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 probability function is $P : S \times S \rightarrow [0,\ 1]$

$$P(x, y) = Pr[X_{n+1} = y | X_n = x]$$

  The chain is said to be homogeneous when this is not a function of time.
- For a homogeneous chain

$$Pr[X_{n+1} = y | X_n = x] = Pr[X_1 = y | X_0 = x],\ \forall n$$

- P satisfies
  - $P(x, y) \geq 0,\ \forall x, y \in S$
  - $\sum_{y \in S} P(x, y) = 1, \forall x \in S$
- If S is finite then P can be represented as a matrix

Recap: Initial State Probabilities

- Initial state probabilities $\pi_0 : S \to [0,\ 1]$

$$\pi_0(x) = Pr[X_0 = x]$$

  It satisfies
  - $\pi_0(x) \geq 0,\ \forall x \in S$
  - $\sum_{x \in S} \pi_0(x) = 1$
- The P and π_0 together determine all joint distributions:

$$Pr[X_0 = x_0, X_1 = x_1, \cdots, X_m = x_m] = \pi_0(x_0) P(x_0, x_1) \cdots P(x_{m-1}, x_m)$$

  Using this we can find joint distribution of any finite number of X_i's

Recap

- The Markov property implies

$$\begin{aligned} Pr[X_{m+n} = y | X_m = x, X_{m-1} \cdots] &= Pr[X_{m+n} = y | X_m = x] \\ &= Pr[X_n = y | X_0 = x] \end{aligned}$$

- Or, in general,

$$f_{X_{m+n}|X_m,\cdots,X_0}(y|x,\cdots) = f_{X_{m+n}|X_m}(y|x)$$

- Further, we can show

$$\begin{aligned} \Pr[X_{m+n} = y | X_m = x,\ X_{m-k} \in A_k,\ k = 1,\cdots,m] &= \\ Pr[X_{m+n} = y | X_m = x] \end{aligned}$$

$$\begin{aligned} Pr[X_{m+n+r} \in B_r,\ r = 0,\cdots,s \mid X_m = x,\ X_{m-k} \in A_k,\ k = 1,\cdots,m] \\ = Pr[X_{m+n+r} \in B_r,\ r = 0,\cdots,s \mid X_m = x] \end{aligned}$$

- The transition probabilities we defined earlier are also called one step transition probabilities

$$P(x,y) = Pr[X_{n+1} = y | X_n = x] = Pr[X_1 = y | X_0 = x]$$

- We can define transition probabilities for multiple steps, that is, $Pr[X_n = y | X_0 = x]$
- Now we get

$$\begin{aligned}
Pr[X_{m+n} = y | X_0 = x] &= \sum_z Pr[X_{m+n} = y, X_m = z | X_0 = x] \\
&= \sum_z Pr[X_{m+n} = y | X_m = z, X_0 = x] \; Pr[X_m = z | X_0 = x] \\
&= \sum_z Pr[X_{m+n} = y | X_m = z] \; Pr[X_m = z | X_0 = x] \\
&= \sum_z Pr[X_n = y | X_0 = z] \; Pr[X_m = z | X_0 = x]
\end{aligned}$$

Chapman-Kolmogorov Equations

- Define: $P^n(x,y) = Pr[X_n = y | x_0 = x]$
- These are called n-step transition probabilities.
- From what we showed,

$$\begin{aligned} P^{m+n}(x,y) &= Pr[X_{m+n} = y | X_0 = x] \\ &= \sum_z Pr[X_n = y | X_0 = z] \; Pr[X_m = z | X_0 = x] \\ &= \sum_z P^m(x,z) P^n(z,y) \end{aligned}$$

- These are known as Chapman-Kolmogorov equations
- This relationship is intuitively clear

- Specifically, using Chapman-Kolmogorov equations,

$$P^2(x,y) = \sum_z P(x,z)P(z,y)$$

- For a finite chain, P is a matrix
- Thus $P^2(x,y)$ is
  the $(x,y)^{th}$ element of the matrix, $P \times P$
- That is why we use P^n for n-step transition probabilities

- Define: $\pi_n(x) = Pr[X_n = x]$.
- Then we get

$$\begin{aligned} \pi_n(y) &= \sum_x Pr[X_n = y | X_0 = x] \; Pr[X_0 = x] \\ &= \sum_x \pi_0(x) P^n(x, y) \end{aligned}$$

- In particular

$$\begin{aligned} \pi_{n+1}(y) &= \sum_x Pr[X_{n+1} = y | X_n = x] \; Pr[X_n = x] \\ &= \sum_x \pi_n(x) P(x, y) \end{aligned}$$

Hitting times

- Let y be a state.
- We define hitting time for y as the random variable

$$T_y = \min\{n > 0 \ : \ X_n = y\}$$

- T_y is the first time that the chain is in state y (after $t = 0$ when the chain is initiated).
- It is easy to see that $Pr[T_y = 1|X_0 = x] = P(x, y)$.
- Notation: $P_z(A) = Pr[A|X_0 = z]$
- We write the above as $P_x(T_y = 1) = P(x, y)$
  Note that

$$P_x[T_y = n] = Pr[T_y = n|X_0 = x] \neq P^n(x, y)$$

$$T_y = \min\{n > 0 \ : \ X_n = y\}$$

- We can now get

$$
\begin{aligned}
P_x(T_y = 2) &= \sum_{z \neq y} P(x,z)\ P(z,y) = \sum_{z \neq y} P(x,z)\ P_z(T_y = 1) \\
P_x(T_y = m) &= Pr[T_y = m | X_0 = x] \\
&= \sum_{z \neq y} Pr[T_y = m | X_1 = z, X_0 = x]\ Pr[X_1 = z | X_0 = x] \\
&= \sum_{z \neq y} P(x,z) Pr[T_y = m | X_1 = z] \\
&= \sum_{z \neq y} P(x,z) P_z(T_y = m-1)
\end{aligned}
$$

- Similarly we can get:

$$P^n(x,y) = \sum_{m=1}^{n} P_x(T_y = m) P^{n-m}(y,y)$$

- We can derive this as

$$\begin{aligned} P^n(x,y) &= Pr[X_n = y | X_0 = x] \\ &= \sum_{m=1}^{n} Pr[T_y = m, X_n = y | X_0 = x] \\ &= \sum_{m=1}^{n} Pr[X_n = y | T_y = m, X_0 = x] \; Pr[T_y = m | X_0 = x] \\ &= \sum_{m=1}^{n} Pr[X_n = y | X_m = y] \; Pr[T_y = m | X_0 = x] \\ &= \sum_{m=1}^{n} P^{n-m}(y,y) \; P_x(T_y = m) \end{aligned}$$

transient and recurrent states

- Define $\rho_{xy} = P_x(T_y < \infty)$.
- It is the probability that starting in x you will visit y
- Note that

$$\rho_{xy} = \lim_{n\to\infty} P_x(T_y < n) = \lim_{n\to\infty} \sum_{m=1}^{n-1} P_x(T_y = m) = \sum_{m=1}^{\infty} P_x(T_y = m)$$

Definition: A state y is called **transient** if $\rho_{yy} < 1$; it is called **recurrent** if $\rho_{yy} = 1$.

- Intuitively, all transient states would be visited only finitely many times while recurrent states are visited infinitely often.
- For any state y define

$$I_y(X_n) = \begin{cases} 1 & \text{if } X_n = y \\ 0 & \text{otherwise} \end{cases}$$

- Now, the total number of visits to y is given by

$$N_y = \sum_{n=1}^{\infty} I_y(X_n)$$

- We can get distribution of N_y as

$$\begin{aligned}
P_x(N_y \geq 1) &= P_x(T_y < \infty) = \rho_{xy} \\
P_x(N_y \geq 2) &= \sum_m P_x(T_y = m) P_y(T_y < \infty) \\
&= \rho_{yy} \sum_m P_x(T_y = m) = \rho_{yy}\ \rho_{xy} \\
P_x(N_y \geq m) &= \rho_{yy}^{m-1}\ \rho_{xy} \\
P_x(N_y = m) &= P_x(N_y \geq m) - P_x(N_y \geq m+1) \\
&= \rho_{yy}^{m-1}\ \rho_{xy} - \rho_{yy}^{m}\ \rho_{xy} = \rho_{xy}\ \rho_{yy}^{m-1}(1 - \rho_{yy}) \\
P_x(N_y = 0) &= 1 - P_x(N_y \geq 1) = 1 - \rho_{xy}
\end{aligned}$$

- Notation: $E_x[Z] = E[Z|X_0 = x]$
- Define

$$\begin{aligned} G(x,y) &\triangleq E_x[N_y] \\ &= E_x\left[\sum_{n=1}^{\infty} I_y(X_n)\right] \\ &= \sum_{n=1}^{\infty} E_x\left[I_y(X_n)\right] \\ &= \sum_{n=1}^{\infty} P^n(x,y) \end{aligned}$$

- $G(x,y)$ is the expected number of visits to y for a chain that is started in x.

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

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

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

Proof of (i): y is transient; $\rho_{yy} < 1$

$$
\begin{aligned}
G(x,y) &= E_x[N(y)] = \sum_m m P_x[N(y) = m] \\
&= \sum_m m\ \rho_{xy}\ \rho_{yy}^{m-1}(1-\rho_{yy}) \\
&= \rho_{xy} \sum_{m=1}^{\infty} m\ \rho_{yy}^{m-1}(1-\rho_{yy}) \\
&= \rho_{xy}\ \frac{1}{1-\rho_{yy}} < \infty, \quad \text{because} \ \ \rho_{yy} < 1
\end{aligned}
$$

$$
\Rightarrow \ \ P_x[N(y) < \infty] = 1
$$

Proof of (ii):

y recurrent $\Rightarrow \ \ \rho_{yy} = 1$. Hence

$$\begin{aligned} P_y[N(y) \geq m] &= \rho_{yy}^m = 1, \ \ \forall m \\ \Rightarrow \ \ P_y[N(y) = \infty] &= \lim_{m \to \infty} P_y[N(y) \geq m] = 1 \\ \Rightarrow \ \ G(y,y) &= E_y[N(y)] = \infty \end{aligned}$$

$$P_x[N(y) \geq m] = \rho_{xy} \, \rho_{yy}^{m-1} = \rho_{xy}, \ \ \forall m$$

Hence $P_x[N(y) = \infty] = \rho_{xy}$

$$\rho_{xy} = 0 \ \ \Rightarrow \ \ P_x[N(y) \geq m] = 0, \ \forall m > 0 \ \ \Rightarrow \ \ G(x,y) = 0$$

$$\rho_{xy} > 0 \ \ \Rightarrow \ \ P_x[N(y) = \infty] > 0 \ \ \Rightarrow \ \ G(x,y) = \infty$$

- Transient states are visited only finitely many times while recurrent states are visited infinitely often
- If S is finite, it should have at least one recurrent state
- If y is transient, then, for all x

$$G(x,y) = \sum_{n=1}^{\infty} P^n(x,y) < \infty \quad \Rightarrow \quad \lim_{n\to\infty} P^n(x,y) = 0$$

- However, $\sum_y P^n(x,y) = 1, \ \forall n, \ \forall x$
- If all $y \in S$ are transient, then we get a contradiction

$$1 = \lim_{n\to\infty} \sum_{y\in S} P^n(x,y) = \sum_{y\in S} \lim_{n\to\infty} P^n(x,y) = 0$$

- A finite chain has to have at least one recurrent state
- An infinite chain can have only transient states

- 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 $\rho_{xy} = \rho_{yx} = 1$.

Proof:

- Take $x \neq y$, wlog. Since $\rho_{xy} > 0$, $\exists n$ s.t. $P^n(x, y) > 0$
- Take least such n. Then we have states $y_1, \cdots, y_{n-1}$, none of which is x (or y) such that

$$P(x, y_1)\ P(y_1, y_2) \cdots P(y_{n-1}, y) > 0$$

- Now suppose, $\rho_{yx} < 1$. Then

$$P(x, y_1)\ P(y_1, y_2) \cdots P(y_{n-1}, y)(1 - \rho_{yx}) > 0$$

is the probability of starting in x but not returning to x.

- But this cannot be because x is recurrent and hence $\rho_{xx} = 1$
- Hence, if x is recurrent and x leads to y then $\rho_{yx} = 1$

- Now, $\exists n_0, n_1$ s.t. $P^{n_0}(x,y) > 0$, $P^{n_1}(y,x) > 0$.

$$\begin{aligned} P^{n_1+n+n_0}(y,y) &= P_y[X_{n_1+n+n_0} = y] \\ &\geq P_y[X_{n_1} = x, X_{n_1+n} = x, X_{n_1+n+n_0} = y] \\ &= P^{n_1}(y,x)P^n(x,x)P^{n_0}(x,y), \quad \forall n \end{aligned}$$

- We know $G(x,x) = \sum_{m=1}^{\infty} P^m(x,x) = \infty$

$$\begin{aligned} \sum_{m=1}^{\infty} P^m(y,y) &\geq \sum_{m=n_0+n_1+1}^{\infty} P^m(y,y) = \sum_{n=1}^{\infty} P^{n_1+n+n_0}(y,y) \\ &\geq \sum_{n=1}^{\infty} P^{n_1}(y,x)P^n(x,x)P^{n_0}(x,y) \\ &= \infty, \quad \text{because } x \text{ is recurrent} \\ &\Rightarrow y \text{ is recurrent} \end{aligned}$$

- What we showed so far is: if x leads to y and x is recurrent, then $\rho_{yx} = 1$ and y is recurrent.
- Now, y is recurrent and y leads to x and hence $\rho_{xy} = 1$.
- This completes proof of the theorem

equivalence relation

- let R be a relation on set A. Note $R \subset A \times A$
- R is called an equivalence relation if it is
  1. reflexive, i.e., $(x, x) \in R, \; \forall x \in A$
  2. symmetric, i.e., $(x, y) \in R \; \Rightarrow \; (y, x) \in R$
  3. transitive, i.e., $(x, y), (y, z) \in R \; \Rightarrow \; (x, z) \in R$

example

- Let $A = \{\frac{m}{n} \mid m, n \text{ are integers}\}$
- Define relation R by

$$\left(\frac{m}{n},\ \frac{p}{q}\right) \in R \text{ if } mq = np$$

- This is the usual equality of fractions
- Easy to check it is an equivalence relation.

Equivalence classes

- Let R be an equivalence relation on A.
- Then, A can be partitioned as

$$A = C_1 + C_2 + \cdots$$

  Where C_i satisfy
  - $x, y \in C_i \;\; \Rightarrow \;\; (x, y) \in R, \; \forall i$
  - $x \in C_i, \; y \in C_j, \; i \neq j \;\; \Rightarrow \;\; (x, y) \notin R$
- In our example, each equivalence class corresponds to a rational number.
- Here, C_i contains all fractions that are equal to that rational number

- The state space of any Markov chain can be partitioned into the transient and recurrent states: $S = S_T + S_R$:

$$S_T = \{y \in S \ : \ \rho_{yy} < 1\} \qquad S_R = \{y \in S \ : \ \rho_{yy} = 1\}$$

- On S_R, consider the relation: 'x leads to y' (i.e., x is related to y if $\rho_{xy} > 0$)
- This is an equivalence relation
  - $\rho_{xx} > 0, \ \forall x \in S_R$
  - $\rho_{xy} > 0 \ \Rightarrow \ \rho_{yx} > 0, \ \forall x, y \in S_R$
  - $\rho_{xy} > 0, \ \rho_{yz} > 0 \ \Rightarrow \ \rho_{xz} > 0$
- Hence we get a partition: $S_R = C_1 + C_2 + \cdots$ where C_i are equivalence classes.

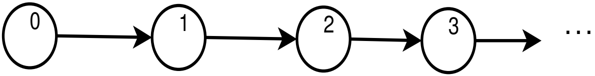
- On S_R, "x leads to y" is an equivalence relation.
- This gives rise to the partition $S_R = C_1 + C_2 + \cdots$
- Since C_i are equivalence classes, they satisfy:
    - $x, y \in C_i \;\;\Rightarrow\;\; x$ leads to y
    - $x \in C_i, y \in C_j, i \neq j \;\;\Rightarrow\;\; \rho_{xy} = 0$
- All states in any C_i lead to each other or communicate with each other
- If $i \neq j$ and $x \in C_i$ and $y \in C_j$, then, $\rho_{xy} = \rho_{yx} = 0$. x and y do not communicate with each other.

- 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.
- We get a partition of recurrent states

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

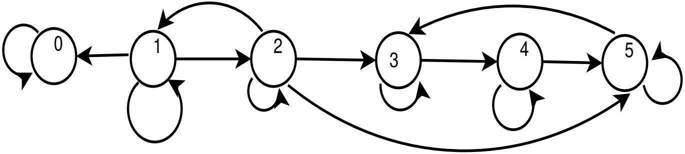
  where each C_i is a closed and irreducible set of states.
- If S is irreducible then the chain is said to be irreducible. (Note that S is trivially closed)

- In an irreducible set of states, if one state is recurrent, then, all states are recurrent.
- We saw that a finite chain has to have at least one recurrent state.
- Thus, a finite irreducible chain is recurrent.
- For example, in the umbrellas problem, the chain is irreducible and hence all states are recurrent.
- An infinite irreducible chain may be wholly transient
- Here is a trivial example of non-irreducible transient chain:



- The state space of any Markov chain can be partitioned into transient and recurrent states.
- We need not calculate ρ_{xx} to do this partition.
- By looking at the structure of transition probability matrix we can get this partition

Example



$$\begin{array}{c|cccccc} & 0 & 1 & 2 & 3 & 4 & 5 \\ \hline 0 & + & - & - & - & - & - \\ 1 & + & + & + & - & - & - \\ 2 & - & + & + & + & - & + \\ 3 & - & - & - & + & + & - \\ 4 & - & - & - & - & + & + \\ 5 & - & - & - & + & - & + \end{array}$$

- State 0 is called an absorbing state. $\{0\}$ is a closed irreducible set.
- $1, 2$ are transient states.
- We get: $S_T = \{1, 2\}$ and $S_R = \{0\} + \{3, 4, 5\}$