An Anytime Algorithm for Good Arm Identification

Marc Jourdan, Clémence Réda

An Anytime Algorithm for Good Arm Identification

Marc Jourdan
Univ. Lille, CNRS, Inria, Centrale Lille,
UMR 9189-CRIStAL,
F-59000 Lille, France

Clémence Réda
Department of Systems Biology and Bioinformatics,
University of Rostock,
G-18051, Rostock, Germany

Abstract

In good arm identification (GAI), the goal is to identify one arm whose average performance exceeds a given threshold, referred to as good arm, if it exists. Few works have studied GAI in the fixed-budget setting, when the sampling budget is fixed beforehand, or 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 are more efficient in detecting the absence of good arms than uniform sampling. 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 good empirical performance of APGAI on synthetic and real-world data. Our work offers an extensive overview of the GAI problem in all settings.

1 INTRODUCTION

Multi-armed bandit algorithms are a family of approaches which demonstrated versatility in solving online allocation problems, where constraints are set 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 arm a whose mean will be denoted by μ_a. Depending on their objectives, agents should have different sampling strategies.

In *pure exploration* problems, the goal is to answer a question about the set of arms. It has been 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 $1-\delta$. In the fixed-budget setting, the objective is to minimize the probability of misidentifying a correct answer with a fixed number of samples T.

While the constraint on δ or T is supposed to be given, properly choosing it 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. When the collected data shows sufficient evidence in favor of one answer, an experiment is often stopped before the initial budget is reached, referred to as *early stopping*. When additional sampling budget have been obtained due to new funding, an experiment can continue after the initial budget has been consumed, referred to as *continuation*. While early stopping and continuation are common practices, both fixed-confidence and fixed-budget settings fail to provide useful guarantees for them. Recently, the *anytime* setting has received increased scrutiny as it fills this gap between theory and practice. In the anytime setting, the agent aims at achieving a low probability of error at any deterministic time (Jun and Nowak, 2016; Zhao et al., 2023; Jourdan et al., 2023b). When the candidate answer has anytime guarantees, the practitioners can use continuation or early stopping (when combined with a stopping rule).

The most studied topic in pure exploration is the *best arm (BAI) / Top-m identification* problem, which

Preliminary work.

aims at determining a subset of m arms with largest means (Karnin et al., 2013; Xu et al., 2018; Tirinzoni and Degenne, 2022). However, in some applications such as investigating treatment protocols, BAI requires too many samples for it to be useful in practice. To avoid wasteful queries, practitioners might be interested in easier tasks that identify 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., 2023b), the agent is interested in an arm which is ε-close to the best one, *i.e.* $\mu_a \geq \max_{k\in\mathcal{A}} \mu_k - \varepsilon$. The larger ε is, the easier the task. However, choosing a meaningful value of ε can be tricky. This is why the focus of this paper is good arm identification (GAI), where the agent aims to obtain a *good arm*, which is defined as an arm whose average performance exceeds a given threshold θ, *i.e.* $\mu_a \geq \theta$. For instance, in our outcome scoring problem (see Section 5), practitioners have enough information about the distributions to define a meaningful threshold beforehand. GAI and variants have been 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. In this paper, we fill this gap by introducing APGAI, an anytime and parameter-free sampling rule for GAI which is independent of a budget T or a confidence δ and can be used in the fixed-budget and fixed-confidence settings.

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

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

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 $\mathcal{F}_t := \sigma(a_1, X_{a_1,1}, \cdots, a_t, X_{a_t,t})$ be the σ-algebra, called *history*, which encompasses all the information available to the agent after t rounds.

Identification Strategy In the anytime setting, an *identification* strategy is defined by 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 $P^{\text{err}}_{\nu,\mathfrak{A}}(t) := \mathbb{P}_\nu(\mathcal{E}^{\text{err}}_{\mathfrak{A}}(t))$ of algorithm $\mathfrak{A}$ on instance μ at time t is the probability of the error event $\mathcal{E}^{\text{err}}_{\mathfrak{A}}(t) = \{\hat{a}_t \in \{\emptyset\} \cup (\mathcal{A} \setminus \mathcal{A}_\theta)\}$ when $\mathcal{A}_\theta \neq \emptyset$, otherwise $\mathcal{E}^{\text{err}}_{\mathfrak{A}}(t) = \{\hat{a}_t \neq \emptyset\}$ when $\mathcal{A}_\theta = \emptyset$.

Those rules have a different objective depending on the considered setting. In anytime GAI, they are designed to ensure that $P^{\text{err}}_{\nu,\mathfrak{A}}(t)$ is small at any time t. In fixed-budget GAI, the goal is to have a low $P^{\text{err}}_{\nu,\mathfrak{A}}(T)$, where T is fixed beforehand. Whereas in fixed-confidence GAI, these two rules are complemented by a stopping rule using a confidence level $1-\delta$ fixed beforehand such that $\mathfrak{A}$ stops sampling after τ_δ rounds. The stopping time τ_δ is also known as the sample complexity of a fixed-confidence algorithm. At stopping time τ_δ, the algorithm should satisfy δ-correctness, which means that $\mathbb{P}_\nu(\{\tau_\delta < +\infty\} \cap \mathcal{E}^{\text{err}}_{\mathfrak{A}}(\tau_\delta)) \leq \delta$ for all instances μ. That requirement leads to a lower bound on the expected sample complexity on 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 well-known change of measure inequality (Kaufmann et al., 2016, Lemma 1).

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

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

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

1.2 Contributions

We propose APGAI, an anytime and parameter-free sampling rule for GAI in stochastic bandits, which is independent of a budget T or a confidence δ. APGAI is the first algorithm which 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 $P^{\text{err}}_{\nu,\text{APGAI}}(t)$ of the order $\exp(-\mathcal{O}(t/H_1(\mu)))$ which holds for any deterministic time t (Theorem 1). Adaptive strategies are more efficient in detecting the absence of good arms than uniform sampling (see Section 3). Second, when combined with a GLR stopping rule, we derive

an upper bound on $\mathbb{E}_\nu[\tau_\delta]$ holding at any confidence level (Theorem 2). In particular, APGAI is asymptotically optimal for GAI with Gaussian distributions when there is no good arm. Finally, APGAI is easy to implement, computationally inexpensive and achieves good empirical performance in both settings on synthetic and real-world data with an outcome scoring problem for RNA-sequencing data (see Section 5). Our work offers an overview of the GAI problem in all settings.

1.3 Related Work

GAI has never been studied in the fixed-budget or anytime setting. In the fixed-confidence setting, several questions have been studied which are closely connected to GAI. Given two thresholds $\theta_L < \theta_U$, Tabata et al. (2020) studies the Bad Existence Checking problem, in which the agent should output "negative" if $\mathcal{A}_{\theta_L}(\mu) = \emptyset$ and "positive" if $\mathcal{A}_{\theta_U}(\mu) \neq \emptyset$. They propose an elimination-based meta-algorithm called BAEC, and analyze its expected sample complexity when combined with several index-policy to define the sampling rule. Kano et al. (2019) considers identifying the whole set of good arms $\mathcal{A}_\theta(\mu)$ with high probability, and returns the good arms in a sequential way. We refer to that problem as AllGAI. In Kano et al. (2019), they introduce three index-based GAI algorithms named APT-G, HDoC and LUCB-G, and show upper bounds on their expected sample complexity. A large number of algorithms from previously mentioned works bear a passing resemblance to the APT algorithm in Locatelli et al. (2016) which tackles the thresholding bandit problem in the fixed-budget setting. The latter should classify all arms into $\mathcal{A}_\theta$ and $\mathcal{A}_\theta^\complement$ at the end of the sampling phase. This resemblance lies in that those algorithms 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.

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 performing some conditioning on the "worst event" (named Murphy Sampling). The empirical pulling proportions are shown to converge towards the allocation realizing $T^\star(\mu)$ in Lemma 1. Another related framework is the identification with high probability of k arms from $\mathcal{A}_\theta(\mu)$ (Katz-Samuels and Jamieson, 2020). They introduce the *unverifiable sample complexity*. It is the minimum number of samples after which the algorithm always outputs a correct answer with high probability. It does not require to certify that the output is correct.

2 ANYTIME PARAMETER-FREE SAMPLING RULE

We propose the APGAI (**A**nytime **P**arameter-free **GAI**) algorithm, which is independent of a budget T or a confidence δ and is 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

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

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 $n_0 \in \mathbb{N}$ times, and we use $n_0 = 1$.

Recommendation Rule At time $t + 1 > n_0 K$, 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 $t + 1 > n_0 K$, 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

Algorithm 1 APGAI

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

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. 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 which is the most likely to be a good arm.

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. BAEC is an elimination-based meta-algorithm which 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 1 shows an upper bound of the order $\exp(-\mathcal{O}(t/H_1(\mu)))$ for $P^{\text{err}}_{\nu,\mathfrak{A}}(t)$ that holds for any deterministic time t.

Theorem 1. *Let $p(x) = x - 0.5\log x$. The APGAI algorithm $\mathfrak{A}$ satisfies that, for all $\nu \in \mathcal{D}^K$ with mean μ such that $\Delta_{\min} > 0$, for all $t > n_0K + 2|\mathcal{A}_\theta|$,*

$$P^{err}_{\nu,\mathfrak{A}}(t) \leq Ke\sqrt{2}\log(e^2t)\exp\left(-p\left(\frac{t - n_0K - 2|\mathcal{A}_\theta|}{2\alpha_{i_\mu}H_1(\mu)}\right)\right)$$

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

Theorem 1 holds for any deterministic time $t > n_0K + 2|\mathcal{A}_\theta|$ and any 1-sub-Gaussian instance ν. In the asymptotic regime where $t \to +\infty$, Theorem 1 shows that $\limsup_{t\to+\infty} t\log(1/P^{\text{err}}_{\nu,\mathfrak{A}}(t))^{-1} \leq 2\alpha_{i_\mu}H_1(\mu)$ for APGAI with $(\alpha_1, \alpha_\theta) = (9, 2)$. We defer the reader to Appendix H for a 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 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 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 only differ by their sampling rule, we can measure the benefits of adaptive sampling. Theorem 4 in Appendix C gives anytime upper bounds on $P^{\text{err}}_{\nu,\text{Unif}}(t)$. In the asymptotic regime, Unif achieves 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 shows that APGAI outperforms Unif on most instances (*e.g.* Figures 1 and 2), and is on par with it otherwise.

Worst-case Lower Bound Degenne (2023) recently studied the existence of a complexity in fixed-budget pure exploration. While there is a complexity $T^\star(\mu)$ as in (2) for the fixed-confidence setting, (Degenne, 2023, Theorem 6) shows that a sequence of fixed-budget algorithms $(\mathfrak{A}_T)_T$ (where $\mathfrak{A}_T$ denotes the algorithm using fixed budget T) cannot have a better asymptotic rate than $KT^\star(\mu)$ on all Gaussian instances

$$\exists \mu \in \mathbb{R}^K,\ \limsup_{T\to+\infty} T\log(1/P^{\text{err}}_{\nu,\mathfrak{A}_T}(T))^{-1} \geq KT^\star(\mu)\,. \tag{4}$$

Unif achieves the rate $KT^\star(\mu)$ when $\mathcal{A}_\theta \neq \emptyset$, but suffers from worse guarantees otherwise. Conversely, APGAI achieves the rate in $T^\star(\mu)$ when $\mathcal{A}_\theta = \emptyset$, but has sub-optimal guarantees otherwise. It does not conflict with (4) *e.g.* considering μ with $\mathcal{A}_\theta \neq \emptyset$ and such that there exists an arm $a \in \mathcal{A}$ with $\Delta_a \leq \max_{a\in\mathcal{A}_\theta} \Delta_a/\sqrt{K/2-1}$. Experiments in Section 5 suggest that the sub-optimal dependency when $\mathcal{A}_\theta \neq \emptyset$ is not aligned with the good practical performance of APGAI. Formally proving better guarantees when $\mathcal{A}_\theta(\mu) \neq \emptyset$ is a direction for future work.

In fixed-budget GAI, a good strategy has highly different sampling modes depending on whether there is a good arm or not. Since wrongfully committing to one of those modes too early will incur higher error, it is challenging to find the perfect trade-off in an adaptive manner. Designing an algorithm whose guarantees are comparable to (4) for all instances is an open problem.

3.1 Benchmark: Other 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 1.

3.1.1 From BAI to GAI Algorithms

Since a BAI algorithm outputs the arm with highest mean, it can be adapted to GAI by comparing the 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 last arm that was not eliminated.

Theorems 5 and 6 in Appendix C give an upper bound on $P^{\mathrm{err}}_{\nu,\text{SR-G}}(T)$ and $P^{\mathrm{err}}_{\nu,\text{SH-G}}(T)$ at the fixed budget T. In the asymptotic regime, their rate is in $4\log(K)\Delta^{-2}_{\min}$ when $\mathcal{A}_\theta(\mu) = \emptyset$, otherwise

$$\mathcal{O}(\log(K)\max\{\max_{a\in\mathcal{A}_\theta}\Delta_a^{-2}, \max_{i>I^\star} i(\max_{a\in\mathcal{A}}\mu_a - \mu_{(i)})^{-2}\})$$

with $I^\star = |\arg\max_{a\in\mathcal{A}}\mu_a|$ and $\mu_{(i)}$ be the i^{th} largest mean in vector μ. Recently, Zhao et al. (2023) have provided a finer analysis of SH. Using their result yields mildly improved rates. We defer the reader to Appendix C for further details. Those rates are better than $2H_1(\mu)$ when there is one good arm with large mean and the remaining arms have means slightly smaller than θ. However, APGAI has better guarantees than SR-G and SH-G when there is one good arm with mean slightly smaller than the largest mean.

Doubling Trick The doubling trick allows the conversion of any fixed-budget algorithm into an anytime algorithm. It considers a sequences of algorithms that are run with increasing budgets $(T_k)_{k\geq 1}$, and recommends the answer outputted by the last instance. Zhao et al. (2023) shows that Doubling SH obtains the same guarantees than SH in BAI, hence Theorem 5 also holds for its GAI counterpart DSH-G (resp. Theorem 6 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 (Fig. 1 and 2).

3.1.2 Prior Knowledge-based GAI Algorithms

Several fixed-budget BAI algorithms assume that the agent has access to some prior knowledge on unknown quantities 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 GAI), described in Appendix D. As for BAEC, PKGAI($\star$) takes as input an index policy $\star$ which is used to define the sampling rule. The 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 $H_\theta(\mu)$.

Table 1: Asymptotic error rate $C(\mu)$ of algorithm $\mathfrak{A}$ on ν, *i.e.* $\limsup_t t\log(1/P^{\mathrm{err}}_{\nu,\mathfrak{A}}(t))^{-1} \leq C(\mu)$. (†) Fixed-budget algorithm $\mathfrak{A}_{T,\nu}$ with prior knowledge on ν. $H_1(\mu)$ as in (1), $\Delta_{\min} := \min_{a\in\mathcal{A}}\Delta_a$, $\overline{\Delta}_{\max} := \max_{a\in\mathcal{A}_\theta}\Delta_a$, $I^\star = |\arg\max_{a\in\mathcal{A}}\mu_a|$ and $\widetilde{\Delta}^{-2} := \max\{\max_{a\in\mathcal{A}_\theta}\Delta_a^{-2}, \max_{i>I^\star} i(\max_{a\in\mathcal{A}}\mu_a - \mu_{(i)})^{-2}\}$.

Algorithm $\mathfrak{A}$	$\mathcal{A}_\theta(\mu) = \emptyset$	$\mathcal{A}_\theta(\mu) \neq \emptyset$
APGAI [Th. 1]	$18H_1(\mu)$	$4H_1(\mu)$
Unif [Th. 4]	$2K\Delta^{-2}_{\min}$	$4K\overline{\Delta}^{-2}_{\max}$
DSR-G [Th. 5]	$16\log K\Delta^{-2}_{\min}$	$4\log K\widetilde{\Delta}^{-2}$
DSH-G [Th. 6]	$16\log K\Delta^{-2}_{\min}$	$4\log K\widetilde{\Delta}^{-2}$
PKGAI($\star$) [Th. 7]†	$2H_1(\mu)$	$2H_1(\mu)$
PKGAI(Unif) [Th. 8]†	$2H_1(\mu)$	$2K\hat{\Delta}^{-2}$

We provide upper confidence bounds on the probability of error at time T holding for any choice of indices (Theorem 7 for PKGAI($\star$)) and for uniform round-robin sampling (Theorem 8 for PKGAI(Unif)). The obtained upper bounds on $P^{\mathrm{err}}_{\nu,\text{PKGAI}}(T)$ are marginally lower than the ones obtained for APGAI, while APGAI does not require the knowledge of $H_1(\mu)$ and $H_\theta(\mu)$.

3.2 Unverifiable Sample Complexity

The *unverifiable sample complexity* was defined in Katz-Samuels and Jamieson (2020) as the smallest stopping time $\tau_{U,\delta}$ after which an algorithm always outputs a correct answer with probability at least $1-\delta$. In GAI, this means that algorithm $\mathfrak{A}$ satisfies $\mathbb{P}_\nu(\bigcup_{t\geq\tau_{U,\delta}}\mathcal{E}^{\mathrm{err}}_{\mathfrak{A}}(t)) \leq \delta$. Compared to the fixed-confidence setting, it does not require to certify that the candidate answer is correct. Authors in Zhao et al. (2023) notice that anytime bounds on the error can imply an unverifiable sample complexity bound. Theorem 3 in Appendix B.3 gives a deterministic upper bound on the unverifiable sample complexity $\tau_{U,\delta}$ of APGAI, *i.e.*

$$U_\delta(\mu) =_{\delta\to 0} 2\alpha_{i_\mu}H_1(\mu)\log(1/\delta) + \mathcal{O}(\log\log(1/\delta))$$

with $i_\mu = 1+(\theta-1)\mathbb{1}(\mathcal{A}_\theta(\mu)\neq\emptyset)$ and $(\alpha_1,\alpha_\theta) = (9,2)$. While such upper bounds are known in BAI (Katz-Samuels and Jamieson, 2020; Zhao et al., 2023; Jourdan et al., 2023b), this is the first result for GAI.

4 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 strategy should define a stopping rule in addition of 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 in Kaufmann et al. (2018). At fixed confidence δ, we stop at $\tau_\delta := \min(\tau_{>,\delta}, \tau_{<,\delta})$

$$\begin{aligned} \text{where } \tau_{>,\delta} &:= \inf\{t \mid \max_{a\in\mathcal{A}} W_a^+(t) \geq \sqrt{2c(t,\delta)}\}\,, \\ \tau_{<,\delta} &:= \inf\{t \mid \min_{a\in\mathcal{A}} W_a^-(t) \geq \sqrt{2c(t,\delta)}\}\,, \end{aligned} \tag{5}$$

and $c : \mathbb{N}\times(0,1) \to \mathbb{R}_+$ is a threshold function. Proven in Appendix G.1, Lemma 2 gives a threshold ensuring that the GLR stopping rule (5) is δ-correct for all $\delta \in (0,1)$, independently of the sampling rule.

Lemma 2. *Let $\overline{W}_{-1}(x) = -W_{-1}(-e^{-x})$ for all $x \geq 1$, where W_{-1} is the negative branch of the Lambert W function. It satisfies $\overline{W}_{-1}(x) \approx x + \log x$. Let $\delta \in (0,1)$. Given any sampling rule, using the threshold*

$$2c(t,\delta) = \overline{W}_{-1}(2\log(K/\delta) + 4\log\log(e^4 t) + 1/2) \tag{6}$$

in the GLR stopping rule (5) yields a δ-correct algorithm for 1-sub-Gaussian distributions.

Non-asymptotic Upper Bound Theorem 2 gives an upper bound on the expected sample complexity of the resulting algorithm holding for any confidence δ.

Theorem 2. *Let $\delta \in (0,1)$. Combined with GLR stopping (5) using threshold (6), the APGAI algorithm is δ-correct and it satisfies that, for all $\nu \in \mathcal{D}^K$ with mean μ such that $\Delta_{\min} > 0$,*

$$\mathbb{E}_\nu[\tau_\delta] \leq C_\mu(\delta) + K\pi^2/6 + 1\,,$$

where $i_\mu := 1 + (\theta - 1)\mathbb{1}\left(\mathcal{A}_\theta(\mu) \neq \emptyset\right)$ and $C_\mu(\delta) :=$

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

with $H_1(\mu)$ and $H_\theta(\mu)$ as in (1). $D_1(\mu)$ and $D_\theta(\mu)$ are defined in Lemmas 16 and 18 in Appendix F, satisfying

$$D_1(\mu) \approx_{\Delta_{\min}\to+\infty} D_\theta(\mu) = \mathcal{O}(H_1(\mu)\log H_1(\mu))\,.$$

In the asymptotic regime, we obtain

$$\limsup_{\delta\to 0} \mathbb{E}_\nu[\tau_\delta]/\log(1/\delta) \leq 2H_{i_\mu}(\mu)\,.$$

since $C_\mu(\delta) =_{\delta\to 0} 2H_{i_\mu}(\mu)\log(1/\delta) + \mathcal{O}(\log\log(1/\delta))$.

Table 2: Asymptotic upper bound $2C(\mu)$ on the expected sample complexity of algorithm $\mathfrak{A}$ on ν, *i.e.* $\limsup_{\delta\to 0} \mathbb{E}_\nu[\tau_\delta]/\log(1/\delta) \leq 2C(\mu)$. (§) Requires an ordering on the possible answers $\mathcal{A}\cup\{\emptyset\}$. $H_1(\mu)$ and $H_\theta(\mu)$ as in (1), $\overline{\Delta}_{\max} := \max_{a\mathcal{A}_\theta} \Delta_a$.

Algorithm $\mathfrak{A}$	$\mathcal{A}_\theta(\mu) = \emptyset$	$\mathcal{A}_\theta(\mu) \neq \emptyset$
APGAI[Th. 2]	$H_1(\mu)$	$H_\theta(\mu)$
S-TaS § (Degenne and Koolen, 2019)	$H_1(\mu)$	$\overline{\Delta}_{\max}^{-2}$
HDoC (Kano et al., 2019)	$H_1(\mu)$	$\overline{\Delta}_{\max}^{-2}$
APT-G, LUCB-G (Kano et al., 2019)	$H_1(\mu)$	–

Most importantly, Theorem 2 holds for any confidence $\delta \in (0,1)$ and any 1-sub-Gaussian instance ν. In the asymptotic regime where $\delta \to 0$, Theorem 2 shows that $\limsup_{\delta\to 0} \mathbb{E}_\nu[\tau_\delta]/\log(1/\delta) \leq 2H_{i_\mu}(\mu)$. This implies that APGAI is asymptotically optimal for Gaussian distributions when $\mathcal{A}_\theta = \emptyset$. When there are good arms, our upper bound scales as $H_\theta(\mu)\log(1/\delta)$, 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 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 large outliers when there are good arms with dissimilar gaps, and that adding forced exploration circumvents this issue (Figure 22 and Table 11 in Appendix I.5). Intuitively, a purely asymptotic analysis of APGAI would 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.2.1 for more details. Compared to asymptotic results, our non-asymptotic guarantees hold for reasonable values of δ, with a δ-independent scaling of the order $\mathcal{O}(H_1(\mu)\log H_1(\mu))$.

Comparison With Existing Upper Bounds Table 2 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 1). However, we emphasize that APGAI is not the best algorithm to tackle fixed-confidence GAI since it is designed for anytime 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 setting 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 5 in Appendix I.2). This partly explains the lack of non-asymptotic guarantees for S-TaS which is asymptotic by nature, while APGAI has non-asymptotic guarantees. For the "bad arm existence" problem, Kaufmann et al. (2018) proves 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 $\lim_{\delta\to 0}\tau_\delta/\log(1/\delta) = T^\star(\mu)$ almost surely, the authors provide no upper bound on the expected sample complexity of Murphy Sampling. Finally, we consider the AllGAI algorithms introduced in Kano et al. (2019) (HDoC, LUCB-G and APT-G) which enjoy theoretical guarantees for some GAI instances as well. 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 in Tabata et al. (2020). Extrapolating the theoretical guarantees of BAEC when $\theta_L \to \theta_U$, one would expect an upper bound on its expected sample complexity of the form $2H_1(\mu)\log(1/\delta) + \mathcal{O}((\log(1/\delta))^{2/3})$.

5 EXPERIMENTS

We assess the empirical performance of the APGAI in terms of empirical error, as well as empirical stopping time. Overall, APGAI perform favorably compared to other algorithms in both settings. Moreover, its empirical performance exceeds what its theoretical guarantees would suggest. This 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. This application is focused 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. Our collaborators tested all treatments, and made RNA-related measurements on treated samples.

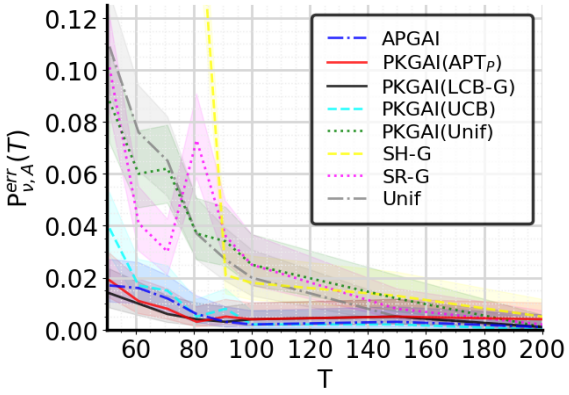

Figure 1: Fixed-budget empirical error on our outcome scoring application (see REALL in Table 3).

Computed on 3 technical replicates, the mean value in $[-1, 1]$ (see Table 3 in Appendix I.1) corresponds to a cosine score computed between gene activity changes in treated and healthy samples. When the mean is higher than $\theta = 0.5$, the treatment is considered significantly positive. Traditional approaches use grid-search with a uniform allocation. We model this application as a Bernoulli instance, *i.e.* observations from arm a are drawn from a Bernoulli distribution with mean $\max(\mu_a, 0)$ (which is 1/2-sub-Gaussian).

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 also considered). Several index policies are considered for PKGAI: Unif, APT$_P$, UCB and LCB-G. At time t, the latter selects among the set $\mathcal{S}_t$ of active candidates $a_t \leftarrow \arg\max_{a\in\mathcal{S}_t}\sqrt{N_a(t)}\mathrm{LCB}(a,t)$, where $\mathrm{LCB}(a,t)$ is the lower confidence bound on $\mu_a - \theta$ at time t. For a budget T up to 200, our results are averaged over $1,000$ runs, and confidence intervals are displayed. 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 actually performs as well as the elimination-based algorithms PKGAI($\star$), while allowing early stopping as well. 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 3.1.1) and Unif. Since DSH-G has poor empirical performance (see Figure 4), we consider the heuristic DSH-G-WR where each SH instance keeps

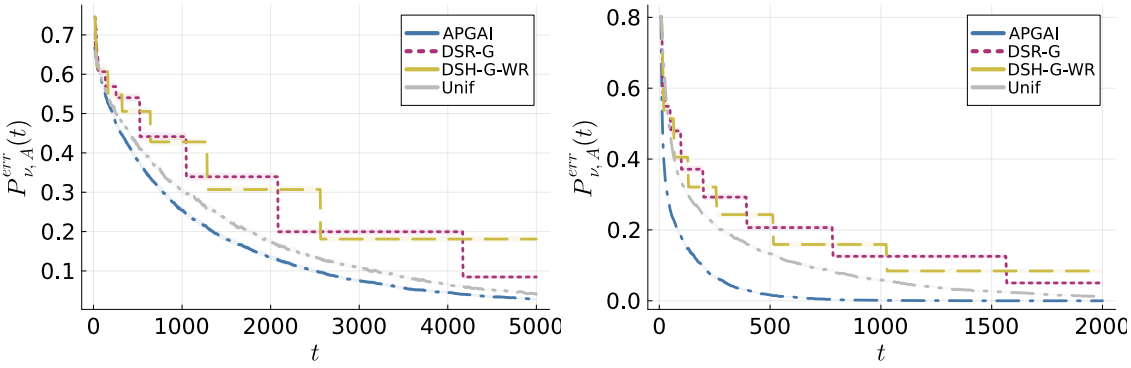

Figure 2: Anytime empirical error on Gaussian instances (a) $\mu \in \{0.55, 0.45\}^{10}$ where $|\mathcal{A}_\theta| = 3$ for $\theta = 0.5$ and (b) $\mu = -(0.1, 0.4, 0.5, 0.6)$ for $\theta = 0$.

its 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 are averaged over $10,000$ runs, and confidence intervals are displayed. 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 1 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 (5) with threshold (6) 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). In Figure 3, we study the impact of the number of good arms by considering Gaussian instances with two groups of arms. Our results are averaged over $1,000$ runs, and the standard deviations are displayed. 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 they suffer from increased $|\mathcal{A}_\theta|$ since they have an exploration bonus for each good arm. In contrast, APGAI is greedy enough to only focus its allocation to one of the good arms. While APGAI achieves the best performance when there is no good arm, it can suffer from large outliers when good arms have dissimilar means (Figure 22 in Appendix I.5). To circumvent this problem, it is enough to add forced exploration to APGAI (Table 11). While APGAI was designed for anytime GAI, it is remarkable that it also has theoretical guarantees in fixed-confidence GAI, and relatively small empirical stopping time.

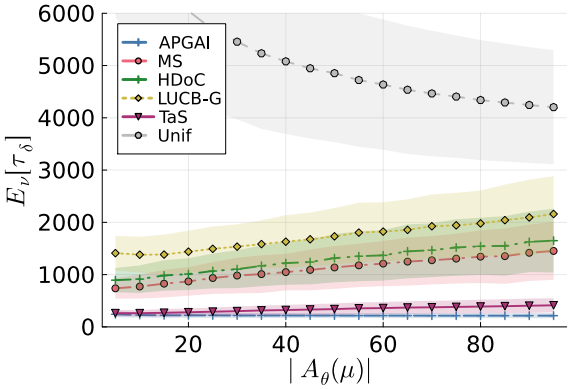

Figure 3: Empirical stopping time ($\delta = 0.01$) on Gaussian instances $\mu \in \{0.5, -0.5\}^{100}$ where $|\mathcal{A}_\theta| \in \{5k\}_{k\in[19]}$ for $\theta = 0$.

6 PERSPECTIVES

We propose APGAI, the first anytime and parameter-free sampling strategy 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 1) and on its expected sample complexity at any confidence δ when combined with the GLR stopping time (5) (Theorem 2). 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 an 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 in a both tractable and meaningful way.

Acknowledgements

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 THIA ANR program "AI_PhD@Lille".

References

Audibert, J.-Y., Bubeck, S., and Munos, R. (2010). Best arm identification in multi-armed bandits. In *COLT*, pages 41–53. Citeseer.

Berry, D. A. (2006). Bayesian clinical trials. *Nature reviews Drug discovery*, 5(1):27–36.

Carpentier, A., Lazaric, A., Ghavamzadeh, M., Munos, R., and Auer, P. (2011). Upper-confidence-bound algorithms for active learning in multi-armed bandits. In *International Conference on Algorithmic Learning Theory*.

Degenne, R. (2019). *Impact of structure on the design and analysis of bandit algorithms*. PhD thesis, Université de Paris.

Degenne, R. (2023). On the existence of a complexity in fixed budget bandit identification. In *Proceedings of Thirty Sixth Conference on Learning Theory*.

Degenne, R. and Koolen, W. M. (2019). Pure exploration with multiple correct answers. *Advances in Neural Information Processing Systems*, 32.

Degenne, R., Koolen, W. M., and Ménard, P. (2019). Non-asymptotic pure exploration by solving games. *Advances in Neural Information Processing Systems*, 32.

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

Gabillon, V., Ghavamzadeh, M., and Lazaric, A. (2012). Best arm identification: A unified approach to fixed budget and fixed confidence. *Advances in Neural Information Processing Systems*, 25.

Garivier, A. and Kaufmann, E. (2016). Optimal best arm identification with fixed confidence. In *Conference on Learning Theory*, pages 998–1027. PMLR.

Garivier, A. and Kaufmann, E. (2021). Non-asymptotic sequential tests for overlapping hypotheses and application to near optimal arm identification in bandit models. *Sequential Analysis*, 40(1):61–96.

Jamieson, K. and Nowak, R. (2014). Best-arm identification algorithms for multi-armed bandits in the fixed confidence setting. In *2014 48th Annual Conference on Information Sciences and Systems (CISS)*, pages 1–6. IEEE.

Jourdan, M., Degenne, R., and Kaufmann, E. (2023a). Dealing with unknown variances in best-arm identification. *International Conference on Algorithmic Learning Theory*.

Jourdan, M., Degenne, R., and Kaufmann, E. (2023b). An ε-best-arm identification algorithm for fixed-confidence and beyond. *arXiv preprint arXiv:2305.16041*.

Jun, K.-S. and Nowak, R. (2016). Anytime exploration for multi-armed bandits using confidence information. In *International Conference on Machine Learning*, pages 974–982. PMLR.

Kano, H., Honda, J., Sakamaki, K., Matsuura, K., Nakamura, A., and Sugiyama, M. (2019). Good arm identification via bandit feedback. *Machine Learning*, 108(5):721–745.

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

Katz-Samuels, J. and Jamieson, K. (2020). The true sample complexity of identifying good arms. In *International Conference on Artificial Intelligence and Statistics*, pages 1781–1791. PMLR.

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

Kaufmann, E., Koolen, W. M., and Garivier, A. (2018). Sequential test for the lowest mean: From thompson to murphy sampling. *Advances in Neural Information Processing Systems*, 31.

Li, L., Jamieson, K., DeSalvo, G., Rostamizadeh, A., and Talwalkar, A. (2017). Hyperband: A novel bandit-based approach to hyperparameter optimization. *The Journal of Machine Learning Research*, 18(1):6765–6816.

Locatelli, A., Gutzeit, M., and Carpentier, A. (2016). An optimal algorithm for the thresholding bandit problem. In *International Conference on Machine Learning*, pages 1690–1698. PMLR.

Mannor, S. and Tsitsiklis, J. (2004). The Sample Complexity of Exploration in the Multi-Armed Bandit Problem. *Journal of Machine Learning Research*, pages 623–648.

Shang, X., Kaufmann, E., and Valko, M. (2018). Adaptive black-box optimization got easier: Hct only needs local smoothness. *European Workshop on Reinforcement Learning*.

Tabata, K., Nakamura, A., Honda, J., and Komatsuzaki, T. (2020). A bad arm existence checking problem: How to utilize asymmetric problem structure? *Machine learning*, 109(2):327–372.

Thompson, W. R. (1933). On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3-4):285–294.

Tirinzoni, A. and Degenne, R. (2022). On elimination strategies for bandit fixed-confidence identification. *Advances in Neural Information Processing Systems*.

Wilson, E. B. (1927). Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association*, 22(158):209–212.

Xu, L., Honda, J., and Sugiyama, M. (2018). A fully adaptive algorithm for pure exploration in linear bandits. In *International Conference on Artificial Intelligence and Statistics*, pages 843–851. PMLR.

Zhao, Y., Stephens, C., Szepesvari, C., and Jun, K.-S. (2023). Revisiting simple regret: Fast rates for returning a good arm. In *Proceedings of the 40th International Conference on Machine Learning*.

Checklist

1. For all models and algorithms presented, check if you include:
   (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. **Yes, in Sections 1 and 2.**
   (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. **Yes, in Sections 2 and 3.**
   (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. **Yes, in the supplementary materials as a zip folder.**

2. For any theoretical claim, check if you include:
   (a) Statements of the full set of assumptions of all theoretical results. **Yes, in Sections 3 and 4.**
   (b) Complete proofs of all theoretical results. **Yes, in Appendices B, C, D and F.**
   (c) Clear explanations of any assumptions. **Yes, in Section 1.**

3. For all figures and tables that present empirical results, check if you include:
   (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). **Yes, in supplementary and Appendix I.2.**
   (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). **Yes, in Appendices I.1 and I.2.**
   (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). **Yes, in Section 5.**
   (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). **Yes, in Appendix I.2.**

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:
   (a) Citations of the creator If your work uses existing assets. **Not Applicable.**
   (b) The license information of the assets, if applicable. **Yes, in supplementary.**
   (c) New assets either in the supplemental material or as a URL, if applicable. **Yes, in supplementary.**
   (d) Information about consent from data providers/curators. **Yes, in supplementary.**
   (e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. **Not Applicable.**

5. If you used crowdsourcing or conducted research with human subjects, check if you include:
   (a) The full text of instructions given to participants and screenshots. **Not Applicable.**
   (b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. **Not Applicable.**
   (c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. **Not Applicable.**

A OUTLINE

The appendices are organized as follows:

- The anytime guarantees of proof APGAI on the probability of error (Theorem 1) are proven in Appendix B.
- Appendix C gathers error guarantees on other algorithms that are used as comparison with the anytime error guarantees of APGAI: UnifB (Theorem 4), SH-G (Theorem 5) and SR-G (Theorem 6).
- We propose the meta-algorithm PKGAI in Appendix D, and analyze its error guarantees for several choices of index policy (Theorems 7 and 8).
- Appendix E gives the proof of Lemma 1. Then, we link the ATP_P index and the GLR stopping rule (5) with the generalized likelihood ratio for GAI.
- The proof of Theorem 2 for APGAI when combined with the GLR stopping (5) using threshold (6) is detailed in Appendix F.
- Appendix G contains the proof of Lemma 2, 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.

B ANALYSIS OF APGAI: THEOREM 1

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 (5). We recall that each arm is pulled n_0 times as initialization.

Proof Strategy Let $\mu \in \mathbb{R}^K$ such that $\mu_a \neq \theta$ for all $a \in \mathcal{A}$. For all $T > n_0 K$ and $\delta \in (0,1)$, let $\tilde{\mathcal{E}}_{T,\delta}$ as in (22) for $s = 0$, *i.e.*

$$\tilde{\mathcal{E}}_{T,\delta} = \left\{ \forall a \in \mathcal{A}, \forall t \leq T,\ |\hat{\mu}_a(t) - \mu_a| < \sqrt{\frac{2\tilde{f}_1(T,\delta)}{N_a(t)}} \right\} , \tag{7}$$
$$\text{with} \quad \tilde{f}_1(T,\delta) = \frac{1}{2}\overline{W}_{-1}(2\log(1/\delta) + 2\log(2+\log T) + 2) ,$$

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

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

Using Lemma 22, we have $\mathbb{P}_\nu(\tilde{\mathcal{E}}_{T,\delta}^{\complement}) \leq K\delta$. Suppose that we have constructed a time $T_\mu(\delta) \geq n_0 K$ such that $\tilde{\mathcal{E}}_{T,\delta} \subseteq \mathcal{E}_\mu^{\mathrm{err}}(T)^{\complement}$ for $T > T_\mu(\delta)$. Then, we obtain

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

where the last inequality is obtained by taking the infimum. To prove Theorem 1, 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 3 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 3. *Let $\delta \in (0,1]$ and $T > n_0 K$. Let $(A_t(T,\delta))_{T \geq t \geq n_0 K}$ be a sequence of events and $(D_a(T,\delta))_{a \in \mathcal{A}}$ be positive thresholds satisfying that, for all $t \in (n_0 K, T] \cap \mathbb{N}$, under the event $A_t(T,\delta)$,*

$$N_{a_{t+1}}(t) \leq D_{a_{t+1}}(T,\delta) \quad \textit{and} \quad N_{a_{t+1}}(t+1) = N_{a_{t+1}}(t) + 1 .$$

Then, we have $\sum_{t=n_0K+1}^{T} \mathbb{1}\left(A_t(T,\delta)\right) \leq \sum_{a \in \mathcal{A}} D_a(T,\delta)$.

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

$$\begin{aligned}\sum_{t=n_0K+1}^{T} \mathbb{1}\left(A_t(T,\delta)\right) &\leq \sum_{t=n_0K+1}^{T} \mathbb{1}\left(N_{a_{t+1}}(t) \leq D_{a_{t+1}}(T,\delta),\ N_{a_{t+1}}(t+1) = N_{a_{t+1}}(t) + 1\right) \\ &\leq \sum_{a \in \mathcal{A}} \sum_{t=n_0K+1}^{T} \mathbb{1}\left(N_a(t) \leq D_a(T,\delta),\ N_a(t+1) = N_a(t) + 1\right) \leq \sum_{a \in \mathcal{A}} D_a(T,\delta) .\end{aligned}$$

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 $D_a(T,\delta)$. □

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 3. Theorem 1 is obtained by combining Lemmas 4 and 8.

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

When $\mathcal{A}_\theta = \emptyset$, we have $\mathcal{E}_\mu^{\mathrm{err}}(T) = \{\hat{a}_T \neq \emptyset\}$. Lemma 4 gives an upper bound on the probability of error based on the recommendation of the APGAI algorithm holding for all time T.

Lemma 4. *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 $T > n_0 K$ such that it has not stopped sampling at time T,*

$$\mathbb{P}_\nu(\hat{a}_T \neq \emptyset) \leq Ke\sqrt{2}(2+\log T)p\left(\frac{T-n_0K}{18H_1(\mu)}\right) .$$

Proof. In order to prove Lemma 4, we show key intermediate properties of the APGAI algorithm when $\mathcal{A}_\theta = \emptyset$.

Error Due to Undersampled Arms At a fixed (T,δ), we define the set of undersampled arms as

$$\forall t \in (n_0K, T] \cap \mathbb{N}, \quad U_t(T,\delta) = \left\{ a \in \mathcal{A} \mid N_a(t) \leq \frac{2\tilde{f}_1(T,\delta)}{\Delta_a^2} \right\} .$$

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 5).

Lemma 5. *For all $T \in \mathbb{N}$, under the event $\tilde{\mathcal{E}}_{T,\delta}$ as in (7), for all $t \in (n_0K, T] \cap \mathbb{N}$, we have*

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

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 (7), we have

$$\theta < \hat{\mu}_{\hat{a}_t}(t) \leq \mu_{\hat{a}_t} + \sqrt{\frac{2\tilde{f}_1(T,\delta)}{N_{\hat{a}_t}(t)}} \quad \text{hence} \quad \hat{a}_t \in U_t(T,\delta) .$$

□

No Remaining Undersampled Arms We show that the events $\{U_t(T,\delta) \neq \emptyset\}$ satisfy the conditions of Lemma 3 (Lemma 6). In other words, if there are still undersampled arms at time t, then a_{t+1} has not been sampled too many times.

Lemma 6. *Let $\delta \in (0,1)$ and $T > n_0K$. Under event $\tilde{\mathcal{E}}_{T,\delta}$, for all $t \in (n_0K, T] \cap \mathbb{N}$ such that $U_t(T,\delta) \neq \emptyset$, we have*

$$N_{a_{t+1}}(t) \leq \frac{18\tilde{f}_1(T,\delta)}{\Delta_{a_{t+1}}^2} \quad \textit{and} \quad N_{a_{t+1}}(t+1) = N_{a_{t+1}}(t) + 1 .$$

Proof. We will be interested in three distinct cases since

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

Case 1. Let $t \in (n_0K, T] \cap \mathbb{N}$ 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 (7), 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

$$N_{a_{t+1}}(t) \leq \frac{2\tilde{f}_1(T,\delta)}{\Delta^2_{a_{t+1}}} \quad \text{and} \quad N_{a_{t+1}}(t+1) = N_{a_{t+1}}(t) + 1 \,.$$

Case 2. Let $t \in (n_0K, T] \cap \mathbb{N}$ 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 (7), we have

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

Using that $W^-_{a_{t+1}}(t) \leq W^-_a(t)$, we have proven that

$$N_{a_{t+1}}(t) \leq \frac{18\tilde{f}_1(T,\delta)}{\Delta^2_{a_{t+1}}} \quad \text{and} \quad N_{a_{t+1}}(t+1) = N_{a_{t+1}}(t) + 1 \,.$$

Case 3. Let $t \in (n_0K, T] \cap \mathbb{N}$ 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 have proven that

$$N_{a_{t+1}}(t) \leq \frac{2\tilde{f}_1(T,\delta)}{\Delta^2_{a_{t+1}}} \quad \text{and} \quad N_{a_{t+1}}(t+1) = N_{a_{t+1}}(t) + 1 \,.$$

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

Lemma 7 provides a time after which all arms are sampled enough, hence no error will be made.

Lemma 7. *Let us define*

$$T_\mu(\delta) = \sup\left\{T \mid T \leq 18H_1(\mu)\tilde{f}_1(T,\delta) + n_0K\right\} \,.$$

For all $T > T_\mu(\delta)$, under the event $\tilde{\mathcal{E}}_{T,\delta}$ as in (7), we have $U_T(T,\delta) = \emptyset$.

Proof. Combining Lemmas 6 and 3, we obtain

$$\sum_{t=n_0K+1}^{T} \mathbb{1}\left(U_t(T,\delta) \neq \emptyset\right) \leq 18H_1(\mu)\tilde{f}_1(T,\delta) \,.$$

For all $a \in \mathcal{A}$, let us define $t_a(T,\delta) = \max\{t \in (n_0K, T] \cap \mathbb{N} \mid a \in U_t(T,\delta)\}$. By definition, we have $a \in U_t(T,\delta)$ for all $t \in (n_0K, t_a(T,\delta)]$ and $a \notin U_t(T,\delta)$ for all $t \in (t_a(T,\delta), T]$. Therefore, for all $t \in (n_0K, \max_{a\in\mathcal{A}} t_a(T,\delta)]$, 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

$$\max_{a\in\mathcal{A}}(t_a(T,\delta) - n_0K) = \sum_{t=n_0K+1}^{T} \mathbb{1}\left(U_t(T,\delta) \neq \emptyset\right) \leq 18H_1(\mu)\tilde{f}_1(T,\delta) \,.$$

Let $T_\mu(\delta)$ defined as in the statement of Lemma 7 and $T > T_\mu(\delta)$. Then, we have

$$T - n_0K > 18H_1(\mu)\tilde{f}_1(T,\delta) \geq \max_{a\in\mathcal{A}}(t_a(T,\delta) - n_0K) \,,$$

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 7. Combining Lemmas 7, 5 and 22, we obtain

$$\mathbb{P}_\nu(\hat{a}_T \neq \emptyset) \leq K \inf\{\delta \mid T > T_\mu(\delta)\} \leq Ke\sqrt{2}(2+\log T)\sqrt{\frac{T-n_0K}{18H_1(\mu)}}\exp\left(-\frac{T-n_0K}{18H_1(\mu)}\right),$$

where the last inequality uses Lemma 26. This concludes the proof of Lemma 4. □

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

When $\mathcal{A}_\theta = \emptyset$, we have $\mathcal{E}_\mu^{\mathrm{err}}(T) = \{\hat{a}_T = \emptyset\} \cup \{\hat{a}_T \in \mathcal{A}_\theta^{\complement}\}$. Lemma 8 gives an upper bound on the probability of error based on the recommendation of APGAI holding for all time T.

Lemma 8. *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 $T > n_0K$ such that it has not stopped sampling at time T,*

$$\mathbb{P}\left(\{\hat{a}_T = \emptyset\} \cup \{\hat{a}_T \in \mathcal{A}_\theta^{\complement}\}\right) \leq Ke\sqrt{2}(2+\log T)p\left(\frac{T-n_0K-2|\mathcal{A}_\theta|}{4H_1(\mu)}\right).$$

Proof. In order to prove Lemma 8, 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), we define the set of under-sampled arms as

$$\forall t \in (n_0K, T] \cap \mathbb{N}, \quad U_t(T,\delta) = \left\{ a \in \mathcal{A} \mid N_a(t) \leq \left(\sqrt{\frac{2\tilde{f}_1(T,\delta)}{\Delta_a^2}} + 1\right)^2 \right\}.$$

Lemma 9 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 $\hat{a}_t \in \mathcal{A}_\theta^{\complement}$ at time t is that this arm is undersampled and will be sampled next, *i.e.* $\hat{a}_t = a_{t+1} \in \mathcal{A}_\theta^{\complement} \cap U_t(T,\delta)$.

Lemma 9. *For all $T \in \mathbb{N}$, under the event $\tilde{\mathcal{E}}_{T,\delta}$ as in (7), for all $t \in (n_0K, T] \cap \mathbb{N}$, we have*

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

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

$$\theta \geq \hat{\mu}_a(t) \geq \mu_a - \sqrt{\frac{2f_1(T,\delta)}{N_a(t)}},$$

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 $\hat{a}_t = a_{t+1} \in \arg\max_{a\in\mathcal{A}} W_a^+(t)$, we have $\hat{\mu}_{\hat{a}_t}(t) > \theta$. Then, we have

$$\theta < \hat{\mu}_{a_{t+1}}(t) \leq \mu_{a_{t+1}} + \sqrt{\frac{2f_1(T,\delta)}{N_{a_{t+1}}(t)}},$$

hence $a_{t+1} \in \mathcal{A}_\theta^{\complement} \cap U_t(T,\delta)$. □

One Good Arm Not Undersampled Lemma 10 shows that the events $\{\mathcal{A}_\theta \subseteq U_t(T,\delta)\}$ are satisfying the conditions of Lemma 3. In other words, having all the good arms undersampled implies that the next arm we will pull was not sampled a lot.

Lemma 10. *Let $\delta \in (0,1)$ and $T > n_0K$. Under event $\tilde{\mathcal{E}}_{T,\delta}$, for all $t \in (n_0K, T] \cap \mathbb{N}$ such that $\mathcal{A}_\theta \subseteq U_t(T,\delta)$, we have*

$$N_{a_{t+1}}(t) \leq D_{a_{t+1}}(T,\delta) \quad \text{and} \quad N_{a_{t+1}}(t+1) = N_{a_{t+1}}(t) + 1,$$

where $D_a(T,\delta) = \left(\Delta_a^{-1}\sqrt{2\tilde{f}_1(T,\delta)} + 1\right)^2$ 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 $t \in (n_0K, T] \cap \mathbb{N}$ such that $\mathcal{A}_\theta \subseteq U_t(T,\delta)$. When $a_{t+1} \in \mathcal{A}_\theta$, we have directly that

$$N_{a_{t+1}}(t) \leq \left(\sqrt{\frac{2\tilde{f}_1(T,\delta)}{\Delta^2_{a_{t+1}}}} + 1\right)^2 \quad \text{and} \quad N_{a_{t+1}}(t+1) = N_{a_{t+1}}(t) + 1\,.$$

In the following, we consider $a_{t+1} \notin \mathcal{A}_\theta$.

We will be interested in three cases since

$$\{\mathcal{A}_\theta \subseteq U_t(T,\delta),\, a_{t+1} \notin \mathcal{A}_\theta\} = \underbrace{\{\mathcal{A}_\theta \subseteq U_t(T,\delta),\, a_{t+1} \notin \mathcal{A}_\theta,\, \max_{a\in\mathcal{A}} \hat{\mu}_a(t) > \theta\}}_{\text{Case 1}}$$

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

Case 1. Let $t \in (n_0K, T] \cap \mathbb{N}$ 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 (7), 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

$$N_{a_{t+1}}(t) \leq \frac{2\tilde{f}_1(T,\delta)}{\Delta^2_{a_{t+1}}} \quad \text{and} \quad N_{a_{t+1}}(t+1) = N_{a_{t+1}}(t) + 1\,.$$

Case 2. Let $t \in (n_0K, T] \cap \mathbb{N}$ 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 (7), for all $a \in \mathcal{A}_\theta$, we have

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

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

$$N_{a_{t+1}}(t) \leq \frac{8\tilde{f}_1(T,\delta)}{\Delta^2_{a_{t+1}}} \quad \text{and} \quad N_{a_{t+1}}(t+1) = N_{a_{t+1}}(t) + 1\,.$$

Case 3. Let $t \in (n_0K, T] \cap \mathbb{N}$ 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

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

Since $a_{t+1} \notin \mathcal{A}_\theta$, we obtain

$$N_{a_{t+1}}(t) \leq \frac{2\tilde{f}_1(T,\delta)}{\Delta^2_{a_{t+1}}} \quad \text{and} \quad N_{a_{t+1}}(t+1) = N_{a_{t+1}}(t) + 1\,.$$

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

Lemma 11 shows that having a good arm that is sampled enough, *i.e.* $\mathcal{A}_\theta \cap U_t(T,\delta)^\complement \neq \emptyset$, is a sufficient condition to recommend a good arm, *i.e.* $\hat{a}_t \in \mathcal{A}_\theta$.

Lemma 11. *Let $\delta \in (0,1)$ and $T > n_0 K$. Under event $\tilde{\mathcal{E}}_{T,\delta}$, for all $t \in (n_0 K, T] \cap \mathbb{N}$ such that $\mathcal{A}_\theta \cap U_t(T,\delta)^\complement \neq \emptyset$, we have $\hat{a}_t \in \mathcal{A}_\theta$.*

Proof. Let $t \in (n_0 K, T] \cap \mathbb{N}$ 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 we have

$$N_a(t) > \left(\sqrt{\frac{2\tilde{f}_1(T,\delta)}{(\mu_a - \theta)^2}} + 1 \right)^2 > \frac{2\tilde{f}_1(T,\delta)}{(\mu_a - \theta)^2} \,. \tag{8}$$

Therefore, under $\tilde{\mathcal{E}}_{T,\delta}$ as in (7), we have

$$\max_{b \in \mathcal{A}} \hat{\mu}_b(t) \geq \hat{\mu}_a(t) \geq \mu_a - \sqrt{\frac{2\tilde{f}_1(T,\delta)}{N_a(t)}} > \theta \,,$$

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 $b \in \mathcal{A}_\theta \cap U_t(T,\delta)^\complement$

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

where the two last inequalities are obtained by using (8) 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

$$\begin{aligned}
\mathcal{A}_\theta \cap U_t(T,\delta)^\complement \neq \emptyset \quad &\Longrightarrow \quad \hat{a}_t \in \arg\max_{a \in \mathcal{A}} W_a^+(t) \ \wedge \ \mathcal{A}_\theta^\complement \cap \arg\max_{a \in \mathcal{A}} W_a^+(t) = \emptyset \\
&\Longrightarrow \quad \hat{a}_t \in \mathcal{A}_\theta \,.
\end{aligned}$$

□

Lemma 12 provides a time after which there exists a good arms which is sampled enough, hence no error will be made.

Lemma 12. *Let us define*

$$S_\mu(\delta) = \sup\left\{ T \mid T \leq 4H_1(\mu)\tilde{f}_1(T,\delta) + n_0 K + 2|\mathcal{A}_\theta| \right\} \,.$$

For all $T > S_\mu(\delta)$, under the event $\tilde{\mathcal{E}}_{T,\delta}$ as in (7), 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 10. Combining Lemmas 10 and 3, we obtain

$$\sum_{t=n_0 K+1}^{T} \mathbb{1}\left(\mathcal{A}_\theta \subseteq U_t(T,\delta)\right) \leq \sum_{a \in \mathcal{A}} D_a(T,\delta) \,.$$

For all $a \in \mathcal{A}_\theta$, let us define $t_a(T,\delta) = \max\{t \in (n_0 K, T] \cap \mathbb{N} \mid a \in U_t(T,\delta)\}$. By definition, we have $a \in U_t(T,\delta)$ for all $t \in (n_0 K, t_a(T,\delta)]$ and $a \notin U_t(T,\delta)$ for all $t \in (t_a(T,\delta), T]$. Therefore, for all $t \in (n_0 K, \min_{a \in \mathcal{A}_\theta} t_a(T,\delta)]$, we have $\mathcal{A}_\theta \subseteq U_t(T,\delta)$ and $\mathcal{A}_\theta \cap U_t(T,\delta)^\complement \neq \emptyset$ for all $t > \max_{a \in \mathcal{A}} t_a(T,\delta)$, hence

$$\min_{a \in \mathcal{A}_\theta}\left(t_a(T,\delta) - K\right) = \sum_{t=n_0 K+1}^{T} \mathbb{1}\left(\mathcal{A}_\theta \subseteq U_t(T,\delta)\right) \leq \sum_{a \in \mathcal{A}} D_a(T,\delta) \,.$$

Let $S_\mu(\delta)$ defined as in the statement of Lemma 12 and $T > S_\mu(\delta)$. Using that $(a+1)^2 \leq 2a^2 + 2$, we have

$$S_\mu(\delta) \geq \sup\left\{ T \mid T \leq \sum_{a\in\mathcal{A}} D_a(T,\delta) + n_0 K \right\} .$$

Then, we have

$$T - n_0 K > \sum_{a\in\mathcal{A}} D_a(T,\delta) \geq \min_{a\in\mathcal{A}_\theta} \left(t_a(T,\delta) - n_0 K\right) ,$$

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 11, we obtain that $\hat{a}_t \in \mathcal{A}_\theta$. This concludes the proof. □

Conclusion Let $S_\mu(\delta)$ as in Lemma 12. Combining Lemmas 12, 9 and 22, we obtain

$$\begin{aligned}
\mathbb{P}_\nu(\{\hat{a}_t = \emptyset\} \cup \{\hat{a}_t \in \mathcal{A}_\theta^{\complement}\}) &\leq K \inf\{\delta \mid T > S_\mu(\delta)\} \\
&\leq K e\sqrt{2}(2+\log T)\sqrt{\frac{T - n_0K - 2|\mathcal{A}_\theta|}{4H_1(\mu)}} \exp\left(-\frac{T - n_0K - 2|\mathcal{A}_\theta|}{4H_1(\mu)}\right) ,
\end{aligned}$$

This concludes the proof. □

B.3 Unverifiable Sample Complexity

In Appendix B.3, we study the unverifiable sample complexity (Katz-Samuels and Jamieson, 2020) (also discussed in Zhao et al. (2023)). This complexity is the minimum number of samples needed for an algorithm to output a correct answer with high probability. In particular, it does not require to check that the output is correct. Theorem 3 shows a deterministic upper bound on the unverifiable sample complexity of APGAI.

Theorem 3. *Let $\delta \in (0,1)$. The APGAI algorithm satisfies that, for any 1-sub-Gaussian distribution with mean μ such that $\mu_a \neq \theta$ for all $a \in \mathcal{A}$,*

$$\begin{aligned}
&\text{if } \mathcal{A}_\theta = \emptyset, \quad \mathbb{P}\left(\forall t > U_\delta(\mu),\ \hat{a}_t = \emptyset\right) \geq 1-\delta , \\
&\text{if } \mathcal{A}_\theta \neq \emptyset, \quad \mathbb{P}\left(\forall t > U_\delta(\mu),\ \hat{a}_t \in \mathcal{A}_\theta\right) \geq 1-\delta .
\end{aligned}$$

The time $U_\delta(\mu)$ is the unverifiable sample complexity of APGAI for GAI defined as

$$U_\delta(\mu) = \begin{cases} h_2(\log(1/\delta), 18H_1(\mu), n_0K) & \text{if } \mathcal{A}_\theta = \emptyset , \\ h_2(\log(1/\delta), 4H_1(\mu), n_0K + 2|\mathcal{A}_\theta|) & \text{otherwise} \quad . \end{cases} \tag{9}$$

The function $h_2(\log(1/\delta), A, B) := A\overline{W}_{-1}(\log(1/\delta) + \frac{B}{A} + \log(A))$ satisfies that $h_2(\log(1/\delta), A, B) =_{\delta\to 0} A\log(1/\delta) + \mathcal{O}(\log\log(1/\delta))$.

Proof. 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 the concentration events $\mathcal{E}_{T,\delta}$ as in (20) for $s = 0$, 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 $f_1(T,\delta) = \log(1/\delta) + \log T + \log K$. Using Lemma 21, we obtain that $\mathbb{P}_\nu(\mathcal{E}_{T,\delta}) \leq \delta$.

Case 1: when $\mathcal{A}_\theta = \emptyset$. Let $T_\mu(\delta)$ defined similarly as in Lemma 7, *i.e.*

$$T_\mu(\delta) := \sup\left\{T \mid T \leq 18H_1(\mu) f_1(T,\delta) + n_0K\right\} .$$

To prove Theorem 1 when $\mathcal{A}_\theta = \emptyset$, we obtain as an intermediary result that: for all $T > T_\mu(\delta)$, $\{\hat{a}_T \neq \emptyset\} \subseteq \mathcal{E}^{\complement}_{T,\delta}$. Using a proof similar to Lemma 27, applying Lemma 25 yields that

$$\begin{aligned} T > T_\mu(\delta) &\iff T > 18H_1(\mu)\log T + 18H_1(\mu)\log(1/\delta) + n_0 K \\ &\iff \frac{T}{18H_1(\mu)} - \log\left(\frac{T}{18H_1(\mu)}\right) > \log(1/\delta) + \frac{n_0 K}{18H_1(\mu)} + \log(18H_1(\mu)) \\ &\iff T > 18H_1(\mu)\overline{W}_{-1}\left(\log(1/\delta) + \frac{n_0 K}{18H_1(\mu)} + \log(18H_1(\mu))\right) , \end{aligned}$$

Let us define $U_\delta(\mu) := h_2(\log(1/\delta), 18H_1(\mu), n_0 K)$, where

$$h_2(\log(1/\delta), A, B) := A\overline{W}_{-1}\left(\log(1/\delta) + \frac{B}{A} + \log A\right)$$

satisfies that $h_2(\log(1/\delta), A, B) =_{\delta\to 0} A\log(1/\delta) + \mathcal{O}(\log\log(1/\delta))$. Hence, we have shown that

$$\{\exists T > U_\delta(\mu),\ \hat{a}_T \neq \emptyset\} \subseteq \mathcal{E}^{\complement}_{T,\delta} .$$

Using that $\mathbb{P}_\nu(\mathcal{E}_{T,\delta}) \leq \delta$, we can conclude that $\mathbb{P}_\nu(\forall T > U_\delta(\mu),\ \hat{a}_T = \emptyset) \geq 1 - \delta$.

Case 2: when $\mathcal{A}_\theta \neq \emptyset$. Let $S_\mu(\delta)$ defined similarly as in Lemma 12, *i.e.*

$$S_\mu(\delta) := \sup\{T \mid T \leq 4H_1(\mu) f_1(T,\delta) + n_0 K + 2|\mathcal{A}_\theta|\} .$$

To prove Theorem 1 when $\mathcal{A}_\theta \neq \emptyset$, we obtain as an intermediary result that: for all $T > T_\mu(\delta)$, $\{\hat{a}_T = \emptyset\} \cup \{\hat{a}_T \in \mathcal{A}_\theta^{\complement}\} \subseteq \mathcal{E}^{\complement}_{T,\delta}$. Using a proof similar to Lemma 27, applying Lemma 25 yields that

$$\begin{aligned} T > S_\mu(\delta) &\iff T > 4H_1(\mu)\log T + 4H_1(\mu)\log(1/\delta) + n_0 K + 2|\mathcal{A}_\theta| \\ &\iff \frac{T}{18H_1(\mu)} - \log\left(\frac{T}{4H_1(\mu)}\right) > \log(1/\delta) + \frac{n_0 K + 2|\mathcal{A}_\theta|}{4H_1(\mu)} + \log(4H_1(\mu)) \\ &\iff T > 4H_1(\mu)\overline{W}_{-1}\left(\log(1/\delta) + \frac{n_0 K + 2|\mathcal{A}_\theta|}{4H_1(\mu)} + \log(4H_1(\mu))\right) , \end{aligned}$$

Let us define $U_\delta(\mu) := h_2(\log(1/\delta), 4H_1(\mu), n_0 K + 2|\mathcal{A}_\theta|)$ where h_2 is as above. Then, we have shown that

$$\{\exists T > U_\delta(\mu),\ \hat{a}_T = \emptyset \ \vee\ \hat{a}_T \in \mathcal{A}_\theta^{\complement}\} \subseteq \mathcal{E}^{\complement}_{T,\delta} .$$

Using that $\mathbb{P}_\nu(\mathcal{E}_{T,\delta}) \leq \delta$, we can conclude that $\mathbb{P}_\nu(\forall T > U_\delta(\mu),\ \hat{a}_T \in \mathcal{A}_\theta) \geq 1 - \delta$. □

C ANALYSIS OF OTHER GAI ALGORITHMS

In Appendix C, we prove anytime guarantees on uniform sampling (Unif) in GAI (Appendix C.1), and 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

$$\hat{a}_T = \emptyset \quad \text{if } \max_{a\in\mathcal{A}} \hat{\mu}_a(T) \le \theta \quad \text{else} \quad \hat{a}_T \in \arg\max_{a\in\mathcal{A}} W_a^+(T)\,. \tag{10}$$

At time t such that $t/K \in \mathbb{N}$, the recommendation of Unif is equivalent to outputing 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 1. Since they have the same recommendation rule, this would allow us to measure the benefit of adaptive sampling.

Theorem 4 shows that the exponential decrease of the probability of error of Unif is linear as a function of time.

Theorem 4. *Let $\mathfrak{A}$ be the Unif algorithm with recommendation rule as in* (10)*. 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 $t/K \in \mathbb{N}$,*

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

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

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

Since the empirical are deterministic and the observations comes from a 1-sub-Gaussian with mean $\mu_a < \theta$, we obtain that for all $a \in \mathcal{A}$

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

Therefore, a direct union bound yields that

$$P^{\mathrm{err}}_{\nu,\mathfrak{A}}(T) \le \sum_{a\in[K]} \exp\left(-\frac{T\Delta_a^2}{2K}\right) \le K \exp\left(-\frac{T \min_{a\in\mathcal{A}} \Delta_a^2}{2K}\right)\,,$$

where we used that $H_6(\mu) = 1/\min_{a\in\mathcal{A}} \Delta_a^2$.

Case 2: $\mathcal{A}_\theta \neq \emptyset$. When $\mathcal{A}_\theta = \emptyset$, we have

$$\begin{aligned}\mathcal{E}^{\mathrm{err}}_\mu(T) &= \{\hat{a}_T = \emptyset\} \cup \{\hat{a}_T \in \mathcal{A}_\theta^\complement\} \\ &= \{\max_{a\in\mathcal{A}} \hat{\mu}_a(T) \le \theta\} \cup \{\max_{a\in\mathcal{A}} \hat{\mu}_a(T) > \theta,\, \arg\max_{a\in\mathcal{A}} W_a^+(T) \cap \mathcal{A}_\theta^\complement \neq \emptyset\}\,.\end{aligned}$$

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

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

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

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

Likewise, we obtain that

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

Therefore, we obtain

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

□

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

$$\hat{a}_T = \emptyset \quad \text{if } \hat{\mu}_{a_T}(T) \leq \theta \quad \text{else} \quad \hat{a}_T = a_T\,. \tag{11}$$

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 5 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 (Zhao et al., 2023, Theorem 1 and 5). Since one component of our proof uses their result, we suffer from this lack of explicit constant in that case.

Theorem 5. *Let $T > K$. Let $\mathfrak{A}_T$ be the SH-G algorithm with recommendation rule as in* (11)*. Then, for any* 1*-sub-Gaussian distribution $\nu \in \mathcal{D}^K$ with mean μ such that $\Delta_{\min} > 0$,*

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

where $I^\star = |\arg\max_{a\in\mathcal{A}}\mu_a|$ and $G_1(\mu)$ is defined in (12)*.*

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

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

Therefore, using $N_{a_T}(T) = N_T \geq \frac{T}{2\lceil \log_2(K)\rceil} - 1$ (since we drop observations from past phases) and similar argument as in the proof of Theorem 4, we obtain

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

Case 2: $\mathcal{A}_\theta \neq \emptyset$. When $\mathcal{A}_\theta = \emptyset$, we have

$$\mathcal{E}^{\mathrm{err}}_\mu(T) = \{\hat{a}_T = \emptyset\} \cup \{\hat{a}_T \in \mathcal{A}_\theta^{\complement}\} = \{a_T \in \mathcal{A}_\theta,\ \hat{\mu}_{a_T}(T) \leq \theta\} \cup \{\hat{a}_T \in \mathcal{A}_\theta^{\complement}\} .$$

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

$$\begin{aligned}\mathbb{P}_\nu(a_T \in \mathcal{A}_\theta,\ \hat{\mu}_{a_T}(T) \leq \theta) &\leq \sum_{a\in\mathcal{A}_\theta} \exp\left(-\frac{\Delta_a^2}{2}\left(\frac{T}{2\lceil\log_2(K)\rceil} - 1\right)\right) \\ &\leq |\mathcal{A}_\theta| e^{\min_{a\in\mathcal{A}_\theta}\Delta_a^2/2} \exp\left(-\frac{T\min_{a\in\mathcal{A}_\theta}\Delta_a^2}{4\lceil\log_2(K)\rceil}\right) .\end{aligned}$$

Let $a^\star \in \arg\max_{a\in\mathcal{A}} \mu_a$. Since $\{\hat{a}_T \notin \mathcal{A}_\theta^{\complement}\} \subset \{\hat{a}_T \neq a^\star\}$, using (Karnin et al., 2013, Theorem 4.1) yields

$$\mathbb{P}_\nu(\hat{a}_T \in \mathcal{A}_\theta^{\complement}) \leq \mathbb{P}_\nu(\hat{a}_T \neq a^\star) \leq 3\log_2(K)\exp\left(-\frac{T}{8\log_2(K)\max_{i>I^\star} i(\max_{a\in\mathcal{A}}\mu_a - \mu_{(i)})^{-2}}\right) .$$

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

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

$$\{\hat{a}_T \in \mathcal{A}_\theta^{\complement}\} = \bigcup_{\varepsilon\in(\max_{a\in\mathcal{A}_\theta}\Delta_a + \min_{b\notin\mathcal{A}_\theta}\Delta_b, \max_{a\in\mathcal{A}_\theta}\mu_a - \min_{b\in\mathcal{A}_\theta}\mu_b)} \{\mu_{\hat{a}_T} < \mu_{a^\star} - \varepsilon\} .$$

Then, using (Zhao et al., 2023, Theorem 1) and taking the infimum over ε yields that

$$\mathbb{P}_\nu(\hat{a}_T \in \mathcal{A}_\theta^{\complement}) \leq \exp\left(-\tilde{\Theta}\left(-\frac{T}{G_1(\mu)}\right)\right) ,$$

with

$$G_1(\mu) = \min_{\varepsilon\in(\max_{a\in\mathcal{A}_\theta}\Delta_a + \min_{b\notin\mathcal{A}_\theta}\Delta_b, \max_{a\in\mathcal{A}_\theta}\mu_a - \min_{b\in\mathcal{A}_\theta}\mu_b)} \max_{i\geq g(\varepsilon)+1} \frac{i}{g(\varepsilon/2)(\mu_{a^\star} - \mu_{(i)})^2} , \tag{12}$$

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

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. (Zhao et al., 2023, Theorem 5) 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 5 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 (11). 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

$$n_k = \left\lceil \frac{T-K}{\overline{\log}(K)(K+1-k)} \right\rceil \quad \text{and} \quad u_T = \sum_{k=1}^{K-1} n_k \,,$$

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

Theorem 6. *Let $T > K$. Let $\mathfrak{A}_T$ be the SR-G algorithm with recommendation rule as in* (11)*. Then, for any* 1*-sub-Gaussian distribution $\nu \in \mathcal{D}^K$ with mean μ such that $\Delta_{\min} > 0$,*

$$\begin{aligned}
&\text{if } \mathcal{A}_\theta = \emptyset, \quad P^{err}_{\nu,\mathfrak{A}_T}(T) \leq K \exp\left(-\frac{T-K}{4\overline{\log}(K)} \min_{a\in\mathcal{A}} \Delta_a^2\right) \,, \\
&\text{if } \mathcal{A}_\theta \neq \emptyset, \quad P^{err}_{\nu,\mathfrak{A}_T}(T) \leq |\mathcal{A}_\theta| \exp\left(-\frac{T-K}{4\overline{\log}(K)} \min_{a\in\mathcal{A}_\theta} \Delta_a^2\right) \\
&+ \min\left\{ K^2 \exp\left(-\frac{T-K}{\overline{\log}(K) \max_{i>I^\star} i(\max_{a\in\mathcal{A}} \mu_a - \mu_{(i)})^{-2}}\right), \frac{(K-1)!}{(|\mathcal{A}_\theta^\complement|-1)!} |\mathcal{A}_\theta^\complement| \exp\left(-\frac{T-K}{4\overline{\log}(K) G_2(\mu)}\right)\right\} \,,
\end{aligned}$$

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

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

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

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

Therefore, using $N_{a_T}(T) = T - u_T$ and similar argument as in the proof of Theorem 4, we obtain

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

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

Case 2: $\mathcal{A}_\theta \neq \emptyset$. When $\mathcal{A}_\theta = \emptyset$, we have

$$\mathcal{E}^{\text{err}}_\mu(T) = \{\hat{a}_T = \emptyset\} \cup \{\hat{a}_T \in \mathcal{A}_\theta^\complement\} = \{a_T \in \mathcal{A}_\theta,\, \hat{\mu}_{a_T}(T) \leq \theta\} \cup \{\hat{a}_T \in \mathcal{A}_\theta^\complement\} \,.$$

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

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

Let $a^\star \in \arg\max_{a\in\mathcal{A}} \mu_a$. Since $\{\hat{a}_T \in \mathcal{A}_\theta^\complement\} \subset \{\hat{a}_T \neq a^\star\}$, using (Audibert et al., 2010, Theorem 2) yields

$$\mathbb{P}_\nu(\hat{a}_T \in \mathcal{A}_\theta^\complement) \leq \mathbb{P}_\nu(\hat{a}_T \neq a^\star) \leq \frac{K(K-1)}{2} \exp\left(-\frac{T-K}{\overline{\log}(K) \max_{i>I^\star} i(\max_{a\in\mathcal{A}} \mu_a - \mu_{(i)})^{-2}}\right) \,.$$

Improved case 2. As in the proof of Theorem 5, using $\{\hat{a}_T \in \mathcal{A}_\theta^\complement\} \subset \{\hat{a}_T \neq a^\star\}$ 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 $\{\hat{a}_T \in \mathcal{A}_\theta^\complement\}$ 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 4, 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

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

where we used that $n_k \geq \frac{T-K}{\overline{\log}(K)(K+1-k)}$ 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 withing 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

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

where

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

□

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 6 can be modified for DSR-G by simply adding the multiplicative factor 4 in front of each hardness constant.

D PRIOR KNOWLEDGE-BASED GAI ALGORITHMS (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 7) and the uniform round-robin version (Theorem 8).

D.1 A Meta-Algorithm For Fixed-Budget GAI

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 $([\widehat{\Delta}_a^-(t), \widehat{\Delta}_a^+(t)])_{t\leq T}$ 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, $\sigma = 1$)

$$\left[\widehat{\Delta}_a^-(t), \widehat{\Delta}_a^+(t)\right] := \left\{ \hat{\mu}_a(t) - \theta \pm \sigma\sqrt{\frac{\beta(t)}{N_a(t)}} \right\} ,$$

where β is a well-chosen threshold function, which is increasing in its argument.

Intuitively, $\widehat{\Delta}_a^-(t)$ (resp. $\widehat{\Delta}_a^+(t)$) 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.

Algorithm 2 PKGAI (**P**rior **K**nowledge-based GAI)

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

$$\begin{aligned}
\text{PKGAI(APT}_P\text{)}: &\quad i_a(t) := \sqrt{N_a(t)}(\hat{\mu}_a(t) - \theta) , \\
\text{PKGAI(UCB)}: &\quad i_a(t) := \widehat{\Delta}_a^+(t) , \\
\text{PKGAI(Unif)}: &\quad i_a(t) := -N_a(t) , \ \text{// minimizing that index} \\
\text{PKGAI(LCB-G)}: &\quad i_a(t) := \sqrt{N_a(t)}\widehat{\Delta}_a^-(t) .
\end{aligned}$$

Sample each arm $a \in \mathcal{A}$ once
Set $t \leftarrow K$, $\mathcal{S}_t \leftarrow \mathcal{A}$, $N_a(t) \leftarrow 1$ and initialize $\widehat{\Delta}_a^-(t), \widehat{\Delta}_a^+(t)$ for $a \in \mathcal{A}$
while $t < T$ **and** $|\mathcal{S}_t| > 0$ **do**
  $a_{t+1} \in \arg\max_{a\in\mathcal{S}_t} i_a(t)$
  Draw arm a_{t+1} and observe $X_{a_{t+1},t+1}$
  Update $\widehat{\Delta}_a^-(t+1)$, $\widehat{\Delta}_a^+(t+1)$ for all $a \in \mathcal{A}$
  $\mathcal{S}_{t+1} \leftarrow \mathcal{S}_t \setminus \{a \in \mathcal{S}_t \mid \widehat{\Delta}_a^+(t+1) < 0\}$
  $t \leftarrow t + 1$
end
if $|\mathcal{S}_t| = 0$ **or** $\max_{a\in\mathcal{S}_T} \widehat{\Delta}_a^-(T) + \widehat{\Delta}_a^+(T) \leq 0$ **then**
  return $\hat{a}_T := \emptyset$
else
  return $\hat{a}_T \in \arg\max_{a\in\mathcal{S}_T} \widehat{\Delta}_a^-(T)$
end

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

$$\widehat{\Delta}_a^-(T) + \widehat{\Delta}_a^+(T) > 0 \ ,$$

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 strategy, named LCB-G, which targets the lower confidence bound.

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 7 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 7 (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 $[\widehat{\Delta}_a^-(t), \widehat{\Delta}_a^+(t)]$ for all arm $a \in \mathcal{A}$ and $t \leq T$ are such that*

$$\mathbb{P}_\nu \left(\bigcup_{\substack{a \in \mathcal{A} \\ t \leq T}} \left\{ |\hat{\mu}_a(t) - \mu_a| \leq \sqrt{\frac{\beta(t)}{N_a(t)}} \right\} \right) \in (0,1) \ , \ \textit{with } \beta(T) \leq \frac{T-K}{4H_1(\mu)} \ . \tag{13}$$

Then, we have $P^{err}_{\nu, PKGAI(*)}(T) \leq 2KTe^{-2\beta(T)}$. This is minimized when Inequality (13) is an equality, hence

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

Furthermore, when considering an uniform round-robin sampling, *i.e.* PKGAI(Unif) (in Line L.7, Algorithm 2)

$$\forall a \in \mathcal{A} \ \forall t \leq T, \ i_a(t) := -N_a(t) \ ,$$

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 8).

Theorem 8 (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*

$$\beta(T) \leq \begin{cases} (T-K)/(4K\hat{\Delta}^{-2}) & \textit{if } \mathcal{A}_\theta(\mu) \neq \emptyset \\ (T-K)/(4H_1(\mu)) & \textit{otherwise} \end{cases} \ . \tag{14}$$

where $\hat{\Delta} := \max_{a \in \mathcal{A}_\theta} \Delta_a + \min_{a \notin \mathcal{A}_\theta} \Delta_a$. Then $P^{err}_{\nu, PKGAI[Unif]}(T) \leq 2KTe^{-2\beta(T)}$. This is minimized when Inequality (14) is an equality, hence

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

This theorem yields a strictly better bound than APGAI and Theorem 7 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 4).

D.3 Proof Sketch

The idea behind the proofs of Theorems 7 and 8 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,

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

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

$$\begin{aligned} P^{\mathrm{err}}_{\nu,\mathfrak{A}}(T) &\leq \mathbb{P}\left[(\mathcal{A}_\theta \neq \emptyset \cap (\hat{a}_\tau \in \{\emptyset\}\cup\mathcal{A}\setminus\mathcal{A}_\theta)\cap\mathcal{E}_T)\cup(\mathcal{A}_\theta = \emptyset \cap \hat{a}_\tau \neq \emptyset \cap \mathcal{E}_T)\right] + \mathbb{P}_\nu(\mathcal{E}_T^{\complement})\,,\\ &\leq \underbrace{\mathbb{P}\left[\mathcal{A}_\theta \neq \emptyset \cap (\hat{a}_\tau \in \{\emptyset\}\cup\mathcal{A}\setminus\mathcal{A}_\theta)\cap\mathcal{E}_T\right]}_{\text{Case 1}} + \underbrace{\mathbb{P}\left[\mathcal{A}_\theta = \emptyset \cap \hat{a}_\tau \neq \emptyset \cap \mathcal{E}_T\right]}_{\text{Case 2}} + \mathbb{P}_\nu(\mathcal{E}_T^{\complement})\,. \end{aligned} \tag{15}$$

For both Theorems 7 and 8, 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, $P^{\mathrm{err}}_{\nu,\mathrm{PKGAI}}(T) \leq \mathbb{P}_\nu(\mathcal{E}_T^{\complement})$.

D.4 Proof Of Theorem 7

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 $\mathbb{P}\left[\mathcal{E}_T^{\text{Case 1}} \cap \hat{a}_\tau \neq \emptyset\right]$ is lesser than

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

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 $b \in \mathcal{S}_T \subseteq \mathcal{A}_\theta^{\complement}$, on $\mathcal{E}_T$

$$\widehat{\Delta}_b^+(T) \geq 0 \implies -\Delta_b + 2\sqrt{\frac{\beta(T)}{N_b(T)}} \geq 0 \implies N_b(T) \leq \frac{4\beta(T)}{\Delta_b^2} < \frac{4\beta(T)}{\Delta_b^2} + 1\,. \tag{16}$$

Moreover, for any arm $c \in \mathcal{S}_T^{\complement} \subseteq \mathcal{A}_\theta^{\complement}$, 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 (16) on round $t_c - 1$,

$$\widehat{\Delta}_c^+(t_c-1) \geq 0 > \widehat{\Delta}_c^+(t_c) \implies N_c(T) - 1 = N_c(t_c-1) \leq \frac{4\beta(t_c-1)}{\Delta_c^2} \leq \frac{4\beta(T)}{\Delta_c^2} \implies N_c(T) \leq \frac{4\beta(T)}{\Delta_c^2} + 1\,. \tag{17}$$

Combining inequalities 16 and 17, 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

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

□

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 $\max_{a \in \mathcal{S}_\tau} \widehat{\Delta}_a^-(\tau) + \widehat{\Delta}_a^+(\tau) \leq 0$ (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

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

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 $\max_{a \in \mathcal{S}_T} \widehat{\Delta}_a^-(T) + \widehat{\Delta}_a^+(T) \leq 0$ on event $\mathcal{E}_T$, then since $\mathcal{E}_T$ holds

$$\forall a \in \mathcal{A}_\theta \subseteq \mathcal{S}_T,\ 2\left(\Delta_a - \sqrt{\frac{\beta(T)}{N_a(T)}}\right) \leq \widehat{\Delta}_a^-(T) + \widehat{\Delta}_a^+(T) \leq 0 \implies N_a(T) \leq \frac{\beta(T)}{\Delta_a^2} < \frac{\beta(T)}{\Delta_a^2} + 1 . \tag{18}$$

Furthermore, as a direct consequence of Inequalities 16 and 17, 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

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

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, $\mathbb{P}\left[\mathcal{E}_T^{\text{Case 2}} \cap \hat{a}_\tau = \emptyset\right] = 0$.

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

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

Furthermore, as Inequalities 16 and 17 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, $\mathbb{P}\left[\mathcal{E}_T^{\text{Case 2}} \cap \hat{a}_T \in \mathcal{A} \setminus \mathcal{A}_\theta\right] = 0$.

□

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 (15), using successively the Hoeffding concentration bounds and union bounds over $\mathcal{A}$ of size K and over $\{1, 2, \ldots, T\}$

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

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 8

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 7 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

$$P^{\text{err}}_{\nu,\text{PKGAI(Unif)}}(T) \leq 2KT\exp(-2\beta(T)) \ ,$$

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 $\mathbb{P}\left[\mathcal{E}_T^{\text{Case 2}} \cap \hat{a}_T = \emptyset\right]$. Then, necessarily, according to Case 2.1 in the proof of Theorem 7

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

$\mathcal{A}_\theta \subseteq \mathcal{S}_T$ (otherwise, we end up with a contradiction with event $\mathcal{E}_T$). Moreover, if $\max_{a\in\mathcal{S}_T} \widehat{\Delta}_a^-(T) + \widehat{\Delta}_a^+(T) \leq 0$, then Inequality (18) applies. Finally, since PKGAI(Unif) uses a uniform sampling, $N_a(T) \geq \left\lfloor \frac{T}{K} \right\rfloor \geq \frac{T}{K} - 1$ for any arm a. Combining all of this yields the following inequalities

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

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 $\mathbb{P}\left[\mathcal{E}_T^{\text{Case 2}} \cap \hat{a}_T \in \mathcal{A}\setminus\mathcal{A}_\theta\right]$. 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 $\mathcal{E}_T$

$$\exists b \notin \mathcal{A}_\theta\ \forall a \in \mathcal{A}_\theta,\ -\Delta_b \geq_{b\notin\mathcal{A}_\theta} \widehat{\Delta}_b^-(T) \geq \widehat{\Delta}_a^-(T) \geq_{a\in\mathcal{A}_\theta} \Delta_a - 2\sqrt{\frac{\beta(T)}{N_a(T)}} \geq \Delta_a - 2\sqrt{\frac{\beta(T)}{\min_{c\in\mathcal{A}_\theta} N_c(T)}}$$

$$\implies \max_{b\notin\mathcal{A}_\theta}(-\Delta_b) \geq \max_{a\in\mathcal{A}_\theta}\Delta_a - 2\sqrt{\frac{\beta(T)}{\min_{c\in\mathcal{A}_\theta} N_c(T)}}$$

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 7, 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

$$P^{\mathrm{err}}_{\nu,\mathrm{PKGAI(Unif)}}(T) \leq 2KT \exp(-2\beta(T)) \ ,$$

which is minimized when the inequalities on $\beta(T)$ are equalities. □

E LOWER BOUND FOR FIXED-CONFIDENCE 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 (5).

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 (Kaufmann et al., 2016, Lemma 1).

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

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

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 strategy. Combining (Kaufmann et al., 2016, Lemma 1) with the δ-correctness of the algorithm and the monotonicity of function kl, for any 1-Gaussian distribution κ of mean $\lambda \in \text{Alt}(\mu)$

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

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

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

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

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

Otherwise, $\mathcal{A}_\theta(\mu) \neq \emptyset$, and then

$$\sup_{\omega \in \triangle_K} \inf_{\lambda \in \text{Alt}(\mu)} \sum_{a \in \mathcal{A}} \omega_a (\mu_a - \lambda_a)^2 = \sup_{\omega \in \triangle_K} \sum_{a \in \mathcal{A}_\theta(\mu)} \omega_a (\mu_a - \theta)^2 = \max_{a \in \mathcal{A}_\theta} \Delta_a^2 \text{ and } \omega_a := \mathbb{1}(a = \arg\max_{a \in \mathcal{A}_\theta} \mu_a) \,.$$

All in all,

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

□

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

$$\mathrm{GLR}_t^{\mathcal{M}}(\Lambda) = \log \frac{\sup_{\tilde{\mu}\in\mathcal{M}} \mathcal{L}_{\tilde{\mu}}(X_1, \dots, X_t)}{\sup_{\lambda\in\Lambda} \mathcal{L}_{\lambda}(X_1, \dots, X_t)} \,.$$

In the case of independent Gaussian distributions with unit variance, the likelihood ratio for two models with mean vectors $\xi, \lambda \in \mathcal{M}$,

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

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

$$\mathrm{GLR}_t^{\mathcal{M}}(\Lambda) = \frac{1}{2} \min_{\lambda\in\Lambda} \sum_{a\in\mathcal{A}} N_a(t)(\hat{\mu}_a(t) - \lambda_a)^2 \,.$$

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 $\mathrm{Alt}(\hat{\mu}(t)) = \bigcup_{a\in\mathcal{A}} \{\lambda \in \mathbb{R}^K \mid \lambda_a > \theta\}$. By direct manipulations similar to the ones in Appendix E.1, the corresponding GLR can be written as

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

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 $\mathrm{Alt}(\hat{\mu}(t), a) = \{\lambda \in \mathbb{R}^K \mid \lambda_a \leq \theta\}$. By direct manipulations similar to the ones in Appendix E.1, the corresponding GLR can be written as

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

F ANALYSIS OF APGAI: THEOREM 2

When combined with the GLR stopping (5) using threshold (6), APGAI becomes dependent of a confidence $\delta \in (0,1)$. However, it is still independent of a budget T, hence T is an analysis parameter in the following.

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 $T > n_0 K$ and $\mathcal{E}_T = \mathcal{E}_{T,1}$ where $\mathcal{E}_{T,\delta}$ as in (20), *i.e.*

$$\mathcal{E}_T = \left\{ \forall a \in \mathcal{A}, \forall t \leq T,\ |\hat{\mu}_a(t) - \mu_a| < \sqrt{\frac{2f_1(T)}{N_a(t)}} \right\} , \tag{19}$$

with $f_1(T) = (1+s)\log T$. Using Lemma 21, we have $\sum_{T>K} \mathbb{P}_\nu(\tilde{\mathcal{E}}_T^{\complement}) \leq K\zeta(s)$ where ζ is the Riemann ζ function. Suppose that we have constructed a time $T_\mu(\delta) > n_0 K$ such that $\mathcal{E}_T \subseteq \{\tau_\delta \leq T\}$ for $T \geq T_\mu(\delta)$. Then, using Lemma 24, we obtain

$$\mathbb{E}_\nu[\tau_\delta] \leq T_\mu(\delta) + K\zeta(s) .$$

To prove Theorem 2, 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 1, our main technical tool is Lemma 3. It is direct to see that Lemmas 7 and 12 can be adapted to hold for $\mathcal{E}_T$ and $f_1(T) = (1+s)\log T$. Combined with Lemma 27, we state those results in a more explicit form, and omit the proof for the sake of space.

Lemma 13. *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 $T_\mu = h_1(18(1+s)H_1(\mu), n_0 K)$ where h_1 is defined in Lemma 27. For all $T > T_\mu$, under the event $\mathcal{E}_T$ as in (19), we have $N_a(T) > 2f_1(T)\Delta_a^{-2}$ for all $a \in \mathcal{A}$.*

Proof. Let us define $\tilde{T}_\mu = \sup\left\{T \mid T \leq 18(1+s)H_1(\mu)\tilde{f}_1(T,\delta) + n_0 K\right\}$. Using Lemma 27, we obtain $\tilde{T}_\mu \leq T_\mu$ where $T_\mu = h_1(18(1+s)H_1(\mu), n_0 K)$. Combined with the proof of Lemma 7, this concludes the proof. □

Lemma 14. *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 $S_\mu = h_1(4(1+s)H_1(\mu), n_0 K + 2|\mathcal{A}_\theta|)$ where h_1 is defined in Lemma 27. For all $T > S_\mu$, under the event $\mathcal{E}_T$ as in (19), we have $\hat{a}_t \in \mathcal{A}_\theta$ and there exists $a \in$ such that $N_a(T) > (\Delta_a^{-1}\sqrt{2\tilde{f}_1(T,\delta)} + 1)^2$.*

Proof. Let us define $\tilde{S}_\mu = \sup\{T \mid T \leq 4(1+s)H_1(\mu)\log T + n_0 K + 2|\mathcal{A}_\theta|\}$. Using Lemma 27, we obtain $\tilde{S}_\mu \leq S_\mu$ where $S_\mu = h_1(4(1+s)H_1(\mu), n_0 K + 2|\mathcal{A}_\theta|)$. Combined with the proof of Lemma 12, this concludes the proof. □

Theorem 2 is obtained by combining Lemmas 16 and 18.

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

When $\mathcal{A}_\theta = \emptyset$, we will have $\tau_\delta = \tau_{<,\delta}$ almost surely and, for T large enough, $\hat{a}_T = \emptyset$ and $a_{T+1} \in \arg\min_{a\in\mathcal{A}} W_a^-(T)$. Lemma 15 formalizes this intuition.

Lemma 15. *Let $s > 1$. Let $T_\mu = h_1(18(1+s)H_1(\mu), n_0 K)$ where h_1 is defined in Lemma 27. For all $T > T_\mu$, $a_{T+1} \in \arg\min_{a\in\mathcal{A}} W_a^-(T)$ and $\hat{a}_T = \emptyset$. Moreover, we have $\tau_\delta = \tau_{<,\delta}$ almost surely.*

Proof. Let T_μ as in Lemma 7. Let $T > T_\mu$. Using Lemma 7, we obtain that $N_a(T) > \frac{2f_1(T)}{(\theta-\mu_a)^2}$ for all $a \in \mathcal{A}$. Then, under $\mathcal{E}_T$ as in (19),

$$\forall a \in \mathcal{A}, \quad \hat{\mu}_a(t) \leq \mu_a + \sqrt{\frac{2f_1(T)}{N_a(T)}} < \theta ,$$

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 $\hat{a}_T = \emptyset$. A direct consequence is that $\tau_{>,\delta} = +\infty$, hence $\tau_{<,\delta} = \tau_\delta$ almost surely. □

When coupled with the GLR stopping (5) using threshold (6), Lemma 16 gives an upper bound on the expected sample complexity of APGAI when $\mathcal{A}_\theta = \emptyset$.

Lemma 16. *Let $\delta \in (0,1)$. Combined with GLR stopping (5) using threshold (6), 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 (1) and $T_\mu = h_1(54H_1(\mu), n_0 K)$ with h_1 is defined in Lemma 27 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 $D_1(\mu) = T_\mu + 2H_1(\mu)\left(\theta - \min_{a\in\mathcal{A}} \mu_a\right)^2 - 6H_1(\mu)\log T_\mu$. In particular, it satisfies

$$\limsup_{\delta \to 0} \mathbb{E}_\nu[\tau_\delta] / \log(1/\delta) \leq 2H_1(\mu) \,.$$

Proof. Let T_μ as in Lemma 15. Let $T > T_\mu$ such that $\mathcal{E}_T \cap \{\tau_\delta > T\}$ holds true. Let $w \in \triangle_K$ 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 15 yields that $a_1 = a_{t_{a_1}+1} \in \arg\min_{a\in\mathcal{A}} W_a^-(t_{a_1})$. Moreover, we have

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

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 (19), using that $a_1 = a_{t_{a_1}+1} \in \arg\min_{a\in\mathcal{A}} W_a^-(t_{a_1})$, we obtain

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

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 $\mathcal{E}_T \subset \{\tau_{<,\delta} \leq T\} = \{\tau_\delta \leq T\}$ (by using Lemma 15). Using Lemma 24, 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 28, direct manipulations show that

$$\limsup_{\delta\to 0} \frac{\mathbb{E}_\nu[\tau_\delta]}{\log(1/\delta)} \leq \limsup_{\delta\to 0} \frac{C_\mu(\delta)}{\log(1/\delta)} \leq 2H_1(\mu)\,.$$

According to Lemma 1, we have proven asymptotic optimality. Lemma 2 gives the δ-correctness of the APGAI algorithm since the recommendation rule of matches the one of Lemma 2.

□

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

When $\mathcal{A}_\theta \neq \emptyset$, we will have $\tau_\delta = \tau_{>,\delta}$ almost surely and, for T large enough, $\hat{a}_T = a_{T+1} \in \arg\max_{a\in\mathcal{A}_\theta} W_a^+(T)$. Lemma 17 formalizes this intuition.

Lemma 17. *Let $s > 1$. Let $S_\mu = h_1(4(1+s)H_1(\mu), n_0K + 2|\mathcal{A}_\theta|)$ where h_1 is defined in Lemma 27. For all $T > S_\mu$, $\hat{a}_T = a_{T+1} \in \arg\max_{a\in\mathcal{A}_\theta} W_a^+(T)$. Moreover, we have $\tau_\delta = \tau_{>,\delta}$ almost surely.*

Proof. Let S_μ as in Lemma 14 Let $T > S_\mu$. Using Lemma 12, there exists $a \in \mathcal{A}_\theta$ such that $N_a(T) > \frac{2f_1(T)}{(\mu_a-\theta)^2}$. Then, we have

$$\hat{\mu}_a(t) \geq \mu_a - \sqrt{\frac{2f_1(T)}{N_a(T)}} > \theta\,,$$

hence $\max_{a\in\mathcal{A}} \hat{\mu}_a(t) > \theta$. Using Lemma 11 and the definition of the recommendation rule when $\max_{a\in\mathcal{A}} \hat{\mu}_a(t) > \theta$, we obtain that $\hat{a}_T = a_{T+1} \in \mathcal{A}_\theta$. Using the definition of the sampling rule when $\max_{a\in\mathcal{A}} \hat{\mu}_a(t) > \theta$, for all $T > S_\mu$, we have $\hat{a}_T = a_{T+1} \in \arg\max_{a\in\mathcal{A}_\theta} W_a^+(T)$. A direct consequence is that $\tau_{<,\delta} = +\infty$, hence $\tau_{>,\delta} = \tau_\delta$ almost surely. □

When coupled with the GLR stopping (5) using threshold (6), Lemma 18 gives an upper bound on the expected sample complexity of APGAI when $\mathcal{A}_\theta \neq \emptyset$.

Lemma 18. *Let $\delta \in (0,1)$. Combined with GLR stopping (5) using threshold (6), 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 (1) and $S_\mu = h_1(12H_1(\mu), n_0K + 2|\mathcal{A}_\theta|)$ with h_1 is defined in Lemma 27 and

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

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

Proof. Let S_μ as in Lemma 17. Let $T > S_\mu$ such that $\mathcal{E}_T \cap \{\tau_\delta > T\}$ holds true. Using Lemma 17, we know that $a_{t+1} \in \mathcal{A}_\theta$ for all $t \in (S_\mu, T]$. Direct summation yields that

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

At time $S_\mu + 1$, let $a_1 \in \mathcal{A}_\theta$ as in Lemma 17, *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

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

Therefore, we have shown that

$$\sum_{a\in\mathcal{A}_\theta} N_a(T) \geq T - g(S_\mu) \quad \text{with} \quad g(S_\mu) = S_\mu - \frac{2f_1(S_\mu)}{\max_{a\in\mathcal{A}_\theta}\Delta_a^2} + 1\,.$$

Let $A_\theta = |\mathcal{A}_\theta|$ and $w \in \triangle_{A_\theta}$ such that $w_a = (\mu_a - \theta)^{-2} H_\theta(\mu)^{-1}$ with $H_\theta(\mu)$ as in (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 us define

$$E_\mu(\delta) = \sup\left\{T \mid T \leq g(S_\mu) + 2H_\theta(\mu) f_1(T)\right\} .$$

Let $T > E_\mu(\delta)$. 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 $D_\mu(\delta) \geq E_\mu(\delta) \geq S_\mu$. Therefore, we have shown that for $T \geq D_\mu(\delta) + 1$, we have $\mathcal{E}_T \subset \{\tau_{>,\delta} \leq T\} = \{\tau_\delta \leq T\}$ (by using Lemma 17). Using Lemma 24, 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 28, direct manipulations show that

$$\limsup_{\delta \to 0} \frac{\mathbb{E}_\nu[\tau_\delta]}{\log(1/\delta)} \leq \limsup_{\delta \to 0} \frac{D_\mu(\delta)}{\log(1/\delta)} \leq 2H_\theta(\mu) .$$

According to Lemma 1, our result is weaker than asymptotic optimality when $|\mathcal{A}_\theta| \geq 2$. Lemma 2 gives the δ-correctness of the APGAI algorithm since the recommendation rule of matches the one of Lemma 2. □

F.2.1 Discussion on Sub-optimal Upper Bound

As discussed in Section 4, Theorem 2 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 (20) for $s=1$ and $\delta=1$, *i.e.*

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

It satisfies $\mathbb{P}_\nu(\mathcal{E}_T^\complement)\leq 2/T$. 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 $t\leq T$,

$$N_2(t)=t-1\quad\text{and}\quad N_1(t)=1\,.$$

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 there exists $(a_1,a_2)\in\mathcal{A}_\theta(\mu)^2$ such that $\min_{a\in\{a_1,a_2\}}N_a(T)\to_{T\to+\infty}+\infty$. Let S_μ as in Lemma 17. Let $T>S_\mu$ such that $\mathcal{E}_T\cap\{\tau_\delta>T\}$ holds true. In the proof of Lemma 18, 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 $\min_{a\in\{a_1,a_2\}}N_a(T)\to_{T\to+\infty}+\infty$, 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 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 $N_a(T)\to_{T\to+\infty}+\infty$.

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.

G CONCENTRATION RESULTS

In Appendix G.1, we prove the δ-correctness of the GLR stopping rule (5) with threshold (6) (Lemma 2). Appendix G.2 gathers sequence of concentration events which are used for our proofs.

G.1 Analysis of the GLR Stopping Rule: Lemma 2

Proving δ-correctness of a GLR stopping rule is done by leveraging concentration results. In particular, we build upon (Jourdan et al., 2023a, Lemma 28). Lemma 19 is obtained as a Corollary of (Jourdan et al., 2023a, Lemma 28) 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 19 (Lemma 28 in Jourdan et al. (2023a))**.** *Let $s > 1$ and $\delta \in (0,1)$. Let $\overline{W}_{-1}(x) = -W_{-1}(-e^{-x})$ for all $x \geq 1$ (see Lemma 25), where W_{-1} is the negative branch of the Lambert W function. Let*

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

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 19 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 19.

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 19.

G.2 Sequence of Concentration Events

Appendix G.2 provides sequence of concentration events which are used for our proofs. Lemma 20 is a standard concentration result for sub-Gaussian distribution, hence we omit the proof.

Lemma 20. *Let X be an observation from a sub-Gaussian distribution with mean 0 and variance $\sigma^2 = 1$. Then, for all $\delta \in (0,1]$,*

$$\mathbb{P}_X\left(|X| \geq \sqrt{2\log(1/\delta)}\right) \leq \delta\,.$$

Lemma 21 gives a sequence of concentration events under which the empirical means are close to their true values.

Lemma 21. *Let $\delta \in (0,1]$ and $s \geq 0$. For all $T > K$, let $f_1(T,\delta) = \log(1/\delta) + (1+s)\log T$ and*

$$\mathcal{E}_{T,\delta} = \left\{\forall a \in \mathcal{A}, \forall t \leq T,\, |\hat{\mu}_a(t) - \mu_a| < \sqrt{\frac{2f_1(T,\delta)}{N_a(t)}}\right\}\,. \tag{20}$$

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

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 $\hat{\mu}_a(t) - \mu_a = \frac{1}{N_a(t)}\sum_{s=1}^t \mathbb{1}(a_s = a)\, X_{s,a}$ and concentration results for sub-Gaussian observations (Lemma 20). □

Lemma 22 provides concentration results on the empirical means, which are tighter than the one obtained in Lemma 21.

Lemma 22. *Let $\delta \in (0,1]$ and $s \geq 0$. Let $\overline{W}_{-1}(x) = -W_{-1}(-e^{-x})$ for all $x \geq 1$ (see Lemma 25), where W_{-1} is the negative branch of the Lambert W function. For all $T > K$, let*

$$\tilde{f}_1(T,\delta) = \frac{1}{2}\overline{W}_{-1}(2\log(1/\delta) + 2s\log T + 2\log(2+\log T) + 2)\,, \tag{21}$$

and

$$\tilde{\mathcal{E}}_{T,\delta} = \left\{\forall a \in \mathcal{A}, \forall t \leq T,\, |\hat{\mu}_a(t) - \mu_a| < \sqrt{\frac{2\tilde{f}_1(T,\delta)}{N_a(t)}}\right\}\,. \tag{22}$$

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

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

$$\overline{M}(t) = \sum_{i\in[D]} w_i \int f_{N_i,\gamma}(x) \exp\left(xS_t - \frac{1}{2}x^2 t\right) dx \,,$$

which satisfies $\overline{M}(0) = 1$. It is direct to see that $M(t) = \exp\left(xS_t - \frac{1}{2}x^2 t\right)$ is a non-negative supermartingale since sub-Gaussian distributions with mean 0 and variance $\sigma^2 = 1$ satisfy

$$\forall \lambda \in \mathbb{R}, \quad \mathbb{E}_X[\exp(sX)] \leq \exp(\lambda^2/2) \,.$$

By Tonelli's theorem, then $\overline{M}(t)$ 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

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

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

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

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

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

Since this upper bound is independent of t, we can optimize it and choose γ as in Lemma 23.

Lemma 23 (Lemma A.3 in Degenne (2019))**.** *For $a, b \geq 1$, the minimal value of $f(\eta) = (1+\eta)(a + \log(b + \frac{1}{\eta}))$ is attained at $\eta^\star$ such that $f(\eta^\star) \leq 1 - b + \overline{W}_{-1}(a+b)$. If $b = 1$, then there is equality.*

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

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

The second inequality is obtained since $D \leq 1 + \frac{\log T}{\log(1+\eta)}$. The last equality is obtained for the choice $\eta^\star = e^2 - 1$, 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

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

Since $3 - 2\log 2 \leq 2$ and $\overline{W}_{-1}$ 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. □

H TECHNICAL RESULTS

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

Methodology Lemma 24 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 24. *Let $(\mathcal{E}_t)_{t>K}$ 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, $\mathbb{E}_\nu[\tau_\delta] \leq T_\mu(\delta) + \sum_{T>K} \mathbb{P}_\nu(\mathcal{E}_T^\complement)$.*

Proof. Since the random variable τ_δ is positive and $\{\tau_\delta > T\} \subseteq \mathcal{E}_n^\complement$ for all $T \geq T_\mu(\delta)$, we have

$$\mathbb{E}_\nu[\tau_\delta] = \sum_{T\geq 0} \mathbb{P}_\nu(\tau_\delta > T) \leq T_\mu(\delta) + \sum_{T \geq T_\mu(\delta)} \mathbb{P}_\nu(\mathcal{E}_T^\complement)\,,$$

which concludes the proof by adding positive terms. □

Inversion Results Lemma 25 gathers properties on the function $\overline{W}_{-1}$, which is used in the literature to obtain concentration results.

Lemma 25 (Jourdan et al. (2023a))**.** *Let $\overline{W}_{-1}(x) = -W_{-1}(-e^{-x})$ for all $x \geq 1$, where W_{-1} is the negative branch of the Lambert W function. The function $\overline{W}_{-1}$ is increasing on $(1, +\infty)$ and strictly concave on $(1, +\infty)$. In particular, $\overline{W}'_{-1}(x) = \left(1 - \frac{1}{\overline{W}_{-1}(x)}\right)^{-1}$ for all $x > 1$. Then, for all $y \geq 1$ and $x \geq 1$,*

$$\overline{W}_{-1}(y) \leq x \quad \iff \quad y \leq x - \log(x)\,.$$

Moreover, for all $x > 1$,

$$x + \log(x) \leq \overline{W}_{-1}(x) \leq x + \log(x) + \min\left\{\frac{1}{2}, \frac{1}{\sqrt{x}}\right\}\,.$$

Lemma 26 is an inversion result to upper bound a probability which is implicitly defined based on times that are implicitly defined.

Lemma 26. *Let $\overline{W}_{-1}$ defined in Lemma 25. Let $A > 0$, $B > 0$, $C > 0$, $E > 0$, $\alpha > 0$, $\beta > 0$ and*

$$D_{A,B,C,E,\alpha,\beta}(\delta) = \sup\left\{x \mid\ x \leq \frac{A}{\alpha}\overline{W}_{-1}\left(\alpha\left(\log(1/\delta) + C\log(\beta + \log x) + E\right)\right) + B\right\}\,.$$

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 25, direct manipulations yield that

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

□

Lemma 27 is an inversion result to upper bound a time which is implicitly defined.

Lemma 27. *Let $\overline{W}_{-1}$ defined in Lemma 25. 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 $h_1(z, y) = z\overline{W}_{-1}\left(y/z + \log z\right)$

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 25 yields that

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

□

Lemma 28 is an inversion result to asymptotically upper bound a time which is implicitly defined.

Lemma 28. *Let $B > 0$ and $A > 0$*

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

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

Proof. Direct manipulations yields that

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

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

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

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

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

Letting γ goes to 0 yields the result. □

I DETAILS ABOUT 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 mean vectors are displayed in Table 3, and the numerical values for the difficulties are reported in Table 4.

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

Name	K	θ	$\lvert\mathcal{A}_\theta\rvert$	Arms									
				1	2	3	4	5	6	7	8	9	10
Thr1	10	0.5	5	0.9	0.9	0.9	0.65	0.55	0.45	0.35	0.1	0.1	0.1
Thr2	6	0.35	3	0.6	0.5	0.4	0.3	0.2	0.1	–	–	–	–
Thr3	10	0.5	3	0.55	0.55	0.55	0.45	0.45	0.45	0.45	0.45	0.45	0.45
Med1	5	0.5	1	0.537	0.469	0.465	0.36	0.34	–	–	–	–	–
Med2	7	1.2	2	1.8	1.6	1.1	1	0.7	0.6	0.5	–	–	–
IsA1	10	0	5	0.5	0.39	0.28	0.17	0.06	−0.06	−0.17	−0.28	−0.39	−0.50
NoA1	5	0	0	−0.5	−0.62	−0.75	−0.88	−1	–	–	–	–	–
IsA2	7	0	3	1.0	0.5	0.1	−0.1	−0.4	−0.5	−0.6	–	–	–
NoA2	4	0	0	−0.1	−0.4	−0.5	−0.6	–	–	–	–	–	–
RealL	18	0.5	6	0.800	0.791	0.676	0.545	0.538	0.506	0.360	0.329	0.306	0.274
				μ_{11}	μ_{12}	μ_{13}	μ_{14}	μ_{15}	μ_{16}	μ_{17}	μ_{18}		
				0.241	0.203	0.112	0.084	0.081	0.007	−0.018	−0.120		

Table 4: Numerical values of difficulty constants. $H_1(\mu)$ and $H_\theta(\mu)$ as in (1), $\hat{\Delta} := \max_{a\in\mathcal{A}_\theta} \Delta_a + \min_{b\notin\mathcal{A}_\theta} \Delta_b$.

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	$4\lvert\mathcal{A}_\theta\rvert$

I.1.1 Real-life Data: 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, our collaborators 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. Those $K = 18$ protocols involve different modes

of injection, three animal ages where injections are made, and three animal-weight-dependent dose intensities (low, moderate and high doses). The score to quantify the effect of a protocol is 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. This score was computed by comparing 3 replicates of each condition. The considered threshold in the main text was $\theta = 0.5$. When the mean is higher than θ, the treatment is considered significantly positive. We model this application as a Bernoulli instance, *i.e.* observations from arm a are drawn from a Bernoulli distribution with mean $\max(\mu_a, 0)$. 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.

I.1.2 Synthetic Data: Gaussian Instances

Along with the above real-life application described above and in Section 5, 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 3.1.1, 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 5 and 6 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 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 $\mathcal{A}_\theta(\mu) \neq \emptyset$.

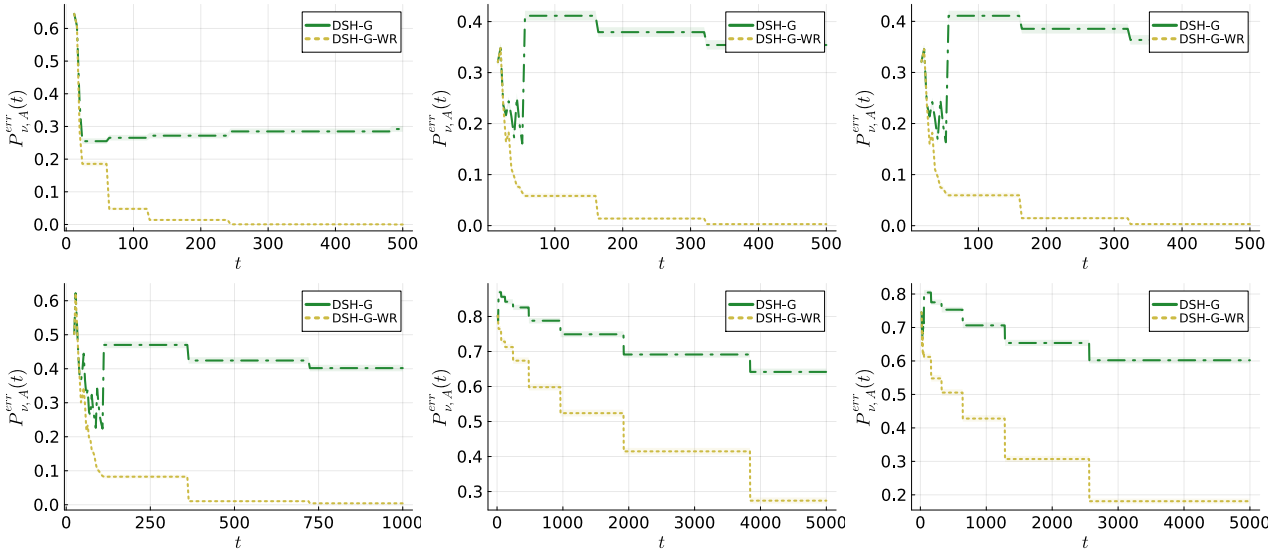

Figure 4: Empirical error on instances (a) NoA1, (b) IsA1, (c) Thr1, (d) RealL, (e) Med1 and (f) Thr3. "-WR" means that each SH instance keeps all its history instead of discarding it.

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 strategy Unif, and different index policies in the prior knowledge-based meta algorithm PKGAI. Those index policies are defined in Section D and recalled below

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

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 7 and 8. We implement the following confidence width, which is theoretically backed by Lemma 22 in Appendix G.2 (for $s = 0$)

$$\beta(t) = \sigma\sqrt{z(T,\delta)/N_a(t)}, \quad \text{where } z(T,\delta) := \overline{W}_{-1}(2\log(K/\delta) + 2\log(2 + \log T) + 2), \text{ using } \delta = 1\%\,. \tag{23}$$

We also consider for algorithms of the PKGAI family the theoretical threshold functions featured in Theorems 7 and 8, *i.e.* relying on problem quantities in practice unavailable at runtime

$$\beta(t) = \sigma\sqrt{q(T,\delta)/N_a(t)}, \quad \text{where } q(T,\delta) := \begin{cases} (T-K)/(4H_1(\mu)) & \text{if } \mathcal{A}_\theta(\mu) = \emptyset \\ (T-K)/(4K\hat{\Delta}^{-2}) & \text{otherwise} \end{cases}, \tag{24}$$

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 (5) is equivalent to their stopping provided that the UCB and LCB are using the same stopping threshold for the bonuses, *i.e.*

$$\max_{a\in\mathcal{A}} W_a^+(t) \geq \sqrt{2c(t,\delta)} \qquad \Longleftrightarrow \qquad \exists a \in \mathcal{A}, \quad \hat{\mu}_a(t) - \sqrt{\frac{2c(t,\delta)}{N_a(t)}} \geq \theta \,,$$

$$\min_{a\in\mathcal{A}} W_a^-(t) \geq \sqrt{2c(t,\delta)} \qquad \Longleftrightarrow \qquad \forall a \in \mathcal{A}, \quad \hat{\mu}_a(t) + \sqrt{\frac{2c(t,\delta)}{N_a(t)}} \leq \theta \,.$$

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 (6)) 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

$$\hat{\mu}_a(t) \pm \sqrt{\frac{2\Lambda_a(t,\delta)}{N_a(t)}} \quad \text{where} \quad \Lambda_a(t,\delta) = \log(4K/\delta) + 2\log N_a(t) \,.$$

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 $c(t,\delta) \sim_{t\to+\infty} 2\log\log t$. Ignoring the δ-dependent terms and the constant, we have a lower bonus as long as $N_a(t) \gtrapprox \log t$. For a fair comparison, we will use the stopping threshold in (6) 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 5, 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

Table 5: Empirical stopping time (± 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.

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	–

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 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 $\max_{a\mathcal{A}} \hat{\mu}_a(t) \geq \theta$, the optimal allocation $w^\star(\hat{\mu}(t))$ to be tracked is a Dirac in $\arg\max_{a\mathcal{A}} \hat{\mu}_a(t)$. Otherwise, using the proof of Lemma 1, the optimal allocation is $w^\star(\hat{\mu}(t))$, which is defined as

$$w^\star(\hat{\mu}(t))_a \propto (\hat{\mu}_a(t) - \theta)^{-2} \,.$$

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 4, the BAEC meta-algorithm is only defined for asymmetric threshold $\theta_U > \theta_L$. Mathematically, it uses the following UCB/LCB indices

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

$$\hat{\mu}_a(t) - \sqrt{\frac{2\Lambda_a^-(t,\delta)}{N_a(t)}} \quad \text{where} \quad \Lambda_a^-(t,\delta) = \log\left(\sqrt{K} N(\delta)/\delta\right) .$$

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).

I.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 (23). 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($\star$), while allowing early stopping as well. Perhaps unsurprisingly, the performance of APGAI are closely related to those of PKGAI(APT_P), as both algorithms share the same sampling rule. In all three instances, although PKGAI has unrealistic assumptions in its theoretical guarantees (Theorems 7 and 8), 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.

Performance on the Real-Life Application We report empirical errors at $T = 200$ in Table 6, at which budget empirical errors for all algorithms seem to converge (see Figure 5).

Table 6: Error across 1,000 runs at $T = 200$.

Algorithm	Error	Conf. intervals	
APGAI	0.001	2.10^{-4}	6.10^{-3}
PKGAI(APT_P)	0.004	2.10^{-3}	0.01
PKGAI(LCB-G)	0.001	2.10^{-4}	6.10^{-3}
PKGAI(UCB)	0.000	0.00	4.10^{-3}
PKGAI(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}

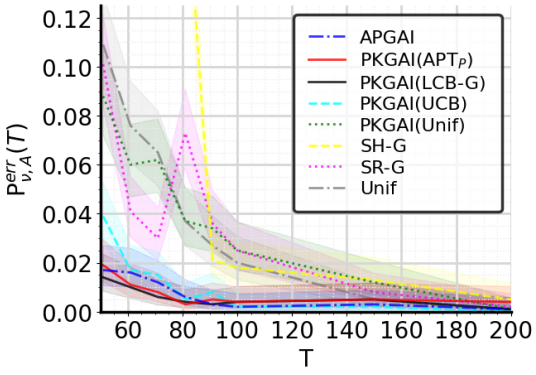

Figure 5: Empirical error on instance RealL.

Performance on Synthetic Datasets ($\mathcal{A}_\theta \neq \emptyset$) We report empirical errors at $T = 700$ in Tables 7 and 8, at which budget empirical errors for all algorithms seem to converge (see Figures 6 and 7). In the figures, the curves of PKGAI(APT_P) and PKGAI(LCB-G) overlap.

Table 7: Error across 1,000 runs at $T = 700$.

Algorithm	Error	Conf. intervals	
APGAI	0.003	1.10^{-3}	9.10^{-3}
PKGAI(APT_P)	0.004	2.10^{-3}	0.01
PKGAI(LCB-G)	0.004	2.10^{-3}	0.01
PKGAI(UCB)	0.000	0.00	4.10^{-3}
PKGAI(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}

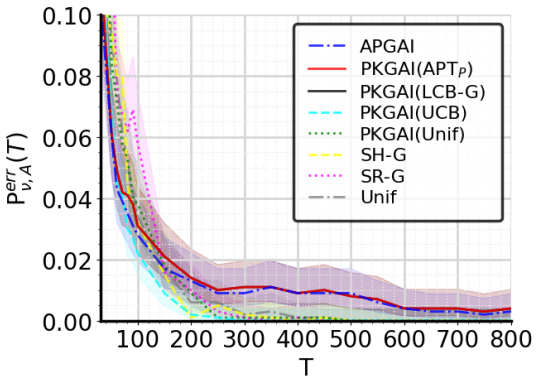

Figure 6: Empirical error on instance IsA1.

Table 8: Error across $1,000$ runs at $T = 700$.

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)	0.000	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}

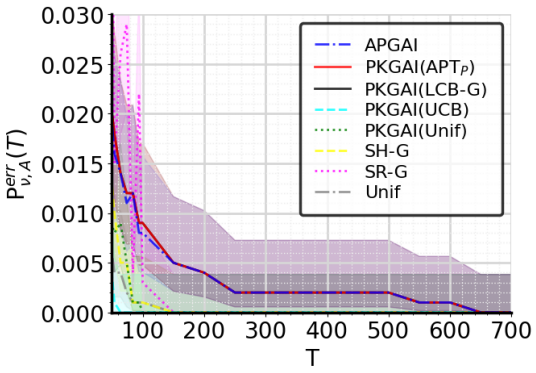

Figure 7: Empirical error on instance IsA2.

Performance on Synthetic Datasets ($\mathcal{A}_\theta = \emptyset$) We report empirical errors at $T = 150$ in Table 9 and $T = 700$ in Table 10, at which budget empirical errors for all algorithms seem to converge (see Figures 8 and 9). In the figures, the curves of PKGAI(APT$_P$) and PKGAI(LCB-G) overlap.

Table 9: Error across $1,000$ runs at $T = 150$.

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

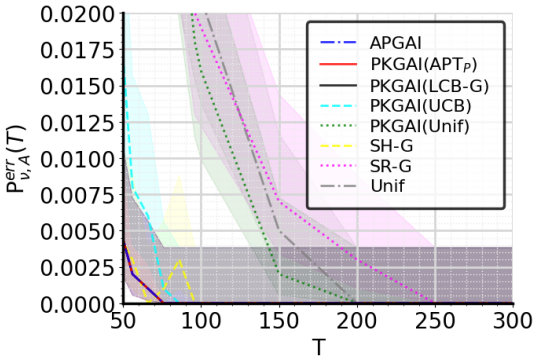

Figure 8: Empirical error on instance NoA1.

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

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

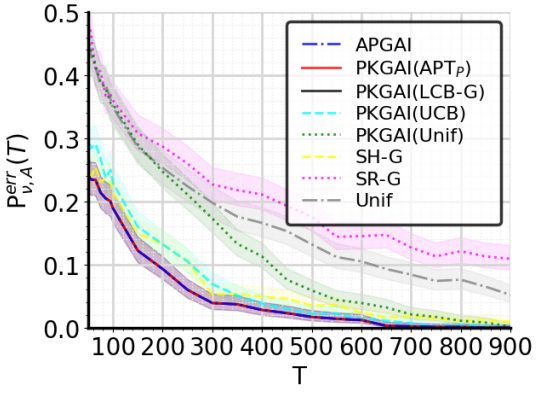

Figure 9: Empirical error on instance NoA2.

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.

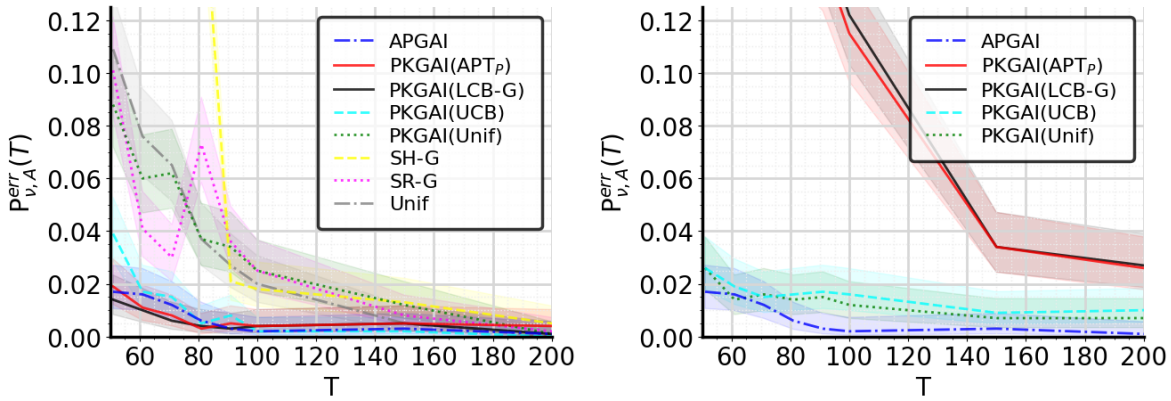

Figure 10: Empirical error on instance RealL. **Left**: with threshold functions from Equation (23). **Right**: with prior knowledge thresholds in Equation (24).

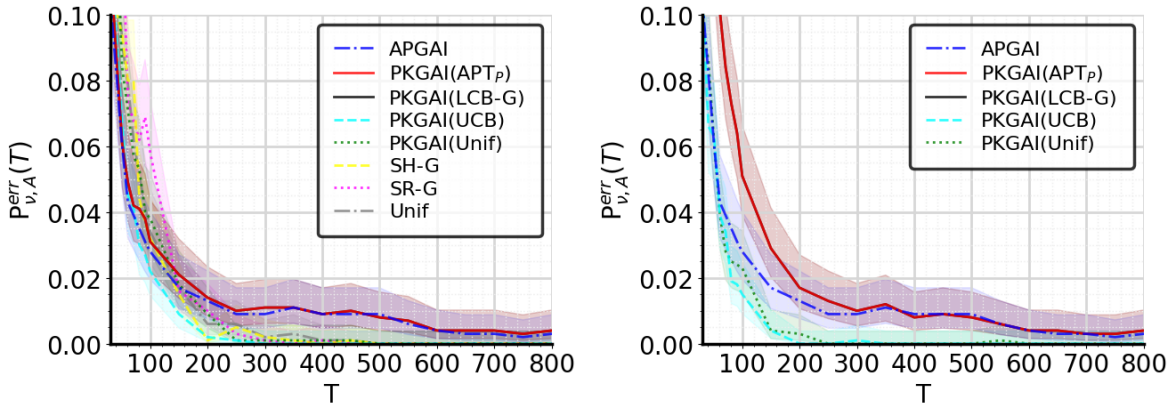

Figure 11: Empirical error on instance IsA1. **Left**: with threshold functions from Equation (23). **Right**: with prior knowledge thresholds in Equation (24).

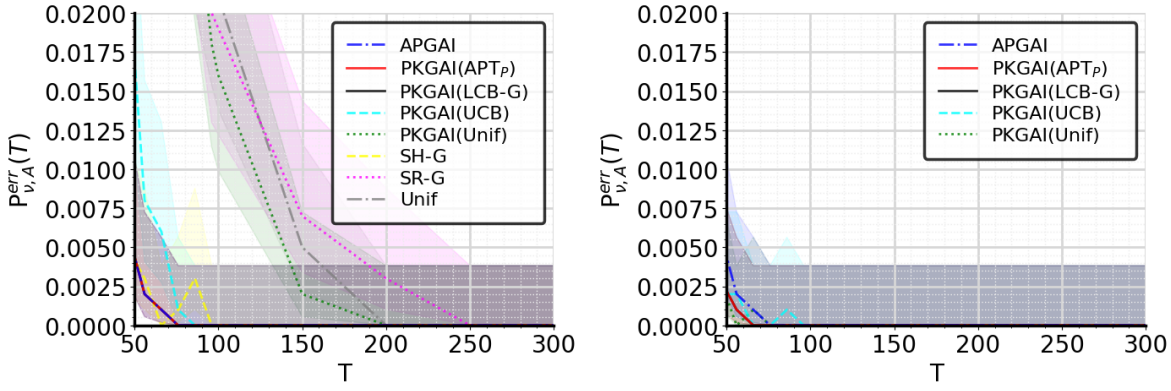

Figure 12: Empirical error on instance NoA1. **Left**: with threshold functions from Equation (23). **Right**: with prior knowledge thresholds in Equation (24).

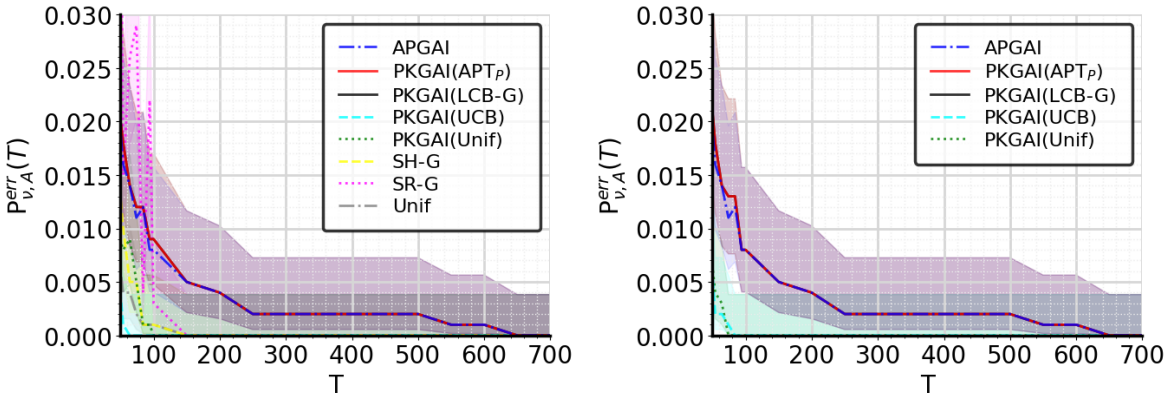

Figure 13: Empirical error on instance IsA2. **Left**: with threshold functions from Equation (23). **Right**: with prior knowledge thresholds in Equation (24).

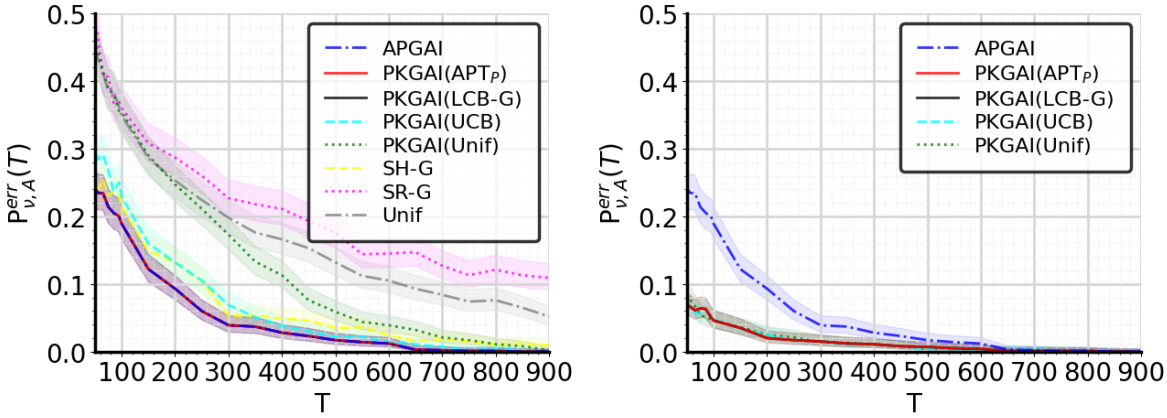

Figure 14: Empirical error on instance NoA2. **Left**: with threshold functions from Equation (23). **Right**: with prior knowledge thresholds in Equation (24).

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 1 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 conly consider the anytime algorithms: APGAI, Unif, DSR-G and DSH-G. As mentioned in Appendix I.2, we 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 1.

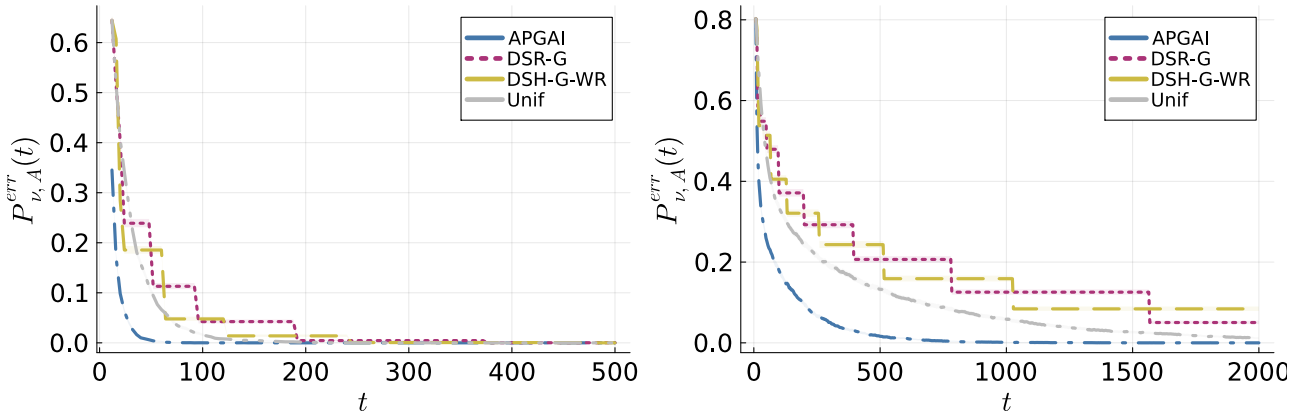

Figure 15: Empirical error on instances (a) NoA1 and (b) NoA2. "-WR" means that each SH instance keeps all its history instead of discarding it.

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 1, 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.

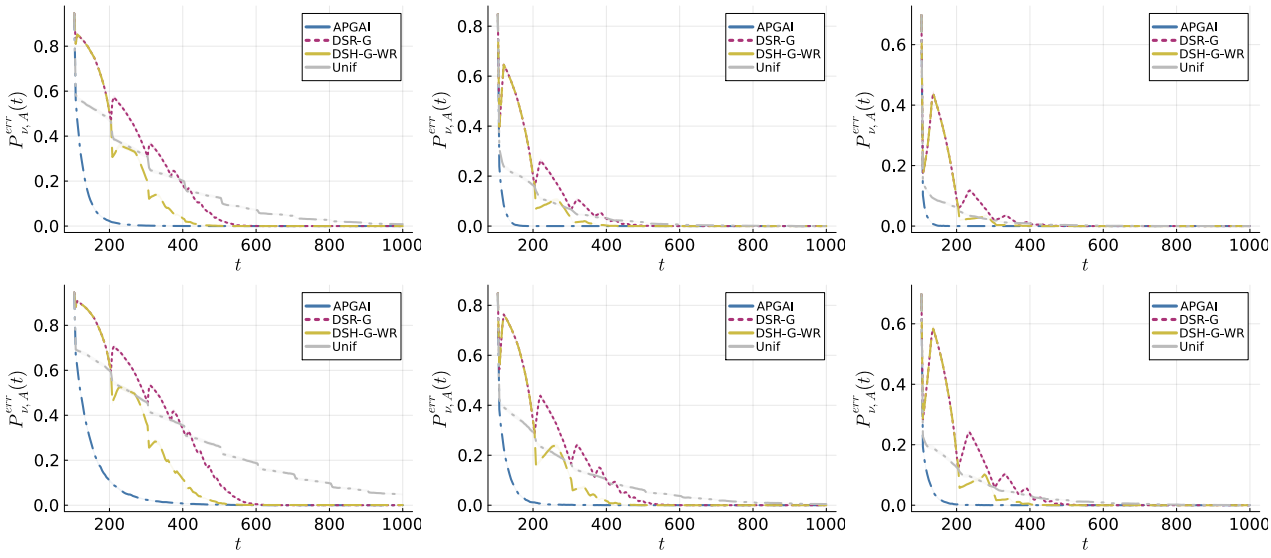

Figure 16: Empirical error for varying number of good arms $|\mathcal{A}_\theta(\mu)| \in \{5, 15, 30\}$ (left to right) among $K = 100$ arms on instances (top) TwoG and (bottom) LinG. "-WR" means that each SH instance keeps all its history instead of discarding it.

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 1 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 inversly 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.

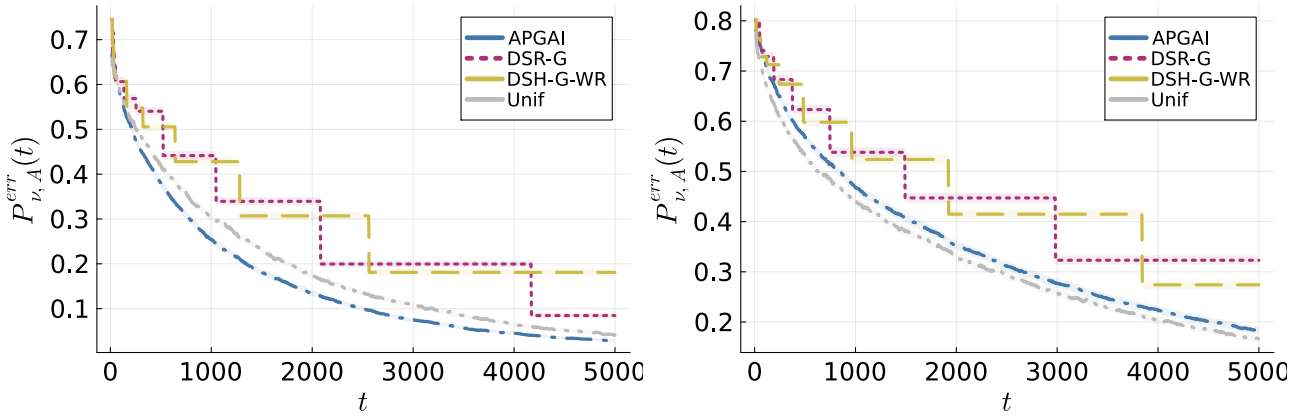

Figure 17: Empirical error on instances (a) Thr3 and (b) Med1. "-WR" means that each SH instance keeps all its history instead of discarding it.

Good Arms With Similar Gaps In light of Table 1, 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 1.

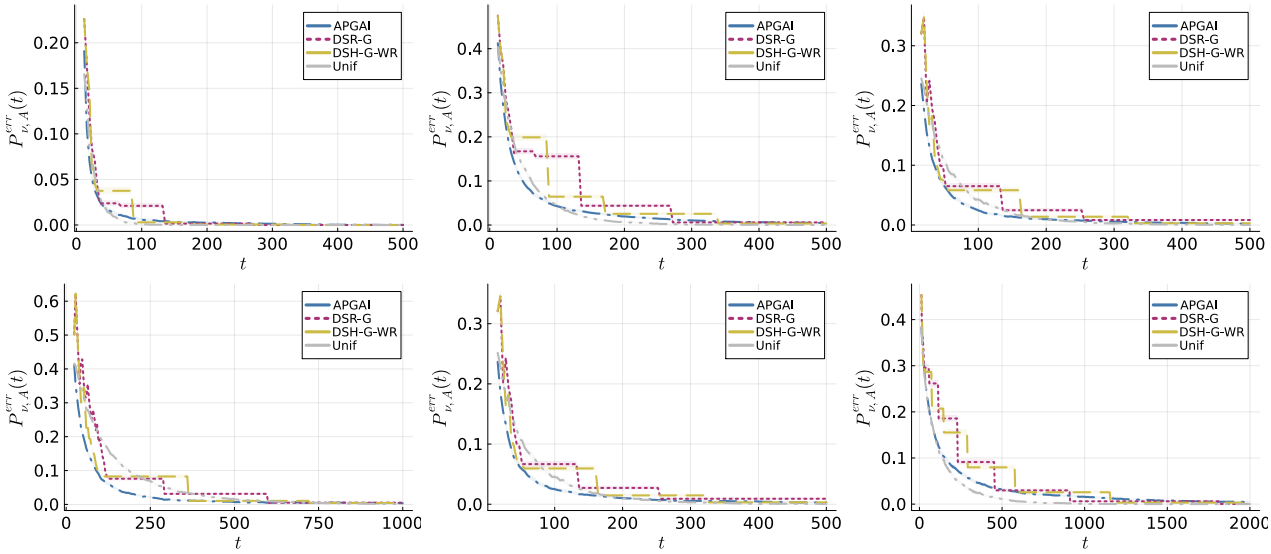

Figure 18: Empirical error on instances (a) IsA2, (b) Med2, (c) IsA1, (d) RealL, (e) Thr1 and (f) Thr2. "-WR" means that each SH instance keeps all its history instead of discarding it.

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 (5) with stopping threshold (6). According to Table 2, 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 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 2. When good arms have dissimilar gaps, APGAI can suffer from large outliers due to the greedyness of it sampling rule. Finally, we shows a simple way to circumvent this limitation by adding forced exloration on top of APGAI.

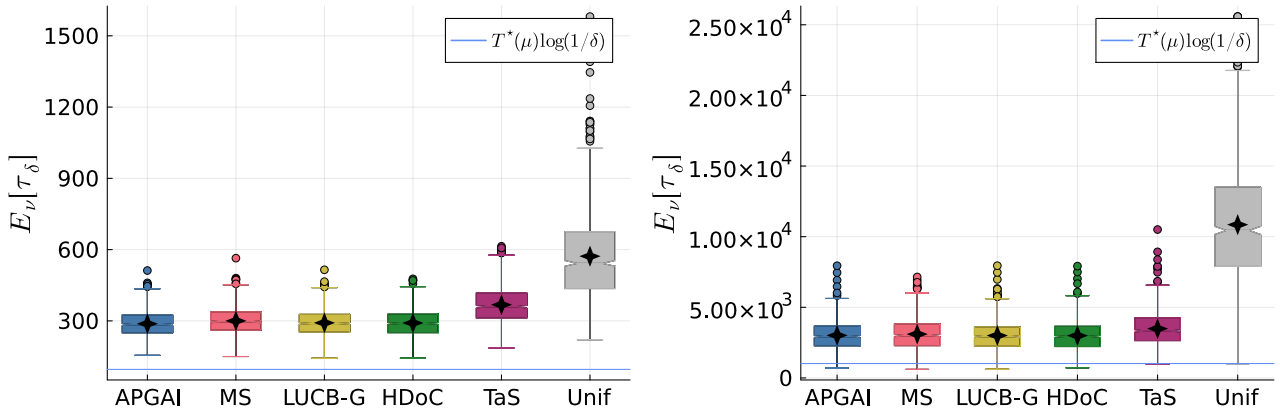

Figure 19: Empirical stopping time ($\delta = 0.01$) on instances (a) NoA1 and (b) NoA2. "MS" is Murphy Sampling, "TaS" is Track-and-Stop and "Unif" is round-robin uniform sampling.

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.

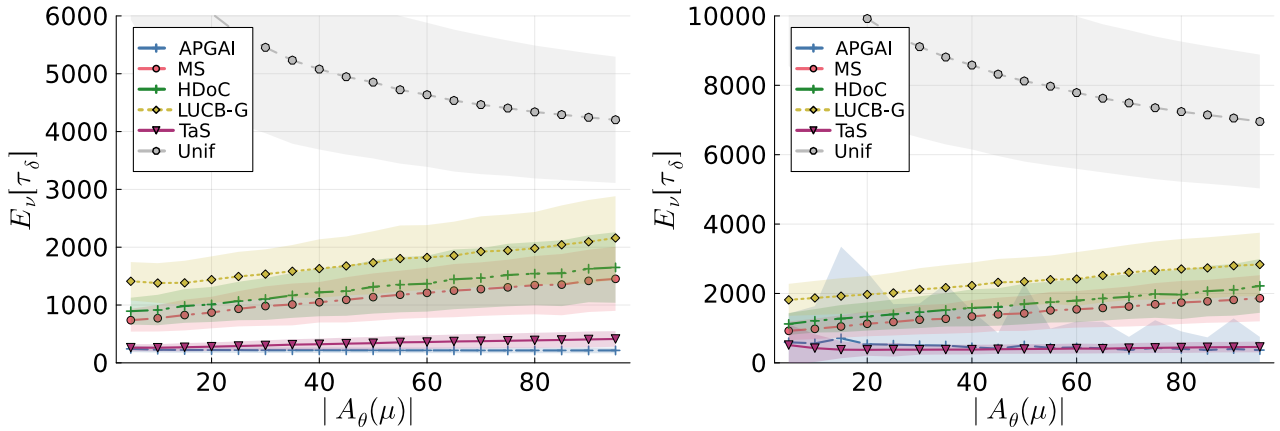

Figure 20: Empirical stopping time ($\delta = 0.01$) for varying number of good arms $|\mathcal{A}_\theta(\mu)| \in \{5k\}_{k\in[19]}$ among $K = 100$ arms on instances (a) TwoG and (b) LinG. "MS" is Murphy Sampling, "TaS" is Track-and-Stop and "Unif" is round-robin uniform sampling.

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 2 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 validate 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.2.1. 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).

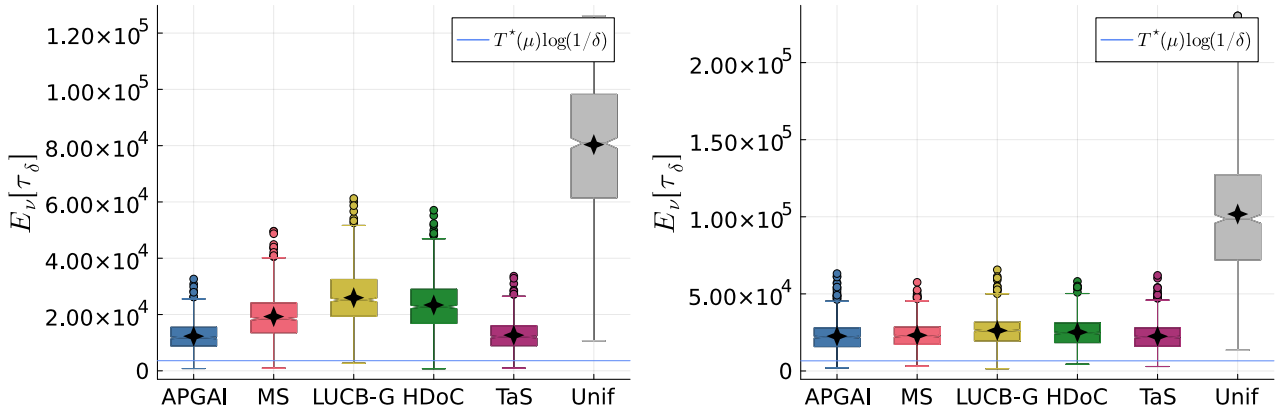

Figure 21: Empirical stopping time ($\delta = 0.01$) on instances (a) Thr3 and (b) Med1. "MS" is Murphy Sampling, "TaS" is Track-and-Stop and "Unif" is round-robin uniform sampling.

Good Arms With Similar Gaps When $\mathcal{A}_\theta(\mu) \neq \emptyset$ and good arms have similar means, Table 2 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.

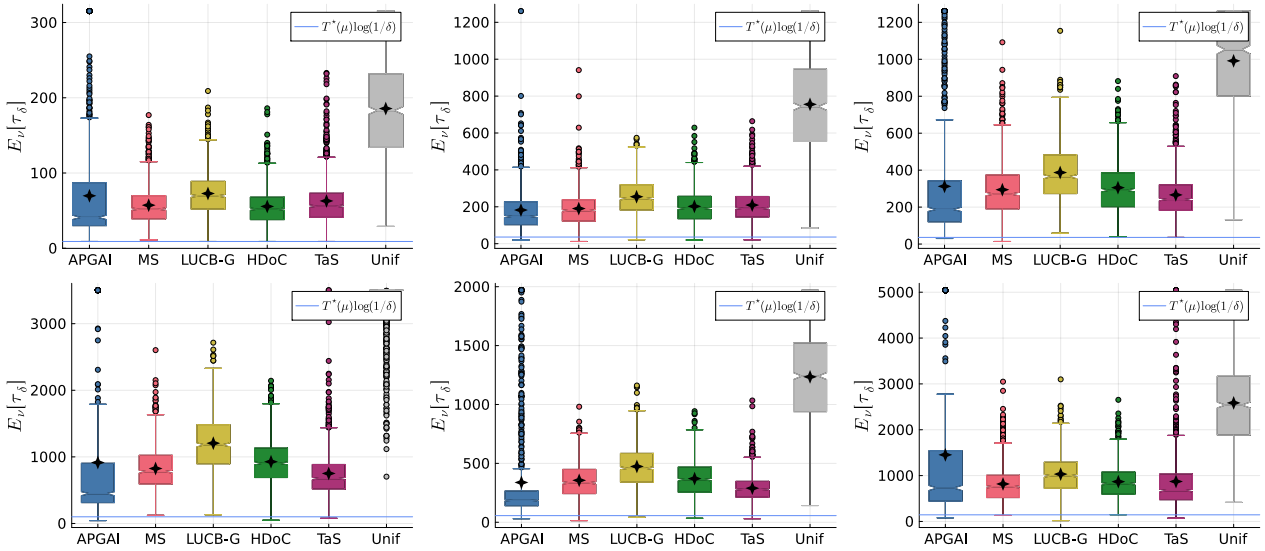

Figure 22: Empirical stopping time ($\delta = 0.01$) on instances (a) IsA2, (b) Med2, (c) IsA1, (d) RealL, (e) Thr1 and (f) Thr2. "MS" is Murphy Sampling, "TaS" is Track-and-Stop and "Unif" is round-robin uniform sampling.

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 2 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 an heavy tail with large outliers.

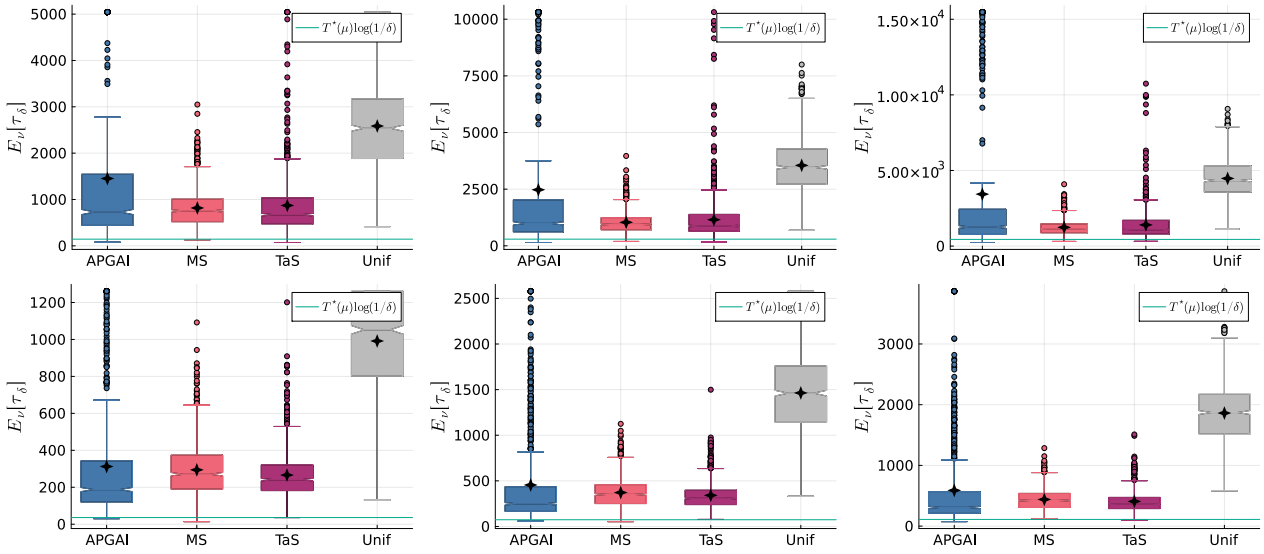

Figure 23: Empirical error for varying confidence level $\delta \in \{10^{-2}, 10^{-4}, 10^{-6}\}$ (left to right) on instances (top) Thr2 and (bottom) IsA1.

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.

Table 11: Empirical stopping time ($\pm$ standard deviation) of APGAI with ("-FE") or without forced exploration.

Ordering	Thr1	Thr2	Thr3	Med1	Med2	IsA1	IsA2	RealL	NoA1	NoA2
APGAI	634	2448	12301	22588	184	544	159	3721	288	3014
	$\pm$2091	$\pm$4269	$\pm$4755	$\pm$9204	$\pm$147	$\pm$1591	$\pm$557	$\pm$12511	$\pm$56	$\pm$1031
APGAI-FE	341	1466	12584	22394	216	341	72	921	287	3022
	$\pm$505	$\pm$2833	$\pm$4818	$\pm$8942	$\pm$106	$\pm$444	$\pm$49	$\pm$1389	$\pm$55	$\pm$1025

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 $\mathcal{U}_t \neq \emptyset$, we pull according to APGAI sampling rule.

Table 11 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.