Chenghao Wang
CS5180 Reinforcement Learning (Fall 2022)
September 12, 2022

Ex0

Question 1. How do you think this compares with your manual policy? (You do not have to run your manual policy for 10^4 steps!) What are some reasons for the difference in performance?

A random policy is used here, where each action has a 25 % probability of being taken. Compared to the manual policy (which reaches reward+1 in about 50 steps), the random policy takes about 1000 steps to achieve reward+1 Figure 1. The reasons behind this difference between performance could be found by calculating the probability that a goal can be reached under a random policy.

Start by labeling the following four rooms: lower left-1, lower right-2, upper left-3, upper right-4. Then by using random policy, the probability of agent getting out of room 1 is 2/27, the probability of agent being in room 2 or 3 and stepping into room 4 is $1/2*(1/26+1/27)$, and the probability of A reaching goal in room 4 is 1/30. After multiplying the above three terms together, we can get the probability of agent reaching goal which is 0.0108%, and this result can explain the difference between random policy and manual policy.

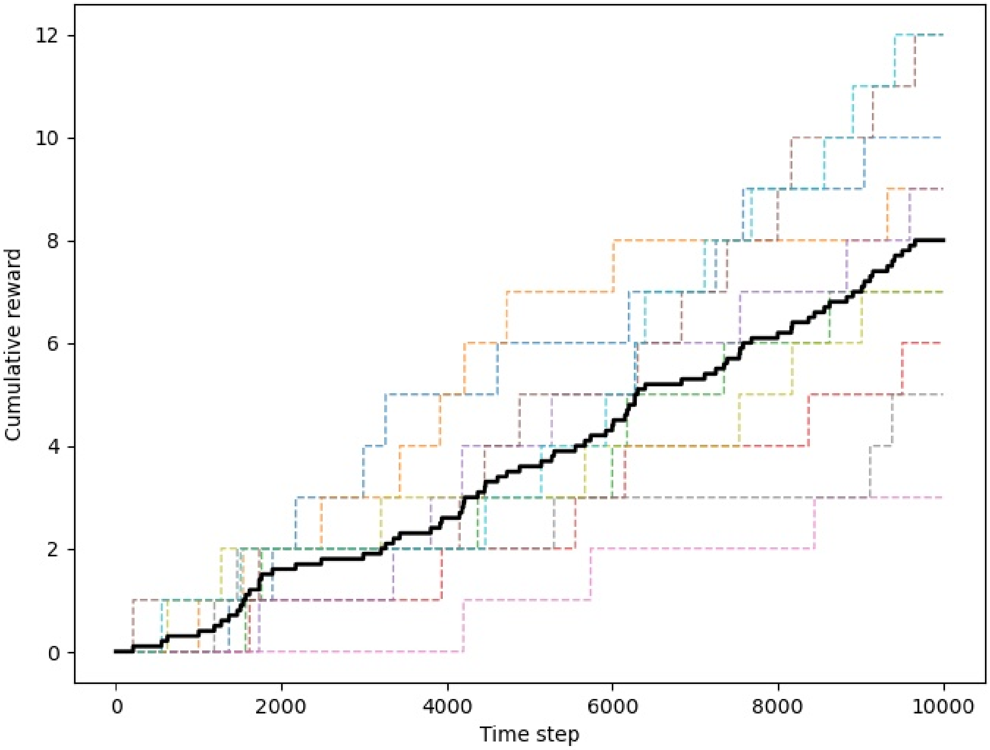

FIGURE 1. Cumulative reward by using random policy.

Question 2. Describe the strategy each policy uses, and why that leads to generally worse/better performance.

Two more policies were used for different performance, one significantly better than the above policy and one significantly worse Figure 2. In the first policy, the probability of moving up and to the right is increased from 25% to 40%, because the goal is in the upper right corner of the map, agent can reach goal in a fast speed under this policy. In the second policy, the probability of upward and rightward actions is reduced from 25% to 22%, and the other two probability are increased to 28, so the speed of reaching the goal is significantly reduced.

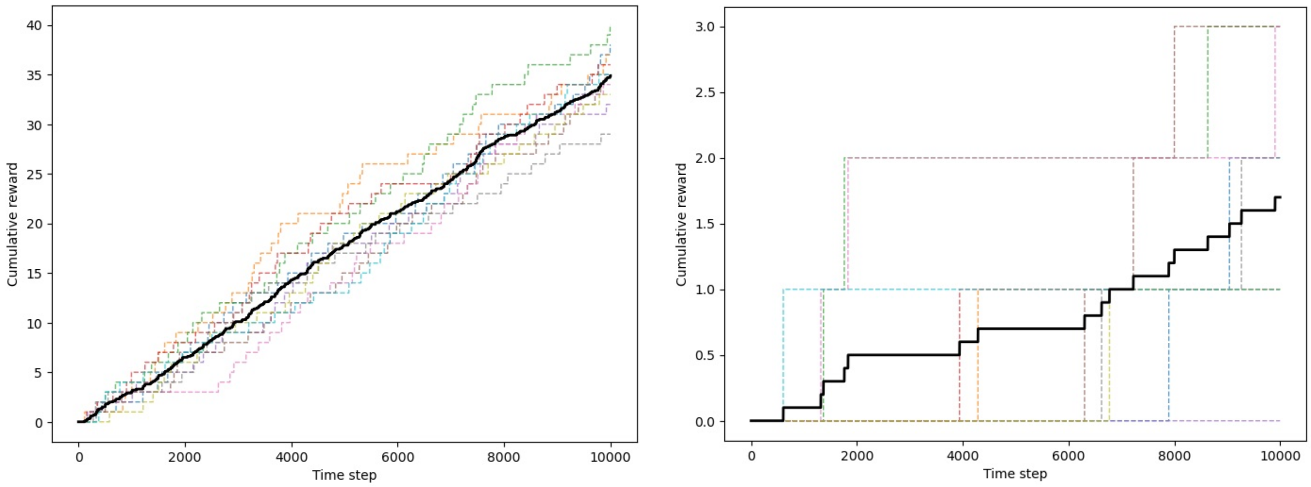

Figure 2. Probability policy (left - better, right - worse)