Fictitious Play and Correlated Learning

K. Sinha
Department of Systems and Information Engineering
University of Virginia

October 27, 2004
Problem Formulation

- Finite set of players: \( N = \{ i : i = 1, 2, \ldots, n \} \)
- Finite set of actions for each player \( i \): \( A_i \)
- Joint Action Set \( A \equiv A_1 \times A_2 \times \ldots \times A_n \)
  - \( A \) is finite with cardinality \( m = |A| \).
    * \( A \) can be enumerated as \( \{ a(1), a(2), \ldots, a(k), \ldots, a(m) \} \).
  - The \( k \)-th joint action can be represented as \( a(k) = (a_i(k), a_{-i}(k)) \)
    * \( a_i(k) \in A_i \) is the action used by players \( i \) in \( a(k) \)
    * \( a_{-i}(k) \in \prod_{j \neq i} A_j \) is the action profile associated with all other players in the joint action \( a(k) \)
- System level cost (disutility) under \( a(k) \in A \): \( U(a(k)) \).
  - Cost is perceived equally by all players (Game of Identical Interests).
  - Every player is a cost minimizer!
Motivation

Consider the optimization problem,

\[
\min U(x)
\]

subject to \( x = (x_1, x_2, \ldots, x_n) \in A_1 \times A_2 \times \ldots \times A_n \).

- Every optimal solution of the optimization problem defined above corresponds to a pure strategy Nash Equilibrium of the corresponding game of identical interests.

- Every pure strategy equilibrium of the game with identical interests corresponds to a local minimum of the optimization problem defined above.
  - What about mixed-strategy equilibria?

It may be worthwhile to explore ways to compute equilibria ...
Probability distributions over $A_i$

Let the set of mixed strategies of player $i \in N$ be

$$\Lambda_i = \left\{ \lambda_i : A_i \mapsto [0, 1] : \sum_{\alpha \in A_i} \lambda_i[\alpha] = 1 \right\}.$$

- Interpretation:
  
  $\lambda_i[\alpha]$ is probability of player $i$ choosing action $\alpha$ from his/her action set $A_i$

- A pure strategy $\alpha \in A_i$ is an extreme point of $\Lambda_i$ i.e., where $\lambda_i[\alpha] = 1$
Empirical Frequency Distributions

Suppose joint actions $\{a^0, a^1, \ldots, a^{t-1}\} \equiv H^t \subset A$ have been observed in $t$ instances of the finite game.

- $H^t$ will also be referred to as the Play History.

- Define

$$\lambda^t_i[\alpha] = \frac{\text{# times action } \alpha \in A_i \text{ is used by player } i \text{ in } H^t}{t}.$$
Fictitious Play (FP)

In FP a player $i \in N$ believes that (a) every player $j \neq i$ plays according to a \textit{stationary} $\lambda_j \in \Lambda_j$ and (b) that these distributions are \textit{independent}.

Definitions

- The (expected) cost of $\lambda \in \prod_{i=1}^{n} \Lambda_i$ is

$$ U(\lambda) = \sum_{k=1}^{m} \left[ \prod_{i=1}^{n} \lambda_i[a_i(k)] \right] U(a(k)) $$

- A best response for Player $i$, to the other players strategies, is defined to be an \textit{action} $\beta_{i}^{fp}(\lambda) \in A_i$ such that

$$ \beta_{i}^{fp}(\lambda) \in B_{i}^{fp}(\lambda) = \arg \min_{\alpha \in A_i} U(\alpha, \lambda_{-i}) $$

  - By $U(\alpha, \lambda_{-i})$ we imply that player $i$’s probability distribution on $A_i$ concentrates all its weight on action $\alpha \in A_i$. 
FP Algorithm


**(Initialization)**

- Start with $t = 0$ and initial play history $H^0 = \{a^0\} \subset A$.

**(Iteration $t \geq 1$)**

- Given $H^{t-1}$ and associated $\lambda^{t-1}_i$,s,
  
  i. each player $i$ computes its individual best response $\beta^{fp}_i(\lambda^{t-1})$,
  
  ii. construct the *aggregate* best response
  \[ x^t = (\beta^{fp}_1(\lambda^{t-1}), \beta^{fp}_2(\lambda^{t-1}), \ldots, \beta^{fp}_n(\lambda^{t-1})) \in A, \]
  
  iii. add $x^t$ to the play history:
  \[ H^t = (H^{t-1}, x^t). \]
Evolution of $\lambda_i^t$ under Fictitious Play

If $\gamma_t = \frac{1}{t+1}$ then

$$\lambda_i^{t+1}[\alpha] = \begin{cases} 
(1 - \gamma_t)\lambda_i^t[\alpha] + \gamma_t, & \text{if } \alpha = \beta_i^f p(\lambda^t), \\
(1 - \gamma_t)\lambda_i^t[\alpha], & \text{otherwise.}
\end{cases}$$
Convergence - I

A game has the fictitious play property if every limit point of every sequence $\lambda^t$ generated by the FP algorithm is a Nash equilibrium (NE) of the game.

- (Monderer and Shapley, 1996) "... fictitious play is said to converge in beliefs to equilibrium if the sequence of beliefs $\lambda^t = (\lambda_{1}^t, \lambda_{2}^t, \ldots, \lambda_{n}^t)$ is asymptotically as close as we desire to the set of equilibria.”
Convergence - II

**Proposition 1** [Fudenberg & Levine 1991]
Under FP, if the empirical distributions $\lambda^t_i$ over each player’s choices converge, the strategy profile corresponding to the product of these empirical distributions is a Nash Equilibrium (NE).

**Proposition 2** [Fudenberg & Levine 1991] (i) If a strict NE, say $s$, is played at time $t$ in the process of FP then $s$ is played at all subsequent times.
(ii) Any pure-strategy steady state of FP must be a NE.

**Proof of Proposition 2 (i):** Note that as the expected value is linear in probabilities $U(s_i, \lambda^{t+1}) = (1 - \gamma_t) U(s_i, \lambda^t) + \gamma_t U(s_i, s_{-i})$. This implies $s$ is a strict best response for $\lambda^{t+1}$.

---

*The Theory of Learning in Games, MIT Press*
Maximizing Payoffs Example - I

An Example [Peyton-Young, 1998]: The Row and Column Players are interested in maximizing their payoffs using FP ...

<table>
<thead>
<tr>
<th>Action</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>((1, \sqrt{2}))</td>
<td>((0, 0))</td>
</tr>
<tr>
<td>2</td>
<td>((0, 0))</td>
<td>((\sqrt{2}, 1))</td>
</tr>
</tbody>
</table>

Mixed Strategy Profile Corresponding to NE \(\{(\lambda_1[1], \lambda_1[2]), (\lambda_2[1], \lambda_2[2])\}\):

\[
\left\{ \begin{pmatrix} \frac{1}{1 + \sqrt{2}}, \frac{\sqrt{2}}{1 + \sqrt{2}} \\ 1 + \sqrt{2} \end{pmatrix}, \begin{pmatrix} \frac{\sqrt{2}}{1 + \sqrt{2}}, \frac{1}{1 + \sqrt{2}} \\ 1 + \sqrt{2} \end{pmatrix} \right\}
\]

\(^a\) Individual Strategy and Social Structure, Princeton University Press
Maximizing Payoffs Example - II

- Player 1’s best reply is row 1 if \( \frac{\lambda_2^t[1]}{1-\lambda_2^t[1]} > \sqrt{2} \)

- Player 2’s best reply is column 1 if \( \frac{\lambda_1^t[1]}{1-\lambda_1^t[1]} > \frac{1}{\sqrt{2}} \)

- If \( \lambda_1^t[1] + \lambda_2^t[1] = 1 \) then \( \lambda_1^{t+1}[1] + \lambda_2^{t+1}[1] = 1 \)

If the game starts from any of the two pure-strategy NE the game stays there.
Now consider the game starting off with a miscoordination say, \( H^0 = \{(1, 2)\} \)...

Note that \( \lambda_1^1[1] + \lambda_2^1[1] = 1 \). It follows then, if Player 1’s best reply is row 1 then Player 2’s best reply will be column 2 (and vice versa). So in each iteration there will be persistent miscoordination !!

What about each player’s empirical frequency distribution?
Figure 1: Player 1’s Empirical Frequency of Playing Action 1.
Figure 2: Player 2’s Empirical Frequency of Playing Action 1.
Remarks

• In FP, the convergence of player’s empirical frequency distributions may not reflect what is realized in the game on a per period basis.
  – A consequence of this could be that the empirical joint distribution of the player’s play can be correlated.

• Surely if the players kept track of the payoffs they obtain they would realize that something was wrong with their world-view?
  – This idea is captured in the concept of $\epsilon$-consistency. Refer [Fudenberg & Levine 1991] for details.
Some Games with the Fictitious Play Property

- Every finite, zero-sum, two-person game.

- Every non-degenerate 2 X 2 matrix game.

- **Every game with Identical Payoff Functions**!

  - Wait a minute, isn’t this good news?
It depends

• Potential of Fictitious Play for optimization has been shown in [Garcia, Smith & Reaume, 2000a]
  – The players (vehicles) in the network are in a game of identical interests, namely to minimize the average trip time.
  – The FP implementation outperformed alternative algorithms.

• There are some issues however ...
  – Exact FP algorithm is unsuitable for problems with large solution spaces ...
    * A best reply might involve enumeration of this entire space. Note a player has to compute \( \arg \min_{\alpha \in A_i} U(\alpha, \lambda_{-i}) \) at each step.
  – Analytical representation of \( U(a(k)) \) may be unavailable.
  – Nash Equilibrium does not account for correlation in players decisions.

Simplifications - I

1. **Sampling** - Instead of evaluating expectations with respect to empirical frequency distributions, we can sample the empirical frequency distribution.

2. **Simulation** - Instead of exact evaluation of the cost, \( U(a(k)) \), associated with a joint action \( a(k) \), we can use a noisy estimate (as from a simulation).

*(Sampled Fictitious Play)* If players independently sample from each others empirical frequency distributions (and independently simulate joint actions) then if the number of samples \( M_t \) taken by each player grows polynomially as \( t \to \infty \) then Monderer and Shapleys set convergence result is preserved.


This approach has its drawbacks ...

- In large scale systems, simulations are expensive.

- Requirement of \( M_t \) growing polynomially leaves us no better off than before.
Simplifications - II

A Wish list ...

1. Why not let players sample the empirical distribution of joint actions?
   - Easily done, as we have the play history, \( H^t = \{ a^0, \ldots, a^{t-1} \} \).

2. and share simulation output?
   - Implicitly done anyway ...
   - Need to account for correlation in players decision making (set convergence to NE is no longer clear)

3. How about taking a finite number of samples \( M_t = M \) ?

Need to move away from the Nash equilibrium solution concept ... Before we do that let us introduce some more notation.
The set of probability distributions over the set of joint actions is
\[
\Lambda = \left\{ \lambda = (\lambda(1), \ldots, \lambda(m)) \in \mathbb{R}_+^m : \sum_{k=1}^{m} \lambda(k) = 1 \right\}.
\]

- Define
\[
\lambda^t(k) = \frac{\text{# times } a(k) \text{ is played in } H^t}{t}.
\]
- Define
\[
\lambda^t = (\lambda^t(1), \lambda^t(2), \ldots, \lambda^t(m)) \in \Lambda.
\]
- Interpretation:
\[
\lambda^t(k) \text{ is the probability that joint action } a(k) \in A \text{ is observed}
\]
Marginal Distributions

Player $i$’s probability of choosing an action $\alpha$ from his or her action set $A_i$, denoted by $\lambda_i[\alpha]$, can be conveniently calculated as

$$\lambda_i[\alpha] = \sum_{k : a_i(k) = \alpha} \lambda(k).$$

Given a profile of actions $a_{-i} \in \prod_{i \neq j} A_i$ (i.e. actions for all players other than player $i$), the probability of experiencing $a_{-i}$ under $\lambda \in \Lambda$ is

$$\lambda_{-i}[a_{-i}] = \sum_{k : a_{-i}(k) = a_{-i}} \lambda(k).$$

Given a play history $H^t$, the marginal relative frequency distributions $\lambda^t_i$ and $\lambda^t_{-i}$ are defined similarly.
Solution Concept -

*Endogenous Correlated Equilibrium*


Endogenous Correlation

Given $\lambda \in \Lambda$, each player $i \in N$ believes all other players will play \textit{(in a possibly correlated fashion)} in accordance with $\lambda_{-i}$.

Definitions:

- The cost of player $i$ unilaterally insisting on a particular action $\alpha \in A_i$ is

$$U(\alpha, \lambda_{-i}) \equiv \sum_{k=1}^{m} \lambda_k U(\alpha, a_{-i}(k)) = \sum_{a_{-i}} \lambda_{-i}[a_{-i}] U(a, a_{-i}).$$

- Player $i$’s \textit{best-reply correspondence} to $\lambda$:

$$B_i(\lambda) = \arg \min_{\alpha \in A_i} U(\alpha, \lambda_{-i}).$$

- Player $i$’s “best response” to $\lambda$:

$$\beta_i(\lambda) \in B_i(\lambda).$$
Endogenous Correlated Equilibrium (ECE)

A probability measure $\lambda^* \in \Lambda$ is an $\eta$–ECE, $\eta > 0$ if and only if,

$$\lambda^*_i[\alpha] > 0 \iff \alpha \in B_i(\lambda^*; \eta)$$

where $\alpha \in B_i(\lambda^*; \eta) \iff U(\alpha, \lambda^*_i) - \min_{a \in A_i} U(a, \lambda^*_i) \leq \eta$

A 0 – ECE will simply be referred to as an ECE.

Contrast to:

- Nash Equilibrium: In two-player games, ECE is equivalent to Nash equilibrium. In general,

  $$\text{Nash Equilibria} \subset ECE$$

  Moreover, there are ECE that Pareto-dominate all Nash equilibria.

- Correlated Equilibrium: In Aumann’s solution concept players correlate their actions via an exogenous random variable (e.g. sunspots).
ECE’s and Optimal Solutions

1. It is not clear a priori that ECE solutions necessarily put positive measure on optimal joint actions.
   - On the other hand, ECE’s by definition put positive measure on actions that are best responses to themselves.

2. We believe that by accounting for correlation, ECE solutions can describe a more “coordinated” response to the problem than Nash equilibria.
Correlated Learning Algorithm [Vanderschraaf, 1997]

(Initialization)
- Start with $t = 0$ and initial play history $H^0 = \{a^0\} \subset A$.

(Iteration $t \geq 1$)
- Given $H^{t-1}$ (and its relative frequency distribution $\lambda^{t-1}$),
  i. each player $i$ computes its individual best response $\beta_i(\lambda^{t-1})$,
  ii. construct the aggregate best response
  $$x^t = (\beta_1(\lambda^{t-1}), \beta_2(\lambda^{t-1}), \ldots, \beta_{n-1}(\lambda^{t-1})) \in A,$$
  iii. add $x^t$ to the play history:
  $$H^t = (H^{t-1}, x^t).$$
Remarks

- Our correlated learning algorithm can be interpreted as a special case of Dirichlet’s Inductive Deliberators [Vanderschraaf, 1997].

- It is possible to show [Vanderschraaf, 1997] that if $\lambda_t \to \lambda^*$, then $\lambda^*$ is an ECE.
  - Unfortunately such convergence is not guaranteed.
Convergence on Subsequences

(Main Result)

**Theorem 1:** The accumulation points of \( \{ \lambda_t \}_{t=1}^{\infty} \subset \Lambda \), generated by the correlated learning algorithm, are ECE.

In other words, if \( \{ \lambda^{t(s)}_t \}_{s=1}^{\infty} \) is a convergent subsequence such that \( \lambda^{t(s)}_t \to \lambda^* \), then \( \lambda^* \) is an ECE. (For any player \( i \) and action \( \alpha \in A_i \), if \( \lambda^*_i[\alpha] > 0 \), then \( \alpha \) is a best response to \( \lambda^* \), i.e. \( \alpha \in B_i(\lambda^*) \).)

Remark

- This result is analogous to Monderer and Shapley’s set convergence result for fictitious play.
Sampled Correlated Learning Algorithm

(Initialization)

- Start with $t = 0$ and initial play history $H^0 = \{a^0\} \subset A$.

(Iteration $t \geq 1$)

- Given $H^{t-1}$ (and its relative frequency distribution $\lambda^{t-1}$),
  i. With Uniform probability, choose $M_t$ elements from the integers \{0, 1, \ldots, t - 1\}.
  ii. each player computes a best response $x_i$ so that
      \[
      \hat{U}(x_i, \lambda^{t-1}_{-i}) \leq \hat{U}(\alpha', \lambda^{t-1}_{-i}), \quad \forall \alpha' \in A_i,
      \]
      where $\hat{U}(\cdot, \lambda^{t-1}_{-i})$ is the sample average w.r.t. the $M_t$ samples,
  iii. aggregate best responses $x^t = (x_1, x_2, \ldots, x_n) \in A$,
  iii. add $x^t$ to the play history: $H^t = (H^{t-1}, x^t)$. 
Observations

At time instant \( t \), the error associated with a player \( i \in N \) best replying to \( \hat{U}(\alpha, \lambda^t_{-i}) \) instead of \( U(\alpha, \lambda^t_{-i}) \) will be denoted by \( \epsilon_{t,i}(\alpha) \) i.e.,

\[
\epsilon_{t,i}(\alpha) = \hat{U}(\alpha, \lambda^t_{-i}) - U(\alpha, \lambda^t_{-i})
\]  

(1)

The properties of the estimator \( \hat{U}(\alpha, \lambda^t_{-i}) \) are summarized in the following lemma:

Lemma 1

\( (a) \quad E[\epsilon_{t,i}(\alpha)] = 0 \)

\( (b) \quad Var[\epsilon_{t,i}(\alpha)] \leq \frac{(O^2+S^2)}{M_t} \)

\( (c) \quad E[\epsilon_{t,i}(\alpha)\epsilon_{t,i}(\alpha')] \geq -\frac{2O^2}{M_t} \)

In the above \( |U(a(k))| \leq O, \forall a(k) \in A \) and \( S \) bounds the variance of zero-mean error term that occurs while estimating any joint action via simulation.
Theorem: 2 For $\eta > 0$, let $M$ be such that $\frac{2(2O^2 + S^2)}{\eta^2 M} < 1$. With probability 1, the accumulation points of $\{\lambda^t\}_{t=1}^\infty$, generated by the sampled algorithm are $\eta$–ECE.
Conclusions

We have developed fictitious-play algorithms for computing ECE solutions.

- We have shown that convergent subsequences under endogenous correlated fictitious play (exact algorithm) lead to ECE solutions.

- We have shown that convergent subsequences under *sampled* endogenous correlated fictitious play has the same property.
  - Moreover we have established conditions on finite number of samples to reach $\eta - ECE$.

Our analytical framework allows us to consider problems involving

- noisy cost evaluation (simulation-based optimization)
Questions?