Proceedings of the 2023 Winter Simulation Conference
C. G. Corlu, S. R. Hunter, H. Lam, B. S. Onggo, J. Shortle, and B. Biller, eds.

BEST ARM IDENTIFICATION WITH FAIRNESS CONSTRAINTS ON SUBPOPULATIONS

Yuhang Wu
Zeyu Zheng
Tingyu Zhu

Department of Industrial Engineering and Operations Research
University of California, Berkeley
Berkeley, CA 94720, USA

ABSTRACT

We formulate, analyze and solve the problem of best arm identification with fairness constraints on subpopulations (BAICS). Standard best arm identification problems aim at selecting an arm that has the largest expected reward where the expectation is taken over the entire population. The BAICS problem requires that a selected arm must be fair to all subpopulations (e.g., different ethnic groups, age groups, or customer types) by satisfying constraints that the expected reward conditional on every subpopulation needs to be larger than some thresholds. The BAICS problem aims at correctly identify, with high confidence, the arm with the largest expected reward from all arms that satisfy subpopulation constraints. We analyze the complexity of the BAICS problem by proving a best achievable lower bound on the sample complexity with closed-form representation. We then design an algorithm and prove the sample complexity to match with the lower bound in terms of order. A brief account of numerical experiments are conducted to illustrate the theoretical findings.

1 INTRODUCTION

Many decision making problems naturally give rise to settings where there are a number of different policies (or systems, designs), each with unknown expected performances, from which the decision maker wants to select the policy with the best expected performance. The decision maker generally has access to observe independent noisy samples of the expected performance of each policy. The statistically principled way of identifying the best policy through the noisy samples has been a fundamental research topic in several research areas. Some early statistical works include Bechhofer (1954) and Bechhofer et al. (1995). In the stochastic simulation literature, the research problem is called *ranking and selection* (R&S); see Hong et al. (2021), Hunter and Nelson (2017), Chick (2006) and Kim and Nelson (2006) for reviews. In the multi-armed bandit literature, the research problem is called *best arm identification* (BAI); see Audibert et al. (2010), Garivier and Kaufmann (2016), Kaufmann et al. (2016), for references. Ma and Henderson (2017) and Glynn and Juneja (2015) have discussed some connections between the two literature. The R&S literature and BAI literature differ in assumptions and analysis tools. Our work is positioned in both literature, and adopts the assumptions and analysis tools in the BAI literature.

In this work, we consider the problem of Best Arm Identification with fairness Constraints on Subpopulations (BAICS). We briefly discuss the problem setting of BAICS and the meaning of fairness constraints on subpopulations. The formal setting with precise mathematical formulation is introduced in Section 2. In BAICS, each arm represents a policy to be used on an entire population, and there are multiple arms in competition. The expected reward of an arm is typically measured on the entire population, and the classical BAI problem aims at identifying the arm with the largest expected reward. However, for some applications, the population consists of several subpopulations. For example, a subpopulation may

979-8-3503-6966-3/23/$31.00 ©2023 IEEE

represent an ethnic group defined through a specific cultural background, or a group of customers defined through a specific consumption need. Fairness constraints on subpopulations refer to that the expected reward of an arm conditional on any subpopulation cannot be lower than some pre-specified threshold. Such fairness constraints imply that a policy is required to be "fair" to all subpopulations and is not allowed to "sacrifice" any subpopulation. Given the constraints on subpopulations, the set of arms are classified as feasible (satisfying the constraints) and infeasble (not satisfying the constraints). The BAICS problem aims at selecting an arm that has the largest expected reward among all the feasible arms.

The BAICS problem and its formulation have direct practical relevance when the decision maker cares about not only the expected total reward, but also the benefit of each subpopulation. One specific example is selecting the best updating plan for a system, under the constraint that the updating plan improves or at least maintains the user experience of customers of all age groups. In particular, the presence of fairness constraints prohibits a decision maker to improve expected reward over the entire population by implicitly exploiting or hurting some subpopulation. The BAICS problem formulation also has relevance to the online controlled experiments (A/B tests) with multiple treatments, where the goal is to select the treatment that provide Pareto improvement to all subpopulations.

We make the following contributions in this work.

- To our best of knowledge, we are the first to consider the fairness constraints of subpopulations in the context of the best arm identification problem with fixed confidence criterion, and we propose a new formulation called Best Arm Identification with fairness Constraints on Subpopulations (BAICS), which incorporates subpopulational fairness constraints into the arm selection process.
- We derive the asymptotic lower bound on the expected stopping time for all algorithms that are guaranteed to solve the problem with a given confidence level. Such lower bounds provide the best achievable sample complexity order for any algorithm that tackles the BAICS problem. We present an explicit formula along with an intuitive interpretation of the sample complexity.
- We design an algorithm that is capable of serving two goals — to identify the best arm and to ensure that it satisfies all subpopulation constraints. We provide theoretical results to show that it achieves the asymptotically optimal sample complexity. We compare our algorithm with two other methods and illustrate its efficiency through numerical experiments.

The theoretical tools that we develop in this work to analyze the lower and upper bounds on the expected sample complexity are partially inspired by the analysis framework proposed in Garivier and Kaufmann (2016) to address the standard BAI problem. In the BAI literature, there are works incorporating other constraints, such as safety constrains (Wang et al. (2022)) and variance constraints (Hou et al. (2022)). To the best of our knowledge, there are no works specifically considering constraints on the arm/policy performances on each subpopulation. Related works in the R&S literature consider the constrained R&S problem; see Andradóttir and Kim (2010), Healey et al. (2014) and Hong et al. (2015) for example. Tsai et al. (2018), He and Kim (2019) and Shi et al. (2022) consider feasibility determination in the R&S framework, and is related to the part of our problem on determining whether an arm satisfies subpopulation constraints (the other part of our problem focuses on selecting the best arm). The analysis framework and tools of the aforementioned works do not focus on developing matching lower and upper bounds for the sampling complexity.

The framework of Garivier and Kaufmann (2016) for the standard BAI problem does not require the outcome distribution of arms to be Gaussian. In this paper, the Gaussian assumption is adopted for simplicity, and we will demonstrate in future works that our results can be generalized to single-parameter exponential families. Further, in the R&S literature, the related selection problem is studied in the setting with unknown variances; see Hong et al. (2021) for a thorough review that includes ranking and selection problems with unknown variance and non-Gaussian distribution.

2 SETTING AND FORMULATION

The mathematical formulation of Best Arm Identification with fairness Constraints on Subpopulations (BAICS) is given as follows. Suppose we have the number of arms $K \geq 2$, the number of subpopulations L, and a vector $\mathbf{q} = (q_1, \cdots, q_L) \in \mathbb{R}^L$ representing the importance of the subpopulations. A typical choice in practice is to take q_l as the proportion of subpopulation l in the total population for $1 \leq l \leq L$. We further make the stochastic assumption that observations from arm k and subpopulation l are i.i.d. random variables drawn from Gaussian distributions with some known variance, and the variances are the same for all k and l. Without loss of generality we assume the variance is 1, so observation of arm k and subpopulation l is given by a normal distribution $P_{\mu_{k,l}} \sim \mathcal{N}(\mu_{k,l}, 1)$ for $k \in [K]$ and $l \in [L]$, here $[K] = \{1, \cdots, K\}$ and $[L] = \{1, \cdots, L\}$. Such distribution assumptions are commonly seen in the best arm identification literature, e.g., Shang et al. (2020), Barrier et al. (2022). The assumption may also be viewed as a special case of the exponential family distribution assumption with one unknown parameter (mean). The quality, or expected performance, of arm k is $\mu_k = \sum_{l=1} q_l \mu_{k,l}$, which is the weighted average of the means of the arm in different subpopulations.

A standard best arm identification problem tends to find the arm k_{BAI} with the maximum quality, i.e.,

$$k_{\text{BAI}} = \arg\max_{k \in [K]} \mu_k.$$

However, arm k_{BAI} may perform bad on some subpopulations. As discussed in Section 1, we hope to find an arm k^* with the best quality among all arms that work well on some given subpopulations. Mathematically, we introduce the definition *feasible arm* as follows: an arm k is feasible if and only if it is in a *feasible set* C, i.e.,

$$k \in C = \{k \in [K] \big| \mu_{k,m} \geq 0, \forall m \in [M]\} \tag{1}$$

with some known M. Our problem can thus be formulated as finding the best arm k^* in the feasible set C:

$$k^* = \arg\max_{k \in C} \mu_k. \tag{2}$$

If $C = \emptyset$, we define $k^* = 0$.

At each step t, the algorithm selects an arm $A_t \in [K]$ and a subpopulation $I_t \in [L]$ based on previous outcomes. After that a sample is drawn from $P_{\mu_{A_t, I_t}}$, which becomes the observation X_t . This naturally defines a filtration generated by all information up to step t denoted $\mathscr{F}_t = \sigma(\{I_s, A_s, X_s\}_{s=1,2,\cdots,t})$. The algorithm then chooses A_{t+1}, I_{t+1} which is $\mathscr{F}_t$-measurable. We further define $N_{a,i}(t) = \sum_{s=1}^t \mathbb{1}(I_s = i, A_s = a)$ and $N_a(t) = \sum_{s=1}^t \mathbb{1}(A_s = a)$. In addition, since the distributions $\left(P_{\mu_{k,l}}\right)_{k \in [K], l \in [L]}$ are assumed to be Gaussian in $\mathscr{P} = \{P \big| P \sim \mathcal{N}(\mu, 1)\}$, we may hence identify any bandit instance with its matrix of means $\boldsymbol{\mu} \in \mathbb{R}^{K \times L}$. Simple calculation shows that the Kullback-Leibler divergence between two Guassian distribution $P \sim \mathcal{N}(\mu, 1)$ and $Q \sim \mathcal{N}(\nu, 1)$ is given by $\mathbf{KL}(P, Q) = \frac{1}{2}(\mu - \nu)^2$.

Denote by $\mathscr{S}$ a set of Gaussian bandit models such that, each bandit model $\boldsymbol{\mu}$ in $\mathscr{S}$ and the feasible set $C(\boldsymbol{\mu})$ satisfy either of the following: (1) There is a unique optimal feasible arm with all constraints strictly satisfied, i.e., $\exists\ k^*(\boldsymbol{\mu}) \in C(\boldsymbol{\mu})$ such that $\mu_{k^*(\boldsymbol{\mu})} > \max_{k \in C}\{\mu_k \big| k \neq k^*(\boldsymbol{\mu})\}$ and $\mu_{k^*(\boldsymbol{\mu}), l} > 0$ for $l \in [M]$. (2) All arms have at least one subpopulation constraint strictly violated, i.e., $C(\boldsymbol{\mu}) = \emptyset$ and $k^*(\boldsymbol{\mu}) = 0$. Note that $k^*(\boldsymbol{\mu})$ is always unique for $\boldsymbol{\mu} \in \mathscr{S}$.

In this paper, we focus on the *fixed-confidence setting* with risk level δ. An algorithm is called δ-PAC if it gives a stopping time τ_δ with respect to $\mathscr{F}_t$, a $\mathscr{F}_{\tau_\delta}$-measurable recommendation $\hat{k}_{\tau_\delta} \in \{0\} \cup [K]$, and

$$\begin{aligned} &\forall \boldsymbol{\mu} \in \mathscr{S}, \mathbb{P}_{\boldsymbol{\mu}}(\tau_\delta < +\infty) = 1, \\ &\mathbb{P}_{\boldsymbol{\mu}}(\hat{k}_{\tau_\delta} \neq k^*(\boldsymbol{\mu})) \leq \delta. \end{aligned} \tag{3}$$

3 LOWER BOUNDS ON THE SAMPLE COMPLEXITY

In this section, we prove and analyze lower bounds on the sample complexity of δ-PAC algorithms for the BAICS problem. The lower bounds represent the best achievable sample complexity for any algorithm that can return the correct solution with the δ-PAC guarantee.

3.1 General Sample Complexity of Best Arm Identification

First, we introduce

$$\mathrm{Alt}(\boldsymbol{\mu}) := \{\boldsymbol{\lambda} \in \mathscr{S} \big| k^*(\boldsymbol{\mu}) \neq k^*(\boldsymbol{\lambda})\}, \tag{4}$$

the set of bandit models where the optimal feasible arm is not the same as in $\boldsymbol{\mu}$, and $\Sigma_{K\times L} = \{\mathbf{w} \in (\mathbb{R}_+ \cup \{0\})^{K\times L} \big| w_1 + \cdots + w_{KL} = 1\}$ the set of probability distributions on $[K]\times[L]$. We have following lower bound for the sample complexity.

Theorem 1 Let $\delta \in (0,1)$ and $\mathbf{q} \in \mathbb{R}^L$. For any δ-PAC policy and any bandit model $\boldsymbol{\mu} \in \mathscr{S}$,

$$\mathbb{E}_{\boldsymbol{\mu}}[\tau_\delta] \geq T^*(\boldsymbol{\mu})\mathrm{kl}(\delta, 1-\delta) \text{ and } \liminf_{\delta\to 0} \frac{\mathbb{E}_{\boldsymbol{\mu}}[\tau_\delta]}{\ln(1/\delta)} \geq T^*(\mu), \tag{5}$$

where

$$T^*(\boldsymbol{\mu})^{-1} = \frac{1}{2} \sup_{\mathbf{w}\in\Sigma_{K\times L}} \inf_{\boldsymbol{\lambda}\in\mathrm{Alt}(\boldsymbol{\mu})} \sum_{k\in[K]} \sum_{l\in[L]} w_{k,l} \left(\mu_{k,l} - \lambda_{k,l}\right)^2, \tag{6}$$

and $\mathrm{kl}(\delta, 1-\delta)$ is the KL divergence of two Bernoulli distributions of parameter δ and $1-\delta$.

The proof of Theorem 1 can be directly adapted from Theorem 1 of Russac et al. (2021) by noting that $\mathbf{KL}(P_{\mu_{k,l}}, P_{\lambda_{k,l}}) = \frac{(\mu_{k,l}-\lambda_{k,l})^2}{2}$. Here, $T^*(\boldsymbol{\mu})$ characterizes the difficulty of the problem, and $\mathbf{w}^*$ that achieves the supreme of (6) can be intuitively understood as the optimal sampling proportions of total samples for each arm and subpopulation. We will also see in next subsection that, the specific structure of $\mathrm{Alt}(\boldsymbol{\mu})$ in the BAICS problem makes $T^*(\boldsymbol{\mu})$ in our problem essentially different from that in the traditional BAI problem.

3.2 Implicit Tradeoff in BAICS Problem

We now focus specifically on our BAICS problem and demonstrate the tradeoff in sampling strategy that arises naturally. To provide an intuition for this tradeoff, we begin with a simple example.

Example 1: Suppose we have $K = 3, L = 2, M = 2$ and $\mathbf{q} = (\frac{1}{2}, \frac{1}{2})$. $\mu_{1,1} = \mu_{1,2} = 1$; $\mu_{2,1} = 4, \mu_{2,2} = -\varepsilon$; and $\mu_{3,1} = 1$, $\mu_{3,2} = 1-\varepsilon$, with some $0 < \varepsilon < 1$. Without constraints, it is easy to see that the BAI problem has $k_{\mathrm{BAI}} = 2$, because $\mu_1 = \frac{1}{2}\mu_{1,1} + \frac{1}{2}\mu_{1,2} = 1$, $\mu_2 = \frac{1}{2}\mu_{1,1} + \frac{1}{2}\mu_{2,2} = 2 - \frac{\varepsilon}{2}$ and $\mu_3 = \frac{1}{2}\mu_{3,1} + \frac{1}{2}\mu_{3,2} = 1 - \frac{\varepsilon}{2}$. Specifically, since the difference between the means of arms 2 and 1 is $\mu_2 - \mu_1 = 1 - \frac{\varepsilon}{2}$, and the difference between the means of arms 2 and 3 is $\mu_2 - \mu_3 = 1$, the gap between the best arm and the other arms is relatively large when ε is much smaller than 1. However, when we consider subpopulation constraints in the BAICS problem, the best arm is now $k^* = 1$. To identify $k^* = 1$, we must discover that $\mu_{2,2} < 0$ because $\mu_2 > \mu_1$, and that $\mu_3 < \mu_1$. When ε goes to 0, these can be substantially more difficult than BAI because both of above gaps are ε.

The simple example above highlights the fundamental difference between BAI and BAICS. In the BAI problem, explorations are used to find the arm with the highest mean. However, in the BAICS problem, explorations introduce an implicit tradeoff between optimality and feasibility. In Example 1 where ε is small, to conclude that $k^* = 1$, people need to estimate μ_1, $\mu_{2,2}$, and μ_3 accurate enough, which leads to a natural problem of allocating samples among arm 1 and arm 3 for optimality, and subpopulation 2 of arm 2 for feasibility. We also point out that Example 1 does not mean BAICS is always more difficult than BAI. In fact, if we slightly change the setting to $\mu_{2,2} = -2-\varepsilon, \mu_{3,2} = -1$ and keep others the same as in Example 1, then it is easy to see $k^* = k_{\mathrm{BAI}} = 1$. This time BAICS is easy because we can easily tell $\mu_{2,2}, \mu_{3,2} < 0$ and arm 1 is feasible, but BAI is hard because $\mu_1 - \mu_2 = \frac{\varepsilon}{2}$ is small when ε is close to 0. We can gain a deeper understanding of the BAICS problem by considering another variant of Example 1. This time, we change $\mu_{2,1} = 1$, $\mu_{3,2} = -1$ and keep others the same. Since now $\mu_1 = 1, \mu_2 = \frac{1-\varepsilon}{2}$ and $\mu_3 = 0$, it is again easy to identify $k^* = 1$. The interesting thing in this example is that it is not necessary to identify the feasible set C before we find $k^* = 1$. Indeed, it is possible that our algorithm can not tell whether $\mu_{2,2} \geq 0$ when ε is small, but it can still recommend $k^* = 1$ with risk at most δ because μ_2 is much smaller than μ_1 and we do not need to know the feasibility of arm 2. In this example, a naive algorithm that attempts to find the feasible set C before searching for the best arm in C can be inefficient in general, so the BAICS problem is not a straightforward synthesis of finding the feasible set and a standard BAI problem.

Above examples and discussions show that the complexity of a BAICS problem need not be related to the corresponding BAI problem without constraints, and the BAICS problem naturally leads to an optimality-feasibility tradeoff and presents unique challenges. We now formally state the theorem that captures this intuition. Based on the theorem, we derive in Section 4 an algorithm that solves the BAICS problem by solving a sequence of optimization subproblems with moderate computational complexity.

Theorem 2 For any $\boldsymbol{\mu} \in \mathscr{S}$, if $k^*(\boldsymbol{\mu}) = 0$,

$$T^*(\boldsymbol{\mu})^{-1} = \frac{1}{2} \max_{\mathbf{w} \in \Sigma_{K \times L}} \min_{k \in [K]} \sum_{l \in [M], \mu_{k,l} < 0} w_{k,l} \mu_{k,l}^2; \tag{7}$$

If $k^*(\boldsymbol{\mu}) \neq 0$, without loss of generality we assume $k^*(\boldsymbol{\mu}) = 1$, then

$$T^*(\boldsymbol{\mu})^{-1} = \frac{1}{2} \max_{\mathbf{w} \in \Sigma_{K \times L}} \min \left(f_{\boldsymbol{\mu}}^{\text{opt}}(\mathbf{w}), f_{\boldsymbol{\mu}}^{\text{fea}}(\mathbf{w}) \right), \tag{8}$$

where

$$f_{\boldsymbol{\mu}}^{\text{opt}}(\mathbf{w}) = \min_{2 \leq k \leq K} \min_{\substack{\boldsymbol{\lambda} \in \mathbb{R}^{K \times L} \\ \lambda_k \geq \lambda_1 \\ \forall l \in [M], \lambda_{k,l} \geq 0}} \left(\sum_{l \in [L]} w_{1,l} (\mu_{1,l} - \lambda_{1,l})^2 + \sum_{l \in [L]} w_{k,l} (\mu_{k,l} - \lambda_{k,l})^2 \right),$$

and

$$f_{\boldsymbol{\mu}}^{\text{fea}}(\mathbf{w}) = \min_{l \in [M]} w_{1,l} \mu_{1,l}^2.$$

The proof of Theorem 2 is given in Section 3.3. In Theorem 2, (7) gives the sample complexity lower bound when there is no feasible arm, i.e. $C(\boldsymbol{\mu}) = \emptyset$. As for the case $C(\boldsymbol{\mu}) \neq \emptyset$, the complexity of the problem $T^*(\boldsymbol{\mu})$ defined in (6) now consists of two terms $f_{\boldsymbol{\mu}}^{\text{opt}}(\mathbf{w})$ and $f_{\boldsymbol{\mu}}^{\text{fea}}(\mathbf{w})$, which reflect the credibility of optimality and the credibility of feasibility, respectively. Briefly, $f_{\boldsymbol{\mu}}^{\text{opt}}(\mathbf{w})$ can be interpreted as a measure of assurance that other feasible arms are not as good as arm 1, and $f_{\boldsymbol{\mu}}^{\text{fea}}(\mathbf{w})$ is a measure of assurance that arm 1 is feasible. The smaller these two values are, the less assurance and therefore the more difficult the problem becomes. The notion $\mathbf{w}$ is the proportions of samples for each arm and each subpopulation. $T^*(\boldsymbol{\mu})^{-1}$ is then obtained through maximizing the minimum of $f_{\boldsymbol{\mu}}^{\text{opt}}(\mathbf{w})$ and $f_{\boldsymbol{\mu}}^{\text{fea}}(\mathbf{w})$, which can be interpreted as a trade-off between minimizing the complexity of optimality and the complexity of feasibility.

3.3 Proof of Theorem 2

In this part, we discuss the proof of Theorem 2. The basic idea is to construct a close-by alternative bandit instance $\boldsymbol{\lambda} \in \text{Alt}(\boldsymbol{\mu})$ such that $\sum_{k \in [K]} \sum_{l \in [L]} w_{k,l} \left(\mu_{k,l} - \lambda_{k,l}\right)^2$ achieves infimum, for fixed $\mathbf{w} \in \Sigma_{K \times J}$.

Recall that

$$T^*(\boldsymbol{\mu})^{-1} = \frac{1}{2} \sup_{\mathbf{w} \in \Sigma_{K \times J}} \inf_{\boldsymbol{\lambda} \in \text{Alt}(\boldsymbol{\mu})} \sum_{k \in [K]} \sum_{l \in [L]} w_{k,l} \left(\mu_{k,l} - \lambda_{k,l}\right)^2.$$

First we consider the case $C(\boldsymbol{\mu}) = \emptyset$, then $k^*(\boldsymbol{\mu}) = 0$, so $\mathrm{Alt}(\boldsymbol{\mu}) = \{\boldsymbol{\lambda} \in \mathscr{S} \big| k^*(\boldsymbol{\lambda}) \neq 0\}$. That is, as long as $\boldsymbol{\lambda}$ has one feasible arm, then it is in the alternative set $\mathrm{Alt}(\boldsymbol{\mu})$. Fix $\mathbf{w}$, for any $i \in [K]$, to make sure $i \in C(\boldsymbol{\mu})$, we only require $\lambda_{i,l} \geq 0$ for all $l \in [M]$, so we only need to set $\lambda_{i,l} = 0$ for those $l \in [M]$ such that $\mu_{i,l} < 0$, and take other $\lambda_{k,l}$ just to be $\mu_{k,l}$, then

$$\inf_{\boldsymbol{\lambda} \in \mathrm{Alt}(\boldsymbol{\mu})} \sum_{k \in [K]} \sum_{l \in [L]} w_{k,l} \left(\mu_{k,l} - \lambda_{k,l}\right)^2 = \min_{k \in [K]} \sum_{l \in [M], \mu_{k,l} < 0} w_{k,l} \mu_{k,l}^2.$$

It is easy to see this is a continuous function of $\mathbf{w}$ and the domain of $\mathbf{w}$ is compact, so the supremum can be attained by some $\mathbf{w}^*(\boldsymbol{\mu})$, and we obtain (7).

As for the case $C(\boldsymbol{\mu}) \neq \emptyset$, without loss of generality we assume $k^*(\boldsymbol{\mu}) = 1$. Again we fix $\mathbf{w}$. To construct an alternative instance $\boldsymbol{\lambda} \in \mathrm{Alt}(\boldsymbol{\mu})$, we have two different ways. The first option consists in taking an arm $i > 1$ and augment means of its subpopulations on the alternative model such that it becomes above arm 1. Otherwise, it is possible to shrink the mean of one subpopulation of arm 1 such that it becomes infeasible on the alternative. We will now consider each of them separately.

For the first option, suppose we want to take arm i $(i \neq 1)$ and augment its means. Then in the alternative model $\boldsymbol{\lambda}$, we would expect $\lambda_k \geq \lambda_1$ and $\forall l \in [M], \lambda_{i,l} \geq 0$, and for other $k \neq 1, i$, we take $\lambda_{k,l} = \mu_{k,l}$ to minimize $\sum_{\substack{k \in [K] \\ k \neq 1, i}} \sum_{l \in [L]} w_{k,l} \left(\mu_{k,l} - \lambda_{k,l}\right)^2$ to be 0. Note here we only require $\lambda_k \geq \lambda_1$ because we can always add a small number to some $\lambda_{k,l}$ to make the inequality strict. Then in this case, we obtain

$$f_{\boldsymbol{\mu}}^{\mathrm{opt}}(\mathbf{w}) = \min_{2 \leq k \leq K} \min_{\substack{\boldsymbol{\lambda} \in \mathbb{R}^{K \times L} \\ \lambda_k \geq \lambda_1 \\ \forall l \in [M], \lambda_{k,l} \geq 0}} \big(\sum_{l \in [L]} w_{1,l} (\mu_{1,l} - \lambda_{1,l})^2 + \sum_{l \in [L]} w_{k,l} (\mu_{k,l} - \lambda_{k,l})^2 \big).$$

For the other way, we want to shrink the mean of one subpopulation of arm 1 to make it infeasible. Thus, we only need to modify $\lambda_{1,l}$ to be 0 for some $l \in [M]$ and set all other $\lambda_{k,l} = \mu_{k,l}$. Again we only need $\lambda_{1,l} = 0$ because we can subtract it by an arbitrarily small number to make it negative. Now by iterating over $l \in [M]$ we can define

$$f_{\boldsymbol{\mu}}^{\mathrm{fea}}(\mathbf{w}) = \min_{l \in [M]} w_{1,l} \mu_{1,l}^2.$$

Combine above two cases together and we obtain

$$\inf_{\boldsymbol{\lambda} \in \mathrm{Alt}(\boldsymbol{\mu})} \sum_{k \in [K]} \sum_{l \in [L]} w_{k,l} \left(\mu_{k,l} - \lambda_{k,l}\right)^2 = \min \left(f_{\boldsymbol{\mu}}^{\mathrm{opt}}(\mathbf{w}), f_{\boldsymbol{\mu}}^{\mathrm{fea}}(\mathbf{w}) \right).$$

It is easy to see $\min \left(f_{\boldsymbol{\mu}}^{\mathrm{opt}}(\mathbf{w}), f_{\boldsymbol{\mu}}^{\mathrm{fea}}(\mathbf{w}) \right)$ is continuous, so the supremum on a compact set can be replaced by the maximum, then

$$T^*(\boldsymbol{\mu})^{-1} = \frac{1}{2} \sup_{\mathbf{w} \in \Sigma_{K \times J}} \inf_{\boldsymbol{\lambda} \in \mathrm{Alt}(\boldsymbol{\mu})} \sum_{k \in [K]} \sum_{l \in [L]} w_{k,l} \left(\mu_{k,l} - \lambda_{k,l}\right)^2 = \frac{1}{2} \max_{\mathbf{w} \in \Sigma_{K \times L}} \min \left(f_{\boldsymbol{\mu}}^{\mathrm{opt}}(\mathbf{w}), f_{\boldsymbol{\mu}}^{\mathrm{fea}}(\mathbf{w}) \right).$$

which finishes the proof.

4 ALGORITHM DESIGN AND COMPLEXITY ANALYSIS

In this section, we develop an algorithm to solve the BAICS problem with δ-PAC guarantee. We prove upper bound on the sample complexity of the proposed algorithm. We show that the upper bound matches the proved lower bound in the order.

To develop our algorithm, we adapt the Track-and-Stop algorithm introduced in Garivier and Kaufmann (2016) to the BAICS problem. We first discuss the sampling rule and its calculation. Then we give the stopping rule, our recommendation of the best feasible arm, and the threshold for stopping. Finally, we give the convergence result of our algorithm to show it is asymptotically optimal in the sense that $\mathbb{E}[\tau_\delta]$ matches the sample complexity lower bound asymptotically as $\delta \to 0$.

4.1 The Sampling Rule and its Calculation

In this part, we first give a high level overview of the sampling rule and then give the details of our implementation. Suppose we are given the number of arms K, subpopulations L, constraints M and weights $\mathbf{q}$. In each round $t = 1, 2, \cdots$, the algorithm first computes the empirical means of all arms and all subpopulations, denoted by $\hat{\boldsymbol{\mu}}(t) \in \mathbb{R}^{KL}$, which is given by $\hat{\mu}_{k,l}(t) = \frac{1}{N_{k,l}(t)} \sum_{s=1}^{t} X_s \mathbb{1}(I_s = l, A_s = k)$. Then the algorithm computes a maximizer $\mathbf{w}_t \in \mathbf{w}^*(\hat{\boldsymbol{\mu}}(t))$ of problem (6) with $\boldsymbol{\mu}$ replaced by $\hat{\boldsymbol{\mu}}(t)$, which can further be simplified to (7) or (8). Here, since the maximizer may not be unique, so $\mathbf{w}^*(\hat{\boldsymbol{\mu}}(t))$ is defined as the set consists of all maximizers, and we can take $\mathbf{w}_t$ to be any element in $\mathbf{w}^*(\hat{\boldsymbol{\mu}}(t))$. Now, we use the C-tracking rule proposed by Garivier and Kaufmann (2016). To be specific, for some given $\varepsilon \in (0, \frac{1}{KL}]$, let $\mathbf{w}_t^{(\varepsilon)}$ be a L^∞ projection of $\mathbf{w}_t$ onto $\Sigma_{K\times L}^{\varepsilon} = \{(w_1, \cdots, w_{KL}) \in [\varepsilon, 1] \big| w_1 + \cdots + w_{KL} = 1\}$. Take $\varepsilon_t = \left(K^2L^2 + t\right)^{-\frac{1}{2}}/2$ and

$$(A_{t+1}, I_{t+1}) \in \arg\max_{(k,l)} \sum_{s=0}^{t} w_{k,l}^{\varepsilon_s}(\hat{\boldsymbol{\mu}}(s)) - N_{k,l}(t). \tag{9}$$

Later we will see, this sampling rule ensures that $N_{k,l}(t)$ is close to $\sum_{s=0}^{t} w_{k,l}^{\varepsilon_s}(\hat{\boldsymbol{\mu}}(s))$ and thus close to $t w_{k,l}^*(\boldsymbol{\mu})$, so it is asymptotically optimal and can achieve the lower bound given by (5).

Now we talk about the calculation of our sampling rule. From above we can see the only challenging aspect is calculating $\mathbf{w}_t$, which is a good approximation of the maximizer of problem (6) with $\boldsymbol{\mu}$ replaced by $\hat{\boldsymbol{\mu}}(t)$. By Theorem 2, solving problem (6) can be further simplified into solving problems (7) and (8) with some given $\boldsymbol{\mu}$. For (7), it is not hard to see we would expect $\sum_{l\in[M], \mu_{k,l}<0} (\mathbf{w}_t)_{k,l} \mu_{k,l}^2$ to be the same for $k \in [K]$. Define $l(k) = \arg\max_{l\in[M], \mu_{k,l}<0} \mu_{k,l}^2$ and break the tie arbitrarily, then under the constraint $\sum_{k\in[K], l\in[L]} w_{k,l} = 1$ we can see

$$(\mathbf{w}_t)_{i,l(i)} = \frac{1}{\mu_{i,l(i)}^2} \left(\sum_{k\in[K]} \frac{1}{\mu_{k,l(k)}^2} \right)^{-1}$$

for $i \in [K]$ and $(\mathbf{w}_t)_{k,l} = 0$ for $l \neq l(k)$. Thus, (7) can be explicitly solved. The optimization problem in (8) is more difficult. We first define $F_{\boldsymbol{\mu}}(\mathbf{w}) = \min\left(f_{\boldsymbol{\mu}}^{\text{opt}}(\mathbf{w}), f_{\boldsymbol{\mu}}^{\text{fea}}(\mathbf{w})\right)$ and calculate $F(\mathbf{w})$ for fixed $\mathbf{w}$ as an optimization problem. Given $\mathbf{w}$, $f_{\boldsymbol{\mu}}^{\text{fea}}(\mathbf{w}) = \min_{l\in[M]} w_{1,l} \mu_{1,l}^2$ is known, so it suffices to calculate $f_{\boldsymbol{\mu}}^{\text{opt}}(\mathbf{w})$. Recall that

$$f_{\boldsymbol{\mu}}^{\text{opt}}(\mathbf{w}) = \min_{2\le k\le K} \min_{\substack{\boldsymbol{\lambda}\in\mathbb{R}^{K\times L} \\ \lambda_k \ge \lambda_1 \\ \forall l\in[M], \lambda_{k,l}\ge 0}} \left(\sum_{l\in[L]} w_{1,l}(\mu_{1,l} - \lambda_{1,l})^2 + \sum_{l\in[L]} w_{k,l}(\mu_{k,l} - \lambda_{k,l})^2 \right),$$

and for fixed $\mathbf{w}$ and each k, the internal minimization programming is a convex quadratic problem with linear constraints, so it can be easily solved through standard optimization methods, say the Lagrangian multiplier method used in Lemma 5 of Russac et al. (2021). Thus $f_{\boldsymbol{\mu}}^{\text{fea}}(\mathbf{w})$ can be calculated through solving $K-1$ optimization subproblems.

Now that $F_{\boldsymbol{\mu}}(\mathbf{w})$ is known, and it is the minimum of several linear functions of $\mathbf{w}$, so it is concave. In addition, from the definition of $F_{\boldsymbol{\mu}}(\mathbf{w})$ we know there exist $\boldsymbol{\lambda}, k \in [K]$ such that $\sum_{l\in[L]} w_{1,l}(\mu_{1,l} - \lambda_{1,l})^2 + \sum_{l\in[L]} w_{k,l}(\mu_{k,l} - \lambda_{k,l})^2 = F_{\boldsymbol{\mu}}(\mathbf{w})$ or there exists $L \in [M]$ such that $w_{1,l}\mu_{1,l}^2 = F_{\boldsymbol{\mu}}(\mathbf{w})$. In both cases we can write $F_{\boldsymbol{\mu}}(\mathbf{w}) = \mathbf{c}(\mathbf{w})^T \mathbf{w}$ for some $\mathbf{c}(\mathbf{w}) \in \mathbb{R}^{KL}$, so we can obtain a subgradient of $-F_{\boldsymbol{\mu}}(\mathbf{w})$ given by $-\mathbf{c}(\mathbf{w})$ by definition. Recall that $F_{\boldsymbol{\mu}}(\mathbf{w})$ can be written as the infimum of linear functions (each $\boldsymbol{\lambda} \in \text{Alt}(\boldsymbol{\mu})$ gives a linear function), so it is concave. Now, by performing projected subgradient method, we can solve the minimization problem $\max_{\mathbf{w}\in\Sigma_{K\times L}} F(\mathbf{w})$ by updating $w^{(n+1)} = \mathbf{P}_{\Sigma_{K\times L}}\left(\mathbf{w}^{(n)} + \alpha_n \mathbf{c}(\mathbf{w}^{(n)})\right)$ iteratively with the projection operator $\mathbf{P}_{\Sigma_{K\times L}}(\cdot)$ and some proper stepsizes $\{\alpha_n\}$. It is known the projected subgradient method converges under mild conditions, see for example Boyd et al. (2003). This finishes the calculation of $\mathbf{w}_t$ and also our sampling rule.

4.2 The Stopping Rule and the Threshold

Following the idea of Garivier and Kaufmann (2016) and Russac et al. (2021), we consider the Chernoff's Generalized Likelihood Ratio statistic:

$$Z(t) = \frac{1}{2} \inf_{\boldsymbol{\lambda} \in \mathrm{Alt}(\boldsymbol{\mu})} \sum_{k \in [K]} \sum_{l \in [L]} N_{k,l}(t) \left(\hat{\mu}_{k,l}(t) - \lambda_{k,l}\right)^2. \tag{10}$$

Note if we define the empirical sampling weights $(\hat{\mathbf{w}}_t)_{k,l} = \frac{N_{k,l}(t)}{t}$, then $Z(t)$ can be written as $Z(t) = \frac{t}{2} F_{\hat{\boldsymbol{\mu}}(t)}(\hat{\mathbf{w}}_t)$, which can be efficiently calculated as dicussed in Section 4.1. For a given risk level $0 < \delta < 1$, we define the stopping time τ_δ as follows:

$$\tau_\delta = \inf_{t \in \mathbb{N}} \{Z(t) > \beta(t, \delta)\}.$$

Here the threshold $\beta(t,\delta)$ should be tuned appropriately. By Proposition 21 of Kaufmann and Koolen (2021), a choice of $\beta(t,\delta) = O\left(L \ln\ln t + \log \frac{K}{\delta}\right)$ would ensure our policy to be δ-PAC, while in practice, as suggested by Garivier and Kaufmann (2016), we use instead the stylized $\ln((1+\ln t)/\delta)$ which is less conservative. The final recommendation $\hat{k}_{\tau_\delta}$ is just the optimal feasible arm in $\hat{\boldsymbol{\mu}}(\tau_\delta)$, i.e.

$$\hat{k}_{\tau_\delta} = \arg \max_{k \in C(\hat{\boldsymbol{\mu}}(\tau_\delta))} \hat{\mu}_k.$$

Again, we take $\hat{k}_{\tau_\delta} = 0$ if $C(\hat{\boldsymbol{\mu}}(\tau_\delta)) = \emptyset$.

4.3 The Convergence Result

We now give the convergence result of our algorithm, which matches the asymptotic optimal lower bound given by (5):

Theorem 3 For every bandit model $\boldsymbol{\mu} \in \mathscr{S}$, our algorithm is δ-PAC and

$$\lim_{\delta \to 0} \frac{\mathbb{E}_{\boldsymbol{\mu}}[\tau_\delta]}{\ln(1/\delta)} = T^*(\boldsymbol{\mu}). \tag{11}$$

The proof of Theorem 3 is given as follows. By applying Lemma 7 of Garivier and Kaufmann (2016) to our C-Tracking rule with KL weights, we have

$$\max_{k \in [K], l \in [L]} \left| N_{k,l}(t) - \sum_{s=0}^{t-1} \mathbf{w}_t \right| \le KL(1 + \sqrt{t}). \tag{12}$$

With force exploration rate ε_t, since $t\varepsilon_t = O(\sqrt{t})$, each subpopulation of each arm would be sampled infinite times as $t \to \infty$, so $\hat{\boldsymbol{\mu}}(t) \to \boldsymbol{\mu}$ almost surely. In addition, since $F_{\boldsymbol{\mu}}(\mathbf{w})$ is concave, so the set of maximizers $\mathbf{w}^*(\boldsymbol{\mu})$ is convex, then by Lemma 6 of Degenne and Koolen (2019) we know $\inf_{\mathbf{w} \in \mathbf{w}^*(\boldsymbol{\mu})} \|\frac{1}{t} \sum_{s=0}^{t-1} \mathbf{w}_t - \mathbf{w}\|_\infty \to 0$ as $t \to \infty$. Combine this with (12) we know that $\inf_{\mathbf{w} \in \mathbf{w}^*(\boldsymbol{\mu})} \left\| \hat{\mathbf{w}}_t - \mathbf{w} \right\|_\infty \to 0$ almost surely, that is, our empirical weights $\hat{\mathbf{w}}_t$ gets close to some oracle weights. Further, considering the numerical error γ_t of solving w_t from an optimization problem as discussed in section 4.1, we note that as long as $\|\gamma_t\|_\infty \to 0$, we have $\|\frac{1}{t} \sum_{s=0}^{t-1} \gamma_t\|_\infty \to 0$, which does not change the convergence result regarding sequences converging to w. Since $\boldsymbol{\mu} \in \mathscr{S}$ so the problem is single-answered, so by Theorem 7 of Degenne and Koolen (2019) we know our algorithm has asymptotically optimal complexity, i.e. $\lim_{\delta \to 0} \frac{\mathbb{E}[\tau_\delta]}{\ln(1/\delta)} = T^*(\boldsymbol{\mu})$. In addition, by our choice of $\beta(t,\delta)$ and Proposition 21 of Kaufmann and Koolen (2021), our algorithm is δ-PAC.

5 NUMERICAL EXPERIMENTS

In this section, we demonstrate the efficiency of the Track-and-Stop with fairness Constraints on Subpopulations (T-a-SCS) strategy for addressing BAICS problems through two examples. In the first example, the arm k of maximum quality is infeasible on one subpopulation l. More specifically, $\mu_{k,l} < 0$ but is close to 0. The second example presents a situation where two arms have maximum quality, but one of them is infeasible, further, there is a third arm which is feasible and has a quality close to the maximum quality. Through these examples, we demonstrate the behavior of the algorithm where there is a tradeoff between testing for optimality and testing for feasibility. In comparison with the T-a-SCS strategy, we consider two other benchmark sampling strategies. The first is the original Track-and-Stop (T-a-S) strategy (Garivier and Kaufmann (2016)), which does not incorporate subpopulation constraints when calculating the weight assignment for each arm. When this strategy is performed, it yields arm A_t to be sampled at iteration t. We then randomly allocate the sample to subpopulation I_t of arm A_t with probability $q_{I_t}/\sum_{l\in[L]} q_l$. The second is the uniform sampling strategy. In each iteration, we sample arm $k(t)$ with probability $1/K$, and randomly allocate the sample to subpopulation I_t of arm A_t with probability $q_{I_t}/\sum_{l\in[L]} q_l$. The choice of sampling strategy does not affect the stopping rule. In our experiment, all three algorithms use the Chernoff's Generalized Likelihood Ratio statistic $Z(t)$ given by (10), and the same threshold $\beta(t,\delta) = \ln((1+\ln t)/\delta)$.

5.1 The First Example

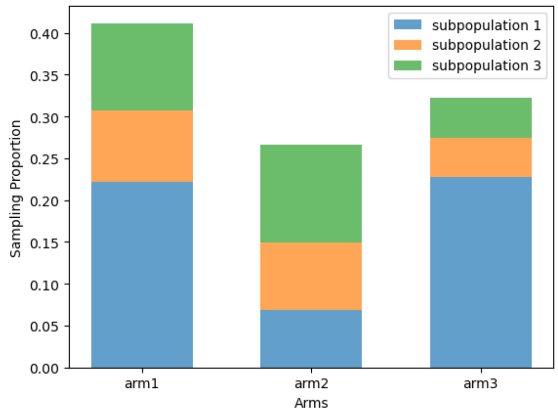
In the first numerical case, we set the number of arms and subpopulations, and the respective arm values on each subpopulation as follows: let $K = 3, L = 3$, $\mu_1 = (0.2, 0.6, 0.8)$, $\mu_2 = (0.4, 0.4, 0.3)$, $\mu_3 = (-0.2, 1, 1.5)$, we have noise level $\sigma = 1$. In calculating the overall quality of an arm, we set for the three subpopulations $q_1 = 0.2$, $q_2 = 0.3$, $q_3 = 0.5$, and $\mu_k = \sum_{l=1}^{3} q_k \mu_{k,l}$. In this case, we have $\mu_1 = 0.62$, $\mu_2 = 0.35$, $\mu_3 = 1.01$, but arm 3 is infeasible because $\mu_{31} < 0$, and arm 1 is the best feasible arm. The probability threshold of correct selection is set as $\delta = 0.1$.

We initialize each arm with 5 draws on each subpopulation. For simplicity, we perform projected subgradient method (see Section 4.1) to update $\mathbf{w}$ one time with stepsize $\alpha = 1$ in each iteration. The optimal weights $\mathbf{w}_t = (w_1, \cdots, w_K)$ for implementing the T-a-S strategy are calculated by solving a rational equation; we refer to the Gaussian case of Garivier and Kaufmann (2016) for details. We also use the C-tracking rule for projecting the T-a-S weights. We run 300 experiments to record the average stopping time $\hat{\tau}_\delta$ and empirical probability of correct selection $\hat{P}_{\boldsymbol{\mu}}(\hat{k}_{\tau_\delta} = k^*(\boldsymbol{\mu}))$. The results are given in table 1.

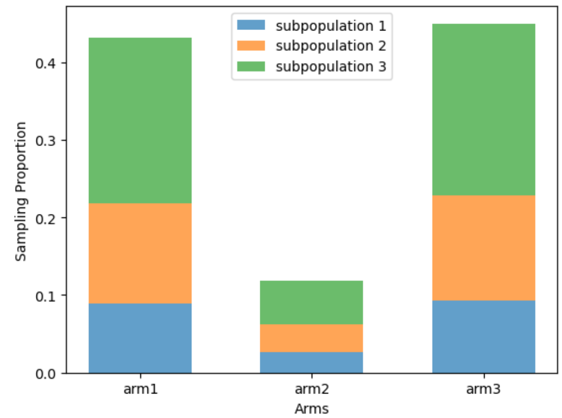
	T-a-SCS	T-a-S	Uniform
$\hat{\tau}_\delta$	530	1703	2432
$\hat{P}_{\boldsymbol{\mu}}$	0.987	0.990	0.983

Table 1: Average stopping time and empirical probability of correct selection of the three sampling strategies in example 1.

We further look at how the samples are allocated to each of the arms and subpopulations in the T-a-SCS strategy, in comparison with the T-a-S strategy. For each experiment, we record the number of samples on each arm and subpopulation, $\{N_{k,l}(\tau_\delta) : k \in [K], l \in [L]\}$, and compute the empirical sampling weights $(\hat{\mathbf{w}})_{k,l} = \frac{N_{k,l}(\tau_\delta)}{\tau_\delta}$. We then take an average over the 300 copies of experiments. The results are given in Figure 1, demonstrating a tradeoff between optimality and feasibility. Compared to the T-a-S strategy, we notice that the T-a-SCS strategy assigns more empirical sampling weights to the subpopulations on which the arm values are close to 0 (e.g., subpopulation 1 of arm 1, subpopulation 1 of arm 3). Further, as the infeasibility of arm 3 is "discovered" by T-a-SCS, it allocates more samples to arm 2 (compared to the T-a-S strategy), which is now the only competitor for arm 1 of being the best feasible arm.



(a) T-a-SCS



(b) T-a-S

Figure 1: Average sample allocation to the arms and the subpopulations in example 1.

5.2 The Second Example

In the second numerical case, we set the number of arms and subpopulations, and the respective arm values on each subpopulation as follows: let $K = 4$, $L = 3$, $\mu_1 = (-0.2, 0.4, 1.2)$, $\mu_2 = (0.2, 0.6, 0.6)$, $\mu_3 = (0.3, 0.3, 0.6)$, $\mu_4 = (-0.6, 0.8, 0.4)$, we have noise level $\sigma = 1$. In calculating the overall quality of an arm, we set for the three subpopulations $q_1 = q_2 = q_3 = 1/3$, and $\mu_k = \sum_{l=1}^{3} q_k \mu_{k,l}$. In this example we have $\mu_1 = \mu_2 = 0.47$, which equals the maximum quality, but arm 1 is infeasible. Also, μ_3 is 0.4, which is close to μ_2. The probability threshold of correct selection is set as $\delta = 0.1$.

The numerical settings are similar to that of Section 5.1. In this example, both the T-a-S and the Uniform sampling strategy exceed the limit of $\tau_{\max} = 15000$ iterations in a large proportion of experiment copies. For T-a-SCS we have $\hat{\tau}_\delta = 3131$ and $\hat{P}_{\boldsymbol{\mu}}(\hat{k}_{\tau_\delta} = k^*(\boldsymbol{\mu})) = 0.980$.

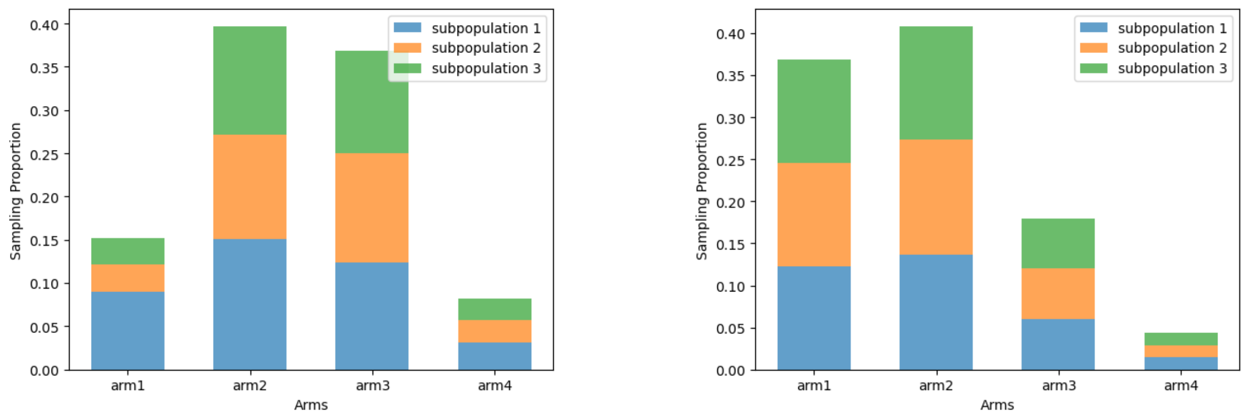
We further look at how the samples are allocated to each of the arms and subpopulations in the T-a-SCS strategy, in comparison with the T-a-S strategy. The results are given in figure 2. In this example, although arm 1 has the maximum quality, the T-a-SCS strategy "realizes" that arm 1 is very likely infeasible because of the negative values on subpopulation 1. The T-a-SCS strategy thus assigns less empirical sampling weight to arm 1 (in comparison to the T-a-S strategy) aside from checking its feasibility on subpopulation 1. This further allows the T-a-SCS strategy to assign more empirical weight to the other feasible arm 3, and to arrive at the conclusion that arm 2 has better quality than arm 3, and is therefore the best feasible arm, with less sampling times.

6 CONCLUSION

We formulate, analyze and solve the problem of best arm identification with fairness constraints on subpopulations (BAICS). The BAICS problem requires that an selected arm must be fair to all subpopulations by satisfying constraints to regulate conditional expected rewards on each subpopulation. The BAICS problem aims at correctly identify the best arm among all feasible arms. We analyze the complexity of the BAICS problem by proving a best achievable lower bound on the sample complexity. We then design an algorithm, and prove the sample complexity to match with the lower bound in terms of order. A brief account of numerical experiments is conducted to illustrate the theoretical findings.

ACKNOWLEDGMENTS

The authors acknowledge the support from the Hellman Fellows Fund. The authors thank Candace Yano for helpful and inspiring discussions.



(a) T-a-SCS (b) T-a-S

Figure 2: Average sample allocation to the arms and the subpopulations in example 2.

REFERENCES

Andradóttir, S., and S.-H. Kim. 2010. "Fully Sequential Procedures for Comparing Constrained Systems via Simulation". *Naval Research Logistics (NRL)* 57(5):403–421.

Audibert, J.-Y., S. Bubeck, and R. Munos. 2010. "Best Arm Identification in Multi-Armed Bandits.". In *COLT*, 41–53.

Barrier, A., A. Garivier, and T. Kocák. 2022. "A Non-asymptotic Approach to Best-Arm Identification for Gaussian Bandits". In *International Conference on Artificial Intelligence and Statistics*, 10078–10109. PMLR.

Bechhofer, R. E. 1954. "A Single-Sample Multiple Decision Procedure for Ranking Means of Normal Populations with Known Variances". *The Annals of Mathematical Statistics*:16–39.

Bechhofer, R. E., T. J. Santner, and D. M. Goldsman. 1995. "Design and Analysis of Experiments for Statistical Selection, Screening, and Multiple Comparison". *(No Title)*.

Boyd, S., L. Xiao, and A. Mutapcic. 2003. "Subgradient Methods". *Lecture Notes of EE392o, Stanford University, Autumn Quarter* 2004:2004–2005.

Chick, S. E. 2006. "Subjective Probability and Bayesian Methodology". *Handbooks in Operations Research and Management Science* 13:225–257.

Degenne, R., and W. M. Koolen. 2019. "Pure Exploration with Multiple Correct Answers". *Advances in Neural Information Processing Systems* 32.

Garivier, A., and E. Kaufmann. 2016. "Optimal Best Arm Identification with Fixed Confidence". In *Conference on Learning Theory*, 998–1027. PMLR.

Glynn, P., and S. Juneja. 2015. "Selecting the Best System and Multi-Armed Bandits". *arXiv preprint arXiv:1507.04564*.

He, J., and S.-H. Kim. 2019. "A New Reward Function for Bayesian Feasibility Determination". In *2019 Winter Simulation Conference (WSC)*, edited by Y.-J. Son, P. Haas, N. Mustafee, M. Rabe, C. Szabo, and S. Lazarova-Molnar, 3480–3491. National Harbor, Maryland: Institute of Electrical and Electronics Engineers, Inc.

Healey, C., S. Andradóttir, and S.-H. Kim. 2014. "Selection Procedures for Simulations with Multiple Constraints Under Independent and Correlated Sampling". *ACM Transactions on Modeling and Computer Simulation (TOMACS)* 24(3):1–25.

Hong, L. J., W. Fan, and J. Luo. 2021. "Review on Ranking and Selection: A New Perspective". *Frontiers of Engineering Management* 8(3):321–343.

Hong, L. J., J. Luo, and B. L. Nelson. 2015. "Chance constrained selection of the best". *INFORMS Journal on Computing* 27(2):317–334.

Hou, Y., V. Y. Tan, and Z. Zhong. 2022. "Almost Optimal Variance-Constrained Best Arm Identification". *IEEE Transactions on Information Theory*.

Hunter, S. R., and B. L. Nelson. 2017. "Parallel Ranking and Selection". In *Advances in Modeling and Simulation*, 249–275. Springer.

Kaufmann, E., O. Cappé, and A. Garivier. 2016. "On the Complexity of Best Arm Identification in Multi-Armed Bandit Models". *Journal of Machine Learning Research* 17:1–42.

Kaufmann, E., and W. M. Koolen. 2021. "Mixture Martingales Revisited with Applications to Sequential Tests and Confidence Intervals". *The Journal of Machine Learning Research* 22(1):11140–11183.

Kim, S.-H., and B. L. Nelson. 2006. "Selecting the Best System". *Handbooks in Operations Research and Management Science* 13:501–534.

Ma, S., and S. G. Henderson. 2017. "An Efficient Fully Sequential Selection Procedure Guaranteeing Probably Approximately Correct Selection". In *2017 Winter Simulation Conference (WSC)*, edited by E. H. Page, G. Wainer, V. Chan, A. D'Ambrogio, G. Zacharewicz, and N. Mustafee, 2225–2236. Las Vegas, Nevada: Institute of Electrical and Electronics Engineers, Inc.

Russac, Y., C. Katsimerou, D. Bohle, O. Cappé, A. Garivier, and W. M. Koolen. 2021. "A/B/n Testing with Control in the Presence of Subpopulations". *Advances in Neural Information Processing Systems* 34:25100–25110.

Shang, X., R. Heide, P. Menard, E. Kaufmann, and M. Valko. 2020. "Fixed-Confidence Guarantees for Bayesian Best-Arm Identification". In *International Conference on Artificial Intelligence and Statistics*, 1823–1832. PMLR.

Shi, Z., Y. Peng, L. Shi, C.-H. Chen, and M. C. Fu. 2022. "Dynamic Sampling Allocation Under Finite Simulation Budget for Feasibility Determination". *INFORMS Journal on Computing* 34(1):557–568.

Tsai, Y.-A., G. Pedrielli, L. Mathesen, Z. B. Zabinsky, H. Huang, A. Candelieri, and R. Perego. 2018. "Stochastic Optimization for Feasibility Determination: an Application to Water Pump Operation in Water Distribution Network". In *2018 Winter Simulation Conference (WSC)*, edited by B. Johansson, S. Jain, O. Rose, M. Rabe, A. Skoogh, N. Mustafee, and A. A. Juan, 1945–1956. Gothenburg, Sweden: Institute of Electrical and Electronics Engineers, Inc.

Wang, Z., A. J. Wagenmaker, and K. Jamieson. 2022. "Best Arm Identification with Safety Constraints". In *International Conference on Artificial Intelligence and Statistics*, 9114–9146. PMLR.

AUTHOR BIOGRAPHIES

YUHANG WU is a second-year Ph.D. student in the Department of Industrial Engineering & Operations Research at the University of California Berkeley. He received his B.S. in Mathematics from Peking University. He has done research in applied probability and statistics. His email address is wuyh@berkeley.edu, and his website is https://wyharturia.github.io/homepage/.

ZEYU ZHENG is an assistant professor in the Department of Industrial Engineering & Operations Research at University of California Berkeley. He received his Ph.D. in Management Science and Engineering, Ph.D. minor in Statistics and M.A. in economics from Stanford University, and B.S. in Mathematics from Peking University. He has done research in simulation, nonstationary stochastic modeling and sequential experiments. His email address is zyzheng@berkeley.edu and his website is https://zheng.ieor.berkeley.edu/.

TINGYU ZHU is a first-year Ph.D. student in the Department of Industrial Engineering & Operations Research at the University of California Berkeley. She received her B.S. in Mathematics from Peking University. She has done research in simulation. Her email address is tingyu_zhu@berkeley.edu.