1. Analysis

Consider a discrete MDP, horizon T, an expert policy π^*

Assume:

$$\mathbb{E}_{p_{\pi^*}(s)}\, \pi_\theta(a \neq \pi^*(s) \mid s) = \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}_{p_{\pi^*}(s_t)}\, \pi_\theta(a_t \neq \pi^*(s_t) \mid s_t) \leq \varepsilon$$

1. Show that $\sum_{s_t} |p_{\pi}(s_t) - p_{\pi^*}(s_t)| \leq 2T\varepsilon$

1) $\frac{1}{T}\sum_{t=1}^{T} \mathbb{E}_{p_{\pi}^*(s_t)}\, \pi_\theta(a_t \neq \pi^*(s_t) \mid s_t) = \frac{1}{T}\cdot\sum_{t=1}^{T}\sum_{s_t} p_{\pi}^*(s_t)\, \pi(a_t \neq \pi^*(s_t) \mid s_t) \leq \varepsilon$

$$\sum_{t=1}^{T}\sum_{s_t} p_{\pi}^*(s_t)\, \pi_\theta(a_t \neq \pi^*(s_t) \mid s_t) \leq T\varepsilon$$

$, \sum_{s_t} p_{\pi^*}(s_t)\, \pi_\theta(a_t \neq \pi^*(s_t) \mid s_t) \leq \varepsilon$

2) By the union bound inequality

$$\sum_{s_t} \pi_\theta(a_t \neq \pi^*(s_t) \mid s_t) \leq \sum_{t=1}^{T}\sum_{s_t} p_{\pi}^*(s_t)\, \pi_\theta(a_t \neq \pi^*(s_t) \mid s_t) \leq T\varepsilon$$

$$p_{\pi_\theta}(s_t) = (1-\varepsilon)^t p_{\pi}^*(s_t) + (1-(1-\varepsilon)^t)\, p_{\pi_{mistake}}(s_t)$$

3) $p_{\pi_\theta}(s_t) = (1-T\varepsilon)^t p_{\pi^*_\theta}(s_t) + (1-(1-T\varepsilon)^t)\, p_{mistake}(s_t)$

$$\sum_{s_t} |p_{\pi_\theta}(s_t) - p_{\pi^*}(s_t)| = \sum_{s_t} (1-(1-T\varepsilon)^t)\, |p_{mistake}(s_t) - p_{\pi^*_\theta}(s_t)| \leq 2T\varepsilon$$

$$\sum_{s_t} |p_{\pi_\theta}(s_t) - p_{\pi}^*(s_t)| = \sum_{s_t} (1-(1-\varepsilon)^t)\, |p_{\pi_{mistake}}(s_t) - p_{\pi^*_\theta}(s_t)|$$

$$\leq \sum_{s_t} 2(1-(1-\varepsilon t)) = \sum_{s_t} 2\varepsilon t$$

$$= 2T\varepsilon \quad (\because \sum_{s_t} t = T)$$

2. Assume $|r(s_t)| \leq R_{max}$, $J(\pi) = \sum_{t=1}^{T} \mathbb{E}_{p_\pi(s_t)}\, r(s_t)$

a) Show that $J(\pi^*) - J(\pi_\theta) = O(T\varepsilon)$ s.t. $r(s_t) = 0 \;\; \forall t < T$

$$J(\pi^*) - J(\pi_\theta) = \sum_{t=1}^{T} \mathbb{E}_{p_\pi^*}(s_t)\, r(s_t) - \sum_{t=1}^{T} \mathbb{E}_{p_{\pi_\theta}}(s_t)\cdot r(s_t)$$

$$= \sum_{t=1}^{T} \left(\sum_{s_t} \left(p_{\pi}^*(s_t) - p_{\pi_\theta}(s_t)\right)\right) \cdot r(s_t)$$

$$= \left(p_{\pi}^*(s_T) - p_{\pi_\theta}(s_T)\right) \cdot r(s_T) \leq 2T\varepsilon \cdot R_{max} \qquad O(T\varepsilon)$$

b) $J(\pi^*) - J(\pi_\theta) = \sum_{t=1}^{T} \sum_{s_t} \left(p_{\pi^*}(s_t) - p_{\pi_\theta}(s_t)\right) \cdot r(s_t)$

$$\leq \sum_{t=1}^{T}\sum_{s_t} |p_{\pi_\theta}(s_t) - p_{\pi^*}(s_t)| \cdot R_{max}$$

$$\leq \sum_{t=1}^{T} 2T\varepsilon \cdot R_{max} = 2R_{max}\cdot T^2 \cdot \varepsilon \qquad O(T^2\varepsilon)$$