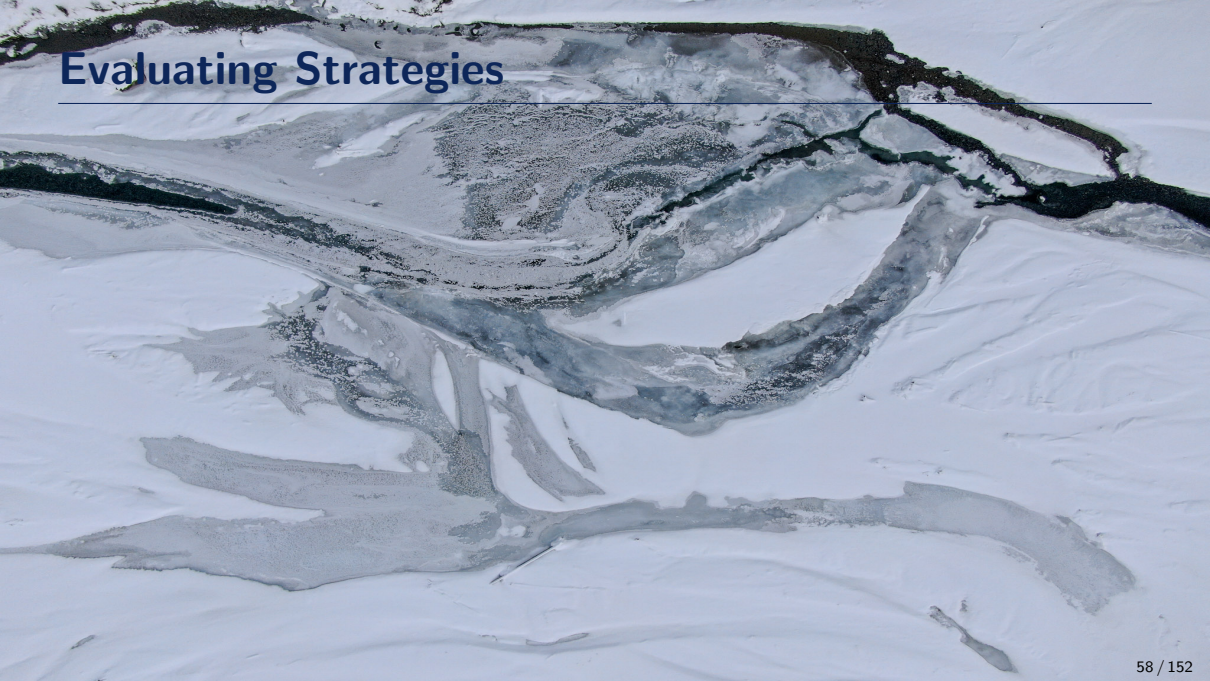
Week 3 - Zero sum & Fictitious Play

Evaluating Strategies



Loosening The Definitions

- Exact solution might be hard (large games, numerical issues, ...)
- Given a strategy profile, we still need to know how "good" it is, even if it's not exactly optimal
- The standard measures tell us how "close" to an optimal policy we are - terms of performance rather then i.e. KL divergence. [2]

[2] Timbers F, Lockhart E, Lanctot M, Schmid M, Schrittwieser J, Hubert T, Bowling M. Approximate exploitability: Learning a best response in large games. arXiv preprint arXiv:2004.09677. 2020 Apr 20.

ϵ-Nash Equilibrium

ϵ-Nash Equilibrium

A strategy profile π is said to be a ϵ-**Nash equilibrium** if for all players i and each his alternate strategy π_i', we have that:

$$u_i(\pi_i, \pi_{-i}) \geq u_i(\pi_i', \pi_{-i}) - \epsilon$$

Standard Metrics

- Player's incentive to deviate is:

$$\delta_i(\pi) = u_i(br_i(\pi_{-i}), \pi_{-i}) - u_i(\pi)$$

NashConv

$$NASHCONV(\pi) = \sum_i \delta_i(\pi)$$

Exploitability

$$EXPLOITABILITY(\pi) = NASHCONV(\pi)/n$$

ϵ-Nash Equilibrium

Policy π for which:

$$max_i \delta_i(\pi) \leq \epsilon$$

Zero-sum Saves the Day

Maximin and Minimax

We will now investigate the important relation between the players' respective maximin values $\underline{v_i}$ and $\underline{v_{-i}}$.

$$\underline{v_i} = \max_{\pi_i} \min_{\pi_{-i}} R_i(\pi_i, \pi_{-i}) \tag{10}$$

$$\underline{v_{-i}} = \max_{\pi_{-i}} \min_{\pi_i} R_{-i}(\pi_i, \pi_i) \tag{11}$$

$$\begin{aligned}
\underline{v_{-i}} &= \max_{\pi_{-i}} \min_{\pi_i} R_{-i}(\pi_i, \pi_i) \\
&= \max_{\pi_{-i}} \min_{\pi_i} -R_i(\pi_i, \pi_i) \\
&= \max_{\pi_{-i}} - \max_{\pi_i} R_i(\pi_i, \pi_i) \\
&= - \min_{\pi_{-i}} \max_{\pi_i} R_i(\pi_i, \pi_i)
\end{aligned}$$

Minimax Theorem

Theorem 12 then states a critical result — the maximin values are in balance $\underline{v_i} = -\underline{v_{-i}}$. We refer to this unique value as the game value and denote it as GV_i.

Theorem

$$\max_{\pi_i} \min_{\pi_{-i}} R_i(\pi_i, \pi_{-i}) = \min_{\pi_{-i}} \max_{\pi_i} R_i(\pi_i, \pi_i) \tag{12}$$

The minimax theorem, proven by John Von Neumann in 1928 has dramatic consequences for two player zero sum games (the proof is for both perfect and imperfect information games). Von Neumann himself later wrote "As far as I can see, there could be no theory of games on these bases without that theorem"

Nash is Maximin

Theorem

For two player zero-sum games:

$$(\pi_1, \pi_2) \text{ is Nash} \implies \pi_1 \text{ is Maximin} \wedge \pi_2 \text{ is Maximin}$$

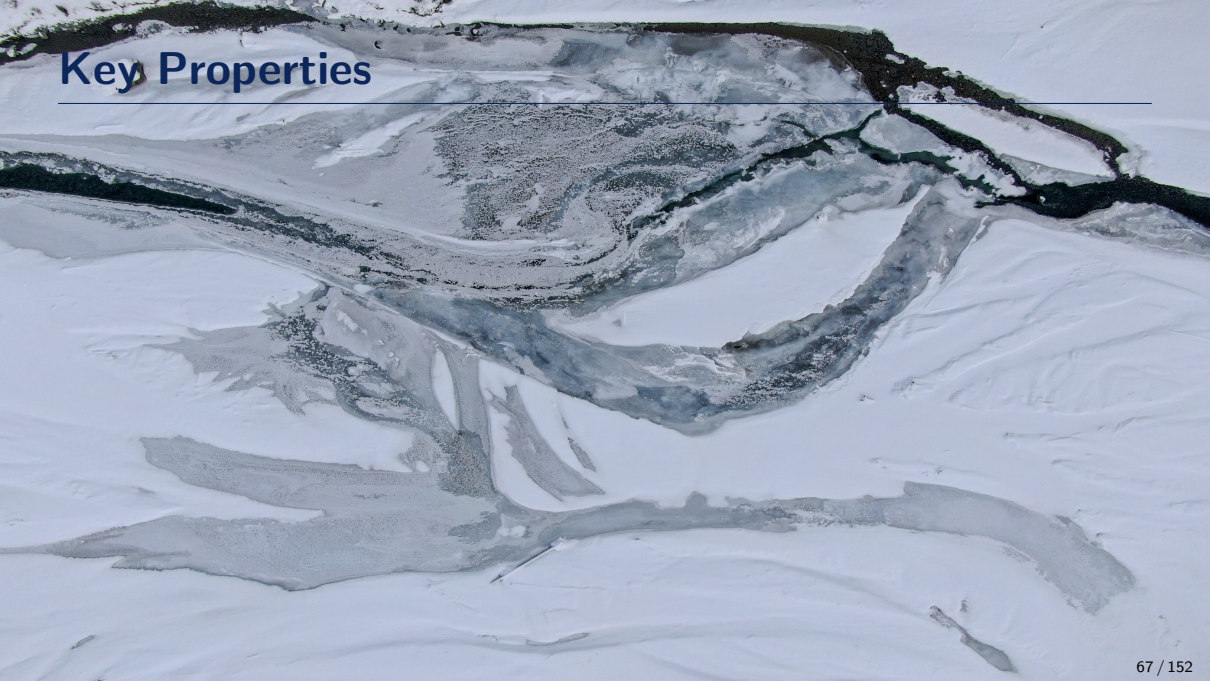
Maximin is Nash

Theorem

For two player, zero-sum games:

$$\pi_1^* \text{ is Maximin} \wedge \pi_2^* \text{ is Maximin} \implies (\pi_1^*, \pi_2^*) \text{ is Nash}$$

Key Properties



Key Motivating Property

Theorem

If we follow an optimal policy when playing both positions, the expected utility against any opponent is greater or equal to zero:

$$(\pi_i, \pi_{-i}) \in \mathit{Nash} : R_i(\pi_i, \pi'_{-i}) + R_{-i}(\pi'_i, \pi_{-i}) \geq 0 \; \forall \pi'_i, \pi'_{-i}$$

Multiplayer

- Equivalent to a two-player non-zero sum game
- How do we convert 3-player zero-sum to 2-player non-zero sum?

Selfplay Learning

Approximate Nash Equilibrium

- For learning algorithms it is hard to find exact Nash equilibrium i n the game.
- We have use some looser concept instead.

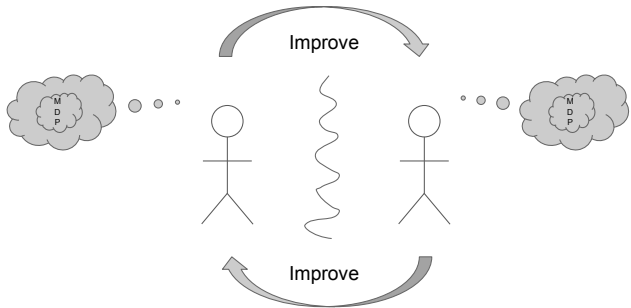
Recall the definition of the NEQ:

- The Nash equilibrium is some strategy profile σ that any player can not improve his utility by changing his strategy.

We will define ϵ-Nash equilbrium:

- The ϵ-**Nash equilibrium** is some strategy profile σ that any player can not improve his utility **more than** ϵ by changing his strategy

Fictitious Play



Fictitious Play Model of Learning

- Given player i 's belief/forecast about his opponents play, he chooses his action at time t to maximize his payoff, i.e. :

$$a_i^t \in \arg\max_{a_i \in A_i} u_i(a_i, \mu_i^t).$$

Remarks:

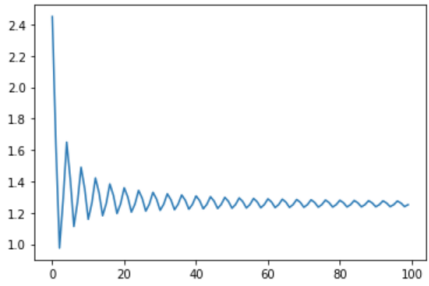
- Even though fictitious play is "belief based," it is also myopic, because players are trying to maximize current payoff without considering their future payoffs. Perhaps more importantly, they are also not learning the "true model" generating the empirical frequencies (that is, how their opponent is actually playing the game).
- In this model, every player plays a pure best response to opponents' empirical distributions.
- Not a unique rule due to multiple best responses. Traditional analysis assumes player chooses any of the pure best responses.

Matching Pennies Example

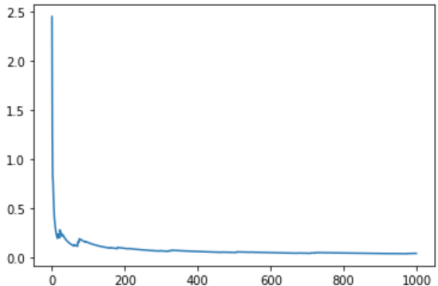
	Heads	Tails
Heads	(1, -1)	(-1, 1)
Tails	(-1, 1)	(1, -1)

Time	n_1^t	n_2^t	(Play)
0	(0,0)	(0,2)	(H,H)
1	(1,0)	(1,2)	(H,H)
2	(2,0)	(2,2)	(H,T)
3	(2,1)	(3,2)	(H,H)
4	(2,2)	(4,2)	(H,H)
5	(2,3)	(4,3)	(H,T)
6	...	...	(T,H)

Current vs Average Convergence



(a) Current



(b) Average

Convergence of Fictitious Play to Pure Strategies

- Let a^t be a sequence of strategy profiles generated by fictitious play (FP). Let us now study the asymptotic behavior of the sequence a^t, i.e., the convergence properties of the sequence a^t as $t \to \infty$.

Definition: Convergence to Pure Strategies

The sequence a^t converges to a if there exists T such that $a^t = a$ for all $t \geq T$.

Theorem

- Let a^t be a sequence of strategy profiles generated by fictitious play. If a^t converges to some a^*, then a^* is a pure strategy Nash equilibrium.
- Suppose that for some t, $a^t = a^*$, where a^* is a **strict** Nash equilibrium. Then $a^{t'} = a^*$for all $t' \geq t$.

Convergence of Fictitious Play to Mixed Strategy

Definition

The sequence a^t converges to mixed strategy profile σ in the time-average sense, if for each player i and for all actions $a_i \in A_i$, we have:

$$lim_{T\to\infty} \frac{\sum_t I(a_i^t = a_i)}{T} = \sigma(a_i)$$

Theorem

Suppose a fictitious play sequence a^t converges to σ in the time-average sense. Then σ is a Nash equilibrium.

Convergence

Fictitious play converges in the time-average sense for the game G under any of the following conditions:

- G is a two player zero-sum game.
- G is a two player nonzero-sum game where each player has at most two strategies.
- G is solvable by iterated strict dominance.
- G is an identical interest game, i.e., all players have the same payoff function.

Week 3 Homework

1. Evaluating policy pair. Given strategy pair $\pi = (\pi_1, \pi_2)$
   - 1.1 Compute Δ_i
   - 1.2 Compute ϵ
   - 1.3 Compute exploitability
2. Maximin $\Leftrightarrow$ Nash. Using the minimax theorem, prove that
   - 2.1 Nash $\Rightarrow$ Maximin
   - 2.2 Maximin $\Rightarrow$ Nash
3. Self-play Methods — fictitious self-play
   - 3.1 Implement naive self-play where you best-respond against the last strategy of the opponent (rather than the averaged one)
   - 3.2 Implement the correct self-play where you best-respond against the average strategy
   - 3.3 Plot the convergence of the exploitability of the averaged strategy for both implementations (as a function of the number of iterations)