

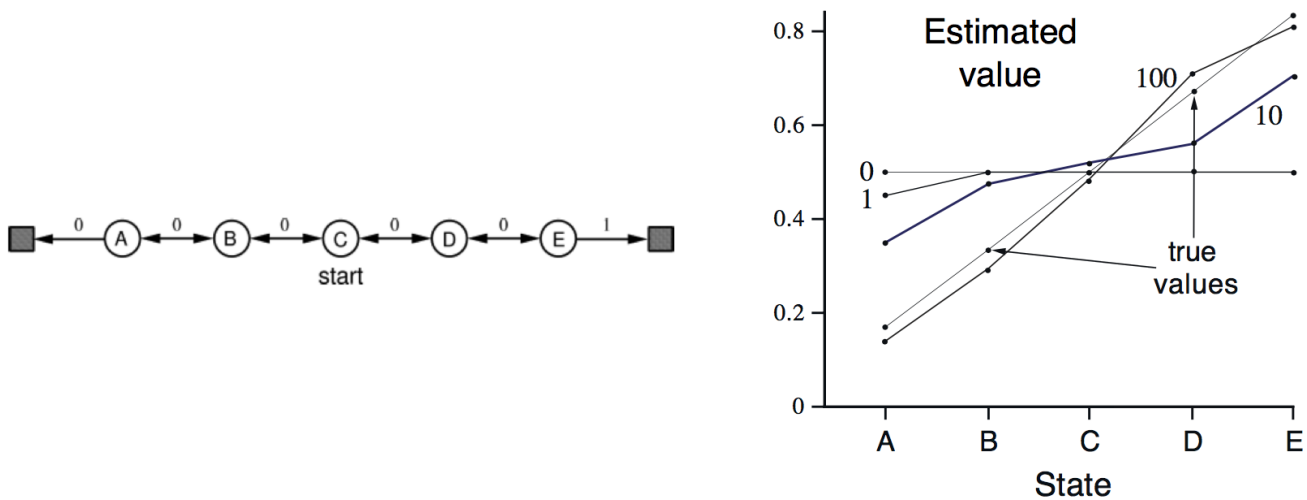
Reinforcement Learning

Exercise 5

Mathias Niepert, Vinh Tong

April 28, 2023

1 Random Walk (2P)



Recall the Random walk example presented in the lecture (lecture 5 slide 12). From the results shown in the right graph above (estimated value) it appears that the first episode results in a change in only $V(A)$. What does this tell you about what happened on the first episode? Why was only the estimate for this one state changed? By exactly how much was it changed (assuming $\alpha = 0.1$)? Support your answers by computing the TD-update.

2 Sarsa and Q-learning on the FrozenLake (8P)

We will again use the FrozenLake environment from gym (https://www.gymnasium.ml/environments/toy_text/frozen_lake/). The code template can be found on Ilias in *ex05-td/ex05-td.py*. It has been tested with gym version 0.18.0 (but should also be stable with version 0.18.3).

- Implement Sarsa and obtain and plot the state-value function, action-value function, and policy for the FrozenLake environment. Plot the average episode length as training continues. (3P)
- Implement Q-learning and obtain and plot the optimal state-value function, action-value function, and policy for FrozenLake. What can you say about performance during training in comparison to the performance of the optimal policy? (3P)
- Explore how your results for a) and b) change if you switch to the non-slippery version (i.e. deterministic environment). (1P)
- Rerun your code for the larger FrozenLake environment. (1P)