Revisiting Simple Regret: Fast Rates for Returning a Good Arm

Yao Zhao [1] **Connor James Stephens** [2] **Csaba Szepesvári** [2] **Kwang-Sung Jun** [1]

Abstract

Simple regret is a natural and parameter-free performance criterion for pure exploration in multi-armed bandits. However, possibly due to the challenge of characterizing it, in the theory community, it has been largely overlooked in favor of studying the probability of selecting the best, or an ϵ-good arm. In this paper, we make significant progress on minimizing simple regret in both data-rich ($T \geq n$) and data-poor regime ($T \leq n$) where n is the number of arms, and T is the number of samples. At its heart is our improved instance-dependent analysis of the well-known Sequential Halving (SH) algorithm, where we bound the probability of returning an arm whose mean reward is not within ϵ from the best (i.e., not ϵ-good) for *any* choice of $\epsilon > 0$, despite ϵ not being input to SH. Our bound not only leads to an optimal worst-case simple regret bound of $\sqrt{n/T}$ up to logarithmic factors but also essentially matches the instance-dependent lower bound for returning an ϵ-good arm reported by Katz-Samuels and Jamieson (2020). For the more challenging data-poor regime, we propose Bracketing SH (BSH) that enjoys the same improvement even without sampling each arm at least once. Our empirical study shows that BSH outperforms existing methods on real-world tasks.

1. Introduction

We consider the pure exploration problem in multi-armed bandits. In this problem, given n arms, the learner sequentially chooses an arm $I_t \in [n] := \{1, \dots, n\}$ at time t to observe a reward $r_t = \mu_{I_t} + \eta_t$ where μ_i is the mean reward of arm i and η_t is stochastic σ^2-sub-Gaussian noise. Without loss of generality, we assume that $1 \geq \mu_1 \geq \cdots \geq \mu_n \geq 0$. After T time steps, the learner is required to output an arm J_T that is estimated to have the largest mean reward. The budget T may or may not be given to the algorithm as input. Such a problem was formalized by Even-Dar et al. (2006) and Bubeck et al. (2009), but similar problems were considered much earlier (Bechhofer, 1958; Chernoff, 1959). Recently, pure exploration has found many applications such as efficient crowdsourcing for cartoon caption contests (Tanczos et al., 2017), hyperparameter optimization (Li et al., 2018; 2020), and improving the time complexity of clustering algorithms (Baharav and Tse, 2019). Hereafter, we take $\sigma^2 = 1$ for ease of exposition, but all our results can easily be extended to σ^2-sub-Gaussian noise.

Among many performance criteria, *simple regret* proposed by Bubeck et al. (2009) is a natural measure of the quality of the estimated best arm J_T:

$$\mathsf{SReg}_T := \mathbb{E}\left[\mu_1 - \mu_{J_T}\right] . \tag{1}$$

Simple regret is an *unverifiable* performance measure, meaning that the algorithm need not verify its performance to a prescribed level. This is in stark contrast to the fixed confidence setting (Even-Dar et al., 2006) such as (ϵ,δ)-PAC that requires the algorithm to *verify* the quality of the estimated best arm to the prescribed target accuracy ϵ and confidence level δ. While certain applications do need such verification, there are many applications where the sampling budget is too limited to show meaningful guarantees, such as cartoon caption contests (Jain et al., 2020) and biological experiments (Jun et al., 2016). In these cases, algorithms with verifiable guarantees may not be meaningful. Furthermore, existing algorithms with verifiable guarantees are vacuous in the data-poor regime of $T \leq n$ as it requires each arm to be pulled at least once (Katz-Samuels and Jamieson, 2020, Section C.1). The fixed budget setting considers an unverifiable measure of $\mathbb{P}\big(\mu_{J_T} \neq \mu_1\big)$ (Audibert et al., 2010), but this quantity inevitably depends on the smallest gap $\mu_1 - \mu_2$ (Carpentier and Locatelli, 2016), which is usually not meaningful until the sample size is very large.

Despite being attractive, the analysis of simple regret has been elusive. Existing studies focus on either achieving minimax simple regret bound (Lattimore et al., 2016) or characterization of how simple regret is different from the cumulative regret (Bubeck et al., 2009). It is not known whether simple algorithms like Sequential Halving (SH) (Karnin

[1]Department of Computer Science, University of Arizona, Tucson, USA [2]Department of Computing Science, University of Alberta, Edmonton, Canada. Correspondence to: Kwang-Sung Jun <kjun@cs.arizona.edu>.

Proceedings of the 40th International Conference on Machine Learning, Honolulu, Hawaii, USA. PMLR 202, 2023. Copyright 2023 by the author(s).

Table 1. Sample complexity comparison with ME (Even-Dar et al., 2006), $\mathcal{P}_3$ (Chaudhuri and Kalyanakrishnan, 2019), and BUCB (Katz-Samuels and Jamieson, 2020). The lower bound for Polynomial(α) instances (Theorem 3, Lemma 3.1) is a contribution of this paper which improves on Katz-Samuels and Jamieson (2020, Theorem 1) to include the correct dependence on δ. The lower bound for EqualGap(m) was first shown by Chaudhuri and Kalyanakrishnan (2017), which works for the fixed budget setting as well. We omit constant and logarithmic factors in n and m. Methods with asterisk consider the fixed confidence setting, and those with dagger have an in expectation bound rather than the high probability one. Anytime algorithms do not require the knowledge of the budget T as input, and parameter-free ones do not require any of $\{\alpha, m, \epsilon, \delta\}$ as input.

	Polynomial(α) with $\alpha > \frac{1}{2}$	EqualGap(m) with $m \le n/4$	Data-poor	Anytime	Parameter-free
ME*	$\frac{n}{\epsilon^2}\log\left(\frac{1}{\delta}\right)$	$\frac{n}{\epsilon^2}\log\left(\frac{1}{\delta}\right)$	✗	✗	✗
$\mathcal{P}_3^*$	$\frac{1}{\epsilon^{2+(1/\alpha)}}\log\left(\frac{1}{\delta}\right) \vee \frac{1}{\epsilon^2}\log^2\left(\frac{1}{\delta}\right)$	$\frac{1}{\epsilon^2}\left(\frac{n}{m}\log\left(\frac{1}{\delta}\right) + \log^2\left(\frac{1}{\delta}\right)\right)$	✓	✗	✗
BUCB†	$\frac{n}{\epsilon^{2-(1/\alpha)}}\log\left(\frac{1}{\delta}\right)$	$\frac{n}{m\epsilon^2}\log\left(\frac{1}{\delta}\right)$	✓	✓	✗
DSH (ours)	$\frac{1}{\epsilon^2}\log\left(\frac{1}{\delta}\right) \vee \frac{n}{\epsilon^{1/\alpha}}$	$\frac{n}{m\epsilon^2}\log\left(\frac{1}{\delta}\right) \vee \frac{n}{\epsilon^2}$	✗	✓	✓
BSH (ours)	$\frac{1}{\epsilon^2}\log\left(\frac{1}{\delta}\right)$	$\frac{n}{m\epsilon^2}\log\left(\frac{1}{\delta}\right)$	✓	✓	✓
Lower bound	$\frac{1}{\epsilon^2}\log\left(\frac{1}{\delta}\right)$	$\frac{n}{m\epsilon^2}\log\left(\frac{1}{\delta}\right)$			

et al., 2013) enjoy the optimal simple regret or not.

In this paper, we revisit simple regret and make two main contributions. First, we provide a novel and tight analysis of Sequential Halving (SH), one of the state-of-the-art algorithms for pure exploration. We analyze a strong performance criterion that we call *ϵ-error probability*:

$$\mathbb{P}\big(\mu_1 - \mu_{J_T} > \epsilon\big) \tag{2}$$

where ϵ is not given to the algorithm as input. This is precisely characterizing the distribution of μ_{J_T}, which was mentioned as an interesting direction of research in Lattimore and Szepesvári (2020, Section 33.4 Note 9). If an algorithm $\mathcal{A}$ does not take ϵ as input yet has a guarantee for any ϵ, we say $\mathcal{A}$ enjoys a *uniform ϵ-error probability bound*. The ϵ-error probability is stronger than simple regret since

$$\mathsf{SReg}_T = \mathbb{E}[\mu_1 - \mu_{J_T}] = \int_0^\infty \mathbb{P}(\mu_1 - \mu_{J_T} > \epsilon)\,\mathrm{d}\epsilon\,. \tag{3}$$

Thus, a uniform ϵ-error probability bound implies a simple regret bound.

We call an arm i to be ϵ-good if $\mu_i \ge \mu_1 - \epsilon$. Let $\Delta_i := \mu_1 - \mu_i$ be the gap of arm i and $g(\epsilon) := \left|\left\{i \in [n] \mid \mu_i \ge \mu_1 - \epsilon\right\}\right|$ be the number of ϵ-good arms. We prove that SH (Karnin et al., 2013) has the following bound for any $\epsilon > 0$ (Theorem 1):

$$\mathbb{P}\big(\mu_{J_T} < \mu_1 - \epsilon\big) \le \exp\left(-\tilde{\Theta}\left(\frac{T}{H_2(\epsilon)}\right)\right), \tag{4}$$

where $\tilde{\Theta}(\cdot)$ hides logarithmic factors and

$$H_2(\epsilon) := \frac{1}{g(\epsilon/2)} \max_{i \ge g(\epsilon)+1} \frac{i}{\Delta_i^2}\,.$$

is the sample complexity function. By setting $\epsilon = 0$, the complexity becomes $H_2(0) = \max_{i\ge 2} i\Delta_i^{-2} =: H_2$, so our bound is never worse than the original bound of SH presented in Karnin et al. (2013): $\mathbb{P}(\mu_{J_T} \ne \mu_1) \le \exp(-\tilde{\Theta}(\frac{T}{H_2}))$. Furthermore, our bound essentially matches the lower bound presented in Katz-Samuels and Jamieson (2020). We also derive a minimax-style bound on the ϵ-error probability, which can be easily integrated out over ϵ to obtain $\mathsf{SReg}_T = \tilde{O}(\sqrt{n/T})$, and further show that a variant of SH results in removing logarithmic factors from the bound and achieve $\mathsf{SReg}_T = O(\sqrt{n/T})$. Finally, we derive an anytime version of SH that we call Doubling SH (DSH), which enjoys the same guarantees without requiring T as input. We present details of these claims in Section 3.

As the second contribution, we propose an improved pure exploration algorithm for the data-poor regime. This is motivated by the fact that the guarantee (4) becomes vacuous in the data-poor regime ($T \le \max_{i\ge g(\epsilon)+1} i\Delta_i^{-2}$ to be precise) as it does not even allow us to pull every arm once. Inspired by Katz-Samuels and Jamieson (2020), we propose Bracketing SH (BSH) that progressively subsamples a larger set of arms (i.e., brackets) for which we invoke SH to find a good arm. Our analysis shows that BSH enjoys a bound that is essentially the same as (4) but works even for $T \le \max_{i\ge g(\epsilon)+1} i\Delta_i^{-2}$. Compared to BUCB (Katz-Samuels and Jamieson, 2020), the state-of-the-art algorithm for the data-poor regime, BSH achieves three improvements: (i) BSH is parameter-free whereas BUCB requires the target error rate δ as input, (ii) BSH has a high probability sample complexity bound as opposed to the expected sample complexity bound of BUCB, (iii) BSH shows a much stronger bound than BUCB. We provide details on BSH in Section 4.

We compare our bounds with existing bounds in various setups in Table 1. For ease of comparison, we consider special instances called Polynomial(α) and EqualGap(m). Polynomial(α) is an instance with gap structure $\Delta_i = \left(\frac{i}{n}\right)^\alpha$ for $2 \le i \le n$, where $\alpha > 0.5$ (Jamieson et al., 2013), and

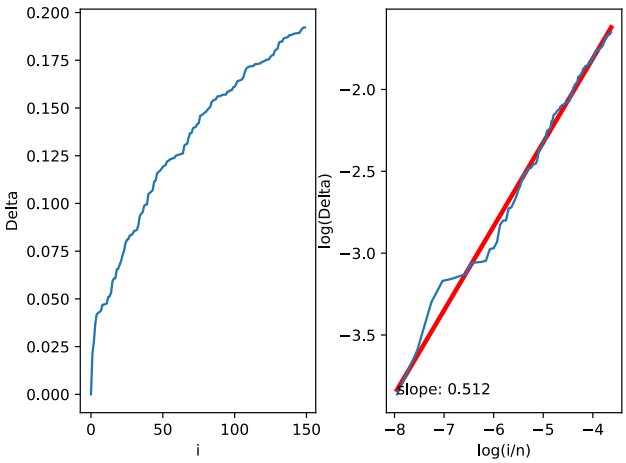

Figure 1. The polynomial fitting for the best 150 arms of the New Yorker Cartoon Caption contest 788.

EqualGap(m) is an instance that has $m-1$ arms with the gap $\epsilon/4$ and $n-m$ arms with the gap $\frac{5}{4}\epsilon$. Both α and m have the role of adjusting the number of good arms. Our bound for BSH matches the lower bounds for the first time to our knowledge, thereby establishing the optimality for these two instances. Note that many real-world instances follow Polynomial(α); e.g., see Figure 1. One interesting aspect of Polynomial(α) is that, although α affects the fraction of good arms, the optimal rate $1/\epsilon^2$ is independent of α. This is because the hardness of returning an ϵ-good arm depends not just on how many good arms we have but also on how far they are from bad arms as we explain in Section 3.

Our result also resolves the open problem of having a factor of $\ln^2(1/\delta)$ rather than $\ln(1/\delta)$ when obtaining accelerated sample complexities under the presence of multiple good arms. Such an issue appeared repeatedly in the literature; e.g., Chaudhuri and Kalyanakrishnan (2019), Katz-Samuels and Jamieson (2020, Section E), and Aziz et al. (2018, Section 5). They all rely on the subsampling device or its variants. Our improved bound can be attributed to the fact that an accelerated sample complexity can be obtained even *without the subsampling device* – we are the first to achieve such a result, to our knowledge.

We evaluate our algorithm on the cartoon caption contest dataset in Section 5 and conclude our paper with future work in Section 6. Due to space constraints, we have the related work section in the appendix, but directly-related works are discussed throughout the paper. All the proofs are deferred to the appendix.

2. Preliminaries

We focus on the fixed budget and anytime setting in pure exploration. In the *fixed budget* setting (Audibert and Bubeck, 2010), the learner is given a budget $T \geq 1$ as input and is required to return an estimated best arm $J_T \in [n]$ at the end of T-th time step. In the *anytime* setting, there is no prescribed T, and the learner is required to output $J_t \in [n]$ for every time step $t \geq 1$. In these settings, unlike the fixed confidence setting, the learner is not required to verify the quality of the returned arm J_T. We collectively call these settings the *unverifiable* setting and refer to the fixed confidence setting as the *verifiable* setting. Note that for any verifiable setting, one can consider an unverifiable version. For example, the verifiable (m, ϵ)-good arm identification problem (Chaudhuri and Kalyanakrishnan, 2017) requires that the learner, given (m, ϵ, δ), identify an arm i that is (m, ϵ)-good (i.e., $\mu_i \geq \mu_m - \epsilon$) w.p. at least $1-\delta$. One can consider the unverifiable (m, ϵ)-good arm identification problem where the goal is to minimize $\mathbb{P}(\mu_{J_T} \geq \mu_m - \epsilon)$ after T time steps.

Sample Complexity. When comparing bounds for simple regret or ϵ-error probability, it is often easier to consider the sample complexity, which measures the number of samples required to achieve the target performance level. For example, the sample complexity for the ϵ-error probability is the smallest time step τ such that $\forall t \geq \tau, \mathbb{P}(\mu_1 - \mu_t > \epsilon) \leq \delta$.

Notations. Throughout the paper, "const" is a universal and positive constant, which may have different values for different expressions. Both $\tilde{\Theta}(\cdot)$ and $\tilde{\mathcal{O}}(\cdot)$ omits logarithmic factors on m, n, but not $1/\delta$.

3. Improved Analyses of Sequential Halving

Sequential Halving (SH) (Karnin et al., 2013) is a simple and parameter-free algorithm that has had an outsize impact on pure exploration since its introduction. Many recent practical algorithms for pure exploration (Aziz et al., 2022; Li et al., 2018; 2020; De Heide et al., 2021) make use of this algorithm, either with slight modifications or as a subroutine. These works have demonstrated that the algorithm's practical and theoretical significance is greater than its original presentation. Even beyond the basic bandit setting, SH also has been used for planning in reinforcement learning (Danihelka et al., 2022), where SH is observed to be easier to tune compared to some UCB style algorithms as it does not have problem-dependent parameters.

We present SH in Algorithm 3.1. SH consists of $\lceil \log_2(n) \rceil$ stages. In each stage ℓ, the algorithm performs a uniform sampling over the surviving arms S_ℓ followed by eliminating the bottom half of S_ℓ w.r.t. empirical means and sets $S_{\ell+1}$ as the resulting arm set.

SH is known to achieve a nearly optimal instance-dependent bound on the $(\epsilon = 0)$-error probability, which we detail in Table 2 in the appendix. However, it is unknown if SH is a good algorithm for returning an ϵ-good arm. For $\epsilon > 0$,

though in the fixed confidence setting, Median Elimination (ME) achieves the sample complexity of $\frac{n}{\epsilon^2}\log\left(\frac{1}{\delta}\right)$ (Even-Dar et al., 2006). Does SH achieve a similar sample complexity for ϵ-error probability? Despite the similarity between SH and ME, the answer is not clear because the sampling scheme of SH is different from ME. In light of the ubiquity of this algorithm and recent empirical success, a stronger characterization of the performance of SH is of great interest to the community.

In this section, via a powerful analysis of the ϵ-error probability incurred by SH, we show that this algorithm achieves an optimal bound for the unverifiable (m,ϵ)-good arm identification problem that further implies a minimax optimal simple regret bound (Lattimore and Szepesvári, 2020, Exercise 33.1). We also provide a near-optimal instance-dependent uniform ϵ-error probability bound for the first time.

Our main theorem shows the following instance-dependent uniform ϵ-error probability bound for SH.

Theorem 1. *For any $\epsilon \in (0,1)$, the ϵ-error probability of SH satisfies*

$$\mathbb{P}\left(\mu_{J_T} < \mu_1 - \epsilon\right) \le \exp\left(-\tilde{\Theta}\left(\frac{T}{H_2(\epsilon)}\right)\right),$$

for $T \ge \tilde{\Theta}\left(\max_{i\ge g(\epsilon)+1}\frac{i}{\Delta_i^2}\right)$.

The proof is in Appendix B.5. Note the requirement of $T \ge \tilde{\Theta}\left(\max_{i\ge g(\epsilon)+1}\frac{i}{\Delta_i^2}\right)$ is not a weakness of our result. One can easily see that if $\Delta_i = \epsilon, \forall i \ge 2$, the standard result of SH (Karnin et al., 2013) becomes vacuous for $T = o(\frac{n}{\epsilon^2})$. We stress that, unlike all existing work that we are aware of, Theorem 1 bounds the ϵ-error probability for an algorithm that does not require ϵ as input. That is to say, ϵ is a parameter for analysis only. This upper bound of ϵ-error probability implies a sample complexity of $\tilde{O}(H_2(\epsilon)\ln(1/\delta))$.

Polynomial gap instances. The class of polynomial gap instances was introduced by Jamieson et al. (2013), who referred to them as 'α-parameterized' instances. The authors studied these instances in the best-arm identification setting, showing that for $\alpha > 1/2$, the sample complexity $H = \sum_{i=2}^n \Delta_i^{-2}$ scales like $\Omega\left(n^{2\alpha}\right)$, with the increase in sample complexity corresponding to the growing number of close-to-optimal arms as α increases. In one of the highlight results of this work, we show that when $\epsilon > 0$ is not pathologically small, SH returns ϵ-good arms on these instances with an optimal sample complexity that is independent of the number of arms n.

Corollary 1.1. *For Polynomial(α) with $\alpha > \frac{1}{2}$, we have*

$$H_2(\epsilon) = \frac{1}{g\left(\frac{\epsilon}{2}\right)}\max_{i\ge g(\epsilon)+1}\frac{i}{\Delta_i^2} < \frac{4}{\epsilon^2}.$$

The proof is in Appendix B.6. While this result initially seems surprising, the lack of dependence on n stems from the ratio between the number of 'well-separated' ϵ-good arms and the sample complexity of filtering out ϵ-bad arms (i.e., non-ϵ-good arms) on these instances. As n increases, due to the polynomial gap structure, the number of well-separated ϵ-good arms (e.g., the collection of $\epsilon/2$-good arms) increases linearly with n, which balances the linear growth in n of the sum over the ϵ-bad arm gaps $\sum_{i=g(\epsilon)+1}^n \Delta_i^{-2}$. This second term is effectively the sample complexity of discarding all ϵ-bad arms. Since we only require one of the good arms to be returned, we get to divide $\sum_{i=g(\epsilon)+1}^n \Delta_i^{-2}$ by $g(\epsilon/2)$.

Lower bounds. We compare our upper bound in Theorem 1 with lower bounds. For brevity, we present an informal version here and defer the full version to Appendix C.3 along with a slightly stronger version.

Theorem 2. *(Informal version of Katz-Samuels and Jamieson (2020, Theorem 1)) For any algorithm, there exists an instance where the expected unverifiable sample complexity for returning an ϵ-good arm with probability greater than 15/16 satisfies*

$$\Omega\left(H^{low}(\epsilon) := \frac{1}{g(\epsilon)}\sum_{i=g(\epsilon)+1}^{n}\left(\frac{1}{\Delta_i^2}\right) - \frac{1}{\Delta_{g(\epsilon)+1}^2}\right).$$

Corollary 2.1 (Comparison with previous lower bound)**.** *In the case that $g(\epsilon)$ and $g\left(\frac{\epsilon}{2}\right)$ have the same order ($\frac{g\left(\frac{\epsilon}{2}\right)}{g(\epsilon)}$ is irrelevant to ϵ), the sample complexity measure $H_2(\epsilon)$ satisfies*

$$H^{low}(\epsilon) \lesssim H_2(\epsilon) \lesssim H^{low}(\epsilon) + \frac{2}{\Delta_{g(\epsilon)+1}^2}$$

where $\lesssim$ hides constants and logarithms of n and $1/\delta$.

The sample complexity implied by Theorem 2 for returning an ϵ-good arm with probability $1-\delta$, with $\delta \in (0, 1/16)$, is $\Omega\left(H^{\text{low}}(\epsilon)\right)$, which does not have the factor $\ln(1/\delta)$. We contribute a new lower bound that suggests a sample complexity of $\Omega\left(\tilde{H}(\epsilon)\log(1/\delta)\right)$, with a complexity term $\tilde{H}(\epsilon)$ that we describe next, and with which we show that Theorem 1 is essentially tight on polynomial gap instances with $\alpha > 1/2$.

For an n-armed bandit instance ν, let

$$\tilde{H}(\nu,\epsilon) = \sum_{i=2}^{n/g(\epsilon)} \Delta_{i\cdot g(\epsilon)}^{-2},$$

Algorithm 1 Sequential Halving (SH)

Input: budget: T, arms: $[n]$
Initialize: $S_1 = [n]$

for $\ell = 1, \dots, \lceil \log_2 n \rceil$ **do**
  Sample each arm $i \in S_\ell$ for T_ℓ times where

$$T_\ell = \left\lfloor \frac{T}{|S_\ell| \lceil \log_2 n \rceil} \right\rfloor$$

  Let $S_{\ell+1}$ be the set of $\lceil S_\ell / 2 \rceil$ arms in S_ℓ with the largest empirical rewards.
Set J_T as the only arm in $S_{\lceil \log_2 n \rceil}$.
Output: J_T

where for simplicity we assume that n is an integer multiple of $g(\epsilon)$.

Theorem 3. *Fix $n > 2$, $\epsilon > 0$ and an n-armed unit-variance Gaussian instance ν. With $a := \tilde{H}(\nu, \epsilon)$ we define $\mathcal{B}_n(a, \epsilon)$, the collection of n-armed bandit instances with unit-variance gaussian arms for which $\tilde{H}(\nu', \epsilon) \le a$ for each $\nu' \in \mathcal{B}_n(a, \epsilon)$. Then for any algorithm π and sample budget $T \in \mathbb{N}$,*

$$\sup_{\nu' \in \mathcal{B}_n(a,\epsilon)} \mathbb{P}_{\nu',\pi}\big(\mu_{J_T} < \mu_1 - \epsilon\big) \ge \max\left\{\frac{1}{2}, \frac{1}{36}\exp\big(-24T/a\big)\right\}.$$

The complexity term $\tilde{H}(\nu, \epsilon)$ plays the same role in our result as that of $H(\nu) = \sum_{i=2}^n \Delta_i^{-2}$ in Carpentier and Locatelli (2016, Theorem 1). The proof of Theorem 3 can be found in Appendix B.8.

Lemma 3.1. *Consider a Polynomial(α) instances with fixed $\alpha > 1/2$. Then for any $\epsilon \le 1/2$, $\tilde{H} \ge \frac{c(\alpha)}{\epsilon^2}$ where $c(\alpha)$ is a problem-dependent constant which does not depend on ϵ.*

The proof is in Appendix B.9. Combining Corollary 1.1 and Lemma 3.1, we see that Theorem 1 is optimal up to logarithmic factors in the sample complexity for polynomial gap instances with $\alpha > 1/2$.

3.1. Unverifiable (m, ϵ)-Good Arm Identification

We show the error probability of SH for unverifiably identifying an (m, ϵ)-good arm (defined in Section 2) and discuss its implications.

Theorem 4. *For any $m \le n$ and any $\epsilon \in (0, 1)$, the error probability of SH for identifying an (m, ϵ)-good arm satisfies*

$$\mathbb{P}\big(\mu_{J_T} < \mu_m - \epsilon\big) \le \log_2 n \cdot \exp\left(-\text{const} \cdot m \cdot \left(\frac{\epsilon^2 T}{4n \log_2^2(2m) \log_2 n} - \ln(4e)\right)\right).$$

Thus, there exist an absolute constant $c_1 > 0$ s.t. if $T \ge c_1 \frac{(\ln(4e) + \ln\ln n) n \log_2^2(2m) \log_2 n}{\epsilon^2} = \tilde{\Theta}\big(\frac{n}{\epsilon^2}\big)$, we have

$$\mathbb{P}\big(\mu_{J_T} < \mu_m - \epsilon\big) \le \exp\left(-\tilde{\Theta}\left(m \frac{\epsilon^2 T}{n}\right)\right).$$

The proof is in Appendix B.3.

Remark 1. *Note that our bound implies a verifiable sample complexity bound for the (m, ϵ)-good arm identification setup as well because, given (m, ϵ, δ), one can work out the sufficient samples size T' to control the RHS of Theorem 4 to be at most δ and then run SH with T'.*

Implications for $m = 1$. The worst-case upper bound of Theorem 4 for $m = 1$ corresponds to the sample complexity of $\tilde{\mathcal{O}}\big(\frac{n}{\epsilon^2} \log \frac{1}{\delta}\big)$. This is the same as the classic lower bound result of the (ϵ, δ)-PAC problem (Mannor and Tsitsiklis, 2004) up to logarithmic factors. While numerous algorithms achieve a matching upper bound, including Even-Dar et al. (2006) and Hassidim et al. (2020), our result in Theorem 1 indicates a tighter bound when there are many good arms.

We now apply (3) to convert a uniform ϵ-error probability bound into a simple regret bound.

Corollary 4.1. *SH satisfies $\mathsf{SReg}_T \le \tilde{\mathcal{O}}\Big(\sqrt{\frac{n}{T}}\Big)$.*

The proof is in Appendix B.2. Note that our bound is minimax optimal up to logarithmic factors (Lattimore and Szepesvári, 2020, Exercise 33.1). To our knowledge, the only algorithm that we are aware of that achieves the minimax simple regret is MOSS (Audibert et al., 2009; Lattimore et al., 2016). However, as MOSS is designed to minimize cumulative regret, it cannot enjoy low instance-dependent bounds for pure exploration tasks (Bubeck et al., 2011). While uniform sampling can also achieve a near-optimal simple regret, one can show that uniform sampling can be arbitrarily worse than SH w.r.t. the $(\epsilon = 0)$-error probability as its sample assignments over the arms are inherently non-adaptive.

Remark 2. *In Appendix B.12, we show that one can adjust T_ℓ in Algorithm 3.1 to achieve the minimax optimality up to a constant factor (i.e., no logarithmic factors), but it is not clear if this variant achieves a similar instance-dependent sample complexity as Algorithm 3.1.*

Implications for $m > 1$. We now show that SH enjoys the optimal worst-case guarantee for the unverifiable (m, ϵ)-good arm identification problem up to logarithmic factors.

Algorithm 2 Doubling Sequential Halving (DSH)

Input arms: $[n]$
Initialize $T_1 = \lceil n \log_2 n \rceil$, $\hat{j} = j$ for some arbitrarily chosen $j \in [n]$, $\hat{\mu}^* = -\infty$. Define the time blocks $\mathcal{T}_1 = \{1, \dots, T_1\}$ and $\mathcal{T}_k = \{T_1(2^{k-1}-1)+1, \dots, T_1(2^k-1)\}$, $\forall k \geq 2$, which satisfies $|\mathcal{T}_k| = T_1 2^{k-1}$. $k = 1$. An instance of SH denoted by $\mathcal{S}_1$ with budget $|\mathcal{T}_1|$.
for $t = 1, 2, \dots$ **do**
  Pull arm I_t according to the recommendation from $\mathcal{S}_k$
  Receive reward R_t and send it to $\mathcal{S}_k$
  if $t = \max \mathcal{T}_k$ **then**
    Set $(\hat{j}, \hat{\mu}^*)$ as the output arm and its empirical mean from the last stage of $\mathcal{S}_k$
    $k \leftarrow k + 1$
    Initialize a new instance of SH denoted by $\mathcal{S}_k$ with budget $|\mathcal{T}_k|$
  Output: $J_t = \hat{j}$ and $M_t = \hat{\mu}^*$.

Our result is especially attractive since the target number of good arms m is not an input to SH.

Our upper bound of Theorem 4 corresponds to a sample complexity of $\tilde{\mathcal{O}}\left(\frac{n}{m\epsilon^2}\log\left(\frac{1}{\delta}\right)\right)$, which matches the lower bound in Chaudhuri and Kalyanakrishnan (2017), thereby closing the gap between the upper and lower bounds from previous work; the algorithms therein have suboptimal upper bounds that scale with $\log^2\left(\frac{1}{\delta}\right)$. Furthermore, these algorithms all require (m, ϵ, δ) as input, which is natural for obtaining verifiable sample complexities but becomes a limitation in other settings. In contrast, Theorem 4 allows (m, ϵ) to be chosen freely to measure the algorithm's performance since the parameters for measuring the performance are no longer necessary for executing the algorithm. Therefore, one can rewrite Theorem 4 as follows:

$$\mathbb{P}\left(\mu_{J_T} < \mu_1 - \epsilon\right) \leq \min_{\{(m', \epsilon'): \Delta_{m'} + \epsilon' \leq \epsilon\}} \exp\left(-\tilde{\Theta}\left(m' \frac{\epsilon'^2 T}{n}\right)\right).$$

3.2. Anytime version of SH

To support anytime simple regret minimization, we combine SH with the doubling trick (Jun and Nowak, 2016), which we call Doubling SH (DSH). Specifically, DSH repeatedly runs SH. The k-th SH receives a doubled budget compared with the $(k-1)$-th SH. Before it finishes k-th SH, it always returns the output of $(k-1)$-th SH. The pseudocode can be found in Algorithm 3.1. The following theorem shows that DSH enjoys the same guarantee as SH up to a constant factor.

Theorem 5. *Let $\epsilon \in (0, 1)$. A single run of DSH satisfies the following ϵ-error probability*

$$\mathbb{P}\left(\mu_{J_T} < \mu_1 - \epsilon\right) \leq \exp\left(-\tilde{\Theta}\left(\frac{T}{H_2(\epsilon)}\right)\right),$$

simultaneously for all $T \geq \tilde{\Theta}\left(\max_{i \geq g(\epsilon)+1} \frac{i}{\Delta_i^2}\right)$

This is a direct consequence of the fact that at time t, the latest finished SH was run with a budget of at least $t/4$. The full proof is in Appendix C.2.

4. Simple Regret in the Data-Poor Regime

Despite the optimality of SH shown in the last section, it requires at least $\tilde{\Theta}\left(\max_{i \geq g(\epsilon)+1} \frac{i}{\Delta_i^2}\right)$ samples, which in the worst case could be $\tilde{\Theta}\left(n/\epsilon^2\right)$. This requirement is unacceptable when the number of arms n is large (or even infinite) while the fraction of the ϵ-good arms is kept constant. This setup is commonly referred to as the data-poor regime, where one would like to return a good arm even before the algorithm samples every arm once.

To cope with the data-poor regime, we take inspiration from BUCB (Katz-Samuels and Jamieson, 2020) and propose Bracketing SH (BSH) that enjoys a similar guarantee as DSH but enjoys non-vacuous sample complexity in the data-poor regime. To understand the bracketing trick, suppose that we uniformly sample a subset of size n/m $(0 < m \leq n)$ from the entire arm set $[n]$. With constant probability, this subset includes at least one of the top m arms. By applying any pure exploration algorithm to this subset, we expect to find the best arm within the subset with the sample complexity that scales with n/m rather than n. Such a subset is called a *bracket*. Note that there is a natural tradeoff here with the bracket size. If the size is large, we are likely to include many good arms, but the sample complexity of identifying a good arm becomes large because we have a lot of arms to pull at least once. On the other hand, if the size is small, we are not likely to include any good arm, although the sample complexity of identifying an arm that is good relative to the bracket is small. As we show later, one can precisely work out such a tradeoff mathematically and find the best bracket size. The challenge is, however, that the best bracket size typically requires knowledge of the number of good arms.

To avoid requiring such knowledge, BSH adopts the bracketing technique of Katz-Samuels and Jamieson (2020) that progressively creates a larger and larger bracket size as the time step t gets larger while invoking a base algorithm for each bracket by cycling through them (i.e., Round Robin). Specifically, BSH uses DSH as the base algorithm. At each time step, BSH takes in the estimated best arms and their empirical means from all the brackets and outputs the one

Algorithm 3 Bracketing SH (BSH)

Input: arms: $[n]$
Initialize: $t = 1$, $B = 0$, $b_1 = 1$. Define $(I(\mathcal{D}), J(\mathcal{D}), M(\mathcal{D}))$ to be the arm to be pulled next, the current estimated best arm, and its empirical mean from an algorithm $\mathcal{D}$, respectively.
for $t = 1, 2, 3 \dots$ **do**
  if $t \geq B2^B$ **then**
    $B \leftarrow B + 1$
    Sample a bracket $A_B = U\big([n], n \vee 2^B\big)$
    Initialize a new instance of DSH, denoted by $\mathcal{D}_B$, with the bracket A_B.
  Pull arm $I_t = I(\mathcal{D}_{b_t})$
  Receive a reward R_t from the environment and send R_t to $\mathcal{D}_{b_t}$
  Output $J_t = J(\mathcal{D}_{\hat{b}})$ where $\hat{b} = \arg\max_{b \in [B]} M(\mathcal{D}_b)$
  $b_{t+1} = \begin{cases} b_t + 1 & \text{if } b_t < B \\ 1 & \text{otherwise} \end{cases}$

with the largest empirical mean. We summarize BSH in Algorithm 4 where the operator $U\big([n], k\big)$ samples k items with replacement from $[n]$ (uniformly at random). If $k \geq n$, $U([n], k)$ returns $[n]$. To define terminology, we say a new bracket is *opened* when a new bracket is sampled. We set the bracket-opening schedule such that the B-th bracket is opened at time step $(B-1) \cdot 2^{B-1}$. The arm pulls between the time step $(B-1) \cdot 2^{B-1}$ and $B \cdot 2^B$ are equally allocated to the opened brackets (total B of them). We say an arm *represents* bracket A_B at time step t if it is the output returned by the DSH on bracket A_B at time step t.

Let us first show the properties of the bracketing technique. The bracketing design ensures the diversity of the size of opened brackets. Specifically, the smallest bracket has 2 arms, and the largest bracket has $\tilde{\Theta}(t)$ arms. Moreover, it also ensures that all the opened brackets have received an (order-wise) equal amount of sampling budget at any time.

We now present the ϵ-error probability bound of BSH.

Theorem 6. *Let $\epsilon \in (0, 1)$. A single run of BSH satisfies the following ϵ-error probability*

$$\mathbb{P}\big(\mu_1 - \mu_{J_t} > \epsilon\big) \leq \exp\left(-\tilde{\Theta}\left(\frac{t}{\max\left\{H_2\big(\frac{\epsilon}{2}\big), \frac{1}{\epsilon^2}\right\}}\right)\right),$$

simultaneously for all $t \geq \tilde{\Theta}\Big(H_2\big(\frac{\epsilon}{2}\big)\Big)$.

The proof is in Appendix C.4. The key difference between Theorem 5 and Theorem 6 is the latter does not require the number of samples to be at least $T \geq \tilde{\Theta}\Big(\max_{i \geq g(\epsilon)+1} \frac{i}{\Delta_i^2}\Big)$.

To show the effectiveness of Theorem 6, we now discuss the implication of our result and compare it with BUCB by Katz-Samuels and Jamieson (2020) using their performance measure called (ϵ, δ)-unverifiable sample complexity.

Definition 1 ((ϵ, δ)-unverifiable sample complexity)**.** *(Katz-Samuels and Jamieson, 2020). For an algorithm π and an instance ρ. Let $\tau_{\epsilon,\delta}$ be a stopping time such that*

$$\mathbb{P}\big(\forall t \geq \tau_{\epsilon,\delta} : \mu_{J_t} > \mu_1 - \epsilon\big) \geq 1 - \delta.$$

Then $\tau_{\epsilon,\delta}$ is called (ϵ, δ)-unverifiable sample complexity of the algorithm with respect to ρ.

The (ϵ, δ)-unverifiable sample complexity indicates how many samples are sufficient for an algorithm to begin to output an ϵ-good arm with high probability. Compared with the verifiable sample complexity, it does not require the algorithm to verify the output is ϵ-good. More discussion can be found in Katz-Samuels and Jamieson (2020). While the verbatim requirement of (ϵ, δ)-unverifiable sample complexity is stronger than the uniform ϵ-error probability bound, we show that the latter implies the former under a mild assumption in Appendix C.8.

To make explicit comparisons, we turn to specific problem instances and show the (ϵ, δ)-unverifiable sample complexity bounds for BUCB (Katz-Samuels and Jamieson, 2020, Theorem 7) and BSH.

Corollary 6.1. *Consider the EqualGap(m) instance. BUCB achieves an expected (ϵ, δ)-unverifiable sample complexity as*

$$\mathbb{E}\big[\tau_{\epsilon,\delta}\big] \leq \tilde{\mathcal{O}}\left(\frac{n}{\epsilon^2 m} \log\left(\frac{1}{\delta}\right)\right).$$

BSH achieves an (ϵ, δ)-unverifiable sample complexity as, with probability $1 - \delta$,

$$\tau_{\epsilon,\delta} \leq \tilde{\mathcal{O}}\left(\frac{n}{\epsilon^2 m} \log\left(\frac{1}{\delta}\right)\right).$$

The proof is in Appendix C.5. In this instance, BUCB and BSH have nearly the same performance guarantees except that BUCB's (ϵ, δ)-unverifiable sample complexity is stated as an expected sample complexity. Interestingly, Katz-Samuels and Jamieson (2020, Section E) remarks that their attempt to derive a high-probability bound resulted in the factor of $\ln^2(1/\delta)$, which we believe is due to the suboptimal bound (i.e., no accelerated rates) of their choice of the base algorithm.

Scaling with n/m instead of n reveals the merit of the (ϵ, δ)-unverifiable sample complexity that was not achieved by the algorithms designed to achieve verifiable sample complexity such as Median Elimination (Even-Dar et al., 2006).

Corollary 6.2. *Consider the Polynomial(α) instance; i.e., $\Delta_i = \left(\frac{i}{n}\right)^\alpha$ with $\alpha > 0.5$. For any $\epsilon \in (0,1)$, let $\hat{\tau}_{\epsilon,\delta}$ be the upper bound of the expected (ϵ,δ)-unverifiable sample complexity reported in Katz-Samuels and Jamieson (2020, Theorem 7) for BUCB. Then, BUCB satisfies*

$$\mathbb{E}\left[\tau_{\epsilon,\delta}\right] \le \hat{\tau}_{\epsilon,\delta} = \tilde{\Theta}\left(\epsilon^{-\frac{2\alpha-1}{\alpha}} n \log\left(\frac{1}{\delta}\right)\right).$$

On the other hand, BSH satisfies, with probability $1-\delta$,

$$\tau_{\epsilon,\delta} \le \tilde{\mathcal{O}}\left(\epsilon^{-2} \log\left(\frac{1}{\delta}\right)\right).$$

The proof is in Appendix C.6. The (ϵ,δ)-unverifiable sample complexity of BSH for the Polynomial(α) instance does not scale with the number of arms n polynomially, unlike BUCB. For large-scale instances, BSH provides a much stronger guarantee than BUCB. Even when n is not too large, as long as $\Delta_2 \le \epsilon$ (which implies $\epsilon^{\frac{1}{\alpha}} \le n$), BSH still has a better sample complexity bound than BUCB.

5. Experiments

We test Bracketing SH on a real-world dataset called The New Yorker Cartoon Caption Contest (Jain et al., 2020). For each cartoon, the editors of The New Yorker collect the evaluation score of n captions from the participants. Specifically, upon arrival of a participant, the algorithm sequentially shows a number of captions and receives the evaluation score of "unfunny", "somewhat funny", or "funny". Following Katz-Samuels and Jamieson (2020), we use the proportion of times the score was "somewhat funny" or "funny" as the ground truth mean reward for each caption. We then set the reward distribution as the Bernoulli. We choose contests 780, 781, and 782, which contain 6509 arms, 5969 arms, and 4389 arms, respectively. As we are especially interested in the data-poor regime, we report the performance of the algorithms up to time step 10,000, which is only about twice larger than the number of arms in our dataset. We ran BSH and BUCB with various δ choices to see the best version of BUCB, which was repeated 500 times with a different random seed. We used a desktop with AMD Ryzen 5 PRO 4650GE CPU and 16GB RAM to conduct the experiment, which took two hours to produce each plot. We summarize the results in Figure 2, where BSH is a clear winner over BUCB. Specifically, in contest 781, at time step $4,000$, Bracketing UCB achieves simple regret about 0.18 for $\delta = 0.2$, while BSH achieves simple regret about 0.11, amounting to 39% improvement. We explain more on the implementation in Appendix D.

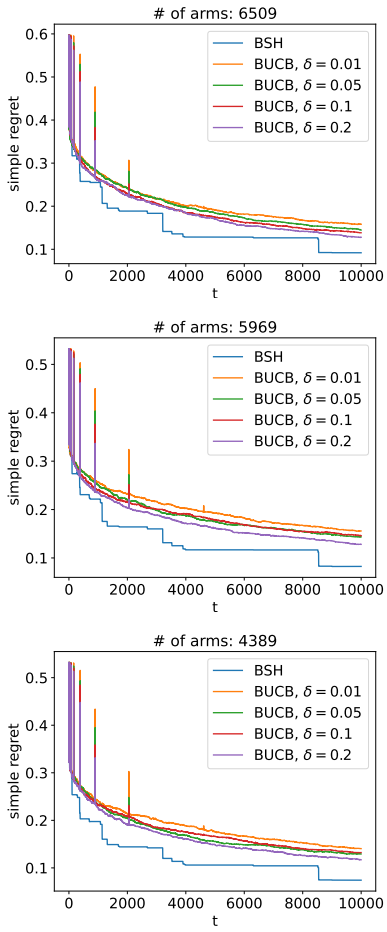

Figure 2. The simple regret comparison between Bracketing UCB and Bracketing SH for the New Yorker Cartoon Caption contest 780, 781, 782.

6. Conclusion

Obtaining a uniform ϵ-error probability bound is precisely a way to characterize the distribution function of Δ_{J_T} induced by a particular algorithm, based on which we believe that a uniform ϵ-error probability bound is a fundamental quantity for any measures for returning an ϵ-good arm.

As such, we believe our paper opens up numerous exciting open problems. First, SH does not achieve an optimal worst-case sample complexity of $\Theta(\frac{n}{\epsilon^2}\ln(1/\delta))$ (due to logarithmic factors). While adjusting its sample allocation scheme does achieve the optimal rate as we report in the appendix, it does not seem to achieve the usual instance-dependent sample complexity of $\tilde{\Theta}(\sum_i(\epsilon \vee \Delta_i)^{-2})$ achieved by many existing algorithms. We wonder if a modified SH or other algorithms can simultaneously be optimal for both sample complexity measures. Second, another practical pure exploration algorithm is track-and-stop by Garivier and Kaufmann (2016) and their variants. They are also

parameter-free if we only take their sampling strategy and not the stopping strategy. It would be interesting to investigate if track-and-stop achieves similar near-optimal guarantees as DSH.

Acknowledgements

We thank Jaehyeok Shin for providing help in the initial development of the improved bound for SH. Kwang-Sung Jun and Yao Zhao were supported by the Eighteenth Mile TRIF Funding from Research, Innovation & Impact at University of Arizona. Csaba Szepesvári gratefully acknowledges funding from the Canada CIFAR AI Chairs Program, through Amii and NSERC.

References

Jean-Yves Audibert and Sébastien Bubeck. Best arm identification in multi-armed bandits. In *Proceedings of the Conference on Learning Theory (COLT)*, pages 41–66, 2010.

Jean-Yves Audibert, Sébastien Bubeck, et al. Minimax policies for adversarial and stochastic bandits. In *Annual Conference on Computational Learning Theory (COLT)*, volume 7, pages 1–122, 2009.

Jean-Yves Audibert, Sébastien Bubeck, and Rémi Munos. Best Arm Identification in Multi-Armed Bandits. In *Proceedings of the Conference on Learning Theory (COLT)*, 2010.

Maryam Aziz, Jesse Anderton, Emilie Kaufmann, and Javed Aslam. Pure exploration in infinitely-armed bandit models with fixed-confidence. In *Algorithmic Learning Theory*, pages 3–24. PMLR, 2018.

Maryam Aziz, Jesse Anderton, Kevin Jamieson, Alice Wang, Hugues Bouchard, and Javed Aslam. Identifying New Podcasts with High General Appeal Using a Pure Exploration Infinitely-Armed Bandit Strategy. In *Proceedings of the 16th ACM Conference on Recommender Systems*, pages 134–144, 2022.

Tavor Baharav and David Tse. Ultra fast medoid identification via correlated sequential halving. *Advances in Neural Information Processing Systems (NeurIPS)*, 32, 2019.

Robert E Bechhofer. A Sequential Multiple-Decision Procedure for Selecting the Best One of Several Normal Populations with a Common Unknown Variance, and Its Use with Various Experimental Designs. *Biometrics*, 14 (3):408–429, 1958.

Sébastien Bubeck, Rémi Munos, and Gilles Stoltz. Pure exploration in multi-armed bandits problems. In *International conference on Algorithmic learning theory*, pages 23–37. Springer, 2009.

Sébastien Bubeck, Rémi Munos, and Gilles Stoltz. Pure exploration in finitely-armed and continuous-armed bandits. *Theoretical Computer Science*, 412(19): 1832–1852, apr 2011. URL https://hal-hec.archives-ouvertes.fr/hal-00609550.

Alexandra Carpentier and Andrea Locatelli. Tight (lower) bounds for the fixed budget best arm identification bandit problem. *Journal of Machine Learning Research*, 49 (June):590–604, 2016.

Arghya Roy Chaudhuri and Shivaram Kalyanakrishnan. PAC identification of a bandit arm relative to a reward quantile. In *Thirty-First AAAI Conference on Artificial Intelligence*, 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 (ICML)*, pages 991–1000. PMLR, 2019.

Herman Chernoff. Sequential design of experiments. *The Annals of Mathematical Statistics*, 30(3):755–770, 1959.

Ivo Danihelka, Arthur Guez, Julian Schrittwieser, and David Silver. Policy improvement by planning with gumbel. In *International Conference on Learning Representations*, 2022.

Eyal Even-Dar, Shie Mannor, Yishay Mansour, and Sridhar Mahadevan. Action elimination and stopping conditions for the multi-armed bandit and reinforcement learning problems. *Journal of machine learning research*, 7(6), 2006.

Aurélien Garivier and Emilie Kaufmann. Optimal best arm identification with fixed confidence. In *Proceedings of the Conference on Learning Theory (COLT)*, pages 998–1027, 2016.

Avinatan Hassidim, Ron Kupfer, and Yaron Singer. An optimal elimination algorithm for learning a best arm. *Advances in Neural Information Processing Systems (NeurIPS)*, 33:10788–10798, 2020.

Rianne de Heide, James Cheshire, Pierre Ménard, and Alexandra Carpentier. Bandits with many optimal arms. *Advances in Neural Information Processing Systems (NeurIPS)*, 34:22457–22469, 2021.

Lalit Jain, Kevin Jamieson, Robert Mankoff, Robert Nowak, and Scott Sievert. The new yorker cartoon caption contest dataset. 2020. URL https://nextml.github.io/caption-contest-data/.

Kevin Jamieson, Matthew Malloy, Robert Nowak, and Sebastien Bubeck. On finding the largest mean among many. *arXiv preprint arXiv:1306.3917*, 2013.

K.-S. Jun and R. Nowak. Anytime exploration for multi-armed bandits using confidence information. In *Proceedings of the International Conference on Machine Learning (ICML)*, volume 48, pages 974–982, 2016. ISBN 9781510829008.

K.-S. Jun, K. Jamieson, R. Nowak, and X. Zhu. Top arm identification in multi-armed bandits with batch arm pulls. In *Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2016.

Zohar Karnin, Tomer Koren, and Oren Somekh. Almost optimal exploration in multi-armed bandits. In *International Conference on Machine Learning (ICML)*, pages 1238–1246. PMLR, 2013.

Julian Katz-Samuels and Kevin Jamieson. The true sample complexity of identifying good arms. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 108, pages 1781–1791, 2020.

Finnian Lattimore, Tor Lattimore, and Mark D Reid. Causal bandits: Learning good interventions via causal inference. *Advances in Neural Information Processing Systems (NeurIPS)*, 2016.

Tor Lattimore and Csaba Szepesvári. *Bandit algorithms*. Cambridge University Press, 2020.

Liam Li, Kevin Jamieson, Afshin Rostamizadeh, Ekaterina Gonina, Jonathan Ben-tzur, Moritz Hardt, Benjamin Recht, and Ameet Talwalkar. A System for Massively Parallel Hyperparameter Tuning. In *Proceedings of Machine Learning and Systems (MLSys)*, pages 230–246. 2020.

Lisha Li, Kevin Jamieson, Giulia DeSalvo, Afshin Rostamizadeh, and Ameet Talwalkar. Hyperband: A Novel Bandit-Based Approach to Hyperparameter Optimization. *Journal of Machine Learning Research*, 18(185):1–52, 2018.

Shie Mannor and John N Tsitsiklis. The sample complexity of exploration in the multi-armed bandit problem. *Journal of Machine Learning Research*, 5(Jun):623–648, 2004.

Ervin Tanczos, Robert Nowak, and Bob Mankoff. A KL-LUCB algorithm for Large-Scale Crowdsourcing. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 5894–5903. 2017.

Appendix

A. Related work

Pure exploration. Pure exploration has a broad position in several closely related research directions, including multi-armed bandits (Audibert and Bubeck, 2010; Karnin et al., 2013; Carpentier and Locatelli, 2016; Aziz et al., 2018; Karnin et al., 2013; Jamieson et al., 2014; Garivier and Kaufmann, 2016; Chen et al., 2017b; Simchowitz et al., 2017; Hassidim et al., 2020) and reinforcement learning (Even-Dar et al., 2006; Azar et al., 2017; Kaufmann et al., 2021). Even more, the Monte-Carlo tree search with tree depth 1 is also a pure exploration problem (Kocsis and Szepesvári, 2006). Even if we only consider multi-armed bandits, there are various types of problems that have been formulated. Our investigation focuses on the standard K-armed bandits model. But we note here pure exploration has also been studied in structured bandits like kernel bandits (Camilleri et al., 2021) and linear bandits (Soare et al., 2014; Zhu et al., 2022).

Started from Even-Dar et al. (2006); Mannor and Tsitsiklis (2004), pure exploration in K-armed bandits is studied with the celebrated (ϵ, δ)-PAC framework. Mannor and Tsitsiklis (2004) show a lower bound, which matches the upper bound of Median Elimination in Even-Dar et al. (2006). The recent work by Hassidim et al. (2020) further shows that it is possible to achieve $\frac{n}{2\epsilon^2}\log\frac{1}{\delta}$ asymptotically, where the constant factor of $1/2$ is the optimal one. Aziz et al. (2018) study (ϵ, δ)-PAC for the infinitely-armed bandits model, where they model the infinitely-armed bandits setting as the arm reservoir. Among them, Even-Dar et al. (2006); Hassidim et al. (2020) focus on the worst-case guarantee. Aziz et al. (2018) provide more instance-dependent bounds.

Best arm identification. The best arm identification is the factually dominating setting for studying the pure exploration problem. Specifically, there are two problem setups of the best arm identification, fixed budget (Audibert and Bubeck, 2010; Karnin et al., 2013; Carpentier and Locatelli, 2016) and fixed confidence (Karnin et al., 2013; Jamieson et al., 2014; Garivier and Kaufmann, 2016; Chen et al., 2017b; Simchowitz et al., 2017; Degenne and Koolen, 2019). In the fixed budget setting, the algorithm takes a budget as the input and is required to return an output before exhausting all the sample budget. The paper by Audibert and Bubeck (2010) is the first one to open up this direction and proposes Successive Rejects algorithm, which achieves the optimality up to logarithmic factors. The follow-up work of Karnin et al. (2013) proposes SH algorithm, which is also an elimination-based algorithm as Successive Reject, but it eliminates empirically bad arms more aggressively. It was not until Carpentier and Locatelli (2016) SH was proven to be optimal up to doubly-logarithmic factors.

Many meaningful results have been developed for the fixed confidence setting. To our knowledge, the paper by Garivier and Kaufmann (2016) is the first one that claims asymptotic optimality. They propose a non-asymptotic lower bound for the sample complexity and also an algorithm called Track-and-Stop that matches the lower bound asymptotically as δ goes to 0. This result is further extended to the setting where the best arm is not unique (Degenne and Koolen, 2019).

Simple regret. In our opinion, simple regret is the least understood measure for pure exploration. Though many works claim simple regret as their target, their performance bounds are either on best arm misidentification or (ϵ, δ)-PAC identification. Of course, by Audibert and Bubeck (2010), one can always use the probability of best arm misidentification to upper bound simple regret. However, in the data-poor regime, i.e., $t < n$, the probability of best arm misidentification is vacuous since we cannot even guarantee each arm has been sampled at least once. For (ϵ, δ)-PAC identification, the algorithm usually requires ϵ as the input of the algorithm. However, simple regret itself is a parameter-free performance measure. Perhaps only Carpentier and Valko (2015) truly deals with simple regret directly. The algorithm of Carpentier and Valko (2015) works for the simple regret minimization problem since it does not take such a predefined threshold ϵ as the input. There is another different problem setup in which the algorithm is required to output one of the top-m arms with ϵ slackness (Chaudhuri and Kalyanakrishnan, 2017; 2019). Though such a problem setup also characterizes how close the output is to the best one, it still needs m and ϵ as the input of the algorithm. Recent work by de2 proposes an algorithm that guarantees that it returns one of the m best arms with the same mean reward without knowledge of m as input to the algorithm. The algorithm of de2 works for both finite and infinite arms settings. However, in light of its limited problem setting that requires the top m arms to have the same mean reward, it is unclear to us if the result can be extended to guarantee simple regret or ϵ-error probability bound as we do. The model of our paper can be regarded as a generalization of de2 because all the results of de2 can be matched with our results up to logarithmic factors if we consider their setting, but our paper does not require the identity among the optimal arms. Furthermore, we provide analysis on ϵ-error probability, a generalization of

misidentification probability, and analysis on simple regret.

(ϵ, δ)-Verifiable and unverifiable sample complexity. The fixed confidence best arm identification shares some similarities with the (ϵ, δ)-PAC identification. For both settings, the algorithm is required to verify the output can meet the corresponding criteria, simple regret of 0 or simple regret of ϵ, with error probability δ. Verifiability requires the algorithm to pull each arm at least once. Therefore the (ϵ, δ)-PAC is inherently impossible for the data-poor regime, e.g., $t < n$. Katz-Samuels and Jamieson (2020) argue (ϵ, δ)-PAC is not aligned with practical usage. They propose the (ϵ, δ)-unverifiable sample complexity, which does not require the algorithm to verify the output is ϵ-good. Instead, the (ϵ, δ)-unverifiable sample complexity represents the number of samplings that the algorithm minimally needs such that it has the ability to output an ϵ-good arm. Note the sample complexity of a fixed budget algorithm (or, to be more accurate, the doubling trick version of the fixed budget algorithm) also captures the nature of the (ϵ, δ)-unverifiable sample complexity. In the reinforcement learning setting, Dann et al. (2017) also discuss the limitation of (ϵ, δ)-PAC and propose a notion called Uniform-PAC.

Top-k arm identification. The top-k arm identification problem requires the algorithm to return k arms with the highest mean reward instead of only the best one (Kalyanakrishnan and Stone, 2010; Kalyanakrishnan et al., 2012; Chen et al., 2017a; Simchowitz et al., 2017). The seminal work is Kalyanakrishnan and Stone (2010). Their problem formulation is a direct extension of the (ϵ, δ)-PAC identification problem. The goal is to return k arms that have mean rewards no less than $\mu_k - \epsilon$. The worst-case lower bound is shown in Kalyanakrishnan et al. (2012). The first instance-dependent lower bound for the k-identification problem is given by Simchowitz et al. (2017) and Chen et al. (2017a) almost at the same time, together with the near-optimal algorithms. Zhou et al. (2014) proposes the aggregate regret to measure the performance of the algorithms for top-k arm identification, which is defined as the gap between the average reward of the top-k arms and that of the output. The best arm identification can be regarded as a special case of the top-k arm identification problem. The similarity is that we are both interested in good arms, but we are measuring the likelihood of identifying one arm out of the top-k rather than the set of top-k arms.

B. Improved Analyses of Sequential Halving

B.1. Implication of Theorem 4 for $m = 1$

The following theorem corresponds to the implication of $m = 1$ of Theorem 4. However, we prove it independently here.

Theorem 7. *For any $\epsilon \in (0, 1)$, the error probability of SH for identifying an ϵ-good arm satisfies,*

$$\mathbb{P}\big(\mu_{J_T} < \mu_1 - \epsilon\big) \le 3 \log_2 n \cdot \exp\left(-\frac{\epsilon^2}{32} \frac{T}{n \log_2 n}\right).$$

Proof. To avoid redundancy and for the sake of readability, we assume n is of a power of 2. It is easy to verify the result for any n. Let $\epsilon_1 = \epsilon/4$, $T' := \frac{T}{\log_2 n}$. And define $\epsilon_{\ell+1} = \frac{3}{4} \cdot \epsilon_\ell$. For each stage ℓ, define the event G_ℓ as

$$G_\ell := \left\{ \max_{i \in S_{\ell+1}} \mu_i \ge \max_{i \in S_\ell} \mu_i - \epsilon_\ell \right\}.$$

Thus as long as $\bigcap_{\ell=1}^{\log_2 n} G_\ell$ happens, we have that the arm returned after the final stage is an ϵ-good arm, because

$$\sum_{\ell=1}^{\log_2 n} \epsilon_\ell < \sum_{\ell=1}^{\infty} \left(\frac{3}{4}\right)^{\ell-1} \cdot \epsilon_1 = \frac{\epsilon}{4} \sum_{\ell=1}^{\infty} \left(\frac{3}{4}\right)^{\ell-1} \le \frac{\epsilon}{4} \lim_{n \to \infty} \frac{1 - (3/4)^n}{1 - 3/4} = \epsilon.$$

Thus, by a union bound,

$$\begin{aligned} \mathbb{P}\big(\mu_{J_T} < \mu_1 - \epsilon\big) &\le \mathbb{P}\left(\left(\bigcap_{\ell=1}^{\log_2 n} G_\ell \right)^c \right) \\ &\le \sum_{\ell=1}^{\log_2 n} \mathbb{P}\big(G_\ell^c\big). \end{aligned} \tag{5}$$

Let a_ℓ be the best arm in S_ℓ,

$$\begin{aligned}\mathbb{P}\big(G_\ell^c\big) &= \mathbb{P}\big(G_\ell^c, \hat{\mu}_{a_\ell} < \mu_{a_\ell} - \epsilon_\ell/2\big) + \mathbb{P}\big(G_\ell^c, \hat{\mu}_{a_\ell} \geq \mu_{a_\ell} - \epsilon_\ell/2\big)\\ &\leq \mathbb{P}\big(\hat{\mu}_{a_\ell} < \mu_{a_\ell} - \epsilon_\ell/2\big) + \mathbb{P}\big(G_\ell^c \mid \hat{\mu}_{a_\ell} \geq \mu_{a_\ell} - \epsilon_\ell/2\big).\\ &\leq \exp\left(-\frac{\epsilon_\ell^2}{2}\frac{T'}{|S_\ell|}\right) + \mathbb{P}\big(G_\ell^c \mid \hat{\mu}_{a_\ell} \geq \mu_{a_\ell} - \epsilon_\ell/2\big).\end{aligned}$$

For the second term,

$$\begin{aligned}\mathbb{P}\big(G_\ell^c \mid \hat{\mu}_{a_\ell} \geq \mu_{a_\ell} - \epsilon_\ell/2\big) &\leq \mathbb{P}\Big(\big|\{i \in S_\ell \mid \hat{\mu}_i > \mu_i + \epsilon_\ell/2\}\big| \geq |S_\ell|/2\Big)\\ &\overset{(a_1)}{\leq} \frac{\mathbb{E}\Big[\big|\{i \in S_\ell \mid \hat{\mu}_i > \mu_i + \epsilon_\ell/2\}\big|\Big]}{|S_\ell|/2}\\ &\leq \frac{|S_\ell| \exp\left(-\frac{\epsilon_\ell^2}{2}\frac{T'}{|S_\ell|}\right)}{|S_\ell|/2}\\ &= 2\exp\left(-\frac{\epsilon_\ell^2}{2}\frac{T'}{|S_\ell|}\right).\end{aligned}$$

For (a_1), we use Markov's inequality. Then,

$$\mathbb{P}\big(G_\ell^c\big) \leq 3\exp\left(-\frac{\epsilon_\ell^2}{2}\frac{T'}{|S_\ell|}\right) = 3\exp\left(-\left(\frac{9}{16}\right)^{\ell-1}\frac{\epsilon^2}{32}\frac{T'}{2^{-(\ell-1)}n}\right) = 3\exp\left(-\left(\frac{9}{8}\right)^{\ell-1}\frac{\epsilon^2}{32}\frac{T'}{n}\right).$$

Taking the above into (5), we have

$$\begin{aligned}\mathbb{P}\big(\mu_{J_T} < \mu_1 - \epsilon\big) &\leq \sum_{\ell=1}^{\log_2 n} \mathbb{P}\big(G_\ell^c\big)\\ &\leq \sum_{\ell=1}^{\log_2 n} 3\exp\left(-\left(\frac{9}{8}\right)^{\ell-1}\frac{\epsilon^2}{32}\frac{T'}{n}\right)\\ &\leq 3\log_2 n \cdot \exp\left(-\frac{\epsilon^2}{32}\frac{T}{n\log_2 n}\right).\end{aligned}$$

□

B.2. Proof of Corollary 4.1

We state the following result that includes two different budget allocation schemes for SH. The appendix B.12 should be read first where we define the two budget allocation schemes:

$$\text{Option 1: } T_\ell = \left\lfloor \frac{T}{|S_\ell|\lceil \log_2 n\rceil}\right\rfloor, \quad \text{Option 2: } T_\ell = \left\lfloor \frac{T}{81n}\cdot\left(\frac{16}{9}\right)^{\ell-1}\cdot \ell\right\rfloor.$$

Corollary 4.1. *The simple regret of SH satisfies*

$$\textit{Option 1: } \mathsf{SReg}_T \leq \mathcal{O}\left(\sqrt{\frac{n\log^3 n}{T}}\right), \quad \textit{Option 2: } \mathsf{SReg}_T \leq \mathcal{O}\left(\sqrt{\frac{n}{T}}\right).$$

Proof. The simple regret can be calculated by considering the following integral with respect to ϵ. We use SH with Option

2. For Option 1, the same analysis holds.

$$\begin{aligned}\mathsf{SReg}_T &= \int_0^\infty \mathbb{P}\big(\mu_1 - \mu_{J_T} > \epsilon\big)\,\mathrm{d}\epsilon \\ &= \int_0^\infty \left(1 \wedge \exp\left(-\epsilon^2 \frac{T}{n}\right)\right)\mathrm{d}\epsilon \\ &\le \int_0^{\sqrt{\frac{n}{T}}} 1\,\mathrm{d}\epsilon + \int_0^\infty \exp\left(-\epsilon^2\frac{T}{n}\right)\mathrm{d}\epsilon \\ &\le \sqrt{\frac{n}{T}} + \int_0^\infty \exp\left(-\epsilon^2\frac{T}{n}\right)\mathrm{d}\epsilon.\end{aligned}$$

We can borrow the result from Gaussian distribution with mean 0 and variance σ^2,

$$\int_0^\infty \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{x^2}{2\sigma^2}\right)\mathrm{d}x = \frac{1}{2}.$$

Taking $\frac{1}{2\sigma^2} = \frac{T}{n}$, we have

$$\int_0^\infty \exp\left(-\epsilon^2\frac{T}{n}\right)\mathrm{d}\epsilon = \frac{\sigma\sqrt{2\pi}}{2} = \frac{\sqrt{\pi}}{2}\sqrt{\frac{n}{T}}.$$

Thus, we have

$$\mathsf{SReg}_T \le \mathcal{O}\left(\sqrt{\frac{n}{T}}\right).$$

□

B.3. Proof of Theorem 4

Theorem 4 is an equivalent result to Theorem 8, so the proof of one implies the other. While Theorem 8 takes a form that is easier to instantiate the bound for specific instances, Theorem 4 takes a form that is easier to prove. Thus, we choose to prove Theorem 4 directly, leaving the proof of Theorem 8 as a consequence of Theorem 4.

Let us first provide an intuitive explanation. Imagine one possible 'typical' scenario where, in each stage, the set of surviving arms for the next stage happens to maintain the fraction of good arms as at least m/n. This means that, after $\Theta(\log_2 m)$ stages, we expect to have at least one good arm in the surviving arm set. The number of surviving arms at that time is around n/m. At this point, the rest of the procedure can be analyzed by the standard analysis of SH where the goal is to find the arm that is the best in the current surviving arm set. Thus, it remains to bound how likely it is to have the 'typical' event above. That is, we would like to bound the failure probability of this event at each stage as tightly as possible. Bounding this failure probability is our key technical innovation (Proposition 9), which promotes the fast rate of the failure probability that improves as a function of the number of good arms. Another key technical step of the proof is a careful design of events that would guarantee the chosen arm's suboptimality in the last stage to be at most ϵ, which requires splitting ϵ into smaller pieces to be distributed over the stages.

Theorem 4. *For any $m \le n$ and any $\epsilon \in (0,1)$, the error probability of SH for identifying an (m,ϵ)-good arm satisfies*

$$\mathbb{P}\big(\mu_{J_T} < \mu_m - \epsilon\big) \le \log_2 n \cdot \exp\left(-\text{const}\cdot m\left(\frac{\epsilon^2 T}{4n\log_2^2(2m)\log_2 n} - \ln(4e)\right)\right).$$

Thus, there exist a positive constant c_1 such that for $T \ge c_1 \frac{(\ln(4e)+\ln\ln n)n\log_2^2(2m)\log_2 n}{\epsilon^2} = \tilde{\Theta}\big(\frac{n}{\epsilon^2}\big)$,

$$\mathbb{P}\big(\mu_{J_T} < \mu_m - \epsilon\big) \le \exp\left(-\tilde{\Theta}\left(m\frac{\epsilon^2 T}{n}\right)\right),$$

where, $\tilde{\Theta}$ means ignoring the logarithmic factors of m, n and constants.

Proof. Let us consider the case of $m \le n/2$ first.

Let $\ell^* = \lceil \log_2 m \rceil$, $\epsilon' = \frac{\epsilon}{2\log_2 m}$, $T' := \left\lceil \frac{T}{\log_2 n} \right\rceil$. And g_ℓ denotes the number of $(m, \ell\cdot\epsilon')$-good arms after finishing stage $\ell \in \left[\lceil \log_2 m \rceil\right]$. For stage ℓ, we define the event G_ℓ as,

$$G_\ell := \left\{ g_\ell \ge 2^{-\ell}\cdot m \right\}.$$

Specifically, we have G_{ℓ^*} to be the event that the number of $(m, \epsilon/2)$-good arms after finishing stage ℓ^* is at least 1. We define G_{ℓ^*+1} as the algorithm succeeds to return an arm in $\text{Top}_m(\epsilon)$,

$$G_{\ell^*+1} := \left\{ \mu_{J_T} \ge \mu_m - \epsilon \right\}.$$

Then the event $\bigcap_{\ell=1}^{\ell^*+1} G_\ell$ is a possible path the algorithm returns an arm in $\text{Top}_m(\epsilon)$ in the end. Thus the probability of missing all of the (m, ϵ)-good arms can be upper bounded as follows. Define $F_\ell = \left(\bigcap_{i=1}^{\ell-1} G_i\right) \cap G_\ell^c$ for $\ell > 1$, $F_1 = G_1^c$. Since $\bigcap_{\ell=1}^{\ell^*+1} G_\ell \subset G_{\ell^*+1}$ implies $G_{\ell^*+1}^c = \left\{ \mu_{J_T} < \mu_m - \epsilon \right\} \subset \left(\bigcap_{\ell=1}^{\ell^*+1} G_\ell\right)^c$,

$$\begin{aligned}
\mathbb{P}\left(\mu_{J_T} < \mu_m - \epsilon\right) &\le \mathbb{P}\left(\left(\bigcap_{\ell=1}^{\ell^*+1} G_\ell\right)^c\right) \\
&= \mathbb{P}\left(\bigcup_{\ell=1}^{\ell^*+1} F_\ell\right) \\
&\overset{(a_1)}{\le} \sum_{\ell=1}^{\ell^*+1} \mathbb{P}\left(G_\ell^c \mid \bigcap_{i=1}^{\ell-1} G_i\right) \\
&= \sum_{\ell=1}^{\ell^*} \mathbb{P}\left(G_\ell^c \mid \bigcap_{i=1}^{\ell-1} G_i\right) + \mathbb{P}\left(G_{\ell^*+1}^c \mid \bigcap_{i=1}^{\ell^*} G_i\right). \qquad (6)
\end{aligned}$$

For (a_1), we use $\mathbb{P}(A, B) \le \mathbb{P}(A|B)$.

For the first term of (6), we apply the result of Proposition 9 to each stage of SH with the parameters therein as $k = 2^{-\ell}\cdot m$, $s =$

$2^{-\ell}\cdot n$, $m' = 2^{-\ell+1}\cdot m$, $\epsilon = \epsilon'$, and $T = T'$. Thus, for stage ℓ,

$$
\begin{aligned}
&\mathbb{P}\big(G_\ell^c \mid G_{\ell-1}\big) \\
&\le \binom{2^{-\ell+1}\cdot m}{2^{-\ell}\cdot m}\exp\left(-2^{-\ell}\cdot m\frac{\epsilon'^2T'}{2\cdot 2^{-(\ell-1)}\cdot n}\right) + \binom{2^{-\ell+1}\cdot n}{2^{-\ell}\cdot(n-m)}\exp\left(-2^{-\ell}\cdot(n-m)\frac{\epsilon'^2T'}{2\cdot 2^{-(\ell-1)}\cdot n}\right) \\
&\overset{(a_3)}{\le} h_{\ell,1}\exp\left(-2^{-\ell}\cdot m\frac{\epsilon'^2T'}{2\cdot 2^{-(\ell-1)}\cdot n}\right) + h_{\ell,2}\exp\left(-2^{-\ell}\cdot(n-m)\frac{\epsilon'^2T'}{2\cdot 2^{-(\ell-1)}\cdot n}\right) \\
&\le h_{\ell,1}\exp\left(-\frac{m}{2}\frac{\epsilon'^2T'}{2n}\right) + h_{\ell,2}\exp\left(-\frac{n-m}{2}\frac{\epsilon'^2T'}{2n}\right) \\
&\overset{(a_4)}{\le} h_{\ell,1}\exp\left(-\frac{m}{2}\frac{\epsilon'^2T'}{2n}\right) + h_{\ell,2}\exp\left(-\frac{n}{4}\frac{\epsilon'^2T'}{2n}\right) \\
&\le \exp\left(-\frac{m}{2}\frac{\epsilon^2T}{2n\log_2^2 m\log_2 n} + \ln(2e)2^{-\ell}\cdot m\right) \\
&\quad + \exp\left(-\frac{n}{4}\frac{\epsilon^2T}{2n\log_2^2 m\log_2 n} + \ln(4e)2^{-\ell}\cdot(n-m)\right) \\
&\le \exp\left(-m\left(\frac{\epsilon^2T}{4n\log_2^2 m\log_2 n} - 2\ln(4e)2^{-\ell}\right)\right) \\
&\quad + \exp\left(-\frac{n}{2}\left(\frac{\epsilon^2T}{4n\log_2^2 m\log_2 n} - 2\ln(4e)2^{-\ell}\right)\right) \\
&\le 2\exp\left(-m\left(\frac{\epsilon^2T}{4n\log_2^2 m\log_2 n} - 2\ln(4e)2^{-\ell}\right)\right)
\end{aligned}
\tag{7}
$$

where (a_3) is $\binom{2^{-\ell+1}\cdot m}{2^{-\ell}\cdot m} \le (2e)^{2^{-\ell}\cdot m} =: h_{\ell,1}$ (Sterling's formula $\binom{x}{y} \le \left(\frac{ex}{y}\right)^y$) and $\binom{2^{-\ell+1}\cdot n}{2^{-\ell}\cdot(n-m)} \le (2en/(n-m))^{2^{-\ell}\cdot(n-m)} \le (4e)^{2^{-\ell}\cdot n} =: h_{\ell,2}$ and (a_4) is by the assumption $m \le n/2$.

Then we have,

$$
\sum_{\ell=1}^{\ell^*}\mathbb{P}\big(G_\ell^c \mid G_{\ell-1}\big) \le \text{const}\cdot\log_2 m\cdot\exp\left(-m\left(\frac{\epsilon^2T}{4n\log_2^2 m\log_2 n} - \ln(4e)\right)\right). \tag{8}
$$

For the second term of (6), $G_{\ell^*+1}^c \mid G_{\ell^*}$ indicates the events where SH fails to return an (m,ϵ)-good arm given the event that there is at least one $(m,\epsilon/2)$-good arm after finishing stage ℓ^*. Note the stages from ℓ^* to the last stage can be regarded as a whole process of SH with a budget $\left(\lceil\log_2 n\rceil - \ell^*\right)T' \ge \text{const}\cdot\log_2\left(\frac{n}{m}\right)\frac{T}{\log_2(n)}$. The initial arms are the surviving arms after finishing stage ℓ^*. Note we have $2^{-\ell^*}\cdot n = n/m$ arms surviving, denoted by S_{ℓ^*}, after finishing stage ℓ^*. Let c be the index of the true best arm in S_{ℓ^*}. Due to the event G_{ℓ^*}, we have $\mu_c \ge \mu_m - \frac{\epsilon}{2}$. Then, by applying the result of Theorem 7,

$$
\begin{aligned}
\mathbb{P}\big(G_{\ell^*+1}^c \mid G_{\ell^*}\big) &= \mathbb{P}\big(\mu_{J_T} < \mu_m - \epsilon \mid G_{\ell^*}\big) \\
&\le \mathbb{P}\left(\mu_{J_T} < \mu_c - \frac{\epsilon}{2} \mid G_{\ell^*}\right) \\
&\le 3\log_2\left(\frac{n}{m}\right)\cdot\exp\left(-\frac{m}{128}\frac{\epsilon^2\left(\lceil\log_2 n\rceil - \ell^*\right)T'}{n\log_2\left(\frac{n}{m}\right)}\right) \\
&\le \log_2\left(\frac{n}{m}\right)\cdot\exp\left(-\text{const}\cdot m\frac{\epsilon^2T}{n\log_2 n}\right).
\end{aligned}
\tag{9}
$$

Bringing (8),(9) into (6), we have,

$$\mathbb{P}\big(\mu_{J_T} < \mu_m - \epsilon\big) \le \log_2 n \cdot \exp\left(-\text{const} \cdot m\left(\frac{\epsilon^2 T}{4n\log_2^2(m)\log_2 n} - \ln(4e)\right)\right).$$

Note the above analysis is for $m > 1$. To incorporate the result of Theorem 7 for $m = 1$, we rewrite the final formula as

$$\mathbb{P}\big(\mu_{J_T} < \mu_m - \epsilon\big) \le \log_2 n \cdot \exp\left(-\text{const} \cdot m\left(\frac{\epsilon^2 T}{4n\log_2^2(2m)\log_2 n} - \ln(4e)\right)\right).$$

Thus, there exist a positive constant c_1 such that for $T \ge c_1 \frac{(\ln(4e)+\ln\ln n) n \log_2^2(2m)\log_2 n}{\epsilon^2} = \tilde{\Theta}\big(\frac{n}{\epsilon^2}\big)$,

$$\mathbb{P}\big(\mu_{J_T} < \mu_m - \epsilon\big) \le \exp\left(-\tilde{\Theta}\left(m\frac{\epsilon^2 T}{n}\right)\right).$$

For the case of $m > n/2$, we consider

$$\mathbb{P}\big(\mu_{J_T} < \mu_m - \epsilon\big) \le \mathbb{P}\big(\mu_{J_T} < \mu_{n/2} - \epsilon\big).$$

Thus, one can repeat the same analysis as above with m replaced by $n/2$. Using $m/2 \le n/2 \le m$, the statement of this theorem holds. □

The result of Theorem 4 directly improves the previous works including $\mathcal{O}\Big(\frac{n}{m\epsilon^2}\log^2\big(\frac{1}{\delta}\big)\Big)$ of Chaudhuri and Kalyanakrishnan (2017), and $\mathcal{O}\Big(\frac{1}{\epsilon^2}\Big(\frac{n}{m}\log\big(\frac{1}{\delta}\big) + \log^2\big(\frac{1}{\delta}\big)\Big)\Big)$ of Chaudhuri and Kalyanakrishnan (2019). They share a similar strategy that, assuming the knowledge of m, ϵ and δ, uniformly samples $\Theta\Big(\frac{n}{m}\log\big(\frac{1}{\delta}\big)\Big)$ arms first and then runs the Median-Elimination algorithm of Even-Dar et al. (2006) or LUCB of Kaufmann and Kalyanakrishnan (2013) for this subset.

B.4. Proof of Theorem 8

Theorem 8. *For any $\epsilon \in (0,1)$, the error probability of SH for identifying an ϵ-good arm satisfies,*

$$\mathbb{P}\big(\mu_{J_T} < \mu_1 - \epsilon\big) \le \min_{\{(m',\epsilon'):\Delta_{m'}+\epsilon' \le \epsilon\}} \log_2 n \cdot \exp\left(-\text{const} \cdot m'\left(\frac{\epsilon'^2 T}{4n\log_2^2(2m')\log_2 n} - \ln(4e)\right)\right).$$

Proof. Note the result we proved in Theorem 4 holds for any $m \le n$ and any $\epsilon \in (0,1)$. Thus $\mathbb{P}\big(\mu_{J_T} < \mu_1 - \epsilon\big)$ can be upper bounded for any $(m', \epsilon') \in \big\{(m', \epsilon') : \Delta_{m'} + \epsilon' \le \epsilon\big\}$. The result is therefore implied. □

B.5. Proof of Theorem 1

Theorem 1. *For any $\epsilon \in [0,1)$, the ϵ-error probability of SH satisfies*

$$\mathbb{P}\big(\mu_{J_T} < \mu_1 - \epsilon\big) \le \exp\left(-\tilde{\Theta}\left(\frac{T}{\max_{i \ge g(\epsilon)+1} \frac{i}{\Delta_i^2} \cdot \frac{1}{g\left(\frac{\epsilon}{2}\right)}}\right)\right),$$

for $T \ge \tilde{\Theta}\Big(\max_{i \ge g(\epsilon)+1} \frac{i}{\Delta_i^2}\Big)$.

Proof. To avoid redundancy and for the sake of readability, we assume n is of a power of 2. It is easy to verify the result for any n. Recall $g\big(\frac{\epsilon}{2}\big)$ is the number of $\frac{\epsilon}{2}$-good arms in the instance, and $g_\ell\big(\frac{\epsilon}{2}\big)$ is the number of $\frac{\epsilon}{2}$-good arms at the beginning of stage ℓ. Thus $g_1\big(\frac{\epsilon}{2}\big) = g\big(\frac{\epsilon}{2}\big)$. Let $\ell' = \max\Big\{1 \le \ell \le \log_2(n) - 2 \mid \Delta_{\frac{n}{2^{\ell+1}}} > \epsilon\Big\}$. There are three cases we need to consider:

- $\ell' = \log_2 n - 2$.
- $1 \le \ell' < \log_2 n - 2$.
- Such an ℓ' does not exist.

The last case that such a ℓ' does not exit means $\epsilon \ge \Delta_{\frac{n}{4}}$, which means that the instance has $\Theta(n)$ amount of ϵ-good arms, therefore an easy instance. We address this case at the end of the proof.

Case 1. $\ell' = \log_2 n - 2$
In this case, we have $\Delta_2 > \epsilon$, then the problem is a fixed budget best arm identification problem rather than an ϵ-good arm identification problem. Since $g(\epsilon) = g\left(\frac{\epsilon}{2}\right) = 1$ for $\Delta_2 > \epsilon$, the theorem statement reduces to

$$\mathbb{P}\left(\mu_1 - \mu_{J_T} > \epsilon\right) \le \exp\left(-\tilde{\Theta}\left(\frac{T}{\max_{i\ge 2} \frac{i}{\Delta_i^2}}\right)\right),$$

which is a trivial consequence of (Karnin et al., 2013, Theorem 4.1) under the theorem's condition on T.

Case 2. $1 \le \ell' < \log_2 n - 2$
For $1 \le \ell' < \log_2 n - 2$, recall that $|A_\ell| = \frac{n}{2^{\ell-1}}$. One can show that

$$|A_{\ell'+1}|/4 \le g(\epsilon) \le |A_{\ell'}|/4 \quad \text{and} \quad \Delta_{|A_{\ell'+1}|/4} \le \Delta_{g(\epsilon)} \le \epsilon < \Delta_{|A_{\ell'}|/4}\,. \tag{10}$$

For $\ell \in [\ell'+1]$, define the good event for stage ℓ as

$$G_\ell := \left\{ g_\ell\left(\frac{\epsilon}{2}\right) \ge g\left(\frac{\epsilon}{2}\right) - \left\lceil (\ell-1)\frac{g\left(\frac{\epsilon}{2}\right)}{2\log_2 n}\right\rceil + 1 \right\}.$$

Note G_1 holds with probability 1. Define $m' := \left\lceil \frac{g\left(\frac{\epsilon}{2}\right)}{2\log_2 n}\right\rceil = \tilde{\Theta}\left(g\left(\frac{\epsilon}{2}\right)\right)$. Intuitively, we allow at most m' of $\frac{\epsilon}{2}$-good arms to be eliminated per stage in the initial phases, i.e., $\ell \in [\ell'+1]$ of running a Sequential Halving. The ϵ-error probability can be expressed as

$$\mathbb{P}\left(\mu_1 - \mu_{J_T} > \epsilon\right) = \mathbb{P}\left(\mu_1 - \mu_{J_T} > \epsilon, \cap_{\ell=2}^{\ell'+1} G_\ell\right) + \mathbb{P}\left(\mu_1 - \mu_{J_T} > \epsilon, \left(\cap_{\ell=2}^{\ell'+1} G_\ell\right)^c\right).$$

Since

$$\mathbb{P}\left(\mu_1 - \mu_{J_T} > \epsilon, \cap_{\ell=2}^{\ell'+1} G_\ell\right) \le \mathbb{P}\left(\mu_1 - \mu_{J_T} > \epsilon \mid \cap_{\ell=2}^{\ell'+1} G_\ell\right),$$

and

$$\mathbb{P}\left(\mu_1 - \mu_{J_T} > \epsilon, \left(\cap_{\ell=2}^{\ell'+1} G_\ell\right)^c\right) \le \mathbb{P}\left(\left(\cap_{\ell=2}^{\ell'+1} G_\ell\right)^c\right) \le \sum_{\ell=1}^{\ell'} \mathbb{P}\left(G_{\ell+1}^c \mid \cap_{i=1}^{\ell} G_i\right),$$

we have

$$\mathbb{P}\left(\mu_1 - \mu_{J_T} > \epsilon\right) \le \mathbb{P}\left(\mu_1 - \mu_{J_T} > \epsilon \mid \cap_{\ell=2}^{\ell'+1} G_\ell\right) + \sum_{\ell=1}^{\ell'} \mathbb{P}\left(G_{\ell+1}^c \mid \cap_{i=1}^{\ell} G_i\right). \tag{11}$$

We bound the two terms in (11) respectively.

We start with the second term. Let $\ell \in [\ell']$. To bound $\mathbb{P}\left(G_{\ell+1}^c \mid \cap_{i=1}^{\ell} G_i\right)$ with Lemma 9.1, we need to relate the quantities in stage ℓ with those in Lemma 9.1.

We first claim that, given $\cap_{i=1}^{\ell} G_i$, the event $G_{\ell+1}^c$ implies that there are at least m' of $\frac{\epsilon}{2}$-good arms that are eliminated during the stage ℓ, i.e., $g_\ell\left(\frac{\epsilon}{2}\right) - g_{\ell+1}\left(\frac{\epsilon}{2}\right) \ge m'$.

To see why, we have by $\cap_{i=1}^{\ell} G_i$,

$$g_\ell\left(\frac{\epsilon}{2}\right) \geq g\left(\frac{\epsilon}{2}\right) - \left\lceil (\ell-1)\frac{g\left(\frac{\epsilon}{2}\right)}{2\log_2 n}\right\rceil + 1,$$

and by $G_{\ell+1}^c$,

$$g_{\ell+1}\left(\frac{\epsilon}{2}\right) \leq g\left(\frac{\epsilon}{2}\right) - \left\lceil \ell\frac{g\left(\frac{\epsilon}{2}\right)}{2\log_2 n}\right\rceil.$$

Therefore, using $\lceil x+y\rceil - \lceil x\rceil + 1 \geq \lceil y\rceil$, we have

$$g_\ell(\frac{\epsilon}{2}) - g_{\ell+1}(\frac{\epsilon}{2}) = \left\lceil \ell\frac{g\left(\frac{\epsilon}{2}\right)}{2\log_2 n}\right\rceil + 1 - \left\lceil (\ell-1)\frac{g\left(\frac{\epsilon}{2}\right)}{2\log_2 n}\right\rceil \geq \left\lceil \frac{g\left(\frac{\epsilon}{2}\right)}{2\log_2 n}\right\rceil = m',$$

which proves the claim.

Define A_ℓ as the set of surviving arms at the beginning of stage ℓ, $A_{\ell,i}$ as i-th best arm in A_ℓ (ties are broken arbitrarily), and $A_\ell[L+1:R] := \left\{A_{\ell,i}, i \in [R]\setminus[L]\right\}$. Let $d_\ell(\epsilon)$ be the smallest gap of the mean rewards between any two arms chosen from $A_\ell\left[\frac{n}{2^{\ell+1}}+1 : \frac{n}{2^{\ell-1}}\right]$ and $A_\ell\left[1 : g_\ell\left(\frac{\epsilon}{2}\right)\right]$ respectively, i.e.,

$$d_\ell(\epsilon) = \min\left\{\mu_i - \mu_j \mid j \in A_\ell\left[\frac{n}{2^{\ell+1}}+1 : \frac{n}{2^{\ell-1}}\right], i \in A_\ell\left[1 : g_\ell\left(\frac{\epsilon}{2}\right)\right]\right\}.$$

The definition of $d_\ell(\epsilon)$ is same as d_q of Lemma 9.1. By Lemma 8.1, we have $d_\ell(\epsilon) \geq \frac{1}{2}\Delta_{\frac{n}{2^{\ell+1}}}$. Define $S_\ell(\epsilon) := \frac{\Delta^2_{\frac{n}{2^{\ell+1}}}}{n\cdot 2^{-\ell+1}}$.

We can now apply Lemma 9.1 with $T = T'$, $s = \frac{n}{2^\ell}$, $k = m'$, and $q = \frac{n}{2^{\ell+1}}$ as follows:

$$\begin{aligned}
&\mathbb{P}\left(G_{\ell+1}^c \mid \cap_{i=1}^{\ell} G_i\right)\\
&\leq \binom{g_\ell\left(\frac{\epsilon}{2}\right)}{m'}\exp\left(-m'\frac{d_\ell^2(\epsilon)T'}{8n\cdot 2^{-\ell+1}}\right) + \binom{\frac{3}{4}n\cdot 2^{-\ell+1}}{\frac{1}{4}n\cdot 2^{-\ell+1}+m'}\exp\left(-\left(\frac{1}{4}n\cdot 2^{-\ell+1}+m'\right)\frac{d_\ell^2(\epsilon)T'}{8n\cdot 2^{-\ell+1}}\right)\\
&\overset{b_1}{\leq} \binom{g_\ell\left(\frac{\epsilon}{2}\right)}{m'}\exp\left(-m'\frac{d_\ell^2(\epsilon)T'}{8n\cdot 2^{-\ell+1}}\right) + \binom{n\cdot 2^{-\ell+1}}{n\cdot 2^{-\ell}}\exp\left(-\frac{n}{2^{\ell+1}}\frac{d_\ell^2(\epsilon)T'}{8n\cdot 2^{-\ell+1}}\right)\\
&\overset{b_2}{\leq} \exp\left(-m'\frac{d_\ell^2(\epsilon)T'}{8n\cdot 2^{-\ell+1}} + m'\ln\left(\frac{eg_\ell\left(\frac{\epsilon}{2}\right)}{m'}\right)\right) + \exp\left(-\frac{n}{2^{\ell+1}}\frac{d_\ell^2(\epsilon)T'}{8n\cdot 2^{-\ell+1}} + \frac{n}{2^\ell}\ln(2e)\right)\\
&\leq \exp\left(-m'\frac{\Delta^2_{\frac{n}{2^{\ell+1}}}T'}{32n\cdot 2^{-\ell+1}} + m'\ln\left(\frac{eg_\ell\left(\frac{\epsilon}{2}\right)}{m'}\right)\right) + \exp\left(-\frac{n}{2^{\ell+1}}\frac{\Delta^2_{\frac{n}{2^{\ell+1}}}T'}{32n\cdot 2^{-\ell+1}} + \frac{n}{2^\ell}\ln(2e)\right) && (d_\ell(\epsilon) \geq \tfrac{1}{2}\Delta_{\frac{n}{2^{\ell+1}}})\\
&\overset{b_3}{\leq} \exp\left(-\tilde{\Theta}\left(m'S_\ell(\epsilon)T\right)\right) + \exp\left(-\tilde{\Theta}\left(\frac{n}{2^{\ell+1}}S_\ell(\epsilon)T\right)\right), && (T \geq \tilde{\Theta}\left(\tfrac{1}{S_\ell(\epsilon)}\right))
\end{aligned} \tag{12}$$

where,

- b_1 is by $\binom{x}{y} < \binom{x'}{y}$ for $x' > x$ and $\max_y \binom{x}{y} = \binom{x}{\lfloor x/2\rfloor} = \binom{x}{\lceil x/2\rceil}$,
- b_2 is by upper bounding the binomial coefficients by Sterling's formula $\binom{x}{y} \leq \left(\frac{ex}{y}\right)^y$ and moving the coefficients into the exponent.

- b_3 is by $T \geq \tilde{\Theta}\Big(\frac{1}{S_\ell(\epsilon)}\Big)$ which is implied by $T \geq \tilde{\Theta}\Big(\max_{i \geq g(\epsilon)+1} \frac{i}{\Delta_i^2}\Big)$ in the theorem statement.

By the definition of ℓ' and the fact of $\ell \in [\ell']$, we have

$$\Delta_{\frac{n}{2^{\ell+1}}} > \epsilon.$$

By the definition of $g\big(\frac{\epsilon}{2}\big)$, we have

$$\Delta_{g(\frac{\epsilon}{2})} \leq \frac{\epsilon}{2}.$$

Thus

$$\Delta_{\frac{n}{2^{\ell+1}}} - \Delta_{g(\frac{\epsilon}{2})} > \frac{\epsilon}{2},$$

which means $\frac{n}{2^{\ell+1}} > g\big(\frac{\epsilon}{2}\big) \geq m'$. We continue to upper bound (12). For $T \geq \tilde{\Theta}\Big(\frac{1}{S_\ell(\epsilon)}\Big)$, we have

$$\begin{aligned}
\mathbb{P}\Big(G_{\ell+1}^c \mid \cap_{i=1}^{\ell} G_i\Big) &\leq \exp\Big(-\tilde{\Theta}\big(m' S_\ell(\epsilon) T\big)\Big) + \exp\Big(-\tilde{\Theta}\Big(\frac{n}{2^{\ell+1}} S_\ell(\epsilon) T\Big)\Big) \\
&\leq \exp\Big(-\tilde{\Theta}\big(m' S_\ell(\epsilon) T\big)\Big) + \exp\Big(-\tilde{\Theta}\big(m' S_\ell(\epsilon) T\big)\Big) \\
&\leq \exp\Big(-\tilde{\Theta}\Big(g\Big(\frac{\epsilon}{2}\Big) S_\ell(\epsilon) T\Big)\Big). \qquad \Big(m' = \tilde{\Theta}\Big(g\big(\tfrac{\epsilon}{2}\big)\Big)\Big)
\end{aligned} \tag{13}$$

Next, we bound the first term in (11). The stages $\ell \geq \ell' + 1$ can be viewed as a new Sequential Halving game with $A_{\ell'+1}$ as the initial arm set, $g_{\ell'+1}\big(\frac{\epsilon}{2}\big)$ of which are $\frac{\epsilon}{2}$-good. When $\cap_{i=1}^{\ell'+1} G_i$ happens, we have, as $\ell' < \log_2 n$,

$$g_{\ell'+1}\Big(\frac{\epsilon}{2}\Big) \geq g\Big(\frac{\epsilon}{2}\Big) - \left\lceil \ell' \frac{g\big(\frac{\epsilon}{2}\big)}{2\log_2 n} \right\rceil + 1 \geq g\Big(\frac{\epsilon}{2}\Big) - \left\lceil \frac{g\big(\frac{\epsilon}{2}\big)}{2} \right\rceil + 1 \geq \frac{g\big(\frac{\epsilon}{2}\big)}{2} = \tilde{\Theta}\Big(g\Big(\frac{\epsilon}{2}\Big)\Big).$$

By Lemma 8.2, a $\Big(g_{\ell'+1}\big(\frac{\epsilon}{2}\big), \frac{\Delta_{g(\epsilon)+1}}{2}\Big)$-good arm in $A_{\ell'+1}$ returned after finishing the final stage of Sequential Halving is an ϵ-good arm. We define $\mu_m(A)$ as the m-th largest value in $\{\mu_j : j \in A\}$. We apply Theorem 4 and set m, n, ϵ of Theorem 4 as $m = g_{\ell'+1}\big(\frac{\epsilon}{2}\big), n = |A_{\ell'+1}|, \epsilon = \frac{\Delta_{g(\epsilon)+1}}{2}$. To apply Theorem 4, we need the condition of $T \geq \tilde{\Theta}\big(\frac{n}{\epsilon^2}\big)$ in Theorem 4 to be true, which means we need $T \geq \tilde{\Theta}\Big(\frac{|A_{\ell'+1}|}{\Delta^2_{g(\epsilon)+1}}\Big)$. Note by (10), we have $\frac{n}{2^{\ell'+2}} \leq g(\epsilon)$, thus $|A_{\ell'+1}| = \frac{n}{2^{\ell'}} \leq 4g(\epsilon) \leq 4\big(g(\epsilon)+1\big)$. Using this bound and the condition on T from the theorem statement, we have $T \geq \tilde{\Theta}\Big(\max_{i \geq g(\epsilon)+1} \frac{i}{\Delta_i^2}\Big) \geq \tilde{\Theta}\Big(\frac{g(\epsilon)+1}{\Delta^2_{g(\epsilon)+1}}\Big) \geq \tilde{\Theta}\Big(\frac{|A_{\ell'+1}|}{\Delta^2_{g(\epsilon)+1}}\Big)$ So, the condition of Theorem 4 is satisfied. Then, we have

$$\begin{aligned}
&\mathbb{P}\Big(\mu_1 - \mu_{J_T} > \epsilon \mid \cap_{i=1}^{\ell'+1} G_i\Big) \\
&\leq \mathbb{P}\Big(\mu_{g_{\ell'+1}(\frac{\epsilon}{2})}(A_{\ell'+1}) - \mu_{J_T} > \frac{1}{2}\Delta_{g(\epsilon)+1} \mid \cap_{i=1}^{\ell'+1} G_i\Big) && \text{(Lemma 8.2)} \\
&\leq \exp\left(-\tilde{\Theta}\left(g_{\ell'+1}\Big(\frac{\epsilon}{2}\Big) \cdot \frac{\Delta^2_{g(\epsilon)+1} T'}{|A_{\ell'+1}|}\right)\right) && \Big(T \geq \tilde{\Theta}\Big(\tfrac{g(\epsilon)+1}{\Delta^2_{g(\epsilon)+1}}\Big)\Big) \\
&\leq \exp\left(-\tilde{\Theta}\left(g\Big(\frac{\epsilon}{2}\Big) \cdot \frac{\Delta^2_{g(\epsilon)+1} T'}{4\big(g(\epsilon)+1\big)}\right)\right). && \Big(g_{\ell'+1}\big(\tfrac{\epsilon}{2}\big) = \tilde{\Theta}\Big(g\big(\tfrac{\epsilon}{2}\big)\Big)\Big)
\end{aligned} \tag{14}$$

The role of considering ℓ' is that as long as the Sequential Halving goes well ($\cap_{i=1}^{\ell'+1} G_i$ happens) the ratio of $\frac{\epsilon}{2}$-good arms to the all surviving arms is increasing to $g\left(\frac{\epsilon}{2}\right)/g(\epsilon)$ after finishing stage ℓ'. Specifically, for the Polynomial(α) instance of our interest where $\Delta_i = \left(\frac{i}{n}\right)^\alpha$ with $\alpha > 0.5$, $g\left(\frac{\epsilon}{2}\right)/g(\epsilon)$ can be lower bounded by a constant. The last step is to combine (13) and (14). For $T \geq \tilde{\Theta}\left(\max_{i \geq g(\epsilon)+1} \frac{i}{\Delta_i^2}\right)$, we have

$$
\begin{aligned}
\mathbb{P}\left(\mu_1 - \mu_{J_T} > \epsilon\right) &\leq \mathbb{P}\left(\mu_1 - \mu_{J_T} > \epsilon \mid \cap_{\ell=2}^{\ell'+1} G_\ell\right) + \sum_{\ell=1}^{\ell'} \mathbb{P}\left(G_{\ell+1}^c \mid \cap_{i=1}^{\ell} G_i\right) \\
&\leq \exp\left(-\tilde{\Theta}\left(g\left(\frac{\epsilon}{2}\right) \cdot \frac{\Delta_{g(\epsilon)+1}^2 T}{4\left(g(\epsilon)+1\right)}\right)\right) + \log(n) \exp\left(-\tilde{\Theta}\left(g\left(\frac{\epsilon}{2}\right) \min_{\ell \leq \ell'} S_\ell(\epsilon) T\right)\right) \\
&\leq \exp\left(-\tilde{\Theta}\left(\min\left\{g\left(\frac{\epsilon}{2}\right) \cdot \frac{\Delta_{g(\epsilon)+1}^2}{4\left(g(\epsilon)+1\right)}, g\left(\frac{\epsilon}{2}\right) \min_{\ell \leq \ell'} S_\ell(\epsilon)\right\} T\right)\right) \qquad \left(T \geq \tilde{\Theta}\left(\max_{i \geq g(\epsilon)+1} \frac{i}{\Delta_i^2}\right)\right) \\
&\leq \exp\left(-\tilde{\Theta}\left(g\left(\frac{\epsilon}{2}\right) \cdot \min\left\{\frac{\Delta_{g(\epsilon)+1}^2}{4\left(g(\epsilon)+1\right)}, \min_{\ell \leq \ell'} \frac{\Delta_{\frac{n}{2^{\ell+1}}}^2}{\frac{4n}{2^{\ell+1}}}\right\} T\right)\right) \\
&\leq \exp\left(-\tilde{\Theta}\left(g\left(\frac{\epsilon}{2}\right) \cdot \min_{i \geq g(\epsilon)+1} \frac{\Delta_i^2}{i} T\right)\right) \\
&= \exp\left(-\tilde{\Theta}\left(\frac{T}{\max_{i \geq g(\epsilon)+1} \frac{i}{\Delta_i^2} \cdot \frac{1}{g\left(\frac{\epsilon}{2}\right)}}\right)\right).
\end{aligned}
$$

Case 3. ℓ' does not exist
This case condition implies that $\Delta_{\frac{n}{4}} \leq \epsilon$, which means the number of ϵ-good arms is $\Theta(n)$, so we can use (14) solely to claim our bound. □

Lemma 8.1. *In the case of $1 \leq \ell' < \log_2 n - 2$, for $\ell \in [\ell']$, define $d_\ell(\epsilon)$ as the minimum gap of the mean rewards between any two arms chosen from $A_\ell\left[\frac{n}{2^{\ell+1}} + 1 : \frac{n}{2^{\ell-1}}\right]$ and $A_\ell\left[1 : g_\ell\left(\frac{\epsilon}{2}\right)\right]$ respectively,*

$$
d_\ell(\epsilon) = \min\left\{\mu_i - \mu_j \mid \forall j \in A_\ell\left[\frac{n}{2^{\ell+1}} + 1 : \frac{n}{2^{\ell-1}}\right], \forall i \in A_\ell\left[1 : g_\ell\left(\frac{\epsilon}{2}\right)\right]\right\}.
$$

Then $d_\ell(\epsilon)$ satisfies

$$
d_\ell(\epsilon) \geq \frac{1}{2} \Delta_{\frac{n}{2^{\ell+1}}}.
$$

Proof. By definition, we have $A_\ell \subset A_1$, which implies the i-th best arm in A_ℓ has an mean reward lower than or equal to the i-th best arm in A_1. Then we have

$$
\begin{aligned}
\max\left\{\mu_i \mid i \in A_\ell\left[\frac{n}{2^{\ell+1}} + 1 : \frac{n}{2^{\ell-1}}\right]\right\} &\leq \max\left\{\mu_i \mid i \in A_1\left[\frac{n}{2^{\ell+1}} + 1 : \frac{n}{2^{\ell-1}}\right]\right\} \\
&\leq \mu_{\frac{n}{2^{\ell+1}}} \\
&= \mu_1 - \Delta_{\frac{n}{2^{\ell+1}}}.
\end{aligned}
$$

Also,

$$
\min\left\{\mu_i, i \in A_\ell\left[1 : g_\ell\left(\frac{\epsilon}{2}\right)\right]\right\} \geq \mu_1 - \frac{\epsilon}{2}.
$$

Therefore for $\forall j \in A_\ell\left[\frac{n}{2^{\ell+1}}+1:\frac{n}{2^{\ell-1}}\right]$ and $\forall i \in A_\ell\left[1:g_\ell\left(\frac{\epsilon}{2}\right)\right]$,

$$\mu_i-\mu_j \geq \mu_1-\frac{\epsilon}{2}-\left(\mu_1-\Delta_{\frac{n}{2^{\ell+1}}}\right)=\Delta_{\frac{n}{2^{\ell+1}}}-\frac{\epsilon}{2}.$$

Since $\ell \in [\ell']$, by the definition of ℓ' we have

$$\Delta_{\frac{n}{2^{\ell+1}}} > \epsilon.$$

Thus

$$\mu_i-\mu_j \geq \Delta_{\frac{n}{2^{\ell+1}}}-\frac{\epsilon}{2}=\frac{\Delta_{\frac{n}{2^{\ell+1}}}}{2}+\frac{\Delta_{\frac{n}{2^{\ell+1}}}}{2}-\frac{\epsilon}{2}>\frac{\Delta_{\frac{n}{2^{\ell+1}}}}{2}.$$

□

Lemma 8.2. *A $\left(g_{\ell'+1}\left(\frac{\epsilon}{2}\right),\frac{\Delta_{g(\epsilon)+1}}{2}\right)$-good arm in $A_{\ell'+1}$ returned after finishing the final stage of Sequential Halving is an ϵ-good arm.*

Proof. We first claim that

$$\Delta_{J_T} > \epsilon \iff \Delta_{J_T} > \Delta_{g(\epsilon)}\ .$$

To see this, for the forward direction, we have $\Delta_{J_T} > \epsilon \geq \Delta_{g(\epsilon)}$. For the other direction, if $\Delta_{J_T} > \Delta_{g(\epsilon)}$ is true, then $J_T > g(\epsilon)$. This means that $\Delta_{J_T} > \epsilon$. Then, we can also show that

$$\Delta_{J_T} > \epsilon \iff \Delta_{J_T} \geq \Delta_{g(\epsilon)+1},$$

which can be shown easily by showing

$$\Delta_{J_T} \geq \Delta_{g(\epsilon)+1} \iff \Delta_{J_T} > \Delta_{g(\epsilon)}.$$

Define $\Delta_m(A)$ to be the m-th smallest value in $\{\Delta_i : i \in A\}$. Then, we have $\Delta_{g_{\ell'+1}\left(\frac{\epsilon}{2}\right)}(A_{\ell'+1}) \leq \frac{1}{2}\epsilon$. Using the fact that $\epsilon < \Delta_{g(\epsilon)+1}$, we have

$$\begin{aligned}
\Delta_{J_T} \geq \Delta_{g(\epsilon)+1} &= \Delta_{g_{\ell'+1}\left(\frac{\epsilon}{2}\right)}(A_{\ell'+1})+\Delta_{g(\epsilon)+1}-\Delta_{g_{\ell'+1}\left(\frac{\epsilon}{2}\right)}(A_{\ell'+1}) \\
&\geq \Delta_{g_{\ell'+1}\left(\frac{\epsilon}{2}\right)}(A_{\ell'+1})+\Delta_{g(\epsilon)+1}-\frac{1}{2}\epsilon \\
&> \Delta_{g_{\ell'+1}\left(\frac{\epsilon}{2}\right)}(A_{\ell'+1})+\frac{1}{2}\Delta_{g(\epsilon)+1}. && (\epsilon < \Delta_{g(\epsilon)+1})
\end{aligned}$$

By defining $\mu_m(A)$ as the m-th largest value in $\{\mu_j : j \in A\}$, the above equivalently means

$$\mu_{J_T} < \mu_{g_{\ell'+1}\left(\frac{\epsilon}{2}\right)}(A_{\ell'+1})-\frac{1}{2}\Delta_{g(\epsilon)+1}.$$

Therefore the output is not a $\left(g_{\ell'+1}\left(\frac{\epsilon}{2}\right),\frac{\Delta_{g(\epsilon)+1}}{2}\right)$-good arm in $A_{\ell'+1}$. Our claim holds by contraposition. □

B.6. Proof of Corollary 1.1

Corollary 1.1. *For the Polynomial(α) instance where $\Delta_i=\left(\frac{i}{n}\right)^\alpha$ with $\alpha > 0.5$, we have*

$$H_2(\epsilon)=\max_{i\geq g(\epsilon)+1}\frac{i}{\Delta_i^2}\cdot\frac{1}{g\left(\frac{\epsilon}{2}\right)}<\frac{4}{\epsilon^2}.$$

Proof.

$$\max_{i\geq g(\epsilon)+1}\frac{i}{\Delta_i^2}\cdot\frac{1}{g\left(\frac{\epsilon}{2}\right)}=\max_{i\geq g(\epsilon)+1} i^{1-2\alpha}n^{2\alpha}\frac{1}{n\left(\frac{\epsilon}{2}\right)^{1/\alpha}}<\left(n\epsilon^{1/\alpha}\right)^{1-2\alpha}n^{2\alpha}\frac{1}{n\left(\frac{\epsilon}{2}\right)^{1/\alpha}}<\frac{4}{\epsilon^2}.$$

where both the last inequality and the last equality are by $\alpha > 0.5$. □

B.7. Proof of Corollary 2.1

Corollary 2.1. *Katz-Samuels and Jamieson (2020) shows a lower bound of the sample complexity for identifying an ϵ-good arm that scales as*

$$\Omega\left(H^{low}(\epsilon) := \frac{1}{g(\epsilon)} \sum_{i=g(\epsilon)+1}^{n} \left(\frac{1}{\Delta_i^2}\right) - \frac{1}{\Delta_{g(\epsilon)+1}^2}\right).$$

In the case that $g(\epsilon)$ and $g\left(\frac{\epsilon}{2}\right)$ have the same order ($\frac{g\left(\frac{\epsilon}{2}\right)}{g(\epsilon)}$ is irrelevant to ϵ), the sample complexity measure $H_2(\epsilon)$ satisfies

$$H^{low}(\epsilon) \lesssim H_2(\epsilon) \lesssim H^{low}(\epsilon) + \frac{2}{\Delta_{g(\epsilon)+1}^2},$$

where $\lesssim$ hides logarithms of n and constants.

Proof.

$$\begin{aligned}
H^{\text{low}}(\epsilon) =& \frac{1}{g(\epsilon)} \sum_{i=g(\epsilon)+1}^{n} \frac{1}{\Delta_i^2} - \frac{1}{\Delta_{g(\epsilon)+1}^2} \\
=& \frac{1}{g(\epsilon)} \sum_{i=g(\epsilon)+1}^{n} \frac{1}{i}\frac{i}{\Delta_i^2} - \frac{1}{\Delta_{g(\epsilon)+1}^2} \\
\leq& \frac{1}{g(\epsilon)} \max_{i\geq g(\epsilon)+1} \frac{i}{\Delta_i^2} \cdot \sum_{i=g(\epsilon)+1}^{n} \frac{1}{i} - \frac{1}{\Delta_{g(\epsilon)+1}^2} \\
\overset{(a_1)}{\leq}& \frac{1}{g(\epsilon)} \max_{i\geq g(\epsilon)+1} \frac{i}{\Delta_i^2} \cdot \log(2n) - \frac{1}{\Delta_{g(\epsilon)+1}^2} \\
=& \frac{g\left(\frac{\epsilon}{2}\right)}{g(\epsilon)} H_2(\epsilon) \cdot \log(2n) - \frac{1}{\Delta_{g(\epsilon)+1}^2} \\
\leq& \frac{g\left(\frac{\epsilon}{2}\right)}{g(\epsilon)} H_2(\epsilon) \cdot \log(2n),
\end{aligned}$$

for (a_1), we use $\sum_{i=g(\epsilon)}^{n} \frac{1}{i} \leq \sum_{i\in[n]} \frac{1}{i} \leq \log(2n)$, which can be found in Audibert and Bubeck (2010). On the other hand,

$$\begin{aligned}
H^{\text{low}}(\epsilon) =& \frac{1}{g(\epsilon)} \sum_{i=g(\epsilon)+1}^{n} \frac{1}{\Delta_i^2} - \frac{1}{\Delta_{g(\epsilon)+1}^2} \\
\geq& \frac{1}{g(\epsilon)} \max_{i\geq g(\epsilon)+1} \frac{i-g(\epsilon)}{\Delta_i^2} - \frac{1}{\Delta_{g(\epsilon)+1}^2} \\
=& \frac{1}{g(\epsilon)} \left(\max_{i\geq g(\epsilon)+1} \frac{i}{\Delta_i^2} - \frac{g(\epsilon)}{\Delta_i^2}\right) - \frac{1}{\Delta_{g(\epsilon)+1}^2} \\
\geq& \frac{1}{g(\epsilon)} \left(\max_{i\geq g(\epsilon)+1} \frac{i}{\Delta_i^2} - \frac{g(\epsilon)}{\Delta_{g(\epsilon)+1}^2}\right) - \frac{1}{\Delta_{g(\epsilon)+1}^2} \\
=& \frac{1}{g(\epsilon)} \max_{i\geq g(\epsilon)+1} \frac{i}{\Delta_i^2} - \frac{2}{\Delta_{g(\epsilon)+1}^2} \\
=& \frac{g\left(\frac{\epsilon}{2}\right)}{g(\epsilon)} H_2(\epsilon) - \frac{2}{\Delta_{g(\epsilon)+1}^2}.
\end{aligned}$$

□

B.8. Proof of Theorem 3

For an n-armed bandit instance ν and a fixed value of $\epsilon > 0$, consider a partition of the arms of ν, $\mathcal{A} = \big\{[g(\epsilon)], [g(\epsilon)+1 : 2{\cdot}g(\epsilon)], \ldots, [n - g(\epsilon) + 1 : n]\big\}$, where we avoid rounding by making the assumption that n is an integer multiple of $g(\epsilon)$. We use $\mathcal{A}_i := [(i-1){\cdot}g(\epsilon) : i{\cdot}g(\epsilon)]$ to denote the subsets of the partition $\mathcal{A}$. Making use of this partition, we define

$$\tilde{H}(\nu, \epsilon) := \sum_{i=2}^{n/g(\epsilon)} \tilde{\Delta}_{\mathcal{A}_i}^{-2},$$

where for an arbitrary set of arms $A \subseteq [n]$ we define $\tilde{\Delta}_A := \max_{j \in A} \Delta_j$, the largest gap among the arms in $\mathcal{A}$.

Proof of Theorem 3. The proof structure follows Carpentier and Locatelli (2016, Theorem 1). Fix $n > 2$, $\epsilon > 0$ an n-armed unit-variance gaussian instance ν and an arbitrary policy π. Let $\mathbb{P}_{\nu,\pi}$ denote the probability measure over interaction histories induced by the interaction of π with instance ν. Without loss of generality, we take the arm means of ν to be ordered such that $\mu_1 \geq \mu_2 \geq \ldots \geq \mu_n$. Understanding $g(\epsilon)$ to refer to the number of ϵ-good arms on ν, we define instances ν^i for $i \in [2 : n/g(\epsilon)]$ from ν by setting $\nu^i_j = \nu_j$ for $j \notin \mathcal{A}_i$, and setting $\mu^i_j := \mu_j + 2\Delta_j$ for $j \in \mathcal{A}_i$. Each instance ν^i has $g(\epsilon)$ ϵ-good arms with indices in $\mathcal{A}_i$, and $\tilde{H}(\nu^i, \epsilon) \leq \tilde{H}(\nu, \epsilon) = a$ for all $i \in [2 : n/g(\epsilon)]$.

We will make use of a change of measure between ν and these alternate instances, restricted on an event on which the interaction histories on ν and ν^i are hard to distinguish from one another:

$$\mathcal{E}_i = \left\{ \sum_{t=1}^{T} \sum_{j=1}^{n} \mathbb{1}\{I_t = j\} \Big[\Lambda^i_t(j) - (1+\rho)\mathsf{KL}(\nu^i_j, \nu_j) \Big] \leq \frac{1}{\rho} \log(1/\delta) \right\},$$

where $\Lambda^i_t(j) = \log\left(\frac{\mathrm{d}\nu_j(X_t)}{\mathrm{d}\nu^i_j(X_t)} \right)$ for $i \in [2 : n/g(\epsilon)]$, and $\rho > 0$ is a parameter to be chosen later.

Lemma 8.3 (Jun and Zhang (2022, Lemma 10))**.** *For all $i \in [2 : n/g(\epsilon)]$ and arbitrary π, $\rho > 0$, and $\delta \in (0, 1)$,*

$$\mathbb{P}_{\nu,\pi}(\mathcal{E}_i) \geq 1 - \delta.$$

We can now prove the result. Denoting the number of times arm $j \in [n]$ has been pulled in T rounds of interaction by $N_j(T)$, let $t_i = \mathbb{E}_{\nu,\pi}\left[\sum_{j \in \mathcal{A}_i} N_j(T)\right]$ be the expected number of pulls that arm group $\mathcal{A}_i$ receives, for $i \in [2 : n/g(\epsilon)]$. By Markov's inequality, for all $i \in [2 : n/g(\epsilon)]$ we have

$$\mathbb{P}_{\nu,\pi}\left(\sum_{j \in \mathcal{A}_i} N_j(T) \geq 6t_i \right) \leq 1/6.$$

Defining the event

$$E_i = \{J_t \in \mathcal{A}_1\} \cap \{\mathcal{E}_i\} \cap \left\{ \sum_{j \in \mathcal{A}_i} N_j(T) \leq 6t_i \right\}$$

for $i \in [2 : n/g(\epsilon)]$ and setting $\delta = 1/6$ in Lemma 8.3, for all $i \in [2 : n/g(\epsilon)]$ and arbitrary π

$$\mathbb{P}_{\nu,\pi}(E_i) \geq 1 - (1/6 + 1/6 + \xi) = 2/3 - \xi,$$

where $\xi := \mathbb{P}_{\nu,\pi}(J_t \notin \mathcal{A}_1)$.

It follows that for $i \in [2 : n/g(\epsilon)]$

$$\begin{aligned} \mathbb{P}_{\nu^i,\pi}(J_t \notin \mathcal{A}_i) &\geq \mathbb{P}_{\nu^i,\pi}(J_t \in \mathcal{A}_1) \\ &\geq \mathbb{P}_{\nu^i,\pi}(E_i). \end{aligned}$$

Making use of a standard change of measure technique for bandits (Audibert et al., 2010), we have

$$\begin{aligned}
\mathbb{P}_{\nu^i,\pi}(E_i) &= \mathbb{E}_{\nu,\pi}\left[\mathbb{1}\{E_i\}\exp\left(-\sum_{t=1}^{T}\sum_{j\in\mathcal{A}_i}\mathbb{1}\{I_t=a\}\Lambda_t^i(j)\right)\right] \\
&\geq \mathbb{E}_{\nu,\pi}\left[\mathbb{1}\{E_i\}\exp\left(-\sum_{t=1}^{T}\sum_{j\in\mathcal{A}_i}\mathbb{1}\{I_t=j\}\,2\mathsf{KL}(\nu_j^i,\nu_j)-\log(6)\right)\right] && \text{(on } E_i\text{, taking } \rho=1 \text{ and } \delta=1/6) \\
&\geq \mathbb{E}_{\nu,\pi}\left[\mathbb{1}\{E_i\}\exp\left(-\sum_{j\in\mathcal{A}_i}4N_j(T)\cdot\max_{k\in\mathcal{A}_i}\Delta_k^2-\log(6)\right)\right] && \text{(evaluating the KL-divergence)} \\
&= \frac{1}{6}\,\mathbb{E}_{\nu,\pi}\left[\mathbb{1}\{E_i\}\exp\left(-\sum_{a\in\mathcal{A}_i}4N_j(T)\tilde{\Delta}_{\mathcal{A}_i}^2\right)\right] && \text{(definition of } \tilde{\Delta}_{\mathcal{A}_i}) \\
&\geq \frac{1}{6}\,\mathbb{E}_{\nu,\pi}\left[\mathbb{1}\{E_i\}\exp\left(-24t_i\tilde{\Delta}_{\mathcal{A}_i}^2\right)\right] && \text{(on } E_i) \\
&= \frac{1}{6}\exp\left(-24t_i\tilde{\Delta}_{\mathcal{A}_i}^2\right)\mathbb{P}_{\nu,\pi}(E_i).
\end{aligned}$$

If $\xi \geq 1/2$, the result follows. Otherwise if $\xi \leq 1/2$, then $\mathbb{P}_\nu(E_i) \geq 1/6$ and it follows that $\mathbb{P}_{\nu^i}(J_t \notin \mathcal{A}_i) \geq \frac{1}{36}\exp\left(-24t_i\tilde{\Delta}_{\mathcal{A}_i}^2\right)$.

Observing that $\tilde{H}(\epsilon,\nu) = \sum_{i=2}^{n/g(\epsilon)}\tilde{\Delta}_{\mathcal{A}_i}^{-2}$, and $\sum_{i=1}^{n/m} t_i = T$, there exists some $i \in [2 : n/g(\epsilon)]$ such that

$$t_i \leq \frac{T}{\tilde{H}(\epsilon,\nu)\tilde{\Delta}_{\mathcal{A}_i}^2} = \frac{T}{a\tilde{\Delta_{\mathcal{A}_i}^2}},$$

which can easily be shown by way of contradiction. For this i, by our display above, we have

$$\max\left\{\mathbb{P}_{\nu^i,\pi}(J_t\notin\mathcal{A}_i),\mathbb{P}_{\nu,\pi}(J_t\notin\mathcal{A}_1)\right\} \geq \max\left\{1/2,\frac{1}{36}\exp\left(-24T/a\right)\right\}.$$

□

B.9. Proof of Lemma 3.1

Proof of Lemma 3.1. To see this, for a given value of n and $\alpha > 1/2$, fix $\epsilon \leq 1/2$. We have

$$\begin{aligned}
\tilde{H}(\epsilon) &= \sum_{i=2}^{n/g(\epsilon)}\tilde{\Delta}_{\mathcal{A}_i}^{-2} \\
&= \sum_{i=2}^{n/g(\epsilon)}\left(\frac{n}{i\cdot g(\epsilon)}\right)^{2\alpha} && \text{(Polynomial}(\alpha)\text{ instance)} \\
&\geq \left(\frac{n}{g(\epsilon)}\right)^{2\alpha}\left[\sum_{i=2}^{n/g(\epsilon)}\int_i^{i+1}x^{-2\alpha}dx\right] && (x^{-2\alpha} \text{ decreases monotonically on } x>0\text{, for } \alpha>0) \\
&= \left(\frac{n}{g(\epsilon)}\right)^{2\alpha}\int_2^{n/g(\epsilon)+1}x^{-2\alpha}dx \\
&\geq \left(\frac{n}{g(\epsilon)}\right)^{2\alpha}\int_2^{\epsilon^{-1/\alpha}+1}x^{-2\alpha}dx && \left(\left(\tfrac{g(\epsilon)}{n}\right)^{\alpha}<\epsilon\text{, by definition of } g(\epsilon).\right) \\
&= \frac{1}{\epsilon^2}\frac{2^{-2\alpha+1}-\left(\epsilon^{-1/\alpha}+1\right)^{-2\alpha+1}}{2\alpha-1} \\
&\geq \frac{1}{\epsilon^2}\left[\frac{2^{-2\alpha+1}-\left(2^{1/\alpha}+1\right)^{-2\alpha+1}}{2\alpha-1}\right]. && \text{(assuming } \epsilon<1/2)
\end{aligned}$$

□

B.10. Analysis of the uniform sampling

Uniform sampling means the algorithm that takes a budget T as the input and equally allocates the budget to all the arms. Define $\mathrm{Top}_m(\epsilon) := \{i : \mu_i \geq \mu_m - \epsilon\}$ and a shortcut of $\mathrm{Top}_m := \mathrm{Top}_m(0)$.

Proposition 9. *Suppose we run the uniform sampling on a n-armed bandit with a budget T and output the s arms with the largest empirical rewards. Let $C_T(s, m', \epsilon)$ be the (m', ϵ)-good arms of the output. Then,*

$$\mathbb{P}\Big(\big|C_T(s,m',\epsilon)\big| \leq k\Big) \leq \binom{m'}{m'-k}\exp\left(-(m'-k)\frac{\epsilon^2 T}{8n}\right) + \binom{n-\big|\mathrm{Top}_{m'}(\epsilon)\big|}{s-k}\exp\left(-(s-k)\frac{\epsilon^2 T}{8n}\right).$$

This proposition is where our key contribution is rooted at. We show that our bound is tight for the uniform sampling in Section B.11.

Proof. **Step 1**
We claim that the intersection of the two conditions below is a sufficient condition for $\big|C_T(s,m',\epsilon)\big| > k$.

- $\left|\left\{i \in \mathrm{Top}_{m'} : \hat{\mu}_i > \mu_i - \frac{\epsilon}{2}\right\}\right| \geq k+1.$

- $\left|\left\{i \in \mathrm{Top}^c_{m'}(\epsilon) : \hat{\mu}_i < \mu_i + \frac{\epsilon}{2}\right\}\right| \geq \big|\mathrm{Top}^c_{m'}(\epsilon)\big| - s + k + 1.$

If all arms in $\left\{i \in \mathrm{Top}_{m'} : \hat{\mu}_i > \mu_i - \frac{\epsilon}{2}\right\}$ are included in the output, the claim naturally holds. Thus we focus on the situation where at least one of the arms in $\left\{i \in \mathrm{Top}_{m'} : \hat{\mu}_i > \mu_i - \frac{\epsilon}{2}\right\}$ is not included in the output. We prove the claim by contradiction. Suppose $\big|C_T(s,m',\epsilon)\big| \leq k$. Then the output includes no less than $s-k$ non-(m',ϵ)-good arms. Since

$$s - k + \big|\mathrm{Top}^c_{m'}(\epsilon)\big| - s + k + 1 = \big|\mathrm{Top}^c_{m'}(\epsilon)\big| + 1 > \big|\mathrm{Top}^c_{m'}(\epsilon)\big|,$$

at least one arm in $\left\{i \in \mathrm{Top}^c_{m'}(\epsilon) : \hat{\mu}_i < \mu_i + \frac{\epsilon}{2}\right\}$ is included in the output. To be more specific, there must exist overlap between $\left\{i \in \mathrm{Top}^c_{m'}(\epsilon) : \hat{\mu}_i < \mu_i + \frac{\epsilon}{2}\right\}$ and the set of the non-(m',ϵ)-good arms included in the output; otherwise the union of these two sets is even larger than the set of all the non-(m',ϵ)-good arms we have in the entire arm set. Additionally, all arms in $\left\{i \in \mathrm{Top}^c_{m'}(\epsilon) : \hat{\mu}_i < \mu_i + \frac{\epsilon}{2}\right\}$ have less empirical rewards than any arms in $\left\{i \in \mathrm{Top}_{m'} : \hat{\mu}_i > \mu_i - \frac{\epsilon}{2}\right\}$. Thus the output includes all arms in $\left\{i \in \mathrm{Top}_{m'} : \hat{\mu}_i > \mu_i - \frac{\epsilon}{2}\right\}$ as well. This leads to the contradiction with our supposition that at least one of the arms in $\left\{i \in \mathrm{Top}_{m'} : \hat{\mu}_i > \mu_i - \frac{\epsilon}{2}\right\}$ are not included in the output.

Step 2

We apply union bound and Hoeffding's inequality for the event that at least one of the two conditions does not hold.

$$\begin{aligned}
\mathbb{P}\Big(\big|C_T(s,m',\epsilon)\big| \le k\Big) &\le \mathbb{P}\left(\left|\left\{i \in \mathrm{Top}_{m'} : \hat{\mu}_i > \mu_i - \frac{\epsilon}{2}\right\}\right| < k+1\right) \\
&\quad + \mathbb{P}\left(\left|\left\{i \in \mathrm{Top}^c_{m'}(\epsilon) : \hat{\mu}_i < \mu_i + \frac{\epsilon}{2}\right\}\right| < \left|\mathrm{Top}^c_{m'}(\epsilon)\right| - s + k + 1\right) \\
&= \mathbb{P}\left(\left|\left\{i \in \mathrm{Top}_{m'} : \hat{\mu}_i \le \mu_i - \frac{\epsilon}{2}\right\}\right| \ge m' - k\right) \\
&\quad + \mathbb{P}\left(\left|\left\{i \in \mathrm{Top}^c_{m'}(\epsilon) : \hat{\mu}_i \ge \mu_i + \frac{\epsilon}{2}\right\}\right| \ge s - k\right) \\
&\le \mathbb{P}\left(\exists \mathcal{A} \subset \mathrm{Top}_{m'}, \text{s.t. } |\mathcal{A}| = m' - k \text{ and } \forall i \in \mathcal{A}, \hat{\mu}_i \le \mu_i - \frac{\epsilon}{2}\right) \\
&\quad + \mathbb{P}\left(\exists \mathcal{A} \subset \mathrm{Top}^c_{m'}(\epsilon), \text{s.t. } |\mathcal{A}| = s - k \text{ and } \forall i \in \mathcal{A}, \hat{\mu}_i \ge \mu_i + \frac{\epsilon}{2}\right) \\
&\le \binom{m'}{m'-k} \exp\left(-(m'-k)\frac{\epsilon^2 T}{8n}\right) + \binom{n - \left|\mathrm{Top}_{m'}(\epsilon)\right|}{s-k} \exp\left(-(s-k)\frac{\epsilon^2 T}{8n}\right).
\end{aligned}$$

□

Lemma 9.1. *Suppose we run the uniform sampling on a n-armed bandit with a budget of T and output the s arms with the largest empirical rewards. Let $C_T(s,\epsilon)$ be the ϵ-good arms in the output. Assume $s > g(\epsilon)$ with $g(\epsilon)$ being the number of ϵ-good arms. Let q be a parameter satisfying $g(\epsilon) < q < s$ and $d_q = \mu_{g(\epsilon)} - \mu_{q+1}$, then*

$$\mathbb{P}\Big(\big|C_T(s,\epsilon)\big| \le g(\epsilon) - k\Big) \le \binom{g(\epsilon)}{k} \exp\left(-k\frac{d_q^2 T}{8n}\right) + \binom{n-q}{s-q+k} \exp\left(-(s-q+k)\frac{d_q^2 T}{8n}\right).$$

Especially for Sequential Halving where $s = \frac{n}{2}$, choosing $q = \frac{n}{4}$ we have

$$\mathbb{P}\left(\left|C_T\left(\frac{n}{2},\epsilon\right)\right| \le g(\epsilon) - k\right) \le \binom{g(\epsilon)}{k} \exp\left(-k\frac{d_q^2 T}{8n}\right) + \binom{\frac{3n}{4}}{\frac{n}{4}+k} \exp\left(-\left(\frac{n}{4}+k\right)\frac{d_q^2 T}{8n}\right).$$

Proof. The event

$$\Big\{\big|C_T(s,\epsilon)\big| \le g(\epsilon) - k\Big\}$$

implies that there are at least k ϵ-good arms eliminated. Since $g(\epsilon) < q < s$, the s output arms include at most $q-k$ arms that are from the best q arms $\{i \mid \mu_i \ge \mu_q\}$. Such that, the s output arms include at least $s-q+k$ arms that are from the worst $n-q$ arms $\{i \mid \mu_i < \mu_q\}$; otherwise, the number of outputs is even less than s. Denote the set of eliminated arms that are ϵ-good as A', and the set of output arms that are from the worst $n-q$ arms as B', we have $|A'| \ge k, |B'| \ge s-q+k$. We then have the existence of A, B satisfying

$$\Big\{\exists A \subset \big[g(\epsilon)\big], |A| = k, \exists B \subset [n] \setminus [q], |B| = s-q+k, \text{s.t., } \forall i \in A, j \in B, \hat{\mu}_i \le \hat{\mu}_j\Big\}.$$

The above implication can also be seen as a generalization of the technique of Karnin et al. (2013). We then have

$$\left\{\exists A \subset \big[g(\epsilon)\big], |A| = k, \exists B \subset [n] \setminus [q], |B| = s-q+k, \text{s.t., } \forall i \in A, j \in B, \hat{\mu}_i \le \mu_i - \frac{d_q}{2} \text{ or } \hat{\mu}_j \ge \mu_j + \frac{d_q}{2}\right\}.$$

By the distributive law of Boolean algebra, we have at least one of the following two holds,

$$\left\{\exists A \subset \big[g(\epsilon)\big], |A| = k, \text{s.t., } \forall i \in A, \hat{\mu}_i \le \mu_i - \frac{d_q}{2}\right\},$$

or

$$\left\{\exists B \subset [n] \setminus [q], |B| = s - q + k, \text{s.t., } j \in B, \hat{\mu}_j \geq \mu_j + \frac{d_q}{2}\right\}.$$

Thereby the probability of $\left\{\left|C_T(s,\epsilon)\right| \leq g(\epsilon) - k\right\}$ can be upper bounded as

$$\begin{aligned}
&\mathbb{P}\Big(\left|C_T(s,\epsilon)\right| \leq g(\epsilon) - k\Big) \\
\leq &\mathbb{P}\left(\exists A \subset \left[g(\epsilon)\right], |A| = k, \text{s.t., } \forall i \in A, \hat{\mu}_i \leq \mu_i - \frac{d_q}{2}\right) \\
&+ \mathbb{P}\left(\exists B \subset [n] \setminus [q], |B| = s - q + k, \text{s.t., } j \in B, \hat{\mu}_j \geq \mu_j + \frac{d_q}{2}\right) \\
\leq &\binom{g(\epsilon)}{k} \exp\left(-k\frac{\left(\frac{d_q}{2}\right)^2 T}{2n}\right) + \binom{n-q}{s-q+k} \exp\left(-(s-q+k)\frac{\left(\frac{d_q}{2}\right)^2 T}{2n}\right) \\
= &\binom{g(\epsilon)}{k} \exp\left(-k\frac{d_q^2 T}{8n}\right) + \binom{n-q}{s-q+k} \exp\left(-(s-q+k)\frac{d_q^2 T}{8n}\right).
\end{aligned}$$

□

B.11. Lower bound for uniform sampling

Theorem 10. *Consider a family $\mathcal{E}(n,m,\epsilon)$ of all two-level univariate Gaussian bandit instances with n arms, of which m arms have mean $\epsilon > 0$ and all other arms have to mean 0. We consider the problem in which samples are drawn from some instance $\theta \in \mathcal{E}(n,m,\epsilon)$ in an off-line, uniform fashion with B samples from each arm (i.e., with a total sampling budget of T samples, $B = T/n$ ignoring integer effects) and a subset of arms $\hat{S} \subseteq [n]$ of size s is then chosen based only on the observed samples (i.e., without knowledge of θ) in a symmetric fashion. Here 'symmetric' is taken to mean that for any subset $S \subseteq [n]$, $\mathbb{P}_\theta(\hat{S} = S) = \mathbb{P}_{\sigma(\theta)}(\hat{S} = \sigma(S))$ where $\sigma \in \Sigma_n$ is an element of the permutation group on sets of size n, and $\sigma(S)$ is taken to mean the element-wise application of σ to the elements of S.[1] That is to say that the choice of $\hat{S}$ is independent of the specific indices of the arms chosen.*

Suppose $m \leq s \wedge n/3$ and $s < \frac{6}{11}n$. Let $\tilde{M}_\theta$ be the number of false negatives (i.e. $\| \geq$) Then, there exists $\theta \in \mathcal{E}(n,m,\epsilon)$ s.t.

$$\mathbb{P}_\theta(\tilde{M}_\theta \geq \frac{1}{4}m) \geq \min\left(\frac{1}{2},\ 2\left(\frac{6}{5}\cdot\frac{n-s}{s}\right)^{\frac{3}{16}m} \exp\left(-\left(1 + 16\cdot\frac{\ln(en/m)}{\ln\left(\frac{6}{5}\cdot\frac{n-s}{s}\right)}\right)mB\epsilon^2\right)\right)$$

where $\mathbb{P}_\theta$ is the probability measure induced by sampling from instance θ.

Proof. Let $\theta = (\epsilon, \ldots, \epsilon, 0, \ldots, 0) \in \mathbb{R}^n$ with $\epsilon > 0$ be a vector of mean rewards for each arm where the first m coordinates are nonzero. Let $\hat{S} \in [n]$ be the output of the algorithm (recall $|\hat{S}| = s$). Let $\tilde{M}_\theta$ be the number of false negatives when taking $\hat{S}$ as the prediction for the true support $[s]$ of θ. Let $\tilde{Q}_a \subset 2^{[n]}$ be the collection of all subsets of $[n]$ of size s such that $\tilde{M}_\theta = a$. For convenience, let $\gamma = \frac{1}{8}m$. Let $k = \frac{3}{4}s$.

$$\begin{aligned}
\xi := \mathbb{P}_\theta(\tilde{M}_\theta \geq m - k) &= \sum_{a=m-k}^{k} \mathbb{P}_\theta(\tilde{M}_\theta = a) \\
&\geq \sum_{a=3\gamma}^{5\gamma} \mathbb{P}_\theta(\tilde{M}_\theta = a).
\end{aligned}$$

[1] Note that the assumption of a symmetric selection $\hat{S}$ is equivalent to considering a lower bound for general selections $\hat{S}$ in the face of a uniform mixture over all permutations of θ (see Simchowitz et al. (2017) for a more in-depth discussion)

Let $\mathsf{first}(\tilde{Q}_a)$ be the first member of $\tilde{Q}_a$ in lexicographic order. For example, $\mathsf{first}(tQ_a) = [a+1 : a+s]$ where $[A : B] := \{A, A+1, \dots, B\}$. Then, by symmetry, one can see that $\mathbb{P}_\theta(\tilde{M}_\theta = a) = |\tilde{Q}_a| \mathbb{P}_\theta(\hat{S} = \mathsf{first}(\tilde{Q}_a))$.

We consider the event on which the 'empirical KL-divergence' concentrates, which can be shown to be true with probability at least $1-\delta$ by Jun and Zhang (2020, Lemma 5):

$$\mathsf{conc}(\theta', \theta) := \left\{ \sum_{i=1}^{n} \sum_{t=1}^{B} \ln\left(\frac{p_{\theta'}(X_{i,t}|A_t = i)}{p_\theta(X_{i,t}|A_t = i)} \right) - (1+\rho) B \sum_{i=1}^{n} \mathsf{KL}(\theta'_i, \theta_i) \le \frac{1}{\rho} \ln(1/\delta) \right\}.$$

We now employ a change of measure argument to $\mathbb{P}_\theta(\hat{S} = \mathsf{first}(\tilde{Q}_a))$ and switch θ with θ' that would result in $\tilde{M}_{\theta'} = a - 3\gamma =: b$. For $a \in [3\gamma : 5\gamma]$, this can be achieved with the choice of $\theta' = (0, \dots, 0, \epsilon, \dots, \epsilon, 0, \dots, 0)$ that is supported on $[3\gamma+1 : m+3\gamma]$. Let Θ be the set of permutations of θ. Then,

$$\begin{aligned}
&\mathbb{P}_\theta(\tilde{M}_\theta = a) \\
&= |\tilde{Q}_a| \mathbb{P}_\theta(\hat{S} = \mathsf{first}(\tilde{Q}_a)) \\
&\ge |\tilde{Q}_a| \mathbb{P}_{\theta'}(\hat{S} = \mathsf{first}(\tilde{Q}_a), \mathsf{conc}(\theta', \theta)) \exp(-(1+\rho) 2mB \cdot (\epsilon^2/2) - \rho^{-1} \ln(1/\delta)) && \text{(change of measure; } \rho > 0) \\
&\ge |\tilde{Q}_a| \mathbb{P}_{\theta'}(\hat{S} = \mathsf{first}(\tilde{Q}_a), \cap_{\sigma \in \Sigma} \mathsf{conc}(\theta', \sigma(\theta'))) \exp(-(1+\rho) mB\epsilon^2 - \rho^{-1} \ln(1/\delta)) && (\Sigma\text{: symmetric group of } [n]) \\
&\ge |\tilde{Q}_a| \mathbb{P}_\theta(\hat{S} = \mathsf{first}(\tilde{Q}_b), \cap_{\sigma \in \Sigma} \mathsf{conc}(\theta, \sigma(\theta))) \exp(-(1+\rho) mB\epsilon^2 - \rho^{-1} \ln(1/\delta)) && \text{(symmetry)} \\
&= \frac{|\tilde{Q}_a|}{|\tilde{Q}_b|} \mathbb{P}_\theta(\tilde{M}_\theta = b, \cap_{\sigma \in \Sigma} \mathsf{conc}(\theta, \sigma(\theta))) \exp(-(1+\rho) mB\epsilon^2 - \rho^{-1} \ln(1/\delta)) && \text{(symmetry)} \\
&\ge \frac{|\tilde{Q}_a|}{|\tilde{Q}_b|} \Big(\mathbb{P}_\theta(\tilde{M}_\theta = b) - |\Theta|\delta \Big) \exp\Big(-(1+\rho) mB\epsilon^2 - \rho^{-1} \ln(1/\delta) \Big) \ . \\
& && (\mathbb{P}(A, B) \ge \mathbb{P}(A) - \mathbb{P}(B^c)\text{; union bound over } \{\sigma(\theta) : \sigma \in \Sigma\})
\end{aligned}$$

Thus,

$$\begin{aligned}
&\sum_{a=3\gamma}^{5\gamma} \mathbb{P}_\theta(\tilde{M}_\theta = a) \\
&\ge \underbrace{\min_{a \in [3\gamma : 5\gamma]} \frac{|\tilde{Q}_a|}{|\tilde{Q}_{a-3\gamma}|}}_{=: Y} \left(\underbrace{\mathbb{P}_\theta(\tilde{M}_\theta \in [0 : 2\gamma])}_{\ge 1-\xi} - (2\gamma+1)|\Theta|\delta \right) \exp\Big(-(1+\rho) mB\epsilon^2 - \rho^{-1} \ln(1/\delta) \Big) \ .
\end{aligned}$$

Let us choose $\delta = \frac{1}{2} \cdot \frac{1-\xi}{(2\gamma+1)|\Theta|}$. Since the LHS above is at most ξ, we have

$$\xi \ge Y \frac{1-\xi}{2} \exp\left(-(1+\rho) mB\epsilon^2 - \rho^{-1} \ln\left(\frac{2(2\gamma+1)|\Theta|}{1-\xi} \right) \right)$$

One can consider two cases, namely $\xi \ge \frac{1}{2}$ and $\xi < \frac{1}{2}$, to arrive at

$$\xi \ge \min\left(\frac{1}{2}, \ \exp\Big(-(1+\rho) mB\epsilon^2 - \rho^{-1} \ln\big(4(2\gamma+1)|\Theta|\big) + \ln(4Y) \Big) \right)$$

It remains to find an appropriate value of ρ. One simple choice is

$$\rho^{-1} = \frac{1}{2} \frac{\ln(4Y)}{\ln\big(4(2\gamma+1)|\Theta|\big)} \ ,$$

which satisfies that $\rho^{-1} > 0$ as show later. Thus, we have

$$\xi \ge \min\left(\frac{1}{2}, \ 2\sqrt{Y} \exp\left(-\left(1 + 2 \frac{\ln\big(4(2\gamma+1)|\Theta|\big)}{\ln(4Y)} \right) mB\epsilon^2 \right) \right)$$

It remains to figure out bounds for Y and $|\Theta|$. For Y, note that $|\tilde{Q}_a| = \binom{m}{m-a}\binom{n-m}{s-(m-a)}$. So, for $a \in [3\gamma : 5\gamma]$ and $b = a - 3\gamma$,

$$\begin{aligned}
\min_a \frac{|\tilde{Q}_a|}{|\tilde{Q}_b|} &= \frac{\binom{m}{m-a}\binom{n-m}{s-m+a}}{\binom{m}{m-b}\binom{n-m}{s-m+b}} \\
&= \frac{(m-b)(m-b-1)\cdots(m-a+1)}{a(a-1)\cdots(b+1)} \cdot \frac{(n-s-b)(n-s-b-1)\cdots(n-s-a+1)}{(s-m+a)(s-m+a-1)\cdots(s-m+b+1)} \\
&\overset{(a)}{\geq} \left(\frac{m-b}{a}\right)^{a-b} \cdot \left(\frac{n-s-b}{s-m+a}\right)^{a-b} \\
&\geq \left(\frac{6}{5}\right)^{3\gamma} \cdot \left(\frac{n-s-2\gamma}{s-3\gamma}\right)^{3\gamma} \geq \left(\frac{6}{5}\right)^{3\gamma} \cdot \left(\frac{n-s}{s}\right)^{3\gamma} \\
\implies Y &= \min_{a\in[3\gamma:5\gamma]} \frac{|\tilde{Q}_a|}{|\tilde{Q}_{a-2\gamma}|} \geq \left(\frac{6}{5}\right)^{3\gamma} \cdot \left(\frac{n-s}{s}\right)^{3\gamma}
\end{aligned}$$

where (a) is due to the fact that $\frac{m-b}{a} > 1$ implies $(m-b-i)/(a-i) \geq (m-b)/a$ for $i \in [0 : a-b-1]$ and, with a similar reasoning, $n \geq s/2 \implies \frac{n-s-a+3\gamma}{s-m+a} > 1 \implies (n-s-b-i)/(s-m+a-i) \geq (n-s-b)/(s-m+a)$. Then, using $|\Theta| = \binom{n}{m} \leq \left(\frac{en}{m}\right)^m$, we have

$$\begin{aligned}
\rho = 2\frac{\ln\left(4(2\gamma+1)|\Theta|\right)}{\ln(4Y)} &\leq 2\frac{\ln(m+4) + m\ln\left(\frac{en}{m}\right)}{\ln(4) + \frac{m}{4}\ln\left(\frac{6}{5}\cdot\frac{n-s}{s}\right)} \\
&\overset{(a)}{\leq} 4\frac{m\ln\left(\frac{en}{m}\right)}{\ln(4) + \frac{m}{4}\ln\left(\frac{6}{5}\cdot\frac{n-s}{s}\right)} \\
&\leq 16\frac{\ln(en/m)}{\ln\left(\frac{6}{5}\cdot\frac{n-s}{s}\right)}
\end{aligned}$$

where (a) is by $m \leq n/3 \implies \ln(m+4) \leq m\ln(en/m)$.

Altogether,

$$\mathbb{P}_\theta(\tilde{M} \geq \frac{1}{4}m) \geq \min\left(\frac{1}{2},\ 2\left(\frac{6}{5}\cdot\frac{n-s}{s}\right)^{\frac{3}{16}m} \exp\left(-\left(1 + 16\cdot\frac{\ln(en/m)}{\ln\left(\frac{6}{5}\cdot\frac{n-s}{s}\right)}\right)mB\epsilon^2\right)\right)$$

To verify that our choice of ρ is nonnegative, it suffices to show that $\frac{6}{5}\cdot\frac{n-s}{s} > 1$, which is true for $s < \frac{6}{11}n$.

□

B.12. A minimax optimal budget allocation scheme for SH

We show that with a different budget allocation scheme $T_\ell = \left\lfloor \frac{T}{81n} \cdot \left(\frac{16}{9}\right)^{\ell-1} \cdot \ell \right\rfloor$, SH can achieve a sample complexity of $\mathcal{O}\left(\frac{n}{\epsilon^2}\log\frac{1}{\delta}\right)$ (without any logarithmic terms of n). We call this budget allocation scheme Option 2 and the original one of Karnin et al. (2013) Option 1. The price of achieving this exact minimax optimality is the potential loss of the near instance optimality (or at least the standard proof technique does not work). We summarize the comparison in Table 2. It is not clear to us whether there exists a different value of T_ℓ in SH that can achieve both the minimax and instance-dependent optimality. It is also not clear whether Option 2 can achieve a similar bound as Theorem 1. Note that for the fixed budget setting, there is still a $\log\log(n)$ gap between the instance-dependent upper and lower bounds.

Theorem 11. *For any $\epsilon \in (0,1)$, with Option 2, there exists an absolute constant c_1, such that the error probability of SH for identifying an ϵ-good arm satisfies,*

$$\mathbb{P}\left(\mu_{J_T} < \mu_1 - \epsilon\right) \leq \exp\left(-\Theta\left(\frac{\epsilon^2 T}{n}\right)\right),$$

for $T > c_1 \cdot \frac{n}{\epsilon^2}$.

Proof. Still let $\epsilon_1 = \epsilon/4$. And define $\epsilon_{\ell+1} = \frac{3}{4} \cdot \epsilon_\ell$. For each stage ℓ, define the event G_ℓ as

$$G_\ell := \left\{ \max_{i \in S_{\ell+1}} \mu_i \geq \max_{i \in S_\ell} \mu_i - \epsilon_\ell \right\}.$$

We have

$$\sum_{\ell=1}^{\log n} \epsilon_\ell < \sum_{\ell=1}^{\infty} \left(\frac{3}{4}\right)^{\ell-1} \cdot \epsilon_1 = \frac{\epsilon}{4} \sum_{\ell=1}^{\infty} \left(\frac{3}{4}\right)^{\ell-1} \leq \frac{\epsilon}{4} \lim_{n\to\infty} \frac{1-(3/4)^n}{1-3/4} = \epsilon.$$

For stage ℓ, we assign budget $T_\ell \cdot n2^{-(\ell-1)} := \frac{T}{81n} \cdot \left(\frac{16}{9}\right)^{\ell-1} \cdot \ell \cdot n2^{-(\ell-1)} = \frac{T}{81} \cdot \left(\frac{8}{9}\right)^{\ell-1} \cdot \ell$. We can verify this allocation scheme does not exceed the budget as follows,

$$\sum_{\ell=1}^{\log n} T_\ell \cdot n2^{-(\ell-1)} \leq \sum_{\ell=1}^{\infty} \frac{T}{81} \cdot \left(\frac{8}{9}\right)^{\ell-1} \cdot \ell \leq \frac{T}{9} \lim_{\ell\to\infty} \frac{1-\left(8/9\right)^\ell}{1-8/9} = T.$$

Let a_ℓ be the best arm in S_ℓ,

$$\begin{aligned}
\mathbb{P}\left(G_\ell^c\right) &= \mathbb{P}\left(G_\ell^c, \hat{\mu}_{a_\ell} < \mu_{a_\ell} - \epsilon_\ell/2\right) + \mathbb{P}\left(G_\ell^c, \hat{\mu}_{a_\ell} \geq \mu_{a_\ell} - \epsilon_\ell/2\right) \\
&\leq \mathbb{P}\left(\hat{\mu}_{a_\ell} < \mu_{a_\ell} - \epsilon_\ell/2\right) + \mathbb{P}\left(G_\ell^c \mid \hat{\mu}_{a_\ell} \geq \mu_{a_\ell} - \epsilon_\ell/2\right).
\end{aligned}$$

For the first term

$$\begin{aligned}
\mathbb{P}\left(\hat{\mu}_{a_\ell} < \mu_{a_\ell} - \epsilon_\ell/2\right) &\leq \exp\left(-\frac{\epsilon_\ell^2}{2} \frac{T_\ell}{|S_\ell|}\right) \\
&\leq \exp\left(-\frac{\epsilon_1^2 \left(\frac{9}{16}\right)^{\ell-1}}{2} \frac{\frac{T}{81n} \cdot \left(\frac{16}{9}\right)^{\ell-1} \cdot \ell \cdot n2^{-(\ell-1)}}{n2^{-(\ell-1)}}\right) \\
&\leq \exp\left(-\frac{\epsilon^2 T}{32 \cdot 81n} \ell\right).
\end{aligned}$$

For the second term,

$$\begin{aligned}
\mathbb{P}\left(G_\ell^c \mid \hat{\mu}_{a_\ell} \geq \mu_{a_\ell} - \epsilon_\ell/2\right) &\leq \mathbb{P}\left(\left|\{i \in S_\ell \mid \hat{\mu}_i > \mu_i + \epsilon_\ell/2\}\right| \geq |S_\ell|/2\right) \\
&\leq \frac{\mathbb{E}\left[\left|\{i \in S_\ell \mid \hat{\mu}_i > \mu_i + \epsilon_\ell/2\}\right|\right]}{|S_\ell|/2} \\
&\leq \frac{|S_\ell| \exp\left(-\frac{\epsilon^2 T}{32 \cdot 81n} \ell\right)}{|S_\ell|/2} \\
&= 2\exp\left(-\frac{\epsilon^2 T}{32 \cdot 81n} \ell\right).
\end{aligned}$$

By contraposition,

$$\begin{aligned}
\mathbb{P}\left(\mu_{J_T} < \mu_1 - \epsilon\right) &\leq \sum_{\ell=1}^{\log n} \mathbb{P}\left(G_\ell^c\right) \\
&\leq \sum_{\ell=1}^{\infty} 3\exp\left(-\frac{\epsilon^2 T}{32 \cdot 81n} \ell\right).
\end{aligned}$$

Let $X := \frac{\epsilon^2 T}{32 \cdot 81n} > 0$,

$$\begin{aligned}
\mathbb{P}\left(\mu_{J_T} < \mu_1 - \epsilon\right) &\leq \sum_{\ell=1}^{\infty} 3\exp(-\ell X) \\
&\leq 3\exp(-X) \frac{1-\exp(-\ell X)}{1-\exp(-X)} \\
&\leq 3\exp(-X) \frac{1}{1-\exp(-X)}.
\end{aligned}$$

Table 2. The ϵ-error probability of SH (worst-case) and $(\epsilon = 0)$-error probability (instance-dependent) for Option 1 and 2 where $H_2 := \max_i \frac{i}{\Delta_i^2}$ is the problem hardness parameter. Here, we omit constant factors. Check marks for the worst-case means that it is optimal, and those for the instance-dependent means that it is optimal up to $\log\log(n)$ factors in the sample complexity (Mannor and Tsitsiklis, 2004; Carpentier and Locatelli, 2016).

	Worst-case	Instance-dependent
Option 1: $T_\ell = \left\lfloor \frac{T}{\lvert S_\ell \rvert \lceil \log_2 n \rceil} \right\rfloor$	✗ $\log_2 n \cdot \exp\left(-\frac{\epsilon^2 T}{n \log_2 n}\right)$	✓ $\log_2 n \cdot \exp\left(-\frac{T}{\log_2 n H_2}\right)$
Option 2: $T_\ell = \left\lfloor \frac{T}{81n} \cdot \left(\frac{16}{9}\right)^{\ell-1} \cdot \ell \right\rfloor$	✓ $\exp\left(-\frac{\epsilon^2 T}{n}\right)$	✗ $\log_2 n \cdot \exp\left(-\frac{T}{n \log_2 n H_2}\right)$

As $\frac{1}{1-\exp(-X)}$ is a monotonically decreasing function, there exists an absolute constant c_1, such that

$$\mathbb{P}\left(\mu_{J_T} < \mu_1 - \epsilon\right) \le \exp\left(-\Theta\left(\frac{\epsilon^2 T}{n}\right)\right),$$

holds for $T > c_1 \cdot \frac{n}{\epsilon^2}$. □

C. Simple Regret in the Data-Poor Regime

C.1. The number of opened brackets

Lemma 11.1. *The number of opened brackets till round t, denoted by L_t, satisfies*

$$0.63 \cdot \log_2\left(1 + \ln(2)t\right) < L_t \le 1 + \log_2\left(1 + \ln(2)t\right).$$

The size of the largest bracket satisfies

$$2^{L_t} \ge \tilde{\Theta}(t).$$

Proof. Recall that whenever $t \ge B2^B$ is true, we open a new bracket and increase B by 1. $B = 0$ before the game starts. First, note that for $t = 1$ we have $L_t = 1$. Then, let us focus on $t \ge 2$. It is not hard to verify that

$$L_t = \max\left\{B : t \ge (B-1)2^{B-1}\right\}.$$

Fix $t \ge 2$. Then, we can define

$$t' := (L_t - 1)2^{L_t - 1}, \tag{15}$$

the first time point that the bracket L_t was opened. Clearly $t' \le t$. Solve (15) for L_t.

$$\begin{aligned} t' &= (L_t - 1)\exp\left(\ln(2)(L_t - 1)\right), \\ \ln(2)t' &= \ln(2)(L_t - 1)\exp\left(\ln(2)(L_t - 1)\right) \\ &=: X\exp(X). \end{aligned}$$

Solving it for X is exactly the Lambert W function. By Lemma 17 of Orabona and Pal (2016), we have

$$0.6321 \cdot \ln\left(1 + \ln(2)t'\right) \le X \le \ln\left(1 + \ln(2)t'\right).$$

This implies that

$$X = \ln(2)(L_t - 1) \le \ln\left(1 + \ln(2)t'\right) \le \ln\left(1 + \ln(2)t\right) \implies L_t \le 1 + \log_2\left(1 + \ln(2)t\right).$$

To obtain a lower bound, define

$$t'' := L_t 2^{L_t},$$

which is strictly larger than t. Using similar techniques as above, we can derive

$$L_t \geq 0.6321 \cdot \log_2\big(1 + \ln(2)t''\big) > 0.63 \cdot \log_2\big(1 + \ln(2)t\big).$$

Together, we have

$$0.63 \cdot \log_2\big(1 + \ln(2)t\big) < L_t \leq 1 + \log_2\big(1 + \ln(2)t\big).$$

Note that we have

$$L_t 2^{L_t} \leq t \leq (L_t + 1)2^{L_t+1}.$$

Thus

$$2^{L_t} \geq \frac{t}{2(L_t+1)} \geq \frac{t}{2\Big(2 + \log_2\big(1 + \ln(2)t\big)\Big)} \geq \frac{t}{4\log_2 t} = \tilde{\Theta}(t). \qquad (t \geq 4)$$

□

C.2. The number of pulls for the representative arms

Lemma 11.2. *At round t, the number of times the representative arm of bracket B has been pulled in the latest finished SH, denoted as $D_{B,t}$, satisfies,*

$$D_{B,t} \geq \text{const} \cdot \frac{t}{\ln(t) \cdot \log_2(n)}.$$

Proof. We call the round interval $(L_t - 1)2^{L_t-1} \leq t < L_t 2^{L_t}$ as *phase* L_t, whose length is $L_t 2^{L_t} - (L_t - 1)2^{L_t-1} = (L_t + 1)2^{L_t-1}$. Correspondingly, sampling budget of $\frac{(L_t+1)2^{L_t-1}}{L_t}$ is assigned to each bracket. For each opened bracket $A_B, B \in [L_t - 1]$, the sampling budget for it is accumulated from phase B to phase L_t, while phase L_t is not finished yet. The sampling budget bracket A_B receives satisfies

$$\begin{aligned} T_{B,t} &= \sum_{i=B}^{L_t-1} \frac{2^{i-1}(i+1)}{i} + \frac{t - 2^{L_t-1}(L_t-1)}{L_t} \\ &> \frac{2^{L_t-2}L_t}{L_t - 1} + \frac{t - 2^{L_t-1}(L_t-1)}{L_t} \\ &> \frac{t}{4L_t} \\ &> \frac{t}{4 + 4\log\big(1 + \ln(2)t\big)} > \frac{t}{4\ln t}. \end{aligned} \qquad (t > 16)$$

Recall we run the SH with the doubling trick on each bracket individually. We claim that the latest finished SH receives a budget of at least $T_{B,t}/4$. To see this, note that we initialize Algorithm 3.1 with budget starting from $\lceil n \log_2 n \rceil$. Then restart the SH algorithm with budget $2 \cdot \lceil n \log_2 n \rceil$, and so on. Before finishing the k-th doubling trick, the best arm from $(k-1)$-th doubling trick is output. Thus the portion of the lasted finished SH ranges in

$$\left[\frac{|\mathcal{T}_{k-1}|}{\sum_{i=1}^{k}|\mathcal{T}_i|}, \frac{|\mathcal{T}_{k-1}|}{\sum_{i=1}^{k-1}|\mathcal{T}_i|}\right),$$

which, by some simple calculations, is

$$\left[\frac{b_{r-1}}{4b_{r-1} - b_0}, \frac{b_{r-1}}{2b_{r-1} - b_0}\right). \qquad (16)$$

That is to say, the most recently finished SH, which gives the current output, has a budget of at least $T_{B,t}/4$. We notice that $T_{B,t}$ is essentially irrelevant to B. This means each bracket receives an order-wise equal amount of budget to query the SH

algorithm. In the SH algorithm, the output arm is pulled a fixed number of times when the number of arms and the sampling budget is fixed. It is pulled $\Theta\left(\frac{T}{\log_2 n}\right)$ times for budget T. We have all we need to determine the lower bound for $D_{B,t}$.

$$D_{B,t} \geq \text{const} \cdot \frac{t}{\ln(t) \cdot \log_2(n)}.$$

□

C.3. Lower bounds and their implications in special instances

While Katz-Samuels and Jamieson (2020) shows a lower bound for the (ϵ, δ)-unverifiable sample complexity that can match the upper bound in special instances, the final result they presented has a negative term, which results in looseness, and the bound can even go to 0 in certain instances. Recall the definition of $\tau_{\epsilon,\delta}$, the (ϵ, δ)-unverifiable sample complexity (Definition 1).

Theorem 12 (Theorem 1 of Katz-Samuels and Jamieson (2020))**.** *Fix $\epsilon > 0, \delta \in (0, 1/16)$, and a vector $\mu \in \mathbb{R}^n$. Consider n arms where rewards from the i-th arm are distributed according to $\mathcal{N}(\mu_i, 1)$. 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 $\pi(\mu)$. Then the (ϵ, δ)-unverifiable sample complexity satisfies,*

$$\mathbb{E}\left[\tau_{\epsilon,\delta}\right] \geq \frac{1}{64}\left(-\left(\mu_1 - \mu_{g(\epsilon)+1}\right)^{-2} + \frac{1}{g(\epsilon)} \sum_{i=g(\epsilon)+1}^{n} (\mu_1 - \mu_i)^{-2}\right),$$

where the expectation is with respect to $\pi \in \mathbb{S}^n$ and $\pi(\mu)$.

In fact, the proof of Katz-Samuels and Jamieson (2020, Theorem 1) shows a stronger bound, which is more useful for deriving lower bounds for specific instances.

Theorem 13 (A stronger version of Theorem 1 of Katz-Samuels and Jamieson (2020))**.** *Fix $\epsilon > 0, \delta \in (0, 1/16)$, and a vector $\mu \in \mathbb{R}^n$. Consider n arms where rewards from the i-th arm are distributed according to $\mathcal{N}(\mu_i, 1)$. 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 $\pi(\mu)$. Then the (ϵ, δ)-unverifiable sample complexity satisfies,*

$$\mathbb{E}\left[\tau_{\epsilon,\delta}\right] \geq \frac{1}{64}\frac{1}{g(\epsilon)} \sum_{i=2g(\epsilon)+1}^{n} (\mu_1 - \mu_i)^{-2},$$

where the expectation is with respect to $\pi \in \mathbb{S}^n$ and $\pi(\mu)$.

Corollary 13.1. *For the EqualGap(m) instance, Proposition 13 shows a lower bound of the (ϵ, δ)-unverifiable sample complexity as*

$$\mathbb{E}[\tau_\pi] \geq \frac{1}{100}\left(\frac{n}{m}\frac{1}{\epsilon^2} - \frac{2}{\epsilon^2}\right).$$

Proof.

$$\begin{aligned}
\frac{1}{64}\frac{1}{g(\epsilon)} \sum_{i=2g(\epsilon)+1}^{n} (\mu_1 - \mu_i)^{-2} &= \frac{1}{64}\frac{1}{m} \sum_{i=2m+1}^{n} \frac{16}{25\epsilon^2} \\
&= \frac{1}{64}\frac{n-2m}{m}\frac{16}{25\epsilon^2} \\
&= \frac{1}{100}\left(\frac{n}{m}\frac{1}{\epsilon^2} - \frac{2}{\epsilon^2}\right).
\end{aligned}$$

□

Corollary 13.2. *For the Polynomial(α) instance, Proposition 13 shows a lower bound of the (ϵ, δ)-unverifiable sample complexity as*

$$\mathbb{E}[\tau_\pi] \geq \frac{1}{128}\frac{1}{2\alpha - 1}\left(2^{1-2\alpha}\epsilon^{-2} - \epsilon^{-\frac{1}{\alpha}}\right).$$

Proof. For the Polynomial(α) instance,

$$\begin{aligned}\frac{1}{64}\frac{1}{g(\epsilon)}\sum_{i=2g(\epsilon)+1}^{n}(\mu_1-\mu_i)^{-2} &= \frac{1}{64}\frac{1}{g(\epsilon)}\sum_{i=2g(\epsilon)+1}^{n}\left(\frac{i}{n}\right)^{-2\alpha}\\ &= \frac{1}{64}\frac{n^{2\alpha}}{n\epsilon^{1/\alpha}}\sum_{i=2g(\epsilon)+1}^{n}\left(\frac{1}{i}\right)^{2\alpha}\\ &= \frac{1}{64}\frac{n^{2\alpha-1}}{\epsilon^{1/\alpha}}\sum_{i=2g(\epsilon)+1}^{n}\left(\frac{1}{i}\right)^{2\alpha}.\end{aligned}$$

The summation of $\sum_{i=2g(\epsilon)+1}^{n}\left(\frac{1}{i}\right)^{2\alpha}$ can be bounded as

$$\begin{aligned}\sum_{i=2g(\epsilon)+1}^{n}\left(\frac{1}{i}\right)^{2\alpha} &\overset{(a_1)}{\geq} \frac{1}{2}\int_{2g(\epsilon)}^{n}\left(\frac{1}{i}\right)^{2\alpha}di\\ &= \frac{1}{2}\Big(f(n)-f\big(2g(\epsilon)\big)\Big) && \left(f(x)=\tfrac{1}{-2\alpha+1}x^{-2\alpha+1}\right)\\ &= \frac{1}{2}\left(\frac{1}{-2\alpha+1}n^{-2\alpha+1}-\frac{1}{-2\alpha+1}(2g(\epsilon))^{-2\alpha+1}\right)\\ &= \frac{1}{2}\left(\frac{1}{-2\alpha+1}n^{-2\alpha+1}-\frac{1}{-2\alpha+1}(2n\epsilon^{1/\alpha})^{-2\alpha+1}\right),\end{aligned}$$

where (a_1) is by the geometric meaning of integral and

$$\frac{1}{2}\left(\frac{1}{i}\right)^{2\alpha}=\left(\frac{1}{2^{\frac{1}{2\alpha}}i}\right)^{2\alpha}\leq\left(\frac{1}{2i}\right)^{2\alpha}\leq\left(\frac{1}{i+1}\right)^{2\alpha}. \qquad (i\geq 1,\alpha>0.5)$$

Thus

$$\begin{aligned}\frac{1}{64}\frac{1}{g(\epsilon)}\sum_{i=2g(\epsilon)+1}^{n}(\mu_1-\mu_i)^{-2} &= \frac{1}{64}\frac{n^{2\alpha-1}}{\epsilon^{1/\alpha}}\sum_{i=2g(\epsilon)+1}^{n}\left(\frac{1}{i}\right)^{2\alpha}\\ &\geq \frac{1}{128}\frac{n^{2\alpha-1}}{\epsilon^{1/\alpha}}\left(\frac{1}{-2\alpha+1}n^{-2\alpha+1}-\frac{1}{-2\alpha+1}(2n\epsilon^{1/\alpha})^{-2\alpha+1}\right)\\ &= \frac{1}{128}\frac{1}{\epsilon^{1/\alpha}}\left(\frac{1}{-2\alpha+1}-\frac{2^{1-2\alpha}}{-2\alpha+1}\epsilon^{\frac{-2\alpha+1}{\alpha}}\right)\\ &= \frac{1}{128}\epsilon^{-\frac{1}{\alpha}}\frac{1}{2\alpha-1}\left(2^{1-2\alpha}\epsilon^{\frac{-2\alpha+1}{\alpha}}-1\right)\\ &= \frac{1}{128}\frac{1}{2\alpha-1}\left(2^{1-2\alpha}\epsilon^{-2}-\epsilon^{-\frac{1}{\alpha}}\right).\end{aligned}$$

□

C.4. Proof of Theorem 6

Theorem 6. *For any $\epsilon\in(0,1)$, the ϵ-error probability of BSH satisfies*

$$\mathbb{P}\big(\mu_1-\mu_{J_t}>\epsilon\big)\leq\exp\left(-\tilde{\Theta}\left(\frac{t}{\frac{1}{\epsilon^2}+\max_{i\geq g(\frac{\epsilon}{2})+1}\frac{i}{\Delta_i^2}\cdot\frac{1}{g(\frac{\epsilon}{4})}}\right)\right),$$

for $t\geq\tilde{\Theta}\Big(H_2\big(\frac{\epsilon}{2}\big)\Big)$.

Proof. Due to the doubling nature of the bracketing scheme, it is not hard to see that each bracket receives an order-wise equal amount of sampling budget $\tilde{\Theta}(t)$ (details in Appendix C.2). Our analysis is centered around finding the ideal bracket whose representative arm is expected to be of the highest quality, which we call the *best bracket*. Intuitively, on the one hand, if the best bracket is too large, the per-arm sampling budget will be too small to guarantee a meaningful output. On the other hand, if the best bracket is too small, it may not well represent the entire instance. The core idea of the proof is balancing this trade-off.

We define the following notations.

- A^*: the best bracket, which is a multiset satisfying $\forall i \in A^*, i \in [n]$. We have $|A^*| = L$ with L being specified later.
- $r = \frac{L}{n}$: the subsampling ratio of the best bracket A^*.
- $\epsilon_k := 2^{k-1} \cdot \frac{\epsilon}{4}$, for $k \in [K]$, where $K = \left\lceil \log_2\left(\frac{8}{\epsilon}\right) \right\rceil$. Also K satisfies $\epsilon_{K-1} \le 1, \epsilon_K \ge 1$.
- $Z_k = \left\{ i \in [n] \mid \Delta_i \in \left[0, \min\{\epsilon_k, 1\}\right] \right\}$.
- $g(\epsilon_k) = |Z_k|$.
- $Z'_k = \left\{ i \in A^* \mid \Delta_i \in \left[0, \min\{\epsilon_k, 1\}\right] \right\}$.
- $g'(\epsilon_k) = \left|Z'_k\right|$.
- $\mu_i(A)$: the mean reward of the i-th best arm in A.
- $\Delta'_i = \max\left\{\mu_j(A^*) \mid j \in A^*\right\} - \mu_i(A^*)$: the suboptimality gap of i-th best arm in A^* with respect to the best arm in A^*.
- $\Delta''_i = \mu_1 - \mu_i(A^*)$: the suboptimality gap of i-th best arm in A^* with respect to the best arm in the entire instance.
- $U_1\left(\frac{\epsilon}{2}\right) = \max_{i \ge g(\epsilon_2)+1} \frac{i}{\Delta_i^2}$, and $U_2\left(\frac{\epsilon}{2}\right) = \max_{i \ge g'(\epsilon_2)+1} \frac{i}{\Delta_i'^2}$. We call this the un-accelerated sample complexity on the entire instance and the best bracket for identifying an $\frac{\epsilon}{2}$-good arm.

We define the following events:

- E_1: $\forall k \in [K], \left| \frac{g'(\epsilon_k)}{L} - \frac{g(\epsilon_k)}{n} \right| < \frac{g(\epsilon_k)}{2n}$.
- E_2: an $\frac{\epsilon}{2}$-good arm represents the best bracket.

The event E_1 ensures that the best bracket can well represent the entire arm set in the sense that, for each peeling piece k, the ratio of the arms therein to all the arms in the best bracket $\frac{g'(\epsilon_k)}{L}$ is close to its original ratio $\frac{g(\epsilon_k)}{n}$. The ϵ-error probability can now be bounded as

$$\begin{aligned}
\mathbb{P}\left(\mu_{J_t} < \mu_1 - \epsilon\right) &= \mathbb{P}\left(\mu_{J_t} < \mu_1 - \epsilon, E_1^c\right) + \mathbb{P}\left(\mu_{J_t} < \mu_1 - \epsilon, E_1, E_2^c\right) + \mathbb{P}\left(\mu_{J_t} < \mu_1 - \epsilon, E_1, E_2\right) \\
&\le \mathbb{P}\left(E_1^c\right) + \mathbb{P}\left(E_1, E_2^c\right) + \mathbb{P}\left(\mu_{J_t} < \mu_1 - \epsilon, E_2\right) \\
&\le \mathbb{P}\left(E_1^c\right) + \mathbb{P}\left(E_2^c \mid E_1\right) + \mathbb{P}\left(\mu_{J_t} < \mu_1 - \epsilon \mid E_2\right). \qquad (17)
\end{aligned}$$

We bound each term separately. For the first term.

$$\begin{aligned}
\mathbb{P}\left(E_1^c\right) &= \mathbb{P}\left(\exists k, \left|\frac{g'(\epsilon_k)}{L} - \frac{g(\epsilon_k)}{n}\right| > \frac{g(\epsilon_k)}{2n}\right) \\
&\leq \sum_{k=1}^{K} \mathbb{P}\left(\left|\frac{g'(\epsilon_k)}{L} - \frac{g(\epsilon_k)}{n}\right| > \frac{g(\epsilon_k)}{2n}\right) \\
&\leq \sum_{k=1}^{K} \mathbb{P}\left(\frac{g'(\epsilon_k)}{L} < \frac{g(\epsilon_k)}{2n} \text{ or } \frac{g'(\epsilon_k)}{L} > \frac{3g(\epsilon_k)}{2n}\right) \\
&\leq \sum_{k=1}^{K} \mathbb{P}\left(\frac{g'(\epsilon_k)}{L} < \frac{g(\epsilon_k)}{2n}\right) + \sum_{k=1}^{K} \mathbb{P}\left(\frac{g'(\epsilon_k)}{L} > \frac{3g(\epsilon_k)}{2n}\right).
\end{aligned}$$

Let $\mathsf{KL}(p, q)$ be the Kullback-Leibler divergence between two Bernoulli distributions with parameters p, q. The first summation can be upper bounded by the additive Chernoff bound for binomial distribution (Arratia and Gordon, 1989),

$$\begin{aligned}
\sum_{k=1}^{K} \mathbb{P}\left(\frac{g'(\epsilon_k)}{L} < \frac{g(\epsilon_k)}{2n}\right) &\leq \sum_{k=1}^{K} \exp\left(-L\mathsf{KL}\left(\frac{g(\epsilon_k)}{2n}, \frac{g(\epsilon_k)}{n}\right)\right) \\
&\leq \sum_{k=1}^{K} \exp\left(-\frac{1}{8} L \frac{\frac{g^2(\epsilon_k)}{n^2}}{\frac{g(\epsilon_k)}{n}}\right) && (\mathsf{KL}(p,q) \geq \tfrac{(p-q)^2}{2\max\{p,q\}}) \\
&= \sum_{k=1}^{K} \exp\left(-\frac{1}{8} L \frac{g(\epsilon_k)}{n}\right).
\end{aligned}$$

The second summation can be upper bounded by

$$\begin{aligned}
\sum_{k=1}^{K} \mathbb{P}\left(\frac{g'(\epsilon_k)}{L} > \frac{3g(\epsilon_k)}{2n}\right) &\leq \sum_{k=1}^{K} \exp\left(-L\mathsf{KL}\left(1 - \frac{3g(\epsilon_k)}{2n}, 1 - \frac{g(\epsilon_k)}{n}\right)\right) \\
&\overset{(a_1)}{\leq} \sum_{k=1}^{K} \exp\left(-L\mathsf{KL}\left(\frac{3g(\epsilon_k)}{2n}, \frac{g(\epsilon_k)}{n}\right)\right) \\
&\leq \sum_{k=1}^{K} \exp\left(-\frac{1}{12} L \frac{\frac{g^2(\epsilon_k)}{n^2}}{\frac{g(\epsilon_k)}{n}}\right) && (\mathsf{KL}(p,q) \geq \tfrac{(p-q)^2}{2\max\{p,q\}}) \\
&\leq \sum_{k=1}^{K} \exp\left(-\frac{1}{12} L \frac{g(\epsilon_k)}{n}\right),
\end{aligned}$$

where (a_1) is by the definition of KL divergence for Bernoulli distribution. Since $g(\epsilon_k)$ is a non-decreasing function of k, we have

$$\begin{aligned}
\mathbb{P}\left(E_1^c\right) &\leq \sum_{k=1}^{K} \exp\left(-\frac{1}{8} L \frac{g(\epsilon_k)}{n}\right) + \sum_{k=1}^{K} \exp\left(-\frac{1}{12} L \frac{g(\epsilon_k)}{n}\right) \\
&\leq 2\left\lceil \log_2\left(\frac{8}{\epsilon}\right)\right\rceil \exp\left(-\frac{1}{12} L \frac{g(\epsilon_1)}{n}\right).
\end{aligned}$$

We choose $L = \tilde{\Theta}\left(\frac{nt}{g(\epsilon_1) H_2\left(\frac{\epsilon}{2}\right)}\right)$. Then

$$\begin{aligned}
\mathbb{P}\left(E_1^c\right) &\leq 2\left\lceil \log_2\left(\frac{8}{\epsilon}\right)\right\rceil \exp\left(-\tilde{\Theta}\left(\frac{t}{H_2\left(\frac{\epsilon}{2}\right)}\right)\right) \\
&\leq \exp\left(-\tilde{\Theta}\left(\frac{t}{H_2\left(\frac{\epsilon}{2}\right)}\right)\right). && \left(t \geq \tilde{\Theta}\left(H_2\left(\tfrac{\epsilon}{2}\right)\right)\right)
\end{aligned} \tag{18}$$

Our choice of L here also ensures that a bracket of size L is opened because the largest bracket is in the size of $\tilde{\Theta}(t)$ by Lemma 11.1. Furthermore,

$$L = \tilde{\Theta}\left(\frac{nt}{g(\epsilon_1)H_2\left(\frac{\epsilon}{2}\right)}\right) = \tilde{\Theta}\left(\frac{nt}{\max_{i\geq g(\epsilon_2)+1}\frac{i}{\Delta_i^2}}\right) \leq \tilde{\Theta}\left(\frac{nt}{n/\Delta_n^2}\right) = \tilde{\Theta}\left(\Delta_n^2 t\right) \leq \tilde{\Theta}(t).$$

The second term of (17) means the probability that SH fails to return an $\frac{\epsilon}{2}$-good arm from bracket A^* given that bracket A^* can well represent the entire instance. Denote $H_2'\left(\frac{\epsilon}{2}\right)$ as the sample complexity of returning an $\frac{\epsilon}{2}$-good from the bracket A^*, i.e.,

$$H_2'\left(\frac{\epsilon}{2}\right) = \max_{i\geq g'(\epsilon_2)+1}\frac{i}{\Delta_i'^2}\cdot\frac{1}{g'(\epsilon_1)}.$$

Let $H_2\left(\frac{\epsilon}{2}\right)$ be the sample complexity of returning an $\frac{\epsilon}{2}$-good arm from the entire instance, i.e.,

$$H_2\left(\frac{\epsilon}{2}\right) = \max_{i\geq g(\epsilon_2)+1}\frac{i}{\Delta_i^2}\cdot\frac{1}{g(\epsilon_1)}.$$

We use Theorem 1 to upper bound the second term of (17). To do so, we need to verify that the best bracket is allocated to enough budget, i.e., the budget allocated to the best bracket should be no less than $U_2\left(\frac{\epsilon}{2}\right)$. By event E_1, we have

$$\frac{1}{2}rg(\epsilon_k) \leq g'(\epsilon_k) \leq 2rg(\epsilon_k). \tag{19}$$

Then, by Lemma 13.1 and our choice of L, we have

$$U_2\left(\frac{\epsilon}{2}\right) = g'(\epsilon_1)H_2'\left(\frac{\epsilon}{2}\right) \leq 2rg(\epsilon_1)H_2'\left(\frac{\epsilon}{2}\right) = 2\frac{L}{n}g(\epsilon_1)H_2'\left(\frac{\epsilon}{2}\right) = \tilde{\Theta}\left(\frac{H_2'\left(\frac{\epsilon}{2}\right)}{H_2\left(\frac{\epsilon}{2}\right)}t\right) = \tilde{\Theta}(t).$$

Meanwhile, each bracket receives $\tilde{\Theta}(t)$ budget by the bracketing design. Thus the best bracket is allocated to enough budget. By applying Theorem 1 to (a_1) and Lemma 13.1 to (a_2) below, we have

$$\mathbb{P}\left(E_2^c \mid E_1\right) \overset{(a_1)}{\leq} \exp\left(-\tilde{\Theta}\left(\frac{t}{H_2'\left(\frac{\epsilon}{2}\right)}\right)\right) \overset{(a_2)}{\leq} \exp\left(-\tilde{\Theta}\left(\frac{t}{H_2\left(\frac{\epsilon}{2}\right)}\right)\right). \tag{20}$$

The third term of (17) means an arm whose mean reward is less than $\mu_1 - \epsilon$ has a larger empirical reward than an arm whose mean reward is larger than $\mu_1 - \epsilon/2$. Note the recently opened brackets may not have finished their first SH yet. In this case, the DSH always returns an empirical reward of negative infinity. Thus the arms returned from these brackets will never be selected by BSH. Denote $\dot{A}(t) := \left\{a \in \bigcup_{k=1, A_k\neq A^*}^{L_t} A_k, \mu_a < \mu_1 - \epsilon\right\}$. We have

$$\left|\dot{A}(t)\right| \leq \sum_{k=1}^{L_t} 2^k \leq 2\cdot\left(2^{1+\log_2\left(1+\ln(2)t\right)} - 1\right) \leq 8\cdot\left(1 + \ln(2)t\right).$$

By the number of arm pulls in Lemma 11.2,

$$\begin{aligned}
\mathbb{P}\left(\mu_{J_t} < \mu_1 - \epsilon \mid E_2\right) &\leq \mathbb{P}\left(\exists a_1 \in A_{r_i^*(t)}, \exists a_2 \in \dot{A}(t), \text{s.t.}, \mu_{a_1} \geq \mu_1 - \epsilon/2, \hat{\mu}_{a_1} < \hat{\mu}_{a_2}\right) \\
&\leq \left|\dot{A}(t)\right|\left|A_{r_i^*(t)}\right|\exp\left(-\tilde{\Theta}\left(\epsilon^2 t\right)\right) \\
&\leq \exp\left(-\tilde{\Theta}\left(\epsilon^2 t\right) + \ln\left(\left|\dot{A}(t)\right|\left|A_{r_i^*(t)}\right|\right)\right) \\
&\leq \exp\left(-\tilde{\Theta}\left(\epsilon^2 t\right) + \ln\left(8\cdot\left(1 + \ln(2)t\right)t\right)\right) \\
&\leq \exp\left(-\tilde{\Theta}\left(\epsilon^2 t\right) + \mathcal{O}(\ln t)\right) \\
&\leq \exp\left(-\tilde{\Theta}\left(\epsilon^2 t\right)\right).
\end{aligned} \tag{21}$$

Given (18), (20) and (21), we have the result proved. □

Lemma 13.1. *Let $H_2'\left(\frac{\epsilon}{2}\right)$ be the complexity measure of returning an $\frac{\epsilon}{2}$-good from the bracket A^*, i.e.,*

$$H_2'\left(\frac{\epsilon}{2}\right) = \max_{i \geq g'(\epsilon_2)+1} \frac{i}{\Delta_i'^2} \cdot \frac{1}{g'(\epsilon_1)}.$$

Let $H_2\left(\frac{\epsilon}{2}\right)$ be the complexity measure of returning an $\frac{\epsilon}{2}$-good from the entire instance, i.e.,

$$H_2\left(\frac{\epsilon}{2}\right) = \max_{i \geq g(\epsilon_2)+1} \frac{i}{\Delta_i^2} \cdot \frac{1}{g(\epsilon_1)}.$$

Given the event E_1 happens, we have

$$H_2'\left(\frac{\epsilon}{2}\right) \leq 64 H_2\left(\frac{\epsilon}{2}\right).$$

Proof. For the entire instance, by Lemma 13.2,

$$\max_{k \in \{3,\cdots,K\}} \frac{g(\epsilon_k)}{\epsilon_k^2} \leq U_1\left(\frac{\epsilon}{2}\right) \leq \max_{k \in \{3,\cdots,K\}} \frac{4g(\epsilon_k)}{\epsilon_k^2}. \tag{22}$$

For the best bracket, by Lemma 13.3, we have

$$\max_{i \geq g'(\epsilon_2)+1} \frac{i}{\Delta_i''^2} \leq \max_{i \geq g'(\epsilon_2)+1} \frac{i}{\Delta_i'^2} \leq 4 \max_{i \geq g'(\epsilon_2)+1} \frac{i}{\Delta_i''^2}.$$

Note the term $\max_{i \geq g'(\epsilon_2)+1} \frac{i}{\Delta_i''^2}$ can be considered the un-accelerated sample complexity on a virtual instance where all the arms are the same as the ones of A^*, except the best arm of A^* is replaced as the best arm in the entire instance (if the best arm of A^* is same as the best arm in the entire instance, then the virtual instance is same as A^*). The above inequalities essentially show that on event E_1, setting the best arm in A^* to have the same mean reward as the best arm of the entire instance does not change its un-accelerated sample complexity (up to a constant of 4). Denote

$$U_2'\left(\frac{\epsilon}{2}\right) = \max_{i \geq g'(\epsilon_2)+1} \frac{i}{\Delta_i''^2}.$$

Thus,

$$U_2'\left(\frac{\epsilon}{2}\right) \leq U_2\left(\frac{\epsilon}{2}\right) \leq 4U_2'\left(\frac{\epsilon}{2}\right). \tag{23}$$

By Lemma 13.2,

$$\max_{k \in \{3,\cdots,K\}} \frac{g'(\epsilon_k)}{\epsilon_k^2} \leq U_2'\left(\frac{\epsilon}{2}\right) \leq \max_{k \in \{3,\cdots,K\}} \frac{4g'(\epsilon_k)}{\epsilon_k^2}.$$

By event E_1, we have for $k \in [K]$

$$\frac{1}{2} r g(\epsilon_k) \leq g'(\epsilon_k) \leq 2r g(\epsilon_k). \tag{24}$$

Thus

$$\max_{k \in \{3,\cdots,K\}} \frac{r g(\epsilon_k)}{2\epsilon_k^2} \leq U_2'\left(\frac{\epsilon}{2}\right) \leq \max_{k \in \{3,\cdots,K\}} \frac{8r g(\epsilon_k)}{\epsilon_k^2}. \tag{25}$$

By (23) and (25),

$$\max_{k \in \{3,\cdots,K\}} \frac{r g(\epsilon_k)}{2\epsilon_k^2} \leq U_2\left(\frac{\epsilon}{2}\right) \leq \max_{k \in \{3,\cdots,K\}} \frac{32r g(\epsilon_k)}{\epsilon_k^2}. \tag{26}$$

Therefore, by (22) and (26), we have

$$\frac{1}{32r} \le \frac{U_1\left(\frac{\epsilon}{2}\right)}{U_2\left(\frac{\epsilon}{2}\right)} \le \frac{8}{r}. \tag{27}$$

By $U_1\left(\frac{\epsilon}{2}\right) = g(\epsilon_1)H_2\left(\frac{\epsilon}{2}\right)$, $U_2\left(\frac{\epsilon}{2}\right) = g'(\epsilon_1)H_2'\left(\frac{\epsilon}{2}\right)$, (24) and (27), we have

$$H_2'\left(\frac{\epsilon}{2}\right) = \frac{g(\epsilon_1)U_2\left(\frac{\epsilon}{2}\right)}{g'(\epsilon_1)U_1\left(\frac{\epsilon}{2}\right)} H_2\left(\frac{\epsilon}{2}\right) \le \frac{2}{r} \cdot 32r \cdot H_2\left(\frac{\epsilon}{2}\right) = 64H_2\left(\frac{\epsilon}{2}\right).$$

□

Lemma 13.2. *Let $U\left(\frac{\epsilon}{2}\right)$ be the un-accelerated sample complexity on any bandit instance for identifying an $\frac{\epsilon}{2}$ good arm (note $\epsilon_2 = \frac{\epsilon}{2}$), i.e.,*

$$U\left(\frac{\epsilon}{2}\right) = \max_{i \ge g(\epsilon_2)+1} \frac{i}{\Delta_i^2}.$$

Then up to a constant of 4, the un-accelerated sample complexity can be approximated by $\max_{k\in\{3,\cdots,K\}} \frac{g(\epsilon_k)}{\epsilon_k^2}$, i.e.,

$$\max_{k\in\{3,\cdots,K\}} \frac{g(\epsilon_k)}{\epsilon_k^2} \le U\left(\frac{\epsilon}{2}\right) \le \max_{k\in\{3,\cdots,K\}} \frac{4g(\epsilon_k)}{\epsilon_k^2}.$$

Proof. Let $i_0 \ge g(\epsilon_2) + 1$ satisfy

$$U\left(\frac{\epsilon}{2}\right) = \max_{i \ge g(\epsilon_2)+1} \frac{i}{\Delta_i^2} = \frac{i_0}{\Delta_{i_0}^2}.$$

Denote k_0 as the smallest index of the peeling interval to which i_0 belongs, i.e., $\epsilon_{k_0-1} < \Delta_{i_0} \le \epsilon_{k_0}$. Since $i_0 \ge g(\epsilon_2) + 1$, we know $k_0 \ge 3$. By $\epsilon_{k_0-1} = \frac{\epsilon_{k_0}}{2}$, we have $\frac{\epsilon_{k_0}}{2} < \Delta_{i_0}$. Thus,

$$U\left(\frac{\epsilon}{2}\right) = \frac{i_0}{\Delta_{i_0}^2} < \frac{4i_0}{\epsilon_{k_0}^2} \le \frac{4g(\epsilon_{k_0})}{\epsilon_{k_0}^2} \le \max_{k\in\{3,\cdots,K\}} \frac{4g(\epsilon_k)}{\epsilon_k^2},$$

where the second inequality is by $i_0 \le g(\epsilon_{k_0})$. By the definition of $U\left(\frac{\epsilon}{2}\right)$, we have

$$U\left(\frac{\epsilon}{2}\right) \ge \max_{k\in\{3,\cdots,K\}} \frac{g(\epsilon_k)}{\Delta_{g(\epsilon_k)}^2} \ge \max_{k\in\{3,\cdots,K\}} \frac{g(\epsilon_k)}{\epsilon_k^2}$$

Thus,

$$\max_{k\in\{3,\cdots,K\}} \frac{g(\epsilon_k)}{\epsilon_k^2} \le U\left(\frac{\epsilon}{2}\right) \le \max_{k\in\{3,\cdots,K\}} \frac{4g(\epsilon_k)}{\epsilon_k^2}.$$

□

Lemma 13.3. *Given that the event E_1 holds, for $i \ge g'(\epsilon_2) + 1$, Δ_i' the suboptimality gap of i-th best arm in the best bracket with respect to the best arm in the best bracket can be approximated by Δ_i'', the suboptimality gap of i-th best arm in the best bracket with respect to the best arm in the entire instance, i.e.,*

$$\frac{1}{\Delta_i''^2} \le \frac{1}{\Delta_i'^2} \le \frac{4}{\Delta_i''^2}.$$

Proof. The first inequality holds trivially, and the equality holds when the best arm out of the entire arm set is included in the best bracket. For the second inequality, as the best arm in the best bracket is $\frac{\epsilon}{4}$-good, we have

$$\Delta_i' - \frac{\epsilon}{4} \ge \Delta_i'' - \frac{\epsilon}{2} \ge \frac{\Delta_i''}{2} - \frac{\epsilon}{4},$$

which results in the second inequality. □

C.5. Proof of Corollary 6.1

Corollary 6.1. *Consider the EqualGap(m) instance. BUCB achieves an expected (ϵ, δ)-unverifiable sample complexity as*

$$\mathbb{E}\left[\tau_{\epsilon,\delta}\right] \le \tilde{\mathcal{O}}\left(\frac{n}{\epsilon^2 m}\log\left(\frac{1}{\delta}\right)\right).$$

BSH achieves an (ϵ, δ)-unverifiable sample complexity as, with probability $1-\delta$,

$$\tau_{\epsilon,\delta} \le \tilde{\mathcal{O}}\left(\frac{n}{\epsilon^2 m}\log\left(\frac{1}{\delta}\right)\right).$$

Proof. For BUCB, Theorem 7 of Katz-Samuels and Jamieson (2020) gives the following instance dependent upper bound for the (ϵ, δ)-unverifiable sample complexity,

$$\mathbb{E}\left[\tau_{\epsilon,\delta}\right] \le \min_{j\in[g(\epsilon)]}\frac{1}{j}\left(\sum_{i=1}^{g(\epsilon)}\left(\Delta_{i\vee j, g(\epsilon)+1}\right)^{-2}\ln\left(\frac{n}{j\delta}\right) + \sum_{i=g(\epsilon)+1}^{n}\Delta_{j,i}^{-2}\ln\left(\frac{1}{\delta}\right)\right).$$

Plug the instance

$$\mu_1 = \frac{5}{4}\epsilon, \mu_i = \epsilon, \forall i \in \{2,\dots,m\}, \mu_i = 0, \forall i \ge m+1 \text{ for some } m \ge 2, \epsilon > 0\ .$$

We easily have

$$\mathbb{E}\left[\tau_{\epsilon,\delta}\right] \le \tilde{\mathcal{O}}\left(\frac{n}{\epsilon^2 m}\log\left(\frac{1}{\delta}\right)\right).$$

For BSH,

$$\max_{i\ge g(\epsilon/2)+1}\frac{i}{\Delta_i^2}\cdot\frac{1}{g\left(\frac{\epsilon}{4}\right)} = \frac{16n}{25m\epsilon^2}.$$

□

C.6. Proof of Corollary 6.2

Corollary 6.2. *Consider the Polynomial(α) instance; i.e., $\Delta_i = \left(\frac{i}{n}\right)^\alpha$ with $\alpha > 0.5$. For any $\epsilon \in (0,1)$, let $\hat{\tau}_{\epsilon,\delta}$ be the upper bound of the expected (ϵ, δ)-unverifiable sample complexity reported in Katz-Samuels and Jamieson (2020, Theorem 7) for BUCB. Then, BUCB satisfies*

$$\mathbb{E}\left[\tau_{\epsilon,\delta}\right] \le \hat{\tau}_{\epsilon,\delta} = \tilde{\Theta}\left(\epsilon^{-\frac{2\alpha-1}{\alpha}} n\log\left(\frac{1}{\delta}\right)\right).$$

On the other hand, BSH satisfies, with probability $1-\delta$,

$$\tau_{\epsilon,\delta} \le \tilde{\mathcal{O}}\left(\epsilon^{-2}\log\left(\frac{1}{\delta}\right)\right).$$

Proof. We first show the result for Bracketing UCB. Theorem 7 of Katz-Samuels and Jamieson (2020) gives the following instance dependent upper bound for the (ϵ, δ)-unverifiable sample complexity,

$$\hat{\tau}_{\epsilon,\delta} = \min_{j\in[g(\epsilon)]}\frac{1}{j}\left(\sum_{i=1}^{g(\epsilon)}\left(\Delta_{i\vee j, g(\epsilon)+1}\right)^{-2}\ln\left(\frac{n}{j\delta}\right) + \sum_{i=g(\epsilon)+1}^{n}\Delta_{j,i}^{-2}\ln\left(\frac{1}{\delta}\right)\right). \tag{28}$$

Plug the instance $\Delta_i = \left(\frac{i}{n}\right)^\alpha$ into the above,

$$
\begin{aligned}
\hat{\tau}_{\epsilon,\delta} &= \min_{j\in[g(\epsilon)]} \frac{1}{j}\left(\sum_{i=1}^{g(\epsilon)} \left(\Delta_{i\vee j, g(\epsilon)+1}\right)^{-2} \ln\left(\frac{n}{j\delta}\right) + \sum_{i=g(\epsilon)+1}^{n} \Delta_{j,i}^{-2} \ln\left(\frac{1}{\delta}\right)\right) \\
&> \min_{j\in[g(\epsilon)]} \frac{1}{j}\left(\sum_{i=1}^{g(\epsilon)} \left(\Delta_{i\vee j, g(\epsilon)+1}\right)^{-2} \ln\left(\frac{n}{j\delta}\right)\right) \\
&> \min_{j\in[g(\epsilon)]} \frac{1}{j} \ln\left(\frac{1}{\delta}\right)\left(\sum_{i=1}^{g(\epsilon)} \left(\Delta_{i\vee j, g(\epsilon)+1}\right)^{-2}\right) \\
&> \min_{j\in[g(\epsilon)]} \frac{1}{j} \ln\left(\frac{1}{\delta}\right)\left(\left(\Delta_{g(\epsilon), g(\epsilon)+1}\right)^{-2}\right) \\
&> \min_{j\in[g(\epsilon)]} \frac{n^{2\alpha}}{j} \ln\left(\frac{1}{\delta}\right)\left(\frac{1}{\left((g(\epsilon)+1)^\alpha - (g(\epsilon))^\alpha\right)^2}\right) \\
&\overset{(a_1)}{>} \min_{j\in[g(\epsilon)]} \frac{n^{2\alpha}}{j} \ln\left(\frac{1}{\delta}\right)\left(\frac{1}{\left(\alpha\left(g(\epsilon)+1\right)^{\alpha-1}\right)^2}\right) \\
&> \min_{j\in[g(\epsilon)]} \frac{n^{2\alpha}}{j} \ln\left(\frac{1}{\delta}\right)\left(\frac{1}{\alpha^2\left(2g(\epsilon)\right)^{2\alpha-2}}\right) \\
&= \frac{2^{-2\alpha+2}\alpha^{-2}n^{2\alpha}}{\left(g(\epsilon)\right)^{2\alpha-1}} \ln\left(\frac{1}{\delta}\right) \\
&\overset{(a_2)}{=} \frac{2^{-2\alpha+2}\alpha^{-2}n}{\epsilon^{\frac{2\alpha-1}{\alpha}}} \ln\left(\frac{1}{\delta}\right).
\end{aligned}
$$

Note here we lower bound the upper bound in Theorem 7 of Katz-Samuels and Jamieson (2020). The inequality (a_1) results by

$$(g+1)^\alpha - g^\alpha = \int_g^{g+1} \alpha x^{\alpha-1}\,\mathrm{d}x \le \alpha(g+1)^{\alpha-1}.$$

The equation (a_2) is because of $g(\epsilon) = n\epsilon^{1/\alpha}$ by the definition of the instance.

We also upper bound (28). Note that for $i > j$, we have

$$i^\alpha - j^\alpha = \int_j^i \alpha x^{\alpha-1}\,\mathrm{d}x \ge \alpha(i-j)\cdot j^{\alpha-1}\ .$$

With this, we have

$$\hat{\tau}_{\epsilon,\delta} \le n^{2\alpha}\left(\min_{j\le g(\epsilon)} \frac{1}{(g(\epsilon)+1-j)^2 j^{2\alpha-2}} + \frac{1}{j}\sum_{i=j+1}^{g(\epsilon)} \frac{1}{(g(\epsilon)+1-i)^2 i^{2\alpha-2}} + \frac{1}{j}\sum_{i=g(\epsilon)+1}^{n} \frac{1}{(i-j)^2 j^{2\alpha-2}}\right)$$

We can bound the third sum by

$$\frac{1}{j}\sum_{i=g(\epsilon)+1}^{n} \frac{1}{(i-j)^2 j^{2\alpha-2}} \lesssim \frac{1}{j^{2\alpha-1}}\cdot\frac{1}{g(\epsilon)+1-j}$$

Take $j = g(\epsilon) - \sqrt{g(\epsilon)}$. This means that $j = \Theta(g(\epsilon))$ and $g(\epsilon) - j = \sqrt{g(\epsilon)}$. With this choice,

$$\begin{aligned}
\hat{\tau}_{\epsilon,\delta} &\lesssim n^{2\alpha}\left(\frac{1}{g(\epsilon)\cdot g(\epsilon)^{2\alpha-2}} + \frac{1}{g(\epsilon)}\cdot\left(\sum_{i=g(\epsilon)-\sqrt{g(\epsilon)}+1}^{g(\epsilon)} \frac{1}{(g(\epsilon)+1-i)^2}\cdot\frac{1}{g(\epsilon)^{2\alpha-2}}\right) + \frac{1}{g(\epsilon)^{2\alpha-1}}\cdot\frac{1}{1+\sqrt{g(\epsilon)}}\right)\\
&\le n^{2\alpha}\left(\frac{1}{g(\epsilon)^{2\alpha-1}} + \frac{1}{g(\epsilon)^{2\alpha-1}}\cdot\frac{\pi^2}{6} + \frac{1}{g(\epsilon)^{2\alpha-1}}\cdot\frac{1}{1+\sqrt{g(\epsilon)}}\right)\\
&\lesssim n^{2\alpha}\cdot\frac{1}{g(\epsilon)^{2\alpha-1}} = n^{2\alpha}\frac{1}{(n\epsilon^{1/\alpha})^{2\alpha-1}} = \frac{n}{\epsilon^{\frac{2\alpha-1}{\alpha}}}.
\end{aligned}$$

For Bracketing SH,

$$\max_{i\ge g(\epsilon/2)+1}\frac{i}{\Delta_i^2}\cdot\frac{1}{g\left(\frac{\epsilon}{4}\right)} = \max_{i\ge g(\epsilon/2)+1} i^{1-2\alpha}n^{2\alpha}\frac{1}{n\left(\frac{\epsilon}{4}\right)^{1/\alpha}} < \left(n\left(\frac{\epsilon}{2}\right)^{1/\alpha}\right)^{1-2\alpha} n^{2\alpha}\frac{1}{n\left(\frac{\epsilon}{4}\right)^{1/\alpha}} < \frac{16}{\epsilon^2}.$$

where both the last inequality and the last equality are by $\alpha > 0.5$.

□

C.7. An alternative upper bound for BSH

The following bound of BSH is similar to the bound of Theorem 6. Though the bound of Theorem 14 is minimax in nature, we show, in Corollary 14.1 and 14.2, that it still achieves the same upper bound as BUCB for the EqualGap(m) instance and a better upper bound for Polynomial(α) than BUCB when n is large enough because it does not scale with n polynomially. Since this bound involves an optimization problem, we speculate it could be tighter than the bound of Theorem 6 for certain instances.

Theorem 14. *For any $\epsilon \in (0,1)$, the error probability of Bracketing SH satisfies*

$$\mathbb{P}\left(\mu_{J_t} < \mu_1 - \epsilon\right) \le \exp\left(-\tilde{\Theta}\left(\min\left\{\max_{i\in[g(\epsilon/2)]}\frac{i\Delta^2_{i,g(\epsilon/2)+1}}{n}, \epsilon^2\right\}t'\right)\right),$$

where $t' = \frac{t}{4\ln t}$. Accordingly, $\tilde{\mathcal{O}}\left(\max\left\{\min_{i\in[g(\epsilon/2)]}\frac{n}{i\Delta^2_{i,g(\epsilon/2)+1}}\log\left(\frac{1}{\delta}\right), \frac{1}{\epsilon^2}\log\left(\frac{1}{\delta}\right)\right\}\right)$ samples are sufficient for Bracketing SH to output an ϵ-good arm with probability $1-\delta$.

Proof. The proof idea is similar to the proof of Theorem 6. The core idea is to find the best bracket whose size is well-balanced. At round t, define the best bracket as bracket $r_i^*(t) = \left\lceil \log_2\left(c\Delta^2_{i,g(\epsilon/2)+1}t'\right)\right\rceil$, where $i \in [g(\epsilon/2)]$ is a free parameter and c is a logarithmic term for shorthand $c = \left(16\ln(4e)\log_2^2\left(\frac{i\Delta^2_{i,g(\epsilon/2)+1}t'}{n}\right)\log_2\left(2\Delta^2_{i,g(\epsilon/2)+1}t'\right)\right)^{-1}$. Thus the size of bracket $r_i^*(t)$ satisfies,

$$c\Delta^2_{i,g(\epsilon/2)+1}t' \le \left|A_{r_i^*(t)}\right| \le 2^{\log_2\left(c\Delta^2_{i,g(\epsilon/2)+1}t'\right)+1} = 2c\Delta^2_{i,g(\epsilon/2)+1}t'.$$

Define the following events:

- E_1: the number of arms with mean reward at least μ_i included in bracket $r_i^*(t)$ is at least j_t, where $j_t = \left\lfloor\frac{ci\Delta^2_{i,g(\epsilon/2)+1}t'}{2n}\right\rfloor$.
- E_2: $\mu_{a_{r_i^*(t)}} \ge \mu_1 - \epsilon/2$, where $a_{r_i^*(t)}$ is the arm representing bracket $r_i^*(t)$ at round t.

Then the error probability of Bracketing SH can be expressed as follows,

$$\begin{aligned}\mathbb{P}\big(\mu_{J_t} < \mu_1 - \epsilon\big) &= \mathbb{P}\big(\mu_{J_t} < \mu_1 - \epsilon, E_1^c\big) + \mathbb{P}\big(\mu_{J_t} < \mu_1 - \epsilon, E_1, E_2^c\big) + \mathbb{P}\big(\mu_{J_t} < \mu_1 - \epsilon, E_1, E_2\big) \\ &\le \mathbb{P}\big(E_1^c\big) + \mathbb{P}\big(E_1, E_2^c\big) + \mathbb{P}\big(\mu_{J_t} < \mu_1 - \epsilon, E_2\big). \end{aligned} \tag{29}$$

We bound the three terms respectively. For the first term, we use the same technique as in the proof sketch.

$$\begin{aligned}\mathbb{P}\big(E_1^c\big) &\le \exp\left(-\left|A_{r_i^*(t)}\right| \cdot \mathsf{KL}\left(\frac{j_t}{\left|A_{r_i^*(t)}\right|}, \frac{i}{n}\right)\right) \\ &\overset{(a_1)}{\le} \exp\left(-\frac{\left|A_{r_i^*(t)}\right| n}{2i}\left(\frac{j_t}{\left|A_{r_i^*(t)}\right|} - \frac{i}{n}\right)^2\right) \\ &\le \exp\left(-\frac{\Delta^2_{i,g(\epsilon/2)+1} t' n}{2i}\left(\frac{\frac{i\Delta^2_{i,g(\epsilon/2)+1} t'}{2n}}{\Delta^2_{i,g(\epsilon/2)+1} t'} - \frac{i}{n}\right)^2\right) \\ &\le \exp\left(-\frac{ci\Delta^2_{i,g(\epsilon/2)+1} t'}{8n}\right) \\ &\le \exp\left(-\tilde{\Theta}\left(\frac{i\Delta^2_{i,g(\epsilon/2)+1} t'}{n}\right)\right). \end{aligned} \tag{30}$$

The inequality (a_1) is due to $\mathsf{KL}(p,q) \ge \frac{(p-q)^2}{2\max\{p,q\}}$. For the second term of (29), we bound the event that SH returns a non-$(j_t, \Delta_{i,g(\epsilon/2)+1})$-good arm, where j_t is with respect to the best bracket. We use $\mu_{(j_t)}$ for the mean reward of the j_t-th best arm in bracket $r_i^*(t)$. Thus $\mu_{(j_t)} - \Delta_{i,g(\epsilon/2)+1} \ge \mu_1 - \epsilon/2$. By Theorem 4,

$$\begin{aligned}\mathbb{P}\big(E_1, E_2^c\big) &\le \mathbb{P}\left(\mu_{a_{r_i^*(t)}} < \mu_1 - \epsilon/2 \mid E_1\right) \\ &\le \mathbb{P}\left(\mu_{a_{r_i^*(t)}} < \mu_{(j_t)} - \Delta_{i,g(\epsilon/2)+1} \mid E_1\right) \\ &\overset{(a_1)}{\le} \log_2\left|A_{r_i^*(t)}\right| \cdot \exp\left(-\text{const} \cdot j_t\left(\frac{\Delta^2_{i,g(\epsilon/2)+1} t'}{4\left|A_{r_i^*(t)}\right| \log_2^2(2j_t) \log_2\left|A_{r_i^*(t)}\right|} - \ln(4e)\right)\right) \\ &\le \log_2\left|A_{r_i^*(t)}\right| \cdot \exp\left(-\text{const} \cdot \frac{ci\Delta^2_{i,g(\epsilon/2)+1} t'}{n}\right) \\ &\le \exp\left(-\tilde{\Theta}\left(\frac{i\Delta^2_{i,g(\epsilon/2)+1} t'}{n}\right)\right). \end{aligned} \tag{31}$$

For the inequality (a_1), we bound the term in parenthesis as a constant,

$$\begin{aligned}
&\frac{\Delta^2_{i,g(\epsilon/2)+1}t'}{4\left|A_{r_i^*(t)}\right|\log_2^2(2j_t)\log_2\left|A_{r_i^*(t)}\right|} - \ln(4e)\\
>&\frac{\Delta^2_{i,g(\epsilon/2)+1}t'}{8c\Delta^2_{i,g(\epsilon/2)+1}t'\log_2^2(2c\frac{i\Delta^2_{i,g(\epsilon/2)+1}t'}{2n})\log_2\left(2c\Delta^2_{i,g(\epsilon/2)+1}t'\right)} - \ln(4e)\\
>&\frac{1}{8c\log_2^2(\frac{i\Delta^2_{i,g(\epsilon/2)+1}t'}{n})\log_2\left(2\Delta^2_{i,g(\epsilon/2)+1}t'\right)} - \ln(4e)\\
>&2\ln(4e) - \ln(4e)\\
=&\ln(4e).
\end{aligned}$$

The third term of (29) means an arm whose mean reward is less than $\mu_1 - \epsilon$ has a larger empirical reward than an arm whose mean reward is larger than $\mu_1 - \epsilon/2$. Note the recently opened brackets may not have finished their first SH yet. In this case, the DSH always returns an empirical reward of negative infinity. Thus the arms returned from these brackets will never be selected by BSH. Denote $\dot{A}(t) := \left\{a \in \bigcup_{k=1,k\neq r_i^*(t)}^{L_t} A_k, \mu_a < \mu_1 - \epsilon\right\}$. We have

$$\left|\dot{A}(t)\right| \leq \sum_{k=1}^{L_t} 2^k \leq 2\cdot\left(2^{1+\log_2(1+\ln(2)t)} - 1\right) \leq 8\cdot\left(1 + \ln(2)t\right).$$

By the number of arm pulls in Lemma 11.2,

$$\begin{aligned}
\mathbb{P}\left(\mu_{J_t} < \mu_1 - \epsilon, E_2\right) \leq& \mathbb{P}\left(\exists a_1 \in A_{r_i^*(t)}, \exists a_2 \in \dot{A}(t), \text{s.t.}, \mu_{a_1} \geq \mu_1 - \epsilon/2, \hat{\mu}_{a_1} < \hat{\mu}_{a_2}\right)\\
\leq& \left|\dot{A}(t)\right|\left|A_{r_i^*(t)}\right|\exp\left(-\left(\frac{\epsilon}{2}\right)^2\cdot\text{const}\cdot\frac{t'}{\log_2 n}\right)\\
\leq& \exp\left(-\text{const}\cdot\epsilon^2\frac{t'}{\log_2 n} + \ln\left(\left|\dot{A}(t)\right|\left|A_{r_i^*(t)}\right|\right)\right)\\
\leq& \exp\left(-\text{const}\cdot\epsilon^2\frac{t'}{\log_2 n} + \ln\left(8\cdot\left(1 + \ln(2)t\right)2\Delta^2_{i,g(\epsilon/2)+1}t'\right)\right)\\
\leq& \exp\left(-\text{const}\cdot\epsilon^2\frac{t'}{\log_2 n} + \mathcal{O}(\ln t)\right)\\
\leq& \exp\left(-\tilde{\Theta}\left(\epsilon^2 t'\right)\right).
\end{aligned} \tag{32}$$

Combine (30)(31)(32) and the fact $i \in [g(\epsilon/2)]$ is a free parameter, then we have

$$\mathbb{P}\left(\mu_{J_t} < \mu_1 - \epsilon\right) \leq \exp\left(-\tilde{\Theta}\left(\min\left\{\max_{i\in[g(\epsilon/2)]}\frac{i\Delta^2_{i,g(\epsilon/2)+1}}{n}, \epsilon^2\right\}t'\right)\right).$$

□

Corollary 14.1. *Consider the EqualGap(m) instance. BUCB achieves an expected (ϵ,δ)-unverifiable sample complexity as*

$$\mathbb{E}\left[\tau_{\epsilon,\delta}\right] \leq \tilde{\mathcal{O}}\left(\frac{n}{\epsilon^2 m}\log\left(\frac{1}{\delta}\right)\right).$$

For BSH, the bound of Theorem 14 shows that BSH achieves an (ϵ,δ)-unverifiable sample complexity as

$$\tau_{\epsilon,\delta} \leq \tilde{\mathcal{O}}\left(\frac{n}{\epsilon^2 m}\log\left(\frac{1}{\delta}\right)\right)$$

with probability $1-\delta$.

Proof. For BUCB, the proof is presented in the proof of Corollary 6.1. For BSH, the result is straightforward by plugging the instance,

$$\min_{i\in[g(\epsilon/2)]} \frac{n}{i\Delta^2_{i,g(\epsilon/2)+1}} \log\left(\frac{1}{\delta}\right) \le \min\left\{\frac{n}{\left(\epsilon/2\right)^2}\log\left(\frac{1}{\delta}\right), \frac{n}{m\epsilon^2}\log\left(\frac{1}{\delta}\right)\right\} = \frac{n}{m\epsilon^2}\log\left(\frac{1}{\delta}\right).$$

□

The following corollary shows that even if we use the upper bound of Theorem 14 that is minimax in nature, we can still achieve an upper bound for the Polynomial(α) instance that does not scale with the instance size n, and for moderately large ϵ or large enough n, the upper bound of Theorem 14 is always better than the upper bound of BUCB.

Corollary 14.2. *Consider the Polynomial(α) instance $\Delta_i = \left(\frac{i}{n}\right)^\alpha$. BUCB achieves an expected (ϵ,δ)-unverifiable sample complexity as*

$$\mathbb{E}\left[\tau_{\epsilon,\delta}\right] \le \hat{\tau}_{\epsilon,\delta} = \tilde{\Theta}\left(\epsilon^{-\frac{2\alpha-1}{\alpha}} n \log\left(\frac{1}{\delta}\right)\right).$$

For BSH, the bound of Theorem 14 shows that BSH achieves an (ϵ,δ)-unverifiable sample complexity as

$$\tau_{\epsilon,\delta} \le \tilde{\mathcal{O}}\left(\epsilon^{-\frac{2\alpha+1}{\alpha}} \log\left(\frac{1}{\delta}\right)\right).$$

with probability $1-\delta$.

Proof. For BUCB, the proof is presented in the proof of Corollary 6.2. For BSH, first, notice that

$$\begin{aligned}\Delta^2_{i,j+1} > \Delta^2_{i,j} =& \left(\left(\frac{j}{n}\right)^\alpha - \left(\frac{i}{n}\right)^\alpha\right)^2 \\ =& \left(\frac{j}{n}\right)^{2\alpha} + \left(\frac{i}{n}\right)^{2\alpha} - 2\left(\frac{ij}{n^2}\right)^\alpha.\end{aligned}$$

The sample complexity satisfies,

$$\begin{aligned}\min_{i\in[g(\epsilon/2)]} \frac{n}{i\Delta^2_{i,g(\epsilon/2)+1}} \log\left(\frac{1}{\delta}\right) \le& \min_{i\in[g(\epsilon/2)]} \frac{n}{i\left(\left(\frac{g(\epsilon/2)}{n}\right)^{2\alpha} + \left(\frac{i}{n}\right)^{2\alpha} - 2\left(\frac{ig(\epsilon/2)}{n^2}\right)^\alpha\right)} \log\left(\frac{1}{\delta}\right) \\ =& \min_{i\in[g(\epsilon/2)]} \frac{n\cdot n^{2\alpha}}{i\left(\left(g(\epsilon/2)\right)^{2\alpha} + i^{2\alpha} - 2\left(ig(\epsilon/2)\right)^\alpha\right)} \log\left(\frac{1}{\delta}\right) \\ \overset{(a_1)}{\le}& \frac{2\cdot n^{2\alpha+1}}{(g(\epsilon/2))^{2\alpha+1}(1+2^{-2\alpha}-2^{-\alpha+1})} \log\left(\frac{1}{\delta}\right) \\ =& \frac{2\cdot n^{2\alpha+1}}{(n(\epsilon/2)^{1/\alpha})^{2\alpha+1}(1+2^{-2\alpha}-2^{-\alpha+1})} \log\left(\frac{1}{\delta}\right) \\ =& 2^{\frac{2\alpha+1}{\alpha}+1}\left(1+2^{-2\alpha}-2^{-\alpha+1}\right)^{-1}\epsilon^{-\frac{2\alpha+1}{\alpha}} \log\left(\frac{1}{\delta}\right).\end{aligned}$$

The inequality (a_1) is by taking $i = g(\epsilon/2)/2$. □

C.8. (ϵ, δ)-unverifiable sample complexity of algorithms with ϵ-error probability bound

Definition 1 ((ϵ, δ)-unverifiable sample complexity (Katz-Samuels and Jamieson, 2020))**.** *For an algorithm π and an instance ρ. Let $\tau_{\epsilon,\delta}$ be a stopping time such that*

$$\mathbb{P}\big(\forall t \geq \tau_{\epsilon,\delta} : \mu_{J_t} > \mu_1 - \epsilon\big) \geq 1 - \delta.$$

Then, $\tau_{\epsilon,\delta}$ is called (ϵ, δ)-unverifiable sample complexity of the algorithm with respect to ρ.

The (ϵ, δ)-unverifiable sample complexity requires all the outputs at and after $t = \tau_{\epsilon,\delta}$ are ϵ-good. This is slightly different from the sample complexity of BSH. However, we show they are order-wise equivalent. To see this, let us consider an anytime algorithm that achieves an exponentially decreasing error probability,

$$\mathbb{P}\big(\mu_{J_t} < \mu_1 - \epsilon\big) \leq \exp(-c{\cdot}t).$$

Let $t_{\epsilon,\delta}$ be the first time step that satisfies $\mathbb{P}\Big(\mu_{J_{t_{\epsilon,\delta}}} < \mu_1 - \epsilon\Big) \leq \delta$. Note this only guarantees the output at round $t_{\epsilon,\delta}$ is ϵ-good with probability $1 - \delta$, instead of all the outputs at and after $t_{\epsilon,\delta}$. We prove our claim by showing the probability that there is a non-ϵ-good output at or after $t_{\epsilon,\delta}$ has an exponentially decreasing rate with the same parameter.

$$\begin{aligned}
\mathbb{P}\big(\exists t \geq t_{\epsilon,\delta} : \mu_{J_t} < \mu_1 - \epsilon\big) &\leq \sum_{t=t_{\epsilon,\delta}}^{\infty} \mathbb{P}\big(\mu_{J_t} < \mu_1 - \epsilon\big) \\
&\leq \sum_{t=t_{\epsilon,\delta}}^{\infty} \exp(-c{\cdot}t) \\
&= \lim_{t\to\infty} \exp\big(-c{\cdot}t_{\epsilon,\delta}\big) \frac{1 - \exp(-c{\cdot}t)}{1 - \exp(-c)} \\
&< \exp\big(-c{\cdot}t_{\epsilon,\delta}\big) \frac{1}{1 - \exp(-c)} \\
&= \exp\big(-\Theta(c{\cdot}t_{\epsilon,\delta})\big).
\end{aligned}$$

D. Discussion on the Practical Algorithm Implementation

The design of DSH (Algorithm 3.1) and BSH (Algorithm 4) involves two key ideas, bracketing and doubling trick. In practice, both two techniques could be implemented in a more efficient way. In addition, the base algorithm SH (Algorithm 3.1) has a commonly used implementation for reusing samples (Baharav and Tse, 2019; Jun and Nowak, 2016). We summarize these strategies as follows for the practitioner's consideration. However, they do not make any order-wise difference (not more than logarithmic factors) in terms of the theoretical guarantee.

- SH: The whole procedure of SH is divided into $\log_2 n$ stages. All the stages are independent of each other since the samples of prior stages are abandoned when a new stage starts, as described in Algorithm 3.1. In practical usage, one can keep all the samples since the first stage for the surviving arms.
- DSH: The common doubling trick does not require keeping the samples after finishing one invocation of the base algorithm. Not requiring to keep the samples is convenient to implement as we only need to repeatedly initialize a new instance of the base algorithm with the doubled budget parameter. In fact, for multi-armed bandit problems, keeping all the samples of prior invocations is beneficial. For the implementation, one can create a class for the base algorithm SH, and we initialize a new instance of the class with initial empirical rewards equal to the empirical rewards saved in the prior instance.
- BSH: The bracketing technique does not promise to avoid overlap with the already-opened brackets. A more practical implementation of BSH is to share the samples of the same arm across different brackets. Such that the empirical reward is more accurate. However, reusing sampling across different brackets is meaningless if we consider the infinitely-armed bandit models. Because, for the infinitely-armed bandit models, we will not draw exactly the same arm more than once. Thus it is barely possible to have overlap among the opened brackets.

Note our implementation reported in section 5 only uses the first reusing strategy for the base algorithm SH.

References

Richard Arratia and Louis Gordon. Tutorial on large deviations for the binomial distribution. *Bulletin of mathematical biology*, 51(1):125–131, 1989.

Jean-Yves Audibert and Sébastien Bubeck. Best arm identification in multi-armed bandits. In *Proceedings of the Conference on Learning Theory (COLT)*, pages 41–66, 2010.

Jean-Yves Audibert, Sébastien Bubeck, and Rémi Munos. Best Arm Identification in Multi-Armed Bandits. In *Proceedings of the Conference on Learning Theory (COLT)*, 2010.

Mohammad Gheshlaghi Azar, Ian Osband, and Rémi Munos. Minimax regret bounds for reinforcement learning. In *International Conference on Machine Learning (ICML)*, pages 263–272. PMLR, 2017.

Maryam Aziz, Jesse Anderton, Emilie Kaufmann, and Javed Aslam. Pure exploration in infinitely-armed bandit models with fixed-confidence. In *Algorithmic Learning Theory*, pages 3–24. PMLR, 2018.

Tavor Baharav and David Tse. Ultra fast medoid identification via correlated sequential halving. *Advances in Neural Information Processing Systems (NeurIPS)*, 32, 2019.

Romain Camilleri, Kevin Jamieson, and Julian Katz-Samuels. High-dimensional experimental design and kernel bandits. In *International Conference on Machine Learning (ICML)*, pages 1227–1237. PMLR, 2021.

Alexandra Carpentier and Andrea Locatelli. Tight (lower) bounds for the fixed budget best arm identification bandit problem. *Journal of Machine Learning Research*, 49(June):590–604, 2016.

Alexandra Carpentier and Michal Valko. Simple regret for infinitely many armed bandits. In *International Conference on Machine Learning (ICML)*, pages 1133–1141. PMLR, 2015.

Arghya Roy Chaudhuri and Shivaram Kalyanakrishnan. PAC identification of a bandit arm relative to a reward quantile. In *Thirty-First AAAI Conference on Artificial Intelligence*, 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 (ICML)*, pages 991–1000. PMLR, 2019.

Lijie Chen, Jian Li, and Mingda Qiao. Nearly instance optimal sample complexity bounds for top-k arm selection. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 101–110. PMLR, 2017a.

Lijie Chen, Jian Li, and Mingda Qiao. Towards instance optimal bounds for best arm identification. In *Conference on Learning Theory (COLT)*, pages 535–592. PMLR, 2017b.

Christoph Dann, Tor Lattimore, and Emma Brunskill. Unifying pac and regret: Uniform pac bounds for episodic reinforcement learning. *Advances in Neural Information Processing Systems (NeurIPS)*, 30, 2017.

Rémy Degenne and Wouter M Koolen. Pure exploration with multiple correct answers. *Advances in Neural Information Processing Systems*, 32, 2019.

Eyal Even-Dar, Shie Mannor, Yishay Mansour, and Sridhar Mahadevan. Action elimination and stopping conditions for the multi-armed bandit and reinforcement learning problems. *Journal of machine learning research*, 7(6), 2006.

Aurélien Garivier and Emilie Kaufmann. Optimal best arm identification with fixed confidence. In *Proceedings of the Conference on Learning Theory (COLT)*, pages 998–1027, 2016.

Avinatan Hassidim, Ron Kupfer, and Yaron Singer. An optimal elimination algorithm for learning a best arm. *Advances in Neural Information Processing Systems (NeurIPS)*, 33:10788–10798, 2020.

Kevin Jamieson, Matthew Malloy, Robert Nowak, and Sébastien Bubeck. lil'UCB: An optimal exploration algorithm for multi-armed bandits. In *Proceedings of the Conference on Learning Theory (COLT)*, pages 423–439, 2014.

K.-S. Jun and R. Nowak. Anytime exploration for multi-armed bandits using confidence information. In *Proceedings of the International Conference on Machine Learning (ICML)*, volume 48, pages 974–982, 2016. ISBN 9781510829008.

Kwang-Sung Jun and Chicheng Zhang. Crush Optimism with Pessimism: Structured Bandits Beyond Asymptotic Optimality. *ICML Workshop on Theoretical Foundations of Reinforcement Learning (arXiv:2006.08754)*, 2020.

Kwang Sung Jun and Chicheng Zhang. PopArt: Efficient Sparse Regression and Experimental Design for Optimal Sparse Linear Bandits. *Advances in Neural Information Processing Systems (NeurIPS)*, 2022-December, 2022.

Shivaram Kalyanakrishnan and Peter Stone. Efficient selection of multiple bandit arms: Theory and practice. In *International Conference on Machine Learning (ICML)*, 2010.

Shivaram Kalyanakrishnan, Ambuj Tewari, Peter Auer, and Peter Stone. PAC subset selection in stochastic multi-armed bandits. In *International Conference on Machine Learning (ICML)*, volume 12, pages 655–662, 2012.

Zohar Karnin, Tomer Koren, and Oren Somekh. Almost optimal exploration in multi-armed bandits. In *International Conference on Machine Learning (ICML)*, pages 1238–1246. PMLR, 2013.

Julian Katz-Samuels and Kevin Jamieson. The true sample complexity of identifying good arms. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 108, pages 1781–1791, 2020.

Emilie Kaufmann and Shivaram Kalyanakrishnan. Information complexity in bandit subset selection. In *Conference on Learning Theory (COLT)*, pages 228–251. PMLR, 2013.

Emilie Kaufmann, Pierre Ménard, Omar Darwiche Domingues, Anders Jonsson, Edouard Leurent, and Michal Valko. Adaptive reward-free exploration. In *Algorithmic Learning Theory*, pages 865–891. PMLR, 2021.

Levente Kocsis and Csaba Szepesvári. Bandit based monte-carlo planning. In *European conference on machine learning*, pages 282–293. Springer, 2006.

Shie Mannor and John N Tsitsiklis. The sample complexity of exploration in the multi-armed bandit problem. *Journal of Machine Learning Research*, 5(Jun):623–648, 2004.

Francesco Orabona and David Pal. Coin betting and parameter-free online learning. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 29, pages 577–585, 2016.

Max Simchowitz, Kevin Jamieson, and Benjamin Recht. The Simulator: Understanding Adaptive Sampling in the Moderate-Confidence Regime. *Proceedings of Machine Learning Research*, 65:1–41, 2017.

Marta Soare, Alessandro Lazaric, and Rémi Munos. Best-arm identification in linear bandits. *Advances in Neural Information Processing Systems (NeurIPS)*, 27, 2014.

Yuan Zhou, Xi Chen, and Jian Li. Optimal PAC multiple arm identification with applications to crowdsourcing. In *International Conference on Machine Learning*, pages 217–225. PMLR, 2014.

Yinglun Zhu, Julian Katz-Samuels, and Robert Nowak. Near instance optimal model selection for pure exploration linear bandits. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 6735–6769. PMLR, 2022.