IE 498: Online Learning and Decision Making, Fall 2019 Sep 12th

Lecture 06: lil' UCB Algorithm

Lecturer: Yuan Zhou *Scribe: Zekun Yang, Peijun Xiao*

1 Introduction

Recall in last lecture, we introduce the successive reject algorithm and show that the sample complexity to solve the best arm identification problem is $\mathcal{O}\left(\sum_{i=2}^{n}(\Delta_i^{-2}(\log n + \log \delta^{-1} + \log\log \Delta_i^{-1}))\right)$. Our goal for this lecture is to remove the extra $\log(n)$ and obtain the complexity $\mathcal{O}\left(\sum_{i=2}^{n}(\Delta_i^{-2}(\log \delta^{-1} + \log\log \Delta_i^{-1}))\right)$. We are only seeing the lower bound, so we do not need to care about the $\log\log$ factor. Basically, the $\log(n)$ factor comes from the successive reject algorithm. Therefore, we usually apply a union bound to all the arms, which give us a $\log(n)$ factor. In this lecture, we will not apply the union bound on all the arms. Instead, we introduce a new algorithm called **LiL' UCB** [1] (in section 3) and apply some techniques to show the algorithm can achieve our desired complexity. Before we present the algorithm and the new techniques, we first show a lemma to give some intuition. This lemma explains why the algorithm is called **LiL' UCB**.

2 Lemma of Law-of-Iterated-Log Bound

Let $X_1, X_2, X_3, ...,$ be a sequence of i.i.d random variables such that $X_i \in [-1, 1]$ and $\mathbb{E}(X_i) = 0$. Then, we have a concentration bound for their sum.

Here we want apply an uniform concentration bound to every sum. If we look at a particular t, by Hoeffding's inequality, we got the upper bound $\sqrt{t\log(t/\delta)}$.

In order to make it hold for every δ, we put a large enough constant 10. Therefore, we have:

$$Pr\left[\forall t = 1, 2, 3, ..., \sum_{i=1}^{t} X_i \leq \sqrt{t \cdot 10\log(t/\delta)}\right] \geq 1 - \delta. \quad (1)$$

The reason why we could do in this way is that we can get a union bound for each t and:

$$\text{for each t:} \quad p(failure) \leq \left(\frac{\delta}{t}\right)^2. \quad (2)$$

By the union bound, we just sum over all the t and the sum would converge to a constant. So the whole thing would be bounded by δ.

Indeed, we can put something even better. The form could be like this:

$$Pr\left[\forall t=1,2,3,...,\ \sum_{i=1}^{t} X_i \leq \sqrt{4t\ln\frac{3\ln(t+1)}{\delta}}\right] \geq 1-\delta. \tag{3}$$

Compared the last line with the previous bound, we have $\ln\ln t$ term. This inequality is called *Finite Sample Law of Iterated Log.*

We can see that:

$$\limsup_{t\to+\infty}\frac{x_1+x_2+...+x_t}{\sqrt{t\ln\ln t}} \leq \mathcal{O}(1). \tag{4}$$

It shows that *Law of Iterated Log* can give us a finite sample bound.

To sum up from the above, we present the lemma of lil' UCB as the following:

Lemma 1(Law-of-Iterated-Log Bound).

Let $X_1, X_2, X_3, ...$ be i.i.d. centered random variables bounded between $[-1,1]$.

For each $\delta \in (0,1/2)$, we have

$$Pr\left[\forall t=1,2,3,...,\ \sum_{i=1}^{t} X_i \leq \sqrt{4t\ln\frac{2.89\ln(t+1)}{\delta}}\right] \geq 1-\delta. \tag{5}$$

Proof.

Let

$$\varepsilon = \left\{\forall k=1,2,3,...,\ \sum_{i=1}^{2^k} X_i \leq \sqrt{2^{k-1}ln(\frac{4k^2}{\delta})}\right\}. \tag{6}$$

By Hoeffding's Inequality:

$$Pr\left[\varepsilon\right] \geq 1-\sum_{k=1}^{+\infty}\frac{\delta}{4k^2} = 1-\frac{\delta}{4}\cdot\frac{\pi^2}{6} \geq 1-\delta/2. \tag{7}$$

For each k=1,2,3,..., let

$$\mathcal{F}_k = \left\{\forall s=1,2,3,...,2^{k-1},\ \sum_{i=2^k+1}^{2^k+s} X_i \leq \sqrt{2^{k-1}\ln(\frac{4k^2}{\delta})}\right\}. \tag{8}$$

By *Hoeffding's maximal inequality* (lemma 3 in section 5):

$$Pr[\mathcal{F}_k] \geq 1-\frac{\delta}{4k^2}. \tag{9}$$

Therefore, via union bound,

$$Pr\left[\varepsilon \cap \wedge_{k=1}^{+\infty} \mathcal{F}_k\right] \geq 1 - \frac{\delta}{2} - \sum_{k=1}^{+\infty} \frac{\delta}{4k^2} \geq 1 - \delta. \tag{10}$$

When $\varepsilon \cap \wedge_{k=1}^{+\infty} \mathcal{F}_k$ happens, for each $t \geq 2$, write $t = 2^k + s$ ($s \in [0, 2^k - 1]$),

$$\sum_{i=1}^{t} x_i = \sum_{i=1}^{2^k} x_i + \sum_{i=2^k+1}^{2^k+s} x_i \leq \sqrt{2^{k-1} \ln \frac{4k^2}{\delta}} + \sqrt{2^{k-1} \ln \frac{4k^2}{\delta}} \tag{11}$$

$$= \sqrt{2 \cdot 2^k \ln \frac{4k^2}{\delta}} \leq \sqrt{2tln \frac{4(\ln t / \ln 2)^2}{\delta}} \leq \sqrt{4t \ln \frac{2.89 \ln(t+1)}{\delta}}. \tag{12}$$

□

Remark 1. *(More refined bound in.[Jamieson et.al' 2014]).*

For any $\varepsilon \in (0,1)$ and $\delta \in (0, ln(1+\varepsilon)/e)$, we have

$$Pr\left[\forall t \geq 1, \sum_{i=1}^{t} x_i \leq (1+\sqrt{\varepsilon})\sqrt{2(1+\varepsilon)t \ln \frac{\ln(1+\varepsilon)t}{\delta}}\right] \geq 1 - \frac{2+\varepsilon}{\varepsilon}\left(\frac{\delta}{\ln(1+\varepsilon)}\right)^{1+\varepsilon}.$$

3 Algorithm and Complexity

We can design the algorithm **LiL UCB** like this:

1.Play each arm once. Set $T_i(t) = 1$ for all i, $t = n$.

2.WHILE $\forall i \in [n],\ T_i(t) < 1 + 80 \sum_{j \neq i} T_j(t)$, DO:

  2a. Let $i_t = \arg\max_{i \in [n]} \{\hat{\mu}_{i,T_i(t)} + 2\mathcal{M}(T_i(t), \delta)\}$.

  2b. Play it and set

$$T_i(t+1) = \begin{cases} T_i(t) & i_t \neq i \\ T_i(t) + 1 & i_t = i \end{cases}; \qquad t \leftarrow t+1$$

3. RETURN $\quad \arg\max_{i \in [n]} \{T_i(t)\}$.

Let us define:

$$\mathcal{M}(t,\delta) = \sqrt{\frac{1}{t} log \frac{t}{\delta}} \Rightarrow \mathcal{M}(t,\delta) = 2\sqrt{\frac{ln(6ln(t+1))/\delta}{t}}.$$

The reason why we put constant 6 here is that we want both upper bound and lowerbound for the union bound.

In the step 2 of the algorithm, there is not a single arm playing for too many times. If we do the UCB, we will always play the best arm. In the step 2a, $\hat{\mu}_{i,T_i(t)}$ means the number of samples we have already made to the *ith* arm and $\mathcal{M}$ function defines our UCB based on our success parameter δ and $T_i(t)$. Finally, we return the arm i played for the maximum number of time which is more than 80 times of that of any other arm.

Theorem 1. *Apply the algorithm **LiL UCB** to solve the best arm identification, the complexity is*

$$\mathcal{O}(\sum_{i=2}^{n}(\Delta_i^{-2}(\log\delta^{-1}+\log\log\Delta_i^{-1}))). \tag{13}$$

4 Proof of theorem 1

How do we prove the theorem about the complexity of **LiL UCB** (theorem 1)? We need to consider two things. One is to bound the sample complexity (in section 4.1). The other is to make sure we can return the best arm (in section 4.2).
First, let's look at the sample complexity.

4.1 Sample Complexity

We need to prove that with probability $1-\delta$, the algorithm terminates before the sum of $T_i(t)$. The challenge is that we cannot do all the union bounds. The lemma has already given us some uniform bounds. Therefore, we do not need to do all the union bounds over t.

For each arm i, let

$$\varepsilon_i(\omega) = \left\{\, \forall t \geq 1 :\ |\hat{\mu}_{i,t} - \mu_i| \leq \mathcal{M}(t,\omega) \right\}.$$

Then we apply the lemma of lil (eq. (5)): for every ω, we claim that

$$Pr\left[\varepsilon_i(\omega)\right] \geq 1-\omega.$$

However, we cannot assume that good events happen for every i. To assume that, we need to put an extra $\log(n)$. So first we need to condition on first event $\varepsilon_1(\omega)$ because we want the best arm not go over the uniform bound. This means that:

$$\forall t :\ \hat{\mu}_{i,t} + \mathcal{M}(t,\delta) \geq \mu_1.$$

Then we look at the favorite sub-optimal arm:

$$\text{Fix } i \in 2,3,...,n,\ \text{ let } q_i = \frac{64\ln(6\ln(1/\Delta_i^2+1)/\delta)}{\Delta_i^2}.$$

q_i is the expected number of plays we made to every sub-optimal arm. We will show that the probabilty we exceed the bound is very small for arm i. The order is $\Delta_i^{-2}(log\delta^{-1}+loglog\Delta_i^{-1})$, which corresponds to the number of samples in our goal. The numerator here has a similar form of that in $\mathcal{M}$ function.

Now let us look at T_i:

For each integer z=1,2,3..., we consider the probability:

$$Pr[T_i > zq_i]$$

We expect that T_i does not go over q_i, so we want the probability be very small.

Then we need to play arm i for $zq_i + 1$ times. Since the algorithm gives us the greatest UCB for arm i, we have

$$Pr\left[T_i > zq_i\right] \leq Pr\left[\hat{\mu}_{i,zq_i} + 2\mathcal{M}(zq_i, \delta) > \mu_1\right]. \tag{14}$$

Then we will do some calculation. We have already known the form of $\mathcal{M}(t, \delta)$, so we have

$$2\mathcal{M}(zq_i, \delta) = 4\sqrt{\frac{\ln(6\ln(zq_i+1))/\delta}{zq_i}} \leq 4\sqrt{\frac{2z\ln(6\ln(1/\Delta_i^2+1))/\delta}{zq_i}}$$

$$\leq 4\sqrt{\frac{2}{64}\cdot\Delta_i^2} = \frac{\Delta_i}{\sqrt{2}}.$$

(We replace q_i with $1/\Delta_i^2$ since we will not increase much and we put z out since $z \geq \log(\log(z))$.)

Then eq. (14) can be written as:

$$Pr\left[T_i > zq_i\right] \leq Pr\left[\hat{\mu}_{i,zq_i} - \mu_i > \frac{\Delta_i}{4}\right]. \tag{15}$$

By Hoeffding's Inequality,we can expand eq. (15) as:

$$\begin{aligned}
Pr[T_i > zq_i] \leq & Pr\left[\hat{\mu}_{i,zq_i} - \mu_i > \frac{\Delta_i}{4}\right] && (16)\\
\leq & \exp\left(-2zq_i(\frac{\Delta_i}{4})^2\right) && (17)\\
= & \exp\left(-\delta\cdot z\cdot ln(6ln(\frac{1}{\Delta_i^2}+1)/\delta)\right) && (18)\\
\leq & \exp\left(-8zln\frac{1}{\delta}\right) && (19)\\
= & \delta^{8z}. && (20)
\end{aligned}$$

Finally, we got

$$Pr[T_i > zq_i] \leq \delta^{8z}. \tag{21}$$

We can see that δ^{8z} increases supper fast. Therefore, we get a very strong upper bound on the number of plays for every arm.

4.2 Correctness

Now we want to make sure we can return the best arm. This is equivalent to make the following claim about correctness.

Claim

$$Pr\left[\sum_{i=2}^{n} T_i \leq 2\sum_{i=2}^{n} q_i\right] \geq 1 - 2\delta. \tag{22}$$

eq. (22) implies that with high probability, the total plays for all sub-optimal arms are no more than a double of the sum of q_i. Since the WHILE condition in section 3 also provides a upper bound of for the optimal arm, then with eq. (22), we can conclude the proof of theorem.

Proof. Let $\tilde{T_i} = T_i - q_i$.

$$\begin{aligned}
\mathbb{E}\left[\tilde{T_i}\right] &\leq \sum_{z=2}^{n} \delta^{8z} z q_i \quad \text{(by } eq. \text{ (21))} && (23)\\
&= \frac{\delta^8}{1-\delta^8} q_i && (24)\\
&\leq 2\delta^8 q_i && (25)
\end{aligned}$$

Remark 2. *Notice that eq. (23) is simplex extension of the following fact:*

For any constant q and a random variable T,

$$\begin{aligned}
\mathbb{E}[T] &\leq q + Pr[T \geq q] + qPr[T \geq 2q] + \cdots && (26)\\
&\leq \sum_{i=1}^{\infty} Pr[T \geq (i-1)q] && (27)
\end{aligned}$$

Therefore,

$$\begin{aligned}
&Pr\left[\sum_{i=2}^{n} T_i \leq 2\sum_{i=2}^{n} q_i\right] && (28)\\
=&Pr\left[\sum_{i=2}^{n} \tilde{T_i} \leq \sum_{i=2}^{n} q_i\right] \quad \text{(subtract } q_i \text{ on both sides)} && (29)\\
\leq&2\delta^8 \quad \text{(by Markov's inequality and eq. (25))} && (30)\\
\leq&\delta \quad (\delta \text{ small enough}). && (31)
\end{aligned}$$

□

Next, we will present a lemma for the correctness of the algorithm.

Lemma 2. *For $i = 2, 3, \cdots,$ define*

$$H_i = \left\{ T_i < 1 + 20 \sum_{j=1}^{i-1} T_j(t), \quad \forall t \right\}. \tag{32}$$

Then

$$Pr\left[\wedge_{i=2}^{n} H_i\right] \geq 1 - \mathcal{O}(\sqrt{\delta}). \tag{33}$$

Notice that the condition in the definition of H_i is a stronger condition than the WHILE condition in section 3. The rest of this subsection is to prove the above lemma.

To prove the lemma, for fixed $i \in \{2, 3, \cdots, n\}$, we define an event

$$G_i = \left\{ \text{at least } \frac{i-1}{2} \text{ of } \varepsilon_1(\delta), \varepsilon_1(\delta), \cdots, \varepsilon_{i-1}(\delta) \text{ happen} \right\}. \tag{34}$$

Remark: Although there are $(i-1)$'s arms better than arm i, we only want $\frac{i-1}{2}$ events rather than all of events happen. The reason is we cannot afford all of the events to happen, because using $(i-1)$ events requires to use the union bound, and this will not lead to our desired complexity.

Next, we want to bound the complement of G_i. Since each ε_i are independent, we apply Chernoff bound (lemma 4) on the the following inequality by setting $\mu = (i-1)\delta, \gamma = \frac{i-1}{2\mu} - 1 = \frac{1}{2\delta} - 1$:

$$\begin{aligned} Pr\left[\ \overline{G_i}\ \right] &= Pr\left[X_1 + X_2 + \cdots + X_{i-1} \geq \frac{i-1}{2}\right] && (35) \\ &\leq \left(\frac{e^{\frac{1}{2\delta} - 1}}{\frac{1}{2\delta}^{\frac{1}{2\delta}}} \right)^{(i-1)\delta} && (36) \\ &< (2e\delta)^{\frac{i-1}{2}}. && (37) \end{aligned}$$

We observe that when $i = 2$, we get roughly $\sqrt{\delta}$, and when i increases, the probability of $\overline{G_i}$ also decreases exponentially.

Now, we will following fact which combines the introduced events together. We will show the proof of the fact at the end of this section.

Fact.

$$G_i \wedge \varepsilon_i(\delta^{i-1}) \Longrightarrow H_i. \tag{38}$$

Notice that δ^{i-1} is much smaller than δ, this means ε_i only fails with much smaller probability.

With this fact, we conclude our proof by the following analysis:

$$Pr\left[\ \overline{H_i}\ \right] \leq Pr\left[\ \overline{G_i}\ \right] + Pr\left[\ \overline{\varepsilon_i(\delta^{i-1})}\ \right] \tag{39}$$

$$\leq (2e\delta)^{i-1} + \delta^{i-1}, \tag{40}$$

which implies

$$Pr\left[\ \wedge_{i=2}^{n} H_i\ \right] \geq 1 - \sum_{i=2}^{n}\left((2e\delta)^{\frac{i-1}{2}} + \delta^{i-1}\right) \tag{41}$$

$$\geq 1 - \mathcal{O}\left(\sqrt{\delta}\right). \tag{42}$$

Conclusion: the key idea of this proof is not to use the union bound directly, but to introduce the events which happens with smaller and smaller probability, then we can apply union bound on these events and get the desired bound.

Finally, we will provide the proof of the fact.

Proof. Suppose at time t, we have $T_i(t) = 80\sum_j^{i-1} T_j(t)$.

We prove the fact by contrapositive. There is equivalent to show there exist arm j less than i, such that

$$\hat{\mu}_{j,T_j(t)} + 2\mathcal{M}(T_j(t), \delta) > \hat{\mu}_{i,T_i(t)} + 2\mathcal{M}(T_i(t), \delta) \tag{43}$$

For simplicity, we introduce a number S such that

$$\frac{i-1}{2}s = \sum_{j=1}^{i-1} T_j(t) \quad \text{or} \quad T_i(t) = 40(i-1)s. \tag{44}$$

Let $Q = \{j < i : \varepsilon_j(\delta)\}$. By G_i (eq. (34)), we know

$$|Q| \geq \frac{i-1}{2}. \tag{45}$$

Then by definition of Q and S, we have

$$\sum_{j\in Q} T_j(t) \geq \sum_{j=i}^{j-1} T_j(t) = \frac{(i-1)}{s}S \tag{46}$$

Combining eq. (45) and eq. (46), we conclude that there exists a j^* such that

$$T_j^* \leq S.$$

In other words, there exist $j = j^*$ such that

$$\varepsilon_{j^*}(\delta), \quad T_{j^*}(t) \leq s.$$

With the above observation, we are ready to prove the desired bound (eq. (43)) by relaxing the inequalities on both sides.

We first relax the LHS of (eq. (43)) by setting $j = j^*$, and obtain

$$\begin{aligned} LHS \geq & (\mu_j - \mathcal{M}(T_j(t), \delta)) + 2\mathcal{M}(T_j(t), \delta) && (47) \\ \geq & \mu_j + \mathcal{M}(s, \delta), && (48) \end{aligned}$$

where the last line holds because the confidence bound is a decreasing function, i.e. the more samples we have, the bound should be smaller.

We then relax the RHS of (eq. (43)), and obtain

$$\begin{aligned} RHS \leq & \mu_i + \mathcal{M}(T_i(t), \delta^{i-1}) + 2\mathcal{M}(T_i(t), \delta) && (49) \\ \leq & \mu_i + 3\mathcal{M}\left(40(i-1)s, \delta^{i-1}\right), && (50) \end{aligned}$$

where the lat line holds because (eq. (44)), and $\mathcal{M}$ decreases when δ increases.

The final step is to show

$$\mu_j + \mathcal{M}(s, \delta) > \mu_i + 3\mathcal{M}\left(40(i-1)s, \delta^{i-1}\right).$$

Since $j < i$, then $\mu_j > \mu_i$. The rest is to show

$$\mathcal{M}(s, \delta)$$

is greater than $3\mathcal{M}\left(40(i-1)s, \delta^{i-1}\right)$. The proof can be done by applying the fact that the confidence bound is a decreasing function and plugging in the parameters. □

5 Appendix

Lemma 3. *Hoeffding's maximal inequality [Hoeffding'1963]).*
Let $X_1, X_2, X_3, ..., X_n$ be random variables such that $\mathbb{E}\, X_i = 0$, and $X_i \in [a_i, b_i]$ for every i=1,2,...,n. Then for each $t \geq 0$,

$$Pr\left[\forall i = 1, 2, ..., n, \ \sum_{j=1}^{i} X_i \leq t\right] \geq 1 - exp\left(\frac{-2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right).$$

Lemma 4. *(Chernoff Bound) Let* $X_1, X_2, \cdots, X_n$ *be independent Bernoulli random variables with* $\mathbb{E}[X_i] = p_i$ *for all* i*. Let* $X = \sum_i^n X_i, \mu = \mathbb{E}[X]$*. Then for any* $\gamma > 0$*,*

$$Pr\left[X \geq (1+\gamma)\mu\right] \leq \left[\frac{e^{\gamma}}{(1+\gamma)^{1+\gamma}})\right] \tag{51}$$

$$Pr\left[X \leq (1+\gamma)\mu\right] \leq \left[\frac{e^{-\gamma}}{(1-\gamma)^{1-\gamma}})\right]. \tag{52}$$

References

[1] Kevin Jamieson, et al. *lil'ucb: An optimal exploration algorithm for multi-armed bandits.*. Conference on Learning Theory.2014.