A Computationally Efficient Implementation of Fictitious Play in a Distributed Setting

BRIAN SWENSON†*, SOUMMYA KAR† AND JOÃO XAVIER*

***Abstract*—The paper deals with distributed learning of Nash equilibria in games with a large number of players. The classical fictitious play (FP) algorithm is impractical in large games due to demanding communication requirements and high computational complexity. A variant of FP is presented that aims to mitigate both issues. Complexity is mitigated by use of a computationally efficient Monte-Carlo based best response rule. Demanding communication problems are mitigated by implementing the algorithm in a network-based distributed setting, in which player-to-player communication is restricted to local subsets of neighboring players as determined by a (possibly sparse, but connected) preassigned communication graph. Results are demonstrated via a simulation example.**

***Index Terms*—Games, Distributed Learning, Fictitious Play, Nash Equilibrium**

I. Introduction

The field of learning in games studies how groups of interacting agents can adaptively learn to coordinate their behavior through repeated interaction. One of the best studied game-theoretic learning algorithms is Fictitious Play (FP) [1]. FP serves as an archetype for many learning algorithms (e.g. [2]–[7]), and has itself been applied in a variety of settings including large-scale optimization [7], [8], dynamic programming [9], traffic routing [10], and cognitive radio [11]–[13].

The importance and usefulness of FP stems from its intuitively simple nature and provable learning properties in certain multiagent games. However, the practical value of FP is limited by the fact that it can become extremely difficult to implement in games with a large number of players.[1]

In this paper we address two issues with FP that make it difficult to implement in large-scale games: demanding communication requirements, and high computational complexity. In particular, classical FP assumes a form of all-to-all communication among players, and has a computational complexity that grows exponentially in the number of players.

In order to implement classical FP in a communication-efficient manner, [6], [14] proposed a network-based implementation of FP, which we refer to as distributed FP. In distributed FP, it is assumed that players are provided with a preassigned communication graph (possibly sparse, but connected) through which they must efficiently communicate all information relevant to the FP process. The work [6] showed that the fundamental learning properties of FP can be retained in this setting.

Several methods have been proposed to mitigate the problem of computational complexity in FP. Of particular interest to the present work is sampled FP [7]— a variant of FP in which players use a Monte Carlo method to approximate the mixed utility each round, thereby obviating any need for direct evaluation of the (computationally expensive) best response optimization problem. The drawback of this approach is that the number of samples that must be drawn each round grows without bound as the algorithm progresses; in particular, in round t the number of samples that must be drawn is of the order $\sqrt{t}$. In [15], the authors present computationally efficient sampled FP (CESFP)—a variant of sampled FP in which the number of samples drawn each round is uniformly bounded; in fact, in CESFP only one sample need be drawn each round of the learning process.

The main contribution of this paper is the presentation of a variant of FP that simultaneously addresses both of the aforementioned problems—i.e., is efficient in terms of both communication and computation requirements. In particular, we present an implementation of the (low complexity) CESFP algorithm in a communication-efficient distributed setting, as studied in [6].

The remainder of the paper is organized as follows. Section II sets up the notation to be used in the subsequent development. Section III reviews classical FP and some of its fundamental shortcomings in large games. Section IV reviews the CESFP algorithm. Section V presents the distributed-information framework that will be used in the distributed implementation of CESFP. Section VI presents our distributed implementation of the CESFP algorithm, states the formal convergence result, and presents the proof of the result. Section VIII demonstrates results via a simulated cognitive radio example. Section IX concludes the paper.

II. Preliminaries

A game in normal form is represented by the triple $\Gamma := (N, (Y_i, u_i)_{i\in N})$, where $N = \{1, \dots, n\}$ denotes the set of players, Y_i denotes the finite set of actions available to player i, and $u_i : \prod_{i\in N} Y_i \to \mathbb{R}$ denotes the utility function of player i. Denote by $Y := \prod_{i\in N} Y_i$ the joint action space.

In order to guarantee the existence of Nash equilibria it is necessary to consider the mixed-extension of Γ in which players are permitted to play probabilistic strategies. Let $m_i := |Y_i|$ be the cardinality of the action space of player i, and let

The work was partially supported by the FCT projects FCT [UID/EEA/5009/2013] and FCT [UID/EEA/50009/2013] through the Carnegie Mellon/Portugal Program managed by ICTI from FCT and by FCT Grant CMU-PT/SIA/0026/2009 and was partially supported by NSF grant ECCS-1306128.

†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/IST), LARSyS, Instituto Superior Técnico, University of Lisbon (jxavier@isr.ist.utl.pt).

[1]Throughout the paper, we use the terms agent and player interchangeably.

$\Delta_i := \{p \in \mathbb{R}^{m_i} : \sum_{k=1}^{m_i} p(k) = 1,\ p(k) \geq 0\ \forall k\}$ denote the set of mixed strategies available to player i—note that a mixed strategy is a probability distribution over the action space of player i. Denote by $\Delta^n := \prod_{i \in N} \Delta_i$, the set of joint mixed strategies. When convenient, we represent a mixed strategy $p \in \Delta^n$ by $p = (p_i, p_{-i})$ where p_i denotes the marginal strategy of player i and p_{-i} is a $(n-1)$ tuple containing the marginal strategies of the other players.

In the context of mixed strategies, we often wish to retain the notion of playing a deterministic action. For this purpose, let $A_i := \{e_1, \ldots, e_{m_i}\}$ denote the set of "pure strategies" of player i, where e_j is the j-th cannonical vector in $\mathbb{R}^{m_i}$ containing a 1 at position j and zeros otherwise.

The mixed utility function of player i is given by

$$U_i(p) := \sum_{y \in Y} u_i(y) p_1(y_1) \ldots p_n(y_n), \tag{1}$$

where $U_i : \Delta^n \to \mathbb{R}$. The set of Nash equilibria is given by $NE := \{p \in \Delta^n : U_i(p_i, p_{-i}) \geq U_i(p'_i, p_{-i}),\ \forall p'_i \in \Delta_i,\ \forall i \in N\}$. The distance of a distribution $p \in \Delta^n$ from a set $S \subset \Delta^n$ is given by $d(p, S) = \inf\{\|p - p'\| : p' \in S\}$. Throughout the paper $\|\cdot\|$ denotes the standard $\mathcal{L}_2$ Euclidean norm unless otherwise specified.

For $i \in N$ and a vector $\xi \in \mathbb{R}^{m_i}$, denote by $P_{\Delta_i}(\xi) := \arg\inf_{p_i \in \Delta_i} \|p_i - \xi\|$, the projection of ξ onto the set Δ_i. Note that since Δ_i is closed and convex, the projection of ξ onto Δ_i is unique.

There exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ rich enough to carry out the construction of the various random variables required in this paper. As a matter of convention, all equalities and inequalities involving random objects are to be interpreted almost surely (a.s.) with respect to the underlying probability measure unless otherwise stated.

This paper considers discrete-time repeated-play learning algorithms. Fix a normal form game Γ. Let players repeatedly face off in the game Γ, and for $t \in \{1, 2, \ldots\}$, denote by $a_i(t) \in A_i$, the action played by player i in round t. Let the n-tuple $a(t) = (a_1(t), \ldots, a_n(t))$ denote the joint action at time t. Let the empirical distribution of the actions of player i be given by $q_i(t) := \frac{1}{t} \sum_{s=1}^{t} a_i(s)$ and let the joint empirical distribution be given by the n-tuple $q(t) = (q_1(t), \ldots, q_n(t))$.

III. Fictitious Play

A sequence of actions $\{a(t)\}_{t \geq 1}$ such that for all[2] $t \geq 1$,

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

is referred to as a *fictitious play process*. Intuitively speaking, this describes a process where player i assumes (perhaps incorrectly) that opponents are using statistically-independent time-invariant mixed strategies, and hence player i picks her next-stage action as a myopic best response to the empirical distribution $q_{-i}(t)$ of opponent action history.

FP has been studied extensively to determine the classes of games in which it leads players to learn Nash equilibrium strategies [2], [3]. Of particular relevance to the multi-agent setting is a class of games known as potential games [16]. A game $\Gamma = (N, (Y_i, u_i(\cdot))_{i \in N})$ is said to be a potential game if there exists a potential function $\phi : Y \to \mathbb{R}$ such that for all $i \in N$, and all $y_{-i} \in Y_{-i}$

$$u_i(y_i, y_{-i}) - u_i(x_i, y_{-i}) = \phi(y_i, y_{-i}) - \phi(x_i, y_{-i}),\ \forall y_i, x_i \in Y_i.$$

The existence of a potential function means, intuitively speaking, that all player's utility functions are aligned in such way that players share a common underlying objective. It has been shown [16], [17] that FP leads players to learn NE strategies in potential games in the sense that $d(q(t), NE) \to 0$ as $t \to \infty$.

A. *Problems With Fictitious Play in Large Multi-Agent Games*

Despite theoretical convergence results, FP is impractical to implement in large multi-agent games. We discuss two major problems with FP that we seek to address in this paper.

1) High Computational Complexity: In order to choose a next-stage action in FP (see (2)), a player must compute $U_i(\alpha_i, q_{-i}(t))$ for each $\alpha_i \in A_i$. Recall from (1) that the mixed utility $U_i(\alpha_i, q_{-i}(t))$ is the expected value of $u_i(\cdot)$ when opponents use the probabilistic strategy $q_{-i}(t)$ (an $(n-1)$ dimensional probability distribution) and player i uses the pure strategy α_i. The complexity of computing this mixed utility grows exponentially in terms of the number of players [18].

2) Demanding Communication Requirements: In order to compute the set of best responses in (2), a player must first have perfect knowledge of $q_{-i}(t)$. In a large multi-agent game this is equivalent to requiring that players be capable of directly observing or instantaneously communicating with all other agents at all times.

In this paper we seek to resolve both of these issues in a unified manner. The Computationally Efficient Sampled FP (CESFP) algorithm of [15] mitigates complexity, but does not address communication requirements. The distributed-information setting presented in [6] addresses communication but does not mitigate complexity. In this paper we present an implementation of the CESFP algorithm of [15] implemented over a (possibly sparse, but connected) communication structure as proposed in [6].

IV. Computationally Efficient Sampled FP

In classical FP, the complexity of the best response computation quickly becomes impractical as the number of players grows large. In Computationally Efficient Sampled FP (CESFP), this problem is mitigated by avoiding direct evaluation of the best response computation (2). Instead, each player i forms an estimate $\hat{U}_i(\alpha_i, t)$ of the utility $U_i(\alpha_i, q_{-i}(t))$ for each $\alpha_i \in A_i$ that is updated recursively using a Monte-Carlo-type approach. CESFP follows the same fundamental behavior rule as FP, but uses the estimate $\hat{U}_i(t)$ as a surrogate for $U_i(\alpha_i, q_{-i}(t))$ in (2).

We refer to a sequence of actions $\{a(t)\}_{t \geq 1}$ generated according to the following algorithm as an *CESFP process*.

[2]In all variants of FP discussed in this paper, the initial action $a_i(1)$ may be chosen arbitrarily for all i.

A. Computationally Efficient Sampled FP Algorithm

Initialize

(i) For each $i \in N$, let $a_i(1)$ be arbitrary, let $q_i(1) = a_i(1)$, and let $\hat{U}(\alpha_i, 0) = 0,\ \forall \alpha_i \in A_i$.

Iterate $(t \geq 1)$

(ii) Player i draws a "test action" $a^*_{-i}(t)$ from the distribution[3] $q_{-i}(t)$, and updates the estimate $\hat{U}_i(\alpha_i, t)$ for each action $\alpha_i \in A_i$ according to the recursion,[4]

$$\hat{U}_i(\alpha_i, t) = (1 - \rho(t))\hat{U}_i(\alpha_i, t-1) + \rho(t)U_i(\alpha_i, a^*_{-i}(t)).$$

The player chooses a next stage action using the rule (compare with (2)):

$$a_i(t+1) \in \arg\max_{\alpha_i \in A_i} \hat{U}_i(\alpha_i, t).$$

(iii) The empirical distribution is updated to reflect the action just taken $q_i(t+1) = \frac{1}{t+1}\sum_{s=1}^{t+1} a_i(s)$, or equivalently, in recursive form $q_i(t+1) = q_i(t) + 1/(t+1)(a_i(t+1) - q_i(t))$.

B. Discussion

In [15], it is shown that that if Γ is a potential game and

A. 1. *The sequence* $\{\rho(t)\}_{t\geq 1}$ *is such that* $0 < \rho(t) \leq 1$, $\sum_{t\geq 1}(\rho(t))^2 < \infty$, *and* $\lim_{t\to\infty} \frac{\log t}{t\rho(t)} = 0$.

holds,[5] where $\{\rho_i(t)\}_{t\geq 1}$ is a weight sequence used to update the estimate $\hat{U}_i(\alpha_i, t)$, (see step (ii), above) then CESFP achieves learning in the sense that $\lim_{t\to\infty} d(q(t), NE) = 0$.

V. Distributed Setup

In the centralized setting it is assumed that agents are able to instantaneously observe the actions of all other agents. In this paper we consider a relaxation of this assumption in which agents are incapable of observing the actions of others, but are endowed with a preassigned communication graph through which they may exchange information with neighboring agents. Formally we assume:

A. 2. *Players are endowed with a preassigned communication graph* $G = (V, E)$*, where the vertices* V *represent the players and the edge set* E *consists of communication links (bidirectional) between pairs of players that can communicate directly. The graph* G *is connected.*

A. 3. *Players directly observe only their own actions.*

A. 4. *A player may exchange information with immediate neighbors, as defined by* G*, at most once for each iteration or round of the repeated play.*

For any learning algorithm, we refer to an implementation that meets **A.2–A.4** as a *distributed implementation* of the algorithm. The distributed implementation of an algorithm is not unique; in particular, there may be many information sharing protocols (e.g., dynamic consensus [19], or asynchronous gossip [20]) that allow agents to learn NE in the distributed setting. In the following section we present a distributed implementation of CESFP in which players disseminate information pertinent to the learning process using a dynamic consensus algorithm.

[3] Note that it is implicitly assumed that each player has precise knowledge of the joint empirical distribution $q(t)$. In order for player i to know $q(t)$ at each time step, she must be aware of the actions $a_j(t)$ of all other agents. In a network setting, this is equivalent to assuming all-to-all communication.

[4] Since the joint strategy $a^*_{-i}(t)$ consists of only pure strategies, the evaluation of the utility is relatively simple.

[5] In [15] the condition $\lim_{t\to\infty} \log t/(t\rho(t)) = 0$ as stated above is exchanged for the slightly weaker condition: $\lim_{t\to\infty} 1/(t\rho(t)) = 0$. However, in this paper, order to ensure convergence in the distributed setting, we require the stronger assumption.

VI. Distributed Implementation of Computationally Efficient Fictitious Play

A. Algorithm Setup

In our distributed implementation of CESFP—henceforth referred to as D-CESFP—players do not have precise knowledge of the empirical distribution $q(t)$; instead, each player i forms an estimate of the empirical distribution by communicating with neighboring players. Denote by $\hat{q}^i_j(t)$, the estimate that player i maintains of the empirical distribution of player j, and denote by $\hat{q}^i(t) := (\hat{q}^i_1(t), \ldots, \hat{q}^i_n(t))$ the estimate that player i maintains of the joint empirical distribution.

Furthermore, in D-CESFP, each player i forms an estimate of the utility $U_i(\cdot, q_{-i}(t))$. Let $\hat{U}_i(\alpha_i, t) \in \mathbb{R}$ denote the estimate player i maintains of $U_i(\alpha_i, q_{-i}(t))$ for each $\alpha_i \in A_i$, $t \in \{0, 1, \ldots\}$.

In D-CESFP, player i forms her estimate of the empirical distribution $\hat{q}^i(t)$ by exchanging information with neighboring players. Let W be a weighting matrix satisfying the following assumption:

A. 5. W *is symmetric, doubly stochastic, aperiodic, irreducible. Furthermore,* W *is conformant to the graph topology* G*; i.e,* $w_{ij} > 0$ *only if* G *contains an edge from* i *to* j*.*

The matrix $W = (w_{ij})_{i,j=1}^n$ will be used to specify the weighting constants in the distributed update of $\hat{q}^i(t)$ (see steps (i) and (iv) of the D-CESFP algorithm below).

B. Some Additional Definitions

The following notation is used to facilitate a compact description of the algorithm. Let $s = \sum_{k\in N} m_k$. Let $q'_i(t)$ be an augmented (zero-stuffed) vector representing the empirical distribution of player i such that

$$q'_i(t) := (0, \ldots 0, nq_i(t), 0 \ldots, 0) \in \mathbb{R}^s.$$

The augmented vector $q'_i(t)$ matches the general structure of $\hat{q}^i(t)$, but in the place of $q^i_i(t)$ we substitute in $n \cdot q_i(t)$ (a scaled copy of the true empirical distribution) and set all other entries to zero.

For $i \in N$, $\xi = (\xi_1, \ldots, \xi_{i-1}, \xi_{i+1}, \ldots, \xi_n)$, $\xi_j \in \mathbb{R}^{m_j}$ let

$$P_{\Delta_{-i}}(\xi) := (P_{\Delta_1}(\xi_1), \ldots, P_{\Delta_n}(\xi_n)),$$

where $P_{\Delta_i}(\xi_i)$ is the projection of ξ_i onto the set Δ_i as defined in Section II.

C. D-CESFP Algorithm

Initialize

(i) Each player i takes an arbitrary initial action $a_i(1)$, and the empirical distribution for player i is initialized as $q_i(1) = a_i(1)$. Player i initializes her local estimate of the joint empirical distribution as

$$\hat{q}^i(1) = \sum_{j \in \Omega_i \cup \{i\}} w_{ij} q'_j(1), \tag{3}$$

where Ω_i is the set of neighbors of player i. The utility estimate is initialized as $\hat{U}_i(\alpha_i, 0) = 0$ for all $\alpha_i \in A_i$.

Iterate $(t \geq 1)$

(ii) Player i draws an action $a^*_{-i}(t)$ as a random sample from the probability mass function[6] $P_{\Delta_{-i}}\left(\hat{q}^i_{-i}(t)\right)$. For each $\alpha_i \in A_i$ player i updates her estimate of the predicted utility according to the rule:

$$\hat{U}_i(\alpha_i, t) = (1 - \rho(t))\hat{U}_i(\alpha_i, t-1) + \rho(t) U_i(\alpha_i, a^*_{-i}(t)). \tag{4}$$

(iii) Player i chooses her stage $(t+1)$ action as a best response to the predicted utility $\hat{U}(\cdot, t)$; i.e.,

$$a_i(t+1) \in \arg\max_{\alpha_i \in A_i} \hat{U}(\alpha_i, t). \tag{5}$$

The local empirical distribution of player i is recursively updated to include the action just taken, i.e.,

$$q_i(t+1) = q_i(t) + \frac{1}{t+1}\left(a_i(t+1) - q_i(t)\right).$$

(iv) Player i updates her estimate of the joint empirical distribution using the following rule:

$$\hat{q}^i(t+1) = \sum_{j \in \Omega_i \cup \{i\}} w_{ij}\left(\hat{q}^j(t) + q'_j(t+1) - q'_j(t)\right). \tag{6}$$

D. Main result

The following result shows that players engaged in a D-CESFP process asymptotically learn a Nash equilibrium.

Theorem 1. *Let Γ be a potential game. Let $\{a(t), q(t)\}_{t \geq 1}$, where $q(t) := (q_1(t), \ldots, q_n(t))$ be computed according to the D-CESFP algorithm of Section VI-C, and assume **A.1**–**A.5** hold. Then players learn a Nash equilibrium in the sense that $\lim_{t\to\infty} d(q(t), NE) = 0$. Furthermore, each player i achieves asymptotic strategy learning in the sense that $\lim_{t\to\infty} \|\hat{q}^i_j(t) - q_j(t)\| = 0$ for all $j \in N$.*

Proof. We will prove the result by showing that there exists a sequence $\{\epsilon_t\}_{t\geq 1}$ such that $\epsilon_t \to 0$ and $U_i(a_i(t+1), q_{-i}(t)) \geq \max_{\alpha_i \in A_i} U_i(\alpha_i, q_{-i}(t)) - \epsilon_t$ for all i. By [4], Corollary 5, this is sufficient to ensure $d(q(t), NE) \to 0$ in potential games.

By Lemma 1, there holds $\lim_{t\to\infty} \|\hat{q}^i(t) - q(t)\| = 0$. Since the map $P_{\Delta_{-i}}$ is Lipschitz continuous, this implies $\lim_{t\to\infty} \|P_{\Delta_{-i}}\left(\hat{q}^i_{-i}(t)\right) - q_{-i}(t)\| = 0$. By Lipschitz continuity of U_i, this implies that

$$\lim_{t\to\infty} |U_i(\alpha_i, P_{\Delta_{-i}}\left(\hat{q}^i_{-i}(t)\right)) - U_i(\alpha_i, q_{-i}(t))| = 0, \ \forall \alpha_i, \ \forall i. \tag{7}$$

Invoking assumption **A.1**, by Lemma 2 of [15] there holds,

$$\lim_{t\to\infty} |\hat{U}_i(\alpha_i, t) - U_i(\alpha_i, P_{\Delta_{-i}}\left(\hat{q}^i_{-i}(t)\right))| = 0, \ \forall \alpha_i, \ \forall i.$$

[6]Given the distributed update rule used to compute $\hat{q}^i(t)$ in step iv, it is possible that $\hat{q}^i(t)$ may sometimes leave the probability simplex. The projection operation guarantees that a player has a valid probability distribution from which to draw the sample $a^*_{-i}(t)$.

Combining this with (7) gives,

$$\lim_{t\to\infty} |\hat{U}_i(\alpha_i, t) - U_i(\alpha_i, q_{-i}(t))| = 0, \ \forall \alpha_i, \ \forall i.$$

Combining this with the D-CESFP action rule (5), we see that there exists a sequence $\{\epsilon_t\}_{t\geq 1}$ such that $\epsilon_t \to 0$ and

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

and the desired result holds. □

E. Discussion

In the above result, the joint empirical distribution $q(t)$ converges to the set of Nash equilibria. Because of **A.3**, player i may not have precise knowledge of the empirical distribution $q(t)$. However, the above result states that player i's estimate of the empirical distribution of opponent j, $\hat{q}^i_j(t)$ converges to the true empirical distribution of opponent j, $q_j(t)$. Hence, each player i achieves asymptotic strategy learning in the sense that her estimate of the joint empirical distribution converges to the set of Nash equilibria.

The distributed implementation of CESFP given above addresses both the problems of computational complexity and demanding communication associated with classical FP. The mitigation in complexity can be seen in (4) and (5)—rather than compute an expected value over $(n-1)$ dimensional space as required in (2), players simply compute the utility of a randomly sampled pure strategy a^*_{-i} (a relatively easy task) and use the recursive rule (4) to obtain an estimate of the utility $U(\alpha_i, q_{-i}(t))$.

The communication problem of classical FP is mitigated by implementing the algorithm in a setting where players are assumed to have limited ability to view the actions of other players, as in **A.3**, and interagent communication is restricted to local neighborhoods, as in **A.2**, **A.4**.

VII. Intermediate Results

Lemma 1. *Assume **A.2**–**A.5**. For $i, j \in N$, let $\hat{q}^i_j(t)$ be as defined in Section VI-B. Let the estimates $\hat{q}^i(t)$ be formed as in Section VI-C. Then $\lim_{t\to\infty} \|\hat{q}^i_j(t) - q_j(t)\| = 0, \ \forall i, j \in N$.*

Proof. Let $\bar{q}'(t) = \frac{1}{n}\sum_{i=1}^{n} q'_i(t)$ be the average of the augmented empirical distributions as defined in Section VI-B. Note that $\bar{q}'(t) \in \mathbb{R}^s$ is, in fact, a vector which stacks the true empirical distributions, $\bar{q}'(t) = (q_1(t), \ \cdots, \ q_n(t))$. Thus, by solving for $\bar{q}'(t)$, players are in fact solving for the true empirical distribution.

Let $k \in \{1, \ldots, s\}$, and let $\tilde{q}_k(t) \in \mathbb{R}^n$ with

$$\tilde{q}_k(t) := (\hat{q}^1(t,k), \ldots, \hat{q}^n(t,k)),$$

where $\hat{q}^i(t,k)$ is the k-th entry of the vector $\hat{q}^i(t)$. Note that the incremental change in any element of $\tilde{q}_k(t)$ is bounded by $1/t$, and that $\tilde{q}_k(t)$ is updated in a manner fitting the template of Lemma 1 of [6]. Invoking Lemma 1 of [6] (where (3) is to be seen as (6) for $t = 0$ and $\hat{q}^i(0) = 0$, $q_i(0) = 0$) gives

$$\lim_{t\to\infty} \|\tilde{q}_k(t) - \mathbf{1}\bar{q}'(t,k)\| = 0,$$

or equivalently, $\lim_{t\to\infty} \|\hat{q}^i(t,k) - \bar{q}'(t,k)\| = 0,\ \forall i$. Since this holds for all $k \in \{1,\ldots,s\}$ and $\bar{q}'(t) = q(t)$, (where $q(t)$ is considered as a vector), this gives the desired result. $\square$

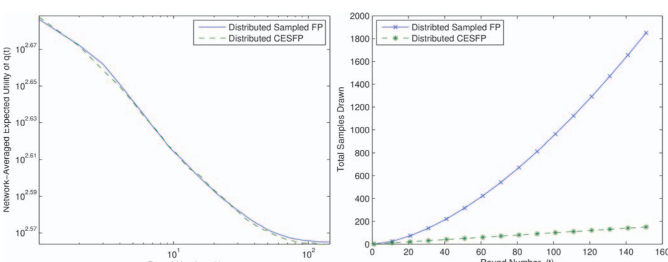
VIII. Simulation Example

We simulated D-CESFP in a simple cognitive radio example. Let $N = \{1,\ldots n\}$ indicate a set of users (or players), and let Ch indicate a finite collection of permissible frequency channels shared by all users (i.e., $Y_i = Ch,\ \forall i \in N$). For $y \in Y$, and $k \in Ch$, let $\sigma_k(y)$ denote the number of users on channel k under the strategy y. Further, for $k \in Ch$ and $\ell \in \{0,1,2,\ldots\}$, let $c_k(\ell)$ denote the cost of using channel k when there are ℓ users occupying the channel. Let the utility for player i be given by $u_i(y) = -c_{y_i}(\sigma(y))$. This game is classified as a congestion game—a known type of potential game.

We simulated the D-CESFP and a distributed implementation of Sampled FP[7] (D-Sampled FP) in the cognitive radio example given above with 30 users and 15 channels. The D-Sampled FP algorithm used a sample-rate parameter of $k_t = \lfloor t^{.6} \rfloor$ and D-CESFP used a parameter $\rho(t) = t^{-.6},\ \forall t$. All simulations used the same randomly generated initial conditions. In the distributed implementation of both algorithms the communication graph was given by the random geometric graph. The weights w_{ij} (see (6)) were derived using the Metropolis-Hastings rule [21].

Figure 1(a) shows a logarithmic plot of the networked-averaged expected utility of $q(t)$ in the distributed algorithms. The trend is consistent with convergence to NE and suggests similar per-iteration performance characteristics for both algorithms.

Figure 1(b) shows a plot of the number of samples drawn at each node up to and including round t. D-CESFP is able to achieve a similar quality solution to D-Sampled FP despite drawing far fewer total samples.



(a) (b)

Fig. 1. (a) Utility of $q(t)$ in Distributed CESFP and Distributed Sampled FP; (b) Total samples drawn by iteration t.

IX. Conclusions

Classical fictitious play (FP) can be difficult to implement in large-scale settings due to demanding communication requirements and high computational complexity. We have presented an approach that mitigates both of these issues. Demanding communication requirements are mitigated by restricting inter-agent communication to a local subset of neighboring players. Computational complexity is mitigated by using a best response rule based on Computationally Efficient Sampled FP.

[7] The distributed implementation of Sampled FP was formed in the same manner as distributed implementation of CESFP in Section VI. Using reasoning similar to the proof of Theorem 1 it can be shown that the distributed implementation of Sampled FP converges to NE in the same manner as in Theorem 1.

References

[1] G. W. Brown, *"Iterative Solutions of Games by Fictitious Play" In Activity Analysis of Production and Allocation*, T. Coopmans, Ed. New York: Wiley, 1951.

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

[3] H. P. Young, *Strategic learning and its limits*. Oxford University Press, 2004.

[4] D. S. Leslie and E. Collins, "Generalised weakened fictitious play," *Games and Economic Behavior*, vol. 56, no. 2, pp. 285–298, 2006.

[5] J. R. Marden, G. Arslan, and J. S. Shamma, "Joint strategy fictitious play with inertia for potential games," *IEEE Transactions on Automatic Control*, vol. 54, no. 2, pp. 208–220, 2009.

[6] B. Swenson, S. Kar, and J. Xavier, "Empirical centroid fictitious play: an approach for distributed learning in multi-agent games," *IEEE Transactions on Signal Processing*, vol. 63, no. 15, pp. 3888–3901.

[7] 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, 2005.

[8] T. J. Lambert and H. Wang, "Fictitious play approach to a mobile unit situation awareness problem," *Univ. Michigan, Tech. Rep*, 2003.

[9] M. Epelman, A. Ghate, and R. L. Smith, "Sampled fictitious play for approximate dynamic programming," *Computers & Operations Research*, vol. 38, no. 12, pp. 1705–1718, 2011.

[10] A. Garcia, D. Reaume, and R. L. Smith, "Fictitious play for finding system optimal routings in dynamic traffic networks," *Transportation Research Part B: Methodological*, vol. 34, no. 2, pp. 147–156, 2000.

[11] K. Dabcevic, A. Betancourt, L. Marcenaro, and C. S. Regazzoni, "A fictitious play-based game-theoretical approach to alleviating jamming attacks for cognitive radios," in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*. IEEE, 2014, pp. 8158–8162.

[12] ——, "Intelligent cognitive radio jamming-a game-theoretical approach," *EURASIP Journal on Advances in Signal Processing*, vol. 2014, no. 1, p. 171, 2014.

[13] B. Wang, Y. Wu, and K. R. Liu, "Game theory for cognitive radio networks: An overview," *Computer networks*, vol. 54, no. 14, pp. 2537–2561, 2010.

[14] B. Swenson, S. Kar, and J. Xavier, "Distributed learning in large-scale multi-agent games: A modified fictitious play approach," in *46th Asilomar Conference on Signals, Systems, and Computers*, Pacifc Grove, CA, USA, 2012, pp. 1490 – 1495.

[15] ——, "A computationally efficient implementation of fictitious play for large-scale games," 2015, submitted for journal publication, http://arxiv.org/abs/1506.04320.

[16] D. Monderer and L. S. Shapley, "Potential Games," *Games and Economic Behavior*, vol. 14, no. 1, pp. 124–143, 1996.

[17] ——, "Fictitious play property for games with identical interests," *Journal of Economic Theory*, vol. 68, no. 1, pp. 258–265, 1996.

[18] N. Nisan, *Algorithmic game theory*. Cambridge University Press, 2007.

[19] S. Rajagopalan and D. Shah, "Distributed averaging in dynamic networks," *IEEE jornal of selected topics in signal processing*, vol. 5, no. 4, pp. 845–854, 2011.

[20] A. G. Dimakis, S. Kar, J. M. Moura, M. G. Rabbat, and A. Scaglione, "Gossip algorithms for distributed signal processing," *Proceedings of the IEEE*, vol. 98, no. 11, pp. 1847–1864, 2010.

[21] S. Chib and E. Greenberg, "Understanding the Metropolis-Hastings algorithm," *The American Statistician*, vol. 49, no. 4, pp. 327–335, 1995.