Updating Strategies Through Observed Play

-

Optimization Under Bounded Rationality [1]

R. Cressman
Department of Mathematics
Wilfrid Laurier University
Waterloo, Ontario, N2L 3C5
Canada

K.H. Schlag
Economic Theory III
University of Bonn
Adenauerallee 24-26
D-53113 Bonn, Germany
www.econ3.uni-bonn.de
schlag@econ3.uni-bonn.de

University of Bonn Discussion Paper No. **B-432**
April, 1998

[1]We wish to thank Dorothea Herreiner for helpful comments. Financial support from the Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 303 at the University of Bonn and from the Natural Sciences and Engineering Research Council of Canada is gratefully acknowledged.

Abstract

Individuals repeatedly face a multi-decision task with unknown payoff distributions. They have minimal memory and update their *strategy* by observing previous *play* (and not strategy) of someone else. We select behavior rules that increase average payoffs as often as possible in a large population where all use the same rule. Here imitation generalizes to a pasting procedure. When decisions within the task are unrelated, individuals eventually learn the efficient strategy but the underlying dynamic is not monotone. However, when choices influence which decisions are subsequently faced in the task, play may not be efficient in the long run as it approaches a Nash equilibrium of the agent normal form.

Keywords: Multi-Armed Bandit, improving, undominated behavioral rule, play-wise imitating, replicator dynamic, monotone dynamics.

JEL classification: C72, C79

1 Introduction

People are constantly faced with choices, that affect their well-being, without a complete knowledge of the consequences of their possible actions. In a society, people in these situations often make their choice through observing (or *sampling*) actions and their consequences of other individuals. As argued by Samuelson (1997), they develop rules of thumb, based on their limited information, which their experience suggests will lead to their best choice. We call these *behavioral rules*.

In extensive-form games, strategies are typically very complex, consisting of a plan of action at each decision node. The ability to observe the entire strategy of other players is particularly unnatural in large extensive-form games. To relax this assumption and to assume that only previous *play* (i.e., choices made on the outcome path) is observable has led to investigations by Fudenberg and Levine (1993) and Nöldeke and Samuelson (1993). In both of these models, individuals are still rather sophisticated, being completely informed about the game's payoff and able to calculate best replies. Models of boundedly rational behavior with more constraints on individual rationality (e.g., imitation models) typically appeal to the normal form when analyzing extensive-form games (e.g., Cressman, 1997; Gale et al., 1995). When motivated by models of imitation, this means that *strategies* are sampled. Only recently have attempts been made to investigate models of imitation based on sampled *play* (Cooper, 1997; Ruebeck, 1997). Results for the repeated Prisoners' Dilemma are shown to be drastically altered. However, these investigations rely on specific individual behavior which leaves open the possibility that no change in prediction arises under appropriately chosen alternative behavior. We set out to investigate whether universal rules of behavior exist that can cope with this imperfect information in an optimal way. For this analysis, we restrict attention to *Multi-Decision Tasks* which are one-player extensive-form games against nature with chance nodes.[1] A strategy describes a plan of action in each node (or decision) and play describes the decisions faced and actions chosen during one task.

We consider a large population of identical individuals, each individual independently and repeatedly facing the same Multi-Decision Task. Each decision in the task has a finite number of choices, play in the task yields payoffs in a known bounded interval. This

[1] Technically, the extensive form game against nature is a perfect information game since each decision is a singleton information set. However, the phrase "perfect information" is not to be confused with the incomplete information inherent in our tasks where the explicit payoff consequences of action(s) chosen is not known.

is all an individual initially knows about the task. Each time play in the task is over, an individual receives information about the play of others. He samples at random an individual from the population and observes last play (not the strategy) and payoff of the sampled individual. Then, given his behavioral rule, the individual determines which strategy to use the next time he faces the task.

Our assumptions of bounded rationality are as follows. We a

ERROR: undefined
OFFENDING COMMAND: con

STACK:

[0.24 0 0 -0.24 18.0 820.0]
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0