BTP - Best Arm Identification

Preethi Malyala(200070041)
Sakshi Heda (200070071)

August 2023

Contents

1 Introduction

We work under the setting of best arm identification and introduce attributes for each arm. Given a set of arms, each characterized by multiple sub-Gaussian attributes, we aim to address the problem of identifying the optimal arm with the highest cumulative reward.
The cumulative reward of an arm is defined as the summation of rewards obtained from its individual attributes. To determine the best arm, we seek to maximize this cumulative reward.
Additionally, a subset of valid arms is established, comprising arms for which the mean reward values of all attributes exceed a predetermined threshold denoted as TH.
The primary objective of this undertaking is to effectively ascertain the optimal arm from within the set of valid arms. This optimal arm is defined as the one possessing the highest cumulative reward among the valid arms.

2 Problem Statement Formalisation

We consider set S of N arms, each having K attributes. Each attribute's distribution is described as sub-gaussian with mean values denoted as μ_{ij}, indicating the mean of the j-th attribute for the i-th arm. The cumulative reward for arm i is expressed as $\mu_i = \sum_{j=1}^{K} \mu_{ij}$.
The best arm would then be

$$F = \{i \in S : \min_j \{\mu_{ij}\} \geq TH\}$$

$$a = \arg\max_i \{\mu_i : i \in F\}$$

We work in a fixed confidence setting. Define $\delta :=$ confidence bound. We find bounds on T, the number of samples taken by an algorithm to identify the best arm with probability $\geq (1 - \delta)$

3 Proposed Solutions

Following are some of the algorithms proposed to solve the above problem

3.1 Action Elimination

A set of potential arms is maintained, which initially contains all the arms. After performing uniform exploration for a pre-defined number of times, sub-optimal arms are eliminated from this set. Sub-optimal, in this case, is decided by

1. the attribute with the smallest mean being higher than the threshold
2. the cumulative mean must be higher than that of a reference arm

The algorithm is as mentioned below:

Algorithm 1 Action Elimination

Input: confidence parameter, δ and uniform exploration parameter, r_k
Define, $\Omega_0 := [N]$
while $\Omega_k > 1$ **do**
  Sample each arm r_k times
  Define reference arm, $a = \arg\max_{i \in [N], j \in [K]} UCB[i][j]$
  $\Omega_{k+1} = \{i \in \Omega_k : LCB[a] < UCB[i] \text{ and } \min_{j \in [k]} UCB[i][j] \geq TH\}$
end while

The LCB and UCB (lower and upper confidence bounds) are calculated using Hoeffding inequality: $\mathbb{P}(|M_n - E[M_n]| \geq t) \leq 2e^{-t^2 n}$

Three different experiments were performed with this algorithm:

3.1.1 Mean of best and second best

Here, we study the accuracy and time taken by the algorithm to identify the best arm in the case that none of the arm's attributes crosses the threshold. The means of the best and second-best arms, however, get closer.

As the distance between the means of the two arms varies between 0 and 0.01, the experiment is performed 10,000 times each for three, four, and five arms, and the outcome is noted. The number of times a correct prediction was made is plotted below in terms of probability. Also, the average number of samples per arm is plotted for each case.

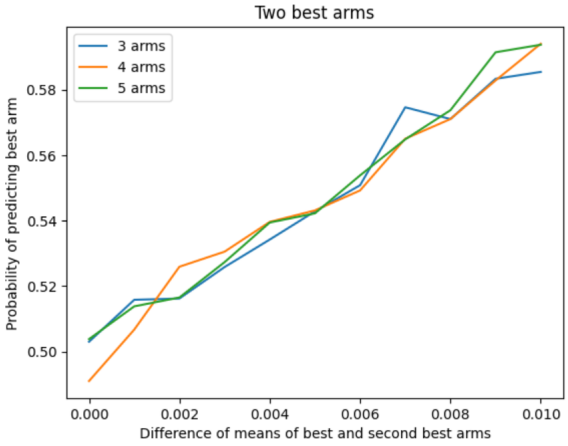


Figure 1: In each of the cases, the best arm has means [0.4, 0.7], and the best sub-optimal arm has means varying from [0.4, 0.69] to [0.4, 0.7]. The threshold is 0.2, and all the attributes of all arms are above the threshold.

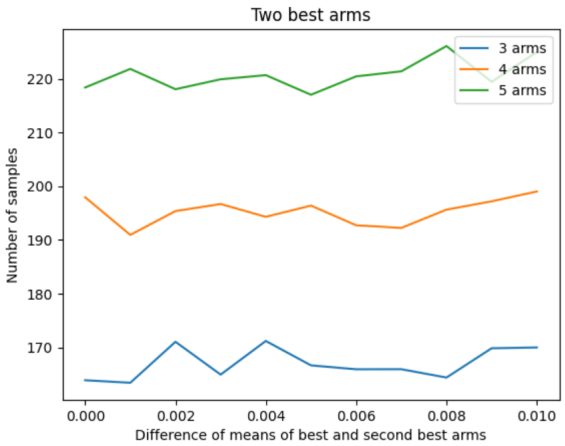


Figure 2: The means of arms are the same as in Fig 1. The total number of samples is plotted

3.1.2 One arm violates

Here, two arms have the best cumulative mean. However, one of these has a single attribute that violates the threshold criterion. The mean of this attribute is taken closer and closer to the threshold, and the performance of the algorithm is measured.

For 10,000 different instances of arms, each for three, four, and five arms, the mean of the violating arm is varied from 0 to 0.2 (0.2 is the threshold), and the number of times the algorithm predicts the right arm is plotted in terms of probability. Similar to the previous case, the average samples per arm are also plotted

3.1.3 Best arm violates threshold

The best arm, in terms of cumulative mean, has an attribute that violates the threshold. The probability of the second-best arm being selected by the algorithm is plotted, as a function of its cumulative mean

3.1.4 Reproducing Results

We chose to look at the very simple case of just n = 6 arms with linearly decreasing means: $\{1, 4/5, 3/5, 2/5, 1/5, 0\}$. All experiments were run with input confidence $\delta = 0.1$. All realizations of the arms were Gaussian random variables with mean μ_i and standard deviation $1/2$. We estimate $\mathbb{P}(I_t = i)$ at every time t by calculating a proportion of the indices pulled in the interval $[t - n + 1, t]$ and average over 5000 trials each algorithm completed.

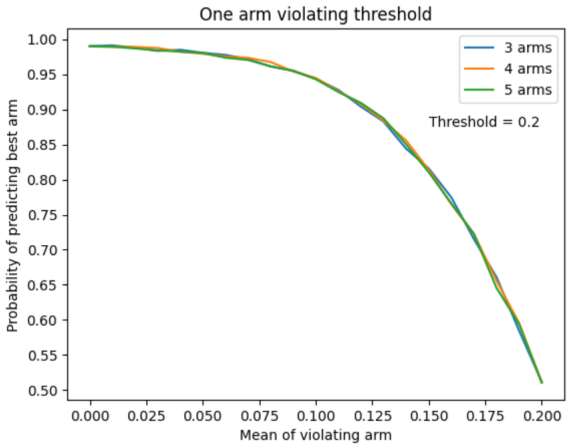


Figure 3: In each of the cases, the required arm has means [0.4, 0.7], and another arm has mean varying from [0, 1.1] to [0.2, 0.9]. The threshold is 0.2, and one of the arms gets closer to the threshold

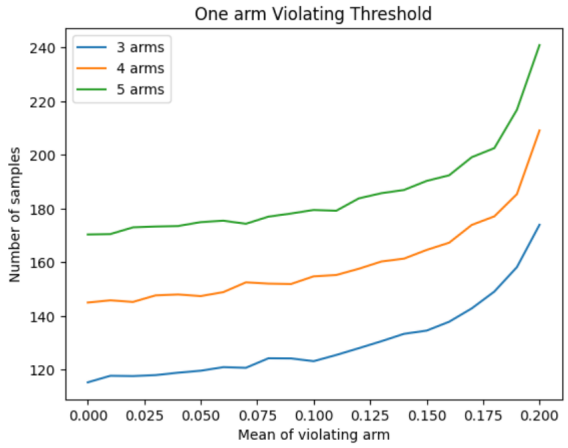


Figure 4: The means of arms are the same as in Fig 3. The total number of samples is plotted against the mean of the violating arm

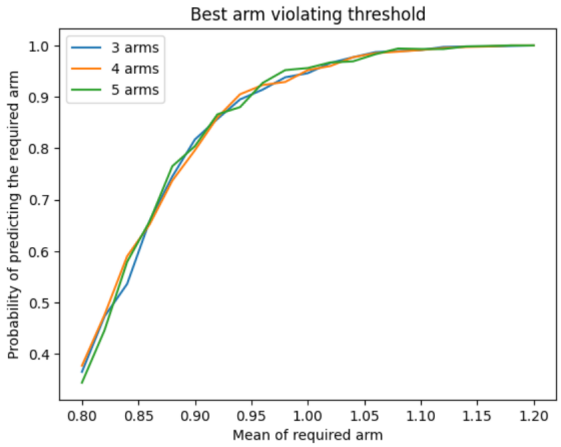


Figure 5: The best arm has mean [0.1, 0.9] (The threshold is 0.2). The second best arm has mean changing between [0.4, 0.4] and [0.4, 0.8]

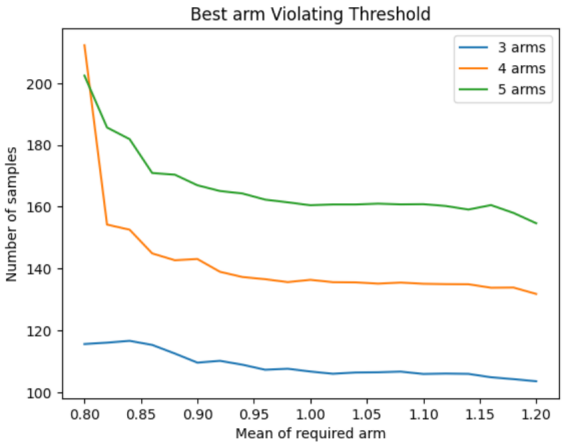


Figure 6: The means of arms are the same as in Fig.5. The total number of samples is plotted against the mean of the required arm

3.2 Modification of Variance-Aware LUCB (VA-LUCB)

Algorithm 2 Modification of VA-LUCB

Sample each of the N arms thrice
Set $\mathcal{F}_N = [N]$
for time steps $t = N+1, N+2, \ldots$ **do**
    Calculate the attribute wise sample mean for attribute j of arm i using $\hat{\mu_{ij}}(t) := \frac{1}{T_i(t)} \sum_{s=1}^{t-1} X_{s,i} \mathbb{1}\{i \in \mathcal{J}_s\}$
    Calculate the sample mean for arm i using $\hat{\mu_i}(t) := \sum_{j=1}^{m} \hat{\mu_{ij}}$
    Calculate the LCB and the UCB corresponding to each arm and attribute using $LCB(t) = \hat{\mu}(t) - \alpha(t,T)$ and $UCB(t) = \hat{\mu}(t) + \alpha(t,T)$ where $\alpha(t) := \sqrt{\frac{1}{2T} \ln(\frac{N(M+1)t^3}{2\delta})}$
    Update the good feasible set according to $\mathcal{F}_{1t} = \{i : LCB_{ij}(t) \geq \mu_{TH} \forall j\}$
    Update the feasible set according to $\mathcal{F}_t = \{UCB_{ij}(t) \geq \mu_{TH} \forall j\}$
    Find $i_t^* := \arg\max\{\hat{\mu}_i(t) : i \in \mathcal{F}_{1t}\}$
    Update the potential set $\mathcal{P}_t$ according to $\mathcal{P}_t = \{i : LCB_{i_t^*}(t) \leq UCB_i(t)\}$
    Set $i_t := \arg\max\{\hat{\mu}_i(t) : i \in \mathcal{F}_t\}$
    Set competitor arm $c_t := \arg\max\{UCB_i(t) : i \in \mathcal{F}_t, i \neq i_t\}$

Figure 7: Result as obtained in the paper

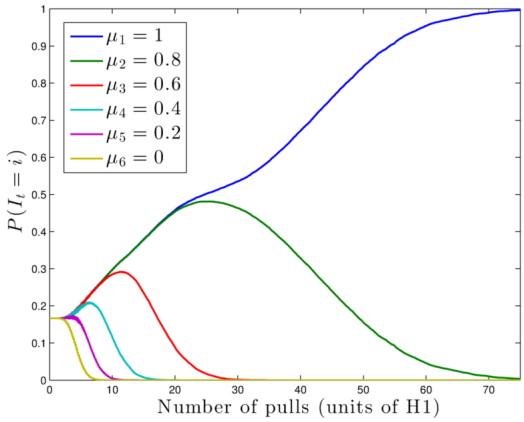
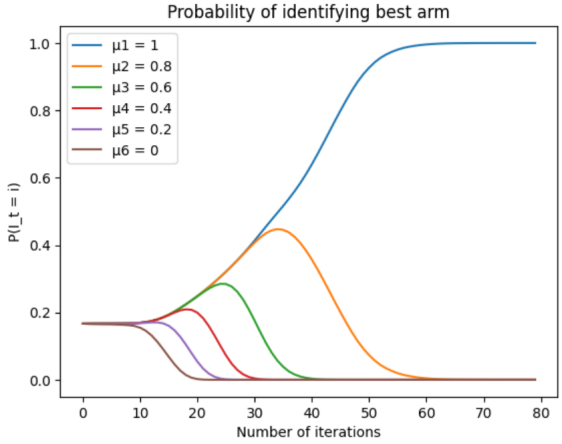


Figure 8: Reproduced results



3.2.1 Simulation Results - for 3 arms with 2 attributes

Both arms valid with means getting closer

In this case, all the three arms are valid, arm 0 having the maximum mean. The mean of arm 1 is increased and taken closer to that of arm 1. With this the success probability (predicting arm 0 as best arm) decrease and the number of samples required increases.

Two arms with same mean, one violating the threshold

In this case, both arm 0 and arm 1 has same total mean (which is more than the third one) but the arm 1 violates the threshold, the mean of one of its attributes is increased from 1 to the threshold. With this, the probability

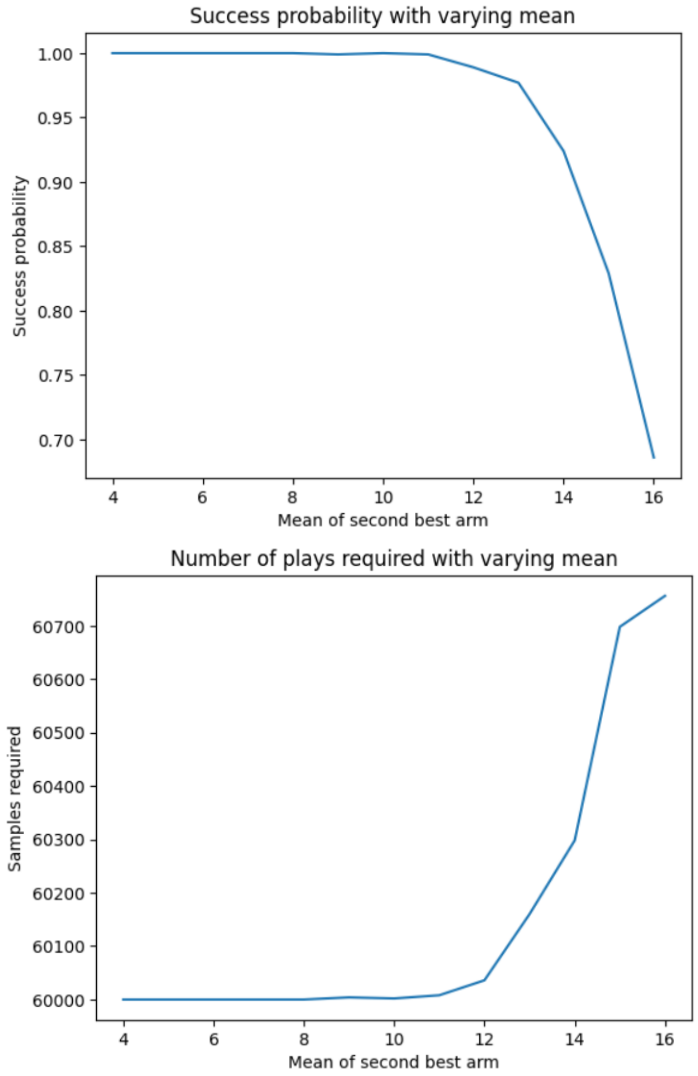


Figure 9: Here the means of all the arms are [20, 10], [x, 13] and [5, 6] with threshold 2 and x varying from 4 to 16

of predicting arm 0 as best arm decreases and the number of samples required increases.

Best arm violates threshold

In this case, arm 1 has maximum total mean but one of its attributes violates the threshold. Arm 0 has 2nd maximum mean with both the arms valid and its

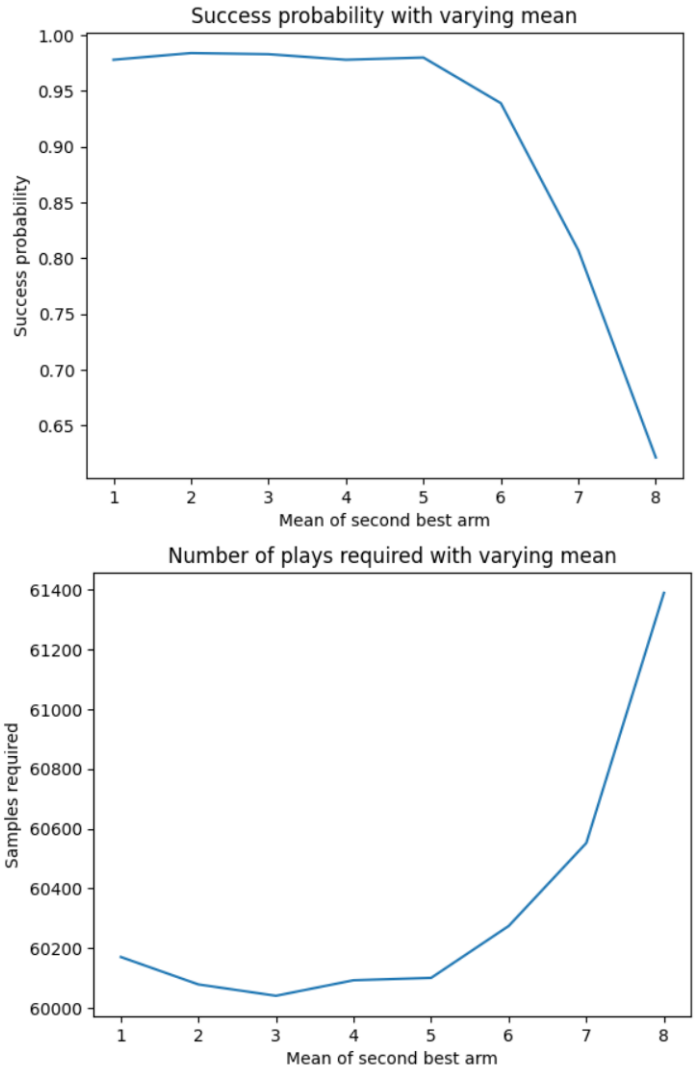


Figure 10: Here the means of all the arms are [20, 10], [x, 30-x] and [10, 10] with threshold 7 and x varying from 1 to 7

mean is increased to cross the mean of the arm 1. Here is the plot of success probability (predicting arm 0) and the number of samples required.

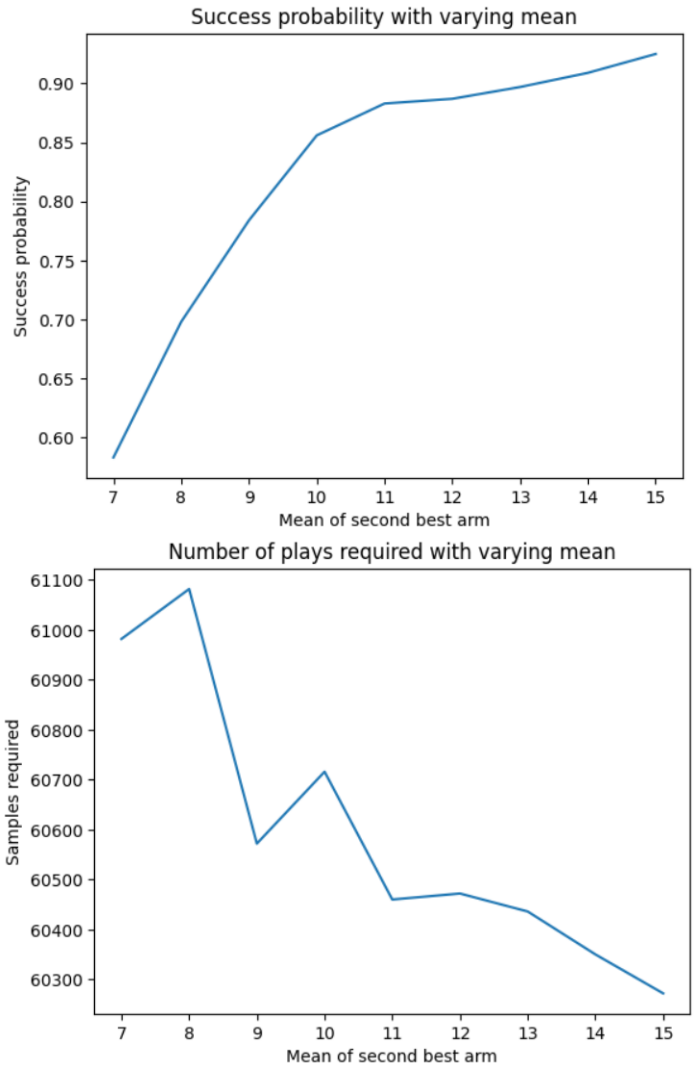


Figure 11: Here the means of all the arms are [x, 8], [5, 16] and [7, 8] with threshold 6 and x varying from 7 to 15

3.2.2 Simulation Results - for 3, 4 and 5 arms and 2 attributes

All arms are valid with the mean of one arm getting closer to the best arm In this case the mean of arm 0 is maximum and that of arm 1 is varied to get it closer to the arm 0. The success probability decreases and number of samples required increases as visible from the following graph

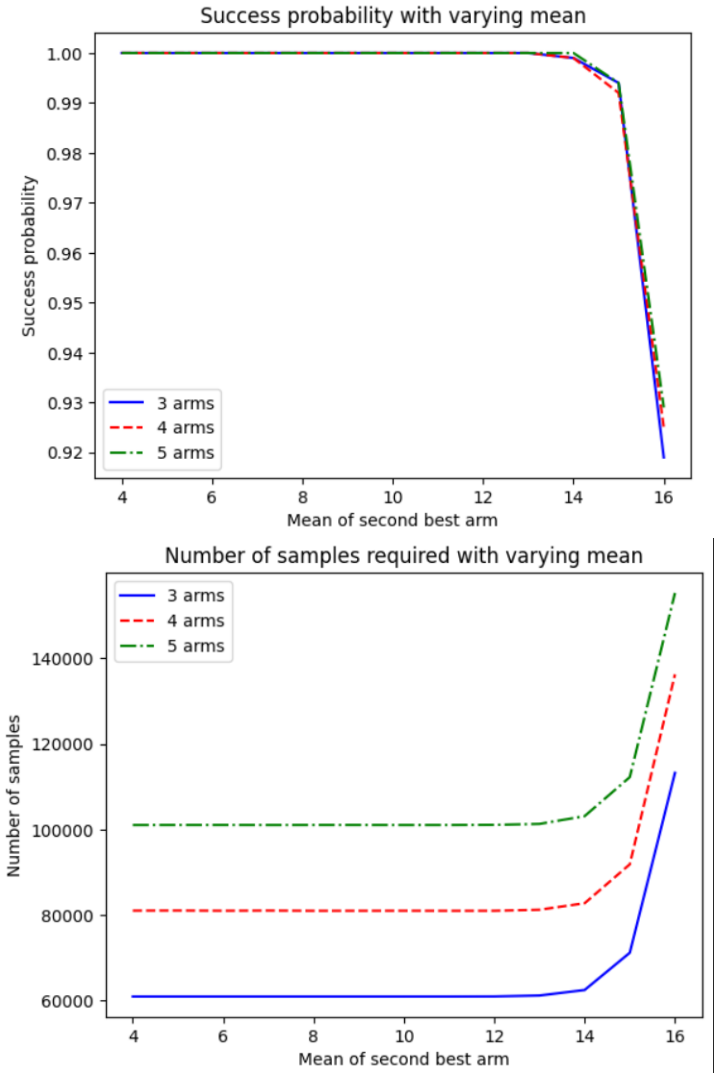


Figure 12: Here the means of all the arms are [20, 10], [x, 13], [5, 6], [7, 8] and [9, 11] with threshold 2 and x varying from 4 to 16

Best arms with same mean, one of them varying the threshold Both arm 0 and arm 1 have maximum total mean but arm 1 violates the threshold. The mean of the attribute of arm 1 is constantly increased to cross the threshold

Best arm violates threshold Arm 1 has maximum mean but one of its attributes violates the threshold. Arm 0 has 2nd maximum mean with both

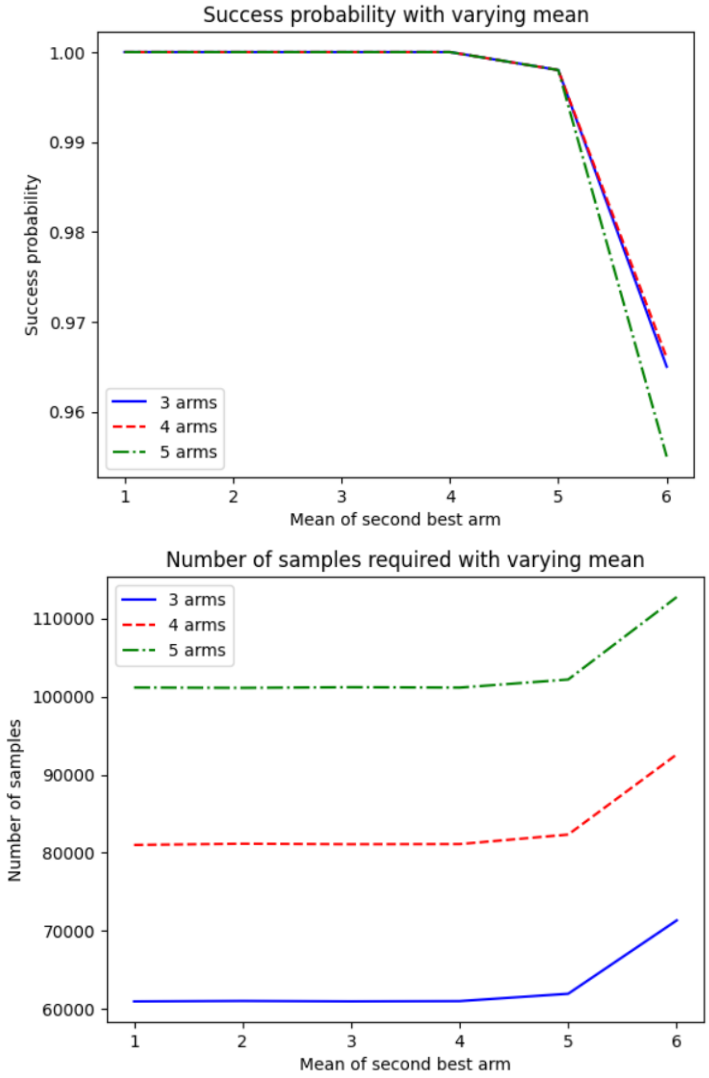


Figure 13: Here the means of all the arms are [20, 10], [x, 30-x], [10, 10], [5, 20] and [7, 15] with threshold 7 and x varying from 1 to 7

arms valid and its mean is gradually increased

3.2.3 Use of H-index

According to the paper the upper bound on algorithm's time complexity is given by $O(H_{id} \ln(\frac{H_{id}}{\delta})$ where H_{id} is the H-index value. We plotted the sample complexity against the time bound and the graph came out to be linear similar

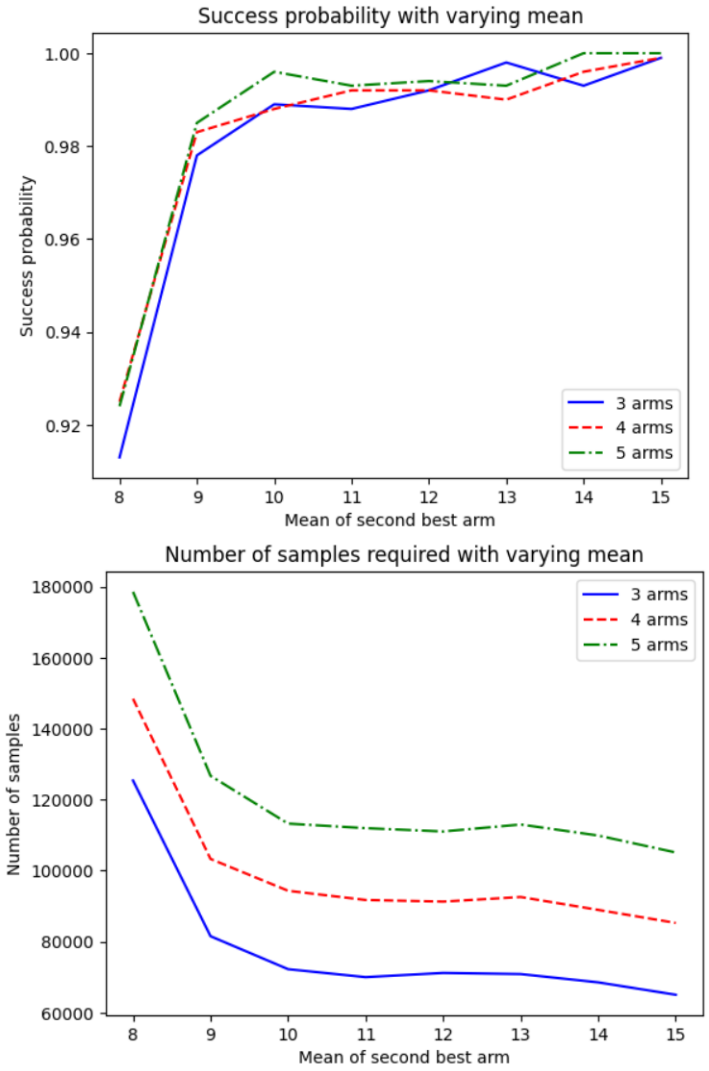
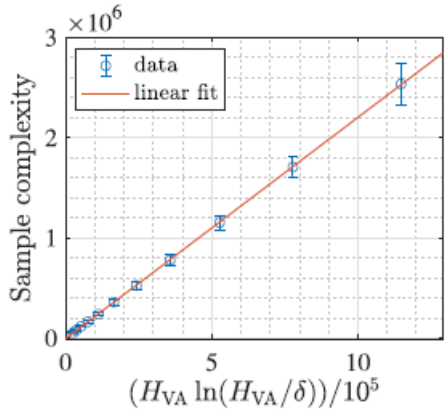


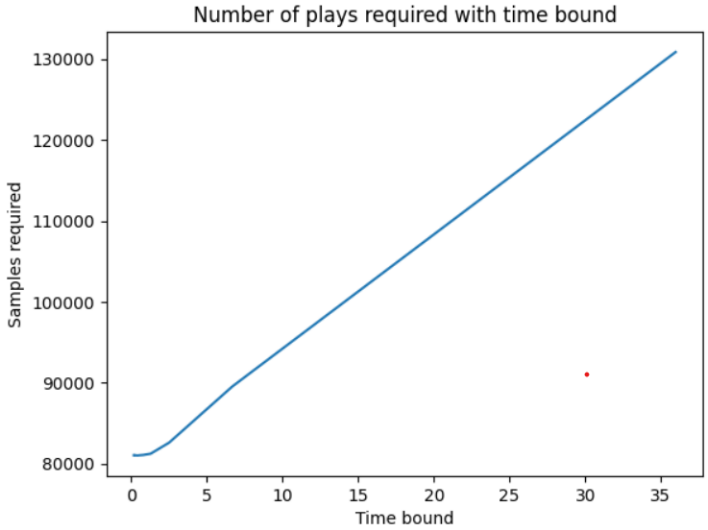
Figure 14: Here the means of all the arms are [x, 8], [5, 16], [7, 8], [5, 10] and [7, 10] with threshold 6 and x varying from 7 to 15

to that shown in the paper

4 Working with Bernoulli distribution

If the value of $p < 0.735$, the Bernoulli distribution is 1/2-subgaussian. Any Bernoulli distribution is subgaussian for variance proxy ≥ 0.849.





The pmf of Bernoulli distribution is given by $P(X = x) = \begin{cases} p & \text{if } x = 1 \\ 1 - p & \text{if } x = 0 \end{cases}$

The most important property of the Bernoulli random variable that was exploited was that it is **bounded in [0, 1].**

In this setting, each of the N arms has K attributes, with the j^{th} attribute of the i^{th} arm being sampled from a Bernoulli distribution with parameter p_{ij}. For an arm i to be valid, it must satisfy the condition $min_{j \in [K]}\{p_{ij}\} \geq TH$

The following experiments were performed with the underlying distribution being Bernoulli:

4.1 Experiment 1a

Δ_{i*} is increased, keeping other terms of H_{VA} constant. All attributes satisfy the threshold, and the mean of the best two arms gets closer

Table 1: Bernoulli parameters for individual attributes — $TH = 0.3$

Arm	Attribute 1	Attribute 2
1	x*	0.6
2	0.5	0.6
3	0.2	0.8
4	0.4	0.5
5	0.5	0.3

*x is varied from 0.6 to 0.9, thus making the problem instance increasingly easier. TH is kept at 0.3, such that none of the attributes violates the threshold. The number of samples required for the algorithm to terminate is plotted against the value of x:

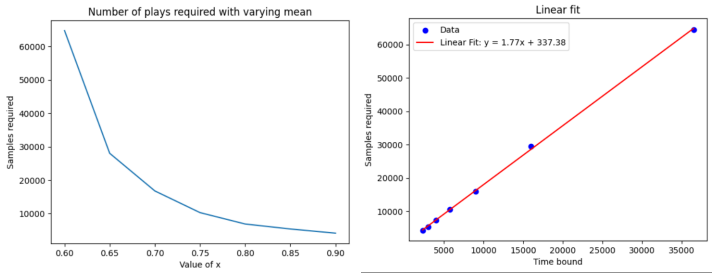


Figure 15: Experiment 1a

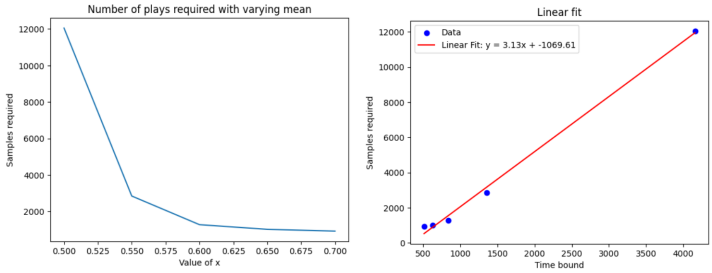
4.2 Experiment 1b

All arms (other than the best one) violate threshold. The mean of best arm varies

Table 2: Bernoulli parameters for individual attributes — $TH = 0.45$

Arm	Attribute 1	Attribute 2
1	x*	1
2	0.4	0.4
3	0.2	0.4
4	0.1	0.5
5	0.5	0.3

*x is varied from 0.5 to 0.7, thus making the problem instance increasingly easier. TH is fixed at 0.45. The number of samples required for the algorithm to terminate is plotted against the value of x:



The number of samples plotted against the time bound derived in the paper is given as follows:

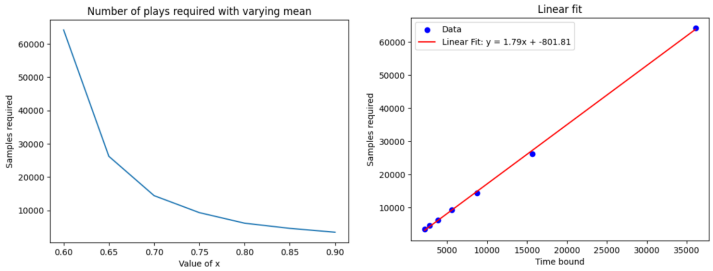
4.3 Experiment 2

The second-best arm violates the threshold. The mean of best arm varies. All the other arms are sub-optimal

Table 3: Bernoulli parameters for individual attributes — $TH = 0.4$

Arm	Attribute 1	Attribute 2
1	x*	0.8
2	0.3	1
3	0.5	0.6
4	0.4	0.5
5	0.1	0.5

*x is varied from 0.6 to 0.9, thus making the problem instance increasingly easier. TH is fixed at 0.4. The best arm remains the same (arm 1) for all values of x. The number of samples required for the algorithm to terminate is plotted against the value of x:



The number of samples plotted against the time bound derived in the paper is given as follows:

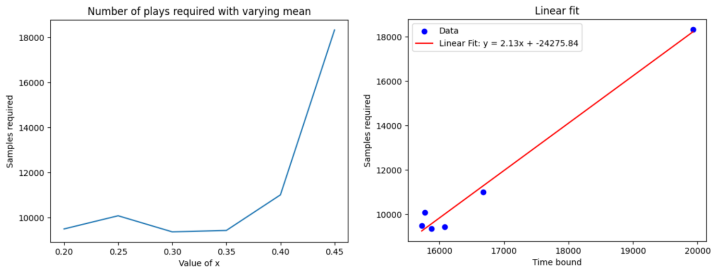
4.4 Experiment 3

The arm with the highest mean is infeasible. Its mean is varied

Table 4: Bernoulli parameters for individual attributes — $TH = 0.5$

Arm	Attribute 1	Attribute 2
1	0.55	0.6
2**	x*	1
3	0.4	0.6
4	0.4	0.5
5	0.1	0.5

*x is varied from 0.2 to 0.45 ** Arm 2 is the best in terms of mean. However, it violates the threshold (TH is set at 0.5) The best arm thus remains arm 1 for all values of x. The number of samples required for the algorithm to terminate is plotted against the value of x:



The number of samples plotted against the time bound derived in the paper is given as follows:

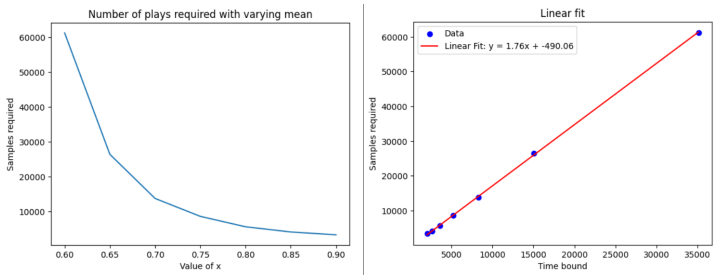
4.5 Experiment 4a

3 arms are infeasible, out of which 1 is risky, 2 are sub-optimal

Table 5: Bernoulli parameters for individual attributes — $TH = 0.4$

Arm	Attribute 1	Attribute 2
1	0.8	x
2	0.3	0.5
3	0.4	0.6
4	0.8	0.5
5	0.2	0.4

*x is varied from 0.6 to 0.9 The number of samples required for the algorithm to terminate is plotted against the value of x:



The number of samples plotted against the time bound derived in the paper is given as follows:

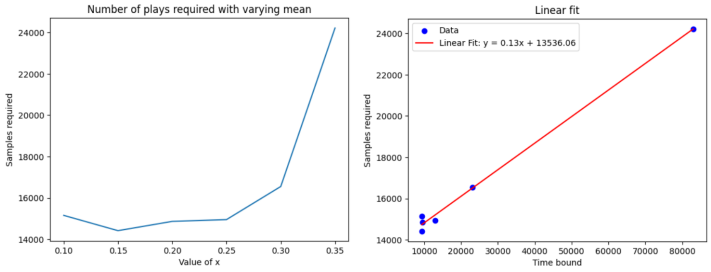
4.6 Experiment 4b

The highest mean arm is infeasible; it gets closer to the best arm

Table 6: Bernoulli parameters for individual attributes — $TH = 0.4$

Arm	Attribute 1	Attribute 2
1	0.5	0.7
2	x	0.8
3	0.3	0.5
4	0.5	0.5
5	0.2	0.4

*x is varied from 0.1 to 0.35, making the problem instance **tougher**. The number of samples required for the algorithm to terminate is plotted against the value of x:



The number of samples plotted against the time bound derived in the paper is given as follows:

5 Working with Beta distribution

The beta function is given by

$$\beta(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

where $\Gamma(a) = (a-1)!$ for integer a.
The pdf of beta distribution is given by

$$f_{a,b}(x) = \frac{x^{(a-1)}(1-x)^{b-1}}{\beta(a,b)}.$$

The values generated from beta distribution are bounded in [0, 1] which is well suited for the bounds used in our algorithm. The mean of beta distribution is given by $\mu = \frac{a}{a+b}$
Following are the results of the algorithm when applied to out problem statement with each arm giving rewards according to this distribution by varying the values

of a and b, thus varying the mean of the rewards. Same experiments as from the bernoulli case are performed for the beta distribution and following are the results.

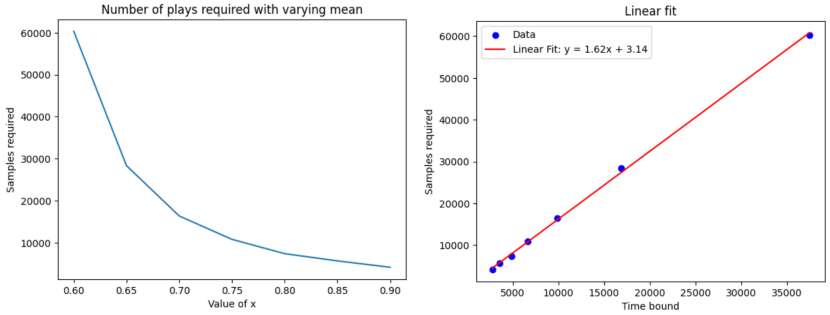


Figure 16: Experiment 1a

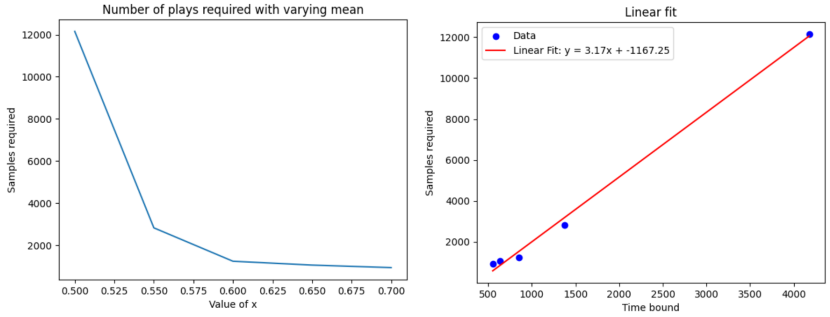


Figure 17: Experiment 1b

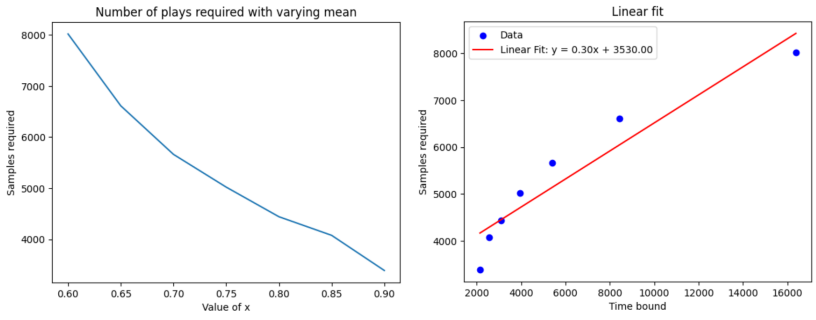


Figure 18: Experiment 2

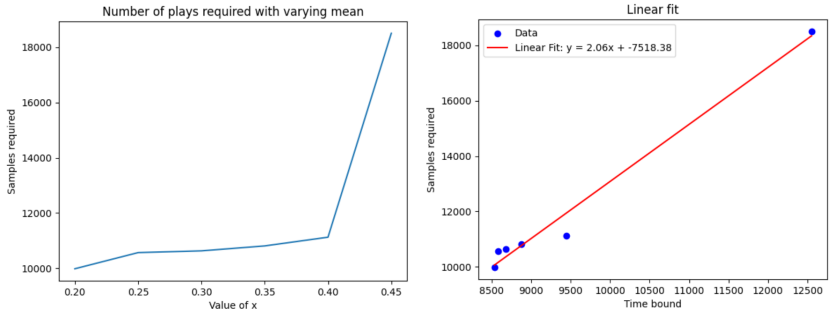


Figure 19: Experiment 3

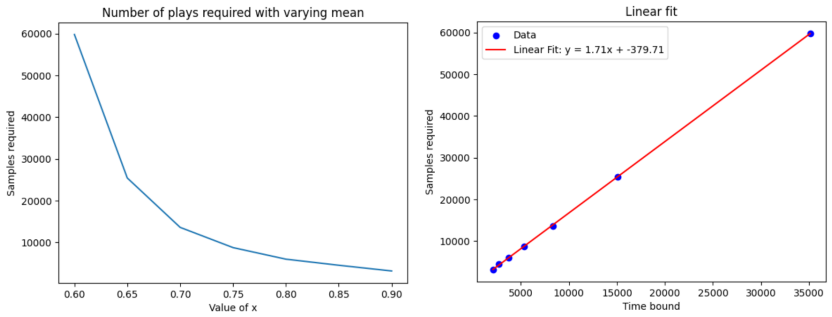


Figure 20: Experiment 4a

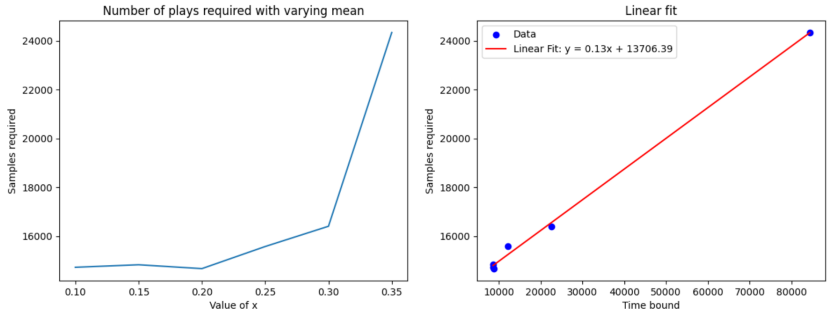


Figure 21: Experiment 4b

6 Proof of the bounds

We observe that the bound provided in the paper works well for the cases when total mean of the arms is varied but it is slightly off when focus is on varying

distance of minimum valued attribute from the threshold (corresponding to the case of varying distance of variance from the threshold in the paper).

Before getting to the upper bound of the number of samples, we define certain notations:

6.1 Based on ground truth

As mentioned in the problem statement, there are N arms, with M attributes each. Each attribute, j of arm i has a reward distribution ν_{ij}. A reward is sampled from arm i, attribute j is $X_{ij} \sim \nu_{ij}$. We consider i.i.d bounded rewards; w.l.o.g., $X_i \in [0,1]$. The expectation and variance of X_{ij} are denoted by μ_{ij} and σ_{ij}^2 respectively.

We impose a lower bound, μ_{TH} on the mean of a distribution. An instance is then defined by the $M \times N$ distributions, and the threshold: *instance* := $(\nu = \{\nu_{ij}, (i,j) \in [N] \times [M]\}, \mu_{TH})$. An arm is said to be feasible if all its attributes have means above the threshold. i.e. Arm i is feasible iff $\mu_{ij} > \mu_{TH} \forall j \in [M]$. The *instance* is said to be feasible iff $\exists i \in [N]$ s.t. arm i is feasible. The instance is infeasible otherwise. Define the feasibility flag, $f := \begin{cases} 1 & \text{instance is feasible} \\ 0 & \text{instance is infeasible} \end{cases}$

Based on the ground truth, we define the following sets :

- Feasible set $\mathcal{F} = \{i \in [N] : \min_j\{\mu_{ij}\} \geq \mu_{TH}\}$.
- Infeasible set $\bar{\mathcal{F}^C} := [N] \backslash \mathcal{F}$
- sub-optimal set $\mathcal{S} := \begin{cases} \{i \in [N] : \mu_i < \mu^\star\} & \mathcal{F} \neq \phi \\ \phi & \mathcal{F} = \phi \end{cases}$
- risky set $\mathcal{R} := [N] \backslash \mathcal{S}$

The **best feasible arm** is $i^\star := argmax\{\mu_i : i \in \mathcal{F}\}$. We want the algorithm to output $i^\star$ with high probability. We also define, $i^{\star\star} := argmax\{\mu_i : i \in \mathcal{S}\}$. Define the mean gap for arm $i \in \mathcal{S}, \Delta_i := \mu_{i^\star} - \mu_i \; if \; \mathcal{F} \neq \phi$. The mean gap for the best arm is $\Delta_{i^\star} := \mu_{i^\star} - \mu_{i^{\star\star}}$. Similarly, we define the mean gaps for each individual attribute, $\Delta_{ij}^{attr} := \mu_{ij} - \mu_{TH}$. We also define, $\Delta_i^{attr} := |\min_j \mu_{ij} - \mu_{TH}|$.

Also define, separator $\bar{\mu} := \begin{cases} (\mu_{i^\star} + \mu_{i^{\star\star}})/2 & \mathcal{F} \neq \phi \, and \, \mathcal{S} \neq \phi \\ -\infty & otherwise \end{cases}$

6.2 Based on the observed samples

Empirical mean of an attribute, $\hat{\mu_{ij}}(t) := \sum_{x=1}^{t} \frac{X_{ij}(x)}{t}$

Empirical mean of an arm, $\hat{\mu_i}(t) := \sum_{j=1}^{M} \frac{\hat{\mu_{ij}}(t)}{M}$

The confidence radius is defined as

$$\alpha(t,T) = \sqrt{\frac{1}{2T}\ln(\frac{N(M+1)t^3}{2\delta})} \tag{1}$$

We define the confidence intervals for each attribute with the lower confidence bound and the upper confidence bound:

$$L_{ij}(t) := \hat{\mu_{ij}}(t) - \alpha(t, T_i(t))$$

$$U_{ij}(t) := \hat{\mu_{ij}}(t) + \alpha(t, T_i(t))$$

Similarly, we define the confidence interval for arm i as follows:

$$L_i(t) := \hat{\mu_i}(t) - \alpha(t, T_i(t))$$

$$U_i(t) := \hat{\mu_i}(t) + \alpha(t, T_i(t))$$

Now, based on these empirically calculated values, we define the following sets:

- empirically feasible set, $\mathcal{F}_t := \{i \in [N] : L_{ij}(t) \geq \mu_{TH}, \forall j \in [M]$
- empirically almost feasible set, $\partial\mathcal{F}_t := \{i \in [N] : U_{ij}(t) \geq \mu_{TH} > L_{ij}(t), \exists j \in [M]$
- possibly feasible set, $\bar{\mathcal{F}}_t := \mathcal{F} \cup \partial\mathcal{F}$
- empirically infeasible set, $\bar{\mathcal{F}}_t^c := [N] \backslash \bar{\mathcal{F}}_t = \{i \in [N] : U_{ij}(t) < \mu_{TH}, \forall j \in [M]$
- potential set, $\mathcal{P}_t := \begin{cases} \{i \in [N] : L_{i_t^\star}(t) \leq U_i(t) & \mathcal{F} \neq \phi \\ [N] & \mathcal{F} = \phi \end{cases}$
- $\mathcal{S}_t := \{i : U_i(t) < \bar{\mu}\}$
  $\mathcal{R}_t := \{i : L_i(t) > \bar{\mu}\}$
  $\mathcal{N}_t := [N] \setminus (\mathcal{S}_t \cup \mathcal{R}_t) = \{i : L_i(t) \leq \bar{\mu} \leq U_i(t)\}$

We define three arms:

- best empirically feasible arm, $i_t^\star := argmax\{\hat{\mu_i}(t) : i \in \mathcal{F}_t\}$
- $i_t := \text{argmax}\left\{\hat{\mu}_i(t) : i \in \overline{\mathcal{F}}_t\right\}$
- $c_t := \text{argmax}\left\{U_i^\mu(t) : i \in \overline{\mathcal{F}}_t, i \neq i_t\right\}$

$$H_{VA} := \frac{1}{\min\{\frac{\Delta_{i^*}}{2}, 2\Delta_{i^*}^{\text{attr}}\}^2} + \sum_{i \in \mathcal{F} \cap \mathcal{S}} \frac{1}{(\frac{\Delta_i}{2})^2} + \sum_{i \in \bar{\mathcal{F}^C} \cap \mathcal{R}} \frac{1}{(\frac{\Delta_i^{\text{attr}}}{2})^2} + \sum_{i \in \bar{\mathcal{F}^C} \cap \mathcal{S}} \frac{1}{\max\{\frac{\Delta_i}{2}, 2\Delta_i^{\text{attr}}\}^2} \tag{2}$$

Theorem 1 (Upper bound)**.** *Given an instance and a confidence parameter δ, with probability at least $1-\delta$, the algorithm succeeds and terminates in $\mathcal{O}(H_{VA} \ln \frac{H_{VA}}{\delta})$*

Proof of Theorem 1. We define the following events with $\hat{\mu}$ being the empirical values and $\alpha(t, T_i(t))$ be the confidence radii defined according to (1)

$$E_i^{\text{tot}}(t) = \{|\hat{\mu_i}(t) - \mu_i| \leq \alpha(t, T_i(t)\} \tag{3}$$

$$E_{ij}^{\text{attr}}(t) = \{|\hat{\mu_{ij}}(t) - \mu_{ij}| \leq \alpha(t, T_i(t)\} \tag{4}$$

$$E_i(t) = E_i^{\text{tot}} \cap (\bigcap_{j=1}^{m} E_{ij}^{\text{attr}}(t)) \tag{5}$$

$$E(t) := \bigcap_{i=1}^{n} E_i(t) \text{ and } E := \bigcap_{t \geq 2} E(t) \tag{6}$$

From Lemma 1 and Lemma 2, we can say that if the algorithm stops then it will succeed with probability atleast $1 - \delta/2$.

□

Lemma 1. *Event E is defined as in* (6)*, then E occurs with probability at least $1 - \delta/2$*

Lemma 2. *Given an instance with confidence parameter δ, on event $E(\tau)$ and that the algorithm terminates, then*

- *if the instance is infeasible, then $\hat{f} = f = 0$*
- *if the instance is feasible, then $i_{out} = i^*, \hat{f} = f = 1$*

This means that if event E occurs, the algorithm is guaranteed to give correct result

Lemma 3. *On the event $E(t)$, if the algorithm does not terminate, then atleast one of the following statements hold:*

- $i_t \in (\partial\mathcal{F}_t \backslash \mathcal{S}_t) \cup (\mathcal{F}_t \cap \mathcal{N}_t)$
- $c_t \in (\partial\mathcal{F}_t \backslash \mathcal{S}_t) \cup (\mathcal{F}_t \cap \mathcal{N}_t)$

Next, we define $u_i(t)$ to be the smallest number of pulls required for arm i till time t such that the confidence radii ($\alpha(t, T_i(t))$ for this arm is less than or equal to Δ_i, i.e.

$$u_i(t) := \lceil \frac{1}{2\Delta_i^2} \ln \frac{N(M+1)t^3}{2\delta} \rceil \tag{7}$$

and we define $v_i(t)$ as the smallest number of pulls required for the confidence radii to be less than or equal to Δ_i^{attr}, i.e,

$$v_i(t) := \lceil \frac{1}{2(\Delta_i^{\text{attr}})^2} \ln \frac{N(M+1)t^3}{2\delta} \rceil \tag{8}$$

Lemma 4. *In the above algorithm, we can say that*

- *For i^*,*

$$\mathbb{P}[T_{i^*}(t) > 16u_{i^*}(t), i^* \notin \mathcal{R}_t] \leq \frac{2\delta}{(\frac{\Delta_{i^*}}{2})^2 N(M+1)t^3} =: A_1(i^*)$$

- *For any arm i in suboptimal set,*

$$\mathbb{P}[T_i(t) > 16u_i(t), i \notin \mathcal{S}_t] \leq \frac{2\delta}{(\frac{\Delta_i}{2})^2 N(M+1)t^3} =: A_2(i)$$

- *For any arm i in feasible set,*

$$\mathbb{P}[T_i(t) > 4v_i(t), i \notin \mathcal{F}_t] \leq \frac{2\delta}{2(\Delta_i^{attr})^2 N(M+1)t^3} =: A_3(i)$$

- *For any infeasible arm i,*

$$\mathbb{P}[T_i(t) > 4v_i(t), i \notin \bar{\mathcal{F}}_t^C] \leq \frac{2\delta}{2(\Delta_i^{attr})^2 N(M+1)t^3} =: A_4(i)$$

Lemma 5. *Let $t^* = 152H_{VA}\ln(H_{VA}/\delta)$. Then for any $t > t^*$, the probability that this algorithm does not terminate is less than $5\delta/t^2$*

Proof of Lemma 1. We will calculate the probability of the event E as follows:

$$\begin{aligned}
\mathbb{P}(E) &= 1 - \mathbb{P}(\overline{E}) \\
&= 1 - \mathbb{P}(\bigcup_{t\geq 2}\bigcup_{i=1}^{n}(\overline{E_i^{\text{tot}}}(t)\bigcup_{j=1}^{m}\overline{E_{ij}^{\text{attr}}(t)})) \\
&\geq 1 - \sum_{t\geq 2}\sum_{i=1}^{n}(M+1)2\exp(-2(\alpha(t,T_i(t)))^2 T_i(t)) \\
&\geq 1 - \frac{\delta}{2}
\end{aligned}$$

□

Remark 1. *Some remarks:*

1. *Lemma 1 proves that event E occurs with a high probability if the algorithm stops*
2. *Lemma 2 proves that if event E occurs, then algorithm will give the correct answer*
3. *Lemma 3 proves that algorithm will NOT stop if "some conditions" (given in the lemma statement) occur*

4. *Lemma 4 and 5 prove that these "some conditions" occur with low probability*

5. *Therefore, using lemma 4, 5, and 3, we can say algorithm terminates with high probability*

6. *And lemma 1 and 2 prove the correctness of the algorithm assuming it terminates*