Stochastic Evolutionary Game Theory

RICHARD BARON\textsuperscript{a,*}, JACQUES DURIEU\textsuperscript{a}, HANS HALLER\textsuperscript{b}, PHILIPPE SOLAL\textsuperscript{a}

January, 2002

\textsuperscript{a}CREUSET, University of Saint-Etienne, 42100 Saint-Etienne, France.
\textsuperscript{b}Department of Economics, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061-0316, USA.

Introduction

Do economic agents really compute optimal solutions to complex social problems? In normative game theory, players are fully rational and can find optimal solutions even to the most complex and demanding problems. Recent trends in game theory study adaptive learning models, in which players or agents have a limited understanding of their environment and act according to boundedly rational behavioral rules. Decisions taken on the basis of myopic responses, reinforcement rules, or other short-sighted rules, coalesce in the long-run into limit sets or conventions. This chapter is organized as follows. In section 1, we give a brief overview of some recent developments in the area of adaptive learning in games or evolutionary game theory. We refer to Walliser (1998) for a more complete presentation. In section 2, we present a general framework to model adaptive learning processes with persistent noise and specifically define the notion of stochastic stability. We provide several leading examples, including a brief account of recent work of ours on control costs and logit rules.

1 Models of Adaptive Learning in Games

Nash equilibrium is the central concept in non-cooperative game theory. There exists a literature on adaptive learning aimed at investigating to what extent learning procedures can lead agents to select among the Nash equilibria of the underlying game. We are going to present three main strands of literature that have emerged in the context of adaptive learning in games.

One strand of the literature concentrates on adaptive procedures based on backward-looking criteria. A procedure is said to be backward-looking
when it does not model the opponents’ future behavior given information about the history of play. For instance in reinforcement procedures, agents compute from their past payoffs an index for each action and then choose a mixed strategy where an action’s probability is strictly increasing in its index. In this way successful behavior is reinforced. An agent following such a procedure need not be aware that he faces other players. These rather naive procedures have been first suggested by psychologists and animal behaviorists in the fifties and have been tested in laboratory settings. Another instance of backward-looking procedures are imitation rules to the extent that agents are unable to form beliefs about their opponents’ future choices. For example, agents may copy the most popular strategy in the last play or imitate the first behavior they observe. One can imagine more sophisticated imitation procedures, where agents observe a sample from a finite history of their opponents’ past play and payoffs and imitate a strategy that yielded the maximum average payoff.

A second strand of the literature focuses on learning procedures that are based on Fictitious Play algorithms, originally for two-player games. In the simplest version, an agent keeps track of the empirical distribution of the opponent’s strategy choices and responds optimally to that distribution. These procedures are forward-looking, because they model the opponent’s future behavior based on past behavior. Learning procedures of this kind have been analysed extensively, especially their convergence properties.

The third strand of literature discusses models in which individual choices play a secondary role. Each agent of a large but finite population is programmed to play a pure strategy. A state of the system is a probability vector where each component is the population share of individuals programmed to play a particular pure strategy. Agents reproduce continuously according to a reproduction rule which assures that the most efficient strategies spread progressively in the population. Payoffs represent thus fitness, measured as the number of offsprings, and each offspring inherits its single parent’s strategy. For instance the replicator rule specifies that the growth rate of a strategy is proportional to relative fitness, that is the fitness of the strategy less the current population average fitness. These population learning models have been introduced by biologists in the sixties.

To conclude this section we state a general result. In the context of a large population where agents are randomly matched, it can be shown that well defined aggregate random sequences generated by some reinforcement, imitation or Fictitious Play procedures are close to the trajectory of the
replicator dynamics. In other words the aggregate dynamics of these models belong to the same class. This suggests that in many cases, the aggregate of an adaptive learning procedure is qualitatively different from the learning process at the level of the individual; see for example Hopkins (1999).

2 Stochastic Stability

We first define the basic strategic situation that individuals are supposed to be learning about. Then we introduce the evolutionary model with persistent noise.

The Recurrent Game

The setting is a finite $n$-player game in strategic form

$$\mathcal{G} = (I, (S_i)_{i \in I}, (u_i)_{i \in I}).$$

$I = \{1, \ldots, n\}$ is the finite set of players. For player $i \in I$, $S_i$ is $i$’s finite strategy set, with generic elements $s_i, s_i', k_i$. Let $S_{-i} = \prod_{j \in I \setminus \{i\}} S_j$ be the set of strategy profiles of all players except $i$, with generic elements $s_{-i} = (s_j)_{j \in I \setminus \{i\}}$. $\mathcal{S} = \prod_{i \in I} S_i$ denotes the set of strategy profiles of all players, with generic elements $s = (s_i)_{i \in I}$. An element $s \in \mathcal{S}$ will also be called a play. In slight abuse of notation, we sometimes write $(s_i, s_{-i})$ instead of $s$. Player $i \in I$ has payoffs represented by the utility function $u_i : \mathcal{S} \to \mathbb{R}$.

For $i \in I$, let $Br_i : S_{-i} \to S_i$ be the $i$th pure best-reply correspondence which maps each strategy profile $s_{-i} \in S_{-i}$ to the nonempty and finite set:

$$Br_i(s_{-i}) = \arg \max_{s_i \in S_i} u_i(s_i, s_{-i}).$$

The combined pure best-reply correspondence $Br : \mathcal{S} \to \mathcal{S}$ of the game $\mathcal{G}$ is defined as the product set of all players’ pure best-reply correspondences

$$Br(s) = \prod_{i \in I} Br_i(s_{-i}).$$

A Nash equilibrium in pure strategies is a fixed point of $Br$, that is a strategy profile $s^* \in \mathcal{S}$ such that $s^* \in Br(s^*)$. The game $\mathcal{G}$ constitutes the recurrent game.
Evolution with Persistent Noise

Let $t = 1, 2, \ldots$ denote successive time periods. The recurrent game $G$ is played once in each period by a population of $n$ boundedly rational agents. In each period $t$, one agent is drawn with probability $q_t > 0$ from this population to adjust its strategy and does so according to a perturbed adaptive rule. An adaptive rule is a procedure which determines a mixed strategy in each period based on the history of past play. A perturbed adaptive rule is an adaptive rule indexed by a level of noise. Typically, the noise is attributed to mistakes or trembles on the part of the agents. With some positive probability, the reviewing agent fails to implement the computed mixed strategy. There is, however, a great deal of variation in how precisely noise is introduced into the adaptive rules. To obtain a general framework which can encompass several of these variants some additional definitions are needed. The (truncated) history or state at time $t \geq m$ is the sequence $h^t = (s^{t-m+1}, \ldots, s^t)$ where $m$ is the memory size of all agents. Let $H = \mathcal{S}^m$ be the finite set of histories of length $m$ and let $h$ be an arbitrary element of this set. For $i \in I$, let $\Delta_i$ denote his set of mixed strategies. A perturbed adaptive rule is a function $p_i^\epsilon : H \times \mathbb{R}_{++} \rightarrow \Delta_i$ which assigns to each state $h \in H$ and $\epsilon \in \mathbb{R}_{++}$ a mixed strategy $p_i^\epsilon(h) = (p_i^{\epsilon, s_i}(h))_{s_i \in S_i} \in \Delta_i$.

We will assume for all $i \in I$,

(i) $p_i^\epsilon$ does not depend on time.

(ii) For any fixed $\epsilon \in \mathbb{R}_{++}$, $p_i^\epsilon(h)$ is completely mixed for all $h \in H$.

(iii) $p_i^\epsilon$ is continuous in $\epsilon \in \mathbb{R}_{++}$ and $p_i^\epsilon = \lim_{\epsilon \to 0} p_i^\epsilon$ exists.

Example 1 Myopic best-reply with Bernoulli trembles

Fix $m = 1$ so that $h^t = s^t$ for every $t$. The parameter $\epsilon \in (0, 1)$ describes the probability that the selected agent $i \in I$ fails to implement his myopic best-reply $Br_i(s_{-i}^t)$. Thus, the probability of playing a best reply $s_i^* \in Br_i(s_{-i})$ against $s_{-i}$ is given by

$$p_i^{\epsilon, s_i^*}(s) = \frac{1 - \epsilon}{|Br_i(s_{-i})|} + \frac{\epsilon}{|S_i|},$$

and the probability that $i$ chooses a non-best reply is

$$p_i^{\epsilon, s_i}(s) = \frac{\epsilon}{|S_i|} \text{ for all } s_i \notin Br_i(s_{-i}).$$

Note that as $\epsilon \to 0$ the probability of playing any strategy that is not a best reply against $s_{-i}$ goes to zero. This class of models of evolution with Bernoulli
trembles have been introduced by Kandori, Mailath, and Rob (1993), Young (1993), Ellison (1993) and in another context by Kirman (1993).

Example 2  Logistic Fictitious Play with bounded memory

Fix $\epsilon \in R_{++}$. Suppose $|I| = 2$, and $m > 1$. Let $b_{-i}(s_{-i})$ be the proportion of time that strategy $s_{-i} \in S_{-i}$ was played by agent $-i$ in the history $h^t$, with $\sum_{s_{-i} \in S_{-i}} b_{-i}(s_{-i}) = 1$. Let $b_{-i}^t = (b_{-i}^t(s_{-i}))_{s_{-i} \in S_{-i}}$ be the probability vector associated with the empirical distribution. At the end of period $t$, the reviewing agent, say $i$, computes the profile $b_{-i}^t$ from $h^t$. This constitutes the belief held by $i$ about $-i$’s play in the next period. Now suppose that the probability of playing the best reply $s_i^*$ against $b_{-i}^t$ is given by

$$p_i^{s_i^*}(h) = \frac{\exp[\tilde{u}_i(s_i^*, b_{-i})/\epsilon]}{\sum_{s_{-i} \in S_{-i}} \exp[\tilde{u}_i(s_{-i}, b_{-i})/\epsilon]}$$

where $\tilde{u}_i(\cdot, b_{-i})$ denotes $i$’s expected payoff with respect to $(\cdot, b_{-i}) \in S_i \times \Delta_{-i}$. This is logistic Fictitious Play with memory $m$: each strategy is played in proportion to an exponential function of the utility it would have yielded historically. It corresponds to the logit rule. As in the previous example, the probability of playing any strategy that is not a best reply against $b_{-i}$ goes to zero as $\epsilon \to 0$. This model is analysed by Fudenberg and Levine (1995).

The logit rule is also used by Blume (1993), Young (1998) and Baron et al. (2002) in games played on graphs\(^1\). In the following example we present a method to endogenize the logit adaptive rule.

Example 3  Control costs and logit rules

Here we follow van Damme and Weibull (1998, 1999). Fix $m = 1$ so that $h^t = s^t$ for every $t$. The key assumption is that an agent can reduce the amount of trembles by expending an extra effort and incurring some disutility or control costs. The specific control costs assumed here will be such that the disutility of eliminating trembles completely will be prohibitive. The details

\(^1\)A graph is a way to define the spatial structure of a game, though in most games played on (non-directed) graphs, the spatial structure represents mainly the matching technology; each player is matched with his neighbours on the graph to play a two-player game.
are as follows. Let each agent \( i \in I \) have a specific control cost function \( E_i : \Delta_i \to \mathbb{R} \) defined by

\[
E_i(p_i) = \ln \alpha_i + \sum_{s_i \in S_i} p_i^{s_i} \ln p_i^{s_i}
\]

for \( p_i = (p_i^{s_i})_{s_i \in S_i} \in \Delta_i \), with \( \alpha_i \geq |S_i| \) and the convention \( 0 \cdot (-\infty) = 0 \) which makes \( E_i \) continuous on its domain.
For \( \epsilon \in \mathbb{R}_{++} \) we associate to \( \bar{G} \) a perturbed game in strategic form,

\[
\bar{G}^\epsilon = (I, (\Delta_i)_{i \in I}, (\bar{u}_i^\epsilon)_{i \in I}).
\]

For \( i \in I \), the payoff function \( \bar{u}_i^\epsilon : \prod_{j \in I} \Delta_j \to \mathbb{R} \) is given by

\[
\bar{u}_i^\epsilon(p) = \bar{u}_i(p) - \epsilon E_i(p_i)
\]

where \( \bar{u}_i(p) \) denotes here \( i \)'s expected payoff with respect to \( p \). The control cost parameter \( \epsilon \) reflects the extend to which the game is perturbed and determines how much best reply play is distorted. Notice that given \( \epsilon \), the costs are minimised when \( i \)'s play is perfectly random, and costs are maximised when \( i \)'s play is deterministic, i.e. a particular pure strategy is played. As before, one randomly determined agent has the opportunity to adjust his strategy. He does so by choosing the myopic best reply in \( \bar{G}^\epsilon \) against the current play of the other agents. If the current state is \( s \) and agent \( i \) is the next one to adjust his strategy, then he chooses \( p_i \in \Delta_i \) that maximises

\[
\bar{u}_i(p_i, s_{-i}) - \epsilon E_i(p_i)
\]

The first order condition yields for all \( s_i \in S_i \) :

\[
p_i^{\epsilon, s_i}(s) = \frac{\exp[u_i(s_i, s_{-i})/\epsilon]}{\sum_{s'_i \in S_i} \exp[u_i(s'_i, s_{-i})/\epsilon]}
\]

(1)

The main feature of the logit rule (1) is that the probability of trembles is related to the loss in payoff from playing a non-best reply. In other words, trembles are state dependent. This is not the case with Bernoulli trembles since in these models the probability \( \epsilon/|S_i| \) of choosing a non-best reply does not depend on the state (see Example 1).

Keeping in mind these examples we can go one step further in the study of our model of evolution. Given a history \( h^t = (s^{t-m+1}, \ldots, s^t) \) at time \( t \), the stochastic process moves to a state of the form \( h^{t+1} = (s^{t-m+2}, \ldots, s^{t+1}) \) in the next period, where \( s^{t+1} = (s'_i, s'_{-i}) \) for some \( i \in I \). Such a state is called
a successor of $h^t$. Our assumptions imply that the process moves from a current state to a successor in each period according to $(q_i)_{i \in I}$ and $(p^t_i)_{i \in I}$. These operations define a finite Markov chain on the state space $H$. Let $M^\varepsilon$ denote its transition matrix. A model of gradual evolution with persistent noise may be summarized as a quadruplet $(G, (q_i)_{i \in I}, (p^t_i)_{i \in I}, M^\varepsilon)$ where for all $i \in I$, $q_i \in (0, 1)$ and $p_i^t$ satisfies properties (i) through (iii).

For any fixed $\varepsilon \in B_{++}$, the Markov chain corresponding to the model is irreducible and aperiodic. Let $(\mu^\varepsilon_h)_{h \in H}$ be its unique stationary distribution i.e. $\mu^\varepsilon$ is a row probability vector on $H$ such that $\mu^\varepsilon = \mu^\varepsilon M^\varepsilon$. Our main concern is the characterisation of the support of the limit stationary distribution $\mu^* = \lim_{\varepsilon \to 0} \mu^\varepsilon$. We write $\mathcal{C}(\mu^*)$ for its support:

$$\mathcal{C}(\mu^*) = \{ h \in H : \mu^*_h > 0 \}$$

The states belonging to $\mathcal{C}(\mu^*)$ will be referred to as stochastically stable states. In some of the literature, they are called (ultra) long-run equilibria. First, observe that $\mu^*$ need not exist. But it does exist in many instances. Suppose it does for the remainder of this chapter. By (iii), $M^* = \lim_{\varepsilon \to 0} M^\varepsilon$ exists. Consequently,

$$\lim_{\varepsilon \to 0} \mu^\varepsilon = \lim_{\varepsilon \to 0} \mu^\varepsilon M^\varepsilon \Rightarrow \mu^* = \mu^* M^*.$$  \hspace{1cm} (2)

$M^*$ is referred to as the transition matrix of the unperturbed process. A direct consequence of (2) is that every stochastically stable state must be contained in a limit set of the unperturbed process. To clarify matters, in Examples 1 and 3 the unperturbed process is the asynchronous and myopic best-reply dynamics. In Example 2, the unperturbed process is the Fictitious Play algorithm with memory $m$. Tree construction algorithms have been developed by Solberg (1975) and Frédlin and Wentzell (1984) to analyse stochastic stability in models of gradual evolution with persistent Bernoulli noise. These algorithms are used and developed by Kandori, Mailath and Rob (1993), Young (1993), Ellison (1993), Kandori and Rob (1995), and Samuelson (1997). Recently, Ellison (2000) proposed a new method which is often computationally more efficient. Its main drawback is that it gives only sufficient conditions for stochastic stability and thus sometimes fails to provide a full characterization.

In the rest of the paper, we present a method to characterize the support of $\mu^*$ when the adaptive rules are given by (1), $m = 1$, and the game has a potential. This method is formally equivalent to simulated annealing.
Potential Games

Potential games are introduced and characterized by Monderer and Shapley (1996). A game \( \mathcal{G} \) is a potential game if there exists a function \( \mathcal{P} : \mathcal{S} \rightarrow \mathbb{R} \) such that for any \( i \in I, \ s \in \mathcal{S}, s'_i \in S_i \),

\[
u_i(s) - u_i(s'_i, s_{-i}) = \mathcal{P}(s) - \mathcal{P}(s'_i, s_{-i}).\]

Then \( \mathcal{P} \) is called a potential of \( \mathcal{G} \). If \( \mathcal{G} \) has a potential, then for all \( \epsilon \in \mathbb{R}_{++} \), the logit rule (1) becomes

\[
p_{i}^{\epsilon,\pi}(s) = \frac{\exp[\mathcal{P}(s_i, s_{-i})/\epsilon]}{\sum_{s'_i \in S_i} \exp[\mathcal{P}(s'_i, s_{-i})/\epsilon]},
\]

and the stationary distribution assumes the Gibbs-Boltzmann form

\[
\mu^{\epsilon} = \frac{\exp[\mathcal{P}(s)/\epsilon]}{\sum_{s' \in \mathcal{S}} \exp[\mathcal{P}(s')/\epsilon]}.
\]

That (3) implies (4) has been shown for special cases by Mazaika (1987), Blume (1993, 1997), Young (1998), Haller and Outkin (1999), among others. It follows from (4) that the stochastically stable states are the maximizers of \( \mathcal{P} \), i.e.

\[
\mathcal{C}(\mu^{*}) = \{ s^{*} \in \mathcal{S} : s^{*} \in \arg\max_{s \in \mathcal{S}} \mathcal{P}(s) \},
\]

and, consequently, constitute a subset of the Nash equilibria of \( \mathcal{G} \). It further follows that all stochastically stable states are equally likely.

References


