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

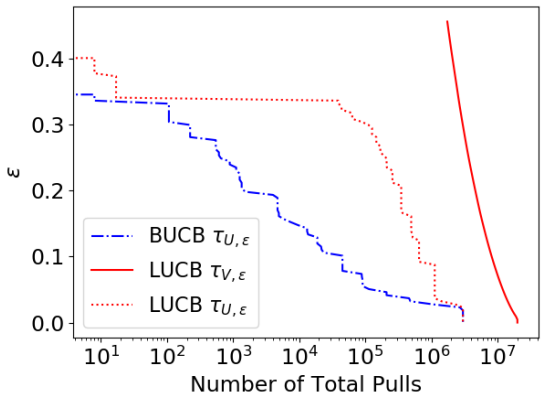

Figure 1: ϵ-GOOD ARM IDENTIFICATION

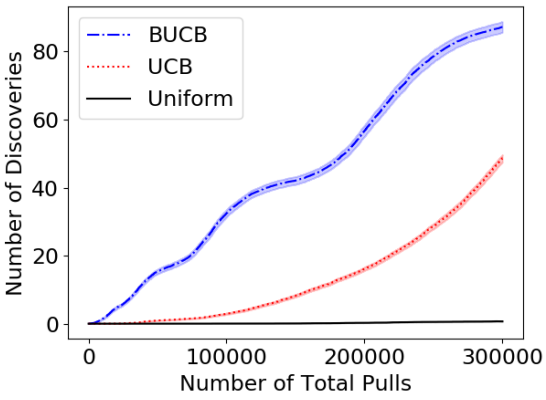

Figure 2: Identifying means above a threshold

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 $\mu_0, \mu_1, \ldots, \mu_n \in [0, 1]$.

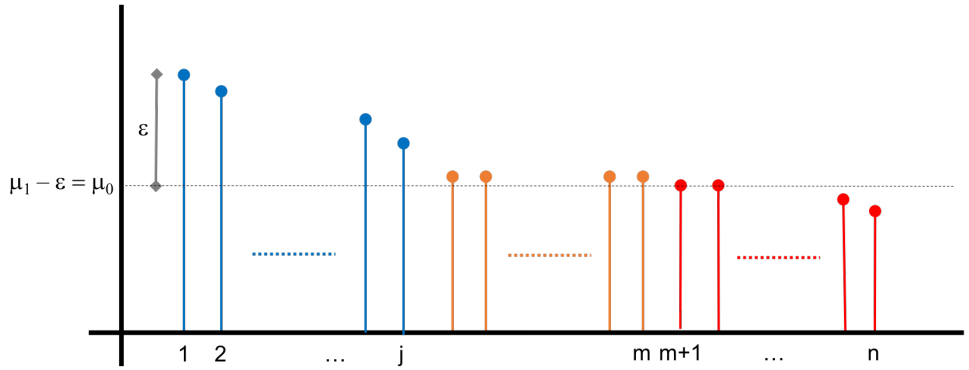

Figure 3: Our sample complexity results rely on picking a bracket of an appropriate size: $\frac{n}{m}$ is too small, n is too large, and $\frac{n}{j}$ appears to be about a good size.

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 $t \in \mathbb{N}$, $\mathbb{E}[\frac{|\mathcal{S}_t \cap \mathcal{H}_0|}{|\mathcal{S}_t| \vee 1}] \le 2\delta$ 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.

A Outline of Supplementary Material

We briefly outline the Supplementary Material. In Section B, we discuss in more detail related work on lower bounds on verifiable sample complexity for ϵ-GOOD ARM IDENTIFICATION. In Section C, we present the proof of our lower bound, namely, Theorem 1. In Section D, we give additional algorithms for the the k-IDENTIFICATIONS PROBLEM that have stronger guarantees but are less practical. In Section E, we prove the upper bound results of this paper including Theorems 2 and 3. In Section F, we provide an algorithm that has nearly optimal verifiable and unverifiable sample complexity (ignoring logarthmic factors). Finally, in Section G, we discuss the details of our experiments.

B Related work: $(\epsilon, \delta) - PAC$ for identifying k ϵ-good arms

Kaufmann et al. (2016) proved the following theorem which characterizes the sample complexity for ϵ-good arm identification $k = 1, m \geq 1$ and multiple identifications above a threshold μ_0 in the special case of $k = m$ (in general, we are interested in any $1 \leq k \leq m$) in the (ϵ, δ)-PAC setting.

Theorem 4 (Kaufmann et al. (2016)). *Fix $\epsilon, \delta > 0$, and a vector $\mu \in \mathbb{R}^n$. Fix a bandit instance ρ of n arms where the ith distribution equals $\rho_i(\mu) = \mathcal{N}(\mu_i, 1)$, a Gaussian distribution with mean μ_i and variance 1. Assume without loss of generality that $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_n$ and let $m = |\{i \in [n] : \mu_i \geq \mu_1 - \epsilon\}|$ so that $\mu_i \geq \mu_1 - \epsilon$ for all $i \in [m]$. If algorithm $\mathcal{A}$ returns $k = 1$ arms of the top m arms and is (ϵ, δ)-PAC on $\mathcal{P} = \{\mathcal{N}(\mu', I) : \mu' \in \mathbb{R}^n\}$ then*

$$\mathbb{E}_\rho\Big[\sum_{i=1}^n T_i(\tau_{PAC})\Big] \geq \tfrac{1}{2}\log(1/2.4\delta)\Big((m-1)\epsilon^{-2} + \sum_{i=m+1}^n (\mu_1 - \mu_i)^{-2}\Big) \qquad (k=1)$$

Under the same conditions, if $\mathcal{A}$ returns $k = m$ arms then

$$\mathbb{E}_\rho\Big[\sum_{i=1}^n T_i(\tau_{PAC})\Big] \geq 2\log(1/2.4\delta)\Big(\sum_{i=1}^m (\mu_i - \mu_{m+1})^{-2} + \sum_{i=m+1}^n (\mu_m - \mu_i)^{-2}\Big) \quad (k=m)$$

Note that by the definition of m we have that $\mu_m - \mu_{m+1} > 0$. We emphasize that the sample complexity of Theorem 4 for both $k = 1$ or $k = m$ is necessarily $\Omega(n)$ regardless of the number of ϵ-good arms m. As discussed below, the $k = 1$ lower bound is achievable up to log log factors Karnin et al. (2013). The special case of $k = m$ is notably the TOP-k identification problem where lower bounds were recently sharpened with additional log factors independently by Simchowitz et al. (2017); Chen et al. (2017). In particular, if for some μ_0 we have $\mu_i = \mu_0 + \epsilon$ for $i \leq m$ and $\mu_i = \mu_0$ for $i < m$ then their lower bounds on the expected sample complexity scale like $k\epsilon^{-2}\log(n-k) + (n-k)\epsilon^{-2}\log(k)$, which is always larger than $n\epsilon^{-2}$ that is predicted by the above theorem.

C Proof of lower bounds

We now briefly provide some intuition behind the proof. Suppose $m > 1$ and $k = 1$ and consider the easier problem where the permutation set averaged over is just the identity permutation $\pi_1 = (1, 2, \ldots, n)$ and the permutation π_2 that swaps $\{1, \ldots, m\}$ and some fixed $\sigma \subset [n] \setminus [m]$ with $|\sigma| = m$. That is, the algorithm knows the instance it is playing is either $\pi_1(\rho) = \rho$ or $\pi_2(\rho)$ where ρ is known but the permutation π_1 or π_2 is not. Information theoretic arguments say that at least $\tau \approx \min_{i\in\sigma}(\mu_1 - \mu_i)^{-2}$ observations from $[m] \cup \sigma$ are necessary in order to determine whether the underlying instance is $\pi_1(\rho)$ versus $\pi_2(\rho)$. But if the algorithm cannot distinguish between π_1 and π_2 with fewer than τ samples, then we can also argue that if π_1 and π_2 are chosen with equal probability, then taking nearly τ samples from the arms in σ with sub-optimal means is unavoidable in expectation. The choice of σ was arbitrary and there are $\frac{n}{m} - 1$ disjoint choices (e.g., $\{m+1, \ldots, 2m\}, \{2m+1, \ldots, 3m\}, \ldots$) resulting in a lower bound of about $\frac{1}{m}\sum_{i=m+1}^n (\mu_1 - \mu_i)^{-2}$.

The $k > 1$ case is trickier because if we used just π_1 and π_2 as above, as soon as we found just one ϵ-good arm (and thus being able to accurately discern whether the instance is $\pi_1(\rho)$ or $\pi_2(\rho)$) the algorithm would immediately know of $m - 1$ other ϵ-good arms. To overcome this, we choose a large enough set $\sigma \subset [m]$ such that $\sigma \cap \widehat{S}$ is non-empty with constant probability on the identity permutation. This way, if we swap this set $\sigma \subset [m]$ with some other set in $[n] \setminus [m]$ of size $|\sigma|$, then the algorithm would error with constant probability on

this alternative permutation. The next lemma guarantees the existence of such a set of size $\lceil m/k \rceil$ and the final result follows from the fact that there are about $\frac{n}{\lceil m/k \rceil}$ such disjoint choices in $[n] \setminus [m]$.

We introduce the following notation: for any $j \leq m$ let $\binom{[m]}{j}$ denote all subsets of $\{1, \ldots, m\}$ of size j.

Lemma 1. *Fix $m \in \mathbb{N}$ and let S be a random subset of size $k \leq m$ drawn from an arbitrary distribution over $\binom{[m]}{k}$. For any $\ell \leq m - k$ there exists a subset $\sigma \subset [m]$ with $|\sigma| = \ell$ such that*

$$\mathbb{P}(\sigma \cap S \neq \emptyset) \geq 1 - \binom{m-k}{\ell} \Big/ \binom{m}{\ell} \geq 1 - e^{-\ell k/m}$$

If $\ell > m - k$ then $\mathbb{P}(\sigma \cap S \neq \emptyset) = 1$.

Proof. Because the max of a set of positive numbers is always at least the average, we have

$$\begin{aligned}
\max_{\sigma \in \binom{[m]}{\ell}} \mathbb{P}(\sigma \cap S \neq \emptyset) &\geq \frac{1}{\binom{m}{\ell}} \sum_{\sigma \in \binom{[m]}{\ell}} \mathbb{P}(\sigma \cap S \neq \emptyset) \\
&= \frac{1}{\binom{m}{\ell}} \sum_{\sigma \in \binom{[m]}{\ell}} \sum_{s \in \binom{[m]}{k}} \mathbb{P}(S = s) \mathbf{1}\{\sigma \cap s \neq \emptyset) \\
&= \frac{1}{\binom{m}{\ell}} \sum_{s \in \binom{[m]}{k}} \mathbb{P}(S = s) \sum_{\sigma \in \binom{[m]}{\ell}} \mathbf{1}\{\sigma \cap s \neq \emptyset) \\
&= \frac{1}{\binom{m}{\ell}} \sum_{s \in \binom{[m]}{k}} \mathbb{P}(S = s) \left(\binom{m}{\ell} - \binom{m-k}{\ell} \right) \\
&= 1 - \binom{m-k}{\ell} \Big/ \binom{m}{\ell}
\end{aligned}$$

where the last line follows from the fact that $\sum_{s \in \binom{[m]}{k}} \mathbb{P}(S = s) = 1$ because it is a probability distribution. Now

$$\begin{aligned}
\binom{m-k}{\ell} \Big/ \binom{m}{\ell} &= \frac{(m-k)!\,(m-\ell)!}{(m-k-\ell)!\,m!} \\
&= \prod_{i=0}^{k-1} \frac{m-i-\ell}{m-i} = \prod_{i=0}^{k-1} \left(1 - \frac{\ell}{m-i}\right) \leq \prod_{i=0}^{k-1} \left(1 - \frac{\ell}{m}\right) \leq e^{-\ell k/m}.
\end{aligned}$$

□

Fix any $\sigma \subset [m]$ with $|\sigma| = \lceil m/k \rceil$ that satisfies $\mathbb{P}_\rho\left(\widehat{S} \cap \sigma \neq \emptyset\right) \geq 1 - e^{-1}$ (which must exist by the above lemma). Now fix any $\sigma' \subset [n] \setminus [m]$ with $|\sigma'| = |\sigma|$ and define ρ' as swapping the arms of σ and σ', maintaining their relative ordering of the indices within the sets. Note that by the correctness assumption at the relative stopping times of ρ and ρ' we have

$$\mathbb{P}_\rho(\widehat{S} \subset [m]) \geq 1 - \delta, \qquad \mathbb{P}_{\rho'}(\widehat{S} \cap \sigma \neq \emptyset) \leq \delta, \qquad \mathbb{P}_\rho(\widehat{S} \cap \sigma \neq \emptyset) \geq 1 - e^{-1}$$

which implies

$$\mathrm{TV}(\mathbb{P}_\rho, \mathbb{P}_{\rho'}) = \sup_{\mathcal{E}} |\mathbb{P}_\rho(\mathcal{E}) - \mathbb{P}_{\rho'}(\mathcal{E})| \geq |\mathbb{P}_\rho(\widehat{S} \cap \sigma \neq \emptyset) - \mathbb{P}_{\rho'}(\widehat{S} \cap \sigma \neq \emptyset)| \geq 1 - \delta - e^{-1}. \tag{9}$$

Remark 1. *Given* (9)*, one is tempted to apply Pinsker's inequality to obtain the right-hand-side of Lemma 1 from Kaufmann et al. (2016) and then provide a lower bound on $\mathbb{E}_\rho[\sum_{i \in \sigma \cup \sigma'} T_i]$. The difficulty here is that once we cover $[n] \setminus [m]$ with alternative σ' sets, they would all share the same σ in this lower bound, which suggests putting all samples on σ and a trivial lower bound. Alternatively, one could consider using the technique of Chen et al. (2017) which compares a given instance to a degenerate instance where the means of σ' would be copied to σ and argue that the probability of error is at least $1/2$ since there truly is no difference. This strategy is successful if $k = 1$ so that $|\sigma| = m$ but breaks down when $k > 1$ because one cannot reason about what the algorithm would have to do if the means of σ were changed like one could if $k = 1$. Consequently, we employ the use of the Simulator argument from Simchowitz et al. (2017) that is much more powerful at the cost of the introduction of some machinery.*

The Simulator (background)

The simulator argument is a kind of thought experiment where the player is playing against a non-stationary distribution. In the real game when the player pulls arm $I_t = i$ arm at time t she observes a sample from the ith distribution of instance ρ: $X_{i,t} \sim \rho_i$. However, when playing against the simulator she observes a sample form the ith distribution of an instance denoted $\text{Sim}(\rho, \{I_1, \dots, I_t\})$ that depends on all past requests: $X_{i,t} \sim \text{Sim}(\rho, \{I_1, \dots, I_t\})_i$ with probability law Q given ρ, $\{I_s = i_s\}_{s=1}^t$. That is, instead of receiving rewards from a stationary distribution ρ at each time t, the simulator is an instance that depends on all the indices of past pulls (but not their values). For any set $A \subset \mathbb{R}$ define

$$\mathbb{P}_{\text{Sim}(\rho,(i_1,\dots,i_t))}\left(X_{i_t,t} \in A\right) \coloneqq Q\left(X_{i_t,t} \in A | \rho, \{I_s = i_s\}_{s=1}^t\right).$$

We allow the algorithm to have internal randomness with probability law P so that for $B \subset [n]$ define

$$\mathbb{P}_{\text{Alg}((i_1,x_1,\dots,i_{t-1},x_{t-1}))}\left(I_t \in B\right) \coloneqq P\left(I_t \in B | \{I_s = i_s, X_{I_s} = x_s\}_{s=1}^{t-1}\right)$$

so that for any event $E \in \mathcal{F}_T$ we define

$$\begin{aligned}
&\mathbb{P}_{\text{Alg},\text{Sim}(\rho)}(E) \\
&\coloneqq \sum_{i_1,\dots,i_T} \int_{x_1,\dots,x_T} \mathbf{1}_E \prod_{t=1}^T Q\left(X_{I_t} = x_t | \rho, \{I_s = i_s\}_{s=1}^t\right) P\left(I_t = i_t | \{I_s = i_s, X_{I_s} = x_s\}_{s=1}^{t-1}\right) dx_1 \dots dx_T \\
&= \sum_{i_1,\dots,i_T} \int_{x_1,\dots,x_T} \mathbf{1}_E \prod_{t=1}^T \mathbb{P}_{\text{Sim}(\rho,(i_1,\dots,i_t))}\left(X_{I_t} = x_t\right) \mathbb{P}_{\text{Alg}((i_1,x_1,\dots,i_{t-1},x_{t-1}))}\left(I_t = i_t\right) dx_1 \dots dx_T
\end{aligned}$$

so that for any T we have $KL\left(\mathbb{P}_{\text{Alg},\text{Sim}(\rho)}, \mathbb{P}_{\text{Alg},\text{Sim}(\rho')}\right) =$

$$\begin{aligned}
&\sum_{i_1,\dots,i_T} \int_{x_1,\dots,x_T} \mathbb{P}_{\text{Alg},\text{Sim}(\rho)}(\{I_s = i_s, X_{I_s} = x_s\}_{s=1}^T) \log\left(\frac{\mathbb{P}_{\text{Alg},\text{Sim}(\rho)}(\{I_s = i_s, X_{I_s} = x_s\}_{s=1}^T)}{\mathbb{P}_{\text{Alg},\text{Sim}(\rho')}(\{I_s = i_s, X_{I_s} = x_s\}_{s=1}^T)}\right) dx_1 \dots dx_T \\
&= \sum_{i_1,\dots,i_T} \int_{x_1,\dots,x_T} \mathbb{P}_{\text{Alg},\text{Sim}(\rho)}(\{I_s = i_s, X_{I_s} = x_s\}_{s=1}^T) \log\left(\frac{\prod_{t=1}^T \mathbb{P}_{\text{Sim}(\rho,(i_1,\dots,i_t))}\left(X_{I_t} = x_t\right)}{\prod_{t=1}^T \mathbb{P}_{\text{Sim}(\rho',(i_1,\dots,i_t))}\left(X_{I_t} = x_t\right)}\right) dx_1 \dots dx_T \\
&= \sum_{t=1}^T \sum_{i_1,\dots,i_T} \int_{x_1,\dots,x_T} \mathbb{P}_{\text{Alg},\text{Sim}(\rho)}(\{I_s = i_s, X_{I_s} = x_s\}_{s=1}^T) \log\left(\frac{\mathbb{P}_{\text{Sim}(\rho,(i_1,\dots,i_t))}\left(X_{I_t} = x_t\right)}{\mathbb{P}_{\text{Sim}(\rho',(i_1,\dots,i_t))}\left(X_{I_t} = x_t\right)}\right) dx_1 \dots dx_T \\
&= \sum_{t=1}^T \sum_{i_1,\dots,i_T} \mathbb{P}_{\text{Alg},\text{Sim}(\rho)}(\{I_s = i_s\}_{s=1}^T) \int_{x_t} \mathbb{P}_{\text{Sim}(\rho,(i_1,\dots,i_t))}\left(X_{I_t} = x_t\right) \log\left(\frac{\mathbb{P}_{\text{Sim}(\rho,(i_1,\dots,i_t))}\left(X_{I_t} = x_t\right)}{\mathbb{P}_{\text{Sim}(\rho',(i_1,\dots,i_t))}\left(X_{I_t} = x_t\right)}\right) dx_t \\
&= \sum_{t=1}^T \sum_{i_1,\dots,i_T} \mathbb{P}_{\text{Alg},\text{Sim}(\rho)}(\{I_s = i_s\}_{s=1}^T) KL\left(\mathbb{P}_{\text{Sim}(\rho,(i_1,\dots,i_t))}, \mathbb{P}_{\text{Sim}(\rho',(i_1,\dots,i_t))}\right) \\
&= \sum_{i_1,\dots,i_T} \mathbb{P}_{\text{Alg},\text{Sim}(\rho)}(\{I_s = i_s\}_{s=1}^T) \sum_{t=1}^T KL\left(\mathbb{P}_{\text{Sim}(\rho,(i_1,\dots,i_t))}, \mathbb{P}_{\text{Sim}(\rho',(i_1,\dots,i_t))}\right) \\
&\le \max_{i_1,\dots,i_T} \sum_{t=1}^T KL\left(\mathbb{P}_{\text{Sim}(\rho,(i_1,\dots,i_t))}, \mathbb{P}_{\text{Sim}(\rho',(i_1,\dots,i_t))}\right)
\end{aligned}$$

The simulator will be defined so that the right hand side is always finite for any T. When it is clear from context we will simply write $\mathbb{P}_\rho(E)$ or $\mathbb{P}_{\text{Sim}(\rho)}(E)$ to represent $\mathbb{P}_{\text{Alg},\rho}(E)$ or $\mathbb{P}_{\text{Alg},\text{Sim}(\rho)}(E)$, respectively. Let $\Omega_t = \{I_1, \dots, I_t\}$ denote the history of all arm pulls requested by the player up to time t. Note that Ω_t is a multi-set so that $|\Omega_t| = t$.

Definition 5. *We say an event W is* truthful *under a simulator* Sim *with respect to instance ρ if for all events $E \in \mathcal{F}_T$*

$$\mathbb{P}_\rho(E \cap W) = \mathbb{P}_{\text{Sim}(\rho,\Omega_T)}(E \cap W).$$

Lemma 2 (Simchowitz et al. (2017)). *Let $\rho^{(1)}$ and $\rho^{(2)}$ be two instances, $\mathrm{Sim}(\cdot,\cdot)$ be a simulator, and let W_i be two truthful $\mathcal{F}_T$-measureable events under $\mathrm{Sim}(\rho^{(i)},\Omega_T)$ for $i=1,2$ where Ω_T is the history of pulls up to a stopping time T. Then*

$$\mathbb{P}_{\rho^{(1)}}(W_1^c)+\mathbb{P}_{\rho^{(2)}}(W_2^c)\geq \mathrm{TV}(\rho^{(1)},\rho^{(2)})-Q\left(KL\left(\mathbb{P}_{\mathrm{Alg},\mathrm{Sim}(\rho^{(1)})},\mathbb{P}_{\mathrm{Alg},\mathrm{Sim}(\rho^{(2)})}\right)\right)$$

where $Q(\beta)=\min\{1-\frac{1}{2}e^{-\beta},\sqrt{\beta/2}\}$.

Constructing the Simulator

Recall the definitions of ρ,ρ' and σ,σ' from above. For some $\tau\in\mathbb{N}$ and multiset Ω of requested arm pulls, define $W_\sigma(\Omega)=\{\sum_{i\in\Omega}\mathbf{1}\{i\in\sigma\}\leq\tau\}$ and $W_{\sigma'}(\Omega)=\{\sum_{i\in\Omega}\mathbf{1}\{i\in\sigma'\}\leq\tau\}$. For these events, an instance $\nu\in\{\rho,\rho'\}$, and any multiset Ω_t denoting the indices the player has played up to the current time t, define a simulator

$$\mathrm{Sim}(\nu,\Omega_t)_i=\begin{cases}\nu_i & \text{if } i\notin\sigma\cup\sigma'\\ \nu_i & \text{if } i\in\sigma\cup\sigma',\ W_\sigma(\Omega_t)\cap W_{\sigma'}(\Omega_t)\\ \nu_i & \text{if } i\in\sigma,\ W_\sigma(\Omega_t)\cup W^c_{\sigma'}(\Omega_t)\\ \nu_i & \text{if } i\in\sigma',\ W^c_\sigma(\Omega_t)\cup W_{\sigma'}(\Omega_t)\\ \rho_i & \text{if } i\in\sigma,\ W_\sigma(\Omega_t)^c\cap W_{\sigma'}(\Omega_t)\\ \rho_{\sigma(\sigma'^{-1}(i))} & \text{if } i\in\sigma',\ W_\sigma(\Omega_t)\cup W^c_{\sigma'}(\Omega_t)\end{cases}$$

where $\sigma(i)$ denotes the ith element of σ and $\sigma^{-1}(i)\in\{1,\dots,|\sigma|\}$ so that $\sigma(\sigma'^{-1}(i))\in\sigma$ for any $i\in\sigma'$.

Observe that $W_{\sigma'}(\Omega_t)$ is truthful under $\mathrm{Sim}(\cdot,\Omega_t)$ with respect to ρ since if $W_{\sigma'}(\Omega_t)$ occurs $\mathrm{Sim}(\rho,\Omega_t)_i=\rho_i$ for all $i\in[n]$ and all $t\in\mathbb{N}$ by construction. Similarly, $W_\sigma(\Omega_t)$ is truthful under $\mathrm{Sim}(\cdot,\Omega_t)$ to ρ'. Note that $\mathrm{Sim}(\rho,\Omega_t)_i=\mathrm{Sim}(\rho',\Omega_t)_i$ for all $i\in[n]\setminus\sigma\cup\sigma'$ and if $\min\{\sum_{j\in\Omega_t}\mathbf{1}\{j\in\sigma\},\sum_{j\in\Omega_t}\mathbf{1}\{j\in\sigma'\}\}>\tau$ then $\mathrm{Sim}(\rho,\Omega_t)_i=\mathrm{Sim}(\rho',\Omega_t)_i$ for all $i\in[n]$. Therefore, we can easily upper bound the KL divergence:

$$\begin{aligned}\max_{i_1,\dots,i_T\in[n]}\sum_{t=1}^T KL\left(\mathrm{Sim}(\rho,\{i_s\}_{s=1}^t),\mathrm{Sim}(\rho',\{i_s\}_{s=1}^t)\right)&\leq\max_{i\in\sigma}\tau KL(\rho_i,\rho'_i)+\max_{j\in\sigma'}\tau KL(\rho_j,\rho'_j)\\&=\max_{i=1,\dots,\ell}\tau(\mu_{\sigma(i)}-\mu_{\sigma'(i)})^2.\end{aligned}$$

As shown in (Simchowitz et al., 2017, Lemma 1) averaging over all permutations is equivalent to constructing a symmeterized version of the algorithm such that given any bandit instance, the algorithm randomly permutes the arms internally and then after making its set selection, returns the set inverted by the randomly chosen permutation. This modified algorithm is symmetric in the sense that

$$\mathbb{P}_\rho((i_1,\dots,i_T,s)=(I_1,\dots,I_T,\widehat{S}))=\mathbb{P}_{\pi(\rho)}((i_1,\dots,i_T,,s)=(\pi(I_1),\dots,\pi(I_T),\pi(\widehat{S}))).$$

In what follows, we assume the algorithm is symmetric which, in particular, implies

$$\mathbb{P}_\rho(W^c_{\sigma'})+\mathbb{P}_{\rho'}(W^c_\sigma)=2\mathbb{P}_\rho(W^c_{\sigma'}).$$

Putting all the pieces together we have

$$\begin{aligned}\mathbb{P}_\rho\left(\sum_{i\in\sigma'}T_i>\tau\right)=\mathbb{P}_\rho(W^c_{\sigma'})&=\frac{1}{2}\left(\mathbb{P}_\rho(W^c_{\sigma'})+\mathbb{P}_{\rho'}(W^c_\sigma)\right)\\&\geq\frac{1}{2}\left(1-\delta-e^{-1}-\sqrt{\tau\max_{i=1,\dots,\ell}(\mu_{\sigma(i)}-\mu_{\sigma'(i)})^2/2}\right)\\&>\frac{1}{2}(1/8-\delta)\end{aligned}$$

if $\tau = \frac{1}{2\max_{i=1,\ldots,\ell}(\mu_{\sigma(i)}-\mu_{\sigma'(i)})^2}$. By Markov's inequality, $\mathbb{E}_\rho[\sum_{i\in\sigma'} T_i] \geq \tau\mathbb{P}_\rho\left(\sum_{i\in\sigma'} T_i > \tau\right)$. Noting that $\sigma' \subset [n]\setminus[m]$ was arbitrary, we apply the above calculation for all connected subsets of size $\lceil m/k\rceil$

$$\begin{aligned}
\mathbb{E}_\rho\left[\sum_{i=m+1}^{n} T_i\right] &\geq \frac{1}{4}(1/8-\delta)\sum_{r=1}^{(n-m)k/m}(\mu_1-\mu_{m+rm/k})^{-2} \\
&\geq \frac{1}{4}(1/8-\delta)\frac{k}{m}\sum_{i=m+m/k+1}^{n}(\mu_1-\mu_i)^{-2} \\
&\geq \frac{1}{4}(1/8-\delta)\left[-(\mu_1-\mu_{m+1})^{-2}+\frac{k}{m}\sum_{i=m+1}^{n}(\mu_1-\mu_i)^{-2}\right] \\
&\geq \frac{1}{64}\left[-(\mu_1-\mu_{m+1})^{-2}+\frac{k}{m}\sum_{i=m+1}^{n}(\mu_1-\mu_i)^{-2}\right]
\end{aligned}$$

where the last line follows since $\delta \in (0, \frac{1}{16})$.

C.1 Unverifiable Sample Complexity of LUCB and Median Elimination

We note that a very wide class of algorithms satisfy the two conditions in the following proposition.

Proposition 1. *Let $\epsilon \in (0,\frac{1}{2})$ and $\delta \in (0,\frac{1}{4})$. Let $\mathcal{A}$ be any algorithm that* (i) *begins by pulling every arm once and* (ii) *for all $t \in \mathbb{N}$, for all $i,j \in [n]$ if $\widehat{\mu}_{i,T_i(t)} > \widehat{\mu}_{j,T_j(t)}$ and $T_i(t) \geq T_j(t)$, then $\widehat{S}_t \neq j$. Then, there exists a problem instance ρ such that*

$$\mathbb{E}_{\pi\sim\mathbb{S}^n}\mathbb{E}_{\pi(\rho)}[\tau_{U,\epsilon,\delta}] \geq \frac{n}{4}, \qquad \mathcal{H}_{\text{low},1}(\epsilon) = \frac{1}{64}\epsilon^{-2}.$$

Proof. Define

$$\rho_i = \begin{cases} \text{bernoulli}(1/2+\epsilon) & i \in [n/2] \\ \text{bernoulli}(1/2) & i \in \{n/2+1,\ldots,n\} \end{cases}.$$

Let $X_{i,j}$ denote the jth iid realization of arm i. Define the event $E = \{\sum_{i\in[n]\setminus[n/2]} X_{i,1} \geq \frac{n}{4}\}$. Note that E occurs with probability at least $1/2$. Consider $t = n$, the round at which $\mathcal{A}$ has pulled all arms once. Define the event $F = \{\widehat{S}_t \in [n/2]\}$, the event that an ϵ-good arm occurs. Since the arms have been randomly permuted before the beginning of the game, notice that all of the arms in $G = \{i \in [n] : X_{i,1} = 1\}$ are statistically indistinguishable. Therefore, at time $t = n$, since the Algorithm outputs one of the arms in G,

$$\mathbb{P}(\widehat{S}_t \notin [n/2]|E) \geq \frac{|G\cap\{n/2+1,\ldots,n\}|}{|G|} \geq \frac{\frac{n}{4}}{n} = \frac{1}{4}.$$

Thus, since $\delta \in (0,1/4)$, we have that conditional on E, $\tau_{U,\epsilon,\delta} \geq n$. This implies that

$$\mathbb{E}_{\pi\sim\mathbb{S}^n}\mathbb{E}_{\pi(\rho)}[\tau_{U,\epsilon,\delta}] \geq \mathbb{E}_{\pi\sim\mathbb{S}^n}\mathbb{E}_{\pi(\rho)}[\tau_{U,\epsilon,\delta}|E]\frac{1}{4} \geq \frac{n}{4}$$

□

D Additional Algorithms

In this section, we briefly introduce two additional algorithms that are very similar to the Algorithm 1 presented earlier but have stronger guarantees for the task of identifying means above a threshold. A FWER-TPR (family-wise error rate-true positive rate) guarantee outputs a set $\mathcal{Q}_t$ such that $\mathbb{P}(\exists t : \mathcal{Q}_t\cap\mathcal{H}_0 \neq \emptyset) \leq c\delta$ and $\mathbb{E}[|\mathcal{Q}_t\cap\mathcal{H}_1|] \geq (1-\delta)k$ for large enough t. A FWER-FWPD (family-wise error rate-family-wise probability of detection) guarantee is stronger since it requires that the outputted set $\mathcal{R}_t$ satisfies $\mathbb{P}(\exists t : \mathcal{R}_t \cap \mathcal{H}_0 \neq \emptyset) \leq c\delta$ and $|\mathcal{R}_t \cap \mathcal{H}_1| \geq k$ for large enough t. For more formal examples of these guarantees, see Theorems 6 and 8.

Algorithm 2 Infinite UCB Algorithm: FWER-TPR and FWER-FWPD

1: $\delta_r = \frac{\delta}{r^2}$, $\delta'_r = \frac{\delta_r}{6.4\log(36/\delta_r)}$ $R_0 = 0$, $\ell = 0$, $\mathcal{S}_0 = \emptyset$, $\mathcal{Q}_0 = \emptyset$
2: **for** $t = 1, 2, \ldots$ **do**
3: **if** $t \geq 2^\ell \ell$ **then**
4: Draw a set $A_{\ell+1}$ uniformly at random from $\binom{[n]}{M_{\ell+1}}$, 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** there exists $i \in A_{R_t} \setminus \mathcal{S}_t$ such that $T_{i,R_t}(t) = 0$ **then**
8: Pull an arm I_t belonging to $\{i \in A_{R_t} \setminus \mathcal{S}_t : T_{i,R_t}(t) = 0\}$
9: **else if** FWER-TPR **then**
10: Pull arm $I_t = \text{argmax}_{i \in A_{R_t} \setminus \mathcal{Q}_t} \widehat{\mu}_{i,R_t,T_{i,R_t}(t)} + U(T_{i,R_t}(t), \delta)$
11: $\mathcal{Q}_{t+1} = \mathcal{Q}_t \cup \{i \in A_{R_t} : \widehat{\mu}_{i,R_t,T_{i,R_t}(t)} - U(T_{i,R_t}(t), \frac{\delta}{|A_{R_t}|R_t^2}) \geq \mu_0\}$ % FWER Thm.6
12: **else if** FWER-FWPD **then**
13: $\xi_{t,R_t} = \max\{2|\mathcal{S}_t \cap A_{R_t}|, \frac{5}{3(1-4\delta_{R_t})}\log(1/\delta_{R_t})R_t^2\}$
14: Pull arm $I_t = \text{argmax}_{i \in A_{R_t} \setminus \mathcal{S}_t} \widehat{\mu}_{i,R_t,T_{i,R_t}(t)} + U(T_{i,R_t}(t), \frac{\delta}{\xi_{t,R_t}})$
15: $s(p) = \{i \in A_{R_t} : \widehat{\mu}_{i,R_t,T_{i,R_t}(t)} - U(T_{i,R_t}(t), \frac{p}{|A_{R_t}|}\delta'_{R_t} \geq \mu_0\}$
16: $\mathcal{S}_{t+1} = \mathcal{S}_t \cup s(\widehat{p})$ where $\widehat{p} = \max\{p \in [|A_{R_t}|] : |s(p)| \geq p\}$
17: **if** $\mathcal{S}_t \cap A_{R_t} \neq \emptyset$ **then**
18: $\nu_{t,R_t} = \max(|\mathcal{S}_t \cap A_{R_t}|, 1)$
19: Pull arm $J_t = \text{argmax}_{i \in \mathcal{S}_t \cap A_{R_t} \setminus \mathcal{R}_t} \widehat{\mu}_{i,R_t,T_{i,R_t}(t)} + U(T_{i,R_t}(t), \frac{\delta_{R_t}}{\nu_{t,R_t}})$
20: $\chi_{t,R_t} = |A_{R_t}| - (1 - 2\delta'_{R_t}(1 + 4\delta'_{R_t}))|\mathcal{S}_t \cap A_{R_t}| + \frac{4(1+4\delta'_{R_t})}{3}\log(5\log_2(|A_{R_t}|/\delta'_{R_t})/\delta'_{R_t})$
21: $\mathcal{R}_{t+1} = \mathcal{R}_t \cup \{i \in \mathcal{S}_t \cap A_{R_t} : \widehat{\mu}_{i,R_t,T_{i,R_t}(t)} - U(T_{i,R_t}(t), \frac{\delta}{\chi_{t,R_t}}) \geq \mu_0\}$ % FWER Thm.8

The algorithm suggests different sets depending on the objective. If FWER-TPR is desired, the algorithm maintains a set $\mathcal{Q}_t$ and adds arms whose lower confidence bounds are above the threshold μ_0 (Line 11). If FWER-FWPD is the goal, then an additional arm J_t is pulled each time based on an upper confidence bound criterion and arms are accepted into the set $\mathcal{R}_{t+1}$ (Line 21) if their lower confidence bound is above the threshold μ_0.

E Proofs of Upper Bounds

The proofs for the FDR-TPR result (the proof of Theorem 5 in Section E.1) should be read first. Then, one can read the proofs for any of the other results. We introduce some notation that we use throughout the proofs. We use c to denote a positive constant whose value may change from line to line. We also define $\log(x) := \max(\ln(x), 1)$. Define

$$\rho_{i,r} = \sup\{\rho \in (0,1] : \cap_{t=1}^{\infty}\{|\widehat{\mu}_{i,r,t} - \mu_i| \leq U(t,\rho)\}\}.$$

We note that $\{\rho_{i,r}\}_{i\in[n],r\in\mathbb{N}}$ are independent and $\mathbb{P}(\rho_{i,r} \leq \delta) \leq \delta$ since by definition of $U(\cdot,\cdot)$ for any bracket $r \in \mathbb{N}$ and $\alpha \in (0,1)$, $\mathbb{P}(\cap_{t=1}^{\infty}\{|\widehat{\mu}_{i,r,t} - \mu_i| \leq U(t,\alpha)) \geq 1 - \alpha$. We define

$$\mathcal{I}_r = \{i \in \mathcal{H}_1 \cap A_r : \rho_{i,r} \leq \delta\}.$$

to be those arms in bracket r whose empirical means concentrate well in the sense that $\rho_{i,r} \leq \delta$. We also define $U^{-1}(\gamma,\delta) = \min(t : U(t,\delta) \leq \gamma)$. It can be shown for a sufficiently large constant c that $U^{-1}(\gamma,\delta) \leq c\gamma^{-2}\log(\log(\gamma^{-2})/\delta)$. Recall that we make that simplifying assumption that $\mu_0, \mu_1, \ldots, \mu_n \in [0,1]$ and that we define $\log(x) := \max(\ln(x), 1)$.

We note that although all of our upper bounds apply to the expectation of a stopping time, it is possible to obtain high-probability bounds by arguing that with high probability there is an appropriately sized bracket with enough "good" arms, e.g., an ϵ-good arm. Unfortunately, this argument would lead to an upper bound that scales as $\log^2(1/\delta)$ and would lose the dependence on the individual gaps of the arms with mean greater than $\mu_1 - \epsilon$ or μ_0.

E.1 Proof of FDR-TPR

Recall the relevant notation that $\Delta_{i,j} \coloneqq \mu_i - \mu_j$ and $\Delta_{j,0} \coloneqq \mu_j - \mu_0$. We restate Theorem 3 from the main body of the paper with the doubly logarithmic terms. We only consider the gap-independent upper bound here; in the following section, we will prove a stronger result, which implies the the gap-dependent upper bound.

Theorem 5. *Let $\delta \leq (0, 1/40)$. Let $k \in [|\mathcal{H}_1|]$. For all $j \in [m]$, define*

$$\tilde{\mathcal{H}}_{\mathrm{id}}(\mu_0; j) \coloneqq \frac{n}{j} k \Delta_{j,0}^{-2} \log\left(\log(\tfrac{n}{j}k)\log(\Delta_{j,0}^{-2})/\delta\right).$$

Let $(\mathcal{F}_t)_{t\in\mathbb{N}}$ be the filtration generated by playing Algorithm 1 on problem ρ. Then, Algorithm 1 has the property that for all $t \in \mathbb{N}$, $\mathbb{E}[\frac{|\mathcal{S}_t \cap \mathcal{H}_0|}{|\mathcal{S}_t| \wedge 1}] \leq 2\delta$ and there exists a stopping time τ_k wrt $(\mathcal{F})_{t\in\mathbb{N}}$ such that

$$\mathbb{E}[\tau_k] \leq c \min_{k \leq j \leq m} \tilde{\mathcal{H}}_{\mathrm{id}}(\mu_0; j) \log(\tilde{\mathcal{H}}_{\mathrm{id}}(\mu_0; j)) \tag{10}$$

where c is a universal constant and for all $t \geq \tau_k$, $\mathbb{E}[|\mathcal{S}_t \cap \mathcal{H}_1|] \geq (1-\delta)k$.

We briefly sketch the proof. Let $j_0 \in \{k, \ldots, m\}$ minimize the upper bound (10). Then, there exists a bracket r_0 with size $\Theta(\frac{n}{j_0}k)$ such that with constant probability A_{r_0} has at least k arms in $[j_0]$ and the empirical means concentrate well enough (defined formally in Lemma 4 as the event $E_{r_0} \coloneqq E_{r_0} \cap E_{0,r_0} \cap E_{1,r_0}$). The argument controls $\mathbb{E}[\tau_k]$ by partitioning the sample space according to which bracket $r_0 + s$ is the first such that the good event E_{r_0+s} occurs, i.e., according to $\{E_{r_0}, E_{r_0}^c \cap E_{r_0+1}, E_{r_0}^c \cap E_{r_0+1}^c \cap E_{r_0+2}, ...\}$. Lemma 4 shows that $\mathbb{E}[\mathbf{1}\{E_{r_0}\}\tau_k]$ has the same upper bound as (10) and that $\mathbb{E}[\mathbf{1}\{E_{r_0+s}\}\tau_k]$ has an upper bound that is larger than line (10) by a factor exponential in s. On the other hand, because the brackets are independent and growing exponentially in size, the probability of $E_{r_0+s} \cap (\cap_{r=0}^{s-1} E_{r_0+r}^c)$ decreases exponentially in s, enabling control of the exponential increase in $\mathbb{E}[\mathbf{1}\{E_{r_0+s}\}\tau_{r_0+s,k}]$ and, by extension, $\mathbb{E}[\tau_k]$.

Lemma 3 bounds the false discovery rate of Algorithm 1.

Lemma 3. *For all $t \in \mathbb{N}$, $\mathbb{E}[\frac{|\mathcal{S}_t \cap \mathcal{H}_0|}{|\mathcal{S}_t| \vee 1}] \leq 2\delta$.*

Proof.

$$\begin{aligned}
\mathbb{E}[\frac{|\mathcal{S}_t \cap \mathcal{H}_0|}{|\mathcal{S}_t| \vee 1}] &\leq \mathbb{E}[\frac{\sum_{l=1}^{\infty} |\mathcal{S}_t \cap A_l \cap \mathcal{H}_0|}{|\mathcal{S}_t| \vee 1}] \\
&\leq \sum_{l=1}^{\infty} \mathbb{E}[\frac{|\mathcal{S}_t \cap A_l \cap \mathcal{H}_0|}{|\mathcal{S}_t \cap A_l| \vee 1}] \\
&\leq \delta \sum_{l=1}^{\infty} \frac{1}{l^2} \\
&= \delta \frac{\pi^2}{6}
\end{aligned}$$

where we used Lemma 1 of Jamieson and Jain (2018).

□

Lemma 4, below, is the key result for establishing Theorem 5. For $k \in [|\mathcal{H}_1|]$ and $j_0 \in \{k, \ldots, |\mathcal{H}_1|\}$, it bounds the expected number of iterations that it takes a bracket r (of size at least $2^r \geq k$) to add k arms to the set $\mathcal{S}_t$ when the events $E_r \cap E_{0,r} \cap E_{1,r}$ occur where

$$\begin{aligned}
E_r &= \{|[j_0] \cap A_r| \geq k\}, \\
E_{0,r} &= \{\sum_{i \in \mathcal{H}_0 \cap A_r} \Delta_{j_0,i}^{-2} \log(\frac{1}{\rho_{i,r}}) \leq 5 \sum_{i \in \mathcal{H}_0 \cap A_r} \Delta_{j_0,i}^{-2} \log(\frac{1}{\delta})\}, \\
E_{1,r} &= \{\sum_{i \in [j_0] \cap A_r} \Delta_{i \vee j_0,0}^{-2} \log(\frac{1}{\rho_{i,r}}) \leq 5 \sum_{i \in [j_0] \cap A_r} \Delta_{i \vee j_0,0}^{-2} \log(\frac{1}{\delta})\}.
\end{aligned}$$

Event E_r says that there are at least k arms in A_r with $\mu_i \geq \mu_{j_0}$. The event $E_{0,r}$ says that the empirical means of the arms in $\mathcal{H}_0 \cap A_r$ concentrate well on the whole; event $E_{1,r}$ makes the analogous claim about $[j_0] \cap A_r$. We remark that the the events $E_{0,r}$ and $E_{1,r}$ allow us to avoid using a union bound.

Lemma 4. *Fix $\delta \in (0, 1/40)$, $k \in [|\mathcal{H}_1|]$, $j_0 \in \{k, \ldots, |\mathcal{H}_1|\}$, and $r \in \mathbb{N}$ such that $2^r \geq k$. 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 τ_k wrt $(\mathcal{F}_t)_{t\in\mathbb{N}}$ such that for all $t \geq \tau_k$, $\mathbb{E}[|\mathcal{S}_t \cap \mathcal{H}_1|] \geq (1-\delta)k$, and*

$$\mathbb{E}[\mathbf{1}\{E_r \cap E_{0,r} \cap E_{1,r}\}\tau_k] \leq c[2^{r-1}(r-1) + |A_r|\Delta_{j_0,0}^{-2} \log(r\frac{\log(\Delta_{j_0,0}^{-2})}{\delta}) \log(|A_r|\Delta_{j_0,0}^{-2} \log(r\frac{\log(\Delta_{j_0,0}^{-2})}{\delta}))] \tag{11}$$

where c is a universal constant.

Proof. **Step 1: Define stopping time.** Define

$$\tau_k = \min(t \in \mathbb{N} \cup \{\infty\} : |A_r \cap [j_0]| \geq k \text{ and } \mathcal{I}_r \cap A_r \cap \mathcal{H}_1 \subset \mathcal{S}_t).$$

Observe that for all $t \geq \tau_k$, $\mathbb{E}[|\mathcal{S}_t \cap \mathcal{H}_1|] \geq (1-\delta)k$ since for $t \geq \tau_k$

$$\mathbb{E}[|\mathcal{S}_t \cap \mathcal{H}_1|] \geq \mathbb{E}[|\mathcal{I}_r \cap A_r \cap \mathcal{H}_1|] \geq (1-\delta)|A_r \cap \mathcal{H}_1| \geq (1-\delta)k.$$

Step 2: Relate to bracket r.

In the interest of brevity, define $E \coloneqq E_r \cap E_{0,r} \cap E_{1,r}$ and since we will only focus on bracket r, write $\widehat{\mu}_{i,t}$, $T_i(t)$, $\mathcal{I}$, and ρ_i instead of $\widehat{\mu}_{i,r,t}$, $T_{i,r}(t)$, $\mathcal{I}_r$, and $\rho_{i,r}$. We will bound the number of rounds until $\mathcal{I} \cap A_r \cap \mathcal{H}_1 \subset \mathcal{S}_t$. Define

$$T = |\{t \in \mathbb{N} : \mathcal{I} \cap A_r \cap [j_0] \not\subset \mathcal{S}_t \text{ and } R_t = r\}|,$$

i.e., the number of rounds that the algorithm works on the rth bracket and $\mathcal{I} \cap A_r \cap \mathcal{H}_1 \not\subset \mathcal{S}_t$.

Next, we bound the number of brackets $r+s$ that are opened before $\mathcal{I} \cap A_r \cap [j_0] \subset \mathcal{S}_t$. The $r+1$ bracket is opened after bracket r is sampled 2^r times and similarly the $r+s$th bracket is opened after bracket r is sampled $\sum_{i=0}^{s-1} 2^{r+i} \geq 2^{r+s-1}$ times. Thus,

$$2^{r+s-1} \leq T \implies r+s-1 \leq \log(T).$$

So while $\mathcal{I} \cap A_r \cap \mathcal{H}_1 \not\subset \mathcal{S}_t$, every time bracket r is sampled, at most $\log(T)$ total brackets are sampled. Thus, we have that once the algorithm starts working on bracket r, after

$$\log(T)T \tag{12}$$

additional rounds, we have that $\mathcal{I} \cap A_r \cap [j_0] \subset \mathcal{S}_t$.

We note that after $2^{r-1}(r-1)$ rounds, the algorithm starts working on bracket r. Thus,

$$\begin{aligned} \mathbf{1}\{E\}\tau_k &\leq [2^{r-1}(r-1) + \mathbf{1}\{E\}\log(T)T] \\ &= [2^{r-1}(r-1) + \log(\mathbf{1}\{E\}T)\mathbf{1}\{E\}T] \end{aligned} \tag{13}$$

Step 3: Bounding $\mathbf{1}\{E\}T$. Note that we can write

$$\begin{aligned}
\mathbf{1}\{E\}T &= \mathbf{1}\{E\}\sum_{t=1}^{\infty}\mathbf{1}\{[j_0]\cap\mathcal{I}\cap A_r\not\subset\mathcal{S}_t, R_t=r\}\\
&= \mathbf{1}\{E\}\sum_{t:R_t=r}^{\infty}\mathbf{1}\{[j_0]\cap\mathcal{I}\cap A_r\not\subset\mathcal{S}_t\}\\
&\le \mathbf{1}\{E\}\sum_{t:R_t=r}^{\infty}\mathbf{1}\{[j_0]\cap\mathcal{I}\cap A_r\not\subset\mathcal{S}_t, I_t\in\mathcal{H}_0\}\\
&\qquad + \mathbf{1}\{[j_0]\cap\mathcal{I}\cap A_r\not\subset\mathcal{S}_t, I_t\in\mathcal{H}_1\cap[j_0]^c\} + \mathbf{1}\{I_t\in[j_0]\}\\
&\le \mathbf{1}\{E\}\sum_{t:R_t=r}^{\infty}\mathbf{1}\{[j_0]\cap\mathcal{I}\cap A_r\not\subset\mathcal{S}_t, I_t\in\mathcal{H}_0\}\\
&\qquad + \mathbf{1}\{[j_0]\cap\mathcal{I}\cap A_r\not\subset\mathcal{S}_t, I_t\in\mathcal{H}_1\cap[j_0]^c, \widehat{\mu}_{I_t,T_{I_t}(t)}<\mu_0+\frac{\Delta_{j_0,0}}{2}\}\\
&\qquad + \mathbf{1}\{I_t\in\mathcal{H}_1\cap[j_0]^c, \widehat{\mu}_{I_t,T_{I_t}(t)}\ge\mu_0+\frac{\Delta_{j_0,0}}{2}\} + \mathbf{1}\{I_t\in[j_0]\}
\end{aligned}$$

To begin, we bound the first sum.

For any $l\in\mathcal{I}\cap[j_0]\cap A_r$ we have $\rho_l\ge\delta$ by definition, so

$$\widehat{\mu}_{l,T_l(t)}+U(T_l(t),\delta)\ge\mu_l-U(T_l(t),\rho_l)+U(T_l(t),\delta)\ge\mu_l\ge\mu_{j_0}.$$

For any $i\in\mathcal{H}_0\cap A_r$,

$$\widehat{\mu}_{i,T_i(t)}+U(T_i(t),\delta)\le\mu_i+U(T_i(t),\rho_i)+U(T_i(t),\delta)\le\mu_i+2U(T_i(t),\rho_i\delta).$$

Thus, $\widehat{\mu}_{i,T_i(t)}+U(T_i(t),\delta)\le\mu_{j_0}$ if $T_i(t)\ge U^{-1}(\frac{\Delta_{j_0,i}}{2},\rho_i\delta)$, so that arm i would not be pulled this many times as long as $[j_0]\cap\mathcal{I}\cap A_r\not\subset\mathcal{S}_t$. Thus,

$$\begin{aligned}
\mathbf{1}\{E\}\sum_{t:R_t=r}^{\infty}\mathbf{1}\{[j_0]\cap\mathcal{I}\cap A_r\not\subset\mathcal{S}_t, I_t\in\mathcal{H}_0\} &\le \mathbf{1}\{E\}\sum_{i\in\mathcal{H}_0\cap A_r}U^{-1}(\frac{\Delta_{j_0,i}}{2},\rho_i\delta)\\
&\le \mathbf{1}\{E\}\sum_{i\in\mathcal{H}_0\cap A_r}c\Delta_{j_0,i}^{-2}\log(\frac{\log(\Delta_{j_0,i}^{-2})}{\delta\rho_i})\\
&= \mathbf{1}\{E\}\sum_{i\in\mathcal{H}_0\cap A_r}c\Delta_{j_0,i}^{-2}\log(\frac{\log(\Delta_{j_0,i}^{-2})}{\delta})+c\Delta_{j_0,i}^{-2}\log(\frac{1}{\rho_i})\\
&\le \mathbf{1}\{E\}\sum_{i\in\mathcal{H}_0\cap A_r}c'\Delta_{j_0,i}^{-2}\log(\frac{\log(\Delta_{j_0,i}^{-2})}{\delta})\\
&\le \sum_{i\in\mathcal{H}_0\cap A_r}c'\Delta_{j_0,i}^{-2}\log(\frac{\log(\Delta_{j_0,i}^{-2})}{\delta}) \qquad (14)
\end{aligned}$$

where the second to last inequality follows from $E_r\subseteq E$.

Next, we consider the second sum. If $[j_0]\cap\mathcal{I}\cap A_r\not\subset\mathcal{S}_t$, for any arm i satisfying $\widehat{\mu}_{i,T_i(t)}<\mu_0+\frac{\Delta_{j_0,0}}{2}$, we have that

$$\widehat{\mu}_{i,T_i(t)}+U(T_i(t),\delta)<\mu_0+\frac{\Delta_{j_0,0}}{2}+U(T_i(t),\delta)$$

so that if $T_i(t)\ge U^{-1}(\frac{\Delta_{j_0,0}}{2},\delta)$, then $\widehat{\mu}_{i,T_i(t)}+U(T_i(t),\delta)<\mu_{j_0}$ and therefore arm i is not pulled again until

$[j_0] \cap \mathcal{I} \cap A_r \subset \mathcal{S}_t$. Thus,

$$\begin{aligned}\sum_{t:R_t=r}^{\infty} \mathbf{1}\{[j_0] \cap \mathcal{I} \cap A_r \not\subset \mathcal{S}_t, I_t \in \mathcal{H}_1 \cap [j_0]^c \cap A_r, \widehat{\mu}_{I_t,T_{I_t}(t)} < \mu_0 + \frac{\Delta_{j_0,0}}{2}\} \\ \leq \sum_{i \in \mathcal{H}_1 \cap [j_0]^c \cap A_r} U^{-1}(\frac{\Delta_{j_0,0}}{2}, \delta) \\ \leq c|\mathcal{H}_1 \cap [j_0]^c \cap A_r|\Delta_{j_0,0}^{-2} \log(\frac{\log(\Delta_{j_0,0}^{-2})}{\delta}).\end{aligned}$$

Next, we bound the final summands

$$\mathbf{1}\{E\} \sum_{t:R_t=r}^{\infty} \mathbf{1}\{I_t \in \mathcal{H}_1 \cap [j_0]^c, \widehat{\mu}_{I_t,T_{I_t}(t)} \geq \mu_0 + \frac{\Delta_{j_0,0}}{2}\} + \mathbf{1}\{I_t \in [j_0]\}.$$

Let $p \leq |A_r|$. If $j \in \mathcal{H}_1 \cap [j_0]^c \cap A_r$ and $\widehat{\mu}_{j,T_j(t)} \geq \mu_0 + \frac{\Delta_{j_0,0}}{2}$, then

$$\widehat{\mu}_{j,T_j(t)} - U(T_j(t), \delta_r' \frac{p}{|A_r|}) \geq \mu_0 + \frac{\Delta_{j_0,0}}{2} - U(T_j(t), \delta_r' \frac{p}{|A_r|})$$

so that $\widehat{\mu}_{j,T_j(t)} - U(T_j(t), \delta_r' \frac{p}{|A_r|}) \geq \mu_0$ if $T_i(t) \geq U^{-1}(\frac{\Delta_{j_0,0}}{2}, \delta_r' \frac{p}{|A_r|})$, which implies that $j \in s(p)$.
Next, if $j \in [j_0] \cap A_r$, then

$$\begin{aligned}\widehat{\mu}_{j,T_j(t)} - U(T_j(t), \delta_r' \frac{p}{|A_r|}) &\geq \mu_j - U(T_j(t), \rho_j) - U(T_j(t), \delta_r' \frac{p}{|A_r|}) \\ &\geq \mu_j - 2U(T_j(t), \rho_j \delta_r' \frac{p}{|A_r|})\end{aligned}$$

so that $\widehat{\mu}_{j,T_j(t)} - U(T_j(t), \delta_r' \frac{p}{|A_r|}) \geq \mu_0$ if $T_i(t) \geq U^{-1}(\frac{\mu_j - \mu_0}{2}, \rho_j \delta_r' \frac{p}{|A_r|})$, which implies that $j \in s(p)$.

While there is *some* p associated with each arm when it is added to $s(p)$ and then consequently to $\mathcal{S}_t$, we don't know the order in or time at which particular arms are added. However, in the worst case, the arms of $\mathcal{H}_1$ are added one at a time to $\mathcal{S}_t$ instead of in a big group so that the first reqires $p = 1$, the second $p = 2$, etc. Letting

$\Gamma = \{f : f : \mathcal{H}_1 \longrightarrow [|H_1|] \text{ is a bijection}\}$,

$$
\begin{aligned}
&\mathbf{1}\{E\} \sum_{t:R_t=r}^{\infty} \mathbf{1}\{I_t \in \mathcal{H}_1 \cap [j_0]^c, \widehat{\mu}_{I_t,T_{I_t}(t)} \geq \mu_0 + \frac{\Delta_{j_0,0}}{2}\} + \mathbf{1}\{I_t \in [j_0]\} \\
&\leq \mathbf{1}\{E\} c \max_{\sigma\in\Gamma} \left(\sum_{j\in\mathcal{H}_1\cap[j_0]^c\cap A_r} U^{-1}(\frac{\Delta_{j_0,0}}{2}, \delta_r' \frac{\sigma(j)}{|A_r|}) + \sum_{j\in[j_0]\cap A_r} U^{-1}(\frac{\mu_j-\mu_0}{2}, \rho_j\delta_r' \frac{\sigma(j)}{|A_r|}) \right) \\
&\leq \mathbf{1}\{E\} c \max_{\sigma\in\Gamma} \left(\sum_{j\in\mathcal{H}_1\cap[j_0]^c\cap A_r} \Delta_{j_0,0}^{-2} \log(\frac{|A_r|}{\sigma(j)} \frac{\log(\Delta_{j_0,0}^{-2})}{\delta_r'}) + \sum_{j\in[j_0]\cap A_r} \Delta_{j,0}^{-2} \log(\frac{|A_r|}{\sigma(j)} \frac{\log(\Delta_{j,0}^{-2})}{\rho_j\delta_r'}) \right) \\
&= \mathbf{1}\{E\} c \max_{\sigma\in\Gamma} \Bigg(\sum_{j\in\mathcal{H}_1\cap[j_0]^c\cap A_r} \Delta_{j_0,0}^{-2} \log(\frac{|A_r|}{\sigma(j)} \frac{\log(\Delta_{j_0,0}^{-2})}{\delta_r'}) \\
&\qquad + \sum_{j\in[j_0]\cap A_r} \Delta_{j,0}^{-2} \log(\frac{|A_r|}{\sigma(j)} \frac{\log(\Delta_{j,0}^{-2})}{\delta_r'}) + \sum_{j\in[j_0]\cap A_r} \Delta_{j,0}^{-2} \log(\frac{1}{\rho_j}) \Bigg) \\
&= \mathbf{1}\{E\} c \max_{\sigma\in\Gamma} \Bigg(\sum_{j\in\mathcal{H}_1\cap[j_0]^c\cap A_r} \Delta_{j_0,0}^{-2} \log(\frac{|A_r|}{\sigma(j)} \frac{\log(\Delta_{j_0,0}^{-2})}{\delta_r'}) \\
&\qquad + \sum_{j\in[j_0]\cap A_r} \Delta_{j,0}^{-2} \log(\frac{|A_r|}{\sigma(j)} \frac{\log(\Delta_{j,0}^{-2})}{\delta_r'}) + 5 \sum_{j\in[j_0]\cap A_r} \Delta_{j,0}^{-2} \log(\frac{1}{\delta}) \Bigg) \\
&\leq c' \max_{\sigma\in\Gamma} \sum_{i\in\mathcal{H}_1\cap A_r} \Delta_{i\vee j_0,0}^{-2} \log(\frac{|A_r|}{\sigma(i)} r^2 \frac{\log(\Delta_{i\vee j_0,0}^{-2})}{\delta}) \qquad (15) \\
&\leq c' \sum_{i=1}^{|\mathcal{H}_1\cap A_r|} \Delta_{j_0,0}^{-2} \log(\frac{|A_r|}{i} r^2 \frac{\log(\Delta_{j_0,0}^{-2})}{\delta}) \\
&\leq c'' |A_r| \Delta_{j_0,0}^{-2} \log(r \frac{\log(\Delta_{j_0,0}^{-2})}{\delta}) \qquad (16)
\end{aligned}
$$

where the last line follows from the fact that for any $p \leq |A_r|$, $\sum_{i=1}^{p} \log(\frac{|A_r|}{i}) \leq |A_r|$.

Step 4: finishing bound (11)**.** Using lines (16) and (13),

$$\mathbf{1}\{E\}\tau_k \leq c'[2^{r-1}(r-1) + \log(|A_r|\Delta_{j_0,0}^{-2} \log(r\frac{\log(\Delta_{j_0,0}^{-2})}{\delta}))|A_r|\Delta_{j_0,0}^{-2} \log(r\frac{\log(\Delta_{j_0,0}^{-2})}{\delta})]$$

deterministically, which yields line (11).

□

Proof of Theorem 5. As in the proof of Lemma 4, define

$$
\begin{aligned}
\tau_k &= \min(t \in \mathbb{N} \cup \{\infty\} : \exists s \text{ such that } |A_s \cap [j_0]| \geq k \text{ and } \mathcal{I}_s \cap A_s \cap \mathcal{H}_1 \subset \mathcal{S}_t), \\
\tau_k^{(r)} &= \min(t \in \mathbb{N} \cup \{\infty\} : |A_r \cap [j_0]| \geq k \text{ and } \mathcal{I}_r \cap A_r \cap \mathcal{H}_1 \subset \mathcal{S}_t)
\end{aligned}
$$

As was argued in Step 1 of the proof of Lemma 4, for all $t \geq \tau_k$, $\mathbb{E}[|\mathcal{S}_t \cap \mathcal{H}_1|] \geq (1-\delta)k$.

Step 1: A lower bound on the probability of a good event. Let $j_0 \in \{k, \ldots, m\}$ minimize (10). Define $E_r = E_r \cap E_{0,r} \cap E_{1,r}$. We note that since $\{\rho_{i,r}\}_{i\in[n],r\in\mathbb{N}}$ and the brackets $\{A_r\}_{r\in\mathbb{N}}$ are independent, $\{E_r\}_{r\in\mathbb{N}}$ are independent events. Let r_0 be the smallest integer such that

$$\min(40\frac{n}{j_0}k, n) \leq 2^{r_0} \leq 80\frac{n}{j_0}k,$$

Note that if $2^{r_0} \geq n$, then the bracket r_0 has n arms.

Next, we bound $\mathbb{P}(E_{r_0}^c)$. If $2^{r_0} \geq n$, then $\mathbb{P}(E_{r_0}^c) = 0$, so assume that $2^{r_0} < n$. Note that since the elements of A_{r_0} are chosen uniformly from $[n]$ and $|A_{r_0}| = 2^{r_0} \geq 40\frac{n}{j_0}k$ we have that

$$\begin{aligned}\mathbb{E}[|[j_0] \cap A_{r_0}|] &= \frac{j_0}{n}|A_{r_0}| \\ &\geq 40k.\end{aligned}$$

Then, by a Chernoff bound for hypergeometric random variables,

$$\mathbb{P}(|[j_0] \cap A_{r_0}| \leq 20k) \leq \exp(-\frac{1}{8}40k) \leq \exp(-5).$$

Thus, E_{r_0} occurs with probability at least $1 - \exp(-5)$. Furthermore, we note that for any $r \geq r_0$, $\mathbb{P}(E_r^c) \leq \exp(-5)$.

Furthermore, by Lemma 8 of Jamieson and Jain (2018), for any $r \in \mathbb{N}$ and $i = 0, 1$,

$$\mathbb{P}(E_{i,r}^c) = \mathbb{E}[\mathbb{P}(E_{i,r}^c | A_r)] \leq \delta.$$

Finally, note that for every $r \geq r_0$ and any $\delta \in (0, 1/40)$ we have

$$\mathbb{P}(E_r^c) \leq \exp(-5) + 2\delta \leq \frac{1}{16}.$$

Furthermore, we claim that $\mathbb{P}(\cap_{l=r_0}^{\infty} E_l^c) = 0$. Let $s \geq r_0$; then, using the independence between brackets,

$$\mathbb{P}(\cap_{l=r_0}^{\infty} E_l^c) \leq \mathbb{P}(\cap_{l=r_0}^{s} E_l^c) = \frac{1}{16^s} \longrightarrow 0$$

as $s \longrightarrow \infty$, proving the claim.

Step 2: Gap-Independent bound on the number of samples. For the sake of brevity, write τ instead of τ_k and $\tau^{(r)}$ instead of $\tau_k^{(r)}$. Then, by the independence between brackets, the fact that $\cup_{r=r_0}^{\infty} E_r \cap (\cap_{r_0 \leq l < r} E_l^c)$ occurs with probability 1, and line 11 of Lemma 4,

$$\begin{aligned}
\mathbb{E}[\tau] &= \mathbb{E}[\tau \mathbf{1}\{\cup_{r=r_0}^{\infty} E_r \cap (\cap_{r_0 \leq l < r} E_l^c)\}] \\
&\leq \sum_{r=r_0}^{\infty} \mathbb{E}[\tau \mathbf{1}\{E_r \cap (\cap_{r_0 \leq l < r} E_l^c)\}] \\
&\leq \sum_{r=r_0}^{\infty} \mathbb{E}[\tau^{(r)} \mathbf{1}\{E_r \cap (\cap_{r_0 \leq l < r} E_l^c)\}] \\
&= \sum_{r=r_0}^{\infty} \mathbb{E}[\tau^{(r)} \mathbf{1}\{E_r\}] \mathbb{P}(\cap_{r_0 \leq l < r} E_l^c) \\
&\leq \sum_{r=r_0}^{\infty} [2^{r-1}(r-1) + \log(|A_r|\Delta_{j_0,0}^{-2} \log(r\frac{\log(\Delta_{j_0,0}^{-2})}{\delta}))|A_r|\Delta_{j_0,0}^{-2} \log(r\frac{\log(\Delta_{j_0,0}^{-2})}{\delta}))] \frac{1}{16^{r-r_0}} \\
&\leq \sum_{s=0}^{\infty} [2^{r_0-1} \cdot 2^s (r_0 + s - 1) \\
&\quad + \log(2^s |A_{r_0}|\Delta_{j_0,0}^{-2} \log((r_0+s)\frac{\log(\Delta_{j_0,0}^{-2})}{\delta}))2^s |A_{r_0}|\Delta_{j_0,0}^{-2} \log((r_0+s)\frac{\log(\Delta_{j_0,0}^{-2})}{\delta}))] \frac{1}{16^s}.
\end{aligned}$$

We bound the first term as follows:

$$\begin{aligned}
\sum_{s=0}^{\infty} \frac{2^{r_0-1} \cdot 2^s (r_0 + s - 1)}{16^s} &= 2^{r_0-1} \sum_{s=0}^{\infty} \frac{(r_0 + s - 1)}{8^s} && (17) \\
&\leq c 2^{r_0} r_0 && (18) \\
&\leq c' \frac{n}{j_0} k \log(\frac{n}{j_0} k) \\
&\leq c'' \tilde{\mathcal{H}}_{\mathrm{id}}(\mu_0; j_0) \log(\tilde{\mathcal{H}}_{\mathrm{id}}(\mu_0; j_0)).
\end{aligned}$$

where the last inequality follows since $\Delta_{i,j}^{-2} \geq 1$ for all $i < j \in [n] \cup \{0\}$ since $\mu_0, \mu_1, \ldots, \mu_n \in [0,1]$.

We note that

$$\begin{aligned}\log((r_0+s)\frac{\log(\Delta_{j_0,0}^{-2})}{\delta})) &\leq c[\log(r_0\frac{\log(\Delta_{j_0,0}^{-2})}{\delta})) + \log(s\frac{\log(\Delta_{j_0,0}^{-2})}{\delta}))] \\ &\leq c'\log(r_0\frac{\log(\Delta_{j_0,0}^{-2})}{\delta})) + c\log(s)\end{aligned}$$

and

$$\begin{aligned}\log(2^s|A_{r_0}|\Delta_{j_0,0}^{-2}\log((r_0+s)\frac{\log(\Delta_{j_0,0}^{-2})}{\delta})) &= \log(|A_{r_0}|\Delta_{j_0,0}^{-2}\log((r_0+s)\frac{\log(\Delta_{j_0,0}^{-2})}{\delta})) + s \\ &\leq \log(|A_{r_0}|\Delta_{j_0,0}^{-2}c'\log(r_0\frac{\log(\Delta_{j_0,0}^{-2})}{\delta}) + c\log(s)) + s \\ &\leq c''\log(|A_{r_0}|\Delta_{j_0,0}^{-2}\log(r_0\frac{\log(\Delta_{j_0,0}^{-2})}{\delta}))) + c'''\log(\log(s)) + s \\ &\leq c''\log(|A_{r_0}|\Delta_{j_0,0}^{-2}\log(r_0\frac{\log(\Delta_{j_0,0}^{-2})}{\delta}))) + c''''s\end{aligned}$$

Then,

$$\begin{aligned}&\sum_{s=0}^{\infty}\log(2^s|A_{r_0}|\Delta_{j_0,0}^{-2}\log((r_0+s)\frac{\log(\Delta_{j_0,0}^{-2})}{\delta}))2^s|A_{r_0}|\Delta_{j_0,0}^{-2}\log((r_0+s)\frac{\log(\Delta_{j_0,0}^{-2})}{\delta}))\frac{1}{16^s} \\ &\leq \sum_{s=0}^{\infty}[c''\log(|A_{r_0}|\Delta_{j_0,0}^{-2}\log(r_0\frac{\log(\Delta_{j_0,0}^{-2})}{\delta}))) + c''''s]|A_{r_0}|\Delta_{j_0,0}^{-2}[c'\log(r_0\frac{\log(\Delta_{j_0,0}^{-2})}{\delta})) + c\log(s)]\frac{1}{8^s} \\ &\leq c''c'\log(|A_{r_0}|\Delta_{j_0,0}^{-2}\log(r_0\frac{\log(\Delta_{j_0,0}^{-2})}{\delta})))|A_{r_0}|\Delta_{j_0,0}^{-2}\log(r_0\frac{\log(\Delta_{j_0,0}^{-2})}{\delta})) \\ &+ c''''\log(|A_{r_0}|\Delta_{j_0,0}^{-2}\log(r_0\frac{\log(\Delta_{j_0,0}^{-2})}{\delta})))|A_{r_0}|\Delta_{j_0,0}^{-2} \\ &+ c''''''|A_{r_0}|\Delta_{j_0,0}^{-2}\log(r_0\frac{\log(\Delta_{j_0,0}^{-2})}{\delta})) + +c''''''|A_{r_0}|\Delta_{j_0,0}^{-2} \\ &\leq c''''''\log(|A_{r_0}|\Delta_{j_0,0}^{-2}\log(r_0\frac{\log(\Delta_{j_0,0}^{-2})}{\delta})))|A_{r_0}|\Delta_{j_0,0}^{-2}\log(r_0\frac{\log(\Delta_{j_0,0}^{-2})}{\delta}))\end{aligned}$$

Plugging in $|A_{r_0}|$ and r_0 yields the gap independent bound.

□

E.2 Proof of FWER-TPR

In this section, we prove an upper bound for the FWER-TPR version of our Algorithm (see Algorithm 2). We note that the gap-dependent upper bound in Theorem 3 follows as a corollary since whenever the FWER-TPR version of our Algorithm 2 accepts an arm, the FDR-TPR version of our Algorithm 1 accepts the same arm.

Theorem 6. *Let $\delta \in (0, 1/40)5$. Let $k \in [|\mathcal{H}_1|]$. For all $j \in \{k, \ldots, |\mathcal{H}_1|\}$ define*

$$\mathcal{H}_{\text{FWER}}(\mu_0; j) := \frac{k}{j}\left(\underbrace{\left\{\sum_{i=1}^{m}\Delta_{i\vee j,0}^{-2}\right\}}_{top\ arms}\log(\frac{nk}{j\delta}\log(\Delta_{i\vee j,0}^{-2})) + \underbrace{\sum_{i=m+1}^{n}\Delta_{j,i}^{-2}}_{bottom\ arms}\log(\frac{\log(\Delta_{j,i}^{-2})}{\delta})\right).$$

Let $(\mathcal{F}_t)_{t\in\mathbb{N}}$ be the filtration generated by playing Algorithm 2 on problem ρ. Then, Algorithm 2 has the property that $\mathbb{P}(\exists t : \mathcal{Q}_t \cap \mathcal{H}_0 \neq \emptyset) \leq 2\delta$ and there exists a stopping time τ_k wrt $(\mathcal{F})_{t\in\mathbb{N}}$ such that

$$\mathbb{E}[\tau_k] \leq c\min_{k\leq j\leq m}\mathcal{H}_{\text{FWER}}(\mu_0; j)\log(\mathcal{H}_{\text{FWER}}(\mu_0; j) + \Delta_{j,0}^{-2}\log(\tfrac{n}{j}k\log(\Delta_{j,0}^{-2})/\delta)) \tag{19}$$

and for all $t \geq \tau_k$, $\mathbb{E}[|\mathcal{Q}_t \cap \mathcal{H}_1|] \geq (1-\delta)k$.

E_r, $E_{0,r}$, and $E_{1,r}$ are defined as in Section E.1.

Lemma 5. *Fix* $\delta \in (0, 1/40)$, $k \in [|\mathcal{H}_1|]$, $j_0 \in \{k, \ldots, m\}$, *and* $r \in \mathbb{N}$ *such that* $2^r \geq k$. *Define*

$$U_r \coloneqq \frac{\min(2^r, n)}{n}[\sum_{i \in \mathcal{H}_1} \Delta_{i \vee j_0, 0}^{-2} \log(\min(2^r, n) r \frac{\log(\Delta_{i \vee j_0, 0}^{-2})}{\delta})) + \sum_{i \in \mathcal{H}_0} \Delta_{j_0, i}^{-2} \log(\frac{\log(\Delta_{j_0, i}^{-2})}{\delta})].$$

Let $(\mathcal{F}_t)_{t \in \mathbb{N}}$ *be the filtration generated by playing Algorithm 2 on problem* ρ. *Then, there exists a stopping time* τ_k *wrt* $(\mathcal{F}_t)_{t \in \mathbb{N}}$ *such that for all* $t \geq \tau_k$, $\mathbb{E}[|\mathcal{Q}_t \cap \mathcal{H}_1|] \geq (1-\delta)k$, *and*

$$\mathbb{E}[\mathbf{1}\{E_r \cap E_{0,r} \cap E_{1,r}\}\tau_k] \leq c[2^{r-1}(r-1) + U_r \log(U_r + \Delta_{j_0,0}^{-2} \log(\min(2^r, n) r \frac{\log(\Delta_{j_0,0}^{-2})}{\delta})))] \tag{20}$$

where c *is a universal constant.*

Remark 2. *Note that* $\mathcal{H}_{\mathrm{FWER}}(\mu_0; j) \approx U_{\lceil \log_2(\frac{nk}{j}) \rceil}$.

Proof. **Step 1: Define stopping time.** Define

$$\tau_k = \min(t \in \mathbb{N} \cup \{\infty\} : |A_r \cap [j_0]| \geq k \text{ and } \mathcal{I}_r \cap A_r \cap \mathcal{H}_1 \subset \mathcal{Q}_t).$$

Observe that for all $t \geq \tau_k$, $\mathbb{E}[|\mathcal{Q}_t \cap \mathcal{H}_1|] \geq (1-\delta)k$ since for $t \geq \tau_k$

$$\mathbb{E}[|\mathcal{Q}_t \cap \mathcal{H}_1|] \geq \mathbb{E}[|\mathcal{I}_s \cap A_s \cap \mathcal{H}_1|] \geq (1-\delta)|A_s \cap \mathcal{H}_1| \geq (1-\delta)k.$$

Let $r \in \mathbb{N}$. Define

$$T = |\{t \in \mathbb{N} : \mathcal{I} \cap A_r \cap [j_0] \not\subset \mathcal{Q}_t \text{ and } R_t = r\}|,$$

By the same argument used in Lemma 4 to obtain line (13),

$$\mathbf{1}\{E\}\tau_k \leq [2^{r-1}(r-1) + \log(\mathbf{1}\{E\}T)\mathbf{1}\{E\}T]. \tag{21}$$

We can use the same argument that was used to obtain line (14) and line (15) in Lemma 4 and the lower bounds $1 \leq \sigma(i)$ and $p \geq 1$ to obtain

$$\begin{aligned}
\mathbf{1}\{E\}T \leq c\Big(& \sum_{i \in \mathcal{H}_0 \cap A_r} \Delta_{j_0,i}^{-2} \log(\frac{\log(\Delta_{j_0,i}^{-2})}{\delta}) && (22)\\
& + |\mathcal{H}_1 \cap [j_0]^c \cap A_r| \Delta_{j_0,0}^{-2} \log(\frac{\log(\Delta_{j_0,0}^{-2})}{\delta}) + \sum_{i \in \mathcal{H}_1 \cap A_r} \Delta_{i \vee j_0, 0}^{-2} \log(|A_r| r^2 \frac{\log(\Delta_{i \vee j_0, 0}^{-2})}{\delta}) \Big) \\
\leq c'[& \sum_{i \in \mathcal{H}_0 \cap A_r} \Delta_{j_0,i}^{-2} \log(\frac{\log(\Delta_{j_0,i}^{-2})}{\delta}) + \sum_{i \in \mathcal{H}_1 \cap A_r} \Delta_{i \vee j_0, 0}^{-2} \log(|A_r| r \frac{\log(\Delta_{i \vee j_0, 0}^{-2})}{\delta})] && (23)\\
\coloneqq c' & S_r && (24)
\end{aligned}$$

where the second inequality follows from the fact that $\Delta_{i \vee j_0, 0} \geq \Delta_{j_0, 0}$ so the third term absorbs the second.

Using lines (23) and (21),

$$\mathbf{1}\{E\}\tau_k \leq c[2^{r-1}(r-1) + \log(S_r) S_r]$$

but note that now the bound depends on the particular random elements of $A_r \cap \mathcal{H}_0$ and $A_r \cap \mathcal{H}_1$.

Step 2: Bounding $\mathbb{E}[\log(S_r)S_r]$. Next, taking the expectation of both sides and focusing on the expectation of the second term,

$$\mathbb{E}[\log(S_r)S_r] = \sum_{i\in\mathcal{H}_0} \Delta_{j_0,i}^{-2}\log(\frac{\log(\Delta_{j_0,i}^{-2})}{\delta})\mathbb{E}[\mathbf{1}\{i\in A_r\}\log(S_r)] + \sum_{i\in\mathcal{H}_1}\Delta_{i\vee j_0,0}^{-2}\log(|A_r|r^2\frac{\log(\Delta_{i\vee j_0,0}^{-2})}{\delta})\mathbb{E}[\mathbf{1}\{i\in A_r\}\log(S_r)].$$

It suffices to bound the first sum since the argument for the second is the same.

$$\begin{aligned}
\mathbb{E}[\mathbf{1}\{j\in A_r\}\log(S_r)] &= \mathbb{E}[\log(S_r)|j\in A_r]\frac{\min(2^r,n)}{n} && (25)\\
&\le \log(\mathbb{E}[S_r|j\in A_r])\frac{\min(2^r,n)}{n} && (26)\\
&= \log(\frac{\min(2^r-1,n-1)}{n-1}[\sum_{i\in\mathcal{H}_1}\Delta_{i\vee j_0,0}^{-2}\log(\min(2^r,n)r\frac{\log(\Delta_{i\vee j_0,0}^{-2})}{\delta})) \\
&\quad + \sum_{i\in\mathcal{H}_0\setminus j}\Delta_{j_0,i}^{-2}\log(\frac{\log(\Delta_{j_0,i}^{-2})}{\delta})] + \Delta_{j_0,j}^{-2}\log(\frac{\log(\Delta_{j_0,j}^{-2})}{\delta}))\frac{\min(2^r,n)}{n} && (27)\\
&\le \log(S_r + \Delta_{j_0,j}^{-2}\log(\frac{\log(\Delta_{j_0,i}^{-2})}{\delta}))\frac{\min(2^r,n)}{n}, && (28)
\end{aligned}$$

where line (25) follows by the law of total expectation, line (26) follows by Jensen's inequality, and line (28) follows since $\frac{a}{b}\le\frac{a+1}{b+1}$ if $a\le b$. Thus, collecting terms,

$$\mathbb{E}[\mathbf{1}\{E_r\cap E_{0,r}\cap E_{1,r}\}\tau_{r,k}] \le U_r\log(U_r + \Delta_{j_0,0}^{-2}\log(\min(2^r,n)r\frac{\log(\Delta_{j_0,0}^{-2})}{\delta}))$$

yielding line (20). □

Proof of Theorem 6. **Step 1: Showing $\mathbb{P}(\exists t : \mathcal{Q}_t\cap\mathcal{H}_0\neq\emptyset)\le 2\delta$.** First, we show that $\mathbb{P}(\exists t : \mathcal{Q}_t\cap\mathcal{H}_0\neq\emptyset)\le 2\delta$.

$$\begin{aligned}
\mathbb{P}(\exists t : \mathcal{Q}_t\cap\mathcal{H}_0\neq\emptyset) &\le \sum_{r=1}^{\infty}\mathbb{P}(\exists t : \mathcal{Q}_t\cap A_r\cap\mathcal{H}_0\neq\emptyset)\\
&\le \sum_{r=1}^{\infty}\mathbb{P}(\exists t\in\mathbb{N}\text{ and } i\in\mathcal{H}_0\cap A_r : \widehat{\mu}_{i,r,T_{i,r}(t)} - U(T_{i,r}(t),\frac{\delta}{|A_r|r^2})\ge\mu_0)\\
&\le \sum_{r=1}^{\infty}\mathbb{P}(\exists t\in\mathbb{N}\text{ and } i\in\mathcal{H}_0\cap A_r : \widehat{\mu}_{i,r,T_{i,r}(t)} - U(T_{i,r}(t),\frac{\delta}{|A_r|r^2})\ge\mu_i)\\
&\le \sum_{r=1}^{\infty}|A_r\cap\mathcal{H}_0|\frac{\delta}{|A_r|r^2}\\
&\le \sum_{r=1}^{\infty}\frac{\delta}{r^2}\\
&\le \delta\frac{\pi^2}{6}
\end{aligned}$$

Step 2: Defining the stopping time. As in the proof of Lemma 5, define

$$\begin{aligned}
\tau_k &= \min(t\in\mathbb{N}\cup\{\infty\} : \exists s\text{ such that } |A_s\cap[j_0]|\ge k\text{ and }\mathcal{I}_s\cap A_s\cap\mathcal{H}_1\subset\mathcal{Q}_t),\\
\tau_k^{(r)} &= \min(t\in\mathbb{N}\cup\{\infty\} : |A_r\cap[j_0]|\ge k\text{ and }\mathcal{I}_r\cap A_r\cap\mathcal{H}_1\subset\mathcal{S}_t).
\end{aligned}$$

As was argued in Step 1 of the proof of Lemma 5, for all $t\ge\tau_k$, $\mathbb{E}[|\mathcal{Q}_t\cap\mathcal{H}_1|]\ge(1-\delta)k$ since for $t\ge\tau_k$.

Step 3: A lower bound on the probability of a good event. Let $j_0 \in \{k, \dots, m\}$ minimize (19). Define $E_r = E_r \cap E_{0,r} \cap E_{1,r}$. We note that since $\{\rho_{i,r}\}_{i\in[n],r\in\mathbb{N}}$ and the brackets $\{A_r\}_{r\in\mathbb{N}}$ are independent, $\{E_r\}_{r\in\mathbb{N}}$ are independent events. Let r_0 be the smallest integer such that

$$\min(40\frac{n}{j_0}k, n) \le 2^{r_0} \le 80\frac{n}{j_0}k,$$

Note that if $2^{r_0} \ge n$, then the bracket r_0 has n arms.

As was argued in the proof of Theorem 5 we have that

$$\mathbb{P}(E_r^c) \le \exp(-5) + 2\delta \le \frac{1}{16}.$$

and that $\mathbb{P}(\cap_{l=r_0}^{\infty} E_l^c) = 0$.

Step 4: Gap-Dependent bound on the number of samples. For the sake of brevity, write τ instead of τ_k and $\tau^{(r)}$ instead of $\tau_k^{(r)}$. Then, by the independence between brackets, the fact that $\cup_{r=r_0}^{\infty} E_r \cap (\cap_{r_0\le l<r} E_l^c)$ occurs with probability 1, and Lemma 5,

$$\begin{aligned}
\mathbb{E}[\tau] &= \mathbb{E}[\tau \mathbf{1}\{\cup_{r=r_0}^{\infty} E_r \cap (\cap_{r_0\le l<r} E_l^c)\}] \\
&\le \sum_{r=r_0}^{\infty} \mathbb{E}[\tau \mathbf{1}\{E_r \cap (\cap_{r_0\le l<r} E_l^c)\}] \\
&\le \sum_{r=r_0}^{\infty} \mathbb{E}[\tau^{(r)} \mathbf{1}\{E_r \cap (\cap_{r_0\le l<r} E_l^c)\}] \\
&= \sum_{r=r_0}^{\infty} \mathbb{E}[\tau^{(r)} \mathbf{1}\{E_r\}] \mathbb{P}(\cap_{r_0\le l<r} E_l^c) \\
&\le \sum_{r=r_0}^{\infty} c[2^{r-1}(r-1) + U_r \log(U_r + \Delta_{j_0,0}^{-2} \log(\min(2^r, n) r \frac{\log(\Delta_{j_0,0}^{-2})}{\delta}))] \frac{1}{16^{r-r_0}} \\
&\le \sum_{r=r_0}^{\infty} c[2^{r-r_0} \cdot 2^{r_0-1}(r-1) + 4^{r-r_0} U_{r_0} \log(4^{r-r_0} U_{r_0} + \Delta_{j_0,0}^{-2} \log(2^{r_0} \cdot 2^{r-r_0} r \frac{\log(\Delta_{j_0,0}^{-2})}{\delta}))] \frac{1}{16^{r-r_0}}
\end{aligned}$$

where we used Lemma 4 and the fact that $4^s U_r \ge U_{r+s}$ for any $s \ge 1$, which holds by the following argument

$$\begin{aligned}
4^s U_r &= 4^s \frac{\min(2^r, n)}{n} [\sum_{i\in\mathcal{H}_1} \Delta_{i\vee j_0,0}^{-2} \log(\min(2^r, n) r \frac{\log(\Delta_{i\vee j_0,0}^{-2})}{\delta})) + \sum_{i\in\mathcal{H}_0} \Delta_{j_0,i}^{-2} \log(\frac{\log(\Delta_{j_0,i}^{-2})}{\delta})] \\
&\ge \frac{\min(2^{r+s}, n)}{n} [\sum_{i\in\mathcal{H}_1} \Delta_{i\vee j_0,0}^{-2} \log(\min(2^{r2^s}, n) r \frac{\log(\Delta_{i\vee j_0,0}^{-2})}{\delta})) + \sum_{i\in\mathcal{H}_0} \Delta_{j_0,i}^{-2} \log(\frac{\log(\Delta_{j_0,i}^{-2})}{\delta})] \\
&\ge \frac{\min(2^{r+s}, n)}{n} [\sum_{i\in\mathcal{H}_1} \Delta_{i\vee j_0,0}^{-2} \log(\min(2^{r+s}, n) r \frac{\log(\Delta_{i\vee j_0,0}^{-2})}{\delta})) + \sum_{i\in\mathcal{H}_0} \Delta_{j_0,i}^{-2} \log(\frac{\log(\Delta_{j_0,i}^{-2})}{\delta})] \\
&= U_{r+s}.
\end{aligned}$$

Next, we can bound the first term using the same argument in line (18):

$$\begin{aligned}
\sum_{r=r_0}^{\infty} 2^{r-r_0} \cdot 2^{r_0-1}(r-1) \frac{1}{16^{r-r_0}} &\le c 2^{r_0} r_0 \\
&\le c' \frac{n}{j_0} k \log(\frac{n}{j_0} k) \\
&\le c'' \mathcal{H}_{\mathrm{FWER}}(\mu_0; j_0) \log(\mathcal{H}_{\mathrm{FWER}}(\mu_0; j_0)).
\end{aligned}$$

where the last inequality follows since $\Delta_{i,j}^{-2} \ge 1$ for all $i < j \in [n] \cup \{0\}$ since $\mu_0, \mu_1, \dots, \mu_n \in [0, 1]$.

Next, we bound the second term.

$$\begin{aligned}
&\sum_{r=r_0}^{\infty} \frac{1}{4^{r-r_0}} U_{r_0} \log(4^{r-r_0} U_{r_0} + \Delta_{j_0,0}^{-2} \log(2^{r_0} \cdot 2^{r-r_0}(s+r_0) \frac{\log(\Delta_{j_0,0}^{-2})}{\delta})))\\
&= \sum_{s=0}^{\infty} \frac{1}{4^{s}} U_{r_0} \log(4^{s} U_{r_0} + \Delta_{j_0,0}^{-2} \log(2^{r_0} \cdot 2^{s}(s+r_0) \frac{\log(\Delta_{j_0,0}^{-2})}{\delta})))\\
&\le \sum_{s=0}^{\infty} \frac{1}{4^{s}} U_{r_0} \log(4^{s} U_{r_0} + c' \Delta_{j_0,0}^{-2} \log(2^{r_0} r_0 \frac{\log(\Delta_{j_0,0}^{-2})}{\delta})) + c's)\\
&\le U_{r_0} \sum_{s=0}^{\infty} \frac{1}{4^{s}} [c'' \log(4^{s} U_{r_0} + c' \Delta_{j_0,0}^{-2} \log(2^{r_0} r_0 \frac{\log(\Delta_{j_0,0}^{-2})}{\delta}))) + c''' \log(s)]\\
&\le U_{r_0} \sum_{s=0}^{\infty} \frac{1}{4^{s}} [c'' \log(U_{r_0} + c' \Delta_{j_0,0}^{-2} \log(2^{r_0} r_0 \frac{\log(\Delta_{j_0,0}^{-2})}{\delta}))) + c'' \log(4^{s}) + c''' \log(s)]\\
&\le c'''' U_{r_0} \log(U_{r_0} + \Delta_{j_0,0}^{-2} \log(2^{r_0} r_0 \frac{\log(\Delta_{j_0,0}^{-2})}{\delta})))\\
&\le c''''' U_{r_0} \log(U_{r_0} + \Delta_{j_0,0}^{-2} \log(2^{r_0} \frac{\log(\Delta_{j_0,0}^{-2})}{\delta})))\\
&\le c'''''' \mathcal{H}_{\mathrm{FWER}}(\mu_0; j_0) \log(\mathcal{H}_{\mathrm{FWER}}(\mu_0; j_0) + \Delta_{j_0,0}^{-2} \log(2^{r_0} \frac{\log(\Delta_{j_0,0}^{-2})}{\delta})))
\end{aligned}$$

where we used $U_{r_0} \le c\mathcal{H}_{\mathrm{FWER}}(\mu_0; j_0)$ and

$$\begin{aligned}
\log(2^{r_0} \cdot 2^{s}(s+r_0) \frac{\log(\Delta_{j_0,0}^{-2})}{\delta})) &= \log(2^{r_0}(s+r_0) \frac{\log(\Delta_{j_0,0}^{-2})}{\delta})) + cs\\
&\le c' \log(2^{r_0} r_0 \frac{\log(\Delta_{j_0,0}^{-2})}{\delta})) + c' \log(2^{r_0} s \frac{\log(\Delta_{j_0,0}^{-2})}{\delta})) + cs\\
&\le c'' \log(2^{r_0} r_0 \frac{\log(\Delta_{j_0,0}^{-2})}{\delta})) + c'''' s
\end{aligned}$$

□

E.3 Proof of ϵ-Good Arm Identification

We restate Theorem 2 with the doubly logarithmic terms.

Theorem 7. *Let $\epsilon > 0$ and $\delta \in (0,1)$. Define $m = |\{i : \mu_i > \mu_1 - \epsilon\}|$. For all $j \in [m]$ define*

$$\mathcal{H}_{\mathrm{g}}(\epsilon; j) := \frac{1}{j} \left(\underbrace{\sum_{i=1}^{m} \Delta_{i \vee j, m+1}^{-2}}_{top\ arms} \log(\frac{n}{j\delta} \log(\Delta_{i \vee j, m+1}^{-2})) + \underbrace{\sum_{i=m+1}^{n} \Delta_{j,i}^{-2}}_{bottom\ arms} \log(\frac{\Delta_{j,i}^{-2}}{\delta}) \right).$$

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 τ wrt $(\mathcal{F})_{t \in \mathbb{N}}$ such that

$$\mathbb{E}[\tau] \le c \min_{j \in [m]} \mathcal{H}_{\mathrm{g}}(\epsilon; j) \log(\mathcal{H}_{\mathrm{g}}(\epsilon; j) + \Delta_{j,m+1}^{-2} \log(\frac{n}{j\delta} \log(\Delta_{j,m+1}^{-2}))) \tag{29}$$

and $\mathbb{P}(\exists s \ge \tau : \mu_{O_s} \le \mu_1 - \epsilon) \le 2\delta$.

Lemma 6 is the key intermediate result in the proof of Theorem 7; its role is similar to that of Lemma 4 in the proof of Theorem 5 and the proof is technically similar to the proof of Lemma 4. For any $r \in \mathbb{N}$ and $j \in [m]$

define the events

$$F_{r,1}^{(j)} = \{A_r \cap [j] \neq \emptyset\}$$
$$F_{r,2}^{(j)} = \{\sum_{i \in A_r : \mu_i < \frac{\mu_j + \mu_{m+1}}{2}} \Delta_{j,i}^{-2} \log(\frac{1}{\rho_{i,r}}) \leq 5 \sum_{i \in A_r : \mu_i < \frac{\mu_j + \mu_{m+1}}{2}} \Delta_{j,i}^{-2} \log(\frac{1}{\delta})\},$$
$$F_{r,3}^{(j)} = \{\exists i_0 \in A_r \cap [j] \text{ s.t.} \forall t \in \mathbb{N} : |\widehat{\mu}_{i_0,r,t} - \mu_{i_0}| \leq U(t,\delta)\}.$$

$F_{r,1}^{(j)}$ says that there is at least one arm in bracket r with mean at least $\mu_j \geq \mu_m$. $F_{r,2}^{(j)}$ allows us to avoid a union bound and says that most of the arms in bracket r with mean at most $\frac{\mu_j + \mu_{m+1}}{2}$ have large $\rho_{i,r}$. Finally, $F_{r,3}^{(j)}$ says that at least one of the arms in the rth bracket with mean at least $\mu_j \geq \mu_m$ that concentrates well in the sense that $\rho_{i,r} \geq \delta$.

Lemma 6. *Let $\epsilon > 0$, $m = |\{i : \mu_i > \mu_1 - \epsilon\}|$, $j_0 \in [m]$, and $r \in \mathbb{N}$. Define*

$$Y_r = \frac{\min(2^r, n)}{n}[\sum_{i=1}^{m} \Delta_{i \vee j_0, m+1}^{-2} \log(|A_r| r \frac{\log(\Delta_{i \vee j_0, m+1}^{-2})}{\delta}) + \sum_{i=m+1}^{n} \Delta_{j_0,i}^{-2} \log(\frac{\log(\Delta_{j_0,i}^{-2})}{\delta})]$$

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 τ wrt $(\mathcal{F}_t)_{t \in \mathbb{N}}$ such that $\mathbb{P}(\exists s \geq \tau : \mu_{O_s} \leq \mu_1 - \epsilon) \leq 2\delta$, and

$$\mathbb{E}[\mathbf{1}\{F_{r,1}^{(j_0)} \cap F_{r,2}^{(j_0)} \cap F_{r,3}^{(j_0)}\}\tau] \leq c[2^{r-1}(r-1) + Y_r \log(Y_r + \Delta_{j_0,m+1}^{-2} \log(|A_r| r \frac{\log(\Delta_{j_0,m+1}^{-2})}{\delta})]. \tag{30}$$

Remark 3. *Note that $\mathcal{H}_{\mathrm{g}}(\epsilon; j) \approx Y_{\lceil \log_2(\frac{n}{j}) \rceil}$.*

Proof. **Step 1: Define stopping time.** Our strategy is to define a stopping time τ that says that some arm i that is ϵ-good has been sampled enough times so that its confidence bound is sufficiently small and then to show that with high probability for all $t \geq \tau$, *(i)* the lower confidence bound of arm i is above μ_{m+1} and *(ii)* the algorithm always outputs an ϵ-good arm. To this end, define

$$\tau = \min\{t \in \mathbb{N} \cup \{\infty\} : \exists s \in \mathbb{N} \text{ and } \exists i \in A_s \text{ s.t. } \mu_i \geq \frac{\mu_{j_0} + \mu_{m+1}}{2} \text{ and } T_{i,s}(t) \geq U^{-1}(\frac{\Delta_{i \vee j_0, m+1}}{4}, \frac{\delta}{|A_s| s^2})\}.$$

We claim that $\mathbb{P}(\exists t \geq \tau : \mu_{O_t} < \mu_1 - \epsilon) \leq 2\delta$. Define the event

$$F = \{\forall t \in \mathbb{N}, s \in \mathbb{N}, \text{ and } i \in A_s : |\widehat{\mu}_{i,s,t} - \mu_i| \leq U(t, \frac{\delta}{|A_s| s^2})\}.$$

By a union bound, F occurs with probability at least $1 - 2\delta$. Suppose F occurs and let $t \geq \tau$. Then, since $t \geq \tau$, there exists a bracket s and an arm $i \in A_s$ such that $\mu_i \geq \frac{\mu_{j_0} + \mu_{m+1}}{2}$ and $T_{i,s}(t) \geq U^{-1}(\frac{\Delta_{i \vee j_0, m+1}}{4}, \frac{\delta}{|A_s| s^2})$. Then by event F,

$$\begin{aligned}
\widehat{\mu}_{i,s,T_{i,s}(t)} - U(T_{i,s}(t), \frac{\delta}{|A_s| s^2}) &\geq \mu_i - 2U(T_{i,s}(t), \frac{\delta}{|A_s| s^2}) \\
&> \mu_i - \frac{\Delta_{i \vee j_0, m+1}}{2} \\
&\geq \mu_{m+1}
\end{aligned}$$

where the last inequality follows by considering separately the cases *(i)* $\mu_i \geq \mu_{j_0}$ and *(ii)* $\mu_i < \mu_{j_0}$. Towards a contradiction, suppose that there exists a bracket $s_0 \in \mathbb{N}$ and another arm $j \in A_{s_0}$ ($j \neq i$) such that $\mu_j \leq \mu_1 - \epsilon$ and the algorithm outputs j at time t. Then, by event F,

$$\mu_j \geq \widehat{\mu}_{j,s_0,T_{j,s_0}(t)} - U(T_{j,s_0}(t), \frac{\delta}{|A_{s_0}| s_0^2}) \geq \widehat{\mu}_{i,s,T_{i,s}(t)} - U(T_{i,s}(t), \frac{\delta}{|A_s| s^2}) > \mu_{m+1} \geq \mu_j,$$

which is a contradiction. Thus, $\mathbb{P}(\exists t \geq \tau : \mu_{O_t} < \mu_1 - \epsilon) \leq 2\delta$.

Step 2: Relating τ to bracket r. Next, we bound $\mathbb{E}[\mathbf{1}\{F_{r,1}^{(j_0)} \cap F_{r,2}^{(j_0)} \cap F_{r,3}^{(j_0)}\}\tau]$. For the sake of brevity, we write $F_{r,i}$ instead of $F_{r,i}^{(j_0)}$ and define $F_r \coloneqq F_{r,1} \cap F_{r,2} \cap F_{r,3}$ and since we will only focus on bracket r, write $\widehat{\mu}_{i,t}$, $T_i(t)$, and ρ_i instead of $\widehat{\mu}_{i,r,t}$, $T_{i,r}(t)$, and $\rho_{i,r}$. Define

$$T = |\{t \in \mathbb{N} : R_t = r \text{ and } \nexists i \in A_r \text{ s.t. } \mu_i \geq \frac{\mu_{j_0} + \mu_{m+1}}{2} \text{ and } T_i(t) \geq U^{-1}(\frac{\Delta_{i \vee j_0, m+1}}{4}, \frac{\delta}{|A_r| r^2})\}|,$$

i.e., the number of rounds that the algorithm works on the rth bracket and there does not exist $i \in A_r$ s.t. $\mu_i \geq \frac{\mu_{j_0}+\mu_{m+1}}{2}$ and $T_i(t) \geq U^{-1}(\frac{\Delta_{i\vee j_0,m+1}}{2}, \frac{\delta}{|A_r|r^2})$. By the same argument given in line (13) in Lemma 4, we have that

$$\mathbf{1}\{F_r\}\tau \leq c[2^{r-1}(r-1) + \log(T\mathbf{1}\{F_r\})T\mathbf{1}\{F_r\}].$$

Step 3: Bounding $T\mathbf{1}\{F_r\}$. In the interest of brevity, define $F(t) = \{\nexists i \in A_r \text{ s.t. } \mu_i \geq \frac{\mu_{j_0}+\mu_{m+1}}{2} \text{ and } T_i(t) \geq U^{-1}(\frac{\Delta_{i\vee j_0,m+1}}{4}, \frac{\delta}{|A_r|r^2})\}$. Then,

$$\begin{aligned}
\mathbf{1}\{F_r\}T &\leq \mathbf{1}\{F_r\}\sum_{t=1}^{\infty} \mathbf{1}\{R_t = r, F(t)\} \\
&\leq \mathbf{1}\{F_r\}\sum_{t:R_t=r}^{\infty} \mathbf{1}\{\mu_{I_t} < \frac{\mu_{j_0}+\mu_{m+1}}{2}\} + \mathbf{1}\{\mu_{I_t} \geq \frac{\mu_{j_0}+\mu_{m+1}}{2}, F(t)\}
\end{aligned}$$

We bound each sum separately. Note that by $F_{r,3}$ there exists an $i_0 \in A_r \cap G_\gamma$ such that

$$\widehat{\mu}_{i_0,T_{i_0}(t)} + U(T_{i_0}(t),\delta) \geq \mu_{i_0} \geq \mu_{j_0}. \tag{31}$$

Let j such that $\mu_j < \frac{\mu_{j_0}+\mu_{m+1}}{2}$. Then,

$$\widehat{\mu}_{j,T_j(t)} + U(T_j(t),\delta) \leq \mu_j + U(T_j(t),\rho_j) + U(T_j(t),\delta) \leq \mu_j + 2U(T_j(t),\rho_j\delta).$$

Thus, line (31) implies that if $T_j(t) \geq U^{-1}(\frac{\Delta_{j_0,j}}{4}, \rho_j\delta)$, arm j is not pulled since in that case

$$\widehat{\mu}_{j,T_j(t)} + U(T_j(t),\delta) \leq \mu_j + 2U(T_j(t),\rho_j\delta) \leq \mu_j + \frac{\Delta_{j_0,j}}{2} \leq \mu_{j_0}.$$

Thus, by arguments made throughout this paper (e.g., line (14) of the proof of Lemma 4) and the event $F_{r,2}$,

$$\sum_{t:R_t=r}^{\infty} \mathbf{1}\{\mu_{I_t} \leq \mu_1 - \epsilon\} \leq c \sum_{j \in A_r : \mu_j < \frac{\mu_{j_0}+\mu_{m+1}}{2}} \Delta_{j_0,j}^{-2} \log(\frac{\log(\Delta_{j_0,j}^{-2})}{\delta})$$

Finally, by event F we clearly have

$$\sum_{t:R_t=r}^{\infty} \mathbf{1}\{\mu_{I_t} \geq \frac{\mu_{j_0}+\mu_{m+1}}{2}, F(t)\} \leq c \sum_{j \in A_r : \mu_j \geq \frac{\mu_{j_0}+\mu_{m+1}}{2}} \Delta_{j_0\vee j,m+1}^{-2} \log(|A_r| r \frac{\log(\Delta_{j_0\vee j,m+1}^{-2})}{\delta})$$

Thus,

$$\begin{aligned}
\mathbf{1}\{F_r\}T &\leq c[\sum_{j \in A_r : \mu_j < \frac{\mu_{j_0}+\mu_{m+1}}{2}} \Delta_{j_0,j}^{-2} \log(\frac{\log(\Delta_{j_0,j}^{-2})}{\delta}) + \sum_{j \in A_r : \mu_j \geq \frac{\mu_{j_0}+\mu_{m+1}}{2}} \Delta_{j\vee j_0,m+1}^{-2} \log(|A_r| r \frac{\log(\Delta_{j\vee j_0,m+1}^{-2})}{\delta})] \\
&\leq c[\sum_{j \in A_r : \mu_j \leq \mu_1 - \epsilon} \Delta_{j_0,j}^{-2} \log(\frac{\log(\Delta_{j_0,j}^{-2})}{\delta}) + \sum_{j \in A_r : \mu_j > \mu_1 - \epsilon} \Delta_{j\vee j_0,m+1}^{-2} \log(|A_r| r \frac{\log(\Delta_{j\vee j_0,m+1}^{-2})}{\delta})] \\
&\coloneqq cX_r
\end{aligned}$$

where we used the fact that for j satisfying $\mu_j < \frac{\mu_{j_0}+\mu_{m+1}}{2}$, it follows that

$$\Delta_{j_0,j} = \mu_{j_0} - \mu_j \geq \frac{\mu_{j_0} - \mu_{m+1}}{2} = \frac{\Delta_{j_0,m+1}}{2}.$$

Then, using the same argument from lines (25)-(28), we have that

$$\mathbb{E}X_r \log(X_r) \leq cY_r \log(Y_r + \Delta_{j_0,m+1}^{-2} \log(|A_r| r \frac{\log(\Delta_{j_1,m+1}^{-2})}{\delta})]$$

Thus, putting it together,

$$\mathbb{E}[\mathbf{1}\{\mathbf{1}\{F_r\}\tau] \leq c[2^{r-1}(r-1) + Y_r \log(Y_r + \Delta_{j_0,m+1}^{-2} \log(|A_r| r \frac{\log(\Delta_{j_0,m+1}^{-2})}{\delta})]$$

□

Proof of Theorem 7. Let $j_0 \in [m]$ minimize the optimization problem in line (29). Let r_0 such that be the smallest integer such that

$$\min(40\frac{n}{j_0}, n) \leq 2^{r_0} \leq 80\frac{n}{j_0}.$$

For the sake of brevity, we write $F_{r_0,i}$ instead of $F_{r_0,i}^{(j_0)}$. We bound $\mathbb{P}((F_{r_0,1} \cap F_{r_0,2} \cap F_{r_0,3})^c)$. By a union bound and the law of total probability,

$$\begin{aligned}
\mathbb{P}((F_{r_0,1} \cap F_{r_0,2} \cap F_{r_0,3})^c) &\leq \mathbb{P}(F_{r_0,1}^c \cap F_{r_0,3}^c) + \mathbb{P}(F_{r_0,2}^c) \\
&\leq \mathbb{P}(F_{r_0,1}^c) + \mathbb{P}(F_{r_0,3}^c | F_{r,1}) + \mathbb{P}(F_{r_0,2}^c) \\
&\leq 2\delta + \mathbb{P}(F_{r_0,1}^c) \\
&\leq 2\delta + \exp(-5) \\
&\leq \frac{1}{16}
\end{aligned}$$

The rest of the proof proceeds as the proof of Theorem 5 starting at step 2. □

E.4 Proof of FWER-FWPD

Finally, we present a Theorem for the FWER-FWPD version of Algorithm 2. Although it is possible to use the ideas from the other upper bound proofs to establish a result that depends on the distribution of the arms in $\mathcal{H}_1$, for simplicity our upper bound is in terms of $\Delta = \min_{i \in \mathcal{H}_1} \mu_i - \mu_0$ and $m := |\{i : \mu_i > \mu_0\}$.

Theorem 8. *Let $\delta \in (0, \frac{1}{600})$. Let $k \in [|\mathcal{H}_1|]$. Define*

$$\begin{aligned}
\widetilde{V}_k &:= (\frac{n}{m}k - k)\Delta^{-2} \log(\max(k, \log\log(\frac{n}{m}k\frac{1}{\delta})) \log(\Delta^{-2}) \log(\frac{n}{m}k)/\delta) \\
&\quad + k \log(\max(\frac{n}{m}k - (1 - 2\delta(1+4\delta))k, \log\log(\frac{n}{m}k\frac{1}{\delta})) \log(\Delta^{-2}) \log(\frac{n}{m}k)/\delta)] \\
&\lesssim (\frac{n}{m}k - k)\Delta^{-2} \log(k/\delta) + k \log(\frac{\frac{n}{m}k - (1 - 2\delta(1+4\delta))k}{\delta})
\end{aligned}$$

Furthermore, define

$$\lambda_k = \min(t \in \mathbb{N} : |\mathcal{R}_t \cap \mathcal{H}_1| \geq k).$$

Then, Algorithm 2 has the property that $\mathbb{P}(\exists t \in \mathbb{N} : \mathcal{R}_t \cap \mathcal{H}_0 \neq \emptyset) \leq 10\delta$ and

$$\mathbb{E}[\lambda_k] \leq c \log(\widetilde{V}_k)\widetilde{V}_k.$$

Lemma 7. *Let $\delta \in (0, .01)$. Let $k \in [|\mathcal{H}_1|]$. Let $r \in \mathbb{N}$ such that $2^r \geq k$. Define*

$$\lambda_r = \min(t \in \mathbb{N} : |\mathcal{R}_t \cap A_r \cap \mathcal{H}_1| \geq k).$$

Define

$$\begin{aligned} V_r \coloneqq (2^r - \min(|\mathcal{H}_1|, \frac{|\mathcal{H}_1|}{n}2^r))\Delta^{-2}\log(\max(\min(|\mathcal{H}_1|, \frac{|\mathcal{H}_1|}{n}2^r), \log\log(r2^r/\delta))\log(\Delta^{-2})r/\delta) \\ + \min(|\mathcal{H}_1|, \frac{|\mathcal{H}_1|}{n}2^r)\log(\max(2^r - (1-2\delta(1+4\delta))\min(|\mathcal{H}_1|, \frac{|\mathcal{H}_1|}{n}2^r), \log\log(\frac{r2^r}{\delta}))\log(\Delta^{-2})r/\delta)] \end{aligned}$$

Then with probability at least $1 - 6\delta - 2\exp(-2^{r-3}) - \mathbb{P}(|A_r \cap \mathcal{H}_1| < k)$,

$$\lambda_r \leq c(2^{r-1}(r-1) + \log(V_r)V_r).$$

Proof. **Step 1: Definitions and events.** Recall R_t is the bracket chosen at time t and define

$$T = |\{t \in \mathbb{N} : A_r \cap \mathcal{H}_1 \not\subset \mathcal{R}_t \text{ and } R_t = r\}|,$$

i.e., the number of rounds that the algorithm works on the rth bracket and $A_r \cap \mathcal{H}_1 \not\subset \mathcal{R}_t$. Define the events

$$\begin{aligned} \Sigma_{r,1} &= \{|A_r \cap \mathcal{H}_1| \geq k\} \\ \Sigma_{r,2} &= \{|A_r \cap \mathcal{H}_1| \leq \min(|\mathcal{H}_1|, \tfrac{|\mathcal{H}_1|}{n}2^{r+1}))\} \\ \Sigma_{r,3} &= \{|A_r \cap \mathcal{H}_1| \geq \min(|\mathcal{H}_1|, \tfrac{|\mathcal{H}_1|}{n}2^{r-1}))\} \end{aligned}$$

If $2^{r+1} \geq n$, then $|A_r \cap \mathcal{H}_1| \leq |\mathcal{H}_1|$ implies $\mathbb{P}(\Sigma_{r,2}^c) = 0$. Therefore, suppose $2^{r+1} < n$. Then, by multiplicative Chernoff for hypergeometric random variables,

$$\mathbb{P}(\Sigma_{r,2}^c) = \mathbb{P}(|A_r \cap \mathcal{H}_1| > \frac{|\mathcal{H}_1|}{n}2^{r+1}) \leq \exp(-\frac{|\mathcal{H}_1|}{n}2^{r-2}) \leq \exp(-2^{r-2})$$

Similarly, if $2^r \geq n$, then $|A_r| = n$ and $\mathbb{P}(\Sigma_{r,2}^c) = 0$. Therefore, suppose $2^r < n$.

$$\mathbb{P}(\Sigma_{r,3}^c) = \mathbb{P}(|A_r \cap \mathcal{H}_1| < \frac{|\mathcal{H}_1|}{n}2^{r-1})) \leq \exp(-\frac{|\mathcal{H}_1|}{n}2^{r-3}) \leq \exp(-2^{r-3})$$

Since the algorithm essentially runs the FWER-FWDP version of the algorithm from Jamieson and Jain (2018) on each bracket r with confidence δ/r^2, we can apply Theorem 4 of Jamieson and Jain (2018) directly to obtain that there exists an event $\Sigma_{r,4}$, which only depends on the samples of the arms in bracket r, such that $\mathbb{P}(\Sigma_{r,4}^c) \leq 6\delta$ and on $\Sigma_{r,4}$

$$\begin{aligned} T \leq c[(|A_r| - |A_r \cap \mathcal{H}_1|)\Delta^{-2}\log(\max(|A_r \cap \mathcal{H}_1|, \log\log(|A_r|/\delta_r))\log(\Delta^{-2})/\delta_r) \\ + |A_r \cap \mathcal{H}_1|\Delta^{-2}\log(\max(|A_r| - (1-2\delta_r(1+4\delta_r)|A_r \cap \mathcal{H}_1|, \log\log(\frac{|A_r|}{\delta_r}))\log(\Delta^{-2})/\delta_r)]. \end{aligned}$$

This roughly says

$$T \lesssim (|A_r| - |A_r \cap \mathcal{H}_1|)\Delta^{-2}\log(|A_r \cap \mathcal{H}_1|/\delta) + |A_r \cap \mathcal{H}_1|\Delta^{-2}\log((|A_r| - |A_r \cap \mathcal{H}_1|)/\delta).$$

Step 2: Bounding λ_r. In what follows, assume $\Sigma_{r,1} \cap \Sigma_{r,2} \cap \Sigma_{r,3} \cap \Sigma_{r,4}$ occurs, which happens with probability at least

$$1 - 6\delta - 2\exp(-2^{r-3}) - \mathbb{P}(\Sigma_{r,1}^c).$$

By the same argument given in lines (12) and (13), event $\Sigma_{r,1}$ implies that

$$\lambda_r \leq c(2^{r-1}(r-1) + \log(T)T).$$

Furthermore, using $\Sigma_{r,2} \cap \Sigma_{r,3} \cap \Sigma_{r,4}$,

$$\begin{aligned}
T \leq & c[(|A_r| - |A_r \cap \mathcal{H}_1|)\Delta^{-2} \log(\max(|A_r \cap \mathcal{H}_1|, \log\log(|A_r|/\delta_r)) \log(\Delta^{-2})/\delta_r) \\
& + |A_r \cap \mathcal{H}_1|\Delta^{-2} \log(\max(|A_r| - (1 - 2\delta_r(1 + 4\delta_r))|A_r \cap \mathcal{H}_1|, \log\log(\frac{|A_r|}{\delta_r})) \log(\Delta^{-2})/\delta_r)] \\
\leq & c'[(|A_r| - |A_r \cap \mathcal{H}_1|)\Delta^{-2} \log(\max(|A_r \cap \mathcal{H}_1|, \log\log(r|A_r|/\delta)) \log(\Delta^{-2})r/\delta) \\
& + |A_r \cap \mathcal{H}_1|\Delta^{-2} \log(\max(|A_r| - (1 - 2\delta(1 + 4\delta))|A_r \cap \mathcal{H}_1|, \log\log(\frac{r|A_r|}{\delta})) \log(\Delta^{-2})r/\delta)] \\
\leq & c''[(2^r - \min(|\mathcal{H}_1|, \frac{|\mathcal{H}_1|}{n}2^r))\Delta^{-2} \log(\max(\min(|\mathcal{H}_1|, \frac{|\mathcal{H}_1|}{n}2^r), \log\log(r2^r/\delta)) \log(\Delta^{-2})r/\delta) \\
& + \min(|\mathcal{H}_1|, \frac{|\mathcal{H}_1|}{n}2^r)\Delta^{-2} \log(\max(2^r - (1 - 2\delta(1 + 4\delta)) \min(|\mathcal{H}_1|, \frac{|\mathcal{H}_1|}{n}2^r), \log\log(\frac{r2^r}{\delta})) \log(\Delta^{-2})r/\delta)]
\end{aligned}$$

□

Proof of Theorem 8. We note that the algorithm essentially runs the FWER-FWDP version of the algorithm from Jamieson and Jain (2018) on each bracket r with confidence δ/r^2. Therefore, by Theorem 4 from Jamieson and Jain (2018),

$$\mathbb{P}(\exists t \in \mathbb{N} : A_r \cap \mathcal{R}_t \cap \mathcal{H}_0 \neq \emptyset) \leq 6\frac{\delta}{r^2}$$

Thus,

$$\begin{aligned}
\mathbb{P}(\exists t \in \mathbb{N} : \mathcal{R}_t \cap \mathcal{H}_0 \neq \emptyset) &\leq \mathbb{P}(\exists t \in \mathbb{N}, r \in \mathbb{N} : A_r \cap \mathcal{R}_t \cap \mathcal{H}_0 \neq \emptyset) \\
&\leq \sum_{r \in \mathbb{N}} \mathbb{P}(\exists t \in \mathbb{N} : A_r \cap \mathcal{R}_t \cap \mathcal{H}_0 \neq \emptyset) \\
&\leq \sum_{r \in \mathbb{N}} 6\frac{\delta}{r^2} \\
&\leq 10\delta.
\end{aligned}$$

Let $r_0 \in \mathbb{N}$ be the smallest integer such that $r_0 \geq 6$ and

$$\min(40\frac{n}{m}k, n) \leq 2^{r_0} \leq 80\frac{n}{m}k.$$

If $2^{r_0} \geq n$, then $\mathbb{P}(|A_r \cap \mathcal{H}_1| < k) = 0$. Otherwise,by multiplicative Chernoff for hypergeometric random variables,

$$\mathbb{P}(|A_r \cap \mathcal{H}_1| < k) \leq \exp(-5).$$

In the interest of brevity, define $\Sigma_r = \Sigma_{r,1} \cap \Sigma_{r,2} \cap \Sigma_{r,3} \cap \Sigma_{r,4}$. Observe that $\{\Sigma_r\}_{r \in \mathbb{N}}$ are mutually independent. Further, using $\delta \in (0, \frac{1}{600})$, for all brackets $r \geq r_0$, the events occur which happens with probability at least

$$\mathbb{P}(\Sigma_r^c) \leq 6\delta + 2\exp(-2^{r-3}) + \mathbb{P}(\Sigma_{r,1}^c) \leq \frac{1}{16}$$

The rest of the proof proceeds as in Step 2 of the proof of Theorem 5. □

F ϵ-Good Arm Identification: Favorable Verifiable and Unverifiable Sample Complexity

One practical concern about the SimplePAC setting is that it is not clear when to stop the algorithm. To address this concern we propose Algorithm 3, which combines Algorithm 1 and LUCB from Kalyanakrishnan et al. (2012) to achieve the best of both worlds of PAC and SimplePAC. Let LUCB(ϵ) denote the LUCB algorithm that terminates once it finds an ϵ-good arm. Let $\beta(t, \delta)$ denote the confidence bound used in Kalyanakrishnan et al. (2012); although, it is possible to tighten these confidence bounds, for the sake of simplicity and brevity we

Algorithm 3 To Verify or not to Verify: ϵ-Good Arm Identification

Input: $\epsilon > 0$
for $t = 1, 2, \ldots$ **do**
  Pull arm according to sampling rule given by the ϵ-good arm identification version of Algorithm 1
  Pull arm according to sampling rule given by LUCB(ϵ)
  Let O_t be the arm returned by the ϵ-good arm identification version of Algorithm 1
  if LUCB(ϵ) terminates **then**
    Let $\widehat{j}$ denote the arm returned by LUCB(ϵ)
    $r_0 = \text{argmax}_{r\in\mathbb{N}} \widehat{\mu}_{O_t,r,T_{O_t,r}(t)} - U(T_{O_t,r}(t), \frac{\delta}{|A_r|r^2})$
    if $\widehat{\mu}_{O_t,r_0,T_{i,r}(t)} - U(T_{O_t,r_0}(t), \frac{\delta}{|A_{r_0}|r^2}) \geq \widehat{\mu}_{\widehat{j},T_{\widehat{j}}(t)} - \beta(T_{\widehat{j}}(t), \delta)$ **then**
      Set $\widehat{i}_t = O_t$
    else
      Set $\widehat{i}_t = \widehat{j}$
    Output $\widehat{i}_t$ and terminate.
  else
    Set $\widehat{i}_t = O_t$
    Output $\widehat{i}_t$

use theirs so that we can appeal to their sample complexity results. Algorithm 3 takes a desired tolerance $\epsilon > 0$ as input, runs LUCB(ϵ) and the ϵ-good arm identification version of Algorithm 1 in parallel without sharing samples between the algorithms,[3] and outputs an arm $\widehat{i}_t$ at every iteration. This arm $\widehat{i}_t$ is the arm O_t suggested by Algorithm 1 for every iteration until the termination condition of LUCB(ϵ) obtains at which point algorithm 3 decides whether to output O_t or the arm suggested by LUCB(ϵ). Let $\widehat{\mu}_{i,t}$ denote the empirical mean at time t of arm i based on the samples collected by LUCB(ϵ) and $T_{i,t}$ denote the number of pulls of arm i at time t by LUCB(ϵ).

Theorem 9. *Let ρ be a problem instance and let $\delta \in (0, 1/40)$ and $\epsilon_1, \epsilon_2 > 0$. Let $(\mathcal{F}_t)_{t\in\mathbb{N}}$ be the filtration generated by running Algorithm 3 with input ϵ_1 on ρ. There is a stopping time τ_{simple} wrt $(\mathcal{F}_t)_{t\in\mathbb{N}}$ such that*

$$\mathbb{E}[\tau_{simple}] \lesssim \min_{\gamma\in(0,\epsilon_2)} U_{\epsilon_2}(\gamma) \log(U_{\epsilon_2}(\gamma) + \Delta^{-2}_{m,\epsilon_2,\gamma}) \tag{32}$$

and $\mathbb{P}(\exists s \geq \tau_{simple} : \mu_{\widehat{i}_s} \leq \mu_1 - \epsilon_2) \leq 2\delta$. Furthermore, there exists a stopping time τ_{PAC} wrt $(\mathcal{F}_t)_{t\in\mathbb{N}}$ such that

$$\mathbb{E}[\tau_{PAC}] \lesssim H^{\epsilon/2} \log(\frac{H^{\epsilon/2}}{\delta}) \tag{33}$$

where $H^\gamma = \sum_{i\in[n]} \max(\mu_1 - \mu_i, \gamma)^{-2}$ and at time τ_{PAC} the Algorithm 3 terminates and returns an arm $\widehat{i}_{\tau_{PAC}}$ such that $\mathbb{P}(\mu_{\widehat{i}_{\tau_{PAC}}} \leq \mu_1 - \min(\epsilon_1, \epsilon_2)) \leq 3\delta$.

To interpret the Theorem 9, suppose that $\epsilon_1 > \epsilon_2 > 0$ are such that $\mathbb{E}[\tau_{simple}] \leq \mathbb{E}[\tau_{PAC}]$. Then, Theorem 9 says that Algorithm 3 with input ϵ_1 starts outputting an ϵ_2-good arm in nearly optimal time *and* certifies that it is an ϵ_1-good arm in nearly optimal optimal. Thus, Algorithm 3 achieves the best of both worlds.

Proof of Theorem 9. Theorem 6 of Kalyanakrishnan et al. (2012) implies that there exists a stopping time τ_{PAC} wrt $(\mathcal{F}_t)_{t\in\mathbb{N}}$ such that at time τ_{PAC} the Algorithm 3 terminates and (33) holds. Theorem 2 implies the existence of stopping time τ_{simple} wrt $(\mathcal{F}_t)_{t\in\mathbb{N}}$ such that (32) holds and $\mathbb{P}(\exists s \geq \tau_{simple} : \mu_{O_s} \leq \mu_1 - \epsilon_2) \leq 2\delta$.

It remains to show that when the Algorithm 3 terminates at $t = \tau_{PAC}$, $\mathbb{P}(\mu_{\widehat{i}_{\tau_{PAC}}} \leq \mu_1 - \min(\epsilon_1, \epsilon_2)) \leq 3\delta$. Define the event

$$F = \{\forall t \in \mathbb{N}, s \in \mathbb{N}, \text{ and } i \in A_s : |\widehat{\mu}_{i,s,t} - \mu_i| \leq U(t, \frac{\delta}{|A_s|s^2})\}.$$

By a union bound, F occurs with probability at least $1 - 2\delta$. By the argument in Step 1 of the proof of Lemma 6, on F, for all $t \geq \tau_{simple}$

$$\max_{r\in\mathbb{N}} \widehat{\mu}_{O_t,r,T_{O_t,r}(t)} - U(T_{O_t,r}(t), \frac{\delta}{|A_r|r^2}) > \max_{i:\mu_i\leq\mu_1-\epsilon_2} \mu_i.$$

[3] Samples should be shared in practice.

Next, define the event

$$E = \{\forall t \in \mathbb{N} \text{ and } \forall i \in [n] : |\widehat{\mu}_{i,t} - \mu_i| \leq \beta(t, \delta)\}$$

By Theorem 1 of Kalyanakrishnan et al. (2012), $\mathbb{P}(E) \geq 1 - \delta$ and on E,

$$\widehat{\mu}_{\widehat{j},T_{\widehat{j}}(\tau_{PAC})} - \beta(T_{\widehat{j}}(\tau_{PAC}), \delta) > \mu_1 - \epsilon_1$$

Suppose F and E occur, which by a union bound occur with probability at least $1 - 3\delta$. Either $\widehat{i}_{\tau_{PAC}} = \widehat{j}$ or $\widehat{i}_{\tau_{PAC}} = O_{\tau_{PAC}}$. Suppose $\widehat{i}_{\tau_{PAC}} = \widehat{j}$. Then,

$$\begin{aligned}
\mu_{\widehat{i}_{\tau_{PAC}}} &= \mu_{\widehat{j}} \\
&\geq \widehat{\mu}_{\widehat{j},T_{\widehat{j}}(\tau_{PAC})} - \beta(T_{\widehat{j}}(\tau_{PAC}), \delta) \\
&> \max_{r \in \mathbb{N}} \widehat{\mu}_{O_t,r,T_{O_t,r}(t)} - U(T_{O_t,r}(t), \frac{\delta}{|A_r| r^2}) \\
&\geq \max_{i:\mu_i \leq \mu_1 - \epsilon_2} \mu_i,
\end{aligned}$$

which implies that $\mu_{\widehat{i}_{\tau_{PAC}}} \geq \mu_1 - \min(\epsilon_1, \epsilon_2)$. A similar argument proves the case $\widehat{i}_{\tau_{PAC}} = O_{\tau_{PAC}}$.

□

G Experiment Details

We used two publicly available datasets to base our simulated experiments on.

G.1 ϵ-good arm identification

For the ϵ-good arm identification experiment, we used the *New Yorker Magazine* Caption Contest data available at `https://github.com/nextml/caption-contest-data`. Specifically, we used contest 641 conducted the first week of December of 2018. Briefly, visitors to the site `nextml.org/captioncontest` are shown a fixed image and one of n captions that they rate as either `Unfunny`, `Somewhat funny`, or `Funny`. When they make their selection, the image stays the same but one of n other captions are shown (uniformly at random for this contest). Contest 641 has $n = 9061$ arms and each one was shown about 155 times. For the ith caption we define $\widehat{\mu}_{i,T_i}$ as the proportion of times `Somewhat funny` of `Funny` was clicked relative to the total number of times it was rated denoted T_i. These empirical means $\widehat{\mu}_{i,T_i}$ were treated as ground truth so that in our experiments a pull of the ith arm was an iid draw from a Bernoulli distribution with mean $\widehat{\mu}_{i,T_i}$. Figure 4 shows the histogram $\widehat{\mu}_{i,T_i}$ and T_i for all $n = 9061$ arms.

To measure $\tau_{U,\epsilon}$, we run LUCB and BUCB for 3 million rounds; 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 $t \in [\tau_{U,\epsilon}, 3 \cdot 10^6]$. To measure $\tau_{V,\epsilon}$ for LUCB, we run LUCB for 20 million rounds and report its guarantee on the returned arm at every t.

G.2 Identifying arms above a threshold

This dataset is from Hao et al. (2008). The study was interested in identifying genes in Drosophila that inhibit virus replication. Essentially, for each individual gene $i \in [n]$ for $n = 13071$ the researchers used RNAi to "knock-out" the gene from a population of cells, infected the cells with a virus connected to a florescing tag, and then measured the amount of florescence after a period of time. The idea is that if a lot of florescence was measured when the ith gene was knocked out, that means that gene was very influential for inhibiting virus replication because more virus was present. A control or baseline amount of florescence μ_0 (and its variance) was established by infecting cells without any genes knocked out. Using these controls, each measurement (pull) from the ith gene (arm) is reported as a Z-score such that under the null (gene i has no impact on virus replication) an observation is normally distributed with mean $\mu_i = \mu_0$ with variance 1. We make the simplifying assumption that if the gene did have non-negligible influence so that $\mu_i > 0$, then the variance was still equal to 1.

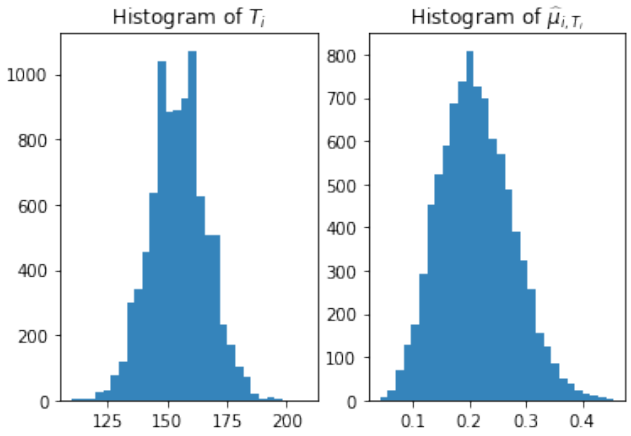

Figure 4: Empirical means and counts from the *New Yorker Magazine* caption contest 641. There were $n = 9061$ arms.

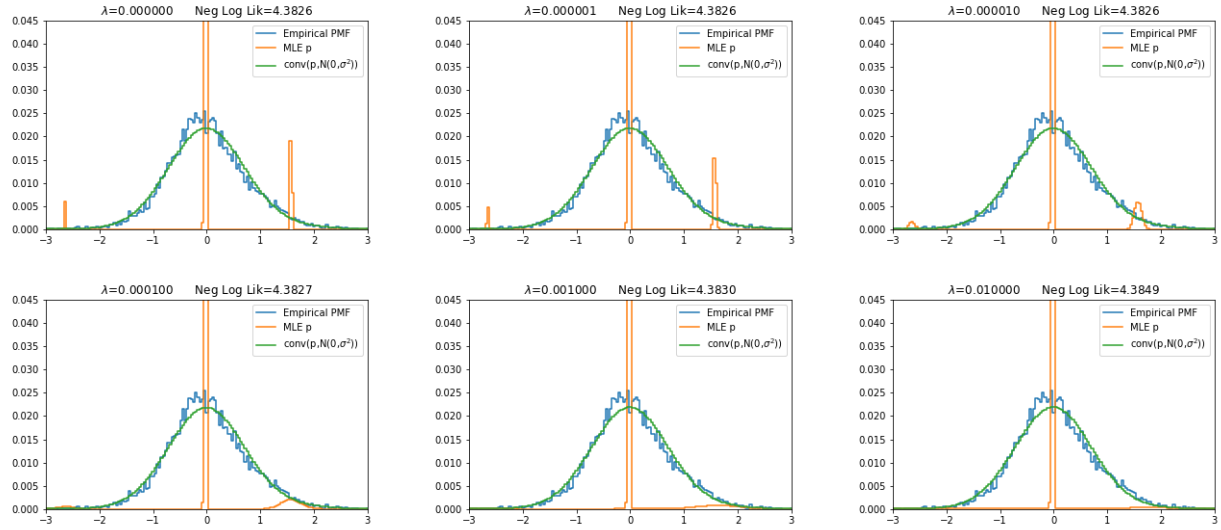

Figure 5: Drosophila data.

As described in Hao et al. (2008), the researchers measured each of the $n = 13071$ genes twice and eliminated all but the 1000 most extreme observations, and then measured each of these 1000 genes 12 times. Finally, they reported the 100 genes that were statistically significant of these 1000 genes measured 12 times. To generate the data for our experiments, we average just the two initial measurements from all $n = 13071$ measurements. Two averaged Z-scores of the ith gene, denoted $\widehat{\mu}_i$, have a variance of $1/2$ which more or less buries any signal in noise. If we adopt the model $\widehat{\mu}_i \sim \mathcal{N}(\mu_i, 1/2)$ then we can perform a a maximum likelihood estimate (MLE) of the original distribution of underlying $\{\mu_i\}_{i=1}^n$ using a fine grid on $[-4, 4]$, the range of the observations. The normalized histogram of $\{\widehat{\mu}_i\}_i$ as well as the MLE of the $\{\mu_i\}_i$ are shown in the first panel of Figure 5. Reassuringly, there is a spike with mass of about .97 at 0 indicating that the vast majority of genes have no influence on inhibiting virus proliferation. The majority of the remaining mass lies in a spike around 1. To encourage the distribution of the means not at 0 to have a bit more shape, we use a small amount of entropic regularization without increasing negative log likelihood too much. For our experiments we used $\lambda = 1e^{-4}$.

G.3 Algorithm Details

We use $\delta = 0.05$ for all of the algorithms. For the implementation of our algorithms, we chose the starting bracket to have size 2^6. We share samples between the brackets and stop opening brackets after a bracket of size

n is opened.

For the ϵ-good arm identification experiment, we change the sampling rule slightly to mirror LUCB in the following sense: at each round, the algorithm pulls both a maximizer of the empirical mean and a maximizer of the upper confidence on the mean. The theory on BUCB directly applies since once of these arms must be the same arm that BUCB would pull. We also use a heuristic where we remove a bracket if its maximum lower confidence bound is less than the maximum lower confidence bound of a larger bracket.

For the experiment concerning the dataset of Hao et al. (2008) we used the FDR-TPR versions of our algorithm and the algorithm of Jamieson and Jain (2018). Following the advice of Jamieson and Jain (2018), we use the Benjamini-Hochberg procedure developed for multi-armed bandits at level δ instead of $O(\delta/\log(1/\delta))$. We used the following two heuristics for our algorithm. First, we give each bracket a point if it pulls an accepted arm more than any of the other brackets. Then, we remove a bracket if its score is less than the score of a larger bracket. Second, we estimate the number of pulls required for each bracket to accept 5 additional arms and choose the bracket with lowest estimate 90% of the time and otherwise cycle through the brackets.[4] We calculate this estimate as follows. For each bracket, we take the 5 arms with the largest empirical means and estimate the remaining number of times that they need to be pulled by

$$\widehat{\mu}_{i,T_i(t)}^{-2} \log[\text{ size of the bracket } \cdot \text{ number of total brackets to open } /\delta] - T_i(t).$$

For the other arms, we estimate the number of times that they need to be pulled before accepting 5 arms with the largest empirical means in the following way. Let λ denote the value of the fifth smallest mean multiplied by a factor of 2, which estimates roughly the value of its upper confidence bound at the point at which it is accepted. Then, the estimate is

$$(\lambda - \widehat{\mu}_{i,T_i(t)})^{-2} \log[\text{ number of total brackets to open } /\delta] - T_i(t).$$

We note that while the above heuristics for removing brackets break the sample complexity guarantees of the algorithms because they may remove a good bracket, the algorithms are still correct in the sense that the confidence bounds hold with high probability. We ran each experiment for 100 trials. We also plot 95% confidence intervals.

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.

[4] We note that we only get slightly worse performance if we pick the bracket with lowest estimate 50% of the time. See Figure 6.

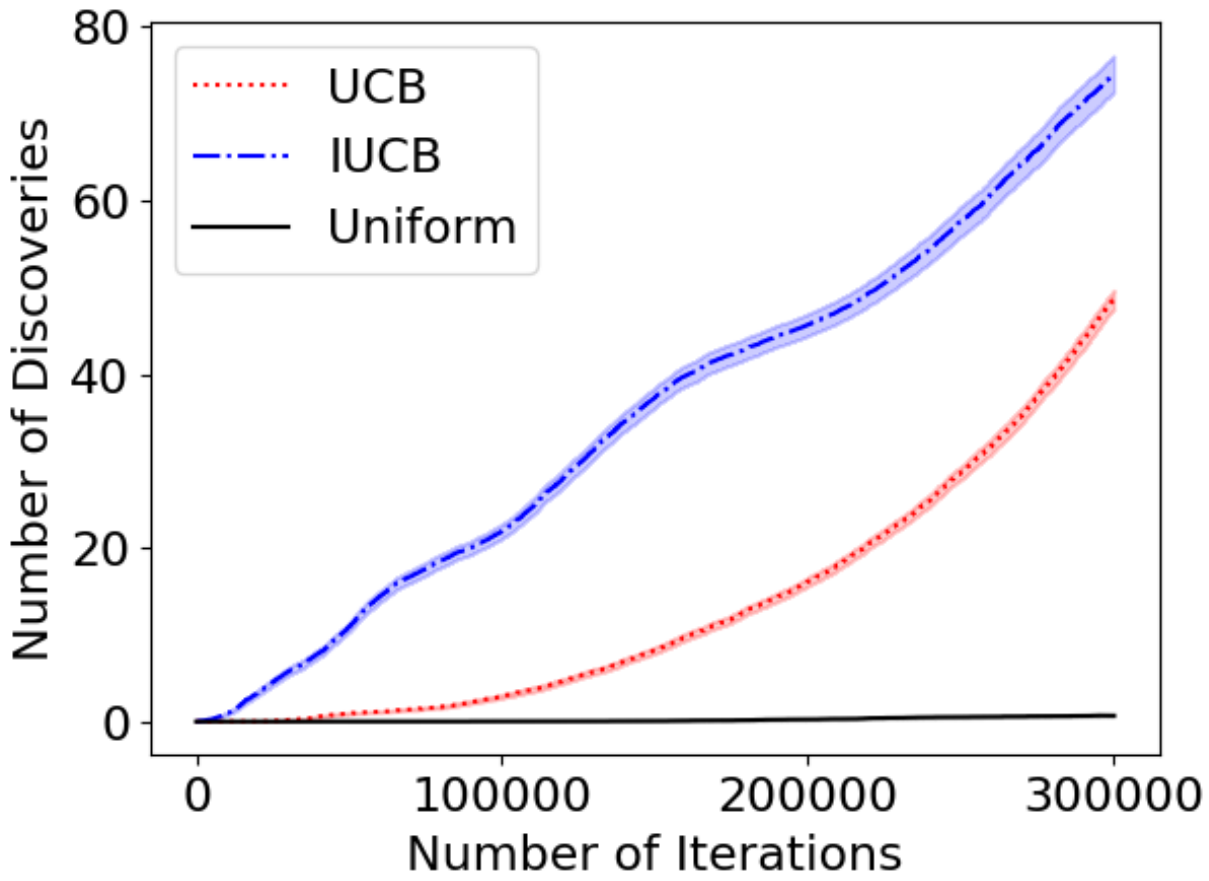

Figure 6: Identifying means above a threshold: pick estimated best bracket 50% of the time.

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.