the problem solving, I will take two problem solve it, and then obviously, the assignments would be accordingly. So, hopefully I will be able to solve one problem in one day which is in the 18th lecture or if possible I will try to finish both this 18th lecture, I will try to utilise to solving both the problems.

(Refer Slide Time: 02:09)



So, what are the key steps in decision tree analysis? So, they are as noted down is basically you have to identify the problem and the alternatives. So, what is the problem? Problem is it to related to find out finding out the best decision in terms of expected value, or is it basically the decision to find out the decision, where the expected values are the same but we want to find the minimum variance, or is a decision were for all the minimums, we try to find out the maximum of them, or is it that we want to find out the, the minimum of, of the decisions such that we make the minimum loss and so on and so forth.

So, obviously it will depend what is the matrix or the criteria based on which we are taking a decision, it can be trying to be seen maximize the profit, it can be trying to minimise the profit minimise the lost my apologies. It can be say for example, try to take a combination of them. Say for example, if I find out the ratio of the mean value with respect to the variance, so what I am trying to do is the that I am trying to take the per unit return with respect to the variance, where we can consider the variance are some sort of loss, loss means some sort of money, which is going out from your pocket.

So, if we have the ratio of the expected value to the variance so will basically rank them from the maximum value to the minimum value, and take the one which has gives us the best benefit with respect to the expected value to the variance. Other way can be say for example, I am try to find out the minimum of the ratio, where the ratio is now, the ratio

of the various with the expected value such that I rank them from the minimum to the maximum and take the one, which is minimum with respect to the criteria which you have set for a our self.

Now, obviously the criterions, which we are talking about can depend on other constraints. So, other constraints if we remember for the DA, to give an example for DA, we have different constraints like say for example, the ratio of in output input is less than equal to 1, the ratio of input output is greater than equal to 1 depending on what type of oriented model you are looking at, either it is input oriented model or output oriented model.

Similarly, you will have different concept of constraints coming up in the scenario of decision trees, but our main focus would be to vary simple problems, where we want to basically find out the expected value and rank them accordingly. So, next point which is mentioned under point 1 is the alternative. So, for basically for any decision, you have defined alternatives. So, alternatives can be say for example, you want to buy a house. So, house can be you can buy in, in, in consider the city of Delhi, you can buy in house cross you can find buy in JK1 complex or JK 2 complex, you can find buy a house in Pitampura, which ever area you want to buy.

Now, the main criteria when you want to buy a house would be alternatives would be it will have different type of, of criterias based on which you want to find out the alternatives. Say for example, you want to buy a house, which is on the cost wise is the least, you will also want to buy the house on, which the safety aspect with the maximum, you want to buy a house where from the point of view after all mode of transportation is nearest to the metro station or nearest to the bus stop or nearest to the auto stop. It can be say for example, you want to buy a house, where it is nearest to the school where your kids are studying, or it can be a house where it is nearest to the market or the bazaar when do when you want to buy your things. So, they would be different criteria.

So, obviously it would may mean there trying to maximize the criterias based on the different criterias, which I mentioned like the market, the school, the safety, the price, the location considering the transportation facilities, all these things may not be possible at the same time. So, obviously, you have to rank them accordingly and find out which is the best alternative such that you find out the best combinations of the criteria, which

will give you the best alternative suiting, your budget in all respect. Budget does not mean only money, it can be all the points which I have mentioned.

Or say for example, you want to buy a car. So, these examples for the car or the house would not be considered here, I will just consider a simple decision trees of say for example, on a marketing firm and basically based on the oil exploration example. So, other example which I was talking about can be the car. Say for example, you want to buy a car, where the car colour is important for you, the price of the car is important for you, the safety of the car is important for you, what is maintenance cost of the car is important for you. May be say for example what is the boot space the luggage space of the car is important for you, or it may be how many number of passengers it carries is important for you, what is the resale value is important for you.

So, all these things when you consider they would be many conflicting once, conflicting criterias, but you are main destination is to consider that all alternative, where the criterias are combined in such a way that you get the best benefit. What is best for you may not be best for me, but you have to basically analyse from the point of view the decision maker.

Now, in the second step you basically delineate the decision tree that means, you find out the alternatives, find out the routes, routes means the way they can go. And, basically draw the diagram in such a way that a following any path we will give you one of the paths you can take to reach that alternative. So, whether is there is good or bad that you are to make a decision later on, but it will basically give you at one go the different type of paths, you can follow to, to reach the decision from the starting stage, which will consider the source to the sink stage, which basically is the alternatives decision which you are going to make.

The third point would be basically to specify the probabilities of the outcomes, because considering an decision trees, there are alternatives which will have probability. So, probabilities can be say for example, if there are three cars you want to buy, the probability may be or with respect to the, the probability of buying considering the cost factor maybe that brand x, you want to buy with probability 70 percent, brand B want to buy with probability of 20 percent, and brand c you want to buy with property 10 percent that would be based on the price. But say for example, on the aspect of safety maybe that

alternative c or c brand may give you the highest this level of safety. So obviously, the probability of safeties would definitely would be reversed where C would be higher than b, then which will be higher than a.

So, obviously, this probability is on cost, on colour, on safety and all this things, which I am mentioned may not go in hand, in hand that means they would be basically at different directions. Direction by the word direction name means that one maybe high in one case that means, the alternatives car a, b, c with respect to the cost a would be best while with respect to safety as I mentioned c would be the best. So, you have to basically make the compromise, compromise giving some quantitative values.

You specify the alternative and you also specify the monetary values. The reason why I am saying the monetary values, because it is much easier for us to understand that how monetary values, of monetary figure can be utilised in order to understand that how the decisions can be taken such that you can rank the alternatives from the best to the worst, or from the worst to the best depending on what your criteria is.

In this in the fourth stage obviously so we have delineated the decision tree, draw it, specify the probability so what is important is the probabilities and the monetary outcomes. And the next stage would be to evaluate the various decision alternatives which are there. So, say for example, for car a, b, c which I mentioned you will be seen take all the paths considering all the alternatives. And basically arrive at the all the decisions such that it will give you the best benefit. The word best I am using specific to any person who is taking the decision.

So, in this case, you should identify the problem and the alternatives that is important. You should delineate the decision tree which is important that means draw it. Then you should specify the probability on the monetary outcome which are respect to the each and every decision you are going to take. And evaluate the various decision alternatives such that you find out the best combination which will give you the best benefit.

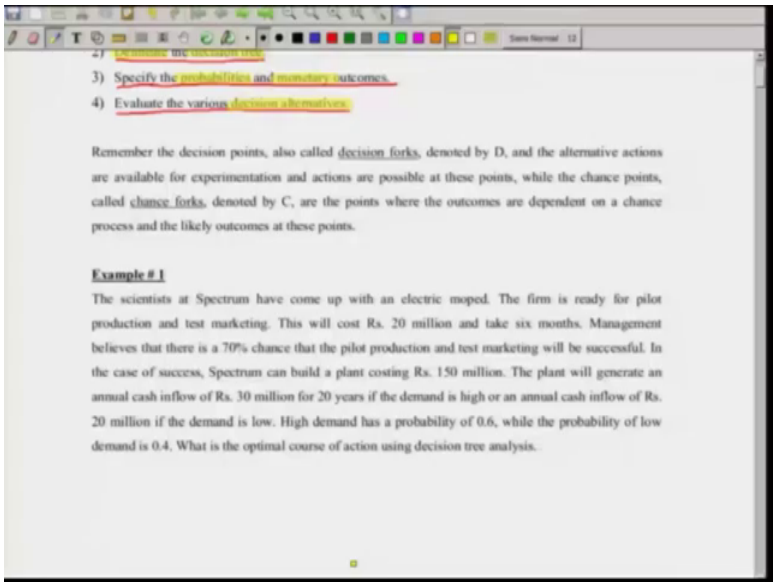
Remember the decision tree also called a decision fork tree. So, they would be forks that means there would be different arrows moving out from one decision. And this forks would basically decision points will be denoted by the capital D, while the alternative actions are available, from where alternative action are available for experimentations,

and different actions are possible at that point. While the chance point where the chance is taken are would be denoted by the symbol capital C.

Now, D would be donated by a square while C would be denoted by a circle. And obviously, they can be different force coming out from D and C accordingly and will mark the probabilities accordingly. Probabilities is basically the third point, which I mentioned remember it. While the chance point points let me continue reading it called the forks denoted by the C or the points where the outcomes are dependent on a chance process and the likely outcomes of these points would be considered in, in a in totality such that we arrive at the best decision.

So, as I said we will consider examples only. So, first let us read one example which is a simple one then we will basically make a little bit more. I would not use the word complicated more involved in the sense that the lot of information would be given, you have to you find out the best alternative such that it will give you the best decision. So, let us read the problem.

(Refer Slide Time: 11:27)



The scientist has Spectrum have come up with an electric moped, because all the concepts which we will use are already considered is very simple case of utility ulary theory application. The scientist has spectrum have come up with an electric moped. The firm is ready for pilot production and test marketing. This will cost 20 million rupees, and this monetary unit which I have mentioning rupees, dollars it is immaterial here, but

I am going to consider the values as given. So, and obviously, one question would be coming up time and again that how we are going to utilise the concept of utility, I will mention that as we proceed.

So, this will cost 20 millions and take a month a time period of 6 months which is a half year. Management believe that there is 70 percent chance that the pilot production and test marketing will be successful. So, I have already read the firm is ready for the pilot production and test marketing and that will take 20 million rupees.

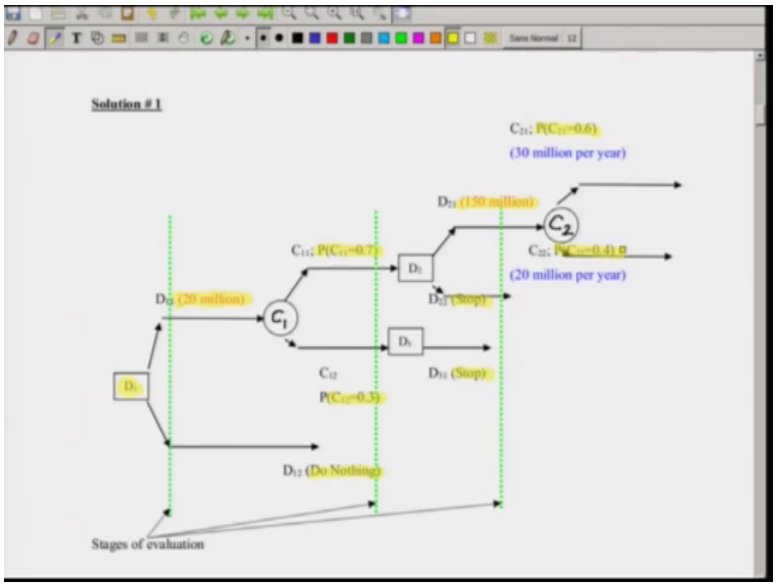
Let me continue reading it. In case of success, Spectrum the company can build a plant, which will cost 150 million rupees. The plant will generate an annual cash inflow of rupees 30 million, continuously for 20 years if the demand is high, because they have already test market read it and, and, and done a pilot production. So, it will be 30 million each year for the coming 20 years if demand is high for which the probability is 60 percent, if you read it I will count come to that later on. And if the demand is low the annual cash inflow of cash would be 20 million and it is also we are considering that a time period is 20 years.

So, 30 million 20 years, 20 million 20 years, but obviously, they would be many very simplistic assumptions which we are slowly relax and tell you the background how you can solve the problem. As I was mentioning high demand has a probability of 60 percent was a probability of low demand is 40 percent. You have to basically held the scientist spectrum to arrive using the decision tree analysis to arrive at the best alternative.

So, again I repeat the main important points. It will, it will do a pilot production and test marketing, which will cost 20 million. I am not using the, the units of rupees, dollars, whatever I am just mentioning the values 20 millions time period is 6 months. It that management believes the 70 percent chance the production and test market will be successful obviously 30 percent is not successful.

In case of success, the company can build a plant costing 150 million. And if the demand is high, it will get 30 million for the each year for the coming 20 years. And or else if the demand is low, it will get 20 millions each year for the coming 20 years. Now, the high demand probabilities is 60 percent and the low demand probability is 40 percent. So, these are the main points. And I am sure if you have listen to the carefully, you would have noted down the important facts which are important.

(Refer Slide Time: 14:39)



Now, I will go into the diagram. So, this maybe a little bit I will try to reduce it, this is the diagram. So, this is the optimum size I can show it to you. Now, consider the diagram and I will come to the values, which I have stated. So, I will go explaining the diagram from my left to the right, but when I solve the problem it from the right to the left, why, I will come to that later.

So, you have the, the green vertical lines are the stages of evaluation. So, this stage 1, stage 2, stage 3 for evaluation if you remember we said that will denote the decisions fork from the square, which will be denoted by D and the chance one would be denoted by C which is a circle. So, it has been denoted likewise. So, for the decisions stage, when you are taking the decision, which will denote it from the left as D 1 so, this is the suffix D 1 I will now highlight it, so it easier for, for us to understand. So, this is D 1.

So, in D 1 when you go there are two paths. Path one is if you remember it would basically do a test marketing for which it will invest 20 million. So, this 20 millions which you are seeing which is red in colour, red means that is the money is going out from my pocket or spectrum companies pocket. So, this D 11 would be 20 millions negative, and D 12 which is note denoted here I will do nothing that means, I will I would not start the business. I will just not want to do a carry on forward, because it may it may it is possible that if I do and invest 20 millions. And 20 millions also out from my

pocket, it may prove that in the long run the overall value of the project is negative, such that it would have been better if I have done nothing for this whole project.

On the second stage, once you have invested 20 millions to do the pilot plant and the marketing, now as I mentioned in the problem they were 70 and 30 percent chance that of probability that you should basically continue with building the main plant and continue with the business and 30 percent chance was that you do nothing. So, obviously, from the point of C, C 12 where you do nothing where you stop, you and you do not take a decision the probability is given as 30 percent, which I have already mentioned.

And if I take the upper route which means I take the C 11 route, so this the initial part was 30 percent was the C 12 route that means I am going down, and the decision would have been D 3. These are the nomenclature of the suffix of D 1, D 2, D 3, D 4 as just mention from the left hand side to the right hand side. So, but obviously, we have a rule that as you follow it from the left hand side to right hand side, we either give the numbering from the top to the bottom, or from the and obviously, from the left to the right.

So, C 11 has a probability of 70 percent which I have already in mentioned. And C 12 has a probability of 30 percent. And once 30 percent that the test for marketing dose not yield anything, I will do nothing and stop this production, which is given as D 31, because D 3 is the decision point which is shown by the square. Now, if 70 percent is true, then obviously, there can be two alternatives, number of alternatives 1 1 would be I build a plant and the cost if you remember is 150 million, and again it is red in colour, because I am investing from my pocket I means the Spectrum company.

And the other alternative means that the probability is high that means, 70 percent is true, but I do nothing that may I just waste that money, and do not continue the business, because it may be possible that if the demand is if I continue the production, it may be possible then in the long run the overall value for continue the business maybe negative. Through in this case, we do nothing, stop. Similarly, hence D 31 will stop and do nothing, stop and do nothing are the same thing.

Now, if we have invested 150 million, it may be possible, and as I mentioned that there is 60 percent, so obviously when you come to that point of C 2. So, this circle is C 2 let me right it down. So, it will be easier for us to. So, this will be C 1, this will be C 2. Now,

when we consider the high demand and the low demand, the high demand and the low demands are given by probabilities of 60 percent and 40 percent, which I have already mentioned.

And the income which is happening for high demand and low demand is respectively as this, 30 million for each year for the coming 20 years, 20 million for each year for the coming 20 years. So, now, you may ask that why it is 20 years for both the case, yes, that is true, that if is high demand it may be possible that 30 million is being is the income we are getting each year for maybe for the coming 20 years. And for the low demand, it may be possible that we get 20 millions per year for the coming say for example 15 years, but I am just taking not very simplistic assumption. So, with this decision, we need to basically find out the best alternative.

Now, there is a concept which is known in optimisation at the stagecoach problem. So, in stagecoach problem a person is going from one east coast to the west coast or west coast to the east coast whichever way you are analysing. So, it a starts from say for example, from area of New Jersey or in the east coast, and then goes to California, and it is following different routes, stagecoach mean this horse drawn carriages.

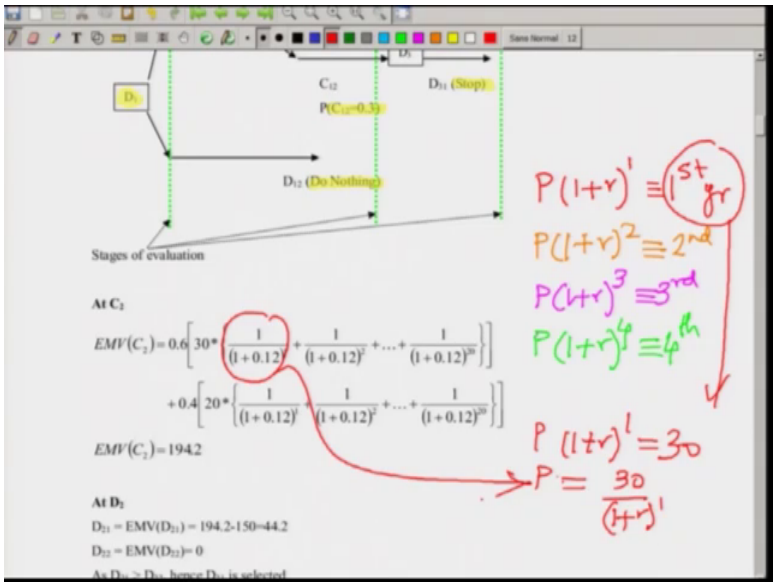
Now, for each routes, they would be cost for travel per unit, they would we say for example, safety issues, they can be waiting time so on and so forth. So, all these things for positive negative inputs are taking in order to arrive at a decision. So, obviously, there are different forks, forks or different routes the stagecoach can take to start from the source, which is the east coast and reach the sink which is the west coast.

Now, when you are trying to analyse the problem, best would be that you start from the end stage and work backwards to find out that if the out alternative, we have taken was say for example, either path 1 or path 2 or path 3, so which paths you would have taken in order to reach path 1, path 2, path 3 that means, you going on stage at a time going backwards, and then slowly trying to find out the best combination in the paths such that you are able to eliminate all the other paths which would not give you the same benefit or the same maximum benefit. So, I will when I solve the problem, this will become creative. So, will start from basically from the right and that is the sink, and proceed on to the left and basically reach the source in order to find out the best alternate.

So, I will pause for 1 minute, have a look, we are starting from D 1 20 million we are investing, and then there is a probability of good business, bad business, 70 percent, 30 percent. 30 percent we do not do, 70 percent we do that means again we invest 150 million to build the actual plant. High demand, low demand being 60 percent, and when what wait 1 minute 60 percent and 40 percent, amount of money coming is 30 million for 20 years and per year 30 million; and another case being 20 millions per year for 20 years.

And do nothing means basically we do not either go for the project or, or stop the project is that means, we just stop at that stage. So, you have to basically balance all the alternatives with, with amongst each others, so arrived at the best decision which is possible.

(Refer Slide Time: 22:41)



Now, let us go one by one. So, we if we go from the right, so what will do is that we try to find out the expected value of C 2, and then work backwards. So, let us find out the expected value of C 2. So, what is the expected value of C 2? Now, at C 2 which is the rightmost circle which is there, there are probabilities 60 percent, 40 percent. So, 60 percent would be multiplied by the net present value of the 30 million which you are getting each year for the coming 20 years plus the 40 percent of the net present value of the 20 million which you are getting each year for the coming 20 years. So, this is how we do the calculations.

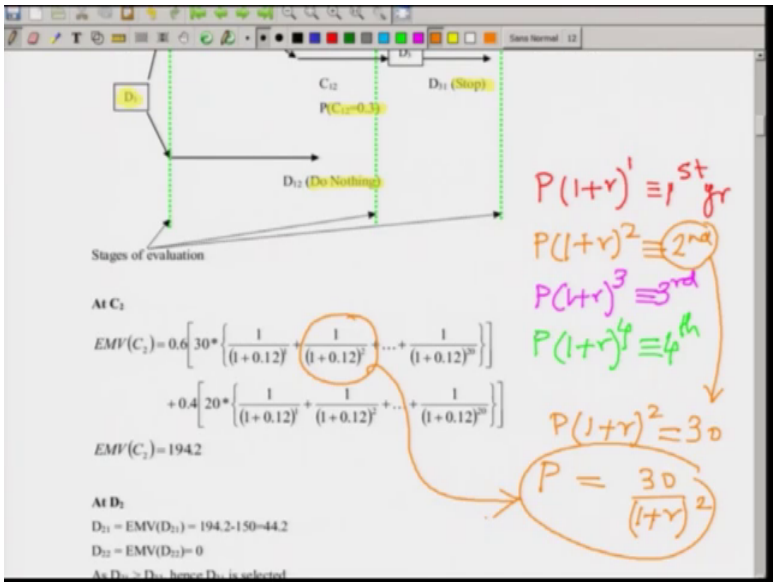
This is the probability. So, this 30 which you are doing we are multiplying by the so called net present value factors for the first year, second year, third year till the twentieth year. That means, if I had a value of P, I am doing it in red colour. Now, now means consider on first of January 2019, the interest rate was r, and I am basically investing that money for 1 year. So, at the end the year 2019 that means on 31 December 2019 or 1st January 2020, the total amount of money which I will have in my pocket would be P into 1 plus r to the power 1. This power 1 goes to the value of 1 plus r.

Say for example, if you so this would be the value after the first year. Say for example, you find out the value after the second year. So, let me use a different colour. So, it will be easy, P plus and I am not taking out the money in between. This is the value which is after the second year; is this is a value, this is a value after the third year, this is plus 1, plus. Similarly, the value which I get after the fourth year that means once I invest, I continue that money. Now, see what why where this calculation used.

Now, it is saying that I receive 30 millions after each year; that means, after the first year if I have a value of 30 million. So, when I do this calculation, so this first year value is 30 1 plus r is which is what is given, I need to find out P. So, P is given by 30 by 1 plus r to the power 1, and this is the value, which I have here.

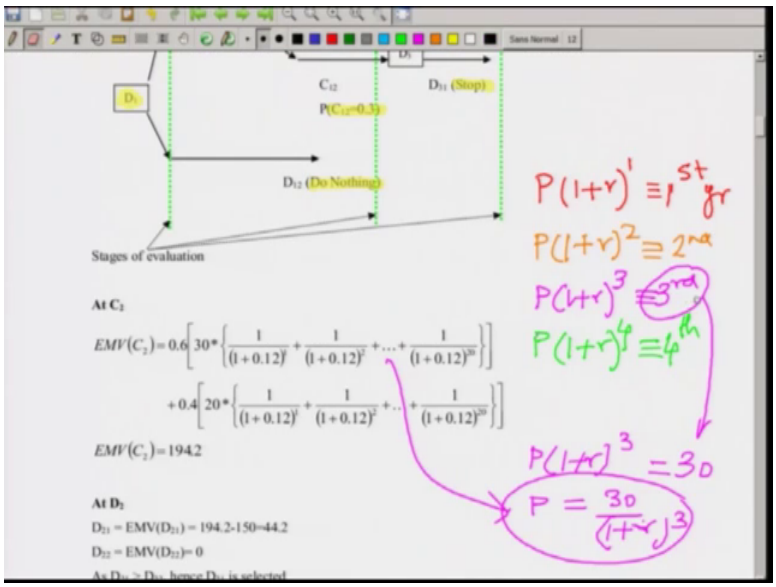
So, just look at it. Now, let us go into the second calculation for the second year. So, let me use the different colour. So, first let me erase it. This is I am going to do the calculation for the second year. So, this goes; so, this was basically first year. Now, I do the calculation for the second year. So, second year value is also after second year which is now 31st December of 2020 or 1st January 2021.

(Refer Slide Time: 26:59)



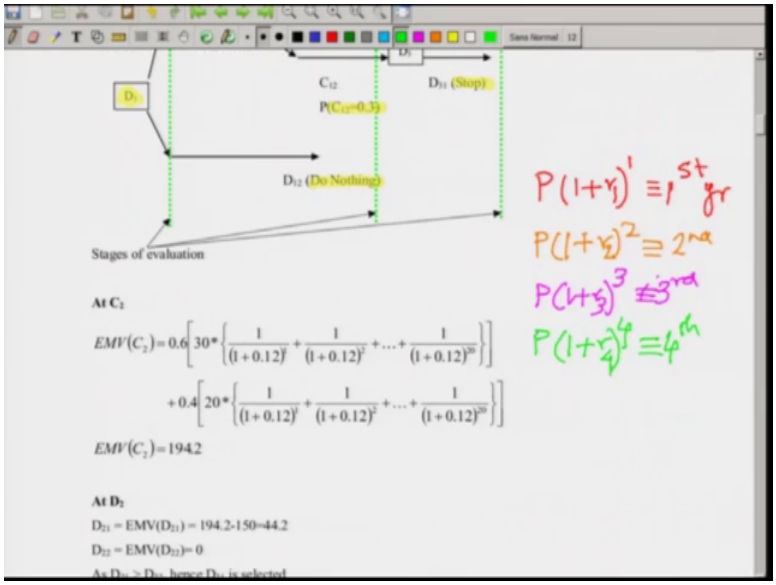
This is 30, P value, P is a variable, so obviously it will be utilized. So, when I use P, it becomes. So, this value which I have is exactly the value which I had here. So, if I come to the third year, so have a look, I am just one thing. So, you may be thinking why am I using the same value of r interested will that is a simplistic assumption interested change, so obviously, technically the r which I am using for the first year, second year, third year, fourth year can change. So, let me do the third part with the same value of r, and come to the change values which I can do.

(Refer Slide Time: 28:01)



So, third year value is again, third year now becomes 301st December 2021 or 1st January 2022 is P 1 plus r cube. So, P is equal to 30 divided by 1 plus r, this value is coming if they would be a third year, so it is not there. So, it is coming. So, obviously, the concept which you are talking about r being different if that is true be r 1, this would be r 2.

(Refer Slide Time: 28:37)



The r 1, r 2 are different. This would be r 3 suffix, and this would be r 4, and you will continue doing the problems accordingly. So, with this I will end the third lecture in the fourth week which is 18th lecture, and continue discussing more of the decision tree in the next lecture and definitely wrap it up in all the details as necessary.

Have a nice day and thank you very much.