# CONTINUOUS TIME MARKOV CHAINS

STEVEN P. LALLEY

## 1. Introduction

Discrete-time Markov chains are useful in simulation, since updating algorithms are easier to construct in discrete steps. They can also be useful as crude models of physical, biological, and social processes. However, in the physical and biological worlds time runs continuously, and so discrete-time mathematical models are not always appropriate. This is especially true in population biology – organisms do not reproduce, infect each other, etc., synchronously, as in the Galton-Watson and Reed-Frost models. In such situations *continuous-time* Markov chains are often more suitable as models.

**Definition 1.1.** A *continuous-time Markov chain* with finite or countable state space $\mathcal{X}$ is a family $\{X_t = X(t)\}_{t\geq 0}$ of $\mathcal{X}-$valued random variables such that

(a) The paths $t \mapsto X(t)$ are right-continuous step functions; and
(b) For any set of times $t_i < t_{i+1} = t_i + s_{i+1}$ and states $x_i \in \mathcal{X}$, with $t_0 = 0$,

$$P(X(t_{k+1}) = x_{k+1} \,|, X(t_i) = x_i \; \forall\, i \leq k) = P(X(s_{k+1}) = x_{k+1} \,|\, X(0) = x_k). \tag{1.1}$$

This is not quite the standard definition: most elementary textbooks omit the condition (a), but then tacitly assume it later. (Ross refers to property (a) as "regularity".) This condition guarantees that the Markov chain makes only finitely many jumps in any finite time interval. There are interesting examples (due to Blackwell) of processes $X(t)$ that satisfy the Markov property (b) but make *infinitely* many jumps in every finite time interval.

Condition (b) is the natural continuous-time analogue of the Markov property. It requires two things: first, that the future is conditionally independent of the past given the present, and second, that the transition probabilities are time-homogeneous. An equivalent way of writing the equation (1.1) is this:

$$P(X(t_j) = x_j \; \forall\, 0 \leq j \leq k+1) = P(X(0) = x_0) \prod_{j=0}^{k} p_{s_{j+1}}(x_j, x_{j+1}) \tag{1.2}$$

where

$$p_t(x,y) := P(X(t) = y \,|\, X(0) = x_0). \tag{1.3}$$

This has an interesting consequence: For any scalar $\Delta > 0$, the discrete sequence $\{X(n\Delta)\}_{n\geq 0}$ is a discrete-time Markov chain, with one-step transition probabilities $p_\Delta(x,y)$.

**Example 1.1.** Let $N(t)$ be the Poisson counting process with rate $\lambda > 0$. Then $N(t)$ is a continuous-time Markov chain on the state space $\mathbb{Z}_+$ of nonnegative integers. This is an easy consequence of the independent-increments property of the Poisson process.

As in the discrete-time case, I will use the notation $P^x$ to denote the probability measure under which the Markov chain starts in state $x$, that is, $P^x\{X(0) = x\} = 1$. Similarly, if $\nu$ is a probability distribution on $\mathcal{X}$ then $P^\nu$ denotes the probability measure under which the initial state is chosen at random according to $\nu$: equivalently,

$$P^\nu = \sum_x \nu_x P^x.$$

The transition probabilities $p_t(x, y)$ can be put into a matrix with rows indexed by $x$ and columns by $y$: this matrix will be denoted $\mathbb{P}_t$. The same simple calculation that gave the Chapman-Kolmogorov equations in the discrete-time case shows that

$$\mathbb{P}_{t+s} = \mathbb{P}_t \mathbb{P}_s \quad \forall\, s, t \geq 0. \tag{1.4}$$

In general, a family of matrices $\{\mathbb{P}_t\}_{t\geq 0}$ with this property is called a *semigroup*. Note that (1.4) requires that $\mathbb{P}_0 = I$.

## 2. First Jump Time

Assume now that $X(t)$ is a continuous-time Markov chain on the state space $\mathcal{X}$. Since the sample paths of a continuous-time Markov chain are right-continuous step functions, the time of the first jump from one state to another is a well-defined, positive random variable $T$:

$$T := \min\{t > 0 \,:\, X(t) \neq X(0)\}. \tag{2.1}$$

**Proposition 2.1.** *For each state $x$, there exists a scalar $\lambda_x > 0$ such that under $P^x$ the first jump time $T$ has the exponential distribution with parameter $\lambda_x$.*

*Proof.* It is easy to give a heuristic argument, based on the Markov property, that the distribution of $T$ has the memoryless property that characterizes the exponential distribution. See Ross, sec. 5.2, for this. However, it is a bit tricky to make this argument rigorous, because as stated the Markov property (1.1) only allows conditioning on finitely many time points. For a rigorous argument, use the embedded discrete-time Markov chains

$$Y_n^m := X(n/2^m), \quad \text{where}\, n = 0, 1, 2, \ldots. \tag{2.2}$$

For each integer $m \geq 0$, define $2^m \tau_m$ to be the smallest integer $n \geq 1$ such that $Y_n^m \neq Y_0^m$ (that is, look at the continuous-time Markov chain $X(t)$ only at times $n/2^m$, and pick the time of the first jump). Notice that increasing $m$ has the effect of chopping up the time interval $[0, \infty)$ more and more finely, making it possible that a jump missed by the coarser partitioning will now show up: hence,

$$\tau_m \geq \tau_{m+1} \quad \forall\, m \geq 0.$$

Eventually, the partitioning of time will be so fine that there will not be more than one jump in any interval before $T$, and so the values $\tau_m$ will decrease to $T$:

$$\lim_{m\to\infty} \tau_m = T. \tag{2.3}$$

On the other hand, since $Y_n^m$ is a discrete-time Markov chain, the Markov property for discrete-time chains implies

$$P^x\{\tau_m > n/2^m\} = p_{2^{-m}}(x, x)^n,$$

that is, $2^m\tau_m$ has a geometric distribution. Since $\tau_m \to T$ it follows that $T$ has an exponential distribution, by the following lemma. $\square$

**Lemma 2.2.** *Suppose that $G_m$ are geometric random variables such that with probability one,*

$$\lim_{m\to\infty} \downarrow 2^{-m} G_m = H \tag{2.4}$$

*as $m \to \infty$. Then either $H = 0$ with probability one, or $H$ has an exponential distribution.*

**Note:** (1) The hypothesis that the random variables $2^{-m}G_m$ are decreasing is unnecessary, but it makes the proof slightly easier. (2) There is a close connection between this lemma and the Poisson approximation to the Binomial distribution. The Poisson approximation theorem states that if you toss a $\lambda/n$ coin about $nt$ times, then the number of heads will be approximately Poisson-$\lambda t$. Lemma 2.2 implies that if you toss a $\lambda/2^m$ coin until the time $2^m\tau_m$ of the first head, then $\tau_m$ will have approximately an exponential-$\lambda$ distribution. The connection is that the exponential-$\lambda$ distribution is the distribution of the time of the first occurrence in a Poisson process with rate $\lambda$.

*Proof.* Let $G_m$ be geometric with parameter $a_m$, that is, $G_m$ is the number of tosses until the first Head for an $a_m-$coin. Since the random variables $2^{-m}G_m$ decrease to $H$, the event $\{H \geq t\}$ is the decreasing limit of the events $\{2^{-m}G_m \geq t\}$. Consequently, for any $t > 0$,

$$P\{H \geq t\} = \lim_{m\to\infty} P\{G_m \geq 2^m t\} = \lim_{m\to\infty} (1 - a_m)^{[2^m t]} \tag{2.5}$$

where $[2^m t]$ denotes the greatest integer $\leq 2^m t$. Either this limit is 0 for all $t > 0$, in which case $H = 0$ with probability one, or there is some value of $t$ for which it is positive. However, the limit must be $< 1$ strictly, because if the limit were $= 1$ for some $t = \varepsilon > 0$, then it would follow that the limit is one for $t = 2\varepsilon, 4\varepsilon, \dots,$ which would imply that $P\{H = \infty\} = 1$. So assume that there is some $\varepsilon > 0$ such that

$$\lim_{m\to\infty} (1 - a_m)^{[2^m \varepsilon]} = e^{-\lambda\varepsilon}$$

for some $\lambda > 0$. Then by (2.5), for every $s > 0$,

$$P\{H \geq s\varepsilon\} = e^{-\lambda\varepsilon s},$$

and so $H$ has an exponential distribution with parameter $\lambda\varepsilon$. $\square$

## 3. The Post-Jump Process

**Proposition 3.1.** *Let $X(t)$ be a continuous-time Markov chain with first jump time $T$. Then there is a stochastic matrix $\mathbb{A} = (a_{x,y})_{x,y\in\mathcal{X}}$ such that for every pair of states $x, y$ and all $t > 0$,*

$$P^x(X(T) = y \,|\, T = t) = a_{x,y}. \tag{3.1}$$

*In other words, the state $X(T)$ entered at the time of the first jump is independent of the time $T$ at which the jump occurs.*

**Note:** It follows from (3.1) that the *unconditional* distribution of $X(T)$ under $P^x$ is also given by the right side of (3.1).

*Proof.* Once again, let $Y_n^m = Y^m(n)$ be the embedded discrete-time Markov chain defined by (2.2), that is, look at the continuous-time Markov chain $X(t)$ only at times that are integer multiples of $2^{-m}$. As in the proof of Proposition 2.1, let $2^m\tau_m$ be the smallest $n \geq 1$ such that $Y_n^m \neq Y_0^m$, and recall that $\tau_m \downarrow T$. Since the paths of the process $X(t)$ are *right*-continuous,

$$X_T = \lim_{m\to\infty} Y^m(2^m\tau_m). \tag{3.2}$$

In other words, for all sufficiently large $m$, the value of $Y^m(\cdot)$ at the time of its first jump will coincide with the value of $X(\cdot)$ at *its* first jump time $T$.

To prove that $X(T)$ and $T$ are independent, I will show that for each time $t > 0$ and each state $y \neq x$ the events $\{T \geq t\}$ and $\{X(T) = y\}$ are independent. Fix $t > 0$, and for each $m \geq 1$ let $n_m = n_m(t)$ be the smallest integer $\geq 2^m t$. Then by (2.3) and (3.2),

$$\begin{aligned}
P^x\{X(T) = y;\, T \geq t\} &= \lim_{m\to\infty} P^x\{Y^m(2^m\tau_m) = y;\, \tau_m \geq t\}\\
&= \lim_{m\to\infty} \sum_{n\geq n_m} p_{2^{-m}}(x,x)^{n-1} p_{2^{-m}}(x,y)\\
\Longrightarrow P^x\{T \geq t\} &= \lim_{m\to\infty} \sum_{n\geq n_m} p_{2^{-m}}(x,x)^{n-1} \sum_{y\neq x} p_{2^{-m}}(x,y)\\
\Longrightarrow P^x\{X(T) = y \,|\, T \geq t\} &= \lim_{m\to\infty} p_{2^{-m}}(x,y) \Big/ \sum_{z\neq x} p_{2^{-m}}(x,z) := a_{x,y}.
\end{aligned}$$

This limit clearly does not depend on the value of $t$. It follows that $X(T)$ is independent of $T$, which implies (3.1). Since the state $X(T)$ entered at the time of the first jump is a random variable with values in the set $\mathcal{X} \setminus \{x\}$, $\sum_{y\neq x} a_{x,y} = 1$. □

**Proposition 3.2.** *Let $X(t)$ be a continuous-time Markov chain with first jump time $T$. Define the* post-$T$ *process*

$$X^*(t) = X(T + t). \tag{3.3}$$

*Then under $P^x$ the post-$T$ process is independent of $T$, and has the same law as does the process $X(t)$ under the measure $P^{\nu_x}$, where $\nu_x(y) = a_{x,y}$ is the distribution of $X(T)$ under $P^x$. More precisely, for any times $t_i < t_{i+1} = t_i + s_{i+1}$ and states $x_i \in \mathcal{X}$, with $t_0 = 0$ and $y = x_0$,*

$$P^x(X^*(t_i) = x_i \,\forall\, 0 \leq i \leq k \,|\, T = t) = a_{x,y} \prod_{j=0}^{k} p_{s_{j+1}}(x_j, x_{j+1}). \tag{3.4}$$

*Proof.* This is very similar to the proof of Proposition 3.1, and therefore is omitted. □

## 4. Structure of a Continuous-Time Markov Chain

Propositions 2.1 and 3.2 have far-reaching implications. Together, they imply that a continuous-time Markov chain $X(t)$ started at $X(0) = x$ can be built as follows: First, generate $T$ – by Proposition 2.1 this can be done simply by generating an exponential-$\lambda_x$

random variable. The Markov chain $X(t)$ will remain at state $x$ until time $T = T_1$. Second, choose a state $X(T)$ at random according to the distribution

$$P^x\{X(T) = y\} = a_{x,y}, \tag{4.1}$$

independent of $T$. Third, given that $X(T) = y$, build an independent Markov chain $X^*(t)$ started at $y$, and attach it to the initial segment $X(t) = x$, $t < T$. This is a recursive description, and so it can be iterated: To build $X^*(t)$, generate $T^*$ $\sim$Exponential-$\lambda_y$ and $X^*(T^*)$ using the law $a_{y,\cdot}$, and continue indefinitely. Let $T_1, T_2, \dots$ be the jump times of the resulting process $X(t)$. By construction, the sequence

$$X(T_1), X(T_2), X(T_3), \dots \tag{4.2}$$

of successive states visited will itself be a discrete-time Markov chain with transition probability matrix $\mathbb{A}$. This discrete-time Markov chain is called the *embedded jump chain.*

By Proposition 3.1 and induction, each successive state $X(T_n)$ is chosen independently of the previous holding time $T_n - T_{n-1}$, conditional on the previous state $X(T_{n-1})$. Thus, the continuous-time Markov chain $X(t)$ can also be constructed as follows: First, generate a discrete-time Markov chain $Z_n$ with initial state $Z_0 = x$ and transition probability matrix $\mathbb{A}$. Second, generate independent, identically distributed exponential-1 random variables $\sigma_n$, $n \geq 1$. Finally, set

$$X(t) = Z_n \quad \text{for } T_n \leq t < T_{n+1} \tag{4.3}$$

where

$$T_n = \sum_{j=1}^{n} \sigma_j / \lambda_{Z_j}. \tag{4.4}$$

(Recall that if $\sigma$ is a unit exponential, then $\sigma/\lambda$ is exponentially distributed with parameter $\lambda$, equivalently, exponentially distributed with mean $1/\lambda$.)

It is natural to ask whether this construction can be reversed. In other words, given a stochastic matrix $\mathbb{A}$ and an assignment of positive parameters $\lambda_x$ for $x \in \mathcal{X}$, if $Z_n$ is a discrete-time Markov chain with transition probability matrix $\mathbb{A}$ and if $\sigma_n$ are i.i.d. unit exponentials, will (4.3)-(4.4) define a continuous-time Markov chain? The answer, as it turns out, is complicated – It takes up a large chunk of the treatise *Markov Chains with Stationary Transition Probabilities* by K. L. Chung. The difficulty is that if the state space is infinite and the jump rates $\lambda_x$ are unbounded, then the terms of the sums (4.4) can become small so fast that

$$\lim_{n\to\infty} T_n < \infty,$$

and so infinitely many jumps occur in a finite time interval. Such events are called *explosions.*

## 5. Kolmogorov's Forward and Backward Equations

**Definition 5.1.** The *infinitesimal generator* (also called the$Q-matrix$) of a continuous-time Markov chain is the matrix $\mathbb{Q} = (q_{x,y})_{x,y\in\mathcal{X}}$ with entries

$$q_{x,y} = \lambda_x a_{x,y} \tag{5.1}$$

where $\lambda_x$ is the parameter of the holding distribution for state $x$ (Proposition 2.1) and $\mathbb{A} = (a_{x,y})_{x,y\in\mathcal{X}}$ is the transition probability matrix of the embedded jump chain (Proposition 3.1).

**Theorem 1.** *The transition probabilities $p_t(x,y)$ of a continuous-time Markov chain satisfy the following differential equations, called the* Kolmogorov equations *(also called the backward and forward equations, respectively):*

$$\frac{d}{dt}p_t(x,y) = \sum_{z\in\mathcal{X}} q(x,z)p_t(z,y) \quad (BW) \tag{5.2}$$

$$\frac{d}{dt}p_t(x,y) = \sum_{z\in\mathcal{X}} p_t(x,z)q(z,y) \quad (FW). \tag{5.3}$$

**Note:** In matrix form the Kolmogorov equations read

$$\frac{d}{dt}\mathbb{P}_t = \mathbb{Q}\mathbb{P}_t \quad \text{(BW)} \tag{5.4}$$

$$\frac{d}{dt}\mathbb{P}_t = \mathbb{P}_t\mathbb{Q} \quad \text{(FW)}. \tag{5.5}$$

*Proof.* I will prove this only for finite state spaces $\mathcal{X}$. To prove the result in general, it is necessary to deal with the technical problem of interchanging a limit and an infinite series – but the basic idea is the same as in the finite state space case. For details, see Karlin & Taylor, *A Second Course in Stochastic Processes*, Ch. 14.

The Chapman-Kolmogorov equations (1.4) imply that for any $t, \varepsilon > 0$,

$$\varepsilon^{-1}(p_{t+\varepsilon}(x,y) - p_{t(x,y)}) = \sum_{z\in\mathcal{X}} \varepsilon^{-1}(p_\varepsilon(x,z) - \delta(x,x))p_t(z,y) \quad \text{(BW)} \tag{5.6}$$

$$= \sum_{z\in\mathcal{X}} \varepsilon^{-1}p_t(x,z)(p_\varepsilon(z,y) - \delta(z,z)) \quad \text{(FW)} \tag{5.7}$$

where $\delta(x,y)$ is the Kronecker $\delta$ (that is, $\delta(x,y) = 1$ if $x = y$ and $= 0$ if $x \neq y$). Since the sum (5.6) has only finitely many terms, it follows that to prove the backward equation (5.2) it suffices to prove that

$$\lim_{\varepsilon\to 0} \varepsilon^{-1}(p_\varepsilon(x,z) - \delta(x,x)) = q_{x,z} = \lambda_x a_{x,z}. \tag{5.8}$$

Consider how the Markov chain might find its way from state $x$ at time 0 to state $z \neq x$ at time $\varepsilon$ when $\varepsilon > 0$ is small: Either there is just one jump, from $x$ to $z$, or there are two or more jumps before time $\varepsilon$. By Proposition 2.1,

$$P^x\{T_1 \leq \varepsilon\} = 1 - e^{-\lambda_x\varepsilon} = \lambda_x\varepsilon + O(\varepsilon^2).$$

Consequently, the chance that there are two or more jumps before time $\varepsilon$ is of order $O(\varepsilon^2)$, and this is not enough to affect the limit (5.8). Thus, when $\varepsilon > 0$ is small,

$$\varepsilon^{-1}p_\varepsilon(x,z) \approx \lambda_x a_{x,z} \quad \text{for } x \neq z, \text{ and} \tag{5.9}$$
$$\varepsilon^{-1}(p_\varepsilon(x,x) - 1) \approx -\lambda_x.$$

Since $q_{x,z} = \lambda_x a_{x,z}$ for $z \neq x$ and $q_{x,x} = -\lambda_x$, this proves the backward equations (5.2) in the case where the state space $\mathcal{X}$ is finite. A similar argument, this time starting from the equation (5.7),proves the forward equations. $\square$

**Definition 5.2.** A probability distribution $\pi = \{\pi_x\}_{x \in \mathcal{X}}$ on the state space $\mathcal{X}$ is called a *stationary distribution* for the Markov chain if for every $t > 0$,

$$\pi^T \mathbb{P}_t = \pi^T \tag{5.10}$$

**Corollary 5.1.** *A probability distribution $\pi$ is stationary if and only if*

$$\pi^T \mathbb{Q} = 0^T. \tag{5.11}$$

*Proof.* Suppose first that $\pi^T$ is stationary. Take the derivative of each side of (5.10) at $t = 0$ to obtain (5.11). Now suppose, conversely, that $\pi$ satisfies (5.11). Multiply both sides by $\mathbb{P}_{\approx}$ to obtain

$$\pi^T \mathbb{Q} \mathbb{P}_t = 0^T \quad \forall\, t \geq 0.$$

By the Kolmogorov backward equations, this implies that

$$\frac{d}{dt} \pi^T \mathbb{P}_t = 0^T \quad \forall\, t \geq 0;$$

but this means that $\pi^T \mathbb{P}_t$ is constant in time $t$. Since $\lim_{t \to 0} \mathbb{P}_t = I$, equation (5.10) follows. $\square$

## 6. Ergodic Theory of Continuous-Time Markov Chains

The large-time behavior of the transition probabilities can easily be deduced from that for discrete-time Markov chains. Just as in the discrete-time case, we say that a continuous-time Markov chain is *irreducible* if for any two states $x, y$ there exists $t > 0$ such that $p_t(x, y) > 0$. Periodicity plays no role in continuous time, by the following proposition.

**Proposition 6.1.** *If for two states $x, y$ there exists $t > 0$ such that $p_t(x, y) > 0$, then for all $t > 0$ it is the case that $p_t(x, y) > 0$. In particular, if a continuous-time Markov chain is irreducible, then for every $t > 0$, all entries of the transition probability matrix $\mathbb{P}_t$ are strictly positive.*

*Proof.* Let $T_1, T_2, \ldots$ be the jump times, as in section 4 above, and let $X(T_1), X(T_2), \ldots$ be the embedded jump chain. Recall that the jump chain is a discrete-time Markov chain with transition probability matrix $\mathbb{A}$. In order that the continuous-time Markov chain $X(t)$ be irreducible, it is clearly necessary and sufficient that the jump chain be irreducible. (Exercise: Explain why.) In fact, if $x, y$ are two states such that $p_t(x, y) > 0$ for some $t > 0$, then there must be a finite sequence of states $x_0, x_1, \ldots, x_m$ connecting $x_0 = x$ to $x_m = y$, such that $a(x_i, x_{i+1}) > 0$ for each $i$. (If not, it would be impossible for the embedded jump chain $X(T_n)$ to ever reach $y$ starting from $x$.)

Suppose that $x, z$ are two states such that $a(x, z) > 0$. Then by equation (5.9) above,

$$p_\varepsilon(x, z) \sim \varepsilon \lambda_x a_{x,z} > 0$$

for all sufficiently small $\varepsilon > 0$. Consequently, if $a(x, z) > 0$ then $p_t(x, z) > 0$ for *all* $t > 0$, because it is possible, with positive probability, for the Markov chain to jump from $x$ to $z$ in time $\varepsilon$ and then to remain at $z$ for the next $t - \varepsilon$ units of time.

Now suppose that $x, y$ are two states that are connected by a finite path $x_i$ such that $a(x_i, x_{i+1}) > 0$ for each step $i$. Then by the argument of the preceding paragraph, $p_{t/m}(x_i, x_{i+1}) > 0$ for each $i$ and for any $t > 0$, and hence

$$p_t(x,y) \geq \prod_{i=0}^{m-1} p_{t/m}(x_i, x_{i+1}) > 0.$$

□

**Corollary 6.2.** *If the continuous-time Markov chain $X(t)$ is irreducible and has a stationary distribution $\pi$, then the stationary distribution is unique, and for all states $x, y$,*

$$\lim_{t\to\infty} p_t(x,y) = \pi(y). \tag{6.1}$$

*Proof.* By Proposition 6.1, if $X(t)$ is irreducible then for every $t > 0$ all entries of the transition probability matrix $\mathbb{P}_t$ are strictly positive. Hence, in particular, the transition probability matrix $\mathbb{P}_1$ is aperiodic and irreducible, and has stationary distribution $\pi$. Thus, by Kolmogorov's theorem for discrete-time Markov chains,

$$\lim_{n\to\infty} p_n(x,z) = \pi(z)$$

for all states $x, z$. It is easy to deduce the result (6.1) from this: Let $n = [t]$ be the integer part of $t$; then as $t \to \infty$,

$$p_t(x,y) = \sum_z p_n(x,z)p_{t-n}(z,y) \longrightarrow \sum_z \pi(z)p_{t-n}(z,y) = \pi(y),$$

the last because $\pi$ is a stationary distribution. (Exercise: Justify the interchange of limit and summation.) Uniqueness of the stationary distribution clearly follows from (6.1). □