The True Sample Complexity of Identifying Good Arms

Julian Katz-Samuels
University of Washington

Kevin Jamieson
University of Washington

Abstract

We consider two multi-armed bandit problems with n arms: *(i)* given an $\epsilon > 0$, identify an arm with mean that is within ϵ of the largest mean and *(ii)* given a threshold μ_0 and integer k, identify k arms with means larger than μ_0. Existing lower bounds and algorithms for the PAC framework suggest that both of these problems require $\Omega(n)$ samples. However, we argue that the PAC framework not only conflicts with how these algorithms are used in practice, but also that these results disagree with intuition that says *(i)* requires only $\Theta(\frac{n}{m})$ samples where $m = |\{i : \mu_i > \max_{j\in[n]} \mu_j - \epsilon\}|$ and *(ii)* requires $\Theta(\frac{n}{m}k)$ samples where $m = |\{i : \mu_i > \mu_0\}|$. We provide definitions that formalize these intuitions, obtain lower bounds that match the above sample complexities, and develop explicit, practical algorithms that achieve nearly matching upper bounds.

1 Introduction

We consider the multi-armed bandit (MAB) problem of ϵ-GOOD ARM IDENTIFICATION. In this problem there are n distributions $\rho_1, \dots, \rho_n$ (also referred to as arms) with means $\mu_1, \dots, \mu_n$; an agent plays a sequential game where at each round t, she chooses (or "pulls") an arm $I_t \in \{1, \dots, n\}$ and observes an i.i.d. realization from ρ_{I_t}. The goal of the game is to use as few total pulls as possible to identify an *ϵ-good arm*, that is, an arm i that satisfies $\mu_i > \max_j \mu_j - \epsilon$ for a given $\epsilon > 0$. In the well-studied PAC framework, the sample complexity of an agent is measured by the total number of pulls until the agent can terminate the game and return an ϵ-good arm with probability at least $1 - \delta$.

Proceedings of the 23[rd]International Conference on Artificial Intelligence and Statistics (AISTATS) 2020, Palermo, Italy. PMLR: Volume 108.

ϵ-GOOD ARM IDENTIFICATION has received much attention in the MAB literature and has many potential applications ranging from clinical trials to crowdsourcing. The literature has focused on designing algorithms that optimize the PAC notion of sample complexity; in this paper, we argue that PAC sample complexities are impractically large even for a modest number of arms. Consider our experiment on the recently crowdsourced New Yorker Caption Contest with 9061 Bernoulli arms (presented in Section 1.3), where the top arm has a mean of about 0.45 and the bottom arm a mean of about 0.04. On this realistic bandit problem, it takes a state-of-the-art ϵ-GOOD ARM IDENTIFICATION algorithm LUCB over 1 million samples to identify an arm as 0.45-good with probability at least 0.95. But, if one simply chose a random arm without taking any samples, then with probability 1 the returned arm would be 0.45-good! As we discuss in detail below, lower bounds show that these impractical sample complexities are unavoidable, scaling like $\Theta(n)$ because the PAC framework requires that the agent *verify* that the returned arm is ϵ-good. For this reason, we also refer to PAC sample complexity as *verifiable sample complexity*.

In this paper, we propose a novel framework for quantifying the sample complexity of an algorithm for ϵ-GOOD ARM IDENTIFICATION. We suppose that the agent outputs an arm $\widehat{i}_t$ at every round t and, informally, we consider the sample complexity of the agent to be the round at which the agent begins to output an ϵ-good arm with high probability at every subsequent round. We call this *unverifiable sample complexity* because, in contrast to the PAC notion of sample complexity, it does not require that the algorithm verify that an arm is ϵ-good. $\widehat{i}_t$ represents the "best guess" of the algorithm and unverifiable sample complexity is the number of rounds until the agent happens to be right with high probability on all subsequent rounds. Through the development of lower bounds and algorithms with nearly matching upper bounds, we show that unverifiable sample complexity can be arbitrarily smaller than PAC sample complexity, scaling like $\Theta(\frac{n}{m})$ where m is the number of ϵ-good arms.

As a corollary to our study of the unverifiable sample complexity of ϵ-GOOD ARM IDENTIFICATION, we obtain results for the intimately related problem of identifying $k \leq n$ arms that satisfy $\mu_i > \mu_0 \in \mathbb{R}$, where μ_0 *is known*. We call this the k-IDENTIFICATIONS PROBLEM. By contrast to the optimization flavor of ϵ-GOOD ARM IDENTIFICATION, this problem can be thought of as akin to *satisficing*, an approach to decision problems that seeks to find acceptable options (Simon, 1956). This problem is relevant to applications where it suffices to find k arms that meet a known standard. For example, consider the task of hiring crowdsourcing workers where a practitioner often wishes to hire a certain number of workers that meet a certain standard (e.g., answer a question correctly with probability at least 0.9). As another example, consider the biological sciences where a scientist is often interested in determining which of a collection of genes are important for a biological process, and is satisfied if she makes a few discoveries (Hao et al., 2008). Although satisficing problems are ubiquitous in applications, they have received far less attention in the MAB pure exploration literature.

1.1 Multi-armed bandits

Define a *multi-armed bandit instance* ρ as a collection of n distributions over $\mathbb{R}$ where the ith distribution ρ_i has expectation $\mathbb{E}_{X\sim\rho_i}[X] = \mu_i$. We assume without loss of generality that $\mu_1 \geq \mu_2 \geq \ldots \geq \mu_n$. At round $t \in \mathbb{N}$ a player selects an index $I_t \in [n] \coloneqq \{1, \ldots, n\}$, immediately observes an independent realization Z_t of ρ_{I_t}, and then outputs $\widehat{S}_t$, which is either a subset of $[n]$ or an element in $[n]$, depending on the problem. Formally, defining the filtrations $(\mathcal{F}_t)_{t\in\mathbb{N}}$ and $(\mathcal{F}_t^-)_{t\in\mathbb{N}}$ where $\mathcal{F}_t = \{(I_s, Z_s, \widehat{S}_s) : 1 \leq s \leq t\}$ and $\mathcal{F}_t^- = \mathcal{F}_{t-1} \cup \{(I_t, Z_t)\}$, we require that I_t is $\mathcal{F}_{t-1}$ measurable while $\widehat{S}_t$ is $\mathcal{F}_t^-$ measurable, each with possibly additional external sources of randomness.

The player strategically chooses an arm I_t at each time t in order to accomplish a goal for $\widehat{S}_t$ as quickly as possible. We consider the following two objectives.

1. **ϵ-good arm identification**: for a given $\epsilon > 0$, minimize τ such that the index $\widehat{S}_t \in [n]$ satisfies $\mu_{\widehat{S}_t} > \max_{i\in[n]} \mu_i - \epsilon$ for all $t \geq \tau$ with high probability.
2. **k-identifications problem**: for a given threshold $\mu_0 \in \mathbb{R}$ and $k \in [n]$, minimize τ_k such that the set $\widehat{S}_t \subseteq [n]$ satisfies $|\widehat{S}_t \cap \{i : \mu_i > \mu_0\}| \geq \min(k, |\{i : \mu_i > \mu_0\}|)$ for every $t \geq \tau_k$ subject to $\widehat{S}_s \cap \{i : \mu_i \leq \mu_0\} = \emptyset$ for all $s \in \mathbb{N}$ with high probability[1].

[1]The constraint $\widehat{S}_s \cap \{i : \mu_i \leq \mu_0\} = \emptyset$ is known as a family-wise error rate (FWER) condition. We will also consider a more relaxed condition known as false discovery rate (FDR) which controls $\mathbb{E}[|\widehat{S}_s \cap \{i : \mu_i \leq \mu_0\}|/|\widehat{S}_s|]$.

When $\epsilon = 0$ and arm 1 is uniquely optimal, ϵ-GOOD ARM IDENTIFICATION is the well-studied problem of *best arm identification*.

Why study both objectives simultaneously? ϵ-GOOD ARM IDENTIFICATION and the k-IDENTIFICATIONS PROBLEM are closely related. If $k = 1$, then the k-IDENTIFICATIONS PROBLEM is essentially ϵ-GOOD ARM IDENTIFICATION where the threshold $\mu_0 = \mu_1 - \epsilon$ is known, but $\epsilon = \mu_1 - \mu_0$ is unknown. The same algorithmic ideas can be applied to both problems, and, indeed, our proposed algorithms and analyses for both problems are very similar.

Furthermore, the fundamental difficulty of the objectives are closely related: for a fixed set of means $\mu_1 \geq \cdots \geq \mu_n$ and any threshold μ_0, we may consider $\epsilon = \mu_1 - \mu_0$ so that $\{\mu_i : \mu_i > \mu_1 - \epsilon\} = \{\mu_i : \mu_i > \mu_0\}$. Thus, identifying k arms above the threshold μ_0 is equivalent to identifying k ϵ-good means for $\epsilon = \mu_1 - \mu_0$. Consequently, if $m = |\{i \in [n] : \mu_i > \mu_1 - \epsilon\}|$ then we can study *lower bounds* on the sample complexity of both problems simultaneously by considering the necessary number of samples required to identify k of the m largest means (i.e., to have $\widehat{S}_t \subset [m]$ with $|\widehat{S}_t| = k$) for any value of $1 \leq k \leq m$. Henceforth, we use m to denote $|\{i \in [n] : \mu_i > \mu_1 - \epsilon\}|$ or $|\{i \in [n] : \mu_i > \mu_0\}|$; the context will leave no ambiguity.

Intuition for unverifiable sample complexity. Suppose that it is *known* that there are m ϵ-good arms and consider the following algorithm: let A be a set of n/m arms chosen uniformly at random from $[n]$ and apply any nearly optimal best arm identification algorithm to A. Observe that one of the arms in A is ϵ-good with constant probability since

$$\mathbb{P}(A \cap [m] = \emptyset) \leq (1 - m/n)^{n/m} \leq \exp(-1).$$

Thus, this algorithm will return an ϵ-good arm with constant probability in a number of samples that scales like n/m (instead of the typical n). Although this algorithm requires knowledge of m, it suggests that when there are m ϵ-good distributions, the unverifiable sample complexity to identify an ϵ-good distribution scales as n/m, not n. In an extreme case, if half the distributions are ϵ-good, then one should expect the number of samples to identify an ϵ-good distribution to be *constant* with respect to n. A similar argument applies to the k-IDENTIFICATIONS PROBLEM : if there are m means above the threshold μ_0, then one would expect that the number of samples required to identify at least $1 \leq k \leq m$ of them scales like $k\frac{n}{m}$, not n.

While considering m is helpful for analysis, it should be stressed that *the algorithm does not know m and must adapt to it.*

Finally, we stress that although the same algorithmic ideas apply to both ϵ-GOOD ARM IDENTIFICATION and k-IDENTIFICATIONS PROBLEM, our notion of unverifiable sample complexity (made rigorous shortly) does not apply to the k-IDENTIFICATIONS PROBLEM because μ_0 is known and, hence, an agent can verify once k arms above μ_0 have been found.

1.2 Revisiting ϵ-good arm identification: an unverifiable sample complexity perspective

We begin by considering the standard verifiable notion of sample complexity from the well-studied PAC framework.

Definition 1. *Fix a class of bandit instances $\mathcal{P}$. Fix an algorithm $\mathcal{A} \equiv (I_t, \widehat{S}_t, \tau_{V,\epsilon,\delta})$ where $\tau_{V,\epsilon,\delta}$ is a stopping time with respect to the filtration $(\mathcal{F}_t)_{t\in\mathbb{N}}$. Then $\mathcal{A}$ is* ***(ϵ,δ)-PAC (Probably Approximately Correct)*** *wrt $\mathcal{P}$ if $\forall \rho \in \mathcal{P}$ $\mathcal{A}$ terminates at $\tau_{V,\epsilon,\delta}$ and $\mathbb{P}_\rho(\mu_{\widehat{S}_{\tau_{V,\epsilon,\delta}}} > \max_i \mu_i - \epsilon) \geq 1-\delta$. We call $\mathbb{E}_\rho[\tau_{V,\epsilon,\delta}]$ the* ***expected (ϵ,δ)-verifiable sample complexity*** *of $\mathcal{A}$ with respect to ρ.*

In words, $\tau_{V,\epsilon,\delta}$ is the point at which an algorithm $\mathcal{A}$ has collected enough data about ρ to declare confidently that a particular arm is ϵ-good. Setting $\mathcal{P} = \{\mathcal{N}(\mu', I) : \mu' \in \mathbb{R}^n\}$, one can show that for a given ϵ, δ, and instance $\rho \in \mathcal{P}$,

$$\mathbb{E}_\rho[\tau_{V,\epsilon,\delta}] \gtrsim \log(1/\delta) \sum_{i=1}^{n} \max(\mu_1 - \mu_i, \epsilon)^{-2}$$

for any (ϵ,δ)-PAC algorithm over $\mathcal{P}$ (Kaufmann et al., 2016; Mannor et al., 2004) (see Appendix B for a formal statement). That is, *the expected verifiable sample complexity $\mathbb{E}[\tau_{V,\epsilon,\delta}]$ is at least $\Omega(n)$, regardless of m.* Intuitively, this is necessary because if there is some unpulled arm j, then no information is known about j and, thus, the algorithm cannot guarantee that $\mu_j < \mu_i + \epsilon$ for any other arm i. We now propose a definition for unverifiable sample complexity.

Definition 2. *Fix an algorithm $\mathcal{A} \equiv (I_t, \widehat{S}_t)$ and an instance ρ. Let $\tau_{U,\epsilon,\delta}$ be a stopping time with respect to the filtration $(\mathcal{F}_t)_{t\in\mathbb{N}}$ such that*

$$P_\rho(\forall t \geq \tau_{U,\epsilon,\delta} : \mu_{\widehat{S}_t} > \max_i \mu_i - \epsilon) \geq 1-\delta \qquad (1)$$

and for any other stopping time τ' with respect to the filtration $(\mathcal{F}_t)_{t\in\mathbb{N}}$ that satsifies (1) $\tau_{U,\epsilon,\delta} \leq \tau'$. Then, $\mathbb{E}_\rho[\tau_{U,\epsilon,\delta}]$ is the ***expected (ϵ,δ)-unverifiable sample complexity*** *of $\mathcal{A}$ with respect to ρ.*

$\tau_{U,\epsilon,\delta}$ is the number of samples until an algorithm begins to recommend an ϵ-good arm with high probability on instance ρ. We emphasize that $\tau_{U,\epsilon,\delta}$ is for *analysis purposes only* and *is unknown to the algorithm.* Clearly, if an algorithm $\mathcal{A}$ is (ϵ,δ)-PAC, then for an instance ρ, we have that $\tau_{U,\epsilon,\delta} \leq \tau_{V,\epsilon,\delta}$. However, as the above discussion suggests, $\mathbb{E}\tau_{U,\epsilon,\delta}$ may be significantly smaller than $\mathbb{E}\tau_{V,\epsilon,\delta}$, even as small as $\mathbb{E}\tau_{U,\epsilon,\delta} = O(1)$ while $\mathbb{E}\tau_{V,\epsilon,\delta} = \Omega(n)$. Henceforth, when there is no ambiguity, we will write τ_U and τ_V instead of $\tau_{U,\epsilon,\delta}$ and $\tau_{V,\epsilon,\delta}$ respectively.

Two of the main contributions in this work are *(i)* an instance-dependent lower bound on $\mathbb{E}\tau_U$ and *(ii)* an Algorithm *BUCB* (Bracketing UCB, see Algorithm 1) that achieves a nearly matching upper bound on $\mathbb{E}\tau_U$.

Practical Considerations. It may be unclear how a practitioner would decide to stop collecting samples without a guarantee that the currently most promising arm $\widehat{S}_t$ is ϵ-good. We address this concern in several ways. First, at each round, our algorithm BUCB provides a high probability confidence lower bound $L_t \in \mathbb{R}$ on the mean of the recommended arm $\mu_{\widehat{S}_t}$. Therefore, a practitioner can assess the quality of $\mu_{\widehat{S}_t}$ using L_t and use this information to decide whether to stop sampling. Second, it is possible to design an algorithm that has nearly optimal verifiable and unverifiable sample complexity (see the Appendix for details). Third, a practitioner can interpret our algorithm BUCB as finding as good an arm as possible in a time horizon T (for any $T \in \mathbb{N}$), that is, as minimizing the high-probability *simple regret* $\mu_1 - \mu_{\widehat{S}_T}$ (Bubeck et al., 2011). Finally, we note that in some applications, practitioners are more interested in finding a good arm quickly than in certifying that a returned arm is ϵ-good.

1.3 Motivating Experiments

Next, we briefly present some illustrative experiments that motivate our framework.

ϵ-good arm identification. The LUCB algorithm of Kalyanakrishnan et al. (2012) is an (ϵ,δ)-PAC algorithm whose sample complexity is within $\log(n)$ of the lower bound of any (ϵ,δ)-PAC algorithm and is known to have excellent empirical performance (Jamieson and Nowak, 2014). LUCB does not use ϵ as a sampling rule (only a stopping condition), and thus can be evaluated after any number of pulls using its empirical best arm. We compare its performance to our algorithm BUCB in this paper designed to optimize unverifiable sample complexity. We obtain a realistic bandit instance of 9061 Bernoulli arms with parameters defined by the empirical means from a recent crowd-sourced *New Yorker Magazine* Caption Contest, where each caption was shown uniformly at random to a participant, and

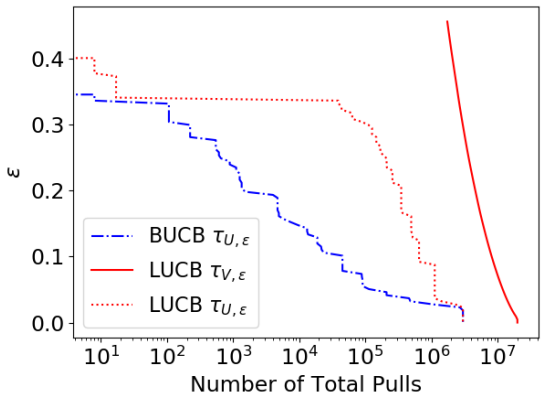

Figure 1: ϵ-GOOD ARM IDENTIFICATION

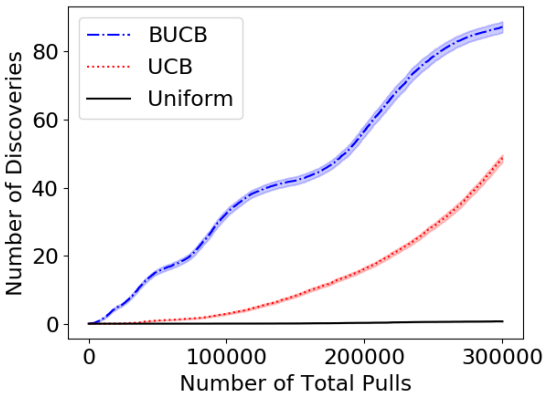

Figure 2: Identifying means above a threshold

received on average 155 votes of funny/unfunny (see Appendix G for details). We run LUCB and BUCB with $\delta = 0.05$ for 100 trials. Figure 1 depicts the results from the experiment. For a given $\epsilon > 0$, $\tau_{U,\epsilon}$ is the first round at which the empirical probability of returning an ϵ-good arm is above $1 - \delta$ at every $t \geq \tau_{U,\epsilon}$. We observe that our proposed algorithm begins to recommend ϵ-good arms with high probability using orders of magnitude fewer samples than LUCB for a large range of values of ϵ. In addition, the verifiable complexity $\tau_{V,\epsilon}$ of LUCB is worse than the unverifiable sample complexity of BUCB by several orders of magnitude.

k-Identifications Problem. The recent work of Jamieson and Jain (2018) proposed an algorithm (UCB) that identifies nearly all m arms above a threshold in a number of samples that is nearly optimal, but has a sample complexity that scales with n. We compare its performance to our algorithm BUCB that optimizes identifying $k < m$ arms. Consider the experimental data of Hao et al. (2008), which aimed to discover genes in Drosophila that inhibit virus replication. Hao et al. (2008) measured 13,071 genes using a total budget of about 38,000 measurements. Figure 2 depicts a simulation of 100 trials based on plug-in estimates of the experimental data of Hao et al. (2008) (described in Appendix G) and shows that our algorithm (BUCB) is able to make discoveries much more quickly than the algorithm from Jamieson and Jain (2018) (UCB). See Appendix G for more details on the experiments.

1.4 Related work

In addition to the lower bounds for the (ϵ, δ)-PAC setting discussed in Section 1.2 (Kaufmann et al., 2016; Mannor et al., 2004), a related line of work has studied the exact PAC sample complexity in the asymptotic regime as $\delta \to 0$ (Degenne and Koolen, 2019; Garivier and Kaufmann, 2019). By contrast, our results concern the moderate confidence regime where δ is treated as a constant (e.g., around 0.05).

Our definition of unverifiable sample complexity may be interpreted as a high probability version of the expected *simple regret* metric (c.f. Bubeck et al. (2011)), however, neither definition subsumes the other. The closest work to our setting is that of Chaudhuri and Kalyanakrishnan (2017, 2019); Aziz et al. (2018) that also aimed to identify multiple arms, but with the critical difference that m is assumed to be *known*. Specifically, given a tolerance $\eta \geq 0$, they say an arm i is (η, m)-optimal if $\mu_i \geq \mu_m - \eta$. The objective, given m and η as inputs to the algorithm, is to identify k (η, m)-optimal arms with probability at least $1 - \delta$. The case when $\eta = 0$ and $m = |\{i : \mu_i > \mu_1 - \epsilon\}|$ coincides with our setting, with the critical difference that in our setting the algorithm never has knowledge of m. With just knowledge of ϵ but not m, as in our setting, there is no guide a priori to how many arms we need to consider in order to get just one ϵ-good arm. However, still relevant from a lower bound perspective, they prove *worst-case* results for $\eta > 0$. In contrast, our work demonstrates instance-specific lower-bounds (i.e., those that depend on the particular means μ) that directly apply to their setting, a contribution of its own.

Algorithms for ϵ-good identification. The last few decades have seen many proposed (ϵ, δ)-PAC algorithms for identifying an ϵ-good arm (Even-Dar et al., 2006; Kalyanakrishnan et al., 2012; Gabillon et al., 2012; Kaufmann and Kalyanakrishnan, 2013; Karnin et al., 2013; Simchowitz et al., 2017; Garivier and Kaufmann, 2019). A closely related problem is known as the *infinite armed-bandit problem* where the player has access to an infinite pool of arms such that when

a new arm is requested, its mean is drawn iid from a distribution ν. In principle, an infinite armed bandit algorithm could solve the problem of interest of this paper by taking $\nu(x) = \frac{1}{n}\sum_{i=1}^n \mathbf{1}\{\mu_i \le x\}$. With the exception of Li et al. (2017), nearly all of the existing work makes parametric[2] assumptions about ν in some way (Berry et al., 1997; Wang et al., 2009; Carpentier and Valko, 2015; Chandrasekaran and Karp, 2014; Jamieson et al., 2016). However, the algorithm of Li et al. (2017) was designed for a much more general setting and therefore sacrifices both theoretical and practical performance, and was not designed to take a fixed confidence δ as input.

Algorithms for identifying means above μ_0. In the *thresholding bandit problem*, the agent is given a budget of T pulls, and the goal is to maximize the probability of identifying *every* arm as either above or below a threshold μ_0 (Locatelli et al., 2016; Mukherjee et al., 2017). These works explicitly assume no arms are *equal* to μ_0 and penalize incorrectly predicting a mean above or below the threshold equally. For our problem setting, the most related work is Jamieson and Jain (2018) which proposes an algorithm that takes a confidence δ and threshold μ_0 as input. The authors characterize the total number of samples the algorithm takes before all $k = m$ arms with means above the threshold are output with probability at least $1-\delta$ for all future times, that is, the k-IDENTIFICATIONS PROBLEM where $k = m$. While this sample complexity is nearly optimal for the $k = m$ case (see the lower bounds of Simchowitz et al. (2017); Chen et al. (2014)) this work is silent on the issue of identifying just a subset of size $k \le m$ means above the threshold (and the algorithm does not generalize to this setting).

2 Lower bounds

For the rest of the paper, we focus on developing lower bounds and algorithms with upper bounds for unverifiable sample complexity, as well as analogous results for the k-IDENTIFICATIONS PROBLEM. We begin by presenting a lower bound. To avoid trivial algorithms that deterministically output an index that happens to be the best arm, we adopt the random permutation model of Simchowitz et al. (2017) and Chen et al. (2017). We say $\pi \sim \mathbb{S}^n$ if π is drawn uniformly at random from the set of permutations over $[n]$, denoted $\mathbb{S}^n$. For any $\pi \in \mathbb{S}^n$, $\pi(i)$ denotes the index that i is mapped to under π. Also, let $T_i(t)$ denote the number of pulls of arm i up to time t. For a bandit instance $\rho = (\rho_1, \dots, \rho_n)$ let $\pi(\rho) = (\rho_{\pi(1)}, \rho_{\pi(2)}, \dots, \rho_{\pi(n)})$ so that $\mathbb{E}_{\pi(\rho)}[T_{\pi(i)}(t)]$ denotes the expected number of samples taken by the algorithm up to time t from the arm with mean $\mu_{\pi(i)}$ when run on instance $\pi(\rho)$. The sample complexity of interest is the expected number of samples taken by the algorithm under $\pi(\rho)$ averaged over all possible $\pi \in \mathbb{S}^n$.

As pointed out in the introduction, there is a one-to-one correspondance between a problem instance for identifying k arms above a threshold μ_0 and a problem instance for identifying k ϵ-good arms, where $\epsilon = \mu_1 - \mu_0$. Thus, if $m = |\{i : \mu_i > \mu_1 - \epsilon\}|$ then a lower bound for identifying k ϵ-good arms or k arms above a threshold μ_0 is implied by a lower bound for identifying k arms among the m largest means for any $1 \le k \le m$. The next theorem handles all $1 \le k \le m$ cases simultaneously for a specific instance (i.e., not worst-case as in (Chaudhuri and Kalyanakrishnan, 2019)).

Theorem 1. *Fix $\epsilon > 0$, $\delta \in (0, 1/16)$, and a vector $\mu \in \mathbb{R}^n$. Consider n arms where rewards from the ith arm are distributed according to $\mathcal{N}(\mu_i, 1)$. Assume without loss of generality that $\mu_1 \ge \mu_2 \ge \cdots \ge \mu_n$ and let $m = |\{i \in [n] : \mu_i > \mu_1 - \epsilon\}|$. For every permutation $\pi \in \mathbb{S}^n$ let $(\mathcal{F}_t^\pi)_{t\in\mathbb{N}}$ be the filtration generated by the algorithm playing on instance $\pi(\rho)$, and let τ_π be a stopping time with respect to $(\mathcal{F}_t^\pi)_{t\in\mathbb{N}}$ at which time the algorithm outputs a set $\widehat{S}_{\tau_\pi} \subseteq [n]$ with $|\widehat{S}_{\tau_\pi}| = k$. If $\mathbb{P}_{\pi(\rho)}(\widehat{S}_{\tau_\pi} \subset \pi([m])) \ge 1 - \delta$, then*

$$\mathbb{E}_{\pi\sim\mathbb{S}^n}\mathbb{E}_{\pi(\rho)}\Big[\tau_\pi\Big] \ge \mathcal{H}_{\mathrm{low},k}(\epsilon)$$
$$:= \frac{1}{64}\Big(-(\mu_1 - \mu_{m+1})^{-2} + \frac{k}{m}\sum_{i=m+1}^n (\mu_1 - \mu_i)^{-2}\Big).$$

Since the theorem applies to any stopping time τ_π that satisfies $\mathbb{P}_{\pi(\rho)}(\widehat{S}_{\tau_\pi} \subset \pi([m])) \ge 1 - \delta$, in particular it yields a lower bound for expected unverifiable sample complexity. Furthermore, by definition, $(\mu_1 - \mu_{m+1})^{-2} \le \epsilon^{-2}$ so aside from pathological cases such as $\mu_1 - \mu_i \gg \epsilon$ for all $i > m+1$ the lower bound will be positive and non-trivial. Consider the following examples.

Example 1. *If $(\mu_1 - \mu_{m+1})^{-2} \le \frac{k}{2m}\sum_{i=m+1}^n (\mu_1 - \mu_i)^{-2}$, then $\mathcal{H}_{\mathrm{low},k}(\epsilon) \ge \frac{1}{128}\epsilon^{-2} + \frac{1}{256}\frac{k}{m}\sum_{i=m+1}^n (\mu_1 - \mu_i)^{-2}$.*

Example 2. *If $\mu_1 = \ldots = \mu_m = \mu_0 + \epsilon$ and $\mu_{m+1} = \ldots \mu_n = \mu_0$, then $\mathcal{H}_{\mathrm{low},k}(\epsilon) \ge \frac{1}{64}\frac{k(n-m)}{m}\epsilon^{-2}$. If in addition $n \ge 2m$, then $\mathcal{H}_{\mathrm{low},k}(\epsilon) \ge \frac{1}{128}\frac{kn}{m}\epsilon^{-2}$.*

Example 2 shows that Theorem 1 yields a lower bound matching our intuition for the n/m scaling of (i) unverifiable sample complexity of ϵ-GOOD ARM IDENTIFICATION, and (ii) the sample complexity of the k-IDENTIFICATIONS PROBLEM.

[2] For example, for a drawn arm with random mean μ it is assumed $\mathbb{P}(\mu \le x) \ge c(x - \mu_*)^\beta$ for some fixed parameters c, μ_*, β that are known (or not).

Algorithm 1 Bracketing UCB: ϵ-GOOD ARM IDENTIFICATION and k-IDENTIFICATIONS PROBLEM

1: $\delta_r = \frac{\delta}{r^2}$, $\delta'_r = \frac{\delta_r}{6.4\log(36/\delta_r)}$, $\ell = 0$, $R_0 = 0$, $\mathcal{S}_0 = \emptyset$
2: **for** $t = 1, 2, \ldots$ **do**
3: **if** $t \geq 2^\ell \ell$ **then**
4: $A_{\ell+1} \sim \text{Uniform}(\binom{[n]}{M_{\ell+1}})$, where $M_\ell := n \wedge 2^\ell$
5: $\ell = \ell + 1$
6: $R_t = 1 + R_{t-1} \cdot \mathbf{1}\{R_{t-1} < \ell\}$
7: **if** $\exists i \in A_{R_t} \setminus \mathcal{S}_t$ such that $T_{i,R_t}(t) = 0$ **then**
8: Pull $I_t \in \{i \in A_{R_t} \setminus \mathcal{S}_t : T_{i,R_t}(t) = 0\}$
9: **else**
10: Pull $I_t = \underset{i \in A_{R_t} \setminus \mathcal{S}_t}{\text{argmax}} \widehat{\mu}_{i,R_t,T_{i,R_t}(t)} + U(T_{i,R_t}(t), \delta)$
11: **if** ϵ-GOOD ARM IDENTIFICATION **then**
12: $O_t = \underset{i \in A_r \text{ for some } r \leq \ell}{\text{argmax}} \widehat{\mu}_{i,r,T_{i,r}(t)} - U(T_{i,r}(t), \frac{\delta}{|A_r| r^2})$
13: **else if** k-IDENTIFICATIONS PROBLEM **then**
14: $s(p) = \{i : \widehat{\mu}_{i,R_t,T_{i,R_t}(t)} - U(T_{i,R_t}(t), \frac{p\delta'_{R_t}}{|A_{R_t}|}) \geq \mu_0\}$ for all $p \in [|A_{R_t}|]$
15: $\mathcal{S}_{t+1} = \mathcal{S}_t \cup s(\widehat{p})$ where $\widehat{p} = \max\{p \in [|A_{R_t}|] : |s(p)| \geq p\}$

The proof of Theorem 1 employs an extension of the *Simulator* argument (Simchowitz et al., 2017). While the $k = 1$ case can be proven using an argument similar to Chen et al. (2017), we needed the Simulator strategy for the $k > 1$ case. The technique may be useful for proving lower bounds for other combinatorial settings where many outcomes are potentially correct (e.g., choose any k of m) (Chen et al., 2014, 2017).

Finally, we close this section by noting that the unverifiable sample complexity of popular algorithms like LUCB or Median Elimination can be greater than $\mathcal{H}_{\text{low},k}(\epsilon)$ by a factor of n (see Appendix C.1). This motivates the development of new algorithms.

3 Algorithm

Algorithm 1 simultaneously handles both ϵ-GOOD ARM IDENTIFICATION (Line 12) and the k-IDENTIFICATIONS PROBLEM (Line 15). To motivate the intuition behind the algorithm, we consider ϵ-GOOD ARM IDENTIFICATION. Suppose the number of ϵ-good arms m were known. Because a random subset A of size $\frac{n}{m}$ contains an ϵ-good arm with constant probability, applying any reasonable best arm identification algorithm to A would achieve our goal of a sample complexity that scales like $\frac{n}{m}$. However, m is not known, so the algorithm applies the doubling trick on the number of ϵ-good arms, subsampling progressively larger random subsets of the arms over time.

We call the random subset $A_\ell \subset [n]$ the ℓth *bracket*. After $(\ell - 1)2^{\ell-1}$ rounds, the bracket A_ℓ is drawn uniformly at random from $\binom{[n]}{M_\ell}$, where $\binom{[n]}{M_\ell}$ denotes all subsets of $[n]$ of size $M_\ell := n \wedge 2^\ell$, at which point we say that ℓth bracket is *open* (Line 4). At each round t, Algorithm 1 chooses one of the open brackets R_t (Line 6) and pulls an arm $I_t \in R_t$ that maximizes an upper confidence bound $\widehat{\mu}_{i,R_t,T_{i,R_t}(t)} + U(T_{i,R_t}(t), \delta)$ on its mean (Line 10). Here, $\widehat{\mu}_{i,r,t}$ denotes the empirical mean of arm i in bracket r after t pulls, $T_{i,r}(t)$ denotes the number of times arm i has been pulled in bracket r up to time t, and finally $U(t, \delta) = c\sqrt{\frac{1}{t}\log(\log(t)/\delta)}$ denotes an anytime confidence bound (thus, satisfying for any $r \in \mathbb{N}$ and $i \in [n]$ $\mathbb{P}(\cap_{t=1}^\infty |\widehat{\mu}_{i,r,t} - \mu_i| \leq U(t, \delta)) \geq 1 - \delta$) based on the law of the iterated logarithm (LIL) (Jamieson et al., 2014; Kaufmann et al., 2016). We note that this sampling rule is similar to the sampling rule of lil'UCB (Jamieson et al., 2014), a nearly optimal algorithm for best arm identification with good empirical performance.

In addition to a sampling rule, we need a recommendation rule. For ϵ-GOOD ARM IDENTIFICATION, the algorithm outputs a maximizer O_t of its lower confidence bound (Line 12). The reason for this is that once an ϵ-good arm i has been pulled roughly $(\mu_i - \mu_{m+1})^{-2}$ times, then with high probability for all subsequent rounds, its confidence lower bound will exceed $\mu_1 - \epsilon$ and the algorithm will only output ϵ-good arms.

For the problem of multiple identifications above a threshold, various suggested sets are possible depending on the desired guarantees. In the main body of the paper, we focus on building a set $\mathcal{S}_t$ that satisfies the following property (Jamieson and Jain, 2018).

Definition 3 (False Discovery Rate, FDR)**.** *Fix some $\delta \in (0, 1)$. We say an algorithm is FDR-δ if for all possible instances (ρ, μ_0), it satisfies $\mathbb{E}[\frac{|\mathcal{S}_t \cap \mathcal{H}_0|}{|\mathcal{S}_t| \wedge 1}] \leq \delta$ for all $t \in \mathbb{N}$, where $\mathcal{H}_0 = \{i \in [n] : \mu_i \leq \mu_0\}$.*

For this goal, the algorithm builds a set $\mathcal{S}_t$ (Line 15) based on the Benjamini-Hochberg procedure developed for multi-armed bandits in Jamieson and Jain (2018). In the Appendix, we present algorithms that satisfy stronger guarantees, but are also less practical.

We note that the above algorithms do not require ϵ or k as an input, and a practitioner can choose to terminate at any point.

4 Upper Bounds

Our upper bounds all have a similar form. They are characterized in terms of $\Delta_{i,j} = \mu_i - \mu_j$, the *gap* between the ith arm and the jth arm. In Appendix E we state our theorems including all factors, but for the purposes of exposition, here we use "$\lesssim$" to hide constants and doubly logarithmic factors. For simplicity, we assume that the distributions are 1-sub-Gaussian and that $\mu_0, \mu_1, \ldots, \mu_n \in [0, 1]$.

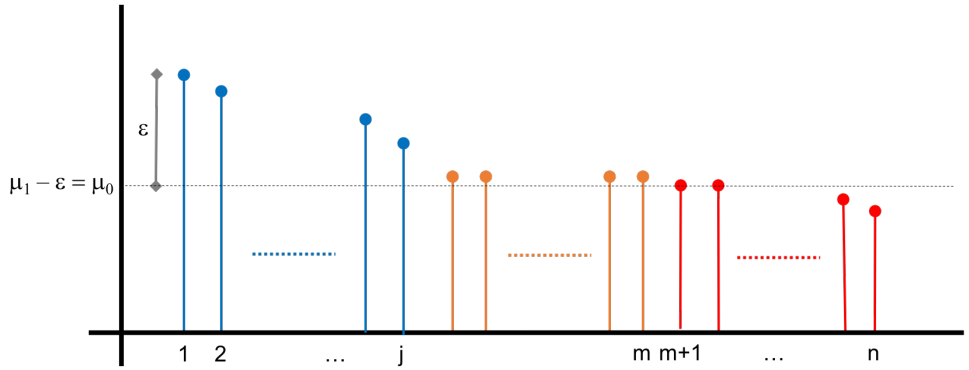

Figure 3: Our sample complexity results rely on picking a bracket of an appropriate size: $\frac{n}{m}$ is too small, n is too large, and $\frac{n}{j}$ appears to be about a good size.

4.1 ϵ-Good Arm Identification

To begin, we state our theorem for the unverifiable sample complexity of ϵ-GOOD ARM IDENTIFICATION in full generality. Next, we state several more accessible corollaries that demonstrate the power of the result.

Theorem 2 (ϵ-good identification)**.** *Let $\delta \le 0.025$ and $\epsilon > 0$. Let $(\mathcal{F}_t)_{t\in\mathbb{N}}$ be the filtration generated by playing Algorithm 1 on problem ρ. Then, there exists a stopping time $\tau_{U,\epsilon}$ wrt $(\mathcal{F}_t)_{t\in\mathbb{N}}$ such that $\mathbb{P}(\exists s \ge \tau_{U,\epsilon} : \mu_{O_s} \le \mu_1 - \epsilon) \le 2\delta$ and*

$$\mathbb{E}[\tau_{U,\epsilon}] \lesssim \min_{j\in[m]} \mathcal{H}_{\mathrm{g}}(\epsilon; j)\ln(\mathcal{H}_{\mathrm{g}}(\epsilon; j) + \Delta_{j,m+1}^{-2}) \quad (2)$$

where $\mathcal{H}_{\mathrm{g}}(\epsilon; j) :=$

$$\frac{1}{j}\left(\sum_{i=1}^{m}(\Delta_{j,i} \vee \Delta_{i,m+1})^{-2}\ln(\frac{n}{j\delta}) + \sum_{i=m+1}^{n}\Delta_{j,i}^{-2}\ln(\frac{1}{\delta})\right).$$

Define $\bar{\mathcal{H}}_\epsilon = \sum_{i=1}^{n}\max(\epsilon, (\mu_1 - \mu_i))^{-2}\ln(\frac{n}{m\delta})$.

Corollary 1. *Let $\mathcal{P} = \{\mathcal{N}(\mu', I) : \mu' \in \mathbb{R}^n\}$ and $\rho \in \mathcal{P}$. Define $m = |\{i : \mu_i > \mu_1 - \epsilon\}|$. Let $\mathcal{A}$ be any $(2\epsilon, \delta)$-PAC algorithm wrt $\mathcal{P}$ and let $\tau_{V,2\epsilon}$ be its associated stopping rule. Then, the $\tau_{U,2\epsilon}$ associated with Algorithm 1 defined in Theorem 2 satisfies*

$$\begin{aligned}\mathbb{E}[\tau_{U,2\epsilon}] &\lesssim \frac{1}{m}\bar{\mathcal{H}}_\epsilon \ln(\frac{1}{m}\bar{\mathcal{H}}_\epsilon) \\ &\lesssim \ln(\frac{1}{m}\mathbb{E}[\tau_{V,2\epsilon}])\ln(n/m)\frac{\mathbb{E}[\tau_{V,2\epsilon}]}{m}.\end{aligned}$$

Corollary 2. *Let $\tau_{U,\epsilon}$ be the stopping time associated with Algorithm 1 defined in Theorem 2. Consider the following inequalities:*

$$\mathbb{E}[\tau_{U,\epsilon}] \lesssim \frac{1}{m}\bar{\mathcal{H}}_\epsilon \ln(\frac{1}{m}\bar{\mathcal{H}}_\epsilon) \quad (3)$$

$$\lesssim \mathcal{H}_{\mathrm{low},1}(\epsilon)\ln(\frac{n}{m\delta})\ln(\mathcal{H}_{\mathrm{low},1}(\epsilon)). \quad (4)$$

(3) holds if $|\{i \in [n] : \mu_i \ge \mu_1 - \epsilon/2\}| \ge \frac{m}{2}$, and (4) holds if $(\mu_1 - \mu_{m+1})^{-2} \le \frac{1}{2m}\sum_{i=m+1}^{n}(\mu_1 - \mu_i)^{-2}$.

Corollary 3. *Suppose $\mu_1 = \ldots = \mu_m = \mu_0 + \epsilon$, $\mu_{m+1} = \ldots = \mu_n = \mu_0$, and $n \ge 2m$. Then, the stopping time $\tau_{U,\epsilon}$ associated with Algorithm 1 defined in Theorem 2 satisfies*

$$\begin{aligned}\mathbb{E}[\tau_{U,\epsilon}] &\lesssim \epsilon^{-2}\frac{n}{m}\ln(\frac{n}{m\delta})\ln(\epsilon^{-2}\frac{n}{m}) \\ &= \mathcal{H}_{\mathrm{low},1}(\epsilon)\ln(\frac{n}{m\delta})\ln(\mathcal{H}_{\mathrm{low},1}(\epsilon)).\end{aligned}$$

Corollary 1 says that Algorithm 1 has an unverifiable sample complexity for identifying a 2ϵ-good arm that is *better than the verifiable sample complexity of any $(2\epsilon, \delta)$-PAC algorithm over $\mathcal{P}$ by a factor of the number of ϵ-good arms* (ignoring logarithmic factors). Corollary 2 gives two general conditions under which the unverifiable sample complexity of Algorithm 1 matches the lower bound from Theorem 1 up to logarithmic factors. In words, these conditions are *(i)* a constant proportion of the ϵ-good arms are $\frac{\epsilon}{2}$-good and *(ii)* the cost of determining that a random set of n/m arms of the bottom $n - m$ arms are not ϵ-good dominates the cost of determining that $\mu_1 > \mu_{m+1}$. Finally, Corollary 3 shows that the unverifiable sample complexity of Algorithm 1 attains the desired n/m scaling on the basic problem where m arms have mean $\mu_0 + \epsilon$ and $n - m$ have mean μ_0.

Theorem 2 Discussion. For $j \in [m]$, $\mathcal{H}_{\mathrm{g}}(\epsilon; j)$ bounds the expected unverifiable sample complexity of a random set of size n/j (call it B_j) identifying an ϵ-good arm conditional on *(i)* an arm in $[j]$ belonging to B_j and *(ii)* the empirical means of the arms in B_j concentrating well. $\ln(\mathcal{H}_{\mathrm{g}}(\epsilon; j) + \Delta_{j,m+1}^{-2})$ is the number of brackets that Algorithm 1 opens by the time B_j unverifiably identifies an ϵ-good arm. The minimization problem in (2) says that Algorithm 1 uses the bracket of size about n/j that minimizes the overall unverifiable sample complexity.

It is worthwhile to consider the tradeoff in the bracket

size at some length. Although a bracket of size $\Theta(\frac{n}{m})$ is sufficiently large to contain an ϵ-good arm with constant probability, it may be advantageous to use a much larger bracket in hopes of getting an ϵ-good arm that is much easier to identify as ϵ-good unverifiably. Informally, if one randomly chooses $\frac{n}{j}$ arms then one expects the highest mean amongst these to have an index J uniformly distributed in $[j]$. Thus, a bracket of size about $\frac{n}{m}$ would require distinguishing $J \sim \text{Uniform}([m])$ from the bottom $n-m$ arms, which could require an enormous number of samples on average if many of the arms in $[m]$ are very close to the means of the bottom $n-m$ arms. Thus, for some problems, it is advantageous to use a bracket of size $\frac{n}{j}$ if μ_j is much easier to distinguish from the bottom $n-m$ arms (see Figure 3 for an illustration of this phenomenon).

Proof Discussion. Algorithm 1 essentially applies lil'UCB to random sets separately, so the analysis may focus on lil'UCB applied to a random set B_j of size n/j. A key observation in our proof is that we can analyze lil'UCB on a *fixed* set B_j such that an ϵ-good arm belongs to B_j and the empirical means of the arms in B_j concentrate well. Then, we can take the expectation with respect to the randomness in B_j, which results in a scaling of n/j because each arm belongs to B_j with probability $1/j$.

4.2 k-identifications problem

$\mathcal{H}_1 := \{i \in [n] : \mu_i > \mu_0\}$ consists of the arms that we wish to identify and $\mathcal{H}_0 := \{i \in [n] : \mu_i \le \mu_0\}$ all the other arms. Let $m = |\mathcal{H}_1|$ and recall $\Delta_{j,0} \coloneqq \mu_j - \mu_0$. We measure the sample complexity of the algorithm in the following way (Jamieson and Jain, 2018).

Definition 4 (True Positive Rate, TPR). *Fix some $\delta \in (0,1)$ and $k \le |\mathcal{H}_1|$. We say an algorithm is* TPR-(k,δ,τ) *on an instance (ρ,μ_0) if $\mathbb{E}[|\mathcal{S}_t \cap \mathcal{H}_1|] \ge (1-\delta)k$ for all $t \ge \tau$.*

In the Appendix, we present algorithms that have stronger guarantees, but are also less practical. Theorem 3 bounds the sample complexity in the above sense while showing the FDR of $\mathcal{S}_t$ in Algorithm 1 is controlled. The subsequent corollaries give more accessible consequences of this result.

Theorem 3 (FDR-TPR). *Let $\delta \in (0, .025)$. Let $k \le |\mathcal{H}_1|$. Let $(\mathcal{F}_t)_{t\in\mathbb{N}}$ be the filtration generated by playing Algorithm 1 on problem ρ. Then, for all $t \in \mathbb{N}$, $\mathbb{E}[\frac{|\mathcal{S}_t \cap \mathcal{H}_0|}{|\mathcal{S}_t| \wedge 1}] \le 2\delta$ and there exists a stopping time τ_k wrt $(\mathcal{F}_t)_{t\in\mathbb{N}}$ such that for all $t \ge \tau_k$, $\mathbb{E}[|\mathcal{S}_t \cap \mathcal{H}_1|] \ge (1-\delta)k$ and*

$$\mathbb{E}[\tau_k] \lesssim \min_{k \le j \le m} \mathcal{H}_{\text{id}}(\mu_0; j) \ln(\mathcal{H}_{\text{id}}(\mu_0; j) + \Delta_{j,0}^{-2}), \quad (5)$$

$$\mathbb{E}[\tau_k] \lesssim \min_{k \le j \le m} \tilde{\mathcal{H}}_{\text{id}}(\mu_0; j) \ln(\tilde{\mathcal{H}}_{\text{id}}(\mu_0; j)) \quad (6)$$

where

$$\mathcal{H}_{\text{id}}(\mu_0; j) \coloneqq \frac{k}{j}\Big(\sum_{i=1}^{m} \Delta_{i \vee j, 0}^{-2} \ln(\frac{nk}{j\delta}) + \sum_{i=m+1}^{n} \Delta_{j,i}^{-2} \ln(\frac{1}{\delta})\Big)$$

$$\tilde{\mathcal{H}}_{\text{id}}(\mu_0; j) \coloneqq \frac{n}{j} k \Delta_{j,0}^{-2} \ln(1/\delta)\,.$$

Corollary 4. *Let τ_k be the stopping time associated with Algorithm 1 defined in Theorem 3. Consider the following inequalities.*

$$\mathbb{E}[\tau_k] \lesssim \frac{k}{m} \bar{\mathcal{H}} \ln(\frac{nk}{m\delta}) \ln(\frac{k}{m} \bar{\mathcal{H}}) \quad (7)$$

$$\lesssim \mathcal{H}_{\text{low},k}(\mu_1 - \mu_0) \ln(\frac{nk}{m\delta}) \ln(\mathcal{H}_{\text{low},k}(\mu_1 - \mu_0)) \quad (8)$$

where $\bar{\mathcal{H}} = m\Delta_{1,0}^{-2} \ln(\frac{nk}{m\delta}) + \sum_{i=m+1}^{n} \Delta_{j,i}^{-2} \ln(\frac{1}{\delta})$. (7) holds if $|\{i \in [m] : \Delta_{i,0} \ge \frac{1}{2}\Delta_{1,0}\}| \ge \frac{m}{2}$, and (8) holds if $(\mu_1 - \mu_{m+1})^{-2} \le \frac{1}{2m} \sum_{i=m+1}^{n} (\mu_1 - \mu_i)^{-2}$.

Corollary 5. *Suppose $\mu_1 = \ldots = \mu_m = \mu_0 + \epsilon$, $\mu_{m+1} = \ldots = \mu_n = \mu_0$, and $n \ge 2m$. Then, the stopping time τ_k defined in Theorem 3 satisfies*

$$\mathbb{E}[\tau_k] \lesssim \mathcal{H}_{\text{low},k}(\mu_1 - \mu_0) \ln(\frac{1}{\delta}) \ln(\mathcal{H}_{\text{low},k}(\mu_1 - \mu_0)).$$

Corollary 4 gives conditions under which our algorithm for identifying k arms above a threshold improves by a factor of $\frac{k}{m}$ on the result of Jamieson and Jain (2018) for identifying *all of the arms* above a threshold. Corollary 5 shows that we improve on the gap-independent version of the bound in Jamieson and Jain (2018) by a factor of $\frac{k}{m}$. In addition, these corollaries give conditions under which the sample complexity of Algorithm 1 is within a logarithmic factor of our lower bound.

Theorem 3 Discussion. (5) gives a gap-dependent bound, while (6) sacrifices the dependence on the individual gaps to remove an additional logarithmic factor on the arms in $\mathcal{H}_1$. $\mathcal{H}_{\text{id}}(\mu_0; j)$ bounds the expected number of samples required by a bracket of size $\Theta(\frac{nk}{j})$ to identify k arms satisfying $\mu_i > \mu_0$ when *(i)* at least k of its arms have means greater than $\mu_j > \mu_0$ and *(ii)* the empirical means of the arms in the bracket concentrate well. $\tilde{\mathcal{H}}_{\text{id}}(\mu_0; j)$ plays a similar role but removes a logarithmic factor on the arms in $\mathcal{H}_1$ at the cost of losing the dependence on the individual gaps. Similarly to ϵ-GOOD ARM IDENTIFICATION, there is a tradeoff in the size of the bracket, and the minimization problem in (5) and (6) shows that the algorithm picks an optimal bracket for the overall sample complexity. The proof is quite similar to the proof of Theorem 2.

Acknowledgements

The authors would like to thank Max Simchowitz for very helpful feedback that substantially improved the clarity of the paper. The authors would also like to thank Clay Scott, Jennifer Rogers, and Andrew Wagenmaker for their very useful comments. We also thank Horia Mania for inspiring the proof of Lemma 1. Julian Katz-Samuels is grateful to Clay Scott for his very generous support, which relied on NSF Grants No. 1422157 and 1838179 and funding from the Michigan Institute for Data Science.

References

Maryam Aziz, Jesse Anderton, Emilie Kaufmann, and Javed Aslam. Pure exploration in infinitely-armed bandit models with fixed-confidence. In *ALT 2018-Algorithmic Learning Theory*, 2018.

Donald A. Berry, Robert W. Chen, Alan Zame, David C. Heath, and Larry A. Shepp. Bandit problems with infinitely many arms. *Ann. Statist.*, 25(5): 2103–2116, 10 1997. doi: 10.1214/aos/1069362389.

S. Bubeck, R. Munos, and G. Stoltz. Pure Exploration in Finitely Armed and Continuous Armed Bandits. *Theoretical Computer Science 412, 1832-1852*, 412: 1832–1852, 2011.

Alexandra Carpentier and Michal Valko. Simple regret for infinitely many armed bandits. *CoRR*, abs/1505.04627, 2015.

Karthekeyan Chandrasekaran and Richard Karp. Finding a most biased coin with fewest flips. In *Conference on Learning Theory*, pages 394–407, 2014.

Arghya Roy Chaudhuri and Shivaram Kalyanakrishnan. Pac identification of a bandit arm relative to a reward quantile. In *AAAI*, pages 1777–1783, 2017.

Arghya Roy Chaudhuri and Shivaram Kalyanakrishnan. Pac identification of many good arms in stochastic multi-armed bandits. In *International Conference on Machine Learning*, pages 991–1000, 2019.

Lijie Chen, Jian Li, and Mingda Qiao. Nearly instance optimal sample complexity bounds for top-k arm selection. In *Artificial Intelligence and Statistics*, pages 101–110, 2017.

Shouyuan Chen, Tian Lin, Irwin King, Michael R. Lyu, and Wei Chen. Combinatorial pure exploration of multi-armed bandits. In *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada*, pages 379–387, 2014.

Rémy Degenne and Wouter M Koolen. Pure exploration with multiple correct answers. In *Advances in Neural Information Processing Systems*, pages 14564–14573, 2019.

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*, 7: 1079–1105, 2006.

Victor Gabillon, Mohammad Ghavamzadeh, and Alessandro Lazaric. Best arm identification: A unified approach to fixed budget and fixed confidence. In F. Pereira, C.J.C. Burges, L. Bottou, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 3212–3220. Curran Associates, Inc., 2012.

Aurélien Garivier and Emilie Kaufmann. Non-asymptotic sequential tests for overlapping hypotheses and application to near optimal arm identification in bandit models. *arXiv preprint arXiv:1905.03495*, 2019.

Linhui Hao, Akira Sakurai, Tokiko Watanabe, Ericka Sorensen, Chairul A Nidom, Michael A Newton, Paul Ahlquist, and Yoshihiro Kawaoka. Drosophila rnai screen identifies host genes important for influenza virus replication. *Nature*, 454(7206):890, 2008.

K. Jamieson and R Nowak. Best-arm identification algorithms for multi-armed bandits in the fixed confidence setting. *Information Sciences and Systems (CISS)*, pages 1–6, 2014.

Kevin Jamieson and Lalit Jain. A bandit approach to multiple testing with false discovery control. In *Advances in Neural Information Processing Systems*, 2018.

Kevin Jamieson, Matthew Malloy, Robert Nowak, and Sébastien Bubeck. lilucb: An optimal exploration algorithm for multi-armed bandits. In *Conference on Learning Theory*, pages 423–439, 2014.

Kevin G Jamieson, Daniel Haas, and Benjamin Recht. The power of adaptivity in identifying statistical alternatives. In *Advances in Neural Information Processing Systems*, pages 775–783, 2016.

Shivaram Kalyanakrishnan, Ambuj Tewari, Peter Auer, and Peter Stone. PAC subset selection in stochastic multi-armed bandits. In *Proceedings of the 29th International Conference on Machine Learning, ICML 2012, Edinburgh, Scotland, UK, June 26 - July 1, 2012*, 2012.

Zohar Karnin, Tomer Koren, and Oren Somekh. Almost optimal exploration in multi-armed bandits. In Sanjoy Dasgupta and David Mcallester, editors, *Proceedings of the 30th International Conference on Machine Learning (ICML-13)*, volume 28, pages

1238–1246. JMLR Workshop and Conference Proceedings, May 2013.

E. Kaufmann and S. Kalyanakrishnan. Information complexity in bandit subset selection. In *Proceeding of the 26th Conference On Learning Theory.*, 2013.

Emilie Kaufmann, Olivier Cappé, and Aurélien Garivier. On the complexity of best-arm identification in multi-armed bandit models. *The Journal of Machine Learning Research*, 17(1):1–42, 2016.

Lisha Li, Kevin G Jamieson, Giulia DeSalvo, Afshin Rostamizadeh, and Ameet Talwalkar. Hyperband: A novel bandit-based approach to hyperparameter optimization. *Journal of Machine Learning Research*, 18:185–1, 2017.

Andrea Locatelli, Maurilio Gutzeit, and Alexandra Carpentier. An optimal algorithm for the thresholding bandit problem. In *International Conference on Machine Learning*, pages 1690–1698, 2016.

Shie Mannor, John N. Tsitsiklis, Kristin Bennett, and Nicol Cesa-bianchi. The sample complexity of exploration in the multi-armed bandit problem. *Journal of Machine Learning Research*, 5:2004, 2004.

Subhojyoti Mukherjee, Naveen Kolar Purushothama, Nandan Sudarsanam, and Balaraman Ravindran. Thresholding bandits with augmented ucb. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, pages 2515–2521. AAAI Press, 2017.

Max Simchowitz, Kevin Jamieson, and Benjamin Recht. The simulator: Understanding adaptive sampling in the moderate-confidence regime. In *Conference on Learning Theory*, pages 1794–1834, 2017.

Herbert A Simon. Rational choice and the structure of the environment. *Psychological review*, 63(2):129, 1956.

Yizao Wang, Jean yves Audibert, and Rémi Munos. Algorithms for infinitely many-armed bandits. In D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou, editors, *Advances in Neural Information Processing Systems 21*, pages 1729–1736. Curran Associates, Inc., 2009.