Lecture VIII: Learning

Markus M. Möbius

March 6, 2003

Readings: Fudenberg and Levine (1998), The Theory of Learning in Games, Chapter 1 and 2

1 Introduction

What are the problems with Nash equilibrium? It has been argued that Nash equilibrium are a reasonable minimum requirement for how people should play games (although this is debatable as some of our experiments have shown). It has been suggested that players should be able to figure out Nash equilibria starting from the assumption that the rules of the game, the players' rationality and the payoff functions are all common knowledge.[1] As Fudenberg and Levine (1998) have pointed out, there are some important conceptual and empirical problems associated with this line of reasoning:

1. If there are multiple equilibria it is not clear how agents can coordinate their beliefs about each other's play by pure introspection.

2. Common knowledge of rationality and about the game itself can be difficult to establish.

[1] We haven't discussed the connection between knowledge and Nash equilibrium. Assume that there is a Nash equilibrium σ^* in a two player game and that each player's best-response is unique. In this case player 1 knows that player 2 will play σ_2^* in response to σ_1^*, player 2 knows that player 1 knows this etc. Common knowledge is important for the same reason that it matters in the coordinated attack game we discussed earlier. Each player might be unwilling to play her prescribed strategy if she is not absolutely certain that the other play will do the same.

3. Equilibrium theory does a poor job explaining play in early rounds of most experiments, although it does much better in later rounds. This shift from non-equilibrium to equilibrium play is difficult to reconcile with a purely introspective theory.

1.1 Learning or Evolution?

There are two main ways to model the processes according to which players change their strategies they are using to play a game. A *learning model* is any model that specifies the learning rules used by individual players and examines their interaction when the game (or games) is played repeatedly. These types of models will be the subject of today's lecture.

Learning models quickly become very complex when there are many players involved. *Evolutionary models* do not specifically model the learning process at the individual level. The basic assumption there is that some unspecified process at the individual level leads the population as a whole to adopt strategies that yield improved payoffs. These type of models will the subject of the next few lectures.

1.2 Population Size and Matching

The natural starting point for any learning (or evolutionary) model is the case of fixed players. Typically, we will only look at 2 by 2 games which are played repeatedly between these two fixed players. Each player faces the task of inferring future play from the past behavior of agents.

There is a serious drawback from working with fixed agents. Due to the repeated interaction in every game players might have an incentive to influence the future play of their opponent. For example, in most learning models players will defect in a Prisoner's dilemma because cooperation is strictly dominated for any beliefs I might hold about my opponent. However, if I interact frequently with the same opponent, I might try to cooperate in order to 'teach' the opponent that I am a cooperator. We will see in a future lecture that such behavior can be in deed a Nash equilibrium in a *repeated game*.

There are several ways in which repeated play considerations can be assumed away.

1. We can imagine that players are locked into their actions for quite

a while (they invest infrequently, can't build a new factory overnight etc.) and that their discount factors (the factor by which they weight the future) is small compared that lock-in length. It them makes sense to treat agents as approximately myopic when making their decisions.

2. An alternative is to dispense with the fixed player assumption, and instead assume that agents are drawn from a large population and are randomly matched against each other to play games. In this case, it is very unlikely that I encounter a recent opponent in a round in the near future. This breaks the strategic links between the rounds and allows us to treat agents as approximately myopic again (i.e. they maximize their short-term payoffs).

2 Cournot Adjustment

In the Cournot adjustment model two fixed players move sequentially and choose a best response to the play of their opponent in the last period.

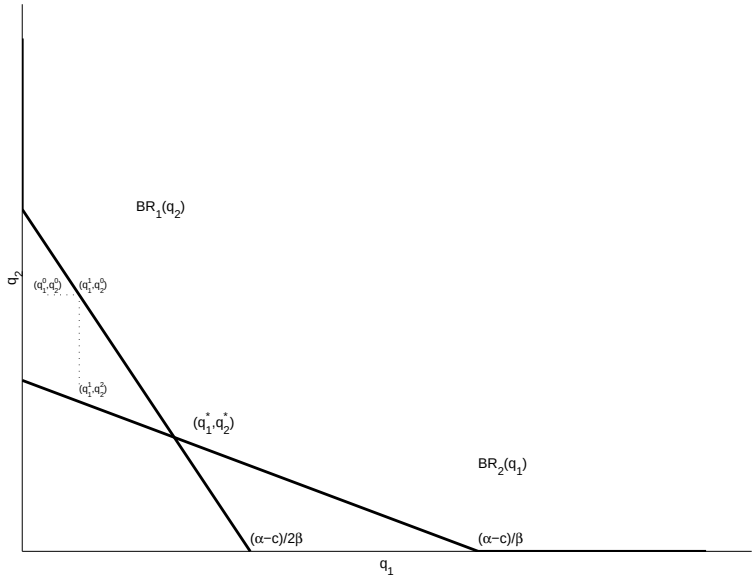
The model was originally developed to explain learning in the Cournot model. Firms start from some initial output combination (q_1^0, q_2^0). In the first round both firms adapt their output to be the best response to q_2^0. They therefore play $(BR_1(q_2^0), BR_2(q_1^0))$.

This process is repeated and it can be easily seen that in the case of linear demand and constant marginal costs the process converges to the unique Nash equilibrium. If there are several Nash equilibria the initial conditions will determine which equilibrium is selected.

2.1 Problem with Cournot Learning

There are two main problems:

- Firms are pretty dim-witted. They adjust their strategies today as if they expect firms to play the same strategy as yesterday.
- In each period play can actually change quite a lot. Intelligent firms should anticipate their opponents play in the future and react accordingly. Intuitively, this should speed up the adjustment process.



Cournot adjustment can be made more realistic by assuming that firms are 'locked in' for some time and that they move alternately. Firms 1 moves in period 1,3,5,... and firm 2 moves in periods 2,4,6,.. Starting from some initial play (q_1^0, q_2^0), firms will play (q_1^1, q_2^0) in round 1 and (q_1^1, q_2^2) in round 2. Clearly, the Cournot dynamics with alternate moves has the same long-run behavior as the Cournot dynamics with simultaneous moves.

Cournot adjustment will be approximately optimal for firms if the lock-in period is large compared to the discount rate of firms. The less locked-in firms are the smaller the discount rate (the discount rate is the weight on next period's profits).

Of course, the problem with the lock-in interpretation is the fact that it is not really a model of learning anymore. Learning is irrelevant because firms choose their optimal action in each period.

3 Fictitious Play

In the process of fictitious play players assume that their opponents strategies are drawn from some stationary but unknown distribution. As in the Cournot adjustment model we restrict attention to a fixed two-player setting. We also assume that the strategy sets of both players are finite.

In fictitious play players choose the best response to their assessment of their opponent's strategy. Each player has some exogenously given weighting function $\kappa_i^0 : S_{-i} \to \Re^+$. After each period the weights are updated by adding 1 to each opponent strategy each time is has been played:

$$\kappa_i^t(s_{-i}) = \begin{cases} \kappa_i^{t-1}(s_{-i}) & \text{if } s_{-i} \neq s_{-i}^{t-1} \\ \kappa_i^{t-1}(s_{-i}) + 1 & \text{if } s_{-i} = s_{-i}^{t-1} \end{cases}$$

Player i assigns probability $\gamma_i^t(s_{-i})$ to strategy profile s_{-i}:

$$\gamma_i^t(s_{-i}) = \frac{\kappa_i^t(s_{-i})}{\sum_{\tilde{s}_{-i} \in S_{-i}} \kappa_i^t(\tilde{s}_{-i})}$$

The player then chooses a pure strategy which is a best response to his assessment of other players' strategy profiles. Note that there is not necessarily a unique best-response to every assessment - hence fictitious play is not always unique.

We also define the empirical distribution $d_i^t(s_i)$ of each player's strategies as

$$d_i^t(s_i) = \frac{\sum_{\tilde{t}=0}^{t} I^{\tilde{t}}(s_i)}{t}$$

The indicator function is set to 1 if the strategy has been played in period $\tilde{t}$ and 0 otherwise. Note, that as $t \to \infty$ the empirical distribution d_j^t of player j's strategies approximate the weighting function κ_i^t (since in a two player game we have $j = -i$).

Remark 1 *The updating of the weighting function looks intuitive but also somewhat arbitrary. It can be made more rigorous in the following way. Assume, that there are n strategy profiles in S_{-i} and that each profile is played by player i's opponents' with probability $p(s_{-i})$. Agent i has a prior belief according to which these probabilities are distributed. This prior is a Dirichlet distribution whose parameters depend on the weighting function. After each round agents update their prior: it can be shown that the posterior belief is again Dirichlet and the parameters of the posterior depend now on the updated weighting function.*

3.1 Asymptotic Behavior

Will fictitious play converge to a Nash equilibrium? The next proposition gives a partial answer.

Proposition 1 *If s is a strict Nash equilibrium, and s is played at date t in the process of fictitious play, s is played at all subsequent dates. That is, strict Nash equilibria are absorbing for the process of fictitious play. Furthermore, any pure-strategy steady state of fictitious play must be a Nash equilibrium.*

Proof : Assume that $s = (s_i, s_j)$ is played at time t. This implies that s_i is a best-response to player i's assessment at time t. But his next period assessment will put higher relative weight on strategy s_j. Because s_i is a BR to s_j and the old assessment it will be also a best-response to the updated assessment. Conversely, if fictitious play gets stuck in some pure steady state then players' assessment converge to the empirical distribution. If the steady state is not Nash players would eventually deviate.

A corollary of the above result is that fictitious play cannot converge to a pure steady state in a game which has only mixed Nash equilibria such as matching pennies.

	H	T
H	1,−1	−1,1
T	−1,1	1,−1

Assume that both players have initial weights weights (1.5,2) and (2,1.5). Then fictitious play cycles as follows: In the first period, 1 and 2 play T, so the weights the next period are (1.5,3) and (2, 2.5). Then 1 plays T and 2 plays H for the next two periods, after which 1's weights are (3.5,3) and 2's are (2,4.5). At this point 1 switches to H, and both players play H for the next three periods, at which point 2 switches to T, and so on. It may not be obvious, but although the actual play in this example cycles, the empirical

distribution over each player's strategies are converging to $\left(\frac{1}{2}, \frac{1}{2}\right)$ - this is precisely the unique mixed Nash equilibrium.

This observation leads to a general result.

Proposition 2 *Under fictitious play, if the empirical distributions over each player's choices converge, the strategy profile corresponding to the product of these distributions is a Nash equilibrium.*

Proof : Assume that there is a profitable deviation. Then in the limit at least one player should deviate - but this contradicts the assumption that strategies converge.

These results don't tell us when fictitious play converges. The next theorem does precisely that.

Theorem 1 *Under fictitious play the empirical distributions converge if the stage has generic payoffs and is 2 2, or zero sum, or is solvable by iterated strict dominance.*

We won't prove this theorem in this lecture. However, it is intuitively clear why fictitious play observes IDSDS. A strictly dominated strategy can never be a best response. Therefore, in the limit fictitious play should put zero relative weight on it. But then all strategies deleted in the second step can never be best responses and should have zero weight as well etc.

3.2 Non-Convergence is Possible

Fictitious play does not have to converge at all. An example for that is due to Shapley.

	L	M	R
T	0,0	1,0	0,1
M	0,1	0,0	1,0
D	1,0	0,1	0,0

The unique mixed NE of the game is $s_1 = s_2 = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$.

If the initial play is (T, M) then the sequence becomes (T, M), (T, R), (M, R), (M, L), (D, L), (D, M), (T, M).

One can show that the number of time each profile is played increases at a fast enough rate such that the play never converges. Also note, that the diagonal entries are never played.

3.3 Pathological Convergence

Convergence in the empirical distribution of strategy choices can be misleading even though the process converges to a Nash equilibrium. Take the following game:

	A	B
A	0,0	1,1
B	1,1	0,0

Assume that the initial weights are $\left(1, \sqrt{2}\right)$ for both players. In the first period both players think the other will play B, so both play A. The next period the weights are $\left(2, \sqrt{2}\right)$, and both play B; the outcome is the alternating sequence (B, B), (A, A), (B, B), and so on. The empirical frequencies of each player's choices converge to 1/2, 1/2, which is the Nash equilibrium. The realized play is always on the diagonal, however, and both players receive payoff 0 each period. Another way of putting this is that the empirical joint distribution on pairs of actions does not equal the product of the two marginal distributions, so the empirical joint distribution corresponds to correlated as opposed to independent play.

This type of correlation is very appealing. In particular, agents don't seem to be smart enough to recognize cycles which they could exploit. Hence the attractive property of convergence to a Nash equilibrium can be misleading if the equilibrium is mixed.