

Lecture 2: January 8, 2024

Lecturer: Yishay Mansour

Scribe: Yoav Nagel, Batya Berzack, Ido Cohen¹

1 Lecture Overview

In this lecture we discuss sampling lower bounds for MAB and we present two regret lower bounds (worst case and instance based). We will begin by covering basic concepts from information theory, L1-norm and KL-Divergence as measures of how one probability distribution is different from a second. The lecture is based on the works of Slivkins[2] and Kleinberg et al.[1].

2 Distance between Distributions

Let P and Q be two probability distributions over the sample space Ω .

Definition The L_1 distance between distributions P and Q is defined as:

$$\|P - Q\|_1 = \sum_{x \in \Omega} |P(x) - Q(x)| \quad (1)$$

Definition The total variation distance (TV) between distribution P and Q is defined as:

$$\|P - Q\|_{TV} = \max_{A \subseteq \Omega} |P(A) - Q(A)| \quad (2)$$

Lemma 1 $\forall A \subseteq \Omega : P(A) - Q(A) = Q(\bar{A}) - P(\bar{A})$

Proof:

$$P(A) + P(\bar{A}) = 1 \quad (3a)$$

$$Q(A) + Q(\bar{A}) = 1 \quad (3b)$$

$$(3a) - (3b) : P(A) - Q(A) - Q(\bar{A}) + P(\bar{A}) = 0 \quad (3c)$$

□

Lemma 2 $\|P - Q\|_1 = 2 \|P - Q\|_{TV}$

Proof: Let $A = \{x \in \Omega : P(x) \geq Q(x)\}$.

$$\begin{aligned} \|P - Q\|_1 &= \sum_{x \in \Omega} |P(x) - Q(x)| \\ &= \sum_{x \in A} P(x) - Q(x) + \sum_{x \notin A} Q(x) - P(x) \\ &\stackrel{\text{Lemma 1}}{=} 2 \cdot \sum_{x \in A} P(x) - Q(x) \end{aligned} \quad (4)$$

□

¹Based on scribe notes of Michael Shaik and Yuval Stein from 2021/22

Lemma 3 For every function $f : \Omega \rightarrow [-1, +1]$:

$$\mathbb{E}_P [f(x)] - \mathbb{E}_Q [f(x)] \leq \|P - Q\|_1$$

Proof:

$$\begin{aligned} |\mathbb{E}_P [f] - \mathbb{E}_Q [f]| &= \left| \sum_{x \in \Omega} f(x) (P(x) - Q(x)) \right| \\ &\leq \sum_{x \in \Omega} |f(x)| \cdot |P(x) - Q(x)| \\ &\leq \sum_{x \in \Omega} |P(x) - Q(x)| \\ &= \|P(x) - Q(x)\|_1 \end{aligned} \tag{5}$$

□

Lemma 4 There exists f such that:

$$\mathbb{E}_P [f(x)] - \mathbb{E}_Q [f(x)] = \|P - Q\|_1$$

Proof: Let f be defined as $f(x) = \mathbf{sign}(P(x) - Q(x))$

$$\begin{aligned} \mathbb{E}_P [f] - \mathbb{E}_Q [f] &= \sum_{x \in \Omega} f(x) (P(x) - Q(x)) \\ &= \sum_{x \in \Omega} |P(x) - Q(x)| \\ &= \|P(x) - Q(x)\|_1 \end{aligned} \tag{6}$$

□

Indeed it is intuitive to see that f maximises the the distance between the expected values.

Since we eventually want to analyse sampling many points, not just one, we will consider sampling m times *i.i.d*:

Definition Given independent distributions $P_1 \dots P_m, Q_1, \dots, Q_m$ we define the product distributions:

$$\begin{aligned} P^m &= P_1 \times P_2 \dots P_m \\ Q^m &= Q_1 \times Q_2 \dots Q_m \end{aligned}$$

Lemma 5 $\|P^m - Q^m\|_1 \leq \sum_{i=1}^m \|P_i^m - Q_i^m\|_1$

Proof: By induction on m .

For $m = 1$: trivial.

For $m > 1$:

Denote $x_{-i} = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_m)$. We can now write:

$$P^m(x) = \prod_{i=1}^m P_i(x_i) = P_m(x_m) P^{m-1}(x_{-m}) \tag{7}$$

$$Q^m(x) = \prod_{i=1}^m Q_i(x_i) = Q_m(x_m) Q^{m-1}(x_{-m}) \tag{8}$$

By our induction assumption:

$$\|P^{m-1} - Q^{m-1}\|_1 \leq \sum_{i=1}^{m-1} \|P_i - Q_i\|_1 \quad (9)$$

Therefore:

$$\begin{aligned} \|P^m - Q^m\|_1 &= \sum_{x_m} \sum_{x_{-m}} |P_m(x_m)P^{m-1}(x_{-m}) - Q_m(x_m)Q^{m-1}(x_{-m})| \\ &= \sum_{x_m} \sum_{x_{-m}} |P^{m-1}(x_{-m}) [P_m(x_m) - Q_m(x_m)] + Q_m(x_m) [P^{m-1}(x_{-m}) - Q^{m-1}(x_{-m})]| \\ &\leq \sum_{x_m} |P_m(x_m) - Q_m(x_m)| \underbrace{\sum_{x_{-m}} P^{m-1}(x_{-m})}_{=1} \\ &\quad + \sum_{x_{-m}} |P^{m-1}(x_{-m}) - Q^{m-1}(x_{-m})| \underbrace{\sum_{x_m} Q(x_m)}_{=1} \\ &\leq \|P_m - Q_m\|_1 + \sum_{i=1}^{m-1} \|P_i - Q_i\|_1 \end{aligned} \quad (10)$$

□

3 Lower Bound for Coin-Toss

In order to prove an *upper bound* we had to show a specific algorithm that attains this bound. To show a *lower bound* we will have to show that for every algorithm the bound holds. We will use the previous lemma to derive a lower bound on the number of tosses needed to distinguish between coins with distributions $Pr_{P^m} \sim Br[\frac{1}{2}]$ and $Pr_{Q^m} \sim Br[\frac{1+\epsilon}{2}]$. It holds that $\|P_i - Q_i\|_1 = \epsilon$.

Assume $f : \Omega \rightarrow \{0, 1\}$ is a function such that:

$$\begin{aligned} P(f = 1) &\leq \delta \\ Q(f = 1) &\geq 1 - \delta \end{aligned}$$

We can think of f as a function that given a series of i.i.d coin tosses returns 1 if the coins came from Q or 0 if they came from P , and is correct with probability $1 - \delta$.

Therefore, for a small constant value for δ :

$$1 - 2\delta \leq E_Q[f] - E_P[f] \stackrel{*}{\leq} \|Q^m - P^m\|_1 \stackrel{**}{\leq} \frac{1}{2}m\epsilon \quad (11)$$

(*) lemma 3

(**) lemma 5

In conclusion, to distinguish between P^m and Q^m with probability $1 - 2\delta$ we need $\Omega(\frac{1}{\epsilon})$ coins. This bound is not tight, since we know from Chernoff bound that the rate should be $\Theta(\frac{1}{\epsilon^2})$. Next we are going to derive a tighter bound using the KL-Divergence.

4 KL-Divergence

The Kullback-Leibler (KL) divergence is a measure of the difference between two distributions P and Q .

Definition The KL Divergence between two distributions P and Q is defined as:

$$KL(P||Q) = \mathbb{E}_P \left[\log \frac{P(x)}{Q(x)} \right] = \sum_{x \in \Omega} P(x) \log \frac{P(x)}{Q(x)} \quad (12)$$

Note that KL-divergence is not bounded and not symmetric, i.e. $KL(P||Q) \neq KL(Q||P)$, and also it is not a norm.

4.1 Properties of KL-divergence

4.1.1 Positive

$KL(P||Q) \geq 0$ and $KL(P||Q) = 0 \iff P = Q$.

Proof: Let $f(y) = y \log y$ for $y > 0$. The function f is convex.

$$\begin{aligned}
 KL(P||Q) &= \sum_{x \in \Omega} P(x) \log \frac{P(x)}{Q(x)} \\
 &= \sum_{x \in \Omega} Q(x) f\left(\frac{P(x)}{Q(x)}\right) = \mathbb{E}_Q \left[f\left(\frac{P(x)}{Q(x)}\right) \right] \\
 &\stackrel{\text{Jensen's inequality}}{\geq} f\left(\sum_{x \in \Omega} Q(x) \frac{P(x)}{Q(x)}\right) = f\left(\sum_{x \in \Omega} P(x)\right) = f(1) = 0
 \end{aligned} \tag{13}$$

The function f is strongly convex, and therefore the inequality will become equality if and only if $\forall x \in \Omega : P(x) = Q(x)$. \square

4.1.2 Chain Rule

For independent probability distributions: $KL(P^m||Q^m) = \sum_{i=1}^m KL(P_i||Q_i)$.

Proof: Let $h_i(x_i) = \log \frac{P_i(x_i)}{Q_i(x_i)}$.

$$\begin{aligned}
 KL(P^m||Q^m) &= \sum_{x \in \Omega} P^m(x) \log \left(\frac{P^m(x)}{Q^m(x)} \right) = \sum_{x \in \Omega} P^m(x) \log \left(\prod_{i=1}^m \frac{P_i(x)}{Q_i(x)} \right) \\
 &= \sum_{x \in \Omega} P^m(x) \left(\sum_{i=1}^m \log \left(\frac{P_i(x)}{Q_i(x)} \right) \right) = \sum_{i=1}^m \sum_{x \in \Omega} P^m(x) h_i(x) \\
 &= \sum_{i=1}^m \sum_{x_i^* \in \Omega} h_i(x_i^*) \sum_{x: x_i = x_i^*} P^m(x) = \sum_{i=1}^m \sum_{x_i^* \in \Omega} P_i(x_i^*) h_i(x_i^*) \\
 &= \sum_{i=1}^m \sum_{x_i^* \in \Omega} P_i(x_i^*) \left(\log \left(\frac{P_i(x_i^*)}{Q_i(x_i^*)} \right) \right) = \sum_{i=1}^m KL(P_i||Q_i)
 \end{aligned} \tag{14}$$

\square

4.1.3 Triangle Inequality

$$\forall A \subseteq \Omega : \sum_{x \in A} P(x) \log \frac{P(x)}{Q(x)} \geq P(A) \log \frac{P(A)}{Q(A)}.$$

Proof: $\forall x \in A$ let $P_A(x) = P(x|A)$ and $Q_A(x) = Q(x|A)$. Remember that $P(x) = P(A)P_A(x)$ and $Q(x) = Q(A)Q_A(x)$.

$$\begin{aligned} \sum_{x \in A} P(x) \log \frac{P(x)}{Q(x)} &= P(A) \sum_{x \in A} P_A(x) \log \frac{P(A)P_A(x)}{Q(A)Q_A(x)} \\ &= P(A) \underbrace{\sum_{x \in A} P_A(x) \log \frac{P_A(x)}{Q_A(x)}}_{KL(P_A||Q_A) \geq 0} + P(A) \log \frac{P(A)}{Q(A)} \underbrace{\sum_{x \in A} P_A(x)}_{=1} \\ &\geq P(A) \log \left(\frac{P(A)}{Q(A)} \right) \end{aligned} \quad (15)$$

□

4.1.4 Pinsker Inequality

$$\forall A \subseteq \Omega : 2(P(A) - Q(A))^2 \leq KL(P||Q).$$

Proof: By applying the 3rd property twice we get:

$$\sum_{x \in A} P(x) \log \frac{P(x)}{Q(x)} \geq P(A) \log \frac{P(A)}{Q(A)} \quad (16a)$$

$$\sum_{x \notin A} P(x) \log \frac{P(x)}{Q(x)} \geq (1 - P(A)) \log \frac{1 - P(A)}{1 - Q(A)} \quad (16b)$$

Denote $a = P(A)$, $b = Q(A)$. By summing the two inequalities (16a), (16b):

$$KL(P||Q) \geq a \log \frac{a}{b} + (1 - a) \log \frac{1 - a}{1 - b} = \int_a^b -\frac{a}{x} + \frac{1 - a}{1 - x} dx = \int_a^b \frac{x - a}{x(1 - x)} dx \quad (17)$$

Since $a, b \in [0, 1]$ we know $x(1 - x) \leq \frac{1}{4}$ and therefore:

$$KL(P||Q) \geq \int_a^b 4(x - a) dx = 2(b - a)^2 = 2(Q(A) - P(A))^2 \quad (18)$$

□

5 Better Lower Bound for Coin-Toss

We will use KL-divergence to improve the lower bound on the number of samples needed to distinguish between coins with distributions $P_i \sim Br(\frac{1}{2})$ and $Q_i \sim Br(\frac{1+\epsilon}{2})$.

$$\begin{aligned} KL(P||Q) &= \frac{1+\epsilon}{2} \log(1+\epsilon) + \frac{1-\epsilon}{2} \log(1-\epsilon) \\ &= \frac{1}{2} \underbrace{\log(1-\epsilon^2)}_{<0} + \frac{\epsilon}{2} \log\left(\frac{1+\epsilon}{1-\epsilon}\right) \leq \frac{\epsilon}{2} \log\left(1 + \frac{2\epsilon}{1-\epsilon}\right) \\ &\stackrel{\text{using } \log(1+x) \leq x}{\leq} \frac{\epsilon}{2} \cdot \frac{2\epsilon}{1-\epsilon} \stackrel{\text{assuming } \epsilon \leq \frac{1}{2}}{\leq} 2\epsilon^2 \end{aligned} \quad (19)$$

By using the chain-rule property of the KL-divergence we get $KL(P^m||Q^m) \leq 2m\epsilon^2$.
By using Pinsker inequality we get:

$$2(1 - 2\delta)^2 \leq KL(P^m||Q^m) \leq 2m\epsilon^2 \quad (20)$$

By rearranging the equation above we obtain an improved lower bound on $m \geq \frac{(1-2\delta)^2}{\epsilon^2} = \Omega(\frac{1}{\epsilon^2})$.

6 Best Arm Identification: Lower Bound

We begin with a short reminder about the Best-Arm-Identification problem:

- There is an action set A . Denote the number of actions as $k = |A|$.
- Each action $a \in A$ has a reward $r_t(a) \in [0, 1]$. Denote $\mu(a) = \mathbb{E}[r_t(a)]$.
- The algorithm uses actions T times and then outputs a guess $y_T \in A$ for the best action.
- The algorithm's goal is to maximize the probability of choosing the best arm a^* : $\Pr[y_T = a^*]$.
- An actions profile is defined as $I = \{\mu(a) : a \in A\}$.
- The profile set which we will consider in our setting is:

$$I_j = \begin{cases} \mu(i) = \frac{1}{2} & i \neq j \\ \mu(i) = \frac{1+\epsilon}{2} & i = j \end{cases}$$

- We will use a constant $\delta = 0.01$ for simplicity. We will want $\forall j \in [1 : k] : \Pr[y_T = j|I_j] \geq 0.99$.
- We will show that a lower bound for best arm identification is $T = \Omega(\frac{k}{\epsilon^2})$.

Lemma 6 *Suppose $T \leq \frac{ck}{\epsilon^2}$ for a small enough c . Then there exists at least $\lceil \frac{k}{3} \rceil$ actions for which $\Pr[y_T = a|I_a] < \frac{3}{4}$. We will prove this lemma for two cases:*

1. For $k = 2$ we prove the bound $T = \Omega(\frac{1}{\epsilon^2})$.
2. For $k \geq 24$ we prove the bound $T = \Omega(\frac{k}{\epsilon^2})$.

6.1 Proof for $k = 2$

There are two actions $\{1, 2\}$. Denote by B the realizations where the algorithm predicts $y_T = 1$. Assume by contradiction that $T = o(\frac{1}{\epsilon^2})$. For correctness, we demand

$$P_1(B) = P(y_T = 1|I_1) \geq \frac{3}{4} \quad (21)$$

$$P_2(B) = P(y_T = 1|I_2) \leq \frac{1}{4} \quad (22)$$

Then:

$$\begin{aligned} 2(P_1(B) - P_2(B))^2 &\leq KL(P_1||P_2) = \sum_{a \in \{1,2\}} \sum_{t=1}^T KL(P_1^{a,t}||P_2^{a,t}) \\ &\leq 2T \cdot 2\epsilon^2 = 4T\epsilon^2 \end{aligned} \quad (23)$$

By requiring in (21),(22) that $P_1(B) - P_2(B) \geq \frac{1}{2}$ we get: $\frac{1}{2} \leq 4T\epsilon^2$.
Therefore $T \geq \frac{1}{8\epsilon^2}$, in contradiction.

6.2 Proof for $k \geq 24$

We define another profile: $I_0 = \{\mu(a) = \frac{1}{2}\}$. Intuitively, this is a fair profile, which assigns the same expected value to all arms (no better action than others), in contrast to the unfair profiles I_j , which give the j^{th} arm's expected value a positive bias. Denote $\mathbb{E}_0[\cdot] = \mathbb{E}[\cdot|I_0]$ and $P_0[\cdot] = \Pr[\cdot|I_0]$.

We make the following claims:

Lemma 7 $\exists K_1 \subseteq A : |K_1| \geq \frac{2}{3}k \wedge \forall j \in K_1 : \mathbb{E}_0[T_j] \leq \frac{3T}{k}$.

Proof: By contradiction, assume that there is a subset $A' \subseteq A$ s.t. $|A'| > \frac{k}{3}$ and $\forall j \in A'$ we get $\mathbb{E}_0[T_j] > \frac{3T}{k}$. This implies that the arms in A' are played strictly more than T times, which is a contradiction. \square

Lemma 8 $\exists K_2 \subseteq A : |K_2| \geq \frac{2}{3}k \wedge \forall j \in K_2 : P_0[y_T = j] \leq \frac{3}{k}$.

Proof: By contradiction, assume that there is a subset $A' \subseteq A$ s.t. $|A'| > \frac{k}{3}$ with $\forall j \in A' : P_0[y_T = j] > \frac{3}{k}$. This implies that the combined probability that the arms in A' are selected to be strictly greater than 1, which is a contradiction. \square

Lemma 9 $\mathbb{E}_0[T_j] \leq \frac{3T}{k} \rightarrow P_0[T_j \geq \frac{24T}{k}] \leq \frac{1}{8}$.

Proof: Directly by applying Markov's inequality. Note that this is equivalent to $\Pr_0[T_j \leq \frac{24T}{k}] \geq \frac{7}{8}$. \square

Lemma 10 $\exists K_3 \subseteq A : |K_3| \geq \frac{1}{3}k \wedge P_0[T_j \leq \frac{24T}{k}] \geq \frac{7}{8} \wedge P_0[y_T = j] \leq \frac{3}{k}$.

Proof: We define $K_3 = K_1 \cap K_2$. Due to the pigeon hole principle we have $|K_1 \cap K_2| \geq \frac{1}{3}k$. It follows from the definition of K_3 and lemmas 7, 8. \square

We define some notations:

- A probability space for t samples of action $a \in A$, $\Omega_a^t = \{0, 1\}^t$. Each $\omega \in \Omega_a^t$ is a vector of length t where each ω_i is a realization of reward from action a .
- The general space is $\Omega = \prod_{a \in A} \Omega_a^T$.
- We choose any action $j \in K_3$ and define a reduced sample space $\Omega^* = \Omega_j^m \times \prod_{a \neq j} \Omega_a^T$, where arm j is played $m = \frac{24T}{k}$ times at most.
- We define for every profile I_l and $\forall B \subseteq \Omega^* : P_l^*(B) = \Pr[B|I_l]$.

We will now bound the distance between P_0^*, P_l^* using the chain rule and Pinsker's rule:

$$\begin{aligned}
2(P_0^*(B) - P_l^*(B))^2 &\leq KL(P_0^* || P_l^*) \\
&= \sum_a \sum_{t=1}^T KL(P_0^{*a,t} || P_l^{*a,t}) \\
&= \underbrace{\sum_{a \neq l} \sum_{t=1}^T KL(P_0^{a,t} || P_l^{a,t})}_{=0} + \sum_{t=1}^m \underbrace{KL(P_0^{*l,t} || P_l^{*l,t})}_{\leq 2\epsilon^2} \\
&\leq 2m\epsilon^2
\end{aligned} \tag{24}$$

Rearranging terms, we get for $m = \frac{24T}{k}, T \leq \frac{ck}{\epsilon^2}, \epsilon > 0$ (for small enough c):

$$\forall B \subseteq \Omega^* : |P_0^*(B) - P_l^*(B)| \leq \epsilon\sqrt{m} = \epsilon\sqrt{\frac{24T}{k}} = \epsilon\sqrt{\frac{24ck}{\epsilon^2 k}} = \sqrt{24c} \leq \frac{1}{8} \tag{25}$$

Notice the role of m in the simplification. We note the bound above holds only for events $B \subseteq \Omega^*$. Therefore, we can not use it the bound the event $B = \{y_T = j\}$ directly. To overcome this, we now define two events:

$$B_1 = \{y_T = j \wedge T_j \leq m\}, B_2 = \{T_j > m\}$$

Note that B_1 and B_2 are distinct and that both of these events are in Ω^* ($B_2 \subseteq \Omega^*$ because whether the algorithm samples action j more than m times is completely determined by the first m realizations). This implies that $P^*(B_1) = P(B_1)$ and $P^*(B_2) = P(B_2)$. Using B_1 and B_2 and $j \in K_3$, we bound $P_j^*(B)$ as follows:

$$P_j^*(B_1) \leq \frac{1}{8} + P_0(B_1) \leq \frac{1}{8} + P_0^*(y_T = j) \leq \frac{1}{8} + \frac{3}{k} \leq \frac{1}{4} \quad (26)$$

$$P_j^*(B_2) \leq \frac{1}{8} + P_0(B_2) \leq \frac{1}{8} + \frac{1}{8} = \frac{1}{4} \quad (27)$$

We can now bound the accuracy for action j :

$$P_j(y_T = j) \leq P_j^*(y_T = j \wedge T_j \leq m) + P_j^*(T_j > m) = P_j^*(B_1) + P_j^*(B_2) \leq \frac{1}{2} \quad (28)$$

This is even a tighter bound than $\frac{3}{4}$ for lemma 6.

For every MAB algorithm with $T \leq \frac{ck}{\epsilon^2}$ we get: $\Pr[y_T \neq a^*] \geq \frac{1}{12}$ ($\frac{1}{3}$ to be in K_3 and $\frac{1}{4}$ to be wrong), which implies for the probability that the algorithm fails to identify the best arm.

7 Regret Lower Bound for MAB

Theorem 11 *Fix any algorithm for best arm identification. Choose an arm a uniformly at random, and run the algorithm on profile I_a . Then:*

$$\mathbb{E}[\text{regret}] = \Omega\left(\sqrt{kT}\right)$$

Proof: Assume $T \leq \frac{ck}{\epsilon^2}, \epsilon > 0$. Fix a random profile I_{a^*} . From lemma 10: $\Pr[a_t \neq a^*] \geq \frac{1}{12}$. For $a_t \neq a^*$, the loss is:

$$\Delta(a_t) = \mu(a^*) - \mu(a_t) = \frac{1 + \epsilon}{2} - \frac{1}{2} = \frac{\epsilon}{2} \quad (29)$$

Therefore,

$$\mathbb{E}[\Delta(a_t)] = \Pr[a_t \neq a^*] \frac{\epsilon}{2} \geq \frac{\epsilon}{24} \quad (30)$$

Hence,

$$\mathbb{E}[\text{regret}] = \sum_{t=1}^T E[\Delta(a_t)] \geq \frac{T\epsilon}{24} \quad (31)$$

We got a lower bound as a function of ϵ . For $\epsilon = \sqrt{\frac{ck}{T}}$ we get:

$$\mathbb{E}[\text{regret}] \geq \frac{T}{24} \sqrt{\frac{ck}{T}} = \Omega\left(\sqrt{kT}\right) \quad (32)$$

□

Remark- in order to get the tightest bound for the regret, we would like to choose ϵ as large as possible, but since $T \leq \frac{ck}{\epsilon^2}$ we get that ϵ is bounded to be $\epsilon = \sqrt{\frac{ck}{T}}$.

8 Instance Dependent Regret Lower Bound

Instance dependent lower bounds follows the proof of Kleinberg et al. (lemma 14[1]). Assume we have an algorithm for MAB that works well on a specific profile, but also for every profile I_i plays non-optimal actions at most $c_1 T^{0.1}$ times. c_1 is a constant which can be dependent on the profile, but not on T .

We will show that in such a case:

$$\mathbb{E}[\text{regret}] = \Omega\left(\frac{k}{\Delta} \log T\right)$$

where Δ is the sub-optimality, defined as:

$$\Delta = \min_i \Delta_i, \quad \Delta_i = |\mu_i - \mu^*|$$

In other words, we want to demonstrate a lower bound on the regret for algorithms which are not "too bad" on all profiles. We will show the bound for $k = 2$.

Define the profiles:

$$\begin{aligned} P &= (\mu_1, \mu_2) : \mu_1 > \mu_2 \\ Q &= (\mu_1, \lambda) : \mu_1 < \lambda \end{aligned}$$

Such that in P the best action is 1 and in Q it is 2. We'll choose λ such that $KL(\mu_2||\lambda) = 1.1KL(\mu_2||\mu_1)$ (we write μ_1, μ_2, λ instead of writing $Br(\mu_1), Br(\mu_2), Br(\lambda)$).

Denote by n_i the number of time action i was chosen. Then:

$$\mathbb{E}_Q[n_1] = \mathbb{E}_Q[T - n_2] \leq c_1 T^{0.1} \quad (33)$$

We will bound the probability for profile Q :

$$\Pr_Q \left[n_2 \leq \frac{0.9 \log T}{KL(\mu_2||\lambda)} \right] = \Pr_Q \left[T - n_2 \geq T - \frac{0.9 \log T}{KL(\mu_2||\lambda)} \right] \stackrel{Markov}{\leq} \frac{\mathbb{E}_Q[T - n_2]}{T - \frac{0.9 \log T}{KL(\mu_2||\lambda)}} \leq \frac{c_1 T^{0.1}}{c_3 T} = c_2 T^{-0.9} \quad (34)$$

Let B be the event: $n_2 < \frac{0.9 \log T}{KL(\mu_2||\lambda)}$.

If $\Pr_P(B) < \frac{1}{3}$ then:

$$\mathbb{E}_P[n_2] \geq \frac{2}{3} \cdot \frac{0.9 \log T}{KL(\mu_2||\lambda)} = \Omega\left(\frac{\log T}{KL(\mu_2||\mu_1)}\right) \quad (35)$$

Using a similar analysis to our analysis above for $Br(\frac{1}{2})$ and $Br(\frac{1+\epsilon}{2})$, we know that in general $KL(\mu_2||\mu_1) \approx (\mu_2 - \mu_1)^2$. Using this we get:

$$\mathbb{E}_P[\text{regret}] = \mathbb{E}_P[n_2](\mu_1 - \mu_2) = \Omega\left(\frac{(\mu_1 - \mu_2) \log T}{KL(\mu_2||\mu_1)}\right) = \Omega\left(\frac{\log T}{\Delta}\right) \quad (36)$$

Otherwise, if $\Pr_P(B) \geq \frac{1}{3}$, by using the triangle inequality for KL-divergence:

$$KL(P||Q) \geq P(B) \log \frac{P(B)}{Q(B)} \geq \frac{1}{3} \log \frac{1/3}{Q(B)} \stackrel{(34)}{\geq} \frac{1}{3} \log \frac{T^{0.9}}{3c_2} = \Omega(\log T) \quad (37)$$

Now we will bound $\mathbb{E}_P[n_2]$. Note that P and Q are identical on action 1, so the only difference is action 2. Every time the algorithm selects action 2 it increases the KL by $KL(\mu_2||\Delta)$. Thus:

$$KL(P||Q) = \mathbb{E}_P[n_2] KL(\mu_2||\lambda) \Rightarrow \mathbb{E}_P[n_2] = \Omega\left(\frac{\log T}{KL(\mu_2||\lambda)}\right) \quad (38)$$

Since $KL(\mu_2||\lambda) = 1.1KL(\mu_2||\mu_1)$, it holds that $\mathbb{E}_P[n_2] = \Omega\left(\frac{\log T}{KL(\mu_2||\mu_1)}\right)$.

The regret will be:

$$\mathbb{E}_P[\text{regret}] = \mathbb{E}_P[n_2](\mu_1 - \mu_2) = \Omega\left((\mu_1 - \mu_2) \frac{\log T}{KL(\mu_2||\mu_1)}\right) = \Omega\left(\frac{\log T}{\Delta}\right) \quad (39)$$

References

- [1] Robert Kleinberg, Alexandru Niculescu-Mizil, and Yogeshwer Sharma. “Regret bounds for sleeping experts and bandits”. In: *Machine Learning* 80.2 (Sept. 2010), pp. 245–272. ISSN: 1573-0565. DOI: 10.1007/s10994-010-5178-7. URL: <https://doi.org/10.1007/s10994-010-5178-7>.
- [2] Aleksandrs Slivkins. “Introduction to Multi-Armed Bandits”. In: *CoRR* abs/1904.07272 (2019). arXiv: 1904.07272. URL: <http://arxiv.org/abs/1904.07272>.