Tabular Method

Data $\{(s,a,r,s')\}$.

→ Assume: each (s,a) appears n times.

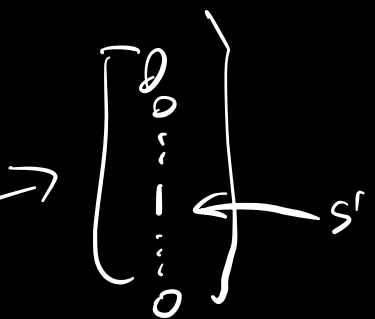
$r \sim R(\cdot|s,a)$, $s' \sim P(\cdot|s,a)$

Let $D_{s,a}$ be the n (r,s') pairs from (s,a).

Alg: "certainty-equivalence". estimate $\hat{M}$.

$$\hat{R}(s,a) = \frac{1}{n} \sum_{(r,s') \in D_{s,a}} r.$$

$$\hat{P}(\cdot|s,a) = \frac{1}{n} \sum_{(r,s') \in D_{s,a}} e_{s'}$$



Output: $\pi^*_{\hat{M}}$. $V_M^{\pi^*_{\hat{M}}}$.

Analysis: $\| V_M^* - V_M^{\pi^*_{\hat{M}}} \|_\infty \leq O\left(\frac{1}{\sqrt{n}}\right)$

w.p. $\geq 1-\delta$.

Define:

$$\varepsilon_R := \max_{s,a} |\hat{R}(s,a) - R(s,a)|.$$

$$\varepsilon_P := \max_{s,a} \| \hat{P}(\cdot|s,a) - P(\cdot|s,a) \|_1.$$

Simulation Lemma: $\forall \pi: S \to A \quad (S \to \Delta(A))$

$$\|V_M^{\pi} - V_{\hat{M}}^{\pi}\|_\infty \le \frac{\epsilon_R}{1-\gamma} + \frac{\gamma \epsilon_P V_{max}}{2(1-\gamma)}$$

$V_{max} = \frac{R_{max}}{1-\gamma}$

① $\|V_M^* - V_M^{\pi_{\hat{M}}^*}\|_\infty \le$ (× 2)

② Prove simu Lemma.

③ Bound ϵ_R, ϵ_P as a fn of n.

$\forall s,$

$$V_M^*(s) - V_M^{\pi_{\hat{M}}^*}(s) = V_M^{\pi_M^*}(s) - V_{\hat{M}}^{\pi_M^*}(s)$$

"decision loss"

$$+ V_{\hat{M}}^{\pi_{\hat{M}}^*}(s) - V_M^{\pi_{\hat{M}}^*}(s)$$

($V_{\hat{M}}^{\pi_M^*}(s) \le V_{\hat{M}}^{\pi_{\hat{M}}^*}$)

$$\le 2 \cdot \max_{\pi: S \to A} \|V_M^{\pi} - V_{\hat{M}}^{\pi}\|_\infty. \quad \text{"evaluation error"}$$

$\forall f \in \mathbb{R}^S \quad \forall s_0 \quad \forall \pi: S \to \Delta(A)$

$f = V_{\hat{M}}^{\pi}(s_0)$

$$f(s_0) - V_M^{\pi}(s_0) = \frac{1}{1-\gamma} \mathbb{E}_{\substack{s \sim d^{\pi, s_0} \\ a \sim \pi(\cdot|s) \\ r \sim R(\cdot|s,a)}} \left[f(s) - r - \gamma f(s') \right].$$

$\mathbb{E}[\ (\ \ )\ |\ s\ ]$

$s' \sim P(\cdot|s,a)$.

$$= \frac{1}{1-\gamma} \mathbb{E}_{s \sim d^{\pi,s_0}} \left[f(s) - \left(R(s,\pi) + \gamma \cdot \mathbb{E}_{s' \sim P(s,\pi)} [f(s')] \right) \right]$$

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

Bellman error / residual. $(f - \mathcal{T} f)$

$$\| f - V_M^\pi \|_\infty \leq \frac{1}{1-\gamma} \| f - \mathcal{T}_M^\pi f \|_\infty$$

$$\| f - V_M^\pi \|_\infty \leq \| f - \mathcal{T}_M^\pi f \|_\infty + \| \mathcal{T}_M^\pi f - \mathcal{T}_M^\pi V_M^\pi \|_\infty$$

$$\leq \gamma \cdot \| f - V_M^\pi \|_\infty$$

$$\frac{1}{1-\gamma}\,\mathbb{E}_{\substack{s\sim d^{\pi,s_0}\\ a\sim\pi(\cdot|s),\\ r\sim R(\cdot|s,a),\\ s'\sim P(\cdot|s,a)}}\left[f(s)-r+\gamma f(s')\right].$$

$$= V_M^{\pi}(s_0).$$

$$\mathbb{E}_{\substack{s\sim d_1^{\pi,s_0}\\ a\sim\pi}}\left[f(s)-\gamma f(s')\right]$$

$$+\gamma\,\mathbb{E}_{\substack{s\sim d_2^{\pi,s_0}\\ a\sim\pi}}\left[f(s)-\gamma f(s')\right]$$

$$+\gamma^2\,\mathbb{E}_{\substack{s\sim d_3^{\pi,s_0}\\ a\sim\pi}}\left[f(s)-\gamma f(s')\right]$$

$$+\ \vdots$$

$$= f(s_0).$$

$$\mathbb{E}_{Q}[f] = \mathbb{E}_{Q}[f] \quad \forall f.$$

$$\left|V_{\hat{M}}^{\pi}(s_0) - V_{M}^{\pi}(s_0)\right| = \left|\frac{1}{1-\gamma}\mathbb{E}_{\substack{(s,a)\sim d^{\pi,s_0} \\ r\sim R(\cdot|s,a) \\ s'\sim P(\cdot|s,a)}}\left[V_{\hat{M}}^{\pi}(s) - r - \gamma V_{\hat{M}}^{\pi}(s')\right]\right|$$

$$\leq \frac{1}{1-\gamma}\max_s \left| V_{\hat{M}}^{\pi}(s) - R(s,\pi) - \gamma \mathbb{E}_{s'\sim P(\cdot|s,\pi)}\left[V_{\hat{M}}^{\pi}(s')\right]\right|$$

$$\leq \frac{1}{1-\gamma}\max_s \left| \hat{R}(s,\pi) + \gamma \mathbb{E}_{s'\sim \hat{P}(\cdot|s,a)}\left[V_{\hat{M}}^{\pi}(s')\right] \right.$$

ε_R. $\quad \left. - R(s,\pi) - \gamma \mathbb{E}_{s'\sim P(\cdot|s,a)}\left[V_{\hat{M}}^{\pi}(s')\right]\right|$

$\gamma \varepsilon_P \cdot V_{max}$

$$= \left| \gamma \langle \hat{P}(\cdot|s,a), V_{\hat{M}}^{\pi} \rangle - \gamma \langle P(\cdot|s,a), V_{\hat{M}}^{\pi} \rangle \right|$$

$V_{\hat{M}}^{\pi} \in [0, V_{max}]$.

$$= \gamma \left| \langle \hat{P}(\cdot|s,a) - P(\cdot|s,a), V_{\hat{M}}^{\pi} \rangle \right|.$$

$$\leq \gamma \underbrace{\|\hat{P}(\cdot|s,a) - P(\cdot|s,a)\|_1}_{\leq \varepsilon_P} \cdot \|V_{\hat{M}}^{\pi}\|_\infty.$$

$$\downarrow$$

$$V_{max} = \frac{R_{max}}{1-\gamma}.$$

Hölder's ineq:

$u, v \in \mathbb{R}^n$.

$|u^T v| \leq \|u\| \cdot \|v\|_*$

for any $\|\cdot\|, \|\cdot\|_*$.

Special case:

$$= \frac{1}{1-\gamma}\left(\varepsilon_R + \gamma \varepsilon_p \frac{V_{max}}{2}\right).$$

$\|\cdot\|_p \overset{\text{dual}}{\hookrightarrow} \|\cdot\|_q$ for $1/p + 1/q = 1$.

$$\sqrt{u^T A u} := \|u\|_A.$$

$\|\cdot\|_A$. $\|\cdot\|_{A^{-1}}$.

$$u^T v = u^T A^{1/2} A^{-1/2} v$$

$$\leq \|u^T A^{1/2}\|_2 \cdot \|\cdot\|_2$$

$[0, V_{max}]$.

$$|\langle \hat{P} - P, V \rangle| \leq |\langle \hat{P} - P, V - \frac{V_{max}}{2} \cdot \mathbb{1} \rangle|$$

$$\left[-\frac{V_{max}}{2}, \frac{V_{max}}{2}\right]$$

$$|\hat{R}(s,a) - R(s,a)| \leq \frac{1}{n} \sum_{r \in D_{s,a}} r \cdot \mathbb{1} \wedge R_{max} \sqrt{\frac{1}{2n} \ln \frac{2}{\delta}} \quad \text{w.p.} \geq 1 - \delta$$

$$\|\hat{P}(\cdot|s,a) - P(\cdot|s,a)\|_1$$

$$= \sum_{\tilde{s}} \left| \hat{P}(\tilde{s}|s,a) - P(\tilde{s}|s,a) \right|.$$

$$\downarrow$$

$$\frac{1}{n} \sum_{s' \in D_{s,a}} \mathbb{I}[s' = \tilde{s}].$$

$$\sqrt{|S| \frac{1}{2n} \ln \frac{2^{|S|}}{\delta}}$$

$$\|\hat{P}(\cdot) - P(\cdot)\|_1.$$

$\hat{\Delta}(S)$ $\quad$ $\Delta(S)$.

$$= \max_{u \in \{-1,1\}^S} u^T (\hat{P} - P).$$

$$\|x\|_* = \max_{\|y\| \le 1} y^T x$$

$$u^T \hat{P} = u^T \frac{1}{n} \sum_{s' \in D} e_{s'}.$$

$$= \frac{1}{n} \sum_{s' \in D_{s,a}} u^T e_{s'}. \qquad \mathbb{E}_{s' \sim p}[u(s')]$$

$$= \frac{1}{n} \sum_{s'} u(s'). \qquad = \langle p, u \rangle.$$

$u(s') \in [-1, 1]$.

$u^T \hat{p} - u^T p.$

Fix any u, $\quad |u^T \hat{p} - u^T p| \le 2 \cdot \sqrt{\frac{1}{2n} \ln \frac{2}{\delta}}.$

$$\|\hat{p} - p\|_1 = \max_{u \in \{-1,1\}^S} |u^T \hat{p} - u^T p|$$

$$\le 2 \cdot \sqrt{\frac{1}{2n} \ln \frac{2 \cdot 2^S}{\delta}}.$$

$\|V_M^* - V_M^{\pi_{\hat{M}}^*}\|_\infty.$ $\qquad \pi_{\hat{M}}^* = \pi_{Q_{\hat{M}}^*}.$

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

$$\|V_M^* - V_M^{\pi_f}\|_\infty \le \frac{2 \|f - Q_M^*\|_\infty}{1 - \gamma}.$$

$$\le 2 \cdot \|Q_{\hat{M}}^* - Q_M^*\|_\infty \le \frac{}{1 - \gamma}.$$

$$\| Q^*_{\hat{M}} - Q^*_M \|_\infty \le \frac{1}{1-\gamma} V_{\max} \sqrt{\frac{1}{n} \ln \frac{2|S \times A|}{\delta}}$$

$$= \| Q^*_{\hat{M}} - T_{\hat{M}} Q^*_M + T_{\hat{M}} Q^*_M - Q^*_M \|_\infty .$$

$$\le \underbrace{\| T_{\hat{M}} Q^*_{\hat{M}} - T_{\hat{M}} Q^*_M \|_\infty}_{\le \gamma \cdot \| Q^*_{\hat{M}} - Q^*_M \|_\infty} + \underbrace{\| T_{\hat{M}} Q^*_M - T_M Q^*_M \|_\infty}_{(II)}$$

$$(II)(s,a) = \hat{R}(s,a) + \gamma \langle \hat{P}(\cdot|s,a), V^*_M \rangle - R(s,a) - \gamma \langle P(\cdot|s,a), V^*_M \rangle$$

$$\frac{1}{n} \sum_{(r,s') \in D_{s,a}} r + \gamma \left\langle \frac{1}{n} \sum_{(r,s') \in D_{s,a}} e_{s'}, V^*_M \right\rangle$$

$$= \frac{1}{n} \sum_{(r,s') \in D_{s,a}} \left(r + \gamma \langle e_{s'}, V^*_M \rangle \right)$$

$$= \frac{1}{n} \sum_{(r,s') \in D_{s,a}} \left(r + \gamma V^*_M(s') \right)$$

$$\left| \frac{1}{n} \sum \left(r + \gamma V_m^*(s') \right) - Q_m^*(s,a) \right|$$

$$\mathop{\mathbb{E}}_{r,s'}[\cdot] = Q_m^*(s,a)$$

$$R_{max} + \gamma \cdot \frac{R_{max}}{1-\gamma} = \frac{R_{max}}{1-\gamma} = V_{max}$$

$$\leq V_{max} \cdot \sqrt{\frac{1}{2n} \ln \frac{2|S \times A|}{\delta}}$$