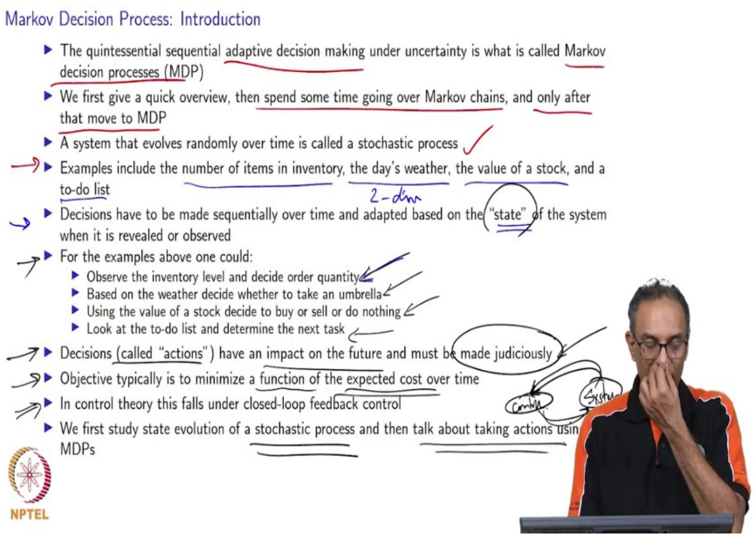
Decision Making Under Uncertainty
Prof. Natarajan Gautam
Department of Industrial and Systems Engineering
Texas A & M University, USA

Lecture - 30
Markov Chains for Decisions

This lecture is about Markov Chains for Decisions.

(Refer Slide Time: 00:21)

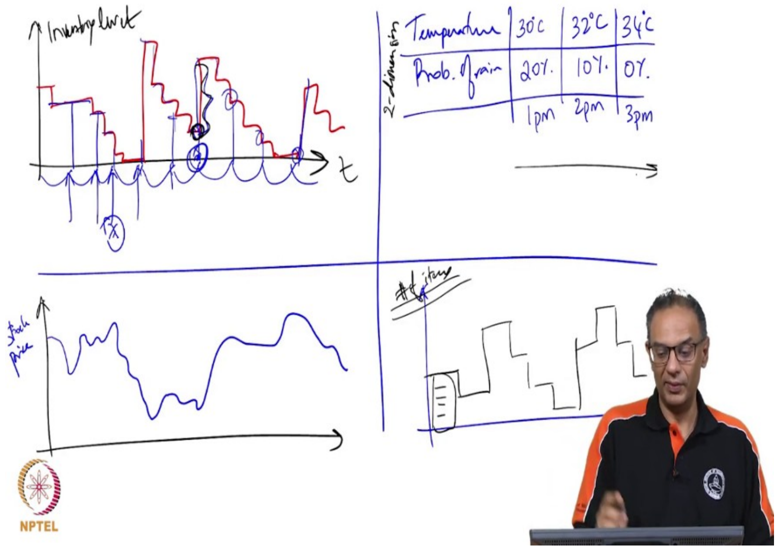


Markov decision process is the most often used as adaptive decision making tool and this is what we use for decision making under uncertainty. There are some restrictions, but we will talk about the Markov decision process as we go along this lecture. Turns out that, we first need to go and give a quick overview of what is going on and then talk a little bit about Markov chains because Markov chains is going to be the basic structural part of a Markov decision process.

So, if you do not do any controls and just let the stochastic system evolve over time, you will get a Markov chain. And once we talk about that, we will move to what is called Markov Decision Process (MDP). A stochastic process is a system that evolves randomly over time; we will give you some examples very soon.

Think about the number of items in an inventory and then the second example is the day's weather, perhaps a value of a stock and a to-do list.

(Refer Slide Time: 01:40)



So, let us draw a little picture. So, let us say the inventory level, we have seen this in the previous topic- inventory level over time. So, maybe the inventory level is here and then a new product came in and goes here and then it stays here for a while and then goes down and goes down very quickly down to 0 and then you get a little shipment that goes here; somebody bought 2.

And then 1 and goes down here and then the reorder goes here, goes down here, becomes empty and then you reorder and so on. Now, this is the way the inventory system goes about. Now, if I were to choose unlike the case we had in topic 3, where we continuously observe the inventory, which is what is happening here. Let us say: we observe the inventory at discrete time points.

Let us say we observe this once a day and then we make a recommendation of whether we want to do nothing. So, let us say this is the beginning of an end of every day. And at the end of every day, you look at your inventory level and make a decision. Now, this is different from what we had before. In the previous case- in topic three, where we were looking at a safety stock; as soon as you hit the safety stock and you know exactly when that happens because you are observing it continuously. As soon as you hit the safety stock, you place an order and then there is a lead time.

Here we will ignore the lead time, when you could incorporate. Because model becomes a little bit harder, you will ignore the lead time for now and all we are going to be worried about is you know we are looking at the system once in a while and make a decision of should we place an order and if yes how much should we order up to? So, for example, the first time period, it will come here, you see how the inventory level is pretty high, I will not do anything. I know I will come back, then end of the next day, I go here I see it looks.

I go here, I see, it looks ok; I go here- oops! The level is 0. So, let me go out and place an order, I place an order, and then the inventory level keeps coming down and then I go here at the end of the next day. I see it is here as I do nothing, then I go here and then I see: well, last time I did poorly by not making the decision here to go ahead and buy. This time I am going to be a bit careful and I will go ahead and I will buy some and then when it arrives, it goes here and the next time I do nothing. I again get a little bit complacent I do nothing here; all of a sudden I hit 0 and then I buy.

So, you make a decision over time about whether to buy something or not. Now, so that is typically how an inventory level over time, this is a stochastic process because these times are random and even sometimes if you even consider the lead times and that also is random.

The second example that we are going to be talking about is if you look at the temperature on some of your phones- temperature and probability of rain, you will see you know a random process that will go like this. It will tell you the temperature right now is 30° Celsius and the probability of rain is 20% and then it will tell you this is right now. This is let us say 1 PM and then at 2 PM the temperature is 32°C and the probability of rain goes with 10% and then it goes to say 34°C. Probability of rain is 0% at 3 PM and so on, every hour it gives you a forecast of the temperature and the probability of rain.

This is a random process because the temperature changes randomly and the probability of rain also changes randomly over time. However, notice that this is 2 dimensional- there are 2 dimensions here and they are not independent. The temperature is probably correlated with the probability of rain. I am not saying there is perfect correlation, but there is a relation. And then if you see the third example that we see here is the example of looking at the value of a stock and then see what we need to do.

So, if you model the stock price of a particular stock, the stock price could possibly be here and then go down and up. I am only am drawing like its continuous, but I do know that it kind of moves discretely over time, but for all practical purposes it is kind of continuous. So, this is the stock price over time, it could be a quantity that goes like this and then lastly the example is well I should probably not model it this way. So, let us say you have a “to-do list” a bunch of items here. So, this is time and then this is your number of items in the list.

So, I have four items in my “to-do list”, I get done with one; I come down to three and I add two more to my list and then I come down by one, I keep doing stuff, then I get added two more items, I add one more item. So, I keep adding items into my “to-do list”. So, my “to-do list” has a bunch of items that I keep increasing and decreasing over time as they keep completing the list, I keep going down to one and as I keep adding I go up by one. As we all know the “to-do list” will never have 0 items- yes that is a joke, but that is in reality though. Now, let us go back to what we were looking at. So, we were looking at these example.

So, to remind you- we looked at the number of items in inventory over time, the day’s weather, the value of a stock, and the “to-do list”. So, the number of items in inventory in the example that we saw, we actually observed it continuously and I said what if you observed at discrete time, which is what we will do at towards the later part of this lecture. You look at daily weather, which had 2 dimensions. So, there are 2 dimensions, we looked at some forecast into the future. We look at a value of just of a particular stock- it went up and down.

And we were also looking at “to-do lists”, where you are adding items to it. You could even think of the entire “to-do list” as a stochastic process as supposed to looking at each item and number of items to list, you could look at each item in the list and that is completely up to you; some things are easily analyzable, some things are difficult.

Now, we also need to make decisions. We were basically except for the inventory example, all the other 3 examples- I was only telling you something about how the system is going to evolve over time randomly and that is called a stochastic process.

So, even in the inventory case, you know we are probably going to be interested in situation, where let us say: you observe and then you take some actions and that yields a certain pattern. Now, I am going to talk about that in a little bit, but you could think of a whole bunch of decisions. We talked about this decision, we are saying: we observe the inventory level and then decide how much to order. So, you look at it here and then you are saying: ok, I am here, how much do I want to order, I want to place an order for these many items, when I was here.

So, at every stage whether or not to get anything and if yes how much to order. So, when to order and how much to order. Now, the key difference; one more time it is very important to understand the difference between this and earlier. Earlier, we were continuously observing, here we are observing at discrete times called periodic review and continuous review is what we saw earlier. In the temperature case, you might decide whether or not to take an umbrella for the day. You look at this and say: ok, this is my process, well it looks like a low chance of rain. I look outside- ah! It is not raining, I won't take an umbrella. On some other days you might say: I will take an umbrella even though the probability of rain is only 20 percent.

So, that depends when you know you are wearing a nice piece of suit and pants and you are thinking maybe I should take an umbrella. So, it depends on your state information as well as what is going on outside. Now, in the stock price situation you can decide whether or not to buy a stock or sell a stock. When the stock prices are high, people typically buy and the stock prices are low they sell, which a lot of people say is not the correct thing to do. You should actually be selling when the price is high and buying when the price is low.

So, that you can make a huge profit. But, you know in practice what happens is that generally does not. In terms of "to do list", you can decide which of my items the "to-do list" should I work on next and that could be a decision assuming that you are not very good at multitasking, we only do one thing at a time which by the way is what I would recommend; is to figure out which one do I work on next. So, these are some of the decisions like I said- when to order and how much to order or whether to order and how much to order and whether to take an umbrella or not, in the rain case or whether to buy or sell or do nothing to a stock right?

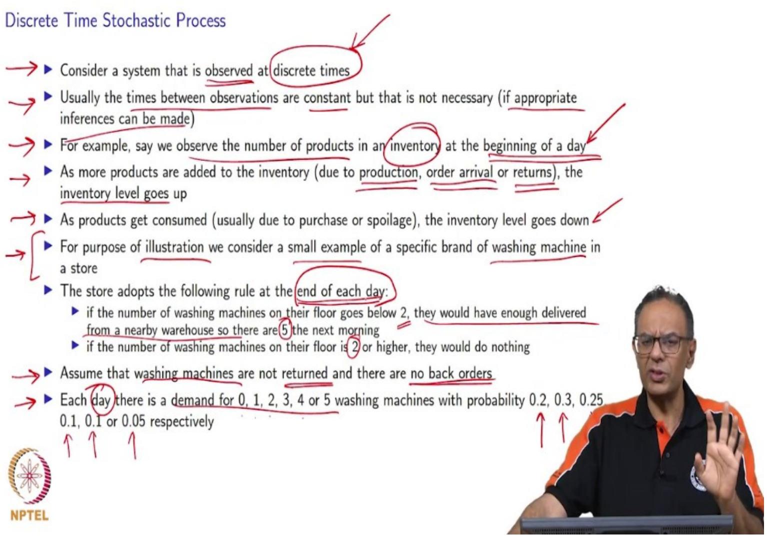
You do not have to do anything either at some time and you have to look at the "to-do list" and see what to do next. So, these are examples of actions. So, the state is evolving randomly over time. So, that is one word we will see a lot in this lecture. And decisions are what are called actions. So, you take an action and that decision could be impactful for not just now, but also into the future. So, that is an important aspect. So, when we model these things, we need to be concerned about not just my immediate decision. So, you should not be very myopic. There are my needs; right now are these, but you need to think about the long run.

Let me give you an example, let us say you are tired and aged like I am and you go through what is called midlife crisis and you are thinking: oh boy, maybe I should buy a very expensive car. Now, when you buy a very expensive car, it does satisfy your immediate need.

However, you know this might be a terrible decision when it comes to your future. So, you have to make these decisions judiciously- not being too myopic, but looking at the long run.

So, the objectives typically in these situations are some type of a minimization of a function, which is usually the expected cost over time, usually. And in control theory, such problems are falling under what is called “closed loop feedback control”. So, you observe the state and then you have a controller, it looks at the state and then it performs an action. So, this is the system and it gauges the state of the system and it will decide what control action to take. So, this is commonly studied in the field of control theory. So, what we will do is, we will first study the stochastic processes then two lectures later, we will talk about taking actions.

(Refer Slide Time: 12: 57)



Now, next thing we will talk about is a discrete time stochastic process. So, we are going to observe a system. So, now we are not taking any action, we are only observing the system and we are saying: we are going to observe the system at discrete time points, so that means, every hour or every day or every minute or every nanosecond depends on how dynamic your system is. You may not observe an inventory system every nanosecond. On the other hand, you know there are systems which is no point and let us say observing the number of messages in a queue in a computer server. Let us say every 5 days, I mean that is too long for considering that dynamic.

So, you need to pick a good time frame and observe the system. Usually the times between observations are constant, but that is not a requirement. You could observe at non constant times, you just have to be very careful. When you make inferences, you have to be very careful, when you observe them at non constant time. But, more often than not, I would probably say 8 out of 10 examples that you would see, you would be observing at in a periodic fashion and that is why even the inventory example is called periodic review.

So, the example of inventory comes next. So, let us say: we observe the number of products of a specific item in the inventory at the beginning of a day. So, the observation times are specified and they are constant at the morning of every single day, I am going to figure out how many items are in inventory. We will say later as to why we observe at the beginning of a day and not at another time. For that, we will wait for a little bit. Now, here is what is important. As you add products to the inventory, this is usually because you either produce or an order arrives.

So, it is either a production inventory system, where you are producing and adding to the inventory or you are ordering and adding to the inventory. That is how usually products are added to the inventory. So you go up, when you add stuff to the inventory or sometimes even when someone returns an item. So, they come back to stores saying: I do not like this item, I would like to return it.

So, that is another way that the inventory level goes up. Then as the items get consumed, the inventory level goes down because no reason for you to believe that inventory level goes down one by one like I had shown in the picture, you could have a situation where somebody comes and buys multiple of an item.

So, the inventory level goes down whenever someone comes and buys. Now, we will take a very specific example and this example is just for illustration. We are going to look at a small one, where we have a washing machine brand in a store. So, there is a specific brand and a specific type of washing machine in a store. The store adopts the following policy or rule. This is their rule: if the number of washing machines on the floor is below two; by the way they are observing it on the night or at the end of a day.

So, at the end of each day, the store manager walks around and then comes to this particular brand of washing machine, looks at the floor and sees how many washing machines they have that are unsold. If that number is below 2, then they will go ahead and call their warehouse and they will make sure that the order enough, so that there are 5 of them in the morning of the next day; they can make a call see that there are 5 in the next day. If the number of washing machines in the floor is 2 or more, they would do nothing. So, number of washing machines is 2 or more, they would do nothing.

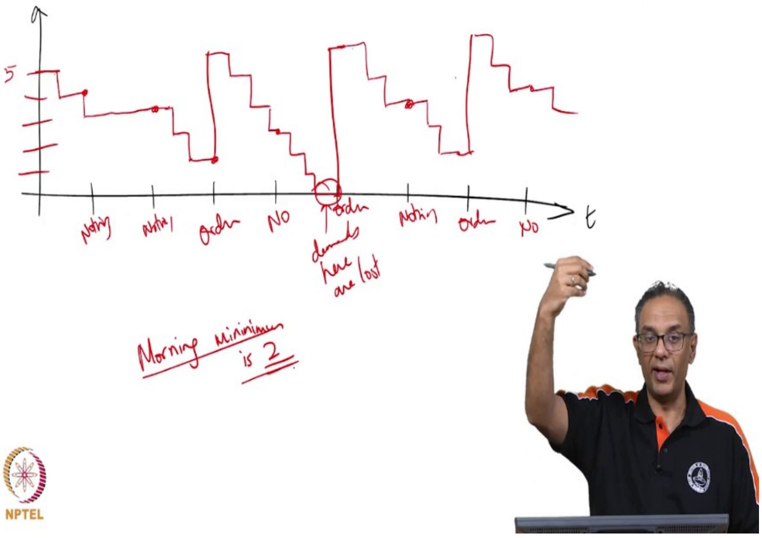
So, let me just repeat what I just said. So, the number of washing machines is below 2, they would order, they would order enough so that they have 5. Essentially there are two possible numbers, it is either 1 or 0. If it is 1 they will order 4, so that the next morning they have 5. If it is 0, they will order 5. So, the next morning there is 5. So, those are the only two things that is done.

Now, we are going to make an assumption to make or model tractable although in practice this is not something that is required. We are going to assume that the washing machines are not returned. Somebody comes and buys the washing machine typically keeps it and there are also no back orders. By that what we mean is let us say there are 0 washing machine in the floor.

If a customer comes in and asks for a washing machine, we are not going to allow back orders, that means, they pay for the washing machine and then they come and pick it up when they get one. That is not going to happen, this is just again for modeling convenience. I am not saying that this never happens in real life, but in a washing machine situation, it is not a bad assumption.

Now, this part is important and I will say this multiple times: the demand for washing machines are 0 or 1 or 2 or 3 or 4 or 5 in each day. So, maximum of 5 washing machines are demanded each day and a minimum of 0 washing machines are demanded. The probabilities respectively are 0.2, 0.3, 0.25, 0.1, 0.1, and 0.05. This is the probability that in each day this will be the demand for washing machine. So, each days demand looks like so.

(Refer Slide Time: 18:34)



I am going to next draw a picture to show you the evolution of the system over time. So, this is t. So, now, you are observing at the end of each day: 1, 2, 3, 4, 5, 6, 7, 8, So that is just 8 days and then the inventory level. So, let us say it is at 5 in the beginning, 4, 3, 2 and 1. So, we start with 5 let us say and then it goes down to 4 and it is at 4 and then here you decide: I will not do anything, we let it go. Then it goes down to 3, and it stays at 3, I am not going to do anything. The next let us say it goes down to 2 and then goes down to 1. Then we say: oops! There is only 1. Let us go ahead and buy, the next morning I am at 5. Then, there is one demand of 1, demand of 2, demand of 3 and then you observe: oh! OK, we are still at 3, we are ok.

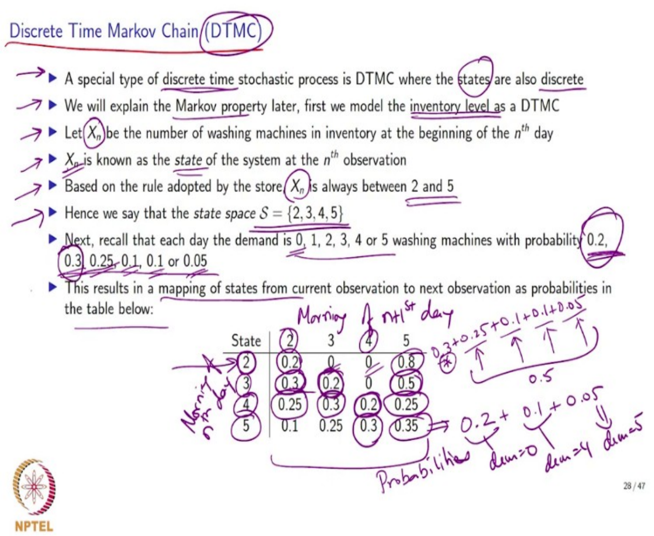
Then you will go from 2 to 1 to 0 and there will be nothing. So, there is any demand here, these demands are lost- here are lost. Then, you come here and then you order 5. So, whenever you place an order, the next morning you will have 5. Then, let us say it goes to 4, 3 and then you observe here and you do nothing, then it goes down to 2 and then you do not order right. Let me just redraw that you order it at this line; the order and next morning you get it. Then, it goes on like that, do nothing and then keeps going on.

So, in these situations you do nothing. Here, you place an order and then here you do nothing, here you order, here there is nothing and here you order, here you would not order. Now, another important thing is: why are we observing the system at the morning of a day and not in the evening?

So, if you look at in the evening times, your stage could be 0, 1, 2, 3, 4, 5 because you could have come as small as 0 like here or it could be as high as 5. These are the 5 numbers that you are going to see. Especially, if you started here initially with 5, which makes sense because maximum is 5. Of course, you could think of a never reaching 0 and do some other cool stuff, which I completely agree. Now turns out, many times they will not do that; we will hit those items in a little bit, but think about the following: the maximum you are ever going to be is 5.

However, in the morning, the minimum you will be is 2 because if you are less than 2; if you were 1 or 0, then for sure you would have placed an order so that the next morning you are at 5. So, it is either going to be 5, 4, 3 or 2. If it is 1, the next morning it is going to be at 5. If it is at 0, the next morning it is going to be at 5. So, you are really seeing the system in the morning at only 4 states. However, if you observe the system at the end of the day, you will see the system in 6 states: 0, 1, 2, 3, 4, 5 and then the morning it is 4 stage. Normally, we like to have as few states as possible because that is good modeling practice. But, you could have just solved this problem assuming that it is actually equal to I mean it is observed at the end of the day, you just have 2 more states.

(Refer Slide Time: 22:22)



Now, let us see how to model this. This system is modeled using what is called a Discrete Time Markov Chain. So, this is a very special type of stochastic process. It is observed at discrete times in our case at the beginning of each day and the states are also discrete. We see the system in states like 2, 3, 4 and 5 is a hand countable discrete. You could have stochastic processes that are neither discrete time nor the states are discrete.

For example the amount of water in a dam could be a stochastic process, which is observed continuously and not a discrete time. It is continuously observed and these states are also continuous; the amount of water in the dam is a continuous quantity. So, you could have something like this as well. So, we are in discrete time Markov chain. It is going to be interested only in discrete states and discrete time processes. Besides that, there is a property called Markov property that we will talk about later, which is why it is called discrete time Markov chain.

So, we will model the inventory level that we saw as a discrete time Markov chain and then we will talk about the Markov property later. So, we are going to use this letter X_n now in the spirit of random variables, where you use an uppercase X to denote something that is random; here too the number of washing machines that you will see in inventory at the beginning of the n^{th} day is actually a random variable. However, unlike the IID random variables, here the random variables are not independent and identically distributed.

So, X_1: if you know that X_2 depends on X_1, it depends on how many I have the previous day in order for me to figure out what I am going to have the next day. So, these are random variables, but they are dependent on each other. X_n is called state of the system- that is the state.

So, when you make an observation- what you observe is a state of the system. The observed quantity during the n^{th} observation is called a state and that is why I call it X_n. The subscript n is the n^{th} observation. Then, in our case like I said a little while ago: X_n is always a number between 2 and 5 because you observe in the morning, after you have replenished your inventory if any. Then, you are either going to be at 5 because you were at 5 and you did not sell anything the previous day or you were at 1 or 0 and you place an order and you will be at 5. You will be at 4, 3 or 2 and those are the only possible states you can be in the next morning.

You will never be at 1 because if you are at 1 you would have already placed an order and the order would have come. Although I did not state it explicitly, we do assume that whenever I place an order, the warehouse is going to have something to deliver. We will never be in a situation where the warehouse runs out of items. Now, that is a multi-echelon inventory control problem, which is another important problem studied by a lot of people. So, anyway this thing: 2, 3, 4 and 5 is what we call state space. The state space is the set of all possible values that X_n can take.

Now, remember that the demand is 0, 1, 2, 3 or 4 or 5 with probability 0.2, 0.3, 0.25, 0.1, 0.1 and 0.05. So, what we want to do is, we want to map the transition from 1 day to another. So, this one let us say the morning of the n^{th} day, this is the state. This is morning of the $n+1^{st}$ day, you observed the morning of the n^{th} day and the morning of the $n+1^{st}$ day.

So, let us say in the morning of the n^{th} today, I am in state 2, that means, I have 2 items at inventory; what could happen? Well I could have a demand of 0, if I had a demand of 0 with probability 0.2, then the next morning, I will be in state 2. Today you have 2, I had demand at 0 then the next morning I am guaranteed to have 2 items in inventory and that happens with probability 0.2. So, all these numbers inside here are probabilities. Now, if I were at state 2 today morning, tomorrow morning there is no way I will be in 3 or 4.

Why is that the case? Well, the worst I could be in at the end of the day today is 1 or 0 right; that is all I could be. And then that happens, the next morning I am guaranteed to be at 5. So, there is no way I can go from 2 to 3. I can only go from 2 to 5. So, the probability that happens is 0.8; 0.8 essentially is equal to 0.3+0.25+0.1+0.1+ 0.05 and why is that the case? Well, if I had a demand of 1 or 2 or 3 or 4 or 5, I will immediately go to 5 the next time I observe it. So, today when I observe and I observe the system is in state 2 that is what it is here right; we started with that and then if I had a demand of 1, I will go down to 1.

If I had 1, I will order 4 more and I will be at 5 that happens with the probability 0.3. If I had 2 and I had demand of 2, I will be down to 0. I place an order for 5 and I will be up at 5 the next day. Then, if I had 2; I had a demand of 3. Now what happens is the demand of 3, 2 of these will get washing machines, the third one's demand will not be fulfilled. Still the next morning you will be at 5. Likewise, if I had a demand a 4 or 5, 2 or 3 people will be rejected, they will not get the washing machines and then the next morning it will be at 5. Let us move to state 3. From state 3, here is all that could happen. If I had absolutely no one coming and buying anything, I will remain in state 3.

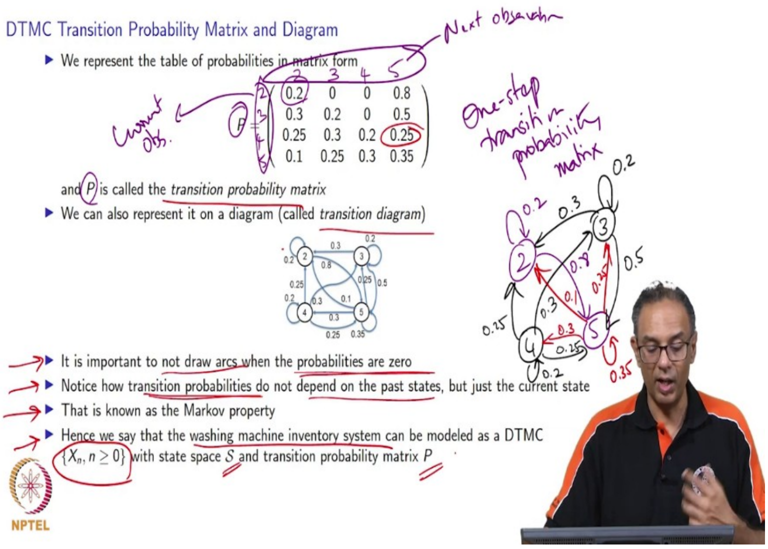
I will remain in state 3 with probability 0.2, which is the probability of 0 demand. If I have a demand of 1, I will come down to 2 which happens with probability 0.3. And then, if I had a demand of 2 or more then I will place an order and the next morning I will be in state 5 and 2 or more has a probability of so, it is these guys added together and that is 0.5: 0.25 plus 0.1, 0.35, 0.45 and 0.5.

So, with probability 0.5 we will go from 3 to 5. Let us look at state 4. In state 4 if I had 0 demand, I will stay in state 4 that happens with probability 0.2. If I had a demand of 1, which happens with probability 0.3, I will go to state 3. If I had a demand of exactly 2 that happens with probability 0.25, then I will go to state 0. Now, if I had a demand of 3, 4 or 5, I am guaranteed to go to state 5 that is because if I go down to 1 or 0, I will surely have to buy. Now state 5 is a little bit more tricky. So, in state 5, one of two things could happen here. This guy is equal to either I am in state 5 and I did not sell anything that happens with probability 0.2. I am in state 5, and I did not sell anything or I had 0 demand.

On the other hand, if I had also a demand, so this is demand of 0. If I had a demand of 4 or 5 with probability 0.1 and 0.2. So, this is demand equals 4 and this is demand equals 5. If I demand a 4, 5 or 0, I am going to go back to 5. This is crucial, we never had this in the other states and if you add them up that probability is 0.35.

Now, this guy 0.3 is the probability that will have demand of 1, which is right here or if I demand of 2 with probability 0.25, I will go to state 3. If I had demand of 2; trigger the number right now actually 3 with probability 0.1 then I will go to state 2. So, that those are my transition probabilities.

(Refer Slide Time: 31:10)



Now, I could write this exact same information in the form of a matrix. So, I am going to put down 4 numbers 2, 3, 4 and 5. All I have done is- I have taken these 16 numbers: 0.2, 0, 0, 0.8, 0.3, 0.2, 0, 0.5 and so on and then I put it down in a matrix form and I call that matrix P. P is what is called a one-step transition probability matrix. Now, we sometimes omit the word one step because that is the default: one observation, the next observation.

So, this is the current observation and one in the top is the next observation. So, the probability of transitioning from one to the other is written in the P matrix. Now, notice that these numbers 2, 3, 4 and 5 is the state space. So, this is the state space and one of those numbers is a state. So, you always write it in this form. So, P is called the transition probability matrix or the one-step transition probability matrix.

So, you take it and put it in a matrix form like this. We could instead also depict it in a figure. So, it is like this. So, from state 2 what could happen: well, I could have one of three things. Either I could remain in state 2 with probability 0.2- that is this guy or I could go to state 5 with probability 0.8.

So, what I do is I take 2 and then I put an arrow back to 2 or another arrow to 5. So, I write this as 0.2 and this as 0.8, then what I do is I am in state 3. So, from state 3, what could happen? From state 3, either I could go to state 2 with probability 0.3. I can remain in state 3 with probability 0.2 or I can go to state 5 with probability 0.5. From state 4, I can either remain in state 4, go to 2, go to 3 or go to 5 from 4 if I remain in 4 the probability is 0.2.

If I have to go to 3, probability is 0.3. If I have to go to 2, it is 0.25 and then if I have to go to 5, it is 0.25. And finally, from state 5, I could either remain in 5, go to 3, go to 2, go to 4 from 5 to 5 the probabilities 0.25. From 5 to 4, the probability is 0.3 and from 5 to 2 the probability is 0.1 and that information is put down in this great transition diagram.

So, we put that down in the transition diagram. So, that is our situation, we could model; these transitions are the way the stochastic process evolves over time either using the transition probability matrix or using the transition diagram. Both have their own pros and cons, you do not have to do both, but it will be good idea in the beginning to do both as far as possible. It is also important to not draw arcs, when the probabilities are 0. So, for example, I do not have an arc from 3 to 4. So, typically we will not draw an arc from 3 to 4 and that is because the probability is 0.

So, you never draw an arc when the probability is 0. Now, notice that the transition probability matrix, the transition probabilities themselves do not depend on the past state. So, for example, I do not care how much inventory I had 2 days ago or 10 days ago. For me to predict how much I will have tomorrow, all I need to know is how much I had today. So, it is called a one-step transition and it also does not depend on the past and that property is called the Markov property.

One more time, let me quickly repeat. The probability that tomorrow I will have 5 given that today I have 4 is 0.25. It does not depend on what was there yesterday, what was there the day before, what was the other day before that and so on. That property is called the Markov property.

Now, once all that is satisfied, we say that the washing machine inventory system can be modeled as a Discrete Time Markov chain (DTMC) and is written in this format. It is all values of X_n for n greater than or equal to 0, with state space and transition probability matrix. So, this is how we would model something as a discrete time Markov chain.

Thank you.