Marginalized IS

IS: $\mathbb{E}_{x\sim q}\left[\frac{p(x)}{q(x)} f(x)\right] = \mathbb{E}_p[f]$

In MDPs: $\mathbb{E}_{\pi_b}\left[\frac{P^{\pi}(s_1,a_1,\dots,s_t,a_t)}{P^{\pi_b}(s_1,a_1,\dots,s_t,a_t)} r_t\right] = \mathbb{E}_{\pi}[r_t]$.

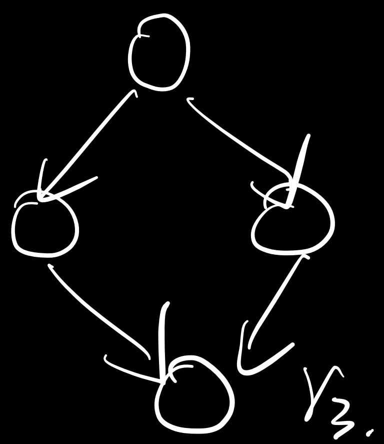
$\prod_{t'=1}^{t} \pi(a_{t'}|s_{t'})/\pi_b(a_{t'}|s_{t'})$

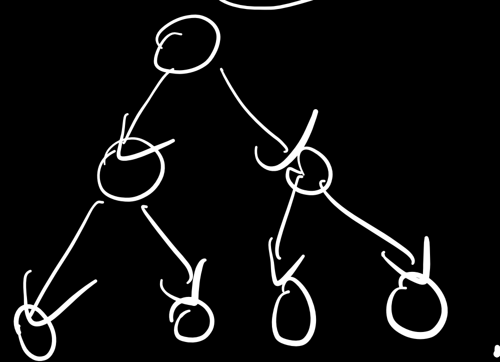
$= \mathbb{E}_{\pi_b}\left[\frac{d_t^{\pi}(s_t,a_t)}{d_t^{\pi_b}(s_t,a_t)} r_t\right]$.

$(x,y)\sim p$. $(x,y)\sim q$.

$$\mathbb{E}_{(x,y)\sim p}[f(x)] = \mathbb{E}_q\left[\frac{p(x,y)}{q(x,y)} f(x)\right].$$

$$= \mathbb{E}_q\left[\frac{p(x)}{q(x)} f(x)\right].$$





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

Assume: $\left.\begin{array}{l} w^{\pi} \in \mathcal{W} \\ \boxed{Q^{\pi} \in \mathcal{F}} \end{array}\right\}$

MQL) $\forall f : \mathbb{R}^{S\times A}$

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

$$= \left|\frac{1}{1-\gamma}\mathbb{E}_\mu\left[\frac{d^\pi(s,a)}{\mu(s,a)}\left((f-\mathcal{T}^\pi f)(s,a)\right)\right]\right| \qquad \mathbb{E}_{\substack{(s,a)\sim\mu,\\ r,s'}}[\delta(s,a,r,s')]$$

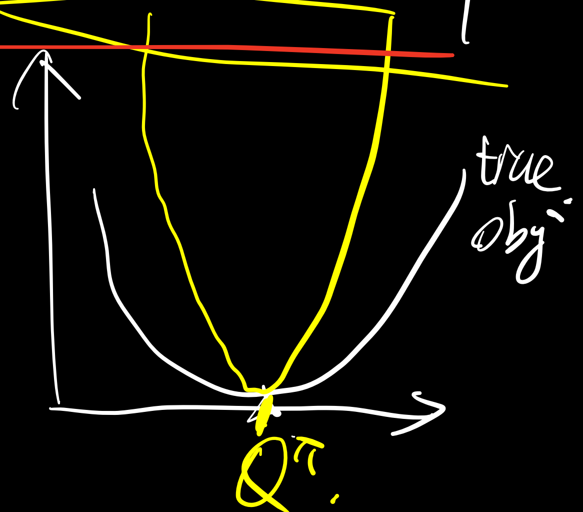
$$= \left|\frac{1}{1-\gamma}\mathbb{E}_{\substack{(s,a)\sim\mu\\ r,s'}}\left[w^\pi(s,a)\big(f(s,a) - r - \gamma f(s',\pi)\big)\right]\right|.$$

$(f - \mathcal{T}^\pi f)$

$w^\pi \in \mathcal{W}$

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

$$\hat{f} := \operatorname*{argmin}_{f\in\mathcal{F}} L_g(f).$$



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

$\mathbb{E}_\mu[w(\cdots)]$.

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

$\downarrow$ $w(s,a) = d(s,a)/\mu(s,a)$

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

$$\left| \frac{1}{1-\gamma} \mathbb{E}_{\mu}\left[w(s,a)\left(Q^{\pi}(s,a) - \gamma Q^{\pi}(s',\pi) \right)\right] - \mathbb{E}_{s\sim d_0}\left[Q^{\pi}(s,\pi)\right] \right|$$

$$\leq \max_{\boxed{f\in\mathcal{F}}} \left| \frac{1}{1-\gamma} \mathbb{E}_{\mu}\left[w(s,a)\left(f(s,a) - \gamma f(s',\pi) \right)\right] - \mathbb{E}_{d_0}\left[f(s,\pi)\right] \right|$$