Date _____

Expt. No. _____ Page No. _____

RL

Homework-3

Ans 1. Monte Carlo Exploring starts, Modified:

Initialize:

1. $\pi(s) \in A(s)$ (arbitrarily), for all $s \in S$
2. $Q(s,a) \in \mathbb{R}$ (arbitrarily), for all $s \in S$, $a \in A(s)$
3. $N(s,a) = 0 \quad \forall\, s \in S,\ a \in A(s)$

Loop forever (for each episode):

4. Choose $S_0 \in S$, $A_0 \in A(S_0)$ randomly such that all pairs have probability > 0
5. Generate an episode from S_0, A_0 following π: $S_0, A_0, R_1, \ldots, S_{T-1}, A_{T-1}, R_{T-1}$
6. $G \leftarrow 0$
7. Loop for each step of episode, $t = T-1, T-2, \ldots, 0$:
8. $G \leftarrow \gamma G + R_{t+1}$
9. ~~Unless the pair S_t, A_t appears in~~ $N(S_t, A_t) = N(S_t, A_t) + 1$
10. $Q(S_t, A_t) = Q(S_t, A_t) + \frac{1}{N(S_t, A_t)} \left(G - Q(S_t, A_t)\right)$
11. $\pi(S_t) \leftarrow \underset{a}{\operatorname{argmax}}\, Q(S_t, a)$

(State, action papers)

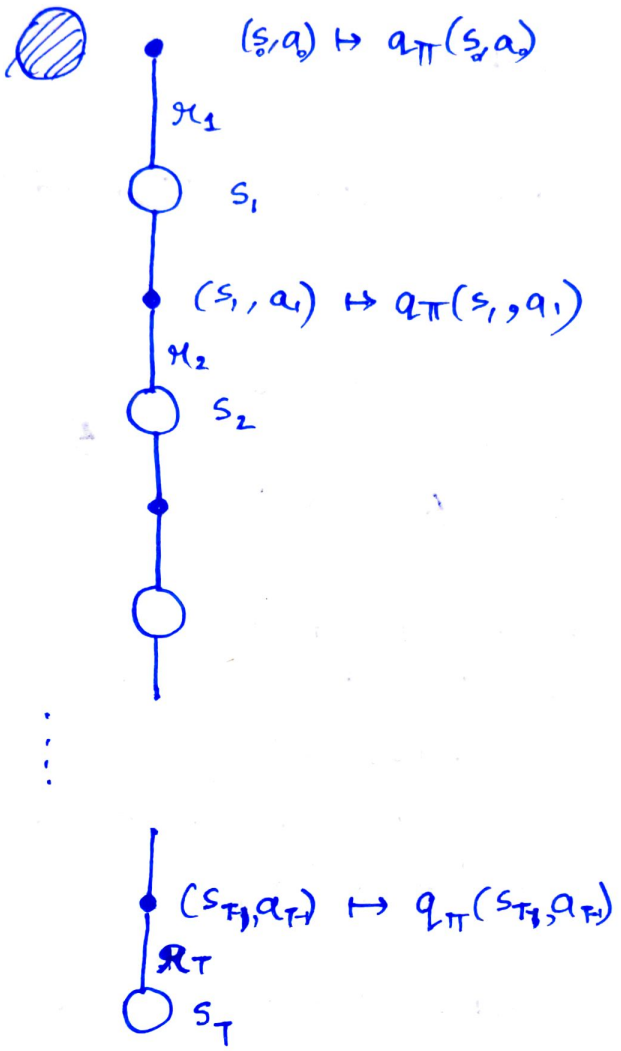
Line 3 initializes the count of $N(s,a)$ $\forall\, s \in S,\ a \in A$

Line 9 increments the count of S_t, A_t by 1.

Line 10 updates the Q value by incremental mean approach

Teacher's Signature : _____

Ans2 Back up diagram for Monte Carlo Estimation of q_π.



Date _____

Expt. No. _____ Page No. _____

Ans 6.3 In the first episode, the terminal state is the left terminal state because of which the value of state A only decreases by $\alpha * (0.5)$ which is 0.45. The estimate of state A changed because the estimated value of terminal state B is 0 while of others is 0.5 and also the reward is 0 in this episode.

Ans 6.4