Discrete – Time Markov Chain and Poisson Processes
Professor Ayon Ganguly
Department of Mathematics,
Indian Institute of Technology, Guwahati
Lecture 06
Module: Chapman– Kolmogorov Equations and Communication
Lecture: Markov Property and Chapman Kolmogorov Equations

Welcome to the lecture six of this course Discrete Time Markov chains and Poisson Processes. Just recall that in the last couple of lectures, we talked about stochastic processes and then we introduce Markov chain. And just recall that for a Markov chain the given, the present and past, the future only depends on present and does not depend on past.

We have talked about this and then we go for the definition of time homogeneous Markov chain that basically mean that probability of moving one step at any time are same . So, with, and then we have seen a couple of examples, where the Markov chain can be used. So, in today's lecture, we will carry forward those things.

So, first we will start with a fact. So, this fact I am not going to prove, but it is a very, very important concept for any kind of stochastic processes. So, the concept here is that see, we generally define our random variable and then any probabilistic nature of the random variable, any probabilistic characteristics of the random variable are specified by its distribution.

So, that is we have talked about earlier that when I am talking about any random variable, if I know the cumulative distribution function of the random variable, all its probabilistic natures probabilistic characteristics are specified by that cumulative distribution function. So, do we have something of the similar nature in case of stochastic process, more precisely in case of Markov chain? And the answer is yes, let us see what is that and this fact is all about that.

So, the main part of the facts is that will that for a stochastic process the probabilistic characteristics are specified by its finite dimensional distributions. A stochastic process is specified, is completely specified that means, it is probably strict characteristics are completely specified by its finite dimensional distribution. So, what is this finite dimensional distribution?

The finite dimensional distribution is nothing, but of this thing. So, see that I have $X_0, X_1, \ldots, X_n$ if I know this probability for all values of n and for all states $i_0, i_1, \ldots, i_n$, then we say that basically we know all the probabilistic characteristics of a stochastic process.

So, these things are important that need to be known for all $n \geq 0$ and for all the states i$i_0, i_1, \ldots, i_n$ belongs to the state space that basically we need to know. So, if we know this probability for all values of n and for all states $i_0, i_1, \ldots, i_n$, then I know all the probabilistic characteristics of a stochastic process, of that stochastic process. So, Markov Chain is a special case of stochastic process that we have already discussed.

So, of course, for the Markov chain also if I know the finite dimensional distributions, I know all the probabilistic characteristics of the Markov chain. So, for Markov chain I need to know what is this probability for all values of $n \geq 0$ and for all state space $i_0, i_1, \ldots, i_n$, if I know that for a Markov chain, if I know that probability for all n, all i, j's, I know what is the probabilistic characteristics of that particular Markov chain.

And for the Markov chain the nice thing is that we can even tell that if I know the one-step transition probability matrix, if I know the one-step transition probability matrix and if I know the initial distribution of a Markov chain, these two things together completely specify the probabilistic characteristics of a Markov chain.

Why? Again, we go back, that is the notice, recall that what is, before going into the proof just recall that what is the initial distribution? The initial distribution of a Markov chain is nothing but the PMF of X_0 that is basically the initial distribution. Now, and what is one step transition probability matrix? The one-step transition probability, we have denoted by this and if I consider this matrix for all $i, j \in S$, that matrix is basically a one-step transition probability matrix.

That is what we have talked about earlier. Now, in case of the Markov chain, in case of Markov chain, the, this fact basically tells us that if I know these two things, that the initial distribution and one-step transition probability matrix, then I know this probability, and if I know this probability that means that these probabilities specify the probabilistic characteristics of a stochastic process, so in, for a Markov chain also.

So, the idea behind this is that exactly this equation, this equation tells that this probability, I can write in terms of one-step transition probability, product of one-step transition probabilities and the initial distribution where μ_i is basically defined like this and then basically I can write in this particular manner, and that is it. So, why this equality is true? Let us discuss why this equality is true.

And the proof of that is very simple, because I can write that

$$P\left(X_0 = i_0, X_1 = i_1, \cdots, X_{n-1} = i_{n-1}, X_n = i_n\right)$$

this probability, you see that using the concept of the conditional probability I can write it

$$P\left(X_n = i_n | X_{n-1} = i_{n-1}, X_{n-2} = i_{n-2}, \cdots, X_0 = i_0\right) P\left(X_{n-1} = i_{n-1}, X_{n-2} = i_{n-2}, \cdots, X_0 = i_0\right)$$

And how do I get that? This is basically nothing but the simple fact that

$$P\left(A \cap B\right) = P(A|B)P(B), \text{ if } P(B) > 0.$$

Exactly that is what I have used here. So, this one I have taken as A and that whole event I have taken as B. So, I can write $P\left(A \cap B\right) = P(A|B)P(B)$. Now, because X is a Markov chain here; that basically mean that given the present and past future only depend on the present, there will be no dependency on the past.

So, that means, I can write

$$P\left(X_n = i_n | X_{n-1} = i_{n-1}\right) P\left(X_{n-1} = i_{n-1}, X_{n-2} = i_{n-2}, \cdots, X_0 = i_0\right)$$

Now, the first quantity is basically nothing but one-step transition probability. So, that means, I can write the first quantity as $P_{i_{n-1}}$, I am moving to i_n and then multiply it by this probability.

And this probability again I can write

$$P\left(X_{n-1} = i_{n-1} | X_{n-2} = i_{n-2}, \cdots, X_0 = i_0\right) P\left(X_{n-2} = i_{n-2}, \cdots, X_0 = i_0\right)$$

Again this quantity actually given the present and past the future only depend on the present.

So, this is again a one-step transition probability matrix. So, I can write this is $P_{i_{n-1}}\ i_n$ then P, I am moving from i_{n-2} to i_{n-1} and then the rest of the probability that $P\left(X_{n-2}=i_{n-2},\cdots,X_0=i_0\right)$. Now, if I proceed this way then I will basically get this product term; I will basically get this whole product term.

And finally what I am left? It basically nothing but, in this way if I keep on proceeding it basically turns out to be

$$p_{i_{n-1},i_n},\cdots,p_{i_0,i_0}P\left(X_0=i_0\right)$$

from here is basically nothing but my μ_{i_0} and the rest of the terms are written in terms of the product here.

So, the main takeaway from this slide is as follows that in case of a Markov chain, the initial distribution and the one-step transition probability matrix completely specify the probabilistic characteristics of a Markov chain. So, with this takeaway, let us proceed.

So, next one is called Markov property and this property I have put as a theorem and that, let us first go through the theorem and then try to understand it line by line. So, the theorem tells that let X_n be a Markov chain with initial distribution μ and transition probability matrix, this transition, when I am writing the transition probability matrix keep in mind this is a one-step transition probability matrix, so one-step transition probability matrix P.

So, these two things together as we discussed completely specify the probabilistic characteristics of the Markov chain X_n. Now, the main point in this theorem is as follows is tells that then conditional on $X_m=i$, $\{X_{m+n}\}_{n\geq 0}$, this stochastic process, this stochastic process is a Markov chain with initial distribution δ_i and one-step transition probability matrix P.

So, let us try to understand what that theorem basically, in practice or intuitively what does it make; what does it tells us, ok. So, suppose a Markov chain starts from here. So, this is time 0, then time 1, so on it is going and here I have the time m and at the time m the state is i, so this is time, this is state. So, state is i, at the time in the state is i; that is basically this condition is telling us. Then what is this part? This part is basically nothing but $m+1$, $m+2$ and so on it is going.

So, this part when I put $n=0$, this is X_m, $n=1$, this is X_{m+1}, $n=2$, this is x_{m+2} and so on so forth. So, this part is nothing but this part of the Markov chain, this part of the Markov chain, from m onwards. So, what I know the conditional on $X_m=i$ that means, I know that at the step m, at the time-step m the state is i. So, this theorem tells us that if I know that I have seen the state i here then starting from here the rest of the part.

So, removing, so I am removing this part, I am only keeping this part that the later part is again a Markov chain, the theorem basically tells that if I know that at the point i, at the at time step m, the state i has occurred, it says that starting from m, the rest of the part is again a Markov chain, not only that, it basically says that the transition probability remains same, this P and this P are same, one-step transition probability remains same, but the initial distribution changes.

Whatever initial distribution you have mu does not matter. Now, the initial distribution is δ_i and recall that what is δ_i, $\delta_i(x)$ is basically 1, if $x=i$ and 0, otherwise, that is what the definition we have taken. So, it is δ_i and that is obvious because I have already seen at the step m it is i, so basically, $P(X_m=i|X_m=i)=1$ because this is conditioning on that, which is always 1.

So, this is the probability mass function; that is the probability mass function of $X_m|X_m = i$. So, basically this means that if I know the state of a Markov chain in any state, in any time step, then starting from this time step the Markov chain restart itself, only thing is that its initial distribution now is δ_i and where i is the state which is visited at that particular time step.

So, that Markov property basically mean that if I know the state at any time step Markov chain restart itself from there as if it started from Xm equals to i and then proceeded in time, then it is proceeds in time. So, that is basically the idea of Markov property and of course, this is a valid property and proof of this one is very, very simple, it is intuitive, because it is a Markov chain.

So, if I know the present, I do not need to know the past to tell about probability of the future and that is exactly, exactly this theorem also tells us, so this is basically the Markov property.

Let us proceed now, let us talk about now n-step transition probability. So, till now we are talking, we talked about one-step transition probability, and one-step transition probability we have defined as $P(X_1 = j|X_0 = i)$ and because it is a time homogeneous Markov chain, we are only considering time homogeneous Markov chain. So, it, I can also write as $P(X_{1+k} = j|X_k = i)$.

Ok, so, it is a time homogeneous Markov chain. So, this is true for all k. So, what is now this n-step transition probability matrix? Just to recall that this is one-step transition probability matrix and in case of one step transition probability matrix is that the currently I am in the i-th step, what is the probability of going to the j-th state in one-step.

Currently I am in the state i, what is the probability that going to state j in exactly one-step that is basically, that conditional probability is basically the one step transition probability. Similarly, when I talk about the n-step transition probability, that is nothing, but currently I am at the state i then given this condition, what is the conditional probability that I will be instead j in exactly n-steps. So, currently I am in state i given this condition or conditional on that event, what is the conditional probability of that?

I will be in state j in exactly n-steps. I will be n-state j in exactly n time steps, that is basically nothing but called n-step transition probability. So, n-step transition probabilities given by this. Now, just like this, because of the time homogeneous Markov chain, I can, this one I can replace by $k + 1$ and this 0 I can replace by k. So, I am just adding k in the indices.

So, it does not matter whether from 0, I am moving to the step 1 or whether from k, I am moving to the $k + 1$, the probability will remain same, if I go from i to j, the probability will remain same. Similarly, in case of the n-step transition probability also, this can be proved that whether I am adding k on both the indices, whether I am going from 0 to n or I am going from 0 to $k + n$, the probability will remain same.

And showing this one is not complicated at least for two step I can show very easily that if I talk about $X_2 = j|X_0 = i$,is exactly same as $X_{2+k} = j|X_k = i$. Why this is true? The reason is that, notice that the later probability, , $P(X_{2+k} = j|X_k = i)$ this quantity I can write as $P(X_{k+2} = j, X_{k+1} = l|X_k = i)$ and I take the summation over l, $l \in S$.

Why these two probabilities are same? The idea is basically nothing but theorem of total probability that at the state $k + l$, the Markov chain has to be in some state. So, I am taking at the step $k + l$, at this time step $k + l$ I am at the state l and then I am taking the sum over all possible values of l. So, that is why these two probability using the theorem of total probability are same. Now, I just use the Markov things, so I can write $l \in S$,

$$P(X_{k+2} = j|X_{k+1} = l, X_k = i) \cdot P(X_{k+1} = l|X_k = i).$$

So, now you see from here is that first one that if I know the present, past does not have any effects. So, first probability can be written as it will be $l \in S$, the first probability can be written as it is nothing but p, in one-step moving from l to j, in one-step moving from i to l and the next probability, this probability is one-step moving from i to l, so p_{il}. So, this probability is same as this, so this probability is same as $P(X_{k+2} = j|X_k = i)$. So, we have proved that this probability can be written in this particular form.

Now, notice that in this particular expression there is no k, there is no k in this particular expression and that basically means that where, whatever value of k I take here, the probability remains same, the probability does not depend on k and that is why I can take $k = 0$ and if I take $k = 0$, I get this particular expression and this one can be generalized to prove exactly this particular equality also.

I can prove this particular equality just generalizing this proof; in this case, in between I have only one state, one time step, so I am, here I am talking about $k + 2$, I have only one time step in between, if I have multiple steps step in between I have to basically take all those time steps here and equate it to $l_1, l_2, l_3, \ldots$ and then take sum over all those lj's and then I am done.

So, the proof will be little bit complicated, but if you can write it, I am sure you will be able to prove this, but the basic idea is that the in between step I have to just incorporate and then I have to take the sum over all possible values of l what demands that l belongs to the state space S. So, this is basically definition of n-step transition probability.

And then this note basically tells us that all it does not matter whether I am moving from 0, time 0 step to the time step n or whether I am moving to time step k to times step $k+n$. As long as I am going from i to j, state i to j in n-step it will remain same whether I am starting from 0 or whether I am starting from any value or any other value, any other time step.

So, with that let us proceed and the next one is basically Chapman Kolmogorov equations, and this equation is a very, very useful equation in in Markov chain. We will see some of the use of this particular equation. So, what is Chapman Kolmogorov equation tells us? Chapman Kolmogorov equation tells the following, I put it as a theorem, it says that, consider a Markov chain having the state space S and S in this case given by $\{0, 1, 2, \ldots\}$.

And one-step transition probabilities p_{ij}, for all $i, j \in S$. This was our standard notation. Then that Chapman Kolmogorov equations are given by this particular equation and here you see that I have i, I have j, I have m, I have n, I have m here, I have n here, I have i and I have j and it says that this particular equation is true for all values of $m, n \geq 0$ and all values of i and j equals to $0, 1, 2, \ldots$, for all $i, j \in S$.

That basically mean that i and j, for all values of $i, j \in S$. So, that is basically called Chapman Kolmogorov equation. Now, let us try to see what this equation actually tells us? This equation actually tells us as this the following thing, see, what is this, what is this quantity, that this quantity is basically nothing but

$$P(X_{m+n} = j|X_0 = i).$$

So, this probability is nothing but in $(m + n)$-step, what is the, this probability is the conditional probability that starting from i, I am moving to j in exactly $(m + n)$-step, after $(m+n)$-step I am in state j, starting from i after $(m+n)$-step I am in step j, that is basically this probability. And what is this individual probabilities here, so that this probability is $P(X_m = k|X_0 = i)$.

And this probability is $P(X_n = k|X_0 = j)$. I'm sorry, I am sorry; it will be other way around; that it will be $P(X_n = j|X_0 = k)$. So, if I look into this one, so what I am doing is that this side probability is that I am at zero time I am in state j, at the $(m+n)$-step, I am in the state j. So, this side is basically saying that, well, in m step I am moving from i to some k.

So, starting from zero, I have m step here, at the m step I am in some state k and then starting from k, I am moving to j in n step. So, in this side I have m step, in this step I have n step, so, sum it up to $(m+n)$ steps. So, that means that this theorem says that starting from zero, I am going to j in $(m+n)$ steps, but that can happen through any state at the time state k because I have the sum here .

So, I am starting from zero, finally, I am reaching the state j in $(m+n)$ steps and that can happen like this starting from 0 at the m-th step I am in k and then starting from k, I am at the state j in $(m+n)$ steps and when I sum them up over k, over all possible values of k that means that I am moving from 0 to $(m+n)$, I am moving to i to j in $(m+n)$ steps.

So, that is basically these theory, these theorems, these equations all about. And the proof of the equation is very, very simple just we did in the last case that I can write this probability which is basically this probability, I can write as $P(X_{m+n} = j, X_m = k|X_0 = i)$ and then I have to take the sum over all possible values of k.

So, k can take value 0 to ∞ because that is my state space. So, basically instead of writing this sum to 0 to ∞, I can also write this sum $k \in S$, it is the same thing. So, all possible state that at the m step I can be in any step, in any state, so I am taking this sum. So now I use again that conditioning technique that this is basically nothing but

$$\sum_{k=0}^{\infty} P(X_{m+n} = j|X_m = k, X_0 = i) \cdot P(X_m = k|X_0 = i).$$

Now, this is the most recent thing I know, this is the past thing.

So, this term I can get rid of and finally, I have,

$$\sum_{k=0}^{\infty} P(X_{m+n} = j|X_m = k) \cdot P(X_m = k|X_0 = i).$$

So, whatever I have discussed here that in the mth step I am in k, I am making that thing here, I keep that thing here and then I am? this k can be anything, this k can be any state, any state so that is why I taken the sum over k and then I am done.

Now, this part is nothing but from m, the time step m to time step $(m+n)$, so this is exactly n steps there. So, n-step I am moving from k to j, so the first part is basically nothing but p, in m-step I am moving from, I am moving from k to j and then in m-step I am moving from i to k. So, sorry, this will be n, this will be n. So, in n-step I am moving from k to j and in m-step I am moving from, I am moving from i to k.

And exactly that is what I have written here p_{ik} in m-step multiplied by p_{kj} in n-step, that is what basically given here. So, this is a very, very useful equation, in Chapman Kolmogorov equation is a very, very useful equation in case of the Markov chain.

So, let us now talk about something called n-step transition probability, we have talked about one-step transition probability that is basically nothing but P, we denote with P, which is basically written as $(p_{ij})_{i,j \in S}$. So, that is basically p if $S = \{0, 1, 2, \ldots\}$, this matrix is $p_{00}, p_{01}, \ldots; p_{10}, p_{11}, \ldots$ and this matrix is of this form, that matrix we have talked about earlier.

So, similarly, I can now talk about n-step transition probability matrix also and the standard notation we are going to use is $P^{(n)}$ and that is basically nothing but I take $p_{00}^{(n)}$, 0 in n-step I am starting from 0 going to 0, $p_{01}^{(n)}$ in n-step, I am starting from 0 moving to 1 in n-step and so on so forth, then $p_{10}^{(n)}$ in n-step, $p_{11}^{(n)}$ in n-step and so on so forth and this way I can fill it up .

So, that is basically n-step transition probability matrix and as I said that the standard notation we are going to use $P^{(n)}$ in the same notation, in this case. So, clearly if I take $P^{(1)}$ that will be nothing but P, that one-step, this is basically the one-step transition probability matrix that is same as our this P what we have used earlier. Now, let us see this one, this one is basically nothing but matrix form of the Chapman Kolmogorov equations.

So, just recall that if I have a matrix A and a B, suppose the matrix A looks like $a_{11}, a_{12}, .., a_{1n}$; $a_{21}, a_{22}, \ldots, a_{2n}$, finally $a_{n1}, a_{n2}, \ldots, a_{nn}$. So, this is the $n \times n$ matrix, similarly B is also $n \times n$ matrix, $b_{11}, b_{12}, \ldots, b_{1n}; a_{21}$ finally $b_{n1}, b_{n2}, \ldots, b_{nn}$. So, both of them are n cross $n \times n$ matrix. Now, if I take the matrix product of these two how do we proceed? This, the first row multiplied by first column here.

So, that is basically mean that a common one, a general term of the product can be written as $\sum_k a_{ik}b_{kj}$. This is the ij-th term, ij-th term of the product matrix, this is the ij-th term of the product matrix. So, now, if you see this Kolmogorov equation called Kolmogorov Chapman equation you see exactly of this form, this is basically plays the role of a_{ik} and this is plays the, this part is plays the role of b_{kj}.

So, that means that this sum can be written in the product matrix form and if we can do that, we can exactly get that. See that P, the entry of $P^{(m+n)}$ is basically nothing but $p_{ij}^{(m+n)}$, p_{ij} in $(m+n)$ step. Now, if I see the entry of $P^{(m)} \cdot P^{(n)}$, so that is basically p_{ij} in n-step multiplied by p_{ij} m-step. And when I do, similarly when I do this that actually give me nothing but $\sum_k p_{ik}^{(n)} p_{kj}^{(m)}$.

And exactly that is what is written, so Kolmogorov equation in the previous slide says that these two quantity are same, these two quantity are same, this quantity and this quantity are same so that means that if I take this product, I will basically get, I will basically get exactly this matrix. So, this is basically nothing but the matrix form of Chapman Kolmogorov equations.

Now, the nice thing here is as follows that; see that because $P^{(1)} = P$, that $P^{(1)}$, the one-step transition probability matrix we have denoted by P. I can actually claim this particular thing; I can actually claim this particular thing. How? The reasoning is very simple. Suppose, I start with $P^{(n)}$, I can write it $P^{(n-1+1)}$.

That $n - 1 + 1 = n$, so I can write in this manner. Now, I just use this matrix form of the Chapman Kolmogorov equation, replacing n by $n-1$ and m by 1. So, I can write this one as $P^{(n-1)} \times P^{(1)}$ and $P^{(1)} = P$ so this turns out to be $P^{(n-1)} \times P$. Now, if I keep doing this again and again finally I get it is P^n.

So, this tells us that if I need to find out the n-th step transition probabilities, I can start with one-step transition probability and take that to the power n. So, I just multiply that. So, basically that means, if I try to find out two-step transition probability $P^{(2)}$, probability is that is nothing but P^2 so this is $P \cdot P$.

So, I start with P, that P, I can find out from the problem, then I product it $P \cdot P$, that is P^2 and this elements of the P^2 are nothing but two-step transition probabilities. With that I stop in this particular lecture, and in the next lecture, we will see some of the examples where we can use this Chapman Kolmogorov equations to find out several kind of probabilities which are of interest in many situations. Thank you for listening.