Sequential Bayesian Search
Appendices

Zheng Wen   Branislav Kveton   Brian Eriksson   Sandilya Bhamidipati

A Proof of Theorem 1

Assume that at the beginning of game t, the system's belief in the user's preference is $\mathbb{P}_t$. Then, the certainty-equivalet user preference during game t is

$$\pi_t^*(i) = \mathbb{E}_{\pi \sim \mathbb{P}_t}\left[\pi(i)\right] \quad \forall i \in \mathcal{I}.$$

Recall we define $\pi_{\min}^* = \min_{i \in \mathcal{I}} \pi^*(i)$, Lemma A-1 formalizes the result that if π_t^* is "close" to π^*, then for any decision tree T, $\mathbb{E}_{i \sim \pi_t^*}\left[N(T,i)\right]$ is "close" to $\mathbb{E}_{i \sim \pi^*}\left[N(T,i)\right]$:

Lemma A-1: *For any decision tree T, we have that*

$$\left|\mathbb{E}_{i \sim \pi^*}\left[N(T,i)\right] - \mathbb{E}_{i \sim \pi_t^*}\left[N(T,i)\right]\right| \leq \frac{\|\pi^* - \pi_t^*\|_\infty}{\pi_{\min}^*} \mathbb{E}_{i \sim \pi^*}\left[N(T,i)\right]. \tag{1}$$

Proof:
Notice that

$$\begin{aligned}
\left|\mathbb{E}_{i \sim \pi^*}\left[N(T,i)\right] - \mathbb{E}_{i \sim \pi_t^*}\left[N(T,i)\right]\right| &= \left|\sum_{i \in \mathcal{I}} \left[\pi^*(i) - \pi_t^*(i)\right] N(T,i)\right| \\
&\leq \sum_{i \in \mathcal{I}} \left|\pi^*(i) - \pi_t^*(i)\right| N(T,i) \\
&= \sum_{i \in \mathcal{I}} \frac{\left|\pi^*(i) - \pi_t^*(i)\right|}{\pi^*(i)} \pi^*(i) N(T,i) \\
&\leq \max_{i \in \mathcal{I}} \left[\frac{\left|\pi^*(i) - \pi_t^*(i)\right|}{\pi^*(i)}\right] \sum_{i \in \mathcal{I}} \pi^*(i) N(T,i) \\
&= \max_{i \in \mathcal{I}} \left[\frac{\left|\pi^*(i) - \pi_t^*(i)\right|}{\pi^*(i)}\right] \mathbb{E}_{i \sim \pi^*}\left[N(T,i)\right] \\
&\leq \frac{\max_{i \in \mathcal{I}} \left|\pi^*(i) - \pi_t^*(i)\right|}{\pi_{\min}^*} \mathbb{E}_{i \sim \pi^*}\left[N(T,i)\right] \\
&= \frac{\|\pi^* - \pi_t^*\|_\infty}{\pi_{\min}^*} \mathbb{E}_{i \sim \pi^*}\left[N(T,i)\right],
\end{aligned}$$

where the first inequality follows from the triangular inequality and the second inequality follows from the Hölder's inequality. Q.E.D.

Note that the bound in Lemma A-1 is tight in the following example. Assume $\mathcal{I} = \{1, 2\}$ and

$$\begin{aligned} N(T,1) &= 0 \\ N(T,2) &= 1 \\ \pi^*(1) &= 1-\varepsilon \\ \pi^*(2) &= \varepsilon \\ \pi_t^*(1) &= 1-2\varepsilon \\ \pi_t^*(2) &= 2\varepsilon. \end{aligned}$$

Then $\mathbb{E}_{i\sim\pi^*}\left[N(T,i)\right] = \varepsilon$, $\mathbb{E}_{i\sim\pi_t^*}\left[N(T,i)\right] = 2\varepsilon$, and therefore

$$\left|\mathbb{E}_{i\sim\pi^*}\left[N(T,i)\right] - \mathbb{E}_{i\sim\pi_t^*}\left[N(T,i)\right]\right| = \varepsilon.$$

On the other hand, we have $\|\pi^* - \pi_t^*\|_\infty = \varepsilon$. Furthermore, for $\varepsilon \leq \frac{1}{2}$, we have that

$$\pi^*_{\min} = \pi^*(2) = \varepsilon.$$

Thus we have

$$\frac{\|\pi^* - \pi_t^*\|_\infty}{\pi^*_{\min}}\mathbb{E}_{i\sim\pi^*}\left[N(T,i)\right] = \frac{\varepsilon}{\varepsilon}\varepsilon = \varepsilon.$$

Hence, the bound in Lemma A-1 is tight in this example.

Throughout this section, we assume the certainty-equivalent (CE) optimization problem is solved exactly, and use T_t^* to denote the solution of the CE optimization problem in game t, $\forall t = 0, 1, \cdots$. Lemma A-2 states that if $\|\pi_t^* - \pi^*\|_\infty$ is "small", the one-game regret (conditioning on π_t^*) is also "small":

Lemma A-2: *If $\|\pi^* - \pi_t^*\|_\infty < \pi^*_{\min}$, then we have*

$$\frac{2\|\pi^* - \pi_t^*\|_\infty}{\pi^*_{\min} - \|\pi^* - \pi_t^*\|_\infty}\mathbb{E}_{i\sim\pi^*}\left[N(T^*,i)\right] \geq \mathbb{E}_{i\sim\pi^*}\left[N(T_t^*,i)\right] - \mathbb{E}_{i\sim\pi^*}\left[N(T^*,i)\right] \geq 0. \quad (2)$$

Proof:
By definition of T_t^*, we have that

$$\mathbb{E}_{i\sim\pi_t^*}\left[N(T^*,i)\right] \geq \mathbb{E}_{i\sim\pi_t^*}\left[N(T_t^*,i)\right].$$

On the other hand, from the inequality (1), we have that

$$\begin{aligned} \mathbb{E}_{i\sim\pi_t^*}\left[N(T_t^*,i)\right] &\geq \mathbb{E}_{i\sim\pi^*}\left[N(T_t^*,i)\right] - \frac{\|\pi^* - \pi_t^*\|_\infty}{\pi^*_{\min}}\mathbb{E}_{i\sim\pi^*}\left[N(T_t^*,i)\right] \\ &= \frac{\pi^*_{\min} - \|\pi^* - \pi_t^*\|_\infty}{\pi^*_{\min}}\mathbb{E}_{i\sim\pi^*}\left[N(T_t^*,i)\right]. \end{aligned}$$

Similarly, we have that

$$\begin{aligned} \mathbb{E}_{i\sim\pi_t^*}\left[N(T^*,i)\right] &\leq \mathbb{E}_{i\sim\pi^*}\left[N(T^*,i)\right] + \frac{\|\pi^* - \pi_t^*\|_\infty}{\pi^*_{\min}}\mathbb{E}_{i\sim\pi^*}\left[N(T^*,i)\right] \\ &= \frac{\pi^*_{\min} + \|\pi^* - \pi_t^*\|_\infty}{\pi^*_{\min}}\mathbb{E}_{i\sim\pi^*}\left[N(T^*,i)\right]. \end{aligned}$$

Combining the above three inequalities, we have that

$$\frac{\pi^*_{\min} + \|\pi^* - \pi^*_t\|_\infty}{\pi^*_{\min}} \mathbb{E}_{i\sim\pi^*}\left[N(T^*, i)\right] \geq \frac{\pi^*_{\min} - \|\pi^* - \pi^*_t\|_\infty}{\pi^*_{\min}} \mathbb{E}_{i\sim\pi^*}\left[N(T^*_t, i)\right]$$

That is

$$\frac{\pi^*_{\min} + \|\pi^* - \pi^*_t\|_\infty}{\pi^*_{\min} - \|\pi^* - \pi^*_t\|_\infty} \mathbb{E}_{i\sim\pi^*}\left[N(T^*, i)\right] \geq \mathbb{E}_{i\sim\pi^*}\left[N(T^*_t, i)\right].$$

So we have

$$\mathbb{E}_{i\sim\pi^*}\left[N(T^*_t, i)\right] - \mathbb{E}_{i\sim\pi^*}\left[N(T^*, i)\right] \leq \frac{2\,\|\pi^* - \pi^*_t\|_\infty}{\pi^*_{\min} - \|\pi^* - \pi^*_t\|_\infty} \mathbb{E}_{i\sim\pi^*}\left[N(T^*, i)\right].$$

Finally, notice that by definition of T^* (i.e. $T^* \in \arg\min_T \mathbb{E}_{i\sim\pi^*}\left[N(T, i)\right]$), we have that

$$\mathbb{E}_{i\sim\pi^*}\left[N(T^*_t, i)\right] \geq \mathbb{E}_{i\sim\pi^*}\left[N(T^*, i)\right].$$

Thus, we have proved Lemma A-2. Q.E.D.

Now we consider the case when the prior belief $\mathbb{P}_0$ of the system is modeled as a Dirichlet distribution with parameter $\alpha \in \Re^M_+$ (henceforth denoted as $\mathrm{Dir}(\alpha)$). Specifically, its probability density function (PDF) over the probability simplex Δ^{M-1} is given by

$$f_{\mathbb{P}_0}(\pi) = \frac{1}{B(\alpha)} \prod_{i\in\mathcal{I}} \pi(i)^{\alpha(i)-1}, \forall \pi \in \Delta^{M-1},$$

where $\pi(i)$ is the probability mass at item i, and $\alpha(i)$ is the associated parameter. $B(\alpha)$ is a normalizing constant given by

$$B(\alpha) = \frac{\prod_{i\in\mathcal{I}} \Gamma\left(\alpha(i)\right)}{\Gamma\left(\sum_{i\in\mathcal{I}} \alpha(i)\right)},$$

where $\Gamma(\cdot)$ is the classical gamma function. The main advantage of Dirichlet prior is that it results in a simple posterior distribution, since it is the *conjugate prior* of the multinomial distribution. Specifically, $\forall t = 0, 1, \cdots$, we define the indicator vector $Z_t \in \Re^M$ as follows:

$$Z_t(i) = \begin{cases} 1 & \text{if } i = i_t \\ 0 & \text{otherwise} \end{cases}$$

Then, based on the Bayes rule, the posterior belief at the beginning of game t is

$$\mathbb{P}_t = \mathrm{Dir}\left(\alpha + \sum_{\tau=0}^{t-1} Z_\tau\right).$$

From the properties of Dirichlet distribution, we have that

$$\pi^*_t(i) = \mathbb{E}_{\pi\sim\mathbb{P}_t}\left[\pi(i)\right] = \frac{\alpha(i) + \sum_{\tau=0}^{t-1} Z_\tau(i)}{\sum_{i'\in\mathcal{I}}\left[\alpha(i') + \sum_{\tau=0}^{t-1} Z_\tau(i')\right]}.$$

Notice that $\sum_{i'\in\mathcal{I}}\sum_{\tau=0}^{t-1} Z_\tau(i') = \sum_{\tau=0}^{t-1}\sum_{i'\in\mathcal{I}} Z_\tau(i') = \sum_{\tau=0}^{t-1} 1 = t$. Furthermore, we define $\alpha_0 = \sum_{i'\in\mathcal{I}} \alpha(i')$. Thus, we have

$$\pi^*_t(i) = \frac{\alpha(i) + \sum_{\tau=0}^{t-1} Z_\tau(i)}{\alpha_0 + t} = \frac{\alpha_0}{\alpha_0 + t}\frac{\alpha(i)}{\alpha_0} + \frac{t}{\alpha_0 + t}\frac{\sum_{\tau=0}^{t-1} Z_\tau(i)}{t}.$$

Throughout this paper, we use the convention that "$\frac{0}{0} = 0$", so for $t = 0$, we have $\pi^*_0(i) = \frac{\alpha(i)}{\alpha_0}$. The above equation has a very nice interpretation: notice that $\frac{\alpha(i)}{\alpha_0}$ is the estimate of $\pi^*(i)$ based on the prior belief, while $\frac{\sum_{\tau=0}^{t-1} Z_\tau(i)}{t}$ is the estimate of $\pi^*(i)$ based on observations, the above equation states that $\pi_t^*(i)$ is a convex combination (weighted average) of these two estimates. Furthermore, the weights depend on t, the index of the current interactive game (or equivalently, the number of past observations).

From Hoeffding's inequality, $\forall \epsilon > 0$, we have that

$$\mathbb{P}\left(\left| \frac{\sum_{\tau=0}^{t-1} Z_\tau(i)}{t} - \pi^*(i) \right| \leq \epsilon \right) \geq 1 - 2\exp(-2\epsilon^2 t).$$

That is, for any $i \in \mathcal{I}$, at the beginning of game t, with probability at least $1 - 2\exp(-2\epsilon^2 t)$, we have that

$$\left| \frac{\sum_{\tau=0}^{t-1} Z_\tau(i)}{t} - \pi^*(i) \right| \leq \epsilon.$$

Let $E_t(i)$ denote the event that $\left| \frac{\sum_{\tau=0}^{t-1} Z_\tau(i)}{t} - \pi^*(i) \right| > \epsilon$. Then we have proved that $\mathbb{P}(E_t(i)) \leq 2\exp(-2\epsilon^2 t)$ for any $i \in \mathcal{I}$. From the union bound of the probability, we have that

$$\mathbb{P}\left(\cup_{i \in \mathcal{I}} E_t(i)\right) \leq \sum_{i \in \mathcal{I}} \mathbb{P}\left(E_t(i)\right) \leq 2M\exp(-2\epsilon^2 t).$$

Thus, with probability at least $1 - 2M\exp(-2\epsilon^2 t)$, we have that

$$\max_{i \in \mathcal{I}} \left| \frac{\sum_{\tau=0}^{t-1} Z_\tau(i)}{t} - \pi^*(i) \right| \leq \epsilon.$$

Finally, notice that $\forall i \in \mathcal{I}$, we have that

$$\begin{aligned} |\pi^*(i) - \pi_t^*(i)| &= \left| \frac{\alpha_0}{\alpha_0 + t} \left(\frac{\alpha(i)}{\alpha_0} - \pi^*(i) \right) + \frac{t}{\alpha_0 + t} \left(\frac{\sum_{\tau=0}^{t-1} Z_\tau(i)}{t} - \pi^*(i) \right) \right| \\ &\leq \frac{\alpha_0}{\alpha_0 + t} \left| \frac{\alpha(i)}{\alpha_0} - \pi^*(i) \right| + \frac{t}{\alpha_0 + t} \left| \frac{\sum_{\tau=0}^{t-1} Z_\tau(i)}{t} - \pi^*(i) \right|. \end{aligned}$$

Thus we have that

$$\begin{aligned} \max_{i \in \mathcal{I}} |\pi^*(i) - \pi_t^*(i)| &\leq \max_{i \in \mathcal{I}} \left[\frac{\alpha_0}{\alpha_0 + t} \left| \frac{\alpha(i)}{\alpha_0} - \pi^*(i) \right| + \frac{t}{\alpha_0 + t} \left| \frac{\sum_{\tau=0}^{t-1} Z_\tau(i)}{t} - \pi^*(i) \right| \right] \\ &\leq \frac{\alpha_0}{\alpha_0 + t} \max_{i \in \mathcal{I}} \left| \frac{\alpha(i)}{\alpha_0} - \pi^*(i) \right| + \frac{t}{\alpha_0 + t} \max_{i \in \mathcal{I}} \left| \frac{\sum_{\tau=0}^{t-1} Z_\tau(i)}{t} - \pi^*(i) \right|. \end{aligned}$$

Notice that $\max_{i \in \mathcal{I}} \left| \frac{\alpha(i)}{\alpha_0} - \pi^*(i) \right|$ is the maximum estimation error based on the prior belief, which is independent of the observations. On the other hand, $\max_{i \in \mathcal{I}} \left| \frac{\sum_{\tau=0}^{t-1} Z_\tau(i)}{t} - \pi^*(i) \right|$ is the maximum estimation error based on observations, which is a random variable.

Lemma A-3 upper bounds the regret in game t:

Lemma A-3: *$\forall t > 0$ and $\forall 0 < \eta \leq \frac{1}{3}$, if*

$$\frac{\alpha_0}{\alpha_0 + t} \max_{i \in \mathcal{I}} \left| \frac{\alpha(i)}{\alpha_0} - \pi^*(i) \right| < \eta \pi^*_{\min},$$

then we have that

$$\mathbb{E}_{T_t^*}\left\{\mathbb{E}_{i\sim\pi^*}\left[N(T_t^*,i)\right]-\mathbb{E}_{i\sim\pi^*}\left[N(T^*,i)\right]\right\}<3\eta\mathbb{E}_{i\sim\pi^*}\left[N(T^*,i)\right]+2M|\mathcal{Q}|\exp\left\{4\eta\pi^*_{\min}\alpha_0\max_{i\in\mathcal{I}}\left|\frac{\alpha(i)}{\alpha_0}-\pi^*(i)\right|-2\left[\eta\pi^*_{\min}\right]^2t\right\}.$$

Proof:
Since $\frac{\alpha_0}{\alpha_0+t}\max_{i\in\mathcal{I}}\left|\frac{\alpha(i)}{\alpha_0}-\pi^*(i)\right|<\eta\pi^*_{\min}$, thus, one sufficient condition to ensure that

$$\max_{i\in\mathcal{I}}|\pi^*(i)-\pi_t^*(i)|\leq\eta\pi^*_{\min}$$

is

$$\max_{i\in\mathcal{I}}\left|\frac{\sum_{\tau=0}^{t-1}Z_\tau(i)}{t}-\pi^*(i)\right|\leq\eta\pi^*_{\min}\left(1+\frac{\alpha_0}{t}\right)-\frac{\alpha_0}{t}\max_{i\in\mathcal{I}}\left|\frac{\alpha(i)}{\alpha_0}-\pi^*(i)\right|. \tag{3}$$

Define $\epsilon=\eta\pi^*_{\min}\left(1+\frac{\alpha_0}{t}\right)-\frac{\alpha_0}{t}\max_{i\in\mathcal{I}}\left|\frac{\alpha(i)}{\alpha_0}-\pi^*(i)\right|$, from the above discussion, we know that inequality (3) holds with probability at least $1-2M\exp(-2\epsilon^2t)$. Furthermore, from Lemma A-2, $\max_{i\in\mathcal{I}}|\pi^*(i)-\pi_t^*(i)|\leq\eta\pi^*_{\min}$ implies that

$$\frac{2\eta}{1-\eta}\mathbb{E}_{i\sim\pi^*}\left[N(T^*,i)\right]\geq\mathbb{E}_{i\sim\pi^*}\left[N(T_t^*,i)\right]-\mathbb{E}_{i\sim\pi^*}\left[N(T^*,i)\right]. \tag{4}$$

Thus, we have proved that with probability at least $1-2M\exp(-2\epsilon^2t)$, inequality (4) holds. In other words, if we define E as the event that inequality (4) holds, then we have that $\mathbb{P}(E)\geq1-2M\exp(-2\epsilon^2t)$

On the other hand, notice that a naive bound on the regret is $\mathbb{E}_{i\sim\pi^*}\left[N(T_t^*,i)\right]-\mathbb{E}_{i\sim\pi^*}\left[N(T^*,i)\right]\leq|\mathcal{Q}|$. With E defined as the event that inequality (4) holds and $\bar{E}$ defined as the complement of E, we have that:

$$\begin{aligned}\mathbb{E}_{T_t^*}\left\{\mathbb{E}_{i\sim\pi^*}\left[N(T_t^*,i)\right]-\mathbb{E}_{i\sim\pi^*}\left[N(T^*,i)\right]\right\}&\leq\mathbb{P}(E)\mathbb{E}_{T_t^*}\left\{\mathbb{E}_{i\sim\pi^*}\left[N(T_t^*,i)\right]-\mathbb{E}_{i\sim\pi^*}\left[N(T^*,i)\right]|E\right\}\\&+\left[1-\mathbb{P}(E)\right]\mathbb{E}_{T_t^*}\left\{\mathbb{E}_{i\sim\pi^*}\left[N(T_t^*,i)\right]-\mathbb{E}_{i\sim\pi^*}\left[N(T^*,i)\right]|\bar{E}\right\}\\&\leq\mathbb{P}(E)\frac{2\eta}{1-\eta}\mathbb{E}_{i\sim\pi^*}\left[N(T^*,i)\right]+\left[1-\mathbb{P}(E)\right]|\mathcal{Q}|.\end{aligned}$$

On the other hand, notice that $\mathbb{E}_{i\sim\pi^*}\left[N(T^*,i)\right]\leq|\mathcal{Q}|$ by definition, and $\eta\leq\frac{1}{3}$ implies that $\frac{2\eta}{1-\eta}\leq1$, thus we have $\frac{2\eta}{1-\eta}\mathbb{E}_{i\sim\pi^*}\left[N(T^*,i)\right]\leq|\mathcal{Q}|$. Together with $\mathbb{P}(E)\geq1-2M\exp(-2\epsilon^2t)$, we have that

$$\begin{aligned}\mathbb{E}_{T_t^*}\left\{\mathbb{E}_{i\sim\pi^*}\left[N(T_t^*,i)\right]-\mathbb{E}_{i\sim\pi^*}\left[N(T^*,i)\right]\right\}&\leq\left[1-2M\exp(-2\epsilon^2t)\right]\frac{2\eta}{1-\eta}\mathbb{E}_{i\sim\pi^*}\left[N(T^*,i)\right]+2M\exp(-2\epsilon^2t)|\mathcal{Q}|\\&<\frac{2\eta}{1-\eta}\mathbb{E}_{i\sim\pi^*}\left[N(T^*,i)\right]+2M|\mathcal{Q}|\exp(-2\epsilon^2t).\end{aligned}$$

Notice that $0<\eta\leq\frac{1}{3}$ implies that $0<\frac{1}{1-\eta}\leq\frac{3}{2}$, thus $0<\frac{2\eta}{1-\eta}\leq3\eta$. Hence we have that

$$\mathbb{E}_{T_t^*}\left\{\mathbb{E}_{i\sim\pi^*}\left[N(T_t^*,i)\right]-\mathbb{E}_{i\sim\pi^*}\left[N(T^*,i)\right]\right\}<3\eta\mathbb{E}_{i\sim\pi^*}\left[N(T^*,i)\right]+2M|\mathcal{Q}|\exp(-2\epsilon^2t).$$

From the definition of ϵ, we have

$$\begin{aligned}\epsilon^2t&=\left[\eta\pi^*_{\min}\sqrt{t}+\frac{\alpha_0}{\sqrt{t}}\left(\eta\pi^*_{\min}-\max_{i\in\mathcal{I}}\left|\frac{\alpha(i)}{\alpha_0}-\pi^*(i)\right|\right)\right]^2\\&=\left[\eta\pi^*_{\min}\right]^2t+2\eta\pi^*_{\min}\alpha_0\left(\eta\pi^*_{\min}-\max_{i\in\mathcal{I}}\left|\frac{\alpha(i)}{\alpha_0}-\pi^*(i)\right|\right)+\frac{\alpha_0^2}{t}\left(\eta\pi^*_{\min}-\max_{i\in\mathcal{I}}\left|\frac{\alpha(i)}{\alpha_0}-\pi^*(i)\right|\right)^2\\&>\left[\eta\pi^*_{\min}\right]^2t-2\eta\pi^*_{\min}\alpha_0\max_{i\in\mathcal{I}}\left|\frac{\alpha(i)}{\alpha_0}-\pi^*(i)\right|,\end{aligned}$$

where the last inequality follows from the fact that $\frac{\alpha_0^2}{t}\left(\eta\pi^*_{\min} - \max_{i\in\mathcal{I}}\left|\frac{\alpha(i)}{\alpha_0} - \pi^*(i)\right|\right)^2 \geq 0$ and $\left[\eta\pi^*_{\min}\right]^2\alpha_0 > 0$. So we have

$$-2\epsilon^2 t < 4\eta\pi^*_{\min}\alpha_0 \max_{i\in\mathcal{I}}\left|\frac{\alpha(i)}{\alpha_0} - \pi^*(i)\right| - 2\left[\eta\pi^*_{\min}\right]^2 t.$$

Thus, we have proved Lemma A-3. Q.E.D.

We define τ_E as

$$\tau_E = \min\left\{t \geq 4 : \frac{\ln(t)}{t} \leq \left(\frac{\pi^*_{\min}}{6}\right)^2 \text{ and } \frac{4}{3}\alpha_0 \max_{i\in\mathcal{I}}\left|\frac{\alpha(i)}{\alpha_0} - \pi^*(i)\right| \leq \left[t\ln(t)\right]^{\frac{1}{2}}\right\}, \tag{5}$$

where $\ln(\cdot)$ is the logarithm function with base e. Notice that for $t \geq 3$, $\frac{\ln(t)}{t}$ is monotonically decreasing. Notice that τ_E depends on (1) $\pi^*_{\min}$ and (2) the "quality" of the prior. Lemma A-4 derives a more useful one-game regret bound based on Lemma A-3 and the definition of τ_E:

Lemma A-4: *$\forall t \geq \tau_E$, we have*

$$\mathbb{E}_{T_t^*}\left\{\mathbb{E}_{i\sim\pi^*}\left[N(T_t^*, i)\right] - \mathbb{E}_{i\sim\pi^*}\left[N(T^*, i)\right]\right\} < \frac{6}{\pi^*_{\min}}\left[\frac{\ln(t)}{t}\right]^{\frac{1}{2}}\mathbb{E}_{i\sim\pi^*}\left[N(T^*, i)\right] + \frac{2M|\mathcal{Q}|}{t^2},$$

where τ_E is defined in Eqn(5).

Proof:

For $\forall t \geq \tau_E$, we choose $\eta = \frac{2}{\pi^*_{\min}}\left[\frac{\ln(t)}{t}\right]^{\frac{1}{2}}$. We first show that this particular η satisfies the conditions of Lemma A-3. Since $\frac{\ln(t)}{t}$ is monotonically decreasing for $t \geq 3$ and $t \geq \tau_E \geq 4$, we have that

$$\eta = \frac{2}{\pi^*_{\min}}\left[\frac{\ln(t)}{t}\right]^{\frac{1}{2}} \leq \frac{2}{\pi^*_{\min}}\left[\frac{\ln(\tau_E)}{\tau_E}\right]^{\frac{1}{2}} \leq \frac{2}{\pi^*_{\min}}\frac{\pi^*_{\min}}{6} = \frac{1}{3}.$$

On the other hand, since $t\ln(t)$ is monotonically increasing, thus, $t \geq \tau_E$ implies that

$$\frac{4}{3}\alpha_0 \max_{i\in\mathcal{I}}\left|\frac{\alpha(i)}{\alpha_0} - \pi^*(i)\right| \leq \left[t\ln(t)\right]^{\frac{1}{2}}.$$

Thus

$$\frac{\alpha_0}{\alpha_0 + t}\max_{i\in\mathcal{I}}\left|\frac{\alpha(i)}{\alpha_0} - \pi^*(i)\right| < \frac{\alpha_0}{t}\max_{i\in\mathcal{I}}\left|\frac{\alpha(i)}{\alpha_0} - \pi^*(i)\right| \leq \frac{3\left[t\ln(t)\right]^{\frac{1}{2}}}{4t} = \frac{2}{\pi^*_{\min}}\left[\frac{\ln(t)}{t}\right]^{\frac{1}{2}}\frac{3\pi^*_{\min}}{8} = \frac{3\pi^*_{\min}}{8}\eta < \eta\pi^*_{\min}.$$

Thus, the conditions of Lemma A-3 are satisfied and we have that

$$\mathbb{E}_{T_t^*}\left\{\mathbb{E}_{i\sim\pi^*}\left[N(T_t^*, i)\right] - \mathbb{E}_{i\sim\pi^*}\left[N(T^*, i)\right]\right\} < 3\eta\mathbb{E}_{i\sim\pi^*}\left[N(T^*, i)\right] + 2M|\mathcal{Q}|\exp\left\{4\eta\pi^*_{\min}\alpha_0\max_{i\in\mathcal{I}}\left|\frac{\alpha(i)}{\alpha_0} - \pi^*(i)\right| - 2\left[\eta\pi^*_{\min}\right]^2 t\right\}.$$

Notice that

$$4\eta\pi^*_{\min}\alpha_0\max_{i\in\mathcal{I}}\left|\frac{\alpha(i)}{\alpha_0} - \pi^*(i)\right| = 8\left[\frac{\ln(t)}{t}\right]^{\frac{1}{2}}\alpha_0\max_{i\in\mathcal{I}}\left|\frac{\alpha(i)}{\alpha_0} - \pi^*(i)\right| \leq 8\left[\frac{\ln(t)}{t}\right]^{\frac{1}{2}}\frac{3}{4}\left[t\ln(t)\right]^{\frac{1}{2}} = 6\ln(t),$$

and

$$2\left[\eta\pi^*_{\min}\right]^2 t = 2\left[\left(\frac{2}{\pi^*_{\min}}\left[\frac{\ln(t)}{t}\right]^{\frac{1}{2}}\right)\pi^*_{\min}\right]^2 t = 2\left[\left(2\left[\frac{\ln(t)}{t}\right]^{\frac{1}{2}}\right)\right]^2 t = 8\ln(t).$$

Thus we have

$$\exp\left\{4\eta\pi^*_{\min}\alpha_0 \max_{i\in\mathcal{I}}\left|\frac{\alpha(i)}{\alpha_0}-\pi^*(i)\right| - 2\left[\eta\pi^*_{\min}\right]^2 t\right\} \leq \exp\left\{6\ln(t) - 8\ln(t)\right\} = \exp\left\{-2\ln(t)\right\} = \frac{1}{t^2}.$$

On the other hand, we have that $3\eta = \frac{6}{\pi^*_{\min}}\left[\frac{\ln(t)}{t}\right]^{\frac{1}{2}}$, thus we have

$$\begin{aligned}
\mathbb{E}_{T^*_t}\left\{\mathbb{E}_{i\sim\pi^*}\left[N(T^*_t,i)\right] - \mathbb{E}_{i\sim\pi^*}\left[N(T^*,i)\right]\right\} &< 3\eta\mathbb{E}_{i\sim\pi^*}\left[N(T^*,i)\right] + 2M|\mathcal{Q}|\exp\left\{4\eta\pi^*_{\min}\alpha_0 \max_{i\in\mathcal{I}}\left|\frac{\alpha(i)}{\alpha_0}-\pi^*(i)\right| - 2\left[\eta\pi^*_{\min}\right]^2 t\right\} \\
&\leq \frac{6}{\pi^*_{\min}}\left[\frac{\ln(t)}{t}\right]^{\frac{1}{2}}\mathbb{E}_{i\sim\pi^*}\left[N(T^*,i)\right] + \frac{2M|\mathcal{Q}|}{t^2}.
\end{aligned}$$

Q.E.D.

In this remainder of this section, we prove Theorem 1:

Proof of Theorem 1:
Notice that a naive bound on $\mathbb{E}_{T^*_t}\left\{\mathbb{E}_{i\sim\pi^*}\left[N(T^*_t,i)\right] - \mathbb{E}_{i\sim\pi^*}\left[N(T^*,i)\right]\right\}$ is

$$\mathbb{E}_{T^*_t}\left\{\mathbb{E}_{i\sim\pi^*}\left[N(T^*_t,i)\right] - \mathbb{E}_{i\sim\pi^*}\left[N(T^*,i)\right]\right\} \leq |\mathcal{Q}|.$$

Thus, for $0 \leq \tau < \tau_E$, we have that

$$\sum_{t=0}^{\tau}\mathbb{E}_{T^*_t}\left\{\mathbb{E}_{i\sim\pi^*}\left[N(T^*_t,i)\right] - \mathbb{E}_{i\sim\pi^*}\left[N(T^*,i)\right]\right\} \leq |\mathcal{Q}|(\tau+1).$$

On the other hand, from Lemma A-4, for $\tau \geq \tau_E$, we have that

$$\begin{aligned}
\sum_{t=0}^{\tau}\mathbb{E}_{T^*_t}\left\{\mathbb{E}_{i\sim\pi^*}\left[N(T^*_t,i)\right] - \mathbb{E}_{i\sim\pi^*}\left[N(T^*,i)\right]\right\} &= \sum_{t=0}^{\tau_E-1}\mathbb{E}_{T^*_t}\left\{\mathbb{E}_{i\sim\pi^*}\left[N(T^*_t,i)\right] - \mathbb{E}_{i\sim\pi^*}\left[N(T^*,i)\right]\right\} \\
&+ \sum_{t=\tau_E}^{\tau}\mathbb{E}_{T^*_t}\left\{\mathbb{E}_{i\sim\pi^*}\left[N(T^*_t,i)\right] - \mathbb{E}_{i\sim\pi^*}\left[N(T^*,i)\right]\right\} \\
&\leq |\mathcal{Q}|\tau_E + \sum_{t=\tau_E}^{\tau}\mathbb{E}_{T^*_t}\left\{\mathbb{E}_{i\sim\pi^*}\left[N(T^*_t,i)\right] - \mathbb{E}_{i\sim\pi^*}\left[N(T^*,i)\right]\right\} \\
&\leq |\mathcal{Q}|\tau_E + \sum_{t=\tau_E}^{\tau}\left[\frac{6}{\pi^*_{\min}}\left[\frac{\ln(t)}{t}\right]^{\frac{1}{2}}\mathbb{E}_{i\sim\pi^*}\left[N(T^*,i)\right] + \frac{2M|\mathcal{Q}|}{t^2}\right],
\end{aligned}$$

where the first inequality follows from the naive bound and the second inequality follows from Lemma A-4. Since $\tau_E > 1$, we have that

$$\sum_{t=\tau_E}^{\tau}\frac{1}{t^2} < \sum_{t=\tau_E}^{\infty}\frac{1}{t^2} < \sum_{t=\tau_E}^{\infty}\frac{1}{(t-1)t} = \sum_{t=\tau_E}^{\infty}\left[\frac{1}{t-1}-\frac{1}{t}\right] = \frac{1}{\tau_E-1}.$$

On the other hand, notice that $\frac{\ln(t)}{t}$ is monotonically decreasing on interval $[\tau_E-1,\infty)$ (Since $\tau_E - 1 \geq 3$), and the derivative of the function $[t\ln(t)]^{\frac{1}{2}}$ is $\frac{1}{2}\left[\frac{1}{t\ln(t)}\right]^{\frac{1}{2}} + \frac{1}{2}\left[\frac{\ln(t)}{t}\right]^{\frac{1}{2}}$, we have that

$$\sum_{t=\tau_E}^{\tau}\left[\frac{\ln(t)}{t}\right]^{\frac{1}{2}} < \int_{\tau_E-1}^{\tau}\left[\frac{\ln(t)}{t}\right]^{\frac{1}{2}}dt < \int_{\tau_E-1}^{\tau}\left(\left[\frac{\ln(t)}{t}\right]^{\frac{1}{2}} + \left[\frac{1}{t\ln(t)}\right]^{\frac{1}{2}}\right)dt = 2\left[\tau\ln(\tau)\right]^{\frac{1}{2}} - 2\left[(\tau_E-1)\ln(\tau_E-1)\right]^{\frac{1}{2}}.$$

Thus, for $\tau \geq \tau_E$, we have that

$$\sum_{t=0}^{\tau} \mathbb{E}_{T_t^*} \left\{ \mathbb{E}_{i\sim\pi^*} \left[N(T_t^*, i) \right] - \mathbb{E}_{i\sim\pi^*} \left[N(T^*, i) \right] \right\} \leq |\mathcal{Q}|\tau_E + \sum_{t=\tau_E}^{\tau} \left[\frac{6}{\pi^*_{\min}} \left[\frac{\ln(t)}{t} \right]^{\frac{1}{2}} \mathbb{E}_{i\sim\pi^*} \left[N(T^*, i) \right] + \frac{2M|\mathcal{Q}|}{t^2} \right]$$

$$< \quad |\mathcal{Q}|\tau_E + \frac{12\mathbb{E}_{i\sim\pi^*} \left[N(T^*, i) \right]}{\pi^*_{\min}} \left\{ [\tau \ln(\tau)]^{\frac{1}{2}} - [(\tau_E - 1)\ln(\tau_E - 1)]^{\frac{1}{2}} \right\} + \frac{2M|\mathcal{Q}|}{\tau_E - 1} = O([\tau \ln(\tau)]^{\frac{1}{2}}).$$

Q.E.D.

B Proof of Theorem 2

Throughout this section, we assume that the certainty-equivalent (CE) optimization problem is solved by the greedy algorithm, and use T_t^{g} to denote the solution based on the greedy algorithm of the CE optimization problem in game t, $\forall t = 0, 1, \cdots$. Note that in the proof, **we still use T_t^* to denote the exact solution of the CE optimization problem in game** t. Lemma A-5 is the counterpart of Lemma A-2 in this case:

Lemma A-5: *If $\|\pi^* - \pi_t^*\|_\infty < \pi_{\min}^*$, then we have*

$$\mathbb{E}_{i\sim\pi^*}\left[N(T_t^{\mathrm{g}}, i)\right] \leq \mathbb{E}_{i\sim\pi^*}\left[N(T^*, i)\right]\left[\frac{\pi_{\min}^* + \|\pi^* - \pi_t^*\|_\infty}{\pi_{\min}^* - \|\pi^* - \pi_t^*\|_\infty}\right]^2 \ln\left(\frac{e}{\pi_{\min}^* - \|\pi^* - \pi_t^*\|_\infty}\right).$$

Proof:
Before proceeding, notice that from Theorem 10 of [1], we have that

$$\mathbb{E}_{i\sim\pi_t^*}\left[N(T_t^{\mathrm{g}}, i)\right] \leq \mathbb{E}_{i\sim\pi_t^*}\left[N(T_t^*, i)\right]\left(\ln\left(\frac{1}{\min_i \pi_t^*(i)}\right) + 1\right),$$

where T_t^* is the exact solution of the CE optimization problem in game t, and T_t^{g} is the approximation solution based on the greedy algorithm. From Lemma A-1, we know that

$$\mathbb{E}_{i\sim\pi_t^*}\left[N(T_t^{\mathrm{g}}, i)\right] \geq \frac{\pi_{\min}^* - \|\pi^* - \pi_t^*\|_\infty}{\pi_{\min}^*}\mathbb{E}_{i\sim\pi^*}\left[N(T_t^{\mathrm{g}}, i)\right],$$

and

$$\mathbb{E}_{i\sim\pi_t^*}\left[N(T_t^*, i)\right] \leq \frac{\pi_{\min}^* + \|\pi^* - \pi_t^*\|_\infty}{\pi_{\min}^*}\mathbb{E}_{i\sim\pi^*}\left[N(T_t^*, i)\right].$$

Thus we have that

$$\begin{aligned} & \frac{\pi_{\min}^* - \|\pi^* - \pi_t^*\|_\infty}{\pi_{\min}^*}\mathbb{E}_{i\sim\pi^*}\left[N(T_t^{\mathrm{g}}, i)\right] \\ \leq \; & \frac{\pi_{\min}^* + \|\pi^* - \pi_t^*\|_\infty}{\pi_{\min}^*}\mathbb{E}_{i\sim\pi^*}\left[N(T_t^*, i)\right]\left(\ln\left(\frac{1}{\min_i \pi_t^*(i)}\right) + 1\right). \end{aligned}$$

We define $c = \frac{\pi_{\min}^* + \|\pi^* - \pi_t^*\|_\infty}{\pi_{\min}^* - \|\pi^* - \pi_t^*\|_\infty}$, thus we have that

$$\mathbb{E}_{i\sim\pi^*}\left[N(T_t^{\mathrm{g}}, i)\right] \leq c\mathbb{E}_{i\sim\pi^*}\left[N(T_t^*, i)\right]\left(\ln\left(\frac{1}{\min_i \pi_t^*(i)}\right) + 1\right).$$

Combining with Lemma A-2, we have that

$$\mathbb{E}_{i\sim\pi^*}\left[N(T_t^{\mathrm{g}}, i)\right] \leq c^2\mathbb{E}_{i\sim\pi^*}\left[N(T^*, i)\right]\left(\ln\left(\frac{1}{\min_i \pi_t^*(i)}\right) + 1\right). \tag{6}$$

Finally, assume that $\min_i \pi_t^*(i) = \pi_t^*(i^*)$, we have that

$$\min_i \pi_t^*(i) = \pi_t^*(i^*) = \pi^*(i^*) + [\pi_t^*(i^*) - \pi^*(i^*)] \geq \pi_{\min}^* - \|\pi^* - \pi_t^*\|_\infty.$$

So we have

$$\ln\left(\frac{1}{\min_i \pi_t^*(i)}\right) \leq \ln\left(\frac{1}{\pi_{\min}^* - \|\pi^* - \pi_t^*\|_\infty}\right).$$

Hence

$$\ln\left(\frac{1}{\min_i \pi_t^*(i)}\right) + 1 \leq \ln\left(\frac{1}{\pi_{\min}^* - \|\pi^* - \pi_t^*\|_\infty}\right) + 1 = \ln\left(\frac{e}{\pi_{\min}^* - \|\pi^* - \pi_t^*\|_\infty}\right).$$

Plug the above inequality into Eqn(6), we have proved Lemma A-5. Q.E.D.

Lemma A-6 upper bounds the scaled regret in game t:

Lemma A-6: $\forall t > 0$ *and* $\forall 0 < \eta < 1$, *if*

$$\left[\frac{1+\eta}{1-\eta}\right]^2 \ln\left(\frac{1}{1-\eta}\right) + \frac{4\eta}{(1-\eta)^2}\ln\left(\frac{e}{\pi^*_{\min}}\right) \leq 1 \quad \textit{and} \quad \frac{\alpha_0}{\alpha_0+t}\max_{i\in\mathcal{I}}\left|\frac{\alpha(i)}{\alpha_0} - \pi^*(i)\right| < \eta\pi^*_{\min},$$

then we have that

$$\begin{aligned}
& \mathbb{E}_{T^{\mathrm{g}}_t}\left\{\mathbb{E}_{i\sim\pi^*}\left[N(T^{\mathrm{g}}_t, i)\right] - \ln\left(\frac{e}{\pi^*_{\min}}\right)\mathbb{E}_{i\sim\pi^*}\left[N(T^*, i)\right]\right\} \\
< & \left\{\left[\frac{1+\eta}{1-\eta}\right]^2 \ln\left(\frac{1}{1-\eta}\right) + \frac{4\eta}{(1-\eta)^2}\ln\left(\frac{e}{\pi^*_{\min}}\right)\right\}\mathbb{E}_{i\sim\pi^*}\left[N(T^*, i)\right] \\
+ & \; 2M|\mathcal{Q}|\exp\left(4\eta\pi^*_{\min}\alpha_0 \max_{i\in\mathcal{I}}\left|\frac{\alpha(i)}{\alpha_0} - \pi^*(i)\right| - 2\left[\eta\pi^*_{\min}\right]^2 t\right).
\end{aligned}$$

Proof:
Since $\frac{\alpha_0}{\alpha_0+t}\max_{i\in\mathcal{I}}\left|\frac{\alpha(i)}{\alpha_0} - \pi^*(i)\right| < \eta\pi^*_{\min}$, thus, one sufficient condition to ensure that

$$\max_{i\in\mathcal{I}}|\pi^*(i) - \pi^*_t(i)| \leq \eta\pi^*_{\min}$$

is

$$\max_{i\in\mathcal{I}}\left|\frac{\sum_{\tau=0}^{t-1} Z_\tau(i)}{t} - \pi^*(i)\right| \leq \eta\pi^*_{\min}\left(1 + \frac{\alpha_0}{t}\right) - \frac{\alpha_0}{t}\max_{i\in\mathcal{I}}\left|\frac{\alpha(i)}{\alpha_0} - \pi^*(i)\right|. \tag{7}$$

We define $\epsilon = \eta\pi^*_{\min}\left(1 + \frac{\alpha_0}{t}\right) - \frac{\alpha_0}{t}\max_{i\in\mathcal{I}}\left|\frac{\alpha(i)}{\alpha_0} - \pi^*(i)\right|$. From the discussion before Lemma A-3, we know that inequality (7) holds with probability at least $1 - 2M\exp(-2\epsilon^2 t)$. Furthermore, from Lemma A-5, $\|\pi^* - \pi^*_t\| \leq \eta\pi^*_{\min}$ implies that

$$\mathbb{E}_{i\sim\pi^*}\left[N(T^{\mathrm{g}}_t, i)\right] \leq \mathbb{E}_{i\sim\pi^*}\left[N(T^*, i)\right]\left[\frac{1+\eta}{1-\eta}\right]^2 \ln\left(\frac{e}{\pi^*_{\min}(1-\eta)}\right).$$

Thus we have that

$$\begin{aligned}
& \mathbb{E}_{i\sim\pi^*}\left[N(T^{\mathrm{g}}_t, i)\right] - \ln\left(\frac{e}{\pi^*_{\min}}\right)\mathbb{E}_{i\sim\pi^*}\left[N(T^*, i)\right] \\
\leq & \left\{\left[\frac{1+\eta}{1-\eta}\right]^2 \ln\left(\frac{e}{\pi^*_{\min}(1-\eta)}\right) - \ln\left(\frac{e}{\pi^*_{\min}}\right)\right\}\mathbb{E}_{i\sim\pi^*}\left[N(T^*, i)\right] \\
= & \left\{\left[\frac{1+\eta}{1-\eta}\right]^2 \ln\left(\frac{1}{1-\eta}\right) + \frac{4\eta}{(1-\eta)^2}\ln\left(\frac{e}{\pi^*_{\min}}\right)\right\}\mathbb{E}_{i\sim\pi^*}\left[N(T^*, i)\right].
\end{aligned} \tag{8}$$

Thus, we have proved that with probability at least $1 - 2M\exp(-2\epsilon^2 t)$, inequality (8) holds. In other words, if we define E as the event that inequality (8) holds, then we have that $\mathbb{P}(E) \geq 1 - 2M\exp(-2\epsilon^2 t)$.

On the other hand, notice that a naive bound on the scaled regret is

$$\mathbb{E}_{i\sim\pi^*}\left[N(T^{\mathrm{g}}_t, i)\right] - \ln\left(\frac{e}{\pi^*_{\min}}\right)\mathbb{E}_{i\sim\pi^*}\left[N(T^*, i)\right] \leq |\mathcal{Q}|.$$

With E defined as the event that inequality (8) holds and $\bar{E}$ defined as the complement of E, we have that:

$$\begin{aligned}
& \mathbb{E}_{T_t^{\mathrm{g}}}\left\{\mathbb{E}_{i\sim\pi^*}\left[N(T_t^{\mathrm{g}},i)\right]-\ln\left(\frac{e}{\pi^*_{\min}}\right)\mathbb{E}_{i\sim\pi^*}\left[N(T^*,i)\right]\right\} \\
\leq\ & \mathbb{P}(E)\mathbb{E}_{T_t^{\mathrm{g}}}\left\{\mathbb{E}_{i\sim\pi^*}\left[N(T_t^{\mathrm{g}},i)\right]-\ln\left(\frac{e}{\pi^*_{\min}}\right)\mathbb{E}_{i\sim\pi^*}\left[N(T^*,i)\right]|E\right\} \\
+\ & \left[1-\mathbb{P}(E)\right]\mathbb{E}_{T_t^{\mathrm{g}}}\left\{\mathbb{E}_{i\sim\pi^*}\left[N(T_t^{\mathrm{g}},i)\right]-\ln\left(\frac{e}{\pi^*_{\min}}\right)\mathbb{E}_{i\sim\pi^*}\left[N(T^*,i)\right]|\bar{E}\right\} \\
\leq\ & \mathbb{P}(E)\left\{\left[\frac{1+\eta}{1-\eta}\right]^2\ln\left(\frac{1}{1-\eta}\right)+\frac{4\eta}{(1-\eta)^2}\ln\left(\frac{e}{\pi^*_{\min}}\right)\right\}\mathbb{E}_{i\sim\pi^*}\left[N(T^*,i)\right]+\left[1-\mathbb{P}(E)\right]|\mathcal{Q}|.
\end{aligned}$$

On the other hand, notice that $\mathbb{E}_{i\sim\pi^*}\left[N(T^*,i)\right]\leq|\mathcal{Q}|$ by definition, and $\left[\frac{1+\eta}{1-\eta}\right]^2\ln\left(\frac{1}{1-\eta}\right)+\frac{4\eta}{(1-\eta)^2}\ln\left(\frac{e}{\pi^*_{\min}}\right)\leq 1$, thus we have

$$\left\{\left[\frac{1+\eta}{1-\eta}\right]^2\ln\left(\frac{1}{1-\eta}\right)+\frac{4\eta}{(1-\eta)^2}\ln\left(\frac{e}{\pi^*_{\min}}\right)\right\}\mathbb{E}_{i\sim\pi^*}\left[N(T^*,i)\right]\leq|\mathcal{Q}|.$$

Together with $\mathbb{P}(E)\geq 1-2M\exp(-2\epsilon^2 t)$, we have that

$$\begin{aligned}
& \mathbb{E}_{T_t^{\mathrm{g}}}\left\{\mathbb{E}_{i\sim\pi^*}\left[N(T_t^{\mathrm{g}},i)\right]-\ln\left(\frac{e}{\pi^*_{\min}}\right)\mathbb{E}_{i\sim\pi^*}\left[N(T^*,i)\right]\right\} \\
\leq\ & \left\{\left[\frac{1+\eta}{1-\eta}\right]^2\ln\left(\frac{1}{1-\eta}\right)+\frac{4\eta}{(1-\eta)^2}\ln\left(\frac{e}{\pi^*_{\min}}\right)\right\}\mathbb{E}_{i\sim\pi^*}\left[N(T^*,i)\right]+2M|\mathcal{Q}|\exp(-2\epsilon^2 t).
\end{aligned}$$

From the definition of ϵ, we have

$$\begin{aligned}
\epsilon^2 t &= \left[\eta\pi^*_{\min}\sqrt{t}+\frac{\alpha_0}{\sqrt{t}}\left(\eta\pi^*_{\min}-\max_{i\in\mathcal{I}}\left|\frac{\alpha(i)}{\alpha_0}-\pi^*(i)\right|\right)\right]^2 \\
&= \left[\eta\pi^*_{\min}\right]^2 t+2\eta\pi^*_{\min}\alpha_0\left(\eta\pi^*_{\min}-\max_{i\in\mathcal{I}}\left|\frac{\alpha(i)}{\alpha_0}-\pi^*(i)\right|\right)+\frac{\alpha_0^2}{t}\left(\eta\pi^*_{\min}-\max_{i\in\mathcal{I}}\left|\frac{\alpha(i)}{\alpha_0}-\pi^*(i)\right|\right)^2 \\
&> \left[\eta\pi^*_{\min}\right]^2 t-2\eta\pi^*_{\min}\alpha_0\max_{i\in\mathcal{I}}\left|\frac{\alpha(i)}{\alpha_0}-\pi^*(i)\right|,
\end{aligned}$$

where the last inequality follows from the fact that $\frac{\alpha_0^2}{t}\left(\eta\pi^*_{\min}-\max_{i\in\mathcal{I}}\left|\frac{\alpha(i)}{\alpha_0}-\pi^*(i)\right|\right)^2\geq 0$ and $\left[\eta\pi^*_{\min}\right]^2\alpha_0>0$. So we have

$$-2\epsilon^2 t<4\eta\pi^*_{\min}\alpha_0\max_{i\in\mathcal{I}}\left|\frac{\alpha(i)}{\alpha_0}-\pi^*(i)\right|-2\left[\eta\pi^*_{\min}\right]^2 t.$$

Thus, we have proved Lemma A-6. Q.E.D.

Before proceeding, we derive a sufficient condition for

$$\left[\frac{1+\eta}{1-\eta}\right]^2\ln\left(\frac{1}{1-\eta}\right)+\frac{4\eta}{(1-\eta)^2}\ln\left(\frac{e}{\pi^*_{\min}}\right)\leq 1$$

that is easy to verify. Notice that $f(\eta)=\left[\frac{1+\eta}{1-\eta}\right]^2\ln\left(\frac{1}{1-\eta}\right)+\frac{4\eta}{(1-\eta)^2}\ln\left(\frac{e}{\pi^*_{\min}}\right)$ is an increasing and continuous function of η on interval $[0,1)$, and $f(0)=0$, $\lim_{\eta\uparrow 1}f(\eta)=\infty$, thus, there exists an $\eta^*\in(0,1)$ such that $f(\eta^*)=1$. Similarly, we can show that $g(\eta)=\left[\frac{1+\eta}{1-\eta}\right]^2\ln\left(\frac{1}{1-\eta}\right)+\frac{4\eta}{(1-\eta)^2}$ is an increasing and continuous function of η, and $g(0.1378)=1$.

We now show that if $\eta \leq \frac{0.1378}{\ln\left(\frac{e}{\pi^*_{\min}}\right)}$, then $f(\eta) \leq 1$. Notice that since $\ln\left(\frac{e}{\pi^*_{\min}}\right) \geq 1$, then we have $\frac{0.1378}{\ln\left(\frac{e}{\pi^*_{\min}}\right)} \leq 0.1378$. Thus we have, for $\eta \leq \frac{0.1378}{\ln\left(\frac{e}{\pi^*_{\min}}\right)}$,

$$f(\eta) \leq f\left[\frac{0.1378}{\ln\left(\frac{e}{\pi^*_{\min}}\right)}\right] \leq g(0.1378) = 1. \tag{9}$$

Thus, one sufficient condition for $\left[\frac{1+\eta}{1-\eta}\right]^2 \ln\left(\frac{1}{1-\eta}\right) + \frac{4\eta}{(1-\eta)^2}\ln\left(\frac{e}{\pi^*_{\min}}\right) \leq 1$ is that $\eta \leq \frac{0.1378}{\ln\left(\frac{e}{\pi^*_{\min}}\right)}$.

We define τ_G as

$$\tau_G = \min\left\{t \geq 4 : \frac{\ln(t)}{t} \leq \left(\frac{0.0689\pi^*_{\min}}{\ln\left(\frac{e}{\pi^*_{\min}}\right)}\right)^2 \text{ and } \frac{4}{3}\alpha_0 \max_{i\in\mathcal{I}}\left|\frac{\alpha(i)}{\alpha_0} - \pi^*(i)\right| \leq [t\ln(t)]^{\frac{1}{2}}\right\}. \tag{10}$$

Lemma A-7 derives a more useful one-game regret bound based on Lemma A-5 and definition of τ_G:

Lemma A-7: $\forall t \geq \tau_G$, *we have that*

$$\begin{aligned}&\mathbb{E}_{T^{\mathrm{g}}_t}\left\{\mathbb{E}_{i\sim\pi^*}\left[N(T^{\mathrm{g}}_t, i)\right] - \ln\left(\frac{e}{\pi^*_{\min}}\right)\mathbb{E}_{i\sim\pi^*}\left[N(T^*, i)\right]\right\} \\ < \quad & \left[8 + 12\ln\left(\frac{e}{\pi^*_{\min}}\right)\right]\frac{1}{\pi^*_{\min}}\left[\frac{\ln(t)}{t}\right]^{\frac{1}{2}}\mathbb{E}_{i\sim\pi^*}\left[N(T^*, i)\right] + \frac{2M|\mathcal{Q}|}{t^2}.\end{aligned}$$

Proof:

For $t \geq \tau_G$, we choose $\eta = \frac{2}{\pi^*_{\min}}\left[\frac{\ln(t)}{t}\right]^{\frac{1}{2}}$. We first show that this particular η satisfies the conditions of Lemma A-6. Since $\frac{\ln(t)}{t}$ is monotonically decreasing for $t \geq 3$ and $t \geq \tau_G \geq 4$, we have that

$$\eta = \frac{2}{\pi^*_{\min}}\left[\frac{\ln(t)}{t}\right]^{\frac{1}{2}} \leq \frac{2}{\pi^*_{\min}}\left[\frac{\ln(\tau_G)}{\tau_G}\right]^{\frac{1}{2}} \leq \frac{0.1378}{\ln\left(\frac{e}{\pi^*_{\min}}\right)}.$$

From the discussion above, we have $\left[\frac{1+\eta}{1-\eta}\right]^2 \ln\left(\frac{1}{1-\eta}\right) + \frac{4\eta}{(1-\eta)^2}\ln\left(\frac{e}{\pi^*_{\min}}\right) \leq 1$ for $\eta = \frac{2}{\pi^*_{\min}}\left[\frac{\ln(t)}{t}\right]^{\frac{1}{2}}$.

On the other hand, since $t\ln(t)$ is monotonically increasing, thus, $t \geq \tau_G$ implies that

$$\frac{4}{3}\alpha_0 \max_{i\in\mathcal{I}}\left|\frac{\alpha(i)}{\alpha_0} - \pi^*(i)\right| \leq [t\ln(t)]^{\frac{1}{2}}.$$

Similarly as the proof for Lemma A-4, we have that

$$\frac{\alpha_0}{\alpha_0 + t}\max_{i\in\mathcal{I}}\left|\frac{\alpha(i)}{\alpha_0} - \pi^*(i)\right| < \frac{\alpha_0}{t}\max_{i\in\mathcal{I}}\left|\frac{\alpha(i)}{\alpha_0} - \pi^*(i)\right| \leq \frac{3[t\ln(t)]^{\frac{1}{2}}}{4t} = \frac{2}{\pi^*_{\min}}\left[\frac{\ln(t)}{t}\right]^{\frac{1}{2}}\frac{3\pi^*_{\min}}{8} = \frac{3\pi^*_{\min}}{8}\eta < \eta\pi^*_{\min}.$$

Thus, the conditions of Lemma A-6 are satisfied. Furthermore, similarly as the proof for Lemma A-4, we have that

$$\exp\left\{4\eta\pi^*_{\min}\alpha_0\max_{i\in\mathcal{I}}\left|\frac{\alpha(i)}{\alpha_0} - \pi^*(i)\right| - 2\left[\eta\pi^*_{\min}\right]^2 t\right\} \leq \exp\{6\ln(t) - 8\ln(t)\} = \exp\{-2\ln(t)\} = \frac{1}{t^2}$$

for $t \geq \tau_G$.

We now bound the term $\left[\frac{1+\eta}{1-\eta}\right]^2 \ln\left(\frac{1}{1-\eta}\right) + \frac{4\eta}{(1-\eta)^2} \ln\left(\frac{e}{\pi^*_{\min}}\right)$. Notice that for $t \geq \tau_G$, we have that

$$\eta = \frac{2}{\pi^*_{\min}} \left[\frac{\ln(t)}{t}\right]^{\frac{1}{2}} \leq \frac{0.1378}{\ln\left(\frac{e}{\pi^*_{\min}}\right)} \leq 0.1378.$$

Thus we have $\left[\frac{1+\eta}{1-\eta}\right]^2 \leq 1.7415 < 2$, and $\frac{1}{(1-\eta)^2} \leq 1.3452 < 1.5$. On the other hand, notice that $\ln\left(\frac{1}{1-\eta}\right) \leq 2\eta$ for $0 \leq \eta \leq \frac{1}{2}$, thus, for $\eta \leq 0.1378$, we have that

$$\left[\frac{1+\eta}{1-\eta}\right]^2 \ln\left(\frac{1}{1-\eta}\right) + \frac{4\eta}{(1-\eta)^2} \ln\left(\frac{e}{\pi^*_{\min}}\right) < 4\eta + 6\eta \ln\left(\frac{e}{\pi^*_{\min}}\right) = \left[4 + 6\ln\left(\frac{e}{\pi^*_{\min}}\right)\right] \frac{2}{\pi^*_{\min}} \left[\frac{\ln(t)}{t}\right]^{\frac{1}{2}}.$$

Combining the above inequalities and the result of Lemma A-6, we have that

$$\begin{aligned} & \mathbb{E}_{T^{\mathrm{g}}_t} \left\{ \mathbb{E}_{i\sim\pi^*}\left[N(T^{\mathrm{g}}_t, i)\right] - \ln\left(\frac{e}{\pi^*_{\min}}\right) \mathbb{E}_{i\sim\pi^*}\left[N(T^*, i)\right] \right\} \\ < & \left[4 + 6\ln\left(\frac{e}{\pi^*_{\min}}\right)\right] \frac{2}{\pi^*_{\min}} \left[\frac{\ln(t)}{t}\right]^{\frac{1}{2}} \mathbb{E}_{i\sim\pi^*}\left[N(T^*, i)\right] + \frac{2M|\mathcal{Q}|}{t^2}, \end{aligned}$$

for $t \geq \tau_G$. Q.E.D.

Finally, we prove Theorem 2.

Proof of Theorem 2:

The proof is similar to Theorem 1. Specifically, for $0 \leq \tau < \tau_G$, we have that

$$\sum_{t=0}^{\tau} \mathbb{E}_{T^{\mathrm{g}}_t} \left\{ \mathbb{E}_{i\sim\pi^*}\left[N(T^{\mathrm{g}}_t, i)\right] - \ln\left(\frac{e}{\pi^*_{\min}}\right) \mathbb{E}_{i\sim\pi^*}\left[N(T^*, i)\right] \right\} \leq |\mathcal{Q}|(\tau + 1).$$

On the other hand, from Lemma A-7, for $\tau \geq \tau_G$, we have that

$$\begin{aligned} & \sum_{t=0}^{\tau} \mathbb{E}_{T^{\mathrm{g}}_t} \left\{ \mathbb{E}_{i\sim\pi^*}\left[N(T^{\mathrm{g}}_t, i)\right] - \ln\left(\frac{e}{\pi^*_{\min}}\right) \mathbb{E}_{i\sim\pi^*}\left[N(T^*, i)\right] \right\} \\ = & \sum_{t=0}^{\tau_G - 1} \mathbb{E}_{T^{\mathrm{g}}_t} \left\{ \mathbb{E}_{i\sim\pi^*}\left[N(T^{\mathrm{g}}_t, i)\right] - \ln\left(\frac{e}{\pi^*_{\min}}\right) \mathbb{E}_{i\sim\pi^*}\left[N(T^*, i)\right] \right\} \\ + & \sum_{t=\tau_G}^{\tau} \mathbb{E}_{T^{\mathrm{g}}_t} \left\{ \mathbb{E}_{i\sim\pi^*}\left[N(T^{\mathrm{g}}_t, i)\right] - \ln\left(\frac{e}{\pi^*_{\min}}\right) \mathbb{E}_{i\sim\pi^*}\left[N(T^*, i)\right] \right\} \\ \leq & |\mathcal{Q}|\tau_G + \sum_{t=\tau_G}^{\tau} \mathbb{E}_{T^{\mathrm{g}}_t} \left\{ \mathbb{E}_{i\sim\pi^*}\left[N(T^{\mathrm{g}}_t, i)\right] - \ln\left(\frac{e}{\pi^*_{\min}}\right) \mathbb{E}_{i\sim\pi^*}\left[N(T^*, i)\right] \right\} \\ \leq & |\mathcal{Q}|\tau_G + \sum_{t=\tau_G}^{\tau} \left[\left[8 + 12\ln\left(\frac{e}{\pi^*_{\min}}\right)\right] \frac{1}{\pi^*_{\min}} \left[\frac{\ln(t)}{t}\right]^{\frac{1}{2}} \mathbb{E}_{i\sim\pi^*}\left[N(T^*, i)\right] + \frac{2M|\mathcal{Q}|}{t^2} \right], \end{aligned}$$

where the first inequality follows from the naive bound and the second inequality follows from Lemma A-7. Since $\tau_G > 1$, we have that

$$\sum_{t=\tau_G}^{\tau} \frac{1}{t^2} < \sum_{t=\tau_G}^{\infty} \frac{1}{t^2} < \sum_{t=\tau_G}^{\infty} \frac{1}{(t-1)t} = \sum_{t=\tau_G}^{\infty} \left[\frac{1}{t-1} - \frac{1}{t}\right] = \frac{1}{\tau_G - 1}.$$

On the other hand, notice that $\frac{\ln(t)}{t}$ is monotonically decreasing on interval $[\tau_G - 1, \infty)$ (Since $\tau_G - 1 \geq 3$), and the derivative of the function $[t\ln(t)]^{\frac{1}{2}}$ is $\frac{1}{2}\left[\frac{1}{t\ln(t)}\right]^{\frac{1}{2}} + \frac{1}{2}\left[\frac{\ln(t)}{t}\right]^{\frac{1}{2}}$, we have that

$$\sum_{t=\tau_G}^{\tau}\left[\frac{\ln(t)}{t}\right]^{\frac{1}{2}} < \int_{\tau_G-1}^{\tau}\left[\frac{\ln(t)}{t}\right]^{\frac{1}{2}}dt < \int_{\tau_G-1}^{\tau}\left(\left[\frac{\ln(t)}{t}\right]^{\frac{1}{2}} + \left[\frac{1}{t\ln(t)}\right]^{\frac{1}{2}}\right)dt = 2\left[\tau\ln(\tau)\right]^{\frac{1}{2}} - 2\left[(\tau_G-1)\ln(\tau_G-1)\right]^{\frac{1}{2}}.$$

Thus, for $\tau \geq \tau_G$, we have that

$$\begin{aligned}
& \sum_{t=0}^{\tau}\mathbb{E}_{T_t^{\mathrm{g}}}\left\{\mathbb{E}_{i\sim\pi^*}\left[N(T_t^{\mathrm{g}}, i)\right] - \ln\left(\frac{e}{\pi^*_{\min}}\right)\mathbb{E}_{i\sim\pi^*}\left[N(T^*, i)\right]\right\} \\
\leq\ & |\mathcal{Q}|\tau_G + \sum_{t=\tau_G}^{\tau}\left[\left[8 + 12\ln\left(\frac{e}{\pi^*_{\min}}\right)\right]\frac{1}{\pi^*_{\min}}\left[\frac{\ln(t)}{t}\right]^{\frac{1}{2}}\mathbb{E}_{i\sim\pi^*}\left[N(T^*, i)\right] + \frac{2M|\mathcal{Q}|}{t^2}\right] \\
<\ & |\mathcal{Q}|\tau_G + \left[16 + 24\ln\left(\frac{e}{\pi^*_{\min}}\right)\right]\frac{\mathbb{E}_{i\sim\pi^*}\left[N(T^*, i)\right]}{\pi^*_{\min}}\left\{\left[\tau\ln(\tau)\right]^{\frac{1}{2}} - \left[(\tau_G-1)\ln(\tau_G-1)\right]^{\frac{1}{2}}\right\} + \frac{2M|\mathcal{Q}|}{\tau_G - 1} \\
=\ & O\left(\left[\tau\ln(\tau)\right]^{\frac{1}{2}}\right).
\end{aligned}$$

Notice that $\ln\left(\frac{e}{\pi^*_{\min}}\right) \geq 1$, so we have

$$\begin{aligned}
& \sum_{t=0}^{\tau}\mathbb{E}_{T_t^{\mathrm{g}}}\left\{\mathbb{E}_{i\sim\pi^*}\left[N(T_t^{\mathrm{g}}, i)\right] - \ln\left(\frac{e}{\pi^*_{\min}}\right)\mathbb{E}_{i\sim\pi^*}\left[N(T^*, i)\right]\right\} \\
<\ & |\mathcal{Q}|\tau_G + 40\ln\left(\frac{e}{\pi^*_{\min}}\right)\frac{\mathbb{E}_{i\sim\pi^*}\left[N(T^*, i)\right]}{\pi^*_{\min}}\left\{\left[\tau\ln(\tau)\right]^{\frac{1}{2}} - \left[(\tau_G-1)\ln(\tau_G-1)\right]^{\frac{1}{2}}\right\} + \frac{2M|\mathcal{Q}|}{\tau_G - 1} \\
=\ & O\left(\left[\tau\ln(\tau)\right]^{\frac{1}{2}}\right).
\end{aligned}$$

Q.E.D.

References

[1] Daniel Golovin and Andreas Krause. Adaptive submodularity: Theory and applications in active learning and stochastic optimization. *Journal of Artificial Intelligence Research*, 42:427–486, 2011.