A. Lower Bound on the Worst Case Sample Complexity to Solve (k, m, n)

Theorem 3.1. *[Lower Bound for (k, m, n)] Let $\mathcal{L}$ be an algorithm that solves (k, m, n). Then, there exists an instance $(\mathcal{A}, n, m, k, \epsilon, \delta)$, with $0 < \epsilon \leq \frac{1}{\sqrt{32}}$, $0 < \delta \leq \frac{e^{-1}}{4}$, and $n \geq 2m$, $1 \leq k \leq m$, on which the expected number of pulls performed by $\mathcal{L}$ is at least $\frac{1}{18375} \cdot \frac{1}{\epsilon^2} \cdot \frac{n}{m-k+1} \ln \frac{\binom{m}{k-1}}{4\delta}$.*

The proof technique for Theorem 3.1 follows a path similar to that of (Kalyanakrishnan et al., 2012, Theorem 8), but differs in the fact that any k of the m (ϵ, m)-optimal arms needs to be returned as opposed to all the m.

A.1. Bandit Instances:

Assume we are given a set of n arms $\mathcal{A} = \{0, 1, 2, \cdots, n-1\}$. Let $I_0 \stackrel{\text{def}}{=} \{0, 1, 2, \cdots, m-k\}$ and $\mathcal{I}_l \stackrel{\text{def}}{=} \{I : I \subseteq \{\mathcal{A} \setminus I_0\} \wedge |I| = l\}$. Also for $I \subseteq \{m-k+1, m-k+2, \cdots, n-1\}$, we define

$$\bar{I} \stackrel{\text{def}}{=} \{m-k+1, m-k+2, \cdots, n-1\} \setminus I.$$

With each $I \in \mathcal{I}_{k-1} \cup \mathcal{I}_m$ we associate an n-armed bandit instance $\mathcal{B}^I$, in which each arm a produces a reward from a Bernoulli distribution with mean μ_a defined as:

$$\mu_a = \begin{cases} \frac{1}{2} & \text{if } a \in I_0 \\ \frac{1}{2} + 2\epsilon & \text{if } a \in I \\ \frac{1}{2} - 2\epsilon & \text{if } a \in \bar{I}. \end{cases} \tag{2}$$

Notice that all the instances in $\mathcal{I}_{k-1} \cup \mathcal{I}_m$ have exactly m (ϵ, m)-optimal arms. For $I \in \mathcal{I}_{k-1}$, all the arms in I_0 are (ϵ, m)-optimal, but for $I \in \mathcal{I}_m$ they are not. With slight overloading of notation we write $\mu(S)$ to denote the multi-set consisting of means of the arms in $S \subseteq \mathcal{A}$.

The key idea of the proof is that without sufficient sampling of each arm, it is not possible to correctly identify k of the (ϵ, m)-optimal arms with high probability.

A.2. Bounding the Error Probability:

We shall prove the theorem by first making the following assumption, which we shall demonstrate leads to a contradiction.

Assumption 1. *Assume, that there exists an algorithm $\mathcal{L}$, that solves each problem instance in (k, m, n) defined on bandit instance $\mathcal{B}^I$, $I \in \mathcal{I}_{k-1}$, and incurs a sample complexity SC_I. Then for all $I \in \mathcal{I}_{k-1}$, $\mathbb{E}[\mathrm{SC}_I] < \frac{1}{18375} \cdot \frac{1}{\epsilon^2} \cdot \frac{n}{m-k+1} \ln \left(\frac{\binom{m}{m-k+1}}{4\delta} \right)$, for $0 < \epsilon \leq \frac{1}{\sqrt{32}}$, $0 < \delta \leq \frac{e^{-1}}{4}$, and $n \geq 2m$, where $C = \frac{1}{18375}$.*

For convenience, we denote by $\Pr_I$ the probability distribution induced by the bandit instance $\mathcal{B}^I$ and the possible randomisation introduced by the algorithm $\mathcal{L}$. Also, let $S_{\mathcal{L}}$ be the set of arms returned (as output) by $\mathcal{L}$, and T_S be the total number of times the arms in $S \subseteq \mathcal{A}$ get sampled until $\mathcal{L}$ stops.

Then, as $\mathcal{L}$ solves (k, m, n), for all $I \in \mathcal{I}_{k-1}$

$$\Pr_I\{S_{\mathcal{L}} \subseteq I_0 \cup I\} \geq 1 - \delta. \tag{3}$$

Therefore, for all $I \in \mathcal{I}_{k-1}$

$$\mathbb{E}_I[T_{\mathcal{A}}] \leq C \frac{n}{(m-k+1)\epsilon^2} \ln \left(\frac{\binom{m}{m-k+1}}{4\delta} \right). \tag{4}$$

A.2.1. Changing $\Pr_I$ to $\Pr_{I \cup Q}$ where $Q \in \bar{I}$ s.t. $|Q| = m - k + 1$:

Consider an arbitrary but fixed $I \in \mathcal{I}_{k-1}$. Consider a fixed partitioning of $\mathcal{A}$, into $\left\lfloor \frac{n}{m-k+1} \right\rfloor$ subsets of size $(m-k+1)$ each. If Assumption (1) is correct, then for the instance $\mathcal{B}^I$, there are at most $\left\lfloor \frac{n}{4(m-k+1)} \right\rfloor - 1$ partitions $B \subset \bar{I}$, such that

$\mathbb{E}_I[T_B] \geq \frac{4C}{\epsilon^2}\ln\left(\frac{1}{4\delta}\right)$. Now, as $\left\lfloor \frac{n-m}{m-k+1} \right\rfloor - \left(\left\lfloor \frac{n}{4(m-k+1)} \right\rfloor - 1\right) \geq \left\lfloor \frac{n}{4(m-k+1)} \right\rfloor + 1 > 0$; therefore, there exists at least one subset $Q \in \bar{I}$ such that $|Q| = m-k+1$, and $\mathbb{E}_I[T_Q] < \frac{4C}{\epsilon^2}\ln\left(\frac{\binom{m}{m-k+1}}{4\delta}\right)$. Define $T^* = \frac{16C}{\epsilon^2}\ln\left(\frac{\binom{m}{m-k+1}}{4\delta}\right)$. Then using Markov's inequality we get:

$$\Pr_I\{T_Q \geq T^*\} < \frac{1}{4}. \tag{5}$$

Let $\Delta = 2\epsilon T^* + \sqrt{T^*}$ and also let K_Q be the total rewards obtained from Q.

Lemma A.1. *If $I \in \mathcal{I}_{k-1}$ and $Q \in \bar{I}$ s.t. $|Q| = m-k+1$, then*

$$\Pr_I\left\{T_Q \leq T^* \wedge K_Q \leq \frac{T_Q}{2} - \Delta\right\} \leq \frac{1}{4}\,.$$

Proof. Let $K_Q(t)$ be the total sum obtained from Q at the end of the trial t. As for $\mathcal{B}^{I_0}$, $\forall j \in Q\ \mu_j = 1/2 - 2\epsilon$, hence selecting and pulling one arm at each trial from Q following any rule (deterministic or probabilistic) is equivalent to selection of a single arm from Q for once and subsequently perform pulls on it. Hence whatever be the strategy of pulling one arm at each trial from Q, the expected reward for each pull will be $1/2 - 2\epsilon$. Let r_i be the i.i.d. reward obtained from the i^{th} trial. Then $K_Q(t) = \sum_{i=1}^{t} r_i$ and $Var[r_i] = \left(\frac{1}{2} - 2\epsilon\right)\left(\frac{1}{2} + 2\epsilon\right) = \left(\frac{1}{4} - 4\epsilon^2\right) < \frac{1}{4}$. As $\forall i : 1 \leq i \leq t$, r_i are i.i.d., we get $Var[K_Q(t)] = \sum_{i=1}^{t} Var(r_i) < \frac{t}{4}$. Now we can write the following:

$$\begin{aligned}
& \Pr_I\left\{\min_{1\leq t\leq T^*}\left(K_Q(t) - t\left(\frac{1}{2} - 2\epsilon\right)\right) \leq -\sqrt{T^*}\right\} \\
\leq\ & \Pr_I\left\{\max_{1\leq t\leq T^*}\left|K_Q(t) - t\left(\frac{1}{2} - 2\epsilon\right)\right| \geq \sqrt{T^*}\right\} \\
\leq\ & \frac{Var[K_Q(T^*)]}{T^*} < \frac{1}{4},
\end{aligned} \tag{6}$$

wherein we have used Kolmogorov's inequality. □

Lemma A.2. *Let $I \in \mathcal{I}_{k-1}$ and $Q \in \mathcal{I}_{m-k+1}$ such that $Q \subseteq \bar{I}$, and let W be some fixed sequence of rewards obtained by a single run of algorithm $\mathcal{L}$ on $\mathcal{B}^I$ such that $T_Q \leq T^*$ and $K_Q \geq \frac{T_Q}{2} - \Delta$, then:*

$$\Pr_{I\cup Q}\{W\} > \Pr_I\{W\} \cdot \exp(-32\epsilon\Delta). \tag{7}$$

Proof. Recall the fact that all the arms in Q have the same mean. Hence, if chosen one at each trial (following any strategy), the expected reward at each trial remains the same. Hence the probability of getting a given reward sequence generated from Q is independent of the sampling strategy. Again as the arms in Q have higher mean in $\mathcal{B}^Q$, the probability of getting the sequence (of rewards) decreases monotonically as the 1-rewards for $\mathcal{B}^{I_0}$ become fewer. So we get

$$\begin{aligned}
\Pr_{I\cup Q}\{W\} &= \Pr_I\{W\}\frac{\left(\frac{1}{2}+2\epsilon\right)^{K_Q}\left(\frac{1}{2}-2\epsilon\right)^{T_Q-K_Q}}{\left(\frac{1}{2}-2\epsilon\right)^{K_Q}\left(\frac{1}{2}+2\epsilon\right)^{T_Q-K_Q}} \\
&\geq \Pr_I\{W\}\frac{\left(\frac{1}{2}+2\epsilon\right)^{\left(\frac{T_Q}{2}-\Delta\right)}\left(\frac{1}{2}-2\epsilon\right)^{\left(\frac{T_Q}{2}+\Delta\right)}}{\left(\frac{1}{2}-2\epsilon\right)^{\left(\frac{T_Q}{2}-\Delta\right)}\left(\frac{1}{2}+2\epsilon\right)^{\left(\frac{T_Q}{2}+\Delta\right)}} \\
&= \Pr_I\{W\}\cdot\left(\frac{\frac{1}{2}-2\epsilon}{\frac{1}{2}+2\epsilon}\right)^{2\Delta} \\
&> \Pr_I\{W\}\cdot\exp(-32\epsilon\Delta)\left[\text{for } 0 < \epsilon \leq \frac{1}{\sqrt{32}}\right].
\end{aligned}$$

□

Lemma A.3. *If* (5) *holds for an* $I \in \mathcal{I}_{k-1}$ *and* $Q \in \mathcal{I}_{m-k+1}$ *such that* $Q \subseteq \bar{I}$, *and if* $\mathcal{W}$ *is the set of all possible reward sequences* W, *obtained by algorithm* $\mathcal{L}$ *on* $\mathcal{B}^I$, *then* $\Pr_{I\cup Q}\{\mathcal{W}\} > \left(\Pr_I\{\mathcal{W}\} - \frac{1}{2}\right) \cdot 4\delta$. *In particular,*

$$\Pr_{I\cup Q}\{S_{\mathcal{L}} \subseteq I_0 \cup I\} > \frac{\delta}{\binom{m}{m-k+1}}. \tag{8}$$

Proof. Let for some fixed sequence (of rewards) W, T_Q^W and K_Q^W respectively denote the total number of samples received by the arms in Q and the total number of 1-rewards obtained before the algorithm $\mathcal{L}$ stopped. Then:

$$\begin{aligned}
&\Pr_{I\cup Q}\{W\} = \Pr_{I\cup Q}(W : W \in \mathcal{W}) \\
&\geq \Pr_{I\cup Q}\left\{W : W \in \mathcal{W} \bigwedge T_Q^W \leq T^* \bigwedge K_Q^W \geq \frac{T_Q^W}{2} - \Delta\right\} \\
&> \Pr_{I}\left\{W : W \in \mathcal{W} \bigwedge T_Q^W \leq T^* \bigwedge K_Q^W \geq \frac{T_Q^W}{2} - \Delta\right\} \cdot \exp(-32\epsilon\Delta) \\
&\geq \left(\Pr_{I}\left\{W : W \in \mathcal{W} \bigwedge T_Q^W \leq T^*\right\} - \frac{1}{4}\right) \cdot \exp(-32\epsilon\Delta) \\
&\geq \left(\Pr_{I}\{\mathcal{W}\} - \frac{1}{2}\right) \cdot \frac{4\delta}{\binom{m}{m-k+1}} \quad \text{for } C = \frac{1}{18375},\ \delta < \frac{e^{-1}}{4}.
\end{aligned}$$

In the above, the 3rd, 4th and the last step are obtained using Lemma A.2, Lemma A.1 and Equation (5) respectively. The inequality (8) is obtained by using inequality (3), as $\Pr_I\{S_{\mathcal{L}} \in I_0\} > 1 - \delta \geq 1 - \frac{e^{-1}}{4} > \frac{3}{4}$. $\square$

A.2.2. SUMMING OVER $\mathcal{I}_{k-1}$ AND $\mathcal{I}_m$

Now, we sum up the probability of errors across all the instances in $\mathcal{I}_{k-1}$ and $\mathcal{I}_m$. If the Assumption 1 is true, using the pigeon-hole principle we show that there exists some instance for which the mistake probability is greater than δ.

$$
\begin{aligned}
&\sum_{J \in \mathcal{I}_m} \Pr_J\{S_{\mathcal{L}} \nsubseteq J\} \\
&\geq \sum_{J \in \mathcal{I}_m} \sum_{\substack{J' \subset J \\ :|J'|=m-k+1}} \Pr_J\{S_{\mathcal{L}} \subseteq \{J \setminus J'\} \cup I_0\} \\
&\geq \sum_{J \in \mathcal{I}_m} \sum_{\substack{J' \subset J \\ :|J'|=m-k+1}} \Pr_J\{\exists a \in I_0 : S_{\mathcal{L}} = \{J \setminus J'\} \cup \{a\}\} \\
&= \sum_{J \in \mathcal{I}_m} \sum_{\substack{J' \subset J \\ :|J'|=m-k+1}} \sum_{I \in \mathcal{I}_{k-1}} \mathbb{1}[I \cup J' = J] \cdot \Pr_J\{S_{\mathcal{L}} \subseteq I \cup I_0\} \\
&= \sum_{J \in \mathcal{I}_m} \sum_{\substack{J' \subset \mathcal{A} \setminus I_0 \\ :|J'|=m-k+1}} \sum_{I \in \mathcal{I}_{k-1}} \mathbb{1}[I \cup J' = J] \cdot \Pr_J\{S_{\mathcal{L}} \subseteq I \cup I_0\} \\
&= \sum_{J \in \mathcal{I}_m} \sum_{I \in \mathcal{I}_{k-1}} \sum_{\substack{J' \subset \mathcal{A} \setminus I_0 \\ :|J'|=m-k+1}} \mathbb{1}[I \cup J' = J] \cdot \Pr_J\{S_{\mathcal{L}} \subseteq I \cup I_0\} \\
&= \sum_{I \in \mathcal{I}_{k-1}} \sum_{J \in \mathcal{I}_m} \sum_{\substack{J' \subset \bar{I} \\ :|J'|=m-k+1}} \mathbb{1}[I \cup J' = J] \cdot \Pr_J\{S_{\mathcal{L}} \subseteq I \cup I_0\} \\
&= \sum_{I \in \mathcal{I}_{k-1}} \sum_{\substack{J' \subset \bar{I} \\ :|J'|=m-k+1}} \sum_{J \in \mathcal{I}_m} \mathbb{1}[I \cup J' = J] \cdot \Pr_J\{S_{\mathcal{L}} \subseteq I \cup I_0\} \\
&= \sum_{I \in \mathcal{I}_{k-1}} \sum_{\substack{J' \subset \bar{I} \\ :|J'|=m-k+1}} \Pr_{I \cup J'}\{S_{\mathcal{L}} \subseteq I \cup I_0\}
\end{aligned}
$$

Recall that $\forall I \in \mathcal{I}_{k-1}$ there exists a set $Q \subset \mathcal{A} \setminus \{I \cup I_0\} : |Q| = (m-k+1)$, such that $T_Q < T^*$. Therefore,

$$
\begin{aligned}
&\sum_{J \in \mathcal{I}_m} \Pr_J\{S_{\mathcal{L}} \nsubseteq J\} \\
&\geq \sum_{I \in \mathcal{I}_{k-1}} \sum_{\substack{J' \subset \bar{I} \\ :|J'|=m-k+1}} \Pr_{I \cup J'}\{S_{\mathcal{L}} \subseteq I \cup I_0\} \\
&> \sum_{I \in \mathcal{I}_{k-1}} \sum_{\substack{J' \subset \bar{I} \\ :|J'|=m-k+1}} \frac{\delta}{\binom{m}{m-k+1}} \\
&\geq \sum_{I \in \mathcal{I}_{k-1}} \binom{n-m}{m-k+1} \cdot \frac{\delta}{\binom{m}{m-k+1}} \\
&\geq \binom{n-(m-k+1)}{k-1} \cdot \binom{n-m}{m-k+1} \cdot \frac{\delta}{\binom{m}{m-k+1}} \\
&= \binom{n-(m+k-1)}{m} \delta \\
&= |\mathcal{I}_m| \delta.
\end{aligned}
$$

Hence, we get a contradiction to Assumption 1, thereby proving the theorem.

B. Analysis of LUCB-k-m

Let at time t, $\hat{p}_a^t$ be the empirical mean of the arm $a \in \mathcal{A}$, and u_a^t be the number of times the arm a has been pulled until (and excluding) time t. For a given $\delta \in (0, 1]$, we define $\beta(u_a^t, t, \delta) = \sqrt{\frac{1}{2u_a^t} \ln \frac{k_1 n t^4}{\delta}}$, where $k_1 = 5/4$. We define upper and lower confidence bound on the estimate of the true mean of arm $a \in \mathcal{A}$ as $ucb(a, t) = \hat{p}_a + \beta(u_a^t, t, \delta)$, and $lcb(a, t) = \hat{p}_a - \beta(u_a^t, t, \delta)$ respectively.

To analyse the sample complexity, first we define some events, at least one of which must occur if the algorithm does not stop at the round t.

PROBABLE EVENTS. Let $a, b \in \mathcal{A}$, such that $\mu_a > \mu_b$. During the run of the algorithm, any of the following five events may occur:
i) The empirical mean of an arm may falls outside the upper or the lower confidence bound. We define it as:

$$CROSS_a^t \stackrel{\text{def}}{=} \{ucb(a, t) < \mu_a \vee lcb(a, t) > \mu_a\}.$$

ii) The empirical mean of arm a may be lesser than that of arm b; we definite as:

$$ErrA(a, b, t) \stackrel{\text{def}}{=} \{\hat{p}_a^t < \hat{p}_b^t\}.$$

iii) The lower and upper confidence bounds of arm a may fall below those of arm b; we define them as:

$$ErrL(a, b, t) \stackrel{\text{def}}{=} \{lcb(a, t) < lcb(b, t)\},$$
$$ErrU(a, b, t) \stackrel{\text{def}}{=} \{ucb(a, t) < ucb(b, t)\}.$$

iv) If an arm's confidence bounds are above a certain radius (say d), we define that event as

$$NEEDY_a^t(d) \stackrel{\text{def}}{=} \{\{lcb(a, t) < \mu_a - d\} \vee \{ucb(a, t) > \mu_a + d\}\}.$$

Let $u^*(a, t) \stackrel{\text{def}}{=} \left\lceil \frac{32}{\max\{\Delta_a, \frac{\epsilon}{2}\}^2} \ln \frac{k_1 n t^4}{\delta} \right\rceil$ for all $a \in \mathcal{A}$, where $k_1 = 5/4$. We show that any arm a, if sampled sufficiently, that is $u_a^t \geq u^*(a, t)$, then occurrence of any of the PROBABLE EVENTS imply occurrence of $CROSS_a^t$. First we show that if $CROSS_a^t$ does not occur for any $a \in \mathcal{A}$, then occurrence of any one of the PROBABLE EVENTS implies the occurrence of $NEEDY_a^t(\cdot)$ or $NEEDY_b^t(\cdot)$.

Lemma B.1. *[Expressing* PROBABLE EVENTS *in terms of $NEEDY_a^t$ and $CROSS_a^t$] To prove that* $\{\neg CROSS_a^t \wedge \neg CROSS_b^t\} \wedge \{ErrA(a, b, t) \vee ErrU(a, b, t) \vee ErrL(a, b, t)\} \implies \{NEEDY_a^t\left(\frac{\Delta_{ab}}{2}\right) \vee NEEDY_b^t\left(\frac{\Delta_{ab}}{2}\right)\}$.

Proof. $\mathbf{ErrA(a, b, t)}$: To prove that $\neg\{CROSS_a^t \vee CROSS_b^t\} \wedge ErrA(a, b, t) \implies NEEDY_a^t\left(\frac{\Delta_{ab}}{2}\right) \vee NEEDY_b^t\left(\frac{\Delta_{ab}}{2}\right)$.

$$\begin{aligned}
&ErrA(a, b, t) \implies \hat{p}_a^t < \hat{p}_b^t \\
&\implies \hat{p}_a^t - (p_a - \beta(u_a^t, t, \delta)) < \hat{p}_b^t - (p_b + \beta(u_b^t, t, \delta) + \\
&(\beta(u_a^t, t, \delta) + \beta(u_b^t, t, \delta)) - \Delta_{ab}/2) \\
&\implies NEEDY_a^t\left(\frac{\Delta_{ab}}{2}\right) \vee NEEDY_b^t\left(\frac{\Delta_{ab}}{2}\right).
\end{aligned}$$

$\mathbf{ErrU(a, b, t)}$: To prove that $\neg\{CROSS_a^t \vee CROSS_b^t\} \wedge ErrU(a, b, t) \implies NEEDY_b^t\left(\frac{\Delta_{ab}}{2}\right)$.

Assuming $\neg CROSS_a^t \wedge \neg CROSS_b^t$ we get

$$\begin{aligned}
&ErrU(a,b,t) \implies \{ucb(b,t) > ucb(a,t)\} \\
&\implies \{\hat{p}_b^t + \beta(u_b^t, t, \delta) > \hat{p}_a^t + \beta(u_a^t, t, \delta)\} \\
&\implies \{\hat{p}_b^t > \mu_b + \beta(u_b^t, t, \delta)\} \vee \{\hat{p}_a^t < \mu_a - \beta(u_a^t, t, \delta)\} \vee \\
&\quad \{2\beta(u_b^t, t, \delta) > \Delta_{ab}\} \\
&\implies NEEDY_b^t \left(\frac{\Delta_{ab}}{2}\right).
\end{aligned}$$

$\mathbf{ErrL(a, b, t)}$: To prove that $\neg\{CROSS_a^t \vee CROSS_b^t\} \wedge ErrL(a,b,t) \implies NEEDY_a^t\left(\frac{\Delta_{ab}}{2}\right)$.
Assuming $\neg CROSS_a^t \wedge \neg CROSS_b^t$ we get

$$\begin{aligned}
&ErrL(a,b,t) \implies \{lcb(b,t) > lcb(a,t)\} \\
&\implies \{\hat{p}_b^t - \beta(u_b^t, t, \delta) > \hat{p}_a^t - \beta(u_a^t, t, \delta)\} \\
&\implies \{\hat{p}_b^t > \mu_b + \beta(u_b^t, t, \delta)\} \vee \{\hat{p}_a^t < \mu_a - \beta(u_a^t, t, \delta)\} \vee \\
&\quad \{2\beta(u_a^t, t, \delta) > \Delta_{ab}\} \\
&\implies NEEDY_a^t \left(\frac{\Delta_{ab}}{2}\right).
\end{aligned}$$

□

We show that given a threshold d, if an arm a is sufficiently sampled, such that $\beta(u_a^t, t, \delta) \leq \frac{d}{2}$, then $NEEDY_a^t$ infers $CROSS_a^t$.

Lemma B.2. *For any $a \in \mathcal{A}$, $\{NEEDY_a^t(d) | \beta(u_a^t, t, \delta) < d/2\} \implies CROSS_a^t$.*

Proof. First, we show that $\{lcb(a,t) < \mu_a - d | \beta(u_a^t, t, \delta) < d/2\} \implies CROSS_a^t$,

$$\begin{aligned}
&\{lcb(a,t) < \mu_a - d | \beta(u_a^t, t, \delta) < d/2\} \\
&\implies \{\hat{p}_a^t - \beta(u_a^t, t, \delta) < \mu_a - d | \beta(u_a^t, t, \delta) < d/2\} \\
&\implies \{\hat{p}_a^t < \mu_a - d + \beta(u_a^t, t, \delta) | \beta(u_a^t, t, \delta) < d/2\} \\
&\implies \{\hat{p}_a^t < \mu_a - d/2 | \beta(u_a^t, t, \delta) < d/2\} \\
&\implies CROSS_a^t.
\end{aligned} \tag{9}$$

The other part follows the similar way. □

By the very definition of confidence bound, at any round t, the probability that the empirical mean of an arm will lie outside it, is very low. In other words, the probability of occurrence $CROSS_a^t$ is very low for all t and $a \in \mathcal{A}$.

Lemma B.3. *[Upper bounding the probability of $CROSS_a^t$] $\forall a \in \mathcal{A}$ and $\forall t \geq 0$, $\Pr\{CROSS_a^t\} \leq \frac{\delta}{knt^4}$. Hence, $P\left[\exists t \geq 0 \wedge \exists a \in \mathcal{A} : CROSS_a^t | u_a^t \geq 0\right] \leq \frac{\delta}{k_1 t^3}$.*

Proof. $\Pr\{CROSS_a^t\}$ is upper bounded by using Hoeffding's inequality, and the next statement gets proved by taking union bound over all arms and t. □

Now, recalling the definition of h_*^t, and l_*^t from Algorithm 1, we present the key logic underlying the analysis of LUCB-k-m. The idea is to show that if the algorithm has not stopped, then one of those PROBABLE EVENTS must have occurred. Then using Lemma B.1, and Lemma B.2, Lemma B.3 we show that beyond a certain number of rounds, the probability that LUCB-k-m will continue is sufficiently small. Lastly, using the argument based on pigeon-hole principle, similar to Lemma 5 of Kalyanakrishnan (2011), we establish the upper bound on the sample complexity. Below we present the core logic that shows, until the algorithm stops one of the PROBABLE EVENTS must occur.

Case 1 $h_*^t \in B_1 \wedge l_*^t \in B_1$

if $\exists b_3 \in A_1^t \cap B_3$ **then**
  Then $ErrL(h_*^t, b_3, t)$ has occurred.
else
  $\exists b_3 \in A_2^t \cap B_3$
  Then $ErrA(h_*^t, b_3, t)$ has occurred.
end if

Case 2 $h_*^t \in B_1 \wedge l_*^t \in B_2$

if $\exists b_3 \in A_1^t \cap B_3$ **then**
  Then $ErrL(h_*^t, b_3^t, t)$ has occurred.
else
  $\exists b_3 \in A_2^t \cap B_3$.
  if $\Delta_{h_*^t l_*^t} \geq \frac{\Delta_{h_*^t}}{2}$ **then**
    Then $NEEDY_{h_*^t}^t(\Delta_{h_*^t}/4) \vee NEEDY_{l_*^t}^t(\Delta_{h_*^t}/4)$ has occurred.
  else
    Then $ErrL(l_*^t, b_3^t, t)$ has occurred.
  end if
end if

Case 3 $h_*^t \in B_1 \wedge l_*^t \in B_3$

Then $NEEDY_{h_*^t}^t(\Delta_{h_*^t}/4) \vee NEEDY_{l_*^t}^t(\Delta_{l_*^t}/4)$ has occurred.

Case 4 $h_*^t \in B_2 \wedge l_*^t \in B_1$

if $\Delta_{h_*^t l_*^t} \geq \frac{\Delta_{h_*^t}}{2}$ **then**
  Then $ErrA(l_*^t, h_*^t, t)$ has occurred.
else
  if $\exists b_3 \in A_1^t \cap B_3$ **then**
    Then $ErrL(h_*^t, b_3^t, t)$ has occurred.
  else
    $\exists b_3 \in A_2^t \cap B_3$
    $\therefore ErrA(l_*^t, b_3, t)$ has occurred.
  end if
end if

Case 5 $h_*^t \in B_2 \wedge l_*^t \in B_2$ and $\Delta_{h_*^t l_*^t} > 0$

Here, $\exists b_1 \in (A_2^t \cup A_3^t) \cap B_1$ and $\exists b_3 \in (A_1^t \cup A_2^t) \cap B_3$
if $|\Delta_{h_*^t l_*^t}| < \Delta_{h_*^t}/2$ **then**
  if $\Delta_{b_1 h_*^t} > \Delta_{b_1}/4$ **then**
    if $b_1 \in A_2^t) \cap B_1$ **then**
      $ErrA(b_1, h_*^t, t)$
    else
      $b_1 \in A_3^t \cap B_1$
      $ErrU(b_1, l_*^t, t)$ has occurred.
    end if
  else
    $\Delta_{b_1 h_*^t} \leq \Delta_{b_1}/4$ and hence $\Delta_{l_*^t b_3} \geq \Delta_{l_*^t}/4$
    if $b_3 \in A_2^t \cap B_3$ **then**
      $ErrA(l_*^t, b_3, t)$ has occurred.
    else
      $b_3 \in A_1^t \cap B_3$
      $ErrL(h_*^t, b_3, t)$ has occurred.
    end if
  end if
else
  $|\Delta_{h_*^t l_*^t}| > \Delta_{h_*^t}/2$
  $NEEDY_{h_*^t}^t(\Delta_{h_*^t}/4) \vee NEEDY_{l_*^t}^t(\Delta_{h_*^t}/4)$ has occurred.
end if

Case 5 (continued) $h_*^t \in B_2 \wedge l_*^t \in B_2$ and $\Delta_{h_*^t l_*^t} \leq 0$

Here, $\exists b_1 \in (A_2^t \cup A_3^t) \cap B_1$ and $\exists b_3 \in (A_1^t \cup A_2^t) \cap B_3$
if $|\Delta_{h_*^t l_*^t}| < \Delta_{h_*^t}/2$ **then**
  if $\Delta_{b_1 l_*^t} > \Delta_{b_1}/4$ **then**
    if $b_1 \in A_2^t \cap B_1$ **then**
      $ErrA(b_1, h_*^t, t)$ has occurred.
    else
      $b_1 \in A_3^t \cap B_1$
      $ErrU(b_1, l_*^t, t)$ has occurred.
    end if
  else
    $\Delta_{b_1 l_*^t} \leq \Delta_{b_1}/4$ and hence $\Delta_{h_*^t b_3} \geq \Delta_{h_*^t}/4$
    if $b_3 \in A_2^t \cap B_3$ **then**
      $ErrA(l_*^t, b_3, t)$ has occurred.
    else
      $b_3 \in A_1^t \cap B_3$
      $ErrL(h_*^t, b_3, t)$ has occurred.
    end if
  end if
else
  $|\Delta_{h_*^t l_*^t}| > \Delta_{h_*^t}/2$
  $NEEDY_{h_*^t}^t(\Delta_{h_*^t}/4) \vee NEEDY_{l_*^t}^t(\Delta_{h_*^t}/4)$ has occurred.
end if

Case 6 $h_*^t \in B_2 \wedge l_*^t \in B_3$

if $\Delta_{h_*^t l_*^t} \geq \frac{\Delta_{l_*^t}}{2}$ **then**
  Then $NEEDY_{h_*^t}^t(\Delta/4) \vee NEEDY_{l_*^t}^t(\Delta_{l_*^t}/4)$ has occurred.
else
  $\Delta_{h_*^t l_*^t} < \frac{\Delta_{l_*^t}}{2}$
  $\therefore \forall b_1 \in \{A_2^t \cup A_3^t\} \cap B_1, \Delta_{b_1 h_*^t} > \frac{\Delta_{b_1}}{2}$.
  if $\exists b_1 \in A_2^t \cap B_1$ **then**
    $ErrA(b_1, h_*^t, t)$ has occurred.
  else
    $\exists b_1 \in A_3^t \cap B_1$.
    Then $ErrU(b_1^t, l_*^t, t)$ has occurred.
  end if
end if

Case 7 $h_*^t \in B_3 \wedge l_*^t \in B_1$

$\therefore ErrA(l_*^t, h_*^t, t)$ has occurred.

Case 8 $h_*^t \in B_3 \wedge l_*^t \in B_2$

if $\Delta_{h_*^t l_*^t} \geq \frac{\Delta_{h_*^t}}{2}$ **then**
  $ErrA(l_*^t, h_*^t, t)$ has occurred.
else
  $\Delta_{h_*^t l_*^t} < \frac{\Delta_{h_*^t}}{2}$
  $\therefore \forall b_1 \in \{A_2^t \cup A_3^t\} \cap B_1, \Delta_{b_1 l_*^t} > \frac{\Delta_{b_1}}{2}$.
  if $\exists b_1 \in A_2^t \cap B_1$ **then**
    $ErrA(b_1, h_*^t, t)$ has occurred.
  else
    $\exists b_1 \in A_3^t \cap B_1$.
    $\therefore ErrU(b_1, l_*^t, t)$ has occurred.
  end if
end if

Case 9 $h_*^t \in B_3 \wedge l_*^t \in B_3$

$\exists b_1 \in \{A_2^t \cup A_3^t\} \cap B_1$
if $\exists b_1 \in A_2^t \cap B_1$ **then**
  $ErrA(b_1, h_*^t, t)$ has occurred.
else
  $\exists b_1 \in A_3^t \cap B_1$
  $\therefore ErrA(b_1, l_*^t, t)$ has occurred.
end if

Lemma B.4 (H). *If $T = CH_\epsilon \ln\left(\frac{H_\epsilon}{\delta}\right)$, then for $C \geq 2732$, the following holds:*

$$T > 2 + 2\sum_{a \in \mathcal{A}} u^*(a, T).$$

Proof. This proof is taken from Appendix B.3 of Kalyanakrishnan (2011).

$$2 + 2\sum_{a \in \mathcal{A}} u^*(a, T) = 2 + 64\sum_{a \in \mathcal{A}} \left\lceil \frac{1}{\max(\Delta_a, (\epsilon/2))^2} \ln \frac{knt^4}{\delta} \right\rceil$$

$$
\begin{aligned}
&\leq 2 + 64n + 64H_\epsilon \ln \frac{knT^4}{\delta} \\
&= 2 + 64n + 64H_\epsilon \ln k + 64H_\epsilon \ln \frac{n}{\delta} + 256H_\epsilon \ln T \\
&< (66 + 64\ln k)H_\epsilon + 64H_\epsilon \ln \frac{n}{\delta} + 256H_\epsilon \left[\ln C + \ln H_\epsilon + \ln\ln \frac{H_\epsilon}{\delta}\right] \\
&< (66 + 64\ln k)H_\epsilon + 64H_\epsilon \ln \frac{n}{\delta} + 256H_\epsilon \left[\ln C + \ln H_\epsilon + \ln\ln \frac{H_\epsilon}{\delta}\right] \\
&< 130H_\epsilon + 64H_\epsilon \ln \frac{n}{\delta} + 256H_\epsilon \left[\ln C + \ln H_\epsilon + \ln \frac{H_\epsilon}{\delta}\right] \\
&< 130H_\epsilon + 64H_\epsilon \ln \frac{H_\epsilon}{\delta} + 256H_\epsilon \left[\ln C + 2\ln \frac{H_\epsilon}{\delta}\right] \\
&< (706 + 256\ln C)H_\epsilon \ln \frac{H_\epsilon}{\delta} < CH_\epsilon \ln \frac{H_\epsilon}{\delta} \quad [\text{For } C \geq 2732] .
\end{aligned}
$$

□

Lemma B.5. *Let $T^* = \left\lceil 2732 H_\epsilon \ln\left(\frac{H_\epsilon}{\delta}\right)\right\rceil$. For every $T > T_1^*$, the probability that the Algorithm 1 has not terminated after T rounds of sampling is at most $\frac{8\delta}{T^2}$.*

Proof. Letting $\bar{T} = \frac{T}{2}$ we define two events for $\bar{T} \leq t \leq T-1$: $E^{(1)} \stackrel{\text{def}}{=} \exists a \in \mathcal{A} : CROSS_a^t$ and $E^{(2)} \stackrel{\text{def}}{=} \exists NEEDY_a^t\left(\frac{\Delta_a}{4}\right)$. If the algorithm stops for $t < \bar{T}$, then there is nothing to prove. On the contrary, let the algorithm has not stopped after $t > \bar{T}$ and neither $E^{(1)}$ nor $E^{(2)}$ has occurred. Letting N_{rounds} be the the required number of rounds beyond $\bar{T}$, we can upper bound it as:

$$
\begin{aligned}
N_{rounds} &= \sum_{t=\bar{T}} \left\{ \mathbb{1}\left[NEEDY_{h_*^t}^t\left(\frac{\Delta_{h_*^t}}{4}\right) \vee NEEDY_{m_*^t}^t\left(\frac{\Delta_{m_*^t}}{4}\right) \vee NEEDY_{l_*^t}^t\left(\frac{\Delta_{l_*^t}}{4}\right)\right]\right\} \\
&\leq \sum_{\bar{T}}^{T-1} \sum_{a\in\mathcal{A}} \mathbb{1}\left[a \in \{h_*^t, m_*^t, l_*^t\} \wedge NEEDY_a^t\left(\frac{\Delta_a}{4}\right)\right] \\
&= \sum_{\bar{T}}^{T-1} \sum_{a\in\mathcal{A}} \mathbb{1}[a \in \{h_*^t, m_*^t, l_*^t\} \wedge (u_a^t < u^*(a,t))] \\
&\leq \sum_{\bar{T}}^{T-1} \sum_{a\in\mathcal{A}} \mathbb{1}[a \in \{h_*^t, m_*^t, l_*^t\} \wedge (u_a^t < u^*(a,t))] \\
&\leq \sum_{a\in\mathcal{A}} \sum_{\bar{T}}^{T-1} \mathbb{1}[(a \in \{h_*^t, m_*^t, l_*^t\}) \wedge (u_a^t < u^*(a,t))] \\
&\leq \sum_{a\in\mathcal{A}} u^*(a,t).
\end{aligned}
$$

Using Lemma B.4, $T \geq T^* \Rightarrow T > 2 + 2\sum_{a\in\mathcal{A}} u^*(a,t)$. Hence, if neither $E^{(1)}$ nor $E^{(2)}$ occurs then the algorithm runs for at most $\bar{T} + N_{rounds} \leq \lceil T/2 \rceil + \sum_{a\in\mathcal{A}} 16u^*(a,t) < T$ number of rounds.

The probability that the algorithm does not stop within T rounds, is upper-bounded by $P[E^{(1)} \vee E^{(2)}]$. Applying Lemma B.2 and Lemma B.3,

$$
P[E^{(1)} \vee E^{(2)}] \leq \sum_{t=\bar{T}}^{T-1} \left(\frac{\delta}{k_1 t^3} + \frac{\delta}{kt^4}\right) \leq \sum_{t=\bar{T}}^{T-1} \frac{\delta}{k_1 t^3}\left(1 + \frac{2}{t}\right) \leq \left(\frac{T}{2}\right)\frac{8\delta}{k_1 T^3}\left(1 + \frac{4}{T}\right) < \frac{8\delta}{T^2}.
$$

□

Theorem 3.2. *[Expected Sample Complexity of* LUCB-*k-m]* LUCB-*k-m solves* (k, m, n) *using an expected sample complexity upper-bounded by* $O\left(H_\epsilon \log \frac{H_\epsilon}{\delta}\right)$.

Using Lemma B.4, and Lemma B.5 the expected sample complexity of the Algorithm 1 can be upper bounded as

$$E[SC] \leq 2\left(T_1^* + \sum_{t=T_1^*}^{\infty} \frac{8\delta}{T^2}\right) \leq 5464 \cdot \left(H_\epsilon \ln\left(\frac{H_\epsilon}{\delta}\right)\right) + 32. \tag{10}$$

C. Proof of Theorem 4.7

Algorithm 4 describes OPTQP. It uses $\mathcal{P}_2$ (Roy Chaudhuri & Kalyanakrishnan, 2017) with MEDIAN ELIMINATION as the subroutine (inside $\mathcal{P}_2$), to select an $[\epsilon, \rho]$-optimal arm with confidence $1-\delta'$. We have assumed $\delta' = 1/4$, in practice the one can choose any sufficiently small value for it, which will merely affect the multiplicative constant in the upper bound.

Algorithm 4 OPTQP

Input: $\mathcal{A}, \epsilon, \delta$, and OPTQF.

Output: A single $[\epsilon, \rho]$-optimal arm

Set $\delta' = 1/4$, $u = \left\lceil \frac{1}{2(0.5-\delta')^2} \cdot \log \frac{2}{\delta} \right\rceil = \left\lceil 8 \log \frac{2}{\delta} \right\rceil$.

Run u copies of $\mathcal{P}_2(\mathcal{A}, \rho, \epsilon/2, \delta')$ and form set S with the output arms.

Return the output from OPTQF $\left(S, u, \lfloor \frac{u}{2} \rfloor, 1, \frac{\epsilon}{2}, \frac{\delta}{2}\right)$.

Theorem C.1. *[Correctness and Sample Complexity of* OPTQP*] If* OPTQF *exists, then* OPTQP *solves* Q-P*, within the sample complexity* $\Theta\left(\frac{1}{\rho\epsilon^2} \log \frac{1}{\delta} + \gamma(\cdot)\right)$.

Proof. First we prove the correctness and then upper bound the sample complexity.

Correctness. First we notice that each copy of $\mathcal{P}_2$ outputs an $[\epsilon/2, \rho]$-optimal arm with probability at least $1-\delta'$. Also, OPTQF outputs an $[\epsilon/2, \rho]$-optimal arm with probability $1-\delta$. Let, $\hat{X}$ be the fraction of sub-optimal arms in S. Then $\Pr\{\hat{X} \geq \frac{1}{2}\} = \Pr\{\hat{X} - \delta' \geq \frac{1}{4}\} \leq \exp(-2 \cdot (\frac{1}{4})^2 \cdot u) = \exp(-2 \cdot \frac{1}{16} \cdot 8 \log \frac{2}{\delta}) < \frac{\delta}{2}$. On the other hand, the mistake probability of OPTQF is upper bounded by $\delta/2$. Therefore, by taking union bound, we get the mistake probability is upper bounded by δ. Also, the mean of the output arm is not less than $\frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$ from the $(1-\rho)$-th quantile.

Sample complexity. First we note that, for some appropriate constant C, the sample complexity (SC) of each of the u copies of $\mathcal{P}_2$ is $\frac{C}{\rho(\epsilon/2)^2}\left(\log \frac{2}{\delta'}\right)^2 \in O\left(\frac{1}{\rho\epsilon^2}\right)$. Hence, SC of all the u copies $\mathcal{P}_2$ together is upper bounded by $\frac{C_1 \cdot u}{\rho\epsilon^2}$, for some constant C_1. Also, for some constant C_2, the sample complexity of OPTQF is upper bounded by $C_2\left(\frac{u}{(u/2)(\epsilon/2)^2} \log \frac{2}{\delta} + \gamma(\cdot)\right) = C_2\left(\frac{8}{\epsilon^2} \log \frac{2}{\delta} + \gamma(\cdot)\right)$. Now, adding the sample complexities, and substituting for u we prove the bound. $\square$