Lecture-18: Continuous Time Markov Chains

1 Markov Process

Definition 1.1. Consider a real-valued stochastic process $X : \Omega \to \mathcal{X}^{\mathbb{R}_+}$ indexed by positive reals and state space $\mathcal{X}$, adapted to its natural filtration $\mathcal{F}_\bullet = (\mathcal{F}_t : t \in \mathbb{R}_+)$ where $\mathcal{F}_t \triangleq \sigma(X_s, s \leqslant t)$ for all $t \in \mathbb{R}_+$. Then, X is a *Markov process* if it satisfies the Markov property. That is, the distribution of the future states conditioned on the present, is independent of the past, and hence for any Borel measurable set $A \in \mathcal{B}(\mathcal{X})$ and times $s \leqslant t \in \mathbb{R}_+$, we have

$$P(\{X_t \in A\} \mid \mathcal{F}_s) = P(\{X_t \in A\} \mid \sigma(X_s)).$$

Definition 1.2. A Markov process $X : \Omega \to \mathcal{X}^{\mathbb{R}_+}$ with countable state space $\mathcal{X}$ is called *continuous time Markov chain (CTMC)*.

Remark 1. The Markov property for the CTMCs can be interpreted as follows. For all times $0 < t_1 < \cdots < t_m < t$ and states $x_1, \ldots, x_m, y \in \mathcal{X}$, we have

$$P(\{X_t = y\} \mid \cap_{k=1}^m \{X_{t_k} = x_k\}) = P(\{X_t = y\} \mid \{X_{t_m} = x_m\}).$$

Example 1.3 (Counting process). Any simple counting process $N : \Omega \to \mathbb{Z}_+^{\mathbb{R}_+}$ with independent increments (possibly non stationary) is a continuous time Markov chain. Countability of the state space is clear from the definition of the counting process. For Markov property, we observe that for $t > s$, the increment $N_t - N_s$ is independent of $\mathcal{F}_s$. Let $\mathcal{F}_\bullet = (\mathcal{F}_t : t \in \mathbb{R}_+)$ be the natural filtration for process N, then

$$\mathbb{E}[\mathbb{1}_{\{N_t=j\}} \mid \mathcal{F}_s] = \sum_{i \in \mathbb{Z}_+} \mathbb{E}[\mathbb{1}_{\{N_t=j, N_s=i\}} \mid \mathcal{F}_s] = \sum_{i \in \mathbb{Z}_+} \mathbb{1}_{\{N_s=i\}} \mathbb{E}\mathbb{1}_{\{N_t - N_s = j-i\}} = \mathbb{E}[\mathbb{1}_{\{N_t=j\}} \mid \sigma(N_s)].$$

1.1 Transition probability kernel

Definition 1.4. We define the *transition probability* from state x at time s to state y at time $t+s$ as

$$P_{xy}(s, s+t) \triangleq P(\{X_{s+t} = y\} \mid \{X_s = x\}).$$

Definition 1.5. The Markov process has *homogeneous* transitions for all states $x, y \in \mathcal{X}$ and all times $s, t \in \mathbb{R}_+$, if

$$P_{xy}(t) \triangleq P_{xy}(0,t) = P_{xy}(s, s+t).$$

We denote the *transition probability kernel* by $P : \mathbb{R}_+ \to [0,1]^{\mathcal{X} \times \mathcal{X}}$, such that the transition probability matrix at time t is $P(t) \triangleq (P_{xy}(t) : x, y \in \mathcal{X})$.

Remark 2. We will mainly be interested in continuous time Markov chains with homogeneous jump transition probabilities. We will assume that the sample path of the process X is right continuous with left limits at each time $t \in \mathbb{R}_+$.

Remark 3. Conditioned on the initial state of the process is x, we denote the conditional probability for any event $A \in \mathcal{F}$ as $P_x(A) \triangleq P(A \mid \{X_0 = x\})$ and the conditional expectation for any random variable $Y : \Omega \to \mathbb{R}$ as $\mathbb{E}_x Y \triangleq \mathbb{E}[Y \mid \{X_0 = x\}]$.

Lemma 1.6 (Stochasticity). *Transition probability kernel $P : \mathbb{R}_+ \to [0,1]^{\mathcal{X} \times \mathcal{X}}$ at each time $t \in \mathbb{R}_+$ is a stochastic matrix, i.e. $P(t) \in \mathcal{M}(\mathcal{X})^{\mathcal{X}}$ for each $t \in \mathbb{R}_+$.*

Proof. From the countable partition of the state space $\mathcal{X}$, we can write $1 = P_x(\{X_t \in \mathcal{X}\}) = \sum_{y\in\mathcal{X}} P_{xy}(t)$ for any state $x \in \mathcal{X}$. □

Lemma 1.7 (Semigroup property). *Transition probability kernel $P : \mathbb{R}_+ \to \mathcal{M}(\mathcal{X})^{\mathcal{X}}$ satisfies the semigroup property, i.e. $P(s+t) = P(s)P(t)$ for all $s,t \in \mathbb{R}_+$.*

Proof. From the Markov property and homogeneity of CTMC, and law of total probability, we can write

$$P_{xy}(s+t) = P_{xy}(0,s+t) = \sum_{z\in\mathcal{X}} P_{xz}(0,s)P_{zy}(s,s+t) = \sum_{z\in\mathcal{X}} P_{xz}(0,s)P_{zy}(0,t) = [P(s)P(t)]_{xy}.$$

□

Lemma 1.8 (Continuity). *Transition probability kernel $P : \mathbb{R}_+ \to \mathcal{M}(\mathcal{X})^{\mathcal{X}}$ for a homogeneous continuous time Markov chain $X : \Omega \to \mathcal{X}^{\mathbb{R}_+}$ is a continuous function of time $t \in \mathbb{R}_+$, such that $\lim_{t\downarrow 0} P(t) = I$, the identity matrix.*

Proof. We will first show the continuity of transition kernel at time $t = 0$. From right continuity of sample paths for process X, we have $\lim_{t\downarrow 0} X_t = X_0$ and from continuity of probability functions we get $\lim_{t\downarrow 0} P_x \{X_t = y\} = P_x \{\lim_{t\downarrow 0} X_t = y\} = I_{xy}$.

Fix a time $t \in \mathbb{R}_+$, to write the difference $P(t+h) - P(t) = P(t)(P(h) - I)$ using the semigroup property of the transition kernel. The continuity of transition kernel at time $t = 0$ and boundedness of $P(t)$ implies continuity of $P(t)$ at all times $t \in \mathbb{R}_+$. □

Remark 4. Consider a time-homogeneous continuous time Markov chain $X : \Omega \to \mathcal{X}^{\mathbb{R}_+}$ with transition kernel $P : \mathbb{R}_+ \to \mathcal{M}(\mathcal{X})^{\mathcal{X}}$. Then, for all times $0 < t_1 < \cdots < t_m$ and states $x_0, x_1, \dots, x_m \in \mathcal{X}$, we have

$$P(\cap_{k=1}^m \{X_{t_k} = x_k\} \,|\, \{X_0 = x_0\}) = P_{x_0x_1}(t_1)P_{x_1x_2}(t_2 - t_1)\dots P_{x_{m-1}x_m}(t_m - t_{m-1}).$$

If the initial distribution is $\nu_0 \in \mathcal{M}(\mathcal{X})$ such that $\nu_0(x) = P\{X_0 = x\}$ for each $x \in \mathcal{X}$, then we observe that all finite dimensional distributions of the continuous time Markov chain X are governed by the initial distribution ν_0 and the transition probability kernel P. That is,

$$P\Big(\cap_{k=1}^m \{X_{t_k} = x_k\}\Big) = \sum_{x_0\in\mathcal{X}} \nu_0(x_0)P_{x_0x_1}(t_1)\dots P_{x_{m-1}x_m}(t_m - t_{m-1}).$$

1.2 Generator matrix

Definition 1.9 (Exponentiation of a matrix). For a matrix $A \in \mathbb{R}^{m\times m}$ or an operator $A : \mathbb{Z}_+ \times \mathbb{Z}_+ \to \mathbb{R}$ with spectral radius less than unity, we can define $e^A \triangleq I + \sum_{n\in\mathbb{N}} \frac{A^n}{n!}$.

Lemma 1.10. *The transition probability kernel for a homogeneous continuous time Markov chain X can be written in terms of a constant matrix $e^Q \triangleq P(1)$, as $P(t) = e^{tQ}$ for all $t \in \mathbb{R}_+$.*

Proof. This follows from the semigroup property and the continuity of transition probability kernel at all times. In particular, we notice that $P(n) = P(1)^n$ and $P(\frac{1}{m}) = P(1)^{\frac{1}{m}}$ for all $m,n \in \mathbb{N}$. Since, any rational number $q \in \mathbb{Q}$ can be expressed as a ratio of integers with no common divisor, we get $P(q) = P(1)^q$ for any $q \in \mathbb{Q}$. Since the rationals are dense in reals, for any real number $t \in \mathbb{R}_+$, we can find a sequence of rationals $q \in \mathbb{Q}^{\mathbb{N}}$ that converges to $t = \lim_{n\in\mathbb{N}} q_n$. Further, from the continuity of P and exponential function, it follows that

$$P(t) = P(\lim_{n\in\mathbb{N}} q_n) = \lim_{n\in\mathbb{N}} P(q_n) = \lim_{n\in\mathbb{N}} P(1)^{q_n} = P(1)^{\lim_{n\in\mathbb{N}} q_n} = P(1)^t.$$

Since the choice of t was arbitrary, the relationship holds for all $t \in \mathbb{R}_+$. The result follows from the definition of Q such that $e^Q = P(1)$. □

Remark 5. From Lemma 1.10 for a homogeneous continuous time Markov chain $X : \Omega \to \mathcal{X}^{\mathbb{R}_+}$, we can write the probability transition kernel function $t \mapsto P(t) = e^{tQ}$, where $e^Q = P(1)$. The matrix $Q \in \mathbb{R}^{\mathcal{X}\times\mathcal{X}}$ is called the generator matrix for the homogeneous continuous time Markov chain X. From the Definition 1.9 for the exponentiation of matrix, this implies that

$$P(t) = I + \sum_{n\in\mathbb{N}} \frac{t^n}{n!} Q^n, \quad t \in \mathbb{R}_+. \tag{1}$$

This relation implies that the probability transition kernel can be written in terms of this fundamental generator matrix Q.

Definition 1.11 (Generator matrix). For a homogeneous continuous time Markov chain $X : \Omega \to \mathcal{X}^{\mathbb{R}_+}$ with transition probability kernel $P : \mathbb{R}_+ \to \mathcal{M}(\mathcal{X})^{\mathcal{X}}$, the *generator matrix* $Q \in \mathbb{R}^{\mathcal{X}\times\mathcal{X}}$ is defined as the following limit when it exists

$$Q \triangleq \lim_{t\downarrow 0} \frac{P(t) - I}{t}.$$

Remark 6. From Eq. (1), it is clear that the generator matrix is the limit defined above.

Remark 7. From the semigroup property of probability kernel function and definition of generator matrix, we get the backward equation for all $t \in \mathbb{R}_+$,

$$\frac{dP(t)}{dt} = \lim_{s\downarrow 0} \frac{P(s+t) - P(t)}{s} = \lim_{s\downarrow 0} \frac{(P(s) - I)}{s} P(t) = QP(t).$$

Similarly, we can also get the forward equation for all $t \in \mathbb{R}_+$,

$$\frac{dP(t)}{dt} = \lim_{s\downarrow 0} \frac{P(s+t) - P(t)}{s} = P(t) \lim_{s\downarrow 0} \frac{(P(s) - I)}{s} = P(t)Q.$$

Both these results need a formal justification for exchange of limits and summation when $\mathcal{X}$ is countable, and we next present a formal proof for these two equations.

Theorem 1.12 (backward equation*). *For a homogeneous continuous time Markov chain $X : \Omega \to \mathcal{X}^{\mathbb{R}_+}$ with transition probability kernel $P : \mathbb{R}_+ \to \mathcal{M}(\mathcal{X})^{\mathcal{X}}$ and generator matrix $Q \in \mathbb{R}^{\mathcal{X}\times\mathcal{X}}$, we have for all $t \in \mathbb{R}_+$.*

$$\frac{dP(t)}{dt} = QP(t).$$

Proof. We fix states $x, y \in \mathcal{X}$ and a finite subset $F \subseteq \mathcal{X}$ containing x. In the following, we consider the lim inf and lim sup of (x, y)th term of $\frac{1}{s}(P(t+s) - P(s)) = \frac{1}{s}(P(s) - I)P(t)$. We observe that

$$P_{xy}(t+s) - P_{xy}(t) = \sum_{z\in\mathcal{X}} (P_{xz}(s) - I_{xz}) P_{zy}(t), \qquad \sum_{z\in F} (P_{xz}(s) - I_{xz}) + \sum_{z\notin F} (P_{xz}(s) - I_{xz}) = 0.$$

We observe that $\sum_{z\in F}(P_{xz}(s) - I_{xz}) \leqslant 0$ iff $x \in F$. Together with this fact and that $P_{zy}(t) \in [0,1]$, we obtain for any $x \in F$

$$\sum_{z\in F} (P_{xz}(s) - I_{xz}) P_{zy}(t) \leqslant \sum_{z\in\mathcal{X}} (P_{xz}(s) - I_{xz}) P_{zy}(t) \leqslant \sum_{z\in F} (P_{xz}(s) - I_{xz}) P_{zy}(t) - \sum_{z\in F} (P_{xz}(s) - I_{xz}). \tag{2}$$

Taking $\liminf_{s\downarrow 0}$ on the both sides of the first inequality in (3) and $\limsup_{s\downarrow 0}$ on the both sides of the second inequality in (3), exchanging limit and finite sum, we obtain that

$$\sum_{z\in F} Q_{xz} P_{zy}(t) \leqslant \liminf_{s\downarrow 0} \sum_{z\in\mathcal{X}} \frac{1}{s} (P_{xz}(s) - I_{xz}) P_{zy}(t) \leqslant \limsup_{s\downarrow 0} \sum_{z\in\mathcal{X}} \frac{1}{s} (P_{xz}(s) - I_{xz}) P_{zy}(t) \leqslant \sum_{z\in F} Q_{xz} P_{zy}(t) - \sum_{z\in F} Q_{xz}.$$

Taking limit over increasing sets F containing x and recognizing that $\sum_{z\in\mathcal{Z}} Q_{xz} = 0$, we get the result. ☐

Theorem 1.13 (forward equation*). *For a homogeneous continuous time Markov chain $X : \Omega \to \mathcal{X}^{\mathbb{R}_+}$ with transition probability kernel $P : \mathbb{R}_+ \to \mathcal{M}(\mathcal{X})^{\mathcal{X}}$ and generator matrix $Q \in \mathbb{R}^{\mathcal{X}\times\mathcal{X}}$, we have for all $t \in \mathbb{R}_+$*

$$\frac{dP(t)}{dt} = P(t)Q.$$

Proof. We fix states $x, y \in \mathcal{X}$ and a finite set $F \subseteq \mathcal{X}$ containing y. In the following, we consider the lim inf and lim sup of (x, y)th term of $\frac{1}{s}(P(t+s) - P(s)) = P(t)\frac{1}{s}(P(s) - I)$. We observe that

$$P_{xy}(t+s) - P_{xy}(t) = \sum_{z\in\mathcal{X}} P_{xz}(t)(P_{zy}(s) - I_{zy}).$$

We observe that $\sum_{z\in F}(P_{zy}(s) - I_{zy}) \leqslant 0$ iff $y \in F$. Together with this fact and that $P_{xz}(t) \in [0,1]$, we obtain

$$\sum_{z\in F} P_{xz}(t)(P_{zy}(s) - I_{zy}) \leqslant \sum_{z\in\mathcal{X}} P_{xz}(t)(P_{zy}(s) - I_{zy}) \leqslant \sum_{z\in F} P_{xz}(t)(P_{zy}(s) - I_{zy}) + \sum_{z\notin F} (P_{zy}(s) - I_{zy}). \tag{3}$$

Taking $\liminf_{s\downarrow 0}$ on the both sides of the first inequality in (3) and $\limsup_{s\downarrow 0}$ on the both sides of the second inequality in (3), exchanging limit and finite sum, we obtain that

$$\sum_{z\in F} P_{xz}(t) Q_{zy} \leqslant \liminf_{s\downarrow 0} \sum_{z\in\mathcal{X}} P_{xz}(t) \frac{1}{s}(P_{zy}(s) - I_{zy}) \leqslant \liminf_{s\downarrow 0} \sum_{z\in\mathcal{X}} P_{xz}(t) \frac{1}{s}(P_{zy}(s) - I_{zy}) \leqslant \sum_{z\in F} P_{xz}(t) Q_{zy} + \sum_{z\notin F} Q_{zy}.$$

Taking limit over increasing sets F containing y, we get the result. We need to justify that $\limsup_{s\downarrow 0} \sum_{z\notin F}(P_{zy}(s) - I_{zy} \leqslant \sum_{z\notin F} Q_{zy}$. ☐

1.3 Strong Markov property

Lemma 1.14. *A continuous time Markov chain $X : \Omega \to \mathcal{X}^{\mathbb{R}_+}$ has the strong Markov property.*

Proof. It follows from the right continuity of the CTMC process X, and the fact that the map $t \mapsto \mathbb{E}[f(X_{t+s}) \mid \sigma(X_t)]$ is right-continuous for any bounded continuous function $f : \mathcal{X} \to \mathbb{R}$. To see the right continuity of the map, we observe that

$$\mathbb{E}[f(X_{t+s}) \mid \sigma(X_t)] = \sum_{x \in \mathcal{X}} \mathbb{1}_{\{X_t = x\}} \sum_{y \in \mathcal{X}} P_{xy}(t, t+s) f(y).$$

Right-continuity of the map follows from the right continuity of the sample paths of process X, right-continuity and boundedness of the kernel function, and boundedness and continuity of f, and bounded convergence theorem. □

Corollary 1.15. *Consider a time homogeneous continuous time Markov chain $X : \Omega \to \mathcal{X}^{\mathbb{R}_+}$. For any stopping time τ adapted to the natural filtration of X, finite $m \in \mathbb{N}$, finite times $0 < t_1 < \cdots < t_m$, any event $H \in \mathcal{F}_\tau$, and states $x_0, x_1, \dots, x_m \in \mathcal{X}$, we have*

$$P(\cap_{k=1}^m \{X_{t_k+\tau} = x_k\} \mid H \cap \{X_\tau = x_0\}) = P_{x_0}(\cap_{k=1}^m \{X_{t_k} = x_k\}).$$

Remark 8. In particular, for a time-homogeneous continuous time Markov chain $X : \Omega \to \mathcal{X}^{\mathbb{R}_+}$, stopping time τ, event $H \in \mathcal{F}_\tau$, states $x, y \in \mathcal{X}$, and time $s \in \mathbb{R}_+$, we have

$$P(\{X_{\tau+s} = y\} \mid \{X_\tau = x\} \cap H) = P_{xy}(s).$$