The Intrinsic Robustness of Stochastic Bandits to Strategic Manipulation

Appendix

A. Useful Definitions and Inequalities

Definition A.1 (σ-sub-Gaussian)**.** *A random variable $X \in \mathbb{R}$ is said to be sub-Gaussian with variance proxy σ^2 if $\mathbb{E}[X] = \mu$ and satisfies,*

$$\mathbb{E}\left[\exp(s(X-\mu))\right] \leq \exp\left(\frac{\sigma^2 s^2}{2}\right), \forall s \in \mathbb{R}$$

Note the distribution defined on $[0,1]$ is a special case of $1/2$-sub-Gaussian.

Fact A.2. *Let $X_1, X_2, \cdots, X_n$ i.i.d drawn from a σ-sub-Gaussian, $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$ and $\mathbb{E}[X]$ be the mean, then*

$$\mathbb{P}\left(\overline{X} - \mathbb{E}[X] \geq a\right) \leq e^{-na^2/2\sigma^2} \quad \text{and } \mathbb{P}\left(\overline{X} - \mathbb{E}[X] \leq -a\right) \leq e^{-na^2/2\sigma^2}$$

Fact A.3 (Harmonic Sequence Bound)**.** *For $t_2 > t_1 \geq 2$, we have*

$$\ln\frac{t_2}{t_1} \leq \sum_{t=t_1}^{t_2}\frac{1}{t} \leq \ln\left(\frac{t_2}{t_1 - 1}\right)$$

Fact A.4. *For a Gaussian distributed random variable Z with mean μ and variance σ^2, for any z,*

$$\mathbb{P}\left(|Z-\mu| > z\sigma\right) \leq \frac{1}{2}e^{-z^2/2}$$

Lemma A.5 (Theorem 3 in (Auer et al., 2002a))**.** *In ε-Greedy, for any arm $k \in [K], t > K, n \in \mathbb{N}_+$, we have*

$$\mathbb{P}\left(\widehat{\mu}_k(t-1) \leq \mu_k - \frac{\Delta_k}{n}\right) \leq x_t \cdot e^{-x_t/5} + \frac{2\sigma^2 n^2}{\Delta_k^2} e^{-\Delta_k^2 \lfloor x_t \rfloor / 2\sigma^2 n^2}, \quad \text{and}$$
$$\mathbb{P}\left(\widehat{\mu}_{i^*}(t-1) \geq \mu_{i^*} + \frac{\Delta_k}{n}\right) \leq x_t \cdot e^{-x_t/5} + \frac{2\sigma^2 n^2}{\Delta_k^2} e^{-\Delta_k^2 \lfloor x_t \rfloor / 2\sigma^2 n^2},$$

where $x_t = \frac{1}{2K}\sum_{s=K+1}^t \varepsilon_s$.

B. Ommited Proofs in Section 3

B.1. Proof of Lemma 3.2

Proof. Let $C_i(T) = \max\left\{\frac{81\sigma^2 \ln T}{\Delta_i^2}, \frac{3B_i}{\Delta_i}\right\}$. By Fact A.2, we have for any $s \geq 1$ and $\ell \geq C_i(T)$

$$\begin{aligned} \forall k, \ \mathbb{P}\left(\mu_k - \widehat{\mu}_k(t-1) \geq 3\sigma\sqrt{\frac{\ln T}{n_k(t-1)}} \middle| n_k(t-1) = s\right) &\leq \frac{1}{T^{9/2}} \\ \mathbb{P}\left(\widehat{\mu}_i(t-1) - \mu_i \geq \frac{\Delta_i}{3} \middle| n_i(t-1) = \ell\right) &\leq \frac{1}{T^{9/2}} \end{aligned} \tag{7}$$

We first decompose $\mathbb{E}[n_i(T)]$ as follows,

$$\begin{aligned} \mathbb{E}\left[n_i(T)\right] &\leq 1 + \mathbb{E}\left[\sum_{t=K+1}^T \mathbb{I}\{I_t = i, n_i(t-1) \leq C_i(T)\}\right] + \mathbb{E}\left[\sum_{t=K+1}^T \mathbb{I}\{I_t = i, n_i(t-1) \geq C_i(T)\}\right] \\ &\leq 1 + C_i(T) + \mathbb{E}\left[\sum_{t=K+1}^T \mathbb{I}\{I_t = i, n_i(t-1) \geq C_i(T)\}\right] \\ &\leq 1 + C_i(T) + \sum_{t=K+1}^T \mathbb{P}\left(\mathrm{UCB}_i(t) + \frac{\beta_{t-1}^{(i)}}{n_i(t-1)} \geq \mathrm{UCB}_{i^*}(t), n_i(t-1) \geq C_i(T)\right) \end{aligned} \tag{8}$$

We then bound the probability $\mathbb{P}\left(\mathrm{UCB}_i(t)+\frac{\beta_{t-1}^{(i)}}{n_i(t-1)}\geq \mathrm{UCB}_{i^*}(t), n_i(t-1)\geq C_i(T)\right)$ by union bound, and decompose this probability term as follows,

$$\begin{aligned}&\mathbb{P}\left(\mathrm{UCB}_i(t)+\frac{\beta_{t-1}^{(i)}}{n_i(t-1)}\geq \mathrm{UCB}_{i^*}(t), n_i(t-1)\geq C_i(T)\right)\\&\leq \sum_{s=1}^{t-1}\sum_{\ell\geq C_i(T)}^{t-1}\mathbb{P}\left(\mathrm{UCB}_i(t)+\frac{\beta_{t-1}^{(i)}}{n_i(t-1)}\geq \mathrm{UCB}_{i^*}(t)\Big| n_i(t-1)=\ell, n_{i^*}(t-1)=s\right).\end{aligned}\tag{9}$$

What remains is to upper bound the summand in the above term. Consider for $1\leq s\leq t-1$ and $C_i(T)\leq \ell\leq t-1$, we have

$$\begin{aligned}&\mathbb{P}\left(\mathrm{UCB}_i(t)+\frac{\beta_{t-1}^{(i)}}{n_i(t-1)}\geq \mathrm{UCB}_{i^*}(t)\Big| n_i(t-1)=\ell, n_{i^*}(t-1)=s\right)\\&\leq \mathbb{P}\left(\widehat{\mu}_i(t-1)+3\sigma\sqrt{\frac{\ln T}{n_i(t-1)}}+\frac{\Delta_i}{3}\geq \widehat{\mu}_{i^*}(t-1)+3\sigma\sqrt{\frac{\ln T}{n_{i^*}(t-1)}}\Big| n_i(t-1)=\ell, n_{i^*}(t-1)=s\right)\\&\leq \mathbb{P}\left(\widehat{\mu}_i(t-1)+\frac{\Delta_i}{3}+\frac{\Delta_i}{3}\geq \widehat{\mu}_{i^*}(t-1)+3\sigma\sqrt{\frac{\ln T}{n_{i^*}(t-1)}}\Big| n_i(t-1)=\ell, n_{i^*}(t-1)=s\right)\end{aligned}$$

The first inequality relies on the fact that $\ell\geq C_i(T)\geq \frac{3B_i}{\Delta_i}\geq \frac{3\beta_{t-1}^{(i)}}{\Delta_i}$ and second inequality holds because $\ell\geq C_i(T)\geq \frac{81\sigma^2\ln T}{\Delta_i^2}$. By union bound and Equation (7), we can further upper bound the last term in the above inequality by

$$\begin{aligned}&\mathbb{P}\left(\widehat{\mu}_i(t-1)-\mu_i\geq\frac{\Delta_i}{3}\Big| n_i(t-1)=\ell\right)+\mathbb{P}\left(\mu_{i^*}-\widehat{\mu}_{i^*}(t-1)\geq 3\sigma\sqrt{\frac{\ln T}{n_{i^*}(t-1)}}\Big| n_{i^*}(t-1)=s\right)\\&\leq\frac{1}{T^{9/2}}+\frac{1}{T^{9/2}}\leq\frac{2}{T^{9/2}}\end{aligned}$$

Combining Equations (8) and the fact that

$$\sum_{t=K+1}^{T}\sum_{s=1}^{t-1}\sum_{\ell\geq C_i(T)}^{t-1}\frac{2}{t^{T/2}}\leq\sum_{t=K+1}^{T}\frac{2}{T^2}\leq 2,$$

we complete the proof. □

B.2. Proof of Theorem 3.3

We begin with a few notations. Let I_t^S denote the arm being pulled at time t for any investment strategy S, and $Z_t^S=\{I_1^S,\cdots,I_t^S\}$ denote the *sequence* of arms being pulled *up to* time t. Note that $Z_t^S=\{I_1^S,\cdots,I_t^S\}$ can be viewed as a stochastic process for any t. Let $S^{(-i)}$ denote the investment strategies of all arms excluding arm i. In addition, we denote by $(\mathtt{LSI}, S^{(-i)})$ the strategy that arm i uses LSI strategy and the other arms adopt $S^{(-i)}$. For each arm $j\neq i$, $S^{(j)}$ only depends on its own history, which means given fixed strategies $S^{(-i)}$, at any time t, each of the arms $j\neq i$ will invest the same budget if it has been pulled the same times and the true rewards are the same up to time t.

Our proof of Theorem 3.3 relies on a carefully chosen coupling of the two stochastic processes $Z_T^{S_1}, Z_T^{S_2}$ induced by different investment strategies S_1, S_2, respectively.

Definition B.1 (Arm Coupling). *Given any two investment strategies S_1, S_2, the* **Arm Coupling** *of $Z_T^{S_1}$ and $Z_T^{S_2}$ is a coupling of these two stochastic processes such that the reward of any arm $k\in[K]$ when pulled for the same times is the same in these two random processes. In this case, we also say $Z_T^{S_1}=\{I_1^{S_1},\cdots,I_T^{S_1}\}$ and $Z_T^{S_2}=\{I_1^{S_2},\cdots,I_T^{S_2}\}$ are* **Arm-Coupled.**

Our goal is to compare $\left(\texttt{LSI}, S^{(-i)}\right)$ and any other strategy $S = (S^{(i)}, S^{(-i)})$ for arm i, using **Arm Coupling**. In the remainder of this proof we will always fix all other arms' manipulation strategy $S^{(-i)}$. Thus for convenience we simply omit $S^{(-i)}$ in the superscript and use $Z_t^{\texttt{LSI}}$ and $Z_t^{S^{(i)}}$ to denote the two stochastic sequences of our interests. Let $Z_{t:t'}^{\texttt{LSI}}$ denote the stochastic process from time t to time t' under $\left(\texttt{LSI}, S^{(-i)}\right)$ manipulation, and similarly for $Z_{t:t'}^{S^{(i)}}$. Similar notations and simplifications are used for n_i. We first show LSI is the dominant strategy for the arm when principal runs UCB algorithm, given any history h_{t-1}. Hence LSI is a dominant-strategy SPE.

The following lemma shows an interesting property about the two arm sequences $Z_t^{\texttt{LSI}}$ and any $Z_t^{S^{(i)}}$ pulled under these two different investment strategies. That is, under **Arm Coupling**, all the arms — except for the special arm i — will be pulled according to the same order after time t, given any history h_{t-1}.

Lemma B.2. *Suppose $t \geq K$ and the principal runs UCB algorithm. Let $Z_{t:t'}^{\texttt{LSI}}(-i)$ [resp. $Z_{t:t'}^{S^{(i)}}(-i)$] denote the subsequence of $Z_{t:t'}^{\texttt{LSI}}$ [resp. $Z_{t:t'}^{S^{(i)}}$] after deleting all i's in the sequence. Then given any history h_{t-1} and time t, under* **Arm Coupling**, *either $Z_{t:t'}^{\texttt{LSI}}(-i)$ is a subsequence of $Z_{t:t'}^{S^{(i)}}(-i)$ or vice versa.*

Proof. We prove by induction on t'. When $t' = t$, if $I_t^{(\texttt{LSI}, S^{(-i)})}$ or I_t^S is i, the conclusion holds trivially. If $I_t^S = k \neq i$, then k is the largest UCB term. Since the history h_{t-1} is fixed, UCB terms of each arm must be the same, thus, if $I_t^S = k$, then $I_t^{\texttt{LSI}} = k$, as desired.

Now, assume the lemma holds for some $t'(> t)$, and we now consider the case $t' + 1$. This follows a case analysis.

If $n_i^{\texttt{LSI}}(t : t') = n_i^{S^{(i)}}(t : t')$, then we know that $Z_{t:t'}^{\texttt{LSI}}(-i)$ and $Z_{t:t'}^{S^{(i)}}(-i)$ have the same length. Since one of them is a subsequence of the other by induction hypothesis, this implies that they are the same sequence. If one of $I_{t'+1}^{\texttt{LSI}}, I_{t'+1}^{S^{(i)}}$ equals i, say, e.g., $I_{t'+1}^{\texttt{LSI}} = i$, then $Z_{t:t'+1}^{\texttt{LSI}}(-i) = Z_{t:t'}^{\texttt{LSI}}(-i) = Z_{t:t'}^{S^{(i)}}(-i)$ which is a subsequence of $Z_{t:t'+1}^{S^{(i)}}(-i)$, as desired. If both $I_{t'+1}^{\texttt{LSI}}, I_{t'+1}^{S^{(i)}}$ are not equal to i, then we claim that they must be the same arm. This is because they are the arm with the highest UCB index after round t. Since $Z_{t:t'}^{\texttt{LSI}}(-i)$ and $Z_{t:t'}^{S^{(i)}}(-i)$ are the same sequence of arms, each arms are pulled by exactly the same time in both stochastic processes from 0 to t', given the fixed history h_{t-1}. Moreover, due to **Arm Coupling**, their rewards are also the same. Given the fixed strategies of the other arms $S^{(-i)}$, their manipulations will also be the same. Therefore, the arm with the highest *modified UCB terms* must also be the same. Therefore, we have $Z_{t:t'+1}^{\texttt{LSI}}(-i) = Z_{t:t'+1}^{S^{(i)}}(-i)$, as desired.

If $n_i^{\texttt{LSI}}(t : t') > n_i^{S^{(i)}}(t : t')$, then we know that $Z_{t:t'}^{\texttt{LSI}}(-i)$ is a strict subsequence of $Z_{t:t'}^{S^{(i)}}(-i)$. Let $l = |Z_{t:t'}^{\texttt{LSI}}(-i)|$ denote the length of $Z_{t:t'}^{\texttt{LSI}}(-i)$, and $\tilde{k}$ denote the $(l+1)$th element in $Z_{t:t'}^{S^{(i)}}(-i)$. We claim that $I_{t'+1}^{\texttt{LSI}}$ must be either i or $\tilde{k}$, which implies $Z_{t:t'+1}^{\texttt{LSI}}(-i)$ is a subsequence of $Z_{t:t'+1}^{S^{(i)}}(-i)$ as desired. In particular, if $I_{t'+1}^{\texttt{LSI}} \neq i$, then the fact that $\tilde{k}$ is the $(l+1)$th element in $Z_{t:t'}^{S^{(i)}}(-i)$ implies that $\tilde{k}$ has the highest *modified UCB term* among all arms in $[K] \setminus \{i\}$ when these arms are pulled according to sequence $Z_t^{S^{(i)}}$. Following the same argument above and **Arm Coupling**, we know that $I_{t'+1}^{\texttt{LSI}}$, the arm with the highest *modified UCB term*, must equal $\tilde{k}$ if it does not equal i.

The case of $n_i^{\texttt{LSI}}(t : t') < n_i^{S^{(i)}}(t : t')$ can be argued similarly. This concludes the proof of the lemma. □

The following lemma shows that under **Arm Coupling**, the number of times that arm i is pulled up to time T under strategy LSI is always at least that under any other investment strategy $S^{(i)}$.

Lemma B.3. *When the principal runs UCB algorithm, under* **Arm Coupling**, *given any history h_{t-1} and time t, we have $n_i^{\texttt{LSI}}(t : T) \geq n_i^{S^{(i)}}(t : T)$ with probability 1 for any investment strategy S and $T \geq t \geq K$.*

Proof. We still prove through induction. Given any fixed $t \geq K$ and history h_{t-1}, for $T = t$, it holds trivially since if $I_t^{S^{(i)}} = i$ then $I_t^{\texttt{LSI}}$ must be i. We assume this lemma is true for $t' = T - 1 > t$. For $t' = T$, we consider the following two cases.

(1) If $n_i^{\texttt{LSI}}(t : T-1) > n_i^{S^{(i)}}(t : T-1)$, then $n_i^{\texttt{LSI}}(t : T) \geq n_i^{\texttt{LSI}}(t : T-1) \geq n_i^{S^{(i)}}(t : T-1) + 1 \geq n_i^{S^{(i)}}(t : T)$, as desired.

(2) If $n_i^{\texttt{LSI}}(t:T-1) = n_i^{S^{(i)}}(t:T-1)$, then Lemma B.2 implies that $Z_{t:T-1}^{\texttt{LSI}}$ and $Z_{t:T-1}^{S^{(i)}}$ are the same sequence. Therefore, the UCB term for each arm $k \in [K]$ (excluding arm i) for LSI and $S^{(i)}$ are the same at time T. For arm i, we have

$$\widehat{\mathrm{UCB}}_i^{(\texttt{LSI},S^{(-i)})}(T) = \mathrm{UCB}_i^{(\texttt{LSI},S^{(-i)})}(T) + \frac{B_i}{n_i^{(\texttt{LSI},S^{(-i)})}(T-1)} = \mathrm{UCB}_i^S(T) + \frac{B_i}{n_i^S(T-1)}$$
$$\geq \mathrm{UCB}_i^S(T) + \frac{\beta_{T-1}^{(i)}}{n_i^S(T-1)} = \widehat{\mathrm{UCB}}_i^S(T),$$

This implies that if $I_T^S = i$, then we must also have $I_T^{(\texttt{LSI},S^{(-i)})} = i$. Then $n_i^{(\texttt{LSI},S^{(-i)})}(T) \geq n_i^S(T)$ still holds.

To sum up, $n_i^{(\texttt{LSI},S^{(-i)})}(t:T) \geq n_i^S(t:T)$ holds with probability 1, concluding the proof. □

B.3. Proof of Theorem 3.4

We show the lower bound of the regret by deriving the upper bound of the expected number of times that arm i^* being pulled, which is summarized in Lemma B.4. Given Lemma B.4 and Eq. (3), it is straightforward to conclude Theorem 3.4 for UCB principal.

Lemma B.4. *Suppose each strategic arm $i(i \neq i^*)$ uses* LSI *and $\underline{\Delta} = \min_{i\neq i^*} \Delta_i$, the expected number of times that optimal arm i^* being pulled up to time T is bounded by,*

$$\mathbb{E}\left[n_{i^*}(T)\right] \leq T - \sum_{i\neq i^*} \frac{B_i}{2\Delta_i} + \mathcal{O}\left(\frac{\ln T}{\underline{\Delta}^2}\right)$$

Proof. Let $\underline{\Delta} = \min_{i\neq i^*} \Delta_i, C(T) = \frac{36\sigma^2 \ln T}{\underline{\Delta}^2}, D_i = \frac{B_i}{2\Delta_i}$. First, by Fact A.2, we have for any $\ell \geq C(T)$, $s \geq 1$ and any i,

$$\begin{aligned}
\mathbb{P}\left(\mu_i - \widehat{\mu}_i(t-1) \geq 3\sigma\sqrt{\frac{\ln T}{n_i(t-1)}} \middle| n_i(t-1) = s\right) &\leq \frac{1}{T^{9/2}} \\
\mathbb{P}\left(\widehat{\mu}_{i^*}(t-1) - \mu_{i^*} \geq \frac{\Delta_i}{2} \middle| n_{i^*}(t-1) = \ell\right) \leq \exp\left(-\frac{\ell\Delta_i^2}{8\sigma^2}\right) &\leq \exp\left(-\frac{C(T)\Delta_i^2}{8\sigma^2}\right) \leq \frac{1}{T^{9/2}}
\end{aligned} \tag{10}$$

First, we decompose $\mathbb{E}\left[n_{i^*}(T)\right]$ as follows,

$$\begin{aligned}
\mathbb{E}\left[n_{i^*}(T)\right] &\leq 1 + \mathbb{E}\left[\sum_{t=K+1}^{T} \mathbb{I}\{I_t = i^*, n_{i^*}(t-1) \leq C(T)\}\right] + \mathbb{E}\left[\sum_{t=K+1}^{T} \mathbb{I}\{I_t = i^*, n_{i^*}(t-1) \geq C(T)\}\right] \\
&\leq 1 + \mathbb{E}\left[\sum_{t=K+1}^{T} \mathbb{I}\{I_t = i^*, n_{i^*}(t-1) \leq C(T)\}\right] \\
&\quad + \mathbb{E}\left[\sum_{t=K+1}^{T} \mathbb{I}\{I_t = i^*, n_{i^*}(t-1) \geq C(T), \forall i \neq i^*, n_i(t-1) \geq D_i\}\right] \\
&\quad + \mathbb{E}\left[\sum_{t=K+1}^{T} \mathbb{I}\left(I_t = i^*, n_{i^*}(t-1) \geq C(T), \exists i \neq i^*, n_i(t-1) \leq D_i\right)\right]
\end{aligned} \tag{11}$$

For the first term in the above decomposition, it can be trivially bounded by $C(T)$. For the second term, since $n_{i^*}(t) \leq T - \sum_{i\neq i^*} n_i(t), \forall t$, we have

$$\begin{aligned}
&\mathbb{E}\left[\sum_{t=K+1}^{T} \mathbb{I}\{I_t = i^*, n_{i^*}(t-1) \geq C(T), \forall i \neq i^*, n_i(t-1) \geq D_i\}\right] \\
&\leq \mathbb{E}\left[\sum_{t=K+1}^{T} \mathbb{I}\{I_t = i^*, n_{i^*}(t-1) \leq T - \sum_{i\neq i^*} D_i\}\right] \leq T - \sum_{i\neq i^*} \frac{B_i}{2\Delta_i}
\end{aligned}$$

What remains is to bound the third term in Equations (11). By union bound, we have

$$\mathbb{E}\left[\sum_{t=K+1}^{T}\mathbb{I}\left(I_t=i^*, n_{i^*}(t-1)\geq C(T), \exists i\neq i^*, n_i(t-1)\leq D_i\right)\right]$$
$$=\sum_{i\neq i^*}\sum_{t=K+1}^{T}\mathbb{P}\left(I_t=i^*, n_{i^*}(t-1)\geq C(T), n_i(t-1)\leq D_i\right)$$

Note $I_t = i^*$ implies $\mathrm{UCB}_{i^*}(t) \geq \widehat{\mathrm{UCB}}_i(t)$, combining the facts that $3\sigma\sqrt{\frac{\ln T}{n_{i^*}(t-1)}} \leq \underline{\Delta}/2$ and $\frac{B_i}{n_i(t-1)} \geq 2\Delta_i$ and standard union bound, we have

$$\begin{aligned}
&\mathbb{P}\left(I_t=i^*, n_{i^*}(t-1)\geq C(T), n_i(t-1)\leq D_i\right)\\
&\leq \sum_{s=1}^{D_i\wedge t-1}\sum_{\ell\geq C(T)}^{t-1}\mathbb{P}\left(\widehat{\mu}_{i^*}(t-1)+3\sigma\sqrt{\frac{\ln T}{n_{i^*}(t-1)}}\geq \mathrm{UCB}_i(t)+\frac{B_i}{n_i(t-1)}\Big| n_{i^*}(t-1)=\ell, n_i(t-1)=s\right)\\
&\leq \sum_{s=1}^{D_i\wedge t-1}\sum_{\ell\geq C(T)}^{t-1}\mathbb{P}\left(\widehat{\mu}_{i^*}(t-1)+\frac{\Delta_i}{2}\geq \widehat{\mu}_i(t-1)+3\sigma\sqrt{\frac{\ln T}{n_i(t-1)}}+2\Delta_i, \Big| n_{i^*}(t-1)=\ell, n_i(t-1)=s\right) \qquad (12)\\
&\leq \sum_{s=1}^{D_i\wedge t-1}\sum_{\ell\geq C(T)}^{t-1}\mathbb{P}\left(\widehat{\mu}_{i^*}(t-1)-\mu_{i^*}\geq \frac{\Delta_i}{2}\Big| n_{i^*}(t-1)=\ell\right)+\mathbb{P}\left(\mu_i-\widehat{\mu}_i(t-1)\geq 3\sigma\sqrt{\frac{\ln T}{n_i(t-1)}}\Big| n_i(t-1)=s\right)
\end{aligned}$$

The last inequality is based on union bound, if both $\widehat{\mu}_{i^*}(t-1)-\mu_{i^*} < \underline{\Delta}/2$ and $\mu_i - \widehat{\mu}_i(t-1) < 3\sigma\sqrt{\frac{\ln T}{n_i(t-1)}}$ hold when $n_{i^*}(t-1)=\ell, n_i(t-1)=s$, then

$$\begin{aligned}
\widehat{\mu}_{i^*}(t-1)+\frac{\Delta_i}{2} &< \mu_{i^*}+\frac{\underline{\Delta}}{2}+\frac{\Delta_i}{2}\leq \mu_i+\Delta_i+\Delta_i\\
&< \widehat{\mu}_i(t-1)+3\sigma\sqrt{\frac{\ln T}{n_i(t-1)}}+2\Delta_i
\end{aligned}$$

Given Equation (10), we have

$$\mathbb{P}\left(I_t=i^*, n_{i^*}(t-1)\geq C(T), n_i(t-1)\leq D_i\right)\leq \sum_{s=1}^{t-1}\sum_{\ell=1}^{t-1}\frac{2}{T^{9/2}}\leq\frac{2}{T^2}$$

Combining Equation (11), we get

$$\begin{aligned}
\mathbb{E}\left[n_{i^*}(T)\right] &\leq 1+C(T)+T-\sum_{i\neq i^*}\frac{B_i}{2\Delta_i}+\sum_{i\neq i^*}\sum_{t=K+1}^{T}\frac{2}{T^2}\\
&= T+\frac{36\sigma^2\ln T}{\underline{\Delta}^2}-\sum_{i\neq i^*}\frac{B_i}{2\Delta_i}+1+\frac{2(K-1)}{T}
\end{aligned}$$

□

Combining Lemma B.4 and Eq/ 3, we complete the proof for Theorem 3.4.

B.4. Proof of Theorem 3.5

To prove Theorem 3.5, we first show the following Lemma.

Lemma B.5. *Suppose all the strategic arms use* LSIBR*, and let time step n be the last time that a strategic arm spend budget for some $n \leq T$. Then for the three algorithms we consider (UCB, ε-Greedy and TS), the expected number of plays of the optimal arm i^* from time $n+1$ to T is bounded by,*

$$\mathbb{E}\left[\sum_{t=n+1}^{T}\mathbb{I}\{I_t=i^*\}\right]\leq \mathbb{E}\left[n_{i^*}^{\mathtt{LSI}}(T)\right]=T-\sum_{i\neq i^*}\frac{B_i}{2\Delta_i}+\mathcal{O}\left(\frac{\ln T}{\underline{\Delta}^2}\right).$$

Proof. The proof follows a simple reduction to the setting with arms using LSI. By using LSIBR, any strategic arm i has no budget to manipulate after (includes) time step $n+1$, which is analogous to the case that arm i has no budget to manipulate after time $K+1$ using LSI in unbounded reward setting. Then after time $n+1$, the $\widetilde{\mu}_i(t-1) = \widehat{\mu}_i(t-1) + \frac{B_i}{n_i(t-1)}, \forall \in [K]$, which shares the same formula with it in LSI setting. Finally, we notice that the proofs of the upper bounds of $\mathbb{E}\left[\sum_{t=K+1}^{T} \mathbb{I}\{I_t = i^*\}\right]$ in LSI settings (Lemma B.4, C.6 and Theorem C.7) don't depend on the starting time step in the summand. Therefore, the proofs in these previous results can be directly applied here. □

Next, we prove Theorem 3.5 using the above Lemma.

Proof of Theorem 3.5. Let n be the last time step that any arm can spend the budget. First we show the upper bound of $\mathbb{E}\left[n_{i^*}^{\texttt{LSIBR}}(T)\right]$. Note, from time 1 to $n-1$, any strategic arm i always promote its reward to 1, which makes arm i the "optimal arm" from time 1 to n (the arm selection at time n only depends on previous feedback). Then following the standard analysis in stochastic MAB alogrithms (UCB, ε-Greedy and Thompson Sampling), $\mathbb{E}\left[n_{i^*}^{\texttt{LSIBR}}(n)\right] \leq O\Big(\frac{\ln n}{(1-\mu_{i^*})^2}\Big)$. Thus, $\mathbb{E}\left[n_{i^*}^{\texttt{LSIBR}}(T)\right]$ can be bounded by,

$$\mathbb{E}\left[n_{i^*}^{\texttt{LSIBR}}(T)\right] \leq \mathbb{E}\left[n_{i^*}^{\texttt{LSI}}(T)\right] + \mathcal{O}\left(\frac{\ln n}{(1-\mu_{i^*})^2}\right).$$

Consequently, we can show the lower bound of regret when all strategic arms use LSIBR, as follows

$$\mathbb{E}\left[R^{\texttt{LSIBR}}(T)\right] \geq \mathbb{E}\left[R^{\texttt{LSI}}(T)\right] - \mathcal{O}\left(\frac{\Delta \ln T}{(1-\mu_{i^*})^2}\right).$$

□

C. Omitted Proofs in Section 4

C.1. Proof of Theorem 4.1

To prove this theorem, we instead prove the following Lemma C.1 to bound $\mathbb{E}[n_i(T)]$ for each arm $i \neq i^*$. Given this Lemma, it is then easy to show Theorem 4.1.

Lemma C.1. *Suppose the principal runs the ε-Greedy algorithm with $\varepsilon_t = \min\{1, \frac{cK}{T}\}$ when $t > K$, where the constant $c = \max\left\{20, \frac{36\sigma^2}{\Delta_i^2}\right\}$. Then for any strategic manipulation strategy S, the expected number of times of arm i being pulled up to time T can be bounded by*

$$\mathbb{E}\left[n_i(T)\right] \leq \frac{3B_i}{\Delta_i} + \mathcal{O}\left(\frac{\ln T}{\Delta_i^2}\right).$$

Proof. Let $C_i = \frac{3B_i}{\Delta_i}, x_t = \frac{1}{2K}\sum_{s=K+1}^{t} \epsilon_s$ and for $t \geq \lfloor cK \rfloor + 1$, Given Fact A.3, we have

$$x_t \geq \sum_{s=K+1}^{\lfloor cK \rfloor} \frac{\epsilon_s}{2K} + \sum_{t=\lfloor cK \rfloor+1}^{t} \frac{\epsilon_s}{2K} \geq \lfloor cK \rfloor - K + \frac{c}{2}\sum_{s=\lfloor cK \rfloor+1}^{t} \frac{1}{s} \geq \lfloor cK \rfloor - K + \frac{c}{2}\ln\frac{t}{\lfloor cK \rfloor + 1} \tag{13}$$

We do the decomposition for $\mathbb{E}[n_i(T)]$ as follows,

$$\begin{aligned}
\mathbb{E}\left[n_i(T)\right] &\leq 1 + \mathbb{E}\left[\sum_{t=K+1}^{T} \mathbb{I}\{I_t = i, n_i(t-1) \leq C_i\}\right] + \mathbb{E}\left[\sum_{t=K+1}^{T} \mathbb{I}\{I_t = i, n_i(t-1) \geq C_i\}\right] \\
&\leq 1 + C_i + \sum_{t=K+1}^{T} \frac{\epsilon_t}{K} + \mathbb{E}\left[\sum_{t=K+1}^{T} (1-\epsilon_t) \cdot \mathbb{I}\Big\{\widetilde{\mu}_i(t-1) \geq \widehat{\mu}_{i^*,t-1}, n_i(t-1) \geq C_i\Big\}\right] \qquad (14) \\
&\leq 1 + C_i + \sum_{t=K+1}^{T} \frac{\epsilon_t}{K} + \sum_{t=\lfloor cK \rfloor+1}^{T} \mathbb{P}\Big(\widehat{\mu}_i(t-1) + \frac{\beta_{t-1}}{n_i(t-1)} \geq \widehat{\mu}_{i^*}(t-1), n_i(t-1) \geq C_i\Big)
\end{aligned}$$

The last inequality holds because $\epsilon_t = 1$ when $t \leq \lfloor cK \rfloor$ and $1 - \epsilon_t \leq 1, \forall t$. What remains is to bound the last term above. Since $n_i(t-1) \geq C_i, \beta_{t-1} \leq B_i, \forall t \leq T$, this term is always upper bounded by

$$\begin{aligned}\mathbb{P}\left(\widehat{\mu}_i(t-1) + \frac{\beta_{t-1}}{n_i(t-1)} \geq \widehat{\mu}_{i^*}(t-1), n_i(t-1) \geq C_i\right) &\leq \mathbb{P}\left(\widehat{\mu}_i(t-1) + \frac{B_i}{C_i} \geq \widehat{\mu}_{i^*}(t-1)\right) \\ &= \mathbb{P}\left(\widehat{\mu}_i(t-1) + \frac{\Delta_i}{3} \geq \widehat{\mu}_{i^*}(t-1)\right)\end{aligned} \tag{15}$$

By union bound, we have $\mathbb{P}\Big(\widehat{\mu}_i(t-1) + \frac{\Delta_i}{3} \geq \widehat{\mu}_{i^*}(t-1)\Big) \leq \mathbb{P}\Big(\widehat{\mu}_i(t-1) \geq \mu_i + \frac{\Delta_i}{3}\Big) + \mathbb{P}\Big(\widehat{\mu}_{i^*}(t-1) \leq \mu_{i^*} - \frac{\Delta_k}{3}\Big)$. Based on Lemma A.5, we have

$$\mathbb{P}\Big(\widehat{\mu}_i(t-1) + \frac{\Delta_i}{3} \geq \widehat{\mu}_{i^*}(t-1)\Big) \quad \leq 2x_t \cdot e^{-x_t/5} + \tfrac{18\sigma^2}{\Delta_i^2} e^{-\Delta_k^2 \lfloor x_t \rfloor / 18\sigma^2} \tag{16}$$

We observe the fact that $x_t \geq \lfloor cK \rfloor - K + \frac{c}{2} \ln \frac{t}{\lfloor cK \rfloor + 1} > 5$. Given $xe^{-x/5} \leq ye^{-y/5}, \forall x \geq y \geq 5$ and $e^{-x} \leq e^{-y}, \forall x \geq y$, we have

$$\begin{aligned} x_t e^{-x_t/5} \leq \left(\lfloor cK \rfloor - K + \tfrac{c}{2} \ln \tfrac{t}{\lfloor cK \rfloor + 1}\right) e^{-\frac{c}{10} \ln \frac{t}{\lfloor cK \rfloor + 1}} &= \left(\lfloor cK \rfloor - K + \tfrac{c}{2} \ln \tfrac{t}{\lfloor cK \rfloor + 1}\right) \cdot \left(\tfrac{\lfloor cK \rfloor + 1}{t}\right)^{c/10} \\ \tfrac{\sigma^2}{\Delta_i^2} e^{-\Delta_i^2 \lfloor x_{t-1} \rfloor / 18\sigma^2} \leq \tfrac{\sigma^2}{\Delta_i^2} e^{-\Delta_i^2 c \ln \frac{t}{\lfloor cK \rfloor + 1} / 36\sigma^2} &= \tfrac{\sigma^2}{\Delta_i^2} \left(\tfrac{\lfloor cK \rfloor + 1}{t}\right)^{c\Delta_i^2/36\sigma^2} \end{aligned}$$

Combining the above inequalities and Fact A.3, we can bound

$$\begin{aligned} &\sum_{t=\lfloor cK \rfloor + 1}^{T} 2x_t \cdot e^{-x_t/5} + \frac{18\sigma^2}{\Delta_i^2} e^{-\Delta_i^2 \lfloor x_t \rfloor / 18\sigma^2} \\ &\leq \sum_{t=\lfloor cK \rfloor + 1}^{T} \left(2\lfloor cK \rfloor - 2K + c \ln\left(\frac{t}{\lfloor cK \rfloor + 1}\right)\right) \cdot \left(\frac{\lfloor cK \rfloor + 1}{t}\right)^2 + \frac{18\sigma^2}{\Delta_i^2} \frac{\lfloor cK \rfloor + 1}{t} \\ &\leq (\lfloor cK \rfloor - K) \cdot \frac{2(\lfloor cK \rfloor + 1)^2 \pi^2}{3} + \left(c + \frac{18\sigma^2}{\Delta_i^2}\right) \sum_{t=\lfloor cK \rfloor + 1}^{T} \frac{\lfloor cK \rfloor + 1}{t} \\ &\leq (\lfloor cK \rfloor - K) \cdot \frac{2(\lfloor cK \rfloor + 1)^2 \pi^2}{3} + (\lfloor cK \rfloor + 1)\left(c + \frac{18\sigma^2}{\Delta_i^2}\right) \ln \frac{T}{\lfloor cK \rfloor} \end{aligned} \tag{17}$$

The first inequality in the above holds because $c \geq \max\{20, \frac{36\sigma^2}{\Delta_i^2}\}$, and the second inequality is based on the fact that $\ln x < x, \forall x > 1$ and $\sum_{t=1}^{T} \frac{1}{t^2} \leq \frac{\pi^2}{3}$. The last inequality is the implication of Fact A.3. Moreover, utilizing Fact A.3, we bound $\sum_{t=K+1}^{T} \frac{\epsilon_t}{K}$ in the following way,

$$\sum_{t=K+1}^{T} \frac{\epsilon_t}{K} = \sum_{t=K+1}^{\lfloor cK \rfloor} \frac{1}{K} + \sum_{t=\lfloor cK \rfloor + 1}^{T} \frac{\epsilon_t}{K} \leq \frac{\lfloor cK \rfloor - K}{K} + c \ln \frac{T}{\lfloor cK \rfloor}, \tag{18}$$

Combining Equations (14), (15), (16) and (18), we complete the proof. □

C.2. Proof of Lemma 4.3

We bound the terms in the decomposition of $\mathbb{E}[n_i(T)]$ in Eq. (6) using Lemma C.2 – Lemma C.5.

Lemma C.2 (Lemma 2.16 in (Agrawal & Goyal, 2017)). *Let $x_i = \mu_i + \frac{\Delta_i}{3}$ and $y_i = \mu_{i^*} - \frac{\Delta_i}{3}$,*

$$\mathbb{E}\left[\sum_{t=K+1}^{T} \mathbb{I}\{I_t = i, E_i^{\mu}(t), \overline{\mathbb{E}_i^{\theta}(t)}\}\right] \leq \frac{18 \ln T}{\Delta_i^2} + 1$$

Lemma C.3 (Eq. (4) in (Agrawal & Goyal, 2017)). $\sum_{t=K+1}^{T} \mathbb{P}\left(I_t = i, E_i^{\mu}(t), \mathbb{E}_i^{\theta}(t)\right) \leq \sum_{s=K+1}^{T-1} \mathbb{E}\left[\frac{1}{p_{i,\tau_{i^*,s}+1}} - 1\right]$

Lemma C.4 (Extension of Lemma 2.13 in (Agrawal & Goyal, 2017)). *Let* $y_i = \mu_{i^*} - \frac{\Delta_i}{3}$,

$$\mathbb{E}\left[\frac{1}{p_{i,\tau_{i^*},s}+1} - 1\right] \leq \begin{cases} e^{11/4\sigma^2} + \frac{\pi^2}{3} & \forall s \\ \frac{4}{T\Delta_i^2} & \text{if } s \geq \frac{72\ln(T\Delta_i^2)\cdot\max\{1,\sigma^2\}}{\Delta_i^2} \end{cases}$$

Proof. This lemma extends Lemma 2.13 in (Agrawal & Goyal, 2017) to our setting, and we mainly emphasize the required changes to the proof. Using the same notation as in (Agrawal & Goyal, 2017), let Θ_j denote the Gaussian random variable follows $\mathcal{N}(\widehat{\mu}_{i^*}(\tau_j+1), \frac{1}{j})$, given $\mathcal{F}_{\tau_j}$. Let G_j be the geometric random variable denoting the number of consecutive independent trials until a sample of Θ_j becomes greater than y_i. Let $\gamma \geq 1$ be an integer and $z = 2\sigma\sqrt{\ln\gamma}$. Then we have $\mathbb{E}\left[\frac{1}{p_{i,\tau_j+1}} - 1\right] = \mathbb{E}[G_j]$. Following the same argument proposed in (Agrawal & Goyal, 2017), we have for any $\gamma > e^{11/4\sigma^2}$,

$$\mathbb{P}(G_j < \gamma) \geq \left(1 - \frac{1}{\gamma^2}\right)\mathbb{P}\left(\widehat{\mu}_{i^*} + \frac{z}{\sqrt{j}} \geq y_i\right)$$

For $n_{i^*}(t-1) = j$, $\mathcal{F}_{\tau_j}$, we have

$$\begin{aligned} \mathbb{P}\left(\widehat{\mu}_{i^*}(\tau_j+1) + \frac{z}{\sqrt{j}} \geq y_i\right) &\geq \mathbb{P}\left(\widehat{\mu}_{i^*}(\tau_j+1) + \frac{z}{\sqrt{j}} \geq \mu_{i^*}\right) \\ &\geq 1 - e^{-\frac{z^2}{2\sigma^2}} \\ &= 1 - e^{-4\sigma^2\ln\gamma/2\sigma^2} = 1 - \left(\frac{1}{\gamma}\right)^2 \end{aligned}$$

Then $\mathbb{P}(G_j < \gamma) \geq 1 - \frac{1}{\gamma^2} - \frac{1}{\gamma^2} = 1 - \frac{2}{t^2}$. Therefore,

$$\mathbb{E}[G_j] = \sum_{\gamma=0}^{T}\mathbb{P}(G_j \geq \gamma) \leq e^{11/4\sigma^2} + \sum_{\gamma\geq 1}\frac{2}{t^2} \leq e^{11/4\sigma^2} + \frac{\pi^2}{3}$$

By the proof of Lemma 2.13 in (Agrawal & Goyal, 2017), we have for any $D_i(T) \geq 0$,

$$\mathbb{E}\left[\frac{1}{p_{i,\tau_j+1}} - 1\right] \leq \frac{1}{\left(1 - \frac{1}{2}e^{-D_i(T)\Delta_i^2/72}\right)\left(1 - e^{-D_i(T)\Delta_i^2/72\sigma^2}\right)}$$

Since $D_i(T) = \frac{72\ln(T\Delta_i^2)\cdot\max\{1,\sigma^2\}}{\Delta_i^2}$, we have both $1 - \frac{1}{2}e^{-D_i(T)\Delta_i^2/72}$ and $1 - e^{-D_i(T)\Delta_i^2/72\sigma^2}$ are larger than or equal to $1 - \frac{1}{T\Delta_i^2}$. Thus, $\mathbb{E}\left[\frac{1}{p_{i,\tau_j+1}} - 1\right]$ can be bounded by $\frac{4}{T\Delta_i^2}$ when $j \geq D_{i,T}$. □

Lemma C.5.

$$\mathbb{E}\left[\sum_{t=K+1}^{T}\mathbb{I}\left\{I_t = i, \overline{E_i^{\mu}(t)}\right\}\right] \leq \max\left\{\frac{6B_i}{\Delta_i}, \frac{144\sigma^2\ln T}{\Delta_i^2}\right\} + 1 \tag{19}$$

Proof. Let $C_i(T) = \max\left\{\frac{6B_i}{\Delta_i}, \frac{144\sigma^2\ln T}{\Delta_i^2}\right\}$. We first decompose the left hand side in Equation (19) as below,

$$\begin{aligned} \mathbb{E}\left[\sum_{t=K+1}^{T}\mathbb{I}\left\{I_t = i, \overline{E_i^{\mu}(t)}\right\}\right] \leq\ & \mathbb{E}\left[\sum_{t=K+1}^{T}\mathbb{I}\left\{I_t = i, \overline{E_i^{\mu}(t)}, n_i(t-1) \leq C_i(T)\right\}\right] \\ &+ \mathbb{E}\left[\sum_{t=K+1}^{T}\mathbb{I}\left\{I_t = i, \overline{E_i^{\mu}(t)}, n_i(t-1) \geq C_i(T)\right\}\right] \end{aligned} \tag{20}$$

The first term in the above decomposition is trivially bounded by $c_i(T)$. What remains is to bound second term

$$\begin{aligned}
&\mathbb{E}\left[\sum_{t=K+1}^{T}\mathbb{I}\{I_t=i,\overline{E_{i,t}^{\mu}},n_i(t-1)\geq c_i(T)\}\right]\\
&\leq\sum_{t=K+1}^{T}\mathbb{P}\left(\overline{E_{i,t}^{\mu}},n_i(t-1)\geq C_i(T)\right)\\
&\leq\sum_{t=K+1}^{T}\mathbb{P}\left(\widehat{\mu}_{i,t-1}+\frac{\beta_{t-1}}{n_i(t-1)}\geq x_i\middle|n_i(t-1)\geq C_i(T)\right)\\
&\leq\sum_{t=K+1}^{T}\mathbb{P}\left(\widehat{\mu}_{i,t-1}+\frac{\beta_{t-1}}{n_i(t-1)}\geq x_i\middle|n_i(t-1)\geq C_i(T)\right)
\end{aligned}$$

By union bound, we have

$$\begin{aligned}
&\mathbb{P}\left(\widehat{\mu}_{i,t-1}+\frac{\beta_{t-1}}{n_i(t-1)}\geq x_i\middle|n_i(t-1)\geq C_i(T)\right)\\
&\leq\sum_{s=c_i(T)}^{t-1}\mathbb{P}\left(\widehat{\mu}_{i,t-1}+\frac{B_i}{n_i(t-1)}\geq x_i\middle|n_i(t-1)=s\right)\\
&\leq\sum_{s=c_i(T)}^{t-1}e^{-\frac{s\cdot\left(x_i-\mu_i-\frac{B_i}{s}\right)^2}{2\sigma^2}}\leq\sum_{s=1}^{t-1}\frac{1}{T^2}
\end{aligned}$$

The last inequality above uses Fact (A.2) and the fact $s\geq c_i(T)\geq\frac{6B_i}{\Delta_i}$ and $s\geq\frac{144\sigma^2\ln T}{\Delta_i^2}$. Then the second term of the right hand side in Equations 20 can be bounded by $\sum_{t=K+1}^{T}\sum_{s=1}^{t-1}\frac{1}{T^2}\leq 1$. □

C.3. Proof of Proposition 4.4

We complete the proofs for ε-Greedy principal and Thompson Sampling separately. Similar to UCB principal, we derive the upper bound of $\mathbb{E}[n_{i^*}(T)]$ when all strategic arms use `LSI` manipulation strategy, shown in Lemma C.6 (for ε-Greedy principal) and Theorem C.7 (for Thompson Sampling). Then Proposition 4.4 is straightforward.

ϵ-Greedy principal.

Lemma C.6. *$\forall t>K$, let $\epsilon_t=\min\{1,\frac{cK}{t}\}$, where a constant $c=\max\left\{20,\frac{16\sigma^2}{\Delta_k^2},\forall k\in[K]\right\}$, B_i be the total budget for strategic arm. The expected number of plays of arm i^* up to time T, if all strategic arms use `LSI`, is bounded by*

$$\mathbb{E}\left[n_{i^*}(T)\right]\leq T-\sum_{i\neq i^*}\frac{B_i}{2\Delta_i}+\mathcal{O}\left(\frac{\ln T}{\underline{\Delta}^2}\right)$$

Proof. Let $C_i=\frac{B_i}{2\Delta_i}$, $x_t=\frac{1}{2K}\sum_{s=K+1}^{t}\epsilon_s$ and for $t\geq\lfloor cK\rfloor+1$, by Equation (13) $x_t\geq\lfloor cK\rfloor-K+\frac{c}{2}\ln\frac{t}{\lfloor cK\rfloor+1}$.

We first bound the probability of $\mathbb{P}\left(\widehat{\mu}_{i^*}(t-1) \geq \widetilde{\mu}_i(t-1) \middle| n_i(t-1) \leq C_i\right)$ for $t \geq K+1$,

$$\begin{aligned}
& \mathbb{P}\Big(\widehat{\mu}_{i^*}(t-1) \geq \widetilde{\mu}_i(t-1), n_i(t-1) \leq C_i\Big) \\
= \; & \mathbb{P}\Big(\widehat{\mu}_{i^*}(t-1) \geq \widehat{\mu}_i(t-1) + \frac{B_i}{n_i(t-1)}, n_i(t-1) \leq C_i\Big) \\
\leq \; & \mathbb{P}\Big(\widehat{\mu}_{i^*}(t-1) \geq \widehat{\mu}_i(t-1) + 2\Delta_i\Big) \\
\leq \; & \mathbb{P}\Big(\widehat{\mu}_{i^*}(t-1) \geq \mu_{i^*} + \frac{\Delta_i}{2}\Big) + \mathbb{P}\Big(\widehat{\mu}_i(t-1) \leq \mu_i - \frac{\Delta_i}{2}\Big) \\
\leq \; & 2x_t \cdot e^{-x_t/5} + \frac{8\sigma^2}{\Delta_i^2} e^{-\Delta_i^2 \lfloor x_t \rfloor / 8\sigma^2} \text{ (By Lemma A.5)}
\end{aligned} \tag{21}$$

We can decompose the expected number of plays of the optimal arm i, $\mathbb{E}[n_{i^*,T}]$, as follows,

$$\begin{aligned}
\mathbb{E}\left[n_{i^*}(T)\right] = 1 + \mathbb{E}\left[\sum_{t=K+1}^{T} \mathbb{I}\{I_t = i^*, \forall i \neq i^* n_i(t-1) \geq C_i\}\right] \\
+ \mathbb{E}\left[\sum_{t=K+1}^{T} \mathbb{I}\{I_t = i^*, \exists i \neq i^*, n_i(t-1) \leq C_i\}\right]
\end{aligned} \tag{22}$$

The first term in the above decomposition can be bounded by $T - \sum_{i \neq i^*} C_i$. This is because

$$\begin{aligned}
& \mathbb{E}\left[\sum_{t=K+1}^{T} \mathbb{I}\big\{I_t = i^*, \forall i \neq i^*, n_i(t-1) \geq C_i\big\}\right] \\
& \qquad \leq \mathbb{E}\left[\sum_{t=K+1}^{T} \mathbb{I}\big\{I_t = i^*, n_{i^*}(t-1) \leq T - \sum_{i \neq i^*} C_i\big\}\right] \leq T - \sum_{i \neq i^*} C_i.
\end{aligned}$$

By union bound, the second term is bounded by $\sum_{i \neq i^*} \mathbb{E}\left[\sum_{t=K+1}^{T} \mathbb{I}\{I_t = i^*, n_i(t-1) \leq C_i\}\right]$. Then, we bound the above summand using Equations (21) and the fact that $1 - \epsilon_t = 0$ when $t \leq \lfloor cK \rfloor$,

$$\begin{aligned}
& \mathbb{E}\left[\sum_{t=K+1}^{T} \mathbb{I}\{I_t = i^*, n_i(t-1) \leq C_i\}\right] \\
\leq & \sum_{t=K+1}^{T} \frac{\epsilon_t}{K} + \sum_{t=K+1}^{T} (1-\epsilon_t) \cdot \mathbb{P}\Big(\widehat{\mu}_{i^*}(t-1) \geq \widetilde{\mu}_i(t-1), n_i(t-1) \leq C_i\Big) \\
\leq & \sum_{t=K+1}^{T} \frac{\epsilon_t}{K} + \sum_{t=\lfloor cK \rfloor + 1}^{T} 2x_t \cdot e^{-x_t/5} + \frac{8\sigma^2}{\Delta_i^2} e^{-\Delta_i^2 \lfloor x_t \rfloor / 8\sigma^2}
\end{aligned} \tag{23}$$

What remains is to bound the last term in the above equations. Following the same arguments and proof procedure in Equations (17), we can bound

$$\begin{aligned}
& \sum_{t=\lfloor cK \rfloor + 1}^{T} 2x_t \cdot e^{-x_t/5} + \frac{8\sigma^2}{\Delta_i^2} e^{-\Delta_i^2 \lfloor x_t \rfloor / 8\sigma^2} \\
& \leq (\lfloor cK \rfloor - K) \cdot \frac{2(\lfloor cK \rfloor + 1)^2 \pi^2}{3} + (\lfloor cK \rfloor + 1)\left(c + \frac{8\sigma^2}{\Delta_i^2}\right) \ln \frac{T}{\lfloor cK \rfloor}
\end{aligned} \tag{24}$$

By Eq. (18), we have

$$\begin{aligned}\mathbb{E}[n_{i^*}(T)] &\leq T - \sum_{i\neq i^*}\frac{B_i}{2\Delta_i} + \frac{\lfloor cK\rfloor}{K} + c\ln\frac{T}{\lfloor cK\rfloor} \\ &+ \sum_{i\neq i^*}\left((\lfloor cK\rfloor - K)\cdot\frac{2(\lfloor cK\rfloor+1)^2\pi^2}{3} + (\lfloor cK\rfloor+1)\left(c+\frac{8\sigma^2}{\Delta_i^2}\right)\ln\frac{T}{\lfloor cK\rfloor}\right) \\ &\leq T - \sum_{i\neq i^*}\frac{B_i}{2\Delta_i} + \mathcal{O}\left(\frac{\ln T}{\underline{\Delta}^2}\right)\end{aligned}$$

□

Thompson Sampling principal. Here we slightly abuse notations, and use $E_{i^*}^{\mu}(t)$ to denote the event that $\widehat{\mu}_{i^*}(t-1)\leq v_i$ whereas $E_{i^*}^{\theta}(t)$ to denote the event that $\theta_{i^*}(t)\leq w_i$, where $\mu_{i^*}<v_i<w_i$.

Theorem C.7.

$$\mathbb{E}[n_{i^*}(T)] \leq T - \sum_{i\neq i^*}\frac{B_i}{2\Delta_i} + \mathcal{O}\left(\frac{\ln T}{\underline{\Delta}^2}\right)$$

Proof. We decompose the expected number of plays of the optimal arm i^* as follows,

$$\begin{aligned}\mathbb{E}[n_{i^*}(T)] \leq 1 &+ \sum_{t=K+1}^{T}\mathbb{P}\left(I_t=i^*,\overline{E_{i^*}^{\mu}(t)}\right) + \sum_{t=K+1}^{T}\mathbb{P}\left(I_t=i^*,\overline{E_{i^*}^{\theta}(t)},E_{i^*}^{\mu}(t)\right) \\ &+ \sum_{t=K+1}^{T}\mathbb{P}\left(I_t=i^*,E_{i^*}^{\theta}(t),E_{i^*}^{\mu}(t)\right)\end{aligned}$$

Then we bound each of the above terms. Lemma C.8, C.9 and C.12 show the upper bound of each term and complete the proof. □

Lemma C.8. *Let* $v_i=\mu_{i^*}+\frac{\Delta_i}{3}$,

$$\sum_{t=K+1}^{T}\mathbb{P}\left(I_t=i^*,\overline{E_{i^*}^{\mu}(t)}\right)\leq\frac{18\sigma^2}{\Delta_i^2}$$

Proof. Following the proof of Lemma 2.11 in (Agrawal & Goyal, 2017), we have

$$\begin{aligned}\sum_{t=K+1}^{T}\mathbb{P}\left(I_t=i^*,\overline{E_{i^*}^{\mu}(t)}\right) &\leq \sum_{s=1}^{T-1}\mathbb{P}\left(\overline{E_{i^*}^{\mu}(\tau_{i^*,s+1})}\right) = \sum_{s=1}^{T-1}\mathbb{P}\left(\widehat{\mu}_{i^*}(\tau_{i^*,s+1})>v_i\right) \\ &\leq \sum_{s=1}^{T-1}\exp\left(-\frac{s(v_i-\mu_{i^*})^2}{2\sigma^2}\right)\leq\frac{2\sigma^2}{(v_i-\mu_{i^*})^2}\end{aligned}$$

The first inequality holds because each summand on the right hand side in this inequality is a fixed number since the distribution of $\widehat{\mu}_{i^*}(\tau_{i^*,s+1})$ only depends on s. The second inequality is based on Fact A.4 and the third inequality goes through because $\sum_{k=1}^{\infty}e^{-kx}\leq\frac{1}{x},\forall x>0$. □

Notice that Lemma C.3 holds *independently with* the identity of the arm. Then the following Lemma can be directly implied.

Lemma C.9. *Let* $v_i=\mu_{i^*}+\frac{\Delta_i}{3}$ *and* $w_i=\mu_{i^*}+\frac{2\Delta_i}{3}$

$$\sum_{t=K+1}^{T}\mathbb{P}\left(I_t=i^*,\overline{E_{i^*}^{\theta}(t)},E_{i^*}^{\mu}(t)\right)\leq\frac{18\ln T}{\Delta_i^2}+1$$

Proof. The proof of Lemma 2.16 in (Agrawal & Goyal, 2017) can be directly applied here by regarding arm i^* as a standard sub-optimal arm i. □

What remains is to bound $\sum_{t=K+1}^{T}\mathbb{P}\left(I_t=i^*,E_{i^*}^{\theta}(t),E_{i^*}^{\mu}(t)\right)$. To this end, we show some auxiliary lemmas in the following. Lemma C.10 mimics Lemma 2.8 in (Agrawal & Goyal, 2017), which bridges the probability that arm i^* will be pulled and the probability that arm i will be pulled at time t. Lemma C.11 bounds the term $\mathbb{E}\left[\frac{1}{q_{i,\tau_{i,s}+1}}-1\right]$ by a reduction to the case shown in Lemma C.3.

Lemma C.10. *For any instantiation F_{t-1} of $\mathcal{F}_{t-1}$, let $q_{i,t}:=\mathbb{P}\left(\theta_i(t)>w_i\middle|F_{t-1}\right)$, we have*

$$\mathbb{P}\left(I_t=i^*,E_{i^*}^{\theta}(t),E_{i^*}^{\mu}(t)\middle|F_{t-1}\right)\le\frac{1-q_{i,t}}{q_{i,t}}\mathbb{P}\left(I_t=i,E_{i^*}^{\theta}(t),E_{i^*}^{\mu}(t)\middle|F_{t-1}\right)$$

Proof. Since $E_{i^*}^{\mu}(t)$ is only determined by the instantiation F_{t-1} of $\mathcal{F}_{t-1}$, we can assume event $E_{i^*}^{\mu}(t)$ is true without loss of generality. Then, it is sufficient to show that for any F_{t-1} we have

$$\mathbb{P}\left(I_t=i^*\middle|E_{i^*}^{\theta}(t),F_{t-1}\right)\le\frac{1-q_{i,t}}{q_{i,t}}\mathbb{P}\left(I_t=i,\middle|E_{i^*}^{\theta}(t),F_{t-1}\right)$$

Note, given $E_{i^*}^{\theta}(t)$, $I_t=i^*$ implies $\theta_j(t)\le w_i,\forall j$, meanwhile, $\theta_i(t)$ is independent with $\theta_j(t),j\neq i$, given $\mathcal{F}_{t-1}=F_{t-1}$. Therefore, we have

$$\begin{aligned}\mathbb{P}\left(I_t=i^*\middle|E_{i^*}^{\theta}(t),F_{t-1}\right)&\le\mathbb{P}\left(\theta_j(t)\le w_i,\forall j\middle|E_{i^*}^{\theta}(t),F_{t-1}\right)\\&=\mathbb{P}\left(\theta_i(t)\le w_i\middle|F_{t-1}\right)\cdot\mathbb{P}\left(\theta_j(t)\le w_i,\forall j\neq i\middle|E_{i^*}^{\theta}(t),F_{t-1}\right)\end{aligned}$$

On the other side,

$$\begin{aligned}\mathbb{P}\left(I_t=i\middle|E_{i^*}^{\theta}(t),F_{t-1}\right)&\ge\mathbb{P}\left(\theta_i(t)>w_i\ge\theta_j(t),\forall j\neq i\middle|E_{i^*}^{\theta}(t),F_{t-1}\right)\\&=\mathbb{P}\left(\theta_i(t)>w_i\middle|F_{t-1}\right)\cdot\mathbb{P}\left(\theta_j(t)\le w_i,\forall j\neq i\middle|E_{i^*}^{\theta}(t),F_{t-1}\right)\end{aligned}$$

Thus, the above two inequalities implies the correctness of the Lemma. □

Lemma C.11. *Let $w_i=\mu_{i^*}+\frac{2\Delta_i}{3}$. For any $s\ge1$, given $n_i(\tau_{i,s})\le\frac{B_i}{2\Delta_i}$, we have*

$$\mathbb{E}\left[\frac{1}{q_{i,\tau_{i,s}+1}}-1\middle|n_i(\tau_{i,s})\le\frac{B_i}{2\Delta_i}\right]\le\begin{cases}e^{11/4\sigma^2}+\frac{\pi^2}{3} & \forall s\\ \frac{1}{T\Delta_i} & \text{if } s\ge L_i(T)\end{cases}$$

where $L_i(T)=\frac{72\ln(T\Delta_i^2)\cdot\max\{1,\sigma^2\}}{\Delta_i^2}$.

Proof. We prove this Lemma by a reduction to Lemma C.4. First, we observe $\theta_i(\tau_{i,s}+1)\sim\mathcal{N}\left(\widetilde{\mu}_i(\tau_{i,s}),\frac{1}{n_i(\tau_{i,s})}\right)$, where $\widetilde{\mu}_i(\tau_{i,s})=\widehat{\mu}_i(\tau_{i,s})+\frac{B_i}{n_i(\tau_{i,s})}$. Given $n_i(\tau_{i,s})\le\frac{B_i}{\Delta_i}$, we have $\widetilde{\mu}_i(\tau_{i,s})\ge\widehat{\mu}_i(\tau_{i,s})+2\Delta_i$. Let $\zeta_i(\tau_{i,s}+1)$ denote the random variable of Gaussian distribution $\mathcal{N}\left(\widehat{\mu}_i(\tau_{i,s}),\frac{1}{n_i(\tau_{i,s})}\right)$. By the fact that a Gaussian random variable $a\sim\mathcal{N}(m,\sigma^2)$ is stochastically dominated by any $b\sim\mathcal{N}(m',\sigma^2)$ when $m<m'$, we have for any F_{t-1} of $\mathcal{F}_{t-1}$

$$\begin{aligned}q_{i,\tau_{i,s}+1}&=\mathbb{P}\left(\theta_i(\tau_{i,s}+1)>w_i\middle|F_{t-1}\right)\ge\mathbb{P}\left(\zeta_i(\tau_{i,s}+1)+2\Delta_i>w_i\middle|F_{t-1}\right)\\&=\mathbb{P}\left(\zeta_i(\tau_{i,s}+1)>\mu_i-\frac{\Delta_i}{3}\middle|F_{t-1}\right):=\eta_{i,\tau_{i,s}+1}\end{aligned}$$

Therefore, $\mathbb{E}\left[\frac{1}{q_{i,\tau_{i,s}+1}}-1\right]\le\mathbb{E}\left[\frac{1}{\eta_{i,\tau_{i,s}+1}}-1\right]$. Denote $u_i:=\mu_i-\frac{\Delta_i}{3}$. Recall

$$p_{i,\tau_{i,s}+1}=\mathbb{P}\left(\theta_{i^*}(\tau_{i^*,s}+1)>\mu_{i^*}-\frac{\Delta_i}{3}\middle|F_{t-1}\right),$$

we observe $\eta_{i,\tau_{i,s}+1}$ is analogous to $p_{i,\tau_{i,s}+1}$ in formula, when we replace μ_i and $\widehat{\mu}_i(\tau_{i,s}+1)$ by μ_{i^*} and $\widehat{\mu}_{i^*}(\tau_{i^*,s}+1)$ respectively (i.e. change arm i by i^*). Recall the proof in Lemma C.3, it only depends on the relationship between $y_i = \mu_{i^*} - \frac{\Delta_i}{3}$ and μ_{i^*}, which is the same as u_i and μ_i in $\eta_{i,\tau_{i,s}+1}$. Thus, the proof of Lemma C.3 can be directly applied here to bound $\mathbb{E}\left[\frac{1}{\eta_{i,\tau_{i,s}+1}} - 1\right]$. □

Lemma C.12.

$$\begin{aligned}
&\sum_{t=K+1}^{T} \mathbb{P}\left(I_t = i^*, E^{\theta}_{i^*}(t), E^{\mu}_{i^*}(t)\right) \\
\leq\ & T - \sum_{i\neq i^*} \frac{B_i}{2\Delta_i} + \sum_{i\neq i^*} \left((e^{11/4\sigma^2} + \frac{\pi^2}{3}) \cdot \frac{72\ln(T\Delta_i^2)\cdot\max\{1,\sigma^2\}}{\Delta_i^2} + \frac{4}{\Delta_i^2} \right)
\end{aligned}$$

Proof. We first decompose the target term by thresholding $n_i(t-1)$ as follows,

$$\begin{aligned}
&\sum_{t=K+1}^{T} \mathbb{P}\left(I_t = i^*, E^{\theta}_{i^*}(t), E^{\mu}_{i^*}(t)\right) \\
\leq\ & \mathbb{E}\left[\sum_{t=K+1}^{T} \mathbb{I}\Big\{I_t = i^*, E^{\theta}_{i^*}(t), E^{\mu}_{i^*}(t), \forall i\neq i^*, n_i(t-1) \geq \frac{B_i}{2\Delta_i}\Big\}\right] \\
&+ \sum_{t=K+1}^{T} \mathbb{P}\left(I_t = i^*, E^{\theta}_{i^*}(t), E^{\mu}_{i^*}(t), \exists i\neq i^*, n_i(t-1) \leq \frac{B_i}{2\Delta_i}\right)
\end{aligned} \tag{25}$$

For the first term in above decomposition, it can be trivially upper bounded by $T - \sum_{i\neq i^*} \frac{B_i}{2\Delta_i}$. By union bound and Lemma C.10, we can bound the second term as follows,

$$\begin{aligned}
&\sum_{t=K+1}^{T} \mathbb{P}\left(I_t = i^*, E^{\theta}_{i^*}(t), E^{\mu}_{i^*}(t), \exists i\neq i^*, n_i(t-1) \leq \frac{B_i}{2\Delta_i}\right) \\
&\leq \sum_{i\neq i^*}\sum_{t=K+1}^{T} \mathbb{P}\left(I_t = i^*, E^{\theta}_{i^*}(t), E^{\mu}_{i^*}(t), \exists i\neq i^*, n_i(t-1) \leq \frac{B_i}{2\Delta_i}\right) \\
&= \sum_{i\neq i^*}\sum_{t=K+1}^{T} \mathbb{E}\left[\mathbb{P}\left(I_t = i^*, E^{\theta}_{i^*}(t), E^{\mu}_{i^*}(t), n_i(t-1) \leq \frac{B_i}{2\Delta_i}\Big|\mathcal{F}_{t-1}\right)\right] \\
&= \sum_{i\neq i^*}\sum_{t=K+1}^{T} \mathbb{E}\left[\frac{1-q_{i,t}}{q_{i,t}}\cdot\mathbb{P}\left(I_t = i, E^{\theta}_{i^*}(t), E^{\mu}_{i^*}(t), n_i(t-1) \leq \frac{B_i}{2\Delta_i}\Big|\mathcal{F}_{t-1}\right)\right] \\
&\leq \sum_{i\neq i^*}\sum_{t=K+1}^{T} \mathbb{E}\left[\frac{1-q_{i,t}}{q_{i,t}}\cdot\mathbb{P}\left(I_t = i, E^{\theta}_{i^*}(t), E^{\mu}_{i^*}(t)\Big|n_i(t-1) \leq \frac{B_i}{2\Delta_i}, \mathcal{F}_{t-1}\right)\right] \\
&= \sum_{i\neq i^*}\sum_{t=K+1}^{T} \mathbb{E}\left[\frac{1-q_{i,t}}{q_{i,t}}\cdot\mathbb{I}\Big\{I_t = i, E^{\theta}_{i^*}(t), E^{\mu}_{i^*}(t)\Big\}\Big|n_i(t-1) \leq \frac{B_i}{2\Delta_i}\right]
\end{aligned}$$

Observe that $q_{i,t} = \mathbb{P}\left(\theta_i(t) > w_i\Big|\mathcal{F}_{t-1}\right)$ changes only at the time step after each pull of arm i. Therefore we can bound

the above term by,

$$\sum_{s=1}^{T-1}\mathbb{E}\left[\frac{1-q_{i,\tau_{i,s}+1}}{q_{i,\tau_{i,s}+1}}\cdot\sum_{t=\tau_{i,s}+1}^{\tau_{i,s+1}}\mathbb{I}\{I_t=i,E_{i^*}^{\theta}(t),E_{i^*}^{\mu}(t)\}\Big|n_i(\tau_{i,s})\le\frac{B_i}{2\Delta_i}\right]$$
$$\le\quad\sum_{s=1}^{T-1}\mathbb{E}\left[\frac{1-q_{i,\tau_{i,s}+1}}{q_{i,\tau_{i,s}+1}}\Big|n_i(\tau_{i,s})\le\frac{B_i}{2\Delta_i}\right]$$

Combining Lemma C.11 and Equation (25), we complete the proof. □

D. Additional Simulations

We report our simulation results for bounded rewards in this section. Similarly, we also consider a stochastic bandit setting with three arms. The reward of each arm lies within the interval $[0,1]$. The distributions of rewards of each arm are $\texttt{Beta}(1,1)$, $\texttt{Beta}(2,1)$ and $\texttt{Beta}(3,1)$ respectively. In ε-Greedy algorithm, we use a different ε_t parameter, i.e. $\varepsilon_t=\min\{1,\frac{20}{t}\}$. We run simulations for the same settings as those in Section 5 and report the results in Figure 3 and 4. These figures illustrate similar performances for bounded rewards as for unbounded rewards.

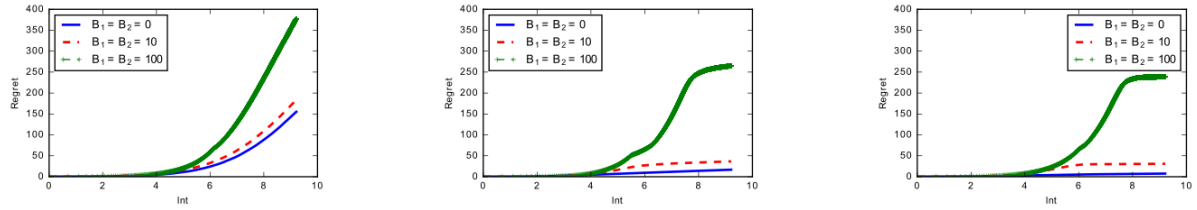

Figure 3: $[0,1]$ bounded rewards: plots of regret with $\ln t$ for UCB principal (left), ε-Greedy principal (middle), and Thompson Sampling principal (right), as B_1 and B_2 vary. We set $B_3=0$ for the three algorithms.

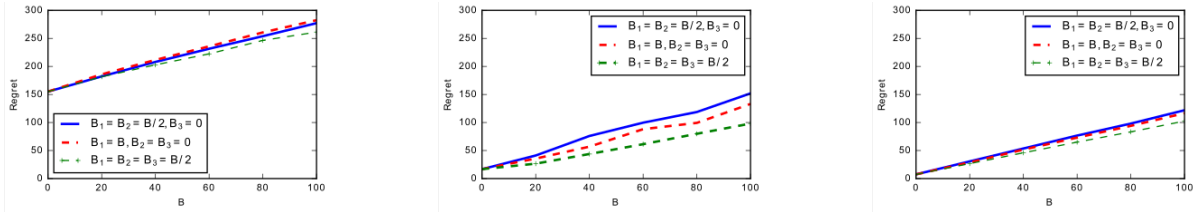

Figure 4: $[0,1]$ bounded rewards: plots of regret with total budget B of strategic arms (arm 1 and 2) for UCB principal (left), ε-Greedy principal (middle), and Thompson Sampling principal (right), as B_i varies.