SUPPLEMENTARY MATERIAL

A. Proofs

A.1. Proof of Theorem 2

The proof of Theorem 2 is closely related to the proof of Proposition 1 in Shin et al. (2019a). However, in this proof, we decouple the bias part from the integrability condition which makes the proof significantly simpler than the one in Shin et al. (2019a).

Under the condition in Theorem 2, we first prove that if the function $\mathcal{D}^*_\infty \mapsto \mathbb{1}(C)/N_k(\mathcal{T})$ is an decreasing function of $X^*_{i,k}$ while keeping all other entries in $\mathcal{D}^*_\infty$ fixed for each i then, for any $t \in \mathbb{N}$ and $y \in \mathbb{R}$, the following inequality holds.

$$\mathbb{E}\left[\frac{\mathbb{1}(C)}{N_k(\mathcal{T})}\mathbb{1}(N_k(\mathcal{T}) < \infty)\mathbb{1}(A_t = k)\left[\mathbb{1}(Y_t \leq y) - F_k(y)\right]\right] \geq 0 \tag{32}$$

Proof of inequality (32). First note that if $\mathcal{D}^*_\infty \mapsto \mathbb{1}(C)/N_k(\mathcal{T})$ is a decreasing function of $X^*_{i,k}$ then the following function is also a decreasing function of $X^*_{i,k}$:

$$D^*_\infty \mapsto \frac{\mathbb{1}(C)}{N_k(\mathcal{T})}\mathbb{1}(N_k(\mathcal{T}) < \infty) := h(\mathcal{D}^*_\infty) \tag{33}$$

Then, by the tabular representation of MAB, we can rewrite the LHS of inequality (32) as follows:

$$\begin{aligned}
&\mathbb{E}\left[\frac{\mathbb{1}(C)}{N_k(\mathcal{T})}\mathbb{1}(N_k(\mathcal{T}) < \infty)\mathbb{1}(A_t = k)\left[\mathbb{1}(Y_t \leq y) - F_k(y)\right]\right] \\
&= \mathbb{E}\left[h(\mathcal{D}^*_\infty)\mathbb{1}(A_t = k)\left[\mathbb{1}(X^*_{N_k(t),k} \leq y) - F_k(y)\right]\right] \\
&= \mathbb{E}\sum_{i=1}^{t}\left[h(\mathcal{D}^*_\infty)\mathbb{1}(A_t = k, N_k(t) = i)\left[\mathbb{1}(X^*_{i,k} \leq y) - F_k(y)\right]\right] \\
&= \sum_{i=1}^{t}\mathbb{E}\left[h(\mathcal{D}^*_\infty)\mathbb{1}(A_t = k, N_k(t) = i)\left[\mathbb{1}(X^*_{i,k} \leq y) - F_k(y)\right]\right],
\end{aligned}$$

where the third equality comes from the fact $N_k(t) \in \{1, \ldots, t\}$. Therefore, to prove the inequality (32), it is enough to show the following inequality:

$$\mathbb{E}\left[\mathbb{1}(A_t = k, N_k(t) = i)\, h(\mathcal{D}^*_\infty)\left[\mathbb{1}(X^*_{i,k} \leq y) - F_k(y)\right]\right] \geq 0. \tag{34}$$

Note that the term $\mathbb{1}(A_t = k, N_k(t) = i)$ does not depend on $X^*_{i,k}$ by the definition of A_t and $N_k(t)$.

Now, let $\mathcal{D}^{*'}_\infty$ be an other the tabular representation which is identical to $\mathcal{D}^*_\infty$ except the (i,k)-th entry of X^*_∞ in $\mathcal{D}^*_\infty$ being replaced with an independent copy $X^{*'}_{i,k}$ from the same distribution P_k.

Since the function h is a decreasing function of $X^*_{i,k}$ while keeping all other entries in $\mathcal{D}^*_\infty$ fixed, we have that

$$\left[h(\mathcal{D}^*_\infty) - h(\mathcal{D}^{*'}_\infty)\right]\left[\mathbb{1}(X^*_{i,k} \leq y) - \mathbb{1}(X^{*'}_{i,k} \leq y)\right] \geq 0, \tag{35}$$

which implies that

$$\begin{aligned}
&h(\mathcal{D}^*_\infty)\left[\mathbb{1}(X^*_{i,k} \leq y) - F_k(y)\right] + h(\mathcal{D}^{*'}_\infty)\left[\mathbb{1}(X^{*'}_{i,k} \leq y) - F_k(y)\right] + \\
&\geq h(\mathcal{D}^{*'}_\infty)\left[\mathbb{1}(X^*_{i,k} \leq y) - F_k(y)\right] + h(\mathcal{D}^*_\infty)\left[\mathbb{1}(X^{*'}_{i,k} \leq y) - F_k(y)\right].
\end{aligned} \tag{36}$$

By multiplying $\mathbb{1}(A_t = k, N_k(t) = i)$ and taking expectations on both sides, we can show the inequality (32) hold as follows:

$$2\mathbb{E}\left[\mathbb{1}\left(A_t = k, N_k(t) = i\right) h(\mathcal{D}_\infty^*)\left[\mathbb{1}(X_{i,k}^* \leq y) - F_k(y)\right]\right] \tag{37}$$

$$\geq 2\mathbb{E}\left[\mathbb{1}\left(A_t = k, N_k(t) = i\right) h(\mathcal{D}_\infty^*)\left[\mathbb{1}(X_{i,k}^{*'} \leq y) - F_k(y)\right]\right] \tag{38}$$

$$= 2\mathbb{E}\left[\mathbb{1}\left(A_t = k, N_k(t) = i\right) h(\mathcal{D}_\infty^*)\right]\mathbb{E}\left[\mathbb{1}(X_{i,k}^{*'} \leq y) - F_k(y)\right] \tag{39}$$

$$\geq 0, \tag{40}$$

where the first equality comes from the independence between $\mathbb{1}\left(A_t = k, N_k(t) = i\right) h(\mathcal{D}_\infty^*)$ and $X_{i,k}^{*'}$, and the second inequality holds since $\mathbb{E}\left[\mathbb{1}(X_{i,k}^{*'} \leq y)\right] = F_k(y)$. □

Based on the inequality (32), we are ready to prove Theorem 2.

Proof of Theorem 2. First, suppose the function $\mathcal{D}_\infty^* \mapsto \mathbb{1}\left(C\right)/N_k(\mathcal{T})$ is an decreasing function of $X_{i,k}^*$ while keeping all other entries in $\mathcal{D}_\infty^*$ fixed for each i. Let $\{L_t\}_{t\in\mathbb{N}}$ be a sequence of random variables defined as follows:

$$L_t := \sum_{s=1}^{t} \frac{\mathbb{1}(C)}{N_k(\mathcal{T})}\mathbb{1}(N_k(\mathcal{T}) < \infty)\mathbb{1}\left(A_s = k\right)\left[\mathbb{1}(Y_s \leq y) - F_k(y)\right], \;\; \forall t \in \mathbb{N}. \tag{41}$$

From the inequality (32), we have

$$\mathbb{E}\left[L_t\right] = \sum_{s=1}^{t}\mathbb{E}\left[\frac{\mathbb{1}(C)}{N_k(\mathcal{T})}\mathbb{1}\left(A_s = k\right)\mathbb{1}(N_k(\mathcal{T}) < \infty)\left[\mathbb{1}(Y_s \leq y) - F_k(y)\right]\right] \geq 0, \;\; \forall t \in \mathbb{N}. \tag{42}$$

Note that $N_k(\mathcal{T}) := \sum_{t=1}^{\mathcal{T}} \mathbb{1}(A_t = k) = \sum_{t=1}^{\infty} \mathbb{1}(A_t = k)$ since it is understood that for $t > \mathcal{T}$, $\mathbb{1}(A_t = k) = 0$. Therefore, we know that, for each $y \in \mathbb{R}$, the sequence of random variables $\{L_t\}_{t\in\mathbb{N}}$ converges to $\left[\widehat{F}_{k,\mathcal{T}}(y) - F_k(y)\right]\mathbb{1}(C)\mathbb{1}(N_k(\mathcal{T}) < \infty)$ almost surely. Also, it can be easily checked for each $t \in \mathbb{N}$, $|L_t|$ is upper bounded by 2. Hence, from the dominated convergence theorem and the inequality (42), we have

$$0 \leq \lim_{t\to\infty}\mathbb{E}[L_t] = \mathbb{E}\left[\left[\widehat{F}_{k,\mathcal{T}}(y) - F_k(y)\right]\mathbb{1}(C)\mathbb{1}(N_k(\mathcal{T}) < \infty)\right] \tag{43}$$

$$= \mathbb{E}\left[\widehat{F}_{k,\mathcal{T}}(y)\mathbb{1}(C)\mathbb{1}(N_k(\mathcal{T}) < \infty)\right] - F_k(y)\mathbb{P}(C \cap \{N_k(\mathcal{T}) < \infty\}). \tag{44}$$

Since $\widehat{F}_{k,\mathcal{T}}(y)\mathbb{1}(N_k(\mathcal{T}) = \infty) = F_k(y)\mathbb{1}(N_k(\mathcal{T}) = \infty)$ almost surely, the last inequality also implies that

$$F_k(y)\mathbb{P}(C) \leq \mathbb{E}\left[\widehat{F}_{k,\mathcal{T}}(y)\mathbb{1}(C)\right] \tag{45}$$

Since we assumed $\mathbb{P}(C) > 0$, by multiplying $1/\mathbb{P}(C)$ on both sides, we have

$$F_k(y) \leq \mathbb{E}\left[\widehat{F}_{k,\mathcal{T}}(y) \mid C\right], \tag{46}$$

as desired. The inequality (46) shows that the underlying distribution of arm k stochastically dominates the empirical distribution of arm k in the conditional expectation. In this case, it is well-known that for any non-decreasing integrable function f, the following inequality holds

$$E_k f \geq \mathbb{E}\left[\widehat{E}_{k,\mathcal{T}} f \mid C\right]. \tag{47}$$

For the completeness of the proof, we formally prove the inequality (47). Since f is integrable, without loss of generality, we may assume $f \geq 0$. For any $x \in \mathbb{R}$, define $f^{-1}(x) := \inf\{y : f(y) > x\}$. Since f is non-decreasing, for any probability measure P, the following equality holds

$$P\left(\{y : f(y) > x\}\right) = P\left(\{y : y > f^{-1}(x)\}\right),$$

for all but at most countably many $x \in \mathbb{R}$ which implies that

$$
\begin{aligned}
E_k f &= \int_0^\infty P_k\left(\{y : f(y) > x\}\right) \mathrm{d}x && (48)\\
&= \int_0^\infty 1 - F_k\left(f^{-1}(x)\right) \mathrm{d}x && (49)\\
&\geq \int_0^\infty 1 - \mathbb{E}\left[\widehat{F}_{k,\mathcal{T}}\left(f^{-1}(x)\right) \mid C\right] \mathrm{d}x && (50)\\
&= \int_0^\infty \mathbb{E}\left[\widehat{P}_{k,\mathcal{T}}\left(\{y : f(y) > x\}\right) \mid C\right] \mathrm{d}x && (51)\\
&= \mathbb{E}\left[\widehat{E}_{k,\mathcal{T}} f \mid C\right], && (52)
\end{aligned}
$$

where the first and last equalities come from the Fubini's theorem with the integrability condition on f, and the first inequality comes from the inequality (46).

From the same argument with reversed inequalities, it can be shown that if the function $\mathcal{D}_\infty^* \mapsto \mathbb{1}(C)/N_k(\mathcal{T})$ is an increasing function of $X_{i,k}^*$ while keeping all other entries in $\mathcal{D}_\infty^*$ fixed for each i, we have

$$
F_k(y) \geq \mathbb{E}\left[\widehat{F}_{k,\mathcal{T}}(y) \mid C\right], \quad \forall y \in \mathbb{R}. \tag{53}
$$

Equivalently, for any non-decreasing integrable function f, we have

$$
E_k f \leq \mathbb{E}\left[\widehat{E}_k f \mid C\right], \tag{54}
$$

which completes the proof of Theorem 2. □

A.2. Proof of Corollary 5

Before proving Corollary 5 formally, we first provide an intuition as to why, for any reasonable and efficient algorithm for the best-arm identification problem, the sample mean and empirical CDF of an arm are negative and positive biases, respectively, conditionally on the event that the arm is not chosen as the best arm.

For any $k \in [K]$ and $i \in \mathbb{N}$, let $\mathcal{D}_\infty^*$ and $\mathcal{D}_\infty^{*'}$ be two collections of all possible arm rewards and external randomness that agree with each other except $X_{i,k}^* \geq X_{i,k}^{*'}$. Since we have a smaller reward from arm k in the second scenario $\mathcal{D}_\infty^{*'}$, if $\kappa \neq k$ under the first scenario $\mathcal{D}_\infty^*$, any reasonable algorithm also would not pick the arm k as the best arm under the more unfavorable scenario $\mathcal{D}_\infty^{*'}$. In this case, we know that $\kappa \neq k$ implies $\kappa' \neq k$. Also note that any efficient algorithm should be able to exploit the more unfavorable scenario $D_\infty^{*'}$ to easily identify arm k as a suboptimal arm and choose another arm as the best one by using less samples from arm k. Therefore, we would have $N_k(\mathcal{T}) \geq N_k'(\mathcal{T}')$. As a result, we can expect that, from any reasonable and efficient algorithm, we would have $\frac{\mathbb{1}(\kappa \neq k)}{N_k(\mathcal{T})} \leq \frac{\mathbb{1}(\kappa' \neq k)}{N_k'(\mathcal{T}')}$ which implies that for each i, the function $\mathcal{D}_\infty^* \mapsto \mathbb{1}(C)/N_k(\mathcal{T})$ is a decreasing function of $X_{i,k}^*$ while keeping all other entries in $\mathcal{D}_\infty^*$ fixed. Then, from Theorem 2, we have that the sample mean and empirical CDF of arm k are negatively and positively biased conditionally on the event $\kappa \neq k$, respectively. Below, we formally verify that this intuition works for the lil'UCB algorithm. The proof is based on the following two facts about the lil'UCB algorithm:

- **Fact 1.** The lil'UCB algorithm has an optimistic sampling rule. That is, for any fixed , $i, t \in \mathbb{N}$ and $k \in [K]$, the function $\mathcal{D}_\infty^* \mapsto N_k(t)$ is an increasing function of $X_{i,k}^*$ while keeping all other entries in $\mathcal{D}_\infty^*$ fixed (see Fact 3 in Shin et al., 2019a).
- **Fact 2.** Let $\mathcal{D}_\infty^*$ and $\mathcal{D}_\infty^{*'}$ be two collections of all possible arm rewards and external randomness that agree with each other except in their k-th column of stacks of rewards X_∞^* and $X_\infty^{*'}$. For $j \in [K]$, let $N_j(t)$ and $N_j'(t)$ be the numbers of draws from arm j under $\mathcal{D}_\infty^*$ and $\mathcal{D}_\infty^{*'}$ respectively. Then for each $t \in \mathbb{N}$, the following implications hold for lil'UCB algorithm (see Fact 3 and Lemma 9 in Shin et al., 2019a):

$$
\begin{aligned}
N_k(t) \leq N_k'(t) &\Rightarrow N_j(t) \geq N_j'(t), \quad \text{for all } j \neq k,\\
N_k(t) \geq N_k'(t) &\Rightarrow N_j(t) \leq N_j'(t), \quad \text{for all } j \neq k,
\end{aligned}
$$

which also implies that

$$N_k(t) = N_k'(t) \Rightarrow N_j(t) = N_j'(t), \quad \text{for all } j \neq k.$$

Proof of Corollary 5. For any given $i \in \mathbb{N}$ and $k \in [K]$, let $\mathcal{D}_\infty^*$ and $\mathcal{D}_\infty^{*'}$ be two collections of all possible arm rewards and external randomness that agree with each other except (i,k)-th entries, $X_{i,k}^*$ and $X_{i,k}^{*'}$ of their stacks of rewards. Let $(N_k(t), N_k'(t))$ denote the numbers of draws from arm k up to time t. Let $(\mathcal{T}, \mathcal{T}')$ be the stopping times and (κ, κ') be choosing functions of the lil'UCB algorithm under $\mathcal{D}_\infty^*$ and $\mathcal{D}_\infty^{*'}$ respectively.

Suppose $X_{i,k}^* \geq X_{i,k}^{*'}$. To prove the claimed bias result, it is enough to show that the function $\mathcal{D}_\infty^* \mapsto \mathbb{1}(C)/N_k(\mathcal{T})$ is a decreasing function of $X_{i,k}^*$ while keeping all other entries in $\mathcal{D}_\infty^*$ fixed which corresponds to prove the following inequality holds:

$$\frac{\mathbb{1}(\kappa \neq k)}{N_k(\mathcal{T})} \leq \frac{\mathbb{1}(\kappa' \neq k)}{N_k'(\mathcal{T}')}. \tag{55}$$

Note that if $\kappa = k$ or $N_k(\mathcal{T}) = \infty$, the inequality (55) holds trivially. Therefore, for the rest of the proof, we assume $\kappa \neq k$ and $N_k(\mathcal{T}) < \infty$.

We will first prove the inequality $N_k(\mathcal{T}) \geq N_k'(\mathcal{T}')$ holds. From Fact 1 and the assumption $X_{i,k}^* \geq X_{i,k}^{*'}$, we have $N_k(t) \geq N_k'(t)$ for any fixed $t > 0$. Then, by Fact 2, we also have $N_j(t) \leq N_j'(t)$ for any $j \neq k$. Since $\sum_{i \neq j} N_i(t) = t - N_j(t)$ for all t, we can rewrite the lil'UCB stopping rule as stopping whenever there exists $j \in [K]$ such that the inequality $N_j(t) \geq \frac{1+\lambda t}{1+\lambda}$ holds. Therefore, from the definition of the stopping rule with the fact $N_j(t) \leq N_j'(t)$ for any $t \geq 1$ and $j \neq k$, at the stopping time $\mathcal{T}$, we have

$$\frac{1+\lambda\mathcal{T}}{1+\lambda} \leq N_j(\mathcal{T}) \leq N_j'(\mathcal{T}), \tag{56}$$

for some $j \neq k$ which also implies that the stopping condition is also satisfied for arm j at time $\mathcal{T}$ under $\mathcal{D}_\infty^{*'}$ which implies that the stopping time under $\mathcal{D}_\infty^{*'}$ must be at most $\mathcal{T}$. Therefore we have $\mathcal{T}' \leq \mathcal{T}$. Now, since the inequality $N_k(t) \geq N_k'(t)$ holds for any $t \geq 1$, we have $N_k(\mathcal{T}) \geq N_k'(\mathcal{T})$. Finally, since $t \mapsto N_k'(t)$ is a non-decreasing function, we can conclude $N_k(\mathcal{T}) \geq N_k'(\mathcal{T}) \geq N_k'(\mathcal{T}')$.

Since we proved $N_k(\mathcal{T}) \geq N_k'(\mathcal{T}')$, to complete the proof of Corollary 5, it is enough to show that $\kappa \neq k$ implies $\kappa' \neq k$. We prove this statement by the proof by contradiction. Suppose $\kappa \neq k$ but $\kappa' = k$. Then, there exists $j \neq k$ such that $\kappa = j$. By the definition of $\mathcal{T}$ and κ, we know that

$$N_j(\mathcal{T}) > N_k(\mathcal{T}). \tag{57}$$

Similarly, we can show that

$$N_j'(\mathcal{T}') < N_k'(\mathcal{T}'). \tag{58}$$

It is important to note that these inequalities are strict. Note that since we draw a singe sample at each time, if $N_j(\mathcal{T}) = N_k(\mathcal{T})$ then at the time $\mathcal{T} - 1$, either arm j or k should satisfy the stopping rule which contradicts to the definition of $\mathcal{T}$.

Recall that, in Equation (56), we showed that if $\kappa = j$, at stopping time $\mathcal{T}$, we have

$$\frac{1+\lambda\mathcal{T}}{1+\lambda} \leq N_j(\mathcal{T}) \leq N_j'(\mathcal{T}),$$

which implies $\mathcal{T}' \leq \mathcal{T}$. By the same argument, at the stopping time $\mathcal{T}'$ with the assumption $\kappa' = k$, we have

$$\frac{1+\lambda\mathcal{T}'}{1+\lambda} \leq N_k'(\mathcal{T}') \leq N_k(\mathcal{T}'),$$

which also implies $\mathcal{T} \leq \mathcal{T}'$. From these two inequalities on stopping times, we have $\mathcal{T}' = \mathcal{T}$. Finally, by combining inequalities between pairs of N_k, N_k', N_j, N_j' with the observation $\mathcal{T}' = \mathcal{T}$, we have

$$N_j'(\mathcal{T}') < N_k'(\mathcal{T}') \leq N_k(\mathcal{T}') = N_k(\mathcal{T}) < N_j(\mathcal{T}) \leq N_j'(\mathcal{T}) = N_j'(\mathcal{T}')$$

where the first inequality comes from the inequality (58). The second inequality come from $N_k' \leq N_k$. The first equality comes from $\mathcal{T}' = \mathcal{T}$ and the third inequality comes from the inequality (57). The last inequality comes from $N_j \leq N_j'$ and the final equality comes from $\mathcal{T} = \mathcal{T}'$.

This is a contradiction, and, therefore, $\kappa \neq k$ implies that $\kappa' \neq k$. This proves that for each i, the function $\mathcal{D}_\infty^* \mapsto \mathbb{1}(C)/N_k(\mathcal{T})$ is a decreasing function of $X_{i,k}^*$ while keeping all other entries in $\mathcal{D}_\infty^*$ fixed and, from Theorem 2, we can conclude that the sample mean and empirical CDF of arm k from the lil'UCB algorithm are negatively and positive biased conditionally on the event the arm k is not chosen as the best arm, respectively. □

A.3. Proofs of Corollary 7 and 8

For any given $i \in \mathbb{N}$ and $k \in [K]$, let $\mathcal{D}_\infty^*$ and $\mathcal{D}_\infty^{*'}$ be two infinite collections of arm rewards and external randomness that agree with each other except on the (i,k)-th entries, $X_{i,k}^*$ and $X_{i,k}^{*'}$ of their stacks of rewards. Without loss of generality, we assume $X_{i,k}^* \leq X_{i,k}^{*'}$. Finally, let $(N_k(t), N_k'(t))$ denote the numbers of draws from arm k up to time $t \leq T$.

We first prove that the following inequality holds:

$$N_k(t) \leq N_k'(t), \;\; \forall t \leq T. \tag{59}$$

From Theorem 2, the above inequality implies that the sample mean and empirical CDF of each arm at a fixed time are negatively and positively biased, respectively as claimed in Corollary 7.

Proof of inequality (59). For each $t \leq T$, let $R_k(t)$ and $R_k'(t)$ be numbers of rounds in which arm k is in the active set up to time t under $\mathcal{D}_\infty^*$ and $\mathcal{D}_\infty^{*'}$, respectively. Since $N_k(t) = \sum_{r=1}^{R_k(t)} m_r$ and $N_k'(t) = \sum_{r=1}^{R_k'(t)} m_r$, it is enough to show $R_k(t) \leq R_k'(t)$. Now, for the sake of contradiction, suppose $R_k(t) > R_k'(t)$. Since arm k is active at rounds $r = 1, \ldots, R_k'(t)$ under both $\mathcal{D}_\infty^*$ and $\mathcal{D}_\infty^{*'}$, we must have that $\widehat{\mu}_k^{(r)} \leq \widehat{\mu}_k^{'(r)}$ for each $r = 1, \ldots, R_k'(t)$, where $\widehat{\mu}_k^{(r)}$ and $\widehat{\mu}_k^{'(r)}$ are averages of the observed samples from arm k at each round r under $\mathcal{D}_\infty^*$ and $\mathcal{D}_\infty^{*'}$, respectively. By the same logic, we have that $\mathcal{A}_r = \mathcal{A}_r'$ for each $r = 1, \ldots, R_k'(t)$, where $\mathcal{A}_r$ and $\mathcal{A}_r'$ are the sets of active arms under $\mathcal{D}_\infty^*$ and $\mathcal{D}_\infty^{*'}$, respectively. In particular, at the $R_k'(t)$-th round, under both scenarios $\mathcal{D}_\infty^*$ and $\mathcal{D}_\infty^{*'}$, we have the exactly same active sets and the same sample averages of all the active arms except for arm k, for which it hols that $\widehat{\mu}_k^{(R_k'(t))} \leq \widehat{\mu}_k^{'(R_k'(t))}$. However, it cannot be the case that arm k is eliminated from the active set after the $R_k'(t)$-th round under $\mathcal{D}_\infty^{*'}$ but the same arm still remains active after the same round under the scenario $\mathcal{D}_\infty^*$ in which the arm k has even a smaller sample average. This proves $R_k(t) \leq R_k'(t)$, which implies the claimed inequality (59) as desired. □

For the proof of Corollary 8, let κ and κ' be indices of the chosen arm under $\mathcal{D}_\infty^*$ and $\mathcal{D}_\infty^{*'}$, respectively. To prove the first two inequalities, we need to to show that

$$\frac{\mathbb{1}(\kappa \neq k)}{N_k(T)} \geq \frac{\mathbb{1}(\kappa' \neq k)}{N_k'(T)}. \tag{60}$$

From Equation (59), we have that $N_k(T) \leq N_k'(T)$. Therefore, it is enough to show that $\mathbb{1}(\kappa \neq k) \geq \mathbb{1}(\kappa' \neq k)$ which is equivalent to showing that

$$\mathbb{1}(\kappa = k) \leq \mathbb{1}(\kappa' = k). \tag{61}$$

For the sake of contradiction, suppose $\kappa = k$ but $\kappa' \neq k$. Recall that $R_k(T)$ and $R_k'(T)$ are the number of rounds in which arm k is in the active set up to time T under $\mathcal{D}_\infty^*$ and $\mathcal{D}_\infty^{*'}$ respectively. Since $\kappa = k$ but $\kappa' \neq k$, we have $\lceil \log_2 K \rceil = R_k(T) > R_k'(T)$. However this contradicts the fact $N_k(T) \leq N_k'(T)$, since $N_k(T) = \sum_{r=1}^{R_k(T)} m_r$ and $N_k'(T) = \sum_{r=1}^{R_k'(T)} m_r$. Hence, inequality (61) is true. By Theorem 2 this implies the first two inequalities in Corollary 8.

Similarly, to prove the last two inequalities in Corollary 8, we need to to show that

$$\frac{\mathbb{1}(\kappa = k)}{N_k(T)} \leq \frac{\mathbb{1}(\kappa' = k)}{N_k'(T)}. \tag{62}$$

If $\kappa \neq k$, the above inequality holds trivially. Hence, we can assume $\kappa = k$ and, from the inequality (61), we also have $\kappa' = k$. Since arm k is chosen as the best one under both $\mathcal{D}_\infty^*$ and $\mathcal{D}_\infty^{*'}$, we have that $N_k(T) = N_k'(T) = \sum_{r=1}^{\lceil \log_2 K \rceil} m_r$ which proves inequality (62). The result again follows from Theorem 2.

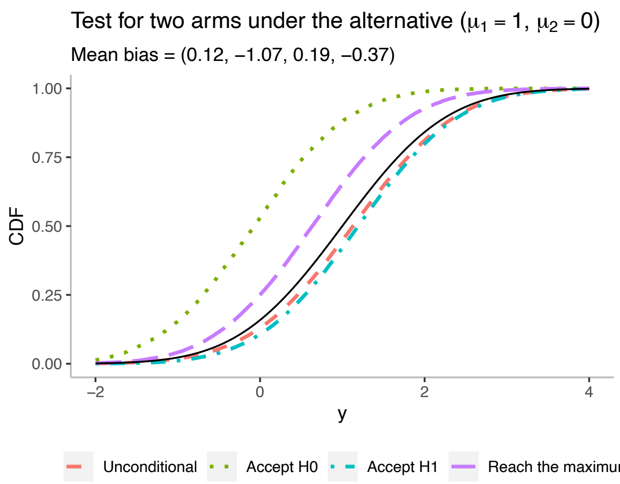

Figure 6. Average of conditional empirical CDFs of arm 1 from repeated sequential tests for two arms under the alternative hypothesis ($\mu_1 = 1, \mu_2 = 0$). See Section 4.2 for the detailed explanation about the sequential test.

B. Additional Simulation Results

In this section, we present additional simulation results for Section 4 and 5 which are omitted from the main part due to the page limit.

B.1. Conditional Bias Under Alternative Hypothesis in Section 4.2

As we conducted in Section 4.2, we have two standard normal arms with means μ_1 and μ_2. Then, we use the following upper and lower stopping boundaries to test whether $\mu_1 \leq \mu_2$ or not:

$$U(t) := z_{\alpha/2}\sqrt{\frac{2}{t}}, \quad \text{and} \quad L(t) = -U(t), \tag{63}$$

where α is set to 0.2 to show the bias better. In contrast to the experiment in Section 4.2 in which the true means are equal to each other, in this experiment, we set $\mu_1 = 1$ and $\mu_2 = 0$ to make the alternative hypothesis is true.

Figure 6 show the marginal and conditional biases of the empirical CDFs and sample means for arm 1 based on 10^5 repetitions of the experiment. The black solid line corresponds to the true underlying CDF. The red dashed line refers to the average of the marginal CDFs, and the purple long-dashed line corresponds to the average of the empirical CDFs conditionally on reaching the maximal time. For these two cases, although the marginal CDF is negatively and the conditional CDF is positively biased, these are not general phenomena and the sign of bias can be changed as we change mean parameters.

However, for the cases corresponding to accepting H_1 (blue dot-dashed line) and accepting H_0 (green dotted line), we can check that signs of biases of CDFs and sample means are consistent with what Theorem 2 and corresponding inequalities (18) to (17) described. Also note that the bias results do not depend on whether the arms are under the null or alternative hypotheses.

B.2. Experiment About the Bias of the Sequential Halving Algorithm in Section 4.4

To verify the conditional bias results in Corollary 7 and 8, we conducted 10^5 trials of the sequential halving algorithm on three unit-variance normal arms with $\mu_1 = 1, \mu_2 = 0.5$ and $\mu_3 = 0$ as we did for the lil'UCB algorithm in Section 4.3. Again, it is important to note that the signs of the biases do not depend on the choice of parameters or of the underlying distributions, but the magnitudes of the biases do. To best illustrates the bias results, we use an unusually small time budget $T = 10$ in this experiment.

The left side of Figure 7 shows the averages of the empirical CDFs of arm 1 (the arm with the largest mean) conditionally

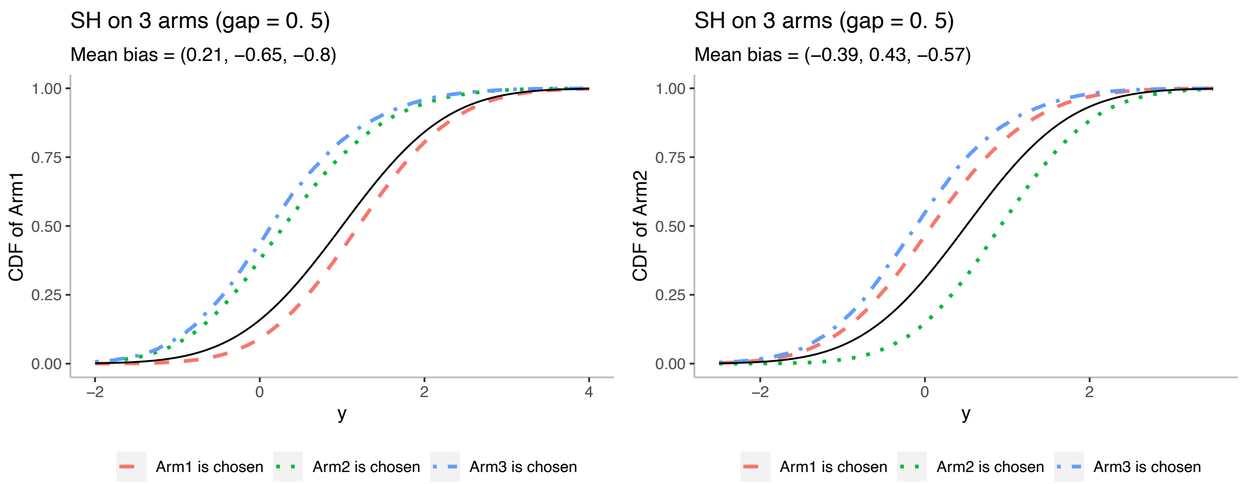

Figure 7. Average of conditional empirical CDFs of arm 1 (left) and arm 2 (right) from 10^5 *the sequential halving algorithm runs on three unit-variance normal arms with* $\mu_1 = 1, \mu_2 = 0.5$ *and* $\mu_3 = 0$*, as described in Appendix B.2. Black solid lines refer to the true CDF of arm 1 and arm 2.*

on each arm being chosen as the best arm. The thick black line corresponds to the true underlying CDF. The red dashed line, which lies below the true CDF, indicates that the empirical CDF of arm 1 conditionally on the event that arm 1 is chosen as the best arm (i.e., $\kappa = 1$) is negatively biased; this then implies that the sample mean of the chosen arm is positively biased. In contrast, the green dotted and blue dot-dashed lines, lying above the true CDF, show that conditionally on the event that arm 1 is not chosen as the best arm (i.e., $\kappa \neq 1$), the empirical CDF is positively biased and the sample mean is negatively biased.

The right side of Figure 7 displays the averages of the empirical CDFs of arm 2. Though arm 2 is not the best arm, we can check that the signs of the conditional biases follow the same pattern as arm 1. Conditionally on the event that arm 2 is chosen as the best arm ($\kappa = 2$), the empirical CDF is negatively biased (green dotted line), but conditionally on the event that arm 2 is not chosen as the best arm ($\kappa \neq 2$), the corresponding CDFs are now positively biased (red dashed and blue dot-dashed lines), as expected.

B.3. Experiments on Conditional Biases of Sample Variance and Median in MABs

As stated in Section 5, characterizing the bias of other important functionals such as sample variance and sample quantiles is an important open problem. In this subsection, we present a simulation study on the bias of sample variance and median.

For a given $n \geq 2$ i.i.d. samples $X_1, \ldots, X_n$ from a distribution P, the sample variance $\widehat{\sigma}^2$ and median $\widehat{m}$ are defined by

$$\widehat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2 \tag{64}$$

$$\widehat{m} = \begin{cases} \frac{1}{2}\left(X_{(n/2)} + X_{(n/2+1)}\right) & \text{if } n \text{ is even}, \\ X_{(\frac{n+1}{2})} & \text{if } n \text{ is odd}, \end{cases} \tag{65}$$

where $\bar{X}_n$ corresponds to the sample mean and $X_{(i)}$ refers to the i-th smallest sample.

It is well-known that for any distribution P with finite variance σ^2, the sample variance $\widehat{\sigma}^2$ is an unbiased estimator of σ^2. Also, though the sample median is not necessarily unbiased, for any symmetric distribution P including the normal distribution as a special case, the sample median is unbiased. However, for adaptively collected data from a MAB experiment, it is unclear whether the sample variance and median are unbiased or not. Furthermore, it is an open question how to characterize the bias of sample variance and median estimators if they are biased estimators.

As an initial step, we conduct the repeated sequential experiments described in Section 4.1 and empirically investigate the biases of the sample variance and median estimators. Figure 8 describes a simulation study on the bias of the sample

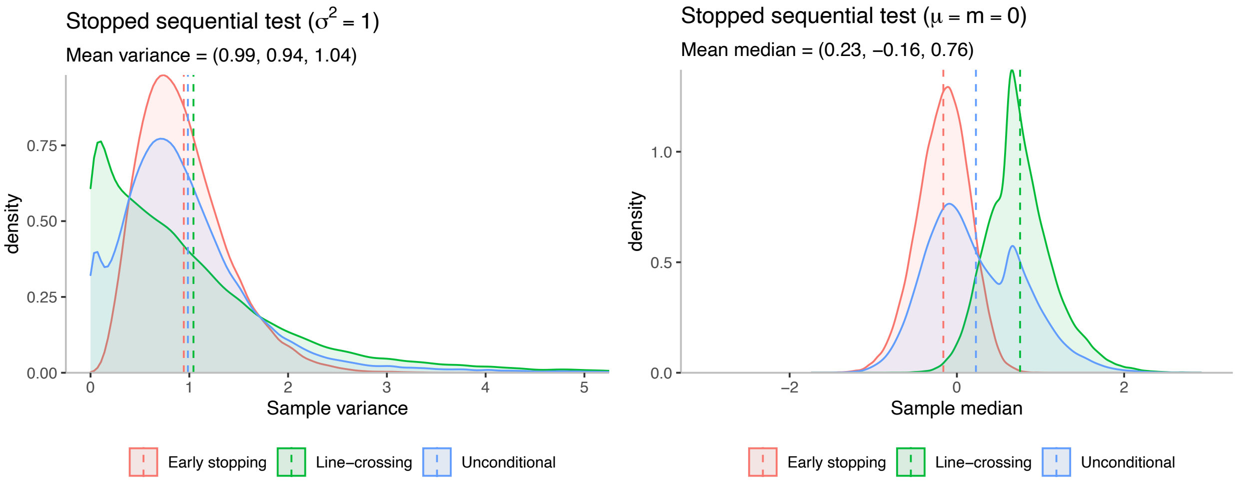

Figure 8. Left: Densities of observed sample variances from repeated stopped sequential test as described in Section 4.1. Right: Densities of observed sample median from the same repeated stopped sequential test. For both figures, vertical dashed lines correspond to averages of sample variances and medians on each conditions.

variance in the sequential testing setting of Section 4.1. Recall that, in this experiment, we have a stream of samples from a standard normal distribution. Each test terminates once either the number of samples reaches a fixed early stopping time $M = 10$ or the sample mean crosses the upper boundary $t \mapsto \frac{z_\alpha}{\sqrt{t}}$ with $\alpha = 0.2$.

Figure 8 shows the marginal and conditional distributions of the sample variance and median from 10^5 stopped sequential tests. Vertical lines correspond to averages of the sample variances and medians over repetitions of the experiment and under different conditions. For the sample variance, the simulation shows that the sample variance is negatively biased marginally and conditionally on the early stopping event. On the other hand, conditionally on the line-crossing event, the sample variance has a heavy right tail and is positively biased. For the sample median, we can check that, marginally and conditionally on the line-crossing event, the sample median is positively biased. In contrast, the sample median is negatively biased conditionally on the early stopping event. Note that, for the sample median, sizes of biases of the sample median are similar to ones from the sample means which were equal to $(0.22, -0.16, 0.75)$.