OPE

IS: exp variance. (unless $\pi \approx \pi_b$).

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

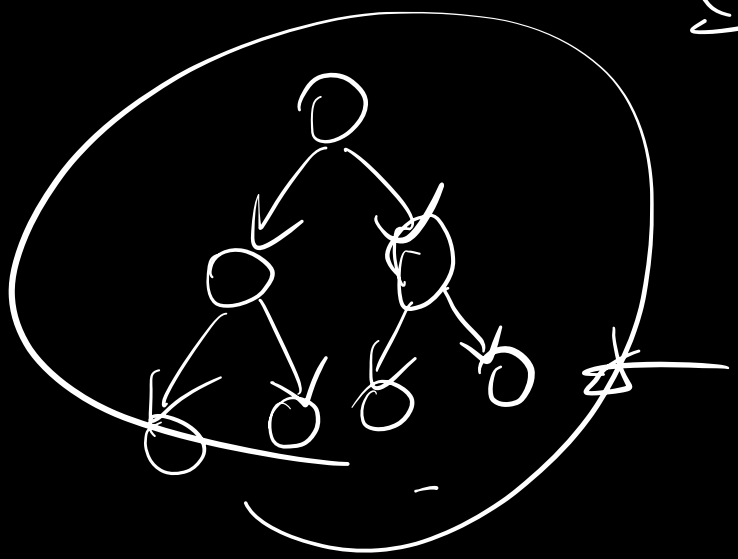
Require $\begin{cases} \mathcal{T}^\pi f \in \mathcal{F} \quad \forall f \in \mathcal{F} \\ \left\| \frac{d^\pi}{\mu} \right\|_\infty \leq C \end{cases}$

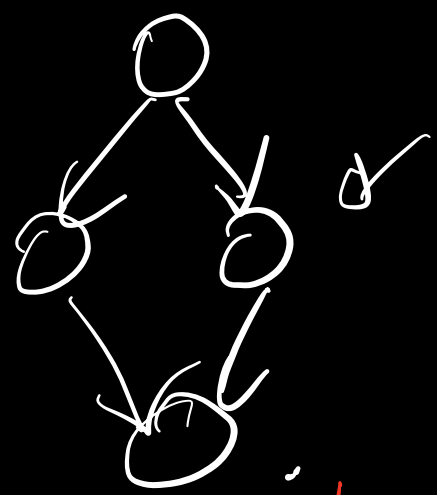
Data: $(s,a) \sim \mu$. $r \sim R(s,a)$ $s' \sim P(\cdot|s,a)$

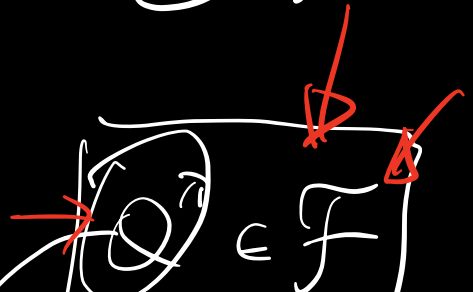
$$J(\pi) = \mathbb{E}_{d_0}[V^\pi(s)] = \frac{1}{1-\gamma} \mathbb{E}_{(s,a) \sim d^\pi}[r],$$

$r \sim R(s,a)$

$$= \frac{1}{1-\gamma} \mathbb{E}_{(s,a) \sim \mu}\left[\frac{d^\pi(s,a)}{\mu(s,a)} \cdot r\right].$$







Setup: data $\sim \mu$ $\quad$ $w^\pi \in \mathcal{W}$ $\quad$ $w^\pi(s,a) = \frac{d^\pi(s,a)}{\mu(s,a)}$

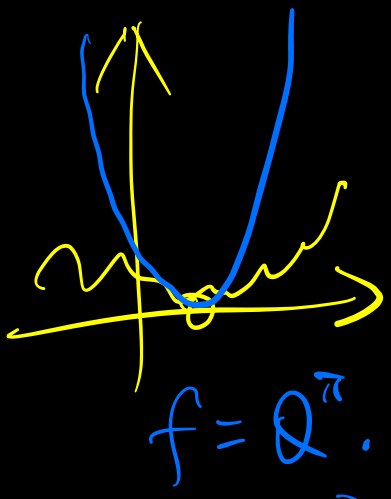
Goal: estimate $J(\pi) = \mathbb{E}_\mu[w^\pi \cdot r]/(1-\gamma)$.

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

$$L(f) = |\cdot| = \frac{1}{1-\gamma}\left|\mathbb{E}_{d^\pi}[(f - \mathcal{T}^\pi f)]\right|$$

$$= \frac{1}{1-\gamma}\left|\mathbb{E}_\mu\left[\frac{d^\pi(s,a)}{\mu(s,a)}(f - \mathcal{T}^\pi f)\right]\right|$$

$$\le \frac{1}{1-\gamma}\max_{w\in\mathcal{W}}\left|\mathbb{E}_\mu[w\cdot(f - \mathcal{T}^\pi f)]\right|$$



$$= \frac{1}{1-\gamma}\max_{w\in\mathcal{W}}\left|\mathbb{E}_{(s,a,r,s')\sim\mu}[w(s,a)(f(s,a) - r - \gamma f(s',\pi))]\right|$$

Alg: $\hat{f} = \arg\min_{f\in\mathcal{F}}\max_{w\in\mathcal{W}}$

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

f: $\forall w \in W$, $\mathbb{E}_\mu[w \cdot (f - \mathcal{T}^\pi f)] = 0$.

$\forall (s,a)\ (f - \mathcal{T}^\pi f)(s,a) = 0$.

$\left| \frac{1}{1-\gamma} \mathbb{E}_\mu\left[\frac{d}{\mu} R\right] - J(\pi) \right|$ $\quad w = \frac{d^\pi}{\mu}$.

$$\frac{1}{1-\gamma}\mathbb{E}_d[R] - J(\pi) = \frac{1}{1-\gamma}\mathbb{E}_{(s,a)\sim d, s'\sim P(\cdot|s,a)}[Q^\pi(s,a) - \gamma Q^\pi(s',\pi)] - \mathbb{E}_{s\sim d_0}[Q^\pi(s,\pi)]$$

$\hat{w} = \operatorname{argmin}_{w \in W} \max_{f \in \mathcal{F}} \left| \quad f \quad \right|$

$\hat{J}(\pi) = \frac{1}{1-\gamma} \mathbb{E}_\mu[\hat{w} \cdot r]$.

$O\left(\|w\|_\infty \|\mathcal{F}\|_\infty \sqrt{\frac{1}{n} \ln \frac{F}{\delta}}\right)$