Assignment 3

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

$Q^{\pi_D}(s,a) = R(s,a) + \sum_{s'} P(s,a,s') \cdot V^{\pi_D}(s')$

$V^{\pi_D}(s) = R(s,\pi_D(s)) + \sum_{s'} P(s,\pi_D(s),s') \cdot V^{\pi_D}(s')$

$Q^{\pi_D}(s,a) = R(s,a) + \sum_{s'} P(s,a,s') \cdot Q^{\pi_D}(s', \pi_D(s'))$

(2) $V^*(s) = \max_{a \in A} \{ R(s,a) + \gamma \cdot \sum_{s' \in N} P(s,a,s') \cdot V^*(s') \}$

$$R(s,a) = P(s,a,s) \cdot (1+a) + P(s,a,s+1) \cdot (1-a)$$
$$= (1-a)(1+a) + a(1-a)$$
$$= (1+2a)(1-a)$$

$$V^*(s) = \max_{a \in A} \{ (1+2a)(1-a) + \tfrac{1}{2}(a V^*(s+1) + (1-a) V^*(s)) \}$$

Suppose V^* were independent of s. Then:

$$V^*(s) = \max_{a \in A} \{ (1+2a)(1-a) + \tfrac{1}{2} V^*(s) \} = \tfrac{1}{2} V^*(s) + \max_{a \in A} \{(1+2a)(1-a)\}$$

$\Rightarrow V^*(s) = 2 \cdot 9/8 = 9/4$ for all s

An optimal deterministic policy $\pi_D(s)$ is:

$\pi_D(s) = 1/4$ for all $s \in S$

Assignment 3 (cont.)

(3) $N = \{1, \ldots, n-1\}$ $T = \{0, n\}$ $S = N \cup T$

$A = \{A, B\}$

$$\mathbb{P}[s' | s, a] = \begin{cases} \frac{s}{n} & \text{if } a = \text{"A" and } s' = s-1 \\ \frac{n-s}{n} & \text{if } a = \text{"A" and } s' = s+1 \\ \frac{1}{n} & \text{if } a = \text{"B" and } s' \neq s \\ 0 & \text{otherwise} \end{cases}$$

Reward of transition (s, a, s') is $+1$ if $s' = n$, and is 0 otherwise.

$\gamma = 1$

The optimal policy is to always choose "A" unless you are on the first lily pad.

(4)

$c = e^{as'}$, where $s' \sim N(s, \sigma^2)$

We want to choose a to minimize $\mathbb{E}[c]$ for any s:

$\mathbb{E}[c] = \mathbb{E}[e^{as'}] = M(a)$ where M is the moment generating function of the distribution of s'. In particular, if $s' \sim N(s, \sigma^2)$, then:

$$\mathbb{E}[c] = M_{normal}(a) = \exp\left(sa + \frac{\sigma^2 a^2}{2}\right)$$

We can minimize by calculating: $\frac{d}{da}\mathbb{E}[c] = (s + \sigma^2 a) M(a) = 0 \Rightarrow \boxed{a = -\frac{s}{\sigma^2}}$