Lecture-7

* **POMDP review** –

$$M = (S, A, \Omega, O, P, R, d_0, \gamma)$$

→ Ω = set of possible observations that agent can make

O = Observation function

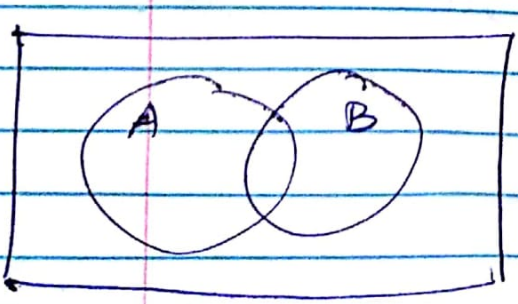
$$O : S \times \Omega \Rightarrow [0, 1]$$

$$O(s, \omega) = P_r(\omega_t = \omega \mid S_t = s)$$

→ Observation at time t

* **Conditional Probability** –

$$P(A = a \mid B = b) = \frac{P(A = a, B = b)}{P(B = b)}$$



$$E[A \mid B = b] = \sum_{a} a \cdot P_r(A = a \mid B = b)$$

Marginalization —

A = sum of two die-roll
B = first die-roll

$$Pr(A=a) = \sum_{b} Pr(B=b) \cdot Pr(A=a \mid B=b)$$

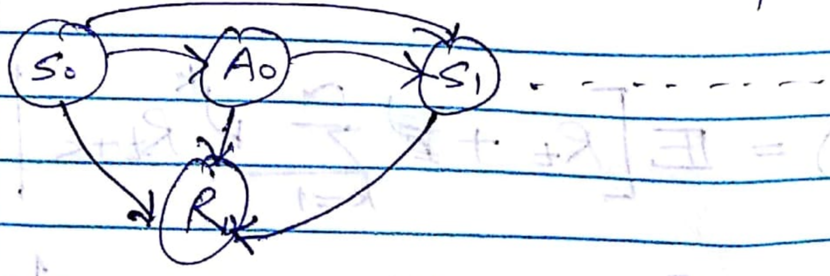
Baye's Theorem —

$$Pr(A=a \mid B=b) = \frac{Pr(A=a, B=b)}{P(B)}$$

$$Pr(B=b \mid A=a) = \frac{Pr(B=b, A=a)}{Pr(A)}$$

$$\therefore \; Pr(A=a \mid B=b) = \frac{Pr(B=b \mid A=a)\, Pr(A=a)}{Pr(B=b)}$$

Baye's Net = (Deseparation)



- Conditional independence:

if X is independant of Z

given Y then,

$$\Pr(X = x \mid Y = y, Z = z)$$
$$= \Pr(X = x \mid Y = y)$$

<u>deseparation</u> : X & Z are deseparated by evidence variables Y iff every undirected path from X to Z is blocked

path blocked if : $\exists\, v \in Y$

① $\leftarrow v \rightarrow$

② $\rightarrow v \rightarrow$

<u>Bellman equation</u> :

$$V^{\pi}(s) = \mathbb{E}\left[R_t + \sum_{k=1}^{\infty} \gamma^k R_{t+k} \,\middle|\, S_t = s, \pi\right]$$

$$= \mathbb{E}\left[R_t + \gamma \sum_{k=0}^{\infty} \gamma^k R_{t+k+1} \,\middle|\, S_t = s, \pi\right]$$

$$= \sum_a \Pr(A_t = a \mid S_t = s, \pi)\, \mathbb{E}\left[R_t + \gamma \sum_{k=0}^{\infty} \gamma^k R_{t+k+1} \,\middle|\, S_t = s,\right.$$

$=$

$$\sum_{a} \pi(s,a) \sum_{s'} Pr\left(s_{t+1}=s' \mid s_t=s, A_t=a, \pi\right)$$

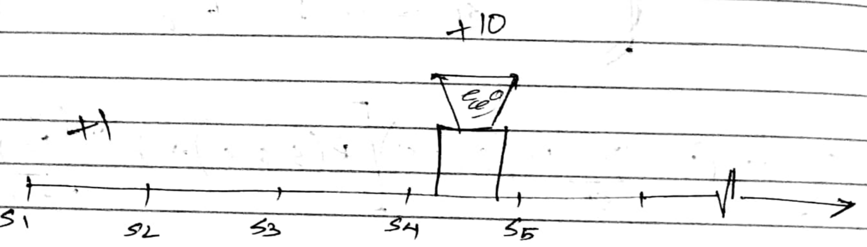
$$\times \left(\mathbb{E}\left[R_t \mid s_t=s, A_t=a, s_{t+1}=s', A_{t+1}=a \right] + \mathbb{E}\left[\gamma \sum_{t=0}^{\infty} \gamma^k R_{t+1+k} \,\middle|\, \begin{matrix} s_t=s \\ a_t=a \\ s_{t+1}=s' \\ a_{t+1}=a \end{matrix} \right] \right)$$

$$= \sum_{a} \pi(s,a) \sum_{s'} P(s,a,s') \left[R(s,a,s') + \gamma \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^k R_{t+1} \,\middle|\, \begin{matrix} s_{t+1}=s' \\ a_{t+1}=a' \end{matrix} \right] \right]$$

$$V^{\pi}(s')$$

$$\therefore = \sum_{a} \pi(s,a) \sum_{s'} P(s,a,s') \left[R(s,a,s') + \gamma V^{\pi}(s') \right]$$

* Example



$$V^{\pi}(s_5) = 0$$

$$V^{\pi}(s_4) = 10$$

$$V^{\pi}(s_1) = R(s_1, a, s_2) + \gamma V^{\pi}(s_2)$$

$$= 1 + 5/4$$

\#

* Optimal value function

$$V^* : S \Rightarrow R$$

$$V^*(s) = \max_{\pi \in \Pi} V^{\pi}(s)$$