Worksheet C1M2

1. Suppose $\gamma = 0.9$ and the reward sequence is $R_1 = 2, R_2 = -2, R_3 = 0$ followed by an infinite sequence of 7s. What are G_1 and G_0?

Answer:

$G_3 = 7 + 0.9 \times 7 + 0.9^2 \times 7 + ... = 7 \times \frac{1}{1-0.9} = 70$
$G_2 = R_3 + 0.9 \times G_3 = 0 + 0.9 \times 70 = 63$
$G_1 = R_2 + 0.9 \times G_2 = -2 + 0.9 \times 63 = 54.7$
$G_0 = R_1 + 0.9 \times G_1 = 2 + 0.9 \times 54.7 = 51.23$

2. (Exercise 2.2 from S&B 2nd edition) Consider a k-armed bandit problem with $k = 4$ actions, denoted 1, 2, 3, and 4. Consider applying to this problem a bandit algorithm using ϵ-greedy action selection, sample-average action-value estimates, and initial estimates of $Q_1(a) = 0$, for all a. Suppose the initial sequence of actions and rewards is $A_1 = 1, R_1 = -1, A_2 = 2, R_2 = 1, A_3 = 2, R_3 = -2, A_4 = 2, R_4 = 2, A_5 = 3, R_5 = 0$. On some of these time steps the ϵ case may have occurred, causing an action to be selected at random. On which time steps did this definitely occur? On which time steps could this possibly have occurred?

3. Assume you have a bandit problem with 4 actions, where the agent can see rewards from the set $\mathcal{R} = \{-3.0, -0.1, 0, 4.2\}$. Assume you have the probabilities for rewards for each action: $p(r|a)$ for $a \in \{1, 2, 3, 4\}$ and $r \in \{-3.0, -0.1, 0, 4.2\}$. How can you write this problem as an MDP? Remember that an MDP consists of $(\mathcal{S}, \mathcal{A}, \mathcal{R}, P, \gamma)$.

More abstractly, recall that a Bandit problem consists of a given action space $\mathcal{A} =$

$\{1, ..., k\}$ (the k arms) and the distribution over rewards $p(r|a)$ for each action $a \in \mathcal{A}$. Specify an MDP that corresponds to this Bandit problem.

Answer:

Action space: $\mathcal{A}$
set of possible rewards: $\mathcal{R}$
$\mathcal{S} = \{1\}$
$p(s', r|s, a) = p(r|a)$ where s', s always equal 1
$\gamma = 0$

4. Prove that the discounted sum of rewards is always finite, if the rewards are bounded: $|R_{t+1}| \leq R_{\max}$ for all t for some finite $R_{\max} > 0$.

$$\left| \sum_{i=0}^{\infty} \gamma^i R_{t+1+i} \right| < \infty \qquad \text{for } \gamma \in [0, 1)$$

Hint: Recall that $|a + b| < |a| + |b|$.

Answer:

$$\begin{aligned}
\left| \sum_{i=0}^{\infty} \gamma^i R_{t+1+i} \right| &\leq \sum_{i=0}^{\infty} \left| \gamma^i R_{t+1+i} \right| \\
&= \sum_{i=0}^{\infty} \gamma^i \left| R_{t+1+i} \right| \\
&\leq \sum_{i=0}^{\infty} \gamma^i R_{\max} \\
&= R_{\max} \sum_{i=0}^{\infty} \gamma^i \\
&= R_{\max} \frac{1}{1 - \gamma}
\end{aligned}$$

$R_{\max}$ and $\frac{1}{1-\gamma}$ are finite.