HW 4

Problem 1)

1) $G_i^s = \sum_{t=0}^{|\tau_i^s|} \gamma^t R(S_{t,i}^s, A_{t,i}^s) \leq \sum_{t=0}^{|\tau_i^s|} \gamma^t R_{max} = \frac{1-\gamma^\tau}{1-\gamma} R_{max}$

Let $\alpha = -\frac{1-\gamma^\tau}{1-\gamma} R_{max}$ and $\beta = \frac{1-\gamma^\tau}{1-\gamma} R_{max}$

Such that $\alpha \leq G_i \leq \beta$

2) $\mathbb{E}[E(s)] = \mathbb{E}\left[\sum_{i=1}^{N^s} G_i^s\right] = \sum_{i=1}^{N^s} \mathbb{E}[G_i^s] = \sum_{i=1}^{N^s} V^\pi(s) = N^s V^\pi(s)$

3) $\mathbb{P}(|E(s) - \mathbb{E}[E(s)]| \geq \varepsilon) \leq 2\exp\left(-\frac{2\varepsilon^2}{N^s(\beta-\alpha)^2}\right)$

where $\alpha = -\frac{1-\gamma^\tau}{1-\gamma} R_{max}$ $\beta = \frac{1-\gamma^\tau}{1-\gamma} R_{max}$

4) $V^\pi(s) = \frac{1}{N^s} E(s)$ $\varepsilon = N^s \varepsilon'$

$\mathbb{P}(|V^\pi(s) - \mathbb{E}[V^\pi(s)]|) \geq \varepsilon') \leq 2\exp\left(-\frac{2N^s\varepsilon'}{(\beta-\alpha)^2}\right)$

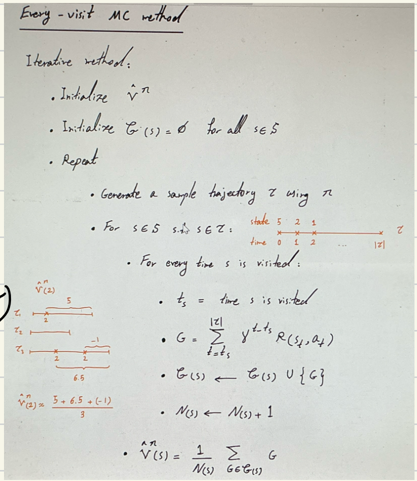
5) $\|\hat{V}^\pi(s) - V^\pi(s)\|_\infty \geq \varepsilon' \iff \bigcup_{s\in S}(|\hat{V}^\pi(s) - V^\pi(s)| \geq \varepsilon')$

$\therefore \mathbb{P}(\|\hat{V}^\pi(s) - V^\pi(s)\|_\infty \geq \varepsilon') \leq \sum_{s\in S} 2\exp\left(-\frac{2N^s\varepsilon'^2}{(\beta-\alpha)^2}\right)$

$2\exp\left(-\frac{2N^s\varepsilon'^2}{(\beta-\alpha)^2}\right) \leq 2\exp\left(-\frac{2N\varepsilon'^2}{(\beta-\alpha)^2}\right)$ where $N = \min_{s\in S} N^s$

$\therefore \mathbb{P}(\|\hat{V}^\pi(s) - V^\pi(s)\|_\infty \geq \varepsilon') \leq \sum_{s\in S} 2\exp\left(-\frac{2N\varepsilon'^2}{(\beta-\alpha)^2}\right)$

$$Q^\pi(s,x) = \frac{1}{N(s)} \sum_{G \in \mathcal{G}(s)} G$$



Problem 2

τ_1: $G_1^1 = \gamma^0(-1) + \gamma^1(2)$ $\quad N=1$

$Q = -1 + \gamma 2$

τ_2: $G_2^1 = \gamma^0(-1) + \gamma(2) + \gamma^2(0) + \gamma^3(-1) + \gamma^4(1.5)$
$+ \gamma^5(1.5)$ $\quad N=2$

$$Q = \frac{G_1^1 + G_2^1}{N} = \frac{-1+\gamma^2 - 1 + \gamma 2 - \gamma^3 + \gamma^4(1.5) + \gamma^5(1.5)}{2}$$

$G_2^2 = -1 + \gamma(1.5) + \gamma^2(1.5)$ $\quad N=3$

$$Q = \frac{G_1^1 + G_2^1 + G_2^2}{N} = \frac{-1+\gamma^2 - 1 + \gamma 2 - \gamma^3 + \gamma^4(1.5) + \gamma^5(1.5) - 1 + \gamma(1.5) + \gamma^2(1.5)}{3}$$

τ_3: $G_3^1 = 0$ $\quad N=3$ $\quad Q =$

Problem 3

1) Each state except terminal states are run multiple times and an average is aggregated to get a more accurate estimate
2) States are randomly selected until all non-terminal states have been visited >= 50 times