Game Dynamics

Best Response: If I a player also can deviate to improve, then one or them does.

e.g. Matching Pennies

No point wise convergence.

\[ \text{AVG} = \frac{1}{T} \left( \frac{\text{# of } H}{\text{played}}, \frac{\text{# of } T}{\text{played}} \right) \]

\[ = \frac{1}{T} \left( \frac{T}{2}, \frac{T}{2} \right) \]

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

\[ \approx \text{In zero-sum games avg. of fictitious play (pure play) deterministic converges to } \text{a NE}. \]

Q: For what games BRD converges point wise? 

1. Games where every player has a dominant strategy.

2. Potential Games.

\[ V_i, \forall i \in \mathcal{P} \]

\[ V_i, \forall i \in \mathcal{P} \]

\[ C_i(q_i, p_{-i}) - C_i(p_i, p_{-i}) = \psi(q_i, p_{-i}) - \psi(p_i, p_{-i}) \]

* General non-zero-sum Games. Not even any way converge!
* No-Regret Dynamics:

* External Regret Model. vs (Adversary).

\[
\text{Decision / player / agent vs adversary.}
\]

* At time \( t = 1, \ldots, T \) (mixed-strategy).

1. Player chooses a prob. dist \( \pi_t \in \Delta(A) \)
2. Adversary picks cost \( c_t : A \to [0, 1] \)
3. Player chooses \( a_t \) w.p. \( \pi_t(a_t) \) & incurs cost \( c_t(a_t) \)

\( \downarrow \)

* Learn only \( c_t(a_t) \) = Bandit model

(\text{can get similar guarantees w/ a bit of loss})

\[ T \to \infty \text{ assumes is covered} \]
Q: How bad does \( \frac{1}{t} \sum_{t=1}^{T} c_t(a_t) \) cost as the player is compared to the best possible?

\[
\frac{1}{T} \sum_{t=1}^{T} c_t(a_t) \geq \min_{a \in A} \sum_{t=1}^{T} c_t(a_t)
\]

e.g. \( A = \{1, 2, 3\} \)

Adversary: if \( p_t^1 \geq p_t^2 \) then set \( c_t = (1, 0) \)
else \( (p_t^1 < p_t^2) \) then set \( c_t = (0, 1) \)

Expected cost \( a \sim p_t \geq \max \{p_t^1, p_t^2\} \geq \frac{1}{2} \cdot \varepsilon \)

\[
\Rightarrow \frac{1}{T} \sum_{t=1}^{T} c_t(a_t) \geq \frac{T}{2}
\]

\( \Rightarrow \) Best, cost = 0.

Q: Compare to \( \min_{a} \sum_{t=1}^{T} c_t(a) \)
Def: Time-avg-Regret w.r.t action a.

\[ \frac{1}{T} \left( \sum_{t=1}^{T} c(a_{t}) - \min_{a \in A} \sum_{t=1}^{T} c(a) \right) \]

Def: Alg A is no-regret if Time-avg-regret w.r.t best action \( \to 0 \) as \( T \to \infty \)

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

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

Q: Would a deterministic strategies (p\textsubscript{b}) work?

No!

\[ \tilde{c}(1), \tilde{c}(2) \]

w.p. 1 then adjacent

\[ (1, 0) \]

\[ \tilde{c}(1), \tilde{c}(2) \]

w.p. 0
0.w.
q, w.p. 0 is his adversary \((0, 1)\)
q' w.p. 1

\[
\text{Cost - player} = T
\]
\[
\text{Cost - best action in hindsight} \leq \frac{T}{2}
\]
\[
\frac{1}{T} - \frac{T}{2} \geq \frac{1}{2} \times 0
\]

Randomization is important.

Eq.: \(A = \{1, 2, 3\}\)
suppose adversary chooses
\[
\ell = (1, 0) \quad \text{w.p. } \frac{1}{2}
\]
\[
= (0, 1) \quad \text{w.p. } \frac{1}{2}
\]

No matter \(pt\), the expected cost = \(\frac{pt}{2} + \frac{pt}{2}\)
= \(\frac{1}{2}\)
\[ E \left[ \text{cost to player} \right] = \frac{1}{2} \quad \text{Expectation} \quad A^\text{Past} \quad \frac{1}{\sqrt{T}} \quad \text{Variance} \quad \sqrt{T} \]

\[ E \left[ \text{cost to best action in hindsight} \right] \leq \frac{1}{2} - \Theta(1/\sqrt{T}) \quad \text{(with constant)} \quad \text{prob.} \]

\[ = \frac{1}{T} \left( \frac{T}{2} - \frac{T}{2} + \Theta(1/\sqrt{T}) \right) \approx \frac{1}{2T} \]

\[ \text{in actions} \approx \mathcal{N} \left( \frac{\ln m}{T} \right) \]

Then: A no-regret algo with avg. regret \( \approx 0 \left( \sqrt{\frac{\ln m}{T}} \right) \)

\[ T = O \left( \frac{\ln m}{\epsilon^2} \right) \]

**Multiplicative Weight Update (MWU)** or Hedge

Intuition: increases prob. of good action

\( \approx \) aggressively punish bad actions.
Player's Algorithm:

1. \( w^0(a) = 1 \quad \forall a \in A \) (initialization)
2. For \( t = 1 \ldots T \)
   
   1. \( p_t \) prop. to \( w_t : \forall a \in A, p_t(a) = \frac{w_t(a)}{\sum_w w_t(w)} \)
   
   where \( r_t = \sum_w w_t(w) \)

3. \( \epsilon \sim p_t \) ...

4. Given \( \epsilon : A \to [0, 1] \), update \( \forall a \in A \)

\[
\begin{aligned}
&w_{t+1}(a) = w_t(a) (1 - \epsilon) c_t(a) \\
\end{aligned}
\]

\( \epsilon \to 1 \quad w_{t+1}(a) = w_t(a) \) : Exploration

\( \epsilon \to 0 \quad \text{B.R.} \) : Exploitation.

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

\[ \Gamma_t = \sum_a w_t(a) \]

Goal: Relate \( \sum_{t=1}^{T} \sum_{a \in A} p_t(a) c_t(a) \) to \( \min_{a \in A} \sum_{t=1}^{T} c_t(a) \) via \( \Gamma \).
\[ \Gamma^T = \sum_{a \in A} \omega^T(a) \geq \omega^T(a^*) = \omega^t(a^*) \prod_{t=1}^{T} (1 - \varepsilon) \]

\[ = (1 - \varepsilon)^{\sum_{t=1}^{T} \delta_t(a^*)} = (1 - \varepsilon)^{\text{opt}} \rightarrow 0 \]

- Expected cost for MWU in round \( t \)

\[ v_t^T = \sum_{a \in A} \rho_t(a) c_t(a) = \sum_{a \in A} \frac{\omega^t(a)}{\Gamma^t} c_t(a) \rightarrow 0 \]

\[ \Gamma^{t+1} = \sum_{a \in A} \omega^{t+1}(a) \]

\[ = \sum_{a \in A} \omega^t(a) (1 - \varepsilon) c_t(a) \]

\[ \leq \sum_{a \in A} \omega^t(a) (1 - \varepsilon c_t(a)) \begin{pmatrix} \delta^{[0, \frac{1}{2}]} \\ \varepsilon \in [0, 1] \\ (1 - \varepsilon) \leq (1 - \varepsilon) \end{pmatrix} \]

\[ = \Gamma^t \left( \sum_{a \in A} \frac{\omega^t(a)}{\Gamma^t} \right) - \varepsilon \sum_{a \in A} \frac{\omega^t(a)}{\Gamma^t} c_t(a) \]

\[ = \Gamma^t (1 - \varepsilon v_t) \rightarrow 0 \rightarrow 0 \]

- \( \Gamma_t \)
\[(1-e^e)^{\text{opt}} \leq \Gamma \leq \Gamma' \leq \prod_{t=1}^{T} (1-e^e) \]

\[
\text{apply } \ln \]

\[
\text{opt } \ln(1-e^e) \leq \ln n + \frac{1}{T} \sum_{t=1}^{T} \ln (1-e^e) \]

\[
\ln x = -x - \frac{x^2}{2} - \frac{x^3}{3} \ldots \quad x \in [0,1] \\
-x - x^2 \leq \ln x \leq -x
\]

\[
\implies \text{opt } (-e-e^e) \leq \ln n + \frac{1}{T} \sum_{t=1}^{T} -e^e
\]

\[
\text{opt } f(1+e) \geq \frac{-\ln n + e \sum_{t=1}^{T} e^e}{e}
\]

\[
E[\text{cost MWU}] = \sum_{t=1}^{T} e^e \leq \frac{\ln n + \text{opt} + e \text{opt}}{e} \leq \frac{\ln n + e \cdot T + \text{opt}}{e}
\]

\[
\text{set } \varepsilon = \sqrt{\frac{\ln n}{T}}
\]
Set $\epsilon = \sqrt{\frac{\ln n}{T}}$

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

$$\Rightarrow \text{agrregret} = \frac{1}{T} \left[ E[\text{cost MWU}] - \text{OPT} \right]$$

$$\leq \frac{1}{\sqrt{T}} \left[ \sqrt{\ln n} \cdot \sqrt{T} \right]$$

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

---

N-player game.

i.e. N moves $S_i$

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

$$\Rightarrow \text{after } T = \frac{\ln n}{\epsilon^2} \text{ rounds } \text{agrregret} \leq \epsilon.$$

Then time avg play = $\epsilon - c\text{CE}$.

$$\text{opt} = \max_{p^t \in A(S_i)}$$
Let player \( i \) plays \( p_i^t \in \Delta (S_i) \)

\[
\delta_t = \prod_{i=1}^{N} p_i^t
\]

\[
\delta = \sum_{t=1}^{T} \delta_t \quad \text{Then}
\]

\[
\delta \quad \text{is} \quad \epsilon - \text{CCE}.
\]

\[E_{s_i} \left[ G_i(s) \right] \leq E_{s_i} \left[ G_i(s_i', s_{-i}) \right] + \epsilon \]

\( i = 1 \text{ to } N \)

\[
E \left[ \text{cost of MWU for player } i \right] \quad E \left[ \text{cost of action } s_i' \right]
\]