論文の内容の要旨

論文題目　Essays on Learning and Preference Evolution in Games

氏　　名　山森　哲雄

In many situations, even if people enter into strategic interactions with others, they follow the simple decision rule, according to which, they react to the environment without elaborating a strategy because their rationality is bounded by various reasons. People cannot always deduce the opponents' actions at a point in time since they often do not know the opponents' preferences or cannot gauge how smart they are. People cannot always calculate the best response since they often do not know all the actions that are available to them or the payoffs of these actions.

Economists have formally modeled the dynamics of human behavior, which describes how people learn to satisfy their preferences under a situation with bounded rationality. The earliest learning process was presented by Cournot's study of duopoly (1838). In his simultaneous best-reply dynamics, at each period, each firm chooses the quantity that maximizes its profit under the assumption that the other firms continue to choose the quantity of the previous period. Brown (1951) introduced a more sophisticated learning process called fictitious play. In this dynamics, at each period, each player predicts that the probability distribution of the opponents' play is the empirical frequency distribution of their past play, and the player simply chooses a best reply to it. A more sophisticated version of fictitious play was also studied by Milgrom and Roberts (1990, 1991).

The main concern addressed in these studies dealing with learning in games is how or whether players can learn to play Nash equilibrium under a particular learning process. In a general class of games, a learning process does not necessarily converge to Nash equilibrium. Therefore, much attention has been devoted to find a class of games in which the learning process converges to some Nash equilibrium. For example, Robinson (1951), Miyasawa (1961), Krishna (1991), and Monderer and Shapley (1996) found several classes of games in which the fictitious play converges to a Nash equilibrium, whereas Shapley (1964) provided a game in which it does not converge.

This thesis studies three types of the dynamics of human behavior. Two of them are learning processes called best-reply dynamics and better-reply dynamics. The other dynamics is a learning process in which the players' preferences are adjusted over time in addition to their strategies. The following situation is common to these dynamics: Finite numbers of myopic players repeatedly play a game in discrete times. Players cannot always change their strategies; instead, each player randomly receives an opportunity to revise his/her strategy at each period. The probability of the revision opportunity is strictly between 0 and 1, and independent over time and across players.

In the best-reply dynamics, if the current strategy of the player with a revision opportunity maximizes his/her payoff given the opponents' strategies, then he/she continues to choose the strategy. Otherwise, he/she switches to one of the best replies to the current strategy profile with equal probability. This dynamics is similar to the best-response dynamics introduced by Gilboa and Matsui (1991). However, there are differences between their dynamics and the one studied in this thesis. They consider a continuum of players and assume that a deterministic fraction of the players always revise their strategies.

Several classes of games have been found to have the global convergence property under the best-reply dynamics: the sequence of the strategy profiles from any initial strategy profile almost surely converges to some Nash equilibrium. Kandori and Rob (1995) showed that the best-reply dynamics globally converges to some Nash equilibrium in every finite, symmetric, strict supermodular game with totally ordered strategy sets. Kukushkin (2004) obtained the same global convergence result in games with additive aggregation.

Chapter 2 introduces the pure Nash equilibrium property (PNEP) as a sufficient condition for the global convergence under the best-reply dynamics. A game has PNEP if there is a pure strategy Nash equilibrium in any game that has been constructed by restricting the strategies of the players in an original game to its subset. Any finite ordinal potential game and any finite supermodular game have the PNEP. We show that any finite, two-player game with the PNEP has the global convergence property under the best-reply dynamics.

It is implicitly assumed that each player is able to calculate the best replies to the current strategy profile in the best-reply dynamics. In many situations, however, a player may be unable to perform this task since he/she often does not know all the actions that are available to him/her or the payoffs of these actions. The following better-reply dynamics is suitable for modeling human behavior under such a situation.

In the better-reply dynamics, each player with a revision opportunity picks up one strategy from his/her strategy set with equal probability and compares his/her current payoff with the payoff that he/she would receive if he/she opted for the new strategy against the current strategy profile. He/she switches to the new strategy if and only if it gives a higher payoff.

Note that the possible sequences of the strategy profiles in the best-reply dynamics are fewer than those in the better-reply dynamics. Therefore, any game that has the global convergence property under the best-reply dynamics also has the same property under the better-reply dynamics.

Chapter 3 focuses on quasi-supermodular games, which include supermodular games and investigates the global convergence under both the better- and best-reply dynamics. We show the following global convergence results: every quasi-supermodular game has the global convergence property under the better-reply dynamics, and every quasi-supermodular game with totally ordered strategy sets has the global convergence property under the best-reply dynamics.

These two results strengthen the result provided by Friedman and Mezzetti (2001), who showed that the better-reply dynamics globally converges to some Nash equilibrium in every supermodular game with totally ordered strategy sets. In the first result, we relax their assumption that strategy sets are totally ordered. In the second result, we maintain the completeness assumption of ordering, but show the global convergence under the best-reply dynamics instead of the better-reply dynamics. Furthermore, we also weaken the condition of supermodularity to quasi-supermodularity in the two results.

The players' preferences, which are given by the payoffs of the underlying game, are invariable through time in the best- and better-reply dynamics. In contrast, Chapter 4 studies the dynamic process in which players learn to behave on the basis of their preferences, which are in turn shaped by natural selection. Now, the payoffs of the underlying game are not the players' preferences, but their fitness. At each period, as in the best-reply dynamics, each player with a revision opportunity plays a best reply to the current strategy profile in terms of his/her preference, which need not match the underlying fitness. After the strategy profile of the next period is determined, some players are randomly selected. If a selected player does not have the highest fitness among the players, his/her preference is replaced by a new one. Such a dynamic process is called the preference evolution or the indirect evolutionary approach.

Studies on the preference evolution in a strategic environment have focused on the linkage between people's fitness and whether or not their preferences can serve as credible commitment devices. For example, even if a player's preference does not match the underlying fitness, when his/her preference is observable and the players are rational, he/she may gain higher fitness than those whose preferences correspond to the underlying fitness by committing to the certain strategy. Ultimately, the offspring who inherits such a preference dominates over the society (e.g., Güth and Yaari (1992), Güth (1995), Bester and Güth (1998)). In contrast, when the population is large and the players' preferences are not observable, the preferences that do not match the underlying fitness are deprived of their capability to make commitments. Consequently, the preferences of the surviving players are consistent with the underlying fitness (e.g., Ok and Vega-Redondo (2001)).

The preference evolution model studied in Chapter 4 shows that preferences have important consequences for the outcomes even if they do not serve as commitment devices. We focus on the underlying game having two actions, where the players' common fitness exhibits the economies of scale. Both the states in which all players choose the same actions are Nash equilibria. We first show that our dynamics globally converges to one of the Nash equilibria of the underlying game. In this sense, the players' preferences are irrelevant to the result of the underlying game. However, we show that if rare mutations are introduced into the process of the preference evolution, the players' preferences may drift without affecting the equilibrium behavior, and these drifts may influence the results of the equilibrium selection in the underlying game.

Note that the process of changing the preferences can be analogized as the process of changing the manager of the firm in a production economy: The manager's preferences may be different from the firm's profits or the owner's preferences; if the firm does not gain the highest profit among all its competitors, then the owner dismisses the current manager and employs a new one. In Chapter 4, our model of the preference evolution is applied in an economic setting.