University of California, Berkeley Lecturer: Anant Sahai

EE 290-S: Advanced Topics in Communications and Information Theory

Sep 25, 2018 Scribes: Heyuan Liu, Renyuan Xu

Lecture 10: Minimax Problem and Adaptation

1 Outline

Today's topics:

- Finding Nash equilibrium via learning zero-sum games.
- Adaption to unknown time horizon T.

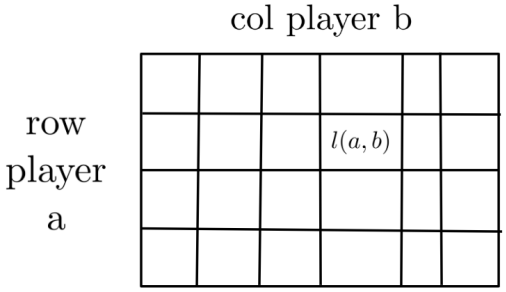
For the Nash equilibrium part, we will see how to use an online learning algorithm to get an approximate Nash equilibrium and to prove that

$$\max_{p_a} \min_{p_b} \bar{l}(p_a, p_b) = \min_{p_a} \max_{p_b} \bar{l}(p_a, p_b)$$

for a linear loss function $\bar{l}(\cdot, \cdot)$.

For the adaption part, we will see what is the strategy when we don't know the time horizon T and how to get an $O(\sqrt{T})$ regret.

2 Nash equilibrium in Zero-Sum Game



Loss matrix for row player is,

$$\bar{l}(p_a, p_b) = \sum_{a,b} p_a(a) p_b(b) l(a, b).$$

Note that in zero-sum game, for other is just negative loss. Since the loss matrix of column player is $-l(a, b)$, what he wants is to maximize $\bar{l}(p_a, p_b)$. Recall the definition of a Nash equilibrium (p_a^*, p_b^*):

$$\bar{l}(p_a^*, p_b^*) \leq \bar{l}(p_a, p_b^*), \forall p_a,$$
$$\bar{l}(p_a^*, p_b^*) \geq \bar{l}(p_a^*, p_b), \forall p_b.$$

In the zero-sum game, the nature questions to ask are: does there exist (p_a^*, p_b^*) satisfying these condition? or does $v_{\max,\min}$ equal to $v_{\min,\max}$? where

$$v_{\max,\min} = \max_{p_a} \min_{p_b} \bar{l}(p_a, p_b),$$
$$v_{\min,\max} = \min_{p_a} \max_{p_b} \bar{l}(p_a, p_b).$$

Next, we will implement a online learning to construct $(\bar{p}_a, \bar{p}_b)$ such that it is close enough to $v_{\max,\min}$ and $v_{\min,\max}$. The brief idea is the following:

1 Player (A) (us) choose any abitrary distribution $p_{a,1}$;

2 Player (B) (nature) choose a distribution $p_{b,1}$ to "hurt" us, that is, to maximize $\bar{l}(p_{a,1}, p_{b,1})$;

3 Player (A) then implement a "no-regret learning", like FTPL or FTRL, to get $p_{a,2}$.

4 ...

Now we formalize our idea to an algorithm to construct $(\bar{p}_a, \bar{p}_b)$.
Goal: find a Nash equilibrium (approximately).
Approach: Pick a sufficiently large T.

- For (A), play "no-regret learning": a strategy with cumulative regret $O(\sqrt{T})$, that is, with average regret $O\left(\frac{1}{\sqrt{T}}\right)$.
- For (B), play $p_{b,t} = \arg\max_{p_b'} \bar{l}(p_{a,t}, p_b')$, or (b) can also implement "no-regret" learning.

Claim: $\bar{p}_a = \frac{1}{T}\sum_{t=1}^{T} p_{a,t}$ and $\bar{p}_a = \frac{1}{T}\sum_{t=1}^{T} p_{a,t}$ are close to Nash equilibrium.
The proof is as following.

Theorem 1 *Suppose we play the game for T rounds, it holds that*

$$v_{\min,\max} \leq v_{\max,\min} + O\left(\frac{1}{\sqrt{T}}\right).$$

Proof Let $\bar{p}_a, \bar{p}_b$ be the time average we get from game. First it holds that

$$\begin{aligned} v_{\min,\max} &= \min_{p_a} \max_{p_b} \bar{l}(p_a, p_b) \leq \max_{p_b} \bar{l}(\bar{p}_a, p_b) \\ &= \max_{p_b} \bar{l}\left(\frac{1}{T}\cdot\sum_{t=1}^{T} p_{a,t}, p_b\right) = \max_{p_b}\left\{\frac{1}{T}\cdot\sum_{t=1}^{T}\bar{l}\left(p_{a,t}, p_b\right)\right\}, \end{aligned}$$

where the first equality comes from the definition of $v_{\min,\max}$, the second equality comes from the definition of $\bar{p}_a$ and the third comes from the linearity of $\bar{l}(\cdot,\cdot)$. Then, we have that

$$\max_{p_b}\left\{\frac{1}{T}\cdot\sum_{t=1}^{T}\bar{l}\left(p_{a,t}, p_b\right)\right\} \leq \frac{1}{T}\cdot\sum_{t=1}^{T}\max_{p_b}\bar{l}\left(p_{a,t}, p_b\right) \leq \frac{1}{T}\cdot\sum_{t=1}^{T}\bar{l}\left(p_{a,t}, p_{b,t}\right),$$

where the second inequality comes from definition of $p_{b,t}$. Since the player A implement a no-regret learning, it holds that

$$\begin{aligned}
\frac{1}{T}\cdot\sum_{t=1}^{T}\bar{l}\left(p_{a,t},p_{b,t}\right) &\leq \min_{p'_a}\left\{\frac{1}{T}\cdot\sum_{t=1}^{T}\bar{l}\left(p'_a,p_{b,t}\right)\right\}+O\left(\frac{1}{\sqrt{T}}\right)\\
&\leq \min_{p'_a}\bar{l}\left(p'_a,\frac{1}{T}\cdot\sum_{t=1}^{T}p_{b,t}\right)+O\left(\frac{1}{\sqrt{T}}\right)\\
&\leq \max_{p'_b}\min_{p'_a}\bar{l}\left(p'_a,p'_b\right)+O\left(\frac{1}{\sqrt{T}}\right)\\
&= v_{\max,\min}+O\left(\frac{1}{\sqrt{T}}\right),
\end{aligned}$$

where the first inequality holds since the total regret is $O(\sqrt{T})$, the second inequality comes from the linearity of $\bar{l}(\cdot,\cdot)$ and the equality comes from the definition of $v_{\max,\min}$. ■

Remark 1 *When $T\to\infty$, $v_{\min,\max}\leq v_{\max,\min}$ and that*

$$\bar{l}(\bar{p}_a,\bar{p}_b)\in\left[v-O\left(\frac{1}{\sqrt{T}}\right),v+O\left(\frac{1}{\sqrt{T}}\right)\right].$$

Also note that $\bar{p}_a$ is a robust choice:it only deviate from p_a^* by $O\left(\frac{1}{\sqrt{T}}\right)$ in regret.

Remark 2 *Theorem 1 also holds when $\bar{l}(\cdot,\cdot)$ is continuous and convex-concave*[1]*. (Linearity of $\bar{l}$ now becomes Jansen's inequality.)*

Theorem 2 *For any arbitrary $f(\cdot,\cdot)$, it holds that*

$$\min_{p_a}\max_{p_b}f(p_a,p_b)\geq\max_{p_b}\min_{p_a}f(p_a,p_b).$$

Proof It holds that

$$\begin{aligned}
&f(p'_a,p_b)\geq\min_{p_a}f(p_a,p_b),\ \forall p'_a\\
\Rightarrow&\max_{p_b}f(p'_a,p_b)\geq\max_{p_b}\min_{p_a}f(p_a,p_b),\ \forall p'_a && \text{(take maximum of } p_b \text{ on both side)}\\
\Rightarrow&\min_{p_a}\max_{p_b}f(p_a,p_b)\geq\max_{p_b}\min_{p_a}f(p_a,p_b) && \text{(by definition)}
\end{aligned}$$

■

Remark 3 *Even if $p_{b,t}$ is also chosen by "no-regret" learning, $p_{a,t}$ and $p_{b,t}$ still may not converge. We can only get the "time average" $\bar{p}_a$ and $\bar{p}_b$ converge to the Nash equilibrium.*

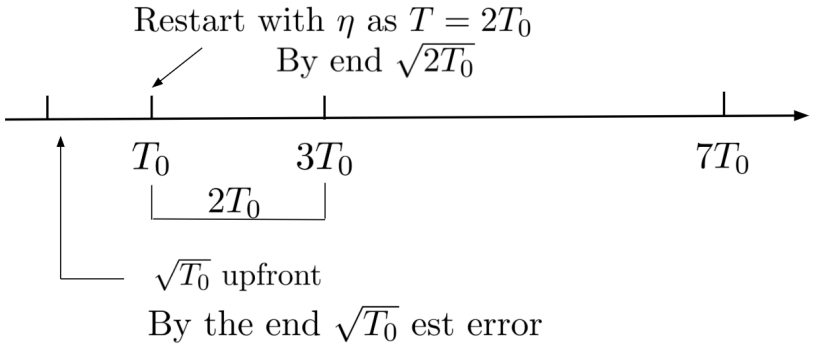
[1] $l(\cdot,\cdot)$ is convex-concave when $l(\cdot,p_b)$ is convex and $l(p_a,\cdot)$ is concave for any fixed p_a, p_b.

3 Adaptation

The Skier's Rule (motivating example): A skiing fan can choose between buying the ski kit with $ 500 and renting the ski kit for $ 50 each time. He does not known when he will stop skiing. That is, there is a random stopping time T for him. For example, if he breaks his leg at time T, then he needs to stop skiing.
Strategy (intuiation): Rent until he has paid buying cost, then buy.

Sequential Learning: Pay 2 error terms: approximation error ($O(\frac{1}{\eta})$ once) + estimation error (η per time-step).

Doubling Trick: start with some η for some T_0. Restart with η as $T = T_0$.



For the first time-interval $[0, T_0]$, we pay $O(\sqrt{T})$ approximation error in the beginning, and in the end, we pay $O(\sqrt{T})$ cumulative estimation error. For the second time-interval $[T_0, 3T_0]$, we pay both $O(\sqrt{2T})$ approximation and estimation error.