Recap: Markov Chain

- Let $X_n, \ n = 0, 1, \cdots$ be a sequence of discrete random variables taking values in S.
- We say it is a Markov chain if

$$Pr[X_{n+1} = x_{n+1} | X_n = x_n, X_{n-1} = x_{n-1} \cdots X_0 = x_0] = Pr[X_{n+1} = x_{n+1} | X_n = x_n],$$

- We can write it as

$$f_{X_{n+1}|X_n,\cdots X_0}(x_{n+1}|x_n, \cdots, x_0) = f_{X_{n+1}|X_n}(x_{n+1}|x_n), \ \forall x_i$$

- For a Markov chain, given the current state, the future evolution is independent of the history of how you reached the current state

Recap: Transition Probabilities

- Transition probabilities: $P(x,y) = Pr[X_{n+1} = y | X_n = x]$
  Chain is homogeneous:
  $Pr[X_{n+1} = y | X_n = x] = Pr[X_1 = y | X_0 = x], \ \forall n$
- Initial probabilities $\pi_0(x) = Pr[X_0 = x]$
- Similarly, $\pi_n(x) = Pr[X_n = x]$

Recap: Chapman-Kolmogorov Equations

- n-step transition probabilities:
  $P^n(x, y) = Pr[X_n = y | X_0 = x]$
- These satisfy Chapman-Kolmogorov equations:

$$P^{m+n}(x, y) = \sum_z P^m(x, z) P^n(z, y)$$

- For a finite chain, the n-step transition probability matrix is n-fold product of the transition probability matrix

Recap: transient and recurrent states

- Hitting time for y: $T_y = \min\{n > 0 \ : \ X_n = y\}$
- $\rho_{xy} = P_x(T_y < \infty)$.
- A state y is called transient if $\rho_{yy} < 1$; it is called recurrent if $\rho_{yy} = 1$.
- $N(y)$ – total number of visits to y
- $G(x, y) = E_x[N(y)]$

Recap

Theorem:
(i). Let y be transient. Then

$$P_x(N(y) < \infty) = 1, \ \forall x \ \text{ and } \ G(x,y) = \frac{\rho_{xy}}{1 - \rho_{yy}} < \infty, \ \forall x$$

(ii) Let y be recurrent. Then

$$P_y[N(y) = \infty] = 1, \ \text{ and } \ G(y,y) = E_y[N(y)] = \infty$$

$$P_x[N(y) = \infty] = \rho_{xy}, \ \text{ and } \ G(x,y) = \begin{cases} 0 & \text{if } \ \rho_{xy} = 0 \\ \infty & \text{if } \ \rho_{xy} > 0 \end{cases}$$

Recap

- Transient states are visited only finitely many times while recurrent states are visited infinitely often
- A finite chain should have at least one recurrent state
- We say, x leads to y if $\rho_{xy} > 0$
  Theorem: If x is recurrent and x leads to y then y is recurrent and $\rho_{xy} = \rho_{yx} = 1$.

Recap: closed and irreducible sets

- A set of states, $C \subset S$ is said to be irreducible if x leads to y for all $x, y \in C$
- An irreducible set is also called a communicating class
- A set of states, $C \subset S$, is said to be closed if $x \in C,\ y \notin C$ implies $\rho_{xy} = 0$.
- Once the chain visits a state in a closed set, it cannot leave that set.

Recap: Partition of state space

- $S = S_T + S_R$, transient and recurrent states and

$$S_R = C_1 + C_2 + \cdots$$

  where C_i are closed and irreducible
- We can calculate absorption probabilities for C_i using

$$\rho_C(x) = \sum_{y \in C} P(x,y) + \sum_{y \in S_T} P(x,y)\ \rho_C(y)$$

Recap: Stationary distribution

- π is said to be a stationary distribution for the Markov chain with transition probabilities P if

$$\pi(y) = \sum_{x \in S} \pi(x) P(x, y), \quad \forall y \in S$$

- For finite chains, $P^T \pi = \pi$
- When π is stationary distribution,
  $\pi_0 = \pi \;\; \Rightarrow \;\; \pi_n = \pi, \; \forall n$
- If $\pi_n = \pi, \; \forall n$ then π is a stationary distribution
- For a finite chain, a stationary distribution always exists.
- The stationary distribution, when it exists, is related to expected fraction of time spent in different states.

- Let $I_y(X_n)$ be indicator of $[X_n = y]$
- Number of visits to y till n: $N_n(y) = \sum_{m=1}^{n} I_y(X_m)$

$$G_n(x, y) \triangleq E_x[N_n(y)] = \sum_{m=1}^{n} E_x[I_y(X_m)] = \sum_{m=1}^{n} P^m(x, y)$$

- Expected fraction of time spent in y till n is

$$\frac{G_n(x, y)}{n} = \frac{1}{n} \sum_{m=1}^{n} P^m(x, y)$$

- We will first establish a limit for the above as $n \to \infty$

- Suppose y is transient. Then we have

$$\lim_{n\to\infty} N_n(y) = N(y)$$

$$\text{and} \quad Pr[N(y) < \infty] = 1 \qquad \lim_{n\to\infty} G_n(x,y) = G(x,y) < \infty$$

$$\Rightarrow \qquad \lim_{n\to\infty} \frac{N_n(y)}{n} = 0 \ (w.p.1) \quad \text{and} \quad \lim_{n\to\infty} \frac{G_n(x,y)}{n} = 0$$

- The expected fraction of time spent in a transient state is zero.
- This is intuitively obvious

- Now, let y be recurrent
- Then, $P_y[T_y < \infty] = 1$
- Define $m_y = E_y[T_y]$
- m_y is mean return time to y
- We will show that $\frac{N_n(y)}{n}$ converges to $\frac{1}{m_y}$ if the chain starts in y.
- Convergence would be with probability one.

- Consider a chain started in y
- let T_y^r be time of r^{th} visit to y, $r \geq 1$

$$T_y^r = \min\{n \geq 1 \ : \ N_n(y) = r\}$$

- Define $W_y^1 = T_y^1 = T_y$ and $W_y^r = T_y^r - T_y^{r-1}, \ \ r > 1$
- Note that $E_y[W_y^1] = E_y[T_y] = m_y$
- Also, $T_y^r = W_y^1 + \cdots + W_y^r$
- W_y^r are the "waiting times"
- By Markovian property we should expect them to be iid
- We will prove this.
- Then T_y^r/r converges to m_y by law of large umbers

- We have

$$Pr[W_y^3 = k_3 | W_y^2 = k_2, W_y^1 = k_1] \;\; =$$
$$Pr[X_{k_1+k_2+j} \neq y, \; 1 \leq j \leq k_3 - 1, \;\; X_{k_1+k_2+k_3} = y \mid B]$$

where $B = [X_{k_1+k_2} = y, \; X_{k_1} = y, \; X_j \neq y, j < k_1 + k_2, j \neq k_1]$

- Using the Markovian property, we get

$$Pr[W_y^3 = k_3 | W_y^2 = k_2, W_y^1 = k_1] \;\; =$$
$$Pr[X_{k_1+k_2+j} \neq y, \; 1 \leq j \leq k_3 - 1, \;\; X_{k_1+k_2+k_3} = y \mid X_{k_1+k_2} = y]$$
$$= Pr[X_j \neq y, \; 1 \leq j \leq k_3 - 1, \;\; X_{k_3} = y \mid X_0 = y]$$
$$= P_y[W_y^1 = k_3]$$

- In general, we get

$$Pr[W_y^r = k_r \mid W_y^{r-1} = k_{r-1}, \cdots, W_y^1 = k_1] = P_y[W_y^1 = k_r]$$

- This shows the waiting time are iid

$$\begin{aligned}
P_y[W_y^2 = k_2] &= \sum_{k_1} P_y[W_y^2 = k_2 \mid W_y^1 = k_1]\; P_y[W_y^1 = k_1] \\
&= \sum_{k_1} P_y[W_y^1 = k_2]\; P_y[W_y^1 = k_1] \\
&= P_y[W_y^1 = k_2]
\end{aligned}$$

$\Rightarrow$ identically distributed

$$\begin{aligned}
P_y[W_y^2 = k_2,\ W_y^1 = k_1] &= P_y[W_y^2 = k_2 \mid W_y^1 = k_1] P_y[W_y^1 = k_1] \\
&= P_y[W_y^1 = k_2]\; P_y[W_y^1 = k_1] \\
&= P_y[W_y^2 = k_2]\; P_y[W_y^1 = k_1]
\end{aligned}$$

$\Rightarrow$ independent

- We have shown $W_y^r, \ r = 1, 2, \cdots$ are iid
- Since $E[W_y^1] = m_y$, by strong law of large numbers,

$$\lim_{k \to \infty} \frac{T_y^k}{k} = \lim_{k \to \infty} \frac{1}{k} \sum_{r=1}^{k} W_y^r = m_y, \quad (w.p.1)$$

- Note that this is true even if $m_y = \infty$

- For all n such that $N_n(y) \geq 1$, we have

$$T_y^{N_n(y)} \leq n < T_y^{N_n(y)+1}$$

- $N_n(y)$ is the number of visits to y till time step n
- Suppose $N_{50}(y) = 8$ – Visited y 8 times till time 50.
- So, the 8^{th} visit occurred at or before time 50.
- The 9^{th} visit has not occurred till 50.
- So, time of 9^{th} visit is beyond 50.

$$T_y^{N_n(y)} \leq n < T_y^{N_n(y)+1}$$

- Now we have

$$\frac{T_y^{N_n(y)}}{N_n(y)} \leq \frac{n}{N_n(y)} < \frac{T_y^{N_n(y)+1}}{N_n(y)}$$

- We know that
  - As $n \to \infty$, $N_n(y) \to \infty$, $w.p.1$
  - As $n \to \infty$, $\frac{T_y^n}{n} \to m_y$, $w.p.1$
- Hence we get

$$\lim_{n\to\infty} \frac{n}{N_n(y)} = m_y, \quad w.p.1$$

or

$$\lim_{n\to\infty} \frac{N_n(y)}{n} = \frac{1}{m_y}, \quad w.p.1$$

- All this is true if the chain started in y.
- That means it is true if the chain visits y once.
- So, we get

$$\lim_{n\to\infty} \frac{N_n(y)}{n} = \frac{I_{[T_y<\infty]}}{m_y}, \quad w.p.1$$

- Since $0 \leq \frac{N_n(y)}{n} \leq 1$, almost sure convergence implies convergence in mean

$$\lim_{n\to\infty} \frac{G_n(x,y)}{n} = \lim_{n\to\infty} E_x\left[\frac{N_n(y)}{n}\right] = \lim_{n\to\infty} \frac{P_x[T_y<\infty]}{m_y} = \frac{\rho_{xy}}{m_y}$$

- The fraction of time spent in each recurrent state is inversely proportional to the mean recurrence time

- Thus we have proved the following theorem
- **Theorem**:
  Let y be recurrent. Then
  1. $$\lim_{n\to\infty} \frac{N_n(y)}{n} = \frac{I_{[T_y<\infty]}}{m_y}, \quad w.p.1$$
  2. $$\lim_{n\to\infty} \frac{G_n(x,y)}{n} = \frac{\rho_{xy}}{m_y}$$

- The limiting fraction of time spent in a state is inversely proportional to m_y, the mean return time.
- Intuitively, the stationary probability of a state could be the limiting fraction of time spent in that state.
- Thus $\pi(y) = \frac{1}{m_y}$ is a good candidate for stationary distribution.
- We first note that we can have $m_y = \infty$.
  Though $P_y[T_y < \infty] = 1$, we can have $E_y[T_y] = \infty$.
- What if $m_y = \infty, \ \forall y$?
- Does not seem reasonable for a finite chain.
- But for infinite chains??
- Let us characterize y for which $m_y = \infty$

- A recurrent state y is called **null recurrent** if $m_y = \infty$.
- y is called **positive recurrent** if $m_y < \infty$
- We earlier saw that the fraction of time spent in a transient state is zero.
- Suppose y is null recurrent. Then

$$\lim_{n \to \infty} \frac{N_n(y)}{n} = \frac{1}{m_y} = 0$$

- Thus the limiting fraction of time spent by the chain in transient and null recurrent states is zero.

- **Theorem:** Let x be positive recurrent and let x lead to y. Then y is positive recurrent.
  Proof
- Since x is recurrent and x leads to y we know $\exists n_0, n_1$ s.t. $P^{n_0}(x,y) > 0$, $P^{n_1}(y,x) > 0$ and

  $$P^{n_1+m+n_0}(y,y) \geq P^{n_1}(y,x)P^m(x,x)P^{n_0}(x,y), \quad \forall m$$

  Summing the above for $m = 1, 2, \cdots n$ and dividing by n

$$\frac{1}{n}\sum_{m=1}^{n} P^{n_1+m+n_0}(y,y) \geq P^{n_1}(y,x) \;\; \frac{1}{n}\sum_{m=1}^{n} P^m(x,x) \;\; P^{n_0}(x,y), \; \forall n$$

  If we now let $n \to \infty$, the RHS goes to $P^{n_1}(y,x) \; \frac{1}{m_x} \; P^{n_0}(x,y) > 0$.

$$\frac{1}{n}\sum_{m=1}^{n} P^{n_1+m+n_0}(y,y) \geq P^{n_1}(y,x) \;\; \frac{1}{n}\sum_{m=1}^{n} P^{m}(x,x) \;\; P^{n_0}(x,y), \;\forall n$$

- We can write the LHS of above as

$$\begin{aligned}
\frac{1}{n}\sum_{m=1}^{n} P^{n_1+m+n_0}(y,y) &= \frac{1}{n}\sum_{m=1}^{n_1+n+n_0} P^{m}(y,y) - \frac{1}{n}\sum_{m=1}^{n_1+n_0} P^{m}(y,y) \\
&= \frac{n_1+n+n_0}{n}\frac{1}{n_1+n+n_0}\sum_{m=1}^{n_1+n+n_0} P^{m}(y,y) - \frac{1}{n}\sum_{m=1}^{n_1+n_0} P^{m}(y,y)
\end{aligned}$$

$$\Rightarrow \quad \lim_{n\to\infty}\frac{1}{n}\sum_{m=1}^{n} P^{n_1+m+n_0}(y,y) = \frac{1}{m_y}$$

$$\Rightarrow \quad \frac{1}{m_y} \geq P^{n_1}(y,x)\,\frac{1}{m_x}\,P^{n_0}(x,y) > 0$$

which implies y is positive recurrent

- Thus, in a closed irreducible set of recurrent states, if one state is positive recurrent then all are positive recurrent.
- Hence, in the partition: $S_R = C_1 + C_2 + \cdots$, each C_i is either wholly positive recurrent or wholly null recurrent.
- We next show that a finite chain cannot have any null recurrent states.

- Let C be a finite closed set of recurrent states.
- Suppose all states in C are null recurrent. Then

$$\lim_{n \to \infty} \frac{1}{n} \sum_{m=1}^{n} P^m(x, y) = 0, \quad \forall x, y \in C$$

- Since C is closed, $\sum_{y \in C} P^m(x, y) = 1, \ \forall m, \ \forall x \in C$.
- Thus we get

$$1 = \frac{1}{n} \sum_{m=1}^{n} \sum_{y \in C} P^m(x, y) = \sum_{y \in C} \frac{1}{n} \sum_{m=1}^{n} P^m(x, y), \ \forall n$$

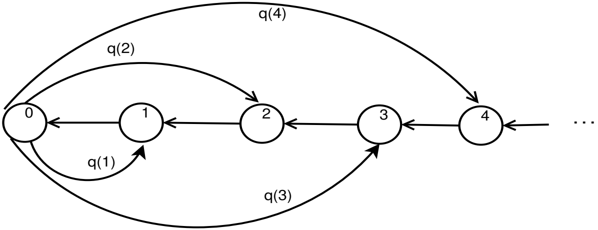
$$\Rightarrow \quad 1 = \lim_{n \to \infty} \sum_{y \in C} \frac{1}{n} \sum_{m=1}^{n} P^m(x, y) = \sum_{y \in C} \lim_{n \to \infty} \frac{1}{n} \sum_{m=1}^{n} P^m(x, y) = 0$$

where we could take the limit inside the sum because C is finite.

- If C is a finite closed set of recurrent states then all states in it cannot be null recurrent.
- Actually what we showed is that any closed finite set must have at least one positive recurrent state.
- Hence, in a finite chain, every closed irreducible set of recurrent states contains only positive recurrent states.
- Hence, a finite chain cannot have a null recurrent state.

Example of null recurrent chain

- Consider the chain with state space $\{0, 1, \cdots\}$ given by



- Here, $q(k) \geq 0, \forall k$ and $\sum_{k=1}^{\infty} q(k) = 1$. We have

$$P_0[T_0 = j+1] = q(j) \quad \Rightarrow m_0 = \sum_{j=2}^{\infty} j \; P_o[T_0 = j] = \sum_{j=2}^{\infty} j \; q(j-1)$$

  (Note that $P_0[T_0 = 1] = 0$)
- So, $m_0 = \infty$ if the $q(\cdot)$ distribution has infinite expectation. For example, if $q(k) = \frac{c}{k^2}$
- Then state 0 is null recurrent. Implies chain is null recurrent

- Suppose π is a stationary distribution.
- Then $\pi(y) = 0$ if y is transient or null recurrent
- We prove this as follows

$$
\begin{aligned}
\pi(y) &= \sum_x \pi(x) P^m(x,y) \;\; \forall m \\
\Rightarrow \;\; \pi(y) &= \frac{1}{n} \sum_{m=1}^{n} \sum_x \pi(x) P^m(x,y) = \sum_x \pi(x) \, \frac{1}{n} \sum_{m=1}^{n} P^m(x,y) \\
\Rightarrow \;\; \pi(y) &= \lim_{n\to\infty} \sum_x \pi(x) \, \frac{1}{n} \sum_{m=1}^{n} P^m(x,y)
\end{aligned}
$$

- The proof is complete if we can take the limit inside the sum

- **Bounded Convergence Theorem**: Suppose $a(x) \geq 0, \ \forall x \in S$ and $\sum_x a(x) < \infty$. Let $b_n(x), \ x \in S$ be such that $|b_n(x)| \leq K, \ \forall x, n$ and suppose $\lim_{n \to \infty} b_n(x) = b(x), \forall x \in S$. Then

$$\lim_{n \to \infty} \sum_{x \in S} a(x) \ b_n(x) = \sum_{x \in S} a(x) \ \lim_{n \to \infty} b_n(x) = \sum_{x \in S} a(x) \ b(x)$$

- We had

$$\pi(y) = \lim_{n \to \infty} \sum_x \pi(x) \ \frac{1}{n} \sum_{m=1}^{n} P^m(x, y)$$

- We have

$$\pi(x) \geq 0; \ \ \sum_x \pi(x) = 1; \ \ 0 \leq \frac{1}{n} \sum_{m=1}^{n} P^m(x, y) \leq 1, \forall x$$

- Hence, if y is transient or null recurrent, then

$$\pi(y) = \sum_x \pi(x) \ \lim_{n \to \infty} \frac{1}{n} \sum_{m=1}^{n} P^m(x, y) = 0$$

- In any stationary distribution π, we would have $\pi(y) = 0$ is y is transient or null recurrent.
- Hence an irreducible transient or null recurrent chain would not have a stationary distribution.

- **Theorem** An irreducible positive recurrent chain has a unique stationary distribution given by

$$\pi(y) = \frac{1}{m_y}, \ \forall y \in S$$

- Suppose $\exists \ \pi$ such that $\pi(y) = \sum_x \pi(x) P(x,y)$. Then

$$\begin{aligned}
\pi(y) &= \sum_x \pi(x) \ P^m(x,y), \ \forall m \\
\Rightarrow \quad \pi(y) &= \sum_x \pi(x) \ \frac{1}{n} \sum_{m=1}^{n} P^m(x,y), \ \forall n \\
\Rightarrow \quad \pi(y) &= \lim_{n\to\infty} \sum_x \pi(x) \ \frac{1}{n} \sum_{m=1}^{n} P^m(x,y) \\
\Rightarrow \quad \pi(y) &= \sum_x \pi(x) \ \lim_{n\to\infty} \ \frac{1}{n} \sum_{m=1}^{n} P^m(x,y) \\
&= \sum_x \pi(x) \ \frac{1}{m_y} = \frac{1}{m_y}
\end{aligned}$$

- To complete the proof, we need to show $\sum_y \frac{1}{m_y} = 1$.
- We also need to show $\frac{1}{m_y} = \sum_x \frac{1}{m_x} \, P(x, y)$
- We skip these steps in the proof.
- The theorem shows that an irreducible positive recurrent chain has a unique stationary distribution
- Corollary: An irreducible chain has a stationary distribution if and only if it is positive recurrent
- An irreducible finite chain has a unique stationary distribution

- If π^1 and π^2 are stationary distributions, then so is $\alpha\pi^1 + (1-\alpha)\pi^2$ (easily verified)
- Let C be a closed irreducible set of positive recurrent states.
  Then there is a unique stationary distribution π that satisfies $\pi(y) = 0, \ \forall y \notin C$.
- Any other stationary distribution of the chain is a convex combination of the stationary distributions concentrated on each of the closed irreducible sets of positive recurrent states.
- This answers all questions about existence and uniqueness of stationary distributions

- Consider an irreducible positive recurrent chain.
- It has a unique stationary distribution and $\frac{1}{n}\sum_{m=1}^{n} P^m(x,y)$ converges to $\pi(y)$.
- The next question is convergence of π_n

$$\lim_{n\to\infty} \pi_n(y) = \lim_{n\to\infty} \sum_x \pi_0(x)\ P^n(x,y) = ?$$

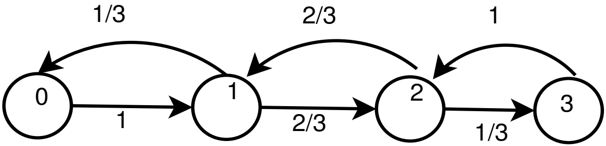
- If $P^n(x,y)$ converges to $g(y)$ then that would be the stationary distribution and π_n converges to it
- But, $\frac{1}{n}\sum_{m=1}^{n} a_m$ may have a limit though $\lim_{n\to\infty} a_n$ may not exist.
  For example, $a_n = (-1)^n$

- Consider a chain with transition probabilities

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1/3 & 0 & 2/3 & 0 \\ 0 & 2/3 & 0 & 1/3 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

- One can show $\pi^T = \left[\frac{1}{8}\ \frac{3}{8}\ \frac{3}{8}\ \frac{1}{8}\right]$
- However, P^n goes to different limits based on whether n is even or odd

- The chain is the folowing



$$P = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1/3 & 0 & 2/3 & 0 \\ 0 & 2/3 & 0 & 1/3 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

- We can return to a state only after even number of time steps
- That is why P^n does not go to a limit
- Such a chain is called a periodic chain

- We define period of a state x as

$$d_x = \gcd\{n \geq 1 \ : \ P^n(x,x) > 0\}$$

- If $P(x,x) > 0$ then $d_x = 1$
- If x leads to y and y leads to x, then $d_x = d_y$
- Let $P^{n_1}(x,y) > 0$, $P^{n_2}(y,x) > 0$. Then $P^{n_1+n_2}(x,x) > 0 \ \Rightarrow \ d_x$ divides $n_1 + n_2$.
- For any n s.t. $P^n(y,y) > 0$, we get $p^{n_1+n+n_2}(x,x) > 0$
- Hence, d_x divides n for all n s.t. $P^n(y,y) > 0 \Rightarrow d_x \leq d_y$
- Similarly, $d_y \leq d_x$ and hence $d_y = d_x$
- All states in an irreducible chain have the same period.
- If the period is 1 then chain is called aperiodic

- The extra condition we need for convergence of π_n is aperiodicity
- For an aperiodic, irreducible, positive recurrent chain, there is a unique stationary distribution and π_n converges to it irrespective of what π_0 is.
- An aperiodic, irreducible, positive recurrent chain is called an ergodic chain