Recap: Markov Chain

- $X_n, \; n = 0, 1, \cdots,$ (with X_i discrete) is a Markov chain if

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

Recap: Markov Chain

- $X_n, \ n = 0, 1, \cdots,$ (with X_i discrete) 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

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

Recap: Markov Chain

- $X_n, \; n = 0, 1, \cdots,$ (with X_i discrete) 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: $P(x, y) = Pr[X_{n+1} = y | X_n = x]$

Recap: Transition Probabilities

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

Recap: Transition Probabilities

- Transition probabilities: $P(x, y) = Pr[X_{n+1} = y | X_n = x]$
  Chain is homogeneous if
  $Pr[X_{n+1} = y | X_n = x] = Pr[X_1 = y | X_0 = x], \ \forall n$
- P satisfies
  - $P(x, y) \geq 0, \ \forall x, y \in S$
  - $\sum_{y \in S} P(x, y) = 1, \forall x \in S$

Recap: Transition Probabilities

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

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

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

Recap: Distributions of X_n

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

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

  It satisfies

  - $\pi_0(x) \geq 0,\ \forall x \in S$

Recap: Distributions of X_n

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

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

It satisfies

  - $\pi_0(x) \geq 0,\ \forall x \in S$
  - $\sum_{x \in S} \pi_0(x) = 1$

Recap: Distributions of X_n

- 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

Recap: Distributions of X_n

- 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
- Similarly, $\pi_n(x) = Pr[X_n = x]$

Recap: Distributions of X_n

- 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
- Similarly, $\pi_n(x) = Pr[X_n = x]$

$$\pi_{n+1}(y) = \sum_x \pi_n(x) P(x, y)$$

Recap: Chapman-Kolmogorov Equations

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

Recap: Chapman-Kolmogorov Equations

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

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

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 (1-step) transition probability matrix

Recap: Hitting times

- Hitting time for y: $\quad T_y = \min\{n > 0 \ : \ X_n = y\}$

Recap: Hitting times

- Hitting time for y: $\quad T_y = \min\{n > 0 \;:\; X_n = y\}$
- We have

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

Recap: Hitting times

- Hitting time for y: $T_y = \min\{n > 0 \ : \ X_n = y\}$
- We have

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

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

Recap: Hitting times

- Hitting time for y: $T_y = \min\{n > 0 \ : \ X_n = y\}$
- We have

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

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

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

Recap: transient and recurrent states

- Define $\rho_{xy} = P_x(T_y < \infty)$.

Recap: transient and recurrent states

- Define $\rho_{xy} = P_x(T_y < \infty)$.
- Note that

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

Recap: transient and recurrent states

- Define $\rho_{xy} = P_x(T_y < \infty)$.
- Note that

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

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

Recap: transient and recurrent states

- Define $\rho_{xy} = P_x(T_y < \infty)$.
- Note that

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

- 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: Number of visits to a state

- $I_y(X_n)$ is indicator rv of $[X_n = y]$

Recap: Number of visits to a state

- $I_y(X_n)$ is indicator rv of $[X_n = y]$
- The total number of visits to y: $\quad N(y) = \sum_{n=1}^{\infty} I_y(X_n)$

Recap: Number of visits to a state

- $I_y(X_n)$ is indicator rv of $[X_n = y]$
- The total number of visits to y: $\quad N(y) = \sum_{n=1}^{\infty} I_y(X_n)$
- Distribution of $N(y)$:

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

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

Recap: Number of visits to a state

- $I_y(X_n)$ is indicator rv of $[X_n = y]$
- The total number of visits to y: $N(y) = \sum_{n=1}^{\infty} I_y(X_n)$
- Distribution of $N(y)$:

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

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

- Expected number of visits:

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

(Notation: $E_x[Z] = E[Z|X_0 = x]$)

Recap

Theorem:

Recap

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

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

Recap

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

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

(ii) Let y be recurrent. Then

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

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

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

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 $\rho_{xy} > 0$

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

- 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 $\rho_{xy} = \rho_{yx} = 1$.
- On the set of recurrent states, 'leads to' is an equivalence relation

Recap: closed and irreducible sets

- A set of states, $C \subset S$ is said to be irreducible if x leads to y for all $x, y \in C$

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

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 $\rho_{xy} = 0$.

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

Recap: Partition of state space

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

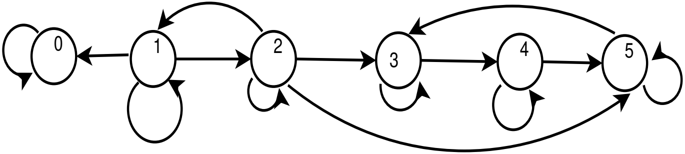
$$S_R = C_1 + C_2 + \cdots$$

where C_i are closed and irreducible

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
- Eventually the chain spends all its time in one of the C_i

Recap: Example of partition of state space



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

- $1, 2$ 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 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.

- 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.

- 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

- 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

- Let C be a closed irreducible set of recurrent states
- T_C – hitting time for C.

- Let C be a closed irreducible set of recurrent states
- T_C – hitting time for C.
  $T_C = \min\{n > 0 \ : \ X_n \in C\}$

- 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

- 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 $\rho_C(x) = P_x[T_C < \infty]$

- 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 $\rho_C(x) = P_x[T_C < \infty]$

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

- 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

- 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

- 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

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

- 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

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

- 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 $\rho_C(x) = P_x[T_C < \infty]$

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

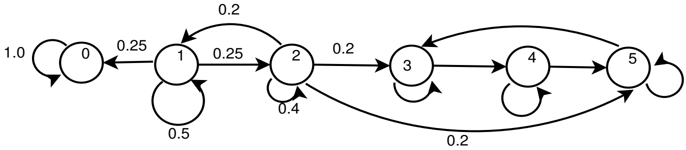
  Because each x is in a closed irreducible set
- Suppose x is transient. Then

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

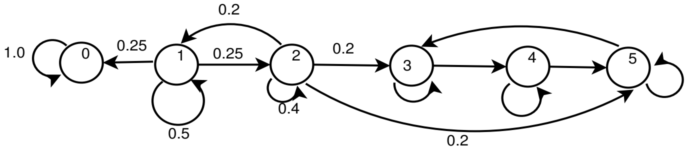
- By solving this set of linear equations we can get $\rho_c(x), \; x \in S_T$

Example: Absorption probabilities

Example: Absorption probabilities

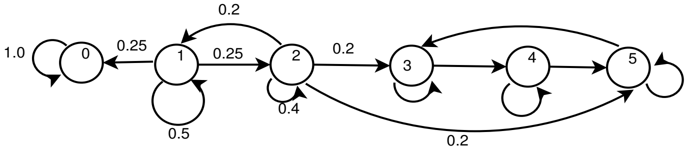


Example: Absorption probabilities



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

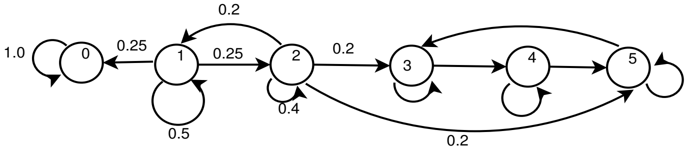
Example: Absorption probabilities



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

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

Example: Absorption probabilities

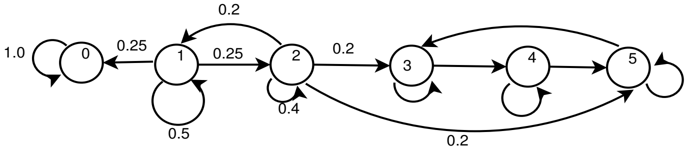


- $S_T = \{1, 2\}$ and $C_1 = \{0\}$, $C_2 = \{3, 4, 5\}$

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

$$\rho_{C_1}(1) \quad = \quad P(1, 0) + P(1, 1)\rho_{C_1}(1) + P(1, 2)\rho_{C_1}(2)$$

Example: Absorption probabilities

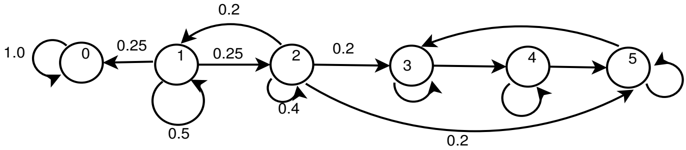


- $S_T = \{1, 2\}$ and $C_1 = \{0\}$, $C_2 = \{3, 4, 5\}$

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

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

Example: Absorption probabilities

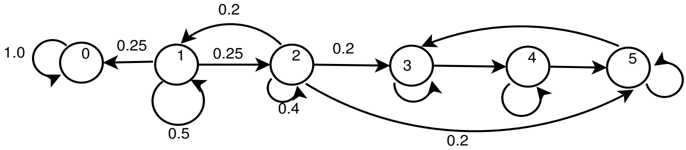


- $S_T = \{1,2\}$ and $C_1 = \{0\}$, $C_2 = \{3,4,5\}$

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

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

Example: Absorption probabilities



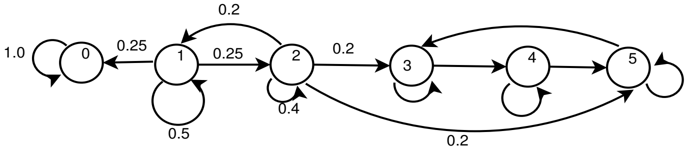
- $S_T = \{1,2\}$ and $C_1 = \{0\}$, $C_2 = \{3,4,5\}$

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

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

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

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

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

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.

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

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.

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

$$B = I + P_T B$$

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 $\sum_{x \in S} \pi(x) = 1$

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

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

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, π, over S, 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.

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, π, over S, 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.

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, π, over S, 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

$$\pi^T = \pi^T \ P$$

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, π, over S, 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

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

- π is a stationary distribution if

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

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

- 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

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

- 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

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

- 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 $\pi_1 = \pi$

- 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 $\pi_n = \pi, \ \forall n$

- 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.

- 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.

- 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.

- 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

- 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

$$\pi(y) = \pi_1(y)$$

- 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

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

- 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

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

- 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 can be represented by a vector

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

$$P^T \pi = \pi$$

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

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

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

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

- Suppose S is finite.
  Then π can be represented by a vector
- 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)

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

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

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

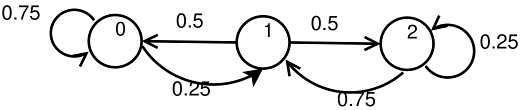
- Note that each column of P^T sums to 1.
- Hence, $\left(P^T - I\right)$ would be singular
  (1 is always an eigen value for a column stochastic matrix)
- A stationary distribution always exists for a finite chain.
- But it may or may not be unique.

- Suppose S is finite.
  Then π can be represented by a vector
- 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?

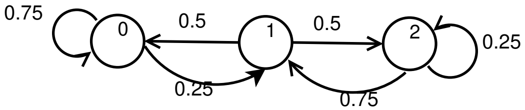
Example



- The stationary distribution has to satisfy

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

Example

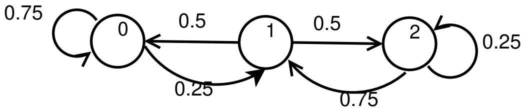


- The stationary distribution has to satisfy

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

- Thus we get the following linear equations

Example



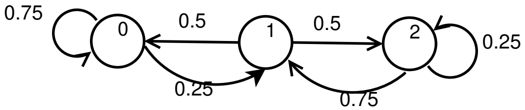
- The stationary distribution has to satisfy

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

- Thus we get the following linear equations

$$0.75\pi(0) + 0.5\pi(1) \;\; = \;\; \pi(0)$$

Example



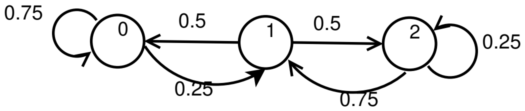
- The stationary distribution has to satisfy

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

- Thus we get the following linear equations

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

Example



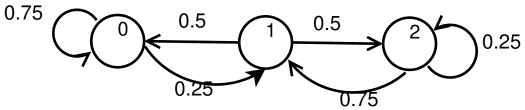
- The stationary distribution has to satisfy

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

- Thus we get the following linear equations

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

Example



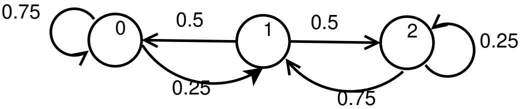
- The stationary distribution has to satisfy

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

- Thus we get the following linear equations

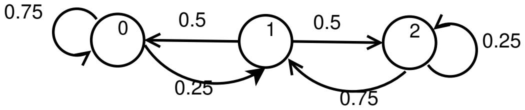
$$\begin{aligned} 0.75\pi(0) + 0.5\pi(1) &= \pi(0) \\ 0.25\pi(0) + 0.75\pi(2) &= \pi(1) \\ 0.5\pi(1) + 0.25\pi(2) &= \pi(2) \end{aligned}$$

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



- We can also write the equations for π as

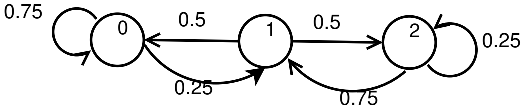
$$\begin{bmatrix} \pi(0) & \pi(1) & \pi(2) \end{bmatrix} \begin{bmatrix} 0.75 & 0.25 & 0 \\ 0.5 & 0 & 0.5 \\ 0 & 0.75 & 0.25 \end{bmatrix} = \begin{bmatrix} \pi(0) & \pi(1) & \pi(2) \end{bmatrix}$$



- We can also write the equations for π as

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

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

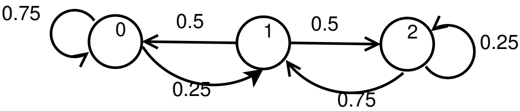


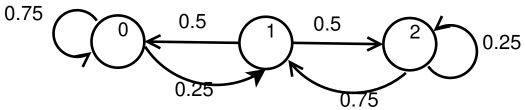
- We can also write the equations for π as

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

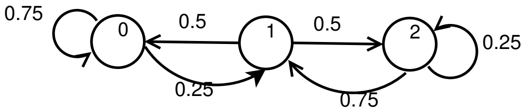
$$\begin{aligned} 0.75\pi(0) + 0.5\pi(1) &= \pi(0) \\ 0.25\pi(0) + 0.75\pi(2) &= \pi(1) \\ 0.5\pi(1) + 0.25\pi(2) &= \pi(2) \end{aligned}$$

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

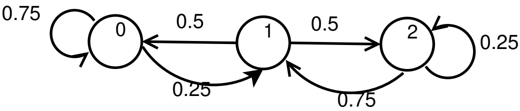

0.75
0
0.5
1
0.5
2
0.25
0.25
0.75



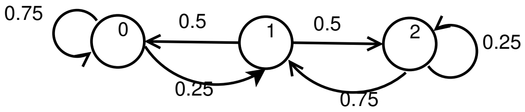
$$
\begin{aligned}
0.75\pi(0) + 0.5\pi(1) &= \pi(0) \\
0.25\pi(0) + 0.75\pi(2) &= \pi(1) \\
0.5\pi(1) + 0.25\pi(2) &= \pi(2) \\
\pi(0) + \pi(1) + \pi(2) &= 1
\end{aligned}
$$



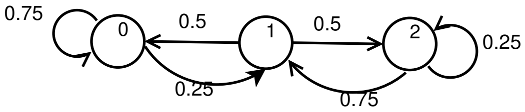
$$
\begin{aligned}
0.75\pi(0) + 0.5\pi(1) &= \pi(0) \quad \Rightarrow \quad \pi(1) = \frac{1}{2}\,\pi(0) \\
0.25\pi(0) + 0.75\pi(2) &= \pi(1) \\
0.5\pi(1) + 0.25\pi(2) &= \pi(2) \\
\pi(0) + \pi(1) + \pi(2) &= 1
\end{aligned}
$$



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

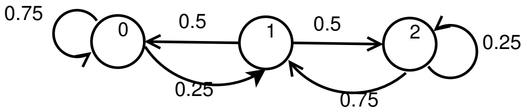


$$
\begin{aligned}
0.75\pi(0) + 0.5\pi(1) &= \pi(0) \;\Rightarrow\; \pi(1) = \frac{1}{2}\,\pi(0) \\
0.25\pi(0) + 0.75\pi(2) &= \pi(1) \;\Rightarrow\; \pi(2) = \frac{1}{3}\pi(0) \\
0.5\pi(1) + 0.25\pi(2) &= \pi(2) \\
\pi(0) + \pi(1) + \pi(2) &= 1 \;\Rightarrow\; \pi(0)\left(1 + \frac{1}{2} + \frac{1}{3}\right)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



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

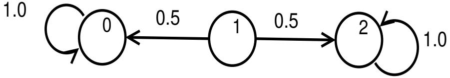
- Now, $\pi(0)\left(1 + \frac{1}{2} + \frac{1}{3}\right) = 1$ gives $\pi(0) = \frac{6}{11}$



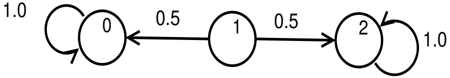
$$
\begin{aligned}
0.75\pi(0) + 0.5\pi(1) &= \pi(0) \quad \Rightarrow \quad \pi(1) = \frac{1}{2}\,\pi(0) \\
0.25\pi(0) + 0.75\pi(2) &= \pi(1) \quad \Rightarrow \quad \pi(2) = \frac{1}{3}\pi(0) \\
0.5\pi(1) + 0.25\pi(2) &= \pi(2) \\
\pi(0) + \pi(1) + \pi(2) &= 1 \quad \Rightarrow \quad \pi(0)\left(1 +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

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

Example2



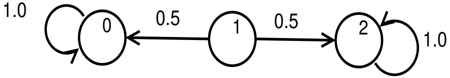
Example2



- The stationary distribution has to satisfy

$$\left[\begin{array}{ccc}\pi(0) & \pi(1) & \pi(2)\end{array}\right]\left[\begin{array}{ccc}1.0 & 0 & 0 \\ 0.5 & 0 & 0.5 \\ 0 & 0 & 1.0\end{array}\right]=\left[\begin{array}{ccc}\pi(0) & \pi(1) & \pi(2)\end{array}\right]$$

Example2

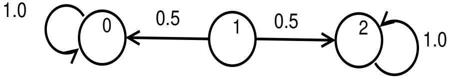


- The stationary distribution has to satisfy

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

- We also have to add the equation $\pi(0) + \pi(1) + \pi(2) = 1$

Example2

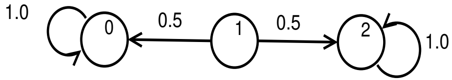


- The stationary distribution has to satisfy

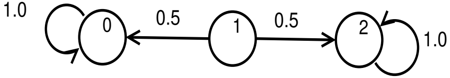
$$\left[\begin{array}{ccc} \pi(0) & \pi(1) & \pi(2) \end{array}\right] \left[\begin{array}{ccc} 1.0 & 0 & 0 \\ 0.5 & 0 & 0.5 \\ 0 & 0 & 1.0 \end{array}\right] = \left[\begin{array}{ccc} \pi(0) & \pi(1) & \pi(2) \end{array}\right]$$

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

Example2

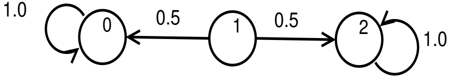


Example2



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

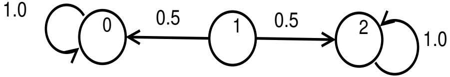
Example2



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

- We get the following linear equations

Example2

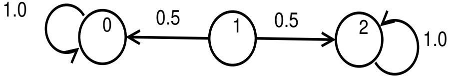


$$\pi(y) = \sum_{x \in S} \pi(x) P(x, y), \quad \forall y \in S$$

- We get the following linear equations

$$\pi(0) + 0.5\pi(1) \quad = \quad \pi(0)$$

Example2

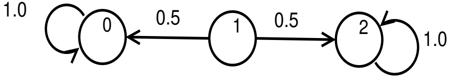


$$\pi(y) = \sum_{x \in S} \pi(x) P(x, y), \quad \forall y \in S$$

- We get the following linear equations

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

Example2

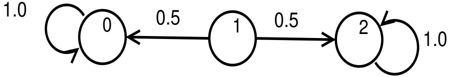


$$\pi(y) = \sum_{x \in S} \pi(x) P(x, y), \quad \forall y \in S$$

- We get the following linear equations

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

Example2

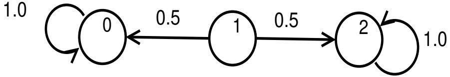


$$\pi(y) = \sum_{x \in S} \pi(x) P(x, y), \quad \forall y \in S$$

- We get the following linear equations

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

Example2

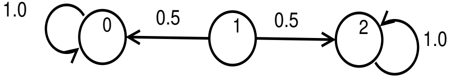


$$\pi(y) = \sum_{x \in S} \pi(x) P(x, y), \quad \forall y \in S$$

- We get the following linear equations

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

Example2

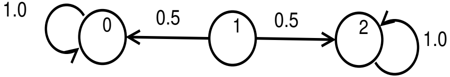


$$\pi(y) = \sum_{x \in S} \pi(x) P(x, y), \quad \forall y \in S$$

- We get the following linear equations

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

Example2



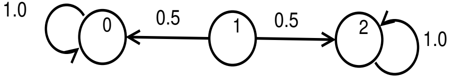
$$\pi(y) = \sum_{x \in S} \pi(x) P(x,y), \quad \forall y \in S$$

- 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.

Example2



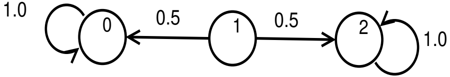
$$\pi(y) = \sum_{x \in S} \pi(x) P(x,y), \quad \forall y \in S$$

- 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 $0 \le a \le 1$ is a 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

- We now explore conditions for existence and uniqueness of stationary distributions
- For finite chains stationary distribution always exists.

- 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.

- 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.

- 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.
- The stationary distribution, when it exists, is related to expected fraction of time spent in different states.

- 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.
- The stationary distribution, when it exists, is related to expected fraction of time spent in different states.
- When the stationary distribution is unique, we also want to know if the chain converges to that distribution starting with any π_0.

- Let $I_y(X_n)$ be indicator of $[X_n = y]$

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

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

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

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

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

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

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

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

- Expected fraction of time spent in y till n is

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

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

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

- 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 $n \to \infty$

- Suppose y is transient. Then we have

- Suppose y is transient. Then we have

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

- Suppose y is transient. Then we have

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

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

- Suppose y is transient. Then we have

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

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

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

- Suppose y is transient. Then we have

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

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

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

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

- Suppose y is transient. Then we have

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

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

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

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

- Now, let y be recurrent

- Now, let y be recurrent
- Then, $P_y[T_y < \infty] = 1$

- Now, let y be recurrent
- Then, $P_y[T_y < \infty] = 1$
- Define $m_y = E_y[T_y]$

- 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

- 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.

- 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

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

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

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

- 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 $W_y^r = T_y^r - T_y^{r-1}, \ \ r > 1$

- 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 $E_y[W_y^1] = E_y[T_y] = m_y$

- 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, $T_y^r = W_y^1 + \cdots + W_y^r$

- 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"

- 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

- 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.

- 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

- We have

$$Pr[W_y^3 = k_3 | W_y^2 = k_2, W_y^1 = k_1] \quad =$$

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

- 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

- 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

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

- 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

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

- 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

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

- We have

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

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

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

- 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

- This shows the waiting time are iid

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

- This shows the waiting time are iid

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

- This shows the waiting time are iid

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

- 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

- 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

$$P_y[W_y^2 = k_2, \; W_y^1 = k_1] = P_y[W_y^2 = k_2 \mid W_y^1 = k_1] P_y[W_y^1 = k_1]$$

- 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

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

- 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

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

- This shows the waiting time are iid

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

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

- We have shown $W_y^r, \; r = 1, 2, \cdots$ are iid

- 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,

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

- 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

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

- 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

- 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 $N_{50}(y) = 8$

- 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.

- 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.

- 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.

- 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.

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

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

- Now we have

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

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

- 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

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

- 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 $n \to \infty$, $N_n(y) \to \infty$, $w.p.1$

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

- 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 $n \to \infty$, $\frac{T_y^n}{n} \to m_y$, $w.p.1$

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

- 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.

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

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

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

- 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

- 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

$$\lim_{n\to\infty} \frac{G_n(x,y)}{n} = \lim_{n\to\infty} E_x\left[\frac{N_n(y)}{n}\right]$$

- 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

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

- 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

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

- 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

- Thus we have proved the following theorem
- **Theorem**:
  Let y be recurrent. Then

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

- 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. $$\lim_{n\to\infty} \frac{G_n(x,y)}{n} = \frac{\rho_{xy}}{m_y}$$

- 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. $$\lim_{n\to\infty} \frac{G_n(x,y)}{n} = \frac{\rho_{xy}}{m_y}$$
- Note that

$$\frac{1}{m_y} = \lim_{n\to\infty} \frac{G_n(y,y)}{n}$$

- 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. $$\lim_{n\to\infty} \frac{G_n(x,y)}{n} = \frac{\rho_{xy}}{m_y}$$
- Note that

$$\frac{1}{m_y} = \lim_{n\to\infty} \frac{G_n(y,y)}{n} = \lim_{n\to\infty} \frac{1}{n}\sum_{m=1}^{n} P^m(y,y)$$

- The limiting fraction of time spent in a state is inversely proportional to m_y, the mean return time.

- 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.

- 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.

- 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 $m_y = \infty$.

- 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 $E_y[T_y] = \infty$.

- 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 $m_y = \infty, \ \forall y$?

- 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.

- 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??

- 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 $m_y = \infty$.

- A recurrent state y is called **null recurrent** if $m_y = \infty$.
- y is called **positive recurrent** if $m_y < \infty$

- 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.

- 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

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

- 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.

- **Theorem:** Let x be positive recurrent and let x lead to y. Then y is positive recurrent.

  Proof

- **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

- **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

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

- **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

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

- **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 $P^{n_1}(y,x) \ \frac{1}{m_x} \ P^{n_0}(x,y) > 0$.

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

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

- We can write the LHS of above as

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

- We can write the LHS of above as

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

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

- We can write the LHS of above as

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

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

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

- We can write the LHS of above as

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

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

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

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

- We can write the LHS of above as

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

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

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

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

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

- We can write the LHS of above as

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

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

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

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

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.

- 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.

- 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.