Lecture 4: Continuous-time Markov Chains

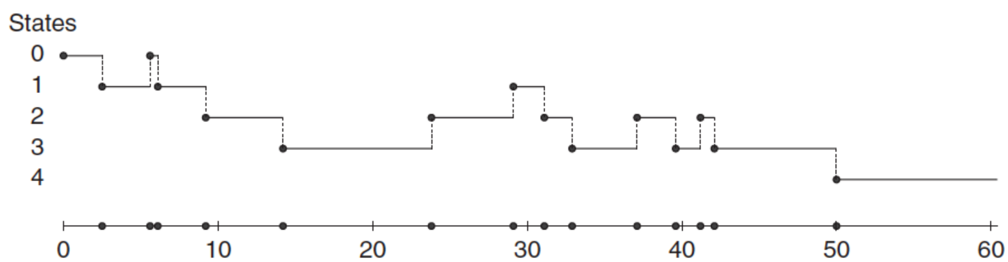
Readings

- Grimmett and Stirzaker (2001) 6.8, 6.9.

Options:

- Grimmett and Stirzaker (2001) 6.10 (a survey of the issues one needs to address to make the discussion below rigorous)
- Norris (1997) Chapter 2,3 (rigorous, though readable; this is the classic text on Markov chains, both discrete and continuous)
- Durrett (2012) Chapter 4 (straightforward introduction with lots of examples)

Many random processes have a discrete state space, but can change their values at any instant of time rather than at fixed intervals: radioactive atoms decaying, the number of molecules in a chemical reaction, populations with birth/death/immigration/emigration, the number of emails in an inbox, etc. The process is piecewise constant, with jumps that occur at continuous times, as in this example showing the number of people in a lineup, as a function of time (from Dobrow (2016)):



The dynamics may still satisfy a continuous version of the Markov property, but they evolve continuously in time. Rather than simply discretize time and apply the tools we learned before, a more elegant model comes from considering a continuous-time Markov chain (ctMC.)

In this class we'll introduce a set of tools to describe continuous-time Markov chains. We'll make the link with discrete-time chains, and highlight an important example called the Poisson process. If time permits, we'll show two applications of Markov chains (discrete or continuous): first, an application to clustering and data science, and then, the connection between MCs, electrical networks, and flows in porous media.

4.1 Definition and Transition probabilities

Definition. Let $X = (X_t)_{t\geq 0}$ be a family of random variables taking values in a finite or countable state space S, which we can take to be a subset of the integers. X is a *continuous-time Markov chain* (ctMC) if it satisfies the *Markov property*:[1]

$$P(X_{t_n} = i_n | X_{t_1} = i_1, \ldots, X_{t_{n-1}} = i_{n-1}) = P(X_{t_n} = i_n | X_{t_{n-1}} = i_{n-1}) \tag{1}$$

for all $i_1, \ldots, i_n \in S$ and any sequence $0 \le t_1 < t_2 < \cdots < t_n$ of times. The process is *time-homogeneous* if the conditional probability does not depend on the current time, so that:

$$P(X_{t+s} = j | X_s = i) = P(X_t = j | X_0 = i), \qquad s \ge 0. \tag{2}$$

We will consider only time-homogeneous processes in this lecture. Additionally, we will consider only processes which are *right-continuous*.

The Markov property (1) says that the distribution of the chain at some time in the future, only depends on the current state of the chain, and not its history. The difference from the previous version of the Markov property that we learned in Lecture 2, is that now the set of times t is continuous – the chain can jump between states at any time, not just at integer times.

There is no exact analogue of the transition matrix P, since there is no natural unit of time. Therefore we consier the transition probabilities as a function of time.

Definition. The *transition probability* for a time-homogeneous chain is

$$P_{ij}(t) = P(X_{t+s} = j | X_s = i), \qquad s,t \ge 0. \tag{3}$$

Write $P(t) = (P_{ij}(t))$ for the matrix of transition probabilities at time t.

Clearly, $P(t)$ is a stochastic matrix.

Remark. For a time-inhomogeneous chain, the transition probability would be a function of two times, as $P_{ij}(s,t) = P(X(t+s) = j | X_s = i)$.

In a similar way to the discrete case, we can show the Chapman-Kolmogorov equations hold for $P(t)$:

Chapman-Kolmogorov Equation. *(time-homogeneous)*

$$P(t+s) = P(t)P(s) \quad \Longleftrightarrow \quad P_{ij}(t+s) = \sum_{k \in S} P_{ik}(t) P_{kj}(s)\ . \tag{4}$$

[1] The Markov property in continuous time can be formulated more rigorously in terms of σ-algebras. Let $(\Omega, \mathscr{F}, P)$ a the probability space and let $\{\mathscr{F}_t\}_{t \ge 0}$ be a filtration: an increasing sequence of σ-algebras such that $\mathscr{F}_t \subseteq \mathscr{F}$ for each t, and $t_1 \le t_2 \Rightarrow \mathscr{F}_{t_1} \subseteq \mathscr{F}_{t_2}$. We suppose the process X_t is *adapted* to the filtration $\{\mathscr{F}_t\}_{t \ge 0}$: each X_t is measurable with respect to $\mathscr{F}_t$. For example, this will be true automatically if we let $\mathscr{F}_t$ be the σ-algebra generated by $(X_s)_{0 \le s \le t}$, i.e. generated by the pre-images $X_s^{-1}(B)$ for Borel sets $B \subset \mathbb{R}$. Then X_t has the Markov property if

$$\mathbb{E}(f(X_t) | \mathscr{F}_s) = \mathbb{E}(f(X_t) | \sigma(X_s))$$

for all $0 \le s \le t$ and bounded, measurable functions f. Another way to say this is $P(X_t \in A | \mathscr{F}_s) = P(X_t \in A | \sigma(X_s))$, where $P(\cdot|\cdot)$ is a *regular conditional probability* (see Koralov and Sinai (2010), p.184.)

Proof.

$$
\begin{aligned}
P_{ij}(s+t) &= P(X_{s+t}=j|X_0=i) \\
&= \sum_k P(X_{s+t}=j|X_t=k,X_0=i)P(X_t=k|X_0=i) && \text{(LoTP)} \\
&= \sum_k P(X_{s+t}=j|X_t=k)P(X_t=k|X_0=i) && \text{(Markov property)} \\
&= \sum_k P_{ik}(t)P_{kj}(s) && \text{(time-homogeneity.)}
\end{aligned}
$$

□

Remark. A set of operators $(T(t))_{t\geq 0}$ such that $T(t+s)=T(s)T(t)$ is a *semigroup*. The transition probabilities $P(t)$ form a semigroup (where $P(t)$ is an operator in the sense that it maps vectors to vectors.) Studying abstract semigroups of operators, sometimes with additional properties like continuity at $t=0$, or uniform continuity, is another approach to studying Markov processes, that is often used in the probability literature. See e.g. Grimmett and Stirzaker (2001), p.256, for more details related to continuous-time MCs, and see Koralov and Sinai (2010); Pavliotis (2014) for a discussion of general Markov processes.

The transition probability can be used to completely characterize the evolution of probability for a continuous-time Markov chain, but it gives too much information. We don't need to know $P(t)$ for all times t in order to characterize the dynamics of the chain. We will consider two different ways of completely characterizing the dynamics:

(i) By the times at which the chain jumps, and the states that it jumps to.

(ii) Through the *generator* Q, which is like an infinitesimal version of P.

We will consider both approaches. Let's start with the second one, since it gives the most fundamental characterization of a Markov process.

4.2 Infinitesimal generator, and forward and backward equations

For the rest of the lecture, we will assume that $|S|<\infty$. Most of the results we derive will also be true when $|S|=\infty$ under certain additional, not-very-restrictive assumptions. Because worrying about these details is technical and beyond the scope of the course, we leave them out and focus on the easier case for now. See Norris (1997) for a rigorous discussion that includes the case $|S|=\infty$.

4.2.1 Infinitesimal generator

The fundamental way to characterize a ctMC is by its generator, which is like its infinitesimal transition rates. Let's go back to the transition probability $P(t)$. By definition, we have that $P(0)=I$, the identity matrix. Let's *assume* that $P(t)$ is right-differentiable at $t=0$. This does not have to be the case for any ctMC, but it will be true in applications and it makes the subsequent theory much easier to derive.[2] This

[2] Actually, the weaker assumption $P(t)\to I$ uniformly as $t\searrow 0$ is sufficient to derive the results; see Grimmett and Stirzaker (2001), section 6.10 or Norris (1997), Chapter 2.

assumption also implies that $P(t)$ is differentiable for all $t > 0$. You can show this (if you are interested) from the Chapman-Kolmogorov equations.

Definition. Let $X = (X_t)_{t\geq 0}$ be a ctMC with transition probabilities $P(t)$. The *generator* or *infinitesimal generator* of the Markov Chain is the matrix

$$Q = \lim_{h\to 0^+} \frac{P(h)-I}{h}. \tag{5}$$

Write its entries as $Q_{ij} = q_{ij}$.

Some properties of the generator that follow immediately from its definition are:

(i) Its rows sum to 0: $\sum_j q_{ij} = 0$.
(ii) $q_{ij} \geq 0$ for $i \neq j$.
(iii) $q_{ii} < 0$

Proof.

(i) $\sum_j P_{ij}(h) = 1$, since $P(h)$ is a transition matrix, and $\sum_j I_{ij} = 1$. Pass to the limit.
(ii) $P_{ij}(h) \geq I_{ij}$, so this is true in the limit.
(iii) Follows from (i) and (ii), since $q_{ii} = -\sum_j q_{ij}$.

□

Remark. If $|S| = \infty$, then most of what follows will be true under the assumption that $\sup_i |q_{ii}| < \infty$. This condition ensures the chain cannot blow up to ∞ in finite time. Therefore, we will consider examples with $|S| = \infty$, even though we won't prove here that the theory works in this case. See Norris (1997), Chapters 2.

How should the entries of the generator be interpreted? Each entry q_{ij} is like a *rate* of jumping from i to j. Recall that a rate measures the average number of events (in the colloquial sense), per unit time. In this case, an event is a transition from i to j. The diagonal entries $-q_{ii}$ are the overall rates of leaving each state i.

Imagine an experiment where we start in state i, and every time we jump out of it we return to i immediately. Then, q_{ij} would be approximated by the total number of jumps to state j, divided by the total time over which we perform the experiment. Indeed, the instantaneous transition rate of hitting $j \neq i$ is

$$\lim_{h\to 0+} \frac{\mathbb{E}[\text{number of transitions to } j \text{ in (t,t+h]} \,|\, X_t{=}i]}{h} = \lim_{h\to 0+} \frac{P(X_{t+h}{=}j|X_t{=}i)}{h} = q_{ij}.$$

(Since, in a small interval of time, the number of transitions is either 0 or 1.) For a similar reason the overall rate of leaving is $-q_{ii} = -\sum_j q_{ij}$, because the overall rate of leaving is the rate at which *any* transition happens, so it is the sum of the individual rates.

Another justification of the word "rate" comes from considering the transition probabilities after a small amount of time h has elapsed:

$$\begin{aligned} P_{ij}(h) &= q_{ij}h + o(h) && (j \neq i) \\ P_{ii}(h) &= 1 + q_{ii}h + o(h) && (j = i) \end{aligned} \tag{6}$$

Then q_{ij} gives the rate at which probability accumulates in state j (given we are in state i), at least instantaneously, and $-q_{ii}$ is the rate at which probability leaves i.

Here are some examples of ctMCs and their corresponding generators. In virtually all applications, you are given the *generator* (or you construct it as a model), and not the set of transition probabilities. The generator therefore is the fundamental object of interest; (5) is only used to compute the transition probabilities from the generator, as we will see in the next section.

Example (Two-state chain)**.** A generic two-state ctMC has generator

$$Q = \begin{pmatrix} -\alpha & \alpha \\ \beta & -\beta \end{pmatrix}$$

where $\alpha, \beta \geq 0$ are parameters. For example, suppose you wish to model a ligand that can bind and unbind to a protein. You might be told that the rate of binding is α binding events per second, and the rate of unbinding is β binding events per second. Your model for the ligand is a continuous-time Markov chain X_t (t is measured in seconds) on the state space $S = \{\text{unbound,bound}\}$, with generator given above.

Example (Birth-death process)**.** Suppose you are modeling the population of black bears in the Catskills. You assume that bears are born with rate λ per year, and they die with rate μ per year. (If there are no more bears, you still assume they are "born" with rate λ; for example this could model immigration from another location.) If X_t is the number of bears in year t, and $X = (X_t)_{t\geq 0}$ is a Markov process, then the generator for this process is

$$Q = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ \vdots \end{array} \begin{array}{c} \begin{array}{ccccc} 0 & 1 & 2 & 3 & \end{array} \\ \begin{pmatrix} -\lambda & \lambda & 0 & 0 & \cdots \\ \mu & -(\lambda+\mu) & \lambda & 0 & \cdot \\ 0 & \mu & -(\lambda+\mu) & \lambda & \cdots \\ 0 & 0 & \mu & -(\lambda+\mu) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \end{array}$$

The generator above is an example of a *birth-death process*, a ctMC on the set of non-negative integers where transitions can only occur to neighbouring states. For a general birth-death process, the birth and death rates can be state-dependent, so that transitions from $i \to i+1$ occur with rate λ_i, and transitions from $i \to i-1$ occur with rate μ_i.

Such a process models a wide range of situations, such as people arriving and leaving a lineup, search requests arriving at a Google server, emails arriving and being deleted from an inbox, the number of infected individuals in a population, etc.

Example (Poisson process)**.** A *Poisson process with rate* λ is a continuous-time Markov chain $N = (N_t)_{t\geq 0}$ on $S = \{0, 1, 2, \ldots\}$ with generator

$$Q = \begin{pmatrix} -\lambda & \lambda & 0 & 0 & \cdots \\ 0 & -\lambda & \lambda & 0 & \cdots \\ 0 & 0 & -\lambda & \lambda & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

and initial condition $N_0 = 0$. The Poisson process is a birth-death process with death rate $\mu = 0$, but it is treated separately because it arises so frequently in applications.

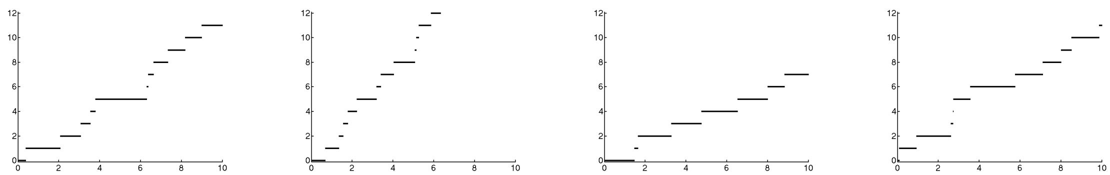


Figure 1: Some realizations of the Poisson process with rate $\lambda = 1$. These were all simulated for 10 time units.

The Poisson process is often used to count the number of events that have happened at time t, if the events occur independently with rate λ. For example, it could count the number of busses that pass a bus stop in a certain time t, if the drivers have absolutely no idea where the other drivers are, and are so delayed by bad traffic that their arrival times are completely random. It could count the number of radioactive atoms that have decayed, or the number of telephone calls that arrive at a call center, or the number of search requests received by Google's servers. It is used to model the growth of a population, say of people, bacteria, or rabbits (if we don't count deaths.[3]) It could be used to model the number of large waves that arrive at an oil platform, or the locations of breaking waves in the ocean.[4]

Remark. Another definition of the Poisson process, which is sometimes more useful in practical calculations and doesn't require the theory of continuous-time Markov chains, is that it is a process $N = (N_t)_{t\geq 0}$ taking values in $S = \{0, 1, 2, \ldots\}$ such that:

(i) $N(0) = 0$;

(ii) N_t has stationary and sindependent increments: for any $0 \leq t_1 < t_2 < \cdots < t_n$, the random variables

$$N_{t_2} - N_{t_1},\ N_{t_3} - N_{t_2},\ \ldots, N_{t_n} - N_{t_{n-1}}$$

are independent ("independent increments"), and for any $t \geq 0$, $s \geq 0$, the distribution of $N_{t+s} - N_t$ is independent of t ("stationary increments");

(iii) $N_t \sim \text{Poisson}(\lambda t)$.

4.2.2 Forward and backward equations

We defined the generator from the full time-dependent transition probabilities. The next question we want to know is: can we recover $P(t)$ from Q? The answer is yes. Let's show this, by deriving an evolution equation for $P(t)$, and then solving it.

We calculate:

$$P'(t) = \lim_{h\to 0^+} \frac{P(t+h) - P(t)}{h} = \lim_{h\to 0^+} \frac{P(t)P(h) - P(t)I}{h} = P(t)\left(\lim_{h\to 0^+} \frac{P(h) - I}{h}\right) = P(t)Q.$$

[3] We could include deaths by modelling them as a Poisson process too, and this would be a birth-death process.

[4] We have only defined the Poisson process on a line, but it can be generalized to a function of multiple spatial coordinates.

We factored $P(t+h) = P(t)P(h)$, using the Chapman-Kolmogorov equations. This gives an evolution equation for $P(t)$.

We may also consider the evolution of a probability distribution $\mu(t)$. As in the discrete-time case, $\mu(t)$ is a row vector, whose components are $\mu_i(t) = P(X_t = i | X_0 \sim \mu(0))$, where $\mu(0)$ is the initial probability distribution. Since $\mu(t) = \mu(0)P(t)$ and therefore $\mu'(t) = \mu(0)P'(t)$, we can multiply the above equation on the left by $\mu(0)$ to obtain an evolution equation for $\mu(t)$. We obtain two versions of the forward equation:

Forward Kolmogorov Equation. *Given a time-homogeneous ctMC with generator Q, the transition probabilities $P(t)$ evolve as*

$$P'(t) = P(t)Q\,, \qquad P(0) = I\,, \tag{7}$$

and the probability distribution $\mu(t)$ evolves as

$$\frac{d\mu}{dt} = \mu(t)Q\,, \qquad \mu(0) = \textit{given initial distribution}\,. \tag{8}$$

Going back to our initial calculation, we can factor out $P(t)$ on the right instead, to get

$$P'(t) = \lim_{h\to 0^+} \frac{P(t+h) - P(t)}{h} = \lim_{h\to 0^+} \frac{P(h)P(t) - IP(t)}{h} = \left(\lim_{h\to 0^+} \frac{P(h) - I}{h}\right) P(t) = QP(t).$$

From this we may consider expectations of functions of the Markov chain. Let $u(t)$ be a column vector with components $u_k(t) = \mathbb{E}_k f(X_t) = \mathbb{E}(f(X_t)|X_0 = k)$. Write $f = (f(1), f(2), \ldots, f(|S|))^T$, so that $u(0) = f$. Then $u(t) = P(t)f$, and $u'(t) = P'(t)f$, so multiplying by f on the right gives an evolution equation for $u(t)$. We obtain two versions of the backward equation:

Backward Kolmogorov Equation. *Given a time-homogeneous ctMC with generator Q, the transition probabilities $P(t)$ evolve as*

$$P'(t) = QP(t)\,, \qquad P(0) = I\,, \tag{9}$$

and the expectation $u(t)$ evolves as

$$\frac{du}{dt} = Qu(t)\,, \qquad u(0) = f\,. \tag{10}$$

From the forward and backward equations we may make several observations. First, putting (7), (9) together, shows that $QP(t) = P(t)Q$, i.e. the transition probability matrix commutes with the generator. We can also solve explicitly for $P(t)$, to obtain

$$P(t) = e^{Qt}P(0) = e^{Qt}\,.$$

Recall that $e^{Qt} \equiv \sum_{n=0}^{\infty} \frac{1}{n!} Q^n t^n$ for any square matrix Q.[5] Therefore, if we know the infinitesimal generator Q, then we can completely determine the transition probabilities $P(t)$ for $t > 0$, so Q completely characterizes the continuous-time Markov chain.

Example (Poisson process)**.** Let's try to calculate the probability distribution of the Poisson process at each point in time. Let $\mu_j(t) = P(N_t = j)$, $j = 0, 1, 2, \ldots$. We solve for $\mu_j(t)$ using the forward Kolmogorov equations (8). When $j = 0$ we have

$$\frac{d\mu_0}{dt} = -\lambda\mu_0, \qquad \mu_0(0) = 1.$$

[5]That this power series converges for a finite generator Q is shown in Norris (1997), section 2.10.

The solution is $\mu_0(t) = e^{-\lambda t}$. The next equation is

$$\frac{d\mu_1}{dt} = \lambda\mu_0 - \lambda\mu_1, \qquad \mu_1(0) = 0.$$

Substituting for $\mu_0(t)$ and solving gives $\mu_1(t) = \lambda t e^{-\lambda t}$. In general, we have

$$\frac{d\mu_j}{dt} = -\lambda\mu_j + \lambda\mu_{j-1}, \qquad \mu_j(0) = 0 \quad (j > 0).$$

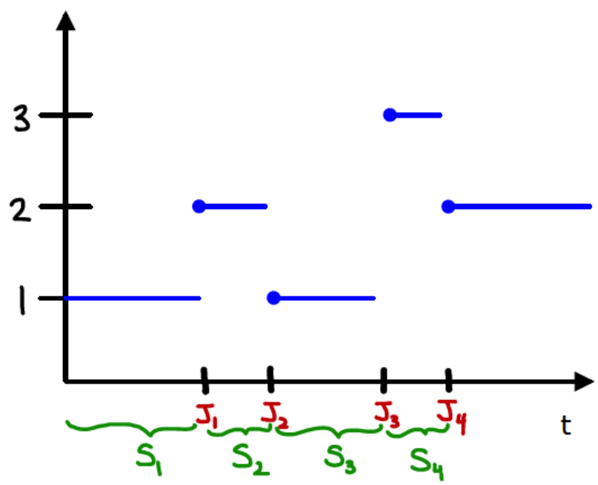
We can solve these by induction to find that

$$\mu_j(t) = \frac{\lambda^j t^j}{j!} e^{-\lambda t}.$$

This shows that at fixed time t, $N(t)$ is a Poisson random variable with parameter λt.

4.3 Transition times and jumps

A continuous-time Markov chain stays in one state for a certain amount of time, then jumps immediately to another state where it stays for another amount of time, etc. A natural way to characterize this process by the the times at which it jumps, and the distributions of the states it jumps to. Such a characterization will also lead to a method to simulate realizations of the process exactly.



Definition. Let us define the following:

- The *jump time* J_m is the time of mth jump. It is defined recursively by

$$J_{m+1} = \inf\{t > J_m : X_t \neq X_{J_m}\}, \qquad J_0 = 0.$$

- The *holding time* S_m is the length of time a ctMC stays in its state, before jumping for the mth time. It is calculated from the jump times as $S_m = J_m - J_{m-1}$.

- The discrete-time process $(Y_n)_{n=0}^{\infty}$ given by $Y_n = X_{J_n}$ is called the *jump process*, *jump chain*, or *embedded chain*.

We will determine the probability distributions of $(J_m)_{m\geq 0}$ and $(Y_m)_{m\geq 0}$. First, we need a couple of technical results.

For a continuous-time process X, a random variable T with values in $[0,\infty]$ is a stopping time if, for each $t \in [0,\infty)$, the event $\{T \leq t\}$ depends only on $(X_s : s \leq t)$.

Theorem. *The jump time J_m is a stopping time of $(X_t)_{t\geq 0}$ for all m.*

For a proof, see Norris (1997), Lemma 6.5.2 p. 226.

Theorem (Strong Markov property)**.** *Let $X = (X_t)_{t\geq 0}$ be a ctMC with generator Q and let T be a stopping time. Assume that $P(T < \infty) = 1$. Then $(X_{T+t})_{t\geq 0}$ is a continuous-time Markov chain with generator Q, and initial condition equal to the distribution of X_T.*

Remark. Proving this theorem (and even properly making sense of a stopping time in the continuous-time case) requires measure theory, so we won't do that here. See Norris (1997) section 6.5, Theorem 6.5.4, p.227.

The Strong Markov Property says that if we stop X at some stopping time, and then consider the subsequent process, then it has exactly the same transition probabilities as the original. In other words,

$$P(X_{T+t} = j | X_T = i, X_s = x_s, \ 0 \leq s \leq T) = P(X_t = j | X_0 = i)\,. \tag{11}$$

Now let's use these results to compute the jump times and jump probabilities. We start with the jump times. Suppose we are in state i after the $m-1$th jump, and we wish to know the distribution of S_m, the time we wait until we jump again.

Theorem. *Suppose $X_{J_{m-1}} = i$. The holding time S_m is an exponentially distributed random variable with parameter $-q_{ii}$.*

Remark. Recall that a random variable Y is exponentially distributed with parameter $\lambda > 0$, if it has probability density function $p(y)dy = \lambda^{-1}e^{-\lambda y}$. Exponential random variables have the "lack-of-memory" property: $P(Y > y + x | Y > y) = P(Y > x)$, and in fact, one can show they are the only continuous random variables with this property (Grimmett and Stirzaker, 2001, p. 259, 140).

Proof. We will show that S_m is memoryless. First, notice that

$$\{S_m > r\} = \{X_{J_{m-1}+u} = i \ \text{ for all } 0 \leq u \leq r\}\,.$$

Therefore

$$\begin{aligned}
P(S_m > r + t \,|\, S_m > r, X_{J_{m-1}}{=}i) &= P(S_m > r + t \,|\, X_{J_{m-1}+u}{=}i, \ 0 \leq u \leq r) && \text{conditioning on same event} \\
&= P(S_m > t + r \,|\, X_{J_{m-1}+r}{=}i) && \text{Strong Markov property} \\
&= P(S_m > t \,|\, X_{J_{m-1}}{=}i) && \text{time-homogeneity}
\end{aligned}$$

Therefore, S_m is memoryless (given we start in state i), and since it is continuous it must have an exponential distribution: $P(S_m > t \,|\, X_{J_{m-1}}{=}i) = e^{-\lambda_i t}$ for some parameter λ_i.

What is λ_i? Let's calculate:

$$\lambda_i = -\frac{d}{dt}\Big|_{t=0} P(S_m > t \,|\, X_{J_{m-1}}{=}i)$$
$$= \lim_{h\to 0} \frac{1 - P(S_m > h|X_{J_{m-1}}{=}i)}{h} = \lim_{h\to 0} \frac{1 - P(X_h{=}i|X_0{=}i) + o(h)}{h} = \lim_{h\to 0} \frac{-q_{ii}h + o(h)}{h} = -q_{ii}.$$

Therefore $\lambda_i = -q_{ii}$, the negative of the diagonal elements of the generator. □

Remark. The same proof can be straightforwardly adapted to show that, if $X_s = i$ for some *deterministic* time s, then the waiting time until the next jump is Exponential$(-q_{ii})$. Just replace J_{m-1} with s in the proof above, and use the (regular) Markov property. In words, because the exponential distribution is memoryless, the waiting time distribution does not depend on the time at which we start counting.

Now we consider the jump process $Y_0, Y_1, \ldots$. By the Strong Markov property and the fact that J_m is a stopping time, this process is a discrete-time Markov chain. It is time-homogeneous because X is. Therefore we can characterize the jump chain by its transition matrix, which we will call $\tilde{P}$. It has elements

$$\tilde{P}_{ij} = P(Y_m = j|Y_{m-1} = i) = P(X_{J_m} = j \,|\, X_{J_{m-1}} = i)\,. \tag{12}$$

Proposition. *The transition matrix of the embedded chain has elements $\tilde{P}_{ij} = -q_{ij}/q_{ii}$.*

Proof. We would like to condition exactly on a jump time. However, conditioning on an event of measure zero can be problematic, so instead we condition on the jump occurring in a small interval $(t, t+h]$, and then let $h \to 0$.

We have

$$\begin{aligned}
\tilde{P}_{ij} &= \lim_{h\to 0+} P(X_{J_{m-1}+t+h}{=}j|X_{J_{m-1}}{=}i, t < S_m \le t+h) \\
&= \lim_{h\to 0+} P(X_{J_{m-1}+t+h}{=}j|X_{J_{m-1}+t}{=}i, t < S_m \le t+h) && \text{Strong Markov property \& defn of } S_m \\
&= \lim_{h\to 0+} P(X_{t+h}{=}j|X_t{=}i, t < U \le t+h) && \text{time homogeneity} \\
&= \lim_{h\to 0+} \frac{P(X_{t+h}{=}j|X_t{=}i) + o(h)}{P(t < U \le t+h|X_t{=}i)} && \text{defn of conditional probability} \\
&= \lim_{h\to 0+} \frac{q_{ij}h + o(h)}{-q_{ii}h + o(h)} && \text{distribution of } U \text{ \& next jump} \\
&= -\frac{q_{ij}}{q_{ii}}\,.
\end{aligned}$$

We defined $U = \inf\{s > t : X_s \ne i\}$ to be the time until the next jump, given the process is currently at state i, and used that $P(X_{t+h}{=}j, t < U \le t+h|X_t{=}i) = P(X_{t+h}{=}j|X_t{=}i) + o(h)$.

In words: Suppose that $X_0 = i$, and that $t < J_1 \le t+h$, and suppose that h is small enough that the chain jumps only once in $(t, t+h]$. Then

$$P(\text{jumps to } j|\text{it first jumps in } (t,t+h]) \approx \frac{p_{ij}(h)}{1 - p_{ii}(h)} \to -\frac{q_{ij}}{q_{ii}} \quad \text{as} \quad h \searrow 0.$$

□

Putting these results together shows that if we are given Q, we can calculate the holding time distributions and $\tilde{P}$, and conversely if we are given the holding time distributions and $\tilde{P}$, we can recover Q.

From the jump chain and holiding times we obtain a method to simulate exact realizations of X.

Kinetic Monte Carlo algorithm (KMC) Also known as the *stochastic simulation algorithm* (SSA), or the *Gillespie algorithm*. Suppose $X_t = i$. Update the process as follows.

- Generate a random variable τ from an exponential distribution with parameter $-q_{ii}$;
- Choose a state j to jump to from the probability distribution given by the ith row of $\tilde{P}$;
- Jump to j at time $t+\tau$, i.e. set $X_s = i$ for $t \leq s < t+\tau$, and $X_{t+\tau} = j$.
- Repeat, starting at state j at time $t+\tau$.

This algorithm is used to simulate a wide range of problems in chemical kinetics and materials science (e.g. surface diffusion, surface growth, defect mobility, etc.) You will see an example on the homework.

4.4 Long-time behaviour

The long-time behaviour of a continuous-time Markov chain is similar to the discrete case.

Definition. A probability distribution μ is a *limiting distribution* of a continuous-time Markov chain if, for all i, j, $\lim_{t\to\infty} P_{ij}(t) = \mu_j$.

Definition. A probability distribution π is a *stationary distribution* of a continuous-time Markov chain if

$$\pi = \pi P(t) \qquad \forall\, t \geq 0. \tag{13}$$

A stationary distribution is, like in the discrete-time case, a probability distribution such that if the Markov chain starts with this distribution, then the distribution never changes. Like for discrete-time, a limiting distribution is a stationary distribution, but the converse is not true in general.

How do we find the stationary distribution? We can find it directly from the generator.

Theorem. *Let $P(t)$ be the transition function for a ctMC. Then a probability distribution π is a stationary distribution iff*

$$\pi Q = 0\,. \tag{14}$$

Proof. Take d/dt of (13) to get $\pi P'(t) = 0$. From (7) this implies $\pi P(t) Q = 0$. But $\pi P(t) = \pi$, so the result follows. $\square$

Equation (14) says that π is a left eigenvector of Q corresponding to eigenvalue 0. Therefore, we can find it by linear algebra.

There is a close link to the formula for a discrete-time Markov chain – recall that if Y_n is a discrete-time Markov chain with transition matrix P, then the stationary distribution solves $(P-I)\pi = 0$. But $(P-I)$ is like a forward-difference approximation to $\frac{dP}{dt}|_t = 0$ using a time step of $\Delta t = 1$. Therefore the discrete and continuous time equations are structurally the same.

Detailed balance is also quite similar to the discrete-time case.

Definition. A ctMC with generator Q and stationary distribution π satisfies *detailed balance with respect to* π, or is *reversible with respect to* π, if it satisfies the *detailed balance equations*

$$\pi_i q_{ij} = \pi_j q_{ji} \,. \tag{15}$$

***Exercise* 4.1.** Find the stationary distribution for the birth-death chain, and give the conditions under which it exists. Is the chain reversible?

Now, when will a stationary distribution be a limiting distribution? Recall the following definition, which is only very slightly modified from the discrete-time case.

Definition. A continuous-time Markov chain is *irreducible* if, for all i, j, there exists $t > 0$ s.t. $P_{ij}(t) > 0$.

Lemma. *If $P_{ij}(t) > 0$ for some $t > 0$, then $P_{ij}(t) > 0$ for all $t > 0$.*

Proof. (From Norris (1997), Theorem 3.2.1 p.111.) If a ctMC is irreducible, then for each i, j there are states $i_0, i_1, \ldots i_n$ with $i_0 = i$, $i_n = j$, and $\tilde{P}_{i_0 i_i}\tilde{P}_{i_1 i_2}\ldots\tilde{P}_{i_{n-1}i_n} > 0$ ($\tilde{P}$ is the transition matrix for the jump chain, defined in the next section), and therefore $q_{i_0 i_i} q_{i_1 i_2}\ldots q_{i_{n-1}i_n} > 0$. If $q_{i'j'} > 0$ then

$$p_{i'j'}(t) \geq P(J_1 \leq t, Y_1 = j', S_2 > t) = (1 - e^{-\lambda_{i'} t})\tilde{p}_{i'j'} e^{-\lambda_{i'} t} > 0$$

for all t. Therefore

$$p_{ij}(t) \geq p_{i_0 i_1}(t/n)\ldots p_{i_{n-1}i_n}(t/n) > 0$$

for all $t > 0$. □

In words: if $P_{ij}(t) > 0$ for some t, then there exists a path in state space from $i \to j$ such that there is a positive probability of performing each step of the path in any positive time, and therefore for any other time $s > 0$ there is a positive probability density of reaching j from i in s time units. Note that there is no notion of periodicity for a continuous-time chain.

Theorem (From Norris (1997), Theorem 3.2.1, p.111)**.** *X is irreducible, iff for each i, j, there exists a sequence of states $i_1, i_1, \ldots, i_n$ with $i_0 = i, i_n = j$, such that $q_{i_0 i_1} q_{i_1 i_2} \cdots q_{i_{n-1} i_n} > 0$.*

In other words, there is a path in the generator leading from i to j; equivalently, the jump chain is irreducible.

Theorem (Ergodic Theorem for continuous-time Markov chains)**.** *Let X be an irreducible continuous-time Markov chain with transition function $P(t)$, such that $P(t) \to I$ as $t \to 0$. Then*

(a) If there exists a stationary distribution π then it is unique, and $P_{ij}(t) \to \pi_j$ as $t \to \infty$, for all i and j;

(b) If there is no stationary distribution then $P_{ij}(t) \to 0$ as $t \to \infty$, for all i and j.

This theorem holds when $|S| = \infty$, with no other conditions. For a sketch of a proof, see Grimmett and Stirzaker (2001, Theorem (21), Section 6.9, p.261). Clearly, the second condition can only hold if $|S| = \infty$. Therefore, for a finite irreducible ctMC, there is always a unique stationary distribution, which is the limiting distribution.

Example. A Poisson process is an example of a ctMC with no stationary distribution. It is not even irreducible, since since the process can only increase.

4.5 Mean first-passage time

Recall that given a set $A \subset S$, the first-passage time to A is the random variable T_A defined by

$$T_A = \inf\{t \geq 0 : X_t \in A\} .$$

The mean first passage time (mfpt) to A, starting at $i \in S$, is $\tau_i = \mathbb{E}_i T_A$. We can solve for the mfpt by solving a system of linear equations.

Mean first-passage time. The mfpt solves the system of equations

$$\begin{cases} \tau_i = 0 & j \in A \\ 1 + \sum_j Q_{ij}\tau_j = 0 & i \notin A \end{cases} \tag{16}$$

Remark. System (16) is sometimes written (heuristically) as

$$Q'\tau = -1, \qquad \tau(A) = 0$$

where Q' is the matrix formed from Q by removing the rows and columns corresponding to states in A. Therefore the mfpt solves the non-homogeneous backward equation with a particular boundary condition.

Remark. Recall that for the discrete-time case, we had $(P-I)\tau = -1$, $\tau(A) = 0$. Again, $(P-I)$ is like a forward-difference approximation to $\frac{dP}{dt}|_t = 0$ using a time step of $\Delta t = 1$.

The proof is slightly less straightforward that in the discrete case, since we need to account for the time spent in each state, not just where we transition to next. It will use the notation of *conditional expectation*. Given two random variables A, B with some joint distribution, the conditional expectation is defined by

$$\mathbb{E}[A|B=b] \quad = \begin{cases} \sum_a aP(A=a|B=b) & \text{if A, B are discrete,} \\ \int_a aP(A=a|B=b)da & \text{if A, B are continuous.} \end{cases}$$

Proof. Clearly $\tau_i = 0$ for $i \in A$. Suppose $X_0 = i \notin A$. Let

$$\begin{aligned} J_1 &= \text{next jump time, given } X_0 = i \\ Y_1 &= \text{next state jump to, given } X_0 = i \end{aligned}$$

Let's calculate the mfpt, by conditioning on the first jump, and then subtracting the time of the first jump. We have

$$\begin{aligned} \tau_i = \mathbb{E}[T_A|X_0{=}i] &= \mathbb{E}[J_1|X_0{=}i] + \mathbb{E}[T_A - J_1|X_0{=}i] \\ &= \mathbb{E}[J_1|X_0{=}i] + \sum_{j\neq i}\mathbb{E}[T_A - J_1, Y_1 = j|X_0{=}i] \\ &= \mathbb{E}[J_1|X_0{=}i] + \sum_{j\neq i}\mathbb{E}[T_A - J_1|Y_1{=}j, X_0{=}i]P(Y_1{=}j|X_0{=}i) \end{aligned}$$

Now, we will use three short calculations:

- $\mathbb{E}[T_A - J_1|Y_1{=}j, X_0{=}i] = \mathbb{E}[T_A|X_0{=}j]$, by the Strong Markov property.
- $P(Y_1{=}j|X_0{=}i) = \tilde{P}_{ij} = \frac{q_{ij}}{-q_{ii}}$.

- $\mathbb{E}[J_1|X_0{=}i] = -1/q_{ii}$, since this is the mean of an exponential random variable with parameter $-q_{ii}$.

Substituting these calculations gives

$$\tau_i = \frac{1}{-q_{ii}} + \sum_{j\neq i} \tau_j \frac{q_{ij}}{-q_{ii}}$$

Rearranging gives the desired equations. □

References

Coifman, R. R., Lafon, S., Lee, A. B., Maggioni, M., Nadler, B., Warner, F., and Zucker, S. W. (2005). Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. *Proceedings of the National Academy of Sciences*, 102(21):7426–7431.

Dobrow, R. P. (2016). *Introduction to Stochastic Processes with R*. Wiley.

Durrett, R. (2012). *Essentials of Stochastic Processes*. Springer, 2nd edition.

Grimmett, G. and Stirzaker, D. (2001). *Probability and Random Processes*. Oxford University Press.

Koralov, L. B. and Sinai, Y. G. (2010). *Theory of Probability and Random Processes*. Springer.

Lafon, S. and Lee, A. B. (2006). Diffusion maps and coarse-graining: A unified framework for dimensionality reduction, graph partitioning, and data set parameterization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(9):1393–1403.

Norris, J. R. (1997). *Markov Chains*. Cambridge University Press.

Pavliotis, G. A. (2014). *Stochastic Processes and Applications*. Springer.