Fictitious Play

Ali Tarabay
Berk Bozkurt
Edwin Meriaux

April 2023

Contents

1 Introduction **2**

2 Fictitious Play Algorithm **2**
2.1 The algorithm 2
2.2 Notions of convergence 4
2.2.1 Convergence to a strategy profile 4
2.2.2 Convergence in the time-average sense 4
2.3 Main results 5
2.3.1 Important results for the analysis 5
2.3.2 Types of games in which convergence is guaranteed . . 6

3 Experiments **7**
3.1 Prisoner's Dilemma 7
3.2 Matching Pennies 8
3.3 Battle of Sexes 8
3.4 Fudenberg-Kreps 9
3.5 Rock Paper Scissors 11
3.5.1 Zero sum RPS and convergence behavior 11
3.5.2 Constant sum RPS and convergence behavior 11
3.5.3 Modified RPS and non-convergence 13

4 Discussion and Conclusion **13**

References **15**

1 Introduction

Fictitious play is a Best Response (BR) algorithm that assumes that the players will pick a strategy in line with an empirical distribution that is updated over time based on every other player's previous choices.

In each round of the game, players choose a strategy based on their current beliefs about the opponent's strategy. Once the round is played, the players then update their beliefs based on the opponent's choice for that round. The details of the algorithm are described further in Section 2.1, however the algorithm intuition is similar to that of Bayesian inference where each player assumes that their opponent's strategy follows a fixed but unknown distribution sampled from a sequence of multinomial random variables [1]. Similar to Bayesian inference, here each player has a prior set of beliefs about that unknown distribution which they then update as they get new information from each round of play.

One important limitation of fictitious play is that it assumes that the opponent is not trying to influence the players strategy choice [1]. This assumption implies that opponents will not deliberately pick a strategy to *trick* the player into believing they will play a certain a way in the future. This assumption can be rationalized by assuming that the opponents are randomly sampled from a sufficiently large but finite population (with similar behavior) where the sampled opponent player does not get the time to execute such a strategy.

While fictitious play has its limitations, which will be expanded on in Section 4, it can be used to solve the difficult problem of Nash equilibrium (NE) for a subset of games using iterative means. This is especially impressive when considering that the algorithm was first introduced in 1951 before the notion of Nash equilibrium was introduced [2, 3].

2 Fictitious Play Algorithm

2.1 The algorithm

As mentioned above, the fictitious play algorithm involves an update of the player's belief profile of their opponents strategy at each round. This is done

[1] Namely, it is assumed that every player is trying to maximize their payoff at each itearation, which might not always be a realistic assumption, especially in a learning setting.

by assigning weights to each of the opponent's strategies. These weights are initialized at some value at the start of the algorithm, and updated at each round. The simplest form of update can be seen in equation 2.1 below where one is added to the weight corresponding to the opponent's strategy choice for that round.

$$\kappa_t^i(s^{-i}) = \kappa_{t-1}^i(s^{-i}) + \mathbb{1}_{\{s_{t-1}^{-i}=s^{-i}\}} \tag{2.1}$$

From the set of weights, the players can construct their belief set on what they expect their opponent to play based on previous rounds of the game. This is simply done by normalizing the weights to generate a probability profile where each of the opponent's strategies is assigned a likelihood probability.

$$\gamma_t^i(s^{-i}) = \frac{\kappa_t^i(s^{-i})}{\sum_{\bar{s}^{-i}} \kappa_t^i(\bar{s}^{-i})} \tag{2.2}$$

Using the weights and probabilities above, an algorithm can be developed to propose a Best Response strategy at each round.

Algorithm 1 2-Player fictitious play

```
Input: The game matrix and initial weights, κ_0^1, κ_0^2
Output: Action history, final weights, κ_T^1, κ_T^2
while t < T do
    γ_t^i(s^{-i}) = κ_t^i(s^{-i}) / Σ_{s̄^{-i}} κ_t^i(s̄^{-i}), ∀i ∈ {1, 2}.
    Calculate the exp. payoff under γ_t^i for all possible actions.
    Find the exp. payoff maximizing action for both players, a_t^1, a_t^2.
    κ_{t+1}^1(a_t^2) = κ_t^1(a_t^2) + 1, κ_{t+1}^2(a_t^1) = κ_t^2(a_t^1) + 1.
    Add the profile (a_t^1, a_t^2) to the history.
end while
```

As can be seen in the pseudo-code above, the algorithm starts with a set of initial weights κ_0^1 and κ_0^2. Based on the weights, a belief set γ is generated and the expected payoff is calculated under that probability for each possible strategy. Based on the expected payoff, the player's Best Response strategy is chosen for this round. Once the round is terminated, the players update their respective weights and belief sets based on the actions played during that round. This new belief set is then used for subsequent rounds. This procedure is repeated T times until the algorithm terminates.

2.2 Notions of convergence

In this section, we present two notions of convergence to be used in the rest of the analysis.

2.2.1 Convergence to a strategy profile

First convergence notion is in regard to the rather intuitive case in which the iterative play at each step of the procedure results in an unchanged strategy profile after a certain point in iterations is reached.

Definition 1 (Convergence to a strategy profile) Let $\{s_t\}_{t\geq 0}$ be the sequence of strategy profiles during fictitious play, i.e.,

$$s_t = (a_t^i, a_t^{-i}).$$

The sequence $\{s_t\}_{t\geq 0}$ is said to be *converging to a strategy profile* s if there exists a T such that $s_t = s,\ \forall t \geq T$.

2.2.2 Convergence in the time-average sense

Second notion of convergence relates to the empirical distributions[2] getting closer to probability a distribution over the action space as time goes to infinity.

Definition 2 (Time-average convergence) The sequence $\{s_t\}_{t\geq 0}$ is said to be *converging the time-average sense* to a strategy profile s if for all i, a_i, we have

$$\lim_{T\to\infty} \frac{1}{T+1}\left[\#\texttt{times}\ \ a_t^i = a^i, t \in \{0,1,\ldots,T\}\right] = \sigma_i(a^i). \tag{2.3}$$

or equivalently,

$$\lim_{T\to\infty} \frac{1}{T+1}\kappa_t^i(a^{-i}) = \sigma_{-i}(a^{-i}), \quad \forall a^{-i}. \tag{2.4}$$

where $\sigma = (\sigma_1, \sigma_2)$ for a game with 2 players.

[2]Note that we are using the terms *empirical distribution* and *belief* over the action space interchangeably. This is not the case generally if the weights are not initialized as zero. However, as time goes to infinity we would expect the beliefs to become sufficiently close to the empirical distributions.

2.3 Main results

In this section, we introduce several propositions which are necessary for the analysis of fictitious play procedure.

2.3.1 Important results for the analysis

Given the definitions in the previous section, it is crucial make concrete connections between the convergence of the algorithm to Nash equilibria (pure and mixed) of the game. For this purpose, we introduce the following Propositions.

Proposition 1 Suppose $s_\star = (a_\star^1, a_\star^2)$ is a *strict* pure Nash equilibrium, and it is played at iteration t. Then, $s_\star$ is played at all subsequent iterations.

Proof. The belief vector for each player evolves according to the following equation for some $\beta \in (0, 1)$.

$$\gamma_{t+1}^i = (1-\beta)\gamma_t^i + \beta\delta\big(a_\star^{-i}\big).$$

As $s_\star$ is played at t and it is a strict NE, $a_\star^i$ should be the BR against γ_t^i at t. Then,

$$U_i(a_\star^i, \gamma_t^i) > U_i(a, \gamma_t^i), \quad \forall a, i.$$

Then, we have

$$\begin{aligned} U_i(a_\star^i, \gamma_{t+1}^i) &= (1-\beta)U_i(a_\star^i, \gamma_t^i) + \beta u_i(a_\star^i, a_\star^{-i}) \\ &> (1-\beta)U_i(a, \gamma_t^i) + \beta u_i(a, a_\star^{-i}) \\ &= U_i(a, \gamma_{t+1}^i), \quad \forall a, i. \end{aligned}$$

which shows that $a_\star^i$ is also the best response at iteration $t+1$, i.e., $a_\star^i = BR(\gamma_{t+1}^i)$. It is then immediate by induction that the actions played by each player will not change after t under fictitious play. □

Proposition 2 Suppose a fictitious play sequence $\{s_t\}_{t\geq 0}$ converges to σ in the time-average sense. Then σ is a Nash equilibrium.

Proof. We prove this for a 2 player game. Let $s = (a_i, a_{-i})$. Assume $\{s_t\}_{t\geq 0}$ converges to σ, but σ is not a NE. Then, for some $a_i, a_i^{'}$, we have

$$\sigma_i(a_i) > 0 \quad \text{and} \quad U_i(a_i^{'}, \sigma_{-i}) > U_i(a_i, \sigma_{-i}).$$

Let

$$2B < U_i(a_i^{'}, \sigma_{-i}) - U_i(a_i, \sigma_{-i})$$

for some $B \in \mathbb{R}$. For this value of B, we can pick a T great enough, such that for any $t \geq T$, we have

$$\sup_s |\gamma_t^i(s) - \sigma_{-i}(s)| \leq C$$

where $C = \frac{B}{\sum_a u_i(a_i, a)}$. Then, for all $t \leq T$, we have

$$\begin{aligned}
U_i(a_i, \gamma_t^i) &= \sum_a u_i(a_i, a)\gamma_t^i(a) \\
&\leq \sum_a u_i(a_i, a)\big[\sigma_{-i}(a) + C\big] \\
&= \sum_a u_i(a_i, a)\sigma_{-i}(a) + C\sum_a u_i(a_i, a) \\
&= U_i(a_i, \sigma_{-i}) + B \\
&< U_i(a_i^{'}, \sigma_{-i}) - B \\
&\leq U_i(a_i^{'}, \gamma_t^i)
\end{aligned}$$

where the last inequality holds as

$$\begin{aligned}
U_i(a_i^{'}, \sigma_{-i}) - B &= \sum_a u_i(a_i^{'}, a)\sigma_{-i}(a) - B \\
&\leq \sum_a u_i(a_i^{'}, a)\big[\gamma_t^i(a) + C\big] - B \\
&= \sum_a u_i(a_i^{'}, a)\gamma_t^i(a) + C\sum_a u_i(a_i^{'}, a) - B \\
&= U_i(a_i^{'}, \gamma_t^i) + B - B \\
&= U_i(a_i^{'}, \gamma_t^i)
\end{aligned}$$

This means that a_i is never played after t, which would result in $\gamma_t^{-i}(a_t) \to 0$ as $t \to \infty$. As the empirical distribution converges to σ_i, this would imply that $\sigma_i(a_i) = 0$, which is a contradiction. □

2.3.2 Types of games in which convergence is guaranteed

Proposition 3 Under fictitious play, empirical distributions converge for the following classes of games [1]

- 2×2 games with generic payoffs[3] [4]
- Zero (constant) sum games [5]
- Games that can be solved by iterative elimination of dominates strategies (IEDS) [6]

Convergence proofs can be found in corresponding papers cited above. IEDS case is pretty straightforward. For a 2 player game, if an action is strictly dominated, then it is never played as a best response to any possible distribution over the opponent's actions. This in turn results in the weight corresponding to that action go to zero, effectively eliminating the play of that action as iterations grow.

3 Experiments

For the sake of this Project, 7 different experiments were run. Each of these tests show different aspects of fictitious play. The examples will cover the different types of convergence previously explained. Also the different cases of Proposition 3 will be discussed. Each game contains 3 important tables or figures. The first one is the game matrix itself. This will display all necessary actions in a given game and the payout associated with each action pair. Then there is the action plot which gives what action each player played at ever time step (iteration). Finally there is the belief graph which shows the belief that each player has on the other player's actions. This is given in terms of a probability between 0-1 based on the actions taken by that other player (and the initial prior).

3.1 Prisoner's Dilemma

Prisoner's Dilemma is a simple game with one Nash Equilibrium (Table 1). That equilibrium is playing the `A` action on both players. This is a case of IEDS in all cases `A` dominates `R`. This game is also simultaneous a case of a generic game. This is due to the fact that there is a unique action with the highest payout for each player given the other player's action. If the action table for this game is observed (Fig 1a) it can be seen that each player immediately played `A`. So the convergence was immediate. This convergence will always happen for this game but the time to convergence is variable. It depends on how good the prior is. In this case, the prior led to both

[3] Games in which each player has a unique best pure strategy as a response to each of their opponent's actions.

players immediately playing the Nash but if the prior was more heavily biased towards making the players play the other action it would have taken some time for this to be corrected.

	A	R
A	-1,-1	0,-3
R	-3,0	-2,-2

Table 1: Prisoner's Dilemma game matrix

3.2 Matching Pennies

Matching Pennies is a simple example of no pure Nash Equilibrium (Table 2). It contains a singular mixed Nash strategy.

	H	T
H	1,-1	-1,1
T	-1,1	1,-1

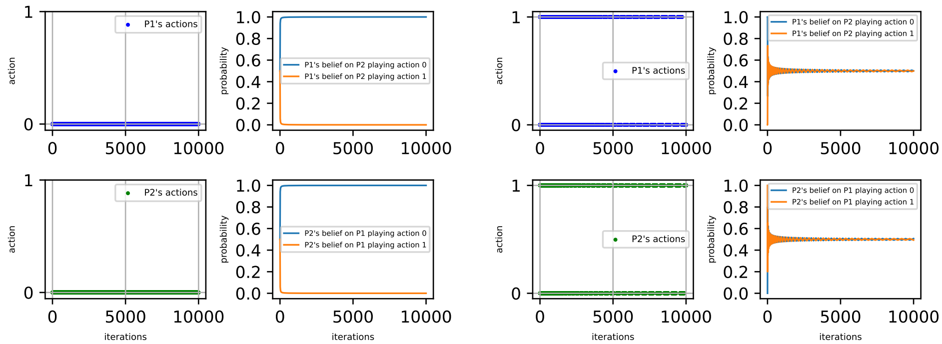
Table 2: Matching Pennies game matrix

The mixed Nash strategy is $(1/2, 1/2)$ for both players. This game is another case of a generic payoff-type game and is 2×2 so with Proposition 3 we have the guarantee of there being a Nash equilibrium. This mixed strategy can be seen in the action plots (Fig 1b). Effectively speaking each player alternates fairly quickly between both possible actions they could take. This does mean that the game does not simply converge to a singular action. As the belief graph shows there is constant altering between thinking about which belief is more likely. But it does converge in the time-averaging sense. As explained this is another equally valid form of convergence for fictitious play.

3.3 Battle of Sexes

Battle of Sexes game represents the first multi-Nash game (Table 3). Any 2×2 game should have an odd number of Nash equilibrium. This includes both mixed and pure strategies.

As can be seen in this game there are 3 Nash Equilibria. The first is both players playing the `F` action, the second is both players playing the `O` action, and finally a mixed strategy of $(2/3, 1/3)$ for player 1 and $(1/3, 2/3)$ for player 2. This game is 2×2 generic payoff game. Since this game has multiple



(a) Prisoner's dilemma simulation (b) Matching pennies simulation

Figure 1: Convergence behavior of PD and MP

	F	O
F	2,1	0,0
O	0,0	1,2

Table 3: Battle of Sexes game matrix

possible Nash solutions a weakness of fictitious play becomes visible. It can only find only Nash every time it is used. The Nash it finds is dependent on the prior it knows. In the case of the graphs shown for this game convergence to action `F` is shown (Fig 2a). This is due to the fact the prior led to this. If the prior was the opposite then the other pure Nash could be found. Finally, the mixed strategy is more complicated and will actually be shown in the example proposed by Fudenberg and Kreps. But the same reason holds for how the mixed Nash will be generated. In the case of the pure strategies, the Nash is found through straight convergence while the mixed strategy has time average convergence.

3.4 Fudenberg-Kreps

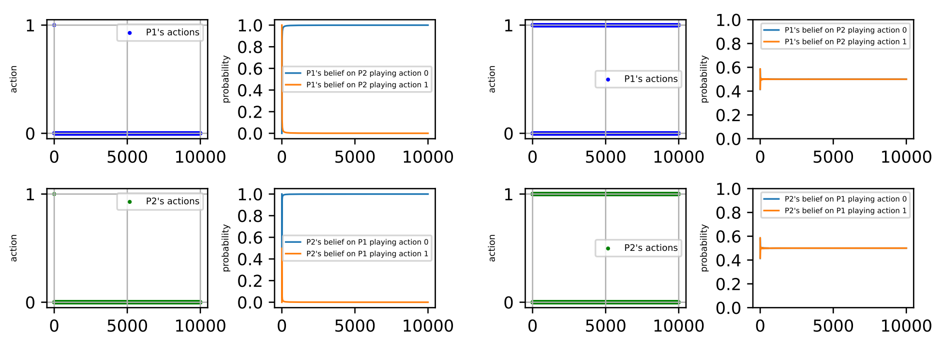
This game is similar to Battle of Sexes (Table 4). It is also a generic payoff game that have straight convergence in the pure strategies and time average convergence in the mixed strategy game. It is a good example of how finding mixed strategies when there is are multiple pure strategies can be complicated.

This game has two simple changes from Battle of Sexes. The first is that on

	F	O
F	0,0	1,1
O	1,1	0,0

Table 4: Fudenberg-Kreps game matrix

each Pure Nash each player has the same reward. One player does not favor one Nash compared to the other. The other change is that the axes on which the reward is given are different. The Nash equilibria in Battle of Sexes are all cases where both players play the same action `(F,F)` or `(O,O)`. But in this game, the equilibria come from when players play opposite actions `(F,O)` and `(O,F)`. The pure Nashes are arrived at the same way in this game compared to Battle of Sexes. This means the initial prior determines which Nash is found. What is interesting though is that the mixed Nash can be obtained by favoring both players to initially play the same action. This causes them to have the same best response which leads them to have the same actions still. This forces each player to be stuck between either possible action (as can been seen in Fig 2b). After the game is over the Nash found is the mixed strategy. An interesting remark is the fact that during this *learning* phase where the Nash was being discovered neither player actually gets any reward. They both perpetually keep ending up on the actions that return them 0 reward. This means the utility over this learning is 0 but the correct expected utility is found.



(a) Battle of Sexes simulation (b) Fudenberg-Kreps simulation

Figure 2: Convergence behavior of BoS and FK

3.5 Rock Paper Scissors

Rock Paper Scissors is a 3×3 game. This means that the guarantees from Proposition 3 do not all hold for this game type. The generic payoff case only holds for 2×2 games. This game can be played in 3 different variations. The first is a zero-sum game that converges, the second is a constant-sum game that converges, and finally the case that does not fall into those categories and does not converge. This game does not have any pure strategy and only has a mixed strategy solution of $(1/3, 1/3, 1/3)$ for all variants of the game.

3.5.1 Zero sum RPS and convergence behavior

Since this game is zero-sum it is know by proposition 3 that it must converge (Table 5). In this case the game is zero-sum because when the game matrix is looked at all the reward combos for players 1 and 2 add up to zero. Winning the game give a reward of 1, losing grants a reward of -1, and tying leads to a reward of 0. As the action plot shows each player plays a combination of all 3 actions. The amount of time spent on each action slowly increases as the value in playing the other actions slowly needs to be compensated in the best response determination (Fig 3a). The belief curve can be seen that leads it to find the mixed strategy of $(1/3, 1/3, 1/3)$ in the time average convergence.

	R	P	S
R	0,0	1,-1	-1,1
P	-1,1	0,0	1,-1
S	1,-1	-1,1	0,0

Table 5: Rock Paper Scissors zero-sum game matrix

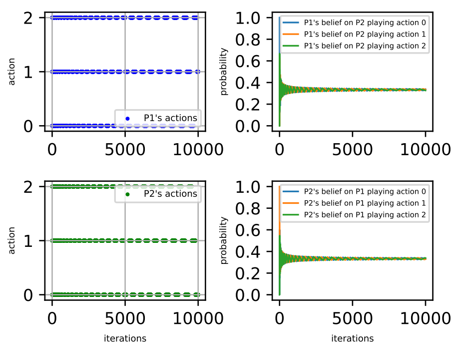
3.5.2 Constant sum RPS and convergence behavior

This game is almost the same as the zero-sum game (Table 6). The only difference is the payouts. Winning the game gives a reward of 1, losing grants a reward of -1, and tying leads to a reward of $1/2$. This makes the game constant-sum because each action pair adds to a reward of 1. The game's convergence to mixed strategy is identical as the previous game. As the action plot shows each player plays a combination of all 3 actions. The amount of time spent on each action slowly increases as the value in playing the other actions slowly needs to be compensated in the best response

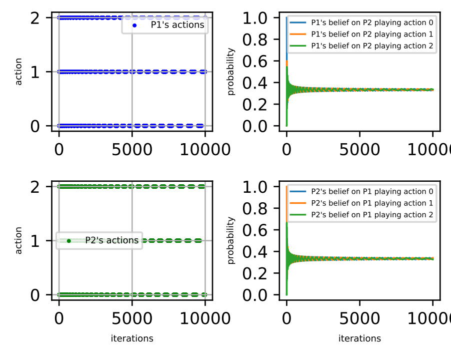
	R	P	S
R	0.5,0.5	1,0	0,1
P	0,1	0.5,0.5	1,0
S	1,0	0,1	0.5,0.5

Table 6: Rock Paper Scissors constant-sum game matrix

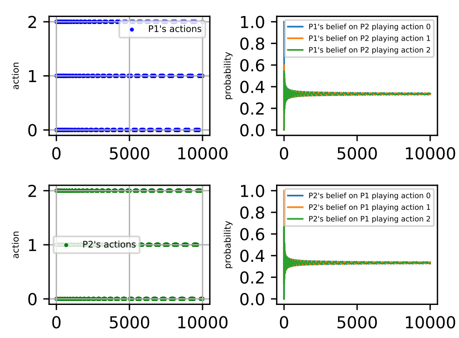
determination (Fig 3b). The belief curve can be seen that leads it to find the mixed strategy of $(1/3, 1/3, 1/3)$ in the time average convergence.



(a) Rock Paper Scissors Convergence with Zero-Sum



(b) Rock Paper Scissors Convergence with constant-sum



(c) Rock Paper Scissors no Convergence

Figure 3: Convergence behavior of Different RPS Games

	R	P	S
R	0,0	1,0	0,1
P	0,1	0,0	1,0
S	1,0	0,1	0,0

Table 7: Rock Paper Scissors (non-zero-sum) game matrix

3.5.3 Modified RPS and non-convergence

The big difference between this version of the game and the previous two is that this is neither constant or zero-sum (Table 7). Winning the game gives a reward of 1, losing grants a reward of -1, and tying leads to a reward of 0. Losing and tying in this game are equally valued. This causes a shift in the action plot. As can be seen the amount of time spent on each action grows very quickly (Fig 3c). The beliefs graph does not show any convergence even after 10,000 iterations. This game does not converge to a solution. It is not inside the bounds of Proposition 3. This is mainly due to the following fact: Switches between different strategy profiles occur in an alternating manner for each player[4]. If the ties and losses are treated equally, the amount of time that is required for each switch grows more rapidly as the weight accumulation needed for each switch is significantly greater since not penalizing a player for the loss reduces the impact of weight changes on the preferences of each player.

4 Discussion and Conclusion

Fictitious play provides a means for finding the Nash Equilibrium for a subset of games (described in Section 2.3.2). However, as it is proven to work for a certain subset of games, it might not be the most efficient method of calculating Nash equilibria given its iterative structue. For example, if the subset of interest is zero-sum games, many well studied and optimized tools (e.g., linear programming) can be used instead. In addition, one issue that is not discussed in the resources provided in the text is the rate at which fictitious play converges to a Nash equilibrium. Given that other methods can be also used as previously mentioned, having rate of convergence guarantees can be the deciding factor when choosing the algorithm.

Moreover, for these subset of games fictitious play in its most basic form

[4]For any given strategy profile sequence, the index of action switched alternate, i.e., the action played by $P1$ changes, then action played by $P2$, then again $P1$'s action etc.

necessitates that the player assumes that the environment of the game is stationary, meaning that, in an environment where players have time depending payoffs, the algorithm would fail to converge. It should also be noted that the strategy choice made in the earlier rounds of the game has a significantly higher impact on how the beliefs are evolving at each iteration[5]. It can be worthwhile to consider modified version of the algorithm to allow more recent events to be weighted more heavily to deal with this shortcoming.

Lastly, even though the literature considers fictitious play as a *learning* algorithm, it only works when all players follow the same scheme, which is unlikely in an online setting as players would have incentives to deviate from the algorithm to deceive others in various ways to maximize their cumulative payoffs in the long run. If the algorithm is merely treated as a computational tool to be used in simulations before the game is played (offline), then there already exists many powerful that would yield more computationally efficient ways of calculating equilibria as previously mentioned.

[5]At each iteration 1 is added to the weight corresponding to the action played by the opponent. Since this increase is time independent, as the number of iterations increase, each step becomes less impactful because of the accumulation of weights through the iterations.

References

[1] D. Fudenberg and D. Levine, *The theory of learning in games*, vol. 36. Sept. 1998.

[2] G. W. Brown, "Iterative solution of games by fictitious play," *Act. Anal. Prod Allocation*, vol. 13, no. 1, p. 374, 1951.

[3] U. Berger, "Brown's Original Fictitious Play," *Journal of Economic Theory*, vol. 135, no. 1, pp. 572–578, 2007.

[4] K. Miyasawa, "On the convergence of the learning process in a 2 x 2 non-zero-sum two-person game," tech. rep., PRINCETON UNIV NJ, 1961.

[5] J. Robinson, "An iterative method of solving a game," *Annals of mathematics*, pp. 296–301, 1951.

[6] J. H. Nachbar, "Evolutionary selection dynamics in games: Convergence and limit properties," *International journal of game theory*, vol. 19, pp. 59–89, 1990.