CS 7545: Machine Learning Theory **Fall 2019**

Lecture 18: UCB Algorithm + Analysis

Lecturer: Jacob Abernethy *Scribes: Weiyang Liu, Dantong Zhu*

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications.*

18.1 Protocol of Bandit Algorithms

Assume that for each $i = 1, ..., n$, D_i is sub-gaussian with variance proxy 1 and the mean of D_i is μ_i. The goal is to find out the distribution D_i that has the highest mean.

Without loss of generality, assume $\mu_1 = \max_{1\le i\le n}\{\mu_i\}$. Let $\Delta_i = \mu_1 - \mu_i$ for $i = 2, ..., n$.

In general, a bandit algorithm selects arm $I_t \in [n]$ at time t. For every $i \in [n]$ and every $t \in [T]$, let X_i^t be the reward on arm i at time t, which is drawn from the distribution D_i at time t; let $N_i^t = \sum_{s=1}^{t-1} 1[I_t = i]$, i.e. the number of times arm i has picked by time t; and let $\hat{\mu_i^t} = \frac{1}{N_i^t}\sum_{i=1}^{t-1} 1[I_t = i] \cdot X_i^t$, i.e. the empirical mean of D_i by time t.

18.2 UCB Algorithm

For every $i \in [n]$, $t \in [T]$, define

$$\mathrm{UCB}_i^t = \hat{\mu_i^t} + \sqrt{\frac{2\log(1/\delta)}{N_i^t}},$$

where δ is a parameter that can be tuned. In each UCB_i^t, $\hat{\mu_i^t}$ is the *estimated reward*, and we call $\sqrt{\frac{2\log(1/\delta)}{N_i^t}}$ the *exploration bonus.*

The UCB algorithm, at each time t, plays

$$I_t = \arg\max_{1\le i\le n}\{\mathrm{UCB}_i^t\}.$$

The goal of the rest of the lecture is to prove the upper bound of the regret of UCB in the following theorem.

Theorem 18.1 *With a suitable choice of δ,*

$$\mathbb{E}(\mathit{Regret}_T) \le 16\sum_{i=2}^{n}\frac{\log(T+1)}{\Delta_i} + 2\sum_{i=2}^{n}\Delta_i.$$

In the rest of this notes, we consider the parameter $\delta > 0$ as a fixed number, which will be defined explicitly at the end of the file.

Claim 18.2 $\mathbb{P}(\mu_i > \mathit{UCB}_i^t) \le \delta$ *for all $i \in [n]$, $t \in [T]$.*

Proof: Let $i \in [n]$ and $t \in [T]$. By definition, $\mathrm{UCB}_i^t = \hat{\mu_i^t} + \sqrt{\frac{2\log(1/\delta)}{N_i^t}}$, meaning that $\mathbb{P}(\mu_i > \mathrm{UCB}_i^t) = \mathbb{P}(\mu_i - \hat{\mu_i^t} > \sqrt{\frac{2\log(1/\delta)}{N_i^t}})$. Let $\epsilon = \sqrt{\frac{2\log(1/\delta)}{N_i^t}}$. By Hoeffding's inequality, it follows that

$$\mathbb{P}(\mu_i > \mathrm{UCB}_i^t) \le \mathbb{P}(\mu_i - \hat{\mu_i^t} > \epsilon) \le \exp(-\frac{1}{2}N_i^t\epsilon^2) \le \exp(-\frac{1}{2}N_i^t\frac{2\log(1/\delta)}{N_i^2}) = \delta.$$

■

Now for each $i \in [n]$, let $k_i = \lceil \frac{8\log(1/\delta)}{\Delta_i^2} \rceil$. For analysis purpose, for each $i \in [n]$ let

$$\hat{\mu_i}^{(k_i)} = \mu_i^t \text{ where } t = \min\{t' : N_i^{t'} = k_i\}, \text{ if } N_i^{T+1} \geq k_i;$$

$$\hat{\mu_i}^{(k_i)} = \frac{1}{k_i}\big(N_i^{T+1}\mu_i^{T+1} + \sum_{j=1}^{k_i - N_i^{T+1}} Y_j\big) \text{ where } Y_j \sim D_i \forall j = 1, ..., k_i - N_i^{T+1}, \text{ if } N_i^{T+1} < k_i.$$

Also define events

$$G_i = \{\mu_1 < \text{UCB}_1^t, \forall t = 1, ..., T\} \cap \{\hat{\mu_i}^{(k_i)} + \sqrt{\frac{2\log(1/\delta)}{k_i}} < \mu_1\}, \forall i = 1, ..., n.$$

Claim 18.3 *If G_i is true for some $i \geq 2$, then $N_i^{T+1} \leq k_i$.*

Proof: For the sake of a contradiction, assume that $N_i^{T+1} > k_i$ for some $i \geq 2$. Let t be the final round where $N_i^t = k_i$. Note this implies that $I_t = i$. Since G_i is true, we know that

$$\text{UCB}_1^t > \mu_1 > \hat{\mu_i}^{(k_i)} + \sqrt{\frac{2\log(1/\delta)}{k_i}}.$$

By the definition of $\hat{\mu_i}^{(k_i)}$ and the choice of t, it follows that

$$\hat{\mu_i}^{(k_i)} + \sqrt{\frac{2\log(1/\delta)}{k_i}} = \hat{\mu_i}^t + \sqrt{\frac{2\log(1/\delta)}{N_i^t}} = \text{UCB}_i^t.$$

This means that $\text{UCB}_1^t > \text{UCB}_i^t$, which is a contradiction since according to the algorithm, $i = I_t = \arg\max_{1\leq j\leq n}\{\text{UCB}_j^t\}$. ■

Notice that $\forall i \geq 2$, we have

$$\begin{aligned}\mathbb{E}(N_i^{T+1}) &= \underbrace{\mathbb{E}(N_i^{T+1} \cdot 1[G_i])}_{<k_i} + \underbrace{\mathbb{E}(N_i^{T+1} \cdot 1[\bar{G}_i])}_{\leq T\cdot\mathbb{P}(\bar{G}_i)} \\ &< k_i + T \cdot \mathbb{P}(\bar{G}_i)\end{aligned}$$

which indicates that we need to bound $\mathbb{P}(\bar{G}_i)$.

Note that we have

$$\bar{G}_i = \Big\{\exists t \leq T \text{ s.t. } \mu_1 \geq \text{UCB}_1^t\Big\} \cup \Big\{\hat{\mu}_i^{(k_i)} + \sqrt{\frac{2\log(\frac{1}{\delta})}{k_i}} \geq \mu_1\Big\}.$$

Let $B_t = \{\mu_1 \geq \text{UCB}_1^t\}$ for all $t \in [T]$ and $C = \{\hat{\mu}_i^{(k_i)} + \sqrt{\frac{2\log(\frac{1}{\delta})}{k_i}} \geq \mu_1\}$. Then using union bound, we have that

$$\mathbb{P}(\bar{G}_i) \leq \mathbb{P}(C) + \sum_{t=1}^{T} \mathbb{P}(B_t).$$

Recall that $\mathbb{P}(\mu_i > \text{UCB}_i^t) \leq \delta$, $\forall i, t$. Then we end up with $\mathbb{P}(\bar{G}_i) \leq \mathbb{P}(C) + T \cdot \delta$.

Note that we have

$$\begin{aligned}\mathbb{P}(C) &= \mathbb{P}\bigg(\hat{\mu}_i^{(k_i)} + \sqrt{\frac{2\log(\frac{1}{\delta})}{k_i}} \geq \mu_1\bigg) \\ &= \mathbb{P}\bigg(\hat{\mu}_i^{(k_i)} + \sqrt{\frac{2\log(\frac{1}{\delta})}{k_i}} \geq \mu_1 + \Delta_i\bigg) \\ &= \mathbb{P}\bigg(\mu_1 - \hat{\mu}_i^{(k_i)} < \sqrt{\frac{2\log(\frac{1}{\delta})}{k_i}} - \Delta_i\bigg).\end{aligned}$$

Because $k_i = \lceil \frac{8\log(\frac{1}{\delta})}{\Delta_i^2} \rceil$, then we have that $\sqrt{\frac{2\log(\frac{1}{\delta})}{k_i}} \leq \sqrt{\frac{2\log(\frac{1}{\delta})}{8\log(\frac{1}{\delta})/\Delta_i^2}} = \frac{\Delta_i}{2}$. Putting this into the expression of $\mathbb{P}(C)$, we will have that

$$\begin{aligned}\mathbb{P}(C) &\leq \mathbb{P}\big(\mu_1 - \hat{\mu}_i^{k_i} < -\frac{\Delta_i}{2}\big) \\ \text{(Hoeffding)} \quad &\leq \exp\left(\frac{(-\frac{\Delta_i}{2})^2 \cdot k_i}{2}\right) \\ &\leq \exp\left(-\frac{1}{8}\Delta_i^2 \cdot \frac{8\log(\frac{1}{\delta})}{\Delta_i^2}\right) = \delta\end{aligned}$$

which results in $\mathbb{P}(\bar{G}_i) \leq \delta + T \cdot \delta = \delta(T+1)$. Therefore, we have that $\mathbb{E}(N_i^{T+1}) < k_i + \delta \cdot T(T+1)$ which leads to

$$\begin{aligned}\mathbb{E}(\text{Regret}_T) &= \sum_{i=2}^{n} \Delta_i \cdot \mathbb{E}(N_i^{T+1}) \\ &= \sum_{i=2}^{n} \Delta_i\big(k_i + \delta \cdot T(T+1)\big) \\ &\leq \sum_{i=2}^{n} \Delta_i\big(\frac{8\log(\frac{1}{\delta})}{\Delta_i^2} + 1 + \delta(T+1)^2\big).\end{aligned}$$

After setting $\delta = \frac{1}{(T+1)^2}$, we obtain that

$$\begin{aligned}\mathbb{E}(\text{Regret}) &\leq \sum_{i=2}^{n} \Delta_i\big(\frac{8\log(\frac{1}{\delta})}{\Delta_i^2} + 1 + 1\big) \\ &= \sum_{i=2}^{n} \big(\frac{16\log(T+1)}{\Delta_i} + 2\Delta_i\big) \\ &= 16\sum_{i=2}^{n} \frac{\log(T+1)}{\Delta_i} + 2\sum_{i=2}^{n} \Delta_i\end{aligned}$$