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 $\pi_0 : S \rightarrow [0,\ 1]$

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

  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

Recap

- The Markov property implies

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

- 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

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

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

Recap: Chapman-Kolmogorov Equations

- The n-step transition probabilities are defined by

$$P^n(x, y) = Pr[X_n = y | X_0 = x]$$

- These n-step transition probabilities satisfy

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

- These are known as Chapman-Kolmogorov equations
- For a finite chain, the n-step transition probability matrix is n-fold product of the transition probability matrix
- We also have

$$\pi_n(x) \triangleq Pr[X_n = x] = \sum_x \pi_0(x) P^n(x, y)$$

Recap: Hitting times

- We define hitting time for y as the random variable

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

- Using this defintion, we can derive

$$P_x(T_y = m) = \sum_{z \neq y} P(x, z) P_z(T_y = m - 1)$$

(Notation: $P_z(A) = Pr[A | X_0 = z]$)

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

Recap: 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

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

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.

Recap

- For any state y define

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

- 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

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

and $P_x(N(y) = 0) = 1 - \rho_{xy}$

Recap

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

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

- $G(x,y)$ 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; $\rho_{yy} < 1$

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

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

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.

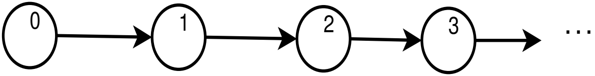
- On S_R, "x leads to y" is an equivalence relation.
- This gives rise to the partition $S_R = C_1 + C_2 + \cdots$
- Since C_i are equivalence classes, they satisfy:
  - $x, y \in C_i \;\; \Rightarrow \;\; x$ leads to y
  - $x \in C_i, y \in C_j, i \neq j \;\; \Rightarrow \;\; \rho_{xy} = 0$
- All states 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

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

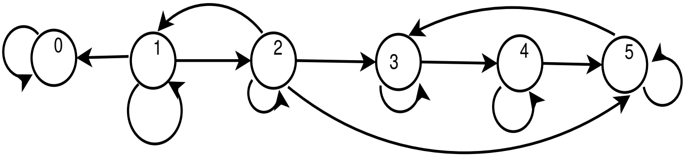
  where each C_i is a closed and irreducible set of states.
- If S is irreducible then the chain is said to be irreducible. (Note that S is trivially closed)

- 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 $S_R = \{0\} + \{3, 4, 5\}$

- If you start the chain in a recurrent state it will stay in the corresponding closed irreducible set
- If you start in one of the transient states, it would eventually get 'absorbed' in one of the closed irreducible sets of recurrent states.
- We want to know the probabilities of ending up in different sets.
- We want to know how long you stay in transient states
- We want to know what is the 'steady state'?

- let C be a closed irreducible set of recurrent states
- T_C – hitting time for C.
  $T_C = \min\{n > 0 \ : \ X_n \in C\}$
  It is the first time instant when the chain is in C
- Define $\rho_C(x) = P_x[T_C < \infty]$

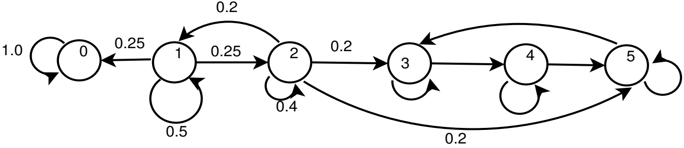
$$\text{If } x \text{ is recurrent, } \ \rho_C(x) = \begin{cases} 1 & \text{if } x \in C \\ 0 & \text{if } x \notin C \end{cases}$$

  Because each x is in a closed irreducible set
- Suppose x is transient. Then

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

- By solving this set of linear equations we can get $\rho_c(x), \ x \in S_T$

Example: Absorption probabilities



- $S_T = \{1, 2\}$ and $C_1 = \{0\}$, $C_2 = \{3, 4, 5\}$

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

$$\begin{aligned}
\rho_{C_1}(1) &= P(1,0) + P(1,1)\rho_{C_1}(1) + P(1,2)\rho_{C_1}(2) \\
&= 0.25 + 0.5\rho_{C_1}(1) + 0.25\rho_{C_1}(2) \\
\rho_{C_1}(2) &= 0 + 0.2\rho_{C_1}(1) + 0.4\rho_{C_1}(2)
\end{aligned}$$

- Solving these, we get $\rho_{C_1}(1) = 0.6, \quad \rho_{C_1}(2) = 0.2$
- What would be $\rho_{C_2}(1)$?

Expected time in transient states

- We consider a simple method to get the time spent in transient states for finite chains
- Let states $1, 2, \cdots, t$ be the transient states
- b_{ij} – the expected number of time instants spent in state j when started in i.
- Then we get

$$b_{ij} = \delta_{ij} + \sum_{k=1}^{t} P(i,k) b_{kj}$$

  where $\delta_{ij} = 1$ if $i = j$ and is zero otherwise
- let B be the $t \times t$ matrix of b_{ij}, I be the $t \times t$ identity matrix and P_T be the submatrix (corresponding to the transient states) of P.
- Then the above in Matrix notation is

$$B = I + P_T B \quad \text{or} \quad B = (I - P_T)^{-1}$$

stationary distributions

- $\pi : S \to [0,\ 1]$ is a probability distribution (mass function) over S if $\pi(x) \geq 0,\ \forall x$ and $\sum_{x \in S} \pi(x) = 1$
- A probability distribution over S, π, is said to be a stationary distribution for the Markov chain with transition probabilities P if

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

- Suppose S is finite. Then π can be represented by a vector.
- The π is stationary if

$$\pi^T = \pi^T\ P \quad \text{or} \quad P^T \pi = \pi$$

- π is a stationary distribution if

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

- Recall $\pi_n(x) \triangleq Pr[X_n = x]$ satisfies

$$\pi_{n+1}(y) = \sum_{x \in S} Pr[X_{n+1} = y | X_n = x] Pr[X_n = x] = \sum_{x \in S} \pi_n(x) P(x, y)$$

- Hence, if $\pi_0 = \pi$ then $\pi_1 = \pi$
  and hence $\pi_n = \pi, \ \forall n$
- Hence the name, stationary distribution.
- It is also called the invariant distribution or the invariant measure

- If the chain is started in stationary distribution then the distribution of X_n is not a function of time, as we saw.
- Suppose for a chain, distribution of X_n is not dependent on n. Then the chain must be in a stationary distribution.
- Suppose $\pi = \pi_0 = \pi_1 = \cdots = \pi_n = \cdots$. Then

$$\pi(y) = \pi_1(y) = \sum_{x \in S} \pi_0(x) P(x, y) = \sum_{x \in S} \pi(x) P(x, y)$$

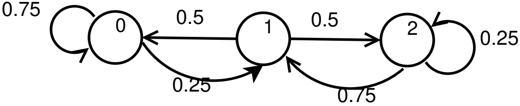
  which shows π is a stationary distribution

- Suppose S is finite.
- Then π is a stationary distribution if

$$P^T \pi = \pi \quad \text{or} \quad \left(P^T - I\right)\, \pi = 0$$

- Note that each column of P^T sums to 1.
- Hence, $\left(P^T - I\right)$ would be singular
  (1 is always an eigen value for a column stochastic matrix)
- A stationary distribution always exists for a finite chain.
- But it may or may not be unique.
- What about infinite chains?

Example



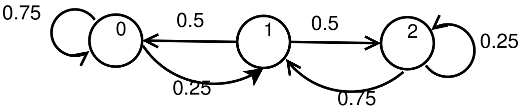
- The stationary distribution has to satisfy

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

- Thus we get the following linear equations

$$\begin{aligned} 0.75\pi(0) + 0.5\pi(1) &= \pi(0) \\ 0.25\pi(0) + 0.75\pi(2) &= \pi(1) \\ 0.5\pi(1) + 0.25\pi(2) &= \pi(2) \end{aligned}$$

in addition, $\quad \pi(0) + \pi(1) + \pi(2) = 1$

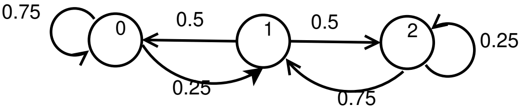


- We can also write the equations for π as

$$\begin{bmatrix} \pi(0) & \pi(1) & \pi(2) \end{bmatrix} \begin{bmatrix} 0.75 & 0.25 & 0 \\ 0.5 & 0 & 0.5 \\ 0 & 0.75 & 0.25 \end{bmatrix} = \begin{bmatrix} \pi(0) & \pi(1) & \pi(2) \end{bmatrix}$$

$$\begin{aligned} 0.75\pi(0) + 0.5\pi(1) &= \pi(0) \\ 0.25\pi(0) + 0.75\pi(2) &= \pi(1) \\ 0.5\pi(1) + 0.25\pi(2) &= \pi(2) \end{aligned}$$

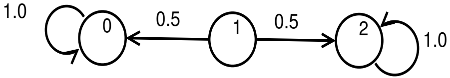
- We have to solve these along with $\pi(0) + \pi(1) + \pi(2) = 1$



$$
\begin{aligned}
0.75\pi(0) + 0.5\pi(1) &= \pi(0) \;\Rightarrow\; \pi(1) = \frac{1}{2}\,\pi(0) \\
0.25\pi(0) + 0.75\pi(2) &= \pi(1) \;\Rightarrow\; \pi(2) = \frac{1}{3}\pi(0) \\
0.5\pi(1) + 0.25\pi(2) &= \pi(2) \\
\pi(0) + \pi(1) + \pi(2) &= 1 \;\Rightarrow\; \pi(0)\left(1 + \frac{1}{2} + \frac{1}{3}\right) = 1
\end{aligned}
$$

- Now, $\pi(0)\left(1 + \frac{1}{2} + \frac{1}{3}\right) = 1$ gives $\pi(0) = \frac{6}{11}$
- We get a unique solution: $\left[\frac{6}{11} \;\; \frac{3}{11} \;\; \frac{2}{11}\right]$

Example2

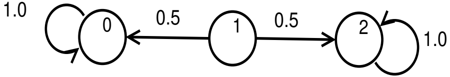


- The stationary distribution has to satisfy

$$\begin{bmatrix} \pi(0) & \pi(1) & \pi(2) \end{bmatrix} \begin{bmatrix} 1.0 & 0 & 0 \\ 0.5 & 0 & 0.5 \\ 0 & 0 & 1.0 \end{bmatrix} = \begin{bmatrix} \pi(0) & \pi(1) & \pi(2) \end{bmatrix}$$

- We also have to add the equation $\pi(0) + \pi(1) + \pi(2) = 1$
- We now do not have a unique stationary distribution

Example2



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

- We get the following linear equations

$$\begin{aligned}
\pi(0) + 0.5\pi(1) &= \pi(0) \Rightarrow \pi(1) = 0 \\
0.5\pi(1) + \pi(2) &= \pi(2) \Rightarrow \pi(1) = 0 \\
\pi(0) + \pi(1) + \pi(2) &= 1 \Rightarrow \pi(0) = 1 - \pi(2)
\end{aligned}$$

- Now there are infinitely many solutions.
- Any distribution $[a \;\; 0 \;\; 1-a]$ with $0 \leq a \leq 1$ is a stationary distribution
- This chain is not irreducible; the previous one is irreducible

- We now explore conditions for existence and uniqueness of stationary distributions
- For finite chains stationary distribution always exists.
- For finite irreducible chains it is unique.
- But for infinite chains, it is possible that stationary distribution does not exist.
- When the stationary distribution is unique, we also want to know if the chain converges to that distribution
- 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: $N_n(y) = \sum_{m=1}^{n} I_y(X_n)$

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

- 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 $n \to \infty$