



HW 4

problem 1)

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

let $\alpha = -\frac{1-\gamma^\tau}{1-\gamma} R_{\max}$ and $\beta = \frac{1-\gamma^\tau}{1-\gamma} R_{\max}$
such that $\alpha \leq G_i^s \leq \beta$

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

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

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

$$4) V^s(s) = \frac{1}{N^s} E[G_i^s] \quad \epsilon' = N^s \epsilon$$

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

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

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

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

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

$$Q^n(S, x) = \frac{1}{N(t)} \sum_{G \in G(t)} G$$

Problem 2

$$\begin{aligned} T_1: \quad G_1' &= \gamma^0(-1) + \gamma^1(2) \quad N=1 \\ Q_1 &= -1 + \gamma 2 \end{aligned}$$

$$\begin{aligned} T_2: \quad G_2' &= \gamma^0(-1) + \gamma^1(2) + \gamma^2(0) + \gamma^3(-1) + \gamma^4(1.5) \\ &\quad + \gamma^5(6.5) \quad N=2 \end{aligned}$$

$$Q_2 = \frac{G_1' + G_2'}{N} = \frac{-1 + \gamma^2 - 1 + \gamma 2 - \gamma^3 + \gamma^4(1.5) + \gamma^5(6.5)}{2}$$

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

$$Q_3 = \frac{G_1' + G_2' + G_2''}{N} = \underbrace{-1 + \gamma^2 - 1 + \gamma 2 - \gamma^3 + \gamma^4(1.5) + \gamma^5(6.5)}_{3} - 1 + \gamma(1.5) + \gamma^2(1.5)$$

$$T_3: \quad G_3' = 0 \quad N=3 \quad G_2 = \overbrace{\hspace{10em}}$$

Every-visit MC method

Iterative method:

- Initialize \hat{V}^n

- Initialize $G(s) = \emptyset$ for all $s \in S$

- Repeat

- Generate a sample trajectory τ using n

- For $s \in S$, $s \in \tau$:


- For every time s is visited:

 - t_s = time s is visited

 - $G = \sum_{t=t_s}^{|T|} \gamma^{t-t_s} R(s_t, a_t)$

 - $G(s) \leftarrow G(s) \cup \{G\}$

 - $N(s) \leftarrow N(s) + 1$

 - $\hat{V}^n(s) = \frac{1}{N(s)} \sum_{G \in G(s)} G$

Problem 3

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