On Universally Optimal Algorithms for A/B Testing

Po-An Wang [§ 1 2] **Kaito Ariu** [2] **Alexandre Proutiere** [1]

Abstract

We study the problem of best-arm identification with fixed budget in stochastic multi-armed bandits with Bernoulli rewards. For the problem with two arms, also known as the A/B testing problem, we prove that there is no algorithm that (i) performs as well as the algorithm sampling each arm equally (referred to as the *uniform sampling* algorithm) in all instances, and that (ii) strictly outperforms uniform sampling on at least one instance. In short, there is no algorithm better than the uniform sampling algorithm. To establish this result, we first introduce the natural class of *consistent* and *stable* algorithms, and show that any algorithm that performs as well as the uniform sampling algorithm in all instances belongs to this class. The proof then proceeds by deriving a lower bound on the error rate satisfied by any consistent and stable algorithm, and by showing that the uniform sampling algorithm matches this lower bound. Our results provide a solution to the two open problems presented in (Qin, 2022). For the general problem with more than two arms, we provide a first set of results. We characterize the asymptotic error rate of the celebrated Successive Rejects (SR) algorithm (Audibert et al., 2010) and show that, surprisingly, the uniform sampling algorithm outperforms the SR algorithm in some instances.

1. Introduction

We study the problem of Fixed-Budget Best-Arm Identification (FB-BAI) in stochastic multi-armed bandits with Bernoulli rewards. In this problem, the learner sequentially pulls an arm and observes a random reward generated according to the corresponding distribution. The expected rewards of the arms are initially unknown. The learner has a fixed budget of $T \in \mathbb{N}$ pulls or samples, and after gathering these samples, she has to return what she believes to be the arm with the highest mean reward. For any $k \in [K] \coloneqq \{1, \ldots, K\}$, we denote by $\mu_k \in (0,1)$ the unknown mean reward of arm k. We assume that the best arm is unique and define the parameter set of the mean rewards as $\Lambda = \{\boldsymbol{\mu} \in (0,1)^K : \exists k : \mu_k > \mu_j, \forall j \neq k\}$. A strategy for fixed-budget best-arm identification consists of a *sampling rule* and a *decision rule*. The sampling rule determines the arm $A_t \in [K]$ to be explored in round t, based on past observations. The corresponding observed reward is $X_t \in \{0,1\}$. The arm A_t selected in round t is $\mathcal{F}_t$ measurable where $\mathcal{F}_t$ demotes the σ-algebra generated by the set of random variables $\{A_1, X_1, \ldots, A_{t-1}, X_{t-1}\}$. After T rounds, the decision rule returns an answer $\hat{\imath} \in [K]$, which is $\mathcal{F}_T$ measurable. The goal is to identifiy a strategy that minimizes the error probability defined as

$$p_{\boldsymbol{\mu},T} \coloneqq \mathbb{P}_{\boldsymbol{\mu}} \left[\hat{\imath} \neq 1(\boldsymbol{\mu})\right],$$

where $1(\boldsymbol{\mu}) \coloneqq \arg\max_k \mu_k$ denotes the unique best arm under the instance $\boldsymbol{\mu}$.

A naive strategy consists in allocating a fixed fraction of the budget to sample each arm. Once the budget is exhausted, the strategy then returns the arm with the highest empirical mean. We refer to such a strategy as a *static* algorithm (in contrast to adaptive algorithms that may select the arm to pull next based on the rewards observed so far). An example of a static strategy is the uniform sampling strategy that allocates the budget fairly among arms. Static algorithms are well-understood and in particular, their asymptotic error rates are known (Glynn & Juneja, 2004). Many adaptive sampling algorithms have been designed, see, e.g., (Audibert et al., 2010; Gabillon et al., 2012; Karnin et al., 2013; Russo, 2020; Komiyama et al., 2022; Wang et al., 2023), with the hope of an improved performance compared to static algorithms. It is still unclear whether this hope can actually be fulfilled.

Despite recent research efforts, the FB-BAI problem remains largely open (Qin, 2022). This contrasts with the two other classical learning tasks in stochastic bandits, namely regret minimization (Lai & Robbins, 1985) and best arm identification with fixed confidence (Garivier & Kaufmann, 2016). Indeed, for these tasks, asymptotic instance-specific

§Work initiated during the internship at CyberAgent. [1]EECS and Digital Futures, KTH, Stockholm, Sweden [2]CyberAgent, Tokyo, Japan. Correspondence to: Po-An Wang <wang9@kth.se>.

Proceedings of the 41st International Conference on Machine Learning, Vienna, Austria. PMLR 235, 2024.

performance limits and matching algorithms have been derived. In this paper, we aim at improving our understanding of the FB-BAI problem and more specifically at answering the following two natural questions, mentioned as open problems in (Qin, 2022).

Open problem 1. Is there an algorithm whose performance is as good as that of the uniform sampling algorithm on all instances and that strictly outperforms the latter on some instances?

Open problem 2. Can we derive an asymptotic and instance-specific error rate lower bound that (i) is satisfied by all algorithms within a wide class $\mathcal{A}$ of algorithms and that (ii) is achieved by a single algorithm in $\mathcal{A}$ on all instances?

We address both open problems in the case of the FB-BAI problem with two arms (also referred to as the A/B testing problem) with Bernoulli rewards. We also provide a first set of results towards addressing these problems in the general setting with more than two arms. More precisely our contributions are as follows.

Contributions.

(a) For the A/B testing problem, we prove that there is no algorithm strictly outperforming the uniform sampling algorithm (Theorem 2.2). To this aim, we first introduce the natural class of consistent and stable algorithms (stability here just means that the algorithm exhibits a symmetric and continuous behavior with respect to the instances). We then show that this class includes any algorithm performing as well as the uniform sampling algorithm on all instances (Theorem 3.3). We finally derive an instance-specific lower bound on the error rate satisfied by any consistent and stable (Theorem 4.1). As it turns out, this lower bound corresponds to the performance of the uniform sampling algorithm. The answer to the question of the open problem 1 is hence negative.

(b) Our analysis further provides a positive answer to the question of the open problem 2. Indeed, it yields an instance-specific error rate lower bound for the class of consistent and stable algorithms, and the performance of the uniform sampling algorithm matches this fundamental limit.

(c) For the FB-BAI problem with more than two arms, we manage to exactly characterize the asymptotic error rate of the celebrated Successive Rejects (SR) algorithm (Audibert et al., 2010) (Theorem 5.1). This contrasts with existing analyses of adaptive algorithms where only upper bounds of the error rate can be derived. Using this characterization, we show that, surprisingly, the uniform sampling algorithm outperforms the SR algorithm in certain instances (Theorem 5.2).

2. Preliminaries and Main Result

In this section, we first present existing results on the performance of static algorithms in the A/B testing problem. We then state our main result: there is no strictly better algorithm than the uniform sampling algorithm.

2.1. Performance of Static Algorithms

In two-arm bandits, a static algorithm is parameterized by a single variable $x \in (0,1)$ specifying the fraction of the budget used to sample the second arm (a static algorithm parameterized by x pulls the first arm $(1-x)T+o(T)$ times and pulls the second arm $xT + o(T)$ times). Defining

$$g(x, \boldsymbol{\mu}) \coloneqq \min_{\lambda \in [0,1]} (1-x)d(\lambda, \mu_1) + xd(\lambda, \mu_2), \quad (1)$$

where $d(a,b)$ is the KL-divergence between two Bernoulli distributions with respective means a and b, (Glynn & Juneja, 2004) shows that under a static algorithm parametrized by x,

$$\lim_{T\to\infty} \frac{T}{\log(1/p_{\boldsymbol{\mu},T})} = \frac{1}{g(x,\boldsymbol{\mu})}.$$

The optimization problem in (1) can be solved by explicitly writing the KKT conditions (Glynn & Juneja, 2004). Its unique solution and value are given by

$$\lambda(x,\boldsymbol{\mu}) = \frac{(\frac{\mu_1}{1-\mu_1})^{1-x}(\frac{\mu_2}{1-\mu_2})^{x}}{1+(\frac{\mu_1}{1-\mu_1})^{1-x}(\frac{\mu_2}{1-\mu_2})^{x}} \in (0,1), \quad (2)$$

$$g(x,\boldsymbol{\mu}) = -\log((1-\mu_1)^{1-x}(1-\mu_2)^{x} + \mu_1^{1-x}\mu_2^{x}). \quad (3)$$

From (3), we can readily verify that $g(x,\boldsymbol{\mu})$ is strictly concave in x, i.e., $\frac{\partial^2 g(x,\boldsymbol{\mu})}{\partial^2 x} < 0$ as long as $\mu_1 \neq \mu_2$ or equivalently $\boldsymbol{\mu} \in \Lambda$.[1] Therefore, $g(x,\boldsymbol{\mu})$ has a unique maximizer denoted by $x^*(\boldsymbol{\mu}) \coloneqq \operatorname{argmax}_x g(x,\boldsymbol{\mu})$. Given the expected rewards $\boldsymbol{\mu}$ of the arms, $x^*(\boldsymbol{\mu})$ corresponds to the static algorithm with the best possible performance. However, under this static algorithm, the fraction of the budget used for each arm depends on the initially unknown $\boldsymbol{\mu}$.

Over the last few years, researchers have tried to determine whether there exists an adaptive algorithm that could achieve the performance of the best static algorithm for any $\boldsymbol{\mu}$. The answer to this question is actually negative, as recently proved in (Degenne, 2023): for any algorithm, there exists an instance $\boldsymbol{\mu}$ such that the considered algorithm performs strictly worse than the best static algorithm on this instance. Refer to Section 6 for additional details. This negative result illustrates the difficulty of devising adaptive and efficient algorithms. We establish even more striking evidence of this challenge. We show that there is no algorithm universally outperforming the uniform sampling algorithm. We formalize this result below.

[1] For completeness, we include proof of (2), (3), and the strict concavity of $g(x,\boldsymbol{\mu})$ in Appendix A.

2.2. Main Result

The performance of the uniform sampling algorithm is characterized by $g(1/2, \boldsymbol{\mu})$. More precisely, under this algorithm,

$$\forall \boldsymbol{\mu} \in \Lambda, \quad \lim_{T\to\infty} \frac{T}{\log(1/p_{\boldsymbol{\mu},T})} = \frac{1}{g(1/2,\boldsymbol{\mu})}.$$

Our main result concerns the class of *better than uniform* algorithms already introduced and discussed in (Qin, 2022). These algorithms are at least as good as the uniform sampling algorithm in all instances.

Definition 2.1. An algorithm is *better than uniform* if

$$\forall \boldsymbol{\mu} \in \Lambda, \quad \overline{\lim_{T\to\infty}} \frac{T}{\log(1/p_{\boldsymbol{\mu},T})} \leq \frac{1}{g(1/2,\boldsymbol{\mu})}.$$

Theorem 2.2. *For any better than uniform algorithm,*

$$\forall \boldsymbol{\mu} \in \Lambda, \quad \lim_{T\to\infty} \frac{T}{\log(1/p_{\boldsymbol{\mu},T})} = \frac{1}{g(1/2,\boldsymbol{\mu})}.$$

As a consequence, surprisingly, one cannot devise an adaptive algorithm that performs as well as the uniform sampling algorithm on all instances and that strictly outperforms it on some instances. This also implies that if an algorithm strictly outperforms the uniform sampling algorithm in at least one instance, then there is an instance where the uniform sampling algorithm strictly outperforms this algorithm. This provides a solution to the open problem 1 (Qin, 2022) presented in the introduction (refer to Section 6.1 for more details).

Theorem 2.2 is proved by combining the results presented in Sections 3 and 4. There, we introduce the class of consistent and stable algorithms, and show that better than uniform algorithms are consistent and stable. We also establish the error rate achieved by the uniform sampling algorithm constitutes an error rate lower bound satisfied by consistent and stable algorithms. Note that these intermediate results towards Theorem 2.2 provide a solution to the open problem 2 (Qin, 2022) presented in the introduction.

2.3. Notation

For each $t \in \{1, 2, \dots, T\}$ and $k \in \{1, 2\}$, define $N_k(t) := \sum_{s=1}^{t} \mathbb{1}\{A_s = k\}$ as the number of times arm k is pulled up to round t, and $\omega_k(t) := N_k(t)/t$ as the proportion of times arm k is pulled.

3. Stable and Consistent Algorithms

In this section, we demonstrate that any better than uniform algorithm is both consistent and stable, as defined below.

Definition 3.1. An algorithm is *consistent* if for all $\boldsymbol{\mu} \in \Lambda$, $\lim_{T\to\infty} p_{\boldsymbol{\mu},T} = 0$.

Definition 3.2. An algorithm is *stable* if for any $a \in (0, 1)$, the following properties hold:
(A) There exists $\{\boldsymbol{\lambda}^{(n)}\}_{n=1}^{\infty} \subset \{\boldsymbol{\lambda} \in \Lambda : \lambda_1 > \lambda_2\}$ such that $\boldsymbol{\lambda}^{(n)} \xrightarrow{n\to\infty} (a, a)$ and

$$\lim_{n\to\infty} \underline{\lim_{T\to\infty}} \mathbb{E}_{\boldsymbol{\lambda}^{(n)}}[\omega_2(T)] = \lim_{n\to\infty} \overline{\lim_{T\to\infty}} \mathbb{E}_{\boldsymbol{\lambda}^{(n)}}[\omega_2(T)] = \frac{1}{2}.$$

(B) There exists $\{\boldsymbol{\pi}^{(n)}\}_{n=1}^{\infty} \subset \{\boldsymbol{\pi} \in \Lambda : \pi_1 < \pi_2\}$ such that $\boldsymbol{\pi}^{(n)} \xrightarrow{n\to\infty} (a, a)$ and

$$\lim_{n\to\infty} \underline{\lim_{T\to\infty}} \mathbb{E}_{\boldsymbol{\pi}^{(n)}}[\omega_2(T)] = \lim_{n\to\infty} \overline{\lim_{T\to\infty}} \mathbb{E}_{\boldsymbol{\pi}^{(n)}}[\omega_2(T)] = \frac{1}{2}.$$

Intuitively, an algorithm is stable if it exhibits a symmetric and continuous behavior with respect to the bandit instances. The notion of stability is natural and just refers to the property of evenly allocating the budget when the arms have very similar mean rewards. It is satisfied by the uniform sampling algorithm (and of course all the adaptive algorithms that evenly select arms in the case of two-armed bandits) and by most *reasonably adaptive* algorithms. We give several families of stable algorithms in Appendix E. For example, stability is guaranteed as soon as the algorithm is designed such that the number of times arm 1 is sampled up to time t closely matches $tf(\hat{\mu}_1(t), \hat{\mu}_2(t))$, where $\hat{\mu}_1(t)$ and $\hat{\mu}_2(t)$ are the current estimates of the mean rewards and $f > 0$ is a continuous function such that $f(a, a) = 1/2$ for any $a \in (0, 1)$.

In addition, as established in the following theorem, better than uniform algorithms are stable.

Theorem 3.3. *A better than uniform algorithm is consistent and stable.*

3.1. Proof of Theorem 3.3

Consistency. In view of Definition 2.1, a better than uniform algorithm is consistent. Indeed, for any $\boldsymbol{\mu} \in \Lambda$, there exists $T_{\boldsymbol{\mu}} \in \mathbb{N}$ such that if $T > T_{\boldsymbol{\mu}}$, then $\frac{T}{\log(1/p_{\boldsymbol{\mu},T})} \leq \frac{2}{g(1/2,\boldsymbol{\mu})}$. As a result, we have $p_{\boldsymbol{\mu},T} \leq e^{-\frac{1}{2}g(1/2,\boldsymbol{\mu})T}$. We conclude the proof by observing that $g(1/2, \boldsymbol{\mu}) > 0$.

Stability. We show by contradiction that a better than uniform algorithm is stable. Suppose there exists $a \in (0, 1)$ such that (B) in Definition 3.2 does not hold (if (A) in Definition 3.2 does not hold, one can obtain a contradiction in a symmetric way). The following lemma is proven in Appendix C.

Lemma 3.4. *Let $a \in (0, 1)$. Assume that the statement* (B) *of Definition 3.2 does not hold. Then for any $\{\boldsymbol{\pi}^{(n)}\}_{n=1}^{\infty} \subset \{\boldsymbol{\pi} \in \Lambda : \pi_1 < \pi_2\}$ such that $\boldsymbol{\pi}^{(n)} \xrightarrow{n\to\infty} (a, a)$, there exists a value $x \in [0, 1]$ and increasing sequences of integers $\{n_m\}_{m=1}^{\infty}, \{T_{m,\ell}\}_{\ell=1}^{\infty} \subset \mathbb{N}$ such that*

$$\lim_{m\to\infty} \lim_{\ell\to\infty} \mathbb{E}_{\boldsymbol{\pi}^{(n_m)}}[\omega_2(T_{m,\ell})] = x \neq \frac{1}{2}.$$

Let x, $\{n_m\}_{m=1}^{\infty}$, and $\{T_{m,\ell}\}_{\ell=1}^{\infty}$ be a real number and sequences satisfying the statement of Lemma 3.4. Using a standard change-of-measure argument (e.g., inequality (6) in (Garivier et al., 2019)), for any $\boldsymbol{\mu} \in \Lambda$ with $\mu_1 > \mu_2$, for each $m, \ell \in \mathbb{N}$,

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{\pi}^{(n_m)}}[N_1(T_{m,\ell})]d(\pi_1^{(n_m)}, \mu_1) \\
&\quad + \mathbb{E}_{\boldsymbol{\pi}^{(n_m)}}[N_2(T_{m,\ell})]d(\pi_2^{(n_m)}, \mu_2) \\
&\geq d(\mathbb{P}_{\boldsymbol{\pi}^{(n_m)}}[\hat{\imath} = 2], \mathbb{P}_{\boldsymbol{\mu}}[\hat{\imath} = 2]) \\
&\geq \mathbb{P}_{\boldsymbol{\pi}^{(n_m)}}[\hat{\imath} = 2] \log\left(1/p_{\boldsymbol{\mu},T_{m,\ell}}\right) - \log 2, \qquad (4)
\end{aligned}$$

where the last inequality stems from the fact that $d(p, q) \geq p \log(1/q) - \log 2$ for all $p, q \in [0, 1]$. By dividing both sides of equation (4) by $T_{m,\ell}$, we can rearrange the inequality as follows:

$$\begin{aligned}
&\frac{T_{m,\ell}}{\mathbb{P}_{\boldsymbol{\pi}^{(n_m)}}[\hat{\imath} = 2] \log(1/p_{\boldsymbol{\mu},T_{m,\ell}})} \\
&\geq \Bigg(\mathbb{E}_{\boldsymbol{\pi}^{(n_m)}}[\omega_1(T_{m,\ell})]d(\pi_1^{(n_m)}, \mu_1) \\
&\qquad + \mathbb{E}_{\boldsymbol{\pi}^{(n_m)}}[\omega_2(T_{m,\ell})]d(\pi_2^{(n_m)}, \mu_2) + \frac{\log 2}{T_{m,\ell}} \Bigg)^{-1}.
\end{aligned}$$

Given that a better than uniform algorithm is consistent, it follows that $\mathbb{P}_{\boldsymbol{\pi}^{(n_m)}}[\hat{\imath} = 2] = 1 - p_{\boldsymbol{\pi}^{(n_m)},T_{m,\ell}} \xrightarrow{\ell \to \infty} 1$ from $\pi_1^{(n_m)} < \pi_2^{(n_m)}$. Driving $\ell \to \infty$ first, and then letting $m \to \infty$, we obtain

$$\begin{aligned}
&\varliminf_{m \to \infty} \varliminf_{\ell \to \infty} \frac{T_{m,\ell}}{\log(1/p_{\boldsymbol{\mu},T_{m,\ell}})} \\
&\geq \frac{1}{(1-x)d(a, \mu_1) + x d(a, \mu_2)}. \qquad (5)
\end{aligned}$$

Next, we use the following lemma related to the function g and prove after completing the proof of Theorem 3.3. Lemma 3.5 is visualized in the left-hand side of Figure 1.

Lemma 3.5. *For any $a \in (0, 1)$, $x \in [0, 1]$ such that $x \neq 1/2$, there exists $\boldsymbol{\mu} \in \Lambda$ such that $\mu_1 > \mu_2$, $\lambda(x, \boldsymbol{\mu}) = a$, and $g(x, \boldsymbol{\mu}) < g(1/2, \boldsymbol{\mu})$.*

Plugging such $\boldsymbol{\mu}$ into (5) yields that

$$\varliminf_{m \to \infty} \varliminf_{\ell \to \infty} \frac{T_{m,\ell}}{\log(1/p_{\boldsymbol{\mu},T_{m,\ell}})} \geq \frac{1}{g(x, \boldsymbol{\mu})} > \frac{1}{g(1/2, \boldsymbol{\mu})}.$$

This contradicts the assumption that the algorithm is better than uniform. □

Proof of Lemma 3.5. We assume that $x \in (1/2, 1]$. The case for $x \in [0, 1/2)$ will be addressed at the end. We first present Proposition 3.6, whose proof is given in Appendix A, and its visualization is shown in the right panel of Figure 1.

Proposition 3.6. *For any $a \in (0, 1), x \in (1/2, 1]$, there exists an instance $\boldsymbol{\mu} \in \Lambda$ such that* (i) $\mu_1 > \mu_2$, $\mu_1 + \mu_2 \geq 1$, (ii) $\lambda(x, \boldsymbol{\mu}) = a$, *and* (iii) $x^*(\boldsymbol{\mu}) < (1/2 + x)/2$.

Let $\boldsymbol{\mu} \in \Lambda$ be an instance satisfying the conditions of Proposition 3.6. If $x^*(\boldsymbol{\mu}) \leq 1/2$, the strict concavity of $g(\cdot, \boldsymbol{\mu})$ immediately implies that $g(1/2, \boldsymbol{\mu}) > g(x, \boldsymbol{\mu})$. On the other hand, if $x^*(\boldsymbol{\mu}) > 1/2$, we can observe that $x^*(\boldsymbol{\mu}) < (1/2 + x)/2 \leq 3/4$, which leads to $\delta = x^*(\boldsymbol{\mu}) - 1/2 \leq \min\{x^*(\boldsymbol{\mu}), 1 - x^*(\boldsymbol{\mu})\}$. We use the following proposition.

Proposition 3.7. *Suppose $\mu_1 > \mu_2$ and $\mu_1 + \mu_2 \geq 1$. For any positive $\delta \leq \min\{x^*(\boldsymbol{\mu}), 1 - x^*(\boldsymbol{\mu})\}$, $g(x^*(\boldsymbol{\mu}) - \delta, \boldsymbol{\mu}) \geq g(x^*(\boldsymbol{\mu}) + \delta, \boldsymbol{\mu})$.*

The proof and visualization of Proposition 3.7 can be found in Appendix B and Figure 2, respectively. Setting $\delta = x^*(\boldsymbol{\mu}) - 1/2$, we obtain the following inequality

$$g(1/2, \boldsymbol{\mu}) = g(x^*(\boldsymbol{\mu}) - \delta, \boldsymbol{\mu}) \geq g(x^*(\boldsymbol{\mu}) + \delta, \boldsymbol{\mu}).$$

Using the strict concavity of $g(\cdot, \boldsymbol{\mu})$ again and the fact that $x^*(\boldsymbol{\mu}) + \delta = 2x^*(\boldsymbol{\mu}) - 1/2 < x$, we derive that $g(1/2, \boldsymbol{\mu}) \geq g(x^*(\boldsymbol{\mu}) + \delta, \boldsymbol{\mu}) > g(x, \boldsymbol{\mu})$. This concludes the proof when $x \in (1/2, 1]$.

Next, we consider the proof when $x \in [0, 1/2)$. To this aim, we use the following symmetrical property of $g(x, \boldsymbol{\mu})$.

Proposition 3.8. *Denote $\bar{\boldsymbol{\mu}} = (1 - \mu_2, 1 - \mu_1)$. For any $x \in (0, 1)$, for any $\boldsymbol{\mu} \in \Lambda$, $g(1 - x, \bar{\boldsymbol{\mu}}) = g(x, \boldsymbol{\mu})$ and $\lambda(1 - x, \bar{\boldsymbol{\mu}}) = 1 - \lambda(x, \boldsymbol{\mu})$.*

The proof of Proposition 3.8 is presented in Appendix A.3. The previous proof (replacing x with $1 - x \in (1/2, 1]$) yields the existence of $\boldsymbol{\mu} \in \Lambda$ such that $g(1 - x, \boldsymbol{\mu}) < g(1/2, \boldsymbol{\mu})$ and $\lambda(1 - x, \boldsymbol{\mu}) = 1 - a$. Let $\bar{\boldsymbol{\mu}} = (1 - \mu_2, 1 - \mu_1)$, Proposition 3.8 and the strict concavity of $g(\cdot, \boldsymbol{\mu})$ imply that

$$g(x, \bar{\boldsymbol{\mu}}) = g(1 - x, \boldsymbol{\mu}) < g(1/2, \boldsymbol{\mu}) = g(1/2, \bar{\boldsymbol{\mu}}),$$

and $\lambda(x, \bar{\boldsymbol{\mu}}) = 1 - \lambda(1 - x, \boldsymbol{\mu}) = a$. This concludes the proof for $x \in [0, 1/2)$, thus completing the proof of Lemma 3.5. □

4. Error Rate of Consistent and Stable Algorithms

In this section, we establish that the performance of any stable and consistent algorithm is either equivalent to or worse than that of uniform sampling.

Theorem 4.1. *If an algorithm is consistent and stable, then*

$$\forall \boldsymbol{\mu} \in \Lambda, \quad \varliminf_{T \to \infty} \frac{T}{\log(1/p_{\boldsymbol{\mu},T})} \geq \frac{1}{g(1/2, \boldsymbol{\mu})}.$$

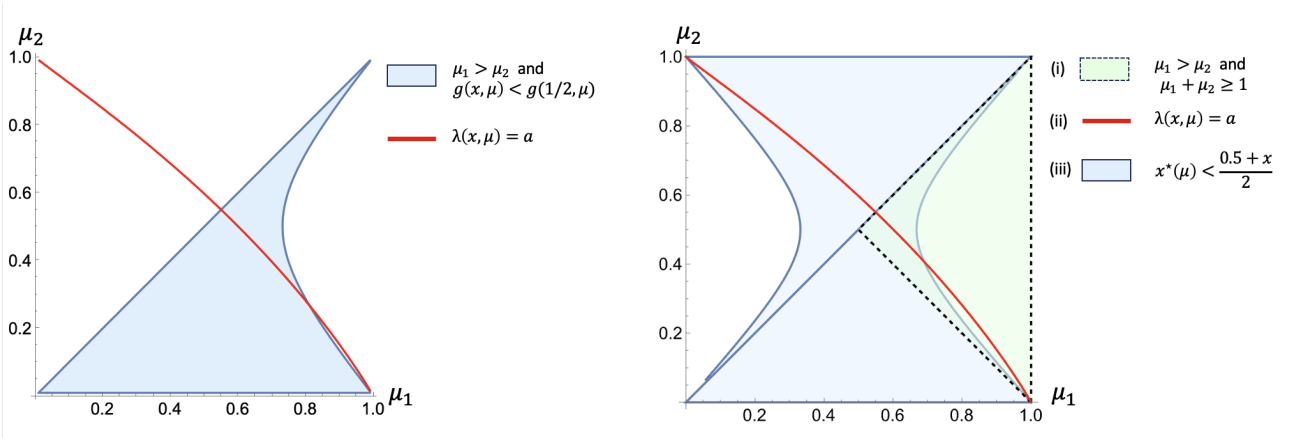

Figure 1. Left: Visualization of Lemma 3.5 with $a = 0.55$ and $x = 0.51$. The blue region indicates where $\mu_1 > \mu_2$ and $g(x, \boldsymbol{\mu}) < g(1/2, \boldsymbol{\mu})$. The red curve represents $\lambda(x, \boldsymbol{\mu}) = a$. The intersection of the blue region and red curve validates Lemma 3.5. Right: Visualization of Proposition 3.6 with $a = 0.55$ and $x = 0.51$. The green region indicates (i) $\mu_1 > \mu_2$, $\mu_1 + \mu_2 \geq 1$. The red curve represents (ii) $\lambda(x, \boldsymbol{\mu}) = a$. The blue region shows (iii) $x^*(\boldsymbol{\mu}) < (\frac{1}{2} + x)/2$. The intersection of the three regions validates Proposition 3.6.

Proof. Without loss of generality, assume $1(\boldsymbol{\mu}) = 1$, namely, $\mu_1 > \mu_2$. For any $\boldsymbol{\pi} \in \Lambda$ such that $\pi_1 < \pi_2$, applying a standard change-of-measure argument, as in the proof of Theorem 3.3, yields that

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{\pi}}[N_1(T)]d(\pi_1, \mu_1) + \mathbb{E}_{\boldsymbol{\pi}}[N_2(T)]d(\pi_2, \mu_2) \\
&\geq d(\mathbb{P}_{\boldsymbol{\pi}}[\hat{\imath} = 2], \mathbb{P}_{\boldsymbol{\mu}}[\hat{\imath} = 2]) \\
&\geq \mathbb{P}_{\boldsymbol{\pi}}[\hat{\imath} = 2] \log(1/p_{\boldsymbol{\mu},T}) - \log 2. \qquad (6)
\end{aligned}$$

Since $\pi_1 < \pi_2$, the consistent assumption yields that $\mathbb{P}_{\boldsymbol{\pi}}[\hat{\imath} = 2] = 1 - p_{\boldsymbol{\pi},T} \xrightarrow{T\to\infty} 1$. Dividing the both sides of (6) by T and taking $T \to \infty$, we obtain

$$\begin{aligned}
&\varliminf_{T\to\infty} \frac{T}{\log(1/p_{\boldsymbol{\mu},T})} \qquad (7) \\
&\geq \varliminf_{T\to\infty} \frac{1}{\mathbb{E}_{\boldsymbol{\pi}}[\omega_1(T)]d(\pi_1, \mu_1) + \mathbb{E}_{\boldsymbol{\pi}}[\omega_2(T)]d(\pi_2, \mu_2)}
\end{aligned}$$

by simply rearranging the terms. Next, we let $a = \lambda(\frac{1}{2}, \boldsymbol{\mu})$. Since the algorithm is stable, there exists $\{\boldsymbol{\pi}^{(n)}\}_{n=1}^{\infty} \subset \Lambda$ such that $\pi_1^{(n)} < \pi_2^{(n)}$, $\boldsymbol{\pi}^{(n)} \xrightarrow{n\to\infty} (a, a)$, and

$$\lim_{n\to\infty} \varliminf_{T\to\infty} \mathbb{E}_{\boldsymbol{\pi}^{(n)}}[\omega_2(T)] = \lim_{n\to\infty} \overline{\lim_{T\to\infty}} \mathbb{E}_{\boldsymbol{\pi}^{(n)}}[\omega_2(T)] = \frac{1}{2}.$$

Notice that previous equations also imply that

$$\begin{aligned}
\lim_{n\to\infty} \varliminf_{T\to\infty} \mathbb{E}_{\boldsymbol{\pi}^{(n)}}[\omega_2(T)] &= \lim_{n\to\infty} \varliminf_{T\to\infty} (1 - \mathbb{E}_{\boldsymbol{\pi}^{(n)}}[\omega_1(T)]) \\
&= 1 - \lim_{n\to\infty} \overline{\lim_{T\to\infty}} \mathbb{E}_{\boldsymbol{\pi}^{(n)}}[\omega_1(T)] \\
&= \frac{1}{2}. \qquad (8)
\end{aligned}$$

Thus, rearranging (8) yields that

$$\lim_{n\to\infty} \overline{\lim_{T\to\infty}} \mathbb{E}_{\boldsymbol{\pi}^{(n)}}[\omega_1(T)] = \frac{1}{2}.$$

Plugging them into (7) and taking n to infinity yields that

$$\begin{aligned}
\varliminf_{T\to\infty} \frac{T}{\log(1/p_{\boldsymbol{\mu},T})} &\geq \lim_{n\to\infty} (A_n)^{-1} = \Big(\lim_{n\to\infty} A_n\Big)^{-1} \\
&\geq \frac{2}{d(a, \mu_1) + d(a, \mu_2)} = \frac{1}{g(1/2, \boldsymbol{\mu})}.
\end{aligned}$$

where A_n is the limit superior of the sequence

$$\big(\mathbb{E}_{\boldsymbol{\pi}^{(n)}}[\omega_1(T)]d(\pi_1^{(n)}, \mu_1) + \mathbb{E}_{\boldsymbol{\pi}^{(n)}}[\omega_2(T)]d(\pi_2^{(n)}, \mu_2)\big)_{T\in\mathbb{N}}.$$

□

We remark that the combination of Theorems 3.3 and 4.1 leads to Theorem 2.2.

5. K-Armed Bandits with $K \geq 3$

Extending our results to the general case where there are more than two arms is challenging. We investigate whether existing adaptive algorithms could be better than uniform algorithms. This question is not easy to answer because existing analyses of these algorithms provide *upper* bounds only on their error rates. Even if these upper bounds are, for some instances, worse than the error rate of the uniform sampling algorithm, it does not imply that the latter performs better on these instances. To answer the question, we also need to derive *lower* bounds on their error rates (which is challenging – refer to (Wang et al., 2023) for a detailed discussion).

In this section, we restrict our attention to the celebrated Successive Rejects (SR) algorithm (Audibert et al., 2010), and

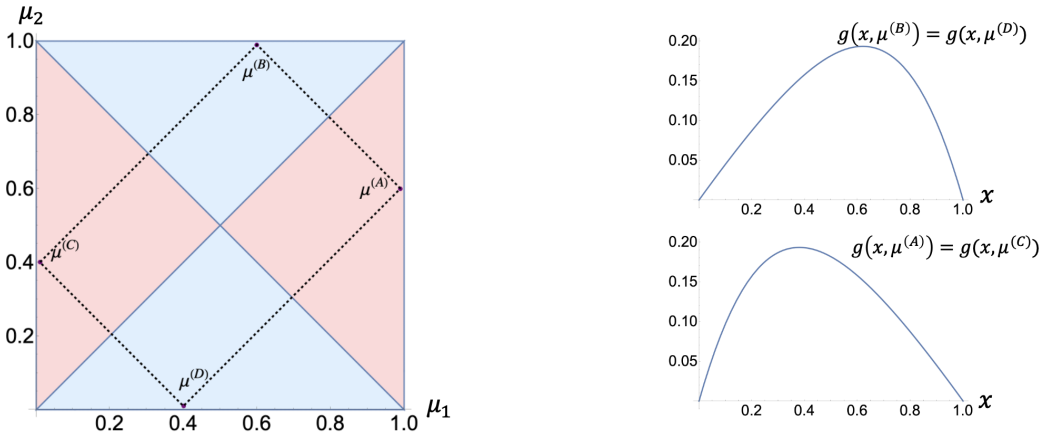

Figure 2. Visualization of the function $g(x, \boldsymbol{\mu})$ properties. The left panel shows the partition of Λ into four regions by $\mu_1 = \mu_2$ and $\mu_1 + \mu_2 = 1$, with blue indicating $x^*(\boldsymbol{\mu}) < 1/2$ and red indicating $x^*(\boldsymbol{\mu}) > 1/2$. Four bandit instances are chosen symmetrically from these regions for further analysis. The right panels show the functions $g(x, \boldsymbol{\mu}^{(A)}) = g(x, \boldsymbol{\mu}^{(C)})$ (top) and $g(x, \boldsymbol{\mu}^{(B)}) = g(x, \boldsymbol{\mu}^{(D)})$ (bottom), demonstrating the asymmetrical property as stated in Proposition 3.7.

we manage to derive the exact expression of its asymptotic error rate. From there, we exhibit instances where surprisingly, the uniform sampling algorithm provably outperforms the SR algorithm.

5.1. The Successive Rejects Algorithm

Algorithm 1 Successive Rejects (SR)

Initialization $\mathcal{C}_K \leftarrow [K]$, $j \leftarrow K$;
for $t = 1, 2, \ldots, T$ **do**
  if ($j > 2$ and $\min_{k \in \mathcal{C}_j} N_k(t) \geq \frac{T}{j\overline{\log} K}$) **then**
    $\ell_j \leftarrow \operatorname{argmin}_{k \in \mathcal{C}_j} \hat{\mu}_k(t)$ (tie broken arbitrarily), $\mathcal{C}_{j-1} \leftarrow \mathcal{C}_j \setminus \{\ell_j\}$, and $j \leftarrow j - 1$
  end if
  Sample $A_t \leftarrow \operatorname{argmin}_{k \in \mathcal{C}_j} N_k(t)$ (tie broken arbitrarily), update $\{N_k(t)\}_{k \in \mathcal{C}_j}$ and $\hat{\boldsymbol{\mu}}(t)$
end for
$\ell_2 \leftarrow \arg\min_{k \in \mathcal{C}_2} \hat{\mu}_k(T)$ (tie broken arbitrarily)
Return $\hat{\imath} \leftarrow \arg\max_{k \in \mathcal{C}_2} \hat{\mu}_k(T)$ (tie broken arbitrarily)

Consider the FB-BAI problem with K arms as described in the introduction. For this problem, the SR algorithm starts by initializing the set of candidate arms as $\mathcal{C}_K = [K]$. The sampling budget is partitioned into $K - 1$ phases. Following each phase, SR discards the empirically determined worst-performing arm from the candidate set. During each phase, SR adopts a uniform sampling strategy for the arms within the candidate set.

The phase lengths are determined as follows. Define $\overline{\log} K := 1/2 + \sum_{k=2}^{K} 1/k$. When the candidate set, denoted by $\mathcal{C}_j$, comprises more than two arms, i.e., $j > 2$, in the corresponding phase, SR works as follows: (i) each arm within $\mathcal{C}_j$ is sampled until the round t at which $\min_{k \in \mathcal{C}_j} N_k(t) \geq T/(j\overline{\log} K)$. (ii) the arm identified as the empirical worst, denoted by ℓ_j, is then discarded, which means $\mathcal{C}_{j-1} = \mathcal{C}_j \setminus \{\ell_j\}$. During the final phase, SR samples the two remaining arms evenly and recommends $\hat{\imath}$, the arm that exhibits the higher empirical mean in $\mathcal{C}_2$. Algorithm 1 presents the pseudo-code of SR.

5.2. Exact Analysis of SR

In Theorem 2 in (Wang et al., 2023), the authors show that SR satisfies for any $\boldsymbol{\mu} \in \Lambda$

$$\varlimsup_{T \to \infty} \frac{T}{\log(1/p_{\boldsymbol{\mu},T})} \leq \max_{j=2,\ldots,K} \frac{j\overline{\log} K}{\Gamma_j(\boldsymbol{\mu})}. \tag{9}$$

where

$$\Gamma_j(\boldsymbol{\mu}) := \min_{J \in \mathcal{J}_j(\boldsymbol{\mu})} \inf_{\boldsymbol{\lambda} \in \Lambda : \lambda_{1(\boldsymbol{\mu})} \leq \min_{k \in J} \lambda_k} \sum_{k \in J} d(\lambda_k, \mu_k),$$

and $\mathcal{J}_j(\boldsymbol{\mu}) := \{J \subseteq [K] : |J| = j, 1(\boldsymbol{\mu}) \in J\}$. We show the bound (9) is in fact tight. Indeed, in the following theorem whose proof is presented in Appendix D, we derive a matching lower bound.

Theorem 5.1. *Under the Successive Rejects algorithm (Audibert et al., 2010), for any $\boldsymbol{\mu} \in \Lambda$,*

$$\lim_{T \to \infty} \frac{T}{\log(1/p_{\boldsymbol{\mu},T})} = \max_{j=2,\ldots,K} \frac{j\overline{\log} K}{\Gamma_j(\boldsymbol{\mu})}.$$

5.3. Instances Where Uniform Sampling Outperforms SR

We can use Theorem 5.1 to assess whether the SR algorithm is better than uniform. The next theorem shows that it is not even for three-armed bandits.

Theorem 5.2. *There exists a three-armed bandit instance in which uniform sampling strictly outperforms SR asymptotically.*

Proof. From Theorem 5.1, for any $\boldsymbol{\mu} \in \Lambda$, the error rate of SR satisfies

$$\lim_{T\to\infty} \frac{T}{\log(1/p_{\boldsymbol{\mu},T})} = \max\left\{\frac{8}{3\Gamma_2(\boldsymbol{\mu})}, \frac{4}{\Gamma_3(\boldsymbol{\mu})}\right\}. \quad (10)$$

As for uniform sampling, the error rate satisfies (Glynn & Juneja, 2004)

$$\lim_{T\to\infty} \frac{T}{\log(1/p_{\boldsymbol{\mu},T})} = \frac{3}{\Gamma_2(\boldsymbol{\mu})}. \quad (11)$$

Thank to Proposition D.4 in Appendix D.1, we can compute $\Gamma_2(0.5, 0.3, 0.3) \approx 0.0426387$ and $\Gamma_3(0.5, 0.3, 0.3) \approx 0.0562588$, which implies

$$(10) \approx 71.1 > 70.3587 \approx (11).$$

We conclude that SR has a higher error rate than the uniform sampling algorithm in the instance $(0.5, 0.3, 0.3)$. □

We can use the results from Proposition D.4 in Appendix D.1 to numerically compare the error rates of the SR and uniform sampling algorithms. In Figure 3, we plot the set of instances $\boldsymbol{\mu}$ such $\mu_1 > \mu_2 = \mu_3$ where the SR algorithm has an higher error rate than the uniform sampling algorithm.

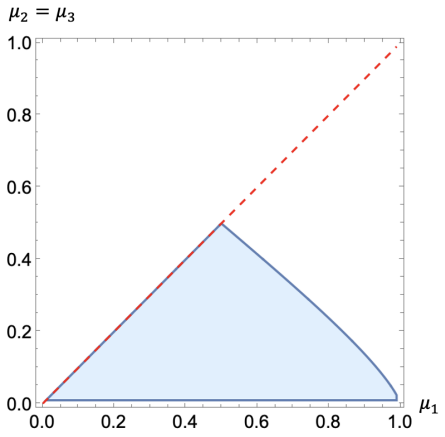

Figure 3. The blue area corresponds to instances $\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_2)$ such that $\mu_1 > \mu_2 = \mu_3$ and such that the uniform sampling algorithm strictly outperforms the SR algorithm. The red dashed line is the set of instances such that $\mu_1 = \mu_2$.

6. Related Work and Discussion

In this section, we start by stating rigorously the two open problems in (Qin, 2022) and that we address in this paper. We then present the related work for the FB-BAI problem in general, and finally discuss existing results for the two-arm case.

6.1. Open Problems Stated in (Qin, 2022)

Problem 1 (Qin, 2022) consists of investigating whether there exists a better than uniform algorithm that strictly outperforms uniform sampling on some instances. Based on Theorem 2.2, we can conclude that no such algorithms exist in two-armed bandits.

Problem 2 (Qin, 2022) consists in investigating whether the two following properties can hold simultaneously:

(a) *Lower bound.* There exist an algorithm class $\mathcal{A}$ and a function $\Gamma^* : \Lambda \mapsto \mathbb{R}$ such that for any algorithm in $\mathcal{A}$,

$$\forall \boldsymbol{\mu} \in \Lambda, \quad \underline{\lim}_{T\to\infty} \frac{T}{\log(1/p_{\boldsymbol{\mu},T})} \geq \Gamma^*(\boldsymbol{\mu}).$$

(b) *Upper bound.* There is a single algorithm in $\mathcal{A}$ satisfies

$$\forall \boldsymbol{\mu} \in \Lambda, \quad \overline{\lim}_{T\to\infty} \frac{T}{\log(1/p_{\boldsymbol{\mu},T})} \leq \Gamma^*(\boldsymbol{\mu}).$$

For problems with more than two arms, (Garivier & Kaufmann, 2016) conjecture that the lower bound discussed above for two-arm bandits can be generalized. However, again, (Ariu et al., 2021) prove that for any algorithm, when the number of arms is large, there exists at least one instance where the algorithm cannot reach the lower bound. Our Theorem 4.1 addresses this open problem by considering $\mathcal{A}$ as the set of consistent and stable algorithms, by setting

$$\Gamma^*(\boldsymbol{\mu}) = \frac{1}{g(1/2, \boldsymbol{\mu})},$$

and by using the fact that the uniform sampling algorithm matches the corresponding lower bound.

6.2. Fixed-Budget Best-Arm Identification

The study of fixed-budget best-arm identification has been relatively recent (Audibert et al., 2010; Bubeck et al., 2011), especially when compared to regret minimization (Lai & Robbins, 1985; Cappé et al., 2013) and fixed-confidence best-arm identification (Chernoff, 1959; Even-Dar et al., 2006; Garivier & Kaufmann, 2016). Since then, algorithms such as Successive Rejects (SR) (Audibert et al., 2010), Sequential Halving (Karnin et al., 2013), and UGapE (Gabillon et al., 2012) have been proposed with performance guarantees.

A first lower bound on the error rate for FB-BAI has been proposed in (Audibert et al., 2010). They prove that for any algorithm that knows the reward distributions of the arms, but does not know the order in which they correspond to the arms, there exists a bandit instance such that the probability of misidentification is lower bounded by

$\exp(-\frac{cT}{H_2})$, where $c > 0$ is some universal constant and $H_2 := \max_{k \neq 1(\boldsymbol{\mu})} \frac{k}{(\mu_{1(\boldsymbol{\mu})} - \mu_k)^2}$.

In (Carpentier & Locatelli, 2016), the authors prove that for any algorithm, there exist bandit instances such that the probability of error for one of the instances is lower bounded by $\exp(-\frac{400T}{H_2 \log K})$, where K is the number of the arms. The authors of (Ariu et al., 2021) revisit this lower bound, and demonstrate that no algorithm can universally achieve the same error probability as the best static algorithm, particularly when the number of arms is large.

(Komiyama et al., 2022) presents a minimax characterization of the error probability. They also conjecture that exploration that adapts to the instance is costly.

The recent study (Wang et al., 2023) investigates the FB-BAI problem using large deviation techniques. They relate the error probability to the Large Deviation Principle satisfied by the stochastic process capturing the empirical proportions of arm pulls and the sample means. Leveraging the connection, they not only enhance the guarantee of the Successive Rejects (SR) but also devise and analyze a novel algorithm with more adaptive rejections, Continuous Rejects (CR). The CR algorithm demonstrates superior performance both theoretically and numerically. Note however, that as SR, the CR algorithm is identical to the uniform sampling algorithm in bandit problems with two arms.

6.3. A/B Testing

For two-armed bandits, in (Kaufmann et al., 2014; 2016), the authors try to provide a characterization of the minimal instance-specific error probability for fixed-budget best-arm identification. For Bernoulli rewards, they establish a lower bound of this probability satisfied by any consistent algorithm:

$$\forall \boldsymbol{\mu} \in \Lambda, \quad \varliminf_{T \to \infty} \frac{T}{\log(1/p_{\boldsymbol{\mu},T})} \geq \min_{x \in (0,1)} \frac{1}{g(x, \boldsymbol{\mu})}. \quad (12)$$

(Kaufmann et al., 2014; 2016) also note, as in (Glynn & Juneja, 2004), that the best static algorithm (that requires the knowledge of $\boldsymbol{\mu}$) matches this lower bound. They do not find an adaptive algorithm that universally matches the lower bound across all bandit instances. Our results state that this is indeed impossible.

Now for Gaussian rewards, (Kaufmann et al., 2014; 2016) derive an error probability lower bound satisfied by consistent algorithms. They also show that when the learner is aware of the variances of the rewards, one may find an algorithm whose performance matches this lower bound on all instances. This algorithm is a static algorithm that pulls the first arm $\sigma_1/(\sigma_1 + \sigma_2)T + o(T)$ times and the second arm $\sigma_2/(\sigma_1 + \sigma_2)T + o(T)$ times, where σ_1^2 and σ_2^2 are the variances of the rewards of arm 1 and 2, respectively.

In (Degenne, 2023), the author shows that, for Bernoulli bandits, a universally optimal algorithm matching the lower bound (12) does not exist. Specifically, for any algorithm, there exists an instance and a static algorithm such that the considered algorithm performs strictly worse than the best static algorithm on the instance. Essentially, characterizing the instance-specific minimal error rate within a class of algorithms that includes all static algorithms is impossible. In this paper, we show that adaptive algorithms cannot even compete with a single static algorithm, namely the uniform sampling algorithm, on all instances.

7. Conclusion

In this paper, we investigated the problem of finding universally optimal algorithms for Fixed-Budget Best-Arm Identification (FB-BAI) in stochastic multi-armed bandits with Bernoulli rewards. We found that, surprisingly, for two-armed bandits (the A/B testing problem), no algorithm strictly outperforms the uniform sampling algorithm. We actually proved that within a natural and wide class of consistent and stable algorithms, uniform sampling is universally optimal. Extending these results to the case of more than two arms is challenging. So far we have not found any adaptive algorithm outperforming uniform sampling for all instances. For example, we were able to exactly characterize the asymptotic error probability of the celebrated Successive Rejects (SR) algorithm. As it turns out, SR is outperformed by uniform sampling in some instances.

Our study advances the understanding of the FB-BAI problem. However, we obtained a complete picture only for A/B testing with Bernoulli rewards. A similar picture is for now out of reach for general reward distributions. Indeed, the minimal error probability is not known even in the case of Gaussian reward distributions with unknown variances or in the case where these distributions are within one parameter exponential family. In the latter case, we conjecture that uniform sampling remains universally optimal. For problems involving more than two arms, the FB-BAI problem becomes even more challenging. We are currently working on extending the notion of stable algorithms, and on comparing their performance to that of the uniform sampling algorithm.

Impact Statement

This paper focuses on discussing the existence of universally optimal algorithms for best-arm identification. It presents work whose goal is to advance the field of Machine Learning. There are many potential societal consequences of our work, none which we feel must be specifically highlighted here.

Acknowledgments

We would like to express our gratitude to Chao Qin for meticulously examining the proofs in our early draft and identifying their incompleteness. His insightful comments led us to refine our definition of stable algorithms. Kaito Ariu's research is supported by JSPS KAKENHI Grant No. 23K19986. Alexandre Proutiere is supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation, the Swedish Research Council (VR) and Digital Futures.

References

Ariu, K., Kato, M., Komiyama, J., McAlinn, K., and Qin, C. Policy choice and best arm identification: Asymptotic analysis of exploration sampling. *arXiv preprint arXiv:2109.08229*, 2021.

Audibert, J.-Y., Bubeck, S., and Munos, R. Best arm identification in multi-armed bandits. In *Proc. of COLT*, 2010.

Bubeck, S., Munos, R., and Stoltz, G. Pure exploration in finitely-armed and continuous-armed bandits. *Theoretical Computer Science*, 412(19):1832–1852, 2011.

Cappé, O., Garivier, A., Maillard, O.-A., Munos, R., Stoltz, G., et al. Kullback–leibler upper confidence bounds for optimal sequential allocation. *Annals of Statistics*, 2013.

Carpentier, A. and Locatelli, A. Tight (lower) bounds for the fixed budget best arm identification bandit problem. In *Proc. of COLT*, 2016.

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

Combes, R. and Proutiere, A. Unimodal bandits: Regret lower bounds and optimal algorithms. In *Proc. of ICML*, 2014.

Degenne, R. On the existence of a complexity in fixed budget bandit identification. In *Proc. of COLT*, 2023.

Even-Dar, E., Mannor, S., Mansour, Y., and Mahadevan, S. Action elimination and stopping conditions for the multi-armed bandit and reinforcement learning problems. *JMLR*, 7(6), 2006.

Gabillon, V., Ghavamzadeh, M., and Lazaric, A. Best arm identification: A unified approach to fixed budget and fixed confidence. *Proc. of NeurIPS*, 2012.

Garivier, A. and Kaufmann, E. Optimal best arm identification with fixed confidence. In *Proc. of COLT*, 2016.

Garivier, A., Ménard, P., and Stoltz, G. Explore first, exploit next: The true shape of regret in bandit problems. *Mathematics of Operations Research*, 44(2):377–399, 2019.

Glynn, P. and Juneja, S. A large deviations perspective on ordinal optimization. In *Proc. of the 2004 Winter Simulation Conference*, 2004.

Karnin, Z., Koren, T., and Somekh, O. Almost optimal exploration in multi-armed bandits. In *Proc. of ICML*, 2013.

Kaufmann, E., Cappé, O., and Garivier, A. On the complexity of A/B testing. In *Proc. of COLT*, 2014.

Kaufmann, E., Cappé, O., and Garivier, A. On the complexity of best-arm identification in multi-armed bandit models. *JMLR*, 2016.

Komiyama, J., Tsuchiya, T., and Honda, J. Minimax optimal algorithms for fixed-budget best arm identification. In *Proc. of NeurIPS*, 2022.

Lai, T. L. and Robbins, H. Asymptotically efficient adaptive allocation rules. *Advances in applied mathematics*, 1985.

Qin, C. Open problem: Optimal best arm identification with fixed-budget. In *Proc. of COLT*, 2022.

Russo, D. Simple bayesian algorithms for best-arm identification. *Operations Research*, 68(6):1625–1647, 2020.

Wang, P.-A., Tzeng, R.-C., and Proutiere, A. Best arm identification with fixed budget: A large deviation perspective. In *Proc. of NeurIPS*, 2023.

A. Proof of Proposition 3.6

This section aims to prove Proposition 3.6, restated below for convenience.

Proposition 3.6. For any $a \in (0,1), x \in (\frac{1}{2}, 1]$, there exists an instance $\boldsymbol{\mu} \in \Lambda$ such that (i) $\mu_1 > \mu_2$, $\mu_1 + \mu_2 \geq 1$, (ii) $\lambda(x, \boldsymbol{\mu}) = a$, and (iii) $x^*(\boldsymbol{\mu}) < (\frac{1}{2} + x)/2$.

Proof. Bernoulli distributions belong to the single-parameter exponential family. Thus, there is a strictly convex function $\phi : \mathbb{R} \mapsto \mathbb{R}$ specific to these distributions. We denote by $\bar{d}$ the corresponding Bregman divergence. More precisely, $\phi(\xi)] := \log(1 + e^{\xi})$ with $\phi'(\xi) = \frac{e^{\xi}}{1+e^{\xi}}$ and $\phi'^{-1}(\mu) = \log \frac{\mu}{1-\mu}$. For a given $\boldsymbol{\mu} \in \Lambda$, there is a parameter $\boldsymbol{\xi} \in \mathbb{R}^2$ with $\xi_1 = \phi'^{-1}(\mu_1)$ and $\xi_2 = \phi'^{-1}(\mu_2)$ (note that ϕ' is an invertible function). Let $\bar{\Lambda} = \{\boldsymbol{\xi} \in \mathbb{R}^2 : \xi_1 \neq \xi_2\} = \phi'^{-1}(\Lambda)$ as the set of all parameters. $d(\mu_1, \mu_2)$ can be written as:

$$d(\mu_1, \mu_2) = \bar{d}(\xi_2, \xi_1) = \phi(\xi_2) - \phi(\xi_1) - (\xi_2 - \xi_1)\phi'(\xi_1). \tag{13}$$

Following this formalism, we can present the conditions $(\bar{\text{i}})$, $(\bar{\text{ii}})$, and $(\bar{\text{iii}})$ that are equivalent to the conditions (i), (ii) and (iii) used in the proposition. The proof and visualization for the lemma below can be found in Appendix C and Figure 4 respectively.

Lemma A.1. *The statement of Proposition 3.6 is equivalent to the following: for any $\alpha \in \mathbb{R}$, $x \in (\frac{1}{2}, 1]$, there exists an instance $\boldsymbol{\xi} \in \bar{\Lambda}$ such that $(\bar{\text{i}})$ $\xi_1 > \xi_2, \xi_1 \geq -\xi_2$, $(\bar{\text{ii}})$ $(1-x)\xi_1 + x\xi_2 = \alpha$, and $(\bar{\text{iii}})$ $\bar{d}(\xi_1, (1-\widetilde{x})\xi_1 + \widetilde{x}\xi_2) > \bar{d}(\xi_2, (1-\widetilde{x})\xi_1 + \widetilde{x}\xi_2)$, where $\widetilde{x} = (\frac{1}{2} + x)/2$.*

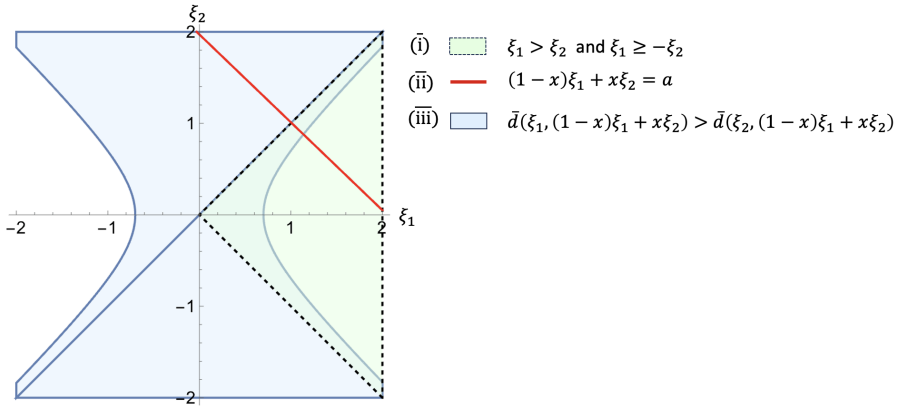

Figure 4. Visualization of Lemma A.1 with $a = 0.55$ and $x = 0.51$. The green region indicates $(\bar{\text{i}})$ $\xi_1 > \xi_2, \xi_1 \geq -\xi_2$. The red curve represents $(\bar{\text{ii}})$ $(1-x)\xi_1 + x\xi_2 = \alpha$. The blue region shows $(\bar{\text{iii}})$ $\bar{d}(\xi_1, (1-\widetilde{x})\xi_1 + \widetilde{x}\xi_2) > \bar{d}(\xi_2, (1-\widetilde{x})\xi_1 + \widetilde{x}\xi_2)$, where $\widetilde{x} = (\frac{1}{2} + x)/2$.

We consider two cases: (a) when $\alpha < 0$ and (b) when $\alpha \geq 0$.

Case (a) ($\alpha < 0$). We show that $\boldsymbol{\xi} = (-\alpha/(2x-1), \alpha/(2x-1))$ satisfies $(\bar{\text{i}})$, $(\bar{\text{ii}})$, $(\bar{\text{iii}})$. As $\alpha < 0, x > 1/2$, $(\bar{\text{i}})$ follows directly. $(\bar{\text{ii}})$ follows as $(1-x)\xi_1 + x\xi_2 = (1-2x)\xi_1 = \alpha$. As for $(\bar{\text{iii}})$, observe that

$$\widetilde{\xi} = (1-\widetilde{x})\xi_1 + \widetilde{x}\xi_2 = \frac{(2\widetilde{x}-1)\alpha}{2x-1} < 0.$$

Hence $\phi'(\widetilde{\xi}) = \frac{e^{\widetilde{\xi}}}{1+e^{\widetilde{\xi}}} < \frac{1}{2}$. As a consequence,

$$\begin{aligned}
\bar{d}(\xi_1, \widetilde{\xi}) - \bar{d}(\xi_2, \widetilde{\xi}) &= \phi(\xi_1) - \phi(\xi_2) - \phi'(\widetilde{\xi})(\xi_1 - \xi_2) \\
&= \log\left(\frac{1+e^{\xi_1}}{1+e^{-\xi_1}}\right) - \phi'(\widetilde{\xi})2\xi_1 \\
&= \xi_1\left(1 - 2\phi'(\widetilde{\xi})\right) > 0.
\end{aligned}$$

Case (b) ($\alpha \geq 0$). We claim that there is a half disk $\mathcal{D} := \{\boldsymbol{\xi} \in \mathbb{R}^2 : \|\boldsymbol{\xi} - (\alpha, \alpha)\|_\infty < \delta,\ \xi_1 > \xi_2\}$ for some $\delta > 0$ such that ($\bar{\text{iii}}$) holds whenever $\boldsymbol{\xi} \in \mathcal{D}$. We then show the intersection of the disk and the line $\mathcal{L} := \{\boldsymbol{\xi} \in \mathbb{R}^2 : (1-x)\xi_1 + x\xi_2 = \alpha\}$ is nonempty and satisfies ($\bar{\text{i}}$) and ($\bar{\text{ii}}$).

Assume that $\xi_1 > \xi_2$. On the one hand, applying Lemma C.3 with $\alpha = \xi_1$ and $\beta = (1-\widetilde{x})\xi_1 + \widetilde{x}\xi_2$ implies that there is $r_1 \in [\xi_2, \xi_1]$ such that

$$\bar{d}(\xi_1, (1-\widetilde{x})\xi_1 + \widetilde{x}\xi_2) = \frac{\phi''(r_1)\widetilde{x}^2(\xi_1 - \xi_2)^2}{2}. \tag{14}$$

On the other hand, applying the same lemma with $\alpha = \xi_2$ and $\beta = (1-\widetilde{x})\xi_1 + \widetilde{x}\xi_2$, we find $r_2 \in [\xi_2, \xi_1]$ such that

$$\bar{d}(\xi_2, (1-\widetilde{x})\xi_1 + \widetilde{x}\xi_2) = \frac{\phi''(r_2)(1-\widetilde{x})^2(\xi_1 - \xi_2)^2}{2}. \tag{15}$$

Since $\phi''(\xi) = \frac{e^\xi}{(1+e^\xi)^2}$ is a continuous function from $\mathbb{R}$ to $\mathbb{R}_{>0}$, Lemma C.4 implies that there exist $\delta > 0$ such that $\min_{r\in[\xi_2,\xi_1]} \phi''(r)\widetilde{x}^2 > \max_{r\in[\xi_2,\xi_1]} \phi''(r)(1-\widetilde{x})^2$ if $\|\boldsymbol{\xi} - (\alpha,\alpha)\|_\infty < \delta$. Thus by (14) and (15), we conclude that if $\boldsymbol{\xi} \in \mathcal{D} := \{\boldsymbol{\xi} \in \mathbb{R}^2 : \|\boldsymbol{\xi} - (\alpha,\alpha)\|_\infty < \delta, \xi_1 > \xi_2\}$, then ($\bar{\text{iii}}$) holds. Obviously, $\mathcal{D} \cap \mathcal{L} = \cup_{s\in(0,\delta/x)}\{(\alpha + xs, \alpha - (1-x)s)\} \neq \emptyset$. Consider $\boldsymbol{\xi} = (\alpha + xs, \alpha - (1-x)s)$ with some $s \in (0, \delta/x)$, we have $\xi_1 > \alpha > \xi_2$ and $\xi_1 + \xi_2 = 2\alpha + (2x-1)s \geq 0$, hence $\boldsymbol{\xi}$ satisfies ($\bar{\text{i}}$). ($\bar{\text{ii}}$) holds directly by definition of $\mathcal{L}$.

□

A.1. Proof for the Closed-Form Expressions of $g(x, \boldsymbol{\mu})$ and $\lambda(x, \boldsymbol{\mu})$

Here, we aim to present the proof of the following closed-form expressions for $g(x, \boldsymbol{\mu})$ and $\lambda(x, \boldsymbol{\mu})$:

Lemma A.2. *For any $\boldsymbol{\mu} \in \Lambda$, for any $x \in (0,1)$, the following equations hold.*

$$\lambda(x, \boldsymbol{\mu}) = \frac{(\frac{\mu_1}{1-\mu_1})^{1-x}(\frac{\mu_2}{1-\mu_2})^x}{1 + (\frac{\mu_1}{1-\mu_1})^{1-x}(\frac{\mu_2}{1-\mu_2})^x}, \tag{2}$$

$$g(x, \boldsymbol{\mu}) = -\log((1-\mu_1)^{1-x}(1-\mu_2)^x + \mu_1^{1-x}\mu_2^x). \tag{3}$$

Proof. Again, we denote $\xi_1 = \phi'^{-1}(\mu_1) = \log(\frac{\mu_1}{1-\mu_1})$, $\xi_2 = \phi'^{-1}(\mu_2) = \log(\frac{\mu_2}{1-\mu_2})$ as in the proof of Proposition 3.6. The minimization problem

$$\min_{\lambda\in[0,1]} \big((1-x)d(\lambda, \mu_1) + xd(\lambda, \mu_2)\big) \tag{16}$$

can be written as:

$$\min_{\lambda\in[0,1]} \big((1-x)\bar{d}(\xi_1, \phi'^{-1}(\lambda)) + x\bar{d}(\xi_2, \phi'^{-1}(\lambda))\big),$$

or equivalently,

$$\min_{\eta\in\mathbb{R}} \big((1-x)\bar{d}(\xi_1, \eta) + x\bar{d}(\xi_2, \eta)\big). \tag{17}$$

One can observe that

$$\frac{\partial}{\partial\eta}\big((1-x)\bar{d}(\xi_1,\eta) + x\bar{d}(\xi_2,\eta)\big) = \big(\eta - (1-x)\xi_1 - x\xi_2\big)\phi''(\eta)$$

has a unique root at the point $\eta = \eta(x, \boldsymbol{\xi}) = (1-x)\xi_1 + x\xi_2$. As ϕ' is an invertible mapping, we conclude that $\phi'(\eta(x, \boldsymbol{\xi}))$ is the unique minimizer to the minimization problem (16). Therefore,

$$\begin{aligned}
\lambda(x, \boldsymbol{\mu}) &= \phi'((1-x)\xi_1 + x\xi_2) \\
&= \phi'\left(\log\left(\left(\frac{\mu_1}{1-\mu_1}\right)^{1-x}\left(\frac{\mu_2}{1-\mu_2}\right)^x\right)\right) \\
&= \frac{(\frac{\mu_1}{1-\mu_1})^{1-x}(\frac{\mu_2}{1-\mu_2})^x}{1 + (\frac{\mu_1}{1-\mu_1})^{1-x}(\frac{\mu_2}{1-\mu_2})^x}.
\end{aligned}$$

Finally, (3) can be obtained as follows.

$$\begin{aligned}
g(x,\boldsymbol{\mu}) &= (1-x)\bar{d}(\xi_1,\eta(x,\boldsymbol{\xi})) + x\bar{d}(\xi_2,\eta(x,\boldsymbol{\xi})) \\
&= (1-x)\left(\phi(\xi_1) - \phi(\eta(x,\boldsymbol{\xi})) - (\xi_1-\eta(x,\boldsymbol{\xi}))\phi'(\eta(x,\boldsymbol{\xi}))\right) + x\left(\phi(\xi_2) - \phi(\eta(x,\boldsymbol{\xi})) - (\xi_2-\eta(x,\boldsymbol{\xi}))\phi'(\eta(x,\boldsymbol{\xi}))\right) \\
&= (1-x)\phi(\xi_1) + x\phi(\xi_2) - \phi(\eta(x,\boldsymbol{\xi})) - ((1-x)\xi_1 + x\xi_2 - \eta(x,\boldsymbol{\xi}))\phi'(\eta(x,\boldsymbol{\xi})) \\
&= (1-x)\phi(\xi_1) + x\phi(\xi_2) - \phi(\eta(x,\boldsymbol{\xi})) \\
&= (1-x)\phi\left(\log(\frac{\mu_1}{1-\mu_1})\right) + x\phi\left(\log(\frac{\mu_2}{1-\mu_2})\right) - \phi\left((1-x)\log(\frac{\mu_1}{1-\mu_1}) + x\log(\frac{\mu_2}{1-\mu_2})\right) \\
&= -\log((1-\mu_1)^{1-x}(1-\mu_2)^x + \mu_1^{1-x}\mu_2^x).
\end{aligned}$$

□

A.2. Proof for the Strong Concavity of $g(x,\boldsymbol{\mu})$

Lemma A.3. *For any $\boldsymbol{\mu}\in\Lambda$, for any $x\in(0,1)$, $\frac{\partial^2}{\partial^2 x}g(x,\boldsymbol{\mu}) < 0$.*

Proof. As shown in the proof of Lemma A.2,

$$g(x,\boldsymbol{\mu}) = (1-x)\phi(\xi_1) + x\phi(\xi_2) - \phi(\eta(x,\boldsymbol{\xi})), \tag{18}$$

where $\xi_1 = \phi'^{-1}(\mu_1)$, $\xi_2 = \phi'^{-1}(\mu_2)$, and $\eta(x,\boldsymbol{\xi}) = (1-x)\xi_1 + x\xi_2$. We differentiate (18) with respect to x:

$$\begin{aligned}
\frac{\partial}{\partial x}g(x,\boldsymbol{\mu}) &= \phi(\xi_2) - \phi(\xi_1) + (\xi_1-\xi_2)\phi'(\eta(x,\boldsymbol{\xi})), \\
\frac{\partial^2}{\partial^2 x}g(x,\boldsymbol{\mu}) &= -(\xi_1-\xi_2)^2\phi''(\eta(x,\boldsymbol{\xi})) < 0.
\end{aligned}$$

□

A.3. Proof of Proposition 3.8

Proposition 3.8. Denote $\bar{\boldsymbol{\mu}} = (1-\mu_2, 1-\mu_1)$. For any $x\in(0,1)$, for any $\boldsymbol{\mu}\in\Lambda$, $g(1-x,\bar{\boldsymbol{\mu}}) = g(x,\boldsymbol{\mu})$ and $\lambda(1-x,\bar{\boldsymbol{\mu}}) = 1-\lambda(x,\boldsymbol{\mu})$.

Proof. From (3), we obtain

$$\begin{aligned}
g(1-x,\bar{\boldsymbol{\mu}}) &= -\log((1-\mu_1)^{1-x}(1-\mu_2)^x + \mu_1^{1-x}\mu_2^x) \\
&= g(x,\boldsymbol{\mu}).
\end{aligned}$$

Lastly, by (2), we have

$$\begin{aligned}
\lambda(1-x,\bar{\boldsymbol{\mu}}) &= \frac{(1-\mu_1)^{1-x}(1-\mu_2)^x}{(1-\mu_1)^{1-x}(1-\mu_2)^x + \mu_1^{1-x}\mu_2^x} \\
&= 1 - \frac{\mu_1^{1-x}\mu_2^x}{(1-\mu_1)^{1-x}(1-\mu_2)^x + \mu_1^{1-x}\mu_2^x} \\
&= 1-\lambda(x,\boldsymbol{\mu}).
\end{aligned}$$

This concludes the proof.

□

B. Proof of Proposition 3.7

Proposition 3.7. Suppose $\mu_1 > \mu_2$ and $\mu_1 + \mu_2 \geq 1$. For any positive $\delta \leq \min\{x^*(\boldsymbol{\mu}), 1 - x^*(\boldsymbol{\mu})\}$,

$$g(x^*(\boldsymbol{\mu}) - \delta, \boldsymbol{\mu}) \geq g(x^*(\boldsymbol{\mu}) + \delta, \boldsymbol{\mu}).$$

Proof. For simplicity, let $x^* = x^*(\boldsymbol{\mu})$, and define

$$\hat{g}(x, \boldsymbol{\mu}) = \exp(-g(x, \boldsymbol{\mu})) = (1-\mu_1)^{1-x}(1-\mu_2)^x + \mu_1^{1-x}\mu_2^x. \tag{19}$$

Consider the function $f(\delta) := \hat{g}(x^* + \delta, \boldsymbol{\mu}) - \hat{g}(x^* - \delta, \boldsymbol{\mu})$. In order to demonstrate that $f(\delta) \geq 0$ using the mean value theorem, we show that the first-order derivative, $f'(\delta)$, is non-negative for all $\delta \in (0, \min\{x^*(\boldsymbol{\mu}), 1 - x^*(\boldsymbol{\mu})\})$.

By definition of x^*, we have $\frac{\partial \hat{g}(x^*, \boldsymbol{\mu})}{\partial x} = 0$, i.e.,

$$(1-\mu_1)^{1-x^*}(1-\mu_2)^{x^*} \log\left(\frac{1-\mu_2}{1-\mu_1}\right) + \mu_1^{1-x^*}\mu_2^{x^*} \log\left(\frac{\mu_2}{\mu_1}\right) = 0.$$

This implies that

$$\frac{(1-\mu_1)^{1-x^*}(1-\mu_2)^{x^*}}{\log\left(\frac{\mu_1}{\mu_2}\right)} = \frac{\mu_1^{1-x^*}\mu_2^{x^*}}{\log\left(\frac{1-\mu_2}{1-\mu_1}\right)}. \tag{20}$$

Let M be the value of (20). $M \geq 0$ as $\mu_1 > \mu_2$. Recalling the definition of $f(\delta)$ and (19), we get

$$f(\delta) = M \log\left(\frac{\mu_1}{\mu_2}\right)\left[\left(\frac{1-\mu_2}{1-\mu_1}\right)^{\delta} - \left(\frac{1-\mu_2}{1-\mu_1}\right)^{-\delta}\right] - M \log\left(\frac{1-\mu_2}{1-\mu_1}\right)\left[\left(\frac{\mu_1}{\mu_2}\right)^{\delta} - \left(\frac{\mu_1}{\mu_2}\right)^{-\delta}\right]$$

and

$$f'(\delta) = M \log\left(\frac{\mu_1}{\mu_2}\right) \log\left(\frac{1-\mu_2}{1-\mu_1}\right)\left[\left(\frac{1-\mu_2}{1-\mu_1}\right)^{\delta} + \left(\frac{1-\mu_2}{1-\mu_1}\right)^{-\delta} - \left(\frac{\mu_1}{\mu_2}\right)^{\delta} - \left(\frac{\mu_1}{\mu_2}\right)^{-\delta}\right].$$

Observe that the first three factors on the r.h.s. of the above expression are all positive, since $\mu_1 > \mu_2$. For the last factor, we first have $\frac{1-\mu_2}{1-\mu_1} \geq \frac{\mu_1}{\mu_2}$, according to Lemma C.5 in Appendix C. Additionally, note that the mapping $z \mapsto z^{\delta} + z^{-\delta}$ is increasing when $z > 0$. From these observations, we can conclude that $f'(\delta) \geq 0$, which completes the proof. □

C. Technical Lemmas

Lemma 3.4. Let $a \in (0,1)$. Assume that the statement (B) of Definition 3.2 does not hold. Then for any $\{\boldsymbol{\pi}^{(n)}\}_{n=1}^{\infty} \subset \{\boldsymbol{\pi} \in \Lambda : \pi_1 < \pi_2\}$ such that $\boldsymbol{\pi}^{(n)} \xrightarrow{n\to\infty} (a,a)$, there exists a value $x \in [0,1]$ and increasing subsequences of integers $\{n_m\}_{m=1}^{\infty}, \{T_{m,\ell}\}_{\ell=1}^{\infty} \subset \mathbb{N}$ such that

$$\lim_{m\to\infty} \lim_{\ell\to\infty} \mathbb{E}_{\boldsymbol{\pi}^{(n_m)}}[\omega_2(T_{m,\ell})] = x \neq \frac{1}{2}.$$

Proof. First, Lemma C.1 shows that the equations

$$\lim_{n\to\infty} \lim_{T\to\infty} \mathbb{E}_{\boldsymbol{\pi}^{(n)}}[\omega_2(T)] = \lim_{n\to\infty} \overline{\lim_{T\to\infty}} \mathbb{E}_{\boldsymbol{\pi}^{(n)}}[\omega_2(T)] = \frac{1}{2}$$

in the statement (B) of Definition 3.2 is equivalent to the statement: $\forall \varepsilon > 0$, $\exists N \in \mathbb{N}$ such that $\forall n \geq N$, $\exists T_n \in \mathbb{N}$, $\forall T \geq T_n$,

$$\left| \mathbb{E}_{\boldsymbol{\pi}^{(n)}}[\omega_2(T)] - \frac{1}{2} \right| < \varepsilon.$$

Its negation is: there exists $\varepsilon \in (0, \frac{1}{2}]$ such that $\forall N \in \mathbb{N}$, $\exists \bar{n} \geq N$ such that $\forall T \in \mathbb{N}$, $\exists \bar{T}_{\bar{n}} \geq T$ such that

$$\left| \mathbb{E}_{\boldsymbol{\pi}^{(\bar{n})}}[\omega_2(\bar{T}_{\bar{n}})] - \frac{1}{2} \right| \geq \varepsilon. \tag{21}$$

In the above statement, let select N arbitrarily, and fix a corresponding $\bar{n}$. If we take $T = 1$, we can find $\bar{T}_{\bar{n}} = \bar{T}_{\bar{n},1} \geq 1$ satisfying (21). Next, we take $T = \bar{T}_{\bar{n},1} + 1$, we can find $\bar{T}_{\bar{n}} = \bar{T}_{\bar{n},2} > \bar{T}_{\bar{n},1}$ satisfying (21). By repeating this operation, we can construct an increasing sequence of integers $\{\bar{T}_{\bar{n},\ell}\}_{\ell=1}^{\infty}$ such that:

$$\forall \ell \in \mathbb{N}, \quad \left| \mathbb{E}_{\boldsymbol{\pi}^{(\bar{n})}}[\omega_2(\bar{T}_{\bar{n},\ell})] - \frac{1}{2} \right| \geq \varepsilon.$$

We have now proved that there exists $\varepsilon \in (0, \frac{1}{2}]$ such that $\forall N \in \mathbb{N}$, $\exists \bar{n} \geq N$ such that

$$\exists \{\bar{T}_{\bar{n},\ell}\}_{\ell=1}^{\infty} : \ \forall \ell \in \mathbb{N}, \bar{T}_{\bar{n},\ell} < \bar{T}_{\bar{n},\ell+1} \text{ and } \left| \mathbb{E}_{\boldsymbol{\pi}^{(\bar{n})}}[\omega_2(\bar{T}_{\bar{n},\ell})] - \frac{1}{2} \right| \geq \varepsilon. \tag{22}$$

Again, in the above statement, if we take $N = 1$, we can find $\bar{n}_1 \geq 1$ satisfying (22). Next, if we take $N = \bar{n}_1 + 1$, we can find $\bar{n} = \bar{n}_2 > \bar{n}_1$ satisfying (22). By repeating this operation, we can construct an increasing sequence of integers $\{\bar{n}_m\}_{m=1}^{\infty}$ satisfying (22). In summary, we have found an increasing sequence of integers $\{\bar{n}_m\}_{m=1}^{\infty}$ and for all m, an other increasing sequence of integers $\{\bar{T}_{\bar{n}_m,\ell}\}_{\ell=1}^{\infty}$ such that:

$$\forall (m,\ell) \in \mathbb{N}^2, \quad \left| \mathbb{E}_{\boldsymbol{\pi}^{(\bar{n}_m)}}[\omega_2(\bar{T}_{\bar{n}_m,\ell})] - \frac{1}{2} \right| \geq \varepsilon. \tag{23}$$

From the Bolzano–Weierstrass theorem, a bounded sequence always contains a convergent subsequence. Thus, for each $m \in \mathbb{N}$, one can always find $\{T_{\bar{n}_m,\ell}\}_{\ell=1}^{\infty} \subset \{\bar{T}_{\bar{n}_m,\ell}\}_{\ell=1}^{\infty}$ such that $\{\mathbb{E}_{\boldsymbol{\pi}^{(\bar{n}_m)}}[\omega_2(T_{\bar{n}_m,\ell})]\}_{\ell=1}^{\infty}$ converges. We denote by $x_{\bar{n}_m}$ its limit, i.e., $\lim_{\ell\to\infty} \mathbb{E}_{\boldsymbol{\pi}^{(\bar{n}_m)}}[\omega_2(T_{\bar{n}_m,\ell})] = x_{\bar{n}_m}$. Note that from (23), with some $\varepsilon \in (0, \frac{1}{2}]$,

$$\forall m \in \mathbb{N}, \quad \left| x_{\bar{n}_m} - \frac{1}{2} \right| \geq \varepsilon. \tag{24}$$

Futher observe that of course, $x_{\bar{n}_m} \in [0,1]$ for all m. Using the Bolzano–Weierstrass theorem again, there exists $\{n_m\}_{m=1}^{\infty} \subset \{\bar{n}_m\}_{m=1}^{\infty}$ such that x_{n_m} converges to $x \in [0,1]$, i.e, $\lim_{m\to\infty} x_{n_m} = x$. From (24), we remark that $|x - \frac{1}{2}| \geq \varepsilon$. The constructed x, $\{n_m\}_{m=1}^{\infty}$, and $\{T_{m,\ell}\}_{\ell=1}^{\infty}$ satisfy the desired claim, which concludes the proof.

□

Lemma C.1. *Let $\{\psi(n,T)\}_{n,T=1}^{\infty}$ be a double sequence of real numbers. The following two statements are equivalent:*

$$(I) \qquad \lim_{n\to\infty} \underline{\lim}_{T\to\infty} \psi(n,T) = \lim_{n\to\infty} \overline{\lim}_{T\to\infty} \psi(n,T) = \frac{1}{2}$$

$$(II) \qquad \forall \varepsilon > 0, \exists N \in \mathbb{N} : \forall n \geq N, \left(\exists T_n \in \mathbb{N}, \forall T \geq T_n, \ \left|\psi(n,T) - \frac{1}{2}\right| < \varepsilon\right).$$

Proof. We first prove that $(II) \Rightarrow (I)$. We rewrite (II) as follows: $\forall \varepsilon > 0$, $\exists N \in \mathbb{N}$ such that $\forall n \geq N$, $\exists T_n \in \mathbb{N}$, $\forall T \geq T_n$,

$$\frac{1}{2} - \varepsilon < \psi(n,T) < \frac{1}{2} + \varepsilon. \tag{25}$$

By taking $\overline{\lim}_{T\to\infty}$ and $\underline{\lim}_{T\to\infty}$ of (25), we obtain the following statement: $\forall \varepsilon > 0$, $\exists N \in \mathbb{N}$ such that $\forall n \geq N$,

$$\frac{1}{2} - \varepsilon \leq \underline{\lim}_{T\to\infty} \psi(n,T) \leq \overline{\lim}_{T\to\infty} \psi(n,T) \leq \frac{1}{2} + \varepsilon.$$

Therefore, we get (I).

Next we prove that $(I) \Rightarrow (II)$. Observe that $\lim_{n\to\infty} \overline{\lim}_{T\to\infty} \psi(n,T) = 1/2$ implies for any $\eta_1 > 0$, there exists $\overline{N} \in \mathbb{N}$, such that $\forall n \geq \overline{N}$,

$$\overline{\lim}_{T\to\infty} \psi(n,T) \leq \frac{1}{2} + \eta_1. \tag{26}$$

As a consequence of (26), for any $\eta_1 > 0$, $\exists \overline{N} \in \mathbb{N}$ such that $\forall n \geq \overline{N}$, for any $\eta_2 > 0$, $\exists \overline{T}_n \in \mathbb{N}$ such that $\forall T \geq \overline{T}_n$,

$$\psi(n,T) \leq \frac{1}{2} + \eta_1 + \eta_2. \tag{27}$$

Similarly, $\lim_{n\to\infty} \underline{\lim}_{T\to\infty} \psi(n,T) = 1/2$ implies: for any $\eta_1 > 0$, $\exists \underline{N} \in \mathbb{N}$ such that $\forall n \geq \underline{N}$, for any $\eta_2 > 0$, $\exists \underline{T}_n \in \mathbb{N}$ such that $\forall T \geq \underline{T}_n$,

$$\psi(n,T) \geq \frac{1}{2} - \eta_1 - \eta_2. \tag{28}$$

Combing (27) and (28), for any $\eta_1 > 0$, $\exists N (= \max\{\overline{N}, \underline{N}\})$ such that $\forall n \geq N$, for any $\eta_2 > 0$, $\exists T_n (= \max\{\overline{T}_n, \underline{T}_n\})$ such that $\forall T \geq T_n$, $\frac{1}{2} - \eta_1 - \eta_2 \leq \psi(n,T) \leq \frac{1}{2} + \eta_1 + \eta_2$. By taking $\eta_1 = \varepsilon/2$ and $\eta_2 = \varepsilon/2$, we obtain the statement: for any $\varepsilon > 0$, $\exists N \in \mathbb{N}$ such that $\forall n \geq N$, $\exists T_n \in \mathbb{N}$ such that $\forall T \geq T_n$, $\frac{1}{2} - \varepsilon \leq \psi(n,T) \leq \frac{1}{2} + \varepsilon$, which completes the proof.

□

Lemma A.1. The statement of Proposition 3.6 is equivalent to the following: for any $\alpha \in \mathbb{R}$, $x \in (\frac{1}{2}, 1]$, there exists an instance $\boldsymbol{\xi} \in \bar{\Lambda}$ such that $(\bar{\text{i}})\, \xi_1 > \xi_2, \xi_1 \geq -\xi_2$, $(\bar{\text{ii}})\, (1-x)\xi_1 + x\xi_2 = \alpha$, and $(\bar{\text{iii}})\, \bar{d}(\xi_1, (1-\widetilde{x})\xi_1 + \widetilde{x}\xi_2) > \bar{d}(\xi_2, (1-\widetilde{x})\xi_1 + \widetilde{x}\xi_2)$, where $\widetilde{x} = (\frac{1}{2} + x)/2$.

Proof. We show the equivalence of (i) to $(\bar{\text{i}})$, (ii) to $(\bar{\text{ii}})$, and (iii) to $(\bar{\text{iii}})$ in the following.

Equivalence of (i) to $(\bar{\text{i}})$. Note that $\mu_1 > \mu_2$ can be rewritten as $\phi'(\xi_1) > \phi'(\xi_2)$. As ϕ' is a strictly increasing function, it holds if and only if $\xi_1 > \xi_2$. As for $\mu_1 + \mu_2 \geq 1$, its equivalent statement is

$$1 \leq \phi'(\xi_1) + \phi'(\xi_2) = \frac{e^{\xi_1} + e^{\xi_2} + 2e^{\xi_1+\xi_2}}{(1+e^{\xi_1})(1+e^{\xi_2})}.$$

By rearranging the above inequality, we obtain $\xi_1 > -\xi_2$.

Equivalence of (ii) to ($\overline{\text{ii}}$). Introduce the following notation.

$$\begin{aligned}\bar{g}(x,\boldsymbol{\xi}) &\coloneqq \inf_{\bar{\lambda}\in\mathbb{R}} (1-x)\bar{d}(\xi_1,\bar{\lambda}) + x\bar{d}(\xi_2,\bar{\lambda}) = g(x,\phi'(\xi_1),\phi'(\xi_2)),\\ \bar{\lambda}(x,\boldsymbol{\xi}) &\coloneqq \operatorname*{argmin}_{\bar{\lambda}\in\mathbb{R}} (1-x)\bar{d}(\xi_1,\bar{\lambda}) + x\bar{d}(\xi_2,\bar{\lambda}) = \phi'^{-1}(\lambda(x,\phi'(\xi_1),\phi'(\xi_2))),\\ \text{and } \bar{x}^*(\boldsymbol{\xi}) &\coloneqq \operatorname*{argmax}_{x\in(0,1)} \bar{g}(x,\boldsymbol{\xi}) = x^*(\phi'(\xi_1),\phi'(\xi_2)).\end{aligned}$$

We use the following lemma from (Degenne, 2023).

Lemma C.2 (Lemma 19 in (Degenne, 2023))**.** *For any $\boldsymbol{\xi}\in\bar{\Lambda}$, $x\in[0,1]$, $\bar{\lambda}(x,\boldsymbol{\xi}) = (1-x)\xi_1 + x\xi_2$ and $\bar{x}^*(\boldsymbol{\xi}) = \frac{\xi_1-\eta(\boldsymbol{\xi})}{\xi_1-\xi_2}$, where $\eta(\boldsymbol{\xi}) = \phi'^{-1}(\frac{\phi(\xi_1)-\phi(\xi_2)}{\xi_1-\xi_2})$.*

The equivalence of (ii) to ($\overline{\text{ii}}$) directly follows from Lemma C.2.

Equivalence of (iii) to ($\overline{\text{iii}}$). From Lemma C.2, $\bar{x}^*(\boldsymbol{\xi}) = x^*(\phi'(\xi_1),\phi'(\xi_2)) < (\frac{1}{2}+x)/2 = \widetilde{x}$ for a given $x\in(\frac{1}{2},1]$ is equivalent to

$$\phi'^{-1}\left(\frac{\phi(\xi_1)-\phi(\xi_2)}{\xi_1-\xi_2}\right) = \eta(\boldsymbol{\xi}) > (1-\widetilde{x})\xi_1 + \widetilde{x}\xi_2.$$

By denoting $\widetilde{\xi} = (1-\widetilde{x})\xi_1 + \widetilde{x}\xi_2$, we then arrange it as:

$$\phi(\xi_1) - \phi(\widetilde{\xi}) - \phi'(\widetilde{\xi})(\xi_1-\widetilde{\xi}) > \phi(\xi_2) - \phi(\widetilde{\xi}) - \phi'(\widetilde{\xi})(\xi_2-\widetilde{\xi}),$$

that is, $\bar{d}(\xi_1,\widetilde{\xi}) > \bar{d}(\xi_2,\widetilde{\xi})$. This concludes the proof of Lemma A.1.

□

Lemma C.3. *Given $\alpha,\beta\in\mathbb{R}$, there exists $\min\{\alpha,\beta\}\le r\le\max\{\alpha,\beta\}$ such that*

$$\bar{d}(\alpha,\beta) = \frac{(\alpha-\beta)^2\phi''(r)}{2}.$$

Proof. The first and second derivatives of $\bar{d}$ are:

$$\frac{\partial}{\partial\alpha}\bar{d}(\alpha,\beta) = \phi'(\alpha) - \phi'(\beta) \text{ and } \frac{\partial^2}{\partial^2\alpha}\bar{d}(\alpha,\beta) = \phi''(\alpha).$$

Using Taylor's equality, we have $r\in[\min\{\alpha,\beta\},\max\{\alpha,\beta\}]$ such that

$$\bar{d}(\alpha,\beta) = \bar{d}(\beta,\beta) + (\alpha-\beta)\frac{\partial}{\partial\alpha}\bar{d}(\beta,\beta) + \frac{(\alpha-\beta)^2}{2}\frac{\partial^2}{\partial^2\alpha}\bar{d}(r,\beta) = \frac{(\alpha-\beta)^2\phi''(r)}{2}.$$

□

Lemma C.4. *Suppose $f:\mathbb{R}\mapsto\mathbb{R}_{>0}$ is a continuous function. For any $\alpha\in\mathbb{R}$, $x\in(\frac{1}{2},1)$, there exist $\delta>0$ s.t. if $\|\boldsymbol{\xi}-(\alpha,\alpha)\|_\infty<\delta$ and $\xi_1>\xi_2$, then*

$$\min_{r\in[\xi_2,\xi_1]} f(r)x^2 > \max_{r\in[\xi_2,\xi_1]} f(r)(1-x)^2.$$

Proof. As $x\in(\frac{1}{2},1)$, we derive $\frac{1}{4x^2}<1<\frac{1}{4(1-x)^2}$. By the continuity of f and $f(\alpha)>0$, there exists $\delta>0$ such that if $|r-\alpha|<\delta$,

$$\frac{f(\alpha)}{4x^2} < f(r) < \frac{f(\alpha)}{4(1-x)^2}.$$

Consequently, when $\|\boldsymbol{\xi} - (\alpha, \alpha)\|_\infty < \delta$ and $\xi_1 > \xi_2$,

$$\min_{r\in[\xi_2,\xi_1]} f(r) \geq \min_{|r-\alpha|<\delta} f(r) > \frac{f(\alpha)}{4x^2} \text{ and } \max_{r\in[\xi_2,\xi_1]} f(r) \leq \max_{|r-\alpha|<\delta} f(r) < \frac{f(\alpha)}{4(1-x)^2},$$

which yields that $\min_{r\in[\xi_2,\xi_1]} f(r)x^2 > \frac{f(\alpha)}{4} > \max_{r\in[\xi_2,\xi_1]} f(r)(1-x)^2$. □

Lemma C.5. *If $\mu_1 > \mu_2$ and $\mu_1 + \mu_2 \geq 1$, then $\frac{1-\mu_2}{1-\mu_1} \geq \frac{\mu_1}{\mu_2}$.*

Proof. Observe that the given two assumptions can be rewritten as $\mu_1 - \frac{1}{2} > \mu_2 - \frac{1}{2}$ and $\frac{1}{2} - \mu_2 \leq \mu_1 - \frac{1}{2}$. Thus, we conclude that

$$(1-\mu_2)\mu_2 = -\left(\mu_2 - \frac{1}{2}\right)^2 + \frac{1}{4} \geq -\left(\mu_1 - \frac{1}{2}\right)^2 + \frac{1}{4} = (1-\mu_1)\mu_1,$$

which is equivalent to the desired conclusion. □

D. Proof of Theorem 5.1

Throughout this section, we assume, without loss of generality, that $\mu_1 > \mu_2 \geq \ldots \geq \mu_K$. Recall that

$$\Gamma_j(\boldsymbol{\mu}) = \min_{J \in \mathcal{J}_j(\boldsymbol{\mu})} \inf_{\boldsymbol{\lambda} \in \Lambda : \lambda_1 \leq \min_{k \in J} \lambda_k} \sum_{k \in J} d(\lambda_k, \mu_k), \tag{29}$$

and $\mathcal{J}_j(\boldsymbol{\mu}) = \{J \subseteq [K] : |J| = j, 1 \in J\}$.

Proof of Theorem 5.1. As mentioned in Section 5.2, (Wang et al., 2023) establishes that

$$\overline{\lim_{T \to \infty}} \frac{T}{\log(1/p_{\boldsymbol{\mu},T})} \leq \max_{j=2,\ldots,K} \frac{j\overline{\log} K}{\Gamma_j(\boldsymbol{\mu})}, \forall \boldsymbol{\mu} \in \Lambda. \tag{9}$$

Hence we just need to prove the following Lemma D.1.

Lemma D.1. *Under Algorithm 1, one has*

$$\underline{\lim_{T \to \infty}} \frac{T}{\log(1/p_{\boldsymbol{\mu},T})} \geq \max_{j=2,\ldots,K} \frac{j\overline{\log} K}{\Gamma_j(\boldsymbol{\mu})}, \forall \boldsymbol{\mu} \in \Lambda. \tag{30}$$

Combining (9) and (30), we then complete the proof. □

Proof of Lemma D.1. For all $j = 2, \ldots, K$, let $\mathcal{E}_j = \{\ell_j = 1\}$ be the event that SR discards the best arm, 1, when there are j candidates remaining. As under the event $\cup_{j=2}^K \mathcal{E}_j$, 1 is removed before the end, we have

$$p_{\boldsymbol{\mu},T} = \sum_{j=2}^K \mathbb{P}_{\boldsymbol{\mu}}[\mathcal{E}_j] \geq \mathbb{P}_{\boldsymbol{\mu}}[\mathcal{E}_j], \quad \forall j = 2, \ldots, K.$$

Consequently, the above inequalities imply that

$$\underline{\lim_{T \to \infty}} \frac{T}{\log(1/p_{\boldsymbol{\mu},T})} \geq \max_{j=2,\ldots,K} \underline{\lim_{T \to \infty}} \frac{T}{\log(1/\mathbb{P}_{\boldsymbol{\mu}}[\mathcal{E}_j])}.$$

The proof of the lemma is reduced to showing that for any $j = 2, \ldots, K$,

$$\underline{\lim_{T \to \infty}} \frac{T}{\log(1/\mathbb{P}_{\boldsymbol{\mu}}[\mathcal{E}_j])} \geq \frac{j\overline{\log} K}{\Gamma_j(\boldsymbol{\mu})}. \tag{31}$$

Proof of (31). Let $j \in \{2, \ldots, K\}$. For convenience, we set μ_{K+1} equal to 0. We then consider some $\boldsymbol{\lambda} \in \Lambda_j$, where

$$\Lambda_j = \left\{ \boldsymbol{\lambda} \in (0,1)^K \times \{0\} : \mu_{j+1} < \lambda_1 \leq \min_{k \in [j]} \lambda_k, \text{ and } \lambda_k = \mu_k, \forall k \geq j+1 \right\}.$$

Applying a standard change-of-measure argument, as in the proof of Theorem 2.2, yields that

$$\sum_{k=1}^K \mathbb{E}_{\boldsymbol{\mu}}[N_k(\lfloor \theta T \rfloor)] d(\lambda_k, \mu_k) \geq d(\mathbb{P}_{\boldsymbol{\lambda}}[\mathcal{E}_j], \mathbb{P}_{\boldsymbol{\mu}}[\mathcal{E}_j]) \geq \mathbb{P}_{\boldsymbol{\lambda}}[\mathcal{E}_j] \log\left(\frac{1}{\mathbb{P}_{\boldsymbol{\mu}}[\mathcal{E}_j]}\right) - \log 2, \tag{32}$$

where the last inequality stems from the fact that $d(p,q) \geq p \log(1/q) - \log 2$ for all $p, q \in [0,1]$. Thanks to law of large number, as $T \to \infty$, $\mathbb{E}_{\boldsymbol{\mu}}[\omega_k(\lfloor \theta T \rfloor)] \to \frac{1}{j\overline{\log} K}$ for each $k = 1, \ldots, j$, and $\mathbb{P}_{\boldsymbol{\lambda}}[\mathcal{E}_j] \to 1$. Taking T to infinity in (32) yields that

$$\underline{\lim_{T \to \infty}} \frac{T}{\log(1/\mathbb{P}_{\boldsymbol{\mu}}[\mathcal{E}_j])} \geq \frac{j\overline{\log} K}{\sum_{k=1}^j d(\lambda_k, \mu_k)}.$$

Since the above inequality holds for all $\boldsymbol{\lambda} \in \Lambda_j$, we have

$$\begin{aligned}\varliminf_{T\to\infty} \frac{T}{\log(1/\mathbb{P}_{\boldsymbol{\mu}}[\mathcal{E}_j])} &\geq \sup_{\boldsymbol{\lambda}\in\Lambda_j} \frac{j\overline{\log}K}{\sum_{k=1}^{j} d(\lambda_k,\mu_k)} \\ &= \frac{j\overline{\log}K}{\inf_{\boldsymbol{\lambda}\in\Lambda_j}\sum_{k=1}^{j} d(\lambda_k,\mu_k)} = \frac{j\overline{\log}K}{\Gamma_j(\boldsymbol{\mu})},\end{aligned}$$

where the last equation directly follows from (39) in Proposition D.4 in Appendix D.1.

□

D.1. Derivation of Computationally Tractable Form of $\Gamma_j(\boldsymbol{\mu})$

Throughout this subsection, we define the function $\Psi_j : \mathbb{R}^j \mapsto \mathbb{R}$ as:

$$\Psi_j(x_1,\ldots,x_j) \coloneqq \inf_{\boldsymbol{\eta}\in\mathbb{R}^j : \eta_1 \leq \min_{k\in[j]}\eta_k} \sum_{k=1}^{j} \bar{d}(x_k,\eta_k). \tag{33}$$

In the following, we establish important properties of Ψ_j for a fixed $j \in \{2,\ldots,K\}$. These properties will help us to understand $\Gamma_j(\boldsymbol{\mu})$.

Proposition D.2. *If $x_1 > x_2 \geq \ldots \geq x_j$, then*

$$\Psi_j(x_1,\ldots,x_j) = \frac{\sum_{k\in\mathcal{I}_j(\boldsymbol{x})} \bar{d}(x_k, \frac{\sum_{k'\in\mathcal{I}_j(\boldsymbol{x})} x_{k'}}{|\mathcal{I}_j(\boldsymbol{x})|})}{|\mathcal{I}_j(\boldsymbol{x})|}, \tag{34}$$

where $\mathcal{I}_j(\boldsymbol{x}) = \left\{ i \in \{2,\ldots,j\} : x_i(j-i+1) < x_1 + \sum_{i<k\leq j} x_k \right\} \cup \{1\}$. Moreover, the minimizer $\boldsymbol{\eta}^*$ of (33) satisfies: $x_j < \eta_1^* \leq \min_{k\in[j]}\eta_k^*$, i.e.,

$$\Psi_j(x_1,\ldots,x_j) = \inf_{\boldsymbol{\eta}\in\mathbb{R}^j : x_j < \eta_1 \leq \min_{k\in[j]}\eta_k} \sum_{k=1}^{j} \bar{d}(x_k,\eta_k). \tag{35}$$

Proof. Observe that in the optimization problem (33), there exists $\boldsymbol{\eta} \in \mathbb{R}^j$ such that all the constraints are strict (satisfying Slater's condition). Thus, the solution of (33) can be identified by verifying the KKT conditions. The corresponding Lagrangian function is

$$\mathcal{L}(\boldsymbol{\eta},\alpha_2,\ldots,\alpha_j) = \sum_{k=1}^{j} \bar{d}(x_k,\eta_k) + \sum_{k=2}^{j} \alpha_k(\eta_1-\eta_k),\ \forall(\boldsymbol{\eta},\alpha_2,\ldots,\alpha_j) \in \mathbb{R}^j \times \mathbb{R}_{\geq 0}^{j-1}.$$

Let $(\boldsymbol{\eta}^*,\alpha_2^*,\ldots,\alpha_j^*)$ be a saddle point of $\mathcal{L}$. It satisfies KKT conditions:

$$\begin{aligned}\eta_1^* \leq \eta_k^*, \forall k = 2,\ldots,j, &\qquad \text{(Primal Feasibility)}\\ \alpha_k^* \geq 0,\ \forall k = 2,\ldots,j, &\qquad \text{(Dual Feasibility)}\\ \frac{\partial}{\partial\eta_k}\mathcal{L}(\boldsymbol{\eta}^*,\alpha_2^*,\ldots,\alpha_j^*) = 0,\ \forall k = 1,\ldots,j, &\qquad \text{(Stationarity)}\\ \alpha_k^*(\eta_1^*-\eta_k^*) = 0,\ \forall k = 2,\ldots,j. &\qquad \text{(Complementarity)}\end{aligned}$$

Recall that $\bar{d}(x,\eta) = \phi(x) - \phi(\eta) - (x-\eta)\phi'(\eta)$ (see (13)), the partial differentiation on the second argument is hence $\frac{\partial}{\partial\eta}\bar{d}(x,\eta) = \phi''(\eta)(\eta - x)$. We rewrite the above stationarity condition as:

$$\phi''(\eta_1^*)(\eta_1^*-x_1) + \sum_{k=2}^{j}\alpha_k^* = 0;\ \phi''(\eta_k^*)(\eta_k^*-x_k) = \alpha_k^*,\ \forall k = 2,\ldots,j, \qquad \text{(Stationarity)}$$

Next observe that $\forall i \in \mathcal{I}_j(x) \setminus \{1\}$, it holds that

$$\frac{\sum_{k\in\mathcal{I}_j(\boldsymbol{x})} x_k}{|\mathcal{I}_j(\boldsymbol{x})|} - x_i = \frac{1}{|\mathcal{I}_j(\boldsymbol{x})|}\left(\sum_{k\in\mathcal{I}_j(\boldsymbol{x})} x_k - |\mathcal{I}_j(\boldsymbol{x})|\, x_i\right)$$
$$\geq \frac{1}{|\mathcal{I}_j(\boldsymbol{x})|}\left(x_1 + \sum_{i<k\leq j} x_k - (j-i+1)x_i\right) > 0, \tag{36}$$

where the first inequality stems from $x_1 > x_2 \geq \ldots \geq x_K$, and the last one holds directly from the definition of $\mathcal{I}_j(\boldsymbol{x})$.

One can verify $(\boldsymbol{\eta}^*, \alpha_2^*, \ldots, \alpha_j^*)$ defined below satisfies the KKT conditions listed above.

$$\eta_i^* = \begin{cases} \frac{\sum_{k\in\mathcal{I}_j(\boldsymbol{x})} x_k}{|\mathcal{I}_j(\boldsymbol{x})|}, & \text{if } i \in \mathcal{I}_j(\boldsymbol{x}), \\ x_i, & \text{otherwise,} \end{cases} \quad \alpha_i^* = \begin{cases} \phi''(\frac{\sum_{k\in\mathcal{I}_j(\boldsymbol{x})} x_k}{|\mathcal{I}_j(\boldsymbol{x})|})\left(\frac{\sum_{k\in\mathcal{I}_j(\boldsymbol{x})} x_k}{|\mathcal{I}_j(\boldsymbol{x})|} - x_i\right), & \text{if } i \in \mathcal{I}_j(\boldsymbol{x}), \\ 0, & \text{otherwise,} \end{cases}$$

where $\alpha_i^* \geq 0$, $\forall i \in \mathcal{I}_j(\boldsymbol{x})$ as (36). As for (35), we observe that $\{1, j\} \in \mathcal{I}_j(\boldsymbol{x})$ as $x_1 > x_j$. Hence the above minimizer $\boldsymbol{\eta}^*$ needs to satisfy that $x_j < \eta_1^* \leq \min_{k\in[j]} \eta_k^*$, and (35) follows directly. □

Proposition D.3. *Let $(x_1, \ldots, x_j) \in \mathbb{R}^j$ be such that $x_1 > x_2 \geq \ldots \geq x_j$. Then, for any $k \neq 1$, we have $\frac{\partial}{\partial x_k}\Psi_j(x_1, \ldots, x_j) \leq 0$. Consequently, if $\boldsymbol{y}, \boldsymbol{y}' \in \{\boldsymbol{z} \in \mathbb{R}^j : z_1 > z_2 \geq \ldots \geq z_j\}$ are such that $y_1 = y_1'$ and $y_k \geq y_k'$ for all $k = 2, \ldots, j$, then it follows that $\Psi_j(y_1, y_2, \ldots, y_j) \leq \Psi_j(y_1, y_2', \ldots, y_j')$.*

Proof. Recall that $\bar{d}(x, \eta) = \phi(x) - \phi(\eta) - (x - \eta)\phi'(\eta)$ (see (13)), together with (34) in Proposition D.2, one can deduce that

$$\Psi_j(\boldsymbol{x}) = \sum_{k\in\mathcal{I}_j(\boldsymbol{x})} \phi(x_k) - |\mathcal{I}_j(\boldsymbol{x})|\, \phi(\frac{\sum_{k\in\mathcal{I}_j(\boldsymbol{x})} x_k}{|\mathcal{I}_j(\boldsymbol{x})|}). \tag{37}$$

From the r.h.s of (37), the partial differential on the k-th coordinate yields that

$$\frac{\partial}{\partial x_k}\Psi_j(x_1, \ldots, x_j) = \begin{cases} 0, & \text{if } k \notin \mathcal{I}_j(\boldsymbol{x}), \\ \phi'(x_k) - \phi'(\frac{\sum_{k\in\mathcal{I}_j(\boldsymbol{x})} x_k}{|\mathcal{I}_j(\boldsymbol{x})|}), & \text{otherwise.} \end{cases}$$

Let $i \in \mathcal{I}_j(\boldsymbol{x}) \setminus \{1\}$, the definition of $\mathcal{I}_j(\boldsymbol{x})$ implies that

$$x_i < \frac{\sum_{i<k\leq j} x_k}{j-i+1} \leq \frac{\sum_{k\in\mathcal{I}_j(\boldsymbol{x})} x_k}{|\mathcal{I}_j(\boldsymbol{x})|}.$$

Because ϕ' is a strictly increasing function, we then conclude that $\frac{\partial}{\partial x_k}\Psi_j(x_1, \ldots, x_j) \leq 0$, $\forall k = 2, \ldots, j$. □

We next present the useful forms for $\Gamma_j(\boldsymbol{\mu})$ for any $j = 2, \ldots, K$.

Proposition D.4. *Let $\xi_k = \phi'^{-1}(\mu_k) = \log\frac{\mu_k}{1-\mu_k}$, $\forall k \in [K]$ with $\mu_1 > \mu_2 \geq \ldots \geq \mu_K$, where ϕ is the strictly convex function shown in Appendix A. Then*

$$\Gamma_j(\boldsymbol{\mu}) = \frac{\sum_{k\in\mathcal{I}_j(\boldsymbol{\mu})} \bar{d}(\xi_k, \frac{\sum_{k'\in\mathcal{I}_j(\boldsymbol{\mu})} \xi_{k'}}{|\mathcal{I}_j(\boldsymbol{\mu})|})}{|\mathcal{I}_j(\boldsymbol{\mu})|}, \tag{38}$$

where $\mathcal{I}_j(\boldsymbol{\mu}) = \left\{i \in \{2, \ldots, j\} : \xi_i(j-i+1) < \xi_1 + \sum_{i<k\leq j} \xi_k\right\} \cup \{1\}$. Moreover, the minimizer $(J^*, \boldsymbol{\lambda}^*)$ of (29) satisfies, $J^* = [j]$, $\mu_{j+1} < \lambda_1^* \leq \min_{k\in[j]} \lambda_k^*$, and $\lambda_k^* = \mu_k$ $\forall k \geq j+1$, i.e.,

$$\Gamma_j(\boldsymbol{\mu}) = \inf_{\boldsymbol{\lambda}\in\Lambda_j} \sum_{k=1}^{j} d(\lambda_k, \mu_k), \tag{39}$$

where

$$\Lambda_j = \left\{ \boldsymbol{\lambda} \in (0,1)^K : \mu_{j+1} < \lambda_1 \leq \min_{k \in [j]} \lambda_k, \text{ and } \lambda_k = \mu_k, \forall k \geq j+1 \right\}.$$

Proof. Recall that $\Gamma_j(\boldsymbol{\mu}) = \min_{J \in \mathcal{J}_j(\boldsymbol{\mu})} \inf_{\boldsymbol{\lambda} \in \Lambda : \lambda_1 \leq \min_{k \in J} \lambda_k} \sum_{k \in J} d(\lambda_k, \mu_k)$. The fact that $d(\lambda_k, \mu_k) = \bar{d}(\xi_k, \phi'^{-1}(\lambda_k))$ (see (13)) implies that

$$\begin{aligned} \Gamma_j(\boldsymbol{\mu}) &= \min_{J \in \mathcal{J}_j(\boldsymbol{\mu})} \inf_{\boldsymbol{\eta} \in \mathbb{R}^j : \lambda_1 \leq \min_{k \in J} \lambda_k} \sum_{k \in J} \bar{d}(\xi_k, \eta_k) \\ &= \min_{J \in \mathcal{J}_j(\boldsymbol{\mu})} \Psi_j(\xi_{J_1}, \ldots, \xi_{J_j}), \end{aligned} \tag{40}$$

where J_k denotes the k-th smallest index in J. Recall that $\mathcal{J}_j(\boldsymbol{\mu}) = \{J \subseteq [K] : |J| = j, 1(\boldsymbol{\mu}) \in J\}$, hence $[j] \in \mathcal{J}_j(\boldsymbol{\mu})$. Therefore, from Proposition D.3 and $\xi_1 > \xi_2 \geq \ldots \geq \xi_K$, we deduce that

$$\min_{J \in \mathcal{J}_j(\boldsymbol{\mu})} \Psi_j(\xi_{J_1}, \ldots, \xi_{J_j}) = \Psi_j(\xi_1, \ldots, \xi_j). \tag{41}$$

(38) follows as the consequence of (40), (41), and (34) in Proposition D.2.
As for (39), (40), (41), and (35) in Proposition D.2 yield that

$$\Gamma_j(\boldsymbol{\mu}) = \inf_{\boldsymbol{\eta} \in \mathbb{R}^j : \xi_j < \eta_1 \leq \min_{k \in [j]} \eta_k} \sum_{k=1}^{j} \bar{d}(\xi_k, \eta_k). \tag{42}$$

Using fact that $\bar{d}(\xi_k, \phi'^{-1}(\lambda_k)) = d(\lambda_k, \mu_k)$ again, one can derive (39) from (42). □

E. Examples of Stable Algorithms

In this section, we present various examples of stable algorithms (Definition 3.2). We show that algorithms following one of the design principles below are stable. We assume that in all cases, there is an initialization phase where each arm is sampled $\lfloor \alpha T \rfloor$ times for some $\alpha > 0$. This ensures that the arm rewards will be estimated accurately and that the algorithms are consistent. In the second phase, the algorithm design can be:

1. *Uniform Sampling if Empirically Close.* The algorithm equally samples arms whenever the estimated gap $|\hat{\mu}_1(\tau) - \hat{\mu}_2(\tau)|$ of the mean arm rewards on the $\tau = \lfloor \alpha T \rfloor$-th round falls below a fixed threshold $\varepsilon > 0$ is stable. No rules are added if the estimated gap is above the threshold. The algorithm could, for example, use the estimated optimal static exploration rate $x^*(\hat{\boldsymbol{\mu}}(t)) = \operatorname{argmax}_x g(x, \hat{\boldsymbol{\mu}}(t))$. The algorithm with such a choice is referred to as ETT (Estimate and Thresholded Tracking), and it is discussed in E.1.

2. *Track a Symmetric Continuous Function of the Empirical Rewards.* Here, the algorithm samples arms so that up to round the t-th round, arm 2 has been sampled $tf(\hat{\boldsymbol{\mu}}(t))$ where f is a continuous function satisfying $f(a, a) = 1/2$ for any a and $\hat{\boldsymbol{\mu}}(t)$ denotes the empirical rewards at round t. We refer to this kind of algorithm as TCSF (Track-a-Continuous-Symmetric-Function), and it is discussed in E.2.

We present these algorithms in detail below and establish their stability. We note that the class of algorithms satisfying one of the above design principles is wide, and this makes the class of stable and consistent algorithms relevant. Simple numerical experiments are presented at the end of this section, in E.4.

E.1. The ETT Algorithm

The pseudo-code of ETT is presented in Algorithm 2.

Lemma E.1. *The algorithm ETT with input $\alpha \in (0, 1/2)$ and $\varepsilon > 0$ is stable.*

Algorithm 2 Estimate and Thresholded Tracking (ETT)

Input: $\alpha > 0, \varepsilon > 0$
Play each arm $\max\{\lfloor \alpha T \rfloor, 1\}$ times
$\tau \leftarrow 2\max\{\lfloor \alpha T \rfloor, 1\}$
Estimate the optimal allocation $\hat{x}^* \leftarrow \operatorname{argmax}_x g(x, \hat{\boldsymbol{\mu}}(\tau))$
if $|\hat{\mu}_1(\tau) - \hat{\mu}_2(\tau)| > \varepsilon$ **then**
  for $t = \tau + 1, \ldots, T$ **do**
    play $A_t \leftarrow \begin{cases} 2 & \text{if } \hat{x}^* > \frac{N_2(t)}{t}, \\ 1 & \text{otherwise} \end{cases}$
  end for
else
  for $t = \tau + 1, \ldots, T$ **do**
    play $A_t \leftarrow \operatorname{argmin}_k N_k(t)$ (tie broken arbitrarily)
  end for
end if
$\hat{\imath} \leftarrow \operatorname{argmax}_{k \in \{1,2\}} \hat{\mu}_k(T)$ (tie broken arbitrarily)
Output: $\hat{\imath}$

Proof of Lemma E.1. From the definition of a stable algorithm (Definition 3.2), it suffices to show $\lim_{T\to\infty} \mathbb{E}_{\boldsymbol{\mu}}[\omega_2(T)] = 1/2$ whenever $|\mu_1 - \mu_2| < \varepsilon/3$. We observe that

$$\begin{aligned} |\hat{\mu}_1(2\lfloor \alpha T \rfloor) - \hat{\mu}_2(2\lfloor \alpha T \rfloor)| &\le |\hat{\mu}_1(2\lfloor \alpha T \rfloor) - \mu_1| + |\mu_1 - \mu_2| + |\mu_2 - \hat{\mu}_2(2\lfloor \alpha T \rfloor)| \\ &\le \frac{\varepsilon}{3} + |\hat{\mu}_1(2\lfloor \alpha T \rfloor) - \mu_1| + |\mu_2 - \hat{\mu}_2(2\lfloor \alpha T \rfloor)|, \end{aligned}$$

where the first inequality is from the triangle inequality. Hence,

$$\{|\hat{\mu}_1(2\lfloor \alpha T \rfloor) - \hat{\mu}_2(2\lfloor \alpha T \rfloor)| > \varepsilon\} \subseteq \left\{|\hat{\mu}_1(2\lfloor \alpha T \rfloor) - \mu_1| > \frac{\varepsilon}{3}\right\} \cup \left\{|\hat{\mu}_2(2\lfloor \alpha T \rfloor) - \mu_2| > \frac{\varepsilon}{3}\right\}. \tag{43}$$

Furthermore, the design of Algorithm 2 yields that if $|\hat{\mu}_1(2\lfloor \alpha T \rfloor) - \hat{\mu}_2(2\lfloor \alpha T \rfloor)| \leq \varepsilon$, then $|\omega_2(T) - 1/2| \leq 1/T$. This fact together with (43) yields that

$$\begin{aligned}\mathbb{P}_{\boldsymbol{\mu}}\left[\left|\omega_2(T) - \frac{1}{2}\right| > \frac{1}{T}\right] &\leq \mathbb{P}_{\boldsymbol{\mu}}\left[|\hat{\mu}_1(2\lfloor \alpha T \rfloor) - \hat{\mu}_2(2\lfloor \alpha T \rfloor)| > \varepsilon\right] \\ &\leq \mathbb{P}_{\boldsymbol{\mu}}\left[|\hat{\mu}_1(2\lfloor \alpha T \rfloor) - \mu_1| > \frac{\varepsilon}{3}\right] + \mathbb{P}_{\boldsymbol{\mu}}\left[|\hat{\mu}_2(2\lfloor \alpha T \rfloor) - \mu_2| > \frac{\varepsilon}{3}\right] \\ &\leq 4\exp\left(\frac{-18\lfloor \alpha T \rfloor}{\varepsilon^2}\right), \end{aligned} \tag{44}$$

where the last inequality is an application of Hoeffding inequality. From (44), we can conclude that $\lim_{T\to\infty} \mathbb{E}_{\boldsymbol{\mu}}[\omega_2(T)] = 1/2$ and hence Algorithm 2 is stable.

□

E.2. The TCSF Algorithm

Let $f : [0,1]^2 \to [0,1]$ be a continuous function such that $f(a,a) = 1/2$ for all a. We propose two versions of the TCSF algorithm, one randomized and one de-randomized. Their pseudo-codes are presented in Algorithms 3 and 4, respectively.

Algorithm 3 Randomized TCSF

Input: function f and $\alpha > 0$
Play each arm $\max\{\lfloor \alpha T \rfloor, 1\}$ times
$\tau \leftarrow 2\max\{\lfloor \alpha T \rfloor, 1\}$
for $t = \tau + 1, \ldots, T$ **do**
  play $A_t \leftarrow \begin{cases} 2 & \text{w.p. } f(\hat{\boldsymbol{\mu}}(t)) \\ 1 & \text{w.p. } 1 - f(\hat{\boldsymbol{\mu}}(t)) \end{cases}$
end for
$\hat{\imath} \leftarrow \operatorname{argmax}_{k\in\{1,2\}} \hat{\mu}_k(T)$ (tie broken arbitrarily)
Output: $\hat{\imath}$

Algorithm 4 De-randomized TCSF

Input: function f and $\alpha > 0$
Play each arm $\max\{\lfloor \alpha T \rfloor, 1\}$ times
$\tau \leftarrow 2\max\{\lfloor \alpha T \rfloor, 1\}$
for $t = \tau + 1, \ldots, T$ **do**
  play $A_t \leftarrow \begin{cases} 2 & \text{if } \omega_2(t) < f(\hat{\boldsymbol{\mu}}(t)) \\ 1 & \text{otherwise} \end{cases}$
end for
$\hat{\imath} \leftarrow \operatorname{argmax}_{k\in\{1,2\}} \hat{\mu}_k(T)$ (tie broken arbitrarily)
Output: $\hat{\imath}$

In the following, we show that Algorithm 3 (resp. Algorithm 4) is stable in Lemma E.2 (resp. Lemma E.3).

Lemma E.2. *If $f : [0,1]^2 \mapsto (0,1)$ be a continuous function satisfying that $f(a,a) = 1/2$, $\forall a \in [0,1]$, then Algorithm 3 is stable.*

Proof. Thanks to Lemma E.4 in Appendix E.3, it suffices to show (53) and (54). In the following, we prove (53), and (54) hold in a similar manner. To this aim, we fix $a \in (0,1)$ and $\varepsilon > 0$. As f is continuous at (a,a) and $f(a,a) = 1/2$, there exists $\eta > 0$ such that $|f(x,y) - 1/2| < \varepsilon$ if $\|(x,y) - (a,a)\|_\infty < \eta$. (53) follows provided that we show

$$\underline{\lim}_{T\to\infty} \mathbb{E}_{\boldsymbol{\mu}}[\omega_2(T)] \geq \frac{1}{2} - \varepsilon, \ \forall \boldsymbol{\mu} \in \Lambda \text{ such that } \|\boldsymbol{\mu} - (a,a)\|_\infty < \frac{\eta}{2}. \tag{45}$$

Let $\boldsymbol{\mu} \in \Lambda$ such that $\|\boldsymbol{\mu} - (a,a)\|_\infty < \frac{\eta}{2}$ and $T \in \mathbb{N}$ such that $\alpha T > 1$. We observe

$$\begin{aligned}\mathbb{E}_{\boldsymbol{\mu}}[\omega_2(T)] &\geq \alpha - \frac{1}{T} + \frac{1}{T}\mathbb{E}_{\boldsymbol{\mu}}[\sum_{t=\tau+1}^{T} \mathbb{1}\{A_t = 2\}] \\ &\geq \alpha - \frac{1}{T} + \frac{1}{T}\mathbb{E}_{\boldsymbol{\mu}}[\sum_{t=\tau+1}^{T} \mathbb{1}\{A_t = 2, \|\hat{\boldsymbol{\mu}}(t) - \boldsymbol{\mu}\|_\infty \leq \eta/2\}] \\ &= \alpha - \frac{1}{T} + \frac{1}{T}\mathbb{E}_{\boldsymbol{\mu}}[\sum_{t=\tau+1}^{T} f(\hat{\boldsymbol{\mu}}(t))\mathbb{1}\{\|\hat{\boldsymbol{\mu}}(t) - \boldsymbol{\mu}\|_\infty \leq \eta/2\}], \end{aligned} \tag{46}$$

where the last inequality is simply from the algorithm design. Notice that if $\|\hat{\boldsymbol{\mu}}(t) - \boldsymbol{\mu}\|_\infty \leq \eta/2$, then $\|\hat{\boldsymbol{\mu}}(t) - (a,a)\|_\infty \leq \eta$, and hence $f(\hat{\boldsymbol{\mu}}(t)) > 1/2 - \varepsilon$. We then derive that

$$\begin{aligned}\mathbb{E}_{\boldsymbol{\mu}}[\sum_{t=\tau+1}^{T} f(\hat{\boldsymbol{\mu}}(t))\mathbb{1}\{\|\hat{\boldsymbol{\mu}}(t) - \boldsymbol{\mu}\|_\infty \leq \eta/2\}] &\geq (\frac{1}{2} - \varepsilon)\mathbb{E}_{\boldsymbol{\mu}}[\sum_{t=\tau+1}^{T} \mathbb{1}\{\|\hat{\boldsymbol{\mu}}(t) - \boldsymbol{\mu}\|_\infty \leq \eta/2\}] \\ &= (\frac{1}{2} - \varepsilon)(T - \tau - \mathbb{E}_{\boldsymbol{\mu}}[\sum_{t=\tau+1}^{T} \mathbb{1}\{\|\hat{\boldsymbol{\mu}}(t) - \boldsymbol{\mu}\|_\infty > \eta/2\}]). \end{aligned} \tag{47}$$

As for each $t > \tau$ and $k \in \{1,2\}$, $N_k(t) \geq \alpha T \geq \alpha(t-\tau)$, an application of Lemma E.5 in Appendix E.3 with $H = \{t > \tau\}$, $\zeta = \alpha$ and $\delta = \eta/2$ yields that

$$\mathbb{E}_{\boldsymbol{\mu}}\left[\sum_{t=\tau+1}^{T} \mathbb{1}\{|\hat{\mu}_k(t) - \mu_k| > \frac{\eta}{2}\}\right] \leq \frac{4}{\alpha\eta^2}, \quad \forall k = 1,2. \tag{48}$$

Using (46)-(47)-(48), we conclude that

$$\mathbb{E}_{\boldsymbol{\mu}}[\omega_2(T)] \geq \alpha - \frac{1}{T} + (\frac{1}{2} - \varepsilon)\frac{(T - \tau - 4/\alpha\eta^2)}{T} \geq \alpha - \frac{1}{T} + (\frac{1}{2} - \varepsilon)\frac{(T - \alpha T + 1 - 4/\alpha\eta^2)}{T},$$

and (45) follows. □

Lemma E.3. *If $f : [0,1]^2 \mapsto (0,1)$ be a continuous function satisfying that $f(a,a) = 1/2$, $\forall a \in [0,1]$, then Algorithm 4 is stable.*

Proof. Thanks to Lemma E.4 in Appendix E.3, it suffices to show (53) and (54). In the following, we prove (54), and (53) hold in a similar manner. To this aim, we fix $a \in (0,1)$ and $\varepsilon > 0$. As f is continuous at (a,a) and $f(a,a) = 1/2$, there exists $\eta > 0$ such that $|f(x,y) - 1/2| < \varepsilon$ if $\|(x,y) - (a,a)\|_\infty < \eta$. (54) follows as long as we show

$$\overline{\lim_{T\to\infty}} \mathbb{E}_{\boldsymbol{\mu}}[\omega_2(T)] \leq \frac{1}{2} + \varepsilon, \, \forall \boldsymbol{\mu} \in \Lambda \text{ such that } \|\boldsymbol{\mu} - (a,a)\|_\infty < \frac{\eta}{2}. \tag{49}$$

Let $\boldsymbol{\mu} \in \Lambda$ such that $\|\boldsymbol{\mu} - (a,a)\|_\infty < \frac{\eta}{2}$ and $T \in \mathbb{N}$ such that $\alpha T > 1$. From the algorithm design, we deduce that

$$\begin{aligned}\mathbb{E}_{\boldsymbol{\mu}}[\omega_2(T)] &\leq \alpha + \frac{1}{T} + \frac{1}{T}\mathbb{E}_{\boldsymbol{\mu}}[\sum_{t=\tau+1}^{T} \mathbb{1}\{\omega_2(t) \leq f(\hat{\boldsymbol{\mu}}(t))\}] \\ &\leq \alpha + \frac{1}{T} + \frac{1}{T}\left(\mathbb{E}_{\boldsymbol{\mu}}[\sum_{t=\tau+1}^{T} \mathbb{1}\{\omega_2(t) \leq f(\hat{\boldsymbol{\mu}}(t)), \|\hat{\boldsymbol{\mu}}(t) - \boldsymbol{\mu}\|_\infty \leq \frac{\eta}{2}\}] + \mathbb{E}_{\boldsymbol{\mu}}[\sum_{t=\tau+1}^{T} \mathbb{1}\{\|\hat{\boldsymbol{\mu}}(t) - \boldsymbol{\mu}\|_\infty > \frac{\eta}{2}\}]\right). \end{aligned} \tag{50}$$

As for each $t > \tau$ and $k \in \{1,2\}$, $N_k(t) \geq \alpha T \geq \alpha(t-\tau)$, an application of Lemma E.5 in Appendix E.3 with $H = \{t > \tau\}$, $\zeta = \alpha$ and $\delta = \eta/2$ yields that

$$\mathbb{E}_{\boldsymbol{\mu}}\left[\sum_{t=\tau+1}^{T} \mathbb{1}\{|\hat{\mu}_k(t) - \mu_k| > \frac{\eta}{2}\}\right] \leq \frac{4}{\alpha\eta^2}, \quad \forall k = 1,2. \tag{51}$$

Next we observe that $\|\hat{\boldsymbol{\mu}}(t) - \boldsymbol{\mu}\|_\infty \le \frac{\eta}{2}$ implies that $\|\hat{\boldsymbol{\mu}}(t) - (a,a)\|_\infty \le \eta$, and $f(\hat{\boldsymbol{\mu}}(t)) < 1/2 + \varepsilon$ thanks to the triangle inequality. Thus, the third term in (50) is bound as:

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\mu}}[\sum_{t=\tau+1}^{T} \mathbb{1}\{\omega_2(t) \le f(\hat{\boldsymbol{\mu}}(t)), \|\hat{\boldsymbol{\mu}}(t) - \boldsymbol{\mu}\|_\infty \le \frac{\eta}{2}\}] &\le \mathbb{E}_{\boldsymbol{\mu}}[\sum_{t=\tau+1}^{T} \mathbb{1}\{\omega_2(t) \le \frac{1}{2} + \varepsilon\}] \\
&= \mathbb{E}_{\boldsymbol{\mu}}[\sum_{t=\tau+1}^{T} \mathbb{1}\{N_2(t) \le t(\frac{1}{2} + \varepsilon)\}] \\
&\le \mathbb{E}_{\boldsymbol{\mu}}[\sum_{t=\tau+1}^{T} \mathbb{1}\{N_2(t) \le T(\frac{1}{2} + \varepsilon)\}] \\
&\le T(\frac{1}{2} + \varepsilon) - \lfloor \alpha T \rfloor \\
&\le T(\frac{1}{2} + \varepsilon) - \alpha T + 1. \qquad (52)
\end{aligned}$$

By (51)-(52), we derive (50) is bounded by

$$\alpha + \frac{1}{T} + \frac{1}{T}\left(T(\frac{1}{2} + \varepsilon) - \alpha T + 1 + \frac{4}{\alpha\eta^2}\right).$$

By taking the limit superior on the above upper bound, we get (49). □

E.3. Technical Lemmas

Lemma E.4. *Suppose that an algorithm satisfies*

$$\lim_{(\mu_1,\mu_2)\to(a,a)} \underline{\lim}_{T\to\infty} \mathbb{E}_{\boldsymbol{\mu}}[\omega_2(T)] = \frac{1}{2}, \quad \forall a \in (0,1), \qquad (53)$$

and

$$\lim_{(\mu_1,\mu_2)\to(a,a)} \overline{\lim}_{T\to\infty} \mathbb{E}_{\boldsymbol{\mu}}[\omega_2(T)] = \frac{1}{2}, \quad \forall a \in (0,1). \qquad (54)$$

Then it is a stable algorithm.

Proof. Let $a \in (0,1)$, we show (A) in Definition 3.2 holds, and (B) follows similarly. Consider a sequence $\{\boldsymbol{\lambda}^{(n)}\}_{n\in\mathbb{N}}$ defined as: $\lambda_1^{(n)} = a + \frac{1-a}{2n}$ and $\lambda_2^{(n)} = a - \frac{a}{2n}$ for all $n \in \mathbb{N}$. The assumption (53) implies that

$$\lim_{n\to\infty} \underline{\lim}_{T\to\infty} \mathbb{E}_{\boldsymbol{\lambda}^{(n)}}[\omega_2(T)] = \frac{1}{2}.$$

On the other hand, the assumption (54) implies that

$$\lim_{n\to\infty} \overline{\lim}_{T\to\infty} \mathbb{E}_{\boldsymbol{\lambda}^{(n)}}[\omega_2(T)] = \frac{1}{2}.$$

Thus, (A) is satisfied with the above sequence $\{\boldsymbol{\lambda}^{(n)}\}_{n\in\mathbb{N}}$. □

Lemma E.5 ((Combes & Proutiere, 2014))**.** *Let $\zeta > 0$ and $H \subset \mathbb{N}$ be a (random) set of rounds such that $\{t \in H\}$ is $\mathcal{F}_{t-1}$-measurable for all $t \ge 1$. Furthermore, we assume for each $t \in H$, we have $N_k(t) \ge \zeta \sum_{s=1}^{t} \mathbb{1}_{\{s\in H\}}$. Then for all $\delta > 0$,*

$$\mathbb{E}_{\boldsymbol{\mu}}\left[\sum_{t\ge 1} \mathbb{1}\{t \in H, |\hat{\mu}_k(t) - \mu_k| > \delta\}\right] \le \frac{1}{\zeta\delta^2}.$$

E.4. Numerical Experiments

We illustrate the performance of the ETT algorithm with $\alpha = 1/4$ and different thresholds ε, and compare it to that of the uniform sampling algorithm and to that of an Oracle algorithm that selects arms using optimal exploration rate $x^*(\boldsymbol{\mu}) = \operatorname{argmax}_x g(x, \boldsymbol{\mu})$. We consider the instance: $\boldsymbol{\mu} = (0.0005, 0.0001)$. For this instance, the optimal budget allocation is approximately $x^*(\boldsymbol{\mu}) \approx 0.43434$.

We first examine how the algorithms behave when the sampling budget varies. Figure 5 illustrates the estimated error probabilities as the budget changes from $T = 6000$ to $T = 40000$. The error probabilities are derived from 40000 trials for each setting and algorithm. In all budget scenarios, the Oracle algorithm outperforms the others, while ETT performs comparably or worse than the uniform sampling algorithm. This observation aligns with our Theorem 4.1.

We then investigate the sensitivity of ETT to the input value ε. Figure 6 displays the error probability with a fixed budget of $T = 20000$ and varying ε from 0 to 0.0008. The error probability is again determined from 40000 trials for each setting and algorithm. Regardless of ε, the performance of ETT is similar to or worse than that of the uniform sampling algorithm, further supporting our Theorem 4.1. Given that $\mu_1 - \mu_2 = 0.0004$, the relatively low performance of ETT with $\varepsilon < 0.0004$ compared to that of the uniform sampling algorithm suggests that relying less on the estimated optimal allocation could yield better results.

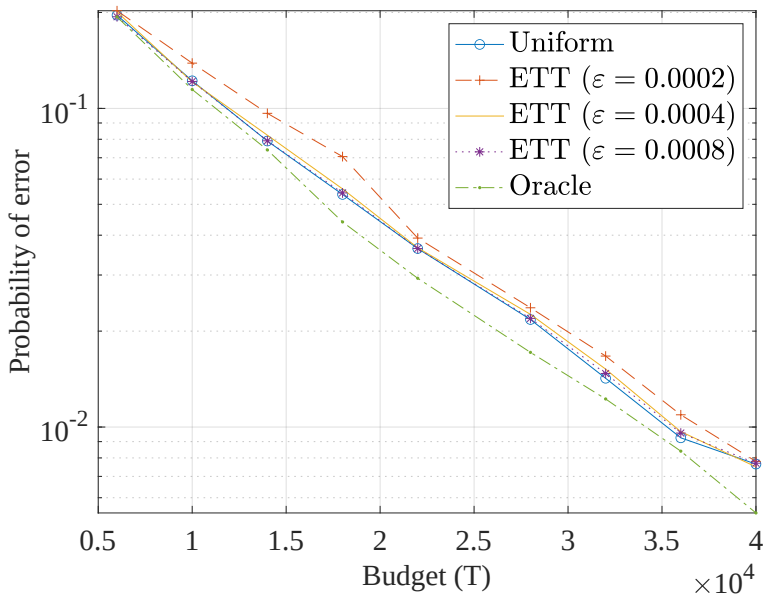

Figure 5. Error probability comparison across algorithms for varying sample budgets ($T = 6000$ to $T = 40000$). Derived from 40000 trials for each setting and algorithm.

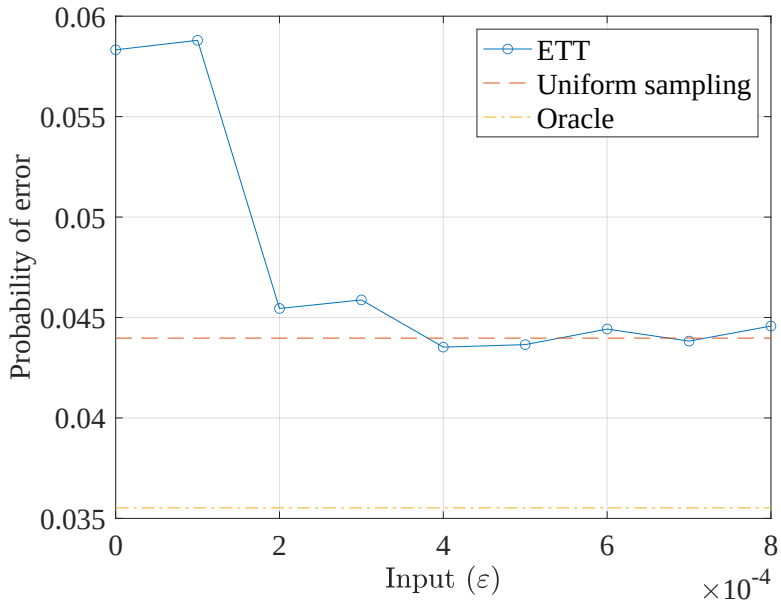

Figure 6. Error probability comparison for varying ETT threshold inputs ($\varepsilon = 0$ to $\varepsilon = 0.0008$). Derived from 40000 trials for each setting and algorithm.