Hidden Markov Models

$X_0, X_1, X_2, \ldots$ (random variables) is a Markov chain if

$$\underbrace{Pr(X_{t+1} \mid X_0, \ldots, X_t) = Pr(X_{t+1} \mid X_t)}.$$

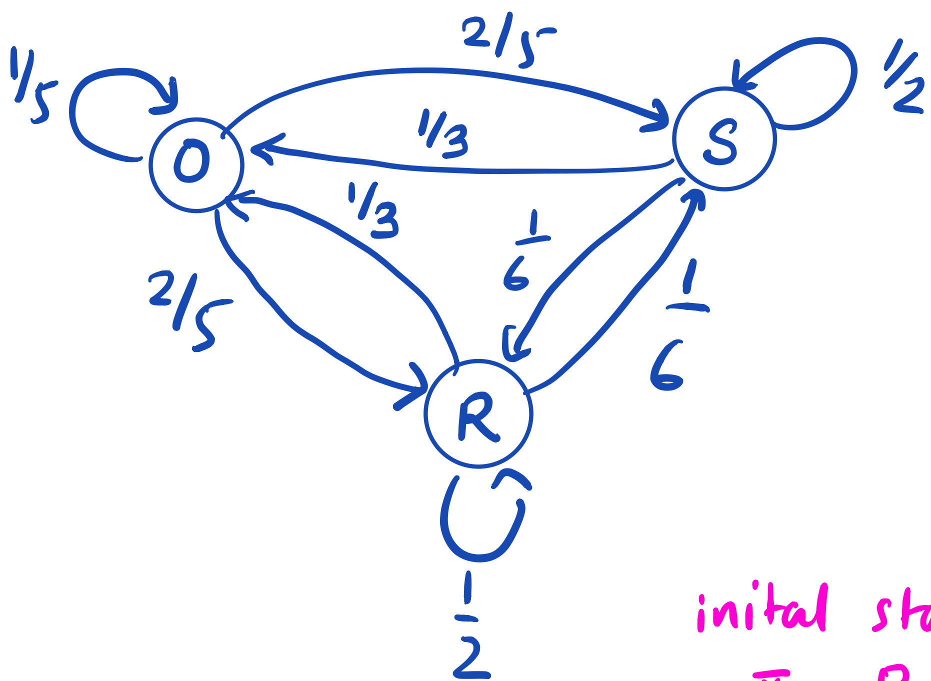
Future independent of past, given the present.

Finite State MC :- N :- # states.

$X_t \in \{s_1, \ldots, s_N\}$, $\quad p_t(j) = Pr(X_t = s_j)$ — ①

$a_{ij} = Pr(X_{t+1} = s_j \mid X_t = s_i)$ — ②

EX :-



$N = 3$,
$\{O, S, R\}$.
$\{s_1, s_2, s_3\}$
$Pr(S \mid R) = \frac{1}{6}$.
...

initial state distribution
$\pi_j = Pr(X_0 = s_j)$.

1. **GOAL** :- Given a_{ij}, and $\boxed{\pi}$, what is $Pr(X_t = s_j)$?

$$Pr(X_{t+1}=s_j) = \sum_{\ell=1}^{N} Pr(X_{t+1}=s_j,\ X_t=s_\ell)$$

$$= \sum_{\ell=1}^{N} Pr(X_t=s_\ell)\cdot Pr(X_{t+1}=s_j \mid X_t=s_\ell)$$

$$= \sum_{\ell=1}^{N} \left(Pr(X_t=s_\ell)\cdot a_{\ell j}\right)$$

```
Prob(t, j)
  { if t = 0
        return π_j ;

    else
        sum = Σ_{ℓ=1}^{N} a_{ℓj} · Prob(t-1, ℓ)
        return sum
  }
```

```
fib(n)
 { if n=0 v n=1
       return 1
   else
     return
       fib(n-1)+fib(n-2)
 }
```

$$Prob(X_{t+1}=s_j) = \sum_{\underline{\underline{X_0}} \cdots \underline{\underline{X_t}}} \underbrace{Pr(X_{t+1}=s_j, X_t, \ldots, X_0)}_{N^t}$$

Dynamic Programming

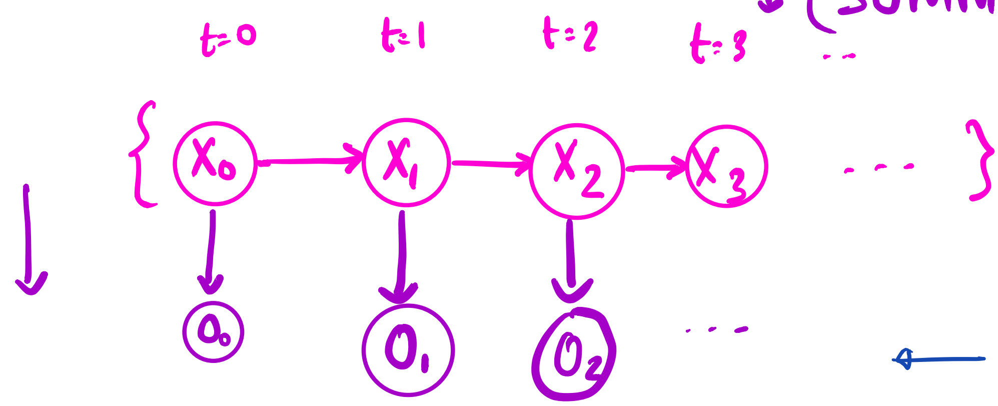
Recall, $p_{t+1}(j) = \sum_{l=1} a_{lj} \cdot \underbrace{p_t(l)}$.

- Compute distb^n @ time t+1, we can remember distb^n @ time t.

$$\left\{ \begin{array}{l} p_0(s) = \pi_s \\ \text{for } i = 1, \ldots, t+1 \\ \quad \text{for } s = 1, \ldots, N \\ \quad\quad \underline{\underline{p_i(s)}} = \sum_{l=1}^{N} p_{i-1}(s) \cdot a_{ls} \end{array} \right. \longrightarrow \boxed{N^2 t \ \underline{\text{vs}} \ N^t}$$

Hidden Markov Models.

(Murphy Ch 17), (Rabiner - tutorial)
↓ (summer to read list).



① $X_0, X_1, \ldots$ a MC. (hidden, a_{ij}, π)...

$O_0, O_1, O_2, \ldots$ ← get to see (observations).

EX :- • speech • communication ...

② $Pr(O_t \mid X_t, X_0 \ldots X_{t-1}, O_0, \ldots, O_{t-1}) = Pr(O_t \mid X_t)$.

N - # states $\qquad O_1 \ldots \in \{\theta_1, \ldots \theta_M\}$.

Let $b_i(j) = Pr(O_t = \theta_j \mid X_t = s_i)$

HMM. $\boxed{\{ \{s_1 \ldots s_N\}, \{\theta_1 \ldots \theta_M\}, a_{ij}, b_i(j), \pi \}}$ ← Model

1. State prob. estimation.

Estimate $Prob(X_t = s_j \mid O_1 \ldots O_t)$?

2. Most Probable Path (MPP). (viterbi Algorithm) (CDMA, auto subtitles)

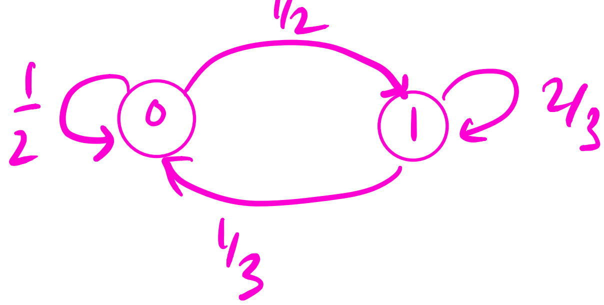
Given $O_1, \ldots O_T$,

$$\max_{s_0 s_1 \ldots s_T} Prob(X_0 = s_0, X_1 = s_1 \ldots X_T = s_T \mid O_1 \ldots O_T).$$

3. Learn the HMM (BW).

$$\max_{\text{models}} Pr(O_1 \ldots O_T \mid \text{model}).$$

EXAMPLE :- $\qquad N = 2. \qquad M = 11$



$\{-5, -4, \ldots, 5\}$.

- if $X_t = 0$, observe an even # @ random $\{4, -2, 0, 2, 4\}$ – $1/5$.
- if $X_t = 1$, observe an odd # @ random $\{-5, -3, -1, 1, 3, 5\} \to 1/6$.

→ easy

1. State prob. estimation.

Estimate $\text{Prob}(X_{t+1} = s_j \mid O_1 \ldots O_t)$?

$\{s_1 \ldots s_N\}$ $\{\theta_1, \ldots \theta_M\}$.

→ $a_{ij} = \Pr(X_{t+1} = s_j \mid X_t = s_i)$, $b_i(\ell) = \Pr(O_t = \theta_\ell \mid X_t = s_i)$

- $\alpha_t(i) := \text{Prob}(O_1, \ldots O_t, X_t = s_i)$, $\sum_{i=1}^{N} \alpha_t(i) = \text{Prob}(O_1 \ldots O_t)$.

$$\alpha_{t+1}(j) = \text{Prob}(O_1 \ldots O_{t+1}, X_{t+1} = s_j)$$

$$= \sum_{i=1}^{N} \text{Prob}(O_1 \ldots O_t, X_t = s_i, O_{t+1}, X_{t+1} = s_j)$$

$$= \sum^{n} \text{Prob}(O_1 \ldots O_t, X_t = s_i) \cdot \Pr(X_{t+1} = s_j \mid X_t = s_i)$$

$$\cdot \Pr(O_{t+1} \mid X_{t+1} = S_j)$$

$$= \boxed{\sum_{i=1}^{N} \alpha_t(i) \cdot a_{ij} \cdot b_j(O_{t+1})} .$$

1. Compute $\alpha_t(1), \dots, \alpha_t(N)$ (time $= N^2 t$)

2. $$\Pr(X_{t+1} = S_j \mid O_0 \dots O_{t+1}) = \frac{\alpha_{t+1}(j)}{\alpha_{t+1}(1) + \dots + \alpha_{t+1}(N)} \quad \text{— ①}$$

$$= \frac{\Pr(O_1 \dots O_{t+1}, X_{t+1} = S_j)}{\sum_{i=1}^{N} \Pr(O_0 \dots O_{t+1}, X_{t+1} = S_i)}$$

---

2. Most Probable Path (MPP).

Given $O_1, \dots O_T$,

$$\max_{S_0 S_1 \dots S_T} \text{Prob}(X_0 = S_0, X_1 = S_1 \dots X_T = S_T \mid O_1 \dots O_T).$$

$$\arg\max_{X_0 \dots X_T} \Pr(X_0 \dots X_T \mid \overbrace{O_0 \dots O_T}) \longrightarrow$$

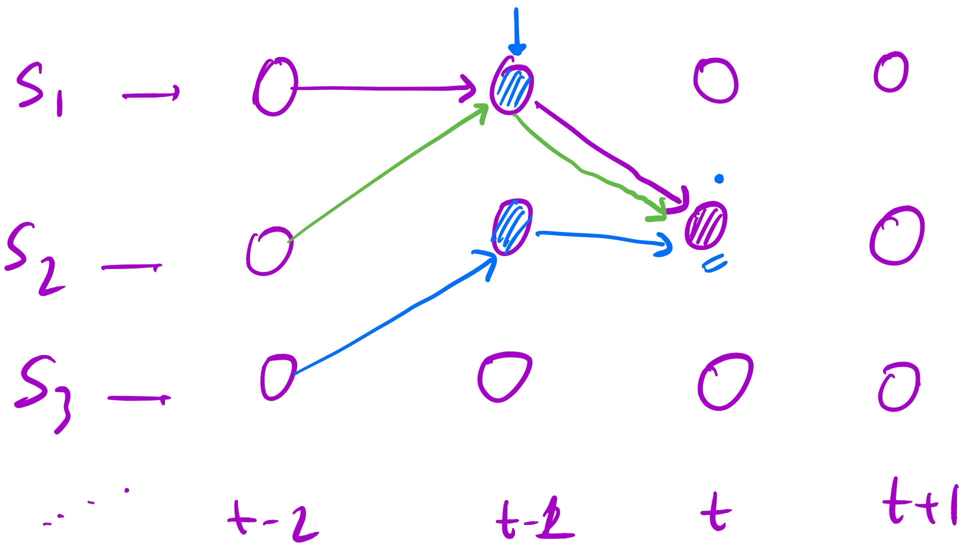
$$= \arg\max_{X_0 \dots X_T} \frac{\text{Prob}(O_0 \dots O_T \mid X_0 \dots X_T) \cdot \Pr(X_0 \dots X_T)}{\cancel{\boxed{\Pr(O_0 \dots O_T)}}}$$

$$= \arg\max_{X_0 \dots X_T} \text{Prob}(O_0 \dots O_T \mid X_0 \dots X_T) \cdot \Pr(X_0 \dots X_T)$$

Let $\rightarrow \delta_t(i) = \max_{x_0 \dots x_{t-1}} \Pr(\underbrace{x_0 \dots x_{t-1}}, x_t = s_i, o_1, \dots, o_t)$

$\boxed{x_t = s_i}$

$N = 3$



$\delta_t(2)$