# * "Assignment-3" *

Nelakuditi Rahul Naga

AI20BTECH11029

1) a) It is given that $\pi$ is fully supported on $\pi_b$. Hence, importance sampling can be used to estimate $v^{\pi}$. As it is given that the dataset consists of a single sample $(a, r)$, the estimate of $v^{\pi}$ is given by:-

$$\text{Estimate of } v^{\pi} = \frac{\pi(a|s)}{\pi_b(a|s)} \cdot r$$

The estimate of $v^{\pi}$ is unbiased since $E_{a\sim\pi_b}\left[\frac{\pi(a|s)}{\pi_b(a|s)} \cdot r\right] = E_{a\sim\pi}[r]$.

As $E_{a\sim\pi}[r]$ is equal to true $v^{\pi}$ ($\because v^{\pi} = E_{\pi}[r|a\sim\pi]$), we can conclude that the IS estimate of $v^{\pi}$ is unbiased.

b) $$E_{a\sim\pi_b}\left[\frac{\pi(a|.)}{\pi_b(a|.)}\right] = \sum_{a\in A} \frac{\pi(a|.)}{\cancel{\pi_b(a|.)}} \cdot \cancel{\pi_b(a|.)}$$

$$= \sum_{a\in A} \pi(a|.)$$

$$= 1$$

c) It is given that $\pi_b(a|s) = \frac{1}{K} \; \forall a \in A$. Also, as policy $\pi$ is deterministic, $\pi(a|s) = 1$ if $\pi(s) = a$ and $\pi(a|s) = 0$ if $\pi(s) \neq a$.

$$\therefore \; IS = \frac{\pi(a|s)}{\pi_b(a|s)} = \begin{cases} K, & \text{if } a = \pi(s) \\ 0, & \text{otherwise} \end{cases}$$

d) We have already calculated in part (a) that IS estimate of $v^{\pi}$ is given by $\frac{\pi(a|s)}{\pi_b(a|s)} \cdot r$. Say $\alpha = \frac{\pi(a|s)}{\pi_b(a|s)} \Rightarrow$ Estimate of $v^{\pi} = \alpha \cdot r$. As $\pi_b$ is a uniformly random policy, we can assume that action 'a' is drawn from uniform distribution $U$ i.e; $a\sim U$. We need to calculate $V[\alpha r | a\sim U]$ where $V[\cdot]$ denotes variance.

$$V[\alpha r | a\sim U] = r^2 \cdot V[\alpha | a\sim U] \quad (\because r \text{ is a constant})$$

→ Now we know that for a random variable $X$, $V[X] = E[X^2] - (E[X])^2$.

Therefore we have :-

$$V[\alpha\gamma \mid a\sim U] = \gamma^2 \left( E[\alpha^2 \mid a\sim U] - (E[\alpha \mid a\sim U])^2 \right)$$

→ We have already proved in part (b) that $E_{a\sim\pi_b}\left[\frac{\pi(a|.)}{\pi_b(a|.)}\right] = E_{a\sim\pi_b}(\alpha) = 1$.

Here we have $\pi_b = U \Rightarrow E_{a\sim U}(\alpha) = E[\alpha \mid a\sim U] = 1$.

$$\therefore V[\alpha\gamma \mid a\sim U] = \gamma^2 \left( E[\alpha^2 \mid a\sim U] - 1 \right)$$

→ Now, $\alpha^2 = \begin{cases} K^2, & \text{if } a = \pi(s) \\ 0, & \text{otherwise} \end{cases}$ (from part (c)). Therefore we have:-

$$E[\alpha^2 \mid a\sim U] = K^{\cancel{2}} \cdot \frac{1}{\cancel{K}} + 0 \cdot \frac{1}{K} + \dots 0 \cdot \frac{1}{K} = K.$$

$$\therefore V[\alpha\gamma \mid a\sim U] = \gamma^2 (K-1).$$

e) We now have a reward distribution which is bounded in the range $[0,1]$. The variance of IS estimate is given by $V[\alpha\gamma \mid a\sim U]$. As $V[X] = E[X^2] - (E[X])^2$ we have:-

$$V[\alpha\gamma \mid a\sim U] = E[\alpha^2\gamma^2 \mid a\sim U] - (E[\alpha\gamma \mid a\sim U])^2 \leq E[\alpha^2\gamma^2 \mid a\sim U]$$

→ As $\gamma \in [0,1]$, we have :-

$$E[\alpha^2\gamma^2 \mid a\sim U] \leq E[\alpha^2 \mid a\sim U] = K \text{ (from part (d))}$$

$$\therefore V[\alpha\gamma \mid a\sim U] \leq K.$$

f) As $P(\tau)$ is a joint distribution over the trajectory $\tau$ induced by target policy $\pi$, we have :-

$$P(\tau) = \mu(s_0) \cdot \prod_{t=0}^{\infty} \left( P'(s_{t+1} \mid s_t, a_t) \cdot \pi(a_t \mid s_t) \right)$$

where $\mu$ is the initial start state distribution and $P'(s_{t+1} \mid s_t, a_t)$ denotes the state transition probability from $s_t$ to $s_{t+1}$ via action $a_t$.

Similarly, $Q(T)$ is given by :-

$$Q(T) = \mu(s_0) \cdot \prod_{t=0}^{\infty} \left( P'(s_{t+1} | s_t, a_t) \cdot \pi_b(a_t | s_t) \right)$$

Therefore IS weight $\frac{P(T)}{Q(T)}$ is given by :-

$$\frac{P(T)}{Q(T)} = \frac{\cancel{\mu(s_0)} \cdot \prod_{t=0}^{\infty} \left( P'(\cancel{s_{t+1}} | s_t, a_t) \cdot \pi(a_t | s_t) \right)}{\cancel{\mu(s_0)} \cdot \prod_{t=0}^{\infty} \left( P'(\cancel{s_{t+1}} | s_t, a_t) \cdot \pi_b(a_t | s_t) \right)} = \prod_{t=0}^{\infty} \left( \frac{\pi(a_t | s_t)}{\pi_b(a_t | s_t)} \right)$$

2) I have submitted my solution to this question in a ipynb notebook named Assignment 3 - AI20BTECH11029. This file is included in the zip file that I have submitted.