भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Stochastic Processes

Markov Chains

Karthik P. N.
Assistant Professor, Department of AI
Email: pnkarthik@ai.iith.ac.in

07 March 2025

Markov Chains

Markov Chain

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$.

Definition (Markov Chain)

A process $\{X_t : t \in \mathcal{T}\}$ is called a Markov chain if for any $t \in \mathcal{T}$,

$$(X_s : s < t) \perp\!\!\!\perp (X_s : s > t) \mid X_t,$$

Markov Chain

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$.

Definition (Markov Chain)

A process $\{X_t : t \in \mathcal{T}\}$ is called a Markov chain if for any $t \in \mathcal{T}$,

$$(X_s : s < t) \perp\!\!\!\perp (X_s : s > t) \mid X_t,$$

i.e., for any $m, n \in \mathbb{N}$, $s_1 < \cdots < s_m < t$, $t < t_1 < \cdots < t_n$, $x_1, \ldots, x_m \in \mathbb{R}$, $y_1, \ldots, y_n \in \mathbb{R}$, and $x \in \mathbb{R}$,

$$\begin{aligned} &\mathbb{P}(\underbrace{X_{s_1} \leq x_1, \ldots, X_{s_m} \leq x_m,}_{\text{before } t} \underbrace{X_{t_1} \leq y_1, \ldots, X_{t_n} \leq y_n}_{\text{after } t} \mid X_t \leq x) \\ &\quad = \mathbb{P}(X_{s_1} \leq x_1, \ldots, X_{s_m} \leq x_m \mid X_t \leq x) \cdot \mathbb{P}(X_{t_1} \leq y_1, \ldots, X_{t_n} \leq y_n \mid X_t \leq x). \end{aligned}$$

Markov Chain

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$.

Definition (Markov Chain)

A process $\{X_t : t \in \mathcal{T}\}$ is called a Markov chain if for any $t \in \mathcal{T}$,

$$(X_s : s < t) \perp\!\!\!\perp (X_s : s > t) \mid X_t,$$

i.e., for any $m, n \in \mathbb{N}$, $s_1 < \cdots < s_m < t, \quad t < t_1 < \cdots < t_n$, $x_1, \ldots, x_m \in \mathbb{R}$, $y_1, \ldots, y_n \in \mathbb{R}$, and $x \in \mathbb{R}$,

$$\mathbb{P}(\underbrace{X_{s_1} \le x_1, \ldots, X_{s_m} \le x_m,}_{\text{before } t} \underbrace{X_{t_1} \le y_1, \ldots, X_{t_n} \le y_n}_{\text{after } t} \mid X_t \le x)$$
$$= \mathbb{P}(X_{s_1} \le x_1, \ldots, X_{s_m} \le x_m \mid X_t \le x) \cdot \mathbb{P}(X_{t_1} \le y_1, \ldots, X_{t_n} \le y_n \mid X_t \le x).$$

In words: given the present value, the past is independent of future.

Discrete-Time Markov Chain Taking Finitely Many Values

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$.

Definition (DTMC taking Finitely Many Values)

Consider a process $\{X_n\}_{n=1}^{\infty}$ taking values in a finite set $\mathcal{X}$.
Then, $\{X_n\}_{n=1}^{\infty}$ is called a discrete time Markov chain (DTMC) on $\mathcal{X}$ if

$$(X_1, \ldots, X_{n-1}) \perp\!\!\!\perp (X_{n+1}, X_{n+2}, \ldots) \mid X_n \qquad \text{for any } n \in \mathbb{N},$$

Discrete-Time Markov Chain Taking Finitely Many Values

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$.

Definition (DTMC taking Finitely Many Values)

Consider a process $\{X_n\}_{n=1}^{\infty}$ taking values in a finite set $\mathcal{X}$.
Then, $\{X_n\}_{n=1}^{\infty}$ is called a discrete time Markov chain (DTMC) on $\mathcal{X}$ if

$$(X_1, \ldots, X_{n-1}) \perp\!\!\!\perp (X_{n+1}, X_{n+2}, \ldots) \mid X_n \qquad \text{for any } n \in \mathbb{N},$$

i.e., for any $n, L \in \mathbb{N}$, $\;n < t_1 < \cdots < t_L$,
$x_1, \ldots, x_{n-1} \in \mathbb{R}$, $y_1, \ldots, y_L \in \mathbb{R}$, and $x \in \mathbb{R}$,

$$\mathbb{P}(\underbrace{X_1 = x_1, \ldots, X_{n-1} = x_{n-1},}_{\text{before } n} \underbrace{X_{t_1} = y_1, \ldots, X_{t_L} = y_L}_{\text{after } n} \mid X_n = x)$$

$$= \mathbb{P}(X_1 = x_1, \ldots, X_{n-1} = x_{n-1} \mid X_n = x) \cdot \mathbb{P}(X_{t_1} = y_1, \ldots, X_{t_L} = y_L \mid X_n = x).$$

Alternate Viewpoint of DTMC

$$\mathbb{P}(\underbrace{X_{t_1} = y_1, \ldots, X_{t_L} = y_L}_{\text{after } n} \mid X_n = x, \underbrace{X_1 = x_1, \ldots, X_{n-1} = x_{n-1}}_{\text{before } n})$$
$$= \mathbb{P}(X_{t_1} = y_1, \ldots, X_{t_L} = y_L \mid X_n = x).$$

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Alternate Viewpoint of DTMC

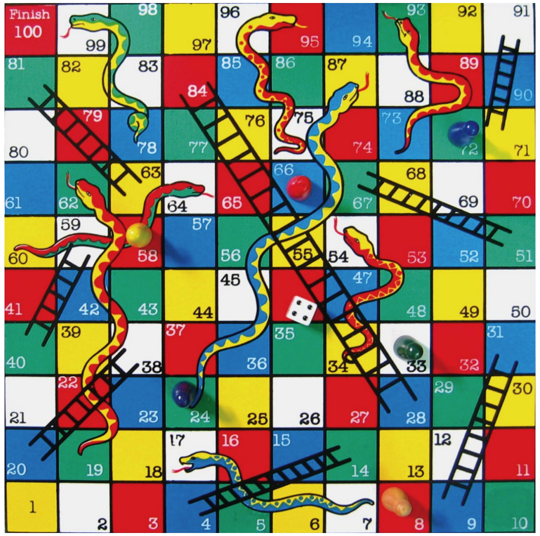
$$\mathbb{P}(\underbrace{X_{t_1} = y_1, \ldots, X_{t_L} = y_L}_{\text{after } n} \mid X_n = x, \underbrace{X_1 = x_1, \ldots, X_{n-1} = x_{n-1}}_{\text{before } n})$$
$$= \mathbb{P}(X_{t_1} = y_1, \ldots, X_{t_L} = y_L \mid X_n = x).$$

Important

To determine X_{n+1} given the history $(X_1, \ldots, X_n)$ up to time n, it suffices to retain only X_n and discard $(X_1, \ldots, X_{n-1})$.

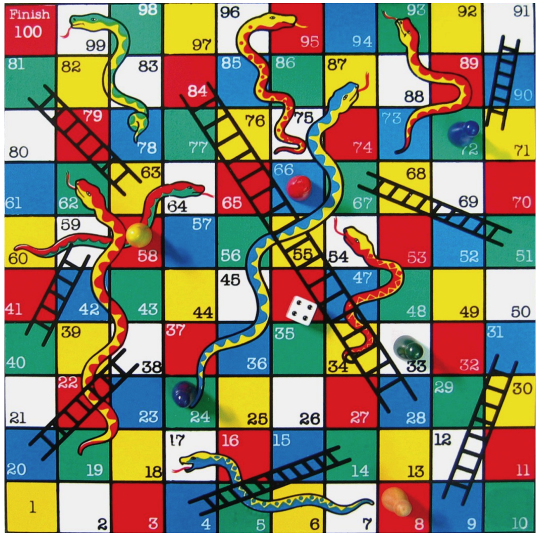
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Example



- $x =$

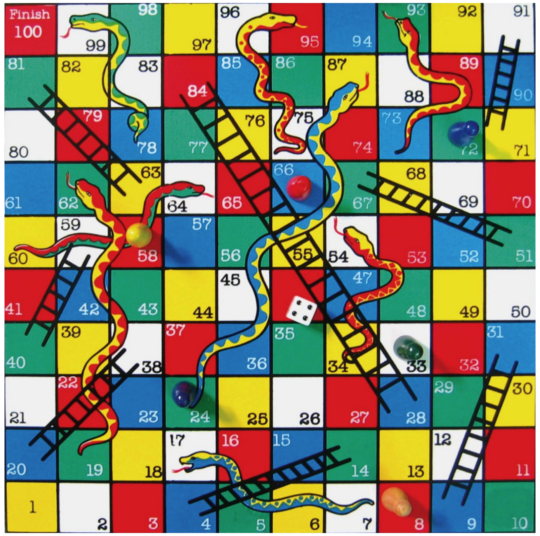
Example



- $\mathcal{X} = \{1, \ldots, 100\}$
- Suppose $X_n = 6$ at some time n
- Then, $X_{n+1} \in$

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Example

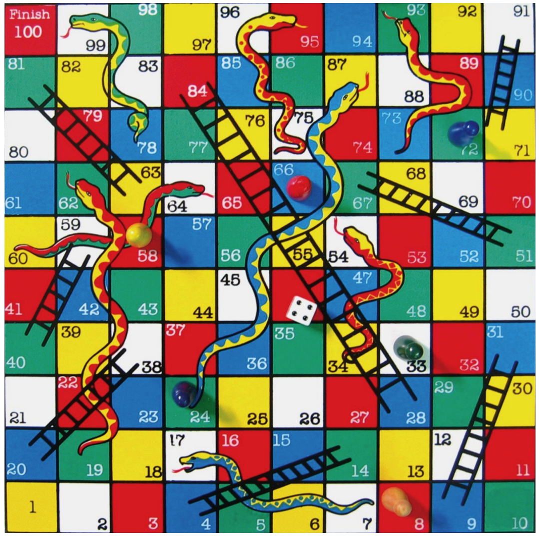


- $\mathcal{X} = \{1, \ldots, 100\}$
- Suppose $X_n = 6$ at some time n
- Then, $X_{n+1} \in \{7, 8, 31, 10, 11, 12\}$
- For all $x, x_1, \ldots, x_{n-1} \in \mathcal{X}$, we have

$$\mathbb{P}(X_{n+1} = x | X_n = 6, \underbrace{X_{n-1} = x_{n-1}, \ldots, X_1 = x_1}_{\text{trajectory before } n})$$

$$= \mathbb{P}(X_{n+1} = x | X_n = 6)$$

Example



- $\mathcal{X} = \{1, \ldots, 100\}$
- Suppose $X_n = 6$ at some time n
- Then, $X_{n+1} \in \{7, 8, 31, 10, 11, 12\}$
- For all $x, x_1, \ldots, x_{n-1} \in \mathcal{X}$, we have

$$\mathbb{P}(X_{n+1} = x | X_n = 6, \underbrace{X_{n-1} = x_{n-1}, \ldots, X_1 = x_1}_{\text{trajectory before } n})$$

$$= \mathbb{P}(X_{n+1} = x | X_n = 6)$$

Remark

The set $\mathcal{X}$ is called the state space of the Markov chain.

Example

Let $\{X_n\}_{n=1}^{\infty}$ be an $\mathbb{N}$-valued i.i.d. process.
For each $n \in \mathbb{N}$, let $S_n = \sum_{i=1}^{n} X_i$.

For any $\mathbf{s} = (s_1, \ldots, s_{n-1})$,

$$\mathbb{P}(S_{n+1} = s_{n+1} \mid S_n = s_n,\ S_{1:n-1} = \mathbf{s}) =$$

Example

Let $\{X_n\}_{n=1}^{\infty}$ be an $\mathbb{N}$-valued i.i.d. process.
For each $n \in \mathbb{N}$, let $S_n = \sum_{i=1}^{n} X_i$.

For any $\mathbf{s} = (s_1, \ldots, s_{n-1})$,

$$\mathbb{P}(S_{n+1} = s_{n+1} \mid S_n = s_n, \; S_{1:n-1} = \mathbf{s}) = \mathbb{P}(X_{n+1} = s_{n+1} - s_n)$$

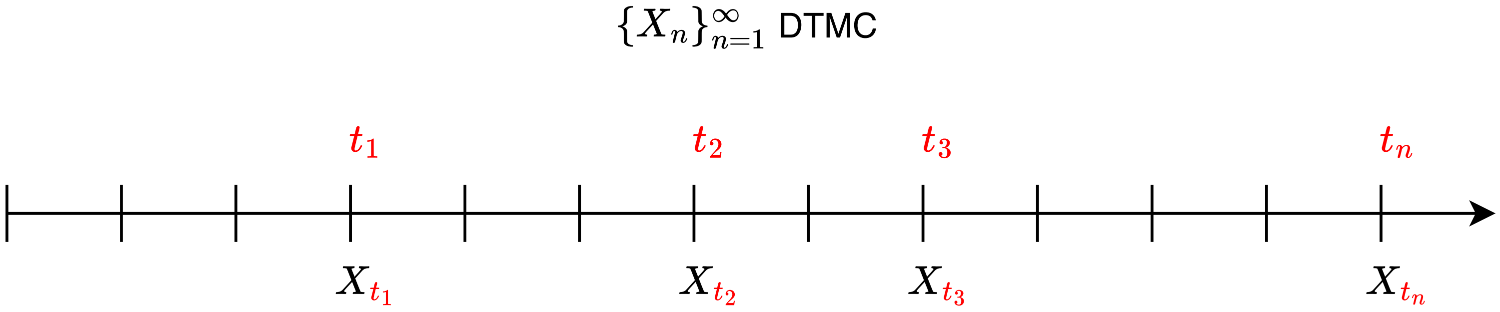
Example

Let $\{X_n\}_{n=1}^{\infty}$ be an $\mathbb{N}$-valued i.i.d. process.
For each $n \in \mathbb{N}$, let $S_n = \sum_{i=1}^{n} X_i$.

For any $\mathbf{s} = (s_1, \ldots, s_{n-1})$,

$$\mathbb{P}(S_{n+1} = s_{n+1} \mid S_n = s_n,\ S_{1:n-1} = \mathbf{s}) = \mathbb{P}(X_{n+1} = s_{n+1} - s_n) = \mathbb{P}(S_{n+1} = s_{n+1} \mid S_n = s_n)$$

Markov Property for Deterministic Sampling Times



Lemma (Markov Property for Deterministic Sampling Times)

Suppose that $\{X_n\}_{n=1}^{\infty}$ is a DTMC with a finite state space $\mathcal{X}$.
For all $n \in \mathbb{N}$, deterministic $t_1 < t_2 < \cdots < t_n$, and $x_1, x_2, \ldots, x_n \in \mathbb{R}$,

$$\mathbb{P}(X_{t_n} = x_n \mid X_{t_{n-1}} = x_{n-1}, \ldots, X_{t_1} = x_1) = \mathbb{P}(X_{t_n} = x_n \mid X_{t_{n-1}} = x_{n-1}).$$

In words: Suffices to retain the most recent information and discard the history.

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Transition Probability Matrix

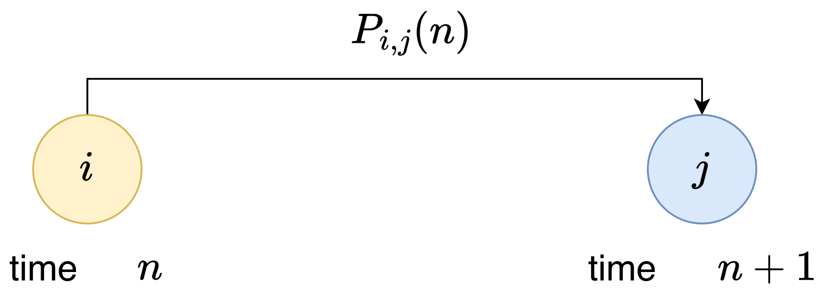
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Transition Probability Matrix

Definition (Transition Probability Matrix)

Let $\{X_n\}_{n=1}^{\infty}$ be a DTMC with discrete state space $\mathcal{X}$.
The transition probability matrix (TPM) of the Markov chain at any time $n \in \mathbb{N}$ is a matrix $P(n) = [P_{i,j}(n)]_{i,j\in\mathcal{X}}$ defined as

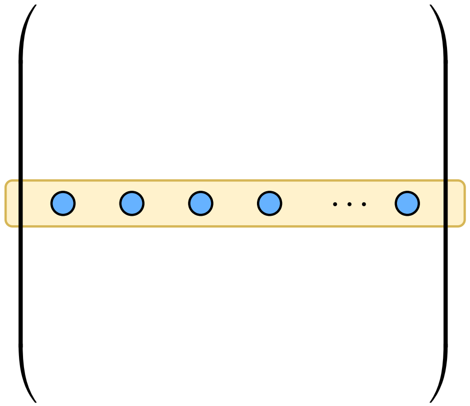
$$P_{i,j}(n) = \mathbb{P}(X_{n+1} = j \mid X_n = i), \qquad i,j \in \mathcal{X}.$$



Transition Probability Matrix

- For each $i \in \mathcal{X}$, $\sum_{j \in \mathcal{X}} P_{i,j}(n) =$

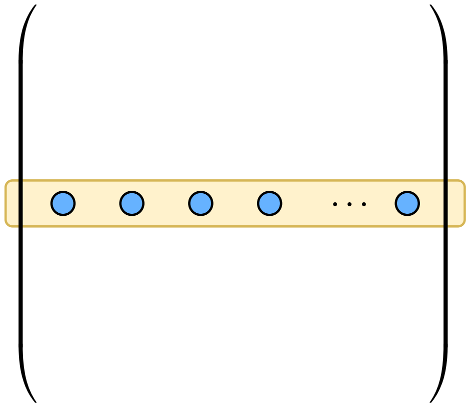
i



Transition Probability Matrix

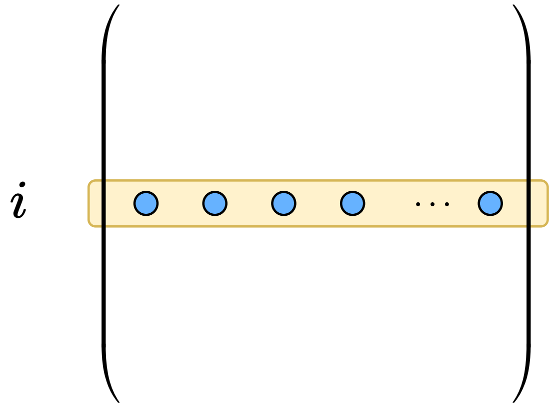
- For each $i \in \mathcal{X}$, $\sum_{j \in \mathcal{X}} P_{i,j}(n) =$

i



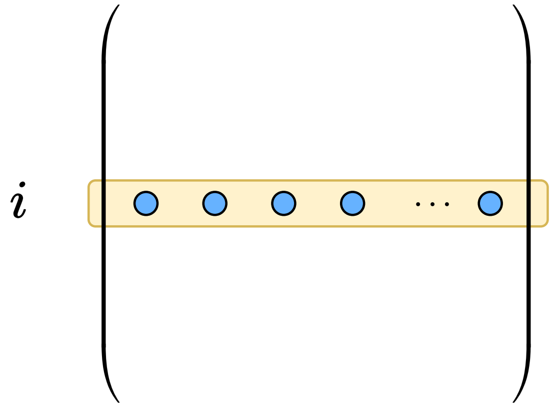
Transition Probability Matrix

- For each $i \in \mathcal{X}$, $\quad \sum_{j \in \mathcal{X}} P_{i,j}(n) = 1.$
- A matrix with non-negative entries and row sums equal to 1 is called a row stochastic matrix



भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

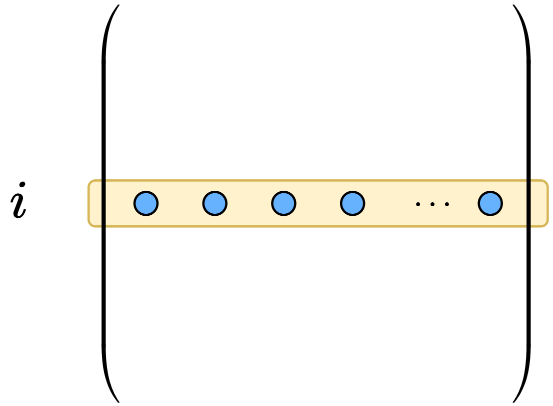
Transition Probability Matrix



- For each $i \in \mathcal{X}$, $\quad \sum_{j \in \mathcal{X}} P_{i,j}(n) = 1.$
- A matrix with non-negative entries and row sums equal to 1 is called a row stochastic matrix
- $P(n)$ is a row stochastic matrix for every n

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

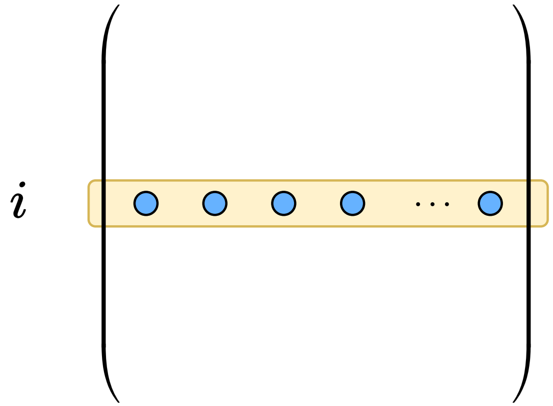
Transition Probability Matrix



- For each $i \in \mathcal{X}$, $\quad \sum_{j \in \mathcal{X}} P_{i,j}(n) = 1.$
- A matrix with non-negative entries and row sums equal to 1 is called a row stochastic matrix
- $P(n)$ is a row stochastic matrix for every n
- Each row of $P(n) =$ PMF on $\mathcal{X}$

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Transition Probability Matrix

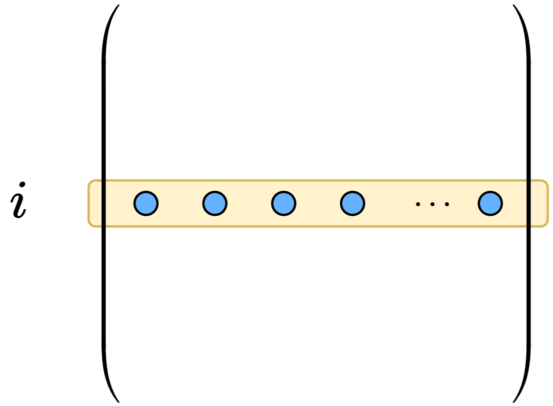


- For each $i \in \mathcal{X}$, $\sum_{j \in \mathcal{X}} P_{i,j}(n) = 1$.
- A matrix with non-negative entries and row sums equal to 1 is called a row stochastic matrix
- $P(n)$ is a row stochastic matrix for every n
- Each row of $P(n)$ = PMF on $\mathcal{X}$
- For instance,

$$\sum_{j \in \mathcal{X}} j^2 \, P_{i,j}(n) = \mathbb{E}[X_{n+1}^2 \mid X_n = i],$$

where $\mathbb{E}$ above is w.r.t. row i of $P(n)$

Transition Probability Matrix

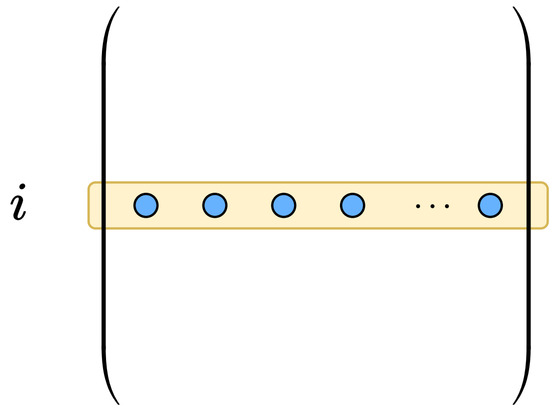


- For each $i \in \mathcal{X}$, $\quad \sum_{j \in \mathcal{X}} P_{i,j}(n) = 1$.
- A matrix with non-negative entries and row sums equal to 1 is called a row stochastic matrix
- $P(n)$ is a row stochastic matrix for every n
- Each row of $P(n)$ = PMF on $\mathcal{X}$
- For instance,

$$\sum_{j \in \mathcal{X}} j^2 \, P_{i,j}(n) = \mathbb{E}[X_{n+1}^2 \mid X_n = i],$$

  where $\mathbb{E}$ above is w.r.t. row i of $P(n)$
- $P(n)$ has a right eigenvector with eigenvalue 1

Transition Probability Matrix



- For each $i \in \mathcal{X}$, $\quad \sum_{j \in \mathcal{X}} P_{i,j}(n) = 1.$
- A matrix with non-negative entries and row sums equal to 1 is called a row stochastic matrix
- $P(n)$ is a row stochastic matrix for every n
- Each row of $P(n)$ = PMF on $\mathcal{X}$
- For instance,

$$\sum_{j \in \mathcal{X}} j^2 \, P_{i,j}(n) = \mathbb{E}[X_{n+1}^2 \mid X_n = i],$$

  where $\mathbb{E}$ above is w.r.t. row i of $P(n)$
- $P(n)$ has a right eigenvector with eigenvalue 1
- $P(n) \cdot \mathbf{1} = \mathbf{1}$, where $\mathbf{1}$ is the all-ones vector

Time Homogeneous DTMC

Definition (Time Homogeneous DTMC

A DTMC with discrete state space $\mathcal{X}$ and TPMs $\{P(n)\}_{n=1}^{\infty}$ is called time homogeneous if

$$P(n) = P(n+1) \qquad \forall n \in \mathbb{N}.$$

In this case, we simply write P to denote the common TPM.

Example

Let $X_1, X_2, \ldots$ be i.i.d. on $\{-1, +1\}$, with $\mathbb{P}(X_1 = 1) = p$.
For each $n \in \mathbb{N}$, let $S_n = \sum_{i=1}^{n} X_i$.
What is the TPM of $\{S_n\}_{n=1}^{\infty}$?

$\{X_n\}_{n=1}^{\infty}$ DTMC ; state space $\mathcal{X}$.

$\{X_n\}_{n=1}^{\infty}$ time homogeneous; TPM P.

$$\mathbb{P}(X_{n+1}=j \mid X_n = i) = P_{i,j} \quad \forall\, n \in \mathbb{N}.$$

$$\mathbb{P}(X_2 = j \mid X_1 = i) = P_{i,j}.$$

$$\mathbb{P}(X_3=j \mid X_1 = i) = \sum_{k\in\mathcal{X}} \mathbb{P}(X_3=j \mid X_2=k,\, X_1=i)\, \mathbb{P}(X_2=k \mid X_1=i)$$

$$= \sum_{k\in\mathcal{X}} \mathbb{P}(X_3=j \mid X_2=k) \cdot \mathbb{P}(X_2=k \mid X_1=i)$$

$$= \sum_{k\in\mathcal{X}} P_{k,j} \cdot P_{i,k} \quad = P^2_{i,j}$$

$$= \underset{P_{i,\cdot}}{\begin{bmatrix} \quad & \quad & \quad \end{bmatrix}} \overset{P_{\cdot,j}}{\begin{bmatrix} \; \\ \; \\ \; \end{bmatrix}}$$

Chapman – Kolmogorov Equation

Consider a time-homogeneous DTMC $\{X_n\}_{n=1}^{\infty}$ with discrete state space $\mathcal{X}$.

For any $n \in \mathbb{N}$, $i, j \in \mathcal{X}$, let

$$p_{i,j}^{(n)} := \mathbb{P}\left(X_{n+1} = j \mid X_1 = i \right).$$

Chapman - Kolmogorov:

Let $P^{(n)} = \left[p_{i,j}^{(n)} \right]_{i,j \in \mathcal{X}}$. Then

$$\boxed{P^{(n)} = P^n \qquad \forall\, n \in \mathbb{N}}$$

Proof: Exercise (induction!)