MASSACHUSETTS INSTITUTE OF TECHNOLOGY

6.436J/15.085J Fall 2018
Lecture 24

INFINITE MARKOV CHAINS. CONTINUOUS TIME MARKOV CHAINS.

Contents

1 INTRODUCTION

In this lecture we discuss infinite state Markov chains. Then we consider finite and infinite state M.c. where the transition between the states occurs during some random time interval, as opposed to unit time steps. Most of the times we state the results without proofs. Our treatment of this material is also very brief. A more in depth analysis of these concepts is devoted by the course 6.262 - Discrete Stochastic Processes.

2 INFINITE STATE MARKOV CHAINS

Suppose we have a (homogeneous) Markov chain whose state space is countably infinite $\mathcal{X} = \{0, 1, 2, \ldots\}$. In this case the theory is similar in some respects to the finite state counterpart, but different in other respects. We denote again by $p_{i,j}$ the probability of transition from state i to state j. Thus we will consider only homogeneous Markov chains, without explicitly saying this. We introduce the notion of i communicates with j, written as $i \to j$, in the same manner as before. Thus again we may decompose the state space into transient states i, namely states such that for some j, $i \to j$ but $j \nrightarrow i$; and the remaining states which are recurrent. However, in the case of infinite M.c. a new complication appears. To discuss it, let us again define a probability distribution π on $\mathcal{X}$ to be stationary if it is time invariant. The necessary and sufficient condition for this is $\pi_i \geq 0, \sum_{i \in \mathcal{X}} \pi_i = 1$ and for every state i

$$\pi_i = \sum_j \pi_j p_{j,i}.$$

As a result, if the M.c. X_n has the property $X_0 \stackrel{d}{=} \pi$, then $X_n \stackrel{d}{=} \pi$ for all n.

Now let us consider the following M.c. on $\mathbb{Z}_+$. A parameter p is fixed. For every integer $i > 0$, $p_{i,i+1} = p, p_{i,i-1} = 1-p$ and $p_{0,1} = p, p_{0,0} = 1-p$. This M.c. is called *random walk with reflection at zero*. Let us try to find a stationary distribution π of this M.c. It must satisfy

$$\pi_i = \pi_{i-1} p_{i-1,i} + \pi_{i+1} p_{i+1,i} = \pi_{i-1} p + \pi_{i+1}(1-p), \quad i \geq 1.$$
$$\pi_0 = \pi_0(1-p) + \pi_1(1-p).$$

From this we obtain $\pi_1 = \frac{p}{1-p}\pi_0$ and iterating

$$\pi_{i+1} = \frac{p}{1-p}\pi_i. \tag{1}$$

This gives $\pi_i = (p/(1-p))^i \pi_0$. Now if $p > 1/2$ then $\pi_i \to \infty$ and we cannot possibly have that $\sum_i \pi_i = 1$. Thus no stationary distribution exists. Note, that however all pairs of states i, j communicate, as we can get from i to $j > i$ in $j-i$ steps with probability $p^{j-i} > 0$, and from j to i in $j-i$ steps with probability $(1-p)^{j-i}$.

We conclude that an infinite state M.c. does not necessarily have a stationary distribution. Recall that in the case of finite state M.c. if i is a recurrent state, then its recurrence time T_i has finite expected value (as it has geometrically decreasing tails). For the case of infinite M.c. the difficulty is the fact that while every state i communicates with every other state j, it is possible that the chain starting from i wanders off to "infinity" for ever without ever returning to i. Furthermore, it is possible that even if the chain returns to i infinitely often with probability one, the *expected* return time from i to i is infinite. Recall, that the return time is defined to be $T_i = \min\{n \geq 1 : X_n = i\}$, when the M.c. starts at i at time 0, when such n exists, and defined to be $T_i = \infty$ when the chain never returns to i.

Definition 1. *Given an infinite M.c.* $X_n, n \geq 1$*, the state* i *is defined to be* transient *if the probability of never returning to* i *is positive. Namely,*

$$\mathbb{P}(X_n \neq i, \forall n \geq 1 | X_0 = i) > 0.$$

Otherwise the state is defined to be recurrent*. It is defined to be* positive recurrent *if* $\mathbb{E}[T_i] < \infty$ *and* null-recurrent *if* $\mathbb{E}[T_i] = \infty$.

Thus, unlike the finite state case, the state is transient if there is a positive probability of no return, as opposed to existence of a state from which the return to starting state has probability zero. It is an exercise to see that the definition above when applied to the finite state case is consistent with the earlier definition. Namely, it is an *implication* of how we defined the transient and recurrent states, rather than the definition. Also, observe that there are no null-recurrent states in the finite state case.

The following theorem holds, the proof of which we skip.

Theorem 1. *Given an infinite M.c.* $X_n, n \geq 1$ *suppose all the states communicate. Then there exists a stationary distribution* π *iff there exists at least one positive recurrent state* i*. In this case in fact all the states are positive recurrent and the stationary distribution* π *is unique. It is given as* $\pi_j = 1/\mathbb{E}[T_j] > 0$ *for every state* j.

We see that in the case when all the states communicate, all states have the same status: positive recurrent, null recurrent or transient. In this case we will say the M.c. itself is positive recurrent, null recurrent, or transient. There is an extension of this theorem to the cases when not all states communicate, but we skip the discussion of those. Similarly, if there are several communicating classes, then there exists at least one stationary distribution per class *which contains at least one positive recurrent state (and as a result all states in the class are positive recurrent).*

Theorem 2. *A random walk with reflection* X_n *on* $\mathbb{Z}_+$ *is positive recurrent if* $p < 1/2$*, null-recurrent if* $p = 1/2$ *and transient if* $p > 1/2$.

Proof. The case $p < 1/2$ will be resolved by exhibiting explicitly at least one steady state distribution π. Since all the states communicate, then by Theorem 1 we know that the stationary distribution is unique and $\mathbb{E}[T_i] = 1/\pi_i < \infty$ for all i. Thus the chain is positive recurrent. In fact we can find explicitly the stationary distribution. Consider again at the recurrence (1), which suggests $\pi_i = (p/(1-p))^i \pi_0$. From this we obtain

$$\pi_0(1 + \sum_{i>0} (p/(1-p))^i) = 1$$

implying $\pi_0 = 1 - p/(1-p) = (1-2p)/(1-p)$ and

$$\pi_i = \frac{1-2p}{1-p}\Big(\frac{p}{1-p}\Big)^i, \;\; i \geq 0.$$

This gives us a probability vector π with $\sum_i \pi_i = 1$ and completes the proof for the case $p < 1/2$.

The case $p \geq 1/2$ will be analyzed using our earlier result on random walk on $\mathbb{Z}$. Recall that for such a r.w. the probability of return to zero is $= 1$ iff $p = 1/2$. In the case $p = 1/2$ we have also established that the expected return time to zero is infinite. Thus suppose $p = 1/2$. A r.w. without reflection makes the first step into 1 or -1 with probability $1/2$ each. Conditioning on $X_1 = 1$ and conditioning on $X_1 = -1$, we have that the expected return time to zero is again infinite. If the first transition is into 1, then the behavior of this r.w. till the first return to zero is the same as of our r.w. with reflection at zero. In particular, the return to zero happens with probability one and the expected return time is infinite. We conclude that the state 0 is null-recurrent.

Finally, suppose $p > 1/2$. We already saw that the M.c. cannot have a stationary distribution. Thus by Theorem 1, since all the states communicate we have that all states are null-recurrent or transient. We just need to refine this result to show that in fact all states are transient.

For the unreflected r.w. we have that with positive probability the walk never returns to zero. Let, T_0 denote return time to 0 - the time it takes to come back to zero for unreflected random walk, when it at zero. We claim that $\mathbb{P}(T_0 = \infty | X_1 = 1) > 0, \mathbb{P}(T_0 = \infty | X_1 = -1) = 0$. Namely, the "no return to zero" happens iff the first step is to the right. First let us see why just the first inequality, namely $\mathbb{P}(T_0 = \infty | X_1 = 1) > 0$ implies our result. Conditioned on the event that the first step is to the right, the r.w. behaves as r.w. with reflection at zero until the first return to zero. The assumption $\mathbb{P}(T_0 = \infty | X_1 = 1) > 0$ means there is a positive probability of no return for random walk without reflection when the first step is to the right. Then there is a positive probability of no return for the reflected r.w. conditioned on $X_1 = 1$. Since the transition from zero to 1 occurs with positive probability p, then there is a positive probability of no return to zero starting from zero for the random walk with reflection and thus state 0 is transient. Since all states communicate, this means that all states of the random walk with reflection are transient.

Now we establish that claim. We have $\mathbb{P}(T_0 = \infty) = p\mathbb{P}(T_0 = \infty | X_1 = 1) + (1-p)\mathbb{P}(T_0 = \infty | X_1 = -1)$. We also have that $\mathbb{P}(T_0 = \infty) > 0$. We now establish that $\mathbb{P}(T_0 = \infty | X_1 = -1) = 0$. This immediately implies $\mathbb{P}(T_0 = \infty | X_1 = 1) > 0$, which we need. Now assume $X_1 = -1$. Consider $Y_n = -X_n$. Observe that, until the first return to zero, Y_n is a reflected r.w. with parameter $q = 1 - p$. Since $q < 1/2$, then, as we established at the beginning of the proof, the process Y_n returns to zero with probability one (moreover the return time has finite expected value). We conclude that X_n returns to zero with probability one, namely $\mathbb{P}(T_0 = \infty | X_1 = -1) = 0$. This completes the proof. $\square$

3 CONTINUOUS TIME MARKOV CHAINS

We consider a stochastic process $X(t)$ which is a function of a real argument t instead of integer n. Let $\mathcal{X}$ be the state space of this process, which is assumed to be finite or countably infinite.

Definition 2. *$X(t)$ is defined to be a continuous time Markov chain if for every $j, i_1, \ldots, i_{n-1} \in \mathcal{X}$ and every sequence of times $t_1 < t_2 < \cdots < t_n$,*

$$\begin{aligned} &\mathbb{P}(X(t_n) = j | X(t_{n-1}) = i_{n-1}, \ldots, X(t_1) = i_1) && (2) \\ &\quad = \mathbb{P}(X(t_n) = j | X(t_{n-1}) = i_{n-1}). && (3) \end{aligned}$$

The process is defined to be homogeneous if $\mathbb{P}(X(t) = j | X(s) = i) = \mathbb{P}(X(t-s) = j | X(0) = i)$ for every i, j and $s < t$.

From now on we assume without explicitly saying that our M.c. is homogeneous. We write $p_{i,j}^{(t)}$ for $\mathbb{P}(X(t) = j | X(0) = i)$. The continuous time Markov chain is a special case of a Markov process, the definition of which we skip. Loosely speaking, a stochastic process is a Markov process if its future trajectory is completely determined by its current state, independently from the past. We already know an example of a continuous time M.c. - Poisson process. It is given as $\mathbb{P}(X(t) = i + k | X(s) = i) \stackrel{d}{=} \text{Pois}(\lambda(t-s)), k \geq 0$ and $\mathbb{P}(X(t) = i + k | X(s) = i) = 0$ for $k < 0$.

Given a state i and time t_0 introduce "holding time" $U(i,t_0)$ as $\inf\{s>0 : X(t_0+s)\neq i\}$, when $X(t_0)=i$. Namely, it is the time that the chain spends in state i after time t_0, assuming that it is in i at time t_0. It might turn out in special cases that $U(i,t_0)=0$ with positive probability. But in many special cases this will not happen. For now we assume that $U(i,t_0)>0$ a.s. In special cases we can establish this directly.

Proposition 1. *For every state i and time t_0, $U(i,t_0)\overset{d}{=}\mathrm{Exp}(\mu_i)$ for some parameter μ_i which depends only on the state.*

Since, per proposition above, the distribution of holding time is exponential, and therefore memoryless, we see that the time till the next transition occurs is independent from the past history of the chain and only depends on the current state i. The parameter μ_i is usually called transition rate out of state i. This is a very fundamental (and useful) property of continuous time Markov chains.

Proof sketch. Consider

$$\mathbb{P}(U(i,t_0)>x+y|U(i,t_0)>x,X(t_0)=i).$$

The event $U(i,t_0)>x,X(t_0)=i$ implies in particular $X(t_0+x)=i$. Since we have a M.c. the trajectory of $X(t)$ for $t\geq t_0+x$ depends only on the state at time t_0+x which is i in our case. Namely

$$\mathbb{P}(U(i,t_0)>x+y|U(i,t_0)>x,X(t_0)=i)=\mathbb{P}(U(i,t_0+x)>y|X(t_0+x)=i).$$

But the latter expression by homogeneity is $\mathbb{P}(U(i,t_0)>y|X(t_0)=i)$, as it is the probability of the holding time being larger than y when the current state is i. We conclude that

$$\mathbb{P}(U(i,t_0)>x+y|U(i,t_0)>x,X(t_0)=i)=\mathbb{P}(U(i,t_0)>y|X(t_0)=i),$$

namely

$$\mathbb{P}(U(i,t_0)>x+y|X(t_0)=i)=\mathbb{P}(U(i,t_0)>y|X(t_0)=i)\mathbb{P}(U(i,t_0)>x|X(t_0)=i).$$

Since the exponential function is the only one satisfying this property, then $U(i,t_0)$ must be exponentially distributed. □

There is an omitted subtlety in the proof. We assumed that for every $t,z>0$ and state i, $\mathbb{P}(X(t+s)=i,\forall\, s\in[0,z]|X(t)=i,\Im_t)=\mathbb{P}(X(t+s)=i,\forall\, s\in$

$[0, z]|X(t) = i)$ where $\Im_t$ denotes the history of the process up to time t. We deduced this based on the assumption (2). This requires a technical proof, which we ignored above.

Thus the evolution of a continuous M.c. $X(t)$ can be described as follows. It stays in a given state i during some exponentially distributed time U_i, with parameter μ_i which only depends on the state. After this time it makes a transition to the next state j. If we consider the process only at the random times of transitions, denoted say by $t_1 < t_2 < \cdots$, then we obtain an *embedded* discrete time process $Y_n = X(t_n)$. It is an exercise to show that Y_n is in fact a homogeneous Markov chain. Denote the transition rates of this Markov chain by $p_{i,j}$. The value $q_{i,j} = \mu_i p_{i,j}$ is called "transition rate" from state i to state j. Note, that the values $p_{i,j}$ were introduced only for $j \neq i$, as they were derived from M.c. changing its state. Define $q_{i,i} = -\sum_{j \neq i} q_{i,j}$. The matrix $G = (q_{i,j}), i, j \in \mathcal{X}$ is defined to be the **generator** of the M.c. $X(t)$ and plays an important role, specifically for the discussion of a stationary distribution.

A stationary distribution π of a continuous M.c. is defined in the same way as for the discrete time case: it is the distribution which is time invariant. The following fact can be established.

Proposition 2. *A vector $(\pi_i), i \in \mathcal{X}$ is a stationary distribution iff $\pi_i \geq 0, \sum_i \pi_i = 1$ and $\sum_j \pi_j q_{j,i} = 0$ for every state i. In vector form $\pi^T G = 0$.*

As for the discrete time case, the theory of continuous time M.c. has a lot of special structure when the state space is finite. We now summarize without proofs some of the basic results. First there always exists a stationary distribution. The condition for uniqueness of the stationary distribution is the same - single recurrence class, with communications between the states defined similarly. A nice "advantage" of continuous M.c. is the lack of periodicity. There is no notion of a period of a state. Moreover, and most importantly, suppose the chain has a unique recurrence class. Then, letting π denote the corresponding unique stationary distribution, the mixing property

$$\lim_{t \to \infty} p_{i,j}^{(t)} = \pi_j$$

holds for all states i, j. For the modeling purposes, it is useful sometimes to consider a continuous as opposed to a discrete M.c.

There is an alternative way to describe a continuous M.c. and the embedded discrete time M.c. Assume that to each pair of states i, j we associate an exponential "clock" - exponentially distributed r.v. $U_{i,j}$ with parameter $\mu_{i,j}$. Each time the process jumps into i all of the clocks turned on simultaneously. Then at time $U_i \triangleq \min U_{i,j}$ the process jumps into state $j = \arg\min_j U_{i,j}$. It is not hard to establish the following: the resulting process is a continuous time finite state M.c. The embedded discrete time M.c. has then transition probabilities $\mathbb{P}(X(t_{n+1}) = j | X(t_n) = i) = \frac{\mu_{i,j}}{\sum_k \mu_{i,k}}$, as the probability that $U_{i,j} = \min_k U_{i,k}$ is given by this expression, when the distribution of $U_{i,j}$ is exponential with parameters $\mu_{i,j}$. The holding time has then the distribution $\mathrm{Exp}(\mu_i)$ where $\mu_i = \sum_k \mu_{i,k}$. Thus we obtain an alternative description of a M.c. The transition rates of this M.c. are $q_{i,j} = \mu_i p_{i,j} = \mu_{i,j}$. In other words, we described the M.c. via the rates $q_{i,j}$ as given.

This description extends to the infinite M.c., when the notion of holding times is well defined (see the comments above).

References

[1] G. R. Grimmett and D. R. Stirzaker, *Probability and Random Processes*, Oxford University Press, 3rd edition, 2001.

[2] D. P. Bertsekas and J. N. Tsitsiklis, *Introduction to probability,* Athena Scientific, 2002.