Linear Multi Armed Bandit With an Eavesdropper

Dean Elimelech
Network Engineering Student
Ben Gurion University
deaneli13@gmail.com

Asaf Cohen
Electrical Engineering Professor
Ben Gurion University
coasaf@gmail.com

Daniel Blozrov
Network Engineering Student
Ben Gurion University
danny8196@gmail.com

***Abstract*—The subject of Multi Armed Bandits is becoming a hot topic in machine learning, and with it there are many problems and algorithms.A rather unique problem is the presence of an eavesdropper in the channel, who attempts to gather information which can disrupt the learning process's security. This paper aims to tackle this problem,using information theory tools to deny the eavesdropper from gathering said information using frequency analysis.We present a probabilistic proof that shows the eavesdropper's probability to select the optimal arm is low.**

I. Introduction

The Multi Armed Bandit problem is a problem that consists of a bandit pulling arms, where a bandit is an entity that could be a person,or a machine and arms are choices that are available. [1] For example, a gambler in a casino may pull levers , the gambler would be the bandit and the levers would be the arms. The arms have a reward for pulling them, which is not known apriori to the bandit. The problem we aim to tackle is how to choose the optimal arm, using a fixed number of steps,while denying an observer who can see which arms we pull,from knowing which arm is the optimal. There are many uses to the multi armed bandit problem, for example figuring out a customer profile for advertisement services from a vendor's perspective. Our focus will be shifted to fixed-budget best arm selection, meaning the amount of arm pulls allowed will be bounded by some constant, during which we need to find out the optimal arm with the best reward, while not letting the eavesdropper, who can see which arms we pull, figure out the best arm.

A. Main contributions:

- We will provide an algorithm using pseudocode, which is not hard to implement in a system.
- There exist a analytic proof that gives bounds on the error for both our bandit and the eavesdropper.

B. Related works and comparisons:

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magnam aliquam quaerat voluptatem. Ut enim aeque doleamus animo, cum corpore dolemus, fieri tamen permagna accessio potest, si aliquod aeternum et infinitum impendere malum nobis opinemur. Quod idem licet transferre in voluptatem, ut.

II. Model Setup

We will denounce the maximal number of steps or arm pulls as T and the set of arms $\mathcal{A}$. We will focus on the stochastic linear case of multi armed bandits. Assume there are k arms, denoted by $A_0, A_1 \ldots A_{k-1}$ where each arm A_i is a vector $A_i \in \mathbb{R}^d$.

Assume there is another vector, $\theta^* \in \mathbb{R}^d$,which is unknown to everyone, and a gaussian random process $\eta(t)$ where $\eta(t)$ is a gaussian random variable with mean $\mu_\eta = 0$ and Var(η) = 1 Now assume a reward function, based on the arm chosen at time t,denoted as $A(t)$:

$$X(t) = \langle \theta^*, A(t) \rangle + \eta \tag{1}$$

The reward function $X(t)$ as a function of the inner product of arm the arm chosen at time t and θ^*,in addition to white gaussian noise η.

The same setting can be found in [2] ,which will form the basis of our paper. We assume that there is a unique arm A^* which will be our optimal arm in terms of the reward, assuming that there was no noise, I.E :

$$i^* = \text{argmax}_{i \in \mathcal{K}} \mathbb{E}[X(i)] \tag{2}$$

Where $\mathcal{K}$ means $\{0, 1, 2, \ldots k-1\}$. We will assume that the set of our vectors $\{A_0, A_1, \ldots A_{k-1}\}$ spans $\mathbb{R}^d$. We will use a modified version to the approach suggested in [2] using the G-optimal design. At each iteration, we will approximate θ^* and try using inner products with the estimated $\hat{\theta}$ and that will save us arm pulls. Let $(A(1), A(2) \ldots A(n))$ be a sequence of arms pulled sequentially and let :

$$\hat{\theta} = V^{-1} \sum_{i=1}^{n} A(i)X(i) \tag{3}$$

$$V = \sum_{i=1}^{n} A(i)A(i)^T \tag{4}$$

Notice that V is positive definite and invertible as it is an empirical covariance matrix. where Equation 3 represents the optimal ordinary least squares estimator for θ^*.

G optimal Design - Let $\pi \in \mathbb{R}^k$ be a vector of probabilities such that :

$$0 \leq \pi(i) \leq 1 \tag{5}$$

$$\sum_{i=1}^{k} \pi(i) = 1 \tag{6}$$

We aim to find a distribution π such that it minimizes :

$$g(\pi) = \max_{i \in \mathcal{K}} \|A(i)\|_{V(\pi)^{-1}} \tag{7}$$

$$V(\pi) = \sum_{i \in \mathcal{K}}^{n} \pi(i) A(i) A(i)^T \tag{8}$$

There are efficient solutions to this convex optimization problem,however for our paper, using an approximation algorithm given by [3] will suffice.

In a setting without an eavesdropper, the **OD-LINBAI** algorithm presented in [2] provides optimal solutions, however in the presence of an adversary, frequency analysis can be used to detect the best arms, and because the algorithm prunes the set of arms such that

$$|\mathcal{A}_{r-1}| = \left\lceil \frac{1}{2} \, |\mathcal{A}_r| \right\rceil \tag{9}$$

This creates a flaw that allows the observer to notice that the best arm will necessarily be one of the 2 arms in the final round. Instead, let us define linear combinations

$$v_1 = A_1 + A_2 + A_3 + ...A_s \tag{10}$$

$$v_2 = A_2 + A_3 + ...A_{s-1} \tag{11}$$

Using the linearity of (1) we can see that

$$X(v_1) = \langle A_1 + A_2 + A_3 + ...A_s, \theta^* \rangle + \eta_1 \tag{12}$$

$$X(v_2) = \langle A_2 + A_3 + ...A_s, \theta^* \rangle + \eta_2 \tag{13}$$

subtracting the equations we can see that :

$$X(A_1) \approx X(v_1) - X(v_2) + \eta_2 - \eta_1 \tag{14}$$

The observer saw that we pulled s arms once, and then $s-1$ arms once, however if we repeat this trial many times, the frequency should look uniform, if we picked good linear combinations. We need to take into account that $\eta_2 - \eta_1 \sim \mathcal{N}(0,2)$ due to the linear combination of independent gaussian random variables, so repeating it many times costs us in higher noise. After finding π we can estimate a good number of arm pulls per arm every round, using more pulls for better arms, for exploitation and minimizing the noise's effect. We will annotate the number of arm i pulls in round r as $T_r(i)$. Generalizing the above idea in (10) will allow us to create random linear combinations. Let $p_i \in \mathbb{R}^K$ be a binary vector such that $p_{ij} \in \{0,1\}$ that are chosen randomly in a uniform matter, $i \in \{0, 1, ...T_r(w) - 1\}$ where w presents the wth arm.

$$y_i = \left\langle \sum_{j=1}^{k} p_{ij} A_j, \theta^* \right\rangle + \eta_i \tag{15}$$

$$y = Px + \eta \tag{16}$$

$$x = (\langle A_1, \theta^* \rangle, \langle A_2, \theta^* \rangle, ...\langle A_k, \theta^* \rangle)^T \tag{17}$$

We have arrived at a linear equation system for which we can solve for x using the least squares method :

$$\hat{x} = \text{argmin}_x \; \| Rx - y \; \|_2^2 \tag{18}$$

$$\hat{x} = (P^T P)^{-1} P^T y \tag{19}$$

III. Algorithm

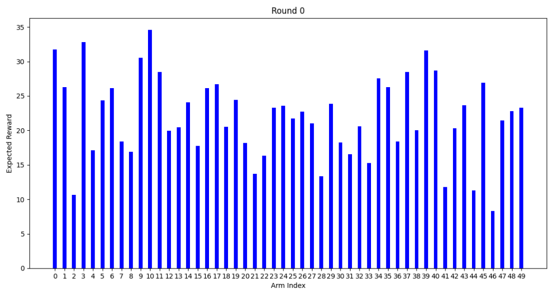
let arms be the set of arms as input, such that arms is a matrix of k vectors each of dimension $\mathbb{R}^d$, and $T \in \mathbb{N}$ the time budget for which we must declare a chosen arm in time $t \leq T$.

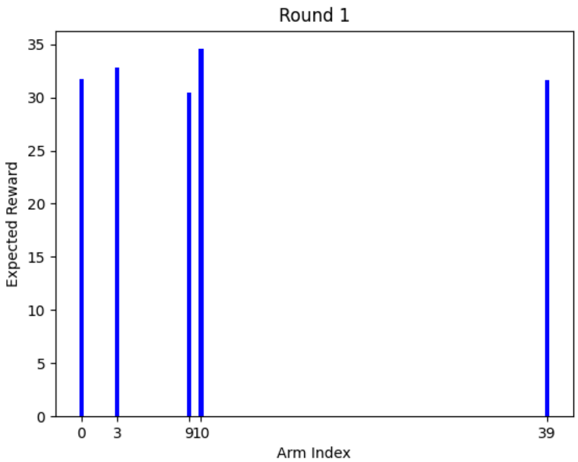
Modified OD-LINBAI (arms,T):

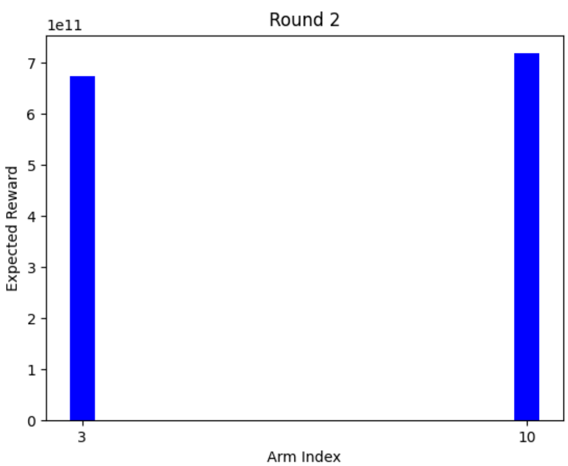
calculate m using (19)
let curr_indexes = $[0, 1, 2, ...k-1]$
for $r = 1$ up **to** $r = \lceil \log_2(d) \rceil$
    find G-optimal design π_r
    set $T_r(i)$ according **to** (19).
    for all arms A_i **if** $i \in$ curr_indexes**:**
        let P be a matrix of coefficients of indexes $\notin$ curr_indexes
        calculate $\mathbb{E}[X(i)]$ using (18)
        set curr_indexes as best $\frac{d}{2^r}$ $\mathbb{E}[X(i)]$ **if** $i \in$ curr_indexes
return arms(only index $\in$ curr_indexes)

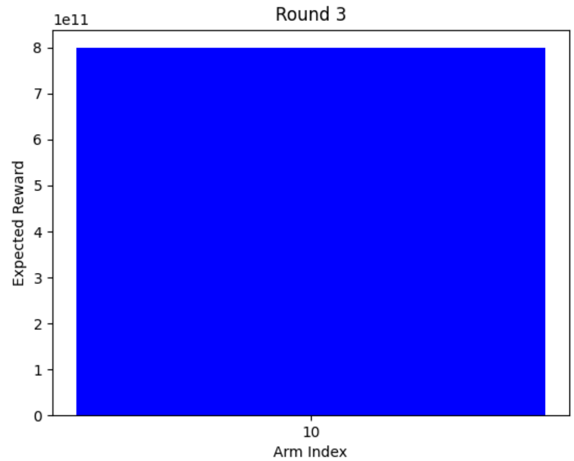
IV. Results

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magnam aliquam quaerat voluptatem. Ut enim aeque doleamus animo, cum corpore dolemus, fieri tamen permagna accessio potest, si aliquod aeternum et infinitum impendere malum nobis opinemur. Quod idem licet transferre in voluptatem, ut postea variari voluptas distinguique possit, augeri amplificarique non possit. At etiam Athenis, ut e patre audiebam facete et urbane Stoicos irridente, statua est in quo a nobis philosophia defensa et collaudata est, cum id, quod maxime placeat, facere possimus, omnis voluptas assumenda est, omnis dolor repellendus. Temporibus autem quibusdam et.









We can see the pruning of the set of arms by half every round, until we only have the optimal arm. Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magnam aliquam quaerat voluptatem. Ut enim aeque doleamus animo, cum corpore dolemus, fieri. In here, we can see the eavesdropper's final histogram of frequencies each arm was pulled.

V. Conclusions

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magnam aliquam quaerat voluptatem. Ut enim aeque doleamus animo, cum corpore dolemus, fieri tamen permagna accessio potest, si aliquod aeternum et infinitum impendere malum nobis opinemur. Quod idem licet transferre in voluptatem, ut postea variari voluptas distinguique possit, augeri amplificarique non possit. At etiam Athenis, ut e patre audiebam facete et urbane Stoicos irridente, statua est in quo a nobis philosophia defensa et collaudata est, cum id, quod maxime placeat, facere possimus, omnis voluptas assumenda est, omnis dolor repellendus. Temporibus autem quibusdam et aut officiis debitis aut rerum necessitatibus saepe eveniet, ut et voluptates repudiandae sint et molestiae non recusandae. Itaque earum rerum defuturum, quas natura non depravata desiderat. Et quem ad me accedis, saluto: 'chaere,' inquam, 'Tite!' lictores, turma omnis chorusque: 'chaere, Tite!' hinc hostis mi Albucius, hinc inimicus. Sed iure Mucius. Ego autem mirari satis non queo unde hoc sit tam insolens domesticarum rerum fastidium. Non est omnino hic docendi locus; sed ita prorsus existimo, neque eum Torquatum, qui hoc primus cognomen invenerit, aut torquem illum hosti detraxisse, ut aliquam ex eo est consecutus? – Laudem et caritatem, quae sunt vitae.

VI. Notations

Symbol	**Meaning**
$A(t)$	Arm pulled at time t,where an arm is a vector $A(t) \in \mathbb{R}^d$
$\mathcal{A}_r$	Set of all arms that are used in round r of the algorithm
θ^*	The unknown vector used to calculate reward.
η	gaussian random variable with mean 0 and Variance 1.
$X(t)$	Reward at time t ,calculated by $X(t) = \langle A(t), \theta^* \rangle + \eta$
π_r	a distribution vector, with the probability of each arm index to be chosen at round r.
m	$\dfrac{T - \min\left(K, \frac{(d)(d+1)}{2}\right) - \sum_{r=1}^{\lceil \log_2(d) \rceil + 1} \left\lceil \frac{d}{2^r} \right\rceil}{\lceil \log_2(d) \rceil}$
$\\|v\\|_A^2$	$v^T A v$ where A is a positive semi definite matrix.
$T_r(i)$	$\lfloor \pi_r(A_r(i)) \cdot m \rfloor$ Which represents the number of times arm A_i was chosen in round r.
V_r	$\sum_{i \in A_{r-1}} T_r(i) A_r(i) A_r(i)^T$

References

[1] K. Jamieson, "Lecture Notes on Multi Armed Bandits," *University of Washington*, 2021.

[2] V. Y. F. T. Junwen Yang, "Minimax Optimal Fixed-Budget Best Arm Identification in Linear Bandits," *Armenian Journal of Proceedings*, vol. 61, pp. 192–219, 2020.

[3] "https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.minimize.html."