Date
DELTA Pg No.

Q1) Table Explanation

$$E[R_{search}] = r_{search} = P_{find-con} \times 1 + P_{no\ con} \times 0$$

$$= P_{find-can}$$

$$\Rightarrow P_{find-con} = r_{search} = P_{find-con} | S = high, A = search$$

$$p(s', r | s, a) = p(s' | s, a) \times p(r | s, a)$$

$$= \alpha \times r_{search}$$

$$\because p(s' | s, a) = \alpha$$

$$p(s', r=0 | s, a) = \alpha - \alpha \times r_{search}$$

Similar logic followed for the rest of the cases

| | s | a | s' | r | p(s',r\|s,a) |
|---|---|---|---|---|---|
| (1) | high | search | high | 1 | $\alpha r_{search}$ |
| (2) | high | search | high | 0 | $\alpha - \alpha r_{search}$ |
| (3) | high | search | low | 1 | $(1-\alpha) r_{search}$ |
| (4) | high | search | low | 0 | $(1-\alpha) - (1-\alpha) r_{search}$ |
| (5) | high | wait | high | 1 | $r_{wait}$ |
| (6) | high | wait | high | 0 | $1 - r_{wait}$ |
| (7) | low | search | high | -3 | $1-\beta$ |
| (8) | low | search | low | 1 | $\beta r_{search}$ |
| (9) | low | search | low | 0 | $\beta - \beta r_{search}$ |
| (10) | low | wait | low | 1 | $r_{wait}$ |
| (11) | low | wait | low | 0 | $1 - r_{wait}$ |
| (12) | low | recharge | high | 0 | 1 |

Exercise 3.15

$$G_t = R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \dots = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1}$$

$$V(s) = E[G_t \mid S_t = s]$$

Replacing each $R_i$ with $R_i + C$

$$G_t' = \sum_{k=0}^{\infty} \gamma^k (R_{t+k+1} + C)$$

$$= \sum_{k=0}^{\infty} \gamma^k R_{t+k+1} + \sum_{k=0}^{\infty} \gamma^k c$$

$$G_t' = G_t + c \sum_{k=0}^{\infty} \gamma^k$$

$$= G_t + \frac{c}{1-\gamma} \quad \text{[sum of infinite GP]}$$

$$V(s) = E[G_t \mid S_t = s] + E\left[\frac{c}{1-\gamma} \mid S_t = s\right]$$

$$= E[G_t \mid S_t = s] + \frac{c}{1-\gamma}$$

$$V(s) = E[G_t \mid S_t = s] + V_c$$

$$V_c = \frac{c}{1-\gamma}$$

Exercise 3.16

$$G_t = R_{t+1} + \gamma R_{t+2} + \dots + \gamma^{n-1} R_{t+n} = \sum_{k=0}^{n-1} \gamma^k R_{t+k+1}$$

Add constant c to rewards.

$$G'_t = \sum_{k=0}^{n-1} \gamma^k (R_{t+k+1} + c)$$

$$\Rightarrow G'_t = G_t + c \sum_{k=0}^{n-1} \gamma^k$$

$$= G_t + c \left[ \frac{1-\gamma^n}{1-\gamma} \right]$$

With an increasing number of time steps, i.e, as n increases, the value of $G'_t$ ~~decreases~~ increases. ~~proportionately~~

For $n=2$,

$$G'_t = G_t + c \left[ \frac{(1-\gamma)(1+\gamma)}{1-\gamma} \right]$$

$$= G_t + c\{1+\gamma\}$$

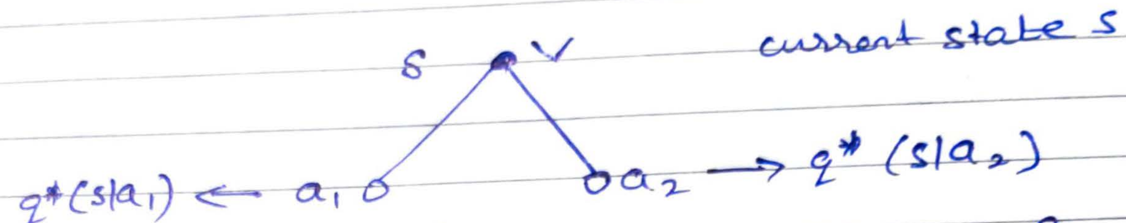
For $n=4$ $$G'_t = G_t + c \left[ \frac{(1-\gamma)(1+\gamma)(1+\gamma^2)}{1-\gamma} \right]$$

$$= G_t + c\,[(1+\gamma)(1+\gamma^2)]$$

In the continuous case, the value added to $G_t$ was always a constant $\frac{c}{1-\gamma}$

**Question 5**

We can visualize this problem as -




set of actions $a_1, a_2, \dots, a_n \in A(s)$

$$v^*(s) = \max_{a \in A(s)} q^*(s|a)$$

P.S → I used • for state and ∘ for action instead of the convention which is the opposite