Practice Midterm

EC 400

October 25, 2021

1. Imagine throwing a fair coin n times to obtain a sequence of heads and tails. Let p_n be the probability that the sequence $HTHTHTHT$ appears in the result. Argue that $\lim_{n\to\infty} p_n = 1$.

2. Consider an MDP with a single state and two actions. The first action gives you a reward of 2, while the second gives you a reward of 3. Compute the Q-values associated both actions under the optimal policy. The discount factor γ should appear in your answer.

3. Consider an MDP with three states, labeled $1, 2, 3$. In state i, you have the option to "stay" or "move right." In each state, staying leaves you at that state with probability 1. At one or two, "move right" brings you to, respectively, two and three, with probability $1/2$; and with probability $1/2$ you stay where you are. At three, "move right" keeps you at 3 with probability 1. The rewards to "move right" are always zero, while the rewards to "stay" are one, two, and four respectively.

   Compute the optimal policy using dynamic programming with $K = 4$ steps and $\gamma = 1$.

4. Perform two steps of policy evaluation on the policy which takes actions uniformly at random at every state on the same MDP. Initialize all the values to be zero. Leave the discount factor γ in your final answer.

5. Perform two steps of Q-learning on the same MDP, assuming the state-action pairs generated are $s_1 = 1$,$a_1 =$"move right",$s_2 = 2$, $a_2=$"stay,", $s_3 = 2$. Initialize the Q-values at zero. Leave the discount factor γ in your final answer.

6. Suppose $\gamma = 1$. Give an example of an MDP where the value of a node is infinite.

7. True or false: if $J \neq T_\pi J$, then $J \neq J_\pi$.

8. True or false: the optimal policy in an MDP is unique.

9. Suppose T is an α-contraction in the infinity norm. Argue that

$$||T^{k+1}x - T^{k+1}y||_\infty \leq \alpha||T^k x - T^k y||_\infty$$