HW2: Problem 1

Mehran Shakerinava

October 2019

For brevity, I'll write $\mathbb{E}_{x\sim p_\theta(X|y)}$ as $\mathbb{E}_{x|y}$.
The following result forms the basis of the solutions to this problem:

$$
\begin{aligned}
\mathbb{E}_{a_t|s_t}\left[\nabla_\theta \log \pi_\theta(a_t|s_t)\right] &= \int \pi_\theta(a_t|s_t)\nabla_\theta \log \pi_\theta(a_t|s_t) da_t \\
&= \int \pi_\theta(a_t|s_t)\frac{\nabla_\theta \pi_\theta(a_t|s_t)}{\pi_\theta(a_t|s_t)} da_t \\
&= \int \nabla_\theta \pi_\theta(a_t|s_t) da_t \\
&= \nabla_\theta \int \pi_\theta(a_t|s_t) da_t \\
&= \nabla_\theta 1 \\
&= 0
\end{aligned}
$$

Part A

$$
\begin{aligned}
\mathbb{E}_\tau\left[\nabla_\theta \log \pi_\theta(a_t|s_t) b(s_t)\right] &= \mathbb{E}_{s_t,a_t}\left[\mathbb{E}_{\tau/s_t,a_t|s_t,a_t}\left[\nabla_\theta \log \pi_\theta(a_t|s_t) b(s_t)\right]\right] \\
&= \mathbb{E}_{s_t,a_t}\left[\nabla_\theta \log \pi_\theta(a_t|s_t) b(s_t)\right] \\
&= \mathbb{E}_{s_t}\left[\mathbb{E}_{a_t|s_t}\left[\nabla_\theta \log \pi_\theta(a_t|s_t) b(s_t)\right]\right] \\
&= \mathbb{E}_{s_t}\left[b(s_t)\mathbb{E}_{a_t|s_t}\left[\nabla_\theta \log \pi_\theta(a_t|s_t)\right]\right] \\
&= 0
\end{aligned}
$$

Part B

a)

Because of the Markov property, given s_t, the distribution of states and actions after time t is independent of states and actions before time t. This implies that $\nabla_\theta \log \pi_\theta(a_t|s_t)b(s_t)$, which is a function of s_t and a_t, is independent of $(s_1, a_1, ..., a_{t-1})$ given s_t, and thus, conditioning on $(s_1, a_1, ..., a_{t-1}, s_t)$ is equivalent to conditioning only on s_t.

b)

$$
\begin{aligned}
\mathbb{E}_\tau\left[\nabla_\theta \log \pi_\theta(a_t|s_t) b(s_t)\right] &= \mathbb{E}_{s_{1:t},a_{1:t-1}}\left[\mathbb{E}_{s_{t+1:T},a_{t:T}|s_{1:t},a_{1:t-1}}\left[\nabla_\theta \log \pi_\theta(a_t|s_t) b(s_t)\right]\right] \\
&= \mathbb{E}_{s_{1:t},a_{1:t-1}}\left[\mathbb{E}_{s_{t+1:T},a_{t:T}|s_t}\left[\nabla_\theta \log \pi_\theta(a_t|s_t) b(s_t)\right]\right] \\
&= \mathbb{E}_{s_{1:t},a_{1:t-1}}\left[\mathbb{E}_{a_t|s_t}\left[\nabla_\theta \log \pi_\theta(a_t|s_t) b(s_t)\right]\right] \\
&= \mathbb{E}_{s_{1:t},a_{1:t-1}}\left[b(s_t)\mathbb{E}_{a_t|s_t}\left[\nabla_\theta \log \pi_\theta(a_t|s_t)\right]\right] \\
&= 0
\end{aligned}
$$