Stochastic Recursive Optimization: A Structured Multi-Armed Bandit Problem

Roland Szögi
Supervisor: Balázs Csanád Csáji

I. Introduction

A multi-armed bandit problem is a problem in which a series of decisions have to be made in order to maximize the expected reward while having partial knowledge of the usefulness of the actions. However, by choosing an action, information can be gained about its usefulness. The multi-armed bandit problem is one of the most studied problems in decision theory [1] with many applications including A/B testing, advert placement and recommendation services [2].

II. Multi-armed Bandits

The multi-armed bandit model consists of a set of arms $\mathcal{A}$ ($n = |\mathcal{A}|$) and to every arm $a \in \mathcal{A}$ belongs a distribution $\nu(a)$. In each round an arm $a \in \mathcal{A}$ is chosen and a reward $R(a)$ is sampled from distribution $\nu(a)$. An arm is called optimal, if it has the highest expected reward. There can be multiple optimal arms, their expected reward is denoted by r^*.

Definition 1. *An arm $a \in \mathcal{A}$ is called ε-optimal if*

$$\mathbb{E}[R(a)] \geq r^* - \varepsilon.$$

One of the most common learning objectives is to find an ε-optimal arm with high probability.

Definition 2. *An algorithm is called an (ε, δ)-PAC (probably approximately correct) algorithm for the multi-armed bandit problem with sample complexity T, if it outputs an ε-optimal arm with probability at least $1 - \delta$ when it terminates, and the number of steps the algorithm performs until termination is bounded by T.*

An (ε, δ)-PAC algorithm is known for the case of binary rewards, called Median Elimination [3].

Algorithm 1 Median Elimination

```
Input: $\varepsilon > 0$, $\delta > 0$
Output: an arm which is $\varepsilon$-optimal with probability at least $1 - \delta$
  Set $S_1 = \mathcal{A}$, $\varepsilon_1 = \varepsilon/4$, $\delta_1 = \delta/2$, $\ell = 1$.
  repeat
    Sample every arm $a \in S_\ell$ for $1/(\varepsilon_\ell/2)^2 \log(3/\delta_\ell)$ times, and let $\hat{p}_a^\ell$ denote its empirical value.
    Find the median of $\hat{p}_a^\ell$, denoted by $m_\ell$.
    $S_{\ell+1} = S_\ell \setminus \{a : \hat{p}_a^\ell < m_\ell\}$
    $\varepsilon_{\ell+1} = \frac{3}{4}\varepsilon_\ell$, $\delta_{\ell+1} = \delta_\ell/2$, $\ell = \ell + 1$
  until $|S_\ell| = 1$
```

Statement 1. *The Median Elimination algorithm is an (ε, δ)-PAC algorithm and its sample complexity is*

$$\mathcal{O}\left(\frac{n}{\varepsilon^2} \log \frac{1}{\delta}\right).$$

In [4] an $\mathcal{O}\left((n/\varepsilon^2)\log(1/\delta)\right)$ lower bound is provided on the expected number of trials under any policy that finds an ε-optimal arm with probability at least $1 - \delta$.

III. Subgaussian Random Variables

More information about subgaussian random variables and the proof of Statement 2 can be found in [2].

Definition 3. *A random variable X is σ-subgaussian if for all $\lambda \in \mathbb{R}$:*

$$\mathbb{E}[\exp(\lambda X)] \leq \exp\left(\frac{\lambda^2\sigma^2}{2}\right).$$

Statement 2. *Assume that $X_i - \mu$ are independent, σ-subgaussian random variables. Then for any $\varepsilon > 0$,*

$$\mathbb{P}(\hat{\mu} \geq \mu + \varepsilon) \leq \exp\left(-\frac{n\varepsilon^2}{2\sigma^2}\right),$$

$$\mathbb{P}(\hat{\mu} \leq \mu - \varepsilon) \leq \exp\left(-\frac{n\varepsilon^2}{2\sigma^2}\right)$$

where $\hat{\mu} = \frac{1}{n}\sum_{i=1}^n X_i$.

Remark 1. *For random variables that are not centered ($\mathbb{E}[X] \neq 0$), the notation is abused by saying that X is σ-subgaussian if the noise $X - \mathbb{E}[X]$ is σ-subgaussian. A distribution is called σ-subgaussian if a random variable drawn from that distribution is σ-subgaussian.*

IV. Arms with a Special Structure

In the previous semesters I investigated a special case of the multi-armed bandit problem, in which we have further knowledge of the structure of the arms. This special case of the multi-armed bandit problem also arises in practice, for example the quantized estimation problem studied in [5] leads to a bandit problem of this kind.

In this case the assumptions on the arms are the following:

Assumption 1. *There are $n = 2^m + 1$, $m \geq 1$ arms numbered from 0 to 2^m: $a_0, a_1, ..., a_{2^m}$.*
(The expectation of arm a_i is denoted by μ_i.)

Assumption 2. *There exists a $k \in \{0, 1, ..., 2^m\}$ such that*

$$\mu_0 < \mu_1 < ... < \mu_{k-1} < \mu_k > \mu_{k+1} > \mu_{k+2} > ... > \mu_{2^m}.$$

Assumption 3. *There exists a $\Delta > 0$ such that*

$$|\mu_{i+1} - \mu_i| \geq \Delta \quad \forall\, i \in \{0, 1, ..., 2^m - 1\},$$

and it is known in advance.

Assumption 4. *The arms are* 1*-subgaussian.*

Algorithm 2

```
Input: $\delta > 0$
Output: an arm which is optimal with probability at least $1 - \delta$
  Set $S_1 = \mathcal{A}$, $\delta_1 = \delta/2$, $\ell = 1$.
  while $|S_\ell| > 3$ do
    Sample the three arms: $a_j$ , $j \in \{i \cdot 2^{m-\ell-1},\ i = 1, 2, 3\}$
    $n_\ell = \lceil \log(4/\delta_\ell)/(2^{2m-2\ell-5}\Delta^2) \rceil$ times each, and let $\hat{\mu}_j^\ell$ denote their empirical values
    $i_\ell^* = \arg\max_j \hat{\mu}_j^\ell$
    $S_{\ell+1} = \left\{ a_i : i_\ell^* - 2^{m-\ell-1} \leq i \leq i_\ell^* + 2^{m-\ell-1} \right\}$
    Renumber the arms from 0 to $2^{m-\ell}$
    $\delta_{\ell+1} = \delta_\ell/2$, $\ell = \ell + 1$
  end while
  Sample each of the three remaining arms $a_j$, $j \in \{0, 1, 2\}$
  $n_m = \lceil \log(4/\delta_m)/(2^{-3}\Delta^2) \rceil$ times, and let $\hat{\mu}_j^m$ denote their empirical values
  $i_m^* = \arg\max_j \hat{\mu}_j^m$
  return $a_{i_m^*}$
```

Theorem 1. *Under Assumptions 1-4, Algorithm 2 finds the optimal arm with probability at least $1 - \delta$ and its sample complexity is*

$$\mathcal{O}\left(\log n + \frac{1}{\Delta^2} \log \frac{n}{\delta} \right).$$

Remark 2. *The general case when $2^{m-1} + 1 < n \leq 2^m + 1$ can be solved using Algorithm 2 with a simple trick.*

V. Infinitely Many Arms with a Concave Structure

In the previous semester I considered the problem in which the arms are the points of the $[0, 1]$ interval and an unknown concave function describes the expectations of the arms. In this case the assumptions are the following:

Assumption 5. *The arms are the points of the $[0, 1]$ interval and the expectation of arm $a \in [0, 1]$ is $f(a)$, where $f : [0, 1] \to \mathbb{R}$ is an unknown concave function.*

Assumption 6. *The arms are* 1*-subgaussian.*

In each round of Algorithm 3 the set of arms is reduced or the algorithm terminates. In round ℓ the set of arms is denoted by S_ℓ, which is divided into four subintervals of equal size by five arms: x_0^ℓ, x_1^ℓ, x_2^ℓ, x_3^ℓ, x_4^ℓ. The expectation of arm x_i^ℓ is denoted by μ_i^ℓ. In round ℓ we have estimates of μ_0^ℓ and μ_4^ℓ from the previous round. We want to estimate μ_1^ℓ, μ_2^ℓ and μ_3^ℓ as well, so we sample x_1^ℓ, x_2^ℓ, x_3^ℓ many times and we estimate the expectations with the sample means. The sample mean of arm x_i^ℓ is denoted by $\hat{\mu}_i^\ell$. If $\hat{\mu}_1^\ell$ and $\hat{\mu}_3^\ell$ are close to $\hat{\mu}_2^\ell$, then the arm x_2^ℓ will be close-to-optimal because of the concavity and the algorithm returns x_2^ℓ. If $\hat{\mu}_1^\ell$ is much bigger than $\hat{\mu}_2^\ell$, then $\mu_1^\ell \geq \mu_2^\ell$ and because of the concavity, the optimal arm cannot be on the right of arm x_2^ℓ and we can remove these arms. If $\hat{\mu}_1^\ell$ is much smaller than $\hat{\mu}_2^\ell$, then $\mu_1^\ell \leq \mu_2^\ell$ and because of the concavity, the optimal arm cannot be on the right of arm x_1^ℓ and we can remove these arms. Similarly, if $\hat{\mu}_3^\ell$ is much smaller or bigger than $\hat{\mu}_2^\ell$, we can also remove arms.

Algorithm 3

```
Input: $\delta > 0, \varepsilon > 0$
Output: an arm which is $\varepsilon$ optimal with probability at least $1 - \delta$
  Set $\delta_0 = \delta/2$,
  $x_0^1 = 0$, $x_1^1 = 0.25$, $x_2^1 = 0.5$, $x_3^1 = 0.75$, $x_4^1 = 1$.
  Sample $n_0 = \lceil 128 \log(4/\delta_0)/\varepsilon^2 \rceil$ times the two arms: $x_0^1$ and $x_4^1$. Let $\hat{\mu}_0^1$ and $\hat{\mu}_4^1$ denote the sample means and $\mu_0^1$, $\mu_4^1$ denote the corresponding expectations.
  Set $S_1 = [0, 1]$, $\delta_1 = \delta_0/2$, $\ell = 1$.
  while TRUE do
    $S_{\ell+1} = S_\ell$.
    Sample $n_\ell = \lceil 128 \log(6/\delta_\ell)/\varepsilon^2 \rceil$ times the three arms: $x_1^\ell$, $x_2^\ell$, $x_3^\ell$. Let $\hat{\mu}_1^\ell, \hat{\mu}_2^\ell, \hat{\mu}_3^\ell$ denote the sample means and $\mu_1^\ell, \mu_2^\ell, \mu_3^\ell$ denote the expectations.
    if $\hat{\mu}_1^\ell, \hat{\mu}_3^\ell \in (\hat{\mu}_2^\ell - \varepsilon/4, \hat{\mu}_2^\ell + \varepsilon/4)$ then
      return $x_2^\ell$
    end if
    if $\hat{\mu}_1^\ell \geq \hat{\mu}_2^\ell + \varepsilon/4$ then
      $S_{\ell+1} = S_{\ell+1} \setminus (x_2^\ell, x_4^\ell]$
    else if $\hat{\mu}_1^\ell \leq \hat{\mu}_2^\ell - \varepsilon/4$ then
      $S_{\ell+1} = S_{\ell+1} \setminus [x_0^\ell, x_1^\ell)$
    end if
    if $\hat{\mu}_3^\ell \geq \hat{\mu}_2^\ell + \varepsilon/4$ then
      $S_{\ell+1} = S_{\ell+1} \setminus [x_0^\ell, x_2^\ell)$
    else if $\hat{\mu}_3^\ell \leq \hat{\mu}_2^\ell - \varepsilon/4$ then
      $S_{\ell+1} = S_{\ell+1} \setminus (x_3^\ell, x_4^\ell]$
    end if
    $x_0^{\ell+1} = \min S_{\ell+1}$, $x_4^{\ell+1} = \max S_{\ell+1}$,
    $x_1^{\ell+1} = \frac{3}{4} \cdot x_0^{\ell+1} + \frac{1}{4} \cdot x_4^{\ell+1}$,
    $x_2^{\ell+1} = \frac{1}{2} \cdot x_0^{\ell+1} + \frac{1}{2} \cdot x_4^{\ell+1}$,
    $x_3^{\ell+1} = \frac{1}{4} \cdot x_0^{\ell+1} + \frac{3}{4} \cdot x_4^{\ell+1}$.
    Let $\hat{\mu}_0^{\ell+1}$ and $\hat{\mu}_4^{\ell+1}$ denote the sample means of $x_0^{\ell+1}$ and $x_4^{\ell+1}$ calculated in round $\ell$, and let $\mu_0^{\ell+1}$ and $\mu_4^{\ell+1}$ denote the the expectations of $x_0^{\ell+1}$ and $x_4^{\ell+1}$.
    $\delta_{\ell+1} = \delta_\ell/2$, $\ell = \ell + 1$
  end while
```

Theorem 2. *Under Assumptions 5-6, Algorithm 3 finds an ε-optimal arm with probability at least $1 - \delta$.*

VI. Lipschitz continuous case

In this semester I have considered the case when there are infinitely many arms and an unknown concave function describes the expectations of the arms, where this function is Lipschitz continuous with a known Lipschitz constant (L):

Assumption 7. *Function f is Lipschitz continuous with Lipschitz constant L:*

$$|f(x) - f(y)| \leq L \cdot |x - y| \quad \forall\, x, y \in [0, 1].$$

In this case if the length of the set of arms is less than or equal to $2 \cdot \varepsilon/L$ and it contains the optimal arm, then the arm in the middle of the interval will be ε-optimal. By modifying Algorithm 3 so that in this case it returns the arm in the middle, we get Algorithm 4.

Algorithm 4

```
Input: $\delta > 0, \varepsilon > 0$
Output: an arm which is $\varepsilon$ optimal with probability at least
  $1 - \delta$
  Set $\delta_0 = \delta/2$,
  $x_0^1 = 0$, $x_1^1 = 0.25$, $x_2^1 = 0.5$, $x_3^1 = 0.75$, $x_4^1 = 1$.
  Sample $n_0 = \lceil 128 \log(4/\delta_0)/\varepsilon^2 \rceil$ times $x_0^1$ and $x_4^1$.
  Let $\hat{\mu}_0^1$, $\hat{\mu}_4^1$ denote the sample means and $\mu_0^1$, $\mu_4^1$ denote the
  expectations.
  Set $S_1 = [0, 1]$, $\delta_1 = \delta_0/2$, $\ell = 1$.
  while TRUE do
    if $|S_\ell| \leq 2 \cdot \varepsilon/L$ then
      return $x_2^\ell$
    end if
    $S_{\ell+1} = S_\ell$.
    Sample $n_\ell = \lceil 128 \log(6/\delta_\ell)/\varepsilon^2 \rceil$ times $x_1^\ell$, $x_2^\ell$ and $x_3^\ell$.
    Let $\hat{\mu}_1^\ell, \hat{\mu}_2^\ell, \hat{\mu}_3^\ell$ denote the sample means and $\mu_1^\ell, \mu_2^\ell, \mu_3^\ell$
    denote the expectations.
    if $\hat{\mu}_1^\ell, \hat{\mu}_3^\ell \in (\hat{\mu}_2^\ell - \varepsilon/4, \hat{\mu}_2^\ell + \varepsilon/4)$ then
      return $x_2^\ell$
    end if
    if $\hat{\mu}_1^\ell \geq \hat{\mu}_2^\ell + \varepsilon/4$ then
      $S_{\ell+1} = S_{\ell+1} \setminus (x_2^\ell, x_4^\ell]$
    else if $\hat{\mu}_1^\ell \leq \hat{\mu}_2^\ell - \varepsilon/4$ then
      $S_{\ell+1} = S_{\ell+1} \setminus [x_0^\ell, x_1^\ell)$
    end if
    if $\hat{\mu}_3^\ell \geq \hat{\mu}_2^\ell + \varepsilon/4$ then
      $S_{\ell+1} = S_{\ell+1} \setminus [x_0^\ell, x_2^\ell)$
    else if $\hat{\mu}_3^\ell \leq \hat{\mu}_2^\ell - \varepsilon/4$ then
      $S_{\ell+1} = S_{\ell+1} \setminus (x_3^\ell, x_4^\ell]$
    end if
    $x_0^{\ell+1} = \min S_{\ell+1}$, $x_4^{\ell+1} = \max S_{\ell+1}$,
    $x_1^{\ell+1} = \frac{3}{4} \cdot x_0^{\ell+1} + \frac{1}{4} \cdot x_4^{\ell+1}$,
    $x_2^{\ell+1} = \frac{1}{2} \cdot x_0^{\ell+1} + \frac{1}{2} \cdot x_4^{\ell+1}$,
    $x_3^{\ell+1} = \frac{1}{4} \cdot x_0^{\ell+1} + \frac{3}{4} \cdot x_4^{\ell+1}$.
    Let $\hat{\mu}_0^{\ell+1}$ and $\hat{\mu}_4^{\ell+1}$ denote the sample means of $x_0^{\ell+1}$ and
    $x_4^{\ell+1}$ calculated in round $\ell$, and let $\mu_0^{\ell+1}$ and $\mu_4^{\ell+1}$ denote
    the the expectations of $x_0^{\ell+1}$ and $x_4^{\ell+1}$.
    $\delta_{\ell+1} = \delta_\ell/2$, $\ell = \ell + 1$
  end while
```

Theorem 3. *Under Assumptions 5-7, Algorithm 4 is an (ε, δ)-PAC algorithm, and its sample complexity is*

$$\mathcal{O}\left(\frac{1}{\varepsilon^2}\left(\left(\log\frac{L}{\varepsilon}\right)^2 + \left(\log\frac{L}{\varepsilon}\right) \cdot \log\frac{1}{\delta}\right)\right).$$

Proof. If f is Lipschitz continuous with Lipschitz constant L and there is an interval with length less than or equal to $2 \cdot \varepsilon/L$ containing the optimal arm (y), then the arm in the middle of the interval (x) is ε-optimal:

$$|x - y| \leq \varepsilon/L \implies |f(x) - f(y)| \leq \varepsilon.$$

We have previously seen that the probability that the optimal arm is removed by Algorithm 3 is less than δ, this way this Algorithm 4 is an (ε, δ)-PAC algorithm.

After ℓ rounds the length of the interval is $(1/2)^\ell$, so

$$\ell = \left\lfloor \log_2 \frac{L}{\varepsilon} \right\rfloor$$

round is enough to get an interval no longer than $2 \cdot \varepsilon/L$. Algorithm 4 terminates in ℓ rounds, so the number of samples required by Algorithm 4 is bounded by:

$$\begin{aligned}
&2\left\lceil \frac{128}{\varepsilon^2} \log\frac{4}{\delta_0} \right\rceil + 3\sum_{i=1}^{\ell} \left\lceil \frac{128}{\varepsilon^2} \log\frac{6}{\delta_i} \right\rceil \\
&\leq 3\ell + 2 + 2 \cdot \frac{128}{\varepsilon^2} \log\frac{4}{\delta_0} + 3 \cdot \frac{128}{\varepsilon^2} \sum_{i=1}^{\ell} \log\frac{6}{\delta_i} \\
&\leq 3\ell + 2 + 3 \cdot \frac{128}{\varepsilon^2} \sum_{i=0}^{\ell} \log\frac{6}{\delta_i} \\
&\leq 3\ell + 2 + 3 \cdot \frac{128}{\varepsilon^2} \sum_{i=1}^{\ell} \log\frac{12 \cdot 2^i}{\delta} \\
&= 3\ell + 2 + 3 \cdot \frac{128}{\varepsilon^2} \log\left(\prod_{i=1}^{\ell} \frac{12 \cdot 2^i}{\delta}\right) \\
&= 3\ell + 2 + 3 \cdot \frac{128}{\varepsilon^2} \log\left(\frac{12^{\ell+1}}{\delta^{\ell+1}} \cdot 2^{\frac{\ell(\ell+1)}{2}}\right) \\
&= 3\ell + 2 + 3 \cdot \frac{128}{\varepsilon^2} \left((\ell + 1)\log\frac{12}{\delta} + \frac{\ell(\ell+1)}{2}\log 2\right) \\
&= \mathcal{O}\left(\frac{1}{\varepsilon^2}\left(\ell^2 + \ell \cdot \log\frac{1}{\delta}\right)\right) \\
&= \mathcal{O}\left(\frac{1}{\varepsilon^2}\left(\left(\log\frac{L}{\varepsilon}\right)^2 + \left(\log\frac{L}{\varepsilon}\right) \cdot \log\frac{1}{\delta}\right)\right).
\end{aligned}$$

□

VII. Finitely Many Arms with a Concave Structure in 1D

In this section we will consider the case when there are finitely many arms and an unknown concave function describes the expectations of the arms. By modifying Algorithm 3 we can achieve to halve the set of arms in each round. First we will see that this modified algorithm can be used to solve the case when there are $2^m + 1$ arms for some $m \in \mathbb{N}$ and then the general case can be solved using this special case. The assumptions are the following:

Assumption 8. *There are $n = 2^m + 1$, $m \geq 1$ arms numbered from 0 to 2^m.*

Assumption 9. *The expectation of arm i is $f(i)$, where $f : \mathbb{R} \to \mathbb{R}$ is an unknown concave function.*

Assumption 10. *The arms are 1-subgaussian.*

The difference between Algorithm 3 and Algorithm 5 is that if $\hat{\mu}_1^\ell \in (\hat{\mu}_2^\ell - \varepsilon/4, \hat{\mu}_2^\ell + \varepsilon/4)$ and $\hat{\mu}_3^\ell \le \hat{\mu}_2^\ell - \varepsilon/4$ then in Algorithm 3 only a quarter of the arms is removed, while in Algorithm 5 the arm $x_{1.5}^\ell = (x_1^\ell + x_2^\ell)/2$ is sampled and based on the results another quarter of the arms is removed. Similarly, if $\hat{\mu}_1^\ell \le \hat{\mu}_2^\ell - \varepsilon/4$ and $\hat{\mu}_3^\ell \in (\hat{\mu}_2^\ell - \varepsilon/4, \hat{\mu}_2^\ell + \varepsilon/4)$ then arm $x_{2.5}^\ell = (x_2^\ell + x_3^\ell)/2$ is sampled in order to remove another quarter of the arms.

Theorem 4. *Under Assumptions 8 - 10, Algorithm 5 finds an ε-optimal arm with probability at least $1 - \delta$.*

Lemma 1. *Using the notation of Algorithm 5:*

$$\mathbb{P}(|\hat{\mu}_0^1 - \mu_0^1| \ge \varepsilon/8 \vee |\hat{\mu}_0^4 - \mu_0^4| \ge \varepsilon/8) \le \delta/2.$$

Proof. Based on Statement 2:

$$\begin{aligned} &\mathbb{P}(|\hat{\mu}_0^1 - \mu_0^1| \ge \varepsilon/8 \vee |\hat{\mu}_0^4 - \mu_0^4| \ge \varepsilon/8) \\ &\quad \le \mathbb{P}(|\hat{\mu}_0^1 - \mu_0^1| \ge \varepsilon/8) + \mathbb{P}(|\hat{\mu}_0^4 - \mu_0^4| \ge \varepsilon/8) \\ &\quad \le 2 \cdot 2 \cdot \exp \cdot \left(-n_0 \cdot (\varepsilon/8)^2/2\right) \\ &\quad \le \delta_0 = \delta/2. \end{aligned}$$

□

Lemma 2. *If x is sampled $n_1 = \lceil 128 \log(6/\delta)/\varepsilon^2 \rceil$ times then*

$$\mathbb{P}(|\hat{\mu} - \mu_i| \ge \varepsilon/8) \le \delta/3.$$

Similarly, if x is sampled $n_2 = \lceil 288 \log(6/\delta)/\varepsilon^2 \rceil$ times then

$$\mathbb{P}(|\hat{\mu} - \mu_i| \ge \varepsilon/12) \le \delta/3.$$

Proof. Based on Statement 2:

$$\mathbb{P}(|\hat{\mu} - \mu_i| \ge \varepsilon/8) \le 2 \exp \cdot \left(-n_1 \cdot (\varepsilon/8)^2/2\right) \le \delta/3.$$

$$\mathbb{P}(|\hat{\mu} - \mu_i| \ge \varepsilon/12) \le 2 \exp \cdot \left(-n_2 \cdot (\varepsilon/12)^2/2\right) \le \delta/3.$$

□

Lemma 3. *Suppose that $|\hat{\mu}_i^\ell - \mu_i^\ell| \le \varepsilon/8$ and $|\hat{\mu}_j^\ell - \mu_j^\ell| \le \varepsilon/8$. Let x^* denote the optimal arm. If $i < j$ and $\hat{\mu}_i^\ell \ge \hat{\mu}_j^\ell + \varepsilon/4$ then $x^* \le x_j^\ell$. Similarly, if $i > j$ and $\hat{\mu}_i^\ell \ge \hat{\mu}_j^\ell + \varepsilon/4$ then $x^* \ge x_j^\ell$.*

Proof. If $i < j$ and $\hat{\mu}_i^\ell \ge \hat{\mu}_j^\ell + \varepsilon/4$ then $\mu_i^\ell \ge \mu_j^\ell$:

$$\begin{aligned} \mu_i^\ell &\ge \hat{\mu}_i^\ell - \varepsilon/8 \\ &\ge \hat{\mu}_j^\ell + \varepsilon/4 - \varepsilon/8 \\ &= \hat{\mu}_j^\ell + \varepsilon/8 \\ &\ge \mu_j^\ell. \end{aligned}$$

Arguing indirectly, assume that $x^* > x_j^\ell$. Then there exists a $t \in [0, 1]$ such that $x_j^\ell = t \cdot x_i^\ell + (1 - t) \cdot x^*$. Because of the concavity:

$$f(x_j^\ell) = f(t \cdot x_i^\ell + (1-t) \cdot x^*) \ge t \cdot f(x_i^\ell) + (1-t) \cdot f(x^*) > f(x_i^\ell).$$

It contradicts the fact that $\mu_i^\ell \ge \mu_j^\ell$.

Algorithm 5

```
Input: $\delta > 0, \varepsilon > 0$
Output: an $\varepsilon$-optimal arm with probability at least $1 - \delta$
  Set $\delta_0 = \delta/2$,
  $x_0^1 = 0$, $x_1^1 = 2^{m-2}$, $x_2^1 = 2^{m-1}$, $x_3^1 = 3 \cdot 2^{m-2}$, $x_4^1 = 2^m$.
  Sample $n_0 = \lceil 128 \log(4/\delta_0)/\varepsilon^2 \rceil$ times $x_0^1$ and $x_4^1$. The sample mean and the expectation of $x_i^\ell$ will be denoted by $\hat{\mu}_i^\ell$ and $\mu_i^\ell$.
  Set $S_1 = \{i : 0 \le i \le 2^m\}$, $\delta_1 = \delta_0/2$, $\ell = 1$.
  while $|S_\ell| > 5$ do
    Sample the three arms: $x_1^\ell$, $x_2^\ell$, $x_3^\ell$, so that each of them will be sampled $n_\ell = \lceil 128 \log(6/\delta_\ell)/\varepsilon^2 \rceil$ times totally.
    if $\hat{\mu}_1^\ell, \hat{\mu}_3^\ell \in (\hat{\mu}_2^\ell - \varepsilon/4, \hat{\mu}_2^\ell + \varepsilon/4)$ then
      return $x_2^\ell$
    else if $\hat{\mu}_1^\ell \ge \hat{\mu}_2^\ell + \varepsilon/4$ then
      $S_{\ell+1} = \{i \in S_\ell : i \le x_2^\ell\}$
    else if $\hat{\mu}_3^\ell \ge \hat{\mu}_2^\ell + \varepsilon/4$ then
      $S_{\ell+1} = \{i \in S_\ell : i \ge x_2^\ell\}$
    else if $\hat{\mu}_1^\ell, \hat{\mu}_3^\ell \le \hat{\mu}_2^\ell - \varepsilon/4$ then
      $S_{\ell+1} = \{i \in S_\ell : x_1^\ell \le i \le x_3^\ell\}$
    else if $\hat{\mu}_1^\ell \in (\hat{\mu}_2^\ell - \varepsilon/4, \hat{\mu}_2^\ell + \varepsilon/4)$ and $\hat{\mu}_3^\ell \le \hat{\mu}_2^\ell - \varepsilon/4$ then
      $x_{1.5}^\ell = (x_1^\ell + x_2^\ell)/2$
      Sample $n = \lceil 288 \log(6/\delta_\ell)/\varepsilon^2 \rceil - \lceil 128 \log(6/\delta_\ell)/\varepsilon^2 \rceil$ times $x_1^\ell$ and $x_2^\ell$ so that they are sampled $\lceil 288 \log(6/\delta_\ell)/\varepsilon^2 \rceil$ times totally. Sample $x_{1.5}^\ell$
      $n = \lceil 288 \log(12/\delta_\ell)/\varepsilon^2 \rceil$ times.
      if $\hat{\mu}_1^\ell \le \hat{\mu}_{1.5}^\ell - \varepsilon/6$ or $\hat{\mu}_2^\ell \ge \hat{\mu}_{1.5}^\ell + \varepsilon/6$ then
        $S_{\ell+1} = \{i \in S_\ell : x_1^\ell \le i \le x_3^\ell\}$
      else if $\hat{\mu}_2^\ell \le \hat{\mu}_{1.5}^\ell - \varepsilon/6$ or $\hat{\mu}_1^\ell \ge \hat{\mu}_{1.5}^\ell + \varepsilon/6$ then
        $S_{\ell+1} = \{i \in S_\ell : i \le x_2^\ell\}$
      else
        return $x_{1.5}^\ell$
      end if
    else if $\hat{\mu}_1^\ell \le \hat{\mu}_2^\ell - \varepsilon/4$ and $\hat{\mu}_3^\ell \in (\hat{\mu}_2^\ell - \varepsilon/4, \hat{\mu}_2^\ell + \varepsilon/4)$ then
      $x_{2.5}^\ell = (x_2^\ell + x_3^\ell)/2$
      Sample $n = \lceil 288 \log(6/\delta_\ell)/\varepsilon^2 \rceil - \lceil 128 \log(6/\delta_\ell)/\varepsilon^2 \rceil$ times $x_2^\ell$ and $x_3^\ell$ so that they are sampled $\lceil 288 \log(6/\delta_\ell)/\varepsilon^2 \rceil$ times totally. Sample $x_{2.5}^\ell$
      $n = \lceil 288 \log(12/\delta_\ell)/\varepsilon^2 \rceil$ times.
      if $\hat{\mu}_2^\ell \le \hat{\mu}_{2.5}^\ell - \varepsilon/6$ or $\hat{\mu}_3^\ell \ge \hat{\mu}_{2.5}^\ell + \varepsilon/6$ then
        $S_{\ell+1} = \{i \in S_\ell : i \le x_2^\ell\}$
      else if $\hat{\mu}_3^\ell \le \hat{\mu}_{2.5}^\ell - \varepsilon/6$ or $\hat{\mu}_2^\ell \ge \hat{\mu}_{2.5}^\ell + \varepsilon/6$ then
        $S_{\ell+1} = \{i \in S_\ell : x_1^\ell \le i \le x_3^\ell\}$
      else
        return $x_{2.5}^\ell$
      end if
    end if
    $x_0^{\ell+1} = \min S_{\ell+1}$, $x_4^{\ell+1} = \max S_{\ell+1}$,
    $x_1^{\ell+1} = (3 \cdot x_0^{\ell+1} + x_4^{\ell+1})/4$,
    $x_2^{\ell+1} = (x_0^{\ell+1} + x_4^{\ell+1})/2$,
    $x_3^{\ell+1} = (x_0^{\ell+1} + 3 \cdot x_4^{\ell+1})/4$.
    $\delta_{\ell+1} = \delta_\ell/2$, $\ell = \ell + 1$
  end while
  Sample the 5 remaining arms so that each is sampled
  $n_\ell = \lceil 128 \log(6/\delta_\ell)/\varepsilon^2 \rceil$ times. Return the arm with the highest sample mean.
```

Similarly, if $i > j$ and $\hat{\mu}_i^\ell \geq \hat{\mu}_j^\ell + \varepsilon/4$ then $\mu_i^\ell \geq \mu_j^\ell$. Arguing indirectly, assume that $x^* < x_j^\ell$. Then there exists a $t \in [0,1]$ such that $x_j^\ell = t \cdot x_i^\ell + (1-t) \cdot x^*$. Because of the concavity:

$$f(x_j^\ell) = f(t \cdot x_i^\ell + (1-t) \cdot x^*) \geq t \cdot f(x_i^\ell) + (1-t) \cdot f(x^*) > f(x_i^\ell).$$

It contradicts the fact that $\mu_i^\ell \geq \mu_j^\ell$.

□

Lemma 4. *Suppose that $|\hat{\mu}_i^\ell - \mu_i^\ell| \leq \varepsilon/12$, $|\hat{\mu}_j^\ell - \mu_j^\ell| \leq \varepsilon/12$. Let x^* denote the optimal arm. If $i < j$ and $\hat{\mu}_i^\ell \geq \hat{\mu}_j^\ell + \varepsilon/6$ then $x^* \leq x_j^\ell$. Similarly, if $i > j$ and $\hat{\mu}_i^\ell \geq \hat{\mu}_j^\ell + \varepsilon/6$ then $x^* \geq x_j^\ell$.*

Proof. It can be proved the same way as Lemma 3. □

Lemma 5. *Suppose that $|\hat{\mu}_j^\ell - \mu_j^\ell| \leq \varepsilon/8$, $j = 0, 1, 2, 3, 4$. If $\hat{\mu}_{i-1}^\ell, \hat{\mu}_{i+1}^\ell \in (\hat{\mu}_i^\ell - \varepsilon/4, \hat{\mu}_i^\ell + \varepsilon/4)$, then*

$$\mu_{i-1}^\ell, \mu_{i+1}^\ell \in [\mu_i^\ell - \varepsilon/2, \mu_i^\ell + \varepsilon/2].$$

Proof.

$$\begin{aligned}
\mu_{i-1}^\ell &\geq \hat{\mu}_{i-1}^\ell - \varepsilon/8 \\
&\geq \hat{\mu}_i^\ell - \varepsilon/4 - \varepsilon/8 \\
&= \hat{\mu}_i^\ell - 3/8 \cdot \varepsilon \\
&\geq \mu_i^\ell - \varepsilon/8 - 3/8 \cdot \varepsilon \\
&= \mu_i^\ell - \varepsilon/2
\end{aligned}$$

$$\begin{aligned}
\mu_{i-1}^\ell &\leq \hat{\mu}_{i-1}^\ell + \varepsilon/8 \\
&\leq \hat{\mu}_i^\ell + \varepsilon/4 + \varepsilon/8 \\
&= \hat{\mu}_i^\ell + 3/8 \cdot \varepsilon \\
&\leq \mu_i^\ell + \varepsilon/8 + 3/8 \cdot \varepsilon \\
&= \mu_i^\ell + \varepsilon/2
\end{aligned}$$

Similarly, $\mu_{i+1}^\ell \in [\mu_i^\ell - \varepsilon/2, \mu_i^\ell + \varepsilon/2]$.

□

Lemma 6. *Suppose that $|\hat{\mu}_j^\ell - \mu_j^\ell| \leq \varepsilon/12$, $j = 0, 1, 2, 3, 4$. If $\hat{\mu}_{i-1}^\ell, \hat{\mu}_{i+1}^\ell \in (\hat{\mu}_i^\ell - \varepsilon/6, \hat{\mu}_i^\ell + \varepsilon/6)$, then*

$$\mu_{i-1}^\ell, \mu_{i+1}^\ell \in [\mu_i^\ell - \varepsilon/3, \mu_i^\ell + \varepsilon/3].$$

Proof.

$$\begin{aligned}
\mu_{i-1}^\ell &\geq \hat{\mu}_{i-1}^\ell - \varepsilon/12 \\
&\geq \hat{\mu}_i^\ell - \varepsilon/6 - \varepsilon/12 \\
&= \hat{\mu}_i^\ell - 3/12 \cdot \varepsilon \\
&\geq \mu_i^\ell - \varepsilon/12 - 3/12 \cdot \varepsilon \\
&= \mu_i^\ell - \varepsilon/3
\end{aligned}$$

$$\begin{aligned}
\mu_{i-1}^\ell &\leq \hat{\mu}_{i-1}^\ell + \varepsilon/12 \\
&\leq \hat{\mu}_i^\ell + \varepsilon/6 + \varepsilon/12 \\
&= \hat{\mu}_i^\ell + 3/12 \cdot \varepsilon \\
&\leq \mu_i^\ell + \varepsilon/12 + 3/12 \cdot \varepsilon \\
&= \mu_i^\ell + \varepsilon/3
\end{aligned}$$

Similarly, $\mu_{i+1}^\ell \in [\mu_i^\ell - \varepsilon/3, \mu_i^\ell + \varepsilon/3]$.

□

Lemma 7. *Suppose that $|\hat{\mu}_i^\ell - \mu_i^\ell| \leq \varepsilon/8$, $i = 1, 2, 3$. If $\hat{\mu}_1^\ell, \hat{\mu}_3^\ell \in (\hat{\mu}_2^\ell - \varepsilon/4, \hat{\mu}_2^\ell + \varepsilon/4)$, then*

$$\mu_2^\ell \geq \max_{x \in S_\ell} f(x) - \varepsilon.$$

Proof. Let $x^* = \arg\max_{x \in S_\ell} f(x)$. Arguing indirectly assume, that $f(x^*) > f(x_2^\ell) + \varepsilon$.

If $x^* < x_1^\ell$, then there exists a $t \in [0,1]$ such that

$$x_1^\ell = (1-t) \cdot x^* + t \cdot x_2^\ell.$$

As x_1^ℓ is closer to x^* than to x_2^ℓ, $t \leq 1/2$. Because of the concavity:

$$\begin{aligned}
f(x_1^\ell) &= f((1-t) \cdot x^* + t \cdot x_2^\ell) \\
&\geq (1-t) \cdot f(x^*) + t \cdot f(x_2^\ell) \\
&> (1-t) \cdot (f(x_2^\ell) + \varepsilon) + t \cdot f(x_2^\ell) \\
&= f(x_2^\ell) + (1-t) \cdot \varepsilon \\
&\geq f(x_2^\ell) + \varepsilon/2.
\end{aligned}$$

It contradicts the fact that $f(x_1^\ell) \leq f(x_2^\ell) + \varepsilon/2$ based on Lemma 5.

If $x_1^\ell < x^* < x_2^\ell$, then there exists a $t \in [0,1]$ such that $x_2^\ell = (1-t) \cdot x^* + t \cdot x_3^\ell$. As x_2^ℓ is closer to x^* than to x_3^ℓ, $t \leq 1/2$. Because of the concavity:

$$\begin{aligned}
f(x_2^\ell) &= f((1-t) \cdot x^* + t \cdot x_3^\ell) \\
&\geq (1-t) \cdot f(x^*) + t \cdot f(x_3^\ell) \\
&> (1-t) \cdot (f(x_2^\ell) + \varepsilon) + t \cdot (f(x_2^\ell) - \varepsilon/2) \\
&= f(x_2^\ell) + (1 - 3/2 \cdot t) \cdot \varepsilon \\
&> f(x_2^\ell).
\end{aligned}$$

It is a contradiction.

If $x_2^\ell < x^* < x_3^\ell$, then there exists a $t \in [0,1]$ such that $x_2^\ell = (1-t) \cdot x_1^\ell + t \cdot x^*$. As x_2^ℓ is closer to x^* than to x_1^ℓ, $1 > t \geq 1/2$. Because of the concavity:

$$\begin{aligned}
f(x_2^\ell) &= f((1-t) \cdot x_1^\ell + t \cdot x^*) \\
&\geq (1-t) \cdot f(x_1^\ell) + t \cdot f(x^*) \\
&> (1-t) \cdot (f(x_2^\ell) - \varepsilon/2) + t \cdot (f(x_2^\ell) + \varepsilon) \\
&= f(x_2^\ell) + (3/2 \cdot t - 1/2) \cdot \varepsilon \\
&> f(x_2^\ell).
\end{aligned}$$

It is a contradiction.

If $x_3^\ell < x^*$, then there exists a $t \in [0,1]$ such that

$$x_3^\ell = (1-t) \cdot x_2^\ell + t \cdot x^*.$$

As x_3^ℓ is closer to x^* than to x_2^ℓ, $t \geq 1/2$. Because of the concavity:

$$\begin{aligned}
f(x_3^\ell) &= f((1-t) \cdot x_2^\ell + t \cdot x^*) \\
&\geq (1-t) \cdot f(x_2^\ell) + t \cdot f(x^*) \\
&> (1-t) \cdot f(x_2^\ell) + t \cdot (f(x_2^\ell) + \varepsilon) \\
&= f(x_2^\ell) + t \cdot \varepsilon \\
&\geq f(x_2^\ell) + \varepsilon/2.
\end{aligned}$$

It contradicts the fact that $f(x_3^\ell) \leq f(x_2^\ell) + \varepsilon/2$ based on Lemma 5.

□

Lemma 8. *Suppose that* $|\hat{\mu}_i^\ell - \mu_i^\ell| \leq \varepsilon/12$, $i = 1, 1.5, 2$. *If* $\hat{\mu}_1^\ell, \hat{\mu}_2^\ell \in (\hat{\mu}_{1.5}^\ell - \varepsilon/6, \hat{\mu}_{1.5}^\ell + \varepsilon/6)$*, then*

$$\mu_{1.5}^\ell \geq \max_{x \in \{i : x_0^\ell \leq i \leq x_3^\ell\}} f(x) - \varepsilon.$$

Similarly, if $|\hat{\mu}_i^\ell - \mu_i^\ell| \leq \varepsilon/12$, $i = 2, 2.5, 3$ *and* $\hat{\mu}_2^\ell, \hat{\mu}_3^\ell \in (\hat{\mu}_{2.5}^\ell - \varepsilon/6, \hat{\mu}_{2.5}^\ell + \varepsilon/6)$ *then*

$$\mu_{2.5}^\ell \geq \max_{x \in \{i : x_1^\ell \leq i \leq x_4^\ell\}} f(x) - \varepsilon.$$

Proof. Let

$$x^* = \operatorname*{arg\,max}_{x \in \{i : x_0^\ell \leq i \leq x_3^\ell\}} f(x).$$

Arguing indirectly assume, that $f(x^*) > f(x_{1.5}^\ell) + \varepsilon$.

If $x^* < x_1^\ell$, then there exists a $t \in [0, 1]$ such that

$$x_1^\ell = (1 - t) \cdot x^* + t \cdot x_{1.5}^\ell.$$

From this $t = \frac{x_1 - x^*}{x_{1.5} - x^*} \leq \frac{2}{3}$. Because of the concavity:

$$\begin{aligned} f(x_1^\ell) &= f((1 - t) \cdot x^* + t \cdot x_{1.5}^\ell) \\ &\geq (1 - t) \cdot f(x^*) + t \cdot f(x_{1.5}^\ell) \\ &> (1 - t) \cdot (f(x_{1.5}^\ell) + \varepsilon) + t \cdot f(x_{1.5}^\ell) \\ &= f(x_{1.5}^\ell) + (1 - t) \cdot \varepsilon \\ &\geq f(x_{1.5}^\ell) + \varepsilon/3. \end{aligned}$$

It contradicts the fact that $f(x_1^\ell) \leq f(x_{1.5}^\ell) + \varepsilon/3$ based on Lemma 6.

If $x_1^\ell < x^* < x_{1.5}^\ell$, then there exists a $t \in [0, 1]$ such that $x_{1.5}^\ell = (1 - t) \cdot x^* + t \cdot x_2^\ell$. Here $t = \frac{x_{1.5} - x^*}{x_2 - x^*} \leq \frac{1}{2}$. Because of the concavity:

$$\begin{aligned} f(x_{1.5}^\ell) &= f((1 - t) \cdot x^* + t \cdot x_2^\ell) \\ &\geq (1 - t) \cdot f(x^*) + t \cdot f(x_2^\ell) \\ &> (1 - t) \cdot (f(x_{1.5}^\ell) + \varepsilon) + t \cdot (f(x_{1.5}^\ell) - \varepsilon/3) \\ &= f(x_{1.5}^\ell) + (1 - 4/3 \cdot t) \cdot \varepsilon \\ &> f(x_{1.5}^\ell). \end{aligned}$$

It is a contradiction.

If $x_{1.5}^\ell < x^* < x_2^\ell$, then there exists a $t \in [0, 1]$ such that $x_{1.5}^\ell = (1 - t) \cdot x_1^\ell + t \cdot x^*$. In this case $t = \frac{x_{1.5} - x_1}{x^* - x_1} \geq \frac{1}{2}$. Because of the concavity:

$$\begin{aligned} f(x_{1.5}^\ell) &= f((1 - t) \cdot x_1^\ell + t \cdot x^*) \\ &\geq (1 - t) \cdot f(x_1^\ell) + t \cdot f(x^*) \\ &> (1 - t) \cdot (f(x_{1.5}^\ell) - \varepsilon/3) + t \cdot (f(x_{1.5}^\ell) + \varepsilon) \\ &= f(x_{1.5}^\ell) + (4/3 \cdot t - 1/3) \cdot \varepsilon \\ &> f(x_{1.5}^\ell). \end{aligned}$$

It is a contradiction.

If $x_2^\ell < x^*$, then there exists a $t \in [0, 1]$ such that

$$x_2^\ell = (1 - t) \cdot x_{1.5}^\ell + t \cdot x^*.$$

In this case $t = \frac{x_2 - x_{1.5}}{x^* - x_{1.5}} \geq \frac{1}{3}$. Because of the concavity:

$$\begin{aligned} f(x_2^\ell) &= f((1 - t) \cdot x_{1.5}^\ell + t \cdot x^*) \\ &\geq (1 - t) \cdot f(x_{1.5}^\ell) + t \cdot f(x^*) \\ &> (1 - t) \cdot f(x_{1.5}^\ell) + t \cdot (f(x_{1.5}^\ell) + \varepsilon) \\ &= f(x_{1.5}^\ell) + t \cdot \varepsilon \\ &\geq f(x_{1.5}^\ell) + \varepsilon/3. \end{aligned}$$

It contradicts the fact that $f(x_2^\ell) \leq f(x_{1.5}^\ell) + \varepsilon/3$ based on Lemma 6. □

Lemma 9. *If* $|S_\ell| = 5$ *and* $|\hat{\mu}_j - \mu_j| \leq \varepsilon/8$, $j = 1, 2, 3, 4, 5$ *then the arm with the highest sample mean is* ε*-optimal.*

Proof. Suppose, that $\hat{\mu}_i$ is the highest sample mean. In this case for $j = 1, 2, 3, 4, 5$:

$$\mu_i \geq \hat{\mu}_i - \varepsilon/8 \geq \hat{\mu}_j - \varepsilon/8 \geq \mu_j - \varepsilon/4.$$

□

Based on the lemmas, the probability that there is a round in which for a sampled arm $|\hat{\mu} - \mu| \geq \varepsilon/8$ (or $|\hat{\mu} - \mu| \geq \varepsilon/12$ when needed) is less than δ. Now consider the case, when $|\hat{\mu} - \mu| \leq \varepsilon/8$ (or $|\hat{\mu} - \mu| \leq \varepsilon/12$ when needed) for all of the arms sampled in any of the rounds. In each round the set of the arms is halved or the algorithm terminates. By Lemma 3 and 4 the optimal arm is never thrown away and by Lemma 7, 8 and 9 an ε-optimal arm is returned when the algorithm terminates. This way Algorithm 5 returns an ε-optimal arm with probability at least $1 - \delta$.

Theorem 5. *The sample complexity of Algorithm 5 when there are* $n = 2^m + 1$ *arms:*

$$\mathcal{O}\left(\frac{1}{\varepsilon^2}\left(m^2 + m \log \frac{1}{\delta}\right)\right).$$

Proof. At the beginning we sample $n_0 = \lceil 128 \log(4/\delta_0)/\varepsilon^2 \rceil$ times 2 arms. If in a round $x_{1.5}$ or $x_{2.5}$ is sampled, then we will count these samples to the next round. We can do this, because otherwise this arm would have been sampled in the next round, but this way it is not needed anymore. This way in each round three arms are sampled, each $n = \lceil 288 \log(6/\delta_\ell)/\varepsilon^2 \rceil$ times max. When only 5 arms remain, then 3 of them is already sampled, so in the last round only 2 arms have to be sampled, $n = \lceil 288 \log(6/\delta_{m-1})/\varepsilon^2 \rceil$ times maximum. So the sample complexity is:

$$\begin{aligned}
&2\cdot\lceil 128\log(4/\delta_0)/\varepsilon^2\rceil + 3\cdot\sum_{i=1}^{m-2}\lceil 288\log(6/\delta_\ell)/\varepsilon^2\rceil\\
&\quad + 2\cdot\lceil 288\log(6/\delta_{m-1})/\varepsilon^2\rceil\\
&\quad \le 3m + 3\cdot\frac{288}{\varepsilon^2}\sum_{i=1}^{m-1}\log\frac{6}{\delta_i}\\
&\quad = 3m + 3\cdot\frac{288}{\varepsilon^2}\log\left(\prod_{i=1}^{m-1}\frac{12\cdot 2^i}{\delta}\right)\\
&\quad = 3m + 3\cdot\frac{288}{\varepsilon^2}\log\left(\frac{12^{m-1}\cdot 2^{(m-1)m/2}}{\delta^{m-1}}\right)\\
&\quad = 3m + 3\cdot\frac{288}{\varepsilon^2}\left((m-1)\log\frac{12}{\delta}+\frac{(m-1)m}{2}\log 2\right)\\
&\quad = \mathcal{O}\left(\frac{1}{\varepsilon^2}\left(m^2 + m\log\frac{1}{\delta}\right)\right).
\end{aligned}$$

□

If there are $2^m + 1 < n < 2^{m+1}+1$ arms, we can do the following:
Run the first round of Algorithm 5 with arms

$$\begin{aligned}
&\left\{i : \left\lfloor\frac{n}{2}\right\rfloor - 2^{m-1} \le i \le \left\lfloor\frac{n}{2}\right\rfloor + 2^{m-1}\right\},\\
&x_0^1 = \left\lfloor\frac{n}{2}\right\rfloor - 2^{m-1},\ x_1^1 = \left\lfloor\frac{n}{2}\right\rfloor - 2^{m-2},\ x_2^1 = \left\lfloor\frac{n}{2}\right\rfloor,\\
&x_3^1 = \left\lfloor\frac{n}{2}\right\rfloor + 2^{m-2},\ x_4^1 = \left\lfloor\frac{n}{2}\right\rfloor + 2^{m-1}
\end{aligned}$$

and $\delta_0 = \delta/2$. If after the first round of Algorithm 5:

- $S_2 = \{i : x_0^1 \le i \le x_2^1\}$, then set
$$S = \{i : 0 \le i \le 2^m\},$$
- $S_2 = \{i : x_1^1 \le i \le x_3^1\}$, then set
$$S = \left\{i : \left\lfloor\frac{n}{2}\right\rfloor - 2^{m-1} \le i \le \left\lfloor\frac{n}{2}\right\rfloor + 2^{m-1}\right\},$$
- $S_2 = \{i : x_2^1 \le i \le x_4^1\}$, then set
$$S = \{i : n - 2^m \le i \le n\}.$$

The length of S is $2^m + 1$ in all cases. Now do Algorithm 5 with S and $\delta_0 = \delta/8$.
Based on the lemmas, the probability that the optimal arm is thrown away in the first round is less than $\frac{3}{4}\delta$. If the optimal arm is not removed in the first round, then Algorithm 5 with S and $\delta_0 = \delta/8$ will return an ε-optimal arm with probability at least $1 - \delta/4$. So the probability that the returned arm is not ε-optimal is less than δ.

At the beginning 2 arms are sampled $\lceil 128\log(4/\delta_0)/\varepsilon^2\rceil$ times each, then we sample 3 arms $\lceil 288\log(6/\delta_\ell)/\varepsilon^2\rceil$ times maximum. After that we do Algorithm 5 with $\delta_0 = \delta/8$ with 2^m+1 arms, so based on our previous calculation, the sample complexity in this case is:

$$\begin{aligned}
&2\cdot\lceil 128\log(4/\delta_0)/\varepsilon^2\rceil + 3\cdot\lceil 288\log(6/\delta_\ell)/\varepsilon^2\rceil\\
&\quad + \mathcal{O}\left(\frac{1}{\varepsilon^2}\left(m^2 + m\log\frac{1}{\delta}\right)\right)\\
&\quad = \mathcal{O}\left(\frac{1}{\varepsilon^2}\left(m^2 + m\log\frac{1}{\delta}\right)\right).
\end{aligned}$$

Similarly to the Lipschitz continuous case, if there is a known Δ such that $|\mu_i - \mu_{i-1}| \le \Delta$, $i = 1, 2, ..., n$, then Algorithm 5 can terminate when the number of arms is $2\cdot\lfloor\frac{\varepsilon}{\Delta}\rfloor + 1$ or less, by returning the arm in the middle (x_j). In this case the returned arm will be ε-optimal because:

$$|\mu_i - \mu_j| \le |i - j|\cdot\Delta \le \left\lfloor\frac{\varepsilon}{\Delta}\right\rfloor\cdot\Delta \le \varepsilon \quad \forall i.$$

In this case the algorithm can terminate after $\left\lfloor\log_2\frac{n}{2\lfloor\varepsilon/\Delta\rfloor+1}\right\rfloor$ rounds so the sample complexity in this case is:

$$\mathcal{O}\left(\frac{\ell^2}{\varepsilon^2} + \frac{\ell}{\varepsilon^2}\log\frac{1}{\delta}\right).$$

where $\ell = \left\lfloor\log_2\frac{n}{2\lfloor\varepsilon/\Delta\rfloor+1}\right\rfloor$.

VIII. Conclusion

In this semester I have considered the special case of the multi-armed bandit problem, in which the arms are the points of the $[0,1]$ interval and an unknown concave function describes the expectations of the arms, which is Lipschitz continuous with a known Lipschitz constant (L). I have created a modified version of Algorithm 3 which is (ε,δ)-PAC with a sample complexity of

$$\mathcal{O}\left(\frac{1}{\varepsilon^2}\left(\left(\log\frac{L}{\varepsilon}\right)^2 + \left(\log\frac{L}{\varepsilon}\right)\cdot\log\frac{1}{\delta}\right)\right).$$

I have also analyzed that case of the multi-armed bandit problem, in which there are finitely many arms and an unknown concave function describes the expectations of the arms. In this case the Median Elimination algorithm could be used to find an ε-optimal arm with probability at least $1-\delta$ with a sample complexity of

$$\mathcal{O}\left(\frac{n}{\varepsilon^2}\log\frac{1}{\delta}\right).$$

However, the algorithm I developed achieves the same result with a sample complexity of

$$\mathcal{O}\left(\frac{1}{\varepsilon^2}\left((\log n)^2 + \log n\cdot\log\frac{1}{\delta}\right)\right).$$

This can be further improved, if there is a known Δ such that $|\mu_i - \mu_{i-1}| \le \Delta$, $i = 1, 2, ..., n$, in this case the sample complexity is

$$\mathcal{O}\left(\frac{\ell^2}{\varepsilon^2} + \frac{\ell}{\varepsilon^2}\log\frac{1}{\delta}\right).$$

where $\ell = \left\lfloor\log_2\frac{n}{2\lfloor\varepsilon/\Delta\rfloor+1}\right\rfloor$.

References

[1] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.

[2] T. Lattimore and C. Szepesvári, *Bandit algorithms*. Cambridge University Press, 2020.

[3] E. Even-Dar, S. Mannor, and Y. Mansour, "Action elimination and stopping conditions for the multi-armed bandit and reinforcement learning problems," *Journal of Machine Learning Research*, vol. 7, no. 39, pp. 1079–1105, 2006.

[4] S. Mannor and J. N. Tsitsiklis, “The sample complexity of exploration in the multi-armed bandit problem,” *Journal of Machine Learning Research*, vol. 5, no. Jun, pp. 623–648, 2004.

[5] B. C. Csáji and E. Weyer, “System identification with binary observations by stochastic approximation and active learning,” in *2011 50th IEEE Conference on Decision and Control and European Control Conference*. IEEE, 2011, pp. 3634–3639.