# Markov Process


**Industrial AI Lab.**

**Prof. Seungchul Lee**

# Sequential Processes

- Most classifiers ignored the sequential aspects of data
- Consider a system which can occupy one of *N* discrete states or categories

$$q_t \in \{S_1, S_2, \ldots, S_N\}$$

- We are interested in stochastic systems, in which state evolution is random
- Any joint distribution can be factored into a series of conditional distributions

$$p(q_0, q_1, \ldots, q_T) = p(q_0)p(q_1|q_0)p(q_2|q_1q_0)\cdots$$

Almost impossible to compute !

# Markov Chain

- Joint distribution can be factored into a series of conditional distributions

$$p(q_0, q_1, \cdots, q_T) = p(q_0)\, p(q_1 \mid q_0)\, p(q_2 \mid q_1, q_0) \cdots$$

- Markovian property (assumption)

$$p(q_{t+1} \mid q_t, \cdots, q_0) = p(q_{t+1} \mid q_t)$$

- Tractable in computation of joint distribution

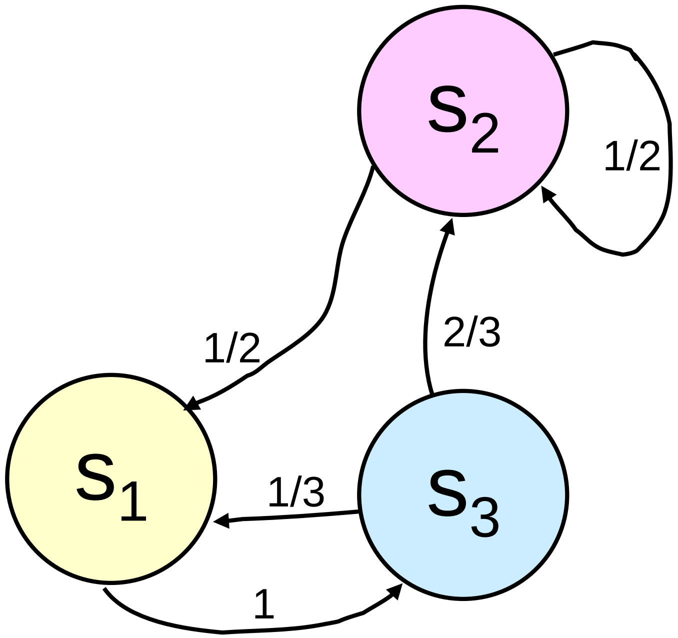
$$\begin{aligned} p(q_0, q_1, \cdots, q_T) &= p(q_0)\, p(q_1 \mid q_0)\, p(q_2 \mid q_1, q_0)\, p(q_3 \mid q_2, q_1, q_0) \cdots \\ &= p(q_0)\, p(q_1 \mid q_0)\, p(q_2 \mid q_1)\, p(q_3 \mid q_2) \cdots \end{aligned}$$

# Markov Process

$$p(q_0, q_1, \dots, q_T) = p(q_0)p(q_1|q_0)p(q_2|q_1q_0)p(q_3|q_2q_1q_0)\cdots$$

$$p(q_0, q_1, \dots, q_T) = p(q_0)p(q_1|q_0)p(q_2|q_1)p(q_3|q_2)\cdots$$

Possible and tractable



# Markov Process

- (Assumption) for a Markov process, the next state depends only on the current state:

$$p(q_{t+1}|q_t, \cdots, q_0) = p(q_{t+1}|q_t)$$

- More clearly

$$P\left(q_{t+1} = s_j | q_t = s_i\right) = P\left(q_{t+1} = s_j | q_t = s_i, \text{any earlier history}\right)$$

- Given current state, the past does not matter
- The state captures all relevant information from the history
- The state is a sufficient statistic of the future

# State Transition Matrix

- For a Markov state $s$ and successor state $s'$, the state transition probability is defined by

$$\mathcal{P}_{ss'} = \mathbb{P}\left[S_{t+1} = s' \mid S_t = s\right]$$

- State transition matrix $P$ defines transition probabilities from all states $s$ to all successor states $s'$,

$$\mathcal{P} \quad = \textit{from} \quad \begin{array}{c} \textit{to} \\ \begin{bmatrix} \mathcal{P}_{11} & \dots & \mathcal{P}_{1n} \\ \vdots & & \\ \mathcal{P}_{n1} & \dots & \mathcal{P}_{nn} \end{bmatrix} \end{array}$$

POSTECH

# Markov Process

- A Markov process is a memoryless random process, i.e., a sequence of random states $s_1, s_2, \cdots$ with the Markov property

**Definition**

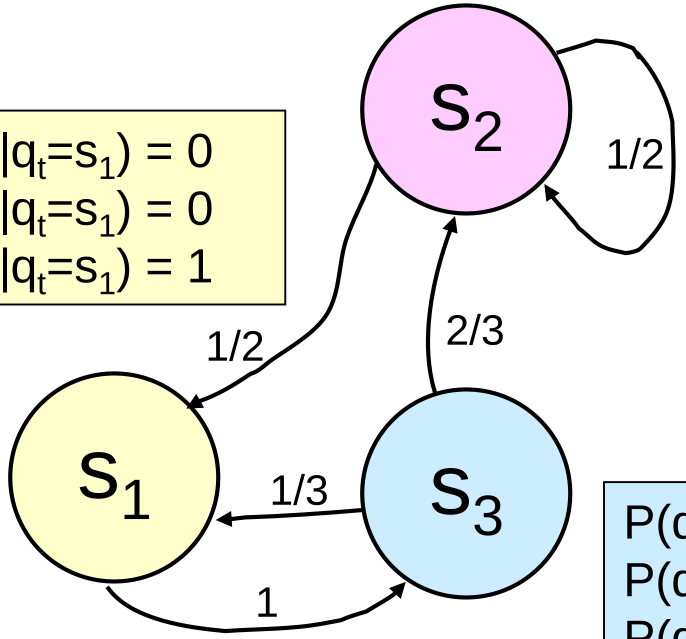
A *Markov Process* (or *Markov Chain*) is a tuple $\langle \mathcal{S}, \mathcal{P} \rangle$

- $\mathcal{S}$ is a (finite) set of states
- $\mathcal{P}$ is a state transition probability matrix,
  $\mathcal{P}_{ss'} = \mathbb{P}\left[S_{t+1} = s' \mid S_t = s\right]$

# State Transition Matrix

$P(q_{t+1}=s_1|q_t=s_2) = 1/2$
$P(q_{t+1}=s_2|q_t=s_2) = 1/2$
$P(q_{t+1}=s_3|q_t=s_2) = 0$

$P(q_{t+1}=s_1|q_t=s_1) = 0$
$P(q_{t+1}=s_2|q_t=s_1) = 0$
$P(q_{t+1}=s_3|q_t=s_1) = 1$



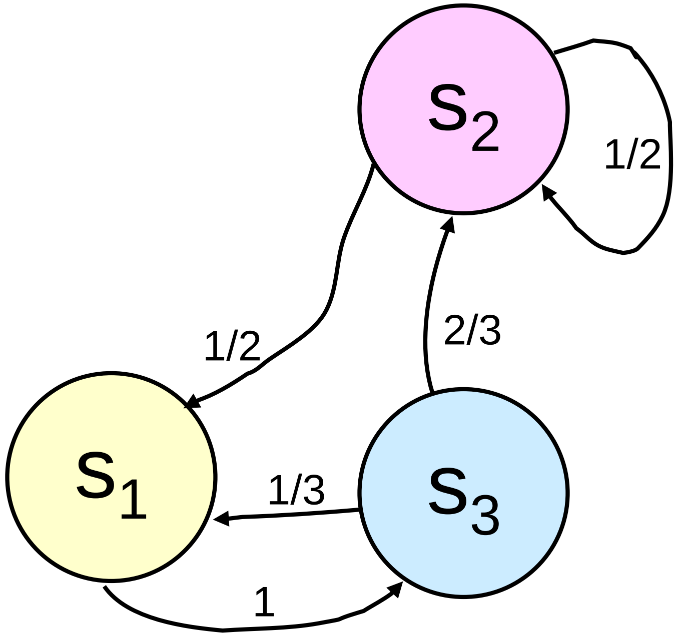

$P(q_{t+1}=s_1|q_t=s_3) = 1/3$
$P(q_{t+1}=s_2|q_t=s_3) = 2/3$
$P(q_{t+1}=s_3|q_t=s_3) = 0$

$$P = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 1/2 & 1/2 & 0 \\ 1/3 & 2/3 & 0 \end{bmatrix}$$

# Property of *P* Matrix

- Sum of the elements on each row yields 1

$$\sum_{j \in S} p_{i,j} = 1$$

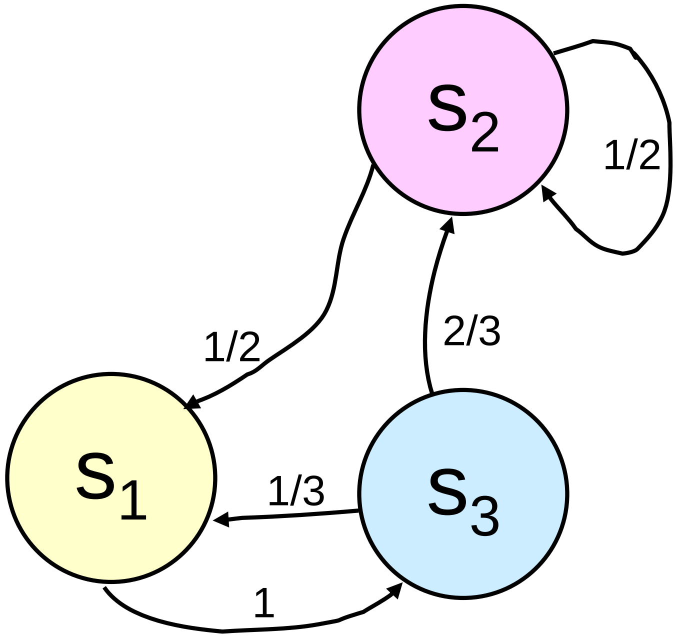



$$P = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 1/2 & 1/2 & 0 \\ 1/3 & 2/3 & 0 \end{bmatrix}$$

# Property of $P$ Matrix

- Sum of the elements on each row yields 1

$$\sum_{j \in S} p_{i,j} = 1$$

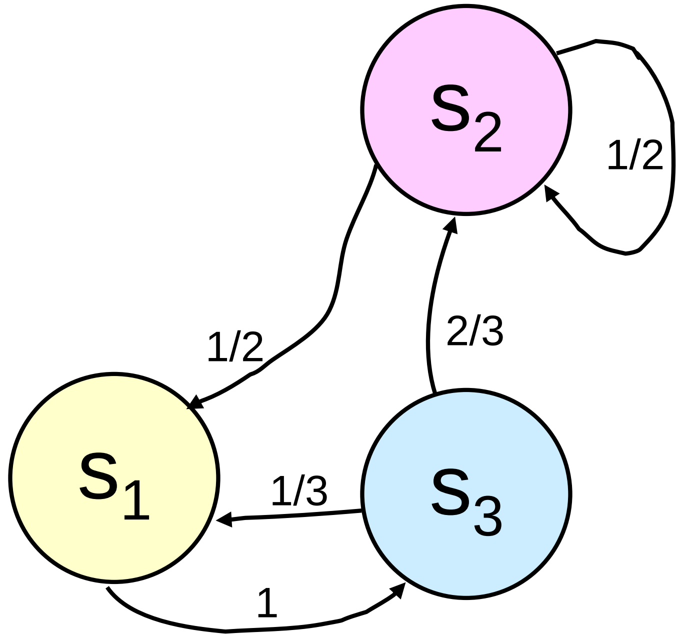



$$P = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 1/2 & 1/2 & 0 \\ 1/3 & 2/3 & 0 \end{bmatrix}$$

- Question: $P^2$ and $P^n$ (will discuss later)

# Markov Chain Components

1. a finite set of $N$ states, S = { $S_1, \cdots, S_N$ }
2. a state transition probability, $P = \{a_{ij}\}_{M \times M}$ , $1 \le i, j \le M$
3. an initial state probability distribution, $\pi = \{\pi_i\}$

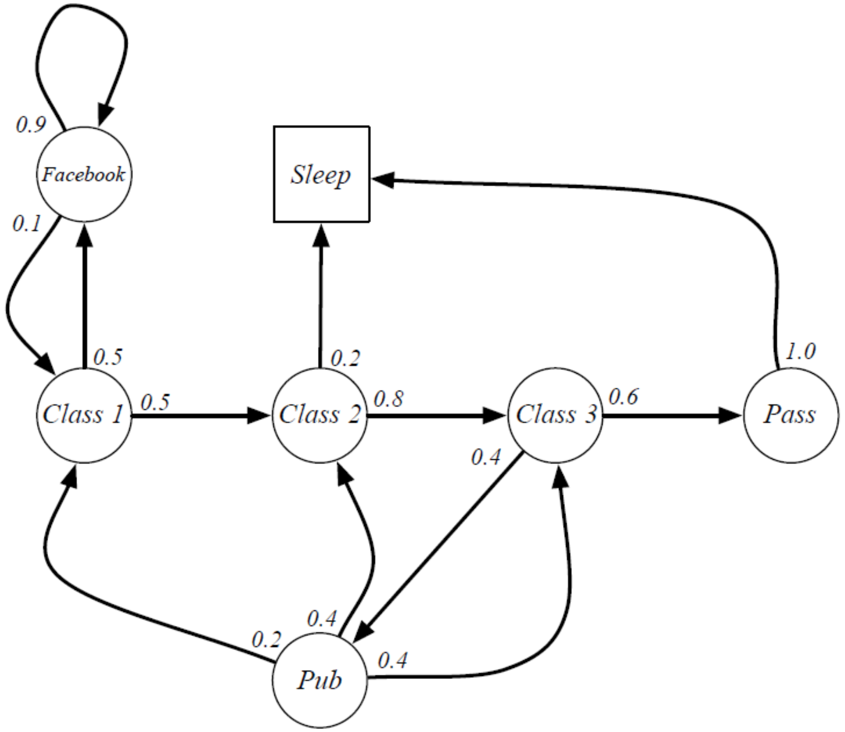



$$P = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 1/2 & 1/2 & 0 \\ 1/3 & 2/3 & 0 \end{bmatrix}$$

- Passive stochastic behavior

# Student Markov Chain Episodes

- Starting from $S_1$ = Class 1




- C1 C2 C3 Pass Sleep
- C1 FB FB C1 C2 Sleep
- C1 C2 C3 Pub C2 C3 Pass Sleep
- C1 FB FB C1 C2 C3 Pub C1 FB FB FB C1 C2 C3 Pub C2 Sleep

$$
\mathcal{P} = \begin{array}{c} \\ C1 \\ C2 \\ C3 \\ Pass \\ Pub \\ FB \\ Sleep \end{array}
\begin{array}{c}
\begin{array}{ccccccc} C1 & C2 & C3 & Pass & Pub & FB & Sleep \end{array} \\
\left[\begin{array}{ccccccc}
 & 0.5 & & & & 0.5 & \\
 & & 0.8 & & & & 0.2 \\
 & & & 0.6 & 0.4 & & \\
 & & & & & & 1.0 \\
0.2 & 0.4 & 0.4 & & & & \\
0.1 & & & & & 0.9 & \\
 & & & & & & 1
\end{array}\right]
\end{array}
$$

# Chapman-Kolmogorov Equation

- (1-step transition probabilities) For a Markov chain on a finite state space, S = { $S_1, \cdots, S_N$}, with transition probability matrix $P$ and initial distribution $\pi = \left\{\pi_i^{(0)}\right\}$ (row vector) then the distribution of $X(1)$ is given by

$$\begin{bmatrix} \pi_1^{(1)} & \pi_2^{(1)} & \pi_3^{(1)} \end{bmatrix} = \begin{bmatrix} \pi_1^{(0)} & \pi_2^{(0)} & \pi_3^{(0)} \end{bmatrix} \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix}$$

# Chapman-Kolmogorov Equation

- (2-step transition probabilities) For a Markov chain on a finite state space, S = { $S_1, \cdots ,S_N$}, with transition probability matrix $P$ and initial distribution $\pi = \left\{\pi_i^{(0)}\right\}$ (row vector) then the distribution of $X(2)$ is given by

$$\begin{bmatrix}\pi_1^{(2)} & \pi_2^{(2)} & \pi_3^{(2)}\end{bmatrix} = \begin{bmatrix}\pi_1^{(1)} & \pi_2^{(1)} & \pi_3^{(1)}\end{bmatrix} P = \begin{bmatrix}\pi_1^{(0)} & \pi_2^{(0)} & \pi_3^{(0)}\end{bmatrix} P^2$$

# Chapman-Kolmogorov Equation

- (n-step transition probabilities) For a Markov chain on a finite state space, S = { $S_1$, ⋯ ,$S_N$}, with transition probability matrix $P$ and initial distribution $\pi = \left\{\pi_i^{(0)}\right\}$ (row vector) then the distribution of $X(n)$ is given by

$$\left[\pi_1^{(n)} \quad \pi_2^{(n)} \quad \pi_3^{(n)}\right] = \left[\pi_1^{(n-1)} \quad \pi_2^{(n-1)} \quad \pi_3^{(n-1)}\right] P = \left[\pi_1^{(0)} \quad \pi_2^{(0)} \quad \pi_3^{(0)}\right] P^n$$

- $P^n$: n-step transition probabilities

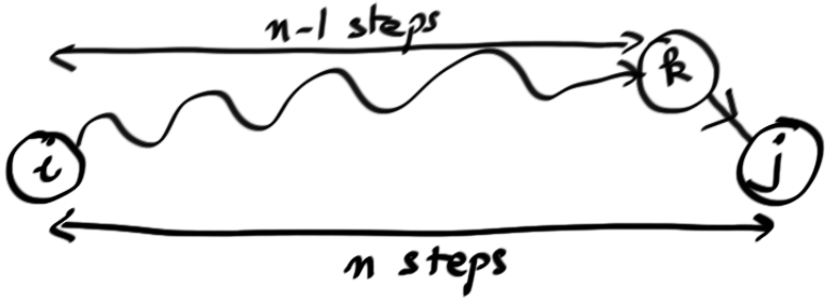
# n-step Transition Probability

- $p_{ij}(n) = P[X_n = j | X_0 = i]$
- $p_{ij} = p_{ij}(1) = P[X_1 = j | X_0 = i]$

- Key recursion:

$$p_{ij}(n) = \sum_{k=1}^{N} p_{ik}(n-1)\, p_{kj}(1)$$

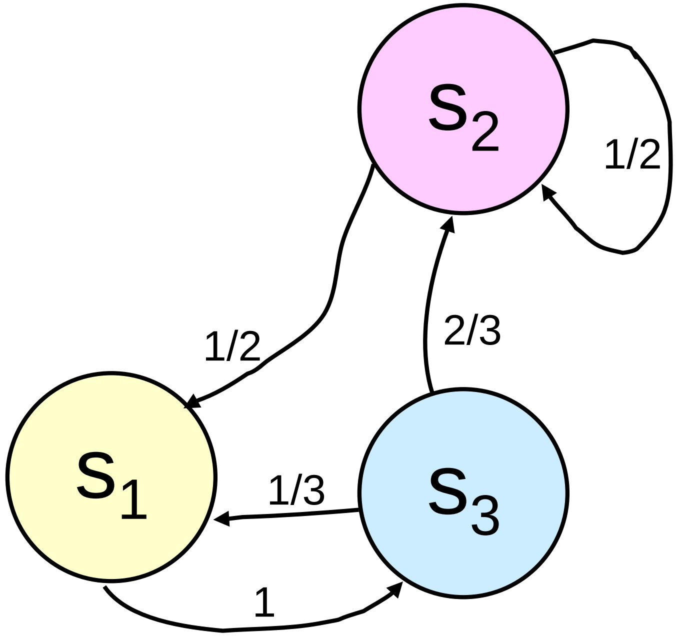
$$i \to k \text{ and } k \to j \text{ imply } i \to j$$



# Example

| | n = 1 | n = 2 | n = 3 |
|---|---|---|---|
| $p_{11}(n)$ | | | |
| $p_{12}(n)$ | | | |
| $p_{13}(n)$ | | | |

$$\begin{bmatrix}\pi_1 & \pi_2 & \pi_3\end{bmatrix} = \begin{bmatrix}1 & 0 & 0\end{bmatrix}$$



# Stationary Distribution

- Steady-state behavior
- Does $p_{ij}(n) = P[X_n = j | X_0 = i]$ converge to some $\pi_j$?

- Take the limit as $n \to \infty$

$$p_{ij}(n) = \sum_{k=1}^{N} p_{ik}(n-1)\, p_{kj}$$

$$\pi_j = \sum_{k=1}^{N} \pi_k\, p_{kj}$$

- Need also $\sum_j \pi_j = 1$

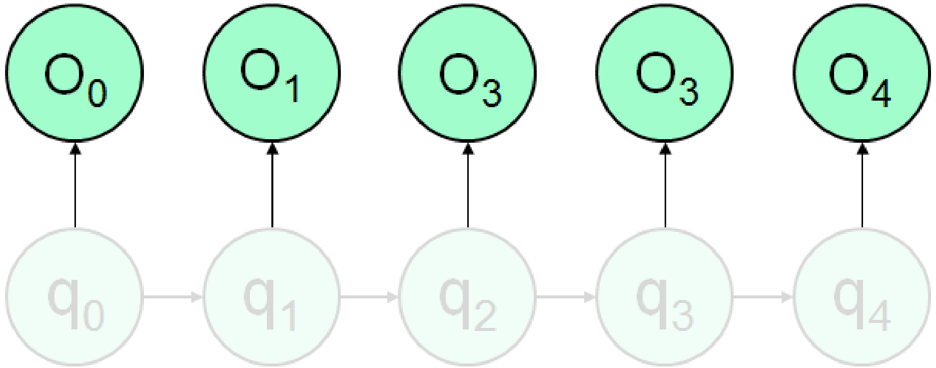
$$\boxed{\pi = \pi P}$$

- How to compute
  - Eigen-analysis
  - Fixed-point iteration

# Hidden Markov Model

# Hidden Markov Model (HMM)

- True state (or hidden variable) follows Markov chain
- Observation emitted from state




- Question: state estimation

What is $p(q_t = s_i \mid O_1, O_2, \cdots, O_T)$

- HMM can do this, but with many difficulties

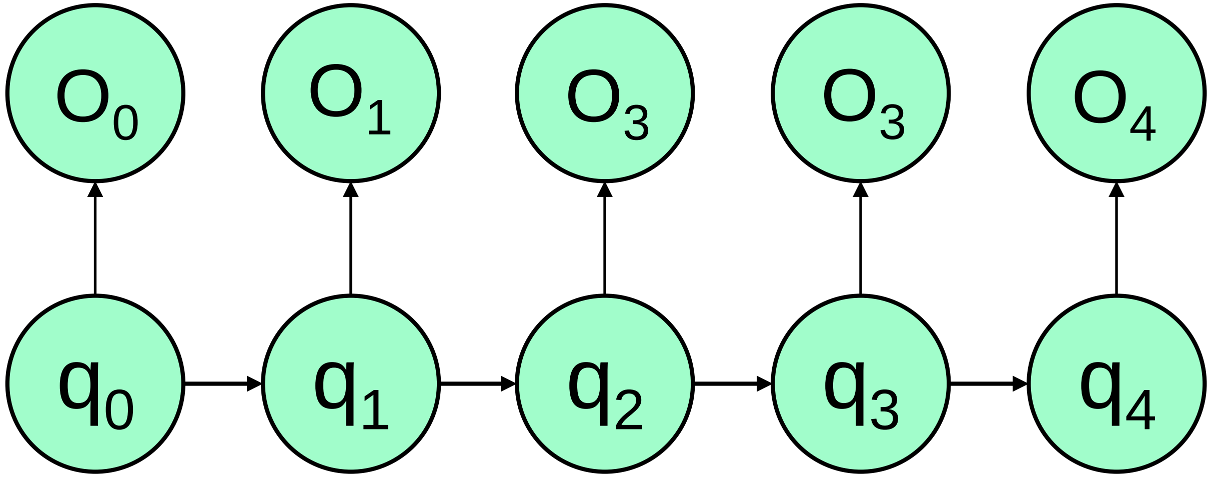
# Hidden State

- Assumption
  - We can observe something that's affected by the true state
  - Natural way of thinking
- Limited sensors (incomplete state information)
  - But still partially related
- Noisy sensors
  - Unreliable

- Observation emitted from $q_t$
  - $O_t$ is noisily determined depending on the current state $q_t$
  - Assume that $O_t$ is conditionally independent of $\{q_{t-1}, q_{t-2}, \dots, q_0, O_{t-1}, O_{t-2}, \dots, O_0\}$ given $q_t$

# Markov Property

1. a finite set of $N$ states, S = $\{ S_1, \cdots ,S_N\}$
2. a state transition probability, $\mathrm{P}=\{a_{ij}\}_{M\times M}$ , $1 \leq i,j \leq M$
3. an initial state probability distribution, $\pi$ = $\{ \pi_i \}$
4. an observation symbol probability distribution, $b_j(O(n))$




Very simplified HMM

# Hidden Markov Models

- Question 1: State Estimation

  What is $P(q_t = Si | O_1 O_2 \cdots O_T)$

Interested for us
- Current state estimation given sequence of observations

- Question 2: Most Probable Path

  Given $O_1 O_2 \cdots O_T$, what is the most probable path that I took? And what is that probability?

- Question 3: Learning HMMs

  Given $O_1 O_2 \ldots O_T$, what is the maximum likelihood HMM that could have produced this string of observations?

# Kalman Filter

# Low-pass Filter in Time

New data $x_k$ comes in

$$\bar{x}_k = \frac{x_1 + x_2 + \cdots + x_{k-1} + x_k}{k}$$

Recursive

$$\begin{aligned}\bar{x}_k &= \frac{k-1}{k}\bar{x}_{k-1} + \frac{1}{k}x_k \\ &= \alpha\bar{x}_{k-1} + (1-\alpha)x_k, \qquad \alpha = \frac{k-1}{k}\end{aligned}$$

# Moving Average Filter

- Use only the latest $n$ data points

$$\bar{x}_k = \frac{x_{k-n+1} + x_{k-n+2} + \cdots + x_k}{n}$$

$$\bar{x}_{k-1} = \frac{x_{k-n} + x_{k-n+1} + \cdots + x_{k-1}}{n}$$

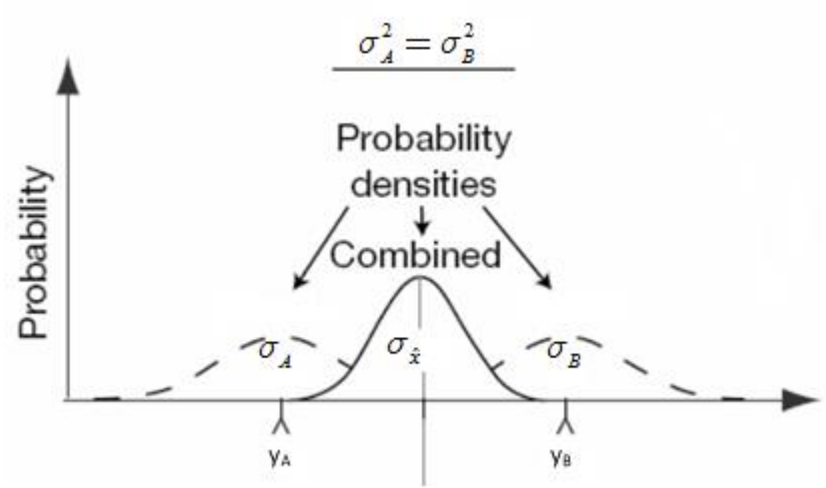
$$\bar{x}_k - \bar{x}_{k-1} = \frac{x_{k-n+1} + x_{k-n+2} + \cdots + x_k}{n} - \frac{x_{k-n} + x_{k-n+1} + \cdots + x_{k-1}}{n} = \frac{x_k - x_{k-n}}{n}$$

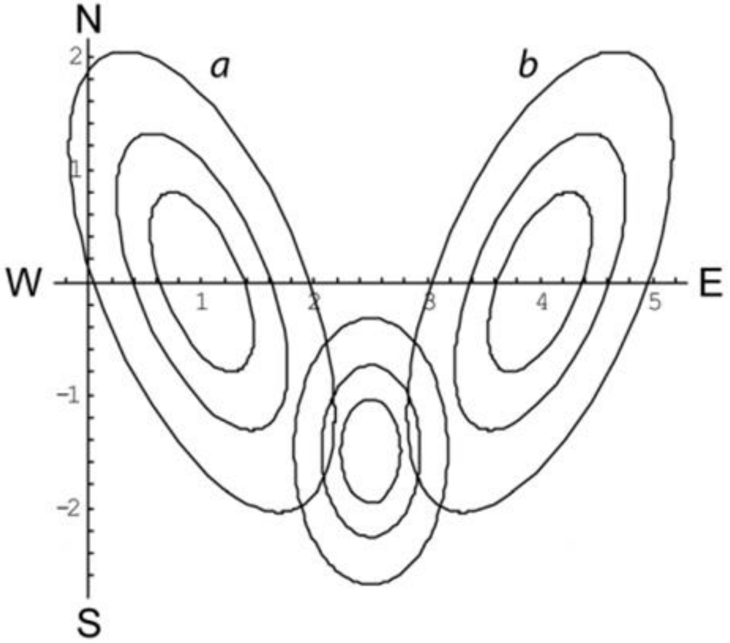
$$\bar{x}_k = \bar{x}_{k-1} + \frac{x_k - x_{k-n}}{n}$$

# Exponentially Weighted Average Filter

$$
\begin{aligned}
\bar{x}_k &= \alpha\bar{x}_{k-1} + (1-\alpha)x_k \\
&= \alpha(\alpha\bar{x}_{k-2} + (1-\alpha)x_{k-1}) + (1-\alpha)x_k \\
&= \alpha^2\bar{x}_{k-2} + \alpha(1-\alpha)x_{k-1} + (1-\alpha)x_k \\
&\vdots \\
&= (1-\alpha)(1x_k + \alpha x_{k-1} + \alpha^2 x_{k-2} + \cdots)
\end{aligned}
$$

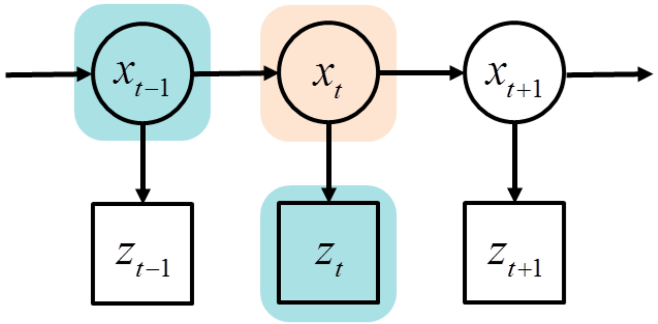
# Sensor Fusion (Two Measured Observations)



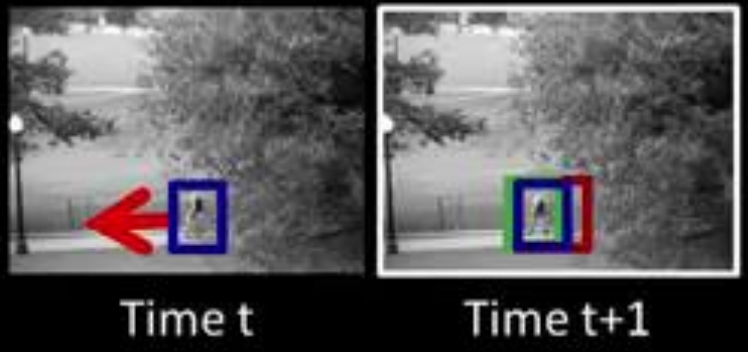
# Kalman Filter

- Linear dynamical system of motion




$$
\begin{aligned}
x_{t+1} &= Ax_t + Bu_t \\
z_t &= Cx_t
\end{aligned}
$$

- *A, B, C* ?


Time t
Time t+1


Tracking with KFs: Gaussians
Initial (prior) estimate
Y
X