Discrete – Time Markov Chain and Poisson Processes
Professor. Subhamay Saha
Department of Mathematics,
Indian Institute of Technology, Guwahati
Lecture 20
Lecture: Limit Theorems I

Hello everyone, so welcome to the 20th lecture of the course Discrete-time Markov Chains and Poisson Processes. So, in the last module, we saw about stationary distributions and I said that it is very important in the context of discrete time Markov chains. Till now, we have seen two important properties or two important results about stationary distributions. The first was if you start with a stationary distribution that means if the initial distribution is the stationary distribution, then the distribution of all accents is also same as the stationary distribution. So if you start from the stationary distribution, you remain there. That is one of the reasons why we call it a stationary distribution. Another thing we saw that if a Markov chain has a, or if an irreducible Markov chain has a stationary distribution then all states are positive recurrent. Now the question is are these all, so are these all about stationary distributions or there are some other important properties of stationary distributions, the answer is yes, there is something more to stationary distributions and that we are going to see in this module, which we are going to start today, which is limiting distributions. So, let us start.

So, first we start with some definitions. So, for any state $i \in S$, we define $V_i^{(n)}$ to be like all t is between 0 and n such that $X_t = i$ or in other words it is $\sum_{k=0} n\delta_i(X_k)$ that means this $\delta_i(X_k)$ is equal to 1 if $X_k = i$, 0 otherwise. So, then what is this $V_i^{(n)}$? This is basically the number of visits to state i during $\{0, 1, 2, \ldots, n\}$. Remember, we have already seen this total number of visits where the summation was from k running from 0 to ∞, which we denoted by V_i, but now we are looking at $V_i^{(n)}$ which we are saying the total number of visits or the number of visits to state i during the time $\{0, 1, 2, \ldots, n\}$, during the first n time units. That is the first definition. Now, the second definition is again for any i in the state space, define $L_i^{(n)} = \frac{V_i^{(n)}}{n+1}$, but then what is this $L_i^{(n)}$, so remember $L_i^{(n)} V_i^{(n)}$ was the number of visits to state i during $\{0, 1, 2, \ldots, n\}$, but now we are dividing by $n + 1$, so we are looking at the proportion, so this $L_i^{(n)}$ is called the empirical distribution at time n or it is also you can think of it as the proportion of times the chain is in state i. Because $V_i^{(n)}$ is the total number of visits between $\{0, 1, 2, \ldots, n\}$ and then I divide by the total time so I get the proportion of times the chain is in state i, fine. So, that i call as $L_i^{(n)}$ and this there is another name to it is also called the empirical distribution at time n. So, if I look at this $L_i^{(n)}$ for all $i \in S$. So if you fix an i, it is the proportion of times the chain is in state i and $L_i^{(n)}$ is called the empirical distribution at time n. Where now i runs from one, like i runs over the entire state space, why it is distribution because you can easily check the way it is defined if you do summation $L_i^{(n)}$, say for example, if you do summation $L_i^{(n)}$ sum over $i \in S$, that will be equal to 1, this each L_i's are obviously greater than or equal to 0, so this is a probability vector, or what is called a probability mass function, that way also you can think of it. So it is, that is why it is actually a distribution and it is called the empirical distribution at time n. Now, we come to the first theorem, which is called the law of large numbers for Markov chain. Now, why am I calling it law of large numbers? Recall, what is like what is law of large numbers say, for example, say if you are tossing a coin, then if I, so look at say the number of times you get i sorry, the number of times you get a head divided by n, so if you look at that. Or the proportion of times you get a head when you toss a coin n times, if I look at that look at this quantity then you know what law of large numbers tells you, it tells you that it goes to actually the probability of head of that particular coin. So, if I am looking at, like if I am tossing a coin infinitely many times and if I am looking at the quantity that the proportion of times the coin lands head in the first n tosses, then if I take n going to infinity, it converges to the probability of head that is one implication of law of large numbers. So, you will see, so here I am calling this theorem as law of large numbers, because it tells us the limit of this $L_i^{(n)}$ which is also the proportion of times the chain is in state i.

So, let $\{X_n\}_{n\geq 0}$ be an irreducible and recurrent Markov chain, fix a state i, suppose that $P(X_0 = i) = 1$ or in other words the initial distribution is δ_i. Then for any state j, if I look at this $L_j^{(n)}$, that means what this is the proportion of time, so the chain is starting from i and you are looking at the proportion of times the chain is in state j up to time n. Then what this theorem is telling you that this converges to $\frac{\gamma_j^i}{E_i(T_i)}$, if $E_i(T_i) < \infty$ and again recall, what is this γ_j^i is what we saw in the previous module that this is basically starting from i the expected number of visits to state j up to the first passage time to i, that means you look at from 0 to $T_i - 1$ where T_i is the first passage time to state i, so that is $\frac{\gamma_j^i}{E_i(T_i)}$, if $E_i(T_i) < \infty$ and it is equal to 0, if if $E_i(T_i) = \infty$. So remember it is a recurrent Markov chain. So each state is recurrent, but now the question is whether each of them is null recurrent or positive recurrent remember it is irreducible. So either all states will be null recurrent or all states will be positive recurrent. So if all of them are positive recurrent then it converges to this quantity $\frac{\gamma_j^i}{E_i(T_i)}$ and if each of them is null recurrent then it converges to 0. Again see this each $L_j^{(n)}$, we have already seen if you sum then it is equal to 1, similarly we also know that, if you say sum, γ_j^i, over all j, if you sum it over all j, you actually get $E_i(T_i)$, that is from the definition of γ_j^i and we are dividing it by that. So actually this quantity also if you sum over all j you actually get 1, because this sum over j, $\gamma_j^i = E_i(T_i)$, we are dividing by the sum so now the total sum becomes 1. So, again this γ_j^i over $E_i(T_i)$ is actually, so at least in this case where $E_i(T_i) < \infty$ it is again a probability mass function or a distribution. And now how is this converge, like what kind of convergence is this, this happens with probability 1 now what is with probability 1 remember this each $L_j^{(n)}$ these are all random variables. Because, so look at the definition so again it is a random variable so it for given each omega in the probability space you will get one value. So, what this means is that for each ω, $L_j^{(n)}(\omega)$ converges to this quantity as $n \to \infty$. So, this is convergence ω, so at each ω or for each again it is not actually for each ω what this with probability 1 tells you like the set of all ω for which this convergence happens that has probability 1. So, it is possible that this convergence fails for some ω but the probability of the set of all omegas where this convergence does not happen, the probability of that set is equal to 0, fine. So, anyway, so the main thing is, this is a random variable, which is telling that for each ω this converges to this quantity. But remember this limit is actually a constant, but again, a constant you can think of it again as a random variable. So, it says that for each ω, $L_j^{(n)}(\omega)$ converges to this constant if $E_i(T_i) < \infty$ and it converges to 0, if $E_i(T_i) = \infty$. Again, although I am saying it for all ω, but again it is actually not for all ω, this statement with probability 1 means the set of all ωs or the set of all bad ωs where this convergence does not happen that has probability 0. So, the set of all ωs for which this happens has probability 1. And in probability as long as something happens on a set with probability 1, we are happy for all purposes it happens for all ω that is why I am using. Again, it is not a very precise statement to say for all ω but from a probabilistic point of view it is actually for all ω because it happens with probability 1, fine. Now, in particular, if i is positive recurrent that means if $E_i(T_i) < \infty$, fine. That means we are in this first case then if I look at $L_i^{(n)}$, now what is $L_i^{(n)}$, $L_i^{(n)}$ is now starting from i, the proportion of time you spend in state i up to time n, but now that means, so this was for any j. Now, if I take $j = i$,, we know that $\gamma_i^i = 1$, we have already seen this when we define this

γ, this quantity γ in the previous module, the first result we saw that there it was k, so it was γ_k, $k = 1$, but again here instead of k, we are taking the state i, so g$\gamma_i^i = 1$ and again it is very easy to see from the definition as well because this is the expected number of visits starting from i to state i up to the first passage time. But since the first passage time is the first time after time 0 the chain hits i so in between it will not hit i. So, the only time it is at i is at time 0, so $\gamma_i^i = 1$. So, the numerator becomes on1, so if I look at $L_i^{(n)}$ this converges to $\frac{1}{E_i(T_i)}$. But now we know that this if i is positive recurrent we are in the setting of irreducible Markov chain so that means the whole chain is positive recurrent. Now, we know from the theorem that we saw in previous module that if you have an irreducible Markov chain and all states are positive recurrent then there exist a unique stationary distribution π_i which is equal to $E_i(T_i)$. So, what this theorem tells you that if you have an irreducible positive recurrent Markov chain then this proportion of times the chain is in state i starting from i this actually converges to the stationary distribution. So, you see in that ways you get another important property of stationary distribution. So again, this is a very natural question to ask that the proportion of time the chain spends or the chain visits the state i within first in visits and it turns out that if the chain is irreducible and positive recurrent then this quantity converges to the i-th component, if we are looking at the ith state then it converges to the i-th component of the stationary distribution. So, you see stationary distribution here makes an appearance as the limit, the limit of what? The limit of the proportion of times the chain visits state i. So, since this is like as $n \to \infty$ so this is also what is called the long run proportion of times long run because we are looking at $n \to \infty$. So, the long, if a chain is reducible and positive recurrent and suppose you fix a state i and you say that the chain starts from state i then the long run proportion of time the chain visits state i at is equal to π_i. Long run proportion means the proportion as $n \to \infty$. So, the long run proportion of time, the chain visit state i or spends at in the state i is π_i which is the stationary distribution. So, you see yet another important property of stationary distribution because it comes as a limit of this proportion of times the chain is in state i or it is the long run proportion of times the chain is in state i. So, another important property of stationary distribution, it makes an appearance as a limit of something.

Now, another important theorem. So, we have already seen that in a Markov chain, the states can be classified as either transient or recurrent now if it is recurrent then it can be further classified into two subclasses positive recurrent and null recurrent. Now, so you fix a state $\imath \in S$ then i is transient if and only if this thing is less than infinity this we have already seen and we have already seen a theorem, we said that i is transient if and only if $\sum_{k=0}^{\infty} p_{ii}^{(k)} < \infty$ and it is recurrent if and only if $\sum_{k=0}^{\infty} p_{ii}^{(k)} = \infty$, in this case it is recurrent. So, we have already seen this theorem. So, this first part is already proved, but now the question is how do I know whether it is positive recurrent or null recurrent. So, I need some further criteria to understand this further subdivision. Now, this theorem talks about that for the criteria. So, what this point 2 is saying, i is null recurrent if and only if this is equal to infinity anyway this is fine, because null recurrent means it is recurrent. So if, so this thing has to happen that this sum should be equal to infinity. And if you look at this quantity $\lim_{n\to\infty} \frac{1}{n+1} \sum_{k=0}^{n} p_{ii}^{(k)}$ that should be equal to 0. And i is positive recurrent if and only again this part is already done and this thing is strictly greater than 0. So, this quantity basically gives you the distinguishing criteria. The first is true for recurrence now it is null recurrent if this quantity is equal to 0 and it is positive recurrent, if this quantity is strictly greater than 0. So, this is one way of checking whether a state is positive recurrent or null recurrent. So, first in order to check whether it is transient or recurrent, you need to check this quantity whether this is equal to infinity or less than infinity. Now, once you get, so if it is less than infinity settle it is transient. If it is equal to infinity then you have to further check whether it is positive recurrent or null recurrent in order to do that you have to look at this quantity. So, if it is equal to 0 then it is null recurrent and if it is positive, if this limit is greater than 0 then it is positive. So, that is what this theorem tells you. So, it, now it gives you criteria for checking all three transient, recurrent, null recurrent, positive recurrent.

Now let us look at the proof of this the proof is more or less simple. So, here you see that I have assumed anything about the Markov chain, that whether it is irreducible or not, how many communicating classes it has, but so it is true in general no matter how many classes it has. So for each state you if you do all this checking then you can say whether it is null recurrent, positive recurrent or transient. So, in addition if you know that it is irreducible then you just need to check for any particular state then all states are same, because all these are class properties. But in the proof, we will just assume irreducibility because that makes the proof little bit simpler, but the result does not need irreducibility. But for the proof, we will just assume irreducibility that because under this assumption the proof is simpler. So we will do proof.

Now, before going to that, now let us see. So, it says that these are if and only, but see if we just prove that i is null recurrent implies, this is true and i is positive recurrent implies this is true, then also we get an, if and only criteria, why? Because suppose, we have shown that i is null recurrent if this is equal to infinity as if this quantity is equal to infinity and this is equal to 0 and positive recurrent if this quantity is equal to infinity and this quantity is strictly greater than 0. Now, for the other side, so suppose we want to show that this plus this implies that it is null recurrent. Now suppose if it is not null recurrent. So, again as soon as we know that it is equal to infinity, we know that it is recurrent now suppose it is not null recurrent, but it is positive recurrent, but in three if we, but if you have already shown that positive recurrent implies that this quantity has to be strictly greater than 0, then we will get a contradiction. So, suppose we start from this, now the claim is it has to be null recurrent, if not null recurrent other option is positive recurrent, but if we have already shown that positive recurrence means this quantity has to be strictly greater than 0, but we have started, we started with that this is equal to 0. So it cannot be positive recurrent but has to be null recurrent. So, if we show this 2 and 3 only one side because this, because it is a dichotomy, a state can either be positive recurrent or null recurrent, so in order to show if and only if, if we just show one side of both the statements that is enough.

So, we proved 2 and 3 together. So, suppose we first look at, suppose i is null recurrent. Now, as i is recurrent, this is equal to infinity, we have already seen this in a previous theorem. Now, by previous theorem the previous theorem I mean this one. so it says that $L_i^{(n)}$, if say for example, if I look at for $j = i$, $L_i^{(n)}$ converges to 0 if $E_i(T_i) < \infty$ or in other words if i is null recurrent. So, $\lim_{n\to\infty} L_i^{(n)} = 0$, now I take expectation on both sides. So that will imply this. But here, I have done a small thing, so here, there is a small certainty, I will not get too much into the detail of that certainty, but I just want to mention, so say if I take expectation here, now expectation of 0 is 0, no problem but you see in this statement, what I have done is, I have now pushed the expectation inside the limit. Now, the certainty is that you can do it in this case, why you can do it in this case, I will not go into the detail of that, but I just wanted to point out that I have actually done that and here it is actually, so there is no issue in doing this. So, generally you need to whenever you are pushing expectation inside the limit or summation inside the limit you need to be careful whether you can do it or not, but in this particular case, we can do that. Why we can do that? So that is because of some theorem, but I do not want to go into the details of that theorem. So, here I have pushed the expectation inside the limit, so this implies that $\lim_{n\to\infty} E_i(L_i^{(n)}) = 0$, but again what was $L_i^{(n)}$, $L_i^{(n)}$ was $\lim_{n\to\infty} \frac{1}{n+1} \sum_{k=0}^{n} p_{ii}^{(k)}$. Now, again expectation of this, this is a finite sum, so I can actually push the expectation inside but now this $\delta_i(X_k)$ these are all random variables, what kind of a random variable, it is a very simple random variable, it is a Bernoulli random variable. So, this is equal to 1, if $X_k = i$, 0, otherwise. So, it takes 2 values 1 and 0. So, again that is why I call it a Bernoulli random variable. So, what is the expectation of a Bernoulli random variable? It is just probability that the random variable takes value 1. So, here it takes value 1, when $X_k = i$, so expectation of, so what I am saying is that, if I look at E_i of, so I am doing here E_i. So, $E_i \delta_i(X_k)$ that is nothing but $P_i(X_k = i)$ because that is precisely, when it takes the value one but what is $P_i(X_k = i)$, that is nothing but $p_{ii}^{(k)}$. So, that is why, and again this n plus 1 is, will remain as it is, so this is, so expectation of this is actually equal to this, but now you already know that this expectation is 0, so that implies $n \to \infty$ is equal to 0. So, we get this from that theorem, which we started with in today's lecture, that $L_i^{(n)} \to 0$, if i is null recurrent. So, then I take expectation on both sides of that and we get the result. Now, suppose that i is positive recurrent, so if i is positive, so for null recurrent, we have shown that this has to be 0. Now, suppose that i is positive recurrent, again as i is recurrent, this thing is true again by previous theorem, we know that in this, so if i is positive recurrent then it converges to $\frac{1}{E_i(T_i)}$. Again, I take expectation on both sides, again expectation can be pushed inside the limit. So, remember this is a constant, so expectation of a constant is just that, so $\lim_{n\to\infty} E_i(L_i^{(n)}) = 0$, but again what was $L_i^{(n)}$, $L_i^{(n)}$ is equal to this, but now as I explained just in this case, this expectation is equal to this. So, what this tells you is that $\lim_{n\to\infty} \frac{1}{n+1} \sum_{k=0}^{n} p_{ii}^{(k)} = \frac{1}{E_i(T_i)}$. But since i is positive recurrent, we know that $E_i(T_i) < \infty$. So, we know that since i is positive recurrent we know that $E_i(T_i) < \infty$, so $\frac{1}{E_i(T_i)} > 0$. So, if i is positive recurrent, we have shown that this is strictly greater than 0. Now, this also says that all these things are if and only because of the reason which I explained before starting the proof that, since now if I have to show that this plus this implies that it is null recurrent, now since this is equal to infinity, we know that this is recurrent, now we have this. Now, if it is positive recurrent, now this part tells you that this has to be positive, but we are starting with that, this is equal to 0, so it cannot be positive recurrent, but has to be null recurrent and similarly you can conclude for positive recurrent. So, these are if and only criteria, so this gives you a complete characterization of a state. So, if $\sum_{k=0}^{n} p_{ii}^{(k)} < \infty$, then it is transient. If it is equal to infinity then it is recurrent, but what kind of recurrent positive or null for that we need to look at $\lim_{n\to\infty} \frac{1}{n+1} \sum_{k=0}^{n} p_{ii}^{(k)}$. And if this limit is equal to 0, then it is null recurrent and if this limit is positive, then it is positive recurrent. So, we will stop this lecture here. Thank you all.