CME241 A3

Yiyun Gu

1 Problem 1

$V_D^\pi(s) = R(s, \pi_D(s)) + \gamma V_D^\pi(s')$ where $P(s, \pi_D(s), s') = 1$
$Q_D^\pi(s, \pi_D(s)) = R(s, \pi_D(s)) + \gamma V_D^\pi(s')$ where $P(s, \pi_D(s), s') = 1$
$Q_D^\pi(s, \pi_D(s)) = R(s, \pi_D(s)) + \gamma Q_D^\pi(s', \pi_D(s'))$ where $P(s, \pi_D(s), s') = 1$
$V_D^\pi(s) = Q_D^\pi(s, \pi_D(s))$

2 Problem 2

$R(s,a) = \mathbf{E}[R_{t+1}|R_t = s, A_t = a] = \sum_{s' \in S} \sum_{r \in D} P_R(s,a,r,s')r$
$= P(s+1|s,a) * (1-a) + P(s|s,a) * (1+a) = (1-a)(1+2a)$
$V^*(s) = \max_{a \in [0,1]}[R(s,a) + \gamma \sum_{s' \in N} P(s,a,s')V^*(s')]$
$= \max_{a \in [0,1]}[R(s,a) + 0.5(P(s,a,s+1)V^*(s+1) + P(s,a,s)V^*(s)]$
$= \max_{a \in [0,1]}[(1-a)(1+2a) + 0.5aV^*(s+1) + 0.5(1-a)V^*(s)]$
$= \max_{a \in [0,1]}[(1-a)(1+2a) + 0.5V^*(s)]$, since the reward and transition are independent of s
This is equivalent to maximizing (1-a)(1+2a).
$a^* = \pi^*(s) = \frac{1}{4}$
$V^*(s) = 2(1-a^*)(1+2a^*) = \frac{9}{4}$

3 Problem 3

1. State space: 0, 1, ..., n
   Action space: A, B
   Transitions function:
   If $i = 1, ..., n-2$: $P_R(i, A, 0, i+1) = \frac{n-i}{n}$, $P_R(i, A, 0, i-1) = \frac{i}{n}$
   If $i = n-1$:$P_R(i, A, \frac{n-i}{n}, i+1) = \frac{n-i}{n}$, $P_R(i, A, 0, i-1) = \frac{i}{n}$
   $P_R(i, B, 0, j) = \frac{1}{n}, j = 0, 1, ..., n-1, j! = i$
   $P_R(i, B, \frac{1}{n}, n) = \frac{1}{n}$
   Reward functions:
   If $i = 1, ..., n-2$: $R(i, A) = 0$
   If $i = n-1$: $R(i, A) = \frac{1}{n^2}$
   $R(i, B) = \frac{1}{n^2}$
2. The pattern is starting with croak B and then keeping croak A.

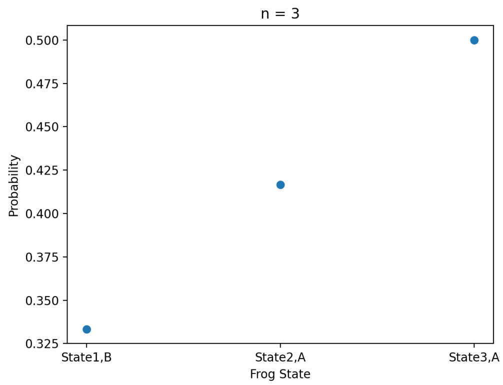

Figure 1: Optimal Escape-Probability for n = 3

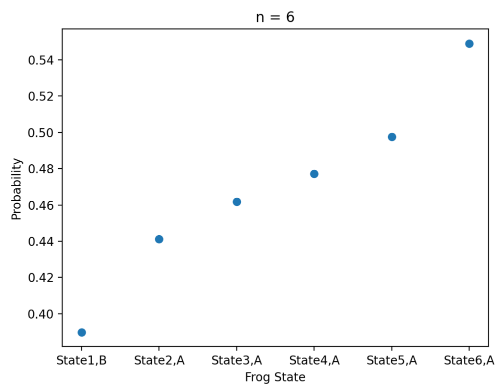

Figure 2: Optimal Escape-Probability for n = 6

4 Problem 4

$V^*(s) = \max_{a \in A} Q^*(s,a)$
$Q^*(S,a) = R(s,a) + \gamma \sum_{s' \in N} P(s,a,s') V^*(s')$
Since $\gamma = 0$, we want to maximize $R(s,a)$.
$R(s,a) = \mathbf{E}[R_{t+1} | R_t = s, A_t = a] = \sum_{s' \in S} \sum_{r \in D} P_R(s,a,r,s') r$
$= \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{-(s'-s)^2}{2\sigma^2}} * e^{as'} \, ds'$
Want $\frac{\partial}{\partial a} \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{-(s'-s)^2}{2\sigma^2}} e^{as'} \, ds' = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{-(s'-s)^2}{2\sigma^2}} e^{as'} s' \, ds' = 0$
$\int_{-\infty}^{\infty} e^{-\frac{-(s'-s)^2}{2\sigma^2}} e^{as'} s' \, ds' = \int_{-\infty}^{\infty} e^{-\frac{(s'+\sigma^2 a - s)^2 - \sigma^4 a^2 + 2\sigma^2 a}{2\sigma^2}} s' \, ds'$
$= \sigma^2 e^{\frac{\sigma^4 a^2 - 2\sigma^2 as}{2\sigma^2}} [\int_{\infty}^{\infty} e^{-\frac{(s'+\sigma^2 a - s)^2}{2\sigma^2}} \, d\frac{(s'+\sigma^2 a - s)^2}{2\sigma^2} - \sqrt{2}\sigma \int_{\infty}^{\infty} e^{-\frac{(s'+\sigma^2 a - s)^2}{2\sigma^2}} (\sigma^2 a - s) \, d\frac{s'+\sigma^2 a - s}{\sqrt{2}\sigma}]$
$= 0 - \sqrt{2}\sigma^3 e^{\frac{\sigma^2 a^2 - 2as}{2}} (\sigma^2 a - s)\sqrt{\pi} = 0$

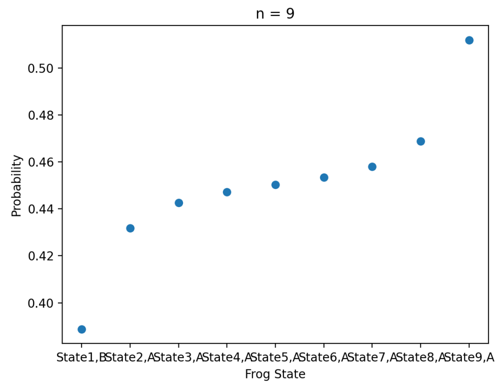

Figure 3: Optimal Escape-Probability for n = 9

$\longrightarrow a^* = \frac{s}{\sigma^2}$