ps4-1-sol

virusdoll

June 16, 2021

(a)

When $\hat{\pi}_0 = \pi_0$,

$$\begin{aligned}
& \mathbb{E}_{\substack{s\sim p(s)\\ a\sim\pi_0(s,a)}} \frac{\pi_1(s,a)}{\hat{\pi}_0(s,a)} R(s,a) \\
= & \sum_{(s,a)} \frac{\pi_1(s,a)}{\hat{\pi}_0(s,a)} R(s,a) p(s,a) \pi_0(s,a) \\
= & \sum_{(s,a)} R(s,a) p(s,a) \pi_1(s,a) \\
= & \mathbb{E}_{\substack{s\sim p(s)\\ a\sim\pi_1(s,a)}} R(s,a)
\end{aligned}$$

(b)

When $\hat{\pi}_0 = \pi_0$,

$$\begin{aligned}
& \frac{\mathbb{E}_{\substack{s\sim p(s)\\ a\sim\pi_0(s,a)}} \frac{\pi_1(s,a)}{\hat{\pi}_0(s,a)} R(s,a)}{\mathbb{E}_{\substack{s\sim p(s)\\ a\sim\pi_0(s,a)}} \frac{\pi_1(s,a)}{\hat{\pi}_0(s,a)}} \\
= & \frac{\sum_{(s,a)} \frac{\pi_1(s,a)}{\hat{\pi}_0(s,a)} R(s,a) p(s,a) \pi_0(s,a)}{\sum_{(s,a)} \frac{\pi_1(s,a)}{\hat{\pi}_0(s,a)} p(s,a) \pi_0(s,a)} \\
= & \frac{\sum_{(s,a)} R(s,a) p(s,a) \pi_1(s,a)}{\sum_{(s,a)} p(s,a) \pi_1(s,a)} \\
= & \frac{\mathbb{E}_{\substack{s\sim p(s)\\ a\sim\pi_1(s,a)}} R(s,a)}{\mathbb{E}_{\substack{s\sim p(s)\\ a\sim\pi_1(s,a)}} 1} \\
= & \mathbb{E}_{\substack{s\sim p(s)\\ a\sim\pi_1(s,a)}} R(s,a)
\end{aligned}$$

(c)

For example, if we have only one single data element such as

$$(s^{(1)}, a^{(1)}, R(s^{(1)}, a^{(1)}))$$

then,

$$\begin{aligned} & \frac{\mathbb{E}_{\substack{s\sim p(s)\\ a\sim\pi_0(s,a)}} \frac{\pi_1(s,a)}{\hat{\pi}_0(s,a)} R(s,a)}{\mathbb{E}_{\substack{s\sim p(s)\\ a\sim\pi_0(s,a)}} \frac{\pi_1(s,a)}{\hat{\pi}_0(s,a)}} \\ = & \frac{\frac{\pi_1(s^{(1)},a^{(1)})}{\hat{\pi}_0(s^{(1)},a^{(1)})} R(s^{(1)}, a^{(1)})}{\frac{\pi_1(s^{(1)},a^{(1)})}{\hat{\pi}_0(s^{(1)},a^{(1)})}} \\ = & R(s^{(1)}, a^{(1)}) \end{aligned}$$

It shows that the result of our weighted importance sampling estimator depends only on this single data element. So it's biased.

(d)

i

When $\hat{\pi}_0 = \pi_0$,

$$\begin{aligned} & \mathbb{E}_{\substack{s\sim p(s)\\ a\sim\pi_0(s,a)}} \big((\mathbb{E}_{a\sim\pi_1(s,a)} \hat{R}(s,a)) + \frac{\pi_1(s,a)}{\hat{\pi}_0(s,a)}(R(s,a) - \hat{R}(s,a))\big) \\ = & \sum_{(s,a)} \big((\sum_a \hat{R}(s,a)\pi_1(s,a)) + \frac{\pi_1(s,a)}{\hat{\pi}_0(s,a)}(R(s,a) - \hat{R}(s,a))\big) p(s)\pi_0(s,a) \\ = & \sum_{(s,a)} \big((\sum_a \hat{R}(s,a)\pi_1(s,a)) p(s)\pi_0(s,a) + (R(s,a) - \hat{R}(s,a)) p(s)\pi_1(s,a)\big) \\ = & \mathbb{E}_{\substack{s\sim p(s)\\ a\sim\pi_1(s,a)}} \mathbb{E}_{a\sim\pi_0(s,a)} \hat{R}(s,a) + \mathbb{E}_{\substack{s\sim p(s)\\ a\sim\pi_1(s,a)}} R(s,a) - \mathbb{E}_{\substack{s\sim p(s)\\ a\sim\pi_1(s,a)}} \hat{R}(s,a) \\ = & \mathbb{E}_{\substack{s\sim p(s)\\ a\sim\pi_1(s,a)}} R(s,a) \end{aligned}$$

ii

When $\hat{R}(s,a) = R(s,a)$,

$$\begin{aligned} & \mathbb{E}_{\substack{s\sim p(s)\\ a\sim\pi_0(s,a)}} \big((\mathbb{E}_{a\sim\pi_1(s,a)} \hat{R}(s,a)) + \frac{\pi_1(s,a)}{\hat{\pi}_0(s,a)}(R(s,a) - \hat{R}(s,a))\big) \\ = & \mathbb{E}_{\substack{s\sim p(s)\\ a\sim\pi_0(s,a)}} \mathbb{E}_{a\sim\pi_1(s,a)} R(s,a) \\ = & \mathbb{E}_{\substack{s\sim p(s)\\ a\sim\pi_1(s,a)}} \mathbb{E}_{a\sim\pi_0(s,a)} R(s,a) \\ = & \mathbb{E}_{\substack{s\sim p(s)\\ a\sim\pi_1(s,a)}} R(s,a) \end{aligned}$$

(e)

Regression estimator:

$$\mathbb{E}_{\substack{s\sim p(s)\\ a\sim\pi_0(s,a)}} \hat{R}(s,a)$$

Importance sampling estimator:

$$\mathbb{E}_{\substack{s\sim p(s)\\ a\sim\pi_0(s,a)}} \frac{\pi_1(s,a)}{\hat{\pi}_0(s,a)} R(s,a)$$

In regression estimator, we want to estimate $\hat{R}(s,a)$. In importance sampling estimator, we want to estimate $\hat{\pi}_0(s,a)$.

i

The interaction between the drug, patient and lifespan is very complicated means $\hat{R}(s,a)$ is hard to estimate. So we use importance sampling estimator.

ii

Drugs are assigned to patients in a very complicated manner means $\hat{\pi}_0(s,a)$ is hard to estimate. So we use regression estimator.