# Markov Chains

- Consider a sequence of discrete random variables $X_0, X_1, X_2, \ldots$ where the index represents discrete time.

- Recall that an infinite sequence of discrete random variables is described by a collection of joint PMFs $P_{X_{t_1}, X_{t_2}, \ldots, X_{t_m}}(x_{t_1}, x_{t_2}, \ldots, x_{t_m})$ for every possible choice of time indices $t_1 < t_2 < \cdots < t_m$ (and any $m$).

- Previously, we considered independent and identically distributed (i.i.d.) random variables, $P_{X_{t_1}, X_{t_2}, \ldots, X_{t_m}}(x_{t_1}, x_{t_2}, \ldots, x_{t_m}) = P_X(x_{t_1})\, P_X(x_{t_2}) \cdots P_X(x_{t_m})$, which have no dependence across time.

- Here, we introduce a simple way to model memory over time. We say the sequence $X_0, X_1, X_2, \ldots$ has the Markov property if, for $t_1 < t_2 < \cdots < t_m$, $X_{t_m}$ conditioned on $X_{t_1}, \ldots, X_{t_{m-1}}$ only depends on the most recent random variable $X_{t_{m-1}}$,

$$P_{X_{t_m} | X_{t_{m-1}}, \ldots, X_{t_1}}(x_{t_m} | x_{t_{m-1}}, \ldots, x_{t_1}) = P_{X_{t_m} | X_{t_{m-1}}}(x_{t_m} | x_{t_{m-1}}).$$

- A discrete-time Markov chain is a sequence of discrete random variables $X_0, X_1, X_2, \ldots$ satisfying the Markov property.
  - → We refer to the random variable $X_i$ as the state at time $i$.
  - → The Markov property is equivalent to the following:
    The next state $X_{n+1}$, conditioned on the full history $X_0, \ldots, X_n$, only depends on the current state $X_n$,

$$\boxed{P_{X_{n+1}|X_n,\ldots,X_0}(x_{n+1} \mid x_n, \ldots, x_0) = P_{X_{n+1}|X_n}(x_{n+1} \mid x_n).}$$

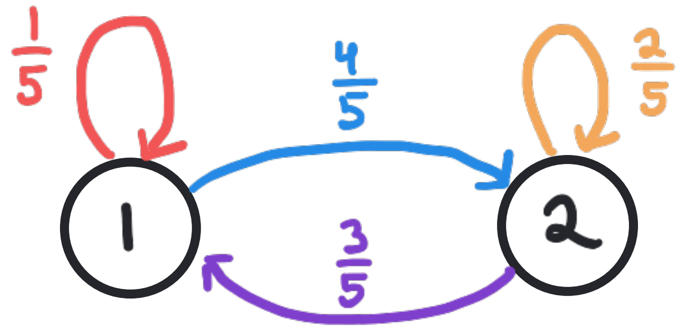
- We will focus on Markov chains with the following properties:
  - → Finite Range (or Finite State Space): To simplify our notation, we will set the range to $R_X = \{1, 2, \ldots, K\}$.
  - → Homogeneous: The conditional PMF $P_{X_{t+1}|X_t}(x_{t+1} \mid x_t)$ is not a function of $t$, just the two values plugged in, $x_{t+1}, x_t$.
- Some applications: Google's PageRank Algorithm, models for population genetics, epidemics, speech, network traffic, mathematical finance, etc.

- We define the transition probabilities $P_{jk}$ for $j,k \in \{1,\dots,K\}$ to be the probability of going from state $j$ to state $k$ in a single step. Specifically, we have $P_{X_{t+1}|X_t}(k \mid j) = P_{jk}$ for all $t$.
  - → Non-negativity: $P_{jk} \geq 0$ → Normalization: $\sum_{k=1}^{K} P_{jk} = 1$

- It can be very useful to visualize Markov chains as follows:
  - → Draw a node for each state in the range $R_X = \{1,\dots,K\}$.
  - → For each positive transition probability $P_{jk} > 0$, draw a directed edge from node $j$ to node $k$. Label this edge with the value of $P_{jk}$. (Do not draw an edge if $P_{jk} = 0$.)

- Example: $R_X = \{1,2\}$

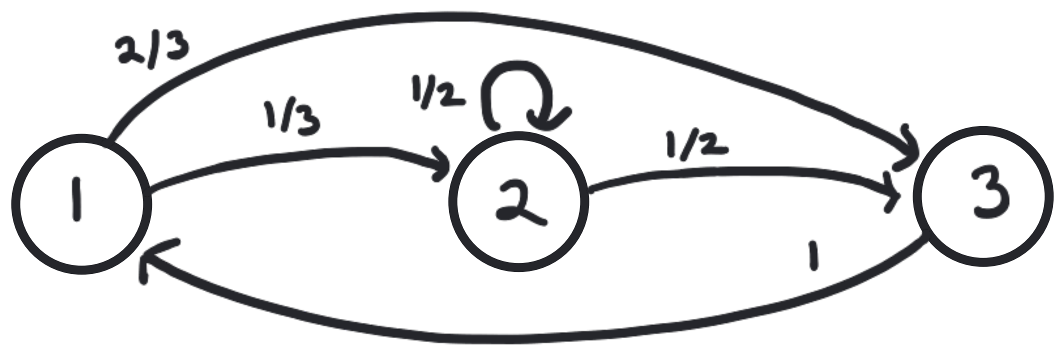
$P_{11} = \frac{1}{5}$ $P_{12} = \frac{4}{5}$ $P_{21} = \frac{3}{5}$ $P_{22} = \frac{2}{5}$




- Example: $R_X = \{1,2,3\}$

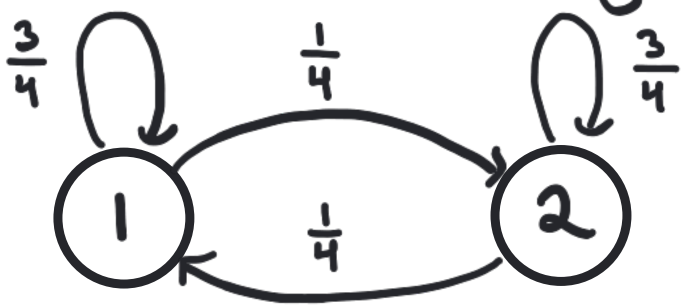
$P_{11} = 0$ $P_{12} = \frac{1}{3}$ $P_{13} = \frac{2}{3}$ $P_{21} = 0$ $P_{22} = \frac{1}{2}$ $P_{23} = \frac{1}{2}$

$P_{31} = 1$ $P_{32} = 0$ $P_{33} = 0$

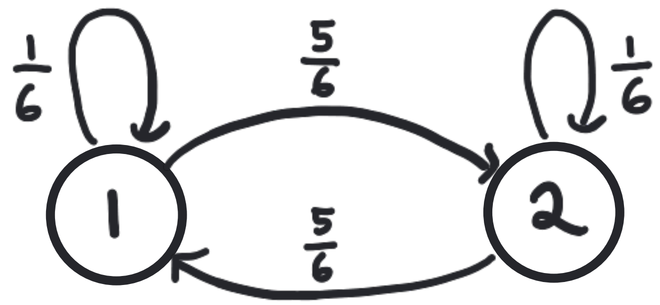


- The initial distribution is the marginal PMF of $X_0$ $P_{X_0}(x_0)$.
  - → If $P_{X_0}(x_0) = \begin{cases} 1 & x_0 = a \\ 0 & x_0 \neq a \end{cases}$, then we say $a$ is the initial state.
  - → The initial distribution $P_{X_0}(x_0)$ along with the transition probabilities $P_{jk}$ provide a full probabilistic description of a Markov chain.

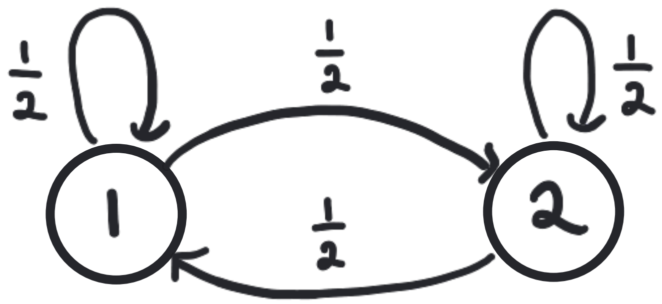
- <u>Example</u>: Let us examine some sample paths for Markov chains with range $R_X = \{1, 2\}$ and initial state 1.



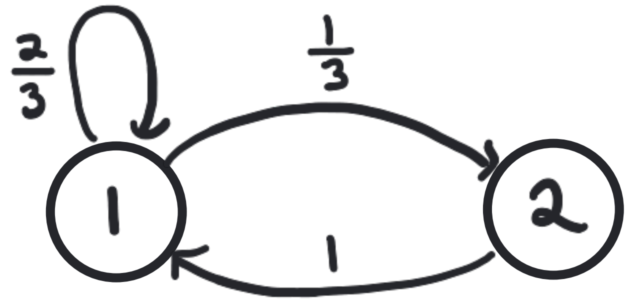

| $i$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $X_i$ | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 1 | 1 | 2 | 2 | 2 |




| $i$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $X_i$ | 1 | 2 | 1 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 2 |




| $i$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $X_i$ | 1 | 1 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 2 | 2 | 1 |




| $i$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $X_i$ | 1 | 1 | 1 | 2 | 1 | 1 | 2 | 1 | 1 | 1 | 2 | 1 |

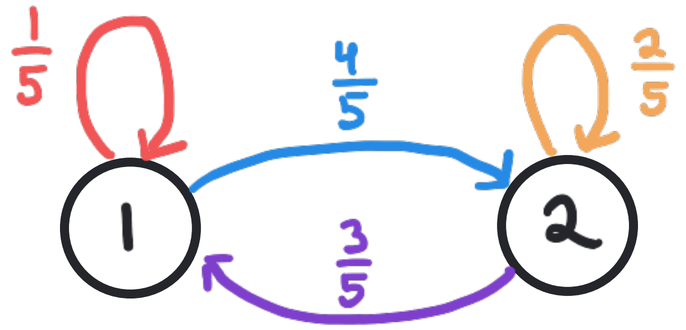
- The transition probabilities $P_{jk}$ tell us the probability of going from state $j$ to state $k$ in one time step. What about two or more time steps?

- The n-step transition probabilities $P_{jk}(n)$ for $j, k \in \{1, \dots, K\}$ specify the probability of going from state $j$ to state $k$ in exactly $n$ time steps. That is, $P_{X_{t+n} | X_t}(k \mid j) = P_{jk}(n)$ for all $t$.
  - → Note that $P_{jk}(1) = P_{jk}$.  → Normalization: $\sum_{k=1}^{K} P_{jk}(n) = 1$
  - → They can be determined using the Chapman - Kolmogorov equations:

$$\boxed{P_{jk}(n+m) = \sum_{i=1}^{K} P_{ji}(n) P_{ik}(m)}$$

- Example: Determine the two-step transition probabilities.

$P_{11} = \frac{1}{5}$  $P_{12} = \frac{4}{5}$  $P_{21} = \frac{3}{5}$  $P_{22} = \frac{2}{5}$




$$P_{jk}(2) = \sum_{i=1}^{2} P_{ji}(1) P_{ik}(1) = P_{j1} P_{1k} + P_{j2} P_{2k}$$

$$P_{11}(2) = P_{11} P_{11} + P_{12} P_{21} = \frac{1}{5} \cdot \frac{1}{5} + \frac{4}{5} \cdot \frac{3}{5} = \frac{13}{25}$$

$$P_{12}(2) = P_{11} P_{12} + P_{12} P_{22} = \frac{1}{5} \cdot \frac{4}{5} + \frac{4}{5} \cdot \frac{2}{5} = \frac{12}{25}$$

$$P_{21}(2) = P_{21} P_{11} + P_{22} P_{21} = \frac{3}{5} \cdot \frac{1}{5} + \frac{2}{5} \cdot \frac{3}{5} = \frac{9}{25}$$

$$P_{22}(2) = P_{21} P_{12} + P_{22} P_{22} = \frac{3}{5} \cdot \frac{4}{5} + \frac{2}{5} \cdot \frac{2}{5} = \frac{16}{25}$$