Advanced Machine Learning
Learning and Games

MEHRYAR MOHRI
MOHRI@
COURANT INSTITUTE & GOOGLE RESEARCH
Outline

- Normal form games
- Nash equilibrium
- von Neumann's minimax theorem
- Correlated equilibrium
- Internal regret
Normal Form Games: Example

- Rock-Paper-Scissors.

<table>
<thead>
<tr>
<th></th>
<th>R</th>
<th>P</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>0,0</td>
<td>-1,1</td>
<td>1,-1</td>
</tr>
<tr>
<td>P</td>
<td>1,-1</td>
<td>0,0</td>
<td>-1,1</td>
</tr>
<tr>
<td>S</td>
<td>-1,1</td>
<td>1,-1</td>
<td>0,0</td>
</tr>
</tbody>
</table>
Be Truly Random

http://goo.gl/3sVFzN

Note: A truly random game of rock-paper-scissors would result in a statistical tie with each player winning, tying and losing one-third of the time. However, people are not truly random and thus can be studied and analyzed. While this computer won’t win all rounds, over time it can exploit a person’s tendencies and patterns to gain an advantage over its opponent.
Normal Form Games

- \( p \) players.

- For each player \( k \in [1, p] \):
  - set of actions (or pure strategies) \( \mathcal{A}_k \).
  - payoff function \( u_k : \prod_{k=1}^{p} \mathcal{A}_k \rightarrow \mathbb{R} \).

- Goal of each player: maximize his payoff in a repeated game.
Prisoner’s Dilemma

- Silence/Betrayal.
  - for each player, the best action is B, regardless of the other player’s action.
  - but, with (B, B), both are worse off than (S, S).

<table>
<thead>
<tr>
<th></th>
<th>S</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>2,2</td>
<td>0,3</td>
</tr>
<tr>
<td>B</td>
<td>3,0</td>
<td>1,1</td>
</tr>
</tbody>
</table>
Matching Pennies

- Player A wins when pennies match, player B otherwise.
  - other versions: penalty kick.
  - no pure strategy Nash equilibrium.

<table>
<thead>
<tr>
<th></th>
<th>H</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>H</td>
<td>1,-1</td>
<td>-1,1</td>
</tr>
<tr>
<td>T</td>
<td>-1,1</td>
<td>1,-1</td>
</tr>
</tbody>
</table>
Battle of The Sexes

- Opera/Football.
  - two pure strategy Nash equilibria.

<table>
<thead>
<tr>
<th></th>
<th>O</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>O</td>
<td>3,2</td>
<td>0,0</td>
</tr>
<tr>
<td>F</td>
<td>0,0</td>
<td>2,3</td>
</tr>
</tbody>
</table>
Mixed Strategies

- **Strategies:**
  - pure strategies: elements of $\prod_{k=1}^{p} \mathcal{A}_k$.
  - mixed strategies: elements of $\prod_{k=1}^{p} \Delta_1(\mathcal{A}_k)$.

- **Payoff:** for each player $k \in [1, p]$, when players play mixed strategies $(p_1, \ldots, p_p)$,

$$
\mathbb{E}_{a_k \sim p_k} \left[ u_k(a) \right] = \sum_{a=(a_1, \ldots, a_p)} p_1(a_1) \cdots p_p(a_p) u_k(a).
$$
Nash Equilibrium

Definition: a mixed strategy \((p_1, \ldots, p_p)\) is a (mixed) Nash equilibrium if for all \(k \in [1, p]\) and \(q_k \in \Delta_1(A_k)\),

\[
  u_k(q_k, p_{-k}) \leq u_k(p_k, p_{-k}).
\]

if for all \(k\), \(p_k\) is a pure strategy, then \((p_1, \ldots, p_p)\) is said to be a pure Nash equilibrium.
Nash Equilibrium: Examples

- Prisoner’s dilemma: (B, B) is a pure Nash equilibrium. Dominant strategy: both better off playing B regardless of the other player’s action.

- Matching Pennies: no pure Nash equilibrium; clear mixed Nash equilibrium: uniform probability for both.

- Battle of The Sexes:
  - pure Nash equilibria: both (O, O) and (F, F).
  - mixed Nash equilibria: ((2/3, 1/3), (1/3, 2/3)).
  - payoff of 2/3 for both in mixed case: less than payoffs in pure cases!
Nash’s Theorem

Theorem: any normal form game with a finite set of players and finite set of actions admits a (mixed) Nash equilibrium.
Proof

Define function \( \Phi: \prod_{k=1}^{p} \Delta_1(\mathcal{A}_k) \rightarrow \prod_{k=1}^{p} \Delta_1(\mathcal{A}_k) \) by

\[
\Phi(p_1, \ldots, p_p) = (p'_1, \ldots, p'_p)
\]

with \( \forall k \in [1, p], j \in [1, n_k], \ p'^{kj}_k = \frac{p^{kj}_k + c^{kj}_k}{1 + \sum_{j=1}^{n_k} c^{kj}_k}, \)

where \( c^{kj}_k = u_k(e_j, p_k) - u_k(p_k, p_k), \ c^{kj}_k^+ = \max(0, c^{kj}_k). \)

\( \Phi \) is a continuous function mapping from a non-empty compact convex to itself, thus, by Brouwer’s fixed-point theorem, there exists \((p_1, \ldots, p_p)\) such that

\[
\Phi(p_1, \ldots, p_p) = (p_1, \ldots, p_p).
\]
Proof

Observe that for any \( k \in [1, p] \),
\[
\sum_{j=1}^{n_k} p_k^j c_k^j = \sum_{j=1}^{n_k} p_k^j u_k(e_j, p_{-k}) - u_k(p_k, p_{-k}) = 0.
\]

Thus, there exists at least one \( j \) such that \( c_k^j = 0 = c_k^{j+} \)
with \( p_k^j > 0 \). For that \( j \),
\[
p_k^j = \frac{p_k^j}{1 + \sum_{j=1}^{n_k} c_k^{j+}} \Rightarrow 1 + \sum_{j=1}^{n_k} c_k^{j+} = 1
\]
\[
\Rightarrow c_k^{j+} = 0, \forall j
\]
\[
\Rightarrow u_k(e_j, p_{-k}) \leq u_k(p_k, p_{-k}), \forall j
\]
\[
\Rightarrow u_k(q_k, p_{-k}) \leq u_k(p_k, p_{-k}), \forall q_k.
\]
Nash Equilibrium: Problems

- Different equilibria:
  - not clear which one will be selected.
  - different payoffs.

- Circular definition.

- Finding any Nash equilibrium is a PPAD-complete (polynomial parity argument directed) problem. NP-complete problem for typical settings.

- Not a natural model of rationality if computationally hard.
Zero-Sum Games: Order of Play

- If row player plays $p$ then column player plays $q$ to minimize

$$F(p) = \min_{q \in \Delta_1(A_2)} \mathbb{E}_{a_1 \sim p, a_2 \sim q} [u_1(a)].$$

- Thus, if row player starts, he plays $p$ to maximize $F(p)$ and the payoff is

$$\max_{p \in \Delta_1(A_1)} \min_{q \in \Delta_1(A_2)} \mathbb{E}_{a_1 \sim p, a_2 \sim q} [u_1(a)].$$

- Similarly, if column player plays first, the expected payoff is

$$\min_{q \in \Delta_1(A_2)} \max_{p \in \Delta_1(A_1)} \mathbb{E}_{a_1 \sim p, a_2 \sim q} [u_1(a)].$$
von Neumann’s Theorem

Theorem (von Neumann’s minimax theorem): for any two-player zero-sum game with finite action sets,

\[
\max_{p \in \Delta_1(A_1)} \min_{q \in \Delta_1(A_2)} \mathbb{E}_{a_1 \sim p, a_2 \sim q} [u_1(a)] = \min_{q \in \Delta_1(A_2)} \max_{p \in \Delta_1(A_1)} \mathbb{E}_{a_1 \sim p, a_2 \sim q} [u_1(a)].
\]

• common value called value of the game.
• mixed Nash equilibria coincide with maximizing and minimizing pairs and they all have the same payoff.
Proof

Playing second is never worse:

$$\max_{p \in \Delta_1(A_1)} \min_{q \in \Delta_1(A_2)} \mathbb{E}[u_1(a)] \leq \min_{q \in \Delta_1(A_2)} \max_{p \in \Delta_1(A_1)} \mathbb{E}[u_1(a)].$$

- straightforward:

$$\forall p \in \Delta_1(A_1), q \in \Delta_1(A_2), \min_{q \in \Delta_1(A_2)} \mathbb{E}[u_1(a)] \leq \mathbb{E}[u_1(a)] \leq \max_{p \in \Delta_1(A_1)} \min_{q \in \Delta_1(A_2)} \mathbb{E}[u_1(a)].$$
Proof

Set-up: at reach round,

- column player selects $q_t$ using RWM.
- row player selects $p_t = \max_{p \in \Delta_1(A_1)} p^T U q_t$.

Thus, letting $T \to +\infty$ in the following completes the proof:

\[
\min_{q \in \Delta_1(A_2)} \max_{p \in \Delta_1(A_1)} \mathbb{E}_{a_1 \sim p, a_2 \sim q} [u_1(a)] = \min_{q \in \Delta_1(A_2)} \max_{p \in \Delta_1(A_1)} p^T U q
\]

\[
\leq \max_{p \in \Delta_1(A_1)} p^T U \left[ \frac{1}{T} \sum_{t=1}^T q_t \right] = \max_{p \in \Delta_1(A_1)} \frac{1}{T} \sum_{t=1}^T p^T U q_t
\]

\[
\leq \frac{1}{T} \sum_{t=1}^T \max_{p \in \Delta_1(A_1)} p^T U q_t = \frac{1}{T} \sum_{t=1}^T p_t^T U q_t = \min_{q} \frac{1}{T} \sum_{t=1}^T p_t^T U q + \frac{R_T}{T}
\]

\[
= \min_{q} \left[ \frac{1}{T} \sum_{t=1}^T p_t^T \right] U q + \frac{R_T}{T} \leq \max_{p} \min_{q} p^T U q + \frac{R_T}{T}.
\]
Notes

- Unique value: all Nash equilibria have the same payoff (less problematic than general case).
- Potentially several equilibria but no need to cooperate.
- Computationally efficient: convergence in $O\left(\sqrt{\frac{\log N}{T}}\right)$.
- Plausible explanation of how an equilibrium is reached — note that both players can play RWM.
- In general non-zero-sum games regret minimization does not lead to an equilibrium.
Yao’s Lemma

Theorem: for any two-player zero-sum game with finite action sets,

\[
\max_{p \in \Delta_1(\mathcal{A}_1)} \min_{a_2 \in \mathcal{A}_2} \mathbb{E} [u_1(a)] = \min_{q \in \Delta_1(\mathcal{A}_2)} \max_{a_1 \in \mathcal{A}_1} \mathbb{E} [u_1(a)].
\]

- consequence: for any distribution \(D\) over the inputs, the cost of a randomized algorithm is lower bounded by the minimum \(D\)-average cost of a deterministic algorithm.
General Finite Games

- Regret notion not relevant: (external) regret minimization may not lead to a Nash equilibrium.

- Notion of equilibrium: several issues related to Nash equilibria.

  ➡️ new notion of equilibrium, new notion of regret.
**Correlated Equilibrium**

**(Aumann, 1974)**

- **Definition:** consider a normal form game with $p < +\infty$ players and finite action sets $\mathcal{A}_k$, $k \in [1, p]$. Then, a probability distribution $p$ over $\prod_{k=1}^p \mathcal{A}_k$ is a correlated equilibrium if for all $k \in [1, p]$, for all $a_k \in \mathcal{A}_k$ with $p(a_k) > 0$ and all $a'_k \in \mathcal{A}_k$,

$$\sum_{a_{-k} \in \mathcal{A}_{-k}} p(a) u_k(a_k, a_{-k}) \geq \sum_{a_{-k} \in \mathcal{A}_{-k}} p(a) u_k(a'_k, a_{-k}),$$

or

$$\sum_{a_{-k} \in \mathcal{A}_{-k}} p(a_{-k} | a_k) u_k(a_k, a_{-k}) \geq \sum_{a_{-k} \in \mathcal{A}_{-k}} p(a_{-k} | a_k) u_k(a'_k, a_{-k}).$$
Think of the distribution as a correlation device.

The set of all correlated equilibria is a convex set (it is a polyhedron): defined by a system of linear inequalities, including the simplex constraints. Solution via solving an LP problem.

The set of Nash equilibria in general is not convex. It is defined by the intersection of the polyhedron of correlated equilibria and the constraints

\[ p(a) = p_1(a_1) \times \cdots \times p_p(a_p). \]
Traffic Lights

- Stop/Go.

<table>
<thead>
<tr>
<th></th>
<th>S</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>4,4</td>
<td>1,5</td>
</tr>
<tr>
<td>G</td>
<td>5,1</td>
<td>0,0</td>
</tr>
</tbody>
</table>

- Pure Nash equilibria: (S, G), (G, S). Mixed Nash equilibrium: ((1/2, 1/2), (1/2, 1/2)).

- Correlated equilibria:

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1/2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1/2</td>
<td>1/3</td>
</tr>
<tr>
<td>1/2</td>
<td>0</td>
<td>1/3</td>
</tr>
<tr>
<td>1/3</td>
<td>1/3</td>
<td>0</td>
</tr>
</tbody>
</table>
Internal Regret

Definition: internal regret, $C_{a,b}$ functions $f : \mathcal{A} \to \mathcal{A}$ leaving all actions unchanged but $a$ which is switched to $b$.

$$R_T = \sum_{t=1}^{T} \mathbb{E}_{a_t \sim p_t} [l(a_t)] - \min_{f \in C_{a,b}} \sum_{t=1}^{T} \mathbb{E}_{a_t \sim p_t} [l(f(a_t))].$$

Definition: swap regret, $C$ family of all functions $f : \mathcal{A} \to \mathcal{A}$.

$$R_T = \sum_{t=1}^{T} \mathbb{E}_{a_t \sim p_t} [l(a_t)] - \min_{f \in C} \sum_{t=1}^{T} \mathbb{E}_{a_t \sim p_t} [l(f(a_t))].$$
Swap Regret and Correlated Eq.

**Theorem**: consider a finite normal form game played repeatedly. Assume that each player follows a swap regret minimizing strategy. Then, the empirical distribution of all plays converges to a correlated equilibrium.
Definition the instantaneous regret of player $k$ at time $t$ as
\[
\hat{r}_{k,t,j,j'} = 1_{a_k,t=j} [l_k(j, a_{-k}, t) - l_k(j', a_{-k}, t)],
\]
and
\[
r_{k,t,j,j'} = p_{k,t,j} [l_k(j, a_{-k}, t) - l_k(j', a_{-k}, t)].
\]

Then, $E[\hat{r}_{k,t,j,j'} | \text{past } \& \text{ other players’ actions}] = r_{k,t,j,j'}$.

Thus, for any $(j, j')$, $(r_{k,t,j,j'} - \hat{r}_{k,t,j,j'})$ is a bounded martingale difference. By Azuma’s inequality and the Borell-Cantelli lemma, for all $k$ and $(j, j')$,
\[
\lim_{T \to +\infty} \frac{1}{T} \sum_{t=1}^{T} r_{k,t,j,j'} - \hat{r}_{k,t,j,j'} = 0 \text{ (a.s.)}.
\]

Therefore,
\[
\forall k, \limsup_{T \to +\infty} \max_{j,j'} \frac{1}{T} \sum_{t=1}^{T} \hat{r}_{k,t,j,j'} \leq 0 \text{ (a.s.)}.
\]
Swap Regret Algorithm

Theorem: there exists an algorithm with $O(\sqrt{NT \log N})$ swap regret.

$R_i$'s external regret minimization algorithm

(Blum and Mansour, 2007)
Proof

- Define for all $t \in [1, T]$ the stochastic matrix

\[ Q_t = (q_{t,i,j})_{(i,j) \in [1,N]^2} = \begin{bmatrix} q_{t,1}^\top \\ \vdots \\ q_{t,N}^\top \end{bmatrix}. \]

- Since $Q_t$ is stochastic, it admits a stationary distribution $p_t$:

\[ p_t^\top = p_t^\top Q_t \iff \forall j \in [1,N], p_{t,j} = \sum_{i=1}^N p_{t,i} q_{t,i,j} \]

- Thus,

\[ \sum_{t=1}^T p_t \cdot 1 = \sum_{j=1}^N p_{t,j} l_{t,j} = \sum_{j=1}^N \sum_{i=1}^N p_{t,i} q_{t,i,j} l_{t,j} = \sum_{i=1}^N q_{t,i} \cdot (p_{t,i} l_t) \leq \sum_{i=1}^N \min_j \sum_{t=1}^T p_{t,i} l_{t,j} + R_{T,i}. \]
Proof

Thus, for any $f : \mathcal{A} \rightarrow \mathcal{A}$,

$$\sum_{t=1}^{T} p_t \cdot 1_t \leq \sum_{i=1}^{N} \sum_{t=1}^{T} p_{t,i} l_{t,f(i)} + R_{T,i}.$$  

For RWM, $R_{T,i} = O(\sqrt{L_{\min,i} \log N})$. Thus, by Jensen’s inequality,

$$\sum_{i=1}^{N} R_{T,i} = N \frac{1}{N} \sum_{i=1}^{N} R_{T,i} \leq O \left( N \sqrt{\frac{1}{N} \sum_{i=1}^{N} L_{\min,i} \log N} \right) \quad \text{(Jensen’s ineq.)}$$

$$\leq O \left( N \sqrt{\frac{1}{N} \sum_{i=1}^{N} T \log N} \right) = O \left( \sqrt{NT \log N} \right).$$
Notes

Surprising result:

- no explicit joint distribution in the game!
- correlation induced by the empirical sequence of plays by the players.

Game matrix:

- no need to know the full matrix (which could be huge with a lot of players).
- only need to know the loss or payoff for actions taken.
Conclusion

Zero-sum finite games:
- external regret minimization algorithms (e.g., RWM).
- Nash equilibrium, value of the game reached.

General finite games:
- internal/swap regret minimization algorithms.
- correlated equilibrium, can be learned.

Questions:
- Nash equilibria.
- extensions: e.g., time selection functions (Blum and Mansour, 2007), conditional correlated equilibrium (MM and Yang, 2014).
References


References


