$\frac{1}{2}$ - Approximate MSNE Computation for bimatrix games

Lecture 6.1

We assume $\underbrace{\text{w.l.o.g}}$ that all the utility values lie in $[0,1]$.

because any strategic form game remains invariant under affine transformation of utility matrices.

<u>Algorithm</u>

- Pick any strategy i for player 1.
- Let j be <u>a</u> <u>best-response</u> strategy of player 2 against i.
- Let k be <u>a</u> best-response strategy of player 1 against j.
- Output $\left(\{ i : \frac{1}{2}, k : \frac{1}{2} \}, j \right)$

Claim: $\Big(\underbrace{\{i:\tfrac{1}{2}, k:\tfrac{1}{2}\}}_{\sigma}, j\Big)$ is a $\frac{1}{2}$-MSNE.

Proof: Let the utility matrices of players 1 and 2 be A and B respectively.

$$\sigma A e_j = \frac{1}{2}\underbrace{e_i A e_j}_{\geqslant 0} + \frac{1}{2} e_k A e_j$$

$$\geqslant \frac{1}{2} e_k A e_j$$

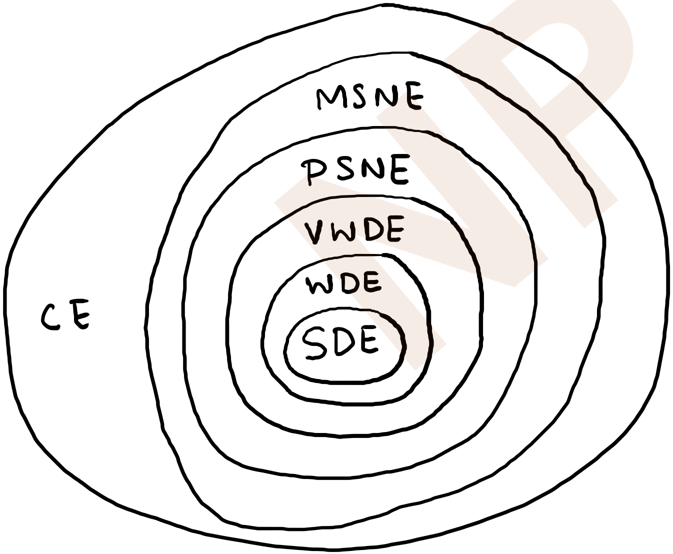
$$\sigma B e_j = \frac{1}{2} e_i B e_j + \underbrace{\frac{1}{2} e_k B e_j}_{\geqslant 0}$$

$$\geqslant \frac{1}{2} \underbrace{e_i B e_j}$$

$$\geqslant \frac{1}{2}$$

Hence $\left(\left\{i:\frac{1}{2}, k:\frac{1}{2}\right\}, j\right)$ is a $\frac{1}{2}$- MSNE.

Since finding an MSNE seems to be computationally intractable, how can we expect real-world players to compute an MSNE? This casts doubt on the predictive power of the concept of mixed strategy equilibrium.



1. There exist games with more than one PSNEs / MSNEs.

→ 2. Computing a PSNE and MSNE seems to be computationally intractable.

CE: Correlated equilibrium.

Theorem: Finding a correlated equilibrium is an efficiently solvable computational problem. ∎

Correlated Equilibrium

Lecture 6.2

Examle: (Traffic light.) Suppose there are two cars at a road junction. If both cars go, then there will be an accident.

	stop	go
stop	0, 0	0, 1
go	1, 0	−100, −100

(Stop, go) and (go, stop) are two PSNEs of this game.

$\sigma : \left((\text{stop}, \text{go}) : \frac{1}{2}, (\text{go}, \text{stop}) : \frac{1}{2} \right) \leftarrow$

Let $\sigma^* \in \Delta\left(\bigtimes_{i=1}^{n} S_i \right)$ be a probability distribution over strategy profiles. A trusted third party samples a strategy profile from σ^*, and it conveys each player its strategy only. Then, σ^* will be called a correlated equilibrium if no player has any incentive to deviate from its "advised" strategy, assuming every other player follow the expert's advice.

$\sigma = \left((\text{stop}, \text{go}) : \frac{1}{2}, (\text{go}, \text{stop}) : \frac{1}{2} \right)$ a correlated equilibrium of the traffic light game.

<u>Definition (CE)</u>: Given a normal form game $\Gamma = \langle N, (S_i)_{i \in N}, (u_i)_{i \in N} \rangle$, a probability distribution $\sigma \in \Delta\left(\bigtimes_{i=1}^{N} S_i \right)$ is called a CE if

$\forall i \in N, \forall s_i, s_i' \in S_i,$

$$\rightarrow \mathop{\mathbb{E}}_{(s_i, s_{-i}) \sim \sigma} \left[u_i(s_i, s_{-i}) \mid s_i \right] \geq \mathop{\mathbb{E}}_{(s_i, s_{-i}) \sim \sigma} \left[u_i(s_i', s_{-i}) \mid s_i \right]$$

Equivalently, for every (switching) function $\delta : S_i \longrightarrow S_i$

$$\mathop{\mathbb{E}}_{(s_i, s_{-i}) \sim \sigma} \left[u_i(s_i, s_{-i}) \right] \geq \mathop{\mathbb{E}}_{(s_i, s_{-i}) \sim \sigma} \left[u_i(\delta(s_i), s_{-i} \right]$$

Theorem: Finding a CE is polynomial time solvable.

Proof: Write a linear program for finding a CE.

Variables: $x(s)$, $s \in \bigtimes_{i=1}^{n} S_i$

$$x(s) \geq 0 \quad \forall s \in \bigtimes_{i=1}^{n} S_i, \quad \sum_{s \in \bigtimes_{i=1}^{n} S_i} x(s) = 1.$$

$\forall s_i, s_i' \in S_i$

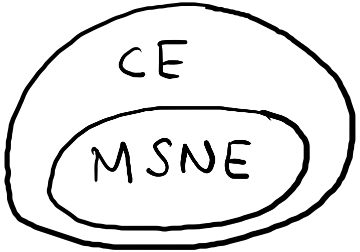
$$\underbrace{\sum_{s_{-i} \in S_{-i}} x(s_i, s_{-i})\, u_i(s_i, s_{-i})}_{\displaystyle \mathbb{E}_{(s_i, s_{-i}) \sim \sigma}\left[u_i(s_i, s_{-i}) \mid s_i\right]} \geq \sum_{s_{-i} \in S_{-i}} x(s_i', s_{-i})\, u_i(s_i, s_{-i})$$

Any feasible solution to the above linear program is a CE. ☒

Q. Does a CE always exist in a finite strategic form game?

Ans. YES! Because, every mixed strategy Nash equilibrium can be equivalently viewed as a CE. ☒

- A correlated equilibrium can be computed in polynomial time.
- Every finite game has a CE.



- $\sigma \in \Delta\left(\bigtimes_{i=1}^{n} S_i\right)$
- A trusted third party samples $(s_i)_{i \in N} \sim \sigma$

- Tells s_i to player i.

- "Non-binding" contract.

- Binding contracts.

Coarse Correlated Equilibrium

- $\sigma \in \Delta\left(\bigtimes_{i=1}^{n} S_i\right)$

Def: $\Gamma = \langle N, (u_i)_{i\in N}, (S_i)_{i\in N} \rangle$. A probability distribution $\sigma \in \Delta\left(\bigtimes_{i=1}^{n} S_i\right)$ is called a coarse correlated equilibrium (CCE) if $\forall i \in N$

$$\mathop{\mathbb{E}}_{s\sim\sigma}\left[u_i(s)\right] \geq \mathop{\mathbb{E}}_{(s_i, s_{-i})\sim\sigma}\left[u_i(s_i', s_{-i})\right] \quad \forall s_i' \in S_i$$

Observation: Every CE is also a CCE.

Corollary: Every finite game has a CCE.

Corollary: A CCE can be computed in polynomial time.

Linear program for finding a CCE:

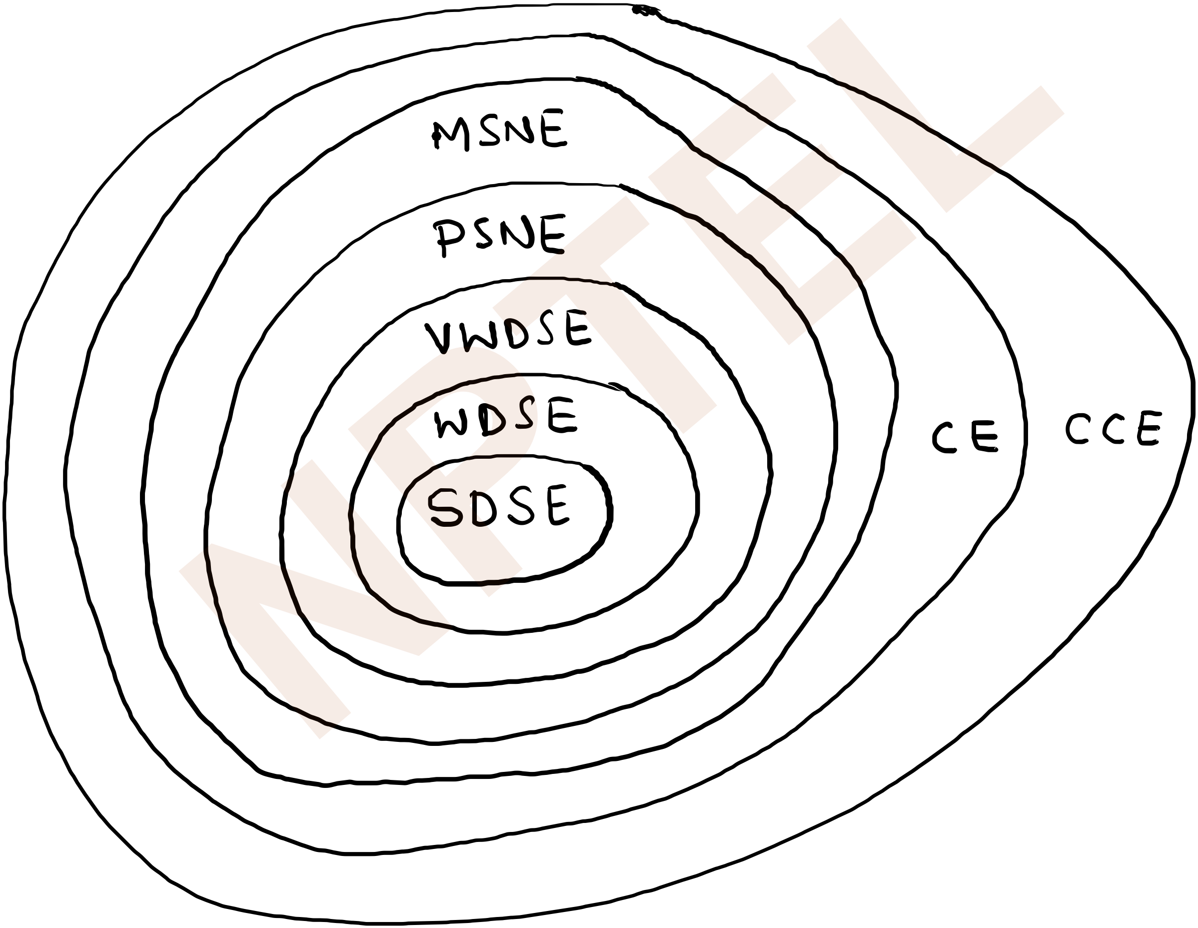
$$x_s \geq 0 \quad \forall s \in \bigtimes_{i=1}^{n} S_i$$

$$\sum_{s \in \bigtimes_{i=1}^{n} S_i} x_s = 1$$

$\forall i \in N$ $$\mathbb{E}_{s \sim \sigma}[u_i(s)] \geq \mathbb{E}_{(s_i, s_{-i}) \sim \sigma}[u_i(s_i', s_{-i})] \quad \forall s_i' \in S_i$$

LP cont.

$$\sum_{s \in \times_{i=1}^{n} S_i} x(s)\, u_i(s) \geq \sum_{(s_i, s_{-i}) \in \times_{i=1}^{n} S_i} x(s_i, s_{-i}) \cdot u_i(s_i', s_{-i})$$


MSNE
PSNE
VWDSE
WDSE
SDSE
CE
CCE

Qn: Does there exist any "natural algorithm" to find a CE or CCE?

Next lectures: Some "natural algorithm" learning dynamics which players can follow iteratively and converge to a CE or CCE.

Lecture 6.4

External Regret Framework

The setting: Iterative process between a player and an adversary.

- For each time step $t = 1, 2, \ldots, T$ $\{$
  - the player picks a probability distribution $p_t \in \Delta(A)$. // A is the set of actions available to the player
  - Adversary picks a utility function $\pi_t : A \to [0,1]$

- The player samples an action $a_t \sim p_t$ and receives a reward $\pi_t(a_t)$.

- The player gets to know π_t.

}

Total expected utility : $\sum_{t=1}^{T} \sum_{a \in A} \pi_t(a) \cdot p_t(a)$

Regret : $\underbrace{\boxed{\left(\sum_{t=1}^{T} \max_{a \in A} \pi_t(a) \right)}}_{\mathcal{B}} - \sum_{t=1}^{T} \sum_{a \in A} \pi_t(a)\, p_t(a)$

Time-averaged regret : $\frac{1}{T}\left[\left[\sum_{t=1}^{T} \max_{a\in A} \pi_t(a)\right] - \left[\sum_{t=1}^{T} \sum_{a\in A} \pi_t(a)\, p_t(a)\right]\right]$

No regret dynamic/algorithm: if time-averaged regret goes to zero when $T \to \infty$

Q: Does there exist any no regret algorithm?

Ans. NO!

$n = |A|$, $n = 2$, $A = \{a_1, a_2\}$

<u>Adversary:</u>

$$\left.\begin{array}{l}\pi_t(a_1) = 1 \\ \pi_t(a_2) = 0\end{array}\right\} \text{ if } p_t(a_1) < p_t(a_2)$$

$$\left.\begin{array}{l}\pi_t(a_1) = 0 \\ \pi_t(a_2) = 1\end{array}\right\} \text{ if } p_t(a_1) \geq p_t(a_2)$$

$B = T$,

Expected utility of the player $\leq \frac{T}{2}$

Time averaged - regret $\geq \frac{1}{T}\left[T - \frac{T}{2}\right] = \frac{1}{2}$

Weaken the benchmark:

$$B = \max_{a \in A} \sum_{t=1}^{T} \pi_t(a)$$: External regret benchmark.

Q. Does there exist any no-external regret algorithm?

Ans. YES!

No-Regret Algorithm

Theorem: Let $|A| = n$. Then there exists a no-regret algorithm whose time averaged regret is $O\left(\sqrt{\frac{\log n}{T}}\right)$.

Corollary: There exists a no-regret algorithm whose expected time averaged regret is at most ε, for any $\varepsilon > 0$, after $O\left(\frac{\log n}{\varepsilon^2}\right)$ iterations.

The algorithm is called multiplicate weight (MW)/ hedge algorithm.

Lecture 6.5

Multiplicative Weight Algorithm

1. $w_0(a) = 1 \quad \forall a \in A$
2. for $t = 1, \dots, T$ {
3. the commited probability distribution is $p_t(a) \propto w_{t-1}(a)$
4. Picks $a \sim p_t$ and receives an utility of $\pi_t(a)$

5. After knowing π_t, update $w_t(a) = w_{t-1}(a) \cdot (1+\varepsilon)^{\pi_t(a)}$ $\forall a \in A$

}

<u>Theorem:</u> Let $|A| = n$. Then the MW algorithm has external regret $O\left(\sqrt{\frac{\log n}{T}}\right)$.

<u>Proof:</u> $\Gamma_t := \sum_{a \in A} w_t(a)$ the sum of the weight values in the t-th iteration.

✓ Expected utility: $\sum_{t=1}^{T} \sum_{a \in A} p_t(a) \cdot \pi_t(a) = \sum_{t=1}^{T} \boxed{\sum_{a \in A} \frac{w_{t-1}(a)}{\Gamma_{t-1}(a)} \cdot \pi_t(a).}$

✓ Benchmark (OPT): $\max_{a \in A} \sum_{t=1}^{T} \pi_t(a) = \sum_{t=1}^{T} \pi_t(a^*)$

$$
\begin{aligned}
\Gamma_T &= \sum_{a \in A} w_t(a) \\
&\geq w_t(a^*) \\
&= w_{t-1}(a^*) \cdot (1+\varepsilon)^{\pi_t(a^*)} \\
&\;\;\vdots \\
&= \underbrace{w_0(a^*)}_{1} \cdot \prod_{t=1}^{T} (1+\varepsilon)^{\pi_t(a^*)}
\end{aligned}
$$

$$\Gamma_T \geq \prod_{t=1}^{T} (1+\varepsilon)^{\pi_t(a^*)}$$

$$= (1+\varepsilon)^{\sum_{t=1}^{T} \pi_t(a^*)}$$

$$= (1+\varepsilon)^{OPT} \quad \cdots \geq \quad \boxed{\Gamma_t \geq (1+\varepsilon)^{OPT}} \quad \text{——— (1)}$$

$$\Gamma_T = \sum_{a \in A} \omega_T(a)$$

$$= \sum_{a \in A} \omega_{T-1}(a) \cdot \underbrace{(1+\varepsilon)^{\pi_t(a)}}$$

$$\leq \sum_{a \in A} \omega_{T-1}(a) \left(1 + \varepsilon \, \pi_t(a)\right)$$

$$\left[\because (1+\varepsilon)^x \leq 1 + \varepsilon x \text{ if } x \in [0,1] \right]$$

$$= \sum_{a \in A} w_{T-1}(a) + \varepsilon \underbrace{\sum_{a \in A} w_{T-1}(a)\, \pi_t(a)}$$

$$= \Gamma_{T-1} + \varepsilon \cdot \Gamma_{T-1} \underbrace{\boxed{\sum_{a \in A} \frac{w_{T-1}(a)}{\Gamma_{T-1}} \cdot \pi_t(a)}}_{\nu_T}$$

Expected utility in the t-th iteration.

$$= \Gamma_{T-1} \left(1 + \varepsilon \nu_T\right)$$

$$\vdots$$

$$\leq \Gamma_0 \prod_{t=1}^{T} \left(1 + \varepsilon \nu_t\right)$$

$$= n \prod_{t=1}^{T} \left(1 + \varepsilon \nu_t\right)$$

$$T_T \leq n \cdot \prod_{t=1}^{T} (1 + \varepsilon \nu_t) \qquad \text{--- (2)}$$

From (1) and (2)

$$(1+\varepsilon)^{OPT} \leq n \cdot \prod_{t=1}^{T} (1+\varepsilon \nu_t)$$

$$\Rightarrow OPT \ln(1+\varepsilon) \leq \ln n + \sum_{t=1}^{T} \ln(1+\varepsilon \nu_t)$$

$$\Rightarrow OPT(\varepsilon - \varepsilon^2) \leq \ln n + \sum_{t=1}^{T} \varepsilon \nu_t$$

$$\Rightarrow OPT(1-\varepsilon) \leq \frac{\ln n}{\varepsilon} + \boxed{\sum_{t=1}^{T} \nu_t}$$

Expected utility.

$\varepsilon \in [0,1]$

$\ln(1+\varepsilon) \geq \varepsilon - \varepsilon^2$

$\ln(1+\varepsilon) \leq \varepsilon$

$$\Rightarrow \quad OPT - \sum_{t=1}^{T} \nu_t \leq \varepsilon \, OPT + \frac{\ln n}{\varepsilon}$$

$$\leq \underline{\varepsilon T} + \frac{\ln n}{\varepsilon} \qquad \left[\because OPT \leq T \right]$$

$$\leq 2\sqrt{T \cdot \ln n} \qquad \left[\text{Put } \varepsilon = \sqrt{\frac{\ln n}{T}} \right]$$

$$\Rightarrow \quad \frac{1}{T}\left(OPT - \sum_{t=1}^{T} \nu_t \right) \leq 2\sqrt{\frac{\ln n}{T}}$$