CONTINUOUS-TIME MARKOV CHAINS

1. Definition and First Properties

Definition 1. A *continuous-time Markov chain* on a finite or countable state space $\mathscr{X}$ is a family of $\mathscr{X}$−valued random variables $X_t = X(t)$ indexed by $t \in \mathbb{R}_+$ such that:

(A) The sample paths $t \mapsto X_t$ are right-continuous $\mathscr{X}$−valued step functions with only finitely many discontinuities (jumps) in any finite time interval; and

(B) The process X_t satisfies the *Markov property*, that is, for any choice of states $x_0, x_1, \dots, x_{n+1}$ and times $t_0 < t_1 < \cdots < t_n < t_{n+1}$,

$$P(X(t_{n+1}) = x_{n+1} \mid X(t_i) = x_i \; \forall \, i \le n) = p_{t_{n+1}-t_n}(x_n, x_{n+1}). \tag{1}$$

Property (B) holds, in particular, for times t_j in an arithmetic progression $\Delta\mathbb{Z}_+$, and so for each $\Delta > 0$ the discrete-time sequence $X(n\Delta)$ is a discrete-time Markov chain with one-step transition probabilities $p_\Delta(x, y)$. It is natural to wonder if every discrete-time Markov chain can be embedded in a continuous-time Markov chain; the answer is *no*, for reasons that will become clear in the discussion of the *Kolmogorov differential equations* below.

The probabilities $p_s(x, y)$ are called the *transition probabilities* for the Markov chain, and for the same reason as in the discrete-time case it is often advantageous to view them as being arranged in matrices

$$\mathbb{P}_s = (p_s(x, y))_{x,y \in \mathscr{X}}. \tag{2}$$

For each $s \ge 0$ the transition probability matrix $\mathbb{P}_s$ is stochastic. The one-parameter family $\{\mathbb{P}_s\}_{s \ge 0}$ is called the *transition semigroup*, because the matrices obey a multiplication law: for any $s, t > 0$

$$\mathbb{P}_{t+s} = \mathbb{P}_t \mathbb{P}_s. \tag{3}$$

(These are the natural analogues of the Chapman-Kolmogorov equations for discrete-time chains.) As for discrete-time Markov chains, we denote the initial state x (or initial probability distribution ν on $\mathscr{X}$) by a superscript P^x or P^ν. With this notation, the Markov property can be written in the following equivalent form:

$$P^x\{X(t_j) = x_j \; \forall \, 1 \le j \le n\} = \prod_{j=1}^{n} p_{t_j - t_{j-1}}(x_{j-1}, x_j) \tag{4}$$

with the convention that $x = x_0$ and $t_0 = 0$.

Proposition 1. *The transition semigroup is continuous (in t), that is,*

$$\lim_{s \to 0} \mathbb{P}_s = I \tag{5}$$

Proof. This is an easy consequence of the right-continuity of sample paths: With P^x—probability one, $X_t = x$ for all t near 0, and so $X_t \to x$ in P^x—probability as $t \to 0$. Therefore, the transition probabilities satisfy

$$\lim_{t\to 0} p_t(x,x) = 1 \quad \text{and}$$
$$\lim_{t\to 0} p_t(x,y) = 0 \quad \text{for all } y \neq x.$$

□

Together with the semigroup property (3), Proposition 1 implies that $\mathbb{P}_{t+s} \to \mathbb{P}_t$ as $s \to 0$ for every $t \geq 0$. This is the principal reason for the restriction (A) in Definition 1. In fact there are *discontinuous* semigroups of transition probability matrices: For example, with $\mathscr{X} = \{1,2\}$,

$$\mathbb{P}_0 = I \quad \text{and} \quad \mathbb{P}_t = \begin{pmatrix} .5 & .5 \\ .5 & .5 \end{pmatrix} \tag{6}$$

It is generally impossible to build processes X_t that satisfy the Markov property (B) whose transition probabilities come from such discontinuous semigroups that are not themselves discontinuous *everywhere*.

Example 1. A constant-rate Poisson counting process is a continuous-time Markov chain on $\mathbb{Z}_+$ with transition probabilities

$$p_t(x,y) = (\lambda t)^{y-x} \frac{(\lambda t)^{y-x} e^{-\lambda t}}{(y-x)!} \quad \text{for } x \leq y.$$

Example 2. Let N_t be a standard unit-intensity Poisson counting process, and let $\xi_1, \xi_2, \dots$ be independent, identically distributed random variables from a probability distribution $\{p_k\}_{k\in\mathbb{Z}}$ on the integers. Assume that the Poisson process N_t is independent of the random variables ξ_i. Define

$$X_t := \sum_{j=1}^{N_t} \xi_j. \tag{7}$$

The process X_t is a continuous-time Markov chain on the integers. Such processes are generically called *compound Poisson processes.* In the special case where $p_1 = p_{-1} = 1/2$, the process X_t is called the *continuous-time simple random walk* on the integers.

Example 3. A *pure birth process* with birth rates $\beta_x > 0$ is a continuous-time Markov chain X_t on the nonnegative integers built from i.i.d. unit exponential random variables ξ_i as follows: For some initial state $x \geq 0$, define

$$X_t = X_t^x = \max\{y \geq x : \sum_{j=x+1}^{y} \beta_j^{-1} \xi_j < t\}. \tag{8}$$

We will see later that if $\sum_{j=1}^{\infty} \beta_j^{-1} = \infty$ then the process X_t is well-defined and satisfies the Markov property. Note that the Poisson process with rate λ is a pure birth process (with $\beta_j = \lambda$). Another example is the *Yule process*, for which $\beta_j = j$.

2. Jump Times and the Embedded Jump Chain

Because the paths of a continuous-time Markov chain are step functions, with only finitely many jumps in any finite time interval, the jumps occur at a discrete set of time points $0 < T = T_1 < T_2 < \dots$. Assume, to avoid trivialities, that there are no absorbing states (that is, states x such that $p_t(x,x) = 1$ for all $t \geq 0$).

Theorem 2. *For every state x there is a positive parameter $\lambda_x > 0$ such that under P^x the distribution of T is exponential with mean $1/\lambda_x$, that is,*

$$P^x\{T > t\} = e^{-\lambda_x t} \quad \forall\, t \geq 0. \tag{9}$$

Furthermore, the state $X(T)$ of the Markov chain at the first jump time T is independent of T, and has distribution

$$P^x\{X(T) = y\} = \lim_{n\to\infty} \frac{p_{2^{-n}}(x,y)}{1 - p_{2^{-n}}(x,x)} \quad \textit{for all } y \neq x. \tag{10}$$

The strategy of the proof will be to use the right-continuity of the sample paths to reduce the problem to proving a similar statement about *discrete-time* Markov chains.

Lemma 3. *Let $\{X_n\}_{n\geq 0}$ be a discrete-time Markov chain on a finite or countable set $\mathscr{X}$, with one-step transition probabilities $p(x,y)$. Define τ to be the first time $n \geq 1$ such that $X_n \neq X_0$ (that is, the time of the first jump). Then for any initial state x, under P^x,*

- (A) *the distribution of τ is geometric with parameter $1 - p(x,x)$; and*
- (B) *the random variable X_τ is independent of τ, and has distribution*

$$\begin{aligned} &P^x\{X_\tau = y\} = p(x,y)/(1-p(x,x)) \quad \textit{for } y \neq x; \\ &P^x\{X_\tau = x\} = 0. \end{aligned}$$

Proof of the Lemma. Let's first show that τ has a geometric distribution. For this, observe that if $X_0 = x$ then the event $\{\tau > n\}$ coincides with the event $\{X_i = x \text{ for all } i = 0,1,2,\dots,n\}$; consequently,

$$\begin{aligned} P^x\{\tau > n\} = P^x\{X_i = x \text{ for all } i \leq n\} = p(x,x)^n \quad \Longrightarrow \\ P^x\{\tau = n+1\} = p^n(x,x)(1-p(x,x)). \end{aligned}$$

This shows that the distribution is geometric with parameter $1 - p(x,x)$.

Now consider the *joint* distribution of τ and X_τ. By the same reasoning as above, for any $y \neq x$

$$\begin{aligned} P^x\{\tau = n+1 \text{ and } X_\tau = y\} &= P^x\{X_j = x \text{ for all } j \leq n \text{ and } X_{n+1} = y\} \\ &= p(x,x)^n p(x,y) \\ &= P^x\{\tau = n+1\}\frac{p(x,y)}{1-p(x,x)}. \end{aligned}$$

This shows that X_τ is independent of τ and that X_τ has distribution

$$P^x\{X_\tau = y\} = \frac{p(x,y)}{1-p(x,x)} \text{ for } y \neq x.$$

□

Proof. The key is that for each $n = 1, 2, \dots$ the discrete-time process $(X(k/2^n))_{k=0,1,2,\dots}$ is a discrete-time Markov chain; thus, Lemma 3 applies for each of these. In particular, if $X(0) = x$ is the initial state and τ_n is the time k of the first jump for the nth process $(X(k/2^n))_{k=0,1,2,\dots}$ then τ_n has the geometric distribution with parameter $1 - p_{2^{-n}}(x, x)$.

Now consider the event $\{T \geq t\}$. Because the sample paths of the process are step functions, the event $\{T \geq t\}$ is the intersection of the events $\{\tau_n \geq [2^n t]\}$, where $[s]$ denotes the greatest integer in s. By Lemma 3,

$$P^x\{\tau_n \geq [2^n t]\} = p_{2^{-n}}(x, x)^{[2^n t]} = \exp\{[2^n t] \log p_{2^{-n}}(x, x)\}.$$

Since $\lim_{n\to\infty} P^x\{\tau_n \geq [2^n t]\} = P^x\{T \geq t\}$, and since $P^x\{T \geq t\}$ is not zero for all $t > 0$ (because this would force $T = 0$ with probability 1) it follows that

$$\lambda_x := \lim_{n\to\infty} -2^n \log p_{2^{-n}}(x, x) \tag{11}$$

exists and is nonnegative, and that equation (9) holds. Moreover, the parameter λ_x must be *strictly* positive, because otherwise $P^x\{T \geq t\} = 1$ for all t, and x would be an absorbing state, contrary to our assumptions. Thus, the first jump time T has the exponential distribution with parameter (11).

A similar argument shows that the random variable $X(T)$ is independent of T. Because the paths of the process $X(t)$ are right-continuous step functions, $X(T) = X(\tau_n/2^n)$ for all sufficiently large n. But for each $n = 1, 2, \dots$ the random variable $X(\tau_n/2^n)$ is independent of τ_n, by Lemma 3, and hence independent of the event $\{\tau_n \geq [2^n t]\}$. Since the event $\{T‘ > t\}$ is the intersection of events $\{\tau_n \geq [2^n t]\}$, we have

$$\begin{aligned} P^x\{X(T) = y \text{ and } T \geq t\} &= \lim_{n\to\infty} P^x\{X(\tau_n/2^n) = y \text{ and } \tau_n \geq [2^n t]\} \\ &= \lim_{n\to\infty} P^x\{X(\tau_n/2^n) = y\} P^x\{\tau_n \geq [2^n t]\} \\ &\lim_{n\to\infty} P^x\{X(\tau_n/2^n) = y\} \lim_{n\to\infty} P^x\{\tau_n \geq [2^n t]\} \\ &= P^x\{X(T) = y\} P^x\{T \geq t\}. \end{aligned}$$

This proves that the random variables $X(T)$ and T are independent. Finally, since $X(T) = X(\tau_n/2^n)$ for all sufficiently large n, it follows by Lemma 3 that for any $y \neq x$,

$$P^x\{X(T) = y\} = \lim_{n\to\infty} P^x\{X(\tau_n/2^n) = y\} = \lim_{n\to\infty} \frac{p_{2^{-n}}(x, y)}{1 - p_{2^{-n}}(x, x)}.$$

□

Theorem 4. *Let $X(t)$ be a continuous-time Markov chain that starts in state $X(0) = x$. Then conditional on T and $X(T) = y$, the* post-jump *process*

$$X^*(s) := X(T + s) \tag{12}$$

is itself a continuous-time Markov chain with the transition probabilities $\mathbb{P}_s$ and initial state y. More precisely, there exists a stochastic matrix $\mathbb{A} = (a_{x,y})$ such that for all times $s \geq 0$ and $0 = t_0 < t_1 < t_2 < \dots$, and all states $x, y = y_0, y_1, \dots$,

$$P^x\{T > s \textit{ and } X^*(t_i) = y_i \ \forall\, 0 \leq i \leq n\} = e^{-\lambda_x s} a_{x,y} \prod_{i=1}^{n} p_{t_i - t_{i-1}}(y_{i-1}, y_i). \tag{13}$$

The proof is similar to that of Theorem 2 and therefore is omitted. Theorem 4 provides a recursive description of a continuous-time Markov chain: Start at x, wait an exponential-λ_x random time, choose a new state y according to the distribution $\{a_{x,y}\}_{y\in\mathscr{X}}$, and then begin again at y. The only information from the past that is retained in this recursion is the state y. Thus, an easy induction argument (on n) proves the following:

Corollary 5. *The* embedded jump chain $Y_n := X(T_n)$ *is itself a discrete-time Markov chain with transition probability matrix* $\mathbb{A}$.

3. Kolmogorov Backward and Forward Equations

3.1. **Kolmogorov Equations.**

Definition 2. The *infinitesimal generator* (also called the *Q—matrix*) of a continuous-time Markov chain is the matrix $\mathbb{Q} = (q_{x,y})_{x,y\in\mathscr{X}}$ with entries

$$q_{x,y} = \lambda_x a_{x,y} \tag{14}$$

where λ_x is the parameter of the holding distribution for state x (Theorem 2) and $\mathbb{A} = (a_{x,y})_{x,y\in\mathscr{X}}$ is the transition probability matrix of the embedded jump chain (Theorem 4).

Theorem 6. *The transition probabilities* $p_t(x,y)$ *of a* finite-state *continuous-time Markov chain satisfy the following differential equations, called the* Kolmogorov equations *(also called the backward and forward equations, respectively):*

$$\frac{d}{dt}p_t(x,y) = \sum_{z\in\mathscr{X}} q(x,z)p_t(z,y) \quad \textit{(BW)} \tag{15}$$

$$\frac{d}{dt}p_t(x,y) = \sum_{z\in\mathscr{X}} p_t(x,z)q(z,y) \quad \textit{(FW)}. \tag{16}$$

The transition probabilities of an infinite-state *continuous-time Markov chain satisfy the backward equations, but not always the forward equations.*

Note: In matrix form the Kolmogorov equations read

$$\frac{d}{dt}\mathbb{P}_t = \mathbb{Q}\mathbb{P}_t \quad \text{(BW)} \tag{17}$$

$$\frac{d}{dt}\mathbb{P}_t = \mathbb{P}_t\mathbb{Q} \quad \text{(FW)}. \tag{18}$$

Proof. I will prove this only for finite state spaces $\mathscr{X}$. To prove the backward equations for infinite-state Markov chains, 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. However, for infinite state Markov chains, the validity of the *forward* equations is a very sticky problem – see K. L. Chung's book *Markov Chains with Stationary Transition Probabilities* for the whole story.

The Chapman-Kolmogorov equations (3) imply that for any $t, \varepsilon > 0$,

$$\varepsilon^{-1}(p_{t+\varepsilon}(x,y) - p_t(x,y)) = \sum_{z\in\mathscr{X}} \varepsilon^{-1}(p_\varepsilon(x,z) - \delta(x,x))p_t(z,y) \quad \text{(BW)} \tag{19}$$

$$= \sum_{z\in\mathscr{X}} \varepsilon^{-1} p_t(x,z)(p_\varepsilon(z,y) - \delta(z,z)) \quad \text{(FW)} \tag{20}$$

where $\delta(x,y)$ is the Kronecker δ (that is, $\delta(x,y)=1$ if $x=y$ and $=0$ if $x\neq y$). Since the sum (19) has only finitely many terms, to prove the backward equation (15) it suffices to prove that

$$\lim_{\varepsilon\to 0}\varepsilon^{-1}(p_\varepsilon(x,z)-\delta(x,x))=q_{x,z}=\lambda_x a_{x,z}. \tag{21}$$

Consider how the Markov chain might find its way from state x at time 0 to state $z\neq x$ at time ε when $\varepsilon>0$ is small: Either there is just one jump, from x to z, or there are two or more jumps before time ε. By Theorem 2,

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

Consequently, the chance that there are two or more jumps before time ε is of order $O(\varepsilon^2)$, and this is not enough to affect the limit (21). Thus, when $\varepsilon>0$ is small,

$$\begin{aligned}\varepsilon^{-1}p_\varepsilon(x,z)&\approx\lambda_x a_{x,z}\quad\text{for } x\neq z,\text{ and}\\ \varepsilon^{-1}(p_\varepsilon(x,x)-1)&\approx-\lambda_x.\end{aligned} \tag{22}$$

Since $q_{x,z}=\lambda_x a_{x,z}$ for $z\neq x$ and $q_{x,x}=-\lambda_x$, this proves the backward equations (15) in the case where the state space $\mathscr{X}$ is finite. A similar argument, this time starting from the equation (20),proves the forward equations. □

3.2. **Stationary Distributions.**

Definition 3. A probability distribution $\pi=\{\pi_x\}_{x\in\mathscr{X}}$ on the state space $\mathscr{X}$ is called a *stationary distribution* for the Markov chain if for every $t>0$,

$$\pi^T\mathbb{P}_t=\pi^T \tag{23}$$

A continuous-time Markov chain is said to be *irreducible* if any two states communicate. It is not difficult to show (exercise!) that a continuous-time Markov chain X_t is irreducible if and only if for each $\Delta>0$ the discrete-time Markov chain $X_{n\Delta}$ is irreducible. Nor is it difficult to show that for every $\Delta>0$ discrete-time Markov chain $X_{n\Delta}$ is *aperiodic* (use Theorem 2).

Corollary 7. *If an irreducible continuous-time Markov chain has a stationary distribution π_x then it is unique, and for each pair of states x,y,*

$$\lim_{t\to\infty}p_t(x,y)=\pi_y \tag{24}$$

Proof. For any $\Delta>0$ the discrete-time chain $X(n\Delta)$ is aperiodic and irreducible, so Kolmogorov's theorem for discrete-time chains implies uniqueness of the stationary distribution, and the convergence

$$\lim_{n\to\infty}p_{n\Delta}(x,y)=\pi_y$$

Continuity of the semigroup $\mathbb{P}_t$ (Proposition 1) therefore implies (24). □

In practice, it is often difficult to calculate stationary distributions by directly solving the equations (23), in part because it isn't always possible to solve the Kolmogorov equations (15)–(16) in a useful closed form. Nevertheless, the Kolmogorov equations lead to another characterization of stationary distributions that often leads to explicit formulas even when the equations (15)–(16) cannot be solved:

Corollary 8. *A probability distribution π is stationary if and only if*

$$\pi^T\mathbb{Q}=0^T. \tag{25}$$

Proof. Suppose first that π^T is stationary. Take the derivative of each side of (23) at $t = 0$ to obtain (25). Now suppose, conversely, that π satisfies (25). Multiply both sides by $\mathbb{P}_t$ 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 (23) follows. □

3.3. **Matrix Exponentials and the Kolmogorov Equations.**

Definition 4. If A is a square matrix then its exponential is defined by

$$e^A := \exp\{A\} := \sum_{n=0}^{\infty} \frac{A^n}{n!} \tag{26}$$

The infinite sum converges, because the matrix norms of the partial sums are bounded, by the triangle inequality and the fact that the power series for the *scalar* exponential function converges:

$$\left\| \sum_{n=m+1}^{m+k} \frac{A^n}{n!} \right\| \leq \sum_{n=m+1}^{m+k} \frac{\|A\|^n}{n!} \leq \sum_{n=m+1}^{\infty} \frac{\|A\|^n}{n!} \longrightarrow 0$$

as $m \to \infty$.

Special Case: Diagonal Matrices. Suppose that A is a diagonal matrix, with diagonal entries λ_i. Then

$$\exp\{A\} = \begin{pmatrix} e^{\lambda_i} & 0 & 0 & \cdots & 0 \\ 0 & e^{\lambda_2} & 0 & \cdots & 0 \\ \cdots & & & & \\ 0 & 0 & 0 & \cdots & e^{\lambda_m}. \end{pmatrix}$$

Special Case: $A = UDU^{-1}$. Suppose that A is similar to D, that is, $A = UDU^{-1}$ for some invertible matrix U. Then

$$\exp\{A\} = \exp\{UDU^{-1}\} = U\exp\{D\}U^{-1}.$$

Consequently, if A can be diagonalized then its exponential e^A can be computed by exponentiating the eigenvalues λ_i.

Proposition 9. *If A and B are square $m \times m$ matrices such that $AB = BA$ then*

$$\exp\{A + B\} = \exp\{A\}\exp\{B\}. \tag{27}$$

Consequently, for any $m \times m$ square matrix A with real entries and all $s, t \in \mathbb{R}$,

$$\exp\{(s + t)A\} = \exp\{sA\}\exp\{tA\}, \tag{28}$$

and so the mapping $t \mapsto e^{At}$ is a group homomorphism from the additive group $(\mathbb{R}, +)$ into the multiplicative group $GL_m(\mathbb{R})$ of invertible $m \times m$ matrices with real entries.

CAUTION: The multiplication law (27) is not generally true if A and B do not commute.

Proof. This should remind you of the calculations we did in proving some of the fundamental connections between the binomial and Poisson distributions (in particular, the superposition and thinning theorems):

$$\begin{aligned}
\exp\{A+B\} &= \sum_{n=0}^{\infty} \frac{(A+B)^n}{n!} \\
&= \sum_{n=0}^{\infty}\sum_{k=0}^{n} \binom{n}{k} \frac{A^k B^{n-k}}{n!} \\
&= \sum_{n=0}^{\infty}\sum_{k=0}^{n} \frac{A^k B^{n-k}}{k!(n-k)!} \\
&= \sum_{k=0}^{\infty}\sum_{m=0}^{\infty} \frac{A^k B^m}{k!m!} \\
&= \sum_{k=0}^{\infty} \frac{A^k}{k!} \sum_{m=0}^{\infty} \frac{B^m}{m!} \\
&= \exp\{A\}\exp\{B\}.
\end{aligned}$$

This proves (27); the relation (28) follows immediately. □

Corollary 10. *For any square matrix A,*

$$\frac{d}{dt}\exp\{tA\} = A\exp\{tA\} = \exp\{tA\}A. \tag{29}$$

Proof. Relation (28) implies that

$$\frac{\exp\{(t+s)A\} - \exp\{tA\}}{s} = \exp\{tA\}\frac{\exp\{sA\} - I}{s} = \frac{\exp\{sA\} - I}{s}\exp\{tA\}$$

But as $s \to 0$,

$$\frac{\exp\{sA\} - I}{s} = \sum_{n=1}^{\infty} \frac{s^n A^n}{s\,n!} \longrightarrow A.$$

(Why?) □

Corollary 11. *The (unique) solution to the Kolmogorov backward equations* (17) *is*

$$\mathbb{P}_t = \exp\{t\mathbb{Q}\}. \tag{30}$$

Proof. The matrix function on the right side satisfies the same first-order differential equations, and the same initial condition. □