Supplementary Material for "Thompson Exploration with Best Challenger Rule in Best Arm Identification"

Jongyeong Lee LEE@MS.K.U-TOKYO.AC.JP
The University of Tokyo, RIKEN AIP

Junya Honda HONDA@I.KYOTO-U.AC.JP
Kyoto University, RIKEN AIP

Masashi Sugiyama SUGI@K.U-TOKYO.AC.JP
RIKEN AIP, The University of Tokyo

Editors: Berrin Yanıkoğlu and Wray Buntine

In this supplementary material, we provide detailed proof of the results presented in the main paper.

Appendix A. Additional notation

Before beginning the proof, we first define good events on estimates $\hat{\mu}_i(t)$ and Thompson samples $\tilde{\mu}_i(t)$ for any $\epsilon > 0$,

$$\begin{aligned}
\mathcal{A}_i(t) = \mathcal{A}_{i,\epsilon}(t) &:= \begin{cases} \{\hat{\mu}_1(t) \geq \mu_1 - \epsilon\}, & \text{if } i = 1, \\ \{\hat{\mu}_i(t) \leq \mu_i + \epsilon\}, & \text{otherwise,} \end{cases} \\
\mathcal{B}_i(t) = \mathcal{B}_{i,\epsilon}(t) &:= \{|\hat{\mu}_i(t) - \mu_i| \leq \epsilon\}, \\
\tilde{\mathcal{B}}_i(t) = \tilde{\mathcal{B}}_{i,\epsilon}(t) &:= \{|\tilde{\mu}_i(t) - \mu_i| \leq \epsilon\}, \\
\mathcal{M}(t) &:= \{m(t) = \tilde{m}(t)\},
\end{aligned}$$

Note that for all $i \in [K]$ and $t \in \mathbb{N}$, $\mathcal{B}_i(t) \subset \mathcal{A}_i(t)$ holds.

Next, let us define another random variables $D_1 = D_{1,\epsilon} := \max_{i \neq 1} D_{i,\epsilon}$ where

$$D_i = D_{i,\epsilon} := \sup_{t \geq 2K+1} \mathbb{1}[\mathcal{B}_{i,\epsilon}^c(t)] N_i(t) d\left(\hat{\mu}_i(t), \hat{\mu}_1(t)\right)$$

denotes the supremum of $N_a(t) d\left(\hat{\mu}_a(t), \hat{\mu}_1(t)\right)$ when $\mathcal{B}_{i,\epsilon}^c(t)$ occurs. In other words,

$$\{N_a(t) d\left(\hat{\mu}_a(t), \hat{\mu}_1(t)\right) \geq D_{i,\epsilon}\} \implies \{\mathbb{1}[\mathcal{B}_{i,\epsilon}(t)] = 1\}.$$

We further define $\underline{d}_1 = d(\mu_1 - \epsilon, \mu_2 + \epsilon)$ and for $i \neq 1$

$$\underline{d}_i = \min_{\substack{\mu \in [\mu_i', \mu_1'], \\ \mu_i' \leq \mu_i + \epsilon,\ \mu_1' \geq \mu_1 - \epsilon, \\ d(\mu_i', \mu) \geq d(\mu_1', \mu)}} d(\mu_i', \mu). \tag{14}$$

Appendix B. Proof of Lemma 1: Subdifferentials

Here, we derive the subdifferential of the objective function g.
Proof By abuse of notation, we define a characteristic function $I_{\Sigma_K} : \mathbb{R}^K \to \mathbb{R}$,

$$I_{\Sigma_K}(x) = \begin{cases} 0, & \text{if } x \in \Sigma_K \\ -\infty, & \text{if } x \notin \Sigma_K. \end{cases}$$

Then, the problem in (3) can be written as

$$\sup_{\boldsymbol{w} \in \Sigma_K} \min_{i \neq 1} f_i(\boldsymbol{w}; \boldsymbol{\mu}) = \max_{\boldsymbol{w} \in \mathbb{R}^K} \left\{ \min_{i \neq 1} f_i(\boldsymbol{w}) + I_{\Sigma_K}(\boldsymbol{w}) \right\}. \tag{15}$$

Then, the set of differential of (15) is

$$\partial \left(\min_{a \neq 1} f_a(\boldsymbol{w}) + I_{\Sigma_K}(\boldsymbol{w}) \right) = \left\{ q + r : q \in \partial \min_{i \neq a} f_a(\boldsymbol{w}),\ r \in \partial I_{\Sigma_K}(\boldsymbol{w}) \right\}.$$

Let $\partial I_{\Sigma_K}(\boldsymbol{w})$ denote the set of subgradient $\boldsymbol{v}$ of I_{Σ_K} at point $(\boldsymbol{w}; \boldsymbol{\mu})$. Then, $\partial I_{\Sigma_K}(\boldsymbol{w})$ is written as

$$\partial I_{\Sigma_K}(\boldsymbol{w}) = \{ \boldsymbol{v} \in \mathbb{R}^K : \forall \boldsymbol{x} \in \mathbb{R}^K, I_{\Sigma_K}(\boldsymbol{x}) \leq I_{\Sigma_K}(\boldsymbol{w}) + \boldsymbol{v}^\top (\boldsymbol{x} - \boldsymbol{w}) \} \tag{16}$$

From the definition of I_{Σ_K}, if $\boldsymbol{x} \notin \Sigma_K$, the inequality constraint in (16) always holds for any $\boldsymbol{v} \in \mathbb{R}^K$. Thus, it suffices to show that

$$\begin{aligned} \partial I_{\Sigma_K}(\boldsymbol{w}) &= \{ \boldsymbol{v} \in \mathbb{R}^K : \forall \boldsymbol{x} \in \Sigma_K, I_{\Sigma_K}(\boldsymbol{x}) \leq I_{\Sigma_K}(\boldsymbol{w}) + \boldsymbol{v}^\top (\boldsymbol{x} - \boldsymbol{w}) \} \\ &= \{ r\mathbf{1} : r \in \mathbb{R} \}, \end{aligned} \tag{17}$$

which implies that all subgradients $\boldsymbol{v}$ can be written as a multiple of the K-dimensional all-one vector $\mathbf{1} = [1, \ldots, 1]$. To show the equivalence, we will show that

$$\begin{aligned} &(B1) : \{ r\mathbf{1} : r \in \mathbb{R} \} \subset \partial I_{\Sigma_K}(\boldsymbol{w}) \\ &(B2) : \{ r\mathbf{1} : r \in \mathbb{R} \} \supset \partial I_{\Sigma_K}(\boldsymbol{w}). \end{aligned}$$

B.1. Case (B1)

Note that $\mathbf{0} \in \partial I_{\Sigma_K}(\boldsymbol{w})$, which implies $\partial I_{\Sigma_K}(\boldsymbol{w}) \neq \emptyset$. Since $\boldsymbol{x} \in \Sigma_K$, $\boldsymbol{v} \in \partial I_{\Sigma_K}(\boldsymbol{w})$ satisfies $0 \leq \boldsymbol{v}^\top (\boldsymbol{x} - \boldsymbol{w})$ for all $\boldsymbol{x} \in \Sigma_K$. One can see that $\{ r\mathbf{1} : r \in \mathbb{R} \} \subset \partial I_{\Sigma_K}(\boldsymbol{w})$ for $\boldsymbol{w} \in \Sigma_K$ since $\sum_{i=1}^K w_i = \sum_{i=1}^K x_i = 1$ from the assumption.

B.2. Case (B2)

Then, we need to show the equality in (17) for $\boldsymbol{w} \in \operatorname{Int} \Sigma_K$. At first, let assume $K \geq 2$ and $\boldsymbol{v} = r\mathbf{1} + \sum_{i=1}^K a_i e_i$, where e_i is a standard basis for $\mathbb{R}^K$ and $a_i \in \mathbb{R}$. Then, $\forall \boldsymbol{x} \in \Sigma_K$,

$$0 \leq \sum_{i=1}^K a_i (x_i - w_i) \tag{18}$$

holds. We will prove the equality in (17) by contradiction, i.e., we assume that there exists $i \neq j \in [K]$ such that $a_i \neq a_j$. From the definition of $\mathrm{Int}\Sigma_K$, we can take a positive constant $\epsilon \in \mathbb{R}_+$ satisfying $0 < \epsilon < \min(\min_i w_i, 1 - \max_i(w_i))$.[1]

Define two K dimensional vectors as

$$\boldsymbol{x}^1 = (x_i)_{i=1}^K = \begin{cases} w_i, & \text{if } i \in [K] \setminus \{i_1, i_2\}, \\ w_i + \epsilon, & \text{if } i = i_1, \\ w_i - \epsilon, & \text{if } i = i_2, \end{cases}$$

and

$$\boldsymbol{x}^2 = (x_i)_{i=1}^K = \begin{cases} w_i, & \text{if } i \in [K] \setminus \{i_1, i_2\}, \\ w_i - \epsilon, & \text{if } i = i_1, \\ w_i + \epsilon, & \text{if } i = i_2, \end{cases}$$

where $i_1 \neq i_2 \in [K]$. Then, both $\boldsymbol{x}^1$ and $\boldsymbol{x}^2$ are in Σ_K. From (18), this implies that two inequalities

$$0 \leq \epsilon(a_{i_1} - a_{i_2}) \text{ and } 0 \leq -\epsilon(a_{i_1} - a_{i_2})$$

hold at the same time. Thus, $a_{i_1} = a_{i_2}$ should hold. However, we can make these kinds of vectors for every pair of bases, which means that $\nexists i \neq j \in [K]$ such that $a_i \neq a_j$. This is a contradiction, and thus (17) holds.

B.3. Conclusion

Consequently, it holds $\forall \boldsymbol{w} \in \mathrm{Int}\Sigma_K$ that

$$\begin{aligned} \partial g &= \left\{ q + r\mathbf{1} : q \in \mathbf{Co} \bigcup \{\partial f_i(\boldsymbol{w}; \boldsymbol{\mu}) : f_i(\boldsymbol{w}; \boldsymbol{\mu}) = g(\boldsymbol{w}; \boldsymbol{\mu})\}, r \in \mathbb{R} \right\} \\ &= \left\{ q + r\mathbf{1} : q \in \mathbf{Co} \bigcup \{\nabla_{\boldsymbol{w}} f_i(\boldsymbol{w}; \boldsymbol{\mu}) : f_i(\boldsymbol{w}) = g(\boldsymbol{w})\}, r \in \mathbb{R} \right\}, \end{aligned}$$

where $\mathbf{Co} \bigcup \{\nabla_{\boldsymbol{w}} f_i(\boldsymbol{w}; \boldsymbol{\mu}) : f_i(\boldsymbol{w}; \boldsymbol{\mu}) = g(\boldsymbol{w}; \boldsymbol{\mu})\}$ is the convex hull of the union of superdifferentials of all active function at $\boldsymbol{w}$. Let us define the set

$$\mathcal{J}(\boldsymbol{w}; \boldsymbol{\mu}) := \underset{i \neq 1}{\arg\min}\, f_i(\boldsymbol{w}; \boldsymbol{\mu}) = \{i \in [K] : f_i = g\},$$

which concludes the proof. ■

Appendix C. Comparison with other optimality notions

In this section, we provide more detail that completes Sections 4 and 5.

1. Note that such ϵ always exists by Archimedean property if w is in the interior of the probability simplex, i.e., $\forall i \in [K], w_i \neq 0, 1$.

C.1. Two-armed bandits

Firstly, let us introduce a function that enables us to derive a more explicit formula for $\boldsymbol{w}^*(\boldsymbol{\mu})$, for any $i \neq 1$,

$$k_i(x;\boldsymbol{\mu}) = d\left(\mu_1, \frac{1}{1+x}\mu_1 + \frac{x}{1+x}\mu_i\right) + xd\left(\mu_i, \frac{1}{1+x}\mu_1 + \frac{x}{1+x}\mu_i\right).$$

As demonstrated in Garivier and Kaufmann (2016), this function is a strictly increasing bijective mapping from $[0,\infty)$ onto $[0, d(\mu_1,\mu_a))$. Therefore, one can define l_i as the inverse function of k_i for any $i \neq 1$ and l_1 as a constant function, which is

$$\begin{aligned} k_i^{-1} = l_i : [0, d(\mu_1,\mu_i)) &\mapsto [0,\infty) \\ l_1 : [0, d(\mu_1,\mu_i)) &\mapsto 1. \end{aligned} \tag{19}$$

Then, Garivier and Kaufmann (2016) provided the following characterization of $\boldsymbol{w}^*(\boldsymbol{\mu})$.

Lemma 5 (Theorem 5 in Garivier and Kaufmann (2016)) *For every $i \in [K]$,*

$$\boldsymbol{w}_i^*(\boldsymbol{\mu}) = \frac{l_i(y^*)}{\sum_{a=1}^K l_a(y^*)},$$

where y^* is the unique solution of the equation $F_{\boldsymbol{\mu}}(y) = 1$, and where

$$F_{\boldsymbol{\mu}} : y \mapsto \sum_{i=2}^K \frac{d\left(\mu_1, \frac{\mu_1 + l_i(y)\mu_i}{1+l_i(y)}\right)}{d\left(\mu_i, \frac{\mu_1 + l_i(y)\mu_i}{1+l_i(y)}\right)}$$

is a continuous, increasing function on $[0, d(\mu_1,\mu_2))$ such that $F_{\boldsymbol{\mu}}(0) = 0$ and $F_{\boldsymbol{\mu}}(y) = \infty$ when $y \to d(\mu_1,\mu_2)$.

However, to derive a more explicit formula for the maximizer of (11), we require another function for any $i \neq 1$

$$h_i(z;\boldsymbol{\mu}) = (1-z)d(\mu_1, (1-z)\mu_1 + z\mu_i) + zd(\mu_i, (1-z)\mu_1 + z\mu_i),$$

whose domain is $[0,1]$. The derivative of this function is

$$h_i'(z;\boldsymbol{\mu}) = d(\mu_i, (1-z)\mu_1 + z\mu_i) - d(\mu_1, (1-z)\mu_1 + z\mu_i).$$

Thus, $h_i(z;\boldsymbol{\mu})$ is a concave function with $h_i(0;\boldsymbol{\mu}) = 0$ and $h_i(1,\boldsymbol{\mu}) = 0$. It reaches its maximum at

$$z_i^*(\boldsymbol{\mu}) : d(\mu_i, (1-z_i^*)\mu_1 + z_i^*\mu_i) = d(\mu_1, (1-z_i^*)\mu_1 + z_i^*\mu_i). \tag{20}$$

Therefore, one can see that $\gamma = z_2^*$. From the definitions of f_i, k_i, and h_i, one can find the following relationship

$$f_i(\boldsymbol{w};\boldsymbol{\mu}) = w_1 k_i\left(\frac{w_i}{w_1};\boldsymbol{\mu}\right) = (w_1 + w_i)h_i\left(\frac{w_i}{w_1+w_i};\boldsymbol{\mu}\right). \tag{21}$$

For $z_i = \frac{w_i}{w_1+w_i}$, the equality between h_i and k_i can be written as

$$h_i(z_i; \boldsymbol{\mu}) = (1 - z_i)k_i\left(\frac{z_i}{1 - z_i}; \boldsymbol{\mu}\right).$$

We further define the problem-dependent constant $\underline{z}_i \in [0, 1]$ for $i \neq 1$ satisfying

$$\underline{z}_i : k_i\left(\frac{\underline{z}_i}{1 - \underline{z}_i}; \boldsymbol{\mu}\right) = k_2\left(\frac{z_2^*}{1 - z_2^*}; \boldsymbol{\mu}\right) \tag{22}$$

and $\underline{z}_1 = \frac{1}{2}$. Here, we have $\underline{z}_2 = z_2^*$ and $\underline{z}_i \leq z_2^*$ since k_i is strictly increasing and $k_i(x; \boldsymbol{\mu}) \leq k_j(x; \boldsymbol{\mu})$ holds for any $x \in \mathbb{R}_+$ if $\mu_i \leq \mu_j$ (see Garivier and Kaufmann, 2016, Appendix A.3.). Based on $\underline{z}_i$, we define a normalized proportion $\underline{w} \in \Sigma_K$ by

$$\underline{w}_i(\boldsymbol{\mu}) = \frac{\frac{\underline{z}_i}{1-\underline{z}_i}}{\sum_{i=1}^K \frac{\underline{z}_i}{1-\underline{z}_i}} = \frac{l_i(\underline{y})}{\sum_{i=1}^K l_i(\underline{y})}, \tag{23}$$

where $\underline{y} = k_i\left(\frac{\underline{z}_i}{1-\underline{z}_i}; \boldsymbol{\mu}\right)$ for any $i \neq 1$. Therefore, Theorem 3 implies that the empirical proportion of arm plays of BC-TE will converge to $\underline{w}$, which is equivalent to $g(\boldsymbol{w}^t; \hat{\boldsymbol{\mu}}(t)) \to g(\underline{w}; \boldsymbol{\mu})$. Here, one can see that $F_{\boldsymbol{\mu}}(\underline{y}) \geq 1$ since

$$\frac{d\left(\mu_1, \frac{\mu_1 + \frac{\underline{z}_2}{1-\underline{z}_2}\mu_2}{1+\frac{\underline{z}_2}{1-\underline{z}_2}}\right)}{d\left(\mu_2, \frac{\mu_1 + \frac{\underline{z}_2}{1-\underline{z}_2}\mu_2}{1+\frac{\underline{z}_2}{1-\underline{z}_2}}\right)} = \frac{d\left(\mu_1, (1 - \underline{z}_2)\mu_1 + \underline{z}_2\mu_2)\right)}{d\left(\mu_2, (1 - \underline{z}_2)\mu_1 + \underline{z}_2\mu_2)\right)} = 1$$

holds from the definition of $\underline{z}_2 = z_2^*$ in (20), which directly implies that $\underline{y} \geq y^*$. However, it is important to note that from $\underline{z}_i \leq z_i^*$, it always hold that for any $i \neq 1$

$$\frac{d\left(\mu_1, (1 - \underline{z}_i)\mu_1 + \underline{z}_i\mu_i)\right)}{d\left(\mu_i, (1 - \underline{z}_i)\mu_1 + \underline{z}_i\mu_2)\right)} \leq \frac{d\left(\mu_1, (1 - z_i^*)\mu_1 + z_i^*\mu_i)\right)}{d\left(\mu_i, (1 - z_i^*)\mu_1 + z_i^*\mu_i)\right)} = 1.$$

This implies that

$$1 \leq F_{\boldsymbol{\mu}}(\underline{y}) \leq K - 1, \tag{24}$$

where the right equality holds only when $\mu_2 = \mu_3 = \ldots = \mu_K$. Here, it is important to note that the left equality is always valid for two-armed bandit problems. In other words, BC-TE is *asymptotically optimal* in the context of two-armed bandit problems.

C.2. Gaussian bandits

Here, we prove Lemma 4 based on the definitions provided in Section C.1.

Proof of Lemma 4 Since $d(\mu, \mu') = \frac{(\mu-\mu')^2}{2\sigma^2}$, for any $i \neq 1$ and $\Delta_i = \mu_1 - \mu_i$

$$k_i(x; \boldsymbol{\mu}) = \left(\frac{x}{1+x}\right)^2 \frac{\Delta_i^2}{2\sigma^2} + \frac{x}{(1+x)^2}\frac{\Delta_i^2}{2\sigma^2} = \frac{x}{1+x}\frac{\Delta_i^2}{2\sigma^2}$$

$$h_i(z; \boldsymbol{\mu}) = z(1 - z)\frac{\Delta_i^2}{2\sigma^2}.$$

Firstly, from (20), the maximizers of h_i, z_i^* satisfies

$$\frac{\Delta_i^2}{2\sigma^2}(1-z_i^*)^2 = \frac{\Delta_i^2}{2\sigma^2}(z_i^*)^2,$$

which implies that $z_i^* = 1/2$ for any $i \neq 1$. Then, for any $i \neq 1$, from the definition of $\underline{z}_i$ in (22), it holds

$$\begin{aligned} k_2(1;\boldsymbol{\mu}) = \frac{\Delta_2^2}{4\sigma^2} &= k_i\left(\frac{\underline{z}_i}{1-\underline{z}_i};\boldsymbol{\mu}\right) \\ &= \frac{\Delta_i^2}{2\sigma^2}\underline{z}_i, \end{aligned}$$

which implies $\underline{z}_i = \frac{\Delta_2^2}{2\Delta_i^2}$ for $i \neq 1$. Therefore, we obtain that $\underline{w}_i = \frac{\frac{\Delta_2^2}{2\Delta_i^2-\Delta_2^2}}{\sum_{a=1}^K \frac{\Delta_2^2}{2\Delta_a^2-\Delta_2^2}}$. By letting $\Delta_1 = \Delta_2$, the objective function g at $\underline{w}$ can be written as

$$g(\underline{w};\boldsymbol{\mu}) = \underline{w}_1 k_i\left(\frac{\underline{z}_i}{1-\underline{z}_i};\boldsymbol{\mu}\right) = \frac{1}{\sum_{a=1}^K \frac{\Delta_2^2}{2\Delta_a^2-\Delta_2^2}}\frac{\Delta_2^2}{4\sigma^2},$$

which implies that

$$\underline{T}(\boldsymbol{\mu}) = \sum_{i=1}^K \frac{4\sigma^2}{\Delta_i^2 + (\Delta_i^2 - \Delta_2^2)}.$$

■

C.3. Additional numerical results

Here, we first provide additional comparisons between $\underline{T}(\boldsymbol{\mu})$ and $T^{1/2}(\boldsymbol{\mu})$.

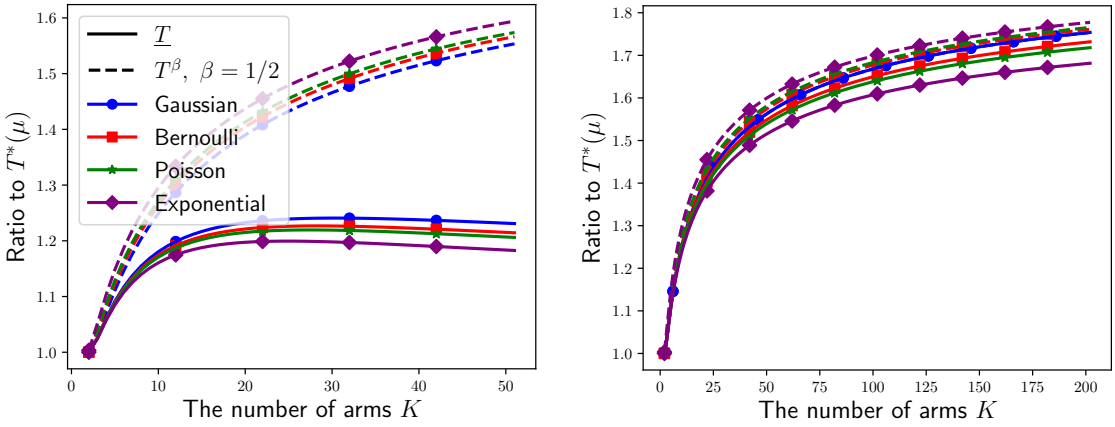
In Figure 3.(a), we zoom in on Figure 1.(a) from the main paper specifically for $K \leq 50$. It can be observed that $\underline{T}(\boldsymbol{\mu}^{(1)})$ is closer to $T^*(\boldsymbol{\mu}^{(1)})$ compared to $T^{1/2}(\boldsymbol{\mu}^{(1)})$. Next, we consider a worst-case instance $\boldsymbol{\mu}'$ based on $\boldsymbol{\mu}^{(1)} = (0.3, 0.21)$, where we add additional arm $\mu_K = \mu_2$ for any K in Figure 3.(b). Therefore, in $\boldsymbol{\mu}'$, all suboptimal arms share the same expected rewards, e.g., $\mu' = (0.3, 0.21, 0.21, 0.21)$ for $K = 4$. This instance is of specific interest since one can observe that $\underline{T}(\boldsymbol{\mu})$ differs from $T^*(\boldsymbol{\mu})$ at most when all suboptimal arms have the same expected rewards according to (24). Even in such cases, $\underline{T}(\boldsymbol{\mu}')$ and $T^{1/2}(\boldsymbol{\mu}')$ exhibit a similar tendency, which would make BC-TE a reasonable policy in general.

Next, for the implementation in Section 5, we focus on T-D in our experiments although there exist two versions of the TaS policy. T-D directly tracks the optimal proportion of arm plays at each round ($N(t) \rightsquigarrow t\boldsymbol{w}^*(\hat{\boldsymbol{\mu}}(t))$), and it has been found to outperform the version with C-tracking in experiments, which tracks the cumulative optimal proportions ($N(t) \rightsquigarrow \sum_{s\leq t} \boldsymbol{w}^*(\hat{\boldsymbol{\mu}}(s))$).

Appendix D. Additional experimental results

In this section, we provide additional experimental results where the rewards follow the exponential distribution and Pareto distribution.

Exponential bandits In the first experiment, we consider the 5-armed Bernoulli bandit instance $\boldsymbol{\mu}_5^{\mathrm{E}} = (0.5, 0.45, 0.43, 0.4, 0.3)$ where $\boldsymbol{w}^*(\boldsymbol{\mu}_5^{\mathrm{B}}) = (0.41, 0.40, 0.13, 0.05, 0.01)$.



(*a*) Instance $\boldsymbol{\mu}^{(1)}$ with small K. (*b*) Worst case instance $\boldsymbol{\mu}'$ with varying K.

Figure 3: The ratio of $\underline{T}(\boldsymbol{\mu})$ and $T^{1/2}(\boldsymbol{\mu})$ to $T^*(\boldsymbol{\mu})$ for different reward distributions.

Pareto bandits In the second experiment, we consider the 4-armed Pareto bandit instance $\boldsymbol{\mu}_4^{\mathrm{P}} = (5.0, 3.0, 2.0, 1.5)$ with unit scale $\sigma = 1$ where $\boldsymbol{w}^*(\boldsymbol{\mu}_4^{\mathrm{P}}) = (0.34, 0.60, 0.04, 0.01)$. The density function of the Pareto distribution with shape $\theta > 0$ and scale $\sigma > 0$ is written as

$$f_{\mathrm{P}}(x; \theta, \sigma) = \frac{\theta \sigma^{\theta}}{x^{\theta+1}}.$$

Notice that since $\sigma = 1$, the shape parameter is given as $\boldsymbol{\theta} = (1.25, 1.5, 2, 3)$, where the first three arms have *infinite* variance. It is worth noting that the sample complexity of T3C for $\delta \in \{0.01, 0.001\}$ becomes extremely larger than other policies (e.g., more than 25,000), we exclude the result of T3C in this section although it performs well in the Gaussian and Bernoulli bandits.

Results The overall results are presented in Table 3. Similarly to the Gaussian and Bernoulli cases, both BC-TE and FWS-TE consistently show a better empirical performance than other optimal policies across most risk parameters, especially when large δ is considered. Although the empirical probability of misidentification (error rate) for each policy is less than given threshold δ for most cases, their error rates exceed the threshold when we considered $\boldsymbol{\mu}_4^{\mathrm{P}}$ with $\delta = 0.001$ as shown in Table 4. This implies that the current choice of stopping rule, $\beta(t, \delta) = \log(\log(t) + 1)/\delta)$, a widely-used heuristic, may be not appropriate when one considers the bandit instance possibly with infinite variance.

Appendix E. Proof of Theorem 2: Convergence of empirical means

We begin the proof of Theorem 2 by introducing two lemmas that show a sufficient condition to occur $\mathcal{B}_i(t)$ for $i = 1$ and $i \neq 1$, respectively.

Lemma 6 *For any constant $M > 0$, assume that*

$$\left\{m(t) = 1,\ j(t) = j,\ i(t) = j,\ \mathcal{A}_1(t), \mathcal{B}_j(t), \mathcal{M}(t),\ N_j(t) > \max\left\{M, D_1/\underline{d}_j\right\}\right\}$$

Table 3: Sample complexity over 3,000 independent runs, where outperforming policies are highlighted in boldface using one-sided Welch's t-test with the significance level 0.05. LB denotes the lower bound in (2), and PLB denotes the practical version of LB considered in Degenne et al. (2019). $\boldsymbol{\mu}_5^{\mathrm{E}}$ denotes 5-armed Exponential bandit instance with means $(0.5, 0.45, 0.43, 0.4, 0.3)$ and $\boldsymbol{\mu}_4^{\mathrm{P}}$ denotes 4-armed Pareto bandit instance with means $(5.0, 3.0, 2.0, 1.5)$ and unit scale.

$\boldsymbol{\mu}$	δ	BC-TE	FWS-TE	FWS	T-D	LMA	RR	PLB	LB
$\boldsymbol{\mu}_5^{\mathrm{B}}$	0.2	**2910**	**2938**	3086	3158	4092	6471	3434	747
	0.1	**3568**	**3623**	3791	3840	4851	7753	4074	1579
	0.01	**5743**	**5849**	5938	5977	7165	12032	6182	4046
	0.001	**7977**	**8010**	**8085**	**8023**	9533	16201	8278	6194
$\boldsymbol{\mu}_4^{\mathrm{P}}$	0.2	**1164**	**1171**	**1178**	1268	1695	2329	937	212
	0.1	**1447**	**1478**	**1457**	1554	2016	2792	1120	449
	0.01	**2396**	**2379**	**2376**	2493	3059	4323	1720	1150
	0.001	**3270**	**3249**	**3174**	3366	4026	5792	2318	1760

Table 4: Error rate for $\boldsymbol{\mu}_4^{\mathrm{P}}$ and $\delta = 0.001$.

BC-TE	FWS-TE	FWS	T-D	LMA	RR
0.004	0.0047	0.0073	0.005	0.008	0.005

occurred for some t. Then, for all $t' \geq t$, we have $\mathbb{1}[\mathcal{B}_1(t')] = 1$ and

$$N_1(t) \geq \frac{\max\{\underline{d}_j M, D_1\}}{d(\mu_1 + \epsilon, \mu_j - \epsilon)}.$$

Lemma 7 *For any constant $M > 0$, assume that*

$$\left\{ m(t) = 1,\ i(t) = 1,\ \mathcal{A}_{j(t)}(t), \mathcal{B}_1(t), \mathcal{M}(t),\ N_1(t) > \max\left\{ M, \max_{i \neq 1} \frac{D_i}{\underline{d}_i} \right\} \right\}$$

occurred for some t. Then, for all $i \neq 1$ and $t' \geq t$, we have $\mathbb{1}[\mathcal{B}_i(t')] = 1$ and

$$N_i(t) \geq \frac{\max\{\underline{d}_i M, D_i\}}{d(\mu_1 + \epsilon, \mu_i - \epsilon)}.$$

Therefore, if both events in Lemmas 6 and 7 occurred until rounds T, only $\{\mathcal{B}_i(t)\}$ occurs for all $i \in [K]$ and $t \geq T$. The proofs of these lemmas are postponed to Section E.1.

Proof of Theorem 2 Firstly, let us define another random variable $T_C \leq T_B$ such that

$$\forall s \geq T_C : \mathbb{1}[\mathcal{B}_1(s)] = 1,$$

which implies that the mean estimate of the optimal arm is close to its true value after T_C rounds. Let $D = \max\left\{M, \frac{D_1}{\min_{a \in [K]} \underline{d}_a}\right\}$ for some positive constant M specified later and $T_M = \max(KD, T_C)$.

Let us consider a subset of rounds with any fixed $T > T_M$

$$\begin{aligned}
S_1(T) &:= \{s \in [T_M, T] \cap \mathbb{N} : m(s) = 1, i(s) = j(s), \mathcal{B}_1(s), \mathcal{B}_{j(s)}(s), \mathcal{M}(s)\} \\
&= \{T_{S_1} =: s_{S_1,1}, s_{S_1,2}, \ldots, s_{S_1,|S_1(T)|}\} \\
S_2(T) &:= \{s \in [T_M, T] \cap \mathbb{N} : m(s) = 1, i(s) = 1, \mathcal{A}_{j(s)}(s), \mathcal{B}_1(s), \mathcal{M}(s)\} \\
&= \{T_{S_2} =: s_{S_2,1}, s_{S_2,2}, \ldots, s_{S_2,|S_2(T)|}\},
\end{aligned}$$

where $s_{S_m,k}$ implies the round when the event occurs k-th time for $m = 1, 2$, respectively.

Similarly, let us define a subset of rounds with any fixed $T > T_M$

$$\begin{aligned}
S_0(T) := \Big\{ & s \in [T_M, T] \cap \mathbb{N} : \{\mathcal{B}_1(s), \mathcal{M}^c(s)\} \cup \{\mathcal{B}_1(s), \mathcal{B}^c_{i(s)}, \mathcal{M}(s)\} \\
& \cup \{m(s) = 1, i(s) = 1, \mathcal{B}_1(s), \mathcal{A}^c_{j(s)}(s), \mathcal{M}(s)\} \\
& \cup \{m(s) \neq 1, i(s) = j(s), \mathcal{B}_1(s), \mathcal{A}^c_{m(s)}(s), \mathcal{B}_{j(s)}(s), \mathcal{M}(s)\} \Big\}
\end{aligned}$$

and a random variable

$$\begin{aligned}
T_S := T_M + \sum_{s=T_M+1}^{T} & \mathbb{1}[\mathcal{B}_1(s), \mathcal{M}^c(s)] + \mathbb{1}[\mathcal{B}_1(s), \mathcal{B}^c_{i(s)}, \mathcal{M}(s)] \\
& + \mathbb{1}[m(s) = 1, i(s) = 1, \mathcal{B}_1(s), \mathcal{A}^c_{j(s)}(s), \mathcal{M}(s)] \\
& + \mathbb{1}[m(s) \neq 1, i(s) = j(s), \mathcal{B}_1(s), \mathcal{B}^c_{m(s)}(s), \mathcal{B}_{j(s)}(s), \mathcal{M}(s)],
\end{aligned}$$

such that $T_S = |S_0(T)| + T_M$ holds.

First objective Here, we first aim to show that for $t \geq T_M$, it holds

$$1 = \mathbb{1}[t \in S_0(T)] + \mathbb{1}[t \in S_1(T)] + \mathbb{1}[t \in S_2(T)].$$

Since $\mathcal{B}_1(s)$ always holds for $s \geq T_M$, it holds that

$$
\begin{aligned}
1 &= \mathbb{1}[\mathcal{B}_1(s)] \\
&= \mathbb{1}[\mathcal{M}^c(s), \mathcal{B}_1(s)] + \mathbb{1}[\mathcal{M}(s), \mathcal{B}_1(s)] \\
&= \mathbb{1}[\mathcal{M}^c(s), \mathcal{B}_1(s)] + \mathbb{1}[\mathcal{M}(s), \mathcal{B}_1(s), m(s) = 1] + \mathbb{1}[\mathcal{M}(s), \mathcal{B}_1(s), m(s) \neq 1] \\
&= \mathbb{1}[\mathcal{M}^c(s), \mathcal{B}_1(s)] \\
&\quad + \mathbb{1}[\mathcal{M}(s), \mathcal{B}_1(s), m(s) = 1, i(s) = 1] + \mathbb{1}[\mathcal{M}(s), \mathcal{B}_1(s), m(s) = 1, i(s) = j(s)] \\
&\quad + \mathbb{1}[\mathcal{M}(s), \mathcal{B}_1(s), m(s) \neq 1, i(s) = m(s), \mathcal{B}_1(s), \mathcal{B}^c_{m(s)}(s)] \\
&\quad\quad + \mathbb{1}[\mathcal{M}(s), \mathcal{B}_1(s), m(s) \neq 1, i(s) = j(s), \mathcal{B}_1(s)\mathcal{A}^c_{m(s)}(s)] \qquad (25) \\
&= \mathbb{1}[\mathcal{M}^c(s), \mathcal{B}_1(s)] \\
&\quad + \mathbb{1}[\mathcal{M}(s), \mathcal{B}_1(s), m(s) = 1, i(s) = 1, \mathcal{A}^c_{j(s)}(s)] \\
&\quad\quad + \mathbb{1}[\mathcal{M}(s), \mathcal{B}_1(s), m(s) = 1, i(s) = 1, \mathcal{A}_{j(s)}(s)] \\
&\quad + \mathbb{1}[\mathcal{M}(s), \mathcal{B}_1(s), m(s) = 1, i(s) = j(s), \mathcal{B}^c_{j(s)}(s)] \\
&\quad\quad + \mathbb{1}[\mathcal{M}(s), \mathcal{B}_1(s), m(s) = 1, i(s) = j(s), \mathcal{B}_{j(s)}(s)] \\
&\quad + \mathbb{1}[\mathcal{M}(s), \mathcal{B}_1(s), m(s) \neq 1, i(s) = m(s), \mathcal{B}^c_{m(s)}(s)] \\
&\quad + \mathbb{1}[\mathcal{M}(s), \mathcal{B}_1(s), m(s) \neq 1, i(s) = j(s), \mathcal{A}^c_{m(s)}(s), \mathcal{B}^c_{j(s)}(s)] \\
&\quad\quad + \mathbb{1}[\mathcal{M}(s), \mathcal{B}_1(s), m(s) \neq 1, i(s) = j(s), \mathcal{A}^c_{m(s)}(s), \mathcal{B}_{j(s)}(s)] \qquad (26) \\
&= \mathbb{1}[\mathcal{M}^c(s), \mathcal{B}_1(s)] \\
&\quad + \mathbb{1}[\mathcal{M}(s), \mathcal{B}_1(s), m(s) = 1, i(s) = 1, \mathcal{A}^c_{j(s)}(s)] \\
&\quad\quad + \mathbb{1}[\mathcal{M}(s), \mathcal{B}_1(s), m(s) = 1, i(s) = 1, \mathcal{A}_{j(s)}(s)] \\
&\quad + \mathbb{1}[\mathcal{M}(s), \mathcal{B}_1(s), m(s) = 1, i(s) = j(s), \mathcal{B}_{j(s)}(s)] \\
&\quad + \mathbb{1}[\mathcal{M}(s), \mathcal{B}_1(s), \mathcal{B}^c_{i(s)}(s)] \\
&\quad\quad + \mathbb{1}[\mathcal{M}(s), \mathcal{B}_1(s), m(s) \neq 1, i(s) = j(s), \mathcal{A}^c_{m(s)}(s), \mathcal{B}_{j(s)}(s)] \\
&= \mathbb{1}[s \in S_0(T)] + \mathbb{1}[s \in S_1(T)] + \mathbb{1}[s \in S_2(T)],
\end{aligned}
$$

where (25) and (26) hold from

$$
\mathbb{1}[m(s) \neq 1, \mathcal{B}_1(s)] = \mathbb{1}[m(s) \neq 1, \mathcal{B}_1(s), \mathcal{B}^c_{m(s)}(s)] = \mathbb{1}[m(s) \neq 1, \mathcal{B}_1(s), \mathcal{A}^c_{m(s)}(s)]. \qquad (27)
$$

The last equality holds from

$$
\begin{aligned}
&\mathbb{1}[\mathcal{M}(s), \mathcal{B}_1(s), \mathcal{B}^c_{i(s)}(s)] \\
&\quad = \mathbb{1}[\mathcal{M}(s), \mathcal{B}_1(s), \mathcal{B}^c_{i(s)}(s), m(s) = 1] + \mathbb{1}[\mathcal{M}(s), \mathcal{B}_1(s), \mathcal{B}^c_{i(s)}(s), m(s) \neq 1] \\
&\quad = \mathbb{1}[\mathcal{M}(s), \mathcal{B}_1(s), \mathcal{B}^c_{i(s)}(s), m(s) = 1, i(s) = j(s)] \\
&\qquad + \mathbb{1}[\mathcal{M}(s), \mathcal{B}_1(s), \mathcal{B}^c_{i(s)}(s), m(s) \neq 1, i(s) = m(s)] \\
&\qquad\quad + \mathbb{1}[\mathcal{M}(s), \mathcal{B}_1(s), \mathcal{B}^c_{i(s)}(s), m(s) \neq 1, i(s) = j(s)] \\
&\quad = \mathbb{1}[\mathcal{M}(s), \mathcal{B}_1(s), m(s) = 1, i(s) = j(s), \mathcal{B}^c_{j(s)}(s)] \\
&\qquad + \mathbb{1}[\mathcal{M}(s), \mathcal{B}_1(s), m(s) \neq 1, i(s) = m(s), \mathcal{B}^c_{m(s)}(s)] \\
&\qquad\quad + \mathbb{1}[\mathcal{M}(s), \mathcal{B}_1(s), m(s) \neq 1, i(s) = j(s), \mathcal{A}^c_{m(s)}(s), \mathcal{B}^c_{j(s)}(s)], \qquad (28)
\end{aligned}
$$

where we used (27) in (28) again. This implies that if $T \geq T_M$, then $[T_M, T] \cap \mathbb{N} = S_0(T) \cup S_1(T) \cup S_2(T)$ holds. Note that if $s = T_M \geq KD$, there exists at least one arm $a \in [K]$ satisfying $N_a(s) \geq D$.

(1) If $N_1(s) \geq D$ Recall the definition $T_{S_1} = \inf S_1(T)$ and $T_{S_2} = \inf S_2(T)$, which implies the first round when the events in Lemmas 6 and 7 occur, respectively.

(1-i) $S_0(T)$ **is a subinterval** If $S_0(T)$ consists of consecutive natural numbers, i.e., the subinterval in $[T_M, T] \cap \mathbb{N}$, then $\min(T_{S_1}, T_{S_2}) \leq T_S + 1$ holds since we can only observe events in $S_1(T)$ or $S_2(T)$ for $s > T_S$.

(1-ii) $S_0(T)$ **is not a subinterval** If $S_0(T)$ is not a subinterval of $[T_M, T] \cap \mathbb{N}$, this directly implies that $\min(T_{S_1}, T_{S_2}) \leq T_S$ from $[T_M, T] \cap \mathbb{N} = S_0(T) \cup S_1(T) \cup S_2(T)$.

(1-iii) Summary What we have shown is $\min(T_{S_1}, T_{S_2}) \leq T_S + 1$. Let us consider the case $T_{S_1} < T_{S_2}$. From the definition of T_{S_1} where $i(t) = j(t)$, we have for $j(t) = j$ and $a \neq 1, j$ that

$$
\begin{aligned}
&(N_1(T_{S_1}) + N_j(T_{S_1}))d(\hat{\mu}_1(T_{S_1}), \hat{\mu}_{1,j}(T_{S_1})) \\
&\qquad \leq N_1(T_{S_1})d(\hat{\mu}_1(T_{S_1}), \hat{\mu}_{1,j}(T_{S_1})) + N_j(T_{S_1})d(\hat{\mu}_j(T_{S_1}), \hat{\mu}_1(T_{S_1})) \\
&\qquad = T_{S_1} f_j(\boldsymbol{w}^t; \hat{\mu}(T_{S_1})) \\
&\qquad \leq T_{S_1} f_a(\boldsymbol{w}^t; \hat{\mu}(T_{S_1})) \\
&\qquad \leq N_a(T_{S_1})d(\hat{\mu}_a(T_{S_1}), \hat{\mu}_1(T_{S_1})).
\end{aligned}
$$

From the assumption $N_1(T_{S_1}) \geq D$, it holds that

$$
\begin{aligned}
N_1(T_{S_1})d(\hat{\mu}_1(T_{S_1}), \hat{\mu}_{1,j}(T_{S_1})) &\geq N_1(T_{S_1})d(\hat{\mu}_1(T_{S_1}), \hat{\mu}_{1,j}(T_{S_1}))\frac{D_1}{\min_{i\in[K]} \underline{d}_i} \\
&\geq D_1 = \max_{i \neq 1} D_i.
\end{aligned}
$$

Therefore,

$$
\max_{i\in[K]} D_i < \min_{i \neq 1} N_a(T_{S_1})d(\hat{\mu}_a(T_{S_1}), \hat{\mu}_1(T_{S_1})). \qquad (29)
$$

Recall the definition $D_i = \sup_t \mathbb{1}[\mathcal{B}^c_i(t)]N_i(t)d(\hat{\mu}_i(t), \hat{\mu}_1(t))$. Thus (29) implies that $\mathcal{B}_a(t)$ holds for all $t \geq T_{S_1}$ and any $i \in [K]$, i.e., $T_B \leq T_{S_1} \leq T_S + 1$. When $T_{S_2} < T_{S_1}$ holds, $T_B \leq T_{S_2} \leq T_S + 1$ can be directly derived from Lemma 7.

(2) If $N_i(s) \geq D$ for $i \neq 1$ From (1), one can expect that T_B will be bounded at least if either $N_{j(s)}(s)$ or $N_1(s)$ satisfies the condition in (29) for any $s \leq T$.

(2-i) $j(s) = i$ holds for some $s \in S_1(T)$ In this case, we have for $a \neq 1, i$

$$N_1(s)d(\hat{\mu}_1(s), \hat{\mu}_{1,i}(s)) + N_i(s)d(\hat{\mu}_i(s), \hat{\mu}_{1,i}(s)) = sf_i < sf_a \leq N_a(s)d(\hat{\mu}_a(s), \hat{\mu}_1(s)),$$

where we denote $\hat{\mu}_{1,i}^{\boldsymbol{w}^s}(s)$ by $\hat{\mu}_{1,i}(s)$ for notational simplicity. From $N_i(s) \geq D$,

$$\max_{a \in [K]} D_a \leq N_i(s)d(\hat{\mu}_i(s), \hat{\mu}_{1,i}(s)) \leq \min_{a \neq 1} N_a(s)d(\hat{\mu}_a(s), \hat{\mu}_1(s)), \tag{30}$$

which implies $T_B \leq s$.

(2-ii) $j(s) \neq a$ holds for all $s \in S_1(T)$ Take arbitrary $t' \in (T_M, \infty) \cap \mathbb{N}$ and assume that there exists an arm $j' \neq 1$ and a round $s' \geq t'$ such that $\mathbb{1}[\mathcal{B}_{j'}^c(s')] = 1$ holds. Note that whenever $N_{j(s)}(s) \geq D$ holds, substituting $a = j(s)$ in (30) leads to the same inequality, which implies $T_B \leq s$.

(2-iii) Summary Therefore, for all $j \neq 1$, $\sum_{s \in S_1(T)} \mathbb{1}[j(s) = j] \leq D$ should hold since $\sum_{s \in S_1(T)} \mathbb{1}[j(s) = j] > D$ admits the existence of $s \in S_1(T)$ such that satisfies (30), which contradicts to the assumption of the existence of such s'. In other words, $\sum_{s \in S_1(T)} \mathbb{1}[j(s) = j] \leq D$ is a necessary condition to satisfy the assumption of the existence of j' and s' satisfying $\mathbb{1}[\mathcal{B}_{j'}^c(s')] = 1$. From the definition of $S_1(T)$, for any $s \in S_1(T)$, $N_{j(s)}(s+1) = N_{j(s)}(s) + 1$ holds. Hence, at worst, if $|S_1(T) \cap [T_M, t')| \geq (K-2)D$ holds at some round t', there exists $s \in S_1(T) \cap [T_M, t')$ such that $N_{j(s)}(s) \geq D$. Therefore, T_B is at most the round until $S_1(T)$ occur $(K-2)D$ times.

Similarly, if the event in $S_2(T)$ occurs D times at some round t'', then $N_1(t'') \geq D$ holds from the sampling rule. This implies that $B_i(s)$ holds for all $i \in [K]$ for $s \geq t''$ from (29), i.e., T_B is at most the round until $S_2(T)$ occur D times.

(3) Conclusion In summary, we have $[T_M, T] \cap \mathbb{N} = S_0(T) \cup S_1(T) \cup S_2(T)$ and there exists an arm i satisfying $N_i(t) \geq D$. If $N_1(s) \geq D$, then $T_B \leq T_S + 1$ holds. If $N_i(s) \geq D$ holds for $i \neq 1$, then T_B is at most the round $s_{S_1,(K-2)D}$ when the event in $S_1(T)$ occurs $(K-2)D$ times or $s_{S_2,D}$ when the event in $S_2(T)$ occur D times. Hence, we have

$$T_B \leq T_S + (K-2)D + D + 1,$$

where $T_S = T_M + |S_0(T)| = \max(T_C, KD) + |S_0(T)|$. Then, we have

$$\begin{aligned}
\mathbb{E}[T_B] &\leq \mathbb{E}[T_S] + (K-1)\mathbb{E}[D] + 1 \\
&\leq \mathbb{E}[T_C] + (2K-1)\mathbb{E}\left[\sup_{i \neq 1} \sup_{s \geq t} \mathbb{1}[B_i^c(s)] N_i(s) d(\hat{\mu}_i(s), \hat{\mu}_1(s))\right] \\
&\quad + \mathbb{E}\Bigg[\sum_{t=T_M}^{T} \mathbb{1}[\mathcal{M}^c(t)] + \mathbb{1}[m(t) = 1, i(t) = 1, \mathcal{B}_1(t), \mathcal{A}_{j(t)}^c(t), \mathcal{M}(t)] \\
&\qquad + \mathbb{1}[m(t) \neq 1, i(t) = j(t), \mathcal{B}_1(t), \mathcal{A}_{m(t)}^c(t), \mathcal{B}_{j(t)}(t), \mathcal{M}(t)] \\
&\qquad + \mathbb{1}[\mathcal{B}_1(t), \mathcal{B}_{i(t)}^c(t), \mathcal{M}(t)]\Bigg] + 1.
\end{aligned}$$

Then, the following five lemmas conclude the proofs. ■

Lemma 8 *For a bounded region of parameters $R \subset \mathbb{R}$, it holds that for arbitrary $\mu' \in R$ and $i \in [K]$*

$$\mathbb{E}\left[\sup_{n\in\mathbb{N},\mu'\in R} \mathbb{1}[|\hat{\mu}_{i,n} - \mu_i| \geq \epsilon] n d(\hat{\mu}_{i,n}, \mu')\right] = \mathcal{O}\left(d_\epsilon^{-1}\right),$$

where $\hat{\mu}_{i,n}$ is the empirical mean reward of the arm i when it is played n times.

Here, note that $\hat{\mu}_{i,n}$ is different from $\hat{\mu}_{a,b}(t)$ that denotes the weighted average of their empirical mean. Lemma 8 provides the finiteness of the expectation of D_i for any $i \in [K]$.

Lemma 9 *For the finite number of arms K and any $T \in \mathbb{N}$, it holds that*

$$\mathbb{E}\left[\sum_{t=1}^{T} \mathbb{1}\left[m(t) = 1, i(t) = 1, \mathcal{B}_1(t), \mathcal{A}^c_{j(t)}(t), \mathcal{M}(t)\right]\right] \leq \mathcal{O}\left(K d_\epsilon^{-1}\right),$$
$$\mathbb{E}\left[\sum_{t=1}^{T} \mathbb{1}\left[i(t) = j(t), \mathcal{A}^c_{m(t)}(t), \mathcal{B}_{j(t)}(t), \mathcal{M}(t)\right]\right] \leq \mathcal{O}\left(K^2 d_\epsilon^{-1}\right).$$

Lemma 10 *For the finite number of arms K and any $T \in \mathbb{N}$, it holds that*

$$\mathbb{E}\left[\sum_{t=1}^{T} \mathbb{1}\left[\mathcal{B}^c_{i(t)}(t), \mathcal{M}(t)\right]\right] \leq \mathcal{O}\left(K d_\epsilon^{-1}\right).$$

The proofs of Lemmas 8–10 are provided in Section E.2.

Lemma 11 *For the finite number of arms K and any $T \in \mathbb{N}$, it holds that*

$$\mathbb{E}\left[\sum_{t=1}^{T} \mathbb{1}[\mathcal{M}^c(t)]\right] \leq \mathcal{O}\left(K^2 d_\epsilon^{-2}\right).$$

The proof of Lemma 11 is given in Section E.3.

Lemma 12 *Under Algorithm 1, it holds for any $\epsilon \in \left(0, \frac{\mu_1 - \mu_2}{2}\right)$ that*

$$\mathbb{E}[T_C] \leq C(\pi_{\mathrm{j}}, \boldsymbol{\mu}, \epsilon) + 4 d_\epsilon^{-3},$$

where $C(\pi_{\mathrm{j}}, \boldsymbol{\mu}, \epsilon)$ specified in Lemma 15.

The proof of Lemma 12 is given in Section E.4, where we adapt the analysis in Korda et al. (2013) to our problem.

E.1. Proofs of technical lemmas for Theorem 2: Sufficient conditions for the convergence of estimates

Here, we provide the proof of Lemmas 6 and 7.

Proof of Lemma 6 Since $i(t) = j$ implies

$$d(\hat{\mu}_j(t), \hat{\mu}_{1,j}(t)) \geq d(\hat{\mu}_1(t), \hat{\mu}_{1,j}(t)),$$

we have

$$d(\hat{\mu}_j(t), \hat{\mu}_{1,j}(t)) \geq \underline{d}_j,$$

from the definition of $\underline{d}_j$ in (14).

Then, we have

$$\begin{aligned} tf_j(\boldsymbol{w}^t, \hat{\boldsymbol{\mu}}(t)) &= N_1(t)d(\hat{\mu}_1(t), \hat{\mu}_{1,j}(t)) + N_j(t)d(\hat{\mu}_j(t), \hat{\mu}_{1,j}(t)) \\ &\geq N_j(t)\underline{d}_j > D_1 \end{aligned}$$

On the other hand, if $|\hat{\mu}_1(t) - \mu_1| \geq \epsilon$ and $|\hat{\mu}_j(t) - \mu_j| \leq \epsilon$, then

$$tf_j(\boldsymbol{w}^t, \hat{\boldsymbol{\mu}}(t)) \leq N_1(t)d(\hat{\mu}_1(t), \hat{\mu}_j(t)) \leq D_1$$

by the definition of $D_1 = \sup_{i \neq 1} D_i$. Therefore, $|\hat{\mu}_1(t) - \mu_1| \geq \epsilon$ cannot hold.

Under $|\hat{\mu}_1(t) - \mu_1| \leq \epsilon$ and $|\hat{\mu}_j(t) - \mu_j| \leq \epsilon$, we see that

$$\begin{aligned} N_j(t)\underline{d}_j \leq tf_j(\boldsymbol{w}^t, \hat{\boldsymbol{\mu}}(t)) &\leq N_1(t)d(\hat{\mu}_1(t), \hat{\mu}_j(t)) \\ &\leq N_1(t)d(\mu_1 + \epsilon, \mu_j - \epsilon), \end{aligned}$$

which completes the proof. ■

Proof of Lemma 7 Since $j(t) = \arg\min_{i \neq m(t)} tf_i(\boldsymbol{w}^t, \hat{\boldsymbol{\mu}}(t))$ and $i(t) = 1$, it holds for all $i \neq 1$ that

$$tf_i(\boldsymbol{w}^t, \hat{\boldsymbol{\mu}}(t)) \geq tf_{j(t)}(\boldsymbol{w}^t, \hat{\boldsymbol{\mu}}(t))$$

and

$$d(\hat{\mu}_1(t), \hat{\mu}_{1,j(t)}(t)) \geq d(\hat{\mu}_{j(t)}(t), \hat{\mu}_{1,j(t)}(t)).$$

Then, we can use the same argument as Lemma 6 by exchanging the role of 1 and j. ■

E.2. Proofs of technical lemmas for Theorem 2: Boundedness of the number of rounds where estimates do not converge

Here, we provide the proof of Lemmas 8–10. Firstly, to prove Lemma 8, we require the lemma below, whose proof is postponed to Section F.1.

Lemma 13 *Let $R \subset \mathbb{R}$ be a bounded region of parameters and fix arbitrary μ_0. Then, there exists $a, b \geq 0$ such that*

$$d(\mu, \mu') \leq ad(\mu, \mu_0) + b$$

for arbitrary $\mu \in \mathbb{R}$ and $\mu' \in R$.

Proof of Lemma 8 Let $P(z) := \mathbb{P}[d(\hat{\mu}_{i,n}, \mu_i) \geq z]$. Then, by Chernoff bound, we have $P(z) \leq 2e^{-nz}$. Therefore,

$$\begin{aligned}
\mathbb{E}\left[\mathbb{1}[|\hat{\mu}_{i,n} - \mu_i| \geq \epsilon] \sup_{\mu' \in R} d(\hat{\mu}_{i,n}, \mu')\right] &\leq \mathbb{E}[\mathbb{1}[|\hat{\mu}_{i,n} - \mu_i| \geq \epsilon](ad(\hat{\mu}_{i,n}, \mu_i) + b)] \\
&\leq 2be^{-nd_\epsilon} + a\int_{d_\epsilon}^{\infty} z\mathrm{d}(-P(z)) \\
&= 2be^{-nd_\epsilon} + a\left(-[zP(z)]_{d_\epsilon}^{\infty} + \int_{d_\epsilon}^{\infty} zP(z)\mathrm{d}z\right) \\
&\leq 2be^{-nd_\epsilon} + 2ad_\epsilon e^{-nd_\epsilon} + a\int_{d_\epsilon}^{\infty} zP(z)\mathrm{d}z \\
&\leq 2be^{-nd_\epsilon} + 2ad_\epsilon e^{-nd_\epsilon} + 2a\left[-\frac{ze^{-nz}}{n} - \frac{e^{-nz}}{n^2}\right]_{d_\epsilon}^{\infty} \\
&\leq 2\left(b + a\left(d_\epsilon + \frac{d_\epsilon}{n} + \frac{1}{n^2}\right)\right)e^{-nd_\epsilon},
\end{aligned}$$

where $d_\epsilon := \min_{i \in [K]}\{d(\mu_i - \epsilon, \mu_i), d(\mu_i + \epsilon, \mu_i)\}$ and the first inequality holds from Lemma 13. Since this quality decays exponentially in n, it is straightforward that

$$\begin{aligned}
\mathbb{E}\left[\sup_{n \in \mathbb{N}, \mu' \in R} \mathbb{1}[|\hat{\mu}_{i,n} - \mu_i| \geq \epsilon] n d(\hat{\mu}_{i,n}, \mu')\right] &\leq \sum_{n=1}^{\infty} \mathbb{E}\left[\mathbb{1}[|\hat{\mu}_{i,n} - \mu_i| \geq \epsilon] \sup_{\mu' \in A} d(\hat{\mu}_{i,n}, \mu')\right] \\
&= \mathcal{O}(d_\epsilon^{-1}).
\end{aligned}$$

■

Proof of Lemma 9 For $j(t) = j$, we first consider

$$D_j = \sup_t \left\{\mathbb{1}[|\hat{\mu}_j(t) - \mu_i| \geq \epsilon] N_j(t) d(\hat{\mu}_j(t), \hat{\mu}_1(t))\right\}.$$

Note that on $\mathcal{B}_1(t)$, $\hat{\mu}_1(t) \in [\mu_1 - \epsilon, \mu_1 + \epsilon]$ is bounded so that we can apply Lemmas 8 and 13. We first show the existence of a bounded constant $c_j^* \in \mathbb{R}_+$ such that

$$N_1(t) \leq c_j^* D_j,$$

where

$$c_j^* = \min\left(c_j, \frac{x_j'}{d_\zeta}\right)$$

for constants c_j, x_j' and d_ζ that depend on models.

(1) When $\hat{\mu}_j(t) \not\approx \hat{\mu}_{m(t)}(t)$ From their definitions, we have

$$0 \leq N_j(t) d(\hat{\mu}_i(t), \hat{\mu}_{1,j}(t)(t)) \leq N_j(t) d(\hat{\mu}_j(t), \hat{\mu}_1(t)) \leq D_i$$

and

$$\begin{aligned}
N_1(t) d(\hat{\mu}_1(t), \hat{\mu}_{1,j}(t)) &\leq N_1(t) d(\hat{\mu}_1(t), \hat{\mu}_{1,j}(t)) + N_j(t) d(\hat{\mu}_i(t), \hat{\mu}_{1,j}(t)) \\
&= t g(\boldsymbol{w}^t; \hat{\boldsymbol{\mu}}(t)).
\end{aligned}$$

Let us consider

$$\psi(x;t) = xd(\hat{\mu}_{m(t)}(t), \hat{\mu}_{m(t),j}(x;t)) + d(\hat{\mu}_j(t), \hat{\mu}_{m(t),j}(x;t)),$$

where $\hat{\mu}_{a,b}(x;t) = \frac{x\hat{\mu}_a(t)+\hat{\mu}_b(t)}{x+1}$. One can see that $\psi(x;t)$ is strictly increasing with respect to x since $\psi'(x;t) = d(\hat{\mu}_{m(t)}(t), \hat{\mu}_{m(t),j}(x;t)) > 0$ and it tends to $d(\hat{\mu}_j(t), \hat{\mu}_{m(t)}(t))$ when x goes to infinity (Garivier and Kaufmann, 2016). Then, under the condition $\{m(t) = 1, j(t) = j\}$, it holds that

$$\begin{aligned} tg(\boldsymbol{w}^t; \hat{\boldsymbol{\mu}}(t)) = N_j(t)\psi\left(\frac{N_1(t)}{N_j(t)}; t\right) &\leq N_j(t)d(\hat{\mu}_j(t), \hat{\mu}_1(t)) \\ &\leq D_j. \end{aligned}$$

Therefore,

$$N_1(t) \leq \frac{1}{d(\hat{\mu}_1(t), \hat{\mu}_{1,j}(t))} D_j.$$

Note that there exists a constant c_j such that $\frac{1}{d(\hat{\mu}_1(t),\hat{\mu}_{1,j}(t))} \leq c_j < \infty$ when $\hat{\mu}_a(t) \not\approx \hat{\mu}_{m(t)}(t)$, which shows the existence of c_j^*.

(2) When $\hat{\mu}_j(t) \approx \hat{\mu}_{m(t)}(t)$ Here, $i(t) = 1$ implies that

$$d\left(\hat{\mu}_1(t), \hat{\mu}_{1,j}^{\boldsymbol{w}^t}(t)\right) \geq d\left(\hat{\mu}_j(t), \hat{\mu}_{1,j}^{\boldsymbol{w}^t}(t)\right). \tag{31}$$

Note that as $\frac{w_1(t)}{w_j(t)}$ increases, RHS of (31) decreases and LHS of (31) increases simultaneously. Therefore,

$$\forall t \in \mathbb{N},\ \exists\, x_{j,t}^* \in \mathbb{R}_+ \text{ s.t. } \frac{w_1(t)}{w_j(t)} = x_{j,t}^* \Leftrightarrow d(\hat{\mu}_1(t), \hat{\mu}_{1,j}^{\boldsymbol{w}^t}(t)) = d(\hat{\mu}_j(t), \hat{\mu}_{1,j}^{\boldsymbol{w}^t}(t)).$$

Note that $x_{j,t}^*$ depends on the distribution of reward and history H_t until round t, e.g., $\forall t \in \mathbb{N}$, $x_{j,t}^* = 1$ for the Gaussian distribution. Since $\hat{\mu}_1(t)$ is bounded under $\{\mathcal{B}_1(t)\}$ and $\hat{\mu}_j(t) \in (\mu_j + \epsilon, \hat{\mu}_1(t)] \subset (\mu_j + \epsilon, \mu_1 + \epsilon]$ holds under $\{\mathcal{B}_1(t), \mathcal{A}_j^c(t), m(t) = 1\}$, there exists $x_j' \in \mathbb{R}_+$ such that for any $t \in \mathbb{N}$

$$N_1(t) > x_j' N_j(t) \implies d(\hat{\mu}_1(t), \hat{\mu}_{1,j}(t)) < d(\hat{\mu}_j(t), \hat{\mu}_{1,j}(t)), \text{ i.e., } i(t) = j.$$

Let consider a bounded region $R = [\mu_1 - \epsilon, \mu_1 + \epsilon] \subset \mathbb{R}$ and a random variable

$$D_j = \sup_{t\in\mathbb{N}} \sup_{\mu'\in A} \left\{ \mathbb{1}[|\hat{\mu}_j(t) - \mu_j| \geq \epsilon] N_j(t) d(\hat{\mu}_j(t), \mu') \right\}, \quad j \in [K] \setminus \{1\}.$$

Since $m(t) = 1$ holds under the condition, we have

$$\sup_{\mu'\in A} d(\hat{\mu}_j(t), \mu') = \max\{d(\hat{\mu}_j(t), \mu_1 - \epsilon), d(\hat{\mu}_j(t), \mu_1 + \epsilon)\}$$

and $\hat{\mu}_1(t) > \hat{\mu}_j(t)$. Let $\zeta(\epsilon) \in A$ be a point such that $d(\zeta, \mu_1 - \epsilon) = d(\zeta, \mu_1 + \epsilon) = d_\zeta$. Then, it holds that

$$\sup_{\mu'\in A} d(\hat{\mu}_j(t), \mu') > d_\zeta.$$

Note that d_ζ and x_j' only depend on the models. Therefore, there exists a constant $c_j^* \in \mathbb{R}_+$ such that

$$N_1(t) \leq \frac{x_j'}{d_\zeta} D_j \leq c_j^* D_j.$$

(3) Conclusion From Lemma 8, we obtain

$$\begin{aligned}\mathbb{E}\left[\sum_{i\in[K]\setminus\{1\}}\sum_{t=1}^{\tau}\mathbb{1}\left[m(t)=1,i(t)=1,\mathcal{B}_1(t),j(t)=i,\mathcal{A}^c_{j(t)}(t),\mathcal{M}(t)\right]\right]&\\ \leq \mathbb{E}\left[\sum_{i\in[K]\setminus\{1\}}\sum_{t=1}^{\infty}\mathbb{1}[i(t)=1,N_1(t)\leq c_i^*D_i]\right]&\\ \leq \sum_{i\in[K]\setminus\{1\}}c_j^*\mathbb{E}[D_j]\leq\mathcal{O}(Kd_\epsilon^{-1}),&\end{aligned}$$

which concludes the first case.

Similarly, the second case can be bounded by considering $R_j=[\mu_j-\epsilon,\mu_j+\epsilon]$ and

$$D_{m(t),j}=\sup_n\sup_{\mu'\in R_j}\{\mathbb{1}[|\hat{\mu}_{m(t)}(n)-\mu_{m(t)}|\geq\epsilon]nd(\hat{\mu}_{m(t)}(n),\mu')\}$$

for every $m(t)\in[K]$ and $j\in[K]\setminus\{m(t)\}$. Since $\hat{\mu}_j(t)\in R_j$ holds under $\{B_j(t)\}$, we can apply Lemmas 8 and 13 by exchanging the role of $m(t)$ and j, which concludes the proof. ■

Proof of Lemma 10 From the Chernoff bound, it holds for any arm $i\in[K]$ that

$$\mathbb{P}[|\hat{\mu}_i(t)-\mu_i|\geq\epsilon|N_i(t)=n]\leq 2e^{-nd_\epsilon},\tag{32}$$

where d_ϵ is defined in (10). One can rewrite the expectation as

$$\begin{aligned}\mathbb{E}\left[\sum_{t=1}^{T}\mathbb{1}\left[\mathcal{B}^c_{i(t)}(t),\mathcal{M}(t)\right]\right]&=\mathbb{E}\left[\sum_{i=1}^{K}\sum_{t=1}^{T}\sum_{n=1}^{\infty}\mathbb{1}\left[i(t)=i,\mathcal{B}^c_{i(t)}(t),\mathcal{M}(t),N_{i(t)}(t)=n\right]\right].\\ &=\mathbb{E}\left[\sum_{i=1}^{K}\sum_{t=1}^{T}\sum_{n=1}^{\infty}\mathbb{1}\left[i(t)=i,\mathcal{B}^c_{i}(t),\mathcal{M}(t),N_{i}(t)=n\right]\right]\end{aligned}$$

For every arm $i\in[K]$, an event $\{i(t)=i,N_i(t)=n\}$ could happen at most once for any $n\in\mathbb{N}$. Therefore, by applying (32), one has

$$\mathbb{E}\left[\sum_{t=1}^{T}\mathbb{1}\left[\mathcal{B}^c_{i(t)}(t),\mathcal{M}(t)\right]\right]\leq\sum_{i=1}^{K}\sum_{n=1}^{\infty}2e^{-nd_\epsilon}\leq\mathcal{O}(Kd_\epsilon^{-1}),$$

which concludes the proof. ■

E.3. Proof of technical lemma for Theorem 2: An upper bound on the number of rounds where TE occurs

Here, we provide the proof of Lemma 11, which shows that the expected number of rounds where Thompson samples and the empirical mean estimates disagree is finite. Before beginning the proof, we present the posterior concentration result when we employ the Jeffreys prior in the SPEF.

Lemma 14 (Theorem 4 in Korda et al. (2013)) *For the Jeffreys prior and d_ϵ defined in (10), there exists constants $C_{1,a} = C_1(\theta_a, A) > 0$, $C_{2,a} = C_2(\theta_a, A, \epsilon) > 0$ and $N(\theta_a, A)$ such that for any $N_a(t) \geq N(\theta_a, A)$,*

$$\mathbb{1}[\mathcal{B}_a(t)]\mathbb{P}[\tilde{\mathcal{B}}_a^c(t)|X_{a,N_a(t)}] \leq 2C_{1,a}N_a(t)e^{-(N_a(t)-1)(1-\epsilon C_{2,a})d_\epsilon}$$

whenever ϵ is such that $1 - \epsilon C_{2,a}(\epsilon) > 0$. Note that A is a convex function in (1).

Proof of Lemma 11 Let us define $L(\theta) := \frac{1}{2}\min(\sup_y p(y|\theta), 1)$ and an event

$$\tilde{E}_a(t) = \left(\exists 1 \leq s' \leq N_a(t) : p(x_{a,s'}|\theta_a) \geq L(\theta_a), \left| \frac{\sum_{s=1,s\neq s'}^{N_a(t)} x_{a,s}}{N_a(t)-1} - \mu_a \right| \leq \epsilon \right).$$

Consider

$$\begin{aligned}\sum_{t=1}^{T} \mathbb{1}[\mathcal{M}^c(t)] &= \sum_{t=1}^{T}\sum_{i\in[K]} \mathbb{1}[i(t)=i, \mathcal{M}^c(t)] \\ &= \sum_{t=1}^{T}\sum_{i\in[K]} \mathbb{1}[i(t)=i, \tilde{E}_a^c(t), \mathcal{M}^c(t)] + \mathbb{1}[i(t)=i, \tilde{E}_a(t), \mathcal{M}^c(t)]\end{aligned}$$

It is shown by Korda et al. (2013) that

$$\begin{aligned}\mathbb{E}\left[\sum_{t=1}^{T} \mathbb{1}[i(t)=i, \tilde{E}_i^c(t), \mathcal{M}^c(t)]\right] &\leq \sum_{t=1}^{\infty} \mathbb{P}(p(x_{i,1}|\theta_a) \leq L(\theta_a))^t + \sum_{t=1}^{\infty} 2te^{-(t-1)d_\epsilon} \\ &\leq \mathcal{O}\left(d_\epsilon^{-2}\right). \end{aligned} \tag{33}$$

Then, consider

$$\begin{aligned}\sum_{t=1}^{T} \mathbb{1}[i(t)=i, \tilde{E}_i(t), \mathcal{M}^c(t)] = \sum_{t=1}^{T} \Bigg(&\mathbb{1}[i(t)=i, \tilde{\mathcal{B}}_i(t), \tilde{E}_i(t), \mathcal{M}^c(t)] \\ &+ \mathbb{1}[i(t)=i, \tilde{\mathcal{B}}_i^c(t), \tilde{E}_i(t), \mathcal{M}^c(t)] \Bigg).\end{aligned}$$

On $\tilde{E}_i(t)$, the following holds for a constant $N(\theta_i, A)$ from Lemma 14.

$$\begin{aligned}&\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i\in[K]} \mathbb{1}[i(t)=i, \tilde{\mathcal{B}}_i^c(t), \tilde{E}_i(t), \mathcal{M}^c(t)]\right] \\ &\quad\leq \sum_{i\in[K]} N(\theta_i, A) + \sum_{i\in[K]} \sum_{\substack{t:i(t)=i \\ N_a(t)\geq N(\theta_i,A)}}^{T} 2C_{1,i}e^{-(N_i(t)-1)(1-\epsilon C_{2,i})d_\epsilon + \log(N_i(t))} \\ &\quad\leq \sum_{i\in[K]} N(\theta_i, A) + \sum_{i\in[K]} \sum_{n=N(\theta_i,A)}^{\infty} 2C_{1,i}ne^{-(n-1)(1-\epsilon C_{2,i})d_\epsilon} \\ &\quad\leq \mathcal{O}\left(Kd_\epsilon^{-2}\right),\end{aligned}$$

where the second inequality holds since $N_i(t)$ increases when $\{i(t) = i\}$ happens.

Finally, we will show that

$$\sum_{t=1}^{T} \sum_{i\in[K]} \mathbb{1}[i(t) = i, \tilde{\mathcal{B}}_i(t), \tilde{E}_i(t), \mathcal{M}^c(t)] \leq \mathcal{O}\left(K^2 d_\epsilon^{-2}\right).$$

On $\mathcal{M}^c(t)$, $i(t) \in \{m(t), \tilde{m}(t)\}$ holds so that

$$\begin{aligned}\sum_{t=1}^{T} \sum_{i\in[K]} \mathbb{1}[i(t) = i, \tilde{\mathcal{B}}_i(t), \tilde{E}_i(t), \mathcal{M}^c(t)] \leq &\sum_{t=1}^{T} \sum_{i\in[K]} \mathbb{1}[i(t) = m(t) = i, \tilde{\mathcal{B}}_i(t), \tilde{E}_i(t), \mathcal{M}^c(t)] \\ &+ \sum_{t=1}^{T} \sum_{i\in[K]} \mathbb{1}[i(t) = \tilde{m}(t) = i, \tilde{\mathcal{B}}_i(t), \tilde{E}_i(t), \mathcal{M}^c(t)].\end{aligned}$$

Let us define $N_A = \max_{a\in[K]} N(\theta_a, A)$. For any $i \in [K]$, we have

$$\begin{aligned}&\sum_{t=1}^{T} \mathbb{1}[i(t) = m(t) = i, \tilde{\mathcal{B}}_i(t), \tilde{E}_i(t), \mathcal{M}^c(t)] \\ &\qquad\leq N_A + \sum_{t=1}^{T} \mathbb{1}[i(t) = m(t) = i, \tilde{\mathcal{B}}_i(t), \tilde{E}_i(t), \mathcal{M}^c(t), N_i(t) \geq N_A]\end{aligned}$$

and

$$\begin{aligned}&\sum_{t=1}^{T} \mathbb{1}[i(t) = \tilde{m}(t) = i, \tilde{\mathcal{B}}_i(t), \tilde{E}_i(t), \mathcal{M}^c(t)] \\ &\qquad\leq N_A + \sum_{t=1}^{T} \mathbb{1}[i(t) = \tilde{m}(t) = i, \tilde{\mathcal{B}}_i(t), \tilde{E}_i(t), \mathcal{M}^c(t), N_i(t) \geq N_A].\end{aligned}$$

Consider

$$\begin{aligned}&\mathbb{1}[i(t) = m(t) = i, \tilde{\mathcal{B}}_i(t), \tilde{E}_i(t), \mathcal{M}^c(t), N_i(t) \geq N_A] = \\ &\qquad\sum_{j\in[K]\setminus\{i\}} \underbrace{\mathbb{1}[i(t) = m(t) = i, \tilde{\mathcal{B}}_i(t), \tilde{E}_i(t), \mathcal{M}^c(t), N_i(t) \geq N_A, \tilde{m}(t) = j, \tilde{E}_j(t)]}_{(\divideontimes)} \\ &\qquad\qquad + \underbrace{\mathbb{1}[i(t) = m(t) = i, \tilde{\mathcal{B}}_i(t), \tilde{E}_i(t), \mathcal{M}^c(t), N_i(t) \geq N_A, \tilde{m}(t) = j, \tilde{E}_j^c(t)]}_{(\star)}.\end{aligned}$$

Similarly to (33), it holds that $\mathbb{E}\left[\sum_t (\star)\right] \leq \mathcal{O}\left(d_\epsilon^{-2}\right)$. On $\mathcal{M}^c(t)$, $\{i(t) = m(t)\}$ implies that $\{N_{m(t)}(t) \leq N_{\tilde{m}(t)}(t)\}$, i.e., $N_j(t) \geq N_i(t) \geq N_A$ so that one can apply Lemma 14. Hence,

$$\begin{aligned}\sum_t \mathbb{E}[(\divideontimes)] \leq \mathcal{O}(d_\epsilon^{-2}) + \sum_t \mathbb{E}\Big[&\mathbb{1}[i(t) = m(t) = i, \tilde{\mathcal{B}}_i(t), \tilde{E}_i(t)] \\ &\cdot \mathbb{1}[\mathcal{M}^c(t), N_i(t) \geq N_A, \tilde{m}(t) = j, \tilde{E}_j(t), \tilde{\mathcal{B}}_j(t)]\Big].\end{aligned}$$

From its definition, on $\tilde{E}_i(t)$, the empirical mean reward of arm i is well concentrated around its true mean. Thus,

$$m(t) = i, \tilde{E}_i(t), \tilde{E}_j(t) \implies i > j.$$

However, on $\{\tilde{\mathcal{B}}_i(t), \tilde{\mathcal{B}}_j(t), \tilde{m}(t) = j\}$, $i < j$ holds, which is a contradiction. Therefore,

$$\mathbb{1}[i(t) = m(t) = i, \tilde{\mathcal{B}}_i(t), \tilde{E}_i(t), \mathcal{M}^c(t), N_i(t) \geq N_A, \tilde{m}(t) = j, \tilde{E}_j(t), \tilde{\mathcal{B}}_j(t)] = 0,$$

which leads to

$$\mathbb{E}\left[\sum_{t=1}^{T} \mathbb{1}[\mathcal{M}^c(t)]\right] = \mathcal{O}\left(K^2 d_\epsilon^{-2}\right). \qquad \blacksquare$$

E.4. Proof of technical lemma for Theorem 2: Analysis with TS

Here, we provide the proof of Lemma 12.

Proof of Lemma 12 Let us define an event

$$\mathcal{C}(t) := \bigcup_{s=t}^{\infty} \{\mathcal{B}_1^c(s)\}$$

so that $\mathcal{C}^c(t) = \bigcap_{s=t}^{\infty} \{\mathcal{B}_1(s)\}$ implies only $\mathcal{B}_1(s)$ occurs for $s \geq t$, meaning that $\mathcal{C}(t) \Leftrightarrow \{T_C \geq t\}$. Therefore.

$$\begin{aligned}\mathbb{E}[T_C] = \sum_{s=1}^{\infty} \mathbb{P}[T_C \geq s] &= \sum_{s=1}^{\infty} \mathbb{P}[\mathcal{C}(s)] \\ &= \sum_{s=1}^{\infty} \mathbb{P}[\mathcal{C}(s), N_1(s) \leq \sqrt{s}] + \mathbb{P}[\mathcal{C}(s), N_1(s) \geq \sqrt{s}].\end{aligned}$$

From the Chernoff bound, we can derive the upper bound of the second term as

$$\begin{aligned}\sum_{s=1}^{\infty} \mathbb{P}[\mathcal{C}(s), N_1(s) \geq \sqrt{s}] &\leq \sum_{s=1}^{\infty} \sum_{n=\sqrt{s}}^{\infty} \mathbb{P}[|\hat{\mu}_{1,n} - \mu_1| \geq \epsilon] \\ &\leq \sum_{s=1}^{\infty} \sum_{n=\sqrt{s}}^{\infty} 2e^{-nd_\epsilon} \\ &\leq \sum_{s=1}^{\infty} \frac{2}{d_\epsilon} e^{-\sqrt{s}d_\epsilon} \\ &\leq \frac{2}{d_\epsilon} \int_0^{\infty} e^{-\sqrt{s}d_\epsilon} \mathrm{d}s = \frac{2}{d_\epsilon} \int_0^{\infty} 2xe^{-d_\epsilon x} \mathrm{d}x \\ &= 4d_\epsilon^{-3}.\end{aligned}$$

Then, the Lemma 15 below concludes the proof. $\blacksquare$

Lemma 15 *For the finite number of arms $K < \infty$, and $\epsilon \in \left(0, \frac{\mu_1 - \mu_2}{2}\right)$, there exists some constants $C(\pi_{\mathrm{j}}, \boldsymbol{\mu}, \epsilon) < \infty$ such that*

$$\sum_{s=1}^{\infty} \mathbb{P}[\mathcal{C}(s), N_1(s) \leq \sqrt{s}] \leq C(\pi_{\mathrm{j}}, \boldsymbol{\mu}, \epsilon).$$

The proof of Lemma 15 is given in F.2.

Appendix F. Proofs of additional lemmas

In this section, we provide proofs of additional lemmas that prove the lemmas for proving Theorem 2.

F.1. Proof of technical lemma for Lemma 8: Lemma 13

Proof of Lemma 13 It holds from the expression of KL divergence that

$$\begin{aligned} d(\mu, \mu') - d(\mu, \mu_0) &= A(\theta(\mu_0)) - A(\theta(\mu')) + (\theta(\mu') - \theta(\mu_0))\mu \\ &\leq A(\theta(\mu_0)) - \inf_{x \in R} A(\theta(x)) + |\mu| \sup_{x \in A} |\theta(x) - \theta(\mu_0)|. \end{aligned}$$

Since $d(\mu, \mu_0)$ is convex with respect to μ, there exist constant $a', b' \geq 0$ such that $|\mu| \leq a' d(\mu, \mu_0) + b'$. Letting $a := 1 + a' \sup_{x \in A} |\theta(x) - \theta(\mu_0)|$ and $b := b' \sup_{x \in A} |\theta(x) - \theta(\mu_0)| + A(\theta(\mu_0)) - \inf_{x \in A} A(\theta(x))$ concludes the proof. ■

F.2. Proof of technical lemma for Lemma 12: Lemma 15

Here, we present the proof of Lemma 15, where we adapt the proof techniques considered in Kaufmann et al. (2012) and Korda et al. (2013). Before beginning, we introduce some results in Korda et al. (2013).

The following Lemma shows the concentration inequality when an arm is played sufficiently.

Lemma 16 (Lemma 10 in Korda et al. (2013)) *For every $a \in [K]$ and $\epsilon > 0$, there exist constants $C'_a = C'(\mu_a, \epsilon, A)$ and N such that for $t \geq N_K$,*

$$\begin{aligned} \mathbb{P}[\exists s \leq t, \exists a \neq 1 : |\hat{\mu}_a(s) - \mu_a| \geq \epsilon, N_a(s) > C'_a \log t] &\leq \frac{2(K-1)}{t^3} \\ \mathbb{P}[\exists s \leq t, \exists a \neq 1 : |\tilde{\mu}_a(s) - \mu_a| \geq \epsilon, N_a(s) > C'_a \log t] &\leq \frac{4(K-1)}{t^3}. \end{aligned}$$

Note that we use the upper bound with the order of $\mathcal{O}(t^{-3})$ differently from the original lemma whose order is $\mathcal{O}(t^{-2})$. This can be done simply by changing the constant term with a multiplication of $3/2$.

The following lemma holds for the SPEF.

Lemma 17 (Lemma 9 in Korda et al. (2013)) *There exists a constant $C = C(\pi_{\mathrm{j}}) < 1$, such that for every (random) interval I and for every positive function ℓ, one has*

$$\mathbb{P}[\forall s \in I, \tilde{\mu}_1(s) \leq \mu_2 + \epsilon, |I| \geq \ell(t)] \leq C^{\ell(t)}.$$

Proof of Lemma 15 Let τ_n denote n-th time when arm 1 is played and $\xi_n = (\tau_{n+1} - 1) - \tau_n$ be the time between $n+1$-th and n-th time of arm 1 playing. From the definition, it holds that

$$\mathbb{P}[N_1(t) \leq \sqrt{t}, \mathcal{C}(t)] \leq \sum_{n=0}^{\lfloor \sqrt{t} \rfloor} \mathbb{P}[\xi_n \geq \sqrt{t} - 1, \mathcal{C}(t)].$$

For simplicity, let us define an event

$$G_n := \{\xi_n \geq \sqrt{t} - 1, \mathcal{C}(t)\} = \{\xi_n \geq \sqrt{t} - 1, \{\exists n \geq N_1(t) : |\hat{\mu}_{1,n} - \mu_1| \geq \epsilon\}\}$$

so that

$$\mathbb{P}[N_1(t) \leq \sqrt{t}, \mathcal{C}(t)] \leq \sum_{n=0}^{\lfloor \sqrt{t} \rfloor} \mathbb{P}[G_n].$$

On G_n, we define an index set I_n and its subset $I_{n,l}$

$$I_n := [\tau_n, \tau_n + \lceil \sqrt{t} - 1 \rceil] \subset [\tau_n, \tau_{n+1}]$$
$$I_{n,l} := \left[\tau_n + \left\lceil \frac{l-1}{K} (\sqrt{t} - 1) \right\rceil, \tau_n + \left\lceil \frac{l}{K} (\sqrt{t} - 1) \right\rceil \right], \quad l \in [K].$$

Note that the inclusion on I_n holds under G_n. In the analysis of Thompson sampling (Agrawal and Goyal, 2012; Kaufmann et al., 2012; Korda et al., 2013), an arm a is called *saturated* if $N_a(t) \geq C'_a \log t$ for a constant C'_a that depends on the model.

In this chapter, we call an arm i is saturated if $N_i(t) \geq \max_{a \in [K]} C_a \log t$ for a constant C_a such that

$$C_a \geq C'_a \frac{d(\mu_2 + \epsilon, \mu_K - \epsilon)}{\underline{d}_a}.$$

Note that C_a's are also constants that only depend on the model, and $C_a \geq C'_a$ holds from the definition of $\underline{d}_a$, so that Lemma 16 is still applicable. For each interval I_n, let introduce

- $F_{n,l}$: the event that by the end of the interval $I_{n,l}$ at least l suboptimal arms are saturated.
- $r_{n,l}$: the number of playing unsaturated suboptimal arms, which is called interruptions during $I_{n,l}$.

Let us consider

$$\mathbb{P}[G_n] = \underbrace{\mathbb{P}[G_n, F_{n,K-1}]}_{(D1)} + \underbrace{\mathbb{P}[G_n, F^c_{n,K-1}]}_{(E1)}. \tag{34}$$

F.2.1. BOUNDS ON (D1)

From the definition, one can rewrite

$$\begin{aligned}(D1) &= \mathbb{P}[\{\exists s \in I_{n,K}, \exists a \neq 1 : \tilde{\mu}_a(s) \geq \mu_2 + \epsilon\}, G_n, F_{n,K-1}] \\ &\quad + \mathbb{P}[\{\forall s \in I_{n,K}, \forall a \neq 1 : \tilde{\mu}_a(s) \leq \mu_2 + \epsilon\}, G_n, F_{n,K-1}] \\ &\leq \frac{2(K-1)}{t^3} + \overbrace{\mathbb{P}[\underbrace{\{\forall s \in I_{n,K}, \forall a \neq 1 : \tilde{\mu}_a(s) \leq \mu_2 + \epsilon\}}_{=:D_{n,K}}, G_n, F_{n,K-1}]}^{(D2)},\end{aligned}$$

where the inequality holds from Lemma 16. Here, (D2) can be decomposed as

$$\begin{aligned}(D2) &= \mathbb{P}[D_{n,K}, G_n, F_{n,K-1}, \{\forall a \neq 1, \exists s \in I_{n,K} : \mathcal{B}^c_a(s) \cup \tilde{\mathcal{B}}^c_a(s)\}] \\ &\quad + \mathbb{P}[D_{n,K}, G_n, F_{n,K-1}, \{\forall a \neq 1, \forall s \in I_{n,K} : \mathcal{B}_a(s) \cap \tilde{\mathcal{B}}_a(s)\}].\end{aligned}$$

From Lemma 16, we obtain

$$
\begin{aligned}
(D2) &\leq \frac{6(K-1)}{t^3} + \mathbb{P}[D_{n,K}, G_n, F_{n,K-1}, \{\forall a \neq 1, \forall s \in I_{n,K} : \mathcal{B}_a(s) \cap \tilde{\mathcal{B}}_a(s)\}] \\
&\leq \frac{6(K-1)}{t^3} \\
&\quad + \mathbb{P}[D_{n,K}, G_n, F_{n,K-1}, \{\forall a \neq 1, \forall s \in I_{n,K} : \mathcal{B}_a(s) \cap \tilde{\mathcal{B}}_a(s), \tilde{m}(s) \neq 1\}] \\
&\quad + \mathbb{P}[D_{n,K}, G_n, F_{n,K-1}, \{\forall a \neq 1, \forall s \in I_{n,K} : \mathcal{B}_a(s) \cap \tilde{\mathcal{B}}_a(s)\} \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad , \{\exists s \in I_{n,K} : \tilde{m}(s) = 1\}] \\
&\leq \frac{6(K-1)}{t^3} + C^{\frac{\sqrt{t}-1}{K}} \\
&\quad + \left. \begin{matrix} \mathbb{P}[D_{n,K}, G_n, F_{n,K-1}, \{\forall a \neq 1, \forall s \in I_{n,K} : \mathcal{B}_a(s) \cap \tilde{\mathcal{B}}_a(s)\} \\ , \{\exists s \in I_{n,K} : \tilde{m}(s) = 1\}] \end{matrix} \right\} (D_3),
\end{aligned}
$$

where the last inequality holds from Lemma 17. Next, one can see

$$
\begin{aligned}
(D3) &= \mathbb{P}[D_{n,K}, G_n, F_{n,K-1}, \{\forall a \neq 1, \forall s \in I_{n,K} : \mathcal{B}_a(s) \cap \tilde{\mathcal{B}}_a(s)\} \\
&\qquad\qquad\qquad\qquad , \{\exists s \in I_{n,K} : \tilde{m}(s) = 1, m(s) = 1\}] \\
&\quad + \mathbb{P}[D_{n,K}, G_n, F_{n,K-1}, \{\forall a \neq 1, \forall s \in I_{n,K} : \mathcal{B}_a(s) \cap \tilde{\mathcal{B}}_a(s)\} \\
&\qquad\qquad\qquad\qquad , \{\exists s \in I_{n,K} : \tilde{m}(s) = 1, m(s) \neq 1\}] \\
&\leq \mathbb{P}[D_{n,K}, G_n, F_{n,K-1}, \{\forall a \neq 1, \forall s \in I_{n,K} : \mathcal{B}_a(s) \cap \tilde{\mathcal{B}}_a(s)\} \\
&\qquad\qquad\qquad\qquad , \{\exists s \in I_{n,K} : \tilde{m}(s) = 1, m(s) = 1\}] \\
&\quad + \mathbb{P}[D_{n,K}, G_n, F_{n,K-1}, \{\text{arm } 1 \text{ is saturated}\}, \{\exists s \in I_{n,K} : \mathcal{B}_1^c(s)\}] \qquad (35)
\end{aligned}
$$

where (35) holds from Thompson exploration since $i(t) \neq 1$ on $\mathcal{M}^c(t)$ implies that $N_1(t) \geq N_{i(t)}$, i.e., arm 1 is saturated. From Lemma 17, it holds that

$$
\begin{aligned}
(D3) &\leq \frac{2(K-1)}{t^3} + \mathbb{P}[D_{n,K}, G_n, F_{n,K-1}, \{\forall a \neq 1, \forall s \in I_{n,K} : \mathcal{B}_a(s) \cap \tilde{\mathcal{B}}_a(s)\} \\
&\qquad\qquad\qquad\qquad , \{\exists s \in I_{n,K} : \tilde{m}(s) = m(s) = 1\}] \\
&= \frac{2(K-1)}{t^3} + (D4),
\end{aligned}
$$

where $(D4)$ denotes the second term. Note that Thompson exploration with $\{m(s) = 1\}$ will choose only $j(s)$ under the event G_n, i.e., only $\{i(s) = j(s)\}$ happens during I_n for any n when $m(s) = \tilde{m}(s)$ holds. It holds that

$$
\begin{aligned}
(D4) \leq & \underbrace{\sum_{s \in I_{n,K}} \sum_{a=2}^{K} \mathbb{P}[m(s) = 1, i(s) = j(s) = a, \mathcal{A}_1(s), \mathcal{B}_a(s), \mathcal{M}(s), G_n]}_{(D5)} \\
& + \underbrace{\sum_{s \in I_{n,K}} \sum_{a=2}^{K} \mathbb{P}[m(s) = 1, i(s) = j(s) = a, \mathcal{A}_1^c(s), \mathcal{B}_a(s), \mathcal{M}(s)]}_{(D6)}.
\end{aligned}
$$

From Lemma 7, if an event in $(D5)$ occurs for some s, then it implies that $\mathcal{B}_1(t)$ holds for all $t \geq s$ such that for all $t \geq N'$, $C_a^* \log t \geq \max\{M, D_1/\underline{d}_a\}$ for all $a \in [K] \setminus \{1\}$ holds, which contradicts to the event G_n that implies the existence of $t \geq s$ such that $\mathcal{B}_1^c(t)$ holds. Therefore, we have

$$(D5) = 0.$$

Note that $(D6)$ is the form considered in Lemma 9. Therefore, we have

$$(D6) \leq \frac{\sqrt{t}-1}{K} \sum_{a=2}^{K} \mathbb{P}\left[N_a(s) \leq c_a^* D_a\right],$$

for some constants c_a^* and random variables D_a in Lemma 9 such that its expectation is finite. Let $N_{\boldsymbol{\mu},A}(\epsilon)$ be a constant that depends on the model and epsilon such that for $t \geq N_{\boldsymbol{\mu},A}(\epsilon)$, it holds for any $a \in \{2, \ldots, K\}$

$$C_a^* \log t \geq c_a^* D_a,$$

i.e., the event in $(D6)$ cannot occur for $t \geq N_{\boldsymbol{\mu},A}(\epsilon)$. Hence, there exist some constant $C_D(\pi_{\mathrm{j}}, \boldsymbol{\mu}, b, \epsilon) < \infty$ such that

$$\begin{aligned}\sum_{t=1}^{T} \sum_{n=0}^{\lfloor\sqrt{t}\rfloor} (D1) &\leq \max\left\{N', N_{\boldsymbol{\mu},A}(\epsilon)\right\} + \sum_{t=N_{\boldsymbol{\mu},A}(\epsilon)+1}^{\infty} \frac{8(K-1)}{t^2\sqrt{t}} + \sqrt{t} C^{\frac{\sqrt{t}-1}{K}} \\ &\leq C_D(\pi_{\mathrm{j}}, \boldsymbol{\mu}, b, \epsilon). \end{aligned} \tag{36}$$

F.2.2. BOUNDS ON (E1)

By adapting the proof of Kaufmann et al. (2012); Korda et al. (2013), we prove $(E1)$ is upper bounded by some constants through the mathematical induction, i.e., we will show

$$\mathbb{P}[G_n, F_{n,K-1}^c] \leq (K-2)\left(\frac{10(K-1)}{t^3} + k(\boldsymbol{\mu}, b, n, t)\right),$$

where k is a function such that $\sum_{t\geq 1}\sum_{n\leq\sqrt{t}} k < \infty$.

First, for the base case, it can be easily seen that for $t \geq N_{\boldsymbol{\mu},b}$ such that

$$\forall t \geq N_{\boldsymbol{\mu},b}, \left\lceil \frac{\sqrt{t}-1}{K^2} \right\rceil \geq C_* \log t,$$

where $C_* = \max_{a\neq 1} C_a$ since only suboptimal arms are selected during $I_{n,l}$ under G_n. Then, for $t \geq N_{\boldsymbol{\mu},b}$,

$$\mathbb{P}[G_n, F_{n,1}^c] = 0.$$

We refer the reader to Kaufmann et al. (2012) for more explanations in the base case. Then, we assume that for some $2 \leq l \leq K-1$ if $t \geq N_{\boldsymbol{\mu},b}$, then

$$\mathbb{P}[G_n, F_{n,l-1}^c] \leq (l-2)\left(\frac{10(K-1)}{t^3} + k(\boldsymbol{\mu}, b, n, t)\right).$$

Therefore, we remain to show that

$$\mathbb{P}[G_n, F_{n,l}^c, F_{n,l-1}] \leq \frac{10(K-1)}{t^3} + k(\boldsymbol{\mu}, b, n, t).$$

On the event $(G_n, F^c_{n,l}, F_{n,l-1})$, there are exactly $l-1$ saturated suboptimal arms at the beginning of interval $I_{n,l}$ and no new arm is saturated during this interval, which implies that $r_{n,l} \leq KC_* \log t$. For the set of saturated suboptimal arms $\mathcal{S}_l$ at the end of $I_{n,l}$, it holds that

$$\begin{aligned}\mathbb{P}[G_n, F^c_{n,l}, F_{n,l-1}] &\leq \mathbb{P}[G_n, F_{n,l-1}, \{r_{n,l} \leq KC_* \log t\}] \\ &\leq \mathbb{P}[G_n, F_{n,l-1}, \{\exists s \in I_{n,l}, a \in \mathcal{S}_{l-1} : \tilde{\mathcal{B}}^c_a(s) \cup \mathcal{B}^c_a(s)\}] \\ &\quad + \left.\begin{array}{l}\mathbb{P}[G_n, F_{n,l-1}, \{r_{n,l} \leq KC_* \log t\}, \\ \quad \{\forall s \in I_{n,l}, a \in \mathcal{S}_{l-1} : \tilde{\mathcal{B}}_a(s) \cap \mathcal{B}_a(s)\}]\end{array}\right\} (E2),\end{aligned}$$

By applying Lemma 16 again, we have

$$\mathbb{P}[G_n, F_{n,l-1}, \{\exists s \in I_{n,l}, a \in \mathcal{S}_{l-1} : \tilde{\mathcal{B}}^c_a(s) \cup \mathcal{B}^c_a(s)\}] \leq \frac{6(K-1)}{t^3}.$$

To bound $(E2)$, we introduce a random interval $\mathcal{J}_k$ for $k \in \{0, \ldots, r_{n,l} - 1\}$ as the time between k-th and $k+1$-th interruption in $I_{n,l}$ and set $\mathcal{J}_k = \emptyset$ for $k \geq r_{n,l}$. On $(E2)$, there is a subinterval where no interruptions occur with length $\lceil \frac{\sqrt{t}-1}{C_* K^2 \log t} \rceil$. Then, it holds that

$$\begin{aligned}(E2) &\leq \mathbb{P}\left[\left\{\exists k \in \{0, \ldots, r_{n,l}\} : |\mathcal{J}_k| \geq \frac{\sqrt{t}-1}{C_* K^2 \log t}\right\}, \right. \\ &\qquad\qquad \left. \{\forall s \in I_{n,l}, a \in \mathcal{S}_l : \tilde{\mathcal{B}}_a(s) \cap \mathcal{B}_a(s)\}, G_n, F_{n,l-1}\right] \\ &\leq \sum_{k=1}^{KC_* \log t} \mathbb{P}\left[\left\{|\mathcal{J}_k| \geq \frac{\sqrt{t}-1}{C_* K^2 \log t}\right\}, \{\forall s \in \mathcal{J}_k, a \in \mathcal{S}_l : \tilde{\mathcal{B}}_a(s) \cap \mathcal{B}_a(s)\}, G_n\right].\end{aligned}$$

Note that on G_n and $\forall s \in \mathcal{J}_k$, only $i(s) \in \mathcal{S}_l$ happens, i.e., $\{m(s) \neq \tilde{m}(s), m(s) \notin \mathcal{S}_l, \tilde{m}(s) \notin \mathcal{S}_l\}$ cannot occur. Therefore, for any $s \in \mathcal{J}_k$ under $\{\forall a \in \mathcal{S}_l : \tilde{\mathcal{B}}_a(s) \cap \mathcal{B}_a(s)\}$, we have

$$\begin{aligned}\mathbb{1}[m(s) \neq \tilde{m}(s), G_n, \tilde{\mathcal{B}}_{\tilde{m}(s)}(s)] &= \mathbb{1}[m(s) \in \mathcal{S}_l, \tilde{m}(s) \in \mathcal{S}_l \setminus \{m(s)\}, G_n, \tilde{\mathcal{B}}_{\tilde{m}(s)}(s)] \\ &\quad + \mathbb{1}[m(s) = 1, \tilde{m}(s) \in \mathcal{S}_l, G_n, \tilde{\mathcal{B}}_{\tilde{m}(s)}(s), \tilde{\mathcal{B}}^c_1(s)] \\ &\quad + \mathbb{1}[m(s) \in \mathcal{S}_l, \tilde{m}(s) = 1, G_n, \tilde{\mathcal{B}}_1(s), \mathcal{B}^c_1(s)].\end{aligned}$$

Here, it holds that

$$\{m(s) \in \mathcal{S}_l, \tilde{m}(s) \in \mathcal{S}_l \setminus \{m(s)\}, G_n, \tilde{\mathcal{B}}_{\tilde{m}(s)}(s)\} \subset \{\tilde{\mu}_1(s) \leq \mu_2 + \epsilon, G_n\}.$$

Similarly to the (D3), $i(s) \neq 1$ implies that arm 1 is already played more than the saturated arm. Let us define an event

$$E2(s) := \{m(s) = \tilde{m}(s) \in \mathcal{S}^c_l \cup \{1\}]\} \cap \{\tilde{\mu}_1(s) \geq \mu_2 + \epsilon\}.$$

Then, from the above inclusive relationship, we have

$$\begin{aligned}
&\mathbb{P}\left[\left\{|\mathcal{J}_k| \geq \frac{\sqrt{t}-1}{C_* K^2 \log t}\right\}, \{\forall s \in \mathcal{J}_k, a \in \mathcal{S}_l : \tilde{\mathcal{B}}_a(s) \cap \mathcal{B}_a(s)\}, G_n\right] \\
&\quad \leq \mathbb{P}\left[\left\{|\mathcal{J}_k| \geq \frac{\sqrt{t}-1}{C_* K^2 \log t}\right\}, \left\{\forall s \in \mathcal{J}_k : \{\forall a \in \mathcal{S}_l : \tilde{\mathcal{B}}_a(s) \cap \mathcal{B}_a(s)\}\right.\right. \\
&\qquad\qquad \left.\left. \cap \{\tilde{\mu}_1(s) \leq \mu_2 + \epsilon\}\right\}, G_n\right] \\
&\quad + \mathbb{P}\left[\left\{|\mathcal{J}_k| \geq \frac{\sqrt{t}-1}{C_* K^2 \log t}\right\}, \{\forall s \in \mathcal{J}_k, a \in \mathcal{S}_l : \tilde{\mathcal{B}}_a(s) \cap \mathcal{B}_a(s)\}, \right. \\
&\qquad\qquad \left. \{\exists s \in \mathcal{J}_k : \mathcal{B}_1^c(s) \cup \tilde{\mathcal{B}}_1^c(s)\}, G_n\right] \\
&\quad + \left.\mathbb{P}\left[\left\{|\mathcal{J}_k| \geq \frac{\sqrt{t}-1}{C_* K^2 \log t}\right\}, \{\forall s \in \mathcal{J}_k, a \in \mathcal{S}_l : \tilde{\mathcal{B}}_a(s) \cap \mathcal{B}_a(s)\} \right.\right. \\
&\qquad\qquad \left.\left. \{\exists s \in \mathcal{J}_k : E2(s)\}, G_n\right]\right\} (E3).
\end{aligned}$$

By applying Lemmas 16 and 17, we have

$$\begin{aligned}
&\mathbb{P}\left[\left\{|\mathcal{J}_k| \geq \frac{\sqrt{t}-1}{C_* K^2 \log t}\right\}, \{\forall s \in \mathcal{J}_k, a \in \mathcal{S}_l : \tilde{\mathcal{B}}_a(s) \cap \mathcal{B}_a(s)\}, G_n\right] \\
&\qquad\qquad \leq C^{\frac{\sqrt{t}-1}{C_* K^2 \log t}} + \frac{6}{t^3} + (E3).
\end{aligned}$$

From the definition of $\mathcal{J}_k$ and G_n, one can see that

$$\begin{aligned}
(E3) &= \mathbb{P}\left[\left\{|\mathcal{J}_k| \geq \frac{\sqrt{t}-1}{C_* K^2 \log t}\right\}, \{\forall s \in \mathcal{J}_k : a \in \mathcal{S}_l : \tilde{\mathcal{B}}_a(s) \cap \mathcal{B}_a(s)\} \right. \\
&\qquad\qquad \left., \{\exists s \in \mathcal{J}_k : E2(s) \cap \{j(s) = i(s) \in \mathcal{S}_l\}\}, G_n\right] \\
&\leq \mathbb{P}\Big[\exists s \in \mathcal{J}_k : m(s) = \tilde{m}(s) \in \mathcal{S}_l^c \cup \{1\}, j(s) \in \mathcal{S}_l, i(s) = j(s), \mathcal{A}_{m(s)}^c \\
&\qquad\qquad , \mathcal{B}_{j(s)}, \tilde{\mu}_1(s) \geq \mu_2 + \epsilon, G_n\Big] \\
&\quad + \mathbb{P}\Big[\exists s \in \mathcal{J}_k : m(s) = \tilde{m}(s) \in \mathcal{S}_l^c \cup \{1\}, j(s) \in \mathcal{S}_l, i(s) = j(s), \mathcal{A}_{m(s)} \\
&\qquad\qquad , \mathcal{B}_{j(s)}, \tilde{\mu}_1(s) \geq \mu_2 + \epsilon, G_n\Big]. \qquad (37) \\
&=: (E4) + (E5).
\end{aligned}$$

The first equation holds since only saturated suboptimal arms have to be played on $\mathcal{J}_k$ when $m(s) = \tilde{m}(s)$ is unsaturated or optimal arm, which makes $j(s) = i(s) \in \mathcal{S}_l$. Let us denote the event in the first term and the second term of RHS in (37) by $(E4)$ and $(E5)$, respectively.

From Lemma 9, we have

$$\begin{aligned}\mathbb{1}[(E4)] &\leq \sum_{s\in\mathcal{J}_k}\sum_{a\in\mathcal{S}_l}\sum_{m\in\mathcal{S}_l\cup\{1\}} \mathbb{1}[m(s)=m, i(s)=j(s)=a, \mathcal{A}_m^c(s), \mathcal{B}_a(s)] \\ &\leq \sum_{s\in\mathcal{J}_k}\sum_{a\in\mathcal{S}_l}\sum_{m\in\mathcal{S}_l\cup\{1\}} \mathbb{1}[N_a(s) \leq c^*_{m,a} D_{m,a}].\end{aligned}$$

Similarly to the case of $(D4)$, there exists some deterministic constant $N_{\boldsymbol{\mu},A}(\epsilon)'$ such that for $t \geq N_{\boldsymbol{\mu},A}(\epsilon)', \forall (m,a) \in (\mathcal{S}_l^c \cup \{1\}, \mathcal{S}_l)$

$$C^*_a \log t \geq c^*_{m,a} D_{m,a},$$

where we replace 1 by m in c^*_a and D_a to define $c^*_{m,a}$ and $D_{m,a}$.

Further, $(E5)$ can be decomposed by

$$(E5) = (E6) + (E7),$$

where

$$(E6) := \mathbb{P}\left[\exists s \in \mathcal{J}_k : m(s) = \tilde{m}(s) \in \mathcal{S}_l^c, j(s) \in \mathcal{S}_l, i(s) = j(s), \mathcal{A}_{m(s)}, \mathcal{B}_{j(s)}, \tilde{\mu}_1(s) \geq \mu_2 + \epsilon, G_n\right]$$

$$(E7) := \mathbb{P}\left[\exists s \in \mathcal{J}_k : m(s) = \tilde{m}(s) = 1, j(s) \in \mathcal{S}_l, i(s) = j(s), \mathcal{A}_1, \mathcal{B}_{j(s)}, \tilde{\mu}_1(s) \geq \mu_2 + \epsilon, G_n\right].$$

Note that on $(E6)$, $\tilde{\mathcal{B}}_m^c(s)$ always holds since $\tilde{\mu}_1 > \mu_2 + \epsilon$ but $\tilde{m}(s) \neq 1$ and $(E5)$ is a subset of the event we consider in Lemma 7, i.e., event $(E6)$ implies the existence of $s \in \mathcal{J}_k$ such that

$$N_m(s) \geq N_{j(s)} \frac{\underline{d}_{j(s)}}{d(\mu_m + \epsilon, \mu_j - \epsilon)} \geq C_* \frac{\underline{d}_{j(s)}}{d(\mu_m + \epsilon, \mu_{j(s)} - \epsilon)} \log t.$$

From the definition of C_* and saturation, it holds that for any $m \in \mathcal{S}_l^c$

$$C_* \frac{\underline{d}_{j(s)}}{d(\mu_m + \epsilon, \mu_{j(s)} - \epsilon)} \geq C_* \frac{\min_{a\neq 1} \underline{d}_a}{d(\mu_2 + \epsilon, \mu_K - \epsilon)} \geq C'_m \log t.$$

As a result, we have

$$\mathbb{P}[(E6)] = \mathbb{P}[\{\exists s \in \mathcal{J}_k, m \in \mathcal{S}_l^c : \tilde{\mathcal{B}}_m^c(s)\} \cap (E5)] \leq \frac{4(K-1)}{t^3}.$$

Similarly to the case of $(D5)$, if the event in $(E7)$ occurs some $s \in \mathcal{J}_k$ for t such that $t \geq N'$, $C^*_a \log t \geq \max\{M, D_1/\underline{d}_a\}$ for all $a \in [K] \setminus \{1\}$, then only $\mathcal{B}_1(t)$ holds for $s \geq t$ holds, which contradicts to the event G_n.

Therefore, for $t \geq N_0 := \max(N_{\boldsymbol{\mu},b}, N_{\boldsymbol{\mu},A}(\epsilon)', N_K, N')$, where N_K in Lemma 16, it holds

$$(E2) \leq KC_* \log t \left(C^{\frac{\sqrt{t}-1}{C_* K^2 \log t}} + \frac{10(K-1)}{t^3} \right) =: k(\boldsymbol{\mu}, b, n, t).$$

Hence, there exists some constants $C_E(\pi_{\mathrm{j}}, \boldsymbol{\mu}, b, \epsilon) < \infty$ such that

$$\begin{aligned}\sum_{T=1}^{\infty}\sum_{t=T+1}^{\infty}\sum_{n=1}^{\lfloor\sqrt{t}\rfloor}(E1) &\leq N_0 + \sum_{T=N_0+1}^{\infty}\sum_{t=T+1}^{\infty}\frac{6(K-1)^2}{t^2\sqrt{t}} \\ &\quad + \sum_{T=N_0+1}^{\infty}\sum_{t=T+1}^{\infty} KC_* \log t\left(\sqrt{t}C^{\frac{\sqrt{t}-1}{C_* K^2 \log t}} + \frac{10(K-1)}{t^2\sqrt{t}}\right) \\ &\leq N_0 + C_E(\pi_{\mathrm{j}}, \boldsymbol{\mu}, b, \epsilon). \end{aligned} \tag{38}$$

F.2.3. CONCLUSION

By combining (36) and (38) with (34), we obtain

$$\begin{aligned}\sum_{T=1}^{\infty}\sum_{t=T+1}^{\infty}\mathbb{P}[N_1(t) \leq \sqrt{t}, \mathcal{C}(t)] &\leq \sum_{T=1}^{\infty}\sum_{t=T+1}^{\infty}\sum_{n=N_1(T+1)}^{\lfloor\sqrt{t}\rfloor}(D1) + (E1) \\ &\leq N_0 + C_D(\pi_{\mathrm{j}}, \boldsymbol{\mu}, b, \epsilon) + C_E(\pi_{\mathrm{j}}, \boldsymbol{\mu}, b, \epsilon) \\ &=: C(\pi_{\mathrm{j}}, \boldsymbol{\mu}, b, \epsilon) < \infty,\end{aligned}$$

which concludes the proof. ■

Appendix G. Proof of Theorem 3: Sample complexity

Here, we derive the upper bound on the sample complexity of BC-TE.

Before beginning the proof, we first provide a technical lemma provided in Garivier and Kaufmann (2016).

Lemma 18 (Lemma 18 in Garivier and Kaufmann (2016)) *For every $\alpha \in [1, \frac{e}{2}]$, for any two constants $c_1, c_2 > 0$,*

$$x = \frac{\alpha}{c_1}\left[\log\left(\frac{c_2 e}{c_1^{\alpha}}\right) + \log\log\left(\frac{c_2}{c_1^{\alpha}}\right)\right]$$

is such that $c_1 x \geq \log(c_2 x^{\alpha})$.

Next, we define a set of bandit instances $\mathcal{S}$ for any $\epsilon > 0$ as follows:

$$\mathcal{S} = \mathcal{S}(\nu, \epsilon) := \{\boldsymbol{\mu}' : |\boldsymbol{\mu}' - \boldsymbol{\mu}| \leq \epsilon\},$$

where $\boldsymbol{\mu}$ denotes the true mean reward vector. For any $i \neq 1$, if $\boldsymbol{\mu}' \in \mathcal{S}$, we have the following inequality:

$$\forall \boldsymbol{w} \in \Sigma_K : \frac{1}{1+\epsilon} f_i(\boldsymbol{w}; \boldsymbol{\mu}) \leq f_i(\boldsymbol{w}; \boldsymbol{\mu}') \leq (1+\epsilon) f_i(\boldsymbol{w}; \boldsymbol{\mu}). \tag{39}$$

From the relationship in (21), (39) is equivalent to

$$\begin{aligned}\forall \boldsymbol{w} \in \Sigma_K : \frac{1}{1+\epsilon} g(\boldsymbol{w}; \boldsymbol{\mu}) &\leq g(\boldsymbol{w}; \boldsymbol{\mu}') \leq (1+\epsilon) g(\boldsymbol{w}; \boldsymbol{\mu}) \\ \forall x \in [0,1] : \frac{1}{1+\epsilon} k_i(x; \boldsymbol{\mu}) &\leq k_i(x; \boldsymbol{\mu}') \leq (1+\epsilon) k_i(x; \boldsymbol{\mu}) \\ \forall z \in [0,1] : \frac{1}{1+\epsilon} h_i(z; \boldsymbol{\mu}) &\leq h_i(z; \boldsymbol{\mu}') \leq (1+\epsilon) h_i(z; \boldsymbol{\mu}).\end{aligned}$$

Notice that that for any $t \geq T_B$, $\hat{\boldsymbol{\mu}}(t) \in \mathcal{S}$ holds from the the definition of T_B in (9).

Therefore, we can assume

$$\frac{1}{1+\epsilon}\frac{z_i^*}{1-z_i^*} \leq \frac{z_i^*(\boldsymbol{\mu}')}{1-z_i^*(\boldsymbol{\mu}')} \leq (1+\epsilon)\frac{z_i^*}{1-z_i^*} \tag{40}$$

$$\frac{1}{1+\epsilon}\frac{\underline{z}_i}{1-\underline{z}_i} \leq \frac{\underline{z}_i(\boldsymbol{\mu}')}{1-\underline{z}_i(\boldsymbol{\mu}')} \leq (1+\epsilon)\frac{\underline{z}_i}{1-\underline{z}_i}. \tag{41}$$

and for $t \geq T_B$ and the definition of a challenger at round t, $j(t)$ in (8),

$$\frac{1}{1+\epsilon}\min_{a\neq 1} f_i(x;\boldsymbol{\mu}) \leq f_{j(t)}(x;\boldsymbol{\mu}) \leq (1+\epsilon)\min_{a\neq 1} f_i(x;\boldsymbol{\mu}). \tag{42}$$

Notice that (42) provides

$$\frac{1}{1+\epsilon}\min_{a\neq 1} k_i(x;\boldsymbol{\mu}) \leq k_{j(t)}(x;\boldsymbol{\mu}) \leq (1+\epsilon)\min_{i\neq 1} k_i(x;\boldsymbol{\mu}). \tag{43}$$

Since $tf_i(\boldsymbol{w}^t;\boldsymbol{\mu}) = (N_1(t)+N_i(t))h_i(z_i^t;\boldsymbol{\mu})$ holds from their relationship in (21) and $z_i^t = \frac{w_i^t}{w_1^t+w_i^t}$, (42) also implies that

$$\begin{aligned}\frac{1}{1+\epsilon}\min_{i\neq 1}(N_1(t)+N_i(t))h_i(z_i^t;\boldsymbol{\mu}) &\leq (N_1(t)+N_{j(t)}(t))h_{j(t)}(z_{j(t)}^t;\boldsymbol{\mu}) \\ &\leq (1+\epsilon)\min_{i\neq 1}(N_1(t)+N_i(t))h_i(z_i^t;\boldsymbol{\mu}).\end{aligned}$$

From the concavity of the objective function, we have the following result, whose proof is provided in Section G.3.

Lemma 19 *For any $i \neq 1$, $tf_i(\boldsymbol{w}^t;\boldsymbol{\mu})$ is non-decreasing with respect to $t \in \mathbb{N}$.*

Proof of theorem 3 We first introduce a positive increasing sequence $(G_m)_{m\in\mathbb{N}}$ and let ψ_m be the first round where $tg(\boldsymbol{w}^t;\boldsymbol{\mu}) > G_m$ holds, which is defined as

$$\psi_m := \inf\{t \in \mathbb{N}_{\geq T_B} : tg(\boldsymbol{w}^t;\boldsymbol{\mu}) \geq G_m\}.$$

Notice that Lemma 19 ensures $\psi_m \leq \psi_{m+1}$ for any $m \in \mathbb{N}$ since $tg(\boldsymbol{w}^t;\boldsymbol{\mu}) = t\min_{i\neq 1} f_i(\boldsymbol{w}^t;\boldsymbol{\mu})$ is non-decreasing.

For notational simplicity, $\underline{g}$ denotes the value of the objective function $g(\boldsymbol{w};\boldsymbol{\mu})$ at $\boldsymbol{w} = \underline{w}$ defined in (23). Then from (21)

$$\forall i \neq 1 : \underline{g} = \underline{w}_1 k_i(\underline{w}_i/\underline{w}_1;\boldsymbol{\mu}) = (\underline{w}_1+\underline{w}_i)h_i(\underline{z}_i;\boldsymbol{\mu}). \tag{44}$$

Here, we set G_1 to satisfy

$$\forall i \in [K] : N_i(T_B) \leq \frac{\underline{w}_i}{\underline{g}}G_1. \tag{45}$$

Then, the stopping time τ_δ can be written as

$$\begin{aligned}\tau_\delta &= \inf\{t \in \mathbb{N} : tg(\boldsymbol{w}^t;\hat{\boldsymbol{\mu}}(t)) \geq \beta(t,\delta)\} \\ &\leq \inf\{t \in \mathbb{N}_{\geq T_B} : \frac{tg(\boldsymbol{w}^t;\boldsymbol{\mu})}{1+\epsilon} \geq \beta(t,\delta)\} \\ &\leq T_B + \inf\left\{\psi_m : \frac{1}{1+\epsilon}G_m \geq \beta(\psi_m,\delta), m \in \mathbb{N}\right\}.\end{aligned} \tag{46}$$

To find the upper bound of the stopping time, we require the relationship between G_m and ψ_m. To do this, we first derive the bounds on the number of plays $N_i(t)$.

G.1. Bounds on the number of plays

Here, we aim to derive the upper bounds on $N_i(t)$ for $t \in [\psi_m, \psi_{m+1})$ and for any $i \in [K]$.

For $t \geq T_B$, only $m(t) = 1$ occurs. Therefore, an arm $i \neq 1$ is played either when TE occurs or when $j(t) = i$ and $d(\hat{\mu}_i(t), \hat{\mu}_{1,i}(t)) \geq d(\hat{\mu}_1(t), \hat{\mu}_{1,i}(t))$ for $t \geq T_B$. Thus, if $j(t) \neq i$ holds for all $t \in [\psi_m, \psi_{m+1})$, then

$$N_i(\psi_{m+1}) = N_i(\psi_m) + M_{i,m},$$

where $M_{i,m}$ denote the number of the arm i being played by TE during $[\psi_m, \psi_{m+1})$, which is

$$M_{i,m} = \sum_{t=\psi_m}^{\psi_{m+1}-1} \mathbb{1}[\mathcal{M}^c(t), i(t) = i].$$

The latter condition can be rewritten as $j(t) = i$ and $z_i^t \leq z_i^*(\hat{\boldsymbol{\mu}}(t))$ from the definition of z_i^* in (20). For notational simplicity, we denote $z_i^*(\hat{\boldsymbol{\mu}}(t))$ and $\underline{z}_i(\hat{\boldsymbol{\mu}}(t))$ by $z_{i,t}^*$ and $\underline{z}_{i,t}$, respectively.

(1) Upper bound for the second-best arm Firstly, let us consider the second-best arm $j^*(\nu)$, which is assumed to be the arm 2 in this chapter. It should be noted that the second-best arm may not be unique. Then let us define a partition of $Q_m := [\psi_m, \psi_{m+1})$

$$(Q1) := \left\{ t \in [\psi_m, \psi_{m+1}) : N_1(t) \leq \frac{\underline{w}_1}{\underline{g}} G_{m+1} \right\}$$
$$(Q2) := \left\{ t \in [\psi_m, \psi_{m+1}) : N_1(t) > \frac{\underline{w}_1}{\underline{g}} G_{m+1} \right\}.$$

Then, we define $\epsilon_1 = \epsilon_1(\epsilon, G_{m+1}/G_m) > \epsilon$ to be a constant satisfying

$$k_2\left((1+\epsilon_1)\frac{\underline{w}_2}{\underline{w}_1}; \boldsymbol{\mu}\right) \geq \frac{G_{m+1}}{G_m}\frac{\underline{g}}{\underline{w}_1}, \tag{47}$$

Here, one can see that $\epsilon_1 \to 0_+$ as $\epsilon \to 0_+$ and $\frac{G_{m+1}}{G_m} \to 1_+$ from (44). Then we will show that if $N_2(t) \geq N' = (1+\epsilon_1)\frac{\underline{w}_2}{\underline{g}} G_{m+1}$, then $i(t) = 2$ holds only when TE occurs.

(1-i) When $t \in (Q1)$ In this case,

$$\begin{aligned}
N_2(t) \geq N' = (1+\epsilon_1)\frac{\underline{w}_2}{\underline{g}} G_m &= (1+\epsilon_1)\frac{\underline{w}_2}{\underline{w}_1}\frac{\underline{w}_1}{\underline{g}} G_m \\
&\geq (1+\epsilon_1)\frac{\underline{w}_2}{\underline{w}_1} N_1(t) && \because t \in (Q1) \\
&= (1+\epsilon_1)\frac{\underline{z}_2}{1-\underline{z}_2} N_1(t) && \text{by definition of } \underline{w} \text{ in (23)} \\
&= (1+\epsilon_1)\frac{z_2^*}{1-z_2^*} N_1(t) && \text{by definition of } \underline{z} \text{ in (22)} \\
&> \frac{z_{2,t}^*}{1-z_{2,t}^*} N_1(t). && \text{by (40) and } \epsilon_1 > \epsilon
\end{aligned}$$

This implies that for $t \in (Q1)$, if $N_2(t) \geq N'$, then $z_2^t > z_{2,t}^*$ holds. Therefore, only $i(t) = 1$ happens unless TE occurs.

(1-ii) When $t \in (Q2)$ From the relationship between f_i and k_i in (21), one can see that $tf_i(\boldsymbol{w}^t; \boldsymbol{\mu}) = N_1(t)k_i(w_i^t/w_1^t; \boldsymbol{\mu})$. Therefore, one can extend Lemma 19 to show that $yk_i(c/y; \boldsymbol{\mu})$ is non-decreasing with respect to $y \geq 0$ for fixed $c > 0$ and any $i \neq 1$. Recall that the $k_i(x; \boldsymbol{\mu})$ is a strictly increasing function with respect to $x > 0$. Then we can obtain that

$$\begin{aligned}
N_1(t)k_2\left(\frac{N_2(t)}{N_1(t)}; \boldsymbol{\mu}\right) &\geq N_1(t)k_2\left(\frac{N'}{N_1(t)}; \boldsymbol{\mu}\right) \\
&\geq G_m \frac{\underline{w}_1}{\underline{g}} k_2\left(N' \frac{\underline{g}}{G_m \underline{w}_1}; \boldsymbol{\mu}\right) && \because t \in (Q2) \\
&= G_m \frac{\underline{w}_1}{\underline{g}} k_2\left((1+\epsilon_1)\frac{\underline{\underline{w}}_2}{\underline{w}_1}; \boldsymbol{\mu}\right) \\
&\geq G_m \frac{\underline{w}_1}{\underline{g}} \frac{G_{m+1}}{G_m} \frac{\underline{g}}{\underline{w}_1} && \text{by definition of } \epsilon_1 \text{ in (47)} \\
&= G_{m+1},
\end{aligned}$$

which contradicts the assumption $t \in (Q2)$.

(1-iii) Conclusion Therefore, for any $t \in Q_m$,

$$\left\{ N_2(t) \geq (1+\epsilon_1)\frac{\underline{\underline{w}}_2}{\underline{g}} G_m \right\} \implies \{j(t) \neq 2\},$$

which directly implies that

$$N_2(t) \leq \max\left(N_2(\psi_m), (1+\epsilon_1)\frac{\underline{\underline{w}}_2}{\underline{g}} G_m \right) + M_{2,m}.$$

Here, from the definition of G_1 in (45), $N_1(t) \leq \frac{\underline{w}_1}{\underline{g}} G_1$ holds for all $t < \psi_1$, which implies that $N_2(\psi_m) \leq (1+\epsilon_1)\frac{\underline{\underline{w}}_2}{\underline{g}} G_m + M_{2,0}$. Therefore, for any $t \in [\psi_m, \psi_{m+1})$,

$$N_2(t) \leq (1+\epsilon_1)\frac{\underline{\underline{w}}_2}{\underline{g}} G_m + M_2(\psi_{m+1})$$

where $M_i(\psi_{m+1}) = \sum_{l=0}^{m} M_{i,l}$ for any $i \in [K]$.

Here, let use define a random variable $M_T = \sum_{t=T_B}^{T} \mathbb{1}[\mathcal{M}^c(t)] = \sum_{i=1}^{K} \sum_m M_{i,m}$, which satisfies $\mathbb{E}[M_T] < \infty$ by Lemma 11. Then we can set G_m sufficiently large to satisfy

$$G_m \geq \frac{\underline{\underline{g}}}{\epsilon} M_T,$$

which directly implies that

$$N_2(t) \leq (1+\epsilon_1)\frac{\underline{\underline{w}}_2}{\underline{g}} G_m + \frac{\epsilon}{\underline{g}} G_m. \tag{48}$$

(2) Lower bound for the optimal arm For any $t \in Q_m$, it holds that

$$\begin{aligned} G_m &\leq N_1(t) \min_{i \neq 1} k_i \left(\frac{N_i(t)}{N_1(t)}; \boldsymbol{\mu} \right) \\ &= \min_{i \neq 1} (N_1(t) + N_i(t)) h_i(z_i^t; \boldsymbol{\mu}) && \text{by (21)} \\ &\leq (N_1(t) + N_2(t)) h_2(z_2^t; \boldsymbol{\mu}) \\ &\leq (N_1(t) + N_2(t)) h_2(\underline{z}_2; \boldsymbol{\mu}) && \text{by } \underline{z}_2 = z_2^* \\ &= \frac{N_1(t) + N_2(t)}{\underline{w}_1 + \underline{w}_2} \underline{g}. && \text{by (44)} \end{aligned}$$

Therefore, for $t = \psi_m$, the upper bound of $N_2(\psi_m)$ in (48) provides

$$N_1(\psi_m) \geq \frac{\underline{w}_1 + \underline{w}_2}{\underline{g}} G_m - (1 + \epsilon_1) \frac{\underline{w}_2}{\underline{g}} G_m - \frac{\epsilon}{\underline{g}} G_m.$$

Since $N_1(t)$ is non-decreasing from its definition, for any $t \geq \psi_m$,

$$N_1(t) \geq \frac{\underline{w}_1}{\underline{g}} G_m - \epsilon_1 \frac{\underline{w}_2}{\underline{g}} G_m - \frac{\epsilon}{\underline{g}} G_m. \tag{49}$$

(3) Upper bound on the challenger arms Based on the results obtained in (1) and (2), we will derive the upper bound of $N_{j(t)}(t)$ for $t \geq T_B$. For $t \in Q_m$, it holds that

$$G_m \leq N_1(t) \min_{i \neq 1} k_i \left(\frac{N_i(t)}{N_1(t)}; \boldsymbol{\mu} \right) < G_{m+1}.$$

Since $j(t) = \arg\min_{i=1} f_i(\boldsymbol{w}^t; \hat{\boldsymbol{\mu}}(t))$, by using (43), one can obtain that

$$\frac{1}{1+\epsilon} k_{j(t)} \left(\frac{N_{j(t)}(t)}{N_1(t)}; \boldsymbol{\mu} \right) \leq \min_{i \neq 1} k_i \left(\frac{N_i(t)}{N_1(t)}; \boldsymbol{\mu} \right).$$

Then, by (49)

$$\begin{aligned} N_1(t) \min_{i \neq 1} k_i \left(\frac{N_i(t)}{N_1(t)}; \boldsymbol{\mu} \right) &\geq \frac{1}{1+\epsilon} N_1(t) k_{j(t)} \left(\frac{N_{j(t)}(t)}{N_1(t)}; \boldsymbol{\mu} \right) \\ &\geq \frac{1}{1+\epsilon} \frac{G_m}{\underline{g}} (\underline{w}_1 - \epsilon_1 \underline{w}_2 - \epsilon) k_{j(t)} \left(\frac{\underline{g} N_{j(t)}(t)}{(\underline{w}_1 - \epsilon_1 \underline{w}_2 - \epsilon) G_m}; \boldsymbol{\mu} \right), \end{aligned}$$

which implies

$$k_{j(t)} \left(\frac{\underline{g} N_{j(t)}(t)}{(\underline{w}_1 - \epsilon_1 \underline{w}_2 - \epsilon) G_m}; \boldsymbol{\mu} \right) < (1+\epsilon) \frac{G_{m+1}}{G_m} \frac{\underline{g}}{\underline{w}_1 - \epsilon_1 \underline{w}_2 - \epsilon}.$$

This directly implies that

$$\begin{aligned} \frac{\underline{g} N_{j(t)}(t)}{(\underline{w}_1 - \epsilon_1 \underline{w}_2 - \epsilon) G_m} &< l_{j(t)} \left((1+\epsilon) \frac{G_{m+1}}{G_m} \frac{\underline{g}}{\underline{w}_1 - \epsilon_1 \underline{w}_2 - \epsilon}; \boldsymbol{\mu} \right) \\ &\leq (1+\epsilon_2) \frac{\underline{w}_{j(t)}}{\underline{w}_1}, \end{aligned}$$

where l_i is the inverse function of k_i defined in (19) and $\epsilon_2 > \epsilon_1$ is a constant such that $\epsilon_2 \to 0_+$ as $\epsilon \to 0_+$ and $\frac{G_{m+1}}{G_m} \to 1_+$. Then, we have for any $t \in Q_m$ that

$$N_{j(t)}(t) < (1+\epsilon_2)\frac{w_{j(t)}}{g}G_m.$$

In other words, if there exist $s \in Q_m$ such that

$$N_i(t) \geq (1+\epsilon_2)\frac{w_i}{g}G_m,$$

then only $j(s) \neq 1$ occurs for $t \in [s, \psi_{m+1})$, which implies that such arm i will be played only when TE occurs until ψ_{m+1}. Therefore, for $t \in Q_m$

$$\begin{aligned} N_i(t) &\leq \max\left(N_i(\psi_m, (1+\epsilon_2)\frac{w_i}{g}G_m\right) + M_{i,m} \\ &\leq (1+\epsilon_2)\frac{w_i}{g}G_m + M_i(\psi_{m+1}) \\ &\leq (1+\epsilon_2)\frac{w_i}{g}G_m + \frac{\epsilon}{g}G_m. \end{aligned}$$

(4) Upper bound on the optimal arm Here, let us assume that there exists $t' \in Q_m$ such that $N_1(t') \geq (1+\epsilon)(1+\epsilon_2)\frac{w_1}{g}G_m$. If there exists no such t', then one can directly obtain that $N_1(t) \leq (1+\epsilon)(1+\epsilon_2)\frac{w_1}{g}G_m$ for all $t \in Q_m$.

Since $N_{j(t)}(t) < (1+\epsilon_2)\frac{w_{j(t)}}{g}G_m$ holds from (G.1), then for any $t \in [t', \psi_{m+1})$

$$\begin{aligned} \frac{N_{j(t)}(t)}{N_1(t)} &< \frac{1}{1+\epsilon}\frac{w_{j(t)}}{w_1} = \frac{1}{1+\epsilon}\frac{z_{j(t)}}{1-z_{j(t)}} \\ &\leq \frac{z_{j(t),t}}{1-z_{j(t),t}}, \qquad \text{by (41)} \end{aligned}$$

which implies that $z^t_{j(t)} < z_{j(t),t} \leq z^*_{j(t),t}$. Since BC-TE plays the optimal arm 1 if $z_{j(t),t} \geq z^*_{j(t),t}$, only $i(t) = j(t)$ is possible unless TE occurs until ψ_{m+1}. Therefore, for $t \in Q_m$, it holds that

$$\begin{aligned} N_1(t) &\leq \max\left(N_1(\psi_m), (1+\epsilon)(1+\epsilon_2)\frac{w_1}{g}G_m\right) + M_{1,m} \\ &\leq (1+\epsilon)(1+\epsilon_2)\frac{w_1}{g}G_m + M_1(\psi_{m+1}) \\ &\leq (1+\epsilon_3)\frac{w_1}{g}G_m + \frac{\epsilon}{g}G_m, \end{aligned}$$

where ϵ_3 is a constant such that $(1+\epsilon)(1+\epsilon_2) = 1+\epsilon_3$. One can see that $\epsilon_3 \to 0_+$ as $\epsilon \to 0_+$ and $\frac{G_{m+1}}{G_m} \to 1_+$.

(5) Conclusion In summary, for any $t \in [\psi_m, \psi_{m+1})$, the results in (1)–(4) imply that for any $i \in [K]$:

$$N_i(t) \leq (1+\epsilon_3)\frac{w_i}{g}G_m + \frac{\epsilon}{g}G_m. \tag{50}$$

G.2. Sample complexity

From the upper bound on the number of plays for each arm in (50), for any $m \in \mathbb{N}$,

$$\begin{aligned}\psi_m = \sum_{i=1}^{K} N_i(\psi_m) &\leq \sum_{i=1}^{K}(1+\epsilon_3)\frac{w_i}{\underline{g}}G_m + \frac{\epsilon}{\underline{g}}G_m \\ &= (1+\epsilon_3)\frac{1}{\underline{g}}G_m + \frac{K\epsilon}{\underline{g}}G_m,\end{aligned}$$

which implies that

$$\frac{\underline{g}\psi_m}{(1+\epsilon_3+K\epsilon)} \leq G_m.$$

Therefore, the stopping time τ_δ in (46) can be written as

$$\begin{aligned}\tau_\delta &\leq T_B + \inf\left\{\psi_m : \frac{1}{1+\epsilon}G_m \geq \beta(\psi_m,\delta)\right\} \\ &\leq T_B + \inf\left\{\psi_m : \frac{1}{1+\epsilon}\frac{\underline{g}\psi_m}{(1+\epsilon_3+K\epsilon)} \geq \beta(\psi_m,\delta)\right\} \\ &\leq T_B + \inf\left\{\psi_m : \frac{\underline{g}\psi_m}{(1+\epsilon_4)} \geq \log\left(\frac{Ct^\alpha}{\delta}\right)\right\},\end{aligned}$$

for some $\epsilon_4 > \epsilon_3$ satisfying $\epsilon_4 \to 0_+$ as $\epsilon \to 0_+$ and $\frac{G_{m+1}}{G_m} \to 1_+$ and constants C and $\alpha \in [1, e/2]$ considered in Section 2.3. Then, by Lemma 18

$$\tau_\delta \leq T_B + \frac{\alpha}{\underline{g}}(1+\epsilon_4)\left[\log\left((1+\epsilon_4)^\alpha\frac{Ce}{\delta\underline{g}^\alpha}\right) + \log\log\left((1+\epsilon_4)^\alpha\frac{C}{\delta\underline{g}^\alpha}\right)\right].$$

Therefore, by taking expectations, we can obtain that

$$\limsup_{\delta\to 0}\frac{\mathbb{E}[\tau_\delta]}{\log(1/\delta)} \leq \frac{\alpha(1+\epsilon_4)}{\underline{g}}$$

since $\mathbb{E}[T_B]$ is finite from Theorem 2. Letting $\epsilon \to 0$ and setting $\frac{G_{m+1}}{G_m} \to 1$ conclude the proof. ■

G.3. Proof of Lemma 19: Non-decreasing objective function

Proof of Lemma 19 From the relation with f_i and h_i in (21), we can rewrite the function $tf_i(\boldsymbol{w}^t;\boldsymbol{\mu})$ as

$$tf_i(\boldsymbol{w}^t;\boldsymbol{\mu}) = (N_1(t)+N_i(t))h_i\left(\frac{N_i(t)}{N_1(t)+N_i(t)};\boldsymbol{\mu}\right).$$

Recall that $h_i(z;\boldsymbol{\mu})$ is a concave function with respect to $z \in [0,1]$ and $h_i(0;\boldsymbol{\mu}) = h_i(1;\boldsymbol{\mu}) = 0$ for any $i \neq 1$. For any $i \neq 1$, let us consider three possible cases (1) $i(t) = 1$, (2) $i(t) = i$, and (3) $i(t) \notin \{1, i\}$.

(1) When the optimal arm is played When $i(t) = 1$ holds, for any $i \neq 1$

$$(t+1)f_i(\boldsymbol{w}^{t+1};\boldsymbol{\mu}) = (N_1(t)+N_i(t)+1)h_i\left(\frac{N_i(t)}{N_1(t)+N_i(t)+1};\boldsymbol{\mu}\right).$$

From the concavity of h_i, we obtain that

$$\begin{aligned} h_i\left(\frac{N_i(t)}{N_1(t)+N_i(t)+1};\boldsymbol{\mu}\right) &= h_i\left(\frac{N_i(t)}{N_1(t)+N_i(t)}\frac{N_1(t)+N_i(t)}{N_1(t)+N_i(t)+1};\boldsymbol{\mu}\right) \\ &\geq \frac{N_1(t)+N_i(t)}{N_1(t)+N_i(t)+1}h_i\left(\frac{N_i(t)}{N_1(t)+N_i(t)};\boldsymbol{\mu}\right) \\ &\quad + \frac{1}{N_1(t)+N_i(t)+1}h_i(0;\boldsymbol{\mu}), \end{aligned}$$

which implies

$$\begin{aligned} &(N_1(t)+N_i(t)+1)h_i\left(\frac{N_i(t)}{N_1(t)+N_i(t)+1};\boldsymbol{\mu}\right) \\ &\qquad \geq (N_1(t)+N_i(t))h_i\left(\frac{N_i(t)}{N_1(t)+N_i(t)};\boldsymbol{\mu}\right) = tf_i(\boldsymbol{w}^t;\boldsymbol{\mu}). \end{aligned}$$

This concludes the case when $i(t) = 1$.

(2) When the suboptimal arm is played When $i(t) = i$ holds,

$$(t+1)f_i(\boldsymbol{w}^{t+1};\boldsymbol{\mu}) = (N_1(t)+N_i(t)+1)h_i\left(\frac{N_i(t)+1}{N_1(t)+N_i(t)+1};\boldsymbol{\mu}\right).$$

By the concavity, again, we obtain that

$$\begin{aligned} &h_i\left(\frac{N_i(t)+1}{N_1(t)+N_i(t)+1};\boldsymbol{\mu}\right) \\ &\quad = h_i\left(\frac{N_i(t)}{N_1(t)+N_i(t)}\frac{N_1(t)+N_i(t)}{N_1(t)+N_i(t)+1} + \frac{1}{N_1(t)+N_i(t)+1};\boldsymbol{\mu}\right) \\ &\quad \geq \frac{N_1(t)+N_i(t)}{N_1(t)+N_i(t)+1}h_i\left(\frac{N_i(t)}{N_1(t)+N_i(t)};\boldsymbol{\mu}\right) + \frac{1}{N_1(t)+N_i(t)+1}h_i(1;\boldsymbol{\mu}) \\ &\quad = \frac{N_1(t)+N_i(t)}{N_1(t)+N_i(t)+1}h_i\left(\frac{N_i(t)}{N_1(t)+N_i(t)};\boldsymbol{\mu}\right), \end{aligned}$$

which concludes the case when $i(t) = i$.

(3) When the other suboptimal arms are played When $i(t) \notin \{1, i\}$, $N_1(t+1) = N_1(t)$ and $N_i(t+1) = N_i(t+1)$ holds. Therefore, $(t+1)f_i(\boldsymbol{w}^{t+1};\boldsymbol{\mu}) = tf_i(\boldsymbol{w}^t;\boldsymbol{\mu})$ holds, which concludes the case when $i(t) \neq 1, i$. ■

References

Shipra Agrawal and Navin Goyal. Analysis of Thompson sampling for the multi-armed bandit problem. In *Annual Conference on Learning Theory*. PMLR, 2012.

Rémy Degenne, Wouter M Koolen, and Pierre Ménard. Non-asymptotic pure exploration by solving games. In *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2019.

Aurélien Garivier and Emilie Kaufmann. Optimal best arm identification with fixed confidence. In *Annual Conference on Learning Theory*. PMLR, 2016.

Emilie Kaufmann, Nathaniel Korda, and Rémi Munos. Thompson sampling: An asymptotically optimal finite-time analysis. In *International Conference on Algorithmic Learning Theory*. Springer, 2012.

Nathaniel Korda, Emilie Kaufmann, and Remi Munos. Thompson sampling for 1-dimensional exponential family bandits. In *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2013.