1. $\sum_{s_t} |P_{\pi_\theta}(s_t) - P_{\pi^*}(s_t)|$ (1)

consider $P_{\pi_\theta}(s_t) = P_{\pi_\theta}(s_t \mid \text{no mistake}) \cdot P_{\pi_\theta}(\text{no mistake})$

$+ P_{\pi_\theta}(s_t \mid \text{mistake}) \cdot P_{\pi_\theta}(\text{mistake})$

$P_{\pi_\theta}(s_t \mid \text{no mistake}) = P_{\pi^*}(s_t)$, $P_{\pi_\theta}(\text{no mistake}) = 1 - P_{\pi_\theta}(\text{mistake})$

Consider $P_{\pi_\theta}(\text{mistake})$

$= P_{\pi_\theta} \cdot \left(\bigcup_{t' \in [1,t]} \{\pi_\theta \text{ makes mistake at } t', \text{ but not before.}\} \right)$

$\leq \sum_{t' \in [1,t]} P_{\pi_\theta} [\pi_\theta \text{ makes mistake at } t', \text{ but not before}]$

Consider a Markov chain of three events

① π_θ makes mistake at t'

② s_t

③ π_θ makes no mistake before t'

$\Rightarrow$ ③ $\rightarrow$ ② $\rightarrow$ ①

$\sum_{t' \in [1,t]} P_{\pi_\theta} [\pi_\theta \text{ makes mistake at } t', \text{ but not before}]$

$= \sum_{t' \in [1,t]} \sum_{s_t} P_{\pi_\theta}(① \wedge ② \wedge ③)$

$= \sum_{t' \in [1,t]} \sum_{s_t} P_{\pi_\theta}(① \mid ②) P_{\pi_\theta}(② \mid ③) \cdot P_{\pi_\theta}(③)$ (2)

Note that $P_{\pi_\theta}(① \mid ②) = \pi_\theta(a_{t'} \neq \pi_\theta(s_{t'}) \mid s_{t'})$

$P_{\pi_\theta}(② \mid ③) = P_{\pi^*}(s_t)$

$P_{\pi_\theta}(③) < 1$

$\therefore$ (2) $\leq \sum_{t' \in [1,t]} E_{P_{\pi^*}(s_t)} \pi_\theta(a_{t'} \neq \pi_\theta(s_{t'}) \mid s_{t'})$

$\leq \sum_{t}^{T} E_{P_{\pi^*}(s_t)} \cdot \pi_\theta(a_t \neq \pi_\theta(s_t) \mid s_t)$

$$(1) = \sum_{s_t} P_{\pi_\theta}(\text{mistake}) \cdot \left| P_{\pi_\theta}(s_t \mid \text{mistake}) - P_{\pi^*}(s_t) \right|$$

$$\leq \underbrace{\sum_t^T E_{p_{\pi^*}(s_t)} \pi_\theta(a_t \neq \pi_\theta(s_t) \mid s_t)}_{\leq \varepsilon T} \cdot \underbrace{\sum_{s_t} \left| P_{\pi_\theta}(s_t \mid \text{mistake}) - P_{\pi^*}(s_t) \right|}_{\leq 2}$$

$$\leq 2 \varepsilon T$$

$\boxed{}$

2.

(a) $$J(\pi^*) - J(\pi_\theta) = E_{p_{\pi^*}(s_T)} r(s_T) - E_{p_{\pi_\theta}(s_T)} r(s_T)$$

$$= \sum_{s_T} \left(P_{\pi^*}(s_T) - P_{\pi_\theta}(s_T) \right) \cdot r(s_T)$$

$$\leq r(s_T) \cdot \sum_{s_T} \left| P_{\pi^*}(s_T) - P_{\pi_\theta}(s_T) \right|$$

$$\leq R_{max} \cdot 2T\varepsilon$$

$$\therefore \quad J(\pi^*) - J(\pi_\theta) = O(T\varepsilon)$$

$\boxed{}$

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

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

$$\leq T \cdot R_{max} \cdot 2T\varepsilon$$

$$\therefore \quad J(\pi^*) - J(\pi_\theta) = O(T^2 \varepsilon)$$