CS 6840 Algorithmic Game Theory October 16 and 18, 2024

Lecture 21|22: Follow the Leader, A Method of Learning

Instructor: Eva Tardos *Scribe: Haripriya Pulyassary, Hannane Yaghoubizade*

In this section, we return to the question of designing low-regret learning algorithms, and we introduce a new method that is usually more flexible than Multiplicative Weight Update. Later we will see some examples where this method becomes more useful than MWU.

1 Introduction

The main idea is to play the historically best strategy at each time. That is, player i at time t selects $s_i^t = \arg\max_s \sum_{\tau=1}^{t-1} u_i^\tau(s, \mathbf{s}_{-i}^\tau)$. In 1956, Julia Robinson showed that for 2-person 0-sum games, the above sequence of plays converges to Nash equilibrium in marginal distribution. However, this idea does not seem to work well in general games.

The Setting For today's analysis, we consider the following setting: We have two players, who we refer to as the "learner" and the "adversary". At each iteration t of the game, the learner chooses a strategy $s^t \in S$ and the adversary selects a utility function $u^t : S \to [0,1]$. Then the utility the learner obtains will be $u^t(s^t)$.

Follow the Leader Based on the idea discussed above, one approach the learner can take is to choose

$$s^t = \arg\max_s \sum_{\tau=1}^{t-1} u^\tau(s)$$

But as expected, this approach does not work well. For example let $S = \{s_1, s_2\}$ and suppose the adversary chooses:

$$u^t(s_1) = \begin{cases} 0 & t \overset{2}{\equiv} 0 \\ 1 & t \overset{2}{\equiv} 1 \end{cases} = (1, 0, 1, 0, \dots) \qquad u^t(s_2) = \begin{cases} \frac{1}{2} + \epsilon & t = 1 \\ \frac{1}{2} & \text{otherwise} \end{cases} = (\frac{1}{2} + \epsilon, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \dots)$$

Using these utility functions for $t \geq 2$ we have

$$\sum_{\tau=1}^{t-1} u^\tau(s_1) = \begin{cases} \frac{t}{2} & t \overset{2}{\equiv} 0 \\ \frac{t-1}{2} & t \overset{2}{\equiv} 1 \end{cases} \qquad \sum_{\tau=1}^{t-1} u^\tau(s_2) = \frac{t-1}{2} + \epsilon$$

Therefore, the learner chooses $s^2 = s_1$, $s^3 = s_2$, $s^4 = s_1$ and will keep alternating between s_1 and s_2. However, he will be choosing s_1 when its utility is 0, and s_2 when s_1's utility is 1. This way, at time T, the total utility is $\approx \frac{T}{4}$, whereas fixing either strategy would give a total utility of about $\approx \frac{T}{2}$ at time T. One can easily come up with similar examples for any deterministic learning algorithm. This gives the motivation to use a "randomized" version of Follow the Leader.

Follow the Perturbed (Noisy) Leader As an attempt to randomize our algorithm, we ask the learner to add a noise while making decisions. To be specific, for a strategy s, let $F_s \geq 0$ be some distribution and $\xi_s^t \sim F_s$ be a non-negative random noise. Distributions F_s are not necessarily identical and random variables $\{\xi_s^t\}_{s\in S, t\in[T]}$ are independent. Then let

$$s^t = \arg\max_s \left(\sum_{\tau=1}^{t-1} u^\tau(s)\right) + \xi_s^t \tag{1}$$

Next, we will show a low-regret guarantee for this algorithm for a suitable choice of F_s.

2 No-regret Analysis

Be the Leader, A Cheat Algorithm Suppose, for a moment, that at time t the learner cheats and chooses

$$s^t = \arg\max_s \sum_{\tau=1}^{t} u^\tau(s). \tag{2}$$

This is indeed unrealistic since $u^t(s)$ is not known before the t-th iteration of the game, but we show that this algorithm has 0 regret and helps us in our analysis.

Claim 1. *Let $s^1, s^2, \dots$ be the sequence produced by the cheat algorithm* (2)*. This sequence is truly no-regret, that is for every t and every $\tilde{s} \in S$*

$$\sum_{\tau=1}^{t} u^\tau(s^\tau) \geq \sum_{\tau=1}^{t} u^\tau(\tilde{s})$$

Proof. The proof proceeds by induction. In the base case where $t = 1$, the claim follows immediately by the choice of s^t. For $t > 1$,

$$\begin{aligned}
\sum_{\tau=1}^{t} u^\tau(s^\tau) &= \sum_{\tau=1}^{t-1} u^\tau(s^\tau) + u^t(s^t) \\
&\geq \sum_{\tau=1}^{t-1} u^\tau(s^t) + u^t(s^t) && \text{(by induction hypothesis for } \tilde{s} = s^t) \\
&= \sum_{\tau=1}^{t} u^\tau(s^t) \\
&\geq \sum_{\tau=1}^{t} u^\tau(\tilde{s}) && \text{(by choice of } s^t)
\end{aligned}$$

■

To get closer to our real algorithm (1), the next step is adding noise to the cheat algorithm. We would then like to argue that doing so does not increase the regret too much.

Be the Perturberd Leader (Noisy Cheat Algorithm) In this version, at time $t = 1$, the learner independently generates a non-negative random noise ξ_s, for each strategy $s \in S$. Then at each time t he chooses

$$s^t = \arg\max_s \left(\sum_{\tau=1}^{t} u^\tau(s)\right) + \xi_s \tag{3}$$

Claim 2. *Let $s^1, s^2, \dots,$ be the sequence produced by noisy cheat algorithm (3), then for every realization of $\{\xi_s\}_{s\in S}$ like $\{\bar{\xi}_s\}_{s\in S}$, every t and every $\tilde{s} \in S$ we have*

$$\sum_{\tau=1}^{t} u^\tau(s^\tau) + \max_s \bar{\xi}_s \geq \left(\sum_{\tau=1}^{t} u^\tau(\tilde{s})\right) + \bar{\xi}_{\tilde{s}} \tag{4}$$

Proof. Consider $\bar{\xi}_s$ as $u^0(s)$, and consider running "Be the Leader"(2) from $t = 0$ instead of $t = 1$. We have

$$\begin{aligned}
\sum_{\tau=1}^{t} u^\tau(s^\tau) + \max_s \bar{\xi}_s &= \sum_{\tau=1}^{t} u^\tau(s^\tau) + u^0(s^0) && \text{(by choice of algorithm at time } t = 0) \\
&= \sum_{\tau=0}^{t} u^\tau(s^\tau) \\
&\geq \sum_{\tau=0}^{t} u^\tau(\tilde{s}) && \text{(by claim 1 for } \tau \text{ starting from 0)} \\
&= \left(\sum_{\tau=1}^{t} u^\tau(\tilde{s})\right) + \bar{\xi}_{\tilde{s}} && \text{(as } u^0(s) = \bar{\xi}_s)
\end{aligned}$$

■

Corollary 3. *Let $s^1, s^2, \dots,$ be the sequence produced by noisy cheat algorithm (3). Then for every t and $\tilde{s} \in S$ we have*

$$\mathbb{E}\left(\sum_{\tau=1}^{t} u^\tau(s^\tau)\right) + \mathbb{E}\left(\max_s \xi_s\right) \geq \left(\sum_{\tau=1}^{t} u^\tau(\tilde{s})\right) + \mathbb{E}(\xi_{\tilde{s}})$$

The above inequality further implies the regret of noisy cheat algorithm is bounded by $\mathbb{E}\left(\max_s \xi_s\right)$.

Proof. by (4) and taking expectation over $\{\xi_s\}_{s\in S}$. ■

Let us consider a better noisy cheat algorithm now. At each time t, the learner will generate a new set of independent[1] noises $\{\xi_s^t\}_{s\in S}$ and plays

$$\bar{s}^t = \arg\max_s \left(\sum_{\tau=1}^{t} u^\tau(s)\right) + \xi_s^t \tag{5}$$

Lemma 4. *Let $\bar{s}^1, \bar{s}^2, \dots,$ be the sequence produced by above algorithm (5), then for every t and every $\tilde{s} \in S$*

$$\mathbb{E}\left(\sum_{\tau=1}^{t} u^\tau(\bar{s}^\tau)\right) + \mathbb{E}\left(\max_s \xi_s^t\right) \geq \left(\sum_{\tau=1}^{t} u^\tau(\tilde{s})\right) + \mathbb{E}\left(\xi_{\tilde{s}}^t\right)$$

where the expectation is taken over values of $\{\xi_s^\tau\}_{s\in S,\tau\in[t]}$.

Proof. We provide two alternative proofs for the lemma.

[1] $\{\xi_s^t\}_{s\in S,t\in[T]}$ are independant.

First Proof For the sake of simplicity, we assume ξ_s^t are discrete random variables. The proof can be easily extended to continuous random variables. Let $s^t_{\{\bar{\xi}_s\}_{s\in S}}$ be result of (3) where the set of generated noises is $\{\bar{\xi}_s\}_{s\in S}$. Then we have:

$$
\begin{aligned}
&\mathbb{E}\left(\sum_{\tau=1}^{t} u^\tau(\bar{s}^\tau)\right) + \mathbb{E}\left(\max_s \xi_s^t\right) \\
=&\sum_{\tau=1}^{t} \mathbb{E}_{\{\xi_s^\tau\}_{s\in S}}\left(u^\tau(\bar{s}^\tau)\right) + \mathbb{E}\left(\max_s \xi_s^t\right) && \text{(by linearity of expectation)} \\
=&\sum_{\tau=1}^{t} \sum_{\{\bar{\xi}_s\}_{s\in S}} \mathbb{P}\left(\xi_s^\tau = \bar{\xi}_s\ \forall s\in S\right)\cdot u^\tau\left(s^\tau_{\{\bar{\xi}_s\}_{s\in S}}\right) + \sum_{\{\bar{\xi}_s\}_{s\in S}} \mathbb{P}\left(\xi_s^t = \bar{\xi}_s\ \forall s\in S\right)\cdot\left(\max_s \bar{\xi}_s\right) \\
& && \text{(by definition of expectation)} \\
=&\sum_{\{\bar{\xi}_s\}_{s\in S}} \mathbb{P}\left(\xi_s = \bar{\xi}_s\ \forall s\in S\right)\left(\sum_{\tau=1}^{t} u^\tau\left(s^\tau_{\{\bar{\xi}_s\}_{s\in S}}\right) + \max_s \bar{\xi}_s\right) && \text{(by independence of noises over time)} \\
\geq&\sum_{\{\bar{\xi}_s\}_{s\in S}} \mathbb{P}\left(\xi_s = \bar{\xi}_s\ \forall s\in S\right)\left(\left(\sum_{\tau=1}^{t} u^\tau(\tilde{s})\right) + \bar{\xi}_{\tilde{s}}\right) && \text{(by claim 2)} \\
=&\left(\sum_{\tau=1}^{t} u^\tau(\tilde{s})\right) + \mathbb{E}(\bar{\xi}_{\tilde{s}})
\end{aligned}
$$

Second Proof We prove by an induction similar to Claim 1. For $t = 1$, by Corollary 3 for every $\tilde{s}\in S$ we have:[2]

$$\mathbb{E}\left(u^1(\bar{s}^1)\right) + \mathbb{E}\left(\max_s \xi_s^1\right) \geq u^1(\tilde{s}) + \mathbb{E}\left(\xi_{\tilde{s}}^1\right)$$

When $t > 1$, for every $\{\xi_s^t\}_{s\in S}$, and $\tilde{s}\in S$ we have

$$
\begin{aligned}
&\mathbb{E}\left(\sum_{\tau=1}^{t} u^\tau(\bar{s}^\tau)\middle|\{\xi_s^t\}_{s\in S}\right) + \mathbb{E}\left(\max_s \xi_s^{t-1}\middle|\{\xi_s^t\}_{s\in S}\right) \\
=&\mathbb{E}\left(\sum_{\tau=1}^{t-1} u^\tau(\bar{s}^\tau)\middle|\{\xi_s^t\}_{s\in S}\right) + \mathbb{E}\left(\max_s \xi_s^{t-1}\middle|\{\xi_s^t\}_{s\in S}\right) + \mathbb{E}\left(u^t(\bar{s}^t)\middle|\{\xi_s^t\}_{s\in S}\right) \\
\geq&\left(\sum_{\tau=1}^{t-1} u^\tau(\bar{s}^t)\right) + \mathbb{E}\left(\xi_{\bar{s}^t}^{t-1}\middle|\{\xi_s^t\}_{s\in S}\right) + \mathbb{E}\left(u^t(\bar{s}^t)\middle|\{\xi_s^t\}_{s\in S}\right) \\
&\text{(as } \{\xi_s^t\}_{s\in S} \text{ is independent from previous noises, by induction hypothesis for } \tilde{s}=\bar{s}^t\text{)} \\
=&\left(\sum_{\tau=1}^{t} u^\tau(\bar{s}^t)\right) + \mathbb{E}\left(\xi_{\bar{s}^t}^{t-1}\middle|\{\xi_s^t\}_{s\in S}\right)
\end{aligned}
$$

Taking expectation over $\{\xi_s^t\}_{s\in S}$ we get

$$\mathbb{E}\left(\sum_{\tau=1}^{t} u^\tau(\bar{s}^\tau)\right) + \mathbb{E}\left(\max_s \xi_s^{t-1}\right)$$

[2] This proof can be done independently from the previous lemmas. In that case, we need to elaborate more on the induction base.

$$\geq \left(\sum_{\tau=1}^{t} u^\tau(\bar{s}^t) \right) + \mathbb{E}\left(\xi_{\bar{s}^t}^{t-1}\right)$$

$$= \left(\sum_{\tau=1}^{t} u^\tau(\bar{s}^t) \right) + \mathbb{E}\left(\xi_{\bar{s}^t}^{t}\right) \qquad \text{(as } \xi_s^t \text{ and } \xi_s^{t-1} \text{ have the same distribution)}$$

$$\geq \left(\sum_{\tau=1}^{t} u^\tau(\tilde{s}) \right) + \mathbb{E}\left(\xi_{\tilde{s}}^{t}\right) \qquad \text{(by choice of } s^t \text{ and taking expectation)}$$

■

Next, we would like to argue that for every t and a suitable choice of ξ_s^t, with a reasonably high probability the maximizer of $\left(\sum_{\tau=1}^{t} u^\tau(s)\right) + \xi_s^t$ is *also* the maximizer of $\left(\sum_{\tau=1}^{t-1} u^\tau(s)\right) + \xi_s^t$. This means that Follow the Perturbed Leader and Be the Perturbed Leader often suggest the learner the same strategy.

Back to Follow the Perturbed Leader For convenience let

$$U^t(s) := \sum_{\tau=1}^{t} u^\tau(s)$$

Recall that $\bar{s}^t := \arg\max_s U^t(s) + \xi_s^t$ is what noisy cheat algorithm plays at time t. As described earlier, in Follow the Perturbed Leader, at every time t, the learner plays

$$\hat{s}^t := \arg\max_s U^{t-1}(s) + \xi_s^t \tag{6}$$

where ξ_s^t are independent. For the following lemma, recall from basic probability that geometric distribution $Geo(\epsilon)$ describes the number of times we need to flip a coin which lands on head w.p. ϵ until a head occurs.

Lemma 5. *For a fixed* $0 < \epsilon < 1$ *and* t *we have*

$$\mathbb{P}\left(\hat{s}^t = \bar{s}^t\right) \geq 1 - \epsilon$$

where $\xi_s^t \sim Geo(\epsilon)$ *are i.i.d. and* $\hat{s}^t$ *and* $\bar{s}^t$ *are defined based on the same* $\{\xi_s^t\}_{s\in S}$.

Proof. Fix a realization of $\hat{s}^t$ like s^*, and also fix a realization $\{\bar{\xi}_s^t\}_{s\neq s^*}$ of noises. Conditioning on $\hat{s}^t = s^*$ and $\{\xi_s^t\}_{s\neq s^*} = \{\bar{\xi}_s^t\}_{s\neq s^*}$ we get

$$\max_{s\neq s^*} U^t(s) + \bar{\xi}_s^t = \max_{s\neq s^*} U^{t-1}(s) + u^t(s) + \bar{\xi}_s^t$$

$$\leq \max_{s\neq s^*} U^{t-1}(s) + \bar{\xi}_s^t + 1 \qquad \text{(as } u^t(s) \leq 1\text{)}$$

And if $\max_{s\neq s^*} U^{t-1}(s) + \bar{\xi}_s^t + 1 \leq U^{t-1}(s^*) + \xi_{s^*}^t$, we get

$$\max_{s\neq s^*} U^t(s) + \bar{\xi}_s^t \leq U^{t-1}(s^*) + \xi_{s^*}^t$$

$$\leq U^t(s^*) + \xi_{s^*}^t \qquad \text{(as } u^t(s) \geq 0\text{)}$$

This implies[3]

$$\bar{s}^t = \arg\max_s U^t(s) + \xi_s^t = s^*$$

[3] We need to be more careful with deriving this conclusion in the case where the maximizer is not unique. One way to solve this issue is by having $u^t(s) \in [0, 1)$.

Therefore

$$
\begin{aligned}
&\mathbb{P}\left(\bar{s}^t = \hat{s}^t \mid \hat{s}^t = s^*, \{\xi_s^t\}_{s\neq s^*}\right) \\
=&\mathbb{P}\left(\bar{s}^t = s^* \middle| \{\xi_s^t\}_{s\neq s^*}, \max_{s\neq s^*} U^{t-1}(s) + \xi_s^t \le U^{t-1}(s^*) + \xi_{s^*}^t\right) \\
\ge&\mathbb{P}\left(\max_{s\neq s^*} U^{t-1}(s) + \xi_s^t + 1 \le U^{t-1}(s^*) + \xi_{s^*}^t \middle| \{\xi_s^t\}_{s\neq s^*}, \max_{s\neq s^*} U^{t-1}(s) + \xi_s^t \le U^{t-1}(s^*) + \xi_{s^*}^t\right) \\
=&\mathbb{P}\left(\max_{s\neq s^*} U^{t-1}(s) - U^{t-1}(s^*) + \xi_s^t + 1 \le \xi_{s^*}^t \middle| \{\xi_s^t\}_{s\neq s^*}, \max_{s\neq s^*} U^{t-1}(s) - U^{t-1}(s^*) + \xi_s^t \le \xi_{s^*}^t\right) \\
=&1-\epsilon
\end{aligned}
$$

Note that in the last step, we used the fact that for $X \sim Geo(\epsilon)$ we have $\mathbb{P}(X \ge i+1 \mid X \ge i) = 1-\epsilon$. Intuitively, this means if the first i coin flips have been tail, with probability $1-\epsilon$ the next one will be also a tail.

Finally from the above inequality and using the law of total probability, we get $\mathbb{P}(\bar{s}^t = \hat{s}^t) \ge 1-\epsilon$. ■

Corollary 6. *For every time t we have*

$$\mathbb{E}(u^t(\hat{s}^t)) \ge \mathbb{E}(u^t(\hat{s}^t) \mid \hat{s}^t = \bar{s}^t) \cdot \mathbb{P}(\hat{s}^t = \bar{s}^t) \ge (1-\epsilon)\cdot \mathbb{E}(u^t(\bar{s}^t))$$

Theorem 7. *For any fixed t and $0 < \epsilon < 1$, let $\hat{s}^1, \hat{s}^2, \ldots,$ be the sequence produced by Follow the Perturbed Leader (6) with noises $\xi_s^t \sim Geo(\epsilon)$, then we have*

$$\mathbb{E}\left(\sum_{\tau=1}^t u^\tau(\hat{s}^\tau)\right) \ge (1-\epsilon)\cdot \max_s \left(\sum_{\tau=1}^t u^\tau(s)\right) - \mathbb{E}\left(\max_s \xi_s^t\right)$$

We will later show $\mathbb{E}\left(\max_s \xi_s^t\right) = O\left(\frac{\ln|S|}{\epsilon}\right)$.

Proof. We have

$$
\begin{aligned}
\mathbb{E}\left(\sum_{\tau=1}^t u^\tau(\hat{s}^\tau)\right) &\ge (1-\epsilon)\cdot \mathbb{E}\left(\sum_{\tau=1}^t u^\tau(\bar{s}^\tau)\right) && \text{(by (6) and linearity of expectation)} \\
&\ge (1-\epsilon)\cdot\left[\max_s \left(\sum_{\tau=1}^t u^\tau(s)\right) - \mathbb{E}\left(\max_s \xi_s^t\right)\right] && \text{(by Lemma 4)} \\
&\ge (1-\epsilon)\cdot \max_s \left(\sum_{\tau=1}^t u^\tau(s)\right) - \mathbb{E}\left(\max_s \xi_s^t\right) && \text{(as } \mathbb{E}\left(\max_s \xi_s^t\right) \ge 0\text{)}
\end{aligned}
$$

■

The above guarantee gives us the same regret bound as MWU by setting $\epsilon = \frac{1}{\sqrt{T\cdot \ln|S|}}$. Finally, we provide a short proof for $\mathbb{E}\left(\max_s \xi_s^t\right) = O\left(\frac{\ln|S|}{\epsilon}\right)$.

Lemma 8. *For $\xi_s \sim Geo(\epsilon)$ and independent, we have $\mathbb{E}(\max_s \xi_s) = O(\frac{\ln|S|}{\epsilon})$.*

Proof. Note that $\mathbb{P}(\xi_s > m) = (1-\epsilon)^m$ by definition, so

$$\mathbb{P}\left(\xi_s > \epsilon^{-1}(a + \ln|S|)\right) = (1-\epsilon)^{\epsilon^{-1}(a+\ln|S|)} \le e^{-(a+\ln|S|)} = \frac{e^{-a}}{|S|}$$

using that $(1-\epsilon)^{-\epsilon} \approx e^{-1}$. By the union bound we have

$$\mathbb{P}\left(\max_s \xi_s > \epsilon^{-1}(a + \ln|S|)\right) \leq e^{-a}$$

Now using this fact, the expected value can be bounded by

$$\begin{aligned}
\mathbb{E}(\max_s \xi_s) &= \sum_{i=1}^{\infty} \mathbb{P}\left(\max_s \xi_s > i\right) \\
&\leq \frac{\ln|S|}{\epsilon} + \frac{1}{\epsilon}\sum_{a=1}^{\infty} \mathbb{P}\left(\max_s \xi_s > \epsilon^{-1}(a + \ln|S|)\right) \\
&\leq \frac{\ln|S|}{\epsilon} + \frac{1}{\epsilon}\sum_{a=1}^{\infty} e^{-a} = O\left(\frac{\ln|S|}{\epsilon}\right)
\end{aligned}$$

■

The Missing Piece So far we assumed the learner has full information about the history. That is, he knows the past utilities for every strategy. This assumption is not very realistic as in many settings the learner only observes the utility of the strategy he plays and not every other strategy. Next time, we will try to solve this issue.