Problem 1.

(a)

$$\nabla_\theta \log \pi_\theta(a|s) = 1 - \frac{1}{10} = \frac{9}{10}$$

$$\nabla_\theta \log \pi_\theta(b|s) = 1 - \frac{5}{10} = \frac{5}{10}$$

$$\nabla_\theta \log \pi_\theta(c|s) = 1 - \frac{4}{10} = \frac{6}{10}$$

$$\hat{\nabla} V = \left[100 \times \begin{bmatrix} \frac{9}{10} \\ -\frac{5}{10} \\ -\frac{4}{10} \end{bmatrix}, \; 98 \times \begin{bmatrix} -\frac{1}{10} \\ \frac{5}{10} \\ -\frac{4}{10} \end{bmatrix}, \; 95 \times \begin{bmatrix} -\frac{1}{10} \\ -\frac{5}{10} \\ \frac{6}{10} \end{bmatrix} \right]$$

$$E[\hat{\nabla} V] = \frac{1}{10} \times 100 \times \begin{bmatrix} \frac{9}{10} \\ -\frac{5}{10} \\ -\frac{4}{10} \end{bmatrix} + \frac{5}{10} \times 98 \times \begin{bmatrix} -\frac{1}{10} \\ \frac{5}{10} \\ -\frac{4}{10} \end{bmatrix} + \frac{4}{10} \times 95 \times \begin{bmatrix} -\frac{1}{10} \\ -\frac{5}{10} \\ \frac{6}{10} \end{bmatrix}$$

$$= \begin{bmatrix} 9 - 4.9 - 3.8 \\ -5 + 24.5 - 19 \\ -4 - 19.6 + 22.8 \end{bmatrix} = \begin{bmatrix} 0.3 \\ 0.5 \\ -0.8 \end{bmatrix}$$

$$\text{Variance} = E\left[[\hat{\nabla} V - E[\hat{\nabla} V]] [\hat{\nabla} V - E[\hat{\nabla} V]]^T \right]$$

$$= \frac{1}{10} \times (8100 + \overset{4100}{2500 + 1600}) + \frac{5}{10} \times 4033.68 + \frac{4}{10} \times 5595.5$$

$$- (0.09 + 0.25 + 0.64)$$

$$= 1220 + 2016.84 + 2238.2 - 0.98 = 5474.06 \; \#$$

(b)

$$V^{\pi_\theta}(s) = \frac{1}{10}\times 100 + \frac{5}{10}\times 98 + \frac{4}{10}\times 95 = 97$$

$$\hat{\nabla} V = \left[3\times\begin{bmatrix}0.9\\-0.5\\-0.4\end{bmatrix},\ 1\times\begin{bmatrix}-0.1\\0.5\\-0.4\end{bmatrix},\ -2\times\begin{bmatrix}-0.1\\-0.5\\0.6\end{bmatrix} \right]$$

$$E[\hat{\nabla} V] = \frac{1}{10}\times 3\times\begin{bmatrix}0.9\\-0.5\\-0.4\end{bmatrix} + \frac{5}{10}\times\begin{bmatrix}-0.1\\0.5\\-0.4\end{bmatrix} + \frac{4}{10}\times(-2)\times\begin{bmatrix}-0.1\\-0.5\\0.6\end{bmatrix}$$

$$= \begin{bmatrix}0.27 - 0.05 + 0.08\\ -0.15 + 0.25 + 0.4\\ -0.12 - 0.2 - 0.48\end{bmatrix} = \begin{bmatrix}0.3\\0.5\\-0.8\end{bmatrix} \#$$

Variance :

$$\frac{1}{10}\times(0.27^2 + 0.15^2 + 0.12^2) + \frac{5}{10}\times(0.01 + 0.25 + 0.16)$$

$$+ \frac{4}{10}\times(0.04 + 1 + 1.44) - (0.09 + 0.25 + 0.64)$$

$$= 0.01098 + 0.21 + 0.992 - 0.98 = 0.23298 \#$$

(c)

Variance:

$$\frac{1}{10}(100-x)^2\overset{1.22}{(0.81+0.25+0.16)} + (98-x)^2\overset{0.42}{(0.01+0.25+0.16)}\frac{5}{10}$$

$$+ (95-x)^2\overset{0.62}{(0.01+0.25+0.36)}\frac{4}{10}$$

$$= 0.122(100-x)^2 + 0.21(98-x)^2 + 0.248(95-x)^2$$

$$= (0.122+0.21+0.248)x^2 - x(\overset{24.4}{200\cdot 0.122} + \overset{41.16}{196\cdot 0.21} + \overset{47.12}{190\cdot 0.248})$$

$$+ \overset{1220}{0.122\cdot 100^2} + \overset{2016.84}{0.21\cdot 98^2} + \overset{2238.2}{0.248\cdot 95^2}$$

194.2758

$$= 0.58x^2 - 112.68x + 5475.04$$

$$x \approx \frac{112.68}{1.16}$$ #

Problem 2.

$$\frac{1}{1-\gamma} \mathop{E}_{s \sim d_\mu^{\pi_\theta}} \left[\sum_a \pi_\theta(a|s) f(s,a) \right]$$

$$= \frac{1}{1-\gamma} \sum_s d_\mu^{\pi_\theta}(s) \sum_a \pi_\theta(a|s) f(s,a)$$

$$= \frac{1}{1-\gamma} \sum_{s_0} \mu(s_0) d_{s_0}^{\pi}(s) \sum_a \pi_\theta(a|s) f(s,a)$$

$$= \frac{1}{1-\gamma} (1-\gamma) \sum_{s_0} \mu(s_0) \sum_{t=0}^{\infty} \gamma^t P(s_t = s | s_0, \pi) \sum_a \pi_\theta(a|s) f(s,a)$$

$$= \mathop{E}_{\tau \sim P_\mu^{\pi_\theta}} \left[\sum_{t=0}^{\infty} \gamma^t f(s_t, a_t) \right]$$

Problem 3.

$$V(s) = R_T P_T + (R_T + R_S)\cdot P_S P_T + (R_T + 2R_S)\cdot P_S^2 P_T + \cdots$$

$$= R_T P_T (1 + P_S + P_S^2 + \cdots) + R_S P_T (P_S + 2P_S^2 + 3P_S^3 + \cdots)$$

$$1 + P_S + P_S^2 + \cdots = \frac{1 - P_S^{\infty}}{1 - P_S} = \frac{1}{P_T}$$

$$P_S + 2P_S^2 + 3P_S^3 + \cdots$$

$$= (P_S + P_S^2 + P_S^3 + \cdots) + (P_S^2 + P_S^3 + P_S^4 + \cdots) + (P_S^3 + P_S^4 + P_S^5 + \cdots) + \cdots$$

$$= \frac{P_S - P_S^{\infty}}{1 - P_S} + \frac{P_S^2 - P_S^{\infty}}{1 - P_S} + \frac{P_S^3 - P_S^{\infty}}{1 - P_S} + \cdots$$

$$= \frac{P_S + P_S^2 + P_S^3 + \cdots}{1 - P_S} = \frac{\left(\frac{P_S - P_S^{\infty}}{1 - P_S}\right)}{1 - P_S} = \frac{P_S}{P_T^2}$$

$$V(s) = R_T P_T \cdot \frac{1}{P_T} + R_S P_T \cdot \frac{P_S}{P_T^2} = \frac{P_S}{P_T} R_S + R_T \quad \#$$

$$E_\tau[\hat{V}_{MC}(S;\tau)]$$

$$= \sum_{k=0}^{\infty} P_T P_S^{k} \left(\frac{R_S + 2R_S + \cdots + kR_S + (k+1)R_T}{k+1} \right)$$

$$= P_T R_T \sum_{k=0}^{\infty} P_S^{k} + \sum_{k=0}^{\infty} P_T P_S^{k} \left(\frac{R_S + \cdots + kR_S}{k+1} \right)$$

$$\frac{\frac{(k+1)kR_S}{2}}{k+1}$$

$$= P_T R_T \cdot \frac{1}{P_T} + \frac{P_T R_S}{2} \sum_{k=0}^{\infty} k P_S^{k}$$

$$= P_T R_T \cdot \frac{1}{P_T} + \frac{P_T R_S}{2} \cdot \frac{P_S}{P_T^{2}}$$

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

#