Machine Learning (2020) 109:327–372
https://doi.org/10.1007/s10994-019-05854-7

A bad arm existence checking problem: How to utilize asymmetric problem structure?

Koji Tabata[1,2] · **Atsuyoshi Nakamura**[1,3] · **Junya Honda**[4] · **Tamiki Komatsuzaki**[1,2,5]

Received: 21 January 2019 / Revised: 25 September 2019 / Accepted: 4 October 2019 /
Published online: 30 October 2019

Abstract

We study a *bad arm existence checking problem* in a stochastic K-armed bandit setting, in which a player's task is to judge whether a *positive* arm exists or all the arms are *negative* among given K arms by drawing as small number of arms as possible. Here, an arm is positive if its expected loss suffered by drawing the arm is at least a given threshold θ_U, and it is negative if that is less than another given threshold $\theta_L (\leq \theta_U)$. This problem is a formalization of diagnosis of disease or machine failure. An interesting structure of this problem is the asymmetry of *positive* and *negative* arms' roles; finding one positive arm is enough to judge positive existence while all the arms must be discriminated as negative to judge whole negativity. In the case with $\Delta = \theta_U - \theta_L > 0$, we propose elimination algorithms with *arm selection policy* (policy to determine the next arm to draw) and *decision condition* (condition to conclude positive arm's existence or the drawn arm's negativity) utilizing this asymmetric problem structure and prove its effectiveness theoretically and empirically.

Keywords Online learning · Bandit problem · Best arm identification

Editors: Po-Ling Loh, Evimaria Terzi, Antti Ukkonen, Karsten Borgwardt.

✉ Atsuyoshi Nakamura
atsu@ist.hokudai.ac.jp

Koji Tabata
ktabata@es.hokudai.ac.jp

Junya Honda
jhonda@k.u-tokyo.ac.jp

Tamiki Komatsuzaki
tamiki@es.hokudai.ac.jp

[1] Research Center of Mathematics for Social Creativity, Research Institute for Electronic Science, Hokkaido University, Kita 20 Nishi 10, Kita-ku, Sapporo 001-0020, Japan

[2] Institute for Chemical Reaction Design and Discovery (WPI-ICReDD), Hokkaido University, Kita 21 Nishi 10, Kita-ku, Sapporo, Hokkaido 001-0021, Japan

[3] Graduate School of Information Science and Technology, Hokkaido University, Kita 14, Nishi 9, Kita-ku, Sapporo, Hokkaido 060-0814, Japan

[4] University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa-shi, Chiba 277-8561, Japan

[5] École Normale Supérieure de Lyon, 46 Allée d'Italie, 69007 Lyon, France

1 Introduction

In the diagnosis of disease or machine failure, the test object is judged as “positive” if some anomaly is detected in at least one of many parts. In the case that the purpose of the diagnosis is the classification into one of the two classes, “positive” or “negative”, then the diagnosis can be terminated right after the first anomaly part has been detected. Thus, fast diagnosis will be realized if one of anomaly parts can be detected as fast as possible in the positive case.

The fast diagnosis of anomaly detection is particularly important in the case that the judgment is done based on measurements using a costly or slow device. For example, a Raman spectral image has been known to be useful for cancer diagnosis (Haka et al. 2009), but its acquisition time is 1–10 seconds per point (pixel)[1] resulting in an order of hours or days per one image (typically 10,000–40,000 pixels), so it is critical to measure only the points necessary for cancer diagnosis in order to achieve fast measurement. A Raman spectrum of each point is believed to be converted to a *cancer index*, which indicates how likely the point is inside a cancer cell, and we can judge the existence of cancer cells from the existence of area with a high cancer index.

The above cancer cell existence checking problem can be formulated as the problem of checking the existence of a grid with a high cancer index for a given area that is divided into grids. By regarding each grid as an arm, we formalize this problem as a loss-version of a stochastic K-armed bandit problem in which the existence of *positive* arms is checked by drawing arms and suffering losses for the drawn arms. In our formulation, given an acceptable error rate $0 < \delta < 1/2$ and two thresholds θ_L and θ_U with $0 < \theta_L \le \theta_U < 1$, a player is required to, with probability at least $1 - \delta$, answer “positive” if positive arms exist and “negative” if all the arms are *negative*. Here, an arm is defined to be positive if its loss mean is at least θ_U, and defined to be negative if its loss mean is less than θ_L. We call player algorithms for this problem as *($\theta_L, \theta_U, \delta$)-BAEC (Bad Arm Existence Checking) algorithms*. The objective of this research is to design a ($\theta_L, \theta_U, \delta$)-BAEC algorithm that minimizes the number of arm draws, that is, an algorithm with the lowest sample complexity. The problem of this objective is said to be a *Bad Arm Existence Checking Problem*.

The bad arm existence checking problem is closely related to the *thresholding bandit problem* (Locatelli et al. 2016), which is a kind of pure-exploration problem such as the *best arm identification problem* (Even-Dar et al. 2006; Audibert et al. 2010). In the thresholding bandit problem, provided a threshold θ and a required precision $\epsilon > 0$, the player's task is to classify each arm into positive (its loss mean is at least $\theta + \epsilon$) or negative (its loss mean is less than $\theta - \epsilon$) by drawing a fixed number of samples, and his/her objective is to minimize the error probability, that is, the probability that positive (resp. negative) arms are wrongly classified into negative (resp. positive). Apart from whether fixed confidence (constraint on error probability to achieve) or fixed budget (constraint on the allowable number of draws), positive and negative arms are treated symmetrically in the thresholding bandit problem while they are dealt with asymmetrically in our problem setting; judgment of one positive arm existence is enough for positive conclusion though all the arms must be judged as negative for negative conclusion. This asymmetry has also been considered in the *good arm identification problem* (Kano et al. 2017), and our problem can be seen as its specialized version though their problem deal with the case with $\theta_L = \theta_U$ only. In their setting, the player's task is to output all the arms of above-threshold means with probability at least $1 - \delta$, and his/her objective is to minimize the number of drawn samples until λ arms are outputted as arms

[1] http://www.horiba.com/en_en/raman-imaging-and-spectroscopy-recording-spectral-images-profiles/.

with above-threshold means for a given λ. In the case with $\lambda = 1$, algorithms for their problem can be used to solve our existence checking problem. Their proposed algorithm, however, does not utilize the asymmetric problem structure. Kaufmann et al. (2018) studied the problem of sequential test for the lowest mean, which is basically the same problem as the bad arm existence checking problem except the difference in the number of thresholds; they also treat the case with $\theta_L = \theta_U$ only. They proposed an algorithm to utilize the asymmetric problem structure: *Murphy Sampling* and asymmetric stopping condition. Our approach to utilize the asymmetric problem structure is different from their approach; our algorithm is an *elimination algorithm* and asymmetric conditions are used not only to stop but also to eliminate the drawn arm.

We consider elimination algorithms BAEC[ASP, LB, UB] that are mainly composed of an *arm-selection policy* $\arg\max_i \mathrm{ASP}(t, i)$ and a *decision condition* $\mathrm{LB}(t) \geq \theta_L$ or $\mathrm{UB}(t) < \theta_U$ at time t. The arm-selection policy decides which arm is drawn at each time t based on loss samples obtained so far. The decision condition is used to conclude positive arm's existence if $\mathrm{LB}(t) \geq \theta_L$ holds or the drawn arm's negativity if $\mathrm{UB}(t) < \theta_U$ holds. If the conclusion is positive arm's existence, then the algorithms stop immediately by returning "positive". In the case that the conclusion is the drawn arm's negativity, the arm is eliminated from the set of positive-arm candidates, which is composed of all the arms initially, and will not be drawn any more. If there remains no positive-arm candidate, then the algorithms stop by returning "negative". To utilize our asymmetric problem structure, we propose a decision condition that uses Δ-dependent *asymmetric* confidence bounds $\underline{\mu}(t)$ and $\overline{\mu}(t)$ of estimated loss means as $\mathrm{LB}(t)$ and $\mathrm{UB}(t)$ in the case with $\Delta = \theta_U - \theta_L > 0$. Here, asymmetric bounds mean that the width of the upper confidence interval is narrower than the width of the lower confidence interval. As an arm selection policy, we propose policy $\mathrm{APT_P}$ that is derived by modifying policy APT (Anytime Parameter-free Thresholding) (Locatelli et al. 2016) so as to favor arms with sample means larger than a single threshold θ (rather than arms with sample means closer to θ as the original APT does). Here, as the single threshold θ used by policy $\mathrm{APT_P}$, we use not the center between θ_L and θ_U but the value closer to θ_U by utilizing the asymmetry of our confidence bounds.

By using Δ-dependent asymmetric confidence bounds as the decision condition, the worst-case bound on the number of samples for each arm is shown to be improved by $\Omega\left(\frac{1}{\Delta^2}\ln\frac{\sqrt{K}}{\Delta^2}\right)$ compared to the case using the conventional symmetric confidence bounds of the successive elimination algorithm (Even-Dar et al. 2006).

Our sample complexity results regarding the asymptotic behavior as $\delta \to 0$ is summarized as Table 1. Reflecting the asymmetric structure of the problem, the existence of a positive arm makes the sample complexity higher. In the case with negative arms only, algorithm BAEC[$*$, $\underline{\mu}$, $\overline{\mu}$], our elimination algorithm with any arm selection policy and Δ-dependent asymmetric confidence bounds, is proved to achieve almost optimal sample complexity. In the case with positive arm existence, the upper bound on the expected number of samples for algorithm BAEC[$\mathrm{APT_P}$, $\underline{\mu}$, $\overline{\mu}$] is proved to be almost optimal when all the positive arms have the same loss mean while that for algorithm BAEC[UCB, $\underline{\mu}$, $\overline{\mu}$] using UCB (Upper Confidence Bound) (Auer et al. 2002) as the arm selection policy like HDoC (Hybrid algorithm for the Dilemma of Confidence) (Kano et al. 2017) is proved to be almost optimal when just one positive arm has the largest loss mean.

The effectiveness of our decision condition using the Δ-dependent asymmetric confidence bounds is demonstrated in simulation experiments. The algorithm using our Δ-dependent asymmetric confidence bounds stops drawing an arm about two times faster than the algorithm using the symmetric confidence bounds when its loss mean is around the center of the

Table 1 Our asymptotic lower and upper bounds on the expected stopping times $\mathbb{E}[T]$ divided by $\ln\frac{1}{\delta}$ as $\delta \to +0$, that is, $\lim_{\delta\to+0}\frac{\mathbb{E}[T]}{\ln\frac{1}{\delta}}$, for bad arm existence checking problem

Case		At least one positive	Negative only
Lower bound (Bernoulli distribution case)		$\frac{1}{d(\mu_1,\theta_L)}\left(\le\frac{1}{2\Delta_1^2}\right)$ (Th. 1)	$\sum_{i=1}^{K}\frac{1}{d(\mu_i,\theta_U)}$ $\left(\le\sum_{i=1}^{K}\frac{1}{2\Delta_i^2}\right)$ (Th. 1)
Upper bound	BAEC[$*$, $\underline{\mu}$, $\overline{\mu}$]	$\sum_{i=1}^{K}\frac{1}{2\Delta_i^2}$ (Th. 6)	$\sum_{i=1}^{K}\frac{1}{2\Delta_i^2}$ (Th. 6)
	BAEC[$\mathrm{APT}_{\mathrm{P}}$, $\underline{\mu}$, $\overline{\mu}$]	$\frac{1}{2\Delta_m^2}$ (Cor. 1)	
	BAEC[UCB, $\underline{\mu}$, $\overline{\mu}$]	$\frac{\lvert\{i \mid \mu_i=\mu_1\}\rvert}{2\Delta_1^2}$ (Cor. 2)	

Without loss of generality, arms i with mean loss μ_i are assumed to be sorted as $\mu_1 \ge \cdots \ge \mu_K$. Function d is Kullback–Leibler divergence for Bernoulli distribution defined as $d(x,y) = x\ln\frac{x}{y} + (1-x)\ln\frac{1-x}{1-y}$, and the parenthesized upper bounds can be obtained by Pinsker's Inequality, which is known to be tight in the worst case. For θ defined by Eq. (6), Δ_i is defined to be $\mu_i - \theta_L$ if $\mu_i \ge \theta$ and $\theta_U - \mu_i$ otherwise, and $m = \lvert\{i \mid \mu_i \ge \theta\}\rvert$

thresholds. Our algorithm BAEC[$\mathrm{APT}_{\mathrm{P}}$, $\underline{\mu}$, $\overline{\mu}$] almost always stops faster than the algorithm BAEC[UCB, $\underline{\mu}$, $\overline{\mu}$], and our algorithm's stopping time is faster or comparable to the stopping time of the algorithm BAEC[ASP, $\underline{\mu}$, $\overline{\mu}$] using LUCB (Lower and Upper Confidence Bounds) (Kalyanakrishnan et al. 2012), Thompson Sampling (Thompson 1933) and Murphy Sampling (Kaufmann et al. 2018) as ASPs in almost all the our simulations using Bernoulli loss distribution with synthetically generated means and means generated from a real-world dataset.

Related work

The bad arm existence checking problem is a kind of *multi-armed bandit problem*, which is a classical problem studied by Thompson (1933) and Robbins (1952). A bandit problem is an *online learning problem* (Littlestone and Warmuth 1994), but a player can obtain partial information only in its setting. In our study, loss (or reward) distribution is assumed to be stochastic, which is easier to deal with than the adversarial setting (Auer et al. 2003). For the bandit problem, depending on problem objectives, two kinds of settings exist: *regret-minimization* setting (Auer et al. 2002) and *pure-exploration* setting (Bubeck et al. 2011). Most pure-exploration problems are *best arm identification problems* (Even-Dar et al. 2006; Audibert et al. 2010; Kalyanakrishnan et al. 2012; Kaufmann and Kalyanakrishnan 2013) which are the problems to identify the arms with the maximum reward means. There are the fixed budget version and the fixed confidence version of best arm identification problems, and algorithms for the fixed confidence version have an arm selection policy and a stopping condition. Some best arm identification algorithms eliminate arms that are estimated not to be the best, and most of those algorithms use uniform sampling as an arm selection policy (Even-Dar et al. 2006; Bubeck et al. 2013). Non-elimination best arm identification algorithms use a more sophisticated adaptive sampling as an arm selection policy (Gabillon et al. 2012; Kalyanakrishnan et al. 2012). Comparison analysis between elimination and non-elimination algorithms was performed by Kaufmann and Kalyanakrishnan (2013). Identification of the above-or-below-threshold arms (Locatelli et al. 2016; Kano et al. 2017; Kaufmann et al. 2018) is a variant of best arm identification, and among these algorithms, only ours and HDoC (Kano et al. 2017) are elimination algorithms using adaptive sampling. This combination is effective for checking existence of above-or-below-threshold arm setting.

2 Preliminaries

For given thresholds $0 < \theta_L \leq \theta_U < 1$, consider the following bandit problem. Let $K (\geq 2)$ be the number of arms, and at each time $t = 1, 2, \ldots,$ a player draws arm $i_t \in \{1, \ldots, K\}$. For $i \in \{1, \ldots, K\}$, $X_i(n) \in [0, 1]$ denotes the loss for the nth draw of arm i, where $X_i(1), X_i(2), \ldots$ are a sequence of i.i.d. random variables generated according to a probability distribution ν_i with mean $\mu_i \in [0, 1]$. We assume independence between $\{X_i(t)\}_{t=1}^{\infty}$ and $\{X_j(t)\}_{t=1}^{\infty}$ for any $i, j \in \{1, \ldots, K\}$ with $i \neq j$. For a distribution set $\boldsymbol{\nu} = \{\nu_i\}$ of K arms, $\mathbb{E}_{\boldsymbol{\nu}}$ and $\mathbb{P}_{\boldsymbol{\nu}}$ denote the expectation and the probability under $\boldsymbol{\nu}$, respectively, and we omit the subscript $\boldsymbol{\nu}$ if it is trivial from the context. Without loss of generality, we can assume that $\mu_1 \geq \cdots \geq \mu_K$ and the player does not know this ordering. Let $n_i(t)$ denote the number of draws of arm i right before the beginning of the round at time t. After the player observed the loss $X_{i_t}(n_{i_t}(t) + 1)$, he/she can choose stopping or continuing to play at time $t + 1$. Let T denote the stopping time.

The player's objective is to check the existence of some *positive* arm(s) with as small a stopping time T as possible. Here, arm i is said to be *positive* if $\mu_i \geq \theta_U$, *negative* if $\mu_i < \theta_L$, and *neutral* otherwise. We consider a *bad arm existence checking problem*, which is a problem of developing algorithms that satisfy the following definition with as small number of arm draws as possible.

Definition 1 Given[2] $0 < \theta_L \leq \theta_U < 1$ and $\delta \in (0, 1/2)$, consider a game that repeats choosing one of K arms and observing its loss at each time t. A player algorithm for this game is said to be a $(\theta_L, \theta_U, \delta)$-*BAEC (Bad Arm Existence Checking) algorithm* if it stops in a finite time outputting "positive" with probability at least $1 - \delta$ in the case that at least one arm is positive, and "negative" with probability at least $1 - \delta$ in the case that all the arms are negative.

Note that the definition of BAEC algorithms requires nothing when arm 1 is neutral. Our problem definition coincides with the highest-mean version problem of *sequential testing for the lowest mean* (Kaufmann et al. 2018) in the case with $\theta_L = \theta_U$. Table 2 is the table of notations used throughout this paper.

[2] Thresholds θ_L and θ_U correspond to $\theta - \epsilon$ and $\theta + \epsilon$, respectively, in thresholding bandit problem (Locatelli et al. 2016) with one threshold θ and precision ϵ, but we use the two thresholds due to convenience for our asymmetric problem structure.

Table 2 Notation list

K : Number of arms

θ_U, θ_L : Upper and lower thresholds. ($0 < \theta_L \leq \theta_U < 1$)

Δ : Gray zone width ($\Delta = \theta_U - \theta_L$)

δ : Acceptable error rate. ($\delta \in (0, 1/2)$)

ν_i : Loss distribution of arm i

$\boldsymbol{\nu}$: Set $\{\nu_i\}$ of loss distributions of K arms

μ_i : Loss mean (expected loss) of arm i. ($\mu_i \in [0, 1]$)

$$\text{Arm } i \text{ is } \begin{cases} \text{positive if } \mu_i \geq \theta_U, \\ \text{neutral if } \theta_L \leq \mu_i < \theta_U, \\ \text{negative if } \mu_i < \theta_L \end{cases}$$

$\Delta_{1i} = \mu_1 - \mu_i$

$\mathbb{E}_{\boldsymbol{\nu}}$: Expectation of some random variable w.r.t. $\boldsymbol{\nu}$

$\mathbb{P}_{\boldsymbol{\nu}}$: Probability of some event w.r.t. $\boldsymbol{\nu}$ ($\boldsymbol{\nu}$ is omitted when it is trivial from the context)

i_t : Drawn arm at time t

$X_i(n)$: Loss suffered by the nth draw of arm i

$n_i(t)$: Number of draws of arm i at the beginning of the round at time t

T : Stopping time

$\hat{\mu}_i(n) = \frac{1}{n} \sum_{s=1}^{n} X_i(s)$

$\underline{\mu}'_i(n) = \hat{\mu}_i(n) - \sqrt{\frac{1}{2n} \ln \frac{2Kn^2}{\delta}} \quad \overline{\mu}'_i(n) = \hat{\mu}_i(n) + \sqrt{\frac{1}{2n} \ln \frac{2Kn^2}{\delta}}$

$N_{\Delta} = \left\lceil \frac{2\mathrm{e}}{(\mathrm{e}-1)\Delta^2} \ln \frac{2\sqrt{K}}{\Delta^2 \delta} \right\rceil \quad T_{\Delta} = \left\lceil \frac{2}{\Delta^2} \ln \frac{\sqrt{K} N_{\Delta}}{\delta} \right\rceil$

$\alpha = \sqrt{1 + \frac{\ln K}{\ln \frac{N_{\Delta}}{\delta}}} \qquad \theta = \theta_U - \frac{1}{1+\alpha} \Delta = \theta_L + \frac{\alpha}{1+\alpha} \Delta$

$\Delta_i = \begin{cases} \mu_i - \theta_L (\mu_i \geq \theta) \\ \theta_U - \mu_i (\mu_i < \theta) \end{cases} \quad T_{\Delta_i} = \left\lceil \frac{2}{\Delta_i^2} \ln \frac{\sqrt{K} N_{\Delta}}{\delta} \right\rceil$

$\underline{\Delta}_i = |\mu_i - \theta|$

m : Number of arms i with $\mu_i \geq \theta$

$\underline{\mu}_i(n) = \hat{\mu}_i(n) - \sqrt{\frac{1}{2n} \ln \frac{K N_{\Delta}}{\delta}} \qquad \overline{\mu}_i(n) = \hat{\mu}_i(n) + \sqrt{\frac{1}{2n} \ln \frac{N_{\Delta}}{\delta}}$

τ_i : Number n of draws of arm i until algorithm BAEC[$*, \underline{\mu}, \overline{\mu}$]'s decision condition ($\underline{\mu}_i(n) \geq \theta_L$ or $\overline{\mu}_i(n) < \theta_U$) is satisfied.

$\hat{i}_1$: First arm that is drawn τ_i times by algorithm BAEC[$\mathrm{APT_P}, \underline{\mu}, \overline{\mu}$]

$\mathcal{E}^+ = \bigcup_{i:\mu_i \geq \theta_U} \bigcap_{n=1}^{T_{\Delta}} \left\{ \overline{\mu}_i(n) \geq \mu_i \right\} \quad \mathcal{E}^- = \bigcap_{i=1}^{K} \bigcap_{n=1}^{T_{\Delta}} \left\{ \underline{\mu}_i(n) < \mu_i \right\}$

$\mathcal{E}_i^{\mathrm{POS}}$: Event that arm i is judged as positive

3 Sample complexity lower bound

In this section, we derive a lower bound on the expected number of samples needed for a $(\theta_L, \theta_U, \delta)$-BAEC algorithm. The derived lower bound is used to evaluate algorithm's sample complexity upper bound in Sects. 5.3 and 6.2.

We let $\mathrm{KL}(\nu, \nu')$ denote Kullback–Leibler divergence from distribution ν' to ν and define $d(x, y)$ as

$$d(x, y) = x \ln \frac{x}{y} + (1-x) \ln \frac{1-x}{1-y}.$$

Note that $\mathrm{KL}(\nu, \nu') = d(\mu_i, \mu_i')$ holds if ν and ν' are Bernoulli distributions with means μ_i and μ_i', respectively.

The following theorem is an extension[3] of Lemma 1 in Kaufmann et al. (2018) to the case with two thresholds.

Theorem 1 *Let $\{\nu_i\}$ be a set of Bernoulli distributions with means $\{\mu_i\}$. Then, the stopping time T of any $(\theta_L, \theta_U, \delta)$-BAEC algorithm with θ_U and θ_L is bounded as*

$$\mathbb{E}(T) > \frac{1-2\delta}{d(\mu_1, \theta_L)} \ln \frac{1-\delta}{\delta} \tag{1}$$

if some arm is positive, and

$$\mathbb{E}(T) > \sum_{i=1}^{K} \frac{1-2\delta}{d(\mu_i, \theta_U)} \ln \frac{1-\delta}{\delta} \tag{2}$$

if all the arms are negative.

Proof See "Appendix A". □

Remark 1 Identification is not needed for checking existence, however, in terms of asymptotic behavior as $\delta \to +0$, the shown expected sample complexity lower bounds of both the tasks are the same; $\lim_{\delta \to +0} \mathbb{E}(T)/\ln(1/\delta) \geq 1/d(\mu_1, \theta_L)$ for both the tasks in the case with some positive arms. The bounds are tight considering the shown upper bounds, so the bad arm existence checking is not more difficult than the good arm identification[4] (Kano et al. 2017) with respect to asymptotic behavior as $\delta \to +0$.

4 Algorithm

4.1 BAEC[ASP, LB, UB] algorithm framework

As $(\theta_L, \theta_U, \delta)$-BAEC algorithms, we consider algorithm BAEC[ASP, LB, UB] shown in Algorithm 1 that, at each time t, chooses an arm i_t from the set A_t of positive-candidate arms by an *arm-selection policy* ASP

$$i_t \leftarrow \arg\max_{i \in A_t} \mathrm{ASP}(t, i)$$

[3] The original lemma treats the problem to decide whether the lowest mean is less than a given one threshold for one-parameter canonical exponential family of K distributions.

[4] The lower bound on the stopping time under the decision of no more positive arm is not analyzed in Kano et al. (2017), and the stopping time in the case with no positive arm is the time of its special case. In good arm identification, the algorithm must stop without falsely identifying any arm as positive in such case with probability at least $1-\delta$, so its task is the same as our bad arm existence checking problem in the case with no positive arm.

Algorithm 1 BAEC[ASP, LB, UB]

```
Parameter Function:
        ASP$(t, i)$: index value of arm $i$ at time $t$ for arm selection
        LB$(t)$, UB$(t)$: lower and upper confidence bounds of arm $i_t$'s estimated loss mean
Input: $K$: the number of arms
        $0 < \theta_L \leq \theta_U < 1$: thresholds
        $\delta \in (0, 1/2)$: acceptable error rate
1: $A_1 \leftarrow \{1, 2, \ldots, K\}, n_i(1) \leftarrow 0$ for $i = 1, \ldots, K$
2: $t \leftarrow 1$
3: while $A_t \neq \emptyset$ do
4:     $i_t \leftarrow \arg\max_{i \in A_t} \mathrm{ASP}(t, i)$
5:     $n_i(t+1) \leftarrow \begin{cases} n_i(t) + 1 & (i = i_t) \\ n_i(t) & (i \neq i_t) \end{cases}$
6:     Draw $i_t$ and suffer a loss $X_{i_t}(n_{i_t}(t+1))$.
7:     $\hat{\mu}_{i_t}(n_{i_t}(t+1)) \leftarrow \frac{\hat{\mu}_{i_t}(n_{i_t}(t)) \times n_{i_t}(t) + X_{i_t}(n_{i_t}(t+1))}{n_{i_t}(t+1)}$
8:     if LB$(t) \geq \theta_L$ then
9:         return "positive"                    ▷ Conclude positive arm's existence
10:    else if UB$(t) < \theta_U$ then
11:        $A_{t+1} \leftarrow A_t \setminus \{i_t\}$                ▷ Conclude Arm $i_t$'s negativity
12:    end if
13:    $t \leftarrow t + 1$
14: end while
15: return "negative"
```

using some index value ASP(t, i) of arm i at time t (Line 4), suffers a loss $X_{i_t}(n_{i_t}(t+1))$ (Line 6) and then checks whether a *decision condition*

$$\mathrm{LB}(t) \geq \theta_L \text{ or } \mathrm{UB}(t) < \theta_U$$

is satisfied (Lines 8 and 10). Here, LB(t) and UB(t) are lower and upper confidence bounds of an estimated loss mean of the current drawn arm i_t, and condition LB$(t) \geq \theta_L$ is the condition for the decision of positive arm's existence , and condition UB$(t) < \theta_U$ is the condition for concluding the drawn arm's negativity and eliminating arm i_t from the set A_{t+1} of positive-candidate arms of time $t + 1$. In addition to the case with positive conclusion, algorithm BAEC[ASP, LB, UB] also stops with negative conclusion when A_t becomes empty.

Define sample loss mean $\hat{\mu}_i(n)$ of arm i with n draws as

$$\hat{\mu}_i(n) = \frac{1}{n} \sum_{s=1}^{n} X_i(s),$$

and we use $\hat{\mu}_{i_t}(n_{i_t}(t+1))$ as an estimated loss mean of the current drawn arm i_t at time t.

4.2 Asymmetric Δ-dependent confidence bounds

As we use the sample mean $\hat{\mu}_i(n)$ as an estimated loss mean, LB(t) and UB(t) are determined by defining lower and upper bounds of a confidence interval of $\hat{\mu}_i(n)$ for $i = i_t$ and $n = n_{i_t}(t+1)$.

As lower and upper confidence bounds of $\hat{\mu}_i(n)$,

$$\underline{\mu}'_i(n) = \hat{\mu}_i(n) - \sqrt{\frac{1}{2n} \ln \frac{2Kn^2}{\delta}} \text{ and } \overline{\mu}'_i(n) = \hat{\mu}_i(n) + \sqrt{\frac{1}{2n} \ln \frac{2Kn^2}{\delta}}, \tag{3}$$

respectively, are generally used[5] in successive elimination algorithms (Even-Dar et al. 2006). Define $\underline{\mu}'(t)$ and $\overline{\mu}'(t)$ as $\underline{\mu}'(t) = \underline{\mu}'_{i_t}(n_{i_t}(t+1))$ and $\overline{\mu}'(t) = \overline{\mu}'_{i_t}(n_{i_t}(t+1))$ for use as $\mathrm{LB}(t)$ and $\mathrm{UB}(t)$.

Consider the case with $\theta_L < \theta_U$, namely, the case that θ_L is strictly smaller than θ_U. In this case, we propose asymmetric bounds $\underline{\mu}_i(n)$ and $\overline{\mu}_i(n)$ defined using a *gray zone width* $\Delta = \theta_U - \theta_L$ as follows:

$$\underline{\mu}_i(n) = \hat{\mu}_i(n) - \sqrt{\frac{1}{2n}\ln\frac{KN_\Delta}{\delta}} \text{ and } \overline{\mu}_i(n) = \hat{\mu}_i(n) + \sqrt{\frac{1}{2n}\ln\frac{N_\Delta}{\delta}}, \tag{4}$$

where

$$N_\Delta = \left\lceil \frac{2\mathrm{e}}{(\mathrm{e}-1)\Delta^2}\ln\frac{2\sqrt{K}}{\Delta^2\delta} \right\rceil.$$

We also let $\underline{\mu}(t)$ and $\overline{\mu}(t)$ denote $\mathrm{LB}(t)$ and $\mathrm{UB}(t)$ using these bounds, that is, $\underline{\mu}(t) = \underline{\mu}_{i_t}(n_{i_t}(t+1))$ and $\overline{\mu}(t) = \overline{\mu}_{i_t}(n_{i_t}(t+1))$.

Note that $\overline{\mu}_i(n) < \overline{\mu}'_i(n)$ for $n > \sqrt{N_\Delta/2K}$ and $\underline{\mu}_i(n) > \underline{\mu}'_i(n)$ for $n > \sqrt{N_\Delta/2}$, so $\overline{\mu}_i(n) - \underline{\mu}_i(n) < \overline{\mu}'_i(n) - \underline{\mu}'_i(n)$ holds for $n \geq \sqrt{N_\Delta/2}$. Both $\overline{\mu}_i(n) - \underline{\mu}_i(n)$ and $\overline{\mu}'_i(n) - \underline{\mu}'_i(n)$ decrease as n increases, and $\mathrm{LB}(t) \geq \theta_L$ or $\mathrm{UB}(t) < \theta_U$ is satisfied for $\mathrm{BAEC}[*, \underline{\mu}, \overline{\mu}]$ and $\mathrm{BAEC}[*, \underline{\mu}', \overline{\mu}']$ when they become at most Δ for $n = n_i(t+1)$, where $\mathrm{ASP} = *$ means that any index function $\mathrm{ASP}(t, i)$ can be assumed.

Remark 2 Condition $\underline{\mu}(t) \geq \theta_L$ essentially identifies non-negative arm i_t. Is there real-valued function LB that can check existence of a non-negative arm without identifying it? The answer is yes. Consider a virtual arm at each time t whose mean loss μ^t is a weighted average over the mean losses μ_i of all the arms i $(i = 1, \ldots, K)$ defined as $\mu^t = \frac{1}{t}\sum_{i=1}^K n_i(t+1)\mu_i$. If $\mu^t \geq \theta_L$, then at least one arm i must be non-negative. Thus, we can check the existence of a non-negative arm by judging whether $\mu^t \geq \theta_L$ or not. Since $\underline{\mu}^t(t)$ defined as

$$\underline{\mu}^t(t) = \frac{1}{t}\sum_{i=1}^{K} n_i(t+1)\hat{\mu}_i(t+1) - \sqrt{\frac{1}{2t}\ln\frac{2t^2}{\delta}}$$

can be considered to be a lower bound of the estimated value of μ^t, $\underline{\mu}^t$ can be used as LB for checking the existence of a non-negative arm without identifying it. Instead of the set of all arms, any arm subset can be considered to be a virtual arm as the stopping condition proposed by Kaufmann et al. (2018) in the case with $\theta_L = \theta_U$. However, the increase of the number of subsets to be considered also makes the required number of each subset's samples increase due to the property of union bound. In this paper, we do not pursue in this direction, and instead focus on the effect investigation of the decision condition using Δ-dependent asymmetric confidence bounds.

The ratio of the width of our upper confidence interval $\left[\hat{\mu}_i(n), \overline{\mu}_i(n)\right]$ to the width of our lower confidence interval $\left[\underline{\mu}_i(n), \hat{\mu}_i(n)\right]$ is $\sqrt{\ln\frac{N_\Delta}{\delta}} : \sqrt{\ln\frac{KN_\Delta}{\delta}} = 1 : \sqrt{1 + \frac{\ln K}{\ln\frac{N_\Delta}{\delta}}}$. Thus, we

[5] Precisely speaking, $\hat{\mu}_i(n) \pm \sqrt{\frac{1}{2n}\ln\frac{4Kn^2}{\delta}}$ is used in successive elimination algorithms for best arm identification problem. A narrower confidence interval is enough to judge whether expected loss is larger than a *fixed* threshold.

define θ as

$$\theta = \theta_U - \frac{1}{1+\alpha}\Delta \quad \text{where } \alpha = \sqrt{1 + \frac{\ln K}{\ln \frac{N_\Delta}{\delta}}}.$$

This θ can be considered to be the balanced center between the thresholds θ_L and θ_U for our asymmetric confidence bounds.

4.3 Arm selection policy APT$_\mathrm{P}$

As arm selection policy ASP, we consider policy APT$_\mathrm{P}$ that uses index function

$$\mathrm{APT}_\mathrm{P}(t,i) = \sqrt{n_i(t)}\left(\hat{\mu}_i(n_i(t)) - \theta\right), \tag{5}$$

where we use $\hat{\mu}_i(n_i(t)) = \theta$ when $n_i(t) = 0$. This arm-selection policy is a modification of the policy of APT (Anytime Parameter-free Thresholding algorithm) (Locatelli et al. 2016), in which an arm

$$\arg\min_i \sqrt{n_i(t)}\left(\left|\hat{\mu}_i(n_i(t)) - \theta\right| + \epsilon\right) \tag{6}$$

is chosen for given threshold θ and accuracy ϵ. In the original APT, arm i with the sample mean $\hat{\mu}_i(n_i(t))$ closest to θ is preferred to be chosen no matter whether $\hat{\mu}_i(n_i(t))$ is larger or smaller than θ. In APT$_\mathrm{P}$, there is at most one arm i whose sample mean $\hat{\mu}_i(n_i(t))$ is larger than θ at any time t because of the above our definition of $\hat{\mu}_j(n_j(t))$ for arms j with $n_j(t) = 0$ and mathematical induction in t, and such unique arm i is always chosen as long as $\hat{\mu}_i(n_i(t)) > \theta$.

5 Theoretical analyses of algorithm BAEC[$*, \underline{\mu}, \overline{\mu}$]

In the following sections, we consider the case with $\theta_L < \theta_U$ ($\Delta > 0$). We first analyze arm's sample complexity for any arm, then analyze algorithm's sample complexity.

5.1 Worst case sample complexity upper bound for any arm

One merit of the two threshold setting with $\theta_L < \theta_U$ is that the number of drawn samples until the decision condition is satisfied, is upper-bounded for any arm by a common constant depending on $\Delta = \theta_U - \theta_L$ and δ. In this subsection, we prove such common constant bound for our Δ-dependent asymmetric confidence bounds and compare it with the corresponding number of samples for the conventional symmetric confidence bounds.

Let τ_i denote the smallest number n of draws of arm i for which the decision condition is met, that is, either $\underline{\mu}_i(n) \geq \theta_L$ or $\overline{\mu}_i(n) < \theta_U$ holds. Define T_Δ as $T_\Delta = \left\lceil \frac{2}{\Delta^2} \ln \frac{\sqrt{K} N_\Delta}{\delta} \right\rceil$. Then, τ_i can be upper-bounded by T_Δ for any arm i as the following theorem.

Theorem 2 *Inequality $\tau_i \leq T_\Delta$ holds for $i = 1, \ldots, K$.*

Proof See "Appendix B". □

How good is the worst case bound T_Δ on the number of samples for each arm compared to the case with LB $= \underline{\mu}'$ and UB $= \overline{\mu}'$ (Eq. 3)? It is shown by the following theorem that, in BAEC$[*, \underline{\mu}', \overline{\mu}']$, the number of arm draws τ_i' for some arm i, which is corresponding to τ_i, can be larger than $T_\Delta' = \lfloor \frac{2}{\Delta^2} \ln \frac{448K}{\Delta^4\delta} \rfloor$, which means $\tau_i' - \tau_i = \Omega\left(\frac{1}{\Delta^2} \ln \frac{\sqrt{K}}{\Delta^2}\right)$ if $\frac{1}{\delta} = o\left(\mathrm{e}^{\sqrt{K}/\Delta^2}\right)$.

Theorem 3 *Consider algorithm* BAEC$[*, \underline{\mu}', \overline{\mu}']$ *and define* $\tau_i' = \min\{n \mid \underline{\mu}_i'(n) \geq \theta_L$ *or* $\overline{\mu}_i'(n) < \theta_U\}$ *for* $i = 1, \ldots, K$. *Then, event* $\tau_i' > T_\Delta'$ *can happen for* $i = 1, \ldots, K$, *where* T_Δ' *is defined as* $T_\Delta' = \lfloor \frac{2}{\Delta^2} \ln \frac{448K}{\Delta^4\delta} \rfloor$. *Furthermore, the difference between the worst case decision times* $\tau_i' - \tau_i$ *is lower-bounded as*

$$\tau_i' - \tau_i > T_\Delta' - T_\Delta > \frac{2}{\Delta^2}\left(\ln \frac{52\sqrt{K}}{\Delta^2} - \ln\ln \frac{3\sqrt{K}}{\Delta^2\delta}\right).$$

Proof See “Appendix C”. □

Remark 3 Theorem 3 says that the difference between the worst case decision times τ_i' and τ_i of arm i is $\Omega\left(\frac{1}{\Delta^2} \ln \frac{\sqrt{K}}{\Delta^2}\right)$ for $\delta = \omega\left(\frac{\sqrt{K}}{\Delta^2}\mathrm{e}^{-\frac{52\sqrt{K}}{\Delta^2}}\right)$ under the condition that $\delta > \frac{3\sqrt{K}}{\Delta^2}\mathrm{e}^{-\frac{52\sqrt{K}}{\Delta^2}}$. In the experimental setting of Sect. 7.1, in which parameters $K = 100$, $(\Delta, \delta) = (0.2, 0.01), (0.2, 0.001), (0.02, 0.01), (0.02, 0.001)$ are used, the lower bounds of $\tau_i' - \tau_i$ calculated using the above inequality are 352.7, 343.4, 56579.7, 55900.7, respectively, which seem relatively large compared to the corresponding $T_\Delta = 684, 808, 93098, 105307$. The range of δ which guarantees that the lower bound of $\tau_i' - \tau_i$ is positive, is $> 1.11 \times 10^{-5643}$ for $\Delta = 0.2$ and 1.12×10^{-564578} for $\Delta = 0.02$.

Remark 4 Instead of $\overline{\mu}_i'(n)$ defined in Eq. (3), $\overline{\mu}_i''(n) = \hat{\mu}_i(n) + \sqrt{\frac{1}{2n} \ln \frac{2n^2}{\delta}}$ can be used because an union bound is not necessary for a positive arm as $\overline{\mu}_i(n)$ defined in Eq. (4). For the algorithm BAEC$[*, \underline{\mu}', \overline{\mu}'']$ using this upper confidence bound $\overline{\mu}_i''(n)$ $(i = 1, \ldots, K)$, the decision time difference from τ_i is still lower-bounded by $\frac{2}{\Delta^2}\left(\ln \frac{2}{K^{\frac{1}{4}}\Delta^2} - \ln\ln \frac{3\sqrt{K}}{\Delta^2\delta}\right)$ by Theorem 9 in “Appendix D”. The values of this lower bound for the experimental setting of Sect. 7.1, that is, $K = 100$, $(\Delta, \delta) = (0.2, 0.01), (0.2, 0.001), (0.02, 0.01), (0.02, 0.001)$, are 17.13, 7.80, 23020.3, 22341.3, respectively. Compared to the corresponding $T_\Delta = 684, 808, 93098, 105307$, the difference seems still large for $\Delta = 0.02$ though it becomes small for $\Delta = 0.2$. The range of δ guaranteeing positiveness of the lower bound is $> 1.03 \times 10^{-5}$ for $\Delta = 0.2$ and 1.63×10^{-682} for $\Delta = 0.02$.

5.2 Algorithm's correctness

In this subsection, we prove that algorithm BAEC$[*, \underline{\mu}, \overline{\mu}]$ is a $(\theta_L, \theta_U, \delta)$-BAEC algorithm.

We define events $\mathcal{E}^+$ and $\mathcal{E}^-$as

$$\mathcal{E}^+ = \bigcup_{i:\mu_i \geq \theta_U} \bigcap_{n=1}^{T_\Delta} \left\{\overline{\mu}_i(n) \geq \mu_i\right\}, \mathcal{E}^- = \bigcap_{i=1}^{K} \bigcap_{n=1}^{T_\Delta} \left\{\underline{\mu}_i(n) < \mu_i\right\}.$$

Note that algorithm BAEC$[*, \underline{\mu}, \overline{\mu}]$ returns “positive” under the event $\mathcal{E}^+$ and returns “negative” under the event $\mathcal{E}^-$. For any event $\mathcal{E}$, we let $\mathbb{1}\{\mathcal{E}\}$ denote an indicator function of $\mathcal{E}$, that is, $\mathbb{1}\{\mathcal{E}\} = 1$ if $\mathcal{E}$ occurs and $\mathbb{1}\{\mathcal{E}\} = 0$ otherwise.

The following proposition is used to prove Lemma 1.

Proposition 1 $T_\Delta \leq N_\Delta$.

Proof See "Appendix E". □

The next lemma says that algorithm's output is correct with probability at least $1 - \delta$ in the cases that at least one positive arm exists or all the arms are negative.

Lemma 1 *For the complementary events* $\overline{\mathcal{E}^+}$, $\overline{\mathcal{E}^-}$ *of events* $\mathcal{E}^+$, $\mathcal{E}^-$, *inequality* $\mathbb{P}\{\overline{\mathcal{E}^+}\} \leq \delta$ *holds when* $\mu_1 \geq \theta_U$ *and inequality* $\mathbb{P}\{\overline{\mathcal{E}^-}\} \leq \delta$ *holds when* $\mu_1 < \theta_L$.

Proof Assume that $\mu_1 \geq \theta_U$. Using De Morgan's laws, $\overline{\mathcal{E}^+}$ can be expressed as

$$\begin{aligned}\overline{\mathcal{E}^+} &= \bigcap_{i:\mu_i \geq \theta_U} \bigcup_{n=1}^{T_\Delta} \left\{\overline{\mu}_i(n) < \mu_i\right\} \\ &= \bigcap_{i:\mu_i \geq \theta_U} \bigcup_{n=1}^{T_\Delta} \left\{\hat{\mu}_i(n) < \mu_i - \sqrt{\frac{1}{2n} \ln \frac{N_\Delta}{\delta}}\right\}.\end{aligned}$$

So, the probability that event $\overline{\mathcal{E}^+}$ occurs is bounded by δ using Hoeffding's inequality:

$$\begin{aligned}\mathbb{P}\{\overline{\mathcal{E}^+}\} &\leq \max_{i:\mu_i \geq \theta_U} \sum_{n=1}^{T_\Delta} \mathbb{P}\left\{\hat{\mu}_i(n) < \mu_i - \sqrt{\frac{1}{2n} \ln \frac{N_\Delta}{\delta}}\right\} \\ &\leq \sum_{n=1}^{T_\Delta} \frac{\delta}{N_\Delta} = \frac{T_\Delta}{N_\Delta}\delta \leq \delta. \quad \text{(by Proposition 1)}\end{aligned}$$

Assume that $\mu_1 < \theta_L$. Using De Morgan's laws, $\overline{\mathcal{E}^-}$ can be expressed as

$$\begin{aligned}\overline{\mathcal{E}^-} &= \bigcup_{i=1}^{K} \bigcup_{n=1}^{T_\Delta} \left\{\underline{\mu}_i(n) \geq \mu_i\right\} \\ &= \bigcup_{i=1}^{K} \bigcup_{n=1}^{T_\Delta} \left\{\hat{\mu}_i(n) \geq \mu_i + \sqrt{\frac{1}{2n} \ln \frac{KN_\Delta}{\delta}}\right\}.\end{aligned}$$

So, the probability that event $\overline{\mathcal{E}^-}$ occurs is bounded by δ using the union bound and Hoeffding's inequality:

$$\begin{aligned}\mathbb{P}\{\overline{\mathcal{E}^-}\} &\leq \sum_{i=1}^{K} \sum_{n=1}^{T_\Delta} \mathbb{P}\left\{\hat{\mu}_i(n) \geq \mu_i + \sqrt{\frac{1}{2n} \ln \frac{KN_\Delta}{\delta}}\right\} \\ &\leq \sum_{i=1}^{K} \sum_{n=1}^{T_\Delta} \frac{\delta}{KN_\Delta} = \frac{T_\Delta}{N_\Delta}\delta \leq \delta. \quad \text{(by Proposition 1)}\end{aligned}$$

□

The following theorem states that algorithm BAEC[$*, \underline{\mu}, \overline{\mu}$] is a $(\theta_L, \theta_U, \delta)$-BAEC algorithm which needs at most KT_Δ samples in the worst case.

Theorem 4 *Algorithm* BAEC[$*, \underline{\mu}, \overline{\mu}$] *is a* $(\theta_L, \theta_U, \delta)$*-BAEC algorithm that stops after at most* KT_Δ *arm draws.*

Proof By the definition of τ_i, algorithm BAEC$[*, \underline{\mu}, \overline{\mu}]$ draws arm i at most τ_i times, which is upper-bounded by T_{Δ} due to Theorem 2. So, algorithm BAEC$[*, \underline{\mu}, \overline{\mu}]$ stops after at most KT_{Δ} arm draws.

When at least one arm is positive, that is, in the case with $\mu_1 \geq \theta_U$, algorithm BAEC$[*, \underline{\mu}, \overline{\mu}]$ returns "positive" if event $\mathcal{E}^+$ occurs. Thus, algorithm BAEC$[*, \underline{\mu}, \overline{\mu}]$ returns "positive" with probability $\mathbb{P}\{\mathcal{E}^+\} = 1 - \mathbb{P}\{\overline{\mathcal{E}^+}\} \geq 1 - \delta$ by Lemma 1. When all the arms are negative, that is, in the case with $\mu_1 < \theta_L$, algorithm BAEC$[*, \underline{\mu}, \overline{\mu}]$ returns "negative" if event $\mathcal{E}^-$ occurs. Thus, algorithm BAEC$[*, \underline{\mu}, \overline{\mu}]$ returns "negative" with probability $\mathbb{P}\{\mathcal{E}^-\} = 1 - \mathbb{P}\{\overline{\mathcal{E}^-}\} \geq 1 - \delta$ by Lemma 1. □

5.3 High-probability and average-case bounds

By Theorem 4, we know worst-case upper bound KT_{Δ} on the number of samples needed for algorithm BAEC$[*, \underline{\mu}, \overline{\mu}]$. In this section, we show a high-probability and an average-case bounds for the algorithm.

We define Δ_i as

$$\Delta_i = \begin{cases} \mu_i - \theta_L & (\mu_i \geq \theta) \\ \theta_U - \mu_i & (\mu_i < \theta) \end{cases}$$

and let T_{Δ_i} denote $T_{\Delta_i} = \left\lceil \frac{2}{\Delta_i^2} \ln \frac{\sqrt{K} N_{\Delta}}{\delta} \right\rceil$.

A high-probability upper bound of the number of samples needed for algorithm BAEC$[*, \underline{\mu}, \overline{\mu}]$ is shown in the next theorem. Compared to worst case bound, KT_{Δ} can be improved to $\sum_{i=1}^{K} T_{\Delta_i}$ in the case with $\mu_1 < \theta_L$, however, only one T_{Δ} is guaranteed to be improved to the maximum T_{Δ_i} among those of positive arms i in the case with $\mu_1 \geq \theta_U$.

Theorem 5 *In algorithm* BAEC$[*, \underline{\mu}, \overline{\mu}]$*, inequality* $\tau_i \leq T_{\Delta_i}$ *holds for at least one positive arm* i *with probability at least* $1 - \delta$ *when* $\mu_1 \geq \theta_U$*. Inequality* $\tau_i \leq T_{\Delta_i}$ *holds for all the arm* $i = 1, \ldots, K$ *with probability at least* $1 - \delta$ *when* $\mu_1 < \theta_L$*. As a result, with probability at least* $1 - \delta$*, the stopping time* T *of algorithm* BAEC$[*, \underline{\mu}, \overline{\mu}]$ *is upper-bounded as* $T \leq \max_{i:\mu_i \geq \theta_U} T_{\Delta_i} + (K-1)T_{\Delta}$ *when* $\mu_1 \geq \theta_U$ *and* $T \leq \sum_{i=1}^{K} T_{\Delta_i}$ *when* $\mu_1 < \theta_L$.

Proof See "Appendix F". □

The last sample complexity upper bound for algorithm BAEC$[*, \underline{\mu}, \overline{\mu}]$ is an upper bound on the expected number of samples. Compared to the high-probability bound, $T_{\Delta_i} = \left\lceil \frac{2}{\Delta_i^2} \ln \frac{\sqrt{K} N_{\Delta}}{\delta} \right\rceil$ is improved to $\frac{1}{2\Delta_i^2} \ln \frac{K N_{\Delta}}{\delta}$ or $\frac{1}{2\Delta_i^2} \ln \frac{N_{\Delta}}{\delta}$.

Theorem 6 *For algorithm* BAEC$[*, \underline{\mu}, \overline{\mu}]$*, the expected value of* τ_i *of each arm* i *is upper-bounded as follows.*

$$\mathbb{E}[\tau_i] \leq \begin{cases} \frac{1}{2\Delta_i^2} \ln \frac{K N_{\Delta}}{\delta} + O\left(\left(\ln \frac{K N_{\Delta}}{\delta}\right)^{2/3}\right) & (\mu_i \geq \theta) \\ \frac{1}{2\Delta_i^2} \ln \frac{N_{\Delta}}{\delta} + O\left(\left(\ln \frac{N_{\Delta}}{\delta}\right)^{2/3}\right) & (\mu_i < \theta) \end{cases}$$

As a result, the expected stopping time $\mathbb{E}[T]$ *of algorithm* BAEC$[*, \underline{\mu}, \overline{\mu}]$ *is upper-bounded as*

$$\mathbb{E}[T] \leq \frac{1}{2} \ln \frac{N_\Delta}{\delta} \sum_{i=1}^{K} \frac{1}{\Delta_i^2} + \frac{\ln K}{2} \sum_{i:\mu_i \geq \theta} \frac{1}{\Delta_i^2} + O\left(K\left(\ln \frac{K N_\Delta}{\delta}\right)^{2/3}\right). \tag{7}$$

The above theorem can be easily derived from the following lemma by setting event $\mathcal{E}$ to a *certain event* (an event that occurs with probability 1).

Lemma 2 *For any event* $\mathcal{E}$*, in algorithm* BAEC$[*, \underline{\mu}, \overline{\mu}]$*, inequality*

$$\mathbb{E}[\tau_i \mathbb{1}\{\mathcal{E}\}] \leq \frac{\mathbb{P}[\mathcal{E}]}{2\Delta_i^2} \ln \frac{K N_\Delta}{\delta} + O\left(\left(\ln \frac{K N_\Delta}{\delta}\right)^{\frac{2}{3}}\right). \tag{8}$$

holds for any arm i *with* $\mu_i \geq \theta$ *and*

$$\mathbb{E}[\tau_i \mathbb{1}\{\mathcal{E}\}] \leq \frac{\mathbb{P}[\mathcal{E}]}{2\Delta_i^2} \ln \frac{N_\Delta}{\delta} + O\left(\left(\ln \frac{N_\Delta}{\delta}\right)^{\frac{2}{3}}\right). \tag{9}$$

holds for any arm i *with* $\mu_i < \theta$.

Proof See “Appendix G”. □

Remark 5 When all the arms have Bernoulli loss distributions with means less than θ_L, by Pinsker’s Inequality $d(x, y) \geq 2(x - y)^2$, the right-hand side of Ineq. (2) in Theorem 1 can be upper-bounded as

$$\sum_{i=1}^{K} \frac{1-2\delta}{d(\mu_i, \theta_U)} \ln \frac{1-\delta}{\delta} \leq \sum_{i=1}^{K} \frac{1-2\delta}{2\Delta_i^2} \ln \frac{1-\delta}{\delta}.$$

Since Pinsker’s Inequality is tight in the worst case, algorithm BAEC$[*, \underline{\mu}, \overline{\mu}]$ is almost asymptotically optimal as $\delta \to +0$. Algorithm BAEC$[*, \underline{\mu}, \overline{\mu}]$ is a kind of elimination algorithm, that is, the arms that satisfy negative decision condition are eliminated. Excluding elimination algorithms, UCB and Murphy Sampling coupled with a *box stopping rule* is known to also have asymptotically optimal stopping time in this case when $\Delta = 0$ (Kaufmann et al. 2018).

6 Sample complexity of algorithm BAEC[APT$_\mathrm{P}$, $\underline{\mu}$, $\overline{\mu}$]

6.1 Sample complexity upper bound

If all the arms are judged as negative in algorithm BAEC[ASP, $\underline{\mu}$, $\overline{\mu}$], that is, drawing arm i is stopped by the decision condition of $\overline{\mu}_i(\tau_i) < \theta_U$ for all $i = 1, \ldots, K$, the stopping time T is $\sum_{i=1}^{K} \tau_i$ regardless of arm-selection policy ASP. In the case that some positive arms exist, however, the stopping time depends on how fast the $(\theta_L, \theta_U, \delta)$-BAEC algorithm can find one of positive arms.

In this subsection, we prove upper bounds on the expected number of samples needed for algorithm BAEC[APT$_\mathrm{P}$, $\underline{\mu}$, $\overline{\mu}$], an instance of algorithm BAEC$[*, \underline{\mu}, \overline{\mu}]$ with specific arm-selection policy APT$_\mathrm{P}$.

Let arm $\hat{i}_1$ denote the first arm that is drawn τ_i times in algorithm BAEC[APT$_\mathrm{P}$, $\underline{\mu}$, $\overline{\mu}$]. In addition to Δ_i, we also use $\underline{\Delta}_i = |\mu_i - \theta|$ in the following analysis. We let m denote the number of arms i with $\mu_i \geq \theta$. The event that arm i is judged as positive is denoted as $\mathcal{E}_i^{\mathrm{POS}}$.

From the following theorem and corollary, we know that, when δ is small, the dominant terms of our upper bound on the expected stopping time of algorithm BAEC[APT$_\mathrm{P}$, $\underline{\mu}$, $\overline{\mu}$], are $\frac{\mathbb{P}\left[\hat{i}_1 = i, \mathcal{E}_i^{\mathrm{POS}}\right]}{2\Delta_i^2} \ln \frac{1}{\delta}$ $(i = 1, \ldots, m)$, whose sum is between $\frac{1}{2\Delta_1^2} \ln \frac{1}{\delta}$ and $\frac{1}{2\Delta_m^2} \ln \frac{1}{\delta}$.

Theorem 7 *If $m \geq 1$ (or $\mu_1 \geq \theta$), then the expected stopping time $\mathbb{E}[T]$ of algorithm* BAEC[APT$_\mathrm{P}$, $\underline{\mu}$, $\overline{\mu}$] *is upper-bounded as*

$$\begin{aligned}\mathbb{E}[T] \leq & \sum_{i=1}^{m} \left(\frac{\mathbb{P}\left[\hat{i}_1 = i, \mathcal{E}_i^{\mathrm{POS}}\right]}{2\Delta_i^2} \ln \frac{K N_{\vartriangle}}{\delta} + \frac{2(m-1)}{\underline{\Delta}_i^4} + \left(\frac{1}{\underline{\Delta}_i^2} + 4\right) \sum_{j=m+1}^{K} \frac{1}{\underline{\Delta}_j^2} \right) \\ & + m(K-m) + O\left(m \left(\ln \frac{K N_{\vartriangle}}{\delta} \right)^{\frac{2}{3}} \right) \\ & + K T_{\vartriangle} \left(\frac{\mathrm{e}^{2\underline{\Delta}_i^2}}{2\underline{\Delta}_i^2} \sum_{i=1}^{m} \left(\frac{\delta}{N_{\vartriangle}} \right)^{\left(\frac{\underline{\Delta}_i}{\max\{\theta_U, 1-\theta_L\}}\right)^2} \right. \\ & \qquad \left. + \left(1 + \frac{1}{2\underline{\Delta}_1^2}\right) \sum_{i=m+1}^{K} \left(\frac{\delta}{N_{\vartriangle}} \right)^{\frac{1}{4}\left(\frac{\underline{\Delta}_i}{\max\{\theta_U, 1-\theta_L\}}\right)^2} \right)\end{aligned}$$

Proof See "Appendix H". □

The next corollary is easily derived from Theorem 7.

Corollary 1 *If $m \geq 1$, then*

$$\lim_{\delta \to +0} \frac{\mathbb{E}[T]}{\ln \frac{1}{\delta}} \leq \sum_{i=1}^{m} \frac{\lim_{\delta \to +0} \mathbb{P}\left[\hat{i}_1 = i, \mathcal{E}_i^{\mathrm{POS}}\right]}{2\Delta_i^2} \leq \frac{1}{2\Delta_m^2}$$

holds for the expected stopping time $\mathbb{E}[T]$ of algorithm BAEC[APT$_\mathrm{P}$, $\underline{\mu}$, $\overline{\mu}$].

6.2 Comparison with BAEC[UCB, $\underline{\mu}$, $\overline{\mu}$]

HDoC (Hybrid algorithm for the Dilemma of Confidence)(Kano et al. 2017) for good arm identification problem uses arm selection policy UCB (Upper Confidence Bound) (Auer et al. 2002), in which

$$\mathrm{UCB}(t, i) = \begin{cases} \infty & (n_i(t) = 0) \\ \hat{\mu}_i(n_i(t)) + \sqrt{\frac{1}{2n_i(t)} \ln t} & (n_i(t) > 0) \end{cases}$$

is used as ASP(t, i). In this section, we analyze a sample complexity upper bound of algorithm[6] BAEC[UCB, $\underline{\mu}$, $\overline{\mu}$] and compare it with that of BAEC[APT$_\mathrm{P}$, $\underline{\mu}$, $\overline{\mu}$].

[6] This is not completely the same algorithm as HDoC because, in the HDoC's decision condition, bounds $\hat{\mu}_i(n_i(t)) \pm \sqrt{\frac{1}{2n_i(t)} \ln \frac{4Kn_i(t)^2}{\delta}}$ are used.

Define Δ_{1i} as $\Delta_{1i} = \mu_1 - \mu_i$. Then, we can obtain the following theorem and corollary, from which, we know that, when δ is small, the dominant terms of our upper bound on the expected stopping time of algorithm BAEC[UCB, $\underline{\mu}, \overline{\mu}$], are $\frac{1}{2\Delta_i^2} \ln \frac{1}{\delta}$ $(i : \mu_i = \mu_1)$, whose sum is $\frac{|\{i|\mu_i=\mu_1\}|}{2\Delta_1^2} \ln \frac{1}{\delta}$.

Theorem 8 *If $m \geq 1$, then expected stopping time $\mathbb{E}[T]$ of algorithm* BAEC[UCB, $\underline{\mu}, \overline{\mu}$] *is upper-bounded as*

$$\begin{aligned}\mathbb{E}[T] \leq & \sum_{i:\mu_i=\mu_1} \left(\frac{1}{2\Delta_i^2} \ln \frac{KN_{\scriptscriptstyle\Delta}}{\delta} + O\left(\left(\ln \frac{KN_{\scriptscriptstyle\Delta}}{\delta} \right)^{\frac{2}{3}} \right) \right) \\ & + \sum_{i:\mu_i<\mu_1} \left(\frac{\ln KT_{\scriptscriptstyle\Delta}}{2\Delta_{1i}^2} + O((\ln KT_{\scriptscriptstyle\Delta})^{\frac{2}{3}}) \right) \\ & + O((\ln KT_{\scriptscriptstyle\Delta})^{\frac{2}{3}} \ln \ln KT_{\scriptscriptstyle\Delta}) + \frac{\mathrm{e}^{2\underline{\Delta}_1^2} KT_{\scriptscriptstyle\Delta}}{2\underline{\Delta}_1^2} \left(\frac{\delta}{N_{\scriptscriptstyle\Delta}} \right)^{\left(\frac{\underline{\Delta}_1}{\max\{\theta_U, 1-\theta_L\}} \right)^2}.\end{aligned}$$

Proof See “Appendix I”. □

Corollary 2 *If $m \geq 1$, then*

$$\lim_{\delta \to +0} \frac{\mathbb{E}[T]}{\ln \frac{1}{\delta}} \leq \frac{|\{i \mid \mu_i = \mu_1\}|}{2\Delta_1^2}$$

holds for the expected stopping time $\mathbb{E}[T]$ of algorithm BAEC[UCB, $\underline{\mu}, \overline{\mu}$].

Remark 6 From the upper bound shown by Ineq. (7), inequality

$$\lim_{\delta \to +0} \frac{\mathbb{E}[T]}{\ln \frac{1}{\delta}} \leq \sum_{i=1}^{K} \frac{1}{2\Delta_i^2}$$

is derived. This means that the expected stopping time upper bounds for algorithm BAEC[APT$_\mathrm{P}$, $\underline{\mu}, \overline{\mu}$] and BAEC[UCB, $\underline{\mu}, \overline{\mu}$] shown in Theorems 7 and 8 are asymptotically smaller than that of algorithm BAEC[$*$, $\underline{\mu}, \overline{\mu}$] as $\delta \to +0$.

Remark 7 When all the arms have Bernoulli loss distributions, the right-hand side of Ineq. (1) in Theorem 1 can be upper-bounded as

$$\frac{1-2\delta}{d(\mu_1, \theta_L)} \ln \frac{1-\delta}{\delta} \leq \frac{1-2\delta}{2\Delta_1^2} \ln \frac{1-\delta}{\delta}$$

by Pinsker’s Inequality. Considering tightness of Pinsker’s Inequality, $\frac{1}{2\Delta_1^2}$ is considered to be a tight upper bound of $\lim_{\delta \to +0} \frac{\mathbb{E}[T]}{\ln \frac{1}{\delta}}$ if Ineq. (1) is tight. There is a large gap between $\sum_{i=1}^{m} \frac{\lim_{\delta \to +0} \mathbb{P}\left[\hat{i}_1 = i, \mathcal{E}_i^{\mathrm{POS}}\right]}{2\Delta_i^2}$ and $\frac{1}{2\Delta_1^2}$, and improvement of the upper bound on the number of samples for APT$_\mathrm{P}$ seems difficult, so the algorithm BAEC with arm selection policy APT$_\mathrm{P}$ does not seem asymptotically optimal unless $\lim_{\delta \to +0} \mathbb{P}\left[\hat{i}_1 = 1, \mathcal{E}_i^{\mathrm{POS}}\right] = 1$. On the other hand, $\lim_{\delta \to +0} \frac{\mathbb{E}[T]}{\ln \frac{1}{\delta}}$ for UCB is upper-bounded by $\frac{1}{2\Delta_1^2}$, that is, asymptotically optimal when $\mu_i < \mu_1$ for all arm $i \neq 1$. In the case with $\mu_i = \mu_1$ for all $i = 1, \ldots, m$, however, $\lim_{\delta\to+0}\frac{\mathbb{E}[T]}{\ln\frac{1}{\delta}} \le \frac{m}{2\Delta_1^2}$ holds for UCB while the corresponding bound for APT$_\mathrm{P}$ is asymptotically optimal, that is, $\lim_{\delta\to+0}\frac{\mathbb{E}[T]}{\ln\frac{1}{\delta}} \le \frac{1}{2\Delta_1^2}$ holds. The stopping time's asymptotic optimality of Murphy Sampling coupled with a *box stopping rule* (Kaufmann et al. 2018) for $\Delta = 0$ is basically the same as that of BAEC[UCB, $\underline{\mu}$, $\overline{\mu}$] for $\Delta > 0$; its stopping time is optimal in the unique-best-arm case but not in the multiple-best-arms case.

Remark 8 Comparing non-dominant terms of BAEC[APT$_\mathrm{P}$, $\underline{\mu}$, $\overline{\mu}$] and BAEC[UCB, $\underline{\mu}$, $\overline{\mu}$], a cause for the large upper bound of the expected stopping time can be the existence of arms i whose loss mean μ_i is close to μ_1 in BAEC[UCB, $\underline{\mu}$, $\overline{\mu}$] while it can be the existence of arms i whose loss mean μ_i is close to θ in BAEC[APT$_\mathrm{P}$, $\underline{\mu}$, $\overline{\mu}$].

7 Experiments

In this section, we report the results of our experiments that were conducted in order to demonstrate the effectiveness of our Δ-dependent asymmetric confidence bounds used in decision condition and arm selection policy on the stopping time.

In all the tables of experimental results, the smallest averaged stopping time in each parameter setting is bolded or italic, and bolded ones mean statistically significant difference.

7.1 Effectiveness of Δ-dependent asymmetric confidence bounds

As upper and lower confidence bounds LB and UB, we proposed $\underline{\mu}$ and $\overline{\mu}$ based on Δ-dependent asymmetric bounds $\overline{\mu}_i(n)$ and $\underline{\mu}_i(n)$ defined by Eq. (4), instead of $\underline{\mu}'$ and $\overline{\mu}'$ based on conventional non-Δ-dependent symmetric bounds $\overline{\mu}_i'(n)$ and $\underline{\mu}_i'(n)$ defined by Eq. (3). In this subsection, we empirically compare the number of draws for an arm with mean μ_i to satisfy the decision condition using those bounds.

In the experiment, an i.i.d. loss sequence $X_i(1),\ldots$ was generated according to a Bernoulli distribution with mean μ_i and we measured the decision time τ_i which is the smallest n that satisfies the decision condition ($\underline{\mu}_i(n) \ge \theta_L$ or $\overline{\mu}_i(n) < \theta_U$). The decision times were averaged over 100 runs for each combination of parameters $\delta = 0.001, 0.01$, $\mu_i = 0.2, 0.4, 0.6, 0.8$ and $(\theta_L, \theta_U) = (0.1, 0.3), (0.3, 0.5), (0.5, 0.7), (0.7, 0.9), (0.19, 0, 21), (0.39, 0.41), (0.59, 0.61), (0.79, 0.81)$. Note that $\Delta = \theta_U - \theta_L = 0.2$ for the first half of the setting and $\Delta = 0.02$ for the last half of the setting. We used $K = 100$ so as to make the bounds asymmetric. As a result, $\alpha = 1.154, 1.186$ for $\delta = 0.001, 0.01$, respectively. So, θ is $(\theta_L + \theta_U)/2 + 0.007$ for $\delta = 0.001$ and $(\theta_L + \theta_U)/2 + 0.009$ for $\delta = 0.01$.

The result is shown in Table 3. As we can see from the table, the decision condition using Δ-dependent asymmetric bounds make the decision time fast compared to that using conventional bounds except in the case with $\Delta = 0.02$ and $\mu_i > \theta$. The effect of the proposed Δ-dependent asymmetric confidence bounds become significant when the arm is neutral or negative, notably, 1.74∼2.08 times faster when $\mu_i \approx \theta$. The reason why the decision condition using conventional bounds performs better for $\Delta = 0.02$ and $\mu_i > \theta$, is that $\underline{\mu}_i'(\tau_i') > \underline{\mu}_i(\tau_i')$ occurs frequently for decision time τ_i' using $\underline{\mu}'$. In fact, $\underline{\mu}_i'(n) > \underline{\mu}_i(n)$ holds for $n < \sqrt{N_\Delta/2}$, and $\sqrt{N_\Delta/2} = 246.99, 264.79$ for $\delta = 0.01, 0.001$, respectively, in the case with $\Delta = 0.02$ and $K = 100$. The width $\hat{\mu}_i(n) - \underline{\mu}_i'(n)$ of the lower confidence interval of $\underline{\mu}_i'(n)$ is 0.206 for $\delta = 0.01$ and $n = 246$, and 0.210 for $\delta = 0.001$ and $n = 264$, Thus, arm i with mean μ_i larger than θ by more than 0.21 is more likely to satisfy condition

Table 3 Number of draws of arm i with loss mean μ_i until the decision condition is satisfied

$\varDelta$	$\delta(T_\varDelta)$	$\begin{pmatrix}\theta_U\\\theta_L\end{pmatrix}$	LB, UB	$\mu_i = 0.2$	$\mu_i = 0.4$	$\mu_i = 0.6$	$\mu_i = 0.8$
0.2	0.01(684)	$\begin{pmatrix}0.3\\0.1\end{pmatrix}$	$\underline{\mu}, \overline{\mu}$	**497.64 ± 28.47**	*88.87 ± 8.61*	*35.16 ± 2.88*	16.96 ± 1.23
			$\underline{\mu}', \overline{\mu}'$	957.81 ± 37.56	104.95 ± 10.33	37.04 ± 3.50	*16.82 ± 1.28*
		$\begin{pmatrix}0.5\\0.3\end{pmatrix}$	$\underline{\mu}, \overline{\mu}$	**63.24 ± 4.88**	**427.10 ± 32.90**	*86.50 ± 8.18*	*30.79 ± 1.91*
			$\underline{\mu}', \overline{\mu}'$	106.67 ± 7.83	889.36 ± 51.33	103.44 ± 9.95	32.26 ± 2.35
		$\begin{pmatrix}0.7\\0.5\end{pmatrix}$	$\underline{\mu}, \overline{\mu}$	**23.52 ± 1.87**	**63.79 ± 6.54**	**435.91 ± 37.37**	**91.56 ± 6.99**
			$\underline{\mu}', \overline{\mu}'$	35.01 ± 2.46	105.13 ± 9.73	885.55 ± 47.61	109.35 ± 7.93
		$\begin{pmatrix}0.9\\0.7\end{pmatrix}$	$\underline{\mu}, \overline{\mu}$	**13.90 ± 0.95**	**24.85 ± 2.32**	**65.07 ± 6.50**	**500.93 ± 29.47**
			$\underline{\mu}', \overline{\mu}'$	17.60 ± 1.36	34.86 ± 3.10	106.25 ± 10.03	963.05 ± 34.02
	0.001(808)	$\begin{pmatrix}0.3\\0.1\end{pmatrix}$	$\underline{\mu}, \overline{\mu}$	**595.24 ± 31.77**	**102.05 ± 8.26**	*37.26 ± 3.11*	18.07 ± 1.19
			$\underline{\mu}', \overline{\mu}'$	1072.65 ± 43.99	123.16 ± 10.42	39.68 ± 3.77	*17.37 ± 1.31*
		$\begin{pmatrix}0.5\\0.3\end{pmatrix}$	$\underline{\mu}, \overline{\mu}$	**75.92 ± 5.17**	**560.31 ± 34.91**	*100.66 ± 9.37*	*38.76 ± 2.50*
			$\underline{\mu}', \overline{\mu}'$	123.73 ± 8.64	980.23 ± 49.95	119.85 ± 11.64	41.10 ± 2.91
		$\begin{pmatrix}0.7\\0.5\end{pmatrix}$	$\underline{\mu}, \overline{\mu}$	**29.43 ± 2.14**	**73.87 ± 6.71**	**546.24 ± 37.43**	**107.93 ± 7.50**
			$\underline{\mu}', \overline{\mu}'$	41.32 ± 2.51	116.51 ± 9.38	969.24 ± 53.78	126.04 ± 8.71
		$\begin{pmatrix}0.9\\0.7\end{pmatrix}$	$\underline{\mu}, \overline{\mu}$	**15.50 ± 1.05**	**29.33 ± 2.50**	**76.96 ± 7.08**	**599.91 ± 29.82**
			$\underline{\mu}', \overline{\mu}'$	19.62 ± 1.33	40.21 ± 3.31	117.49 ± 10.16	1075.36 ± 39.92

Table 3 continued

Δ	$\delta(T_\Delta)$	$\begin{pmatrix}\theta_U\\ \theta_L\end{pmatrix}$	LB, UB	$\mu_i = 0.2$	$\mu_i = 0.4$	$\mu_i = 0.6$	$\mu_i = 0.8$
0.02	0.01(93098)	$\begin{pmatrix}0.21\\ 0.19\end{pmatrix}$	$\underline{\mu}, \overline{\mu}$	**70.99 ± 3.76**$_{\times 10^3}$	231.03 ± 16.83	65.58 ± 4.97	27.32 ± 1.57
			$\underline{\mu}', \overline{\mu}'$	142.60 ± 4.86$_{\times 10^3}$	*228.81 ± 18.52*	**55.12 ± 5.12**	**20.77 ± 1.62**
		$\begin{pmatrix}0.41\\ 0.39\end{pmatrix}$	$\underline{\mu}, \overline{\mu}$	**187.22 ± 14.34**	**66.31 ± 4.15**$_{\times 10^3}$	232.18 ± 19.90	62.03 ± 4.34
			$\underline{\mu}', \overline{\mu}'$	232.78 ± 17.24	133.22 ± 5.86$_{\times 10^3}$	*229.23 ± 21.52*	**52.51 ± 4.30**
		$\begin{pmatrix}0.61\\ 0.59\end{pmatrix}$	$\underline{\mu}, \overline{\mu}$	*51.55 ± 3.51*	**182.52 ± 16.05**	**63.71 ± 4.06**$_{\times 10^3}$	247.35 ± 15.52
			$\underline{\mu}', \overline{\mu}'$	55.97 ± 3.89	226.82 ± 20.01	130.31 ± 5.76$_{\times 10^3}$	*244.38 ± 17.50*
		$\begin{pmatrix}0.81\\ 0.79\end{pmatrix}$	$\underline{\mu}, \overline{\mu}$	24.05 ± 1.56	*48.43 ± 4.12*	**178.58 ± 15.18**	**69.11 ± 3.48**$_{\times 10^3}$
			$\underline{\mu}', \overline{\mu}'$	*23.52 ± 1.86*	52.85 ± 4.95	226.95 ± 20.13	140.78 ± 4.57$_{\times 10^3}$
	0.001(105307)	$\begin{pmatrix}0.21\\ 0.19\end{pmatrix}$	$\underline{\mu}, \overline{\mu}$	**81.82 ± 4.09**$_{\times 10^3}$	268.02 ± 19.15	71.79 ± 5.47	33.38 ± 1.85
			$\underline{\mu}', \overline{\mu}'$	152.20 ± 5.39$_{\times 10^3}$	*264.47 ± 20.82*	*63.83 ± 5.26*	**26.72 ± 2.11**
		$\begin{pmatrix}0.41\\ 0.39\end{pmatrix}$	$\underline{\mu}, \overline{\mu}$	**209.27 ± 14.12**	**76.89 ± 4.03**$_{\times 10^3}$	*263.45 ± 19.38*	66.83 ± 4.03
			$\underline{\mu}', \overline{\mu}'$	258.32 ± 16.77	146.09 ± 5.55$_{\times 10^3}$	263.72 ± 20.73	**56.92 ± 3.98**
		$\begin{pmatrix}0.61\\ 0.59\end{pmatrix}$	$\underline{\mu}, \overline{\mu}$	*56.98 ± 3.91*	**200.61 ± 15.39**	**75.96 ± 4.42**$_{\times 10^3}$	263.74 ± 15.30
			$\underline{\mu}', \overline{\mu}'$	61.08 ± 4.43	249.15 ± 18.51	143.81 ± 5.85$_{\times 10^3}$	*262.54 ± 17.30*
		$\begin{pmatrix}0.81\\ 0.79\end{pmatrix}$	$\underline{\mu}, \overline{\mu}$	26.41 ± 1.63	*60.67 ± 5.08*	**202.98 ± 16.90**	**79.98 ± 3.91**$_{\times 10^3}$
			$\underline{\mu}', \overline{\mu}'$	*25.99 ± 1.86*	66.23 ± 5.93	252.81 ± 19.55	149.41 ± 5.07$_{\times 10^3}$

The numbers are averaged over 100 runs and the intervals determined by '±' with its following numbers are their 99% confidence intervals

$\underline{\mu}'_i(n_i(t)) \geq \theta_L$ before satisfying condition $\underline{\mu}_i(n_i(t)) \geq \theta_L$. This indicates that, in the case with very small Δ, decision condition using conventional bounds is better for arms far from θ.

7.2 Effectiveness of arm selection policy $\mathrm{APT_P}$

7.2.1 Simulation using synthetic distribution parameters

In this experiment, we first generated distribution means $\mu_1, \ldots, \mu_{100}$ of 100 arms, and then ran algorithm BAEC[$\mathrm{APT_P}$, $\underline{\mu}$, $\overline{\mu}$] simulating an arm-i draw by generating a loss according to a Bernoulli distribution with mean μ_i.

For given natural number m and a threshold pair (θ_L, θ_U), m distribution means were generated according to a uniform distribution over $[\theta, 1]$ and $100 - m$ distribution means were generated according to a uniform distribution over $[0, \theta)$, where $\theta = \theta_U - \frac{1}{1+\alpha}\Delta$.

For each set of 100 distribution means, we also ran algorithms BAEC[ASP, $\underline{\mu}$, $\overline{\mu}$] for ASP = UCB, LUCB, TS (Thompson sampling) and MS (Murphy sampling)[7] in addition to for ASP = $\mathrm{APT_P}$ by generating the same i.i.d. loss sequence for the same arm, which can be realized by feeding a same seed to a random number generator for the same arm. Here, arm selection policy LUCB uses

$$\mathrm{LUCB}(t,i) = \begin{cases} \infty & (n_i(t) = 0) \\ \hat{\mu}_i(n_i(t)) & (n_i(t) > 0, t \text{ is odd}) \\ \hat{\mu}_i(n_i(t)) + \sqrt{\frac{1}{2n_i(t)} \ln \frac{5Kt^4}{4\delta}} & (n_i(t) > 0, t \text{ is even}). \end{cases}$$

Note that LUCB[8] (Kalyanakrishnan et al. 2012) is an algorithm for the best k arm identification problem, and the above policy is exactly the same arm-selection policy as original LUCB for $k = 1$.

Both of TS and MS decide the arm to select at each round t based on samples $\tilde{\mu}_i^t$ drawn from $[0, 1]$ according to each arm's posterior loss-mean distribution π_i^t $(i = 1, \ldots, K)$. TS chooses the arm $i \in A_t$ with $\tilde{\mu}_i^t = \max_j \tilde{\mu}_j^t$ without any condition while MS similarly chooses[9] the maximum-sampled-mean arm $i \in A_t$ under the condition[10] that the $\max_j \tilde{\mu}_j^t > \theta$. We used independent uniform distribution over $[0, 1]$ for each arm as the prior loss-mean distribution of TS and MS.

For each $m = 0, 1, 25, 50, 100$, we generated 100 sets[11] of 100 distribution means, and ran the three algorithms for each set and for each combination of parameters $\delta = 0.01, 0.001$ and $(\theta_L, \theta_U) = (0.19, 0.21), (0.49, 0.51), (0.79, 0.81), (0.1, 0.3), (0.4, 0.6), (0.7, 0.9)$. As

[7] Note that BAEC[MS, $\underline{\mu}$, $\overline{\mu}$] is an elimination algorithm though original Murphy sampling does not eliminate arms.

[8] LUCB means that both of LCB (lower confidence bound) and UCB (upper confidence bound) are used in the algorithm. In fact, it chooses the arm i with the smallest LCB among the arms with the largest m sample means when $m \geq 2$.

[9] The original Murphy sampling is an algorithm for checking the existence of negative arms and the procedure of MS here is completely opposite to the original one.

[10] This conditioned sampling is realized by rejecting a condition-unsatisfied set of samples and drawing another one repeatedly until a condition-satisfied set of samples is drawn.

[11] Note that the results shown in Table 4 are the averaged decision times not for a specific set of Bernoulli distributions but for 100 sets of Bernoulli distributions with means generated from certain uniform distributions. So, the decision times obtained in this experiment are not a direct experimental evaluation of the theoretically analyzed decision times.

for threshold pairs (θ_L, θ_U), $\Delta = 0.02$ for the first three and $\Delta = 0.2$ for the last three. Stopping times were averaged over 100 runs.

The result is shown in Table 4. In the case with large $\Delta(= 0.2)$, the averaged stopping time for $\mathrm{APT_P}$ is the smallest for all the combinations of parameters in this experiment. In the case with small $\Delta(= 0.02)$, BAEC[$\mathrm{APT_P}$, $\underline{\mu}$, $\overline{\mu}$] also stopped first, on average, for more than half of the combinations of parameters. For this small Δ, MS, TS and LUCB also performed well to some extent, and in fact, MS and TS stopped first for most of small m ($m = 1, 25$), and LUCB's stopping time was shortest for about a quarter of the parameter combinations. BAEC[$\mathrm{APT_P}$, $\underline{\mu}$, $\overline{\mu}$] stopped first even when $m = 0$, that is, in the case that all the loss means are below θ. In such case, some gray zone arms can be judged as positive and make the algorithm stop. BAEC[$\mathrm{APT_P}$, $\underline{\mu}$, $\overline{\mu}$] is considered to have found such gray zone arms faster.

7.2.2 Simulation based on real dataset

In this experiment, as loss distribution means, we used estimated ad click rates by users in the same category calculated from Real-Time Bidding dataset provided by iPinYou (Zhang et al. 2014). From the training dataset of the second season of iPinYou dataset, we chose 20 most frequently appeared user categories (sets of user profile ids) and calculated the click rate by the users in the category for each of them using the impression and click logs. Since the click rates are smaller than 0.001, we used the values multiplied by 100 as loss means. The loss means $\mu_1, \ldots, \mu_{20}$ used in the experiment are followings:

$$\begin{array}{lllll}
\mu_1 : 0.06232, & \mu_5 : 0.04124, & \mu_9 : 0.03792, & \mu_{13} : 0.02535, & \mu_{17} : 0.02183, \\
\mu_2 : 0.05549, & \mu_6 : 0.04060, & \mu_{10} : 0.03764, & \mu_{14} : 0.02498, & \mu_{18} : 0.02055, \\
\mu_3 : 0.05011, & \mu_7 : 0.04031, & \mu_{11} : 0.03054, & \mu_{15} : 0.02203, & \mu_{19} : 0.01255, \\
\mu_4 : 0.04587, & \mu_8 : 0.03907, & \mu_{12} : 0.02594, & \mu_{16} : 0.02197, & \mu_{20} : 0.01033.
\end{array}$$

In this experiment, 5 thresholds $(\theta_L, \theta_U) = (\theta_{m'} - 0.01, \theta_{m'} + 0.01)$ for $m' = 0, 1, 5, 10, 19$ are used so as to let the loss means of about m' arms be at least θ, where $\theta_0 = \mu_1 + \frac{\mu_1 - \mu_2}{2}$, $\theta_{m'} = \frac{\mu_{m'} + \mu_{m'+1}}{2}$ for $m' = 1, 5, 10, 19$. For these (θ_L, θ_U)s, $\theta = 0.06649, 0.05966, 0.04168, 0.03485, 0.01220$ when $\delta = 0.001$, and $\theta = 0.06659, 0.05976, 0.04178, 0.03495, 0.01230$ when $\delta = 0.01$. For these θs, the number of arms whose loss mean is at least θ is $0, 1, 4, 10, 19$. For each combination of parameters $\delta = 0.01, 0.001$, $(\theta_L, \theta_U) = (\theta_{m'} - 0.01, \theta_{m'} + 0.01)$ ($m' = 0, 1, 5, 10, 19$), we ran algorithm BAEC[ASP, $\underline{\mu}$, $\overline{\mu}$] with three arm selection policies ASP = $\mathrm{APT_P}$, LUCB and UCB 100 times and calculated their stopping times averaged over the 100 runs.

The result is shown in Table 5. For $m = 1$, the stopping times for $\mathrm{APT_P}$ are significantly small compared with those for the other four arm selection policies. Shortest averaged stopping time was achieved by MS and TS for $m = 4, 10$ and by LUCB for $m = 19$ though the differences from $\mathrm{APT_P}$'s stopping times are not significant except for the stopping time of MS and TS in the case with $\delta = 0.001$, $m = 10$. When $m = 0$, the stopping times of the three algorithms are equal, which means that all the arms including the unique neutral arm μ_1 were always judged as negative arms in the experiment.

8 Conclusions

We theoretically and empirically studied sample complexity of a *bad arm existence checking problem* (BAEC problem), whose objective is to judge whether some arms are bad (having loss mean at least θ_U) or all the arms are good (having loss mean less than θ_L) correctly

Table 4 The average stopping times $\times 10^{-3}$ of five algorithms, and their 99% confidence intervals in the simulations using synthetic distribution parameters

		(θ_L, θ_U)	Policy	$m = 0$	$m = 1$	$m = 25$	$m = 50$	$m = 100$
$\Delta = 0.2$	$\delta = 0.01$ $(KT_\Delta = 68.4)$	$(0.1, 0.3)$ $\theta = 0.2085$	$\mathrm{APT_P}$	*10.14 ± 1.74*	*1.28 ± 0.46*	**0.06 ± 0.02**	**0.04 ± 0.01**	**0.03 ± 0.01**
			UCB	14.66 ± 0.69	3.47 ± 0.80	0.27 ± 0.00	0.33 ± 0.00	0.45 ± 0.01
			LUCB	11.00 ± 1.54	1.47 ± 0.48	0.14 ± 0.00	0.14 ± 0.00	0.14 ± 0.00
			TS	12.52 ± 1.09	2.44 ± 0.53	0.13 ± 0.01	0.13 ± 0.01	0.16 ± 0.01
			MS	NA†	2.44 ± 0.53	0.13 ± 0.01	0.14 ± 0.01	0.15 ± 0.01
		$(0.4, 0.6)$ $\theta = 0.5085$	$\mathrm{APT_P}$	*6.16 ± 0.71*	*1.86 ± 0.53*	**0.11 ± 0.03**	**0.06 ± 0.01**	**0.07 ± 0.02**
			UCB	7.28 ± 0.37	3.14 ± 0.57	0.44 ± 0.01	0.63 ± 0.01	1.04 ± 0.02
			LUCB	6.59 ± 0.58	2.25 ± 0.56	0.21 ± 0.01	0.22 ± 0.01	0.21 ± 0.01
			TS	6.86 ± 0.47	2.68 ± 0.53	0.22 ± 0.01	0.25 ± 0.01	0.31 ± 0.02
			MS	NA†	2.68 ± 0.53	0.22 ± 0.01	0.25 ± 0.01	0.34 ± 0.01
		$(0.7, 0.9)$ $\theta = 0.8085$	$\mathrm{APT_P}$	*4.83 ± 0.43*	*1.92 ± 0.49*	**0.22 ± 0.03**	**0.18 ± 0.02**	**0.17 ± 0.02**
			UCB	5.28 ± 0.26	3.14 ± 0.38	1.65 ± 0.03	2.61 ± 0.03	4.49 ± 0.05
			LUCB	5.05 ± 0.35	2.39 ± 0.46	0.51 ± 0.03	0.58 ± 0.05	0.61 ± 0.07
			TS	5.05 ± 0.33	2.39 ± 0.44	0.51 ± 0.02	0.73 ± 0.03	1.08 ± 0.05
			MS	NA†	2.39 ± 0.44	0.51 ± 0.02	0.74 ± 0.03	1.09 ± 0.05
	$\delta = 0.001$ $(KT_\Delta = 80.8)$	$(0.1, 0.3)$ $\theta = 0.2072$	$\mathrm{APT_P}$	*13.30 ± 2.09*	*1.81 ± 0.57*	**0.06 ± 0.02**	**0.05 ± 0.01**	**0.04 ± 0.02**
			UCB	17.59 ± 0.94	4.46 ± 1.05	0.28 ± 0.00	0.36 ± 0.00	0.50 ± 0.01
			LUCB	14.31 ± 1.87	1.92 ± 0.48	0.15 ± 0.00	0.15 ± 0.01	0.16 ± 0.01
			TS	15.31 ± 1.48	3.08 ± 0.69	0.15 ± 0.01	0.15 ± 0.01	0.18 ± 0.01
			MS	NA†	3.07 ± 0.68	0.14 ± 0.01	0.15 ± 0.01	0.19 ± 0.01

Table 4 continued

		(θ_L, θ_U)	Policy	$m = 0$	$m = 1$	$m = 25$	$m = 50$	$m = 100$
		(0.4, 0.6) $\theta = 0.5072$	$\mathrm{APT_P}$	*7.35 ± 0.95*	*2.11 ± 0.66*	**0.11 ± 0.03**	**0.07 ± 0.01**	**0.08 ± 0.02**
			UCB	8.72 ± 0.51	3.25 ± 0.66	0.48 ± 0.01	0.72 ± 0.01	1.13 ± 0.02
			LUCB	7.80 ± 0.82	2.38 ± 0.67	0.22 ± 0.01	0.22 ± 0.01	0.23 ± 0.02
			TS	8.27 ± 0.64	2.76 ± 0.63	0.22 ± 0.01	0.27 ± 0.01	0.36 ± 0.02
			MS	NA†	2.75 ± 0.63	0.22 ± 0.01	0.27 ± 0.01	0.37 ± 0.02
		(0.7, 0.9) $\theta = 0.8072$	$\mathrm{APT_P}$	*5.99 ± 0.49*	*2.42 ± 0.62*	**0.24 ± 0.03**	**0.20 ± 0.02**	**0.22 ± 0.04**
			UCB	6.46 ± 0.30	3.84 ± 0.48	1.79 ± 0.04	2.87 ± 0.04	5.06 ± 0.07
			LUCB	6.17 ± 0.41	3.05 ± 0.57	0.56 ± 0.04	0.60 ± 0.06	0.61 ± 0.06
			TS	6.19 ± 0.38	2.97 ± 0.55	0.55 ± 0.02	0.79 ± 0.03	1.17 ± 0.06
			MS	NA†	2.96 ± 0.55	0.54 ± 0.02	0.77 ± 0.03	1.17 ± 0.06
$\Delta = 0.02$	$\delta = 0.01$ ($KT_\Delta = 9309.9$)	(0.19, 0.21) $\theta = 0.2006$	$\mathrm{APT_P}$	*298.58 ± 27.17*	7.23 ± 3.68	0.47 ± 0.68	*0.12 ± 0.07*	*0.15 ± 0.17*
			UCB	324.62 ± 25.06	8.65 ± 2.47	0.30 ± 0.00	0.40 ± 0.01	0.59 ± 0.01
			LUCB	315.27 ± 25.44	*4.27 ± 1.43*	*0.16 ± 0.01*	0.17 ± 0.01	0.17 ± 0.01
			TS	303.77 ± 25.72	4.93 ± 1.30	*0.16 ± 0.01*	0.18 ± 0.01	0.23 ± 0.01
			MS	NA†	4.93 ± 1.30	*0.16 ± 0.01*	0.18 ± 0.01	0.23 ± 0.01
		(0.49, 0.51) $\theta = 0.5006$	$\mathrm{APT_P}$	*130.83 ± 17.64*	8.19 ± 3.42	0.51 ± 0.51	0.30 ± 0.16	0.41 ± 0.67
			UCB	138.03 ± 17.96	8.05 ± 2.02	0.61 ± 0.02	0.94 ± 0.02	1.50 ± 0.03
			LUCB	140.86 ± 18.53	7.89 ± 2.49	*0.27 ± 0.01*	*0.29 ± 0.02*	*0.31 ± 0.03*
			TS	134.45 ± 17.42	6.21 ± 1.56	*0.27 ± 0.01*	0.35 ± 0.02	0.48 ± 0.02
			MS	NA†	*6.19 ± 1.55*	*0.27 ± 0.01*	0.35 ± 0.02	0.48 ± 0.02

Table 4 continued

	(θ_L, θ_U)	Policy	$m = 0$	$m = 1$	$m = 25$	$m = 50$	$m = 100$
$\delta = 0.001$ ($KT_{\vartriangle} = 10530.7$)	(0.79, 0.81) $\theta = 0.8006$	$\mathrm{APT_P}$	*101.37 ± 17.65*	14.59 ± 5.93	1.05 ± 0.29	*1.06 ± 0.75*	*1.01 ± 0.52*
		UCB	102.37 ± 17.78	16.20 ± 3.74	3.01 ± 0.05	5.28 ± 0.07	9.71 ± 0.14
		LUCB	102.40 ± 17.79	19.21 ± 5.04	0.93 ± 0.04	1.18 ± 0.11	1.35 ± 0.17
		TS	101.93 ± 17.68	11.89 ± 2.95	*0.80 ± 0.04*	1.23 ± 0.06	1.93 ± 0.09
		MS	NA[†]	*11.87 ± 2.94*	*0.80 ± 0.04*	1.21 ± 0.07	1.92 ± 0.09
	(0.19, 0.21) $\theta = 0.2005$	$\mathrm{APT_P}$	*350.82 ± 34.25*	9.65 ± 4.00	0.21 ± 0.14	**0.08 ± 0.05**	*0.12 ± 0.08*
		UCB	379.43 ± 28.55	9.08 ± 2.44	0.31 ± 0.01	0.41 ± 0.01	0.62 ± 0.01
		LUCB	367.20 ± 29.76	5.43 ± 1.70	*0.17 ± 0.01*	0.18 ± 0.01	0.18 ± 0.01
		TS	361.85 ± 30.26	*5.38 ± 1.34*	*0.17 ± 0.01*	0.19 ± 0.01	0.24 ± 0.01
		MS	NA[†]	*5.38 ± 1.34*	*0.17 ± 0.01*	0.19 ± 0.01	0.24 ± 0.01
	(0.49, 0.51) $\theta = 0.5005$	$\mathrm{APT_P}$	*150.82 ± 21.59*	9.10 ± 4.65	*0.26 ± 0.11*	*0.27 ± 0.18*	**0.17 ± 0.05**
		UCB	156.21 ± 21.28	8.44 ± 2.29	0.63 ± 0.02	1.01 ± 0.02	1.66 ± 0.03
		LUCB	157.72 ± 21.52	8.19 ± 2.82	0.30 ± 0.02	0.30 ± 0.02	0.33 ± 0.03
		TS	153.63 ± 21.28	*6.41 ± 1.74*	0.28 ± 0.01	0.37 ± 0.02	0.52 ± 0.03
		MS	NA[†]	*6.41 ± 1.74*	0.29 ± 0.02	0.37 ± 0.02	0.53 ± 0.03
	(0.79, 0.81) $\theta = 0.8005$	$\mathrm{APT_P}$	*113.76 ± 19.46*	15.28 ± 6.67	1.54 ± 0.67	1.54 ± 0.78	*1.08 ± 0.43*
		UCB	117.81 ± 20.14	16.89 ± 3.93	3.19 ± 0.06	5.72 ± 0.08	10.42 ± 0.16
		LUCB	119.57 ± 20.72	20.44 ± 5.40	1.02 ± 0.06	*1.25 ± 0.11*	1.34 ± 0.16
		TS	114.57 ± 19.54	12.37 ± 3.05	*0.86 ± 0.05*	*1.25 ± 0.05*	1.99 ± 0.09
		MS	NA[†]	*12.36 ± 3.04*	*0.86 ± 0.05*	1.26 ± 0.05	1.96 ± 0.10

[†] For MS, we were not able to measure the stopping time due to the large computation time in the case with $m = 0$ in which the rejection probability in a posterior distribution of a loss-mean set approaches to one

Table 5 The average stopping times $\times 10^{-3}$ of the five algorithms and their 99% confidence intervals in the simulations based on real dataset. Note that $(\theta_L, \theta_U) = (\theta_{m'} - 0.01, \theta_{m'} + 0.01)$ for $m' = 0, 1, 5, 10, 19$

δ $(KT_\vartriangle)$	Policy	$\theta_0 = 0.06573$	$\theta_1 = 0.05890$	$\theta_5 = 0.04092$	$\theta_{10} = 0.03409$	$\theta_{19} = 0.01144$
		$(m = 0)$	$(m = 1)$	$(m = 4)$	$(m = 10)$	$(m = 19)$
$\delta = 0.01$ (1776.1)	APT$_\mathrm{P}$	*149.9 ± 2.4*	**62.2 ± 4.3**	24.9 ± 5.9	21.6 ± 3.9	9.5 ± 1.7
	UCB	*149.9 ± 2.4*	149.6 ± 6.0	52.5 ± 2.5	41.2 ± 2.1	21.5 ± 1.1
	LUCB	*149.9 ± 2.4*	123.8 ± 8.8	28.1 ± 2.4	19.3 ± 1.9	*9.0 ± 0.8*
	TS	*149.9 ± 2.4*	72.9 ± 3.8	*21.9 ± 1.4*	*17.1 ± 1.3*	9.9 ± 0.7
	MS	NA†	72.4 ± 3.6	*21.9 ± 1.4*	*17.1 ± 1.3*	9.9 ± 0.7
$\delta = 0.001$ (2021.0)	APT$_\mathrm{P}$	*174.2 ± 2.2*	**66.0 ± 3.4**	29.0 ± 6.1	25.5 ± 4.5	9.9 ± 1.9
	UCB	*174.2 ± 2.2*	157.4 ± 5.6	56.5 ± 2.7	44.5 ± 2.4	22.8 ± 1.2
	LUCB	*174.2 ± 2.2*	131.6 ± 7.7	29.6 ± 2.6	21.4 ± 1.8	*9.2 ± 0.8*
	TS	*174.2 ± 2.2*	77.1 ± 3.1	*22.9 ± 1.4*	*18.3 ± 1.4*	10.6 ± 0.8
	MS	NA†	77.0 ± 3.1	*22.9 ± 1.4*	*18.3 ± 1.4*	10.6 ± 0.8

†For MS, we were not able to measure the stopping time due to the large computation time in the case with $m = 0$ in which the rejection probability in a posterior distribution of a loss-mean set approaches to one with probability at least $1-\delta$ for given thresholds $0 < \theta_L \leq \theta_U < 1$ and a given acceptable error rate $0 < \delta < 1/2$. In the case with $\Delta = \theta_U - \theta_L > 0$, we proposed algorithm BAEC[$\mathrm{APT_P}$, $\underline{\mu}$, $\overline{\mu}$] that utilizes asymmetry of positive and negative arms' roles in this problem; the algorithm with a *decision condition* for each arm i with the current number of draws n using Δ-dependent asymmetric confidence bounds $\underline{\mu}_i(n)$ and $\overline{\mu}_i(n)$, and arm selection policy $\mathrm{APT_P}$ that uses a single threshold θ closer to θ_U instead of the center between θ_L and θ_U. Effectiveness of our decision condition was shown empirically and theoretically. Algorithm BAEC[$\mathrm{APT_P}$, $\underline{\mu}$, $\overline{\mu}$] stopped faster or comparably fast as algorithms BAEC[ASP, $\underline{\mu}$, $\overline{\mu}$] for ASP = LUCB, UCB, TS (Thompson Sampling) and MS (Murphy Sampling) in almost all the our simulations. We also showed an asymptotic upper bound of the expected stopping time for BAEC[$\mathrm{APT_P}$, $\underline{\mu}$, $\overline{\mu}$] which is smaller than that for BAEC[UCB, $\underline{\mu}$, $\overline{\mu}$] in the case that there are multiple positive arms and all the positive arms have the same loss means. Current theoretical support for our arm selection policy $\mathrm{APT_P}$ is very limited, and further theoretical analysis that explains its empirically observed small stopping times is our future work.

Acknowledgements This work was partially supported by JST CREST Grant Numbers JPMJCR1662 and JPMJCR18K3, JSPS KAKENHI Grant Numbers JP18H05413 and JP19H04161.

A Proof of Theorem 1

We use the following lemma to prove our lower bound on the number of samples needed for a $(\theta_L, \theta_U, \delta)$-BAEC algorithm.

Lemma 3 *(Kaufmann et al. 2016) Let $\boldsymbol{\nu}$ and $\boldsymbol{\nu}'$ be two loss distribution sets of K arms such that distributions ν_i and ν_i' are mutually absolutely continuous for $i = 1, \ldots, K$. For any almost-surely finite stopping time T and any event $\mathcal{E}$, the following inequality holds.*

$$\sum_{i=1}^{K} \mathbb{E}_{\boldsymbol{\nu}}[n_i(T)]\mathrm{KL}(\nu_i, \nu_i') \geq d(\mathbb{P}_{\boldsymbol{\nu}}(\mathcal{E}), \mathbb{P}_{\boldsymbol{\nu}'}(\mathcal{E})).$$

Proof of Theorem 1. Consider a set $\boldsymbol{\nu}$ of Bernoulli distributions ν_i with mean μ_i for which some positive arms exist, that is, the case with $\mu_1 \geq \theta_U$. Let k be the number of arms i with $\mu_i \geq \theta_L$ in $\{\nu_i\}$, that means $\mu_1 \geq \cdots \geq \mu_k \geq \theta_L > \mu_{k+1} \geq \cdots \geq \mu_K$. For an arbitrary fixed $\epsilon > 0$, let $\{\nu_i'\}$ be the set of Bernoulli distributions with means μ_i' defined as

$$\mu_i' = \begin{cases} \theta_L - \epsilon & (i \leq k) \\ \mu_i & (i > k) \end{cases}$$

For any $(\theta_L, \theta_U, \delta)$-BAEC algorithm, $\mathcal{E}_{\mathrm{POS}}$ denotes the event that its output is "positive". Since some positive arms exist for the distribution set $\boldsymbol{\nu}$, the probability that the event $\mathcal{E}_{\mathrm{POS}}$ occurs must be at least $1-\delta$ by Definition 1, that is, inequality $\mathbb{P}_{\boldsymbol{\nu}}(\mathcal{E}_{\mathrm{POS}}) \geq 1-\delta$ holds. All the arms are negative in the distribution set $\boldsymbol{\nu}' = \{\nu_i'\}$, likewise by Definition 1, inequality $\mathbb{P}_{\boldsymbol{\nu}'}(\mathcal{E}_{\mathrm{POS}}) < \delta$ holds. Thus,

$$\begin{aligned}
\sum_{i=1}^{K} \mathbb{E}[n_i(T)]KL(\nu_i, \nu_i') &= \sum_{i=1}^{k} \mathbb{E}[n_i(T)]d(\mu_i, \mu_i') \quad (\text{by } d(\mu_i, \mu_i) = 0) \\
&= \sum_{i=1}^{k} \mathbb{E}[n_i(T)]d(\mu_i, \theta_L - \epsilon) \\
&\geq d(\mathbb{P}_{\boldsymbol{\nu}}(\mathcal{E}_{\mathrm{POS}}), \mathbb{P}_{\boldsymbol{\nu}'}(\mathcal{E}_{\mathrm{POS}})) \quad (\text{by Lemma 3})
\end{aligned}$$

$$>d(1-\delta,\delta)$$

holds. From the fact that $\max_{i\in\{1,\ldots,k\}} d(\mu_i,\theta_L-\epsilon)=d(\mu_1,\theta_L-\epsilon)$,

$$\mathbb{E}[T]=\sum_{i=1}^{K}\mathbb{E}[n_i(T)]>\frac{d(1-\delta,\delta)}{d(\mu_1,\theta_L-\epsilon)}=\frac{1-2\delta}{d(\mu_1,\theta_L-\epsilon)}\ln\frac{1-\delta}{\delta}$$

holds, which leads to Ineq. (1) by considering its limit as $\epsilon\to+0$.

Next, consider a set $\boldsymbol{\nu}$ of Bernoulli distributions ν_i with mean μ_i for which all the arms are negative, that is, the case with $\mu_1<\theta_L$. Fix $j\in\{1,\ldots,K\}$ arbitrarily. For arbitrary $\epsilon>0$, let $\boldsymbol{\nu}'$ be a set of Bernoulli distributions ν_i' with mean μ_i' defined as

$$\mu_i'=\begin{cases}\theta_U+\epsilon & (i=j)\\ \mu_i & (i\neq j)\end{cases}$$

For any $(\theta_L,\theta_U,\delta)$-BAEC algorithm, $\mathcal{E}_{\mathrm{NEG}}$ denotes the event that its output is “negative”. Then, inequalities $\mathbb{P}_{\boldsymbol{\nu}}(\mathcal{E}_{\mathrm{NEG}})\geq 1-\delta$ and $\mathbb{P}_{\boldsymbol{\nu}'}(\mathcal{E}_{\mathrm{NEG}})<\delta$ hold by Definition 1 because all the arms are negative in $\boldsymbol{\nu}$ and arm j is positive in $\boldsymbol{\nu}'$. Thus, by Lemma 3,

$$\mathbb{E}[n_j(T)]d(\mu_j,\theta_U+\epsilon)\geq d(\mathbb{P}_{\boldsymbol{\nu}}(\mathcal{E}_{\mathrm{NEG}}),\mathbb{P}_{\boldsymbol{\nu}'}(\mathcal{E}_{\mathrm{NEG}}))>d(1-\delta,\delta)$$

holds, that is, for each $j=1,\ldots,K$,

$$\mathbb{E}[n_j(T)]>\frac{d(1-\delta,\delta)}{d(\mu_j,\theta_U+\epsilon)}=\frac{1-2\delta}{d(\mu_j,\theta_U+\epsilon)}\ln\frac{1-\delta}{\delta}$$

holds. This leads to Ineq. (2) by considering its limit as $\epsilon\to+0$ and the summation over $j=1,\ldots,K$. □

B Proof of Theorem 2

We prove Theorem 2 using the following proposition.

Proposition 2 *For any $x>0$, the following inequality holds:*

$$\sqrt{4+x}\leq\sqrt{1+x}+1\leq\sqrt{4+2x}.$$

Proof Since

$$\sqrt{1+x}+1=\sqrt{(\sqrt{1+x}+1)^2}=\sqrt{2+x+2\sqrt{1+x}}$$

holds,

$$\sqrt{4+x}=\sqrt{2+x+2}\leq\sqrt{1+x}+1$$

and

$$\sqrt{4+2x}=\sqrt{2+x+2\left(1+\frac{x}{2}\right)}\geq\sqrt{1+x}+1$$

hold for $x>0$. □

Proof of Theorem 2. We prove this theorem by contradiction. Assume that $\overline{\mu}_i(T_\Delta) \geq \theta_U$ and $\theta_L > \underline{\mu}_i(T_\Delta)$. Then,

$$\overline{\mu}_i(T_\Delta) - \underline{\mu}_i(T_\Delta) > \theta_U - \theta_L = \Delta \tag{10}$$

holds. On the other hand,

$$\begin{aligned}
\Delta =& \sqrt{\frac{2}{\frac{2}{\Delta^2}\ln\frac{\sqrt{K}N_\Delta}{\delta}}\ln\frac{\sqrt{K}N_\Delta}{\delta}} \\
\geq& \sqrt{\frac{4}{2T_\Delta}\ln\frac{\sqrt{K}N_\Delta}{\delta}} \\
=& \sqrt{\frac{1}{2T_\Delta}\ln\frac{N_\Delta}{\delta}}\sqrt{4 + \frac{2\ln K}{\ln\frac{N_\Delta}{\delta}}} \\
\geq& \sqrt{\frac{1}{2T_\Delta}\ln\frac{N_\Delta}{\delta}}\left(\sqrt{1 + \frac{\ln K}{\ln\frac{N_\Delta}{\delta}}} + 1\right) \quad \text{(by Proposition 2)} \\
=& \sqrt{\frac{1}{2T_\Delta}\ln\frac{KN_\Delta}{\delta}} + \sqrt{\frac{1}{2T_\Delta}\ln\frac{N_\Delta}{\delta}} = \overline{\mu}_i(T_\Delta) - \underline{\mu}_i(T_\Delta)
\end{aligned}$$

holds, which contradicts Ineq. (10). □

C Proof of Theorem 3

If $\overline{\mu}'_i(T'_\Delta) - \underline{\mu}'_i(T'_\Delta) > \Delta$ holds, then $\overline{\mu}'_i(n) - \underline{\mu}'_i(n) > \Delta$ holds for $n = 1, \ldots, T'_\Delta$. In this case, $\underline{\mu}'_i(n) < \theta_L$ and $\overline{\mu}'_i(n) \geq \theta_U$ hold for $n = 1, \ldots, T'_\Delta$ when $\theta_U - (\overline{\mu}'_i(n) - \underline{\mu}'_i(n))/2 \leq \hat{\mu}_i(n) < \theta_L + (\overline{\mu}'_i(n) - \underline{\mu}'_i(n))/2$, which means $\tau'_i > T'_\Delta$. In fact, Inequality $\overline{\mu}'_i(T'_\Delta) - \underline{\mu}'_i(T'_\Delta) > \Delta$ holds because

$$\begin{aligned}
&\overline{\mu}'_i(T'_\Delta) - \underline{\mu}'_i(T'_\Delta) = 2\sqrt{\frac{1}{2T'_\Delta}\ln\frac{2KT'^2_\Delta}{\delta}} \\
=&2\sqrt{\frac{1}{2\lfloor\frac{2}{\Delta^2}\ln\frac{448K}{\Delta^4\delta}\rfloor}\ln\frac{2K\lfloor\frac{2}{\Delta^2}\ln\frac{448K}{\Delta^4\delta}\rfloor^2}{\delta}} \\
\geq&2\sqrt{\frac{1}{2\frac{2}{\Delta^2}\ln\frac{448K}{\Delta^4\delta}}\ln\frac{2K\left(\frac{2}{\Delta^2}\ln\frac{448K}{\Delta^4\delta}\right)^2}{\delta}} \quad \left(\text{because } f(x) = \frac{\ln x}{x} \text{ is decreasing for } x \geq \mathrm{e}\right) \\
=&\Delta\sqrt{\frac{1}{\ln\frac{448K}{\Delta^4\delta}}\ln\left(\frac{8K}{\Delta^4\delta}\left(\ln\frac{448K}{\Delta^4\delta}\right)^2\right)} \\
>&\Delta\sqrt{\frac{1}{\ln\frac{448K}{\Delta^4\delta}}\ln\left(\frac{8K}{\Delta^4\delta}\cdot 56\right)} \quad \left(\text{by } \left(\ln\frac{448K}{\Delta^4\delta}\right)^2 > \left(\ln\frac{448\cdot 2}{1^4(\frac{1}{2})}\right)^2 = 56.11\cdots > 56\right)
\end{aligned}$$

$$=\Delta\sqrt{\frac{1}{\ln\frac{448K}{\Delta^4\delta}}\ln\frac{448K}{\Delta^4\delta}}=\Delta.$$

The difference between the worst case stopping times $\tau_i'-\tau_i$ is lower-bounded as

$$\begin{aligned}
\tau_i'-\tau_i >& T_{\Delta}'-T_{\Delta}=\left\lfloor\frac{2}{\Delta^2}\ln\frac{448K}{\Delta^4\delta}\right\rfloor-\left\lceil\frac{2}{\Delta^2}\ln\frac{\sqrt{K}N_{\Delta}}{\delta}\right\rceil\\
>&\frac{2}{\Delta^2}\ln\frac{448K}{\Delta^4\delta}-\frac{2}{\Delta^2}\ln\frac{\sqrt{K}N_{\Delta}}{\delta}-2\\
=&\frac{2}{\Delta^2}\ln\frac{448\sqrt{K}}{\Delta^4N_{\Delta}\mathrm{e}^{\Delta^2}}\\
>&\frac{2}{\Delta^2}\ln\frac{448\sqrt{K}e^{-\Delta^2}}{\Delta^4\left(\frac{2\mathrm{e}}{(\mathrm{e}-1)\Delta^2}\ln\frac{2\sqrt{K}}{\Delta^2\delta}+1\right)}\\
=&\frac{2}{\Delta^2}\ln\frac{448\sqrt{K}e^{-\Delta^2}\cdot\frac{(\mathrm{e}-1)}{2\mathrm{e}\Delta^2}}{\ln\frac{2\sqrt{K}}{\Delta^2\delta}+\frac{(\mathrm{e}-1)\Delta^2}{2\mathrm{e}}}\\
=&\frac{2}{\Delta^2}\ln\frac{\frac{224\sqrt{K}e^{-\Delta^2-1}(\mathrm{e}-1)}{\Delta^2}}{\ln\frac{2\sqrt{K}}{\Delta^2\delta}\mathrm{e}^{\frac{(\mathrm{e}-1)\Delta^2}{2\mathrm{e}}}}\\
>&\frac{2}{\Delta^2}\ln\frac{224\sqrt{K}e^{-2}(\mathrm{e}-1)/\Delta^2}{\ln\frac{2\sqrt{K}}{\Delta^2\delta}\mathrm{e}^{\frac{\mathrm{e}-1}{2\mathrm{e}}}}\quad(\text{by }\Delta<1)\\
>&\frac{2}{\Delta^2}\ln\frac{52\sqrt{K}/\Delta^2}{\ln\frac{3\sqrt{K}}{\Delta^2\delta}}\quad\left(\text{by }224\mathrm{e}^{-2}(\mathrm{e}-1)>52\text{ and }2\mathrm{e}^{\frac{\mathrm{e}-1}{2\mathrm{e}}}<3\right)\\
=&\frac{2}{\Delta^2}\left(\ln\frac{52\sqrt{K}}{\Delta^2}-\ln\ln\frac{3\sqrt{K}}{\Delta^2\delta}\right).
\end{aligned}$$

D Theorem refered in Remark 4

Define $\overline{\mu}_i''(n)$ as

$$\overline{\mu}_i''(n)=\hat{\mu}_i(n)+\sqrt{\frac{1}{2n}\ln\frac{2n^2}{\delta}}. \tag{11}$$

Then, the following theorem holds.

Theorem 9 *Consider algorithm* BAEC$[*,\underline{\mu}',\overline{\mu}'']$ *and define* $\tau_i''=\min\{n\mid\underline{\mu}_i'(n)\geq\theta_L$ *or* $\overline{\mu}_i''(n)<\theta_U\}$ *for* $i=1,\ldots,K$. *Then, event* $\tau_i''>T_{\Delta}''$ *can happen for* $i=1,\ldots,K$, *where* T_{Δ}'' *is defined as* $T_{\Delta}''=\lfloor\frac{2}{\Delta^2}\ln\frac{366K^{1/4}}{\Delta^4\delta}\rfloor$. *Furthermore, the difference between the worst case stopping times* $\tau_i''-\tau_i$ *is lower-bounded as*

$$\tau_i''-\tau_i>T_{\Delta}''-T_{\Delta}>\frac{2}{\Delta^2}\left(\ln\frac{2}{K^{\frac{1}{4}}\Delta^2}-\ln\ln\frac{3\sqrt{K}}{\Delta^2\delta}\right).$$

Proof If $\overline{\mu}''_i(T''_\Delta) - \underline{\mu}'_i(T''_\Delta) > \Delta$ holds, then $\overline{\mu}''_i(n) - \underline{\mu}'_i(n) > \Delta$ holds for $n = 1, \ldots, T''_\Delta$. In this case, $\underline{\mu}'_i(n) < \theta_L$ and $\overline{\mu}''_i(n) \geq \theta_U$ hold for $n = 1, \ldots, T''_\Delta$ when $\theta_U - (\overline{\mu}''_i(n) - \underline{\mu}'_i(n))/2 \leq \hat{\mu}_i(n) < \theta_L + (\overline{\mu}''_i(n) - \underline{\mu}'_i(n))/2$, which means $\tau''_i > T''_\Delta$. In fact, Inequality $\overline{\mu}''_i(T''_\Delta) - \underline{\mu}'_i(T''_\Delta) > \Delta$ holds because

$$\begin{aligned}
\overline{\mu}''_i(T''_\Delta) - \underline{\mu}'_i(T''_\Delta) &= \sqrt{\frac{1}{2T''_\Delta}\ln\frac{2T''^2_\Delta}{\delta}} + \sqrt{\frac{1}{2T''_\Delta}\ln\frac{2KT''^2_\Delta}{\delta}} \\
=&\sqrt{\frac{1}{2T''_\Delta}\ln\frac{2T''^2_\Delta}{\delta}}\left(1+\sqrt{1+\frac{\ln K}{\ln\frac{2T''^2_\Delta}{\delta}}}\right) \\
\geq&\sqrt{\frac{1}{2T''_\Delta}\ln\frac{2T''^2_\Delta}{\delta}}\sqrt{4+\frac{\ln K}{\ln\frac{2T''^2_\Delta}{\delta}}} \quad \text{(by Proposition 2)} \\
=&\sqrt{\frac{2}{T''_\Delta}\ln\frac{2K^{\frac{1}{4}}T''^2_\Delta}{\delta}} \\
\geq&\sqrt{\frac{\Delta^2}{\ln\frac{366K^{\frac{1}{4}}}{\Delta^4\delta}}\ln\frac{8K^{\frac{1}{4}}\left(\ln\frac{366K^{\frac{1}{4}}}{\Delta^4\delta}\right)^2}{\Delta^4\delta}} \\
>&\sqrt{\frac{\Delta^2}{\ln\frac{366K^{\frac{1}{4}}}{\Delta^4\delta}}\ln\frac{366K^{\frac{1}{4}}}{\Delta^4\delta}} \quad \left(\text{by } 8\left(\ln\frac{366K^{\frac{1}{4}}}{\Delta^4\delta}\right)^2 > 8\left(\ln\frac{366\cdot 2^{\frac{1}{4}}}{1^4\cdot\frac{1}{2}}\right)^2 > 366.56.. > 366\right) \\
=&\Delta
\end{aligned}$$

The difference between the worst case stopping times $\tau''_i - \tau_i$ is lower-bounded as

$$\begin{aligned}
\tau''_i - \tau_i >& T''_\Delta - T_\Delta = \left\lfloor\frac{1}{2\Delta^2}\ln\frac{366K^{\frac{1}{4}}}{\Delta^4\delta}\right\rfloor - \left\lceil\frac{1}{2\Delta^2}\ln\frac{\sqrt{K}N_\Delta}{\delta}\right\rceil \\
>&\frac{1}{2\Delta^2}\ln\frac{366K^{\frac{1}{4}}}{\Delta^4\delta} - \frac{1}{2\Delta^2}\ln\frac{\sqrt{K}N_\Delta}{\delta} - 2 \\
=&\frac{1}{2\Delta^2}\ln\frac{366}{\Delta^4 N_\Delta \mathrm{e}^{4\Delta^2}K^{\frac{1}{4}}} \\
>&\frac{1}{2\Delta^2}\ln\frac{366e^{-4\Delta^2}}{\Delta^4 K^{\frac{1}{4}}\left(\frac{2\mathrm{e}}{(\mathrm{e}-1)\Delta^2}\ln\frac{2\sqrt{K}}{\Delta^2\delta}+1\right)} \\
=&\frac{1}{2\Delta^2}\ln\frac{366e^{-4\Delta^2}\cdot\frac{(\mathrm{e}-1)}{2\mathrm{e}\Delta^2}\cdot K^{-\frac{1}{4}}}{\ln\frac{2\sqrt{K}}{\Delta^2\delta}+\frac{(\mathrm{e}-1)\Delta^2}{2\mathrm{e}}} \\
=&\frac{1}{2\Delta^2}\ln\frac{\frac{183e^{-4\Delta^2-1}(\mathrm{e}-1)}{\Delta^2}}{K^{\frac{1}{4}}\ln\frac{2\sqrt{K}}{\Delta^2\delta}\mathrm{e}^{\frac{(\mathrm{e}-1)\Delta^2}{2\mathrm{e}}}}
\end{aligned}$$

$$
\begin{aligned}
&> \frac{1}{2\Delta^2} \ln \frac{183e^{-5}(\mathrm{e}-1)/\Delta^2}{K^{\frac{1}{4}} \ln \frac{2\sqrt{K}}{\Delta^2\delta} \mathrm{e}^{\frac{\mathrm{e}-1}{2\mathrm{e}}}} \quad (\text{by } \Delta < 1)\\
&> \frac{1}{2\Delta^2} \ln \frac{2/\Delta^2}{K^{\frac{1}{4}} \ln \frac{3\sqrt{K}}{\Delta^2\delta}} \quad \left(\text{by } 183e^{-5}(\mathrm{e}-1) = 2.118.. > 2 \text{ and } 2\mathrm{e}^{\frac{\mathrm{e}-1}{2\mathrm{e}}} = 2.74\cdots < 3\right)\\
&= \frac{1}{2\Delta^2} \left(\ln \frac{2}{K^{\frac{1}{4}}\Delta^2} - \ln\ln \frac{2\sqrt{K}}{\Delta^2\delta} \right).
\end{aligned}
$$

□

E Proof of Proposition 1

The following proposition is needed to prove Proposition 1.

Proposition 3 *For* $0 < a < 1$*, any* $t \geq \frac{\mathrm{e}}{(\mathrm{e}-1)a} \ln \frac{1}{a}$ *satisfies the following inequality.*

$$at \geq \ln t.$$

Proof For $0 < a < 1$, let $f(t) = at - \ln t$. When $a > \frac{1}{\mathrm{e}}$, $f(t)$ is always positive for any $t > 0$ since $f(t)$ takes minimum value $1 - \ln \frac{1}{a}$ at $t = \frac{1}{a}$.

When $a \leq \frac{1}{\mathrm{e}}$, if $t = \frac{\mathrm{e}}{(\mathrm{e}-1)a} \ln \frac{1}{a}$,

$$at - \ln t = \left(\frac{1}{\mathrm{e}-1} \ln \frac{1}{a} - \ln \frac{\mathrm{e}}{\mathrm{e}-1} \right) - \ln\ln \frac{1}{a} \geq 0$$

holds because $y = \frac{1}{\mathrm{e}-1}x - \ln \frac{\mathrm{e}}{\mathrm{e}-1}$ is a tangential line of $y = \ln x$ at $x = \mathrm{e}-1$. If $t > \frac{\mathrm{e}}{(\mathrm{e}-1)a} \ln \frac{1}{a} \left(\geq \frac{\mathrm{e}}{(\mathrm{e}-1)a} > \frac{1}{a} \right)$, $\frac{df(t)}{dt} = a - \frac{1}{t}$ is positive. Therefore, for $t \geq \frac{\mathrm{e}}{(\mathrm{e}-1)a} \ln \frac{1}{a}$, $at - \ln t \geq 0$. □

Proof of Proposition 1. The following inequality is derived from Proposition 3 by setting a to $\frac{\Delta^2\delta}{2\sqrt{K}}$ that means $t = \frac{\sqrt{K}N_\Delta}{\delta} \geq \frac{2\mathrm{e}\sqrt{K}}{(\mathrm{e}-1)\Delta^2\delta} \ln \frac{2\sqrt{K}}{\Delta^2\delta}$,

$$\ln \frac{\sqrt{K}N_\Delta}{\delta} \leq \frac{\Delta^2\delta}{2\sqrt{K}} \cdot \frac{\sqrt{K}N_\Delta}{\delta} = \frac{\Delta^2 N_\Delta}{2}.$$

Thus,

$$N_\Delta \geq \frac{2}{\Delta^2} \ln \frac{\sqrt{K}N_\Delta}{\delta}$$

holds, and so

$$N_\Delta \geq \left\lceil \frac{2}{\Delta^2} \ln \frac{\sqrt{K}N_\Delta}{\delta} \right\rceil = T_\Delta$$

holds. □

F Proof of Theorem 5

Consider the case that $\mu_1 \geq \theta_U$ and event $\mathcal{E}^+$ occurs. In this case, $\bigcap_{n=1}^{T_\Delta}\{\overline{\mu}_i(n) \geq \mu_i\}$ holds for some i with $\mu_i \geq \theta_U$. Assume $T_{\Delta_i} < \tau_i$ for this i. Then, $\overline{\mu}_i(T_{\Delta_i}) \geq \mu_i \geq \theta_U$ and $\underline{\mu}_i(T_{\Delta_i}) < \theta_L$ hold. However,

$$\begin{aligned}
\underline{\mu}_i(T_{\Delta_i}) =&\hat{\mu}_i - \sqrt{\frac{1}{2T_{\Delta_i}} \ln \frac{KN_\Delta}{\delta}} \\
\geq&\mu_i - \sqrt{\frac{1}{2T_{\Delta_i}} \ln \frac{N_\Delta}{\delta}} - \sqrt{\frac{1}{2T_{\Delta_i}} \ln \frac{KN_\Delta}{\delta}} \quad (\text{by } \overline{\mu}_i(T_{\Delta_i}) \geq \mu_i) \\
=&\mu_i - \sqrt{\frac{1}{2T_{\Delta_i}} \ln \frac{N_\Delta}{\delta}} \left(\sqrt{1 + \frac{\ln K}{\ln \frac{N_\Delta}{\delta}}} + 1\right) \\
\geq&\mu_i - \sqrt{\frac{1}{2T_{\Delta_i}} \ln \frac{N_\Delta}{\delta}} \sqrt{4 + \frac{2\ln K}{\ln \frac{N_\Delta}{\delta}}} \quad (\text{by Proposition 2}) \\
=&\mu_i - \sqrt{\frac{4}{2T_{\Delta_i}} \ln \frac{\sqrt{K}N_\Delta}{\delta}} \geq \mu_i - \Delta_i = \theta_L \quad \left(\text{by } T_{\Delta_i} \geq \frac{2}{\Delta_i^2} \ln \frac{\sqrt{K}N_\Delta}{\delta}\right)
\end{aligned}$$

holds, which contradicts the fact that $\underline{\mu}_i(T_{\Delta_i}) < \theta_L$. Thus, $\tau_i \leq T_{\Delta_i}$ holds for at least one positive arm i with probability $\mathbb{P}\{\mathcal{E}^+\}$ which is at least $1 - \delta$ by Lemma 1.

Consider the case that $\mu_1 < \theta_L$ holds and event $\mathcal{E}^-$ occurs. Assume $T_{\Delta_i} < \tau_i$ for $i = 1, \ldots, K$. Then, $\overline{\mu}_i(T_{\Delta_i}) \geq \theta_U$ and $\underline{\mu}_i(T_{\Delta_i}) < \mu_i < \theta_L$ hold. However,

$$\begin{aligned}
\overline{\mu}_i(T_{\Delta_i}) =&\hat{\mu}_i + \sqrt{\frac{1}{2T_{\Delta_i}} \ln \frac{N_\Delta}{\delta}} \\
<&\mu_i + \sqrt{\frac{1}{2T_{\Delta_i}} \ln \frac{N_\Delta}{\delta}} + \sqrt{\frac{1}{2T_{\Delta_i}} \ln \frac{KN_\Delta}{\delta}} \quad (\text{by } \underline{\mu}_i(T_{\Delta_i}) < \mu_i) \\
=&\mu_i + \sqrt{\frac{1}{2T_{\Delta_i}} \ln \frac{N_\Delta}{\delta}} \left(\sqrt{1 + \frac{\ln K}{\ln \frac{N_\Delta}{\delta}}} + 1\right) \\
\leq&\mu_i + \sqrt{\frac{1}{2T_{\Delta_i}} \ln \frac{N_\Delta}{\delta}} \sqrt{4 + \frac{2\ln K}{\ln \frac{N_\Delta}{\delta}}} \quad (\text{by Proposition 2}) \\
=&\mu_i + \sqrt{\frac{4}{2T_{\Delta_i}} \ln \frac{\sqrt{K}N_\Delta}{\delta}} \leq \mu_i + \Delta_i = \theta_U \quad \left(\text{by } T_{\Delta_i} \geq \frac{2}{\Delta_i^2} \ln \frac{\sqrt{K}N_\Delta}{\delta}\right)
\end{aligned}$$

holds, which contradicts the fact that $\overline{\mu}_i(T_{\Delta_i}) \geq \theta_U$. Thus, $\tau_i \leq T_{\Delta_i}$ holds for all arms i with probability $\mathbb{P}\{\mathcal{E}^-\}$ which is at least $1 - \delta$ by Lemma 1.

G Proof of Lemma 2

Let ϵ be an arbitrary real that satisfies $0 < \epsilon < \Delta/2(1+\alpha)$.

Consider the case with $\mu_i \geq \theta$. Define n_i as $n_i = \frac{1}{2(\Delta_i - \epsilon)^2} \ln \frac{KN_\Delta}{\delta}$. Then,

$$
\begin{aligned}
\mathbb{E}[\tau_i \mathbb{1}\{\mathcal{E}\}] =& \sum_{n=1}^{\infty} n\mathbb{P}[\tau_i = n, \mathcal{E}] = \sum_{n=1}^{\infty} \mathbb{P}[\tau_i \geq n, \mathcal{E}] \\
\leq& \sum_{n=2}^{\infty} \mathbb{P}[\underline{\mu}_i(n-1) < \theta_L, \mathcal{E}] + 1 \\
=& \sum_{n=1}^{\infty} \mathbb{P}[\underline{\mu}_i(n) < \theta_L, \mathcal{E}] + 1 \\
\leq& \sum_{n=1}^{\lfloor n_i \rfloor} \mathbb{P}[\mathcal{E}] + \sum_{n=\lfloor n_i \rfloor + 1}^{\infty} \mathbb{P}[\underline{\mu}_i(n) < \theta_L] + 1 \\
& \left(\text{because} \mathbb{P}[\underline{\mu}_i(n) < \theta_L, \mathcal{E}] \leq \min\{\mathbb{P}[\underline{\mu}_i(n) < \theta_L], \mathbb{P}[\mathcal{E}]\}\right) \\
\leq& \mathbb{P}[\mathcal{E}]n_i + \sum_{n=\lfloor n_i \rfloor + 1}^{\infty} \mathbb{P}\left[\hat{\mu}_i(n) - \sqrt{\frac{1}{2n_i} \ln \frac{KN_\Delta}{\delta}} < \theta_L\right] + 1 \\
& \left(\text{because } \sqrt{\frac{1}{2n_i} \ln \frac{KN_\Delta}{\delta}} \geq \sqrt{\frac{1}{2n} \ln \frac{KN_\Delta}{\delta}} \text{ for } n \geq n_i\right) \\
=& \mathbb{P}[\mathcal{E}]n_i + \sum_{n=\lfloor n_i \rfloor + 1}^{\infty} \mathbb{P}[\hat{\mu}_i(n) < \mu_i - \epsilon] + 1 \quad \left(\text{because } \sqrt{\frac{1}{2n_i} \ln \frac{KN_\Delta}{\delta}} = \Delta_i - \epsilon\right) \\
\leq& \mathbb{P}[\mathcal{E}]n_i + \sum_{n=\lfloor n_i \rfloor + 1}^{\infty} \mathrm{e}^{-2n\epsilon^2} + 1 \quad \text{(by Hoeffding's Inequality)} \\
\leq& \mathbb{P}[\mathcal{E}]n_i + \frac{1}{\mathrm{e}^{2\epsilon^2} - 1} + 1 \leq \mathbb{P}[\mathcal{E}]n_i + \frac{1}{2\epsilon^2} + 1 \quad (\text{because } \mathrm{e}^x - 1 \geq x \text{ for any real } x) \\
=& \frac{\mathbb{P}[\mathcal{E}]}{2(\Delta_i - \epsilon)^2} \ln \frac{KN_\Delta}{\delta} + \frac{1}{2\epsilon^2} + 1
\end{aligned}
$$

holds. Since $\frac{1}{\Delta_i^2} + \frac{6\epsilon}{\Delta_i^3} - \frac{1}{(\Delta_i - \epsilon)^2} = \frac{\epsilon(\Delta_i - 2\epsilon)(4\Delta_i - 3\epsilon)}{\Delta_i^3(\Delta_i - \epsilon)^2} \geq 0$ holds for $0 < \epsilon \leq \frac{\Delta_i}{2}$, $\frac{1}{(\Delta_i - \epsilon)^2} \leq \frac{1}{\Delta_i^2} + \frac{6\epsilon}{\Delta_i^3}$ holds for $0 < \epsilon < \Delta/2(1+\alpha) \leq \Delta_i/2$. Thus, Ineq. (8) can be obtained by setting ϵ to $O((\ln \frac{KN_\Delta}{\delta})^{-1/3})$.

Next, consider the case with $\mu_i < \theta$. Define n_i as $n_i = \frac{1}{2(\Delta_i - \epsilon)^2} \ln \frac{N_\Delta}{\delta}$. Then,

$$
\begin{aligned}
\mathbb{E}[\tau_i \mathbb{1}\{\mathcal{E}\}] =& \sum_{n=1}^{\infty} n\mathbb{P}[\tau_i = n, \mathcal{E}] = \sum_{n=1}^{\infty} \mathbb{P}[\tau_i \geq n, \mathcal{E}] \\
\leq& \sum_{n=2}^{\infty} \mathbb{P}[\overline{\mu}_i(n-1) \geq \theta_U, \mathcal{E}] + 1 \\
=& \sum_{n=1}^{\infty} \mathbb{P}[\overline{\mu}_i(n) \geq \theta_U, \mathcal{E}] + 1
\end{aligned}
$$

holds. Similar calculation leads to Inequality (9).

H Proof of Theorem 7

Define $\mathrm{apt}_{\mathrm{P}}(n, i)$ as $\mathrm{apt}_{\mathrm{P}}(n, i) = \sqrt{n}(\hat{\mu}_i(n) - \theta)$ for convenience. Note that $\mathrm{APT}_{\mathrm{P}}(t, i) = \mathrm{apt}_{\mathrm{P}}(n_i(t), i)$. Random variables Y_i and $N_i(a)$ are defined as

$$\begin{aligned} Y_i &= \min_{n \in \{1,\ldots,\tau_i\}} \mathrm{apt}_{\mathrm{P}}(n, i) \text{ and} \\ N_i(a) &= \min\left(\{n | n \in \{1, \ldots, \tau_i - 1\}, \mathrm{apt}_{\mathrm{P}}(n, i) < a\} \cup \{\tau_i\}\right). \end{aligned}$$

To obtain an upper bound of the expected stopping time $\mathbb{E}[T]$ for algorithm BAEC[$\mathrm{APT}_{\mathrm{P}}, \underline{\mu}, \overline{\mu}$], we consider the case that, for some arm i with $\mu_i \geq \theta$, arm i is the first arm that satisfies decision condition and $\underline{\mu}_i(\tau_i) \geq \theta_L$, that is, the case that event $\{\hat{i}_1 = i, \mathcal{E}_i^{\mathrm{POS}}\}$ occurs for $i \leq m$. In the case with no such arm i, stopping time T is upper-bounded by the worst case bound $KT_{\vartriangle}$ (Theorem 4) and the decreasing order of the occurrence probability of this case as $\delta \to +0$ can be proved to be small compared to the increasing order of $KT_{\vartriangle}$ (for the case with $\hat{i}_1 = i \geq m + 1$ by Lemma 13 and for the case that event $\overline{\mathcal{E}_i^{\mathrm{POS}}}$ occurs for $i \leq m$ by Lemma 14), so it can be ignored asymptotically as $\delta \to +0$. An upper bound of $\mathbb{E}[T \mathbb{1}\{\hat{i}_1 = i, \mathcal{E}_i^{\mathrm{POS}}\}]$ for arm i with $\mu_i \geq \theta$ is proved in Lemma 10. When event $\{\hat{i}_1 = i, \mathcal{E}_i^{\mathrm{POS}}\}$ occurs for arm i with $\mu_i \geq \theta$, the number of arm draws is τ_i for arm i, at most $N_j(Y_i)$ for arm $j \neq i$ if $Y_i \leq 0$ and at most $N_j(0)$ for arm $j \neq i$ if $Y_i > 0$. So, to prove the upper bound in Lemma 10, we upper bound $\mathbb{E}[\tau_i \mathbb{1}\{\hat{i}_1 = i, \mathcal{E}_i^{\mathrm{POS}}\}]$ by Lemma 2, $\mathbb{E}[N_j(Y_i)\mathbb{1}\{Y_i \leq 0, \hat{i}_1 = i, \mathcal{E}_i^{\mathrm{POS}}\}]$ for $j \neq i$ by Lemmas 5 and 8 and $\mathbb{E}[N_j(0)\mathbb{1}\{Y_i > 0, \hat{i}_1 = i, \mathcal{E}_i^{\mathrm{POS}}\}]$ for $j \neq i$ by Lemma 9.

Lemma 4 BAEC[$\mathrm{APT}_{\mathrm{P}}, \underline{\mu}, \overline{\mu}$] *satisfies*

$$\sum_{n=1}^{\infty} \mathbb{P}[N_j(a) \geq n] < \frac{2}{\underline{\Delta}_j^4} \textit{ for } j \leq m \textit{ and } a \leq 0.$$

Proof

$$\begin{aligned} \sum_{n=1}^{\infty} \mathbb{P}[N_j(a) \geq n] &= \sum_{n=1}^{\infty} \sum_{t=n}^{\infty} \mathbb{P}[N_j(a) = t] \\ &\leq \sum_{n=1}^{\infty} \sum_{t=n}^{\infty} \mathbb{P}[\mathrm{apt}_{\mathrm{P}}(t, j) < a] \leq \sum_{n=1}^{\infty} \sum_{t=n}^{\infty} \mathbb{P}[\mathrm{apt}_{\mathrm{P}}(t, j) < 0] \\ &= \sum_{n=1}^{\infty} \sum_{t=n}^{\infty} \mathbb{P}[\sqrt{t}(\hat{\mu}_j(t) - \theta) < 0] \\ &= \sum_{n=1}^{\infty} \sum_{t=n}^{\infty} \mathbb{P}[\hat{\mu}_j(t) < \mu_j - \underline{\Delta}_j] \\ &\leq \sum_{n=1}^{\infty} \sum_{t=n}^{\infty} \mathrm{e}^{-2t\underline{\Delta}_j^2} \quad \text{(by Hoeffding's Inequality)} \\ &= \sum_{n=1}^{\infty} \frac{\mathrm{e}^{-2n\underline{\Delta}_j^2}}{1 - \mathrm{e}^{-2\underline{\Delta}_j^2}} = \frac{\mathrm{e}^{-2\underline{\Delta}_j^2}}{(1 - \mathrm{e}^{-2\underline{\Delta}_j^2})^2} = \frac{\mathrm{e}^{2\underline{\Delta}_j^2}}{(\mathrm{e}^{2\underline{\Delta}_j^2} - 1)^2} \end{aligned}$$

$$< \frac{e^2}{4\underline{\Delta}_j^4} < \frac{2}{\underline{\Delta}_j^4} \quad \text{(because } \underline{\Delta}_j < 1\text{)}$$

□

Lemma 5 BAEC[APT$_\mathrm{P}$, $\underline{\mu}$, $\overline{\mu}$] *satisfies*

$$\mathbb{E}[N_j(Y_i)\mathbb{1}\{Y_i \le 0\}] \le \frac{2}{\underline{\Delta}_j^4}\mathbb{P}[Y_i \le 0]$$

for $i = 1, \dots, K$ *and* $j \le m$ $(i \ne j)$.

Proof Define $\mathbb{F}_i(a)$ as $\mathbb{F}_i(a) = \mathbb{P}[Y_i \le a]$. Then,

$$\begin{aligned}
\mathbb{E}[N_j(Y_i)\mathbb{1}\{Y_i \le 0\}] &= \sum_{n=1}^{\infty} n\mathbb{P}[N_j(Y_i) = n, Y_i \le 0] \\
&= \sum_{n=1}^{\infty} \mathbb{P}[N_j(Y_i) \ge n, Y_i \le 0] \\
&= \int_{-\infty}^{0} \sum_{n=1}^{\infty} \mathbb{P}[N_j(Y_i) \ge n \mid Y_i = a]\mathrm{d}\mathbb{F}_i(a) \\
&= \int_{-\infty}^{0} \sum_{n=1}^{\infty} \mathbb{P}[N_j(a) \ge n]\mathrm{d}\mathbb{F}_i(a) \\
&\le \frac{2}{\underline{\Delta}_j^4} \int_{-\infty}^{0} \mathrm{d}\mathbb{F}_i(a) \quad \text{(by Lemma 4)} \\
&= \frac{2}{\underline{\Delta}_j^4}[\mathbb{P}[Y_i \le a]]_{-\infty}^{0} = \frac{2}{\underline{\Delta}_j^4}\mathbb{P}[Y_i \le 0]
\end{aligned}$$

holds.

□

Lemma 6 BAEC[APT$_\mathrm{P}$, $\underline{\mu}$, $\overline{\mu}$] *satisfies*

$$\sum_{n=1}^{\infty} \mathbb{P}[N_j(a) \ge n] < \frac{4a^2}{\underline{\Delta}_j^2} + \frac{4}{\underline{\Delta}_j^2} + 1$$

for $j \ge m + 1$ *and* $a \le 0$.

Proof Define n_0 as $n_0 = \frac{4a^2}{\underline{\Delta}_j^2}$. Note that $\underline{\Delta}_j + \frac{a}{\sqrt{n}} > \frac{\underline{\Delta}_j}{2}$ for $n > n_0$. Then,

$$\begin{aligned}
\sum_{n=1}^{\infty} \mathbb{P}[N_j(a) \ge n] \le \sum_{n=1}^{\infty} \mathbb{P}[\mathrm{apt}_\mathrm{P}(n-1, j) \ge a] &\le \sum_{n=1}^{\infty} \mathbb{P}[\mathrm{apt}_\mathrm{P}(n, j) \ge a] + 1 \\
&= \sum_{n=1}^{\infty} \mathbb{P}[\sqrt{n}(\hat{\mu}_j(n) - \theta) \ge a] + 1 \\
&= \sum_{n=1}^{\infty} \mathbb{P}\left[\hat{\mu}_j(n) \ge \theta + \frac{a}{\sqrt{n}}\right] + 1
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{n=1}^{\infty} \mathbb{P}\left[\hat{\mu}_j(n) \geq \mu_j + \underline{\Delta}_j + \frac{a}{\sqrt{n}}\right] + 1 \\
&\leq \sum_{n=1}^{\lfloor n_0 \rfloor} 1 + \sum_{n=\lfloor n_0 \rfloor + 1}^{\infty} \mathbb{P}\left[\hat{\mu}_j(n) \geq \mu_j + \underline{\Delta}_j + \frac{a}{\sqrt{n}}\right] + 1 \\
&\leq n_0 + \sum_{n=\lfloor n_0 \rfloor + 1}^{\infty} \mathrm{e}^{-2n\left(\frac{\underline{\Delta}_j}{2}\right)^2} + 1
\end{aligned}
$$

$\Big($by Hoeffding's Inequality and the fact that $\underline{\Delta}_j + \frac{a}{\sqrt{n}} > \frac{\underline{\Delta}_j}{2}$ for $n > n_0\Big)$

$$
\begin{aligned}
&\leq \frac{4a^2}{\underline{\Delta}_j^2} + \frac{\mathrm{e}^{-n_0 \frac{\underline{\Delta}_j^2}{2}}}{1 - \mathrm{e}^{-\frac{\underline{\Delta}_j^2}{2}}} + 1 = \frac{4a^2}{\underline{\Delta}_j^2} + \frac{\mathrm{e}^{\frac{\underline{\Delta}_j^2}{2}}}{\mathrm{e}^{\frac{\underline{\Delta}_j^2}{2}} - 1} \mathrm{e}^{-2a^2} + 1 \\
&\leq \frac{4a^2}{\underline{\Delta}_j^2} + \frac{2\mathrm{e}^{\frac{\underline{\Delta}_j^2}{2}}}{\underline{\Delta}_j^2} + 1 \leq \frac{4a^2}{\underline{\Delta}_j^2} + \frac{2\mathrm{e}^{\frac{1}{2}}}{\underline{\Delta}_j^2} + 1 < \frac{4a^2}{\underline{\Delta}_j^2} + \frac{4}{\underline{\Delta}_j^2} + 1
\end{aligned}
$$

□

Lemma 7 BAEC[APT$_\mathrm{P}$, $\underline{\mu}$, $\overline{\mu}$] *satisfies*

$$
\mathbb{P}[Y_i \leq a] \leq \frac{\mathrm{e}^{-2a^2}}{2\underline{\Delta}_i^2} \textit{ for } i \leq m \textit{ and } a \leq 0.
$$

Proof

$$
\begin{aligned}
\mathbb{P}[Y_i \leq a] &\leq \mathbb{P}\left[\bigcup_{n=1}^{\infty} \{\mathrm{apt}_\mathrm{P}(n, i) \leq a\}\right] \\
&\leq \sum_{n=1}^{\infty} \mathbb{P}[\mathrm{apt}_\mathrm{P}(n, i) \leq a] \\
&= \sum_{n=1}^{\infty} \mathbb{P}[\sqrt{n}(\hat{\mu}_i(n) - \theta) \leq a] \\
&= \sum_{n=1}^{\infty} \mathbb{P}\left[\hat{\mu}_i(n) \leq \theta + \frac{a}{\sqrt{n}}\right] \\
&= \sum_{n=1}^{\infty} \mathbb{P}\left[\hat{\mu}_i(n) \leq \mu_i - \underline{\Delta}_i + \frac{a}{\sqrt{n}}\right] \\
&\leq \sum_{n=1}^{\infty} \mathrm{e}^{-2n\left(\underline{\Delta}_i - \frac{a}{\sqrt{n}}\right)^2} \\
&\leq \mathrm{e}^{-2a^2} \sum_{n=1}^{\infty} \mathrm{e}^{-2n\underline{\Delta}_i^2} = \mathrm{e}^{-2a^2} \frac{1}{\mathrm{e}^{2\underline{\Delta}_i^2} - 1} \leq \frac{\mathrm{e}^{-2a^2}}{2\underline{\Delta}_i^2}.
\end{aligned}
$$

□

Lemma 8 *For $i \le m$ and $j \ge m+1$,* BAEC[APT$_\mathrm{P}$, $\underline{\mu}$, $\overline{\mu}$] *satisfies*

$$\mathbb{E}[N_j(Y_i)\mathbb{1}\{Y_i \le 0\}] \le \frac{1}{\underline{\Delta}_i^2\underline{\Delta}_j^2} + \left(\frac{4}{\underline{\Delta}_j^2}+1\right)\mathbb{P}[Y_i \le 0].$$

Proof Define $\mathbb{F}_i(a)$ as $\mathbb{F}_i(a) = \mathbb{P}[Y_i \le a]$. Then,

$$\begin{aligned}
\mathbb{E}[N_j(Y_i)\mathbb{1}\{Y_i \le 0\}] &= \sum_{n=1}^{\infty}\mathbb{P}[N_j(Y_i) \ge n, Y_i \le 0]\\
&= \int_{-\infty}^{0}\sum_{n=1}^{\infty}\mathbb{P}[N_j(Y_i) \ge n \mid Y_i = a]\mathrm{d}\mathbb{F}_i(a)\\
&= \int_{-\infty}^{0}\sum_{n=1}^{\infty}\mathbb{P}[N_j(a) \ge n]\mathrm{d}\mathbb{F}_i(a)\\
&\le \int_{-\infty}^{0}\left(\frac{4a^2}{\underline{\Delta}_j^2}+\frac{4}{\underline{\Delta}_j^2}+1\right)\mathrm{d}\mathbb{F}_i(a) \quad \text{(by Lemma 6)}\\
&= \frac{4}{\underline{\Delta}_j^2}\int_{-\infty}^{0}a^2\mathrm{d}\mathbb{F}_i(a) + \left(\frac{4}{\underline{\Delta}_j^2}+1\right)\int_{-\infty}^{0}\mathrm{d}\mathbb{F}_i(a)\\
&= \frac{4}{\underline{\Delta}_j^2}\left(\left[a^2\mathbb{P}[Y_i \le a]\right]_{-\infty}^{0} - \int_{-\infty}^{0}2a\mathbb{P}[Y_i \le a]\mathrm{d}a\right) + \left(\frac{4}{\underline{\Delta}_j^2}+1\right)[\mathbb{P}[Y_i \le a]]_{-\infty}^{0}\\
&\quad \text{(using integration by parts)}\\
&= -\frac{4}{\underline{\Delta}_j^2}\int_{-\infty}^{0}2a\mathbb{P}[Y_i \le a]\mathrm{d}a + \left(\frac{4}{\underline{\Delta}_j^2}+1\right)\mathbb{P}[Y_i \le 0]\\
&\le -\frac{2}{\underline{\Delta}_i^2\underline{\Delta}_j^2}\int_{-\infty}^{0}2a\mathrm{e}^{-2a^2}\mathrm{d}a + \left(\frac{4}{\underline{\Delta}_j^2}+1\right)\mathbb{P}[Y_i \le 0] \quad \text{(by Lemma 7)}\\
&= \frac{2}{\underline{\Delta}_i^2\underline{\Delta}_j^2}\left[\frac{e^{-2a^2}}{2}\right]_{-\infty}^{0} + \left(\frac{4}{\underline{\Delta}_j^2}+1\right)\mathbb{P}[Y_i \le 0]\\
&= \frac{1}{\underline{\Delta}_i^2\underline{\Delta}_j^2} + \left(\frac{4}{\underline{\Delta}_j^2}+1\right)\mathbb{P}[Y_i \le 0].
\end{aligned}$$

□

Lemma 9 *For $i \le m$,* BAEC[APT$_\mathrm{P}$, $\underline{\mu}$, $\overline{\mu}$] *satisfies*

$$\mathbb{E}[N_j(0)\mathbb{1}\{Y_i > 0\}] \le \begin{cases} \frac{2}{\underline{\Delta}_j^4}\mathbb{P}[Y_i > 0] & (j \le m)\\ \left(\frac{4}{\underline{\Delta}_j^2}+1\right)\mathbb{P}[Y_i > 0] & (j \ge m+1). \end{cases}$$

Proof

$$\mathbb{E}[N_j(0)\mathbb{1}\{Y_i > 0\}] = \sum_{n=1}^{\infty}\mathbb{P}[N_j(0) \ge n, Y_i > 0]$$

$$
\begin{aligned}
&= \sum_{n=1}^{\infty} \mathbb{P}[N_j(0) \geq n]\mathbb{P}[Y_i > 0] \\
&\qquad \text{(because } N_j(0) \text{ and } Y_i \text{ are independent)} \\
&\leq \begin{cases} \frac{2}{\Delta_j^4}\mathbb{P}[Y_i > 0] & (j \leq m) \ \text{(by Lemma 4)} \\ \left(\frac{4}{\Delta_j^2} + 1\right)\mathbb{P}[Y_i > 0] & (j \geq m+1) \ \text{(by Lemma 6)} \end{cases}
\end{aligned}
$$

□

Lemma 10 *For $i \leq m$ and any event $\mathcal{E}$,* BAEC[APT$_\mathrm{P}$, $\underline{\mu}$, $\overline{\mu}$] *satisfies*

$$
\mathbb{E}[T\mathbb{1}\{\hat{i}_1 = i, \mathcal{E}\}] \leq \frac{\mathbb{P}[\hat{i}_1 = i, \mathcal{E}]}{2\Delta_i^2} \ln \frac{KN_\Delta}{\delta} + O\left(\left(\ln \frac{KN_\Delta}{\delta}\right)^{\frac{2}{3}}\right)
$$

$$
+ \sum_{j \leq m, j \neq i} \frac{2}{\Delta_j^4} + \sum_{j=m+1}^{K} \left\{ \frac{1}{\Delta_i^2 \Delta_j^2} + \left(\frac{4}{\Delta_j^2} + 1\right)\right\}.
$$

Proof In the case that the decision condition is satisfied first by one of arms i with $\mu_i \geq \theta$ ($i \leq m$), that is, $\hat{i}_1 = i$, the stopping time T is at most $\tau_i + \sum_{j \neq i} N_j(Y_i)$ if $Y_i \leq 0$ and at most $\tau_i + \sum_{j \neq i} N_j(0)$ if $Y_i > 0$. Thus, for $i \leq m$,

$$
\begin{aligned}
&\mathbb{E}[T\mathbb{1}\{\hat{i}_1 = i, \mathcal{E}\}] \\
\leq &\mathbb{E}\left[\left(\tau_i + \sum_{j \neq i} N_j(Y_i)\right)\mathbb{1}\{Y_i \leq 0, \hat{i}_1 = i, \mathcal{E}\}\right] + \mathbb{E}\left[\left(\tau_i + \sum_{j \neq i} N_j(0)\right)\mathbb{1}\{Y_i > 0, \hat{i}_1 = i, \mathcal{E}\}\right] \\
= &\mathbb{E}[\tau_i \mathbb{1}\{\hat{i}_1 = i, \mathcal{E}\}] + \sum_{j \neq i} \mathbb{E}[N_j(Y_i)\mathbb{1}\{Y_i \leq 0, \hat{i}_1 = i, \mathcal{E}\}] + \sum_{j \neq i} \mathbb{E}[N_j(0)\mathbb{1}\{Y_i > 0, \hat{i}_1 = i, \mathcal{E}\}] \\
\leq &\mathbb{E}[\tau_i \mathbb{1}\{\hat{i}_1 = i, \mathcal{E}\}] + \sum_{j \neq i} \mathbb{E}[N_j(Y_i)\mathbb{1}\{Y_i \leq 0\}] + \sum_{j \neq i} \mathbb{E}[N_j(0)\mathbb{1}\{Y_i > 0\}] \\
\leq &\frac{\mathbb{P}[\hat{i}_1 = i, \mathcal{E}]}{2\Delta_i^2} \ln \frac{KN_\Delta}{\delta} + O\left(\left(\ln \frac{KN_\Delta}{\delta}\right)^{\frac{2}{3}}\right) + \sum_{j \leq m, j \neq i} \frac{2}{\Delta_j^4}\mathbb{P}[Y_i \leq 0] \quad \text{(by Lemma 2 \& 5)} \\
&+ \sum_{j=m+1}^{K} \left\{ \frac{1}{\Delta_i^2 \Delta_j^2} + \left(\frac{4}{\Delta_j^2} + 1\right)\mathbb{P}[Y_i \leq 0]\right\} \quad \text{(by Lemma 8)} \\
&+ \sum_{j \leq m, j \neq i} \frac{2}{\Delta_j^4}\mathbb{P}[Y_i > 0] + \sum_{j=m+1}^{K} \left(\frac{4}{\Delta_j^2} + 1\right)\mathbb{P}[Y_i > 0] \quad \text{(by Lemma 9)} \\
\leq &\frac{\mathbb{P}[\hat{i}_1 = i, \mathcal{E}]}{2\Delta_i^2} \ln \frac{KN_\Delta}{\delta} + O\left(\left(\ln \frac{KN_\Delta}{\delta}\right)^{\frac{2}{3}}\right) \\
&+ \sum_{j \leq m, j \neq i} \frac{2}{\Delta_j^4} + \sum_{j=m+1}^{K} \left\{ \frac{1}{\Delta_i^2 \Delta_j^2} + \left(\frac{4}{\Delta_j^2} + 1\right)\right\}
\end{aligned}
$$

holds.

□

Define $n_{\Delta,\delta}$ as $n_{\Delta,\delta} = \left\lceil \frac{1}{2(\max\{\theta_U, 1-\theta_L\})^2} \ln \frac{N_\Delta}{\delta} \right\rceil$. Then, τ_i for any arm $i = 1, \ldots, K$ is bounded by $n_{\Delta,\delta}$ from below.

Lemma 11 *In algorithm* BAEC[$*, \underline{\mu}, \overline{\mu}$], $\tau_i \geq n_{\Delta,\delta}$ *holds for any arm* $i = 1, \ldots, K$.

Proof By the definition of τ_i, $\overline{\mu}_i(\tau_i) < \theta_U$ or $\underline{\mu}_i(\tau_i) \geq \theta_L$ must be satisfied for any arm i. In the case with $\overline{\mu}_i(\tau_i) < \theta_U$,

$$\hat{\mu}_i(\tau_i) + \sqrt{\frac{1}{2\tau_i} \ln \frac{N_\Delta}{\delta}} < \theta_U$$

holds. Since $\hat{\mu}_i(\tau_i) \geq 0$,

$$\sqrt{\frac{1}{2\tau_i} \ln \frac{N_\Delta}{\delta}} < \theta_U$$

holds. So, we obtain

$$\tau_i > \frac{1}{2\theta_U^2} \ln \frac{N_\Delta}{\delta}.$$

In the case with $\underline{\mu}_i(\tau_i) \geq \theta_L$,

$$\hat{\mu}_i(\tau_i) - \sqrt{\frac{1}{2\tau_i} \ln \frac{KN_\Delta}{\delta}} \geq \theta_L$$

holds. Since $\hat{\mu}_i(\tau_i) \leq 1$,

$$1 - \sqrt{\frac{1}{2\tau_i} \ln \frac{KN_\Delta}{\delta}} \geq \theta_L$$

holds. So, we obtain

$$\tau_i \geq \frac{1}{2(1-\theta_L)^2} \ln \frac{KN_\Delta}{\delta}.$$

Therefore,

$$\tau_i \geq \min\left\{\frac{1}{2\theta_U^2} \ln \frac{N_\Delta}{\delta}, \frac{1}{2(1-\theta_L)^2} \ln \frac{KN_\Delta}{\delta}\right\} \geq \frac{1}{2(\max\{\theta_U, 1-\theta_L\})^2} \ln \frac{N_\Delta}{\delta}$$

holds. Since τ_i is a natural number,

$$\tau_i \geq \left\lceil \frac{1}{2(\max\{\theta_U, 1-\theta\})^2} \ln \frac{N_\Delta}{\delta} \right\rceil$$

holds. □

Lemma 12 BAEC[APT$_\mathrm{P}$, $\underline{\mu}, \overline{\mu}$] *satisfies*

$$\mathbb{P}\left[Y_i \geq -\frac{\Delta_i}{2}\sqrt{n_{\Delta,\delta}}\right] \leq \mathrm{e}^{-n_{\Delta,\delta}\frac{\Delta_i^2}{2}} \leq \left(\frac{\delta}{N_\Delta}\right)^{\frac{1}{4}\left(\frac{\Delta_i}{\max\{\theta_U, 1-\theta_L\}}\right)^2}$$

for $i \geq m+1$.

Proof

$$\mathbb{P}\left[Y_i \geq -\frac{\Delta_i}{2}\sqrt{n_{\Delta,\delta}}\right] = \mathbb{P}\left[\bigcap_{n=1}^{\tau_i}\left\{\mathrm{apt}_\mathrm{P}(n,i) \geq -\frac{\Delta_i}{2}\sqrt{n_{\Delta,\delta}}\right\}\right]$$

$$
\begin{aligned}
&\le \mathbb{P}\left[\mathrm{apt}_{\mathrm{P}}(n_{\Delta,\delta}, i) \ge -\frac{\Delta_i}{2}\sqrt{n_{\Delta,\delta}}\right] \quad \text{(by Lemma 11)}\\
&= \mathbb{P}\left[\sqrt{n_{\Delta,\delta}}(\hat{\mu}_i(n_{\Delta,\delta}) - \theta) \ge -\frac{\Delta_i}{2}\sqrt{n_{\Delta,\delta}}\right]\\
&= \mathbb{P}\left[\hat{\mu}_i(n_{\Delta,\delta}) \ge \mu_i + \frac{\Delta_i}{2}\right]\\
&\le \mathrm{e}^{-2n_{\Delta,\delta}\left(\frac{\Delta_i}{2}\right)^2} = \mathrm{e}^{-n_{\Delta,\delta}\frac{\Delta_i^2}{2}}\\
&\le \mathrm{e}^{-\frac{1}{4}\left(\frac{\Delta_i}{\max\{\theta_U, 1-\theta_L\}}\right)^2 \ln\frac{N_\Delta}{\delta}} = \left(\frac{\delta}{N_\Delta}\right)^{\frac{1}{4}\left(\frac{\Delta_i}{\max\{\theta_U, 1-\theta_L\}}\right)^2}.
\end{aligned}
$$

□

Lemma 13 *For* $m \ge 1$ *and* $i \ge m+1$, $\mathrm{BAEC}[\mathrm{APT}_{\mathrm{P}}, \underline{\mu}, \overline{\mu}]$ *satisfies*

$$
\mathbb{P}[\hat{i}_1 = i] \le \left(1 + \frac{1}{2\underline{\Delta}_1^2}\right)\left(\frac{\delta}{N_\Delta}\right)^{\frac{1}{4}\left(\frac{\Delta_i}{\max\{\theta_U, 1-\theta_L\}}\right)^2}.
$$

Proof Define $\mathbb{F}_i(a)$ as $\mathbb{F}_i(a) = \mathbb{P}[Y_i \ge a]$. Then,

$$
\begin{aligned}
\mathbb{P}[\hat{i}_1 = i] &= \mathbb{P}\left[\hat{i}_1 = i, Y_i \ge -\frac{\Delta_i}{2}\sqrt{n_{\Delta,\delta}}\right] + \mathbb{P}\left[\hat{i}_1 = i, Y_i < -\frac{\Delta_i}{2}\sqrt{n_{\Delta,\delta}}\right]\\
&\le \mathbb{P}\left[Y_i \ge -\frac{\Delta_i}{2}\sqrt{n_{\Delta,\delta}}\right] + \mathbb{P}\left[Y_1 \le Y_i, Y_i < -\frac{\Delta_i}{2}\sqrt{n_{\Delta,\delta}}\right]. \qquad (12)
\end{aligned}
$$

The second term is bounded as

$$
\begin{aligned}
&\mathbb{P}\left[Y_1 \le Y_i, Y_i < -\frac{\Delta_i}{2}\sqrt{n_{\Delta,\delta}}\right]\\
&= \int_{-\frac{\Delta_i}{2}\sqrt{n_{\Delta,\delta}}}^{-\infty} \mathbb{P}[Y_1 \le Y_i \mid Y_i = a]\mathrm{d}\mathbb{F}_i(a)\\
&\le \int_{-\frac{\Delta_i}{2}\sqrt{n_{\Delta,\delta}}}^{-\infty} \mathbb{P}[Y_1 \le a]\mathrm{d}\mathbb{F}_i(a)\\
&\le \int_{-\frac{\Delta_i}{2}\sqrt{n_{\Delta,\delta}}}^{-\infty} \frac{e^{-2a^2}}{2\underline{\Delta}_1^2}\mathrm{d}\mathbb{F}_i(a) \quad \text{(by Lemma 7)}\\
&= \frac{1}{2\underline{\Delta}_1^2}\left(\left[e^{-2a^2}\mathbb{P}[Y_i \ge a]\right]_{-\frac{\Delta_i}{2}\sqrt{n_{\Delta,\delta}}}^{-\infty} + \int_{-\frac{\Delta_i}{2}\sqrt{n_{\Delta,\delta}}}^{-\infty} 4ae^{-2a^2}\mathbb{P}[Y_i \ge a]\mathrm{d}a\right)\\
&\quad \text{(using integration by parts)}\\
&\le \frac{1}{2\underline{\Delta}_1^2}\left(-\mathrm{e}^{-n_{\Delta,\delta}\frac{\Delta_i^2}{2}}\mathbb{P}\left[Y_i \ge -\frac{\Delta_i}{2}\sqrt{n_{\Delta,\delta}}\right] + \int_{-\frac{\Delta_i}{2}\sqrt{n_{\Delta,\delta}}}^{-\infty} 4ae^{-2a^2}\mathrm{d}a\right)\\
&\le \frac{1}{2\underline{\Delta}_1^2}\int_{-\frac{\Delta_i}{2}\sqrt{n_{\Delta,\delta}}}^{-\infty} 4ae^{-2a^2}\mathrm{d}a\\
&= -\frac{1}{2\underline{\Delta}_1^2}\left[\mathrm{e}^{-2a^2}\right]_{-\frac{\Delta_i}{2}\sqrt{n_{\Delta,\delta}}}^{-\infty} = \frac{1}{2\underline{\Delta}_1^2}\mathrm{e}^{-n_{\Delta,\delta}\frac{\Delta_i^2}{2}} \le \frac{1}{2\underline{\Delta}_1^2}\left(\frac{\delta}{N_\Delta}\right)^{\frac{1}{4}\left(\frac{\Delta_i}{\max\{\theta_U, 1-\theta_L\}}\right)^2}. \qquad (13)
\end{aligned}
$$

Thus, by Ineqs. (12), (13) and Lemma 12,

$$\mathbb{P}[\hat{i}_1 = i] \le \left(1 + \frac{1}{2\underline{\Delta}_1^2}\right)\left(\frac{\delta}{N_\Delta}\right)^{\frac{1}{4}\left(\frac{\Delta_i}{\max\{\theta_U, 1-\theta_L\}}\right)^2}$$

holds. □

Lemma 14 *For the complementary events $\overline{\mathcal{E}_i^{\mathrm{POS}}}$ of event $\mathcal{E}_i^{\mathrm{POS}}$, inequality*

$$\mathbb{P}\left[\overline{\mathcal{E}_i^{\mathrm{POS}}}\right] \le \frac{\mathrm{e}^{2\underline{\Delta}_i^2}}{2\underline{\Delta}_i^2}\left(\frac{\delta}{N_\Delta}\right)^{\left(\frac{\Delta_i}{\max\{\theta_U, 1-\theta_L\}}\right)^2}$$

holds when $i \le m$.

Proof In the case with $\hat{\mu}_i(\tau_i) \ge \theta$, arm i is judged as positive because $\underline{\mu}_i(\tau_i) \ge \theta_L$ holds whenever $\overline{\mu}_i(\tau_i) < \theta_U$ holds.[12] This is because $\theta_U - \theta : \theta - \theta_L = \overline{\mu}_i(\tau_i) - \hat{\mu}_i(\tau_i) : \hat{\mu}_i(\tau_i) - \underline{\mu}_i(\tau_i) = 1 : \alpha$ holds. Thus,

$$\begin{aligned}
\mathbb{P}\left[\overline{\mathcal{E}_i^{\mathrm{POS}}}\right] \le & \mathbb{P}\left[\bigcup_{n=n_{\Delta,\delta}}^{T_\Delta}\{\hat{\mu}_i(n) < \theta\}\right] \\
= & \mathbb{P}\left[\bigcup_{n=n_{\Delta,\delta}}^{T_\Delta}\{\hat{\mu}_i(n) < \mu_i - \underline{\Delta}_i\}\right] \\
\le & \sum_{n=n_{\Delta,\delta}}^{T_\Delta}\mathbb{P}[\hat{\mu}_i(n) < \mu_i - \underline{\Delta}_i] \\
\le & \sum_{n=n_{\Delta,\delta}}^{\infty}\mathrm{e}^{-2n\underline{\Delta}_i^2} = \frac{\mathrm{e}^{2\underline{\Delta}_i^2}\mathrm{e}^{-2n_{\Delta,\delta}\underline{\Delta}_i^2}}{\mathrm{e}^{2\underline{\Delta}_i^2} - 1} \le \frac{\mathrm{e}^{2\underline{\Delta}_i^2}}{2\underline{\Delta}_i^2}\left(\frac{\delta}{N_\Delta}\right)^{\left(\frac{\Delta_i}{\max\{\theta_U, 1-\theta_L\}}\right)^2}
\end{aligned}$$

holds. □

Proof of Theorem 7

$$\begin{aligned}
\mathbb{E}[T] = & \sum_{i=1}^{m}\mathbb{E}\left[T\mathbb{1}\left\{\hat{i}_1 = i, \mathcal{E}_i^{\mathrm{POS}}\right\}\right] + \sum_{i=1}^{m}\mathbb{E}\left[T\mathbb{1}\left\{\hat{i}_1 = i, \overline{\mathcal{E}_i^{\mathrm{POS}}}\right\}\right] + \sum_{i=m+1}^{K}\mathbb{E}[T\mathbb{1}\{\hat{i}_1 = i\}] \\
\le & \sum_{i=1}^{m}\left(\frac{\mathbb{P}\left[\hat{i}_1 = i, \mathcal{E}_i^{\mathrm{POS}}\right]}{2\Delta_i^2}\ln\frac{KN_\Delta}{\delta} + O\left(\left(\ln\frac{KN_\Delta}{\delta}\right)^{\frac{2}{3}}\right) + \sum_{j \le m, j \ne i}\frac{2}{\underline{\Delta}_j^4}\right. \\
& \left. + \sum_{j=m+1}^{K}\left\{\frac{1}{\underline{\Delta}_i^2\underline{\Delta}_j^2} + \left(\frac{4}{\underline{\Delta}_j^2} + 1\right)\right\}\right) \\
& \text{(by Lemma 10)} \\
& + KT_\Delta\sum_{i=1}^{m}\mathbb{P}\left[\overline{\mathcal{E}_i^{\mathrm{POS}}}\right] + KT_\Delta\sum_{i=m+1}^{K}\mathbb{P}[\hat{i}_1 = i]
\end{aligned}$$

[12] An arm is judged as positive when both the positive and negative decision conditions are satisfied simultaneously.

$$
\begin{aligned}
&\leq \sum_{i=1}^{m}\Bigg(\frac{\mathbb{P}\left[\hat{i}_1 = i, \mathcal{E}_i^{\mathrm{POS}}\right]}{2\Delta_i^2}\ln\frac{KN_\Delta}{\delta} + O\left(\left(\ln\frac{KN_\Delta}{\delta}\right)^{\frac{2}{3}}\right) + \frac{2(m-1)}{\underline{\Delta}_i^4} \\
&\qquad\qquad + \left(\frac{1}{\underline{\Delta}_i^2}+4\right)\sum_{j=m+1}^{K}\frac{1}{\underline{\Delta}_j^2} + (K-m)\Bigg) \\
&\quad + KT_\Delta\Bigg(\frac{\mathrm{e}^{2\underline{\Delta}_i^2}}{2\underline{\Delta}_i^2}\sum_{i=1}^{m}\left(\frac{\delta}{N_\Delta}\right)^{\left(\frac{\Delta_i}{\max\{\theta_U,1-\theta_L\}}\right)^2} \qquad \text{(by Lemma 14)} \\
&\quad + \left(1+\frac{1}{2\underline{\Delta}_1^2}\right)\sum_{i=m+1}^{K}\left(\frac{\delta}{N_\Delta}\right)^{\frac{1}{4}\left(\frac{\Delta_i}{\max\{\theta_U,1-\theta_L\}}\right)^2}\Bigg) \qquad \text{(by Lemma 13)} \\
&= \sum_{i=1}^{m}\Bigg(\frac{\mathbb{P}\left[\hat{i}_1 = i, \mathcal{E}_i^{\mathrm{POS}}\right]}{2\Delta_i^2}\ln\frac{KN_\Delta}{\delta} + \frac{2(m-1)}{\underline{\Delta}_i^4} + \left(\frac{1}{\underline{\Delta}_i^2}+4\right)\sum_{j=m+1}^{K}\frac{1}{\underline{\Delta}_j^2}\Bigg) \\
&\quad + m(K-m) + O\left(m\left(\ln\frac{KN_\Delta}{\delta}\right)^{\frac{2}{3}}\right) \\
&\quad + KT_\Delta\Bigg(\frac{\mathrm{e}^{2\underline{\Delta}_i^2}}{2\underline{\Delta}_i^2}\sum_{i=1}^{m}\left(\frac{\delta}{N_\Delta}\right)^{\left(\frac{\Delta_i}{\max\{\theta_U,1-\theta_L\}}\right)^2} \\
&\quad + \left(1+\frac{1}{2\underline{\Delta}_1^2}\right)\sum_{i=m+1}^{K}\left(\frac{\delta}{N_\Delta}\right)^{\frac{1}{4}\left(\frac{\Delta_i}{\max\{\theta_U,1-\theta_L\}}\right)^2}\Bigg).
\end{aligned}
$$

□

I Proof of Theorem 8

We consider event $\bigcup_{i:\mu_i=\mu_1}\mathcal{E}_i^{\mathrm{POS}}$, that is, the event that one of the best arm i is judged as positive. In the case that event $\bigcup_{i:\mu_i=\mu_1}\mathcal{E}_i^{\mathrm{POS}}$ does not occur, stopping time T is upper-bounded by the worst case bound KT_Δ (Theorem 4) and the decreasing order of the occurrence probability of this case as $\delta \to +0$ can be proved to be small compared to the increasing order of KT_Δ (Lemma 14), so it can be ignored asymptotically as $\delta \to +0$. When event $\bigcup_{i:\mu_i=\mu_1}\mathcal{E}_i^{\mathrm{POS}}$ occurs, non-optimal arms i with $\mu_i < \mu_1$ is drawn in the case of μ_i's overestimation ($\mathrm{UCB}(t,i) \geq \mu_1 - \epsilon$) or in the case of μ_1's underestimation ($\mathrm{UCB}(t,1) < \mu_1 - \epsilon$). So, $\mathbb{E}[T\mathbb{1}\{\bigcup_{i:\mu_i=\mu_1}\mathcal{E}_i^{\mathrm{POS}}]$ is upper-bounded by upper bounding $\mathbb{E}[\tau_i\mathbb{1}\{\bigcup_{i:\mu_i=\mu_1}\mathcal{E}_i^{\mathrm{POS}}\}]$ for optimal arms i with $\mu_i = \mu_1$ by Lemma 2, the expected number of overestimations $\mathbb{E}\left[\sum_{t=1}^{KT_\Delta}\mathbb{1}[\mathrm{UCB}(t,i) \geq \mu_1 - \epsilon, i_t = i]\right]$ for non-optimal arms i with $\mu_i < \mu_1$ by Lemma 15, and the expected number of underestimations $\mathbb{E}\left[\sum_{t=1}^{KT_\Delta}\mathbb{1}[\mathrm{UCB}(t,1) < \mu_1 - \epsilon]\right]$ for the optimal arm 1 by Lemma 16.

Lemma 15 *For an arbitrary* $\epsilon > 0$, BAEC[UCB, $\underline{\mu}$, $\overline{\mu}$] *satisfies*

$$
\mathbb{E}\left[\sum_{t=1}^{KT_\Delta}\mathbb{1}[\mathrm{UCB}(t,i) \geq \mu_1 - \epsilon, i_t = i]\right] \leq \frac{\ln KT_\Delta}{2(\Delta_{1i}-2\epsilon)^2} + \frac{1}{2\epsilon^2} + 1
$$

for $i = 2, \ldots, K$ *with* $\mu_i < \mu_1$.

Proof Let $n'_i = \frac{\ln KT_\Delta}{2(\Delta_{1i}-2\epsilon)^2}$. Then,

$$\begin{aligned}
\sum_{t=1}^{KT_\Delta} \mathbb{1}[\mathrm{UCB}(t,i) \geq \mu_1 - \epsilon, i_t = i] &= \sum_{t=1}^{KT_\Delta} \sum_{n=0}^{KT_\Delta - 1} \mathbb{1}\left[\hat{\mu}_i(n) + \sqrt{\frac{\ln t}{2n}} \geq \mu_1 - \epsilon, n_i(t) = n, i_t = i\right] \\
&= \sum_{n=0}^{KT_\Delta - 1} \mathbb{1}\left[\bigcup_{t=1}^{KT_\Delta} \left\{\hat{\mu}_i(n) + \sqrt{\frac{\ln t}{2n}} \geq \mu_1 - \epsilon, n_i(t) = n, i_t = i\right\}\right] \\
&\leq \sum_{n=0}^{KT_\Delta - 1} \mathbb{1}\left[\hat{\mu}_i(n) + \sqrt{\frac{\ln KT_\Delta}{2n}} \geq \mu_1 - \epsilon\right] \\
&\leq \sum_{n=0}^{\lfloor n'_i \rfloor} 1 + \sum_{n=\lfloor n'_i \rfloor + 1}^{\infty} \mathbb{1}\left[\hat{\mu}_i(n) + \sqrt{\frac{\ln KT_\Delta}{2 \cdot \frac{\ln KT_\Delta}{2(\Delta_{1i}-2\epsilon)^2}}} \geq \mu_1 - \epsilon\right] \\
&\leq \frac{\ln KT_\Delta}{2(\Delta_{1i} - 2\epsilon)^2} + 1 + \sum_{n=1}^{\infty} \mathbb{1}\left[\hat{\mu}_i(n) \geq \mu_i + \epsilon\right]
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathbb{E}\left[\sum_{t=1}^{KT_\Delta} \mathbb{1}[\mathrm{UCB}(t,i) \geq \mu_1 - \epsilon, i_t = i]\right] &\leq \frac{\ln KT_\Delta}{2(\Delta_{1i} - 2\epsilon)^2} + 1 + \sum_{n=1}^{\infty} \mathbb{P}\left[\hat{\mu}_i(n) \geq \mu_i + \epsilon\right] \\
&= \frac{\ln KT_\Delta}{2(\Delta_{1i} - 2\epsilon)^2} + 1 + \sum_{n=1}^{\infty} \mathrm{e}^{-2n\epsilon^2} \\
&= \frac{\ln KT_\Delta}{2(\Delta_{1i} - 2\epsilon)^2} + 1 + \frac{1}{\mathrm{e}^{2\epsilon^2} - 1} \\
&\leq \frac{\ln KT_\Delta}{2(\Delta_{1i} - 2\epsilon)^2} + \frac{1}{2\epsilon^2} + 1.
\end{aligned}$$

□

Lemma 16 *For* BAEC[UCB, $\underline{\mu}$, $\overline{\mu}$]*, the following inequality holds.*

$$\mathbb{E}\left[\sum_{t=1}^{KT_\Delta} \mathbb{1}[\mathrm{UCB}(t,1) < \mu_1 - \epsilon]\right] \leq \frac{1}{\epsilon^2} + \frac{1}{4\epsilon^2} \ln \frac{1}{2\epsilon^2}$$

for $0 < \epsilon \leq 1$.

Proof

$$\begin{aligned}
\sum_{t=1}^{KT_\Delta} \mathbb{1}[\mathrm{UCB}(t,1) < \mu_1 - \epsilon] &= \sum_{t=1}^{KT_\Delta} \sum_{n=0}^{KT_\Delta - 1} \mathbb{1}\left[\hat{\mu}_1(n) + \sqrt{\frac{\ln t}{2n}} < \mu_1 - \epsilon, n_1(t) = n\right] \\
&= \sum_{n=0}^{KT_\Delta - 1} \sum_{t=1}^{KT_\Delta} \mathbb{1}\left[t < \mathrm{e}^{2n(\mu_1 - \hat{\mu}_1(n) - \epsilon)^2}, \hat{\mu}_1(n) < \mu_1 - \epsilon, n_1(t) = n\right] \\
&\leq \sum_{n=1}^{KT_\Delta - 1} \mathrm{e}^{2n(\mu_1 - \hat{\mu}_1(n) - \epsilon)^2} \mathbb{1}\left[\hat{\mu}_1(n) \leq \mu_1 - \epsilon\right].
\end{aligned}$$

Define $\mathbb{F}_n(x)$ as $\mathbb{F}_n(x) = \mathbb{P}\{\hat{\mu}_1(n) \le x\}$. Note that $\mathbb{F}_n(x) \le \mathrm{e}^{-2n(\mu_1-x)^2}$ for $x < \mu_1$ by Hoeffding's Inequality. Then,

$$
\begin{aligned}
&\mathbb{E}\left[\sum_{t=1}^{KT_\Delta} \mathbb{1}[\mathrm{UCB}(t,1) < \mu_1 - \epsilon]\right] \\
&\le \sum_{n=1}^{KT_\Delta-1} \mathbb{E}\left[\mathrm{e}^{2n(\mu_1-\hat{\mu}_1(n)-\epsilon)^2}\mathbb{1}\left[\hat{\mu}_1(n) \le \mu_1 - \epsilon\right]\right] \\
&= \sum_{n=1}^{KT_\Delta-1} \int_{-\infty}^{\mu_1-\epsilon} \mathrm{e}^{2n(\mu_1-x-\epsilon)^2}\mathrm{d}\mathbb{F}_n(x) \\
&= \sum_{n=1}^{KT_\Delta-1} \left(\left[\mathrm{e}^{2n(\mu_1-x-\epsilon)^2}\mathbb{F}_n(x)\right]_{-\infty}^{\mu_1-\epsilon} + \int_{-\infty}^{\mu_1-\epsilon} 4n(\mu_1-x-\epsilon)\mathrm{e}^{2n(\mu_1-x-\epsilon)^2}\mathbb{F}_n(x)\mathrm{d}x\right) \\
&\le \sum_{n=1}^{KT_\Delta-1} \left(\mathbb{F}_n(\mu_1-\epsilon) + \int_{-\infty}^{\mu_1-\epsilon} 4n(\mu_1-x-\epsilon)\mathrm{e}^{2n(\mu_1-x-\epsilon)^2}\mathrm{e}^{-2n(\mu_1-x)^2}\mathrm{d}x\right) \\
&\le \sum_{n=1}^{KT_\Delta-1} \left(\mathrm{e}^{-2n\epsilon^2} + \int_{-\infty}^{\mu_1-\epsilon} 4n(\mu_1-x-\epsilon)\mathrm{e}^{-2n\epsilon(2\mu_1-2x-\epsilon)}\mathrm{d}x\right) \\
&= \sum_{n=1}^{KT_\Delta-1} \left(\mathrm{e}^{-2n\epsilon^2} + \frac{1}{4n\epsilon^2}\left[\{4n\epsilon(\mu_1-x-\epsilon)+1\}\mathrm{e}^{-2n\epsilon(2\mu_1-2x-\epsilon)}\right]_{-\infty}^{\mu_1-\epsilon}\right) \\
&= \sum_{n=1}^{KT_\Delta-1} \left(\mathrm{e}^{-2n\epsilon^2} + \frac{1}{4n\epsilon^2}\mathrm{e}^{-2n\epsilon^2}\right) \\
&\le \frac{1}{e^{2\epsilon^2}-1} + \frac{-\ln(1-e^{-2\epsilon^2})}{4\epsilon^2} \quad \left(\text{because} \sum_{n=1}^{\infty} \frac{\left(\mathrm{e}^{-2\epsilon^2}\right)^n}{n} = -\ln\left(1-\mathrm{e}^{-2\epsilon^2}\right)\right) \\
&\le \frac{1}{2\epsilon^2} + \frac{2\epsilon^2 + \ln\frac{1}{e^{2\epsilon^2}-1}}{4\epsilon^2} \\
&\le \frac{1}{2\epsilon^2} + \frac{1}{2} + \frac{1}{4\epsilon^2}\ln\frac{1}{2\epsilon^2} \le \frac{1}{\epsilon^2} + \frac{1}{4\epsilon^2}\ln\frac{1}{2\epsilon^2}.
\end{aligned}
$$

□

Proof of Theorem 8 Let ϵ be $0 < \epsilon \le \min_{i:\Delta_{1i}>0} \Delta_{1i}/4$.

$$
\begin{aligned}
\mathbb{E}[T] =&\mathbb{E}\left[T\mathbb{1}\left\{\bigcup_{i:\mu_i=\mu_1}\mathcal{E}_i^{\mathrm{POS}}\right\}\right] + \mathbb{E}\left[T\mathbb{1}\left\{\bigcap_{i:\mu_i=\mu_1}\overline{\mathcal{E}_i^{\mathrm{POS}}}\right\}\right] \\
\le&\mathbb{E}\left[\sum_{i:\mu_i=\mu_1}\tau_i\mathbb{1}\left\{\bigcup_{i:\mu_i=\mu_1}\mathcal{E}_i^{\mathrm{POS}}\right\}\right] + \mathbb{E}\left[\sum_{t=1}^{KT_\Delta}\mathbb{1}\left\{\mu_{i_t} < \mu_1, \bigcup_{i:\mu_i=\mu_1}\mathcal{E}_i^{\mathrm{POS}}\right\}\right] \\
&+ \mathbb{E}\left[T\mathbb{1}\left[\bigcap_{i:\mu_i=\mu_1}\overline{\mathcal{E}_i^{\mathrm{POS}}}\right]\right]
\end{aligned}
$$

$$
\begin{aligned}
&\le \sum_{i:\mu_i=\mu_1} \mathbb{E}\left[\tau_i \mathbb{1}\left\{\bigcup_{i:\mu_i=\mu_1} \mathcal{E}_i^{\mathrm{POS}}\right\}\right] \\
&\quad + \mathbb{E}\left[\sum_{t=1}^{KT_{\vartriangle}} \mathbb{1}\left[\{\mathrm{UCB}(t,i_t) \ge \mu_1-\epsilon, \mu_{i_t} < \mu_1\} \cup \{\mathrm{UCB}(t,1) < \mu_1-\epsilon\}\right]\right] + \mathbb{E}\left[T\mathbb{1}\left\{\overline{\mathcal{E}_1^{\mathrm{POS}}}\right\}\right] \\
&\le \sum_{i:\mu_i=\mu_1} \mathbb{E}\left[\tau_i \mathbb{1}\left\{\bigcup_{i:\mu_i=\mu_1} \mathcal{E}_i^{\mathrm{POS}}\right\}\right] + \sum_{i:\mu_i<\mu_1} \mathbb{E}\left[\sum_{t=1}^{KT_{\vartriangle}} \mathbb{1}\left[\mathrm{UCB}(t,i) \ge \mu_1-\epsilon, i_t = i\}\right]\right] \\
&\quad + \mathbb{E}\left[\sum_{t=1}^{KT_{\vartriangle}} \mathbb{1}\left[\mathrm{UCB}(t,1) < \mu_1-\epsilon\right]\right] + \mathbb{E}\left[T\mathbb{1}\left\{\overline{\mathcal{E}_1^{\mathrm{POS}}}\right\}\right] \\
&\le \sum_{i:\mu_i=\mu_1}\left(\frac{\mathbb{P}\left[\bigcup_{i:\mu_i=\mu_1}\mathcal{E}_i^{\mathrm{POS}}\right]}{2\Delta_i^2}\ln\frac{KN_{\vartriangle}}{\delta} + O\left(\left(\ln\frac{KN_{\vartriangle}}{\delta}\right)^{\frac{2}{3}}\right)\right) \quad \text{(by Lemma 2)} \\
&\quad + \sum_{i:\mu_i<\mu_1}\left(\frac{\ln KT_{\vartriangle}}{2(\Delta_{1i}-2\epsilon)^2} + \frac{1}{2\epsilon^2} + 1\right) + \frac{1}{\epsilon^2} + \frac{1}{4\epsilon^2}\ln\frac{1}{2\epsilon^2} + KT_{\vartriangle}\mathbb{P}\left[\overline{\mathcal{E}_1^{\mathrm{POS}}}\right]. \\
&\quad \text{(by Lemmas 15 and 16, and Theorem 4)}
\end{aligned}
$$

Since $\frac{1}{\Delta_{1i}^2}+\frac{12\epsilon}{\Delta_{1i}^3}-\frac{1}{(\Delta_{1i}-2\epsilon)^2}=\frac{4\epsilon(\Delta_{1i}-4\epsilon)(2\Delta_{1i}-3\epsilon)}{\Delta_{1i}^3(\Delta_{1i}-2\epsilon)^2}\ge 0$ holds for $0<\epsilon\le\frac{\Delta_{1i}}{4}$, $\frac{1}{(\Delta_{1i}-2\epsilon)^2}\le\frac{1}{\Delta_{1i}^2}+\frac{12\epsilon}{\Delta_{1i}^3}$ holds. Thus, by setting ϵ to $O((\ln KT_{\vartriangle})^{-1/3})$, we have

$$
\begin{aligned}
\mathbb{E}[T] &\le \sum_{i:\mu_i=\mu_1}\left(\frac{1}{2\Delta_i^2}\ln\frac{KN_{\vartriangle}}{\delta} + O\left(\left(\ln\frac{KN_{\vartriangle}}{\delta}\right)^{\frac{2}{3}}\right)\right) + \sum_{i:\mu_i<\mu_1}\left(\frac{\ln KT_{\vartriangle}}{2\Delta_{1i}^2} + O((\ln KT_{\vartriangle})^{\frac{2}{3}})\right) \\
&\quad + O((\ln KT_{\vartriangle})^{\frac{2}{3}}\ln\ln KT_{\vartriangle}) + \frac{\mathrm{e}^{2\underline{\Delta}_1^2}KT_{\vartriangle}}{2\underline{\Delta}_1^2}\left(\frac{\delta}{N_{\vartriangle}}\right)^{\left(\frac{\Delta_1}{\max\{\theta_U,1-\theta_L\}}\right)^2}. \quad \text{(by Lemma 14)}
\end{aligned}
$$

□

References

Audibert, J., Bubeck, S., & Munos, R. (2010). Best arm identification in multi-armed bandits. In *Proceedings of the 23rd conference on learning theory* (pp. 41–53).

Auer, P., Cesa-Bianchi, N., & Fischer, P. (2002). Finite-time analysis of the multiarmed bandit problem. *Machine Learning*, *47*(2–3), 235–256.

Auer, P., Cesa-Bianchi, N., Freund, Y., & Schapire, R. E. (2003). The nonstochastic multiarmed bandit problem. *SIAM Journal on Computing*, *32*(1), 48–77.

Bubeck, S., Munos, R., & Stoltz, G. (2011). Pure exploration in finitely-armed and continuous-armed bandits. *Theoretical Computer Science*, *412*(19), 1832–1852.

Bubeck, S., Wang, T., & Viswanathan, N. (2013). Multiple identifications in multi-armed bandits. In *Proceedings of the 30th international conference on machine learning, proceedings of machine learning research*, vol 28 (pp. 258–265).

Even-Dar, E., Mannor, S., & Mansour, Y. (2006). Action elimination and stopping conditions for the multi-armed bandit and reinforcement learning problems. *Journal of Machine Learning Research*, *7*, 1079–1105.

Gabillon, V., Ghavamzadeh, M., & Lazaric, A. (2012). Best arm identification: A unified approach to fixed budget and fixed confidence. *Advances in Neural Information Processing Systems*, *25*, 3212–3220.

Haka, A. S., Volynskaya, Z., Gardecki, J. J. A., Nazemi, J., Shenk, R., Wang, N., et al. (2009). Diagnosing breast cancer using Raman spectroscopy: Prospective analysis. *Journal of Biomedical Optics*, *14*(5), 054023.

Kalyanakrishnan, S., Tewari, A., Auer, P., & Stone, P. (2012). Pac subset selection in stochastic multi-armed bandits. In *Proceedings of the 29th international conference on machine learning* (pp. 655–662).

Kano, H., Honda, J., Sakamaki, K., Matsuura, K., Nakamura, A., & Sugiyama, M. (2017). Good arm identification via bandit feedback. arXiv e-prints arXiv:1710.06360.

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

Kaufmann, E., & Kalyanakrishnan, S. (2013). Information complexity in bandit subset selection. In *Proceedings of the 26th annual conference on learning theory, proceedings of machine learning research*, vol 30 (pp. 228–251).

Kaufmann, E., Koolen, W. M., & Garivier, A. (2018). Sequential test for the lowest mean: From Thompson to Murphy sampling. In *Proceedings of the 32nd conference on neural information processing systems* (pp. 6333–6343).

Littlestone, N., & Warmuth, M. K. (1994). The weighted majority algorithm. *Information Computation*, *108*(2), 212–261.

Locatelli, A., Gutzeit, M., & Carpentier, A. (2016). An optimal algorithm for the thresholding bandit problem. In *Proceedings of the 33rd international conference on machine learning*, vol PMLR 48 (pp. 1690–1698).

Robbins, H. (1952). Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society*, *58*(5), 527–535.

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

Zhang, W., Yuan, S., Wang, J., & Shen, X. (2014). Real-time bidding benchmarking with iPinYou dataset. arXiv e-prints. arXiv:1407.7073.