

Game Theory Project

Mateusz Pyla

Github [Repo](#)

Incremental Learning, Game Theory, and Applications 2022

Master IASD, Université Paris Dauphine-PSL

ABSTRACT

We consider a typical non-collaborative game theory setup; n (ε -rational) players want to learn strategies for a repeated (finite) stage game. The players use the standard statistical tools in order to validate their beliefs. We assume no prior knowledge, and we do not know the opponents' behaviour nor payoffs.

Dean Foster and H. Peyton Young design learning strategies that come close to subgame equilibrium play.¹

Keywords: Game Theory, Learning, Nash Equilibrium, Hypothesis test, Repeated game

1. INTRODUCTION

1.1 Notation

The paper itself is demanding so in order to enhance the clarity of its message we decided to build a small glossary.

Glossary

G (finite) Stage game. [3](#), [8](#)

G^∞ Infinitely repeated stage game. [3](#), [8](#)

n Number of players. [3](#), [8](#)

i Player i. [3](#), [8](#)

t index for time. [3](#), [8](#)

T time limit for finite game. [3](#), [8](#)

ε Given a non-negative parameter, please refer to [subsection 1.2](#). [3](#), [8](#)

λ Degree of conservation. [8](#)

σ Degree of smoothing - we talk about σ_i -optimal response. [8](#)

τ Test Tolerance. [8](#)

s Amount of data collected. [8](#)

m Memory (we say a model has at most memory m). [8](#)

ω History of play. [8](#)

ω^t Vector of ω_i^t , the actions taken in period t. [3](#), [8](#)

$\bar{\omega}^t$ Sequence of actions taken in periods from 1 to t. [8](#)

$\bar{\omega}$ Sequence of actions in the whole history. 8

$\Omega(\bar{\omega}^t)$ Set of all continuations of the initial history $\bar{\omega}^t$. 8

Δ The set of probability measures over the set of the actions.. 8

Δ^M The set of probability measures over the set of the actions accounting the memory M. 8

X Action space, i.e. $X = \prod X_i$. 8

\vec{a} Vector of responses; n-tuple of strategies. 8

\mathcal{A}_i Response space. 8

$f_i(\cdot|\bar{\omega}^t)$ Response space. 8

\mathcal{A} Space of strategies. 8

$A_i^{\sigma_i}$ Continuous mapping from i's model space to i's response space - smoothed best response function. 8

\vec{u} Vector of utility functions, $u_i : X \rightarrow R$. 8

u_i Utility function for player i. 8

u_i^1 Utility (function) from time period 1 for player i. 8

$u(\cdot)$ applied to (x_i, ϕ_i^t) - Expected utility for i, discounted to time t, from playing x_i in each period from t on, give the model ϕ_i . 8

ϕ_i Model. 8

Φ_i Space of all models having memory at most m, $\Phi_i = \prod_{j \neq i} \Delta_j^M$. 8

Φ_i Space of vectors of models, $\Phi = \prod_i \Phi_i$. 8

P mapping all players current responses to the correct models, function taking a vector of responses and returning the vector of correct models $\prod_{j \neq i} (a_j)$. 8

$A^{\vec{\sigma}}$ Mapping all players believe to their responses, function taking a vector of models (at time t) and returning the vector of distributions generating the process $A_i^{\sigma_i}(\phi_i)$. 8

H_0 Null hypothesis. 8

H_1 Alternative hypothesis. 8

ν_0 Null distribution. 8

ν True distribution. 8

$p_i^t(x_{-i}|\cdot)$ Conditional probability distribution. 8

$\beta_{i,s_i,\tau}$ Probability of making type-II error. 8

$\beta_{i,s_i,\tau}$ Least upper bound on making a type-II error when the true distribution ν is more than τ away from the null ν_0 . 8

R Rejection set. 8

ρ Discount factor. 8

h minimum probability that i plays. 8

1.2 Nature of the game

Let us suppose that n players want to play a repetitive finite-action game, which we denote as stage game G . In general we do not want to bound ourselves by the finite horizon T hence we consider G^∞ . Each player i wants to maximize their own profit hence we deal with non-cooperative game theory. Each game is in normal form meaning that the actions are performed simultaneously. For each period t , each player chooses their own action ω^t , observe the opponents' actions and record their own payoff.

Although we assumed that the game is of perfect information, this is not a strong constraint. On contrary, we **make no assumption** on prior (common) knowledge, opponents' strategies nor payoffs (including the distribution). Hence we aim at finding robust (subsection 1.4) algorithm for our problem. We call player to be ε -good predictor if they do the predictions such that the mean square error is almost surely bounded by ε as we go with t to infinity, i.e.

$$\lim_{T \rightarrow \infty} \sup \frac{1}{T} \sum_{t=1}^T (\phi_i^t - B_i(\vec{a}^t))^2 \leq \varepsilon$$

Assuming the environment as stated above, our ultimate goal is to model the strategic interactions between a set of players. In particular, we will focus on learning how to make predictions for almost rational players.

1.3 Learning

The question of learning within game theory is not precisely defined. This may be caused by the fact that we have numerous of settings, and depending on requirements one may find various learning procedures compelling. In the simplest terms, for our case we are interested in finding the algorithm that learns how to map the history of previous games to predictions for the next games. One may point out that this is a very challenging task as the world to be learnt constantly changes under our actions.

1.3.1 Motivation

The framework of multiple agents interacting with each other has found many applications, including economics, business, biology. With the advance of widely understood artificial intelligence, the game theory also plays a role in designing the framework, especially due to its compatibility with reinforcement learning² and generative methods^{3,4}.

1.4 Previous approaches

The authors of the papers are well-known experts in the field. It is highly likely that they were familiar with all the state of the art methods. They highlighted the following work on learning the strategies:

1. Trivial algorithm (simple search the space of mixed strategies) dynamically converging to Nash equilibrium of the stage game is not adequate from the global picture's perspective. There are two main reasons:
 - The process is not decentralized meaning that the learning is not performed for the individual
 - Finding such Nash equilibrium would require the relaxation of no prior assumption of the payoff knowledge.
2. For some games, using fictitious play leads to Nash equilibrium.
 - For many situations the convergence is not useful for predictions.
 - There are still some examples of games in which the algorithm diverges.
3. Knowing the appropriate Bayesian Nash equilibrium prior makes all limit points of the posterior to be also Bayesian Nash equilibrium. Authors of our paper claim that the problem of solving the equilibrium is not solved, but rather rephrased.

4. Assigning positive probability in players' prior belief to all events with actual positive probability leads to guarantee of reaching ε -equilibrium ε close enough with probability one. The authors of the main paper claim there is no robust procedure to determine such priors on other players' strategies which makes best-response strategies continuous. Instead, they propose another constraint
5. Predicting the previous already realized events, i.e. conditional regret minimization. Given the history of games one can question whether playing one action would be superior over another. There are numerous methods although they only guarantee convergence to the set of correlated equilibria.

1.5 Hypothesis testing

[TO-DO/NOT-SURE]

2. COMPREHENSION

2.1 Theorems

We firstly present our ultimate goal:

THEOREM 1 (FOSTER & YOUNG, 1998-2003). *Let G be a finite, normal-form, n -person game and let $\varepsilon \geq 0$. If the players are almost rational, use sufficiently powerful hypothesis tests with comparable amount of data, and are flexible in their adoption of new hypothesis, then at least $1 - \varepsilon$ of the time:*

I Their repeated-game strategies are ε -close to subgame perfect equilibrium,

II All players are ε -good predictors.

In order to reach the stated results we will use the intermediate results that we will use as a milestones to prove the main theorem.

LEMMA 2.1 (POWERFUL FAMILY OF TESTS). *We define the family of hypothesis tests to be powerful if there are positive tolerance functions k_i and r_i such that for every tolerance $\tau > 0$ there is at least one test in the family such that*

$$\begin{aligned}\alpha_{i,s_i} &\leq k_i(\tau)e^{-r_i(\tau)s_i} && \text{(Upper bound on type-I error)} \\ \beta_{i,s_i} &\leq k_i(\tau)e^{-r_i(\tau)s_i} && \text{(Upper bound on type-II error)}\end{aligned}$$

LEMMA 2.2 (UPPER BOUND ON TYPE-I ERROR). *Assume that null hypothesis imposes the conditional probability of observing any vector of actions at any time is at least ξ . Then:*

$$\forall \tau > 0 \quad \alpha_{i,s_i,\tau} \leq \left(1 + \frac{\tau}{\xi}\right)^{s_i} \alpha_{i,s_i}$$

COROLLARY 2.2.1 (UPPER BOUND OF TYPE-I AND TYPE-II ERRORS).

$$\beta_{i,s_i,\tau} \leq k_i(\tau)e^{-r_i(\tau)s_i/2}$$

There exists $c_i(\tau) \leq \tau$ such that

$$\alpha_{i,s_i,c_i(\tau)} \leq k_i(\tau)e^{-r_i(\tau)s_i/2}$$

LEMMA 2.3 (FAIRLY GOOD MODEL VECTOR). *We call model vector $\vec{\phi}$ to be fairly good if it is good for all responsive players. We denote it to be good if it is good for all players, meaning $|\phi_i - P_i(A^{\vec{\sigma}}(\vec{b}))| \leq \tau$ holds for player i (otherwise it is bad for i). Finally, we call it bad if it is not fairly good.*

The model vector $\vec{\phi}^t$ is fairly good at least $1 - \varepsilon$ of the time.

THEOREM 2.4 (GUARANTEES FOR REPOSPPONSES). *We assume that for the conditions described in 1, the participants use M memory hypotheses for which they employ suitable hypothesis tests with comparable amount of data.*

Moreover we assume that the engaged players have σ_i -smoothed best response functions.

Let $\varepsilon > 0$. There exist functions $\sigma(\varepsilon)$, $\tau(\varepsilon, \sigma)$ and $s(\varepsilon, \sigma, \tau)$ bounding the corresponding parameters.

Then at least $1 - \varepsilon$ of the time t

$$I \quad |a^t - A^{\vec{\sigma}}(P(\vec{a}^t))| \leq \varepsilon/2$$

$$II \quad |U_i^t(a_i^t, P_i(\vec{a}^t)) - \max_{a_i'} U_i^t(a_i', P_i(\vec{a}^t))| \leq \varepsilon \text{ for all players } i$$

$$III \quad |b_i^t - P_i(A^{\sigma_i}(\vec{\phi}^t))| \leq \varepsilon \text{ for all players } i$$

COROLLARY 2.4.1 (GUARANTEE FOR REPEATED GAME). *The second statement of the [Theorem 2.4](#) is equivalent to claiming that the repeated-game strategies are ε -close to the equilibria of the repeated game $G^\infty(\vec{u}, X)$ for at least $1 - \varepsilon$ of the time.*

The third statement imposes that all players are ε -good predictors.

2.1.1 Intuition

Although the statements above are aligned in the correct order, it is essential to gain the intuition behind the formulas.

We have two big ingredients: fixed point model and hypothesis test tools.

Firstly, fixed points are equilibria. We are certain about their existences since we have two continuous maps between compact convex spaces to each other. Those two spaces are namely, the space of vector of models $\Phi \equiv \prod_i \Phi_i$ and the space of tuples of set of responses for players, i.e. strategies \mathcal{A} .

[Theorem 2.3](#) yields that with high probability, the players will reject the hypothesis when at the bad model vectors. Once we reach the fixed point, the probability of escaping it is negligible.

By putting up the claims listed in [subsection 2.1](#) together, we build up the proof to finally reach the main theorem results.

2.2 Proofs

[TO-DO ON SATURDAY]

3. IMPLEMENTATION

[TO-DO ON SATURDAY]

4. CONCLUSION

Once again, the biggest achievement of the paper is to propose a robust way of learning the players' behaviour strategies. For each player we only assume that the player opponents' actions are public. The obtained results are satisfactory and they are based on well-grounded apparatus mathematical.

4.1 Learning Outcomes

The project was a fantastic opportunity to learn about the challenges that researchers in the field of game theory are facing. The paper is relatively old, nevertheless it has not yet had a natural successor to my best knowledge. This reflects the scale of the difficulty of the problem.

The paper was admittedly quiet tough from the mathematical perspective therefore we are satisfied with the further mathematical horizon expansion and with idea of proofs present in the field.

We had also a good occasion to get familiar with gambit, python's library for game theory.

Overall, the field of game theory serves an elegant way of modelling the interactions between the players and we are extremely delighted to learn the basics. The knowledge can be easily transferred to another problems and there are many links with other areas of science to be explored. Game theory itself is a relatively young subfield of mathematics leaving many doors still open for researchers. Aligning with my current interest, there are numerous options to combine the game theory framework and artificial intelligence. For instance, one can consider its role in the generative setting or reinforcement learning.

4.2 Further Work

Given the time constraints, we were not able to conduct all experiments we would like to do. Given more time in the future I would like to additionally:

- Increase the number of experiments and their complexity using gambit library.
- Work on more practical guarantees.

REFERENCES

- [1] Foster, D. P. and Young, H. P., “Learning, hypothesis testing, and nash equilibrium,” *Games and Economic Behavior* **45**(1), 73–96 (2003).
- [2] Nowé, A., Vrancx, P., and Hauwere, Y.-M. D., “Game theory and multi-agent reinforcement learning,” in [*Reinforcement Learning*], 441–470, Springer (2012).
- [3] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y., “Generative adversarial nets,” *Advances in neural information processing systems* **27** (2014).
- [4] Zhou, Y., Kantarcioglu, M., and Xi, B., “A survey of game theoretic approach for adversarial machine learning,” *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* **9**(3), e1259 (2019).

5. APPENDIX

For the sake of clarity of the proof we decided to slightly change the enumerations of theorems and lemmas and modify the notation to be more intuitive.

For the clarity of the statements we decided to assume that for every player i we do the prediction modelling that matters by at least ε . The original formulation of the last sentence of Theorem 1 would be "All players for whom prediction matters are ε -good predictors.", whereas for instance last point of Theorem 2.4 would be: "[...] for every player i for whom prediction matters by at least ε ."

Will use them: $G, G^\infty, \text{nit}, T, \varepsilon, \lambda, \sigma, \tau, s, m, \omega, \omega^t, \bar{\omega}, \bar{\omega}^t, \Omega(\bar{\omega}^t), \Delta, \Delta^M, X, \vec{d}, \mathcal{A}_i, f_i(\cdot|\bar{\omega}^t), \mathcal{A}, A_i^{\sigma_i}, \vec{u}, u_i, u_i^1, u(\cdot), \phi_i, \Phi_i, \Phi_i, H_0, \nu_0, \nu, H_1, p_i^t(x_{-i}|), P, A^{\vec{\sigma}}, \rho, h, \beta_{i,s_i,\tau}, \beta_{i,s_i,\tau}, R$