DMS625: Introduction to stochastic processes and their applications

Sourav Majumdar

Continuous-time Markov chain

October 14, 2024

1 Introduction

So far we have discussed Discrete-time Markov chains (DTMC) with discrete state spaces. We will now consider a generalization where time will become continuous, however the state space will continue to stay discrete.

We say $X(t), t \in [0, \infty)$, is a continuous-time Markov chain (CTMC) if it satisfies the **Markov property** given by,

$$\mathbb{P}(X(t+s) = x_{t+s}|X(s) = x_s, X(u) = x_u, 0 \leq u < s) = \mathbb{P}(X(t+s) = x_{t+s}|X(s) = x_s)$$

Here $x_i \in \mathcal{S}, 0 \leq i \leq t+s$, are the values $X(t)$ attains on the state space $\mathcal{S}$ at each time. This definition of the Markov property is similar to the one for DTMC that we considered earlier. Since in a CTMC time is continuous, this has been accounted for here in the conditioning event on the LHS. The advantages of CTMC are obvious in that now we can model continuously over time. Let us look at an example.

Example 1.1 (Poisson Process)**.** *The Poisson process $N(t)$ with rate λ is a CTMC. We need to show that $N(t)$ satisfies the Markov property,*

$$\begin{aligned}
&\mathbb{P}(N(t+s) = x_{t+s}|N(s) = x_s, N(u) = x_u, 0 \leq u < s)\\
=&\mathbb{P}(N(t+s) = x_{t+s}|N(s) = x_s, N(t_k) = x_{t_k}, N(t_{k-1}) = x_{t_{k-1}}, \dots, N(t_1) = x_{t_1}, N(0) = 0)\\
&\textit{(For an arbitrary } k\textit{, consider any sequence of times such that } 0 < t_1 < t_2 < \ldots < t_k < s\textit{)}\\
=&\mathbb{P}(N(t+s) - N(s) = x_{t+s} - x_s|N(s) - N(t_k) = x_s - x_{t_k}, N(t_k) - N(t_{k-1}) = x_{t_k} - x_{t_{k-1}}, \dots, N(t_1) - N(0) = x_{t_1} - 0)\\
=&\mathbb{P}(N(t+s) - N(s) = x_{t+s} - x_s)\\
&\textit{(Follows from the independent increments property)}\\
=&\mathbb{P}(N(t+s) = x_{t+s}|N(s) = x_s)
\end{aligned}$$

Therefore, $N(t)$ is a CTMC.

Example 1.2 (Non-homogeneous Poisson process)**.** *Consider a Non-homogeneous Poisson process N(t) with rate $\lambda(t)$. One can show that it satisfies the Markov property as above. We note that to show the Poisson process satisfied the Markov property above, we only required the independent increments property. Since the non-homogeneous Poisson process possesses the independent increments property, the proof will follow similarly here.*

We say $X(t)$ is a **time-homogeneous** CTMC, if,

$$\mathbb{P}(X(t+s) = x_{t+s} | X(s) = x_s) = \mathbb{P}(X(t) = x_{t+s} | X(0) = x_s)$$

This is similar to the time-homogeneity condition for DTMC, where we say the probability of going from one state to another only depends on the difference of the times of their transition.

Exercises

1. Verify that the Poisson process $N(t)$ with rate λ is a time-homogeneous CTMC.

2. Verify that the non-homogeneous Poisson process with rate $\lambda(t)$ need not be a time-homogeneous CTMC.

Remark 1. *All CTMCs henceforth will be assumed to be time-homogeneous CTMCs unless otherwise stated.*

2 Time spent in a state

Let T_i denote the time spent by a CTMC in state i. T_i is also referred to as the *sojourn time.* Consider a Poisson process $N(t)$ that is modelling the number of Earthquakes. Let's say the first Earthquake occurred on year $t = 1$, and the second earthquake occurred on year $t = 3$. Then the time spent by Poisson process on state $i = 1$ is $T_i = 3 - 1 = 2$ years. We want to understand the distribution of T_i for CTMCs in general.

0?

$$\begin{aligned}
\mathbb{P}(T_i > t + s | T_i > s) &= \mathbb{P}(X(v) = i, v \in (s, s + t] | X(u) = i, i \in [0, s]) \\
&= \mathbb{P}(X(v) = i, v \in (0, t] | X(0) = i) \\
&\text{(Follows from time-homogeneity)} \\
&= \mathbb{P}(T_i > t)
\end{aligned}$$

Therefore, T_i follows the memoryless property and hence T_i is exponentially distributed. Similarly consider,

$$\begin{aligned}
\mathbb{P}(T_i > t_i, T_j > t_j) = \mathbb{P}(T_j > t_j | T_i > t_i)\mathbb{P}(T_i > t_i)& \\
&\text{(Without loss of generality assume that the chain visits } i \text{ before } j) \\
&= \mathbb{P}(X(v) = j, v \in (s, s + t_j] | X(s) = j, X(u) = i, u \in (r, r + t_i])\mathbb{P}(T_i > t_i) \\
&= \mathbb{P}(X(v) = j, v \in (s, s + t_j] | X(s) = j)\mathbb{P}(T_i > t_i) \\
&\text{(Markov property)} \\
&= \mathbb{P}(T_j > t_j)\mathbb{P}(T_i > t_i)
\end{aligned}$$

We have shown that T_i, T_j are also independent. We can therefore claim the following result.

Theorem 2.1. *In a CTMC, T_i's are independent and exponentially distributed, $\forall i \in \mathcal{S}$.*

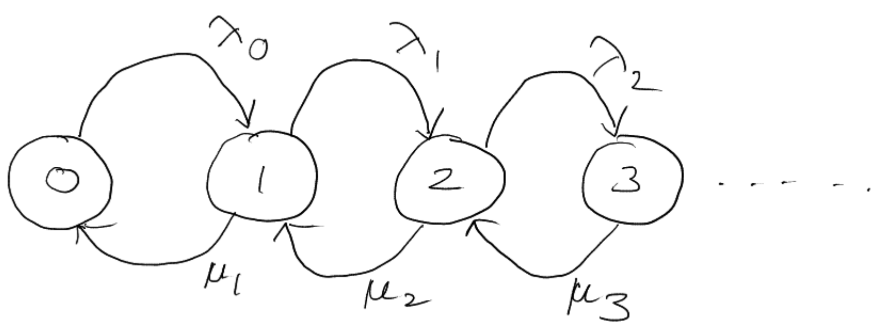
Remark 2. *Note that even though T_i's are exponentially distributed, the rate of the exponential distribution could vary across i. We will see such examples next.*

3 Birth and Death process

We will now look at a class of processes known as the *Birth and Death processes*. These are continuous-time generalisations of the Birth and Death chains we looked at earlier during our study of Markov chains. The state space of the Birth and Death process is $\mathcal{S} = \{0, 1, 2, \ldots\}$.

This means that if the CTMC is at state n, it may either move to state $n-1$ or state $n+1$ with some probability for each, as in the case of Birth and Death chains.

Suppose the CTMC is at state n, then it could move to state $n+1$ after a time $B_n \sim \text{Exp}(\lambda_n)$, i.e, another unit will be *born* after a random time B_n which is exponentially distributed with rate λ_n. Alternatively, it could move to state $n-1$ after a time $D_n \sim \text{Exp}(\mu_n)$, i.e., a unit will *die* after a random time D_n which is exponentially distributed with rate μ_n. We shall assume that the time for a birth and death at state n are independent, i.e., B_n and D_n are independent.



Let's answer a few questions about this system,

- What is the distribution of the time spent by the chain at state n?

  Let us denote this as T_n (time spent in a state, that we introduced above). There could be a birth or a death, therefore the time spent in state n is the time until either a birth or death happens. This is nothing but,

$$T_n = \min(B_n, D_n)$$

  since B_n, D_n are independent exponential random variables, it follows that $T_n \sim \text{Exp}(\lambda_n + \mu_n)$ (see the section on Exponential distribution, in the Poisson process notes).

- What is the probability that the chain moves to state $n+1$?

  Let us denote the probability of moving to state $n+1$ from state n as $p_{n,n+1}$. If the chain moves to state $n+1$ from n, this implies that a birth happened before death. Therefore,

$$p_{n,n+1} = \mathbb{P}(B_n < D_n) = \frac{\lambda_n}{\lambda_n + \mu_n}$$

  By similar logic,

$$p_{n,n-1} = \mathbb{P}(D_n < B_n) = \frac{\mu_n}{\lambda_n + \mu_n}$$

Suppose you record observations from a Birth and Death process such as the one we described above. Each time a transition happens we record the state, say one such realization could be

$0, 1, 2, 1, 2, 3, 2$, where the chain started at 0, followed by a birth to 1, then 2, then a death to 1, followed by births to 2 and 3, and again death to 2. The chain naturally would have spent some time at each of these states but we choose to ignore these times for the moment. Then the transitions from each state to another is called the **embedded Markov chain** with transition matrix for this embedded Markov chain given by $p_{i,j}$ as defined above.

Remark 3. *The idea of the embedded Markov chain can be applied to CTMCs in general. What is implied is that each CTMC can be thought of as a Markov chain (the embedded Markov chain) along with some random time for transition between the states.*

So where do these Birth and Death processes arise?

Example 3.1 (M/M/1 Queue)**.** *Consider a shop with a clerk and customers line up at the shop to get their requirements serviced. Customers arrive at a rate of λ_n and customers are serviced by the clerk at a rate of μ_n. Here n denotes the number of customers in the system, i.e., the customer getting serviced and those waiting in the queue to get serviced.*

Note here that the arrival rate, λ_n, is state dependent (depends on n). This in application means that if a customer were to see a queue with 100 people, they would be less likely to join it, than if there were 5 people in the queue (given that customers generally prefer to wait less). A similar interpretation can be made for the service rate being state dependent.

This is the simplest kind of queue, referred to in the queuing notation (also referred to as the Kendall notation) as M/M/1 queue. The first "M" denotes that the arrival rate has the Memoryless or Markovian property. The second "M" denotes that the service rate has the Memoryless or Markovian property. The third "1" denotes the number of servers in the system. Several other choices and parameters can be used to describe more general queues, but we shall not cover them in this course.

Example 3.2 (M/M/s Queue)**.** *A simple extension of the M/M/1 queue is one with s servers instead of just 1. Customers arrive at rate λ_n. Suppose each of the s servers can service with rate μ. Then convince yourself that,*

$$\mu_n = \begin{cases} n\mu & 1 \le n \le s \\ s\mu & n > s \end{cases}$$

Calculate $p_{n,n+1}$ and $p_{n,n-1}$.

Example 3.3 (Pure Birth process)**.** *Suppose $\mu_n = 0, \forall n$. Then this is called a pure birth process.*

The Birth and Death processes are a general framework for modelling in continuous-time. They can be applied to model queues as considered above, stock prices, limit order books, epidemics and many more phenomena.

Let T_i^{i+1} denote the time taken by the Birth and Death process to go from state i to $i+1$. We want to understand $\mathbb{E}[T_i^{i+1}]$. The idea behind the subsequent calculations is the following. There are two possible cases,

- There is a direct transition from state i to $i+1$.
- There is a transition first from i to $i-1$, in which case the CTMC would first have to return to i and then go to $i+1$.

Define a random variable I_i,

$$I_i = \begin{cases} 1 & i \to i+1 \\ 0 & i \to i-1 \end{cases}$$

I_i takes the value 1, if the chain first moves from i to $i+1$, and 0 if it moves to $i-1$. Note also that $\mathbb{P}(I_i = 1) = p_{i,i+1}$ and $\mathbb{P}(I_i = 0) = p_{i,i-1}$. Therefore,

- $\mathbb{E}[T_i^{i+1}|I_i = 1] = \mathbb{E}[T_i] = \frac{1}{\lambda_i+\mu_i}$
- $\mathbb{E}[T_i^{i+1}|I_i = 0] = \mathbb{E}[T_i] + \mathbb{E}[T_{i-1}^i] + \mathbb{E}[T_i^{i+1}] = \frac{1}{\lambda_i+\mu_i} + \mathbb{E}[T_{i-1}^i] + \mathbb{E}[T_i^{i+1}]$

And,

$$\mathbb{E}[T_i^{i+1}] = \mathbb{P}(I_i = 1)\mathbb{E}[T_i^{i+1}|I_i = 1] + \mathbb{P}(I_i = 0)\mathbb{E}[T_i^{i+1}|I_i = 0]$$

Also, note that $\mathbb{E}[T_0^1] = \frac{1}{\lambda_0}$.
Therefore, you can show that (Verify!) for $\lambda_n = \lambda$ and $\mu_n = \mu \; \forall n$,

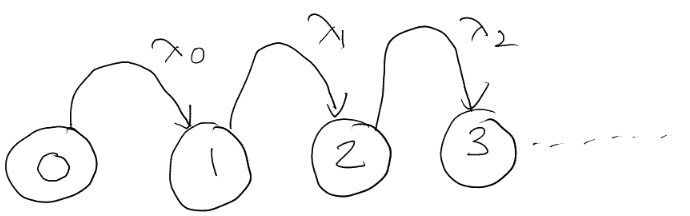
$$\mathbb{E}[T_i^{i+1}] = \begin{cases} \frac{1-(\mu/\lambda)^{i+1}}{\lambda-\mu} & \lambda \neq \mu \\ \frac{(i+1)}{\lambda} & \lambda = \mu \end{cases}$$

Transition Probability Function

The **transition probability function** of a CTMC is defined as,

$$P_{i,j}(t) = \mathbb{P}(X(t+s) = j|X(s) = i)$$

Note that $\sum_j P_{ij}(t) = 1$.
The transition probability function can be thought of as the analogous quantity for a CTMC to the $n-$step transition probability function, $\mathbf{P}^{(n)}(i,j)$ that we came across while studying Markov chains.



Consider a pure birth process defined as above. If $\lambda_1 = \lambda_2 = \ldots = \lambda$, then the pure birth process is a Poisson process (Verify!).
The transition probability function if $X(t)$ is a Poisson process can be calculated,

$$\begin{aligned} \mathbb{P}(X(t+s) = j|X(s) = i) &= \mathbb{P}(X(t+s) - X(s) = j - i|X(s) = i) \\ &= \mathbb{P}(X(t+s) - X(s) = j - i) \\ &\text{(From the independent increments property)} \\ &= \frac{e^{-\lambda t}(\lambda t)^{(j-i)}}{(j-i)!} \end{aligned}$$

Now let us consider the general case of a pure birth process, where $\lambda_i' s$ need not all be identical. It is easy to see that for a pure birth process,

$$T_i \sim \text{Exp}(\lambda_i)$$

We now want to understand the transition probability function of a pure birth process.

$$\begin{aligned}
&P_{i,j}(t)\\
=&\mathbb{P}(X(t+s)=j|X(s)=i)\\
=&\mathbb{P}(X(t+s)<j+1|X(s)=i)-\mathbb{P}(X(t+s)<j|X(s)=i)\\
=&\mathbb{P}(T_i+T_{i+1}+\ldots+T_j>t)-\mathbb{P}(T_i+T_ii+1+\ldots+T_{j-1}>t)
\end{aligned}$$

We know T_i's are independent, but are not identically distributed since λ_i's could be different. The sum of exponential distribution with non-identical parameters follows what is called as the Hypo-exponential distribution. We shall consider a more general approach towards studying transition probability function and return to this question for the pure birth process.

Remark 4. *Note that $p_{i,j}$ that we discussed earlier is the probability of the CTMC going to j from i before any other state. However $P_{ij}(t)$ is the probability of going from i to j in some time t. These both are different probabilities, don't confuse them.*

Chapman-Kolmogorov Equation

Just like we had the Chapman-Kolmogorov equation for the Markov chain, we have an analogous result for CTMCs.

Proposition 3.1. $P_{i,j}(t+s)=\sum_{k=0}^{\infty}P_{i,k}(t)P_{k,j}(s)$

Proof. Left as an exercise. The proof follows similarly to that of Chapman-Kolmogorov equation for the Markov chain. □

Recall that if we wanted to calculate the probability of going from state i to j in n steps, i.e., $\mathbf{P}^{(n)}(i,j)$, we would multiply the transition matrix $\mathbf{P}(i,j)$ to itself n times and consider its $(i,j)-th$ entry. This result was a consequence of the Chapman-Kolmogorov equation for Markov chain.

For CTMCs, unfortunately there is no such simple way for evaluating $P_{i,j}(t)$. However, like the Markov chain we will use the Chapman-Kolmogorov equation to study the transition probability function for CTMCs.

Instantaneous transition rates

We define the **instantaneous transition rate** of a CTMC as,

$$q_{i,j}=v_ip_{i,j}$$

Here v_i is the rate parameter of the distribution of T_i, i.e.,

$$T_i\sim \text{Exp}(v_i)$$

It will be shortly clear why $q_{i,j}$ is referred as the *instantaneous* transition rate.

Example 3.4. *1. Poisson process*

$T_i\sim Exp(\lambda)$, therefore, $v_i=\lambda_i$ and $p_{i,i+1}=1$ (Why?). Therefore,

$$q_{i,i+1}=\lambda$$

Find $q_{i,i-1}$

2. *Pure birth process*

   Find $q_{i,i+1}$ *and* $q_{i,j}$*, where* $j > i+1$.

3. *Birth and Death process*

   $T_i \sim Exp(\lambda_i + \mu_i)$*, therefore,* $v_i = \lambda_i + \mu_i$ *and* $p_{i,i+1} = \frac{\lambda_i}{\lambda_i+\mu_i}$. *Therefore,*

$$q_{i,i+1} = \lambda_i$$

   Find $q_{i,i-1}$.

Lemma 3.1.

$$\lim_{h\to 0} \frac{1 - P_{i,i}(h)}{h} = v_i$$

Proof. Assume that h is small, such that after time h, the CTMC which starts at state i continues to be in state i.

Then,

$$1 - P_{i,i}(h) = \mathbb{P}(T_i < h) = \int_0^h v_i e^{-v_i t} dt = 1 - e^{-v_i h}$$

It follows that $1 - P_{i,i}(h) = v_i h + o(h)$ (Verify!).

Therefore,

$$\lim_{h\to 0} \frac{1 - P_{i,i}(h)}{h} = \lim_{h\to 0} \frac{v_i h + o(h)}{h} = v_i$$

□

and the following Lemma is why we refer to $q_{i,j}$ as the instantaneous transition rate.

Lemma 3.2.

$$\lim_{h\to 0} \frac{P_{i,j}(h)}{h} = q_{i,j}$$

Proof. Again assume that h is small, such that you reach j from i in one jump in some time h and stay at j.

$$P_{i,j}(h) = \mathbb{P}(X(h) = j | X(0) = i) = p_{i,j}\mathbb{P}(T_i < h)$$

This therefore becomes the probability of making the jump ($p_{i,j}$) multiplied by the probability that the jump happens in some time less than h ($\mathbb{P}(T_i < h)$). Hence,

$$\begin{aligned} P_{i,j}(h) &= p_{i,j}\mathbb{P}(T_i < h) \\ &= p_{i,j}(v_i h + o(h)) \\ \text{(See the previous proof)} \qquad &= q_{i,j} h + o(h) \end{aligned}$$

Therefore,

$$\lim_{h\to 0} \frac{P_{i,j}(h)}{h} = q_{i,j}$$

□

Kolmogorov Backward Equation

We will now establish an important tool for computing the transition probability function.

Theorem 3.3. *The Kolmogorov Backward equation for a CTMC is given by,*

$$P'_{i,j}(t) = \sum_{k \neq i} q_{i,k} P_{k,j}(t) - v_i P_{i,j}(t)$$

Proof.

$$\begin{aligned} P_{i,j}(t+h) - P_{i,j}(t) &= \sum_{k=0}^{\infty} P_{i,k}(h) P_{k,j}(t) - P_{i,j}(t) \\ &\quad \text{(Applying the Chapman-Kolmogorov equation to the first term)} \\ &= P_{i,i}(h) P_{i,j}(t) + \sum_{k \neq i} P_{i,k}(h) P_{k,j}(t) - P_{i,j}(t) \\ &= \sum_{k \neq i} P_{i,k}(h) P_{k,j}(t) - P_{i,j}(t)(1 - P_{i,i}(h)) \end{aligned}$$

Now divide by h on both sides and take the limit $h \to 0$ to obtain the result of the theorem. □

References

1. Sheldon Ross, Introduction to Probability Models, Academic Press, 2024.
2. Rick Durrett, Essentials of Stochastic Processes, Springer, 1999.