Lecture 2: Markov Chains

Tianpei Xie

Feb. 1st., 2023

Contents

1 Markov Chain

1.1 Basic Concepts

- ***Markov Chain*** $(X_t)_t$ is a ***probabilistic graphical model*** over a chain graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}_C)$, where each random variable X_t only has exactly one children X_{t+1} and one parent X_{t-1}. Denote the index of variable t as the time. Markov chain $(X_t)_t$ is also a ***stochastic process***.
- By Markov property,

$$P(X_{t+1}|X_t, X_{t-1}, \dots, X_1) = P(X_{t+1}|X_t).$$

  It is seen that the transition probability does not depend on the time t, i.e. Markov chain is **time-invariant**.
- We can see that the joint distribution on $\boldsymbol{X}_{0:t} = [X_0, \dots, X_t]$ can be factorized by transition probabilities

$$P(\boldsymbol{X}_{1:t}) = P(X_0) \prod_{s=1}^{t} P(X_s|X_{s-1})$$

  by Markov property. Denote $\pi_0(i) := P(X_0 = i)$ as the ***initial probability***.
- Define the ***transition kernel*** of Markov Chain as the *time-invariant* ***transition probability***

$$K(x, y) = p(x, y) := P(X_{t+1} = y|X_t = x). \tag{1}$$

  Then the ***m-step transition probability*** is defined as

$$K^m(x, y) = P(X_{t+m} = y|X_t = x). \tag{2}$$

- In *general setting* [Robert and Casella, 1999], $K : \mathcal{X} \times \mathcal{B}(\mathcal{X}) \to \mathbb{R}$ is a function so that $K(x, \cdot)$ is a ***probability measure*** for all $x \in \mathcal{X}$ and $K(\cdot, A)$ is a ***measurable function*** for all $A \in \mathcal{B}(\mathcal{X})$. K can also be considered as a ***functional*** such that

$$K(h(x)) = \int h(y)K(x, dy), \quad h \in L_1(\lambda).$$

  where λ is the dominated measure.
- For $X_t \in \mathcal{X} := \{1, \dots, n\}$ as discrete random variable with $|\mathcal{X}| = n$, we can define the ***transition matrix***

$$\boldsymbol{K} = [K(i, j)]_{n \times n} \tag{3}$$

- We can see that $\boldsymbol{X}_{(t+1):(t+m-1)} := [X_{t+1}, \dots, X_{t+m-1}]$, the m-step transition can be computed using transition kernel

$$
\begin{aligned}
P(X_{t+m}|X_t) &= \sum_{\boldsymbol{X}_{(t+1):(t+m-1)}} P(X_{t+m}|\boldsymbol{X}_{(t+1):(t+m-1)}, X_t) P(\boldsymbol{X}_{(t+1):(t+m-1)}) \\
&= \sum_{\boldsymbol{X}_{(t+1):(t+m-1)}} P(X_{t+m}|X_{t+m-1}) P(\boldsymbol{X}_{(t+1):(t+m-1)}) \\
&= \dots \\
&= \sum_{\boldsymbol{X}_{(t+1):(t+m-1)}} \prod_{i=0}^{m-1} P(X_{t+i+1}|X_{t+i}) && (4) \\
\Rightarrow [K^m(i,j)] &= [K(l,m)]^m = \boldsymbol{K}^m && (5)
\end{aligned}
$$

- We have the ***Chapman-Kolmogorov equation*** [Ross, 2014]:

$$
\begin{aligned}
K^{m+n}(x,y) &= \sum_z K^m(x,z) K^n(z,y), \quad \forall x, y \in \mathcal{X} && (6) \\
\Rightarrow \boldsymbol{K}^{m+n} &= \boldsymbol{K}^m \boldsymbol{K}^n
\end{aligned}
$$

  That is, we split the $(m+n)$-step path from $x \to y$ into all possible combination of a m-step path from $x \to z$ and a n-step path from $z \to y$ for some intermediate state z.

- And the marginal distribution on state X_t can be computed as

$$
\begin{aligned}
P(X_t) &= \sum_{X_{t-1} \in \mathcal{X}} P(X_t|X_{t-1}) P(X_{t-1}) \\
\Rightarrow \boldsymbol{\pi}_t &= \boldsymbol{K} \boldsymbol{\pi}_{t-1} && (7)
\end{aligned}
$$

  where $\boldsymbol{\pi}_t := [P(X_t = i)]$

1.2 Hitting time

- **Definition** Define $T_j = \min\{t \geq 1 : X_t = j\}$ as the time steps for Markov Chain $(X_t)_t$ to *hit* state j for the ***first time***. T_j is called the state j's ***first hitting time***.

  Denote $f_{i,j}$ be the ***probability of ever hitting state*** j (***within finite time***) ***starting from state*** i. That is

$$
f_{i,j} := P(T_j < \infty | X_0 = i) \tag{8}
$$

  Denote $f_{i,j}^{(m)}$ be the *probability of hitting at state* j *at time* m starting from state i

$$
f_{i,j}^{(m)} := P(T_j = m | X_0 = i) \tag{9}
$$

  We can generalize the hitting time for a set of states $T_A := \min\{t \geq 1 : X_t \in A\}$.

- We can the following the ***kernel recursion formula***

$$
\begin{aligned}
K^m(x,y) &= \sum_{n=1}^{m} P(T_y = n | X_0 = x) K^{m-n}(y,y) && (10) \\
&:= \sum_{n=1}^{m} f_{x,y}^{(n)} K^{m-n}(y,y).
\end{aligned}
$$

That is, we *categorize* all possible m-step path from $x \to y$ according to the first time the path visiting y. (This is called the ***First-Step analysis***)

- Similarly, we have the ***hitting time recursion formula***:

$$f_{x,y}^{(m)} := P(T_y = m | X_0 = x) = \sum_{z \neq i} P(T_y = m-1 | X_0 = z)\, K(x,z). \tag{11}$$
$$:= \sum_{z \neq i} f_{z,y}^{(m-1)}\, K(x,z).$$

This formula break down the m-step path from $x \to y$ into two parts: a path from $x \to z$ and a $(m-1)$-step path from intermediate state $z \to y$ (This is also the *First-Step analysis*).

- Define $N(y) := \sum_{t=0}^{\infty} \mathbb{1}\{X_t = y\}$ is the ***total number of times hitting the state*** y.

$$P(N(y) \geq 1 |\, X_0 = x) = P(T_y < \infty | X_0 = x) = f_{x,y} \tag{12}$$
$$P(N(y) \geq m |\, X_0 = x) = P(N(y) \geq 1 |\, X_0 = x) P^{m-1}(N(y) \geq 1 |\, X_0 = y) \tag{13}$$
$$= f_{x,y}\, f_{y,y}^{m-1}$$

Note that in order to visit y at least m times, we need to visit y first time and stating from y recurrently visit y $(m-1)$ times.

The random variable $N(x)|X_0 = x$ follows a ***geometric distribution*** with mean $1/(1 - f_{x,x})$.

$$P(N(x) = m |\, X_0 = x) = (1 - f_{x,x}) f_{x,x}^{m-1} \tag{14}$$

- Define $G(x,y) := \mathbb{E}\left[N(y)|X_0 = x\right]$ as the ***expected number of total visits*** of y starting from x.

$$\begin{aligned} G(x,y) &:= \mathbb{E}\left[N(y)|X_0 = x\right] \\ &= \mathbb{E}\left[\sum_{t=0}^{\infty} \mathbb{1}\{X_t = y\} | X_0 = x\right] \\ &= \sum_{t=0}^{\infty} \mathbb{E}\left[\mathbb{1}\{X_t = y\} | X_0 = x\right] = \sum_{t=0}^{\infty} K^t(x,y) \end{aligned} \tag{15}$$

Note that $\mathbb{E}\left[\sum_{t=0}^{\infty} \mathbb{1}\{X_t = y\} | X_0 = x\right] = \sum_{t=0}^{\infty} \mathbb{E}\left[\mathbb{1}\{X_t = y\} | X_0 = x\right]$ is true since $Z_t := \mathbb{1}\{X_t = y\}$ is non-negative random variable.

Since $N(y)$ is geometric distributed, we can compute $G(x,y)$ via

$$G(x,y) = \frac{f_{x,y}}{1 - f_{y,y}} \tag{16}$$

- Define $G(x,x) = \mathbb{E}\left[N(x)|X_0 = x\right]$ as the *expected number of* ***total returns*** starting from state x.

2 Classification of States

2.1 Equivalence class by communication

- **Definition** For any pair $x, y \in \mathcal{X}$, if there exists $n \in \mathbb{N}_+$ so that $K^n(x, y) > 0$, then the state y is ***accessible*** from state x. This is equivalent to say that the probability of hitting time of y being finite starting from x is above zero, i.e. $\underline{f_{x,y} > 0}$.
- If x is accessible from y, and y is accessible from x, then we say that x and y ***coummunicate***, $x \leftrightarrow y$. It is easy to check that this is an ***equivalence relation***:
    1. $x \leftrightarrow x$;
    2. If $x \leftrightarrow y$, then $y \leftrightarrow x$;
    3. If $x \leftrightarrow z$ and $z \leftrightarrow y$, then $x \leftrightarrow y$
- Thus we can partition the state space $\mathcal{X}$ into several ***equivalence classes*** $\mathcal{X} = \bigcup_k \mathcal{X}^k$ and within each class, all states communicate to each other.

  Equivalently, it means that the kernel $\boldsymbol{K}$ can be *rearranged* into a ***block-diagonal matrix***.
- **Definition** A Markov Chain is ***irreducible*** if it has **only one equivalence class**, i.e. all states in $\mathcal{X}$ communicate to each other.
- Based on hitting time, we can categorize states into two groups:
    - **Definition** A state i is ***recurrent*** if and only if $\underline{f_{i,i} = P(T_i < \infty | X_0 = i) = 1}$, i.e. the Markov Chain will definitely revisit the state i after stating from i.

      Note that it follows from (15) that

      Proposition 2.1 *(Characterization of recurrence via n-**step return probabilities**) A state i is recurrent if and only if $\sum_{t=0}^{\infty} K^t(i, i) = \infty$.*
    - **Definition** A recurrent state i is ***positive recurrent*** if the *expected returning time* $\mathbb{E}\left[T_i | X_0 = i\right] < \infty$; otherwise we say it is ***null recurrent***.
    - **Definition** A state i is called ***transient*** if $f_{i,i} < 1$.
- **Proposition 2.2** *The following conditions are* **equivalent***:*
    1. *state i is recurrent state;*
    2. *The ever returning probability $f_{i,i} = 1$;*
    3. *The probability of total number of visiting is $P(N(i) = \infty | \, X_0 = j) = f_{j,i}$ and $P(N(i) = \infty | \, X_0 = i) = 1$;*
    4. *The expected total number of returning is infinite $G(i, i) = \infty$;*
    5. *The sum of all n-step return probabilities $\sum_{t=0}^{\infty} K^t(i, i) = \infty$.*
- **Proposition 2.3** *If i is recurrent, and $i \to j$, then also $j \to i$.*
- **Proposition 2.4** *If i is positive recurrent, and $i \leftrightarrow j$, then j is also positive recurrent.*

- **Proposition 2.5** *If i is recurrent, and $i \to j$, then j is also recurrent. Therefore, in any **equivalent class**, **either all states are recurrent** or **all are transient**. In particular, if the chain is* **irreducible**, *then either all states are recurrent or all are transient.*

  Based above proposition, we can classify ***each class**, and **an irreducible Markov Chain** as recurrent or transient.*

- **Proposition 2.6** *If a closed subset $S_0 \subset \mathcal{X}$ only has finitely many states, then there must be at least one recurrent state. In particular any finite Markov chain must contain at least one positive recurrent state.*

  Proposition 2.7 *An **irreducible finite state Markov chain** must be **positive recurrent**.*

- **Proposition 2.8** *Any recurrent class is a **closed** subset of states.*

- Let S_T be a set of ***transient** states* and C be a closed set of ***irreducible***, ***recurrent** state*, the ***absorption probability*** is defined as

$$p_C(x) = P(T_C < \infty | X_0 = x), \quad \forall x \in S_T. \tag{17}$$

  It is the probability of hitting recurrent state set starting from a transient state.

- **Theorem 2.9** *Suppose S_T is a set of **transient** states and C is a closed irreducible set of **recurrent** state, then the following system of equations has **unique** solution,*

$$f(x) = \sum_{y \in C} K(x, y) + \sum_{y \in S_T} K(x, y) f(y), \quad \forall\, x \in S_T \tag{18}$$

  and the unique solution is the absorption probability $f(x) = p_C(x)$.

- The recurrence definition ("with *infinite number of visits*") can be generalized as the ***Harris recurrence*** in general theory [Robert and Casella, 1999].

  Definition A set A is ***Harris recurrent*** if $P(N_A = \infty | X_0 = x) = 1$ for all $x \in A$, where $N_A := \sum_{t=0}^{\infty} \mathbb{1}\{X_t \in A\}$. The chain $(X_t)_t$ is ***Harris recurrent*** if there exists a measure p such that $(X_t)_t$ is *p-irreducible* and for every set A with $p(A) > 0$, A is *Harris recurrent*.

2.2 Foster's theorem and Poke's lemma

- **Theorem 2.10** ***(Foster's theorem)***
  Consider an irreducible Markov chain $(X_t)_t$ with state space $\mathcal{X} = \{0, 1, \ldots\}$ and transition matrix $\boldsymbol{K}$ and suppose there exists a function $h : \mathcal{X} \to \mathbb{R}$ such that

  (1) $\inf_{x \in \mathcal{X}} h(x) > -\infty$

  (2) $\sum_{y \in \mathcal{X}} K(x, y)\, h(y) < \infty \quad \forall x \in \mathcal{S}$

  (3) $\sum_{y \in \mathcal{X}} K(x, y)\, h(y) < h(x) - \epsilon \quad \forall x \notin \mathcal{S}$

  *for some finite set $\mathcal{S} \subset \mathcal{X}$ and some $\epsilon > 0$, then the Markov chain $(X_t)_t$ is **positive recurrent**.*

- **Lemma 2.11** ***(Poke's lemma)***
Consider an irreducible Markov chain $(X_t)_t$ with state space $\mathcal{X} = \{0, 1, \ldots\}$ and transition matrix $\boldsymbol{K}$. Assume that for all $x \in \mathcal{X}$ and all $t \geq 0$, $\mathbb{E}[X_{t+1}|X_t = x] < \infty$ and $\lim_{i\to\infty} \sup_{j\geq i} \mathbb{E}[X_{t+1} - X_t|X_t = j] < 0$. Then the Markov chain $(X_t)_t$ is ***positive recurrent****.*

3 Limiting and stationary distribution

3.1 Property of limiting distributions

- **Definition** The probability of states $\{\pi(x), \forall x \in \mathcal{X}\}$ is a ***stationary distribution*** if and only if

$$\pi(y) = \sum_{x\in X} K(x, y)\pi(x), \forall y \in \mathcal{X} \tag{19}$$

$$\boldsymbol{\pi}^T = \boldsymbol{\pi}^T \boldsymbol{K} \tag{20}$$

That is, $\boldsymbol{\pi}$ is the eigenvector of stochastic matrix $\boldsymbol{K}$ corresponding to eigenvalue $\lambda_0 = 1$.

- A stationary distribution is also called an ***invariant distribution*** [Robert and Casella, 1999, Liu, 2001], ***steady-state distribution*** [Ross, 2014] or ***equilibrium distribution*** [Brooks et al., 2011, Ross, 2014]. This is due to its *time invariant property* or *the global balance equation* (23).

- Let the initial distribution be the stationary distribution $P(X_0 = x) = \pi(x)$. Note that

$$\pi_1(y) = P(X_1 = y) = \sum_{x} K(x, y)\pi(x) = \pi(y), \forall y \in \mathcal{X}. \tag{21}$$

In other word, ***the stationary distribution does not change over time***.

In measure theory, the invariant measure π satisfies:

$$\pi(B) = \int_{\mathcal{X}} K(x, B)\pi(dx), \quad \forall B \in \mathcal{B}(\mathcal{X}).$$

- **Proposition 3.1** *Suppose that the* ***limiting distribution*** $\lim_{t\to\infty} P(X_t = y)$ *exists, and*

$$\lim_{t\to\infty} K^t(x, y) = \pi(y), \quad \forall x, y \in \mathcal{X}$$

which is independent of where it starts from, then the Markov Chain has a ***unique stationary distribution*** *and*

$$\lim_{t\to\infty} P(X_t = y) = \pi(y), \quad \forall y \in \mathcal{X} \tag{22}$$

i.e. the limit distribution is stationary distribution.

Note that $P(X_t = y) = \sum_{x\in\mathcal{X}} K^t(x, y)\pi_0(x)$.

- **Proposition 3.2** ***(Global Balance Equation)***
  The stationary distribution $\{\pi(x), \forall x \in \mathcal{X}\}$ satisfies the following ***global balance equation:***

$$\sum_{j\in\mathcal{X}} \pi(i)K(i,j) = \sum_{j\in\mathcal{X}} \pi(j)K(j,i). \tag{23}$$

  This means the total flow out of i (LHS) is equal to the total flow into i (RHS) in steady state.

- **Proposition 3.3** ***(Detailed Balance Equation)***
  For distribution $\{\pi(x), \forall x \in \mathcal{X}\}$, if the following ***detailed balance equation*** *is satisfied*

$$\pi(i)K(i,j) = \pi(j)K(j,i), \quad \forall i,j \in \mathcal{X} \tag{24}$$

  then $\{\pi(x), \forall x \in \mathcal{X}\}$ is a stationary distribution.

- **Definition** Define $\mu_i := \mathbb{E}\left[T_i | X_0 = i\right]$ as the ***expected first return time***, i.e. the number of transition that it takes for Markov chain when starting from state i and returning to that state.

- Let $G^{(n)}(x,y) = \mathbb{E}\left[N^{(n)}(y)|X_0 = x\right]$ where $N^{(n)}(y) = \sum_{t=0}^{n} \mathbb{1}\{X_t = y\}$. $N^{(n)}(y)$ is the total amount of time staying at state y within n transitions. Then

  - **Theorem 3.4** *For* ***transient state*** *y*

$$\begin{aligned} &\lim_{n\to\infty} N^{(n)} < \infty, \quad \textit{(w.p.1)} \\ \Rightarrow &\lim_{n\to\infty} \frac{N^{(n)}}{n} = 0, \quad \textit{(w.p.1)} \\ &\lim_{n\to\infty} G^{(n)}(x,y) = \frac{f_{x,y}}{1 - f_{y,y}} < \infty \\ \Rightarrow &\lim_{n\to\infty} \frac{G^{(n)}(x,y)}{n} = 0, \quad \forall x \in \mathcal{X} \end{aligned}$$

    That is, the frequency of visiting transient state y goes to 0 as $n \to \infty$.

  - **Theorem 3.5** *For* ***recurrent state*** *y*

$$\begin{aligned} \lim_{n\to\infty} \frac{N^{(n)}}{n} &= \frac{\mathbb{1}\{T_y < \infty\}}{\mu_y}, \quad \textit{(w.p.1)} \\ \lim_{n\to\infty} \frac{G^{(n)}(x,y)}{n} &= \frac{f_{x,y}}{\mu_y}, \quad \forall x \in \mathcal{X} \end{aligned}$$

    where $\mu_y := \mathbb{E}\left[T_y | X_0 = y\right]$ is the ***expected first return time*** *of state y. That is, the frequency of visiting* ***positive recurrent*** *state y converge to $\frac{1}{\mu_y}$ as $n \to \infty$; otherwise for* ***null recurrent state*** *y, it converges to zero.*

- **Theorem 3.6** *(Stationary distribution for transient and null recurrent states)*
  Let $\{\pi(x), \forall x \in \mathcal{X}\}$ be stationary distribution. If $x \in \mathcal{X}$ is ***transient*** *or* ***null recurrent*** *state, then*

$$\pi(x) = 0.$$

- **Theorem 3.7** *(**Kac's Theorem**)[Ross, 2014]*
An ***irreducible recurrent*** *Markov Chain has a* ***unique stationary distribution*** $\{\pi(x)\}$*, given*

$$\pi(x) = \frac{1}{\mu_x}, \quad \forall x \in \mathcal{X} \tag{25}$$

where $\mu_x := \mathbb{E}\left[T_x | X_0 = x\right]$ *is the* ***expected first return time*** *of state* x*.*

It implies that as $n \to \infty$, for any state $x \in \mathcal{X}$, the fraction of time that Markov Chain stays at x is unchanged and is the reciprocal of the expected first return time.

3.2 Ergodicity

- Under what condition we have $\forall y \in \mathcal{X}$,

$$\lim_{t\to\infty} P(X_t = y) = \pi(y)? \tag{26}$$

This is the question that ergodicity property tries to answer.

- **Definition** The ***periodicity*** of a state $x \in \mathcal{X}$ is defined as

$$d(x) = \text{g.c.d.}\left\{t \geq 0 : K^t(x,x) > 0\right\} \tag{27}$$

where g.c.d. is the ***greatest common divisor***.

- **Theorem 3.8** *If* $x \leftrightarrow y$ *(i.e.* $f_{x,y} > 0$ *and* $f_{y,x} > 0$*), then* $d(x) = d(y)$*.*

- **Definition** If $d(x) \geq 2$, then state x is ***periodic***. If $d(x) = 1$, then state x is ***aperiodic***

Based on above theorem, the periodicity property is *closed* under the equivalence class C.

- **Definition** A Markov Chain is ***irreducible***, ***positive recurrent*** and ***aperiodic***, then it is called ***ergodic***.

- **Theorem 3.9** *A Markov Chain is* ***irreducible and positive recurrent*** *having stationary distribution* $\boldsymbol{\pi}$*.*

  - *If the Markov Chain is also* ***aperiodic****, then*

$$\lim_{t\to\infty} K^t(x,y) = \pi(y), \quad \forall x, y \in \mathcal{X} \tag{28}$$

  - *If the Markov chain is* ***periodic*** *with period* d*, then there exists* $r \in \mathbb{N}_+$*,* $0 \leq r \leq d-1$ *such that*

$$K^t(x,y) = 0, \quad \forall x, y \in \mathcal{X} \tag{29}$$

unless $t = m\,d + r$ *for some* $m \in \mathbb{N}_+$ *and*

$$\lim_{m\to\infty} K^{m\,d+r}(x,y) = d\,\pi(y), \quad \forall x, y \in \mathcal{X} \tag{30}$$

Note that periodicity only appears on discrete time Markov chain.

Based on the Theorem 3.9 and Proposition 3.1, when a Markov chain is ergodic, its marginal state distribution will converge to the stationary distribution.

3.3 Mean hitting time formula

- **Definition** Let $(X_t)_t$ be a stochastic process and let $\{\mathcal{F}_t, t \geq 0\}$ be an increasing family of σ-field.

  A random variable $T : (\Omega, \mathcal{F}) \to (\mathbb{N}_+ \cup \{+\infty\}, 2^{\mathbb{N}_+ \cup \{+\infty\}})$ is called a ***stopping time*** with respect to $\{\mathcal{F}_t, t \geq 0\}$, if $\forall k \geq 0$, $\mathbb{1}\{T = k\}$ is $\mathcal{F}_k$**-measurable** (i.e. $\overline{\{T = k\} \in \mathcal{F}_k}$ and $\mathcal{F} = \bigcup_{t \geq 0} \mathcal{F}_t$.)

  For Markov chain $(X_t)_t$, the ***first hitting time*** is defined as $T_A^{(1)} = \min\{t > 0 : X_t \in A\}$. Also we can define n-th hitting time as $T_A^{(n)} = \min\left\{t > T_A^{(n-1)} : X_t \in A\right\}$. All of these $\left\{T_A^{(1)}, \ldots, T_A^{(n)}, \ldots\right\}$ are all ***stopping time***.

- **Theorem 3.10** ***(Strong Markov property)*** *[Robert and Casella, 1999]*
  For every initial distribution $\boldsymbol{\pi}$ *and every stopping time* τ *which is almost surely finite,*

$$\mathbb{E}\left[h(X_{\tau+1}, X_{\tau+2}, \ldots) \,|\, x_1 \ldots, x_\tau\right] = \mathbb{E}\left[h(X_1, X_2, \ldots)\right], \tag{31}$$

  provided the expectations exist.

  We can thus condition on a random number of instants while keeping the fundamental properties of a Markov chain. We can proof that for the intervals $\tau_1 = T_x^{(1)}$, $\tau_i := \tau_x^{(i)} - \tau_x^{(i-1)}$, $i = 2, \ldots,$, then $\{\tau_1, \ldots, \tau_n, \ldots\}$ are i.i.d.

- **Theorem 3.11** *Let* $(X_t)_t$ *be a* ***positive recurrent*** *Markov chain with state space* $\mathcal{X}$ *and stationary distribution* $\boldsymbol{\pi}$. *Let* T *be any* ***stopping time*** *of* $(X_t)_t$ *such that for arbitrary* $x \in \mathcal{X}$, $X_T = x$. *Then for all* $y \in \mathcal{X}$,

$$\mathbb{E}_T\left[\sum_{t=0}^{T-1} K^t(x, y) \,|\, x\right] = \mathbb{E}\left[\sum_{t=0}^{T-1} \mathbb{1}\{X_t = y\} \mid X_0 = x\right] = \pi(y)\,\mathbb{E}\left[T | X_0 = x\right]. \tag{32}$$

- **Theorem 3.12** *Let* $i \neq j$ *and* T *be the first time returning* i *after visiting* j, $T = \min\{t > \tau_j, X_t = i | X_0 = i\}$, $\tau_j = \min\{t > 0 : X_t = j\}$ *and* $\tau_i = \min\{t > 0 : X_t = i\}$. *Then*

  (1) $\mathbb{E}\left[T | X_0 = i\right] = \mathbb{E}\left[\tau_i | X_0 = i\right] + \mathbb{E}\left[\tau_j | X_0 = i\right]$

  (2) $\mathbb{E}\left[\sum_{t=0}^{T-1} \mathbb{1}\{X_t = j\} \mid X_0 = i\right] = \mathbb{E}\left[\sum_{t=\tau_j}^{T-1} \mathbb{1}\{X_t = j\} \mid X_0 = i\right] + \mathbb{E}\left[\sum_{t=0}^{\tau_i - 1} \mathbb{1}\{X_t = j\} \mid X_0 = j\right]$

  (3) $\mathbb{E}\left[\sum_{t=0}^{\tau_i - 1} \mathbb{1}\{X_t = j\} \mid X_0 = j\right] = \pi(j)\left(\mathbb{E}\left[\tau_j \,|\, X_0 = i\right] + \mathbb{E}\left[\tau_i \,|\, X_0 = j\right]\right)$

  The "*number of visits to* j *before first returning to* i" is geometric distributed with mean $p := P(\tau_j > \tau_i | X_0 = i)$, thus (3) can be computed as

$$\mathbb{E}\left[\sum_{t=0}^{\tau_i - 1} \mathbb{1}\{X_t = j\} \mid X_0 = j\right] = \frac{1}{P(\tau_j > \tau_i | X_0 = i)}.$$

- **Definition** For finite state, ergodic Markov chain $(X_t)_t$ with stationary distribution $\boldsymbol{\pi}$, define the ***fundamental matrix*** as

$$
\begin{aligned}
\boldsymbol{Z} &:= \left(\boldsymbol{I} - \left(\boldsymbol{K} - \boldsymbol{1}\boldsymbol{\pi}^T\right)\right)^{-1} \\
&= \boldsymbol{I} + \sum_{t \geq 0} \left(\boldsymbol{K}^t - \boldsymbol{1}\boldsymbol{\pi}^T\right)
\end{aligned} \tag{33}
$$

and its (i, j) element is

$$
Z_{i,j} = \sum_{t=0}^{\infty} (K^t(i,j) - \pi_j) \tag{34}
$$

Note that $\boldsymbol{Z} = (\boldsymbol{I} - \boldsymbol{Q})^{-1} = \sum_{t=0}^{\infty} \boldsymbol{Q}^t$, where $\boldsymbol{Q} := \left(\boldsymbol{K} - \boldsymbol{1}\boldsymbol{\pi}^T\right)$.

- **Theorem 3.13** ***(Mean hitting time formula)***
For finite state, ergodic Markov chain $(X_t)_t$ with stationary distribution $\boldsymbol{\pi}$, and $Z_{i,j}$ is defined as (34)*, then*

$$
Z_{i,i} = \pi(i)\, \mathbb{E}\left[\tau_i | X_0 \sim \boldsymbol{\pi}\right], \quad i \in \mathcal{X}, \tag{35}
$$

$$
Z_{j,j} - Z_{i,j} = \pi(j)\, \mathbb{E}\left[\tau_j | X_0 = i\right], \quad i, j \in \mathcal{X}, \tag{36}
$$

where $\tau_i := \min\{t \geq 0 : X_t = i\}$ is the stopping time/first hitting time. Thus

$$
\begin{aligned}
\mathbb{E}\left[\tau_i | X_0 \sim \boldsymbol{\pi}\right] &= \frac{Z_{i,i}}{\pi(i)} \\
\mathbb{E}\left[\tau_j | X_0 = i\right] &= \frac{(Z_{j,j} - Z_{i,j})}{\pi(j)}
\end{aligned}
$$

4 Time-reversible Markov Chain

- **Definition** A Markov chain $(X_t)_t$ is called ***time-reversible***, if it has stationary distribution $\boldsymbol{\pi}$ and the detailed balance equation is satisfied:

$$
\pi(i)K(i,j) = \pi(j)K(j,i), \quad \forall i, j \in \mathcal{X}. \tag{37}
$$

- From this definition, we can see that reversibility implies that the stationary distribution exists, but not *vice versa*.

- The reversed process $(Y_k)_k := (X_{t-k})_k$ is a Markov chain and its transition probability

$$
Q(i,j) = \frac{\pi(j)K(j,i)}{\pi(i)} \tag{38}
$$

Note that $(Y_k)_k$ and $(X_t)_t$ are statistically equivalent since $Q(i,j) = K(i,j)$.

- **Theorem 4.1** *An ergodic Markov chain $(X_t)_t$ for which $K(i,j) = 0$ whenever $K(j,i) = 0$ is* ***time-reversible*** *if and only if starting from any state i, any path back to i has the* ***same probability*** *as its reverse path. That is, for path $i \to i_1 \to i_2 \to \ldots \to i_k \to i$ and its reverse path $i \leftarrow i_1 \leftarrow i_2 \leftarrow \ldots \leftarrow i_k \leftarrow i$*

$$
K(i,i_1)\, K(i_1,i_2)\, \ldots\, K(i_k,i) = K(i,i_k)\, \ldots\, K(i_2,i_1)\, K(i_1,i), \quad \forall i, i_1, \ldots, i_k \in \mathcal{X} \tag{39}
$$

- **Theorem 4.2** ***(Reversal Test)***
  Let $\boldsymbol{K}$ be a stochastic matrix indexed by a countable set $\mathcal{X}$ and let $\boldsymbol{\pi}$ be a probability distribution on $\mathcal{X}$. Let $\boldsymbol{Q}$ be a stochastic matrix indexed by $\mathcal{X}$ such that

$$\pi(i)Q(i,j) = \pi(j)K(j,i), \quad \forall i,j \in \mathcal{X}. \tag{40}$$

  Then $\boldsymbol{\pi}$ is a stationary distribution of $\boldsymbol{K}$

- **Proposition 4.3** *For finite state, ergodic Markov chain $(X_t)_t$ with stationary distribution $\boldsymbol{\pi}$, and $Z_{i,j}$ is defined as (34), then $(X_t)_t$ is time-reversible if and only if*

$$\pi(i)Z_{i,j} = \pi(j)Z_{j,i}, \quad \forall i,j \in \mathcal{X}. \tag{41}$$

  Note that $\pi(i)\mathbb{E}\left[\tau_j | X_0 = i\right] \neq \pi(j)\mathbb{E}\left[\tau_i | X_0 = j\right]$.

- **Theorem 4.4** ***(Cycle-tour property)***
  For states $(i_0, i_1, \ldots, i_m) \subset \mathcal{X}$ of a time-reversible Markov chain,

$$\begin{aligned}
&\mathbb{E}\left[\tau_{i_1} | X_0 = i_0\right] + \mathbb{E}\left[\tau_{i_2} | X_0 = i_1\right] + \ldots + \mathbb{E}\left[\tau_{i_0} | X_0 = i_m\right] \\
&= \mathbb{E}\left[\tau_{i_m} | X_0 = i_0\right] + \mathbb{E}\left[\tau_{i_{m-1}} | X_0 = i_m\right] + \ldots + \mathbb{E}\left[\tau_{i_0} | X_0 = i_1\right]
\end{aligned} \tag{42}$$

5 Ergodic Theorem and Central Limit Theorem

- Consider the empirical mean of samples generated by Markov Chain

$$S_T(h) = \frac{1}{T}\sum_{t=1}^{T} h(X_t). \tag{43}$$

  We are considering the limit behavior of (43).

- **Theorem 5.1** ***(Ergodic Theorem)*** *[Robert and Casella, 1999]*
  If $(X_t)_t$ is Harris recurrent with a σ-finite invariant measure π, then for any $f, g \in L_1(\pi)$ with $\mathbb{E}_\pi\left[g\right] \neq 0$,

$$\lim_{T\to\infty} \frac{S_T(f)}{S_T(g)} = \frac{\mathbb{E}_\pi\left[f\right]}{\mathbb{E}_\pi\left[g\right]} = \frac{\int f(x)d\pi(x)}{\int g(x)d\pi(x)} \tag{44}$$

  It can be shown that if $(X_t)_t$ is ***Harris positive*** with **stationary distribution** π and if $S_T(h)$ *converges* μ_0-almost surely (μ_0 a.s.) to $\mathbb{E}_\pi\left[h\right]$, for an initial distribution μ_0, this convergence occurs for ***every initial distribution*** μ

  Corollary 5.2 *[Liu, 2001]*
  *If a **finite state-space** Markov chain $(X_t)_t$ is irreducible and aperiodic with stationary distribution π, then $S_T(h)$ converges to $\mathbb{E}_\pi\left[h\right]$ **almost surely** for any initial distribution μ.*

- **Theorem 5.3** ***(Central Limit Theorem for discrete atoms)***

If $(X_t)_t$ is Harris positive recurrent with an atom α such that

$$\mathbb{E}\left[T_\alpha^2 | X_0 \in \alpha\right] < \infty, \qquad \mathbb{E}\left[\left(\sum_{t=1}^{T_\alpha} |h(X_t)|\right)^2 \mid X_0 \in \alpha\right] < \infty,$$

$$\textit{and the variance } \sigma_h^2 := \pi(\alpha)\, \mathbb{E}\left[\left(\sum_{t=1}^{T_\alpha} \{h(X_t) - \mathbb{E}_{\boldsymbol{\pi}}[h(X)]\}\right)^2 \mid X_0 \in \alpha\right] > 0,$$

*then the **Central Limit Theorem** applies:*

$$\sqrt{T}\,(S_T(h) - \mathbb{E}_\pi[h]) \overset{\mathcal{L}}{\sim} \mathcal{N}(0, \sigma_h^2) \tag{45}$$

Corollary 5.4 *[Liu, 2001]*
*For **finite state-space**, irreducible and aperiodic Markov chain $(X_t)_t$, the Central Limit Theorem holds, i.e. $\sqrt{T}\,(S_T(h) - \mathbb{E}_\pi[h]) \overset{\mathcal{L}}{\sim} \mathcal{N}(0, \sigma_h^2)$ for any initial distribution μ.*

- **Theorem 5.5** ***(Central Limit Theorem for reversible chains)***
  *If $(X_t)_t$ is **aperiodic, irreducible, and reversible** with stationary distribution $\boldsymbol{\pi}$, the **Central Limit Theorem** applies when*

$$0 < \sigma_h^2 = \mathbb{E}_{\boldsymbol{\pi}}\left[h^2(X_t)\right] + 2\sum_{s=1}^{\infty} \mathbb{E}_{\boldsymbol{\pi}}\left[h(X_t)\, h(X_{t+s})\right] < \infty. \tag{46}$$

References

Steve Brooks, Andrew Gelman, Galin Jones, and Xiao-Li Meng. *Handbook of markov chain monte carlo*. CRC press, 2011.

Jun S Liu. *Monte Carlo strategies in scientific computing*, volume 10. Springer, 2001.

Christian P Robert and George Casella. *Monte Carlo statistical methods*, volume 2. Springer, 1999.

Sheldon M Ross. *Introduction to probability models*. Academic press, 2014.