McGill University

Learning in games

Raihan Seraj

March 19, 2019

Abstract

In this project we study learning in games. Learning in game theoretic framework is interesting since it does not necessarily guarantee convergence to Nash equilibrium or any concepts of equilibria in general. In this work we analyze one of the simplest and earliest form of learning algorithm known as the fictitious play. We do a generic study of fictitious play and empirically validate some of its properties. We then outline one example where fictitious play is used for finding system optimal routings in dynamic traffic networks. This project summarizes the contributions from the first two chapters of the book "The Theory of Learning in Games" by Drew Fudenberg and David K. Levine. We then critically evaluate the paper "Fictitious play for finding system optimal routings in dynamic traffic networks" by Alfredo Garcia, Daniel Reaume and Robert L. Smith, which uses fictitious play to optimally route dynamic traffic network.

Key words: Game theory, learning, fictitious play, Nash equilibria

CONTENTS

1 INTRODUCTION

Learning in games provides an alternative explanation and interpretation to equilibrium concepts. Majority of the studies regrading non cooperative games rely on the analysis of Nash equilibria or its variants as a solution concept. Under learning framework, convergence analysis is paramount as it does not always guarantee convergence to Nash equilibria. In such cases, the understanding of equilibria or a solution concept takes an alternative view. In the learning framework, equilibria is often achieved as the long run outcome of a process in which less than fully rational players grope for optimality over time. Learning in games is often posed as a difficult problem since learning process has to be strategy proof.

Fictitious play first introduced in [1] is a popular game-theoretic model of learning in games. At each iteration in a fictitious play, the players choose a best response to their opponents' average strategy. A formal analysis of fictitious play is presented in the second chapter of the book [2]. Thus before going into deeper analysis of fictitious play, it is important to understand the fundamental assumptions behind the learning models. This motivated us to formally analyze the first two chapters of the book [2] in Sections 2 and 3 respectively. We then summarize the paper [3] which uses fictitious play to find system optimal routing in dynamic traffic network in Section 4. In section **??** we empirically analyze and assess the limitations of fictitious play as a whole. Additionally, Section **??** critiques the paper and highlights important discussions in the overall result of the paper.

2 LEARNING ASSUMPTIONS AND EQUILIBRIUM CONCEPTS

In this section we briefly summarize the contents of Chapter 1 of the book [2]. It primarily focuses on the long run properties of the learning model. To appreciate the assumptions made by the learning models, we present an example of a two player game given by the payoff matrix.

	L	R
U	1, 0	3, 2
D	2, 1	4, 0

Table 2.1: Payoff matrix for two player games.

Considering a two player game which follows a ***fixed player model*** where the players try to anticipate each others play through repeated observations of past play. Both the players need to consider their opponent's current play as well as how they influence their future play. For the learning process to work, the players need to play this game repeatedly. The payoff matrix shows that the row player has a dominant action D. The row player without considering repeated games can always play action D, the best response to which by the column player would be action L and the system converge to a Nash equilibrium in pure strategy where the row player always plays action D and the column player plays action L where both the players receive a payoff of 2 and 1 respectively. However, if the row player is patient enough and knows that the column player naively chooses its best response given its forecast of the row player's action, the row player can deviate and choose action U. This will result in the column player playing action R which will yield a payoff of 3 and 2 for the row and the column player respectively. Hence, this will lead to a higher payoff for the row player had it always played its dominant strategy. Thus, most of the learning framework abstracts from such models with the fundamental assumption that the opponent has no incentive to try to alter its future play or such alterations are small enough to be negligible.

2.1 MODELS OF PLAYER INTERACTION

Different models of player interaction can be used to embed any two or N player games, a variety of models are considered depending on how the players interact and what particular information is revealed to everyone by means of this interaction.

1. *Single pair model*: In this modelling framework a player is chosen at random at each period. At the end of the round, the actions are revealed to the everyone in the population. Under such a modelling framework it is unlikely that the same player will be chosen to play given that the population size is large. Therefore, the players will play a strategy that will maximize their current utility as a result myopic play will be optimal. Thus, the players do not tend to take strategies that will influence the future play of their opponents.
2. *Aggregate statistic model*: Pplayers in this modelling framework are randomly matched with their opponent. At the end of the play, the population aggregate is announced to everyone. Given that the size of the population is large each player will have little effect in driving the population aggregate alone by deviating. Therefore, myopic play is optimal.
3. *Random matching model*: In each period, players are randomly matched and at the end of each round the actions between the matched pair are only revealed between themselves. With a large population size, it is unlikely that the exact same player will be matched or a player will be matched with another opponent that has faced the current opponent. Therefore, the way a player will act at the current time is not going to influence the future play of its opponent.

2.2 CORRELATED EQUILIBRIUM

Game-theoretic analysis in economics deal with Nash equilibria or its refinements. Fictitious play under certain technical conditions converges to a Nash or a correlated equilibrium. We defer the discussion of these technical conditions to Section 3 however, we believe that it is important to understand the core concepts of correlated equilibrium which is more general than the well known Nash equilibrium. Although the book does include a technical definition of a correlated equilibrium, we believe that such a solution concept is best understood by means of an example which is not outlined in the book.

	Dare	Chicken out
Dare	0, 0	7, 2
Chicken out	2, 7	6, 6

Table 2.2: Payoff matrix for game of Dare and Chicken.

Consider a two player game where the payoff matrix is given by Table 2.2. The game has 2 unique Nash equilibria in pure strategies given by (Dare, Chicken out) and (Chicken out, Dare) with a payoff of $(7,2)$ and $(2,7)$ respectively. This suggests that no player will benefit more by playing other strategy other than "Chicken out" if its opponent plays "Dare". In a correlated equilibrium, each player chooses their strategy based on a common public signal. A strategy assigns an action to every possible observation that a player can make. Consider a third party which assigns strategy $(C,D),(D,C),(C,C)$ where C and D are shorthand notation for actions "Chicken out" and "Dare" respectively. Suppose this third party assigns equal probability to each of the strategy so that the probability of either choosing $(C,D),(D,C),(D,D)$ is $\frac{1}{3}$. The strategy recommended by the third party is only private information to each of the player, i.e, none of the players get to know the strategy that the third party has assigned to their opponents. Assuming Player 1 (row player) gets the strategy D as a recommendation from this third party. Player 1 is unlikely to deviate, since it knows that the other opponent must have received strategy C as the set of strategy that is recommended by the third party does not include (D,D). So Player 1 is not going to deviate from action D. Now suppose Player 2 (column player) gets a prescribed strategy C from the third party. Player 2 does not know the strategy of Player 1 which can be either C or D. So the expected payoff of Player 2 if it plays strategy C as prescribed by the third party is given by $(\frac{1}{2}\times 2)+(\frac{1}{2}\times 6)=4$. However, if Player 2 deviates and plays strategy D instead its expected payoff would be $(0+\frac{1}{2}\times 7)=3.5$ Therefore, Player 2 is unlikely to deviate from its prescribed strategy C. Since neither player has an incentive to deviate, this is a correlated equilibrium.

3 FICTITIOUS PLAY

In this section we summarize the contents of chapter 2 of the book, we then provide the mathematical definition of fictitious play for two players and outline the conditions under which fictitious play converges to a Nash or correlated equilibria.

Fictitious play or sometimes referred to as *Brown Robinson learning process* was introduced by Brown in [1] for finding the value of a zero sum game. A more generic description of fictitious play include 2 players playing a finite game repeatedly. In fictitious play, each player plays a myopic best response against the empirical distribution of its opponent's strategy. The agents believe that they are facing a stationary but unknown distribution of opponent's strategy.

3.1 TWO PLAYER FICTITIOUS PLAY

In a two player game, each player maintains an exogenous weight function at each time given by K_t^i. Let the weight function of player i at time $t = 0$ be given by K_0^i. Let S_t^{-i} be a random variable which defines the strategy played by the opponent at time t. The weight function of player i follows the recursive update rule given by,

$$K_t^i(s^{-i}) = K_{t-1}^i(s^{-i}) + \begin{cases} 1 & \text{if } S_{t-1}^{-i} = s^{-i} \\ 0 & \text{otherwise.} \end{cases}$$

The probability that player i assigns to player $-i$ at time t is given by

$$\gamma_t^i(s^{-i}) = \frac{K_t^i(s^{-i})}{\sum_{\tilde{s}^{-i} \in s^{-i}} K_t^i(\tilde{s}^{-i})}.$$

Here, γ_t^i is also known as the player i's assessment of the distribution of past play of its opponent. Fictitious play is therefore defined as any rule $\rho_t^i(\gamma_t^i) \in BR^i(\gamma_t^i)$ where BR is the best response. In the case where best response is used as a rule, we can see that the behavior prescribed by fictitious play is a discontinuous function of player's assessment.

One important difference between fictitious play and if a player always takes a best response to the action seen in the previous play is that in a fictitious play, players maintain a prior over the distribution of their opponents' strategies which gets updated periodically. Often compared to the best response, fictitious play moves slowly over time as the ratio of new observations to old ones becomes small.

3.2 ASYMPTOTIC BEHAVIOR OF FICTITIOUS PLAY

Fictitious play undergoes certain asymptotic behavior which describes when a particular play converges to a Nash. These properties are outlined as follows:

1. If a strategy S is a strict Nash equilibrium which is played at time t of the process, then it is played in the subsequent process. In other words strict Nash equilibria are absorbing to the process of fictitious play.
2. Any pure strategy resulting from steady state of fictitious play, must be a Nash equilibrium.
3. Fictitious play cannot converge to a pure strategy profile of a game all of whose equilibria are mixed.
4. If the empirical distribution D_t^i of each player choices converge, the strategy profile corresponding to the product of these distributions is a Nash equilibrium.
5. In a zero um game, the empirical distribution generated by fictitious play must converge to Nash equilibria.

One of the limitations of fictitious play is that players only keep track of the frequency of opponent's play. This does not take into account any cycles within the play since the conditional probabilities are not tracked. Therefore, a notion of consistency is used which if followed ensures that fictitious play successfully learns the frequency distribution. This learned frequency distribution would yield the same utility that would be achieved if the players knew this distribution in advance. Thus, the maximum utility that is achieved against the empirical distribution is given by

$$\hat{U}_t^i = \max_{\sigma^i} u^i(\sigma^i, D_t^{-i}),$$

where σ^i is some profile of player i and D_t^{-i} is a distribution over player i's opponent.
The time average of player i's realized payoff is given by

$$\bar{U}_t^i = \frac{1}{t} \sum_{\tau=1}^{t} u^i(s_\tau^i, s_\tau^{-i}),$$

where s_τ^i is the action played by player i at time $\tau = t$.

A fictitious play is said to be ε-consistent along a history if there exist a T such that for any $t \geq T$, $\bar{U}_t^i + \varepsilon \geq \hat{U}_t^i$. Thus consistency means that a player does well as it might if it knew the frequency of its opponent's play in advance. If for some reason $\bar{U}_t^i$ remains less than $\hat{U}_t^i$ player i should eventually realize that something is wrong with its model of the environment and adjust its play accordingly. Whenever a player has only two strategies, if the empirical distribution D_t^i is consistent, then the converged solution reaches a correlated equilibria. In Section **??** we perform an empirical study in order to verify these asymptotic behavior of fictitious play.

4 FICTITIOUS PLAY FOR FINDING SYSTEM OPTIMAL ROUTING

Applications of fictitious play in network routing problems has been observed in the literature. In this section, we outline an example that appeared in the work of Alfredo Garcia on fictitious play for finding system optimal routings in dynamic traffic networks [3].

The model considers the average trip time experienced in the network as the utility of the players, and fictitious play is used to compute system optimal routings. The repeated play of the fictitious game is shown to converge to a local optimal routing value. A large-scale computational test is also included in the paper.

The network routing problem appears in the engineering perspective as an optimization problem. However, when the user optimality is considered, and decision is given in a decentralized structure, the problem evolves into a multi-agent decision making problem. Focusing on the average trip time experienced in the network to be minimized, the paper demonstrates that an equilibrium is reached with fictitious play.

Fictitious play can be interpreted as an iterative routing-assignment algorithm in dynamic traffic network game problem. During this procedure, at each step, for each player, time dependent shortest paths are computed given that other players' decisions are distributed according to the historical frequency of their earlier routing decisions.

One of the main limitations of other iterative techniques is that they cannot guarantee convergence. Even though convergence of fictitious play has been established when players share a common objective function, the resulting routing decisions are not system optimal. These decisions are locally optimal in the sense that each of the players cannot further reduce their average time by choosing any other routing protocols.

4.1 PROPOSED MODEL

In the setting of traffic network game, vehicles are denoted as players and each players' payoffs are computed through an assignment mapping. This assignment mapping calculates the travel times given the routing decisions for all players. A routing map assigns each vehicle a time-dependent shortest path from its origin to destination given the route choices of all other vehicles. The routing map is a best response for each player to the routing decisions of the other players.

Using the same notation used in the paper, traffic network game is defined where $N = \{1, 2, \ldots, n\}$ is the index set of vehicles and for every $i \in N$ there is a finite set R_i of possible routes to take. Set of all vehicles' feasible route choices becomes $R = \prod_{i \in N} R_i$.

Assignment mapping is denoted by $A : R \to \mathscr{R}^n$.

$\Delta_i = \left\{ f_i : R_i \to [0,1] \text{ such that } \sum_{r_i \in R_i} f^i(r_i) = 1 \right\}$ is the set of mixed routing decisions.

$A_i\left(f\right) = \sum_{r \in R} A_i(r) \cdot f_1 \cdot f_1(r_1) \cdot f_1(r_1) \cdots f_1(r_1)$ is the expected value of total travel time for vehicle i when all vehicles adopt mixed routing strategy f.

Routing assignment f is Nash Equilibrium iff for every player $i \in N$, the probabilities assigned to the routes yield its minimum expected total travel time, provided that f_i^* the mixed routing choice of all other vehicles is held fixed.

$f_i^* \in \underset{f_i \in \Delta_i}{\operatorname{argmax}} A_i(f_i, f_{-i}^*)$

Monderer and Shapley [5] have demonstrated that when players share a common objective function, fictitious play does converge. In order to use this result, the paper defines a new game by aver-

aging the payoffs of all players so that it becomes a common payoff function.

$$U(f) = \sum_{i \in N} \frac{A_i(f)}{n}$$

Algorithm

1. Set $t = 0$. Pick an initial pure routing strategy .
2. Compute a best reply for each $i \in N$:
   $r_i(t+1) \in \text{argmin}_{r_i \in R_i} \left[U\left(r_i, f_t^{-i}\right) \right]$
3. Update historical frequencies of route choices, f_t.
4. If $f_{t+1} - f_t \leq \varepsilon$ then Stop, otherwise, set $t = t+1$ and go to 2.

4.2 Main results

The algorithm has been implemented using Alliance software package. Including travel patterns, 16500 vehicles are simulated. A comparison of the performance of Alliance to SAVaNT [4], which is an iterative routing-assignment procedure intended to compute user optimal routings, is made. The procedure is perceived to provide (whenever it converges) reasonably good routings in terms of average trip time in the network.

For the simulation, three classes of vehicles are defined to account for the impact of different market penetration levels . Class 1, consisted of those vehicles following the free flow shortest path. Class 2, consisted of those vehicles that perform a periodic update of the free flow shortest paths, and finally, Class 3 vehicles are using the fictitious play algorithm implemented with the Alliance package. The initial routing for all vehicles are denoted as the shortest paths under free flow conditions. The first test conducted with Class 3 vehicles account for 25% of the total number of vehicles, to simulate high market penetration. Alliance computes routings to have similar average trip time to SAVaNT. Moreover Alliance requires fewer iterations and less computational time.

Test 1: Average trip time (min)				
	Class 1	Class 2	Class3	#Iteration
Alliance	8.85988	8.85677	8.71779	14
SAVaNT	8.87245	8.84866	8.68266	34

The second test is conducted with Class 3 vehicles account for 5% of the total number of vehicles to simulate a low market penetration. Travel time of Class 3 vehicles is decreased in this scenario. Results also show the computational load is less than SAVaNT.

Test 1: Average trip time (min)				
	Class 1	Class 2	Class3	#Iteration
Alliance	17.30430	15.59930	17.21905	20
SAVaNT	17.49250	15.49160	17.39240	68

Following graph shows the decrease in travel time of Class 3 vehicles, and indicates the speed of convergence.

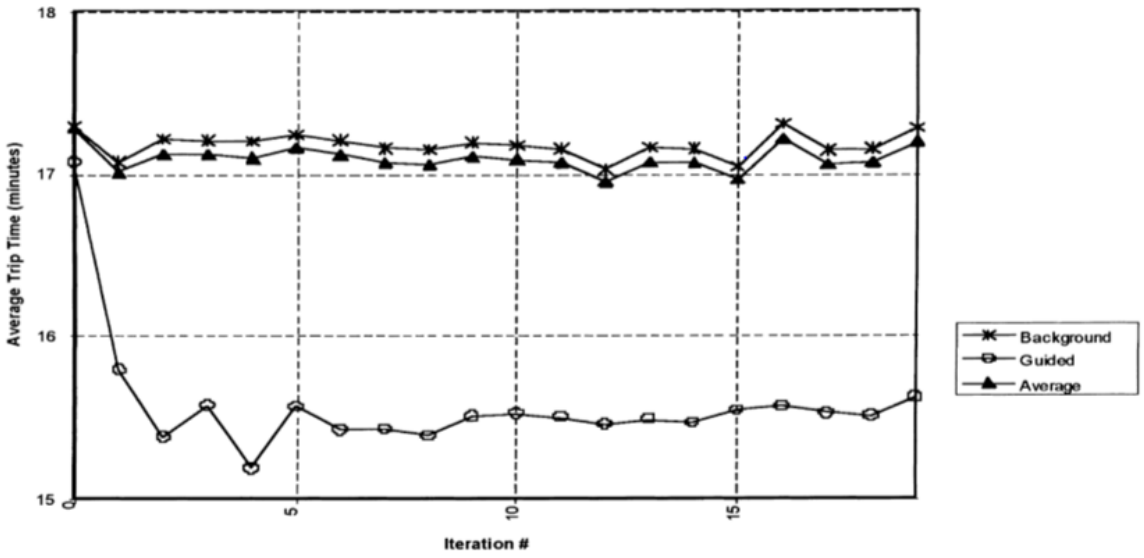

Fig. 1. Illustration of speed of convergence.

4.3 Discussions

This paper uses theory of learning in games in dynamic traffic game to compute system optimal routings and presents a comparison with other iterative routing-assignment procedure. Considering the shared payoff, the problem can be formulated as a centralized optimization problem. However, the paper presents an idea of decentralizing the optimization so as to parallelize the process. Instead of a main solver to compute the optimal routing, it distributes the processing to each agent, so that after few iterations optimal routing is achieved. The comparison with another algorithm is also given that shows computational efficiency and the decrease of number of iterations achieved using fictitious play.

5 Critical Analysis

In this section we perform a critical analysis of the concepts that we have discussed in earlier sections. In the first part of this section, we perform an empirical study of fictitious play demonstrated in simple two player games. We aim to validate the properties exhibited by such an algorithm. In the second part of this section, we critique the paper that used fictitious play to find optimal routing decisions in a dynamic traffic network.

5.1 Empirical Study

Section 3 demonstrates certain asymptotic properties of fictitious play. In our first empirical study, we consider a two player game with the following payoff

	L	R
U	1, 0	3, 2
D	2, 1	4, 0

Table 5.1: Payoff matrix for two player games.

We implemented fictitious play where the prior distribution of each player on their opponent's frequency of play is drawn from a Dirichlet distribution. At each time, the players take a best response based on its belief or assessment of its opponent's play. For this game we run fictitious play for $T = 500$ We observe that for the above game, Nash equilibria in pure strategy is given by action (D, L). Using fictitious play, we obtain the following results.

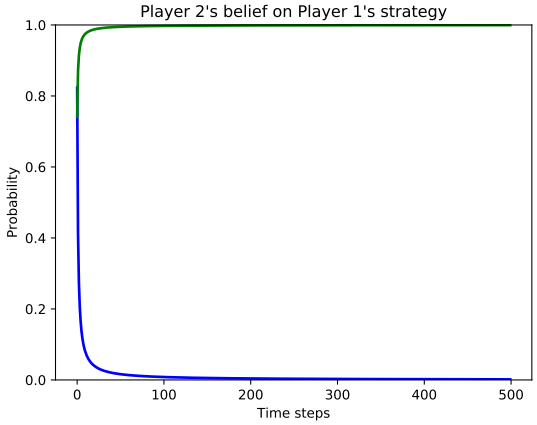

Figure 5.1

Considering the row player as Player 1 and the column player as Player 2, the result in Figure 5.1 demonstrates the assessment of Player 1's strategy made by Player 2. The blue line outlines Player 2's belief that Player 1 will be playing action U at each time instant. The green line shows Player 2's belief that Player 1 would be playing action D. Thus through repeated play, Player 2 determines that Player 1 is going to play action D with probability 1. As a result, Player 2 plays action L as a best response to its assessment. The game therefore converges to Nash equilibrium (DL) in pure strategy. Our analysis confirms, property 1 and 2 outlined in Section 3. Note that the game above also have Nash equilibrium in mixed strategy. However our analysis show that in such cases, fictitious play only converges to a Nash equilibrium to pure strategy (if it exists)

In order to validate property $(3, 4, 5)$ describing the asymptotic behavior of fictitious play, we use

fictitious play in matching pennies which is a two player symmetric zero sum game. Let the row player be denoted as Player 1 and the column player as Player 2.

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

Table 5.2: Payoff matrix for matching pennies.

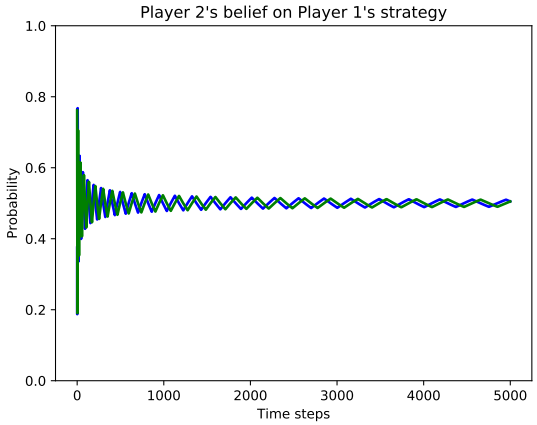

Figure 5.2

The results in Figure 5.2 shows that the assessment made by Player 2 on Player 1's strategy converges to an empirical distribution of $(\frac{1}{2}, \frac{1}{2})$. Since this game has no Nash equilibria in pure strategies. The game has only one Nash equilibrium in mixed strategy, where both the players will play each action with equal probability. The obtain result is also consistent and we show that fictitious play manage to converge to the only Nash equilibria present in mixed strategy, this validates property $(3,4,5)$ outlined in Section 3. For the game mentioned in Table 5.2 we run fictitious play for $T = 5000$, the blue line in Figure 5.2 shows Players 2's assessment that Player 1 will play action T and the green line shows Player 2's assessment that Player 1 will play action H.

5.2 Critique of the presented example

Considering the use of fictitious play for optimal routing problem, the main result highlighted in the paper was convergence obtained by fictitious play using common payoff function. Since the network routing assignment is an optimization problem, it is justified to use multi-agent system analysis. The proposed model distributes the decision making criteria to each player and assigns the payoff of each player based on the round trip time. Assigning a common payoff function to each player is often not practical. Therefore it would be more practical where each agents have an independent payoff function (round trip time) which they try to minimize. Thus, a better modification of the payoff function can be made, where players try to minimize their own payoff using game theoretic approach. The proposed algorithm only compares with a single iterative model which they named as SAVaNT. However, they did not mention whether SAVaNT is a current state of the art, neither have they performed any analysis with a range of other iterative procedures so that a clear understanding on the robustness of the proposed methodology is made.

REFERENCES

[1] G. W. Brown. Iterative solution of games by fictitious play. *Activity analysis of production and allocation*, 13(1):374–376, 1951.

[2] D. Fudenberg and D. Levine. The theory of learning in games. *The MIT Press, Cambridge, MA*, 13, 1998.

[3] A. Garcia, D. Reaume, and R. L. Smith. Fictitious play for finding system optimal routings in dynamic traffic networks. *Transportation Research Part B: Methodological*, 34(2):147–156, 2000.

[4] D. E. Kaufman, R. L. Smith, and K. E. Wunderlich. Dynamic user-equilibrium properties of fixed points in iterative routing/assignment methods. *IVHS Technical Report*, 92-12, 1992.

[5] D. Monderer and L. S. Shapley. Fictitious play property for games with identical interests. *Journal of economic theory*, 68(1):258–265, 1996.