Data: $(s,a) \overset{iid}{\sim} \mu$, $r \sim R(s,a)$, $s' \sim P(\cdot|s,a)$ $\quad \mathcal{T} f_{k-1}$

FQI: $$f_k \leftarrow \underset{f \in \mathcal{F}}{\operatorname{argmin}} \sum_{(s,a,r,s')} \left(f(s,a) - r - \gamma \max_{a'} f_{k-1}(s',a') \right)^2.$$

$\pi = \pi_{f_k}$

Assumptions: ① $\mathcal{T} f \in \mathcal{F}, \forall f \in \mathcal{F}$.

② $\frac{d_f^{\pi}(s,a)}{\mu(s,a)} \le C, \forall s,a,t,\pi$.

FQE: $$f_k \leftarrow \underset{f \in \mathcal{F}}{\operatorname{argmin}} \sum_{(s,a,r,s')} \left(f(s,a) - r - \gamma f_{k-1}(s',\pi) \right)^2$$

$J(\pi) = \mathbb{E}_{s \sim d_0}[V^{\pi}(s)] = \mathbb{E}_{s \sim d_0}[Q^{\pi}(s,\pi)]$.

$\hat{J}(\pi) = \mathbb{E}_{s \sim d_0}[f_k(s,\pi)]$.

Assum: ① $\mathcal{T}^{\pi} f \in \mathcal{F} \quad \forall f \in \mathcal{F}$.

② $\frac{d_f^{\pi}(s,a)}{\mu(s,a)} \le C \quad \forall s,a,t$.

$$L_D(f; f') = \frac{1}{|D|} \sum_{(s,a,r,s') \in D} \left(f(s,a) - r - \gamma f'(s',\pi) \right)^2.$$

$f \to \mathcal{T} f_k$, $f' \to \mathcal{T} f_{k-1}$

$$f_k \leftarrow \underset{f \in \mathcal{F}}{\operatorname{argmin}} L_D(f; f_{k-1})$$

$$L_{\mu}(f; f') = \mathbb{E}_D[L_D(f; f')].$$

Concentration analysis | define. | $\mathbb{E}_\mu[(\cdot)^2] = \|\cdot\|_{2,\mu}^2$

$$\mathbb{E}_\mu\left[\left(f_k - \mathcal{T}^\pi f_{k-1}\right)^2\right] =: \|f_k - \mathcal{T}^\pi f_{k-1}\|_{2,\mu}^2.$$

Fact: $\|f_k - \mathcal{T}^\pi f_{k-1}\|_{2,\mu}^2 = \mathcal{L}_\mu(f_k; f_{k-1})$ ✓

$$- \mathcal{L}_\mu(\mathcal{T}^\pi f_{k-1}; f_{k-1})$$ ✓

Hoeffding: $|\mathcal{L}_D(f; f') - \mathcal{L}_\mu(f; f')| \le \varepsilon$.

$\forall f, f' \in \mathcal{F}$. $\quad \varepsilon \propto \sqrt{\frac{1}{n} \log \frac{|\mathcal{F}|}{\delta}}$

$$\le \mathcal{L}_D(f_k; f_{k-1}) - \mathcal{L}_D(\mathcal{T}^\pi f_{k-1}; f_{k-1}) + 2\varepsilon.$$

$f \in \mathbb{R}^{S \times A}$

$\mathcal{T}^\pi f_{k-1} \in \mathcal{F}$

$$\le 2\varepsilon.$$

$$X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \qquad \|X\|_2^2 = x_1^2 + x_2^2 + x_3^2$$

$$\mu = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$$

$$\|x\|_{2,\mu} = \sqrt{\mathbb{E}_{i\sim\mu}[x_i^2]} = \sqrt{\frac{x_1^2+x_2^2+x_3^2}{3}}.$$

$$\|x\|_1 \geq \|x\|_2 \geq \|x\|_\infty.$$

$$\|x\|_{1,\mu} \leq \|x\|_{2,\mu} \leq \|x\|_\infty.$$

$p = 1, 2, \infty.$

$\|x\|_{p,\mu}$ ← distribution.

"nu"

$$\|x\|_{p,\nu} \leq C^{1/p} \|x\|_{p,\mu} \quad \text{where} \quad \frac{\nu(i)}{\mu(i)} \leq C \ \forall i.$$

$$\|x\|_{p,\nu}^p := \mathbb{E}_{i\sim\nu}[|x_i|^p].$$

$$= \sum_i \nu(i) |x_i|^p$$

$$= \sum_i \frac{\nu(i)}{\mu(i)} \mu(i) |x_i|^p$$

$$\leq C \|x\|_{p,\mu}^p.$$

$$\int \cdots d\pi$$

$$\left|\hat{J}(\pi) - J(\pi)\right| \qquad d_0 = d_1.$$

$$= \left| \mathbb{E}_{s\sim d_0}[f_K(s,\pi)] - \mathbb{E}_{s\sim d_0}[Q^\pi(s,\pi)] \right|$$

$$= \Big| \mathbb{E}_{s\sim d_0}\Big[f_K(s,\pi) - (\mathcal{T}^\pi f_{K-1})(s,\pi) + (\mathcal{T}^\pi f_{K-1})(s,\pi) \Big] - \mathbb{E}_{s\sim d_0}\big[(\mathcal{T}^\pi Q^\pi)(s,\pi) \big] \Big|$$

$$\le \Big| \mathbb{E}_{\substack{s\sim d_0 \\ a\sim\pi(\cdot|s)}} \big[(f_K(s,a) - (\mathcal{T}^\pi f_{K-1})(s,a)) \big] \Big| \quad (\mathrm{I})$$

$$+ \Big| \mathbb{E}_{s\sim d_0}\big[R(s,\pi) + \gamma \mathbb{E}_{s'\sim P(\cdot|s,\pi)}[f_{K-1}(s',\pi)] \big] - \mathbb{E}_{s\sim d_0}\big[R(s,\pi) + \gamma \mathbb{E}_{s'\sim P(\cdot|s,\pi)}[Q^\pi(s',\pi)] \big] \Big| \quad (\mathrm{II})$$

$$(\mathrm{I}) \le \|f_K - \mathcal{T}^\pi f_{K-1}\|_{1,\, d_0\times\pi}$$

$$\le \|f_K - \mathcal{T}^\pi f_{K-1}\|_{2,\, d_0\times\pi} \qquad (s,a)\sim d_1^\pi$$

$$\le \|f_K - \mathcal{T}^\pi f_{K-1}\|_{2,\,\mu} \times \sqrt{C}.$$

$$(\text{II}) = \Big| \mathbb{E}_{s\sim d_0}\Big[R(s,\pi) + \gamma \mathbb{E}_{s'\sim P(\cdot|s,\pi)}[f_{k-1}(s',\pi)] \Big] - \mathbb{E}_{s\sim d_0}\Big[R(s,\pi) + \gamma \mathbb{E}_{s'\sim P(\cdot|s,\pi)}[Q^{\pi}(s',\pi)] \Big] \Big|$$

$$= \gamma \Big| \mathbb{E}_{s'\sim d_2^{\pi}}\Big[f_{k-1}(s',\pi) - Q^{\pi}(s',\pi) \Big] \Big| \leq \cdots \leq O(\gamma^k V_{max})$$

$$\mathbb{E}_{s\sim d_0}[f_K(s,\pi_1)] - J(\pi) = \left(\sum_{t=1}^{K} \gamma^{t-1} \mathbb{E}_{d_t^{\pi}}[f_{K-t+1} - \mathcal{T}^{\pi_{t+1}} f_{K-t}] \right) - \mathbb{E}[\textstyle\sum_{t=K+1}^{\infty} \gamma^{t-1} r_t | \pi, d_0].$$

$\{f_1, f_2, \cdots, f_K\}$. $\pi_1, \pi_2, \cdots, \pi_{K+1} = \pi$.

$$\mathbb{E}_{s\sim d_0}[f(s,\pi)] - J(\pi) = \frac{1}{1-\gamma} \mathbb{E}_{d^{\pi}}[f - \mathcal{T}^{\pi} f].$$

FQI: $\|f_k - \mathcal{T} f_{k-1}\|_{2,\mu} \le \epsilon$.

$$J(\pi) - J(\hat{\pi}) \le \sum_{t=1}^{K} \gamma^{t-1} \left(\mathbb{E}_{d_t^{\pi}}[\mathcal{T} f_{K-t} - f_{K-t+1}] + \mathbb{E}_{d_t^{\hat{\pi}}}[f_{K-t+1} - \mathcal{T} f_{K-t}] \right) + \gamma^K V_{\max}.$$

$\pi_{f_1}, \pi_{f_2}, \ldots$: $\hat{\pi} = \{\pi_t\}_{t=1:K}$, $\pi_t = \pi_{f_{K-t+1}}$.

$\hat{\pi} = \pi_{f_K}$.

$$J(\pi^*) - J(\pi_{f_K})$$

$$= \frac{1}{1-\gamma} \mathbb{E}_{s \sim d^{\pi_{f_K}}} [V^*(s) - Q^*(s, \pi_{f_K})].$$

$$\le \frac{1}{1-\gamma} \mathbb{E}_{s \sim d^{\pi_{f_K}}} [V^*(s) - V^{\hat{\pi}}(s)].$$

$$= \frac{1}{1-\gamma} J_{d^{\pi_{f_K}}}(\pi^*) - J_{d^{\pi_{f_K}}}(\hat{\pi}).$$