① $\ell: \mathcal{X} \to \mathbb{R}$. Def $\|\ell\|_{p,\mu} := \left(\mathbb{E}_\mu[|\ell|^p]\right)^{1/p}$.

for $1 \le p \le q$, $\|\ell\|_{p,\mu} \le \|\ell\|_{q,\mu}$.

② for $\mu, \nu \in \Delta(\mathcal{X})$. Assume $\left\|\frac{\nu}{\mu}\right\|_\infty = \max_x \frac{\nu(x)}{\mu(x)} \le C$.

then $\|\ell\|_{p,\nu} \le C^{1/p} \|\ell\|_{p,\mu}$

③ Regression: let $\{(X_i, Y_i)\}_{i=1}^n \overset{\text{i.i.d}}{\sim} P_{X,Y}$.

$\hat{f} := \operatorname{argmin}_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^n (f(X_i) - Y_i)^2 =: \hat{\mathbb{E}}[(f(X)-Y)^2]$

& $f^* = \left(x \mapsto \mathbb{E}[Y \mid X = x]\right) \in \mathcal{F}$.

then ① $\mathbb{E}[(\hat{f}(X) - Y)^2] \le \mathbb{E}[(f^*(X) - Y)^2] + \varepsilon$.

② $\mathbb{E}[(\hat{f} - f^*)^2] \le \varepsilon$.

---

def: $\hat{L}(f) := \hat{\mathbb{E}}[(f(X)-Y)^2]$, $L(f) = \mathbb{E}[\cdots]$.

$\hat{f} = \operatorname{argmin}_{f \in \mathcal{F}} \hat{L}(f)$. $\quad L(\hat{f}) \le L(f^*) + \varepsilon$.

$\forall f \in \mathcal{F}$: w.p. $\ge 1-\delta$. $\quad |\hat{L}(f) - L(f)| \le O\left(\sqrt{\frac{1}{n} \ln \frac{1}{\delta}}\right)$.

$\Downarrow$

w.p. $\ge 1-\delta$, $\forall f \in \mathcal{F}$. $\quad \le O\sqrt{\frac{1}{n} \ln \frac{|\mathcal{F}|}{\delta}} := \varepsilon'$

$$\Rightarrow L(\hat{f}) \leq \hat{L}(\hat{f}) + \varepsilon'$$
$$\leq \hat{L}(f^*) + \varepsilon'$$
$$\leq L(f^*) + 2\varepsilon'.$$

FQE: Given (s,a,r,s'), $s,a \sim \mu$. & π.

learn Q^π : for $k=1,2,\cdots$.

$$f_k = \underset{f \in \mathcal{F}}{\operatorname{argmin}} \hat{\mathbb{E}}\left[\left(f(s,a) - r - \gamma f_{k-1}(s',\pi)\right)^2\right]$$

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

② $\|d_t^\pi / \mu\|_\infty \le C \quad \forall t$.

Concentration: for each regression:

$$f_k^* = \mathbb{E}\left[r + \gamma f_{k-1}(s',\pi) \mid s,a\right]$$

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

$$\Rightarrow \mathbb{E}_\mu\left[(f_k - \mathcal{T}^\pi f_{k-1})^2\right] \le \varepsilon.$$

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

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

$$\left|\hat{J}(\pi) - J(\pi)\right| \le \mathbb{E}_{s \sim d_0}\left[f_k(s,\pi) - Q^\pi(s,\pi)\right].$$

$$\leq \mathbb{E}_{(s,a)\sim d_1^{\pi}}\left[\left| f_k(s,a) - Q^{\pi}(s,a)\right|\right] \rightarrow \|\cdot\|_{1,d_1^{\pi}}$$

$$\leq \sqrt{\mathbb{E}_{(s,a)\sim d_1^{\pi}}\left[\left(f_k(s,a)-Q^{\pi}(s,a)\right)^2\right]} \rightarrow \|\cdot\|_{2,d_1^{\pi}}$$

$$= \|f_k - Q^{\pi}\|_{2,d_1^{\pi}}$$

$$\leq \|f_k - \mathcal{T}^{\pi} f_{k-1}\|_{2,d_1^{\pi}} + \|\mathcal{T}^{\pi} f_{k-1} - \mathcal{T}^{\pi} Q^{\pi}\|_{2,d_1^{\pi}}$$

$$\leq \gamma \|f_{k-1} - Q^{\pi}\|_{2,d_2^{\pi}}$$

$$\leq \sqrt{C}\cdot \|f_k - \mathcal{T}^{\pi} f_{k-1}\|_{2,\mu} + \sqrt{\mathbb{E}_{(s,a)\sim d_1^{\pi}}\left[\left(R(s,a)+\gamma \mathbb{E}_{s'\sim P(\cdot|s,a)}\left[f_{k-1}(s',\pi)\right] - R(s,a) - \gamma \mathbb{E}_{s'\sim P(\cdot|s,a)}\left[Q^{\pi}(s',\pi)\right]\right)^2\right]}$$

$$\leq \sqrt{C}\cdot\sqrt{\varepsilon} + \gamma\sqrt{\mathbb{E}_{(s,a)\sim d_1^{\pi}} \mathbb{E}_{s'\sim P(\cdot|s,a)}\left[\left(f_{k-1}(s',\pi) - Q^{\pi}(s',\pi)\right)^2\right]}$$

$$= \sqrt{C}\cdot\sqrt{\varepsilon} + \gamma\sqrt{\mathbb{E}_{(s,a)\sim d_2^{\pi}}\left[\left(f_{k-1}(s,a) - Q^{\pi}(s,a)\right)^2\right]}$$

FQE: Given (s,a,r,s'), $s,a \sim \cdot$ & π.

learn Q^π : $\nearrow T^\pi f$.

$$\hat{f} = \underset{f\in F}{\operatorname{argmin}} \max_{g\in F} \hat{\mathbb{E}}\left[(f(s,a)-r-\gamma f(s',\pi))^2\right] - \hat{\mathbb{E}}\left[(g(s,a)-r-\gamma f(s',\pi))^2\right]$$

$$\Rightarrow \mathbb{E}_{\mu}\left[(\hat{f}-T^\pi \hat{f})^2\right] \le \varepsilon.$$

Assumptions: ① $T^\pi f \in F, \ \forall f \in F.$

② $\|d^\pi/\mu\|_\infty \le C. \quad \forall t.$

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

$$= \left|\mathbb{E}_{(s,a)\sim d^\pi}\left[(\hat{f}-T^\pi\hat{f})(s,a)\right]\right|$$

$$\le \|\hat{f}-T^\pi\hat{f}\|_{2,d^\pi}$$

$$\le \sqrt{C}\cdot\|\hat{f}-T^\pi\hat{f}\|_{2,\mu} \le \sqrt{C\cdot\varepsilon}.$$

FQI (minimax ver). Given (s,a,r,s'), $s,a \sim \mu$

learn Q^*:

$$\hat{f} = \underset{f \in F}{\operatorname{argmin}} \max_{g \in F} \hat{\mathbb{E}}\left[\left(f(s,a) - r - \gamma f(s', \pi_f)\right)^2\right] - \hat{\mathbb{E}}\left[\left(g(s,a) - r - \gamma f(s', \pi_f)\right)^2\right]$$

$$\Rightarrow \boxed{\mathbb{E}_\mu\left[\left(\hat{f} - \mathcal{T}\hat{f}\right)^2\right] \le \varepsilon.}$$

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

② $\|d^\pi / \mu\|_\infty \le C. \quad \forall \pi.$

$$J(\pi^*) - J(\pi_f) \le \frac{1}{1-\gamma}\left(\mathbb{E}_{d^{\pi^*}}[\mathcal{T} f - f] + \mathbb{E}_{d^{\pi_f}}[f - \mathcal{T} f]\right)$$

$$J(\pi^*) - J(\pi_{\hat{f}}) \le$$

$$\le \frac{1}{1-\gamma}\left(\|\mathcal{T}\hat{f} - \hat{f}\|_{2,d^{\pi^*}} + \|\mathcal{T}\hat{f} - \hat{f}\|_{2,d^{\hat{f}}}\right)$$

$$\le \frac{2\sqrt{C}}{1-\gamma} \underbrace{\|\mathcal{T}\hat{f} - \hat{f}\|_{2,\mu}} \to \sqrt{\varepsilon}.$$

FQI: Given (s,a,r,s'), $s,a \sim \mu$. $\mathbb{E}_\mu[(f_k - \mathcal{T} f_{k-1})^2] \leq \varepsilon$.

learn Q^*: for $k = 1, 2, \cdots$

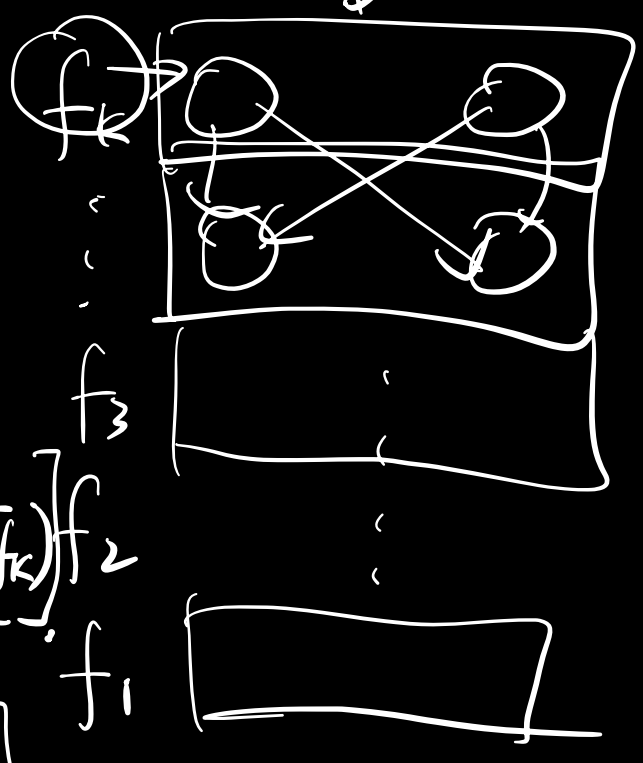
$$f_k = \operatorname*{argmin}_{f \in \mathcal{F}} \hat{\mathbb{E}}\left[\left(f(s,a) - r - \gamma \max_{a'} f(s',a')\right)^2\right].$$

Assumptions: ① $\mathcal{T} f \in \mathcal{F} \quad \forall f \in \mathcal{F}$. (incl. non-stat policies.)

② $\|d_t^\pi / \mu\|_\infty \leq C \quad \forall t, \pi$

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

greedy w.r.t. $f_k \circ f_{k-1} \circ f_{k-2} \circ \cdots \circ$



FQI policy: π_{f_k}.

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

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

$$\leq \frac{1}{1-\gamma} \mathbb{E}_{\square}[V^*(s) - V^{\hat{\pi}}(s)].$$