Strong Convergence to Mixed Equilibria in Fictitious Play

BRIAN SWENSON[†*], SOUMMYA KAR[†] AND JOÃO XAVIER[⋆]

***Abstract*—Learning processes that converge to mixed-strategy equilibria often exhibit learning only in the weak sense in that the time-averaged empirical distribution of players' actions converges to a set of equilibria. A stronger notion of learning mixed equilibria is to require that players period-by-period strategies converge to a set of equilibria. A simple and intuitive method is considered for adapting algorithms that converge in the weaker sense in order to obtain convergence in the stronger sense. The adaptation is applied to the the well-known fictitious play (FP) algorithm, and the adapted version of FP is shown to converge to the set of Nash equilibria in the stronger sense for games known to have the FP property.**

***Index Terms*—Games, Learning, Mixed Equilibria, Fictitious Play, Nash Equilibria,**

I. Introduction

The theory of learning in games is concerned with investigating how dynamical systems induced by repeated play of a normal-form game can lead players to learn equilibrium strategies. The manner in which agents learn equilibrium strategies may be classified into two general categories. We say a learning process converges *weakly* to a set of equilibria if the time-averaged empirical distribution of players' actions converges to an equilibrium set. A well-noted shortcoming of this mode of convergence is that it does not imply that players period-by-period strategies ever converge to an equilibrium set themselves. In contrast to this, we say a learning process converges *strongly* to a set of equilibria if the period-by-period strategies of agents converge to an equilibrium set.

Weak convergence is the traditional mode in which players are said to learn mixed-strategy equilibria. It is often interpreted as a convergence of players 'beliefs' to equilibrium. An underlying issue with this mode of convergence is that, at mixed strategy equilibria, the best response correspondence is not lower semicontinuous [1]; even though a player's 'beliefs' may be converging to equilibrium, the best response associated with each belief may be a unique pure strategy at every step of the repeated play. In such scenarios, it is often the case that the period-by-period strategies cycle in such a way as to drive the time-averaged empirical distribution of actions to equilibrium, while the period-by-period strategies never approach equilibrium themselves [2],[3].

Strong convergence, on the other hand, implies that the strategies actually employed by players are asymptotically optimal in the sense that they converge to a set of equilibria.

Fictitious play (FP) is a classic algorithm which typifies a large class of learning algorithms[1] and is known to converge weakly to the set of Nash equilibria. Our main contribution is to present a simple and intuitive adaptation of FP that converges strongly to the set of Nash equilibria. In our strongly convergent variant of FP, players gradually and independently transition from using the FP best response rule to determine the next-iteration action, to using their current empirical distribution as a probability mass function (pmf) from which they sample to determine their next-iteration action.

From the perspective of a player, this may be seen as a gradual transition from 'learning' to 'implementation'. The idea is that as long as the rate of transition from learning to implementation is sufficiently slow, the empirical distribution will continue to move toward equilibrium, per the FP process. Since players are increasingly likely to draw their next-iteration action as a random sample from their empirical distribution, their period-by-period strategies also approach equilibrium.

While the algorithm presented here deals specifically with FP, an appealing aspect of our approach is that it may be readily generalized as a method for obtaining strong convergence in other weakly convergent learning algorithms. The key property which enables strong convergence in our approach is the robustness property of FP (see section 1). Our analysis may be readily extended to other weakly convergent algorithms that can be shown to possess a similar robustness property.

The work was partially supported by the FCT project [PTDC/EMS-CRO/2042/2012] through the Carnegie-Mellon/Portugal Program managed by ICTI from FCT and by FCT Grant CMU-PT/SIA/0026/2009.

[†]Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15213, USA (brianswe@andrew.cmu.edu and soummyak@andrew.cmu.edu).

[⋆]Institute for Systems and Robotics (ISR), Instituto Superior Tecnico (IST), University of Lisbon, Portugal (jxavier@isr.ist.utl.pt).

[1]FP is considered prototypical of most learning algorithms which rely on best response dynamics.

Related research ([2],[1]) has studied various problems associated with learning mixed-strategy equilibria, including the issue of weak convergence. Stochastic FP (SFP) is an adaptation of FP which seeks to address the problem of weak convergence and provide an explanation for why players might wilfully choose to play randomized strategies in a learning process. SFP has been shown to converge strongly in various classes of games ([2],[4],[5],[6],[7]). Leslie et al. [7] present a payoff-based actor-critic implementation of FP that is shown to converge strongly in games known to have the FP property.

The remainder of the paper is organized as follows. Section II sets up the notation to be used in the subsequent development and presents the traditional FP algorithm. Section III presents our strongly convergent adaptation of FP. The algorithm is proved to converge strongly to the set of Nash equilibria in the same section. Section IV concludes the paper.

II. Preliminaries

A. Setup and Notation

We consider a normal form game Γ with set of players $N = \{1, \ldots, n\}$. The set actions, or pure strategies, for player i is given by A_i. The set of mixed strategies for player i, denoted by Δ_i, is the convex hull of A_i. The set of joint mixed strategies is given by $\Delta^n = \prod_{i \in N} \Delta_i$.

The utility function for player i is given by $U_i(\cdot) : \Delta^n \to \mathbb{R}$. When convenient we sometimes write $U(p)$ as $U_i(p_i, p_{-i})$, where p_i denotes the mixed strategy of player i and p_{-i} denotes the mixed strategies of all other players. The set of Nash equilibria is given by $NE := \{p \in \Delta^n : U_i(p_i, p_{-i}) \geq U_i(p'_i, p_{-i}),\ \forall i \in N\}$. We denote by $BR_i(p_{-i}) = \{\arg\max_{p_i \in \Delta_i} U_i(p_i, p_{-i})\}$ a players set of best responses.

We consider repeated play of Γ. Let $\{g_i(t)\}_{t \geq 1}$, $g_i(t) \in \Delta_i$ be a sequence of mixed strategies for player i. Let $g(t) = (g_1(t), \ldots, g_n(t))$ be the n-tuple containing the mixed strategy of each player for round t. Let $\{a_i(t)\}_{t \geq 1}$ be a sequence of actions such that $a_i(t)$ is obtained by sampling $g_i(t)$. The normalized histogram (empirical distribution) of player i's actions is denoted by $q_i(t) := \frac{1}{t} \sum_{s=1}^{t} a_i(s)$. Let $q(t) = (q_1(t), \ldots, q_n(t))$ be the n-tuple containing each player's marginal empirical distribution.

Unless otherwise stated, $d(\cdot, \cdot)$ denotes the standard Euclidean norm. We say a process $\{a(t)\}_{t \geq 1}$ converges *weakly* to equilibrium if $d(q(t), NE) \to 0$ almost surely (a.s.) as $t \to \infty$, and we say a process $\{a(t)\}_{t \geq 1}$ converges *strongly*[2] to equilibrium if $d(g(t), NE) \to 0$ a.s. as $t \to \infty$.

B. Fictitious Play

FP may be intuitively understood as follows. Players repeatedly face off in a stage game Γ. In any given stage of the game, players choose a next-stage action by assuming (inaccurately) that opponents are using stationary and independent strategies. Thus, in FP, players use the marginal empirical distribution of each opponent's past play as a prediction of the opponent's behavior in the subsequent round and choose a next-round strategy which is a best response against this prediction.

A sequence of actions $\{a(t)\}_{t \geq 1}$ such that[3]

$$a_i(t+1) \in \arg\max_{\alpha_i \in A_i} U_i(\alpha_i, q_{-i}(t)),\ \forall i, \forall t \geq 2$$

is referred to as a *fictitious play process*. FP has been studied extensively ([9],[10],[11],[5]) to determine the classes of games for which it can be said to converge (weakly) to the set of Nash equilibria. We summarize these results in the following theorem.

Theorem 1. *Let $\Gamma = (N, \{u_i(\cdot)\}_{i \in N}, Y^n)$ be a two-player zero-sum game, potential game, or generic 2 by m game, and let $\{a(t)\}$ be a fictitious play process on Γ. Then $d(q(t), NE) \to 0$ as $t \to \infty$.*

An interesting generalization of FP is to consider a scenario where players are permitted some asymptotically decaying error in their understanding of the empirical distribution. Such generalizations have been studied, amongst other places, in [2],[7] and [12].

An important property of FP to be used in the proof of our main result is that the convergence results of Theorem 1 still hold in the presence of such asymptotically decaying errors. We refer to this as the robustness property of FP. To make this precise, we say a sequence of actions $\{a(t)\}_{t \geq 1}$ is a *perturbed fictitious play process* if

$$a_i(t+1) \in \arg\max U_i(\alpha_i, (q_{-i}(t) + \epsilon_i^t))$$

where $q_{-i}(t) + \epsilon_i^t \in \Delta_{-i}$. The following lemma states that, if the magnitude of the perturbations decays to zero, then a perturbed fictitious play process will converge for the same class of games given in Theorem 1.

Lemma 1. *(Robustness of Fictitious Play) Let $\Gamma = (N, \{u_i(\cdot)\}_{i \in N}, Y^n)$ be a two-player zero-sum game,*

[2]The notion of strong convergence presented in this paper is comparable to the notions of 'convergence in intended behavior' given in [2], and 'convergence in strategic intentions' given in [8].

[3]In all variants of FP discussed in this paper $a_i(1)$ may be chosen arbitrarily for all i.

potential game, or generic 2 by m game. Let $\{a(t)\}_{t\geq 1}$ be a perturbed FP process on Γ and let $\|\epsilon_i^t\| \to 0$ as $t \to \infty$. Then $d(q(t), NE) \to 0$ as $t \to \infty$.

Proof: Note that $U_i(\cdot)$ is multilinear and therefore Lipschitz continuous. Hence, there exists a positive constant K such that $|U_i(p) - U_i(p')| \leq K\|p - p'\|$ for all $p, p' \in \Delta^n$. Claim that $\max_{\alpha_i \in A_i} U_i(\alpha_i, q_{-i}(t)) - U_i(a_i(t+1), q_{-i}(t)) \leq 2K\|\epsilon_i^t\|$.

To show this, let $\alpha_i^* \in \{\arg\max_{\alpha_i \in A_i} U_i(\alpha_i, q_{-i}(t))\}$ and observe that

$$\begin{aligned} U_i(\alpha_i^*, q_{-i}(t)) &\leq U_i(\alpha_i^*, q_{-i}(t) + \epsilon_i^t) + K\|\epsilon_i^t\| \\ &\leq U_i(a_i(t+1), q_{-i}(t) + \epsilon_i^t) + K\|\epsilon_i^t\|, \end{aligned}$$

Since $a_i(t+1) \in \arg\max_{\alpha_i \in A_i} U_i(\alpha_i, q_{-i}(t) + \epsilon_i^t)$, a symmetric argument yields $U_i(a_i(t+1), q_{-i}(t) + \epsilon_i^t) \leq U_i(\alpha_i^*, q_{-i}(t)) + K\|\epsilon_i^t\|$, and hence, $|U_i(a_i(t+1), q_{-i}(t) + \epsilon_i^t) - U_i(\alpha_i^*, q_{-i}(t))| \leq K\|\epsilon_i^t\|$. It follows that,

$$\begin{aligned} &U_i(\alpha_i^*, q_{-i}(t)) - U_i(a_i(t+1), q_{-i}(t)) \\ \leq &|U_i(a_i(t+1), q_{-i}(t)) - U_i(a_i(t+1), q_{-i}(t) + \epsilon_i^t)| \\ &+ |U_i(a_i(t+1), q_{-i}(t) + \epsilon_i^t) - U_i(\alpha_i^*, q_{-i}(t))| \\ \leq &2K\|\epsilon_i^t\|, \end{aligned}$$

and the claim holds. This implies that $a_i(t+1)$ is a $(2L\|\epsilon_i^t\|)$-best response, as defined in [7]. Since $\|\epsilon_i^t\| \to 0$ as $t \to \infty$ for all i, the process $\{a(t)\}_{t\geq 1}$ is a *generalized weakened fictitious play* (GWFP) process as defined in [7]. By [7], Corollary 5, any GWFP process converges to the set of Nash equilibria in the sense that $d(q(t), NE) \to 0$ as $t \to \infty$ in two-player zero-sum games, potential games, and generic 2 by m games. ■

III. Strong Convergence in Fictitious Play

Consider a variant of FP where the action for player i at time $t+1$ is chosen according to the following (random) rule

$$a_i(t+1) \sim g_i(t+1), \tag{1}$$

where $g_i(t+1)$ is given by[4]

$$g_i(t+1) := \begin{cases} BR_i(q_{-i}(t)) & \text{with probability } \frac{1}{(t+1)^r} \\ q_i(t) & \text{otherwise,} \end{cases}$$

and $a_i(t) \sim g_i(t)$ indicates that the action $a_i(t)$ is drawn as a random sample from the probability mass function $g_i(t) \in \Delta_i$. In this variant of FP we define the 'empirical distribution' of a players actions $q_i(t)$ to be the empirical average (only) over rounds when a player chooses to play a best response (BR). Formally, let X_i^t be a random variable such that

$$X_i^t = \begin{cases} 1, & \text{if a BR is chosen by } i \text{ at time } t \\ 0, & \text{otherwise.} \end{cases}$$

Let $\ell_i(t) := \sum_{k=1}^{t} X_i^k$. The empirical distribution $q_i(t)$ is defined recursively as

$$q_i(t+1) = \begin{cases} q_i(t) + \frac{1}{\ell(t+1)}\left(a_i(t+1) - q_i(t)\right), & \text{if } X_i^{t+1} = 1 \\ q_i(t) & \text{otherwise.} \end{cases} \tag{2}$$

The joint empirical distribution is given by $q(t) = (q_1(t), \ldots, q_n(t))$.

A. Main Result

The following theorem asserts that this variant of FP will converge *strongly* to the set of Nash equilibria, so long as the rate at which agents transition from 'learning' to 'implementation' is sufficiently slow (i.e., $0 < r < \frac{1}{2}$).

Theorem 2. *Let Γ be a two-player zero-sum game, potential game, or generic 2 by m game. Let $\{a(t)\}_{t\geq 1}$ be a sequence of actions chosen according to* (1) *with $0 < r < \frac{1}{2}$, and let $\{q(t)\}_{t\geq 1}$ be updated according to* (2)*. Then the process converges strongly to the set Nash equilibria. That is, $d(g(t), NE) \to 0$ almost surely as $t \to \infty$.*

Proof: For each $i \in N$, define the sequence $\{\tau_i(s)\}_{s\geq 1}$ and let $\tau_i(1)$ represent the round t when player i picks a best response for the first time, let $\tau_i(2)$ represent the round t when player i picks a best response for the second time, and so on.

Let $\{\hat{q}^i(s)\}_{s\geq 1}$ be a sequence such that $\hat{q}^i(s) := (q_1(\tau_i(s+1) - 1), \ldots, q_n(\tau_i(s+1) - 1))$.

Let $\{\tilde{a}_i(s)\}_{s\geq 1}$ be a sequence such that $\tilde{a}_i(s) := a_i(\tau_i(s))$ and let $\{\tilde{a}(s)\}_{s\geq 1}$ be the corresponding joint sequence such that $\tilde{a}(s) = (a_1(\tau_i(s)), \ldots, a_n(\tau_n(s)))$.

Let $\tilde{q}_i(s) := \frac{1}{s}\sum_{k=1}^{s} \tilde{a}_i(k)$ and $\tilde{q}(s) := \frac{1}{s}\sum_{k=1}^{s} \tilde{a}(k)$.

The sequence of actions $\{\tilde{a}(s)\}_{s\geq 1}$ may be thought of as a perturbed fictitious play process II-B. That is, the sequence of actions can be thought of as following the rule

$$U_i(\tilde{a}_i(s+1), \hat{q}_{-i}^i(s)) = \max_{\alpha_i \in \Delta_i} U_i(\alpha_i, \hat{q}_{-i}^i(s)).$$

The distribution $\tilde{q}(s)$ represents the 'true' empirical distribution of play in this process, and $\hat{q}^i(s)$ represents the estimate which player i has of $\tilde{q}(s)$.

[4]Clearly, $BR_i(p_{-i}) := \{\arg\max_{p_i \in \Delta_i} U_i(p_i, p_{-i})\}$ is a set. In an abuse of notation, when we say $g_i(t+1) = BR_i(q_{-i}(t))$ we mean that $g_i(t+1) \in BR_i(q_{-i}(t))$.

By Lemma 2, $\|\hat{q}^i(s) - \tilde{q}(s)\| = O\left(s^{r-1/2}\right)$ a.s. Therefore, by Lemma 1,

$$d(\tilde{q}(s), NE) \to 0 \text{ as } s \to \infty \text{ a.s.} \tag{3}$$

To show $d(q(t), NE) \to 0$ as $t \to \infty$ a.s., let $\varepsilon > 0$ be given. Let $\hat{\tilde{q}}_i(s) := q(\tau^i(s)) = (q_1(\tau_i(s)), \ldots, q_n(\tau_i(s)))$. By Lemma 3, for each $i \in N$, there exists a random time $S_i > 0$ such that $\forall s \geq S_i$, $\|\hat{\tilde{q}}^i(s) - \tilde{q}(s)\| < \frac{\varepsilon}{2}$ a.s. Let $S^{'} = \max\{S_i\}$. By (3) there exists a random time $S^{''}$ such that $\forall s \geq S^{''}$, $d(\tilde{q}(s), NE) < \frac{\varepsilon}{2}$ a.s. Let $S = \max\{S^{'}, S^{''}\}$. Then $\forall i \in N$, $\forall s \geq S$, $d(\hat{\tilde{q}}^i(s), NE) < \varepsilon$ a.s.

Let $T = \max_i\{\tau_i(S)\}$.[5] Note that for some i, $q(T) = \hat{\tilde{q}}^i(S)$ and therefore

$$d(q(T), NE) < \varepsilon \text{ a.s.} \tag{4}$$

Also note that for any $t_0 > T$, it holds that $\ell_i(t_0) \geq S$, and moreover

$$X_i^{t_0} = 1 \text{ for some } i \;\Rightarrow\; q(t_0) = \hat{\tilde{q}}^i(\ell_i(t_0)), \quad \text{with } \ell_i(t_0) \geq S \tag{5}$$

$$X_i^{t_0} = 0 \text{ for all } i \;\Rightarrow\; q(t_0) = q(t_0 - 1).$$

Consider $t \geq T$. If for some i, $X_i^t = 1$, then by (5), $d(q(t), NE) = d(\hat{\tilde{q}}^i(\ell_i(t)), NE) < \varepsilon$ a.s. Otherwise, if $X_i^t = 0\ \forall i$, then $q(t) = q(t-1)$.

Iterate this argument m times until either (i) $X_i^{t-m} = 1$ for some i, or (ii), $t - m = T$. In the case of (i), $d(q(t), NE) = d(q(t-m), NE) = d(\hat{\tilde{q}}^i(\ell_i(t-m)), NE) < \varepsilon$ a.s. Or, in the case of (ii), $d(q(t), NE) = d(q(T), NE) < \varepsilon$ a.s, where the inequality follows from (4). Hence $d(q(t), NE) \to 0$ a.s. as $t \to \infty$.

Finally, note that by (1), $\|g_i(t) - q_i(t)\| \to 0$ as $t \to \infty$, from which it follows that $d(g(t), NE) \to 0$ a.s. as $t \to \infty$. ∎

IV. Conclusions

In traditional fictitious play (FP), learning occurs only in the weak sense that the time-averaged empirical distribution of players' actions converges to the set of Nash equilibria. We present a simple adaptation of FP that converges in the stronger sense that players' period-by-period strategies converge to the set of Nash equilibria. In our strongly convergent variant of FP, players gradually and independently transition from using the FP best response rule to determine the next-iteration action, to using their current empirical distribution as a probability mass function from which they sample to determine their next-iteration action. An interpretation of this procedure is to say that each player gradually and independently transitions from learning to implementation.

We show that this approach converges strongly to the set of Nash equilibria due to the robustness property of FP (see sec. III). An interesting future research direction will be to investigate similar adaptations of other weakly convergent learning algorithms which can be shown to possess a comparable robustness property.

[5] Note that the almost sure occurrences mentioned before ($\|\hat{\tilde{q}}^i(s) - \tilde{q}(s)\| \to 0$ a.s., and $d(\tilde{q}(s), NE) \to 0$ a.s.) are all dependent on the (almost sure) occurrence of one event: $\|\ell_i(t) - E[\ell_i(t)]\| = O(t^{\frac{1}{2}})\ \forall i$ (see Lemma 4). Note that if this event occurs (which it does, a.s.) then, in addition to implying the previous mentioned occurrences, it also implies that $T < \infty$ a.s.

V. Appendix

A. Intermediate Results

Lemma 2. *Let $\hat{q}^i(s)$ and $\tilde{q}(s)$ be defined as in the proof of Theorem 2. Then $\|\hat{q}^i(s) - \tilde{q}(s)\| = O\left(s^{r-1/2}\right)$ a.s.*

Proof: Let $\hat{\tilde{q}}^i(s) = (q_1(\tau_i(s)), \ldots, q_n(\tau_i(s)))$. Then,

$$\begin{aligned}\|\hat{q}^i(s-1) - \hat{\tilde{q}}^i(s)\| &= \|q(\tau_i(s) - 1) - q(\tau_i(s)\| \\ &\leq \frac{1}{\tau_i(s)} \leq \frac{1}{s},\end{aligned}$$

where the first inequality follows from the fact that the step size of $q(t)$ is $\frac{1}{t}$ and the second inequality follows from the fact that $s \leq \tau_i(s)$. By Lemma 3 it holds that $\|\hat{\tilde{q}}^i(s) - \tilde{q}(s)\| = O\left(s^{r-1/2}\right)$ almost surely. Hence, by the triangle inequality,

$$\begin{aligned}&\|\hat{q}^i(s) - \tilde{q}(s)\| \leq \\ &\underbrace{\|\hat{q}^i(s) - \hat{q}^i(s-1)\|}_{=O(s^{-1})} + \underbrace{\|\hat{q}^i(s-1) - \hat{\tilde{q}}^i(s)\|}_{=O(s^{-1})} \\ &+ \underbrace{\|\hat{\tilde{q}}^i(s) - \tilde{q}(s)\|}_{=O(s^{r-1/2}) \text{ a.s.}} \\ &= O(s^{r-1/2}) \text{ a.s.}\end{aligned}$$

∎

Lemma 3. *Let $\hat{\tilde{q}}^i(s) = (q_1(\tau_i(s)), \ldots, q_n(\tau_i(s)))$ and let $\tilde{q}(s)$ be defined as in the proof of Theorem 2. Then $\|\hat{\tilde{q}}^i(s) - \tilde{q}(s)\| = O\left(s^{r-1/2}\right)$ a.s.*

Proof: Consider the pairwise difference $\|\hat{\tilde{q}}_j^i(s) - \tilde{q}_j(s)\|$; it is sufficient to show that for each $j \in N$, $\|\hat{\tilde{q}}_j^i(s) - \tilde{q}_j(s)\| = O\left(s^{r-1/2}\right)$ a.s.

Let X_i^t, $\ell_i(t)$, and $\tau_i(s)$ be defined as before. Note that for any $t \in \mathbb{N}$, $q_j(t) = \tilde{q}_j(\ell_j(t))$. Also note that $\ell_i(\tau_i(s)) = s$. Therefore,

$$\begin{aligned}\|\hat{\tilde{q}}_j^i(s) - \tilde{q}_j(s)\| &= \|q_j(\tau_i(s)) - \tilde{q}_j(s)\| \\ &= \|\tilde{q}_j(\ell_j(\tau_i(s))) - \tilde{q}_j(s)\| \\ &= \|\tilde{q}_j(\ell_j(\tau_i(s))) - \tilde{q}_j(\ell_i(\tau_i(s)))\| \\ &\leq \frac{|\ell_j(\tau_i(s)) - \ell_i(\tau_i(s))|}{\min\{\ell_j(\tau_i(s)), \ell_i(\tau_i(s))\}} \\ &= \frac{|\ell_j(t) - \ell_i(t)|}{\min\{\ell_j(t), \ell_i(t)\}},\end{aligned}$$

where in the last step we let $\tau_i(s) = t$, and without loss of generality we let $\max_{p,p' \in \Delta_j} \|p - p'\| = 1$. Thus, it is sufficient to show that, $\frac{|\ell_j(t)-\ell_i(t)|}{\min\{\ell_j(t),\ell_i(t)\}} = O\left(s^{r-1/2}\right)$ a.s.

Let event A be the event that $|\ell_j(t) - E[\ell_j(t)]| = O(t^{\frac{1}{2}})$. Let event B be the event that $|\ell_i(t) - E[\ell_i(t)]| = O(t^{\frac{1}{2}})$.

By Lemma 4, event A and event B occur a.s., and therefore $A \cap B$ occurs a.s. Hence, there exist a random time T and a non-negative random variable M, such that $\forall t > T$, $|\ell_i(t) - E[\ell_i(t)]| \leq Mt^{\frac{1}{2}}$ and $|\ell_j(t) - E[\ell_j(t)]| \leq Mt^{\frac{1}{2}}$ almost surely. Thus, $\forall\ t > T$,

$$\frac{|\ell_j(t) - \ell_i(t)|}{\min\{\ell_j(t), \ell_i(t)\}} \leq \frac{2Mt^{\frac{1}{2}}}{E[\ell(t)] - Mt^{\frac{1}{2}}} \text{ a.s.,} \qquad (6)$$

where, noting that $E[\ell_i(t)] = E[\ell_j(t)]\ \forall i, j$, we define $E[\ell(t)] := E[\ell_i(t)] = E[\ell_j(t)]$. Note that

$$E[\ell(t)] = E[\sum_{k=1}^{t} X_i^t] = \sum_{k=1}^{t} E[X_i^t] = \sum_{k=1}^{t} \frac{1}{k^r}$$

and moreover,

$$\begin{aligned}&\frac{1}{1-r}\left((t+1)^{(1-r)} - 1\right) = \int_0^t \frac{1}{(x+1)^r} dx \\ &< \sum_{k=1}^{t} \frac{1}{k^r} < \int_0^t \frac{1}{x^r} dx = \frac{1}{1-r} t^{(1-r)}. \qquad (7)\end{aligned}$$

Therefore, (6) can be strengthened to

$$\begin{aligned}\frac{|\ell_j(t) - \ell_i(t)|}{\min\{\ell_j(t), \ell_i(t)\}} &\leq \frac{2Mt^{\frac{1}{2}}}{\frac{1}{1-r}\left((t+1)^{(1-r)} - 1\right) - Mt^{\frac{1}{2}}} \\ &= O\left(t^{r-1/2}\right) \text{ a.s.}\end{aligned}$$

Noting that $t = \tau_i(s) \geq s$ we get

$$\frac{|\ell_j(t) - \ell_i(t)|}{\min\{\ell_j(t), \ell_i(t)\}} = O\left(s^{r-1/2}\right) \text{ a.s.,}$$

which, by the above explanation implies $\|\hat{\tilde{q}}_j^i(s) - \tilde{q}_j(s)\| = O\left(s^{r-1/2}\right)$ almost surely. ■

Lemma 4. $\limsup_{t\to\infty} \frac{|\ell_i(t) - E[\ell_i(t)]|}{t^{\frac{1}{2}}} \leq 1$ *a.s.*

Proof: Let X_i^t and $\ell_i(t)$ be defined as before. Let $Y_i^t := X_i^t - E[X_i^t]$. Note that $\ell_i(t) - E[\ell_i(t)] = \sum_{k=1}^{t} Y_i^k$, and that $E[Y_i^t] = 0$, and that $E[(Y_i^t)^2] = \frac{1}{t^r}(1 - \frac{1}{t^r})$. Let $B^t = \sum_{k=1}^{t} E[(Y_i^k)^2]$. The sequence $\{Y_i^t\}$ meets the necessary assumptions of the (Kolmogorov) law of the iterated logarithm [13], and therefore

$$\lim_{t\to\infty} \sup \frac{|\sum_{k=1}^{t} Y_i^t|}{(2B^t \log\log B^t)^{\frac{1}{2}}} = 1, \text{ a.s.}$$

Note that

$$\begin{aligned}B^t &= \sum_{k=1}^{t} E[(Y_i^k)^2] = \sum_{k=1}^{t} \frac{1}{k^r}\left(1 - \frac{1}{k^r}\right) \\ &< \sum_{k=1}^{t} \frac{1}{k^r} < \frac{1}{1-r} t^{1-r}.\end{aligned}$$

where the last inequality follows from (7). Hence,

$$\begin{aligned}&\lim_{t\to\infty} \sup \frac{|\ell_i(t) - E[\ell_i(t)]|}{\left(2(\frac{1}{1-r}t^{1-r}) \log\log(\frac{1}{1-r}t^{1-r})\right)^{\frac{1}{2}}} \\ &= \lim_{t\to\infty} \sup \frac{|\sum_{k=1}^{t} Y_i^t|}{\left(2(\frac{1}{1-r}t^{1-r}) \log\log(\frac{1}{1-r}t^{1-r})\right)^{\frac{1}{2}}} \\ &\leq \lim_{t\to\infty} \sup \frac{|\sum_{k=1}^{t} Y_i^t|}{(2B^t \log\log B^t)^{\frac{1}{2}}} = 1 \text{ a.s.}\end{aligned}$$

Note that for $0 < r < \frac{1}{2}$, there exists a T such that for all $t > T$, $\left(2(\frac{1}{1-r}t^{1-r}) \log\log(\frac{1}{1-r}t^{1-r})\right)^{\frac{1}{2}} < t^{\frac{1}{2}}$. Therefore,

$$\begin{aligned}&\limsup_{t\to\infty} \frac{|\ell_i(t) - E[\ell_i(t)]|}{t^{\frac{1}{2}}} \leq \\ &\limsup_{t\to\infty} \frac{|\ell_i(t) - E[\ell_i(t)]|}{\left(2(\frac{1}{1-r}t^{1-r}) \log\log(\frac{1}{1-r}t^{1-r})\right)^{\frac{1}{2}}} \leq 1 \text{ a.s.}\end{aligned}$$

■

REFERENCES

[1] J. S. Jordan, "Three problems in learning mixed-strategy Nash equilibria," *Games and Economic Behavior*, vol. 5, no. 3, pp. 368–386, Jul. 1993.

[2] D. Fudenberg, "Learning Mixed Equilibria," *Games and Economic Behavior*, vol. 5, pp. 320–367, 1993.

[3] H. P. Young, "The evolution of conventions," *Econometrica: Journal of the Econometric Society*, pp. 57–84, 1993.

[4] D. Fudenberg and D. K. Levine, *The theory of learning in games*. MIT press, 1998, vol. 2.

[5] M. Benaım and M. W. Hirsch, "Mixed equilibria and dynamical systems arising from fictitious play in perturbed games," *Games and Economic Behavior*, vol. 29, no. 1, pp. 36–72, Oct. 1999.

[6] J. Hofbauer and W. H. Sandholm, "On the global convergence of stochastic fictitious play," *Econometrica*, vol. 70, no. 6, pp. 2265–2294, 2002.
[7] D. S. Leslie and E. Collins, "Generalised weakened fictitious play," *Games and Economic Behavior*, vol. 56, no. 2, pp. 285–298, Aug. 2006.
[8] H. P. Young, *Strategic learning and its limits*. Oxford University Press on Demand, 2004, vol. 2002.
[9] J. Robinson, "An iterative method of solving a game," *The Annals of Mathematics*, vol. 54, no. 2, pp. 296–301, Sep. 1951.
[10] K. Miyasawa, "On the convergence of the learning process in a 2 x 2 non-zero-sum two-person game," DTIC Document, Tech. Rep., 1961.
[11] D. Monderer and L. S. Shapley, "Potential Games," *Games and Economic Behavior*, vol. 14, no. 1, pp. 124–143, May 1996.
[12] T. J. Lambert, M. A. Epelman, and R. L. Smith, "A fictitious play approach to large-scale optimization," *Operations Research*, vol. 53, no. 3, pp. 477–489, May 2005.
[13] V. V. Petrov, *Limit theorems of probability theory*. Oxford Science Publications, 1995.