# Electronic Companion to Adaptivity and Confounding in Multi-Armed Bandit Experiments

Chao Qin and Daniel Russo

May 20, 2022

This paper provides an electronic companion to Qin and Russo [2022]. It contains the proof of a proposition regarding the asymptotic behavior of deconfounded Thompson sampling. The proof is omitted from that paper, since it is closely related to analyses in Russo [2020], Qin et al. [2017], and Shang et al. [2020]. While the analysis here is quite technical, we have tried to give a relatively clear proof template that could be useful in future work.

## 1 Statement of the proposition

Our goal is to prove the following proposition. Part 2 is Proposition 7. Most of our work is to prove part 1, however, as the second part follows from the first.

**Proposition 1.** *Suppose DTS is applied with $\beta_t$ set by Algorithm 2 and condition on the event that $\theta = \theta_0$ for some $\theta_0 \in \Theta$. Then, the following properties hold.*

1. *For every $\epsilon > 0$, there exists a random time $T$ with $\mathbb{E}[T \mid \theta = \theta_0] < \infty$ such that for each $t \geqslant T$,*

$$|p_{t,i} - p_i^*(\theta_0)| \leqslant \epsilon \quad \textit{for all} \;\; i \in [k].$$

2. *For every $\epsilon > 0$, there exists a random time $T$ with $\mathbb{E}[T \mid \theta = \theta_0] < \infty$ such that for each $t \geqslant T$,*

$$\left| \frac{Z^2_{t,\hat{I}_t,i}}{t} - 2\Gamma^{-1}_{\theta_0} \right| \leqslant \epsilon \quad \textit{for all} \;\; i \in [k].$$

The proof extends the analysis in Qin et al. [2017] and Shang et al. [2020] for problems without contexts to those where context vectors are drawn i.i.d. from some context distribution. In addition, we provide a new proof template of proving sufficient exploration for bandit algorithms, which is of independent interest.

## 2 Outline

The proofs from Section 8 are organized as follows.

1. Section 3 provides the notation used in the proofs.
2. Section 4 formalizes a notion of convergence called "strong convergence" and its properties that are frequently used in the proof of Proposition 1.
3. Section 5 introduces the maximal inequalities for controlling stochastic contexts, random observations and randomized action selections, and uses them to derive the accuracy and confidence of beliefs with sufficient samples.

4. Section 6 shows that under DTS, each arm is sufficiently explored.
5. Section 7 proves that under DTS, the empirical proportions allocated to each arm strongly converge to the optimal context-independent sampling frequencies.
6. Section 8 completes the proofs in Section 5.
7. Section 9 includes technical lemmas used in the proofs.

# 3 Notation

To simplify the exposition, instead of $\theta_0$, we use $\theta$ to denote the fixed but unknown problem parameter, and denote $\tilde{\theta}$ as a random vector drawn from some distribution of interest (e.g., posterior beliefs). In addition, define $\mathbb{P}_t(\cdot) = \mathbb{P}(\cdot|H_t)$ as the posterior measure, under which $\theta$ is a random variable, and $\mathbb{P}_\theta = \mathbb{P}(\cdot \mid \theta)$.

**Notation for minimum gap.** We define the minimum value between the expected rewards of two arms under the population distribution:

$$\Delta_{\min}(\theta) \triangleq \min_{i \neq j} |\mu(\theta, i, w) - \mu(\theta, j, w)|.$$

Under the parameter class in in Equation (12), $\Delta_{\min}(\theta) > 0$.

**Two measures of cumulative effort.** For $(t, i) \in \mathbb{N} \times [k]$, we denote the number of samples allocated to sampling arm $i$ before time $t$ as

$$N_{t,i} \triangleq \sum_{\ell=1}^{t-1} \mathbb{1}\{I_\ell = i\}. \tag{1}$$

For randomized algorithms such as DTS, we define the probability of measuring arm $i$ at time $t$ as $\psi_{t,i} \triangleq \mathbb{P}(I_t = i \mid H_t)$ and an alternative measure of the the cumulative effort

$$\Psi_{t,i} \triangleq \sum_{\ell=1}^{t-1} \psi_{\ell,i}. \tag{2}$$

Note that under DTS, the probability of measuring arm $i$ at time $t$ has the following expression:

$$\psi_{t,i} = \alpha_{t,i} \left( \beta_t + (1 - \beta_t) \sum_{j \neq i} \frac{\alpha_{t,j}}{1 - \alpha_{t,j}} \right).$$

**Uniformly strictly bounded tuning sequence.** We say that the sequence $\{\beta_t\}_{t \in \mathbb{N}}$ is *uniformly strictly bounded* if there exists a constant $\beta_{\min} > 0$ such that with probability 1,

$$\inf_{t \in \mathbb{N}} \min\{\beta_t, 1 - \beta_t\} \geqslant \beta_{\min}. \tag{3}$$

As shown in Lemma 17, the tuning sequence given by Algorithm 2 is uniformly strictly bounded.

**Notation for context distribution.** Recall $\Lambda = \mathbb{E}[X_1 X_1^\top] \succ 0$ by Assumption 1. We define the smallest eigenvalue of $\sigma^{-2}\Lambda$ as $b_{\min} \triangleq \lambda_{\min}(\sigma^{-2}\Lambda) > 0$. Also Assumption 1 states that the context distribution has bounded support, i.e., there exists $b_{\max} > 0$ such that $\sigma^{-2}\|X_1\|^2 \leqslant b_{\max}$. This implies the largest eigenvalue of $\sigma^{-2}\Lambda$ is upper bounded by $b_{\max}$. Therefore,

$$\sigma^2 b_{\min} \leqslant \lambda_{\min}(\Lambda) \leqslant \lambda_{\max}(\Lambda) \leqslant \sigma^2 b_{\max}.$$

**Notation for prior covariance matrices.** Let $p_{\min}, p_{\max} > 0$ such that for any $i \in [k]$,

$$p_{\min} \leqslant \lambda_{\min}\left(\Sigma_{1,i}^{-1}\right) \leqslant \lambda_{\max}\left(\Sigma_{1,i}^{-1}\right) \leqslant p_{\max}.$$

# 4 Properties of strong convergence

Proposition 1 relies on bounding time until $p_{t,i}$ reaches and remains close to $p_i^*$. Here we formalize a corresponding notion of convergence, which we call "strong convergence".

Define the $p$-norm $\|X\|_p = \left(\mathbb{E}\left[|X|^p\right]\right)^{1/p}$ and let the space $\mathcal{L}^p$ consist of all measurable $X$ with $\|X\|_p < \infty$. We define a set of random variables that is 'light-tailed', in the sense that it is in $\mathcal{L}^p$ for any $p \geqslant 1$.

**Definition 1.** *For a real valued random variable $X$, we say $X \in \mathbb{M}$ if and only if $\|X\|_p < \infty$ for all $p \geqslant 1$. Equivalently, $\mathbb{M} = \cap_{p \geqslant 1} \mathcal{L}_p$.*

With this notion in place, we define a custom notion of convergence for random variables. To understand this, it is helpful to start with the usual definition of almost sure convergence. For a sequence of random variables $\{X_n\}$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we say $X_n \to x$ almost surely if

$$\mathbb{P}\left(\omega : \lim_{n\to\infty} X_n(\omega) = x\right) = 1.$$

If we explicitly write out the definition of a limit for a deterministic sequence, the condition for almost sure convergence becomes

$$\mathbb{P}\left(\omega : \forall \epsilon > 0\, \exists N(\omega) \in \mathbb{N} \text{ s.t. } \forall n \geqslant N(\omega)\ |X_n(\omega) - x| \leqslant \epsilon\right) = 1. \tag{4}$$

This definition says that random quantities (e.g. empirical arm means) converge to a neighborhood of their limit (e.g. population mean) *eventually*. A subtle issue for our analysis is that the expected time one needs to wait could be infinite; that is, one could have $\mathbb{E}[N(\omega)] = \infty$ in (4). To bound quantities like the expected stopping time of our best-arm algorithms, we rely on the following stronger notion of convergence.

**Definition 2.** *For a sequence of real valued random variables $\{X_n\}_{n\in\mathbb{N}}$ and a scalar $x \in \mathbb{R}$, we say $X_n \xrightarrow{\mathbb{M}} x$ if*

$$\textit{for all } \epsilon > 0 \textit{ there exists } N \in \mathbb{M} \textit{ such that for all } n \geqslant N,\ |X_n - x| \leqslant \epsilon.$$

*We say $X_n \xrightarrow{\mathbb{M}} \infty$ if*

$$\textit{for all } c > 0 \textit{ there exists } N \in \mathbb{M} \textit{ such that for all } n \geqslant N,\ X_n \geqslant c.$$

*Similarly, we say $X_n \xrightarrow{\mathbb{M}} -\infty$ if $-X_n \xrightarrow{\mathbb{M}} \infty$. For a sequence of random vectors $\{X_n\}_{n\in\mathbb{N}}$ taking values in $\mathbb{R}^d$ and a vector $x \in \mathbb{R}^d$, we say $X_n \xrightarrow{\mathbb{M}} x$ if $X_{n,i} \xrightarrow{\mathbb{M}} x_i$ for all $i \in [d]$. Similarly, a sequence of random matrices converges strongly to a fixed matrix if each element converges strongly.*

To show the asymptotic convergence of the proposed algorithm's sampling proportions on any sample path, we could rely on a variety of powerful asymptotic tools to simplify our arguments. Our aim is to develop a strategy for establishing this mode of convergence that inherits some of the elegance of sample path analysis. To this end, we will develop here a number of convenient properties of the class of random variables $\mathbb{M}$ and the strong convergence notion in Definition 2. We start by showing $\mathbb{M}$ is closed under many natural operations. The most notable exception is exponentiation: a Gaussian random variable $X \in \mathbb{M}$, but $e^{e^X} \notin \mathbb{M}$.

**Lemma 1** (Closedness of $\mathbb{M}$)**.** *Let any non-negative $X, Y \in \mathbb{M}$.*

*A)* $aX + bY \in \mathbb{M}$ *for any scalars* $a, b \in \mathbb{R}$.

*B)* $XY \in \mathbb{M}$.

*C)* $\max\{X, Y\} \in \mathbb{M}$.

*D)* $X^q \in \mathbb{M}$ *for any* $q \geqslant 0$.

*E) If* $g : \mathbb{R} \to \mathbb{R}$ *satisfies* $\sup_{x \in -[c,c]} |g(x)| < \infty$ *for all* $c \geqslant 0$ *and* $|g(x)| = O(|x|^q)$ *as* $|x| \to \infty$ *for some* $q \geqslant 0$, *then* $g(X) \in \mathbb{M}$.

*F) If* $g : \mathbb{R} \to \mathbb{R}$ *is continuous and* $|g(x)| = O(|x|^q)$ *as* $|x| \to \infty$ *for some* $q \geqslant 0$, *then* $g(X) \in \mathbb{M}$.

*Proof.* Parts A) and B) follow from subadditivity and submultiplicativity, respectively. Part C) holds due to $\max\{X, Y\} \leqslant X + Y$ and part A). Part D) follows from Hölder's inequality. To show part E) carefully, take constants $C_1$ and $C_2$ such that $|g(x)| \leqslant C_2|x|^q$ for all $x$ with $|x| > C_1$, and thus $|g(X)| \leqslant \sup_{x \in [-C_1, C_1]} |g(x)| + C_2|X|^q$. Then Part E) holds due to part D). Part F) follows from part E), since continuous functions are bounded on compact sets. □

Clearly, many of these properties can be extended by induction to any finite collection of random variables in $\mathbb{M}$. For example, if $X_i \in \mathbb{M}$ for each $i \in [d]$ then

$$X_1 + \cdots + X_d \in \mathbb{M} \quad \text{and} \quad \max\{X_1, \ldots, X_d\} \in \mathbb{M}.$$

For example, when $d = 3$ we see that $X_1 + X_2 + X_3 = (X_1 + X_2) + X_3 \in \mathbb{M}$ and $\max\{X_1, X_2, X_3\} = \max\{X_1, \max\{X_2, X_3\}\} \in \mathbb{M}$. Proceeding by induction in this manner establishes the claim. We can also repeatedly compose many of the operations described in the above lemma. For example, by parts C), D) and F), we have $\log(1 + (\max\{X, Y\})^2) \in \mathbb{M}$ whenever $X, Y \in \mathbb{M}$. We will freely use such properties in our analysis.

Leveraging the properties of $\mathbb{M}$ established in the previous lemma allows for conclusions about the notion of strong convergence in Definition 2. First, we show an equivalence between pointwise convergence and convergence in the maximum norm.

**Lemma 2.** *For a sequence of random vectors* $\{X_n\}_{n \in \mathbb{N}}$ *taking values in* $\mathbb{R}^d$ *and a vector* $x \in \mathbb{R}^d$, *if* $X_n \xrightarrow{\mathbb{M}} x$, *then* $\|X_n - x\|_\infty \xrightarrow{\mathbb{M}} 0$. *Equivalently, if* $X_{n,i} \xrightarrow{\mathbb{M}} x_i$ *for all* $i \in [d]$, *then for all* $\epsilon > 0$, *there exists* $N \in \mathbb{M}$ *such that* $n \geqslant N$ *implies* $|X_{n,i} - x_i| \leqslant \epsilon$ *for all* $i \in [d]$.

*Proof.* Fix $\epsilon > 0$. For each $i \in [d]$, choose $N_i$ such that $n \geqslant N_i$ implies $|X_{n,i} - x_i| \leqslant \epsilon$. Then defining $N \triangleq \max\{N_1, N_2, \ldots N_d\}$, we have that $n \geqslant N$ implies $\|X_n - x\|_\infty \leqslant \epsilon$. That $N \in \mathbb{M}$ follows from Lemma 1. □

Now we note a continuous mapping theorem for this stronger notion of convergence.

**Lemma 3** (Continuous Mapping)**.** *For random sequence* $\{X_n\}_{n \in \mathbb{N}}$ *and* $x$ *taking values in a normed vector space:*

*A) If* $g$ *is continuous at* $x$, *then* $X_n \xrightarrow{\mathbb{M}} x$ *implies* $g(X_n) \xrightarrow{\mathbb{M}} g(x)$.

*B) If the range of function* $g$ *belongs to* $\mathbb{R}$ *and* $g(y) \to \infty$ *as* $y \to \infty$, *then* $X_n \xrightarrow{\mathbb{M}} \infty$ *implies* $g(X_n) \xrightarrow{\mathbb{M}} \infty$.

*Proof.* We prove only part A) and the proof of part B) is similar. Fix any $\epsilon > 0$. By continuity, there is some $\delta > 0$ such that $\|x' - x\| \leqslant \delta$ implies $|g(x') - g(x)| \leqslant \epsilon$. Since $X_n \xrightarrow{\mathbb{M}} x$, there exists $N \in \mathbb{M}$ such that $n \geqslant N$ implies $\|X_n - x\| \leqslant \delta$. Hence, for $n \geqslant N$ we have $|g(X_n) - g(x)| \leqslant \epsilon$ as desired. Here $\|\cdot\|$ can be any norm on the vector space. □

# 5 Accuracy and confidence of beliefs in terms of maximal inequalities

In this section, we first introduce the maximal inequalities for controlling stochastic contexts, random observations and randomized action selections. Then we can derive the accuracy and confidence of beliefs in terms of these maximal inequalities.

## 5.1 Maximal inequalities

To control the impact of random contexts and observation noises, we define the following path-dependent random variable

$$W_1 \triangleq \sup_{(t,i)\in\mathbb{N}\times[k]} \frac{|m_{t,i} - \mu(\theta, i, w)|}{s_{t,i}\sqrt{\log(N_{t,i}+e)}}. \tag{5}$$

where we call the numerator *prediction error*. The prediction error is close to the standard error of the estimate if observations were i.i.d. The term $s_{t,i}$ in the denominator is the posterior standard deviation, which captures the natural scale of error we'd might expect at a single time period. The term $\sqrt{\log(N_{t,i}+e)}$ corrects for the fact that we are maximizing over times $t$.

Similarly, to control the impact of randomness in action selection, we introduce the path-dependent random variable

$$W_2 \triangleq \sup_{(t,i)\in\mathbb{N}\times[k]} \frac{|N_{t,i} - \Psi_{t,i}|}{\sqrt{(t+1)\log(t+e^2)}} \tag{6}$$

where two measures of cumulative effort $N_{t,i}$ and $\Psi_{t,i}$ are defined in Equations (1) and (2).

Finally, the action selection is context independent. By Assumption 1, contexts are drawn i.i.d. with second moment $\Lambda = \mathbb{E}\left[X_1 X_1^\top\right] \succ 0$. We would expect that when $N_{t,i}$ is large, the posterior covariance matrix

$$\frac{1}{N_{t,i}+1}\Sigma_{t,i}^{-1} = \frac{1}{N_{t,i}+1}\left[\Sigma_{1,i}^{-1} + \sigma^{-2}\sum_{\ell=1}^{t-1}\mathbb{1}\{I_\ell = i\}X_\ell X_\ell^\top\right] \to \sigma^{-2}\Lambda.$$

To control the impact of i.i.d. contexts in updating the posterior covariance matrices, we define the following path-dependent random variable

$$W_3 \triangleq \sup_{(t,i)\in\mathbb{N}\times[k]} \frac{\left\|\Sigma_{t,i}^{-1} - A_{t,i}^{-1}\right\|}{\sqrt{(N_{t,i}+1)\log(N_{t,i}+e)}} \quad \text{where} \quad A_{t,i}^{-1} \triangleq \sigma^{-2}\Lambda(N_{t,i}+1). \tag{7}$$

The next lemma ensures that these maximal deviations are almost surely finite and light-tailed, in the sense that all their moments are finite.

**Lemma 4.** *Conditioned on $\theta$, the random variables $W_1, W_2$ and $W_3$ are elements of $\mathbb{M}$. That is $\mathbb{E}[|W_i|^p \mid \theta = \theta_0] < \infty$ for any $\theta_0$ and any $p \geqslant 0$.*

The detailed proof of this result is presented in Section 8. By the definition of $W_2$ and this result, we have the following corollary.

**Corollary 1.** *For each arm $i \in [k]$,*

$$\frac{N_{t,i}}{t} - \frac{\Psi_{t,i}}{t} \xrightarrow{\mathbb{M}} 0.$$

## 5.2 Accuracy and confidence of beliefs with sufficient samples

We first provide the upper and lower bounds of $s_{t,i}^2$, the posterior variance of the average reward of arm $i$ at time $t$, in terms of $W_3$.

**Lemma 5.** *For any* $(t,i) \in \mathbb{N} \times [k]$,

$$\frac{\|X_{\text{pop}}\|^2}{b_{\max} N_{t,i} + p_{\max}} \leqslant s_{t,i}^2 \leqslant \frac{\|X_{\text{pop}}\|^2 (W_3 + p_{\min})}{p_{\min} b_{\min} (N_{t,i}+1)^{1/4}}.$$ [1]

*Proof.* The first inequality follows from

$$\|\Sigma_{t,i}\| = \left\| \left( \Sigma_{1,i}^{-1} + \sigma^{-2} \sum_{\ell=1}^{t-1} \mathbb{1}\{I_\ell = i\} X_\ell X_\ell^\top \right)^{-1} \right\| \geqslant \frac{1}{p_{\max} + b_{\max} N_{t,i}}$$

where we use $\sigma^{-2}\|X_\ell X_\ell^\top\| = \sigma^{-2}\|X_\ell\|^2 \leqslant b_{\max}$. This completes the proof of the lower bound.

Now we are going to show the upper bound. By the submultiplicative property of the spectral norm,

$$\begin{aligned}
\|\Sigma_{t,i} - A_{t,i}\| &= \left\| \Sigma_{t,i} A_{t,i} \left( A_{t,i}^{-1} - \Sigma_{t,i}^{-1} \right) \right\| \\
&\leqslant \|\Sigma_{t,i}\| \, \|A_{t,i}\| \left\| \Sigma_{t,i}^{-1} - A_{t,i}^{-1} \right\| \\
&\leqslant \frac{W_3}{p_{\min} b_{\min}} \sqrt{\frac{\log(N_{t,i}+e)}{N_{t,i}+1}} \\
&\leqslant \frac{W_3}{p_{\min} b_{\min} (N_{t,i}+1)^{1/4}}
\end{aligned}$$

where the second inequality follows from

$$\|\Sigma_{t,i}\| \leqslant \|\Sigma_{1,i}\| \leqslant \frac{1}{p_{\min}} \quad \text{and} \quad \|A_{t,i}\| = \frac{\|\sigma^2 \Lambda^{-1}\|}{N_{t,i}+1} \leqslant \frac{1}{b_{\min}(N_{t,i}+1)}$$

as well as the definition of $W_3$; the last inequality uses $\log(x+e) \leqslant (x+1)^{1/2}$ for $x \geqslant 0$. Then by triangle inequality,

$$\begin{aligned}
\|\Sigma_{t,i}\| &\leqslant \|\Sigma_{t,i} - A_{t,i}\| + \|A_{t,i}\| \\
&\leqslant \frac{W_3}{p_{\min} b_{\min} (N_{t,i}+1)^{1/4}} + \frac{1}{b_{\min}(N_{t,i}+1)} \\
&\leqslant \frac{W_3 + p_{\min}}{p_{\min} b_{\min} (N_{t,i}+1)^{1/4}}.
\end{aligned}$$

where the last inequality uses $(x+1)^{-1} \leqslant (x+1)^{-1/4}$ for $x \geqslant 0$. This completes the proof. □

By the definition of $W_1$ and Lemma 5, we can provide the following upper bound for each prediction error in terms of the corresponding number of samples.

**Corollary 2.** *For any* $(t,i) \in \mathbb{N} \times [k]$,

$$|m_{t,i} - \mu(\theta, i, w)| \leqslant W_1 \|X_{\text{pop}}\| \sqrt{\frac{(W_3 + p_{\min}) \log(N_{t,i}+e)}{p_{\min} b_{\min} (N_{t,i}+1)^{1/4}}}.$$

*Proof.* By the definition of $W_1$,

$$|m_{t,i} - \mu(\theta, i, w)| \leqslant W_1 s_{t,i} \sqrt{\log(N_{t,i}+e)} \leqslant W_1 \|X_{\text{pop}}\| \sqrt{\frac{(W_3 + p_{\min}) \log(N_{t,i}+e)}{p_{\min} b_{\min} (N_{t,i}+1)^{1/4}}}$$

[1] In the denominator of the upper bound, the exponent $1/4$ is not essential. It is just an arbitrarily chosen number in $(0, 1/2)$.

where the last inequality uses Lemma 5. □

The above result implies the prediction error of the reward of an arm can be as small as possible if the number of samples allocated to this arm is large enough, which is formalized in the next result.

**Lemma 6.** *For all $\epsilon > 0$, there exists $S_\epsilon \in \mathbb{M}$ such that for all $(t, i) \in \mathbb{N} \times [k]$,*

$$N_{t,i} \geqslant S_\epsilon \quad \Longrightarrow \quad |m_{t,i} - \mu(\theta, i, w)| \leqslant \epsilon.$$

*Proof.* The result follows from Corollary 2 since there exists $S_\epsilon \in \mathbb{M}$ such that for all $(t, i) \in \mathbb{N} \times [k]$,

$$N_{t,i} \geqslant S_\epsilon \quad \Longrightarrow \quad W_1 \left\| X_{\text{pop}} \right\| \sqrt{\frac{(W_3 + p_{\min}) \log (N_{t,i} + e)}{p_{\min} b_{\min} (N_{t,i} + 1)^{1/4}}} \leqslant \epsilon.$$

□

The next lemma says that if at least two arms have been sampled a sufficient number of times, then the posterior probability assigned to the worse arm is exponentially small.

**Lemma 7.** *There exists $S \in \mathbb{M}$ such that for all $t \in \mathbb{N}$ and arms $i \neq j$ such that $m_{t,j} - m_{t,i} \geqslant 0$,*

$$\min \left\{ N_{t,i}, N_{t,j} \right\} \geqslant S \quad \Longrightarrow \quad m_{t,j} - m_{t,i} \geqslant \Delta_{\min}/2 \quad \text{and} \quad \alpha_{t,i} \leqslant e^{-Q_{t,i,j}^2/2}$$

*where $\Delta_{\min} = \Delta_{\min}(\theta) = \min_{i \neq j} |\mu(\theta, i, w) - \mu(\theta, j, w)| > 0$ and*

$$Q_{t,i,j} \triangleq \frac{\Delta_{\min}}{2 \left\| X_{\text{pop}} \right\| \left( \frac{W_3 + p_{\min}}{p_{\min} b_{\min}} \right)^{1/2} \left( \frac{1}{N_{t,i}^{1/4} + 1} + \frac{1}{N_{t,j}^{1/4} + 1} \right)^{1/2}}.$$

*Proof.* Fix $t \in \mathbb{N}$ and $i \neq j$ such that $m_{t,j} - m_{t,i} \geqslant 0$. Let $\tilde{\theta}_t$ be a sample drawn independently from $\mathbb{P}_t(\theta \in \cdot)$. Then

$$\alpha_{t,i} \leqslant \mathbb{P}_t \left( \mu(\tilde{\theta}_t, i, w) - \mu(\tilde{\theta}_t, j, w) \geqslant 0 \right) = \Phi \left( -\frac{m_{t,j} - m_{t,i}}{\sqrt{s_{t,i}^2 + s_{t,j}^2}} \right).$$

By Lemma 5,

$$\sqrt{s_{t,i}^2 + s_{t,j}^2} \leqslant \left\| X_{\text{pop}} \right\| \left( \frac{W_3 + p_{\min}}{p_{\min} b_{\min}} \right)^{1/2} \left( \frac{1}{N_{t,i}^{1/4} + 1} + \frac{1}{N_{t,j}^{1/4} + 1} \right)^{1/2}.$$

Next we lower bound $m_{t,j} - m_{t,i}$.

Set $\epsilon = \Delta_{\min}/4$ and take $S_{\Delta_{\min}/4}$ in Lemma 6. If $\min \left\{ N_{t,i}, N_{t,j} \right\} \geqslant S_{\Delta_{\min}/4}$, then

$$\begin{aligned}
m_{t,j} - m_{t,i} &= \left| m_{t,j} - m_{t,i} \right| \\
&= \left| (\mu(\theta, j, w) - \mu(\theta, i, w)) + \left( m_{t,j} - \mu(\theta, j, w) \right) - (m_{t,i} - \mu(\theta, i, w)) \right| \\
&\geqslant |\mu(\theta, j, w) - \mu(\theta, i, w)| - \left| m_{t,j} - \mu(\theta, j, w) \right| - |m_{t,i} - \mu(\theta, i, w)| \\
&\geqslant \Delta_{\min} - \Delta_{\min}/4 - \Delta_{\min}/4 \\
&= \Delta_{\min}/2
\end{aligned}$$

where the first inequality uses the triangle inequality and the second inequality follows from Lemma 6, and

thus

$$\alpha_{t,i} \leqslant \Phi\left(-\frac{m_{t,j} - m_{t,i}}{\sqrt{s_{t,i}^2 + s_{t,j}^2}}\right) \leqslant \Phi\left(-Q_{t,i,j}\right) \leqslant e^{-Q_{t,i,j}^2/2}$$

where the last inequality uses Lemma 25. Taking $S = S_{\Delta_{\min}/4}$ completes the proof. □

# 6 Proof of sufficient exploration

As the first step to proving Proposition 1, we first show that under DTS, each arm is sufficiently explored. This lets us apply results of an asymptotic style — for example results on the concentration of beliefs in Section 7— throughout the remaining proofs of Proposition 1.

Note that, were we only interested in asymptotic results, we could guarantee a result like this by simply interleaving random exploration phases in between phases where DTS is applied. This simplifies proofs and is done in other papers. Here we are trying ot keep the algorithm simple, at the expense of considerable effort in the proofs. **The details of this proof are not used elsewhere, so the reader may, on first reading, skip certain details.**

**Proposition 2.** *Under DTS applied with any uniformly strictly bounded $\{\beta_t\}$, there exists $T \in \mathbb{M}$ such that for all $t \geqslant T$,*

$$\min_{i \in [k]} N_{t,i} \geqslant t^{3/16}.$$[2]

To prove this proposition, we need two major parts, and each of them requires a sequence of results. At a high level the proof proceeds as follows.

1. In part 1, we define a set of insufficiently sampled arms. The complement of this set, the sufficiently sampled set, contains arms that have been measured much more than the least sampled arm and for which further measurement is not valuable at the moment.

2. In part 2, we show one of the two most promising arms – i.e. one of the two arms with highest posterior probability of being optimal – is always insufficiently sampled.

3. In part 3, we use this to show DTS assigns a constant probability to measuring an arm from the insufficiently sampled set. This is intuitive, as DTS is a randomized algorithm that tends to favor the arms identified in part 2.

4. Part 4 uses the the pigeonhole principle. If we're frequently measuring the insufficiently sampled arms, then on average total the arms in that set must have received high measurement effort in the past.

5. Part 5 is mostly technical. We rewrite our results about the "measurement effort" assigned to arm $i$, captured in terms of the chance of measuring an arm as in $\Psi_{t,i}$, in terms of the actual number of times it was measured $N_{t,i}$.

6. In part 6, we observe that the definition of the insufficiently sampled set, requires that no arm in that set is measured too much more than others. If parts 4/5 show some arms in the set are measured a lot, they must *all* have been measured a lot. In particular, the least sampled arm is measured frequently.

## 6.1 Part 1: Definition of insufficiently sampled arms

We define a set of arms that are "insufficiently sampled", in the sense that they have not been sampled too much more than the least sampled arm. The intuition is that further sampling of *sufficiently sampled arms*, would not be valuable, as the quality of these arms is comparatively very well understood.

[2]Similar to the exponent 1/4 in Lemma 5, the exponent 3/16 here is also arbitrarily chosen. We are not interested in the growth rate of $N_{t,i}$ (which later results reveal is linear), but in having as long as it is of the order $t^c$ for some $c > 0$.

Precisely, for $t \in \mathbb{N}$ and random variable $R \geqslant 1$ (almost surely), we define the insufficiently sampled set $U_t^R$ based on the least sampled arm $J_t^{\min}$:

$$U_t^R \triangleq \left\{ i \in [k] \,:\, N_{t,i}^{1/4} \leqslant R \left( N_{t,J_t^{\min}} + 1 \right) \right\} \quad \text{where} \quad J_t^{\min} \in \arg\min_{j \in [k]} N_{t,j}. \tag{8}$$

Note that $R \geqslant 1$ ensures $J_t^{\min} \in U_t^R$ for all $t \in \mathbb{N}$. Here $1/4$ is the same exponent in Lemma 5, which was arbitrarily chosen among $(0, 1/2)$. The dependence of this set on a (yet to be defined) random scalar $R$ is a subtlety required for establishing strong convergence. We will work with arbitrary $R$ for now and later need to show an adequate choice of $R$ exists as some implicit function of $W_1, W_2$ and $W_3$.

## 6.2 Part 2: An insufficiently sampled arm is always one of the two most promising arms

We define the following auxiliary arms, representing the two "most promising arms":

$$J_t^{(1)} \in \arg\max_{i \in [k]} \alpha_{t,i} \quad \text{and} \quad J_t^{(2)} \in \arg\max_{i \neq J_t^{(1)}} \alpha_{t,i}. \tag{9}$$

The first part of the proof focuses on showing that either $J_t^{(1)}, J_t^{(2)}$ or both belong to the insufficiently sampled set parametrized by some $R \in \mathbb{M}$.

**Lemma 8.** *There exists $R \in \mathbb{M}$ such that for any $t \in \mathbb{N}$,*

$$J_t^{(1)} \notin U_t^R \implies J_t^{(2)} \in U_t^R.$$

To prove this result, we are concerned about the case where both of the most promising arms are sufficiently sampled. The core of the proof is then contained in the next technical lemma, which effectively rules out this concerning case. It shows that only one sufficiently sampled arm could ever be more promising than all insufficiently sampled arms: the arm with highest estimated mean among those that were sufficiently sampled.

To see the intuition behind this result, let set $\overline{U_t^R}$ be the complement of insufficiently sampled set $U_t^R$ and imagine that sufficiently sampled arms in $\overline{U_t^R}$ had in fact been sampled an infinite number of times. In that case, the posterior mean $m_{t,i}$ would be the arm true quality $\mu(\theta, i, w)$ and the posterior variance $s_{t,i}$ would equal zero. Under the parameter class in Equation (12), each $\mu(\theta, i, w)$ is unique, so every sufficiently sampled arm in $\overline{U_t^R}$ other than $\arg\max_{i \in \overline{U_t^R}} \mu(\theta, i, w)$ would have a zero chance of being optimal, while insufficiently sampled arms in $U_t^R$, having more uncertain quality, would still be believed to have some chance of being optimal. The proof is a technical version of this argument that uses that uncertainty about sufficiently sampled arms is *an order of magnitude lower* than uncertainty about some other arms (namely $J_t^{\min}$).

**Lemma 9.** *There exists $R \in \mathbb{M}$ such that for any $t \in \mathbb{N}$, if $\overline{U_t^R}$ is nonempty,*

*A)* $\arg\max_{i \in \overline{U_t^R}} m_{t,i}$ *is unique, and let* $\overline{I_t^R} \triangleq \arg\max_{i \in \overline{U_t^R}} m_{t,i}$

*B) for any arm* $i \in \overline{U_t^R} \setminus \left\{ \overline{I_t^R} \right\}$, $\alpha_{t,i} < \max_{j \in U_t^R} \alpha_{t,j}$.

*Proof.* Fix $t \in \mathbb{N}$. We first prove part A). Pick $S$ as in Lemma 7 and consider $R \geqslant S^{1/4}$. By the definition of set $\overline{U_t^R}$,

$$N_{t,i} > \left[ R \left( N_{t,J_t} + 1 \right) \right]^4 \geqslant S, \quad \forall i \in \overline{U_t^R}. \tag{10}$$

Then by Lemma 7, $\arg\max_{i \in \overline{U_t^R}} m_{t,i}$ is unique, and let $\overline{I_t^R} = \arg\max_{i \in \overline{U_t^R}} m_{t,i}$.

Now we are going to show part B). Fix arm $i \in \overline{U_t^R} \setminus \left\{\overline{I_t^R}\right\}$. By Equation (10), $\min\left\{N_{t,i}, N_{t,\overline{I_t^R}}\right\} > S$, and thus by Lemma 7,

$$\alpha_{t,i} \leqslant e^{-Q^2_{t,i,\overline{I_t^R}}/2} \quad \text{where} \quad Q_{t,i,\overline{I_t^R}} = \frac{\Delta_{\min}}{2\left\|X_{\text{pop}}\right\|\left(\frac{W_3+p_{\min}}{p_{\min}b_{\min}}\right)^{1/2}\left(\frac{1}{N_{t,i}^{1/4}+1}+\frac{1}{N_{t,\overline{I_t^R}}^{1/4}+1}\right)^{1/2}}.$$

Recall the least sampled arm $J_t^{\min}$ defined in Equation (8). By the definition of $U_t^R$ in (8), $R \geqslant 1$ ensures $J_t^{\min} \in U_t^R$, so $\max_{j\in U_t^R}\alpha_{t,j} \geqslant \alpha_{t,J_t^{\min}}$. Now it suffices to show $\alpha_{t,J_t^{\min}}$ is greater than the upper bound of $\alpha_{t,i}$ above. Denote $\tilde{\theta}_t$ as a sample drawn independently from $\mathbb{P}_t(\theta \in \cdot)$ and recall $\hat{I}_t \in \arg\max_{j\in[k]} m_{t,j}$. Then

$$\begin{aligned}
\alpha_{t,J_t^{\min}} &= \mathbb{P}_t\left(\mu(\tilde{\theta}_t, J_t^{\min}, w) - \mu(\tilde{\theta}_t, j, w) \geqslant 0,\ \forall j \neq J_t^{\min}\right) \\
&\geqslant \mathbb{P}_t\left(\mu(\tilde{\theta}_t, J_t^{\min}, w) \geqslant m_{t,\hat{I}_t};\ \mu(\tilde{\theta}_t, j, w) \leqslant m_{t,\hat{I}_t},\ \forall j \neq J_t^{\min}\right) \\
&= \mathbb{P}_t\left(\mu(\tilde{\theta}_t, J_t^{\min}, w) \geqslant m_{t,\hat{I}_t}\right) \prod_{j\neq J_t^{\min}} \mathbb{P}_t\left(\mu(\tilde{\theta}_t, j, w) \leqslant m_{t,\hat{I}_t}\right) \\
&\geqslant 2^{-k+1}\Phi\left(-\frac{m_{t,\hat{I}_t} - m_{t,J_t^{\min}}}{s_{t,J_t^{\min}}}\right)
\end{aligned}$$

where the last equality follows from the independent posterior distributions, and the final inequality holds because for each $j \neq J_t^{\min}$,

$$m_{t,j} \leqslant m_{t,\hat{I}_t} \quad \Longrightarrow \quad \mathbb{P}_t\left(\mu(\tilde{\theta}_t, j, w) \leqslant m_{t,\hat{I}_t}\right) \geqslant 1/2.$$

Now we are going to lower bound the argument $-\frac{m_{t,\hat{I}_t} - m_{t,J_t^{\min}}}{s_{t,J_t^{\min}}}$. By Lemma 5,

$$s_{t,J_t^{\min}} \geqslant \frac{\left\|X_{\text{pop}}\right\|}{\left(b_{\max}N_{t,J_t^{\min}} + p_{\max}\right)^{1/2}},$$

and by Corollary 2,

$$\begin{aligned}
m_{t,\hat{I}_t} - m_{t,J_t^{\min}} &= \left[\mu(\theta, \hat{I}_t, w) - \mu(\theta, J_t^{\min}, w)\right] + \left[m_{t,\hat{I}_t} - \mu(\theta, \hat{I}_t, w)\right] - \left[m_{t,J_t^{\min}} - \mu(\theta, J_t^{\min}, w)\right] \\
&\leqslant \Delta_{\max} + W_1\left\|X_{\text{pop}}\right\|\sqrt{\frac{W_3+p_{\min}}{p_{\min}b_{\min}}}\left(\sqrt{\frac{\log\left(N_{t,\hat{I}_t}+e\right)}{(N_{t,\hat{I}_t}+1)^{1/4}}} + \sqrt{\frac{\log\left(N_{t,J_t^{\min}}+e\right)}{(N_{t,J_t^{\min}}+1)^{1/4}}}\right) \\
&\leqslant \Delta_{\max} + c_1 W_1\left\|X_{\text{pop}}\right\|\sqrt{W_3+p_{\min}}
\end{aligned}$$

where $\Delta_{\max}(\theta) \triangleq \max_{j\in[k]}\mu(\theta, j, w) - \min_{j\in[k]}\mu(\theta, j, w)$; the last inequality follows from $\frac{\log(x+e)}{(x+1)^{1/4}} \leqslant 1.2$ for

$x \geqslant 0$ and $c_1 \triangleq \frac{3}{\sqrt{p_{\min} b_{\min}}}$. Hence,

$$\begin{aligned}
\alpha_{t,J_t^{\min}} &\geqslant 2^{-k+1}\Phi\left(-\frac{m_{t,\hat{I}_t} - m_{t,J_t^{\min}}}{s_{t,J_t^{\min}}}\right) \\
&\geqslant 2^{-k+1}\Phi\left(-P_{t,J_t^{\min}}\right) \\
&\geqslant 2^{-k+1} e^{-\left(P_{t,J_t^{\min}} + \sqrt{2\pi}\right)^2/2}
\end{aligned}$$

where the last inequality uses the lower bound of Gaussian tail, (see Lemma 25), and

$$P_{t,J_t^{\min}} \triangleq \frac{\left(\Delta_{\max} + c_1 W_1 \|X_{\text{pop}}\| \sqrt{W_3 + p_{\min}}\right)\left(b_{\max} N_{t,J_t^{\min}} + p_{\max}\right)^{1/2}}{\|X_{\text{pop}}\|}.$$

The final step is using Lemma 10 below this proof to show the upper bound of $\alpha_{t,i}$ is no more than the lower bound of $\alpha_{t,J_t^{\min}}$. Disregarding terms like $W_1 \geqslant 0$ which are non-negative, we have

$$P_{t,J_t^{\min}} \geqslant c_2 \quad \text{where} \quad c_2 \triangleq \frac{\Delta_{\max} p_{\max}^{1/2}}{\|X_{\text{pop}}\|} > 0.$$

We notice that there exists $\widetilde{R} \in \mathbb{M}$ (whose choice depends on constants like $\Delta_{\max}$ and is polynomial in $W_1$ and $W_3$, but independent of arms $i$, $\overline{I_t^R}$ and $J_t^{\min}$) such that for $R \geqslant \widetilde{R}$,

$$i, \overline{I_t^R} \in \overline{U_t^R} \quad \text{i.e.} \quad \min\left\{N_{t,i}^{1/4}, N_{t,\overline{I_t^R}}^{1/4}\right\} > R\left(N_{t,J_t^{\min}} + 1\right)$$

implies

$$\frac{Q_{t,i,\overline{I_t^R}}}{P_{t,J_t^{\min}}} = \frac{\Delta_{\min}}{2\left(\frac{W_3 + p_{\min}}{p_{\min} b_{\min}}\right)^{1/2}\left(\Delta_{\max} + c_1 W_1 \|X_{\text{pop}}\| \sqrt{W_3 + p_{\min}}\right)\left(\frac{b_{\max} N_{t,J_t^{\min}} + p_{\max}}{N_{t,i}^{1/4} + 1} + \frac{b_{\max} N_{t,J_t^{\min}} + p_{\max}}{N_{t,\overline{I_t^R}}^{1/4} + 1}\right)^{1/2}} \geqslant g_{c_2, 2^{1-k}}$$

where $g_{c_2,2^{1-k}}$ is defined in Lemma 10 below. Then by Lemma 10, the upper bound of $\alpha_{t,i}$ is no more than the lower bound of $\alpha_{t,J_t^{\min}}$, and thus $\alpha_{t,i} < \alpha_{t,J_t^{\min}}$. Taking $R = \max\left\{1, \widetilde{R}, S^{1/4}\right\}$ completes the proof. □

**Lemma 10** (Comparison of two exponential functions.). *For any $a, b > 0$, there exists $g_{a,b} \in (0, \infty)$ such that*

$$x \geqslant a \quad \text{and} \quad \frac{y}{x} \geqslant g_{a,b} \quad \Longrightarrow \quad \frac{e^{-y^2/2}}{b \cdot e^{-\left(x + \sqrt{2\pi}\right)^2/2}} < 1.$$

We omit the proof of Lemma 10 and instead proceed to complete the proof of Lemma 8.

*Proof of Lemma 8.* Fix $t \in \mathbb{N}$. Pick $R$ as in Lemma 9, and we are going to show if $J_t^{(1)} \notin U_t^R$ then $J_t^{(2)} \in U_t^R$. If $\overline{U_t^R}$ is empty, we are done. Now suppose $\overline{U_t^R}$ is nonempty. By Lemma 9, the two "most promising arms"

defined in Equation (9) can be rewritten as follows:

$$J_t^{(1)} = \arg\max_{i\in[k]} \alpha_{t,i} = \arg\max_{i\in U_t^R\cup\{\overline{I_t^R}\}} \alpha_{t,i}, \tag{11}$$

$$J_t^{(2)} = \arg\max_{i\neq J_t^{(1)}} \alpha_{t,i} = \arg\max_{i\in U_t^R\cup\{\overline{I_t^R}\}\setminus\{J_t^{(1)}\}} \alpha_{t,i}. \tag{12}$$

Suppose $J_t^{(1)} \notin U_t^R$. By Equation (11), $J_t^{(1)} = \overline{I_t^R}$ , in which case Equation (12) implies $J_t^{(2)} \in U_t^R$. This completes the proof. □

### 6.3 Part 3: DTS assigns constant effort to the insufficiently sampled set in each period

Lemma 8 shows that at any time $t \in \mathbb{N}$, either $J_t^{(1)}, J_t^{(2)}$ or both belong to the insufficiently sampled set $U_t^R$. We denote such an insufficiently sampled arm as

$$\widetilde{I}_t^R = \begin{cases} J_t^{(1)} & \text{if } J_t^{(1)} \in U_t^R, \\ J_t^{(2)} & \text{otherwise .} \end{cases}$$

Note that the identity of $\widetilde{I}_t^R$ can change over time. In this part, we show that at any time $t$, DTS allocates a decent amount of effort to $\widetilde{I}_t^R$. Think of this insufficiently sampled arm $\widetilde{I}_t^R$ as a tool for showing that the insufficiently sample *set* $U_t^R$ is measured.

**Lemma 11.** *Take $R$ in Lemma 9. Under DTS applied with any uniformly strictly bounded $\{\beta_t\}$, for any $t \in \mathbb{N}$,*

$$\psi_{t,\widetilde{I}_t^R} \geqslant \frac{\beta_{\min}}{k^2}$$

*where $\beta_{\min}$ is defined in Equation (3).*

*Proof.* Fix $t \in \mathbb{N}$. Recall for any arm $i \in [k]$, $\alpha_{t,i} = \mathbb{P}(I^* = i \mid H_t)$ and under DTS,

$$\psi_{t,i} = \alpha_{t,i}\left(\beta_t + (1-\beta_t)\sum_{j\neq i}\frac{\alpha_{t,j}}{1-\alpha_{t,j}}\right).$$

By the definition of two "most promising arms" $J_t^{(1)}$ and $J_t^{(2)}$ in Equation (9), we have

$$\alpha_{t,J_t^{(1)}} \geqslant \frac{1}{k} \quad \text{and} \quad \frac{\alpha_{t,J_t^{(2)}}}{1-\alpha_{t,J_t^{(1)}}} \geqslant \frac{1}{k-1}.$$

Hence, if $\widetilde{I}_t^R = J_t^{(1)}$,

$$\psi_{t,\widetilde{I}_t^R} = \psi_{t,J_t^{(1)}} \geqslant \alpha_{t,J_t^{(1)}}\beta_t \geqslant \frac{\beta_{\min}}{k};$$

otherwise,

$$\psi_{t,\widetilde{I}_t^R} = \psi_{t,J_t^{(2)}} \geqslant \alpha_{t,J_t^{(1)}}\frac{\alpha_{t,J_t^{(2)}}}{1-\alpha_{t,J_t^{(1)}}}(1-\beta_t) \geqslant \frac{\beta_{\min}}{k(k-1)}.$$

This completes the proof. □

## 6.4 Part 4: By the pigeonhole principle, in total a lot of measurement effort is assigned to insufficiently sampled arms

Recall that $\Psi_{t,i} = \sum_{\ell=1}^{t-1} \psi_{\ell,i}$ is the total probability assigned to measuring arm $i$ prior to time $t$. By taking the average, we suggest that at most times, the "under-sampled" arm $\widetilde{I}_\ell^R$ has actually had a lot of effort assigned to it in the past. Think of the arm $\widetilde{I}_\ell^R$ as a tool for helping us lower bound how much a *representative* arm in insufficiently-sampled set has been sampled; the construction of the insufficiently-sampled set rules out cases where one arm in this set is measured too much more than others, as will be made formal in the next step of the proof.

**Lemma 12.** *Take $R$ in Lemma 9. Under DTS applied with any uniformly strictly bounded $\{\beta_t\}$, for any $t \in \mathbb{N}$,*

$$\frac{1}{t}\sum_{\ell=1}^{t} \Psi_{\ell,\widetilde{I}_\ell^R} \geqslant \frac{\beta_{\min}(t-1)}{2k^3}.$$

*Proof.* For any $(t,i) \in \mathbb{N} \times [k]$, we define

$$S_{t,i} \triangleq \sum_{\ell=1}^{t-1} \mathbf{1}\left\{i = \widetilde{I}_\ell^R\right\},$$

and then

$$\Psi_{t,i} = \sum_{\ell=1}^{t-1} \psi_{\ell,i} \geqslant \sum_{\ell=1}^{t-1} \mathbf{1}\left\{i = \widetilde{I}_\ell^R\right\} \psi_{\ell,\widetilde{I}_\ell^R} \geqslant \frac{\beta_{\min}}{k^2} S_{t,i}$$

where the last inequality follows from Lemma 11. Hence, for any $t \in \mathbb{N}$,

$$\sum_{\ell=1}^{t} \Psi_{\ell,\widetilde{I}_\ell^R} \geqslant \frac{\beta_{\min}}{k^2} \sum_{\ell=1}^{t} S_{\ell,\widetilde{I}_\ell^R}.$$

Now we are going to lower bound the RHS above.

$$\begin{aligned}
\sum_{\ell=1}^{t} S_{\ell,\widetilde{I}_\ell^R} = \sum_{\ell=1}^{t}\sum_{i=1}^{k} \mathbf{1}\left\{\widetilde{I}_\ell^R = i\right\} S_{\ell,i} = \sum_{i=1}^{k}\sum_{\ell=1}^{t} \mathbf{1}\left\{i = \widetilde{I}_\ell^R\right\} S_{\ell,i} &= \sum_{i=1}^{k}\sum_{h=0}^{S_{t,i}} h \\
&= \frac{1}{2}\sum_{i=1}^{k}(S_{t,i}+1)S_{t,i} \\
&\overset{(a)}{\geqslant} \frac{1}{2}\left[\frac{\left(\sum_{i=1}^{k} S_{t,i}\right)^2}{k} + \sum_{i=1}^{k} S_{t,i}\right] \\
&\overset{(b)}{=} \frac{1}{2}\left[\frac{(t-1)^2}{k} + (t-1)\right] \\
&= \frac{(t-1)(t-1+k)}{2k} \\
&\geqslant \frac{(t-1)t}{2k}
\end{aligned}$$

where step (a) applies Jensen's inequality and step (b) follows from $\sum_{i=1}^{k} S_{t,i} = t-1$.

□

## 6.5 Part 5: From measurement effort to realized measurements

The next Lemma uses the definition of $W_2$ from Section 5.1 to control deviations between the measurement effort $\Psi_{t,i}$ and the actual number of measurements collected $N_{t,i}$.

**Lemma 13.** *Take $R$ as in Lemma 9. Under DTS applied with any uniformly strictly bounded $\{\beta_t\}$, there exist (nonrandom) constant $c_1, c_2 \geqslant 0$ such that for any $t \in \mathbb{N}$,*

$$\frac{1}{t}\sum_{\ell=1}^{t} N_{\ell,\widetilde{I}_\ell^R} \geqslant \frac{\beta_{\min}(t-1)}{2k^3} - W_2\left(c_1 t^{3/4} + c_2\right).$$

*Proof.* There exist (nonrandom) constants $c_1, c_2 \geqslant 0$ such that for any $t \in \mathbb{N}$,

$$c_1 t^{3/4} + c_2 \geqslant \sqrt{(t+1)\log\left(t+e^2\right)},$$

and thus by the definition of $W_2$, for any $(t,i) \in \mathbb{N} \times [k]$,

$$N_{t,i} \geqslant \Psi_{t,i} - W_2\sqrt{(t+1)\log\left(t+e^2\right)} \geqslant \Psi_{t,i} - W_2\left(c_1 t^{3/4} + c_2\right).$$

Then take $R$ as in Lemma 9, and for any $t \in \mathbb{N}$,

$$\frac{1}{t}\sum_{\ell=1}^{t} N_{\ell,\widetilde{I}_\ell^R} \geqslant \frac{1}{t}\sum_{\ell=1}^{t} \Psi_{\ell,\widetilde{I}_\ell^R} - \frac{1}{t}\sum_{\ell=1}^{t} W_2\left(c_1 \ell^{3/4} + c_2\right) \geqslant \frac{\beta_{\min}(t-1)t}{2k^3} - W_2\left(c_1 t^{3/4} + c_2\right)$$

where the second inequality applies Lemma 12 and that $\ell^{3/4}$ is increasing in $\ell$. □

## 6.6 Part 6: Completing the proof by using the definition of the under-sampled set

We have just shown that lot of measurements are collected across time from some representative arm $\widetilde{I}_t^R$ chosen within the insufficiently sampled set. To complete the proof, we use the definition of of the insufficiently sampled set to relate the number of measurements collected from the representative arm across time to the minimal number of measurements taken from any arm.

*Proof of Proposition 2.* Take $R, c_1, c_2$ as in Lemma 13. Fix $t \in \mathbb{N}$. We have

$$R^4\left(\min_{i\in[k]} N_{t,i} + 1\right)^4 \geqslant \frac{1}{t}\sum_{\ell=1}^{t} R^4\left(\min_{i\in K} N_{\ell,i} + 1\right)^4 \geqslant \frac{1}{t}\sum_{\ell=1}^{t} N_{\ell,\widetilde{I}_\ell^R}$$

where the first inequality follows from that $\min_{i\in[k]} N_{\ell,i}$ is increasing in $\ell$; the second inequality holds since by Lemma 8 and the definition of $\widetilde{I}_\ell^R$, we have $\widetilde{I}_\ell^R \in U_\ell^R$, i.e., $R^4\left(\min_{i\in[k]} N_{t,i} + 1\right)^4 \geqslant N_{t,\widetilde{I}_t^R}$.

Therefore,

$$\min_{i\in[k]} N_{t,i} + 1 \geqslant \frac{1}{R}\left(\frac{1}{t}\sum_{\ell=1}^{t} N_{\ell,\widetilde{I}_\ell^R}\right)^{1/4}.$$

Then by Lemma 13,

$$\min_{i\in[k]} N_{t,i} + 1 \geqslant \frac{1}{R}\left(\frac{1}{t}\sum_{\ell=1}^{t} N_{\ell,\widetilde{I}_\ell^R}\right)^{1/4} \geqslant \frac{1}{R}\left[\frac{\beta_{\min}(t-1)}{2k^3} - W_2\left(c_1 t^{3/4} + c_2\right)\right]^{1/4}.$$

Hence, $\min_{i\in[k]} N_{t,i} \geqslant t^{3/16}$ whenever

$$\frac{\beta_{\min}(t-1)}{2k^3} - W_2\left(c_1 t^{3/4} + c_2\right) \geqslant R^4\left(t^{3/16}+1\right)^4.$$

We only need $t \geqslant \left(c_3 R^4 + c_4 W_2 + c_5\right)^4$ where $c_4, c_4, c_5 \geqslant 0$ are nonrandom constants. This completes the proof, as $\left(c_3 R^4 + c_4 W_2 + c_5\right)^4 \in \mathbb{M}$. □

# 7 Proof of Proposition 1: Strong convergence to the optimal proportions

In this section, we complete the proof of Proposition 1. Restated in terms of, strong convergence notation, our goal is to show two limits:

$$p_{t,i} \xrightarrow{\mathbb{M}} p_i^* \quad \text{and} \quad \frac{Z_{t,\hat{I}_t,i}}{t} \xrightarrow{\mathbb{M}} 2\Gamma_\theta^{-1}, \quad \forall i \in [k]$$

Here $p^*$ and $\Gamma_\theta$ are the optimal sampling ratios and exponent given in (19).

To prove this proposition, we need a sequence of results that can be categorized into the following parts.

1. As shown in Subsection 6, DTS sufficiently explores all arms. In part 1, we use this to control the asymptotic behavior of posterior means and variances under DTS.
2. In part 2, we show that under DTS, the fraction of samples allocated to the true best arm converges to the optimal sampling proportion.
3. In part 3, building on the results in the previous parts, we show that under DTS, the fraction of samples allocated to each suboptimal arm also converges to the optimal sampling proportion.

## 7.1 Part 1: Strong convergence of posterior means and variances under DTS

As shown in Subsection 6, under DTS, each arm is sufficiently explored. As the number of samples collected from each arm tends to infinity, we expect estimated arm-means to converge to the truth. The next result establishes this using the notion of strong convergence.

**Lemma 14.** *Under DTS applied with any uniformly strictly bounded $\{\beta_t\}$,*

$$m_{t,i} \xrightarrow{\mathbb{M}} \mu(\theta, i, w), \quad \forall i \in [k].$$

*Proof.* Fix arm $i \in [k]$. Fix any $\epsilon > 0$. Take $T$ as in Proposition 2 and $S_\epsilon$ in Lemma 6. Then

$$t \geqslant T_\epsilon \triangleq \max\left\{T, S_\epsilon^{16/3}\right\} \quad \Longrightarrow \quad N_{t,i} \geqslant t^{3/16} \geqslant S_\epsilon \quad \Longrightarrow \quad |m_{t,i} - \mu(\theta, i, w)| \leqslant \epsilon.$$

□

Lemma 14 and the definition of strong convergence immediately leads to that after enough periods have elapsed, the empirically best arm, $\hat{I}_t = \arg\max_{i\in[k]} m_{t,i}$, is uniquely determined and always the true best one, $I^* = \arg\max_{i\in[k]} \mu(\theta, i, w)$. Recall that $I^*$ is unique (see the parameter class in Equation (12)).

**Corollary 3.** *Under DTS applied with any uniformly strictly bounded $\{\beta_t\}$, there exists $T \in \mathbb{M}$ such that for any $t \geqslant T$, $\hat{I}_t = I^*$.*

Later to prove the results in part 3, we need to better control each arm's posterior variance. Recall that $s_{t,i}^2 = X_{\text{pop}}^\top \Sigma_{t,i} X_{\text{pop}}$ is the posterior variance evaluated in the direction of the population-weights, $X_{\text{pop}}$. Recall also that the posterior covariance matrix $\Sigma_{t,i}^{-1} = \Sigma_{1,i}^{-1} + \sigma^{-2} \sum_{\ell=1}^{t-L} \mathbb{1}\{I_\ell = i\} X_\ell X_\ell^\top$. The next result shows that as the number of time periods tends to infinity (and so does the number of sampled allocated to each arm by Proposition 2), the posterior variance $\sigma_{t,i}^2$ scales inversely with the number of samples allocated to it, $N_{t,i}$. This is a faster rate than in the crude bound of Lemma 5.

**Lemma 15.** *Under DTS applied with any uniformly strictly bounded* $\{\beta_t\}$,

$$N_{t,i} s_{t,i}^2 \xrightarrow{\mathbb{M}} \sigma^2 \|X_{\text{pop}}\|_{\Lambda^{-1}}^2, \quad \forall i \in [k].$$

*Proof.* Fix arm $i \in [k]$. By the definition of $W_3$ in Equation (7),

$$\left\| \Sigma_{t,i}^{-1} - \sigma^{-2} \Lambda (N_{t,i} + 1) \right\| \leqslant W_3 \sqrt{(N_{t,i} + 1) \log (N_{t,i} + e)},$$

which implies

$$\left\| \frac{\Sigma_{t,i}^{-1}}{N_{t,i} + 1} - \sigma^{-2} \Lambda \right\| \leqslant W_3 \sqrt{\frac{\log (N_{t,i} + e)}{N_{t,i} + 1}} \xrightarrow{\mathbb{M}} 0.$$

Since strong convergence is a new convergence concept, we justify last step explicitly. Fix any $\epsilon > 0$. Fix a (nonrandom) constant $n_0$ such that for $n \geqslant n_0$, $\sqrt{\log(n+e)/(n+1)} \leqslant n^{-1/4}$. Take $T \in \mathbb{M}$ as in Proposition 2 and define $S_\epsilon \triangleq \max\left\{n_0, (W_3/\epsilon)^4\right\}$. Then

$$t \geqslant T_\epsilon \triangleq \max\left\{T, S_\epsilon^{16/3}\right\} \quad \Longrightarrow \quad N_{t,i} \geqslant t^{3/16} \geqslant S_\epsilon \quad \Longrightarrow \quad \left\| \frac{\Sigma_{t,i}^{-1}}{N_{t,i} + 1} - \sigma^{-2} \Lambda \right\| \leqslant \frac{W_3}{N_{t,i}^{1/4}} \leqslant \epsilon.$$

We have shown that $(N_{t,i} + 1)^{-1} \Sigma_{t,i}^{-1} \xrightarrow{\mathbb{M}} \sigma^{-2} \Lambda$. Since the function $f(E) = E^{-1}$ is continuous at each point that is invertible, and $\Lambda$ is invertible, the continuous mapping theorem in Lemma 3, implies $(N_{t,i} + 1)\Sigma_{t,i} \xrightarrow{\mathbb{M}} \sigma^2 \Lambda^{-1}$. Again by the continuous mapping theorem, $(N_{t,i} + 1) X_{\text{pop}}^\top \Sigma_{t,i} X_{\text{pop}} \xrightarrow{\mathbb{M}} \sigma^2 X_{\text{pop}}^\top \Lambda^{-1} X_{\text{pop}} = \sigma^2 \|X_{\text{pop}}\|_{\Lambda^{-1}}^2$. Recalling $s_{t,i}^2 = X_{\text{pop}}^\top \Sigma_{t,i} X_{\text{pop}}$ give the result.

□

## 7.2 Part 2: DTS ensures strong convergence to optimal proportion for the best arm

In this section, we are going to show that under DTS applied with Algorithm 2, the empirical proportion of the best arm strongly converges to its optimal proportion:

**Proposition 3.** *Under DTS applied with Algorithm 2,*

$$\frac{N_{t,I^*}}{t} \xrightarrow{\mathbb{M}} p_{I^*}^*.$$

By Corollary 1 that connects $N_{t,i}$ and $\Psi_{t,i}$, it suffices to show

**Lemma 16.** *Under DTS applied with Algorithm 2,*

$$\frac{\Psi_{t,I^*}}{t} \xrightarrow{\mathbb{M}} p_{I^*}^*.$$

To prove this lemma, we need a sequence of results. The following one studies the limiting behavior of the posterior beliefs $(\alpha_{t,i} : i \in [k])$ about the identity of the optimal arm. We show that this distribution converges strongly to a point mass at the true best arm, $I^*$. This can be thought of as a result about the

asymptotic consistency of the posterior distribution, but here stated in terms of a bespoke convergence notion.

**Lemma 17.** *Under DTS applied with any uniformly strictly bounded* $\{\beta_t\}$,

$$\alpha_{t,I^*} \xrightarrow{\mathbb{M}} 1.$$

*Proof.* It is equivalent of showing $\alpha_{t,i} \xrightarrow{\mathbb{M}} 0$ for any $i \neq I^*$.

Fix $i \in [k]$. Take $T$ in Corollary 3 as $T_1$. For any $t \geqslant T_1$, $\hat{I}_t = I^*$, and then by Lemma 7,

$$N_{t,i}, N_{t,I^*} \geqslant S \quad \Longrightarrow \quad \alpha_{t,i} \leqslant e^{-Q_{t,i,I^*}^2/2}$$

where

$$Q_{t,i,I^*} = \frac{\Delta_{\min}}{2\left\|X_{\text{pop}}\right\| \left(\frac{W_3+p_{\min}}{p_{\min}b_{\min}}\right)^{1/2} \left(\frac{1}{N_{t,i}^{1/4}+1} + \frac{1}{N_{t,I^*}^{1/4}+1}\right)^{1/2}}.$$

Note that there exists a (nonrandom) constant $s_\epsilon > 0$ such that

$$N_{t,i}, N_{t,I^*} \geqslant s_\epsilon, \quad \Longrightarrow \quad e^{-Q_{t,i,I^*}^2/2} \leqslant \epsilon.$$

Now take $T$ in Proposition 2 as $T_2$. We have

$$t \geqslant \max\left\{T_1, S^{16/3}, s_\epsilon^{16/3}, T_2\right\} \quad \Longrightarrow \quad N_{t,i}, N_{t,I^*} \geqslant \max\{S, s_\epsilon\} \quad \Longrightarrow \quad \alpha_{t,i} \leqslant e^{-Q_{t,i,I^*}^2/2} \leqslant \epsilon.$$

This completes the proof.

□

In Algorithm 2, the optimal long-run sampling ratios $\hat{p} = p^*(\mu_t)$ is continuous in the plug-in posterior mean vector $\mu_t$. By Lemma 14, we immediately have the following result on the tuning parameter $\beta_t = \hat{p}_{\hat{I}_t}$

**Lemma 18.** *Under DTS applied with Algorithm 2,*

$$\beta_t \xrightarrow{\mathbb{M}} p_{I^*}^*.$$

*Proof.* Take $T$ as in Corollary 3. For any $t \geqslant T$, $\hat{I}_t = I^*$, and then Lemma 5 gives

$$\beta_t = \frac{1}{1 + \sum_{i \neq I^*} \left(\Delta_{t,i}^2 y_t - 1\right)^{-1}}$$

where $\Delta_{t,i} = m_{t,I^*} - m_{t,i}$ for $i \in [k]$ and $y_t$ satisfies $\sum_{i \neq I^*} \left(\Delta_{t,i}^2 y_t - 1\right)^{-2} = 1$. Similarly, by the proof of Lemma 5 in Section D,

$$p_{I^*}^* = \frac{1}{1 + \sum_{i \neq I^*} \left(\Delta_i^2 y^* - 1\right)^{-1}}$$

where $\Delta_i = \mu(\theta, I^*, w) - \mu(\theta, i, w)$ for $i \in [k]$ and $y^*$ satisfies $\sum_{i \neq I^*} \left(\Delta_i^2 y^* - 1\right)^{-2} = 1$. By applying Lemma 14 and Lemma 3 (continuous mapping theorem), $\Delta_{t,i}^2$

$y_t \xrightarrow{\mathbb{M}} y^*$ implies $\beta_t \xrightarrow{\mathbb{M}} p_{I^*}^*$. Hence, it suffices to prove $y_t \xrightarrow{\mathbb{M}} y^*$.

Fix any $\epsilon > 0$. By Lemma 14 and Lemma 3 (continuous mapping theorem), there exists $T_\epsilon \in \mathbb{M}$ such that for any $t \geqslant T_\epsilon$,

$$|\Delta_{t,i}^2 - \Delta_i^2| \leqslant \delta \quad \text{where} \quad \delta \triangleq \frac{\epsilon}{y^* + \epsilon} \min_{i \in [k]} \Delta_i^2.$$

Next we are going to prove by contradiction that for any $t \geqslant \max\{T, T_\epsilon\}$, $|y_t - y^*| \leqslant \epsilon$.
Suppose $y_t > y^* + \epsilon$. Then

$$\sum_{i \neq I^*} \left(\Delta_{t,i}^2 y_t - 1\right)^{-2} \leqslant \sum_{i \neq I^*} \left[(\Delta_i^2 - \delta)(y^* + \epsilon) - 1\right]^{-2} < \sum_{i \neq I^*} \left(\Delta_i^2 y^* - 1\right)^{-2} = 1,$$

which contradicts $\sum_{i \neq I^*} \left(\Delta_{t,i}^2 y_t - 1\right)^{-2} = 1$.
Now suppose $y_t < y^* - \epsilon$. Then

$$\sum_{i \neq I^*} \left(\Delta_{t,i}^2 y_t - 1\right)^{-2} \geqslant \sum_{i \neq I^*} \left[(\Delta_i^2 + \delta)(y^* - \epsilon) - 1\right]^{-2} > \sum_{i \neq I^*} \left(\Delta_i^2 y^* - 1\right)^{-2} = 1,$$

which again contradicts $\sum_{i \neq I^*} \left(\Delta_{t,i}^2 y_t - 1\right)^{-2} = 1$. This completes the proof. □

Now we are ready to complete the proof of Lemma 16.

*Proof of Lemma 16.* Under DTS, for any $t \in \mathbb{N}$,

$$\psi_{t,I^*} = \alpha_{t,I^*} \left[\beta_t + (1 - \beta_t) \sum_{i \neq I^*} \frac{\alpha_{t,i}}{1 - \alpha_{t,i}}\right] := f(\alpha_t, \beta_t).$$

The right hand side is a continuous function, $f$, of $\beta_t$ and $\alpha_t := (\alpha_{t,i} : i \in [k])$. We know $\beta_t \xrightarrow{\mathbb{M}} p_{I^*}^*$ and $\alpha_t \xrightarrow{\mathbb{M}} e_{I^*}$, where $e_i$ is the $i$th standard basis vector. These follow by Corollary 18 and Lemma 17, respectively. Therefore, by the continuous mapping theorem in Lemma 3, $\psi_{t,I^*} \xrightarrow{\mathbb{M}} f(e_{I^*}, p_{I^*}^*) = p_{I^*}^*$.

We conclude by showing that $\psi_{t,I^*} \xrightarrow{\mathbb{M}} p_{I^*}^*$ implies strong convergence of the Cesàro mean, in the sense that $\frac{1}{t}\sum_{\ell=1}^t \psi_{\ell,I^*} \xrightarrow{\mathbb{M}} p_{I^*}^*$. To show this, fix arbitrary $\epsilon > 0$ and pick $\tilde{T}_\epsilon \in \mathbb{M}$ such that for any $t \geqslant \tilde{T}_\epsilon$, $|\psi_{t,I^*} - p_{I^*}^*| \leqslant \epsilon/2$. Then for any $t \geqslant T_\epsilon \triangleq \max\left\{\tilde{T}_\epsilon, 2\tilde{T}_\epsilon/\epsilon\right\}$,

$$\left|\frac{1}{t}\sum_{\ell=1}^t \psi_{\ell,i} - p_{I^*}^*\right| \leqslant \left|\frac{1}{t}\sum_{\ell=1}^{\tilde{T}_\epsilon} (\psi_{\ell,I^*} - p_{I^*}^*)\right| + \left|\frac{1}{t}\sum_{\ell=\tilde{T}_\epsilon+1}^{t} (\psi_{\ell,I^*} - p_{I^*}^*)\right| \leqslant \frac{\tilde{T}_\epsilon}{t} + \frac{\epsilon}{2} \leqslant \epsilon.$$

Recalling $\Psi_{t,I^*} = \sum_{\ell=1}^{t-1} \psi_{\ell,I^*}$ completes the proof.

□

## 7.3 Part 3: Strong convergence to optimal proportions under DTS

In this section, we are going to show that under DTS, the empirical proportion of each arm strongly converges to its optimal proportion, which completes the proof of Proposition 1.

The next lemma looks at the "z-scores" which appear from pairwise comparison between arms under posterior beliefs. As Lemma 17 shows that posterior concentrates on the unique true best arm $I^*$, it is natural to consider comparisons against this arm. Recall in Equation (13), we define the z-scores

$$Z_{t,I^*,j} = \frac{m_{t,I^*} - m_{t,j}}{\sqrt{s_{t,I^*}^2 + s_{t,j}^2}}, \quad \forall j \neq I^*.$$

We show that asymptotically each z-score is approximated by a deterministic function of the proportion of samples taken from the sub-optimal arm being compared. The key to the result is that Proposition 3 already fixes the fraction of samples allocated to the best arm $I^*$.

**Lemma 19.** *For any arm $j \neq I^*$,*

$$\frac{Z_{t,I^*,j}}{\sqrt{t} f_j\left(p_{t,j}\right)} \xrightarrow{\mathbb{M}} 1 \quad \textit{where} \quad p_{t,j} = \frac{N_{t,j}}{t} \quad \textit{and} \quad f_j(p_{t,j}) \triangleq \frac{\mu(\theta, I^*, w) - \mu(\theta, j, w)}{\sigma \|X_{\text{pop}}\|_{\Lambda^{-1}} \sqrt{(p^*_{I^*})^{-1} + p_{t,j}^{-1}}}.$$

*Proof.* Fix $j \neq I^*$. By Proposition 3, $p_{t,I^*} \xrightarrow{\mathbb{M}} p^*_{I^*} > 0$, and by Proposition 2, there exists $T \in \mathbb{M}$ such that for $t \geqslant T$, $p_{t,j} = N_{t,j}/t > 0$. Then we have

$$\begin{aligned}
\frac{Z_{t,I^*,j}}{\sqrt{t} f_j(p_{t,j})} &= \frac{m_{t,I^*} - m_{t,j}}{\mu(\theta, I^*, w) - \mu(\theta, j, w)} \cdot \frac{\sigma \|X_{\text{pop}}\|_{\Lambda^{-1}} \sqrt{(p^*_{I^*})^{-1} + p_{t,j}^{-1}}}{\sqrt{t s^2_{t,I^*} + t s^2_{t,j}}} \\
&= \frac{m_{t,I^*} - m_{t,j}}{\mu(\theta, I^*, w) - \mu(\theta, j, w)} \cdot \frac{\sigma \|X_{\text{pop}}\|_{\Lambda^{-1}} \sqrt{(p^*_{I^*})^{-1} + p_{t,j}^{-1}}}{\sqrt{\frac{N_{t,I^*} s^2_{t,I^*}}{p_{t,I^*}} + \frac{N_{t,j} s^2_{t,j}}{p_{t,j}}}} \xrightarrow{\mathbb{M}} 1
\end{aligned}$$

where the last step follows from Lemmas 14 and 15 and Proposition 3 imply

$$m_{t,j} \xrightarrow{\mathbb{M}} \mu(\theta, j, w), \quad N_{t,j} s^2_{t,j} \xrightarrow{\mathbb{M}} \sigma^2 \|X_{\text{pop}}\|^2_{\Lambda^{-1}} \quad \text{and} \quad p_{t,I^*} \xrightarrow{\mathbb{M}} p^*_{I^*}.$$

□

The result below is the key one for this part. It shows that, under DTS, if an arm has been sampled more in the past than is prescribed under the optimal allocation, then its probability of being sampled this period is exponentially small.

**Lemma 20.** *For any $\epsilon > 0$, there exist a deterministic constant $c_\epsilon > 0$ and a random variable $T_\epsilon \in \mathbb{M}$ such that for any $t \geqslant T_\epsilon$ and $i \neq I^*$,*

$$\frac{\Psi_{t,i}}{t} \geqslant p^*_i + \epsilon \quad \Longrightarrow \quad \psi_{t,i} \leqslant \exp(-c_\epsilon t)$$

*Proof.* Fix $\epsilon > 0$. It suffices to show that for any arm $i \neq I^*$, there exist a deterministic constant $c^{(i)}_\epsilon > 0$ and a random variable $T^{(i)}_\epsilon \in \mathbb{M}$ such that for $t \geqslant T^{(i)}_\epsilon$,

$$\frac{\Psi_{t,i}}{t} \geqslant p^*_i + \epsilon \quad \Longrightarrow \quad \psi_{t,i} \leqslant \exp\left(-c^{(i)}_\epsilon t\right),$$

since taking $T_\epsilon \triangleq \max_{i \neq I^*} T^{(i)}_\epsilon$ and $c_\epsilon \triangleq \min_{i \neq I^*} c^{(i)}_\epsilon$ completes the proof.

Throughout the remaining proof, we fix arm $i \neq I^*$. Under DTS,

$$\psi_{t,i} = \alpha_{t,i}\left[\beta_t + (1-\beta_t)\sum_{j \neq i} \frac{\alpha_{t,j}}{1-\alpha_{t,j}}\right] \leqslant \alpha_{t,i}\beta_t + \alpha_{t,i}(1-\beta_t)\frac{1}{1-\alpha_{t,I^*}} \leqslant \frac{\alpha_{t,i}}{1-\alpha_{t,I^*}}.$$

Denote $\tilde{\theta}_t$ as a sample drawn from $\mathbb{P}_t(\theta \in \cdot)$. Then we have

$$\alpha_{t,i} \leqslant \mathbb{P}_t\left(\mu(\tilde{\theta}_t, i, w) \geqslant \mu(\tilde{\theta}_t, I^*, w)\right)$$

and

$$1 - \alpha_{t,I^*} = \mathbb{P}_t\left(\exists j \neq I^* : \mu(\tilde{\theta}_t, j, w) \geqslant \mu(\tilde{\theta}_t, I^*, w)\right) \geqslant \max_{j \neq I^*} \mathbb{P}_t\left(\mu(\tilde{\theta}_t, j, w) \geqslant \mu(\tilde{\theta}_t, I^*, w)\right).$$

Therefore

$$\psi_{t,i} \leqslant \frac{\mathbb{P}_t\left(\mu(\tilde{\theta}_t, i, w) \geqslant \mu(\tilde{\theta}_t, I^*, w)\right)}{\max_{j \neq I^*} \mathbb{P}_t\left(\mu(\tilde{\theta}_t, J_t, w) \geqslant \mu(\tilde{\theta}_t, I^*, w)\right)} = \frac{\Phi\left(-Z_{t,I^*,i}\right)}{\max_{j \neq I^*} \Phi\left(-Z_{t,I^*,j}\right)}. \tag{13}$$

where each z-scores $Z_{t,I^*,j}$ is defined in Equation (13).

Now suppose the hypothesized condition $\frac{\Psi_{t,i}}{t} \geqslant p_i^* + \epsilon$ holds. Recall that

1. Corollary 1 states $p_{t,i} - \Psi_{t,i}/t \xrightarrow{\mathbb{M}} 0$ where $p_{t,i} = N_{t,i}/t$;
2. Proposition 3 states $p_{t,I^*} \xrightarrow{\mathbb{M}} p_{I^*}^*$, which is equivalent to $\sum_{j \neq I^*} p_{t,j} \xrightarrow{\mathbb{M}} \sum_{j \neq I^*} p_j^*$.

Combining these two facts, we know there exists $\tilde{T}_\epsilon \in \mathbb{M}$ such that for any $t \geqslant \tilde{T}_\epsilon$,

$$\frac{\Psi_{t,i}}{t} \geqslant p_i^* + \epsilon \quad \Longrightarrow \quad p_{t,i} \geqslant p_i^* + \frac{\epsilon}{2} \quad \text{and} \quad \exists J_t \neq I^* \text{ s.t. } p_{t,J_t} \leqslant p_{J_t}^*. \tag{14}$$

In words, if arm $i$ has been measured in a proportion of rounds that strictly exceeds the optimal proportion, then some other arm $J_t$ must have been under-sampled. Then by Equation (13) and Lemma 19, for any $\delta > 0$, there exists $T_\delta \in \mathbb{M}$ such that $T_\delta \geqslant \tilde{T}_\epsilon$ and for any $t \geqslant T_\delta$,

$$\frac{\Psi_{t,i}}{t} \geqslant p_i^* + \epsilon \quad \Longrightarrow \quad \psi_{t,i} \leqslant \frac{\Phi\left(-Z_{t,I^*,i}\right)}{\Phi\left(-Z_{t,I^*,J_n}\right)} \leqslant \frac{\Phi\left(-\sqrt{t} f_i(p_{t,i}) \cdot (1-\delta)\right)\right)}{\Phi\left(-\sqrt{t} f_{J_t}(p_{t,J_t}) \cdot (1+\delta)\right)}.$$

Since each $f_j(\cdot)$ is a strictly increasing function, when Equation (14) holds, we have $f_i(p_{t,i}) \geqslant f_i(p_i^* + \epsilon/2)$ and $f_{J_t}(p_{t,J_t}) \leqslant f_{J_t}(p_{J_t}^*) = f_i(p_i^*)$. The final equality uses that $f_j(p_j^*) = f_i(p_i^*)$ for any $j \neq i$, which is the defining property of the vector $p^*$ given in Equation (20). Therefore, for any $\delta > 0$, for $t \geqslant T_\delta$,

$$\frac{\Psi_{t,i}}{t} \geqslant p_i^* + \epsilon \implies \psi_{t,i} \leqslant \frac{\Phi\left(-\sqrt{t} f_i(p_i^* + \epsilon/2) \cdot (1-\delta)\right)\right)}{\Phi\left(-\sqrt{t} f_i(p_i^*) \cdot (1+\delta)\right)}$$

Pick a sufficiently small $\delta$ as a function of $\epsilon$ such that we have

$$c_1 \triangleq f_i(p_i^* + \epsilon/2) \cdot (1-\delta) > \ f_i(p_i^*) \cdot (1+\delta) \triangleq c_2.$$

The result then follows by the limiting approximation $\Phi(-\sqrt{t}c_1)/\Phi(-\sqrt{t}c_2) \approx \exp(-t(c_1^2 - c_2^2)/2)$ for large $t$, which can be made precise through the fact that $\frac{1}{t}\log\Phi(-\sqrt{t}x) \to -x^2/2$ as $t \to \infty$.

□

The result above shows that once enough time has passed, almost no further measurement effort is allocated to arms which have been sampled more than its optimal proportion. One expects then that sampling proportions should self-correct; those that have been sampled too much are not sampled until their proportions re-align to the desired level and they cannot become over-sampled again. The proof of this result formalized this intuition.

**Lemma 21.** *For any $\epsilon > 0$, there exists $T_\epsilon \in \mathbb{M}$ such that for any $t \geqslant T_\epsilon$ and $i \in [k]$,*

$$\frac{\Psi_{t,i}}{t} \leqslant p_i^* + \epsilon.$$

*Proof.* It suffices to prove this statement only for $i \neq I^*$ since Lemma 16 already handle the cases for $i = I^*$.

Fix some $i \neq I^*$ and $\epsilon > 0$. By Lemma 20, there exists $c_\epsilon > 0$ and $\tilde{T}_\epsilon \in \mathbb{M}$ (whose choice depends on $\epsilon$) such that for any $t \geqslant \tilde{T}_\epsilon$ and $i \neq I^*$,

$$\frac{\Psi_{t,i}}{t} \geqslant p_i^* + \frac{\epsilon}{2} \quad \Longrightarrow \quad \psi_{t,i} \leqslant \exp\left(-c_\epsilon t\right).$$

Define $\kappa_\epsilon \triangleq \sum_{\ell=1}^{\infty} \exp\left(-c_\epsilon \ell\right) < \infty$.

We provide two different upper bounds on $\frac{1}{t}\Psi_{t,i}$ for $t \geqslant \tilde{T}_\epsilon$, by separately considering two cases.

In the first case, suppose that $\forall \ell \in \left\{\tilde{T}_\epsilon, \tilde{T}_\epsilon + 1, \ldots, t-1\right\}, \frac{\Psi_{\ell,i}}{\ell} \geqslant p_i^* + \frac{\epsilon}{2}$. Then we have the bound

$$\begin{aligned}\frac{\Psi_{t,i}}{t} = \frac{1}{t}\sum_{\ell=1}^{\tilde{T}_\epsilon - 1} \psi_{\ell,i} + \frac{1}{t}\sum_{\ell=\tilde{T}_\epsilon}^{t-1} \psi_{\ell,i} = \frac{1}{t}\sum_{\ell=1}^{\tilde{T}_\epsilon - 1} \psi_{\ell,i} + \frac{1}{t}\sum_{\ell=\tilde{T}_\epsilon}^{t-1} \psi_{\ell,i}\mathbf{1}\left(\frac{\Psi_{\ell,i}}{\ell} \geqslant p_i^* + \frac{\epsilon}{2}\right) &\leqslant \frac{1}{t}\left[\tilde{T}_\epsilon - 1 + \sum_{\ell=\tilde{T}_\epsilon}^{t-1} \exp\left(-c_\epsilon \ell\right)\right] \\ &\leqslant \frac{1}{t}\left(\tilde{T}_\epsilon - 1 + \kappa_\epsilon\right).\end{aligned}$$

In the alternative case, where $\exists \ell \in \left\{\tilde{T}_\epsilon, \tilde{T}_\epsilon + 1, \ldots, t-1\right\}, \frac{\Psi_{\ell,i}}{\ell} < p_i^* + \frac{\epsilon}{2}$, define

$$L_t \triangleq \max\left\{\ell \in \left\{\tilde{T}_\epsilon, \tilde{T}_\epsilon + 1, \ldots, t-1\right\} : \frac{\Psi_{\ell,i}}{\ell} < p_i^* + \frac{\epsilon}{2}\right\}.$$

Then we have the bound

$$\begin{aligned}\frac{\Psi_{t,i}}{t} = \frac{1}{t}\sum_{\ell=1}^{L_t - 1} \psi_{\ell,i} + \frac{\psi_{L_t,i}}{t} + \frac{1}{t}\sum_{\ell=L_t+1}^{t-1} \psi_{\ell,i} &= \frac{\Psi_{L_t,i}}{t} + \frac{\psi_{L_t,i}}{t} + \frac{1}{t}\sum_{\ell=L_t+1}^{t-1} \psi_{\ell,i}\mathbf{1}\left(\frac{\Psi_{\ell,i}}{\ell} \geqslant p_i^* + \frac{\epsilon}{2}\right) \\ &\leqslant \left(p_i^* + \frac{\epsilon}{2}\right) + \frac{1}{t} + \frac{1}{t}\sum_{\ell=L_t+1}^{t-1} \exp\left(-c_\epsilon \ell\right) \\ &\leqslant \left(p_i^* + \frac{\epsilon}{2}\right) + \frac{1}{t}\left(1 + \kappa_\epsilon\right)\end{aligned}$$

Putting these together the two cases, we find that for any $t \geqslant \tilde{T}_\epsilon$,

$$\frac{\Psi_{t,i}}{t} \leqslant \max\left\{\frac{1}{t}\left(\tilde{T}_\epsilon - 1 + \kappa_\epsilon\right), \left(p_i^* + \frac{\epsilon}{2}\right) + \frac{1}{t}\left(1 + \kappa_\epsilon\right)\right\} \leqslant \left(p_i^* + \frac{\epsilon}{2}\right) + \frac{\tilde{T}_\epsilon + 2\kappa_\epsilon}{t}.$$

Then we have

$$t \geqslant T_\epsilon \triangleq \frac{2\left(\tilde{T}_\epsilon + 2\kappa_\epsilon\right)}{\epsilon} \quad \Longrightarrow \quad \frac{\Psi_{t,i}}{t} \leqslant p_i^* + \epsilon.$$

Since $\tilde{T}_\epsilon \in \mathbb{M}$, so does $T_\epsilon$. This completes the proof. □

The above result shows that no arm is over sampled in the asymptotic regime, which further implies that no arm is under sampled since the summation of empirical sampling proportions always equals one. With this, we complete the proof of Proposition 1 part 1.

*Proof of Proposition 1 part 1.* Fix $i \in [k]$. By Lemma 21, there exists $\tilde{T}_\epsilon \in \mathbb{M}$ such that for any $t \geqslant \tilde{T}_\epsilon$,

$$\frac{\Psi_{t,j}}{t-1} \leqslant p_j^* + \frac{\epsilon}{k-1}, \quad \forall j \in [k],$$

and thus

$$\frac{\Psi_{t,i}}{t-1} = 1 - \sum_{j \neq i} \frac{\Psi_{t,j}}{t-1} \geqslant 1 - \sum_{j \neq i}\left(p_j^* + \frac{\epsilon}{k-1}\right) = p_i^* - \epsilon.$$

Hence, $\frac{\Psi_{t,i}}{t} \xrightarrow{\mathbb{M}} p_i^*$, and by Corollary 1, $\frac{N_{t,i}}{t} \xrightarrow{\mathbb{M}} p_i^*$.

□

### 7.4 Proof of Proposition 1 part 2: Strong convergence of z-scores

*Proof of Proposition 1 part 2.* Fix $j \in [k]$. By Corollary 3, there exists $T \in \mathbb{M}$ such that

$$t \geqslant T \quad \Longrightarrow \quad \hat{I}_t = \arg\max_{i\in[k]} m_{t,i} = I^* \quad \Longrightarrow \quad Z_{t,\hat{I}_t,j} = Z_{t,I^*,j} = \frac{m_{t,I^*} - m_{t,j}}{\sqrt{s_{t,I^*}^2 + s_{t,j}^2}} = \sqrt{2t\Gamma_{t,j}^{-1}}.$$

By Lemmas 14 and 15 and Proposition 1 part 1, for $i \in \{j, I^*\}$,

$$m_{t,i} \xrightarrow{\mathbb{M}} \mu(\theta, i, w) \quad \text{and} \quad ts_{t,i}^2 = \frac{t}{N_{t,i}} N_{t,i} s_{t,i}^2 \xrightarrow{\mathbb{M}} \frac{\sigma^2 \|X_{\text{pop}}\|_{\Lambda^{-1}}^2}{p_i^*},$$

Since the above expression $Z_{t,I^*,j}^2$ is a continuous function of $\left(m_{t,I^*}, m_{t,j}, ts_{t,I^*}^2, ts_{t,j}^2\right)$, we have that

$$\frac{Z_{t,I^*,i}^2}{t} \xrightarrow{\mathbb{M}} \frac{(\mu(\theta, I^*, w) - \mu(\theta, j, w))^2}{\sigma^2 \|X_{\text{pop}}\|_{\Lambda^{-1}}^2 \left[(p_{I^*}^*)^{-1} + (p_j^*)^{-1}\right]} = 2\Gamma_\theta^{-1}.$$

□

## 8 Proof of Lemma 4 in Section 5

In this section, we show that for each (path-dependent) random variable $W_1, W_2, W_3$ defined in Section 5, its moment generating function is bounded, which completes the proof of Lemma 4 in Section 5.

### 8.1 Maximal inequality for prediction errors

Recall that to control the impact of random contexts and observation noises, we introduce the following path-dependent random variable in Equation (5):

$$W_1 = \sup_{(t,i)\in\mathbb{N}\times[k]} \frac{|m_{t,i} - \mu(\theta, i, w)|}{s_{t,i}\sqrt{\log(N_{t,i} + e)}}.$$

The following lemma extends Lemma 5 in Qin et al. [2017] to our problem here with i.i.d. contexts.

**Lemma 22.** *Under Assumption 1, for any $\lambda \in \mathbb{R}$, $\mathbb{E}[e^{\lambda W_1}] < \infty$.*

As discussed in Subsubsection F.1.1, to simulate an algorithm, we could generate all the randomness upfront by using the so-called latent reward and context tables. We define the following counterpart of $W_1$:

$$\tilde{W}_1 \triangleq \sup_{(n,i)\in\mathbb{N}_0\times[k]} \frac{|\tilde{m}_{n,i} - \mu(\theta, i, w)|}{\tilde{s}_{n,i}\sqrt{\log(n + e)}}.$$

where $\tilde{m}_{n,i}$ and $\tilde{s}_{n,i}$ are defined in Equation (43). If every arm is played infinitely often, $W_1 = \tilde{W}_1$. One always has $W_1 \leqslant \tilde{W}_1$, so it suffices to prove the moment generating functions for $\tilde{W}_1$ is bounded.

*Proof of Lemma 22.* We define

$$\tilde{W}_{1,1} \triangleq \sup_{(n,i)\in\mathbb{N}_0\times[k]} \frac{\left|\tilde{m}_{n,i} - \mu(\theta, i, w) - \tilde{B}_{n,i}\right|}{\tilde{s}_{n,i}\sqrt{\log(n + e)}} \quad \text{and} \quad \tilde{W}_{1,0} \triangleq \sup_{(n,i)\in\mathbb{N}_0\times[k]} \frac{\left|\tilde{B}_{n,i}\right|}{\tilde{s}_{n,i}\sqrt{\log(n + e)}}$$

where the bias $\tilde{B}_{n,i}$ is defined in Lemma 21. By triangle inequality, we have $\tilde{W}_1 \leqslant \tilde{W}_{1,0} + \tilde{W}_{1,1}$, and thus it suffices to show the boundedness of moment generating functions for $\tilde{W}_{1,0}$ and $\tilde{W}_{1,1}$, respectively.

We first bound $\tilde{W}_{1,0}$. By the first bound in Lemma 21,

$$\tilde{W}_{1,0} \leqslant \sup_{(n,i)\in\mathbb{N}_0\times[k]} \frac{\|\tilde{\mu}_{0,i} - \theta^{(i)}\|_{\tilde{\Sigma}_{0,i}^{-1}}}{\sqrt{\log(n+e)}} = \max_{i\in[k]} \|\tilde{\mu}_{0,i} - \theta^{(i)}\|_{\tilde{\Sigma}_{0,i}^{-1}}.$$

Since $\tilde{W}_{1,0}$ is bounded by a constant, its moment generating function of $\tilde{W}_{1,0}$ is bounded.

Next we analyze $\tilde{W}_{1,1}$. Note that when $n = 0$,

$$\left|\tilde{m}_{0,i} - \mu(\theta, i, w) - \tilde{B}_{0,i}\right| = 0, \quad \forall i \in [k].$$

Hence, we only need to consider $n \geqslant 1$, and thus

$$\tilde{W}_{1,1} = \sup_{(n,i)\in\mathbb{N}\times[k]} \frac{\left|\tilde{m}_{n,i} - \mu(\theta, i, w) - \tilde{B}_{n,i}\right|}{\tilde{s}_{n,i}\sqrt{\log(n+e)}}.$$

where we replace $\mathbb{N}_0$ with $\mathbb{N}$.

Let $X \triangleq \left\{\tilde{X}_{n',i'}\right\}_{(n',i')\in\mathbb{N}\times[k]}$ be the context table. Then for all $x \geqslant 2$,

$$\begin{aligned}
\mathbb{P}\left(\tilde{W}_{1,1} \geqslant 2x \mid X\right) =& \mathbb{P}\left(\exists (n,i) \in \mathbb{N}\times[k] : \frac{\left|\tilde{m}_{n,i} - \mu(\theta, i, w) - \tilde{B}_{n,i}\right|}{\tilde{s}_{n,i}\sqrt{\log(n+e)}} \geqslant 2x \mid X\right) \\
\leqslant& \sum_{i\in[k]}\sum_{n\in\mathbb{N}} \mathbb{P}\left(\frac{\left|\tilde{m}_{n,i} - \mu(\theta, i, w) - \tilde{B}_{n,i}\right|}{\sqrt{\tilde{s}_{n,i}^2 - \left\|\tilde{\Sigma}_{n,i} X_{\text{pop}}\right\|_{\tilde{\Sigma}_{0,i}^{-1}}^2}} \geqslant 2x\sqrt{\log(n+e)} \mid X\right) \\
\leqslant& 2k \sum_{n\in\mathbb{N}} \exp\left(-2x^2\log(n+e)\right) \\
\leqslant& 2k \sum_{n\in\mathbb{N}} \exp\left(-x^2 - 2\log(n+e)\right) \\
=& ce^{-x^2}
\end{aligned}$$

where $c \triangleq 2k\sum_{n\in\mathbb{N}}(n+e)^{-2} < \infty$; the inequalities follow from the union bound, Lemmas 21 and 25, and $ab \geqslant a + b$ when $a, b \geqslant 2$, respectively. By integrating over the context table $X = \left\{\tilde{X}_{n,i}\right\}_{(n,i)\in\mathbb{N}\times[k]}$,

$$\mathbb{P}\left(\tilde{W}_{1,1} \geqslant 2x\right) \leqslant ce^{-x^2}, \quad \forall x \geqslant 2.$$

It is clear that $\mathbb{E}\left[e^{\lambda\tilde{W}_{1,1}}\right] < \infty$ for $\lambda \leqslant 0$. Then for $\lambda > 0$,

$$\begin{aligned}
\mathbb{E}\left[e^{\lambda\tilde{W}_{1,1}}\right] = \int_{x=1}^{\infty} \mathbb{P}\left(e^{\lambda\tilde{W}_{1,1}} \geqslant x\right) \mathrm{d}x \overset{(*)}{=} \int_{u=0}^{\infty} \mathbb{P}\left(e^{\lambda\tilde{W}_{1,1}} \geqslant e^{2\lambda u}\right) 2\lambda e^{2\lambda u}\mathrm{d}u \\
= 2\lambda\int_{u=0}^{2} \mathbb{P}\left(\tilde{W}_{1,1} \geqslant 2u\right) e^{2\lambda u}\mathrm{d}u + 2\lambda\int_{u=2}^{\infty} \mathbb{P}\left(\tilde{W}_{1,1} \geqslant 2u\right) e^{2\lambda u}\mathrm{d}u \\
\leqslant (e^{4\lambda} - 1) + 2\lambda c\int_{u=2}^{\infty} e^{-u^2}\cdot e^{2\lambda u}\mathrm{d}u \ < \infty
\end{aligned}$$

where in step $(*)$, we have substituted $x = e^{2\lambda u}$. This concludes the proof. □

## 8.2 Maximal inequality for randomized action selections

To control the impact of randomness in action selection, we introduce the path-dependent random variable in Equation (6):

$$W_2 = \sup_{(t,i)\in\mathbb{N}\times[k]} \frac{|N_{t,i} - \Psi_{t,i}|}{\sqrt{(t+1)\log(t+e^2)}}$$

where two measures of cumulative effort $N_{t,i}$ and $\Psi_{t,i}$ are defined in Equations (1) and (2).

Although Shang et al. [2020] studies the problem without contexts, the following lemma also applies to our contextual problem.

**Lemma 23** (Lemma 4 in Shang et al. [2020])**.** *For any $\lambda \in \mathbb{R}$, $\mathbb{E}[e^{\lambda W_2}] < \infty$.*

## 8.3 Maximal inequality for posterior covariance matrices

To control the impact of i.i.d. contexts in updating the posterior covariance matrices, we introduce the following path-dependent random variable in Equation (7):

$$W_3 = \sup_{(t,i)\in\mathbb{N}\times[k]} \frac{\left\|\Sigma_{t,i}^{-1} - A_{t,i}^{-1}\right\|}{\sqrt{(N_{t,i}+1)\log(N_{t,i}+e)}}$$

where

$$\Sigma_{t,i}^{-1} = \Sigma_{1,i}^{-1} + \sigma^{-2}\sum_{\ell=1}^{t-1} \mathbb{1}\{I_\ell = i\} X_\ell X_\ell^\top \quad \text{and} \quad A_{t,i}^{-1} = \sigma^{-2}\Lambda(N_{t,i}+1).$$

Recall that $\Lambda = \mathbb{E}[X_1 X_1^\top]$. The following result shows that its moment generating function is bounded.

**Lemma 24.** *Under Assumption 1, for any $\lambda \in \mathbb{R}$, $\mathbb{E}[e^{\lambda W_3}] < \infty$.*

As discussed in Subsubsection F.1.1, to simulate an algorithm, we could generate all the randomness upfront by using the so-called latent reward and context tables. We define the following counterpart of $W_3$: Recall in Equation (42),

$$\tilde{\Sigma}_{n,i} = \left(\tilde{\Sigma}_{0,i}^{-1} + \sigma^{-2}\sum_{\ell=1}^{n} \tilde{X}_{\ell,i}\tilde{X}_{\ell,i}^\top\right)^{-1}.$$

We define the counterpart of $W_3$:

$$\tilde{W}_3 \triangleq \sup_{(n,i)\in\mathbb{N}_0\times[k]} \frac{\left\|\tilde{\Sigma}_{n,i}^{-1} - \sigma^{-2}\Lambda(n+1)\right\|}{\sqrt{(n+1)\log(n+e)}}.$$

where

$$\tilde{\Sigma}_{n,i}^{-1} = \tilde{\Sigma}_{0,i}^{-1} + \sigma^{-2}\sum_{\ell=1}^{n} \tilde{X}_{\ell,i}\tilde{X}_{\ell,i}^\top.$$

is defined in Equation (42). If every arm is played infinitely often, $W_3 = \tilde{W}_3$. One always has $W_3 \leqslant \tilde{W}_3$, so it suffices to prove the moment generating function of $\tilde{W}_3$ is bounded.

*Proof.* We further decompose $\tilde{W}_3$ into

$$\tilde{W}_{3,0} \triangleq \sup_{(n,i)\in\mathbb{N}_0\times[k]} \frac{\left\|\tilde{\Sigma}_{0,i}^{-1} - \sigma^{-2}\Lambda\right\|}{\sqrt{(n+1)\log(n+e)}} \quad \text{and} \quad \tilde{W}_{3,1} \triangleq \sup_{(n,i)\in\mathbb{N}\times[k]} \frac{\|Q_{n,i}\|}{\sqrt{(n+1)\log(n+e)}}$$

where

$$Q_{n,i} \triangleq \sum_{\ell=1}^{n} \left( \sigma^{-2} \tilde{X}_{\ell,i} \tilde{X}_{\ell,i}^\top - \sigma^{-2} \Lambda \right).$$

By triangle inequality, we have $\tilde{W}_3 \leqslant \tilde{W}_{3,0} + \tilde{W}_{3,1}$. To prove Lemma 24, it suffices to show the boundedness of the moment generating functions of $\tilde{W}_{3,0}$ and $\tilde{W}_{3,1}$, respectively. Since $\tilde{W}_{3,0} = \max_{i \in [k]} \left\| \tilde{\Sigma}_{0,i}^{-1} - \sigma^{-2} \Lambda \right\|$ is a constant, its moment generating function is clearly bounded.

Now we analyze $\tilde{W}_{3,1}$. By triangle inequality,

$$\left\| \sigma^{-2} \tilde{X}_{\ell,i} \tilde{X}_{\ell,i}^\top - \sigma^{-2} \Lambda \right\| \leqslant \left\| \sigma^{-2} \tilde{X}_{\ell,i} \tilde{X}_{\ell,i}^\top \right\| + \left\| \sigma^{-2} \Lambda \right\| \leqslant 2 b_{\max}$$

where $b_{\max}$ is defined in Section 3. This gives

$$\left( \sigma^{-2} \tilde{X}_{\ell,i} \tilde{X}_{\ell,i}^\top - \sigma^{-2} \Lambda \right)^2 \leqslant 4 b_{\max}^2 I.$$

Let $X \triangleq \left\{ \tilde{X}_{n',i'} \right\}_{(n',i') \in \mathbb{N} \times [k]}$ be the context table. By applying Lemma 26 (Matrix Hoeffding), for all $x \geqslant 0$,

$$\begin{aligned} \mathbb{P}\left( \|Q_{n,i}\| \geqslant x \mid X \right) &= \mathbb{P}\left( \left\| \sum_{\ell=1}^{n} \left( \sigma^{-2} \tilde{X}_{\ell,i} \tilde{X}_{\ell,i}^\top - \sigma^{-2} \Lambda \right) \right\| \geqslant x \mid X \right) \\ &\leqslant 2d \cdot \exp\left( \frac{-x^2}{32 n b_{\max}^2} \right). \end{aligned}$$

Then for all $x \geqslant 16 b_{\max}$,

$$\begin{aligned} \mathbb{P}\left( \tilde{W}_{3,1} \geqslant x \mid X \right) =& \mathbb{P}\left( \exists (n,i) \in \mathbb{N} \times [k] \,:\, \|Q_{n,i}\| \geqslant x \sqrt{(n+1)\log(n+e)} \mid X \right) \\ \leqslant& \sum_{i \in [k]} \sum_{n \in \mathbb{N}} \mathbb{P}\left( \|Q_{n,i}\| \geqslant x \sqrt{(n+1)\log(n+e)} \mid X \right) \\ \leqslant& 2dk \sum_{n \in \mathbb{N}} \exp\left( -\frac{x^2 \log(n+e)}{32 b_{\max}^2} \right) \\ \leqslant& 2dk \sum_{n \in \mathbb{N}} \exp\left( -\frac{x^2}{64 b_{\max}^2} - 2\log(n+e) \right) \\ =& c e^{-\frac{x^2}{64 b_{\max}^2}} \end{aligned}$$

where $c \triangleq 2dk \sum_{n \in \mathbb{N}} (n+e)^{-2} < \infty$ is a constant; the first and third inequalities follow from the union bound and $ab \geqslant a + b$ when $a, b \geqslant 2$, respectively. By integrating over the context table $X = \left\{ \tilde{X}_{n,i} \right\}_{(n,i) \in \mathbb{N} \times [k]}$,

$$\mathbb{P}\left( \tilde{W}_{3,1} \geqslant x \right) \leqslant c e^{-\frac{x^2}{64 b_{\max}^2}}, \quad \forall x \geqslant 0.$$

It is clear that $\mathbb{E}\left[e^{\lambda \tilde{W}_{3,1}}\right] < \infty$ for $\lambda \leqslant 0$. Then for $\lambda > 0$,

$$\begin{aligned}
\mathbb{E}\left[e^{\lambda \tilde{W}_{3,1}}\right] = \int_{x=1}^{\infty} \mathbb{P}\left(e^{\lambda \tilde{W}_{3,1}} \geqslant x\right) \mathrm{d}x &\overset{(*)}{=} \int_{u=0}^{\infty} \mathbb{P}\left(e^{\lambda \tilde{W}_{3,1}} \geqslant e^{\lambda u}\right) \lambda e^{\lambda u} \mathrm{d}u \\
&= \lambda \int_{u=0}^{16 b_{\max}} \mathbb{P}\left(\tilde{W}_{3,1} \geqslant u\right) e^{\lambda u} \mathrm{d}u + \lambda \int_{u=16 b_{\max}}^{\infty} \mathbb{P}\left(\tilde{W}_{3,1} \geqslant u\right) e^{\lambda u} \mathrm{d}u \\
&\leqslant \lambda \int_{u=0}^{16 b_{\max}} e^{\lambda u} \mathrm{d}u + \lambda c \int_{u=16 b_{\max}}^{\infty} e^{-\frac{u^2}{64 b_{\max}^2}} \cdot e^{\lambda u} \mathrm{d}u \ < \infty
\end{aligned}$$

where in step $(*)$, we have substituted $x = e^{\lambda u}$. This concludes the proof.

□

# 9 Technical lemmas

**Lemma 25** (Upper and Lower Bounds of Gaussian Tail)**.** *Let $X \sim N(0,1)$. Then, for all $x \geqslant 0$,*

$$e^{-\left(x+\sqrt{2\pi}\right)^2/2} \leqslant \mathbb{P}(X \leqslant -x) = \mathbb{P}(X \geqslant x) \leqslant e^{-x^2/2}.$$

*Proof.* The upper bound is well-known. Now we are going to prove this version of the lower bound.

$$\begin{aligned}
\mathbb{P}(X \geqslant x) &= \int_x^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \mathrm{d}z \\
&= \int_0^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(x+u)^2/2} \mathrm{d}u \\
&\geqslant \int_0^{\sqrt{2\pi}} \frac{1}{\sqrt{2\pi}} e^{-(x+u)^2/2} \mathrm{d}u \\
&\geqslant \int_0^{\sqrt{2\pi}} \frac{1}{\sqrt{2\pi}} e^{-\left(x+\sqrt{2\pi}\right)^2/2} \mathrm{d}u \\
&\geqslant e^{-\left(x+\sqrt{2\pi}\right)^2/2}.
\end{aligned}$$

□

**Lemma 26** (Matrix Hoeffding [Tropp, 2012])**.** *Consider a finite sequence $\{X_n\}$ of independent, random, self-adjoint matrices with dimension $d$ and a sequence $\{Y_n\}$ of fixed self-adjoint matrices. Assume that each random matrix satisfies*

$$\mathbb{E}[X_n] = 0 \quad \textit{and} \quad X_n^2 \preceq Y_n^2 \quad \textit{almost surely.}$$

*Then, for all $x \geqslant 0$,*

$$\mathbb{P}\left(\left\|\sum_n X_n\right\| \geqslant x\right) \leqslant 2d \cdot \exp\left(\frac{-x^2}{8\left\|\sum_n Y_n^2\right\|}\right).$$

# References


Chao Qin and Daniel Russo. Adaptivity and confounding in multi-armed bandit experiments. *arXiv preprint arXiv:2202.09036*, 2022.

Chao Qin, Diego Klabjan, and Daniel Russo. Improving the expected improvement algorithm. *Advances in Neural Information Processing Systems*, 2017:5382–5392, 2017.

Daniel Russo. Simple bayesian algorithms for best-arm identification. *Operations Research*, 2020.

Xuedong Shang, Rianne de Heide, Pierre Menard, Emilie Kaufmann, and Michal Valko. Fixed-confidence guarantees for bayesian best-arm identification. In Silvia Chiappa and Roberto Calandra, editors, *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, volume 108 of *Proceedings of Machine Learning Research*, pages 1823–1832. PMLR, 26–28 Aug 2020.

Joel A. Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of computational mathematics*, 12(4):389–434, 2012.