Stochastic Fictitious Play with Continuous Action Sets

S.Perkins[a], D.S.Leslie[a]

[a] *School of Mathematics, University of Bristol, University Walk, Clifton, Bristol, BS8 1TW, England*

Abstract

Continuous action space games form a natural extension to normal form games with finite action sets. However, whilst learning dynamics in normal form games are now well studied, it is not until recently that their continuous action space counterparts have been examined. We extend stochastic fictitious play to the continuous action space framework. In normal form games the limiting behaviour of a discrete time learning process is often studied using its continuous time counterpart via stochastic approximation. In this paper we study stochastic fictitious play in games with continuous action spaces using the same method. This requires the asymptotic pseudo-trajectory approach to stochastic approximation to be extended to Banach spaces. In particular the limiting behaviour of stochastic fictitious play is studied using the associated smooth best response dynamics on the space of finite signed measures. Using this approach, stochastic fictitious play is shown to converge to an equilibrium point in single population negative definite games, two-player zero-sum games and N-player potential games, when they have Lipschitz continuous rewards.

Keywords: stochastic fictitious play, learning in games, continuous action set games, abstract stochastic approximation.

1. Introduction

Continuous action space games form a natural extension to normal form games with finite action sets. However whilst learning dynamics in normal form games are now well studied (e.g. Fudenberg and Levine, 1998) it is not until recently that their continuous action space counterparts have been examined. Oechssler and Riedel (2001) and Lahkar (2012) provide existence and uniqueness results for two of the most commonly studied evolutionary dynamics; the replicator dynamics and logit best response, in the single population scenario. Further results along similar lines are given by Oechssler and Riedel (2002); Seymour (2002); Cressman (2005); Cressman et al. (2006) and Hofbauer et al. (2009). Although these dynamics have been studied in continuous time for games with continuous action spaces there are few existing convergence result for discrete time learning: Hofbauer and Sorin (2006) show that in a two-player zero-sum continuous action games the average action of a fictitious play process converges to a global attractor of the best response dynamic; Chen and White (1998) investigate a seemingly impossible stochastic fictitious play model (see Section 2.2 for a discussion). We study a stochastic fictitious play learning processes in continuous action space games, developing the necessary stochastic approximation tools in the process. Extending the existing dynamical systems results of Lahkar (2012) to the N-player case and combining this with our enhanced stochastic approximation theory we study stochastic fictitious play for single-population games and N-player games. Convergence of the single population is shown in negative definite games and convergence of beliefs to mixed strategies is shown for two-player zero-sum and N-player potential games with continuous action spaces and Lipschitz continuous reward function. This extends the previous results of Hofbauer and Sandholm (2002) and Hofbauer and Hopkins (2005) for stochastic fictitious play in normal form games with finite action sets.

In a continuous action space game all actions are selected from an uncountably infinite action set S. Often S is taken to be a compact subset of $\mathbb{R}$, and this is the approach we take throughout. In discrete-action games a mixed strategy is described by a probability mass function on the action set. In continuous action space games this approach must be extended. Let $\mathcal{B}$ denote the Borel σ-algebra on S. Define $\mathcal{M}^e(S,\mathcal{B})$ as the space of finite signed measures on S, which is a Banach space when endowed with a suitable topology, and let $\mathcal{P}(S,\mathcal{B})$ denote the subset of $\mathcal{M}^e(S,\mathcal{B})$ consisting of all probability measures. In a continuous action space game a mixed strategy is a probability measure in $\mathcal{P}(S,\mathcal{B})$.

As with normal form games, if the population interpretation is being used then $P \in \mathcal{P}(S,\mathcal{B})$ is a particular population and selecting an action $s \in S$ using P is interpreted as selecting a member in the population who plays pure strategy s. Alternatively, the strategy interpretation can be used meaning that $\pi \in \mathcal{P}(S,\mathcal{B})$ is the strategy of a particular player and an action $s \in S$ is a random action selected using π. In Section 3 the single population interpretation is taken whilst in Section 4 we consider the strategy interpretation in N-player games.

Dynamical systems, such as the replicator dynamics or logit best response, can be used to study a population P as a process evolving in continuous time on the set of probability measures $\mathcal{P}(S,\mathcal{B})$. The evolution of a population on a continuous action space has been studied in recent years by Oechssler and Riedel (2001, 2002); Seymour (2002); Cressman (2005); Cressman et al. (2006); Hofbauer et al. (2009) and Lahkar (2012). However, here we are interested stochastic fictitious play in a continuous action game. In our stochastic fictitious play learning process a player uses the logit best response to select an action. Each player observes the joint action profile and directly uses these point observations to update their beliefs of each player strategy. Let $P_n \in \mathcal{P}(S,\mathcal{B})$ be the beliefs at iteration n and let $s_{n+1} \in S$ be the action selected at iteration $n+1$. A stochastic fictitious play process for continuous action space games is given by the recursion,

$$P_{n+1} = P_n + \alpha_{n+1}\Big[\delta_{s_{n+1}} + P_n\Big],$$

where δ_x is a Dirac measure at $x \in S$. This is the natural extension to the traditional, normal form game, stochastic fictitious play of Fudenberg and Kreps (1993), Benaïm and Hirsch (1999), Hofbauer and Sandholm (2002) and Hofbauer and Hopkins (2005).

Many discrete time learning processes fit the stochastic approximation framework of Benaïm (1999). An iterative process $\{\theta_n\}_{n\in\mathbb{N}}$ is described by the process,

$$\theta_{n+1} = \theta_n + \alpha_{n+1}\Big[F(\theta_n) + U_{n+1}\Big], \tag{1.1}$$

where $F(\cdot) : \Theta \to \Theta$ is a continuous map, $\{U_n\}_{n\in\mathbb{N}}$ is a noise sequence and $\{\alpha_n\}_{n\in\mathbb{N}}$ are learning rates. If $\Theta = \mathbb{R}^K$, and under some mild conditions, standard stochastic approximation results (e.g. Benaïm, 1999) give that the limiting behaviour of (1.1) can be studied using the ordinary differential equation (ODE)

$$\frac{d\theta}{dt} = F(\theta). \tag{1.2}$$

This is commonly known as the ODE method of stochastic approximation, originally proposed by Ljung (1977) and extended by many authors including Kushner and Clark (1978), Kushner and Yin (1987a,b) Borkar (1997, 1998, 2008), Benaïm (1999) and Benaïm et al. (2003).

In order to produce a framework to study the limiting behaviour of an infinite dimensional, discrete time, stochastic learning process we extend the standard stochastic approximation framework to allow the dynamics to be on a general Banach space Θ and, in particular, the space of finite signed measures $\mathcal{M}^e(S, \mathcal{B})$. This approach has previously been taken by Walk (1977); Berger (1986); Walk and Zsidó (1989); Shwartz and Berman (1989); Koval (1998); Dippon and Walk (2006) for general Hilbert or Banach spaces. Using standard functional analysis techniques the differential equation (1.2) can be extended for $\theta \in \Theta$ (Luenberger, 1969). In particular Shwartz and Berman (1989) use a similar approach to the ODE method described above to extend stochastic approximation to general Banach spaces. We give an update of stochastic approximation on a general Banach space (often called abstract stochastic approximation) to the now common asymptotic pseudo-trajectory approach of Benaïm (1999). The assumptions used for standard stochastic approximation are generalised for a process in the form of (1.1) with $\theta_n \in \Theta$. Under these generalised assumptions the linear interpolation of the $\{\theta_n\}_{n\in\mathbb{N}}$ process is an asymptotic pseudo-trajectory to the ordinary differential equation on the Banach space Θ given by (1.2).

One of the more challenging aspect of stochastic approximation is to verify that the noise process $\{U_n\}_{n\in\mathbb{N}}$ satisfies the appropriate assumptions originally stated by Kushner and Clark (1978); see assumption (A3) below. This is an area where the previous work on abstract stochastic approximation has struggled. For example Koval (1998) considers the simple case when $\{U_n\}_{n\in\mathbb{N}}$ is an i.i.d. noise process whilst Shwartz and Berman (1989) prove a very weak convergence result for a particular process which again uses independent noise. We provide criteria analogous to the martingale noise assumptions in $\mathbb{R}^K$ which guarantee that the noise condition holds on a class of Banach spaces. This result is extended for a suitable noise condition on the space of finite signed measures with the bounded Lipschitz norm. This gives criteria under which we can study the limiting behaviour of noisy learning processes on $\mathcal{P}(S, \mathcal{B}) \subset \mathcal{M}^e(S, \mathcal{B})$ using a deterministic dynamical system.

We combine our stochastic approximation results with the dynamical systems result of Lahkar (2012) to prove convergence of a stochastic fictitious play-like process in negative definite, single population games with Lipschitz continuous reward structure. In addition, we extend the existence and uniqueness results of Lahkar (2012) to the N-population case. This allows us to study a stochastic fictitious play algorithm for N-player continuous action space games. We prove the global convergence of the logit best response dynamics for two-player zero-sum and N-player potential games with continuous action spaces and Lipschitz continuous reward function. Convergence of the stochastic fictitious play algorithm follows by applying our stochastic approximation results.

This paper is organised in the following manner. Section 2 contains the background and an extension to the asymptotic pseudo-trajectory approach of Benaïm (1999) for stochastic approximation on a Banach space. This builds on the previous work in this area by Shwartz and Berman (1989) and links the classical approach to abstract stochastic approximation with the now more common asymptotic pseudo-trajectory approach to stochastic approximation on $\mathbb{R}^K$. Importantly in Section 2.2 we give a set of conditions, similar to the standard martingale difference noise assumptions on $\mathbb{R}^K$, such that we can control the noise term for stochastic approximation on $\mathcal{M}^e(S, \mathcal{B})$. In Section 3 we consider the logit best response dynamic for single population, continuous action games. The previous work on the continuous time dynamical system by Lahkar (2012) is briefly reviewed before a learning variant is presented. Using the stochastic approximation framework of Section 2 and the noise condition on $\mathcal{M}^e(S, \mathcal{B})$ we are able to conclude that for a wide class of games the limiting behaviour of our discrete time, stochastic learning process can be studied using the deterministic, continuous time dynamics of Lahkar (2012). In Section 4 we extend this analysis to N-player, continuous action space games. We show that stochastic fictitious play will converge to an equilibrium for two-player zero-sum and N-player potential games with continuous action spaces and Lipschitz continuous reward function. Throughout this work many of the proofs are relegated to an appendix.

2. Stochastic Approximation on a Banach Space

In normal form games stochastic approximation is used to study the limiting behaviour of stochastic fictitious play via the continuous time smooth best response dynamics. In a continuous action space game the beliefs will be a probability measure in the set $\mathcal{P}(S, \mathcal{B}) \subset \mathcal{M}^e(S, \mathcal{B})$ which will evolve over time. When associated with an appropriate distance metric $\mathcal{M}^e(S, \mathcal{B})$ is a Banach space containing all finite signed measures. Therefore, in order to study the limiting behaviour of stochastic fictitious play in continuous action space games the standard stochastic approximation framework is extended to the Banach space setting.

The ideas in this paper build on a rich history of work on stochastic approximation. In particular we make use of the asymptotic pseudo-trajectory framework of Benaïm (1999). Throughout this section we will use (M, d) to represent a metric space and $(\Theta, \|\cdot\|_\Theta)$ will denote a Banach space. For simplicity these are often written as M and Θ respectively. Since any Banach space is a metric space it should be clear that any statements for M also hold for Θ, whilst the reverse statement is not true in general.

2.1. Asymptotic Pseudo-Trajectory Approach

To begin we present and describe asymptotic pseudo-trajectories which were first introduced in stochastic approximation by Benaïm and Hirsch (1996) and expanded upon by Benaïm and Hirsch (1999) and Benaïm (1999).

Definition 2.1. A *semiflow* Φ on M is a continuous map $\Phi : \mathbb{R}^+ \times M \to M$, $(t, \theta) \to \Phi_t(\theta)$, such that,

(1) $\Phi_0(\theta) = \{\theta\}$;

(2) $\Phi_{t+s}(x) = \Phi_t\big(\Phi_s(\theta)\big)$, for any $t, s \geq 0$.

We assume unless otherwise stated that Φ is a semiflow.

Definition 2.2. A continuous function $\psi : \mathbb{R}^+ \to M$ is an *asymptotic pseudo-trajectory* for Φ if for any $T > 0$,

$$\lim_{t\to\infty} \sup_{0\le s\le T} d\Big(\psi(t+s), \Phi_s\big(\psi(t)\big)\Big) = 0.$$

Intuitively for any $T > 0$, $s \in [0, T]$ the map $\psi(t + s)$ remains close to the semiflow Φ with arbitrary precision for large enough t.

The *ω-limit set* of a semiflow Φ and the *limit set* of an asymptotic pseudo-trajectory ψ are defined in the same manner,

$$\omega(\Phi) := \bigcap_{t\ge 0} \overline{\Phi_{[t,\infty]}} \quad \text{and} \quad L(\psi) := \bigcap_{t\ge 0} \overline{\psi\big([t,\infty]\big)}.$$

Luenberger (1969) outlines the standard functional analysis techniques which can be used to extend the differential equation (1.2) defined on a Euclidean space to the more general Banach space setting. If Θ is a Banach space and we have a uniformly continuous map $F(\cdot) : \Theta \to \Theta$ then as in (1.2) let

$$\frac{d\theta}{dt} = F(\theta). \tag{2.1}$$

As in the standard Euclidean space case (2.1) will define a semiflow on Θ, and the limiting behaviour of an asymptotic pseudo-trajectory to this semiflow can be studied through the deterministic differential equation (2.1).

We now define a discrete time process on Θ in the form of a stochastic approximation process of Benaïm (1999). Let $\theta_0 \in \Theta$. Define $\{\theta_n\}_{n\in\mathbb{N}}$ via the recursive process,

$$\theta_{n+1} = \theta_n + \alpha_{n+1}\Big[F(\theta_n) + U_{n+1}\Big], \tag{2.2}$$

where $U_{n+1} \in \Theta$ is a random term in Θ. This will mean that for all $n \in \mathbb{N}$, $\theta_n \in \Theta$. Let $\tau_0 := 0$, $\tau_n := \sum_{k=1}^{n} \alpha_k$ and $m(t) := \sup\{k \ge 0; t \ge \tau_k\}$. Define the continuous time interpolation of $\{\theta_n\}_{n\in\mathbb{N}}$ such that for $s \in [0, \alpha_{n+1})$,

$$\bar{\theta}(\tau_n + s) := \theta_n + \frac{s}{\alpha_{n+1}}\Big[\theta_{n+1} - \theta_n\Big]. \tag{2.3}$$

We will need the following assumptions;

(A1) For all $n \in \mathbb{N}$, $\theta_n \in \Omega$, where $\Omega \subset \Theta$ is compact.

(A2) $F(\cdot) : \Theta \to \Theta$ is a uniformly continuous map such that for all $\theta \in \Omega$, $\|F(x)\|_\Theta < C$ for some $0 < C < \infty$.

(A3) For all $T > 0$

$$\lim_{n\to\infty} \sup_{k} \left\{ \left\| \sum_{i=n}^{k-1} \alpha_{i+1} U_{i+1} \right\|_\Theta ; k = n+1, \ldots, m(\tau_n + T) \right\} = 0.$$

(A4) A unique solution to the differential equation in (2.1) exists in Ω given any initial choice of $\theta_0 \in \Omega$.

Theorem 2.3. *Under the assumptions (A1)-(A4), $\bar{\theta}(\cdot) : \mathbb{R}^+ \to \Theta$, defined in* (2.3)*, is an asymptotic pseudo-trajectory to the semiflow Φ induced by the differential equation* (2.1)*.*

Proof. The proof is omitted as, other than the Euclidean norm $\|\cdot\|$ being replaced with the norm $\|\cdot\|_\Theta$, the proof is identical to Benaïm (1999, Proposition 4.1). □

We note that assumptions (A1)-(A4) are extensions to those used by Benaïm (1999) for standard stochastic approximation and are similar to those given by Shwartz and Berman (1989) for their convergence result for Banach space stochastic approximation. However, verifying these for a general Banach space as opposed to a Euclidean space can be difficult. In particular in Section 2.2 we provide conditions to verify the challenging assumption in (A3) for martingale noise in a Banach space which is suitable for studying learning in continuous action games.

Many further results can be taken from Benaïm (1999) to characterise the behaviour of asymptotic pseudo-trajectories. Here we present an example of how a Lyapunov function can be used to prove the global convergence of an asymptotic pseudo-trajectory. Initially we introduce a notion of stability for a dynamical system on M before presenting a result from Benaïm (1999).

Definition 2.4. A set A is *positively invariant* if $\Phi_t(A) \subset A$ for any $t \in \mathbb{R}^+$. The set A is *invariant* if $\Phi_t(A) = A$ for all $t \in \mathbb{R}^+$. A point $\tilde{\theta} \in M$ is an *equilibrium point* if $\Phi_t(\tilde{\theta}) = \tilde{\theta}$ for all $t \geq 0$.

Definition 2.5. Let $\Lambda \subset M$ be a compact invariant set for the semiflow Φ. A continuous function $V : M \to \mathbb{R}$ is called a *Lyapunov function* for Λ if for all $t \in \mathbb{R}^+$ the function $V(\Phi_t(\theta))$ is constant for all $\theta \in \Lambda$ and strictly decreasing for all $\theta \in M \backslash \Lambda$.

Theorem 2.6. *Let $\Lambda \subset M$ be a compact invariant set for the semiflow Φ and $V : M \to \mathbb{R}$ be a Lyapunov function for Λ, where $V(\Lambda) \subset \mathbb{R}$ has an empty interior. If $\psi(\cdot) : \mathbb{R}^+ \to M$ is an asymptotic pseudo-trajectory to Φ then both the ω-limit set of Φ and the limit set of ψ are contained in Λ.*

In Theorem 2.6 we have a method of determining the stability of a dynamical system on Θ and Theorem 2.3 gives criteria under which a discrete time stochastic process can be studied using this dynamical system on Θ. The remaining challenge is to show that the assumptions of Theorem 2.3 hold. In particular, (A3) is a challenging condition to verify. This is the focus of the following section.

2.2. Noise Criteria: The Space of Finite Signed Measures

It is natural to represent probability distributions on S as a measure since continuity does not have to be assumed as is the case when using density functions. In particular atoms, corresponding to positive probability on a particular action $s \in S$, can be accommodated. With $S \subset \mathbb{R}$ and $\mathcal{B}$ the Borel σ-algebra on S, define $\mathcal{M}^e(S, \mathcal{B})$ to be the vector space of finite signed measures. This means that for $\mu \in \mathcal{M}^e(S, \mathcal{B})$ there exists two finite measures on $(S, \mathcal{B})$, ν_1 and ν_2, such that for all $A \in \mathcal{B}$, $\mu(A) = \nu_1(A) - \nu_2(A)$.

The space of probability measures can be viewed as a subset of the vector space of finite signed measures. By using an appropriate norm we can consider the Banach space of finite signed measures, and in particular the subset of probability measures on this space.

The notion of convergence of a probability measure depends on the distance metric used on $\mathcal{M}^e(S, \mathcal{B})$. Often this space is assigned the variational norm metric which induces the strong topology, see Oechssler and Riedel (2001) or Seymour (2002). However, we take the now more common approach to use the weak topology and many of the reasons for this approach are discussed in detail by Oechssler and Riedel (2002).

Two metrics have commonly been used to induce the weak topology on $\mathcal{M}^e(S, \mathcal{B})$. Oechssler and Riedel (2002) and Hofbauer et al. (2009) use the Prohorov distance metric whilst Lahkar (2012) uses the bounded Lipschitz norm. Convergence in either of these metrics implies convergence in the other, meaning they induce the same topology. We follow the approach of Lahkar (2012) in using the bounded Lipschitz norm. For a bounded, Lipschitz continuous function $g : S \to \mathbb{R}$ define

$$\|g\|_{BL} := \sup_{x\in S} |g(x)| + \sup_{x\neq y} \frac{|g(x) - g(y)|}{|x - y|}. \tag{2.4}$$

Now let

$$\mathcal{BL} := \{g; g \text{ bounded \& Lipschitz continuous with } \|g\|_{BL} \leq 1\}, \tag{2.5}$$

be the set of bounded Lipschitz continuous functions with BL-norm bounded by 1. The dual BL^*-norm on $\mathcal{M}^e(S, \mathcal{B})$ is defined for $\mu \in \mathcal{M}^e(S, \mathcal{B})$ as

$$\|\mu\|_{BL^*} := \sup_{g\in\mathcal{BL}} \left| \int_S g(x)\mu(\,\mathrm{d}x) \right|. \tag{2.6}$$

Here we consider the space $(\mathcal{M}^e(S, \mathcal{B}), \|\cdot\|_{BL^*})$ which Lahkar (2012) shows is a Banach space. For the remainder of this paper we refer to $\mathcal{M}^e(S, \mathcal{B})$ as opposed to $(\mathcal{M}^e(S, \mathcal{B}), \|\cdot\|_{BL^*})$ with the understanding that $\mathcal{M}^e(S, \mathcal{B})$ will always be equipped with the weak topology induced by the bounded Lipschitz norm.

Let $\mathcal{P}(S, \mathcal{B})$ be the subset of $\mathcal{M}^e(S, \mathcal{B})$ consisting of the probability measures. We can use the abstract stochastic approximation result of Theorem 2.3 for a process $\{\theta_n\}_{n\in\mathbb{N}}$ such that $\theta_n \in \mathcal{P}(S, \mathcal{B})$ for all $n \in \mathbb{N}$. Under the weak topology on $\mathcal{M}^e(S, \mathcal{B})$

$$\|\mu\|_{BL^*} = 1,$$

for any probability measure $\mu \in \mathcal{P}(S, \mathcal{B})$, meaning the boundedness assumption in (A2) will be satisfied for a process in $\mathcal{P}(S, \mathcal{B})$. One additional reason to use the weak topology which is not explicitly given by Oechssler and Riedel (2002) is that under the weak topology the set of probability measures, $\mathcal{P}(S, \mathcal{B})$, is a compact subset of the space of signed measures $\mathcal{M}^e(S, \mathcal{B})$. Assumption (A1) requires that the iterative process remains in a compact set, which is clearly satisfied for processes in $\mathcal{P}(S, \mathcal{B})$ when $\mathcal{M}^e(S, \mathcal{B})$ is equipped with the weak topology. As observed by Oechssler and Riedel (2002), this compactness result does not remain true for the strong topology which immediately makes it more difficult to combine the strong topology on $\mathcal{M}^e(S, \mathcal{B})$ with the stochastic approximation framework from Section 2.

On the Banach space $\mathcal{M}^e(S, \mathcal{B})$ the noise term U_{n+1} is a random signed measure. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\{\mathcal{F}_n\}_{n\in\mathbb{N}}$ a filtration on $\mathcal{F}$.

Proposition 2.7. *Consider the stochastic approximation process given in* (2.2) *on the Banach space* $\mathcal{M}^e(S,\mathcal{B})$. *Assume that*

(B1) $\{\alpha_n\}_{n\in\mathbb{N}}$ *is deterministic with* $\sum_{n\in\mathbb{N}} \alpha_n^2 < \infty$.

(B2) For all $n \in \mathbb{N}$, U_{n+1} *is adapted and measurable with respect to* $\mathcal{F}_n$.

(B3) For all n

$$U_{n+1} = \delta_{s_{n+1}} - P_n,$$

where $P_n \in \mathcal{P}(S,\mathcal{B})$ *is a bounded, absolutely continuous measure which is deterministic given* $\mathcal{F}_n$ *and with associated density* p_n. $s_{n+1} \in S$ *is randomly drawn from the probability density function* p_n.

Then

$$\lim_{n\to\infty} \sup \left\{ \left\| \sum_{i=n}^{k-1} \alpha_{i+1} U_{i+1} \right\|_{BL^*} ; k = n+1, \ldots, m(\tau_n + T) \right\} = 0, \quad \textit{w.p. 1.}$$

The proof of this result comes from approximating the Dirac delta in U_n with a spike centered on the Dirac measure for all iterates. This spike and the absolutely continuous function p_n can then be studied in L^2. By applying the convergence result in Proposition A.1 for L^2 we conclude. This proof is given in Appendix B. We note that the earlier result of Chen and White (1998) assumes that beliefs are updated towards a symmetric, absolutely continuous distribution centered on the observed action $s_{n+1} \in S$ to ensure that $P_n \in L^2$ for all $n \in \mathbb{N}$. However, no such distribution exists if s_{n+1} is on the boundary of S, so their process is actually impossible. Instead we consider distributional beliefs, and approximate these with L^2 spikes only in the proof of Proposition 2.7.

As already noted the space of probability measures, $\mathcal{P}(S,\mathcal{B})$, is compact under the weak topology on $\mathcal{M}^e(S,\mathcal{B})$. Providing the map $F(\cdot) : \mathcal{P}(S,\mathcal{B}) \to \mathcal{P}(S,\mathcal{B})$ is uniformly continuous (A1) and (A2) will be satisfied. Proposition 2.7 can be used to verify (A3). As already demonstrated by Lahkar (2012), the infinite dimensional Picard-Lindelof theorem, proved by Zeidler (1986), provides a method for verifying (A4), the uniqueness of the solution to the differential equation on $\mathcal{M}^e(S,\mathcal{B})$ corresponding to (2.1). Hence, we have straightforward conditions under which we can apply the abstract stochastic approximation results of Theorem 2.3 to study the limiting behaviour of an iterative process (2.2) which remains in $\mathcal{P}(S,\mathcal{B})$. This will be our approach to analysing discrete time learning in continuous action space games.

3. Single Population, Continuous Action Space Games

In this section we present an application of Theorem 2.3 and Proposition 2.7 to a stochastic fictitious play-like learning algorithm, based on the logit best response, in single population continuous action games. This therefore generalises the results of Hofbauer and Sandholm (2002).

Recall that in a single population continuous action game, actions are taken from a compact action set $S \subset \mathbb{R}$ and let $\mathcal{B}$ be the Borel σ-algebra on S. A population is given by the combined play of its members and is a probability measure defined over the measurable space $(S, \mathcal{B})$. So for $A \in \mathcal{B}$, $P(A)$ is the proportion of players using a strategy in A. $\mathcal{P}(S, \mathcal{B})$ represents the set of all populations on S.

Finally, the game has a payoff function $r(x, y)$ for $x, y \in S$. In a single realisation of this game two players are selected randomly from the population. Each player has an associated pure strategy in S which, combined with $r(\cdot, \cdot)$, is used to determine the reward of each player. The expected payoff of an action $x \in S$ against a population $Q \in \mathcal{P}(S, \mathcal{B})$ is

$$E(x, Q) = \int_S r(x, y) Q(\,\mathrm{d}y). \tag{3.1}$$

In a similar manner, the expected payoff of the population P against a population Q is

$$E(P, Q) = \int_S \int_S r(x, y) P(\,\mathrm{d}x) Q(\,\mathrm{d}y). \tag{3.2}$$

In a standard abuse of notation we write $E(x, Q) \equiv E(\delta_x, Q)$ where δ_x is a Dirac measure which places all the probability mass at the action $x \in S$.

3.1. The Logit Best Response Dynamics

The stochastic fictitious play-like learning algorithm presented in Section 3.2 will be shown to approximate the logit best response dynamic. This dynamical system has been studied for continuous action, one population games by Lahkar (2012) and is discussed here. In normal form games the logit best response is the most studied of the smooth best response functions. Smooth best response functions arise from maximising perturbed payoff functions. When the perturbations are the entropy function the smooth best response becomes the logit best response (see Hofbauer and Sandholm (2002) for a full discussion). The full construction of the logit best response is given by Lahkar (2012) for the single population, continuous action case and we extend these details in Section 4 for the N-player scenario.

For $A \in \mathcal{B}$ and fixed $\eta > 0$ the logit best response to a population P is given by

$$LBR_\eta(P)(A) := \frac{\int_A \exp\{\eta^{-1} E(z, P)\} \,\mathrm{d}z}{\int_S \exp\{\eta^{-1} E(u, P)\} \,\mathrm{d}u}. \tag{3.3}$$

Lahkar (2012) shows that for any population $LBR_\eta(P)$ is absolutely continuous and therefore has a density,

$$l_\eta(P)(x) := \frac{\exp\{\eta^{-1} E(x, P)\}}{\int_S \exp\{\eta^{-1} E(u, P)\} \,\mathrm{d}u}. \tag{3.4}$$

Let $P(t)$ denote the population at time $t > 0$ and $P(t)(A)$ denotes the mass of the probability measure on the set $A \in \mathcal{B}$ at time $t > 0$. The logit best response dynamic on $\mathcal{M}^e(S, \mathcal{B})$ is given by the differential equation,

$$\dot{P}(t)(A) = LBR_\eta\big(P(t)\big)(A) - P(t)(A). \tag{3.5}$$

Let $\mathcal{LE}_\eta$ be the set of logit equilibria of (3.5) for a single population game. Lahkar (2012) gives criteria for $\mathcal{LE}_\eta$ to be non-empty and for the logit best response dynamic to have a unique solution. We show in Corollary C.3 that these conditions are satisfied if $r(x, y)$ is Lipschitz continuous in y for all $x \in S$, and throughout this section we shall assume that this is the case.

In addition, Lahkar (2012) provides a global convergence result for the logit best response. Firstly, we require the following definition which is taken from Hofbauer et al. (2009). We note that Lahkar (2012) presents this definition in the more general setting where games can be non-linear which is the extension to the notion of a stable game used by Hofbauer and Sandholm (2007). It suffices here to present the definition for linear games of Hofbauer et al. (2009).

Definition 3.1. A linear, single population game is *negative definite* if for all $P, Q \in \mathcal{P}(S, \mathcal{B})$

$$E(P - Q, P - Q) < 0.$$

The game is *negative semi-definite* if the inequality is replaced with a non-strict inequality.

As discussed by Hofbauer et al. (2009) the class of negative semi-definite games includes many common games including, for example, symmetric zero-sum games. The following result is taken from Lahkar (2012), where we note that by Corollary C.3 the assumptions of Lahkar (2012, Theorem 7.2) are satisfied under the assumptions here.

Theorem 3.2. *Assume that $r(x, y)$ is Lipschitz continuous in y for all $x \in S$, and that the game is negative semi-definite. Then a Lyapunov function, $V_\eta(\cdot) : \mathcal{P}(S, \mathcal{B}) \to \mathbb{R}$, exists for the logit best response dynamics* (3.5) *with attracting set of $\mathcal{LE}_\eta$ and $V_\eta(\mathcal{LE}_\eta)$ has an empty interior.*

The stochastic approximation framework in Section 2 will allow Theorem 3.2 to be applied to our stochastic fictitious play-like learning algorithm in Section 3.2.

3.2. Logit Best Response Learning

In this section we consider a single population stochastic fictitious play-like process analogous to that of Hofbauer and Sandholm (2002) but with a continuous action space. On each iteration of the game a proportion of the population revise their strategy, and adjust to a new action which is selected according to a logit best response to the current population. Alternative interpretations of the learning process are available (see Hofbauer and Sandholm (2002) for details). This discrete time logit best response learning process is therefore given by

$$P_{n+1} = P_n + \alpha_{n+1} \left[\delta_{s_{n+1}} - P_n\right], \tag{3.6}$$

where the action s_{n+1} is an action selected randomly from the logit best response density, $l_\eta(P_n)$.[1] The following theorem states the convergence result for the iterative process in

[1] We note that if $r(x, y) = xg(y) + h(y)$ for bounded $g(\cdot), h(\cdot) : S \to \mathbb{R}$ the cumulative distribution function (CDF) inversion method can used to select an action from the density $l_\eta(P_n)$ and if $r(x, y) = f(y)x^2 + xg(y) + h(y)$, for bounded, positive $f(\cdot) : S \to \mathbb{R}$, then an action can be selected from $l_\eta(P_n)$ using a truncated normal distribution. For other reward functions more advanced techniques would be required, which we do not address here.

(3.6) which is based upon the logit best response dynamic.

Theorem 3.3. *Assume that $r(x, y)$ is bounded and Lipschitz continuous in y for all $x \in S$ and*

$$\sum_{n \in \mathbb{N}} \alpha_n = \infty, \quad \sum_{n \in \mathbb{N}} \alpha_n^2 < \infty.$$

Then a linear interpolation of the stochastic fictitious play-like process (3.6) *is an asymptotic pseudo-trajectory to the logit best response dynamic* (3.5).

Proof. Firstly, note that for

$$V_{n+1} := \delta_{s_{n+1}} - LBR_\eta(P_n),$$

(3.6) can be written as

$$P_{n+1} = P_n + \alpha_{n+1}\Big[\Big(LBR_\eta(P_n) - P_n\Big) + V_{n+1}\Big], \tag{3.7}$$

which fits the general stochastic approximation form used in Theorem 2.3. Hence the claim will follow if we verify that (A1)-(A4) hold for (3.7).

We have already shown the compactness assumption (A1) to hold and the mean field continuity and boundedness assumption (A2) will also hold since the logit best response mean field, $LBR_\eta(P_n) - P_n$ is uniformly continuous (Lahkar, 2012).

With $r(x, y)$ Lipschitz continuous in y for all $x \in S$ we have proved the existence, uniqueness and continuity of solutions in Corollary C.3. This verifies that the existence of solution assumption (A4) holds.

Finally, we need to verify that the assumptions in Proposition 2.7 hold to show that the noise assumption (A3) holds. Let $\mathcal{F}_n$ represent the history of the iterative process (3.6) up to iteration $n \in \mathbb{N}$. The learning rate assumption (B1) is true since $\sum_n \alpha_n^2 < \infty$ by choice of $\{\alpha_n\}_{n \in \mathbb{N}}$. Clearly for all $n \in \mathbb{N}$, V_{n+1} is measurable with respect to $\mathcal{F}_n$ and hence will satisfy (B2). Now since $r(x, y)$ is bounded we know a maximum and a minimum value, $r_{\max}$ and $r_{\min}$, exist. As a result, for any $n \in \mathbb{N}$ and $x \in S$

$$l_\eta(P_n)(x) < \frac{\exp\{\eta^{-1} r_{\max}\}}{|S| \exp\{\eta^{-1} r_{\min}\}} < \infty.$$

This confirms that $\{LBR_\eta(P_n)\}_{n \in \mathbb{N}}$ is bounded. The absolute continuity of this measure is shown by Lahkar (2012). Hence the noise process fits the structure given in (B3). Applying Proposition 2.7 gives that the noise assumption (A3) holds. Applying Theorem 2.3 concludes the proof. □

Theorem 3.4. *If $r(x, y)$ is Lipschitz continuous in y for all $x \in S$, and the game is negative semi-definite, then the single population stochastic fictitious play-like process in* (3.6) *will converge to $\mathcal{LE}_\eta$.*

Proof. The result follows immediately by combining Theorem 3.2 and Theorem 3.3 with Theorem 2.6. □

To conclude this section we present a simple example of a classic evolutionary game which has been extended to a continuous action space for which CDF inversion can be used to generate samples from $l_\eta(P_n)$.

Example 3.5. We consider a linear extension to the standard hawk-dove game. For $C > V$ let

$$r(x, y) = \Big[1 - (y - x)\Big]\frac{V}{2} - xy\frac{C}{2},$$

and $S = [0, 1]$. An action in S corresponds to an aggression level, so $x = 1$ corresponds to playing 'hawk' in the classic game and $x = 0$ corresponds to playing 'dove'. The first term in the reward function represents the likelihood of obtaining a resource of value V and the second term is the 'injuries' sustained in contesting the resource. Both of these terms depend linearly on how aggressively each of the players contests the resource.

It is straightforward to show that this continuous hawk-dove game is negative semi-definite in the sense of Definition 3.1. A consequence of Theorem 3.4 is that the stochastic fictitious play-like process in (3.6) will converge to a logit equilibrium.

This game has infinitely many Nash equilibria but has a unique logit equilibrium.[2] Let $\tilde{P} \in \mathcal{P}(S, \mathcal{B})$ be a logit equilibrium, with associated density $\tilde{p}$. Because the reward function $r(x, y)$ for this continuous hawk-dove game is a linear in x it follows that for any $P \in \mathcal{P}(S, \mathcal{B})$, $l_\eta(P)(x) \propto e^{kx}$, for some $k \in \mathbb{R}$. Normalisation implies that

$$\tilde{p}(x) = \frac{ke^{kx}}{e^k - 1}. \tag{3.8}$$

Knowing that $\tilde{p}(x) = l_\eta(\tilde{P})(x)$ for all $x \in S$ and combining it with (3.8) and (3.4) allows us to calculate the exact value of k. Note that when $C = 2V$, a uniform distribution over S is the logit equilibrium. A simulation of this game with $V = 1$, $C = 4$ and $\eta = 0.005$ is shown in Figure 1. In the discrete action hawk-dove game with $V = 1$, $C = 4$ the Nash equilibrium is for 3/4 of the population to play 'dove'. With these same parameters the logit equilibrium here is also skewed towards the 'dove' action, capturing this particular feature of the traditional hawk-dove game. In Figure 1 the population starts as a uniform distribution and after 4000 iterations this population, which evolves according to the stochastic fictitious play-like process in (3.6), is close to the logit equilibrium for the game.

4. N-Player, Continuous Action Space Games

For finite $N \in \mathbb{N}$, N-player and N-population games can be defined in the same manner with differing interpretations. In an N-player (N-population), continuous action space game game we use $i = 1, \ldots, N$ to denote the players (populations) and in a standard abuse of notation if i denotes one player (population) then use $-i$ to denote all the others. Let $S_i \subset \mathbb{R}$ be a compact action set and let $\mathcal{B}_i$ denote the Borel σ-algebra on S_i. Following the notation of Section 3, let $\mathcal{P}(S_i, \mathcal{B}_i)$ be the set of all probability measures on S_i for $i = 1, \ldots, N$. In an N-population game $\mathcal{P}(S_i, \mathcal{B}_i)$ is thought of as the set of all possible distributions of population i; in a N-player game it should be thought of as the set of all possible strategies for Player i. In this section we follow the strategy interpretation, although we note that the logit best response dynamic remains the same

[2] It is straightforward to check that $P = 1/2\delta_{V/C+\varepsilon} + 1/2\delta_{V/C-\varepsilon}$ is a Nash equilibrium for any $\varepsilon \in [0, \min\{V/C, 1 - V/C\}]$, which shows that there are infinitely many Nash equilibria.

under the population interpretation. π^i is the strategy of Player i where $\pi^i(A_i)$ denotes the probability of player i selecting an action in the set $A_i \in \mathcal{B}_i$.

For $i = 1, \dots, N$, $\mathcal{P}(S_i, \mathcal{B}_i)$ is a subset of the space of signed measures $\mathcal{M}^e(S_i, \mathcal{B}_i)$. As in Section 3 we equip $\mathcal{M}^e(S_i, \mathcal{B}_i)$ with the weak topology using the bounded Lipschitz norm from (2.6). $\mathcal{BL}_i$ and $\| \cdot \|_{BL_i^*}$ are defined as $\mathcal{BL}$ and $\| \cdot \|_{BL^*}$ in (2.5) and (2.6) respectively, but with all integrals over S_i as opposed to S. As previously noted, when equipped with the BL_i^*-norm the space of finite signed measures is a Banach space and $\mathcal{P}(S_i, \mathcal{B}_i)$ is a compact subset of $\mathcal{M}^e(S_i, \mathcal{B}_i)$.

For $s_i \in S_i$ and $\underline{s} = (s_1, \dots, s_N)$, an N-player game has a reward structure $r^i(\underline{s}) \in \mathbb{R}$, $i = 1, \dots, N$. We will assume throughout that for $i = 1, \dots, N$, $r^i(\underline{s})$ is Lipschitz continuous in the joint action $\underline{s}$. This can be used to define the expected reward similarly to (3.1) and (3.2). If we have strategies $\pi^i \in \mathcal{P}(S_i, \mathcal{B}_i)$,

$$E^i(\pi^1, \dots, \pi^N) = E^i(\underline{\pi}) = \int_{S_1} \dots \int_{S_N} r^i(\underline{s}) \pi^1(\,\mathrm{d}s_1) \dots \pi^N(\,\mathrm{d}s_N),$$

and as before for $s_i \in S_i$, $E^i(s_i, \pi^{-i}) = E^i(\delta_{s_i}, \pi^{-i})$.

Now consider the Cartesian product $\Sigma := \mathcal{M}^e(S_1, \mathcal{B}_1) \times \dots \times \mathcal{M}^e(S_N, \mathcal{B}_N)$. If $\underline{\pi} := (\pi^1, \dots, \pi^N)$ represents an element of Σ we use the norm

$$\|\underline{\pi}\|_\Sigma := \max\{\|\pi^1\|_{BL_1^*}, \dots, \|\pi^N\|_{BL_N^*}\},$$

to metrize Σ. Hence if we have $\underline{\pi}, \underline{\rho} \in \Sigma$ the distance between these points in Σ is given by,

$$\|\underline{\pi} - \underline{\rho}\|_\Sigma = \max\{\|\pi^1 - \rho^1\|_{BL_1^*}, \dots, \|\pi^N - \rho^N\|_{BL_N^*}\}.$$

Under this norm Σ is a Banach space. This represents the product topology on Σ; hence a sequence $\underline{\pi}_n \to \underline{\pi}$ if and only if for $i = 1, \dots, N$, $\pi_n^i \to \pi^i$ on $\mathcal{M}^e(S_i, \mathcal{B}_i)$. It will be useful to let $\Delta := \mathcal{P}(S_1, \mathcal{B}_1) \times \dots \times \mathcal{P}(S_N, \mathcal{B}_N)$. Δ is a compact subset of Σ and the joint strategies for a N-player game will evolve on Δ.

4.1. N-Player, Logit Best Response Dynamics

Now we will consider the logit best response dynamic for N-player games. We note that the dynamics for N-population games is the same, although the situations in which these arise will be different. This section is a natural extension of the single population logit best response dynamic described in Section 3.1. We can now define the logit best response dynamic for joint strategies on Σ. For $A_i \in \mathcal{B}_i$ and fixed $\eta > 0$ the logit best response of player i to strategy π^{-i} is given by

$$LBR_\eta^i(\pi^{-i})(A_i) := \frac{\int_{A_i} \exp\{\eta^{-1} E^i(z, \pi^{-i})\} \,\mathrm{d}z}{\int_{S_i} \exp\{\eta^{-1} E^i(u, \pi^{-i})\} \,\mathrm{d}u}. \tag{4.1}$$

For convenience we will take $LBR_\eta(\underline{\pi}) := (LBR_\eta^1(\pi^{-1}), \dots, LBR_\eta^N(\pi^{-N}))$. It is well known that

$$LBR_\eta^i(\pi^{-i}) := \underset{\pi^i \in \mathcal{P}(S_i, \mathcal{B}_i)}{\arg\max} \left\{ E^i(\pi^i, \pi^{-i}) - \eta \nu^i(\pi^i) \right\}.$$

where $\nu^i(\pi^i)$ is the entropy of π^i. The full construction of this, including the definition of $\nu^i(\cdot)$, is given in Appendix D. Following identical logic to Lahkar (2012) the logit best response for player i as defined in (4.1) has an associated density function which is denoted

$$l^i_\eta(\pi^{-i})(x) := \frac{\exp\{\eta^{-1} E^i(x, \pi^{-i})\}}{\int_S \exp\{\eta^{-1} E^i(u, \pi^{-i})\}\,\mathrm{d}u}. \tag{4.2}$$

Let $\underline{\pi}(t) \in \Delta$ denote the joint strategy at time $t > 0$ and $\underline{\pi}(t)(A)$ denote the mass of the probability measure on the set $A = (A_1, \ldots, A_N)$ at time $t > 0$, where $A_i \in \mathcal{B}_i$. The N-player logit best response dynamic on Σ is given by

$$\dot{\underline{\pi}}(t)(A) = LBR_\eta\big(\underline{\pi}(t)\big)(A) - \underline{\pi}(t)(A). \tag{4.3}$$

Lahkar (2012) proves the existence and uniqueness of a solution to the single population logit best response dynamics in (3.5) under certain conditions. These proofs are extended to the N-player scenario in Appendix C.

Let $\mathcal{LE}^N_\eta$ be the set of logit equilibria for the N-player dynamical system (4.3). In Appendix C we show that $\mathcal{LE}^N_\eta$ will be non-empty for jointly Lipschitz continuous rewards. We proceed to describe the continuous action space extensions of two prominent discrete action games.

Definition 4.1. A continuous action space game with reward functions $r^i(\cdot) : S_1 \times \ldots \times S_N \to \mathbb{R}$ for $i = 1, \ldots, N$ is

(1) a *two-player zero-sum game* if $N = 2$ and

$$r^1(\underline{s}) = -r^2(\underline{s}),$$

for every $\underline{s} \in S_1 \times S_2$.

(2) an *N-player identical interest game* if for every $i = 1, \ldots, N$ and $\underline{s} \in S_1 \times \ldots \times S_N$,

$$r^i(\underline{s}) = r(\underline{s}),$$

for some $r(\cdot) : S_1 \times \ldots \times S_N \to \mathbb{R}$.[3]

It remains to present a global convergence result for the logit best response dynamics (4.3) in two-player zero-sum and N-player potential games with continuous action spaces and Lipschitz continuous reward function. These are the natural extensions of the discrete action case given by Hofbauer and Sandholm (2002) and Hofbauer and Hopkins (2005) to the continuous action space.

[3] As discussed by Hofbauer and Sandholm (2002), *potential games* often include payoffs which are common up to a shift. As in the discrete action case a payoff shift for player i which is of the form $r^i(\underline{s}) = r(\underline{s}) + u^i(s^{-i})$, for $u^i(\cdot) : S^{-i} \to \mathbb{R}$, will not affect the continuous action logit best response. Hence the results for identical payoffs here naturally extend to weighted potential games akin to those described by Monderer and Shapley (1996).

Theorem 4.2. *Assume that for $i = 1, 2$, $r^i(x, y)$ is Lipschitz continuous in the joint action (x, y). For $\underline{\pi} = (\pi^1, \pi^2) \in \Delta$ take*

$$V_\eta(\underline{\pi}) := \sum_{i=1}^{2} \Big[E^i\big(LBR^i_\eta(\pi^{-i}), \pi^{-i}\big) - \eta\nu^i\big(LBR^i_\eta(\pi^{-i})\big)\Big] - \sum_{i=1}^{2} \Big[E^i\big(\pi^i, \pi^{-i}\big) - \eta\nu^i\big(\pi^i\big)\Big].$$

Then $V_\eta(\cdot)$ is a Lyapunov function for (4.3) *for any two-player zero-sum, continuous action game and* (4.3) *has attracting set $\mathcal{LE}^2_\eta$.*

The proof combines elements of the proof by Lahkar (2012, Theorem 7.2) for negative definite, single population games and the work on stochastic fictitious play by Hofbauer and Hopkins (2005) and is contained in Appendix D.

Theorem 4.3. *Assume that for $i = 1, \ldots, N$, $r^i(x, y)$ is Lipschitz continuous in the joint action (x, y). For $\underline{\pi} = (\pi^1, \ldots, \pi^N) \in \Delta$ take*

$$V_\eta(\underline{\pi}) := E\big(\underline{\pi}\big) - \eta \sum_{i=1}^{N} \nu^i\big(\pi^i\big).$$

Then $V_\eta(\cdot)$ is a Lyapunov function for (4.3) *for any N-player identical interest game with continuous actions and* (4.3) *has attracting set $\mathcal{LE}^N_\eta$.*

The proof follows on from the proof of Theorem 4.2 and is also given in Appendix D.

4.2. Stochastic Fictitious Play

Now we will consider a stochastic fictitious play algorithm based on the logit best response. Using the stochastic approximation framework from Section 2 we analyse the limiting behaviour of this stochastic fictitious play process using the logit best response dynamic in (4.3).

The beliefs of each player at iteration n, denoted by σ^i_n, are the empirical frequencies with which the actions in S_i have been played. In an N-player game we assume that Player i selects an action $s^i_{n+1} \in S_i$ at iteration $n + 1$ using the logit best response $LBR^i_\eta(\sigma^{-i}_n)$. For $i = 1, \ldots, N$ and $\alpha_n = 1/(n + 1)$, the beliefs of the stochastic fictitious play process are given by

$$\sigma^i_{n+1} = \sigma^i_n + \alpha_{n+1} \left[\delta_{s^i_{n+1}} - \sigma^i_n\right].$$

Let $\underline{\sigma}_n := (\sigma^1_n, \ldots, \sigma^N_n) \in \Delta$. We can consider $\underline{\sigma}_n$ as an iterative process on a compact subset of the Banach space Σ. This will mean that the logit variant of stochastic fictitious play is defined as

$$\underline{\sigma}_{n+1} = \underline{\sigma}_n + \alpha_{n+1} \left[\left(\delta_{s^1_{n+1}}, \ldots, \delta_{s^N_{n+1}}\right) - \underline{\sigma}_n\right], \quad \text{where } s^i_{n+1} \sim l^i_\eta(\sigma^{-i}_n). \tag{4.4}$$

We note that it is costless to consider a more general learning rate process $\{\alpha_n\}_{n\in\mathbb{N}}$, and so we do so in the following.

Theorem 4.4. *Assume that for every $i = 1, \dots, N$, $r^i(\underline{s})$ is bounded and Lipschitz continuous in the joint action $\underline{s}$ and*

$$\sum_{n \in \mathbb{N}} \alpha_n = \infty, \quad \sum_{n \in \mathbb{N}} \alpha_n^2 < \infty.$$

Then a linear interpolation of the stochastic fictitious play process (4.4) *is an asymptotic pseudo-trajectory to the N-player, logit best response dynamics* (4.3).

Proof. The proof here is the natural extension to the proof of Theorem 3.3 and as such follows a similar pattern.

Firstly, we note that Σ is a Banach space with respect to $\|\cdot\|_\Sigma$ and that Δ is a compact subset of Σ. This will imply that the compactness assumption (A1) will always hold for a learning process which remains in Δ. For

$$V_{n+1}^i = \delta_{s_{n+1}^i} - LBR_\eta^i(\sigma_n^{-i}),$$

and $V_{n+1} = (V_{n+1}^1, \dots, V_{n+1}^N)$, (4.4) can be written as,

$$\underline{\sigma}_{n+1} = \underline{\sigma}_n + \alpha_{n+1}\left[\Big(LBR_\eta(\underline{\sigma}_n) - \underline{\sigma}_n\Big) + V_{n+1}\right], \tag{4.5}$$

which fits the general stochastic approximation form used in Theorem 2.3. Hence the claim will follow if we verify that (A1)-(A4) hold for (4.5).

We have already shown the compactness assumption (A1) to hold. The boundedness assumption (A2) will hold since the logit best response mean field in (4.5), $(LBR_\eta(\underline{\sigma}_n) - \underline{\sigma}_n)$, is uniformly continuous.

With $r^i(\underline{s})$ Lipschitz continuous in the joint action $\underline{s}$ for $i = 1, \dots, N$ the existence and uniqueness results in in Appendix C verify that the existence of solution assumption (A4) holds.

In order to verify that the noise assumption (A3) holds we will show that for $i = 1, \dots, N$ the assumptions of Proposition 2.7 hold for $\{V_n^i\}_{n \in \mathbb{N}}$. Let $\mathcal{F}_n$ represent the history of the iterative process (4.4) up to iteration $n \in \mathbb{N}$. The learning rate assumption (B1) is true since $\sum_n \alpha_n^2 < \infty$ by choice of $\{\alpha_n\}_{n \in \mathbb{N}}$. Clearly for all $n \in \mathbb{N}$, V_{n+1}^i is measurable with respect to $\mathcal{F}_n$ and hence will satisfy (B2). Using the same arguments as in the proof of Theorem 3.3 we ascertain that $\{LBR_\eta^i(\underline{\sigma}_n)\}_{n \in \mathbb{N}}$ is a bounded and absolutely continuous measure on Σ and hence the noise process $\{V_n^i\}_{n \in \mathbb{N}}$ fits the structure given in (B3). Applying Proposition 2.7 gives that for $i = 1, \dots, N$,

$$\lim_{n \to \infty} \sup \left\{ \left\| \sum_{j=n}^{k-1} \alpha_{j+1} V_{j+1}^i \right\|_{BL_i^*} ; k = n+1, \dots, m(\tau_n + T) \right\} = 0, \quad \text{w.p. } 1.$$

Since

$$\left\| \sum_{j=n}^{k-1} \alpha_{j+1} V_{j+1} \right\|_\Sigma = \max \left\{ \left\| \sum_{j=n}^{k-1} \alpha_{j+1} V_{j+1}^1 \right\|_{BL_1^*}, \dots, \left\| \sum_{j=n}^{k-1} \alpha_{j+1} V_{j+1}^N \right\|_{BL_N^*} \right\},$$

we conclude that (A3) holds. Applying Theorem 2.3 concludes the proof. □

The following theorem provides a global convergence result for the logit variant of stochastic fictitious play in (4.4) for two-player zero-sum and N-player identical interest continuous action space games with jointly Lipschitz continuous rewards. This extends the well known result for stochastic fictitious play in normal form zero-sum and potential games originally presented by Hofbauer and Sandholm (2002) and Hofbauer and Hopkins (2005).

Theorem 4.5. *In a continuous action space game assume that for every $i = 1, \dots, N$, $r^i(\underline{s})$ is Lipschitz continuous in the joint action $\underline{s}$. If*

(1) the game is a two-player zero-sum then the beliefs in the logit variant of stochastic fictitious play from (4.4) *will converge to $\mathcal{LE}^2_\eta$;*

(2) the game is an N-player identical interest game and $\mathcal{LE}^N_\eta$ is at most countably infinite then the beliefs in the logit variant of stochastic fictitious play from (4.4) *will converge to $\mathcal{LE}^N_\eta$.*

Proof. In zero-sum games, the Lyapunov function takes value 0 at all elements of $\mathcal{LE}^2_\eta$. The additional assumption here on the countability of $\mathcal{LE}^N_\eta$ for identical interest games ensures that $V(\mathcal{LE}^N_\eta)$ has empty interior. Therefore, the results follow immediately by combining Theorem 4.4 and Theorem 4.2/Theorem 4.3 with Theorem 2.6. □

As in Section 3, we conclude this section with an example which extends two-player matching pennies to the continuous action case.

Example 4.6. We consider a linear extension to the standard matching pennies games. Let $r^1(x, y) = (x - 1/2)(y - 1/2)$ and $r^2(x, y) = -r^1(x, y)$, where $S_i = [0, 1]$ for $i = 1, 2$. There is a unique logit equilibrium with $\tilde{\sigma}^i \sim Unif(0, 1)$ for $i = 1, 2$. Theorem 4.5 gives that the beliefs in the stochastic fictitious play process (4.4) will converge to this equilibrium. This convergence is demonstrated in Figure 2.

5. Discussion

In this work we present a method for studying the limiting behaviour of iterative learning processes with an uncountably infinite action space. This extends the work of Fudenberg and Kreps (1993), Benaïm and Hirsch (1999), Hofbauer and Sandholm (2002) and Hofbauer and Hopkins (2005) to games with actions selected from a continuous set.

To achieve this we have developed new tools for stochastic approximation. These build on the asymptotic pseudo-trajectory approach to stochastic approximation of Benaïm (1999) and extend the abstract stochastic approximation approach presented by Shwartz and Berman (1989) to this now more common framework. As a suitable space for continuous strategies we study the space of finite signed measures $\mathcal{M}^e(S, \mathcal{B})$ under the weak topology induced by the bounded Lipschitz norm. Unlike Shwartz and Berman (1989) we provide simple conditions in the spirit of Benaïm (1999) as to when the difficult martingale noise assumption will hold on $\mathcal{M}^e(S, \mathcal{B})$.

In Section 3 we present the motivation and key application of this framework to single population games. A mixture of theoretical results and simulations are provided for the stochastic fictitious play-like algorithm to demonstrate convergence in continuous action space games, analogous to the convergence already known to occur in finite action spaces (Fudenberg and Kreps, 1993; Benaïm and Hirsch, 1999; Hofbauer and Sandholm, 2002; Hofbauer and Hopkins, 2005).

Our framework can also be used to study learning variants of the replicator dynamics, such as in Narendra and Thathachar (1989), Börgers and Sarin (1997) and Leslie (2003, Chapter 2). Although the replicator dynamics are studied more frequently, especially in the continuous action space literature (see Oechssler and Riedel (2001, 2002); Seymour (2002); Cressman (2005)), the framework has limitations. In particular, the replicator dynamics does not generate new strategies, so that the limit point of a trajectory will always depend on the initial conditions. Even when exploration of the state space can be guaranteed convergence of an associated learning process is generally very slow. In contrast, the logit best response is absolutely continuous and assigns some probability to every part of the action space which makes it more straightforward to study associated learning processes. This is true of the discrete action case, where fictitious play and stochastic fictitious play are more frequently studied than similar discrete time variants of the replicator dynamics, and remains true for the continuous action case.

In addition, in Section 4 we extend the existence and uniqueness results of Lahkar (2012) to the N-player case. As a consequence we can study a logit variant of stochastic fictitious play for continuous action space games. We prove the convergence of stochastic fictitious play for two-player zero-sum and N-player potential games with continuous action spaces and Lipschitz continuous reward function. This extends the previous results of Hofbauer and Sandholm (2002) and Hofbauer and Hopkins (2005) for stochastic fictitious play in normal form games with finite action sets.

Appendices

A. Noise Criteria: L^2

It is important for us to be able to verify that the noise condition $(A3)$ on $\mathcal{M}^e(S, \mathcal{B})$ holds so that stochastic approximation can be performed on this Banach space. Doing so is not straightforward and requires us to use the intermediate result, proving that $(A3)$ holds on L^2, presented in this appendix. When stochastic approximation is performed in a Euclidean space the noise process is often assumed to be a martingale difference sequence. The proof for Euclidean space (Benaïm, 1999, Proposition 4.2) relies on the Burkholder-Davis-Gundy inequality to study the martingale difference sequence and this inequality can be extended for certain Banach spaces, notably L^2. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\{\mathcal{F}_n\}_{n\in\mathbb{N}}$ a filtration on $\mathcal{F}$.

Proposition A.1. *Consider the stochastic approximation process given in* (2.2) *on the Banach space of L^2 functions with associated norm $\|\cdot\|_{L^2}$. If for some $q \geq 2$,*

(1) $\{\alpha_n\}_{n\in\mathbb{N}}$ is deterministic with $\sum_{n\in\mathbb{N}} \alpha_n^{1+q/2} < \infty$,

(2) $\{U_n\}_{n\in\mathbb{N}}$ is adapted and measurable with respect to $\mathcal{F}_n$ for all n such that

$$\mathbb{E}\left[U_{n+1}|\mathcal{F}_n\right] = 0, \quad \sup_n \mathbb{E}\left[\|U_n\|_{L^2}^q\right] < \infty,$$

then

$$\lim_{n\to\infty} \sup\left\{\left\|\sum_{i=n}^{k-1} \alpha_{i+1}U_{i+1}\right\|_{L^2}; k = n+1, \ldots, m(\tau_n + T)\right\} = 0, \quad \textit{w.p. 1.}$$

Because the Burkholder-Davis-Gundy inequality can be extended for L^2-valued martingales the proof of this result follows a similar pattern to of Benaïm (1999, Proposition 4.2) for martingales in Euclidean space.

Proof. Brzeźniak (1997) shows that the Burkholder-Davis-Gundy inequality can be extended for a class of Banach spaces which includes L^2. This means that for any $q \in (1, \infty)$ there exists a $C_q > 0$ such that for an L^2-valued martingale $\{Y_n\}_{n\in\mathbb{N}}$ and stopping time $N > 0$ then

$$\mathbb{E}\left[\sup_{n\le N} \|Y_n\|_{L^2}^q\right] \le C_q\mathbb{E}\left[\left(\sum_{n=1}^{N} \|Y_n - Y_{n-1}\|_{L^2}^2\right)^{q/2}\right]. \tag{A.1}$$

The remainder of the proof is an extension to the proof of Benaïm (1999, Proposition 4.2) for L^2 using (A.1) rather than the original Burkholder-Davis-Gundy inequality. Let $W_{n,m} := \sum_{i=n}^{m} \alpha_{i+1}U_{i+1}$ for $m \ge n$. Since $U_{i+1} \in L^2$ then $W_{n,m} \in L^2$. Now fixing $n \in \mathbb{N}$, we have that

$$\mathbb{E}\left[W_{n,m}|\mathcal{F}_m\right] = W_{n,m-1},$$

and hence $W_{n,m}$ is a martingale in L^2. Using (A.1) there exists some constant $C_q > 0$ such that

$$\mathbb{E}\left[\sup_{n\le k\le m(\tau_n+T)} \|W_{n,k}\|_{L^2}^q\right] \le C_q\mathbb{E}\left[\left(\sum_{i=n+1}^{m(\tau_n+T)} \|W_{n,i} - W_{n,i-1}\|_{L^2}^2\right)^{q/2}\right].$$

Now by noticing that $\|W_{n,i} - W_{n,i-1}\|_{L^2} = \alpha_{i+1}\|U_{i+1}\|_{L^2}$ and using the original definition of $W_{n,m}$ gives

$$\mathbb{E}\left[\sup_{n\le k\le m(\tau_n+T)} \left\|\sum_{i=n}^{k} \alpha_{i+1}U_{i+1}\right\|_{L^2}^q\right] \le C_q\mathbb{E}\left[\left(\sum_{i=n}^{m(\tau_n+T)-1} \alpha_{i+1}^2\|U_{i+1}\|_{L^2}^2\right)^{q/2}\right].$$

The remainder of the proof continues exactly as in Benaïm (1999, Proposition 4.2) with the only difference being the norm used here is $\|\cdot\|_{L^2}$ and in Benaïm (1999, Proposition 4.2) this is the standard Euclidean norm. □

Remark A.2. Brzeźniak (1997) shows that the Burkholder-Davis-Gundy inequality can be extended for a class of Banach spaces which includes L^p-spaces for any $p \in [2, \infty)$. The result in Proposition A.1 for L^2 can be extended to any Banach space in the class for which Brzeźniak's result holds with no alteration to the proof.

B. Proof of Proposition 2.7

If we could show that $\mathcal{M}^e(S, \mathcal{B})$ is in the class of Banach spaces for which Brzeźniak (1997) proves the Burkholder-Davis-Gundy inequality can be extended then the proof of Proposition 2.7 would be identical to the proof of Proposition A.1. However, we have been unable to show this and so we are forced to take a different approach. Here we approximate all Dirac measures with a spike of fixed width. This will allow us to consider functions with proper density functions in L^2. We are then able to use Proposition A.1 to show the convergence and show the additional error term from this approach is not significant.

Proof of Proposition 2.7. Fix $\gamma > 0$. We will approximate atoms in S by measures that have a spike density with base width 2γ. Hence we need to consider an expanded space $\bar{S}$ to accommodate spikes near the boundary of S.[4]. Suppose $S \subseteq [a, b] \subset \mathbb{R}$, and define $\bar{S} := [a - \gamma, b + \gamma]$. Let $\| \cdot \|_{BL}$, $\bar{\mathcal{BL}}$ and $\| \cdot \|_{\bar{BL}^*}$ be as defined in (2.4)-(2.6) but with integrals over $\bar{S}$ and $\bar{\mathcal{B}}$ the Borel σ-algebra over $\bar{S}$. $\mathcal{M}^e(\bar{S}, \bar{\mathcal{B}})$ is the space of finite signed measures on $\bar{S}$ and is equipped with the weak topology using $\| \cdot \|_{\bar{BL}^*}$.

Consider arbitrary $\tilde{z} \in S \subset \bar{S}$, let $\bar{h}$ be a spike density on $\bar{S}$ centered on $\tilde{z}$ defined as

$$\bar{h}(z, \tilde{z}) := \begin{cases} \frac{1}{\gamma}\big(z - (\tilde{z} - \gamma)\big), & z \in [\tilde{z} - \gamma, \tilde{z}] \\ 1 - \frac{1}{\gamma}\big(z - \tilde{z}\big), & z \in (\tilde{z}, \tilde{z} + \gamma] \\ 0, & \text{otherwise} \end{cases}, \tag{B.1}$$

and let $\bar{H}(\tilde{z})$ be the measure in $\mathcal{M}^e(\bar{S}, \bar{\mathcal{B}})$ associated with density $\bar{h}(\cdot, \tilde{z})$. We will firstly show that

$$\|\delta_{\tilde{z}} - \bar{H}(\tilde{z})\|_{\bar{BL}^*} \leq \gamma$$

First note that if $\bar{g} \in \bar{\mathcal{BL}}$ then $\bar{g}$ has a Lipschitz constant that is not more than 1, so $|z - y| \leq \gamma \Rightarrow |\bar{g}(z) - \bar{g}(y)| \leq \gamma$. Conversely if $|z - y| \geq \gamma$ then $\bar{h}(z, y) = 0$. Hence

[4] This is a particular issue which Chen and White (1998) did not address when using probability densities on L^2.

$$\begin{aligned}\left|\int_{\bar{s}} \bar{g}(z)\Big(\delta_{\tilde{z}} - H(\tilde{z})\Big)\,\mathrm{d}z\right| &= \left|\bar{g}(\tilde{z}) - \int_{\bar{s}} \bar{g}(z)\bar{h}(z,\tilde{z})\,\mathrm{d}z\right|, \\ &= \left|\int_{\bar{s}} \Big[\bar{g}(\tilde{z}) - \bar{g}(z)\Big]\bar{h}(z,\tilde{z})\,\mathrm{d}z\right|, \\ &\leq \int_{\tilde{z}-\gamma}^{\tilde{z}+\gamma} \left|\bar{g}(\tilde{z}) - \bar{g}(z)\right| \bar{h}(z,\tilde{z})\,\mathrm{d}z, \\ &= \int_{\tilde{z}-\gamma}^{\tilde{z}+\gamma} \gamma\bar{h}(z,\tilde{z})\,\mathrm{d}z, \\ &= \gamma. \end{aligned} \tag{B.2}$$

Hence

$$\|\delta_{\tilde{z}} - \bar{H}(\tilde{z})\|_{\bar{BL}^*} = \sup_{\bar{g}\in\bar{\mathcal{BL}}} \left|\int_{\bar{S}} \bar{g}(z)\Big(\delta_{\tilde{z}} - \bar{H}(\tilde{z})\Big)(\,\mathrm{d}z)\right| \leq \gamma.$$

To use this within our stochastic approximation process we define $\bar{H}_n := \bar{H}(s_n)$ so that for all $n \in \mathbb{N}$,

$$\|\delta_{s_n} - \bar{H}_n\|_{\bar{BL}^*} \leq \gamma. \tag{B.3}$$

To examine the convergence of $\{U_n\}_{n\in\mathbb{N}}$ we also need to consider the convolution

$$\bar{q}_n(z) := \int_S \bar{h}(z,y)p_n(y)\,\mathrm{d}y, \tag{B.4}$$

where $\bar{h}$ is the spike density defined above and p_n is the density of measure P_n and $z \in \bar{S}$. Let $\bar{Q}_n$ be the measure on $\mathcal{M}^e(\bar{S},\bar{\mathcal{B}})$ associated with $\bar{q}_n$. This is useful since $\bar{h}_{n+1}$ can be viewed as an L^2-valued random variable and importantly we have that

$$\mathbb{E}\Big[\bar{h}_{n+1}|\mathcal{F}_n\Big] = \bar{q}_n. \tag{B.5}$$

Since both $\bar{h}_{n+1}$ and $\bar{q}_n$ are in $L^2(\bar{S})$, (B.5) will mean that $\bar{h}_{n+1} - \bar{q}_n$ is an L^2-valued martingale.

Finally, we need to define $\bar{U}_{n+1} = \delta_{s_{n+1}} - \bar{P}_n$, which is the extension of U_{n+1} to $\bar{S}$, where

$$\bar{P}_n := \begin{cases} P_n, \text{on } S \\ 0, \quad \text{on } \bar{S}\backslash S \end{cases},$$

and $\bar{P}_n$ has a density function $\bar{p}_n$. For fixed $T > 0$ it is clear for $k = n+1,\ldots,m(\tau_n + T)$ that

$$\begin{aligned}\left\|\sum_{i=n}^{k-1}\alpha_{i+1}U_{i+1}\right\|_{BL^*} &= \left\|\sum_{i=n}^{k-1}\alpha_{i+1}\bar{U}_{i+1}\right\|_{\bar{BL}^*}\\ &\leq \left\|\sum_{i=n}^{k-1}\alpha_{i+1}\left(\delta_{s_{i+1}}-\bar{H}_{i+1}\right)\right\|_{\bar{BL}^*}+\left\|\sum_{i=n}^{k-1}\alpha_{i+1}\left(\bar{H}_{i+1}-\bar{Q}_{i}\right)\right\|_{\bar{BL}^*}\\ &\quad+\left\|\sum_{i=n}^{k-1}\alpha_{i+1}\left(\bar{Q}_{i}-\bar{P}_{i}\right)\right\|_{\bar{BL}^*}. \qquad \text{(B.6)}\end{aligned}$$

We address each of these terms in turn. Using (B.3) we see that

$$\left\|\sum_{i=n}^{k-1}\alpha_{i+1}\left(\delta_{s_{i+1}}-\bar{H}_{i+1}\right)\right\|_{\bar{BL}^*} \leq \sum_{i=n}^{k-1}\alpha_{i+1}\gamma \approx T\gamma. \qquad \text{(B.7)}$$

The definition of $\|\cdot\|_{\bar{BL}^*}$ implies that $\|\cdot\|_{\bar{BL}^*} \leq \|\cdot\|_{L^1}$, and a standard result for the L^1-norm, which follows from Hölder's inequality, is that if $\bar{S}$ is a compact subset of $\mathbb{R}$ then $\|\cdot\|_{L^1} \leq |\bar{S}|^{1/2}\,\|\cdot\|_{L^2}$. Hence

$$\left\|\sum_{i=n}^{k-1}\alpha_{i+1}\left(\bar{H}_{i+1}-\bar{Q}_{i}\right)\right\|_{\bar{BL}^*} \leq |\bar{S}|^{1/2}\left\|\sum_{i=n}^{k-1}\alpha_{i+1}\left(\bar{h}_{i+1}-\bar{q}_{i}\right)\right\|_{L^2}. \qquad \text{(B.8)}$$

We have already observed that $\bar{h}_{n+1}-\bar{q}_n$ is an L^2-valued martingale sequence, and under the assumptions of Proposition 2.7, since p_n is bounded $\bar{q}_n$ is also bounded, and there exists $C > 0$ such that

$$\sup_n \mathbb{E}\Big[\left\|\bar{h}_{n+1}-\bar{q}_n\right\|_{L^2}^2\Big] \leq \sup_n\Big\{\mathbb{E}\Big[\left\|\bar{h}_{n+1}\right\|_{L^2}^2\Big]\Big\}+\sup_n\Big\{\mathbb{E}\Big[\left\|\bar{q}_n\right\|_{L^2}^2\Big]\Big\} < \frac{1}{\gamma^2}|\bar{S}|+C^2|\bar{S}| < \infty$$

Using Proposition A.1 immediately gives

$$\lim_{n\to\infty}\sup\left\{\left\|\sum_{i=n}^{k-1}\alpha_{i+1}\left(\bar{h}_{i+1}-\bar{q}_{i}\right)\right\|_{L^2}; k=n+1,\ldots,m(\tau_n+T)\right\}=0. \qquad \text{(B.9)}$$

Finally, following a similar approach to that used in obtaining (B.2), the definition of $\bar{q}_n(\cdot)$ in (B.4) tells us that for any $\bar{g} \in \bar{BL}$

$$\begin{aligned}
\left|\int_{\bar S} \bar g(z)\Big[\bar q_n(z) - \bar p_n(z)\Big]\,\mathrm{d}z\right| &= \left|\int_{\bar S} \bar g(z)\Big[\int_{\bar S} \bar h(z,y)\bar p_n(y)\,\mathrm{d}y - \bar p_n(z)\Big]\,\mathrm{d}z\right|, \\
&= \left|\int_{\bar S}\Big[\int_{\bar S} \bar g(z)\bar h(z,y)\,\mathrm{d}z - \bar g(y)\Big]\bar p_n(y)\,\mathrm{d}y\right|, \\
&\le \int_{\bar S}\left|\int_{\bar S} \bar g(z)\bar h(z,y)\,\mathrm{d}z - \bar g(y)\right|\bar p_n(y)\,\mathrm{d}y, \\
&= \int_{\bar S}\left|\int_{\bar S} \Big[\bar g(z) - \bar g(y)\Big]\bar h(z,y)\,\mathrm{d}z\right|\bar p_n(y)\,\mathrm{d}y, \\
&\le \int_{\bar S} \gamma \bar p_n(y)\,\mathrm{d}y, \\
&= \gamma.
\end{aligned}$$

It then follows that

$$\|\bar Q_n - \bar P_n\|_{\bar{BL}^*} \le \gamma, \tag{B.10}$$

and from this

$$\left\|\sum_{i=n}^{k-1} \alpha_{i+1}\left(\bar Q_i - \bar P_i\right)\right\|_{\bar{BL}^*} \le \sum_{i=n}^{k-1} \alpha_{i+1}\left\|\bar Q_i - \bar P_i\right\|_{\bar{BL}^*} \le T\gamma. \tag{B.11}$$

Taking the appropriate limit and supremum of (B.6) and substituting (B.7)-(B.9) and (B.11) gives,

$$\begin{aligned}
&\lim_{n\to\infty} \sup\left\{\left\|\sum_{i=n}^{k-1}\alpha_{i+1}U_{i+1}\right\|_{BL^*} ; k = n+1,\ldots,m(\tau_n+T)\right\} \\
&\quad \le 2T\gamma + |\bar S|^{1/2} \lim_{n\to\infty} \sup\left\{\left\|\sum_{i=n}^{k-1}\alpha_{i+1}\left(\bar h_{i+1} - \bar q_i\right)\right\|_{L^2} ; k = n+1,\ldots,m(\tau_n+T)\right\}, \\
&\quad = 2T\gamma.
\end{aligned}$$

Noting that the initial choice of $\gamma > 0$ was arbitrary completes the proof. □

C. Logit BR Dynamics for Multiple Populations

In this section the existence and uniqueness results of Lahkar (2012) are extended for two-population games. It is clear from the construction here that these techniques naturally extend to N-population games for finite $N \in \mathbb{N}$, but the additional complication to the notation is not required here. By noting that the logit best response dynamic for N-population games and N-player games are identical (simply different interpretations) we obtain the existence and uniqueness of a solution to the logit best response dynamics (4.3) from the results in this section.

The framework for dual population games is identical to the notation used in Section 4 for two-player games with the change in interpretation of $\mathcal{P}(S_i, \mathcal{B}_i)$ to be the set of all populations. To emphasize the difference between Section 4 we let $P \in \mathcal{P}(S_1, \mathcal{B}_1)$, $Q \in \mathcal{P}(S_2, \mathcal{B}_2)$ represent the two populations throughout. The distance between two elements $(P_1, Q_1), (P_2, Q_2) \in \Sigma$ is given by

$$\|(P_1, Q_1) - (P_2, Q_2)\|_\Sigma = \max\{\|P_1 - P_2\|_{BL_1^*}, \|Q_1 - Q_2\|_{BL_2^*}\}.$$

The dual population logit best response dynamic in this case is similar to (4.3) and is given by

$$\begin{aligned} \dot{P}(A_1) &= LBR_\eta^1(Q)(A_1) - P(A_1), \\ \dot{Q}(A_2) &= LBR_\eta^2(P)(A_2) - Q(A_2). \end{aligned} \tag{C.1}$$

With the framework for two population, continuous action space games established we focus on proving existence and uniqueness of solutions to the dynamical system (C.1). For $i = 1, 2$ let

$$\mathcal{M}_i^2 := \{P \in \mathcal{M}^e(S_i, \mathcal{B}_i) : \|P\|_{BL_i^*} \leq 2\}.$$

The following definition is taken form Lahkar (2012) but is extended for two populations.

Definition C.1. The expected payoff $E^i(\cdot, \cdot) : \Sigma \to \mathbb{R}$ is *Lipschitz continuous on $\mathcal{M}_j^2$, $j \neq i$, uniformly in $z \in S_i$, with respect to the weak topology* if there exists a constant $K > 0$ such that

$$|E^i(z, P) - E^i(z, Q)| \leq K\|P - Q\|_{BL_j^*},$$

for all $P, Q \in \mathcal{M}_j^2$, $j \neq i$, $z \in S_i$.

When Definition C.1 is satisfied Lahkar (2012) shows the existence and uniqueness of a solution to the single population logit best response dynamics given in (3.5). We present a new result which gives a straightforward criteria on the reward function for Definition C.1 to be satisfied. This holds for both the single and dual population cases.

Lemma C.2. *Assume that for $x \in S_1$ ($y \in S_2$) $r^1(x, z)$ ($r^2(z, y)$) is Lipschitz continuous in z. Then for all $x \in S_1$ ($y \in S_2$) Definition C.1 will hold for $E^1(x, \cdot)$ ($E^2(\cdot, y)$).*

Proof. Fix $x \in S_1$. Firstly, from the definition we have

$$\left|E^1(x, Q_1) - E^1(x, Q_2)\right| = \left|\int_{S_2} r^i(x, z)(Q_1 - Q_2)(\,\mathrm{d}z)\right|. \tag{C.2}$$

Now $r^1(x, z)$ is Lipschitz continuous in z with Lipschitz constant $C(x)$. In addition let $m(x) := \max_{z \in S_2}\{|r^1(x, z)|\} < \infty$ which exists since S_2 is compact and $r^1(x, z)$ is Lipschitz continuous in z. Now define

$$C := \max_{x \in S_1} C(x) < \infty, \quad \text{and} \quad M := \max_{x \in S_1} m(x) < \infty.$$

Now for all $x \in S_1$, $y \in S_2$ let

$$\tilde{r}^1(x,y) := \frac{r^1(x,y)}{M+C}.$$

It is straightforward to show that for all $x \in S_1$, $\tilde{r}^1(x,z)$ is Lipschitz continuous in z with a Lipschitz constant $C/(M+C)$ and maximum value $M/(M+C)$. This gives that for all $x \in S_1$, $\tilde{r}^1(x,z) \in \mathcal{BL}_2$. Now we extend (C.2) using $\tilde{r}^1(x,y)$,

$$\begin{aligned}
\left|E^1(x,Q_1) - E^1(x,Q_2)\right| &= (C+M)\left|\int_{S_2} \tilde{r}^i(x,z)(Q_1 - Q_2)(\,\mathrm{d}z)\right|, \\
&\le (C+M)\sup_{g\in\mathcal{BL}_2}\left|\int_{S_2} g(z)(Q_1 - Q_2)(\,\mathrm{d}z)\right|, \\
&\le (C+M)\|Q_1 - Q_2\|_{BL_2^*}.
\end{aligned}$$

Finally, we note that no restrictions were placed on Q_1, Q_2 and hence this will hold for $Q_1, Q_2 \in \mathcal{M}_2^2$. This completes the proof for $x \in S_1$ and Lipschitz continuous $r^1(x,z)$. The proof for $y \in S_2$ and Lipschitz continuous $r^2(z,y)$ follows an identical structure. □

Before we proceed with the dual population analysis we state a corollary of Lemma C.2 for the single population case which is used throughout Section 3.

Corollary C.3. *In the single population game described in Section 3, if for all $x \in S$ $r(x,y)$ is Lipschitz continuous in y then from each initial condition $P(0) \in \mathcal{P}(S,\mathcal{B})$ there exists a unique solution to the logit best response dynamic* (3.5) *for all time. Furthermore, the semiflow induced by this dynamical system is continuous with respect to the weak topology on $\mathcal{M}^e(S,\mathcal{B})$ and $\mathcal{LE}$ is non-empty.*

Proof. By Lemma C.2 the reward structure will satisfy Definition C.1 and the claim follows immediately from Lahkar (2012, Theorem 4.1, Theorem 5.2). □

Now we proceed to show the existence of a logit equilibrium on Σ and the uniqueness and continuity of a solution for the dual population dynamical system (C.1). Let $\mathcal{LE}^2_\eta \subset \Delta$ be the set of logit equilibria for the dual population logit best response dynamics in (C.1). The following proofs follow directly from the corresponding single population results of Lahkar (2012).

Theorem C.4. *If for all $x \in S_1$, $r^1(x,y)$ is Lipschitz continuous in y and for all $y \in S_2$, $r^2(x,y)$ is Lipschitz continuous in x, then $\mathcal{LE}^2_\eta$ is non-empty.*

Proof. The proof is omitted since it is a straightforward extension to the proof of Lahkar (2012, Theorem 4.1). □

Theorem C.5. *Assume that for all $x \in S_1$, $r^1(x,y)$ is Lipschitz continuous in y and that for all $y \in S_2$, $r^2(x,y)$ is Lipschitz continuous in x. Then for each initial population in $(P(0),Q(0)) \in \Delta$ there exists a unique solution $(P(t),Q(t))$ on Σ to the dual population logit best response differential equation* (C.1) *for all time $t \in [0,\infty)$. In addition the semiflow $\Phi_t\big((P(0),Q(0))\big) := (P(t),Q(t))$ is continuous with respect to the topology on Σ.*

Proof. The proof is omitted since it is a straightforward extension to the proof of Lahkar (2012, Theorem 5.2). □

D. Convergence of the Logit Best Response Dynamic

In this section we prove the global convergence results of Theorem 4.2 and Theorem 4.3 for the N-player logit best response dynamics in (4.3). These proofs are extensions to similar work of Lahkar (2012, Appendix A.2). We begin by introducing some useful notation and proving two lemmas which will be used to prove Theorem 4.2 and Theorem 4.3.

A general smooth best response is defined as the probability measure maximising a perturbed payoff function

$$\beta^i(\pi^{-i}) := \underset{\pi^i \in \mathcal{P}(S_i,\mathcal{B}_i)}{\arg\max} \left\{ E^i(\pi^i, \pi^{-i}) - \eta \nu^i(\pi^i) \right\}.$$

The logit best response is a particular form of the smooth best response which uses the entropy as a perturbation term. Firstly, we note that if π^i is not absolutely continuous then we can still define a sequence of absolutely continuous probability measures π^i_k with associated densities p^i_k such that $\{\pi^i_k\}_{k\in\mathbb{N}}$ convergences in distribution to π^i. It is convenient here to use the density notation p^i, p^i_k in place of the measure notation π^i, π^i_k. From this the logit best response can be defined using the perturbation term,

$$\nu^i(\pi^i) := \begin{cases} \int_{S_i} p^i(z) \log\left(p^i(z)\right) \mathrm{d}z & \pi^i \text{ absolutely continuous,} \\ \lim_{k\to\infty} \int_{S_i} p^i_k(z) \log\left(p^i_k(z)\right) \mathrm{d}z & \pi^i \text{ not absolutely continuous.} \end{cases} \tag{D.1}$$

Now following the arguments of Lahkar (2012), if π^i is not absolutely continuous then $\nu^i(\pi^i) = \infty$. As a consequence, when using the logit best response we can restrict ourselves to the case where π^i is absolutely continuous.

Define the tangent space for player i as

$$T_i\mathcal{P}(S_i,\mathcal{B}_i) := \left\{ \mu \in \mathcal{M}^e(S_i,\mathcal{B}_i); \int_{S_i} \mu(\,\mathrm{d}z) = 0 \right\}.$$

Recall from (4.2) that the logit best response density is given by

$$l^i_\eta(\pi^{-i})(z) := \frac{\exp\{\eta^{-1} E^i(z,\pi^{-i})\}}{\int_S \exp\{\eta^{-1} E^i(u,\pi^{-i})\} \,\mathrm{d}u}. \tag{D.2}$$

Let $\underline{\pi}$ be absolutely continuous and let the time derivative of $l^i_\eta(\pi^{-i})$ be $\dot{l}^i_\eta(\pi^{-i})$. Clearly $\dot{l}^i_\eta(\pi^{-i})$ is in the tangent space. The existence of $\dot{l}^i_\eta(\pi^{-i})$ follows from the absolute continuity of $\underline{\pi}$ (Lahkar, 2012).

Similarly we can take the derivative of the $\nu^i(\pi^i)$ terms with respect to π^i. Firstly, let $\nabla\nu^i(\pi^i) : \mathcal{P}(S_i,\mathcal{B}_i) \to T_i\mathcal{P}(S_i,\mathcal{B}_i)$ be the derivative of $\nu^i(\pi^i)$ which maps from $\mathcal{M}^e(S_i,\mathcal{B}_i)$ to the tangent space. Extend the definition from (D.1) by letting $\nu^i_z(\pi^i) := p^i(z)\log(p^i(z))$. We note from Lahkar (2012) that

$$\nabla\nu^i_z(\pi^i) = \log\left(p^i(z)\right) - \frac{\int_{S_i} \log\left(p^i(u)\right) \mathrm{d}u}{|S_i|}. \tag{D.3}$$

Lemma D.1. *Let $\underline{\pi}$ be absolutely continuous. Then for $i = 1, \dots, N$*

$$E^i\big(\dot{l}^i_\eta(\pi^{-i}), \pi^{-i}\big) = \eta \int_{S_i} \nabla\nu^i_z(LBR^i_\eta(\pi^{-i}))\dot{l}^i_\eta(\pi^{-i})(z)\,\mathrm{d}z.$$

Proof. The proof is omitted since it is a straightforward extension to the first part of the proof of Lahkar (2012, Lemma A.1). □

Lemma D.2. *Let $\underline{\pi}$ be absolutely continuous. Then for $i = 1, \dots, N$*

$$\int_{S_i} \Big[E^i(z, \pi^{-i}) - \eta\nabla\nu^i_z(\pi^i)\Big]\dot{\pi}^i(\,\mathrm{d}z) \geq 0$$

with equality only when $\underline{\pi} = LBR_\eta(\underline{\pi})$.

Proof. The proof is omitted since it is a straightforward extension to the proof of Lahkar (2012, Lemma A.1). □

Proof of Theorem 4.2. Firstly it is straightforward to show that $V_\eta(\underline{\pi}) \geq 0$ with equality when $\underline{\pi} = LBR_\eta(\underline{\pi})$. Now, let

$$\begin{aligned}
w^1_\eta(\underline{\pi}) &= E^1\big(LBR^1_\eta(\pi^2), \pi^2\big) - \eta\nu^1\big(LBR^1_\eta(\pi^2)\big) - E^1(\pi^1, \pi^2) + \eta\nu^1\big(\pi^1\big),\\
w^2_\eta(\underline{\pi}) &= E^2\big(\pi^1, LBR^2_\eta(\pi^1)\big) - \eta\nu^2\big(LBR^2_\eta(\pi^1)\big) - E^2(\pi^1, \pi^2) + \eta\nu^2\big(\pi^2\big).
\end{aligned}$$

This will give that

$$V_\eta(\underline{\pi}) = w^1_\eta(\underline{\pi}) + w^2_\eta(\underline{\pi}),$$

and hence

$$\dot{V}_\eta(\underline{\pi}) = \dot{w}^1_\eta(\underline{\pi}) + \dot{w}^2_\eta(\underline{\pi}).$$

Taking the derivative of these terms we begin with

$$\frac{d}{dt}E^i(\pi^1, \pi^2) = E^i(\dot{\pi}^1, \pi^2) + E^i(\pi^1, \dot{\pi}^2), \tag{D.4}$$

Recalling that $\dot{l}^i_\eta(\pi^{-i}) \in T_i\mathcal{P}(S_i, \mathcal{B}_i)$ is the time derivative of $l^i_\eta(\pi^{-i})$,

$$\frac{d}{dt}E^1\big(LBR^1_\eta(\pi^2), \pi^2\big) = E^1\big(LBR^1_\eta(\pi^2), \dot{\pi}^2\big) + E^1\big(\dot{l}^1_\eta(\pi^2), \pi^2\big), \tag{D.5}$$

and similarly

$$\frac{d}{dt}E^2\big(\pi^1, LBR^2_\eta(\pi^1)\big) = E^2\big(\dot{\pi}^1, LBR^2_\eta(\pi^1)\big) + E^2\big(\pi^1, \dot{l}^2_\eta(\pi^1)\big). \tag{D.6}$$

Taking the derivative of $\nu^i(\pi^i)$ gives

$$\frac{d}{dt}\nu^i(\pi^i) = \int_{S_i} \nabla\nu^i_z(\pi^i)\dot{\pi}^i(\,\mathrm{d}z), \tag{D.7}$$

and similarly

$$\frac{d}{dt}\nu^i(LBR^i_\eta(\pi^{-i})) = \int_{S_i} \nabla\nu^i_x(LBR^i_\eta(\pi^{-i}))\dot{l}^i_\eta(\pi^{-i})(x)\,\mathrm{d}x, \tag{D.8}$$

Combining the results from (D.4)-(D.8) we get that

$$\begin{aligned}
\dot{w}^1_\eta(\underline{\pi}) &= E^1\big(LBR^1_\eta(\pi^2), \dot{\pi}^2\big) + E^1\big(\dot{l}^1_\eta(\pi^2), \pi^2\big) - \eta\int_{S_1} \nabla\nu^1_x(LBR^1_\eta(\pi^2))\dot{l}^1_\eta(\pi^2)(x)\,\mathrm{d}x \\
&\quad - E^1(\dot{\pi}^1, \pi^2) - E^1(\pi^1, \dot{\pi}^2) + \eta\int_{S_1} \nabla\nu^1_x(\pi^1)\dot{\pi}^1(\,\mathrm{d}x), \\
\dot{w}^2_\eta(\underline{\pi}) &= E^2\big(\dot{\pi}^1, LBR^2_\eta(\pi^1)\big) + E^2\big(\pi^1, \dot{l}^2_\eta(\pi^1)\big) - \eta\int_{S_2} \nabla\nu^2_y(LBR^2_\eta(\pi^1))\dot{l}^2_\eta(\pi^1)(y)\,\mathrm{d}y \\
&\quad - E^2(\pi^1, \dot{\pi}^2) - E^2(\dot{\pi}^1, \pi^2) + \eta\int_{S_2} \nabla\nu^2_y(\pi^2)\dot{\pi}^2(\,\mathrm{d}y),
\end{aligned}$$

Now consider $\dot{w}^1_\eta(\underline{\pi})$. It is straightforward to show that

$$E^1\big(LBR^1_\eta(\pi^2), \dot{\pi}^2\big) - E^1(\pi^1, \dot{\pi}^2) = E^1(\dot{\pi}^1, \dot{\pi}^2).$$

Using this and Lemma D.1 will give,

$$\dot{w}^1_\eta(\underline{\pi}) = E^1(\dot{\pi}^1, \dot{\pi}^2) - E^1(\dot{\pi}^1, \pi^2) + \eta\int_{S_1} \nabla\nu^1_x(\pi^1)\dot{\pi}^1(\,\mathrm{d}x).$$

Expanding the final two terms and applying Lemma D.2 gives,

$$\begin{aligned}
\dot{w}^1_\eta(\underline{\pi}) &= E^1(\dot{\pi}^1, \dot{\pi}^2) - \int_{S_1} \Big[E^1(x, \pi^2) - \eta\nabla\nu^1_x(\pi^1)\Big]\dot{\pi}^1(\,\mathrm{d}x), \\
&\leq E^1(\dot{\pi}^1, \dot{\pi}^2),
\end{aligned}$$

with equality only when $\underline{\pi} = LBR_\eta(\underline{\pi})$. Identical arguments will give that

$$\dot{w}^2_\eta(\underline{\pi}) \leq E^2(\dot{\pi}^1, \dot{\pi}^2).$$

Following this we get

$$\dot{V}_\eta(\underline{\pi}) \leq E^1(\dot{\pi}^1, \dot{\pi}^2) + E^2(\dot{\pi}^1, \dot{\pi}^2) \leq 0,$$

with the final cancellation following because the game is zero sum. Equality holds only when $\underline{\pi} = LBR_\eta(\underline{\pi})$.

This shows that $V_\eta(\underline{\pi})$ is a Lyapunov function for (4.3) for any two-player zero-sum, continuous action space game when $\underline{\pi}$ is absolutely continuous. The continuity of the solution flow of the logit best response dynamics from Theorem C.5 completes the proof. □

Proof of Theorem 4.3. Assume $\underline{\pi}$ is absolutely continuous. With

$$V_\eta(\underline{\pi}) := E(\underline{\pi}) - \eta\sum_{i=1}^{N} \nu^i(\pi^i).$$

we take the derivative of each of the terms as in (D.4) and (D.7) to give

$$\begin{aligned}\dot{V}_\eta(\underline{\pi}) &= \sum_{i=1}^N \left[E\big(\dot{\pi}^i, \pi^{-i}\big) - \eta \int_{S_i} \nabla \nu^i_z(\pi^i) \dot{\pi}^i(\,\mathrm{d}z) \right], \\ &= \sum_{i=1}^N \int_{S_i} \left[E\big(z, \pi^{-i}\big) - \eta \nabla \nu^i_z(\pi^i) \right] \dot{\pi}^i(\,\mathrm{d}z).\end{aligned}$$

Using Lemma D.2 will give that $\dot{V}_\eta(\underline{\pi}) \geq 0$ with equality only when $\underline{\pi} = LBR_\eta(\underline{\pi})$. The continuity of the solution flow of the logit best response dynamics from Theorem C.5 completes the proof. □

M. Benaïm. Dynamics of stochastic approximation algorithms. *Le Séminaire de Probabilités. Lecture Notes*, 1709:1–68, 1999.

M. Benaïm and M. Hirsch. Asymptotic pseudotrajectories and chain recurrent flows. *J. Dynam. Differential Equations*, 8:141–176, 1996.

M. Benaïm and M. Hirsch. Mixed equilibria and dynamical systems arising from fictitious play in perturbed games. *Games and Economic Behaviour*, 29:36–72, 1999.

M. Benaïm, J. Hofbauer, and S. Sorin. Stochastic approximations and differential inclusions. *SIAM J. on Control and Optimization*, 44:328–348, 2003.

E. Berger. Asymptotic behaviour of a class of stochastic approximation procedures. *Probability Theory and Related Fields*, 71:517–552, 1986.

T. Börgers and R. Sarin. Learning through reinforcement and replicator dynamics. *J. of Economic Theory*, 77:1–14, 1997.

V. Borkar. Stochastic approximation with two time scales. *System & Control Letters*, 29:291–294, 1997.

V. Borkar. Asynchronous stochastic approximations. *SIAM J. on Control and Optimization*, 36:840–851, 1998.

V. Borkar. *Stochastic Approximation: A Dynamical Systems Viewpoint.* Cambridge University Press, 2008.

Z. Brzeźniak. On stochastic convolution in Banach spaces and applications. *Stochastics and Stochastic Reports*, 61:245–295, 1997.

X. Chen and H. White. Nonparametric adaptive learning with feedback. *J. of Economic Theory*, 82: 190–222, 1998.

R. Cressman. Stability of the replicator equation with continuous strategy space. *Mathematical Social Sciences*, 50:127–147, 2005.

R. Cressman, J. Hofbauer, and F. Riedel. Stability of the replicator equation for a single species with a multi-dimensional continuous trait space. *J. of Theoretical Biology*, 239:273–288, 2006.

J. Dippon and H. Walk. The averaged Robbins–Monro method for linear problems in a Banach space. *J. of Theoretical Probability*, 19:166–189, 2006.

D. Fudenberg and D. Kreps. Learning mixed equilibria. *Games and Economic Behavior*, 5:320–367, 1993.

D. Fudenberg and D. Levine. *Theory of Learning in Games.* Cambridge: MIT Press, 1998.

J. Hofbauer and E. Hopkins. Learning in perturbed asymmetric games. *Games and Economic Behavior*, 52:133 – 152, 2005.

J. Hofbauer and W. Sandholm. On the global convergence of stochastic fictitious play. *Econometrica*, 70:2265–2294, 2002.

J. Hofbauer and W. Sandholm. Evolution in games with randomly distributed payoffs. *J. of Economic Theory*, 132:47–69, 2007.

J. Hofbauer and S. Sorin. Best response dynamics for continuous zero-sum games. *Discrete and Continuous Dynamical Systems*, 6:215–224, 2006.

J. Hofbauer, J. Oechssler, and F. Riedel. Brown-von Neumann-Nash dynamics: the continuous strategy case. *Games and Economic Behavior*, 65:406–429, 2009.

V. Koval. Rate of convergence of stochastic approximation procedures in a Banach space. *Cybernetics and Systmes Analysis*, 34:386–394, 1998.

H. Kushner and D. Clark. *Stochastic approximation methods for constrained and unconstrained systems.* Springer-Verlag, 1978.

H. Kushner and G. Yin. Asymptotic properties of distributed and communicating stochastic approximation algorithms. *SIAM J. on Control and Optimization*, 25:1266–1290, 1987a.

H. Kushner and G. Yin. Stochastic approximation algorithms for parallel and distributed processing. *Stochastics*, 22:219–250, 1987b.

R. Lahkar. The continuous logit dynamic and price dispersion. *preprint*, 2012.

D. Leslie. Reinforcement learning in games. *Ph.D. thesis. University of Bristol*, 2003.

L. Ljung. Analysis of recursive stochastic algorithms. *IEEE Transactions on Automatic Control*, 22: 551–575, 1977.

D. Luenberger. *Optimization by vector space methods.* Prentice Hall, 1969.

D. Monderer and L. Shapley. Potential games. *Games and Economic Behavior*, 14:124–143, 1996.

K. Narendra and M. Thathachar. *Learning Automata: An Introduction.* New York: John Wiley & Sons, 1989.

J. Oechssler and F. Riedel. Evolutionary dynamics on infinite strategy spaces. *J. of Economic Theory*, 17:141–162, 2001.

J. Oechssler and F. Riedel. On the dynamic foundation of evolutionary stability in continuous models. *J. of Economic Theory*, 107:223–252, 2002.

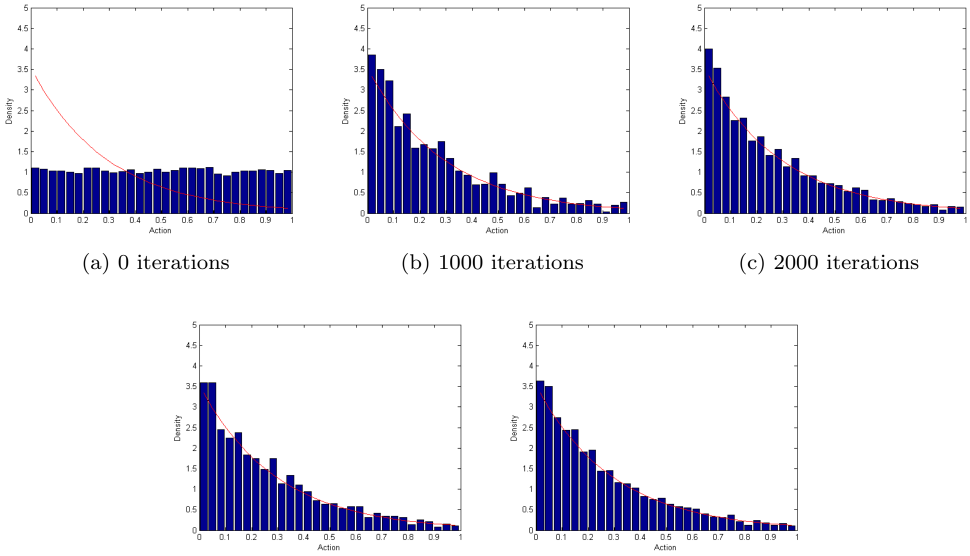
R. Seymour. Dynamics for infinite dimensional games. *mimeo, University College London*, 2002.

A. Shwartz and N. Berman. Abstract stochastic approximations and applications. *Stochastic Processes and their Applications*, 31:133–149, 1989.

H. Walk. An invariance principle for the Robbins–Monro process in a Hilbert space. *Probability Theory and Related Fields*, 39:135–150, 1977.

H. Walk and L. Zsidó. Convergence of Robbins–Monro method for linear problems in Banach space. *J. of Mathematical Analysis and Applications*, 139:152–177, 1989.

E. Zeidler. *Nonlinear Functional Analysis and its Applications.* Springer Verlag, 1986.



(a) 0 iterations

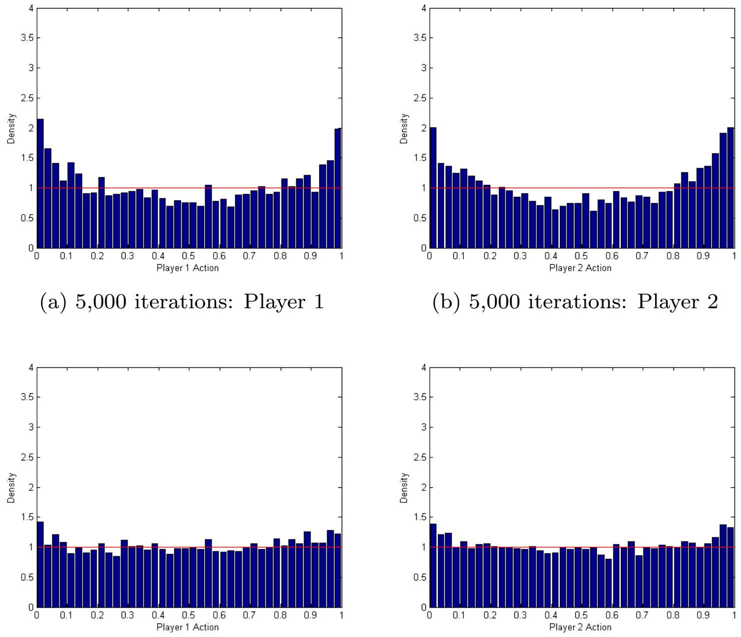
(b) 1000 iterations

(c) 2000 iterations

(d) 3000 iterations

(e) 4000 iterations

Figure 1: Evolution of our stochastic fictitious play-like process for Example 3.5 with $V = 1$, $C = 4$, $\alpha_n = (n + 20)^{-1}$ and $\eta = 0.005$. In each plot a sample from the population is shown along with the logit equilibrium of the game.



(a) 5,000 iterations: Player 1

(b) 5,000 iterations: Player 2

(c) 20,000 iterations: Player 1

(d) 20,000 iterations: Player 2

Figure 2: Evolution of our stochastic fictitious play process for Example 4.6 with $\sigma_0^1 = \delta_0$, $\sigma_0^2 = \delta_1$, $\alpha_n = (n+20)^{-1}$ and $\eta = 0.005$. In each plot a sample from the beliefs of the associated player is shown along with the logit equilibrium of the game.