CS861: Theoretical Foundations of Machine Learning Lecture 1 - 10/30/2023

University of Wisconsin–Madison, Fall 2023

Lecture 24: EXP3, Lower Bounds for adversarial bandits

Lecturer: Kirthevasan Kandasamy *Scribed by: Bo-Hsun Chen, Zexuan Sun*

In this lecture, we will continue `the proof of EXP-3 Theorem` from the previous lecture, then discuss `lower bounds for adversarial bandits`, and finally introduce and define `contextual bandit problem`.

Proof (proof of EXP-3 Theorem)

$$\text{Recall the lemma} \begin{cases} \mathbb{E}[\hat{\ell}_t(a) \mid p_t] = \ell_t(a) \\ \\ [\hat{\ell}_t^2(a) \mid p_t] = \frac{\ell_t^2(a)}{p_t(a)} \end{cases}$$

We will apply the first result from the Hedge theorem with $a = a^*$ and losses ℓ_t. Since $p_t^T \ell_t = \ell_t(A_t) \leq 1$ and the losses are non-negative, we can apply this result. We have,

$$\sum_{t=1}^{T} p_t^T \hat{\ell}_t - \sum_{t=1}^{T} \hat{\ell}_t(a^*) \leq \frac{\log(K)}{\eta} + \eta \sum_{t=1}^{T} p_t^T \hat{\ell}_t{}^2, \tag{1}$$

where $a^* = \arg\min_{a \in [K]} \sum_{t=1}^{T} \ell_t(a)$.

First note that $\mathbb{E}[\hat{\ell}_t(a^*) \mid p_t] = \ell_t(a^*)$ by the lemma above. Next, applying the Lemma again,

$$\begin{aligned} \mathbb{E}[p_t^T \hat{\ell}_t \mid p_t] &= p_t^T \mathbb{E}[\hat{\ell}_t \mid p_t] \\ &= p_t^T \ell_t \text{ (by Lemma (i))} \\ &= \mathbb{E}[\ell_t(A_t) \mid p_t] \end{aligned}$$

Finally, applying the second result of the above lemma,

$$\begin{aligned} \mathbb{E}[p_t^T \hat{\ell}_t^2 \mid p_t] &= \mathbb{E}[\sum_{a=1}^{K} p_t(a) \hat{\ell}_t^2(a) \mid p_t] \\ &= \sum_{a=1}^{K} \Big(p_t(a) \mathbb{E}[\hat{\ell}_t^2(a) \mid p_t] \Big) \\ &= \sum_{a=1}^{K} \ell_t^2(a) \leq K \end{aligned}$$

Now, taking expectations on both sides of (1), we have

$$\begin{aligned}
\mathbb{E}[LHS] &= \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{E}[p_t^T \hat{\ell}_t \mid p_t] - \sum_{t=1}^{T} \mathbb{E}[\hat{\ell}_t(a^*) \mid p_t]\right] \\
&= \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{E}[\ell_t(A_t) \mid p_t] - \sum_{t=1}^{T} \ell_t(a^*)\right] \\
&= \mathbb{E}\left[\sum_{t=1}^{T} \ell_t(A_t)\right] - \sum_{t=1}^{T} \ell_t(a^*) \\
&= R_T(\pi, \underline{\ell}), \\
\mathbb{E}[RHS] &= \frac{\log(K)}{\eta} + \eta \sum_{t=1}^{T} \mathbb{E}\left[\mathbb{E}[p_t^T \hat{\ell}_t{}^2 \mid p_t]\right] \\
&\leq \frac{\log(K)}{\eta} + \eta K T \text{ (since } \mathbb{E}[p_t^T \hat{\ell}_t{}^2 \mid p_t] \leq K)
\end{aligned}$$

This proves the first statement of the EXP-3 Theorem, and the second statement follows by optimizing over η. □

1 Lower bounds for adversarial bandits

The following theorem provides a lower bound for the minimax rate of regret of the adversarial multi-armed bandit problem.

Theorem 1. *For the adversarial multi-armed bandit problem, the minimax regret satisfies,*

$$\inf_{\pi} \sup_{\underline{\ell} \in [0,1]^{K \times T}} R_T(\pi, \underline{\ell}) \in \Omega(\sqrt{KT})$$

Remark Recall the minimax lower bound for stochastic bandits is

$$\inf_{\pi} \sup_{\nu \in \mathcal{P}} R_T^{stoch}(\pi, \nu) \in \Omega(\sqrt{KT})$$

Note that the adversarial bandit problem is is applicable in more general settings than the stochastic bandit problem. Moreover, the regret definitions are different for the adversarial bandit and stochastic bandit problems. For the adversarial bandits, the regret depends on the best action in hindsight. While, the regret of stochastic bandits depends on the arm with the lowest expected mean value. Despite this, we find that the minimax regret is similar for both problems. This is because the hardest stochastic bandit problems are as hard as the hardest adversarial bandit problems. In fact, the proof of this lower bound will rely on similar techniques to the proof of the lower bound for stochastic bandits.

Our proof will consider stochastic losses and show that the expected regret is large. Then, there is at least one sequence of losses for which the regret should be large.

Proof Let π be given. We will consider two stochastic loss models $\nu^{(1)} = (\nu_1^{(1)}, \nu_2^{(1)}, ..., \nu_K^{(1)})$ and $\nu^{(2)} = (\nu_1^{(2)}, \nu_2^{(2)}, ..., \nu_K^{(2)})$ to be defined shortly. Let $P^{(1)}$ and $P^{(2)}$ denote the probability distributions of the action loss sequence $(A_1, \ell_1(A_1), ..., A_T, \ell_1(A_T))$ due to π's interaction with $\nu^{(1)}$ and $\nu^{(2)}$ respectively. Let $\mathbb{E}^{(1)}$ and $\mathbb{E}^{(2)}$ denote the corresponding expectations.

Let $\nu^{(1)}$ be defined as,

$$\nu_1^{(1)} = \text{Bern}\left(\frac{1}{2} - \delta\right) \text{ and } \nu_i^{(1)} = \text{Bern}\left(\frac{1}{2}\right), \forall \text{i} \neq 1,$$

Here, $\delta < 1/8$ is a parameter that we will specify later. So, the means of $\nu_1^{(1)}$ will be $\big(1/2 - \delta, \quad 1/2, \quad 1/2, \quad ..., \quad 1/2\big)$. Since $\sum_{a=1}^{K} \mathbb{E}[N_{a,T}] = T$, where $N_{a,T} = \sum_{t=1}^{T} \mathbb{1}(A_t = a)$, we know

$$\exists j \in 2, 3, ..., K \text{ s.t. } \mathbb{E}^{(1)}[N_{j,T}] \leq \frac{T}{K-1}$$

. We will next define $\nu^{(2)}$ as follows:

$$\nu_j^{(2)} = \text{Bern}\left(\frac{1}{2} - 2\delta\right) \text{ and } \nu_i^{(2)} = \nu_i^{(1)}, \forall i \neq j,$$

So, the means of $\nu^{(2)}$ would be $(1/2-\delta, 1/2, 1/2, ..., \underbrace{1/2 - 2\delta}_{j-th}, ..., 1/2)$. Let $R'_T(\pi, \ell) = \sum_t \ell_t(A_t) - \min_{a\in[K]} \sum_t \ell_t(a)$ and let $\mathbb{E}_\pi$ denote randomness w.r.t. policy, so that $R_T(\pi, \underline{\ell}) = \mathbb{E}_\pi[R'_T(\pi, \underline{\ell})]$. We can now bound the worst-case regret for policy π as follows:

$$\begin{aligned}
R^\star(\pi) &= \sup_{\ell\in[0,1]^K} R_T(\pi, \ell) = \sup_{\ell\in[0,1]^K} \mathbb{E}_\pi[R'_T(\pi, \ell)] \\
&\geq \mathbb{E}_{i\sim\text{Unif}(\{1,2\})} \mathbb{E}_{\ell\sim\nu^{(i)}} \mathbb{E}_\pi[R'_T(\pi, \ell)] \\
&= \frac{1}{2}\mathbb{E}_\pi\left[\mathbb{E}_{\ell\sim\nu^{(1)}}\left[R'_T(\pi, \ell)\right]\right] + \frac{1}{2}\mathbb{E}_\pi\left[\mathbb{E}_{\ell\sim\nu^{(2)}}\left[R'_T(\pi, \ell)\right]\right]
\end{aligned}$$

where the inequality follows from the fact that $i \sim \text{Unif}(\{1, 2\})$, $\ell \sim \nu^{(i)}$ is a distribution of $\{0, 1\}^{K\times T} \subseteq [0, 1]^{K\times T}$ and the fact that the maximum is larger than the average. Next, consider

$$\begin{aligned}
\mathbb{E}_{\ell\sim\nu^{(i)}}\left[R'_T(\pi, \ell)\right] &= \mathbb{E}_{\ell\sim\nu^{(i)}}\left[\sum_{t=1}^{T} \ell_t(A_t) - \min_{a\in[K]} \sum_{t=1}^{T} \ell_t(a)\right. \\
&\geq \mathbb{E}_{\ell\sim\nu^{(i)}}\left[\sum_{t=1}^{T} \ell_t(A_t)\right] - \min_{a\in[K]} \mathbb{E}_{\ell\sim\nu^{(i)}} \sum_{t=1}^{T} \ell_t(a) \\
&= \mathbb{E}_{\ell\sim\nu^{(i)}}\left[\sum_{t=1}^{T} \ell_t(A_t)\right] - T\mu^\star(\nu^{(i)})
\end{aligned}$$

where $\mu^\star(\nu) = \min_{a\in[K]} \mathbb{E}_{X\sim\nu_a}[X]$. The inequality is by Jensen's inequality and noting that the pointwise minimum is concave, i.e. $\mathbb{E}[\min_i x_i] \leq \min_i \mathbb{E}[x_i]$. Therefore, we have

$$\mathbb{E}_\pi\left[\mathbb{E}_{\ell\sim\nu^{(i)}}\left[R'_T(\pi, \ell)\right]\right] \geq \mathbb{E}_\pi \mathbb{E}_{\ell\sim\nu^{(i)}}\left[\sum_{t=1}^{T} \ell_t(A_t)\right] - T\mu^\star(\nu^{(i)}) = R_T^{\text{stoc}}(\pi, \nu^{(i)})$$

here $R_T^{\text{stoc}}(\pi, \nu^{(i)})$ is the stochastic bandit regret of policy π on the stochastic bandit model $\nu^{(i)}$. Therefore, we have

$$R^\star(\pi) \geq \frac{1}{2}\left(R_T^{\text{stoc}}(\pi, \nu^{(1)}) + R_T^{\text{stoc}}(\pi, \nu^{(2)})\right)$$

The term inside the paranthesis is similar to the quantity we obtained when proving lower bounds for stochastic bandits. You will complete the remainder of the proof in the homework.You can show $(\star) \in \Omega(\sqrt{KT})$ for an appropriate choice of δ. □