HW5

Bellman equations for deterministic policy ~~distribution~~

General

$$v_\pi(s) = \sum_a \pi(a|s) \sum_{s'} p(s'|s,a)\left[r(s,a,s') + \gamma v_\pi(s')\right]$$

$$q_\pi(s,a) = \sum_{s'} p(s'|s,a)\left[r(s,a,s') + \gamma \sum_{a'} \pi(a'|s')\, q_\pi(s',a')\right]$$

① Deterministic policy

Under the policy π, each state is mapped to unique action a_s

$$\pi : s \longmapsto a_s$$

Hence, summation over action collapses in singleton.

$$v_\pi(s) = \sum_{s'} p(s'|s,a_s)\left[r(s,a_s,s') + \gamma v(s')\right]$$

$q_\pi(s,a)$ = value when taking action a in state s (action a is arbitrary, not necessarily dictated by policy!) and THEN following policy π!

$$= \sum_{s'} p(s'|s,a)\left[r(s,a,s') + \gamma q(s', a_{s'})\right]$$

Notice: $v_\pi(s) = \sum_a \pi(a|s)\, q(s,a) = q_\pi(s,a_s)$

② Deterministic policy and Transition.

We now have the following deterministic mappings.

$$s \xmapsto{\pi} a_s \quad \text{unique}$$

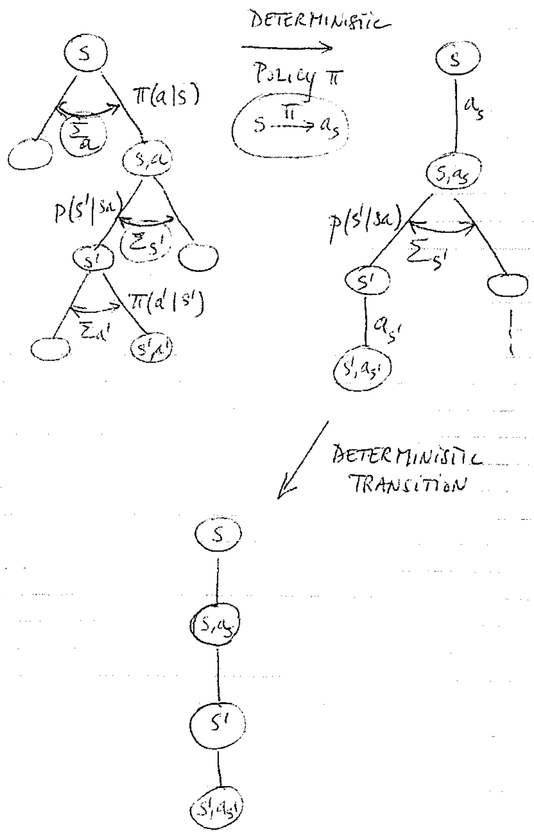
$$\Big\downarrow$$

$$s \xmapsto{a_s} s_{a_s} \quad \text{unique.}$$

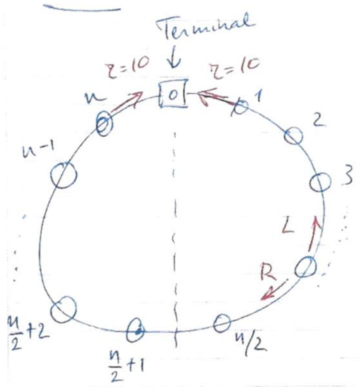
$$v_\pi(s) = r(s, a_s, s_{a_s}) + \gamma v(s_{a_s})$$

$$q_\pi(s, a) = r(s, a, s_a) + \gamma q(s_a, a_{s_a})$$

Backup diagrams



MDP 1 Circular state space



Non-terminal transitions yield reward $r_{NT} = 0$.

n = even
$\Downarrow$
$\frac{n}{2}$ integer.

Node 0 = absorbing : transition yields $r = 10$.

$\gamma = 1$ (no discounting)

π = equiprobable policy ; each action has prob $1/2$
two actions in each node
move clockwise (R)
move counterclockwise (L)

(1) Since transitions btw non-terminal states incur no cost and the agent will eventually end up in absorbing state 0 we conclude:

$$v_\pi(s) = 10 \qquad \forall s$$
$$q_\pi(s,a) = 10 \qquad \forall s, a.$$

② Optimal policy: any policy that ensures eventual absorption in state 0.

$$q^*(s,a) = 10 \qquad v^*(s) = 10.$$

Not unique. (policy).

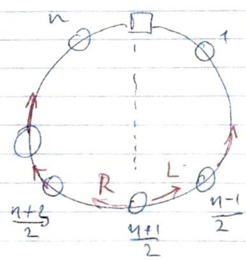
③ If $r_{NT} = -1$: optimal → go to terminal state asap.

π^*: if $s \leq n/2$: action: go L with prob = 1
if $s \geq n/2 + 1$ — go R —

Unique optimal policy.

④ If $\gamma < 1$: go to terminal asap to be optimal.
(similar to 3).

⑤ If n = odd ($r_{NT} = -1$, $\gamma = 1$)



Optimal policy is NO LONGER unique!

if: $s \leq \frac{n-1}{2}$ → go L.

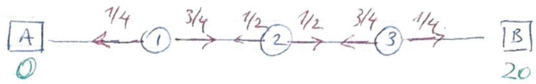
if $s > \frac{n+1}{2}$ → go R

if $s = \frac{n+1}{2}$ one can choose b/w L and R.

8

HW5 ; question 4 : MDP3

(1) State value f^{ion} $v_\pi(s)$ under given policy π



Since the transitions are deterministic we can simplify the Bellman eq:

$$s \xrightarrow{a} s_a$$

$$v_\pi(s) = \sum_a \pi(a|s) \sum_{s'} p(s'|s,a)\left[r(s,a,s') + \gamma v_\pi(s')\right]$$

$$= \sum_a \pi(a|s)\left[r(s,a,s_a) + \gamma v(s_a)\right]$$

Denote: $v_\pi(1) = v_1$, $v_\pi(2) = v_2$, $v_\pi(3) = v_3$

Notice: $v_\pi(A) = 0 = v_\pi(B)$

Then:

$$v_1 = \frac{1}{4}(0+0) + \frac{3}{4}(-2+v_2) = -\frac{3}{2} + \frac{3}{4}v_2$$

$$v_2 = \frac{1}{2}(-2+v_1) + \frac{1}{2}(-2+v_3) = \frac{v_1+v_3}{2} - 2$$

$$v_3 = \frac{3}{4}(-2+v_2) + \frac{1}{4}(20+0) = \frac{3}{4}v_2 + 5 - \frac{3}{2}$$

$$= \frac{3}{4}v_2 + \frac{7}{2}$$

Summing v_1 and v_3:

$$v_1 + v_3 = \left(-\frac{3}{2} + \frac{3}{4} v_2\right) + \left(\frac{3}{4} v_2 + \frac{7}{2}\right)$$

$$= \frac{3}{2} v_2 + 2$$

Substituting this into eq. for v_2:

$$v_2 = \frac{1}{2}(v_1 + v_3) - 2 = \frac{1}{2}\left(\frac{3}{2} v_2 + 2\right) - 2$$

$$= \frac{3}{4} v_2 - 1$$

$$\Rightarrow \boxed{v_2 = -4}$$

$$\Rightarrow \boxed{v_1 = -\frac{3}{2} + \frac{3}{4} v_2 = -\frac{3}{2} + \frac{3}{4}(-4) = -\frac{9}{2}}$$

$$\rightarrow \boxed{v_3 = \frac{3}{4} v_2 + \frac{7}{2} = \frac{3}{4}(-4) + \frac{7}{2} = \frac{1}{2}}$$

(2) Compute state-action value $q_\pi(2,R)$ and $q_\pi(3,L)$

$$q_\pi(2,R) = -2 + v_\pi(3) = -2 + \frac{1}{2} = -3/2$$

$$q_\pi(3,L) = -2 + v_\pi(2) = -2 + \left(-\frac{9}{2}\right) = -\frac{13}{2}$$

(3) Optimal policy: go R in each state.
Unique!

Shapley: Apex game:

5 players → $\underset{\text{Big}}{\underline{1}},\ \underbrace{2, 3, 4, 5}_{\text{small}}$.

$v(B, S_i) = 1$ $\qquad$ $1 = \text{big}$.

$v(S_1 \dots S_4) = 1$

$v = 0$ otherwise.

$v(C) = 1$ if $1 \in C$ & $\#C \geq 2$

$\quad = 1$ if $\#C \geq 4$.

$\quad 0$ otherwise.

<u>Solution 1</u>: all permutations

① If perm $= \overset{1}{|}\ |\ \underbrace{\qquad B \qquad}_{B}\ |\ \overset{5}{|}$ $\qquad$ B in position 2, 3, 4

$\delta_B = 1$

② If perm $= \underset{B}{|\,B\,|}\ \_\_\_\_\_\_\_\_\_\_\_\_ \ |$ $\qquad$ $\# = 4!$
B in position 1

$\delta_B = 0$

③ If perm $= |\ |\ \_\_\_\_\_\_\ |\,B\,|$ $\qquad$ $\# = 4!$
B in position 5

$\delta_B = 0$

Hence

$$\varphi_B = \varphi_1 = \frac{2.4!\,0 + 3.4!\,1}{5!} = \frac{3.4!}{5!} = \frac{3}{5}$$

Because of symmetry the small players have the same value:

$$\sum \varphi_i = v(N) = 1$$

$$\Rightarrow \sum_{i=2}^{5} \varphi_i = 1 - \frac{3}{5} = \frac{2}{5} \Rightarrow \varphi_i = \frac{1}{4} \cdot \frac{2}{5} = \frac{1}{10}$$

$(i = 2, \dots 5)$.

Solution 2

$$\varphi_i = \frac{1}{n} \sum_{S \subset N \setminus i} \binom{n-1}{\# S}^{-1} \delta_i(S)$$

denote $s = \# S$

$i=1$ Big player.

$$s = 0 \Rightarrow S = \phi \qquad \delta_i(S) = 0$$

$$\binom{n-1}{s} = \binom{4}{0} = 4.$$

Solution 2

$$\varphi_i = \frac{1}{n} \sum_{S \subset N \setminus i} \binom{n-1}{\# S}^{-1} \delta_i(S)$$

denote $s = \# S$

$i=1$ <u>Big player.</u>

$s = 0 \Rightarrow S = \phi \qquad \delta_i(S) = 0$

$$\binom{n-1}{s} = \binom{4}{0} = 4 .$$

<u>$s=1$</u> : $S = \underbrace{\{2\} \text{ or } \{3\} \text{ or } \{4\} \text{ or } \{5\}}$

$\delta_i(S) = 1$ $\qquad \binom{n-1}{s}$ sets

w $\qquad \binom{n-1}{s} = \binom{4}{1} = 4$

<u>$s=2$</u> : $S = \{2,3\}, \{2,4\}, \{2,5\}, \{3,4\}, \{3,5\}$
$\{4,5\}$

number of sets $(\# S = 2)$

$\hookrightarrow 6. = \binom{n-1}{2} = \binom{4}{2} = \frac{4.3}{2.1} = 6$

<u>etc.</u>

Conclusion:

δ #sets $\binom{n-1}{s}$

$$\varphi_1 = \varphi_B = \frac{1}{5}\Big[0 \cdot 1 \cdot \frac{1}{1} + 1 \cdot 4 \cdot \frac{1}{4} + 1 \cdot 6 \frac{1}{6}$$

$$+ 1 \cdot 4 \cdot \frac{1}{4} + 0 \cdot 1 \cdot \frac{1}{1}\Big]$$

$$= \frac{1}{5}\big[3\big] = \frac{3}{5}.$$