Posterior Tracking Algorithm for Classification Bandits

Koji Tabata
Hokkaido University

Junpei Komiyama
New York University

Atsuyoshi Nakamura
Hokkaido University

Tamiki Komatsuzaki
Hokkaido University

Abstract

The classification bandit problem aims to determine whether a set of given K arms contains at least L good arms or not. Here, an arm is said to be good if its expected reward is no less than a specified threshold. To solve this problem, we introduce an asymptotically optimal algorithm, named P-tracking, based on posterior sampling. Unlike previous asymptotically optimal algorithms that require solving a linear programming problem with an exponentially large number of constraints, P-tracking solves an equivalent optimization problem that can be computed in time linear in K. Additionally, unlike existing algorithms, P-tracking does not require forced exploration steps. Empirical results show that P-tracking outperforms existing algorithms in sample efficiency.

1 INTRODUCTION

We consider an adaptive combinatorial exploration problem based on a stochastic bandit setting, which is called the classification bandit (Tabata et al., 2021). This problem involves a set of K arms, and each arm is associated with a reward distribution. The agent at each time step chooses one of the arms and receives a sample from the corresponding distribution. The goal of the agent is to classify the model[1] as a whole into *positive* or *negative*. If there are L or more arms with means at least a (predefined) threshold ξ, then the model is defined to be positive. Otherwise, the model is negative.

The classification bandit problem is a versatile model that can be applied to many real-world problems, including the following two examples.

Example 1.1. (Cytodiagnosis, Helal et al. 2019) *Suppose a screening test for cytodiagnosis using a device that requires high cost and long time for accurate measurements such as Raman microscopy. In this test, a pathological diagnosis is made based on whether the ratio of malignant cells to the total number of cells in the specimen exceeds a specified threshold. The malignancy of each cell is quantified by averaging the malignancy measures of all pixels within the cell on the Raman image. If the number of malignant cell is small (i.e., the ratio below a certain threshold) under immune system, a doctor may not necessarily diagnose as the patient being at the fatal, malignant stage. However, if the number of malignant cells is high enough, the patient is diagnosed as being at a certain stage requiring further inspection.*

Example 1.2. (Quality inspection) *Suppose a quality control that requires a verification of the overall quality of items prior to shipment. An inspector, tasked with this responsibility, selects a random sample of items to manually inspect for defects at a factory. To minimize bias in the random sampling, the inspector is required to confirm that no more than L out of K items fail to meet the desired quality, while ensuring the required accuracy of inspection.*

The two examples above are pure exploration problems in which the goal is to determine the overall quality of the set of populations using as few samples as possible. Considering an affinity for these tasks, we deal with the fixed confidence setting, in which the confidence level $\delta \in (0,1)$ is given, and the agent stops the the sampling process immediately once his/her confidence on the correct answer reaches $1-\delta$.

The classification bandit problem is a special instance of general structured exploration problems (Huang et al., 2017; Degenne and Koolen, 2019). In some class of problems (including classification bandit problem), general algorithms called C-Tracking and D-Tracking (Garivier and Kaufmann, 2016; Degenne and Koolen, 2019) are known. These algorithms "track" the opti-

[1] We use the word model to refer a set of K (unknown) distributions.

Proceedings of the 26th International Conference on Artificial Intelligence and Statistics (AISTATS) 2023, Valencia, Spain. PMLR: Volume 206. Copyright 2023 by the author(s).

mal allocation; while the optimal allocation requires the true model parameters, these algorithms use plug-in estimators instead. Although these tracking algorithms are asymptotically optimal in a small δ regime, they have two drawbacks. First, to stabilize the allocation, C-Tracking and D-Tracking involve forced exploration over all arms, which does not necessarily balance the exploration and exploitation. Second, to determine the allocation, these algorithms need to solve a computationally intensive convex optimization for each round. To address the second issue, Wang et al. (2021) proposed an algorithm based on first-order optimization. Although this algorithm is free of convex optimization, it requires solving linear programming whose number of constraints is exponential in L in our case.

Tabata et al. (2021) explored a class of problems that is slightly more general than the classification bandit problem considered in the paper. However, the algorithm they proposed lacks regret analysis and is not designed to be asymptotically optimal. The classification bandit problem is a generalization of Sequential Test for the Lowest Mean (Kaufmann et al., 2018) and Bad Arm Existence Checking (Tabata et al., 2020), which correspond to the case of $L = K$ and $L = 1$, respectively.

The classification bandit problem is also closely related to several problems, such as the thresholding bandit problem (Locatelli et al., 2016; Xu et al., 2020). The goal of the thresholding bandit problem is to determine whether each arm is above or below the threshold, which predicates the class (i.e., positive or negative) of the model. Therefore, the use of a threshold bandit algorithm suffices for the classification bandits. Although there is no exact inclusion relationship, solving the top-L subset selection problem (Kalyanakrishnan et al., 2012) usually suffices to solve the classification bandit problem. At first glance, this seems to imply that there is no particular demand for tailoring an algorithm for the classification bandit problem. However, as we show in this paper, if we focus on classification, the sample complexity of the problem is much smaller than that for solving the thresholding bandit problem. The sample complexity of the thresholding bandit problem and the top-L subset selection problem is $O(K \log 1/\delta)$, whereas the sample complexity of the classification bandit problem is only $O(L \log 1/\delta)$ (resp. $O((K - L) \log 1/\delta)$) for "positive" (resp. "negative") case as we show in this paper. In summary, (1) C-Tracking and D-Tracking algorithms are feasible but not very computationally efficient and involve forced exploration. (2) Thresholding bandit and top-L subset selection algorithms are sample inefficient when $L < K$. These facts incentivize us to invent an algorithm that is optimized for the classification bandit problem. The methodological contributions of this paper are as follows:

- We characterize the sample complexity of the classification bandit problem. Although optimizations in the fixed confidence best arm identification problem (Garivier and Kaufmann, 2016; Degenne and Koolen, 2019) as well as many structured bandit problems (e.g., Magureanu et al. (2014); Komiyama et al. (2015, 2016); Lattimore and Szepesvári (2017)) are known as a form of linear semi-infinite programming[2]. Solving such an optimization for each round limits the utility of the method. Instead, we propose an equivalent discrete optimization that runs in $\tilde{O}(K)$ time.
- We propose a posterior-tracking (P-Tracking) algorithm that has an optimal sample complexity when $\delta \to +0$. Similarly to C-Tracking and D-Tracking, we follow the optimal allocation. Unlike C-Tracking and D-Tracking, P-Tracking does not require forced exploration. Using the posterior sample instead of the empirical mean, P-Tracking conducts implicit exploration.
- We conduct an extensive set of simulations. In particular, we tested many different values of L, K. We verify the advantage of P-Tracking in terms of dependence on L (over thresholding bandit algorithms and top-L identification algorithms) as well as in terms of amount of forced exploration (over C-Tracking and D-Tracking that are also asymptotically optimal).

A byproduct of our proposed method is the use of posterior samples in fixed confidence structured bandit problems. While posterior sampling is widely used to balance exploration and exploitation in an online objective (i.e., Thompson sampling (TS) for the multi-armed bandit problem), limited results are known about its use in the best arm identification, partly due to its challenge in the analysis. One of the well-known versions of TS for the best arm identification is the top-two Thompson sampling (TTTS Russo 2016; Shang et al. 2020). However, TTTS is not directly applicable to structured pure-exploration. More importantly, we use TS in a different way than TTTS. While TTTS uses posterior samples to (implicitly) calculate the optimal allocation, we use posterior sampling so that we do not require forced exploration given an optimal allocation. Such a use of posterior sampling potentially benefits many other structured pure exploration problems.

[2] Linear programming with infinite number of constraints.

2 CLASSIFICATION BANDITS

In this section, we formulate the Classification Bandit problem. We give a notation list in Table A in the supplementary material.

2.1 Problem Setting

Let K be the number of arms. The reward of each arm $i \in [K] := \{1, 2, \dots, K\}$ follows a distribution with mean μ_i. In this paper, we assume that the distributions are Bernoulli. Parameters $\boldsymbol{\mu} := \{\mu_i\}_{i=1}^K \in (0,1)^K$ are unknown to the agent.

Definition 1 (good arm and bad arm)**.** *For a given threshold $\xi \in (0,1)$, we define an arm with its mean larger than or equal to (smaller than) ξ as a good arm (bad arm), respectively.*

We consider the following adaptive setting, which is common in the structured pure exploration literature. For any time $t \geq 1$, the agent chooses an arm i_t from $[K] = \{1, 2, \dots, K\}$, and observes a reward from the underlying distribution associated with i_t. The agent can stop anytime and returns "positive" or "negative" upon stopping time τ. We use the term *algorithm* to describe the strategy for choosing i_t and τ that the agent uses.

Definition 2 (δ-correct)**.** *An algorithm is called δ-correct if the algorithm returns a correct answer "positive" or "negative" with probability at least $1-\delta$.*

In the following, we limit our interest to δ-correct algorithms. That is, we consider the following problem.

Problem 1. *For given $L \in [K]$, $\delta \in (0, 0.5)$, and the threshold ξ, if the number of good arms is larger than or equal to L, answer "positive" with probability at least $1-\delta$ as few samples as possible. Otherwise, answer "negative" with probability at least $1-\delta$ as few samples as possible.*

3 OPTIMAL ALLOCATION

In this section, we discuss the asymptotically optimal allocation of arm selection and its computation.

First, we denote the number of good arms with parameters $\boldsymbol{\mu}$ by $\hat{M}(\boldsymbol{\mu})$, that is, $\hat{M}(\boldsymbol{\mu}) = \sum_{i\in[K]} \mathbf{1}\{\mu_i \geq \xi\}$. For the vector $\boldsymbol{\mu}$ of the arms' expected rewards, a set $\mathrm{Alt}(\boldsymbol{\mu})$ is defined as

$$\mathrm{Alt}(\boldsymbol{\mu}) = \begin{cases} \{\boldsymbol{\nu} \mid \hat{M}(\boldsymbol{\nu}) < L\} & (\hat{M}(\boldsymbol{\mu}) \geq L) \\ \{\boldsymbol{\nu} \mid \hat{M}(\boldsymbol{\nu}) \geq L\} & (\hat{M}(\boldsymbol{\mu}) < L), \end{cases}$$

that is, the set of models $\boldsymbol{\nu}$ for which the answer of Problem 1 is negative (resp. positive) when the true answer is positive (resp. negative).

3.1 Lower Bound

Recent papers (Garivier and Kaufmann, 2016; Degenne and Koolen, 2019) shows that we can construct an asymptotic lower bound of the expected stopping time τ_δ for structured pure-exploration problems. In particular, the bound is represented as follows:

$$\liminf_{\delta\to 0} \frac{\mathbb{E}_{\boldsymbol{\mu}}[\tau_\delta]}{\log(1/\delta)} \geq \left(\sup_{\boldsymbol{w}\in\Delta_K} \inf_{\boldsymbol{\mu}'\in\mathrm{Alt}(\boldsymbol{\mu})} \sum_{i=1}^K w_i d(\mu_i, \mu'_i) \right)^{-1}, \tag{1}$$

where Δ_K is $\{\boldsymbol{w} \in [0,1]^K : \sum_{i=1}^K w_i = 1\}$, and $d(x,y)$ is a Kullback-Leibler divergence of two Bernoulli distributions with mean x and y.

The maximizer $\boldsymbol{w}$ in the above expression corresponds to the optimal allocation for arm selection, that is, when the arms are chosen so that the number of selections of each arm i is proportional to w_i, the stopping time is asymptotically optimal for $\delta \to 0$.

As described in Wang et al. (2021), RHS of equation 1 is a convex programming. However, it requires care on the discontinuities, and the method proposed by Wang et al. (2021) requires solving a linear programming of a size exponential to L. In the following, we show a more explicit formula on the solution of this problem.

3.2 Equivalent Optimization

For a set of estimated parameters $\boldsymbol{\nu} = (\nu_1, \nu_2, \dots, \nu_K)$. We denote the sorted indices by $\nu_{(1)} \geq \nu_{(2)} \geq \dots \geq \nu_{(L)} \geq \dots \geq \nu_{(\hat{M}(\boldsymbol{\nu}))} \geq \xi > \nu_{(\hat{M}(\boldsymbol{\nu})+1)} \geq \dots \geq \nu_{(K)}$.

Here, we assume $\hat{M}(\boldsymbol{\nu}) \geq L$ (positive model) since the following remark holds.

Remark 1. *In the case of $\hat{M}(\boldsymbol{\nu}) < L$ (negative), we are able to convert the model into an equivalent positive model with flipped variables $(\boldsymbol{\nu}', \xi', L')$ such that*

$$\begin{aligned} \nu'_i &= 1 - \nu_i, \\ \xi' &= 1 - \xi, \\ L' &= K - L + 1. \end{aligned}$$

Consider the following optimization problem:

$$\begin{aligned} &\min_{\boldsymbol{S}=(S_1,S_2,\dots,S_K)} \sum_{i=1}^K S_i \\ &\text{subject to} \\ &\quad \min_{U\subseteq[\hat{M}(\boldsymbol{\nu})]:|U|=\hat{M}(\boldsymbol{\nu})-L+1} \sum_{i\in U} S_{(i)} d(\nu_{(i)}, \xi) \geq 1 \\ &\quad S_i \geq 0 \;\; (i = 1, \dots, K). \end{aligned} \tag{2}$$

The solution of the optimization of equation 2 can be given as an explicit formula as shown by the following theorem.

Theorem 1. *Let*

$$L^*(\boldsymbol{\nu}) = \operatorname*{arg\,min}_{l \in [\hat{M}(\boldsymbol{\nu})] \setminus [L-1]} \frac{1}{l-L+1} \sum_{i=1}^{l} \frac{1}{d(\nu_{(i)}, \xi)}. \quad (3)$$

$\boldsymbol{S}(\boldsymbol{\nu})$ defined as follows is an optimal solution[3] of the optimization problem of equation 2:

$$S_{(i)}(\boldsymbol{\nu}) = \begin{cases} \frac{1}{L^*(\boldsymbol{\nu})-L+1} \frac{1}{d(\nu_{(i)},\xi)}, & \text{if } i \le L^*(\boldsymbol{\nu}), \\ 0 & \text{otherwise.} \end{cases} \quad (4)$$

and the optimal value is

$$Q^*(\boldsymbol{\nu}) = \frac{1}{L^*(\boldsymbol{\nu})-L+1} \sum_{i=1}^{L^*(\boldsymbol{\nu})} \frac{1}{d(\nu_{(i)}, \xi)}. \quad (5)$$

In the context where the plug-in parameter is clear (usually, $\boldsymbol{\nu} = \boldsymbol{\mu}$), we omit $\boldsymbol{\nu}$. For example, we use $\boldsymbol{S} = (S_1, S_2, \dots, S_K)$ to describe the optimal solution.

The equivalence of equation 2 and the right-hand side of inequality 1 is guaranteed by the following proposition.

Proposition 2. *The solution $\boldsymbol{S}(\boldsymbol{\mu})$ of equation 2 for $\boldsymbol{\nu} = \boldsymbol{\mu}$ and the solution $w_1^*, \dots, w_K^*$ of the optimization problem $\max_{\boldsymbol{w} \in \Delta_K} \inf_{\boldsymbol{\mu}' \in \mathrm{Alt}(\boldsymbol{\mu})} \sum_{i=1}^{K} w_i d(\mu_i, \mu_i')$ have relation $w_i^* = S_i(\boldsymbol{\mu})/Q^*(\boldsymbol{\mu})$.*

Theorem 1 and Proposition 2 state that the optimal sample complexity of the classification bandit problem is $Q^*(\boldsymbol{\mu}) \log(1/\delta)$, which is achieved when we draw each arm proportional to w_i^*.

Remark 2. (implication of the optimization) *The case of $L^*(\boldsymbol{\nu}) = L$ corresponds to the top-L identification: It draws i-th best arm for* $\log(1/\delta)/d(\boldsymbol{\nu}_{(i)}, \xi)$*, which corresponds to identifying it as a good arm. The more interesting case is $L^*(\boldsymbol{\nu}) > L$. In this case, none of the top-L arms are identified as a good arm, but one can derive that there are at least L good arms with a confidence level δ.*

Remark 3. ($O(K \log K)$ runtime, $O(L)$ sample complexity) *Unlike the original optimization, the alternative optimization of equation 4 no longer requires a convex optimization. We can run the alternative optimization in $O(K \log K)$ time in the following procedure: First, we sort the means to obtain $\nu_{(1)} \ge \nu_{(2)} \ge \dots \ge \nu_{(K)}$, which runs in $O(K \log K)$ time. Second, let*

$$Q_l = \frac{1}{l-L+1} \sum_{i=1}^{l} \frac{1}{d(\nu_{(i)}, \xi)}.$$

We first start with $l = L$ and compute Q_L, and repeat comparing Q_l and Q_{l+1} as long as $l < \hat{M}(\boldsymbol{\nu})$ and $Q_{l+1} < Q_l$ holds.[4] This yields $Q^*(\boldsymbol{\mu}) = \arg\min_l Q_l$ and runs in $O(K)$ time.

Note also that the above discussion implies the sample complexity of $Q^*(\boldsymbol{\mu}) \le Q_L(\boldsymbol{\mu}) = O(L \log(1/\delta)/\Delta_L^2)$, where $\Delta_L^2 = (\nu_{(L)} - \xi)^2$.

Algorithm 1 P-Tracking for Classification Bandit

```
Require: $K$, $\xi$, $L$, $\delta$
Ensure: "positive" (if $|\{i \in [K] \mid \mu_i \ge \xi\}| \ge L$), "negative" (otherwise) with prob. at least $1-\delta$
$a_i \leftarrow 1, b_i \leftarrow 1$ for $i = 1, 2, \dots, K$
for $t = 1, 2, \dots$ do
    $\theta_i(t) \sim \mathrm{Beta}(a_i, b_i)$ for $i \in [K]$
    if $\sum_i \mathbf{1}[\theta_i(t) \ge \xi] \ge L$ then
        $\boldsymbol{S}(\boldsymbol{\theta}(t)) \leftarrow \mathrm{PO}(K, \xi, L, \boldsymbol{\theta}_i(t))$
    else
        $\boldsymbol{S}(\boldsymbol{\theta}(t)) \leftarrow \mathrm{PO}(K, 1-\xi, K-L+1, (1-\theta_i(t))_{i \in [K]})$
    $i_t \leftarrow \arg\max_i S_i(\boldsymbol{\theta}(t))/N_i(t)$ (posterior sampling)
    $a_{i_t} \leftarrow a_{i_t} + r_t$; $b_{i_t} \leftarrow b_{i_t} + 1 - r_t$
    if $\sum_i \mathbf{1}[\hat{\mu}_i(t) \ge \xi] \ge L$ then
        if $\mathrm{SC}(K, \xi, L, (\hat{\mu}_i(t))_{i \in [K]})$ = "true" then
            Return "positive"
    else
        if $\mathrm{SC}(K, 1-\xi, K-L+1, (1-\hat{\mu}_i(t))_{i \in [K]})$ = "true" then
            Return "negative"
```

Algorithm 2 PosteriorOptimization (PO)

```
Require: $K$, $\xi$, $L$, $(\theta_i(t))_{i \in [K]}$
Return $\boldsymbol{S}(\boldsymbol{\theta}(t)) \leftarrow$ Solution of equation 4
```

The optimization above requires the true parameters $\boldsymbol{\mu}$ that requires estimation via some exploration. In the next section, we propose P-Tracking that adopts posterior sampling for balancing exploration and exploitation.

4 P-TRACKING

In this section, we introduce Posterior Tracking (P-Tracking, Algorithm 1), a conceptually simple Bayesian algorithm. Following the literature on the Bayesian multi-armed bandit problem (i.e., Thompson sampling), we adopt the uniform prior $\mathrm{Beta}(1, 1)$. Regarding arm selection, it calculates $\boldsymbol{S}(\boldsymbol{\theta})$ based on a

[3] While the optimal solution can be non-unique, in the analysis (section 4) we consider the case of the unique optimal solution.

[4] Parameter l satisfying this condition is guaranteed to be $L^*(\boldsymbol{\nu})$ by Proposition 12(1) and (2) in Supplementary Material.

Algorithm 3 StoppingCondition (SC)

```
Require: $K$, $\xi$, $L$, $(\hat{\mu}_i(t))_{i\in[K]}$
Ensure: "true" (if $L$ positive arms), "false" (otherwise).
$Z_i \leftarrow \begin{cases} N_i(t)d(\hat{\mu}_i,\xi), & \text{if } \hat{\mu}_i > \xi \\ 0, & \text{otherwise} \end{cases}$.
Sort them such that $Z_{(1)} \ge Z_{(2)} \ge \cdots \ge Z_{(K)}$.
if $\sum_{k=L}^{K} Z_{(k)} \ge \beta(t,\delta)$ then
    Return "true"
Return "false"
```

posterior sample $\boldsymbol{\theta}$, and tries to draw arms proportionally to $\boldsymbol{S}(\boldsymbol{\theta})$ (Algorithm 2). Once we receive a reward, we update the posterior of the selected arm. Regarding the stopping condition, it stops once the empirical log-likelihood, which is defined in terms of the empirical means, reaches $\beta(t,\delta)$ (Algorithm 3). The value $\beta(t,\delta)$ should be an anytime confidence bound $\beta(t,\delta)$, which satisfies the following properties.

$$\mathbb{P}\left[\bigcup_{t\ge 1,\,\hat{\mathcal{I}}\in 2^{[K]}}\left\{\sum_{i\in\hat{\mathcal{I}}} N_i(t)d(\hat{\mu}_i(t),\mu_i) \ge \beta(t,\delta)\right\}\right] \le \delta \tag{6}$$

$$\exists C_1, C_2 > 0 \quad \beta(t,\delta) \le C_1 + \log\left(\frac{C_2\log(t+1)}{\delta}\right), \tag{7}$$

where $N_i(t) = \sum_{s<t}\mathbf{1}[i_t = i]$ and $\hat{\mu}_i(t) = \sum_{s<t}\mathbf{1}[i_t = i, r_t = 1]/N_i(t)$ are the number of selections and the empirical mean of arm i by time t, respectively. An example of such an anytime confidence bound is found in Proposition 21 in Kaufmann and Koolen (2021).

In the following, we derive the correctness (Theorem 5) and an optimal stopping time (Theorem 6) of Algorithm 1.

4.1 Assumptions on the True Parameters

In the following, we state the assumptions on the true parameters $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_K)$. Let $M := \hat{M}(\boldsymbol{\mu})$.

Assumption 1. (True positive) *We assume $M \ge L$ (i.e., positive model).*[5] *Without loss of generality, we assume the true parameters $\boldsymbol{\mu}$ satisfies $\mu_1 \ge \mu_2 \ge \mu_3 \ge \cdots \ge \mu_M > \xi > \mu_{M+1} \ge \cdots \ge \mu_K$.*

In view of Remark 1, Assumption 1 can be posed without loss of generality because a true negative model is flipped into a true positive model.

Assumption 2. (Identifiable parameters) *we assume that no parameter is exactly the same as ξ. That is, $\mu_i \ne \xi$ for all $i \in [K]$.*

Assumption 3. (Non-degenerate solution) *We assume that the solution $\boldsymbol{S}(\boldsymbol{\mu})$ is unique.*

Assumption 3 implies the existence of the following margin constant that guarantees the quality of the solution. For model $\boldsymbol{\nu}$, let the "solution set" $\hat{\mathcal{I}}(\boldsymbol{\nu}) = \{i \in [K] : S_i(\boldsymbol{\nu}) > 0\}$, which is the subset of good arms $\{i \in [K] : \nu_i > \xi\}$ where the optimal solution includes drawing these arms. We also denote $\mathcal{I} = \hat{\mathcal{I}}(\boldsymbol{\mu})$ to be the true solution set.

Remark 4. (Existence of margin constant) *For a $D_{\min} > 0$, let*

$$\Theta_{D_{\min}} = \left\{\boldsymbol{\nu} : \forall_{i\in[K]} |\nu_i - \mu_i| \le D_{\min}\right\}. \tag{8}$$

By choosing a sufficiently small[6] *$D_{\min} > 0$, for any $\boldsymbol{\nu} \in \Theta_{D_{\min}}, i \in [K]$, it holds*

$$\frac{1}{2}S_i(\boldsymbol{\mu}) \le S_i(\boldsymbol{\nu}) \le 2S_i(\boldsymbol{\mu}). \tag{9}$$

Intuitively speaking, if $|\nu_i - \mu_i| \le D_{\min}$ holds for all arms, then the solution $\boldsymbol{S}(\boldsymbol{\nu})$ is a constant-ratio approximation of the true solution $\boldsymbol{S}(\boldsymbol{\mu})$. Note that equation 9 implies $\hat{\mathcal{I}}(\boldsymbol{\nu}) = \mathcal{I}$. Remark 4 is trivially derived by using the uniqueness and continuity of the solution $\boldsymbol{S}(\boldsymbol{\mu})$.

Definition 3. (Subset margin region) *We define the region*

$$\Theta_{D_{\min},\mathcal{I}} = \left\{\boldsymbol{\nu} : \bigcap_{i\in\mathcal{I}}\{|\nu_i - \mu_i| \le D_{\min}\}\right\}.$$

Definition 3 states that parameter ν_i of arms in $\mathcal{I}$ are close to the true parameter μ_i. By definition, $\Theta_{D_{\min},\mathcal{I}} \supset \Theta_{D_{\min}}$.

In fact, Remark 4 can be strengthened to the following lemma.

Lemma 3. (Non-interference) *Let $i \notin \mathcal{I}$. Let $\boldsymbol{\nu} \in \Theta_{D_{\min},\mathcal{I}}$ and $\nu_i \le \mu_i + D_{\min}$. Then, it holds that*

$$S_i(\boldsymbol{\nu}) = 0.$$

Lemma 3 states that if the arms in $\mathcal{I}$ are $D_{\min}$-accurate and arm $i \notin \mathcal{I}$ is not very good, then the arm i is not included in the optimal set regardless of the other arms.

4.2 Characterization on the Posterior Sample

Lemma 4. (geometry) *There exists a (distribution-dependent) constant $C = C(\boldsymbol{\mu})$ such that, for any $\hat{\boldsymbol{\mu}}(t)$,*

[5] We also drop $\boldsymbol{\mu}$ for many functions.

[6] Here, the choice of $D_{\min}$ depends on $\boldsymbol{\mu}$ but is independent of the confidence level δ. Following the literature, we consider $\boldsymbol{\mu}$ as a set of constants.

we have

$$\mathbb{P}\left[\boldsymbol{\theta}(t) \in \Theta_{D_{\min}}\right] \geq C(\boldsymbol{\mu}) \exp\left(-\sum_{i\in[K]} N_i(t) d(\hat{\mu}_i(t), \mu_i)\right).$$

Lemma 4 is one of the most important lemmas that guarantees the amount of exploration in posterior sampling. Essentially, it states if the underestimation of the arms is of q-th quantile[7], then with probability $\Omega(1/q)$, the posterior $\boldsymbol{\theta}(t)$ is around the true value $\boldsymbol{\mu}$. This property is used to guarantee the amount of exploration.

4.3 Main Theorems

This section introduces two main theorems. Theorem 5 states that Algorithm 1 is δ-correct. Theorem 6 shows the optimal sample complexity of the algorithm.

Theorem 5. (Main theorem 1, correctness of the output) *With probability at least $1-\delta$, Algorithm 1 stops and outputs the correct classification result (i.e., positive/negative).*

Proof of Theorem 5. We defer the fact that the algorithm stops almost surely to Lemma 11. Assuming that, we here show that the probability algorithm makes an incorrect identification is at most δ. Let

$$Z_i^{\text{flip}} = \begin{cases} N_i(t) d(\hat{\mu}_i, \xi), & \text{if } \hat{\mu}_i < \xi \\ 0, & \text{otherwise} \end{cases}.$$

Let the sorted values be

$$Z_{(1)}^{\text{flip}}(t) \geq Z_{(2)}^{\text{flip}}(t) \geq \cdots \geq Z_{(K)}^{\text{flip}}(t).$$

The algorithm outputs "negative" if

$$\sum_{i=K-L+1}^{K} Z_{(i)}^{\text{flip}}(t) \geq \beta(t,\delta). \tag{10}$$

Here, we have the following:

$$\left\{\sum_{i=K-L+1}^{K} Z_{(i)}^{\text{flip}}(t) \geq \beta(t,\delta)\right\} \subset \left\{\sum_{i\in[L]} Z_i^{\text{flip}}(t) \geq \beta(t,\delta)\right\}$$

(equation 10 implies that $\sum_{i\in\mathcal{S}} Z_i^{\text{flip}}(t) \geq \beta(t,\delta)$ holds for any $\mathcal{S} \subset [K] : |\mathcal{S}| \leq L$)

$$\subset \left\{\sum_{i\in[L]} N_i(t) d(\hat{\mu}_i(t), \mu_i) \geq \beta(t,\delta)\right\} \tag{11}$$

(by $\hat{\mu}_i < \xi < \mu_i$ for $i \in [L]$). (12)

The anytime confidence bound of equation 6 implies that

$$\mathbb{P}\left[\bigcup_{t\geq 1}\left\{\sum_{i\in[L]} N_i(t) d(\hat{\mu}_i(t), \mu_i) \geq \beta(t,\delta)\right\}\right] \leq \delta,$$

which states that the stopping with "negative" occurs at most probability δ. □

Theorem 6. (Main theorem 2, sample complexity) *Let τ be the stopping time of the algorithm (i.e., the round when the algorithm returns "positive" or "negative"). Then,*

$$\mathbb{E}[\tau] \leq Q^*(\boldsymbol{\mu}) \log(1/\delta) + o(1/\delta). \tag{13}$$

Theorem 6 bounds the expected number of samples required by the algorithm.

Proof of Theorem 6. First, we define the following events. Let $D(\delta) = \sqrt{\log(1/\delta)}$. Let the events be

$$\mathcal{D}(t) = \left\{\sum_{i\in[K]} N_i(t) d(\hat{\mu}_i(t), \mu_i) \geq \log(D(\delta))\right\}, \tag{14}$$

$$\mathcal{E}_i(t) = \left\{i_t = i\right\}, \tag{15}$$

$$\mathcal{E}(t) = \bigcup_{i\notin\mathcal{I}} \mathcal{E}_i(t), \tag{16}$$

$$\mathcal{H}(t) = \left\{\boldsymbol{\theta}(t), \hat{\boldsymbol{\mu}}(t) \in \Theta_{D_{\min},\mathcal{I}}\right\}. \tag{17}$$

Event $\mathcal{D}(t)$ states that the empirical divergence is large enough, which should not occur frequently (i.e., $o(1)$ as $\delta \to 0$). Event $\mathcal{E}(t)$ states that one of the arms outside the solution set $\mathcal{I}$ is drawn. Event $\mathcal{H}(t)$ states that the posterior solution set is a constant ratio approximation of the true solution. For an event $\mathcal{A}$, we use $\mathcal{A}^c$ to denote the complement event.

Let $C_{\text{bnd}} > 0$ be a constant that we define later in Lemma 11. Roughly speaking, the algorithm is terminated in $C_{\text{bnd}} \log(1/\delta) + O(1)$ rounds almost surely. We have

$$\tau = \sum_t \mathbf{1}[t \leq \tau] \tag{18}$$

$$= \sum_{t \leq C_{\text{bnd}}\log(1/\delta)} \mathbf{1}[t \leq \tau] + \sum_{t > C_{\text{bnd}}\log(1/\delta)} \mathbf{1}[t \leq \tau] \tag{19}$$

$$= \sum_{t \leq C_{\text{bnd}}\log(1/\delta)} \mathbf{1}[t \leq \tau, \mathcal{D}^c(t), \mathcal{E}^c(t), \mathcal{H}(t)] \tag{20}$$

$$+ \sum_{t \leq C_{\text{bnd}}\log(1/\delta)} \mathbf{1}[t \leq \tau, \mathcal{D}^c(t), \mathcal{E}(t), \mathcal{H}(t)] \tag{21}$$

$$+ \sum_{t \leq C_{\text{bnd}}\log(1/\delta)} \mathbf{1}[t \leq \tau, \mathcal{D}(t)] \tag{22}$$

[7] Here, $q = \exp\left(\sum_{i\in[K]} N_i(t) d(\hat{\mu}_i(t), \mu_i)\right)$

$$+ \sum_{t \le C_{\mathrm{bnd}} \log(1/\delta)} \mathbf{1}[t \le \tau, \mathcal{D}^c(t), \mathcal{H}^c(t)] \qquad (23)$$

$$+ \sum_{t > C_{\mathrm{bnd}} \log(1/\delta)} \mathbf{1}[t \le \tau]. \qquad (24)$$

Lemma 7. (Leading term)

$$\sum_{t \le C_{\mathrm{bnd}} \log(1/\delta)} \mathbb{P}[t \le \tau, \mathcal{D}^c(t), \mathcal{E}^c(t), \mathcal{H}(t)] \le Q^*(\boldsymbol{\mu}) \log(1/\delta) + o(\log(1/\delta)). \qquad (25)$$

Lemma 7 is the leading term assuming that the model parameters are accurately estimated.

Lemma 8. (Drawing nonsolution arms) *We have*

$$\sum_{t \le C_{\mathrm{bnd}} \log(1/\delta)} \mathbb{P}[t \le \tau, \mathcal{D}^c(t), \mathcal{E}(t), \mathcal{H}(t)] = o(\log(1/\delta)). \qquad (26)$$

Lemma 8 bounds the case where one of the nonsolution arms $\mathcal{I}$ is drawn.

Lemma 9. (Unusual divergence) *We have*

$$\sum_{t \le C_{\mathrm{bnd}} \log(1/\delta)} \mathbb{P}[t \le \tau, \mathcal{D}(t)] = o(\log(1/\delta)). \qquad (27)$$

Lemma 9 considers the case where the empirical divergence is large, which is infrequent, and thus does not affect the leading $O(\log(1/\delta))$ term.

Lemma 10. (Unsaturated rounds)

$$\sum_{t \le C_{\mathrm{bnd}} \log(1/\delta)} \mathbb{P}[t \le \tau, \mathcal{D}^c(t), \mathcal{H}^c(t)] = o(\log(1/\delta)). \qquad (28)$$

Lemma 10 bounds the case that optimal set is undersampled.

Lemma 11. (Drawing more than $O(\log \delta^{-1})$ times) *There exists a constant C_{bnd} such that*

$$\sum_{t > C_{\mathrm{bnd}} \log(1/\delta)} \mathbb{P}[t \le \tau] = O(1). \qquad (29)$$

Lemma 11 bounds the case where the stopping time τ is unusually large.

Lemmas 7–11 in the appendix bound each term of equation 24 in expectation. Applying these lemmas immediately yields equation 13, and thus the proof is completed. □

5 EXPERIMENTAL RESULTS

We conducted experiments with artificial data with $K = 100$ arms. We use $\beta(t, \delta) = \log\left(\frac{\log(t+1)}{\delta}\right)$ and δ is set to 10^{-16} and ξ is set to 0.5. Although this choice of $\beta(t, \delta)$ does not guarantee the δ-correctness, it has been used in existing papers such as Garivier and Kaufmann (2016); Wang et al. (2021), and we follow them. To make the comparison fair, we use the same bound and confidence level among all algorithms.

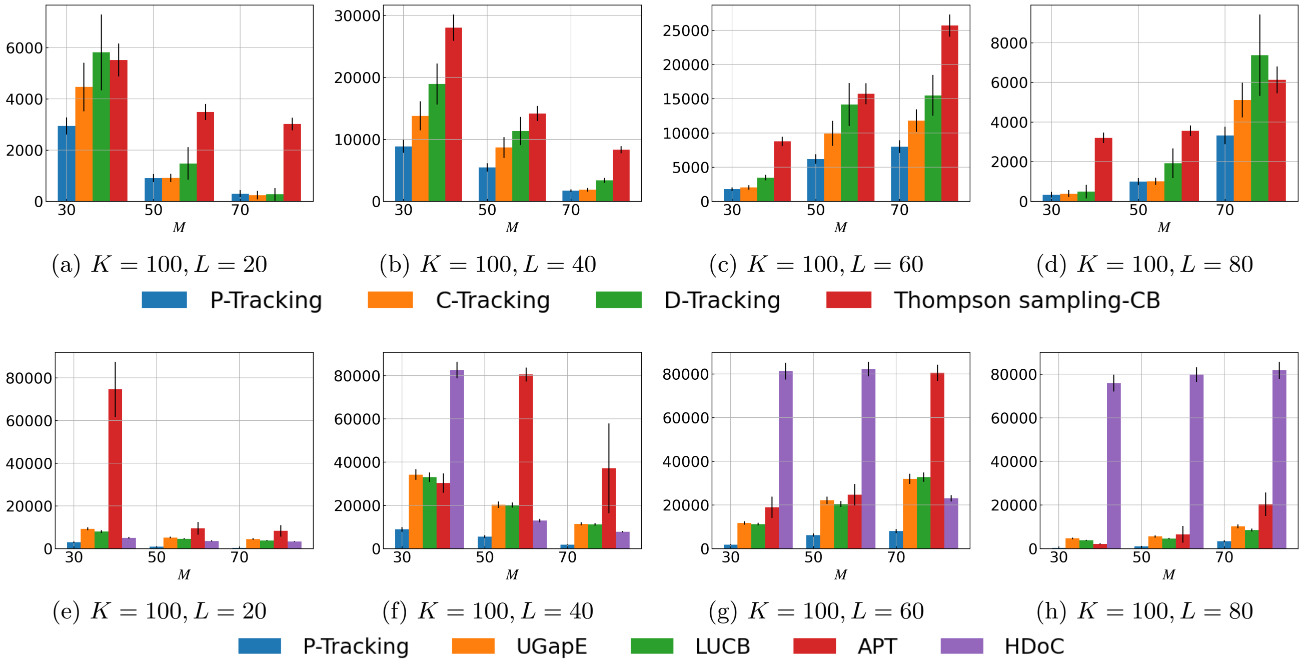
To compare P-Tracking with existing algorithms, we implemented C-Tracking, D-Tracking (Garivier and Kaufmann, 2016), Thompson sampling-CB (Tabata et al., 2021), LUCB (Kalyanakrishnan et al., 2012), UGapE (Gabillon et al., 2012), APT (Locatelli et al., 2016), and HDoC (Kano et al., 2019). We modified some of them for a fair comparison. For C-Tracking and D-Tracking, we adapted the optimization based on $\boldsymbol{S}(\hat{\boldsymbol{\mu}})$ that we have proposed. For Thompson sampling-CB, LUCB, UGapE, APT and HDoC, we changed the stopping condition so that each algorithm does not pull an arm after it is identified as good or bad and stops when it finds the L good arms or $K - L + 1$ bad arms. For each arm i, when $N_i(t) d(\hat{\mu}, \xi) \ge \beta(t, \delta)$ is satisfied, that arm i is classified as good or bad depending on whether $\hat{\mu} > \xi$ or $\hat{\mu} < \xi$. The parameters of LUCB and UGapE were set to find the top-L arms among K arms. Whenever good arm is found, that arm is eliminated from the candidates and L is decremented.[8] Ties are broken by choosing the arm with the lowest index.

We tested several values of L, and the number of good arms M for each of $K = 100$. We set the mean reward of each arm so that $\mu_1, \mu_2, \ldots, \mu_{K-M}$ and $\mu_{K-M+1}, \mu_{K-M+2}, \ldots, \mu_K$ are equally spaced, respectively. Here, we did not assign μ_i in the interval $(\xi - 0.05, \xi + 0.05) = (0.45, 0.55)$ to avoid a very large sample complexity. In summary, the value of μ_i is set as follows:

$$\mu_i = \begin{cases} \frac{0.45i}{K-M-1}, & i \le K - M \\ 0.55 + \frac{0.45(i-K+M-1)}{M-1}, & i \ge K - M + 1 \end{cases}.$$

Figure 1 shows the average and standard deviation of stopping times for 100 runs of each algorithm for different L and M. P-Tracking consistently outperforms (1)

[8] Sometimes, the value L is equal to the number of the remaining arms. This happens when $K - L$ bad arms are identified (note that $K - L + 1$ bad arms should be identified to satisfy the stopping condition). If this is the case, LUCB and UGapE cannot choose the next arm because they require at least $L + 1$ arms among which they choose L. In this case, we decrement L to (the number of remaining arms) $- 1$ so that the algorithm does not crash.



(a) $K = 100, L = 20$

(b) $K = 100, L = 40$

(c) $K = 100, L = 60$

(d) $K = 100, L = 80$

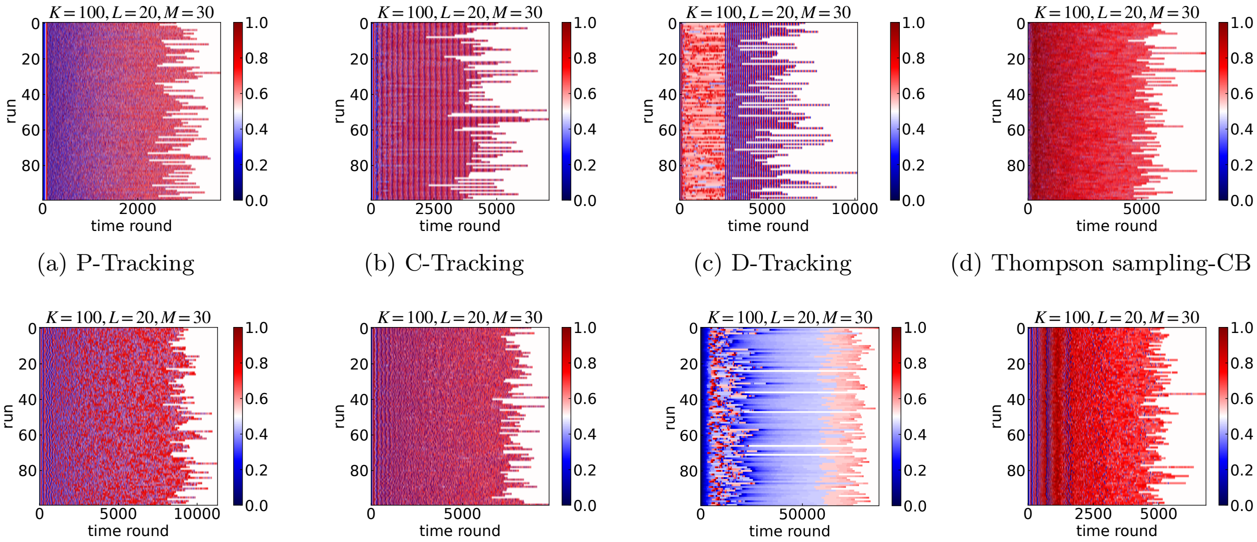
(e) $K = 100, L = 20$

(f) $K = 100, L = 40$

(g) $K = 100, L = 60$

(h) $K = 100, L = 80$

Figure 1: Stopping time of each algorithm for $K = 100$. We tested $K = 100$, $L = 20, 40, 60, 80$ and the number of good arms $M = 30, 50, 70$. (a)-(d) P-Tracking and the other algorithms designed for the classification bandit problem. (e)-(h) P-Tracking and the algorithms with suboptimal sample complexity.



(a) P-Tracking

(b) C-Tracking

(c) D-Tracking

(d) Thompson sampling-CB

(e) UGapE

(f) LUCB

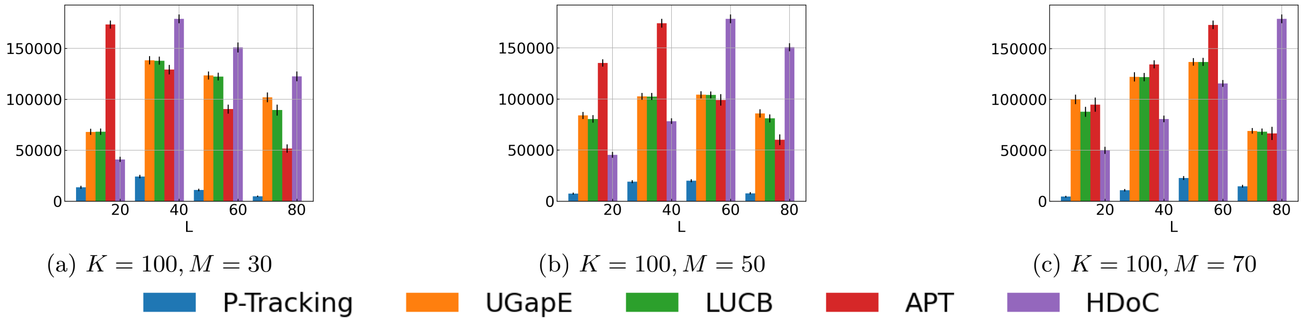
(g) APT

(h) HDoC

Figure 2: Heat maps of arm selection of each algorithm: each x axis corresponds to time step t and each y axis corresponds to the number of each run. Note the range of time step in each figure differs to each other. The color indicated the mean reward μ_i of the chosen arm at that time. (a) Initially, P-Tracking draws 70 bad arms and 30 good arms evenly, gradually reducing the number of bad arm choices. The exploration is distributed evenly thanks to randomized exploration. (b) C-Tracking conducts exploration of ratio $2/\sqrt{K^2+t}$, which is visualized by the stripes. (c) D-Tracking initiates forced exploration after the round $t : \sqrt{t} \geq K/2$.

Thompson sampling-CB, UGapE, LUCB, APT, and HDoc that do not have asymptotically optimal sample complexity, as well as (2) C-Tracking and D-Tracking that have an asymptotically optimal sample complexity.

Figure 2 shows which arm is chosen by each algorithm



(a) $K = 100, M = 30$

(b) $K = 100, M = 50$

(c) $K = 100, M = 70$

Figure 3: Stopping time of each algorithm for K = 100, where μ_i is set to 0.4 for each $i = 1, 2, ..., K - M$ and 0.6 for each $i = K - M + 1, K - M + 2, ..., K$. The advantage of P-Tracking is particularly noticeable when $|M - L|$ is large.

in each time round t over 100 independent runs. One can see that forced exploration in C-Tracking and D-Tracking compromises their performance even with a very small δ. We also conducted experiments with $K = 20$, and the results are essentially the same as those obtained with $K = 100$. These results are presented in the Supplemental Material.

In order to demonstrate the advantage of P-Tracking with respect to the dependency on L over algorithms designed for other purposes such as best arm identification, we also conducted experiments in the setting where all the M good arms has expected mean 0.6, and all the $K - M$ bad arms have expected mean 0.4. The result is shown in the fig. 3. P-Tracking stops earlier when L deviates from M, while the other methods exhibit less variation.

6 CONCLUSION

We have considered the classification bandit problem where the goal is to judge whether there are enough good arms or not. We have introduced P-Tracking and show its advantage in theory and empirical performance. We show that an equivalent optimization for determining the optimal allocation runs in $O(K \log K)$ time. As a result, P-Tracking runs computationally efficiently. Possible future work includes expanding the use of posterior sampling algorithm in a wider class of structured pure-exploration problems.

References

Degenne, R. and Koolen, W. M. (2019). Pure exploration with multiple correct answers. In Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, volume 32.

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

Garivier, A. and Kaufmann, E. (2016). Optimal best arm identification with fixed confidence. In Proceedings of the 29th Conference on Learning Theory, COLT 2016, volume 49 of JMLR Workshop and Conference Proceedings, pages 998–1027. JMLR.org.

Helal, K. M., Taylor, J. N., Cahyadi, H., Okajima, A., Tabata, K., Itoh, Y., Tanaka, H., Fujita, K., Harada, Y., and Komatsuzaki, T. (2019). Raman spectroscopic histology using machine learning for nonalcoholic fatty liver disease. FEBS Letters, 593(18):2535–2544.

Huang, R., Ajallooeian, M. M., Szepesvári, C., and Müller, M. (2017). Structured best arm identification with fixed confidence. CoRR, abs/1706.05198.

Kalyanakrishnan, S., Tewari, A., Auer, P., and Stone, P. (2012). PAC subset selection in stochastic multi-armed bandits. In Proceedings of the 29th International Conference on Machine Learning, ICML 2012.

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

Kaufmann, E. and Koolen, W. M. (2021). Mixture martingales revisited with applications to sequential tests and confidence intervals. J. Mach. Learn. Res., 22:246:1–246:44.

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

Komiyama, J., Honda, J., and Nakagawa, H. (2015). Regret lower bound and optimal algorithm in finite stochastic partial monitoring. In Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, pages 1792–1800.

Komiyama, J., Honda, J., and Nakagawa, H. (2016). Copeland dueling bandit problem: Regret lower bound, optimal algorithm, and computationally efficient algorithm. In Proceedings of the 33nd International Conference on Machine Learning, ICML 2016, volume 48 of JMLR Workshop and Conference Proceedings, pages 1235–1244. JMLR.org.

Lattimore, T. and Szepesvári, C. (2017). The end of optimism? an asymptotic analysis of finite-armed linear bandits. In Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, AISTATS 2017, volume 54 of Proceedings of Machine Learning Research, pages 728–737. PMLR.

Locatelli, A., Gutzeit, M., and Carpentier, A. (2016). An optimal algorithm for the thresholding bandit problem. CoRR, abs/1605.08671.

Magureanu, S., Combes, R., and Proutière, A. (2014). Lipschitz bandits: Regret lower bound and optimal algorithms. In Proceedings of The 27th Conference on Learning Theory, COLT 2014, volume 35 of JMLR Workshop and Conference Proceedings, pages 975–999. JMLR.org.

Russo, D. (2016). Simple bayesian algorithms for best arm identification. In 29th Annual Conference on Learning Theory, volume 49 of Proceedings of Machine Learning Research, pages 1417–1418. PMLR.

Shang, X., de Heide, R., Ménard, P., Kaufmann, E., and Valko, M. (2020). Fixed-confidence guarantees for bayesian best-arm identification. In The 23rd International Conference on Artificial Intelligence and Statistics, AISTATS 2020, volume 108 of Proceedings of Machine Learning Research, pages 1823–1832. PMLR.

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

Tabata, K., Nakumura, A., and Komatsuzaki, T. (2021). Classification bandits: Classification using expected rewards as imperfect discriminators. In Pacific-Asia Conference on Knowledge Discovery and Data Mining, pages 57–69. Springer.

Wang, P., Tzeng, R., and Proutière, A. (2021). Fast pure exploration via frank-wolfe. In Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, pages 5810–5821.

Xu, Y., Chen, X., Singh, A., and Dubrawski, A. (2020). Thresholding bandit problem with both duels and pulls. In The 23rd International Conference on Artificial Intelligence and Statistics, AISTATS 2020, volume 108 of Proceedings of Machine Learning Research, pages 2591–2600. PMLR.

Supplementary Material: Posterior Tracking Algorithm for Classification Bandits

A NOTATION

The following table summarizes our notation.

symbol	definition
K	number of the arms
i_t	arm drawn at round t
$[K]$	$= \{1, 2, \ldots, K\}$
$\boldsymbol{\mu} \in (0,1)^K$	true model parameters
μ_i	i-th component of $\boldsymbol{\mu}$
$\hat{\boldsymbol{\mu}}(t) \in [0,1]^K$	empirical means
$\hat{\mu}_i(t)$	i-th component of $\hat{\boldsymbol{\mu}}(t)$
$\boldsymbol{\theta}(t) \in (0,1)^K$	posterior sample
$\theta_i(t)$	i-th component of $\boldsymbol{\theta}(t)$
$\delta \in (0, 1/2)$	required confidence level
$d(p,q)$	$= p \log(p/q) + (1-p) \log((1-p)/(1-q))$
$\hat{M}(\boldsymbol{\nu})$	$= \lvert\{i \in [K] : \nu_i \geq \xi\}\rvert$
M	$= \hat{M}(\boldsymbol{\mu})$
L	model parameter (if $M \geq L$, then the model is identified as positive)
$\mathcal{I}$	$= \{i \in [K] : S_i(\boldsymbol{\mu}) > 0\}$
$L^* \in \{L, L+1, L+2, \ldots, M\}$	$= \lvert\mathcal{I}\rvert$
$N(C_G, \delta)$	$\min(C_G, C_{\min}(\delta))$
$\beta(t, \delta)$	anytime confidence bound (equation 6)
$D_{\min} > 0$	a sufficiently small constant (Remark 4)
$\Theta_{D_{\min}} \subset (0,1)^K$	defined in Remark 4
$\Theta_{D_{\min},\mathcal{I}} \subset \Theta_{D_{\min}}$	defined in Definition 3
$C_{\text{bnd}} > 0$	a constant such that the algorithm stops before $C_{\text{bnd}} \log(1/\delta)$ (defined in Lemma 11)
$Q^*(\boldsymbol{\mu})$	$= \sum_i S_i(\boldsymbol{\mu})$

B LEMMAS

B.1 Proof of Proposition 2

Proof of Proposition 2. First, we consider the following optimization problem instead of Problem (2):

$$\begin{aligned} &\min_{\boldsymbol{S}=(S_1,S_2,\dots,S_K)} \sum_{i=1}^K S_i \\ &\text{subject to} \\ &\inf_{\boldsymbol{\mu}'\in\mathrm{Alt}(\boldsymbol{\mu})} \sum_{i=1}^K S_i d(\mu_i,\mu_i') \geq 1 \\ &S_i \geq 0 \ \ (i=1,\dots,K). \end{aligned} \tag{30}$$

We show that $w_i^* = S_i^*/\sum_{i=1}^K S_i^*$ $(i=1,\dots,K)$ holds for the solution $S_1^*,\dots,S_K^*$ of this problem, and

$$\sum_{i=1}^K S_i^* = \left(\max_{\boldsymbol{w}\in\Delta_K}\inf_{\boldsymbol{\mu}'\in\mathrm{Alt}(\boldsymbol{\mu})}\sum_{i=1}^K w_i d(\mu_i,\mu_i')\right)^{-1} \tag{31}$$

also holds for the optimal value $\sum_{i=1}^K S_i^*$ of this problem.

The solution $S_1^*,\dots,S_K^*$ of (30) satisfies $\inf_{\boldsymbol{\mu}'\in\mathrm{Alt}(\boldsymbol{\mu})}\sum_{i=1}^K S_i^* d(\mu_i,\mu_i') = 1$ because if $\inf_{\boldsymbol{\mu}'\in\mathrm{Alt}(\boldsymbol{\mu})}\sum_{i=1}^K S_i^* d(\mu_i,\mu_i') > 1$, then $\sum_{i=1}^K S_i$ defined by $S_i = \frac{S_i^*}{\inf_{\boldsymbol{\mu}'\in\mathrm{Alt}(\boldsymbol{\mu})}\sum_{i=1}^K S_i^* d(\mu_i,\mu_i')}$ $(i=1,\dots,K)$ is smaller than $\sum_{i=1}^K S_i^*$, which contradicts the fact that $S_1^*,\dots,S_K^*$ is a solution of (30). Furthermore, there are no $S_1,\dots,S_K$ with $\sum_{i=1}^K S_i = \sum_{i=1}^K S_i^*$ that satisfies $\inf_{\boldsymbol{\mu}'\in\mathrm{Alt}(\boldsymbol{\mu})}\sum_{i=1}^K S_i d(\mu_i,\mu_i') > 1$ by the similar reason. This means that $T^* = \sum_{i=1}^K S_i^*$ is a solution of equation

$$\max_{\boldsymbol{S}\in T\Delta_K}\inf_{\boldsymbol{\mu}'\in\mathrm{Alt}(\boldsymbol{\mu})}\sum_{i=1}^K S_i d(\mu_i,\mu_i') = 1 \tag{32}$$

for T, and $S_1^*,\dots,S_K^*$ is the solution of $\max_{\boldsymbol{S}\in T^*\Delta_K}\inf_{\boldsymbol{\mu}'\in\mathrm{Alt}(\boldsymbol{\mu})}\sum_{i=1}^K S_i d(\mu_i,\mu_i')$. Since $\boldsymbol{\mu}' \neq \boldsymbol{\mu}$ for $\boldsymbol{\mu}' \in \mathrm{Alt}(\boldsymbol{\mu})$, $f(T) = \max_{\boldsymbol{S}\in T\Delta_K}\inf_{\boldsymbol{\mu}'\in\mathrm{Alt}(\boldsymbol{\mu})}\sum_{i=1}^K S_i d(\mu_i,\mu_i')$ is a strictly increasing function of T for $T > 0$, the solution T^* of Eq. (32) coincides with the optimal value $\sum_{i=1}^K S_i^*$ of (30), and the solution $S_1',\dots,S_K'$ of $\max_{\boldsymbol{S}\in T^*\Delta_K}\inf_{\boldsymbol{\mu}'\in\mathrm{Alt}(\boldsymbol{\mu})}\sum_{i=1}^K S_i d(\mu_i,\mu_i')$ is also a solution of (30). Thus, solving optimization problem (30) for the optimal value $\sum_{i=1}^K S_i^*$ and the solution $S_1^*,\dots,S_K^*$ is equivalent to solving Eq. (32) for the solution $T^*, S_1^*,\dots,S_K^*$.

Trivially, the solution $w_1^*,\dots,w_K^*$ of $\max_{\boldsymbol{w}\in\Delta_K}\inf_{\boldsymbol{\mu}'\in\mathrm{Alt}(\boldsymbol{\mu})}\sum_{i=1}^K w_i d(\mu_i,\mu_i')$ and the solution $T^*, S_1^*,\dots,S_K^*$ of (32) have a relation $w_i^* = S_i^*/T^*$, and

$$T^*\max_{\boldsymbol{w}\in\Delta_K}\inf_{\boldsymbol{\mu}'\in\mathrm{Alt}(\boldsymbol{\mu})}\sum_{i=1}^K w_i d(\mu_i,\mu_i') = \max_{S_i\in T^*\Delta_K}\inf_{\boldsymbol{\mu}'\in\mathrm{Alt}(\boldsymbol{\mu})}\sum_{i=1}^K S_i d(\mu_i,\mu_i') = 1$$

holds, which implies Eq. (31).

Next, we show the equivalence of Problems (2) and (30). To show the equivalence, we prove the following equation:

$$\inf_{\boldsymbol{\mu}'\in\mathrm{Alt}(\boldsymbol{\mu})}\sum_{i=1}^K S_i d(\mu_i,\mu_i') = \min_{\substack{U\subseteq[M],\\ |U|=M-L+1}}\sum_{i\in U} S_{(i)} d(\mu_{(i)},\xi) \tag{33}$$

for $M = \hat{M}(\boldsymbol{\mu})$. Note that (i) is the index j of the ith largest parameter μ_j, that is, $\mu_{(1)} \geq \mu_{(2)} \geq \cdots \geq \mu_{(K)}$.

Let $\mathbb{U}$ denote a family of subsets U of $[K]$ with $|U| \geq K - L + 1$. Define $V(\boldsymbol{\mu}')$ as $V(\boldsymbol{\mu}') = \{i \mid \mu_i' < \xi\}$. Then,

$$\mathrm{Alt}(\boldsymbol{\mu}) = \bigcup_{U \in \mathbb{U}} \{\boldsymbol{\mu}' \mid V(\boldsymbol{\mu}') = U\}$$

holds. Let $\mathbb{U}_M = \{U \mid |U| = M - L + 1, U \subseteq \{(i) \mid i \in [M]\}\}$. Then, Eq. (33) is rewritten as

$$\inf_{\boldsymbol{\mu}' \in \mathrm{Alt}(\boldsymbol{\mu})} \sum_{i=1}^{K} S_i d(\mu_i, \mu_i') = \min_{U \in \mathbb{U}_M} \sum_{i \in U} S_i d(\mu_i, \xi). \tag{34}$$

Note that $\overline{\mathbb{U}}_M \coloneqq \{U \subseteq [K] \mid U = U_M \cup \{i \mid \mu_i < \xi\} \text{ for some } U_M \in \mathbb{U}_M\}$ are a subset of $\mathbb{U}$.

Let $D_{\boldsymbol{S}}(\boldsymbol{\mu}') = \sum_{i=1}^{K} S_i d(\mu_i, \mu_i')$ and define $\boldsymbol{\mu}_U$ as $\boldsymbol{\mu}_U = \arg\inf_{\boldsymbol{\mu}' : V(\boldsymbol{\mu}') = U} D_{\boldsymbol{S}}(\boldsymbol{\mu}')$. Then, for any $\boldsymbol{S} = (S_1, \ldots, S_K) : S_1, \ldots, S_K \geq 0$,

$$(\mu_U)_i = \begin{cases} \mu_i & (\mu_i < \xi, i \in U \text{ or } \mu_i \geq \xi, i \notin U) \\ \xi & (\mu_i \geq \xi, i \in U \text{ or } \mu_i < \xi, i \notin U) \end{cases}$$

holds. Thus, for any $U \in \mathbb{U}$

$$\inf_{\boldsymbol{\mu}' : V(\boldsymbol{\mu}') = U} D_{\boldsymbol{S}}(\boldsymbol{\mu}') = D_{\boldsymbol{S}}(\boldsymbol{\mu}_U) = \sum_{\substack{i:\ \mu_i \geq \xi, i \in U \text{ or} \\ \mu_i < \xi, i \notin U}} S_i d(\mu_i, \xi)$$

holds. Let $U' \in \overline{\mathbb{U}}_M$ be $U' = U_M \cup \{i \mid \mu_i < \xi\}$ with $U_M \in \mathbb{U}_M$ satisfying $U_M \subseteq \{i \mid \mu_i \geq \xi, i \in U\}$, then

$$\begin{aligned} D_{\boldsymbol{S}}(\boldsymbol{\mu}_U) &= \sum_{\substack{i:\ \mu_i \geq \xi, i \in U \text{ or} \\ \mu_i < \xi, i \notin U}} S_i d(\mu_i, \xi) \\ &\geq \sum_{i : \mu_i \geq \xi, i \in U} S_i d(\mu_i, \xi) \\ &\geq \sum_{i \in U_M} S_i d(\mu_i, \xi) = D_{\boldsymbol{S}}(\boldsymbol{\mu}_{U'}) \end{aligned}$$

holds. Therefore,

$$\inf_{\boldsymbol{\mu}' \in \mathrm{Alt}(\boldsymbol{\mu})} \sum_{i=1}^{K} S_i d(\mu_i, \mu_i') = \min_{U \in \mathbb{U}} \inf_{\boldsymbol{\mu}' : V(\boldsymbol{\mu}') = U} D_{\boldsymbol{S}}(\boldsymbol{\mu}') = \min_{U \in \overline{\mathbb{U}}_M} D_{\boldsymbol{S}}(\boldsymbol{\mu}_U) = \min_{U_M \in \mathbb{U}_M} \sum_{i \in U_M} S_i d(\mu_i, \xi)$$

holds. □

B.2 Proof of Theorem 1

Proposition 12. *Let L and M be positive integers with $0 < L \leq M$. For $0 < C_1 \leq C_2 \leq \cdots \leq C_M$, define l^* as*

$$l^* = \operatorname*{arg\,min}_{l \in \{L, L+1, \ldots, M\}} \frac{1}{l - L + 1} \sum_{i=1}^{l} C_i.$$

Then, the followings hold. If l^* is unique, Ineq. (1),(2) and (3) hold strictly.

(1) $\frac{1}{l^* - L + 1} \sum_{i=1}^{l^*} C_i \geq C_l$ *for* $l = 1, \ldots, l^*$

(2) $\frac{1}{l^* - L + 1} \sum_{i=1}^{l^*} C_i \leq C_l$ *for* $l = l^* + 1, \ldots, M$

(3) $\frac{1}{l^* - L + 1 - |U|} \sum_{i \in [l^*] \setminus U} C_i \geq C_l$ *for* $U \subseteq [l^*]$ *with* $|U| \leq l^* - L$ *and* $l = 1, \ldots, l^*$

(4) For unique $l^* < M$,

$$\begin{aligned} l^* < l \Leftrightarrow \exists l' &\in \{L, L+1, \dots, l-1\} \\ s.t.\ & \frac{1}{l'-L+1} \sum_{i=1}^{l'} C_j < C_l. \end{aligned} \tag{35}$$

Proof of Proposition 12. Unique-l^* case proofs for the strict versions of Ineq. (1),(2) and (3) can be obtained from the following proofs by replacing "$\leq$" and "$\geq$" with "$<$" and "$>$", respectively.

(Proof of (1)) Ineq. (1) trivially holds for $l^* = L$. Assume that $L < l^*$. By the definition of l^*,

$$\begin{aligned} \frac{1}{l^*-L} \sum_{i=1}^{l^*-1} C_i \geq \frac{1}{l^*-L+1} \sum_{i=1}^{l^*} C_i &\Leftrightarrow \sum_{i=1}^{l^*-1} C_i \geq (l^*-L) C_{l^*} \\ &\Leftrightarrow \sum_{i=1}^{l^*} C_i \geq (l^*-L+1) C_{l^*} \\ &\Leftrightarrow \frac{1}{l^*-L+1} \sum_{i=1}^{l^*} C_i \geq C_{l^*} \\ &\Leftrightarrow \frac{1}{l^*-L+1} \sum_{i=1}^{l^*} C_i \geq C_l \quad (l = 1, \dots, l^*) \end{aligned}$$

holds.

(Proof of (2)) If $l^* = M$, there exists nothing to prove, thus we can assume $l^* < M$. By the definition of l^*,

$$\begin{aligned} \frac{1}{l^*-L+1} \sum_{i=1}^{l^*} C_i \leq \frac{1}{l^*-L+2} \sum_{i=1}^{l^*+1} C_i &\Leftrightarrow \sum_{i=1}^{l^*} C_i \leq (l^*-L+1) C_{l^*+1} \\ &\Leftrightarrow \frac{1}{l^*-L+1} \sum_{i=1}^{l^*} C_i \leq C_{l^*+1} \\ &\Leftrightarrow \frac{1}{l^*-L+1} \sum_{i=1}^{l^*} C_i \leq C_l \quad (l = l^*+1, \dots, M) \end{aligned}$$

holds.

(Proof of (3)) By Ineq. (1),

$$\begin{aligned} \frac{1}{l^*-L+1} \sum_{i=1}^{l^*} C_i \geq C_{l^*} &\Leftrightarrow \sum_{i=1}^{l^*} C_i \geq (l^*-L+1) C_{l^*} \\ &\Leftrightarrow \sum_{i \in [l^*] \setminus U} C_i \geq (l^*-L+1-|U|) C_{l^*} + \left(|U| C_{l^*} - \sum_{i \in U} C_i \right) \\ &\qquad\qquad \text{for } U \subseteq [l^*] \text{ with } |U| \leq l^* - L \\ &\Leftrightarrow \sum_{i \in [l^*] \setminus U} C_i \geq (l^*-L+1-|U|) C_{l^*} \quad \text{for } U \subseteq [l^*] \text{ with } |U| \leq l^* - L \\ &\Leftrightarrow \frac{1}{l^*-L+1-|U|} \sum_{i \in [l^*] \setminus U} C_i \geq C_{l^*} \quad \text{for } U \subseteq [l^*] \text{ with } |U| \leq l^* - L \\ &\Leftrightarrow \frac{1}{l^*-L+1-|U|} \sum_{i \in [l^*] \setminus U} C_i \geq C_l \ (l = 1, \dots, l^*) \\ &\qquad\qquad \text{for } U \subseteq [l^*] \text{ with } |U| \leq l^* - L \end{aligned}$$

holds.

(Proof of (4)) ($\Rightarrow$) Assume $l^* < l$. By the strict version of Ineq. (2),

$$\frac{1}{l^* - L + 1} \sum_{i=1}^{l^*} C_i < C_{l^*+1}$$

holds. Thus, for $l' = l^* < l$, $\frac{1}{l'-L+1} \sum_{i=1}^{l'} C_j < C_l$ holds.

($\Leftarrow$) Assume that, for some $l' \in \{L, L+1, \dots, l-1\}$,

$$\frac{1}{l' - L + 1} \sum_{i=1}^{l'} C_j < C_l$$

holds. Then, by the minimality for l^*,

$$\frac{1}{l^* - L + 1} \sum_{i=1}^{l^*} C_j < C_l$$

holds. Assume $l^* \geq l$. Then, Ineq. (1) holds, which is a contradiction to the above inequality. Therefore $l^* < l$. □

Proof of Theorem 1. We show that the solution of equation

$$\max_{\boldsymbol{S} \in T\Delta_K} \min_{\substack{U \subset [\hat{M}(\boldsymbol{\nu})], \\ |U| = \hat{M}(\boldsymbol{\nu}) - L + 1}} \sum_{i \in U} S_{(i)} d(\nu_{(i)}, \xi) = 1, \tag{36}$$

is $T = Q^*(\boldsymbol{\nu})$ and $\boldsymbol{S} = \boldsymbol{S}(\boldsymbol{\nu})$, where Eq. (36)'s equivalence to Problem (2) can be shown as with Eq. (32)'s equivalence to problem (30). We only show that the solution of

$$\max_{\boldsymbol{S} \in Q^*(\boldsymbol{\nu})\Delta_K} \min_{\substack{U \subset [\hat{M}(\boldsymbol{\nu})], \\ |U| = \hat{M}(\boldsymbol{\nu}) - L + 1}} \sum_{i \in U} S_{(i)} d(\nu_{(i)}, \xi) \tag{37}$$

is $\boldsymbol{S}(\boldsymbol{\nu})$ because if so, the fact

$$\min_{\substack{U \subset [\hat{M}(\boldsymbol{\nu})], \\ |U| = \hat{M}(\boldsymbol{\nu}) - L + 1}} \sum_{i \in U} S(\boldsymbol{\nu})_{(i)} d(\nu_{(i)}, \xi) = \min_{\substack{U \subset [\hat{M}(\boldsymbol{\nu})], \\ |U| = \hat{M}(\boldsymbol{\nu}) - L + 1}} \sum_{i \in U \cap [L^*(\boldsymbol{\nu})]} \frac{1}{L^*(\boldsymbol{\nu}) - L + 1} = 1$$

implies that $T = Q^*(\boldsymbol{\nu})$ is the solution of Eq. (36). Note that $Q^*(\boldsymbol{\nu}) = \frac{1}{L^*(\boldsymbol{\nu})-L+1} \sum_{i=1}^{L^*(\boldsymbol{\nu})} \frac{1}{d(\nu_{(i)},\xi)}$ can be calculated without solving Eq. (36) for T.

Let

$$Q^*(\boldsymbol{\nu})\Delta_K^{\ell} = \{\boldsymbol{S} \in Q^*(\boldsymbol{\nu})\Delta_K \mid S_{(\ell+1)} = \cdots = S_{(K)} = 0\}.$$

We first prove

$$\max_{\boldsymbol{S} \in Q^*(\boldsymbol{\nu})\Delta_K} \min_{\substack{U \subset [\hat{M}(\boldsymbol{\nu})], \\ |U| = \hat{M}(\boldsymbol{\nu}) - L + 1}} \sum_{i \in U} S_{(i)} d(\nu_{(i)}, \xi) = \max_{\boldsymbol{S} \in Q^*(\boldsymbol{\nu})\Delta_K^{L^*(\boldsymbol{\nu})}} \min_{\substack{U \subset [\hat{M}(\boldsymbol{\nu})], \\ |U| = \hat{M}(\boldsymbol{\nu}) - L + 1}} \sum_{i \in U} S_{(i)} d(\nu_{(i)}, \xi). \tag{38}$$

Since $\boldsymbol{S}$ that maximizes $\min_{U \subset [\hat{M}(\boldsymbol{\nu})], |U| = \hat{M}(\boldsymbol{\nu}) - L + 1} \sum_{i \in U} S_{(i)} d(\mu_{(i)}, \xi)$ is trivially in $Q^*(\boldsymbol{\mu})\Delta_K^{\hat{M}(\boldsymbol{\nu})}$, we show

$$\max_{\boldsymbol{S} \in Q^*(\boldsymbol{\nu})\Delta_K^{\hat{M}(\boldsymbol{\nu})}} \min_{\substack{U \subset [\hat{M}(\boldsymbol{\nu})], \\ |U| = \hat{M}(\boldsymbol{\nu}) - L + 1}} \sum_{i \in U} S_{(i)} d(\nu_{(i)}, \xi) = \max_{\boldsymbol{S} \in Q^*(\boldsymbol{\nu})\Delta_K^{L^*(\boldsymbol{\nu})}} \min_{\substack{U \subset [\hat{M}(\boldsymbol{\nu})], \\ |U| = \hat{M}(\boldsymbol{\nu}) - L + 1}} \sum_{i \in U} S_{(i)} d(\nu_{(i)}, \xi). \tag{39}$$

In the case with $L^*(\boldsymbol{\nu}) = \hat{M}(\boldsymbol{\nu})$, Eq. (39) trivially holds, thus we assume that $L^*(\boldsymbol{\nu}) < \hat{M}(\boldsymbol{\nu})$ holds. For any $\boldsymbol{S} \in Q^*(\boldsymbol{\nu})\Delta_K^{\hat{M}(\boldsymbol{\nu})}$, define $\boldsymbol{S}' \in Q^*(\boldsymbol{\nu})\Delta_K^{L^*(\boldsymbol{\nu})}$ as

$$S'_{(i)} = \begin{cases} S_{(i)} + \sum_{j=L^*(\boldsymbol{\nu})+1}^{\hat{M}(\boldsymbol{\nu})} \frac{\frac{1}{L^*(\boldsymbol{\nu})-L+1} \frac{1}{d(\nu_i,\xi)}}{Q^*(\boldsymbol{\nu})} S_{(j)} & ((i) \leq L^*(\boldsymbol{\nu})) \\ 0 & ((i) \geq L^*(\boldsymbol{\nu})+1). \end{cases}$$

Then, for any $U \subset [\hat{M}(\boldsymbol{\nu})]$ with $|U| = \hat{M}(\boldsymbol{\nu}) - L + 1$,

$$\begin{aligned} \sum_{i \in U} S'_{(i)} d(\nu_{(i)}, \xi) &= \sum_{i \in U \cap [L^*(\boldsymbol{\nu})]} \left(S_{(i)} d(\nu_{(i)}, \xi) + \sum_{j=L^*(\boldsymbol{\nu})+1}^{\hat{M}(\boldsymbol{\nu})} \frac{\frac{1}{L^*(\boldsymbol{\nu})-L+1}}{Q^*(\boldsymbol{\nu})} S_{(j)} \right) \\ &= \sum_{i \in U \cap [L^*(\boldsymbol{\nu})]} S_{(i)} d(\nu_{(i)}, \xi) + |U \cap [L^*(\boldsymbol{\nu})]| \sum_{j=L^*(\boldsymbol{\nu})+1}^{\hat{M}(\boldsymbol{\nu})} \frac{\frac{1}{L^*(\boldsymbol{\nu})-L+1}}{Q^*(\boldsymbol{\nu})} S_{(j)} \\ &\geq \sum_{i \in U \cap [L^*(\boldsymbol{\nu})]} S_{(i)} d(\nu_{(i)}, \xi) + \frac{|U \cap [L^*(\boldsymbol{\nu})]|}{L^*(\boldsymbol{\nu}) - L + 1} \sum_{j=L^*(\boldsymbol{\nu})+1}^{\hat{M}(\boldsymbol{\nu})} S_{(j)} d(\nu_{(j)}, \xi) \\ &\geq \sum_{i \in U} S_{(i)} d(\nu_{(i)}, \xi) \end{aligned}$$

holds, where the first inequality holds by inequality $Q^*(\boldsymbol{\nu}) \leq \frac{1}{d(\nu_{(j)},\xi)}$ for $j = L^*(\boldsymbol{\nu}) + 1, \ldots, \hat{M}(\boldsymbol{\nu})$, which is derived from the definition of $L^*(\boldsymbol{\nu})$ and $Q^*(\boldsymbol{\nu})$ using Proposition 12(2), and the second inequality holds by $|U \cap [L^*(\boldsymbol{\nu})]| \geq L^*(\boldsymbol{\nu}) - L + 1$. Therefore $\min_{U \subset [\hat{M}(\boldsymbol{\nu})], |U| = \hat{M}(\boldsymbol{\nu}) - L + 1} \sum_{i \in U} S_{(i)} d(\nu_{(i)}, \xi) \leq \min_{U \subset [\hat{M}(\boldsymbol{\nu})], |U| = \hat{M}(\boldsymbol{\nu}) - L + 1} \sum_{i \in U} S'_{(i)} d(\nu_{(i)}, \xi)$ holds. Thus, Eq. (39) holds.

Finally, we prove

$$\boldsymbol{S}(\boldsymbol{\nu}) \in \operatorname*{arg\,max}_{\boldsymbol{S} \in Q^*(\boldsymbol{\nu})\Delta_K^{L^*(\boldsymbol{\nu})}} \min_{\substack{U \subset [\hat{M}(\boldsymbol{\nu})], \\ |U| = \hat{M}(\boldsymbol{\nu}) - L + 1}} \sum_{i \in U} S_{(i)} d(\nu_{(i)}, \xi).$$

Since

$$\max_{\boldsymbol{S} \in Q^*(\boldsymbol{\nu})\Delta_K^{L^*(\boldsymbol{\nu})}} \min_{\substack{U \subset [\hat{M}(\boldsymbol{\nu})], \\ |U| = \hat{M}(\boldsymbol{\nu}) - L + 1}} \sum_{i \in U} S_{(i)} d(\nu_{(i)}, \xi) = \max_{\boldsymbol{S} \in Q^*(\boldsymbol{\nu})\Delta_K^{L^*(\boldsymbol{\nu})}} \min_{\substack{U \subset [L^*(\boldsymbol{\nu})], \\ |U| = L^*(\boldsymbol{\nu}) - L + 1}} \sum_{i \in U} S_{(i)} d(\nu_{(i)}, \xi)$$

holds trivially, we only have to prove

$$\boldsymbol{S}(\boldsymbol{\nu}) \in \operatorname*{arg\,max}_{\boldsymbol{S} \in Q^*(\boldsymbol{\nu})\Delta_K^{L^*(\boldsymbol{\nu})}} \min_{\substack{U \subset [L^*(\boldsymbol{\nu})], \\ |U| = L^*(\boldsymbol{\nu}) - L + 1}} \sum_{i \in U} S_{(i)} d(\nu_{(i)}, \xi). \tag{40}$$

Let $\boldsymbol{S}$ be any element in $Q^*(\boldsymbol{\nu})\Delta_K^{L^*(\boldsymbol{\nu})}$. Let $i_1, \ldots, i_{L^*(\boldsymbol{\nu})}$ be a permutation of $1, ..., L^*(\boldsymbol{\nu})$ such that $S_{(i_1)} d(\nu_{(i_1)}, \xi) \leq S_{(i_2)} d(\nu_{(i_2)}, \xi) \leq \cdots \leq S_{(i_{L^*(\boldsymbol{\nu})})} d(\nu_{(i_{L^*(\boldsymbol{\nu})})}, \xi)$. Define $\bar{S}^{(c)}_{(i_j)}$ as

$$\bar{S}^{(c)}_{(i_j)} = S(\boldsymbol{\nu})_{(i_j)} + \sum_{h=1}^{c-1} (\bar{S}^{(h)}_{(i_h)} - S_{(i_h)}) \frac{1/d(\nu_{(i_j)}, \xi)}{\sum_{i \in [L^*(\boldsymbol{\nu})] \setminus \{i_1, \ldots, i_h\}} 1/d(\nu_{(i)}, \xi)} \tag{41}$$

for $c = 1, \ldots, L^*(\boldsymbol{\nu}) - L + 1$ and $j \in [L^*(\boldsymbol{\nu})] \setminus \{i_1, \ldots, i_{c-1}\}$. Then,

$$\sum_{j=c}^{L^*(\boldsymbol{\nu})} \bar{S}^{(c)}_{(i_j)} = \sum_{j=1}^{L^*(\boldsymbol{\nu})} S(\boldsymbol{\nu})_{(i_j)} - \sum_{j=1}^{c-1} S_{(i_j)} = Q^*(\boldsymbol{\nu}) - \sum_{j=1}^{c-1} S_{(i_j)} = \sum_{j=c}^{L^*(\boldsymbol{\nu})} S_{(i_j)} \text{ and}$$

$$\bar{S}^{(c)}_{(i_j)} d(\nu_{(i_j)}, \xi) = S(\boldsymbol{\nu})_{(i_j)} d(\nu_{(i_j)}, \xi) + \sum_{h=1}^{c-1} (\bar{S}^{(h)}_{(i_h)} - S_{(i_h)}) \frac{1}{\sum_{i \in [L^*(\boldsymbol{\nu})] \setminus \{i_1, \ldots, i_h\}} 1/d(\nu_{(i)}, \xi)}$$

$$=\frac{1}{L^*(\boldsymbol{\nu})-L+1}+\sum_{h=1}^{c-1}(\bar{S}^{(h)}_{(i_h)}-S_{(i_h)})\frac{1}{\sum_{i\in[L^*(\boldsymbol{\nu})]\setminus\{i_1,\ldots,i_h\}}1/d(\nu_{(i)},\xi)} \quad (j=c,\ldots,L^*(\boldsymbol{\nu}))$$

hold for $c=1,\ldots,L^*(\boldsymbol{\nu})-L+1$. For $\bar{S}^{(c)}_{(i_c)}$ defined by Eq. (41), $S_{(i_c)}d(\nu_{(i_c)},\xi)\leq\bar{S}^{(c)}_{(i_c)}d(\nu_{(i_c)},\xi)$, that is, $S_{(i_c)}\leq\bar{S}^{(c)}_{(i_c)}$ holds because $\sum_{j=c}^{L^*(\boldsymbol{\nu})}\bar{S}^{(c)}_{(i_j)}<\sum_{j=c}^{L^*(\boldsymbol{\nu})}S_{(i_j)}$ holds otherwise, which contradicts the fact $\sum_{j=c}^{L^*(\boldsymbol{\nu})}\bar{S}^{(c)}_{(i_j)}=\sum_{j=c}^{L^*(\boldsymbol{\nu})}S_{(i_j)}$. Therefore,

$$\begin{aligned}
\sum_{j=1}^{L^*(\boldsymbol{\nu})-L+1}S_{(i_j)}d(\nu_{(i_j)},\xi)=&\sum_{j=1}^{L^*(\boldsymbol{\nu})}(\bar{S}^{(j)}_{(i_j)}-(\bar{S}^{(j)}_{(i_j)}-S_{(i_j)}))d(\nu_{(i_j)},\xi)\\
=&\sum_{j=1}^{L^*(\boldsymbol{\nu})-L+1}\bar{S}^{(j)}_{(i_j)}d(\nu_{(i_j)},\xi)-\sum_{j=1}^{L^*(\boldsymbol{\nu})-L+1}(\bar{S}^{(j)}_{(i_j)}-S_{(i_j)})d(\nu_{(i_j)},\xi)\\
=&1+\sum_{j=1}^{L^*(\boldsymbol{\nu})-L+1}(\bar{S}^{(j)}_{(i_j)}-S_{(i_j)})\frac{L^*(\boldsymbol{\nu})-L+1-j}{\sum_{i\in[L^*(\boldsymbol{\nu})]\setminus\{i_1,\ldots,i_j\}}1/d(\nu_{(i)},\xi)}\\
&-\sum_{j=1}^{L^*(\boldsymbol{\nu})-L+1}(\bar{S}^{(j)}_{(i_j)}-S_{(i_j)})d(\nu_{(i_j)},\xi)\\
\leq&1+\sum_{j=1}^{L^*(\boldsymbol{\nu})-L}(\bar{S}^{(j)}_{(i_j)}-S_{(i_j)})d(\nu_{(i_j)},\xi)-\sum_{j=1}^{L^*(\boldsymbol{\nu})-L+1}(\bar{S}^{(j)}_{(i_j)}-S_{(i_j)})d(\nu_{(i_j)},\xi)\\
\leq&1=\sum_{j=1}^{L^*(\boldsymbol{\nu})-L+1}S(\boldsymbol{\nu})_{(i_j)}d(\nu_{(i_j)},\xi)
\end{aligned}$$

holds, where the first inequality is derived using inequality $\sum_{i\in[L^*(\boldsymbol{\nu})]\setminus\{i_1,\ldots,i_j\}}1/d(\nu_{(i)},\xi)>\frac{L^*(\boldsymbol{\nu})-L+1-j}{d(\nu_{(i_j)},\xi)}$ for $j=1,\ldots,L^*(\boldsymbol{\nu})-L$, which can be proved by Proposition 12 (3) and the definition of $L^*(\boldsymbol{\nu})$. Thus, Expression (40) holds. □

B.3 Lemmas: Posterior sample probability

Lemma 13. (Anti-concentration) *Let* $\mu\in(0,1)$ *and* $X\sim\mathrm{Beta}(1+n\mu,1+n(1-\mu))$. *Then,*

$$\mathbb{P}\left[u\geq X\geq l\right]\geq\frac{(u-l)\sqrt{n}}{25}\min_{\nu\in[l,u]}e^{-nd(\mu,\nu)}. \tag{42}$$

Lemma 13 is a useful anti-concentration result when $u-l$ is small.

Proof of Lemma 13. We have

$$\begin{aligned}
&\mathbb{P}\left[u\geq X\geq l\right] &(43)\\
&:=\frac{1}{B(1+n\mu,1+n(1-\mu))}\int_l^u x^{n\mu}(1-x)^{n(1-\mu)}dx &(44)\\
&\geq\frac{u-l}{B(1+n\mu,1+n(1-\mu))}\min_{\nu\in[l,u]}\nu^{n\mu}(1-\nu)^{n(1-\mu)} &(45)\\
&=\frac{(u-l)\Gamma(2+n)}{\Gamma(1+n\mu)\Gamma(1+n(1-\mu))}\min_{\nu\in[l,u]}\nu^{n\mu}(1-\nu)^{n(1-\mu)}\qquad\text{(by definition)} &(46)\\
&\geq\frac{(u-l)}{e^{1/6}\sqrt{2\pi}}\frac{(n+2)^{n+3/2}}{(n\mu+1)^{n\mu+1/2}(n(1-\mu)+1)^{n(1-\mu)+1/2}}\min_{\nu\in[l,u]}\nu^{n\mu}(1-\nu)^{n(1-\mu)}, &(47)
\end{aligned}$$

where, in the last transformation we used the Stirling's formula

$$\sqrt{2\pi}\leq\frac{\Gamma(z)}{z^{z-1/2}e^{-z}}\leq\sqrt{2\pi}e^{1/12}.$$

Moreover, for any ν we have

$$\frac{(u-l)}{e^{1/6}\sqrt{2\pi}}\frac{(n+2)^{n+3/2}}{(n\mu+1)^{n\mu+1/2}(n(1-\mu)+1)^{n(1-\mu)+1/2}}\nu^{n\mu}(1-\nu)^{n(1-\mu)} \tag{48}$$

$$\geq \frac{(u-l)}{e^{1/6}\sqrt{2\pi}}\sqrt{\frac{(n+2)^3}{(n\mu+1)(n(1-\mu)+1)}}\frac{e^{-nd(\mu,\nu)}}{\left(1+\frac{1}{n\mu}\right)^{n\mu}\left(1+\frac{1}{n(1-\mu)}\right)^{n(1-\mu)}} \tag{49}$$

$$\geq \frac{(u-l)}{e^{2+1/6}\sqrt{2\pi}}\sqrt{\frac{(n+2)^3}{(n\mu+1)(n(1-\mu)+1)}}e^{-nd(\mu,\nu)} \tag{50}$$

$$\geq \frac{(u-l)\sqrt{n}}{e^{2+1/6}\sqrt{2\pi}}e^{-nd(\mu,\nu)} \tag{51}$$

$$\text{(by using } (1+x/n)^n \leq e^x) \tag{52}$$

$$\geq \frac{(u-l)\sqrt{n}}{25}e^{-nd(\mu,\nu)}, \tag{53}$$

which completes the proof. □

Lemma 14. (Concentration of beta distribution) *Let* $X \sim \mathrm{Beta}(1+n\mu, 1+n(1-\mu))$*. Then,*

$$\mathbb{P}\left[|X-\mu| \geq \frac{1}{n}+\epsilon\right] \leq 2\exp\left(-\frac{n\epsilon\min(\epsilon, 1/\sqrt{n})}{4}\right) \tag{54}$$

for any $\epsilon > 0$*.*

Proof of Lemma 14. Theorem 1 in Skorski (2021) states that

$$\mathbb{P}\left[|X-\mathbb{E}[X]| \geq \epsilon\right] \leq 2\exp\left(-\frac{\epsilon^2}{2v^2+2c\epsilon}\right), \tag{55}$$

where

$$\alpha = 1+n(1-\mu) \tag{56}$$

$$\beta = 1+n\mu \tag{57}$$

$$v^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \tag{58}$$

$$c = \max\left\{\frac{|\beta-\alpha|}{(\alpha+\beta)(\alpha+\beta+2)}, \sqrt{\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+2)}}\right\}. \tag{59}$$

By using $v \leq 1/n, c \leq 1/\sqrt{n}$, and

$$|\mathbb{E}[X]-\mu| = \left|\frac{1+n\mu}{2+n}-\mu\right| \leq \frac{1}{n},$$

it holds that

$$\mathbb{P}\left[|X-\mu| \geq \frac{1}{n}+\epsilon\right] \leq 2\exp\left(-\frac{n\epsilon\min(\epsilon, 1/\sqrt{n})}{4}\right). \tag{60}$$

□

Note that if $\epsilon \leq 1/\sqrt{n}$ then the equation above is $\exp(-O(n\epsilon^2))$, which is similar to standard concentration inequality of a mean.

Lemma 15. (convergence of $\theta_i(t), \hat{\mu}_i(t)$) *Let* $\delta' \in (0,1)$ *and* $n, T \geq 1$ *be arbitrary. Let arm* $i \in [K]$ *be arbitrary. Then, the following inequality holds:*

$$\mathbb{P}\left[\bigcup_{t:N_i(t)\geq n, t\leq T}\left\{|\theta_i(t)-\mu_i| \geq \frac{6\log(4T/\delta')}{\sqrt{n}} \cup |\hat{\mu}_i(t)-\mu_i| \geq \sqrt{\frac{\log(4T/\delta')}{2n}}\right\}\right] \leq \delta'. \tag{61}$$

Lemma 15 is the mean and posterior convergence that uniformly holds over all rounds.

Proof of Lemma 15. Union bound and Hoeffding's inequality imply

$$\mathbb{P}\left[\bigcup_{t:N_i(t)\geq n, t\leq T}\left\{|\hat{\mu}_i(t)-\mu_i|\geq\sqrt{\frac{\log(4T/\delta')}{2n}}\right\}\right]\leq\sum_{t:N_i(t)\geq n, t\leq T}\mathbb{P}\left[|\hat{\mu}_i(t)-\mu_i|\geq\sqrt{\frac{\log(4T/\delta')}{2n}}\right]\leq T\times 2\times\frac{\delta'}{4T}=\frac{\delta'}{2}. \tag{62}$$

Lemma 14 implies

$$\sum_{t\leq T}\mathbb{P}\left[|\theta_i(t)-\hat{\mu}_i(t)|\geq\frac{1}{N_i(t)}+\frac{4\log(4T/\delta')}{\sqrt{N_i(t)}}\middle| N_i(t)\right]\leq T\times 2\times\frac{\delta'}{4T}=\frac{\delta'}{2}. \tag{63}$$

By using these and $|\theta_i(t)-\mu_i|\leq|\hat{\mu}_i(t)-\mu_i|+|\theta_i(t)-\hat{\mu}_i(t)|$, we have

$$\sum_{t\leq T}\mathbb{P}\left[|\theta_i(t)-\mu_i|\geq\frac{6\log(4T/\delta')}{\sqrt{N_i(t)}}\right]\leq\sum_{t\leq T}\mathbb{P}\left[|\theta_i(t)-\mu_i|\geq\frac{4\log(4T/\delta')}{\sqrt{N_i(t)}}+\frac{1}{N_i(t)}+\sqrt{\frac{\log(4T/\delta')}{2N_i(t)}}\right] \tag{64}$$

$$\leq\frac{\delta'}{2}+\frac{\delta'}{2}\leq\delta'. \tag{65}$$

□

B.4 Proof of Lemma 3

Proof of Lemma 3. Let $\boldsymbol{\nu}\in\Theta_{D_{\min},\mathcal{I}}$, and let $i\notin\mathcal{I}$ and $\nu_i\leq\mu_i+D_{\min}$. Assume that ν_i is the jth largest value in $\{\nu_k \mid k\in[K]\}$, that is, $\nu_{(j)}=\nu_i$. If $\nu_i<\xi$, then $j>\hat{M}(\boldsymbol{\nu})\geq L^*(\boldsymbol{\nu})$. Thus, $S_i(\boldsymbol{\nu})=0$ by Eq. (4). Assume that $\nu_i\geq\xi$. In this case, $j\leq\hat{M}(\boldsymbol{\nu})$ holds. Define $\boldsymbol{\nu}'$ as $\nu'_k=\min\{\nu_k,\mu+D_{\min}\}$. Note that $\nu'_i=\nu_i$. Then, by the definition of $D_{\min}$ (Remark 4), $\hat{\mathcal{I}}(\boldsymbol{\nu}')=\mathcal{I}$ holds[9]. Since $\nu_k\geq\nu'_k$ holds,

$$\frac{1}{L^*(\boldsymbol{\nu})-L+1}\sum_{k=1}^{L^*(\boldsymbol{\nu}')}\frac{1}{d(\nu_{(k)},\xi)}\leq\frac{1}{L^*(\boldsymbol{\nu}')-L+1}\sum_{k=1}^{L^*(\boldsymbol{\nu}')}\frac{1}{d(\nu'_{(k)},\xi)}$$

holds. Thus,

$$\frac{1}{L^*(\boldsymbol{\nu})-L+1}\sum_{k=1}^{L^*(\boldsymbol{\nu})}\frac{1}{d(\nu_{(k)},\xi)}\leq\frac{1}{L^*(\boldsymbol{\nu}')-L+1}\sum_{k=1}^{L^*(\boldsymbol{\nu}')}\frac{1}{d(\nu'_{(k)},\xi)} \tag{66}$$

holds. Since $i\notin\mathcal{I}=\hat{\mathcal{I}}(\boldsymbol{\nu}')$, by the strict version Ineq. (2) of Proposition 12,

$$\frac{1}{L^*(\boldsymbol{\nu}')-L+1}\sum_{k=1}^{L^*(\boldsymbol{\nu}')}\frac{1}{d(\nu'_{(k)},\xi)}<\frac{1}{d(\nu_i,\xi)} \tag{67}$$

holds. Therefore, by combining Ineqs. (66) and (67) and using the fact that $\nu_i=\nu_{(j)}$, we obtain

$$\frac{1}{L^*(\boldsymbol{\nu})-L+1}\sum_{k=1}^{L^*(\boldsymbol{\nu})}\frac{1}{d(\nu_{(k)},\xi)}<\frac{1}{d(\nu_{(j)},\xi)}.$$

Thus, by Proposition 12 (4), $L^*(\boldsymbol{\nu})<j$ holds, which means $S_i(\boldsymbol{\nu})=S_{(j)}(\boldsymbol{\nu})=0$ by Eq. (4). □

B.5 Proof of Lemma 4

Proof of Lemma 4. In the following, we derive the following inequality for any $i\in[K]$:

$$\mathbb{P}\left[\theta_i(t)\in[\mu_i-D_{\min},\mu_i+D_{\min}]\right]\geq C_1(\boldsymbol{\mu})\exp\left(-N_i(t)d(\hat{\mu}_i(t),\mu_i)\right) \tag{68}$$

[9] $\nu_k<\mu_k-D_{\min}$ may hold for some $k\notin\mathcal{I}$, but that does not affect $\boldsymbol{S}(\boldsymbol{\nu})$.

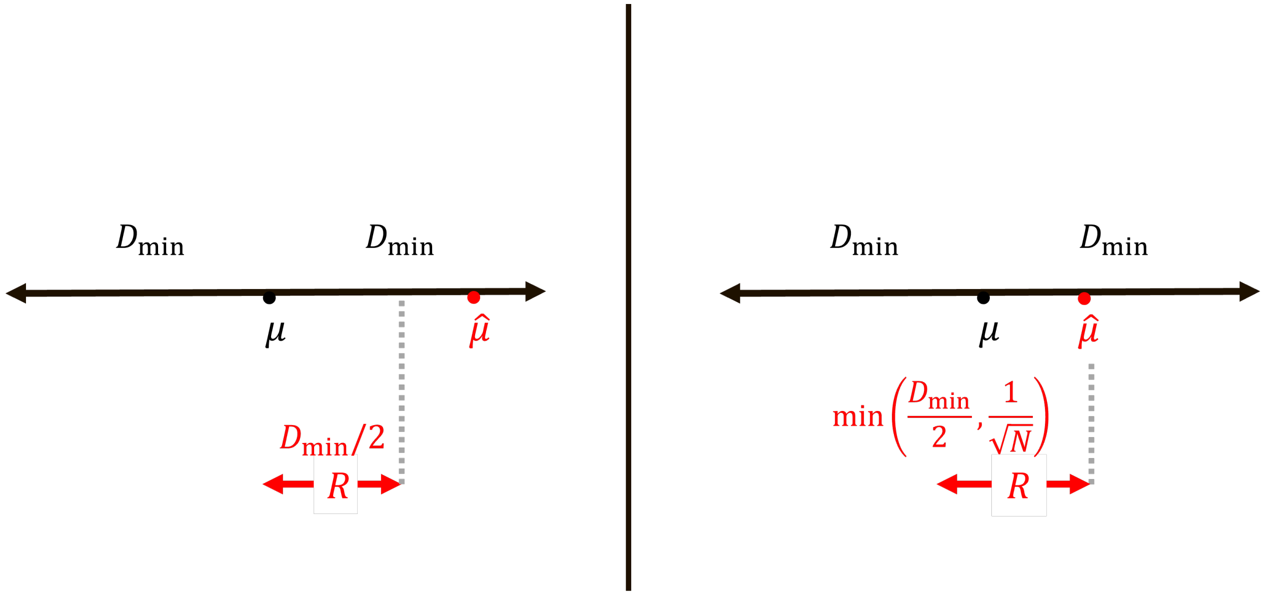


Figure 4: Illustration of Lemma 4 for each i. We consider two cases and derive a bound for each case. The left illustration is the case of $|\hat{\mu}_i(t) - \mu_i| \geq D_{\text{min}}/2$, and the right illustration is the case of $|\hat{\mu}_i(t) - \mu_i| < D_{\text{min}}/2$.

for some distribution-dependent constant $C_1 = C_1(\boldsymbol{\mu}) > 0$. Remark 4 implies that $\theta_i(t) \in (\mu_i - D_{\text{min}}, \mu_i + D_{\text{min}})$ for all $i \in [K]$ suffices to be $\boldsymbol{\theta}(t) \in \Theta_{\mathcal{I}}$. Therefore, multiplying equation 68 for all i immediately yields Lemma 4 with $C = (C_1)^K$.

We derive equation 68 for the two cases separately. Namely, $|\hat{\mu}_i(t) - \mu_i| \geq D_{\text{min}}/2$ or $|\hat{\mu}_i(t) - \mu_i| < D_{\text{min}}/2$.

Case 1 ($|\hat{\mu}_i(t) - \mu_i| \geq D_{\text{min}}/2$): Let R be the line segment defined by two points

$$\mu_i \text{ and } \mu_i - \text{sgn}(\mu_i - \hat{\mu}_i(t)) D_{\text{min}}/2$$

which is a subset of $[\mu_i - D_{\text{min}}, \mu_i + D_{\text{min}}]$ (c.f. Figure 4 left). Therefore,

$$\begin{aligned}
&\mathbb{P}[\theta_i(t) \in [\mu_i - D_{\text{min}}, \mu_i + D_{\text{min}}]] && (69)\\
&\geq \mathbb{P}[\theta_i(t) \in R] && (70)\\
&\geq \left(\frac{(D_{\text{min}}/2)\sqrt{N_i(t)}}{25}\right) \min_{\nu \in R} \exp\left(-N_i(t) d(\hat{\mu}_i(t), \nu_i)\right) \quad \text{(by Lemma 13)} && (71)\\
&= \left(\frac{(D_{\text{min}}/2)\sqrt{N_i(t)}}{25}\right) \exp\left(-N_i(t) d(\hat{\mu}_i(t), \mu_i)\right) \quad \text{(by monotonicity of KL divergence)} && (72)\\
&= \Omega(1) \times \exp\left(-N_i(t) d(\hat{\mu}_i(t), \mu_i)\right). && (73)
\end{aligned}$$

Case 2 ($|\hat{\mu}_i(t) - \mu_i| < D_{\text{min}}/2$): Let the line segment R be defined by two points

$$\hat{\mu}_i(t) \text{ and } \hat{\mu}_i(t) + \text{sgn}(\mu_i - \hat{\mu}_i(t)) \min\left(D_{\text{min}}/2, 1/\sqrt{N_i(t)}\right),$$

which is a subset of $[\mu_i - D_{\text{min}}, \mu_i + D_{\text{min}}]$ (c.f., Figure 4 right). Note that $D_{\text{min}} \leq \mu_i \leq 1 - D_{\text{min}}$ and $|\hat{\mu}_i(t) - \mu|, |\nu - \mu| \leq D_{\text{min}}/2$ hold for an $\nu \in R$, and thus

$$D_{\text{min}}/2 \leq \nu,\ \hat{\mu}_i(t) \leq 1 - D_{\text{min}}/2. \quad (74)$$

Therefore,

$$\mathbb{P}[\theta_i(t) \in [\mu_i - D_{\text{min}}, \mu_i + D_{\text{min}}]] \quad (75)$$

$$\geq \mathbb{P}[\theta_i(t) \in R] \tag{76}$$

$$\geq \left(\frac{\min\left(D_{\min}/2, 1/\sqrt{N_i(t)}\right)\sqrt{N_i(t)}}{25} \right) \min_{\boldsymbol{\nu}\in R} \exp\left(-N_i(t) d(\hat{\mu}_i(t), \nu_i)\right) \qquad \text{(by Lemma 13)} \tag{77}$$

$$= \Omega(1) \times \min_{\boldsymbol{\nu}\in R} \exp\left(-N_i(t) d(\hat{\mu}_i(t), \nu_i)\right) \tag{78}$$

$$\geq \Omega(1) \times \min_{\boldsymbol{\nu}\in R} \exp\left(-N_i(t) \frac{1}{2(D_{\min}/2)(1-D_{\min}/2)} \left(\frac{1}{\sqrt{N_i(t)}} \right)^2 \right) \tag{79}$$

$$\text{(by equation 74 and } d(p,q) \leq (p-q)^2/(2x(1-x)) \text{ for } p,q \in [x, 1-x]) \tag{80}$$

$$= \Omega(1) \times \Omega(1) = \Omega(1). \tag{81}$$

Lemma 16. (Minimum nonzero value, to be deleted) *For any $i, \boldsymbol{\nu}$*

$$S_i(\boldsymbol{\nu}) = 0 \text{ or } S_i(\boldsymbol{\nu})\beta(t,\delta) > C_{\min}(\delta) \tag{82}$$

holds for some $C_{\min}(\delta) = \Theta(\log(1/\delta))$.

It is easy to confirm Lemma 16 with $C_{\min}(\delta) = \min_p \frac{\beta(1,\delta)}{Kd(p,\xi)} = \min\left(\frac{\beta(1,\delta)}{Kd(0,\xi)}, \frac{\beta(1,\delta)}{Kd(1,\xi)} \right)$. Lemma 16 states that, if the arm is included in the solution set, then the number of draw required is at least $C\log(1/\delta)$ for some universal constant $C > 0$. This property contributes to the stability of the solution.

□

B.6 Proof of Lemma 7

In this section, we first propose Lemmas 17 and 18. By using them, we derive Lemma 7.

Let

$$\mathcal{F}(t) = \left\{ \boldsymbol{S}(\boldsymbol{\theta}) \in \Theta_{D_{\min},\mathcal{I}},\, \cap_{j\notin\mathcal{I}} S_j(\boldsymbol{\theta}) = 0 \right\} \tag{83}$$

$$\mathcal{G}(t, C_G) = \left\{ \sum_{s<t} \mathbf{1}\{\mathcal{F}(t)\} \geq \frac{4Q^*(\boldsymbol{\mu})}{\min_i S_i(\boldsymbol{\mu})} C_G + K \right\}. \tag{84}$$

Lemma 17. *Event $\mathcal{G}(t, C_G)$ implies*

$$N_i(t) \geq N(C_G, \delta) := \min(C_G, C_{\min}(\delta)) \tag{85}$$

for all $i \in \mathcal{I}$.

Note that the value $\frac{S_i(\boldsymbol{\mu})}{Q^*(\boldsymbol{\mu})}$ is the minimum ratio of draw of arm i under the optimal solution $\boldsymbol{S}(\boldsymbol{\mu})$. Intuitively speaking, under $\mathcal{H}(t)$, $\boldsymbol{S}(\boldsymbol{\nu})$ is a constant-ratio approximation of $\boldsymbol{S}(\boldsymbol{\mu})$, and thus it draws every arm $i \in \mathcal{I}$ in accordance of the ratio. The minimum with $C_{\min}(\delta)$ in the RHS of equation 85 is derived from the fact that the algorithm attempts to draw any arm $i : S_i(\mu) > 0$ at least $C_{\min}(\delta)$ time by Lemma 16. The following proof make this discussion rigorous.

Proof of Lemma 17. Consider a subsequence $\{s < t : \mathcal{H}(s), s \leq C_{\min}(\delta))\}$ (subset of rounds). Let $N_i^{\mathrm{sub}}(t)$ be the number of draw of arm i assuming the algorithm faces this subsequence. Under $\mathcal{H}(s)$ at round s, $\frac{1}{2} \leq \frac{S_i(\boldsymbol{\theta}(s))}{S_i(\boldsymbol{\mu})}, \frac{S_j(\boldsymbol{\theta}(s))}{S_j(\boldsymbol{\mu})} \leq 2$ holds, and thus

$$\frac{S_i(\boldsymbol{\theta}(s))}{S_j(\boldsymbol{\theta}(s))} \geq \frac{S_i(\boldsymbol{\mu})}{4S_j(\boldsymbol{\mu})}$$

always holds on this subsequence. This implies that,

$$N_i^{\mathrm{sub}}(t) \geq S_i(\boldsymbol{\mu}) \frac{N_j^{\mathrm{sub}}(t)}{4S_j(\boldsymbol{\mu})} - 1$$

holds for any $i, j \in [K]$. From this and $Q^*(\boldsymbol{\mu}) = \sum_j S_j(\boldsymbol{\mu})$ we can obtain

$$N_i^{\mathrm{sub}}(t) \geq \frac{S_i(\boldsymbol{\mu})}{4Q^*(\boldsymbol{\mu})} \times \frac{4Q^*(\boldsymbol{\mu})}{\min_i S_i(\boldsymbol{\mu})} C_G \geq C_G. \tag{86}$$

Next, for any sequence of rounds, adding another round at any position never decreases $N_i^{\mathrm{sub}}(t)$. For two vectors $\boldsymbol{N'}(s), \boldsymbol{N}(s)$ of size $[K]$, we say $\boldsymbol{N'}(s)$ weakly dominates $\boldsymbol{N}(s)$ if $N'_j(s) \geq N_j(s)$ holds for each coordinate $j \in [K]$. Then, it is easy to derive that the weak domination is preserved if we add another round at any position of the subsequence. This implies that $N_i(t) \geq N_i^{\mathrm{sub}}(t)$, which completes the proof.

□

The following lemma bounds the case where non-solution arm j has positive (non-zero) value of $S_j(\boldsymbol{\theta})$.

Lemma 18.

$$\sum_{t \leq C_{\mathrm{bnd}} \log(1/\delta)} \mathbb{P}[t \leq \tau, \mathcal{D}^c(t), \mathcal{H}(t), \mathcal{F}^c(t)] = o(\log(1/\delta)) \tag{87}$$

Proof of Lemma 18.

$$\sum_{t \leq C_{\mathrm{bnd}} \log(1/\delta)} \mathbb{P}\left[t \leq \tau, \mathcal{D}^c(t), \mathcal{H}(t), \mathcal{F}^c(t)\right] \leq \sum_{j \notin \mathcal{I}} \sum_{t \leq C_{\mathrm{bnd}} \log(1/\delta)} \mathbb{P}\left[t \leq \tau, \mathcal{D}^c(t), \mathcal{H}(t), S_j(t) > 0\right] \tag{88}$$

The event $\mathcal{H}(t), S_j(\boldsymbol{\theta}) > 0$ implies $S_j(\boldsymbol{\theta})/Q^*(\boldsymbol{\theta}) \geq 1/K^2$ because otherwise $S_j(\boldsymbol{\theta}) = 0$. By using a subsequence discussion similarly to Lemma 17, we have

$$N_j(t) \geq K^2 C + K^2$$

after the event $t \leq \tau, \mathcal{D}^c(t), \mathcal{H}(t), S_j(t) > 0$ occurs C times. However, after the round

$$t : N_j(t) > \frac{\log(D(\delta))}{d(\mu_j + D_{\min}/2, \xi)} = \Theta(\log\log(1/\delta)),$$

$\mathcal{D}^c(t)$ implies $\hat{\mu}_j(t) \leq \mu_j + D_{\min}/2$, and

$$\mathbb{P}\left[\theta_j(t) \geq \mu_j + D_{\min} \middle| \hat{\mu}_j(t) \leq \mu_j + D_{\min}/2, N_j(t) = \Omega(\log\log(1/\delta))\right] = o(1)$$

by Lemma 14. By using this, we finally have

$$\sum_{t \leq C_{\mathrm{bnd}} \log(1/\delta)} \mathbb{P}\left[t \leq \tau, \mathcal{D}^c(t), \mathcal{H}(t), \mathcal{F}^c(t)\right] \leq K^2 \frac{\log(D(\delta))}{d(\mu_j + D_{\min}/2, \xi)} + K^2 + C_{\mathrm{bnd}} \log(1/\delta) \times o(1) = o(\log(1/\delta)). \tag{89}$$

□

Proof of Lemma 7. Let $\mathcal{G}(t) = \mathcal{G}(t, (\log(1/\delta))^{1/2})$ and $N(\delta) = N((\log(1/\delta))^{1/2}, \delta) = o((\log(1/\delta))^{1/2})$. We have

$$\begin{aligned}
\sum_{t \leq C_{\mathrm{bnd}} \log(1/\delta)} \mathbb{P}[t \leq \tau, \mathcal{D}^c(t), \mathcal{E}^c(t), \mathcal{H}(t)] &\leq \sum_{t \leq C_{\mathrm{bnd}} \log(1/\delta)} \mathbb{P}[t \leq \tau, \mathcal{D}^c(t), \mathcal{E}^c(t), \mathcal{F}(t)] + o(\log(1/\delta)) && (90)\\
&\quad \text{(by Lemma 18)} && (91)\\
&\leq \sum_{t \leq C_{\mathrm{bnd}} \log(1/\delta)} \mathbb{P}[t \leq \tau, \mathcal{D}^c(t), \mathcal{E}^c(t), \mathcal{G}(t)] + o(\log(1/\delta)) && (92)\\
&\leq \sum_{t \leq C_{\mathrm{bnd}} \log(1/\delta)} \mathbb{P}[t \leq \tau, \mathcal{E}^c(t), \mathcal{G}(t)] + o(\log(1/\delta)) && (93)\\
&\leq \sum_{t \leq C_{\mathrm{bnd}} \log(1/\delta)} \mathbb{P}\left[t \leq \tau, \mathcal{E}^c(t), \bigcap_{i \in \mathcal{I}} \{N_i(t) \geq N(\delta)\}\right] + o(\log(1/\delta)) && (94)
\end{aligned}$$

$$\text{(by Lemma 17).} \tag{95}$$

Lemma 15 implies that, with probability $1 - K\delta'$, we have

$$|\theta_i(t) - \mu_i|, |\hat{\mu}_i(t) - \mu_i| \leq \frac{6\log(4C_{\text{bnd}}\log(1/\delta)/\delta')}{\sqrt{N(\delta)}} := V(\delta, \delta') \tag{96}$$

for all $i \in \mathcal{I}, t \leq C_{\text{bnd}} \log(1/\delta)$ such that $N_i(t) \geq N(\delta)$ holds. Here,

$$V(\delta, \delta') = O\left(\frac{\log(\log(1/\delta)/\delta')}{(\log(1/\delta))^{1/2}}\right)$$

Choose δ' to be sufficiently small.[10] This makes $V(\delta, \delta') = o(1)$, and thus for a sufficiently small[11] $\delta > 0$, we have $V(\delta, \delta') \leq D_{\min}$. Inequality $V(\delta, \delta') \leq D_{\min}$ implies $\hat{\boldsymbol{\mu}}(t), \boldsymbol{\theta}(t) \in \Theta_{D_{\min}, \mathcal{I}}$ for all rounds $t < C_{\text{bnd}} \log(1/\delta)$ such that equation 96 and $\mathcal{H}(t)$ hold. Continuity of the solution around $\boldsymbol{S}(\boldsymbol{\mu})$ implies that

$$|S_i(\boldsymbol{\mu}) - S_i(\hat{\boldsymbol{\mu}}(t))|, |S_i(\boldsymbol{\mu}) - S_i(\boldsymbol{\theta}(t))| \leq o(1), \tag{97}$$

for all i, t.

In this case, the algorithm draws arm $i \in \mathcal{I}$ no more than

$$(1 + o(1))S(\boldsymbol{\mu})\beta(C_{\text{bnd}}\log(1/\delta), \delta) = (1 + o(1))S_i(\boldsymbol{\mu})\log(1/\delta) + o(\log(1/\delta)) \tag{98}$$

times because once the algorithm draws arms proportional to $S_i(\boldsymbol{\theta}(t))$ and stops once it draws each arm $i \in \mathcal{I}$ for $S_i(\hat{\boldsymbol{\mu}}(t))\beta(C_{\text{bnd}}\log(1/\delta), \delta) = S_i(\hat{\boldsymbol{\mu}}(t))\log(1/\delta)(1 + o(1))$ times. Moreover, equation 96 holds with probability at least $1 - K\delta' = 1 - o(1)$. By using the discussion above, equation 94 is bounded as:

$$\sum_{t \leq C_{\text{bnd}}\log(1/\delta)} \mathbb{P}\left[t \leq \tau, \mathcal{E}^c(t), \bigcap_{i \in \mathcal{I}} \{N_i(t) \geq N(\delta)\}\right] \leq Q^*(\boldsymbol{\mu})\log(1/\delta) + o(\log(1/\delta)) \tag{99}$$

$$+ C_{\text{bnd}}\log(1/\delta) \times K\delta' \tag{100}$$

$$= Q^*(\boldsymbol{\mu})\log(1/\delta) + o(\log(1/\delta)). \tag{101}$$

□

B.7 Proof of Lemma 8

In the proof, we show that it is unlikely to draw arm $i \notin \mathcal{I}$ at most $o(\log(1/\delta))$ times.

Proof. We have,

$$\sum_{t \leq C_{\text{bnd}}\log(1/\delta)} \mathbb{P}[t \leq \tau, \mathcal{D}^c(t), \mathcal{E}(t), \mathcal{H}(t)] \leq \sum_{i \notin \mathcal{I}} \sum_{t \leq C_{\text{bnd}}\log(1/\delta)} \mathbb{P}[t \leq \tau, \mathcal{D}^c(t), \mathcal{E}_i(t), \mathcal{H}(t)]. \tag{102}$$

Lemma 3 implies that, under $\mathcal{H}(t)$, if $\theta_i(t) \leq \mu_i + D_{\min}$ then arm $i \notin \mathcal{I}$ is not drawn. Let

$$N_{D_{\min}} = \max\left(\frac{2}{D_{\min}}, \frac{36}{D_{\min}^2}\log\left(4C_{\text{bnd}}(\log(1/\delta))^2\right)\right),$$

which is $O(\log\log(1/\delta))$. By using this,

$$\sum_{t \leq C_{\text{bnd}}\log(1/\delta)} \mathbb{P}[t \leq \tau, \mathcal{D}^c(t), \mathcal{E}_i(t), \mathcal{H}(t)] \tag{103}$$

$$\leq \sum_{t \leq C_{\text{bnd}}\log(1/\delta)} \mathbb{P}[t \leq \tau, \mathcal{E}_i(t), \theta_i(t) > \mu_i + D_{\min}] \tag{104}$$

[10] For example, $\delta' = (\log(1/\delta))^{-1}$.

[11] While the discussion here is asymptotic, we can obtain a (involved) finite-time bound if desired.

$$\leq N_{D_{\min}} + \sum_{t \leq C_{\mathrm{bnd}} \log(1/\delta)} \mathbb{P}[t \leq \tau, \theta_i(t) > \mu_i + D_{\min}, N_i(t) \geq N_{D_{\min}}] \tag{105}$$

$$\text{(by } \{\mathcal{E}_i(t), i_t = n\} \text{ occurs at most once)} \tag{106}$$

$$\leq N_{D_{\min}} + C_{\mathrm{bnd}} \log(1/\delta) \times \mathbb{P}\left[\bigcup_{t \leq C_{\mathrm{bnd}} \log(1/\delta)} \{\mathcal{E}_i(t), \mathcal{H}(t), N_i(t) \geq N_{D_{\min}}\}\right] \tag{107}$$

$$\leq N_{D_{\min}} + C_{\mathrm{bnd}} \log(1/\delta) \times \frac{1}{\log(1/\delta)} \tag{108}$$

$$\text{(by Lemma 15 with } \delta' = \log(\delta)) \tag{109}$$

$$\leq N_{D_{\min}} + C_{\mathrm{bnd}} = O(\log\log(1/\delta)) + O(1). \tag{110}$$

□

B.8 Proof of Lemma 9

Proof. We have

$$\sum_{t \leq C_{\mathrm{bnd}} \log(1/\delta)} \mathbb{P}[t \leq \tau, \mathcal{D}(t)] \leq C_{\mathrm{bnd}} \log(1/\delta) \times \mathbb{P}\left[\bigcup_{t < C_{\mathrm{bnd}} \log(1/\delta)} \mathcal{D}(t)\right] \tag{111}$$

$$\leq C_{\mathrm{bnd}} \log(1/\delta) q, \tag{112}$$

where $q = q(\delta)$ be such that $\beta(C_{\mathrm{bnd}} \log(1/\delta), q) = \log(D(\delta))$. By using equation 6,

$$\log(q) = C_1 + \log(C_2 \log(C_{\mathrm{bnd}} \log(1/\delta) + 1)) - \log(D(\delta)), \tag{113}$$

which implies $q^{-1} = O((\log(1/\delta))^{1/2})$, and thus

$$C_{\mathrm{bnd}} \log(1/\delta) q = O((\log(1/\delta))^{1/2}).$$

□

B.9 Proof of Lemma 10

Proof of Lemma 10.

$$\sum_{t \leq C_{\mathrm{bnd}} \log(1/\delta)} \mathbb{P}[t \leq \tau, \mathcal{D}^c(t), \mathcal{H}^c(t)]$$

$$\leq \sum_{t \leq C_{\mathrm{bnd}} \log(1/\delta)} \mathbb{P}[t \leq \tau, \mathcal{D}^c(t), \mathcal{G}^c(t, (\log\log(1/\delta))^4)] + \sum_{t \leq C_{\mathrm{bnd}} \log(1/\delta)} \mathbb{P}[t \leq \tau, \mathcal{D}^c(t), \mathcal{G}(t, (\log\log(1/\delta))^4), \mathcal{H}^c(t)] \tag{114}$$

Lemma 4 states that, conditioned on $\mathcal{D}^c(t)$, with probability at least $C(\boldsymbol{\mu})/D(\delta)$ we have $\boldsymbol{\theta}(t) \in \Theta_{D_{\min}}$. By using this fact we can bound the first term of equation 114 as:

$$\sum_{t \leq C_{\mathrm{bnd}} \log(1/\delta)} \mathbb{P}[t \leq \tau, \mathcal{D}^c(t), \mathcal{G}^c(t, (\log\log(1/\delta))^4)] \leq (C(\boldsymbol{\mu}))^{-1} D(\delta) \times \frac{4Q^*(\boldsymbol{\mu})}{\min_i S_i(\boldsymbol{\mu})} (\log\log(1/\delta))^4 \tag{115}$$

$$= O((\log(1/\delta))^{1/2} \times (\log\log(1/\delta))^4). \tag{116}$$

Moreover, we bound the probability of $\cup_t |\theta_i(t) - \mu_i| \geq D_{\min}$ uniformly as follows:

$$\sum_{t \leq C_{\mathrm{bnd}} \log(1/\delta)} \mathbb{P}[t \leq \tau, \mathcal{D}^c(t), \mathcal{G}(t, (\log\log(1/\delta))^4), \mathcal{H}^c(t)] \tag{117}$$

$$\leq \sum_{t \leq C_{\mathrm{bnd}} \log(1/\delta)} \mathbb{P}\left[t \leq \tau, \mathcal{D}^c(t), \bigcap_{i \in \mathcal{I}} \{N_i(t) \geq N((\log\log(1/\delta))^4, \delta)\}, \mathcal{H}^c(t)\right] \tag{118}$$

$$\text{(by Lemma 17)} \tag{119}$$

$$\leq C_{\mathrm{bnd}} \log(1/\delta) \times \left(4C_{\mathrm{bnd}} \log(1/\delta) \exp\left(-\frac{\sqrt{N((\log\log(1/\delta))^4, \delta)} D_{\min}^2}{36}\right)\right) \tag{120}$$

$$\text{(by Lemma 15, transformed from } \frac{6\log(4T/\delta')}{\sqrt{N((\log\log(1/\delta))^4, \delta)}} = D_{\min} \text{ with } T = C_{\mathrm{bnd}} \log(1/\delta) \text{ to obtain } \delta'\text{)} \tag{121}$$

$$= o(1), \tag{122}$$

$$\text{(by } \sqrt{N((\log\log(1/\delta))^4, \delta)} = \Omega((\log\log(1/\delta))^2) \text{ and } (\exp(x))^2 \exp(-cx^2) = o(1) \text{ for } x = \log\log(1/\delta)\text{)} \tag{123}$$

which bounds the second term of equation 114. □

B.10 Proof of Lemma 11

Let $\mathcal{L}_i(t)$ be the event that arm i is the $(L^*(\boldsymbol{\theta}))$-th largest arm on posterior $\boldsymbol{\theta}$. By the property of the solution, $\mathcal{L}_i(t)$ implies $\max_j S_j(\boldsymbol{\theta}) = S_i(\boldsymbol{\theta}) \leq 1/d(\mu_i, \xi)$. Let

$$\mathcal{U}_i(t) = \left\{\mathcal{L}_i(t),\, |\theta_i(t) - \xi| \leq \frac{|\mu_i - \xi|}{2}\right\} \tag{124}$$

$$\mathcal{U}(t) = \bigcup_i \mathcal{U}_i(t). \tag{125}$$

Intuitively speaking, $\mathcal{U}_i(t)$ states that $S_i(t)$ can be extremely large[12], which does not occur often if $N_i(t)$ is sufficiently large. Note that, under $\mathcal{U}^c(t)$, $S_i(\boldsymbol{\theta})$ for any $j \in [K]$ is not extremely large,[13] and thus $S_i(\boldsymbol{\theta})/S_j(\boldsymbol{\theta})$ for any i, j is bounded by a constant, which we denote C_U.

In the following, we derive Lemma 11 by following the steps below. First, we define an event $\mathcal{W}(\eta)$, which holds with probability at least $1 - O(\eta)$. In the case of $\mathcal{W}(\eta)$, we can expect all means and posteriors are bounded with respect to the number of draws and time step t. We bound the stopping time by $T^{\mathrm{stop}}(\eta, \delta)$ under $\mathcal{W}$ by using several lemmas. Finally, integrating $T^{\mathrm{stop}}(\eta, \delta)$ over the η yields the bound of τ in expectation.

Proof of Lemma 11. Let $\hat{\mu}_{i,n}$ is the value of $\hat{\mu}_i(t)$ when $N_i(t) = n$. Let

$$\mathcal{W}(\eta) := \left\{\bigcap_{i,n}\left\{|\hat{\mu}_{i,n} - \mu_i| \leq \sqrt{\frac{\log(n^2/\eta)}{2n}}\right\}, \bigcap_{i,t}\left\{|\theta_i(t) - \hat{\mu}_i(t)| \leq \frac{1}{N_i(t)} + \frac{4\log(t^2/\eta)}{\sqrt{N_i(t)}}\right\}\right\}. \tag{126}$$

By Hoeffding's inequality,

$$\mathbb{P}\left[\bigcup_{i,n}\left\{|\hat{\mu}_{i,n} - \mu_i| > \sqrt{\frac{\log(n^2/\eta)}{2n}}\right\}\right] \leq \sum_{i,n} \mathbb{P}\left[|\hat{\mu}_{i,n} - \mu_i| > \sqrt{\frac{\log(n^2/\eta)}{2n}}\right] \leq K \sum_n \eta/n^2 = K\eta\pi^2/6 \tag{127}$$

Moreover, by Lemma 14, we have

$$\mathbb{P}\left[\bigcup_{i,t}\left\{|\theta_i(t) - \hat{\mu}_i(t)| > \frac{1}{N_i(t)} + \frac{4\log(t^2/\eta)}{\sqrt{N_i(t)}}\right\}\right] \leq \sum_{i,t} \mathbb{P}\left[|\theta_i(t) - \hat{\mu}_i(t)| > \frac{1}{N_i(t)} + \frac{4\log(t^2/\eta)}{\sqrt{N_i(t)}}\right] \tag{128}$$

$$\leq \sum_{i,t} \frac{\eta}{t^2} = \frac{K\eta\pi^2}{6}. \tag{129}$$

[12] $S_i(t)$ can be arbitrarily large if $\theta_i(t)$ is very close to ξ.

[13] Bounded by a constant $1/\min(d(\xi - \mu_i/2, \xi), d(\xi + \mu_i/2, \xi))$.

In summary,

$$\Pr[\mathcal{W}(\eta)] \geq 1 - \frac{K\eta\pi^2}{6} - \frac{K\eta\pi^2}{6} = 1 - \frac{K\eta\pi^2}{3}. \tag{130}$$

In the following, we bound the stopping time τ under event $\mathcal{W}(\eta)$.

Lemma 19. *Let $N_U(t) = \sum_{s=1}^{t-1} \mathbf{1}[\mathcal{U}_i(t)]$. Then,*

$$N_i(t) \geq \frac{N_U(t)}{K} \tag{131}$$

holds.

Lemma 20. *Assume that $\mathcal{W}(\eta)$ holds. For any T, we have*

$$\sum_{t=1}^{T} \mathbf{1}[\mathcal{U}_i(t)] \leq K \left(\frac{12 \log(T^2/\eta)}{|\mu_i - \xi|} \right)^2. \tag{132}$$

Lemma 21. *Let $\hat{\mathcal{I}} \in 2^{[K]}$ and $\Theta_{\hat{\mathcal{I}}} = \{\boldsymbol{\nu} \in [0,1] : \{i : S_i(\boldsymbol{\nu}) > 0\} = \hat{\mathcal{I}}\}\}$, which is the set of parameters where arms $\hat{\mathcal{I}}$ are drawn. Let C_U be an upper bound of $KS_i(\boldsymbol{\theta})/S_j(\boldsymbol{\theta})$ under $\mathcal{U}^c(t)$, which is a constant when we view $\boldsymbol{\mu}, \xi$ as constants. Let*

$$T_{\hat{\mathcal{I}}}(t) = \sum_{t'=1}^{t-1} \mathbf{1}[\boldsymbol{\theta}(t) \in \Theta_{\hat{\mathcal{I}}}, \mathcal{U}^c(t)]. \tag{133}$$

Then, for any $i \in \hat{\mathcal{I}}$ and t, $N_i(t) \geq T_{\hat{\mathcal{I}}}(t)/C_U - 1$ holds.

Lemma 22. *Under $\mathcal{W}(\eta)$, if there exists $\hat{\mathcal{I}} \in 2^{[K]}$ such that*

$$T_{\hat{\mathcal{I}}}(t) \geq C_U \left(\frac{\beta(t,\delta)}{d\left(\frac{\mu_i+\xi}{2}, \xi\right)} + \left(\frac{12 \log(t^2/\eta)}{|\mu_i - \xi|} \right)^2 \right) =: C_U \, T_U(t, \eta),$$

then the algorithm stops.

Lemmas 19–22 are used to derive the following Lemma 23.

Lemma 23. *Under $\mathcal{W}(\eta)$, there exists a function $T^{\mathrm{stop}}(\eta, \delta) = O(\log(1/\delta) + (\log(1/\eta))^2)$ such that $\tau \leq T^{\mathrm{stop}}(\eta, \delta)$ holds.*

By using Lemma 23, we finally bound the stopping time. Let $C_{\mathrm{bnd}} > 0$ be a (distribution-dependent) constant such that

$$T^{\mathrm{stop}}(\eta, \delta) \leq C_{\mathrm{bnd}}((\log(1/\eta))^2 + \log(1/\delta)).$$

We have

$$\sum_{t=C_{\mathrm{bnd}}\log(1/\delta)+1} \mathbb{P}[t \leq \tau] \leq 1 + \int_{t'} \mathbb{P}[\mathcal{W}^c(\eta) : C_{\mathrm{bnd}} \log(1/\eta) = t'] dt' \tag{134}$$

$$\text{(by Lemma 23, } t' = t - C_{\mathrm{bnd}} \log(1/\delta)) \tag{135}$$

$$= 1 + \int_{\eta=0}^{1} \mathbb{P}[\mathcal{W}^c(\eta)] \frac{2C_{\mathrm{bnd}} \log(1/\eta)}{\eta} d\eta \tag{136}$$

$$\text{(by } \eta = e^{-\sqrt{t'/C_{\mathrm{bnd}}}}) \tag{137}$$

$$\leq 1 + \int_{\eta=0}^{1} \frac{K\eta\pi^2}{3} \frac{2C_{\mathrm{bnd}} \log(1/\eta)}{\eta} d\eta \tag{138}$$

$$\text{(by equation 130)} \tag{139}$$

$$= 1 + \frac{2C_{\mathrm{bnd}} K \pi^2}{3} = O(1), \tag{140}$$

$$\text{(by } \int_0^1 \log(1/\eta) d\eta = 1) \tag{141}$$

$$\tag{142}$$

which completes the proof. □

B.10.1 Proofs of the lemmas used by Lemma 11

Proof of Lemma 19. Similarly to Lemma 17, we use the "subsequence" argument. If we count $N_i(t)$ on the subsequence $\{t : \mathcal{U}_i(t)\}$, then $N_i(t) \geq N_j(t)$ holds for any $j \in [K]$ because under this subsequence $S_i(\boldsymbol{\theta}(t)) > S_j(\boldsymbol{\theta}(t))$ always holds, which implies i is drawn at least $\frac{N_U(t)}{K}$ times on the subsequence. The number of draws on the full sequence $N_i(t)$ is always larger than that of subsequence, which is $N_i(t) \geq \frac{N_U(t)}{K}$. □

Proof of Lemma 20. Under $\mathcal{W}$, we have

$$|\theta_i(t) - \mu_i| \leq \sqrt{\frac{\log(t^2/\eta)}{2N_i(t)}} + \frac{1}{N_i(t)} + \frac{4\log(t^2/\eta)}{\sqrt{N_i(t)}} \leq \frac{6\log(t^2/\eta)}{\sqrt{N_i(t)}}.$$

Assume that for some round $t \leq T$ we have

$$N_U(t) \geq K\left(\frac{12\log(T^2/\eta)}{|\mu_i - \xi|}\right)^2.$$

Then, by Lemma 19 we have,

$$N_i(t) \geq \left(\frac{12\log(T^2/\eta)}{|\mu_i - \xi|}\right)^2. \tag{143}$$

Then

$$\begin{aligned} |\theta_i(t) - \xi| \geq |\mu_i - \xi| - |\theta_i(t) - \mu_i| &\geq |\mu_i - \xi| - \frac{6\log(t^2/\eta)}{\sqrt{N_i(t)}} && (144)\\ &\geq |\mu_i - \xi| - \frac{|\mu_i - \xi|}{2} \quad \text{(by equation 143)} && (145)\\ &\geq \frac{|\mu_i - \xi|}{2}, && (146) \end{aligned}$$

and thus $\mathcal{U}_i(t)$ never occurs. □

Proof of Lemma 21. Lemma 21 is proven by using a subsequence argument similar to Lemmas 17 and 19. □

Proof of Lemma 22. If

$$T_{\hat{\mathcal{I}}}(t) \geq C_U\, T_U(t, \eta) \tag{147}$$

then by Lemma 22 it holds that

$$N_i(t) \geq T_U(t, \eta) \tag{148}$$

for all $i \in \hat{\mathcal{I}}$. Similar discussion as equation 146 yields $|\hat{\mu}_i(t) - \xi| \leq |\mu_i - \xi|/2$, and if

$$N_i(t) \geq \frac{\beta(t,\delta)}{d\left(\frac{\mu_i+\xi}{2}, \xi\right)} \geq \frac{\beta(t,\delta)}{d\left(\hat{\mu}_i(t), \xi\right)} \quad \text{(by } |\hat{\mu}_i(t) - \xi| \leq |\mu_i - \xi|/2)$$

holds for all $i \in \hat{\mathcal{I}}$, then the algorithm stops, which completes the proof. □

Proof of Lemma 23. Let

$$T^{\text{stop}}(\eta, \delta) = \min_t \left\{ t \geq 2^K C_U\, T_U(t, \eta) + K^2\left(\frac{12\log(t^2/\eta)}{|\mu_i - \xi|}\right)^2 \right\}. \tag{149}$$

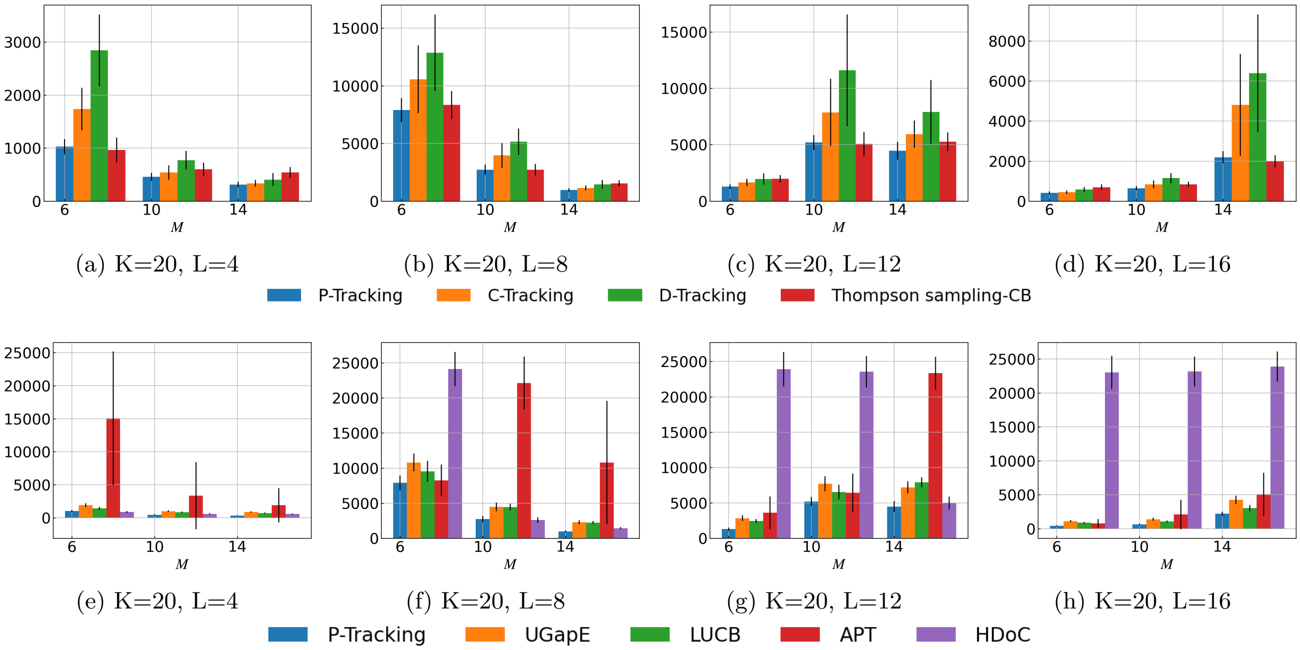
Since $T_U(t,\eta) = O((\log t)^2 + \log(1/\delta) + (\log(1/\eta))^2)$, it holds that $T^{\mathrm{stop}}(\eta,\delta) = O(\log(1/\delta) + (\log(1/\eta))^2)$. Lemma 20 implies

$$\sum_{t'=1}^{t} \mathbf{1}[\mathcal{U}(t)] \leq K^2 \left(\frac{12 \log(t^2/\eta)}{|\mu_i - \xi|} \right)^2, \tag{150}$$

and thus there exists at least one $\hat{\mathcal{I}} \in 2^{[K]}$ such that

$$T_{\hat{\mathcal{I}}}(t) \geq C_U \, T_U(t,\eta)$$

at round $t = T^{\mathrm{stop}}$, and by Lemma 22 the algorithm stops. □



(a) K=20, L=4

(b) K=20, L=8

(c) K=20, L=12

(d) K=20, L=16

(e) K=20, L=4

(f) K=20, L=8

(g) K=20, L=12

(h) K=20, L=16

Figure 5: Stopping time of each algorithm for $K = 20$. We tested $L = 4, 8, 12, 16$ and $M = 6, 10, 14$. In this experiments, the arms $1, 2, \ldots, K - M$ are bad arms and whereas the arms $K - M + 1, \ldots, K - 1, K$ are good arms. (a)-(d) P-Tracking and the asymptotically optimal algorithms. (e)-(h) P-Tracking and the algorithms with suboptimal sample complexity.

C SUPPLEMENTARY EXPERIMENTS

We conducted simulation experiments similar to those presented in the main text by changing the number of arms to $K = 20$. The results for the stopping time of each algorithm are shown in the Figure 5. In these experiments, we found that Thompson sampling-CB works better for $K = 20$ when compared to $K = 100$. However, we can observe that P-Tracking still outperforms the other algorithms in many cases, as observed in the results for $K = 100$.

References

Skorski, M. (2021). Bernstein-type bounds for beta distribution.