ECE 598 - Fall 2025

Lecture 9: Minimax Lower Bounds IV

25th September, 2025

Instructor: Shubhanshu Shekhar

In the previous lecture, we introduced a method for obtaining minimax lower bounds proceeds by reducing estimation to multiple binary hypothesis tests. We will now discuss how this idea is convenient for studying the minimax regret in interactive decision-making problems.

1 Interactive Decision-Making Problems

Consider a collection of distributions $\{P_{\theta,a} : \theta \in \Theta, \text{ and } a \in \mathcal{A}\}$ indexed by a parameter space Θ and an "action space" $\mathcal{A}$. We take the perspective of an agent whose goal is to interactively select actions $A_1, A_2, \ldots, A_n$ according to a policy $\pi \equiv (\pi_1, \ldots, \pi_n)$, where $\pi_t(\cdot \mid H_{t-1}) = P_{A_t|H_{t-1}}$ is a stochastic kernel or conditional probability distribution of the t^{th} action A_t given the history $H_{t-1} = (A_1, X_1, \ldots, A_{t-1}, X_{t-1})$. In each round, having selected the action A_t, the agent observes $X_t \sim P_{\theta,A_t}$. Finally, at the end of n rounds of interaction, the agent may optionally make a terminal decision according to the rule $\rho \equiv P_{W|H_n}$. In many cases, the terminal decision also lies in the same action space $\mathcal{A}$, and we will work under this simplifying assumption for the rest of this lecture. This interaction protocol is summarized in Algorithm 1.

for $t = 1$ **to** n **do**
  agent samples $A_t \sim \pi_t(\cdot \mid H_{t-1})$
  agent observe $X_t \sim P_{\theta,A_t}$
  update history $H_t \leftarrow (H_{t-1}, A_t, X_t)$
(Optional) decision $A_{n+1} \sim \rho(\cdot \mid H_n)$

Algorithm 1: Interaction protocol (stateless)

We will focus mainly on two types of problems:

Regret minimization. In each round, the agent incurs a loss according to some loss function $\ell : \mathcal{X} \times \mathcal{A} \to \mathbb{R}$. The expected loss (or risk) associated with an action $a \in \mathcal{A}$ and parameter $\theta \in \Theta$, and an optimal action associated with $\theta \in \Theta$ are defined as

$$L(\theta, a) = \mathbb{E}_{X \sim P_{\theta,a}}[\ell(X, a)], \quad a^*(\theta) \in \operatorname*{argmin}_{a \in \mathcal{A}} L(\theta, a).$$

Finally, this can be used to define the regret associated with a policy π as

$$\mathrm{Reg}_n(\pi, \theta) = \mathbb{E}_{(\theta,\pi)}\left[\sum_{t=1}^{n} L(\theta, A_t) - L(\theta, a^*(\theta))\right] = \mathbb{E}_{(\theta,\pi)}\left[\sum_{t=1}^{n} \ell(X_t, A_t) - \ell(X_t, a^*(\theta))\right].$$

In the regret minimization problem, there is no terminal action ρ. The goal is to simply study the minimax regret

$$\mathrm{Reg}_n(\mathcal{A}, \Theta) = \inf_{\pi} \sup_{\theta \in \Theta} \mathrm{Reg}_n(\pi, \theta).$$

Pure Exploration. In pure exploration, after the completion of n rounds, the agent outputs $A_{n+1} \sim \rho(\cdot \mid H_n)$, and incurs a pure-exploration risk of

$$r_n(\pi, \rho, \theta) = \mathbb{E}_{(\theta,\pi,\rho)} \left[L(\theta, A_{n+1}) - L(\theta, a^*(\theta)) \right].$$

This can be used to then define the minimax pure-exploration risk

$$r_n(\pi, \rho, \Theta) = \inf_{\pi,\rho} \sup_{\theta \in \Theta} r_n(\pi, \rho, \theta).$$

We now make a simple observation that will allow us to relate the minimax lower bounds on pure-exploration risk to that of regret. In particular, from any given sampling policy π, we can obtain a terminal decision rule ρ_π as

$$\rho_\pi(\cdot \mid H_n) = \frac{1}{n} \sum_{t=1}^{n} \pi_t(\cdot \mid H_{t-1}).$$

This terminal decision rule is simply the average of all the sampling distributions. The pure-exploration risk associated with this rule then, is equal to

$$\begin{aligned} r_n(\pi, \rho_\pi, \theta) &= \mathbb{E}_{(\theta,\pi,\rho_\pi)} \left[L(\theta, A_{n+1}) - L(\theta, a^*(\theta)) \right] = \frac{1}{n} \sum_{t=1}^{n} \mathbb{E}_{(\theta,\pi)} \left[L(\theta, A_t) - L(\theta, a^*(\theta)) \right] \\ &= \frac{1}{n} \mathrm{Reg}_n(\pi, \theta). \end{aligned} \tag{1}$$

This fact leads to the following simple observation.

Proposition 1.1. *The following is true:*

$$\inf_{\pi,\rho} \sup_{\theta} r_n(\pi, \rho, \theta) \geq \psi_n \quad \Longrightarrow \quad \inf_{\pi} \sup_{\theta} \mathrm{Reg}_n(\pi, \theta) \geq n\psi_n.$$

In other words, establishing a minimax lower bound on the pure-exploration risk also gives us a lower bound on the minimax regret.

Proof. The proof of this result is immediate from our earlier discussion:

$$\inf_{\pi,\rho} \sup_{\theta} r_n(\pi, \rho, \theta) \leq \inf_{\pi,\rho_\pi} \sup_{\theta} r_n(\pi, \rho_\pi, \theta) = \inf_{\pi} \sup_{\theta} \frac{1}{n} \mathrm{Reg}_n(\pi, \theta).$$

The inequality is due to the restriction of ρ to ρ_π, and the equality follows from (1). □

2 Application to Linear Bandits

Our first example is the problem of linear bandits. In this case, we have $\Theta = \mathcal{A} = \{x \in \mathbb{R}^d : \|x\|_2 \leq 1\}$, and assume that $P_{\theta,a} = N(\langle \theta, a \rangle, \sigma^2)$. This is equivalent to the observation model

$$X_t = \langle \theta, A_t \rangle + \sigma \varepsilon_t, \quad \text{where} \quad (\varepsilon_t)_{t=1}^n \overset{\text{i.i.d.}}{\sim} N(0,1).$$

The risk and optimal action are defined as

$$L(\theta, a) = \mathbb{E}[-\langle a, X \rangle] = -\langle a, \theta \rangle, \quad \text{which means} \quad a^*(\theta) = \operatorname*{argmax}_{a:\|a\|_2 \leq 1} \langle a, \theta \rangle = \theta / \|\theta\|.$$

The worst-case risk associated with a policy π and decision rule ρ in linear bandits is defined as

$$r_n(\pi, \rho, \Theta) = \sup_{\theta \in \Theta} \mathbb{E}_{(\theta,\pi,\rho)} \left[-\langle A_{n+1}, \theta \rangle + \|\theta\| \right],$$

where we used the fact that $a^*(\theta) = \theta/\|\theta\|$. We will obtain a lower bound on this using Assouad's method.

Indexed parameter class. The first step is to identify a class of problem instances (parameters) indexed by the Hamming cube. We do this by fixing a $\delta \in (0,1)$ whose exact value will be specified later, and then defining

$$\theta_v = \frac{\delta}{\sqrt{d}} v, \quad \text{for all} \quad v \in \mathcal{H}_d = \{-1, +1\}^d.$$

The $\sqrt{d}$ term in the denominator ensures that $\|\theta_v\| \leq \delta \leq 1$. As in the previous lecture, we will use $v^{(j+)}$, $v^{(j-)}$ to denote elements of $\mathcal{H}_d$ that agree with v on all coordinates $i \neq j$, and have $+1$ (resp. -1) in the j^{th} coordinate.

Class of Joint Distributions. Let us fix a pair of (π, ρ), and denote the set of joint distributions $Q_\theta \equiv Q_{\theta,\pi,\rho}$ such that for any (h_n, a_{n+1}), we have

$$Q_\theta(h_n, a_{n+1}) = \pi_1(a_1) P_{\theta,a_1}(x_1) \pi_2(a_2 \mid h_1) P_{\theta,a_2}(x_2) \ldots \pi_n(a_n \mid h_{n-1}) P_{\theta,a_n}(x_n) \rho(a_{n+1} \mid h_n).$$

For any $v \in \mathcal{H}_d$, we will use the notation Q_v to denote $Q_{\theta_v} \equiv Q_{\theta_v,\pi,\rho}$.

Hamming Separation Condition. To apply Assouad's method, we need to verify that the problem instances satisfy the Hamming separation condition. In particular, let $E : \mathcal{A} \to \mathcal{H}_d$ denote the "sign encoding map" satisfying $E(\theta_v) = v$ for all $v \in \mathcal{H}_d$, and observe that

$$L(\theta_v, a) - L(\theta_v, a^*(\theta_v)) = \langle \theta_v, a^*(\theta_v) - a \rangle \geq \sum_{j=1}^d$$

To see why the inequality holds, note that

$$\langle \theta_v, a^*(\theta_v) - a \rangle = \sum_{j=1}^d \theta_v[j] \left(a^*(\theta_v)[j] - a[j] \right) = \sum_{j=1}^d \frac{\delta v[j]}{\sqrt{d}} \left(\frac{v[j]}{\sqrt{d}} - a[j] \right).$$

In the second equality, we used the fact that $a^*(\theta_v) = v/\sqrt{d}$. This can be rewritten as

$$\langle \theta_v, a^*(\theta_v) - a \rangle = \delta - \delta \frac{\langle v, a \rangle}{\sqrt{d}} = \delta \left(1 - \frac{\langle v, a \rangle}{\sqrt{d}}\right).$$

Now, let m denote the Hamming distance between $E(a)$ and v; that is, m is the number of coordinates in which the sign of a and v differ. Then, we can bound $\langle v, a \rangle$ with

$$\langle v, a \rangle \leq \sup_{a' \in \mathcal{A}: d_H(E(a'), v) = m} \langle v, a' \rangle = \sqrt{d - m}.$$

The supremum is achieved by an a' that places 0 (or arbitrarily small) mass on the coordinates at which signs of v and a differ, and equal mass on the remaining points. Hence, we have obtained

$$\langle \theta_v, a^*(\theta_v) - a \rangle = \delta \left(1 - \frac{\langle v, a \rangle}{\sqrt{d}}\right) \geq \delta \left(1 - \sqrt{1 - \frac{m}{d}}\right), \quad \text{where} \quad m = d_H(E(a), \theta_v).$$

Finally, we use the fact that with $x = m/d \in [0, 1]$, we have

$$\left(1 - \frac{x}{2}\right)^2 = 1 - x + \frac{x^2}{4} \geq 1 - x \quad \implies \quad 1 - \frac{x}{2} \geq \sqrt{1 - x} \quad \implies \quad 1 - \sqrt{1 - x} \geq \frac{x}{2}.$$

This gives us the required Hamming separation condition

$$\langle \theta_v, a^*(\theta_v) - a \rangle \geq \frac{\delta}{2d} d_H(v, E(a)).$$

Applying Assouad's Lemma. We can now apply the version of Assouad's Lemma along with Pinsker's to conclude that

$$\begin{aligned}
\sup_{\theta} r_n(\pi, \rho, \theta) &\geq \frac{\delta}{4d} \sum_{j=1}^{d} (1 - TV(Q_{j+}, Q_{j-})) = \frac{\delta}{4} \left(1 - \frac{1}{d} \sum_{j=1}^{d} TV(Q_{j+}, Q_{j-})\right) \\
&\geq \frac{\delta}{4} \left(1 - \frac{1}{\sqrt{d}} \times \left(\sum_{j=1}^{d} TV(Q_{j+}, Q_{j-})^2\right)^{1/2}\right) \\
&\geq \frac{\delta}{4} \left(1 - \frac{1}{\sqrt{d}} \times \left(\frac{1}{2} \sum_{j=1}^{d} D_{\mathrm{KL}}(Q_{j+}, Q_{j-})\right)^{1/2}\right) \qquad (2)
\end{aligned}$$

Now, the standard decomposition of the relative entropy under adaptive sampling tells us

$$D_{\mathrm{KL}}(Q_{v^{(j+)}} \parallel Q_{v^{(j-)}}) = \sum_{t=1}^{n} \mathbb{E}_{Q_{v^{(j+)}}} \left[\frac{(\langle A_t, \theta_{v^{(j+)}} - \theta_{v^{(j-)}} \rangle)^2}{2\sigma^2}\right] = \frac{2\delta^2}{d\sigma^2} \sum_{t=1}^{n} \mathbb{E}_{Q_{v^{(j+)}}} \left[A_t[j]^2\right]$$

On summing over $j \in [d]$, and using the fact that $D_{\mathrm{KL}}(Q_{j+} \parallel Q_{j-}) \leq \max_v D_{\mathrm{KL}}(Q_{v^{(j+)}} \parallel Q_{v^{(j-)}})$, we get the bound

$$\frac{1}{2} \sum_{j=1}^{d} D_{\mathrm{KL}}(Q_{j+}, Q_{j-}) \leq \frac{2n\delta^2}{d\sigma^2}.$$

Plugging this into (2), we get

$$\sup_{\theta} r_n(\pi, \rho, \theta) \geq \frac{\delta}{4}\left(1 - \frac{\delta\sqrt{n}}{\sigma d}\right).$$

By selecting $\delta = \sigma d/2\sqrt{n}$, we get a lower bound on the pure-exploration risk

$$\sup_{\theta \in \Theta} r_n(\pi, \rho, \theta) \geq \frac{\sigma d}{16\sqrt{n}}.$$

This completes the proof of the lower bound on the pure-exploration risk. By Proposition 1.1 this also implies the bound on the regret

$$\inf_{\pi} \sup_{\theta} \mathrm{Reg}_n(\pi, \theta) \geq \frac{\sigma d\sqrt{n}}{16}.$$