IE 498: Online Learning and Decision Making, Fall 2019 09/10/2019

Lecture 05: Expectation of Sample Complexity and Improved SR

Lecturer: Dr. Yuan Zhou *Scribe: Jiaxin Wu, Bochao Li*

1 From Sample Complexity to Expectation

During last lecture, we have talked about Successive Rejection (SR) algorithm, which is an algorithm that can 1. returns the best arm and 2. have sample complexity $\lesssim \sum_{i=2}^{n} \Delta_i^{-2}(\log \frac{n}{\delta} + \log\log \Delta_i^{-1})$ with probability $1-\delta$. In this lecture, we introduce the theorem that given an algorithm with a high probability sample complexity bound, one can design a new algorithm $\mathcal{B}$ with the similar complexity bound, along with its proof.

Theorem 1. *Suppose an algorithm $\mathcal{A}$, with a probability $\geq 1-\delta$, ensures that:*

- *$\mathcal{A}$ returns the best arm.*
- *sample complexity $\leq f \log 1/\delta + g$, $\forall \delta > 0$ where f, g depends on δ.*

Then there exists an algorithm $\mathcal{B}$:

- *returns the best arm with probability $\geq 1-\delta$*
- *$\mathbb{E}$(sample complexity of $\mathcal{B}$) $\lesssim f \log 1/\delta + g$, for $f :=$ the coefficient of $\log \Delta$ and $g :=$ the remaining part of the sample complexity of $\mathcal{A}$ (in the case for SR, $f = \sum_{i=2}^{n} \Delta_i^{-2}$ and $g = \sum_{i=2}^{n} \Delta_i^{-2}(\log n + \log\log \Delta_i^{-1})$).*

The symbol $x \lesssim y$ means $x \leq Cy$ for some constant c

2 Proof for Theorem 1

Proof. Consider a simple new algorithm $\mathcal{B}$ as following:

Algorithm 1: Ancillary algorithm for proof of Theorem 1

```
Result: the best arm a
for r ← 1, 2, 3, 4, ... do
    for i ← 1, 2, ..., log_2 r do
        if 2^{i-1} | r then
            run A_i for A with δ/2^i until:
            if A_i needs a sample then
                sample the arm and feed the observation to A_i ;
            else if A_i outputs arm a then
                output arm a ;
        end
    end
end
```

Then following $\mathcal{B}$, $\mathcal{A}_1$ is ran for each round, $\mathcal{A}_2$ is ran for each 2 rounds, and $\mathcal{A}_3$ is ran for each 4 rounds ....

First, let us show the correctness of algorithm 1. Define event $E_i \coloneqq \{\mathcal{A}_i$ return the best arm with $f \log \frac{2^i}{\delta} + g$ samples$\}$ and $E \coloneqq E_1 \wedge E_2 \wedge E_3 \wedge \ldots$. Then based on algorithm 1, we can have:

$$Pr[E] \geq 1 - \frac{\delta}{2} - \frac{\delta}{4} - \frac{\delta}{8} - \cdots = 1 - \delta \tag{1}$$

and this implies that if this event hold then the new constructed $\mathcal{B}$ following algorithm 1 will not be wrong with a probability of at least $1-\delta$. Next let's consider about one flow chart of the occurrence of E_i as shown in figure 1.

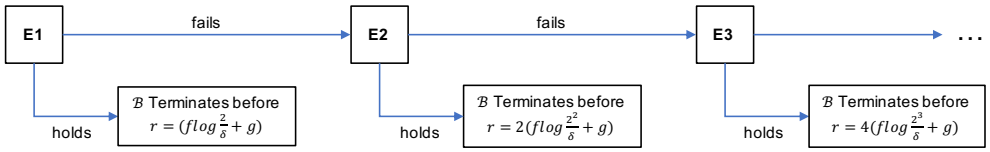

Figure 1: Decision tree of E_i and their number of runs.

This diagram provides us with a intuition that the expected rounds of running algorithm 1 can be found by calculating the weighted sum of the number of rounds r as shown in the figure. Thus, now let's define the situation that all E_i fails as $F_i \coloneqq \overline{E_1} \wedge \overline{E_2} \wedge \overline{E_3} \wedge \cdots \wedge \overline{E_{i-1}} \wedge E_i$ and denote the number of rounds used by $\mathcal{B}$ as r^*. Then we can derive the expectation of r^* as:

$$\mathbb{E}\, r^* = \sum_{i=1}^{+\infty} \mathbb{E}\,[r^*|F_i] Pr[F_i] \leq \sum_{i=1}^{+\infty} 2^{i-1} f \log\left(\frac{2^i}{\delta} + g\right) \prod_{j=1}^{i-1} \frac{\delta}{2^j} \tag{2}$$

where $\sum_{i=1}^{+\infty} 2^{i-1} f \log\left(\frac{2^i}{\delta} + g\right)$ is the upper bound of $\mathbb{E}\,[r^*|F_i]$ and $Pr[F_i]$ is bounded by $\prod_{j=1}^{i-1} \frac{\delta}{2^j}$. The right hand side of the above inequality can be reorganized and expanded to be:

$$(f \log \frac{1}{\delta} + g) \sum_{i=1}^{+\infty} 2^{i-1} \prod_{j=1}^{i-1} \frac{\delta}{2^j} + f \sum_{i=1}^{+\infty} 2^{i-1} \log 2^i \prod_{j=1}^{i-1} \frac{\delta}{2^j} \tag{3}$$

We can see that for both terms in equation 3, the magnitude of the constants are dominated by the non-constant composites and equation 3 can be simplified as:

$$f \log \frac{1}{\delta} + g + f \tag{4}$$

After plug equation 4 into equation 2, the expectation of number of runs for $\mathcal{B}$ is upper bounded by:

$$\mathbb{E}\, r^* \lesssim f \log \frac{1}{\delta} + g + f \tag{5}$$

which is consistent with the claim in theorem 1. □

3 Summary of Complexity

Table 1 summarizes the complexity of the algorithms we have covered so far, as shown in below:

	Minimax	Parameter Dependent
Best Arm Identification	Uniform sampling: $O(\frac{n}{\epsilon^2}\log(\frac{1}{\delta}))$ lower bound: $\Omega(\frac{n}{\epsilon^2}\log\frac{1}{\delta})$	Successive Rejection: $O(\sum_{i=2}^{n}\Delta_i^{-2}(\log\frac{n}{\delta}+\log\log\frac{1}{\Delta_i}))$ lower bound: $\Omega(\sum_{i=2}^{n}\Delta_i^{-2}\log\frac{1}{\delta})$
Regret Minimization	UCB: $O(\sqrt{nT\log T})$ lower bound: $\Omega(\sqrt{nT})$	UCB: $O(\sum_{i=2}^{n}\Delta_i^{-1}\log T)$ lower bound: $\Omega(\sum_{i=2}^{n}\Delta_i^{-1}\log T)$

Table 1: Summary for the complexities of the algorithms covered in lecture so far

4 Improved SR Method

As we can see in the table, the minimax regret bound we get using algorithm taught in previous lecture is not optimal. And in this part we gives an new successive rejection type algorithm that gives us a sub-optimal minimax regret bound and an optimal parameter dependent bound.

Algorithm 2: Improved Successive Rejection Algorithm(ISR)

Result:
initially set $S_0 \leftarrow [n], t \leftarrow 0$
while $t \le T$ **do**
  $t \leftarrow t+1$;
  $\epsilon^t \leftarrow 2^{-t}$;
  play each arm in S_{t-1} for $\frac{\log(2T\epsilon_t^2)}{2\epsilon_t^2}$ times ;
  set $S_t \leftarrow \{i : i \in S_{t-1}, \hat{\mu}_i^t \ge \max_{j\in S_{t-1}}\hat{\mu}_j^t - \epsilon_t\}$;
end

Theorem 2. *When running ISR algorithm, $\forall\lambda \ge \sqrt{\frac{4}{T}}$, we have*

$$\mathbb{E}\, R_T \lesssim \sum_{i=2}^{n} \min\{\Delta_i^{-1}\log(T\Delta_i^2), \lambda^{-1}\log(T\lambda^2)\} \tag{6}$$

(1) when setting $\lambda = \sqrt{\frac{n \log n}{T}}$*, we have*

$$\mathbb{E}\, R_T \lesssim n\lambda^{-1} \log(T\lambda^2) + \lambda T + \frac{1}{\lambda} = O(\sqrt{nT \log n}) \tag{7}$$

An sub-optimal minimax regret bound.
(2)When setting $\lambda = \sqrt{4}T$*, we have*

$$\mathbb{E}[R_T] \lesssim \sum_{i=2}^{n} \Delta_i^{-1} \log T \tag{8}$$

The optimal parameter dependent bound.
T is the number of rounds. If we sort the arms in descendent order (i.e. $\mu_1 > \mu_2 \geq \cdots \geq \mu_n$, while μ_i is the average reward for choosing arm i). We define $\Delta_i = \mu_1 - \mu_i$ the difference of arm i compared with the optimal arm, like in previous lecture.

Proof. Set event

$$E_t = \{1 \in S_t, \forall j \neq 1, \mu_j < \mu_1 - \epsilon_t \text{ implies } j \notin S_t\} = \{i \in S_t, \forall j \neq 1, j \in S_t \text{ implies } \mu_j > \mu_1 - \epsilon_t\}$$

This Event E_t says that in t round, all the sub-optimal arms that are ϵ_t worse than the best arm will be rejected. And if we can prove every E_t happens with large probability in a round when the error reach our desired accuracy, we can say our algorithm successfully output the best arm with high probability. It is easy to see $Pr[E_0] = 1$
We claim that

$$\forall t > 0, Pr[E_t | E_{t-1}] \geq Pr[\{\forall i \in S_{t-1} : |\mu_i - \hat{\mu}_i^t| \leq \epsilon_t/2\} | E_{t-1}]$$

The trick we use here is similar to the one used when proving regret bound for Successive reject algorithm in previous lecture. For convenience, we name event

$$E_t' = \{\forall i \in S_{t-1} : |\mu_i - \hat{\mu}_i^t| \leq \epsilon_t/2\}$$

where $\hat{\mu}_i^t$ is the estimated value of μ_i using samples collected from round t. When E_t' happens, we have $\forall i \in S_{t-1}, |\mu_i - \hat{\mu}_i| \leq \epsilon_t/2$. When E_t' happens and using the elimination rule claimed in the algorithm, we have $\forall i \in S_t, \mu_i > \hat{\mu}_i - \epsilon_t/2 > \max_{j \in S_{t-1}} \hat{\mu}_j^t - 3\epsilon_t/2 > \mu_1 - 2\epsilon_t$. We can further bound the probability as

$$Pr[E_t' | E_{t-1}] \geq \mathbb{E}[1 - \frac{|S_{r-1}|}{T\epsilon_t^2} | E_{t-1}] \geq 1 - \frac{n_{t-1}}{T\epsilon_t^2}$$

While n_t =number of arms with $\{i : \mu_1 - \mu_i < 2\epsilon_t\}$. The first inequality is because to make E_t' happens, using concentration inequality like Hoeffding inequality,if we sample each arm i in S_{t-1} $\frac{\log(2T\epsilon_t^2)}{2\epsilon_t^2}$ times, we have $Pr[|\hat{\mu}_i^t - \mu_i| \leq \epsilon_t/2] \geq 1 - \frac{1}{T\epsilon_t^2}$. We have to bound $|S_{t-1}|$ arms and they are independent, so we have $Pr[E_t' | E_{t-1}] \geq \mathbb{E}[1 - \frac{|S_{r-1}|}{T\epsilon_t^2} | E_{t-1}]$. The expectation is because $|S_{t_1}|$ itself is a random variable. And the second inequality is because the expectation of $|S_{t-1}|$ is n_{t-1}, and it is easy to see that the condition expectation on E_{t-1}, $\mathbb{E}[1 - \frac{|S_{r-1}|}{T\epsilon_t^2} | E_{t-1}]$ is greater than condition expectation on $\overline{E}_{t-1}$, $\mathbb{E}[1 - \frac{|S_{r-1}|}{T\epsilon_t^2} | \overline{E}_{t-1}]$.
Now lets define the event $F_t = E_0 \wedge E_1 \cdots E_t$. Which is the event that we do successful reject in all first t rounds. Then $\forall t^* > 0$, we have

$$\mathbb{E}[R_T] = \sum_{t=1}^{t^*-1} \mathbb{E}[R_T | F_t \wedge \overline{E}_{t+1}] Pr[F_t \wedge \overline{E}_{t+1}] + \mathbb{E}[R_T | F_{t^*}] Pr\, \mathbb{E}[F_{t^*}]$$

For simplicity we name $\sum_{t=1}^{t^*-1} \mathbb{E}[R_T | F_t \wedge \overline{E}_{t+1}] Pr[F_t \wedge \overline{E}_{t+1}]$ as (1) and $\sum_{t=1}^{t^*-1} \mathbb{E}[R_T | F_{t^*}] Pr\, \mathbb{E}[F_{t^*}]$ as (2).
For (1), for each t, first we notice that $Pr[F_t \wedge \overline{E}_{t+1} \leq Pr[E_t \wedge \overline{E}_{t+1}] \leq Pr[\overline{E}_{t+1} | E_t]$. The first inequality

is because for two event A, B, we always have $Pr[A \wedge B] \leq Pr[A]$. The second inequality is because the definition of conditional distribution: $Pr[A|B] = \frac{Pr[A \wedge B]}{Pr[B]}$. And then we can further upper bound this probability by the result we get before: $Pr[\overline{E}_{t+1}|E_t] \leq \frac{n_t}{T\epsilon_{t+1}^2}$

And we can also bounded the conditional expectation of regret R_T. Because we know that if F_t holds, the mean reward μ_i for any arm i lefts in our decision set S_t is ϵ_t close to the optimal arm 1, so our regret is less or equal to $T\epsilon_t$.

So we can bound (1) as

$$(1) \leq \sum_{t=1}^{t^*-1} T\epsilon_t \frac{n_t}{T\epsilon_{t+1}^2} \lesssim \sum_{t=1}^{t^*-1} \frac{n_t}{\epsilon_{t+1}} \lesssim \sum_{i=2}^{n} \sum_{t=1}^{\min\{t^*-1, \lceil \log_2 \Delta_i^{-1} \rceil\}} \frac{1}{\epsilon_t} \lesssim \sum_{i=2}^{n} \min\{\epsilon_{t^*}^{-1}, \Delta_i^{-1}\}$$

The first two inequality are easy to check. For the third inequality, we notice that because n_t is defined as number of ϵ_t sub-optimal arms. Using the ISR algorithm, if t^* is large enough, arm i will be eliminated after $\log_2 \Delta_i^{-1}$ rounds, if t^* is not large enough, then it will be counted in n_t for all $t = 1 \cdots r^* - 1$. And for the forth inequality, we notice that sequence $\frac{1}{\epsilon_t}$ is a geometric sequence with ratio 2, and $\sum_{i=1}^{n} 2^i \leq 2^{n+1}$.

So now we have gives an upper bound for (1), for (2), we notice that

$$(2) \leq \mathbb{E}[R_T|F_{t^*}] \lesssim \sum_{i=2}^{n} \sum_{t=1}^{\min\{t^*-1, \lceil \log_2 \Delta_i^{-1} \rceil\}} \Delta_i \log(T\epsilon_t^2)/\epsilon_t^2 + \mathbb{I}_{\{\epsilon^{t^*} > \Delta_2\}} T\epsilon_{t^*}$$

The $\mathbb{I}_{\{\}}$ is the identity function and get 1 when the condition in $\{\}$ is true, and get 0 if it is false.

The second inequality is because each time we sample arm i, the expectation of regret for this sample is Δ_i. Like discussed before, if t^* is large enough, arm i will be eliminated after $\log_2 \Delta_i^{-1}$ rounds, and will not be sampled anymore. if t^* is not large enough, then it will be counted in n_t for all $t = 1 \cdots r^*$, and will be sampled in each round. So the total number of samples for arm i is $\sum_{t=1}^{\min\{t^*-1, \lceil \log_2 \Delta_i^{-1} \rceil\}} \log(T\epsilon_t^2)/\epsilon_t^2$, and those the regret caused by sampling arm i is $\sum_{t=1}^{\min\{t^*-1, \lceil \log_2 \Delta_i^{-1} \rceil\}} \Delta_i \log(T\epsilon_t^2)/\epsilon_t^2$. And the $\mathbb{I}_{\{\epsilon^{t^*} > \Delta_2\}} T\epsilon_{t^*}$ term is because after t^* rounds, because we know F_{t^*} happens, if ϵ_{t^*} is small compared with Δ_2 and no sub-optimal arms left in set S_{t^*}, we will only choose the best arm 1 in the following $T - t^*$ rounds. If ϵ_{t^*} is not so small and there still are some sub-optimal arms left in set S_{t^*}, we know regret for the following $T - t^*$ rounds will not exceed $T\epsilon_{t^*}$

And also notice that

$$\begin{aligned}
&\sum_{i=2}^{n} \sum_{t=1}^{\min\{t^*-1, \lceil \log_2 \Delta_i^{-1} \rceil\}} \Delta_i \log(T\epsilon_t^2)/\epsilon_t^2 \\
&\lesssim \sum_{i=2}^{n} \min\{\epsilon_{t^*}^{-2} \log(T\epsilon_{t^*}^2), \Delta_i^{-2} \log(T\Delta_i^2)\}\Delta_i \\
&\leq \sum_{i=2}^{n} \min\{\epsilon_{t^*}^{-1} \log(T\epsilon_{t^*}^2), \Delta_i^{-1} \log(T\Delta_i^2)\}
\end{aligned}$$

The first inequality is because when $T > 4/\epsilon_{t'}^2$, we have $\sum_{t=1}^{t'-1} \log(T\epsilon_t^2)/\epsilon_t^2 \leq \log(T\epsilon_{t'}^2)/\epsilon_{t'}^2$. And the second inequality is because when $\Delta_i < \epsilon_{t^*}$, the second term will be smaller than the first term, and we simply eliminate an Δ_i. and when $\Delta_i > \epsilon_{t^*}$, the first term will be smaller and we can replace an $\epsilon_{t^*}^{-1}$ using Δ_i. So in total we have $\mathbb{E}[R_T|F_{t^*}]Pr\,\mathbb{E}[F_{t^*}] \lesssim \sum_{i=2}^{n} \min\{\epsilon_{t^*}^{-1} \log(T\epsilon_{t^*}^2), \Delta_i^{-1} \log(T\Delta_i^2)\} + \mathbb{I}_{\{\epsilon^{t^*} > \Delta_2\}} T\epsilon_{t^*}$

Now we have regret bound for the two parts (1) and (2), it is easy to see that the second one is the dominate one, so we have $\mathbb{E}[R_t] \lesssim \sum_{i=2}^{n} \min\{\epsilon_{t^*}^{-1}\log(T\epsilon_{t^*}^2), \Delta_i^{-1}\log(T\Delta_i^2)\} + \mathbb{I}_{\{\epsilon^{t^*}>\Delta_2\}} T\epsilon_{t^*}$

$\forall \lambda \geq \sqrt{\frac{4}{T}}$, if we define $r^* = \lceil \log_2(1/\lambda) \rceil$, we get $\mathbb{E}[R_t] \lesssim \sum_{i=2}^{n} \min\{\lambda^{-1}\log(T\lambda^2), \Delta_i^{-1}\log(T\Delta_i^2)\} + \mathbb{I}_{\{\lambda>\Delta_2\}}\lambda T$.

(1)If we set $\lambda = \sqrt{(n\log n)/T}$, we have

$$\mathbb{E}[R_T] \lesssim \sum_{i=2}^{n} \sqrt{T/(n\log n)}\log(Tn\log n/T) + T\sqrt{(n\log n)/T} \lesssim n\sqrt{T/(n\log n)}\log(n\log n) + \sqrt{Tn\log n} \lesssim \sqrt{nT\log n}$$

an sub-optimal minmax lower bound.

(2) And If we set $\lambda = \sqrt{\frac{4}{T}}$, when $\Delta_2 < \lambda$ (i.e $\Delta_2^{-1} > \sqrt{\frac{T}{4}}$), we have $\mathbb{I}_{\{\epsilon^{t^*}>\Delta_2\}} T\epsilon_{t^*} T\epsilon_{t^*} = \sqrt{4T} \lesssim \Delta_2^{-1} \lesssim \sum_{i=2}^{n} \Delta_i^{-1}$ we get $\mathbb{E}[R_T] \lesssim \sum_{i=2}^{n} \Delta_i^{-1}\log T + \sum_{i=2}^{n} \Delta_i^{-1} \lesssim \sum_{i=2}^{n} \Delta_i^{-1}\log T$. When $\min_i \Delta_i > \lambda$ we simply get $\mathbb{E}[R_T] \lesssim \sum_{i=2}^{n} \Delta_i^{-1}\log T$. So in both condition we get an optimal parameter dependent bound. □

The high level idea of this algorithm is that we do SR algorithm at first t^* rounds, and that gives an small sets of arm S_{t^*}, and all arms in this set is sub-optimal and thus will not cause large regret in the following rounds. And because we only do successive rejection in the first few rounds, we can make sure the regret obtained in the sampling round not too large.