Q1)

Solⁿ a) Firstly, we ~~can prove~~ need to prove

$$E_\pi\left[E_\pi\left[G_{t+1}|S_{t+1}\right]|S_t\right] = E_\pi\left[G_{t+1}|S_t\right]$$

Let $G_{t+1}=g'$, $S_{t+1}=s'$, $S_t=s$, $G_t=g$

$$\therefore E_\pi\left[E_\pi\left[g'|s'\right]|s\right]$$

$$= E_\pi\left[E_\pi\left[g'|s',s\right]\right]$$

$$= E_\pi\left[\sum_a \pi(a|s) \sum_{g'} g' \, p(g'|s',s)\right]$$

$$= \sum_{s'} \left[\sum_a \pi(a|s) \sum_{g'} g' \, p(g'|s',s)\right] \cdot p(s'|s)$$

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

$$p(g'|s',s) = \frac{p(g',s',s)}{p(s',s)} \quad ; \quad p(s'|s) = \frac{p(s',s)}{p(s)}$$

$$\therefore E_\pi\left[E_\pi\left[g'|s'\right]|s\right] = \sum_a \pi(a|s) \sum_{s'} \sum_{g'} g' \, \frac{p(g',s',s)}{p(s',s)} \, \frac{p(s',s)}{p(s)}$$

$$= \sum_a \pi(a|s) \sum_{s'} \sum_{g'} g' \, p(g'|s)$$

$$= E_\pi\left[g'|s\right]$$

$$= E_\pi\left[G_{t+1}|S_t\right] \quad \text{— (A)}$$

Now, we know,

$$V_\pi(s) = E_\pi\left[r + \gamma \sum_{k=0}^{\infty} \gamma^k R_{t+k+2} \,|\, S_t = s\right]$$

$$= E_\pi\left[r + \gamma G_{t+1} \,|\, S_t = s\right]$$

$$= E_\pi\left[r|s\right] + \gamma E_\pi\left[G_{t+1}|S_t\right]$$

Scanned by CamScanner

$\therefore$ From (A)

$$V_\pi(s) = E_\pi[r|s] + \gamma\, E_\pi[E_\pi[G_{t+1} | S_{t+1}] | S_t]$$

$$= E_\pi[r + \gamma V_\pi(s') | S_t]$$

$$\therefore V_\pi(s) = E_\pi[r + \gamma V_\pi(s') | s]$$

~~This is the state-value function of for Bellman~~

This is the Bellman's equation for state-value f

b) For this, we need to prove firstly,

$$E_\pi[E_\pi[G_{t+1} | S_{t+1}, A_{t+1}] | S_t, A_t] = E_\pi[G_{t+1} | S_t, A_t]$$

Let $G_{t+1} = g'$, $S_{t+1} = s'$, $A_{t+1} = a'$, $G_t = g$; $S_t = s$, $A_t = a$

$$\therefore E_\pi[E_\pi[g' | s', a'] | s, a]$$

$$= E_\pi[E_\pi[g' | s', a', s, a]]$$

$$= E_\pi\left[\sum_a \pi(a|s) \sum_{g'} g'\, p(g' | s', a', s, a)\right]$$

$$= \sum_a \pi(a|s) \sum_{s', a'}\left[\sum g'\, p(g' | s', a', s, a)\right] p(s', a' | s, a)$$

$$p(g' | s', a', s, a) = \frac{p(g', s', a', s, a)}{p(s', a', s, a)}$$

$$p(s', a' | s, a) = \frac{p(s', a', s, a)}{p(s, a)}$$

$$\therefore E_\pi[E_\pi[g' | s', a'] | s, a]$$

$$= \sum_a \pi(a|s) \sum_{s', a'} \sum_{g'} g'\, \frac{p(g', s', a', s, a)}{\cancel{p(s', a', s, a)}} \cdot \frac{\cancel{p(s', a', s, a)}}{p(s, a)}$$

$$= \sum_a \pi(a|s) \sum_{g'} g' \sum_{s',a'} \frac{p(g',s',a'|s,a)}{p(s,a)}$$

$$= \sum_a \pi(a|s) \sum_{g'} g' \; p(g'|s,a)$$

$$= E_\pi [g'| s,a]$$

$$= E_\pi [G_{t+1} | s_t, a_t]$$

Now, we know, $q_\pi(s,a)$

$$q_\pi(s,a) = E_\pi \left[r + \gamma \sum_{k=0}^{\infty} \gamma^k R_{t+k+2} \,\middle|\, S_t = s, A_t = a \right]$$

$$= E_\pi [r + \gamma G_{t+1} | s,a]$$

$$= E_\pi [r|s,a] + \gamma E_\pi [g'|s,a]$$

$$= E_\pi [r|s,a] + \gamma E_\pi [E_\pi[g'|s',a'] | s',a]$$

$$= E_\pi [r + \gamma q_\pi(s',a') | s,a]$$

$$\therefore q_\pi(s,a) = E_\pi [r + \gamma q_\pi(s',a') | S_t = s, A_t = a]$$

This is the Bellman's equation for action-value function.

Q2)

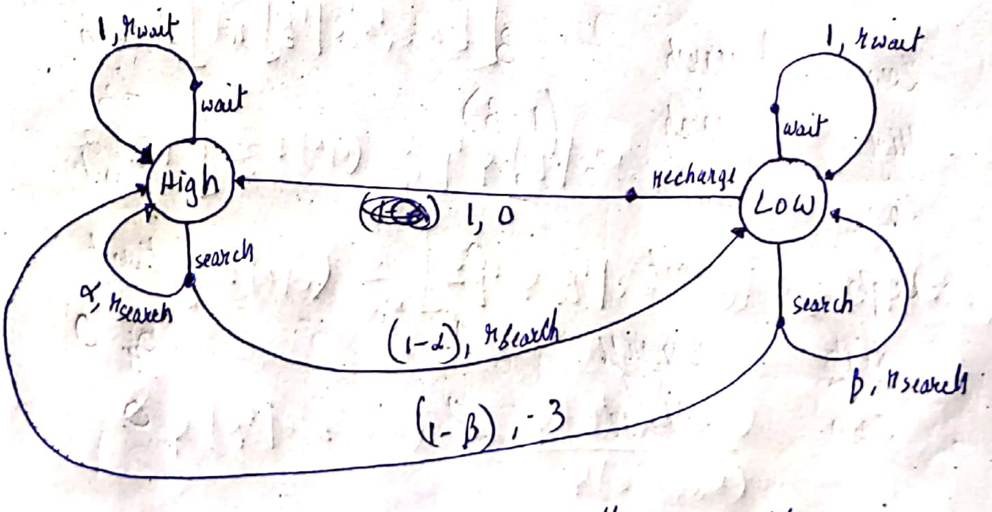
Sol^n:- a) For the can collecting example, the state transition table is as follows -

s	s'	a	$p(s'\|s,a)$	$r(s,a,s')$
High	High	search	α	r_{search}
High	Low	search	$1-\alpha$	r_{search}
High	High	wait	1	r_{wait}
High	Low	wait	0	r_{wait}
Low	Low	search	β	r_{search}
Low	High	search	$(1-\beta)$	-3
Low	Low	wait	1	r_{wait}
Low	High	wait	0	r_{wait}
Low	High	recharge	1	0
Low	Low	recharge	0	0

For the other example, the state transition table is as follows -

s	s'	a	$P(s' \mid s, a)$	$r(s, a, s')$
Good	Good	Stay	$1/2$	$+3$
Good	Bad	~~Move~~ Stay	$1/2$	-1
Good	Good	Move	0	$+3$
Good	Bad	Move	1	-1
Bad	Good	Stay	0	$+3$
Bad	Bad	Stay	1	-1
Bad	Good	Move	1	$+3$
Bad	Bad	Move	0	-1

b) The state space diagram for can collecting bot example is as follows -



The state space diagram for the other example is as follows -