Lecture 6 Markov Chains *

Tiejun Li

Markov process is one of the most important stochastic processes in application. Roughly speaking, A Markov process is independent of the past, knowing the present state. In this lecture, we only consider the finite state Markov chain. The readers may be refereed to [2] for further information.

1 Markov Chains

Example 1. *(1D Random Walk) Let ξ_i are i.i.d. random variables such that $\xi_i = \pm 1$ with probability $\frac{1}{2}$, and let*

$$X_n = \xi_1 + \xi_2 + \ldots + \xi_n$$

$\{X_n\}$ represents a unconstrained unbiased random walk on $\mathbb{Z}$, the set of integers. Given $X_n = i$, we have

$$\begin{aligned} P\{X_{n+1} = i \pm 1 |\ X_n = i\} &= \frac{1}{2}, \\ P\{X_{n+1} = \textit{anything else} |\ X_n = i\} &= 0. \end{aligned}$$

We see that the distribution of X_{n+1} depends only on the value of X_n.

The result above can be restated as the *Markov property*

$$P\{X_{n+1} = i_{n+1} |\ \{X_m = i_m\}_{m=1}^n\} = P\{X_{n+1} = i_{n+1} |\ X_n = i_n\},$$

and the sequence $\{X_n\}_{n=1}^\infty$ is called a realization of a Markov process.

Example 2 (Ehrenfest's diffusion model)**.** *An urn contains a mixture of red and black balls. At each time $1, 2, \ldots$ a ball is picked at random from the urn and replaced by a ball of the other colour. The total number of balls in the urn is therefore a constant N, say. Let the state X_n of the system at time n be the number of black balls in the urn.*

As will be stated below, the one-step transition matrix can be given as

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ \frac{1}{N} & 0 & \frac{N-1}{N} & 0 & 0 & 0 \\ 0 & \frac{2}{N} & 0 & \frac{N-2}{N} & 0 & 0 \\ 0 & 0 & \ddots & 0 & \ddots & 0 \\ 0 & \ldots & 0 & \frac{N-1}{N} & 0 & \frac{1}{N} \\ 0 & \ldots & 0 & 0 & 1 & 0 \end{pmatrix} \tag{1}$$

*School of Mathematical Sciences, Peking University, Beijing 100871, P.R. China

Example 3. *(Finite state Markov chain) Suppose a Markov chain only takes a finite set of possible values, without loss of generality, we let the state space be* $\{1, 2, \ldots, N\}$*. Define the* **transition probabilities**

$$p_{jk}^{(n)} = P\{X_{n+1} = k | X_n = j\}$$

This uses the Markov property that the distribution of X_{n+1} *depends only on the value of* X_n*.*

Proposition 1. *(Chapman-Kolmogorov equation)*

$$P(X_n = j | X_0 = i) = \sum_k P(X_n = j | X_m = k) P(X_m = k | X_0 = i), \quad 1 \leq m \leq n-1.$$

Definition 1. *(Time-stationary, or time homogeneous) A Markov chain is called stationary if* p_{jk}^n *is independent of* n*. From now on we will discuss only stationary Markov chains and let* $P = (p_{jk})_{j,k=1}^N$*.* P *is called the transition probability matrix(TPM).*

Markov property implies that

$$P\{X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n\} = (\mu_0)_{i_0} p_{i_0 i_1} p_{i_1 i_2} \ldots p_{i_{n-1} i_n}$$

where $(\mu_0)_{i_0}$ is defined by the intial distribution $(\mu_0)_{i_0} = P\{X_0 = i_0\}$.

From this we get

$$\begin{aligned} P\{X_n = i_n | X_0 = i_0\} &= \sum_{i_1, \ldots, i_{n-1}} p_{i_0 i_1} p_{i_1 i_2} \ldots p_{i_{n-1} i_n} \\ &= (P^n)_{i_0 i_n} \end{aligned}$$

The last quantity denotes the (i_0, i_n)-th entry of the matrix P^n.

P is also called a *stochastic matrix*, in the sense that

$$p_{ij} \geq 0, \quad \sum_{j=1}^N p_{ij} = 1.$$

Given the initial distribution of the Markov chain μ_0, the distribution of X_n is then given by

$$\mu_n = \mu_0 P^n$$

Example 4. μ_n *satisfies the recurrence relation* $\mu_n = \mu_{n-1} P$*. This equation can also be rewritten as*

$$(\mu_n)_i = (\mu_{n-1})_i (1 - \sum_{j \neq i} p_{ij}) + \sum_{j \neq i} (\mu_{n-1})_j p_{ji}.$$

The interpretation is clear.

The following two questions are of special interest.

- Is there an invariant distribution? π is called an invariant distribution if

  $$\pi = \pi P$$

  This is equivalent to say that there exists a nonnegative left eigenvector of P with eigenvalue equal to 1. Notice that 1 is always an eigenvalue of P since it always has the right eigenvector $(1, \ldots, 1)^T$.

- When is the invariant distribution unique?

To answer these questions, it is useful to recall some general results on nonnegative matrices.

Definition 2. *(Reducibility) If there exists a permutation matrix Q such that*

$$QPQ^T = \begin{pmatrix} A_1 & B \\ 0 & A_2 \end{pmatrix}$$

then P is called reducible. Otherwise P is called irreducible.

Example 5. *(Graph representation of Markov chains) Any Markov chain can be sketched by their graph representation as in Figure 1. The arrows and real numbers show the transition probability of the Markov*

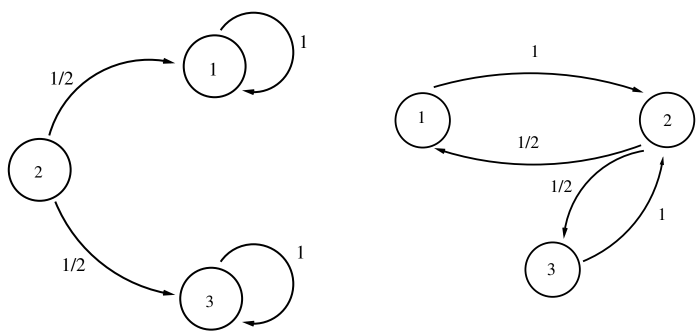

Figure 1: Graph representation of Markov chains. Left panel: chain 1, right panel: chain 2.

chain. The TPM corresponds to left panel is

$$P = \begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 1 \end{bmatrix},$$

It's quite clear that P is a reducible matrix, and it has two invariant distributions $\pi_1 = (1,0,0)$ and $\pi_2 = (0,0,1)$.

The TPM corresponds to the right panel is

$$P = \begin{bmatrix} 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \end{bmatrix}.$$

It's a irreducible matrix, and the only invariant distribution is $\pi = (\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$.

The following theorem is a key answer for invariant distribution of a Markov chain

Theorem 1. *(Perron-Frobenius) Let A be an irreducible nonnegative matrix, and let $\rho(A)$ be its spectral radius: $\rho(A) = \max_\lambda |\lambda|$, where λ is an eigenvalue of A. Then,*

1. *There exists a positive right eigenvector x of A, such that*

$$Ax = \rho(A)x$$

$x = (x_1, \ldots, x_N)^T, x_i > 0.$

2. *$\lambda = \rho(A)$ is an eigenvalue of multiplicity 1.*

Coming back to Markov chains, we obtain as a consequence of the Perron-Frobenius Theorem that

- If P is irreducible, then there exists exactly one invariant distribution.
- If P is reducible, then there are some cases that we can decompose the state space into ergodic components for the Markov chain. On each component there exists a unique in variant distribution. Arbitrary convex combinations of these invariant distributions on each component are invariant distributions for the whole chain. However in this case, the invariant distribution for the whole chain is clearly not unique. One typical example may be as follows:

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 & 0 \\ 0.3 & 0 & 0.4 & 0.3 & 0 \\ 0 & 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0 & 0.5 & 0.5 \end{bmatrix}.$$

  In this case, states $1, 2$ and $4, 5$ form two closed irreducible sub-chains, but P is reducible. There are infinite many invariant distributions. But reducibility itself is not a sufficient condition for the non-uniqueness of the invariant distribution, e.g.

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0.3 & 0.4 & 0.3 & 0 & 0 \\ 0.3 & 0 & 0.4 & 0.3 & 0 \\ 0 & 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0 & 0.5 & 0.5 \end{bmatrix}.$$

  Though the invariant distribution has some zero components which are related to the transience of the states, it is unique.

Irreducibility is equivalent to the property that all nodes on the chain *communicate*, i.e. given any pair (i, j) we have

$$p_{ik_1} p_{k_1 k_2} \cdots p_{k_s j} > 0,$$

for some $(k_1, k_2, \dots, k_s)$ (if there is only transition from $i \to k_1 \to \cdots \to k_j \to j$, we say that j is accessible from i).

The following theorem gives the asymptotic states of a Markov chain

Theorem 2. *Assume that for any pairs (i, j), there exists an s such that $(P^s_{i,j}) > 0$ (irreducible). Then*

1. *There exists a unique invariant distribution π. π is strictly positive.*
2. *For any μ_0,*

$$\pi_n = \mu_0 \bar{P}_n \to \pi \qquad \textit{exponentially fast as } n \to \infty,$$

   where

$$\bar{P}_n = \frac{1}{n} \sum_{j=1}^{n} P^j.$$

Remark 1. *A stronger assumption is "primitive" which says that there exist an natural number s, such that*

$$(P^s)_{ij} > 0, \quad \text{for all } i, j$$

and a stronger convergence theorem $\mu_n = \mu_0 P^n \to \pi$ can be obtained. A critical example is that

$$P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$

which is called a periodic chain. Actually we have primitive $\Leftrightarrow$ irreducible + aperiodic for finite Markov chains.

Theorem 3. *Assume that the Markov chain is primitive. Then for any initial distribution $\boldsymbol{\mu}_0$*

$$\boldsymbol{\mu}_n = \boldsymbol{\mu}_0 \boldsymbol{P}^n \to \boldsymbol{\pi} \quad \text{exponentially fast as } n \to \infty,$$

where $\boldsymbol{\pi}$ is the unique invariant distribution.

Proof. Given two distributions, $\boldsymbol{\mu}_0$ and $\tilde{\boldsymbol{\mu}}_0$, we define the *total variation distance* by

$$d(\boldsymbol{\mu}_0, \tilde{\boldsymbol{\mu}}_0) = \frac{1}{2} \sum_{i \in S} |\mu_{0,i} - \tilde{\mu}_{0,i}|.$$

Since

$$0 = \sum_{i \in S} (\mu_{0,i} - \tilde{\mu}_{0,i}) = \sum_{i \in S} (\mu_{0,i} - \tilde{\mu}_{0,i})^+ - \sum_{i \in S} (\mu_{0,i} - \tilde{\mu}_{0,i})^-,$$

where $a^+ = \max(a, 0)$ and $a^- = \max(-a, 0)$. We also have

$$\begin{aligned} d(\boldsymbol{\mu}_0, \tilde{\boldsymbol{\mu}}_0) &= \frac{1}{2} \sum_{i \in S} (\mu_{0,i} - \tilde{\mu}_{0,i})^+ + \frac{1}{2} \sum_{i \in S} (\mu_{0,i} - \tilde{\mu}_{0,i})^- \\ &= \sum_{i \in S} (\mu_{0,i} - \tilde{\mu}_{0,i})^+ \leq 1. \end{aligned}$$

Let $\boldsymbol{\mu}_s = \boldsymbol{\mu}_0 \boldsymbol{P}^s$, $\tilde{\boldsymbol{\mu}}_s = \tilde{\boldsymbol{\mu}}_0 \boldsymbol{P}^s$ and consider $d(\boldsymbol{\mu}_s, \tilde{\boldsymbol{\mu}}_s)$. We have

$$\begin{aligned} d(\boldsymbol{\mu}_s, \tilde{\boldsymbol{\mu}}_s) &= \sum_{i \in S} \left[\sum_{j \in S} \Big(\mu_{0,j} (\boldsymbol{P}^s)_{ji} - \tilde{\mu}_{0,j} (\boldsymbol{P}^s)_{ji} \Big) \right]^+ \\ &\leq \sum_{j \in S} \Big(\mu_{0,j} - \tilde{\mu}_{0,j} \Big)^+ \sum_{i \in B_+} (\boldsymbol{P}^s)_{ji}, \end{aligned}$$

where B_+ is the subset of indices where $\sum_{j \in S} (\mu_{0,j} - \tilde{\mu}_{0,j}) (\boldsymbol{P}^s)_{ji} > 0$. We note that B_+ cannot contain all the elements of S, otherwise one must have $(\boldsymbol{\mu}_0 \boldsymbol{P}^s)_i > (\tilde{\boldsymbol{\mu}}_0 \boldsymbol{P}^s)_i$ for all i, and

$$\sum_{i \in S} (\boldsymbol{\mu}_0 \boldsymbol{P}^s)_i > \sum_{i \in S} (\tilde{\boldsymbol{\mu}}_0 \boldsymbol{P}^s)_i,$$

which is impossible since both sides sum to 1. Therefore at least one element is missing in B_+. By assumption, there exists an $s > 0$ and $\alpha \in (0, 1)$ such that $(\boldsymbol{P}^s)_{ij} \geq \alpha$ for all pairs (i, j). Hence $\sum_{i \in B_+} (\boldsymbol{P}^s)_{ji} \leq (1 - \alpha) < 1$. Therefore

$$d(\boldsymbol{\mu}_s, \tilde{\boldsymbol{\mu}}_s) \leq d(\boldsymbol{\mu}_0, \tilde{\boldsymbol{\mu}}_0)(1 - \alpha),$$

i.e. the Markov chain is contractive after every s steps. Similarly for any $m \geq 0$

$$d(\boldsymbol{\mu}_n, \boldsymbol{\mu}_{n+m}) \leq d(\boldsymbol{\mu}_{n-sk}, \boldsymbol{\mu}_{n+m-sk})(1-\alpha)^k \leq (1-\alpha)^k,$$

where k is the largest integer such that $n - sk \geq 0$. If n is sufficiently large the right hand side can be made arbitrarily small. Therefore the sequence $\{\boldsymbol{\mu}_n\}_{n=0}^{\infty}$ is a Cauchy sequence. Hence it has to converge to a limit π, which satisfies

$$\boldsymbol{\pi} = \lim_{n\to\infty} \boldsymbol{\mu}_0 \boldsymbol{P}^{n+1} = \lim_{n\to\infty} (\boldsymbol{\mu}_0 \boldsymbol{P}^n)\boldsymbol{P} = \boldsymbol{\pi}\boldsymbol{P}.$$

Such a $\boldsymbol{\pi}$ satisfying such a property is also unique. For if there were two such distributions, $\boldsymbol{\pi}^{(1)}$ and $\boldsymbol{\pi}^{(2)}$, then $d(\boldsymbol{\pi}^{(1)}, \boldsymbol{\pi}^{(2)}) = d(\boldsymbol{\pi}^{(1)}\boldsymbol{P}^s, \boldsymbol{\pi}^{(2)}\boldsymbol{P}^s) < d(\boldsymbol{\pi}^{(1)}, \boldsymbol{\pi}^{(2)})$. This implies $d(\boldsymbol{\pi}^{(1)}, \boldsymbol{\pi}^{(2)}) = 0$, i.e $\boldsymbol{\pi}^{(1)} = \boldsymbol{\pi}^{(2)}$. □

Remark 2. *We do not discuss the convergence speed here. But in fact it is exponential, which depends on the spectral gap of the transition probability matrix P. The readers may be referred to [3, 4].*

Theorem 4 (Ergodic theorem)**.** *let X_n be an irreducible, positive recurrent Markov chain with invariant distribution $\pi(x)$, and f be a bounded function, then*

$$\frac{1}{N}\sum_{n=1}^{N} f(X_n) \to \langle f \rangle_\pi, \quad a.s.$$

1.1 Time Reversal

Theorem 5. *Assume that the Markov chain $\{X_n\}_{n\geq 0}$ admits a unique invariant distribution $\boldsymbol{\pi}$ and is also initially distributed according to $\boldsymbol{\pi}$. Denote by $\boldsymbol{P}$ its transition probability matrix. Define a new Markov chain $\{Y_n\}_{0\leq n\leq N}$ by $Y_n = X_{N-n}$ where $N \in \mathbb{N}$ is fixed. Then $\{Y_n\}_{0\leq n\leq N}$ is also an Markov chain with invariant distribution $\boldsymbol{\pi}$. Its transition probability matrix $\hat{\boldsymbol{P}}$ is given by*

$$\hat{p}_{ij} = \frac{\pi_j}{\pi_i} p_{ji}. \tag{2}$$

Proof. It is straightforward to check that $\hat{\boldsymbol{P}}$ is a stochastic matrix with an invariant distribution $\boldsymbol{\pi}$. To prove that $\{Y_n\}$ is Markov with transition probability matrix $\hat{\boldsymbol{P}}$, it is enough to observe that

$$\begin{aligned} \mathbb{P}(Y_0 = i_0, Y_1 = i_1, \ldots, Y_N = i_N) &= \mathbb{P}(X_N = i_0, X_{N-1} = i_1, \ldots, X_0 = i_N) \\ &= \pi_{i_N} p_{i_N i_{N-1}} \cdots p_{i_1 i_0} = \pi_{i_0} \hat{p}_{i_0 i_1} \cdots \hat{p}_{i_{N-1} i_N} \end{aligned}$$

for any $i_0, i_1, \ldots, i_N$. □

A particularly important class of Markov chains are those that satisfy the condition of *detailed balance*

$$\pi_i p_{ij} = \pi_j p_{ji} \tag{3}$$

In this case, we have $\hat{p}_{ij} = p_{ij}$. We call the chain *reversible*. The reversible chain can be equipped with variational structure and has nice spectral properties. Define the matrix

$$\boldsymbol{L} = \boldsymbol{P} - \boldsymbol{I}$$

and correspondingly its action on any function f

$$(\boldsymbol{L}f)(i) = \sum_{j \in S} p_{ij}(f(j) - f(i)).$$

Let $L^2_{\boldsymbol{\pi}}$ be the space of square summable functions f endowed with the $\boldsymbol{\pi}$-weighted scalar product

$$(f, g)_{\boldsymbol{\pi}} = \sum_{i \in S} \pi_i f(i) g(i). \tag{4}$$

Denote the *Dirichlet form* or energy of a function f by

$$D(f) = \sum_{i,j \in S} \pi_i p_{ij} (f(j) - f(i))^2.$$

One can show that $D(f) = (f, -\boldsymbol{L}f)_{\boldsymbol{\pi}}$. These formulations are particularly useful in potential theory for Markov chains.

1.2 Hitting time distribution

Example 6. *(Hitting time distribution of a Markov chain) Consider TPM of a 4-state Markov chain(1,2,3,4):*

$$P = \begin{bmatrix} \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}.$$

Define the first hitting time $n_* = \inf\{n |\ X_n = 3 \text{ or } 4\}$ *and the hitting time probability* $q(m) = \text{Prob}\{n_* = m\}$*, an interesting question is to ask how to obtain* $q(m)$*. The idea is to modify the chain to a 3-state chain*

$$\tilde{P} = \begin{bmatrix} \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{3} & 0 & \frac{2}{3} \\ 0 & 0 & 1 \end{bmatrix}.$$

then

$$1 - (\mu_n)_3 = \sum_{m=n+1}^{\infty} q(m),$$

hence

$$q(n) = (\mu_n)_3 - (\mu_{n-1})_3 = \mu_0 \cdot (\tilde{P}^n - \tilde{P}^{n-1}) \cdot \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix},$$

2 Continuous time Markov chains

2.1 Poisson Process

Definition 3. *(Poisson Process) Let* $X(t)$ *be the number of calls received up to time* t*, and assume the follows:*

1. $X(0) = 0$*;*

2. *$X(t)$ has independent increments, i.e. for any $0 \le t_1 < t_2 < \cdots < t_n$,*

$$X(t_2) - X(t_1), X(t_3) - X(t_2), \ldots, X(t_n) - X(t_{n-1})$$

are independent;

3. *for any $t \ge 0, s \ge 0$, we have the distribution of the increment $X(t+s) - X(t)$ is independent of s(time-homogeneous);*

4. *for any $t \ge 0, h > 0$, we have*

$$\begin{aligned} P\{X(t+h) = X(t) + 1|\ X(t)\} &= \lambda h + o(h), \\ P\{X(t+h) = X(t)|X(t)\} &= 1 - \lambda h + o(h), \\ P\{X(t+h) \ge X(t) + 2\} &= o(h), \end{aligned}$$

where λ is called the rate.

Then $X(t)$ is called a Poisson process.

Let $p_m(t) = P\{X(t) = m\}$, then

$$p_0(t+h) = p_0(t)p_0(h) = p_0(t)(1 - \lambda h) + o(h).$$

This gives

$$\frac{p_0(t+h) - p_0(t)}{h} = -\lambda p_0(t) + o(1).$$

As $h \to 0$, we obtain

$$\frac{dp_0(t)}{dt} = -\lambda p_0(t), \quad p_0(0) = 1.$$

The solution is given by

$$p_0(t) = e^{-\lambda t}.$$

For $m > 0$, we have

$$p_m(t+h) = p_m(t)p_0(h) + p_{m-1}(t)p_1(h) + \sum_{i=2}^{m} p_{m-i}(t)p_i(h).$$

From the definition of Poisson process, we get

$$p_m(t+h) = p_m(t)(1 - \lambda h) + p_{m-1}(t)\lambda h + o(h).$$

Taking the limit as $h \to 0$, we get

$$\frac{dp_m(t)}{dt} = -\lambda p_m(t) + \lambda p_{m-1}(t)$$

Using the fact $p_m(0) = 0 (m > 0)$, we get

$$p_m(t) = \frac{(\lambda t)^m}{m!} e^{-\lambda t}$$

by induction method. This means that for any fixed t, the distribution of $X(t)$ is Poisson with parameter λt.

The waiting times can be obtained in the following way. Define

$$\mu_t = P\{\text{Waiting time} \geq t\},$$

then $\mu_0 = 1$, and it obeys $\mu_t - \mu_{t+h} = \mu_t \lambda h + o(h)$, thus $\mu_t' = -\lambda \mu_t$, we get

$$\mu_t = e^{-\lambda t}.$$

i.e. The waiting times are i.i.d. exponentially distributed with rate λ.

2.2 Q-Process

Now let us turn to general continuous time Markov chains. We will restrict only on finite state space case in this text. We define

$$p_{ij}(t) = \text{Prob}\{X(t+s) = j | X(s) = i\}.$$

Here we also assumed the stationarity of the Markov chain, i.e. the right hand side is independent of s. By definition we have

$$p_{ij}(t) \geq 0, \qquad \sum_{j=1}^{N} p_{ij}(t) = 1.$$

In addition we require that

$$p_{ii}(h) = 1 - \lambda_i h + o(h), \quad \lambda_i > 0, \tag{5}$$

$$p_{ij}(h) = \lambda_{ij} h + o(h), \quad j \neq i. \tag{6}$$

(5) is a statement about the regularity in time of the Markov chain; together with the obvious constraint that $p_{jj}(0) = 1$. (6) states that if the process is in state j at time t and a change occurs between t and $t + h$, the process must have jumped to some state $i \neq j$; λ_{ij} is the rate of switching from state i to state j.

From the non-negativity and normalization condition of the probability, we have

$$\lambda_{ij} \geq 0, \qquad \sum_{j=1, j\neq i}^{N} \lambda_{ij} = \lambda_i. \tag{7}$$

The Markov property of the process requires the Chapman-Kolmogorov equation

$$p_{ij}(t+s) = \sum_{k=1}^{N} p_{ik}(t) p_{kj}(s). \tag{8}$$

Using matrix notation $P(t) = (p_{ij}(t))$, we can express the Chapman-Kolmogorov relation as

$$P(t+s) = P(t)P(s) = P(s)P(t).$$

Similarly, if we define

$$Q = \lim_{h \to 0+} h^{-1}(P(h) - I), \tag{9}$$

and denote $Q = (q_{ij})$, (5), (6) and (7) can be stated as

$$q_{ii} = -\lambda_i, \qquad q_{ij} = \lambda_{ij} \quad (i \neq j), \qquad \sum_{j=1}^{N} q_{ij} = 0.$$

Q is called the *generator* of the Markov chain.

Since

$$\frac{P(t+h) - P(t)}{h} = \frac{P(h) - I}{h} P(t)$$

as $s \to 0+$, we get

$$\frac{dP(t)}{dt} = QP(t) = P(t)Q \tag{10}$$

The solution of this equation is given by

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

since $P(0) = I$.

Next we discuss how the distribution of the Markov chain evolves in time. Let $\nu(t)$ be the distribution of $X(t)$. Then

$$\begin{aligned} \nu_j(t + dt) &= \sum_{i \neq j} \nu_i(t) p_{ij}(dt) + \nu_j(t) p_{jj}(dt) \\ &= \sum_{i \neq j} \nu_i(t) q_{ij} dt + \nu_j(t)(1 + q_{jj} dt) + o(dt) \end{aligned}$$

for infinitesimal dt. This gives

$$\frac{d\nu(t)}{dt} = \nu(t) Q, \tag{11}$$

which is called the forward Kolmogorov equation for the distribution. Its solution can be given as

$$\nu_j(t) = \sum_{i=1}^{N} \nu_i(0) p_{ij}(t),$$

or, in matrix notation,

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

Similar as the Poisson process, we can consider the waiting time distribution for each state j,

$$\mu_j(t) = \text{Prob}\{\tau \geq t | X(0) = j\}.$$

The same procedure as previous section leads to

$$\frac{d\mu_j(t)}{dt} = q_{jj} \mu_j(t), \qquad \mu_j(0) = 1.$$

Thus the waiting time at state j is exponentially distributed with rate $-q_{jj} = \sum_{k \neq j} q_{jk}$. From the memoryless property of exponential distribution, the waiting time can be counted from any starting point.

It is interesting to investigate the probability

$$\begin{aligned} p(\theta, j | 0, i) d\theta := \text{Prob}\{&\text{The jump time } \tau \text{ is in } [\theta, \theta + d\theta) \\ &\text{and } X(\tau) = j \text{ given } X(0) = i\}. \end{aligned}$$

We have

$$\begin{aligned}p(\theta, j|0, i)d\theta =&\text{Prob}\{\text{No jump occurs in } [0,\theta) \text{ given } X(0)=i\}\\ &\times \text{Prob}\{\text{One jump occurs from } i \text{ to } j \text{ in } [\theta, \theta+d\theta)\}\\ =&\mu_i(\theta)q_{ij}d\theta = \exp(q_{ii}\theta)q_{ij}d\theta. \end{aligned} \tag{12}$$

Thus we obtain the marginal probability

$$\text{Prob}(X(\tau)=j|X(0)=i) = p(j|0,i) = -\frac{q_{ij}}{q_{ii}} = \frac{q_{ij}}{\sum_{j\neq i} q_{ij}}$$

where τ is the waiting time. These results are particularly useful for the numerical simulation of the trajectories of the Q-process.

Define the *jump times* of $(X_t)_{t\geq 0}$

$$J_0 = 0, \quad J_{n+1} = \inf\{t : t \geq J_n, X_t \neq X_{J_n}\}, \quad n \in \mathbb{N}$$

where we take the convention $\inf \emptyset = \infty$, and *holding times*

$$H_n = \begin{cases} J_n - J_{n-1}, & \text{if } J_{n-1} < \infty,\\ \infty, & \text{otherwise.}\end{cases}$$

for $n = 1, 2, \ldots$. We define $X_\infty = X_{J_n}$ if $J_{n+1} = \infty$. Define the *jump chain* induced by X_t

$$Y_n = X_{J_n}, \quad n \in \mathbb{N}.$$

From Strong Markov property and the derivation of $p(\theta, j|0, i)$, we know that the holding times $H_1, H_2, \ldots$ are independent exponential random variables with parameters $q_{Y_0}, q_{Y_1}, \ldots$, respectively, and the jump chain Y_n is a Markov chain with $\tilde{\boldsymbol{Q}}$ as the transition probability matrix, where $\tilde{\boldsymbol{Q}} = (\tilde{q}_{ij})$ defined as

$$\tilde{q}_{ij} = \begin{cases} q_{ij}/q_i, & \text{if } i \neq j \text{ and } q_i > 0,\\ 0, & \text{if } i \neq j \text{ and } q_i = 0,\end{cases} \tag{13}$$

$$\tilde{q}_{ii} = \begin{cases} 0, & \text{if } q_i > 0,\\ 1, & \text{if } q_i = 0.\end{cases} \tag{14}$$

It is called the *jump matrix*, and the corresponding Markov chain is called the *embedded chain* or *jump chain* of the original Q-process.

It is natural to consider the invariant distribution for the Q-processes as in the discrete time Markov chains. From the forward Kolmogorov equation (11), the invariant distribution must satisfy

$$\pi Q = 0, \quad \pi \cdot \mathbf{1}^T = 1.$$

But to ensure the convergence $\nu(t) \to \pi$, we need the following theorem on the finite state space.

Theorem 6 (Convergence to equilibrium)**.** *Suppose the matrix Q is irreducible with invariant distribution π, then for all states i, j we have*

$$p_{ij}(t) \to \pi_j \text{ as } t \to \infty.$$

Note that we do NOT need the primitive condition since in the continuous time case if $q_{ij} > 0$ we have

$$p_{ij}(t) \geq \mathbb{P}_i(J_1 \leq t, Y_1 = j, H_2 > t) = \int_0^t e^{-q_i u} q_{ij} du \cdot e^{-q_j t} = \frac{q_{ij}}{q_i}(1 - e^{-q_i t})e^{-q_j t} > 0.$$

Similarly we also have the ergodic theorem

Theorem 7 (Ergodic theorem)**.** *Suppose the matrix Q is irreducible with invariant distribution π, then for any bounded function f we have*

$$\frac{1}{t}\int_0^t f(X(s))ds \to \langle f \rangle_\pi, \qquad a.s.$$

We should remark that the irreducibility condition is not enough to establish the above ergodic theorems in the countable state space case. We need the so-called positive recurrent condition in both theorems.

3 Homeworks

- HW1. Discuss the invariant distribution of the Ehrenfest's model.
- HW2. Rederive the distribution of Poisson process through characteristic function method.
- HW3. Let f be a function defined on the state space, and let

  $$h_i(t) = \mathbb{E}^i f(X(t))$$

  Derive an equation for h(t).
- HW4. Consider the following binomial process: we repeatedly throw an unfair coin with parameter p (say, the proability that the HEAD appears) with time unit τ. If the HEAD appears, we denote it as a jump. Then we let $p, \tau \to 0$ and consider the limiting process. In which regime you can intuitively get the Poisson process with parameter λ?
- HW5. For the Poisson process, if the condition 3 is removed, and the rate λ depends on t. That is, λ is replaced with $\lambda(t)$ in conidtion 4, then what about $p_m(t)$ and the waiting time distribution μ_s conditioned at the current time t?

References

[1] K.L. Chung, A Course in Probability Theory, Academic Press, New York, 1974.

[2] R. Durret, Probability: theory and examples (3rd edition), Thomson Learning, 2005.

[3] C.P. Robert and G. Casella, Monte Carlo statistical methods, Second edition, Springer-Verlag, 2004.

[4] G. Winkler, Image analysis, random fields and dynamic Monte Carlo methods, Springer-Verlag, Berlin and Heidelberg, 1995.