CME 241 Assignment-3

Halil Ibrahim Gulluk ID: 06454540

February 14, 2022

QUESTION-1: For a deterministic policy π_D we have the following Bellman Equations

$$V^{\pi_D}(s) = \mathcal{R}(s, \pi_D(s)) + \gamma \cdot \sum_{s' \in \mathcal{N}} \mathcal{P}(s, \pi_D(s), s') V^{\pi_D}(s') \tag{1}$$

$$V^{\pi_D}(s) = Q^{\pi_D}(s, \pi_D(s)) \tag{2}$$

$$Q^{\pi_D}(s, a) = 0, a \neq \pi_D(s) \tag{3}$$

$$Q^{\pi_D}(s, \pi_D(s)) = \mathcal{R}(s, \pi_D(s)) + \gamma \sum_{s' \in \mathcal{N}} \mathcal{P}(s, \pi_D(s), s') V^{\pi_D}(s') \tag{4}$$

$$Q^{\pi_D}(s, \pi_D(s)) = \mathcal{R}(s, \pi_D(s)) + \gamma \sum_{s' \in \mathcal{N}} \mathcal{P}(s, \pi_D(s), s') Q^{\pi_D}(s', \pi_D(s')) \tag{5}$$

QUESTION 2: Bellman optimality equation : $V^*(s) = \max_{a \in \mathcal{A}} \{\mathcal{R}(s, a) + \gamma \sum_{s' \in \mathcal{N}} \mathcal{P}(s, a, s') V^*(s')\}$

$$\mathcal{R}(s, a) = \mathbf{Prob}[s+1|s, a]\mathcal{R}(s, s+1) + \mathbf{Prob}[s|s, a]\mathcal{R}(s, s) = a(1-a) + (1-a)(1+a) \implies \tag{6}$$

$$V^*(s) = \max_a 1 + a - 2a^2 + \frac{1}{2}[aV^*(s+1) + (1-a)V^*(s)] \tag{7}$$

Observe that $V^*(s) = V^*(s+1)$ because actions space and rewards of the states, and the game continues to infinity. So, it does not matter where we start actually. Then, $V^*(s) = \max_a 1 + a - 2a^2 + \frac{1}{2}V^*(s)$, this is quadratic wrt a, with a negative leading coefficient, so it takes maximum value at $a = \frac{1}{4} \longrightarrow V^*(s) = 1 + \frac{1}{4} - \frac{1}{8} + \frac{1}{2}V^*(s)$, then $V^*(s) = \frac{9}{4}$, then optimal policy $\pi_D(s, a) = 1$ if $a = \frac{1}{4}$, otherwise $\pi_D(s, a) = 0$.

QUESTION 4: $V^*(s) = \max_a \mathcal{R}(s, a)$ as $\gamma = 0$. We know that

$$\mathcal{R}(s, a) = \int_{s'} f(s, a, s')\mathcal{R}(s, a, s')ds' = \int_{-\infty}^{\infty} e^{\frac{-(s'-s)^2}{2\sigma^2}} \cdot (-e^{as'})ds' \tag{8}$$

Then we will solve the following problem

$$\min_a \int_{-\infty}^{\infty} e^{\frac{-(s'-s)^2}{2\sigma^2}} e^{as'} ds' = \min_a \int_{-\infty}^{\infty} e^{\frac{-1}{2\sigma^2}((s')^2 - 2s's - 2s'\sigma^2 a + s^2)} ds' \tag{9}$$

$$= \min_a \int_{-\infty}^{\infty} e^{\frac{-1}{2\sigma^2}((s'-(s+\sigma^2 a))^2 - 2s\sigma^2 a - \sigma^4 a^2)} ds' = \min_a \int_{-\infty}^{\infty} e^{\frac{-1}{2\sigma^2}(s'-(s+\sigma^2 a))^2} e^{\frac{1}{2\sigma^2}(2s\sigma^2 a + \sigma^4 a^2)} ds' \tag{10}$$

$$= \min_a e^{\frac{1}{2\sigma^2}(2s\sigma^2 a + \sigma^4 a^2)} \int_{-\infty}^{\infty} e^{\frac{-1}{2\sigma^2}(s'-(s+\sigma^2 a))^2} ds' = \min_a e^{\frac{1}{2\sigma^2}(2s\sigma^2 a + \sigma^4 a^2)} \tag{11}$$

As the inside of the integral is the integral of a pdf of a gaussian distribution, which is equal to 1.

In order to minimize $e^{\frac{1}{2\sigma^2}(2s\sigma^2 a + \sigma^4 a^2)}$ we need to minimize the exponential term $\implies$ $\min_a 2sa + \sigma^2 a^2 \implies a^* = \frac{-s}{\sigma^2}$ and the corresponding cost is $e^{\frac{-s^2}{2\sigma^2}}$

QUESTION-3: We develop the MDP as follows : $s_t = k, k = 0, 1, 2, .., n$ stands for the place that the player lies on. $a_t =' a'$ or $a_t =' b'$, these are the only actions we have.

$\mathcal{P}(i,' a', i+1) = \frac{n-i}{n}$, $\mathcal{P}(i,' a', i-1) = \frac{i}{n}$ $\mathcal{P}(i,' b', j) = 1/n$ for $j = 0, 1, .., i-1, i+1, .., n$

As the reward function I choose (one can make different choices as well)

$\mathcal{R}(i,j) = j - i$ for $j = 1, 2, .., n$ and $\mathcal{R}(i, 0) = -n$, these two equations hold for $i = 0, 1, .., n$.

I implemented the code, and it prints the optimal value function and the optimal policy.