2.3 Minmax Optimal Sample Complexity

We'll show that the model based approach is minmax optimal.

1. The discounted case

Theorem 2.6. For $\delta > 0$ and an approximately chosen absolute constant c, we have that

(i) Value estimation: with prob. $\geq 1-\delta$

$$\|Q^* - \hat{Q}^*\|_\infty \leq \gamma \sqrt{\frac{c}{(1-\gamma)^3} \frac{\log(c|S||A|/\delta)}{N}} + \frac{c\gamma}{(1-\gamma)^3} \frac{\log(c|S||A|/\delta)}{N}$$

(ii) Sub-optimality: If $N \geq \frac{1}{(1-\gamma)^2}$, with prob. $\geq 1-\delta$

$$\|Q^* - Q^{\hat{\pi}}\|_\infty \leq \gamma \sqrt{\frac{c}{(1-\gamma)^3} \frac{\log(c|S||A|/\delta)}{N}}$$

Corollary 2.7 Provided that $\varepsilon \leq 1$ and that

$$\begin{aligned}\# \text{samples from generative model} &= |S||A|N \\ &\geq \frac{c|S||A|}{(1-\gamma)^3} \frac{\log(c|S||A|/\delta)}{\varepsilon^2}\end{aligned}$$

then with prob. $\geq 1-\delta$,

$$\|Q^* - \hat{Q}^*\|_\infty \leq \varepsilon.$$

Furthermore, provided $\varepsilon \leq \sqrt{\frac{1}{1-\gamma}}$ and that

$$\# \text{samples from model} = |S||A||N| \geq \frac{c|S||A|}{(1-\gamma)^3} \frac{\log(c|S||A|/\delta)}{\varepsilon^2}$$

then with prob. $\geq 1-\delta$,

$$\|Q^* - Q^{\hat{\pi}}\|_\infty \leq \varepsilon$$

Lower Bounds: an algorithm A is $(\varepsilon-\delta)$-good if

$$\|Q^* - Q^{\hat{\pi}}\|_\infty \leq \varepsilon \quad \text{with prob.} \geq 1-\delta.$$

Theorem For $\varepsilon < \sqrt{1/(1-\gamma)}$, provided $N \geq \frac{c}{(1-\gamma)^3} \frac{\log(cSA/\delta)}{\varepsilon^2}$ then with prob. $\geq 1-\delta$, $\|Q^* - Q^{\hat{\pi}}\|_\infty \leq \varepsilon$

Pf: From Lemma 2.5

$$Q^* - \hat{Q}^* \leq \gamma \|(I - \gamma \hat{P}^{\pi^*})^{-1}(P - \hat{P})V^*\|_\infty$$

From Bernstein's ineq., with prob. $\geq 1-\delta$. we have

$$|(P-\hat{P})V^*| \leq \sqrt{\frac{2\log(2SA/\delta)}{N}} \sqrt{Var_P(V^*)} + \frac{1}{1-\gamma} \frac{2\log(2SA/\delta)}{3N} \vec{1}$$

$$\Rightarrow Q^* - \hat{Q}^* \leq \gamma \sqrt{\frac{2\log(2SA/\delta)}{N}} \|(I - \gamma \hat{P}^{\pi^*})^{-1} \sqrt{Var_P(V^*)}\|_\infty$$

+ lower order term

$$Var_P(V)(s,a) := Var_{P(s,a)}(V)$$

$$Var_P(V) := P(V)^2 - (PV)^2$$

Define $\Sigma_M^\pi(s,a) := E\left[\left(\sum_{t=0}^{\infty} \gamma^t r(s_t,a_t) - Q_M^\pi(s,a)\right)^2 \middle| s_0=s, a_0=a\right]$

Bellman equation for Σ_M^π (total variance of discounted reward)

$$\Sigma_M^\pi = \gamma^2 Var_P(V_M^\pi) + \gamma^2 P^\pi \Sigma_M^\pi$$

Lemma: For any π, M

$$\|(I - \gamma P^\pi)^{-1} \sqrt{Var_P(V_M^\pi)}\|_\infty \leq \sqrt{\frac{2}{(1-\gamma)^3}}$$

$$\|(I - \gamma \hat{P}^{\pi^*})^{-1} \sqrt{Var_P(V^*)}\|_\infty = \|(I - \gamma P_{\hat{M}}^{\pi^*})^{-1} \sqrt{Var_P(V_M^{\pi^*})}\|_\infty$$

need to quantify $\sqrt{Var_p(V_M^{\pi^*})} \approx \sqrt{Var_p(V_{\hat{M}}^{\pi^*})}$

$$\leq \| (I - \gamma P_{\hat{M}}^{\pi^*})^{-1} \sqrt{Var_p(V_{\hat{M}}^{\pi^*})} \|_\infty + \text{"lower order"}$$

$$\leq \sqrt{\frac{2}{(1-\gamma)^3}} + \text{"lower order"}$$

$$\Rightarrow \|Q^* - \hat{Q}^*\|_\infty \leq \gamma \sqrt{\frac{2}{(1-\gamma)^3} \frac{2\log(SA/\delta)}{N}}$$