Lecture 8: December 3

Lecturer: Yishay Mansour *Scribes: Oren Sheffer, Shiri Antaki, Tavor Lipman*

8.1 Lecture Overview

In this lecture we discuss sample size lower bounds, mostly on regret. For this we need some background in Information Theory.

8.2 Distance between Distributions

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

$$\|P - Q\|_1 = \sum_x |P(x) - Q(x)|$$

From the definition it follows that:

$$\begin{aligned}\|P - Q\|_1 &= \sum_x |P(x) - Q(x)| \\ &= \sum_{x:P(x)\geq Q(x)} P(x) - Q(x) + \sum_{x:Q(x)\geq P(x)} Q(x) - P(x) \\ &= 2 \cdot \sum_{x:Q(x)\geq P(x)} Q(x) - P(x)\end{aligned}$$

The expression $\sum_{x:Q(x)\geq P(x)} Q(x) - P(x) = \|Q - P\|_{TV}$ is called the *total variation distance.*

Lemma 8.1 *For every function $f : X \to [-1, +1]$:*

$$E_{x\sim Q}[f(x)] - E_{x\sim P}[f(x)] \leq \|P - Q\|_1$$

and there exists f such that:

$$E_{x\sim Q}[f(x)] - E_{x\sim P}[f(x)] = \|P - Q\|_1$$

We can think of f as deciding if x coming from Q or P. The larger $\|P - Q\|_1$ is, the easier it is to distinguish between P and Q.

[1]Based on the scribe notes from Amit Zohar, Amir Barda, Olya Sirkin, Fall 2017

Proof: For any f:

$$\begin{aligned} |E_P[f] - E_Q[f]| &= |\sum_x f(x)(P(x) - Q(x))| \\ &\leq \sum_x |f(x)||(P(x) - Q(x))| \\ &= \sum_x |P(x) - Q(x)| \\ &= \|P - Q\|_1 \end{aligned}$$

Now, let f be defined as $f(x) = sign(P(x) - Q(x))$

$$\begin{aligned} E_P[f] - E_Q[f] &= \sum_x f(x)(P(x) - Q(x)) \\ &= \sum_x |P(x) - Q(x)| \\ &= \|P - Q\|_1 \end{aligned}$$

□

We can think about f giving a correct answer as $Pr_Q[f(x) = 1]$ and of f giving a wrong answer as $Pr_P[f(x) = 1]$. We would like the difference between the two to be large so that f could distinguish between P and Q.
More specifically if $Pr_Q[f(x) = 1] \geq 1 - \delta$ and $Pr_P[f(x) = 1] \leq \delta$ we distinguish with probability $1 - 2\delta$. In this case $E_Q[f] \geq 1 - 2\delta$ and $E_P[f] \leq -1 + 2\delta$ so $E_Q[f] - E_P[f] \geq 2 - 4\delta$.

Definition Given P_i, Q_i independent distributions for every $i \in \{1, ..., m\}$, define:

$$P^m = P_1 \times P_2 \ldots P_m$$

$$Q^m = Q_1 \times Q_2 \times \ldots Q_m$$

Lemma 8.2

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

Proof: We are going to prove the lemma by induction on m.
For $m = 1$:

$$\|P^m - Q^m\|_1 = \|P_1 - Q_1\|_1$$

This follows trivially from definition of P^m and Q^m.
For $m > 1$:

$$\|P^m - Q^m\|_1 = \sum_{x_1} \sum_{x_2} \cdots \sum_{x_m} |\prod_i P_i(x_i) - \prod_i Q_i(x_i)|$$

Let us define $x_{-1} = x_2 \dots x_m$ and $\alpha(x_{-1}) = \prod_{x=2}^{m} P_i(x_i)$ and also $\beta(x_{-1}) = \prod_{x=2}^{m} Q_i(x_i)$.

We use the following identity:

$$P_1(x_1)\alpha(x_{-1}) - Q_1(x_1)\beta(x_{-1}) = \alpha(x_{-1})(P_1(x_1) - Q_1(x_1)) + Q_1(x_1)(\alpha(x_{-1}) - \beta(x_{-1}))$$

And have that:

$$\begin{aligned} \|P^m - Q^m\|_1 &\leq \sum_{x_1}\sum_{x_{-1}} \alpha(x_{-1})|P_1(x_1) - Q_1(x_1)| + \sum_{x_{-1}}\sum_{x_1} Q_1(x_1)|\alpha(x_{-1}) - \beta(x_{-1})| \\ &= \sum_{x_1} |P_1(x_1) - Q_1(x_1)| \sum_{x_{-1}} \alpha(x_{-1}) + \sum_{x_{-1}} |\alpha(x_{-1}) - \beta(x_{-1})| \sum_{x_1} Q(x_1) \end{aligned}$$

Note that $\sum_{x_{-1}} \alpha(x_{-1}) = 1$ and $\sum_{x_1} Q_1(x_1) = 1$ because α and Q_1 are distributions. Therefore:

$$\begin{aligned} \|P^m - Q^m\|_1 &\leq \sum_{x_1} |P_1(x_1) - Q_1(x_1)| + \sum_{x_{-1}} |\alpha(x_{-1}) - \beta(x_{-1})| \\ &\leq \|P_1 - Q_1\|_1 + \sum_{i=2}^{m} \|P_i - Q_i\|_1 \\ &= \sum_{i=1}^{m} \|P_i - Q_i\|_1 \end{aligned}$$

where in the second inequality we used the inductive hypothesis.

□

We'll use the above lemma to derive a lower bound on m.

Suppose $P_i \sim Br(\frac{1}{2})$, $Q_i \sim Br(\frac{1+\epsilon}{2})$. Where $Br(p)$ is a Bernoulli random variable of probability p. Then,

$$\|P_i - Q_i\|_1 = \epsilon$$

To distinguish between P^m and Q^m with probability $1 - 2\delta$ we need $\|P^m - Q^m\|_1 \geq 2 - 4\delta$. Therefore we get:

$$2 - 4\delta \leq \|P^m - Q^m\|_1 \leq m\epsilon$$

$$m \geq \frac{2 - 4\delta}{\epsilon} = \Omega(\frac{1}{\epsilon})$$

This bound is not tight, since we know that the right rate is $\Theta(\frac{1}{\epsilon^2})$.

Next we are going to derive tighter bound using KL-Divergence.

8.3 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:

$$KL(P||Q) = E_{x\sim P}[\log \frac{P(x)}{Q(x)}] = \sum_x P(x)\log(\frac{P(x)}{Q(x)})$$

Note that KL divergence is not symmetric; i.e., $KL(P||Q) \neq KL(Q||P)$.

8.3.1 Properties of KL divergence.

1. Positivity: $KL(P||Q) \geq 0$ and $KL(P||Q) = 0$ if and only if $P = Q$.

   Proof: Let $f(y) = ylog(y)$ for $y > 0$. The function f is convex since $f''(y) = \frac{1}{y} > 0$, therefore:

   $$KL(P||Q) = \sum_x P(x)log(\frac{P(x)}{Q(x)}) = \sum_x Q(x)f(\frac{P(x)}{Q(x)}) = E_{x\sim Q}[f(\frac{P(x)}{Q(x)})]$$

   $$\geq f(\sum_x Q(x)\frac{P(x)}{Q(x)}) = f(\sum_x P(x)) = f(1) = 0$$

   The inequality will become equality if and only if one of the following applies:

   (a) f is linear

   (b) $\forall x P(x) = Q(x)$

   Since f is not linear, we have that $KL(P||Q) = 0$ iff $P = Q$. □

2. Chain Rule for KL-divergence. Under independence assumption it holds that:

   $$KL(P^m||Q^m) = \sum_{i=1}^{m} KL(P_i||Q_i)$$

   Proof: Let us define $h_i(x_i) = log(\frac{P_i(x_i)}{Q_i(x_i)})$ and $x = (x_1 \ldots x_m)$, x_i $i.i.d$

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

□

3. The Pinsker Inequality:

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

Before proving the inequality, we first begin with a helpful lemma:

Lemma 8.3 $\sum_{x \in A} P(x) \log \frac{P(x)}{Q(x)} \geq P(A) \log \frac{P(A)}{Q(A)}$

Proof: For every: $\forall x \in A$ let $P_A(x) = P[x|A] = \frac{P(x)}{P(A)}$ and $Q_A(x) = Q[x|A] = \frac{Q(x)}{Q(A)}$. Therefore:

$$\sum_{x \in A} P(x) \log \frac{P(x)}{Q(x)} = P(A) \cdot \sum_{x \in A} P_A(x) \log \frac{P(A) \cdot P_A(x)}{Q(A) \cdot Q_A(x)} =$$

$$= P(A) \cdot \sum_{x \in A} P_A(x) \log \frac{P_A(x)}{Q_A(x)} + P(A) \log \frac{P(A)}{Q(A)} \cdot \sum_{x \in A} P_A(x)$$

Since $\sum_{x \in A} P_A(x) = 1$ and $P_A(x) \log \frac{P_A(x)}{Q_A(x)} = KL(P_A||Q_A) \geq 0$, we have

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

□

We now prove the Pinsker Inequality:

Proof: We apply the previous lemma twice and get:

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

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

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

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

Since $x(1-x) \leq \frac{1}{4}$:

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

□

8.4 A Better Lower Bound

Now we can get a much better bound than before. Assume that there are 2 possible distributions $Br(\frac{1+\epsilon}{2})$, and $Br(\frac{1}{2})$, and let μ is equal to each one of them in probability $\frac{1}{2}$.

We flip the coin with distribution μ for m times and we have a decision rule:

$$Rule : \{0,1\}^m \rightarrow \{high, low\}$$

We want:

$$P[rule(obs) = high \mid \mu = Br(\frac{1+\epsilon}{2})] \geq 1-\delta$$

and

$$P[Rule(obs) = high|\mu = Br(\frac{1}{2})] \leq \delta$$

meaning:

$$P[rule(obs) = high \mid \mu = Br(\frac{1+\epsilon}{2})] - P[Rule(obs) = high|\mu = Br(\frac{1}{2})] \geq 1-2\delta$$

Our goal now is to lower bound m as a function of ϵ.

Lemma 8.4 *We need $m \geq \frac{(1-2\delta)^2}{\epsilon^2} = \Omega(\frac{1}{\epsilon^2})$ to distinguish between the 2 coins.*

Proof:
Denote: $A_0 = \{obs : Rule(obs) = high\}$ We want:

$$P[A_0 \mid \mu = Br(\frac{1+\epsilon}{2})] - P[A_0|\mu = Br(\frac{1}{2})] \geq 1-2\delta$$

We define 2 distributions:

P_1 distribution for m coin flips with $Br(\frac{1+\epsilon}{2})$.

P_2 distribution for m coin flips with $Br(\frac{1}{2})$.

So

$$P_1(A_0) = P[A_0 \mid \mu = \ Br(\frac{1+\epsilon}{2})]$$

$$P_2(A_0) = P[A_0 \mid \mu = \ Br(\frac{1}{2})]$$

As we have seen before from pinker inequality and the chian rule:

$$2(P_1(A_0) - P_2(A_0))^2 \leq KL(P_1||P_2) \ = \sum_{i=1}^{m} KL(P_1(i)||P_2(i)) \leq 2\epsilon^2 m$$

$\implies$

$$1 - 2\delta \leq |P_1(A_0) - P_2(A_0)| \leq \epsilon\sqrt{m}$$

$\implies$

$$m \geq \frac{(1-2\delta)^2}{\epsilon^2} = \Omega(\frac{1}{\epsilon^2})$$

□

8.5 A Lower Bound for Best Arm Identification

We begin with a short reminder about the Best-Arm-Identification problem (for simplicity we assume the distribution for each arm is Bernouli):

- The algorithm uses actions T times and then outputs a guess $y_T \in A$ for the best action.
- We focus only on the quality of the guess - the probability for success: $Pr[y_T = a^*]$ where y_T is the algorithm output and a^* is the best action.
- Each action $a \in A$ has a reward $r_t(a) \in [0, 1]$.
- For every action $a \in A$ at time t, denote $\mu(a) = E[r_t(a)]$.
- An actions profile is defined as $I = \{\mu(a) : a \in A\}$
- The criterion for success (we ignore δ for simplicity): $Pr[y_T = a_I^*|I] \geq 0.99$ where $a_I^* = argmax_a \ \mu(a)$

We'll look at a profiles of the following form:

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

We'll want: $Pr[y_T = j|I_j] \geq 0.99$

Lemma 8.5 *Suppose $T \leq \frac{ck}{\epsilon^2}$ for a small enough c. Then there exist at least $\lceil \frac{k}{3} \rceil$ actions for which $Pr[y_T = a|I_a] < \frac{3}{4}$*

We prove for two cases (k is the number of actions):

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

8.6 Proof for $K = 2$

Proof: There are two actions $\{1, 2\}$. Denote by A the realizations where the algorithm predicts 1: $A = \{w \in \Omega : y_T = 1\}$. For correctness, we demand $P_1(A) = P(y_T = 1 \mid I_1) \geq \frac{3}{4}$ and $P_2(A) = P(y_T = 1 \mid I_2) \leq \frac{1}{4}$.

Then:

$$2(P_1(A) - P_2(A))^2 \leq KL(P_1||P_2) = \sum_{a \in \{1,2\}} \sum_{t=1}^{T} KL(P_1^{a,t}||P_2^{a,t})$$
$$\leq 2T \cdot 2\epsilon^2 = 4T\epsilon^2$$

By requiring $|P_1(A) - P_2(A)| \geq \frac{1}{2}$ we get:

$\frac{1}{2} \leq |P_1(A) - P_2(A)| \leq \epsilon\sqrt{2T}$

Therefore: $T = \Omega(\frac{1}{\epsilon^2})$ □

8.7 Proof for $K \geq 24$

Proof: 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 $E[A|I_0] = E_0[A]$ and $Pr[A = j|I_0] = Pr_0[A = j]$

We make the following claims:

Claim 8.6 $\exists K_1 \subseteq K : |K_1| \geq \frac{2}{3}|K| \wedge \forall j \ E_0[T_j] \leq \frac{3T}{K}$

Proof: By contradiction, assume that there is a subset $K' \subseteq K \ s.t. \ |K'| > |K|/3$ and $\forall j \in K$ we get $E_0[T_j] > 3T/K$. This implies that the arms in K' are played strictly more than T times, which is a contradiction. □

Claim 8.7 $\exists K_2 \subseteq K : |K_2| \geq \frac{2}{3}|K| \wedge \forall j \in K_2 : Pr_0[y_T = j] \leq \frac{3}{K}$

Proof: By contradiction, assume that there is a group $K' \subseteq K : |K'| > |K/3|$ arms with $\forall j \in K', Pr_0[y_T = j] > \frac{3}{K}$. This implies that the combined probability that the arms in K' are selected to be y_T is strictly greater than 1, which is a contradiction. □

Now, we use Markov's inequality and $E_0[T_j] \leq \frac{3}{K}T$ from Claim 8.6 to deduce that $Pr_0[T_j \geq \frac{24}{K}T] \leq \frac{1}{8}$. This is equivalent to $Pr_0[T_j \leq \frac{24}{K}T] \geq \frac{7}{8}$

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 Claims 8.6 and 8.7 that:

$$\forall j \in K_3 : Pr_0[T_j \leq \frac{24}{K}T] \geq \frac{7}{8} \wedge Pr_0[y_T = j] \leq \frac{3}{K}$$

We define some notations:
A probability space $\Omega_a^t = \{0,1\}^t$, and each $w \in \Omega$ in length t is a realization of $(r_s(a) : s \leq t)$.
The general space is $\Omega = \Pi_{a \in K}\Omega_a^T$
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 only $m = \frac{24T}{K}$ times.
We define for every profile I_l and $\forall A \subseteq \Omega^*$:

$$P_l^*(A) = Pr[A|I_l]$$

Where $P_l^*(A)$ is $P_l(A)$ restricted to Ω^*. Now we can compute the KL-Divergence for each A in Ω^*.
Using Pinsker's rule and the chain rule:

$$2(P_0^*(A) - P_j^*(A))^2 \leq KL(P_0^*||P_j^*) = \sum_{a \neq j}\sum_t KL(P_0^{*a,t}||P_j^{*a,t}) + \sum_{t=1}^{m} KL(P_0^{*j,t}||P_j^{*j,t}) \leq 2m\epsilon^2$$

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

$$\forall A : |P_0^*(A) - P_j^*(A)| \leq \epsilon\sqrt{m} = \epsilon\sqrt{\frac{24T}{K}} = \epsilon\sqrt{\frac{24cK}{\epsilon^2 K}} = \sqrt{24c} \leq \frac{1}{8} \qquad (8.1)$$

Notice the role of m in the simplification.

We note the bound in (8.1) holds only for events $A \subseteq \Omega^*$. Therefore, we can not use it the bound the event $A = \{y_T = j\}$ directly. To overcome this, we now define the following two events:

$$A = \{y_T = j \wedge T_j \leq m\}, A' = \{T_j > m\}$$

both of these events are in Ω^* ($A' \subseteq \Omega^*$ because whether the algorithm samples action j more than m times is completely determined by the first m coin tosses)

This implies that $P^*(A) = P(A)$ and $P^*(A') = P(A')$
Using A and A', we bound $P_j^*(A)$ as follows:

$$P_j^*(A) \leq \frac{1}{8} + P_0^*(A) \leq \frac{1}{8} + P_0^*(y_T = j) \leq \frac{1}{8} + \frac{3}{K} \leq \frac{1}{4}$$

$$P_j^*(A') \leq \frac{1}{8} + P_0^*(A') \leq \frac{1}{8} + \frac{1}{8} = \frac{1}{4}$$

We can now bound the accuracy for action j:

$$P_j(y_T = j) \leq P_j^*(y_T = j \wedge T_j \leq m) + P_j^*(T_j > m) = P_j^*(A) + P_j^*(A') \leq \frac{1}{2}$$

This is even a tighter bound than $\frac{3}{4}$ for j

□

Conclusion: For every MAB algorithm with $T \leq \frac{ck}{\epsilon^2}$, by choosing an arm j and running the profile I_j, we get: $P(y_T \neq a*) \geq \frac{1}{12}$
Which $P(y_T \neq a*)$ implies for the probability that the algorithm failed to identify the best arm. □

8.8 Proving a Lower Bound on Multi Armed Bandits Algorithms

Theorem 8.8 (MAB Lower Bound) *For every MAB algorithm which picks an action a randomly and runs the profile I_a:*

$$E[regret] = \Omega(\sqrt{TK})$$

Proof: Assume $T \leq \frac{cK}{\epsilon^2},\ \epsilon > 0$

For every round t, we refer to the action a_t as a guess for choosing the best arm. So for a random profile I_a: $P(a_t \neq a) \geq \frac{1}{12}$
For $a_t \neq a$, the loss of a_t:

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

Therefore,

$$E[\Delta(a_t)] = P[a_t \neq a]\frac{\epsilon}{2} \geq \frac{\epsilon}{24}$$

Hence,

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

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

$$E[regret] \geq \frac{T}{24}\sqrt{\frac{ck}{T}} = \frac{\sqrt{c}}{24}\sqrt{kT} = \Omega(\sqrt{kT})$$

(In order to get the tightest bound for the regret, we would like to choose ϵ as large as possible, but since $T \leq \frac{ck}{\epsilon^2}$ from *Lemma* 8.5, we get that ϵ is bounded to be $\epsilon = \sqrt{\frac{ck}{T}}$)

□