Mudit Garg
2016057

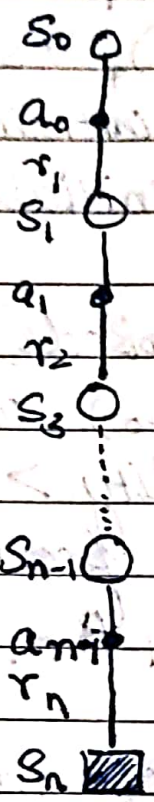
Ans 1)

Initialize:
$\pi(s) \in A(s)$ (arbitrarily), $\forall s \in S$
$Q(s,a) \in \mathbb{R}$ (arbitrarily), $\forall s \in S, \forall a \in A$
Returns_sum$(s,a) \in \mathbb{R}$ (initialize all values to 0)
Returns_count$(s,a) \in \mathbb{R}$ (initialize all values to 0)

loop forever (for each episode):
choose $S_0 \in S, A_0 \in A(S_0)$ randomly such that all pairs have probability > 0
generate an episode from S_0, A_0 following π: $S_0, A_0, R_1, \ldots, S_{T-1}, A_{T-1}, R_T$
$G \leftarrow 0$
loop for each step of the episode, $t = T-1, \ldots, 0$
$G \leftarrow \gamma G + R_{t+1}$
Returns_sum$(S_t, A_t) += G$
Returns_count$(S_t, A_t) += 1$
$Q(S_t, A_t) =$ Returns_sum(S_t, A_t) / Returns_count(S_t, A_t)
$\pi(S_t) = \underset{a}{\text{argmax}}\ (Q(S_t, a))$

Ans 2 $q_\pi(s,a) = r + V_\pi(s',a)$

$= q_\pi(s,a) + \alpha(G_t - q_\pi(s,a))$



Ans 3 $q_\pi(s,a) = E_\pi(G_t \mid S_t = s, A_t = a)$

Now, $Pr(S_{t+1}, A_{t+1} \dots S_T \mid S_t = s, A_t = a, A_{t+1:T-1} \sim \pi)$

$= p(s_{t+1} \mid s,a)\ \pi(A_{t+1} \mid S_{t+1}) \dots \pi(A_{T-1} \mid S_{T-1}) * p(S_T \mid A_{T-1}, S_{T-1})$

$$\Rightarrow P_r = \prod_{k=t+1}^{T-1} \pi(A_k|S_k) P(S_{k+1}|S_k, A_k)$$

$$\Rightarrow \rho_{t+1:T-1} = \prod_{k=t+1}^{T-1} \frac{\pi(A_k|S_k)}{b(A_k|S_k)}$$

Also, $t+1 : T-1$ can be written as $t \in \tau(s,a)$ where $\tau(s,a)$ is trajectory after starting from s, a

$$\Rightarrow q_\pi(s,a) = \frac{\sum_{t \in \tau(s,a)} \rho_{t:T(t)-1} G_t}{\sum_{t \in \tau(s,a)} \rho_{t:T(t)-1}}$$