Learning efficient equilibria in repeated games

Sam Jindani
National University of Singapore

Institut de Mathématiques de Toulouse
Workshop on learning in games
3 July 2024

Motivation

The folk theorem for infinitely repeated games creates a problem of indeterminacy:

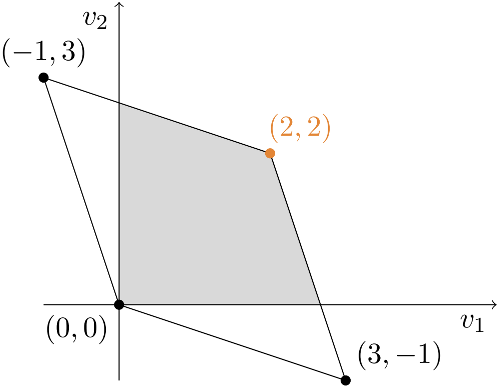
- Many payoff profiles are possible
- For a given payoff profile, many strategy profiles are possible

$\rightarrow$ What is a reasonable prediction?

Motivation

An infinitely repeated prisoner's dilemma:

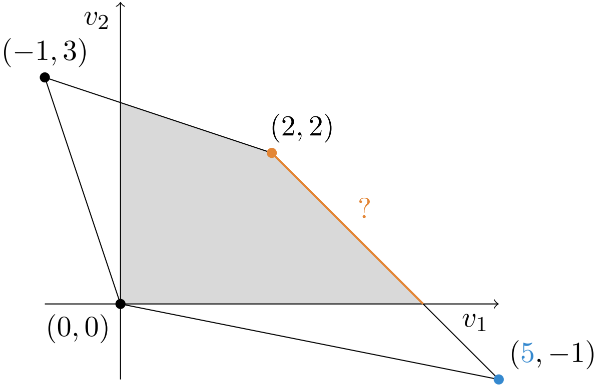
	C	D
C	$2, 2$	$-1, 3$
D	$3, -1$	$0, 0$



Motivation

An infinitely repeated asymmetric prisoner's dilemma:

	C	D
C	$2, 2$	$-1, 3$
D	$5, -1$	$0, 0$



Literature

▸ References

Learning stage-game actions: Well-known selection results for:

- risk-dominant equilibria: Kandori, Mailath and Rob 1993; Young 1993
- efficient equilibria: Robson and Vega-Redondo 1996; Arieli and Babichenko 2012; Pradelski and Young 2012; Juang and Sabourian 2021

Learning repeated-game strategies:

- *Bayesian learning:* Kalai and Lehrer 1993; Jordan 1995; Nachbar 1997; Nyarko 1998; Sandroni 1998; Nachbar 2005; Norman 2021
- *Hypothesis testing:* Foster and Young 2003.

$\rightarrow$ Suggest that players may converge to an equilibrium, but are silent on selection between equilibria.

This paper

A model of learning in two-player, infinitely repeated games.

Players act according to a non-Bayesian heuristic (Foster and Young 2003):

- form beliefs based on evidence and reject them if conflict with observed behaviour
- usually best-respond to their beliefs

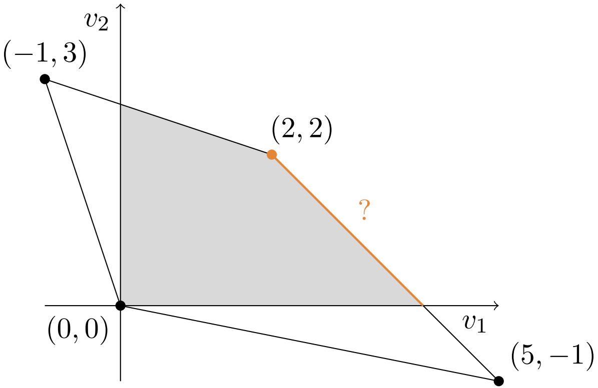
The heuristic is uncoupled (Hart and Mas-Colell 2003) and uses bounded-memory strategies (Aumann and Sorin 1989)

The model selects a subgame-perfect equilibrium with efficient payoffs.

$\rightarrow$ Provides a rationale for equilibrium selection in a learning framework

Learning and bargaining

	C	D
C	$2, 2$	$-1, 3$
D	$5, -1$	$0, 0$



This problem is reminiscent of a bargaining problem

$\rightarrow$ Intuitive specifications of the learning rule select two important bargaining solutions (the Kalai–Smorodinsky and maxmin bargaining solutions)

Stage game

Stage game: $\mathscr{G} = \langle \{1,2\}, (A_i), (u_i) \rangle$

Probability distributions on A_i: Δ_i

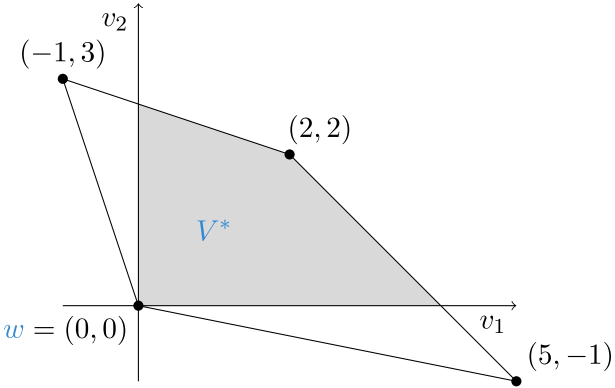
Feasible payoff profiles: V

Pure minmax payoffs: $w_i = \min_{a_j \in A_j} \max_{a_i \in A_i} u_i(a_i, a_j)$

Individually rational payoff profiles: $V^* = \{v \in V : v \gg (w_1, w_2)\}$

Example

	C	D
C	$2, 2$	$-1, 3$
D	$5, -1$	$0, 0$



Repeated game

Repeated game $\mathscr{H}$, discount factor $\delta \in (0,1)$

Players use memory-m strategies

- For any two histories whose m most recent action profiles are the same, the strategy prescribes the same (mixed) action
- Defined by a map from m-tuples of action profiles to Δ_i.

Set of memory-m strategies: $\Sigma_i = \Delta_i^{|A|^m}$

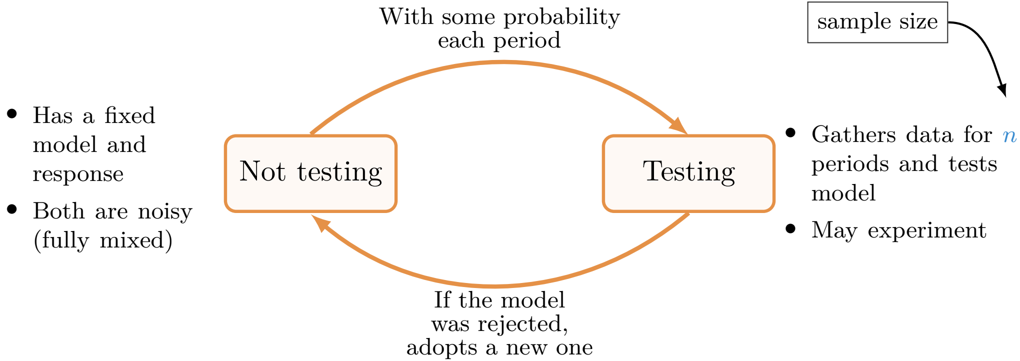
Set of strategy profiles: $\Sigma = \Sigma_1 \times \Sigma_2$

Overview

Players follow a non-Bayesian heuristic related to Foster and Young's learning by hypothesis testing (2003)

- Players do not update their beliefs each period but test them periodically

→ Inertia makes the model tractable (Foster and Young 2003; Young 2009; Arieli and Babichenko 2012; Pradelski and Young 2012)



Learning rule

noisiness

Set of strategies with probability at least ν on each action at every m-tuple: Σ_i^ν

Corresponding set of strategy profiles: $\Sigma^\nu = \Sigma_1^\nu \times \Sigma_2^\nu$

In any period, player i has

- A model $\hat{\sigma}_j \in \Sigma_j^\nu$ of her opponent's behaviour
- A response $\sigma_i \in \Sigma_i^\nu$

Each period, each player not currently testing starts a test with probability $1/n$

Testing (1/2)

Suppose player i is conducting a test

- Model: $\hat{\sigma}_j \in \Sigma_j^\nu$
- Sample: $h = (a^1, a^2, \ldots, a^n) \in A^n$

For any $h' \in A^m$ be observed in h:

- Distribution over A_j implied by i's model: $\hat{\sigma}_j(h') \in \Delta_j$
- Empirical distribution observed in the sample: $\bar{\sigma}_j(h') \in \Delta_j$

Player i rejects her model if there exists some $h' \in A^m$ observed in h such that $||\hat{\sigma}_j(h') - \bar{\sigma}_j(h')|| > \tau$.

tolerance

Testing (2/2)

Additionally, i may experiment

Average undiscounted payoff received in h: $v_i^h = \frac{1}{n}\sum_{t=1}^{n} u_i(a^t)$

Even if the model matches the observed distribution, i rejects her model with probability $\varepsilon^{f_i(v_i^h)}$

- f_i is strictly positive, strictly increasing, and continuous
- A player who received lower payoffs is more likely to experiment
- Consistent with experimental evidence about deviations from optimal behaviour (Lim and Neary 2016; Mäs and Nax 2016)

Updating

If i rejects, a new model and response are chosen according to some measure $\mu_i(h)$ on Σ^ν

- Assume $\mu_i(h)$ is diffuse: the measure of any ζ-ball in Σ^ν is at least $\mu_*(\zeta)$, where $\mu_*(\zeta) > 0$ depends only on ζ
- Interpret $\mu_i(h)$ as
    - placing most of the weight on models that are more likely given j's actions in h
    - placing most of the weight on responses that are approximate best responses to the chosen model

Definitions

Strategy profile $\sigma \in \Sigma$ is η-close to being a subgame-perfect equilibrium if there exists some subgame-perfect equilibrium $\sigma' \in \Sigma$ such that $||\sigma - \sigma'|| \leq \eta$

Any fully mixed $\sigma \in \text{int}(\Sigma)$ implies a unique limiting distribution on A

→ Average (undiscounted) payoff under this distribution: $v_i(\sigma)$

Result

Theorem 1. *Suppose that $f_1(x_1) = f_2(x_2)$ for some strongly Pareto efficient $x \in V^*$. For any $\eta \in (0,1)$, if τ is small enough (given η), if m and δ are large enough and ν is small enough (given η and τ), if ε is small enough (given η, τ, m, δ, and ν), and if n is large enough (given η, τ, m, δ, ν, and ε), then, at least $1-\eta$ of the time, players act according to strategies that*

1. *are η-close to being a subgame-perfect equilibrium and*
2. *yield average payoffs within η of x.*

Payoffs

$x \in V^*$ is strongly Pareto efficient and satisfies $f_1(x_1) = f_2(x_2)$

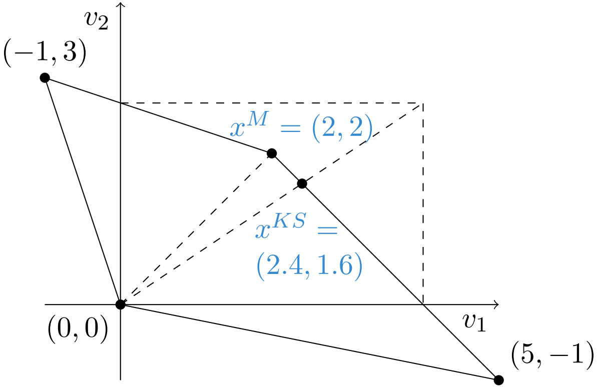
Since each f_i is strictly increasing, x is unique.

Two natural specifications:

- $f_i(x_i) = x_i$ for each i $\Rightarrow x^*$ is the maxmin bargaining solution
- $f_i(x_i) = x_i/\bar{v}_i$ for each i $\Rightarrow x^*$ is the Kalai–Smorodinsky bargaining solution

Example

	C	D
C	$2, 2$	$-1, 3$
D	$5, -1$	$0, 0$



This establishes a novel noncooperative foundation for two important bargaining solutions.

Intuition

A state is a pair $(\sigma, \hat{\sigma}) \in \Sigma^\nu \times \Sigma^\nu$

In the asymmetric prisoner's dilemma, suppose $f_i(x_i) = x_i/\bar{v}_i$

- Consider a state in which the players' models are approximately correct and the responses yield average payoffs close to $(2, 2)$
  - If n is large, the probability of a model being rejected by a test is small
  - $\bar{v}_1 = 4$ and $\bar{v}_2 = 8/3$, so the probabilities of experimenting are approximately $\varepsilon^{2/4} = \varepsilon^{0.5}$ and $\varepsilon^{2/(8/3)} = \varepsilon^{3/4}$
  - If ε is small, the former is (relatively) much larger, so it 'dominates'
- Consider a state in which the players' models are approximately correct and the responses yield average payoffs close to $(2.4, 1.6)$
  - The probability of experimenting is approximately $\varepsilon^{2.4/4} = \varepsilon^{0.6}$ and $\varepsilon^{1.6/(8/3)} = \varepsilon^{0.6}$
  - If ε is small, $\varepsilon^{0.5}$ is (relatively) much larger than $\varepsilon^{0.6}$

$\rightarrow$ A state that doesn't equalise the probabilities that players update their models is 'unstable'

Strategies

The learning rule selects strategies that are forgiving

In the asymmetric prisoner's dilemma, suppose $m = 1$ and $f_i(x_i) = x_i$

- Consider a state in which the models are approximately correct and the responses are perturbed grim triggers:
  - Each period, with probability ν play an action at random
  - Otherwise, play C iff the most recent action profile is (C, C)
- The possible states of the process are $\{CC, CD, DC, DD\}$, with stationary distribution $\to (0, 0, 0, 1)$ as $\nu \to 0$
  - Intuitively, going from CC to DD takes one experimentation but the other direction takes two

  $\to$ If ν is small and n large, the average payoffs are close to 0

$\to$ A non-forgiving equilibrium is 'unstable'

Proof outline (1/4)

A state is bad if, for some player i,

- for some $h \in A^m$, $||\sigma_i(h) - \hat{\sigma}_i(h)|| > 2\tau$ or
- $v_i(\sigma) < x_i - 2\alpha$

Choose τ and α small enough that if a state is not bad, then it is η-close to being a subgame-perfect equilibrium

A state is good if, for each player i,

- for all $h \in A^m$, $||\sigma_i(h) - \hat{\sigma}_i(h)|| \leq \tau/2$ and
- $v_i(\sigma) \geq x_i - \beta/2$.

Choose $\beta < \alpha$ such that $f^* = \min_{i=1,2} f_i(x_i - \beta) > \max_{i=1,2} f_i(x_i - \alpha) = f_*$

$\rightarrow$ Show that when ε is small and n large, the probability of going from a bad to a good state is arbitrarily higher than the probability of leaving a good state.

Proof outline (2/4)

> **2. Lemma.** *For any $\varepsilon \in (0,1)$, there exists n_1 such that, for any $n \geq n_1$, if (i) the state is bad for some player i, (ii) no player is conducting a test at t, and (iii) i begins a test at t, then the probability that the model is rejected is at least ε^{f_*}.*

- If i's model is bad, the law of large numbers implies that, for n large, the observed distribution $\bar{\sigma}_j$ will be far from the model $\hat{\sigma}_j$ with high probability
- If i's payoff is bad, the law of large numbers implies that, for n large, the sample average payoff $\bar{v}_i^h$ will be less than $x_i - \alpha$ with high probability
- In either case, the probability of rejecting the model is at least $\varepsilon^{f_i(x_i-\alpha)} \geq \varepsilon^{f_*}$

Proof outline (3/4)

3. Lemma. *For any $\varepsilon \in (0,1)$, there exists n_3 such that, for any $n \geq n_3$, if (i) the state in some period t is good, (ii) no player is conducting a test at t, and (iii) some player i begins a test at t, then the probability that the model is rejected is at most ε^{f^*}.*

- If n is large, the observed distribution $\bar{\sigma}_j$ will be close to the model $\hat{\sigma}_j$ with high probability
- If n is large, the sample average payoff $\bar{v}_i^h$ will be at least $x_i - \beta/2$ with high probability
- We can choose n so that the total probability of rejecting the model is at most $\varepsilon^{f_i(x_i-\beta)} \leq \varepsilon^{f^*}$

Proof outline (4/4)

- If ε is small, ε^{f_*} is (relatively) much larger than ε^{f^*}
- We can choose ε and n such that the probability of going from a good to a bad state is arbitrarily higher than the probability of leaving a good state
- We can use this to show that that the fraction of time spent in bad states is arbitrarily small
    - Note that just showing that going from a good state to a bad state is unlikely would be insufficient

  → We also have to rule out going from a good state to a bad state indirectly via a state that is neither good nor bad

Conclusion

This paper studies repeated interactions in which players learn independently

- Existing work looks at convergence to equilibrium, but is silent on selection between equilibria

→ This paper presents a learning rule that yields sharp predictions

The learning rule selects subgame-perfect equilibria with forgiving strategies and efficient payoffs

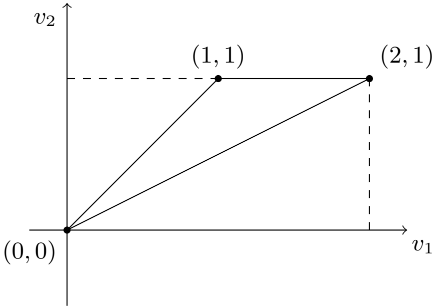
- The exact payoffs selected depend on how players update their beliefs

Additional slides

Example

Back

	C	D
C	$1, 1$	$0, 0$
D	$0, 0$	$2, 1$



If $f_i(v_i) = v_i$, then at the unique strongly Pareto efficient point $(2, 1)$, $f_1(2) \neq f_2(1)$

- We can say that on the equilibrium path each player will get at least 1
- But we can't say anything about beliefs or payoffs off the equilibrium path, because it may take low-probability experiments to reach such a state

References

Back

Arieli, I. and Y. Babichenko. 2012. 'Average testing and Pareto efficiency'. *Journal of Economic Theory* 147:2376–2398.

Aumann, R. J. and S. Sorin. 1989. 'Cooperation and bounded recall'. *Games and Economic Behavior* 1:5–39.

Foster, D. P. and H. P. Young. 2003. 'Learning, hypothesis testing, and Nash equilibrium'. *Games and Economic Behavior* 45:73–96.

Hart, S. and A. Mas-Colell. 2003. 'Uncoupled dynamics do not lead to Nash equilibrium'. *American Economic Review* 93:1830–1836.

Jordan, J. S. 1995. 'Bayesian learning in repeated games'. *Games and Economic Behavior* 9:8–20.

Juang, W.-T. and H. Sabourian. 2021. 'Rules and mutation: A theory of how efficiency and Rawlsian egalitarianism/symmetry may emerge'. Working paper.

References

Back

Kalai, E. and E. Lehrer. 1993. 'Rational learning leads to Nash equilibrium'. *Econometrica* 61:1019–1045.

Kandori, M., G. J. Mailath and R. Rob. 1993. 'Learning, mutation, and long run equilibria in games'. *Econometrica* 61:29–56.

Lim, W. and P. R. Neary. 2016. 'An experimental investigation of stochastic adjustment dynamics'. *Games and Economic Behavior* 100:208–219.

Mäs, M. and H. H. Nax. 2016. 'A behavioral study of "noise" in coordination games'. *Journal of Economic Theory* 162:195–208.

Nachbar, J. H. 1997. 'Prediction, optimization, and learning in repeated games'. *Econometrica* 65:275–309.

———. 2005. 'Beliefs in repeated games'. *Econometrica* 73:459–480.

References

Back

Norman, T. W. 2021. ‘The possibility of Bayesian learning in repeated games’. Working paper.

Nyarko, Y. 1998. ‘Bayesian learning and convergence to Nash equilibria without common priors’. *Economic Theory* 11:643–655.

Pradelski, B. S. R. and H. P. Young. 2012. ‘Learning efficient Nash equilibria in distributed systems’. *Games and Economic Behavior* 75:882–897.

Robson, A. J. and F. Vega-Redondo. 1996. ‘Efficient equilibrium selection in evolutionary games with random matching’. *Journal of Economic Theory* 70:65–92.

Sandroni, A. 1998. ‘Necessary and sufficient conditions for convergence to Nash equilibrium: The almost absolute continuity hypothesis’. *Games and Economic Behavior* 22:121–147.

References

Back

Young, H. P. 1993. 'The evolution of conventions'. *Econometrica* 61:57–84.

———. 2009. 'Learning by trial and error'. *Games and Economic Behavior* 65:626–643.