Comparison of Q(σ), Tree Backup and Expected Sarsa

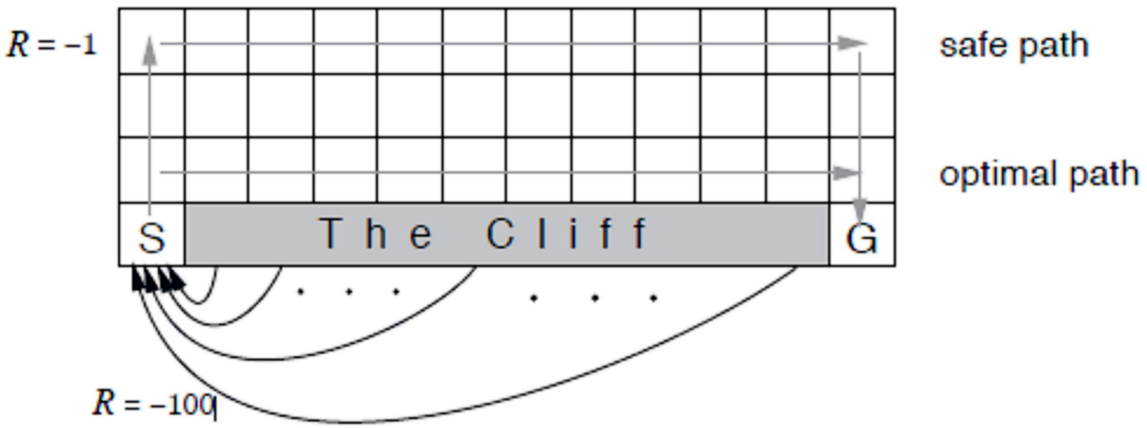
Lucas Berry

Comp 767

February 18th, 2017

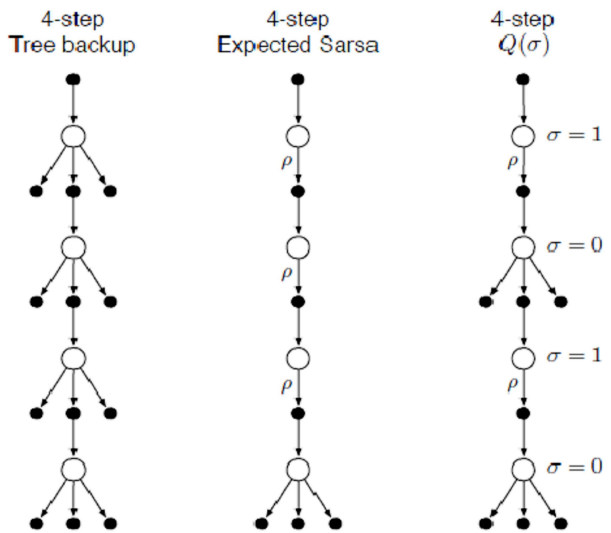
Comparison

These slides contain a comparison of the said methods on the cliff Grid World example on page 141 from Reinforcement Learning: An Introduction by Richard Sutton and Andrew Barto.



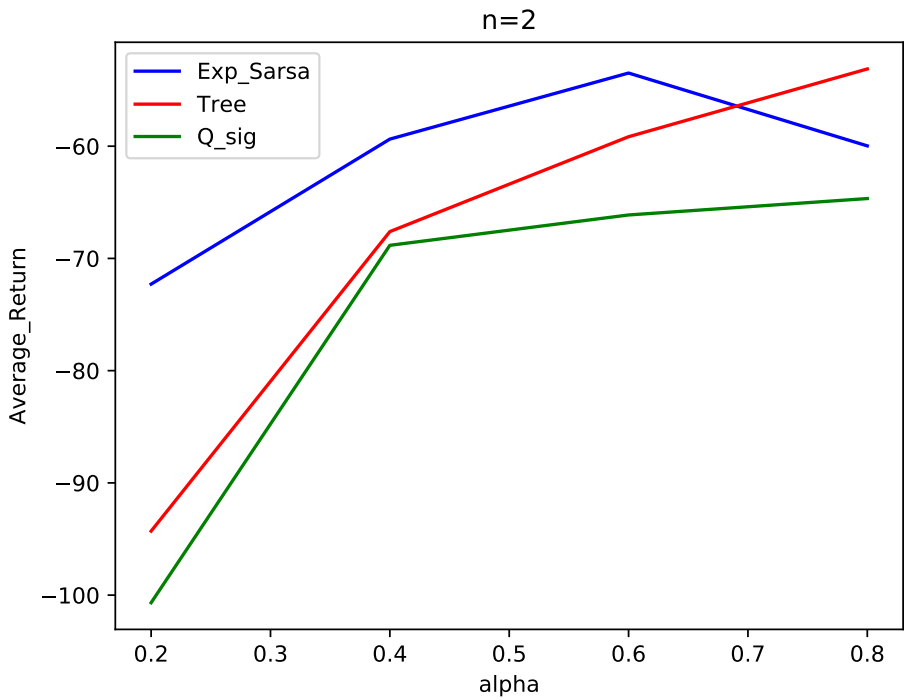
Backup Diagrams

The backup diagrams for the three methods are contained below. Please note that σ refers to amount we sample, with 1 being full sampling and 0 begin the expectation or no sampling. All three algorithms are described in chapter 7 of the previously mentioned book.



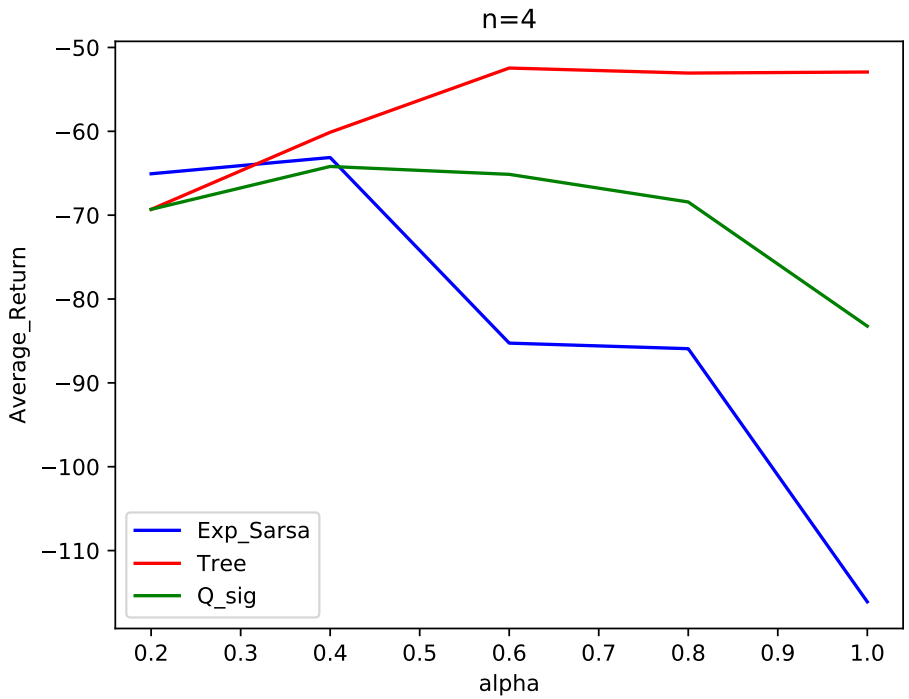
Average Returns

The following graph depicts the average return with varying alphas of the three methods. Note 100 episodes were run 5 times and averaged to obtain the values with $\epsilon = .1$. Please note σ was set to alternating fashion as in the previous diagram.



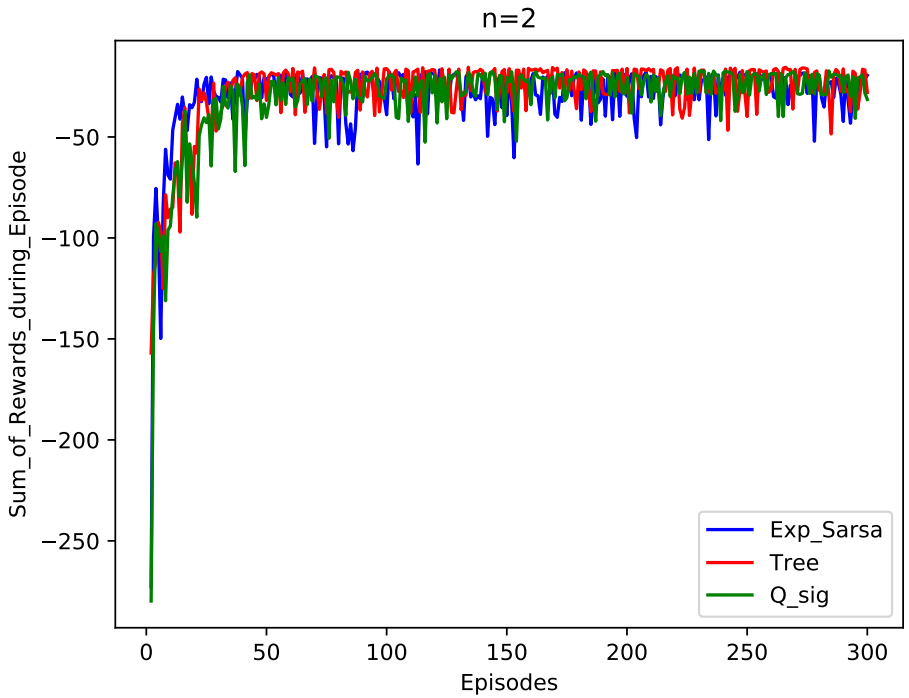
Average Returns

Now changing our n parameter provides the following graph.



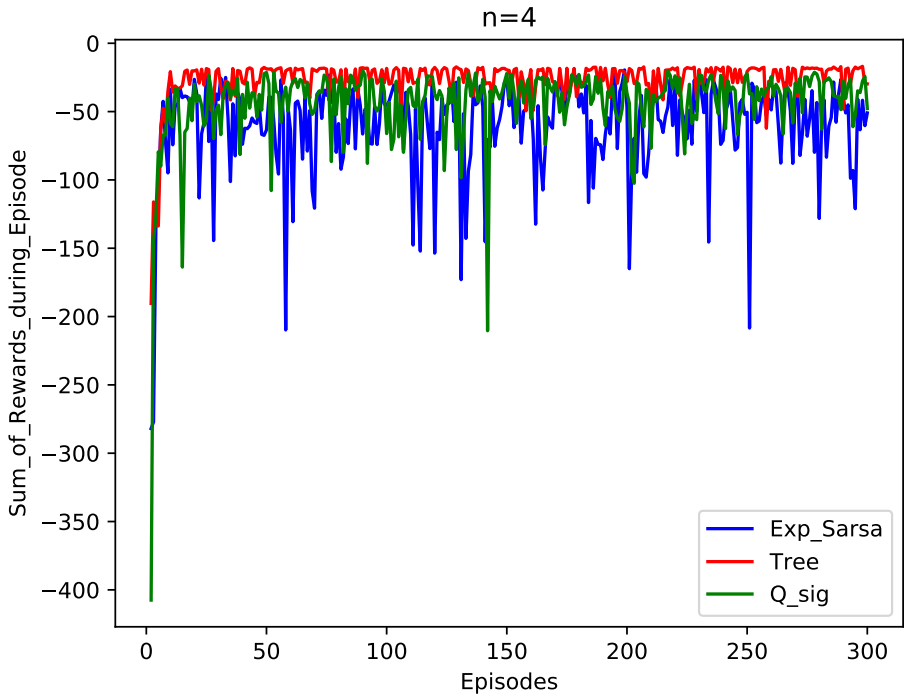
Sum of Rewards

Next we will look at how each method performs for different number of episodes. Note that each method was run 10 times and averaged for more accurate results ($\alpha = .6$, $\epsilon = .1$). Here Expected Sars is performing the best and Q(σ) the worst.



Sum of Rewards

Now changing our n parameter provides the following graph ($\alpha = .8$, $\epsilon = .1$). Here tree backup is performing the best and Expected Sarsa the worst.



Explanation

The previous two graphs coincide with the two α graphs. With $n = 4$ all Expected Sarsa and Q(σ) seem to be a lot more volatile. Therefore more episodes might be needed. This suggests that larger n may be a bad idea for these two methods.