Game Dynamics

Thursday, November 10, 2022 10:55 AM

★ **Best Response**: If $\exists$ a player who can deviate & improve, then one of them does. ←

eg. Matching Pennies

$$\begin{array}{c|cc} P_1 \backslash P_2 & H & T \\ \hline H & 1,-1 & -1,1 \\ T & -1,1 & 1,-1 \end{array}$$

$P_2 \quad P_1 \quad P_1 \quad P_2$

No pointwise convergence.

$$\text{Avg.} = \frac{1}{T}\left(\#\text{time } H \text{ played},\ \#\text{times } T \text{ played}\right)$$

$$= \frac{1}{T}\left(\sim\frac{T}{2},\ \sim\frac{T}{2}\right)$$

$$= \left(\frac{1}{2}, \frac{1}{2}\right)$$

≈ In zero-sum games avg. of Fictitious Play (pure play) deterministic

converges to a NE.

Q: For what games BRD converges point wise?

1. Games where every player has a dominant strategy
2. Potential Games.

$$\forall i,\ \forall q_i \in \mathcal{S}_i,\ \forall P \in \underset{i}{X}\, \mathcal{S}_i \qquad C_i(q_i, P_{-i}) - C_i(P_i, P_{-i}) = \phi(q_i, P_{-i}) - \phi(P_i, P_{-i})$$

$\Downarrow$

V_i

★ General non-zero-sum Games?

- Not even avg. may converge!

*

- Not even any way converge!

$$\begin{array}{c|c|c} & T & F \\ \hline T & 1,2 & 0,0 \\ \hline F & 0,0 & 2,1 \end{array}$$

Simultaneous BR. does not converge.

- Req. coordination.
- Req. knowledge of everyone's play

⋮

---

* No-Regret Dynamics:

* External Regret Model. vs (Adversary).

Decision maker / player / agent vs adversary.

has action set A.

* At time $t = 1 \ldots, T$

(mixed-strategy)

1. player chooses a prob. dist $p^t \in \Delta(A)$
2. adversary picks cost $c^t : A \to [0,1]$
3. player chooses a^t w.p. $p^t(a^t)$ & incurs cost $c^t(a^t)$

   learns entire c^t.

(Learn only $c^t(a^t)$ = Bandit model
can get similar guarantees w/ a bit of loss)

... the player is compared

Q: How bad $\sum_{t=1}^{T} c^t(a^t)$ = cost of the player is compared to the best possible?

$$\sum_{t=1}^{T} \min_{a \in A} c^t(a)$$

eg. $A = \{1, 2\}$.
$t = 1 \ldots T$

Adversary: if $p_1^t \geq p_2^t$ then set $c^t = (1, 0)$
Else $(p_1^t < p_2^t)$ then set $c^t = (0, 1)$

Expected cost $a^t \sim p^t \geq \max\{p_1^t, p_2^t\}$

$$\geq 1/2 .$$

$$\Rightarrow \sum_{t=1}^{T} c^t(a^t) \geq \frac{T}{2}$$

$\Rightarrow$ Best total cost = 0.

Q: Compare to Best action in hind sight.

$$\min \sum^{T} c^t(a)$$

$$\boxed{\min_{a \in A} \sum_{t=1}^{T} c^t(a)}$$

Def: Time-avg-Regret wrt action a.

$$\frac{1}{T}\left(\sum_{t=1}^{T} c^t(a^t) - \sum_{t=1}^{T} c^t(a)\right)$$

Def: Algo A is no-regret if Time-avg-regret wrt best action $\to 0$ as $T \to \infty$

$$\lim_{T\to\infty} \frac{1}{T}\left[\text{Expected cost of } A - \text{cost of best action}\right] \to 0$$

$$\lim_{T\to\infty} \frac{1}{T}\left[\sum_{t=1}^{T}\sum_{a\in A} p^t(a)\, c^t(a) - \min_{a\in A}\sum_{t=1}^{T} c^t(a)\right]$$

Q: Would a deterministic strategies (p^t) work?

NO!

a_1 w.p. 1 then adversary $\begin{array}{cc} c^t(1) & c^t(2) \\ (1, & 0) \end{array}$

& a_2 w.p. 0

o.w.

a_1 w.p. 0   then adversary $(0,1)$

a_2 w.p. 1

cost-player = T #times play a_1 + #times play a_2 $= T$

cost-best-action in hindsight? $\Rightarrow \min\{\#\text{times } a_1, \#\text{times } a_2\}$

$\leq$ $\leq \frac{T}{2}$

$T/2$

$$\frac{1}{T}\left(T - \frac{T}{2}\right) \geq \frac{1}{2} \not\to 0$$

$\boxed{\text{Randomization is important.}}$

eg.: $A = \{1, 2\}$
suppose adversary chooses

$c^t = (1, 0)$ w.p. $1/2$
$= (0, 1)$ w.p. $1/2$.

No matter p^t, the expected cost $= \frac{p^t_1}{2} + \frac{p^t_2}{2}$

$= \frac{1}{2}$

$$\mathbb{E}\left[\begin{array}{c}\text{cost of} \\ \text{player}\end{array}\right] = \frac{T}{2} \leftarrow \text{Expection of } \sum_t c^t = \frac{1}{2}$$

variance $\sqrt{T}$

$$\mathbb{E}\left[\begin{array}{c}\text{cost of} \\ \text{best action} \\ \text{in hindsight}\end{array}\right] \leq \frac{T}{2} - \theta(\sqrt{T}) \quad \left(\text{with constant prob.}\right)$$

$$= \frac{1}{T}\left(\frac{T}{2} - \frac{T}{2} + \theta(\sqrt{T})\right) \approx \Omega\left(\frac{1}{\sqrt{T}}\right)$$

if n actions $\approx \Omega\left(\sqrt{\frac{\ln n}{T}}\right)$.

Thm: $\exists$ no-regret algo with avg. regret $\leq O\left(\sqrt{\frac{\ln n}{T}}\right) = \varepsilon$

$$T = O\left(\frac{\ln n}{\varepsilon^2}\right) \Leftarrow$$

★ Multiplicative - Weight - Update (MWU)
or
Hedge
or
expert advise.

Intuition: increase prob. of good action
≡ aggressively punish bad actions.

$\equiv$ aggresively punish bad actions.

Player's Algorithm:

- $w^t(a) = 1 \quad \forall a \in A$ (initialization)
- For $t = 1 \dots T$

① p^t prop. to w^t : $\forall a \in A$, $p^t(a) = \frac{w^t(a)}{\Gamma^t}$

where $\Gamma^t = \sum_{a \in A} w^t(a)$

$a^t \sim p^t \quad \dots$

② Given $c^t : A \to [0, 1]$, update $\forall a \in A$

$$\boxed{w^{t+1}(a) = w^t(a) \, (1 - \epsilon)^{c^t(a)}}$$

$\epsilon \to 1 \quad w^{t+1}(a) \approx w^t(a)$: Exploration

$\epsilon \to 0 \quad$ B.R. : Exploitation.

$\epsilon \in [0, 1/2]$ pick later.

Pf: $\Gamma^t = \sum_{a \in A} w^t(a)$.

Goal: Relate $\sum_{t=1}^{T} \sum_{a \in A} p^t(a) \, c^t(a)$ to $\underbrace{\overset{a^*}{\underset{a \in A}{\arg\min}} \sum_{t=1}^{T} c^t(a)}_{OPT}$

via Γ^T

$$\Gamma^T = \sum_{a\in A} w^T(a) \geq w^T(a^*)$$

$$= w^1(a^*) \prod_{t=1}^{T} (1-\varepsilon)^{c^t(a^*)}$$

$$= (1-\varepsilon)^{\sum_{t=1}^{T} c^t(a^*)} = (1-\varepsilon)^{OPT} \rightarrow 1$$

- Expected cost of MWU in time t

$$\boxed{\nu^t = \sum_{a\in A} p^t(a)\, c^t(a) = \sum_{a\in A} \frac{w^t(a)}{\Gamma^t}\, c^t(a).} \rightarrow 2$$

$$\Gamma^{t+1} = \sum_{a\in A} w^{t+1}(a)$$

$$= \sum_{a\in A} w^t(a)\, (1-\varepsilon)^{c^t(a)}$$

$$\leq \sum_{a\in A} w^t(a)\, \left(1-\varepsilon c^t(a)\right) \qquad \left(\because\ \varepsilon \in [0, 1/2],\ x \in [0,1],\ (1-\varepsilon)^x \leq (1-\varepsilon x)\right)$$

$$= \Gamma^t \left(\underbrace{\sum_{a\in A} \frac{w^t(a)}{\Gamma^t}}_{=1} - \varepsilon \underbrace{\sum_{a\in A} \frac{w^t(a)}{\Gamma^t}\, c^t(a)}_{=\nu^t} \right)$$

$$= \Gamma^t (1 - \varepsilon \nu^t) \quad (\because 2) \rightarrow 3$$

$$(1-\epsilon)^{OPT} \underset{(\because ①)}{\le} \Gamma^T \underset{(\because ③)}{\le} \underset{n}{\overset{\shortparallel}{\Gamma^1}} \prod_{t=1}^{T} (1-\epsilon\, v^t)$$

apply ln $\Rightarrow$ $$OPT \ln(1-\epsilon) \le \ln n + \sum_{t=1}^{T} \ln(1-\epsilon\, v^t)$$

$$\ln x = \left(-x - \frac{x^2}{2} - \frac{x^3}{2} \cdots\right) \quad x \in [0,1]$$

$$-x - x^2 \le \ln x \le -x$$

$$\Rightarrow \quad OPT(-\epsilon-\epsilon^2) \le \ln n + \sum_{t=1}^{T} -(\epsilon v^t)$$

$$\overset{*(-1)}{\Rightarrow} \quad OPT\, \cancel{\epsilon}(1+\epsilon) \ge \frac{-\ln n}{\epsilon} + \cancel{\epsilon} \sum_{t=1}^{T} v^t$$

$$E[\text{cost MWU}] = \sum_{t=1}^{T} v^t \le \frac{\ln n}{\epsilon} + OPT + \epsilon\, OPT$$

$$\le \frac{\ln n}{\epsilon} + \underline{\epsilon} \cdot T + OPT$$

set $\boxed{\epsilon = \sqrt{\frac{\ln n}{T}}}$

set $\boxed{\epsilon = \sqrt{\frac{\ln n}{T}}}$

$$\mathbb{E}[\text{cost MWU}] \le \sqrt{\ln n \cdot T} + \sqrt{\ln n \cdot T} + OPT$$

$$\Rightarrow \text{avg Regret} = \frac{1}{T}\Big[\mathbb{E}[\text{cost MWU}] - OPT\Big]$$

$$\le \frac{1}{\cancel{T}_{\sqrt{T}}}\Big[2\sqrt{\ln n} \cdot \cancel{\sqrt{T}}\Big]$$

$$= 2\sqrt{\frac{\ln n}{T}}$$

N-player game.

$i \in N$ moves S_i

Each player plays as per MWU.
(no-regret algo).

$\equiv$ after $T = \frac{\ln n}{\epsilon^2}$ rounds avg regret $\le \epsilon$.

Then time avg play $= \epsilon$-CCE.

$p_i^t \in \Delta(S_i)$

~~Proof~~: Let player i plays $p_i^t \in \Delta(S_i)$

$$\sigma^t = \prod_{i=1}^{N} p_i^t$$

$$\sigma = \sum_{t=1}^{T} \sigma^t . \quad \text{Then}$$

$\underline{\sigma}$ is ϵ-CCE.

$$\forall i, \forall s_i' \in S_i \quad \mathbb{E}_{s\sim\sigma}\left[C_i(s)\right] \leq \mathbb{E}_{s\sim\sigma}\left[C_i(s_i', s_{-i})\right] + \epsilon$$

exe.

$$\mathbb{E}_{s\sim\sigma}\left[C_i(s)\right] = \mathbb{E}\begin{bmatrix}\text{cost of} \\ \text{MWU for} \\ \text{player } i\end{bmatrix}$$

$$\mathbb{E}_{s\sim\sigma}\left[C_i(s_i', s_{-i})\right] = \mathbb{E}\begin{bmatrix}\text{cost of} \\ \text{action } s_i'\end{bmatrix}$$