**CS 475/675 Machine Learning** **Spring 2023**

# Lecture 15: Minimax Lower Bounds for Stochastic Multi-Armed Bandits

*Lecturer: Thanh Nguyen-Tang* *Scribe: Thanh Nguyen-Tang*

**Acknowledgement:** *LaTeX template courtesy of UC Berkeley EECS dept.*



## 15.1 Introduction

The UCB algorithm we considered in the previous lecture yields the regret bound $\mathcal{O}(\sqrt{nk\log n})$ which holds regardless of any bandit model $\mu = (\mu_1, \ldots, \mu_k)$. A natural question we might ask is whether we can do better than $\mathcal{O}(\sqrt{nk\log n})$. In other words, how can we justify the tightness of a regret bound in terms of $n$ and $k$? One way to answer this kind of questions is to construct a minimax lower bound. A minimax lower bound $f(n,k)$ essentially establishes that for *any* MAB algorithm $\mathcal{A}$, there exists an bandit instance $M$ in the instance family $\mathcal{M}$ such that the regret of that algorithm cannot be smaller than the minimax lower bound $f(n,k)$. More formally, we want to show the result of the following flavor:

$$\sup_{\mathcal{A}} \inf_{\mu\in\mathcal{M}} \mathbb{E}_M[R_n] \geq f(n,k).$$

On the basis of the minimax lower bound $f(n,k)$, we can justify if the regret bound $\mathcal{O}(\sqrt{nk\log n})$ is tight. In fact, in this lecture, we show that $(n,k) = c\sqrt{n(k-1)}$ for some absolute constant $c > 0$. Thus, the UCB algorithm is minimax-optimal up to log factors.

We now look at minimax lower bounds for stochastic MAB algorithms.

## 15.2 Hypothesis Testing

**Lemma 15.1 (Neyman Pearson)** *Let $x_1, \ldots, x_n \in \mathcal{X}^n$ the random variable that is distributed according to either $P$ and $Q$. For any test function $f : \mathcal{X}^n \to \{0,1\}$, we have*

$$P\left(f(x_1,\ldots,x_n) = 0\right) + Q(f(x_1,\ldots,x_n) = 1) \geq 1 - \sqrt{\frac{1}{2}\mathrm{KL}(P,Q)}.$$

## 15.3 Minimax Lower Bounds

Recall that a $k$-armed bandit model is fully characterized by the probability distributions $(p_1, \ldots, p_k)$ with means $(\mu_1, \ldots, \mu_k)$. An algorithm is *fully characterized* by the sequence of random variables $(a_1, \ldots, a_n)$

[0]These notes are partially based on those of Patrick Rebeschini.

where each $a_i$ depends on the data prior to iteration $i$, i.e. $\{z_{j,a_j}\}_{j\leq i-1}$. It is more helpful to take a functional perspective by viewing each action $a_i$ as a mapping $a_i : \mathbb{R}^{i-1} \to [k]$ where $a_i = a_i(\{z_{j,a_j}\}_{j\leq i-1}) \in [k]$ (rather than a fixed value). Thus, the algorithm does not depend on the bandit model.

**Theorem 15.2** *Let $\mathcal{M}$ be the set of $k$-armed bandit problems $\mu = (\mu_1, \ldots, \mu_k)$. For any $n \geq (k-1)/2$, we have*

$$\inf_{(a_1,\ldots,a_n)} \sup_{\mu\in\mathcal{M}} \mathbb{E}_\mu[R_n] \geq c\sqrt{n(k-1)},$$

*for some absolute constant $c > 0$.*

The $\inf_{(a_1,\ldots,a_n)} \sup_{\mu\in\mathcal{M}}$ in Theorem 15.2 reads that for any algorithm $(a_1, \ldots, a_n)$, there exists a $k$-armed bandit problem $\mu \in \mathcal{M}$ such that $\mathbb{E}_\mu[R_n] \geq c\sqrt{n(k-1)}$.

**Proof of Theorem 15.2**

Consider any algorithm $(a_1, \ldots, a_n)$. The random variable that we are concerned with is the composite r.v. $Z = (z_{1,a_1}, \ldots, z_{n,a_n}) \in \mathbb{R}^n$. The randomness of $Z$ is fully characterized by the algorithm $(a_1, \ldots, a_n)$ and the underlying bandit model.

**Step 1: Reduction to Testing.** The idea to proving minimax lower bounds is that we construct two bandit models $\mu$ and $\nu$ such that any algorithm $(a_1, \ldots, a_n)$ must commit a high regret in either of these bandits. A key to implement this intuition is to relate it to hypothesis testing. Let $P_\mu$ (respectively, $P_\nu$) be the probability distribution of $Z$ under the bandit model $\mu$ ($\nu$, respectively) and the considered algorithm $(a_1, \ldots, a_n)$. We denote $\mathbb{E}_\mu$ as the expectation under $P_\mu$.

The main technical difficulty is in how to construct $\mu$ and $\nu$. The criteria to have a useful construction is that (i) $\mu$ and $\nu$ have different optimal actions so that they can potentially confuse any algorithm $(a_1, \ldots, a_n)$ (in the sense that the algorithm cannot obtain small regret in both problems simultaneously), and (ii) $\mu$ and $\nu$ need to be similar to each other enough so we can have a tight lower bound and can compute $\mathrm{KL}(P_\mu, P_\nu)$ conveniently. To this end, we consider $\mu$ and $\nu$ be Bernoulli distributions with the means of the following form:

$$\begin{aligned}
\mu &= (\frac{1}{2} + \Delta, \frac{1}{2}, \ldots, \frac{1}{2}) \in \mathbb{R}^k \\
\nu &= (\frac{1}{2} + \Delta, \frac{1}{2}, \ldots, \frac{1}{2}, \frac{1}{2} + 2\Delta, \frac{1}{2}, \ldots, \frac{1}{2}) \in \mathbb{R}^k
\end{aligned}$$

where $\Delta \in (0, 1/4)$ and $\nu$ has the optimal action at $b > 1$ with mean $1 + 2\Delta$, where $b$ will be tuned later.

We have

$$\begin{aligned}
\mathbb{E}_\mu[R_n] &= \mathbb{E}_\mu\left[R_n | N_{n,1} \leq \frac{n}{2}\right] P_\mu(N_{n,1} \leq \frac{n}{2}) + \mathbb{E}_\mu\left[R_n | N_{n,1} \geq \frac{n}{2}\right] P_\mu(N_{n,1} \geq \frac{n}{2}) \\
&\geq \mathbb{E}_\mu\left[R_n | N_{n,1} \leq \frac{n}{2}\right] P_\mu(N_{n,1} \leq \frac{n}{2}) \\
&\geq \frac{\Delta n}{2} P_\mu(N_{n,1} \leq \frac{n}{2})
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}_\nu[R_n] &= \mathbb{E}_\nu\left[R_n | N_{n,1} \leq \frac{n}{2}\right] P_\nu(N_{n,1} \leq \frac{n}{2}) + \mathbb{E}_\nu\left[R_n | N_{n,1} \geq \frac{n}{2}\right] P_\nu(N_{n,1} \geq \frac{n}{2}) \\
&\geq \mathbb{E}_\nu\left[R_n | N_{n,1} \geq \frac{n}{2}\right] P_\nu(N_{n,1} \geq \frac{n}{2})
\end{aligned}$$

$$\geq \frac{\Delta n}{2} P_\nu(N_{n,1} \geq \frac{n}{2})$$

Thus, we have

$$\begin{aligned}
\sup_{m\in\mathcal{M}} \mathbb{E}_m[R_n] &\geq \max_{m\in\{\mu,\nu\}} \mathbb{E}_m[R_n] \\
&\geq \frac{1}{2}(\mathbb{E}_\mu[R_n] + \mathbb{E}_\nu[R_n]) \\
&\geq \frac{\Delta n}{4}\left(P_\mu(N_{n,1} \leq \frac{n}{2}) + P_\nu(N_{n,1} \geq \frac{n}{2})\right) \\
&\geq \frac{\Delta n}{4}\left(1 - \sqrt{\frac{1}{2}\mathrm{KL}(P_\mu, P_\nu)}\right) \qquad \text{(Lemma 15.1)} \qquad (15.1)
\end{aligned}$$

**Step 2: Optimizing** $\mathrm{KL}(P_\mu, P_\nu)$**.** Note that the inequality in Eq. (15.1) holds for any $\Delta \in (0, 1/4)$ and $2 \leq b \leq k$. The idea now is to optimize over $\Delta$ and $b$ to maximize $\frac{\Delta n}{2}\left(1 - \sqrt{\frac{1}{2}\mathrm{KL}(P_\mu, P_\nu)}\right)$ (so that we have the tightest lower bound as possible). We have

$$\mathrm{KL}(P_\mu, P_\nu) = \sum_{a\in[k]} \mathrm{KL}(P_{\mu,a}, P_{\nu,a})\mathbb{E}_\mu[N_{\mu,a}] = \mathrm{KL}(P_{\mu,b}, P_{\nu,b})\mathbb{E}_\mu[N_{n,b}].$$

**Step 2.1: Optimizing over** $b$**.** To make $\mathrm{KL}(P_{\mu,b}, P_{\nu,b})\mathbb{E}_\mu[N_{n,b}]$ small, it is natural to select $b$ as $b = \arg\min_{a>1} \mathbb{E}_\mu[N_{n,a}]$. As a result, we have

$$\mathbb{E}_\mu[N_{n,b}] \leq \frac{1}{k-1}\sum_{a=2}^{k} \mathbb{E}_\mu[N_{n,a}] \leq \frac{1}{k-1}\sum_{a=1}^{k} \mathbb{E}_\mu[N_{n,a}] = \frac{n}{k-1}.$$

**Step 2.2: Optimizing over** $\Delta$**.** We also have

$$\begin{aligned}
\mathrm{KL}(P_{\mu,b}, P_{\nu,b}) &= \frac{1}{2}\log\frac{1/2}{1/2+\Delta} + \frac{1}{2}\log\frac{1/2}{1/2-\Delta} \\
&= -\frac{1}{2}\log(1 - 16\Delta^2) \\
&\leq 32\Delta^2 \qquad (-\log(1-x) \leq 2x, \forall x \in (0,1)).
\end{aligned}$$

Overall we have

$$\mathrm{KL}(P_\mu, P_\nu) = \mathrm{KL}(P_{\mu,b}, P_{\nu,b})\mathbb{E}_\mu[N_{n,b}] \leq 32\frac{\Delta^2 n}{k-1}.$$

Thus, we have

$$\frac{\Delta n}{2}\left(1 - \sqrt{\frac{1}{2}\mathrm{KL}(P_\mu, P_\nu)}\right) \geq \frac{\Delta n}{2}(1 - \frac{1}{2}\Delta\sqrt{32\frac{n}{k-1}}). \qquad (15.2)$$

The maximizer of the RHS is $\Delta = \frac{1}{4}\sqrt{\frac{k-1}{2n}}$. Plugging Eq. (15.2) into Eq. (15.1) with $\Delta = \frac{1}{4}\sqrt{\frac{k-1}{2n}}$ completes the proof. ■