AI2613 随机过程	2020–2021 春季学期

Lecture 10 – Continuous-Time Markov Chains

2021 年 5 月 8 日

Lecturer: 张驰豪 *Scribe*: 杨宽

Review of Syllabus

Markovian Process:

	discrete space	continuous space
discrete time	DTMC	Langevin Dynamics
continuous time	Poisson / CTMC	Brownian Motion

Non-Markovian Process:

Martingale

1 Definition of CTMC

Recall that for a (discrete-time) Markov chain, we have

$$\mathbf{Pr}\left[X_{t+1} = j \mid X_t = i_t, X_t = i_{t-1}, \dots, X_0 = i_0\right] = \mathbf{Pr}\left[X_{t+1} = j \mid X_t = i_t\right].$$

Similarly, it is easy to define the following *continuous-time* stochastic process with the lack of memory (Markovian) property.

Definition 1 (Continuous-Time Markov Chain). We say $X(t)$ is a *continuous-time Markov chain* if $\forall\, s, t \geq 0$, $\forall\, 0 < s_0 < \cdots < s_n < s$,

$$\mathbf{Pr}\left[X_{s+t} = j \mid X_s = i, X_{s_0} = i_0, \dots, X_{s_n} = i_n\right] = \mathbf{Pr}\left[X_{s+t} = j \mid X_s = i\right] = \mathbf{Pr}\left[X_t = j \mid X_0 = i\right].$$

We now introduce a significant continuous-time Markov chain.

Example 2 (Continuous-Time Markov Chain described by Poisson Process). Suppose $\{Y_n\}$ is a discrete-time Markov chain with transition probability given by $u(i, j)$, and $N(t)$ is a Poisson process with rate λ. If $\{Y_n\}$ and $N(t)$ are two independent processes, then

$$X(t) = Y_{N(t)}$$

is a continuous-time Markov chain.

Remark. In fact, we will show later that every continuous-time Markov chain has a definition described by a Poisson process.

Intuitively, we can imagine that $X(t)$ is still a Poisson process, where each arrival is associated with a random variable that only depends on the last arrival. It is clear to verify that $X(t)$ is indeed a continuous-time Markov chain.
Recall that discrete-time Markov chains can be described by their transition probabilities $\mathbf{P}(i, j)$. We now similarly introduce the transition probability in the continuous-time case.

Definition 3 (Transition Probability). The *transition probability* $\mathbf{P}_t(i, j)$ is defined as

$$\mathbf{P}_t(i, j) = \mathbf{Pr}\left[X(t) = j \mid X(0) = i\right].$$

Note that in the discrete-time case, the transition probabilities after t steps are given by $\mathbf{P}^t$. In the continuous-time case we also have a similar result.

Theorem 1 (Chapman–Kolmogorov Equation). $\mathbf{P}_{s+t} = \mathbf{P}_s \cdot \mathbf{P}_t$. *That is,*

$$\mathbf{P}_{s+t}(i, j) = \sum_{k\in\mathcal{S}} \mathbf{P}_s(i, k) \cdot \mathbf{P}_t(k, j).$$

Proof. The proof is applying straightforwardly the law of total probability and the Markovian property.

$$\begin{aligned}
\mathbf{P}_{s+t}(i, j) &= \mathbf{Pr}\left[X(s+t) = j \mid X(0) = i\right] \\
&= \sum_k \mathbf{Pr}\left[X(s+t) = j, X(s) = k \mid X(0) = i\right] \\
&= \sum_k \mathbf{Pr}\left[X(s+t) = j \mid X(s) = k, X(0) = i\right] \cdot \mathbf{Pr}[X(s) = k \mid X(0) = i] \\
&= \sum_k \mathbf{P}_t(k, j) \cdot \mathbf{P}_s(i, k).
\end{aligned}$$

□

Chapman–Kolmogorov Equation shows that if there exists $t_0 > 0$ such that we know the transition probability for all $t < t_0$, we know it for all $t > 0$. This observation suggests that (roughly speaking) the probabilities $\mathbf{P}_h(i, j)$ where $h \to 0$ play the key role. We now introduce and will justify later the following quantity which determines the transition probabilities.

Definition 4 (Jump Rate). The *jump rate* from i to j (where $i \neq j$) is given by

$$q(i,j) \triangleq \lim_{h \to 0} \frac{\mathbf{P}_h(i,j)}{h}$$

if the limit exists.

As an example, we now compute the jump rates for the continuous-time Markov chain defined in Example 2.

For any $h > 0$, we have

$$\mathbf{P}_h(i,j) = \sum_{n=0}^{\infty} \mathrm{e}^{-\lambda h} \cdot \frac{(\lambda h)^n}{n!} \cdot u^n(i,j)\,.$$

Remark. Note that here u is actually a matrix, so $u^n(i,j)$ is not $\big(u(i,j)\big)^n$, but $\big(u^n\big)(i,j)$.

Since $u^0(i,j) = 0$ if $i \neq j$, it follows that

$$\mathbf{P}_h(i,j) = \mathrm{e}^{-\lambda h} \cdot \lambda h \cdot u(i,j) + (\lambda h)^2 \sum_{n=2}^{\infty} \mathrm{e}^{-\lambda h} \cdot \frac{(\lambda h)^{n-2}}{n!} \cdot u^n(i,j)\,.$$

Thus,

$$\lim_{h \to 0} \frac{\mathbf{P}_h(i,j)}{h} = \lim_{h \to 0} \mathrm{e}^{-\lambda h} \cdot \lambda \cdot u(i,j) + \lambda^2 h \sum_{n=2}^{\infty} \mathrm{e}^{-\lambda h} \cdot \frac{(\lambda h)^{n-2}}{n!} \cdot u^n(i,j) = \lambda \cdot u(i,j)\,.$$

Then we compute jump rates in other examples. We start from a simple one.

Example 5 (Poisson Process). Let $X(t)$ be the number of total arrivals up to time t in a Poisson process with rate λ.

It is clear that $q(n, n+1) = \lambda$ and $q(i,j) = 0$ otherwise.

Recall that we introduce the notation of queueing theory in the last lecture. We consider the following queue.

Example 6 ($M/M/s$ Queue). $M/M/s$ stands for a model where each customer arrives according to a Poisson process, the end of servings are also Poisson processes (i.e., serving times have exponential distributions), and the number of servers is s.

Suppose that the Poisson process of arrivals has rate λ and serving times have exponential distributions with rate μ. We would like to compute the number of customers being served or waiting in the queue.

Similarly to Example 5 we have $q(n, n+1) = \lambda$. But there is one more case in this example. The number of customers being served or waiting in the queue may decrease — if someone completes the serving and leaves.

Note that the number of customers being served is $\min\{n,s\}$ if there is n customers. So the first leaving occurs after the minimum time among $\min\{n,s\}$ exponential random times. Recall the result of the *exponential races*. Let $X_1 \sim \mathsf{Exponential}(\lambda_1)$ and $X_2 \sim \mathsf{Exponential}(\lambda_2)$ be two independent random variables. Then $Y \triangleq \min\{X_1, X_2\} \sim \mathsf{Exponential}(\lambda_1+\lambda_2)$. So it follows that

$$q(n,n-1) = \mu \cdot \min\{n,s\}.$$

2 Constructing a CTMC with Given Jump Rates

Given jump rates, a natural question is to ask if we can construct a continuous-time Markov chain. In other words, we would like to check if jump rates indeed contain sufficient information on a continuous-time Markov chain so that if we know the jump rates then we can recover the CTMC.

Intuitively, consider the CTMC in Example 2. We already know that the jump rates from i to j is $\lambda \cdot u(i,j)$. So we may construct a CTMC by choosing λ and $u(i,j)$ properly.

However, there is a technical problem. We define $q(i,j)$ only for $i \neq j$, but we should choose a proper $u(i,i)$ for each i.

Recall that in Example 2, the transition of the CTMC has 2 steps: at any time t and state $X(t) = i$, we first choose $s \sim \mathsf{Exponential}(\lambda)$ to determine the next jump time and then choose the next state $X(t+s) \sim u(i,\cdot)$.

Note that if the next state is i, it is equivalent to stay still. So for each state i, we may choose distinct λ_i so that each jump will move to a state different from the current state. Now we are ready to begin our construction.

Let $\lambda_i = \sum_{j\neq i} q(i,j)$, and

$$u(i,j) \triangleq \frac{q(i,j)}{\lambda_i}.$$

It is clear that λ_i and $u(i,j)$ are proper ones. We should note here that different i may have different λ_i. If we require that λ_i are identical, we may choose $\Lambda = \sup_i \lambda_i$ and add self-loops in the Markov chain (i.e., choose $u(i,i) > 0$).

Note that we are cheating here. We only computed the jump rates of the CTMC in Example 2 with identical λ. We haven't know jump rates if λ are distinct for different states. But the proof is easy and we do not show it rigorously here.

Now we've constructed a CTMC with given jump rates. Then another natural question is how to compute the transition probabilities. Given $i \neq j$, if we view $\mathbf{P}_t(i,j)$ as a function of t, it is clear that $q(i,j)$ is its derivative at $t=0$. We now consider $\mathbf{P}_{t+h}(i,j)-\mathbf{P}_t(i,j)$. Applying the Chapman–Kolmogorov Equation we have

$$\begin{aligned}\mathbf{P}_{t+h}(i,j)-\mathbf{P}_t(i,j) &= \sum_k \mathbf{P}_h(i,k)\cdot\mathbf{P}_t(k,j)-\mathbf{P}_t(i,j)\\ &= \sum_{k\neq i}\mathbf{P}_h(i,k)\cdot\mathbf{P}_t(k,j)+\big(\mathbf{P}_h(i,i)-1\big)\cdot\mathbf{P}_t(i,j)\,.\end{aligned}$$

It implies that

$$\lim_{h\to 0}\frac{1}{h}\cdot\Big(\mathbf{P}_{t+h}(i,j)-\mathbf{P}_t(i,j)\Big)=\lim_{h\to 0}\frac{1}{h}\cdot\Big(\sum_{k\neq i}\mathbf{P}_h(i,k)\cdot\mathbf{P}_t(k,j)+\big(\mathbf{P}_h(i,i)-1\big)\cdot\mathbf{P}_t(i,j)\Big)\,.$$

Note that

$$\mathsf{LHS}=\frac{\mathrm{d}\mathbf{P}_t(i,j)}{\mathrm{d}t}\,.$$

Let $A=\sum_{k\neq i}\mathbf{P}_h(i,k)\cdot\mathbf{P}_t(k,j)$, $B=\big(\mathbf{P}_h(i,i)-1\big)\cdot\mathbf{P}_t(i,j)$, and $\mathbf{P}'_t(i,j)$ denote $\frac{\mathrm{d}\mathbf{P}_t(i,j)}{\mathrm{d}t}$. Then we have

$$\begin{aligned}\lim_{h\to 0}\frac{1}{h}\cdot A &= \sum_{k\neq i}\mathbf{P}_t(k,j)\cdot\lim_{h\to 0}\frac{1}{h}\cdot\mathbf{P}_h(i,k)\\ &= \sum_{k\neq i}q(i,k)\cdot\mathbf{P}_t(k,j)\,,\\ \lim_{h\to 0}\frac{1}{h}\cdot B &= \lim_{h\to 0}\frac{1}{h}\Big(-\sum_{k\neq i}\mathbf{P}_h(i,k)\Big)\cdot\mathbf{P}_t(k,j)\\ &= -\sum_{k\neq i}q(i,k)\cdot\mathbf{P}_t(i,j)\ \ = -\lambda_i\cdot\mathbf{P}_t(i,j)\,.\end{aligned}$$

It follows that

$$\mathbf{P}'_t(i,j)=\sum_{k\neq i}q(i,k)\cdot\mathbf{P}_t(k,j)-\lambda_i\cdot\mathbf{P}_t(i,j)\,.$$

Let $\mathbf{Q}$ be a matrix where $\mathbf{Q}(i,j)=q(i,j)$ if $i\neq j$, and $\mathbf{Q}(i,i)=-\lambda_i$ for all i. The equation above can be written as

$$\mathbf{P}'_t=\mathbf{Q}\cdot\mathbf{P}_t\,.$$

If $\mathbf{P}_t$ were a function of a real number t and $\mathbf{Q}$ were a constant number, we could have

$$\frac{\mathbf{P}'_t}{\mathbf{P}_t}=\mathbf{Q}\implies\int\frac{\mathbf{P}'_t}{\mathbf{P}_t}\mathrm{d}t=\int\mathbf{Q}\,\mathrm{d}t\implies\ln\mathbf{P}_t=\mathbf{Q}t+\mathbf{C}\implies\mathbf{P}_t=\mathrm{e}^{\mathbf{Q}t+\mathbf{C}}\,.$$

Although the argument does not hold for matrix, the result remains the same, where the exponential of a matrix $\mathbf{M}$ is given by

$$\mathrm{e}^{\mathbf{M}}=\sum_{n=0}^{\infty}\frac{\mathbf{M}^n}{n!}\,.$$

Since $\mathbf{P}_0$ is the identity matrix, we have $\mathbf{C} = \mathbf{0}$, and thus

$$\mathbf{P}_t = \mathrm{e}^{\mathbf{Q}t} = \sum_{n=0}^{\infty} \frac{(t \cdot \mathbf{Q})^n}{n!}.$$

Remark. We should point out here that the real exponential function $\exp(\cdot)$ can be characterized in a variety of equivalent ways, but the above definition given by power series might be the most common and the most natural one. The above power series always converges, even for matrix. This justifies that the exponential of $\mathbf{M}$ is well-defined. The definition is also compatible with the ordinary exponential function since if $\mathbf{M}$ is a 1×1 matrix, the matrix exponential of $\mathbf{M}$ is a 1×1 matrix whose single element is the ordinary exponential of the single element of $\mathbf{M}$.

According to the above discussion, we obtain the following result.

Theorem 2 (Kolmogorov's Backward Equation).

$$\mathbf{P}_t' = \mathbf{Q} \cdot \mathbf{P}_t.$$

Similarly, we also have the forward equation.

Theorem 3 (Kolmogorov's Forward Equation).

$$\mathbf{P}_t' = \mathbf{P}_t \cdot \mathbf{Q}.$$

These two equations also show that $\mathbf{Q} \cdot \mathbf{P}_t$ is commutative. Note that $\mathbf{P}_t$ and $\mathbf{Q}$ are not invertible so the result is not trivial.

Example 7 (Poisson Process). Let $X(t)$ be the number of total arrivals up to time t in a Poisson process with rate λ.

We've already known in Example 5 that its jump rates are $q(i,j) = \lambda$ if $j = i+1$ and $q(i,j) = 0$ otherwise. We now compute $\mathbf{P}_t$ and verify Kolmogorov's backward equation. We first compute them directly, that is,

$$\mathbf{P}_t(i,j) = \mathrm{e}^{-\lambda t} \cdot \frac{(\lambda t)^{j-i}}{(j-i)!},$$

$$\mathbf{P}_t'(i,j) = \begin{cases} -\lambda \mathrm{e}^{-\lambda t}, & \text{if } i = j; \\ -\lambda \mathrm{e}^{-\lambda t} \cdot \frac{(\lambda t)^{j-i}}{(j-i)!} + \mathrm{e}^{-\lambda t} \cdot \frac{\lambda^{j-i} t^{j-i-1}}{(j-i-1)!}, & \text{otherwise}. \end{cases}$$

On the other hand, Kolmogorov's backward equation shows that

$$\mathbf{P}_t'(i,j) = \lambda \cdot \mathbf{P}_t(i+1,j) - \lambda \cdot \mathbf{P}_t(i,j).$$

It is easy to verify the backward equation.

Example 8 (Two-state Markov Chain). Consider the following Markov chain: $\mathcal{S} = \{1,2\}$, $q(1,2) = \lambda$ and $q(2,1) = \mu$. Then it is clear that

$$\mathbf{Q} = \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix}.$$

Our goal is to compute $\mathbf{P}_t(1,2)$.

Applying Kolmogorov's backward equation, we have

$$\begin{pmatrix} \mathbf{P}'_t(1,1) & \mathbf{P}'_t(1,2) \\ \mathbf{P}'_t(2,1) & \mathbf{P}'_t(2,2) \end{pmatrix} = \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix} \begin{pmatrix} \mathbf{P}_t(1,1) & \mathbf{P}_t(1,2) \\ \mathbf{P}_t(2,1) & \mathbf{P}_t(2,2) \end{pmatrix}.$$

It implies that

$$\begin{cases} \mathbf{P}'_t(1,1) = -\lambda \cdot \mathbf{P}_t(1,1) + \lambda \cdot \mathbf{P}_t(2,1) \\ \mathbf{P}'_t(2,1) = \mu \cdot \mathbf{P}_t(1,1) - \mu \cdot \mathbf{P}_t(2,1) \end{cases}.$$

Let $f(t) = \mathbf{P}'_t(1,1) - \mathbf{P}'_t(2,1)$. It follows that

$$f'(t) = (-\lambda - \mu) \cdot f(t).$$

Hence we have

$$\ln f(t) = \int \frac{f'(t)}{f(t)} \, \mathrm{d}t = \int (-\lambda - \mu) \, \mathrm{d}t = (-\lambda - \mu)t + c.$$

Plugging in $f(0) = \mathbf{P}_0(1,1) - \mathbf{P}_0(2,1) = 1$, it yields that $c = 0$ and thus

$$\mathbf{P}_t(1,1) - \mathbf{P}_t(2,1) = f(t) = \mathrm{e}^{-(\lambda+\mu)t}.$$

Therefore, we conclude that

$$\mathbf{P}_t(1,1) = \mathbf{P}_0(1,1) - \int_0^t \mathbf{P}'_s(1,1) \, \mathrm{d}s = 1 - \int_0^t -\lambda \cdot \mathrm{e}^{-(\lambda+\mu)t} \, \mathrm{d} = \frac{\mu}{\lambda+\mu} + \frac{\lambda}{\lambda+\mu} \cdot \mathrm{e}^{-(\lambda+\mu)t}.$$

3 Comparing to the DTMC

We've already introduced variety of properties in the DTMC part, such as (I) (irreducible), (A) (aperiodic), (S) (exitsing stationary distributions) and so on. We now focus on how to state these properties in the continuous case. Note that the difference between CTMC and DTMC should be their *limiting behavior*.

Definition 9 (Irreducibility). A CTMC $X(t)$ is *irreducible* if for all state i and j, there exists a finite number n and n states $i = k_1, k_2, \ldots, k_n = j$ such that

$$\forall\, 1 \le t < n, \qquad q(k_t, k_{t+1}) > 0\,.$$

Note that any CTMC is "*aperiodic*" since it has self-loops. So it does not make sense to define *aperiodicity* for CTMC. We now consider the stationary distribution.

Definition 10 (Stationary Distribution). We say π is a *stationary distribution* iff

$$\forall\, t > 0, \qquad \pi^{\mathsf{T}} \cdot \mathbf{P}_t = \pi^{\mathsf{T}}\,.$$

However, this condition is not easy to verify. As we mentioned before, we believe that jump rates characterize all information on a CTMC. So it is natural to find the following proposition.

Proposition 4. *π is a stationary distribution iff*

$$\pi^{\mathsf{T}} \cdot \mathbf{Q} = \mathbf{0}\,. \tag{♠}$$

Intuitively, equation (♠) is equivalent to

$$\forall\, j, \qquad \sum_{i \ne j} \pi(i) \cdot q(i,j) - \lambda_j \cdot \pi(j) = 0\,.$$

So it suffices to show that

$$\forall\, j, \qquad \sum_{i \ne j} \pi(i) \cdot q(i,j) = \lambda_j \cdot \pi(j)\,.$$

The LHS is the total rates coming into j, and the RHS is the total rates going out of j. Thus π is the stationary distribution iff LHS = RHS.

Proof. We first show that if π is a stationary distribution, then equation (♠) holds. Since π is a stationary distribution, we have

$$\forall\, j, \qquad \pi^{\mathsf{T}} \cdot \mathbf{P}_t(j) = \pi(j)\,.$$

Taking the derivative with respect to t on both sides, it follows that

$$\sum_i \pi(i) \cdot \mathbf{P}_t'(i,j) = 0\,.$$

Applying Kolmogorov's forward equation, we obtain that

$$
\begin{aligned}
0 &= \sum_i \pi(i) \sum_k \mathbf{P}_t(i,k) \cdot \mathbf{Q}(k,j) \\
&= \sum_k \sum_i \pi(i) \cdot \mathbf{P}_t(i,k) \cdot \mathbf{Q}(k,j) \\
&= \sum_k \pi(k) \cdot \mathbf{Q}(k,j) \,.
\end{aligned}
$$

Next, we consider the other direction. Taking the derivative of $\pi^\top \cdot \mathbf{P}_t(j)$ with respect to t, and applying Kolmogorov's backward equation, we have that

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} \pi^\top \cdot \mathbf{P}_t(j) &= \sum_i \pi(i) \cdot \mathbf{P}'_t(i,j) \\
&= \sum_i \pi(i) \sum_k \mathbf{Q}(i,k) \cdot \mathbf{P}_t(k,j) \\
&= \sum_k \mathbf{P}_t(k,j) \sum_i \pi(i) \cdot \mathbf{Q}(i,k) \ = 0 \,.
\end{aligned}
$$

It justifies that $\pi^\top \cdot \mathbf{P}_t(j)$ is a constant for any t. Since $\pi^\top \cdot \mathbf{P}_0(j) = \pi(j)$, we conclude that $\pi^\top \cdot \mathbf{P}_t(j) = \pi(j)$ for all j. □

Now, similarly to the fundamental theorem of Markov chains in the discrete case, a natural question is to ask under which condition a CTMC has the unique stationary and converges to it.

Theorem 5. *If $X(t)$ is irreducible and has a stationary distribution π, then*

$$
\lim_{t \to \infty} \mathbf{P}_t(i,j) = \pi(j) \,.
$$

Note that the difference between this theorem and the fundamental theorem of Markov chains is that we no longer require the Markov chain to be aperiodic.
Finally, let's review the detailed balanced condition. We say a DTMC is reversible if

$$
\forall\, x, y, \qquad \pi(x) \cdot \mathbf{P}(x,y) = \pi(y) \cdot \mathbf{P}(y,x) \,.
$$

The equation above is called the detailed balanced condition. In the continuous case, we have a similar result as well.

Definition 11 (Detailed Balanced Condition for CTMC). We say a distribution π satisfies the *detailed balanced condition for CTMC* if

$$
\forall\, i, j, \qquad \pi(i) \cdot q(i,j) = \pi(j) \cdot q(j,i) \,.
$$

Theorem 6. *If a distribution π satisfies the detailed balanced condition for CTMC, π is the stationary distribution.*

Proof. Fix j. It suffices to show that $\left(\pi^{\mathsf{T}} \cdot \mathbf{Q}\right)(j) = 0$. Using the detailed balanced condition, it is easy to verify that

$$\begin{aligned}
\left(\pi^{\mathsf{T}} \cdot \mathbf{Q}\right)(j) &= \sum_{i} \pi(i) \cdot \mathbf{Q}(i,j) \\
&= \sum_{i \neq j} \pi(i) \cdot q(i,j) - \lambda_j \cdot \pi(i) \\
&= \sum_{i \neq j} \pi(j) \cdot q(j,i) - \lambda_j \cdot \pi(j) \\
&= \pi(j) \cdot \left(\sum_{i \neq j} \cdot q(j,i) - \lambda_j \right) = 0 .
\end{aligned}$$

□