Games and Economic Behavior 80 (2013) 10–38

Contents lists available at SciVerse ScienceDirect

Games and Economic Behavior

www.elsevier.com/locate/geb

Reinforcement learning in population games

Ratul Lahkar [a,*,1], Robert M. Seymour [b,2]

[a] IFMR, 24, Kothari Road, Nungambakkam, Chennai, 600 034, India
[b] Department of Mathematics, University College London, Gower Street, London, WC1E 6BT, UK

ARTICLE INFO

Article history:
Received 4 April 2011
Available online 28 February 2013

JEL classification:
C72
C73

Keywords:
Reinforcement learning
Continuity equation
Replicator dynamics

ABSTRACT

We study reinforcement learning in a population game. Agents in a population game revise mixed strategies using the Cross rule of reinforcement learning. The population state—the probability distribution over the set of mixed strategies—evolves according to the replicator continuity equation which, in its simplest form, is a partial differential equation. The replicator dynamic is a special case in which the initial population state is homogeneous, i.e. when all agents use the same mixed strategy. We apply the continuity dynamic to various classes of symmetric games. Using 3×3 coordination games, we show that equilibrium selection depends on the variance of the initial strategy distribution, or initial population heterogeneity. We give an example of a 2×2 game in which heterogeneity persists even as the mean population state converges to a mixed equilibrium. Finally, we apply the dynamic to negative definite and doubly symmetric games.

1. Introduction

There is a significant literature that models human behavior in strategic situations through a process of experience-based (or heuristic) learning (see Young, 2005 for a review of this literature). Such learning models are not forward looking in the sense that agents employing them do not seek to influence the future play of their current rivals. Much of this literature, with certain exceptions like Hopkins (1999) and Fudenberg and Takahashi (2011), is framed in the context of "global information" models (Fudenberg and Takahashi, 2011). Fudenberg and Takahashi (2011) give examples of such models. Among them are models in which the same players are repeatedly matched (for example, Fudenberg and Kreps, 1993; Börgers and Sarin, 1997), or a random matching large population model, but in which agents observe all matchings in the population (Ellison and Fudenberg, 2000). Such models have, of course, yielded significant insights into whether the dynamics of heuristic learning can lead to equilibrium behavior. However, there is an inherent incompatibility in the use of heuristic learning under the assumption of global information. Agents can exploit such an information structure to employ more strategic forward looking behavior instead of being content to let their past experience guide them.

One way to resolve this incompatibility is to adopt the idea originally put forward by Fudenberg and Kreps (1993), and subsequently applied by Hopkins (1999) and Fudenberg and Takahashi (2011). In order to justify the myopia inherent in heuristic learning, they appeal to the setting of a large population of agents who are randomly matched to play a game. Each agent has a sequence of different opponents and almost certainly never encounters the same opponent twice. Also, agents only possess "local information" (Fudenberg and Takahashi, 2011) in the sense that they only observe the outcomes of their own matches. Therefore, there is little incentive for them to develop more sophisticated strategies. This paper

* Corresponding author.
E-mail address: r.lahkar@ifmr.ac.in (R. Lahkar).
[1] My coauthor passed away before the final version of the paper could be submitted. I dedicate this paper to his memory.
[2] Deceased, July 24, 2012.

0899-8256/$ – see front matter
http://dx.doi.org/10.1016/j.geb.2013.02.006

develops a general framework to formalize this setting for heuristic learning models and analyzes the population dynamics resulting from such individual behavior. We then apply this framework to analyze reinforcement learning which is perhaps the simplest and most extensively studied form of all formal heuristic learning models.

Under reinforcement learning, an agent carries an internal mixed strategy, construed as the agent's *behavioral disposition*. If an action has yielded a high payoff in the past, then the probability assigned to it increases in the current round; or the behavior associated with the action gets reinforced. Reinforcement protocols have considerable empirical support (e.g. Erev and Roth, 1998). In this paper, we apply one particular form of reinforcement learning—the Cross (1973) rule of learning. Under this rule, the decision maker increases the probability of the action he chose in the last round in proportion to the payoff received, while reducing the probability of the other actions proportionately. The Cross rule has since been analyzed by other authors including Börgers and Sarin (1997, 2000) and Börgers et al. (2004) who characterize this rule as the prototype of a whole class of reinforcement learning schemes. Further, the Cross rule has an interesting implication, as shown in Börgers and Sarin (1997), which we use in this paper. Under this rule, the expected change in the mixed strategy weight of an action is identical to the replicator dynamic from evolutionary game theory.

In order to apply reinforcement type learning to population games, we need to assume that agents use individual mixed strategies. We then define the *population state* as a probability measure over the space of mixed strategies. This population state changes over time in response to agents' learning. The main challenge is to develop techniques to analyze the evolution of the population state. To resolve this question, we consider random matching in a two player symmetric game. Matchings last for one period and in each new matching, players revise their mixed strategies using reinforcement learning. This changes the population state. By making the time difference between successive matches go to zero, we can track the change in the population state by using a generalization to a probability measure setting of a first-order partial differential equation system akin to the *continuity equations* commonly encountered in physics in the study of conserved quantities, such as bulk fluids.[3] When applied to the Cross learning rule, this generates the *replicator continuity equation*, the name reflecting the connection between this learning rule and the replicator dynamic as shown by Börgers and Sarin (1997).

As a first application of this approach, we establish techniques to follow the evolution of the mean of the population state. The mean population state is simply the distribution of agents across different actions and, therefore, corresponds to the classical definition of population state in evolutionary game theory. We show that this mean dynamic, which we call the mean replicator dynamic, is the classical replicator dynamic adjusted by the addition of a covariance term. We also use the mean dynamic to show convergence to Nash equilibria of the mean population state in negative definite games and doubly symmetric games.

Furthermore, generalizing the notion of a population state to a distribution over mixed strategies allows us to address a very specific question—the impact of *heterogeneity* in agents' behavior on social evolution. This extends the classical evolutionary setting of homogeneity in which all agents use a fixed mixed strategy or, equivalently, a fixed mixture of pure strategies. Indeed, this approach allows us to establish the classical replicator dynamic as a very particular case of the replicator continuity equation, generated when the initial population state is homogeneous. We also show, using simple examples of 2×2 games and 3×3 pure coordination games, that heterogeneity has significant implications on social evolution. For 2×2 games, introducing heterogeneity slows down the speed of evolution but does not affect the long-run mean population state. But for 3×3 pure coordination games, heterogeneity even influences equilibrium selection. In these and more general games, from the same initial mean state, the population may converge to different pure equilibria depending upon the initial distribution over strategies. We have already noted the equivalence of the imitation based classical replicator dynamic to the particular case of an initially homogeneous population of Cross reinforcement learners. Our examples therefore reveal that evolutionary conclusions based on the classical replicator dynamic do not necessarily generalize if there exists a significant degree of heterogeneity in the population.

Hines (1980) and Zeeman (1981) are two of the pioneering papers in the biology literature to use the replicator dynamic to study the evolution of mixed strategies. In the literature on heuristic learning, the paper that is most closely related to our work is Hopkins (1999). Hopkins (1999) also applies reinforcement learning in a local information context. In fact, that paper anticipates our results on the dynamics of the population mean. Hopkins (1999) uses the relationship between the replicator dynamic and Cross reinforcement learning (Börgers and Sarin, 1997) to obtain the mean replicator dynamic and also establishes convergence of the mean population state to an evolutionarily stable state. Our approach in this paper is more general since we first derive the dynamics of the probability measure and then derive the mean replicator dynamic as a corollary of the replicator continuity equation. Hopkins (1999) establishes a local convergence result to evolutionarily stable states for negative definite games. In contrast, we establish global convergence of the mean population state to Nash equilibrium in negative definite games.

Our more fundamental approach also allows us to compute the asymptotic distribution to which a population state may converge. Of course, in the case of convergence of the mean population state to a pure Nash equilibrium, the asymptotic distribution is trivial. However, in case of convergence of the mean to a mixed equilibrium, the asymptotic distribution may

[3] In physics, the continuity equation is a linear partial differential equation that describes the rate of change in the mass of fluid in any part of the medium through which it is flowing. See, for example, Margenau and Murphy (1962). However, our continuity equations concern the change in probability mass of agents in any part of the mixed strategy space, and differ from classical versions encountered in physics in that they contain non-linearities. See Ramsza and Seymour (2010) for an application of continuity equation techniques to track the evolution of fictitious play updating weights in a population game. Our paper provides a more general method of constructing continuity equations that can be used for a variety of learning algorithms.