Constrained Best Arm Identification in Grouped Bandits

Sakshi Heda*, Malyala Preethi Sravani*, and Sharayu Moharir
Department of Electrical Engineering, IIT Bombay
sakshiheda.iitb@gmail.com, preethimalyala@gmail.com, sharayum@ee.iitb.ac.in

***Abstract*—We study a grouped bandit setting where each arm comprises multiple independent sub-arms referred to as attributes. Each attribute of each arm has an independent stochastic reward. We impose the constraint that for an arm to be deemed feasible, the mean reward of all its attributes should exceed a specified threshold. The goal is to find the arm with the highest mean reward averaged across attributes among the set of feasible arms in the fixed confidence setting. We propose a confidence interval-based policy to solve this problem and provide analytical guarantees for the policy. We compare the performance of the proposed policy with that of two suitably modified versions of action elimination via simulations.**

***Index Terms*—Stochastic multi-armed bandits, best arm identification, fixed confidence setting**

I. Introduction

Many amenities are a collection of independent services and each customer of such a service may use one or more at a time. For instance, a typical auto garage may offer car wash services, AC servicing, tyre and wheel care services, car inspections, etc. Similarly, a typical salon offers hair services, skin services, nail services, make-up services, etc. To evaluate such services, it makes sense to have customers rate each service separately and maintain a rating for each service. A reasonable metric for evaluating such a service as a whole is the (weighted) average of the ratings of the different services. In addition, for an amenity to be deemed acceptable, it may be desirable that the ratings for each service exceed a threshold.

Motivated by this, we consider a grouped bandit setting where each arm is a group of independent sub-arms. We refer to these sub-arms as attributes. All the attributes of all arms are modeled as an independent stochastic process with a corresponding mean reward. A group is said to be feasible if the mean reward of all its attributes exceeds a given threshold. In this work, we focus on the problem of identifying the feasible group with the highest mean reward averaged across all its attributes in the fixed confidence setting [1]. We consider the setting where the learning agent can choose to sample a specific attribute of a specific arm.

A. Our Contributions

We propose a policy that determines which arm-attribute pair(s) to sample at each round. We provide analytical performance guarantees for this policy. Further, we empirically compare the performance of the proposed policy with suitably adapted versions of a widely studied policy – action elimination. Our numerical results show that our algorithm outperforms these other algorithms.

B. Related Work

The best arm identification problem for multi-armed bandits has been widely studied. The two settings of this problem that have received the most attention are the fixed confidence setting [1] and the fixed budget setting [2]. As mentioned above, in this work, we focus on the former. In the fixed confidence setting, we are given a confidence parameter $\delta \in [0, 1]$ as an input parameter. The goal of the learner is to identify the best arm with a probability of at least $1 - \delta$ using as few samples of the arms as possible. A detailed survey of various algorithms proposed for this setting can be found in [3].

The two key features of the problem we are interested in are: *(i)* each arm is a group of independent sub-arms, and *(ii)* identifying the best arm among those which satisfy a "feasibility" constraint. We now discuss existing literature that looks at the best arm identification problem with these two features.

A version of the grouped arm problem is the focus of [4], where the goal is to identify the group with the highest minimum mean reward. Another grouped arm problem is studied in [5], where authors discuss an algorithm to identify the best arm in each group. In [6], the authors lay emphasis on identifying the best m arms that attain the highest rewards out of the group of arms. Another related problem is called the categorized multi-armed bandits [7]. Here, arms are grouped into different categories with an existing order between these categories, and the knowledge of the group structure is known. The goal is to find the overall best arm.

Multiple works have added a feasibility constraint to the best arm identification problem, where the learning agent is expected to select an optimal arm that satisfies a feasibility rule. In [8], the authors consider the best arm identification problem with linear and then monotonic safety constraints. A method to solve a feasibility-constrained best-arm identification (FC-BAI) problem for a general feasibility rule is shown in [9]. In [10], the authors solve the FC-BAI problem with a constraint on arm mean using the track and stop technique [1]. In [11], the feasibility constraint is in terms of the variance of an arm. Under this constraint, an arm is feasible if its variance is below a given threshold. We use ideas from [11] in the design and analysis of our algorithm.

This work is supported by a SERB grant on Leveraging Edge Resources for Service Hosting.

* Equal contribution

TABLE I: Illustrative Example

Arm	Attribute 1	Attribute 2	Mean Reward
1	0.6	0.4	0.5
2	0.2	1	0.6
3	0.4	0.4	0.4

TABLE II: Notation

Notation	Description
$X_{ij}(t)$	Reward for arm i, attribute j in round t
μ_{ij}	Mean reward of arm i, attribute j
μ_i	Mean reward of arm i
$\mathcal{F}$	Feasible Set
μ_{TH}	Threshold
f	Feasibility flag
$i^\star$	Index of the best feasible arm

II. Problem Setting

There are N arms and $[N] = \{1, 2, \cdots, N\}$ denotes the set of all arms. Each arm has M attributes, denoted by $[M] = \{1, 2, \cdots, M\}$. The reward corresponding to each attribute of each arm is modelled as an independent stochastic process. We use ν_{ij} to denote the distribution of attribute $j \in [M]$ of arm $i \in [N]$. The reward of attribute j of arm i in round t is a stochastic random variable denoted by $X_{ij}(t) \sim \nu_{ij}$. We define $\mu_{ij} = \mathbb{E}[X_{ij}(t)]$.

Further, we are given a threshold, μ_{TH}, which determines the feasibility of each arm. An arm i is said to be feasible if and only if the mean reward of each of its attributes is at least μ_{TH}. We define the set of feasible arms, referred to as *Feasible Set*, denoted by $\mathcal{F}$ as follows:

$$\mathcal{F} \coloneqq \{i \in [N] : \min_j \mu_{ij} \geq \mu_{\text{TH}}\}. \tag{1}$$

A problem instance is said to be feasible if $\mathcal{F} \neq \emptyset$ and is called infeasible otherwise. We define the feasibility flag f as follows:

$$f \coloneqq \begin{cases} 1 & \text{if the instance is feasible} \\ 0 & \text{otherwise.} \end{cases}$$

For a feasible instance, the best arm $i^\star$ is defined as the arm with the highest mean reward, μ_i in $\mathcal{F}$. An arm's mean reward, here denotes the average of the mean rewards of all of its attributes. i.e.,

$$i^\star \coloneqq \arg\max_{i \in \mathcal{F}} \mu_i, \text{ where, } \mu_i \coloneqq \left(\sum_{j=1}^{M} \mu_{ij}\right)/M.$$

For example, consider the problem instance in Table I. Here, we have three arms, each with two attributes. Let $\mu_{\text{TH}} = 0.3$. In this case, Arm 1 and Arm 3 are feasible, with Arm 1 being the best feasible arm. Arm 2 has the highest mean reward but is infeasible since the mean reward for Attribute 1 is less than μ_{TH}.

We assume that the best arm, if it exists, is unique. The rewards are considered to be bounded in $[0, 1]$. The algorithmic challenge for the learner is to decide which arm-attribute pairs to play in each round. Table II summarizes the notation used in the section.

III. Our Algorithm: Confidence Set Sampling

We propose an LUCB-style algorithm [11], where we sample multiple arms in each round. Note that we have an extra degree of freedom as we can choose the attributes to be sampled. We divide the arm-attribute pairs into subsets based on their potential feasibility and explore arms for which we are still determining the feasibility. We also explore the arms with assured feasibility to get tighter confidence bounds for their mean rewards. We stop when we ascertain that the mean reward of the current best feasible arm is greater than that of any other feasible arm.

The algorithm starts with a uniform exploration for each attribute of each arm. Rounds are indexed by t, and the total number of pulls till round t is denoted by $k(t)$. Let $\mathcal{J}_t$ denote the set of arm-attribute pairs pulled in round t. We define $T_{ij}(t)$ as the number of samples of attribute j of arm i taken till round t. Similarly, $T_i(t)$ is the number of times the least explored attribute of arm i is sampled till time t. Formally,

$$T_{ij}(t) \coloneqq \sum_{s=1}^{t-1} \mathbb{1}_{(i,j) \in \mathcal{J}_s}, \; T_i(t) \coloneqq \min_{j \in [M]} T_{ij}(t).$$

The empirical mean of the reward of attribute j of arm i is denoted by $\hat{\mu}_{ij}(t)$. The empirical average reward of arm i is denoted by $\hat{\mu}_i(t)$. Formally,

$$\begin{aligned} \hat{\mu}_{ij}(t) &\coloneqq \frac{1}{T_{ij}(t)} \sum_{s=1}^{t-1} X_{ij}(s) \mathbb{1}_{(i,j) \in \mathcal{J}_s}, \\ \hat{\mu}_i(t) &\coloneqq \left(\sum_{j=1}^{M} \hat{\mu}_{ij}(t)\right)/M. \end{aligned} \tag{2}$$

Confidence radii for the arms and attributes is defined as

$$\alpha(t, T(t), k(t)) \coloneqq \sqrt{\frac{1}{2T(t)} \ln\left(\frac{4NM(k(t))^3}{\delta}\right)},$$

where δ is the confidence parameter and $T(t)$ corresponds to the arm or arm-attribute pair for which we are calculating the confidence radii.

We define the confidence intervals for each attribute with the lower confidence bound (LCB, denoted by $L_{ij}(t)$) and the upper confidence bound (UCB, denoted by $U_{ij}(t)$) as follows:

$$\begin{aligned} L_{ij}(t) &\coloneqq \hat{\mu}_{ij}(t) - \alpha(t, T_{ij}(t), k(t)), \\ U_{ij}(t) &\coloneqq \hat{\mu}_{ij}(t) + \alpha(t, T_{ij}(t), k(t)). \end{aligned} \tag{3}$$

Similarly, we define the confidence interval for arm i with the lower confidence bound (LCB, denoted by $L_i(t)$) and the upper confidence bound (UCB, denoted by $U_i(t)$) as follows:

$$\begin{aligned} L_i(t) &\coloneqq \hat{\mu}_i(t) - \alpha(t, T_i(t), k(t)), \\ U_i(t) &\coloneqq \hat{\mu}_i(t) + \alpha(t, T_i(t), k(t)). \end{aligned} \tag{4}$$

Based on these confidence intervals, we define the following subsets for the arm-attribute pairs for round t:

1) *Perfectly Feasible Attribute Set:* the set of arm-attribute pairs whose lower confidence bound is above the threshold μ_{TH}. Formally,

$$\mathcal{F}^A_{Pt} \coloneqq \{(i, j) \in [N] \times [M] : L_{ij}(t) \geq \mu_{\text{TH}}\}.$$

2) *Almost Feasible Attribute Set:* the set of arm-attribute pairs whose lower confidence bound is less than the threshold μ_{TH} and the upper confidence bound is higher than the threshold μ_{TH}. In other words, the threshold lies within the confidence interval for these arm-attribute pairs. Formally,

$$\partial\mathcal{F}_t^A \coloneqq \{(i,j) \in [N]\times[M] : L_{ij}(t) < \mu_{\text{TH}} \leq U_{ij}(t)\}.$$

3) *Feasible Attribute Set:* the union of the *Perfectly Feasible Attribute Set* and the *Almost Feasible Attribute Set* and is denoted by $\mathcal{F}_t^A$.

Based on the confidence intervals, we define the following subsets of the set of all arms:

1) *Perfectly Feasible Set:* the set of arms with all the attributes in the *Perfectly Feasible Attribute Set*. Formally,

$$\mathcal{F}_{Pt} \coloneqq \{i \in [N] : (i,j) \in \mathcal{F}_{Pt}^A,\ \forall j \in [M]\}.$$

2) *Feasible Set:* the set of arms with all the attributes in the *Feasible Attribute Set*, i.e., all the attributes have UCB higher than μ_{TH}. Formally,

$$\mathcal{F}_t \coloneqq \{i \in [N] : (i,j) \in \mathcal{F}_t^A,\ \forall j \in [M]\}.$$

3) *Almost Feasible Set:* the set of arms which are in the *Feasible Set* but not in the *Perfect Feasible Set* and is denoted by $\partial\mathcal{F}_t$.
4) *Potential Set:* the set of arms with UCB higher than the LCB of the empirically best arm. Formally,

$$\mathcal{P}_t \coloneqq \begin{cases} \{i \in [N] : L_{i_t^\star}(t) \leq U_i(t)\} & \mathcal{F} \neq \phi \\ [N] & \mathcal{F} = \phi, \end{cases} \quad (5)$$

where $i_t^\star$ is the best arm from *Perfect Feasible Set*

These notations are summarized in Table III[1].

TABLE III: Notation used in algorithm.

Notation	Description
$k(t)$	Total number of pulls
$T_{ij}(t)$	Number of pulls - arm i, attribute j
$T_i(t)$	Number of pulls of arm i
$\hat{\mu}_{ij}(t)$	Empirical mean reward of arm i, attribute j
$\hat{\mu}_i(t)$	Empirical mean reward of arm i
$\alpha(t)$	Confidence radii
$L_{ij}(t), U_{ij}(t)$	Attribute confidence bounds
$L_i(t), U_i(t)$	Arm confidence bounds
$\mathcal{F}_{Pt}^A$	Perfectly feasible attribute set
$\partial\mathcal{F}_t^A$	Almost feasible attribute set
$\mathcal{F}_t^A$	Feasible attribute set
$\mathcal{F}_{Pt}$	Perfectly feasible arm set
$\partial\mathcal{F}_t$	Almost feasible arm set
$\mathcal{F}_t$	Feasible arm set
$\mathcal{P}_t$	Potential set
$\hat{f}$	Empirical feasibility flag
i_t	Best arm from perfect feasible set
$i_t^\star$	Best arm from feasible set

[1]Parameters that evolve over rounds are defined using the notation $*(t)$ or $*_t$.

Algorithm 1 Confidence Set Sampling (CSS-LUCB)

$t \leftarrow 1, T_{ij}(t) \leftarrow 1\ \forall\ (i,j) \in [N]\times[M], T_i(t) \leftarrow 1\ \forall\ i$
Initialize $\partial\mathcal{F}_t^A$, $\mathcal{F}_t^A$, $\mathcal{F}_{Pt}$, $\partial\mathcal{F}_t$, $\mathcal{F}_t, \mathcal{P}_t$ as ϕ
Sample each of the N arms once
Set $\mathcal{F}_{M\times N} = [N]$
for time steps $t > M \times N$ **do**
  Calculate $\hat{\mu}_{ij}$, $\hat{\mu}_i$ $\forall i, j$ using (2)
  Calculate confidence bounds using (3) and (4)
  Update $\partial\mathcal{F}_t^A$, $\mathcal{F}_t^A$, $\mathcal{F}_{Pt}$, $\partial\mathcal{F}_t$, $\mathcal{F}_t$ according to III
  Find $i_t^\star \coloneqq \arg\max\{\hat{\mu}_i(t) : i \in \mathcal{F}_{Pt}\}$
  Update $\mathcal{P}_t$ according to (5)
  Set $i_t \coloneqq \arg\max\{\hat{\mu}_i(t) : i \in \mathcal{F}_t\}$
  if $\mathcal{F}_t \cap \mathcal{P}_t = \phi$ **then**
    if $\mathcal{F}_t \neq \phi$ **then**, set $i_{out} = i_t$, $\hat{f} = 1$
    else Set $\hat{f} = 0$
    end if
    break
  end if
  if $|\mathcal{F}_t| = 1$ **then**
    Pull (i_t, j) such that $(i_t, j) \in \mathcal{F}_t^A$
  else
    Find i_t and c_t
    Pull (i,j) such that $i \in \{i_t, c_t\}$, $(i,j) \in \mathcal{F}_t^A$
  end if
end for

A. Sampling Criteria

We consider two cases. The first case is when a single arm is in the *Feasible Set.* In this case, we find i_t defined as the arm with the highest empirical mean reward from the *Feasible Set.* i.e., $i_t \coloneqq \arg\max\{\hat{\mu}_i(t) : i \in \mathcal{F}_t\}$.

We then check the arm-attribute pairs in the *Feasible Attribute Set* and pull all the attributes of arm i_t in this set. In the second case, i.e., when there is more than one arm in the *Feasible Set,* i_t is defined as the best arm from the *Feasible Set,* and c_t is the *potentially competitor arm*, that is an arm other than i_t from the *Feasible Set,* which has the highest UCB. It is defined as follows:

$$c_t \coloneqq \arg\max\{U_i(t) : i \in \mathcal{F}_t, i \neq i_t\}.$$

We then check the *Feasible Attribute Set* and pull all the arm-attribute pairs corresponding to arms i_t and c_t.

B. Stopping Criteria

The algorithm stops when there are no competitor arms to be pulled, that is, the set $\mathcal{F}_t \cap \mathcal{P}_t = \phi$. We then check the feasible set; if $\mathcal{F}_t = \phi$, then the given instance is declared to be infeasible, and the feasibility flag $\hat{f}$ is set to 0. Otherwise, the feasibility flag is set to 1, and the arm i_t is declared the best feasible arm. The pseudo-code of the above algorithm is given in Alg 1.

IV. Analytical Performance Guarantees

In this section, we provide analytical guarantees for the performance of our algorithm. We define the following sets based on the ground truth:

1) *Sub-optimal Set:* the set of arms with mean reward less than that of the best feasible arm. Formally,

$$\mathcal{S} := \begin{cases} \{i \in [N] : \mu_i < \mu_{i^\star}\} & \mathcal{F} \neq \phi \\ \phi & \mathcal{F} = \phi. \end{cases}$$

2) *Risky Set:* the set of arms with mean reward more than the best feasible arm. Note that by definition, all these arms are infeasible. Formally,

$$\mathcal{R} := [N] \setminus \{\mathcal{S} \cup \{i^\star\}\}.$$

We define $i^{\star\star} := \arg\max_{i \in \mathcal{S}} \mu_i$, and for all $i \in \mathcal{S} \setminus \{i^\star\}$, $\Delta_i := |\mu_{i^\star} - \mu_i|$. Further,

$$\Delta_{i^\star} := |\mu_{i^\star} - \mu_{i^{\star\star}}|. \tag{6}$$

Similarly, $\Delta_{ij}^{\text{attr}} := |\mu_{ij} - \mu_{\text{TH}}|$. We also define $\Delta_i^{\text{attr}} := |\min_j \mu_{ij} - \mu_{\text{TH}}|$.

Further, the separator $\bar{\mu}$ is defined as follows:

$$\bar{\mu} := \begin{cases} (\mu_{i^\star} + \mu_{i^{\star\star}})/2 & \mathcal{F} \neq \phi \text{ and } \mathcal{S} \neq \phi \\ -\infty & \text{otherwise.} \end{cases} \tag{7}$$

Next, we define the hardness index of a problem instance, denoted by H_{id} as

$$H_{\text{id}} := \frac{1}{\min\{\frac{\Delta_{i^\star}}{2}, 2\Delta_{i^\star}^{\text{attr}}\}^2} + \sum_{i \in \mathcal{F} \cap \mathcal{S}} \frac{1}{\left(\frac{\Delta_i}{2}\right)^2} + \sum_{i \in \bar{\mathcal{F}} \cap \mathcal{R}} \frac{1}{\left(\frac{\Delta_i^{\text{attr}}}{2}\right)^2} + \sum_{i \in \bar{\mathcal{F}} \cap \mathcal{S}} \frac{1}{\max\{\frac{\Delta_i}{2}, 2\Delta_i^{\text{attr}}\}^2}.$$

We also have Empirically *Sub-optimal Set*, *Risky Set*, and *Neutral Set* depending on the separator value and are defined according to the equations given below:

$$\begin{aligned} \mathcal{S}_t &:= \{i : U_i(t) < \bar{\mu}\} \\ \mathcal{R}_t &:= \{i : L_i(t) > \bar{\mu}\} \\ \mathcal{N}_t &:= [N] \setminus (\mathcal{S}_t \cup \mathcal{R}_t) = \{i : L_i(t) \leq \bar{\mu} \leq U_i(t)\}. \end{aligned}$$

We now state an upper bound on the number of samples required by CSS-LUCB in the following theorem.

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

The proof outline is as follows:

– We first define an event E where the means of all arms and attributes lie within their confidence intervals for all rounds $t \geq 2$. We prove that this event occurs with "high probability" if the algorithm terminates.
– Next, we prove that the algorithm will give the correct answer if event E occurs.
– We then prove that if the algorithm does not terminate, one of the two following conditions is satisfied:
  - $i_t \in (\partial\mathcal{F}_t \backslash \mathcal{S}_t) \cup (\mathcal{F}_t \cap \mathcal{N}_t)$
  - $c_t \in (\partial\mathcal{F}_t \backslash \mathcal{S}_t) \cup (\mathcal{F}_t \cap \mathcal{N}_t)$

  We then show that these conditions occur with "low probability".
– Finally, we prove that for some $t > K \times H_{\text{id}} \ln(H_{\text{id}}/\delta)$, for a constant K, the algorithm terminates with "high probability".

These steps show that the algorithm terminates and identifies the best arm with a probability of at least $1 - \delta$.

V. Numerical Results

In this section, we present our numerical results. We compare the performance of our policy with two suitably adapted variants of the widely studied action elimination algorithm [3]. More specifically, along with the usual elimination of arms whose mean reward is low, we also eliminate arms that have one or more attributes whose UCB is less than the given threshold (μ_{TH}). In addition, we also simulate an algorithm that divides the problem into two sub-tasks. The first task is to eliminate arms that are infeasible, followed by the second task, which is to identify the best arm in the set of feasible arms. We use action elimination for the second task. We refer to this approach as *Feasibility then BAI*.

TABLE IV: Experimental Data

Experiment 1: x varies from 0.31 to 0.35, $\mu_{\text{TH}} = 0.3$

Arm	Attr1	Attr2	Attr3	Attr4	Attr5	Mean
1	**x**	0.6	0.7	0.6	0.7	**0.58 - 0.59**
2	0.4	0.2	0.4	0.4	0.45	**0.37**
3	0.15	0.7	0.8	0.9	0.9	**0.69**
4	**x**	0.35	0.35	0.35	0.4	**0.35 - 0.36**
5	**x**	0.35	0.35	0.35	0.35	**0.34 - 0.35**

Experiment 2: x varies from 0.5 to 0.7, $\mu_{\text{TH}} = 0.3$

Arm	Attr1	Attr2	Attr3	Attr4	Attr5	Mean
1	0.7	0.6	0.8	0.7	0.9	**0.74**
2	0.7	**x**	0.7	0.7	0.8	**0.68 - 0.72**
3	0.15	0.7	0.8	0.9	0.9	**0.69**
4	0.15	0.9	0.9	0.9	0.8	**0.73**
5	0.1	0.9	0.9	0.8	0.8	**0.7**

Experiment 3: x varies from 0.05 to 0.3, $\mu_{\text{TH}} = 0.35$

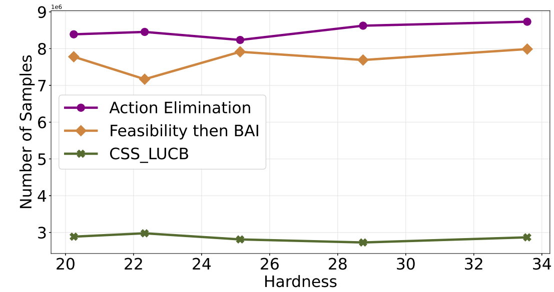
Arm	Attr1	Attr2	Attr3	Attr4	Attr5	Mean
1	0.5	0.6	0.6	0.5	0.8	**0.6**
2	0.7	0.5	0.4	0.4	0.6	**0.52**
3	**x**	0.5	0.9	0.8	0.9	**0.63 - 0.68**
4	0.6	0.2	0.4	0.7	0.6	**0.5**
5	0.3	0.7	0.4	0.9	0.5	**0.56**

For the results presented in this section, the reward of each attribute of each arm is an independent stochastic process with the beta distribution. We perform three different experiments:

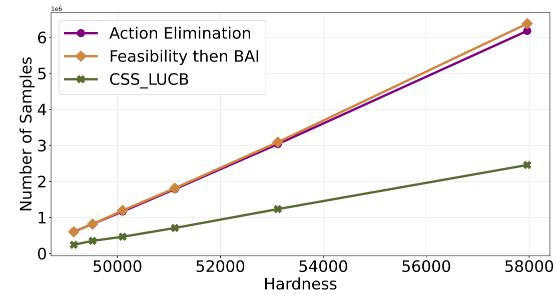
1) We vary the mean reward of an attribute of the best arm, making it more difficult to assess the feasibility of that arm.
2) We vary the mean reward of a suboptimal arm, which increases the hardness of detecting the arm with highest mean.
3) The arm with the highest mean is infeasible (*Risky arm).* We vary the mean reward of one of its attributes, making it difficult to detect its in-feasibility.

We set $\delta = 0.1$, $N = 5$, and $M = 5$. The values of various parameters are given in Table IV. (The shaded rows depict infeasible arms.)

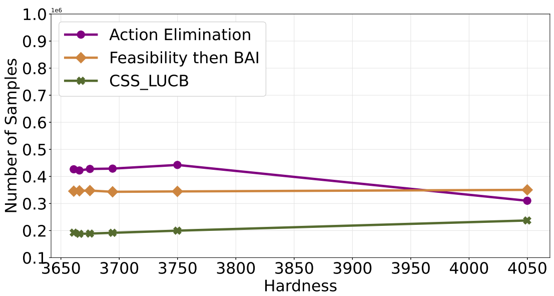
The number of samples required by each algorithm is plotted against the hardness index, H_{id}. The results are shown in Fig.1. Our algorithm outperforms the standard algorithms in all the experiments.



(a) Experiment 1



(b) Experiment 2



(c) Experiment 3

Fig. 1: Sample-complexity as a function of the hardness

Next, we vary the number of arms and attributes, comparing the results of the CSS-LUCB algorithm in the case where the arm with the highest mean was infeasible, and the results are shown in Fig. 2. In Fig 2a, we plot the sample complexity as a function of the number of arms keeping the number of attributes constant. For $N \in \{4, 5, 6\}$, we consider arms $1, 2, \cdots, N$, shown in Table V. In Fig 2b, we vary the number of attributes for a fixed number of arms. For $M \in \{2, 3, 4\}$, we consider the first two, three, and four attributes, respectively, from Table V. As expected, sample complexity increases as the number of arms or attributes increases.

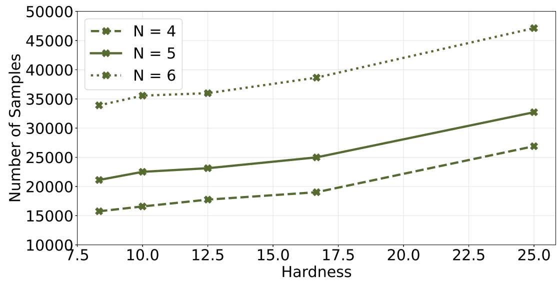
TABLE V: Data for varying instance parameters

Varying N, fixed $M = 2$, $\mu_{\text{TH}} = 0.5$, x varies from 0.38 to 0.46

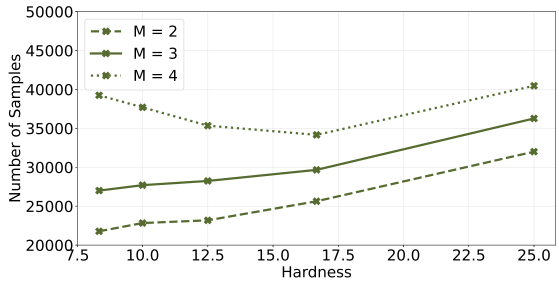
Arm	Attribute 1	Attribute 2	Mean
1	0.6	0.7	**0.65**
2	**x**	0.9	**0.64 - 0.68**
3	0.3	0.55	**0.425**
4	0.55	0.55	**0.55**
5	0.2	0.4	**0.3**
6	0.55	0.6	**0.575**

Varying M, fixed $N = 5$, $\mu_{\text{TH}} = 0.5$, x varies from 0.38 to 0.46

Arm	Attr 1	Attr 2	Attr 3	Attr 4	Mean
1	0.55	0.6	0.65	0.7	**0.625**
2	**x**	0.9	0.7	0.8	**0.695 - 0.715**
3	0.3	0.55	0.4	0.6	**0.463**
4	0.55	0.55	0.55	0.55	**0.55**
5	0.2	0.4	0.3	0.55	**0.363**



(a) Varying number of arms



(b) Varying number of attributes

Fig. 2: Sample complexity of CSS-LUCB on varying the number of arms and attributes

References

[1] Aurélien Garivier and Emilie Kaufmann. Optimal best arm identification with fixed confidence, 2016.

[2] Chao Qin. Open problem: Optimal best arm identification with fixed-budget. In *CoLT*, pages 5650–5654. PMLR, 2022.

[3] Kevin Jamieson and Robert Nowak. Best-arm identification algorithms for multi-armed bandits in the fixed confidence setting. In *2014 48th Annual Conference on Information Sciences and Systems (CISS)*, pages 1–6, 2014.

[4] Zhenlin Wang and Jonathan Scarlett. Max-min grouped bandits. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 8603–8611, 2022.

[5] Victor Gabillon, Mohammad Ghavamzadeh, Alessandro Lazaric, and Sébastien Bubeck. Multi-bandit best arm identification. *Advances in Neural Information Processing Systems*, 24, 2011.

[6] Shivaram Kalyanakrishnan, Ambuj Tewari, Peter Auer, and Peter Stone. Pac subset selection in stochastic multi-armed bandits. In *ICML*, volume 12, pages 655–662, 2012.

[7] Matthieu Jedor, Vianney Perchet, and Jonathan Louedec. Categorized bandits. *Advances in Neural Information Processing Systems*, 32, 2019.

[8] Zhenlin Wang, Andrew Wagenmaker, and Kevin G. Jamieson. Best arm identification with safety constraints. *CoRR*, 2021.

[9] Julian Katz-Samuels and Clayton Scott. Top feasible arm identification. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 1593–1601. PMLR, 2019.

[10] Yuhang Wu, Zeyu Zheng, and Tingyu Zhu. Best arm identification with fairness constraints on subpopulations, 2023.

[11] Yunlong Hou, Vincent YF Tan, and Zixin Zhong. Almost optimal variance-constrained best arm identification. *IEEE Transactions on Information Theory*, 69(4):2603–2634, 2022.