12.1 Recap

Consider a discrete-time Markov chain: \((X_0, X_1, X_2, \ldots) \sim \text{Markov}(\lambda, P)\). If the Markov chain is irreducible and positive recurrent, there exists a unique invariant distribution:

\[ \pi_i := \frac{1}{E_i(R_i)}, \quad i \in I. \]

Furthermore, a Markov chain is aperiodic if \( \forall i \in I, \exists N < \infty \) such that:

\[ p^{(n)}_{ii} > 0 \quad \forall \ n \geq N \]

**Theorem 12.1.** Let \( P \) be irreducible and positive recurrent with unique invariant measure \( \pi \). Also assume that \( P \) is aperiodic. Then,

\[
\mathbb{P}(X_n = j) \xrightarrow{n \to \infty} \pi_j \quad \forall \ j \in I
\]

In particular, \( p^{(n)}_{ij} \xrightarrow{n \to \infty} \pi_j \quad \forall \ i, j \in I \)

i.e. \( P^n \to \begin{bmatrix} \pi_1 & \ldots & \pi_i & \ldots \\ \pi_1 & \ldots & \pi_i & \ldots \\ \pi_1 & \ldots & \pi_i & \ldots \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix} \)

Before we prove the theorem, let’s look at coupling technique.

**Definition 12.2.** A coupling between two stochastic processes \((X_i, i \in \Lambda)\) and \((Y_i, i \in \Lambda)\) is a joint distribution \(((\tilde{X}_i, \tilde{Y}_i), i \in \Lambda)\) such that:

\[ (\tilde{X}_i, i \in \Lambda) \overset{d}{=} (X_i, i \in \Lambda) \]

\[ (\tilde{Y}_i, i \in \Lambda) \overset{d}{=} (Y_i, i \in \Lambda) \]

The goal of coupling is to define the processes in the same space. There is no unique coupling. Coupling examples include:

1. Independent coupling.
2. Equal coupling: When \((X_i, i \in \Lambda) \overset{d}{=} (Y_i, i \in \Lambda)\), we can take \(\tilde{Y}_i = \tilde{X}_i = X_i\).

Proof. Consider two discrete-time Markov chains:

\[(X_0, X_1, X_2, \ldots) \sim \text{Markov}(\lambda, P)\]
\[(Y_0, Y_1, Y_2, \ldots) \sim \text{Markov}(\pi, P)\]

Here \(Y\) is stationary since we start with \(\pi\), and \(X\) and \(Y\) are independent of each other.

![Figure 12.1](image)

Figure 12.1: When a Markov chain \(X\) hits a new Markov chain \(Y\), we can follow the new one for future states. This is due to the property that future states in the Markov chain do not depend on the past states.

Thus, to prove theorem [12.1] we have to show that \(X\) and \(Y\) will hit each other with probability 1.

**Step 1**: Define a 2-dimensional Markov chain \(W_i = (X_i, Y_i)\): \((W_0, W_1, W_2, \ldots) \sim \text{Markov}(\tilde{\lambda}, \tilde{P})\). Here, the initial distribution can be represented as:

\[\tilde{\lambda}(i, j) = \lambda_i \pi_j, \quad (i, j) \in I \times I\]

Furthermore, since the chains are independent of each other:

\[\tilde{p}_{ij,kl} = p_{ik}p_{jl}, \quad (i, j) \in I \times I, \quad (k, l) \in I \times I\]

**Exercise 12.1.** Show that \((W_i, i \geq 0)\) is a Markov chain.

**Step 2**: Fix \(b \in I\) such that \(T = \inf\{n \geq 0 \mid X_n = Y_n = b\} = R_{bb}\)

Here \(T\) is the hitting time for \(bb\), therefore \(T\) is a stopping time. Here choice of \(b\) is not important. To improve the stopping time we can consider the stopping time with \(X_n = Y_n\) for the first time. We want to show that \(T\) is finite, i.e.:

\[P(T < \infty) = 1\]

**Step 3**: The new Markov chain \(W\) is irreducible. Fix \(ij, kl\). Then,

\[p_{ij,kl}^{(n)} = p_{ik}^{(n)} p_{jl}^{(n)}\]

since the chains \(X\) and \(Y\) are independent.
Exercise 12.2. For an irreducible and aperiodic Markov chain, \( i, j \in N \) such that \( p^{(n)}_{ij} > 0, \forall n \geq N \).

Using the result from the above exercise:

For \( ij \), \( \exists N \) such that \( p^{(n)}_{ij} > 0, \forall n \geq N \)

For \( kl \), \( \exists M \) such that \( p^{(n)}_{kl} > 0, \forall n \geq M \)

\[ \implies p^{(n)}_{ij} p^{(n)}_{kl} > 0, \forall n \geq \max(N, M) \]

Step 4: The new Markov chain is positive recurrent:

\[ \hat{\pi}_{kl} = \pi_k \pi_l, \quad (k, l) \in I \times I \]

Here, \( \hat{\pi} \) is an invariant measure for \( \hat{p} \). This implies that \( \hat{p} \) is positive recurrent.

Thus, combining the results from steps 3 and 4, we see that \( X \) and \( Y \) will hit \( b \) in a finite time, i.e.:

\[ \mathbb{P}(T < \infty) = 1 \]

Step 5: By strong Markov property,

\[ (W_{T+0}, W_{T+1}, \ldots) \sim \text{Markov}(\delta_{bb}, \hat{P}) \]

i.e., the process after the hitting time is a Markov chain with initial distribution as \( \delta_{bb} \) and the same transition matrix as earlier. Furthermore, the future states do not depend on (are independent of) the past states \( (W_0, W_1, \ldots, W_T) \).

Step 6: Define a new Markov chain \( \hat{W}_i = (Y_i, X_i) \). Here:

\[ (\hat{W}^n_{T+n}, n \geq 0) \overset{d}{=} (W_{T+n}, n \geq 0) \]

i.e., after the chains hit at time \( T \), we can flip them and still have the same distribution.

\[ (W_0, W_1, \ldots, W_T, W_{T+1}, \ldots) \overset{d}{=} (W_0, W_1, \ldots, W_T, \hat{W}_{T+1}, \ldots) \sim \text{Markov}(\hat{\lambda}, \hat{P}) \]

A similar result applies for the 1-dimensional Markov chain \( X \). Thus, after the two chains hit at \( T \), \( X \) and \( Y \) will have the same distribution.

\[ (X_0, X_1, \ldots) \overset{d}{=} (X_0, X_1, \ldots, X_T, Y_{T+1}, \ldots) \sim \text{Markov}(\lambda, P) \]

We define

\[ Z_n = \begin{cases} X_n & \text{if } n \leq T \\ Y_n & \text{if } n > T. \end{cases} \]

This operation is often referred to as a surgery of the chains.
Step 7:
Finally, we prove that as $n \uparrow \infty$, the Markov chain $X$ converges to the stationary Markov chain $Y$:

$$P(X_n = j) \to \pi_j \quad \forall \ j \in I$$

Looking at the Markov chains individually, we have:

$$P(X_n = j) = P(Z_n = j) = P(T \leq n, Y_n = j) + P(T > n, X_n = j)$$

$$\pi_j = P(Y_n = j) = P(T \leq n, Y_n = j) + P(T > n, Y_n = j)$$

Thus, on combining the above two equations, we have

$$|P(X_n = j) - \pi_j| = |P(T > n, X_n = j) - P(T > n, Y_n = j)| \leq P(T > n) \xrightarrow{n \to \infty} P(T = \infty) = 0$$

\[\blacksquare\]

Remark 12.3.
- For rate of convergence, we need to estimate $P(T > n)$.
- For optimal rate, do a better coupling and use the stopping time when $X_n = Y_n$ for the first time.

Example 12.4. Lazy random walk on $n$-cycle: A regular Markov chain can be converted to a lazy random walk as follows:

$$P \to \frac{1}{2}P + \frac{1}{2}Id.$$ 

Here $Id.$ stands for an identity matrix. Thus we obtain a lazy Markov chain and it is always aperiodic.

Let’s consider the lazy random walk shown below:

We can observe that it irreducible, positive recurrent and aperiodic. Furthermore:

$$\pi = \frac{1}{N} \quad \forall i$$

is the unique invariant distribution.

Consider the specific example below, where $n = 5$:

Here let’s set up the system such that when $X$ stays at same point, $Y$ must move and vice-versa. Therefore, the distance at every step will always change by 1.

With probability 0.5:
Figure 12.2: Lazy Random Walk

Figure 12.3: Lazy Random Walk with \( n = 5 \)

\[
X_1 = X_0 \\
Y_1 = \begin{cases} 
Y_0 - 1 & \text{with probability } 0.5 \\
Y_0 + 1 & \text{with probability } 0.5 
\end{cases}
\]

and with probability 0.5:

\[
Y_1 = Y_0 \\
X_1 = \begin{cases} 
X_0 - 1 & \text{with probability } 0.5 \\
X_0 + 1 & \text{with probability } 0.5 
\end{cases}
\]

Exercise 12.3. Show that: \( E_k(T) = k(n - k) \leq \frac{n^2}{4} \)