Online Companion: Adaptive Sequential Experiments with Unknown Information Arrival Processes

Yonatan Gur
Stanford University

Ahmadreza Momeni
Stanford University

May 3, 2022

A Proofs of Main Results

A.1 Notation

For any policy π and profiles $\boldsymbol{\nu}$ and $\boldsymbol{\nu}'$, let $\mathbb{P}^{\pi}_{\boldsymbol{\nu},\boldsymbol{\nu}^{\text{aux}}}$, $\mathbb{E}^{\pi}_{\boldsymbol{\nu},\boldsymbol{\nu}^{\text{aux}}}$, and $\mathbb{R}^{\pi}_{\boldsymbol{\nu},\boldsymbol{\nu}^{\text{aux}}}$ denote the probability, expectation, and regret when rewards are distributed according to $\boldsymbol{\nu}$, and auxiliary observations are distributed according to $\boldsymbol{\nu}^{\text{aux}}$. Notation of counters and empirical means is provided in the following table:

Table 2: Notation for counters and empirical means (extended)

Definition	**Description**
$n^{\text{known}}_{k,t} \coloneqq \sum_{s=1}^{t-1} \mathbb{1}\{\pi_s = k\} + \sum_{s=1}^{t} \frac{\sigma^2}{\hat{\sigma}^2} h_{k,s}$	Number of pulls (known mappings case)
$\bar{X}^{\text{known}}_{k,t} \coloneqq \bar{X}^{\text{known}}_{k,n^{\text{known}}_{k,t}} \coloneqq \frac{\sum_{s=1}^{t-1} \frac{1}{\sigma^2}\mathbb{1}\{\pi_s=k\}X_{k,s} + \sum_{s=1}^{t} \frac{1}{\hat{\sigma}^2} h_{k,s} Z_{k,s}}{\sum_{s=1}^{t-1} \frac{1}{\sigma^2}\mathbb{1}\{\pi_s=k\} + \sum_{s=1}^{t} \frac{1}{\hat{\sigma}^2} h_{k,s}}$	Number of pulls (known mappings case)
$n^{\pi}_{k,t} \coloneqq \sum_{s=1}^{t-1} \mathbb{1}\{\pi_s = k\}$	Number of pulls
$\bar{X}^{\pi}_{k,t} \coloneqq \bar{X}^{\pi}_{k,n^{\pi}_{k,t}} \coloneqq \left(\sum_{s=1}^{t-1} \mathbb{1}\{\pi_s = k\} X_{k,s}\right) / \left(\max\{1, n^{\pi}_{k,t}\}\right)$	Empirical mean reward
$n^{\text{aux}}_{k,t} \coloneqq \sum_{s=1}^{t} h_{k,s}$	Number of auxiliary observations
$\bar{Y}_{k,t} \coloneqq \bar{Y}_{k,n^{\text{aux}}_{k,t}} \coloneqq \left(\sum_{s=1}^{t} \sum_{m=1}^{h_{k,s}} Y_{k,s,m}\right) / \left(\max\{1, n^{\text{aux}}_{k,t}\}\right)$	Empirical mean of auxiliary obs.
$n^{\pi,\text{aux}}_{k,t} \coloneqq n^{\pi}_{k,t} + \frac{\sigma^2}{\bar{\alpha}^2\hat{\sigma}^2} n^{\text{aux}}_{k,t}$	Weighted number of observations
$\bar{X}^{\pi,\text{aux}}_{k,t} \coloneqq \bar{X}^{\pi,\text{aux}}_{k,n^{\pi}_{k,t},n^{\text{aux}}_{k,t}} \coloneqq \left(n^{\pi}_{k,t}\bar{X}^{\pi}_{k,n^{\pi}_{k,t}} + \frac{\sigma^2}{\bar{\alpha}^2\hat{\sigma}^2} n^{\text{aux}}_{k,t}(\bar{\alpha}\bar{Y}_{k,t} + \bar{\beta})\right) / \left(\max\{1, n^{\pi,\text{aux}}_{k,t}\}\right)$	Optimistic empirical mean reward

*Correspondence: ygur@stanford.edu, amomenis@stanford.edu.

A.2 Proof of Theorem 1

Step 1 (Preliminaries). The proof adapts ideas of identifying worst-case nature strategy (see, e.g., Bubeck et al. 2013) to our setting in order to identify the precise change in the achievable performance as a function of the entries of information arrival matrix $\mathbf{H}$. For $m, q \in \{1, \ldots, K\}$ define the distribution profiles $\boldsymbol{\nu}^{(m,q)}$:

$$\nu_k^{(m,q)} = \begin{cases} \mathcal{N}(0, \sigma^2) & \text{if } k = m \\ \mathcal{N}(+\Delta, \sigma^2) & \text{if } k = q \neq m \\ \mathcal{N}(-\Delta, \sigma^2) & \text{o.w.} \end{cases}.$$

For example, for $m = 1$, one has

$$\boldsymbol{\nu}^{(1,1)} = \begin{pmatrix} \mathcal{N}(0, \sigma^2) \\ \mathcal{N}(-\Delta, \sigma^2) \\ \mathcal{N}(-\Delta, \sigma^2) \\ \vdots \\ \mathcal{N}(-\Delta, \sigma^2) \end{pmatrix}, \; \boldsymbol{\nu}^{(1,2)} = \begin{pmatrix} \mathcal{N}(0, \sigma^2) \\ \mathcal{N}(+\Delta, \sigma^2) \\ \mathcal{N}(-\Delta, \sigma^2) \\ \vdots \\ \mathcal{N}(-\Delta, \sigma^2) \end{pmatrix}, \; \boldsymbol{\nu}^{(1,3)} = \begin{pmatrix} \mathcal{N}(0, \sigma^2) \\ \mathcal{N}(-\Delta, \sigma^2) \\ \mathcal{N}(+\Delta, \sigma^2) \\ \vdots \\ \mathcal{N}(-\Delta, \sigma^2) \end{pmatrix}, \ldots, \; \boldsymbol{\nu}^{(1,K)} = \begin{pmatrix} \mathcal{N}(0, \sigma^2) \\ \mathcal{N}(-\Delta, \sigma^2) \\ \mathcal{N}(-\Delta, \sigma^2) \\ \vdots \\ \mathcal{N}(+\Delta, \sigma^2) \end{pmatrix}.$$

Similarly, assume that the auxiliary information $Y_{k,t,m}$ is distributed according to the reward distribution $\hat{\nu}_k^{(m,q)} = \nu_k^{(m,q)}$, and hence we use the notation $\mathbb{P}_{\boldsymbol{\nu}}^{\pi}$, $\mathbb{E}_{\boldsymbol{\nu}}^{\pi}$, and $\mathbb{R}_{\boldsymbol{\nu}}^{\pi}$ instead of $\mathbb{P}_{\boldsymbol{\nu},\boldsymbol{\nu}^{\text{aux}}}^{\pi}$, $\mathbb{E}_{\boldsymbol{\nu},\boldsymbol{\nu}^{\text{aux}}}^{\pi}$, and $\mathbb{R}_{\boldsymbol{\nu},\boldsymbol{\nu}^{\text{aux}}}^{\pi}$.

Step 2 (Lower bound decomposition). We note that

$$\mathcal{R}_{\mathcal{S}}^{\pi}(\mathbf{H}, T) \geq \max_{m,q \in \{1,\ldots,K\}} \left\{ \mathcal{R}_{\boldsymbol{\nu}^{(m,q)}}^{\pi}(\mathbf{H}, T) \right\} \geq \frac{1}{K} \sum_{m=1}^{K} \max_{q \in \{1,\ldots,K\}} \left\{ \mathcal{R}_{\boldsymbol{\nu}^{(m,q)}}^{\pi}(\mathbf{H}, T) \right\}. \tag{4}$$

Step 3 (A naive lower bound for $\max_{q \in \{1,\ldots,K\}} \{\mathcal{R}_{\boldsymbol{\nu}^{(m,q)}}^{\pi}(\mathbf{H}, T)\}$). We note that

$$\max_{q \in \{1,\ldots,K\}} \{\mathcal{R}_{\boldsymbol{\nu}^{(m,q)}}^{\pi}(\mathbf{H}, T)\} \geq \mathcal{R}_{\boldsymbol{\nu}^{(m,m)}}^{\pi}(\mathbf{H}, T) = \Delta \cdot \sum_{k \in \mathcal{K} \setminus \{m\}} \mathbb{E}_{\boldsymbol{\nu}^{(m,m)}}[n_{k,T+1}^{\pi}]. \tag{5}$$

Step 4 (An information theoretic lower bound). For any profile $\boldsymbol{\nu}$, denote by $\boldsymbol{\nu}_t$ the distribution of the observed rewards up to t under $\boldsymbol{\nu}$. By Lemma 6, for any $q \neq m$, one has

$$\text{KL}(\boldsymbol{\nu}_t^{(m,m)}, \boldsymbol{\nu}_t^{(m,q)}) = \frac{2\Delta^2}{\sigma^2} \cdot \mathbb{E}_{\boldsymbol{\nu}^{(m,m)}}[n_{q,t}] = \frac{2\Delta^2}{\sigma^2} \left(\mathbb{E}_{\boldsymbol{\nu}^{(m,m)}}[n_{q,t}^{\pi}] + \sum_{s=1}^{t} \frac{\sigma^2}{\hat{\sigma}^2} h_{q,s} \right). \tag{6}$$

One obtains:

$$
\begin{aligned}
\max_{q\in\{1,\dots,K\}}\{\mathcal{R}^{\pi}_{\boldsymbol{\nu}^{(m,q)}}(\mathbf{H},T)\} &\geq \frac{1}{K}\mathcal{R}^{\pi}_{\boldsymbol{\nu}^{(m,m)}}(\mathbf{H},T)+\frac{1}{K}\sum_{q\in\mathcal{K}\setminus\{m\}}\mathcal{R}^{\pi}_{\boldsymbol{\nu}^{(m,q)}}(\mathbf{H},T)\\
&\geq \frac{\Delta}{K}\sum_{t=1}^{T}\sum_{k\in\mathcal{K}\setminus\{m\}}\mathbb{P}_{\boldsymbol{\nu}^{(m,m)}}\{\pi_t=k\}+\frac{\Delta}{K}\sum_{q\in\mathcal{K}\setminus\{m\}}\sum_{t=1}^{T}\mathbb{P}_{\boldsymbol{\nu}^{(m,q)}}\{\pi_t\neq q\}\\
&= \frac{\Delta}{K}\sum_{t=1}^{T}\sum_{q\in\mathcal{K}\setminus\{m\}}\left(\mathbb{P}_{\boldsymbol{\nu}^{(m,m)}}\{\pi_t=q\}+\mathbb{P}_{\boldsymbol{\nu}^{(m,q)}}\{\pi_t\neq q\}\right)\\
&\overset{(a)}{\geq} \frac{\Delta}{2K}\sum_{t=1}^{T}\sum_{q\in\mathcal{K}\setminus\{m\}}\exp(-\mathrm{KL}(\boldsymbol{\nu}_t^{(m,m)},\boldsymbol{\nu}_t^{(m,q)}))\\
&\overset{(b)}{=} \frac{\Delta}{2K}\sum_{t=1}^{T}\sum_{q\in\mathcal{K}\setminus\{m\}}\exp\left[-\frac{2\Delta^2}{\sigma^2}\left(\mathbb{E}_{\boldsymbol{\nu}^{(m,m)}}[n^{\pi}_{q,t-1}]+\sum_{s=1}^{t}\frac{\sigma^2}{\hat{\sigma}^2}h_{q,s}\right)\right]\\
&= \sum_{q\in\mathcal{K}\setminus\{m\}}\frac{\Delta\cdot\exp\left(-\frac{2\Delta^2}{\hat{\sigma}^2}\cdot\mathbb{E}_{\boldsymbol{\nu}^{(m,m)}}[n^{\pi}_{q,T+1}]\right)}{2K}\sum_{t=1}^{T}\exp\left(-\frac{2\Delta^2}{\sigma^2}\sum_{s=1}^{t}h_{q,s}\right), \qquad (7)
\end{aligned}
$$

where (a) follows from Lemma 7, (b) holds by (6) and the fact that $n^{\pi}_{q,T+1}\geq n^{\pi}_{q,t}$ for $t\in\mathcal{T}$.

Step 5 (Unifying the lower bounds in steps 3 and 4). Using (5), and (7), we establish

$$
\begin{aligned}
&\max_{q\in\{1,\dots,K\}}\{\mathcal{R}^{\pi}_{\boldsymbol{\nu}^{(m,q)}}(\mathbf{H},T)\}\\
&\quad\geq \frac{\Delta}{2}\sum_{k\in\mathcal{K}\setminus\{m\}}\left(\mathbb{E}_{\boldsymbol{\nu}^{(m,m)}}[n^{\pi}_{k,T+1}]+\frac{\exp\left(-\frac{2\Delta^2}{\sigma^2}\cdot\mathbb{E}_{\boldsymbol{\nu}^{(m,m)}}[n^{\pi}_{k,T+1}]\right)}{2K}\sum_{t=1}^{T}\exp\left(-\frac{2\Delta^2}{\hat{\sigma}^2}\sum_{s=1}^{t}h_{k,s}\right)\right)\\
&\quad\geq \frac{\Delta}{2}\sum_{k\in\mathcal{K}\setminus\{m\}}\min_{x\geq 0}\left(x+\frac{\exp\left(-\frac{2\Delta^2}{\sigma^2}\cdot x\right)}{2K}\sum_{t=1}^{T}\exp\left(-\frac{2\Delta^2}{\hat{\sigma}^2}\sum_{s=1}^{t}h_{k,s}\right)\right)\\
&\quad\overset{(a)}{\geq} \frac{\sigma^2}{4\Delta}\sum_{k\in\mathcal{K}\setminus\{m\}}\log\left(\frac{\Delta^2}{\sigma^2 K}\sum_{t=1}^{T}\exp\left(-\frac{2\Delta^2}{\hat{\sigma}^2}\sum_{s=1}^{t}h_{k,s}\right)\right), \qquad (8)
\end{aligned}
$$

where (a) follows from $x+\gamma e^{-\kappa x}\geq\frac{\log\gamma\kappa}{\kappa}$ for $\gamma,\kappa,x>0$. (Note that the function $x+\gamma e^{-\kappa x}$ is a convex function and we can find its minimum by finding the root of its derivative) The result is then established by putting together (4), and (8). ■

A.3 Proof of Theorem 2

Fix a problem instance $(\boldsymbol{\nu},\boldsymbol{\nu}^{\mathrm{aux}})\in\mathcal{S}$ with the mean rewards $\boldsymbol{\mu}$ and the mean $\mathbf{y}$ for auxiliary observations. Consider a suboptimal arm $k\neq k^*$. If arm k is pulled at epoch t, then $\bar{X}^{\mathrm{known}}_{k,t}+\sqrt{\frac{c\sigma^2\log t}{n^{\mathrm{known}}_{k,t}}}\geq\bar{X}^{\mathrm{known}}_{k^*,t}+\sqrt{\frac{c\sigma^2\log t}{n^{\mathrm{known}}_{k^*,t}}}$. Therefore, at least one of the following three events must occur:

$$\mathcal{E}_{1,t} \coloneqq \left\{ \bar{X}_{k,t}^{\text{known}} \geq \mu_k + \sqrt{\frac{c\sigma^2 \log t}{n_{k,t}^{\text{known}}}} \right\}, \quad \mathcal{E}_{2,t} \coloneqq \left\{ \bar{X}_{k^*,t}^{\text{known}} \leq \mu_{k^*} - \sqrt{\frac{c\sigma^2 \log t}{n_{k^*,t}^{\text{known}}}} \right\}, \quad \mathcal{E}_{3,t} \coloneqq \left\{ \Delta_k \leq 2\sqrt{\frac{c\sigma^2 \log t}{n_{k,t}^{\text{known}}}} \right\}.$$

To see why this is true, assume that all the above events fail. Then, we have

$$\bar{X}_{k,t}^{\text{known}} + \sqrt{\frac{c\sigma^2 \log t}{n_{k,t}^{\text{known}}}} < \mu_k + 2\sqrt{\frac{c\sigma^2 \log t}{n_{k,t}^{\text{known}}}} < \mu_k + \Delta_k = \mu_{k^*} < \bar{X}_{k^*,t}^{\text{known}} + \sqrt{\frac{c\sigma^2 \log t}{n_{k^*,t}^{\text{known}}}},$$

which is in contradiction with the assumption that arm k is pulled at epoch t. For any sequence $\{l_{k,t}\}_{t\in\mathcal{T}}$, and $\{\hat{l}_{k,t}\}_{t\in\mathcal{T}}$, such that $\hat{l}_{k,t} \geq l_{k,t}$ for all $t \in \mathcal{T}$, one has

$$\begin{aligned}
\mathbb{E}^{\pi}_{\boldsymbol{\nu},\boldsymbol{\nu}^{\text{aux}}}\left[n^{\pi}_{k,T+1}\right] &= \mathbb{E}^{\pi}_{\boldsymbol{\nu},\boldsymbol{\nu}^{\text{aux}}}\left[\sum_{t=1}^{T} \mathbb{1}\left\{\pi_t = k\right\}\right] \\
&= \mathbb{E}^{\pi}_{\boldsymbol{\nu},\boldsymbol{\nu}^{\text{aux}}}\left[\sum_{t=1}^{T} \mathbb{1}\left\{\pi_t = k, n_{k,t}^{\text{known}} \leq l_{k,t}\right\} + \mathbb{1}\left\{\pi_t = k, n_{k,t}^{\text{known}} > l_{k,t}\right\}\right] \\
&\leq \mathbb{E}^{\pi}_{\boldsymbol{\nu},\boldsymbol{\nu}^{\text{aux}}}\left[\sum_{t=1}^{T} \mathbb{1}\left\{\pi_t = k, n_{k,t}^{\text{known}} \leq \hat{l}_{k,t}\right\} + \mathbb{1}\left\{\pi_t = k, n_{k,t}^{\text{known}} > l_{k,t}\right\}\right] \\
&\leq \mathbb{E}^{\pi}_{\boldsymbol{\nu},\boldsymbol{\nu}^{\text{aux}}}\left[\sum_{t=1}^{T} \mathbb{1}\left\{\pi_t = k, n_{k,t}^{\pi} \leq \hat{l}_{k,t} - \sum_{s=1}^{t} \frac{\sigma^2}{\hat{\sigma}^2} h_{k,s}\right\}\right] + \sum_{t=1}^{T} \mathbb{P}\left\{\pi_t = k, n_{k,t}^{\text{known}} > l_{k,t}\right\} \\
&\overset{(a)}{\leq} \max_{1\leq t\leq T}\left\{\hat{l}_{k,t} - \sum_{s=1}^{t} \frac{\sigma^2}{\hat{\sigma}^2} h_{k,s}\right\} + \sum_{t=1}^{T} \mathbb{P}\left\{\pi_t = k, n_{k,t}^{\text{known}} > l_{k,t}\right\},
\end{aligned}$$

where (a) follows from the following lemma:

Lemma 1 *Let $\{f_n\}_{n=1}^N$ and $\{g_n\}_{n=1}^N$ be two sequences such that for all $n \in \{1,\dots,N\}$, $f_n = \sum_{j=1}^{n-1} x_j$ with $0 \leq x_j \leq M$ for all $j \in \{1,\dots,N\}$ and some $M \geq 0$. Then, $\sum_{n=1}^N x_n \cdot \mathbb{1}\{f_n \leq g_n\} \leq \max\{0, \max_{1\leq n\leq N} g_n + M\}$.*

Proof. If $\max_{1\leq n\leq N} g_n + M < 0$ then, the result is immediate. Assume $\max_{1\leq n\leq N} g_n + M \geq 0$. We prove the result through contradiction. Suppose that $\sum_{n=1}^N x_n \cdot \mathbb{1}\{f_n \leq g_n\} > \max_{1\leq n\leq N} g_n + M$. Let

$$\tau = \min\left\{1 \leq n \leq N : \sum_{j=1}^{n} x_j \cdot \mathbb{1}\{f_j \leq g_j\} > \max_{1\leq n\leq N} g_n + M\right\}.$$

Note that one must have $f_\tau \leq g_\tau$. Since all x_j's are non-negative, one has

$$\sum_{j=1}^{\tau} x_j \cdot \mathbb{1}\{f_j \leq g_j\} \leq \sum_{j=1}^{\tau-1} x_j + x_\tau \leq f_\tau + M \leq g_\tau + M \leq \max_{1\leq n\leq N} g_n + M$$

which is a contradiction. This concludes the proof. ■

Set the values of $l_{k,t} \coloneqq \frac{4c\sigma^2 \log(t)}{\Delta_k^2}$ and $\hat{l}_{k,t} \coloneqq \frac{4c\sigma^2 \log\left(\tau_{k,t}\right)}{\Delta_k^2}$, where $\tau_{k,t} \coloneqq \sum_{s=1}^{t} \exp\left(\frac{\Delta_k^2}{4c\hat{\sigma}^2} \sum_{\tau=s}^{t} h_{k,\tau}\right)$. To have $n_{k,t}^{\text{known}} > l_{k,t}$, it must be the case that $\mathcal{E}_{3,t}$ does not occur. Therefore,

$$\begin{aligned}
\mathbb{E}^{\pi}_{\boldsymbol{\nu},\boldsymbol{\nu}^{\text{aux}}}\left[n^{\pi}_{k,T+1}\right] &\leq \max_{1\leq t\leq T}\left\{\hat{l}_{k,t} - \sum_{s=1}^{t}\frac{\sigma^2}{\hat{\sigma}^2}h_{k,s}\right\} + \sum_{t=1}^{T}\mathbb{P}\left\{\pi_t = k, \mathcal{E}_{3,t}^{c}\right\} \\
&\leq \max_{1\leq t\leq T}\left\{\hat{l}_{k,t} - \sum_{s=1}^{t}\frac{\sigma^2}{\hat{\sigma}^2}h_{k,s}\right\} + \sum_{t=1}^{T}\mathbb{P}\left\{\mathcal{E}_{1,t}\cup\mathcal{E}_{2,t}\right\} \\
&\leq \max_{1\leq t\leq T}\left\{\hat{l}_{k,t} - \sum_{s=1}^{t}\frac{\sigma^2}{\hat{\sigma}^2}h_{k,s}\right\} + \sum_{t=1}^{T}\frac{2}{t^{c/2-1}}, \qquad (9)
\end{aligned}$$

where the last inequality follows from Lemma 8 and the union bound. Plugging the value of $\hat{l}_{k,t}$ into (9), one obtains:

$$\begin{aligned}
\mathbb{E}^{\pi}_{\boldsymbol{\nu},\boldsymbol{\nu}^{\text{aux}}}\left[n^{\pi}_{k,T+1}\right] &\leq \max_{1\leq t\leq T}\left\{\frac{4c\sigma^2}{\Delta_k^2}\log\left(\sum_{s=1}^{t}\exp\left(\frac{\Delta_k^2}{4c\hat{\sigma}^2}\sum_{\tau=s}^{t}h_{k,\tau} - \frac{\Delta_k^2}{4c\hat{\sigma}^2}\sum_{\tau=1}^{t}h_{k,\tau}\right)\right)\right\} + \sum_{t=1}^{T}\frac{2}{t^{c/2-1}} \\
&\leq \frac{4c\sigma^2}{\Delta_k^2}\log\left(\sum_{t=1}^{T}\exp\left(\frac{-\Delta_k^2}{4c\hat{\sigma}^2}\sum_{s=1}^{t-1}h_{k,s}\right)\right) + \sum_{t=1}^{T}\frac{2}{t^{c/2-1}},
\end{aligned}$$

which concludes the proof. ■

A.4 Proof of Theorem 3

We adapt the proof of the upper bound for Thompson sampling with Gaussian priors in Agrawal and Goyal (2013a) to accomodate auxiliary observations.

Step1 (Notations and definitions). Fix a problem instance $(\boldsymbol{\nu}, \boldsymbol{\nu}^{\text{aux}}) \in \mathcal{S}$ with the mean rewards $\boldsymbol{\mu}$ and the mean $\mathbf{y}$ for auxiliary observations. For every suboptimal arm k, we consider three parameters x_k, y_k, and u_k such that $\mu_k < x_k < y_k < u_k < \mu_{k^*}$. We will specify these three parameters at the end of the proof. Also, define the events $\mathcal{E}_{k,t}^{\mu}$, and $\mathcal{E}_{k,t}^{\theta}$ to be the events on which $\bar{X}_{k,n_{k,t}^{\text{known}}} \leq x_k$, and $\theta_{k,t} \leq y_k$, respectively. In words, the events $\mathcal{E}_{k,t}^{\mu}$, and $\mathcal{E}_{k,t}^{\theta}$ happen when the estimated mean reward, and the sample mean reward do not deviate from the true mean, respectively. Define the history $\mathcal{H}_t \coloneqq \left(\{X_{\pi_s,s}\}_{s=1}^{t-1}, \{\pi_s\}_{s=1}^{t-1}, \{\mathbf{Z}_s\}_{s=1}^{t}, \{\boldsymbol{h}_s\}_{s=1}^{t}\right)$ for all $t = 1, \ldots$. Finally define $p_{k,t} \coloneqq \mathbb{P}^{\pi}_{\boldsymbol{\nu},\boldsymbol{\nu}^{\text{aux}}}\{\theta_{k^*,t} > y_k \mid \mathcal{H}_t\}$.

Step2 (Preliminaries). We use the following lemmas from Agrawal and Goyal (2013a) throughout this proof. The proofs are (mostly) skipped since they are simple adaptation of their analysis to our setting.

Lemma 2 *For any suboptimal arm k, $\sum_{t=1}^{T}\mathbb{P}^{\pi}_{\boldsymbol{\nu},\boldsymbol{\nu}^{\text{aux}}}\left\{\pi_t = k, \mathcal{E}_{k,t}^{\mu}, \mathcal{E}_{k,t}^{\theta}\right\} \leq \sum_{j=0}^{T-1}\mathbb{E}\left[\frac{1-p_{k,t_j+1}}{p_{k,t_j+1}}\right]$, where $t_0 = 0$ and for $j > 0$, t_j is the epoch at which the optimal arm k^* is pulled for the jth time.*

Proof. The proof can be found in the analysis of Theorem 1 in Agrawal and Goyal (2013a). However, we bring the proof for completeness.

$$\begin{aligned}
\sum_{t=1}^{T} \mathbb{P}^{\pi}_{\boldsymbol{\nu},\boldsymbol{\nu}^{\text{aux}}} \left\{ \pi_t = k, \mathcal{E}^{\mu}_{k,t}, \mathcal{E}^{\theta}_{k,t} \right\} &= \sum_{t=1}^{T} \mathbb{E}^{\pi}_{\boldsymbol{\nu},\boldsymbol{\nu}^{\text{aux}}} \left[\mathbb{P} \left\{ \pi_t = k, \mathcal{E}^{\mu}_{k,t}, \mathcal{E}^{\theta}_{k,t} \,\middle|\, \mathcal{H}_t \right\} \right] \\
&\overset{(a)}{\leq} \sum_{t=1}^{T} \mathbb{E}^{\pi}_{\boldsymbol{\nu},\boldsymbol{\nu}^{\text{aux}}} \left[\frac{1-p_{k,t}}{p_{k,t}} \mathbb{P} \left\{ \pi_t = 1, \mathcal{E}^{\mu}_{k,t}, \mathcal{E}^{\theta}_{k,t} \,\middle|\, \mathcal{H}_t \right\} \right] \\
&= \sum_{t=1}^{T} \mathbb{E}^{\pi}_{\boldsymbol{\nu},\boldsymbol{\nu}^{\text{aux}}} \left[\mathbb{E} \left[\frac{1-p_{k,t}}{p_{k,t}} \mathbb{1} \left\{ \pi_t = 1, \mathcal{E}^{\mu}_{k,t}, \mathcal{E}^{\theta}_{k,t} \,\middle|\, \mathcal{H}_t \right\} \right] \right] \\
&= \sum_{t=1}^{T} \mathbb{E}^{\pi}_{\boldsymbol{\nu},\boldsymbol{\nu}^{\text{aux}}} \left[\frac{1-p_{k,t}}{p_{k,t}} \mathbb{1} \left\{ \pi_t = 1, \mathcal{E}^{\mu}_{k,t}, \mathcal{E}^{\theta}_{k,t} \right\} \right] \\
&\leq \sum_{j=1}^{T-1} \mathbb{E}^{\pi}_{\boldsymbol{\nu},\boldsymbol{\nu}^{\text{aux}}} \left[\frac{1-p_{k,t_j+1}}{p_{k,t_j+1}} \sum_{t=t_j+1}^{t_{j+1}} \mathbb{1} \left\{ \pi_t = 1 \right\} \right] = \sum_{j=0}^{T-1} \mathbb{E} \left[\frac{1-p_{k,t_j+1}}{p_{k,t_j+1}} \right],
\end{aligned}$$

where (a) follows from Lemma 1 in Agrawal and Goyal (2013a). ■

Lemma 3 (Agrawal and Goyal 2013a, Lemma 2) *For any suboptimal arm k, $\sum_{t=1}^{T} \mathbb{P}^{\pi}_{\boldsymbol{\nu},\boldsymbol{\nu}^{\text{aux}}} \left\{ \pi_t = k, \bar{\mathcal{E}}^{\mu}_{k,t} \right\} \leq 1 + \frac{\sigma^2}{(x_k - \mu_k)^2}$.*

Step 3 (Regret decomposition). Fix a profile $\boldsymbol{\nu}$. For each suboptimal arm k, we will decompose the number of times it is pulled by the policy as follows and upper bound each term separately:

$$\mathbb{E}^{\pi}_{\boldsymbol{\nu},\boldsymbol{\nu}^{\text{aux}}} \left[n^{\pi}_{k,T+1} \right] = \underbrace{\sum_{t=1}^{T} \mathbb{P}^{\pi}_{\boldsymbol{\nu},\boldsymbol{\nu}^{\text{aux}}} \left\{ \pi_t = k, \mathcal{E}^{\mu}_{k,t}, \mathcal{E}^{\theta}_{k,t} \right\}}_{J_{k,1}} + \underbrace{\sum_{t=1}^{T} \mathbb{P}^{\pi}_{\boldsymbol{\nu},\boldsymbol{\nu}^{\text{aux}}} \left\{ \pi_t = k, \mathcal{E}^{\mu}_{k,t}, \bar{\mathcal{E}}^{\theta}_{k,t} \right\}}_{J_{k,2}} + \underbrace{\sum_{t=1}^{T} \mathbb{P}^{\pi}_{\boldsymbol{\nu},\boldsymbol{\nu}^{\text{aux}}} \left\{ \pi_t = k, \bar{\mathcal{E}}^{\mu}_{k,t} \right\}}_{J_{k,3}}. \tag{10}$$

Step 4 (Analysis of $J_{k,1}$). Given Lemma 2, $J_{k,1}$ can be upper bounded by analyzing $\frac{1-p_{k,t_j+1}}{p_{k,t_j+1}}$:

$$\begin{aligned}
p_{k,t_j+1} &= \mathbb{P} \left\{ \mathcal{N} \left(\bar{X}_{k*,n_{k*,t_j+1}}, c\sigma^2 (n_{k*,t_j+1}+1)^{-1} \right) > y_k \right\} \\
&\geq \mathbb{P} \left\{ \mathcal{N} \left(\bar{X}_{k*,n_{k*,t_j+1}}, \frac{c\sigma^2}{j+1} \right) > y_k \,\middle|\, \bar{X}_{k*,n_{k*,t_j+1}} \geq u_k \right\} \cdot \mathbb{P} \left\{ \bar{X}_{k*,n_{k*,t_j+1}} \geq u_k \right\} =: Q_1 \cdot Q_2.
\end{aligned}$$

First, we analyze Q_1:

$$Q_1 \geq \mathbb{P} \left\{ \mathcal{N} \left(u_k, \frac{c\sigma^2}{j+1} \right) > y_k \right\} \geq 1 - \exp \left(-\frac{(j+1)(u_k - y_k)^2}{2c\sigma^2} \right),$$

where the last inequality follows from Chernoff-Hoeffding bound in Lemma 8. The term Q_2 can also be bounded from below using Chernoff-Hoeffding in Lemma 8 bound as follows:

$$Q_2 \geq 1 - \exp\left(-\frac{j(\mu_1 - u_k)^2}{2\sigma^2}\right).$$

Define $\delta_k \coloneqq \min\{(\mu_1 - u_k), (u_k - y_k)\}$. The last three displays along with Lemma 3 yield

$$\begin{aligned}
J_{k,1} &\leq \frac{1}{\mathbb{P}\{\mathcal{N}(0, c\sigma^2) > y_k\}} - 1 + \sum_{j=1}^{T} \frac{2\exp(-j\delta_k^2/2\sigma^2(c \vee 1))}{1 - \exp(-j\delta_k^2/2\sigma^2(c \vee 1))} \\
&\leq \frac{1}{\mathbb{P}\{\mathcal{N}(0, c\sigma^2) > y_k\}} - 1 + \int_1^{\infty} \frac{2\exp(-x\delta_k^2/2\sigma^2(c \vee 1))}{1 - \exp(-x\delta_k^2/2\sigma^2(c \vee 1))} dx \\
&\leq \frac{1}{\mathbb{P}\{\mathcal{N}(0, c\sigma^2) > y_k\}} - 1 - \frac{4\sigma^2(c \vee 1)}{\delta_k^2} \log\left(1 - \exp(-\delta_k^2/2\sigma^2(c \vee 1))\right). \\
&\leq \frac{1}{\mathbb{P}\{\mathcal{N}(0, c\sigma^2) > y_k\}} - 1 + \frac{4\sigma^2(c \vee 1)}{\delta_k^2} \log\left(\frac{2\sigma^2(c \vee 1)}{\delta_k^2}\right). \qquad (11)
\end{aligned}$$

Step 5 (Analysis of $J_{k,2}$). First, we upper bound the following conditional probability

$$\begin{aligned}
\mathbb{P}^{\pi}_{\boldsymbol{\nu},\boldsymbol{\nu}^{\text{aux}}}\left\{\pi_t = k, \mathcal{E}^{\mu}_{k,t}, \bar{\mathcal{E}}^{\theta}_{k,t} \;\middle|\; n^{\text{known}}_{k,t}\right\} &\leq \mathbb{P}^{\pi}_{\boldsymbol{\nu},\boldsymbol{\nu}^{\text{aux}}}\left\{\mathcal{E}^{\mu}_{k,t}, \bar{\mathcal{E}}^{\theta}_{k,t} \;\middle|\; n^{\text{known}}_{k,t}\right\} \\
&\leq \mathbb{P}^{\pi}_{\boldsymbol{\nu},\boldsymbol{\nu}^{\text{aux}}}\left\{\bar{\mathcal{E}}^{\theta}_{k,t} \;\middle|\; \mathcal{E}^{\mu}_{k,t}, n^{\text{known}}_{k,t}\right\} \\
&= \mathbb{P}\left\{\mathcal{N}\left(\bar{X}^{\text{known}}_{k,t}, c\sigma^2\left(n^{\text{known}}_{k,t} + 1\right)^{-1}\right) > y_k \;\middle|\; \bar{X}^{\text{known}}_{k,t} \leq x_k, n^{\text{known}}_{k,t}\right\} \\
&\leq \mathbb{P}\left\{\mathcal{N}\left(x_k, c\sigma^2\left(n^{\text{known}}_{k,t} + 1\right)^{-1}\right) > y_k \;\middle|\; n^{\text{known}}_{k,t}\right\} \\
&\overset{(a)}{\leq} \exp\left(\frac{-n^{\text{known}}_{k,t}(y_k - x_k)^2}{2c\sigma^2}\right), \qquad (12)
\end{aligned}$$

where (a) follows from Chernoff-Hoeffding's bound in Lemma 8. Define

$$n^{\theta}_{k,t} \coloneqq \sum_{s=1}^{t} \frac{\sigma^2}{\hat{\sigma}^2} h_{k,s} + \sum_{s=1}^{t-1} \mathbb{1}\left\{\pi_s = k, \mathcal{E}^{\mu}_{k,s}, \bar{\mathcal{E}}^{\theta}_{k,s}\right\} \leq n^{\text{known}}_{k,t}.$$

By (12), and the definition above, one obtains:

$$\mathbb{P}^{\pi}_{\boldsymbol{\nu},\boldsymbol{\nu}^{\text{aux}}}\left\{\pi_t = k, \mathcal{E}^{\mu}_{k,t}, \bar{\mathcal{E}}^{\theta}_{k,t} \;\middle|\; n^{\theta}_{k,t}\right\} \leq \exp\left(-\frac{n^{\theta}_{k,t}(y_k - x_k)^2}{2c\sigma^2}\right)$$

for every $t \in \mathcal{T}$. Note that by definition,

$$n^{\theta}_{k,t} = n^{\theta}_{k,t-1} + \frac{\sigma^2}{\hat{\sigma}^2} h_{k,t} + \mathbb{1}\left\{\pi_{t-1} = k, \mathcal{E}^{\mu}_{k,t-1}, \bar{\mathcal{E}}^{\theta}_{k,t-1}\right\}.$$

By the above two displays, we can conclude that

$$\mathbb{E}^{\pi}_{\boldsymbol{\nu},\boldsymbol{\nu}^{\text{aux}}}\left[n^{\theta}_{k,t} \;\middle|\; n^{\theta}_{k,t-1}\right] = n^{\theta}_{k,t-1} + \frac{\sigma^2}{\hat{\sigma}^2}h_{k,t} + \mathbb{P}^{\pi}_{\boldsymbol{\nu},\boldsymbol{\nu}^{\text{aux}}}\left\{\pi_{t-1}=k, E^{\mu}_{k,t-1}, \bar{E}^{\theta}_{k,t-1} \;\middle|\; n^{\theta}_{k,t-1}\right\}$$
$$\leq n^{\theta}_{k,t-1} + \frac{\sigma^2}{\hat{\sigma}^2}h_{k,t} + e^{-\frac{n^{\theta}_{k,t-1}(y_k-x_k)^2}{2c\sigma^2}}.$$

The following lemma will be deployed to analyze $n^{\theta}_{k,T+1}$:

Lemma 4 *Fix an integer $T \geq 1$. For every $t \in \{1,\dots,T\}$, let $A_t \in \{0,1\}$ be random variables, and $w_t \geq 0$ be deterministic constants with $w_T = 0$. Define $N_t \coloneqq W_t + \sum_{s=1}^{t-1} A_s$, where $W_t \coloneqq \sum_{s=1}^{t} w_t$, and assume $\mathbb{E}[A_t \mid N_t] \leq p^{N_t+1}$ for some $p \leq 1$. Then, for every $t \in \{1,\dots,T+1\}$:*

$$\mathbb{E}[N_t] \leq \log_{\frac{1}{p}}\left(1 + \sum_{s=1}^{t-1} p^{W_s}\right) + W_t.$$

Proof. Consider the sequence $\{p^{-N_t} : t = 1,\dots,T\}$. For every $1 \leq t \leq T$,

$$\mathbb{E}\left[p^{-N_t} \;\middle|\; p^{-N_{t-1}}\right] = \mathbb{P}\{A_{t-1}=1 \mid N_{t-1}\}\cdot p^{-(N_{t-1}+w_t+1)} + \mathbb{P}\{A_{t-1}=0 \mid N_{t-1}\}\cdot p^{-(N_{t-1}+w_t)}$$
$$\leq p^{N_{t-1}+1}\cdot p^{-(N_{t-1}+w_t+1)} + p^{-(N_{t-1}+w_t)} = p^{-w_t}\cdot\left(1+p^{-N_{t-1}}\right).$$

Using this inequality and applying a simple induction, one obtains $\mathbb{E}\left[p^{-N_t}\right] \leq \sum_{s=1}^{t} p^{-\sum_{\tau=s}^{t} w_\tau}$, for every $t \in \{1,\dots,T\}$. Finally, by Jensen's inequality, one has $p^{-\mathbb{E}[N_t]} \leq \mathbb{E}\left[p^{-N_t}\right]$, which implies

$$\mathbb{E}[N_t] \leq \log_{\frac{1}{p}}\left(\sum_{s=1}^{t} p^{-\sum_{\tau=s}^{t} w_\tau}\right) = \log_{\frac{1}{p}}\left(1 + \sum_{s=1}^{t-1} p^{\sum_{\tau=1}^{s} w_\tau}\right) + \sum_{s=1}^{t} w_s.$$

This concludes the proof of the lemma. ■

Now, we can apply Lemma 4 to the sequence $\left\{n^{\theta}_{k,t}\right\}_{0\leq t\leq T}$, and obtain

$$\mathbb{E}^{\pi}_{\boldsymbol{\nu},\boldsymbol{\nu}^{\text{aux}}}\left[n^{\theta}_{k,T+1}\right] \leq \log_{\exp((y_k-x_k)^2/2c\sigma^2)}\left(1 + \sum_{t=1}^{T}\exp\left(-\frac{(y_k-x_k)^2}{2c\hat{\sigma}^2}\sum_{s=1}^{t} h_{k,s}\right)\right) + \sum_{t=1}^{T}\frac{\sigma^2}{\hat{\sigma}^2}h_{k,t}$$
$$\leq \frac{2c\sigma^2}{(y_k-x_k)^2}\log\left(1 + \sum_{t=1}^{T}\exp\left(-\frac{(y_k-x_k)^2}{2c\hat{\sigma}^2}\sum_{s=1}^{t} h_{k,s}\right)\right) + \sum_{t=1}^{T}\frac{\sigma^2}{\hat{\sigma}^2}h_{k,t}.$$

By this inequality and the definition of $n^{\theta}_{k,t}$, we have

$$J_{k,2} = \sum_{t=1}^{T}\mathbb{P}^{\pi}_{\boldsymbol{\nu},\boldsymbol{\nu}^{\text{aux}}}\left\{\pi_t = k, \mathcal{E}^{\mu}_{k,t}, \bar{\mathcal{E}}^{\theta}_{k,t}\right\} \leq \frac{2c\sigma^2}{(y_k-x_k)^2}\log\left(1 + \sum_{t=1}^{T}\exp\left(-\frac{(y_k-x_k)^2}{2c\hat{\sigma}^2}\sum_{s=1}^{t} h_{k,s}\right)\right). \tag{13}$$

Step 6 (Analysis of $J_{k,3}$). The term $J_{k,3}$ can be upper bounded using Lemma 3.

Step 7 (Determining the constants) Finally, let $x_k = \mu_k + \frac{\Delta_k}{4}$, $y_k = \mu_k + \frac{2\Delta_k}{4}$, and $u_k = \mu_k + \frac{3\Delta_k}{4}$. Then, by putting (10), (11), (13), and Lemma 3 back together, the result is established. ■

A.5 Proof of Theorem 4

In order to prove the theorem, we need to repeat the following argument for each arm k: Consider a new problem instance $(\tilde{\boldsymbol{\nu}}, \tilde{\boldsymbol{\nu}}^{\mathrm{aux}}) \in \mathcal{S}$ in Theorem 5, for which arm k is the optimal arm. The only difference between $(\tilde{\boldsymbol{\nu}}, \tilde{\boldsymbol{\nu}}^{\mathrm{aux}})$ and $(\boldsymbol{\nu}, \boldsymbol{\nu}^{\mathrm{aux}})$ is the distributions of rewards and auxiliary observations of arm k:

- If $\bar{\alpha} \cdot y_k + \bar{\beta} < \mu^*$ then, $(\tilde{\nu}_k, \tilde{\nu}_k^{\mathrm{aux}}) = (\mathcal{N}(\mu^* + \epsilon_k, \sigma^2), \mathcal{N}(\frac{\mu^* - \bar{\beta} + \epsilon_k}{\bar{\alpha}}, \hat{\sigma}^2))$ with $\epsilon_k = \delta_k = \mu^* - \bar{\alpha} \cdot y_k - \bar{\beta}$;
- If $\mu^* + \Delta_k \geq \bar{\alpha} \cdot y_k + \bar{\beta} > \mu^*$ then, $(\tilde{\nu}_k, \tilde{\nu}_k^{\mathrm{aux}}) = (\mathcal{N}(\mu^* + \epsilon_k, \sigma^2), \mathcal{N}(y_k, \hat{\sigma}^2))$ with $\epsilon_k = -\delta_k = \bar{\alpha} \cdot y_k + \bar{\beta} - \mu^*$;
- If $\bar{\alpha} \cdot y_k + \bar{\beta} > \mu^* + \Delta_k$ then, $(\tilde{\nu}_k, \tilde{\nu}_k^{\mathrm{aux}}) = (\mathcal{N}(\mu^* + \epsilon_k, \sigma^2), \mathcal{N}(y_k, \hat{\sigma}^2))$ with $\epsilon_k = \Delta_k$.

This concludes the proof. ■

A.6 Regret lower bound with unknown mappings and Bernoulli observations

Before providing the regret lower bound for the case with Bernoulli observations we need the following lemma which provides bounds on the KL divergence of Bernoulli distributions (see Lemma 4.1 in Rigollet and Zeevi (2010) for example).

Lemma 5 *For any pair of Bernoulli distributions ν_p and ν_q with parameters p and q, respectively, one has*

$$2(p-q)^2 \leq \mathrm{KL}(\nu_p, \nu_q) \leq \frac{(p-q)^2}{q(1-q)}.$$

In particular, if $q \in (\frac{1}{2} \pm \tau)$ for some $\tau \in (0, \frac{1}{2})$, then $\mathrm{KL}(\nu_p, \nu_q) \leq \frac{(p-q)^2}{\frac{1}{4} - \tau^2}$.

Proof. The lower bound follows from Pinsker's inequality. For the upper bound, one may note that

$$\mathrm{KL}(\nu_p, \nu_q) = p \log(\frac{p}{q}) + (1-p) \log(\frac{1-p}{1-q}) \leq p \left(\frac{p-q}{q}\right) - (1-p) \left(\frac{p-q}{1-q}\right) = \frac{(p-q)^2}{q(1-q)}.$$

■

Theorem 7 (Regret lower bound with unknown mappings and Bernoulli observations) *Assume that both reward and auxiliary observations are Bernoulli random variables. Let $\tau \in (0, \frac{1}{8}]$ and assume that for all arms $k \in \mathcal{K}$, $\mu_k \in (\frac{1}{2} - \tau, \frac{1}{2} + \tau)$ and $y_k \in (\frac{1}{2} - 4\tau, \frac{1}{2} + 4\tau)$ and that for any $\mu \in (\frac{1}{2} - 3\tau, \frac{1}{2} + 3\tau)$, one has $\frac{\mu - \bar{\beta}}{\bar{\alpha}} \in (\frac{1}{2} - 4\tau, \frac{1}{2} + 4\tau)$. Then, for any $T \geq 1$ and $k \in \mathcal{K}$ and for any policy $\pi \in \mathcal{P}^{\mathrm{IAO}}$, the followings hold with $\delta_k = \mu^* - \bar{\alpha} \cdot y_k - \bar{\beta}$:*

1. *If* $\bar{\alpha} \cdot y_k + \bar{\beta} > \mu^*$ *then,* $\mathbb{E}^{\pi}_{\boldsymbol{\nu},\boldsymbol{\nu}^{\text{aux}}_t}[n^{\pi}_{k,T+1}] \geq \frac{C_{15}}{\Delta_k^2} \log\left(\frac{C_{16}\min\{4\Delta_k^4,(\Delta_k-\delta_k)^2\delta_k^2\}}{K\log T}T\right)$;

2. *If* $\bar{\alpha} \cdot y_k + \bar{\beta} < \mu^*$ *then,* $\mathbb{E}^{\pi}_{\boldsymbol{\nu},\boldsymbol{\nu}^{\text{aux}}_t}[n^{\pi}_{k,T+1}] \geq \frac{C_{15}}{\Delta_k^2} \log\left(\frac{C_{16}(\Delta_k+\delta_k)^2\delta_k^2}{K\log T}\sum_{t=1}^{T}\exp\left(-C_{17}\delta_k^2\cdot\sum_{s=1}^{t}h_{k,s}\right)\right)$,

where C_{15}*,* C_{16}*, and* C_{17} *are positive constants that only depend on* τ*,* σ*,* $\hat{\sigma}$*, and* $\bar{\alpha}$*.*

Proof. In order to prove the theorem, we need to repeat the following argument for each arm k and apply Lemma 5: Consider a new problem instance $(\tilde{\boldsymbol{\nu}}, \tilde{\boldsymbol{\nu}}^{\text{aux}}) \in \mathcal{S}$ in Theorem 5, for which arm k is the optimal arm. The only difference between $(\tilde{\boldsymbol{\nu}}, \tilde{\boldsymbol{\nu}}^{\text{aux}})$ and $(\boldsymbol{\nu}, \boldsymbol{\nu}^{\text{aux}})$ is the distributions of rewards and auxiliary observations of arm k:

- If $\bar{\alpha} \cdot y_k + \bar{\beta} < \mu^*$ then, $(\tilde{\nu}_k, \tilde{\nu}_k^{\text{aux}}) = (\text{Ber}(\mu^* + \epsilon_k), \text{Ber}(\frac{\mu^* - \bar{\beta} + \epsilon_k}{\bar{\alpha}},))$ with $\epsilon_k = \delta_k = \mu^* - \bar{\alpha} \cdot y_k - \bar{\beta}$;
- If $\mu^* + \Delta_k \geq \bar{\alpha} \cdot y_k + \bar{\beta} > \mu^*$ then, $(\tilde{\nu}_k, \tilde{\nu}_k^{\text{aux}}) = (\text{Ber}(\mu^* + \epsilon_k), \text{Ber}(y_k))$ with $\epsilon_k = -\delta_k = \bar{\alpha} \cdot y_k + \bar{\beta} - \mu^*$;
- If $\bar{\alpha} \cdot y_k + \bar{\beta} > \mu^* + \Delta_k$ then, $(\tilde{\nu}_k, \tilde{\nu}_k^{\text{aux}}) = (\text{Ber}(\mu^* + \epsilon_k), \text{Ber}(y_k))$ with $\epsilon_k = \Delta_k$.

This concludes the proof. ■

A.7 Proof of Theorem 5

Step 1 (Preliminaries). The proof adapts ideas of identifying worst-case nature strategy to our setting in order to identify the precise change in the achievable performance as a function of the entries of information arrival matrix $\mathbf{H}$. Fix a problem instance $(\boldsymbol{\nu}, \boldsymbol{\nu}^{\text{aux}}) \in \mathcal{S}$ with the mean rewards $\boldsymbol{\mu}$ and the mean $\mathbf{y}$ for auxiliary observations, and recall that $\Delta_k = \mu^* - \mu_k$. Consider a suboptimal arm k. We consider a new problem instance $(\tilde{\boldsymbol{\nu}}, \tilde{\boldsymbol{\nu}}^{\text{aux}}) \in \mathcal{S}$ for which arm k is the optimal arm such that $\mathbb{E}_{X \sim \tilde{\nu}_k}[X] = \mu^* + \epsilon_k$.

Step 2 (Distance between problem instances). This step is based on the following lemma that states how informative the history can be about the true underlying problem instance.

Lemma 6 *Let* $\mathcal{H}_t \coloneqq (\boldsymbol{h}_1, \mathbf{Z}_1, \pi_1, X_{\pi_1,1}, \ldots, \boldsymbol{h}_{t-1}, \mathbf{Z}_{t-1}, \pi_{t-1}, X_{\pi_{t-1},t-1}, \boldsymbol{h}_t, \mathbf{Z}_t)$ *be the history up to* t*. For any problem instance* $(\boldsymbol{\nu}, \boldsymbol{\nu}^{\text{aux}})$*, denote by* $(\mathbb{\nu}, \mathbb{\nu}^{\text{aux}})_t$ *the law of* $\mathcal{H}_t$ *under the problem instance* $(\boldsymbol{\nu}, \boldsymbol{\nu}^{\text{aux}})$*. Then, for any problem instances* $(\boldsymbol{\nu}, \boldsymbol{\nu}^{\text{aux}})$ *and* $(\tilde{\boldsymbol{\nu}}, \tilde{\boldsymbol{\nu}}^{\text{aux}})$ *and any* $t \in \mathcal{T}$*,*

$$\text{KL}\left((\mathbb{\nu}, \mathbb{\nu}^{\text{aux}})_t, (\tilde{\mathbb{\nu}}, \tilde{\mathbb{\nu}}^{\text{aux}})_t\right) = \sum_{k \in \mathcal{K}} \left[\text{KL}\left(\nu_k, \tilde{\nu}_k\right) \cdot \mathbb{E}_{(\boldsymbol{\nu}_t, \boldsymbol{\nu}_t^{\text{aux}})}[n^{\pi}_{k,t}] + \text{KL}\left(\nu_k^{\text{aux}}, \tilde{\nu}_k^{\text{aux}}\right) \cdot \sum_{s=1}^{t} h_{k,s}\right].$$

Proof. This type of statement is well-known in the MAB literature (see, e.g., Auer et al. (2002), Garivier et al. (2019), or Gerchinovitz and Lattimore (2016)). However, we show how it is derived for the sake

of completeness. For any random variables X and Y, we denote by $(\mathbb{\nu}, \mathbb{\nu}^{\mathrm{aux}})^{(X)}$ the law of X, and by $(\mathbb{\nu}, \mathbb{\nu}^{\mathrm{aux}})^{(X|Y)}$ the law of X conditional on Y under the problem instance $(\boldsymbol{\nu}, \boldsymbol{\nu}^{\mathrm{aux}})$. By the chain rule of KL-divergence, one has

$$\begin{aligned}\mathrm{KL}\left((\mathbb{\nu}, \mathbb{\nu}^{\mathrm{aux}})_t, (\tilde{\mathbb{\nu}}, \tilde{\mathbb{\nu}}^{\mathrm{aux}})_t\right) &= \mathrm{KL}\left((\mathbb{\nu}, \mathbb{\nu}^{\mathrm{aux}})_{t-1}, (\tilde{\mathbb{\nu}}, \tilde{\mathbb{\nu}}^{\mathrm{aux}})_{t-1}\right) \\ &\quad + \mathrm{KL}\left((\mathbb{\nu}, \mathbb{\nu}^{\mathrm{aux}})^{((\pi_{t-1}, X_{\pi_{t-1},t-1}, \boldsymbol{h}_t, \mathbf{Z}_t)|\mathcal{H}_{t-1})}, (\tilde{\mathbb{\nu}}, \tilde{\mathbb{\nu}}^{\mathrm{aux}})^{((\pi_{t-1}, X_{\pi_{t-1},t-1}, \boldsymbol{h}_t, \mathbf{Z}_t)|\mathcal{H}_{t-1})}\right).\end{aligned}$$

Applying the chain rule to the second quantity on the right hand side of the above equality, one obtains

$$\begin{aligned}&\mathrm{KL}\left((\mathbb{\nu}, \mathbb{\nu}^{\mathrm{aux}})^{((\pi_{t-1}, X_{\pi_{t-1},t-1}, \boldsymbol{h}_t, \mathbf{Z}_t)|\mathcal{H}_{t-1})}, (\tilde{\mathbb{\nu}}, \tilde{\mathbb{\nu}}^{\mathrm{aux}})^{((\pi_{t-1}, X_{\pi_{t-1},t-1}, \boldsymbol{h}_t, \mathbf{Z}_t)|\mathcal{H}_{t-1})}\right) = \\ &\mathrm{KL}\left(\nu_k, \tilde{\nu}_k\right) \cdot \mathbb{P}_{(\boldsymbol{\nu}_t, \boldsymbol{\nu}_t^{\mathrm{aux}})}\left\{\pi_t = k\right\} + \mathrm{KL}\left(\nu_k^{\mathrm{aux}}, \tilde{\nu}_k^{\mathrm{aux}}\right) \cdot h_{k,t}.\end{aligned}$$

By reiterating the above two displays, the desired result is derived. This concludes the proof. ■

The above lemma yields

$$\mathrm{KL}\left((\mathbb{\nu}, \mathbb{\nu}^{\mathrm{aux}})_t, (\tilde{\mathbb{\nu}}, \tilde{\mathbb{\nu}}^{\mathrm{aux}})_t\right) = \mathrm{KL}\left(\nu_k, \tilde{\nu}_k\right) \cdot \mathbb{E}_{(\boldsymbol{\nu}_t, \boldsymbol{\nu}_t^{\mathrm{aux}})}[n_{k,t}^{\pi}] + \mathrm{KL}\left(\nu_k^{\mathrm{aux}}, \tilde{\nu}_k^{\mathrm{aux}}\right) \cdot \sum_{s=1}^{t} h_{k,s}. \tag{14}$$

Step 3 (A constraint for $\mathbb{E}^{\pi}_{(\boldsymbol{\nu}, \boldsymbol{\nu}_t^{\mathrm{aux}})}[n^{\pi}_{k,T+1}]$). Note that by the assumption that π is IAO, one has

$$\begin{aligned}&\frac{C_{\pi,\mathcal{S}} K \sigma^2}{\epsilon^2} \log T \geq \sum_{t=1}^{T} \mathbb{P}^{\pi}_{(\tilde{\boldsymbol{\nu}}, \tilde{\boldsymbol{\nu}}^{\mathrm{aux}})}\left\{\pi_t \neq k\right\} \\ &\overset{(a)}{\geq} \sum_{t=1}^{T} \frac{1}{2} \exp\left[-\mathrm{KL}\left((\mathbb{\nu}, \mathbb{\nu}^{\mathrm{aux}})_t, (\tilde{\mathbb{\nu}}, \tilde{\nu}^{\mathrm{aux}})_t\right)\right] - \mathbb{P}^{\pi}_{(\boldsymbol{\nu}, \boldsymbol{\nu}^{\mathrm{aux}})}\left\{\pi_t = k\right\} \\ &\overset{(b)}{\geq} \sum_{t=1}^{T} \frac{1}{2} \exp\left[-\mathrm{KL}\left(\nu_k, \tilde{\nu}_k\right) \cdot \mathbb{E}_{(\boldsymbol{\nu}_t, \boldsymbol{\nu}_t^{\mathrm{aux}})}[n_{k,t}^{\pi}] - \mathrm{KL}\left(\nu_k^{\mathrm{aux}}, \tilde{\nu}_k^{\mathrm{aux}}\right) \cdot \sum_{s=1}^{t} h_{k,s}\right] - \mathbb{P}^{\pi}_{(\boldsymbol{\nu}, \boldsymbol{\nu}^{\mathrm{aux}})}\left\{\pi_t = k\right\} \\ &\geq \exp\left[-\mathrm{KL}\left(\nu_k, \tilde{\nu}_k\right) \cdot \mathbb{E}^{\pi}_{(\boldsymbol{\nu}, \boldsymbol{\nu}^{\mathrm{aux}})}[n^{\pi}_{k,T+1}]\right] \cdot \sum_{t=1}^{T} \frac{1}{2} \exp\left[-\mathrm{KL}\left(\nu_k^{\mathrm{aux}}, \tilde{\nu}_k^{\mathrm{aux}}\right) \cdot \sum_{s=1}^{t} h_{k,s}\right] - \mathbb{E}^{\pi}_{(\boldsymbol{\nu}, \boldsymbol{\nu}_t^{\mathrm{aux}})}[n^{\pi}_{k,T+1}],\end{aligned}$$

where (a) follows from Lemma 7 and (b) holds by (14). The above inequality provides a constraint over the quantity of interest, $\mathbb{E}^{\pi}_{(\boldsymbol{\nu}, \boldsymbol{\nu}_t^{\mathrm{aux}})}[n^{\pi}_{k,T+1}]$.

Step 4 (Casting the lower bound of $\mathbb{E}^{\pi}_{(\boldsymbol{\nu}, \boldsymbol{\nu}_t^{\mathrm{aux}})}[n^{\pi}_{k,T+1}]$ as a convex optimization problem). Given the above inequality one can see that $\mathbb{E}^{\pi}_{(\boldsymbol{\nu}, \boldsymbol{\nu}_t^{\mathrm{aux}})}[n^{\pi}_{k,T+1}]$ is lower bounded by the solution of the following convex optimization problem:

$$\begin{aligned}
\min_{x} \quad & x \\
\text{subject to} \quad & \exp\left[-\mathrm{KL}\left(\nu_k, \tilde{\nu}_k\right) \cdot x\right] \cdot \sum_{t=1}^{T} \frac{1}{2} \exp\left[-\mathrm{KL}\left(\nu_k^{\mathrm{aux}}, \tilde{\nu}_k^{\mathrm{aux}}\right) \cdot \sum_{s=1}^{t} h_{k,s}\right] - x \leq \frac{C_{\pi,\mathcal{S}} K \sigma^2}{\epsilon^2} \log T.
\end{aligned}$$

The solution of the above optimization problem is then, lower bounded by that of the following for any $\lambda \geq 0$:

$$\begin{aligned}
\min_{x} \quad & x + \lambda \left(\exp\left[-\mathrm{KL}\left(\nu_k, \tilde{\nu}_k\right) \cdot x\right] \cdot \sum_{t=1}^{T} \frac{1}{2} \exp\left[-\mathrm{KL}\left(\nu_k^{\mathrm{aux}}, \tilde{\nu}_k^{\mathrm{aux}}\right) \cdot \sum_{s=1}^{t} h_{k,s}\right] - x - \frac{C_{\pi,\mathcal{S}} K \sigma^2}{\epsilon^2} \log T \right) \\
\text{subject to} \quad & \exp\left[-\mathrm{KL}\left(\nu_k, \tilde{\nu}_k\right) \cdot x\right] \cdot \sum_{t=1}^{T} \frac{1}{2} \exp\left[-\mathrm{KL}\left(\nu_k^{\mathrm{aux}}, \tilde{\nu}_k^{\mathrm{aux}}\right) \cdot \sum_{s=1}^{t} h_{k,s}\right] - x \leq \frac{C_{\pi,\mathcal{S}} K \sigma^2}{\epsilon^2} \log T.
\end{aligned}$$

Letting $\lambda = \frac{\epsilon^2}{C_{\pi,\mathcal{S}} K \sigma^2 \log T}$ and noting that $x + \gamma e^{-\kappa x} \geq \frac{\log \gamma\kappa}{\kappa}$ for $\gamma, \kappa, x > 0$, one obtains

$$\mathbb{E}^{\pi}_{(\boldsymbol{\nu}, \boldsymbol{\nu}_t^{\mathrm{aux}})}[n^{\pi}_{k,T+1}] \geq \frac{1}{\mathrm{KL}\left(\nu_k, \tilde{\nu}_k\right)} \log\left(\frac{\mathrm{KL}\left(\nu_k, \tilde{\nu}_k\right) \epsilon^2}{C_{\pi,\mathcal{S}} K \sigma^2 \log T} \sum_{t=1}^{T} \exp\left[-\mathrm{KL}\left(\nu_k^{\mathrm{aux}}, \tilde{\nu}_k^{\mathrm{aux}}\right) \cdot \sum_{s=1}^{t} h_{k,s}\right] \right) - 1.$$

This concludes the proof. ■

A.8 Proof of Theorem 6

Step 1 (Preliminaries). Fix a profile $(\boldsymbol{\nu}, \boldsymbol{\nu}^{\mathrm{aux}})$. Note that it is unlikely for the UCB of the optimal arm to be less than the largest mean reward. We will carry out the analysis based on this observation. To be more precise, define $\mathcal{E}_t^* \coloneqq \{U_{k^*,t} < \mu^*\}$ to be the event on which the UCB of the optimal arm is less than the largest mean reward at epoch t. Note that by the Chernoff-Hoefding bound in Lemma 8, one has $\mathbb{P}^{\pi}_{\boldsymbol{\nu}, \boldsymbol{\nu}^{\mathrm{aux}}}\{\mathcal{E}_t^*\} \leq \frac{1}{t^{\frac{c}{2}}}$ for all $t \geq K+1$. Hence, defining $\mathcal{E}^* \coloneqq \bigcup_{t=1}^{T} \mathcal{E}_t^*$, one obtains

$$\begin{aligned}
\mathcal{R}^{\pi}_{\boldsymbol{\nu}, \boldsymbol{\nu}^{\mathrm{aux}}} &\leq \sum_{k \in \mathcal{K}} \Delta_k + \mathbb{E}^{\pi}_{\boldsymbol{\nu}, \boldsymbol{\nu}^{\mathrm{aux}}} \left[\sum_{t=K+1}^{T} \left(\mu^* - \mu_{\pi_t}\right) \,\middle|\, \bar{\mathcal{E}}^* \right] + \sum_{t=K+1}^{T} \frac{\max_{k \in \mathcal{K}} \Delta_k}{t^{\frac{c}{2}}} \\
&\leq \sum_{k \in \mathcal{K}} \Delta_k + \sum_{k \in \mathcal{K}} \Delta_k \cdot \mathbb{E}^{\pi}_{\boldsymbol{\nu}, \boldsymbol{\nu}^{\mathrm{aux}}} \left[n^{\pi}_{k,T+1} \,\middle|\, \bar{\mathcal{E}}^* \right] + \frac{\max_{k \in \mathcal{K}} \Delta_k}{\frac{c}{2} - 1}, \qquad (17)
\end{aligned}$$

where we have used the assumption that $c > 2$. Consider a suboptimal arm $k \neq k^*$. We will bound $\mathbb{E}^{\pi}_{\boldsymbol{\nu}, \boldsymbol{\nu}^{\mathrm{aux}}} \left[n^{\pi}_{k,T+1} \,\middle|\, \bar{\mathcal{E}}^* \right]$, the expected number of times the suboptimal arm k is pulled conditional on the event $\bar{\mathcal{E}}^*$. Our analysis is based on the fact that if arm k is pulled at t, then

$$U_{k,t} \geq U_{k^*,t}. \qquad (18)$$

We give two separate upper bounds for $\mathbb{E}^{\pi}_{(\boldsymbol{\nu}, \boldsymbol{\nu}^{\mathrm{aux}})} \left[n^{\pi}_{k,T+1} \,\middle|\, \bar{\mathcal{E}}^* \right]$. The first upper bound holds for any suboptimal arm k, and the second one is specific to the suboptimal arms $k \in \mathcal{K}^*$. The following lemma

will be the main tool in our analysis.

Step 2 (Regret upper bound for any suboptimal arms $k \in \mathcal{K} \setminus \{k^*\}$). In this step, we deploy the UCB constructed based on reward observations. That is, (18) yields that if arm k is pulled at epoch t then, conditional on the event $\bar{\mathcal{E}}^*$, one has $\bar{X}^\pi_{k,t} + \sqrt{\frac{c\sigma^2 \log t}{n^\pi_{k,t}}} = U^\pi_{k,t} \geq \mu^*$. Therefore, at least one of the following events must occur:

$$\mathcal{E}^\pi_{1,t} \coloneqq \left\{ \bar{X}^\pi_{k,t} \geq \mu_k + \sqrt{\frac{c\sigma^2 \log t}{n^\pi_{k,t}}} \right\}, \quad \mathcal{E}^\pi_{2,t} \coloneqq \left\{ \Delta_k \leq 2\sqrt{\frac{c\sigma^2 \log t}{n^\pi_{k,t}}} \right\}.$$

To see why this is true, assume that all the above events fail. Then, we have

$$\bar{X}^\pi_{k,t} + \sqrt{\frac{c\sigma^2 \log t}{n^\pi_{k,t}}} < \mu_k + 2\sqrt{\frac{c\sigma^2 \log t}{n^\pi_{k,t}}} < \mu_k + \Delta_k = \mu^*,$$

which is in contradiction with the assumption that arm k is pulled at epoch t conditional on the event $\bar{\mathcal{E}}^*$. For any sequence $\{l_{k,t}\}_{t\in\mathcal{T}}$, one has

$$\begin{aligned}
\mathbb{E}^\pi_{\boldsymbol{\nu},\boldsymbol{\nu}^{\text{aux}}}\left[n^\pi_{k,T+1} \;\middle|\; \bar{\mathcal{E}}^* \right] &= \mathbb{E}^\pi_{\boldsymbol{\nu},\boldsymbol{\nu}^{\text{aux}}}\left[\sum_{t=1}^T \mathbb{1}\left\{\pi_t = k\right\} \;\middle|\; \bar{\mathcal{E}}^* \right] \\
&= \mathbb{E}^\pi_{\boldsymbol{\nu},\boldsymbol{\nu}^{\text{aux}}}\left[\sum_{t=1}^T \mathbb{1}\left\{\pi_t = k, n^\pi_{k,t} \leq l_{k,t}\right\} + \mathbb{1}\left\{\pi_t = k, n^\pi_{k,t} > l_{k,t}\right\} \;\middle|\; \bar{\mathcal{E}}^* \right] \\
&\leq \mathbb{E}^\pi_{\boldsymbol{\nu},\boldsymbol{\nu}^{\text{aux}}}\left[\sum_{t=1}^T \mathbb{1}\left\{\pi_t = k, n^\pi_{k,t} \leq \hat{l}_{k,t}\right\} + \mathbb{1}\left\{\pi_t = k, n^\pi_{k,t} > l_{k,t}\right\} \;\middle|\; \bar{\mathcal{E}}^* \right] \\
&\leq 1 + \max_{1\leq t\leq T}\left\{\hat{l}_{k,t}\right\} + \sum_{t=1}^T \mathbb{P}^\pi_{\boldsymbol{\nu},\boldsymbol{\nu}^{\text{aux}}}\left\{\pi_t = k, n^\pi_{k,t} > l_{k,t} \;\middle|\; \bar{\mathcal{E}}^*\right\},
\end{aligned}$$

where the last inequality follows from Lemma 1. Let $l_{k,t} \coloneqq \frac{4c\sigma^2 \log(t)}{\Delta_k^2}$. To have $n^\pi_{k,t} > l_{k,t}$, it must be the case that $\mathcal{E}^\pi_{2,t}$ does not occur. Therefore,

$$\sum_{t=1}^T \mathbb{P}^\pi_{\boldsymbol{\nu},\boldsymbol{\nu}^{\text{aux}}}\left\{\pi_t = k, n^\pi_{k,t} > l_{k,t} \;\middle|\; \bar{\mathcal{E}}^*\right\} \leq \sum_{t=1}^T \mathbb{P}^\pi_{\boldsymbol{\nu},\boldsymbol{\nu}^{\text{aux}}}\left\{\pi_t = k, \bar{\mathcal{E}}^\pi_{2,t} \;\middle|\; \bar{\mathcal{E}}^*\right\} \leq \sum_{t=1}^T \mathbb{P}^\pi_{\boldsymbol{\nu},\boldsymbol{\nu}^{\text{aux}}}\left\{\mathcal{E}^\pi_{1,t} \;\middle|\; \bar{\mathcal{E}}^*\right\} \overset{(a)}{\leq} \sum_{t=1}^T \frac{1}{t^{\frac{c}{2}}} \overset{(b)}{\leq} \frac{1}{\frac{c}{2}-1}, \tag{19}$$

where (a) follows from the Chernoff-Hoeffding bound in Lemma 8 and (b) follows from the assumption that $c > 2$. The last two displays yield

$$\mathbb{E}^\pi_{\boldsymbol{\nu},\boldsymbol{\nu}^{\text{aux}}}\left[n^\pi_{k,T+1} \;\middle|\; \bar{\mathcal{E}}^* \right] \leq \frac{4c\sigma^2 \log(T)}{\Delta_k^2} + \frac{1}{\frac{c}{2}-1} + 1.$$

Step 3 (Regret upper bound for suboptimal arms $k \in \mathcal{K} \setminus \{k^*\}$ which satisfy $\delta_k \geq 0$)). For this set of arms, we deploy the UCB that incorporate both reward observations and auxiliary observations. That is, (18) yields that if arm k is pulled at epoch t then, conditional on the event $\bar{\mathcal{E}}^*$, one has

$$\bar{X}_{k,t}^{\pi,\mathrm{aux}} + \sqrt{\frac{c\sigma^2 \log t}{n_{k,t}^{\pi,\mathrm{aux}}}} = U_{k,t}^{\pi,\mathrm{aux}} \geq \mu^*.$$

Therefore, at least one of the following events must occur:

$$\mathcal{E}_{1,t}^{\pi,\mathrm{aux}} \coloneqq \left\{ \bar{X}_{k,t}^{\pi,\mathrm{aux}} \geq \bar{\alpha} y_k + \sqrt{\frac{c\sigma^2 \log t}{n_{k,t}^{\pi,\mathrm{aux}}}} \right\}, \quad \mathcal{E}_{2,t}^{\pi,\mathrm{aux}} \coloneqq \left\{ \Delta_{k,t}^{\pi,\mathrm{aux}} \leq 2\sqrt{\frac{c\sigma^2 \log t}{n_{k,t}^{\pi,\mathrm{aux}}}} \right\}.$$

A similar analysis as in the proof of Theorem 2, except for replacing Δ_k with δ_k throughout the analysis, results in the following upper bound:

$$\mathbb{E}_{\boldsymbol{\nu},\boldsymbol{\nu}^{\mathrm{aux}}}^{\pi} \left[n_{k,T+1}^{\pi} \,\middle|\, \bar{\mathcal{E}}^* \right] \leq 1 + \frac{4c\sigma^2}{\delta_k^2} \log \left(\sum_{t=0}^{T} \exp \left(\frac{-\delta_k^2}{4c\bar{\alpha}^2\hat{\sigma}^2} \sum_{s=1}^{t} h_{k,s} \right) \right) + \frac{1}{\frac{c}{2} - 1}.$$

Step 4 (Regret upper bound for suboptimal arms $k \in \mathcal{K} \setminus \{k^*\}$ which satisfy $\delta_k \geq \frac{\Delta_k}{2}$)). For this set of arms, we deploy the UCB that incorporate both reward observations and auxiliary observations. That is, (18) yields that if arm k is pulled at epoch t then, conditional on the event $\bar{\mathcal{E}}^*$, one has

$$\bar{X}_{k,t}^{\pi,\mathrm{aux}} + \sqrt{\frac{c\sigma^2 \log t}{n_{k,t}^{\pi,\mathrm{aux}}}} = U_{k,t}^{\pi,\mathrm{aux}} \geq \mu^*.$$

Therefore, at least one of the following events must occur:

$$\mathcal{E}_{1,t}^{\pi,\mathrm{aux}} \coloneqq \left\{ \bar{X}_{k,t}^{\pi,\mathrm{aux}} \geq \mu_{k,t}^{\pi,\mathrm{aux}} + \sqrt{\frac{c\sigma^2 \log t}{n_{k,t}^{\pi,\mathrm{aux}}}} \right\}, \quad \mathcal{E}_{2,t}^{\pi,\mathrm{aux}} \coloneqq \left\{ \Delta_{k,t}^{\pi,\mathrm{aux}} \leq 2\sqrt{\frac{c\sigma^2 \log t}{n_{k,t}^{\pi,\mathrm{aux}}}} \right\},$$

where we define

$$\mu_{k,t}^{\pi,\mathrm{aux}} \coloneqq \mu_{k,n_{k,t}^{\pi},n_{k,t}^{\mathrm{aux}}}^{\pi,\mathrm{aux}} = \frac{\mu_k \cdot n_{k,t}^{\pi} + (\bar{\alpha} \cdot y_k + \bar{\beta}) \cdot \sum_{s=1}^{t} \frac{\sigma^2}{\bar{\alpha}^2\hat{\sigma}^2} h_{k,s}}{n_{k,t}^{\pi,\mathrm{aux}}}; \qquad \Delta_{k,t}^{\pi,\mathrm{aux}} \coloneqq \Delta_{k,n_{k,t}^{\pi},n_{k,t}^{\mathrm{aux}}}^{\pi,\mathrm{aux}} = \mu^* - \mu_{k,t}^{\pi,\mathrm{aux}}.$$

To see why this is true, assume that all the above events fail. Then, we have

$$\bar{X}_{k,t}^{\pi,\mathrm{aux}} + \sqrt{\frac{c\sigma^2 \log t}{n_{k,t}^{\pi,\mathrm{aux}}}} < \mu_{k,t}^{\pi,\mathrm{aux}} + 2\sqrt{\frac{c\sigma^2 \log t}{n_{k,t}^{\pi,\mathrm{aux}}}} < \mu_{k,t}^{\pi,\mathrm{aux}} + \Delta_{k,t}^{\pi,\mathrm{aux}} = \mu^*,$$

which is in contradiction with the assumption that arm k is pulled at epoch t conditional on the event $\bar{\mathcal{E}}^*$. For any sequence $\{l_{k,t}\}_{t \in \mathcal{T}}$, one has

$$\begin{aligned}
\mathbb{E}^{\pi}_{\boldsymbol{\nu},\boldsymbol{\nu}^{\text{aux}}}\left[n^{\pi}_{k,T+1}\;\middle|\;\bar{\mathcal{E}}^*\right] &= \mathbb{E}^{\pi}_{\boldsymbol{\nu},\boldsymbol{\nu}^{\text{aux}}}\left[\sum_{t=1}^{T}\mathbb{1}\left\{\pi_t=k\right\}\;\middle|\;\bar{\mathcal{E}}^*\right] \\
&= \mathbb{E}^{\pi}_{\boldsymbol{\nu},\boldsymbol{\nu}^{\text{aux}}}\left[\sum_{t=1}^{T}\mathbb{1}\left\{\pi_t=k, n^{\pi,\text{aux}}_{k,t}\le l_{k,t}\right\}+\mathbb{1}\left\{\pi_t=k, n^{\pi,\text{aux}}_{k,t}> l_{k,t}\right\}\;\middle|\;\bar{\mathcal{E}}^*\right] \\
&\le \mathbb{E}^{\pi}_{\boldsymbol{\nu},\boldsymbol{\nu}^{\text{aux}}}\left[\sum_{t=1}^{T}\mathbb{1}\left\{\pi_t=k, n^{\pi,\text{aux}}_{k,t}\le l_{k,t}\right\}+\mathbb{1}\left\{\pi_t=k, n^{\pi,\text{aux}}_{k,t}> l_{k,t}\right\}\;\middle|\;\bar{\mathcal{E}}^*\right] \\
&\le \mathbb{E}^{\pi}_{\boldsymbol{\nu},\boldsymbol{\nu}^{\text{aux}}}\left[\sum_{t=1}^{T}\mathbb{1}\left\{\pi_t=k, n^{\pi,\text{aux}}_{k,t}\le l_{k,t}\right\}\;\middle|\;\bar{\mathcal{E}}^*\right] \\
&\quad+\sum_{t=1}^{T}\mathbb{P}^{\pi}_{\boldsymbol{\nu},\boldsymbol{\nu}^{\text{aux}}}\left\{\pi_t=k, n^{\pi,\text{aux}}_{k,t}> l_{k,t}\;\middle|\;\bar{\mathcal{E}}^*\right\}. \qquad (20)
\end{aligned}$$

We will upper bound each term on the right hand side of the above inequality separately. Set the values of $l_{k,t}$ as $l_{k,t}=\frac{4c\sigma^2\log(t)}{(\Delta^{\pi,\text{aux}}_{k,t})^2}$. In order to upper bound the term $\mathbb{E}^{\pi}_{\boldsymbol{\nu},\boldsymbol{\nu}^{\text{aux}}}\left[\sum_{t=1}^{T}\mathbb{1}\left\{\pi_t=k, n^{\pi,\text{aux}}_{k,t}\le l_{k,t}\right\}\;\middle|\;\bar{\mathcal{E}}^*\right]$ in (20), note that the equality $\Delta^{\pi,\text{aux}}_{k,t}=\frac{\Delta_k\cdot n^{\pi}_{k,t}+\frac{\delta_k\sigma^2}{\bar{\alpha}^2\hat{\sigma}^2}n^{\text{aux}}_{k,t}}{n^{\pi}_{k,t}+\frac{\sigma^2}{\bar{\alpha}^2\hat{\sigma}^2}n^{\text{aux}}_{k,t}}$ yields

$$n^{\pi,\text{aux}}_{k,t}\le l_{k,t}\;\Leftrightarrow\Delta_k^2\left(n^{\pi}_{k,t}\right)^2+\left(\frac{2\Delta_k\delta_k\sigma^2}{\bar{\alpha}^2\hat{\sigma}^2}n^{\text{aux}}_{k,t}-4c\sigma^2\log t\right)n^{\pi}_{k,t}+\left(\frac{\delta_k\sigma^2}{\bar{\alpha}^2\hat{\sigma}^2}n^{\text{aux}}_{k,t}\right)^2-\left(4c\sigma^2\log t\right)\left(\frac{\sigma^2}{\bar{\alpha}^2\hat{\sigma}^2}n^{\text{aux}}_{k,t}\right)\le 0. \qquad (21)$$

Since the left-hand side of the above inequality is quadratic in $n^{\pi}_{k,t}$, with a positive coefficient for the quadratic term then, the above inequality holds only if

$$n^{\pi}_{k,t}\le\frac{4c\sigma^2\log t-\frac{2\Delta_k\delta_k\sigma^2}{\bar{\alpha}^2\hat{\sigma}^2}n^{\text{aux}}_{k,t}+\sqrt{\left(4c\sigma^2\log t\right)^2+4\Delta_k(\Delta_k-\delta_k)\left(4c\sigma^2\log t\right)\left(\frac{\sigma^2}{\bar{\alpha}^2\hat{\sigma}^2}n^{\text{aux}}_{k,t}\right)}}{2\Delta_k^2}.$$

Now, in order to upper bound the right hand side of this inequality, note that if $\delta_k=\xi_k\Delta_k$, $\frac{1}{2}\le\xi_k\le 1$, one obtains

$$-\frac{2\Delta_k\delta_k\sigma^2}{\bar{\alpha}^2\hat{\sigma}^2}n^{\text{aux}}_{k,t}+\sqrt{\left(4c\sigma^2\log t\right)^2+4\Delta_k(\Delta_k-\delta_k)\left(4c\sigma^2\log t\right)\left(\frac{\sigma^2}{\bar{\alpha}^2\hat{\sigma}^2}n^{\text{aux}}_{k,t}\right)}\le-\frac{2\left(\frac{2\xi_k-1}{\xi_k^2}\right)\delta_k^2\sigma^2}{\bar{\alpha}^2\hat{\sigma}^2}n^{\text{aux}}_{k,t}+4c\sigma^2\log t$$

That is, the last two displays yield that (21) holds only if

$$\begin{aligned}
n^{\pi}_{k,t}\le\frac{4c\sigma^2\log t-\frac{\left(\frac{2\xi_k-1}{\xi_k^2}\right)\delta_k^2\sigma^2}{\bar{\alpha}^2\hat{\sigma}^2}n^{\text{aux}}_{k,t}}{\Delta_k^2}&\le\frac{4c\sigma^2\log\tau_{k,t}-\frac{\left(\frac{2\xi_k-1}{\xi_k^2}\right)\delta_k^2\sigma^2}{\bar{\alpha}^2\hat{\sigma}^2}n^{\text{aux}}_{k,t}}{\Delta_k^2} \\
&=\frac{4c\sigma^2}{\Delta_k^2}\log\left(\sum_{s=1}^{t}\exp\left(-\frac{\left(\frac{2\xi_k-1}{\xi_k^2}\right)\delta_k^2}{4c\bar{\alpha}^2\hat{\sigma}^2}\sum_{m=1}^{s}h_{m,k}\right)\right)=:\hat{l}_{k,t},
\end{aligned}$$

where we define $\tau_{k,t} \coloneqq \sum_{s=1}^{t} \exp\left(\frac{\left(\frac{2\xi_k-1}{\xi_k^2}\right)\delta_k^2}{4c\bar{\alpha}^2\hat{\sigma}^2} \sum_{m=s}^{t-1} h_{k,m}\right) \geq t$, and the equality follows from $n_{k,t}^{\text{aux}} = \sum_{s=1}^{t} h_{k,s}$.

One obtains

$$\begin{aligned}
\mathbb{E}^{\pi}_{\boldsymbol{\nu},\boldsymbol{\nu}^{\text{aux}}}\left[\sum_{t=1}^{T} \mathbb{1}\left\{\pi_t = k, n_{k,t}^{\pi,\text{aux}} \leq l_{k,t}\right\} \,\middle|\, \bar{\mathcal{E}}^*\right] &\leq \mathbb{E}^{\pi}_{\boldsymbol{\nu},\boldsymbol{\nu}^{\text{aux}}}\left[\sum_{t=1}^{T} \mathbb{1}\left\{\pi_t = k, n_{k,t}^{\pi} \leq \hat{l}_{k,t}\right\} \,\middle|\, \bar{\mathcal{E}}^*\right] \\
&\overset{(a)}{\leq} \max\left\{0, 1 + \max_{1\leq t\leq T} \hat{l}_{k,t}\right\} \overset{(b)}{\leq} \max\left\{0, 1 + \hat{l}_{k,T}\right\} \\
&\leq 1 + \frac{4c\sigma^2}{\Delta_k \delta_k} \log\left(\sum_{t=0}^{T} \exp\left(\frac{-\left(\frac{2\xi_k-1}{\xi_k^2}\right)\delta_k^2}{4c\bar{\alpha}^2\hat{\sigma}^2} \sum_{s=1}^{t} h_{k,s}\right)\right), \quad (22)
\end{aligned}$$

where (a) follows from Lemma 1, and (b) follows from the fact that $\{\hat{l}_{k,t}\}_t$ is an increasing sequence.

In order to upper bound the term $\sum_{t=1}^{T} \mathbb{P}^{\pi}_{\boldsymbol{\nu},\boldsymbol{\nu}^{\text{aux}}}\left\{\pi_t = k, n_{k,t}^{\pi,\text{aux}} > l_{k,t} \,\middle|\, \bar{\mathcal{E}}^*\right\}$ in (20), we note that to have $n_{k,t}^{\pi,\text{aux}} > l_{k,t}$, it must be the case that $\bar{\mathcal{E}}_{2,t}^{\pi,\text{aux}}$ does not occur. On the other hand, if arm k is pulled and $\mathcal{E}_{2,t}^{\pi,\text{aux}}$ does not occur, then it must be the case that $\mathcal{E}_{1,t}^{\pi,\text{aux}}$ occurs. Therefore,

$$\begin{aligned}
\sum_{t=1}^{T} \mathbb{P}^{\pi}_{\boldsymbol{\nu},\boldsymbol{\nu}^{\text{aux}}}\left\{\pi_t = k, n_{k,t}^{\pi,\text{aux}} > l_{k,t} \,\middle|\, \bar{\mathcal{E}}^*\right\} &= \sum_{t=1}^{T} \mathbb{P}^{\pi}_{\boldsymbol{\nu},\boldsymbol{\nu}^{\text{aux}}}\left\{\pi_t = k, \bar{\mathcal{E}}_{2,t}^{\pi,\text{aux}} \,\middle|\, \bar{\mathcal{E}}^*\right\} \\
&\leq \sum_{t=1}^{T} \mathbb{P}^{\pi}_{\boldsymbol{\nu},\boldsymbol{\nu}^{\text{aux}}}\left\{\mathcal{E}_{1,t}^{\pi,\text{aux}} \,\middle|\, \bar{\mathcal{E}}^*\right\} \leq \sum_{t=1}^{T} \frac{1}{t^{c/2}} \leq \frac{1}{\frac{c}{2}-1},
\end{aligned}$$

Putting back together (20), (22), and the above display, one obtains

$$\mathbb{E}^{\pi}_{\boldsymbol{\nu},\boldsymbol{\nu}^{\text{aux}}}\left[n_{k,T+1}^{\pi} \,\middle|\, \bar{\mathcal{E}}^*\right] \leq 1 + \frac{4c\sigma^2}{\Delta_k^2} \log\left(\sum_{t=0}^{T} \exp\left(\frac{-\left(\frac{2\xi_k-1}{\xi_k^2}\right)\delta_k^2}{4c\bar{\alpha}^2\hat{\sigma}^2} \sum_{s=1}^{t} h_{k,s}\right)\right) + \frac{1}{\frac{c}{2}-1}.$$

This concludes the proof. ■