Available online at www.sciencedirect.com

SCIENCE DIRECT®

Games and Economic Behavior 53 (2005) 110–125

GAMES and Economic Behavior

www.elsevier.com/locate/geb

Attainability of boundary points under reinforcement learning

Ed Hopkins [a,*], Martin Posch [b]

[a] *Department of Economics, University of Edinburgh, Edinburgh EH8 9JY, UK*
[b] *Department of Medical Statistics, Medical University of Vienna, Vienna 1090, Austria*

Received 20 December 2002

Available online 5 November 2004

Abstract

This paper investigates the properties of the most common form of reinforcement learning (the "basic model" of Erev and Roth) [Amer. Econ. Rev. 88 (1998) 848–881]. Stochastic approximation theory has been used to analyse the local stability of fixed points under this learning process. However, as we show, when such points are on the boundary of the state space, for example, pure strategy equilibria, standard results from the theory of stochastic approximation do not apply. We offer what we believe to be the correct treatment of boundary points, and provide a new and more general result: this model of learning converges with zero probability to fixed points which are unstable under the Maynard Smith or adjusted version of the evolutionary replicator dynamics. For two player games these are the fixed points that are linearly unstable under the standard replicator dynamics.

JEL classification: C72; C73; D83

Keywords: Learning in games; Reinforcement learning; Stochastic approximation; Replicator dynamics

* Corresponding author.
E-mail addresses: e.hopkins@ed.ac.uk (E. Hopkins), martin.posch@meduniwien.ac.at (M. Posch).
URLs: http://homepages.ed.ac.uk/ehk/, http://www.meduniwien.ac.at/user/martin.posch.

0899-8256/$ – see front matter
doi:10.1016/j.geb.2004.08.002

1. Introduction

Whilst equilibrium analysis has been the mainstay of economic theory for many years, economists have more recently turned to non-equilibrium explanations of human behaviour based on learning models. This approach has found considerable success in explaining how people behave in economic experiments (Roth and Erev, 1995; Erev and Roth, 1998; Camerer and Ho, 1999). This in turn leads to the intriguing prospect of using adaptive learning models in economic applications in the wider world outside the laboratory, for example, to explain consumer behaviour (Erev and Haruvy, 2001; Hopkins, 2003) or to design economic mechanisms robust to bounded rationality (Sandholm, 2002). However, such applications are hampered by the fact that learning models are more difficult to work with than equilibrium analysis: one has first to calculate the equilibria and then consider such issues as stability and convergence. Such difficulties are compounded when one works with stochastic rather than deterministic systems. Therefore, recent results that indicate that stochastic learning models can behave in the same way, at least asymptotically, as deterministic models such as the evolutionary replicator dynamics are particular valuable.[1]

In this paper, we highlight some previously hidden technical difficulties in the application of this methodology, and offer some solutions. In particular, we show that two important results in the theory of stochastic approximation cannot be easily applied to the reinforcement learning model popularised by Erev and Roth (1998). This implies that existing results cannot rule out the possibility that this learning process converges to a state which is not a Nash equilibrium, something that is known to be impossible for the deterministic replicator dynamics. Beggs (2002) makes significant progress on this issue. He shows that for single person decision problems the process cannot converge to a suboptimal action.[2] We give a more general result applicable to all normal form games and show that starting from a position where all strategies are played with positive probability this learning process will indeed converge with probability zero to any point which is linearly unstable under the Maynard Smith or adjusted version of the evolutionary replicator dynamics. This rules out survival of suboptimal actions in decision problems and, in games, rules out convergence to points which are not Nash equilibria. We go on to show how these results can be used in a simple practical application.

In this work we are able to clarify some earlier misunderstandings concerning the convergence of reinforcement learning to boundary points. Stochastic approximation examines the behaviour of a learning process by investigation of an ordinary differential equation or ODE derived from the expected motion of the learning process. One classic result is that if the ODE has a global attractor, the learning process converges with probability one to that point. An example of such a result is Corollary 6 of Theorem 4 of Benveniste et al. (1990, pp. 45–46; also Theorem 17, p. 239). However, the evolutionary replicator dynam-

[1] Analysis of single person decision making is found in Arthur (1993), Rustichini (1999) and Sarin and Vahid (1999), for games in Börgers and Sarin (1997), Posch (1997), Ianni (2000) and Hopkins (2002), and for both in Laslier et al. (2001) and Beggs (2002).

[2] He also proves that, in games, strategies which are dominated by a mixed strategy or removed by iterated deletion of dominated strategies are eliminated by reinforcement learning. Additionally he shows, for constant sum games, that in the limit the players play Nash.