A. Proof of Lemma 2

The proof of Lemma 2 relies on the folloing lemma:

Lemma 7. *For positive measurable functions f_i, we have:*

$$\Big(\int\prod_{i=1}^{K}f_i(x)dx\Big)^K\leq\prod_{i=1}^{K}\int f_i^K(x)dx. \tag{4}$$

Proof. By Cauchy-Schwarz inequality, Eq. (4) is correct for $K=2$. We use mathematical induction to prove that it still holds for $K>2$. Specifically, assume InEq. (4) is true for $K-1$, considering case K, by Holder's inequality we have:

$$\begin{aligned}\Big(\int\prod_{i=1}^{K}f_i(x)dx\Big)^K=&\Big(\int f_K(x)\prod_{i=1}^{K-1}f_i(x)dx\Big)^K\\ \leq&\Big(\Big(\int f_K^K(x)dx\Big)^{\frac{1}{K}}\Big(\int(\prod_{i=1}^{K-1}f_i(x))^{\frac{K}{K-1}}dx\Big)^{\frac{K-1}{K}}\Big)^K\\ =&\Big(\int f_K^K(x)dx\Big)\Big(\int(\prod_{i=1}^{K-1}f_i(x))^{\frac{K}{K-1}}dx\Big)^{K-1}\\ \leq&\prod_{i=1}^{K}\int f_i^K(x)dx.\end{aligned}$$

□

Now we present the proof of Lemma 2.

Proof. With straight-forward computations, we have:

$$\sum_{k=1}^{K}P_k(\psi=k)=\sum_{k=1}^{K}\int_{\psi=k}dP_k=\sum_{k=1}^{K}\int_{\psi=k}\frac{dP_k}{dP_1}dP_1.$$

Because we need a bound for any measurable function ψ, we need to construct the ψ that minimizes the last expression. Obviously, the last expression is minimized for $\psi(x)=\arg\min_{k\leq K}\frac{dP_k}{dP_1}(x)$, so that

$$\sum_{k=1}^{K}P_k(\psi=k)\geq\int dP_1\min_k\frac{dP_k}{dP_1}=\int\min_k dP_k.$$

For vector $\bar{P}=\{dP_1,\cdots,dP_K\}$, define $r_i(\bar{P})$ be the i-th smallest value in $\bar{P}$, we have

$$\Big(\int(\prod_k dP_k)^{\frac{1}{K}}\Big)^K=\Big(\int(\prod_k r_k(\bar{P}))^{\frac{1}{K}}\Big)^K\leq\prod_k\int r_k(\bar{P})=\int\min_k dP_k\prod_{k\geq 2}\int r_k(\bar{P}).$$

The above inequality is proven by Lemma 7.Note that $\sum_k\int dP_k\leq K$, so $\prod_{k\geq 2}\int r_k(\bar{P})\leq(\frac{K}{K-1})^{K-1}<e$. With Jensen's inequality, it yields that:

$$\begin{aligned}\sum_{k=1}^{K}P_k(\psi=k)\geq&\frac{1}{e}\Big(\int(\prod_k dP_k)^{\frac{1}{K}}\Big)^K\\ =&\frac{1}{e}\exp\Big\{K\log\int(\prod_k dP_k)^{\frac{1}{K}}\Big\}\\ =&\frac{1}{e}\exp\Big\{K\log\int(\prod_k\frac{dP_k}{dP_1})^{\frac{1}{K}}dP_1\Big\}\\ \geq&\frac{1}{e}\exp\Big\{\sum_k\int dP_1\log\frac{dP_k}{dP_1}\Big\}\\ =&\frac{1}{e}\exp\Big\{-\sum_{k=2}^{K}KL(P_1,P_k)\Big\}.\end{aligned}$$

□

B. Proof of Lemma 3

Proof. With Hoeffding's inequality and InEq. (2) in the main text, we have

$$\begin{aligned}
P\left(i:\frac{1}{u_i'}\sum_{i=1}^{u_i'}Y_i^s-\mu_s>\beta(u_i,t_i)\right) &\leq \sum_{i\geq 1}P\left(\frac{1}{u_i'}\sum_{i=1}^{u_i'}Y_i^s-\mu_s>\beta(u_i,t_i)\right)\\
&\leq \sum_{i\geq 1}\exp\left\{-2u_i'\beta(u_i,t_i)^2\right\}\\
&\leq \sum_{i\geq 1}\exp\left\{-2u_i\beta(u_i,t_i)^2\right\}\\
&\leq \frac{\delta}{2K}.
\end{aligned}$$

Then, applying the union bound, we complete the proof. □

C. Proof of Lemma 4

In this section, we are going to show that $\mathbb{E}[\gamma - T(t_\gamma)] \leq O(H_c(j)\log\gamma)$ to complete the proof of Lemma 4. The proof is the same as that of Theorem 1 in (Auer et al., 2002), except some constants.

Proof. For simplicity, let s_1 denote $s_c^*(j)$. Considering $s:\mu_s<\mu_{s_1}$, let $ev(s,\gamma)$ be the event that s is pulled at line 22 by Alg. 1 at round γ. Let $T_s(t_\gamma)$ denote the number of pulls on arm s in line 22 Alg. 1 when Alg. 1 selects column j for the γ-th time. Then, $\mathbb{E}[T_s(t_\gamma)]$ can be bounded as follows:

$$\begin{aligned}
\mathbb{E}\left[T_s(t_\gamma)\right] &= \sum_{\gamma'=1}^{\gamma}\mathbb{1}[\text{The algorithm pulls arm } s \text{ in line 22 when Alg. 1 selects column } j \text{ for the } \gamma\text{-th round}]\\
&\leq l+\sum_{\gamma'=1}^{\gamma}\mathbb{1}\left[T_s(t_\gamma'-1)\geq l, ev(s,\gamma')\right]\\
&\leq l+\sum_{\gamma'=1}^{\gamma}\mathbb{1}\left[T_s(t_\gamma'-1)\geq l, \bar{\mu}_{s_1}+\sqrt{2\frac{\log t}{T_{s_1}(t_\gamma'-1)}}\leq \bar{\mu}_s+\sqrt{2\frac{\log\gamma'}{T_s(t_\gamma'-1)}}\right]\\
&\leq l+\sum_{\gamma'=1}^{\gamma}\mathbb{1}\left[T_s(t_\gamma'-1)\geq l, \min_{\gamma_1\in[1,\gamma'-1]}\bar{\mu}_{s_1}+\sqrt{2\frac{\log\gamma}{\gamma_1}}\leq \max_{\gamma_2\in[l,\gamma-1]}\bar{\mu}_s+\sqrt{2\frac{\log\gamma}{\gamma_2}}\right]\\
&\leq l+\sum_{\gamma'=1}^{\gamma}\sum_{\gamma_1=0}^{\gamma-1}\sum_{\gamma_2=l}^{\gamma-1}\mathbb{1}\left[\bar{\mu}_{s_1}+\sqrt{2\frac{\log\gamma'}{\gamma_1}}\leq\bar{\mu}_s+\sqrt{2\frac{\log\gamma'}{\gamma_2}}\right].
\end{aligned}$$

Since $\bar{\mu}_{s_1}+\sqrt{2\frac{\log\gamma'}{\gamma_1}}\leq\bar{\mu}_s+\sqrt{2\frac{\log\gamma'}{\gamma_2}}$, at least one of the following three events happens:

- $\bar{\mu}_{s_1}\leq\mu_{s_1}-\sqrt{2\frac{\log\gamma'}{\gamma_1}}$;
- $\bar{\mu}_s\leq\mu_s-\sqrt{2\frac{\log\gamma'}{\gamma_2}}$;
- $\mu_{s_1}\leq\mu_s+2\sqrt{2\frac{\log\gamma'}{\gamma_2}}$.

By Hoeffding's inequality, the probability that the first and the second events happen is at most $2\gamma'^{-4}$, and for $l\geq\frac{8}{(\mu_s-\mu_{s_1})^2}\log\gamma$, the third event will not happen. So we have $\mathbb{E}[T_s(t_\gamma)]=O\left(\frac{1}{(\mu_s-\mu_{s_1})^2}\log\gamma\right)$. Then, with straightforward computations we complete the proof. □

D. Proof of Lemma 5

Proof. Suppose $s^* \neq none$. If at round γ, Alg.1 does not select $R_{s^*[1]}$ or $C_{s^*[2]}$, then at least one of following events happens:

- $\exists s \notin \{none, s^*\}, s = NE(\bar{\mathbf{M}}_t)$;
- $NE(\bar{\mathbf{M}}_t) = none$.

For the first part, it is easy to see that $s \neq s_r^*(s[1])$ or $s \neq s_c^*(s[2])$. So by Lemma 4, the size of this part is at most $2(\sum_i \Lambda(H_r(i)) + \sum_j \lambda(H_c(j)))$.

Similarly, if the second case happens, then $s^* \neq \arg\min_{s' \in row(s^*)} \bar{\mu}_{s'}$ or $s^* \neq \arg\max_{s' \in col(s^*)} \bar{\mu}_{s'}$. When the second case happens for the γ-th round, due to Alg.1, we selected $R_{s^*[1]}$ for at least $\lfloor \gamma/m \rfloor$ times and $C_{s^*[2]}$ for at least $\lfloor \gamma/n \rfloor$ times. So by Lemma 4, the size of the second part is at most $\lceil (m+n)(\Lambda(H_r(s^*[1])) + \Lambda(H_c(s^*[2]))) \rceil$. □

E. Proof of Lemma 6

This proof is the same as the proof of Theorem 6 in Kalyanakrishnan et al. (2012), except the statement and some constants.

Proof. Without loss of generality, consider a bandit model $v = \{s_1, \cdots, s_k\}$. Suppose $\mu_{s_i} > \mu_{s_{i+1}}$. At each round, we pull arms in v as from line 21 to line 23 in Alg. 1. Let $H = \sum_{i=2} \frac{1}{(\mu_{s_1} - \mu_{s_i})^2}$. Let $s_1(\gamma) = \arg\max_{s \in v} \bar{\mu}_s$ after round γ, and $s_2(\gamma) = \arg\max_{s \in v \setminus s_1(\gamma)} U(s, |T_s(\tau)|, |\tau|)$. Denote $\tau(\gamma)$ be the set of time steps t in the first γ round. If after γ_v round, we have $L(s_1(\gamma), |T_{s_1(\gamma)}(\tau)|, |\tau|) > U(s_2(\gamma), |T_{s_2(\gamma)}(\tau)|, |\tau|)$, we now show that $\mathbb{E}\gamma_v = O(H_v \log \frac{H_v}{\delta})$.

Now let us introduce some notations. Define $c = \frac{\mu_{s_1} + \mu_{s_2}}{2}$ and

$$\Delta_i := \begin{cases} \mu_{s_1} - \mu_{s_2} & i = 1, \\ \mu_{s_1} - \mu_{s_i} & i \geq 2 \end{cases}$$

During round γ, we partition the set of arms into three subsets:

- $Above^\gamma := \{s \in v : \bar{\mu}_s - \beta(T_s(\gamma), \gamma) > c\}$.
- $Below^\gamma := \{s \in v : \bar{\mu}_s + \beta(T_s(\gamma), \gamma) < c\}$.
- $Middle^\gamma := v \setminus (Above^\gamma \cup Below^\gamma)$.

And by Lemma 2 in Kalyanakrishnan et al. (2012), if we cannot identify $\arg\max_{s \in v} \mu(s)$ after round γ, then $s_1(\gamma) \in Middle^\gamma$ or $s_2(\gamma) \in Middle^\gamma$. And with similar computation to Lemma 4 in Kalyanakrishnan et al. (2012), if for all arm s_i, we pulled at least $4\lceil \frac{1}{2\Delta_i} \ln \frac{mn\gamma^4}{4\delta} \rceil$ times, then $\forall s, s \notin Middle^\gamma$ with probability at least $1 - \frac{3\delta H_v}{mn\gamma^4}$. The rest is all the same as Theorem 6 in Kalyanakrishnan et al. (2012).

□

F. Proof of Theorem 5

Proof. With the same argument as Theorem 8 in Even-Dar et al. (2006), for s_1, s_2, suppose $\mu_{s_1} - \mu_{s_2} = \Delta > 0$. Then with probability at least $1 - \delta/(mn)$, after round $\gamma = O(\frac{\ln(mn/\delta\Delta)}{\Delta^2})$, we have $\bar{\mu}_{s_1} - \bar{\mu}_{s_2} > 2\beta_1(\gamma)$. So by the union bound, the racing algorithm stops after finite time with probability at least $1 - \delta$.

Then we show that Alg. 2 is δ-PAC.

According to Lemma 3, the probability that there is an arm s such that $|\bar{\mu}_s - \mu_s| > \beta_1(\gamma)$ for some γ is at most δ. Suppose $\forall s, \gamma, |\bar{\mu}_s - \mu_s| > \beta_1(\gamma)$.

For the case $NE(\mathbb{M}) = none$, consider arm s. If $s \notin \{s_r^*(s[1]), s_c^*(s[2])\}$, clearly, it will be eliminated when its confidence bounds are disjoint with the confidence bounds of $s_r^*(s[1])$ and $s_c^*(s[2])$. Otherwise, if $s \in \{s_r^*(s[1]), s_c^*(s[2])\}$, without loss of generality, suppose $s = s_r^*(s[1])$. Obviously, there is a sequence $S = \{s_1, s_2, \cdots, s_{2k}\}$ such that (1) $s = s_1$; (2) $s_{2i+1} = s_r^*(s_{2i+1}[1])$ and $s_{2i} = s_c^*(s_{2i}[2])$ for all i, where $s_j = s_{(j-1)\%(2k)+1}$; (3) $s_{2i+1} \in col(s_{2i})$ and $s_{2i+2} \in row(s_{2i+1})$. And according to Alg.2, all the arms in S will be eliminated when their confidence bounds are disjoint.

For the case $s^* = NE(\mathbb{M}) \neq none$, obviously, it won't be eliminated by our elimination rule, so Algorithm is δ-PAC.

Up to now we have proven the two statements about Alg. 2 in Theorem 5. □

G. Correctness of baseline

Here, We present the stopping and recommendation rules of our baseline algorithm in detail.

Stopping and recommendation rules: In each round, we pull all arms. Let $\beta_2(\gamma) = \sqrt{\log(5mn\gamma^2/4\delta)/\gamma}$. After the t-th round, if one of the following event happens, the algorithm stops:

- For some arm s, if for all $s' \in row(s)\backslash s, \exists\gamma_{s'}$, after $\gamma_{s'}$ rounds, $\bar{\mu}_s + 2\beta_2(\gamma_{s'}, \delta) \leq \bar{\mu}_{s'}$ and for all $s' \in col(s)\backslash s, \exists\gamma(s')$, after $\gamma_{s'}$ rounds, $\bar{\mu}_s \geq \bar{\mu}_{s'} + 2\beta_2(\gamma_{s'}, \delta)$, then the recommendation rule recommends s as the NE.
- For all arm s, if $\exists s' \in row(s)$, $\bar{\mu}_{s'} + \beta_2(t, \delta) \leq \bar{\mu}_s - \beta_2(t, \delta)$ or $\exists s' \in col(s)$, $\bar{\mu}_{s'} - \beta_2(t, \delta) \geq \bar{\mu}_s + \beta_2(t, \delta)$, then the recommendation rule determines that the underlying game does not have a NE.

Obviously, this algorithm is δ-PAC and the proof for this statement is the same as the proof for Theorem 5.