Machine Learning Theory (CS 6783)

Lecture 18: Lower Bounds via Stochastic Multi-armed Bandit

1 Stochastic Multi-armed Bandit Problem

We already looked at adversarial Multi-armed bandit problem. In this lecture we switch to the stochastic version of the problem. In this version, As before we have N arms or N experts. We only receive as feedback on any round the loss of the arm or expert we chose and not the other. Only difference is that there is a fixed distribution $\mathcal{D}$ over losses of the N arms. That is, $\mathcal{D} \in \Delta([0,1]^N)$ (we can also use losses between -1 and +1, I am simply choosing this one). We will define regret as

$$n\mathrm{Reg}_n = \sum_{t=1}^n \ell_t[I_t] - \sum_{t=1}^n \ell_t[i^*]$$

where $i^* = \operatorname*{argmin}_{i \in [N]} \mathbb{E}_{\ell \sim \mathcal{D}}[\ell[i]]$. Note that the usual notion of regret we would pick i^* that minimizes $\sum_{t=1}^n \ell_t[i^*]$, but due to concentration, using i^* that minimizes expected loss is actually pretty close (at most $1/\sqrt{n}$ away). Hence we will use this version of regret as a proxy. We will quickly review algorithms for stochastic multi-armed bandit problems next lecture but for now let us do lower bounds as lower bounds here are also lower bounds for adversarial setting naturally.

A quick note before we proceed, for every $k \in [N]$ let $\mu_k = \mathbb{E}_{\ell \sim \mathcal{D}}[\ell[k]]$ and define $\Delta_k = \mu_k - \mu_{i^*}$, the sub-optimality of the k'th arm. Now note that

$$n\mathbb{E}_{\mathcal{D},\mathbf{A},n}[\mathrm{Reg}_n] = \sum_{t=1}^n \mathbb{E}_{\mathcal{D},\mathbf{A},n}[\ell[I_t]] - n\mu_{i^*} = \sum_{t=1}^n \mathbb{E}_{\mathcal{D},\mathbf{A},n}[\mu_{I_t} - \mu_{i^*}] = \sum_{t=1}^n \mathbb{E}_{\mathcal{D},\mathbf{A},n}[\Delta_{I_t}] = \sum_{k=1}^N \mathbb{E}_{\mathcal{D},\mathbf{A},n}[n_{k,n}]\,\Delta_k$$

where $n_{k,n}$ is the number of times arm $k \in [N]$ was pulled amongst the n rounds.

2 Lower Bound Tools

For this section just to ease notation, assume losses on each arm take on a finite set of values (for what we care you can just assume that losses are 0 or 1). To deal with more general case you will need to deal with Radon Nykodym derivative and differential entropy etc.

Key Idea: The main idea in these lower bound proofs is to find two distributions $\mathcal{D}$ and $\mathcal{D}'$ and an event A such that,

$$\mathbb{E}_{\mathcal{D},\mathbf{A},n}[\mathrm{Reg}_n | A] \geq \frac{1}{2}\Delta \qquad \text{and} \qquad \mathbb{E}_{\mathcal{D}',\mathbf{A},n}[\mathrm{Reg}_n | A^c] \geq \frac{1}{2}\Delta$$

where $\mathbf{A}$ refers to the bandit algorithm. That is, even A is "bad" under $\mathcal{D}$ and A^c is bad under $\mathcal{D}'$. For instance, event A could be that a sub-optimal arm according to $\mathcal{D}$ is pulled more than $n/2$ number of times. With this, note that,

$$\mathbb{E}_{\mathcal{D},\mathbf{A},n}[\mathrm{Reg}_n] = \mathbb{E}_{\mathcal{D},\mathbf{A},n}[\mathrm{Reg}_n|A]\,P_{\mathcal{D},\mathbf{A},n}(A) + \mathbb{E}_{\mathcal{D},\mathbf{A},n}[\mathrm{Reg}_n|A^c]\,P_{\mathcal{D},\mathbf{A},n}(A^c) \geq \frac{\Delta}{2} P_{\mathcal{D},\mathbf{A},n}(A)$$

where $P_{\mathcal{D},\mathbf{A},n}$ is the probability distribution over observed losses and arms chosen induced by the algorithms $\mathbf{A}$ when run for n rounds and when losses are drawn iid from distribution $\mathcal{D}$. Similarly,

$$\mathbb{E}_{\mathcal{D}',\mathbf{A},n}\left[\mathrm{Reg}_n\right] \geq \frac{\Delta}{2} P_{\mathcal{D}',\mathbf{A},n}(A^c)$$

Hence we have that

$$\begin{aligned}
\mathbb{E}_{\mathcal{D},\mathbf{A},n}\left[\mathrm{Reg}_n\right] + \mathbb{E}_{\mathcal{D}',\mathbf{A},n}\left[\mathrm{Reg}_n\right] &\geq \frac{\Delta}{2}\left(P_{\mathcal{D},\mathbf{A},n}(A) + P_{\mathcal{D}',\mathbf{A},n}(A^c)\right) \\
&= \frac{\Delta}{2}\left(1 + P_{\mathcal{D},\mathbf{A},n}(A) - P_{\mathcal{D}',\mathbf{A},n}(A)\right) \\
&\geq \frac{\Delta}{2}\left(1 - \left|P_{\mathcal{D},\mathbf{A},n}(A) - P_{\mathcal{D}',\mathbf{A},n}(A)\right|\right) \\
&\geq \frac{\Delta}{2}\left(1 - \|P_{\mathcal{D},\mathbf{A},n} - P_{\mathcal{D}',\mathbf{A},n}\|_{\mathrm{TV}}\right)
\end{aligned}$$

where, the total variation distance between two distributions P, P' is defined as:

$$\|P - P'\|_{TV} = \sup_{A \in \mathcal{F}} \left|P(A) - P'(A)\right|$$

where $\mathcal{F}$ above is the sigma algebra. Now the key idea is that if we can find such $\mathcal{D}, \mathcal{D}'$ that are close in terms of KL we would be done.

It is often easier to work with KL divergence than TV distance and for this reason the following relationships bounding TV distance in terms of KL is useful:

1. Pinsker's Inequality
$$\|P - P'\|_{TV} \leq \sqrt{\frac{1}{2}\mathrm{KL}(P|P')}$$

2. An exponential bound:
$$\|P - P'\|_{TV} \leq \sqrt{1 - \exp\left(-\mathrm{KL}(P|P')\right)}$$

First, let us start with a bound on KL divergence over distributions of runs under two different product dsitribution over losses of arms.

Lemma 1. *Let $\mathcal{D}$ and $\mathcal{D}'$ be two distributions on the losses with $(\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_N)$ and $(\mathcal{D}'_1, \mathcal{D}'_2, \ldots, \mathcal{D}'_N)$ being the corresponding marginal distributions for the N arms. Let $\mathbf{A}$ be any stochastic bandit algorithm (that is, the algorithm picks the next action to play (possibly randomly) based on past actions and losses observed). In this case,*

$$\mathrm{KL}(P_{\mathbf{A},\mathcal{D},n}|P_{\mathbf{A},\mathcal{D}',n}) = \sum_{j=1}^{N} \mathbb{E}_{R\sim P_{\mathbf{A},\mathcal{D},n}}\left[n_{j,n}\right] \mathrm{KL}(\mathcal{D}_k|\mathcal{D}'_k)$$

Proof. Consider any sequence R of action losses pairs we encounter, say

$$R = (I_1, \ell_1[I_1], \ldots, I_n, , \ell_n[I_n])$$

Now

$$P_{\mathbf{A},\mathcal{D},n}(R) = \prod_{t=1}^{n} P_{\mathcal{A}}(I_t|I_1, \ell_1[I_1], \ldots, I_{t-1}, \ell_{t-1}[I_{t-1}]) \cdot P_{\mathcal{D}_{I_t}}(\ell_t[I_t])$$

and similarly,

$$P_{\mathbf{A},\mathcal{D}',n}(R) = \prod_{t=1}^{n} P_{\mathcal{A}}(I_t|I_1, \ell_1[I_1], \ldots, I_{t-1}, \ell_{t-1}[I_{t-1}]) \cdot P_{\mathcal{D}'_{I_t}}(\ell_t[I_t])$$

(I am abusing notation here to make life simpler. Basically we want to look at probability of the run under the two distributions). Hence we can conclude that:

$$\begin{aligned}
&\mathrm{KL}(P_{\mathbf{A},\mathcal{D},n}|P_{\mathbf{A},\mathcal{D}',n}) \\
&= \mathbb{E}_{R\sim P_{\mathbf{A},\mathcal{D},n}}\left[\log\left(\frac{\prod_{t=1}^{n} P_{\mathcal{A}}(I_t|I_1, \ell_1[I_1], \ldots, , I_{t-1}, \ell_{t-1}[I_{t-1},]) \cdot P_{\mathcal{D}_{I_t}}(\ell_t[I_t])}{\prod_{t=1}^{n} P_{\mathcal{A}}(I_t|I_1, \ell_1[I_1], \ldots, I_{t-1}, \ell_{t-1}[I_{t-1}]) \cdot P_{\mathcal{D}'_{I_t}}(\ell_t[I_t])}\right)\right] \\
&= \mathbb{E}_{R\sim P_{\mathbf{A},\mathcal{D},n}}\left[\sum_{t=1}^{n}\log\left(\frac{P_{\mathcal{D}_{I_t}}(\ell_t[I_t])}{P_{\mathcal{D}'_{I_t}}(\ell_t[I_t])}\right)\right] \\
&= \mathbb{E}_{R\sim P_{\mathbf{A},\mathcal{D},n}}\left[\sum_{t=1}^{n}\mathbb{E}_{\ell_t[I_t]\sim D_{I_t}}\left[\log\left(\frac{P_{\mathcal{D}_{I_t}}(\ell_t[I_t])}{P_{\mathcal{D}'_{I_t}}(\ell_t[I_t])}\right)\right]\right] \\
&= \mathbb{E}_{R\sim P_{\mathbf{A},\mathcal{D},n}}\left[\sum_{t=1}^{n}\mathrm{KL}(P_{\mathcal{D}_{I_t,S_t}}|P_{\mathcal{D}'_{I_t,S_t}})\right] \\
&= \sum_{j=1}^{N}\mathbb{E}_{R\sim P_{\mathbf{A},\mathcal{D},n}}[n_{j,n}]\left(\mathrm{KL}(\mathcal{D}_{k,s}|\mathcal{D}'_{k,s})\right)
\end{aligned}$$

□

3 Instance Independent Lower Bound

Theorem 2. *For all $N \geq 2$ and any $n > N$, there exists a distribution over losses of arms such that:*

$$\mathbb{E}[\mathrm{Reg}_n] \geq \frac{1}{8}\sqrt{\frac{N-1}{8n}}$$

Proof. Let $D_1 = B_{1/2-\Delta}$ and for all other $k \in [2, \ldots, N]$, $D_i = B_{1/2}$ where B_θ is the Bernoulli distribution with probability θ of a 1 and $1-\theta$ of producing 0. Now the way we produce D' is as follows. Pick $j^* = \underset{k\in[N]:k\neq 1}{\operatorname{argmin}} \mathbb{E}_{P_{\mathbf{A},\mathcal{D},n}}[n_{k,n}]$. Note that, $\mathbb{E}_{P_{\mathbf{A},\mathcal{D},n}}[n_{j^*,n}] \leq \frac{n}{N-1}$ Now set D' to be such that for all $k \neq j^*$, $D'_k = D_k$ and $D'_{j^*} = D_{1/2-2\Delta}$ Note that for D' optimal arm is j^* while for D it is 1. Define the event A to be

$$A = \{\omega \in \Omega : n_{1,n} \leq n/2\}$$

That is, the event that arm 1 is played less than 1/2 the times. Now note two things. First,

$$\mathbb{E}_{\mathcal{D}}[\mathrm{Reg}_n|A] \geq \frac{1}{2}\Delta$$

and

$$\mathbb{E}_{\mathcal{D}'}\left[\text{Reg}_n | A^c\right] \geq \frac{1}{2}\Delta$$

The first is true because we play optimal arm less than 1/2 the number of times and so we are Δ suboptimal more than 1/2 the time. The second is also true because when we are in complement of A, and distribution is D', then we are playing again a Δ sub-optimal arm more than 1/2 the times. Hence we conclude that:

$$\begin{aligned}\mathbb{E}_{\mathcal{D},\mathbf{A},n}\left[\text{Reg}_n\right] + \mathbb{E}_{\mathcal{D}',\mathbf{A},n}\left[\text{Reg}_n\right] &\geq \frac{\Delta}{2}\left(1 - \left|P_{\mathbf{A},\mathcal{D}',n}(A) - P_{\mathbf{A},\mathcal{D},n}(A)\right|\right) \\ &\geq \frac{\Delta}{2}\left(1 - \sqrt{\frac{1}{2}\text{KL}(P_{\mathbf{A},\mathcal{D},n} | P_{\mathbf{A},\mathcal{D}',n})}\right)\end{aligned}$$

Now using lemma from section 2 we have for this case that

$$\text{KL}(P_{\mathbf{A},\mathcal{D},n} | P_{\mathbf{A},\mathcal{D}',n}) = \mathbb{E}_{P_{\mathbf{A},\mathcal{D},n}}\left[n_{j^*,n}\right]\text{KL}(B_{1/2} | B_{1/2-2\Delta}) \leq \frac{n}{N-1}\text{KL}(B_{1/2} | B_{1/2-2\Delta})$$

However,

$$\begin{aligned}&\text{KL}(B_{1/2} | B_{1/2-2\Delta}) \\ &= \frac{1}{2}\log\left(\frac{1/2}{1/2-2\Delta}\right) + \frac{1}{2}\log\left(\frac{1/2}{1/2+2\Delta}\right) \\ &= \frac{1}{2}\log\left(\frac{1/4}{1/4-4\Delta^2}\right) \\ &= \frac{1}{2}\log\left(1 + \frac{4\Delta^2}{1/4-4\Delta^2}\right)\end{aligned}$$

since $1 + x < e^x$,

$$\leq \frac{4\Delta^2}{2(1/4-4\Delta^2)} \leq 16\Delta^2$$

Hence we conclude that:

$$\mathbb{E}_{\mathcal{D}}\left[\text{Reg}_n\right] + \mathbb{E}_{\mathcal{D}'}\left[\text{Reg}_n\right] \geq \frac{\Delta}{2}\left(1 - \sqrt{\frac{16n}{N-1}\Delta^2}\right)$$

Hence we can conclude that:

$$\max\{\mathbb{E}_{\mathcal{D}}\left[\text{Reg}_n\right], \mathbb{E}_{\mathcal{D}'}\left[\text{Reg}_n\right]\} \geq \frac{\Delta}{4}\left(1 - \sqrt{\frac{16n}{N-1}\Delta^2}\right)$$

Setting $\Delta = \sqrt{\frac{N-1}{64n}}$ yields the result that:

$$\max\{\mathbb{E}_{\mathcal{D}}\left[\text{Reg}_n\right], \mathbb{E}_{\mathcal{D}'}\left[\text{Reg}_n\right]\} \geq \frac{\Delta}{8} = \frac{1}{64}\sqrt{\frac{N-1}{n}}$$

and hence lower bound.

□

4 Instance Dependent Lower Bound

We use similar style proof technique now to provide an instance specific lower bound. This one needs a more careful statement.

Theorem 3. *For any Stochastic bandit algorithm (with binary losses) that guarantees a regret bound of:* $\mathbb{E}\left[\mathrm{Reg}_n\right] \leq \mathrm{rate}(n)$*, we have that for any distribution* $\mathcal{D}$*,*

$$\mathbb{E}_{\mathcal{D}}\left[\mathrm{Reg}_n\right] \geq \frac{2}{n} \sum_{k:\Delta_k \geq 4e\ \mathrm{rate}(n)} \frac{1}{\Delta_k}$$

Proof. Given a distribution $\mathcal{D}$ with mean losses of arms $\mu = (\mu_1, \ldots, \mu_N)$ where $\mu_k = \mathbb{E}_{\ell \sim D}\left[\ell[k]\right]$. Note that since the losses are binary, the marginal distribution of loss of each arm k is a Bernoulli distribution B_{μ_k}. Given an arm k, consider the product distribution $\mathcal{D}'_k = (B_{\mu_1}, \ldots, B_{\mu_{k-1}}, B_{\mu_k - 2\Delta_k}, B_{\mu_{k+1}}, \ldots, B_{\mu_N})$. That is, for this new distribution $\mathcal{D}'_k$, arm k is now the optimal arm with margin Δ_k. Now similar to the previous section, define event:

$$A = \{\omega \in \Omega : n_{k,n} > n/2\}$$

Note that conditioned on this event, clearly, regret under distribution $\mathcal{D}$ is larger than $\Delta_k/2$. On the other hand, conditioned on the complement of A, regret under distribution $\mathcal{D}'_k$ is lower bounded by $\Delta_k/2$. Hence, we have that

$$\begin{aligned}
\mathbb{E}_{\mathcal{D},\mathbf{A},n}\left[\mathrm{Reg}_n\right] + \mathbb{E}_{\mathcal{D}'_k,\mathbf{A},n}\left[\mathrm{Reg}_n\right] &\geq \frac{\Delta_k}{2}\left(1 - \|P_{\mathbf{A},\mathcal{D},n} - P_{\mathbf{A},\mathcal{D}'_k,n}\|_{TV}\right) \\
&\geq \frac{\Delta_k}{2}\left(1 - \sqrt{1 - \exp\left(-\mathrm{KL}\left(P_{\mathbf{A},\mathcal{D},n} | P_{\mathbf{A},\mathcal{D}'_k,n}\right)\right)}\right) \\
&\geq \frac{\Delta_k}{2}\exp\left(-\frac{1}{2}\mathrm{KL}\left(P_{\mathbf{A},\mathcal{D},n} | P_{\mathbf{A},\mathcal{D}'_k,n}\right)\right)
\end{aligned}$$

However, since $\max\{\mathbb{E}_{\mathcal{D}'_k,\mathbf{A},n}\left[\mathrm{Reg}_n\right], \mathbb{E}_{\mathcal{D},\mathbf{A},n}\left[\mathrm{Reg}_n\right]\} \leq \mathrm{rate}(n)$ (as the algorithm has a regret guarantee), we have that:

$$\mathrm{KL}\left(P_{\mathbf{A},\mathcal{D},n} | P_{\mathbf{A},\mathcal{D}'_k,n}\right) \geq 2\log\left(\frac{\Delta_k}{4\ \mathrm{rate}(n)}\right)$$

On the other hand, note that by Lemma 1, we have that

$$\begin{aligned}
\mathrm{KL}\left(P_{\mathbf{A},\mathcal{D},n} | P_{\mathbf{A},\mathcal{D}'_k,n}\right) &= \sum_{i=1}^{N} \mathbb{E}_{R \sim P_{\mathbf{A},\mathcal{D},n}}\left[n_{i,n}\right] \mathrm{KL}(\mathcal{D}_i | \mathcal{D}'_i) \\
&= \mathbb{E}_{R \sim P_{\mathbf{A},\mathcal{D},n}}\left[n_{k,n}\right] \mathrm{KL}(\mathcal{D}_k | \mathcal{D}'_k) \\
&= \mathbb{E}_{R \sim P_{\mathbf{A},\mathcal{D},n}}\left[n_{k,n}\right] \mathrm{KL}(B_{\mu_k} | B_{\mu_k - 2\Delta_k})
\end{aligned}$$

Now as long as $\mu_k \in (0.1, 0.9)$ we can conclude that

$$\mathrm{KL}(B_{\mu_k} | B_{\mu_k - 2\Delta_k}) \leq \Delta_k^2$$

Hence putting it all together we can conclude that for any k,:

$$\mathbb{E}_{R\sim P_{\mathbf{A},\mathcal{D},n}}\left[n_{k,n}\right] \geq \frac{2\log\left(\frac{\Delta_k}{4\ \mathrm{rate}(n)}\right)}{\Delta_k^2}$$

Using this with the fact that

$$\mathbb{E}_{\mathcal{D}}\left[\mathrm{Reg}_n\right] = \frac{1}{n}\sum_{i:\Delta_i\neq 0}\Delta_i\mathbb{E}\left[n_{i,n}\right] \geq = \frac{1}{n}\sum_{i:\Delta_i\geq 4\ e\ \mathrm{rate}(n)}\Delta_i\mathbb{E}\left[n_{i,n}\right]$$

Hence, we can conclude that:

$$\mathbb{E}_{\mathcal{D}.\mathbf{A},n}\left[\mathrm{Reg}_n\right] \geq \frac{1}{n}\sum_{k:\Delta_k\geq 4e\ \mathrm{rate}(n)}\frac{2\log\left(\frac{\Delta_k}{4\ \mathrm{rate}(n)}\right)}{\Delta_k} \geq \frac{1}{n}\sum_{k:\Delta_k\geq 4e\ \mathrm{rate}(n)}\frac{2}{\Delta_k}$$

□

Two things about the above lower bound.

1. First, note that if we consider the asymptotic bound, that is when $n\to\infty$, then we see that $\mathrm{rate}(n)\to 0$ for any consistent algorithm and so, for such an algorithm,

$$\lim_{n\to\infty}\mathbb{E}_{\mathcal{D}.\mathbf{A},n}\left[n\mathrm{Reg}_n\right] \geq \sum_{k:\Delta_k\neq 0}\frac{2}{\Delta_k}$$

   But in fact, take a closer look at the proof above, since $\mathrm{rate}(n)\geq\sqrt{1/n}$, if we actually work this out carefully we can conclude that,

$$\lim_{n\to\infty}\mathbb{E}_{\mathcal{D}.\mathbf{A},n}\left[\frac{n\mathrm{Reg}_n}{\log(n)}\right] \geq \sum_{k:\Delta_k\neq 0}\frac{2}{\Delta_k}$$

2. Second, note that the instance independent bound can be recovered easily by setting only one optimal arm and using all other non-optimal arms to have $\Delta_k = 4e\ \mathrm{rate}(n)$ so that for this specific distribution,

$$\mathbb{E}_{\mathcal{D}.\mathbf{A},n}\left[n\mathrm{Reg}_n\right] \geq \frac{2(N-1)}{4e\ \mathrm{rate}(n)}$$

   However, since $\mathrm{rate}(n)$ is an upper bound on every regret, we can conclude that

$$n\ \mathrm{rate}(n) \geq \mathbb{E}_{\mathcal{D}.\mathbf{A},n}\left[n\mathrm{Reg}_n\right] \geq \frac{2(N-1)}{4e\ \mathrm{rate}(n)}$$

   and hence we can conclude that $\mathrm{rate}(n)\geq\sqrt{\frac{N-1}{2en}}$