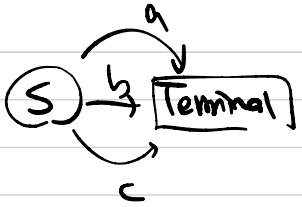
Problem(1) a. Find the mean Vector of $E[\hat{\nabla}v]$



$r_a=100 \quad \theta_a=0$
$r_b=98 \quad \theta_b=\ln 5$
$r_c=95 \quad \theta_c=\ln 4$
$\Rightarrow \begin{cases} \pi(a|s)=0.1 \\ \pi(b|s)=0.5 \\ \pi(c|s)=0.4 \end{cases}$

$\because \pi_\theta(a|s)=\frac{\exp(0)}{\exp(0)+\exp(\ln 5)+\exp(\ln 4)} \Rightarrow \log \pi_\theta(a|s)=\log e^0-\log(e^0+e^{\ln 5}+e^{\ln 4})$

可得 $\begin{cases} \frac{\partial}{\partial \theta_a}\left[\log \pi_\theta(a|s)\right]=1-\pi_\theta(a|s)=1-0.1=0.9 \\ \frac{\partial}{\partial \theta_b}\left[\log \pi_\theta(a|s)\right]=-\pi(b|s)=-0.5 \\ \frac{\partial}{\partial \theta_c}\left[\log \pi_\theta(a|s)\right]=-\pi(c|s)=-0.4 \end{cases} \Rightarrow \nabla_\theta \log \pi_\theta(a|s)=\begin{bmatrix}0.9\\-0.5\\-0.4\end{bmatrix}$ 同理可得

$\nabla_\theta \log \pi_\theta(b|s)=\begin{bmatrix}-0.1\\0.5\\-0.4\end{bmatrix}$

$\nabla_\theta \log \pi_\theta(c|s)=\begin{bmatrix}-0.1\\-0.5\\0.6\end{bmatrix}$

可得 $E[\alpha]=\underset{\tau\sim p_M^{\pi_\theta}}{E}\left[\sum_{t=0}^{T-1}\gamma^t Q^{\pi_\theta}(s_t,a_t)\nabla_\theta \log \pi_\theta(a_t|s_t)\right]$

$\because Q^{\pi_\theta}(s_t,a_t)=r(s,a)+\gamma\sum_{s'} P_{ss'}^{a}\, V(s'=\text{terminal})=r(s,a)$

故 $E[\hat{\nabla}v]=\pi_\theta(a|s)\left[r(s,a)\times\nabla_\theta \log \pi_\theta(a|s)\right]+\pi_\theta(b|s)\left[r(s,b)\times\nabla_\theta \log \pi_\theta(b|s)\right]+\pi_\theta(c|s)\left[r(s,c)\times\nabla_\theta \log \pi_\theta(c|s)\right]$

$=0.1\times\left[100\times\begin{bmatrix}0.9\\-0.5\\-0.4\end{bmatrix}\right]+0.5\times\left[98\times\begin{bmatrix}-0.1\\0.5\\-0.4\end{bmatrix}\right]+0.4\times\left[95\times\begin{bmatrix}-0.1\\-0.5\\0.6\end{bmatrix}\right]$

$=\begin{bmatrix}9\\-5\\-6\end{bmatrix}+\begin{bmatrix}-4.9\\24.5\\-19.6\end{bmatrix}+\begin{bmatrix}-3.8\\-19\\22.8\end{bmatrix}=\begin{bmatrix}0.3\\0.5\\-0.8\end{bmatrix}$ //

Problem(1)-a Find the Covariance Matrix of $\hat{\nabla}v$

Since $E[\hat{\nabla}v]=\begin{bmatrix}0.3\\0.5\\-0.8\end{bmatrix}$, and $\left(\hat{\nabla}v_a,\hat{\nabla}v_b,\hat{\nabla}v_c\right)=\left(100\times\begin{bmatrix}0.9\\-0.5\\-0.4\end{bmatrix},98\times\begin{bmatrix}-0.1\\0.5\\-0.4\end{bmatrix},95\times\begin{bmatrix}-0.1\\-0.5\\0.6\end{bmatrix}\right)$

可得 $\left(\hat{\nabla}v_a-E[\hat{\nabla}v],\hat{\nabla}v_b-E[\hat{\nabla}v],\hat{\nabla}v_c-E[\hat{\nabla}v]\right)=\left(\begin{bmatrix}89.7\\-50.5\\-39.2\end{bmatrix},\begin{bmatrix}-10.1\\48.5\\-38.4\end{bmatrix},\begin{bmatrix}-9.8\\-48\\57.8\end{bmatrix}\right)$

$\therefore E\left[\left(\hat{\nabla}v-E[\hat{\nabla}v]\right)\left(\hat{\nabla}v-E[\hat{\nabla}v]\right)^T\right]$

$=0.1\times\begin{bmatrix}89.7\\-50.5\\-39.2\end{bmatrix}\begin{bmatrix}89.7,-50.5,39.2\end{bmatrix}+0.5\times\begin{bmatrix}-10.1\\48.5\\-38.4\end{bmatrix}\begin{bmatrix}-10.1,48.5,-38.4\end{bmatrix}+0.4\times\begin{bmatrix}-9.8\\-48\\57.8\end{bmatrix}\begin{bmatrix}-9.8,-48,57.8\end{bmatrix}$

$=\begin{bmatrix}894.03,-599.75,-384.28\\-599.75,2252.75,-1843\\-384.28,-1843,2227.28\end{bmatrix}$ //

Problem (2)

$$RHS = \frac{1}{1-\gamma} E_{s\sim d_M^{\pi_\theta}} E_{a\sim\pi_\theta(\cdot|s)}[f(s,a)]$$

$$= \frac{1}{1-\gamma} \sum_s d_M^{\pi_\theta}(s) E_{a\sim\pi_\theta(\cdot|s)}[f(s,a)]$$

$$= \frac{1}{1-\gamma} \sum_s \sum_a \pi(a|s) \times f(s,a) \times \boxed{d_M^{\pi_\theta}(s)} \quad \boxed{d_M^{\pi_\theta}(s)} \to E_{s_0\sim M}[d_M^{\pi_\theta}(s)] = E_{s_0\sim M}\left[(1-\gamma)\sum_{t=0}^{\infty}\gamma^t P(s_t=s|s_0,\pi_\theta)\right]$$

$$= \frac{1}{1-\gamma} E_{s_0\sim M}\left[(1-\gamma)\sum_s \sum_a \pi_\theta(a|s) \times f(s,a) \sum_{t=0}^{\infty}\gamma^t P(s_t=s|s_0,\pi_\theta)\right]$$

$$= E_{s_0\sim M}\left[\sum_s \sum_a \pi_\theta(a|s) \sum_{t=0}^{\infty} P(s_t=s|s_0,\pi_\theta)\cdot\gamma^t \times f(s_t,a_t)\right]$$

$$= \sum_\tau \sum_{t=0}^{\infty} P(\tau|M,\pi_\theta)\,\gamma^t \cdot f(s_t,a_t)$$

$$= E_{\tau\sim p_M^{\pi_\theta}}\left[\sum_{t=0}^{\infty}\gamma^t f(s_t,a_t)\right] \quad \#$$

Problem (3)

Since only have one state S and Terminal T. So we only visit once

Property 1: $\therefore V(S) = P_S(R_S + V(S)) + P_T(R_T + V_T)$

$$\therefore (1-P_S)V(S) = P_S R_S + P_T R_T$$

$$\therefore V(S) = \frac{P_S}{P_T} R_S + R_T$$

Property 2: $E\left[\hat{V}_{MC}(S;\tau)\right] = \sum_k P(k) \times E_\tau[x(x)|k]$

$$= \sum_k P_T P_S^k \left(\frac{R_S + 2R_S + 3R_S + \dots + kR_S + (k+1)R_T}{k+1}\right)$$

$$= \sum_k P_T P_S^k \left(\frac{k}{2} R_S + R_T\right)$$

$$= \frac{P_S}{2P_T} R_S + R_T$$