Best Arm Identification with Resource Constraints

Zitian Li
Department of ISEM,
National University of Singapore

Wang Chi Cheung
Department of ISEM,
National University of Singapore

Abstract

Motivated by the cost heterogeneity in experimentation across different alternatives, we study the Best Arm Identification with Resource Constraints (BAIwRC) problem. The agent aims to identify the best arm under resource constraints, where resources are consumed for each arm pull. We make two novel contributions. We design and analyze the Successive Halving with Resource Rationing algorithm (SH-RR). The SH-RR achieves a near-optimal non-asymptotic rate of convergence in terms of the probability of successively identifying an optimal arm. Interestingly, we identify a difference in convergence rates between the cases of deterministic and stochastic resource consumption.

1 Introduction

Best arm identification (BAI) is a fundamental multi-armed bandit formulation on pure exploration. The over-arching goal is to identify an optimal arm through a sequence of adaptive arm pulls. The efficiency of the underlying strategy is typically quantified by the number of arm pulls. On one hand, the number of arm pulls provides us *statistical insights* into the strategy's performance. On the other hand, the number of arm pulls does not necessarily provide us with the *economic insights* into the total cost of the arm pulls in the scenario of *arm cost heterogeneity*, where the costs of arm pulls differ among different arms.

Arm cost heterogeneity occurs in a variety of applications. For example, consider a retail firm experimenting two marketing campaigns: (a) advertising on an online platform for a day, (b) providing \$5 vouchers to a selected group of recurring customers. The firm wishes to identify the more profitable one out of campaigns (a, b). The executions of (a, b) lead to different costs. A cost-aware retail firm would desire to control the total cost in experimenting these campaign choices, rather than the number of try-outs. Arm cost heterogeneity also occurs in many other operations research applications. Process design decisions in business domains such as supply chain, service operations, pharmaceutical tests typically involve experimenting a collection of alternatives and identifying the best one in terms of profit, social welfare or any desired metric. Compared to the total number of try-outs, it is more natural to keep the total cost in experimentation in check. What is the relationship between the total arm pulling cost and the probability of identifying the best arm?

We make three contributions to shed light on the above question. Firstly, we construct the Best Arm Identification with Resource Constraints (BAIwRC) problem model, which features arm cost heterogeneity in a fundamental pure exploration setting. In the BAIwRC, pulling an arm generates a random reward, while consuming resources. The agent, who is endowed with finite amounts of resources, aims to identify an arm of highest mean reward, subject to the resource constraints. In the marketing example, the resource constraint could be the agent's financial budget for experimenting different campaign, and the agent's goal is to identify the most profitable campaign while not exceeding the budget.

Secondly, we design and analyze the Successive Halving with Resource Rationing (SH-RR) algorithm. SH-RR eliminates sub-optimal arms in phases, and rations an adequate amount of resources to each phase to ensure sufficient exploration in all phases. We derive upper bounds for SH-RR on its Pr(fail BAI), the probability failing to identify a best arm. Our bounds decay exponentially to zero when the amounts of endowed resources increases. In addition, our bounds depends on the mean rewards and consumptions of the arms. Our bounds match the state-of-the-art when specialized to the fixed budget BAI setting. Crucially, we demonstrate the near-optimality of SH-RR by establishing lower bounds on Pr(fail BAI) by any strategy.

Proceedings of the 27$^{\text{th}}$ International Conference on Artificial Intelligence and Statistics (AISTATS) 2024, Valencia, Spain. PMLR: Volume 238.

Thirdly, our results illustrate a fundamental difference between the deterministic and the stochastic consumption settings. Campaign (a) in the marketing example has a deterministic consumption, since the cost for advertisement is determined and fixed (by the ad platform) before (a) is executed. Campaign (b) has a random consumption, since the total cost is proportional to the number of recurring customers who redeem the vouchers, which is random. More precisely, for a BAIwRC instance Q, our results imply that $-\log(\Pr(\text{fail BAI}))$ under an optimal strategy is proportional to $\gamma^{\text{det}}(Q)$ or $\gamma^{\text{sto}}(Q)$, in the cases when Q is a deterministic consumption instance or a stochastic consumption instance respectively. The complexity terms $\gamma^{\text{det}}(Q), \gamma^{\text{sto}}(Q)$ are defined in our forthcoming Section 4, and $\gamma^{\text{det}}, \gamma^{\text{sto}}$ differ in how the resource consumption are encapsulated in their respective settings. Finally, our theoretical findings are corroborated with numerical simulations, which demonstrate the empirical competitiveness of SH-RR compared to existing baselines.

Literature Review. The BAI problem has been actively studied in the past decades, prominently under the two settings of fixed confidence and fixed budget. In the fixed confidence setting, the agent aims to minimize the number of arm pulls, while constraining Pr(fail BAI) to be at most an input confidence parameter. In the fixed budget setting, the agent aims to minimize Pr(fail BAI), subject to an upper bound on the number of arm pulls. The fixed confidence setting is studied in (Even-Dar et al., 2002; Mannor and Tsitsiklis, 2004; Audibert and Bubeck, 2010; Gabillon et al., 2012; Karnin et al., 2013; Jamieson et al., 2014; Kaufmann et al., 2016; Garivier and Kaufmann, 2016), and surveyed in (Jamieson and Nowak, 2014). The fixed budget setting is studied in (Gabillon et al., 2012; Karnin et al., 2013; Kaufmann et al., 2016; Carpentier and Locatelli, 2016). The BAI problem is also studied in the anytime setting (Audibert and Bubeck, 2010; Jun and Nowak, 2016), where a BAI strategy is required to recommend an arm after each arm pull. A related objective to BAI is the minimization of simple regret, which is the expected optimality gap of the identified arm, is studied Bubeck et al. (2009); Audibert and Bubeck (2010); Zhao et al. (2022). Despite the volume of studies on pure exploration problems on multi-armed bandits, existing works focus on analyzing the total number of arm pulls. We provide a new perspective by considering the *total cost of arm pulls*.

BAI problems with constraints have been studied in various works. Wang et al. (2022) consider a BAI objective where the identified arm must satisfy a safety constraint. Hou et al. (2022) consider a BAI objective where the identified arm must have a variance below a pre-specified threshold. Different from these works that impose constraints on the identified arm, we impose constraints on the exploration process. Sui et al. (2015, 2018) study BAI problems in the Gaussian bandit setting, with the constraints that each sampled arm must lie in a latent safety set. Our BAI formulation is different in that we impose cumulative resource consumption constraints across all the arm pulls, rather than constraints on each individual pull. In addition, Sui et al. (2015, 2018) focus on the comparing against the best arm within a certain reachable arm subset, different from our objective of identifying the best arm out of all arms.

Our work is thematically related to the Bandits with Knapsack problem (BwK), where the agent aims to maximize the total reward instead of identifying the best arm under resource constraints. The BwK problem is proposed in Badanidiyuru et al. (2018), and an array of different BwK models have been studied (Agrawal and Devanur, 2014, 2016; Sankararaman and Slivkins, 2018). In the presence of resource constraints, achieving the optimum under the BwK objective does not lead to BAI. For example, in a BwK instance with single resource constraint, it is optimal to pull an arm with the highest mean reward per unit resource consumption, which is generally not an arm with the highest mean reward, when resource consumption amounts differ across arms. Li et al. (2023) propose a BAI problem with BwK setting and shares some settings with us, assuming multiple resources with random consumptions, a finite arm set. But their task is to identify the index set $\mathcal{X}^*$ of all optimal arms in an LP relaxation to a BwK problem. $\mathcal{X}^*$ depends on both the mean reward and mean consumption of the arms. The difference of the target marks a significant departure from our methodologies and results in the field.

Lastly, our work is related to the research on cost-aware Bayesian optimization (BO). In this area, an arm might correspond to a hyper-parameter or a combination of hyper-parameters. A widely adopted idea in the BO community is to set up different acquisition functions to guide the selection of sampling points. One of the most popular choices is Expected Improvement(EI) (Frazier, 2018), without considering the heterogeneous resource consumptions like time or energy. To make EI cost-aware, which is correlated to our setting of a single resource, a common way is to divide it by an approximated cost function $c(x)$ (Snoek et al., 2012; Poloczek et al., 2017; Swersky et al., 2013), calling it Expected Improvement per unit (EIpu). However, Lee et al. (2020) shows this division may encourage the algorithm to explore domains with low consumption, leading to a worse performance when the optimal point consumes more resources. Then Lee et al. (2020) de-

signs the Cost Apportioned BO (CArBo) algorithm, whose acquisition function gradually evolves from EIpu to EI. For better performance, Guinet et al. (2020) develops Contextual EI to achieve Pareto Efficiency. Abdolshah et al. (2019) discusses Pareto Front when there are multiple objective functions. Luong et al. (2021) considers EI and EIpu as two arms in a multi-arm bandits problem, using Thompson Sampling to determine which acquisition function is suitable in each round. These works focus on minimizing the total cost, different from our resource constrained setting. In addition, we allow the resource consumption model to be unknown, random and heterogeneous among arms. In comparison, existing works either assume one unit of resource consumed per unit pulled (Jamieson and Talwalkar, 2016; Li et al., 2020; Bohdal et al., 2022; Zappella et al., 2021), or assume heterogeneity (and deterministic) resource consumption but with the resource consumption (or a good estimate of it) of each arm known Snoek et al. (2012); Ivkin et al. (2021); Lee et al. (2020). Some alternative cost-aware BO require the multi-fidelity or other grey-box assumption (Forrester et al., 2007; Kandasamy et al., 2017; Wu et al., 2020; Foumani et al., 2023; Belakaria et al., 2023), which are not consistent with our settings.

Notation. For an integer $K > 0$, denote $[K] = \{1, \ldots, K\}$. For $d \in [0, 1]$, we denote $\text{Bern}(d)$ as the Bernoulli distribution with mean d.

2 Model

An instance of Best Arm Identification with Resource Constraints (BAIwRC) is specified by the triple $Q = ([K], C, \nu = \{\nu_k\}_{k\in[K]})$. The set $[K]$ represents the collection of K arms. There are L types of different resources. The quantity $C = (C_\ell)_{\ell=1}^L \in \mathbb{R}_{>0}^L$ is a vector, and C_ℓ is the amount of type ℓ resource units available to the agent. For each arm $k \in [K]$, ν_k is the probability distribution on the $(L+1)$-variate outcome $(R_k; D_{1,k}, \ldots, D_{L,k})$, which is received by the agent when s/he pulls arm k once. By pulling arm k once, the agent earns a random amount R_k of reward, and consumes a random amount $D_{\ell,k}$ of the type-ℓ resource, for each $\ell \in \{1, \ldots, L\}$. We allow $R_k, D_{1,k}, \ldots, D_{L,k}$ to be arbitrarily correlated. We assume that R_k is a 1-sub-Gaussian random variable, and $D_{\ell,k} \in [0, 1]$ almost surely for every $k \in [K], \ell \in [L]$.

We denote the mean reward $\mathbb{E}[R_k] = r_k$ for each $k \in [K]$, and denote the mean consumption $\mathbb{E}[D_{\ell,k}] = d_{\ell,k}$ for each $\ell \in [L], k \in [K]$. Similar to existing works on BAI, we assume that there is a unique arm with the highest mean reward, and without loss of generality we assume that $r_1 > r_2 \geq \ldots, \geq r_K$. We call arm 1 the optimal arm. We emphasize that the mean consumption amounts $\{d_{\ell,k}\}_{k=1}^K$ on any resource ℓ need not be ordered in the same way as the mean rewards. We assume that $d_{\ell,k} > 0$ for all $k \in [K], \ell \in [L]$. Crucially, the quantities $r_k, d_{\ell,k}, \nu_k$ for any k, ℓ are not known to the agent.

Dynamics. The agent pulls arms sequentially in time steps $t = 1, 2, \ldots$, according to a non-anticipatory policy π. We denote the arm pulled at time t as $A(t) \in [K]$, and the corresponding outcome as $O(t) = (R(t); D_1(t), \ldots, D_L(t)) \sim \nu_{A(t)}$. A non-anticipatory policy π is represented by the sequence $\{\pi_t\}_{t=1}^\infty$, where π_t is a function that outputs the arm $A(t)$ by inputting the information collected in time $1, \ldots, t-1$. More precisely, we have $A(t) = \pi_t(H(t-1))$, where $H(t-1) = \{O(s)\}_{s=1}^{t-1}$. The agent stops pulling arms at the end of time step τ, where τ is a finite stopping time[1] with respect to the filtration $\{\sigma(H(t))\}_{t=1}^\infty$. Upon stopping, the agent identifies arm $\psi \in [K]$ to be the best arm, using the information $H(\tau)$. Altogether, the agent's strategy is represented as (π, τ, ψ).

Objective. The agent aims to choose a strategy (π, τ, ψ) to maximize $\Pr(\psi = 1)$, the probability of BAI, subject to the resource constraint that $\sum_{t=1}^\tau D_\ell(t) \leq C_\ell$ holds for all $\ell \in [L]$ with certainty. We distinguish between two problem model settings, namely the **stochastic consumption setting** and the **deterministic consumption setting**. The former is precisely as described above, where we allow $\{D_{\ell,k}\}_{\ell,k}$ to be arbitrary random variables bounded between 0 and 1. The latter is a special case where $\Pr(D_{\ell,k} = d_{\ell,k}) = 1$ for all $\ell \in [L], k \in [K]$, meaning that all the resource consumption amounts are deterministic. In the special case when $L = 1$ and $\Pr(D_{1,k} = 1 \text{ for all } k \in [K]) = 1$, the deterministic consumption setting specializes to the fixed budget BAI problem.

We focus on bounding the failure probability $\Pr(\text{fail BAI}) = \Pr(\psi \neq 1)$ in terms of the underlying parameters in Q. The forthcoming bounds are in the form of $\exp(-\gamma(Q))$, where $\gamma(Q) > 0$ can be understood as a complexity term that encodes the difficulty of the underlying BAIwRC instance Q. To illustrate, in the case of $L = 1$, we aim to bound $\Pr(\psi \neq 1)$ in terms of $\exp(-C_1/H)$, where $H > 0$ depends on the latent mean rewards and resource consumption amounts. In the subsequent sections, we establish upper bounds on $\Pr(\psi \neq 1)$ for our proposed strategy SH-RR, as well as lower bounds on $\Pr(\psi \neq 1)$ for any feasible strategy. We demonstrate that the complexity term $\gamma(Q)$ crucially on if Q has deterministic or stochastic consumption.

[1] For any t, the event $\{\tau = t\}$ is $\sigma(H(t))$-measurable, and $\Pr(\tau = \infty) = 0$

Algorithm 1 Sequential Halving with Resource Rationing (SH-RR)

1: **Input:** Total budget C, arm set $[K]$.
2: **Initialize** $\tilde{S}^{(0)} = [K]$, $t = 1$.
3: **Initialize** $\mathsf{Ration}_\ell^{(0)} = \frac{C_\ell}{\lceil \log_2 K \rceil}$ for each $\ell \in [L]$.
4: **for** $q = 0$ **to** $\lceil \log_2 K \rceil - 1$ **do**
5: **Initialize** $I_\ell^{(q)} = 0 \ \forall \ell \in [L]$, $H^{(q)} = J^{(q)} = \emptyset$.
6: **while** $I_\ell^{(q)} \leq \mathsf{Ration}_\ell^{(q)} - 1$ for all $\ell \in [L]$ **do**
7: Identify the arm index $a(t) \in \{1, \ldots, |\tilde{S}^{(q)}|\}$ such that $a(t) \equiv t \bmod |\tilde{S}^{(q)}|$.
8: Pull arm $A(t) = k_{a(t)}^{(q)} \in \tilde{S}^{(q)}$.
9: Observe the outcome $O(t) \sim \nu_{A(t)}$.
10: Update $I_\ell^{(q)} \leftarrow I_\ell^{(q)} + D_\ell(t)$ for each $\ell \in [L]$.
11: Update $H^{(q)} \leftarrow H^{(q)} \cup \{(A(t), O(t))\}$.
12: Update $J^{(q)} \leftarrow J^{(q)} \cup \{t\}$.
13: Update $t \leftarrow t + 1$.
14: **end while**
15: Use $\cup_{m=0}^{q} H^{(m)}$ to compute empirical means $\{\hat{r}_k^{(q)}\}_{k \in \tilde{S}^{(q)}}$, see (1).
16: Set $\tilde{S}^{(q+1)}$ be the set of top $\lceil |\tilde{S}^{(q)}|/2 \rceil$ arms with highest empirical mean.
17: Set $\mathsf{Ration}_\ell^{(q+1)} = \frac{C_\ell}{\lceil \log_2 K \rceil} + (\mathsf{Ration}_\ell^{(q)} - I_\ell^{(q)})$.
18: **end for**
19: Output the arm in $\tilde{S}^{(\lceil \log_2 K \rceil)}$.

3 The SH-RR Algorithm

Our proposed algorithm, dubbed Sequential Halving with Resource Rationing (SH-RR), is displayed in Algorithm 1. SH-RR iterates in phases $q \in \{0, \ldots, \lceil \log_2 K \rceil\}$. Phase q starts with a *surviving arm set* $\tilde{S}^{(q)} \subseteq [K]$. After the arm pulling in phase q, a subset of arms in $\tilde{S}^{(q)}$ is eliminated, giving rise to $\tilde{S}^{(q+1)}$. After the final phase, the surviving arm set $\tilde{S}^{(\lceil \log_2 K \rceil)}$ is a singleton set, and its only constituent arm is recommended as the best arm. We denote $\tilde{S}^{(q)} = \{k_1^{(q)}, \ldots, k_{|\tilde{S}^{(q)}|}^{(q)}\}$. In each phase q, the agent pulls arms in $\tilde{S}^{(q)}$ in a round-robin fashion. At a time step t, the agent first identifies (see Line 7) the arm index $a(t) \in \{1, \ldots, |\tilde{S}^{(q)}|\}$ and pulls the arm $k_{a(t)}^{(q)} \in \tilde{S}^{(q)}$. The round robin schedule ensures that the arms in $\tilde{S}^{(q)}$ are uniformly explored. SH-RR keeps track of the amount of type-ℓ resource consumption via $I_\ell^{(q)}$. The **while** condition (see Line 6) ensures that at the end of phase q, the total amount $I_\ell^{(q)}$ of type-ℓ resource consumption during phase q lies in $(\mathsf{Ration}_\ell^{(q)} - 1, \mathsf{Ration}_\ell^{(q)}]$ for each $\ell \in [L]$. The lower bound ensures sufficient exploration on $\tilde{S}^{(q)}$, while the upper bound ensures the feasibility of SH-RR to the resource constraints, as formalized in the following claim:

Claim 1. *With certainty, SH-RR consumes at most C_ℓ units of resource ℓ, for each $\ell \in [L]$.*

Proof of Claim 1 is in Appendix B.1. Crucially, SH-RR maintains the observation history $H^{(q)}$ that is used to determined the arms to be eliminated from $\tilde{S}^{(q)}$. After exiting the **while** loop, the agent computes (in Line 15) the empirical mean

$$\hat{r}_k^{(q)} = \frac{\sum_{m=0}^{q} \sum_{t \in J^{(m)}} R(t) \cdot \mathbb{1}(A(t) = k)}{\max\{\sum_{m=0}^{q} \sum_{t \in J^{(m)}} \mathbb{1}(A(t) = k), 1\}} \quad (1)$$

for each $k \in \tilde{S}^{(q)}$. The surviving arm set $\tilde{S}^{(q+1)}$ in the next phase of phase $q+1$ consists of the $\lceil |\tilde{S}^{(q)}|/2 \rceil$ arms in $\tilde{S}^{(q)}$ with the highest empirical means, see Line 16. The amounts of resources rationed for phase $q+1$ is in Line 17.

4 Performance Guarantees of SH-RR

We start with the **deterministic consumption setting**, and some necessary notation. For each $k \in \{2, \ldots, K\}$, we denote $\Delta_k = r_1 - r_k \in [0, 1]$. We also denote $\Delta_1 = r_1 - r_2 = \Delta_2$. Consequently, we have $\Delta_1 = \Delta_2 \leq \Delta_3 \leq \ldots \leq \Delta_K$. For each resource type $\ell \in [L]$, we denote $d_{\ell,(1)}, d_{\ell,(2)}, \ldots, d_{\ell,(K)}$ as a permutation of $d_{\ell,1}, d_{\ell,2}, \ldots, d_{\ell,K}$ such that $d_{\ell,(1)} \geq d_{\ell,(2)} \geq \ldots \geq d_{\ell,(K)}$. We define

$$H_{2,\ell}^{\mathrm{det}}(Q) = \max_{k \in \{2,\ldots,K\}} \left\{ \frac{\sum_{j=1}^{k} d_{\ell,(j)}}{\Delta_k^2} \right\}, \quad (2)$$

which encodes the difficulty of the instance. When we specialize to the fixed budget BAI problem by setting $L = 1$ and $\Pr(D_{1,k} = 1) = 1$ for all $k \in [K]$, the quantity $H_{2,1}^{\mathrm{det}}(Q)$ is equal to a quantity H_2, which a complexity term defined for the fixed budget BAI setting (Audibert and Bubeck, 2010; Karnin et al., 2013). Our first main result is an upper bound on $\Pr(\text{fail BAI}) = \Pr(\psi \neq 1)$ for our proposed SH-RR in the deterministic consumption setting.

Theorem 2. *Consider a BAIwRC instance Q in the deterministic consumption setting. SH-RR (Algorithm 1) has BAI failure probability $\Pr(\psi \neq 1)$ at most*

$$\lceil \log_2 K \rceil K \exp\left(-\frac{1}{4\lceil \log_2 K \rceil} \cdot \gamma^{det}(Q) \right) \quad (3)$$

where $\gamma^{det}(Q) = \min_{\ell \in [L]} \{C_\ell / H_{2,\ell}^{det}(Q)\}$, and $H_{2,\ell}^{det}(Q)$ is defined in (2).

Theorem 2 is proved in Appendix B.2. The performance guarantee of SH-RR improves when the complexity term $\gamma^{\mathrm{det}}(Q)$ increases. We provide intuitions in the special case of $L = 1$, so $\ell = 1$ always. The upper bound (3) decreases when C_1 increases, since more resource units allows more experimentation, hence a

lower failure probability. The upper bound (3) increases when $H_{2,1}^{\det}(Q)$ increases. Indeed, when $d_{\ell,(k)}$ increases, the agent consumes more resource units when pulling the arm with the k-th highest consumption on resource ℓ, which leads to less arm pulls under a fixed budget. In addition, when Δ_k decreases, more arm pulls are needed to distinguish between arms $1, k$, leading to a higher Pr(fail BAI). Observe that $H_{2,1}^{\det}(Q)$ involves the mean consumption $d_{1,(1)}, \ldots, d_{1,(K)}$ in a non-increasing order, providing a worst-case hardness measure over all permutations of the arms. One could wonder if the definition of $H_{2,\ell}^{\det}$ can be refined in the non-ordered way, i.e. $\max_{k\in\{2,\ldots,K\}}\{\sum_{j=1}^{k} d_{\ell,j}/\Delta_k^2\}$. Our analysis in appendix B.7 shows such a refinement is unachievable.

The insights above carry over to the case of general L. The complexity term $\gamma^{\det}(Q)$ involves a minimum over all resource types $[L]$, meaning that the failure probability depends on the bottleneck resource type(s). Finally, when we specialize to the fixed budget BAI setting, the upper bound (3) matches (up to a multiplicative absolute constant) the BAI failure probability upper bound of the Successive Halving algorithm Karnin et al. (2013).

At first sight, it seems Theorem 2 should hold in the **stochastic consumption setting**. Indeed, if an arm's pull consumes Bern(d) units of a resource (Let's assume $L = 1$ for the discussion), then N arm pulls consume at most $Nd + 2\sqrt{Nd\log(1/\delta)}$ units with probability $\geq 1-\delta$, for any $\delta \in (0,1)$. With a large enough N, for example when C_1/d is sufficiently large, we expect $Nd \geq 2\sqrt{Nd\log(1/\delta)}$. That is, with probability $\geq 1-\delta$ the realized consumption is at most twice of Nd, the consumption with N pulls where each pull consumes d units with certainty instead of Bern(d). It then transpires that (3) should hold, modulo a different constant (from 1/4) in the exponent.

Despite the intuition, a simulation on two instances $Q^{\det}, Q^{\text{sto}}$ suggests the otherwise. Instances $Q^{\det}, Q^{\text{sto}}$ both have with $K = 2, L = 1, C = 2$. Instances $Q^{\det}, Q^{\text{sto}}$ share the same Bernoulli rewards with means $r_1 = 0.5, r_2 = 0.4$ and the same mean resource consumption $d_1 = d_2 = d$, where d varies. In $Q^{\det}$, an arm pull consumes d units with certainty, while in Q^{sto} it consumes Bern(d) per pull. We plot $\log(\Pr(\psi \neq 1))$ under SH-RR against the varying d in Figure 1, while other model parameters are fixed. Figure 1 shows that the $\Pr(\psi \neq 1)$ for $Q^{\det}$ is always less than that for Q^{sto}. In addition, $\Pr(\psi \neq 1)$'s for $Q^{\det}, Q^{\text{sto}}$ diverge when d shrinks, which is in contrary to the previous mentioned intuition. The left panel shows that the plotted $\log(\Pr(\psi \neq 1))$ does not decreases linearly as $1/d$ grows, which implies that the bound in (3) does not hold for Q^{sto} when d is sufficiently small.

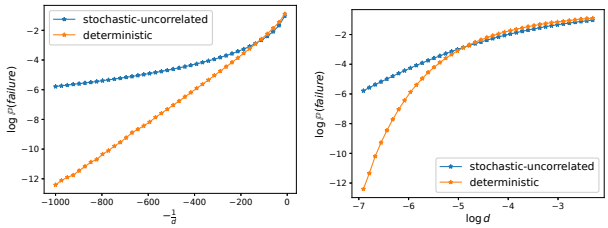

Figure 1: Convergence rates of $\log(\Pr(\psi \neq 1))$, with 10^7 repeated trials

It turns out the **stochastic consumption setting** needs a different characterization. For an instance Q with stochastic consumption, define

$$H_{2,\ell}^{\text{sto}}(Q) = \max_{k\in\{2,\ldots,K\}} \left\{ \frac{\sum_{j=1}^{k} f(d_{\ell,(j)})}{\Delta_k^2} \right\}, \qquad (4)$$

where the function $f : (0,1] \to (0, e^2]$ is defined as

$$f(d) = \begin{cases} e^2 \cdot d & \text{if } d \in [e^{-2}, 1], \\ 2(-\log d)^{-1} & \text{if } d \in (0, e^{-2}). \end{cases} \qquad (5)$$

The function f is continuous and increasing in $(0,1]$.

Theorem 3. *Consider a BAIwRC instance Q in the stochastic consumption setting. The SH-RR algorithm has BAI failure probability $\Pr(\psi \neq 1)$ at most*

$$7LK(\log_2 K)\exp\left(-\frac{1}{8\lceil\log_2 K\rceil}\cdot\gamma^{sto}(Q)\right), \qquad (6)$$

where $\gamma^{sto}(Q) = \min_{\ell\in[L]}\{C_\ell/H_{2,\ell}^{sto}(Q)\}$, and $H_{2,\ell}^{sto}(Q)$ is defined in (4).

Theorem 3 is proved in Appendix B.3. The upper bound in (6) has a similar form to (3), except that $\gamma^{\det}(Q)$ is replaced with $\gamma^{\text{sto}}(Q)$. Crucially, the expected consumption $d_{\ell,(j)}$ in $H_{2,\ell}^{\det}(Q)$ is replaced with *effective consumption* $f(d_{\ell,(j)})$ in $H_{2,\ell}^{\text{sto}}(Q)$. For an arm $k \in [K]$, the effective consumption $f(d_{\ell,k})$ encapsulates the magnitude of the random consumption through the mean $d_{\ell,k}$. The non-linearity of f encapsulates the impact of randomness in resource consumption. The function $f(d)$ is increasing in d, meaning that a higher mean consumption leads to a higher level of utilization on a resource. Note that $f(d) > d$, and $\lim_{d\to 0} f(d)/d = \infty$, which bears the following implications. Consider a stochastic consumption instance and a deterministic consumption instance that have the same $\{C_\ell\}_{\ell\in[L]}, \{r_k\}_{k\in[K]}, \{d_{\ell,k}\}_{k\in[K],\ell\in[L]}$. The upper bound (6) on $\Pr(\psi \neq 1)$ for the stochastic instance converges at a strictly slower rate to zero than the upper bound (3) for the deterministic instance. In addition, when all of $\{d_{\ell,k}\}_{k\in[K],\ell\in[L]}$ tend to zero, the ratio between the two upper bounds grows arbitrarily. These

implications are depicted by the diverging curves in Figure 1, and the upper bound (6) is in fact consistent with the curve for Q^{sto} in Figure 1.

5 Lower Bounds to $\Pr(\psi \neq 1)$

While the deterioration in the upper bound to Pr(fail BAI) could appear to be a limitation to SH-RR, we establish lower bound results for BAIwRC, which demonstrate the near optimality of SH-RR. In what follows, we consider instances where arm 1 needs not be optimal, different from our development of SH-RR. Our lower bound results involve constructing K instances $\{Q^{(i)}\}_{i=1}^K$, where the resource consumption models are carefully crafted to (nearly) match the upper bound results of SH-RR.

Deterministic consumption setting. Each instance involves the set $[K]$ of K arms and L types of resources $[L]$. Let $\{r_k\}_{k=1}^K$ be any sequence such that (a) $1/2 = r_1 \geq r_2 \geq \cdots \geq r_K \geq 1/4$, and let $\{\{d_{\ell,(k)}\}_{k=1}^K\}_{\ell\in[L]}$ be a fixed but arbitrary collection of L sequences such that (b) $d_{\ell,(1)} \geq d_{\ell,(2)} \geq \cdots \geq d_{\ell,(K)}$ for all $\ell \in [L]$, and $d_{\ell,(k)} \in (0,1]$ for all $\ell \in [L], k \in [K]$.

In instance $Q^{(i)}$, pulling arm $k \in [K]$ generates a random reward $R_k \sim \text{Bern}(r_k^{(i)})$, where

$$r_k^{(i)} = \begin{cases} r_k & \text{if } k \neq i, \\ 1 - r_k & \text{if } k = i, \end{cases}.$$

Pulling arm $k \in [K]$ consumes

$$d_{\ell,k} = \begin{cases} d_{\ell,(2)} & \text{if } k = 1, \\ d_{\ell,(1)} & \text{if } k = 2, \\ d_{\ell,(k)} & \text{if } k \in \{3,\ldots,K\} \end{cases} \tag{7}$$

units of resource ℓ for each $\ell \in [L]$ with certainty.

In instance $Q^{(i)}$, arm i is the uniquely optimal arm. All instances $Q^{(1)},\ldots,Q^{(K)}$ have identical resource consumption model, since the consumption amounts (7) do not depend on the instance index i. This ensures that no strategy can extract information about the reward from an arm's consumption. In addition, the consumption amounts in (7) are designed to ensure that (a) instance $Q^{(1)}$ is the hardest among $\{Q^{(i)}\}_{i=K}^K$ in the sense that $H_{2,\ell}^{\text{det}}(Q^{(1)}) = \max_{i\in[K]} H_{2,\ell}^{\text{det}}(Q^{(i)})$ for every $\ell \in [L]$, (b) The ordering $d_{\ell,1} \leq d_{\ell,2} \geq d_{\ell,3} \geq \ldots, \geq d_{\ell,K}$ makes $Q^{(1)}$ a hard instance in the sense that it cost the most to distinguish the second best arm (arm 2) from the best arm. More generally, for each resource ℓ, the consumption amounts are designed such that a sub-optimal arm is more costly to pull when its mean reward is closer to the optimum. Our construction leads to the following lower bound on the performance of any strategy:

Theorem 4. *Consider deterministic consumption instances $Q^{(1)},\ldots,Q^{(K)}$ constructed as above, with $\{r_k\}_{k\in[K]}, \{d_{\ell,(k)}\}_{\ell,k}$ being fixed but arbitrary sequences of parameters that satisfy properties (a, b) respectively. When $C_1,\ldots,C_L$ are sufficiently large, for any strategy there exists an instance $Q^{(i)}$ (where $i \in [K]$) such that*

$$\Pr_i(\psi \neq i) \geq \frac{1}{6}\exp\left(-122\cdot\gamma^{det}(Q^{(i)})\right),$$

where $\Pr_i(\cdot)$ is the probability measure over the trajectory $\{(A(t),O(t))\}_{t=1}^\tau$ under which the arms are chosen according to the strategy and the outcomes are modeled by $Q^{(i)}$, and $\gamma^{det}(Q)$ is as defined in Theorem 2.

Theorem 4 is proved in Appendix B.5. Theorems 2, 4 demonstrate the **near-optimality of SH-RR**, and the fundamental importance of the quantity $\gamma^{\text{det}}(Q)$ for the BAIwRC problem with deterministic consumption. Indeed, both the BAI failure probability upper bound (of SH-RR) in Theorem 2 and the BAI failure probability lower bound in Theorem 4 decay to zero exponentially, with rates linear in $\gamma^{\text{det}}(Q)$. More precisely, the bounds in Theorems 2, 4 imply

$$\sup_{\text{strategy}} \inf_{\substack{\text{det inst } Q: \\ \gamma^{\text{det}}(Q)\geq\kappa^{\text{det}}}} \left\{\frac{-\log\left(\Pr(\text{fail BAI})\right)}{\gamma^{\text{det}}(Q)}\right\} \in \tag{8}$$

$$\left[\frac{1}{16\log_2 K}, 123\right], \tag{9}$$

where $\kappa^{\text{det}} = 32(\log(2K))^2$. The supremum is over all feasible strategy, and the infimum is over all instances Q where $\gamma^{\text{det}}(Q) \geq \kappa^{\text{det}}$, i.e. instances with sufficiently large capacities $C_1,\ldots,C_L$. In the special case of fixed-budget BAI, (Audibert and Bubeck, 2010; Carpentier and Locatelli, 2016) imply that the right hand side in (9) can be $[\frac{1}{8\log_2 K}, \frac{400}{\log K}]$. Pinning down the correct dependence on $\log K$ in (9) is an interesting open question.

Stochastic consumption setting. We construct instances $\{Q^{(i)}\}_{i=1}^K$ in a similar way to the case in deterministic consumption setting, except replacing the consumption model (colored in blue) with the following: Pulling arm $k \in [K]$ consumes $D_{\ell,k}^{(i)} \sim \text{Bern}(d_{\ell,k})$ units of resource ℓ, where $d_{\ell,k}$ is defined in (7). In addition, the reward R_k and $D_{\ell,1},\ldots D_{\ell,K}$ are jointly independent. We have the following lower bound result:

Theorem 5. *Consider a fixed but arbitrary function $g : [0,+\infty) \to [0,+\infty)$ that is increasing and $\lim_{d\to 0^+} \frac{1}{g(d)\log\frac{1}{d}} = +\infty$, $g(0) = 0$, as well as any fixed $\{r_k\}_{k=1}^K \subset (0,1)$, $\frac{1}{2} = r_1 > r_2 \geq \cdots \geq r_K = \frac{1}{4}$, any fixed $\{d_{\ell,(k)}^0\}_{k=1,\ell=1}^{K,L} \subset \mathbb{R}$, $d_{\ell,(1)}^0 \geq d_{\ell,(2)}^0 \geq \cdots \geq d_{\ell,(K)}^0$, and any fixed $i \in \{2,\cdots,K\}$. We can identify $\bar{c} \in$*

$(0,1)$, *such that for any* $c \in (0,\bar{c})$ *and large enough* $\{C_\ell\}_{\ell=1}^L$, *by taking* $d_{\ell,(j)} = cd^0_{\ell,(j)}, \forall j \in [K], \forall \ell \in [L]$, *we can construct corresponding instances* $Q^{(j)}$*: (1) pulling arm* $k \in [K]$ *generates a random reward* $R_k \sim \mathcal{N}(r_k^{(j)}, 1)$, *where* $r_k^{(j)} = \begin{cases} r_k & \text{if } k \neq j, \\ 1 - r_k & \text{if } k = j, \end{cases}$, *(2) pulling arm* $k \in [K]$ *consumes* $D_\ell \sim Bern(d_{\ell,k})$, $d_{\ell,k} = \begin{cases} d_{\ell,(2)} & \text{if } k = 1, \\ d_{\ell,(1)} & \text{if } k = 2, \\ d_{\ell,(k)} & \text{if } k \in \{3,\ldots,K\} \end{cases}$ *for* $\ell \in [L]$. *The following performance lower bound holds for any strategy:*

$$\max_{j\in\{1,i\}} \Pr_{Q^{(j)}}(\psi \neq j) \geq \exp\left(-2\tilde{\gamma}^{sto}(Q^{(j)})\right),$$

where $\tilde{\gamma}^{sto}(Q^{(j)}) = \min_{\ell\in[L]} \frac{C_\ell}{\tilde{H}^{sto}_{2,\ell}}$, *and* $\tilde{H}^{sto}_{2,\ell} = \max_{k\in\{2,3,\cdots,K\}} \frac{\sum_{j=1}^k g(d_{\ell,(j)})}{\Delta_k^2}$.

In Theorem 5, which is proved in Appendix B.6, we establish a lower bound for stochastic consumption in multiple resource scenarios. This theorem, alongside Theorem 3, illustrates the near-optimality of the SH-RR approach in stochastic cases, similar to the conclusion in deterministic settings.

In Theorem 3, we introduce a novel complexity measure, $H^{\text{sto}}_{2,\ell}(Q)$. This measure is notable for incorporating a term $\frac{1}{\log\frac{1}{d}}$, which exceeds d when d is small. This results in a larger and possibly weaker upper bound compared to the deterministic case and is also different from traditional BAI literature. A pertinent question arises: Is it feasible to refine this term from $\frac{1}{\log\frac{1}{d}}$ to d?

Theorem 5 directly addresses this question, clarifying that such a refinement should not be expected to hold for any given sets $\{r_k\}_{k=1}^K$ and $\{d_{\ell,k}\}_{k=1,\ell=1}^{K,L}$. This result decisively indicates that the term $\frac{1}{\log\frac{1}{d}}$ in the definition of $H^{\text{sto}}_{2,\ell}(Q)$ is irreplaceable with $\frac{1}{(\log\frac{1}{d})^{1+\epsilon}}$ for any $\varepsilon > 0$, and certainly not with d itself. This pivotal finding emphasizes a crucial aspect: stochastic consumption scenarios are inherently more complex than deterministic ones, especially in cases where the mean consumptions are extremely low.

It is crucial to highlight the difference in Theorem 5: it asserts $\max_{j\in 1,i}$ in its conclusion, diverging from the conventional form of $\max_{j\in[K]}$. Additionally, the ratio $\frac{\tilde{H}^{\text{sto}}_{2,\ell}(Q)}{H^{\text{sto}}_{2,\ell}(Q)}$ can approach to 0, given the function g and sufficiently small $\{d_{\ell,(k)}\}_{k=1,\ell=1}^{K,L}$. This gap that might stem from how the pulling times of arm i are approximated. The current derivation, based on the assumption that all resources are allocated to arm i, somehow replace the step 2 in appendix B.5. This suggests that there is room for improvement in the approximation. A tighter approximation in the exponential term is to be explored.

6 Numerical Experiments

We conducted a performance evaluation of the SH-RR method on both synthetic and real-world problem sets. Our evaluation included a comparison of SH-RR against four established baseline strategies: Anytime-LUCB (AT-LUCB) (Jun and Nowak, 2016), Upper Confidence Bounds (UCB) (Bubeck et al., 2009), Uniform Sampling, and Sequential Halving (Karnin et al., 2013) augmented with the doubling trick. Unlike fixed confidence and fixed budget strategies, these baseline methods are *anytime* algorithms, which recommend an arm as the best arm after each arm pull, continuously, until a specified resource constraint is violated. The evaluation was carried out until a resource constraint was breached, at which point the last recommended arm was returned as the identified arm. Fixed confidence and fixed budget strategies were deemed inapplicable to the BAIwRC problem as they necessitate an upper bound on the BAI failure probability and an upper limit on the number of arm pulls in their respective settings. Further details regarding the experimental set-ups are elaborated in appendix C.

Synthesis problems. We investigated the performance of our algorithm across various synthetic settings, each with distinct reward and consumption dynamics. (1) *High match High* (HmH) where higher mean rewards correspond to higher mean consumption, (2) *High match Low* (HmL) where they correspond to lower mean consumption, and (3) *Mixture* (M) where each arm consumes less of one resource while consuming more of another, applicable when $L = 2$. Additionally, resource consumption variability was categorized into deterministic, correlated (random and correlated with rewards), and uncorrelated (random but independent of rewards), with the deterministic setting omitted for $L = 2$ due to similar results to $L = 1$. Reward variability across arms was explored through four settings: One Group of Sub-optimal, Trap, Polynomial, and Geometric, analogous to Karnin et al. (2013). The various combinations of these settings are illustrated in Figure 2, with more detailed descriptions provided in Appendices C.1 and C.2.

Figure 2 presents the failure probability of the different strategies in different setups, under $K = 256$, $L = 1, 2$ with an initial budget of 1500 for each resource. Each strategy was executed over 1000 independent trials, with the failure probability quantified as (# trials that fails BAI)/1000. Our analysis anticipated a higher difficulty level for the HmH instances, a notion substantiated by the experimental outcomes.

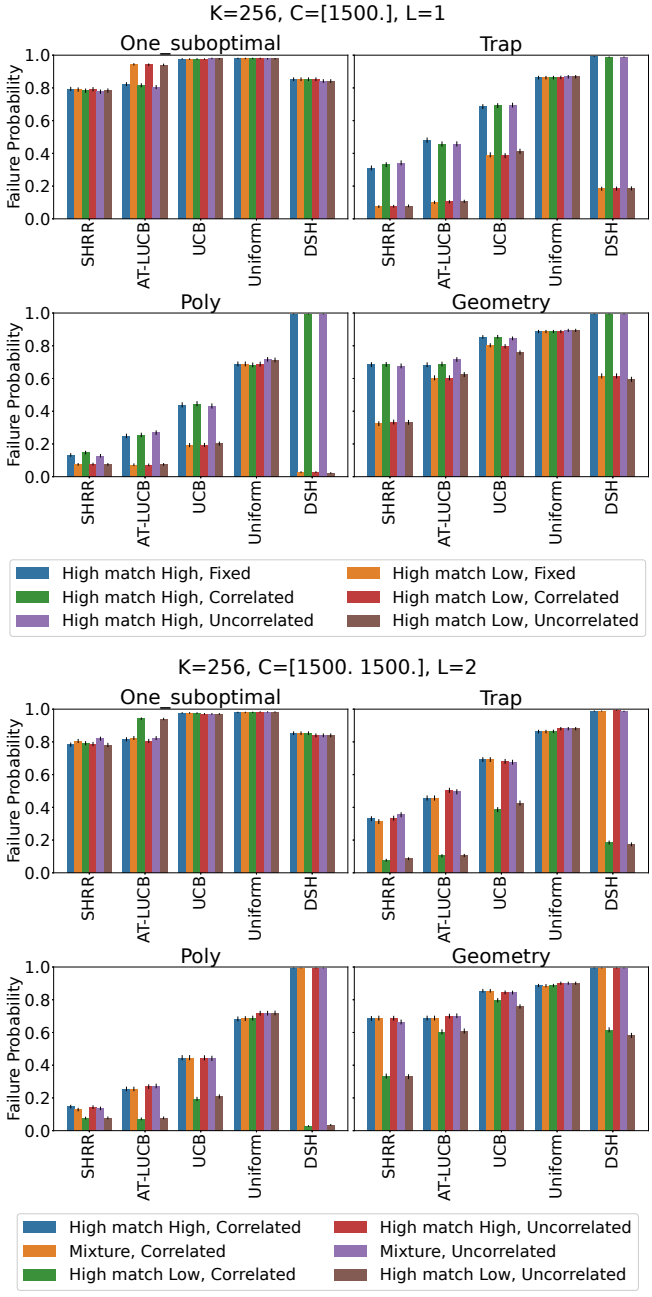

Figure 2: Comparison of SH-RR and anytime baselines in different setups

The bottom panel illustrates comparable performances on M and HmH setups, suggesting the scarcer resource could predominantly influence the performance. This observed behavior aligns with the utilization of the min operator in the definitions of γ^{det} and γ^{sto}.

Our proposed algorithm SH-RR is competitive compared to these state-of-the-art benchmarks. SH-RR achieves the best performance by a considerable margin for HmL, while still achieve at least a matching performance compared to the baselines for HmH. SH-RR favors arms with relatively high empirical reward, thus when those arms consume less resources, SH-RR can achieve a higher probability of BAI. For confidence bound based algorithms such as AT-LUCB and UCB, resource-consumption-heavy sub-optimal arms are repeatedly pulled, leading to resource wastage and a higher failure probability. These empirical results demonstrate the necessity of SH-RR algorithm for BAI-wRC in order to achieve competitive performance in a variety of settings.

Real-world problems. We implemented different machine learning models, each adorned with various hyperparameter combinations, as distinct arms. The overarching goal is to employ diverse BAI algorithms to unravel the most efficacious model and hyperparameter ensemble for tackling supervised learning tasks. There is a single constraint on running time for each BAI experiment. To meld simplicity with time-efficiency, we orchestrated implementations of four quintessential, yet straightforward machine learning models: K-Nearest Neighbour, Logistic Regression, Adaboost, and Random Forest. Each model is explored with eight unique hyperparameter configurations. We considered 5 classification tasks, including (1) Classify labels 3 and 8 in part of the MNIST dataset (MNIST 3&8). (2) Optical recognition of handwritten digits data set (Handwritten). (3) Classify labels -1 and 1 in the MADELON dataset (MADELON). (4) Classify labels -1 and 1 in the Arcene dataset (Arcene). (5) Classify labels on weight conditions in the Obesity dataset (Obesity). See appendix C.3 for details on the set-up.

We designated the arm with the lowest empirical mean cross-entropy, derived from a combination of machine learning models and hyperparameters, as the best arm. Our BAI experiments were conducted across 100 independent trials. During each arm pull in a BAI experiment round—i.e., selecting a machine learning model with a specific hyperparameter combination—we partitioned the datasets randomly into training and testing subsets, maintaining a testing fraction of 0.3. The training subset was utilized to train the machine learning models, and the cross-entropy computed on the testing subset served as the realized reward.

The results, showcased in Table 1 and 2, delineate the failure probability in identifying the optimal machine learning model and hyperparameter configuration for each BAI algorithm. Amongst the tested algorithms, SH-RR emerged as the superior performer across all experiments. This superior performance can be attributed to two primary factors: (1) classifiers with lower time consumption, such as KNN and Random Forest, yielded lower cross-entropy, mirroring the HmL setting; and (2) the scant randomness in realized Cross-Entropy ensured that after each half-elimination in SH-RR, the best arm was retained, underscoring the algorithm's efficacy.

Table 1: Failure Probability of different BAI strategies on Real-life datasets

Algorithm	MNIST 3&8	Handwritten
SHRR	**0**	**0.12**
ATLUCB	0.21	0.23
UCB	0.21	0.34
Uniform	0.21	0.25
DSH	0.14	0.20

Table 2: Failure Probability of different BAI strategies on Real-life datasets, cont.

Algorithm	Arcene	Obesity	MADELON
SHRR	**0.38**	**0.31**	**0**
ATLUCB	0.6	0.43	0.42
UCB	0.71	0.43	0.30
Uniform	0.81	0.56	0.29
DSH	0.67	0.54	0.12

Acknowledgement

The authors are very thankful to Kwang-Sung Jun for his detailed explanations on the AT-LUCB policy and the provision of codes. In addition, the authors would like to thank the reviewing team for the constructive suggestions to strengthen the results. The research is partially funded by a Singapore Ministry of Education AcRF Tier 2 Grant (Project ID: MOE-000238-00, Award Number: MOE-T2EP20121-0012).

References

Abdolshah, M., Shilton, A., Rana, S., Gupta, S., and Venkatesh, S. (2019). Cost-aware multi-objective bayesian optimisation. *arXiv preprint arXiv:1909.03600*.

Agrawal, S. and Devanur, N. (2016). Linear contextual bandits with knapsacks. In *Advances in Neural Information Processing Systems*, volume 29.

Agrawal, S. and Devanur, N. R. (2014). Bandits with concave rewards and convex knapsacks. In *Proceedings of the Fifteenth ACM Conference on Economics and Computation*, page 989–1006. Association for Computing Machinery.

Audibert, J.-Y. and Bubeck, S. (2010). Best arm identification in multi-armed bandits. In *Proceedings of the 23rd Annual Conference on Learning Theory (COLT)*.

Badanidiyuru, A., Kleinberg, R., and Slivkins, A. (2018). Bandits with knapsacks. *Journal of ACM*, 65(3).

Belakaria, S., Doppa, J. R., Fusi, N., and Sheth, R. (2023). Bayesian optimization over iterative learners with structured responses: A budget-aware planning approach. In *International Conference on Artificial Intelligence and Statistics*, pages 9076–9093. PMLR.

Bohdal, O., Balles, L., Ermis, B., Archambeau, C., and Zappella, G. (2022). Pasha: Efficient hpo with progressive resource allocation. *arXiv preprint arXiv:2207.06940*.

Bubeck, S., Munos, R., and Stoltz, G. (2009). Pure exploration in multi-armed bandits problems. In *Algorithmic Learning Theory*, pages 23–37, Berlin, Heidelberg. Springer Berlin Heidelberg.

Carpentier, A. and Locatelli, A. (2016). Tight (lower) bounds for the fixed budget best arm identification bandit problem. In *29th Annual Conference on Learning Theory*, volume 49, pages 590–604. PMLR.

Csiszár, I. (1998). The method of types [information theory]. *IEEE Transactions on Information Theory*, 44(6):2505–2523.

Even-Dar, E., Mannor, S., and Mansour, Y. (2002). Pac bounds for multi-armed bandit and markov decision processes. In *Proceedings of the 15th Annual Conference on Computational Learning Theory*, COLT '02, page 255–270. Springer-Verlag.

Forrester, A. I., Sóbester, A., and Keane, A. J. (2007). Multi-fidelity optimization via surrogate modelling. *Proceedings of the royal society a: mathematical, physical and engineering sciences*, 463(2088):3251–3269.

Foumani, Z. Z., Shishehbor, M., Yousefpour, A., and Bostanabad, R. (2023). Multi-fidelity cost-aware bayesian optimization. *Computer Methods in Applied Mechanics and Engineering*, 407:115937.

Frazier, P. I. (2018). A tutorial on bayesian optimization. *arXiv preprint arXiv:1807.02811*.

Gabillon, V., Ghavamzadeh, M., and Lazaric, A. (2012). Best arm identification: A unified approach to fixed

budget and fixed confidence. In *Advances in Neural Information Processing Systems*, volume 25.

Garivier, A. and Kaufmann, E. (2016). Optimal best arm identification with fixed confidence. In *29th Annual Conference on Learning Theory*, volume 49 of *Proceedings of Machine Learning Research*, pages 998–1027. PMLR.

Guinet, G., Perrone, V., and Archambeau, C. (2020). Pareto-efficient acquisition functions for cost-aware bayesian optimization. *arXiv preprint arXiv:2011.11456*.

Hou, Y., Tan, V. Y., and Zhong, Z. (2022). Almost optimal variance-constrained best arm identification. *arXiv preprint arXiv:2201.10142*.

Ivkin, N., Karnin, Z., Perrone, V., and Zappella, G. (2021). Cost-aware adversarial best arm identification.

Jamieson, K., Malloy, M., Nowak, R., and Bubeck, S. (2014). lil' ucb : An optimal exploration algorithm for multi-armed bandits. In *Proceedings of The 27th Conference on Learning Theory*, volume 35 of *Proceedings of Machine Learning Research*, pages 423–439. PMLR.

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.

Jamieson, K. and Talwalkar, A. (2016). Non-stochastic best arm identification and hyperparameter optimization. In *Artificial intelligence and statistics*, pages 240–248. PMLR.

Jun, K.-S. and Nowak, R. (2016). Anytime exploration for multi-armed bandits using confidence information. In *Proceedings of The 33rd International Conference on Machine Learning*, pages 974–982. PMLR.

Kandasamy, K., Dasarathy, G., Schneider, J., and Póczos, B. (2017). Multi-fidelity bayesian optimisation with continuous approximations. In *International Conference on Machine Learning*, pages 1799–1808. PMLR.

Karnin, Z. S., Koren, T., and Somekh, O. (2013). Almost optimal exploration in multi-armed bandits. In *Proceedings of the 30th International Conference on Machine Learning, ICML*, volume 28, pages 1238–1246. JMLR.org.

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):1–42.

Lattimore, T. and Szepesvári, C. (2020). *Bandit Algorithms*. Cambridge University Press.

Lee, E. H., Perrone, V., Archambeau, C., and Seeger, M. (2020). Cost-aware bayesian optimization. *arXiv preprint arXiv:2003.10870*.

Li, L., Jamieson, K., Rostamizadeh, A., Gonina, E., Ben-Tzur, J., Hardt, M., Recht, B., and Talwalkar, A. (2020). A system for massively parallel hyperparameter tuning. *Proceedings of Machine Learning and Systems*, 2:230–246.

Li, S., Zhang, L., Yu, Y., and Li, X. (2023). Optimal arms identification with knapsacks.

Luong, P., Nguyen, D., Gupta, S., Rana, S., and Venkatesh, S. (2021). Adaptive cost-aware bayesian optimization. *Knowledge-Based Systems*, 232:107481.

Mannor, S. and Tsitsiklis, J. N. (2004). The sample complexity of exploration in the multi-armed bandit problem. *J. Mach. Learn. Res.*, 5:623–648.

Poloczek, M., Wang, J., and Frazier, P. (2017). Multi-information source optimization. *Advances in neural information processing systems*, 30.

Sankararaman, K. A. and Slivkins, A. (2018). Combinatorial semi-bandits with knapsacks. In *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics*, volume 84 of *Proceedings of Machine Learning Research*, pages 1760–1770. PMLR.

Snoek, J., Larochelle, H., and Adams, R. P. (2012). Practical bayesian optimization of machine learning algorithms. *Advances in neural information processing systems*, 25.

Sui, Y., Gotovos, A., Burdick, J., and Krause, A. (2015). Safe exploration for optimization with gaussian processes. In *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pages 997–1005. PMLR.

Sui, Y., Zhuang, V., Burdick, J. W., and Yue, Y. (2018). Stagewise safe bayesian optimization with gaussian processes. In *International Conference on Machine Learning*.

Swersky, K., Snoek, J., and Adams, R. P. (2013). Multi-task bayesian optimization. *Advances in neural information processing systems*, 26.

Wang, Z., Wagenmaker, A. J., and Jamieson, K. (2022). Best arm identification with safety constraints. In *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics*, volume 151 of *Proceedings of Machine Learning Research*, pages 9114–9146. PMLR.

Wu, J., Toscano-Palmerin, S., Frazier, P. I., and Wilson, A. G. (2020). Practical multi-fidelity bayesian optimization for hyperparameter tuning. In *Uncertainty in Artificial Intelligence*, pages 788–798. PMLR.

Zappella, G., Salinas, D., and Archambeau, C. (2021). A resource-efficient method for repeated hpo and nas problems. *arXiv preprint arXiv:2103.16111*.

Zhao, Y., Stephens, C., Szepesvári, C., and Jun, K. (2022). Revisiting simple regret minimization in multi-armed bandits. *CoRR*, abs/2210.16913.

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/No/Not Applicable]
   Yes. Please check the section 2 and 3.
   (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. [Yes/No/Not Applicable]
   Yes. Please check the section 4.
   (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. [Yes/No/Not Applicable]
   Yes. Please check the supplementary file.

2. For any theoretical claim, check if you include:
   (a) Statements of the full set of assumptions of all theoretical results. [Yes/No/Not Applicable]
   Yes. Please check the section 4 and 5.
   (b) Complete proofs of all theoretical results. [Yes/No/Not Applicable]
   Yes. Please check the appendix.
   (c) Clear explanations of any assumptions. [Yes/No/Not Applicable]
   Yes.

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/No/Not Applicable]
   Yes. Please check the appendix.
   (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). [Yes/No/Not Applicable]
   (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/No/Not Applicable]
   Yes. Please check the appendix.
   (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). [Yes/No/Not Applicable]
   Yes. Please check the appendix.

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. [Yes/No/Not Applicable]
   Yes.
   (b) The license information of the assets, if applicable. [Yes/No/Not Applicable]
   Yes.
   (c) New assets either in the supplemental material or as a URL, if applicable. [Yes/No/Not Applicable]
   Not Applicable.
   (d) Information about consent from data providers/curators. [Yes/No/Not Applicable]
   Not Applicable.
   (e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. [Yes/No/Not Applicable]
   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. [Yes/No/Not Applicable]
   Not Applicable.
   (b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. [Yes/No/Not Applicable]
   Not Applicable.
   (c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. [Yes/No/Not Applicable]
   Not Applicable.

A Auxiliary Results

Lemma 6 (Chernoff). *Let $\{X_n\}_{n=1}^N$ be i.i.d random variables, where $X_n \in [0,1]$ almost surely, with common mean $\mathbb{E}[X_1] = \mu \in [0,1]$. For any $\mu_+ \in (\mu, 1]$, it holds that*

$$\mathbb{P}\left[\frac{1}{N}\sum_{n=1}^{N} X_n \geq \mu_+\right] \leq \exp\left(-N \cdot KL(\mu_+, \mu)\right),$$

where $KL(p,q) = p\log\frac{p}{q} + (1-p)\log\frac{1-p}{1-q}$ for $p, q \in [0,1]$. In addition, for any $\epsilon \in (0, 2e-1)$, it holds that

$$\mathbb{P}\left[\frac{1}{N}\sum_{n=1}^{N} X_n \geq (1+\epsilon)\mu\right] \leq \exp\left(-\frac{N\mu\epsilon^2}{4}\right).$$

Lemma 6 can be extracted from Exercise 10.3 in Lattimore and Szepesvári (2020).

B Proofs

B.1 Proof of Claim 1

Proof of Claim 1. The total type-ℓ resource consumption is

$$\begin{aligned}
&\sum_{q=0}^{\lceil \log_2 K\rceil - 1} I_\ell^{(q)} \\
=&\sum_{q=0}^{\lceil \log_2 K\rceil - 1} \left[\frac{C_\ell}{\lceil \log_2 K\rceil} + (\mathsf{Ration}_\ell^{(q)} - \mathsf{Ration}_\ell^{(q+1)})\right] \\
=&C_\ell + (\mathsf{Ration}_\ell^{(0)} - \mathsf{Ration}_\ell^{(\lceil \log_2 K\rceil)}).
\end{aligned}$$

We complete the proof by showing that $\mathsf{Ration}_\ell^{(0)} = \frac{C_\ell}{\lceil \log_2 K\rceil} \leq \mathsf{Ration}_\ell^{(q)}$ with certainty for every q. Indeed, with certainty we have $I_\ell^{(q-1)} \leq \mathsf{Ration}_\ell^{(q-1)}$ for every $q \geq 1$. The **while** loop maintains that $I^{(q-1)} \leq \mathsf{Ration}_\ell^{(q-1)} - 1$, which ensures that $I^{(q-1)} \leq \mathsf{Ration}_\ell^{(q-1)}$ when the **while** loop ends, and consequently $\frac{C_\ell}{\lceil \log_2 K\rceil} \leq \mathsf{Ration}_\ell^{(q)}$ by Line 17. Altogether, the claim is shown. □

B.2 Proof of Theorem 2

Denote $T^{(q)}$ as the pulling times at the q^{th} phase, $S^{(q)}$ as the surviving set at the q^{th} phase. The size of $S^{(q)}$ is always $\lceil \frac{K}{2^q} \rceil$. Assume the pulling times of each arm in each phase are the same (the difference is at most 1). $T^{(q)}$ is a random number. Conditioned on $S^{(q)}$, $T^{(q)} = \min_\ell \frac{C_\ell}{\lceil \log_2 K\rceil \sum_{k\in S^{(q)}} d_{k,\ell}}$. Thus we can assert $\mathbb{P}\left(T^{(q)} \geq \min_\ell \frac{C_\ell}{\lceil \log_2 K\rceil \sum_{k=1}^{\lceil \frac{K}{2^q}\rceil} d_{\ell,(k)}}\right) = 1$. Define $\tilde{T}^{(q)} = T^{(1)} + T^{(2)} + \cdots + T^{(q)}$, then we can assert $\mathbb{P}\left(\tilde{T}^{(q)} \geq \min_\ell \frac{C_\ell}{\lceil \log_2 K\rceil \sum_{k=1}^{\lceil \frac{K}{2^q}\rceil} d_{\ell,(k)}}\right) = 1$.

Denote set $E_q := \{i : \hat{r}_{\tilde{T}_q} > \hat{r}_{1,\tilde{T}_q}\}$, and define the bad event

$$B^{(q)} = \{|E_q| \geq \lceil \frac{K}{2^{q+1}} \rceil\} \tag{10}$$

We assert that, for any phase q

$$\mathbb{P}(B^{(q)}) \leq K\exp\left(-\min_\ell \frac{C_\ell}{4\lceil \log_2 K\rceil H_{2,\ell}}\right). \tag{11}$$

Once the main assertion (11) is shown, the Theorem can be proved by a union bound over all phases:

$$\begin{aligned}\Pr(\psi \neq 1) \leq & \mathbb{P}(\cup_{q=1}^{\lceil \log_2 K \rceil} B^{(q)}) \\ \leq & \sum_{q=1}^{\lceil \log_2 K \rceil} \mathbb{P}(B^{(q)}) \\ \leq & \lceil \log_2 K \rceil K \exp\left(- \min_{\ell} \frac{C_\ell}{4 \lceil \log_2 K \rceil H_{2,\ell}^{\text{det}}} \right).\end{aligned}$$

In what follows, we establish the main claim (11).

For any q, we have

$$\begin{aligned}& \mathbb{P}(B^{(q)}) \\ \leq & \mathbb{P}\left(\exists k \geq \lceil \frac{K}{2^{q+1}} \rceil, \hat{r}_{k,\tilde{T}_q} > \hat{r}_{1,\tilde{T}_q} \right) \\ \leq & \mathbb{P}\left(\exists k \geq \lceil \frac{K}{2^{q+1}} \rceil, \exists N \geq \min_{\ell} \frac{C_\ell}{\lceil \log_2 K \rceil \sum_{k=1}^{\lceil \frac{K}{2^q} \rceil} d_{\ell,(k)}}, \hat{r}_{k,N} > \hat{r}_{1,N} \right) \\ \leq & \sum_{k=\lceil \frac{K}{2^{q+1}} \rceil}^{K} \mathbb{P}\left(\exists N \geq \min_{\ell} \frac{C_\ell}{\lceil \log_2 K \rceil \sum_{k=1}^{\lceil \frac{K}{2^q} \rceil} d_{\ell,(k)}}, \hat{r}_{k,N} > \hat{r}_{1,N} \right).\end{aligned}$$

Denote $N_0 := \min_\ell \frac{C_\ell}{\lceil \log_2 K \rceil \sum_{k=1}^{\lceil \frac{K}{2^q} \rceil} d_{\ell,(k)}}$. For any k, let $\{R_{k,n}, R_{1,n}\}_{n=1}^{\infty}$ be i.i.d samples of rewards under arm k, arm 1. Define $G_n = R_{k,n} - R_{1,n} + \Delta_k$, then $\mathbb{E}G_n = 0$. And $\hat{r}_{k,N} = \frac{1}{N}\sum_{n=1}^{N} R_{k,n}$, $\hat{r}_{1,N} = \frac{1}{N}\sum_{n=1}^{N} R_{1,n}$. $\hat{r}_{k,N} > \hat{r}_{1,N} \Rightarrow \frac{\sum_{n=1}^{N} G_n}{N} > \Delta_k$. Take $\lambda = \Delta_k$,

$$\begin{aligned}& \mathbb{P}\left(\exists N \geq N_0, \frac{\sum_{n=1}^{N} G_n}{N} > \Delta_k \right) \\ = & \mathbb{P}\left(\exists N \geq N_0, \exp(\lambda \sum_{n=1}^{N} G_n) > \exp(N\lambda\Delta_k) \right) \\ = & \mathbb{P}\left(\sup_{N \geq N_0} \frac{\exp(\lambda \sum_{n=1}^{N} G_n)}{\exp(N\lambda\Delta_k)} > 1 \right).\end{aligned}$$

Since $\mathbb{E}\left[\frac{\exp(\lambda \sum_{n=1}^{N+1} G_n)}{\exp((N+1)\lambda\Delta_k)} | G_1, G_2, \cdots, G_N \right] \leq \frac{\exp(\lambda \sum_{n=1}^{N} G_n)}{\exp(N\lambda\Delta_k)} \frac{\exp(\frac{\lambda^2}{2})}{\exp(\lambda\Delta_k)} \leq \frac{\exp(\lambda \sum_{n=1}^{N} G_n)}{\exp(N\lambda\Delta_k)}$, by Doob's optional stopping

theorem, see Theorem 3.9 in Lattimore and Szepesvári (2020) for example.

$$\begin{aligned}
&\mathbb{P}\left(\sup_{N\geq N_0}\frac{\exp(\lambda\sum_{n=1}^{N}G_n)}{\exp(N\lambda\Delta_k)}>1\right)\\
\leq&\mathbb{E}\frac{\exp(\lambda\sum_{n=1}^{N_0}G_n)}{\exp(N_0\lambda\Delta_k)}\\
\leq&\frac{\exp(\frac{N_0\lambda^2}{2})}{\exp(N_0\lambda\Delta_k)}\\
=&\exp(-\frac{N_0\Delta_k^2}{2})\\
=&\exp\left(-\min_{\ell}\frac{C_\ell}{\lceil\log_2 K\rceil\sum_{k=1}^{\lceil\frac{K}{2^q}\rceil}d_{\ell,(k)}}\frac{(r_1-r_k)^2}{2}\right)\\
\leq&\exp\left(-\min_{\ell}\frac{C_\ell}{\lceil\log_2 K\rceil\sum_{k=1}^{\lceil\frac{K}{2^q}\rceil}d_{\ell,(k)}}\frac{(r_1-r_k)^2}{2}\right).
\end{aligned}$$

Thus

$$\mathbb{P}(B^{(q)})\leq K\exp\left(-\min_{\ell}\frac{C_\ell}{4\lceil\log_2 K\rceil H_{2,\ell}^{\mathrm{det}}}\right).$$

Altogether, the Theorem is proved. □

B.3 Proof of Theorem 3

Before the proof, we need a lemma to bound the pulling times of arms.

Lemma 7. *Let $\{X_n\}_{n=1}^{\infty}$ be i.i.d random variable, $\mathbb{P}(X_n\in[0,1])=1$, and denote $\mathbb{E}X_i=d\in[0,1]$. For any positive integer N, it holds that*

$$\mathbb{P}\left(\frac{1}{N}\sum_{n=1}^{N}X_n>f(d)\right)\leq\exp\left(-\frac{N}{3}\right),$$

where the function f is defined in (5).

We start by defining $\bar{T}^{(q)}=T^{(1)}+\ldots T^{(q)}$ and set $E_q:=\{i:\hat{r}_{i,T^{(q)}}>\hat{r}_{1,T^{(q)}}\}$. The bad event is

$$B^{(q)}=\{|E_q|\geq\lceil\frac{K}{2^{q+1}}\rceil\}. \tag{12}$$

We assert that, for any phase q, it holds with certainty that

$$\mathbb{P}\left(B^{(q)}\right)\leq 2LK\exp\left(-\frac{1}{12}\min_{\ell\in[L]}\{\frac{C_\ell}{\lceil\log_2 K\rceil H_{2,\ell}^{\mathrm{sto}}}\}\right). \tag{13}$$

Remark $\{\hat{k}=1\}\supset\cup_{q=1}^{\log_2 K}\neg B^{(q)}$. Once (13) is shown, the Theorem can be proved by a union bound over all phases:

$$\begin{aligned}
&\mathbb{P}(\hat{k}\neq 1)\\
\leq&\mathbb{P}(\cup_{q=1}^{\log_2 K}B^{(q)})\\
\leq&\sum_{q=1}^{\log_2 K}\mathbb{P}(B^{(q)})\\
\leq&2LK(\log_2 K)\exp\left(-\frac{1}{12}\min_{\ell\in[L]}\{\frac{C_\ell}{\lceil\log_2 K\rceil H_{2,\ell}^{\mathrm{sto}}}\}\right)
\end{aligned}$$

In what follows, we establish the main claim (13). For our analysis, we define $\beta_\ell^{(q)} := \frac{C_\ell}{\lceil \log_2 K \rceil \cdot \sum_{k=1}^{\lceil \frac{K}{2^q} \rceil} f(d_{\ell,(k)})}$, and $\bar{\beta}^{(q)} = \min_{\ell \in [L]} \{\beta_\ell^{(q)}\}.$, then we can split $\mathbb{P}\left(B^{(q)}\right)$ into two parts.

$$
\begin{aligned}
&\mathbb{P}(\mathcal{E}_q) \\
\leq &\mathbb{P}\left(\exists k \geq \lceil \frac{K}{2^{q+1}} \rceil, \hat{r}_{k,\tilde{T}^{(q)}} > \hat{r}_{1,\tilde{T}^{(q)}}\right) \\
\leq &\underbrace{\mathbb{P}\left(\exists k \geq \lceil \frac{K}{2^{q+1}} \rceil, \hat{r}_{k,\tilde{T}^{(q)}} > \hat{r}_{1,\tilde{T}^{(q)}}, \tilde{T}^{(q)} \geq \bar{\beta}^{(q)}\right)}_{(\P)} \\
&+ \underbrace{\mathbb{P}\left(\tilde{T}^{(q)} < \bar{\beta}^{(q)}\right)}_{(\ddagger)}
\end{aligned}
$$

To facilitate our discussions, we denote $\{\tilde{D}_{\ell,k}^{(q)}(n)\}_{n=1}^{\infty}$ as i.i.d. samples of the random consumption of resource ℓ by pulling arm k. For the term (‡), we have

$$
\begin{aligned}
&\mathbb{E}\left[\mathbb{1}\left(T^{(q)} < \bar{\beta}^{(q)}\right) \mid \tilde{S}^{(q)}\right] \\
=&\mathbb{E}\left[\mathbb{1}\left(\sum_{n=1}^{\lceil \bar{\beta}^{(q)} \rceil} \sum_{k \in \tilde{S}^{(q)}} \tilde{D}_{\ell,k}^{(q)}(n) > \mathsf{Ration}_\ell^{(q)} \text{ for some } \ell \in [L]\right) \mid \tilde{S}^{(q)}\right] \\
\leq &\mathbb{E}\left[\mathbb{1}\left(\sum_{n=1}^{\lceil \bar{\beta}^{(q)} \rceil} \sum_{k \in \tilde{S}^{(q)}} \tilde{D}_{\ell,k}^{(q)}(n) > \frac{C_\ell}{\lceil \log_2 K \rceil} \text{ for some } \ell \in [L]\right) \mid \tilde{S}^{(q)}\right] &(14)\\
\leq &\sum_{\ell \in [L]} \sum_{k \in \tilde{S}^{(q)}} \mathbb{E}\left[\mathbb{1}\left(\sum_{n=1}^{\lceil \bar{\beta}^{(q)} \rceil} \tilde{D}_k^{(q)}(n) > \frac{C_\ell f(d_{\ell,k})}{\lceil \log_2 K \rceil \cdot \sum_{k' \in \tilde{S}^{(q)}} f(d_{\ell,k'})}\right) \mid \tilde{S}^{(q)}\right] &(15)\\
\leq &\sum_{\ell \in [L]} \sum_{k \in \tilde{S}^{(q)}} \mathbb{E}\left[\mathbb{1}\left(\sum_{n=1}^{\lceil \beta_\ell^{(q)} \rceil} \tilde{D}_k^{(q)}(n) > \beta_\ell^{(q)} f(d_{\ell,k})\right) \mid \tilde{S}^{(q)}\right] &(16)\\
\leq &\sum_{\ell \in [L]} \sum_{k \in \tilde{S}^{(q)}} \exp\left(-\frac{\beta_\ell^{(q)}}{3}\right) &(17)\\
\leq &L \cdot |\tilde{S}^{(q)}| \cdot \exp\left(-\frac{\bar{\beta}^{(q)}}{3}\right). &(18)
\end{aligned}
$$

Step (14) is by the invariance $\mathsf{Ration}_\ell^{(q)} \leq \frac{C_\ell}{\lceil \log_2 K \rceil}$ maintained by the **while** loop of SH-RR. Step (15) is by the pigeonhole principle. Step (16) is by the definition of $\beta_\ell^{(q)}$. Step (17) is by applying Lemma 7.

Next, we analyze the term (¶), we denote

$$
(\dagger)_k^{(q)} = \mathbb{P}\left(\hat{r}_1^{(q)} < \hat{r}_k^{(q)}, \mathbb{1}(T^{(q)} > \bar{\beta}^{(q)})\right).
$$

We remark that $(\P) \leq \sum_{k=\lceil \frac{K}{2^{q+1}} \rceil}^{K} (\dagger)^{(q)}$. Let $\{R_k^{(q)}(n)\}_{n=1}^{\infty}, \{R_{1,n}^{(q)}\}_{n=1}^{\infty}$ be i.i.d. samples of the rewards under arm k and arm 1 respectively. For each n, we define $W(n) = R_k^{(q)}(n) - R_1^{(q)}(n) + \Delta_k$, where we recall that $\Delta_k = r_1 - r_k$.

Clearly, $\mathbb{E}[W(n)] = 0$, and $W(n)$ are i.i.d. 1-sub-Gaussian. For any $\lambda > 0$, we have

$$\begin{aligned}
&(\dagger)_k^{(q)} \\
\leq &\mathbb{P}\left(\exists N \geq \bar{\beta}^{(q)}, \frac{\sum_{n=1}^N G_n}{N} > \Delta_k\right) \\
= &\mathbb{P}\left(\sup_{N \geq \bar{\beta}^{(q)}} \frac{\exp(\lambda \sum_{n=1}^N G_n)}{\exp(N\lambda\Delta_k)} > 1\right) \\
\leq &\mathbb{E}\left[\frac{\exp\left(\lambda \sum_{n=1}^{\lceil \bar{\beta}^{(q)} \rceil} W(n)\right)}{\exp\left(\lambda \lceil \bar{\beta}^{(q)} \rceil \Delta_k\right)}\right] \qquad (19) \\
\leq &\frac{\exp\left(\frac{\lambda^2 \lceil \bar{\beta}^{(q)} \rceil}{2}\right)}{\exp\left(\lambda \lceil \bar{\beta}^{(q)} \rceil \Delta_k\right)}. \qquad (20)
\end{aligned}$$

Step (19) is by the maximal inequality for (super)-martingale, which is a Corollary of the Doob's optional stopping Theorem, see Theorem 3.9 in Lattimore and Szepesvári (2020) for example. Step (20) is by the fact that $G(n)$ is 1-sub-Gaussian. Finally, applying $\lambda = \Delta_k$, (20) leads us to $(\dagger)_k^{(q)} \leq \exp(-\bar{\beta}^{(q)}\Delta_k^2/2)$,, meaning

$$(\P) \leq K \exp\left(-\min_{\ell \in [L]}\{\beta_\ell^{(q)}\} \frac{(r_1 - r_{\lceil \frac{K}{2^{q+1}} \rceil})^2}{2}\right). \qquad (21)$$

Step (21) is by the assumption that Δ_k is not decreasing. Altogether, combining the upper bounds (17,21) to $(\ddagger), (\P)$ respectively, leads us to the proof of (13).

$$\begin{aligned}
&\mathbb{P}(B^{(q)}) \\
\leq &\sum_{k=\lceil \frac{K}{2^{q+1}} \rceil}^{K} \mathbb{P}\left(\hat{r}_{k,\tilde{T}_q} > \hat{r}_{1,\tilde{T}_q}, \mathbb{1}(\bar{T}_q \geq \bar{\beta}^{(q)})\right) + L \cdot K \cdot \exp\left(-\frac{1}{3}\min_{\ell \in [L]}\{\beta_\ell^{(q)}\}\right) \\
\leq &K \exp\left(-\min_{\ell \in [L]}\{\beta_\ell^{(q)}\} \frac{(r_1 - r_{\lceil \frac{K}{2^{q+1}} \rceil})^2}{2}\right) + L \cdot K \cdot \exp\left(-\frac{1}{3}\min_{\ell \in [L]}\{\beta_\ell^{(q)}\}\right) \\
\leq &2LK \exp\left(-\frac{1}{3}\min_{\ell \in [L]}\{\frac{C_\ell}{4\lceil \log_2 K \rceil H_{2,\ell}^{\text{sto}}}\}\right) \\
= &2LK \exp\left(-\frac{1}{12}\min_{\ell \in [L]}\{\frac{C_\ell}{\lceil \log_2 K \rceil H_{2,\ell}^{\text{sto}}}\}\right).
\end{aligned}$$

B.4 Proof of Lemma 7

The proof involves the consideration of two cases: $d \in (e^{-2}, 1]$ and $d \in (0, e^{-2}]$.

Case 1: $d \in (e^{-2}, 1]$**.** In this case, we have $f(d) > 3d$. Consequently,

$$\begin{aligned}
&\mathbb{P}\left(\frac{1}{N}\sum_{n=1}^N X_n > f(d)\right) \\
\leq &\mathbb{P}\left(\frac{1}{N}\sum_{n=1}^N X_n > 3d\right) \leq \exp\left(-\frac{2^2 N d}{4}\right) \qquad (22) \\
= &\exp(-Nd) \leq \exp\left(-\frac{N}{3}\right). \qquad (23)
\end{aligned}$$

Step (22) is by the Chernoff inequality (see Lemma 6), and step (23) is by the case assumption that $d \geq e^{-2}$. Altogether, **Case 1** is shown.

Case 2: $d \in (0, e^{-2})$**.** In this case, note that we still have $f(d) = 2(\log(1/d))^{-1} \geq 2d > d$. We assert that $\mathrm{KL}(f(d), d) \geq 1/2$. Given the assertion, applying Lemma 6 gives

$$\begin{aligned}
&\mathbb{P}\left(\frac{1}{N}\sum_{n=1}^{N} X_n > f(d)\right) \\
\leq &\exp\left(-N \cdot \mathrm{KL}(f(d), d)\right) \\
\leq &\exp\left(-\frac{N}{2}\right) \leq \exp\left(-\frac{N}{3}\right),
\end{aligned}$$

which establishes the desired inequality. In the remaining, we show the assertion, which is equivalent to the assertion

$$\left(\frac{1}{2}\log\frac{1}{d}\right) \cdot \mathrm{KL}(f(d), d) \leq \frac{1}{4}\log\frac{1}{d}. \tag{24}$$

To demonstrate (24), we start with the left hand side of (24):

$$\begin{aligned}
&\left[\frac{1}{2}\log\frac{1}{d}\right] \cdot \mathrm{KL}\left(\frac{2}{\log\frac{1}{d}}, d\right) \\
=&\log\left(\frac{2}{d\log\frac{1}{d}}\right) + \frac{1}{2}\log\frac{1}{d}\left(1 - \frac{2}{\log\frac{1}{d}}\right)\log\frac{1-\frac{2}{\log\frac{1}{d}}}{1-d} \\
=&\log 2 + \log\frac{1}{d} - \log\log\frac{1}{d} \\
&\underbrace{-\left[\frac{1}{2}\log\frac{1}{d} - 1\right] \cdot \log\left(1 + \frac{\frac{2}{\log\frac{1}{d}} - d}{1 - \frac{2}{\log\frac{1}{d}}}\right)}_{(\dagger)}.
\end{aligned} \tag{25}$$

We argue that $(\dagger) \leq 1$. Indeed,

$$\begin{aligned}
(\dagger) \leq &\left[\frac{1}{2}\log\frac{1}{d} - 1\right] \cdot \frac{1}{1 - \frac{2}{\log\frac{1}{d}}} \cdot \left(\frac{2}{\log\frac{1}{d}} - d\right) &(26)\\
= &\left[\frac{1}{2}\log\frac{1}{d}\right] \cdot \left(\frac{2}{\log\frac{1}{d}} - d\right) \\
= &1 - \frac{d}{2}\log\frac{1}{d} \leq 1. &(27)
\end{aligned}$$

Step (26) is by the fact that $\log(1+x) \leq x$ for all $x > -1$. Step (27) is by the case assumption that $d \in (0, e^{-2})$. Next, we apply the bound $(\dagger) \leq 1$ to (25), which yields

$$\begin{aligned}
&\left[\frac{1}{2}\log\frac{1}{d}\right] \cdot \mathrm{KL}\left(\frac{2}{\log\frac{1}{d}}, d\right) \\
\geq &\log 2 + \log\frac{1}{d} - \log\log\frac{1}{d} - 1 \\
\geq &\log\frac{1}{d} - \log\log\frac{1}{d} - 0.5 \\
\geq &\frac{1}{4}\log\frac{1}{d}.
\end{aligned} \tag{28}$$

Step (28) follows from the fact that $\frac{1}{4}\log\frac{1}{d} \geq 0.5$ and $\frac{1}{2}\log\frac{1}{d} \geq \log\log\frac{1}{d}$ hold for any $d \in (0, e^{-2})$. Altogether, **Case 2** is shown and the Lemma is proved. □

B.5 Proof of Theorem 4

To facilitate our discussion, we denote $\mathbb{E}_i[\cdot]$ as the expectation operator corresponding to the probability measure $\Pr_i$. Theorem 4 is proved in the following two steps.

Step 1. We show that, under the assumption $\Pr_1(\psi \neq 1) < 1/2$, for every $i \in \{2, \ldots, K\}$ it holds that

$$\Pr_i(\psi \neq i) \geq \frac{1}{6} \exp\left(-60 t_i \left(\frac{1}{2} - r_i \right)^2 - 2\sqrt{T \log(12KT)} \right), \tag{29}$$

where

$$t_i = \mathbb{E}_1[T_i], \quad T_i = \sum_{t=1}^{\tau} \mathbf{1}(A(t) = i) \tag{30}$$

is the number of times pulling arm i, and

$$T = \min_{\ell \in [L]} \left\{ \lfloor \frac{C_\ell}{d_{\ell,(K)}} \rfloor \right\} \tag{31}$$

is an upper bound to the number of arm pulls by any policy that satisfies the resource constraints with certainty. Note that if the assumption $\Pr_1(\psi \neq 1) < 1/2$ is violated, the conclusion in Theorem 4 immediately holds for $Q^{(1)}$.

Step 2. We show that there exists $i \in \{2, \ldots K\}$ such that

$$\begin{aligned} t_i (1/2 - r_i)^2 &\leq \min_{\ell \in [L]} \left\{ \frac{2C_\ell}{H^{\det}_{\ell,2}(Q_1)} \right\} \\ &\leq \min_{\ell \in [L]} \left\{ \frac{2C_\ell}{H^{\det}_{\ell,2}(Q_i)} \right\}. \end{aligned} \tag{32}$$

This step crucially hinges on the how the consumption model is set in (7). Finally, Theorem 4 follows by taking $C_1, \ldots, C_L$ so large that

$$\min_{i \in [K], \ell \in [L]} \left\{ \frac{2C_\ell}{H^{\det}_{\ell,2}(Q_i)} \right\} \geq \sqrt{T \log(12KT)}.$$

Such $C_1, \ldots, C_L$ exist. For example, we can take $C_1 = \ldots = C_L = C$, then the left hand side of the above condition grows linearly with C, while the right hand side only grows linearly with $\sqrt{C \log C}$. Altogether, the Theorem is shown, and it remains to establish **Steps 1, 2**.

Establishing on Step 1. To establish (29), we follow the approach in (Carpentier and Locatelli, 2016) and consider the event

$$\mathcal{E}_i = \{\psi = 1\} \cap \{T_i \leq 6t_i\} \cap \{\xi\}$$

for $i \in [K]$. The quantities T_i, t_i are as defined in (30), and ξ is an event concerning an empirical estimate on a certain KL divergence term. To define ξ, it requires some set up. Denote $\nu_k^{(i)}$ as the outcome distribution of arm k in instance $Q^{(i)}$ (recall that the outcome consists of the reward R_k and the consumption $D_{1,k}, \ldots D_{\ell,k}$, where R_k has different distributions under different $Q^{(i)}$, while the distribution of $D_{1,k}, \ldots D_{\ell,k}$ is invariant across $Q^{(1)}, \ldots, Q^{(K)}$). Define

$$\begin{aligned} \mathrm{KL}_i &= \mathrm{KL}(\nu_i^{(i)}, \nu_i^{(1)}) \\ &= \mathrm{KL}(\mathrm{Bern}(r_k), \mathrm{Bern}(1 - r_k)) \qquad (33) \\ &= \mathrm{KL}(\mathrm{Bern}(1 - r_k), \mathrm{Bern}(r_k)) \qquad (34) \\ &= (1 - 2r_k) \log\left(\frac{1 - r_k}{r_k} \right), \end{aligned}$$

where (33) is by the fact that the outcomes $\nu_i^{(i)}, \nu_i^{(1)}$ are identical in the resource consumption but only different in reward. In addition, for each $i \in [K], t \in [T_i]$, we define

$$\widehat{\mathrm{KL}}_{i,t} = \frac{1}{t} \sum_{s=1}^{t} (1 - 2\tilde{R}_i(s)) \log \frac{1 - r_k}{r_k},$$

where $\tilde{R}_i(1), \ldots, \tilde{R}_i(T_i)$ are arm i rewards received during the T_i pulls of arm i in the online dynamics (recall the definition of T_i in (30)). Note that $T_i \leq T$ with certainty. Finally, define confidence radius

$$\mathsf{rad}(t) = (\sqrt{2}\log 3) \cdot \sqrt{\frac{\log 12KT}{t}},$$

and the event ξ is defined as

$$\xi = \left\{\forall i \in [K], t \in [T_i], |\widehat{\mathrm{KL}}_{i,t}| - \mathrm{KL}_i \leq \mathsf{rad}(t)\right\}. \tag{35}$$

The event ξ is also considered in (Carpentier and Locatelli, 2016) when T is part of the problem input instead of a set parameter in (31), and the following result from (Carpentier and Locatelli, 2016) still carries over:

Lemma 8 (Lemma 4 in Carpentier and Locatelli (2016))**.** *$\mathbb{P}_{\mathcal{G}^i}(\xi) \geq \frac{5}{6}$ holds for all $i \in [K]$.*

For every $i \in \{2, \ldots, K\}$, we have

$$\begin{aligned}
&\Pr_i(\psi \neq i)\\
\geq&\mathbb{P}_i(\mathcal{E}_i)\\
=&\mathbb{E}_1\left(\mathbb{1}\{\mathcal{E}_i\}\exp(-T_i\widehat{\mathrm{KL}}_{i,T_i})\right) &(36)\\
\geq&\mathbb{E}_1\left(\mathbb{1}\{\mathcal{E}_i\}\exp(-T_iKL_i - 2\sqrt{T_i\log(12KT)})\right) &(37)\\
\geq&\exp\left(-6t_i\mathrm{KL}_i - 2\sqrt{T\log(12KT)}\right)\cdot\Pr_1(\mathcal{E}_i) &(38)\\
\geq&\left[\frac{2}{3} - \Pr_1(\psi \neq 1)\right]\cdot\exp\left(-6t_i\mathrm{KL}_i - 2\sqrt{T\log(12KT)}\right) &(39)\\
\geq&\left[\frac{2}{3} - \Pr_1(\psi \neq 1)\right]\cdot\exp\left(-60t_i\left(\frac{1}{2} - r_i\right)^2 - 2\sqrt{T\log(12KT)}\right). &(40)
\end{aligned}$$

The above calculations establishes **Step 1**, and we conclude the discussion on **Step 1** by justifying steps (36-39).

Step (36) is by a change-of-measure identity frequently used in the MAB literature. For example, it is established in equation (6) in Audibert and Bubeck (2010) and Lemma 18 in Kaufmann et al. (2016). The identity is described as follows. Let τ be a stopping time with respect to $\{\sigma(H(t))\}_{t=1}^{\infty}$, where we recall that $H(t)$ is the historical observation up to the end of time step t. For any event $\mathcal{E} \in \sigma(H(\tau))$ and any instance index $i \in \{2, \ldots K\}$, it holds that

$$\mathbb{P}_{\mathcal{G}^i}(\mathcal{E}) = \mathbb{E}_{\mathcal{G}^1}\left[\mathbb{1}\{\mathcal{E}\}\exp(-T_i\widehat{\mathrm{KL}}_{i,T_i})\right].$$

Consequently, step (36) holds by the fact that the choice of arm ψ only depends on the observed trajectory $\sigma(H(\tau))$, and evidently both T_i and ξ are both $\sigma(H(\tau))$-measurable. Step (37) is by the event ξ. Step (38) is by the event that $T_i \leq 6t_i$.

Step (39) is by the following calculations:

$$\begin{aligned}
&\Pr_1(\mathcal{E}_i) = 1 - \Pr_1(\neg\mathcal{E}_i)\\
\geq&1 - \mathbb{P}_{\mathcal{G}^1}(\neg\{\psi = 1\}) - \mathbb{P}_{\mathcal{G}^1}(\neg\{\xi\}) - \mathbb{P}_{\mathcal{G}^1}(\neg\{T_k \leq 6t_k\})\\
\geq&\frac{2}{3} - \Pr_1(\psi \neq 1). &(41)
\end{aligned}$$

Step (41) follows from Lemma 9 which shows $\Pr_1(\xi) \geq 5/6$, and the Markov inequality that shows that for any $i \in \{2, \ldots, K\}$:

$$\Pr_i(T_i \geq 6t_i) \leq \frac{1}{6}.$$

Finally, step (40) is by the fact that $r_i \in [\frac{1}{4}, \frac{1}{2}]$ for all $i \in [K]$, leading to $0 \leq \mathrm{KL}_i \leq 10(1 - r_i)^2$ for all $i \in [K]$.

Establishing Step 2. To proceed with **Step 2**, we first define a complexity term $H_{1,\ell}^{\det}(Q)$, which is similar to $H_{2,\ell}^{\det}(Q)$ but the former aids our analysis. For a deterministic consumption instance Q (whose arms are not necessarily ordered as $r_1 \geq r_2 \geq \ldots, r_K$), we denote $\{r_{(k)}\}_{k=1}^K$ as a permutation of $\{r_k\}_{k=1}^K$ such that $r_{(1)} > r_{(2)} \geq \ldots \geq r_{(K)}$. For example, when $Q = Q^{(i)}$, we can have $r_{(1)} = r_i^{(i)}$, $r_{(j+1)} = r_j^{(i)}$ for $j \in \{1, \ldots, i-1\}$, and $r_{(j)} = r_j^{(i)}$ for $j \in \{i+1, \ldots, K\}$. Similarly, denote $\{d_{\ell,(k)}\}_{k=1}^K$ as a permutation of $\{d_{\ell,k}\}_{k=1}^K$ such that $d_{\ell,(1)} \geq d_{\ell,(2)} \geq \ldots \geq d_{\ell,(K)}$. Define $\Delta_{(1)} = \Delta_{(2)} = r_{(1)} - r_{(2)}$, and define $\Delta_{(k)} = r_{(1)} - r_{(k)}$ for $k \in \{3, \ldots, K\}$. Now, we are ready to define

$$H_{\ell,1}^{\det}(Q) = \sum_{k=1}^{K} \frac{d_{\ell,(k)}}{\Delta_{(k)}^2}. \tag{42}$$

In the special case of $L = 1$ and $d_{1,k} = 1$ for all $k \in [K]$, the quantity $H_{\ell,1}^{\det}(Q)$ is equal to the complexity term H_1 defined for BAI in the fixed confidence setting (Audibert and Bubeck, 2010) (the term H_1 is relabeled as H in subsequent research works Karnin et al. (2013); Carpentier and Locatelli (2016)). Observe that for any deterministic consumption instance Q, we always have

$$H_{\ell,2}^{\det}(Q) \leq H_{\ell,1}^{\det}(Q). \tag{43}$$

In addition, we observe that for any $i \in \{2, \ldots, K\}$ and any $\ell \in [L]$, it holds that

$$H_{\ell,1}^{\det}(Q^{(1)}) \geq H_{\ell,1}^{\det}(Q^{(i)}), \tag{44}$$

$$H_{\ell,2}^{\det}(Q^{(1)}) \geq H_{\ell,2}^{\det}(Q^{(i)}). \tag{45}$$

After defining $H_{\ell,1}^{\det}(Q)$, we are ready to proceed to establishing **Step 2**. Recall that $T_i = \sum_{t=1}^{\tau} \mathbf{1}(A(t) = i)$ is the number of arm pulls on arm i. By the requirement of feasibility and the definition of $d_{\ell,k}$ in (7), we know that

$$T_1 d_{\ell,(2)} + T_2 d_{\ell,(1)} + \sum_{k=3}^{K} T_k d_{\ell,(k)} \leq C_\ell$$

holds for all $\ell \in [L]$. Taking expectation $\mathbb{E}_1$ and recalling the definition $t_i = \mathbb{E}_1[T_i]$ in (30), we show that

$$t_1 d_{\ell,(2)} + t_2 d_{\ell,(1)} + \sum_{k=3}^{K} t_k d_{\ell,(k)} \leq C_\ell$$

holds for all ℓ. From our definition of $H_{\ell,1}^{\det}(Q^{(1)})$, for every $\ell \in [L]$ we have

$$\frac{d_{\ell,1}}{H_{\ell,1}^{\det}(Q^{(1)})(\frac{1}{2} - r_2)^2} + \sum_{k=2}^{K} \frac{d_{\ell,(k)}}{H_{\ell,1}^{\det}(Q^{(1)})(\frac{1}{2} - r_k)^2} = 1,$$

which implies that

$$\begin{aligned} &\frac{2C_\ell d_{\ell,(1)}}{H_{\ell,1}^{\det}(Q^{(1)})(\frac{1}{2} - r_2)^2} + \sum_{k=3}^{K} \frac{C_\ell d_{\ell,(k)}}{H_{\ell,1}^{\det}(Q^{(1)})(\frac{1}{2} - r_k)^2} \\ \geq &t_1 d_{\ell,(2)} + t_2 d_{\ell,(1)} + t_3 d_{\ell,(3)} + \cdots + t_K d_{\ell,(K)}. \end{aligned} \tag{46}$$

holds for any ℓ. Inequality (46) implies that for any $\ell \in [L]$, it is either the case that $\frac{2C_\ell \cdot d_{\ell,(1)}}{H_{\ell,1}^{\det}(Q^{(1)})(\frac{1}{2} - r_2)^2} \geq t_2 d_{\ell,(1)}$, or there exists $k_\ell \in \{3, \ldots, K\}$ such that $\frac{C_\ell d_{\ell,(k_\ell)}}{H_{\ell,1}^{\det}(Q^{(1)})(\frac{1}{2} - r_{k_\ell})^2} \geq t_{k_\ell} d_{\ell,(k_\ell)}$. Collectively, the implication is equivalent to saying that for all $\ell \in [L]$, there exists $k_\ell \in \{2, \ldots, K\}$ such that

$$t_{k_\ell} \left(\frac{1}{2} - r_{k_\ell}\right)^2 \leq \frac{2C_\ell}{H_{\ell,1}^{\det}(Q^{(1)})},$$

or more succinctly there exists $i \in \{2, \ldots, K\}$ such that

$$t_i \left(\frac{1}{2} - r_i\right)^2 \leq \min_{\ell \in [L]} \left\{ \frac{2C_\ell}{H_{\ell,1}^{\det}(Q^{(1)})} \right\}.$$

Finally, **Step 2** is established by the observations (43, 44, 45).

B.6 Proof of Theorem 5

Denote τ as the total number of arm pulls. Before proving theorem theorem 5 we need to firstly procide a high probability upper bound to τ, with the lemma 2 in Csiszár (1998).

Lemma 9 (Csiszár (1998))**.** *Denote $\{D_t\}_{t=1}^{\infty}$ be i.i.d. random variables distributed as Bern(d), where $d \in (0,1)$. Let C be a positive real number. Define random variable $\rho = \min\{T : \sum_{t=1}^{T} D_t \geq C\}$. For any integer $t' \in (C, C/d)$, it holds that*

$$\Pr(\rho \leq t') \geq \frac{\exp(-t' \log 2 \cdot KL(C/t', d))}{t'+1},$$

where we denote $KL(p,q) = p\log(p/q) + (1-p)\log((1-p)/(1-q))$.

We assert the following lemma.

Lemma 10. *Denote T_i as the pulling times of arm $i \in \{3, \cdots, K\}$,i.e $T_i = \sum_{s=1}^{\tau} \mathbb{1}(A_s = i)$ and $\{d_{\ell_0,(k)}\}_{k=1}^{K} \subset (0,1)^K$. For any $\ell_0 \in [K]$, if $\forall k, g(d_{\ell_0,(k)}) < \frac{1}{\log \frac{1}{d_{\ell_0,(i)}}}$, $\frac{(r_1-r_i)^2}{(r_1-r_2)^2 \log \frac{1}{d_{\ell_0,(i)}}} + \sum_{k=3}^{K} \frac{(r_1-r_i)^2}{(r_1-r_k)^2 \log \frac{1}{d_{\ell_0,(i)}}} < 1$ and $\log \frac{1}{1-d_{\ell_0,(i)}} < \frac{1}{2}$ all hold, we have*

$$\Pr_1\left(T_i > \frac{C_{\ell_0}}{\frac{g(d_{\ell_0,(1)})}{(r_1-r_2)^2} + \sum_{k=3}^{K} \frac{g(d_{\ell_0,(k)})}{(r_1-r_k)^2}} \frac{1}{(r_1-r_i)^2}\right) \leq 1 - \exp\left(-\frac{C_{\ell_0}}{\frac{1}{4\log \frac{1}{d_{\ell_0,(i)}}}}\right)$$

Remarks: We can derive a similar conclusion For the case that $i = 2$, with assumptions $\forall k, g(d_{\ell_0,(k)}) < \frac{1}{\log \frac{1}{d_{\ell_0,(1)}}}$, $\frac{(r_1-r_i)^2}{(r_1-r_2)^2 \log \frac{1}{d_{\ell_0,(1)}}} + \sum_{k=3}^{K} \frac{(r_1-r_i)^2}{(r_1-r_k)^2 \log \frac{1}{d_{\ell_0,(1)}}} < 1$ and $\log \frac{1}{1-d_{\ell_0,(1)}} < \frac{1}{2}$. The details are omitted here.

Proof. For simplicity, denote $\bar{T}_i := \frac{C_{\ell_0}}{\frac{g(d_{\ell_0,(1)})}{(r_1-r_2)^2} + \sum_{k=3}^{K} \frac{g(d_{\ell_0,(k)})}{(r_1-r_k)^2}} \frac{1}{(r_1-r_i)^2}$, $h = \left(\frac{1}{(r_1-r_2)^2} + \sum_{k=3}^{K} \frac{1}{(r_1-r_k)^2}\right)(r_1-r_i)^2$. Easy to see $h \geq 1$. By simple calculation, we have

$$\begin{aligned}
\Pr_1\left(T_i \geq \bar{T}_i\right) =& \Pr_1\left(T_i \geq \bar{T}_i, \sum_{s=1}^{T_i} D_{i,\ell_0,s} < C_{\ell_0}, \sum_{s=1}^{\bar{T}_i} D_{i,\ell_0,s} < C_{\ell_0}\right) \\
\leq& \Pr_1\left(\sum_{s=1}^{\bar{T}_i} D_{i,\ell_0,s} < C_1\right) \\
=& 1 - \Pr_1\left(\sum_{s=1}^{\bar{T}_i} D_{i,\ell_0,s} \geq C_1\right) \\
\leq& 1 - \Pr_1\left(\sum_{s=1}^{\frac{C_{\ell_0}}{h}\log \frac{1}{d_{\ell_0,(i)}}} D_{i,\ell_0,s} \geq C_{\ell_0}\right).
\end{aligned}$$

where $\{D_{i,\ell_0,s}\}_{s=1}^{+\infty} \overset{i.i.d}{\sim} Bern(d_{\ell_0,(i)})$. The last inequality is from the fact that $g(d_{\ell_0,k}) < \frac{1}{\log \frac{1}{d_{\ell_0,(i)}}}$ holds for all k, further $\bar{T}_i > \frac{C_{\ell_0}}{h} \log \frac{1}{d_{\ell_0,(i)}}$.

As $\frac{h}{\log\frac{1}{d_{\ell_0,(i)}}} > \frac{1}{\log\frac{1}{d_{\ell_0,(i)}}} > d_{\ell_0,(i)}$, we can apply lemma 9.

$$\Pr_1\left(\sum_{s=1}^{\frac{C_{\ell_0}}{h}\log\frac{1}{d_{\ell_0,(i)}}} D_{i,\ell_0,s} \geq C_1\right) \tag{47}$$

$$\geq \frac{\exp\left(-\frac{C_{\ell_0}}{h}\log\frac{1}{d_{\ell_0,(i)}}\mathrm{KL}\left(\frac{h}{\log\frac{1}{d_{\ell_0,(i)}}}, d_{\ell_0,(i)}\right)\right)}{\frac{C_{\ell_0}}{h}\log\frac{1}{d_{\ell_0,(i)}}+1} \tag{48}$$

$$\geq \frac{\exp\left(-(\frac{C_{\ell_0}}{h}\log\frac{1}{d_{\ell_0,(i)}})h - \frac{C_{\ell_0}}{h}\log\frac{1}{d_{\ell_0,(i)}}\log\frac{1}{1-d_{\ell_0,(i)}}\right)}{\frac{C_{\ell_0}}{h}\log\frac{1}{d_{\ell_0,(i)}}+1} \tag{49}$$

$$\geq \exp\left(-\frac{C_{\ell_0}}{\frac{1}{\log\frac{1}{d_{\ell_0,(i)}}}} - \frac{C_{\ell_0}}{\frac{1}{\log\frac{1}{d_{\ell_0,(i)}}}}\frac{1}{2} - \frac{C_{\ell_0}}{\frac{1}{\log\frac{1}{d_{\ell_0,(i)}}}}\right) \tag{50}$$

$$\geq \exp\left(-\frac{C_{\ell_0}}{\frac{1}{4\log\frac{1}{d_{\ell_0,(i)}}}}\right). \tag{51}$$

Step (48) is by lemma 9. Step (49) is by the following fact

$$\begin{aligned}
\mathrm{KL}(\frac{M}{\log\frac{1}{d}}, d) =& \frac{M}{\log\frac{1}{d}}\log\frac{\frac{M}{\log\frac{1}{d}}}{d} + \left(1-\frac{M}{\log\frac{1}{d}}\right)\log\frac{1-\frac{M}{\log\frac{1}{d}}}{1-d}\\
=& \frac{M}{\log\frac{1}{d}}\log\frac{M}{\log\frac{1}{d}} + M + \left(1-\frac{M}{\log\frac{1}{d}}\right)\log(1-\frac{M}{\log\frac{1}{d}}) + \left(1-\frac{M}{\log\frac{1}{d}}\right)\log\frac{1}{1-d}\\
\leq& 0 + M + 0 + \left(1-\frac{M}{\log\frac{1}{d}}\right)\log\frac{1}{1-d}\\
=& M + \log\frac{1}{1-d}
\end{aligned}$$

holds for any M, d such that $\frac{M}{\log\frac{1}{d}}, d \in (0,1)$. Step (50) is due to $h \geq 1$, $\log\frac{1}{1-d_{\ell_0,(1)}} < \frac{1}{2}$ and inequality $e^x \geq x+1$. □

Now we are ready to prove theorem 5. We firstly introduced some notations. Define $\tilde{H}^{\mathrm{sto}}_{1,\ell} := \frac{g(d_{\ell,(1)})}{(r_1-r_2)^2} + \sum_{k=3}^{K}\frac{g(d_{\ell,(k)})}{(r_1-r_k)^2}$, $H^{\mathrm{sto}}_{1,\ell} = \frac{g(d_{\ell,(1)})}{(r_1-r_2)^2} + \sum_{k=2}^{K}\frac{g(d_{\ell,(k)})}{(r_1-r_k)^2}$. Easy to see $2\tilde{H}^{\mathrm{sto}}_{1,\ell} \geq H^{\mathrm{sto}}_{1,\ell} > \tilde{H}^{\mathrm{sto}}_{2,\ell}$. This implies once we prove

$$\max_{j\in\{1,i\}} \Pr_{Q^{(j)},alg}(failure) \geq \exp\left(-\min_{\ell\in[L]}\frac{C_\ell}{\tilde{H}^{\mathrm{sto}}_{1,\ell}}\right),$$

for small enough $\bar{c}$ and large enough $\{C_\ell\}_{\ell=1}^L$, we prove theorem 5. The constructions of $\bar{c}$ and $\{C_\ell\}_{\ell=1}^L$ are as follows. As $\lim_{d\to 0^+}\frac{1}{g(d)\log\frac{1}{d}} = +\infty$ is equivalent to $\lim_{d\to 0^+} g(d)\log\frac{1}{d} = 0$, we can further conclude $\lim_{c\to 0^+} g(cd^0_{\ell,(i)})\log\frac{1}{cd^0_{\ell,(j)}} = \lim_{c\to 0^+} g(cd^0_{\ell,(i)})\log\frac{1}{cd^0_{\ell,(i)}} + g(cd^0_{\ell,(i)})\log\frac{d^0_{\ell,(i)}}{d^0_{\ell,(j)}} = 0, \forall i,j\in[K], \forall\ell\in[L]$, thus we can find a $\bar{c}$, such that when $0 < c < \bar{c}$,

- $16g(d_{\ell,(j)}) = 16g(cd^0_{\ell,(j)}) < \frac{1}{\log\frac{1}{cd^0_{\ell,(i)}}} = \frac{1}{\log\frac{1}{d_{\ell,(i)}}}$ for all $j\in[K], \ell\in[L]$,
- $\log\frac{1}{1-d_{\ell,(j)}} < \frac{1}{2}$, $\log\frac{1}{d_{\ell,(j)}} > 1$ for all $j\in[K]$, $\ell\in[L]$,
- $64\left(\frac{g(d_{\ell,(1)})}{(r_1-r_2)^2} + \sum_{k=3}^{K}\frac{g(d_{\ell,(k)})}{(r_1-r_k)^2}\right)\log\frac{1}{d_{\ell,(i)}} < 2$, for all $\ell\in[L]$,

- $\frac{(r_1-r_i)^2}{(r_1-r_2)^2\log\frac{1}{d_{\ell,(i)}}}+\sum_{k=3}^{K}\frac{(r_1-r_i)^2}{(r_1-r_k)^2\log\frac{1}{d_{\ell,(i)}}}<1$, for all $\ell\in[L]$.

We also require $\{C_\ell\}_{\ell=1}^L$ are large enough, such that for the above given $\{d_{\ell,(k)}\}_{k=1,\ell=1}^{K,L}$, we have

- $\frac{C_\ell}{16\tilde{H}^{\text{sto}}_{1,\ell}}>\sqrt{\frac{4}{16}\frac{C_1}{\tilde{H}^{\text{sto}}_{1,\ell}}\log\frac{1}{16}\frac{C_1}{\tilde{H}^{\text{sto}}_{1,\ell}}\frac{1}{(r_1-r_i)^2}}$ holds for all $\ell\in[L]$,
- $C_\ell\geq\log 64$ for all $\ell\in[L]$.

For these c, $\{d_{\ell,(k)}\}_{k=1,\ell=1}^{K,L}$, and $\{C_\ell\}$, we can start the analysis. Define $\bar{T}_i=\min_{\ell\in[L]}\frac{1}{16}\frac{C_\ell}{\tilde{H}^{\text{sto}}_{1,\ell}}\frac{1}{(r_1-r_i)^2}$, $\hat{KL}_{i,s}=\log\frac{f_i(R_{i,s})}{f_{i'}(R_{i,s})}$, where f_i is the density function of $\mathcal{N}(r_i,1)$, $f_{i'}$ is the density function of $\mathcal{N}(1-r_i,1)$, and $R_{i,s}\sim\mathcal{N}(r_i,1)$. Easy to see

$$\begin{aligned}\log\frac{f_i(R_{i,s})}{f_{i'}(R_{i,s})}=&\log e^{-\frac{(R_{i,s}-r_i)^2-(R_{i,s}-(1-r_i))^2}{2}}\\=&-\frac{(R_{i,s}-r_i-R_{i,s}+(1-r_i))(R_{i,s}-r_i+R_{i,s}-(1-r_i))}{2}\\=&-\frac{2(\frac{1}{2}-r_i)(2R_{i,s}-1)}{2}\\=&-\frac{2(r_1-r_i)(2R_{i,s}-1)}{2}\sim\mathcal{N}(2(r_i-r_1)^2,4(r_1-r_i)^2).\end{aligned}$$

Define $\xi_i=\left\{t\in[\bar{T}_i],\widehat{\text{KL}}_{i,t}-2(r_i-r_1)^2\leq 2|r_1-r_i|\cdot\sqrt{\frac{\min_{\ell\in[L]}\frac{C_\ell}{\tilde{H}^{\text{sto}}_{1,\ell}}+\log\bar{T}_i}{t}}\right\}$, easy to derive the following inequality by Chernoff and union bounds.

$$\Pr_1(\neg\xi_i)\leq\sum_{t=1}^{\bar{T}_i}\exp\left(-\frac{4(r_1-r_a)^2\left(\frac{\min_{\ell\in[L]}\frac{C_\ell}{\tilde{H}^{\text{sto}}_{1,\ell}}+\log\bar{T}_i}{t}\right)}{4(r_1-r_i)^2}t\right)=\exp\left(-\min_{\ell\in[L]}\frac{C_\ell}{\tilde{H}^{\text{sto}}_{1,\ell}}\right).$$

Apply the transportaion equality just like section B.5, we have

$$\begin{aligned}&\Pr_i(\psi\neq i)\\\geq&\mathbb{E}_1\left(\mathbb{1}\{\psi\neq i\}\mathbb{1}\{T_i\leq\bar{T}_i\}\mathbb{1}(\xi_i)\exp(-T_i\widehat{\text{KL}}_{i,T_i})\right)&(52)\\\geq&\mathbb{E}_1\left(\mathbb{1}\{\psi\neq i\}\mathbb{1}\{T_i\leq\bar{T}_i\}\mathbb{1}(\xi_i)\exp\left(-\bar{T}_iKL_i-2(r_1-r_i)\sqrt{\bar{T}_i\left(\min_{\ell\in[L]}\frac{C_\ell}{\tilde{H}^{\text{sto}}_{1,\ell}}+\log\bar{T}_i\right)}\right)\right)&(53)\\\geq&\mathbb{E}_1\left(\mathbb{1}\{\psi\neq i\}\mathbb{1}\{T_i\leq\bar{T}_i\}\mathbb{1}(\xi_i)\exp\left(-\bar{T}_iKL_i-\sqrt{4(r_1-r_i)^2\bar{T}_i\min_{\ell\in[L]}\frac{C_\ell}{\tilde{H}^{\text{sto}}_{1,\ell}}}-\sqrt{4(r_1-r_i)^2\bar{T}_i\log\bar{T}_i}\right)\right)&(54)\\=&\mathbb{E}_1\Bigg(\mathbb{1}\{\psi\neq i\}\mathbb{1}\{T_i\leq\bar{T}_i\}\mathbb{1}(\xi_i)\\&\exp\left(-\frac{2}{16}\min_{\ell\in[L]}\frac{C_\ell}{\tilde{H}^{\text{sto}}_{1,\ell}}-\frac{8}{16}\min_{\ell\in[L]}\frac{C_\ell}{\tilde{H}^{\text{sto}}_{1,\ell}}-\sqrt{\frac{4}{16}\min_{\ell\in[L]}\frac{C_\ell}{\tilde{H}^{\text{sto}}_{1,\ell}}\log\frac{1}{16}\min_{\ell\in[L]}\frac{C_\ell}{\tilde{H}^{\text{sto}}_{1,\ell}}\frac{1}{(r_1-r_i)^2}}\right)\Bigg)&(55)\\\geq&\Pr_1\left((\psi\neq i)\text{ and }(T_i\leq\bar{T}_i)\text{ and }\xi_i\right)\exp\left(-\frac{11}{16}\min_{\ell\in[L]}\frac{C_\ell}{\tilde{H}^{\text{sto}}_{1,\ell}}\right).&(56)\end{aligned}$$

Step (54) is by the inequality $\sqrt{a+b}\leq\sqrt{a}+\sqrt{b}$ for $a,b\geq 0$. Step (56) is by the requirement $\frac{C_\ell}{16\tilde{H}^{\text{sto}}_{1,\ell}}>\sqrt{\frac{4}{16}\frac{C_1}{\tilde{H}^{\text{sto}}_{1,\ell}}\log\frac{1}{16}\frac{C_1}{\tilde{H}^{\text{sto}}_{1,\ell}}\frac{1}{(r_1-r_i)^2}}$ holds for all $\ell\in[L]$.

We can further derive the lower bound of the probabilistic term,

$$\begin{aligned}
&\Pr_1\left((\psi \neq i) \text{ and } (T_i \leq \bar{T}_i) \text{ and } \xi_i\right) \\
\geq& 1 - \Pr_1(\psi = i) - \Pr_1\left(T_i > \bar{T}_i\right) - \Pr_1(\neg\xi_i) \\
\geq& 1 - \exp\left(-\min_{\ell\in[L]} \frac{C_\ell}{\tilde{H}^{\text{sto}}_{1,\ell}}\right) - \Pr_1\left(T_i > \bar{T}_i\right) - \exp\left(-\min_{\ell\in[L]} \frac{C_\ell}{\tilde{H}^{\text{sto}}_{1,\ell}}\right).
\end{aligned}$$

The last inequality is by the assumption $\Pr_1(\psi = i) \leq \exp\left(-\min_{\ell\in[L]} \frac{C_\ell}{\tilde{H}^{\text{sto}}_{1,\ell}}\right)$. If this assumption doesn't hold, then we have completed the proof. Denote $\ell_0 = \arg\min_{\ell\in[L]} \frac{C_\ell}{\frac{g(d_{\ell,(1)})}{(r_1-r_2)^2} + \sum_{k=3}^K \frac{g(\ell,d_{(k)})}{(r_1-r_k)^2}}$, apply lemma 10 to $\Pr_1\left(T_i > \bar{T}_i\right)$, we get

$$\begin{aligned}
&\Pr_1\left((\psi \neq i) \text{ and } (T_i \leq \bar{T}_i) \text{ and } \xi_i\right) \\
\geq& 1 - \exp\left(-\frac{C_{\ell_0}}{\tilde{H}^{\text{sto}}_{1,\ell_0}}\right) - \left(1 - \exp\left(-\frac{C_{\ell_0}}{\frac{1}{4\log\frac{1}{d_{\ell_0,(i)}}}}\right)\right) - \exp\left(-\frac{C_{\ell_0}}{\tilde{H}^{\text{sto}}_{1,\ell_0}}\right) \\
=& \exp\left(-\frac{C_{\ell_0}}{\frac{1}{4\log\frac{1}{d_{\ell_0,(i)}}}}\right) - \exp\left(-\frac{C_{\ell_0}}{\tilde{H}^{\text{sto}}_{1,\ell_0}}\right) - \exp\left(-\frac{C_{\ell_0}}{\tilde{H}^{\text{sto}}_{1,\ell_0}}\right).
\end{aligned}$$

By the property of $c, d_{\ell_0,(i)}$, easy to see

$$\begin{aligned}
&64\left(\frac{g(d_{\ell_0,(1)})}{(r_1-r_2)^2} + \sum_{k=3}^K \frac{g(d_{\ell_0,(k)})}{(r_1-r_k)^2}\right)\log\frac{1}{d_{\ell_0,(i)}} < 2, C_{\ell_0} > \log 64, \log\frac{1}{d_{\ell_0,(i)}} > 1 \\
\Rightarrow& \exp\left(-\frac{C_{\ell_0}}{\tilde{H}^{\text{sto}}_{1,\ell_0}}\right) \leq \frac{1}{4}\exp\left(-\frac{C_{\ell_0}}{\frac{1}{4\log\frac{1}{d_{\ell_0,(i)}}}}\right).
\end{aligned}$$

Thus, we can conclude $\Pr_1\left((\psi \neq i) \text{ and } (T_i \leq \bar{T}_i) \text{ and } \xi_i\right) \geq \frac{1}{2}\exp\left(-\frac{C_{\ell_0}}{\frac{1}{4\log\frac{1}{d_{\ell_0,(i)}}}}\right)$, further

$$\Pr_i(\psi \neq i) \geq \frac{1}{2}\exp\left(-\frac{C_{\ell_0}}{\frac{1}{4\log\frac{1}{d_{\ell_0,(i)}}}}\right)\exp\left(-\frac{11}{16}\frac{C_{\ell_0}}{\tilde{H}^{\text{sto}}_{1,\ell_0}}\right).$$

That implies

$$\begin{aligned}
&\frac{\Pr_i(\psi \neq i)}{\exp\left(-\frac{C_{\ell_0}}{\tilde{H}^{\text{sto}}_{1,\ell_0}}\right)} \\
\geq& \exp\left(\frac{5C_{\ell_0}}{16\tilde{H}^{\text{sto}}_{1,\ell_0}} - \frac{C_{\ell_0}}{\frac{1}{4\log\frac{1}{d_{\ell_0,(i)}}}}\right) \\
=& \exp\left(\frac{\left(5 - 64\left(\frac{g(d_{\ell_0,(1)})}{(r_1-r_2)^2} + \sum_{k=3}^K \frac{g(d_{\ell_0,(k)})}{(r_1-r_k)^2}\right)\log\frac{1}{d_{\ell_0,(i)}}\right)C_{\ell_0},}{16\tilde{H}^{\text{sto}}_{1,\ell_0}}\right) \\
\geq& \exp\left(\frac{C_{\ell_0}}{16\tilde{H}^{\text{sto}}_{1,\ell_0}}\right) > 1.
\end{aligned}$$

The last second inequality is from the property $64\left(\frac{g(d_{\ell,(1)})}{(r_1-r_2)^2}+\sum_{k=3}^{K}\frac{g(d_{\ell,(k)})}{(r_1-r_k)^2}\right)\log\frac{1}{d_{\ell,(i)}}<2$, for all $\ell\in[L]$. The overall inequality suggests

$$\Pr_i(\psi\neq i)\geq\exp\left(-\frac{C_{\ell_0}}{\tilde{H}^{\text{sto}}_{1,\ell_0}}\right)=\exp\left(-\min_{\ell\in[L]}\frac{C_\ell}{\tilde{H}^{\text{sto}}_{1,\ell}}\right)\geq\exp\left(-\min_{\ell\in[L]}\frac{C_\ell}{\frac{1}{2}\left(\frac{g(d_{\ell,(1)})}{(r_1-r_2)^2}+\sum_{k=2}^{K}\frac{g(d_{\ell,(k)})}{(r_1-r_k)^2}\right)}\right),$$

which is our target.

B.7 Improvement of $H^{\mathbf{det}}_{2,\ell}(Q)$ is Unachievable

For a problem instance Q with mean reward $\{r^Q_k\}_{k=1}^K, r^Q_1\geq r^Q_2\geq\cdots\geq r^Q_K$, mean consumption $\{d^Q_{\ell,k}\}_{k=1,\ell=1}^{K,L}$ and budget $\{C_\ell\}_{\ell=1}^L$, we define

$$\tilde{H}^{det}_{1,\ell}(Q)=\frac{d_{\ell,1}}{(r^Q_1-r^Q_2)^2}+\sum_{k=2}^{K}\frac{d_{\ell,k}}{(r^Q_1-r^Q_k)^2} \tag{57}$$

$$\tilde{H}^{det}_{2,\ell}(Q)=\max_{2\leq k\leq K}\frac{\sum_{j=1}^{k}d_{\ell,j}}{(r^Q_1-r^Q_k)^2}. \tag{58}$$

Easy to see $\tilde{H}^{det}_{1,\ell}(Q)\leq H^{det}_{1,\ell}(Q)$, $\tilde{H}^{det}_{2,\ell}(Q)\leq H^{det}_{2,\ell}(Q)$. We want to know whether we can find an algorithm such that for any problem instance Q, we can achieve the following upper bound of the failure probability.

$$\begin{aligned}&\Pr_Q(failure)\\ \leq&poly(K)\exp\left(-\frac{O(1)}{\log_2 K}\min_{\ell\in[L]}\left\{\frac{C_\ell}{\tilde{H}^{det}_{2,\ell}(Q)}\right\}\right).\end{aligned} \tag{59}$$

The answer is No. And the analysis method we used is similar to appendix B.5. We can construct a list of problem instance $Q^{(i)}$, and prove a lower bound that could be larger than the right side of (59), as K and $\{C_\ell\}_{\ell=1}^L$ are large enough.

We focus on $L=1$. Assume there are C units of the resource. Given K, let $d_1=\frac{1}{2^{K-2}}, d_k=\frac{1}{2^{K-k}}, k\geq 2$, $r_1=\frac{1}{2}, r_k=\frac{1}{2}-2^{\frac{k-K-4}{2}}, k\geq 2$. Easy to see $d_1=d_2\leq\cdots\leq d_K$, $\frac{1}{2}=r_1\geq r_2\geq\cdots\geq r_K=\frac{1}{4}$. Then we construct K problem instances $\{Q^{(i)}\}_{i=1}^K$. For problem instance $Q^{(1)}$, the mean reward of k^{th} arm is r_k, following the Bernoulli distribution. And the deterministic consumption of k^{th} arm is d_k. For problem instance $Q^{(i)}, 2\leq i\leq K$, the mean reward of $k^{th}\neq i$ arm is r_k, the mean reward of i^{th} arm is $1-r_i$. And the deterministic consumption of k^{th} arm is d_k. For $i\in[K]$, the best arm of $Q^{(i)}$ is always the i^{th} arm. Then we can calculate $\tilde{H}^{det}_{1,\ell=1}(Q^{(i)})$ and $\tilde{H}^{det}_{2,\ell=1}(Q^{(i)})$ for $i\in[K]$. Easy to derive

$$\begin{aligned}&\tilde{H}^{det}_{1,\ell=1}(Q^{(1)})\\ =&\frac{d_1}{(r_1-r_2)^2}+\sum_{k=2}^{K}\frac{d_k}{(r_1-r_k)^2}\\ =&\frac{\frac{K}{2}}{2^{K-3}\frac{1}{2^{K+2}}}=16K.\end{aligned}$$

$$\tilde{H}^{det}_{2,\ell=1}(Q^{(1)})=\max_{k\geq 2}\frac{\sum_{t=1}^{k}d_t}{(r_1-r_k)^2}=32. \tag{60}$$

For $2 \le i \le K$,

$$\begin{aligned}
&\tilde{H}^{det}_{1,\ell=1}(Q^{(i)}) \\
=&\frac{d_i + d_1}{(1-r_i-r_1)^2} + \sum_{t=2}^{i-1} \frac{d_t}{(1-r_i-r_t)^2} + \\
&\sum_{t=i+1}^{K} \frac{d_t}{(1-r_i-r_t)^2} \\
=&\frac{\frac{1}{2^{K-i}} + \frac{1}{2^{K-2}}}{(2^{\frac{i-K-4}{2}})^2} + \sum_{t=2}^{i-1} \frac{\frac{1}{2^{K-t}}}{(2^{\frac{i-K-4}{2}} + 2^{\frac{t-K-4}{2}})^2} + \\
&\sum_{t=i+1}^{K} \frac{\frac{1}{2^{K-t}}}{(2^{\frac{i-K-4}{2}} + 2^{\frac{t-K-4}{2}})^2} \\
=&\frac{2^i+4}{2^{i-4}} + \sum_{t=2}^{i-1} \frac{2^t}{2^{i-4} + 2^{\frac{i+t}{2}-3} + 2^{t-4}} + \\
&\sum_{t=i+1}^{K} \frac{2^t}{2^{i-4} + 2^{\frac{i+t}{2}-3} + 2^{t-4}}.
\end{aligned} \tag{61}$$

$$\begin{aligned}
&\tilde{H}^{det}_{2,\ell=1}(Q^{(i)}) \\
=&\max\{\frac{d_i + d_1}{(1-r_i-r_1)^2}, \max_{2\le t\le i-1} \frac{d_i + \sum_{l=1}^{t} d_l}{(1-r_i-r_t)^2}, \max_{i+1\le t\le K} \frac{\sum_{l=1}^{t} d_l}{(1-r_i-r_t)^2}\} \\
=&\max\{\frac{\frac{1}{2^{K-i}} + \frac{1}{2^{K-2}}}{(2^{\frac{i-K-4}{2}})^2}, \max_{2\le t\le i-1} \frac{\frac{1}{2^{K-i}} + \frac{1}{2^{K-t-1}}}{(2^{\frac{i-K-4}{2}} + 2^{\frac{t-K-4}{2}})^2}, \max_{i+1\le t\le K} \frac{\frac{1}{2^{K-t-1}}}{(2^{\frac{i-K-4}{2}} + 2^{\frac{t-K-4}{2}})^2}\} \\
=&\max\{\frac{2^i+4}{2^{i-4}}, \max_{2\le t\le i-1} \frac{2^i + 2^{t+1}}{2^{i-4} + 2^{\frac{i+t}{2}-3} + 2^{t-4}}, \max_{i+1\le t\le K} \frac{2^{t+1}}{2^{i-4} + 2^{\frac{i+t}{2}-3} + 2^{t-4}}\}.
\end{aligned} \tag{62}$$

With a simple calculation, for $2 \le i \le K$, we have $\frac{2^i+4}{2^{i-4}} \le 32$, $\frac{2^i+2^{t+1}}{2^{i-4}+2^{\frac{i+t}{2}-3}+2^{t-4}} \le 32$ and $\frac{2^{t+1}}{2^{i-4}+2^{\frac{i+t}{2}-3}+2^{t-4}} \le 32$. Thus we can conclude $\tilde{H}^{det}_{2,\ell=1}(Q^{(i)}) \le 32 = \tilde{H}^{det}_{2,\ell=1}(Q^{(1)})$. On the other hand, easy to check $\tilde{H}^{det}_{1,\ell=1}(Q^{(i)}) \le \tilde{H}^{det}_{1,\ell=1}(Q^{(1)})$ from the definition of $\tilde{H}^{det}_{1,\ell=1}$.

Following the step 1 in appendix B.5, we can conclude for every $i \in \{2, \dots, K\}$ it holds that

$$\begin{aligned}
&\Pr_i(\psi \ne i) \\
\ge&\frac{1}{6} \exp\left(-60 t_i \left(\frac{1}{2} - r_i\right)^2 - 2\sqrt{T \log(12KT)}\right),
\end{aligned} \tag{63}$$

where

$$t_i = \mathbb{E}_1[T_i], \quad T_i = \sum_{t=1}^{\tau} \mathbf{1}(A(t) = i) \tag{64}$$

is the number of times pulling arm i, and

$$T = \lfloor \frac{C}{d_1} \rfloor \tag{65}$$

is an upper bound to the number of arm pulls by any policy that satisfies the resource constraints with certainty.

Following the step 2 in appendix B.5, recall $d_1 = d_2 = \frac{1}{2^{K-2}}$, we can derive

$$\sum_{k=2}^{K} \frac{2d_k}{\tilde{H}^{det}_{1,\ell=1}(Q^{(1)})(r_1 - r_k)^2} \ge 1.$$

Since $\sum_{k=1}^{K} t_k d_k \leq C$, we can further conclude

$$\sum_{k=2}^{K} \frac{2Cd_k}{\tilde{H}^{det}_{1,\ell=1}(Q^{(1)})(r_1 - r_k)^2} \geq \sum_{k=1}^{K} t_k d_k$$

which implies there exists $i \geq 2$, such that $\frac{2Cd_i}{\tilde{H}^{det}_{1,\ell=1}(Q^{(1)})(r_1-r_i)^2} \geq t_i d_i$. For this i,

$$\begin{aligned}
&\mathbb{P}_{\mathcal{G}^i}(\hat{k} \neq i) \\
\geq &\frac{1}{6}\exp\Bigg(-120\frac{C}{\tilde{H}^{det}_{1,\ell=1}(Q^{(1)})} - \\
&\sqrt{2}\log 3\sqrt{\lfloor\frac{C}{d_{(k)}}\rfloor \log(12\lfloor\frac{C}{d_{(k)}}\rfloor K)}\Bigg).
\end{aligned}$$

When C is large enough, we can assume $120\frac{C}{H_1(Q^{(1)})} > \sqrt{2}\log 3\sqrt{\lfloor\frac{C}{d_{(k)}}\rfloor \log(12\lfloor\frac{C}{d_1}\rfloor K)}$, for any bandit strategy that returns the arm $\hat{k}$,

$$\begin{aligned}
\max_{2\leq i\leq K} \mathbb{P}_{\mathcal{G}^i}(\hat{k} \neq i) \geq &\frac{1}{6}\exp\left(-240\frac{C}{\tilde{H}^{det}_{1,\ell=1}(Q^{(1)})}\right) \\
= &\frac{1}{6}\exp\left(-480\frac{C}{K\tilde{H}^{det}_{2,\ell=1}(Q^{(1)})}\right) \\
\geq &\frac{1}{6}\exp\left(-480\frac{C}{K\tilde{H}^{det}_{2,\ell=1}(Q^{(i)})}\right).
\end{aligned}$$

That means we should **never** expect to use the right side of 59 as a general upper bound.

C Details on the Numerical Experiment Set-ups

In what follows, we provide details about how the numerical experiments are run. All the numerical experiments were run on the Kaggle servers.

C.1 Single Resource, i.e, $L = 1$

The details about Figure 2 is as follows. Firstly, the bars in the plot are more detailedly explained as follows: From left to right, the 1st blue column is matching high reward and high consumption, considering deterministic resource consumption. The 2nd orange column is matching high reward and low consumption, considering deterministic consumption. The 3rd green column is matching high reward and high consumption, considering correlated reward and consumption. The 4th red column is matching high reward and low consumption, considering correlated reward and consumption. The 5th purple column is matching high reward and high consumption, considering uncorrelated reward and consumption. The 6th brown column is matching high reward and low consumption, considering uncorrelated reward and consumption.

Next, we list down the detailed about the setup.

1. One group of suboptimal arms, High match High
   $r_1 = 0.9$; $r_i = 0.8, i = 2, \cdots, 256$; $d_{\ell=1,i} = 0.9, i = 1, \cdots, 128$; $d_{\ell=1,i} = 0.1, i = 129, \cdots, 256$
2. One group of suboptimal arms, High match Low
   $r_1 = 0.9$; $r_i = 0.8, i = 2, \cdots, 256$; $d_{\ell=1,i} = 0.1, i = 1, \cdots, 128$; $d_{\ell=1,i} = 0.9, i = 129, \cdots, 256$
3. Trap, High match High
   $r_1 = 0.9$; $r_i = 0.8, i = 2, \cdots, 32$; $r_i = 0.1, i = 33, \cdots, 256$; $d_{\ell=1,i} = 0.9, i = 1, \cdots, 128$; $d_{\ell=1,i} = 0.1, i = 129, \cdots, 256$

4. Trap, High match Low

   $r_1 = 0.9$; $r_i = 0.8, i = 2, \cdots, 32$; $r_i = 0.1, i = 33, \cdots, 256$; $d_{\ell=1,i} = 0.1, i = 1, \cdots, 128$; $d_{\ell=1,i} = 0.9, i = 129, \cdots, 256$

5. Polynomial, High match High

   $r_1 = 0.9, r_i = 0.9(1 - \sqrt{\frac{i}{256}}), i \geq 2$. $d_{\ell=1,i} = 0.9, i = 1, \cdots, 128$; $d_{\ell=1,i} = 0.1, i = 129, \cdots, 256$

6. Polynomial, High match Low

   $r_1 = 0.9, r_i = 0.9(1 - \sqrt{\frac{i}{256}}), i \geq 2$. $d_{\ell=1,i} = 0.1, i = 1, \cdots, 128$; $d_{\ell=1,i} = 0.9, i = 129, \cdots, 256$

7. Geometric, High match High

   $r_1 = 0.9, r_{256} = 0.1$, $\{r_i\}_{i=1}^{256}$ is geometric, $r_i = 0.9 * (\frac{1}{9})^{\frac{i-1}{255}}$. $d_{\ell=1,i} = 0.9, i = 1, \cdots, 128$; $d_{\ell=1,i} = 0.1, i = 129, \cdots, 256$

8. Geometric, High match Low

   $r_1 = 0.9, r_{256} = 0.1$, $\{r_i\}_{i=1}^{256}$ is geometric, $r_i = 0.9 * (\frac{1}{9})^{\frac{i-1}{255}}$. $d_{\ell=1,i} = 0.1, i = 1, \cdots, 128$; $d_{\ell=1,i} = 0.9, i = 129, \cdots, 256$

There are three kinds of consumption.

1. Deterministic Consumption. The consumption of each arm are deterministic.
2. Uncorrelated Consumption. When we pull an arm, the consumption and reward follow Bernoulli Distribution and are independent.
3. Correlated Consumption. When we pull the arm i, the consumption is $D_{\ell=1,i} = \mathbb{1}(U \leq d_{\ell=1,i})$, $D_{\ell=2,i} = \mathbb{1}(U \leq d_{\ell=2,i})$, $R = \mathbb{1}(U \leq r_i)$, where U follows uniform distribution on $[0, 1]$

C.2 Multiple Resources

Similarly, in multiple resources cases, we still considered different setups of mean reward, consumption, and consumption setups.

1. One group of suboptimal arms, High match High

   $r_1 = 0.9$; $r_i = 0.8, i = 2, \cdots, 256$; $d_{\ell=1,i} = 0.9, i = 1, \cdots, 128$; $d_{\ell=1,i} = 0.1, i = 129, \cdots, 256$; $d_{\ell=2,i} = 0.9, i = 1, \cdots, 128$; $d_{\ell=2,i} = 0.1, i = 129, \cdots, 256$

2. One group of suboptimal arms, Mixture

   $r_1 = 0.9$; $r_i = 0.8, i = 2, \cdots, 256$; $d_{\ell=1,i} = 0.1, i = 1, \cdots, 128$; $d_{\ell=1,i} = 0.9, i = 129, \cdots, 256$; $d_{\ell=2,i} = 0.9, i = 1, \cdots, 128$; $d_{\ell=2,i} = 0.1, i = 129, \cdots, 256$

3. One group of suboptimal arms, High match Low

   $r_1 = 0.9$; $r_i = 0.8, i = 2, \cdots, 256$; $d_{\ell=1,i} = 0.1, i = 1, \cdots, 128$; $d_{\ell=1,i} = 0.9, i = 129, \cdots, 256$; $d_{\ell=2,i} = 0.1, i = 1, \cdots, 128$; $d_{\ell=2,i} = 0.9, i = 129, \cdots, 256$

4. Trap, High match High

   $r_1 = 0.9$; $r_i = 0.8, i = 2, \cdots, 32$; $r_i = 0.1, i = 33, \cdots, 256$; $d_{\ell=1,i} = 0.9, i = 1, \cdots, 128$; $d_{\ell=1,i} = 0.1, i = 129, \cdots, 256$; $d_{\ell=2,i} = 0.9, i = 1, \cdots, 128$; $d_{\ell=2,i} = 0.1, i = 129, \cdots, 256$;

5. Trap, Mixture

   $r_1 = 0.9$; $r_i = 0.8, i = 2, \cdots, 32$; $r_i = 0.1, i = 33, \cdots, 256$; $d_{\ell=1,i} = 0.1, i = 1, \cdots, 128$; $d_{\ell=1,i} = 0.9, i = 129, \cdots, 256$; $d_{\ell=2,i} = 0.9, i = 1, \cdots, 128$; $d_{\ell=2,i} = 0.1, i = 129, \cdots, 256$;

6. Trap, High match Low

   $r_1 = 0.9$; $r_i = 0.8, i = 2, \cdots, 32$; $r_i = 0.1, i = 33, \cdots, 256$; $d_{\ell=1,i} = 0.1, i = 1, \cdots, 128$; $d_{\ell=1,i} = 0.9, i = 129, \cdots, 256$; $d_{\ell=2,i} = 0.1, i = 1, \cdots, 128$; $d_{\ell=2,i} = 0.9, i = 129, \cdots, 256$

7. Polynomial, High match High

   $r_1 = 0.9, r_i = 0.9(1 - \sqrt{\frac{i}{256}}), i \geq 2$. $d_{\ell=1,i} = 0.9, i = 1, \cdots, 128$; $d_{\ell=1,i} = 0.1, i = 129, \cdots, 256$; $d_{\ell=2,i} = 0.9, i = 1, \cdots, 128$; $d_{\ell=2,i} = 0.1, i = 129, \cdots, 256$

8. Polynomial, Mixture

   $r_1 = 0.9, r_i = 0.9(1 - \sqrt{\frac{i}{256}}), i \geq 2$. $d_{\ell=1,i} = 0.1, i = 1, \cdots, 128$; $d_{\ell=1,i} = 0.9, i = 129, \cdots, 256$; $d_{\ell=2,i} = 0.9, i = 1, \cdots, 128$; $d_{\ell=2,i} = 0.1, i = 129, \cdots, 256$

9. Polynomial, High match Low

   $r_1 = 0.9, r_i = 0.9(1 - \sqrt{\frac{i}{256}}), i \geq 2$. $d_{\ell=1,i} = 0.1, i = 1, \cdots, 128$; $d_{\ell=1,i} = 0.9, i = 129, \cdots, 256$; $d_{\ell=2,i} = 0.1, i = 1, \cdots, 128$; $d_{\ell=2,i} = 0.9, i = 129, \cdots, 256$

10. Geometric, High match High

    $r_1 = 0.9, r_{256} = 0.1$, $\{r_i\}_{i=1}^{256}$ is geometric, $r_i = 0.9 * (\frac{1}{9})^{\frac{i-1}{255}}$. $d_{\ell=1,i} = 0.9, i = 1, \cdots, 128$; $d_{\ell=1,i} = 0.1, i = 129, \cdots, 256$; $d_{\ell=2,i} = 0.9, i = 1, \cdots, 128$; $d_{\ell=2,i} = 0.1, i = 129, \cdots, 256$;

11. Geometric, Mixture

    $r_1 = 0.9, r_{256} = 0.1$, $\{r_i\}_{i=1}^{256}$ is geometric, $r_i = 0.9 * (\frac{1}{9})^{\frac{i-1}{255}}$. $d_{\ell=1,i} = 0.1, i = 1, \cdots, 128$; $d_{\ell=1,i} = 0.9, i = 129, \cdots, 256$; $d_{\ell=2,i} = 0.9, i = 1, \cdots, 128$; $d_{\ell=2,i} = 0.1, i = 129, \cdots, 256$;

12. Geometric, High match Low

    $r_1 = 0.9, r_{256} = 0.1$, $\{r_i\}_{i=1}^{256}$ is geometric, $r_i = 0.9 * (\frac{1}{9})^{\frac{i-1}{255}}$. $d_{\ell=1,i} = 0.1, i = 1, \cdots, 128$; $d_{\ell=1,i} = 0.9, i = 129, \cdots, 256$; $d_{\ell=2,i} = 0.1, i = 1, \cdots, 128$; $d_{\ell=2,i} = 0.9, i = 129, \cdots, 256$

There are two kinds of consumption.

1. Uncorrelated Consumption. When we pull an arm, the consumption and reward follow Bernoulli Distribution and are independent.

2. Correlated Consumption. When we pull the arm i, the consumption is $D_{\ell=1,i} = \mathbb{1}(U \leq d_{\ell=1,i})$, $D_{\ell=2,i} = \mathbb{1}(U \leq d_{\ell=2,i})$, $R = \mathbb{1}(U \leq r_i)$, where U follows uniform distribution on $[0, 1]$

C.3 Detailed Setting of Real-World Dataset

We adopted K Nearest Neighbour, Logistic Regression, Random Forest, and Adaboost as our candidates for the classifiers. And we applied each combination of machine learning model and its hyper-parameter to each supervised learning task with 500 independent trials. We identified the combination with the lowest empirical mean cross-entropy as the best arm.

Our BAI experiments were conducted across 100 independent trials. During each arm pull in a BAI experiment round—i.e., selecting a machine learning model with a specific hyperparameter combination—we partitioned the datasets randomly into training and testing subsets, maintaining a testing fraction of 0.3. The training subset was utilized to train the machine learning models, and the cross-entropy computed on the testing subset served as the realized reward. We flattened the 2-D image as a vector if the dataset is consists of images. All the experiments are deploied on the Kaggle Server with default CPU specifications.

The details of the real-world datasets we used are as follows

- To classify labels 3 and 8 in part of the MNIST Dataset. (MNIST 3&8)

  Number of label 3: 1086, Number of label 8: 1017, Number of Atrributes: 784.

  Time budget: 60 seconds.

  Link of dataset: `https://www.kaggle.com/competitions/digit-recognizer`.

- Optical Recognition of Handwritten Digits Data Set. (Handwritten)

  Number of Instances: 3823, Number of Attributes: 64

  Time budget: 60 seconds.

  Link of dataset: `https://archive.ics.uci.edu/ml/datasets/optical+recognition+of+handwritten+digits`

- To classify labels -1 and 1 in the MADELON dataset. (MADELON)

  Number of Instances: 2000, Number of Attributes: 500.

  Time budget: 80 seconds.

  Link of dataset: `https://archive.ics.uci.edu/ml/datasets/Madelon`.

- To classify labels -1 and 1 in the Arcene dataset. (Arcene)

  Number of Instances: 200, Number of Attributes:10000

  Time budget: 150 seconds.

  Link of dataset: `https://archive.ics.uci.edu/ml/datasets/Arcene` (Arcene)

- To classify labels of weight conditions in the Obesity dataset. (Obesity)

  Number of Instances: 2111, Number of Attributes: 16.

  Time budget: 20 seconds.

  Link of dataset: `https://archive.ics.uci.edu/dataset/544/estimation+of+obesity+levels+based+on+eating+habits+and+physical+condition`.

The machine learning models and candidate hyperparameters, aka arms, are as follows. We implemented all these models through the scikit-learn package in https://scikit-learn.org/stable/index.html

- K Nearest Neighbour
  - n_neighbours = 5, 15, 25, 35, 45, 55, 65, 75
- Logistic Regression
  - Regularization = "l2" or None
  - Intercept exists or not exists
  - Inverse value of regularization coefficient= 1, 2
- Random Forest
  - Fix max_depth = 5
  - n_estimators = 10, 20, 30, 50
  - criterion = "gini" or "entropy"
- Adaboost
  - n_estimators = 10, 20, 30, 40
  - learning rate = 1.0, 0.1