A Notations

Table 1: Summary of notations

A_b	set of arms played during block b
A_t	set of arms played during time step t
$A_{u,t}$	arm played by user u during time step t
A_*	set of top U arms
A^-	set of arms $A \backslash A_*$
B	number of blocks
$B_{a,b}$	number of blocks arm a is played up till block b
K	number of arms
$L_{b,i}$	set of arms in A_b^- where $B_{a,b-1} \leq m_{b,i}$
N_a	number of sub-optimal sets played that includes a
p_a	probability density of reward for arm a
$p_{\pi\nu}$	probability density for the interaction between π and ν
$S_{u,t}$	cumulative reward for user u after t time step
X_t	set of rewards obtained during time step t
$X_{a,t}$	t-th reward obtained from playing arm a
$X_{u,t}$	reward obtained by user u during time step t
R_T	regret after T time steps
$R_{\pi\nu}$	regret of running π on instance ν for T time steps
T	time horizon
$T_{a,t}$	number of time steps arm a is played up till time step t
U	number of users
$\mathcal{E}_b$	event that μ_a is close to $\hat{\mu}_a$ for all $a \in [K]$
$\mathcal{F}_b$	event that A_b is sub-optimal but "not too bad"
$\mathcal{G}_{b,i}$	sub-event of $\mathcal{F}_b$ used in the regret analysis
$\mathcal{G}_{b,i,a}$	arm-dependent variant of $\mathcal{G}_{b,i}$
$\mathcal{H}$	sets in $\mathbf{\Gamma}$ are played for at most $B/2$ times
$\mathcal{V}$	set of all 1-subgaussian `EgalMAB` instances
$\mathbb{P}_a$	probability law of reward for arm a
$\mathbb{P}_{\pi\nu}$	probability law for the interaction between π and ν
Δ_A	difference between μ_* and μ_A
$\Delta_{\max}$	difference between μ_U and μ_{U+1}
$\Delta_{\min}$	difference between $\mu_1 + \cdots + \mu_U$ and $\mu_{K-U+1} + \cdots + \mu_K$
Λ	set of arms $[K] \backslash A_*$
α	technical constant used for the proof
β	technical constant used for the proof
$\epsilon_{b,b'}$	confidence radius of playing b' blocks after block b
γ	technical constant used for the proof
λ	Lebesgue measure
μ_a	expectation for distribution $\mathbb{P}_a$
μ_A	sum of the expected reward over the arms in A
μ_*	sum of the expected reward over the top U arms
$\hat{\mu}_{a,b}$	empirical estimate of μ_a after playing a for b blocks
ν	`EgalMAB` instance
π	policy for `EgalMAB`
ρ	counting measure

B Detailed Algorithm

Algorithm 2: `EgalUCB` with implementation details

initialize current block $b = 0$
initialize current time step $t = 0$
foreach $a \in [K]$ **do**
    let number of blocks $B_{a,b} = 0$
    let cumulative reward $S_{a,t} = 0$
    let upper confidence bound $\mathrm{UCB}_{a,b} = \infty$
end
while $b \leq T/U$ **do**
    update $b = b + 1$
    let $A_b \subseteq_U [K]$ be a set of U arms with highest $\mathrm{UCB}_{a,b-1}$
    let $\mathrm{ind} = (1, \ldots, U)$
    foreach $i \in [U]$ **do**
        update $t = t + 1$
        foreach $u \in [U]$ **do**
            let $A_{u,t} = A_b[\mathrm{ind}[u]]$
        end
        play $(A_{1,t}, \ldots, A_{U,t})$ and receive $(X_{1,t}, \ldots, X_{U,t})$
        foreach $u \in [U]$ **do**
            let $a = A_b[\mathrm{ind}[u]]$
            let $S_{a,t} = S_{a,t-1} + X_{u,t}$
        end
        circular shift ind by one to the right
    end
    foreach $a \in A_b$ **do**
        let $B_{a,b} = B_{a,b-1} + 1$
    end
    foreach $a \in [K]$ **do**
        let $\mathrm{UCB}_{a,b} = \dfrac{S_{a,t}}{B_{a,b}U} + \sqrt{\dfrac{6\ln(bU)}{B_{a,b}U}}$
    end
end

C Proofs of Upper Bounds

This section contains the proof for the upper bounds.

Lemma 1. *Let (ν, T, U) be a 1-subgaussian `EgalMAB`. Then, for all blocks $b \in [B]$,*

$$\mathbb{P}(\mathcal{E}_b^{\mathsf{c}}) \leq \frac{2K}{b^2 U^3}.$$

Proof. Since the rewards are 1-subgaussian, we have

$$\mathbb{P}\big(|\hat{\mu}_{a,b'} - \mu_a| > \epsilon_{b,b'}\big) \leq 2 \exp\left(-\frac{1}{2} b' U \epsilon_{b,b'}^2\right)$$

for any $b' \in [b]$ due to Chernoff's bound. Then, by applying the union bound over all arms and all possible values of $B_{a,b-1}$, we have

$$\begin{aligned}
\mathbb{P}(\mathcal{E}_b^{\mathsf{c}}) &\leq \sum_{a=1}^{K} \sum_{b'=1}^{b} \mathbb{P}\big(|\hat{\mu}_{a,b'} - \mu_a| > \epsilon_{b,b'}\big) \\
&\leq \sum_{a=1}^{K} \sum_{b'=1}^{b} 2 \exp\left(-\frac{1}{2} b' U \epsilon_{b,b'}^2\right) \\
&\leq \sum_{a=1}^{K} \sum_{b'=1}^{b} \frac{2}{(bU)^3} \\
&\leq \frac{2K}{b^2 U^3}.
\end{aligned}$$

□

Lemma 2. *Let $b \in [B]$. If the set of arms A_b played during block b is sub-optimal and $\mathcal{E}_b$ occurs, then $\mathcal{F}_b$ also occurs.*

Proof. Since A_b is assumed to be sub-optimal, we already have $\Delta_{A_b} > 0$. Denote A_*^- to be the set $A_* \backslash A_b$. Observe that

$$\Delta_{A_b} = \sum_{a \in A_*} \mu_a - \sum_{a \in A_b} \mu_a = \sum_{a \in A_*^-} \mu_a - \sum_{a \in A_b^-} \mu_a.$$

since the terms that are associated to arms in $A_b \cap A_*$ cancel out. Furthermore, since `EgalUCB` chooses A_b instead of A_*, we have

$$\sum_{a \in A_b^-} \hat{\mu}_{a,B_{a,b-1}} + \epsilon_{b-1,B_{a,b-1}} \geq \sum_{a \in A_*^-} \hat{\mu}_{a,B_{a,b-1}} + \epsilon_{b-1,B_{a,b-1}}.$$

Using these observations, we have

$$\begin{aligned}
\sum_{a \in A_b^-} \mu_a + 2\epsilon_{b-1,B_{a,b-1}} &\geq \sum_{a \in A_b^-} \hat{\mu}_{a,B_{a,b-1}} + \epsilon_{b-1,B_{a,b-1}} \\
&\geq \sum_{a \in A_*^-} \hat{\mu}_{a,B_{a,b-1}} + \epsilon_{b-1,B_{a,b-1}} \\
&\geq \sum_{a \in A_*^-} \mu_a
\end{aligned}$$

where the first and last inequality holds due to the assumption that $\mathcal{E}_b$ occurs. Rearranging this, we have

$$\begin{aligned}\Delta_{A_b} = \sum_{a \in A_*^-} \mu_a - \sum_{a \in A_b^-} \mu_a &\leq 2 \sum_{a \in A_b^-} \epsilon_{b-1, B_{a,b-1}} \\ &\leq 2 \sum_{a \in A_b^-} \epsilon_{B, B_{a,b-1}}\end{aligned}$$

where the last inequality holds because $b - 1 \leq B$. □

Lemma 3. *Let (ν, T, U) be a 1-subgaussian* `EgalMAB`*. Then, after T time steps, for all users $u \in [U]$, we have*

$$R_{u,T} \leq \sum_{b=1}^{B} \mathbb{E}[\Delta_{A_b} \mathbb{I}\{\mathcal{F}_b\}] + \frac{\pi^2 K \Delta_{\max}}{3U^3}.$$

Proof. We begin by decomposing $R_{u,T}$ into

$$R_{u,T} = \frac{T\mu_*}{U} - \mathbb{E}[S_{u,T}] = \sum_{b=1}^{B} \mu_* - \mathbb{E}[\mu_{A_b}] = \sum_{b=1}^{B} \mathbb{E}[\Delta_{A_b}].$$

Since $\mathbb{I}\{\mathcal{E}\} + \mathbb{I}\{\mathcal{E}^{\mathsf{c}}\} = 1$ almost surely for any event $\mathcal{E}$, we can split

$$\sum_{b=1}^{B} \mathbb{E}[\Delta_{A_b}] = \underbrace{\sum_{b=1}^{B} \mathbb{E}[\Delta_{A_b} \mathbb{I}\{\mathcal{E}_b\}]}_{(\spadesuit)} + \underbrace{\sum_{b=1}^{B} \mathbb{E}[\Delta_{A_b} \mathbb{I}\{\mathcal{E}_b^{\mathsf{c}}\}]}_{(\clubsuit)}.$$

To bound $(\spadesuit)$, we use $\mathbb{I}\{\Delta_{A_b} = 0\} + \mathbb{I}\{\Delta_{A_b} > 0\} = 1$ to get

$$\sum_{b=1}^{B} \mathbb{E}[\Delta_{A_b} \mathbb{I}\{\mathcal{E}_b\} \mathbb{I}\{\Delta_{A_b} = 0\}] + \sum_{b=1}^{B} \mathbb{E}[\Delta_{A_b} \mathbb{I}\{\mathcal{E}_b\} \mathbb{I}\{\Delta_{A_b} > 0\}].$$

Since $\Delta_{A_b} \mathbb{I}\{\Delta_{A_b} = 0\} = 0$ almost surely, the first term is 0. To deal with the second term, observe for any events $\mathcal{E}, \mathcal{E}', \mathcal{F}$, if $\mathcal{E}$ and $\mathcal{E}'$ implies $\mathcal{F}$, then $\mathbb{I}\{\mathcal{E}_1 \cap \mathcal{E}_2\} = \mathbb{I}\{\mathcal{E}_1\}\mathbb{I}\{\mathcal{E}_2\} \leq \mathbb{I}\{\mathcal{F}\}$ almost surely. As such, we have

$$\sum_{b=1}^{B} \mathbb{E}[\Delta_{A_b} \mathbb{I}\{\mathcal{E}_b\} \mathbb{I}\{\Delta_{A_b} > 0\}] \leq \sum_{b=1}^{B} \mathbb{E}[\Delta_{A_b} \mathbb{I}\{\mathcal{F}_b\}]$$

by Lemma 2, thus concluding the proof for $(\spadesuit)$. To bound $(\clubsuit)$, observe that

$$\sum_{b=1}^{B} \mathbb{E}[\Delta_{A_b} \mathbb{I}\{\mathcal{E}_b^{\mathsf{c}}\}] = \sum_{b=1}^{B} \mathbb{E}[\Delta_{A_b} | \mathcal{E}_b^{\mathsf{c}}] \mathbb{P}(\mathcal{E}_b^{\mathsf{c}}).$$

The expectation term can be bounded by

$$\begin{aligned}\mathbb{E}[\Delta_{A_b} | \mathcal{E}_b^{\mathsf{c}}] &= \sum_{A \subseteq_U [K]} \mathbb{P}(A_b = A | \mathcal{E}_b^{\mathsf{c}}) \, \Delta_A \\ &\leq \sum_{A \subseteq_U [K]} \mathbb{P}(A_b = A | \mathcal{E}_b^{\mathsf{c}}) \, \Delta_{\max} \\ &= \Delta_{\max}.\end{aligned}$$

Furthermore, we know that $\mathbb{P}(\mathcal{E}_b^{\mathsf{c}}) \le 2K/b^2U^3$ from Lemma 1. Substituting these results back, we have

$$\begin{aligned}\sum_{b=1}^{B} \mathbb{E}[\Delta_{A_b} | \mathcal{E}_b^{\mathsf{c}}] \mathbb{P}(\mathcal{E}_b^{\mathsf{c}}) &\le \sum_{b=1}^{B} \Delta_{\max} \cdot \frac{2K}{b^2U^3} \\ &\le \frac{2K\Delta_{\max}}{U^3} \sum_{b=1}^{\infty} \frac{1}{b^2} \\ &= \frac{\pi^2 K \Delta_{\max}}{3U^3},\end{aligned}$$

thus concluding the proof for (♣). □

Lemma 4. *Assume that $b > K/U$. On the event $\mathcal{F}_b$, exactly one of the events in $\{\mathcal{G}_{b,i}\}_i$ occurs.*

Proof. It is clear by definition that at most one of $\{\mathcal{G}_{b,i}\}_i$ can happen. We are left to show that at least one of $\{\mathcal{G}_{b,i}\}_i$ must happen. Suppose that none of $\{\mathcal{G}_{b,i}\}_i$ happens. Thus $|L_{b,i}| < \beta^i U$ for all $i \in \mathbb{N}$. First, we claim that all arms $a \in A_b^-$ are played at least once after block K/U. To see this, observe that the radius $\epsilon_{a,b} = \infty$ until a is played; and after which $\epsilon_{a,b} < \infty$. This claim implies that there exist some sufficiently large $j \in \mathbb{N}$ such that the set $L_{b,j} = \emptyset$. Moreover, since $\{L_{b,i}\}_i$ is a non-increasing sequence of sets, we have that for all $b > K/U$, all arms $a \in A_b^-$ must lie in exactly one $L_{b,i-1} \backslash L_{b,i}$. Thus, we have

$$\begin{aligned}\sum_{a \in A_b^-} \frac{1}{\sqrt{B_{a,b-1}}} &= \sum_{i=1}^{\infty} \sum_{a \in L_{b,i-1} \backslash L_{b,i}} \frac{1}{\sqrt{B_{a,b-1}}} \\ &< \sum_{i=1}^{\infty} \sum_{a \in L_{b,i-1} \backslash L_{b,i}} \frac{1}{\sqrt{m_{b,i}}} \\ &= \sum_{i=1}^{\infty} \frac{|L_{b,i-1}| - |L_{b,i}|}{\sqrt{m_{b,i}}} \\ &= \frac{|L_{b,0}|}{\sqrt{m_{b,1}}} + \sum_{i=1}^{\infty} |L_{b,i}| \cdot \left(\frac{1}{\sqrt{m_{b,i+1}}} - \frac{1}{\sqrt{m_{b,i}}} \right) \\ &< \frac{\beta^0 U}{\sqrt{m_{b,1}}} + \sum_{i=1}^{\infty} \beta^i U \cdot \left(\frac{1}{\sqrt{m_{b,i+1}}} - \frac{1}{\sqrt{m_{b,i}}} \right) \\ &= U \sum_{i=1}^{\infty} \frac{\beta^{i-1} - \beta^i}{\sqrt{m_{b,i}}}.\end{aligned}$$

Substituting $m_{b,i}$ into the inequality, we have

$$U \sum_{i=1}^{\infty} \frac{\beta^{i-1} - \beta^i}{\sqrt{m_{b,i}}} = U \sum_{i=1}^{\infty} \frac{\beta^{i-1} - \beta^i}{\sqrt{\gamma \alpha^i U \ln(BU)/\Delta_{A_b}^2}}$$

Rearranging the terms and evaluating the constants, we have

$$\begin{aligned}U \sum_{i=1}^{\infty} \frac{\beta^{i-1} - \beta^i}{\sqrt{\gamma \alpha^i U \ln(BU)/\Delta_{A_b}^2}} &= \frac{1-\beta}{\beta} \sqrt{\frac{U}{\gamma \ln(BU)}} \sum_{i=1}^{\infty} \left(\frac{\beta}{\sqrt{\alpha}} \right)^i \cdot \Delta_{A_b} \\ &< \sqrt{\frac{U}{24 \ln(BU)}} \cdot \Delta_{A_b}.\end{aligned}$$

Since $\mathcal{F}_b$ happens, we have

$$\Delta_{A_b} \le \sum_{a \in A_b^-} \sqrt{\frac{24 \ln(BU)}{B_{a,b-1} U}} < \Delta_{A_b}$$

which is a contradiction. Hence, at least one of the events $\{\mathcal{G}_{b,i}\}_i$ must happen. ☐

Lemma 5. *Let $\nu = (p_1, \ldots, p_K)$. Suppose that p_a is the density for a 1-subgaussian distribution for all $a \in [K]$. Then, after T time steps,*

$$\begin{aligned}\sum_{b=1}^{B} \mathbb{E}[\Delta_{A_b}\mathbb{I}\{\mathcal{F}_b\}] &= \sum_{i=1}^{\infty}\sum_{b=b_0}^{B} \Delta_{A_b}\mathbb{I}\{\mathcal{G}_{b,i}\} + \sum_{b=1}^{b_0-1} \Delta_{A_b}\mathbb{I}\{\mathcal{F}_b\} \\ &\leq 2136\ln(BU)\sum_{a\in\Lambda}\frac{1}{\Delta_{a,N_a}} + \frac{K\Delta_{\max}}{U} \\ &\leq \frac{2136(K-U)\ln(BU)}{\Delta_{\min}} + \frac{K\Delta_{\max}}{U}.\end{aligned}$$

for all users $u \in [U]$.

Proof. For each arm $a \in [K]$, each index $i \in \mathbb{N}$, and each block $b \in [B]$, let

$$\mathcal{G}_{b,i,a} = \mathcal{G}_{b,i} \cap \left\{a \in A_b^-\right\} \cap \left\{B_{a,b-1} \leq m_{b,i}\right\}$$

be an arm-dependent variant of the event $\mathcal{G}_{b,i}$. Since at least $\beta^i U$ arms satisfy $\mathcal{G}_{b,i,a}$ when $\mathcal{G}_{b,i}$ occurs, we have

$$\mathbb{I}\{\mathcal{G}_{b,i}\} \leq \frac{1}{\beta^i U}\sum_{a\in\Lambda}\mathbb{I}\{\mathcal{G}_{b,i,a}\}$$

where $\Lambda \coloneqq [K]\backslash A_*$ is the set of arms that are not in A_*. For each arm $a \in \Lambda$, let

$$N_a \coloneqq \left|\{A_b : b \in [B] \text{ and } a \in A_b^- \text{ and } \Delta_{A_b} > 0\}\right|$$

be the number of distinct sub-optimal sets played that includes a, and let $A_{a,1}, \ldots, A_{a,N_a}$ be these sets sorted by non-ascending order of its sub-optimality gap. In other words, if we denote

$$\Delta_{a,j} \coloneqq \Delta_{A_{a,j}},$$

then $\Delta_{a,1} \geq \cdots \geq \Delta_{a,N_a}$. Let $b_0 = K/U + 1$. Almost surely we have

$$\begin{aligned}\sum_{b=b_0}^{B}\Delta_{A_b}\mathbb{I}\{\mathcal{G}_{b,i}\} &\leq \sum_{a\in\Lambda}\sum_{b=b_0}^{B}\frac{\Delta_{A_b}}{\beta^i U}\mathbb{I}\{\mathcal{G}_{b,i,a}\} \\ &= \sum_{a\in\Lambda}\sum_{b=b_0}^{B}\sum_{j=1}^{N_a}\frac{\Delta_{a,j}}{\beta^i U}\mathbb{I}\{\mathcal{G}_{b,i,a}\}\mathbb{I}\{A_b = A_{a,j}\} \\ &= \sum_{a\in\Lambda}\sum_{j=1}^{N_a}\frac{\Delta_{a,j}}{\beta^i U}\sum_{b=b_0}^{B}\mathbb{I}\{\mathcal{G}_{b,i,a}\}\mathbb{I}\{A_b = A_{a,j}\}\end{aligned}$$

We can upper bound this expression by considering the worst-case realization of the number of blocks each set $A_{a,j}$ is played. Let us start with $j = 1$. Since $A_{a,1}$ has the largest gap $\Delta_{a,1}$, the worst case realization is when $A_{a,1}$ is played as many times as the event $\mathcal{G}_{b,i,a}$ allows. Recall that $\mathcal{G}_{b,i,a}$ implies that $B_{a,b-1} \leq m_{b,i}$. It follows that we can play $A_{a,1}$ for at most $\gamma\alpha^i U\ln(BU)/\Delta_{a,1}^2$ blocks. We can use this argument to find the worst-case realization on the number of blocks $A_{a,2}$ is played. This works out to be

$$\frac{\gamma\alpha^i U\ln(BU)}{\Delta_{a,2}^2} - \frac{\gamma\alpha^i U\ln(BU)}{\Delta_{a,1}^2} = \gamma\alpha^i U\ln(BU)\cdot\left(\frac{1}{\Delta_{a,2}^2} - \frac{1}{\Delta_{a,1}^2}\right).$$

Repeating this argument, we have

$$\begin{aligned}&\sum_{a\in\Lambda}\sum_{j=1}^{N_a}\frac{\Delta_{a,j}}{\beta^i U}\sum_{b=b_0}^{B}\mathbb{I}\{\mathcal{G}_{b,i,a}\}\mathbb{I}\{A_b = A_{a,j}\} \\ &\leq \sum_{a\in\Lambda}\frac{\gamma\alpha^i\ln(BU)}{\beta^i}\cdot\left(\frac{1}{\Delta_{a,1}} + \sum_{j=2}^{N_a}\Delta_{a,j}\left(\frac{1}{\Delta_{a,j}^2} - \frac{1}{\Delta_{a,j-1}^2}\right)\right).\end{aligned}$$

The terms within the bracket can be further bounded by

$$\begin{aligned}
&\frac{1}{\Delta_{a,1}} + \sum_{j=2}^{N_a} \Delta_{a,j} \left(\frac{1}{\Delta_{a,j}^2} - \frac{1}{\Delta_{a,j-1}^2} \right) \\
&= \frac{1}{\Delta_{a,N_a}} + \sum_{j=1}^{N_a-1} \frac{\Delta_{a,j} - \Delta_{a,j+1}}{\Delta_{a,j}^2} \\
&\leq \frac{1}{\Delta_{a,N_a}} + \sum_{j=1}^{N_a-1} \frac{\Delta_{a,j} - \Delta_{a,j+1}}{\Delta_{a,j} \cdot \Delta_{a,j+1}} \\
&= \frac{1}{\Delta_{a,N_a}} + \sum_{j=1}^{N_a-1} \left(\frac{1}{\Delta_{a,j+1}} - \frac{1}{\Delta_{a,j}} \right) \\
&< \frac{2}{\Delta_{a,N_a}}.
\end{aligned}$$

By combining these results, we have, almost surely, that

$$\begin{aligned}
\sum_{b=1}^{B} \Delta_{A_b} \mathbb{I}\{\mathcal{F}_b\} &= \sum_{b=1}^{b_0-1} \Delta_{A_b} \mathbb{I}\{\mathcal{F}_b\} + \sum_{b=b_0}^{B} \Delta_{A_b} \mathbb{I}\{\mathcal{F}_b\} \\
&\leq \frac{K\Delta_{\max}}{U} + \sum_{i=1}^{\infty} \sum_{b=b_0}^{B} \Delta_{A_b} \mathbb{I}\{\mathcal{G}_{b,i}\} \\
&\leq \frac{K\Delta_{\max}}{U} + \sum_{i=1}^{\infty} \sum_{a\in\Lambda} \frac{2\gamma\alpha^i \ln(BU)}{\beta^i \Delta_{a,N_a}} \\
&\leq \frac{K\Delta_{\max}}{U} + 2\gamma \ln(BU) \sum_{a\in\Lambda} \frac{1}{\Delta_{a,N_a}} \sum_{i=1}^{\infty} \left(\frac{\alpha}{\beta}\right)^i \\
&< \frac{K\Delta_{\max}}{U} + 2136 \ln(BU) \sum_{a\in\Lambda} \frac{1}{\Delta_{a,N_a}} \\
&\leq \frac{K\Delta_{\max}}{U} + \frac{2136(K-U)\ln(BU)}{\Delta_{\min}}
\end{aligned}$$

where the last inequality holds because for all arms $a \in \Lambda$, we have $\Delta_{a,N_a} \geq \mu_U - \mu_a \geq \Delta_{\min}$. As such,

$$\sum_{a\in\Lambda} \frac{1}{\Delta_{a,N_a}} \leq \frac{K-U}{\Delta_{\min}}. \qquad \square$$

Note. Observe that when $U = 1$, we have $\Delta_{a,N_a} = \Delta_a$. As such, we can replace the last inequality using

$$\ln(BU) \sum_{a\in\Lambda} \frac{1}{\Delta_{a,N_a}} = \sum_{a:\Delta_a>0} \frac{\ln(T)}{\Delta_a}$$

to obtain the bound for the classic `UCB1` algorithm.

Theorem 1 (Problem-Dependent Upper Bound)**.** *Let (ν, T, U) a 1-subgaussian `EgalMAB`. After running `EgalUCB` for T time steps, we have*

$$R_T \leq \frac{2136(K-U)\ln(T)}{\Delta_{\min}} + \frac{4K\Delta_{\max}}{U}.$$

Proof. We can bound the regret by

$$\begin{aligned}
R_{u,T} &\leq \sum_{b=1}^{B} \mathbb{E}[\Delta_{A_b} \mathbb{I}\{\mathcal{F}_b\}] + \frac{\pi^2 K \Delta_{\max}}{3U^3} \\
&\leq \frac{2136(K-U)\ln(BU)}{\Delta_{\min}} + \frac{K\Delta_{\max}}{U} + \frac{\pi^2 K \Delta_{\max}}{3U^3} \\
&\leq \frac{2136(K-U)\ln(T)}{\Delta_{\min}} + \frac{4K\Delta_{\max}}{U}
\end{aligned}$$

where the first inequality holds due to Lemma 3 and the second inequality holds due to Lemma 5. □

Theorem 2 (Problem-Independent Upper Bound)**.** *Let (ν, T, U) be a 1-subgaussian `EgalMAB` with $\mu_a \in [0,1]$ for all arms $a \in [K]$. After running `EgalUCB` for T time steps, we have*

$$R_T \leq \sqrt{\frac{8544(K-U)\,T\ln(T)}{U}} + \frac{4K\min\{U, K-U\}}{U}.$$

Proof. Set

$$\delta = \sqrt{\frac{2136(K-U)\ln(BU)}{B}}.$$

Since $\mathbb{I}\{\Delta_{A_b} < \delta\} + \mathbb{I}\{\Delta_{A_b} \geq \delta\} = 1$, we have

$$\begin{aligned}
\sum_{b=b_0}^{B} \mathbb{E}[\Delta_{A_b}\mathbb{I}\{\mathcal{F}_b\}] = &\sum_{b=b_0}^{B} \mathbb{E}[\Delta_{A_b}\mathbb{I}\{\mathcal{F}_b\}\mathbb{I}\{\Delta_{A_b} < \delta\}] \\
&+ \sum_{b=b_0}^{B} \mathbb{E}[\Delta_{A_b}\mathbb{I}\{\mathcal{F}_b\}\mathbb{I}\{\Delta_{A_b} \geq \delta\}].
\end{aligned}$$

We can trivially bound $\mathbb{E}[\Delta_{A_b}\mathbb{I}\{\mathcal{F}_b\}\mathbb{I}\{\Delta_{A_b} < \delta\}]$ by δ. The second term can be bounded similarly as in Theorem 5 by using $\Delta_{A_b} \geq \delta$ instead of $\Delta_{A_b} \geq \Delta_{\min}$. Then, we have

$$\begin{aligned}
\sum_{b=b_0}^{B} \mathbb{E}[\Delta_{A_b}\mathbb{I}\{\mathcal{F}_b\}] &\leq B\delta + \frac{2136(K-U)\ln(BU)}{\delta} \\
&= \sqrt{8544(K-U)\cdot B\ln(BU)}.
\end{aligned}$$

Furthermore, since $\mu_a \in [0,1]$ for all $a \in [K]$, we have

$$\Delta_{\max} \leq \min\{U, K-U\}.$$

To understand this bound, note that $\Delta_{\max} \leq U$ generally, but when $K < 2U$, this can be tightened because there are overlaps between the top U arms and the bottom U arms. This works out to $\Delta_{\max} \leq K - U$. By considering the remaining terms in the regret, we have

$$\begin{aligned}
R_T &\leq \sqrt{8544(K-U)\cdot B\ln(BU)} + \frac{4K\Delta_{\max}}{U} \\
&\leq \sqrt{\frac{8544(K-U)\cdot T\ln(T)}{U}} + \frac{4K\min\{U, K-U\}}{U}.
\end{aligned}$$ □

D Proofs of Lower Bound

Lemma 6. *Let ν and ν' be the `EgalMAB` instances defined by equation 1 and equation 2. Under the assumptions of Theorem 3, we have*

$$\begin{aligned} R_{\pi\nu} + R_{\pi\nu'} &> \frac{\Delta T}{4}\big(\mathbb{P}_{\pi\nu}(\mathcal{H}) + \mathbb{P}_{\pi\nu'}(\mathcal{H}^{\mathrm{c}})\big) \\ &\geq \frac{\Delta T}{8}\exp\big(-D_{\mathrm{KL}}(\mathbb{P}_{\pi\nu}\|\mathbb{P}_{\pi\nu'})\big). \end{aligned}$$

Proof. Note that whenever we play some $A \notin \mathbf{\Gamma}$ under ν, there will be at least $U/2$ users who will incur an instantaneous regret of at least Δ. Under $\mathcal{H}$, the total number of sub-optimal arms played across all users and all time steps is at least $TU/4$. By the pigeonhole principle, we know that at least one user played sub-optimal arms for at least $T/4$ times. As such, the regret is at least $\Delta T/4$. A similar argument can be used to show that under ν' and $\mathcal{H}^{\mathrm{c}}$, the regret is at least $\Delta T/4$. Thus

$$\begin{aligned} R_{T,\pi,\nu} + R_{T,\pi,\nu'} &> \frac{\Delta T}{4}\big(\mathbb{P}_{\pi\nu}(\mathcal{H}) + \mathbb{P}_{\pi\nu'}(\mathcal{H}^{\mathrm{c}})\big) \\ &\geq \frac{\Delta T}{8}\exp\big(-D_{\mathrm{KL}}(\mathbb{P}_{\pi\nu}\|\mathbb{P}_{\pi\nu'})\big) \end{aligned}$$

where the last inequality holds due to the Bretagnolle–Huber inequality. ☐

Lemma 7. *Let ν and ν' be the `EgalMAB` instances defined by equation 1 and equation 2. Under the assumptions of Theorem 3, we have*

$$D_{\mathrm{KL}}(\mathbb{P}_{\pi\nu}\|\mathbb{P}_{\pi\nu'}) \leq 4\Delta^2 \sum_{a'\in A'} \mathbb{E}_{\pi\nu}[T_{a',T}].$$

Proof. Using the definition of the KL-divergence and applying the chain rule, we have

$$\begin{aligned} D_{\mathrm{KL}}(\mathbb{P}_{\pi\nu}\|\mathbb{P}_{\pi\nu'}) &= \mathbb{E}_{\pi\nu}\left[\ln\left(\frac{\mathrm{d}\mathbb{P}_{\pi\nu}}{\mathrm{d}\mathbb{P}_{\pi\nu'}}\right)\right] \\ &= \mathbb{E}_{\pi\nu}\left[\ln\left(\frac{\mathrm{d}\mathbb{P}_{\pi\nu}/\mathrm{d}(\rho^U\times\lambda^U)^T}{\mathrm{d}\mathbb{P}_{\pi\nu'}/\mathrm{d}(\rho^U\times\lambda^U)^T}\right)\right]. \end{aligned}$$

We then substitute the Radon-Nikodym derivatives and simplify the terms to get

$$
\begin{aligned}
&\mathbb{E}_{\pi\nu}\left[\ln\left(\frac{\mathrm{d}\mathbb{P}_{\pi\nu}/\mathrm{d}(\rho^U \times \lambda^U)^T}{\mathrm{d}\mathbb{P}_{\pi\nu'}/\mathrm{d}(\rho^U \times \lambda^U)^T}\right)\right] \\
&= \mathbb{E}_{\pi\nu}\left[\ln \prod_{t=1}^{T} \frac{\pi(A_t|A_1, X_1, \ldots, A_{t-1}, X_{t-1})}{\pi(A_t|A_1, X_1, \ldots, A_{t-1}, X_{t-1})} \prod_{u=1}^{U} \frac{p_{A_{u,t}}(X_{u,t})}{p'_{A_{u,t}}(X_{u,t})}\right] \\
&= \mathbb{E}_{\pi\nu}\left[\sum_{t=1}^{T}\sum_{u=1}^{U} \ln \frac{p_{A_{u,t}}(X_{u,t})}{p'_{A_{u,t}}(X_{u,t})}\right] \\
&= \sum_{t=1}^{T}\sum_{u=1}^{U} \mathbb{E}_{\pi\nu}\left[\mathbb{E}_{\pi\nu}\left[\ln \frac{p_{A_{u,t}}(X_{u,t})}{p'_{A_{u,t}}(X_{u,t})} \,\middle|\, A_{u,t}\right]\right] \\
&= \sum_{t=1}^{T}\sum_{u=1}^{U} \mathbb{E}_{\pi\nu}\left[D_{\mathrm{KL}}(\mathbb{P}_{A_{u,t}} \| \mathbb{P}'_{A_{u,t}})\right] \\
&= \sum_{t=1}^{T} \mathbb{E}_{\pi\nu}\left[\sum_{a \in A_t} D_{\mathrm{KL}}(\mathbb{P}_a \| \mathbb{P}'_a)\right] \\
&= \sum_{t=1}^{T} \mathbb{E}_{\pi\nu}\left[\sum_{A \subseteq_U [K]} \sum_{a \in A} D_{\mathrm{KL}}(\mathbb{P}_a \| \mathbb{P}'_a) \cdot \mathbb{I}[A_t = A]\right] \\
&= \sum_{A \subseteq_U [K]} \sum_{a \in A} D_{\mathrm{KL}}(\mathbb{P}_a \| \mathbb{P}'_a) \sum_{t=1}^{T} \mathbb{E}_{\pi\nu}\left[\mathbb{I}[A_t = A]\right] \\
&= \sum_{A \subseteq_U [K]} \sum_{a \in A} D_{\mathrm{KL}}(\mathbb{P}_a \| \mathbb{P}'_a) \cdot \mathbb{E}_{\pi\nu}[T_{A,T}]
\end{aligned}
$$

Note that the KL-divergence between two Gaussian measures with mean μ_1 and μ_2 and variance 1 is $(\mu_1 - \mu_2)^2$. Thus, we have

$$
\begin{aligned}
&\sum_{A \subseteq_U [K]} \sum_{a \in A} D_{\mathrm{KL}}(\mathbb{P}_a \| \mathbb{P}'_a) \cdot \mathbb{E}_{\pi\nu}[T_{A,T}] \\
&= \sum_{A \subseteq_U [K]} \mathbb{E}_{\pi\nu}[T_{A,T}] \sum_{a \in A} D_{\mathrm{KL}}(\mathbb{P}_a \| \mathbb{P}'_a) \\
&= \sum_{A \subseteq_U [K]} \mathbb{E}_{\pi\nu}[T_{A,T}] \cdot 4\Delta^2 |\{a' \in A' | a' \in A\}| \\
&= 4\Delta^2 \sum_{a' \in A'} \sum_{A : a' \in A} \mathbb{E}_{\pi\nu}[T_{A,T}] \\
&= 4\Delta^2 \sum_{a' \in A'} \mathbb{E}_{\pi\nu}[T_{a',T}].
\end{aligned}
$$

□

Lemma 8. *Let ν and ν' be the `EgalMAB` instances defined in equation 1 and equation 2. Under the assumptions of Theorem 3, we have*

$$\sum_{a \in A'} T_{a,T} \leq \frac{TU^2}{K - U}$$

almost surely.

Proof. Suppose, for sake of contradiction, that

$$\sum_{a \in A'} T_{a,T} > \frac{TU^2}{K-U}.$$

Note that since A' is the set of least played arms, we have

$$\begin{aligned}\sum_{A \subseteq_U [K]\setminus[U]} \sum_{a \in A} T_{a,T} &> \sum_{A \subseteq_U [K]\setminus[U]} \frac{TU^2}{K-U} \\ &= \binom{K-U}{U} \frac{TU^2}{K-U}.\end{aligned}$$

Furthermore, the same quantity can be upper bounded by

$$\begin{aligned}\sum_{A \subseteq_U [K]\setminus[U]} \sum_{a \in A} T_{a,T} &= \sum_{a \in [K]\setminus[U]} \sum_{A : a \in A} T_{a,T} \\ &= \sum_{a \in [K]\setminus[U]} \binom{K-U}{U} \frac{U}{K-U} T_{a,T} \\ &= \binom{K-U}{U} \frac{U}{K-U} \sum_{a \in [K]\setminus[U]} T_{a,T} \\ &\le \binom{K-U}{U} \frac{U}{K-U} \sum_{a \in [K]} T_{a,T} \\ &= \binom{K-U}{U} \frac{TU^2}{K-U},\end{aligned}$$

which is a contradiction. □

Theorem 3 (Policy-Independent Lower Bound)**.** *Suppose $K \ge 2U$. For any policy π, there exist an* `EgalMAB` *instance $\nu \in \mathcal{V}$ with regret*

$$R_{\pi\nu} \ge \frac{\sqrt{(K-U)T}}{76U}.$$

Proof. We have

$$\begin{aligned}R_{T,\pi,\nu} + R_{T,\pi,\nu'} &\ge \frac{\Delta T}{8} \exp\left(-4\Delta^2 \sum_{a' \in A'} \mathbb{E}_{\pi\nu}[T_{a',T}]\right) \\ &\ge \frac{\Delta T}{8} \exp\left(-\frac{4\Delta^2 TU^2}{K-U}\right) \\ &= \frac{\Delta T}{8} \exp(-1/2) \\ &> \frac{\sqrt{T(K-U)}}{38U}.\end{aligned}$$

Since $2\max\{R_{T,\pi,\nu}, R_{T,\pi,\nu'}\} \ge R_{T,\pi,\nu} + R_{T,\pi,\nu'}$, dividing by 2 concludes the proof. □