

An Anytime Algorithm for Good Arm Identification

Marc Jourdan*

EPFL, Lausanne, Switzerland

MARC.JOURDAN@EPFL.CH

Andrée Delahaye-Duriez

Université Paris Cité, Inserm, NeuroDiderot, UMR-1141, 75019, Paris, France; Unité fonctionnelle de médecine génomique et génétique clinique, Hôpital Jean Verdier, Assistance Publique des Hôpitaux de Paris, 93140, Bondy, France; Université Sorbonne Paris Nord, 93000, Bobigny, France

ANDREE.DELAHAYE@INSERM.FR

Clémence Réda[†]

BioComp, Institut de Biologie de l'ENS (IBENS UMR 8197), Département de biologie, École normale supérieure, CNRS, Inserm, Université PSL, 75005 Paris, France

REDA@BIO.ENS.PSL.EU

+ Corresponding author

Abstract

In good arm identification (GAI), the goal is to identify one arm whose average performance exceeds a given threshold, referred to as a good arm, if it exists. Few works have studied GAI in the fixed-budget setting when the sampling budget is fixed beforehand, or in the anytime setting, when a recommendation can be asked at any time. We propose APGAI, an anytime and parameter-free sampling rule for GAI in stochastic bandits. APGAI can be straightforwardly used in fixed-confidence and fixed-budget settings. First, we derive upper bounds on its probability of error at any time. They show that adaptive strategies can be more efficient in detecting the absence of good arms than uniform sampling in several diverse instances. Second, when APGAI is combined with a stopping rule, we prove upper bounds on the expected sampling complexity, holding at any confidence level. Finally, we show the good empirical performance of APGAI on synthetic and real-world data. Our work offers an extensive overview of the GAI problem in all settings.

Keywords: multi-armed bandits, pure exploration, good arm identification, fixed-budget setting, anytime setting.

1 Introduction

Multi-armed bandit algorithms are a family of approaches which demonstrated versatility in solving online allocation problems, where constraints exist on the possible allocations: *e.g.* randomized clinical trials (Thompson, 1933; Berry, 2006), hyperparameter optimization (Li et al., 2017; Shang et al., 2018), or active learning (Carpentier et al., 2011). The agents face a black-box environment, upon which they can sequentially act through actions called *arms*. After sampling an arm $a \in \mathcal{A}$, they receive output from the environment through a scalar observation, which is a realization from the unknown probability distribution ν_a of the arm a whose mean will be denoted by μ_a . Depending on their objectives, agents should have different sampling strategies.

*. This work was mainly done when Marc Jourdan was a PhD student in the Scool team of Inria Lille.

†. This work was mainly done when Clémence Réda was a postdoctoral fellow at Université Paris Cité, Inserm, NeuroDiderot, UMR-1141, 75019, Paris, France, and at the Department of Systems Biology and Bioinformatics, University of Rostock, G-18051, Rostock, Germany.

In *pure exploration* problems, the goal is to answer a question about the set of arms. It is studied in two major theoretical frameworks (Audibert et al., 2010; Gabillon et al., 2012; Jamieson and Nowak, 2014; Garivier and Kaufmann, 2016): the *fixed-confidence* and *fixed-budget* setting. In the fixed-confidence setting, the agent aims at minimizing the number of samples used to identify a correct answer with confidence $1 - \delta$, where $\delta \in (0, 1)$ is a *risk* parameter. In the fixed-budget setting, the objective is to minimize the probability of misidentifying a correct answer with a fixed number of samples T , where $T \in \mathbb{N}$ is a *budget* parameter.

While δ or T are assumed given, choosing them is challenging for the practitioner since a “good” choice typically depends on unknown quantities. Moreover, in medical applications (*e.g.* clinical trials or outcome scoring), the maximal budget is limited but might not be fixed beforehand. Independently of the preliminary data, medical applications are prone to reductions in funding or new sources of funding. Therefore, an experiment might stop before the initial budget has been used, referred to as *early stopping*, or continue after it has been consumed, referred to as *continuation*. When the collected data shows sufficient evidence in favor of one answer, an experiment often stops before reaching the initial budget. Given that this early stopping is a data-dependent random variable, it differs fundamentally from the early stopping due to funding shortfalls. While early stopping and continuation are common in practice, both fixed-confidence and fixed-budget settings fail to provide meaningful guarantees for them. Recently, the *anytime* setting has received increased scrutiny as it fills this gap between theory and practice. In the anytime setting, for any fixed deterministic time t that is unknown for the learner, the agent aims at achieving a low probability of error at time t (Jun and Nowak, 2016; Zhao et al., 2023; Jourdan et al., 2024). While T is fixed and known in the fixed-budget setting, t is fixed and unknown in the anytime setting. When the candidate answer has anytime guarantees, the practitioners can use data-independent continuation and early stopping. When combined with a stopping rule, the early stopping can be made data-dependent.

The most studied topic in pure exploration is the *best arm (BAI) / Top- m identification* problem, which aims at determining a subset of m arms with the largest means (Karnin et al., 2013; Xu et al., 2018; Tirinzoni and Degenne, 2022). However, in some applications (*e.g.* investigating treatment protocols), BAI requires too many samples for it to be useful in practice. To avoid wasteful queries, practitioners focus on simpler tasks, *i.e.* identifying one “good enough” option. For instance, in ε -BAI (Mannor and Tsitsiklis, 2004; Even-Dar et al., 2006; Garivier and Kaufmann, 2021; Jourdan et al., 2024), the agent is interested in an arm which is ε -close to the best one, *i.e.* $\mu_a \geq \max_{k \in A} \mu_k - \varepsilon$. The larger ε is, the easier the task. However, choosing a meaningful value of ε can be tricky. In this work, we focus on good arm identification (GAI), where the agent aims to obtain a *good arm*, defined as an arm whose average performance exceeds a given threshold θ , *i.e.* $\mu_a \geq \theta$. GAI and variants are studied in the fixed-confidence setting (Kaufmann et al., 2018; Kano et al., 2019; Tabata et al., 2020), but algorithms for fixed-budget or anytime GAI are missing, despite their practical relevance. We fill this gap by introducing APGAI, an anytime and parameter-free sampling rule for GAI. APGAI is independent of a budget T or a risk δ and is performant in the fixed-budget and fixed-confidence settings.

Our work is motivated by a real-life outcome scoring problem to determine the best treatment protocol for treating the encephalopathy of prematurity in newborns with stem cell injections, in collaboration with the PREMSTEM consortium (see Section 6). In that

case, practitioners have enough information about the distributions associated with each treatment protocol to define a meaningful threshold beforehand.

1.1 Problem Statement

We denote by $\mathcal{D}$ a set to which the distributions of the arms are known to belong. We suppose that all distributions in $\mathcal{D}$ are σ -sub-Gaussian. A distribution ν_0 is σ -sub-Gaussian of mean μ_0 if it satisfies $\mathbb{E}_{X \sim \nu_0}[e^{\lambda(X-\mu_0)}] \leq e^{\sigma^2 \lambda^2 / 2}$ for all $\lambda \in \mathbb{R}$. By rescaling, we assume $\sigma_a = 1$ for all $a \in \mathcal{A}$. Let $\mathcal{A}$ be the set of arms of size K . A bandit instance is defined by unknown distributions $\nu := (\nu_a)_{a \in \mathcal{A}} \in \mathcal{D}^K$ with means $\mu := (\mu_a)_{a \in \mathcal{A}} \in \mathbb{R}^K$. Given a threshold $\theta \in \mathbb{R}$, the set of good arms is defined as $\mathcal{A}_\theta(\mu) := \{a \in \mathcal{A} \mid \mu_a \geq \theta\}$, which we shorten to $\mathcal{A}_\theta$ when μ is unambiguous. In the remainder of the paper, we assume that $\mu_a \neq \theta$ for all $a \in \mathcal{A}$. Let the gap of arm a compared to θ be $\Delta_a := |\mu_a - \theta| > 0$. Let $\Delta_{\min} = \min_{a \in \mathcal{A}} \Delta_a$ be the minimum gap over all arms. Let

$$H_1(\mu) := \sum_{a \in \mathcal{A}} \Delta_a^{-2} \quad \text{and} \quad H_\theta(\mu) := \sum_{a \in \mathcal{A}_\theta(\mu)} \Delta_a^{-2}. \quad (1)$$

At time t , the agent chooses an arm $a_t \in \mathcal{A}$ based on past observations and receives a sample $X_{a_t, t}$, random variable with conditional distribution ν_{a_t} given a_t . Let $\mathcal{F}_t := \sigma(a_1, X_{a_1, 1}, \dots, a_t, X_{a_t, t})$ be the σ -algebra, called *history*, which encompasses all the information available to the agent after t rounds.

Identification algorithm. In the anytime setting, an *identification* algorithm defines two rules which are $\mathcal{F}_t$ -measurable at time t : a sampling rule $a_{t+1} \in \mathcal{A}$ and a recommendation rule $\hat{a}_t \in \mathcal{A} \cup \{\emptyset\}$. In GAI, the probability of error $P_{\nu, \mathfrak{A}}^{\text{err}}(t) := \mathbb{P}_\nu(\mathcal{E}_{\mathfrak{A}}^{\text{err}}(t))$ of algorithm $\mathfrak{A}$ on instance μ at time t is the probability of the error event $\mathcal{E}_{\mathfrak{A}}^{\text{err}}(t) = \{\hat{a}_t \in \{\emptyset\} \cup (\mathcal{A} \setminus \mathcal{A}_\theta)\}$ when $\mathcal{A}_\theta \neq \emptyset$, otherwise $\mathcal{E}_{\mathfrak{A}}^{\text{err}}(t) = \{\hat{a}_t \neq \emptyset\}$ when $\mathcal{A}_\theta = \emptyset$.

Those rules have a different objective depending on the considered setting. In fixed-budget GAI, given a fixed and known budget T , the goal is to have a low $P_{\nu, \mathfrak{A}_T}^{\text{err}}(T)$, where $\mathfrak{A}_T$ highlights the dependency in T of the algorithm. In anytime GAI, the objective is to ensure that $P_{\nu, \mathfrak{A}}^{\text{err}}(t)$ is small at any fixed time t , that is unknown for $\mathfrak{A}$. Whereas in fixed-confidence GAI, these two rules are complemented by a stopping rule using a confidence level $1 - \delta$ fixed beforehand such that the algorithm stops sampling after τ_δ rounds. The stopping time τ_δ is also known as the (verifiable) *sample complexity* of a fixed-confidence algorithm. A fixed-confidence algorithm $\mathfrak{A}_\delta$ always depends on δ due to the stopping time. When the sampling and recommendation rules are independent of δ (*i.e.* anytime) as in APGAI, the notation $\mathfrak{A}$ is used. At stopping time τ_δ , the algorithm should satisfy δ -correctness, which means that $\mathbb{P}_\nu(\{\tau_\delta < +\infty\} \cap \mathcal{E}_{\mathfrak{A}}^{\text{err}}(\tau_\delta)) \leq \delta$ for all instances ν . That requirement leads to a lower bound on the expected sample complexity for any instance. The following lemma is similar to other bounds derived in various settings linked to GAI (Kaufmann et al., 2018; Tabata et al., 2020). The proof in Appendix E.1 relies on the change of measure inequality in Lemma 1 from Kaufmann et al. (2016).

Lemma 1 *Let $\delta \in (0, 1)$. For all δ -correct algorithm and all Gaussian instances $\nu_a = \mathcal{N}(\mu_a, 1)$ with $\mu_a \neq \theta$, we have $\liminf_{\delta \rightarrow 0} \mathbb{E}_\nu[\tau_\delta] / \log(1/\delta) \geq T^*(\mu)$, where*

$$T^*(\mu) := 2 \min_{a \in \mathcal{A}_\theta(\mu)} \Delta_a^{-2} \quad \text{if } \mathcal{A}_\theta(\mu) \neq \emptyset, \quad \text{and} \quad 2H_1(\mu) \quad \text{otherwise.} \quad (2)$$

A fixed-confidence algorithm is *asymptotically optimal* if it is δ -correct, and its expected sample complexity matches the lower bound, *i.e.* $\limsup_{\delta \rightarrow 0} \mathbb{E}_\nu[\tau_\delta] / \log(1/\delta) \leq T^*(\mu)$.

Introduced in Katz-Samuels and Jamieson (2020), the *unverifiable sample complexity* $\tau_{U,\delta}$ is the minimum number of samples after which the algorithm always outputs a correct answer with probability at least $1 - \delta$, namely $\mathbb{P}_\nu(\bigcup_{t \geq \tau_{U,\delta}} \mathcal{E}_\mathbf{A}^{\text{err}}(\tau_\delta)) \leq \delta$ for all instances ν . Compared to the fixed-confidence setting, the unverifiable sample complexity of a strategy is not sufficient to stop and certify a correct output with confidence $1 - \delta$.

Notation. For two probability distributions $\mathbb{P}$ and $\mathbb{Q}$ on the measurable space $(\Omega, \mathcal{G})$, the Total Variation (TV) distance is $\text{TV}(\mathbb{P}, \mathbb{Q}) := \sup_{A \in \mathcal{G}} |\mathbb{P}(A) - \mathbb{Q}(A)|$ and the Kullback-Leibler (KL) divergence is $\text{KL}(\mathbb{P}, \mathbb{Q}) := \int \log\left(\frac{d\mathbb{P}}{d\mathbb{Q}}(\omega)\right) d\mathbb{P}(\omega)$, when $\mathbb{P} \ll \mathbb{Q}$, and $+\infty$ otherwise. For any stopping time τ , let $\mathbb{P}_\nu^\tau$ be the restriction of $\mathbb{P}_\nu$ to the σ -algebra generated by τ . For any τ -measurable event E , we have $\mathbb{P}_\nu^\tau(E) = \mathbb{P}_\nu(E)$.

1.2 Contributions

We introduce APGAI (Algorithm 1 in Section 2), an anytime and parameter-free sampling rule for GAI in stochastic bandits, which is independent of a budget T or a risk δ . APGAI is the first algorithm that can be employed without modification for fixed-budget GAI (and without prior knowledge of the budget) and fixed-confidence GAI. Furthermore, it enjoys guarantees in both settings. As such, APGAI allows both continuation and early stopping. First, we show an upper bound on the probability of error of APGAI at any fixed and unknown time t of the order $\exp(-\mathcal{O}(t/H_1(\mu)))$ which holds for any deterministic time t (Theorem 2 in Section 3). Adaptive strategies are more efficient in detecting the absence of good arms than uniform sampling. Second, we obtain a deterministic upper bound on the unverifiable sample complexity of APGAI holding at any confidence level and scaling as $\mathcal{O}(H_1(\mu) \log(H_1(\mu)/\delta))$ (Theorem 4 in Section 4). Third, when combined with a GLR stopping rule (Lemma 7), we derive a non-asymptotic upper bound on the expected sample complexity of APGAI, whose δ -independent term scales as $\mathcal{O}(H_1(\mu) \log H_1(\mu))$ (Theorem 8 in Section 5). For GAI with Gaussian distributions, APGAI is asymptotically optimal when there is no good arm, yet it is suboptimal when there are good arms. Forth, when there exists a unique good arm and the risk is moderate, we show that a linear dependence in K on the number of samples allocated to suboptimal arms is actually unavoidable (Theorem 5, Corollaries 6 and 9). Fifth, APGAI is easy to implement, computationally inexpensive, and has good empirical performance in both settings on synthetic and real-world data with an outcome scoring problem for RNA-sequencing data (see Section 6). Finally, we provide extensive theoretical and empirical comparisons with other GAI algorithms in all settings, while deriving new guarantees for them as well. For clarity, the lower bounds are summarized in Table 1 and the upper bounds are compared in Tables 2, 3 and 4. Overall, our work offers a compelling overview of the GAI problem, which has previously received little attention despite its practical relevance.

1.3 Related Works

GAI is not studied in a fixed-budget or anytime setting yet. In the fixed-confidence setting, several problems are considered that are similar to GAI.

Setting		Performance Metric	$\mathcal{A}_\theta(\mu) = \emptyset$	$\mathcal{A}_\theta(\mu) \neq \emptyset$
FB	[Thm 3]	$\max_{a \in [K]} \limsup_{T \rightarrow +\infty} \frac{T}{-\log P_{\nu^{(a)}, \mathfrak{A}_T}^{\text{err}}(T)}$	—	$\frac{2K}{(\Delta + \varepsilon)^2}$
UC	[Cor 6]	$\max_{a \in [K]} \mathbb{E}_{\nu^{(a)}}[\tau_{U, \delta} - N_a(\tau_{U, \delta})]$	—	$\frac{K-1}{64(\Delta + \varepsilon)^2}$
FC	[Cor 9]	$\max_{a \in [K]} \mathbb{E}_{\nu^{(a)}}[\tau_\delta - N_a(\tau_\delta)]$	—	$\frac{K-1}{64(\Delta + \varepsilon)^2}$
	[Lem 1]†	$\liminf_{\delta \rightarrow 0} \frac{\mathbb{E}_\nu[\tau_\delta]}{\log(1/\delta)}$	$H_1(\mu)$	$\frac{1}{\Delta_{\max}^2}$

Table 1: Lower bound on the performance of any GAI algorithm for different objectives and metrics of performance: FC (fixed confidence), FB (fixed budget) and UC (unverifiable sample complexity). Let $(\nu^{(a)})_{a \in [K]}$ be the Gaussian instances defined in Theorem 3 based on $(\Delta, \varepsilon) \in (\mathbb{R}_+^*)^2$, namely, for all $a \in [K]$, $\mathcal{A}_\theta(\nu^{(a)}) = \{a\}$, $\Delta_a = \Delta$ and $\Delta_b = \varepsilon$ for all $b \neq a$. (†) Holds for any instance ν . $H_1(\mu)$ as in Eq. (1), $\Delta_{\min} := \min_{a \in \mathcal{A}} \Delta_a$ and $\Delta_{\max} := \max_{a \in \mathcal{A}_\theta} \Delta_a$. $N_a(t)$ is the number of samples pulled from arm a up to time t included.

Given two thresholds $\theta_L < \theta_U$, Tabata et al. (2020); Hayashi et al. (2024) study the Bad Existence Checking problem, in which the agent should output “negative” if $\mathcal{A}_{\theta_L}(\mu) = \emptyset$ and “positive” if $\mathcal{A}_{\theta_U}(\mu) \neq \emptyset$. In particular, Tabata et al. (2020) proposes an elimination-based meta-algorithm called BAEC, and analyzes its expected sample complexity when combined with several index policies to define the sampling rule. Hayashi et al. (2024) focus on classification bandits with margin, which is a variant of the problem where the expected rewards are sampled from a Gaussian process prior, and describe a similar phased-elimination meta-algorithm that leverages the prior assumption.

Kano et al. (2019) considers identifying the whole set of good arms $\mathcal{A}_\theta(\mu)$ with high probability, and returns λ good arms sequentially, where $\lambda \in \{1, 2, \dots, |\mathcal{A}_\theta(\mu)|\}$. We refer to that problem as AllGAI. Kano et al. (2019) introduce three index-based GAI algorithms named APT-G, HDoC, and LUCB-G, and show upper bounds on their expected sample complexity. In the fixed-confidence setting and for Bernoulli distributions, Tsai et al. (2024) built upon the HDoC algorithm for AllGAI, by fine-tuning the number of uniform pulls at the start of the HDoC algorithm. Their contribution is targeted at instances when one of the arms has an expected reward close to the threshold θ or if two arms have similar expected rewards. A variant of the HDoC algorithm copes for structured versions of fixed-confidence AllGAI, *e.g.* see Tsai et al. (2025) for linear bandits where the expected reward depends on the arm’s feature vector.

Numerous algorithms from previously mentioned works bear a passing resemblance to the APT algorithm proposed by Locatelli et al. (2016) to tackle the thresholding bandit problem in the fixed-budget setting. The latter should classify all arms into $\mathcal{A}_\theta$ and $\mathcal{A}_\theta^c$ at the end of the sampling phase. The resemblance to the APT algorithm lies in that those prior works rely on an arm index for sampling. The arm indices in BAEC (Tabata et al., 2020), APT-G, HDoC and LUCB-G (Kano et al., 2019) are reported in Algorithm 2 in Appendix D. However, it should be noted that our contribution APGAI does not feature an elimination algorithm as those algorithms do and that those prior works hold in a fixed-confidence setting and do not convert straightforwardly to the GAI problem. Moreover, our analyses strongly differ from those present in these prior works. For linear bandits, this problem has also recently received

attention in the fixed-confidence setting as well (Rivera and Tewari, 2024). Other structured versions of thresholding bandits have also been recently considered. For instance, Cheshire et al. (2021) considered specific shape constraints on μ , such as monotonic increasing or concave series of means, in a fixed-budget setting. Leveraging these strong assumptions on the ordering of arm means, authors showed that a lower bound on the asymptotic rate on the error probability roughly scales with $\Delta_{\min}^{-2}$, without dependency on the number of arms K , and that nearly-matching-up to logarithmic factors-algorithms based on binary search exist. Mason et al. (2022) studied linear kernel thresholding bandits in a fixed-confidence setting, where the arm means can be approximated in a Reproducing Kernel Hilbert Space (RKHS) with a known level of misspecification and proposed a nearly-matching algorithm for the linear (kernelized) setting. However, in our paper, we make no assumption on the structure of the bandit instance.

More loosely related works include the all- ε good arm identification problem in a fixed-confidence setting, where the goal is to identify all arms in $\{a \mid \mu_a \geq \max_i \mu_i - \varepsilon\}$ with high probability $1 - \delta$ (Mason et al., 2020). In the moderate confidence regime, Mason et al. (2020) derive a lower bound scaling as $H_1(\mu)$, where the sample complexity average over several instances whose best arm is separated by at least 2β from the other arms. Their proof builds on a reduction to the isolated instance testing problem (see Appendix D), where the goal is to detect whether an arm has mean β or $-\beta$, while the other means are smaller than $-\beta$. It is possible to adapt Mason et al. (2020, Algorithm 4) to solve isolated instance testing with a GAI algorithm for $\theta = 0$, with provable guarantees only on instances with a unique good arm. Leveraging this reduction, Mason et al. (2020, Theorem D.6) yields a lower bound scaling as $H_1(\mu)$ on at least one of these instances with a unique good arm. Mason et al. (2020, Theorem D.6) is derived by using the *Simulator* argument of Simchowitz et al. (2017) that builds non-stationary bandit instances. Keeping the core idea of non-stationarity, Al Marjani et al. (2022) proposed a simpler proof technique to obtain lower bounds with a linear dependency in K . Poiani et al. (2025) adapted their arguments to study BAI on Unimodal instances, where the mean vector is an unimodal function of its indices. While Lemma 1 suggests that only one arm should be sampled asymptotically for GAI with good arms, at most 3 arms are needed according to the asymptotic lower bound for Unimodal BAI. However, Poiani et al. (2025, Theorem 2.3) shows that a linear dependence in K is unavoidable. Building on their proof technique, we derive a general lower bound for any strategy whose stopping time satisfies a lower bound constraint on the TV distance between the distributions generated by interacting with instances having different answers (Theorem 5).

Finally, Degenne and Koolen (2019) addressed the “any low arm” problem, which is a GAI problem for threshold $-\theta$ on instance $-\mu$. They introduce Sticky Track-and-Stop, which is asymptotically optimal in the fixed-confidence setting. In Kaufmann et al. (2018), the “bad arm existence” problem aims to answer “no” when $\mathcal{A}_{-\theta}(-\mu) = \emptyset$, and “yes” otherwise. They propose an adaptation of Thompson Sampling conditioning on the “worst event” (named Murphy Sampling). The empirical pulling proportions converge towards the allocation that realizes $T^*(\mu)$ in Lemma 1. Another related framework is the identification with a high probability of k arms from $\mathcal{A}_\theta(\mu)$ (Katz-Samuels and Jamieson, 2020).

Algorithm 1 APGAI

- 1: **Input:** threshold θ , set of arms $\mathcal{A}$
 - 2: **Initialization:** Draw each arm once
 - 3: **Update:** empirical means $\hat{\mu}(t)$ and empirical transportation costs $W_a^\pm(t)$ as in Eq. (3)
 - 4: **if** $\max_{a \in \mathcal{A}} \hat{\mu}_a(t) \leq \theta$ **then**
 - 5: $\hat{a}_t := \emptyset$ and $a_{t+1} \in \arg \min_{a \in \mathcal{A}} W_a^-(t)$
 - 6: **else**
 - 7: $\hat{a}_t := a_{t+1} \in \arg \max_{a \in \mathcal{A}} W_a^+(t)$
 - 8: **end if**
 - 9: **return** arm to pull a_{t+1} and recommendation $\hat{a}_t$
-

2 Anytime Parameter-free Sampling Rule

We propose the APGAI (Anytime **P**arameter-free **G**AI) algorithm, which is independent of a budget T or a risk δ and summarized in Algorithm 1.

Notation. Let $N_a(t) = \sum_{s \leq t} \mathbb{1}(a_s = a)$ be the number of times arm a is sampled at the end of round t , and $\hat{\mu}_a(t) = \frac{1}{N_a(t)} \sum_{s \leq t} \mathbb{1}(a_s = a) X_{a,s}$ be its empirical mean. For all $a \in \mathcal{A}$ and all $t \geq K$, let us define

$$W_a^+(t) = \sqrt{N_a(t)} \Delta_a(t)_+ \quad \text{and} \quad W_a^-(t) = \sqrt{N_a(t)} (-\Delta_a(t))_+, \quad (3)$$

where $(x)_+ := \max(x, 0)$ and $\Delta_a(t) := \hat{\mu}_a(t) - \theta$. If arm a were a σ_a -sub-Gaussian distribution, the rescaling boils down to using $\Delta_a(t)/\sigma_a$ instead of $\Delta_a(t)$. This empirical transportation cost $W_a^+(t)$ (resp. $W_a^-(t)$) represents the amount of information collected so far in favor of the hypothesis that $\{\mu_a > \theta\}$ (resp. $\{\mu_a < \theta\}$). It is linked with the generalized likelihood ratio (GLR) as detailed in Appendix E.2. As initialization, we pull each arm once.

Recommendation rule. At time $t + 1 > K$, the recommendation rule depends on whether the highest empirical mean lies below the threshold θ or not. When $\max_{a \in \mathcal{A}} \hat{\mu}_a(t) \leq \theta$, we recommend the empty set, *i.e.* $\hat{a}_t = \emptyset$. Otherwise, our candidate answer is the arm which is the most likely to be a good arm given the collected evidence, *i.e.* $\hat{a}_t \in \arg \max_{a \in \mathcal{A}} W_a^+(t)$.

Sampling rule. The next arm to pull is based on the APT_P indices introduced by Tabata et al. (2020) as a modification to the APT indices (Locatelli et al., 2016). At time $t + 1 > K$, we pull arm $a_{t+1} \in \arg \max_{a \in \mathcal{A}} \sqrt{N_a(t)} (\hat{\mu}_a(t) - \theta)$. To emphasize the link with our recommendation rule, this sampling rule can also be written as $a_{t+1} \in \arg \min_{a \in \mathcal{A}} W_a^-(t)$ when $\max_{a \in \mathcal{A}} \hat{\mu}_a(t) \leq \theta$, and $a_{t+1} \in \arg \max_{a \in \mathcal{A}} W_a^+(t)$ otherwise. Ties are broken arbitrarily at random, up to the constraint that $\hat{a}_t = a_{t+1}$ when $\max_{a \in \mathcal{A}} \hat{\mu}_a(t) > \theta$. This formulation better highlights the dual behavior of APGAI, which is reminiscent of the expression of the characteristic time $T^*(\mu)$ in Lemma 1. When $\max_{a \in \mathcal{A}} \hat{\mu}_a(t) \leq \theta$, APGAI collects additional observations to verify that there are no good arms, hence pulling the arm which is the least likely to not be a good arm. Otherwise, APGAI gathers more samples to confirm its current belief that there is at least one good arm, hence pulling the arm that is the most likely to be a good arm. In contrast to indices solely based on the empirical means, the APT_P indices are linked to the empirical transportation costs, which account for the empirical counts.

Memory and computational cost. APGAI needs to maintain in memory the values $N_a(t), \hat{\mu}_a(t), W_a^\pm(t)$ for each arm $a \in \mathcal{A}$, hence the total memory cost is in $\mathcal{O}(K)$. The computational cost of APGAI is in $\mathcal{O}(K)$ per iteration, and its update cost is in $\mathcal{O}(1)$.

Differences to BAEC. While both APGAI and BAEC(APT_P) rely on the APT_P indices (Tabata et al., 2020), they differ significantly and we proceed differently from Tabata et al. (2020) in the analysis of APGAI, partially due to the lack of elimination in the latter. BAEC is an elimination-based meta-algorithm that samples active arms and discards arms whose upper confidence bounds (UCB) on the empirical means are lower than θ_U . The recommendation rule of BAEC is only defined at the stopping time, and it depends on lower confidence bounds (LCB) and UCB. Since the UCB/LCB indices depend inversely on the gap $\theta_U - \theta_L > 0$ and on the confidence δ , BAEC is neither anytime nor parameter-free. More importantly, APGAI can be used without modification for fixed-confidence or fixed-budget GAI. In contrast, BAEC can solely be used in the fixed-confidence setting when $\theta_U > \theta_L$, hence not for GAI itself (*i.e.* $\theta_U = \theta_L$).

3 Anytime Guarantees on the Probability of Error

To allow continuation or (deterministic) early stopping, the candidate answer of APGAI should be associated with anytime theoretical guarantees. Theorem 2 shows an upper bound of the order $\exp(-\mathcal{O}(t/H_1(\mu)))$ for $P_{\nu, \mathfrak{A}}^{\text{err}}(t)$ that holds for any deterministic time t .

Theorem 2 *The APGAI algorithm $\mathfrak{A}$ satisfies that, for all $\nu \in \mathcal{D}^K$ with mean μ such that $\Delta_{\min} > 0$, for all $t > K + 2|\mathcal{A}_\theta|$,*

$$P_{\nu, \mathfrak{A}}^{\text{err}}(t) \leq Ke\sqrt{2} \log(e^2 t) \exp \left(-p \left(\frac{t - K - 2|\mathcal{A}_\theta|}{2\alpha_{i_\mu} H_1(\mu)} \right) \right) \quad \text{with} \quad p(x) = x - 0.5 \log x,$$

where $H_1(\mu)$ as in Eq. (1), $(\alpha_1, \alpha_\theta) = (9, 2)$ and $i_\mu = 1 + (\theta - 1)\mathbb{1}(\mathcal{A}_\theta(\mu) \neq \emptyset)$.

While anytime upper bounds on the probability of error exist in (ε) -BAI (Zhao et al., 2023; Jourdan et al., 2024), Theorem 2 is the first result of its kind for GAI. Our result holds for any deterministic time $t > K + 2|\mathcal{A}_\theta|$ and any 1-sub-Gaussian instance ν . In the asymptotic regime where $t \rightarrow +\infty$, Theorem 2 shows that $\limsup_{t \rightarrow +\infty} t \log(1/P_{\nu, \mathfrak{A}}^{\text{err}}(t))^{-1} \leq 2\alpha_{i_\mu} H_1(\mu)$ for APGAI with $(\alpha_1, \alpha_\theta) = (9, 2)$. We defer the reader to Appendix B for detailed proof.

Comparison with uniform sampling. Despite the practical relevance of anytime and fixed-budget guarantees, APGAI is the first algorithm enjoying guarantees on the probability of error in GAI at any time t (hence at a given budget T). As a baseline, we consider the uniform round-robin algorithm, named Unif, which returns the best empirical arm at time t if its empirical mean is higher than θ , and returns $\emptyset$ otherwise. At a time t such that $t/K \in \mathbb{N}$, the recommendation of Unif is equivalent to the one used in APGAI, *i.e.* $\arg \max_{a \in \mathcal{A}} W_a^+(t) = \arg \max_{a \in \mathcal{A}} \hat{\mu}_a(t)$ since $N_a(t) = t/K$. As the two algorithms differ in their sampling rule, we can measure the benefits of adaptive sampling. Theorem 21 in Appendix C.1.1 gives anytime upper bounds on $P_{\nu, \text{Unif}}^{\text{err}}(t)$, and we compare it to the ones of Theorem 2. In the asymptotic regime, the upper bound for Unif has a rate in $2K\Delta_{\min}^{-2}$ when $\mathcal{A}_\theta(\mu) = \emptyset$, and $4K \min_{a \in \mathcal{A}_\theta(\mu)} \Delta_a^{-2}$ otherwise. While the latter rate is better than $2H_1(\mu)$ when arms have dissimilar gaps, APGAI has better guarantees than Unif when there is no good arm. Our experiments show that APGAI can outperform Unif in many instances (*e.g.* Figures 1 and 2, and the experiments in Appendix I), and is on par with it otherwise. In particular, the upper bound derived for APGAI when $\mathcal{A}_\theta \neq \emptyset$ is not aligned with its good

empirical performance. We conjecture that APGAI could have a better dependency than $H_1(\mu)$ when there are good arms, yet our non-asymptotic analysis is not tight enough to reveal it. Proving this conjecture is an interesting direction for future work that requires finer non-asymptotic arguments. Even with the tightest analysis, Theorem 3 below shows that APGAI can not dominate Unif in all instances.

3.1 Lower Bound with Dependence on the Number of Arms

Degenne (2023) recently studied the existence of complexity in fixed-budget pure exploration. For the fixed-budget GAI problem, Degenne (2023, Theorem 6) shows that uniform sampling is asymptotically minimax optimal with a maximal difficulty ratio equal to K . While there is a complexity $T^*(\mu)$ as in Eq. (2) for the fixed-confidence setting, Degenne (2023, Theorem 6) refutes its existence for fixed-budget GAI if the class of algorithms contains the static proportions algorithms: the asymptotic rate on the probability of error cannot be smaller than $KT^*(\mu)$ on all Gaussian instances ν . Based on Degenne (2023, Corollary 4), Theorem 3 states the intermediate result supporting this negative result.

Theorem 3 (Theorem 6 in Degenne (2023)) *Let $(\theta, \Delta, \varepsilon) \in \mathbb{R} \times (\mathbb{R}_+^*)^2$. For $a \in [K]$, let $\nu^{(a)} := \mathcal{N}(\mu^{(a)}, I_K)$ where $\mu_a^{(a)} = \theta + \Delta$ and $\mu_b^{(a)} = \theta - \varepsilon$ if $b \neq a$. Let $\nu^{(\emptyset)} := \mathcal{N}(\mu^{(\emptyset)}, I_K)$ where $\mu_a^{(\emptyset)} = \theta - \varepsilon$ for all $a \in [K]$. For any sequence of fixed-budget algorithms $(\mathfrak{A}_T)_T$, we have either $-\log P_{\nu^{(\emptyset)}, \mathfrak{A}_T}^{err}(T) =_{T \rightarrow +\infty} o(T)$ or*

$$\exists a \in [K], \quad \limsup_{T \rightarrow +\infty} \frac{T}{-\log P_{\nu^{(a)}, \mathfrak{A}_T}^{err}(T)} \geq \frac{2K}{(\Delta + \varepsilon)^2} = \frac{KT^*(\mu^{(a)})}{(1 + \varepsilon/\Delta)^2}. \quad (4)$$

Proof Obtaining Eq. (4) from Degenne (2023, Corollary 4) is done by using the definitions therein, Lemma 1 and the symmetry of the KL divergence for Gaussian distributions with known variance. $\blacksquare$

While not being valid for all instances, Theorem 3 holds for the class of all algorithm families, which includes the static algorithms. Intuitively, an initial exploration phase is necessary: any algorithm has to sample all arms before starting to recommend the unique good arm (*i.e.* the best one). As an arm a is sampled less than the others, the algorithm is slower on $\nu^{(a)}$. Similarly, for fixed-budget BAI with $K = 2$ and Bernoulli distributions, Wang et al. (2024a) showed that an adaptive algorithm that performs as well as the uniform sampling algorithm on all instances can not outperform it in some instances. Within a large class of consistent and stable algorithms, uniform sampling is universally optimal. Extending their result to an arbitrary number of arms is challenging, yet SR is worse than uniform sampling in some 3-armed instances by comparing an asymptotic lower bound for the former with an upper bound for the latter. Based on these prior results, one has little hope for a better bound in fixed-budget GAI for an arbitrary number of arms.

Unif achieves the rate $KT^*(\mu)$ when $\mathcal{A}_\theta \neq \emptyset$, but suffers from worse guarantees otherwise. Conversely, APGAI achieves the rate in $T^*(\mu)$ when $\mathcal{A}_\theta = \emptyset$, but has sub-optimal guarantees otherwise. It does not conflict with Eq. (4) *e.g.* considering μ with $\mathcal{A}_\theta \neq \emptyset$ and an arm $a \in \mathcal{A}$ with $\Delta_a \leq \max_{a \in \mathcal{A}_\theta} \Delta_a / \sqrt{K/2 - 1}$.

In fixed-budget GAI, a “good” algorithm has highly different sampling modes depending on whether there is a good arm or not. Since committing to one of those modes too early

will incur higher error, it is challenging to find the perfect trade-off adaptively. As uniform sampling is asymptotically minimax optimal with a maximal difficulty ratio equal to K , even the smartest adaptive sampling strategy will not dominate Unif in all instances. Even the “perfect” adaptive trade-off is unable to consistently outperform the most “naive” choice. We emphasize that the above comment also hold for anytime GAI.

Trade-off between the anytime and fixed-budget setting. The negative result of Theorem 3 does not explicitly leverage the fact that the sequence of fixed-budget algorithms $(\mathfrak{A}_T)_T$ have prior knowledge on the budget T . Therefore, it trivially holds for any anytime algorithm $\mathfrak{A}$. To the best of our understanding, it is challenging to incorporate this prior knowledge into the current information-theoretic proofs. When considering the asymptotic rate, we conjecture that the knowledge of T is “irrelevant”. For large T , the probability of error is exponentially small: the algorithm already “knows” the unknown instance’s correct answer. For small budget T , fixed-budget algorithms might have an “hedge” over anytime algorithms. Intuitively, any adaptive algorithm should behave closely to uniform sampling when T is small compared to the difficulty of the instance (*i.e.* too small for identification). Any deviation from this “naive” choice would incur a large probability of error on at least one alternative instance whose answer is different. Since the difficulty of the encountered instance is unknown, a fixed-budget algorithm should determine whether it has enough budget to be “smarter” than uniform, while staying close to it in case the budget is insufficient. An anytime algorithm should also understand whether it can be “smarter” than uniform sampling that is minimax optimal (Theorem 3). Yet, it does not know when it will be evaluated (*i.e.* t is fixed but unknown). However, given the knowledge of T , a fixed-budget algorithm might anticipate this evaluation. If the budget is close to be reached without “knowing” the difficulty of the instance, it could behave closer to uniform sampling to minimize the probability of error by collecting information on all the arms. Despite being intuitive, we emphasize that the above distinction has no theoretical grounding yet (to the best of our knowledge). Given our current non-asymptotic techniques, it seems almost impossible to derive theoretical guarantees that truly capture this subtlety between the behaviors of anytime and fixed-budget algorithms.

3.2 Benchmark: Other Fixed-budget GAI Algorithms

To go beyond the comparison with Unif, we propose and analyze additional GAI algorithms. A summary of the comparison with APGAI is shown in Table 2.

3.2.1 FROM BAI TO GAI ALGORITHMS

Since a BAI algorithm outputs the arm with the highest mean, its GAI counterpart compares the empirical mean of the returned arm to the known threshold. We study the GAI adaptations of two fixed-budget BAI algorithms: Successive Rejects (SR) (Audibert et al., 2010) and Sequential Halving (SH) (Karnin et al., 2013). SR-G and SH-G return $\hat{a}_T = \emptyset$ when $\hat{\mu}_{a_T}(T) \leq \theta$ and $\hat{a}_T = a_T$ otherwise, where a_T is the arm that would be recommended for the BAI problem, *i.e.* the arm that remains.

Theorems 24 and 25 in Appendix C give an upper bound on $P_{\nu, \text{SR-G}}^{\text{err}}(T)$ and $P_{\nu, \text{SH-G}}^{\text{err}}(T)$ at the fixed budget T . In the asymptotic regime, their rate is in $4 \log(K) \Delta_{\min}^{-2}$ when $\mathcal{A}_\theta(\mu) = \emptyset$, otherwise $\mathcal{O}(\log(K) \max\{\max_{a \in \mathcal{A}_\theta} \Delta_a^{-2}, \max_{i > I^*} i(\max_{a \in \mathcal{A}} \mu_a - \mu_{(i)})^{-2}\})$ with $I^* = |\arg \max_{a \in \mathcal{A}} \mu_a|$ and $\mu_{(i)}$ be the i^{th} largest mean in vector μ . We emphasize that the

notation $\tilde{\Delta}^{-2}$ in Table 2 “hides” the linear dependency in K of this quantity. Audibert et al. (2010, Section 6.1) shows that

$$\max_{i>I^*} i(\max_{a \in \mathcal{A}} \mu_a - \mu_{(i)})^{-2} = \tilde{\Theta} \left(I^* (\max_{a \in \mathcal{A}} \mu_a - \mu_{(I^*+1)})^{-2} + \sum_{i>I^*} (\max_{a \in \mathcal{A}} \mu_a - \mu_{(i)})^{-2} \right), \quad (5)$$

where $\tilde{\Theta}(\cdot)$ hides a $\overline{\log(K)}$ factor. Recently, Zhao et al. (2023) provides a finer analysis of SH. Using their results yields mildly improved rates. When there is one good arm with a large mean and the remaining arms have means slightly smaller than θ , those rates are better than $2H_1(\mu)$. However, APGAI has better guarantees than SR-G and SH-G when there is at least another good arm with mean slightly smaller than the largest mean as $\tilde{\Delta}^{-2}$ can become arbitrarily large. See the third column in Table 2.

Proof Sketch. When $\mathcal{A}_\theta(\mu) = \emptyset$, the error event $\{\hat{a}_T \neq \emptyset\}$ implies that the last active arm a_T satisfies $\hat{\mu}_{a_T}(T) > \theta$, even though $\mu_{a_T} \leq \theta$. As a_T is sampled linearly, this event has low probability. When $\mathcal{A}_\theta(\mu) \neq \emptyset$, the error event $\{\hat{a}_T = \emptyset\} \cup \{\hat{a}_T \in \mathcal{A}_\theta(\mu)^c\}$ implies that either (1) the last active arm a_T satisfies $\hat{\mu}_{a_T}(T) \leq \theta$ and $\mu_{a_T} > \theta$, or (2) the last active arm a_T satisfies $\mu_{a_T} < \theta$, even though $\max_{a \in \mathcal{A}} \mu_a > \theta$. The first case is unlikely with the same argument as above. The second case is unlikely since it implies that the best arm has been eliminated, *i.e.* this fixed-budget BAI algorithm has an error. Using known upper bound on the probability of error for SR (Audibert et al., 2010) and SH (Karnin et al., 2013) concludes the proof. We defer the reader to Appendices C.2 and C.3 for more details.

Doubling trick. The doubling trick allows the conversion of any fixed-budget algorithm into an anytime algorithm. It considers a sequence of algorithms that are run with increasing budgets $(T_k)_{k \geq 1}$ and recommends the answer returned by the last instance. Zhao et al. (2023) shows that Doubling SH obtains the same guarantees as SH in BAI. Theorem 24 also holds for its GAI counterpart DSH-G (resp. Theorem 25 for DSR-G) at the cost of a multiplicative factor 4 in the rate. Empirically, our experiments show that APGAI is always better than DSR-G and DSH-G (Figures 1 and 2).

Other BAI algorithms. While we consider SR and SH as examples, most fixed-budget BAI algorithms can be converted into GAI algorithms. For example, Wang et al. (2024b) recently introduced and analyzed two algorithms named CR-C and CR-A, where CR-C enjoys a better asymptotic rate than SR. However, the analysis of Wang et al. (2024b) is purely asymptotic as they leverage the Large Deviation Principle. Therefore, it departs from our objective to provide non-asymptotic upper bounds. For completeness, we still provide an asymptotic analysis of SR-G using their tools (see Appendix C.3.1).

3.2.2 PRIOR KNOWLEDGE-BASED GAI ALGORITHMS

Several fixed-budget BAI algorithms assume that the agent has access to some prior knowledge, for instance, of the unknown quantity $H_1(\nu)$, to design upper/lower confidence bounds (UCB/LCB), *e.g.* UCB-E (Audibert et al., 2010) and UGapEb (Gabillon et al., 2012). While this assumption is often not realistic, it yields better guarantees. We investigate those approaches for fixed-budget GAI. We propose an elimination-based meta-algorithm for fixed-budget GAI called PKGAI (**P**rior **K**nowledge-based **G**AI), described in Appendix D. As for BAEC, PKGAI($\star$) takes as input an index policy $\star$ which is used to define the sampling rule. At each sampling round $t < T$, PKGAI($\star$) samples the arm a_t which maximizes

Algorithm $\mathfrak{A}$	$\mathcal{A}_\theta(\mu) = \emptyset$	$\mathcal{A}_\theta(\mu) \neq \emptyset$	Dominance over APGAI when $\mathcal{A}_\theta(\mu) \neq \emptyset$
APGAI [Th 2]	$18H_1(\mu)$	$4H_1(\mu)$	$-$ (anytime)
Unif [Th 21]	$2K\Delta_{\min}^{-2}$	$4K\bar{\Delta}_{\max}^{-2}$	$\succsim$ (anytime)
DSR-G [Th 24]	$16\overline{\log}(K)\Delta_{\min}^{-2}$	$16\overline{\log}(K)\tilde{\Delta}^{-2}$	$\succsim$ (anytime)
DSH-G [Th 25]	$16\lceil\log_2(K)\rceil\Delta_{\min}^{-2}$	$32\lceil\log_2(K)\rceil\tilde{\Delta}^{-2}$	$\succsim$ (anytime)
PKGAI($\star$) [Th 27] $\dagger$	$2H_1(\mu)$	$2H_1(\mu)$	$\succ$ (fixed-budget)
PKGAI(Unif) [Th 28] $\dagger$	$2H_1(\mu)$	$2K\hat{\Delta}^{-2}$	$\succ$ (fixed-budget)

Table 2: Asymptotic error rate $C(\mu)$ of algorithm $\mathfrak{A}$ on ν , *i.e.* $\limsup_t t(\log(1/P_{\nu,\mathfrak{A}}^{\text{err}}(t)))^{-1} \leq C(\mu)$. ($\dagger$) Fixed-budget algorithm $\mathfrak{A}_{T,\nu}$ with prior knowledge of $H_1(\nu)$. $H_1(\mu)$ as in Eq. (1), $\Delta_{\min} := \min_{a \in \mathcal{A}} \Delta_a$, $\bar{\Delta}_{\max} := \max_{a \in \mathcal{A}_\theta} \Delta_a$, $I^* = |\arg \max_{a \in \mathcal{A}} \mu_a|$, $\tilde{\Delta}^{-2} := \max\{\max_{a \in \mathcal{A}_\theta} \Delta_a^{-2}, \max_{i > I^*} i(\max_{a \in \mathcal{A}} \mu_a - \mu_{(i)})^{-2}\}$ that depends linearly on K as shown by Eq. (5), $\hat{\Delta} := \max_{a \in \mathcal{A}_\theta} \Delta_a + \min_{a \notin \mathcal{A}_\theta} \Delta_a$ and $\overline{\log}(K) := \frac{1}{2} + \sum_{i=2}^K \frac{1}{i}$. The dominance of a bandit strategy is defined by the comparison of their *known upper bounds* (smaller means better): $\prec$ (dominated), $\succ$ (dominant) and $\succsim$ (Pareto equivalent: in some cases dominant, in others dominated).

the sampling index $\star$, updates the estimated upper and lower confidence bounds on the difference $\mu_{a_t} - \theta$, and eliminates any arm a such that $\mu_{a_t} - \theta < 0$ with high probability. The first main difference to BAEC lies in the definition of the UCB/LCB since they depend both on the budget T and on knowledge of $H_1(\mu)$ and $H_\theta(\mu)$. We provide upper confidence bounds on the probability of error at time T holding for any choice of indices (Theorem 27 for PKGAI($\star$)) and uniform round-robin sampling (Theorem 28 for PKGAI(Unif)). The obtained upper bounds on $P_{\nu,\text{PKGAI}}^{\text{err}}(T)$ are marginally lower than the ones obtained for APGAI, while APGAI does not require the knowledge of $H_1(\mu)$ and $H_\theta(\mu)$.

The PKGAI($\star$) meta-algorithm allows us to convert prior fixed-confidence algorithms for related problems (Kano et al., 2019; Tabata et al., 2020) into fixed-budget problems. The second main difference with fixed-confidence prior works resides in the stopping rule. In the fixed-budget setting, we should accommodate for the data-poor regime where the number of possible samples T is too small (Line 14 in Algorithm 2). If, at the end of the sampling phase, no remaining arm seems good, then we return the empty set. This additional condition penalizes fixed-confidence algorithms when the budget is too small. As such, PKGAI($\star$) represents a theoretically-supported baseline for our main algorithmic contribution APGAI, which is otherwise missing from the literature due to the lack of work on fixed-budget and anytime settings.

4 Non-asymptotic Guarantees on the Unverifiable Sample Complexity

The *unverifiable sample complexity* was defined by Katz-Samuels and Jamieson (2020) as the smallest stopping time $\tau_{U,\delta}$ after which an algorithm $\mathfrak{A}$ always outputs a correct answer with probability at least $1 - \delta$, *i.e.* $\mathbb{P}_\nu(\bigcup_{t \geq \tau_{U,\delta}} \mathcal{E}_{\mathfrak{A}}^{\text{err}}(t)) \leq \delta$. Compared to the fixed-confidence setting, it does not require to certify that the candidate answer is correct. Zhao et al.

Algorithm $\mathfrak{A}$	$\mathcal{A}_\theta(\mu) = \emptyset$	$\mathcal{A}_\theta(\mu) \neq \emptyset$	Dominance over APGAI when $\mathcal{A}_\theta(\mu) \neq \emptyset$
APGAI [Th 4]	$18H_1(\mu)$	$4H_1(\mu)$	– (anytime)
Unif [Th 22]	$K\Delta_{\min}^{-2}$	$4K\bar{\Delta}_{\max}^{-2}$	$\succsim$ (anytime)

Table 3: Asymptotic upper bound $2C(\mu)$ on the deterministic upper bound $U_\delta(\mu)$ on the unverifiable sample complexity $\tau_{U,\delta}$ of algorithm $\mathfrak{A}$ on ν , *i.e.* $\limsup_{\delta \rightarrow 0} U_\delta(\mu)/\log(1/\delta) \leq 2C(\mu)$. $H_1(\mu)$ as in Eq. (1), $\Delta_{\min} := \min_{a \in \mathcal{A}} \Delta_a$, $\bar{\Delta}_{\max} := \max_{a \in \mathcal{A}_\theta} \Delta_a$. The dominance of a bandit strategy is defined by the comparison of their *known upper bounds* (smaller means better): $\prec$ (dominated), $\succ$ (dominant) and $\succsim$ (Pareto equivalent).

(2023) notice that anytime bounds on the error can imply an unverifiable sample complexity bound. Therefore, anytime guarantees on the probability of error are more fine-grained. Theorem 4 gives a deterministic upper bound $U_\delta(\mu)$ on the unverifiable sample complexity $\tau_{U,\delta}$ of APGAI for GAI for any risk δ (see Appendix B.3 for a proof). While upper bounds on the unverifiable sample complexity $\tau_{U,\delta}$ are known in BAI (Katz-Samuels and Jamieson, 2020; Zhao et al., 2023; Jourdan et al., 2024), Theorem 4 is the first result for GAI.

Theorem 4 *Let $\delta \in (0, 1)$. The APGAI algorithm satisfies that, for any 1-sub-Gaussian distribution with mean μ such that $\Delta_{\min} > 0$, we have $\mathbb{P}_\nu(\bigcup_{t > U_\delta(\mu)} \mathcal{E}_{\mathfrak{A}}^{\text{err}}(t)) \leq \delta$ with*

$$U_\delta(\mu) = h_2(\delta, 6\alpha_{i_\mu}H_1(\mu), K + 2|\mathcal{A}_\theta|),$$

where α_{i_μ} as in Theorem 2, and $h_2(\delta, A, B) := A\bar{W}_{-1}\left(\frac{1}{3}\log\left(\frac{K\pi^2}{6\delta}\right) + B/A + \log(A)\right)$ satisfies that $h_2(\delta, A, B) =_{\delta \rightarrow 0} A\log(1/\delta)/3 + \mathcal{O}(\log\log(1/\delta))$. Moreover, $U_\delta(\mu) =_{\Delta_{\min} \rightarrow +\infty} \mathcal{O}(H_1(\mu)\log H_1(\mu))$ and $\limsup_{\delta \rightarrow 0} U_\delta(\mu)/\log(1/\delta) \leq 2\alpha_{i_\mu}H_1(\mu)$.

Intuitively, Theorem 4 is an aggregated counterpart to Theorem 2. Instead of stating that the probability of error is low at any fixed time, the probability that there exists any error after a large enough time should be low. However, Theorem 4 is not a direct corollary of Theorem 2 obtained by applying a naive union bound. From a technical perspective, both results are a by-product of the same lower-level statement: for large enough time t , the error event $\mathcal{E}_{\mathfrak{A}}^{\text{err}}(t)$ implies the concentration event does not hold, *i.e.* the empirical means deviate significantly from their means.

Comparison with uniform sampling. We compare Theorem 4 for APGAI with the deterministic upper bound on the unverifiable sample complexity of Unif for GAI given by Theorem 22 in Appendix C.1.2. Similarly as in Table 2, in the asymptotic regime described in Table 3, the upper bound for Unif has a rate in $K\Delta_{\min}^{-2}$ when $\mathcal{A}_\theta(\mu) = \emptyset$, and $4K\min_{a \in \mathcal{A}_\theta(\mu)} \Delta_a^{-2}$ otherwise. While the latter rate is better than $2H_1(\mu)$ when arms have dissimilar gaps, APGAI has better guarantees than Unif when there is no good arm.

Time-uniform probability of error. Going one step further, one might be interested in controlling the probability that there exists any error, *i.e.* $\mathbb{P}_\nu(\bigcup_{t \geq t_0} \mathcal{E}_{\mathfrak{A}}^{\text{err}}(t))$ where t_0 is an initialization time. Corollary 20 in Appendix B.4 gives an upper bound on the time-uniform probability of error for APGAI. Its proof combines Theorems 2 and 4, by using a union

bound for the time $t \leq U_\delta(\mu)$ and taking an infimum over δ . While time-uniform guarantees are appealing, they seem to be unrealistic, at least for challenging instances. Therefore, we conjecture our bound is vacuous for hard instances, *i.e.* bigger than one when $H_1(\mu)$ is large. An interesting direction for future work is to characterize the maximal hardness of an instance on which an algorithm can obtain time-uniform guarantees.

4.1 Lower Bound with Dependence on the Number of Arms

When there is a unique good arm, Theorem 4 shows that the unverifiable sample complexity of APGAI is upper bounded by a quantity scaling linearly with K , both when the risk δ is moderate or arbitrarily small. This dependency stems from the initial exploration fostered by APGAI, during which it samples suboptimal arms significantly when its collected observations are “unlucky”. We show that a linear dependence in K is actually unavoidable for moderate risk. Namely, for any risk δ and any GAI algorithm, we exhibit an instance on which the expected number of samples allocated to suboptimal arms scales at least linearly with K , see Corollary 6 below. Similar results exist in the BAI literature, *i.e.* Simchowitz et al. (2017); Al Marjani et al. (2022); Poiani et al. (2025). In particular, we adapt the techniques used in Poiani et al. (2025, Theorem 2), inspired by Al Marjani et al. (2022), and show a more general lower bound, *i.e.* Theorem 5 proven in Appendix E.3.1. It holds for any strategy whose stopping time satisfies a lower bound constraint on the TV distance between the distributions generated by interacting with instances having different answers.

Theorem 5 *Let $(\theta, \Delta, \varepsilon) \in \mathbb{R} \times (\mathbb{R}_+^*)^2$ and $(\nu^{(a)})_{a \in [K]}$ as in Theorem 3. For all $\delta \in (0, 1/4]$, let τ_δ be any stopping time satisfying that $\min_{a \in [K], b \in [K] \setminus \{a\}} \text{TV}(\mathbb{P}_{\nu^{(a)}}^{\tau_\delta}, \mathbb{P}_{\nu^{(b)}}^{\tau_\delta}) \geq 1 - 2\delta$. Then,*

$$\frac{1}{K} \sum_{a \in [K]} \mathbb{E}_{\nu^{(a)}}[\tau_\delta - N_a(\tau_\delta)] \geq \frac{K-1}{64(\Delta + \varepsilon)^2}.$$

Combining Theorem 5 with the definition of unverifiable sample complexity yields Corollary 6.

Corollary 6 *Let $(\theta, \Delta, \varepsilon) \in \mathbb{R} \times (\mathbb{R}_+^*)^2$ and $(\nu^{(a)})_{a \in [K]}$ as in Theorem 3. For any $\delta \in (0, 1/4]$ and any strategy with unverifiable sample complexity $\tau_{U,\delta}$, there exists $a \in [K]$ such that $\mathbb{E}_{\nu^{(a)}}[\tau_{U,\delta} - N_a(\tau_{U,\delta})] \geq \frac{K-1}{64(\Delta + \varepsilon)^2}$.*

Proof Since $\{\hat{a}_{\tau_{U,\delta}} \neq a\}$ is $\tau_{U,\delta}$ -measurable and satisfies that

$$\{\hat{a}_{\tau_{U,\delta}} \neq a\} \subseteq \{\exists t \geq \tau_{U,\delta}, \hat{a}_t \neq a\} \quad \text{and} \quad \{\forall t \geq \tau_{U,\delta}, \hat{a}_t = b\} \subseteq \{\hat{a}_{\tau_{U,\delta}} \neq a\},$$

we obtain that $\min_{a \in [K], b \in [K] \setminus \{a\}} \text{TV}(\mathbb{P}_{\nu^{(a)}}^{\tau_{U,\delta}}, \mathbb{P}_{\nu^{(b)}}^{\tau_{U,\delta}}) \geq 1 - 2\delta$. Using Theorem 5 concludes the proof, see Appendix E.3.2 for more details. $\blacksquare$

Corollary 6 is not valid for any instance. Among K specific instances with one good arm, any algorithm should sample the suboptimal arms at least $\frac{K-1}{64(\Delta + \varepsilon)^2}$ times on at least one of those instances. Intuitively, an initial exploration phase is necessary: the algorithm has to sample all arms before starting to recommend the unique good arm (*i.e.* the best one). As an arm a is sampled less than the others, the algorithm is slower on $\nu^{(a)}$.

5 Non-asymptotic Fixed Confidence Guarantees

In some applications, the practitioner has a strict constraint on the confidence δ associated with the candidate answer. This constraint simultaneously supersedes any limitation on the sampling budget and allows early stopping when enough evidence is collected (random since data-dependent). In the fixed-confidence setting, an identification algorithm should define a stopping rule in addition to the sampling and recommendation rules.

Stopping rule. We couple APGAI with the GLR stopping rule (Garivier and Kaufmann, 2016) for GAI (see Appendix E.2), which coincides with the Box stopping rule introduced by Kaufmann et al. (2018). At fixed confidence δ , we stop at $\tau_\delta := \min(\tau_{>,\delta}, \tau_{<,\delta})$ with

$$\tau_{>,\delta} := \inf\{t \mid \max_{a \in \mathcal{A}} W_a^+(t) \geq \sqrt{2c(t, \delta)}\} \quad \text{and} \quad \tau_{<,\delta} := \inf\{t \mid \min_{a \in \mathcal{A}} W_a^-(t) \geq \sqrt{2c(t, \delta)}\}, \quad (6)$$

where $c : \mathbb{N} \times (0, 1) \rightarrow \mathbb{R}_+$ is a threshold function. Proven in Appendix G.1, Lemma 7 gives a threshold ensuring that the GLR stopping rule Eq. (6) is δ -correct for all $\delta \in (0, 1)$, independently of the sampling rule.

Lemma 7 *Let $\bar{W}_{-1}(x) = -W_{-1}(-e^{-x})$ for all $x \geq 1$, where W_{-1} is the negative branch of the Lambert W function. It satisfies $\bar{W}_{-1}(x) \approx x + \log x$. Let $\delta \in (0, 1)$. Given any sampling rule, using the threshold*

$$2c(t, \delta) = \bar{W}_{-1}(2 \log(K/\delta) + 4 \log \log(e^4 t) + 1/2) \quad (7)$$

in the GLR stopping rule Eq. (6) yields a δ -correct algorithm for 1-sub-Gaussian distributions.

Non-asymptotic upper bound. Theorem 8 gives an upper bound on the expected sample complexity of the resulting algorithm holding for any risk δ . First, we give an implicitly defined upper bound $C_\mu(\delta)$ holding for any stopping threshold $c(t, \delta)$ ensuring δ -correctness. Second, thanks to approximations, we provide a closed-form upper bound $C'_\mu(\delta)$ on $C_\mu(\delta)$ for $c(t, \delta)$ defined in Eq. (7), which is free from large constants in the δ -independent term. The related proofs are given in Appendix F.

Theorem 8 *Let $\delta \in (0, 1)$. Combined with GLR stopping Eq. (6) using threshold Eq. (7), APGAI is δ -correct and it satisfies that, for all $\nu \in \mathcal{D}^K$ with mean μ such that $\Delta_{\min} > 0$,*

$$\mathbb{E}_\nu[\tau_\delta] \leq C_\mu(\delta) + \frac{K\pi^2}{6} + 1 \quad \text{with} \quad C_\mu(\delta) := \sup\{t \mid t \leq 2H_{i_\mu}(\mu)(\sqrt{c(t, \delta)} + \sqrt{3 \log t})^2 + D_{i_\mu}(\mu)\}$$

where $i_\mu := 1 + (\theta - 1)\mathbb{1}(\mathcal{A}_\theta(\mu) \neq \emptyset)$, $H_1(\mu)$ and $H_\theta(\mu)$ as in Eq. (1). $D_1(\mu)$ and $D_\theta(\mu)$ are defined in Lemmas 35 and 37 in Appendix F, satisfying $D_1(\mu) \approx_{\Delta_{\min} \rightarrow +\infty} D_\theta(\mu) = \mathcal{O}(H_1(\mu) \log H_1(\mu))$. In the non-asymptotic regime, the δ -independent dominating dependency is $C_\mu(\delta) = \mathcal{O}(H_1(\mu) \log H_1(\mu))$ even when there are good arms. In the asymptotic regime, we obtain $\limsup_{\delta \rightarrow 0} \mathbb{E}_\nu[\tau_\delta] / \log(1/\delta) \leq 2H_{i_\mu}(\mu)$ since $C_\mu(\delta) =_{\delta \rightarrow 0} 2H_{i_\mu}(\mu) \log(1/\delta) + \mathcal{O}(\log \log(1/\delta))$. We can also provide an explicit and closed-form upper-bound on the constant $C_\mu(\delta)$, namely $C_\mu(\delta) \leq C'_\mu(\delta)$ with

$$C'_\mu(\delta) := h(15H_{i_\mu}(\mu), 4H_{i_\mu}(\mu) (\log(K/\delta) + 15/4 - 2 \log(2H_{i_\mu}(\mu))) + D_{i_\mu}(\mu))$$

where $h(x, y) := y + x \log(x) + x \log(y/x + \log(x)) + x/2$.

Most importantly, Theorem 8 holds for any risk $\delta \in (0, 1)$ and any 1-sub-Gaussian instance ν . In the asymptotic regime where $\delta \rightarrow 0$, Theorem 8 shows that $\limsup_{\delta \rightarrow 0} \mathbb{E}_\nu[\tau_\delta] / \log(1/\delta) \leq 2H_{i_\mu}(\mu)$. Therefore, APGAI is asymptotically optimal for Gaussian distributions when $\mathcal{A}_\theta = \emptyset$. When there are good arms, our upper bound scales as $H_\theta(\mu) \log(1/\delta)$ asymptotically, which is better than the scaling in $H_1(\mu) \log(1/\delta)$ obtained for the unverifiable sample complexity. However, when $\mathcal{A}_\theta \neq \emptyset$, our upper bound is asymptotically sub-optimal compared to $2 \min_{a \in \mathcal{A}} \Delta_a^{-2}$ (see Lemma 1). This sub-optimal scaling stems from the greediness of APGAI when $\mathcal{A}_\theta \neq \emptyset$ since there is no mechanism to detect an arm that is easiest to verify, *i.e.* $\arg \max_{a \in \mathcal{A}_\theta} \Delta_a$. Empirically, we observe that APGAI can suffer from poor outliers when there are good arms with dissimilar gaps and that adding forced exploration circumvents this issue (Figure 22 and Table 14 in Appendix I.5). Intuitively, a purely asymptotic analysis of APGAI might yield the dependency $2 \max_{a \in \mathcal{A}_\theta} \Delta_a^{-2}$ which is independent from $|\mathcal{A}_\theta|$. This intuition is supported by empirical evidence (Figure 3), and we defer the reader to Appendix F.3.1 for more details.

Compared to purely asymptotic results, our non-asymptotic upper bound holds for any reasonable values of δ . It is dominated by the δ -independent term $D_{i_\mu}(\mu)$ that scales as $\mathcal{O}(H_1(\mu) \log H_1(\mu))$, even when there are good arms. Intuitively, we show that no error occur at time $T = \Omega(H_1(\mu) \log H_1(\mu))$, provided the empirical means do not deviate from their mean until time T (Lemmas 32 and 33). The dependency $H_1(\mu)$ is the same as previously obtained for the probability of error (Theorem 2) and the unverifiable sample complexity (Theorem 4). Similarly, as observed in our previous guarantees on APGAI (Theorems 2 and 4), our non-asymptotic proof techniques do not allow to capture the differences in the behavior of APGAI when interacting with instances having good arms or not. However, in the asymptotic regime, our arguments are sufficient to differentiate between both behaviors, as the non-asymptotic δ -independent term $\mathcal{O}(H_1(\mu) \log H_1(\mu))$ vanish in comparison, even though it dominates for moderate risk δ . We refer the reader to Appendix F for a detailed discussion with intuition. Our experiments reveal that the stopping time distribution of APGAI is right-skewed on instances with good arms having dissimilar gaps, suggesting that the scaling in $H_\theta(\mu)$ or $H_1(\mu)$ might not be improvable.

Comparison with uniform sampling. Combined with the same GLR stopping rule Eq. (6) using threshold Eq. (7), we compare Theorem 8 for APGAI with the non-asymptotic upper bound on the expected sample complexity of Unif for GAI given by Theorem 23 in Appendix C.1.3. In contrast to APGAI, the non-asymptotic and asymptotic dominating terms for Unif are scaling similarly. In both cases, the behavior is different when interacting with instances having good arms or not: $K \Delta_{\min}^{-2}$ when $\mathcal{A}_\theta(\mu) = \emptyset$, and $K \min_{a \in \mathcal{A}_\theta} \Delta_a^{-2}$ otherwise. Even by accounting for the right-skewness, our experiments show that APGAI outperforms Unif on average in all the considered instances.

Asymptotic dependency $H_\theta(\mu)$ instead of $H_1(\mu)$. While the δ -independent dominating term scales as $\mathcal{O}(H_1(\mu) \log H_1(\mu))$ in Lemma 37, the asymptotic dependency is $2H_\theta(\mu)$ when there are good arms. To understand this improvement over the asymptotic dependency $2H_1(\mu)$ when there are no good arms (Lemma 35), we provide some intuition behind the technical arguments used in the proof of Lemma 37. First, as in Lemma 35, the δ -dependency in Lemma 37 comes solely from a probabilistic statement involving the GLR stopping rule as in Eq. (6) whose stopping threshold as in Eq. (7) depends on the algorithmic risk parameter δ , see the definition of $D_\mu(\delta)$. Second, using Lemma 36 when $\mathcal{A}_\theta \neq \emptyset$, we know that there is

Algorithm $\mathfrak{A}$	$\mathcal{A}_\theta(\mu) = \emptyset$	$\mathcal{A}_\theta(\mu) \neq \emptyset$	Dominance over APGAI, $\mathcal{A}_\theta(\mu) \neq \emptyset$
APGAI [Th 8] [†]	$H_1(\mu)$	$H_\theta(\mu)$	– (anytime)
Unif [Th 23]	$K\Delta_{\min}^{-2}$	$K\bar{\Delta}_{\max}^{-2}$	$\succsim$ (anytime)
S-TaS § (Degenne and Koolen, 2019)	$H_1(\mu)$	$\bar{\Delta}_{\max}^{-2}$	$\succ$ (fixed-confidence)
HDoC (Kano et al., 2019)	$H_1(\mu)$	$\bar{\Delta}_{\max}^{-2}$	$\succ$ (fixed-confidence)
APT-G, LUCB-G (Kano et al., 2019)	$H_1(\mu)$	–	– (fixed-confidence)
SEE (Li and Cheung, 2025)	$\mathcal{O}(H_1(\mu))$	$\mathcal{O}(\bar{\Delta}_{\max}^{-2})$	$\succ$ (fixed-confidence)

Table 4: Asymptotic upper bound $2C(\mu)$ on the expected sample complexity of algorithm $\mathfrak{A}$ on ν , *i.e.* $\limsup_{\delta \rightarrow 0} \mathbb{E}_\nu[\tau_\delta]/\log(1/\delta) \leq 2C(\mu)$. (†) The δ -independent non-asymptotic bound scales as $\mathcal{O}(H_1(\mu) \log H_1(\mu))$ even when there are good arms. (§) Requires an ordering on the possible answers $\mathcal{A} \cup \{\emptyset\}$. $H_1(\mu)$ and $H_\theta(\mu)$ as in Eq. (1), $\bar{\Delta}_{\max} := \max_{a \in \mathcal{A}_\theta} \Delta_a$. The dominance of a bandit strategy is defined by the comparison of their *known upper bounds* (smaller means better): $\prec$ (dominated), $\succ$ (dominant) and $\succsim$ (Pareto equivalent).

no error at time T and that the “bad” arms are not sampled anymore for large enough T (yet independent of δ), provided concentration holds. When $\mathcal{A}_\theta = \emptyset$, this is in stark contrast with Lemma 34 that only states that there are no errors, yet any arms can continue to be sampled. Third, our non-asymptotic method builds on the technique used to obtain non-asymptotic upper bounds on TTUCB in Jourdan and Degenne (2023). Using the pigeonhole principle, for T large enough (yet independent of δ), there exists a good arm $a \in \mathcal{A}_\theta$ that was sampled more than $T/(\Delta_a H_\theta(\mu))$ at time T . By considering the last time where this arm was sampled, its transportation cost is simultaneously smaller than $\sqrt{2c(T, \delta)}$ (not stopped yet) and larger than $\sqrt{T/H_\theta(\mu)}$ (concentration result). Inverting this inequality concludes the proof, *i.e.* $T \lesssim 2H_\theta(\mu)c(T, \delta)$. When $\mathcal{A}_\theta = \emptyset$, based on Lemma 34, the pigeonhole principle only shows that there exists an arm $a \in [K]$ that was sampled more than $T/(\Delta_a H_1(\mu))$ at time T . Unfolding the same technical argument yields $T \lesssim 2H_1(\mu)c(T, \delta)$. This explains the difference in asymptotic behavior when there are good arms. The above discussion also glimpses why it is challenging to improve on $2H_\theta(\mu)$ with our non-asymptotic proof technique. Our proof does not control the event that all the good arms are sampled linearly, *e.g.*, in a round-robin fashion.

5.1 Lower Bound with Dependence on the Number of Arms

When there is a unique good arm, Theorem 8 shows that the expected sample complexity of APGAI is upper bounded by a quantity scaling linearly with K , when the risk δ is moderate. When the risk is arbitrarily small, the lower bound in Lemma 1 shows the independence in K of the expected sample complexity of any asymptotically optimal algorithm. Building on Theorem 5, we show that a linear dependence in K is actually unavoidable in fixed-confidence GAI (Corollary 9).

Corollary 9 *Let $(\theta, \Delta, \varepsilon) \in \mathbb{R} \times (\mathbb{R}_+^*)^2$ and $(\nu^{(a)})_{a \in [K]}$ as in Theorem 3. For any $\delta \in (0, 1/4]$ and any δ -correct strategy, there exists $a \in [K]$ such that $\mathbb{E}_{\nu^{(a)}}[\tau_\delta - N_a(\tau_\delta)] \geq \frac{K-1}{64(\Delta+\varepsilon)^2}$.*

Proof Since $\{\hat{a}_{\tau_\delta} = a\}$ is τ_δ -measurable and satisfies that $\{\hat{a}_{\tau_\delta} = a\} \subseteq \{\hat{a}_{\tau_\delta} \neq b\}$, we obtain that $\min_{a \in [K], b \in [K] \setminus \{a\}} \text{TV}(\mathbb{P}_{\nu(a)}^{\tau_\delta}, \mathbb{P}_{\nu(b)}^{\tau_\delta}) \geq 1 - 2\delta$. Using Theorem 5 concludes the proof, see Appendix E.3.3 for more details. $\blacksquare$

Corollary 9 is similar to Corollary 6, hence the same comments hold. Based on Katz-Samuels and Jamieson (2020, Theorem 5.6), Li and Cheung (2025, Theorem 5.6) gives a lower bound on $\mathbb{E}_{\nu(a)}[\tau_\delta]$ that resembles Corollary 9. Since it does not imply that suboptimal arms are sampled significantly, our lower bound is stronger.

While being slightly different probabilistic properties, both the δ -unverifiability and the δ -correctness ensures a $1 - 2\delta$ lower bound on the TV distance between the distributions generated by interacting with instances having different unique good arm. Deriving information-theoretic arguments that differentiate between both properties is an interesting direction for future research.

5.2 Benchmark: Other fixed-confidence GAI Algorithms

Table 4 summarizes the asymptotic scaling of the upper bound on the expected sample complexity of existing GAI algorithms. While most GAI algorithms have better asymptotic guarantees when $\mathcal{A}_\theta(\mu) \neq \emptyset$, APGAI is the only one of them which has anytime guarantees on the probability of error (Theorem 2). However, we emphasize that APGAI is designed for anytime GAI and is not the best algorithm for fixed-confidence GAI. Sticky Track-and-Stop (S-TaS) is asymptotically optimal for the “any low arm” problem (Degenne and Koolen, 2019), hence for GAI as well. Even though GAI is one of the few settings where S-TaS admits a computationally tractable implementation, its empirical performance heavily relies on the fixed ordering for the set of possible answers (see Table 8 in Appendix I.2). This explains the lack of non-asymptotic guarantees for S-TaS that is asymptotic by nature, while APGAI has non-asymptotic guarantees. For the “bad arm existence” problem, Kaufmann et al. (2018) prove that the empirical proportion $(N_a(t)/t)_{a \in \mathcal{A}}$ of Murphy Sampling converges almost surely towards the optimal allocation realizing the asymptotic lower bound of Lemma 1. While their result implies that $\lim_{\delta \rightarrow 0} \tau_\delta / \log(1/\delta) = T^*(\mu)$ almost surely, the authors provide no upper bound on the expected sample complexity of Murphy Sampling. Finally, we consider the AllGAI algorithms introduced by Kano et al. (2019) (HDoC, LUCB-G, and APT-G) having theoretical guarantees for some GAI instances. When $\mathcal{A}_\theta(\mu) = \emptyset$, all three algorithms have an upper bound of the form $2H_1(\mu) \log(1/\delta) + \mathcal{O}(\log \log(1/\delta))$. When $\mathcal{A}_\theta(\mu) \neq \emptyset$, only HDoC admits an upper bound on the expected number of time to return one good arm, which is of the form $2 \min_{a \in \mathcal{A}_\theta} \Delta_a^{-2} \log(1/\delta) + \mathcal{O}(\log \log(1/\delta))$.

The indices used for the elimination and recommendation in BAEC (Tabata et al., 2020) have a dependence in $\mathcal{O}(-\log(\theta_U - \theta_L))$, hence BAEC is not defined for GAI where $\theta_U = \theta_L$. While it is possible to use UCB/LCB which are agnostic to the gap $\theta_U - \theta_L > 0$, these choices have not been studied by Tabata et al. (2020). Extrapolating the theoretical guarantees of BAEC when $\theta_L \rightarrow \theta_U$, one would expect an upper bound on its expected sample complexity of the form $2H_1(\mu) \log(1/\delta) + \mathcal{O}((\log(1/\delta))^{2/3})$. In recent concurrent work, Li and Cheung (2025) propose the Sequential-Exploration-Exploitation (SEE) algorithm that proceeds in phases and alternates between exploration and exploitation subphases. Up to the constant multiplicative factor, the upper bounds on the expected sample complexity of SEE are better than the ones obtained for APGAI. Li and Cheung (2025, Theorem 5.3) shows a scaling as

$\mathcal{O}(H_1(\mu) \log(1/\delta))$ when $\mathcal{A}_\theta(\mu) = \emptyset$, and as $\mathcal{O}(\min_{a \in \mathcal{A}_\theta} \Delta_a^{-2} \log(1/\delta))$ when $\mathcal{A}_\theta(\mu) \neq \emptyset$. For fixed-confidence GAI, the above discussion exhibits adaptive algorithms that consistently outperform uniform sampling on all instances, *i.e.* the “perfect” adaptive trade-off exist.

6 Experiments

We assess the empirical performance of the APGAI in terms of empirical error, as well as empirical stopping time. Overall, APGAI performs favorably compared to other algorithms in both settings. While its empirical stopping time seems to align with Theorem 8, its (anytime) empirical error is lower than what Theorem 2 would suggest when there are good arms. This partial discrepancy between theory and practice paves the way for interesting future research. We present a fraction of our experiments and defer the reader to Appendix I for supplementary experiments.

Outcome scoring application. Our real-life motivation is outcome scoring from gene activity (transcriptomic) data (further described in Appendix I.1.1). This application focuses on the treatment of encephalopathy of prematurity in infants. The goal is to determine the optimal protocol for the administration of stem cells among $K = 18$ realistic possibilities. In collaboration with the PREMSTEM consortium, all treatments were tested on a rat model of encephalopathy of prematurity. Rat brain RNA-related measurement data were generated using high-throughput sequencing. Computed on 3 technical replicates, the mean value in $[-1, 1]$ (see Table 6 in Appendix I.1.1) corresponds to a cosine score computed between gene activity changes in treated and healthy samples. Traditional approaches use grid-search with a uniform allocation and select the best cosine score to determine the optimal protocol. Here, to model the stochasticity of the scores that would have been obtained for each protocol in a sequential approach, we applied a Bernoulli instance and considered treatment as significantly efficient when the mean score is higher than $\theta = 0.5$. In other words, observations from arm a are drawn from a Bernoulli distribution with mean $\max(\mu_a, 0)$ (which is 1/2-sub-Gaussian) using the real cosine score of this treatment protocol as μ_a .

Fixed-budget empirical error. The APGAI algorithm is compared to fixed-budget GAI algorithms: SR-G, SH-G, PKGAI and Unif. For a fair comparison, the threshold functions in PKGAI do not use prior knowledge (see Appendix I.2.2, where theoretical thresholds are used). We compare several index policies for PKGAI: Unif, APT_P, UCB, and LCB-G. At time t , the latter selects among the set $\mathcal{S}_t$ of active candidates $a_t \leftarrow \arg \max_{a \in \mathcal{S}_t} \sqrt{N_a(t)} \text{LCB}(a, t)$, where $\text{LCB}(a, t)$ is the lower confidence bound on $\mu_a - \theta$ at time t . For a budget of T up to 200, our results average over 1,000 runs, with associated confidence intervals. On our outcome scoring application, Figure 1 first shows that all uniform samplings (SH-G, SR-G, Unif, and PKGAI(Unif)) are less efficient at detecting one of the good arms contrary to the adaptive strategies. Moreover, APGAI performs as well as the elimination-based algorithms PKGAI($\star$), while allowing early stopping. These performances constitute a relevant advantage for outcome scoring and other medical applications such as clinical trials. In Appendix I.3, we confirm the good performance of APGAI in terms of fixed-budget empirical error on other instances.

Anytime empirical error. The APGAI algorithm is compared to anytime GAI algorithms: DSR-G, DSH-G (see Section 3.2.1) and Unif. Since DSH-G has poor empirical performance

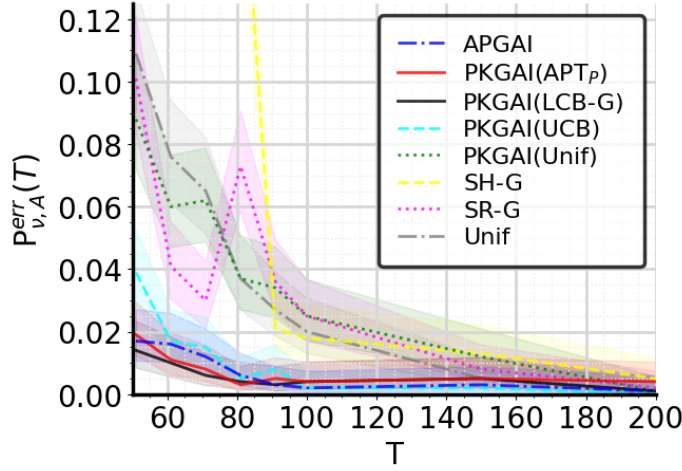
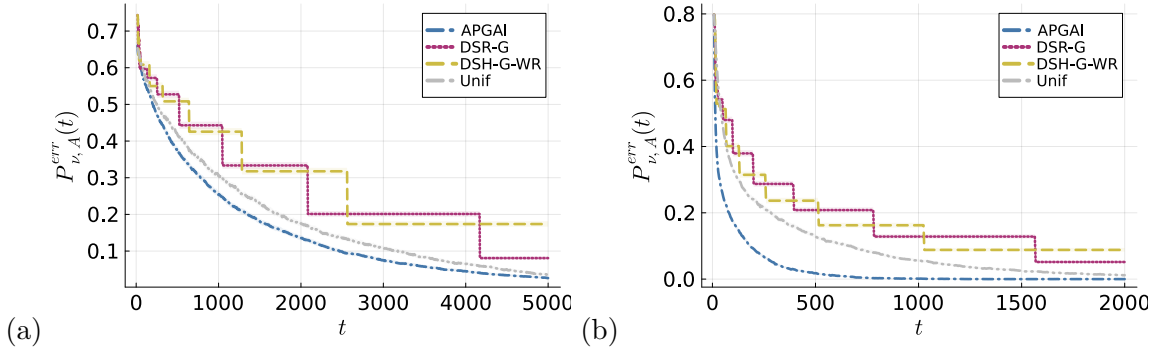


Figure 1: Fixed-budget empirical error for outcome scoring (see REALL in Table 6).


 Figure 2: Anytime empirical error on Gaussian instances (a) $\mu \in \{0.55, 0.45\}^{10}$ where $|\mathcal{A}_\theta| = 3$ for $\theta = 0.5$ and (b) $\mu = -(0.1, 0.4, 0.5, 0.6)$ for $\theta = 0$.

(see Figure 4), we consider the heuristic DSH-G-WR that relies on the whole history instead of discarding it. On two Gaussian instances ($\mathcal{A}_\theta(\mu) \neq \emptyset$ and $\mathcal{A}_\theta(\mu) = \emptyset$), Figure 2 shows that APGAI has significantly smaller empirical error compared to Unif, which is itself better than DSR-G and DSH-G-WR. Our results average over 10,000 runs, with associated confidence intervals. In Appendix I.4, we confirm the good performance of APGAI in terms of anytime empirical error on other instances, *e.g.* when $\mathcal{A}_\theta(\mu) \neq \emptyset$ (Figure 18) and when $|\mathcal{A}_\theta(\mu)|$ varies (Figure 16). Overall, APGAI appears to have better empirical performance than suggested by Theorem 2 when $\mathcal{A}_\theta(\mu) \neq \emptyset$.

Empirical stopping time. The APGAI algorithm is compared to fixed-confidence GAI algorithms using the GLR stopping rule Eq. (6) with threshold Eq. (7) and confidence $\delta = 0.01$: Murphy Sampling (MS) (Kaufmann et al., 2018), HDoC, LUCB-G (Kano et al., 2019), Track-and-Stop for GAI (TaS) (Garivier and Kaufmann, 2016) and Unif (see Appendix I.2.3). While SEE is omitted from our benchmarks as concurrent work, the experiments in (Li and Cheung, 2025) showcase that it performs on par with TaS on the considered instances. In Figure 3, we study the impact of the number of good arms by

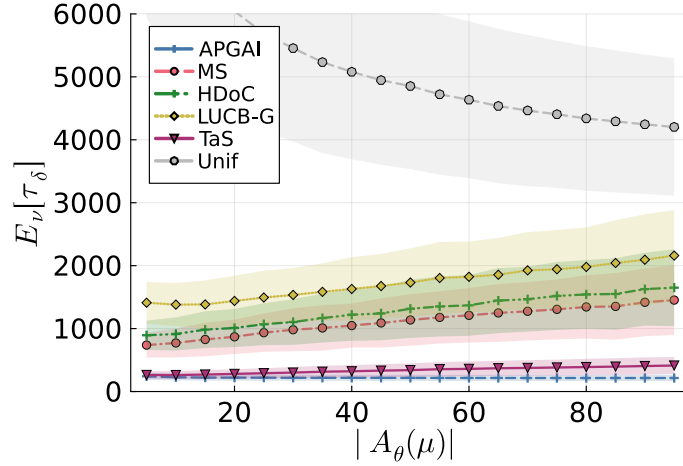


Figure 3: Empirical stopping time ($\delta = 0.01$) on Gaussian instances $\mu \in \{0.5, -0.5\}^{100}$ where $|\mathcal{A}_\theta| \in \{5k\}_{k \in [19]}$ for $\theta = 0$.

considering Gaussian instances with two groups of arms. Our results average over 1,000 runs, with associated standard deviations. Figure 3 shows that the empirical performance of APGAI is invariant to varying $|\mathcal{A}_\theta|$, and comparable to the one of TaS. In comparison, the other algorithms have worse performance and suffer from increased $|\mathcal{A}_\theta|$ since an exploration bonus exists for each good arm. In contrast, APGAI can be greedy enough to only focus its allocation to one of the good arms. Consistent with our guarantees in Theorem 8, APGAI achieves the best performance when there is no good arm. When good arms have dissimilar means (with potentially many arms), APGAI seems to suffer from poor outliers (Figures 20(b) and 22 in Appendix I.5). Given that outliers greatly impact the averaged stopping time, this behavior seems to be consistent with our suboptimal upper bound on the expected sample complexity, *i.e.* scaling as $H_1(\mu)$ for moderate δ and as $H_\theta(\mu)$ instead of $(\max_{a \in \mathcal{A}_\theta} \Delta)^{-2}$ when $|\mathcal{A}_\theta| > 1$ asymptotically (see Theorem 8). To circumvent this problem, it is enough to add forced exploration to APGAI (Table 14). While APGAI is anytime GAI algorithm, it is remarkable that it also has theoretical guarantees in fixed-confidence GAI and relatively small empirical stopping time.

7 Perspectives

We propose APGAI, the first anytime and parameter-free sampling algorithm for GAI in stochastic bandits, which is independent of a budget T or a confidence δ . In addition to showing its good empirical performance, we also provided guarantees on its probability of error at any deterministic time t (Theorem 2) and on its expected sample complexity at any confidence δ when combined with the GLR stopping time (6) (Theorem 8). As such, APGAI allows both continuation and early stopping. We reviewed and analyzed a large number of baselines for each GAI setting for comparison.

While we considered unstructured multi-armed bandits, many applications have a known structure. Investigating the GAI problem on *e.g.* linear or infinitely-armed bandits would be interesting subsequent work. In particular, working in a structured framework when

facing a possibly infinite number of arms would bring out more compelling questions about how to explore the arm space both in a tractable and meaningful way.

Acknowledgments and Disclosure of Funding

Experiments presented in this paper were carried out using the Grid’5000 testbed, supported by a scientific interest group hosted by Inria and including CNRS, RENATER and several Universities as well as other organizations (see <https://www.grid5000.fr>). This work has been partially supported by the Institut national de la santé et de la recherche médicale, the Université Sorbonne Paris Nord, the University Paris Cité, the French National Research Agency (the THIA ANR program “AI PhD@Lille” M.J.; ANR-21-RHUS-009, C.R., A.D-D; ANR-23-IAHU-0010, A.D-D), and Horizon 2020 Framework Program of the European Union (grant agreement no. 874721/PREMSTEM, A.D-D, C.R; grant agreement no. 101102016/RECeSS, C.R.). Laboratory experiments from which the outcome scoring application data were obtained were carried out by Cindy Bokobza in NeuroDiderot laboratory under the direction of Pierre Gressens for the PREMSTEM Consortium study. The comprehensive optimization study of human mesenchymal stem cell protocols including the new transcriptome dataset is the subject of unpublished research that will be released by the PREMSTEM consortium.

Appendix A. Outline

The appendices are organized as follows:

- The anytime guarantees of proof APGAI on the probability of error (Theorem 2) are proven in Appendix B. It also contains the proof of Theorem 4 (Appendix B.3) and Corollary 20 (Appendix B.4).
- Appendix C gathers error guarantees on other algorithms that are used as comparison with the anytime error guarantees of APGAI: Unif (Theorem 21), SH-G (Theorem 24) and SR-G (Theorem 25). For Unif algorithm, we also derive a deterministic upper bound on its unverifiable sample complexity for GAI (Theorem 22) and upper bound its expected sample complexity when combined with the GLR stopping (6) using threshold (7) (Theorem 23).
- We propose the meta-algorithm PKGAI in Appendix D, and analyze its error guarantees for several choices of index policy (Theorems 27 and 28).
- Appendix E gives the proof of our lower bounds: Lemma 1, Theorem 5, Corollaries 6 and 9. We link the ATP_P index and the GLR stopping rule (6) with the generalized likelihood ratio for GAI.
- The proof of Theorem 8 for APGAI when combined with the GLR stopping (6) using threshold (7) is detailed in Appendix F.
- Appendix G contains the proof of Lemma 7, and provides sequence of concentration events which are used for our proofs.
- Appendix H gathers existing and new technical results which are used for our proofs.
- In Appendix I, we provide more details on our experimental study, as well as additional experiments.

Appendix B. Analysis of APGAI: Proof of Theorem 2

The APGAI algorithm is independent of a budget T or a confidence δ which would define a stopping condition. In the following, we consider the behavior of APGAI when it is sampling *forever*. Therefore, we provide guarantees at all time T , where T can be seen as an analysis parameter. In order to upper bound the probability of the complementary of the concentration event at time T , we use an analytical parameter denoted by δ which will be inverted to obtain an upper bound on the probability of error. We emphasize that the δ used in Appendix B is not the same δ than the one to calibrate the stopping thresholds used in the GLR stopping Eq. (6). We recall that each arm is pulled once as initialization.

Proof strategy. Let $\mu \in \mathbb{R}^K$ such that $\mu_a \neq \theta$ for all $a \in \mathcal{A}$. For all $T > K$ and $\delta \in (0, 1)$, let $\tilde{\mathcal{E}}_{T,\delta}$ as in Eq. (23) for $s = 0$, i.e.

$$\tilde{\mathcal{E}}_{T,\delta} = \left\{ \forall a \in \mathcal{A}, \forall t \leq T, |\hat{\mu}_a(t) - \mu_a| < \sqrt{\frac{2\tilde{f}_1(T,\delta)}{N_a(t)}} \right\}, \quad (8)$$

with $\tilde{f}_1(T,\delta) = \frac{1}{2}\overline{W}_{-1}(2\log(1/\delta) + 2\log(2 + \log T) + 2),$

Recall that the error event $\mathcal{E}_\mu^{\text{err}}(T)$ is defined as

$$\mathcal{E}_\mu^{\text{err}}(T) := \{(\mathcal{A}_\theta \neq \emptyset \cap (\hat{a}_T = \emptyset \cup \mu_{\hat{a}_T} < \theta)) \cup (\mathcal{A}_\theta = \emptyset \cap \hat{a}_T \neq \emptyset)\}.$$

Using Lemma 41, we have $\mathbb{P}_\nu(\tilde{\mathcal{E}}_{T,\delta}^c) \leq K\delta$. Suppose that we have constructed a time $T_\mu(\delta) \geq K$ such that $\tilde{\mathcal{E}}_{T,\delta} \subseteq \mathcal{E}_\mu^{\text{err}}(T)^c$ for $T > T_\mu(\delta)$. Then, we obtain

$$\forall T > T_\mu(\delta), \quad P_{\nu,\cdot}^{\text{err}}(T) = \mathbb{P}_\nu(\mathcal{E}_\mu^{\text{err}}(T)) \leq K\delta \quad \text{hence} \quad P_{\nu,\cdot}^{\text{err}}(T) \leq K \inf\{\delta \mid T > T_\mu(\delta)\},$$

where the last inequality is obtained by taking the infimum. To prove Theorem 2, we will distinguish between instances μ such that $\mathcal{A}_\theta = \emptyset$ (Appendix B.1) and instances μ such that $\mathcal{A}_\theta \neq \emptyset$ (Appendix B.2).

Lemma 10 is the key technical tool on which our proofs rely on. It assumes the existence of a sequence of “bad” events such that, under each “bad” event, the arm selected to be pulled next was not sampled a lot yet. Then, it shows that the number of times those “bad” events occur is small.

Lemma 10 *Let $\delta \in (0, 1]$ and $T > K$. Let $(A_t(T, \delta))_{T \geq t \geq K}$ be a sequence of events and $(D_a(T, \delta))_{a \in \mathcal{A}}$ be positive thresholds satisfying that, for all $t \in (K, T] \cap \mathbb{N}$, under the event $A_t(T, \delta)$, $N_{a_{t+1}}(t) \leq D_{a_{t+1}}(T, \delta)$ and $N_{a_{t+1}}(t+1) = N_{a_{t+1}}(t) + 1$. Then, we have $\sum_{t=K+1}^T \mathbb{1}(A_t(T, \delta)) \leq \sum_{a \in \mathcal{A}} D_a(T, \delta)$.*

Proof Using the inclusion of events given by the assumption on $(A_t(T, \delta))_{T \geq t > K}$, we obtain

$$\begin{aligned} \sum_{t=K+1}^T \mathbb{1}(A_t(T, \delta)) &\leq \sum_{t=K+1}^T \mathbb{1}(N_{a_{t+1}}(t) \leq D_{a_{t+1}}(T, \delta), N_{a_{t+1}}(t+1) = N_{a_{t+1}}(t) + 1) \\ &\leq \sum_{a \in \mathcal{A}} \sum_{t=K+1}^T \mathbb{1}(N_a(t) \leq D_a(T, \delta), N_a(t+1) = N_a(t) + 1) \leq \sum_{a \in \mathcal{A}} D_a(T, \delta). \end{aligned}$$

The second inequality is obtained by union bound. The third inequality is direct since the number of times one can increment by one a quantity that is positive and bounded by $D_a(T, \delta)$ is at most $D_a(T, \delta)$. $\blacksquare$

In our proofs, we derive necessary conditions for a mistake to be made and show that having those conditions that hold is a “bad” event satisfying the condition of Lemma 10. Theorem 2 is obtained by combining Lemmas 11 and 15.

B.1 Instances where $\mathcal{A}_\theta = \emptyset$

When $\mathcal{A}_\theta = \emptyset$, we have $\mathcal{E}_\mu^{\text{err}}(T) = \{\hat{a}_T \neq \emptyset\}$. Lemma 11 gives an upper bound on the probability of error based on the recommendation of the APGAI algorithm holding for all time T .

Lemma 11 *Let $p(x) = \sqrt{x} \exp(-x)$. For all $\mu \in \mathbb{R}^K$ such that $\max_{a \in \mathcal{A}} \mu_a < \theta$, the APGAI satisfies, for all $T > K$ such that it has not stopped sampling at time T ,*

$$\mathbb{P}_\nu(\hat{a}_T \neq \emptyset) \leq K e \sqrt{2} (2 + \log T) p\left(\frac{T - K}{18 H_1(\mu)}\right).$$

Proof In order to prove Lemma 11, we show key intermediate properties of the APGAI algorithm when $\mathcal{A}_\theta = \emptyset$.

Error due to undersampled arms. At a fixed (T, δ) , the set of undersampled arms is

$$\forall t \in (K, T] \cap \mathbb{N}, \quad U_t(T, \delta) = \left\{ a \in \mathcal{A} \mid N_a(t) \leq \frac{2 \tilde{f}_1(T, \delta)}{\Delta_a^2} \right\}.$$

We show that a necessary condition for an error to occur at time t , i.e. $\hat{a}_t \neq \emptyset$, is that there exists undersampled arms, i.e. $U_t(T, \delta) \neq \emptyset$ (Lemma 12).

Lemma 12 *For all $T \in \mathbb{N}$, under the event $\tilde{\mathcal{E}}_{T, \delta}$ as in Eq. (8), for all $t \in (K, T] \cap \mathbb{N}$, we have*

$$\hat{a}_t \neq \emptyset \implies U_t(T, \delta) \neq \emptyset.$$

Proof Not recommending $\emptyset$ only happens when the largest empirical mean exceeds θ , i.e. $\max_{a \in \mathcal{A}} \hat{\mu}_a(t) > \theta$. Let $\hat{a}_t = \arg \max_{a \in \mathcal{A}} W_a^+(t)$ which satisfies $\hat{\mu}_{\hat{a}_t}(t) > \theta$. Under $\tilde{\mathcal{E}}_{T, \delta}$ as in Eq. (8), we have $\theta < \hat{\mu}_{\hat{a}_t}(t) \leq \mu_{\hat{a}_t} + \sqrt{2 \tilde{f}_1(T, \delta) / N_{\hat{a}_t}(t)}$, hence $\hat{a}_t \in U_t(T, \delta)$. ■

No remaining undersampled arms. We show that the events $\{U_t(T, \delta) \neq \emptyset\}$ satisfy the conditions of Lemma 10, hence applying it yields Lemma 13. In other words, if there are still undersampled arms at time t , then a_{t+1} has not been sampled too many times.

Lemma 13 *Let $\delta \in (0, 1)$ and $T > K$. Under event $\tilde{\mathcal{E}}_{T, \delta}$, for all $t \in (K, T] \cap \mathbb{N}$ such that $U_t(T, \delta) \neq \emptyset$, we have $N_{a_{t+1}}(t) \leq 18 \tilde{f}_1(T, \delta) / \Delta_{a_{t+1}}^2$ and $N_{a_{t+1}}(t + 1) = N_{a_{t+1}}(t) + 1$.*

Proof We will be interested in three distinct cases since

$$\begin{aligned} \{U_t(T, \delta) \neq \emptyset\} &= \underbrace{\{U_t(T, \delta) \neq \emptyset, \max_{a \in \mathcal{A}} \hat{\mu}_a(t) > \theta\}}_{\text{Case 1}} \cup \underbrace{\{U_t(T, \delta) \neq \emptyset, \max_{a \in \mathcal{A}} \hat{\mu}_a(t) < \theta\}}_{\text{Case 2}} \\ &\quad \cup \underbrace{\{U_t(T, \delta) \neq \emptyset, \max_{a \in \mathcal{A}} \hat{\mu}_a(t) = \theta\}}_{\text{Case 3}} \end{aligned}$$

Case 1. Let $t \in (K, T] \cap \mathbb{N}$ such that $U_t(T, \delta) \neq \emptyset$ and $\max_{a \in \mathcal{A}} \hat{\mu}_a(t) > \theta$. Let $c = \arg \max_{a \in \mathcal{A}} \hat{\mu}_a(t)$. Since $W_c^+(t) > 0$ and $a_{t+1} \in \arg \max_{a \in \mathcal{A}} W_a^+(t)$, we obtain $\hat{\mu}_{a_{t+1}}(t) > \theta$.

Then, under $\tilde{\mathcal{E}}_{T,\delta}$ as in Eq. (8), we have

$$\begin{aligned}\sqrt{N_{a_{t+1}}(t)}(\hat{\mu}_{a_{t+1}}(t) - \theta)_+ &= \sqrt{N_{a_{t+1}}(t)}(\hat{\mu}_{a_{t+1}}(t) - \theta) \\ &\leq \sqrt{N_{a_{t+1}}(t)}(\mu_{a_{t+1}} - \theta) + \sqrt{2\tilde{f}_1(T, \delta)}.\end{aligned}$$

Using that $W_{a_{t+1}}^+(t) > 0$, we obtain $N_{a_{t+1}}(t) \leq \frac{2\tilde{f}_1(T, \delta)}{\Delta_{a_{t+1}}^2}$ and $N_{a_{t+1}}(t+1) = N_{a_{t+1}}(t) + 1$.

Case 2. Let $t \in (K, T] \cap \mathbb{N}$ such that $U_t(T, \delta) \neq \emptyset$ and $\max_{a \in \mathcal{A}} \hat{\mu}_a(t) < \theta$. Let $a_{t+1} \in \arg \min_{a \in \mathcal{A}} W_a^-(t)$ and $a \in U_t(T, \delta)$. Then, under $\tilde{\mathcal{E}}_{T,\delta}$ as in Eq. (8), we have

$$\begin{aligned}\sqrt{N_{a_{t+1}}(t)}(\theta - \mu_{a_{t+1}}) - \sqrt{2\tilde{f}_1(T, \delta)} &\leq \sqrt{N_{a_{t+1}}(t)}(\theta - \hat{\mu}_{a_{t+1}}(t)) = \sqrt{N_{a_{t+1}}(t)}(\theta - \hat{\mu}_{a_{t+1}}(t))_+ \\ \sqrt{N_a(t)}(\theta - \hat{\mu}_a(t))_+ &= \sqrt{N_a(t)}(\theta - \hat{\mu}_a(t)) \leq \sqrt{N_a(t)}(\theta - \mu_a) + \sqrt{2\tilde{f}_1(T, \delta)} \leq 2\sqrt{2\tilde{f}_1(T, \delta)}\end{aligned}$$

Using that $W_{a_{t+1}}^-(t) \leq W_a^-(t)$, we obtain $N_{a_{t+1}}(t) \leq \frac{18\tilde{f}_1(T, \delta)}{\Delta_{a_{t+1}}^2}$ and $N_{a_{t+1}}(t+1) = N_{a_{t+1}}(t) + 1$.

Case 3. Let $t \in (K, T] \cap \mathbb{N}$ such that $U_t(T, \delta) \neq \emptyset$ and $\max_{a \in \mathcal{A}} \hat{\mu}_a(t) = \theta$. Then, $\arg \min_{a \in \mathcal{A}} W_a^-(t) = \{a \in \mathcal{A} \mid \hat{\mu}_a(t) = \theta\}$. Therefore, we have $\hat{\mu}_{a_{t+1}}(t) = \theta$ hence $\theta = \hat{\mu}_{a_{t+1}}(t) \leq \mu_{a_{t+1}} + \sqrt{\frac{2\tilde{f}_1(T, \delta)}{N_{a_{t+1}}(t)}}$. Therefore, we obtain $N_{a_{t+1}}(t) \leq \frac{2\tilde{f}_1(T, \delta)}{\Delta_{a_{t+1}}^2}$ and $N_{a_{t+1}}(t+1) = N_{a_{t+1}}(t) + 1$.

Summary. Combing the three above cases yields the result. ■

Lemma 14 provides a time after which all arms are sampled enough, hence no error will be made.

Lemma 14 *Let us define $T_\mu(\delta) = \sup \left\{ T \mid T \leq 18H_1(\mu)\tilde{f}_1(T, \delta) + K \right\}$. For all $T > T_\mu(\delta)$, under the event $\tilde{\mathcal{E}}_{T,\delta}$ as in Eq. (8), we have $U_T(T, \delta) = \emptyset$.*

Proof Combining Lemmas 13 and 10, we obtain $\sum_{t=K+1}^T \mathbb{1}(U_t(T, \delta) \neq \emptyset) \leq 18H_1(\mu)\tilde{f}_1(T, \delta)$. For all $a \in \mathcal{A}$, let us define $t_a(T, \delta) = \max\{t \in (K, T] \cap \mathbb{N} \mid a \in U_t(T, \delta)\}$. By definition, we have $a \in U_t(T, \delta)$ for all $t \in (K, t_a(T, \delta)]$ and $a \notin U_t(T, \delta)$ for all $t \in (t_a(T, \delta), T]$. Therefore, for all $t \in (K, \max_{a \in \mathcal{A}} t_a(T, \delta)]$, we have $U_t(T, \delta) \neq \emptyset$ and $U_t(T, \delta) = \emptyset$ for all $t > \max_{a \in \mathcal{A}} t_a(T, \delta)$, hence

$$\max_{a \in \mathcal{A}} (t_a(T, \delta) - K) = \sum_{t=K+1}^T \mathbb{1}(U_t(T, \delta) \neq \emptyset) \leq 18H_1(\mu)\tilde{f}_1(T, \delta).$$

Let $T_\mu(\delta)$ defined as in the statement of Lemma 14 and $T > T_\mu(\delta)$. Then, we have

$$T - K > 18H_1(\mu)\tilde{f}_1(T, \delta) \geq \max_{a \in \mathcal{A}} (t_a(T, \delta) - K),$$

hence $T > \max_{a \in \mathcal{A}} t_a(T, \delta)$. This concludes the proof that $U_T(T, \delta) = \emptyset$. ■

Conclusion Let $T_\mu(\delta)$ as in Lemma 14. Combining Lemmas 14, 12 and 41, we obtain

$$\forall T > T_\mu(\delta), \quad \{\hat{a}_T \neq \emptyset\} \cap \tilde{\mathcal{E}}_{T,\delta} = \emptyset \quad \text{and} \quad \mathbb{P}_\nu(\tilde{\mathcal{E}}_{T,\delta}^c) \leq K\delta \quad \text{hence}$$

$$\mathbb{P}_\nu(\hat{a}_T \neq \emptyset) \leq K \inf\{\delta \mid T > T_\mu(\delta)\} \leq Ke\sqrt{2}(2 + \log T) \sqrt{\frac{T-K}{18H_1(\mu)}} \exp\left(-\frac{T-K}{18H_1(\mu)}\right),$$

where the last inequality uses Lemma 45. This concludes the proof of Lemma 11. $\blacksquare$

B.2 Instances where $\mathcal{A}_\theta \neq \emptyset$

When $\mathcal{A}_\theta = \emptyset$, we have $\mathcal{E}_\mu^{\text{err}}(T) = \{\hat{a}_T = \emptyset\} \cup \{\hat{a}_T \in \mathcal{A}_\theta^c\}$. Lemma 15 gives an upper bound on the probability of error based on the recommendation of APGAI holding for all time T .

Lemma 15 *Let $p(x) = \sqrt{x} \exp(-x)$. For all $\mu \in \mathbb{R}^K$ such that $\mathcal{A}_\theta \neq \emptyset$ and $\mu_a \neq \theta$ for all $a \in \mathcal{A}$, the APGAI satisfies, for all $T > K$ such that it has not stopped sampling at time T ,*

$$\mathbb{P}\left(\{\hat{a}_T = \emptyset\} \cup \{\hat{a}_T \in \mathcal{A}_\theta^c\}\right) \leq Ke\sqrt{2}(2 + \log T)p\left(\frac{T-K-2|\mathcal{A}_\theta|}{4H_1(\mu)}\right).$$

Proof In order to prove Lemma 15, we show key intermediate properties of the APGAI algorithm when $\mathcal{A}_\theta \neq \emptyset$.

Error due to undersampled arms. At a fixed (T, δ) , the set of under-sampled arms is

$$\forall t \in (K, T] \cap \mathbb{N}, \quad U_t(T, \delta) = \left\{a \in \mathcal{A} \mid N_a(t) \leq \left(\sqrt{\frac{2\tilde{f}_1(T, \delta)}{\Delta_a^2}} + 1\right)^2\right\}.$$

Lemma 16 shows that a necessary condition to recommend $\emptyset$ at time t is that all the good arms are undersampled arms, *i.e.* $\mathcal{A}_\theta \subseteq U_t(T, \delta)$. It also shows that a necessary condition to recommend $\hat{a}_t \in \mathcal{A}_\theta^c$ at time t is that this arm is undersampled and will be sampled next, *i.e.* $\hat{a}_t = a_{t+1}$ and $a_{t+1} \in \mathcal{A}_\theta^c \cap U_t(T, \delta)$.

Lemma 16 *For all $T \in \mathbb{N}$, under the event $\tilde{\mathcal{E}}_{T,\delta}$ as in Eq. (8), for all $t \in (K, T] \cap \mathbb{N}$, we have*

$$\begin{aligned} \hat{a}_t = \emptyset &\implies \mathcal{A}_\theta \subseteq U_t(T, \delta), \\ \hat{a}_t \in \mathcal{A}_\theta^c &\implies \hat{a}_t = a_{t+1} \text{ and } a_{t+1} \in \mathcal{A}_\theta^c \cap U_t(T, \delta). \end{aligned}$$

Proof Case 1. Suppose that $\hat{a}_t = \emptyset$, hence $\max \hat{\mu}_a(t) \leq \theta$. Then, for all $a \in \mathcal{A}_\theta$, we have $\theta \geq \hat{\mu}_a(t) \geq \mu_a - \sqrt{2\tilde{f}_1(T, \delta)/N_a(t)}$, hence $\mathcal{A}_\theta \subseteq U_t(T, \delta)$.

Case 2. Suppose that $\hat{a}_t \notin \mathcal{A}_\theta$, hence $\max \hat{\mu}_a(t) > \theta$. Since $a_{t+1} \in \arg \max_{a \in \mathcal{A}} W_a^+(t)$ and $\hat{a}_t = a_{t+1}$, we have $\hat{\mu}_{\hat{a}_t}(t) > \theta$. Then, we have $\theta < \hat{\mu}_{a_{t+1}}(t) \leq \mu_{a_{t+1}} + \sqrt{2\tilde{f}_1(T, \delta)/N_{a_{t+1}}(t)}$, hence $a_{t+1} \in \mathcal{A}_\theta^c \cap U_t(T, \delta)$. $\blacksquare$

One good arm not undersampled. Lemma 17 shows that the events $\{\mathcal{A}_\theta \subseteq U_t(T, \delta)\}$ are satisfying the conditions of Lemma 10. In other words, having all the good arms undersampled implies that the next arm we will pull was not sampled a lot.

Lemma 17 *Let $\delta \in (0, 1)$ and $T > K$. Under event $\tilde{\mathcal{E}}_{T,\delta}$, for all $t \in (K, T] \cap \mathbb{N}$ such that $\mathcal{A}_\theta \subseteq U_t(T, \delta)$, we have $N_{a_{t+1}}(t) \leq D_{a_{t+1}}(T, \delta)$ and $N_{a_{t+1}}(t+1) = N_{a_{t+1}}(t) + 1$, where $D_a(T, \delta) = (\Delta_a^{-1} \sqrt{2\tilde{f}_1(T, \delta)} + 1)^2$ for all $a \in \mathcal{A}_\theta$ and $D_a(T, \delta) = 2\tilde{f}_1(T, \delta)\Delta_a^{-2}$ for all $a \notin \mathcal{A}_\theta$.*

Proof Let $t \in (K, T] \cap \mathbb{N}$ such that $\mathcal{A}_\theta \subseteq U_t(T, \delta)$. When $a_{t+1} \in \mathcal{A}_\theta$, we have directly that $N_{a_{t+1}}(t) \leq (\sqrt{2\tilde{f}_1(T, \delta)/\Delta_{a_{t+1}}^2} + 1)^2$ and $N_{a_{t+1}}(t+1) = N_{a_{t+1}}(t) + 1$. In the following, we consider $a_{t+1} \notin \mathcal{A}_\theta$. We will be interested in three cases since

$$\begin{aligned} \{\mathcal{A}_\theta \subseteq U_t(T, \delta), a_{t+1} \notin \mathcal{A}_\theta\} &= \underbrace{\{\mathcal{A}_\theta \subseteq U_t(T, \delta), a_{t+1} \notin \mathcal{A}_\theta, \max_{a \in \mathcal{A}} \hat{\mu}_a(t) > \theta\}}_{\text{Case 1}} \\ &\cup \underbrace{\{\mathcal{A}_\theta \subseteq U_t(T, \delta), a_{t+1} \notin \mathcal{A}_\theta, \max_{a \in \mathcal{A}} \hat{\mu}_a(t) < \theta\}}_{\text{Case 2}} \cup \underbrace{\{\mathcal{A}_\theta \subseteq U_t(T, \delta), a_{t+1} \notin \mathcal{A}_\theta, \max_{a \in \mathcal{A}} \hat{\mu}_a(t) = \theta\}}_{\text{Case 3}} \end{aligned}$$

Case 1. Let $t \in (K, T] \cap \mathbb{N}$ such that $\mathcal{A}_\theta \subseteq U_t(T, \delta)$, $a_{t+1} \notin \mathcal{A}_\theta$ and $\max_{a \in \mathcal{A}} \hat{\mu}_a(t) > \theta$. Let $c = \arg \max_{a \in \mathcal{A}} \hat{\mu}_a(t)$. Since $W_c^+(t) > 0$ and $a_{t+1} \in \arg \max_{a \in \mathcal{A}} W_a^+(t)$, we have $\hat{\mu}_{a_{t+1}}(t) > \theta$. Since $a_{t+1} \notin \mathcal{A}_\theta$, under $\tilde{\mathcal{E}}_{T,\delta}$ as in Eq. (8), we have

$$\sqrt{N_{a_{t+1}}(t)}(\hat{\mu}_{a_{t+1}}(t) - \theta)_+ = \sqrt{N_{a_{t+1}}(t)}(\hat{\mu}_{a_{t+1}}(t) - \theta) \leq \sqrt{N_{a_{t+1}}(t)}(\mu_{a_{t+1}} - \theta) + \sqrt{2\tilde{f}_1(T, \delta)}$$

Using that $W_{a_{t+1}}^+(t) > 0$, we obtain $N_{a_{t+1}}(t) \leq 2\tilde{f}_1(T, \delta)/\Delta_{a_{t+1}}^2$ and $N_{a_{t+1}}(t+1) = N_{a_{t+1}}(t) + 1$.

Case 2. Let $t \in (K, T] \cap \mathbb{N}$ such that $\mathcal{A}_\theta \subseteq U_t(T, \delta)$, $a_{t+1} \notin \mathcal{A}_\theta$ and $\max_{a \in \mathcal{A}} \hat{\mu}_a(t) < \theta$. Let $a_{t+1} \in \arg \min_{a \in \mathcal{A}} W_a^-(t)$. Since $a_{t+1} \notin \mathcal{A}_\theta$, under $\tilde{\mathcal{E}}_{T,\delta}$ as in Eq. (8), for all $a \in \mathcal{A}_\theta$, we have

$$\begin{aligned} \sqrt{N_{a_{t+1}}(t)}(\theta - \mu_{a_{t+1}}) - \sqrt{2\tilde{f}_1(T, \delta)} &\leq \sqrt{N_{a_{t+1}}(t)}(\theta - \hat{\mu}_{a_{t+1}}(t)) = \sqrt{N_{a_{t+1}}(t)}(\theta - \hat{\mu}_{a_{t+1}}(t))_+ \\ \sqrt{N_a(t)}(\theta - \hat{\mu}_a(t))_+ &= \sqrt{N_a(t)}(\theta - \hat{\mu}_a(t)) \leq \sqrt{N_a(t)}(\theta - \mu_a) + \sqrt{2\tilde{f}_1(T, \delta)} \leq \sqrt{2\tilde{f}_1(T, \delta)}. \end{aligned}$$

Combining both inequality by using that $W_{a_{t+1}}^-(t) \leq W_a^-(t)$ yields $\sqrt{N_{a_{t+1}}(t)}(\theta - \mu_{a_{t+1}}) \leq 2\sqrt{2\tilde{f}_1(T, \delta)}$, hence $N_{a_{t+1}}(t) \leq 8\tilde{f}_1(T, \delta)/\Delta_{a_{t+1}}^2$ and $N_{a_{t+1}}(t+1) = N_{a_{t+1}}(t) + 1$.

Case 3. Let $t \in (K, T] \cap \mathbb{N}$ such that $\mathcal{A}_\theta \subseteq U_t(T, \delta)$, $a_{t+1} \notin \mathcal{A}_\theta$ and $\max_{a \in \mathcal{A}} \hat{\mu}_a(t) = \theta$. Then, $a_{t+1} \in \arg \min_{a \in \mathcal{A}} W_a^-(t) = \{a \in \mathcal{A} \mid \hat{\mu}_a(t) = \theta\}$. Therefore, we have $\theta = \hat{\mu}_{a_{t+1}}(t) \leq \mu_{a_{t+1}} + \sqrt{2\tilde{f}_1(T, \delta)/N_{a_{t+1}}(t)}$. Since $a_{t+1} \notin \mathcal{A}_\theta$, we obtain $N_{a_{t+1}}(t) \leq 2\tilde{f}_1(T, \delta)/\Delta_{a_{t+1}}^2$ and $N_{a_{t+1}}(t+1) = N_{a_{t+1}}(t) + 1$.

Summary. Combing the three above cases yields the result. ■

Lemma 18 shows that having a good arm that is sampled enough, *i.e.* $\mathcal{A}_\theta \cap U_t(T, \delta)^c \neq \emptyset$, is a sufficient condition to recommend a good arm, *i.e.* $\hat{a}_t \in \mathcal{A}_\theta$.

Lemma 18 *Let $\delta \in (0, 1)$ and $T > K$. Under event $\tilde{\mathcal{E}}_{T,\delta}$, for all $t \in (K, T] \cap \mathbb{N}$ such that $\mathcal{A}_\theta \cap U_t(T, \delta)^c \neq \emptyset$, we have $\hat{a}_t \in \mathcal{A}_\theta$.*

Proof Let $t \in (K, T] \cap \mathbb{N}$ such that $\mathcal{A}_\theta \cap U_t(T, \delta)^\complement \neq \emptyset$. Let $a \in \mathcal{A}_\theta \cap U_t(T, \delta)^\complement$, hence

$$N_a(t) > \left(\sqrt{\frac{2\tilde{f}_1(T, \delta)}{(\mu_a - \theta)^2}} + 1 \right)^2 > \frac{2\tilde{f}_1(T, \delta)}{(\mu_a - \theta)^2}. \quad (9)$$

Therefore, under $\tilde{\mathcal{E}}_{T, \delta}$ as in Eq. (8), we have $\max_{b \in \mathcal{A}} \hat{\mu}_b(t) \geq \hat{\mu}_a(t) \geq \mu_a - \sqrt{2\tilde{f}_1(T, \delta)/N_a(t)} > \theta$, hence $\hat{a}_t = a_{t+1} \in \arg \max_{a \in \mathcal{A}} W_a^+(t)$.

Suppose towards contradiction that $\mathcal{A}_\theta^\complement \cap \arg \max_{a \in \mathcal{A}} W_a^+(t) \neq \emptyset$. Let $a \in \mathcal{A}_\theta^\complement \cap \arg \max_{a \in \mathcal{A}} W_a^+(t) \neq \emptyset$. It is direct to see that $\hat{\mu}_a(t) > \theta$, otherwise there is a contradiction. Then, using that $a \in \mathcal{A}_\theta^\complement$ (i.e. $\mu_a \leq \theta$), we have for all $b \in \mathcal{A}_\theta \cap U_t(T, \delta)^\complement$

$$\begin{aligned} \sqrt{2\tilde{f}_1(T, \delta)} &\geq \sqrt{N_a(t)}(\mu_a - \theta) + \sqrt{2\tilde{f}_1(T, \delta)} \geq \sqrt{N_a(t)}(\hat{\mu}_a(t) - \theta) = \sqrt{N_a(t)}(\hat{\mu}_a(t) - \theta)_+, \\ \sqrt{N_b(t)}(\hat{\mu}_b(t) - \theta)_+ &= \sqrt{N_b(t)}(\hat{\mu}_b(t) - \theta) \geq \sqrt{N_b(t)}(\mu_b - \theta) - \sqrt{\frac{2\tilde{f}_1(T, \delta)}{N_b(t)}} \\ &> \left(\sqrt{N_b(t)} - 1 \right) (\mu_b - \theta) > \sqrt{2\tilde{f}_1(T, \delta)}, \end{aligned}$$

where the two last inequalities are obtained by using Eq. (9) first the smaller thresholds, then the one in-between. Since $a \neq b$ and $W_a^+(t) \geq W_b^+(t)$, combining the above yields $\sqrt{2\tilde{f}_1(T, \delta)} > \sqrt{2\tilde{f}_1(T, \delta)}$ which is a contradiction. Therefore, we have proven that

$$\mathcal{A}_\theta \cap U_t(T, \delta)^\complement \neq \emptyset \implies \hat{a}_t \in \arg \max_{a \in \mathcal{A}} W_a^+(t) \wedge \mathcal{A}_\theta^\complement \cap \arg \max_{a \in \mathcal{A}} W_a^+(t) = \emptyset$$

which implies that $\hat{a}_t \in \mathcal{A}_\theta$. ■

Lemma 19 provides a time after which there exists a good arms which is sampled enough, hence no error will be made.

Lemma 19 Let us define $S_\mu(\delta) = \sup \left\{ T \mid T \leq 4H_1(\mu)\tilde{f}_1(T, \delta) + K + 2|\mathcal{A}_\theta| \right\}$. For all $T > S_\mu(\delta)$, under the event $\tilde{\mathcal{E}}_{T, \delta}$ as in Eq. (8), we have $\mathcal{A}_\theta \cap U_T(T, \delta)^\complement \neq \emptyset$ and $\hat{a}_T \in \mathcal{A}_\theta$.

Proof Let $(D_a(T, \delta))_{a \in \mathcal{A}}$ as in Lemma 17. Combining Lemmas 17 and 10, we obtain $\sum_{t=K+1}^T \mathbb{1}(\mathcal{A}_\theta \subseteq U_t(T, \delta)) \leq \sum_{a \in \mathcal{A}} D_a(T, \delta)$. For all $a \in \mathcal{A}_\theta$, let us define $t_a(T, \delta) = \max\{t \in (K, T] \cap \mathbb{N} \mid a \in U_t(T, \delta)\}$. By definition, we have $a \in U_t(T, \delta)$ for all $t \in (K, t_a(T, \delta)]$ and $a \notin U_t(T, \delta)$ for all $t \in (t_a(T, \delta), T]$. Therefore, for all $t \in (K, \min_{a \in \mathcal{A}_\theta} t_a(T, \delta)]$, we have $\mathcal{A}_\theta \subseteq U_t(T, \delta)$ and $\mathcal{A}_\theta \cap U_t(T, \delta)^\complement \neq \emptyset$ for all $t > \max_{a \in \mathcal{A}_\theta} t_a(T, \delta)$, hence

$$\min_{a \in \mathcal{A}_\theta} (t_a(T, \delta) - K) = \sum_{t=K+1}^T \mathbb{1}(\mathcal{A}_\theta \subseteq U_t(T, \delta)) \leq \sum_{a \in \mathcal{A}} D_a(T, \delta).$$

Let $S_\mu(\delta)$ defined as in the statement of Lemma 19 and $T > S_\mu(\delta)$. Using that $(a+1)^2 \leq 2a^2 + 2$, we have $S_\mu(\delta) \geq \sup\{T \mid T \leq \sum_{a \in \mathcal{A}} D_a(T, \delta) + K\}$. Then, we have

$$T - K > \sum_{a \in \mathcal{A}_\theta} D_a(T, \delta) \geq \min_{a \in \mathcal{A}_\theta} (t_a(T, \delta) - K),$$

hence $T > \min_{a \in \mathcal{A}_\theta} t_a(T, \delta)$. Therefore, we have $\mathcal{A}_\theta \cap U_T(T, \delta)^\complement \neq \emptyset$. Using Lemma 18, we obtain that $\hat{a}_T \in \mathcal{A}_\theta$. This concludes the proof. $\blacksquare$

Conclusion. Let $S_\mu(\delta)$ as in Lemma 19. Combining Lemmas 19, 18 and 41, we obtain

$$\begin{aligned} \forall T > S_\mu(\delta), \quad & \left(\{\hat{a}_T = \emptyset\} \cup \{\hat{a}_T \in \mathcal{A}_\theta^\complement\} \right) \cap \tilde{\mathcal{E}}_{T,\delta} = \emptyset \quad \text{and} \quad \mathbb{P}_\nu(\tilde{\mathcal{E}}_{T,\delta}^\complement) \leq K\delta \quad \text{hence} \\ & \mathbb{P}_\nu(\{\hat{a}_T = \emptyset\} \cup \{\hat{a}_T \in \mathcal{A}_\theta^\complement\}) \leq K \inf\{\delta \mid T > S_\mu(\delta)\} \\ & \leq Ke\sqrt{2(2 + \log T)} \sqrt{\frac{T - K - 2|\mathcal{A}_\theta|}{4H_1(\mu)}} \exp\left(-\frac{T - K - 2|\mathcal{A}_\theta|}{4H_1(\mu)}\right), \end{aligned}$$

where the last inequality uses Lemma 45. This concludes the proof. $\blacksquare$

B.3 Unverifiable Sample Complexity: Proof of Theorem 4

In Appendix B.1 and B.2, we consider the concentration event $\tilde{\mathcal{E}}_{T,\delta}$ that involved tighter concentration results with thresholds $\tilde{f}_1(T, \delta)$. Let $T > K$ and $\delta \in (0, 1)$. It is direct to see that the same argument holds for the concentration events $\mathcal{E}_{T,\delta}$ as in Eq. (21) for $s = 2$, i.e.,

$$\mathcal{E}_{T,\delta} = \left\{ \forall a \in \mathcal{A}, \forall t \leq T, |\hat{\mu}_a(t) - \mu_a| < \sqrt{\frac{2f_1(T, \delta)}{N_a(t)}} \right\},$$

where $f_1(T, \delta) = \log(1/\delta) + 3 \log T + \log(K\pi^2/6)$. Let $U_\delta(\mu) > K$ to be specified below. Using Lemma 40, we obtain that

$$\mathbb{P}_\nu \left(\bigcup_{T > U_\delta(\mu)} \mathcal{E}_{T,\delta}^\complement \right) \leq \sum_{T > U_\delta(\mu)} \mathbb{P}_\nu(\mathcal{E}_{T,\delta}^\complement) \leq \frac{\delta}{\zeta(2)} \sum_{T > U_\delta(\mu)} \frac{1}{T^2} \leq \delta.$$

Suppose that $U_\delta(\mu)$ is chosen such that $\mathcal{E}_{\mathfrak{A}}^{\text{err}}(T) \cap \mathcal{E}_{T,\delta} = \emptyset$ for all $T > U_\delta(\mu)$. Then, we have

$$\begin{aligned} \mathbb{P}_\nu \left(\bigcup_{T > U_\delta(\mu)} \mathcal{E}_{\mathfrak{A}}^{\text{err}}(T) \right) & \leq \mathbb{P}_\nu \left(\left(\bigcup_{T > U_\delta(\mu)} \mathcal{E}_{\mathfrak{A}}^{\text{err}}(T) \right) \cap \left(\bigcap_{T > U_\delta(\mu)} \mathcal{E}_{T,\delta} \right) \right) + \mathbb{P}_\nu \left(\bigcup_{T > U_\delta(\mu)} \mathcal{E}_{T,\delta}^\complement \right) \\ & \leq \mathbb{P}_\nu \left(\bigcup_{T > U_\delta(\mu)} (\mathcal{E}_{\mathfrak{A}}^{\text{err}}(T) \cap \mathcal{E}_{T,\delta}) \right) + \delta \leq \delta. \end{aligned}$$

Therefore, we can conclude the proof by exhibiting $U_\delta(\mu)$ satisfying the above property.

Case 1: when $\mathcal{A}_\theta = \emptyset$. Let $T_\mu(\delta)$ defined similarly as in Lemma 14, i.e.

$$T_\mu(\delta) := \sup \{T \mid T \leq 18H_1(\mu)f_1(T, \delta) + K\}.$$

To prove Theorem 2 when $\mathcal{A}_\theta = \emptyset$, we obtain as an intermediary result that: for all $T > T_\mu(\delta)$, $\{\hat{a}_T \neq \emptyset\} \subseteq \mathcal{E}_{T,\delta}^c$. Using a proof similar to Lemma 46, applying Lemma 44 yields that

$$\begin{aligned} T &> T_\mu(\delta) \\ \iff T &> 54H_1(\mu) \log T + 18H_1(\mu) \log \left(\frac{K\pi^2}{6\delta} \right) + K \\ \iff \frac{T}{54H_1(\mu)} - \log \left(\frac{T}{54H_1(\mu)} \right) &> \frac{1}{3} \log \left(\frac{K\pi^2}{6\delta} \right) + \frac{K}{54H_1(\mu)} + \log(54H_1(\mu)) \\ \iff T &> 54H_1(\mu) \overline{W}_{-1} \left(\frac{1}{3} \log \left(\frac{K\pi^2}{6\delta} \right) + \frac{K}{54H_1(\mu)} + \log(54H_1(\mu)) \right), \end{aligned}$$

Let us define $U_\delta(\mu) := h_2(\delta, 54H_1(\mu), K)$, where

$$h_2(\delta, A, B) := A \overline{W}_{-1} \left(\frac{1}{3} \log \left(\frac{K\pi^2}{6\delta} \right) + \frac{B}{A} + \log A \right)$$

satisfies that $h_2(\delta, A, B) =_{\delta \rightarrow 0} \frac{A}{3} \log(1/\delta) + \mathcal{O}(\log \log(1/\delta))$. Hence, we have shown that $\{\hat{a}_T \neq \emptyset\} \cap \mathcal{E}_{T,\delta} = \emptyset$ for all $T > U_\delta(\mu)$. This concludes the proof when $\mathcal{A}_\theta = \emptyset$.

Case 2: when $\mathcal{A}_\theta \neq \emptyset$. Let $S_\mu(\delta)$ defined similarly as in Lemma 19, *i.e.*

$$S_\mu(\delta) := \sup \{T \mid T \leq 4H_1(\mu) f_1(T, \delta) + K + 2|\mathcal{A}_\theta|\}.$$

To prove Theorem 2 when $\mathcal{A}_\theta \neq \emptyset$, we obtain as an intermediary result that: for all $T > S_\mu(\delta)$, $\{\hat{a}_T = \emptyset\} \cup \{\hat{a}_T \in \mathcal{A}_\theta^c\} \subseteq \mathcal{E}_{T,\delta}^c$. Using a proof similar to Lemma 46, applying Lemma 44 yields that

$$\begin{aligned} T &> S_\mu(\delta) \\ \iff T &> 12H_1(\mu) \log T + 4H_1(\mu) \log \left(\frac{K\pi^2}{6\delta} \right) + K + 2|\mathcal{A}_\theta| \\ \iff \frac{T}{12H_1(\mu)} - \log \left(\frac{T}{12H_1(\mu)} \right) &> \frac{1}{3} \log \left(\frac{K\pi^2}{6\delta} \right) + \frac{K + 2|\mathcal{A}_\theta|}{12H_1(\mu)} + \log(12H_1(\mu)) \\ \iff T &> 12H_1(\mu) \overline{W}_{-1} \left(\frac{1}{3} \log \left(\frac{K\pi^2}{6\delta} \right) + \frac{K + 2|\mathcal{A}_\theta|}{12H_1(\mu)} + \log(12H_1(\mu)) \right), \end{aligned}$$

Let us define $U_\delta(\mu) := h_2(\delta, 12H_1(\mu), K + 2|\mathcal{A}_\theta|)$ where h_2 is as above. Then, we have shown that $(\{\hat{a}_T = \emptyset\} \cup \{\hat{a}_T \in \mathcal{A}_\theta^c\}) \cap \mathcal{E}_{T,\delta} = \emptyset$ for all $T > U_\delta(\mu)$. This concludes the proof when $\mathcal{A}_\theta \neq \emptyset$. ■

B.4 Time Uniform Probability of Error

Corollary 20 gives an upper bound on the time-uniform probability of error of APGAI.

Corollary 20 *Let α_{i_μ} as in Theorem 2. The APGAI algorithm $\mathfrak{A}$ satisfies that, for all $\nu \in \mathcal{D}^K$ with mean μ such that $\Delta_{\min} > 0$,*

$$\mathbb{P}_\nu \left(\bigcup_{t > K + 2|\mathcal{A}_\theta|} \mathcal{E}_{\mathfrak{A}}^{err}(t) \right) \leq \inf_{\delta \in (0,1)} \{ \delta + K\alpha_{i_\mu} H_1(\mu) e\sqrt{2}\gamma_\mu(\delta) \},$$

where $\gamma_\mu(\delta)$ as in Eq. (10) satisfies that $\limsup_{\delta \rightarrow 0} \gamma_\mu(\delta) < +\infty$.

Proof Combining Theorems 2 and 4, one can easily upper bound the time-uniform probability or error to obtain Corollary 20. Let $\delta \in (0, 1)$ and $U_\delta(\mu)$ as in Theorem 4. Let $p(x) = x - 0.5 \log x$ and α_{i_μ} as in Theorem 2. Using Theorems 2 and 4, a union bound yields

$$\begin{aligned}
 & \mathbb{P}_\nu \left(\bigcup_{T > K+2|\mathcal{A}_\theta|} \mathcal{E}_{\mathfrak{A}}^{\text{err}}(T) \right) \\
 & \leq \mathbb{P}_\nu \left(\bigcup_{K+2|\mathcal{A}_\theta| < T \leq U_\delta(\mu)} \mathcal{E}_{\mathfrak{A}}^{\text{err}}(T) \right) + \mathbb{P}_\nu \left(\bigcup_{T > U_\delta(\mu)} \mathcal{E}_{\mathfrak{A}}^{\text{err}}(T) \right) \\
 & \leq \delta + K e \sqrt{2} \sum_{K+2|\mathcal{A}_\theta| < T \leq U_\delta(\mu)} \log(e^2 T) \sqrt{\frac{T - K - 2|\mathcal{A}_\theta|}{2\alpha_{i_\mu} H_1(\mu)}} \exp \left(-\frac{T - K - 2|\mathcal{A}_\theta|}{2\alpha_{i_\mu} H_1(\mu)} \right) \\
 & \leq \delta + K \alpha_{i_\mu} H_1(\mu) e \sqrt{2} \int_{(0, x_\mu(\delta))} (2 + \log(2\alpha_{i_\mu} H_1(\mu)x + K + 2|\mathcal{A}_\theta|)) \sqrt{x} e^{-x} dx,
 \end{aligned}$$

where $x_\mu(\delta) := \frac{U_\delta(\mu) - K - 2|\mathcal{A}_\theta|}{2\alpha_{i_\mu} H_1(\mu)}$. The last inequality uses that $T \leq U_\delta(\mu)$ and bounds the summation by the integral with the change of variable $x = \frac{T - K - 2|\mathcal{A}_\theta|}{2\alpha_{i_\mu} H_1(\mu)}$. The lower incomplete gamma function is defined $\gamma(s, x) = \int_{x \in (0, x)} t^{s-1} \exp(-t) dt$. Let $(\alpha, \beta) \in (\mathbb{R}_+)^2$ such that $\beta > 1$. Then, we define

$$\tilde{\gamma}(s, x, \alpha, \beta) := \int_{x \in (0, x)} (2 + \log(\alpha t + \beta)) t^{s-1} \exp(-t) dt,$$

Therefore, we have shown that

$$\begin{aligned}
 & \mathbb{P}_\nu \left(\bigcup_{T > K+2|\mathcal{A}_\theta|} \mathcal{E}_{\mathfrak{A}}^{\text{err}}(T) \right) \leq \delta + K \alpha_{i_\mu} H_1(\mu) e \sqrt{2} \gamma_\mu(\delta), \\
 & \text{where } \gamma_\mu(\delta) := \tilde{\gamma} \left(\frac{3}{2}, \frac{U_\delta(\mu) - K - 2|\mathcal{A}_\theta|}{2\alpha_{i_\mu} H_1(\mu)}, 2\alpha_{i_\mu} H_1(\mu), K + 2|\mathcal{A}_\theta| \right). \quad (10)
 \end{aligned}$$

Taking the infimum over $\delta \in (0, 1)$ concludes the proof. Up to multiplicative constant depending on (α, β) , $\tilde{\gamma}$ behaves similarly as γ when $x \rightarrow +\infty$, as the behavior of $t \mapsto \log(\alpha t + \beta) t^{s-1} e^{-t}$ resembles the one of $t \mapsto t^{s-1} e^{-t}$. Since $\lim_{x \rightarrow +\infty} \gamma(s, x) = \Gamma(s)$ where Γ is the gamma function, we have $\limsup_{\delta \rightarrow 0} \gamma_\mu(\delta) < +\infty$ and we conjecture that

$$\limsup_{\delta \rightarrow 0} \tilde{\gamma} \left(\frac{3}{2}, \frac{U_\delta(\mu) - K - 2|\mathcal{A}_\theta|}{2\alpha_{i_\mu} H_1(\mu)}, 2\alpha_{i_\mu} H_1(\mu), K + 2|\mathcal{A}_\theta| \right) = \mathcal{O}(\log H_1(\mu)).$$

■

Appendix C. Analysis of Other GAI Algorithms

In Appendix C, we give extensive guarantees for uniform sampling (Unif) in GAI (Appendix C.1) anytime guarantees (Appendix C.1.1), unverifiable sample complexity bounds

(Appendix C.1.2) and fixed confidence guarantees (Appendix C.1.3). We also provide fixed-budget guarantees of Sequential Halving and Successive Reject when modified to tackle GAI (SH-G in Appendix C.2 and SR-G in Appendix C.3).

C.1 Uniform Sampling (Unif)

Uniform sampling (Unif) combines a uniform round-robin sampling rule with the recommendation rule used by APGAI, namely

$$\hat{a}_T = \emptyset \quad \text{if } \max_{a \in \mathcal{A}} \hat{\mu}_a(T) \leq \theta \quad \text{else} \quad \hat{a}_T \in \arg \max_{a \in \mathcal{A}} W_a^+(T). \quad (11)$$

At time t such that $t/K \in \mathbb{N}$, the recommendation of Unif is equivalent to outputting the arm with the largest empirical mean when $\max_{a \in \mathcal{A}} \hat{\mu}_a(T) > \theta$ since $\arg \max_{a \in \mathcal{A}} W_a^+(t) = \arg \max_{a \in \mathcal{A}} \hat{\mu}_a(t)$ and $N_a(t) = t/K$ for all $a \in \mathcal{A}$. The goal is to compare the rate obtained in the exponential decrease of the probability of error with the one in Theorem 2. Since they have the same recommendation rule, this would allow us to measure the benefit of adaptive sampling.

C.1.1 ANYTIME GUARANTEES ON THE PROBABILITY OF ERROR

Theorem 21 shows that the exponential decrease of the probability of error of Unif is linear as a function of time.

Theorem 21 *Let $\mathfrak{A}$ be Unif with recommendation rule Eq. (11). Then, for any 1-sub-Gaussian distribution $\nu \in \mathcal{D}^K$ with mean μ such that $\Delta_{\min} > 0$, and for all $t > K$ such that $t/K \in \mathbb{N}$,*

$$\begin{aligned} \text{if } \mathcal{A}_\theta = \emptyset, \quad P_{\nu, \mathfrak{A}}^{\text{err}}(t) &\leq K \exp \left(-\frac{t \min_{a \in \mathcal{A}} \Delta_a^2}{2K} \right), \\ \text{if } \mathcal{A}_\theta \neq \emptyset, \quad P_{\nu, \mathfrak{A}}^{\text{err}}(t) &\leq (|\mathcal{A}_\theta^c| + 1) \exp \left(-\frac{T \max_{a \in \mathcal{A}_\theta} \Delta_a^2}{4K} \right). \end{aligned}$$

Proof For the sake of simplicity, we consider only times t that are multiples of K . Therefore, at time T , we have $N_a(T) = T/K$ for all arms $a \in \mathcal{A}$. We distinguish between the cases (1) $\mathcal{A}_\theta = \emptyset$ and (2) $\mathcal{A}_\theta \neq \emptyset$.

Case 1: $\mathcal{A}_\theta = \emptyset$. When $\mathcal{A}_\theta = \emptyset$, we have $\mathcal{E}_\mu^{\text{err}}(T) = \{\hat{a}_T \neq \emptyset\} = \{\max_{a \in \mathcal{A}} \hat{\mu}_a(T) > \theta\} = \bigcup_{a \in \mathcal{A}} \{\hat{\mu}_a(T) > \theta\}$. Since the empirical are deterministic and the observations comes from a 1-sub-Gaussian with mean $\mu_a < \theta$, we obtain that for all $a \in \mathcal{A}$

$$\mathbb{P}_\nu(\hat{\mu}_a(T) > \theta) = \mathbb{P} \left(\frac{K}{T} \sum_{s=1}^{T/K} X_s > \Delta_a \right) \leq \exp \left(-\frac{T \Delta_a^2}{2K} \right).$$

Using that $H_6(\mu) = 1/\min_{a \in \mathcal{A}} \Delta_a^2$, a direct union bound yields that

$$P_{\nu, \mathfrak{A}}^{\text{err}}(T) \leq \sum_{a \in [K]} \exp \left(-\frac{T \Delta_a^2}{2K} \right) \leq K \exp \left(-\frac{T \min_{a \in \mathcal{A}} \Delta_a^2}{2K} \right).$$

Case 2: $\mathcal{A}_\theta \neq \emptyset$. When $\mathcal{A}_\theta = \emptyset$, we have $\mathcal{E}_\mu^{\text{err}}(T) = \{\hat{a}_T = \emptyset\} \cup \{\hat{a}_T \in \mathcal{A}_\theta^{\mathbb{C}}\}$, hence

$$\mathcal{E}_\mu^{\text{err}}(T) = \{\max_{a \in \mathcal{A}} \hat{\mu}_a(T) \leq \theta\} \cup \{\max_{a \in \mathcal{A}} \hat{\mu}_a(T) > \theta, \arg \max_{a \in \mathcal{A}} W_a^+(T) \cap \mathcal{A}_\theta^{\mathbb{C}} \neq \emptyset\}.$$

Let $a^* \in \arg \max_{a \in \mathcal{A}} \mu_a$. By inclusion, we have $\{\max_{a \in \mathcal{A}} \hat{\mu}_a(T) \leq \theta\} \subset \{\hat{\mu}_{a^*}(T) \leq \theta\}$. Therefore, since $N_{a^*}(T) = T/K$ using similar argument as above yields that

$$\mathbb{P}_\nu(\hat{\mu}_{a^*}(T) \leq \theta) \leq \exp\left(-\frac{T \max_{a \in \mathcal{A}} \Delta_a^2}{2K}\right).$$

Since $N_a(T) = T/K$ for all $a \in \mathcal{A}$, we have $\arg \max_{a \in \mathcal{A}} W_a^+(T) = \arg \max_{a \in \mathcal{A}} \hat{\mu}_a(T)$. Therefore, we have

$$\{\max_{a \in \mathcal{A}} \hat{\mu}_a(T) > \theta, \arg \max_{a \in \mathcal{A}} W_a^+(T) \cap \mathcal{A}_\theta^{\mathbb{C}} \neq \emptyset\} \subseteq \bigcup_{b \notin \mathcal{A}_\theta} \{\hat{\mu}_b(T) \geq \hat{\mu}_{a^*}(T)\}.$$

Likewise, we obtain that

$$\mathbb{P}_\nu(\hat{\mu}_b(T) \geq \hat{\mu}_{a^*}(T)) = \mathbb{P}\left(\frac{K}{T} \sum_{s=1}^{T/K} (X_s - Y_s) \geq \mu_{a^*} - \mu_b\right) \leq \exp\left(-\frac{T(\mu_{a^*} - \mu_b)^2}{4K}\right).$$

Therefore, we obtain

$$\begin{aligned} P_{\nu, \mathfrak{A}}^{\text{err}}(T) &\leq \exp\left(-\frac{T \max_{a \in \mathcal{A}} \Delta_a^2}{2K}\right) + \sum_{a \notin \mathcal{A}_\theta} \exp\left(-\frac{T(\mu_{a^*} - \mu_b)^2}{4K}\right) \\ &\leq \exp\left(-\frac{T \max_{a \in \mathcal{A}_\theta} \Delta_a^2}{2K}\right) + |\mathcal{A}_\theta^{\mathbb{C}}| \exp\left(-\frac{T(\max_{a \in \mathcal{A}_\theta} \Delta_a + \min_{b \notin \mathcal{A}_\theta} \Delta_b)^2}{4K}\right) \\ &\leq (|\mathcal{A}_\theta^{\mathbb{C}}| + 1) \exp\left(-\frac{T \max_{a \in \mathcal{A}_\theta} \Delta_a^2}{4K}\right). \end{aligned}$$

■

C.1.2 UNVERIFIABLE SAMPLE COMPLEXITY

Theorem 22 gives a deterministic upper bound $U_\delta(\mu)$ on the unverifiable sample complexity $\tau_{U, \delta}$ of Unif for GAI. Its proof is similar to the one of Theorem 4 by adapting the arguments used in Theorem 21.

Theorem 22 *Let $\delta \in (0, 1)$. The Unif algorithm satisfies that, for any 1-sub-Gaussian distribution with mean μ such that $\Delta_{\min} > 0$, we have $\mathbb{P}_\nu(\bigcup_{t \geq U_\delta(\mu)} \mathcal{E}_{\mathfrak{A}}^{\text{err}}(t)) \leq \delta$ where*

$$U_\delta(\mu) = \begin{cases} h_2\left(\delta, \frac{6K}{\min_{a \in \mathcal{A}} \Delta_a^2}, K\right) & \text{if } \mathcal{A}_\theta = \emptyset \\ h_2\left(\delta, \frac{24K}{\max_{a \in \mathcal{A}_\theta} \Delta_a^2}, K\right) & \text{if } \mathcal{A}_\theta \neq \emptyset \end{cases},$$

where h_2 is defined in Theorem 4.

Proof Let $T > K$ and $\delta \in (0, 1)$. Let $\mathcal{E}_{T,\delta}$ as in Eq. (21) for $s = 2$, i.e.,

$$\mathcal{E}_{T,\delta} = \left\{ \forall a \in \mathcal{A}, \forall t \leq T, |\hat{\mu}_a(t) - \mu_a| < \sqrt{\frac{2f_1(T, \delta)}{N_a(t)}} \right\},$$

where $f_1(T, \delta) = \log(1/\delta) + 3 \log T + \log(K\pi^2/6)$. Let $U_\delta(\mu) > K$ to be specified below. Suppose that $U_\delta(\mu)$ is chosen such that $\mathcal{E}_{\mathfrak{A}}^{\text{err}}(T) \cap \mathcal{E}_{T,\delta} = \emptyset$ for all $T > U_\delta(\mu)$. Using the same arguments as in Theorem 4, we can conclude the proof by exhibiting $U_\delta(\mu)$ satisfying the above property since

$$\mathbb{P}_\nu \left(\bigcup_{T > U_\delta(\mu)} \mathcal{E}_{\mathfrak{A}}^{\text{err}}(T) \right) \leq \mathbb{P}_\nu \left(\bigcup_{T > U_\delta(\mu)} (\mathcal{E}_{\mathfrak{A}}^{\text{err}}(T) \cap \mathcal{E}_{T,\delta}) \right) + \delta \leq \delta.$$

By definition of Unif, we have $N_a(T) \geq \lfloor T/K \rfloor \geq T/K - 1$.

Case 1: when $\mathcal{A}_\theta = \emptyset$. Using the same arguments as in Theorem 21, one can show that

$$\{\hat{a}_T \neq \emptyset\} \cap \mathcal{E}_{T,\delta} \subseteq \bigcup_{a \in \mathcal{A}} \left\{ \mu_a + \sqrt{\frac{2f_1(T, \delta)}{N_a(T)}} > \theta \right\} \subseteq \left\{ \frac{2Kf_1(T, \delta)}{\min_{a \in \mathcal{A}} \Delta_a^2} + K > T \right\}.$$

Let $T_\mu(\delta) := \sup\{T \mid T \leq \frac{2Kf_1(T, \delta)}{\min_{a \in \mathcal{A}} \Delta_a^2} + K\}$. Then, we have $\{\hat{a}_T \neq \emptyset\} \cap \mathcal{E}_{T,\delta} = \emptyset$ for all $T > T_\mu(\delta)$. Let h_2 as in Theorem 4 and $U_\delta(\mu) := h_2\left(\delta, \frac{6K}{\min_{a \in \mathcal{A}} \Delta_a^2}, K\right)$. Applying Lemma 44 as in Theorem 4, we obtain $T > T_\mu(\delta)$ if and only if $T > U_\delta(\mu)$. This concludes the proof when $\mathcal{A}_\theta = \emptyset$.

Case 2: when $\mathcal{A}_\theta \neq \emptyset$. Let $a^* \in \arg \max_{a \in \mathcal{A}} \mu_a$. Then, $\max_{a \in \mathcal{A}_\theta} \Delta_a^2 = \Delta_{a^*}^2$ and $\min_{b \notin \mathcal{A}_\theta} (\mu_{a^*} - \mu_b) \geq \Delta_{a^*}^2$. Using the same arguments as in Theorem 21 and the same manipulation as above, one can show that

$$\begin{aligned} & \left(\{\hat{a}_T = \emptyset\} \cup \{\hat{a}_T \in \mathcal{A}_\theta^c\} \right) \cap \mathcal{E}_{T,\delta} \\ & \subseteq \left\{ \mu_{a^*} - \sqrt{\frac{2f_1(T, \delta)}{N_{a^*}(T)}} < \theta \right\} \cup \bigcup_{b \notin \mathcal{A}_\theta} \left\{ \mu_b + \sqrt{\frac{2f_1(T, \delta)}{N_b(T)}} > \mu_{a^*} - \sqrt{\frac{2f_1(T, \delta)}{N_{a^*}(T)}} \right\} \\ & \subseteq \left\{ \frac{8Kf_1(T, \delta)}{\max_{a \in \mathcal{A}_\theta} \Delta_a^2} + K > T \right\}. \end{aligned}$$

Taking $U_\delta(\mu) := h_2\left(\delta, \frac{24K}{\max_{a \in \mathcal{A}_\theta} \Delta_a^2}, K\right)$ concludes the proof for the case $\mathcal{A}_\theta \neq \emptyset$, similarly as above. ■

C.1.3 FIXED CONFIDENCE GUARANTEES

Theorem 23 gives an upper bound on the expected sample complexity of the Unif algorithm coupled with the GLR stopping rule Eq. (6) with threshold Eq. (7) holding for any confidence δ . Its proof resembles the one of Theorem 8. Using similar manipulation as in

Appendix F.3, one could obtain more explicit upper bound $C_\mu(\delta)$. While we omit those statements for simplicity, they would show that the δ -independent scaling of the upper bound is $\mathcal{O}\left(\frac{K}{\min_{a \in \mathcal{A}} \Delta_a^2} \log\left(\frac{K}{\min_{a \in \mathcal{A}} \Delta_a^2}\right)\right)$ when $\mathcal{A}_\theta = \emptyset$, and $\mathcal{O}\left(\frac{K}{\max_{a \in \mathcal{A}_\theta} \Delta_a^2} \log\left(\frac{K}{\max_{a \in \mathcal{A}_\theta} \Delta_a^2}\right)\right)$ otherwise.

Theorem 23 *Let $\delta \in (0, 1)$. Combined with GLR stopping Eq. (6) using threshold Eq. (7), Unif is δ -correct and it satisfies that, for all $\nu \in \mathcal{D}^K$ with mean μ such that $\Delta_{\min} > 0$,*

$$\begin{aligned} \mathbb{E}_\nu[\tau_\delta] &\leq C_\mu(\delta) + \frac{K\pi^2}{6} + 1 \quad \text{where} \\ C_\mu(\delta) &:= \begin{cases} \sup \left\{ t \mid t \leq \frac{2K}{\min_{a \in \mathcal{A}} \Delta_a^2} (\sqrt{c(t, \delta)} + \sqrt{3 \log t})^2 + K \right\} & \text{if } \mathcal{A}_\theta = \emptyset \\ \sup \left\{ t \mid t \leq \frac{2K}{\max_{a \in \mathcal{A}_\theta} \Delta_a^2} (\sqrt{c(t, \delta)} + \sqrt{3 \log t})^2 + K \right\} & \text{if } \mathcal{A}_\theta \neq \emptyset \end{cases}, \\ \text{and } \limsup_{\delta \rightarrow 0} \frac{\mathbb{E}_\nu[\tau_\delta]}{\log(1/\delta)} &\leq \begin{cases} \frac{2K}{\min_{a \in \mathcal{A}} \Delta_a^2} & \text{if } \mathcal{A}_\theta = \emptyset \\ \frac{2K}{\max_{a \in \mathcal{A}_\theta} \Delta_a^2} & \text{if } \mathcal{A}_\theta \neq \emptyset \end{cases}. \end{aligned}$$

Proof The δ -correctness property is a direct consequence of Lemma 7.

For all $T > K$, let $\mathcal{E}_T = \mathcal{E}_{T,1}$ where $\mathcal{E}_{T,\delta}$ as in Eq. (21) with $s = 2$, *i.e.*

$$\mathcal{E}_T = \left\{ \forall a \in \mathcal{A}, \forall t \leq T, |\hat{\mu}_a(t) - \mu_a| < \sqrt{2f_1(T)/N_a(t)} \right\},$$

with $f_1(T) = 3 \log T$. Using Lemma 40, we have $\sum_{T > K} \mathbb{P}_\nu(\mathcal{E}_T^c) \leq K\pi^2/6$. Suppose that we have constructed a time $T_\mu(\delta) > K$ such that $\mathcal{E}_T \subseteq \{\tau_\delta \leq T\}$ for $T \geq T_\mu(\delta)$. Then, using Lemma 43, we obtain $\mathbb{E}_\nu[\tau_\delta] \leq T_\mu(\delta) + K\pi^2/6$. Therefore, one can conclude the proof by exhibiting such $T_\mu(\delta)$. By definition of Unif, we have $N_a(T) \geq \lfloor T/K \rfloor \geq T/K - 1$.

Case 1: when $\mathcal{A}_\theta = \emptyset$. By definition of τ_δ , we have $\tau_\delta \leq \tau_{<,\delta}$ almost surely. Under $\mathcal{E}_T$, we obtain, for all $a \in \mathcal{A}$,

$$\sqrt{N_a(T)}(\theta - \hat{\mu}_a(T)) \geq \sqrt{N_a(T)}(\theta - \mu_a) - \sqrt{2f_1(T)} \geq \sqrt{T/K - 1} \min_{a \in \mathcal{A}} \Delta_a - \sqrt{6 \log(T)}.$$

Then, under $\mathcal{E}_T \cap \{\tau_\delta > T\}$, we obtain

$$\sqrt{2c(T, \delta)} \geq \min_{a \in \mathcal{A}} \sqrt{N_a(T)}(\theta - \hat{\mu}_a(T))_+ \geq \left(\sqrt{T/K - 1} \min_{a \in \mathcal{A}} \Delta_a - \sqrt{6 \log(T)} \right)_+$$

Let us define

$$C_\mu(\delta) := \sup \left\{ t \mid t \leq \frac{2K}{\min_{a \in \mathcal{A}} \Delta_a^2} (\sqrt{c(t, \delta)} + \sqrt{3 \log t})^2 + K \right\}.$$

By re-ordering the above equation, we obtain $\mathcal{E}_T \cap \{\tau_\delta > T\} = \emptyset$ for all $T > C_\mu(\delta)$. Therefore, taking $T_\mu(\delta) = C_\mu(\delta) + 1$ concludes the proof when $\mathcal{A}_\theta = \emptyset$.

Case 2: when $\mathcal{A}_\theta \neq \emptyset$. By definition of τ_δ , we have $\tau_\delta \leq \tau_{>,\delta}$ almost surely. Let $a^* \in \arg \max_{a \in \mathcal{A}} \mu_a$. Then, we have $\Delta_{a^*} = \max_{a \in \mathcal{A}_\theta} \Delta_a$. Under $\mathcal{E}_T$, we obtain,

$$\sqrt{N_{a^*}(T)}(\hat{\mu}_{a^*}(T) - \theta) \geq \sqrt{N_{a^*}(T)}(\mu_{a^*} - \theta) - \sqrt{2f_1(T)} \geq \sqrt{T/K - 1} \max_{a \in \mathcal{A}_\theta} \Delta_a - \sqrt{6 \log(T)}.$$

Then, under $\mathcal{E}_T \cap \{\tau_\delta > T\}$, we obtain

$$\sqrt{2c(T, \delta)} \geq \max_{a \in \mathcal{A}} \sqrt{N_a(T)} (\hat{\mu}_a(T) - \theta)_+ \geq \left(\sqrt{T/K - 1} \max_{a \in \mathcal{A}_\theta} \Delta_a - \sqrt{6 \log(T)} \right)_+$$

Let us define

$$C_\mu(\delta) := \sup \left\{ t \mid t \leq \frac{2K}{\max_{a \in \mathcal{A}} \Delta_a^2} (\sqrt{c(t, \delta)} + \sqrt{3 \log t})^2 + K \right\}.$$

By re-ordering the above equation, we obtain $\mathcal{E}_T \cap \{\tau_\delta > T\} = \emptyset$ for all $T > C_\mu(\delta)$. Therefore, taking $T_\mu(\delta) = C_\mu(\delta) + 1$ concludes the proof when $\mathcal{A}_\theta \neq \emptyset$.

The asymptotic upper bounds are a direct consequence of Lemma 47. $\blacksquare$

C.2 Sequential Halving for GAI (SH-G)

In Appendix C.2, we study the SH (Karnin et al., 2013) algorithm where instead of recommending the last active arm a_T , we recommend

$$\hat{a}_T = \emptyset \quad \text{if } \hat{\mu}_{a_T}(T) \leq \theta \quad \text{else} \quad \hat{a}_T = a_T. \quad (12)$$

We refer to this modified SH algorithm as SH-G. In SH, there are two arms (a_1, a_2) at the last of the $\lceil \log_2(K) \rceil$ phases. Then, both arms are pulled $N_T = \left\lfloor \frac{T}{2^{\lceil \log_2(K) \rceil}} \right\rfloor$ times. Since SH drops the sampled collected in the previous phase, the last active arm a_T is based on the comparison of the empirical mean of each arm after N_T samples.

Theorem 24 shows that the exponential decrease of the probability of error of SH-G is linear as a function of time. The notation $\tilde{\Theta}(\cdot)$ hides logarithmic factors which were not made explicit in Theorems 1 and 5 from Zhao et al. (2023). Since one component of our proof uses their result, we suffer from this lack of explicit constant in that case.

Theorem 24 *Let $T > K$. Let $\mathfrak{A}_T$ be the SH-G algorithm with recommendation rule as in Eq. (12). Then, for any 1-sub-Gaussian distribution $\nu \in \mathcal{D}^K$ with mean μ such that $\Delta_{\min} > 0$,*

$$\begin{aligned} \text{if } \mathcal{A}_\theta = \emptyset, \quad P_{\nu, \mathfrak{A}_T}^{\text{err}}(T) &\leq K \exp \left(-\frac{T \min_{a \in \mathcal{A}} \Delta_a^2}{4 \lceil \log_2(K) \rceil} + \min_{a \in \mathcal{A}} \Delta_a^2 / 2 \right), \\ \text{if } \mathcal{A}_\theta \neq \emptyset, \quad P_{\nu, \mathfrak{A}_T}^{\text{err}}(T) &\leq |\mathcal{A}_\theta| \exp \left(-\frac{T \min_{a \in \mathcal{A}_\theta} \Delta_a^2}{4 \lceil \log_2(K) \rceil} + \min_{a \in \mathcal{A}_\theta} \Delta_a^2 / 2 \right) + \\ &\min \left\{ 3 \log_2(K) \exp \left(-\frac{T}{8 \log_2(K) \max_{i > I^*} i (\max_{a \in \mathcal{A}} \mu_a - \mu_{(i)})^{-2}} \right), \exp \left(-\tilde{\Theta} \left(\frac{T}{G_1(\mu)} \right) \right) \right\} \end{aligned}$$

where $I^* = |\arg \max_{a \in \mathcal{A}} \mu_a|$ and $G_1(\mu)$ is defined in Eq. (13).

Proof We distinguish between the cases (1) $\mathcal{A}_\theta = \emptyset$ and (2) $\mathcal{A}_\theta \neq \emptyset$.

Case 1: $\mathcal{A}_\theta = \emptyset$. When $\mathcal{A}_\theta = \emptyset$, we have

$$\mathcal{E}_\mu^{\text{err}}(T) = \{\hat{a}_T \neq \emptyset\} = \{\hat{a}_T \neq \emptyset, \hat{\mu}_{a_T}(T) > \theta\} \subseteq \{\hat{\mu}_{a_T}(T) > \theta\} = \bigcup_{a \in \mathcal{A}} \{a_T = a, \hat{\mu}_a(T) > \theta\}.$$

Therefore, using $N_{a_T}(T) = N_T \geq \frac{T}{2^{\lceil \log_2(K) \rceil}} - 1$ (drop observations from past phases) and similar argument as in the proof of Theorem 21, we obtain

$$P_{\nu, \mathfrak{A}_T}^{\text{err}}(T) \leq \sum_{a \in \mathcal{A}} \exp\left(-\frac{N_T}{2} \Delta_a^2\right) \leq K e^{\min_{a \in \mathcal{A}} \Delta_a^2/2} \exp\left(-\frac{T \min_{a \in \mathcal{A}} \Delta_a^2}{4^{\lceil \log_2(K) \rceil}}\right).$$

Case 2: $\mathcal{A}_\theta \neq \emptyset$. When $\mathcal{A}_\theta \neq \emptyset$, we have $\mathcal{E}_\mu^{\text{err}}(T) = \{\hat{a}_T = \emptyset\} \cup \{\hat{a}_T \in \mathcal{A}_\theta^{\mathbb{C}}\}$. By definition of the recommendation rule of SH-G in Eq. (12), we obtain

$$\begin{aligned} \{\hat{a}_T = \emptyset\} &= \{\hat{a}_T = \emptyset, a_T \in \mathcal{A}_\theta, \hat{\mu}_{a_T}(T) \leq \theta\} \cup \{\hat{a}_T = \emptyset, a_T \in \mathcal{A}_\theta^{\mathbb{C}}, \hat{\mu}_{a_T}(T) \leq \theta\} \\ &\subseteq \{a_T \in \mathcal{A}_\theta, \hat{\mu}_{a_T}(T) \leq \theta\} \cup \{a_T \in \mathcal{A}_\theta^{\mathbb{C}}, \hat{\mu}_{a_T}(T) \leq \theta\}, \\ \{\hat{a}_T \in \mathcal{A}_\theta^{\mathbb{C}}\} &= \{\hat{a}_T \in \mathcal{A}_\theta^{\mathbb{C}}, a_T \in \mathcal{A}_\theta^{\mathbb{C}}, \hat{\mu}_{a_T}(T) > \theta\} \subseteq \{a_T \in \mathcal{A}_\theta^{\mathbb{C}}\}. \end{aligned}$$

The dichotomy on whether the last active arm a_T is a good arm or not is crucial when $\hat{a}_T = \emptyset$. When $a_T \in \mathcal{A}_\theta$, having $\hat{a}_T = \emptyset$ implies that this arm was not sampled enough to ensure that $\hat{\mu}_{a_T}(T) > \theta$, even though it satisfies $\mu_{a_T} > \theta$. Since it is sampled linearly, it means that the budget T is not large enough compared to the difficulty $1/\min_{a \in \mathcal{A}_\theta} \Delta_a^2$. When $a_T \notin \mathcal{A}_\theta$, having $\hat{a}_T = \emptyset$ implies that all the good arms have been eliminated in previous phases. Therefore, SH has eliminated the best arm in previous phases, namely we have

$$\{a_T \in \mathcal{A}_\theta^{\mathbb{C}}\} \subseteq \{a_T \notin a^*(\mu)\} \quad \text{where} \quad a^*(\mu) := \arg \max_{a \in [K]} \mu_a \subseteq \mathcal{A}_\theta.$$

Using existing analysis of SH, $\{a_T \notin a^*(\mu)\}$ is known to have a low probability of occurring. Putting everything together, we have shown that

$$\mathcal{E}_\mu^{\text{err}}(T) \subseteq \{a_T \in \mathcal{A}_\theta, \hat{\mu}_{a_T}(T) \leq \theta\} \cup \{a_T \in \mathcal{A}_\theta^{\mathbb{C}}\} \subseteq \{a_T \in \mathcal{A}_\theta, \hat{\mu}_{a_T}(T) \leq \theta\} \cup \{a_T \notin a^*(\mu)\}.$$

Since $N_{a_T}(T) = N_T \geq \frac{T}{2^{\lceil \log_2(K) \rceil}} - 1$, using similar argument as above yields that

$$\begin{aligned} \mathbb{P}_\nu(a_T \in \mathcal{A}_\theta, \hat{\mu}_{a_T}(T) \leq \theta) &\leq \sum_{a \in \mathcal{A}_\theta} \exp\left(-\frac{N_T}{2} \Delta_a^2\right) \\ &\leq |\mathcal{A}_\theta| e^{\min_{a \in \mathcal{A}_\theta} \Delta_a^2/2} \exp\left(-\frac{T \min_{a \in \mathcal{A}_\theta} \Delta_a^2}{4^{\lceil \log_2(K) \rceil}}\right). \end{aligned}$$

Using Theorem 4.1 from Karnin et al. (2013) for SH yields

$$\mathbb{P}_\nu(a_T \notin a^*(\mu)) \leq 3 \log_2(K) \exp\left(-\frac{T}{8 \log_2(K) \max_{i > I^*} i (\max_{a \in \mathcal{A}} \mu_a - \mu_{(i)})^{-2}}\right)$$

where $I^* = |\arg \max_{a \in \mathcal{A}} \mu_a|$.

Improved case 2. Instead of simply using Theorem 4.1 Karnin et al. (2013), we can use recent results from Zhao et al. (2023) by noting that

$$\{a_T \in \mathcal{A}_\theta^{\mathbb{C}}\} = \bigcup_{\varepsilon \in (\max_{a \in \mathcal{A}_\theta} \Delta_a + \min_{b \notin \mathcal{A}_\theta} \Delta_b, \max_{a \in \mathcal{A}_\theta} \mu_a - \min_{b \in \mathcal{A}_\theta} \mu_b)} \{\mu_{a_T} < \mu_{a^*} - \varepsilon\}.$$

Then, using Theorem 1 from Zhao et al. (2023) and taking the infimum over ε yields that $\mathbb{P}_\nu(a_T \in \mathcal{A}_\theta^c) \leq \exp\left(-\tilde{\Theta}\left(\frac{T}{G_1(\mu)}\right)\right)$ with

$$G_1(\mu) = \min_{\varepsilon \in (\max_{a \in \mathcal{A}_\theta} \Delta_a + \min_{b \notin \mathcal{A}_\theta} \Delta_b, \max_{a \in \mathcal{A}_\theta} \mu_a - \min_{b \in \mathcal{A}_\theta} \mu_b)} \max_{i \geq g(\varepsilon)+1} \frac{i}{g(\varepsilon/2)(\mu_{a^*} - \mu_{(i)})^2}, \quad (13)$$

where $g(\varepsilon) = |\{a \in \mathcal{A} \mid \mu_a \geq \mu_{a^*} - \varepsilon\}|$. ■

Doubling SH. It is possible to convert the fixed-budget SH-G algorithm into an anytime algorithm by using the doubling trick. It considers a sequences of algorithms that are run with increasing budgets $(T_k)_{k \geq 1}$, with $T_{k+1} = 2T_k$ and $T_1 = 2K \lceil \log_2 K \rceil$, and recommend the answer outputted by the last instance that has finished to run. Theorem 5 from Zhao et al. (2023) shows that Doubling SH achieves the same guarantees than SH for any time t , where the “cost” of doubling is hidden by the $\tilde{\Theta}(\cdot)$ notation. It is well know that the “cost” of doubling is to have a multiplicative factor 4 in front of the hardness constant. The first two-factor is due to the fact that we forget half the observations. The second two-factor is due to the fact that we use the recommendation from the last instance of SH that has finished. Therefore, Theorem 24 can be modified for DSH-G by simply adding this multiplicative factor 4.

While it might look to be a mild cost, this intervenes inside the exponential hence we need four times as many samples to achieves the same error. For application where sampling is limited, this price is to high to be paid in practice. Moreover, since past observations are dropped when reached budget T_k , doubling-based algorithms are known to have empirical performances that decreases by steps.

C.3 Successive Reject for GAI (SR-G)

In Appendix C.3, we study the SR (Audibert et al., 2010) algorithm where instead of recommending the last active arm a_T , we use the recommendation Eq. (12). We refer to this modified SR algorithm as SR-G. In SR, there is only one arm a_T at time T since we eliminated all but one arm after $K - 1$ phases. Let us denote by $n_k = \left\lceil \frac{T-K}{\log(K)(K+1-k)} \right\rceil$ and $u_T = \sum_{k=1}^{K-1} n_k$, where $\overline{\log}(K) = \frac{1}{2} + \sum_{i=2}^K \frac{1}{i}$. Therefore, we have $N_{a_T}(T) = T - u_T$.

Theorem 25 shows that the exponential decrease of the probability of error of SR-G is linear as a function of time.

Theorem 25 *Let $T > K$. Let $\mathfrak{A}_T$ be the SR-G algorithm with recommendation rule as in Eq. (12). Then, for any 1-sub-Gaussian distribution $\nu \in \mathcal{D}^K$ with mean μ such that $\Delta_{\min} > 0$,*

$$\begin{aligned} \text{if } \mathcal{A}_\theta = \emptyset, \quad P_{\nu, \mathfrak{A}_T}^{\text{err}}(T) &\leq K \exp\left(-\frac{T-K}{4\overline{\log}(K)} \min_{a \in \mathcal{A}} \Delta_a^2\right), \\ \text{if } \mathcal{A}_\theta \neq \emptyset, \quad P_{\nu, \mathfrak{A}_T}^{\text{err}}(T) &\leq |\mathcal{A}_\theta| \exp\left(-\frac{T-K}{4\overline{\log}(K)} \min_{a \in \mathcal{A}_\theta} \Delta_a^2\right) + \\ &\quad K^2 \exp\left(-\frac{T-K}{\overline{\log}(K) \max_{i > I^*} i (\max_{a \in \mathcal{A}} \mu_a - \mu_{(i)})^{-2}}\right), \end{aligned}$$

where $I^* = |\arg \max_{a \in \mathcal{A}} \mu_a|$.

Proof We distinguish between the cases (1) $\mathcal{A}_\theta = \emptyset$ and (2) $\mathcal{A}_\theta \neq \emptyset$.

Case 1: $\mathcal{A}_\theta = \emptyset$. When $\mathcal{A}_\theta = \emptyset$, we have

$$\mathcal{E}_\mu^{\text{err}}(T) = \{\hat{a}_T \neq \emptyset\} = \{\hat{a}_T \neq \emptyset, \hat{\mu}_{a_T}(T) > \theta\} \subseteq \{\hat{\mu}_{a_T}(T) > \theta\} = \bigcup_{a \in \mathcal{A}} \{a_T = a, \hat{\mu}_a(T) > \theta\}.$$

Therefore, using $N_{a_T}(T) = T - u_T$ and similar argument as in the proof of Theorem 21, we obtain

$$P_{\nu, \mathfrak{A}_T}^{\text{err}}(T) \leq \sum_{a \in \mathcal{A}} \exp\left(-\frac{T - u_T}{2} \Delta_a^2\right) \leq K \exp\left(-\frac{T - K}{4\log(K)} \min_{a \in \mathcal{A}} \Delta_a^2\right),$$

where the last inequality uses that $T - u_T \geq \frac{T-K}{2\log(K)}$.

Case 2: $\mathcal{A}_\theta \neq \emptyset$. When $\mathcal{A}_\theta \neq \emptyset$, we have $\mathcal{E}_\mu^{\text{err}}(T) = \{\hat{a}_T = \emptyset\} \cup \{\hat{a}_T \in \mathcal{A}_\theta^c\}$. By definition of the recommendation rule of SR-G in Eq. (12), we obtain

$$\begin{aligned} \{\hat{a}_T = \emptyset\} &= \{\hat{a}_T = \emptyset, a_T \in \mathcal{A}_\theta, \hat{\mu}_{a_T}(T) \leq \theta\} \cup \{\hat{a}_T = \emptyset, a_T \in \mathcal{A}_\theta^c, \hat{\mu}_{a_T}(T) \leq \theta\} \\ &\subseteq \{a_T \in \mathcal{A}_\theta, \hat{\mu}_{a_T}(T) \leq \theta\} \cup \{a_T \in \mathcal{A}_\theta^c\}, \\ \{\hat{a}_T \in \mathcal{A}_\theta^c\} &= \{\hat{a}_T \in \mathcal{A}_\theta^c, a_T \in \mathcal{A}_\theta^c, \hat{\mu}_{a_T}(T) > \theta\} \subseteq \{a_T \in \mathcal{A}_\theta^c\}. \end{aligned}$$

The dichotomy on whether the last active arm a_T is a good arm or not is crucial when $\hat{a}_T = \emptyset$. When $a_T \in \mathcal{A}_\theta$, having $\hat{a}_T = \emptyset$ implies that this arm was not sampled enough to ensure that $\hat{\mu}_{a_T}(T) > \theta$, even though it satisfies $\mu_{a_T} > \theta$. Since it is sampled linearly, it means that the budget T is not large enough compared to the difficulty $1/\min_{a \in \mathcal{A}_\theta} \Delta_a^2$. When $a_T \notin \mathcal{A}_\theta$, having $\hat{a}_T = \emptyset$ implies that all the good arms have been eliminated in previous phases. Therefore, SR has eliminated the best arm in previous phases, namely we have

$$\{a_T \in \mathcal{A}_\theta^c\} \subseteq \{a_T \notin a^*(\mu)\} \quad \text{where} \quad a^*(\mu) := \arg \max_{a \in [K]} \mu_a \subseteq \mathcal{A}_\theta.$$

Using existing analysis of SR, $\{a_T \notin a^*(\mu)\}$ is known to have a low probability of occurring. Putting everything together, we have shown that

$$\mathcal{E}_\mu^{\text{err}}(T) \subseteq \{a_T \in \mathcal{A}_\theta, \hat{\mu}_{a_T}(T) \leq \theta\} \cup \{a_T \in \mathcal{A}_\theta^c\} \subseteq \{a_T \in \mathcal{A}_\theta, \hat{\mu}_{a_T}(T) \leq \theta\} \cup \{a_T \notin a^*(\mu)\}.$$

Since $N_{a_T}(T) = T - u_T \geq \frac{T-K}{2\log(K)}$, using similar argument as above yields that

$$\mathbb{P}_\nu(a_T \in \mathcal{A}_\theta, \hat{\mu}_{a_T}(T) \leq \theta) \leq \sum_{a \in \mathcal{A}_\theta} \exp\left(-\frac{(T-K)\Delta_a^2}{4\log(K)}\right) \leq |\mathcal{A}_\theta| \exp\left(-\frac{T-K}{4\log(K)} \min_{a \in \mathcal{A}_\theta} \Delta_a^2\right).$$

Using Theorem 2 from Audibert et al. (2010) for SR yields

$$\mathbb{P}_\nu(a_T \notin a^*(\mu)) \leq \frac{K(K-1)}{2} \exp\left(-\frac{T-K}{\log(K) \max_{i > I^*} i(\max_{a \in \mathcal{A}} \mu_a - \mu_{(i)})^2}\right).$$

where $I^* = |\arg \max_{a \in \mathcal{A}} \mu_a|$.

Improved case 2: $\mathcal{A}_\theta \neq \emptyset$. As in the proof of Theorem 24, using $\{\hat{a}_T \in \mathcal{A}_\theta^c\} \subset \{\hat{a}_T \neq a^*\}$ can lead to highly sub-optimal rate on some instances. Inspired by the recent analysis of SH conducted in Zhao et al. (2023), we believe that improved guarantees can also be achieved for SR. Namely, it should be able to control $\mathbb{P}_\nu(\mu_{a_T} < \max_{a \in \mathcal{A}} \mu_a - \varepsilon)$ for any $\varepsilon > 0$. Proving such improved guarantees on SR is beyond the scope of this paper, hence we let this question as open problem. However, it is possible to get some intuition on the dependency we would get for GAI.

The core argument of the analysis of SR is to say that if we make a mistake at time T , then there exists a phase k such that the best arm was eliminated at the end of phase k . This argument can be adapted to GAI. A necessary condition for the event $\{\hat{a}_T \in \mathcal{A}_\theta^c\}$ to occurs is that all arms $a \in \mathcal{A}_\theta$ are eliminated. By definition, all arms are eliminated if and only if there exists a set of phases $\{k_a\}_{a \in \mathcal{A}_\theta}$ such that, any arm $a \in \mathcal{A}_\theta$ is eliminated at the end of phase k_a . Let $\{k_a\}_{a \in \mathcal{A}_\theta}$ be a given set of phases and $a \in \mathcal{A}_\theta$. A necessary condition for an arm a to be eliminated at the end of phase k_a is that $\hat{\mu}_a(n_{k_a}) \leq \max_{b \notin \mathcal{A}_\theta} \hat{\mu}_b(n_{k_a})$. Since both arms have been sampled n_{k_a} times, using similar arguments as the one in the proof of Theorem 21, we obtain that

$$\mathbb{P}_\nu(\hat{\mu}_a(n_{k_a}) \leq \max_{b \notin \mathcal{A}_\theta} \hat{\mu}_b(n_{k_a})) \leq \exp\left(-\frac{n_{k_a}}{4}(\Delta_a + \min_{b \notin \mathcal{A}_\theta} \Delta_b)^2\right).$$

Therefore, by union bound and inclusion of event, we have shown that

$$\mathbb{P}_\nu(\hat{a}_T \in \mathcal{A}_\theta^c) \leq |\mathcal{A}_\theta^c| \sum_{\{k_a\}_{a \in \mathcal{A}_\theta}} \exp\left(-\frac{T-K}{4\log(K)} \max_{a \in \mathcal{A}_\theta} \frac{(\Delta_a + \min_{b \notin \mathcal{A}_\theta} \Delta_b)^2}{K+1-k_a}\right).$$

where we used that $n_k \geq \frac{T-K}{\log(K)(K+1-k)}$ and $\mathbb{P}_\nu(\bigcap_i A_i) \leq \min_i \mathbb{P}_\nu(A_i)$. A simple combinatorial argument yields that there are $\binom{K-1}{|\mathcal{A}_\theta|}$ possibilities to define a set of $|\mathcal{A}_\theta|$ phases within the $K-1$ total phases where an arm can be eliminated. Accounting for the $|\mathcal{A}_\theta|!$ possible re-ordering, we have $|\mathcal{A}_\theta|! \binom{K-1}{|\mathcal{A}_\theta|} = \frac{(K-1)!}{(K-1-|\mathcal{A}_\theta|)!}$ possible set of phases $\{k_a\}_{a \in \mathcal{A}_\theta}$ that eliminate all arms in $\mathcal{A}_\theta$. By upper bounding all the above probability by their smallest term, we obtain that

$$\mathbb{P}_\nu(\hat{a}_T \in \mathcal{A}_\theta^c) \leq \frac{(K-1)!}{(|\mathcal{A}_\theta^c| - 1)!} |\mathcal{A}_\theta^c| \exp\left(-\frac{T-K}{4\log(K)G_2(\mu)}\right)$$

where $G_2(\mu) = \max_{\substack{p: \mathcal{A}_\theta \rightarrow [K-1] \\ p \text{ injective}}} \min_{a \in \mathcal{A}_\theta} \frac{K+1-p(a)}{(\Delta_a + \min_{b \notin \mathcal{A}_\theta} \Delta_b)^2}$. ■

Doubling SR. Likewise, it is possible to convert the fixed-budget SR-G algorithm into an anytime algorithm by using the doubling trick. Therefore, Theorem 25 can be modified for DSR-G by simply adding the multiplicative factor 4 in front of each hardness constant.

C.3.1 LARGE DEVIATION ANALYSIS

A key benefit of Theorem 25 is that it holds for any moderate budget T . When one is only interested by the asymptotic error rate $C(\mu)$ of SR-G, as reported in Table 2, one can leverage asymptotic results such as the Large Deviation Principle (LDP). We build on the

recent analysis proposed by Wang et al. (2024b) to provide improved asymptotic error rate for SR-G and DSR-G. Namely, we combine the arguments presented in the proof of their Theorem 2 in Section 3.4 with the proof of Theorem 25. In both cases, we recover exactly the asymptotic upper bound obtained in Theorem 25.

Theorem 26 *Let $T > K$. Let $\mathfrak{A}_T$ be the SR-G algorithm with recommendation rule as in Eq. (12). Then, for any 1-sub-Gaussian distribution $\nu \in \mathcal{D}^K$ with mean μ such that $\Delta_{\min} > 0$,*

$$\begin{aligned} \text{if } \mathcal{A}_\theta = \emptyset, \quad & \liminf_{T \rightarrow +\infty} \frac{1}{T} \log \frac{1}{P_{\nu, \mathfrak{A}_T}^{\text{err}}(T)} \geq \frac{\Delta_{\min}^2}{4\log(K)}, \\ \text{if } \mathcal{A}_\theta \neq \emptyset, \quad & \liminf_{T \rightarrow +\infty} \frac{1}{T} \log \frac{1}{P_{\nu, \mathfrak{A}_T}^{\text{err}}(T)} \geq \frac{1}{4\log(K)G_2(\mu)}. \end{aligned}$$

where

$$G_2(\mu) = \max_{\substack{p: \mathcal{A}_\theta \rightarrow [K-1] \\ p \text{ injective}}} \min_{a \in \mathcal{A}_\theta} \frac{K+1-p(a)}{(\Delta_a + \min_{b \notin \mathcal{A}_\theta} \Delta_b)^2}. \quad (14)$$

Proof Case 1: $\mathcal{A}_\theta = \emptyset$. Since the lower order terms disappear asymptotically, we have

$$\liminf_{T \rightarrow +\infty} \frac{1}{T} \log \frac{1}{\mathbb{P}_\nu(\hat{a}_T \neq \emptyset)} \geq \min_{a \in \mathcal{A}} \liminf_{T \rightarrow +\infty} \frac{1}{T} \log \frac{1}{\mathbb{P}_\nu(a_T = a, \hat{\mu}_a(T) > \theta)}.$$

Recall that $N_{a_T}(T) = T - u_T \geq \frac{T-K}{2\log(K)}$. Let $\varepsilon > 0$ and T_ε such that $1 - u_T/T \geq \frac{1-\varepsilon}{2\log(K)}$ for all $T \geq T_\varepsilon$. Let $T \geq T_\varepsilon$. The event $\{a_T = a, \hat{\mu}_a(T) > \theta\}$ implies that $\{\hat{\mu}(T) \in \mathcal{S}_a, N(T)/T \in \mathcal{W}_a\}$ where $\mathcal{S}_a = \{\lambda \in \mathbb{R}^K \mid \lambda_a > \theta\}$ and $\mathcal{W}_a = \{w \in \Delta_K \mid w_a \geq \frac{1-\varepsilon}{2\log(K)}\}$. Applying the useful corollary (c) of Theorem 1 in Wang et al. (2024b) yields that

$$\liminf_{T \rightarrow +\infty} \frac{1}{T} \log \frac{1}{\mathbb{P}_\nu(\hat{\mu}(T) \in \mathcal{S}_a, N(T)/T \in \mathcal{W}_a)} \geq \inf_{w \in \mathcal{W}_a} \inf_{\lambda \in \text{cl}(\mathcal{S}_a)} \Psi(\lambda, w) = \frac{1-\varepsilon}{4\log(K)} \Delta_a^2,$$

where the last equality is obtained by direct computation since $\Psi(\lambda, w) = \sum_{a \in \mathcal{A}} w_a (\mu_a - \lambda_a)^2/2$ and $\mu_a < \theta$. Combining the above inequalities and taking the limit when $\varepsilon \rightarrow 0$, we conclude that $\liminf_{T \rightarrow +\infty} \frac{1}{T} \log \frac{1}{\mathbb{P}_\nu(\hat{a}_T \neq \emptyset)} \geq \frac{\Delta_{\min}^2}{4\log(K)}$.

Case 2: $\mathcal{A}_\theta \neq \emptyset$. We re-use the arguments from the “Improved case 2” paragraph of the proof of Theorem 25. Let $\mathcal{C}_j$ be the set active arms at phase j and ℓ_j be the empirical worst arm at the end of phase j , i.e. $\mathcal{C}_{j+1} = \mathcal{C}_j \setminus \{j\}$. Similarly, we obtain that

$$\liminf_{T \rightarrow +\infty} \frac{1}{T} \log \frac{1}{\mathbb{P}_\nu(\hat{a}_T \notin \mathcal{A}_\theta)} \geq \min_{\substack{p: \mathcal{A}_\theta \rightarrow [K-1] \\ p \text{ injective}}} \liminf_{T \rightarrow +\infty} \frac{1}{T} \log \frac{1}{\mathbb{P}_\nu(\forall a \in \mathcal{A}_\theta, \ell_{p(a)} = a, \mathcal{C}_{p(a)} \setminus \mathcal{A}_\theta \neq \emptyset)}$$

where we used that there is a finite number of such injective mapping to swap the limit and the sum. Moreover, we have $\mathbb{P}_\nu(\forall a \in \mathcal{A}_\theta, \ell_{p(a)} = a, \mathcal{C}_{p(a)} \setminus \mathcal{A}_\theta \neq \emptyset) \leq \mathbb{P}_\nu(\ell_{p(a)} = a, \mathcal{C}_{p(a)} \setminus \mathcal{A}_\theta \neq \emptyset)$ for all $a \in \mathcal{A}_\theta$. Let $\mathcal{J}_a = \{(G, B) \subseteq \mathcal{A}_\theta \times \mathcal{A}_\theta^c \mid a \in G, B \neq \emptyset, |G \cup B| = K - p(a) + 1\}$. By

union bound, we obtain that

$$\begin{aligned} & \liminf_{T \rightarrow +\infty} \frac{1}{T} \log \frac{1}{\mathbb{P}_\nu(\ell_{p(a)} = a, \mathcal{C}_{p(a)} \setminus \mathcal{A}_\theta \neq \emptyset)} \\ & \geq \min_{(G,B) \in \mathcal{J}_a} \liminf_{T \rightarrow +\infty} \frac{1}{T} \log \frac{1}{\mathbb{P}_\nu(\ell_{p(a)} = a, \mathcal{C}_{p(a)} = G \cup B)}, \end{aligned}$$

where we used that $|\mathcal{J}_a| < +\infty$ to swap the limit and the sum. Recall that the active arms have been sampled n_k times at the end of phase k , where $n_k \geq \frac{T-K}{\log(K)(K+1-k)}$. For all $k \in [K-1]$, let $\alpha_k > 0$ such that the end of phase k corresponds to a time $\alpha_k T$ (assumed to be integer for simplicity). Let $\varepsilon > 0$ and T_ε such that $n_k/(\alpha_k T) \geq \frac{1-\varepsilon}{\alpha_k \log(K)(K+1-k)}$ for all $T \geq T_\varepsilon$ and all $k \in [K-1]$. Let $T \geq T_\varepsilon$. The event $\{\ell_{p(a)} = a, \mathcal{C}_{p(a)} \setminus \mathcal{A}_\theta \neq \emptyset\}$ implies that $\{\hat{\mu}(\alpha_{p(a)} T) \in \mathcal{S}_a, N(\alpha_{p(a)} T)/(\alpha_{p(a)} T) \in \mathcal{W}_a\}$ where $\mathcal{S}_a = \{\lambda \in \mathbb{R}^K \mid \lambda_a \leq \min_{b \in B} \lambda_b\}$ and $\mathcal{W}_a = \{w \in \Delta_K \mid \forall b \in B \cup \{a\}, w_b \geq \frac{1-\varepsilon}{\alpha_{p(a)} \log(K)(K+1-p(a))}\}$. Applying the useful corollary (c) of Theorem 1 in Wang et al. (2024b) yields that

$$\begin{aligned} & \liminf_{T \rightarrow +\infty} \frac{1}{\alpha_{p(a)} T} \log \frac{1}{\mathbb{P}_\nu(\hat{\mu}(\alpha_{p(a)} T) \in \mathcal{S}_a, N(\alpha_{p(a)} T)/(\alpha_{p(a)} T) \in \mathcal{W}_a)} \geq \inf_{w \in \mathcal{W}_a} \inf_{\lambda \in \text{cl}(\mathcal{S}_a)} \Psi(\lambda, w) \\ & = \frac{1-\varepsilon}{2\alpha_{p(a)} \log(K)(K+1-p(a))} \inf \left\{ \sum_{b \in B \cup \{a\}} (\mu_b - \lambda_b)^2 \mid \lambda \in \mathcal{S}_a \right\} \\ & \geq \frac{1-\varepsilon}{4\alpha_{p(a)} \log(K)(K+1-p(a))} \min_{b \in B} (\mu_a - \mu_b)^2, \end{aligned}$$

where we solved explicitly the infimum after using that $\sum_{c \in B \cup \{a\}} (\mu_c - \lambda_c)^2 \geq \sum_{c \in \{a,b\}} (\mu_c - \lambda_c)^2$ for all $b \in B$. Combining the above inequalities and taking the limit when $\varepsilon \rightarrow 0$, we conclude that

$$\liminf_{T \rightarrow +\infty} \frac{1}{T} \log \frac{1}{\mathbb{P}_\nu(\hat{a}_T \notin \mathcal{A}_\theta)} \geq \min_{\substack{p: \mathcal{A}_\theta \rightarrow [K-1] \\ p \text{ injective}}} \max_{a \in \mathcal{A}_\theta} \min_{b \notin \mathcal{A}_\theta} \frac{(\mu_a - \mu_b)^2}{4\log(K)(K+1-p(a))} = \frac{1}{4\log(K)G_2(\mu)}$$

■

Appendix D. Prior Knowledge-based GAI Algorithm (PKGAI)

In this section, we describe a meta-algorithm for fixed-budget GAI called PKGAI (**P**rior **K**nowledge-based GAI, shown in Algorithm 2). This meta-algorithm can be used to convert fixed-confidence GAI algorithms from prior works. As previously mentioned, the sampling rule in this algorithm depends on an index policy $(i_a(t))_{a \in \mathcal{A}, t \leq T}$. We provide guarantees on the error probability for both the partially specified algorithm (without a specific index policy, Theorem 27) and the uniform round-robin version (Theorem 28).

D.1 A Meta-algorithm for Fixed-budget GAI

Algorithm 2 PKGAI (Prior Knowledge-based GAI)

```

1: Input: budget  $T \geq K$ , threshold  $\theta$ 
2: Define: for all  $a \in \mathcal{A}$ , confidence intervals  $([\hat{\Delta}_a^-(t), \hat{\Delta}_a^+(t)])_{t \leq T}$  on  $\mu_a - \theta$ 
3: Define: for all  $a \in \mathcal{A}$  and  $t \leq T$ , sampling index  $i_a(t) : \mathcal{A} \times \mathbb{N} \rightarrow \mathbb{R}$ .
   Possible index policies:

           PKGAI(APTP) :  $i_a(t) := \sqrt{N_a(t)}(\hat{\mu}_a(t) - \theta)$ ,
           PKGAI(UCB) :  $i_a(t) := \hat{\Delta}_a^+(t)$ ,
           PKGAI(Unif) :  $i_a(t) := -N_a(t)$ ,
           PKGAI(LCB-G) :  $i_a(t) := \sqrt{N_a(t)}\hat{\Delta}_a^-(t)$ .

4: Sample each arm  $a \in \mathcal{A}$  once
5: Set  $t \leftarrow K$ ,  $\mathcal{S}_t \leftarrow \mathcal{A}$ ,  $N_a(t) \leftarrow 1$  and initialize  $\hat{\Delta}_a^-(t), \hat{\Delta}_a^+(t)$  for  $a \in \mathcal{A}$ 
6: while  $t < T$  and  $|\mathcal{S}_t| > 0$  do
7:    $a_{t+1} \in \arg \max_{a \in \mathcal{S}_t} i_a(t)$ 
8:   Draw arm  $a_{t+1}$  and observe  $X_{a_{t+1}, t+1}$ 
9:   Update  $\hat{\Delta}_a^-(t+1), \hat{\Delta}_a^+(t+1)$  for all  $a \in \mathcal{A}$ 
10:   $\mathcal{S}_{t+1} \leftarrow \mathcal{S}_t \setminus \{a \in \mathcal{S}_t \mid \hat{\Delta}_a^+(t+1) < 0\}$ 
11:   $t \leftarrow t + 1$ 
12: end while
13: end
14: if  $|\mathcal{S}_t| = 0$  or  $\max_{a \in \mathcal{S}_T} \hat{\Delta}_a^-(T) + \hat{\Delta}_a^+(T) \leq 0$  then
15:   return  $\hat{a}_T := \emptyset$ 
16: else
17:   return  $\hat{a}_T \in \arg \max_{a \in \mathcal{S}_T} \hat{\Delta}_a^-(T)$ 
18: end if
    
```

The meta-algorithm PKGAI –where the sampling index is unspecified– is shown in Algorithm 2. Similarly to fixed-confidence GAI algorithms proposed in the literature (Kano et al., 2019; Tabata et al., 2020), it relies on confidence bounds $([\hat{\Delta}_a^-(t), \hat{\Delta}_a^+(t)])_{t \leq T}$ on gap $\mu_a - \theta$ for any arm a and phased elimination (Line L.11) on the corresponding σ -sub-Gaussian distribution (in our paper, $\sigma = 1$) $[\hat{\Delta}_a^-(t), \hat{\Delta}_a^+(t)] := \left\{ \hat{\mu}_a(t) - \theta \pm \sigma \sqrt{\beta(t)/N_a(t)} \right\}$, where β is a well-chosen threshold function, which is increasing in its argument.

Intuitively, $\hat{\Delta}_a^-(t)$ (resp. $\hat{\Delta}_a^+(t)$) represents an lower (resp. upper) bound on the amount of information towards decision $\{a \in \mathcal{A}_\theta\}$. In the elimination step, all unsuitable candidates are removed at the end of the sampling round; that is, arms which corresponding upper confidence bound is below 0. We assume in the remainder of the section that the sampling budget T is at least equal to K .

Recommendation rule. This algorithm enables early stopping, as if there is no suitable candidate left (*i.e.* $\mathcal{S}_t = \emptyset$), then PKGAI returns the empty set (Line L.13). If there is no suitable candidate a such that $\hat{\Delta}_a^-(T) + \hat{\Delta}_a^+(T) > 0$, it also returns the empty set –when considering symmetrical confidence intervals, it is equivalent to testing whether $\hat{\mu}_a(t) > \theta$ (L.13). Otherwise, it returns one of the arms maximizing the lower confidence bound (L.16).

Sampling rule. As initialization, each arm $a \in \mathcal{A}$ is pulled once. PKGAI combines upper/lower confidence bounds-based sampling (Kano et al., 2019; Kaufmann et al., 2018), and exploitation-oriented approaches (Locatelli et al., 2016; Tabata et al., 2020). Several sampling rules, some inspired by prior fixed-confidence algorithms, are described in Algorithm 2. We also propose another exploration algorithm, named LCB-G, which targets the lower confidence bound. We denote PKGAI(*) the meta-algorithm where the sampling rule remains undefined.

Comparison with prior works. Note that, contrary to APGAI, this algorithm requires the knowledge of instance-dependent quantities to define the confidence bounds, and of T , thus not permitting continuation. This meta-algorithm is related to algorithms proposed in fixed-confidence variants of the GAI problem (e.g. BAEC (Tabata et al., 2020) for PKGAI(APT_P), HDoC and LUCB-G (Kano et al., 2019) for PKGAI(UCB)), albeit not entirely similar. To adapt to the fixed-budget constraint, Lines L.14 and L.16 are introduced, corresponding to cases where the allocated budget is probably too small to assess with certainty whether $\mathcal{A}_\theta = \emptyset$.

D.2 Fixed-budget Guarantees for PKGAI

Theorem 27 shows that for any sampling index (at Line L.7) and if we have access to $H_1(\mu)$ and $H_\theta(\mu)$ —which is quite a strong assumption in practice—using the structure as in PKGAI ensures that the error probability is upper bounded by roughly $\exp(-T/H_1(\mu))$ in all cases, which matches optimality when $\mathcal{A}_\theta = \emptyset$.

Theorem 27 (Proof in Section D.4) *Let $T > K$ and consider any 1-sub-Gaussian distribution with mean $\mu \in \mathbb{R}^K$ such that $\mu_a \neq \theta$ for all $a \in \mathcal{A}$. If confidence intervals $[\hat{\Delta}_a^-(t), \hat{\Delta}_a^+(t)]$ for all arm $a \in \mathcal{A}$ and $t \leq T$ are such that*

$$\mathbb{P}_\nu\left(\bigcup_{a \in \mathcal{A}, t \leq T} \{|\hat{\mu}_a(t) - \mu_a| \leq \sqrt{\beta(t)N_a(t)}\}\right) \in (0, 1), \text{ with } \beta(T) \leq \frac{T - K}{4H_1(\mu)}. \quad (15)$$

Then, we have $P_{\nu, \text{PKGAI}()}^{\text{err}}(T) \leq 2KT e^{-2\beta(T)}$. This is minimized when Inequality (15) is an equality, hence*

$$P_{\nu, \text{PKGAI}(*)}^{\text{err}}(T) \leq 2KT \exp\left(-\frac{T - K}{2H_1(\mu)}\right).$$

Furthermore, when considering an uniform round-robin sampling, i.e. PKGAI(Unif) (in Line L.7, Algorithm 2) $i_a(t) := -N_a(t)$ for all $a \in \mathcal{A}$ and $t \leq T$, the error probability is upper bounded by a term of order $\exp(-T/H_1(\mu))$ when $\mathcal{A}_\theta = \emptyset$ or $\hat{a}_T = \emptyset$, and of order $\exp(-T/(K\hat{\Delta}^{-2}))$ otherwise, where $\hat{\Delta} := \max_{a \in \mathcal{A}_\theta} \Delta_a + \min_{a \notin \mathcal{A}_\theta} \Delta_a$ (Theorem 28).

Theorem 28 (Proof in Section D.5) *Let $T > K$ and consider any 1-sub-Gaussian distribution with mean $\mu \in \mathbb{R}^K$ such that $\mu_a \neq \theta$ for all $a \in \mathcal{A}$. Let $\beta(T)$ satisfying*

$$\beta(T) \leq \begin{cases} (T - K)/(4K\hat{\Delta}^{-2}) & \text{if } \mathcal{A}_\theta(\mu) \neq \emptyset \\ (T - K)/(4H_1(\mu)) & \text{otherwise} \end{cases}. \quad (16)$$

where $\hat{\Delta} := \max_{a \in \mathcal{A}_\theta} \Delta_a + \min_{a \notin \mathcal{A}_\theta} \Delta_a$. Then $P_{\nu, \text{PKGAI}[U_{\text{unif}}]}^{\text{err}}(T) \leq 2KT e^{-2\beta(T)}$. This is minimized when Inequality (16) is an equality, hence

$$P_{\nu, \text{PKGAI}[U_{\text{unif}}]}^{\text{err}}(T) \leq \begin{cases} 2KT \exp\left(-\frac{T-K}{2K\hat{\Delta}^{-2}}\right) & \text{if } \mathcal{A}_\theta \neq \emptyset, \\ 2KT \exp\left(-\frac{T-K}{2H_1(\mu)}\right) & \text{otherwise.} \end{cases}$$

This theorem yields a strictly better bound than APGAI and Theorem 27 for instances such that $\mathcal{A}_\theta \neq \emptyset$ and

$$K\hat{\Delta}^{-2} = K \left(\max_{a \in \mathcal{A}_\theta} \Delta_a + \min_{b \notin \mathcal{A}_\theta} \Delta_b \right)^{-2} < H_1(\mu) := \sum_{a \in \mathcal{A}} \Delta_a^{-2},$$

e.g. in all but one instances among those we have considered (see Table 7).

D.3 Proof Sketch

The idea behind the proofs of Theorems 27 and 28 is to consider each recommendation case, and to determine a value of $\beta(T)$ which prevents an error in PKGAI when confidence intervals hold. As a consequence,

$$P_{\nu, \text{PKGAI}(\ast)}^{\text{err}}(T) \leq \mathbb{P}_\nu(\mathcal{E}_T^c) \text{ where } \mathcal{E}_T := \bigcap_{\substack{a \in \mathcal{A} \\ t \leq T}} \left\{ |\hat{\mu}_a(t) - \mu_a| \leq \sqrt{\frac{\beta(t)}{N_a(t)}} \right\}.$$

Let us denote the last round in PKGAI, for any sampling index $\tau := T \wedge \inf_{t \leq T} \{|\mathcal{S}_t| = 0\}$, i.e. the number of samples after which the recommendation rule is applied. The probability of error of any algorithm $\mathfrak{A}$ with the same structure as PKGAI can be decomposed as follows by union bound

$$\begin{aligned} P_{\nu, \mathfrak{A}}^{\text{err}}(T) &\leq \mathbb{P}[(\mathcal{A}_\theta \neq \emptyset \cap (\hat{a}_\tau \in \{\emptyset\} \cup \mathcal{A} \setminus \mathcal{A}_\theta) \cap \mathcal{E}_T) \cup (\mathcal{A}_\theta = \emptyset \cap \hat{a}_\tau \neq \emptyset \cap \mathcal{E}_T)] + \mathbb{P}_\nu(\mathcal{E}_T^c), \\ &\leq \underbrace{\mathbb{P}[\mathcal{A}_\theta \neq \emptyset \cap (\hat{a}_\tau \in \{\emptyset\} \cup \mathcal{A} \setminus \mathcal{A}_\theta) \cap \mathcal{E}_T]}_{\text{Case 1}} + \underbrace{\mathbb{P}[\mathcal{A}_\theta = \emptyset \cap \hat{a}_\tau \neq \emptyset \cap \mathcal{E}_T]}_{\text{Case 2}} + \mathbb{P}_\nu(\mathcal{E}_T^c). \end{aligned}$$

For both Theorems 27 and 28, we will then proceed by considering two cases, $\mathcal{A}_\theta = \emptyset$ and $\mathcal{A}_\theta \neq \emptyset$, assuming that $\mathcal{E}_T$ holds. In both cases, the goal is to determine the form of appropriate confidence intervals which prevent an error in PKGAI when $\mathcal{E}_T$ holds (by proving a contradiction), such that ultimately, $P_{\nu, \text{PKGAI}}^{\text{err}}(T) \leq \mathbb{P}_\nu(\mathcal{E}_T^c)$.

D.4 Proof of Theorem 27

D.4.1 CASE $\mathcal{A}_\theta(\mu) = \mathcal{A}_\theta = \emptyset$

Proof Let $\nu \in \mathcal{D}^K$ be any instance of mean vector μ such that $\mathcal{A}_\theta(\mu) = \emptyset$. Let us denote $\mathcal{E}_T^{\text{Case 1}} := \{\mathcal{E}_T \cap \mathcal{A}_\theta = \emptyset\}$. The error probability $\mathbb{P}[\mathcal{E}_T^{\text{Case 1}} \cap \hat{a}_\tau \neq \emptyset]$ is lesser than

$$\mathbb{P}\left[\mathcal{E}_T^{\text{Case 1}} \cap \exists a \in \mathcal{A}, \hat{\Delta}_a^+(\tau) + \hat{\Delta}_a^-(\tau) \geq 0\right] \text{ (Line L.13).}$$

Since $\mathcal{S}_\tau \neq \emptyset$ (otherwise, $\hat{a}_T = \emptyset$), then necessarily $\tau = T$. Here, the contradiction will involve the number of samples drawn from each arm during the sampling phase. For any arm $b \in \mathcal{S}_T \subseteq \mathcal{A}_\theta^c$, on $\mathcal{E}_T$

$$\hat{\Delta}_b^+(T) \geq 0 \implies -\Delta_b + 2\sqrt{\frac{\beta(T)}{N_b(T)}} \geq 0 \implies N_b(T) \leq \frac{4\beta(T)}{\Delta_b^2} < \frac{4\beta(T)}{\Delta_b^2} + 1. \quad (17)$$

Moreover, for any arm $c \in \mathcal{S}_T^c \subseteq \mathcal{A}_\theta^c$, it means that c has been eliminated after exactly $K + 1 \leq t_c \leq T$ rounds, and is no longer sampled after round t_c (i.e. $N_c(T) = N_c(t_c)$). By a reasoning similar to the one that led to Inequality (17) on round $t_c - 1$,

$$\begin{aligned} \hat{\Delta}_c^+(t_c - 1) \geq 0 > \hat{\Delta}_c^+(t_c) &\implies N_c(T) - 1 = N_c(t_c - 1) \leq \frac{4\beta(t_c - 1)}{\Delta_c^2} \leq \frac{4\beta(T)}{\Delta_c^2} \\ &\implies N_c(T) \leq \frac{4\beta(T)}{\Delta_c^2} + 1. \end{aligned} \quad (18)$$

Combining inequalities 17 and 18, since $\mathcal{S}_T \neq \emptyset$, $T = \sum_{k \in \mathcal{A}} N_k(T) < \sum_{a \in \mathcal{A}} \left(\frac{4\beta(T)}{\Delta_a^2} + 1 \right) \leq 4H_1(\mu)\beta(T) + K$.

That is, any choice of β such that $\beta(T) \leq (T - K)/(4H_1(\mu))$ automatically yields a contradiction. Then $\mathbb{P} \left[\mathcal{E}_T^{\text{Case 1}} \cap \exists a \in \mathcal{A}, \hat{\Delta}_a^+(\tau) + \hat{\Delta}_a^-(\tau) \geq 0 \right] = 0$. $\blacksquare$

D.4.2 CASE $\mathcal{A}_\theta(\mu) = \mathcal{A}_\theta \neq \emptyset$

Proof Now, we consider any instance $\nu \in \mathcal{D}^K$ of mean vector μ such that $\mathcal{A}_\theta(\mu) = \emptyset$. Let us denote $\mathcal{E}_T^{\text{Case 2}} := \mathcal{E}_T \cap (\mathcal{A}_\theta \neq \emptyset)$. The error probability of PKGAI when $\mathcal{A}_\theta \neq \emptyset$ can be decomposed as follows

$$\mathbb{P} \left[\mathcal{E}_T^{\text{Case 2}} \cap (\hat{a}_\tau \in \{\emptyset\} \cup \mathcal{A} \setminus \mathcal{A}_\theta) \right] = \underbrace{\mathbb{P} \left[\mathcal{E}_T^{\text{Case 2}} \cap \hat{a}_\tau = \emptyset \right]}_{\text{Case 2.1 (L.14 in Algorithm 2)}} + \underbrace{\mathbb{P} \left[\mathcal{E}_T^{\text{Case 2}} \cap \hat{a}_\tau \in \mathcal{A} \setminus \mathcal{A}_\theta \right]}_{\text{Case 2.2 (L.16)}}.$$

Case 2.1. Necessarily, either $\mathcal{S}_\tau = \emptyset$ or $\max_{a \in \mathcal{S}_\tau} \hat{\Delta}_a^-(\tau) + \hat{\Delta}_a^+(\tau) \leq 0$ (L.13).

- If $\mathcal{S}_\tau = \emptyset$, then it means in particular that for any good arm $a \in \mathcal{A}_\theta$, if $\mathcal{E}_T$ holds, then

$$\exists t_a < \tau, \hat{\Delta}_a^+(t_a) < 0 \implies (\mu_a - \theta) = \Delta_a < 0,$$

which contradicts $a \in \mathcal{A}_\theta$. Then, good arms cannot be eliminated at any round on event $\mathcal{E}_T$, that is, $\mathbb{P} \left[\mathcal{E}_T^{\text{Case 2}} \cap \hat{a}_T = \emptyset \cap \mathcal{S}_\tau \neq \emptyset \right] = 0$.

- In that case, $\tau = T$. If $\max_{a \in \mathcal{S}_T} \hat{\Delta}_a^-(T) + \hat{\Delta}_a^+(T) \leq 0$ on event $\mathcal{E}_T$, then since $\mathcal{E}_T$ holds, for all $a \in \mathcal{A}_\theta \subseteq \mathcal{S}_T$

$$2 \left(\Delta_a - \sqrt{\frac{\beta(T)}{N_a(T)}} \right) \leq \hat{\Delta}_a^-(T) + \hat{\Delta}_a^+(T) \leq 0 \implies N_a(T) \leq \frac{\beta(T)}{\Delta_a^2} < \frac{\beta(T)}{\Delta_a^2} + 1. \quad (19)$$

Furthermore, as a direct consequence of Inequalities 17 and 18, for any $b \notin \mathcal{A}_\theta$, $N_b(T) \leq \frac{4\beta(T)}{\Delta_b^2} + 1$. From these upper bounds on the number of samples drawn from each arm, we

can again build a contradiction

$$T = \sum_{a \in \mathcal{A}} N_a(T) < \beta(T) (H_\theta(\mu) + 4(H_1(\mu) - H_\theta(\mu))) + K = \beta(T) (4H_1(\mu) - 3H_\theta(\mu)) + K.$$

That is, any choice of β such that $\beta(T) \leq \frac{1}{4}(T - K)/(H_1(\mu) - \frac{3}{4}H_\theta(\mu))$ automatically yields a contradiction. In that case, $\mathbb{P}[\mathcal{E}_T^{\text{Case 2}} \cap \hat{a}_T = \emptyset] = 0$.

Case 2.2. The only remaining case is when $\tau = T$ (Line L.16). On event $\mathcal{E}_T$, since $\mathcal{A}_\theta \subseteq \mathcal{S}_T$, for all $a \in \mathcal{A}_\theta$,

$$0 >_{\hat{a}_T \notin \mathcal{A}_\theta} -\Delta_{\hat{a}_T} \geq \hat{\Delta}_{\hat{a}_T}^-(T) \geq_{\text{Line L.16}} \hat{\Delta}_a^-(T) \geq \Delta_a - 2\sqrt{\frac{\beta(T)}{N_a(T)}} \implies N_a(T) < 4\beta(T)\Delta_a^{-2}.$$

Furthermore, as Inequalities 17 and 18 hold, for any $b \notin \mathcal{A}_\theta$, $N_b(T) \leq 4\beta(T)\Delta_b^{-2} + 1$. All in all, $T < 4H_1(\mu)\beta(T) + K - |\mathcal{A}_\theta|$. That is, any choice of β such that $\beta(T) \leq \frac{T-K+|\mathcal{A}_\theta|}{4H_1(\mu)}$ automatically yields a contradiction. In that case, $\mathbb{P}[\mathcal{E}_T^{\text{Case 2}} \cap \hat{a}_T \in \mathcal{A} \setminus \mathcal{A}_\theta] = 0$. ■

D.4.3 FINAL STEP

Combining all previous cases, it suffices to consider β such that $\beta(T) \leq \frac{T-K}{4H_1(\mu)}$, to obtain the following upper bound on the error probability from Inequality (19), using successively the Hoeffding concentration bounds and union bounds over $\mathcal{A}$ of size K and over $\{1, 2, \dots, T\}$, $P_{\nu, \text{PKGAI}^*}^{\text{err}}(T) \leq 2KT \exp(-2\beta(T))$. In particular, the right-hand term is minimized for $\beta(T) = \frac{T-K}{4H_1(\mu)}$, and in that case $P_{\nu, \text{PKGAI}^*}^{\text{err}}(T) \leq 2KT \exp\left(-\frac{T-K}{2H_1(\mu)}\right)$. ■

D.5 Proof of Theorem 28

D.5.1 CASE $\mathcal{A}_\theta(\mu) = \mathcal{A}_\theta = \emptyset$

Proof Since PKGAI(Unif) belongs to the family of PKGAI algorithms, then Theorem 27 applies, and conditioned on the fact that $\beta(T) \leq (T - K)/(4H_1(\mu))$, the upper bound on the error probability for any instance $\nu \in \mathcal{D}^K$ in that case is $P_{\nu, \text{PKGAI}(\text{Unif})}^{\text{err}}(T) \leq 2KT \exp(-2\beta(T))$, and is minimized when the previous inequality on $\beta(T)$ is an equality. ■

D.5.2 CASE $\mathcal{A}_\theta(\mu) \neq \emptyset$ AND $\hat{a}_T = \emptyset$

Proof However, when $\mathcal{A}_\theta \neq \emptyset$, we will take into account the sampling rule in order to find a tighter upper bound on the probability $\mathbb{P}[\mathcal{E}_T^{\text{Case 2}} \cap \hat{a}_T = \emptyset]$. Then, necessarily, according to Case 2.1 in the proof of Theorem 27

$$\mathbb{P}[\mathcal{E}_T^{\text{Case 2}} \cap \hat{a}_T = \emptyset] = \mathbb{P}\left[\mathcal{E}_T^{\text{Case 2}} \cap \max_{a \in \mathcal{S}_T} \hat{\Delta}_a^-(T) + \hat{\Delta}_a^+(T) \leq 0\right].$$

$\mathcal{A}_\theta \subseteq \mathcal{S}_T$ (otherwise, we end up with a contradiction with event $\mathcal{E}_T$). Moreover, if $\max_{a \in \mathcal{S}_T} \hat{\Delta}_a^-(T) + \hat{\Delta}_a^+(T) \leq 0$, then Inequality (19) applies. Finally, since PKGAI(Unif)

uses a uniform sampling, $N_a(T) \geq \lfloor \frac{T}{K} \rfloor \geq \frac{T}{K} - 1$ for any arm a . Combining all of this yields the following inequalities

$$\forall a \in \mathcal{A}_\theta, \quad \frac{T}{K} - 1 \leq N_a(T) < \frac{\beta(T)}{\Delta_a^2} + 1 \implies \beta(T) > \frac{T - 2K}{K} \max_{a \in \mathcal{A}_\theta} \Delta_a^2 = \frac{T - 2K}{K (\max_{a \in \mathcal{A}_\theta} \Delta_a)^{-2}}.$$

Then any choice of β such that $\beta(T) \leq (T - 2K)/(K(\max_{a \in \mathcal{A}_\theta} \Delta_a)^{-2})$ would lead to a contradiction. ■

D.5.3 CASE $\mathcal{A}_\theta \neq \emptyset$ AND $\hat{a}_T \neq \emptyset$

Proof Let us find a tighter upper bound on the error probability $\mathbb{P}[\mathcal{E}_T^{\text{Case 2}} \cap \hat{a}_T \in \mathcal{A} \setminus \mathcal{A}_\theta]$. This necessarily implies that the recommendation rule at Line L.16 is fired ($\tau = T$) and that the algorithm makes a mistake ($\hat{a}_T \notin \mathcal{A}_\theta$). On event $\mathcal{E}_T$, there exists $b \notin \mathcal{A}_\theta$, for all $a \in \mathcal{A}_\theta$,

$$\begin{aligned} -\Delta_b \geq_{b \notin \mathcal{A}_\theta} \hat{\Delta}_b^-(T) &\geq \hat{\Delta}_a^-(T) \geq_{a \in \mathcal{A}_\theta} \Delta_a - 2\sqrt{\frac{\beta(T)}{N_a(T)}} \geq \Delta_a - 2\sqrt{\frac{\beta(T)}{\min_{c \in \mathcal{A}_\theta} N_c(T)}} \\ \implies \max_{b \notin \mathcal{A}_\theta} (-\Delta_b) &\geq \max_{a \in \mathcal{A}_\theta} \Delta_a - 2\sqrt{\frac{\beta(T)}{\min_{c \in \mathcal{A}_\theta} N_c(T)}} \end{aligned}$$

Reordering terms and since PKGAI(Unif) uses a uniform sampling

$$2\sqrt{\frac{\beta(T)}{T/K - 1}} \geq \hat{\Delta} := \max_{a \in \mathcal{A}_\theta} \Delta_a + \min_{b \notin \mathcal{A}_\theta} \Delta_b \implies \beta(T) \geq \frac{T - K}{4K\hat{\Delta}^{-2}}.$$

Then any choice of β such that $\beta(T) < \frac{T - K}{4K\hat{\Delta}^{-2}}$ would lead to a contradiction.

D.5.4 FINAL STEP

All in all, similarly to the proof of Theorem 27, if the following inequality is satisfied for $\nu \in \mathcal{D}^K$ of mean vector μ

$$\beta(T) \leq W_\mu(T) := \begin{cases} (T - K)/(4H_1(\mu)) & \text{if } \mathcal{A}_\theta(\mu) = \emptyset \\ (T - K)/(4K\hat{\Delta}^{-2}) & \text{otherwise} \end{cases},$$

where $\hat{\Delta} := \max_{a \in \mathcal{A}_\theta} \Delta_a + \min_{b \notin \mathcal{A}_\theta} \Delta_b$, then we end with the following upper bound on the error probability $P_{\nu, \text{PKGAI}(\text{Unif})}^{\text{err}}(T) \leq 2KT \exp(-2\beta(T))$, which is minimized when the inequalities on $\beta(T)$ are equalities. ■

Appendix E. Lower Bounds for GAI and Generalized Likelihood Ratio

In Appendix E.1, we prove Lemma 1 which is an asymptotic lower bound on the expected sample complexity of a fixed-confidence GAI algorithm. In Appendix E.2, we present the generalized likelihood ratios for GAI, which relate to the APT_P index policy and the GLR stopping rule Eq. (6). In Appendix E.3, we prove lower bounds showing that a linear dependence in K is actually unavoidable, even when there is a unique good arm: Theorem 5 (Appendix E.3.1), Corollary 6 (Appendix E.3.2) and Corollary 9 (Appendix E.3.3).

E.1 Asymptotic Lower Bound for GAI in Fixed Confidence Setting

Lemma 1 gives an asymptotic lower bound on the expected sample complexity in fixed-confidence GAI, and relies on the well-known change of measure inequality (Lemma 1 from Kaufmann et al. (2016)).

Lemma 29 (Lemma 1) *Let $\delta \in (0, 1)$. For all δ -correct algorithm and all Gaussian instances $\nu_a = \mathcal{N}(\mu_a, 1)$, with $\mu_a \neq \theta$, $\liminf_{\delta \rightarrow 0} \mathbb{E}_\nu[\tau_\delta] / \log(1/\delta) \geq T^*(\mu)$, where*

$$T^*(\mu) := \begin{cases} 2 \min_{a \in \mathcal{A}_\theta(\mu)} \Delta_a^{-2} & \text{if } \mathcal{A}_\theta(\mu) \neq \emptyset, \\ 2H_1(\mu) & \text{otherwise.} \end{cases}$$

Proof Let $\delta \in (0, 1)$. Let us consider any Gaussian instance $\nu_a = \mathcal{N}(\mu_a, 1)$, where $\mu_a \neq \theta$. We define the following sets of alternative instances, depending on $\mathcal{A}_\theta(\mu)$

$$\text{Alt}(\mu) := \begin{cases} \{\lambda \in \mathbb{R}^K \mid \exists a \in \mathcal{A}, \lambda_a \geq \theta\} = \bigcup_{a \in \mathcal{A}} \{\lambda \in \mathbb{R}^K \mid \lambda_a \geq \theta\} & \text{if } \mathcal{A}_\theta(\mu) = \emptyset, \\ \bigcap_{a \in \mathcal{A}_\theta(\mu)} \{\lambda \in \mathbb{R}^K \mid \lambda_a < \theta\} & \text{otherwise.} \end{cases}$$

Let us call kl the binary relative entropy. Let us consider any δ -correct algorithm. Combining Lemma 1 from Kaufmann et al. (2016) with the δ -correctness of the algorithm and the monotonicity of function kl , for any 1-Gaussian distribution κ of mean $\lambda \in \text{Alt}(\mu)$

$$\frac{1}{2} \sum_{a \in \mathcal{A}} \mathbb{E}_\nu[N_a(\tau_\delta)] (\mu_a - \lambda_a)^2 \geq \text{kl}(P_{\nu, \mathfrak{A}}^{\text{err}}(\tau_\delta), P_{\kappa, \mathfrak{A}}^{\text{err}}(\tau_\delta)) \geq \text{kl}(1 - \delta, \delta) \geq \log(1/(2.4\delta)).$$

As it holds for any alternative instance κ , if $\Delta_K := \{p \in [0, 1]^K \mid \sum_i p_i = 1\}$, it yields that

$$\mathbb{E}_\nu[\tau_\delta] = \sum_{a \in \mathcal{A}} \mathbb{E}_\nu[N_a(\tau_\delta)] \geq 2 \underbrace{\left(\sup_{\omega \in \Delta_K} \inf_{\lambda \in \text{Alt}(\mu)} \sum_{a \in \mathcal{A}} \omega_a (\mu_a - \lambda_a)^2 \right)^{-1}}_{=T^*(\mu)} \log(1/(2.4\delta)).$$

If $\mathcal{A}_\theta(\mu) = \emptyset$, then using the definition of $\text{Alt}(\mu)$ in that case and since $\Delta_a := |\mu_a - \theta|$

$$\sup_{\omega \in \Delta_K} \inf_{\lambda \in \text{Alt}(\mu)} \sum_{a \in \mathcal{A}} \omega_a (\mu_a - \lambda_a)^2 = \sup_{\omega \in \Delta_K} \min_{a \in \mathcal{A}} \omega_a (\mu_a - \theta)^2 = \sup_{\omega \in \Delta_K} \min_{a \in \mathcal{A}} \omega_a \Delta_a^2 = \left(\sum_{a \in \mathcal{A}} \Delta_a^{-2} \right)^{-1},$$

and $\omega_a := \frac{\Delta_a^{-2}}{\sum_{b \in \mathcal{A}} \Delta_b^{-2}}$. Otherwise, $\mathcal{A}_\theta(\mu) \neq \emptyset$, and then

$$\sup_{\omega \in \Delta_K} \inf_{\lambda \in \text{Alt}(\mu)} \sum_{a \in \mathcal{A}} \omega_a (\mu_a - \lambda_a)^2 = \sup_{\omega \in \Delta_K} \sum_{a \in \mathcal{A}_\theta(\mu)} \omega_a (\mu_a - \theta)^2 = \max_{a \in \mathcal{A}_\theta} \Delta_a^2,$$

and $\omega_a := \mathbb{1}(a = \arg \max_{a \in \mathcal{A}_\theta} \mu_a)$. This concludes the proof for $T^*(\mu)$ as in Eq. (2). $\blacksquare$

E.2 Generalized Likelihood Ratio (GLR)

While we consider 1-sub-Gaussian distributions $\nu \in \mathcal{D}^K$ with mean μ in all generality, the ATP_P index and the GLR stopping rule stem from generalized likelihood ratios for Gaussian distributions with unit variance. In the following, we consider Gaussian distributions $\nu_a = \mathcal{N}(\mu_a, 1)$ which are uniquely characterized by their mean parameter μ_a .

The generalized log-likelihood ratio between the whole model space $\mathcal{M}$ and a subset $\Lambda \subseteq \mathcal{M}$ is $\text{GLR}_t^{\mathcal{M}}(\Lambda) = \log \frac{\sup_{\tilde{\mu} \in \mathcal{M}} \mathcal{L}_{\tilde{\mu}}(X_1, \dots, X_t)}{\sup_{\lambda \in \Lambda} \mathcal{L}_{\lambda}(X_1, \dots, X_t)}$. In the case of independent Gaussian distributions with unit variance, the likelihood ratio for two models with mean vectors $\xi, \lambda \in \mathcal{M}$,

$$\log \frac{\mathcal{L}_{\xi}(X_1, \dots, X_t)}{\mathcal{L}_{\lambda}(X_1, \dots, X_t)} = \frac{1}{2} \sum_{a \in \mathcal{A}} N_a(t) ((\hat{\mu}_a(t) - \lambda_a)^2 - (\hat{\mu}_a(t) - \xi_a)^2).$$

When $\hat{\mu}(t) \in \mathcal{M}$, the maximum likelihood estimator $\tilde{\mu}(t)$ coincide with the empirical mean, otherwise it is $\tilde{\mu}(t) = \arg \min_{\lambda \in \mathcal{M}} \sum_{a \in \mathcal{A}} N_a(t) (\hat{\mu}_a(t) - \lambda_a)^2$. In the following, we consider the case where $\hat{\mu}(t) \in \mathcal{M}$. The GLR for set Λ is $\text{GLR}_t^{\mathcal{M}}(\Lambda) = \frac{1}{2} \min_{\lambda \in \Lambda} \sum_{a \in \mathcal{A}} N_a(t) (\hat{\mu}_a(t) - \lambda_a)^2$.

When $\max_{a \in \mathcal{A}} \hat{\mu}_a(t) \leq \theta$, the recommendation is $\hat{a}_t = \emptyset$. Therefore, the set of alternative parameters (*i.e.* admitting a different recommendation) is $\text{Alt}(\hat{\mu}(t)) = \bigcup_{a \in \mathcal{A}} \{\lambda \in \mathbb{R}^K \mid \lambda_a > \theta\}$. By direct manipulations similar to the ones in Appendix E.1, the corresponding GLR can be written as

$$2\text{GLR}_t^{\mathcal{M}}(\text{Alt}(\hat{\mu}(t))) = \min_{a \in \mathcal{A}} N_a(t) (\theta - \hat{\mu}_a(t))^2 = (\min_{a \in \mathcal{A}} W_a^-(t))^2.$$

When $\max_{a \in \mathcal{A}} \hat{\mu}_a(t) > \theta$, the recommendation is $\hat{a}_t \in \mathcal{A}_\theta(\hat{\mu}(t))$. For each possible answer $a \in \mathcal{A}_\theta(\hat{\mu}(t))$, the set of alternative parameters (*i.e.* admitting a different recommendation) is $\text{Alt}(\hat{\mu}(t), a) = \{\lambda \in \mathbb{R}^K \mid \lambda_a \leq \theta\}$. By direct manipulations similar to the ones in Appendix E.1, the corresponding GLR can be written as

$$\forall a \in \mathcal{A}_\theta(\hat{\mu}(t)), \quad 2\text{GLR}_t^{\mathcal{M}}(\text{Alt}(\hat{\mu}(t), a)) = N_a(t) (\hat{\mu}_a(t) - \theta)^2 = W_a^+(t)^2.$$

E.3 Lower Bounds with Dependence on the Number of Arms

E.3.1 PROOF OF THEOREM 5

All arms are Gaussian with variance 1. These are instances such that $\mathcal{A}_\theta(\nu^{(a)}) = \{a\}$. Let $\mathbb{P}_\nu^\tau$ be the restriction of $\mathbb{P}_\nu$ to the σ -algebra generated by τ . For any τ -measurable event E (*e.g.*, $\{N_b(\tau) > n\}$), we have $\mathbb{P}_\nu^\tau(E) = \mathbb{P}_\nu(E)$.

A bandit model by specifying the law of each successive reward from each arm: the first rewards queried from arm a will have a given distribution, then the second reward will have a (possibly different) distribution, etc. The sequence of distributions is an array of reward laws. In true bandit models, the distribution is stationary, *i.e.* it does not change. However, for the construction of the lower bound, we will use arrays where the distribution changes after some number n , *i.e.* non-stationary distribution. For all $n \in \mathbb{N}$ and $(a, b) \in [K]^2$ with $a \neq b$, we write $\eta_{a \rightarrow b}^n$ for the following array of reward laws:

- For $k \notin \{a, b\}$, $\eta_{a \rightarrow b, k}^n$ is constant equal to $\mathcal{N}(\theta - \varepsilon, 1)$.
- $\eta_{a \rightarrow b, a}^n$ is constant equal to $\mathcal{N}(\theta + \Delta, 1)$.
- For the first n rewards, $\eta_{a \rightarrow b, b}^n$ is $\mathcal{N}(\theta - \varepsilon, 1)$. For the next rewards, $\eta_{a \rightarrow b, b}^n$ is $\mathcal{N}(\theta + \Delta, 1)$.

Since TV is symmetric and satisfies the triangle inequality, we have

$$\text{TV}(\mathbb{P}_{\nu^{(a)}}^\tau, \mathbb{P}_{\nu^{(b)}}^\tau) \leq \text{TV}(\mathbb{P}_{\nu^{(a)}}^\tau, \mathbb{P}_{\eta_{a \rightarrow b}^n}^\tau) + \text{TV}(\mathbb{P}_{\nu^{(b)}}^\tau, \mathbb{P}_{\eta_{b \rightarrow a}^n}^\tau) + \text{TV}(\mathbb{P}_{\eta_{a \rightarrow b}^n}^\tau, \mathbb{P}_{\eta_{b \rightarrow a}^n}^\tau).$$

Using Pinsker's inequality and the data-processing inequality, we obtain

$$\text{TV}(\mathbb{P}_{\eta_{a \rightarrow b}^n}^\tau, \mathbb{P}_{\eta_{b \rightarrow a}^n}^\tau) \leq \sqrt{\text{KL}(\mathbb{P}_{\eta_{a \rightarrow b}^n}^\tau, \mathbb{P}_{\eta_{b \rightarrow a}^n}^\tau)/2} \leq \sqrt{\text{KL}(\mathbb{P}_{\eta_{a \rightarrow b}^n}^\tau, \mathbb{P}_{\eta_{b \rightarrow a}^n}^\tau)/2} = \sqrt{n(\Delta + \varepsilon)^2/2}.$$

An application of the general property that conditioning increases f -divergences yields Lemma 30, proved in Lemma C.4 in Poiani et al. (2025).

Lemma 30 (Lemma C.4 in Poiani et al. (2025))

$$\forall n \in \mathbb{N}, \forall a \in [K], \forall b \in [K] \setminus \{a\}, \quad \text{TV}(\mathbb{P}_{\nu^{(a)}}^\tau, \mathbb{P}_{\eta_{a \rightarrow b}^n}^\tau) \leq \mathbb{P}_{\nu^{(a)}}(N_b(\tau) > n).$$

Combining the above inequalities with Lemma 30 yields

$$\text{TV}(\mathbb{P}_{\nu^{(a)}}^\tau, \mathbb{P}_{\nu^{(b)}}^\tau) \leq \mathbb{P}_{\nu^{(a)}}(N_b(\tau) > n) + \mathbb{P}_{\nu^{(b)}}(N_a(\tau) > n) + \sqrt{n(\Delta + \varepsilon)^2/2},$$

which is exactly Lemma C.6 in Poiani et al. (2025). Summing these inequalities over $a \in [K]$ and $b \in [K] \setminus \{a\}$, we obtain

$$\begin{aligned} & \sum_{a \in [K], b \neq a} \text{TV}(\mathbb{P}_{\nu^{(a)}}^\tau, \mathbb{P}_{\nu^{(b)}}^\tau) - K(K-1)\sqrt{n(\Delta + \varepsilon)^2/2} \\ & \leq \sum_{a \in [K], b \neq a} (\mathbb{P}_{\nu^{(a)}}(N_b(\tau) > n) + \mathbb{P}_{\nu^{(b)}}(N_a(\tau) > n)) \leq \frac{2}{n} \sum_{a \in [K]} \mathbb{E}_{\nu^{(a)}}[\tau - N_a(\tau)]. \end{aligned}$$

where the second inequality uses $\mathbb{E}_{\nu^{(a)}}[\tau - N_a(\tau)] = \sum_{b \neq a} \mathbb{E}_{\nu^{(a)}}[N_b(\tau)]$ and Markov's inequality, i.e., $\mathbb{P}_{\nu^{(a)}}(N_b(\tau) > n) \leq \mathbb{E}_{\nu^{(a)}}[N_b(\tau)]/n$ for all $a \neq b$. Summing the inequalities obtained by assumption on the stopping time τ_δ and re-ordering, we obtain

$$\frac{1}{K} \sum_{a \in [K]} \mathbb{E}_{\nu^{(a)}}[\tau_\delta - N_a(\tau_\delta)] \geq \frac{n(K-1)}{2} \left(1 - 2\delta - \sqrt{n(\Delta + \varepsilon)^2/2}\right).$$

Taking $n = \frac{2}{(\Delta + \varepsilon)^2} \left(\frac{1-2\delta}{2}\right)^2$ concludes the proof since

$$\frac{1}{K} \sum_{a \in [K]} \mathbb{E}_{\nu^{(a)}}[\tau_\delta - N_a(\tau_\delta)] \geq \frac{K-1}{(\Delta + \varepsilon)^2} \left(\frac{1}{2} - \delta\right)^3 \geq \frac{K-1}{64(\Delta + \varepsilon)^2} \left(\frac{1}{2} - \delta_0\right)^3,$$

where the last inequality uses that $\delta \rightarrow \left(\frac{1}{2} - \delta\right)^3$ is decreasing on $(0, 1/4]$ and $\delta \leq 1/4$.

E.3.2 PROOF OF COROLLARY 6

Let $(\theta, \Delta, \varepsilon) \in \mathbb{R} \times (\mathbb{R}_+^*)^2$ and $(\nu^{(a)})_{a \in [K]}$ as in Theorem 5. All arms are Gaussian with variance 1. These are instances such that $\mathcal{A}_\theta(\nu^{(a)}) = \{a\}$. Let $\delta \in (0, 1/4]$. Let $\tau_{U,\delta}$ be the unverifiable sample complexity of a given strategy. For all $a \in [K]$ and all $b \in [K] \setminus \{a\}$,

$$\mathbb{P}_{\nu^{(a)}}(\exists t \geq \tau_{U,\delta}, \hat{a}_t \neq a) \leq \delta \quad \text{and} \quad \mathbb{P}_{\nu^{(b)}}(\forall t \geq \tau_{U,\delta}, \hat{a}_t = b) \geq 1 - \delta.$$

For any $\tau_{U,\delta}$ -measurable event E , we have $\mathbb{P}_{\nu^{(a)}}^{\tau_{U,\delta}}(E) = \mathbb{P}_{\nu^{(a)}}(E)$. Since $\{\hat{a}_{\tau_{U,\delta}} \neq a\}$ is $\tau_{U,\delta}$ -measurable and satisfies that

$$\{\hat{a}_{\tau_{U,\delta}} \neq a\} \subseteq \{\exists t \geq \tau_{U,\delta}, \hat{a}_t \neq a\} \quad \text{and} \quad \{\forall t \geq \tau_{U,\delta}, \hat{a}_t = b\} \subseteq \{\hat{a}_{\tau_{U,\delta}} \neq a\},$$

we obtain

$$\text{TV}(\mathbb{P}_{\nu^{(a)}}^{\tau_{U,\delta}}, \mathbb{P}_{\nu^{(b)}}^{\tau_{U,\delta}}) \geq \mathbb{P}_{\nu^{(b)}}(\hat{a}_{\tau_{U,\delta}} \neq a) - \mathbb{P}_{\nu^{(a)}}(\hat{a}_{\tau_{U,\delta}} \neq a) \geq 1 - 2\delta.$$

Applying Theorem 5 concludes the proof since

$$\max_{a \in [K]} \mathbb{E}_{\nu^{(a)}}[\tau_{U,\delta} - N_a(\tau_{U,\delta})] \geq \frac{1}{K} \sum_{a \in [K]} \mathbb{E}_{\nu^{(a)}}[\tau_{U,\delta} - N_a(\tau_{U,\delta})] \geq \frac{K-1}{64(\Delta + \varepsilon)^2}.$$

E.3.3 PROOF OF COROLLARY 9

Let $(\theta, \Delta, \varepsilon) \in \mathbb{R} \times (\mathbb{R}_+^*)^2$ and $(\nu^{(a)})_{a \in [K]}$ as in Theorem 5. All arms are Gaussian with variance 1. These are instances such that $\mathcal{A}_\theta(\nu^{(a)}) = \{a\}$. Let $\delta \in (0, 1/4]$. Let τ_δ be the sample complexity of a δ -correct strategy. For all $a \in [K]$ and all $b \in [K] \setminus \{a\}$,

$$\mathbb{P}_{\nu^{(a)}}(\hat{a}_{\tau_\delta} = a) \geq 1 - \delta \quad \text{and} \quad \mathbb{P}_{\nu^{(b)}}(\hat{a}_{\tau_\delta} \neq b) \leq \delta.$$

For any τ_δ -measurable event E , we have $\mathbb{P}_{\nu^{(a)}}^{\tau_\delta}(E) = \mathbb{P}_{\nu^{(a)}}(E)$. Since $\{\hat{a}_{\tau_\delta} = a\}$ is τ_δ -measurable and satisfies that $\{\hat{a}_{\tau_\delta} = a\} \subseteq \{\hat{a}_{\tau_\delta} \neq b\}$, we obtain

$$\text{TV}(\mathbb{P}_{\nu^{(a)}}^{\tau_\delta}, \mathbb{P}_{\nu^{(b)}}^{\tau_\delta}) \geq \mathbb{P}_{\nu^{(a)}}(\hat{a}_{\tau_\delta} = a) - \mathbb{P}_{\nu^{(b)}}(\hat{a}_{\tau_\delta} = a) \geq 1 - 2\delta.$$

Applying Theorem 5 concludes the proof since

$$\max_{a \in [K]} \mathbb{E}_{\nu^{(a)}}[\tau_\delta - N_a(\tau_\delta)] \geq \frac{1}{K} \sum_{a \in [K]} \mathbb{E}_{\nu^{(a)}}[\tau_\delta - N_a(\tau_\delta)] \geq \frac{K-1}{64(\Delta + \varepsilon)^2}.$$

Appendix F. Analysis of APGAI: Proof of Theorem 8

When combined with the GLR stopping Eq. (6) using threshold Eq. (7), APGAI becomes dependent of a risk $\delta \in (0, 1)$.

Remark 31 (Risk δ : algorithmic (Appendix F) or analysis (Appendix B)) *The risk parameter δ is only present in the probabilistic statements that involves the GLR stopping rule Eq. (6) due to the stopping threshold $c(T, \delta)$ as in Eq. (7) that depends on the risk δ . The*

risk δ is a parameter of the algorithm ensuring the δ -correctness of the resulting algorithm by Lemma 7. We highlight the difference with the analysis of the probability of error for APGAI detailed in Appendix B. The parameter δ is only used for the analysis to define a similar sequence of concentration events $(\tilde{\mathcal{E}}_{T,\delta})_{T>K}$. While the non-asymptotic analysis of the expected sample complexity only requires coarse upper bound on $\sum_{T>K} \mathbb{P}_\nu(\mathcal{E}_T^c)$ by Lemma 43, the non-asymptotic analysis of the probability of error requires a small upper bound on each $\mathbb{P}_\nu(\tilde{\mathcal{E}}_{T,\delta}^c)$. Therefore, it is not necessary to introduce a similar analysis parameter $\tilde{\delta}$ here, and we simply take $\tilde{\delta} := 1$. The purpose of the analysis parameter δ in Appendix B is to quantify how small $\mathbb{P}_\nu(\tilde{\mathcal{E}}_{T,\delta}^c)$ is. As we show that the error event is included in $\tilde{\mathcal{E}}_{T,\delta}^c$ for T large enough (as a function of δ), we can invert the upper bound based on Lemma 45.

Proof strategy. Let $\mu \in \mathbb{R}^K$ such that $\mu_a \neq \theta$ for all $a \in \mathcal{A}$. Let $s > 1$. For all $T > K$ and $\mathcal{E}_T = \mathcal{E}_{T,1}$ where $\mathcal{E}_{T,\delta}$ as in Eq. (21), i.e.

$$\mathcal{E}_T = \left\{ \forall a \in \mathcal{A}, \forall t \leq T, |\hat{\mu}_a(t) - \mu_a| < \sqrt{2f_1(T)/N_a(t)} \right\}, \quad (20)$$

with $f_1(T) = (1+s)\log T$. The sequence of concentration events $(\mathcal{E}_T)_{T>K}$ will be used to derive probabilistic statements on the APGAI sampling and recommendation rules, holding provided concentration holds. Crucially, while these events are independent of the risk δ , the probability that $\bigcup_{T>K} \mathcal{E}_T^c$ can still be upper bounded. Namely, combining a direct union bound with Lemma 40, we have $\sum_{T>K} \mathbb{P}_\nu(\mathcal{E}_T^c) \leq K\zeta(s)$ where ζ is the Riemann ζ function.

Suppose that we have constructed a time $T_\mu(\delta) > K$ such that $\mathcal{E}_T \subseteq \{\tau_\delta \leq T\}$ for $T \geq T_\mu(\delta)$. Then, using Lemma 43, we obtain $\mathbb{E}_\nu[\tau_\delta] \leq T_\mu(\delta) + K\zeta(s)$. To prove Theorem 8, we will distinguish between instances μ such that $\mathcal{A}_\theta = \emptyset$ (Appendix F.1) and instances μ such that $\mathcal{A}_\theta \neq \emptyset$ (Appendix F.2).

As for the proof of Theorem 2, our main technical tool is Lemma 10. It is direct to see that Lemmas 14 and 19 can be adapted to hold for $\mathcal{E}_T$ and $f_1(T) = (1+s)\log T$. Combined with Lemma 46, we state those results in a more explicit form, and omit the details of the proof. Since the concentration event $\mathcal{E}_T$ is independent of the risk δ , the time T_μ and S_μ in Lemmas 32 and 33 are independent of δ . Since both T_μ and S_μ scale as $\mathcal{O}(H_1(\mu) \log H_1(\mu))$, the δ -independent non-asymptotic bound for APGAI will scale as $\mathcal{O}(H_1(\mu) \log H_1(\mu))$ even when there are good arms. The independence in δ is crucial to differentiate the asymptotic behavior of APGAI when there are good arms. If $T_\mu(\tilde{\delta})$ and $S_\mu(\tilde{\delta})$ were used, we would obtain a dependency in $\mathcal{O}(H_1(\mu) \log(H_1(\mu)/\tilde{\delta}))$, which is undesirable when the analysis parameter $\tilde{\delta}$ is chosen as the algorithmic parameter δ . Taking $\tilde{\delta} = 1$ circumvents this issue.

Lemma 32 (Lemma 14: concentration event $\mathcal{E}_T$ instead of $\tilde{\mathcal{E}}_{T,\delta}$) *Let $\mu \in \mathbb{R}^K$ such that $\mathcal{A}_\theta = \emptyset$ and $\mu_a \neq \theta$ for all $a \in \mathcal{A}$. Let $s > 1$. Let $T_\mu = h_1(18(1+s)H_1(\mu), K)$ where h_1 is defined in Lemma 46. For all $T > T_\mu$, under the event $\mathcal{E}_T$ as in Eq. (20), we have $N_a(T) > 2(1+s)\Delta_a^{-2} \log(T)$ for all $a \in \mathcal{A}$.*

Proof Let us define $\tilde{T}_\mu = \sup \{T \mid T \leq 18(1+s)H_1(\mu) \log(T) + K\}$. Using Lemma 46, we obtain $\tilde{T}_\mu \leq T_\mu$ where $T_\mu = h_1(18(1+s)H_1(\mu), K)$. Combined with the proof of Lemma 14, this concludes the proof. $\blacksquare$

Lemma 33 (Lemma 19: concentration event $\mathcal{E}_T$ instead of $\tilde{\mathcal{E}}_{T,\delta}$) Let $\mu \in \mathbb{R}^K$ such that $\mathcal{A}_\theta \neq \emptyset$ and $\mu_a \neq \theta$ for all $a \in \mathcal{A}$. Let $s > 1$. Let $S_\mu = h_1(4(1+s)H_1(\mu), K + 2|\mathcal{A}_\theta|)$ where h_1 is defined in Lemma 46. For all $T > S_\mu$, under the event $\mathcal{E}_T$ as in Eq. (20), we have $\hat{a}_T \in \mathcal{A}_\theta$ and there exists $a \in \mathcal{A}_\theta$ such that $N_a(T) > (\Delta_a^{-1} \sqrt{2(1+s) \log(T)} + 1)^2$.

Proof Let us define $\tilde{S}_\mu = \sup \{T \mid T \leq 4(1+s)H_1(\mu) \log(T) + K + 2|\mathcal{A}_\theta|\}$. Lemma 46 yields $\tilde{S}_\mu \leq S_\mu$ where $S_\mu = h_1(4(1+s)H_1(\mu), K + 2|\mathcal{A}_\theta|)$. Combined with the proof of Lemma 19, this concludes the proof. $\blacksquare$

Theorem 8 is obtained by combining Lemmas 35 and 37.

F.1 Instances where $\mathcal{A}_\theta = \emptyset$

When $\mathcal{A}_\theta = \emptyset$, provided concentration event $\mathcal{E}_T$ holds, we have $\hat{a}_T = \emptyset$ and $a_{T+1} \in \arg \min_{a \in \mathcal{A}} W_a^-(T)$ for $T > T_\mu$. As detailed above, we have $T_\mu = \mathcal{O}(H_1(\mu) \log H_1(\mu))$, yet is independent of the risk δ . Lemma 34 formalizes this intuition.

Lemma 34 Let $s > 1$. Let $T_\mu = h_1(18(1+s)H_1(\mu), K)$ where h_1 is defined in Lemma 46. For all $T > T_\mu$, under $\mathcal{E}_T$ as in Eq. (20), $a_{T+1} \in \arg \min_{a \in \mathcal{A}} W_a^-(T)$ and $\hat{a}_T = \emptyset$.

Proof Let T_μ as in Lemma 14. Let $T > T_\mu$. Using Lemma 32, under $\mathcal{E}_T$ as in Eq. (20), we obtain that $N_a(T) > \frac{2f_1(T)}{(\theta - \mu_a)^2}$ for all $a \in \mathcal{A}$. Then $\hat{\mu}_a(T) \leq \mu_a + \sqrt{2f_1(T)/N_a(T)} < \theta$ for all $a \in \mathcal{A}$, hence $\max_{a \in \mathcal{A}} \hat{\mu}_a(T) < \theta$. Using the definition of the sampling rule when $\max_{a \in \mathcal{A}} \hat{\mu}_a(T) < \theta$, for all $T > T_\mu$, we have $a_{T+1} \in \arg \min_{a \in \mathcal{A}} W_a^-(T)$ and $\hat{a}_T = \emptyset$. $\blacksquare$

When coupled with the GLR stopping Eq. (6) using threshold Eq. (7), Lemma 35 gives an upper bound on the expected sample complexity of APGAI when $\mathcal{A}_\theta = \emptyset$. Since it involves the stopping threshold Eq. (7), the upper bound $C_\mu(\delta)$ depends on the risk δ . It satisfies $\limsup_{\delta \rightarrow 0} C_\mu(\delta) / \log(1/\delta) \leq 2H_1(\mu)$ and its δ -independent dominating dependency scales as $\mathcal{O}(H_1(\mu) \log H_1(\mu))$.

Lemma 35 Let $\delta \in (0, 1)$. Combined with GLR stopping Eq. (6) using threshold Eq. (7), the APGAI algorithm is δ -correct and it satisfies that, for all $\nu \in \mathcal{D}^K$ with mean μ such that $\mathcal{A}_\theta(\mu) = \emptyset$ and $\Delta_{\min} > 0$, $\mathbb{E}_\nu[\tau_\delta] \leq C_\mu(\delta) + K\pi^2/6 + 1$, with $H_1(\mu)$ as in Eq. (1) and $T_\mu = h_1(54H_1(\mu), K)$ with h_1 is defined in Lemma 46 and

$$\begin{aligned} C_\mu(\delta) &= \sup \left\{ T \mid \frac{T - T_\mu}{2H_1(\mu)} \leq \left(\sqrt{c(T, \delta)} + \sqrt{3 \log T} \right)^2 + \left(\theta - \min_{a \in \mathcal{A}} \mu_a \right)^2 - 3 \log T_\mu \right\} \\ &= \sup \{ t \mid t \leq 2H_1(\mu)(\sqrt{c(t, \delta)} + \sqrt{3 \log t})^2 + D_1(\mu) \}, \end{aligned}$$

where $D_1(\mu) = T_\mu + 2H_1(\mu)(\theta - \min_{a \in \mathcal{A}} \mu_a)^2 - 6H_1(\mu) \log T_\mu$. In particular, it satisfies $\limsup_{\delta \rightarrow 0} \mathbb{E}_\nu[\tau_\delta] / \log(1/\delta) \leq 2H_1(\mu)$.

Proof Let T_μ as in Lemma 34. Let $T > T_\mu$ such that $\mathcal{E}_T \cap \{\tau_\delta > T\}$ holds true. Let $w \in \Delta_K$ such that $w_a = (\theta - \mu_a)^{-2} H_1(\mu)^{-1}$ for all $a \in \mathcal{A}$. Using the pigeonhole principle, at time T there exists $a_1 \in \mathcal{A}$ such that $N_{a_1}(T) - N_{a_1}(T_\mu) \geq (T - T_\mu)w_{a_1}$. Let $T \geq T_\mu + (\min_{a \in \mathcal{A}} w_a)^{-1}$,

hence we have $N_{a_1}(T) - N_{a_1}(T_\mu) \geq w_{a_1} / \min_{a \in \mathcal{A}} w_a \geq 1$. Therefore, arm a_1 has been sampled at least once in (T_μ, T) . Let $t_{a_1} \in (T_\mu, T)$ be the last time at which arm a_1 was selected to be pulled next, *i.e.* $a_{t_{a_1}+1} = a_1$ and $N_{a_1}(T) = N_{a_1}(t_{a_1} + 1) = N_{a_1}(t_{a_1}) + 1$. Since $t_{a_1} > T_\mu$, Lemma 34 yields that $a_1 = a_{t_{a_1}+1} \in \arg \min_{a \in \mathcal{A}} W_a^-(t_{a_1})$. Moreover, we have

$$N_{a_1}(t_{a_1}) = N_{a_1}(T) - 1 \geq (T - T_\mu)w_{a_1} + N_{a_1}(T_\mu) - 1 \geq Tw_{a_1} + \frac{2f_1(T_\mu) - T_\mu H_1(\mu)^{-1}}{(\theta - \mu_{a_1})^2} - 2,$$

where we used that $N_{a_1}(T_\mu) \geq N_{a_1}(T_\mu + 1) - 1 > 2f_1(T_\mu + 1)\Delta_{a_1}^{-2}$ and f_1 is increasing. Under $\mathcal{E}_T$ as in Eq. (20), using that $a_1 = a_{t_{a_1}+1} \in \arg \min_{a \in \mathcal{A}} W_a^-(t_{a_1})$, we obtain

$$\begin{aligned} W_{a_1}^-(t_{a_1}) &= \sqrt{N_{a_1}(t_{a_1})}(\theta - \hat{\mu}_a(t_{a_1}))_+ = \sqrt{N_{a_1}(t_{a_1})}(\theta - \hat{\mu}_a(t_{a_1})) \\ &\geq \sqrt{N_{a_1}(t_{a_1})}(\theta - \mu_{a_1}) - \sqrt{2f_1(T)} \\ &\geq \sqrt{(Tw_{a_1}(\theta - \mu_{a_1})^2 + 2f_1(T_\mu) - T_\mu H_1(\mu)^{-1} - 2(\theta - \mu_{a_1})^2) - \sqrt{2f_1(T)}} \\ &= \sqrt{(T - T_\mu)H_1(\mu)^{-1} + 2f_1(T_\mu) - 2(\theta - \mu_{a_1})^2 - \sqrt{2f_1(T)}}. \end{aligned}$$

Since $a_1 = a_{t_{a_1}+1} \in \arg \min_{a \in \mathcal{A}} W_a^-(t_{a_1})$, using that the condition of the stopping rule is not met at time t_{a_1} yields

$$\begin{aligned} \sqrt{2c(T, \delta)} &\geq \sqrt{2c(\delta, t_{a_1})} \geq \min_{b \in \mathcal{A}} W_b^-(t_{a_1}) = W_{a_1}^-(t_{a_1}) \quad \text{hence} \\ \sqrt{2c(T, \delta)} &\geq \sqrt{(T - T_\mu)H_1(\mu)^{-1} + 2f_1(T_\mu) - 2(\theta - \mu_{a_1})^2 - \sqrt{2f_1(T)}}. \end{aligned}$$

Using $\mu_{a_1} \geq \min_{a \in \mathcal{A}} \mu_a$, the above inequality can be rewritten as

$$T - T_\mu \leq 2 \left(\sqrt{c(T, \delta)} + \sqrt{f_1(T)} \right)^2 H_1(\mu) + 2H_1(\mu) \left((\theta - \min_{a \in \mathcal{A}} \mu_a)^2 - f_1(T_\mu) \right).$$

Let us define

$$C_\mu(\delta) = \sup \left\{ T \mid \frac{T - T_\mu}{2H_1(\mu)} \leq \left(\sqrt{c(T, \delta)} + \sqrt{f_1(T)} \right)^2 + (\theta - \min_{a \in \mathcal{A}} \mu_a)^2 - f_1(T_\mu) \right\}.$$

It is direct to notice that $T_\mu + (\min_{a \in \mathcal{A}} w_a)^{-1} = T_\mu + (\theta - \min_{a \in \mathcal{A}} \mu_a)^2 H_1(\mu) \leq C_\mu(\delta)$. Therefore, we have shown that for $T \geq C_\mu(\delta) + 1$, we have $\mathcal{E}_T \subset \{\tau_{<, \delta} \leq T\} \subseteq \{\tau_\delta \leq T\}$ since $\tau_\delta := \min\{\tau_{>, \delta}, \tau_{<, \delta}\} \leq \tau_{<, \delta}$ almost surely by definition. Using Lemma 43, we obtain $\mathbb{E}_\nu[\tau_\delta] \leq C_\mu(\delta) + K\zeta(s) + 1$. Taking $s = 2$, using that $\zeta(2) = \pi^2/6$ and $f_1(T) = 3 \log T$ yields the second part of the result. Using Lemma 47, direct manipulations show that $\limsup_{\delta \rightarrow 0} \frac{\mathbb{E}_\nu[\tau_\delta]}{\log(1/\delta)} \leq \limsup_{\delta \rightarrow 0} \frac{C_\mu(\delta)}{\log(1/\delta)} \leq 2H_1(\mu)$. According to Lemma 1, we have proven asymptotic optimality. Lemma 7 gives the δ -correctness of the APGAI algorithm due to our recommendation rule. $\blacksquare$

F.2 Instances where $\mathcal{A}_\theta \neq \emptyset$

When $\mathcal{A}_\theta \neq \emptyset$, provided concentration event $\mathcal{E}_T$ holds, we have $\hat{a}_T = a_{T+1}$ and $a_{T+1} \in \arg \max_{a \in \mathcal{A}_\theta} W_a^+(T)$ for $T > S_\mu$. As detailed above, we have $S_\mu = \mathcal{O}(H_1(\mu) \log H_1(\mu))$, yet it is independent of the risk δ . Lemma 36 formalizes this intuition.

Lemma 36 *Let $s > 1$. Let $S_\mu = h_1(4(1+s)H_1(\mu), K + 2|\mathcal{A}_\theta|)$ where h_1 is defined in Lemma 46. For all $T > S_\mu$, under $\mathcal{E}_T$ as in Eq. (20), $\hat{a}_T = a_{T+1}$ and $a_{T+1} \in \arg \max_{a \in \mathcal{A}_\theta} W_a^+(T)$.*

Proof Let S_μ as in Lemma 33. Let $T > S_\mu$. Using Lemma 33, under $\mathcal{E}_T$ as in Eq. (20), we have $\hat{a}_T \in \mathcal{A}_\theta$ and there exists $a \in \mathcal{A}_\theta$ such that $N_a(T) > \frac{2f_1(T)}{(\mu_a - \theta)^2}$. Then, we have $\hat{\mu}_a(T) \geq \mu_a - \sqrt{2f_1(T)/N_a(T)} > \theta$, hence $\max_{a \in \mathcal{A}_\theta} \hat{\mu}_a(T) > \theta$. Using Lemma 18 and the definition of the recommendation rule when $\max_{a \in \mathcal{A}} \hat{\mu}_a(T) > \theta$, we obtain that $\hat{a}_T = a_{T+1}$, hence $a_{T+1} \in \mathcal{A}_\theta$. This concludes the proof. $\blacksquare$

When coupled with the GLR stopping Eq. (6) using threshold Eq. (7), Lemma 37 gives an upper bound on the expected sample complexity of APGAI when $\mathcal{A}_\theta \neq \emptyset$. Since it involves the stopping threshold Eq. (7), the upper bound $C_\mu(\delta)$ depends on the risk δ . It satisfies $\limsup_{\delta \rightarrow 0} C_\mu(\delta)/\log(1/\delta) \leq 2H_\theta(\mu)$. However, its δ -independent dominating dependency scales as $\mathcal{O}(H_1(\mu) \log H_1(\mu))$, *i.e.* the same dependency as when there are no good arms.

Lemma 37 *Let $\delta \in (0, 1)$. Combined with GLR stopping Eq. (6) using threshold Eq. (7), the APGAI algorithm is δ -correct and it satisfies that, for all $\nu \in \mathcal{D}^K$ with mean μ such that $\mathcal{A}_\theta \neq \emptyset$ and $\Delta_{\min} > 0$, $\mathbb{E}_\nu[\tau_\delta] \leq C_\mu(\delta) + K\pi^2/6 + 1$, where $H_\theta(\mu)$ as in Eq. (1) and $S_\mu = h_1(12H_1(\mu), K + 2|\mathcal{A}_\theta|)$ with h_1 is defined in Lemma 46 and*

$$C_\mu(\delta) = \sup \left\{ T \mid \frac{T - S_\mu - 1}{2H_\theta(\mu)} \leq \left(\sqrt{c(T, \delta)} + \sqrt{3 \log T} \right)^2 - \frac{3 \log S_\mu}{H_\theta(\mu) \max_{a \in \mathcal{A}_\theta} \Delta_a^2} \right\} \\ = \sup \{ t \mid t \leq 2H_\theta(\mu)(\sqrt{c(t, \delta)} + \sqrt{3 \log t})^2 + D_\theta(\mu) \},$$

where $D_\theta(\mu) = S_\mu + 1 - \frac{6 \log S_\mu}{\max_{a \in \mathcal{A}_\theta} \Delta_a^2}$. It satisfies $\limsup_{\delta \rightarrow 0} \mathbb{E}_\nu[\tau_\delta]/\log(1/\delta) \leq 2H_\theta(\mu)$.

Proof Let S_μ as in Lemma 36. Let $T > S_\mu$ such that $\mathcal{E}_T \cap \{\tau_\delta > T\}$ holds true. Using Lemma 36, we know that $a_{t+1} \in \mathcal{A}_\theta$ for all $t \in (S_\mu, T]$. Direct summation yields that

$$T - S_\mu = \sum_{a \in \mathcal{A}_\theta} (N_a(T) - N_a(S_\mu)) + \sum_{t \in (S_\mu, T]} \mathbb{1}(a_{t+1} \notin \mathcal{A}_\theta) = \sum_{a \in \mathcal{A}_\theta} (N_a(T) - N_a(S_\mu)).$$

At time $S_\mu + 1$, let $a_1 \in \mathcal{A}_\theta$ as in Lemma 36, *i.e.* such that $N_{a_1}(S_\mu + 1) > \frac{2f_1(S_\mu + 1)}{(\mu_{a_1} - \theta)^2}$. Using that f_1 is increasing, we obtain

$$\sum_{b \in \mathcal{A}_\theta} N_b(S_\mu) \geq N_{a_1}(S_\mu + 1) - 1 > \frac{2f_1(S_\mu + 1)}{(\mu_{a_1} - \theta)^2} - 1 \geq \frac{2f_1(S_\mu)}{\max_{a \in \mathcal{A}_\theta} (\mu_a - \theta)^2} - 1.$$

Let $g(S_\mu) = S_\mu - 2f_1(S_\mu)/\max_{a \in \mathcal{A}_\theta} \Delta_a^2 + 1$. Therefore, we have shown that $\sum_{a \in \mathcal{A}_\theta} N_a(T) \geq T - g(S_\mu)$. Let $A_\theta = |\mathcal{A}_\theta|$ and $w \in \Delta_{A_\theta}$ such that $w_a = (\mu_a - \theta)^{-2} H_\theta(\mu)^{-1}$ with $H_\theta(\mu)$

as in Eq. (1). Using the pigeonhole principle, there exists $a_0 \in \mathcal{A}_\theta$ such that $N_{a_0}(T) \geq w_{a_0}(T - g(S_\mu)) = \Delta_{a_0}^{-2} H_\theta(\mu)^{-1} (T - g(S_\mu))$. Let $E_\mu = \sup \{T \mid T \leq g(S_\mu) + 2H_\theta(\mu)f_1(T)\}$. Let $T > E_\mu$. Then, we have $N_{a_0}(T) \geq \Delta_{a_0}^{-2} H_\theta(\mu)^{-1} (T - g(S_\mu)) > 2f_1(T)\Delta_{a_0}^{-2}$, hence $\mu_{a_0}(T) > \theta$. Using that the condition of the stopping rule is not met at time T , we obtain

$$\sqrt{2c(T, \delta)} \geq \max_{a \in \mathcal{A}} W_a^+(T) \geq W_{a_0}^+(T) = \sqrt{N_{a_0}(T)}(\hat{\mu}_{a_0}(T) - \theta)_+ = \sqrt{N_{a_0}(T)}(\hat{\mu}_{a_0}(T) - \theta).$$

Then, we obtain

$$\begin{aligned} \sqrt{2c(T, \delta)} &\geq \sqrt{N_{a_0}(T)}(\mu_{a_0} - \theta) - \sqrt{2f_1(T)} \geq \sqrt{T - g(S_\mu)}\sqrt{w_{a_0}(\mu_{a_0} - \theta)^2} - \sqrt{2f_1(T)} \\ &= \sqrt{T - g(S_\mu)}H_\theta(\mu)^{-1/2} - \sqrt{2f_1(T)}. \end{aligned}$$

The above can be rewritten as $T \leq 2 \left(\sqrt{c(T, \delta)} + \sqrt{f_1(T)} \right)^2 H_\theta(\mu) + g(S_\mu)$. Using that $g(S_\mu) = S_\mu - \frac{2f_1(S_\mu)}{\max_{a \in \mathcal{A}_\theta} \Delta_a^2} + 1$, let us define

$$D_\mu(\delta) = \sup \left\{ T \mid \frac{T - S_\mu - 1}{2H_\theta(\mu)} \leq \left(\sqrt{c(T, \delta)} + \sqrt{f_1(T)} \right)^2 - \frac{f_1(S_\mu)}{H_\theta(\mu) \max_{a \in \mathcal{A}_\theta} \Delta_a^2} \right\}.$$

It is direct to see that $D_\mu(\delta) \geq E_\mu \geq S_\mu$. Therefore, we have shown that for $T \geq D_\mu(\delta) + 1$, we have $\mathcal{E}_T \subset \{\tau_{>, \delta} \leq T\} \subseteq \{\tau_\delta \leq T\}$ since $\tau_\delta := \min\{\tau_{>, \delta}, \tau_{<, \delta}\} \leq \tau_{>, \delta}$ almost surely by definition. Using Lemma 43, we obtain $\mathbb{E}_\nu[\tau_\delta] \leq D_\mu(\delta) + K\zeta(s) + 1$. Taking $s = 2$, using that $\zeta(2) = \pi^2/6$ and $f_1(T) = 3 \log T$ yields the second part of the result. Using Lemma 47, direct manipulations show that $\limsup_{\delta \rightarrow 0} \frac{\mathbb{E}_\nu[\tau_\delta]}{\log(1/\delta)} \leq \limsup_{\delta \rightarrow 0} \frac{D_\mu(\delta)}{\log(1/\delta)} \leq 2H_\theta(\mu)$. According to Lemma 1, our result is weaker than asymptotic optimality when $|\mathcal{A}_\theta| \geq 2$. Lemma 7 gives the δ -correctness of the APGAI algorithm, since the recommendation rule of matches the one of Lemma 7. $\blacksquare$

F.3 Explicit Non Asymptotic Upper Bound

In the above, we have shown the following implicit upper bound on the sample complexity τ_δ of the APGAI algorithm, namely $\mathbb{E}_\nu[\tau_\delta] \leq C_\mu(\delta) + K\pi^2/6 + 1$ with

$$C_\mu(\delta) := \sup\{t \mid t \leq 2H_{i_\mu}(\mu)(\sqrt{c(t, \delta)} + \sqrt{3 \log t})^2 + D_{i_\mu}(\mu)\},$$

where $i_\mu := 1 + (\theta - 1)\mathbb{1}(\mathcal{A}_\theta(\mu) \neq \emptyset)$. Since $C_\mu(\delta)$ is defined implicitly, we provide an explicit upper bound by leveraging some (loose) approximations. Using that $(x + y)^2 \leq 2(x^2 + y^2)$ and $\overline{W}_{-1}(y) \leq x$ if and only if $y \leq x - \log(x)$ (see Lemma 44), we obtain $C_\mu(\delta) \leq$

$$\begin{aligned} &\sup\{t \mid t \leq 2H_{i_\mu}(\mu)\overline{W}_{-1}(2 \log(K/\delta) + 4 \log \log(e^4 t) + 1/2) + 12H_{i_\mu}(\mu) \log(t) + D_{i_\mu}(\mu)\} \\ &\leq \sup \left\{ t \mid \frac{t - 12H_{i_\mu}(\mu) \log(t) - D_{i_\mu}(\mu)}{2H_{i_\mu}(\mu)} \leq 2 \log \left(\frac{Ke^{1/4}}{\delta} \right) + \right. \\ &\quad \left. \log \left((4 + \log t)^4 \frac{t - 12H_{i_\mu}(\mu) \log(t) - D_{i_\mu}(\mu)}{2H_{i_\mu}(\mu)} \right) \right\}. \end{aligned}$$

Numerically, we observe that $\frac{3}{2} \log(x) + 7 \geq \log(x(4 + \log x)^4)$ for all $x \geq 0.0015$. Since $C_\mu(\delta) \geq 1$ and $\log\left((4 + \log t)^4 \frac{t - 12H_{i_\mu}(\mu) \log(t) - D_{i_\mu}(\mu)}{2H_{i_\mu}(\mu)}\right) \leq \log((4 + \log t)^4 t) - \log(2H_{i_\mu}(\mu))$, we obtain that

$$\begin{aligned} C_\mu(\delta) &\leq \sup \left\{ t \mid t \leq 4H_{i_\mu}(\mu) \log \left(\frac{Ke^{15/4}}{2H_{i_\mu}(\mu)\delta} \right) + 15H_{i_\mu}(\mu) \log(t) + D_{i_\mu}(\mu) \right\} \\ &\leq h_1 \left(15H_{i_\mu}(\mu), 4H_{i_\mu}(\mu) \log \left(\frac{Ke^{15/4}}{2H_{i_\mu}(\mu)\delta} \right) + D_{i_\mu}(\mu) \right), \end{aligned}$$

where the last inequality uses Lemma 46 with h_1 is defined therein as $h_1(x, y) := x\bar{W}_{-1}(y/x + \log(x))$. This upper bound is fully explicit since the function h_1 depends on $\bar{W}_{-1}$. Finally, we can use the approximation $\bar{W}_{-1}(x) \leq x + \log(x) + \min(1/\sqrt{x}, 1/2)$ (see Lemma 44), hence

$$C_\mu(\delta) \leq h(15H_{i_\mu}(\mu), 4H_{i_\mu}(\mu) (\log(K/\delta) + 15/4 - 2\log(2H_{i_\mu}(\mu))) + D_{i_\mu}(\mu))$$

where $h(x, y) := y + x \log(x) + x \log(y/x + \log(x)) + x/2$. ■

F.3.1 DISCUSSION ON SUB-OPTIMAL UPPER BOUND

As discussed in Section 5, Theorem 8 has a sub-optimal scaling when $\mathcal{A}_\theta(\mu) \neq \emptyset$. Instead of $2 \min_{a \in \mathcal{A}_\theta(\mu)} \Delta_a^{-2}$, our asymptotic upper bound on the expected sample complexity scales only as $2H_\theta(\mu)$. It is quite natural to wonder whether we could improve on this dependency, and whether $2 \min_{a \in \mathcal{A}_\theta(\mu)} \Delta_a^{-2}$ is achievable by APGAI. In the following, we provide intuition on why we could improve up to $2 \max_{a \in \mathcal{A}_\theta(\mu)} \Delta_a^{-2}$, but not till $2 \min_{a \in \mathcal{A}_\theta(\mu)} \Delta_a^{-2}$.

On the impossibility to achieve $2 \min_{a \in \mathcal{A}_\theta(\mu)} \Delta_a^{-2}$. We argue that whenever $\mathcal{A}_\theta(\mu) \neq \arg \max_{a \in \mathcal{A}_\theta(\mu)} \Delta_a$, there is no mechanism to avoid that the sampling rule of APGAI focuses all its samples on an arm $a \in \mathcal{A}_\theta(\mu) \setminus \arg \max_{a \in \mathcal{A}_\theta(\mu)} \Delta_a$. Therefore, it is not possible to achieve $2 \min_{a \in \mathcal{A}_\theta(\mu)} \Delta_a^{-2}$.

For the sake of presentation, we consider the most simple case where this impossibility result occur. Let ν be a two-arms instance with mean μ such that $\mu_1 > \mu_2 > \theta = 0$. Let $(X_s)_{s \geq 1}$ and $(Y_s)_{s \geq 1}$ be i.i.d. observations from ν_1 and ν_2 . APGAI initializes by sampling each arm once. Let $\varepsilon \in (0, \mu_2)$ and $T \in \mathbb{N}$ such that

$$T > n_\varepsilon(T) := \sup\{t \mid \sqrt{t-1}\mu_2 - 2\sqrt{\log T} \leq \mu_2 - \varepsilon\}.$$

By conditional independence, the event $\mathcal{G}_{\varepsilon, T} = \{X_1 < \mu_2 - \varepsilon \leq \min_{1 \leq s \leq n_\varepsilon(T)} Y_s\}$ has probability $\mathbb{P}_\nu(\mathcal{G}_{\varepsilon, T}) = \mathbb{P}_{X \sim \nu_1}(X < \mu_2 - \varepsilon)(1 - \mathbb{P}_{Y \sim \nu_2}(Y < \mu_2 - \varepsilon))^{n_\varepsilon(T)}$. Under $\mathcal{G}_{\varepsilon, T}$, we have $a_{t+1} = 2$ for all $2 \leq t \leq n_\varepsilon(T)$, hence $N_2(t) = t - 1$ and $N_1(t) = 1$. Let $\mathcal{E}_T$ as in Eq. (21) for $s = 1$ and $\delta = 1$, i.e.

$$\mathcal{E}_T = \{\forall a \in \{1, 2\}, \forall t \leq T, |\hat{\mu}_a(t) - \mu_a| < \sqrt{4 \log(T)/N_a(t)}\}.$$

It satisfies $\mathbb{P}_\nu(\mathcal{E}_T^c) \leq 2/T$. We will show that by induction that $N_2(t) = t - 1$ and $N_1(t) = 1$ under $\mathcal{E}_T \cap \mathcal{G}_{\varepsilon, T}$. Under $\mathcal{G}_{\varepsilon, T}$, we know that the property holds for all $2 \leq t \leq n_\varepsilon(T)$. Suppose it is true at time $T > t > n_\varepsilon(T)$, we will show that $a_{t+1} = 2$ hence it is true at time $t + 1$. Under $\mathcal{E}_T \cap \mathcal{G}_{\varepsilon, T}$, we have $W_2^+(t) = \sqrt{N_2(t)\hat{\mu}_2(t)} > \sqrt{N_2(t)\mu_2} - 2\sqrt{\log T} =$

$\sqrt{t-1}\mu_2 - 2\sqrt{\log T} > \mu_2 - \varepsilon \geq W_1^+(2) = W_1^+(t)$. Therefore, we have $a_{t+1} = 2$. This concludes the proof by induction that, under $\mathcal{E}_T \cap \mathcal{G}_{\varepsilon,T}$, for all $t \leq T$, $N_2(t) = t - 1$ and $N_1(t) = 1$. Since $\mathcal{E}_T$ and $\mathcal{G}_{\varepsilon,T}$ are both likely events, it is reasonable to expect $\mathcal{E}_T \cap \mathcal{G}_{\varepsilon,T}$ to be likely as well. Under this likely event, we see that APGAI focuses its sampling allocation to the arm 2 instead of the arm 1. The greediness of APGAI prevents it to switch the arm that is easiest to verify.

While the above argument considers only two arms and is not formally proven, it gives some intuition as regards what prevents APGAI from reaching $2 \min_{a \in \mathcal{A}_\theta(\mu)} \Delta_a^{-2}$. It is not possible to recover from one unlucky first draw for the best arm if a sub-optimal arm has no unlucky first draws. Formally proving such a negative result is an interesting direction for future work.

Towards reaching $2 \max_{a \in \mathcal{A}_\theta(\mu)} \Delta_a^{-2}$ asymptotically. We argue that APGAI focuses its sampling allocation to only one of the good arm $a \in \mathcal{A}_\theta(\mu)$, after a long enough time. Therefore, it should be possible to achieve $2 \max_{a \in \mathcal{A}_\theta(\mu)} \Delta_a^{-2}$.

Suppose towards contradiction that

$$\exists (a_1, a_2) \in \mathcal{A}_\theta(\mu)^2, \quad \min_{a \in \{a_1, a_2\}} N_a(T) \rightarrow_{T \rightarrow +\infty} +\infty.$$

Let S_μ as in Lemma 36. Let $T > S_\mu$ such that $\mathcal{E}_T \cap \{\tau_\delta > T\}$ holds true. In the proof of Lemma 37, we have shown that

$$\max_{a \in \mathcal{A}} W_a^+(T) \geq \sqrt{T - g(S_\mu)H_\theta(\mu)^{-1/2}} - \sqrt{2f_1(T)}.$$

At time $S_\mu + 1$, we have $\max_{a \in \mathcal{A}} W_a^+(S_\mu + 1) \geq W_{a_1}^+(S_\mu + 1)$. Since the transportation costs are independent to the other arms, we will show that sampling two arms an infinite number of times implies that the transportation costs are bounded. Given that we have shown they are growing towards $+\infty$, this is a contradiction. Using our assumption that $\min_{a \in \{a_1, a_2\}} N_a(T) \rightarrow_{T \rightarrow +\infty} +\infty$, we have that there exists an infinite number of intervals $(t_i^L, t_i^U)_{i \in \mathbb{N}}$ such that $a_{t+1} = a_1$ for all $t \in \bigcup_{i \in \mathbb{N}} [t_i^L, t_i^U)$, otherwise $a_{t+1} \neq a_1$. Let $i \in \mathbb{N}$. Using that a_1 is the only arm that is sampled in $[t_i^L, t_i^U)$ and that is not sampled at t_i^U , we obtain that

$$W_{a_1}^+(t_i^L) \geq \max_{a \in \mathcal{A} \setminus \{a_1\}} W_a^+(t_i^L) = \max_{a \in \mathcal{A} \setminus \{a_1\}} W_a^+(t_i^U) \geq W_{a_1}^+(t_i^U).$$

Since it is not sampled until t_{i+1}^L , we obtain that $W_{a_1}^+(t_i^U) = W_{a_1}^+(t_{i+1}^L)$. By induction is direct to see that

$$W_{a_1}^+(S_\mu + 1) \geq \max_{i, t_i^L \geq S_\mu + 1} W_{a_1}^+(t_i^L) \geq \sqrt{t_i^L - g(S_\mu)H_\theta(\mu)^{-1/2}} - \sqrt{2f_1(t_i^L)}.$$

Since the right-hand side converges towards infinity, there is a contradiction. Therefore, there exists a unique arm $a \in \mathcal{A}_\theta(\mu)$ such that $N_a(T) \rightarrow_{T \rightarrow +\infty} +\infty$.

While the above argument is not formally proven, it gives some intuition as regards why APGAI can reach $2 \max_{a \in \mathcal{A}_\theta(\mu)} \Delta_a^{-2}$. It is not possible to sample two good arms an infinite number of times since it would imply that the transportation costs are simultaneously bounded and converge towards infinity.

Appendix G. Concentration Results

In Appendix G.1, we prove the δ -correctness of the GLR stopping rule Eq. (6) with threshold Eq. (7) (Lemma 7). Appendix G.2 gathers sequence of concentration events which are used for our proofs.

G.1 Analysis of the GLR Stopping Rule: Proof of Lemma 7

Proving δ -correctness of a GLR stopping rule is done by leveraging concentration results. In particular, we build upon Lemma 28 Jourdan et al. (2023). Lemma 38 is obtained as a Corollary of Lemma 28 from Jourdan et al. (2023) by using a union bound over arms $a \in \mathcal{A}$. While it was only proven for Gaussian distributions, the concentration results also holds for sub-Gaussian distributions with variance $\sigma^2 = 1$ since we have $\mathbb{E}_X[\exp(sX)] \leq \exp(\lambda^2/2)$ for all $\lambda \in \mathbb{R}$.

Lemma 38 (Lemma 28 in Jourdan et al. (2023)) *Let $s > 1$ and $\delta \in (0, 1)$. Let $\overline{W}_{-1}(x) = -W_{-1}(-e^{-x})$ for all $x \geq 1$ (see Lemma 44), where W_{-1} is the negative branch of the Lambert W function. Let*

$$c(T, \delta) = \frac{1}{2} \overline{W}_{-1}(2 \log(K/\delta) + 2s \log(2s + \log T) + 2g(s)) ,$$

with $g(s) = \log(\zeta(s)) + s(1 - \log(2s)) + 1/2$ and ζ be the Riemann ζ function. Then,

$$\mathbb{P} \left(\exists T \in \mathbb{N}, \exists a \in \mathcal{A}, \sqrt{N_a(T)} |\mu_a(T) - \mu_a| > \sqrt{2c(T, \delta)} \right) \leq \delta .$$

We distinguish between the two cases $\mathcal{A}_\theta = \emptyset$ and $\mathcal{A}_\theta \neq \emptyset$. For the sake of simplicity, we use Lemma 38 for $s = 2$ and use that $2g(2) = 2 \log(\pi^2/6) + 5 - 4 \log(4) \leq 1/2$, which can be easily checked numerically.

Case 1. When $\mathcal{A}_\theta = \emptyset$, we have to show $\mathbb{P}_\nu(\tau_\delta < +\infty, \hat{a}_{\tau_\delta} \neq \emptyset) \leq \delta$. We recommend $\hat{a}_T \neq \emptyset$ only when $\max_{a \in \mathcal{A}} \hat{\mu}_a(t) > \theta$. In that case, we have $\hat{a}_T \in \arg \max_{a \in \mathcal{A}} W_a^+(T)$ where $W_a^+(T) = \sqrt{N_a(T)}(\mu_a(T) - \theta)_+$. Therefore, direct manipulations yield that

$$\begin{aligned} & \mathbb{P}_\nu(\tau_\delta < +\infty, \hat{a}_{\tau_\delta} \neq \emptyset) \\ & \leq \mathbb{P} \left(\exists T \in \mathbb{N}, \exists a \in \mathcal{A}, \hat{\mu}_a(t) > \theta, \sqrt{N_a(T)}(\hat{\mu}_a(T) - \theta)_+ \geq \sqrt{2c(T, \delta)} \right) \\ & \leq \mathbb{P} \left(\exists T \in \mathbb{N}, \exists a \in \mathcal{A}, \sqrt{N_a(T)}(\hat{\mu}_a(T) - \mu_a) + \sqrt{N_a(T)}(\mu_a - \theta) \geq \sqrt{2c(T, \delta)} \right) \\ & \leq \mathbb{P} \left(\exists T \in \mathbb{N}, \exists a \in \mathcal{A}, \sqrt{N_a(T)}(\hat{\mu}_a(T) - \mu_a) \geq \sqrt{2c(T, \delta)} \right) \leq \delta/2 . \end{aligned}$$

The second inequality uses that $\hat{\mu}_a(t) > \theta$ before dropping this condition. The third inequality uses that $\mu_a - \theta \leq 0$ since $\mathcal{A}_\theta = \emptyset$. The last inequality uses Lemma 38.

Case 2. When $\mathcal{A}_\theta \neq \emptyset$, we have to show $\mathbb{P}_\nu(\{\tau_\delta < +\infty\} \cap (\{\hat{a}_{\tau_\delta} = \emptyset\} \cup \{\hat{a}_{\tau_\delta} \notin \mathcal{A}_\theta\})) \leq \delta$. As above, when we recommend $\hat{a}_T \notin \mathcal{A}_\theta$, direct manipulations yield that

$$\begin{aligned} & \mathbb{P}_\nu(\tau_\delta < +\infty, \hat{a}_{\tau_\delta} \notin \mathcal{A}_\theta) \\ & \leq \mathbb{P}\left(\exists T \in \mathbb{N}, \exists a \notin \mathcal{A}_\theta, \hat{\mu}_a(t) > \theta, \sqrt{N_a(T)}(\hat{\mu}_a(T) - \theta)_+ \geq \sqrt{2c(T, \delta)}\right) \\ & \leq \mathbb{P}\left(\exists T \in \mathbb{N}, \exists a \notin \mathcal{A}_\theta, \sqrt{N_a(T)}(\hat{\mu}_a(T) - \mu_a) + \sqrt{N_a(T)}(\mu_a - \theta) \geq \sqrt{2c(T, \delta)}\right) \\ & \leq \mathbb{P}\left(\exists T \in \mathbb{N}, \exists a \notin \mathcal{A}_\theta, \sqrt{N_a(T)}(\hat{\mu}_a(T) - \mu_a) \geq \sqrt{2c(T, \delta)}\right) \leq \delta/2. \end{aligned}$$

The third inequality uses that $\mu_a - \theta \leq 0$ since $a \notin \mathcal{A}_\theta$.

Similarly, we recommend $\hat{a}_T = \emptyset$ only when $\max_{a \in \mathcal{A}} \hat{\mu}_a(t) \leq \theta$. In that case, we consider $W_a^-(T) = \sqrt{N_a(T)}(\theta - \mu_a(T))_+$. Therefore, direct manipulations yield that

$$\begin{aligned} & \mathbb{P}_\nu(\tau_\delta < +\infty, \hat{a}_{\tau_\delta} = \emptyset) \\ & \leq \mathbb{P}\left(\exists T \in \mathbb{N}, \forall a \in \mathcal{A}, \hat{\mu}_a(t) \leq \theta, \sqrt{N_a(T)}(\theta - \hat{\mu}_a(T))_+ \geq \sqrt{2c(T, \delta)}\right) \\ & \leq \mathbb{P}\left(\exists T \in \mathbb{N}, \forall a \in \mathcal{A}_\theta, \sqrt{N_a(T)}(\theta - \mu_a) + \sqrt{N_a(T)}(\mu_a - \hat{\mu}_a(T)) \geq \sqrt{2c(T, \delta)}\right) \\ & \leq \mathbb{P}\left(\exists T \in \mathbb{N}, \forall a \in \mathcal{A}_\theta, \sqrt{N_a(T)}(\mu_a - \hat{\mu}_a(T)) \geq \sqrt{2c(T, \delta)}\right) \leq \delta/2. \end{aligned}$$

The second inequality uses that $\hat{\mu}_a(t) \leq \theta$ before dropping this condition, and restrict to $a \in \mathcal{A}_\theta$. The third inequality uses that $\mu_a - \theta > 0$ since $a \in \mathcal{A}_\theta$. The last inequality uses Lemma 38. $\blacksquare$

G.2 Sequence of Concentration Events

Appendix G.2 provides sequence of concentration events which are used for our proofs. Lemma 39 is a standard concentration result for sub-Gaussian distribution, hence we omit the proof.

Lemma 39 *Let X be an observation from a sub-Gaussian distribution with mean 0 and variance $\sigma^2 = 1$. Then, for all $\delta \in (0, 1]$, $\mathbb{P}_X(|X| \geq \sqrt{2\log(1/\delta)}) \leq \delta$.*

Lemma 40 gives a sequence of concentration events under which the empirical means are close to their true values.

Lemma 40 *Let $\delta \in (0, 1]$ and $s \geq 0$. For all $T > K$, let us defined*

$$\mathcal{E}_{T, \delta} = \{\forall a \in \mathcal{A}, \forall t \leq T, |\hat{\mu}_a(t) - \mu_a| < \sqrt{2f_1(T, \delta)/N_a(t)}\}. \quad (21)$$

with $f_1(T, \delta) = \log(1/\delta) + (1 + s) \log T$. Then, for all $T > K$, $\mathbb{P}_\nu((\mathcal{E}_{T, \delta})^c) \leq \frac{K\delta}{T^s}$.

Proof Let $(X_s)_{s \in [T]}$ be i.i.d. observations from one sub-Gaussian distribution with mean 0 and variance $\sigma^2 = 1$. Then, $\frac{1}{m} \sum_{i=1}^m X_i$ is sub-Gaussian with mean 0 and variance $\sigma^2 = 1/m$.

By union bound over $\mathcal{A}$ and over $m \in [T]$, we obtain

$$\begin{aligned} & \mathbb{P}_\mu \left(\exists a \in \mathcal{A}, \exists t \leq T, |\hat{\mu}_a(t) - \mu_a| < \sqrt{\frac{2f_1(T, \delta)}{N_a(t)}} \right) \\ & \leq \sum_{a \in \mathcal{A}} \sum_{m \in [T]} \mathbb{P} \left(\left| \frac{1}{m} \sum_{s \in [m]} X_s \right| \geq \sqrt{\frac{2f_1(T, \delta)}{m}} \right) \leq \delta \sum_{a \in \mathcal{A}} \sum_{m \in [T]} T^{-(1+s)} = K\delta T^{-s}, \end{aligned}$$

where we used that $\hat{\mu}_a(t) - \mu_a = \frac{1}{N_a(t)} \sum_{s=1}^t \mathbf{1}(a_s = a) X_{s,a}$ and concentration results for sub-Gaussian observations (Lemma 39). $\blacksquare$

Lemma 41 provides concentration results on the empirical means, which are tighter than the one obtained in Lemma 40.

Lemma 41 *Let $\delta \in (0, 1]$ and $s \geq 0$. Let $\bar{W}_{-1}(x) = -W_{-1}(-e^{-x})$ for all $x \geq 1$ (see Lemma 44), where W_{-1} is the negative branch of the Lambert W function. For all $T > K$,*

$$\tilde{f}_1(T, \delta) = \frac{1}{2} \bar{W}_{-1}(2 \log(1/\delta) + 2s \log T + 2 \log(2 + \log T) + 2), \quad (22)$$

$$\text{and } \tilde{\mathcal{E}}_{T, \delta} = \{\forall a \in \mathcal{A}, \forall t \leq T, |\hat{\mu}_a(t) - \mu_a| < \sqrt{2\tilde{f}_1(T, \delta)/N_a(t)}\}. \quad (23)$$

Then, for all $T > K$, $\mathbb{P}_\nu((\tilde{\mathcal{E}}_{T, \delta})^c) \leq \frac{K\delta}{T^s}$.

Proof Let $(X_s)_{s \in [T]}$ be i.i.d. observations from one sub-Gaussian distribution with mean 0 and variance $\sigma^2 = 1$. Let $S_t = \sum_{s \in [t]} X_s$. To derive the concentration result, we use peeling.

Let $\eta > 0$, $\gamma > 0$ and $D = \lceil \frac{\log(T)}{\log(1+\eta)} \rceil$. For all $i \in [D]$, let $N_i = (1 + \eta)^{i-1}$. For all $i \in [D]$, we define the family of priors $f_{N_i, \gamma}(x) = \sqrt{\frac{\gamma N_i}{2\pi}} \exp\left(-\frac{x^2 \gamma N_i}{2}\right)$ with weights $w_i = \frac{1}{D}$ and process $\bar{M}(t) = \sum_{i \in [D]} w_i \int f_{N_i, \gamma}(x) \exp\left(xS_t - \frac{1}{2}x^2t\right) dx$, which satisfies $\bar{M}(0) = 1$. It is direct to see that $M(t) = \exp\left(xS_t - \frac{1}{2}x^2t\right)$ is a non-negative supermartingale since sub-Gaussian distributions with mean 0 and variance $\sigma^2 = 1$ satisfy $\mathbb{E}_X[\exp(sX)] \leq \exp(\lambda^2/2)$ for all $\lambda \in \mathbb{R}$. By Tonelli's theorem, then $\bar{M}(t)$ is also a non-negative supermartingale of unit initial value.

Let $i \in [D]$ and consider $t \in [N_i, N_{i+1})$. For all x , $f_{N_i, \gamma}(x) \geq \sqrt{\frac{N_i}{t}} f_{t, \gamma}(x) \geq \frac{1}{\sqrt{1+\eta}} f_{t, \gamma}(x)$. Direct computations shows that

$$\int f_{t, \gamma}(x) \exp\left(xS_t - \frac{1}{2}x^2t\right) dx = \frac{1}{\sqrt{1+\gamma^{-1}}} \exp\left(\frac{S_t^2}{2(1+\gamma)t}\right).$$

Minoring $\bar{M}(t)$ by one of the positive term of its sum, we obtain

$$\bar{M}(t) \geq \frac{1}{D} \frac{1}{\sqrt{(1+\gamma^{-1})(1+\eta)}} \exp\left(\frac{S_t^2}{2(1+\gamma)t}\right),$$

Using Ville's maximal inequality for non-negative supermartingale, we have that with probability greater than $1 - \delta$, $\log \bar{M}(t) \leq \log(1/\delta)$. Therefore, with probability greater than $1 - \delta$, for all $i \in [D]$ and $t \in [N_i, N_{i+1})$,

$$S_t^2/t \leq (1 + \gamma) (2 \log(1/\delta) + 2 \log D + \log(1 + \gamma^{-1}) + \log(1 + \eta)) .$$

Since this upper bound is independent of t , we can optimize it and choose γ as in Lemma 42.

Lemma 42 (Lemma A.3 in Degenne (2019)) *For $a, b \geq 1$, the minimum of $f(\eta) = (1 + \eta)(a + \log(b + \frac{1}{\eta}))$ is attained at η^* such that $f(\eta^*) \leq 1 - b + \bar{W}_{-1}(a + b)$. If $b = 1$, then there is equality.*

Therefore, with probability greater than $1 - \delta$, for all $i \in [D]$ and $t \in [N_i, N_{i+1})$,

$$\begin{aligned} \frac{S_t^2}{t} &\leq \bar{W}_{-1} (1 + 2 \log(1/\delta) + 2 \log D + \log(1 + \eta)) \\ &\leq \bar{W}_{-1} (1 + 2 \log(1/\delta) + 2 \log(\log(1 + \eta) + \log T) - 2 \log \log(1 + \eta) + \log(1 + \eta)) \\ &= \bar{W}_{-1} (2 \log(1/\delta) + 2 \log(2 + \log T) + 3 - 2 \log 2) \end{aligned}$$

The second inequality is obtained since $D \leq 1 + \frac{\log T}{\log(1+\eta)}$. The last equality is obtained for the choice $\eta^* = e^2 - 1$, which minimizes $\eta \mapsto \log(1 + \eta) - 2 \log(\log(1 + \eta))$. Since $[T] \subseteq \bigcup_{i \in [D]} [N_i, N_{i+1})$ and $N_a(t)(\hat{\mu}_a(t) - \mu_a) = \sum_{s \in [N_a(t)]} X_{s,a}$ (unit-variance), this yields

$$\mathbb{P} \left(\exists m \leq T, \left| \frac{1}{m} \sum_{s=1}^m X_s \right| \geq \sqrt{\frac{1}{m} \bar{W}_{-1} (2 \log(1/\delta) + 2 \log(2 + \log(T)) + 3 - 2 \log 2)} \right) \leq \delta .$$

Since $3 - 2 \log 2 \leq 2$ and $\bar{W}_{-1}$ is increasing, taking δT^{-s} instead of δ yields

$$\mathbb{P}_\mu \left(\exists t \leq T, \sqrt{N_a(t)} |\hat{\mu}_a(t) - \mu_a| \geq \sqrt{2 \tilde{f}_1(T, \delta)} \right) \leq \delta T^{-s} .$$

Doing a union bound over arms yields the result. ■

Appendix H. Inversion Lemmas and Other Technical Results

Appendix H gathers existing and new technical results which are used for our proofs.

Methodology. Lemma 43 is a standard result to upper bound the expected sample complexity of an algorithm, *e.g.* see Lemma 1 in Degenne et al. (2019). This is a key method extensively used in the literature.

Lemma 43 *Let $(\mathcal{E}_t)_{t \geq K}$ be a sequence of events and $T_\mu(\delta) > K$ be such that for $T \geq T_\mu(\delta)$, $\mathcal{E}_T \subseteq \{\tau_\delta \leq T\}$. Then, $\mathbb{E}_\nu[\tau_\delta] \leq T_\mu(\delta) + \sum_{T > K} \mathbb{P}_\nu(\mathcal{E}_T^c)$.*

Proof Since the random variable τ_δ is positive and $\{\tau_\delta > T\} \subseteq \mathcal{E}_T^c$ for all $T \geq T_\mu(\delta)$, we have $\mathbb{E}_\nu[\tau_\delta] = \sum_{T \geq 0} \mathbb{P}_\nu(\tau_\delta > T) \leq T_\mu(\delta) + \sum_{T \geq T_\mu(\delta)} \mathbb{P}_\nu(\mathcal{E}_T^c)$, which concludes the proof. ■

Inversion results. Lemma 44 gathers properties on the function $\bar{W}_{-1}$, which is used in the literature to obtain concentration results.

Lemma 44 (Jourdan et al. (2023)) *Let $\bar{W}_{-1}(x) = -W_{-1}(-e^{-x})$ for all $x \geq 1$, where W_{-1} is the negative branch of the Lambert W function. The function $\bar{W}_{-1}$ is increasing on $(1, +\infty)$ and strictly concave on $(1, +\infty)$. In particular, $\bar{W}'_{-1}(x) = \left(1 - \frac{1}{\bar{W}_{-1}(x)}\right)^{-1}$ for all $x > 1$. Then, for all $y \geq 1$ and $x \geq 1$, $\bar{W}_{-1}(y) \leq x$ if and only if $y \leq x - \log(x)$. Moreover, for all $x > 1$, $x + \log(x) \leq \bar{W}_{-1}(x) \leq x + \log(x) + \min\left\{\frac{1}{2}, \frac{1}{\sqrt{x}}\right\}$.*

Lemma 45 is an inversion result to upper bound a probability which is implicitly defined based on times that are implicitly defined.

Lemma 45 *Let $\bar{W}_{-1}$ defined in Lemma 44. Let $A, B, C, E, \alpha, \beta > 0$ and $D_{A,B,C,E,\alpha,\beta}(\delta) = \sup\{x \mid x \leq \frac{A}{\alpha} \bar{W}_{-1}(\alpha(\log(1/\delta) + C \log(\beta + \log x) + E)) + B\}$. Then,*

$$\inf\{\delta \mid x > D_{A,B,C,E,\alpha,\beta}(\delta)\} \leq e^E \left(\alpha \frac{x-B}{A}\right)^{1/\alpha} (\beta + \log x)^C \exp\left(-\frac{x-B}{A}\right).$$

Proof Using Lemma 44, direct manipulations yield that

$$\begin{aligned} x > D_{A,B,C,E,\alpha,\beta}(\delta) &\iff \alpha \frac{x-B}{A} > \bar{W}_{-1}(\alpha(\log(1/\delta) + C \log(\beta + \log x) + E)) \\ &\iff \frac{x-B}{A} - \frac{1}{\alpha} \log\left(\alpha \frac{x-B}{A}\right) > \log(1/\delta) + C \log(\beta + \log x) + E \\ &\iff \delta < e^E \left(\alpha \frac{x-B}{A}\right)^{1/\alpha} (\beta + \log x)^C \exp\left(-\frac{x-B}{A}\right). \end{aligned}$$

■

Lemma 46 is an inversion result to upper bound an implicitly defined time.

Lemma 46 *Let $\bar{W}_{-1}$ defined in Lemma 44. Let $A > 0$, $B > 0$ such that $B/A + \log A > 1$ and $C(A, B) = \sup\{x \mid x < A \log x + B\}$. Then, $C(A, B) < h_1(A, B)$ with $h_1(z, y) = z \bar{W}_{-1}(y/z + \log z)$*

Proof Since $B/A + \log A > 1$, we have $C(A, B) \geq A$, hence

$$C(A, B) = \sup\{x \mid x < A \log(x) + B\} = \sup\{x \geq A \mid x < A \log(x) + B\}.$$

Using Lemma 44 yields that

$$x \geq A \log x + B \iff \frac{x}{A} - \log\left(\frac{x}{A}\right) \geq \frac{B}{A} + \log A \iff x \geq A \bar{W}_{-1}\left(\frac{B}{A} + \log A\right).$$

■

Lemma 47 is an inversion result to asymptotically upper bound an implicit time.

Lemma 47 *Let $B > 0$ and $A > 0$*

$$D(\delta) = \sup \left\{ T \mid \frac{T-B}{A} \leq \left(\sqrt{\frac{1}{2} \bar{W}_{-1} (2 \log(2K/\delta) + 4 \log(4 + \log T) + 1)} + \sqrt{3 \log T} \right)^2 \right\}$$

Then, we have $\limsup_{\delta \rightarrow 0} D(\delta)/\log(1/\delta) \leq A$.

Proof Direct manipulations yields that

$$\begin{aligned} \frac{T-B}{A} &> \left(\sqrt{\frac{1}{2} \bar{W}_{-1} (2 \log(2K/\delta) + 4 \log(4 + \log T) + 1)} + \sqrt{3 \log T} \right)^2 \\ \iff 2 \left(\sqrt{\frac{T-B}{A}} - \sqrt{3 \log T} \right)^2 &> \bar{W}_{-1} (2 \log(2K/\delta) + 4 \log(4 + \log T) + 1) \\ \iff \log(1/\delta) < \frac{T-B}{A} - 6 \log T \sqrt{\frac{T-B}{A}} + 3 \log T - \log \left(\sqrt{\frac{T-B}{A}} - \sqrt{3 \log T} \right) \\ &\quad - 2 \log(4 + \log T) - \frac{1 + 3 \log 2}{2} - \log K. \end{aligned}$$

Let $\gamma > 0$. There exists T_γ , which depends on (B, A) , such that

$$\begin{aligned} \frac{T-B}{A} - 6 \log T \sqrt{\frac{T-B}{A}} + 3 \log T - \log \left(\sqrt{\frac{T-B}{A}} - \sqrt{3 \log T} \right) \\ - 2 \log(4 + \log T) - \frac{1 + 3 \log 2}{2} - \log K \geq \frac{T}{A(1+\gamma)}. \end{aligned}$$

Therefore, we have $D(\delta) \leq T_\gamma + C(\delta)$ where $C(\delta) = \sup \left\{ T \mid \frac{T}{A(1+\gamma)} \leq \log(1/\delta) \right\}$. Then, we have

$$\limsup_{\delta \rightarrow 0} \frac{C(\delta)}{\log(1/\delta)} \leq A(1+\gamma) \quad \text{hence} \quad \limsup_{\delta \rightarrow 0} \frac{D(\delta)}{\log(1/\delta)} \leq A(1+\gamma).$$

Letting γ goes to 0 yields the result. ■

Appendix I. Details on the Experimental Study

In this appendix, we detail the benchmark instances in Appendix I.1 and the implementation details in Appendix I.2. Then, we provide supplementary experiments to assess the performance of the APGAI algorithm on the empirical error both for fixed-budget (Appendix I.3) and anytime algorithms (Appendix I.4), as well as on the empirical stopping time (Appendix I.5).

I.1 Benchmark Instances

We detail our real-life instance based on an outcome scoring application in Appendix I.1.1, as well as synthetic instances in Appendix I.1.2. For all the experiments considered below, the parameters and the mean vectors are respectively displayed in Table 5 and Table 6. The numerical values for the difficulties are reported in Table 7.

Name	THR1	THR2	THR3	MED1	MED2	ISA1	NOA1	ISA2	NOA2	REALL
K	10	6	10	5	7	10	5	7	4	18
θ	0.5	0.35	0.5	0.5	1.2	0	0	0	0	0.5
$ \mathcal{A}_\theta $	5	3	3	1	2	5	0	3	0	6

Table 5: Parameters in synthetic and real-life instances.

Arms										
	1	2	3	4	5	6	7	8	9	10
THR1	0.9	0.9	0.9	0.65	0.55	0.45	0.35	0.1	0.1	0.1
THR2	0.6	0.5	0.4	0.3	0.2	0.1	—	—	—	—
THR3	0.55	0.55	0.55	0.45	0.45	0.45	0.45	0.45	0.45	0.45
MED1	0.537	0.469	0.465	0.36	0.34	—	—	—	—	—
MED2	1.8	1.6	1.1	1	0.7	0.6	0.5	—	—	—
ISA1	0.5	0.39	0.28	0.17	0.06	−0.06	−0.17	−0.28	−0.39	−0.50
NOA1	−0.5	−0.62	−0.75	−0.88	−1	—	—	—	—	—
ISA2	1.0	0.5	0.1	−0.1	−0.4	−0.5	−0.6	—	—	—
NOA2	−0.1	−0.4	−0.5	−0.6	—	—	—	—	—	—
REALL	0.800	0.791	0.676	0.545	0.538	0.506	0.360	0.329	0.306	0.274
	11	12	13	14	15	16	17	18		
	0.241	0.203	0.112	0.084	0.081	0.007	−0.018	−0.120		

 Table 6: Synthetic and real-life mean vector instances (scores for the real-life instance are rounded up to the 3rd decimal place).

Name	$H_1(\mu)$	$H_\theta(\mu)$	$\min_{a \in \mathcal{A}_\theta(\mu)} \Delta_a^{-2}$	$\max_{a \in \mathcal{A}_\theta(\mu)} \Delta_a^{-2}$	$K \hat{\Delta}^{-2}$
THR1	926	463	6	400	49
THR2	921	460	16	400	67
THR3	4000	1200	400	400	1000
MED1	2677	730	730	730	1081
MED2	143	9	3	6	14
ISA1	533	266	3	225	23
NOA1	30	—	—	—	55
ISA2	218	104	1	100	5
NOA2	113	—	—	—	399
REALL	29206	29019	11	27778	93
TwoG	$4K$	K	4	4	$4 \mathcal{A}_\theta $

 Table 7: Numerical values of difficulty constants. $H_1(\mu)$ and $H_\theta(\mu)$ as in Eq. (1), $\hat{\Delta} := \max_{a \in \mathcal{A}_\theta} \Delta_a + \min_{b \notin \mathcal{A}_\theta} \Delta_b$.

I.1.1 REAL-LIFE DATASET: OUTCOME SCORING APPLICATION

Premature birth is known to induce moderate to severe neuronal dysfunction in newborns. Human mesenchymal stem cells might help repair and protect neurons from the injury induced by the inflammation. The goal is to determine whether one among possible therapeutic protocols exerts a strong enough positive effect on patients.

In order to answer this question, in collaboration with the PREMSTEM consortium, we have considered a rat model of perinatal neuroinflammation, which mimics brain injuries due to premature birth. Here, the set of arms are considered protocols for the injection of human mesenchymal stem cells (HuMSCs) in rats. Briefly, rat pups received intraperitoneal IL-1 β injections (20 μ g/kg) twice daily from post-natal day (P)1-P4 and once at P5 to model preterm brain injury, and controls received only PBS. Human umbilical cord-derived MSCs (HuMSCs, Chiesi Pharmaceuticals/Lonza) were administered using 18 different protocols testing three doses (20, 50, 125 M cells/kg), three time points (P5, P10, P20), and two delivery routes (intranasal vs intravenous).

Animals were sacrificed 48 hours post-treatment, and microglia were isolated from brain tissue using anti-CD11b/c magnetic beads (Miltenyi Biotec). RNA was extracted using NucleoSpin RNA XS Plus kit, with quality assessed by fragment analyzer (≥ 7 cutoff). Libraries were prepared using TruSeq Stranded mRNA kit and sequenced on NextSeq 500 (75 bp single reads, ≈ 27 M reads/sample). Reads were aligned to rnor6 genome using STAR, processed with samtools and HTSeq-count. Treatment efficacy was evaluated by comparing gene expression signatures between injured-to-treated groups versus injured-to-control groups using characteristic direction differential expression analysis Clark et al. (2014) and cosine similarity scoring ($N = 3$ per protocol). Those score quantifies the effect of each protocol using a cosine score on gene activity measurement profiles between model animals injected with HuMSCs and control animals, which have not been exposed to the inflammation. The cosine score is between -1 and 1. The closer this score is to 1, the more similar the gene activity changes of the treated group are to those of control group. We considered a threshold of $\theta = 0.5$ for treatment efficiency.

Traditional approaches use grid-search with a uniform allocation and select the best cosine score to determine the optimal protocol. Here, to model the stochasticity of the scores that would have been obtained for each protocol in a sequential approach, we applied a Bernoulli instance. In this application observations from arm a for one treatment are drawn from a Bernoulli distribution with mean $\max(\mu_a, 0)$ using the real cosine score of this treatment protocol as μ_a . Bernoulli distributions are here more realistic with respect to our real-life application, while our algorithms can still be applied to this instance, as a Bernoulli distribution is 1/2-sub-Gaussian. One must nevertheless note that in real life, the data generation were carried out sequentially into several batches, with each treatment protocol tested in triplicate, but only once in the same batch. The real stochasticity of such data is unknown and would require costly and heavier laboratory experiments and sequencing.

I.1.2 SYNTHETIC DATASET: GAUSSIAN INSTANCES

Along with the above real-life application described above and in Section 6, we have also considered several Gaussian instances with unit variance.

Mimicking the experiments conducted in Kano et al. (2019), we consider their three synthetic instances, referred to as THR1 (three group setting), THR2 (arithmetically progressive setting) and THR3 (close-to-threshold setting), as well as their two medical instances, referred to as MED1 (dose-finding of secukinumab for rheumatoid arthritis with satisfactory effect) and MED2 (dose-finding of GSK654321 for rheumatoid arthritis with satisfactory effect). While some instances were studied in Kano et al. (2019) for Bernoulli distributions, here we only consider Gaussian instances. For MED2, the Gaussian instances have variance $\sigma^2 = 1.44$.

Mimicking the experiments conducted in Kaufmann et al. (2018), we consider instances whose means are linearly spaced with and without good arms. ISA1 is linearly space between 0.5 and -0.5 with $K = 10$, and NOA1 between -0.5 and -1 with $K = 5$. In addition, we complement those synthetic experiments with two instances with and without good arms, named ISA2 and NOA2.

Finally, as done in Kaufmann et al. (2018), we study the impact of the number of good arms $|\mathcal{A}_\theta|$ among $K = 100$ arms on the performance. We will consider $|\mathcal{A}_\theta| \in \{5k\}_{k \in [19]}$, with $\theta = 0$. In the TWOG instances, we have $\mu_a = 0.5$ for all $a \in \mathcal{A}_\theta$, otherwise $\mu_a = -0.5$. In the LING instances, we have $\mu_a = -0.5$ for all $a \notin \mathcal{A}_\theta$, and the $|\mathcal{A}_\theta|$ good arms have a strictly positive mean which is linearly spaced up to $\max_{a \in \mathcal{A}} \mu_a = 0.5$.

I.2 Implementation Details

We provide details about the implementation of the considered algorithms for the anytime setting (Appendix I.2.1), fixed-budget setting (Appendix I.2.2) and the fixed-confidence setting (Appendix I.2.3). The reproducibility of our experiments is addressed in Appendix I.2.4.

I.2.1 ANYTIME ALGORITHMS

As described in Section 3.2.1, we modify Successive Reject (SR) (Audibert et al., 2010) and Sequential Halving (SH) (Karnin et al., 2013) to tackle GAI. We derived upper bound on the probability of errors of those modified algorithms (Theorems 24 and 25 in Appendix C). As a reminder, SR eliminates one arm with the worst empirical mean at the end of each phase, and SH eliminated half of them but drops past observations between each phase. Within each phase, both algorithms use a round-robin uniform sampling rule on the remaining active arms. SR-G and SH-G return $\hat{a}_T = \emptyset$ when $\hat{\mu}_{a_T}(T) \leq \theta$ and $\hat{a}_T = a_T$ otherwise, where a_T is the arm that would be recommended for the BAI problem, *i.e.* the last arm that was not eliminated. Then, we convert the fixed-budget SH-G and SR-G algorithms into anytime algorithms by using the doubling trick. It considers a sequences of algorithms that are run with increasing budgets $(T_k)_{k \geq 1}$, with $T_{k+1} = 2T_k$ and $T_1 = 2K \lceil \log_2 K \rceil$, and recommend the answer outputted by the last instance that has finished to run. It is well know that the “cost” of doubling is to have a multiplicative factor 4 in front of the hardness constant. The first two-factor is due to the fact that we forget half the observations. The second two-factor is due to the fact that we use the recommendation from the last instance of SH that has finished. The doubling version of SR-G and SH-G are named Doubling SR-G (DSR-G) and Doubling SH (DSH-G).

Compared to SR, the empirical performance of SH suffers from the fact that it drops observation between phases. While the impact of this forgetting step is relatively mild for

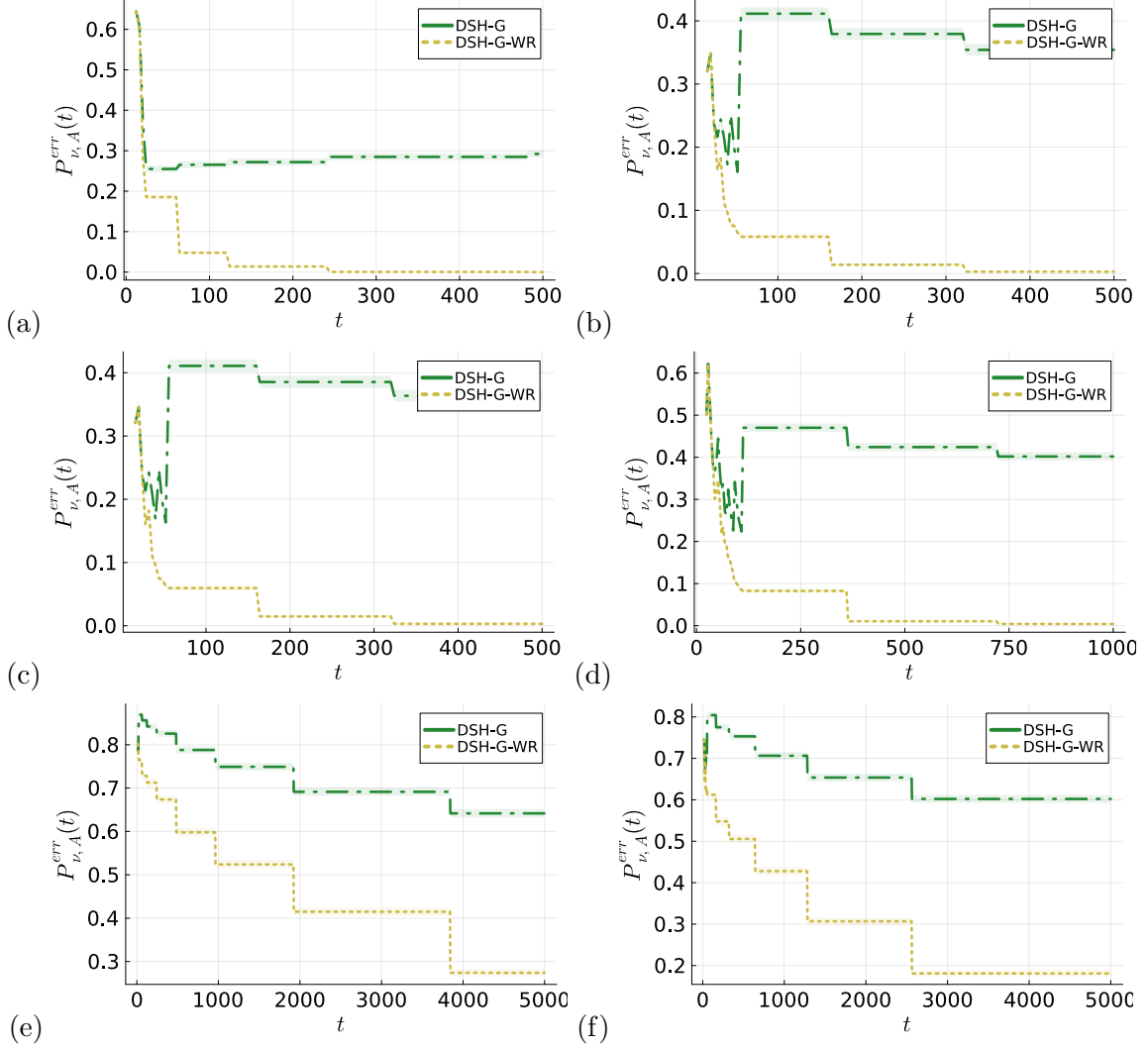


Figure 4: Empirical error on instances (a) NOA1, (b) ISA1, (c) THR1, (d) REALL, (e) MED1 and (f) THR3. “-WR” means that each SH instance keeps all its history instead of discarding it.

BAI where all the arms are sampled linearly, it is larger for GAI since arms are not sampled linearly. In order to assess the impact of this forgetting step, we implement the DSH-G-WR (“without refresh”) algorithm in which each SH-G instance keeps all the observations at the end of each phase. To the best of our knowledge, there is no theoretical analysis of this version of SH, even in the recent analysis of Zhao et al. (2023). Figure 4 highlights the dramatic increase of the empirical error incurred by dropping past observations. This phenomenon occurs in almost all of our experiments, both when $\mathcal{A}_\theta(\mu) = \emptyset$ and when $\mathcal{A}_\theta(\mu) \neq \emptyset$.

I.2.2 FIXED-BUDGET ALGORITHMS

We compare the fixed-budget performances of APGAI with the GAI versions SH-G and SR-G of SH and SR as described in Subsection I.2.1, the uniform round-robin algorithm Unif, and different index policies in the prior knowledge-based meta algorithm PKGAI. Those index policies are defined in Section D and recalled below

$$\begin{aligned} \text{PKGAI}(\text{APT}_P) : \quad & i_a(t) := \sqrt{N_a(t)}(\hat{\mu}_a(t) - \theta) , \\ \text{PKGAI}(\text{UCB}) : \quad & i_a(t) := \hat{\mu}_a(t) - \theta + \sqrt{\frac{\beta(t)}{N_a(t)}} , \\ \text{PKGAI}(\text{Unif}) : \quad & i_a(t) := -N_a(t) , \\ \text{PKGAI}(\text{LCB-G}) : \quad & i_a(t) := \sqrt{N_a(t)}(\hat{\mu}_a(t) - \theta) + \sqrt{\beta(t)} . \end{aligned}$$

Note that, contrary to APGAI and Unif, the other algorithms require the definition of the sampling budget T . For the sake of fairness, we do not use the theoretical value for β as in Theorems 27 and 28. We implement the following confidence width, which is theoretically backed by Lemma 41 in Appendix G.2 (for $s = 0$),

$$\beta(t) = \sigma \sqrt{z(T, \delta)/N_a(t)}, \quad \text{where } z(T, \delta) := \bar{W}_{-1}(2 \log(K/\delta) + 2 \log(2 + \log T) + 2) , \quad (24)$$

using $\delta = 0.01$.

We also consider for algorithms of the PKGAI family the theoretical threshold functions featured in Theorems 27 and 28, *i.e.* relying on problem quantities in practice unavailable at runtime

$$\beta(t) = \sigma \sqrt{q(T, \delta)/N_a(t)}, \quad \text{where } q(T, \delta) := \begin{cases} (T - K)/(4H_1(\mu)) & \text{if } \mathcal{A}_\theta(\mu) = \emptyset \\ (T - K)/(4K\hat{\Delta}^{-2}) & \text{otherwise} \end{cases} , \quad (25)$$

where $\hat{\Delta} := \max_{a \in \mathcal{A}_\theta(\mu)} \Delta_a + \min_{b \notin \mathcal{A}_\theta(\mu)} \Delta_b$.

I.2.3 FIXED-CONFIDENCE ALGORITHMS

Link between GLR stopping and UCB/LCB stopping. In Kano et al. (2019), all the algorithms (HDoC, LUCB-G and APT-G) use a stopping rule which is based on UCB/LCB indices. Namely, they return an arm a as soon as its associated LCB exceeds the threshold θ . Since we consider GAI instead of AllGAI, this condition becomes a stopping rule. The second stopping condition is to return $\emptyset$ as soon as all the arms are eliminated, and an arm is eliminated when its UCB is lower than the threshold θ . Direct manipulations show that the GLR stopping Eq. (6) is equivalent to their stopping provided that the UCB and LCB are using the same stopping threshold for the bonuses, *i.e.*

$$\begin{aligned} \max_{a \in \mathcal{A}} W_a^+(t) \geq \sqrt{2c(t, \delta)} & \iff \exists a \in \mathcal{A}, \quad \hat{\mu}_a(t) - \sqrt{\frac{2c(t, \delta)}{N_a(t)}} \geq \theta , \\ \min_{a \in \mathcal{A}} W_a^-(t) \geq \sqrt{2c(t, \delta)} & \iff \forall a \in \mathcal{A}, \quad \hat{\mu}_a(t) + \sqrt{\frac{2c(t, \delta)}{N_a(t)}} \leq \theta . \end{aligned}$$

In Kano et al. (2019), they consider bonuses that only depend on the pulling count $N_a(t)$ instead of depending on the global time t . This ensures that the UCB remains constant

once the arm has been eliminated. In contrast, using a UCB which depends on the global time t (such as our stopping threshold in Eq. (7)) implies that this elimination step does not ensure that the condition on this arm still hold at stopping time. Mathematically, they use the following UCB/LCB, $\hat{\mu}_a(t) \pm \sqrt{2\Lambda_a(t, \delta)/N_a(t)}$ where $\Lambda_a(t, \delta) = \log(4K/\delta) + 2 \log N_a(t)$. Since Kano et al. (2019) consider Bernoulli distributions which are 1/2-sub-Gaussian, we modified the bonuses to match the ones for 1-sub-Gaussian (by using that the proper scaling is in $\sqrt{2\sigma^2}$).

While both stopping threshold c and $(\Lambda_a)_{a \in \mathcal{A}}$ have the same dominating δ -dependency in $\log(1/\delta)$, it is worth noting that the time dependency of c is significantly better since $c(t, \delta) \sim_{t \rightarrow +\infty} 2 \log \log t$. Ignoring the δ -dependent terms and the constant, we have a lower bonus as long as $N_a(t) \gtrsim \log t$. For a fair comparison, we will use the stopping threshold in Eq. (7) for the UCB/LCB used by HDoC and LUCB-G (both in the sampling and stopping rule) instead of the larger bonuses $(\Lambda_a)_{a \in \mathcal{A}}$ considered in Kano et al. (2019).

Limits of existing algorithms. The APT-G algorithm introduced in Kano et al. (2019) samples $a_{t+1} = \arg \min_{a \in \mathcal{A}_t} \sqrt{N_a(t)} |\hat{\mu}_a(t) - \theta|$, where $\mathcal{A}_t$ is the set of active arms. This index policy is tailored for the Thresholding setting, where one needs to classify all the arms as above or below the threshold θ . Intuitively, a good algorithm for Thresholding will perform poorly on the GAI setting since it must pay $H_1(\mu)$ even when $\mathcal{A}_\theta$. This is confirmed by the experiments in Kano et al. (2019), as well as our own experiments. Since it is not competitive, we omitted its empirical performance from our experiments.

The Sticky Track-and-Stop (S-TaS) algorithm introduced in Degenne and Koolen (2019) admits a computationally tractable implementation for GAI. To the best of our knowledge, this is one of the few setting where this holds, *e.g.* it is not tractable for ε -BAI. The major limitation of S-TaS lies in its dependency on an ordering $\mathcal{O}$ on the set of candidate answers $\mathcal{A} \cup \{\emptyset\}$. Informally, S-TaS computes a set of admissible answer based on a confidence region on the true mean, and sticks to the answer with the lowest ranking in the ordering $\mathcal{O}$. Then, S-TaS samples according to the optimal allocation for this specific answer. Depending on the choice of this ordering, the empirical performance can change drastically, especially for instances such that $\mathcal{A}_\theta(\mu) \neq \emptyset$. We consider two orderings to illustrate this. The ASC considers the ordering $\mathcal{O}$ such that $o_a = a$ for all $a \in \mathcal{A}$, and $a_{K+1} = \emptyset$. The DESC considers the ordering $\mathcal{O}$ such that $o_a = K - a + 1$ for all $a \in \mathcal{A}$, and $a_{K+1} = \emptyset$. In Table 8, we can see that S-TaS performs considerably better for ASC compared to DESC. This can be explained by the fact that in all our instances the means are ordered, so that lower indices correspond to higher mean. Since higher means are easier to verify, this explains the improved performance for ASC.

The Murphy Sampling (MS) algorithm introduced in Kaufmann et al. (2018) uses a rejection step on top of a Thompson Sampling procedure. For Gaussian instances, the posterior distribution $\Pi_{t,a}$ of the arm $a \in \mathcal{A}$ for the improper prior $\Pi_{0,a} = \mathcal{N}(0, +\infty)$ is $\Pi_{t,a} = \mathcal{N}(\hat{\mu}_a(t), 1/\sqrt{N_a(t)})$. Let $\Pi_t = (\Pi_{t,a})_{a \in \mathcal{A}}$. Then, MS samples $\lambda \sim \Pi_t$ until $\max_{a \in \mathcal{A}} \lambda_a > \theta$, and samples arm $\arg \max_{a \in \mathcal{A}} \lambda_a$ for this realization. This rejection steps is equivalent to conditioning on the fact that $\mathcal{A}_\theta(\mu) \neq \emptyset$. As noted in Kaufmann et al. (2018), this rejection step can be computationally costly when $\mathcal{A}_\theta(\mu) = \emptyset$. Intuitively, we need to draw many vectors before observing λ such that $\mathcal{A}_\theta(\lambda) \neq \emptyset$ once the posterior Π_t has converged close to the Dirac distribution on μ when $\mathcal{A}_\theta(\mu) = \emptyset$. Empirically, we observed

Ordering	THR1	THR2	THR3	MED1	MED2	ISA1	ISA2	REALL
ASC	183	435	11787	20488	114	120	33	341
	± 68	± 163	± 4539	± 7972	± 41	± 41	± 10	± 122
DESC	20574	19960	71057	60275	3087	16469	4539	—
	± 5835	± 5885	± 11684	± 16112	± 1293	± 4680	± 1434	—

Table 8: Empirical stopping time ($\pm$ standard deviation) of Sticky Track-and-Stop depending on the ordering on the set of candidate answers $\mathcal{A} \cup \{\emptyset\}$. “—” means that the algorithm didn’t stop after 10^5 steps.

this phenomenon on the NoA2 instance. While all the other algorithms has a CPU running time of the order of 10 milliseconds, MS reached a CPU running time of 10^5 milliseconds.

We consider the Track-and-Stop (TaS) algorithm for GAI. It is direct to adapt the ideas of the original Track-and-Stop introduced in Garivier and Kaufmann (2016) for BAI. When $\max_{a \in \mathcal{A}} \hat{\mu}_a(t) \geq \theta$, the optimal allocation $w^*(\hat{\mu}(t))$ to be tracked is a Dirac in $\arg \max_{a \in \mathcal{A}} \hat{\mu}_a(t)$. Otherwise, using the proof of Lemma 1, the optimal allocation is $w^*(\hat{\mu}(t))$, which is defined as $w^*(\hat{\mu}(t))_a \propto (\hat{\mu}_a(t) - \theta)^{-2}$. On top of the C-Tracking procedure used to target the average optimal allocation, Track-and-Stop relies on a forced exploration procedure which samples under-sampled arms, *i.e.* arms in $\{a \in \mathcal{A} \mid N_a(t) \leq \sqrt{t} - K/2\}$. Without the forced exploration, TaS would have worse empirical performance since it would be too greedy.

As mentioned in Sections 2 and 5, the BAEC meta-algorithm is only defined for asymmetric threshold $\theta_U > \theta_L$. Mathematically, it uses the following UCB/LCB indices

$$\begin{aligned}
 \hat{\mu}_a(t) + \sqrt{\frac{2\Lambda_a^+(t, \delta)}{N_a(t)}} \quad & \text{where} \quad \Lambda_a^+(t, \delta) = \log(N(\delta)/\delta) \quad \text{and} \\
 N(\delta) &:= \left\lceil \frac{2e}{(e-1)(\theta_U - \theta_L)^2} \log\left(\frac{2\sqrt{K}}{(\theta_U - \theta_L)^2 \delta}\right) \right\rceil, \\
 \hat{\mu}_a(t) - \sqrt{\frac{2\Lambda_a^-(t, \delta)}{N_a(t)}} \quad & \text{where} \quad \Lambda_a^-(t, \delta) = \log(\sqrt{K}N(\delta)/\delta).
 \end{aligned}$$

In the GAI setting, those indices will infinite, hence BAEC is not defined properly. Instead of using asymmetric threshold, one could simply use symmetric ones which are independent of $(\theta_U - \theta_L)^{-2}$. In that case, BAEC coincide with the HDoC and LUCB-G algorithms introduced in Kano et al. (2019).

1.2.4 REPRODUCIBILITY

Experiments on fixed-budget empirical error. The benchmark was implemented in Python 3.9, and run on a personal computer (configuration: processor Intel Core i7 – 8750H, 12 cores @2.20GHz, RAM 16GB). The code, along with assets for the real-life instance –where the exact treatment protocols have been replaced with placeholder names– are available in a .zip file under MIT (code) and Creative Commons Zero (assets) licenses. Commands

which have generated plots and tables in this paper can be found in the Bash file named `experiments.sh`.

Experiments on anytime empirical error and empirical stopping time. Our code is implemented in `Julia 1.9.0`, and the plots are generated with the `StatsPlots.jl` package. Other dependencies are listed in the `Readme.md`. The `Readme.md` file also provides detailed julia instructions to reproduce our experiments, as well as a `script.sh` to run them all at once. The general structure of the code (and some functions) is taken from the `tidnabbil` library. This library was created by Degenne et al. (2019), see <https://bitbucket.org/wmkoolen/tidnabbil>. No license were available on the repository, but we obtained the authorization from the authors. Our experiments are conducted on an institutional cluster with 4 Intel Xeon Gold 5218R CPU with 20 cores per CPU and an x86_64 architecture.

I.3 Supplementary Results on Fixed-budget Empirical Error

Recall that we use here the prior-knowledge-agnostic threshold functions defined in Equation Eq. (24). We report in Figures 5, 6, 7, 8 and 9 the empirical error curves for all algorithms described in Subsection I.2.2 on real-life instance `REALL`, along with two synthetic instances `ISA1` and `ISA2` where $\mathcal{A}_\theta \neq \emptyset$, and two other instances where $\mathcal{A}_\theta = \emptyset$ (`NOA1` and `NOA2`). Results are averaged over 1,000 runs. In plots, we display the mean empirical error and shaded area corresponds to Wilson confidence intervals (Wilson, 1927) with confidence 95%. Those Wilson confidence intervals are also reported on the corresponding tables.

In the real-life instance along with the instances with no good arms, uniform samplings (`SH-G`, `SR-G`, `Unif` and `PKGAI(Unif)`) are noticeably less efficient at detecting the presence or absence of good arms, contrary to the adaptive strategies. Moreover, except for instance `ISA2`, `APGAI` actually performs as well as more complex, elimination-based algorithms `PKGAI(★)`, while allowing early stopping as well. Perhaps unsurprisingly, the performance of `APGAI` are closely related to those of `PKGAI(APTP)`, as both algorithms share the same sampling rule. In all three instances, although `PKGAI` has unrealistic assumptions in its theoretical guarantees (Theorems 27 and 28), its performance actually turns out to be the best of all algorithms. In particular, using the UCB sampling rule seems to be the most efficient. This shows that adaptive strategies can fare better than uniform samplings, which are more present in prior works in fixed-budget.

Remark 48 *Our experiments below highlight that an algorithm which only aims at allocating most of the budget to the best arm (e.g. based on UCB indices) would be efficient on instances with a good arm with large gap. However, it would be heavily penalized in instances where there are no good arms, or in instances where the gap between the good and the bad arms is small.*

Performance on the real-life application. We report empirical errors at $T = 200$ in Table 9, at which budget empirical errors for all algorithms seem to converge (see Figure 5).

Performance on synthetic datasets ($\mathcal{A}_\theta \neq \emptyset$). We report empirical errors at $T = 700$ in Tables 10 and 11, at which budget empirical errors for all algorithms seem to converge (see Figures 6 and 7). In the figures, the curves of `PKGAI(APTP)` and `PKGAI(LCB-G)` overlap.

Performance on synthetic datasets ($\mathcal{A}_\theta = \emptyset$). We report empirical errors at $T = 150$ in Table 12 and $T = 700$ in Table 13, at which budget empirical errors for all algorithms

seem to converge (see Figures 8 and 9). In the figures, the curves of $\text{PKGAI}(\text{APT}_P)$ and $\text{PKGAI}(\text{LCB-G})$ overlap.

Algorithm	Error	Conf. intervals	
APGAI	0.001	2.10^{-4}	6.10^{-3}
$\text{PKGAI}(\text{APT}_P)$	0.004	2.10^{-3}	0.01
$\text{PKGAI}(\text{LCB-G})$	0.001	2.10^{-4}	6.10^{-3}
$\text{PKGAI}(\text{UCB})$	0.000	0.00	4.10^{-3}
$\text{PKGAI}(\text{Unif})$	0.001	2.10^{-4}	6.10^{-3}
SH-G	0.005	2.10^{-3}	1.10^{-2}
SR-G	0.002	5.10^{-4}	7.10^{-3}
Unif	0.000	0.00	4.10^{-3}

Table 9: Error across 1,000 runs at $T = 200$.

Algorithm	Error	Conf. intervals	
APGAI	0.003	1.10^{-3}	9.10^{-3}
$\text{PKGAI}(\text{APT}_P)$	0.004	2.10^{-3}	0.01
$\text{PKGAI}(\text{LCB-G})$	0.004	2.10^{-3}	0.01
$\text{PKGAI}(\text{UCB})$	0.000	0.00	4.10^{-3}
$\text{PKGAI}(\text{Unif})$	0.000	0.00	4.10^{-3}
SH-G	0.000	0.00	4.10^{-3}
SR-G	0.000	0.00	4.10^{-3}
Unif	0.000	0.00	4.10^{-3}

Table 10: Error across 1,000 runs at $T = 700$.

Algorithm	Error	Conf. intervals	
APGAI	0.000	0.00	4.10^{-3}
$\text{PKGAI}(\text{APT}_P)$	0.000	0.00	4.10^{-3}
$\text{PKGAI}(\text{LCB-G})$	0.000	0.00	4.10^{-3}
$\text{PKGAI}(\text{UCB})$	0.000	0.00	4.10^{-3}
$\text{PKGAI}(\text{Unif})$	0.000	0.00	4.10^{-3}
SH-G	0.000	0.00	4.10^{-3}
SR-G	0.000	0.00	4.10^{-3}
Unif	0.000	0.00	4.10^{-3}

Table 11: Error across 1,000 runs at $T = 700$.

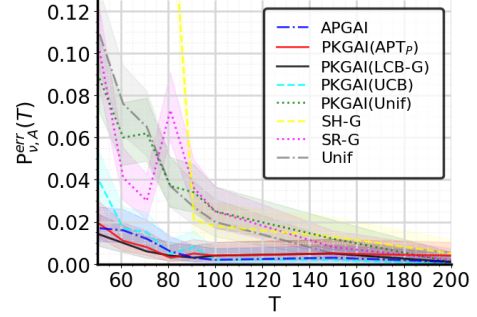


Figure 5: Empirical error (REALL).

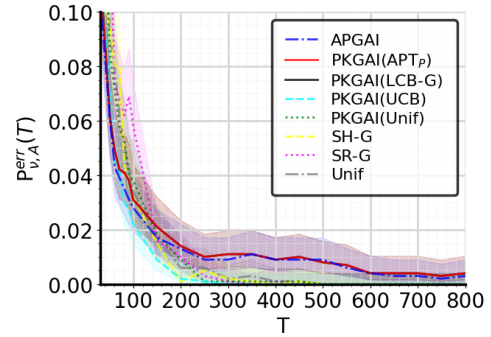


Figure 6: Empirical error (ISA1).

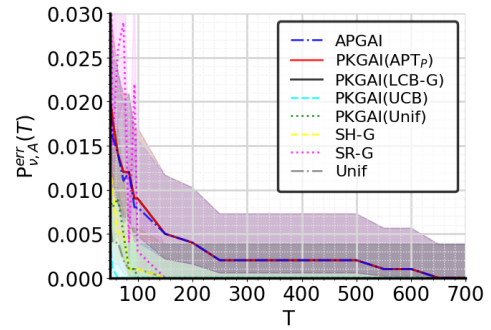


Figure 7: Empirical error (ISA2).

Algorithm	Error	Conf. intervals	
APGAI	0.000	0.00	4.10^{-3}
PKGAI(APT_P)	0.000	0.00	4.10^{-3}
PKGAI(LCB-G)	0.000	0.00	4.10^{-3}
PKGAI(UCB)	0.000	0.00	4.10^{-3}
PKGAI(Unif)	0.002	5.10^{-4}	7.10^{-3}
SH-G	0.000	0.00	4.10^{-3}
SR-G	0.007	3.10^{-3}	0.01
Unif	0.005	2.10^{-3}	0.01

 Table 12: Error across 1,000 runs at $T = 150$.

Algorithm	Error	Conf. intervals	
APGAI	0.002	5.10^{-4}	7.10^{-3}
PKGAI(APT_P)	0.002	5.10^{-4}	7.10^{-3}
PKGAI(LCB-G)	0.002	5.10^{-4}	7.10^{-3}
PKGAI(UCB)	0.007	3.10^{-3}	0.01
PKGAI(Unif)	0.021	0.01	0.03
SH-G	0.018	0.01	0.03
SR-G	0.127	0.11	0.15
Unif	0.084	0.07	0.10

 Table 13: Error across 1,000 runs at $T = 700$.

On prior-knowledge based threshold functions. For the sake of completeness, we have also iterated those experiments using the prior-knowledge threshold functions (in practice, they are unavailable) in algorithms belonging to the PKGAI family.

In those figures, when plotting the empirical curves for PKGAI-like algorithms, we also report on the same plot the corresponding curve for our contribution APGAI (which is not expected to be different from the one on the left-hand plot, as the change in thresholds only affects PKGAI-like algorithms). As expected, the use of the prior-knowledge-based thresholds considerably improves the performance of PKGAI algorithms across most of the considered instances (except for REALL in Figure 10 where the performance of index policies APT_P and LUCB-G is severely impacted). However, more specifically in instances ISA2 (Figure 13), NoA1 (Figure 12), ISA1 (Figure 11) and REALL (Figure 10), we can notice that the gap in performance between APGAI and algorithms from the PKGAI (and more surprisingly, PKGAI(Unif)) is not very large. This means that the theoretical gap in Table 2 does not necessarily translate into practice and highlights the need for more refined tools for the analysis of these algorithms.

I.4 Supplementary Results on Anytime Empirical Error

Since we are interested in the empirical error holding for any time, we only consider the anytime algorithms: APGAI, Unif, DSR-G and DSH-G. As mentioned in Appendix I.2, we

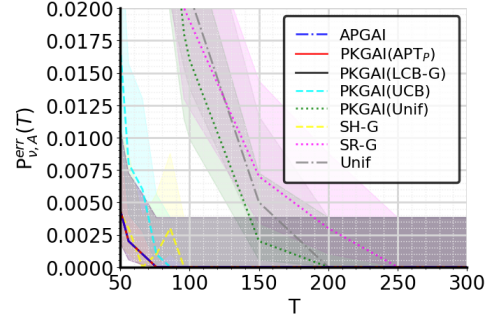


Figure 8: Empirical error (NoA1).

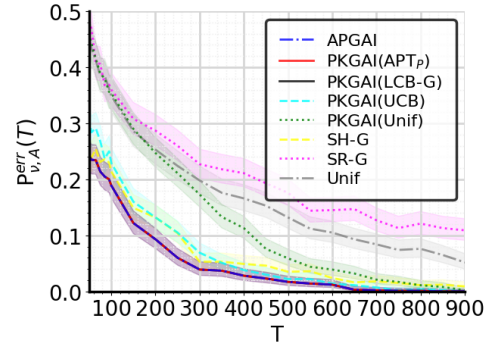


Figure 9: Empirical error (NoA2).

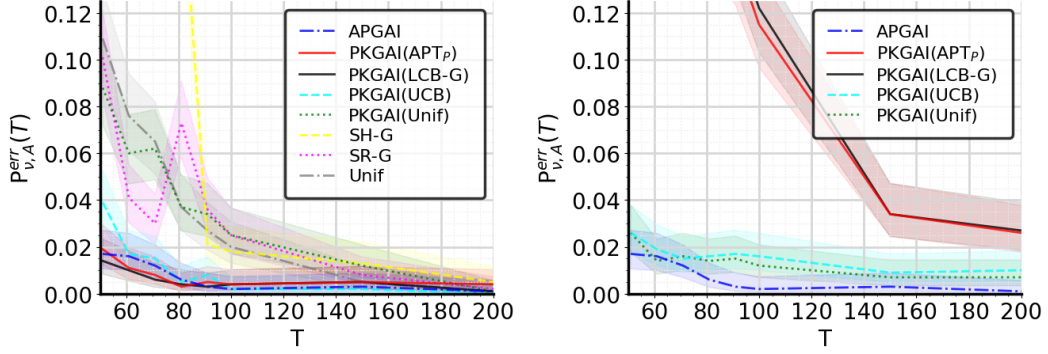


Figure 10: Empirical error on instance REALL. **Left:** with threshold functions from Equation Eq. (24). **Right:** with prior knowledge thresholds in Equation Eq. (25).

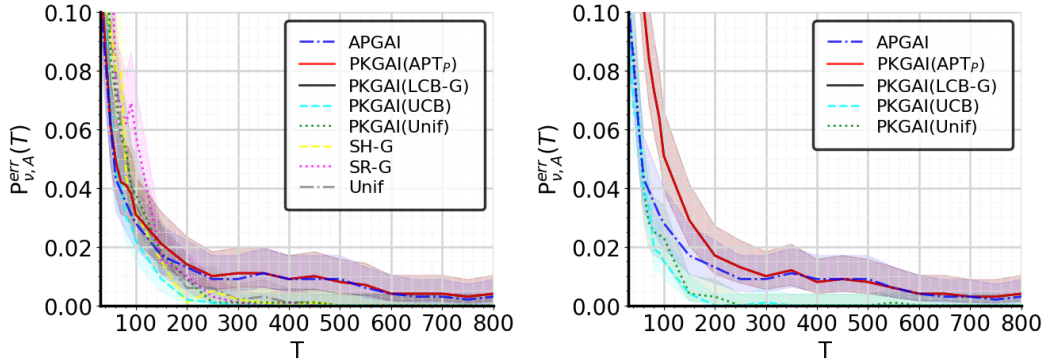


Figure 11: Empirical error on instance ISA1. **Left:** with threshold functions from Equation Eq. (24). **Right:** with prior knowledge thresholds in Equation Eq. (25).

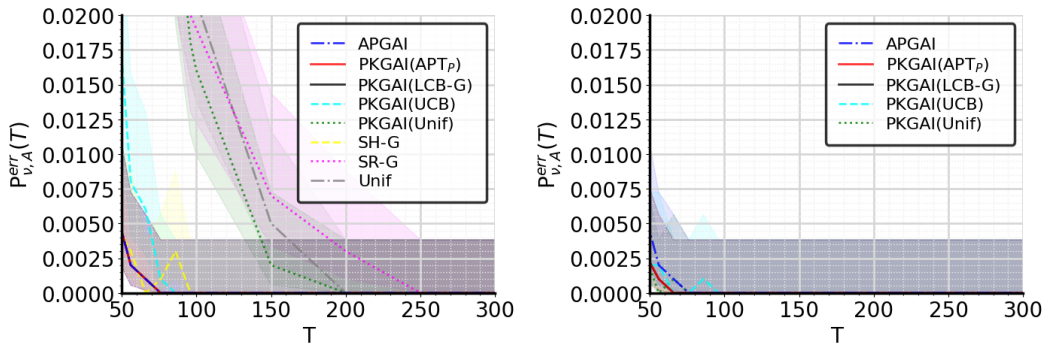


Figure 12: Empirical error on instance NOA1. **Left:** with threshold functions from Equation Eq. (24). **Right:** with prior knowledge thresholds in Equation Eq. (25).

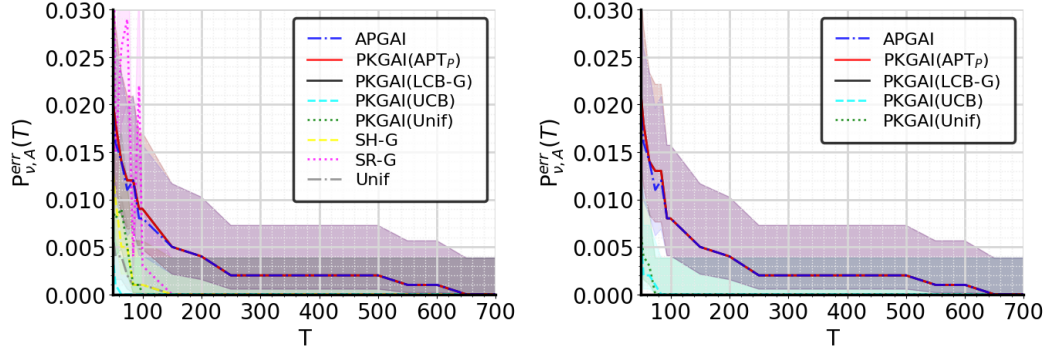


Figure 13: Empirical error on instance Isa2. **Left:** with threshold functions from Equation Eq. (24). **Right:** with prior knowledge thresholds in Equation Eq. (25).

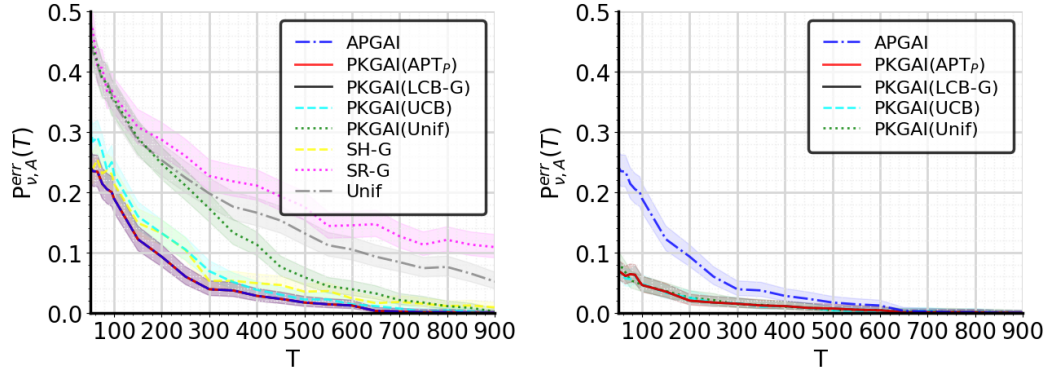


Figure 14: Empirical error on instance NOA2. **Left:** with threshold functions from Equation Eq. (24). **Right:** with prior knowledge thresholds in Equation Eq. (25).

consider the implementation DSH-G-WR (“without refresh”) which keeps all the history within each SH instance. We repeat our experiments over 10000 runs. We display the mean empirical error and shaded area corresponds to Wilson confidence intervals (Wilson, 1927) with confidence 95%.

In summary, our experiments show that APGAI significantly outperforms all the other anytime algorithms when $\mathcal{A}_\theta(\mu) = \emptyset$. When $\mathcal{A}_\theta(\mu) \neq \emptyset$, APGAI has always better performance than DSR-G and DSH-G, and it performs on par with Unif. Our empirical results suggest that APGAI enjoys better empirical performance than suggested by the theoretical guarantees summarized in Table 2.

No good arms. Since APGAI has arguably the best theoretical guarantees when $\mathcal{A}_\theta(\mu) = \emptyset$, we expect it to have superior empirical performance on the instances NOA1 and NOA2. Figure 15 validates empirically that APGAI significantly outperform all the other anytime algorithms by a large margin. While Unif has the “worse” theoretical guarantees in Table 2, the empirical study shows that it outperforms both DSR-G and DSH-G-WR. This phenomenon is mainly due to the doubling trick. Converting a fixed-budget algorithm to an anytime algorithm forces the algorithm to forget past observations, hence considerably impacting the empirical performance.

Varying number of good arms. In Figure 16, we study the impact of an increased number of good arms on the empirical error. While Table 2 suggests that APGAI is not benefiting from increased $|\mathcal{A}_\theta(\mu)|$, we see that the empirical error is decreasing significantly as $|\mathcal{A}_\theta(\mu)|$ increases. This suggests that better theoretical guarantees could be obtained when $\mathcal{A}_\theta(\mu) \neq \emptyset$. It is an interesting direction for future research to show an asymptotic rate featuring a complexity inversely proportional to $|\mathcal{A}_\theta(\mu)|$. In addition, we observe that APGAI outperforms all the other anytime algorithms by a large margin. Intuitively, APGAI is greedy enough when $\mathcal{A}_\theta(\mu) \neq \emptyset$ to avoid sampling the arms which are not good.

Good arms with similar gaps. In light of Table 2, one might expect that APGAI has worse empirical performance when $\mathcal{A}_\theta(\mu) \neq \emptyset$ compared to other anytime algorithms. To assess this fact empirically, we first consider instances where the good arms have similar gaps, *e.g.* THR3 and MED1. In Figure 17, we see that APGAI is better than Unif on THR3, but worse on MED1. In both cases, APGAI outperforms both DSR-G and DSH-G-WR. Therefore, we see that APGAI has better empirical performance compared to the ones suggested by the theoretical guarantees summarized in Table 2.

Good arms with dissimilar gaps. In Figure 18, we consider instances where $\mathcal{A}_\theta(\mu) \neq \emptyset$ and good arms have dissimilar gaps. Overall, APGAI always performs better than DSR-G and DSH-G-WR. While Unif seems to outperform APGAI on some instances (*e.g.* THR2 and MED2), it has worse performance on other instances (*e.g.* REALL and THR1).

I.5 Supplementary Results on Empirical Stopping Time

While APGAI is designed to tackle anytime GAI, it also enjoys theoretical guarantees in the fixed-confidence setting when combined with the GLR stopping rule Eq. (6) with stopping threshold Eq. (7). According to Table 4, we expect that APGAI has good empirical performance when $\mathcal{A}_\theta(\mu) = \emptyset$, and sub-optimal ones when $\mathcal{A}_\theta(\mu) \neq \emptyset$. Since we are interested in the empirical performance for moderate regime of confidence, we take $\delta = 0.01$ in the

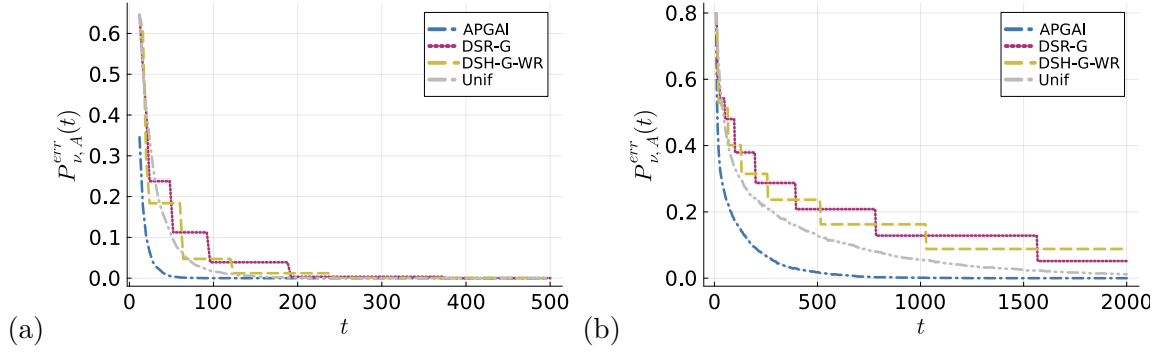


Figure 15: Empirical error on instances (a) NoA1 and (b) NoA2. “-WR” means that each SH instance keeps all its history instead of discarding it.

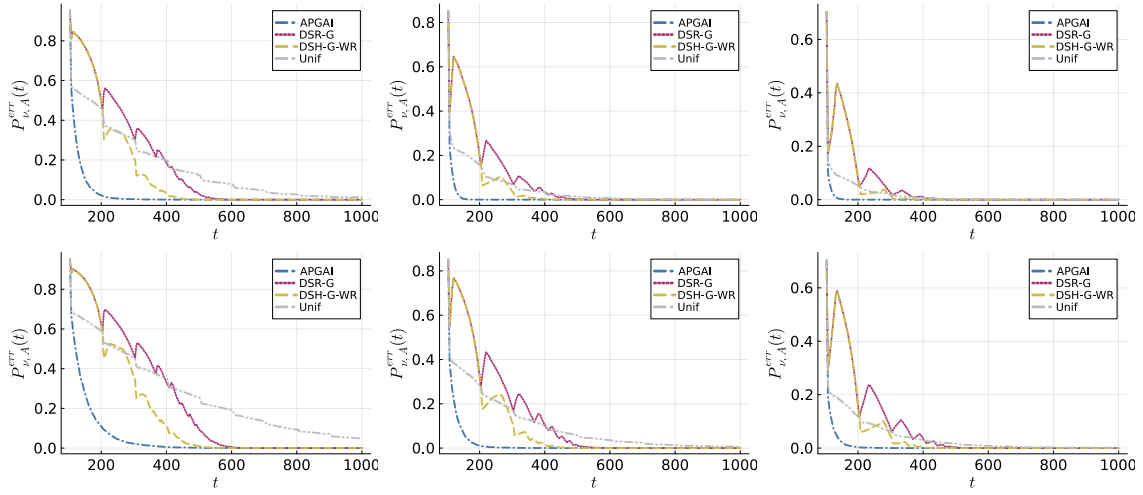


Figure 16: Empirical error for varying number of good arms $|\mathcal{A}_{\theta}(\mu)| \in \{5, 15, 30\}$ (left to right) among $K = 100$ arms on instances (top) TwoG and (bottom) LING. “-WR” means that each SH instance keeps all its history instead of discarding it.

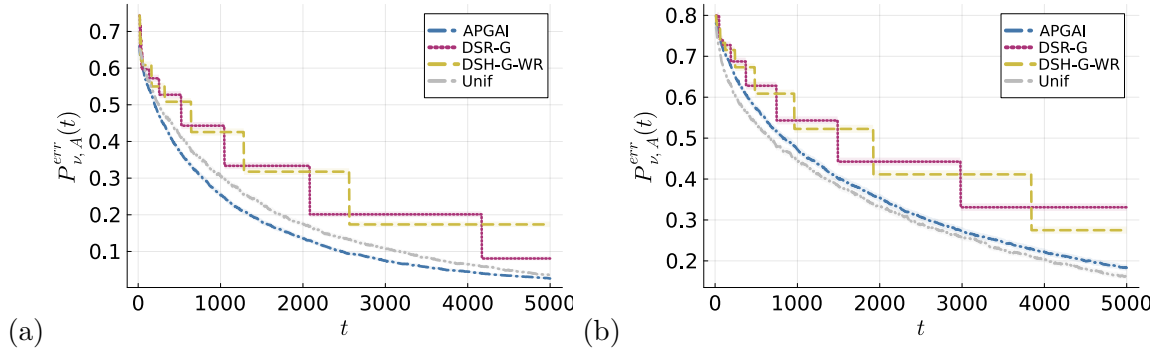


Figure 17: Empirical error on instances (a) THR3 and (b) MED1. “-WR” means that each SH instance keeps all its history instead of discarding it.

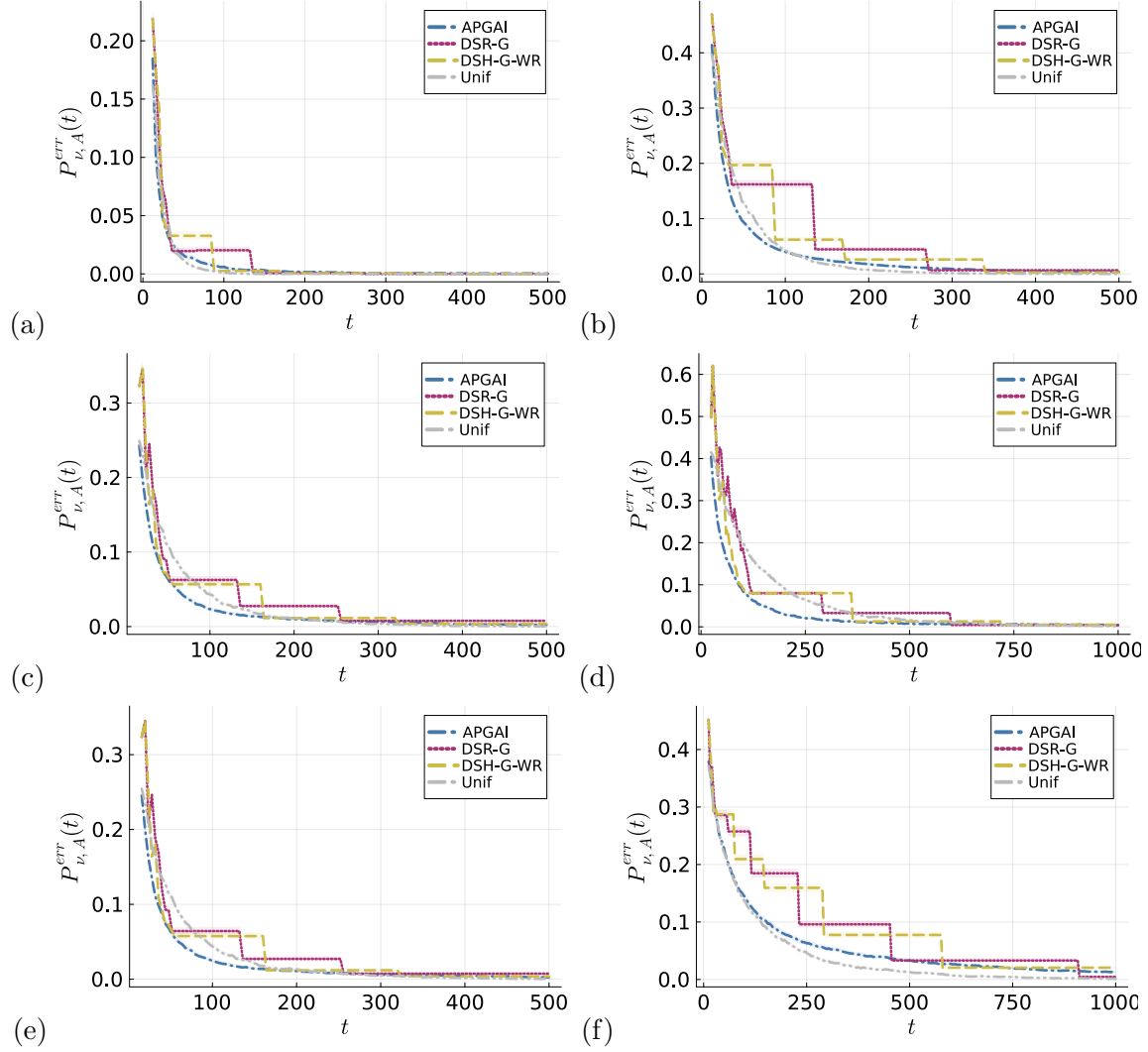


Figure 18: Empirical error on instances (a) ISA2, (b) MED2, (c) ISA1, (d) REALL, (e) THR1 and (f) THR2. “-WR” means that each SH instance keeps all its history instead of discarding it.

following. We repeat our experiments over 1000 runs. We either display the boxplots or the mean with standard deviation as shaded area.

In summary, our experiments show that APGAI performs on par with all the other fixed-confidence algorithms when $\mathcal{A}_\theta(\mu) = \emptyset$. When $\mathcal{A}_\theta(\mu) \neq \emptyset$, APGAI has good performance only when the good arms have similar gaps. Importantly, its performance does not scale linearly with $|\mathcal{A}_\theta(\mu)|$ as suggested by Table 4. When good arms have dissimilar gaps, APGAI can suffer from large outliers due to the greediness of its sampling rule. Finally, we show a simple way to circumvent this limitation by adding forced exploration on top of APGAI.

No good arms. Since APGAI is asymptotically optimal when $\mathcal{A}_\theta(\mu) = \emptyset$, we expect it to perform well on the instances NOA1 and NOA2. Figure 19 shows that APGAI has comparable performance with existing fixed-confidence GAI algorithms on such instances, and that uniform sampling performs poorly.

Varying number of good arms. In Figure 20, we study the impact of an increased number of good arms on the empirical error. While Table 4 suggests that APGAI is suffering from increased $|\mathcal{A}_\theta(\mu)|$ due to the dependency in $H_\theta(\mu)$, we see that the empirical stopping time remains the same when $|\mathcal{A}_\theta(\mu)| \in \{5k\}_{k \in [19]}$. Therefore, Figure 20 empirically validates our theoretical intuition that APGAI can achieve an asymptotic upper bound of the order $2 \max_{a \in \mathcal{A}_\theta(\mu)} \Delta_a^{-2} \log(1/\delta)$ as discussed in Appendix F.3.1. On the LING, we also observe that APGAI can have large outliers due to the good arms with small gaps (see below for more details).

Good arms with similar gaps. When $\mathcal{A}_\theta(\mu) \neq \emptyset$ and good arms have similar means, Table 4 suggests that APGAI could be competitive with other algorithms. Figure 21 validates this observation empirically. On the THR3 instance, APGAI achieves better performance than the other fixed-confidence algorithms, except for Track-and-Stop which has similar performance.

Good arms with dissimilar gaps. In Figure 22, we consider instances where $\mathcal{A}_\theta(\mu) \neq \emptyset$ and good arms have dissimilar gaps. Table 4 suggests that APGAI can have poor empirical performance on such instances. Empirically, we see that APGAI can suffer from very large outliers on such instances. Depending on the initial draws, the greedy sampling rule of APGAI can focus on a good arm with small gap Δ_a instead of verifying a good arm with large gap Δ_a . Since those arms are significantly harder to verify, APGAI will incur a large empirical stopping time in that case. This explains why the distribution of the empirical stopping time has a heavy tail with large outliers. A right-skewed stopping time distribution is not a desirable property in practical application, APGAI is not a good fixed-confidence GAI algorithm on instances with good arms have dissimilar gaps.

In Figure 23, we study the impact of a varying confidence level on instances where APGAI suffers from large outliers. For a fair comparison, we only consider fixed-confidence algorithm whose sampling rule is independent of δ (*i.e.* excluding LUCB-G and HDoC). As expected, the large outliers phenomenon also increases when δ decreases.

Fixing APGAI with forced exploration. In the fixed-confidence setting, APGAI can suffer from large outliers when good arms have dissimilar means since it can greedily focus on good arms with small gaps. To fix this limitation, we propose to add forced exploration on top of APGAI, which we refer to as APGAI-FE. Let $\mathcal{U}_t = \{a \in \mathcal{A} \mid N_a(t) \leq \sqrt{t} - K/2\}$. When $\mathcal{U}_t \neq \emptyset$, we pull $a_{t+1} \in \arg \min_{a \in \mathcal{U}_t} N_a(t)$. When $\mathcal{U}_t = \emptyset$, we pull according to APGAI sampling rule.

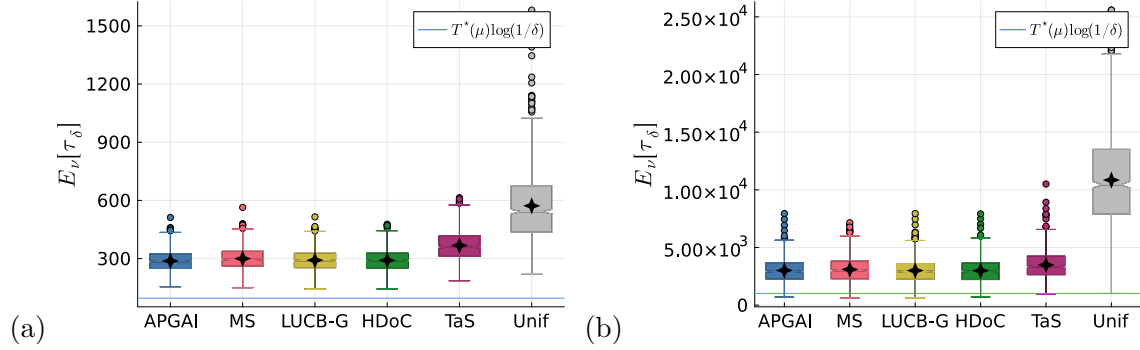


Figure 19: Empirical stopping time ($\delta = 0.01$) on instances (a) NoA1 and (b) NoA2. “MS” is Murphy Sampling, “TaS” is Track-and-Stop and “Unif” is round-robin uniform sampling.

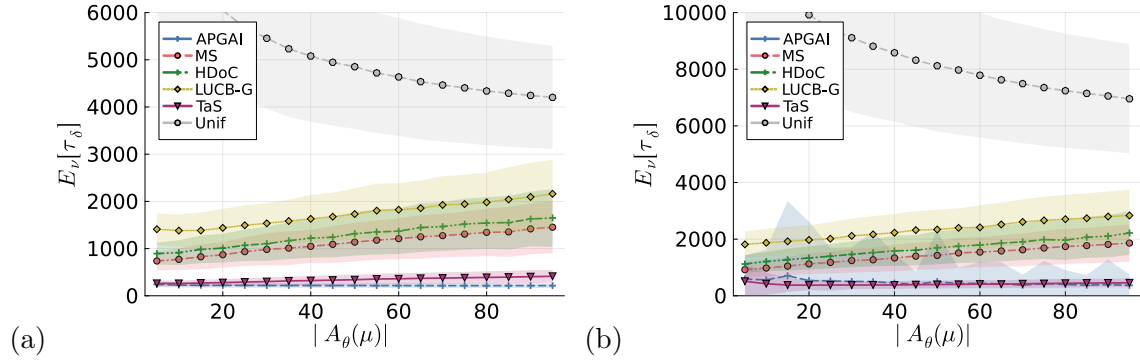


Figure 20: Empirical stopping time ($\delta = 0.01$) for varying number of good arms $|A_\theta(\mu)| \in \{5k\}_{k \in [19]}$ among $K = 100$ arms on instances (a) TWOG and (b) LING. “MS” is Murphy Sampling, “TaS” is Track-and-Stop and “Unif” is round-robin uniform sampling.

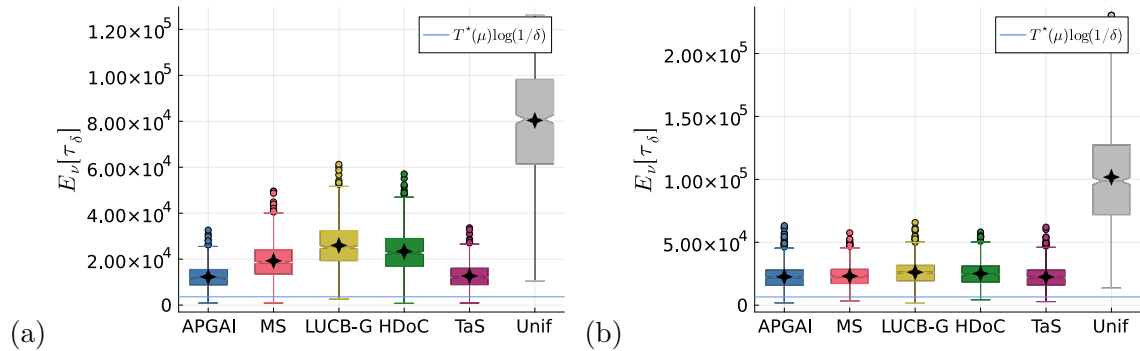


Figure 21: Empirical stopping time ($\delta = 0.01$) on instances (a) THR3 and (b) MED1. “MS” is Murphy Sampling, “TaS” is Track-and-Stop and “Unif” is round-robin uniform sampling.

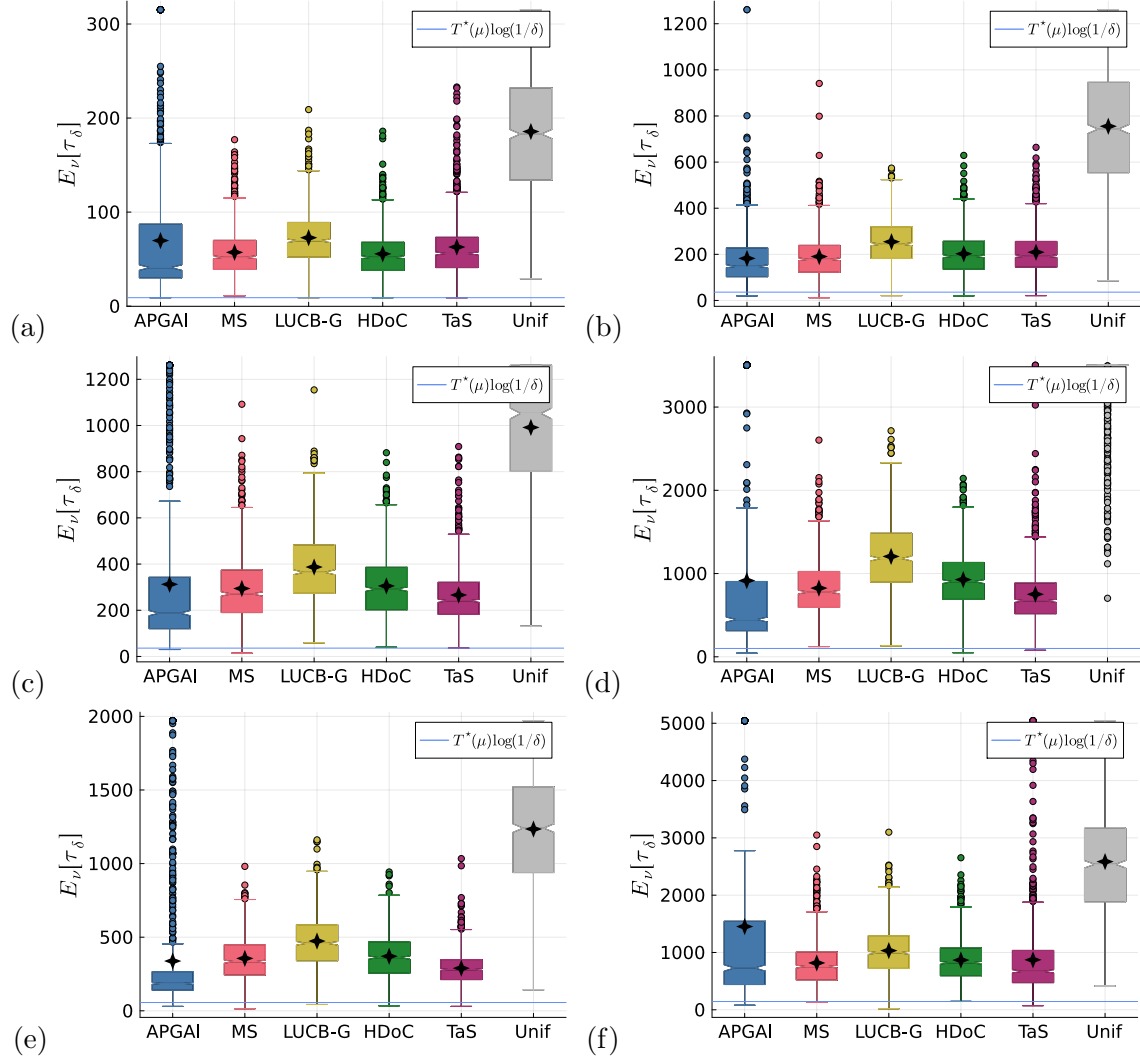


Figure 22: Empirical stopping time ($\delta = 0.01$) on instances (a) Isa2, (b) MED2, (c) Isa1, (d) REALL, (e) THR1 and (f) THR2. “MS” is Murphy Sampling, “TaS” is Track-and-Stop and “Unif” is round-robin uniform sampling.

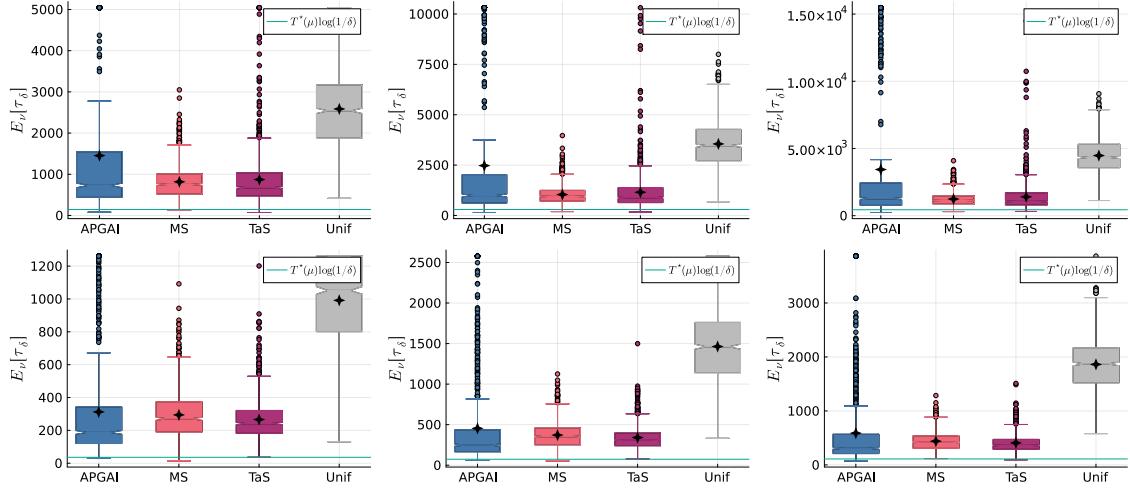


Figure 23: Empirical error for varying confidence level $\delta \in \{10^{-2}, 10^{-4}, 10^{-6}\}$ (left to right) on instances (top) THR2 and (bottom) ISA1.

FE	THR1	THR2	THR3	MED1	MED2	ISA1	ISA2	REALL	NOA1	NOA2
No	634	2448	12301	22588	184	544	159	3721	288	3014
	± 2091	± 4269	± 4755	± 9204	± 147	± 1591	± 557	± 12511	± 56	± 1031
Yes	341	1466	12584	22394	216	341	72	921	287	3022
	± 505	± 2833	± 4818	± 8942	± 106	± 444	± 49	± 1389	± 55	± 1025

Table 14: Empirical stopping time ($\pm$ standard deviation) of APGAI with or without forced exploration.

Table 14 shows that adding forced exploration significantly reduce the mean and the variance of the stopping time on instances where APGAI was prone to large outliers. For instances where APGAI had no large outliers, APGAI-FE has the same empirical performance. Therefore, adding forced exploration allows to circumvent the empirical shortcomings of APGAI in the fixed-confidence setting.

References

- A. Al Marjani, T. Kocak, and A. Garivier. On the complexity of all ε -best arms identification. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 317–332. Springer, 2022.
- J.-Y. Audibert, S. Bubeck, and R. Munos. Best arm identification in multi-armed bandits. In *Conference on Learning Theory*, 2010.
- D. A. Berry. Bayesian clinical trials. *Nature reviews Drug discovery*, 5(1):27–36, 2006.
- A. Carpentier, A. Lazaric, M. Ghavamzadeh, R. Munos, and P. Auer. Upper-confidence-bound algorithms for active learning in multi-armed bandits. In *International Conference*

- on *Algorithmic Learning Theory*, 2011.
- J. Cheshire, P. Ménard, and A. Carpentier. Problem dependent view on structured thresholding bandit problems. In *International Conference on Machine Learning*, pages 1846–1854. PMLR, 2021.
- N. R. Clark, K. S. Hu, A. S. Feldmann, Y. Kou, E. Y. Chen, Q. Duan, and A. Ma’ayan. The characteristic direction: a geometrical approach to identify differentially expressed genes. *BMC bioinformatics*, 15:1–16, 2014.
- R. Degenne. *Impact of structure on the design and analysis of bandit algorithms*. PhD thesis, Université de Paris, 2019.
- R. Degenne. On the existence of a complexity in fixed budget bandit identification. In *Conference on Learning Theory*, 2023.
- R. Degenne and W. M. Koolen. Pure exploration with multiple correct answers. *Advances in Neural Information Processing Systems*, 32, 2019.
- R. Degenne, W. M. Koolen, and P. Ménard. Non-asymptotic pure exploration by solving games. *Advances in Neural Information Processing Systems*, 32, 2019.
- E. Even-Dar, S. Mannor, Y. Mansour, and S. Mahadevan. Action elimination and stopping conditions for the multi-armed bandit and reinforcement learning problems. *Journal of machine learning research*, 7(6), 2006.
- V. Gabillon, M. Ghavamzadeh, and A. Lazaric. Best arm identification: A unified approach to fixed budget and fixed confidence. *Advances in Neural Information Processing Systems*, 25, 2012.
- A. Garivier and E. Kaufmann. Optimal best arm identification with fixed confidence. In *Conference on Learning Theory*, pages 998–1027. PMLR, 2016.
- A. Garivier and E. Kaufmann. Non-asymptotic sequential tests for overlapping hypotheses and application to near optimal arm identification in bandit models. *Sequential Analysis*, 40(1):61–96, 2021.
- T. Hayashi, N. Ito, K. Tabata, A. Nakamura, K. Fujita, Y. Harada, and T. Komatsuzaki. Gaussian process classification bandits. *Pattern Recognition*, 149:110224, 2024.
- K. Jamieson and R. Nowak. Best-arm identification algorithms for multi-armed bandits in the fixed confidence setting. In *Conference on Information Sciences and Systems (CISS)*, pages 1–6. IEEE, 2014.
- M. Jourdan and R. Degenne. Non-asymptotic analysis of a ucb-based top two algorithm. *Thirty-Seventh Conference on Neural Information Processing Systems*, 2023.
- M. Jourdan, R. Degenne, and E. Kaufmann. Dealing with unknown variances in best-arm identification. *International Conference on Algorithmic Learning Theory*, 2023.

- M. Jourdan, R. Degenne, and E. Kaufmann. An ε -best-arm identification algorithm for fixed-confidence and beyond. *Advances in Neural Information Processing Systems*, 36, 2024.
- K.-S. Jun and R. Nowak. Anytime exploration for multi-armed bandits using confidence information. In *International Conference on Machine Learning*, pages 974–982. PMLR, 2016.
- H. Kano, J. Honda, K. Sakamaki, K. Matsuura, A. Nakamura, and M. Sugiyama. Good arm identification via bandit feedback. *Machine Learning*, 108(5):721–745, 2019.
- Z. Karnin, T. Koren, and O. Somekh. Almost optimal exploration in multi-armed bandits. In *International Conference on Machine Learning*, pages 1238–1246. PMLR, 2013.
- J. Katz-Samuels and K. Jamieson. The true sample complexity of identifying good arms. In *International Conference on Artificial Intelligence and Statistics*, pages 1781–1791. PMLR, 2020.
- E. Kaufmann, O. Cappé, and A. Garivier. On the complexity of best arm identification in multi-armed bandit models. *Journal of Machine Learning Research*, 17:1–42, 2016.
- E. Kaufmann, W. M. Koolen, and A. Garivier. Sequential test for the lowest mean: From thompson to murphy sampling. *Advances in Neural Information Processing Systems*, 31, 2018.
- L. Li, K. Jamieson, G. DeSalvo, A. Rostamizadeh, and A. Talwalkar. Hyperband: A novel bandit-based approach to hyperparameter optimization. *The Journal of Machine Learning Research*, 18(1):6765–6816, 2017.
- Z. Li and W. C. Cheung. Near optimal non-asymptotic sample complexity of 1-identification. In *Forty-second International Conference on Machine Learning*, 2025.
- A. Locatelli, M. Gutzeit, and A. Carpentier. An optimal algorithm for the thresholding bandit problem. In *International Conference on Machine Learning*, pages 1690–1698. PMLR, 2016.
- S. Mannor and J. Tsitsiklis. The Sample Complexity of Exploration in the Multi-Armed Bandit Problem. *Journal of Machine Learning Research*, pages 623–648, 2004.
- B. Mason, L. Jain, A. Tripathy, and R. Nowak. Finding all ε -good arms in stochastic bandits. *Advances in Neural Information Processing Systems*, 33:20707–20718, 2020.
- B. Mason, L. Jain, S. Mukherjee, R. Camilleri, K. Jamieson, and R. Nowak. Nearly optimal algorithms for level set estimation. In *International Conference on Artificial Intelligence and Statistics*, pages 7625–7658. PMLR, 2022.
- R. Poiani, M. Jourdan, E. Kaufmann, and R. Degenne. Best-arm identification in unimodal bandits. In *The 28th International Conference on Artificial Intelligence and Statistics*, 2025.

- E. O. Rivera and A. Tewari. Optimal thresholding linear bandit. *arXiv preprint arXiv:2402.09467*, 2024.
- X. Shang, E. Kaufmann, and M. Valko. Adaptive black-box optimization got easier: Hct only needs local smoothness. *European Workshop on Reinforcement Learning*, 2018.
- M. Simchowitz, K. G. Jamieson, and B. Recht. The simulator: Understanding adaptive sampling in the moderate-confidence regime. In *International Conference On Learning Theory (COLT)*, 2017.
- K. Tabata, A. Nakamura, J. Honda, and T. Komatsuzaki. A bad arm existence checking problem: How to utilize asymmetric problem structure? *Machine learning*, 109(2):327–372, 2020.
- W. R. Thompson. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3-4):285–294, 1933.
- A. Tirinzoni and R. Degenne. On elimination strategies for bandit fixed-confidence identification. *Advances in Neural Information Processing Systems*, 2022.
- T.-H. Tsai, Y.-D. Tsai, and S.-D. Lin. lil’hdod: an algorithm for good arm identification under small threshold gap. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 78–89. Springer, 2024.
- Y.-D. Tsai, T.-H. Tsai, and S.-D. Lin. Differentiable good arm identification. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 253–264. Springer, 2025.
- P.-A. Wang, K. Ariu, and A. Proutiere. On universally optimal algorithms for a/b testing. In *Forty-first International Conference on Machine Learning*, 2024a.
- P.-A. Wang, R.-C. Tzeng, and A. Proutiere. Best arm identification with fixed budget: A large deviation perspective. *Advances in Neural Information Processing Systems*, 36, 2024b.
- E. B. Wilson. Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association*, 22(158):209–212, 1927.
- L. Xu, J. Honda, and M. Sugiyama. A fully adaptive algorithm for pure exploration in linear bandits. In *International Conference on Artificial Intelligence and Statistics*, pages 843–851. PMLR, 2018.
- Y. Zhao, C. Stephens, C. Szepesvari, and K.-S. Jun. Revisiting simple regret: Fast rates for returning a good arm. In *International Conference on Machine Learning*, 2023.