Regret Bound for UCB
(Adapted from the proof by Auer et al., 2002)

Shivaram Kalyanakrishnan

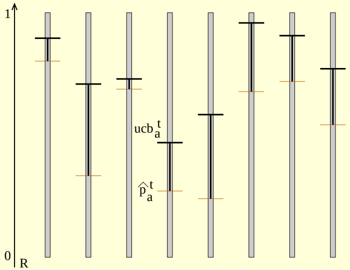
shivaram@cse.iitb.ac.in

Department of Computer Science and Engineering
Indian Institute of Technology Bombay

August 2018

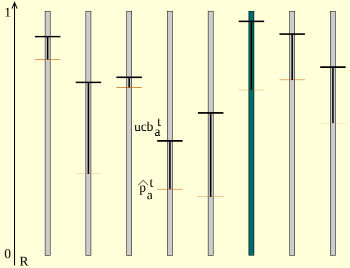
UCB Algorithm

- **UCB**
  - Pull each arm once.
  - At time $t \in \{n, n+1, \dots\}$, for every arm a, $\text{ucb}_a^t \stackrel{\text{def}}{=} \hat{p}_a^t + \sqrt{\frac{2\ln(t)}{u_a^t}}$; pull $\text{argmax}_a \text{ucb}_a^t$.



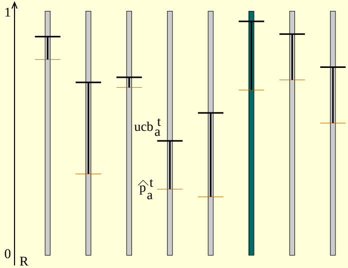
UCB Algorithm

- UCB
  - Pull each arm once.
  - At time $t \in \{n, n+1, \ldots\}$, for every arm a, $\text{ucb}_a^t \stackrel{\text{def}}{=} \hat{p}_a^t + \sqrt{\frac{2\ln(t)}{u_a^t}}$; pull $\text{argmax}_a \, \text{ucb}_a^t$.



UCB Algorithm

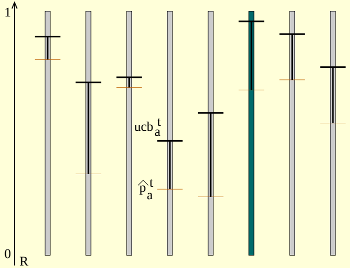
- UCB
  - Pull each arm once.
  - At time $t \in \{n, n+1, \dots\}$, for every arm a, $\text{ucb}_a^t \stackrel{\text{def}}{=} \hat{p}_a^t + \sqrt{\frac{2\ln(t)}{u_a^t}}$; pull $\text{argmax}_a\, \text{ucb}_a^t$.



- Recall that $R^T = Tp_\star - \sum_{t=0}^{T-1} \mathbb{E}[r^t]$.

UCB Algorithm

- UCB
  - Pull each arm once.
  - At time $t \in \{n, n+1, \dots\}$, for every arm a, $\text{ucb}_a^t \stackrel{\text{def}}{=} \hat{p}_a^t + \sqrt{\frac{2\ln(t)}{u_a^t}}$; pull $\text{argmax}_a \, \text{ucb}_a^t$.



- Recall that $R^T = Tp_\star - \sum_{t=0}^{T-1} \mathbb{E}[r^t]$.
- We shall show that UCB achieves $R^T = O\left(\sum_{a:p_a \neq p_\star} \frac{1}{p_\star - p_a} \log(T)\right)$.

Notation

- $\Delta_a \stackrel{\text{def}}{=} p_\star - p_a$ (instance-specific **constant**).

Notation

- $\Delta_a \stackrel{\text{def}}{=} p_\star - p_a$ (instance-specific **constant**).

- Let Z_a^t be the **event** that arm a is pulled at time t.

Notation

- $\Delta_a \stackrel{\text{def}}{=} p_\star - p_a$ (instance-specific **constant**).

- Let Z_a^t be the **event** that arm a is pulled at time t.
- Let z_a^t be a **random variable** that takes value 1 if arm a is pulled at time t, and 0 otherwise.

Notation

- $\Delta_a \stackrel{\text{def}}{=} p_\star - p_a$ (instance-specific **constant**).

- Let Z_a^t be the **event** that arm a is pulled at time t.
- Let z_a^t be a **random variable** that takes value 1 if arm a is pulled at time t, and 0 otherwise. Observe that $\mathbb{E}[z_a^t] = \mathbb{P}\{Z_a^t\}(1) + (1 - \mathbb{P}\{Z_a^t\})(0) = \mathbb{P}\{Z_a^t\}$.

Notation

- $\Delta_a \stackrel{\text{def}}{=} p_\star - p_a$ (instance-specific **constant**).

- Let Z_a^t be the **event** that arm a is pulled at time t.
- Let z_a^t be a **random variable** that takes value 1 if arm a is pulled at time t, and 0 otherwise. Observe that $\mathbb{E}[z_a^t] = \mathbb{P}\{Z_a^t\}(1) + (1 - \mathbb{P}\{Z_a^t\})(0) = \mathbb{P}\{Z_a^t\}$.

- As in the algorithm, u_a^t is a **random variable** that denotes the number of pulls arm a has received up to (and excluding) time t:

$$u_a^t = \sum_{i=0}^{t-1} z_a^i.$$

Notation

- $\Delta_a \stackrel{\text{def}}{=} p_\star - p_a$ (instance-specific **constant**).

- Let Z_a^t be the **event** that arm a is pulled at time t.
- Let z_a^t be a **random variable** that takes value 1 if arm a is pulled at time t, and 0 otherwise. Observe that $\mathbb{E}[z_a^t] = \mathbb{P}\{Z_a^t\}(1) + (1 - \mathbb{P}\{Z_a^t\})(0) = \mathbb{P}\{Z_a^t\}$.

- As in the algorithm, u_a^t is a **random variable** that denotes the number of pulls arm a has received up to (and excluding) time t:

$$u_a^t = \sum_{i=0}^{t-1} z_a^i.$$

- We define an instance-specific **constant**

$$\bar{u}_a^T \stackrel{\text{def}}{=} \left\lceil \frac{8}{(\Delta_a)^2} \ln(T) \right\rceil$$

that will serve in our proof as a "sufficient" number of pulls of arm a for horizon T.

Step 1: Show that $R^T = \sum_{a:p_a \neq p_\star} \mathbb{E}[u_a^T]\Delta_a$.

Step 1: Show that $R^T = \sum_{a:p_a \neq p_\star} \mathbb{E}[u_a^T]\Delta_a$.

$$R^T = Tp_\star - \sum_{t=0}^{T-1} \mathbb{E}[r^t]$$

Step 1: Show that $R^T = \sum_{a: p_a \neq p_\star} \mathbb{E}[u_a^T] \Delta_a$.

$$
\begin{aligned}
R^T &= Tp_\star - \sum_{t=0}^{T-1} \mathbb{E}[r^t] \\
&= Tp_\star - \sum_{t=0}^{T-1} \sum_{a \in A} \mathbb{P}\{Z_a^t\} \mathbb{E}[r^t | Z_a^t]
\end{aligned}
$$

Step 1: Show that $R^T = \sum_{a:p_a \neq p_\star} \mathbb{E}[u_a^T]\Delta_a$.

$$
\begin{aligned}
R^T &= Tp_\star - \sum_{t=0}^{T-1} \mathbb{E}[r^t] \\
&= Tp_\star - \sum_{t=0}^{T-1} \sum_{a \in A} \mathbb{P}\{Z_a^t\} \mathbb{E}[r^t | Z_a^t] \\
&= Tp_\star - \sum_{t=0}^{T-1} \sum_{a \in A} \mathbb{E}[z_a^t] p_a
\end{aligned}
$$

Step 1: Show that $R^T = \sum_{a:p_a \neq p_\star} \mathbb{E}[u_a^T]\Delta_a$.

$$
\begin{aligned}
R^T &= Tp_\star - \sum_{t=0}^{T-1} \mathbb{E}[r^t] \\
&= Tp_\star - \sum_{t=0}^{T-1} \sum_{a \in A} \mathbb{P}\{Z_a^t\} \mathbb{E}[r^t | Z_a^t] \\
&= Tp_\star - \sum_{t=0}^{T-1} \sum_{a \in A} \mathbb{E}[z_a^t] p_a \\
&= \left(\sum_{a \in A} \mathbb{E}[u_a^T] \right) p_\star - \sum_{a \in A} \mathbb{E}[u_a^T] p_a
\end{aligned}
$$

Step 1: Show that $R^T = \sum_{a:p_a \neq p_\star} \mathbb{E}[u_a^T]\Delta_a$.

$$
\begin{aligned}
R^T &= Tp_\star - \sum_{t=0}^{T-1} \mathbb{E}[r^t] \\
&= Tp_\star - \sum_{t=0}^{T-1} \sum_{a \in A} \mathbb{P}\{Z_a^t\}\mathbb{E}[r^t | Z_a^t] \\
&= Tp_\star - \sum_{t=0}^{T-1} \sum_{a \in A} \mathbb{E}[z_a^t] p_a \\
&= \left(\sum_{a \in A} \mathbb{E}[u_a^T] \right) p_\star - \sum_{a \in A} \mathbb{E}[u_a^T] p_a \\
&= \sum_{a \in A} \mathbb{E}[u_a^T](p_\star - p_a)
\end{aligned}
$$

Step 1: Show that $R^T = \sum_{a: p_a \neq p_\star} \mathbb{E}[u_a^T]\Delta_a$.

$$
\begin{aligned}
R^T &= Tp_\star - \sum_{t=0}^{T-1} \mathbb{E}[r^t] \\
&= Tp_\star - \sum_{t=0}^{T-1} \sum_{a \in A} \mathbb{P}\{Z_a^t\}\mathbb{E}[r^t | Z_a^t] \\
&= Tp_\star - \sum_{t=0}^{T-1} \sum_{a \in A} \mathbb{E}[z_a^t] p_a \\
&= \left(\sum_{a \in A} \mathbb{E}[u_a^T] \right) p_\star - \sum_{a \in A} \mathbb{E}[u_a^T] p_a \\
&= \sum_{a \in A} \mathbb{E}[u_a^T](p_\star - p_a) \\
&= \sum_{a: p_a \neq p_\star} \mathbb{E}[u_a^T]\Delta_a.
\end{aligned}
$$

Step 1: Show that $R^T = \sum_{a:p_a \neq p_\star} \mathbb{E}[u_a^T]\Delta_a$.

$$
\begin{aligned}
R^T &= Tp_\star - \sum_{t=0}^{T-1} \mathbb{E}[r^t] \\
&= Tp_\star - \sum_{t=0}^{T-1} \sum_{a \in A} \mathbb{P}\{Z_a^t\}\mathbb{E}[r^t | Z_a^t] \\
&= Tp_\star - \sum_{t=0}^{T-1} \sum_{a \in A} \mathbb{E}[z_a^t] p_a \\
&= \left(\sum_{a \in A} \mathbb{E}[u_a^T] \right) p_\star - \sum_{a \in A} \mathbb{E}[u_a^T] p_a \\
&= \sum_{a \in A} \mathbb{E}[u_a^T](p_\star - p_a) \\
&= \sum_{a:p_a \neq p_\star} \mathbb{E}[u_a^T]\Delta_a.
\end{aligned}
$$

To show the regret bound, we shall show for each sub-optimal arm a that

$$
\mathbb{E}[u_a^T] = O\left(\frac{1}{(\Delta_a)^2} \log(T) \right).
$$

Step 2: Split sub-optimal pulls into two regimes.

Step 2: Split sub-optimal pulls into two regimes.

To prove $\mathbb{E}[u_a^T] = O\left(\frac{1}{\Delta_a^2}\log(T)\right)$, we show $\mathbb{E}[u_a^T] \leq \bar{u}_a^T + C$ for some constant C.

Step 2: Split sub-optimal pulls into two regimes.

To prove $\mathbb{E}[u_a^T] = O\left(\frac{1}{\Delta_a^2}\log(T)\right)$, we show $\mathbb{E}[u_a^T] \leq \bar{u}_a^T + C$ for some constant C.

$$\mathbb{E}[u_a^T] = \sum_{t=0}^{T-1} \mathbb{E}[z_a^t]$$

Step 2: Split sub-optimal pulls into two regimes.

To prove $\mathbb{E}[u_a^T] = O\left(\frac{1}{\Delta_a^2}\log(T)\right)$, we show $\mathbb{E}[u_a^T] \leq \bar{u}_a^T + C$ for some constant C.

$$\begin{aligned}\mathbb{E}[u_a^T] &= \sum_{t=0}^{T-1} \mathbb{E}[z_a^t] \\ &= \sum_{t=0}^{T-1} \mathbb{P}\{Z_a^t\}\end{aligned}$$

Step 2: Split sub-optimal pulls into two regimes.

To prove $\mathbb{E}[u_a^T] = O\left(\frac{1}{\Delta_a^2}\log(T)\right)$, we show $\mathbb{E}[u_a^T] \leq \bar{u}_a^T + C$ for some constant C.

$$\begin{aligned}
\mathbb{E}[u_a^T] &= \sum_{t=0}^{T-1} \mathbb{E}[z_a^t] \\
&= \sum_{t=0}^{T-1} \mathbb{P}\{Z_a^t\} \\
&= \sum_{t=0}^{T-1} \mathbb{P}\{Z_a^t \text{ and } (u_a^t < \bar{u}_a^T)\} + \sum_{t=0}^{T-1} \mathbb{P}\{Z_a^t \text{ and } (u_a^t \geq \bar{u}_a^T)\}
\end{aligned}$$

Step 2: Split sub-optimal pulls into two regimes.

To prove $\mathbb{E}[u_a^T] = O\left(\frac{1}{\Delta_a^2}\log(T)\right)$, we show $\mathbb{E}[u_a^T] \leq \bar{u}_a^T + C$ for some constant C.

$$\begin{aligned}
\mathbb{E}[u_a^T] &= \sum_{t=0}^{T-1} \mathbb{E}[z_a^t] \\
&= \sum_{t=0}^{T-1} \mathbb{P}\{Z_a^t\} \\
&= \sum_{t=0}^{T-1} \mathbb{P}\{Z_a^t \text{ and } (u_a^t < \bar{u}_a^T)\} + \sum_{t=0}^{T-1} \mathbb{P}\{Z_a^t \text{ and } (u_a^t \geq \bar{u}_a^T)\} \\
&= A + B.
\end{aligned}$$

Step 2: Split sub-optimal pulls into two regimes.

To prove $\mathbb{E}[u_a^T] = O\left(\frac{1}{\Delta_a^2}\log(T)\right)$, we show $\mathbb{E}[u_a^T] \leq \bar{u}_a^T + C$ for some constant C.

$$\begin{aligned}
\mathbb{E}[u_a^T] &= \sum_{t=0}^{T-1} \mathbb{E}[z_a^t] \\
&= \sum_{t=0}^{T-1} \mathbb{P}\{Z_a^t\} \\
&= \sum_{t=0}^{T-1} \mathbb{P}\{Z_a^t \text{ and } (u_a^t < \bar{u}_a^T)\} + \sum_{t=0}^{T-1} \mathbb{P}\{Z_a^t \text{ and } (u_a^t \geq \bar{u}_a^T)\} \\
&= A + B.
\end{aligned}$$

We show A is upper-bounded by $\bar{u}_a^T$ and B is upper-bounded by a constant.

Step 3: Bounding A.

Step 3: Bounding A.

$$A = \sum_{t=0}^{T-1} \mathbb{P}\{Z_a^t \text{ and } (u_a^t < \bar{u}_a^T)\}$$

Step 3: Bounding A.

$$A = \sum_{t=0}^{T-1} \mathbb{P}\{Z_a^t \text{ and } (u_a^t < \bar{u}_a^T)\}$$

$$= \sum_{t=0}^{T-1} \sum_{m=0}^{\bar{u}_a^T - 1} \mathbb{P}\{Z_a^t \text{ and } (u_a^t = m)\}$$

Step 3: Bounding A.

$$A = \sum_{t=0}^{T-1} \mathbb{P}\{Z_a^t \text{ and } (u_a^t < \bar{u}_a^T)\}$$
$$= \sum_{t=0}^{T-1} \sum_{m=0}^{\bar{u}_a^T - 1} \mathbb{P}\{Z_a^t \text{ and } (u_a^t = m)\}$$
$$= \sum_{m=0}^{\bar{u}_a^T - 1} \sum_{t=0}^{T-1} \mathbb{P}\{Z_a^t \text{ and } (u_a^t = m)\}$$

Step 3: Bounding A.

$$A = \sum_{t=0}^{T-1} \mathbb{P}\{Z_a^t \text{ and } (u_a^t < \bar{u}_a^T)\}$$

$$= \sum_{t=0}^{T-1} \sum_{m=0}^{\bar{u}_a^T - 1} \mathbb{P}\{Z_a^t \text{ and } (u_a^t = m)\}$$

$$= \sum_{m=0}^{\bar{u}_a^T - 1} \sum_{t=0}^{T-1} \mathbb{P}\{Z_a^t \text{ and } (u_a^t = m)\}$$

$$= \sum_{m=0}^{\bar{u}_a^T - 1} \mathbb{P}\{(Z_a^0 \text{ and } (u_a^0 = m)) \text{ or } (Z_a^1 \text{ and } (u_a^1 = m)) \text{ or } \ldots \text{ or } (Z_a^{T-1} \text{ and } (u_a^{T-1} = m))\}$$

Step 3: Bounding A.

$$
\begin{aligned}
A &= \sum_{t=0}^{T-1} \mathbb{P}\{Z_a^t \text{ and } (u_a^t < \bar{u}_a^T)\} \\
&= \sum_{t=0}^{T-1} \sum_{m=0}^{\bar{u}_a^T - 1} \mathbb{P}\{Z_a^t \text{ and } (u_a^t = m)\} \\
&= \sum_{m=0}^{\bar{u}_a^T - 1} \sum_{t=0}^{T-1} \mathbb{P}\{Z_a^t \text{ and } (u_a^t = m)\} \\
&= \sum_{m=0}^{\bar{u}_a^T - 1} \mathbb{P}\{(Z_a^0 \text{ and } (u_a^0 = m)) \text{ or } (Z_a^1 \text{ and } (u_a^1 = m)) \text{ or } \ldots \text{ or } (Z_a^{T-1} \text{ and } (u_a^{T-1} = m))\} \\
&\leq \sum_{m=0}^{\bar{u}_a^T - 1} 1
\end{aligned}
$$

Step 3: Bounding A.

$$
\begin{aligned}
A &= \sum_{t=0}^{T-1} \mathbb{P}\{Z_a^t \text{ and } (u_a^t < \bar{u}_a^T)\} \\
&= \sum_{t=0}^{T-1} \sum_{m=0}^{\bar{u}_a^T - 1} \mathbb{P}\{Z_a^t \text{ and } (u_a^t = m)\} \\
&= \sum_{m=0}^{\bar{u}_a^T - 1} \sum_{t=0}^{T-1} \mathbb{P}\{Z_a^t \text{ and } (u_a^t = m)\} \\
&= \sum_{m=0}^{\bar{u}_a^T - 1} \mathbb{P}\{(Z_a^0 \text{ and } (u_a^0 = m)) \text{ or } (Z_a^1 \text{ and } (u_a^1 = m)) \text{ or } \ldots \text{ or } (Z_a^{T-1} \text{ and } (u_a^{T-1} = m))\} \\
&\leq \sum_{m=0}^{\bar{u}_a^T - 1} 1 \\
&= \bar{u}_a^T.
\end{aligned}
$$

Step 3: Bounding A.

$$
\begin{aligned}
A &= \sum_{t=0}^{T-1} \mathbb{P}\{Z_a^t \text{ and } (u_a^t < \bar{u}_a^T)\} \\
&= \sum_{t=0}^{T-1} \sum_{m=0}^{\bar{u}_a^T - 1} \mathbb{P}\{Z_a^t \text{ and } (u_a^t = m)\} \\
&= \sum_{m=0}^{\bar{u}_a^T - 1} \sum_{t=0}^{T-1} \mathbb{P}\{Z_a^t \text{ and } (u_a^t = m)\} \\
&= \sum_{m=0}^{\bar{u}_a^T - 1} \mathbb{P}\{(Z_a^0 \text{ and } (u_a^0 = m)) \text{ or } (Z_a^1 \text{ and } (u_a^1 = m)) \text{ or } \ldots \text{ or } (Z_a^{T-1} \text{ and } (u_a^{T-1} = m))\} \\
&\leq \sum_{m=0}^{\bar{u}_a^T - 1} 1 \\
&= \bar{u}_a^T.
\end{aligned}
$$

We have used the fact that for $0 \leq i < j \leq t-1$, $(Z_a^i \text{ and } (u_a^i = m))$ and $(Z_a^j \text{ and } (u_a^j = m))$ are mutually exclusive.

Step 4.1: Bounding B.

Step 4.1: Bounding B.

$$B = \sum_{t=0}^{T-1} \mathbb{P}\{Z_a^t \text{ and } (u_a^t \geq \bar{u}_a^T)\}$$

Step 4.1: Bounding B.

$$B = \sum_{t=0}^{T-1} \mathbb{P}\{Z_a^t \text{ and } (u_a^t \geq \bar{u}_a^T)\}$$
$$\leq \sum_{t=0}^{T-1} \mathbb{P}\left\{\left(\hat{p}_a^t + \sqrt{\frac{2}{u_a^t}\ln(t)} \geq \hat{p}_\star^t + \sqrt{\frac{2}{u_\star^t}\ln(t)}\right) \text{ and } (u_a^t \geq \bar{u}_a^T)\right\}$$

Step 4.1: Bounding B.

$$B = \sum_{t=0}^{T-1} \mathbb{P}\{Z_a^t \text{ and } (u_a^t \geq \bar{u}_a^T)\}$$

$$\leq \sum_{t=0}^{T-1} \mathbb{P}\left\{\left(\hat{p}_a^t + \sqrt{\frac{2}{u_a^t}\ln(t)} \geq \hat{p}_\star^t + \sqrt{\frac{2}{u_\star^t}\ln(t)}\right) \text{ and } (u_a^t \geq \bar{u}_a^T)\right\}$$

$$\leq \sum_{t=0}^{T-1} \sum_{x=\bar{u}_a^T}^{t} \sum_{y=1}^{t} \mathbb{P}\left\{\hat{p}_a(x) + \sqrt{\frac{2}{x}\ln(t)} \geq \hat{p}_\star(y) + \sqrt{\frac{2}{y}\ln(t)}\right\} \text{ where}$$

$\hat{p}_a(x)$ is the empirical mean of the first x pulls of arm a, and

$\hat{p}_\star(y)$ is the empirical mean of the first y pulls of arm $\star$.

Step 4.2: Bounding B.

- Fix $x \in \{\bar{u}_a^T, \bar{u}_a^T + 1, \ldots, t\}$ and $y \in \{1, 2, \ldots, t\}$.

Step 4.2: Bounding B.

- Fix $x \in \{\bar{u}_a^T, \bar{u}_a^T + 1, \ldots, t\}$ and $y \in \{1, 2, \ldots, t\}$.
- We have:

$$\hat{p}_a(x) + \sqrt{\frac{2}{x}\ln(t)} \geq \hat{p}_\star(y) + \sqrt{\frac{2}{y}\ln(t)}$$
$$\implies \left(\hat{p}_a(x) + \sqrt{\frac{2}{x}\ln(t)} \geq p_\star\right) \text{ or } \left(\hat{p}_\star(y) + \sqrt{\frac{2}{y}\ln(t)} < p_\star\right).$$

Step 4.2: Bounding B.

- Fix $x \in \{\bar{u}_a^T, \bar{u}_a^T + 1, \ldots, t\}$ and $y \in \{1, 2, \ldots, t\}$.
- We have:

$$\hat{p}_a(x) + \sqrt{\frac{2}{x}\ln(t)} \geq \hat{p}_\star(y) + \sqrt{\frac{2}{y}\ln(t)}$$
$$\implies \left(\hat{p}_a(x) + \sqrt{\frac{2}{x}\ln(t)} \geq p_\star\right) \text{ or } \left(\hat{p}_\star(y) + \sqrt{\frac{2}{y}\ln(t)} < p_\star\right).$$

- Since $x \geq \bar{u}_a^T$, we have $\sqrt{\frac{2}{x}\ln(t)} \leq \sqrt{\frac{2}{\bar{u}_a^T}\ln(t)} \leq \frac{\Delta_a}{2}$, and so

$$\hat{p}_a(x) + \sqrt{\frac{2}{x}\ln(t)} \geq p_\star \implies \hat{p}_a(x) \geq p_a + \frac{\Delta_a}{2}.$$

Step 4.2: Bounding B.

- Fix $x \in \{\bar{u}_a^T, \bar{u}_a^T + 1, \dots, t\}$ and $y \in \{1, 2, \dots, t\}$.
- We have:

$$\hat{p}_a(x) + \sqrt{\frac{2}{x}\ln(t)} \geq \hat{p}_\star(y) + \sqrt{\frac{2}{y}\ln(t)}$$

$$\implies \left(\hat{p}_a(x) + \sqrt{\frac{2}{x}\ln(t)} \geq p_\star\right) \text{ or } \left(\hat{p}_\star(y) + \sqrt{\frac{2}{y}\ln(t)} < p_\star\right).$$

- Since $x \geq \bar{u}_a^T$, we have $\sqrt{\frac{2}{x}\ln(t)} \leq \sqrt{\frac{2}{\bar{u}_a^T}\ln(t)} \leq \frac{\Delta_a}{2}$, and so

$$\hat{p}_a(x) + \sqrt{\frac{2}{x}\ln(t)} \geq p_\star \implies \hat{p}_a(x) \geq p_a + \frac{\Delta_a}{2}.$$

- In summary:

$$\hat{p}_a(x) + \sqrt{\frac{2}{x}\ln(t)} \geq \hat{p}_\star(y) + \sqrt{\frac{2}{y}\ln(t)} \implies \left(\hat{p}_a(x) \geq p_a + \frac{\Delta_a}{2}\right) \text{ or } \left(\hat{p}_\star(y) < p_\star - \sqrt{\frac{2}{y}\ln(t)}\right).$$

Step 4.3: Bounding B.

Continuing from Step 4.1, and now invoking Hoeffding's Inequality:

$$B \leq \sum_{t=0}^{T-1} \sum_{x=\bar{u}_a^T}^{t} \sum_{y=1}^{t} \mathbb{P}\left\{\hat{p}_a(x) + \sqrt{\frac{2}{x}\ln(t)} \geq \hat{p}_\star(y) + \sqrt{\frac{2}{y}\ln(t)}\right\}$$

Step 4.3: Bounding B.

Continuing from Step 4.1, and now invoking Hoeffding's Inequality:

$$B \leq \sum_{t=0}^{T-1} \sum_{x=\bar{u}_a^T}^{t} \sum_{y=1}^{t} \mathbb{P}\left\{\hat{p}_a(x) + \sqrt{\frac{2}{x}\ln(t)} \geq \hat{p}_\star(y) + \sqrt{\frac{2}{y}\ln(t)}\right\}$$

$$\leq \sum_{t=0}^{T-1} \sum_{x=\bar{u}_a^T}^{t} \sum_{y=1}^{t} \left(\mathbb{P}\left\{\hat{p}_a(x) \geq p_a + \frac{\Delta_a}{2}\right\} + \mathbb{P}\left\{\hat{p}_\star(y) < p_\star - \sqrt{\frac{2}{y}\ln(t)}\right\}\right)$$

Step 4.3: Bounding B.

Continuing from Step 4.1, and now invoking Hoeffding's Inequality:

$$B \leq \sum_{t=0}^{T-1} \sum_{x=\bar{u}_a^T}^{t} \sum_{y=1}^{t} \mathbb{P}\left\{\hat{p}_a(x) + \sqrt{\frac{2}{x}\ln(t)} \geq \hat{p}_\star(y) + \sqrt{\frac{2}{y}\ln(t)}\right\}$$

$$\leq \sum_{t=0}^{T-1} \sum_{x=\bar{u}_a^T}^{t} \sum_{y=1}^{t} \left(\mathbb{P}\left\{\hat{p}_a(x) \geq p_a + \frac{\Delta_a}{2}\right\} + \mathbb{P}\left\{\hat{p}_\star(y) < p_\star - \sqrt{\frac{2}{y}\ln(t)}\right\}\right)$$

$$\leq \sum_{t=0}^{T-1} \sum_{x=\bar{u}_a^T}^{t} \sum_{y=1}^{t} \left(e^{-2x\left(\frac{\Delta_a}{2}\right)^2} + e^{-2y\left(\sqrt{\frac{2}{y}\ln(t)}\right)^2}\right)$$

Step 4.3: Bounding B.

Continuing from Step 4.1, and now invoking Hoeffding's Inequality:

$$B \leq \sum_{t=0}^{T-1} \sum_{x=\bar{u}_a^T}^{t} \sum_{y=1}^{t} \mathbb{P}\left\{\hat{p}_a(x) + \sqrt{\frac{2}{x}\ln(t)} \geq \hat{p}_\star(y) + \sqrt{\frac{2}{y}\ln(t)}\right\}$$

$$\leq \sum_{t=0}^{T-1} \sum_{x=\bar{u}_a^T}^{t} \sum_{y=1}^{t} \left(\mathbb{P}\left\{\hat{p}_a(x) \geq p_a + \frac{\Delta_a}{2}\right\} + \mathbb{P}\left\{\hat{p}_\star(y) < p_\star - \sqrt{\frac{2}{y}\ln(t)}\right\}\right)$$

$$\leq \sum_{t=0}^{T-1} \sum_{x=\bar{u}_a^T}^{t} \sum_{y=1}^{t} \left(e^{-2x\left(\frac{\Delta_a}{2}\right)^2} + e^{-2y\left(\sqrt{\frac{2}{y}\ln(t)}\right)^2}\right)$$

$$\leq \sum_{t=0}^{T-1} \sum_{x=\bar{u}_a^T}^{t} \sum_{y=1}^{t} \left(e^{-4\ln(t)} + e^{-4\ln(t)}\right) \leq \sum_{t=0}^{T-1} t^2 \left(\frac{2}{t^4}\right) \leq \sum_{t=0}^{\infty} \frac{2}{t^2} = \frac{\pi^2}{3}.$$

Step 4.3: Bounding B.

Continuing from Step 4.1, and now invoking Hoeffding's Inequality:

$$B \leq \sum_{t=0}^{T-1} \sum_{x=\bar{u}_a^T}^{t} \sum_{y=1}^{t} \mathbb{P}\left\{\hat{p}_a(x) + \sqrt{\frac{2}{x}\ln(t)} \geq \hat{p}_\star(y) + \sqrt{\frac{2}{y}\ln(t)}\right\}$$

$$\leq \sum_{t=0}^{T-1} \sum_{x=\bar{u}_a^T}^{t} \sum_{y=1}^{t} \left(\mathbb{P}\left\{\hat{p}_a(x) \geq p_a + \frac{\Delta_a}{2}\right\} + \mathbb{P}\left\{\hat{p}_\star(y) < p_\star - \sqrt{\frac{2}{y}\ln(t)}\right\}\right)$$

$$\leq \sum_{t=0}^{T-1} \sum_{x=\bar{u}_a^T}^{t} \sum_{y=1}^{t} \left(e^{-2x\left(\frac{\Delta_a}{2}\right)^2} + e^{-2y\left(\sqrt{\frac{2}{y}\ln(t)}\right)^2}\right)$$

$$\leq \sum_{t=0}^{T-1} \sum_{x=\bar{u}_a^T}^{t} \sum_{y=1}^{t} \left(e^{-4\ln(t)} + e^{-4\ln(t)}\right) \leq \sum_{t=0}^{T-1} t^2 \left(\frac{2}{t^4}\right) \leq \sum_{t=0}^{\infty} \frac{2}{t^2} = \frac{\pi^2}{3}.$$

We are done.