Lecture 4: Discrete-Time Markov Chain – Part 2

Instructor: Yen-Chi Chen

These notes are partially based on those of Mathias Drton.

4.1 Limiting behavior

We have learned the conditions for the existence of a stationary distribution of a Markov chain. The stationary distribution characterizes the behavior of a Markov chain in a steady state, which will be something we expect when running the chain for a long time.

Now consider an interesting case. If we obtain a sequence of RVs from a Markov chain and try to calculate some statistics based on these RVs, what will we get? This is the case when we are running the famous MCMC (Markov Chain Monte Carlo) approach to evaluate some quantity from a distribution. Because the stationary distribution characterizes the *steady behavior* of a Markov chain, we would expect the computed statistics to be related to the stationary distribution. To understand how stationary distribution and the statistics are related, we begin a journey on investigating the limiting behavior of a Markov chain. To simplify the problem, we consider the most common statistic – the sample average. We will show that when averaging the output from a Markov chain, the average will converges almost surely to the average of the stationary distribution! This is also known as the *Ergodic theorem.*

Before we proceed, we first introduce a very useful lemma.

Lemma 4.1 (Regenerative Cycles) *Let* 0 *be a recurrent state of the homogeneous Markov chain* $\{X_n\}$*, and let* $\tau_0 = 0, \tau_1, \tau_2, \ldots$ *be the successive times of return to* 0*. Define* $Z_k = \{X_{\tau_k}, X_{\tau_k+1}, \ldots, X_{\tau_{k+1}-1}\}$*.*

Then the trajectories $Z_1, Z_2, \cdots$ *are IID. In particular, the times between returns* $\tau_1, \tau_2 - \tau_1, \tau_3 - \tau_2, \ldots$ *are i.i.d.*

Proof: Follows from the strong Markov property. ■

The Regenerative Cycles lemma shows that for a Markov chain, each time when we return to the initial state, the chain behaves 'as if' we are starting over again.

The Regenerative Cycles lemma gives us some hint on why the sample average of a Markov chain converges – the chunk of the chain within each return time behaves like independent chain so when averaging over the entire chain, we can view the average as averaging across several independent components.

To make our argument more concrete, let $X_1, \cdots, X_N$ be a Markov chain and we consider the average

$$\frac{1}{N}\sum_{i=1}^{N} f(X_i), \tag{4.1}$$

where f is some function. Note that if $X_1, \cdots, X_N$ are IID and $\mathbb{E}|f(X_1)| < \infty$, the strong law of large number states that

$$\frac{1}{N}\sum_{i=1}^{N} f(X_i) \overset{a.s.}{\to} \mathbb{E}(f(X_1)).$$

We would like to know if the same convergence occurs for a Markov chain.

Proposition 4.2 *Let $\{X_n\}$ be an irreducible, recurrent, and homogeneous Markov chain and* $\mathbf{y}$ *be an invariant measure of $\{X_n\}$ with*

$$y_i = \mathbb{E}_0\left[\sum_{n=1}^{\infty} I(X_n = i)I(n \leq T_0)\right].$$

Let $\nu(n) = \sum_{k=1}^{n} I(X_k = 0)$. If $f : S \to \mathbb{R}$ with $\sum_{i\in S} |f(i)|y_i < \infty$, then

$$\frac{1}{\nu(N)}\sum_{j=1}^{N} f(X_j) \overset{a.s.}{\to} \sum_{i\in S} f(i)y_i \text{ when } N \to \infty.$$

Note that this is not the usual average we calculate, since we are dividing by the number of returns to state 0 in N steps, which explains the y_i (rather than π_i) on the RHS.

Proof: Let τ_j be the j-th time that the Markov chain returns to 0. Note that $\tau_1 = T_0$ is the return time and $\tau_0 = 0$ by definition.

We consider the time between $\tau_j + 1, \cdots, \tau_{j+1}$. Define

$$U_j = \sum_{n=\tau_{j-1}+1}^{\tau_j} f(X_n).$$

By the Regenerative Cycles lemma, $U_1, U_2, \cdots$ are IID.

To see the convergence, we first study the expectation:

$$\begin{aligned}
\mathbb{E}(U_1) &= \mathbb{E}_0\left(\sum_{n=\tau_0+1}^{\tau_1} f(X_n)\right) \\
&= \mathbb{E}_0\left(\sum_{n=1}^{T_0} f(X_n)\right) \\
&= \mathbb{E}_0\left(\sum_{n=1}^{T_0}\sum_{i\in S} f(i)I(X_n = i)\right) \\
&= \sum_{i\in S} f(i)\mathbb{E}_0\left(\sum_{n=1}^{T_0} I(X_n = i)\right) \\
&= \sum_{i\in S} f(i)\mathbb{E}_0\left(\sum_{n=1}^{\infty} I(X_n = i)I(n \leq T_0)\right) \\
&= \sum_{i\in S} f(i)y_i.
\end{aligned}$$

By the strong law of large number and the fact that $U_1, U_2, \cdots$, are IID,

$$\frac{1}{n}\sum_{j=1}^{n} U_j \overset{a.s.}{\to} \sum_{i\in S} f(i)y_i.$$

Using the fact that $U_j = \sum_{n=\tau_{j-1}+1}^{\tau_j} f(X_n)$, we conclude that

$$\frac{1}{n}\sum_{j=1}^{n} U_j = \frac{1}{n}\sum_{i=1}^{\tau_n} f(X_i) \overset{a.s.}{\to} \sum_{i\in S} f(i)y_i. \tag{4.2}$$

This is still not quiet yet what we want because we want the denominator to be $\nu(N)$ and the summation upper bound to be N.

Because $\nu(N) = \sum_{i=1}^{N} I(X_i = 0)$ is the number of visits of state 0, $\tau_{\nu(N)} \leq N \leq \tau_{\nu(N)+1}$ and thus,

$$\sum_{i=1}^{\tau_{\nu(N)}} f(X_i) \leq \sum_{i=1}^{N} f(X_i) \leq \sum_{i=1}^{\tau_{\nu(N)+1}} f(X_i)$$

and

$$\frac{1}{\nu(N)}\sum_{i=1}^{\tau_{\nu(N)}} f(X_i) \leq \frac{1}{\nu(N)}\sum_{i=1}^{N} f(X_i) \leq \frac{1}{\nu(N)}\sum_{i=1}^{\tau_{\nu(N)+1}} f(X_i).$$

To finish the proof, we need to show that the middle one converges almost surely to $\sum_{i\in S} f(i)y_i$. We prove this by showing that the left and the right terms (lower and upper bound) converge almost surely. Note that the fact that the chain is recurrent ($P_0(T_0 < \infty) = 1$) implies that $\nu(N) \to \infty$ almost surely. For the lower bound, by identifying $n = \nu(N)$, equation (4.2) implies

$$\frac{1}{\nu(N)}\sum_{i=1}^{\tau_{\nu(N)}} f(X_i) \overset{a.s.}{\to} \sum_{i\in S} f(i)y_i$$

so we only need to work on the upper bound.

For the upper bound, note that the Markov chain is recurrent so $P_0(T_0 < \infty) = 1$ so

$$\frac{1}{\nu(N)}\sum_{i=1}^{T_0} f(X_i) \overset{a.s.}{\to} 0.$$

Thus,

$$\begin{aligned}
\frac{1}{\nu(N)}\sum_{i=1}^{\tau_{\nu(N)+1}} f(X_i) &= \frac{1}{\nu(N)}\sum_{i=1}^{\tau_{\nu(N)}} f(X_i) + \frac{1}{\nu(N)}\sum_{i=\tau_{\nu(N)}+1}^{\tau_{\nu(N)+1}} f(X_i) \\
&= \frac{1}{\nu(N)}\sum_{i=1}^{\tau_{\nu(N)}} f(X_i) + \frac{1}{\nu(N)}\sum_{i=1}^{T_0} f(X_i) \\
&\overset{a.s.}{\to} \sum_{i\in S} f(i)y_i + 0,
\end{aligned}$$

which proves that the upper bound also converges to the same limit so the result follows.

■

With the above proposition, we are able to derive the Ergodic theorem of a Markov chain.

Theorem 4.3 (Ergodic Theorem) *Let $\{X_n\}$ be an irreducible, homogeneous, and positive recurrent Markov chain on state-space S with stationary distribution π. Let $f : S \to \mathbb{R}$ such that $\sum_{i\in S} |f(i)|\pi_i < \infty$.*

Then, for any initial distribution,

$$\frac{1}{N}\sum_{j=1}^{N} f(X_j) \overset{a.s.}{\to} \sum_{i\in S} f(i)\pi_i$$

as $N \to \infty$*.*

Proof: Applying Proposition 4.2 with $f(x) = 1$, we obtain

$$\frac{1}{\nu(N)}\sum_{j=1}^{N} f(X_j) = \frac{N}{\nu(N)} \overset{a.s.}{\to} \sum_{i\in S} y_i.$$

Thus,

$$\begin{aligned}\frac{1}{N}\sum_{j=1}^{N} f(X_j) &= \frac{\nu(N)}{N}\frac{1}{\nu(N)}\sum_{j=1}^{N} f(X_j)\\ &\overset{a.s.}{\to} \frac{1}{\sum_{\ell\in S} y_\ell}\sum_{i\in S} f(i)y_i\\ &= \sum_{i\in S} f(i)\pi_i\end{aligned}$$

because positive recurrence implies $\pi_i = \frac{y_i}{\sum_{\ell\in S} y_\ell}$.

■

Ergodic theorem is also called the strong law of large number of Markov chain, which shows that the empirical average converges to the average of the stationary distribution.

4.1.1 Convergence of the entire distribution

The Ergodic theorem is very powerful – it tells us that the empirical average of the output from a Markov chain converges to the 'population' average that the population is described by the stationary distribution.

However, convergence of the average statistic is not the only quantity that the Markov chain can offer us. Under suitable condition, the Markov chain behaves like RVs from the stationary distribution regardless of the initial state.

Let X and Y be two random variables on the state space S with probability mass functions p_X and p_Y. We define the *total variation norm* between the distributions of X and Y as

$$d_{TV}(p_X, p_Y) = \frac{1}{2}\sum_{i\in S} |p_X(i) - p_Y(i)|.$$

Theorem 4.4 (Basic Limit Theorem) *Let* $\{X_n\}$ *be irreducible, positive recurrent and aperiodic homogeneous Markov chain with transition probability matrix* $\mathbf{P}$*. Then, for any initial distributions* μ *and* ν

$$\lim_{n\to\infty} d_{TV}(\mu^T\mathbf{P}^n, \nu^T\mathbf{P}^n) = 0.$$

In particular, if $\mu(k) = I(k = i)$ *for some fixed* i *and* $\nu = \pi$ *is the stationary distribution, then*

$$\lim_{n\to\infty}\sum_{j\in S} |p_{ij}^{(n)} - \pi_j| = 0.$$

Proof: See Brémaud (1999, p. 130). ■

The Basic Limit Theorem states that no matter what the initial state is, or how we randomize our starting point, in the long run the Markov chain will behave like random points from the stationary distribution.

With the Basic Limit theorem, one may be wondering if this implies that the t.p.m. $\mathbf{P}^n$ will have some meaningful limit when $n \to \infty$. To investigate this, we first introduce the following proposition.

Proposition 4.5 *Let* $\mathbf{P}$ *denote the t.p.m. of a finite Markov chain with state space* $S = \{1, ..., s\}$. *Suppose that the eigenvalues* $\lambda_1, ..., \lambda_s$ *of* $\mathbf{P}$ *are distinct, with corresponding right (column) eigenvectors* $\mu_1^T, ..., \mu_s^T$*; that is*

$$\mathbf{P}\mu_k^T = \lambda_k \mu_k^T, \qquad k = 1, ..., s.$$

Then there exists $s \times s$ *matrices* $A_1, ..., A_s$ *such that*

$$\mathbf{P}^n = \sum_{k=1}^{s} \lambda_k^n A_k, \qquad n = 0, 1, ...$$

Proof: Define $s \times s$ matrices,

$$U = [\mu_1 | \mu_2 | ... | \mu_s], \qquad \Lambda = \text{diag}[\lambda_1, ..., \lambda_s],$$

so that

$$\mathbf{P}\mu_k = \lambda_k \mu_k \quad \Rightarrow \quad \mathbf{P}U = U\Lambda \quad \Rightarrow \quad \mathbf{P} = U\Lambda U^{-1},$$

since distinct λ_k's ensure that the μ_k's are linearly independent and hence that U is non-singular.

Therefore,

$$\mathbf{P}^n = (U\Lambda U^{-1}) \times (U\Lambda U^{-1}) \times ... \times (U\Lambda U^{-1}) \quad \Rightarrow \quad \mathbf{P}^n = U\Lambda^n U^{-1},$$

for $n = 0, 1, 2, ...$ where $\Lambda^n = \text{diag}[\lambda_1^n, ..., \lambda_s^n]$.

Let ν_k^T denote the k-th row of μ^{-1} so that

$$\begin{aligned}
\mathbf{P}^n &= [\mu_1 | \mu_2 | ... | \mu_s] \begin{bmatrix} \lambda_1^n & & & \\ & \lambda_2^n & & \\ & & \ddots & \\ & & & \lambda_s^n \end{bmatrix} \begin{bmatrix} \nu_1^T \\ \nu_2^T \\ \vdots \\ \nu_s^T \end{bmatrix} \\
&= \lambda_1^n \mu_1 \nu_1^T + \cdots + \lambda_s^n \mu_s \nu_s^T \\
&= \lambda_1^n A_1 + \cdots + \lambda_s^n A_s,
\end{aligned}$$

say, where $A_k = \mu_k \nu_k^T$ for $k = 1, \cdots, s$ is a fixed $s \times s$ matrix. ■

With the above proposition, we obtain the following theorem, which quantifies the convergence rate of $\mathbf{P}^n$ toward a t.p.m. formed by the stationary distribution.

Theorem 4.6 *If* $\mathbf{P}$ *is a transition probability matrix of an irreducible and aperiodic Markov chain on a finite state-space, then*

$$\mathbf{P}^n = \mathbf{1}\pi^T + O\left(n^{m_2 - 1} |\lambda_2|^n\right),$$

where $1 = \lambda_1 > |\lambda_2| \geq \cdots \geq |\lambda_s|$ *are eigenvalues of* $\mathbf{P}$ *and* m_2 *is the multiplicity of* λ_2.

Proof: Follows directly from Perron-Frobenius theorem (see Brémaud, 1999, p. 197). ■

The above theorem says that irreducible and aperiodic Markov chains on finite state-spaces converge to the stationary distribution at a *geometric rate*, determined by the second largest eigenvalue modulus.

4.1.2 Example: Two-State Markov Chains

Let

$$\mathbf{P} = \begin{bmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{bmatrix}$$

a stochastic matrix with $\alpha, \beta \in (0,1)$. The eigenvalues of $\mathbf{P}$ are the solutions of $\det(\mathbf{P} - \lambda \mathbf{I}) = 0$; i.e.

$$\begin{vmatrix} 1-\alpha-\lambda & \alpha \\ \beta & 1-\beta-\lambda \end{vmatrix} = 0 \Rightarrow (\lambda - 1)[\lambda - (1-\alpha-\beta)] = 0,$$

so we can take $\lambda_1 = 1$ and $\lambda_2 = 1 - \alpha - \beta$.

Let $\mathbf{1}^T = (1,1)$. Any stochastic matrix must have a unit eigenvalue since $\mathbf{P1} = \mathbf{1}$. If we exclude the trivial case $\alpha = \beta = 0$ (for which $\mathbf{P} = \mathbf{I}$ and the unit eigenvalue is repeated), then $\lambda_1 \neq \lambda_2$ and the proposition implies that

$$\mathbf{P}^n = A_1 + (1-\alpha-\beta)^n A_2, \quad n = 0, 1, ...$$

where A_1 and A_2 are 2×2 matrices. We can determine A_1 and A_2 by setting

$$\begin{aligned} n = 0 &\quad \Rightarrow \quad \mathbf{I} = A_1 + A_2 \\ n = 1 &\quad \Rightarrow \quad \mathbf{P} = A_1 + (1-\alpha-\beta) A_2 \end{aligned}$$

and solving.

Hence,

$$\mathbf{P}^n = \frac{1}{\alpha+\beta} \begin{bmatrix} \beta & \alpha \\ \beta & \alpha \end{bmatrix} + \frac{(1-\alpha-\beta)^n}{\alpha+\beta} \begin{bmatrix} \alpha & -\alpha \\ -\beta & \beta \end{bmatrix}.$$

We have shown previously that

$$\pi^T = \left[\frac{\beta}{\alpha+\beta}, \frac{\alpha}{\alpha+\beta} \right]$$

is the stationary distribution of the corresponding Markov chain, which is confirmed by the expression for $\mathbf{P}^n$ above. Since $-1 < -\beta < 1 - \alpha - \beta < 1$,

$$\lim_{n \to \infty} \mathbf{P}^n = \frac{1}{\alpha+\beta} \begin{bmatrix} \beta & \alpha \\ \beta & \alpha \end{bmatrix} = \mathbf{1}\pi^T.$$

Consider the case:

$$\mathbf{P} = \begin{bmatrix} 1-\alpha & \alpha \\ 0 & 1 \end{bmatrix}$$

so that state 1 is absorbing. State 0 is transient since it can occur finitely many times before absorption into state 1. Let Z be the time to absorption. Then we can find $E_0[Z]$ by first-step analysis (iterated expectation conditioning on first move):

$$\begin{aligned} \mathbb{E}_0[Z] &= \mathbb{E}[Z|X_0 = 0] = \mathbb{E}_0 \{\mathbb{E}[Z|X_1, X_0 = 0]\} \\ &= P_0(X_1 = 0)\mathbb{E}[Z|X_1 = 0, X_0 = 0] + P_0(X_1 = 1)\mathbb{E}[Z|X_1 = 1, X_0 = 0] \\ &= (1-\alpha)\{1 + \mathbb{E}[Z|X_1 = 0]\} + \alpha \times 1 \\ &= (1-\alpha)\{1 + \mathbb{E}_0[Z]\} + \alpha \times 1 \end{aligned}$$

so $\mathbb{E}_0[Z] = 1/\alpha$.

First-step analysis can determine the entire distribution of Z via its *moment generating function* (mgf), $\psi(t) = \mathbb{E}[t^Z|X_0 = 0]$ (note that this is the same as $\mathbb{E}(e^{sZ}|X_0 = 0)$ with $s = \log t$):

$$\begin{aligned} \psi(t) &= P_0(X_1 = 0)\mathbb{E}_0[t^Z|X_1 = 0] + P_0(X_1 = 1)\mathbb{E}_0[t^Z|X_1 = 1] \\ &= (1-\alpha)\mathbb{E}_0[t^{1+Z}] + \alpha t^1 \\ &= (1-\alpha)t\psi(t) + \alpha t, \end{aligned}$$

so

$$\psi(t) = \frac{\alpha t}{1-(1-\alpha)t} = \alpha t + (1-\alpha)\alpha t + (1-\alpha)\alpha t^2 + (1-\alpha)^2\alpha t^3 + ...$$

which is the mgf of a Geometric(α) distribution. Thus, the time to absorption follows a Geometric(α) distribution.

4.2 Reversibility

Reversibility is a property of Markov chain such that it has some reversible properties with respect to its transition probability matrix and stationary distribution. This property involves the concept of being *detailed balance*, which is related to the global balance property we have encountered. Recall that a probability vector π over state space S satisfies **global balance** if

$$\pi^T = \pi^T \mathbf{P}.$$

A probability vector π satisfies **detailed balance** if

$$\pi_i p_{ij} = \pi_j p_{ji} \quad \text{for all } i, j. \tag{4.3}$$

The detailed balance and the global balance are linked by the following proposition.

Proposition 4.7 *Detailed balance $\Rightarrow$ global balance.*

Proof:

$$\sum_j \pi_j p_{ji} = \sum_j \pi_i p_{ij} = \pi_i \sum_j p_{ij} = \pi_i$$

for every i, j so the global balance property is satisfied. ■

A Markov chain with a stationary distribution satisfying the detailed balance is called a *reversible Markov chain*. Note that there are some disagreements among textbook authors about reversibility; some requires that the reversible chain needs to have a initial distribution that is the same as the stationary distribution; some requires the chain to be irreducible.

To see why we called it a reversible Markov chain, consider a homogeneous Markov chain $\{X_0, \cdots, X_n\}$ that is irreducible and positive recurrent with a t.p.m. $\mathbf{P}$ and a stationary distribution π. Now assume that the initial distribution $X_0 \sim \pi$, note that this implies that $X_t \sim \pi$ (marginal distribution of X_t is π) for every $t = 0, 1, \cdots, n$ because π is the stationary distribution.

Now we define $Y_k = X_{n-k}$ for $k \leq n$. The chain $\{Y_0, \cdots, Y_n\}$ is the chain in a reversing time. Note that $\{Y_t : t = 0, \cdots, n\}$ is also a Markov chain (using the graphical representation). Then the conditional probability

$$\begin{aligned} P(Y_k = i | Y_{k-1} = j) &= P(X_{n-k} = i | X_{n-k+1} = j) \\ &= \frac{P(X_{n-k} = i, X_{n-k+1} = j)}{P(X_{n-k+1} = j)} \\ &= \frac{\pi_i p_{ij}}{\pi_j}. \end{aligned}$$

Now consider the conditional probability of the chain $\{X_0, \cdots, X_n\}$:

$$P(X_k = i | X_{k-1} = j) = p_{ji}.$$

The detailed balance property requires $p_{ji} = \frac{\pi_i p_{ij}}{\pi_j)}$, which implies

$$P(Y_k = i | Y_{k-1} = j) = P(X_k = i | X_{k-1} = j).$$

Here you see how the chain is 'reversible'.

In the above example, you see that we require the chain to have an initial distribution being the stationary distribution and is an irreducible chain. This is why some authors would argue that we should add these properties into the definition of a reversible chain. However, the Basin Limit Theorem shows that under suitable assumptions, the (marginal) distribution of X_n from a Markov chain will eventually converge to the stationary distribution. Thus, the future part of the chain behaves just like the chain starting with an initial distribution being the stationary distribution. So even if we do not assume the initial distribution to be the stationary one, the chain will have the reversible property asymptotically.

In addition to making a chain to be reversible, the detailed balance provides an alternative way to find the stationary distribution as illustrated in the following two examples.

Example: Ehrenfest Model of Diffusion. Recall that Ehrenfest model's transition probabilities are

$$p_{ij} = \begin{cases} \frac{i}{N} & \text{if } j = i-1 \\ 1 - \frac{i}{N} & \text{if } j = i+1. \end{cases}$$

$\{X_n\}$ is an irreducible Markov chain on a finite state space, so the chain is positive recurrent with a unique stationary distribution. However, we do not know the stationary distribution. One way to find this stationary distribution is to solve the global balance equations $\pi^T \mathbf{P} = \pi^T$. Alternatively, we can try to "guess" that at equilibrium $X_n \sim \text{Bin}(N, \frac{1}{2})$ and verify this candidate stationary distribution via the detailed balance equation. Notice we do not know whether the Ehrenfest chain is reversible, but we'll go ahead with the detailed balance check anyway. Entries of our candidate probability vector are

$$\pi_i = \binom{N}{i} \left(\frac{1}{2}\right)^i \left(1 - \frac{1}{2}\right)^{N-i} = \binom{N}{i} \frac{1}{2^N} \tag{4.4}$$

Since X_n can only increase or decrease by one at each time step, we need to check detailed balance only for i and $j = i+1$.

$$\begin{aligned} \pi_i p_{i,i+1} &= \frac{1}{2^N} \binom{N}{i} \frac{N-i}{N} = \frac{1}{2^N} \frac{N!}{i!(N-i)!} \frac{N-i}{N} \\ &= \frac{1}{2^N} \frac{N!}{(i+1)!(N-i-1)!} \frac{i+1}{N} \\ &= \binom{N}{i+1} \frac{1}{2^N} \frac{i+1}{N} = \pi_{i+1} p_{i+1,i}, \end{aligned}$$

confirming our guess.

Example: Random Walk on a Graph. Let $G = (\mathcal{V}, \mathcal{E})$ be a graph, where $\mathcal{V}$ is the vertex set (assumed to be finite) and $\mathcal{E} = \{(i,j) : i, j \in \mathcal{V}\}$ is the edge set. Let $d(i)$ be the degree of note i; the degree is the number of edges connecting to vertex i. Assume that there is no edge from a node to itself. A common model for a random walk on a graph is a Markov chain on $\mathcal{V}$ with the transition probabilities

$$p_{ij} = \begin{cases} \frac{1}{d(i)}, & \text{if } (i,j) \in \mathcal{E} \\ 0, & \text{if } (i,j) \notin \mathcal{E}. \end{cases}$$

Note that a random walk on a graph model is used to spectral analysis; it is related to the spectral clustering, a popular clustering method. If we assume that the graph is connected, then the chain $\{X_n\}$ is irreducible and positive recurrent because $\mathcal{V}$ is finite. To find the stationary distribution, again we use the detailed balance. For $(i,j) \in \mathcal{E}$,

$$\pi_i p_{ij} = \pi_j p_{ji} \Rightarrow \pi_i \frac{1}{d(i)} = \pi_j \frac{1}{d(j)} \Rightarrow \frac{\pi_i}{d(i)} = c$$

for some constant c. To find c,

$$\pi_i = c \cdot d(i) \Rightarrow 1 = \sum_i \pi_i = c \sum_i d(i) \Rightarrow c = \frac{1}{2m},$$

where $m = |\mathcal{E}|$ is the total number of edges. Thus, the stationary distribution is

$$\pi_i = \frac{d(i)}{2m}.$$

4.3 Fundamental matrices

Finally, we introduce the concept of fundamental matrices. Fundamental matrices can be defined for both absorbing and irreducible Markov chains and are powerful tools in deriving certain properties. In a Markov chain, a state i is called absorbing if $p_{ii} = 1$. A Markov chain is called an absorbing Markov chain if every state can reach an absorbing state. Note that if a Markov chain contains transient states and recurrent states forming a few communicating classes, we can merge a communicating class of recurrent states into an absorbing state. This reduces the problem into an absorbing Markov chain problem.

- For absorbing Markov chain, fundamental matrices can be used to compute the probabilities of absorption of recurrent states.
- For irreducible Markov chains, fundamental matrices can be used to calculate $\mathbb{E}_i(T_j)$, the mean return time to state j from state i, and the asymptotic variance of an ergodic estimator (equation (4.1)).

4.3.1 Absorbing Markov chain

Let $\{X_n\}$ be a Markov chain with state space $S = \{1, 2, \cdots, s\}$ and we assume that T, A are a partition of S such that

- $T = \{1, 2, \cdots, m\}$ are the transient states, and
- $A = \{m+1, m+2, \cdots, s\}$ are absorbing states.

Namely, $p_{ij} = 0$ for all $i \in A$ and $j \in T$ and for every $i \in T$, there exists $j \in A$ such that $i \to j$ but $j \not\leftarrow j$.

Because $S = \{T, A\}$, we can write the transition probability matrix in block matrix form as

$$\mathbf{P} = \begin{bmatrix} \mathbf{Q} & \mathbf{R} \\ \mathbf{0} & \mathbf{S} \end{bmatrix}, \tag{4.5}$$

where $\mathbf{Q}$ is an $m \times m$ matrix, $\mathbf{R}$ is an $m \times (s-m)$ matrix, and $\mathbf{S}$ is an $(s-m) \times (s-m)$ matrix.

Example: Gambler's ruin. Consider a gambler's ruin problem with $a + b = 4$. Recall that X_n is the amount of money that the first player has at turn n. In this case, $S = \{0, 1, 2 \cdots, 4\}$ so there are totally

5 states. States 0 and 4 are absorbing and $1, 2, 3$ are transient. The transition probability matrix can be expressed as

$$\mathbf{P} = \begin{array}{c} \nearrow \\ 1 \\ 2 \\ 3 \\ 0 \\ 4 \end{array} \begin{array}{c} \begin{matrix} 1 & 2 & 3 & 0 & 4 \end{matrix} \\ \begin{bmatrix} 0 & p & 0 & q & 0 \\ q & 0 & p & 0 & 0 \\ 0 & q & 0 & 0 & p \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array} = \begin{bmatrix} \mathbf{Q} & \mathbf{R} \\ \mathbf{0} & \mathbf{S} \end{bmatrix},$$

We are often interested in the probability of starting from a transient state to an absorbing state. For instance, in gambler's ruin problem, we are interested the probability of player 1 wins with certain amount of starting money. The winning probability can be viewed as the probability of starting from a transient state (the initial amount of money) to the absorbing state $a + b$ (total amount of money).

Let $i \in T$ be a transient state and $j \in A$ be an absorbing sate and $h_{ij} = P(A_j | X_0 = i, i \in T)$, where A_j is the event "absorbed through j". Using the first-step analysis, we can decompose h_{ij} :

$$\begin{aligned} h_{ij} &= P(A_j | X_0 = i, i \in T) \\ &= P(A_j | X_1 = j, X_0 = i, i \in T) P(X_1 = j | X_0 = i, i \in T) \\ &\quad + \sum_{k \neq j} P(A_j | X_1 = k, X_0 = i, i \in T) P(X_1 = k | X_0 = i, i \in T) \\ &= 1 \cdot p_{ij} + \sum_{k \neq j} P(A_j | X_1 = k) p_{ik} \\ &= p_{ij} + \sum_{k=1}^{m} P(A_j | X_1 = k) p_{ik} \quad \text{(since the first } m \text{ states are transient)} \\ &= p_{ij} + \sum_{k=1}^{m} p_{ik} h_{kj}. \end{aligned}$$

Using the matrices $\mathbf{Q}, \mathbf{R}$, and $\mathbf{P}$, we can reexpress the above equality as

$$\mathbf{H} = \mathbf{R} + \mathbf{Q}\mathbf{H} \Rightarrow \mathbf{H} = (\mathbb{I}_m - \mathbf{Q})^{-1} \mathbf{R},$$

where $\mathbf{H} = \{h_{ij}\}$ is an $m \times (s - m)$ matrix.

In the right-handed side we use the fact that $(\mathbb{I}_m - \mathbf{Q})^{-1}$ exists. Will the inverse always exists? The follow proposition provides a condition for the existence of the inverse.

Proposition 4.8 *Let* $\mathbf{Q}$ *be an* $m \times m$ *square matrix with* $\lim_{n\to\infty} \mathbf{Q}^n = 0$. *Then* $(\mathbb{I}_m - \mathbf{Q})^{-1}$ *exists and* $(\mathbb{I}_m - \mathbf{Q})^{-1} = \sum_{n=0}^{\infty} \mathbf{Q}^n$.

Proof: Since

$$(\mathbb{I}_m - \mathbf{Q})(\mathbb{I}_m + \mathbf{Q} + \cdots + \mathbf{Q}^{n-1}) = \mathbb{I}_m - \mathbf{Q}^n, \tag{4.6}$$

taking determinant in both sides, we obtain

$$\mathsf{det}(\mathbb{I}_m - \mathbf{Q}) \cdot \mathsf{det}(\mathbb{I}_m + \mathbf{Q} + \cdots + \mathbf{Q}^{n-1}) = \mathsf{det}(\mathbb{I}_m - \mathbf{Q}^n).$$

Because determinant is a continuous function, we have

$$\lim_{n\to\infty} \mathsf{det}(\mathbb{I}_m - \mathbf{Q}^n) = \mathsf{det}\left(\lim_{n\to\infty} (\mathbb{I}_m - \mathbf{Q}^n)\right) = \mathsf{det}(\mathbb{I}_m - 0) = 1.$$

Since the limit of determinant is non-zero, there exists some k such that $\mathsf{det}(\mathbb{I}_m - \mathbf{Q}^k) \neq 0$ which implies $\mathsf{det}(\mathbb{I}_m - \mathbf{Q}) \neq 0$ so $(\mathbb{I}_m - \mathbf{Q})^{-1}$ exists. Moreover, taking limits in both sides of equation (4.6),

$$(\mathbb{I}_m - \mathbf{Q}) \cdot \sum_{n=0}^{\infty} \mathbf{Q}^n = \lim_{n\to\infty} \mathbb{I}_m - \mathbf{Q}^n = \mathbb{I}_m,$$

which implies

$$(\mathbb{I}_m - \mathbf{Q})^{-1} = \sum_{n=0}^{\infty} \mathbf{Q}^n.$$

■

Now we come back to our Markov chain with both transient and absorbing states. Can we apply the above proposition? The answer is yes! To see this, we need to show that $\lim_{n\to\infty} \mathbf{Q}^n = 0$. Recall that in equation (4.5), the transition probability matrix has a block form. It turns out that the n-step transition probability matrix also has a block form with

$$\mathbf{P}^n = \begin{bmatrix} \mathbf{Q}^n & \mathbf{R}^{(n)} \\ \mathbf{0} & \mathbf{S}^n \end{bmatrix},$$

where $\mathbf{R}^{(n)}$ is some complex matrix (think about why we obtain the above form). Therefore, for each $i, j \in T$, $p_{ij}^{(n)} = q_{ij}^{(n)}$ with $\mathbf{Q}^n = \{q_{ij}^{(n)}\}$. Recall that if j is a transient state, $\lim_{n\to\infty} p_{ij}^n = 0$ so $q_{ij}^{(n)} \to 0$ for every $j \in T$, which implies $\lim_{n\to\infty} Q^n = 0$. Therefore, the equation $\mathbf{H} = (\mathbb{I}_m - \mathbf{Q})^{-1}\mathbf{R}$ is well-defined.

The quantity $(\mathbb{I}_m - \mathbf{Q})^{-1} = \sum_{n=0}^{\infty} \mathbf{Q}^n$ is called the **fundamental matrix** of the absorbing Markov chain.

The fundamental matrix can also be used to solve the hitting time problem. Let t_{ij} be the expected number of visits to transient state j from transient state i. Namely,

$$t_{ij} = \mathbb{E}_i \left(\sum_{n=0}^{\infty} I(X_n = j) \right)$$

for $i, j \in T$. Using the first step analysis, we obtain

$$t_{ij} = I(i = j) + \sum_{k=1}^{m} p_{ik} t_{kj} = I(i = j) + \sum_{k=1}^{m} q_{ik} t_{kj}.$$

Using the matrix form $\mathbf{T} = \{t_{ij}\}$, we obtain

$$\mathbf{T} = \mathbb{I}_m + \mathbf{QT} \Rightarrow \mathbf{T} = (\mathbb{I}_m - \mathbf{Q})^{-1}.$$

Also, if we are thinking about the expected number of steps to get absorbed from transient state i, we can also use the fundamental matrix to help. Note that $t_i = \sum_{j=1}^{m} t_{ij}$ is the expected number of steps to get absorbed (think about why). Let $\mathbf{t}^T = (t_1, \cdots, t_m)$. Then

$$\mathbf{t} = \mathbf{T}\mathbf{1}_m = (1 - \mathbf{Q})^{-1}\mathbf{1}_m,$$

where $\mathbf{1}_m^T = (1, 1, \cdots, 1) \in \mathbb{R}^s$.

Example: Gambler's ruin.
Consider the gambler's ruin problem with $a + b = 4$ (we have discussed about the block matrix form of this problem). Assume further $p = q = 1/2$. Then

$$\mathbf{Q} = \begin{bmatrix} 0 & 1/2 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1/2 & 0 \end{bmatrix}, \quad \mathbf{R} = \begin{bmatrix} 1/2 & 0 \\ 0 & 0 \\ 0 & 1/2 \end{bmatrix}, \quad \mathbf{S} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

so that

$$\mathbb{I}_m - \mathbf{Q} = \begin{bmatrix} 1 & -1/2 & 0 \\ -1/2 & 1 & 1/2 \\ 0 & -1/2 & 1 \end{bmatrix}$$

and

$$\mathbf{T} = (\mathbb{I}_m - \mathbf{Q})^{-1} = \begin{bmatrix} 3/2 & 1 & 1/2 \\ 1 & 2 & 1 \\ 1/2 & 1 & 3/2 \end{bmatrix}$$

So, for example, $t_{11} = 3/2$ is the expected number of visits to state 1 starting from 1 and $t_{13} = 1/2$ is the expected number of visits to state 3 starting from state 1. The hitting probabilities are

$$\mathbf{H} = (\mathbb{I}_m - \mathbf{Q})^{-1}\mathbf{R} = \begin{bmatrix} 3/2 & 1 & 1/2 \\ 1 & 2 & 1 \\ 1/2 & 1 & 3/2 \end{bmatrix} \begin{bmatrix} 1/2 & 0 \\ 0 & 0 \\ 0 & 1/2 \end{bmatrix} = \begin{bmatrix} 3/4 & 1/4 \\ 1/2 & 1/2 \\ 1/4 & 3/4 \end{bmatrix}$$

Hence, if we start in state 1 there is a probability of 3/4 that we are absorbed through state 0 and 1/4 through 4.

Example: Monitoring Deaths in the SIS model.
Suppose an individual in a population has four possible states: S, susceptible, I, infectious, D_D, dead from disease complications, D_O, dead from other causes. The individual moves between these over the state-space $E = \{S, I, D_O, D_D\}$ according to a Markov chain with the following transition probability matrix:

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & 0 \\ p_{21} & p_{22} & 0 & p_{24} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

In our absorbing Markov chain notation

$$\mathbf{Q} = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}, \qquad \mathbf{R} = \begin{bmatrix} p_{13} & 0 \\ 0 & p_{24} \end{bmatrix}, \qquad \mathbf{S} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

The fundamental matrix is

$$(\mathbf{I} - \mathbf{Q})^{-1} = \frac{1}{(1-p_{11})(1-p_{22}) - p_{12}p_{21}} \begin{bmatrix} 1-p_{22} & p_{12} \\ p_{21} & 1-p_{11} \end{bmatrix}.$$

The hitting probabilities can be computed as

$$\mathbf{H} = (\mathbf{I} - \mathbf{Q})^{-1} \begin{bmatrix} p_{13} & 0 \\ 0 & p_{24} \end{bmatrix}$$

and hitting times as $\mathbf{T} = (\mathbf{I} - \mathbf{Q})^{-1}$.

4.3.2 Irreducible Markov chain

Now we consider the Markov chain $\{X_n\}$ to be irreducible. Moreover, we will assume that $\{X_n\}$ is also aperiodic and has a finite state space. Note that an irreducible and aperiodic Markov chain is also called an ergodic Markov chain. Let $\mathbf{P}$ be its transition matrix and π be its stationary distribution. Then the vector

$$Z = (\mathbb{I}_s - \mathbf{P} + \mathbf{1}_s\pi^T)^{-1} \tag{4.7}$$

is called the fundamental matrix of the irreducible Markov chain $\{X_n\}$.

Proposition 4.9 *The fundamental matrix Z of the above is well-defined and can be expressed as*

$$Z = \mathbb{I}_s + \sum_{n=1}^{\infty} (\mathbf{P}^n - \mathbf{1}_s \pi^T).$$

Proof: Before proceeding to the proof, we first note the following fact about the matrix $\mathbf{1}_s\pi^T$:

$$\begin{aligned} \mathbf{1}_s\pi^T\mathbf{P} &= \mathbf{1}_s\pi^T \\ \mathbf{P}\mathbf{1}_s\pi^T &= \mathbf{1}_s\pi^T \Rightarrow \mathbf{P}^k\mathbf{1}_s\pi^T = \mathbf{1}_s\pi^T \\ \mathbf{1}_s\pi^T\mathbf{1}_s\pi^T &= \mathbf{1}_s\pi^T \Rightarrow (\mathbf{1}_s\pi^T)^k = \mathbf{1}_s\pi^T. \end{aligned}$$

Because $\mathbf{1}_s\pi^T\mathbf{P} = \mathbf{P}\mathbf{1}_s\pi^T$, we can expand

$$\begin{aligned} (\mathbf{P} - \mathbf{1}_s\pi^T)^n &= \sum_{k=1}^{n} \binom{n}{k} (-1)^{n-k} \mathbf{P}^k (\mathbf{1}_s\pi^T)^{n-k} \\ &= \mathbf{P}^n + \underbrace{\sum_{k=1}^{n-1} \binom{n}{k} (-1)^{n-k}}_{=1} \underbrace{\mathbf{P}^k\mathbf{1}_s\pi^T}_{=\mathbf{1}_s\pi^T} \underbrace{(\mathbf{1}_s\pi^T)^{n-k-1}}_{=\mathbf{1}_s\pi^T} \\ &= \mathbf{P}^n - \mathbf{1}_s\pi^T. \end{aligned} \tag{4.8}$$

Note that we use the fact that

$$0 = (1-1)^n = \sum_{k=1}^{n} \binom{n}{k} (-1)^{n-k} = \sum_{k=1}^{n-1} \binom{n}{k} (-1)^{n-k} - 1.$$

Using the fact that for any square matrix A,

$$(\mathbb{I}_s - A)(\mathbb{I}_s + A + \cdots + A^{n-1}) = \mathbb{I}_s - A^n,$$

we conclude

$$(\mathbb{I}_s - (\mathbf{P} - \mathbf{1}_s\pi^T))(\mathbb{I}_s + (\mathbf{P} - \mathbf{1}_s\pi^T) + \cdots + (\mathbf{P} - \mathbf{1}_s\pi^T)^{n-1}) = \mathbb{I}_s - (\mathbf{P} - \mathbf{1}_s\pi^T)^n = \mathbb{I}_s - \mathbf{P}^n + \mathbf{1}_s\pi^T.$$

Because $\lim_{n\to\infty} \mathbf{P}^n + \mathbf{1}_s\pi^T = 0$, taking $\lim_{n\to\infty}$ in both sides, we conclude

$$(\mathbb{I}_s - (\mathbf{P} - \mathbf{1}_s\pi^T)) \left(\mathbb{I}_s + \sum_{k=1}^{\infty} (\mathbf{P} - \mathbf{1}_s\pi^T)^k \right) = \mathbb{I}_s.$$

Therefore, using equation (4.8),

$$Z = \mathbb{I}_s - (\mathbf{P} - \mathbf{1}_s\pi^T)^{-1} = \mathbb{I}_s + \sum_{k=1}^{\infty} (\mathbf{P} - \mathbf{1}_s\pi^T)^k = \mathbb{I}_s + \sum_{n=1}^{\infty} (\mathbf{P}^n + \mathbf{1}_s\pi^T).$$

■

The fundamental matrix allows us to compute some useful quantities as described in the following two theorems.

Theorem 4.10 *Let $\{X_i\}$ be a ergodic Markov chain with a finite state space and let Z be the fundamental matrix defined in equation (4.7). Then for all $i \neq j$, the mean return time to state j from state i is*

$$\mathbb{E}_i(T_j) = \frac{Z_{jj} - Z_{ij}}{\pi_j}.$$

Proof: See Brémaud, 1999, p. 231 (Theorem 6.4). ■

Theorem 4.11 *Let $\{X_i\}$ be a ergodic Markov chain with a finite state space and let Z be the fundamental matrix defined in equation (4.7). For any function $f : S \mapsto \mathbb{R}$, define $f_i = f(i)$ for each $i \in S$. Then regardless of the initial distribution,*

$$\lim_{N\to\infty} N\mathsf{Var}\left(\frac{1}{N}\sum_{i=1}^{N} f(X_i)\right) = 2\sum_{i,j} \pi_i Z_{ij} f_i f_j - \sum_{i,j} \pi_i (\mathbb{I}_s - \mathbf{1}_s \pi^T)_{ij} f_i f_j.$$

Proof: See Brémaud, 1999, p. 232 (Theorem 6.5). ■