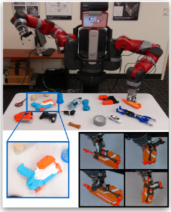
Assouad, Fano, and Le Cam with Interaction: A Unifying Lower Bound Framework and Characterization for Bandit Learnability

Fan Chen, Dylan J. Foster, Yanjun Han, Jian Qian, Alexander Rakhlin, Yunbei Xu

https://arxiv.org/abs/2410.05117



Examples: robotics, games, clinical trails, reinforcement learning, etc.

Goal of this work

Understand the fundamental limits for statistical estimation and interactive decision making problems.

Interaction protocol

For each round $t = 1, ..., T$:

- The learner selects a decision $\pi^t \in \Pi$, where Π is the decision space.
- The learner receives an observation $o^t \in \mathcal{O}$ sampled via $o^t \sim M^\star(\pi^t)$, where $\mathcal{O}$ is the observation space.

After T rounds, the learner outputs $\widehat{\pi} \in \Pi$ and incurs a loss $L(M^\star, \widehat{\pi})$.

- $M^\star : \Pi \to \Delta(\mathcal{O})$ is the *model* of the environment
- Learner is given a model class $\mathcal{M} \subseteq (\Pi \to \Delta(\mathcal{O}))$ containing $M^\star$
- Example: structured bandits/contextual bandits, episodic RL, etc.

Interaction protocol

For each round $t = 1, ..., T$:

- The learner selects a decision $\pi^t \in \Pi$, where Π is the decision space.
- The learner receives an observation $o^t \in \mathcal{O}$ sampled via $o^t \sim M^\star(\pi^t)$, where $\mathcal{O}$ is the observation space.

After T rounds, the learner outputs $\widehat{\pi} \in \Pi$ and incurs a loss $L(M^\star, \widehat{\pi})$.

- $M^\star : \Pi \to \Delta(\mathcal{O})$ is the *model* of the environment
- Learner is given a model class $\mathcal{M} \subseteq (\Pi \to \Delta(\mathcal{O}))$ containing $M^\star$
- Example: structured bandits/contextual bandits, episodic RL, etc.
- Special case: statistical estimation
  - Non-interactive: $o^1, \cdots, o^T \sim M^\star$ independently

Minimax Risk

Minimax criterion

The T-round minimax risk is defined as

$$\min_{\texttt{ALG}} \max_{M \in \mathcal{M}} \mathbb{E}^{M,\texttt{ALG}} L(M, \widehat{\pi})$$

- $\min$ over all possible T-round algorithm $\texttt{ALG}$ with output $\widehat{\pi}$
- $\max$ over the *worst-case* model $M \in \mathcal{M}$

Minimax criterion

The T-round minimax risk is defined as

$$\min_{\mathtt{ALG}} \max_{M \in \mathcal{M}} \mathbb{E}^{M,\mathtt{ALG}} L(M, \widehat{\pi})$$

- $\min$ over all possible T-round algorithm $\mathtt{ALG}$ with output $\widehat{\pi}$
- $\max$ over the *worst-case* model $M \in \mathcal{M}$

Statistical estimation (non-interactive):

- Standard and well-understood in statistics
- Proving upper bound: choosing a particular algorithm
- Proving lower bound: requires specialized techniques
  - Le Cam's two-point method
  - Assouad's lemma
  - Fano's inequality

Minimax criterion

The T-round minimax risk is defined as

$$\min_{\texttt{ALG}} \max_{M \in \mathcal{M}} \mathbb{E}^{M,\texttt{ALG}} L(M, \widehat{\pi})$$

- $\min$ over all possible T-round algorithm $\texttt{ALG}$ with output $\widehat{\pi}$
- $\max$ over the *worst-case* model $M \in \mathcal{M}$

Beyond statistical estimation:

- Upper & lower bounds: case-by-case
- Foster et al. [2021] proposes Decision-Estimation Coefficient (DEC) framework, providing both lower and upper bounds for *any* DMSO problem
  - DEC approach is seemingly different from the classical techniques
  - The DEC lower & upper bounds have a gap related to the complexity of *estimation* [Foster et al., 2021, 2023]

Contributions

- A unifying framework for information-theoretic lower bound in statistical estimation and interactive decision making, which recovers
  - Le Cam's two-point method, Assouad's lemma, Fano's inequality
  - The DEC lower bound approach
- A novel complexity measure, the *Fractional Covering Number*
  - A new lower bound for interactive decision making (and complements the DEC lower bound)
  - A unified characterization of learnability for *any* structured stochastic bandit problem
  - Polynomially matching lower and upper bounds for any convex model class

Fractional covering number

$$\mathsf{N}_{\mathrm{frac}}(\mathcal{M}, \Delta) := \inf_{p \in \Delta(\Pi)} \sup_{M \in \mathcal{M}} \frac{1}{p(\pi : L(M, \pi) \leq \Delta)}.$$

- Measuring the best possible coverage over Δ-optimal decisions

Fractional covering number

Fractional covering number

$$\mathsf{N}_{\mathrm{frac}}(\mathcal{M}, \Delta) := \inf_{p \in \Delta(\Pi)} \sup_{M \in \mathcal{M}} \frac{1}{p(\pi : L(M, \pi) \leq \Delta)}.$$

- Measuring the best possible coverage over Δ-optimal decisions
- Dual form of the *fractional cover*

Theorem (Fractional covering number lower bound; Informal)

For a T-round algorithm to achieve risk $\leq \Delta$, it is necessary that

$$T \geq \Omega(\log \mathsf{N}_{\mathrm{frac}}(\mathcal{M}, \Delta) / C_{\mathrm{KL}}),$$

where C_{KL} is the radius of the model class $\mathcal{M}$ under KL divergence.

- Complementary to the DEC lower bound

Fractional covering number

$$\mathsf{N}_{\mathrm{frac}}(\mathcal{M}, \Delta) := \inf_{p \in \Delta(\Pi)} \sup_{M \in \mathcal{M}} \frac{1}{p(\pi : L(M, \pi) \le \Delta)}.$$

- **Application 1:** bandit learnability (and beyond)
- Observation: fractional covering number also provides an upper bound!
- There is a brute-force algorithm that returns a 2Δ-optimal decision using $T \le \widetilde{O}\left(\frac{\mathsf{N}_{\mathrm{frac}}(\mathcal{M},\Delta)}{\Delta^2}\right)$ rounds

Theorem (Bandit learnability)

A class $\mathcal{M}$ of stochastic bandits (with Gaussian rewards) is learnable with finite T if and only if $\mathsf{N}_{\mathrm{frac}}(\mathcal{M}, \Delta) < +\infty$ for all $\Delta > 0$.

Fractional covering number

$$\mathsf{N}_{\mathrm{frac}}(\mathcal{M}, \Delta) := \inf_{p \in \Delta(\Pi)} \sup_{M \in \mathcal{M}} \frac{1}{p(\pi : L(M, \pi) \leq \Delta)}.$$

- **Application 1:** bandit learnability (and beyond)
- **Application 2:** tighter upper bound for convex class
- $\Rightarrow$ Polynomially matching lower and upper bounds for convex model class

Thanks!

References I

Dylan J Foster, Sham M Kakade, Jian Qian, and Alexander Rakhlin. The statistical complexity of interactive decision making. *arXiv preprint arXiv:2112.13487*, 2021.

Dylan J Foster, Noah Golowich, and Yanjun Han. Tight guarantees for interactive decision making with the decision-estimation coefficient. In *The Thirty Sixth Annual Conference on Learning Theory*, pages 3969–4043. PMLR, 2023.