Lecturer: Hung Q. Ngo
Scribe: Peng Lin

Lecture 6: Convergence to equilibrium, Ergodic Theorem

1 Definitions

We first give out the following definitions:

Definition 1.1 (Measures and distributions). A vector $\lambda = (\lambda_i : i \in I)$ consisting of coordinates $\lambda_i \geq 0, i \in I$, is called a *measure* on I. If the sum of all coordinates is 1, then it is called a *distribution*.

Definition 1.2 (Invariant measure and distribution). Let X be an irreducible recurrent markov chain with transition matrix P, if there exists some vector λ s.t. $\lambda = \lambda P$, then λ is called an *invariant measure* with respect to P for X. Similarly, if it holds that

$$\sum_{i \in I} \lambda_i P_{ij} = \lambda_j$$

, it's called an *invariant distribution*.

With the above mentioned properties of invariant distribution, we can come up with the following propositions:

Proposition 1.3. *If $(X_n)_{n\geq 0}$ is a Markov chain with (λ, P) and λ is invariant with respect to P, then $(X_{N+n})_{n\geq 0}$ is also a Markov chain with (λ, P) for any N. It's also called* stationary, equilibrium.

Proof. From the definition of invariant distribution, we have $\lambda = \lambda P$. By applying the property recursively, it is easy to see that

$$\text{Distribution of } X_{N+n} = \lambda P^N = \lambda P(P^{N-1}) = \lambda P^{N-1} = \cdots = \lambda \tag{1}$$

□

Proposition 1.4. *Suppose $P_{ij}^{(n)}$ tends to π_j as $n \to \infty, \forall j \in I$, i.e., the probability of ending up at j is independent on the starting state, then $\pi = (\pi_j : j \in I)$ is an invariant distribution.*

Proof. We want to show that

$$\pi_j = \sum_{i \in I} \pi_i P_{ij}$$

, Since by definition, $lim_{n\to\infty} P_{ij}^{(n)} = \pi_j$, we have

$$\begin{aligned}
\sum_{i \in I} \pi_i P_{ij} &= \sum_{i \in I} (lim_{n\to\infty} P_{ki}^{(n)}) P_{ij} \\
&= lim_{n\to\infty} (\sum_{i \in I} P_{ki}^{(n)}) P_{ij} \\
&= lim_{n\to\infty} P_{kj}^{(n+1)} \\
&= \pi_j
\end{aligned}$$

□

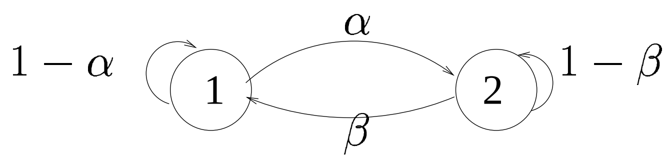

Figure 1: A two-state Markov chain

Example 1.5 (convergency). If a Markov chain is convergent, it must have an invariant distribution.

Let's consider a two-state Markov chain as illustrated in Figure 1. The transition matrix

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

Ignoring the case when $\alpha = \beta = 0, \alpha = \beta = 1$, we have

$$P^n = \begin{pmatrix} \frac{\beta}{\alpha+\beta} & \frac{\alpha}{\alpha+\beta} \\ \frac{\beta}{\alpha+\beta} & \frac{\alpha}{\alpha+\beta} \end{pmatrix}$$

when $n \to \infty$.

From 1.3, we can see that the row components form an invariant distribution. If β is larger, the final state will most likely end up at state 1. However if $\alpha = \beta = 0$, then will end up either at 1 or 2, and the Markov chain will not converge.

Next, we will answer the question "when does a Markov chain have an invariant distribution".

Theorem 1.6. *If P is irreducible, then the following claims are equivalent:*

1. *All states are positive recurrent.*
2. *Some state is positive recurrent.*
3. *P has an invariant distribution π.*

Before showing the proof for the theorem, we first give out two lemmas. For a certain state k, we can find the expected time spent in state i before visiting k:

$$\gamma_i^k = E_k \sum_{n=0}^{T_k - 1} 1_{\{X_n = i\}}$$

where T_k is the first passage time to k.

Lemma 1.7. *If P is irreducible and recurrent, then*

1. $\gamma_k^{(k)} = 1$.
2. $\gamma^{(k)} P = \gamma^{(k)}$.
3. $0 < \gamma_i^{(k)} < \infty, \forall i \in I$.

Lemma 1.8. *If P is irreducible, λ is invariant measure with $\lambda_k = 1$, then $\lambda \geq \gamma^{(k)}$. Moreover, if P is recurrent, then $\lambda = \gamma^{(k)}$.*

With the lemmas given above, we can show the proofs for the theorem as follows:

Proof. 1. $(i) \Rightarrow (ii)$. This is obvious.

2. $(ii) \Rightarrow (iii)$. Suppose i is positive recurrent, then P is recurrent. Since i is positive recurrent, the expected total number of steps before visits to state i: $\mu_i = \sum_{j\in I} \gamma_j^{(i)} < \infty$. Take $\pi = \frac{\gamma^{(i)}}{\mu_i}$, it's easy to see $\sum_{i\in I} \pi_i = 1$. From lemma1.7, π is an invariant distribution.

3. $(iii) \Rightarrow (i)$. Take any state j, $\pi_j = \sum_{i\in I} \pi_i P_{ij}^{(n)}, \forall n \geq 1$. And since π is a distribution, $\sum_{i\in I} \pi_i = 1$. Since P is irreducible, i.e., $\exists k, \mid \pi_k, P_{kj} > 0$, and from the definition of π_j, we have $\pi_j > 0$, $\forall j$. Let $\lambda = (\frac{\pi_1}{\pi_k}, \frac{\pi_2}{\pi_k}, \dots)$, then λ is an invariant measure with $\lambda_k = 1$. Hence, we have $\lambda \geq \gamma^{(k)}$ by Lemma1.8. It follows that:

$$\mu_k = \sum_{j\in I} \gamma_j^{(k)} \leq \sum_{j\in I} \frac{\pi_j}{\pi_k} = \frac{1}{\pi_k} < \infty$$

That is, k is positive recurrent as long as $\pi_k > 0$. And since $\pi_j > 0, \forall j$, we are done.

□

Next, we will look at the problem when a markov chain will converge.

First, we show that periodic is not a good property: Consider the example Markov chain given in figure 1, $P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. $P^2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. $P^3 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = P$. Hence the Markov chain has a period of 2 and will not converge.

Theorem 1.9. *If P is irreducible, aperiodic and has an invariant distribution π, then*

$$\lim_{n\to\infty} P[X_n = j] = \pi_j$$

In other words,

$$lim_{n\to\infty} P_{ij}^{(n)} = \pi_j$$

Given P, $i \in I$, define d_i = period of i $= gcd\{n \geq 1, P_{ii}^{(n)} > 0\}$. If $d_i \times n$, ($\times$ means not divisible by, $|$ means divisible), then $P_{ii}^{(n)} = 0$

If $d_i = 1$, we say i is aperiodic.

Proposition 1.10. *1. If i,j are in same class, then $d_i = d_j$*

2. If i is aperiodic, then $\exists n_o, \mid P_{ii}^{(n)} > 0, \forall n \geq n_o$.

Proof. 1. Since P is irreducible, $\exists n, m > 0 \mid P_{ij}^{(n)}, P_{ji}^{(m)} > 0$. $P_{ii}^{(n+m)} > 0 \Rightarrow d_i \mid (n+m)$. $P_{ii}^{n+m+L} > 0$ if $P_{jj}^{L} > 0 \Rightarrow d_j \mid L, d_j \mid (n+m) \Rightarrow d_j \mid d_i$

2. Let $S = \{n \geq 1 : P_{ii}^{(n)} > 0\}$ then $d_i = gcd(S) = 1$. Note:

   (a) If $x, y \in S$, then $ax + by \in S$,$\forall a, b \in Z^+$.

   (b) $\exists x_1, x_2, \dots, x_m \in S$, $gcd(x_1, x_2, \dots, x_m) = 1$. Take $x_1 \in S$ and $x_1 = P_1^{e_1} \dots P_k^{e_k}$, Let $S_1 = \{n \in S : P_1 \times n\} \neq \emptyset, \dots, S_k = \{n \in S : P_k \times n\} \neq \emptyset$

3. $\exists a_1, \dots, a_m \in Z \mid a_1x_1 + \cdots + a_mx_m = 1$. Say $a_1, \dots, a_k > 0$, $a_{k+1}, \dots, a_m < 0$. Define $b = a_1x_1 + \cdots + a_kx_k = (-a_{k+1})x_{k+1} + \cdots + (-a_m)x_m + 1$. Obviously, $b, b-1 \in S$, we can infer that $2b \in S, 2b-1 \in S, 2b-2 \in S$. In general, we have $(kb-k)$ to $kb \in S$. If k is larger enough (i.e., $k > b$), all consecutive numbers are covered in the range.

□

Now we show the proof to theorem 1.9.

Proof. We use a coupling argument. Let X_n be Markov(λ, P), Y_n be Markov(π, P), and they are independent. Let T be a reference point: $T = inf\{n \geq 1 \mid X_n = Y_n\}$.

1. Define $Z_n = (x_n, y_n)$, $(Z_n)_{n\geq 0}$ is Markov($\lambda\pi, \bar{P}$) with state space $I \times I$. We have

$$P[Z_0 = (i,j)] = P[x_0 = i]P[y_0 = j] = \lambda_i \pi_j$$

$$\bar{P}_{(i,j)(k,l)} = P_{ik}P_{jl}$$

   We want to show $Prob[T < \infty] = 1$.

2. $\bar{P}$ is irreducible and positive recurrent. (proof omitted.)

3. From Lemma1.7, we know $prob[T_{kk} < \infty] = 1, \forall k \in I \Rightarrow prob[T < \infty] = 1$

4. To show that x_n, y_n have the "same" distribution after T. Note that x_n, y_n have the same transition matrix P. We have

   (a)

$$\begin{aligned} P[x_n = i, T \leq n] &= \sum_{m=1}^{n} \sum_{j \in I} P[x_n = i, x_m = j, T = m] \\ &= \sum_{m=1}^{n} \sum_{j \in I} P[x_m = j, T = m] P[x_n = i \mid x_m = j] \\ &= \sum_{m=1}^{n} \sum_{j \in I} P[y_m = j, T = m] P[y_n = i \mid y_m = j] \\ &= P[y_n = i, T \leq n] \end{aligned}$$

   (b)

$$\begin{aligned} P[x_n = i] &= P[x_n = i, T \leq n] + P[x_n = i, T > n] \\ &= P[y_n = i, T \leq n] + P[x_n = i, T > n] \\ &\leq P[y_n = i] + P[x_n = i, T > n] \end{aligned}$$

   Based on the same arguments, we have

$$P[y_n = i] \leq P[x_n = i] + P[y_n = i, T > n]$$

   Hence,

$$|P[x_n = i] - P[y_n = i]| \leq P[x_n = i, T > n] + P[y_n = i, T > n]$$

   Summing up on both sides,

$$\sum_{i \in I} |P[x_n = i] - P[y_n = i]| \leq 2P[T > n]$$

   Taking limits as n tends to infinity, the second term on the left tends to π_i, the term on the right hand side tends to 0. Hence, we have

$$\lim_{n \to \infty} P[x_n = i] = \pi_i$$

□

Theorem 1.11 (Ergodic Theorem). *Let P be irreducible and let X_n be Markov(λ, P), then*

$$P[\frac{V_i(n)}{n} \rightarrow \frac{1}{\mu_{ii}} \text{ as } n \rightarrow \infty] = 1$$

where $V_i(n) = \sum_{m=0}^{n-1} 1_{\{x_m = i\}}$.