Risk Bounds and Rademacher Complexity in Batch Reinforcement Learning

Yaqi Duan, Chi Jin, Zhiyuan Li

Princeton University

ICML, 2021

Batch Reinforcement Learning (RL)

- Episodic Markov decision process $MDP(\mathcal{S}, \mathcal{A}, H, \mathbb{P}, r)$.

Dataset
$\mathcal{D} = \{(s, a, r, s', h)\}$

→

Optimal policy $\hat{\pi}$

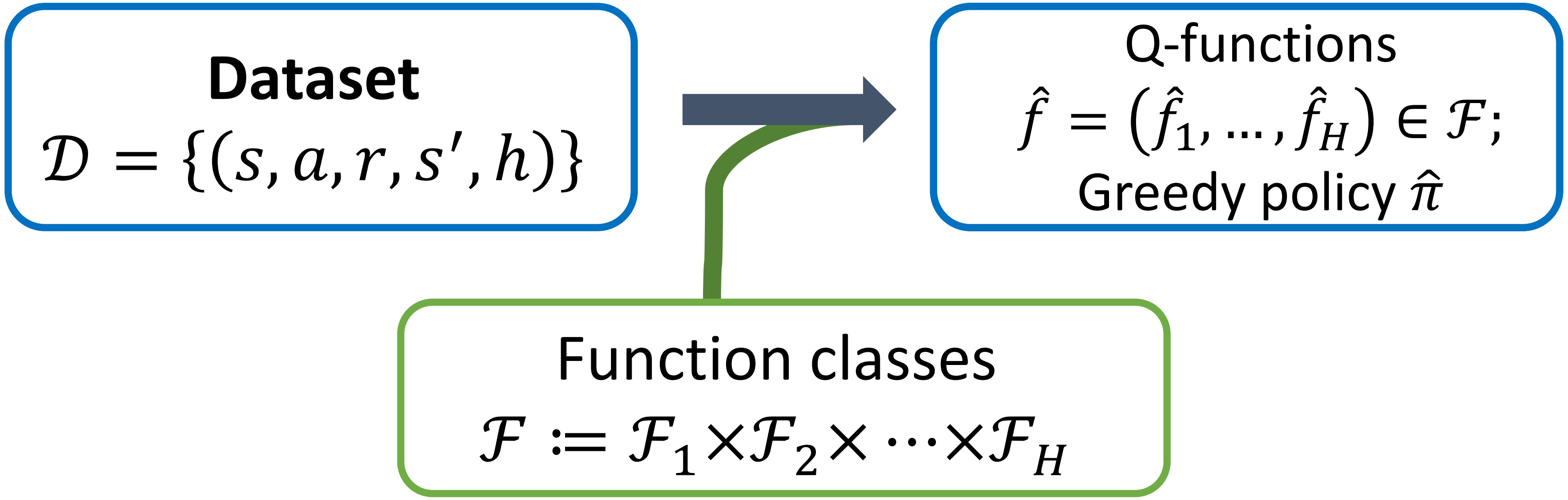
At the h^{th} step, collect
n i.i.d. samples with $(s, a) \sim \mu_h$.

Aim to maximize
$V_1^{\pi}(s_1) := \mathbb{E}\left[\sum_{h=1}^{H} r_h \,\middle|\, s_1, \pi\right]$.

★ **Goal:** Control the value suboptimality $V_1^{\star}(s_1) - V_1^{\hat{\pi}}(s_1)$.

Value-Based Method

1. Approximate optimal Q-function $Q_h^\star$ by $\hat{f}_h \in \mathcal{F}_h$.
2. Output the greedy policy $\hat{\pi}$ associated with $\hat{f}$.



★ **Goal:** $V_1^\star(s_1) - V_1^{\hat{\pi}}(s_1) \lesssim$ complexity of $\mathcal{F}$.

Bellman Error

★ **Goal:** $V_1^\star(s_1) - V_1^{\hat{\pi}}(s_1) \lesssim$ complexity of $\mathcal{F}$.

Bellman error of $f = (f_1, f_2, \dots, f_H)$:

$$\mathcal{E}(f) \coloneqq \frac{1}{H} \sum_{h=1}^{H} \| f_h - \mathcal{T}_h^\star f_{h+1} \|_{\mu_h}^2.$$

optimal Bellman operator ($\mathcal{T}_h^\star$)

data distribution (μ_h)

Bellman Error

★ **Goal:** $V_1^\star(s_1) - V_1^{\hat{\pi}}(s_1) \lesssim$ complexity of $\mathcal{F}$.

Bellman error of $f = (f_1, f_2, \dots, f_H)$:

$$\mathcal{E}(f) := \frac{1}{H} \sum_{h=1}^{H} \| f_h - \mathcal{T}_h^\star f_{h+1} \|_{\mu_h}^2.$$

Key reduction: $V_1^\star(s_1) - V_1^{\hat{\pi}}(s_1) \lesssim \sqrt{\mathcal{E}(\hat{f})}.$

Framework of Analysis

Value suboptimality
$V_1^\star(s_1) - V_1^{\widehat{\pi}}(s_1)$

↓

Bellman error $\mathcal{E}(f)$

↓

(Local) Rademacher complexity
of $\mathcal{F} := \mathcal{F}_1 \times \mathcal{F}_2 \times \cdots \times \mathcal{F}_H$

↓

Concrete bounds for specific $\mathcal{F}$

finite classes,
linear spaces,
kernel spaces,
sparse linear features,
etc.

Recall ★ **Goal:** $\mathcal{E}(\hat{f}) \lesssim$ complexity of $\mathcal{F}$.

Minimax lower bound

curse of dim!

$$\inf_{\text{alg.}} \sup_{\substack{\text{MDP with } \boldsymbol{S} \text{ states,} \\ \text{function class } \mathcal{F}}} \mathbb{E}\mathcal{E}(\hat{f}) - \min_{f \in \mathcal{F}} \mathcal{E}(f) = \Omega\left(\min\left\{1, \frac{\sqrt{S}}{n}\right\}\right).$$

☹

Completeness assumptions help get rid of this.

Remedy: Completeness Assumption

- Empirical loss

$$L(f) := \frac{1}{nH} \sum_{(s,a,r,s',h)\in\mathcal{D}} \left(f_h(s,a) - r - \max_{a'} f_{h+1}(s',a')\right)^2 .$$

$$\mathcal{E}(f) = \mathbb{E}\, L(f) - \boxed{\text{variance}} .$$ ← **need to learn $\mathcal{T}^\star f$**

- Define **ϵ-completeness**: approximate $\mathcal{T}^\star f$ using $\mathcal{F}$,

$$\sup_{f\in\mathcal{F}_{h+1}} \inf_{g\in\mathcal{F}_h} \|g - \mathcal{T}^\star f\|_{\mu_h}^2 \le \epsilon.$$

Upper Bounds with Rademacher Complexities

- When ϵ-complete:

$$\mathcal{E}(\hat{f}) \leq \min_{f \in \mathcal{F}} \mathcal{E}(f) + \epsilon + \text{Rademacher complexity.}$$

- Acceleration by **localization** (from $n^{-\frac{1}{2}}$ to n^{-1}):

$$\mathcal{E}(\hat{f}) \leq \min_{f \in \mathcal{F}} \mathcal{E}(f) + \epsilon + \text{critical radius of } \textbf{local} \text{ Rad. comp.}$$

Recap: Framework of Analysis

Value suboptimality
$V_1^\star(s_1) - V_1^{\hat{\pi}}(s_1)$

↓

Bellman error $\mathcal{E}(f)$

↓

(Local) Rademacher complexity
of $\mathcal{F} := \mathcal{F}_1 \times \mathcal{F}_2 \times \cdots \times \mathcal{F}_H$

↓

Concrete bounds for specific $\mathcal{F}$

finite classes,
linear spaces,
kernel spaces,
sparse linear features,
etc.

Thanks!