CSE 547 - Assignment 4

Philip Pham

June 5, 2018

Problem 0

List of collaborators: I have not collaborated with anyone.

List of acknowledgements: None.

Certify that you have read the instructions: I have read and understood these policies.

Problem 1: Logarithmic Regret of UCB

We will consider the multi-armed bandit setting discussed in class, where the actions $a \in \{1, \cdots, K\}$, μ_a is the mean reward provided by arm a, and X_t is reward observed at time t if we pull arm a. As in class, we assume that the observed rewards are bounded a $0 \leq X_t \leq 1$ almost surely.

Recall $\mu_* = \max_a \mu_a$, and let a_* be the index of an optimal arm. Define Δ_a as:

$$\Delta_a = \mu_* - \mu_a \tag{1}$$

and define:

$$\Delta_{\min} = \min_{a \neq a_*} \Delta_a. \tag{2}$$

In this problem, we seek to prove the following theorem:

Theorem 1. *The UCB algorithm (with an appropriate setting of the parameters) has a regret bound that is:*

$$T\mu_* - \mathbb{E}\left[\sum_{t \leq T} X_t\right] \leq c\frac{K \log T}{\Delta_{\min}}, \tag{3}$$

where c is a universal constant.

Let's prove this!

Let $N_{a,t}$ be the number of times we pulled arm a up to time t. Recall from class that by Hoeffding's bound (and the union bound), we can provide a confidence bound for an arbitrary algogrithm as follows: with probability greater than $1 - \delta$, we have that fo all arms and for all time steps $K \leq t \leq T$:

$$\mathbb{P}\left(\forall t, a, |\hat{\mu}_{a,t} - \mu_a| \leq c_2\sqrt{\frac{\log\left(T/\delta\right)}{N_{a,t}}}\right) \geq 1 - \delta, \tag{4}$$

where c_2 is some universal constant. Note that the algorithm starts the first K steps by sampling each arm once, so we can assume $t \geq K$.

1. Now consider the UCB algorithm using this confidence interval. Argue that with probability greater thatn $1-\delta$, the total number of times that an sub-optimal arm a will be pulled up to time T will be bounded as follows:

$$N_{a,T} \leq c_3 \frac{\log(T/\delta)}{\Delta_a^2} \tag{5}$$

for some constant c_3.

Solution

Proof. This bound follows from Equation 4. Then, for any a and $t \in [K, T]$ with probability greater than $1-\delta$, we have that

$$\begin{aligned} |\hat{\mu}_{a,t} - \mu_a| &\leq c_2 \sqrt{\frac{\log(T/\delta)}{N_{a,t}}} \\ \sqrt{N_{a,t}} &\leq c_2 \frac{\sqrt{\log(T/\delta)}}{|\hat{\mu}_{a,t} - \mu_a|} \\ N_{a,t} &\leq c_2^2 \frac{\log(T/\delta)}{(\hat{\mu}_{a,t} - \mu_a)^2}. \end{aligned}$$

If we let $c_3 = c_2^2$, substitute $\Delta_a^2 = (\hat{\mu}_{a,t} - \mu_a)^2$, and fix $t = T$, we have Equation 5 with probability greater than $1-\delta$ as desired. □

2. Argue that the observed regret of UCB is bounded as follows: with probability greater than $1-\delta$, we have that:

$$T\mu_* - \sum_{t \leq T} \mu_{at} \leq c_3 \sum_{a \neq a_*} \frac{\log(T/\delta)}{\Delta_a}, \tag{6}$$

where a_t is the arm chosen by the algorithm at time t.

Solution

Proof. Equation 6 follows from Equation 5, noting that $\sum_{a=1}^{K} N_{a,T} = T$, and seeing that $\Delta_{a_*} = \mu_* - \mu_{a_*} = 0$

We have that

$$
\begin{aligned}
T\mu_* - \sum_{t \le T} \mu_{a_t} &= \sum_{a=1}^{K} N_{a,T} \left(\mu_* - \mu_a\right) p \\
&= \sum_{a=1}^{K} \Delta_a N_{a,T} \\
&= \sum_{a \neq a_*} \Delta_a N_{a,T} \\
&\le \sum_{a \neq a_*} \Delta_a \left(c_3 \frac{\log\left(T/\delta\right)}{\Delta_a^2} \right) \\
&= \sum_{a \neq a_*} c_3 \frac{\log\left(T/\delta\right)}{\Delta_a},
\end{aligned}
$$

which gives Equation 6 with probability $1-\delta$ as desired. □

3. Now show that the expected regret of UCB is bounded as:

$$
T\mu_* - \mathbb{E}\left[\sum_{t \le T} X_t\right] \le c_4 \sum_{a \neq a_*} \frac{\log\left(T\right)}{\Delta_a}. \tag{7}
$$

Solution

Proof. Fix $\delta = 1/T^2$ as in the proof of Lemma 3.1 from the lecture notes.

We have that

$$
\begin{aligned}
\mathbb{E}\left[\sum_{t \le T} \left(\mu_* - X_t\right)\right] &= T\mu_* - \mathbb{E}\left[\sum_{t \le T} X_t\right] = T\mu_* - \sum_{t \le T} \mu_{a_t} \\
&\le \left(1-\delta\right) c_3 \sum_{a \neq a_*} \frac{\log\left(T/\delta\right)}{\Delta_a} + \delta T \\
&= \left(1 - \frac{1}{T^2}\right) c_3 \sum_{a \neq a_*} \frac{\log\left(T\right) + 2\log\left(T\right)}{\Delta_a} + \frac{1}{T} \\
&= 3c_3 \sum_{a \neq a_*} \frac{\log\left(T\right)}{\Delta_a} - \frac{3c_3}{T^2} \sum_{a \neq a_*} \frac{\log\left(T\right)}{\Delta_a} + \frac{1}{T} \\
&= O\left(\sum_{a \neq a_*} \frac{\log\left(T\right)}{\Delta_a} \right)
\end{aligned}
$$

asymptotically since the other terms decay with T. Thus, it follows that there exists some c_4 such that

$$
T\mu_* - \mathbb{E}\left[\sum_{t \le T} X_t\right] \le c_4 \sum_{a \neq a_*} \frac{\log\left(T\right)}{\Delta_a}.
$$

□

4. Now argue that the theorem follows and specify what the UCB algorithm is (with parameters set appropriately).

Solution

Proof. Applying Equation 2 to Equation 7, we have that

$$\begin{aligned} T\mu_* - \mathbb{E}\left[\sum_{t \le T} X_t\right] &\le c_4 \sum_{a \ne a_*} \frac{\log(T)}{\Delta_a} \\ &\le c_4 \sum_{a \ne a_*} \frac{\log(T)}{\Delta_{\min}} \\ &\le c_4 K \frac{\log(T)}{\Delta_{\min}}, \end{aligned}$$

which gives us Equation 3 if we define $c = c_4$. □

Thus, we have the following UCB algorithm.

(1) Try each of the K arms once.

(2) Fix t. Calculate

$$U_{a,t} = \hat{\mu}_{a,t} + c_2 \sqrt{3 \frac{\log T}{N_{a,t}}} \tag{8}$$

for all $a = 1, 2, \ldots, K$. Pull arm $a_*^{(t)} = \arg\max_a U_{a,t}$.

(3) Repeat Step (2) T times.