Introduction to Queueing Theory

Prof. N. Selvaraju
Department of Mathematics
Indian Institute of Technology Guwahati, India

Lecture - 06

Markov Chains: Definition, Transition Probabilities

Hi and hello, everyone. In this lecture, what we are going to start is our discussion on the discrete-time Markov chains. We have already seen an overview of stochastic processes, and we said at that point of time that the Markov chains, a Markov process in general, play a critical role in our analysis of queueing systems. So, at least all the basic queueing systems are any queueing system that you try to model in the beginning as a first level; let it always relies on this Markov process. So, it is; basically, we need to understand and equip with the necessary background on these Markov chains. Of course, this is not intended to be a complete analysis of the complete theory of Markov processes. We will just see what the Markov process is and some basic properties and whatever the relevant things that are more important for our purpose. So, that part is what then we are going to see. So, as we already defined it, you may recall the definition that says a stochastic process here since we already assumed this to be discrete-time, and since the Markov chain, it is the state space is already, we assumed to be discrete. Because it is a discrete-time as well, so, we said that we would be calling it as in $\{X_n, n \geq 0\}$ form rather than $X(t)$. It is X_n is said to be a Discrete-time Markov chain, or simply we will call this as a Markov chain without the adjective which is discrete time if

$$P\{X_{n+1} = j | X_n = i,\, X_{n-1} = i_{n-1},\, \ldots,\, X_0 = i_0\} = P\{X_{n+1} = j | X_n = i\}$$

for all states $i_0,\, i_1,\, \ldots,\, i_{n-1},\, i$ and j in S and for all $n \geq 0$ (provided the conditional probabilities are defined).

So, basically, what we are assuming?

$$P\{X_{n+1} = j | X_n = i,\, X_{n-1} = i_{n-1},\, \ldots,\, X_0 = i_0\} = P\{X_{n+1} = j | X_n = i\}$$

we know as the Markov property, and this Markov property should hold true for this process in order for us to call this X_n as a discrete time Markov chain or simply a Markov chain.

As we already pointed out here, both parameters and state spaces are discrete and what this says is that what this means is $P\{X_{n+1} = j | X_n = i,\, X_{n-1} = i_{n-1},\, \ldots,\, X_0 = i_0\}$, meaning finding the process in the state j at time $n+1$, given that the process was in state i at time n and at state i_{n-1} at time $n-1$ and so on and it started at the state i_0. This probability would be the same as the $P\{X_{n+1} = j | X_n = i\}$. So, what does that mean? The conditional distribution, this is what is a conditional probability, and since we require it for all j and all i and all states $i_0, i_1, \ldots, i_{n-1}$, so, this is basically you are looking for all possible i, j, i_{n-1} and so on.

So, that means the conditional distribution of X_{n+1} you can think of like a timeline that you are standing at time n and $n-1$ is in the past and $n+1$ in the future. So, the conditional distribution of X_{n+1} given the past and the present depends only on the present and not on the past. So, the knowledge of the past history of the process plays no role in

the further evolution of the process under study and whatever is required for its future evolution; all those information is contained in the description of the state of the process at the present time that is it. So, if you are given that $X_n = i$, so, that is sufficient enough to go further to predict what is going to be the future behaviour of this process. So, that is what essentially $P\{X_{n+1} = j|X_n = i\}$ tells.
So, this is an informal way, like the past has no bearing on the future, only the present has. So, which may not be true in many applications, but which is true, which is going to be true at least in our queueing systems; in most cases, that property holds true. So, these are very simple. As we already pointed out, it is a very simple form of dependency that dependency going back in time only one unit of time, and it is a very useful form of dependency among the random variables; that is why its popularity is there for the Markov process. Everywhere you will find applications. Even if it is not really the things that, this does not hold that is not independent of the past $X_0, X_1, \ldots, X_{n-1}$; in any analysis, many a time, you tend to start with such an assumption that only the present matters for the future, not what happened in the past.

So, this is what is the Markov chain, the definition. Now, this particular quantity that we are seeing here, so, $P\{X_{n+1} = j|X_n = i\}$ quantity because this what is relevant. This what in terms of this only you are defining the property. Now, $P\{X_{n+1} = j|X_n = i\}$ we denote by $p_{ij}(n)$, and we refer to this as transition probability or one-step transition probability. And what does this give? This $p_{ij}(n)$ gives the conditional probability of making a transition from state i to state j from time n to time $n + 1$. So, you are standing at time n and at this point of time, your state is i. Now, $p_{ij}(n)$ then the description is that what is the probability that in $n + 1$ it will be in state j that is what is this basically. So, that is why this is a one-step transition probability, many a time; you might refer to the simplest transition probabilities; transition probability simply means it is always one step. And if it so happens that this $p_{ij}(n) = p_{ij}$ irrespective of what is this n is, then the Markov chain is said to be stationary; Markov chain is said to have stationary transition probabilities, or the Markov chain is called as a time-homogeneous Markov chain or simply a homogeneous Markov chain. Of course, in our case, we consider, as we said already, like when we are talking about queueing system, we are looking at only the time stationary elements as far as our things our analysis is concerned. So, the Markov chains that arise in such models are basically time-homogeneous Markov chains. So, this is what is we are going to look at it, and if it is so happened, which mean that it is independent of whether you move from time step 2 to 3 or 27 to 28 in the time points because we always refer to the parameter as time points for easy reference. So, p_{ij} remains constant irrespective of what it is. So, you have this. Of course, if you make it dependent, the whole theory will go through, but it is a little bit more complex there. So, the matrix $P = ((p_{ij}))_{i,j\in S}$ is called (one-step) transition probability matrix (TPM). And so, basically, what do we have here? We have a P here, which is basically p_{ij} the matrix if I assume that state space to be 1, 2, 3, and so on for simplicity. So, suppose if I assume this to be $S = \{1, 2, 3, \ldots\}$, then I will have here $p_{11}, p_{12}, \ldots, p_{21}, p_{22}, \ldots$ and so on. Now, this is what the matrix is.

$$P = (p_{ij}) = \begin{pmatrix} p_{11} & p_{12} & \cdots \\ p_{21} & p_{22} & \cdots \\ \vdots & \vdots & \vdots \end{pmatrix}$$

So, that is what you will observe. You will observe that each $p_{ij} \geq 0$ for all i and j and $\sum_{j\in S} p_{ij} = 1$ for all i. So, such a matrix is what is called a stochastic matrix. So, in a matrix theory like this is in matrix theory language, this will be called a stochastic matrix or random matrix in some, but a random matrix may have a slightly different meaning in general.

So, you may refer to this simply as a stochastic matrix means each entry is nonnegative, and the row sums are 1 is called a stochastic matrix. Of course, it has some special properties, like one of its eigenvalues is 1. So, you will have some relevance when we are trying to analyze a Markov chain in general. In addition, if the column sums need not be 1 in general, but if it has, if it so happened that the column sums are also equal to 1, then it is called a doubly stochastic matrix. Again for a Markov chain that has a doubly stochastic matrix, you have some special properties; that is why it might be of some use if you encounter such a scenario, but in general, this is what will happen, nonnegative and row sums are 1. This is what is called the transition probability matrix, which basically is what is the Markov chains all the probabilities you are putting in the matrix form. And if the number of states here is finite so, we are going to have a finite matrix; if not, this will be an infinite-dimensional matrix that you may you will have. So, this description of the Markov chain, instead of keeping the probabilities 1 for each state and so on, if you give this P, describes the Markov chain in one way. For Markov chain X n, you need a corresponding transition probability. So, that transition probability is simply this transition probability matrix.

Now, let us look at some examples here.

Example 1.
An MC whose state space is given by set of integers is said to be simple random walk (SRW) if for some $0 < p < 1$, $p_{i,i+1} = p = 1 - p_{i,i-1}$. It is simple symmetric random walk (SSRW) if $p = 1/2$.

So, this is all one dimension. This is a Markov chain right because it does not depend on the previous cases where two-time step before whether it was in 1 or it was in 0 does not matter because, at each time step, it makes an independent time step. So, the random walk, in general, would fall into this Markov chain category.

Example 2.
In a sequence of coin tosses, the number X_n of heads in the first n tosses.

Example 3.
Consider a communication system that transmit the digits 0 and 1. Each digit, transmitted must pass several stages, at each stage of which there is a probability p that the digit entered will remain unchanged when it leaves.

Example 4. *(A Gambling Model)*
Consider a gambler who at each play of the game either wins Re. 1 with probability p or loses Re. 1 with probability $1 - p$. Suppose that the gambler quits plays either when he goes broke or he attains a fortune of Rs. N.

Suppose if assume that the opponent has an infinite amount of money. I mean, you do not restrict, but he has a finite amount of money; suppose if you think, then things will change depending upon the assumptions that you are making it. So, in that case, what might happen or both players put together if they have total money is, say N suppose if that is the scenario if you are looking at it. Then there will be another model. So, depending upon what, this is not precise, but you have to say how much money each player had at the beginning, where they are starting, and what happens when one of them goes broke or if there is an infinite amount of money between either one of them or both of them, then things will go on. So, that is why the state space will also come into play accordingly, but whatever it is like it is, it is a Markov chain it will turn out to be. Suppose if you assume that you know both of them put together to have a total of N amount and one starts with some x amount and the other starts with some $N - x$, now they play the game. Now, the game will be played until one of them go broke, which means that one of them loses all the money the other one gets all the money which means the state 0 and state N would be one particular nature it will behave. Once you go there, the chain will remain there forever, and so on. So, we will see that so, that kind of thing. So, different kinds of variations can be made out from such a gambling scenario, and many a time, you would see the Markov chain the different phases or different features just by considering such a model and by tweaking the underlying assumptions; what happens when a player is getting ruined in some sense all those things can be analyzed further. So, these are some simple examples.

Now, you know a random variable is probabilistically specified by its distribution. You know completely about a random variable if you know its distribution. Likewise, we said that the stochastic process, at least its distributional properties, are specified by its finite-dimensional distributions. So, what happens in the case of the Markov chain, whether you can have a complete description? So, it is possible that is what we said that the point of time in many situations like many interesting processes. This can be the complete specification can be given in terms of some simple quantities; that is what you know we will see in the case of the Markov chain as well. What are we saying here? A Markov chain is specified by its initial distribution and its transition probabilities. So, transition probabilities; if you know and if you know its initial distribution, then a Markov chain is completely specified. What we mean by that is the following.

Suppose the initial distribution is $P\{X_0 = i\}$ and also you have p_ij's, which are the transition probabilities. Suppose if you call $P\{X_0 = i\}$ as μ_i, then what we were saying is that

$$P\{X_0 = i_0,\, X_1 = i_1,\, \ldots,\, X_n = i_n\} = \left(\prod_{k=0}^{n-1} p_{i_k i_{k+1}}\right) \mu_{i_0}.$$

.

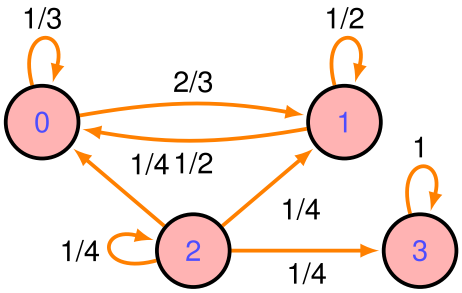
So that means that if I know this transition probability matrix and the initial distribution, then I can get the distribution. Now, how do you see this? This is very simple.

$$\begin{aligned}
LHS &= P\{X_n = i_n | X_0 = i_0,\, X_1 = i_1,\, \ldots,\, X_{n-1} = i_{n-1}\} P\{X_0 = i_0,\, X_1 = i_1,\, \ldots,\, X_{n-1} = i_{n-1}\} \\
&= P\{X_n = i_n | X_{n-1} = i_{n-1}\} P\{X_0 = i_0,\, X_1 = i_1,\, \ldots,\, X_{n-1} = i_{n-1}\} \\
&= p_{i_{n-1} i_n} P\{X_0 = i_0,\, X_1 = i_1,\, \ldots,\, X_{n-1} = i_{n-1}\} \\
&\vdots \\
&= p_{i_{n-1} i_n} p_{i_{n-2} i_{n-1}} \cdots p_{i_0 i_1} P\{X_0 = i_0\} \\
&= RHS
\end{aligned}$$

And you see here in this particular case of Markov chain; it is completely specified. So, each Markov chain is specified by; is associated with its stochastic matrix, and for each stochastic matrix, you can find a Markov chain. Now, along with this, if you specify the initial distribution, then you are completely specified. If the initial distribution is different, then the Markov chain behavior would be different, but this P would remain the same. Now, one useful tool which is used in the Markov chain and Markov process or continuous-time Markov chain theory is what we call a state diagram. A state transition diagram is commonly used, a state transition diagram or transition probability diagram, or simply a state diagram or transition graph. What is it?

For a Markov chain a state transition diagram (or simply state diagram or transition graph or transition probability diagram) for an MC is a directed graph where the nodes represent the states and the edges represent possible one-step transitions. More precisely, the state diagram contains an edge from node i to node j if and only if $p_{ij} > 0$.

This is a very useful tool in visualizing a Markov chain and studying its properties. As you go along, you will see how it is easy if you know this transition diagram in terms of it to look at the Markov chain, and we will use this extensively in the case of Markov chains that we are encountering. And in the applied fields like, mostly in engineering and say computer site networks or network you are trying to model it is a very complex Markov chain. And how do you describe the Markov chain? They will simply describe it in terms of such a diagram.

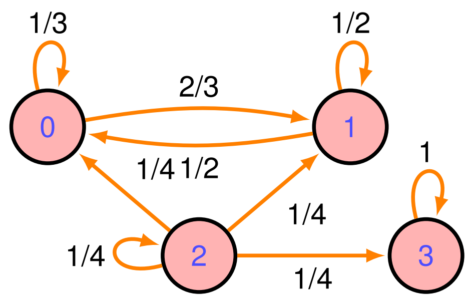


So, this is what is relevant. You will see many pieces of literature in the papers or research articles where a Markov chain, they will say like this is a Markov chain whose transition or diagram or transition graph is as given below finish. So, that is you will get some such thing that is what it means. So, now, let us see for this particular example how we are getting it here. So, you have

$S = \{0, 1, 2, 3\}$

and with TPM

$$P = \begin{bmatrix} 1/3 & 2/3 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$



So, this diagram is a very useful tool for your easy understanding of the Markov chain, which means when you have discrete states, then you can make it as 0, 1, 2, 3, and both discrete-time and continuous-time like this will turn out to be a very useful tool in understanding. So, for understanding, you can easily depict the Markov chain's behavior through this transition diagram. So, here it will be probabilities when you go to continuous time; it will be rates in terms. So, that is why sometimes that is to make a distinction; it says transition probability diagram or graph. So, remember the about this state thing. Now, we have talked so far only about the Markov chain and its one step. Now, we may have interest in knowing if today is our current time point if it is in the particular state now, what is going to be the process state after say 2-time units or 10-time units or 5-time units, not just 1-time units, that is that would be our interest.

So, basically, what we mean is that we have a Markov chain with state-space S and one-step transition probabilities p_{ij} for $i,\, j \in S$. Let us define

$$p_{ij}^{(n)} = P\left\{X_n = j | X_0 = i\right\} = P\left\{X_{n+k} = j | X_k = i\right\}.$$

This is as opposed to $p_{ij}(n)$ that we defined earlier in a non-homogeneous case; that is not the thing here. Here we have this $p_{ij}^{(n)}$ means $P\left\{X_n = j | X_0 = i\right\} = P\left\{X_{n+k} = j | X_k = i\right\}$. These are called as the n-step transition probabilities.

We have originally one-step transition probabilities, which is p_{ij}; now, this is n-step transition probabilities which is $p_{ij}^{(n)}$. Now, how do we compute these n-step transition probabilities? You can condition easily as you can easily you can work it out to see that; suppose if $P\left\{X_n = j | X_0 = i\right\}$ I have to compute, what can I do? I can use my total probability law to condition on what was the process at time n minus 1 and multiply by the corresponding probabilities total probability law that I can use. So, I can use it at any step time point. So, ultimately what will you reach? You will arrive $\sum_{k\in S} p_{ik}^{(m)} p_{kj}^{(n)}$. This is what is known as Chapman-Kolmogorov (CK) equations, which is very simple.

$$p_{ij}^{(m+n)} = \sum_{k\in S} p_{ik}^{(m)} p_{kj}^{(n)}$$

for all $m, n \geq 0$ and all $i,\ j \in S$.

So, you can write it down explicitly and see how this is true because this is one of the fundamental relationships that we will have in the Markov chain when we are trying to compute multi-step transition probabilities. Now, if we denote n-step transition probability matrix by $P^{(n)}$, then

$$P^{(n+m)} = P^{(n)} P^{(m)} \implies P^{(n)} = P^n.$$

Note: $P^{(0)} = I$. So, this one, so, you, for example, you can look at first to understand two-step that will be in terms of one step which is basically P. So, $P \times P$ is what then you are going to get here. So, that is P^2. Easily you can see that my $P^2 = P \times P = P^2$. And similarly, you would arrive at in general as this. So, now, this makes my life easy. Like if I want the n-step transition probabilities to be computed, what will I do? I will take the transition probability matrix, and I will raise that matrix to the required power, and the corresponding entries will give me the corresponding number of steps in which the Markov chain will make the transition from one particular state to the other particular state. Obviously, like if P is finite, you can do this nicely, but if P is infinite, of course, you have little difficulties, but that will be expected. Now, let us look at some examples of such scenarios.

Example.

:-Suppose that the chance of rain tomorrow depends on previous weather conditions only through whether or not it is raining today and not on past weather condition. Suppose that if it is raining today, then it will rain tomorrow with probability 0.75. If it is not raining today, then it will rain tomorrow with probability 0.40. Calculate the probability that it will rain four days from today given that it is raining today.

$$P = \begin{bmatrix} 0.75 & 0.25 \\ 0.40 & 0.60 \end{bmatrix}, \quad P^{(4)} = P^4 = \begin{bmatrix} \mathbf{0.6212} & 0.3788 \\ 0.6061 & 0.3938 \end{bmatrix}$$

And there is another example.

Example.

Suppose that balls are successively distributed among 8 urns, with each ball being equally likely to be put in any of these urns. What is the probability that there will be exactly 3 occupied urns after 9 balls have been distributed?

If X_n is the number of nonempty urns after n balls have been distributed, then $\{X_n\}$ is a MC with states $\{0, 1, 2, \ldots, 8\}$ and with $p_{ii} = i/8 = 1 - p_{i,i+1}, i = 0, 1, 2, \ldots, 8$ and the desired probability is $p_{03}^{(9)}$ which can be computed as **0.00756** using P^9.

But, for our problem, observe that the first transition is deterministic (from 0 to 1) and hence the required probability is equal to $p_{13}^{(8)}$, we can simplify the problem by letting $Y_n = \max\{X_n + 1, 4\}, n \geq 0$ with state space $\{1, 2, 3, 4\}$ and TPM P as given below.

$$P = \begin{bmatrix} 1/8 & 7/8 & 0 & 0 \\ 0 & 2/8 & 6/8 & 0 \\ 0 & 0 & 3/8 & 5/8 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad P^{(4)} = P^4 = \begin{bmatrix} 0.0002 & 0.0256 & 0.2563 & 0.7178 \\ 0 & 0.0039 & 0.0952 & 0.9009 \\ 0 & 0 & 0.0198 & 0.9802 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Now, $p_{13}^{(8)} = \sum_{j=1}^{4} p_{1j}^{(4)} p_{j3}^{(4)} = \mathbf{0.00756}$.

Now, apart from these transition probabilities, we will also be interested in what are called as state probabilities. So, which we define;

Consider an MC having state space S and one-step transition probabilities p_{ij} for $i,\, j \in S$. Let us define

$$\pi_j^{(n)} = P\left\{X_n = j\right\}$$

as the probability of finding the system in state j at time n (also known as state probabilities).

Remember, if I write the $p\{x_n = j | X_0 = i\}$, then that becomes the n-step transition probabilities, but I am not worried about where I started; I am worried about where the process is at the time n, which means what is the probability that I would find this process in state j at time n. So, that is what is $\pi_j^{(n)}$. So, this is all that is called as state probabilities which is what would be we will be interested in mainly in queueing theory as well. So, these are known as n-step state probabilities. It can be shown that

$$\pi_j^{(m)} = \sum_{i \in S} \pi_i^{(m-1)} p_{ij}$$

for all $m \geq 1$ and all $i,\, j \in S$.

In matrix notation,

$$\boldsymbol{\pi}^{(m)} = \boldsymbol{\pi}^{(m-1)} \mathbf{P}.$$

This means that

$$\boldsymbol{\pi}^{(m)} = \boldsymbol{\pi}^{(m-1)} \mathbf{P} = \boldsymbol{\pi}^{(m-2)} \mathbf{P}^2 = \cdots = \boldsymbol{\pi}^{(0)} \mathbf{P}^m,$$

where $\boldsymbol{\pi}^{(0)}$ is the initial state distribution. So, if I know the initial distribution and the transition probability matrix, then not just the n step transition probability matrix but the n-step state probabilities also, I can compute it. So, this is a vector now. This $\boldsymbol{\pi}$ is a vector, and P is a matrix here; that is what you will see. So, this is what you call state probabilities. Now, you can see how one can compute those.

Example. Suppose that $\{X_n, n \geq 0\}$ is a MC with three states $0, 1, 2$ and with TPM

$$P = \begin{bmatrix} 0.75 & 0.25 & 0 \\ 0.25 & 0.50 & 0.25 \\ 0 & 0.75 & 0.25 \end{bmatrix}$$

and the initial distribution $P\{X_0 = i\} = 1/3, i = 0, 1, 2$.

We have $P\{X_3 = 1|X_2 = 2\} = P\{X_7 = 1|X_6 = 2\} = 0.75$.
And, $P\{X_2 = 2, X_1 = 1|X_0 = 2\} = P\{X_2 = 2|X_1 = 1\}P\{X_1 = 1|X_0 = 2\} = (1/4)(3/4) = 3/16$.

Since $P^2 = \begin{bmatrix} 5/8 & 5/16 & 1/16 \\ 5/16 & 1/2 & 3/16 \\ 3/16 & 9/16 & 1/4 \end{bmatrix}$, *we have* $p_{01}^{(2)} = P\{X_{n+2} = 1|X_n = 0\} = 5/16$ *for* $n \geq 0$.
And, $P\{X_2 = 1, X_0 = 0\} = P\{X_2 = 1|X_0 = 0\}P\{X_0 = 0\} = 5/48$.

Exercise.

Determine $P\{X_2 = 1\},\ \ P\{X_0 = 0|X_2 = 1\},\ \ P\{X_3 = 0\}, etc.$

So, this way, like you know, one can compute this multi-step, one-step probability, state probabilities, and try to answer the questions related to this. So, these are elementary stuff that we have seen in this lecture. We will see more about its properties in the next lecture.
Thank you, bye.