Ryan Finn
CSC 449
Nov. 30, 2022
Exam II

90%

-5

1. **procedure** SARSA(number of episodes $N \in \mathbb{N}$
discount factor $\lambda \in (0, 1]$
learning rate $\alpha_n = 1 / \lg(n + 1)$)

Initialize matrices $Q(s, a)$ and $n(s, a)$ to 0, $\forall\ s, a$

for episode $k \in 1, 2, 3, \ldots, \cancel{n}N$[1] **do**

$t \leftarrow 1$

Initialize s_t

~~Choose a_t from a uniform distribution over the actions~~

Choose a_t from s_t using μ_t: an ε-greedy policy with respect to Q[2]

while Episode k is not finished **do**

Take action a_t: observe reward r_{t+1} and next state s_{t+1}

Choose a_{t+1} from s_{t+1} using μ_t: an ε-greedy policy with respect to Q

if The current state is terminal **then**

$y_t = 0$

else

$\cancel{y_t = r_{t+1} + \max_a Q(s_{t+1}, a)}$

$y_t = r_{t+1} + \lambda Q(s_{t+1}, a_{t+1})$[3]

endif

$n(s_t, a_t) \leftarrow n(s_t, a_t) + 1$

Update Q function:

$\cancel{Q(s_{t+1}, a_{t+1}) \leftarrow Q(s_t, a_t) - \alpha_{n(st, at)}(y_t - Q(s_t, a_t))}$

$Q(s_t, a_t) \leftarrow Q(s_t, a_t) - \alpha_{n(st, at)}(y_t - Q(s_t, a_t))$[4]

$t \leftarrow t + 1$

end while

end for

end procedure

[1]: A bit of a technicality, but N is actually defined as the number of episodes, not n which is a matrix.
[2]: This probably isn't a necessary change, since Q is already initialized to all 0, so a_1 will just end up as the very first or last action in the action space. But, if Q is ever initialized as a non-zero matrix this change could be important. That's also how SARSA is defined in the book.
[3]: $y_t = r_{t+1} + \max_a Q(s_{t+1}, a)$ is the target updater for a Q-Learning algorithm, not a SARSA algorithm.
[4]: $Q(s_t, a_t)$ is what should be updated, not $Q(s_{t+1}, a_{t+1})$.

2.

a. Greedy deterministic

-5

b. So long as ε-greedy is used with an $\varepsilon > 0$, then yes, the Q values will converge as the number of time steps increases towards infinity. This is because every action must eventually be sampled infinite times, as time increases, for any positive ε, as dictated by the Law of Large Numbers.