

## Lecture 4: January 29, 2024

Lecturer: Yishay Mansour

Scribe: Tomer Porian, Mikey Shechter, Rachel Mikulinsky

Based on scribe notes of Oz Granit and Ohad Rubinfeld (2021/22)

# 1 Lipschitz Bandits

A bandit model with continuous actions - CAB.

- Actions:  $X = [0, 1]$
- Rewards:  $\forall x \in X$ , where  $\mu(x) \in [0, 1]$
- Assume that the expectation of the rewards are Lipschitz:

$$\forall x, y \in X \quad |\mu(x) - \mu(y)| \leq L|x - y|$$

The Lipschitz constant  $L$  is known to the algorithm.

The profile of the problem is defined by  $(T, L, \mu(\cdot))$ .

**Question:** How can we use an algorithm that assumes a finite number of actions?

*Simple solution:* Constant Discretization.

Choose a finite  $S \subseteq X$ . Run a MAB algorithm on  $S$ : UCB, Successive Arm Elimination. Our source of error is double:

1. From  $S$  (Like regular MAB).
2. From not using the entire action space  $X$ , but only using  $S$ .

Of course we can increase the size of  $S$ . But this increases the error from (1) and decreases the error from (2).

The error from discretization:

$$\begin{aligned} \mu^*(X) &= \sup_{x \in X} \mu(x) \\ \mu^*(S) &= \max_{x \in S} \mu(x) \\ \text{DE}(S) &= \mu^*(X) - \mu^*(S). \end{aligned}$$

## 1.1 Regret analysis

Let  $W(\text{ALG})$  be the reward expectation of the algorithm and  $R_S(T)$  the regret expectation on  $S$  in  $T$  steps.

$$\begin{aligned} \mathbb{E}[\text{regret}] &= \mu^*(X)T - W(\text{ALG}) \\ &= [\mu^*(S)T - W(\text{ALG})] - (\mu^*(X) - \mu^*(S))T \\ &= R_S(T) + T \cdot \text{DE}(S) \end{aligned}$$

There exists algorithms such that  $R_S(T) = O(\sqrt{|S|T \log T})$

Using a uniform discretization we have

$$\varepsilon = \frac{1}{k+1}, \quad S = \{y_i = \varepsilon \cdot i\}, \quad \text{DE}(S) = \varepsilon L$$

and we can now derive the following regret bound:

**Theorem 1** For  $\varepsilon = \left(\frac{\log T}{TL^2}\right)^{\frac{1}{3}}$  we get  $\mathbb{E}[\text{regret}] = O\left(L^{1/3}T^{2/3} \log^{1/3} T\right)$

## 1.2 Lower Bound

We show a lower bound  $\Omega(T^{\frac{2}{3}})$  on the regret. We will use a reduction to the MAB setting of lecture 2.

The previous lower bound used profiles  $I_j$  for each action  $j$ , where

$$I_j(i) = \begin{cases} Br(1/2) & i \neq j \\ Br(1/2 + \varepsilon) & i = j \end{cases}.$$

For the continuous case, our profile will assign, for every action in  $x^* \in X$ , the following function:

$$\mu(x) = \begin{cases} 1/2 & |x - x^*| \geq \varepsilon/L \\ 1/2 + \varepsilon - L|x - x^*| & \text{otherwise} \end{cases},$$

as illustrated in fig. 1.

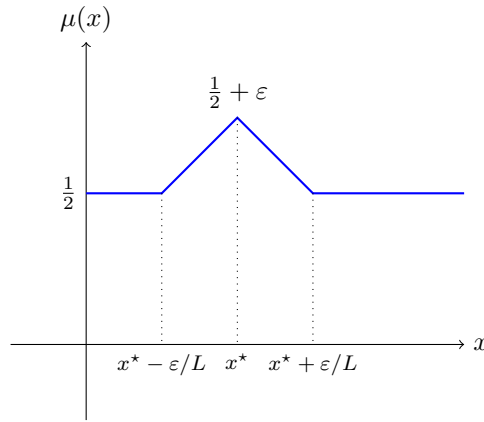


Figure 1: Example of a function in profile  $I(x^*, \varepsilon)$

*Important:* Note that  $\mu$  is  $L$ -Lipschitz and  $\frac{1}{2} \leq \mu(x) \leq \frac{1}{2} + \varepsilon$ .

**Theorem 2** For every CAB algorithm, there exists a profile  $I = I(x^*, \varepsilon)$  such that  $\mathbb{E}[\text{regret} \mid I] = \Omega(L^{1/3}T^{2/3})$

**Proof sketch:** We will partition  $X = [0, 1]$  into  $k$  equal parts. Set  $\varepsilon = \frac{1}{2k}$ , and  $\mathbb{K} = \{x_i = 2\varepsilon i\}$ . For every interval  $[x_i, x_{i+1}]$  we have  $x_i^* = x_i + \varepsilon$ . We will use the profiles  $I(x_i^*, \varepsilon)$ . In general, every time the algorithm plays  $a \in [x_i, x_{i+1}]$ , it should play  $x_i^*$  instead. If the algorithm only plays  $x_i^*$  then it is a MAB with  $k$  actions.

This is exactly the set of profiles for the lower bound from lecture 2. Recall that for every  $\varepsilon \leq \sqrt{c \frac{k}{T}}$  in discrete MAB there is a profile  $I$  such that  $\mathbb{E}[\text{regret}] = \Omega(\varepsilon T) = \Omega(\sqrt{kT})$ .

**Proof:** We will build a MAB algorithm for  $\mathbb{K}$  using the CAB algorithm. When the CAB algorithm gives us a point  $x$  to play, we select a point  $z(x)$  to play. Given the reward for  $z(x)$  denoted  $r(z(x))$  we need to return to MAB  $r_x$ .

We need to define how we choose  $z(x)$  given  $x$  and  $r_x$  given  $r(z(x))$ .

- Choosing  $z(x)$ : if  $x \in [2i\varepsilon, 2(i+1)\varepsilon]$  then  $z(x) = (2i+1)\varepsilon = x_i^*$
- Return  $r_x$ : We need to turn the profit of  $z(x)$  that is  $r \in \{0, 1\}$  into  $r_x \in \{0, 1\}$ . The problem is that we do not know the true  $\mu(x)$ . The solution: with probability  $p_x$  we return  $r$  and otherwise we flip a fair coin and return it. Namely,

$$r_x = \begin{cases} r & \text{w.p } p_x \\ Br(\frac{1}{2}) & \text{otherwise} \end{cases}.$$

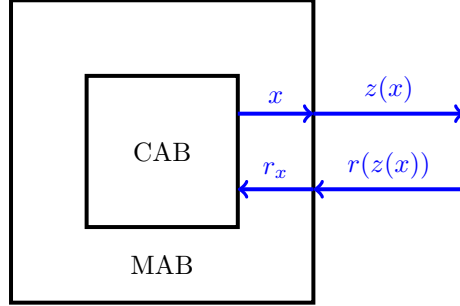


Figure 2: CAB reduction to MAB

The expectation of  $r_x$  is

$$\begin{aligned} \mathbb{E}[r_x | x] &= p_x \mu(z(x)) + (1 - p_x) \frac{1}{2} \\ &= \frac{1}{2} + \left( \mu(z(x)) - \frac{1}{2} \right) p_x \\ &= \begin{cases} 1/2 & z(x) \neq x^* \\ 1/2 + \varepsilon p_x & z(x) = x^* \end{cases} \end{aligned}$$

so in order to get  $\mathbb{E}[r_x | x] = \mu(x)$  we need to choose

$$\varepsilon p_x = \varepsilon - |x - z(x)|L \implies p_x = 1 - \frac{|x - z(x)|L}{\varepsilon}.$$

Let  $x_t$  be the action of the CAB in time  $t$  and  $a_t = z(x_t)$  the action of the MAB in time  $t$ . From the definition of our profiles, we have that  $\mu(x_t) \leq \mu(a_t)$ . This implies that,

$$\mathbb{E}[\text{regret-CAB}] \geq \mathbb{E}[\text{regret-MAB}] \geq \Omega(\sqrt{kT}).$$

We can set

$$\varepsilon = T^{-\frac{1}{3}}, \quad \varepsilon \leq \sqrt{\frac{k}{cT}}, \quad k = \frac{1}{2\varepsilon} = \frac{1}{2}T^{1/3}$$

which gives a lower bound

$$\mathbb{E}[\text{regret-CAB}] = \Omega\left(T^{2/3}\right).$$

□

### 1.3 Lipschitz MAB

We extend from the single dimensional case  $X = [0, 1]$  into a general set and metric. For a metric  $D$  the reward  $\mu$  will hold:

$$\forall x, y \in X \quad |\mu(x) - \mu(y)| \leq LD(x, y)$$

For simplicity assume that  $L = 1$  and that  $D(x, y) \leq 1$ . A Metric spaces  $D : X \times X \rightarrow \mathbb{R}$  has the following properties

1. Non negativity:  $D(x, y) \geq 0$
2. Distance of zero implies Equality :  $x = y \iff D(x, y) = 0$
3. Symmetry:  $(x, y) = D(y, x)$
4. Triangle inequality:  $D(x, z) \leq D(x, y) + D(y, z)$

Examples for metric spaces:

- $X = [0, 1]^d$  with the  $p$ -norm  $\ell_p(x, y) = \|x \cdot y\|_p = \left( \sum_{i=1}^d (x_i \cdot y_i)^p \right)^{\frac{1}{p}}$
- $D(x, y) =$  length of the shortest path from  $x$  to  $y$  where  $X = \{\text{vertices}\}$

Discretization: like before we want that  $\text{DE}(S) \leq \epsilon$  and that  $|S|$  will be small.

Example:  $X = [0, 1]^d$  and  $\ell_p$  norm with  $p \geq 1$ . For a uniform discretization we have:  $|S| = \left\lceil \frac{1}{\epsilon} \right\rceil^d$  and  $\text{DE}(S) = c_{p,d}\epsilon$ . Where  $c_{p,d}$  is dependent on  $p$  and  $d$ . For  $p = 1$  we have  $c_{1,d} = 1$ . And for  $\epsilon = \left( \frac{\log T}{T} \right)^{\frac{1}{d+2}}$  we get:

$$\mathbb{E}[\text{regret}] = O\left(\sqrt{\left(\frac{1}{\epsilon}\right)^d T \log T + \epsilon T}\right) = O\left(T^{\frac{d+1}{d+2}} \log^{\frac{1}{d+2}} T\right)$$

### 1.3.1 General spaces

1.  $\epsilon$ -mesh: A set  $S \subseteq X$  is a  $\epsilon$ -mesh if  $\forall x \in X \exists y \in S : D(x, y) \leq \epsilon$
2. Diameter:  $\text{diam}(X) = \sup_{x, y \in X} D(x, y)$
3.  $\epsilon$ -covering: A set of subsets  $X_i \subseteq X$  are a  $\epsilon$ -covering if  $\bigcup_i X_i = X$  and  $\text{diam}(X_i) \leq \epsilon$
4. Covering number: The size of the smallest  $\epsilon$ -covering. We denote it by  $N_\epsilon(X)$ .

If  $\{X_1 \dots X_N\}$  is a  $\epsilon$ -cover and  $x_i \in X_i$  then  $\{x_1 \dots x_N\}$  is a  $\epsilon$ -mesh.  
The covering dimension of  $X$  with a multiplier  $c > 0$  is

$$\text{cov}_c(X) = \inf_{d \geq 0} \{N_\epsilon(X) \leq c \cdot \epsilon^{-d} \quad \forall \epsilon > 0\}$$

We can now derive regret in general spaces. Let  $d = \text{cov}_c(X)$ , there exists a set  $S$  that is an  $\epsilon$ -mesh for  $X$  where  $|S| = N_\epsilon(X) \leq c/\epsilon^d$ .

#### Theorem 3

$$\mathbb{E}[\text{regret}] = O\left(T^{\frac{d+1}{d+2}} (c \log T)^{\frac{1}{d+2}}\right)$$

We can also extend the lower bound and have

**Theorem 4** For  $([0, 1]^d, \ell_2)$ , for any algorithm:

$$\mathbb{E}[\text{regret}] = \Omega\left(T^{\frac{d+1}{d+2}}\right)$$

## 2 Zoom Algorithm

The zoom algorithm uses adaptive discretization.

$$\text{DE}(S) \leq D(S, x^*) = \min_{x \in S} D(X, x),$$

where  $x^* \in \arg \min_{x \in S} D(S, x)$ .

We want to make  $S$  small while keeping  $\text{DE}(S)$  below  $D(S, x^*)$ . We want to discretize more areas with high reward, but clearly we cannot avoid the lower bound. However, if our profile is simple we will get improved performance.

The zoom algorithm includes:

- Techniques from the UCB algorithm.
- Adaptive discretization.

- The definition of the good event.

The Zoom algorithm keeps a set  $S \subseteq X$  of active actions at every time step:

1. There are actions that become active (i.e., they are added to  $S$ ), and we do not remove actions from  $S$ .
2. We have a “selection rule” and according to it we chose which actions in  $S$  we select.

We must define:

1. When will an action be added to  $S$ .
2. Which action from  $S$  we will pick.

*Assumption:*  $\mu(\cdot)$  is 1-Lipschitz.

**Definition** Confidence radius/ball: For a time  $t$  and an active action  $x \in S$ , Let  $n_t(x)$  be the number of times we played action  $x$  (before time  $t$ ). Let  $\bar{\mu}_t(x)$  be the average reward expectation for action  $x$ , up to time  $t$ . The radius is:

$$r_t(x) = 2\sqrt{\frac{2 \log T}{n_t(x) + 1}}$$

This assures that with high probability we have  $|\mu(x) - \bar{\mu}_t(x)| \leq r_t(x)$ . The confidence ball of  $x$  at time  $t$  (assuming  $D(x, y) \leq 1$ ) is

$$B_t(x) = \{y \in X : D(x, y) \leq r_t(x)\}$$

## 2.1 Activation Rule

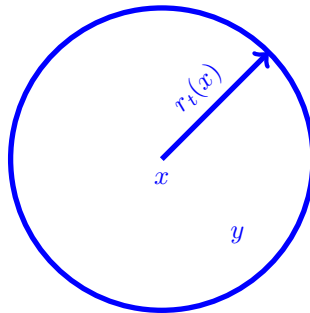


Figure 3: Illustration of  $B_t(x)$ . Note that in a general space the ball will not necessary look like a ball

*Motivation:* If  $y \in B_t(x)$  then,

$$\begin{aligned} D(x, y) &\leq r_t(x) \\ \mu(y) &\leq \mu(x) + r_t(x) \end{aligned}$$

This implies that if the difference between  $\mu(x)$  and  $\mu(y)$  is small.

*Invariant:* At each step we have:  $\bigcup_{x \in S} B_t(x) = X$ .

When the algorithm plays  $x \in X$  then the ball for  $x$  shrinks. If there exists some  $y \in X$  where  $y \notin \bigcup_{x \in S} B_t(x)$ , we will add  $y$  into  $S$ .

Note: If  $n_t(y) = 0$  then  $B_t(y) = X$  so we will add at most one element into  $S$ .

## 2.2 Selection Rule

Similarly to UCB we define:

$$\begin{aligned} \text{index}_t(x) &= \bar{\mu}_t + 2r_t(x) \\ a_t &= \underset{x \in \mathcal{S}}{\text{argmax}} \text{index}(x) \end{aligned}$$

Note: if  $x$  was never played, define  $\bar{\mu}_t(x) = 0$

## 2.3 Zoom Algorithm

---

### Algorithm 1: Zoom

---

```

1 Initialization:  $S \leftarrow \emptyset$ 
2 for  $t = 1, \dots$  do
3   if  $\exists y \in X$  s.t.  $\forall x \in S : y \notin B_t(x)$  then
4      $S \leftarrow S \cup \{y\}$ 
5   end
6   Play  $a_t = \underset{x \in \mathcal{S}}{\text{argmax}} \text{index}(x)$ 
7 end
```

---

Some intuition: when  $y$  is added, it has a big  $r_t$  and therefore probably will be played, but  $B_t(y)$  covers large area so no new actions are added for a while, until  $r_t(y)$  shrinks.

## 2.4 Analysis: The good event

We would like to show:

$$\forall x \in X \quad |\mu(x) - \bar{\mu}_t(x)| \leq r_t(x)$$

The problem:  $X$  is infinite. Suppose we sampled all the rewards in advance. For each  $x \in X$  the good event is,

$$\mathcal{E}_x = \{|\mu(x) - \bar{\mu}_t(x)| \leq r_t(x) \quad \forall t \leq T\}$$

The good event  $\mathcal{E}$  is,

$$\mathcal{E} = \bigcap_{x \in X} \mathcal{E}_x$$

**Claim 5** Assume the rewards are  $\{0, 1\}$ , then

$$\Pr[\mathcal{E}] \geq 1 - 1/T^2$$

Comment: For  $x \notin S$  ( $n(x) = 0$ ) then  $\mathcal{E}_x$  holds because  $1 \leq r_t(x)$ .

**Proof:** By Hoeffding: for any  $x \in X$  and  $t \in [T]$  we have

$$\Pr[|\mu(x) - \bar{\mu}_t(x)| \leq r_t(x)] \geq 1 - 1/T^4.$$

Thus, applying union bound over the time steps gets us to

$$\forall x \in X \quad \Pr[\mathcal{E}_x] \geq 1 - 1/T^3$$

The problem is performing union bound over  $X$  - the set  $X$  is too large.

Fix a profile of Lipschitz MAB. Let  $X_0$  be the set of all actions that can enter  $S$ . This set is finite (!), since the algorithm is deterministic and there are only two possible rewards  $\{0, 1\}$  (this is the only place we use the fact that the rewards are a small finite set)

Let  $N$  be the number of active actions (size  $S$  at the end of run)

Let  $y_j$  be the  $j$ -th action to become active (=entered  $S$ ). Both  $N$  and  $y_j$  are random variables.

We complete  $y_j$  to  $y_T$  by setting  $y_j = y_N$  for  $j > N$

The good event is  $\mathcal{E} = \bigcap_{j \in [T]} \mathcal{E}_{y_j}$

We'll show that  $\mathcal{E}_{y_j}$  occurs in high probability.

We set  $x \in X_0$  and  $j \in [T]$ , the event  $\{y_j = x\}$  is determined by the rewards of the other actions (as up until now  $x$  was inactive).

The event  $\mathcal{E}_x$  is determined only by the rewards of  $x$ . If  $Pr[y_j = x] > 0$  then

$$Pr[\mathcal{E}_{y_j} | y_j = x] = Pr[\mathcal{E}_x | y_j = x] = Pr[\mathcal{E}_x] \geq 1 - 1/T^3$$

Summing over  $x \in X_0$

$$Pr[\mathcal{E}_{y_j}] = \sum_{x \in X_0} Pr[y_j = x] \cdot Pr[\mathcal{E}_{y_j} | y_j = x] \geq (1 - 1/T^3) \cdot \sum_{x \in X_0} Pr[y_j = x] = 1 - 1/T^3$$

Union bound over the time  $j \in [T]$

$$Pr[\mathcal{E}_{y_j}, j \in [T]] \geq 1 - 1/T^2$$

□

From now on we assume the good event occurs.

## 2.5 Analysis: Bad Actions

An action is bad if its reward is significantly lower than  $\mu^*$ . We will show for active bad actions that:

- They are far away from each other
- They are played a small number of times

Let,

$$\mu^* = \sup_{x \in X} \mu(x)$$

$$\Delta(x) = \mu^* - \mu(x)$$

$$n(x) = n_{T+1}(x)$$

**Lemma 6**  $\forall x \in X, \forall t \in [T], \Delta(x) \leq 3r_t(x)$

**Proof:** Assume that at time  $t$  we played  $x$ . There exists  $y \in S$  such that  $x^* \in B_t(y)$

$$\text{index}(x) \geq \text{index}(y) = \underbrace{\bar{\mu}_t(y) + r_t(y)}_{\geq \mu(y)} + r_t(y) \underbrace{\geq}_{\text{Lipshitz}} \mu(x^*) = \mu^*$$

(this happens with high probability, as we assumed before) For  $x$ :

$$\text{index}(x) = \underbrace{\bar{\mu}_t(x)}_{\mu(x) + r_t(x)} + 2r_t(x) \leq \mu(x) + 3r_t(x)$$

Hence:

$$\mu(x) + 3r_t(x) \geq \mu^* \longrightarrow 3r_t(x) \geq \Delta(x)$$

If we didn't play  $x$  at time  $t$ :

- If we've never played  $x$ , then  $r_t(x) > 1$ .
- If we've played  $x$  at time  $\tau$  last, then  $r_t(x) = r_\tau(x) \geq \frac{\Delta(x)}{3}$ .

□

## 2.6 Corollaries

1. **Corollary 7** For every two active actions  $x, y$  then  $D(x, y) > \frac{1}{3} \min \{\Delta(x), \Delta(y)\}$

Assume  $x$  became active first, at time  $\tau$  action  $y$  became active, then  $-D(x, y) > r_\tau(x)$  (since  $y$  became active),  $r_t(x) \geq \frac{\Delta(x)}{3}$

2. **Corollary 8**  $\forall x \in X, n(x) \leq \frac{O(\log T)}{\Delta^2(x)}$

$$\Delta(x) \leq 3r_{T+1}(x) = 3\sqrt{\frac{2\log T}{n(x)}}$$

$$n(x) \leq \frac{18\log T}{\Delta^2(x)}$$

## 2.7 Analysis: Covering number

We partition the actions using  $\Delta(x)$

$$\forall r > 0, X_r = \{x \in X : r \leq \Delta(x) < 2r\}$$

We define  $Y_i = X_r$  for  $r = 2^{-i}$  ( $\bigcup_{i \geq 0} Y_i = X$ )

Let  $Z_i \subseteq Y_i$  be the actions that became active.  $D(x, y) \geq r/3 : x, y \in Z_i$

Say we've added  $x$  before  $y$ , at the time of adding  $y$ :  $\Delta(x, y) > r_t(x) \geq \frac{\Delta(x)}{3} \geq \frac{r}{3}$ , if we cover  $Y_i$  with groups of diameter  $r/3$  then  $x, y$  are not at the same group,  $Y_i$  is cover-able by  $N_{r/3}(Y_i)$  therefore  $|Z_i| \leq N_{r/3}(Y_i)$ .

## 2.8 Analysis: Regret

$$R_i(T) = \sum_{x \in Z_i} \Delta(x)n(x) \leq \frac{O(\log T)}{\Delta(x)} N_{r/3}(Y_i) = O\left(\frac{\log T}{r} N_{r/3}(Y_i)\right)$$

for  $\delta > 0$  we'll separate to  $\Delta(\cdot) \leq \delta$  and  $\Delta(\cdot) > \delta$

$$R(T) \leq \delta \cdot T + \sum_{i: 2^{-i} > \delta} R_i(T) \leq \delta \cdot T + \sum_{i: 2^{-i} = r > \delta} \frac{O(\log T)}{r} N_{r/3}(Y_i)$$

$$\leq \delta \cdot T + O(c \cdot \log T) \left(\frac{1}{\delta}\right)^{d+1}$$

when  $\forall r > 0$   $N_{r/3}(X_r) \leq c \cdot r^{-d}$   
*covering number*

## 2.9 Zoom dimension

$$\inf_{d > 0} \{N_{r/3}(X_r) \leq c \cdot r^{-d}, \forall r > 0\}$$

**Theorem 9**  $\mathbb{E}[\text{regret}] = O\left(T^{\frac{d+1}{d+2}} (c \log T)^{\frac{1}{d+2}}\right)$

The theorem follows by  $\delta = \left(\frac{\log T}{T}\right)^{\frac{1}{d+2}}$

Note: this is mostly interesting for space with some constraints. For euclidean space we get this bound with uniform discretization as we saw at the beginning of the lecture.

Zoom vs Covering dim:

Covering - a property of the metric space.

Zoom - a property of the profile  $(X_r)$

### 3 Product Pricing

In the product pricing model, every buyer  $t$  has a value  $v_t$  which is sampled from the distribution  $D$ . The buyer is offered a price  $p_t$ , if  $p_t \leq v_t$  there is a sale, and the seller gains  $p_t$ . Otherwise there is no sale, and the seller gains zero reward.

The Goal is to maximize the revenue function which is just the summation of the rewards (there is no

production cost):  $\sum_{t=1}^T p_t \cdot \mathbb{I}(p_t \leq v_t)$

We will assume  $D$  is defined over  $[0, 1]$ .

Notice that the function is not Lipschitz because it's not even continuous.

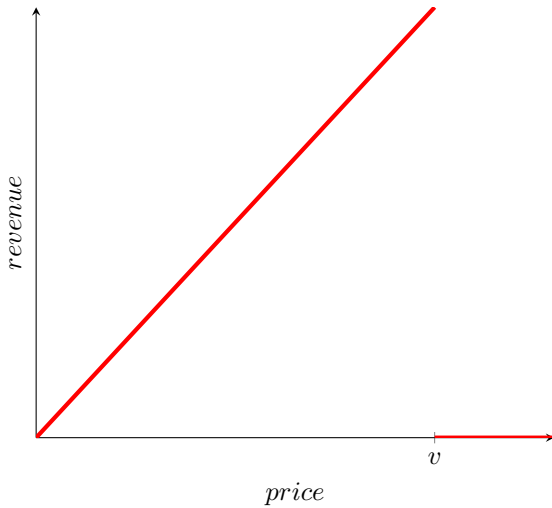
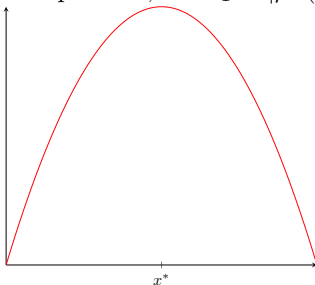


Figure 4: The revenue function  $x \cdot \mathbb{I}(x \leq v)$

We would like to find the optimal price given  $D$ . Meaning, we would like to find  $x$  that maximizes the expectation of the revenue at price  $x$ :

$$\mu(x) = x \cdot \Pr_{v \sim D}[x \leq v]$$

Assume  $\mu$  is Concave which is common practice in economics, i.e.,  $\mu''(x) < 0$ . Also assume it is bound and positive, i.e.  $c_1 < |\mu''(x)| < c_2$



### 3.1 MAB approach

We will perform uniform discretization to  $k$  prices  $\{\frac{1}{k}, \frac{2}{k}, \dots, 1 - \frac{1}{k}, 1\}$  and run a MAB algorithm (say UCB or Succ-Arm Elimination) with  $k$  actions. We would like to find out what will be the regret, and how to choose  $k$ .

Now we will see how this setting translates to a discrete MAB setting:

We will have  $k$  actions  $\frac{i}{k}$  for  $i \in \{1, \dots, k\}$ . Each action will have the expectation:

$$\mu_i = \mu\left(\frac{i}{k}\right) = \frac{i}{k} \cdot \Pr_{v \sim D}\left[\frac{i}{k} \leq v\right]$$

The optimal action will have expectation:  $\mu^* = \max_i \mu_i$ . And for every action  $i$ :  $\Delta_i = \mu^* - \mu_i$ .

We would like to analyze the revenue of UCB:  $Rev^{UCB} = \text{revenue of UCB}$ .

The idea:

1. Bound the regret when only using the  $k$  action  $\mu^*T - Rev^{UCB}$
2. Bound the cost of discretization  $(\mu(x^*) - \mu^*)T$

Overall  $\mathbb{E}[\text{regret}] = 1. + 2.$

**Lemma 10**  $c_1(x^* - x)^2 < \mu(x^*) - \mu(x) < c_2(x^* - x)^2$ .

**Proof:** From the Taylor series, there exists  $z \in [x, x^*]$  such that

$$\mu(x) = \mu(x^*) + (x - x^*)\mu'(x^*) + \frac{(x - x^*)^2}{2}\mu''(z)$$

since  $x^*$  is the maximum of  $\mu$  then  $\mu'(x^*) = 0$ , and we assumed  $c_1 < |\mu''(z)| < c_2$  □

### 3.2 Corollaries

1. **Corollary 11**  $\Delta_i \geq c_1(x^* - \frac{i}{k})^2$
2. **Corollary 12**  $\mu^* > \mu(x^*) - \frac{c_2}{k^2}$  because there exists  $\frac{i}{k}$  such that  $\frac{1}{k} > |x^* - i/k|$

### 3.3 Analysis of the regret

For the UCB algorithm

$$\begin{aligned} \mu^*T - Rev^{UCB} &\leq O(\log T) \sum_{i: \mu_i < \mu^*} \frac{1}{\Delta_i} \\ &\leq O(\log T) \frac{4k^2}{c_1} \sum_{i=1}^{\infty} i^{-2} = O\left(\frac{k^2}{c_1} \log T\right). \end{aligned}$$

For the discretization,

$$\begin{aligned} (\mu(x^*) - \mu^*)T &\leq \frac{c_2}{k^2}T \\ \mu(x^*)T - Rev^{UCB} &= O\left(\frac{k^2}{c_1} \log T + \frac{c_2}{k^2}T\right) = O(k^2 \log T + \frac{T}{k^2}) \end{aligned}$$

Choosing  $k$ :

$$k = \left(\frac{T}{\log T}\right)^{1/4}$$

**Theorem 13**  $\mathbb{E}[\text{regret}] = O(\sqrt{T \log T})$ .

Notes:

- $\mathbb{E}[\text{regret}] = \Omega(\sqrt{T})$
- We got this result because we assumed the function is concave.

## 4 References

1. Aleksandrs Slivkins. “Introduction to Multi-Armed Bandits”. chapter 4.
2. Product Pricing: Kleinberg, Leighton, 2003.